Energies and pressures in viruses: contribution of nonspecific electrostatic interactions†

Antonio Šiber,[*a,c], **Anže Lošdorfer Božič**[b] **and Rudolf Podgornik**[b,c,d]

We summarize some aspects of electrostatic interactions in the context of viruses. A simplified but, within well defined limitations, reliable approach is used to derive expressions for electrostatic energies and the corresponding osmotic pressures in single-stranded RNA viruses and double-stranded DNA bacteriophages. The two types of viruses differ crucially in the spatial distribution of their genome charge which leads to essential differences in their free energies, depending on the capsid size and total charge in a quite different fashion. Differences in the free energies are trailed by the corresponding characteristics and variations in the osmotic pressure between the inside of the virus and the external bathing solution.

1 Introduction

Viruses are abundant and ubiquitous[1] and it is possible that there are no forms of life immune to the effect of viruses[2], which may be advantageous in the fight against disease[3–5]. It appears that there are even viruses that initiate their "lifecycle" exclusively in combination with some *other* viruses, often "stealing" the protein material of those viruses and diverting the cellular processes they initiated to their own advantage[6,7] - the viruses are thus parasites even of their own kin. Although we know the exact nature (the shape and the genome) of only about a hundred[8] viruses, it seems that they are almost as diverse as life itself, so that they represent a type of index to the library of life forms that they parasitize upon. The viruses are indeed only abbreviated, indexed, crippled representation of life and they can hardly be classified as life. They are most often viewed by physicists (or "physical virologists"[9–12] *) as hetero-macromolecular complexes, i.e. complexes of viral proteins and the genome molecule (DNA or RNA) that are reasonably stable in extra-cellular conditions and that initiate a complicated sequence of molecular interactions and transformations once they enter a suitable cell[13]. According to such a view, a virus must in its structure somehow "encode" the crucial steps of its replication process. For example, the proteins that make its protective shell (virus capsid) must have such geometric and chemical characteristics as to activate the appropriate receptors on the cell membrane so that they can attach to and penetrate its interior[14,15]. The reverse of this process when mature viruses are released from the infected cell implies membrane adsorbtion and budding[16,17] to a large extent promoted by electrostatic interactions[18]. The virus needs to be sufficiently stable in the extracellular conditions, yet sufficiently unstable once it enters the cell, so that it can disassemble and deliver its genome molecule to the cellular replication machinery[19]. Once it fulfills walking this tight rope of incipient instability, the manufacturing of virus components in the cell proceeds, leading eventually to new viruses[20]. It is certainly of interest to elucidate the nature of interactions in viruses that enable it to function "between a rock and a hard place", equilibrating on the border of stability, not just from a fundamental scientific point of view but also technologically[21,22].

In this review, we shall concentrate on the description of virus structure in terms of the electrostatic interactions, i.e. we shall be interested in the corresponding *energies* of single viruses (e.g. the energy required to assemble a virus from its constituents) and osmotic *pressures* acting in a virus, though electrostatic interactions are just as important for understanding the interactions *between* the viruses[23]. In fact, the relevance of strong electrostatic interactions for the stability of tobacco mosaic virus (TMV) gels was invoked already in the seminal work of Bernal and Fankuchen[24†], while the Poisson-

---

† They refer to electrostatic interactions as being "probably due to the ionic atmospheres surrounding [viruses]".

Botzmann theory of electrostatic interactions was applied to the case of viruses[25] soon after its publication[26] and even before it was applied to lipid membranes[27].

We shall examine the relation of the formation energy and osmotic pressure of a virus to its structure. Our emphasis is on the electrostatic interaction, first because it contributes a significant part to overall energetics of the viruses[28–31] but also because the electrostatic part of the free energy is the part that can be calculated with a fair amount of precision[32]. Subtler properties of viruses, such as the nature and extent of the ordering of packed DNA/RNA molecule[33,34] may depend on other physical effects, such as genome molecule entropy, yet an important contribution to the energy of protein-genome packaging is of electrostatic nature[35].

The aim of this work is to show the usefulness of elementary physical concepts describing electrostatic interactions as they apply to viruses[36]. Although we will explain some of the intricacies involved in the modeling of viruses, our emphasis will be on the essential physics involved. Although the simple (and necessarily approximate) expressions that we shall expose and employ in this review can be derived from previous, more elaborate publications, we intend to use mostly scaling-based reasoning in their derivation and application in the context of viruses.

Electrostatic interactions in the context of viruses are complex and bring together various theoretical approaches extending from the theory of polyelectrolytes[38,39] and then all the way to the theory of highly charged Coulomb fluids[40,41], sometimes exploiting elaborate theoretical concepts and formally demanding approaches. In this review we shall thus try to retain only most easily understood and applicable concepts[42], yet sufficiently reliable when it comes to description of biological system in general and viruses in particular[43,44]. This approach necessarily excludes the details of some of the more arcane aspects of the theory of electrostatic interactions in aqueous solutions (non-linear salt screening, effects of polyvalent counterion correlations, overcharging by the mobile charge and similar[45]). We do explain the essence of these aspects and their possible relevance in context of viruses in Section 7.

We shall emphasize the *non-specific* aspects of the contributions electrostatic interactions and will thus avoid a detailed exposition of the ion-specific effects[46,47]. This means that we do not deal with chemical specificity of different ions that may drastically modify electrostatic interactions. These effects may be of importance for initiation of assembly, or for speeding up of the assembly, but they are not of primary interest to this review. Our point of view concerns the more robust aspects of virus energetics that can be understood in a sufficiently generic (and simple) physical framework.
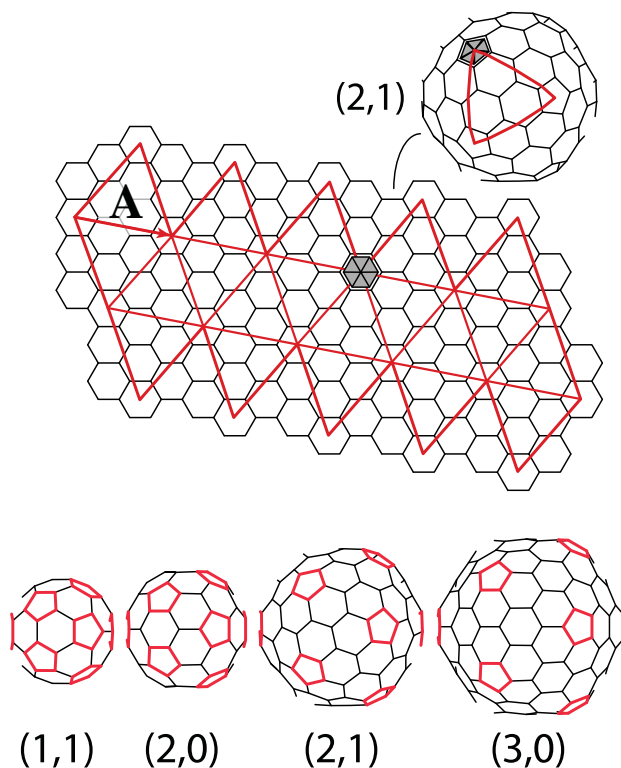

Fig. 1 The geometry behind the Caspar-Klug (CK) classification of viruses. Icosahedral viruses that obey the CK principle can be "cut out" of the lattice of protein hexamers, as shown in the figure. Upon folding of the cut-out piece, twelve of protein hexamers are transformed in pentamers. The CK viruses are described with two integers, $h = 2$ and $k = 1$) in the case shown, which parametrize the vector **A**. The T-number of the capsid is related to h and k as $T = h^2 + hk + k^2$, and the number of protein subunits is $60T$; see Refs.[6,37] for details. The bottom row of images displays the CK structures with $T = 3, 4, 7,$ and 9 (from left to right).

2 A simple description of a virus

The emphasis of this review is not on the symmetry and shapes of viruses - there are already many good reviews on this subject (see e.g. Ref.[6]). For our purposes it is enough to state that all viruses are made of two essential parts: protein coating or a *capsid* and viral genome (of DNA or RNA type) situated in the capsid interior. There are also viruses that in addition to these two essential components need an additional "wrapper", i.e. a piece of cellular membrane, in order to function properly and fuse with the cellular membrane surface[48]. These viruses are referred to as *enveloped* (in contrast to *non-enveloped* viruses which do not have a membrane coating). Because of severe restrictions on the length of their genome encoding the viral shell proteins, the virus capsids are made of *many copies* of one or at most a few types of proteins which are arranged in a highly symmetrical manner as first proposed in the seminal

work by Crick and Watson [49].

Nearly spherical viruses, also called *icosahedral viruses*, show mostly but not always [50] icosahedral order and the proteins that make them can be arranged in the clusters of five (pentamers) or six (hexamers; see Fig. 1 for details and definition of Caspar-Klug classification) *. This arrangement may be only conceptual, but may also have a physical meaning that the interactions in clusters (capsomeres) are somewhat stronger than the interactions between the clusters [6,51]. Crick and Watson surmise that nearly all viruses can be classified either as nearly spherical, i.e. of icosahedral symmetry, or elongated of helical symmetry [49].

Icosahedral viruses tend to look more polyhedral when larger [6,52,53]. There are also non-icosahedral viruses that do not fit in CK classification. Here are some prominent examples: Capsids of some bacteriophages (viruses that infect only bacteria) are "elongated" (prolate) icosahedra [54], i.e. the icosahedral sides around the equator are not equilateral, but isosceles triangles [55]. Capsids of some plant viruses (e.g. tobacco mosaic virus) are (open and hollow) cylinders and their genome molecule is situated in the empty cylindrical space formed by proteins. HIV virus is also non-icosahedral, but is not an elongated icosahedron [56]. Its capsid typically looks conical, being elongated and narrower on one side [57]. Furthermore, even when the viruses are spherical, it is sometimes difficult to classify them according to CK scheme and the typical pentamer-hexamer ordering is not evident [58,59]. Some viruses are multi-layered, i.e. they consist of several protein capsids each of which may be built from different protein [60]. Each of these capsid layers may individually conform to the CK principle [61]. An alternative to CK classification has recently been proposed that apparently contains the CK shapes as the subset of all possible shapes, including those that do not show a clear pentamer-hexamer pattern [58,59].

There is a certain universality in the size of capsid proteins. By analyzing more than 80 different viruses (with T numbers from 1 to 25), we have found that the area of a protein in a capsid is fairly conserved and amounts to ~ 25 nm^2 [62]. The thickness of the protein, i.e. the thickness of the virus capsid in question varies more, but is typically in the interval ~ 2 - 5 nm. The "typical" virus protein can thus be imagined as a disk / cylinder of radius ~ 3 nm and thickness ~ 3 nm. In some viruses these "disks" have positively charged protein "tails" that protrude in the capsid interior and whose role is to bind to a a negatively charged genome molecule (typically ssRNA; see Sec. 4 and Ref. [63]). There is some universality in the distribution of charges along and within the virus capsid, Fig. 2 ‡.

While Caspar-Klug dipoles, corresponding to a bimodal distribution of positive charges on the hypotopal side and negative charges on the epitopal side of the capsid can be observed [65], and we recognize them also for the CCMV capsid in Fig. 2, this is certainly not a rule and other, e.g. monomodal, distributions can be observed just as well. The in-plane angular distribution of charges along the capsid thickness also shows complicated variations within the constraints of the icosahedral symmetry group [62], see Fig. 2. Last but not least, the magnitude of the charges on the surface of the capsomeres is regulated by the dissociation equilibrium while for the buried charges it would have to be estimated from quantum chemical calculations [66,67].

Thus far, we have talked about the order of virus proteins and we have said nothing about the distribution and ordering of the virus DNA or RNA molecule inside the capsid. This will be discussed in the following sections. The virus genome molecule codes for the proteins of the capsid, but also for other proteins needed in the process of virus replication, depending on a virus in question. ssRNA viruses need to code for protein that replicates the virus ssRNA and some viruses also encode the regulatory proteins that are required for correct assembly (scaffolding proteins, see e.g. Ref. [68]) and the proteins required for release of viruses from the infected cell [14]. The amount of information that is required constrains the length of the genome molecule from below.

3 Self-assembling viruses and their energies

Many viruses can self-assemble [69] though the details of the assembly pathways are seldom well understood [70]. This means that the "ingredients" for a virus, individual proteins and its genome molecule, can spontaneously form (assemble in) closed, functional viruses, even outside the cellular environment, in conditions of appropriate pH factor and salt con-

---

* There are also viruses with icosahedral order, yet containing only pentamers. Such is the polio virus [6]

‡ The iso-surfaces obtained here were calculated by first assigning one elementary positive charge to charge residue of *each* lysine and arginine amino acid and protein N-termini, one negative charge to charge residue of each aspar-

---

tic acid and glutamic acid and protein C-termini, and 0.1 positive charges to histidine charge residues (this corresponds to the histidine fractional charge at neutral pH). We thus obtained the 3D coordinates of charges in the capsid, yet this discrete data does not give a visually clear insight in the three-dimensional nature of the charge distribution. To obtain an informative visual representation, we used this data to construct spatial scalar fields of positive and negative charge distributions (separately), by assigning to each charge residue a scalar density, $\tau(r) = |q/e_0| \left\{ 1 - [\min(r, W)/W]^2 \right\}^2$, where q is the charge of the residue, e_0 is the elementary charge, r is the distance from the charge residue, and W is the parameter specifying the extent of the density. The total density field was constructed by summing the scalar densities of all the charges (separately for positive and negative distributions). From the thus obtained space field, an isosurface showing the value of the field at t is plotted. The objects obtained in this way are known in the computer graphics community as metaballs or blobs [64]. In Figs. 2 and 5, $W = 1.34$ nm and $t = 0.85$. Amino acids in the protein interior, i.e. in the capsid wall may not carry charge, as we have assumed, but due to the nature of graphical representation we have chosen, panels a) and b) of Fig. 2 (but not panel c) would ad equately represent the (surface) charge distributions also in the case when only amino acids sufficiently close to the capsid surface were charged.
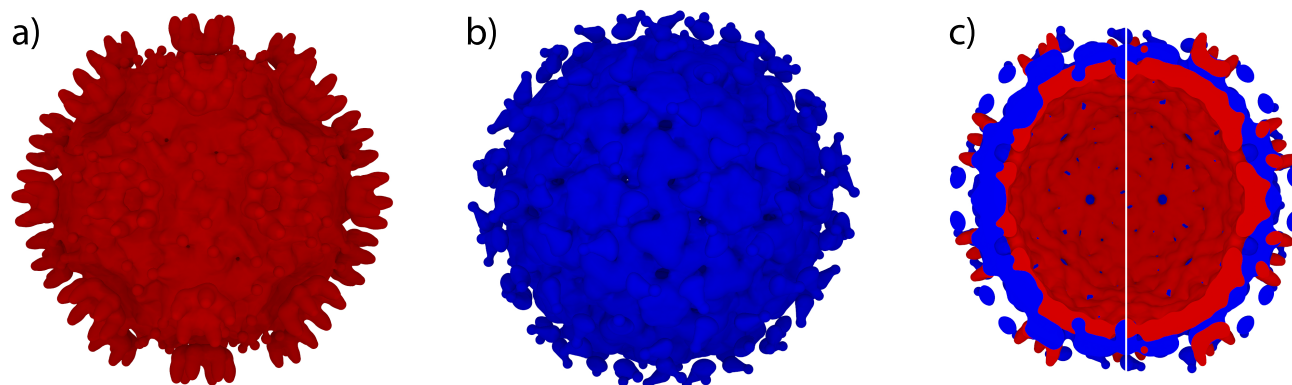
Fig. 2 The calculated representation of the charge distribution on the capsid of cowpea chlorotic mottle (ssRNA) virus as explained in the text, based on the 1CWP entry in the RCSB Protein Data Bank: a) The isosurface of positive charge (red). b) The isosurface of negative charge (blue). c) The combined isosurfaces of positive and negative charges shown in the capsid cut in half so that its interior is seen. On the lef-hand side of image in panel c) (the left of the white vertical line), the (cut) isosurface of negative charge (blue) is translated infinitesimally closer to the viewer, while it is the opposite on the right-hand side of the image.

centration. This is typical for viruses that contain ssRNA molecule and it was first demonstrated in tobacco mosaic virus by Fraenkel-Conrat and Williams[71]. They were able to produce infectious virus particles by simply mixing two solutions, one of them containing only virus proteins, and the other virus RNA molecule.

The process of self-assembly can proceed spontaneously only if the free energy (F) of the assembled virus is lower than the free energy of the disassembled state. This means that, some viruses at least, can be viewed as thermodynamically optimal structures, i.e. they represent minima of the free energy[72].

Upon assembly of proteins and RNA in a virus, their entropy (S) decreases, so the process can not proceed spontaneously unless there is a gain in the internal energy (U) of the system, i.e. there is a favorable "binding" energy of ingredients once they form a virus. Even then, the process of assembly can proceed only when the concentration of the ingredients is sufficiently large, i.e. above a critical concentration required for assembly[73]. Below that concentration, the entropic contribution to the free energy dominates and although the binding energy (enthalpy) of the ingredients in a capsid is favorable, the net gain in free energy is not, and the proteins and RNA molecules remain in a disassembled state.

Not all viruses can self-assemble. Some types of viruses, even when all the ingredients are available, can form only in the cellular environment, i.e. they require some of the cellular mechanisms for the assembly. The typical requirement is ATP energy which suggests that those viruses do not correspond to the simple free energy minima, but are rather examples of free energy driven structures at some elevated position/plateau in the free energy landscape. The use of ATP energy for assembly is typical for bacteriophage viruses as we shall see in the following.

3.1 Energies and assembly of empty viruses

The viruses are kept together by the same interactions as those governing the "living matter"[5,74,75]. The interactions are thus many different guises of electromagnetic force, sometimes as "direct" interactions between entities (e.g. electrostatic, van der Waals (vdW), steric repulsion interaction,...)[76] and sometimes as "indirect" or effective interactions which manifest as forces but arise only in presence of the bathing medium at finite temperature, T (e.g. hydrophobic or hydration interactions)[77,78]. The direct interactions are also modified by the presence and nature of the bathing medium, yet they survive outside it.

A very important characteristic of virus genome molecules is the *negative charge* they carry, due to dissociation of phosphate groups on RNA and DNA bases[79]. A compactification of a highly negatively charged genome in a capsid interior requires energy. That is why the proteins in contact with the DNA or RNA are often *positively charged*. This may not be the case for the protein as a whole, but it is quite often the case for parts of the surface of proteins in direct contact with the DNA or RNA[63,80] (see Fig. 2; note that the capsid interior is mostly red, i.e. positively charged). These parts of proteins are often spatially extended and they in some cases form highly positive "tails", see next section. The complementary charges on the viral proteins and genome increase the electrostatic interaction of the complex and decrease its energy, enabling an easier assembly. That is the reason the viral genome codes for such proteins, but there are viruses (e.g. polyomavirus)

which utilize *cellular*, very positively charged proteins (histones) to associate them with their genome so as to reduce the electrostatic energy and allow for an easier packing [14]. All this indicates the importance of electrostatic interactions for virus assembly.

The charge on the proteins (and on the DNA/RNA) depends on the pH factor of the solution. Modifications of pH factor, and the corresponding dissociation/saturation equilibrium of ionic bonds in proteins leads to variation of their charge [78,81] and a change in the electrostatic energy of the nucleoprotein complexes, as will be seen in the following. Modification of the pH factor may occur in the cellular environment in endosomes/vesicles i.e. pieces of cellular membrane that the viruses carry along with them as they penetrate the cell. Some viruses assemble in specific subcellular compartments (called "virus factories" [82] or inclusion bodies [14]) which may also have an adequately altered, local pH value.

3.1.1 A simple model of a virus capsid and the corresponding energy.

The simplest representation of the protein charge distribution would be a positive charge homogeneously distributed on a protein. A homogeneous distribution of positive charge will tend to keep proteins apart, yet it has been experimentally observed that proteins of some viruses assemble in (empty) capsids under appropriate conditions. Such is a virus of hepatitis B that has been experimentally studied in considerable detail by A. Zlotnick and coworkers [51]. Since the viral proteins do assemble in empty capsids, there must be an attractive contribution to their interaction. The source of this contribution is a combination of hydrophobic and van der Waals interactions [77], with hydrophobic component playing a dominant role. This conclusion can be read out from the experiments performed by Ceres and Zlotnick [51]. They observed that the strength of protein-protein interactions in capsids *increases* with temperature and this suggests an important entropic contribution to the interaction, hydrophobic interaction being the obvious candidate [65].

To understand the cohesive energy of a virus capsid, we first need to evaluate a seemingly simple problem: obtain the electrostatic self-energy of a uniformly (positively) charged (with surface charge density σ), permeable, infinitely thin sphere of radius R - this is the zeroth level description of a capsid. This problem can be solved on the mean-field level by treating the ions as ideal gas, which adjust to the external potential and contribute to it via their charge density. This is the Poisson-Boltzmann (PB) approach which yields nonlinear differential equation for the electrostatic potential, ϕ [42,76,83]. It can be derived by minimizing the appropriate free energy [84] which has the form

$$F_{PB}[\phi(r), \nabla\phi(r), c^i(r)] = \int f_{PB}(\phi(r), \nabla\phi(r), c^i(r)) d^3r, \quad (1)$$

where the free energy density is given by

$$f_{PB}(\phi(r), \nabla\phi(r), c^i(r)) = -\tfrac{1}{2}\varepsilon_0\varepsilon\nabla\phi(r)^2 + \sum_{i=\pm} e_i c^i(r)\phi(r) +$$
$$+ \sum_{i=\pm} k_B T \left[c^i(r)\ln c^i(r) - c^i(r) - \left(c_0^i \ln c_0^i - c_0^i\right)\right] +$$
$$+ e_0 \rho_p(r)\phi(r). \quad (2)$$

Here $e_0\rho_p(r)$ is the charge density of the capsid, T is temperature, k_B is the Boltzmann constant, c^i are the concentrations of (monovalent) salt ions, with c_0^i their bulk concentrations, $\varepsilon\varepsilon_0$ is the permittivity of water, and e_0 is the electron charge. Minimizing the above free energy w.r.t. $\phi(r), \nabla\phi(r)$ as well as $c^i(r)$, leads to the PB equation of the form

$$-\varepsilon\varepsilon_0\nabla^2\phi(r) = \sum_{i=\pm} e_i c_0^i e^{-\beta e_i \phi(r)} + e_0 \rho_p(r), \quad (3)$$

where $\beta^{-1} = k_B T$ and e_i is the charge of the ions, i.e. $e_i = \pm e_0$.

When the electrostatic potentials in the solution are small, $e_0\beta\phi \ll 1$ and we are dealing with a symmetric system, e.g. 1-1 electrolyte, that has $c_0^+ = c_0^- = c_0$, the PB equation can be linearized yielding the Debye-Hückel (DH) equation for the potential [85], of the form

$$-\nabla^2\phi(r) = \frac{\beta}{\varepsilon\varepsilon_0}\left(\sum_{i=\pm} e_i^2 c_0^i\right)\phi(r) + \frac{e_0\rho_p(r)}{\varepsilon\varepsilon_0} + \dots, \quad (4)$$

where we took into account that the salt is assumed to be uni-univalent and thus $\sum_{i=\pm} e_i c_0^i = 0$. At this point one standardly introduces the inverse (Debye-Hückel) screening length κ^{-1}, with $\kappa^2 = \beta\left(\sum_{i=\pm} e_i^2 c_0^i\right)/\varepsilon\varepsilon_0$ [26,86].

The linearity of DH equation renders it amenable to several ways of solving, including the Green function method [87]. But the simplest way to think of the DH approximation is in terms of the renormalization of electrostatic interactions (in vacuo) by salt ions. The effective interaction between the charges Q_1 and Q_2, separated by $\mathbf{r}_1 - \mathbf{r}_2$ in the solution of monovalent ions (with concentration c_0) of (relative) dielectric constant ε is given by the DH potential of the screened exponential form $U(\mathbf{r}_1 - \mathbf{r}_2)$. The easiest way to obtain the self energy of a uniformly charged sphere in the DH approximation is to sum the pair DH interactions over the sphere surface,

$$
\begin{aligned}
F_{DH} &= \frac{\sigma^2}{2}\int dS_1 \int dS_2\, U(\mathbf{r}_1 - \mathbf{r}_2) = \\
&= \frac{1}{2}\frac{\sigma^2}{4\pi\varepsilon_0\varepsilon}\int dS_1 \int dS_2 \frac{\exp(-\kappa|\mathbf{r}_1 - \mathbf{r}_2|)}{|\mathbf{r}_1 - \mathbf{r}_2|}, \quad (5)
\end{aligned}
$$

where dS_1 and dS_2 are infinitesimal elements of the sphere surface centered around vectors $\mathbf{r}_1$ and $\mathbf{r}_2$, and factor of $1/2$ accounts for double counting of the pair interactions. Since the radii of viruses are typically of the order of 20 nm, and the

DH screening length in the physiological conditions ($c_0 \sim 150$ mM) is $\kappa^{-1} \sim 1$ nm, the limit of $\kappa R \gg 1$ is often implied. The range of integration is effectively cut on the scale of κ^{-1}. In case of interactions on a sphere, this defines a spherical cap. But, when $\kappa R \gg 1$, this spherical cap transforms in a disk, and this renders the two integrations independent:

$$\lim_{\kappa R \gg 1} F_{DH} = \frac{\sigma^2}{4\pi\varepsilon_0\varepsilon}\frac{1}{2}\int dS_1 \int_0^\infty dr_2 \int_0^{2\pi} d\phi_2 \exp(-\kappa r_2). \quad (6)$$

Note that the integration over r_2 can be extended to infinity, since κ in the exponential function acts as a cutoff parameter[87]. The self-energy of the capsid is thus[65]

$$\lim_{\kappa R \gg 1} F_{DH} = \frac{\pi\sigma^2 R^2}{\varepsilon_0\varepsilon\kappa}. \quad (7)$$

We have denoted the self-energy using letter F to indicate that it corresponds to free energy, containing also the entropy of salt ions. One should note the meaning of this quantity: it is the energy required to bring *infinitesimal* charges from infinite separations in the solution to the capsid. When the capsid consists of many weakly charged proteins, one may think of this quantity as being approximately the electrostatic contribution to the *assembly free energy*. To estimate this quantity, we need an information on the surface charge density of the virus proteins. This can be estimated from their amino acid content, and typically $\sigma \sim 1$ e_0/nm^2 [65]. Taking now $R \sim 20$ nm and $c_0 = 100$ mM, we obtain $F_{DH} \sim 10^4$ $k_B T$.

The hydrophobic energy can be estimated from the area of proteins engaged in protein-protein contacts, i.e. from the volume slice around the proteins one water molecules thick. The capsid of 20 nm typically have $T = 3$ triangulation number, i.e. they consist of 180 protein subunits. The length of protein contacts in such a capsid is about 3000 nm. This gives the total area exposed to the protein contacts of about 6000 nm^2 if we assume the capsid thickness of 2 nm. The energy of attractive protein interactions (hydrophobic *and* van der Waals) per unit area of the exposed protein surface is typically of the order of 10 mJ / m^2 [88]. Multiplying the estimated exposed area with this energy, the estimated attractive hydrophobic interaction is obtained as, $F_{HP} \sim 10^4$ $k_B T$. This is of the same order of magnitude as the electrostatic repulsion and one is led to conclude that the interactions tending to dissolve a capsid and those keeping it together are in a tight balance (Ceres and Zlotnick have measured the free energy of the hepatitis B capsid to be $- \sim 1.5 \ 10^3$ $k_B T$ i.e. about 5-6 $k_B T$ per inter-protein contact[51]). The reason for this is most likely the fact that viruses, in addition to being able to assemble, need also to disassemble and deliver their genome molecule to the cell. In simple viruses this may be triggered by variation of pH and ionic concentration in different cell regions which increase the electrostatic repulsion in the capsid. Sufficiently large changes

in pH lead to variation (increase) of protein charge[89], i.e. σ, while reduction of ionic concentration leads to decrease of κ, and as Eq. (7) shows, both effect modify the electrostatic self-energy of the capsid. Ceres and Zlotnick[51] have experimentally demonstrated that the free energy of hepatitis B capsids increases with concentration of mono-valent salt. This may be interpreted as the screening of repulsive protein-protein electrostatic interaction[65,90]. A sufficiently large modification of the electrostatic interaction may lead to enthalpic instability of the capsid and to its disassembly[87]. Experimental studies of virus assembly[51,91,92] clearly show the importance of pH factor and the salinity of the solution for the assembly of complete capsids. For sufficiently large concentration of virus proteins, the assembly depends on the values of both of these parameters[91,92].

3.1.2 Refined models of virus capsids. The simplest approach to capsid electrostatics presented above may be improved in several respects. The easiest one is to examine the complete DH solution to the problem, i.e. in a whole range of κR values, valid in particular for lower ionic concentrations when $\kappa < R$. In fact one can derive a general formula[90]

$$F_{DH} = \frac{2\pi\sigma^2 R^2}{\varepsilon_0\varepsilon\kappa[1 + \coth(\kappa R)]} \quad (8)$$

that describes the capsid electrostatic free energy for any κ within the range of validity of the DH approximation..

A next level of refinement is to abandon the assumption of smallness of electrostatic potential $\beta e_0 \phi \ll 1$ and to solve the PB equation, Eq. 3, for the potential and obtain the capsid energy in this way. This is still not the "exact" solution to the problem, since the PB approach is an approximation of the mean-field genre and it neglects ionic correlations[40]. Yet, the PB approach is more reliable for large surface charge densities and smaller ionic concentration, i.e. in cases where ϕ is not necessarily small. In the limit $\beta e_0 \phi \ll 1$ it of course reduces back to the DH limit. The detailed study and comparison of PB and DH approaches to capsid electrostatics has been performed in Ref.[90]. The PB capsid energies are always smaller from the corresponding DH values, yet the functional dependence of F on σ, R, and c_0 is similar in both approaches, the R^2 dependence in particular. The DH results are very reliable quantitatively when $R \sim 20$ nm, $\sigma < 0.8$ e_0/nm^2 and $c_0 > 50$ mM.

The calculations that adopt a refined representation of the capsid with regards to its finite thickness, δ, have also been performed. The distribution of charge across the thickness of the capsid is such that the positive charges are often concentrated on the capsid interior surface, while, typically negative charges are concentrated on the capsid exterior surface (see Fig. 2). There are also a few but functionally non-negligible[93] net charges embedded in the interior of the capsid proteins[78,94], while most of the charges there are partial

charges due to electronic charge redistribution in chemical bond formation[67,95]. To account for the finite thickness of the capsid, it was treated as a dielectric shell with relative permittivity ε_p, impermeable to ions, with interior and exterior surfaces which are uniformly charged[90] with surface charge densities σ_1 and σ_2, respectively (see Fig. 3). This model shall still prove useful, so we discuss it briefly in the following.
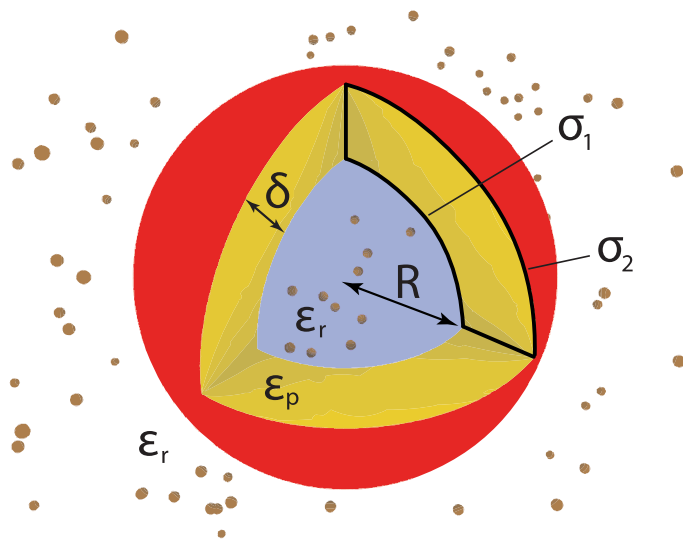

Fig. 3 An illustration of an electrostatic model of a viral capsid with finite thickness, δ. Salt ions are represented by small spheres. The interior of the capsid contains water and salt ions, but the salt ions are not present in the viral capsid shell.

Intriguingly, for this model also, in the regime $\kappa R \gg 1$ the capsid free energy scales with the second power of capsid radius, both in the DH and PB cases[90]. DH expressions for the electrostatic potential and free energy can be obtained analytically, yet they are sufficiently transparent only in certain limits. A limit of interest to us is $\kappa R \gg 1$, $\varepsilon \gtrsim \varepsilon_p$, $\delta \ll R$. In this case,

$$F_{DH}(\sigma_1, \sigma_2, \delta) = 2\pi R^2 \frac{\varepsilon_p (\sigma_1 + \sigma_2)^2 + \varepsilon (\sigma_1^2 + \sigma_2^2) \kappa \delta}{\varepsilon_0 \varepsilon \kappa (2\varepsilon_p + \varepsilon \kappa \delta)}. \quad (9)$$

For viruses in physiological conditions, we may further take $\varepsilon_p \ll \varepsilon$ (for proteins, $\varepsilon_p \sim 5$[78]), and $\kappa \delta \sim 1$, which simplifies Eq. (9) to

$$F_{DH} = \frac{2\pi (\sigma_1^2 + \sigma_2^2) R^2}{\varepsilon_0 \varepsilon \kappa}. \quad (10)$$

By comparing this equation with Eq. 7) one concludes that the qualitative behavior of the capsid free energy is not importantly modified by an introduction of the finite capsid thickness and that the free energy is of the same order of magnitude as in the case of infinitely thin shell.

Angularly nonuniform distribution of the capsid charge (still positioned on a perfectly spherical, infinitely thin capsid), in accordance with its icosahedral symmetry, can be introduced in the electrostatic model. This complicates the treatment and introduces special function series, but the electrostatic free energy still retains the already familiar functional DH behavior in the $\kappa R \gg 1$ limit. In particular, Marzec and Day[96] have obtained that

$$F = \frac{\pi \Sigma_Q^2 R^2}{\varepsilon_0 \varepsilon \kappa}, \quad (11)$$

where Σ_Q^2 is the average square charge density, $\Sigma_Q^2 \equiv (4\pi)^{-1} \int d\Omega \Sigma_Q^2(\Omega)$, and Ω is the spatial angle.

A completely numerical PB approaches are also possible. These start with a determination of the spatial distribution of capsid charge based on the capsid amino acid content and the atomic coordinates determined from X-ray studies[28]. The PB equation is then solved on a three-dimensional grid using advanced numerical routines[28]. Such approaches are, however, less transparent concerning the scaling of energies with various parameters of the system, e.g. virus radius, its total charge, salt concentration, and similar.

4 Energies of ssRNA viruses

Viruses that contain the single-stranded RNA molecule (ssRNA) often self-assemble[97]. A simplified description of the ssRNA molecule in these viruses characterizes it as a generic flexible polyelectrolyte[81,98]. The ssRNA flexibility is described by its effective *persistence length* i.e. the length on which the ssRNA refuses to bend. For flexible polyelectrolytes, this effective length is of the order of monomer (nucleotide) separation ($a \sim 0.5$ nm) and they can thus be well described within the framework of the Edwards - de Gennes *flexible chain model*[99,100]. It is the connectivity of the chain that gives an essential imprint to the behavior of monomers of the chain as opposed to free particles in solution[101]. The long range interactions between monomers, such as Coulomb interactions in the case of ssRNA, are modified in an essential way when coupled to the connectivity of the chain[102]. The main difference between the flexible chain model describing the salient features of the ssRNA chain and the semiflexible chains, such as dsDNA, is that the elasticity of the chain is purely entropic[103] in the former while being enthalpic in the latter case[104,105]. The connectivity of the chain introduces many important features also in the behavior of the chain in external fields as is the case in adsorption to charged surfaces[106] and the consequent bridging interactions present between two apposed charged surfaces[107].

It is an experimental fact[108,109] (and any reasonable theory should account for it) that the ssRNA viruses contain the

ssRNA molecule in a thin shell closely distanced from the interior capsid radius. In the flexible polyelectrolyte theory the thickness of this adsorbed ssRNA polyelectrolyte layer is of the order of a[110]. The thickness of this ssRNA shell results in general from a relatively complicated interplay of all the interaction energies and chain entropy involved in the problem[111]: the *electrostatic contribution* to the interactions between negatively charged ssRNA and positively charged hypotope of the capsid and the *entropic contribution* of the constrained polyelectrolyte to the free energy[112,113].

4.1 Electrostatic interactions and energies of ssRNA viruses

The total free energy in this case is composed of the electrostatic interactions due to ions and charges on the polyelectrolyte, as well as the entropy of the polyelectrolyte chain[102,111]. The electrostatic part of the free energy, F_{es}, contains the PB functional, Eq. (2), augmented by the presence of the charges on the polyelectrolyte chain expressed in terms of the polyelectrolyte monomer concentration, $\rho(r)$ and reads[114]

$$F_{es}[\phi(r), \nabla\phi(r), \rho(r)] = \int f_{es}(\phi(r), \nabla\phi(r), \rho(r))d^3r - \\ - \mu\left(\int d^3r\rho(r) - N\right), \quad (12)$$

where μ is the Lagrange multiplier enforcing the condition of fixed number of monomers, N, of the polyelectrolyte chain, with

$$f_{es}(\phi(r), \nabla\phi(r), \rho(r)) = f_{PB}(\phi(r), \nabla\phi(r)) - pe_0\rho(r)\phi(r). \quad (13)$$

Here $f_{PB}(r)$ was defined in Eq. 2 and pe_0 is the charge per monomer, with e_0 is the electron charge and $0 < p < 1$. The part of the free energy due to the entropy of the flexible polyelectrolyte can be approximated in the so-called *ground state dominance*[98,100] as

$$F_{ent}[\rho(r), \nabla\rho(r)] = k_BT\frac{a^2}{6}\int d^3r\frac{[\nabla\rho(r)]^2}{\rho(r)}. \quad (14)$$

Minimizing the sum of the electrostatic and entropic contributions then leads to a polyelectrolyte PB equation that can be solved numerically in the spherical geometry of the capsid[111].

Numerical solutions of the polyelectrolyte PB theory are complicated but they invariably point to the existence of an adsorption layer next to the internal positively charged wall of the capsid[111]. This leads to the conclusion that ssRNA should be non-uniformly distributed within the capsid, showing a relatively dense surface layer and a depleted core. The existence of the adsorbed layer along the periphery of the capsid then engenders the attractive polyelectrolyte bridging interactions of the type observed to act between planar charged surfaces[107,115] but here act between different parts of the spherical hypotope and thus stabilize the protein shell.

4.1.1 Scaling approach to the ssRNA packing inside a capsid.
A scaling estimate of the electrostatic free energy of ssRNA packing inside a capsid can again be obtained in a suitably simplified framework and in appropriate range of parameters. It has been found that the optimal virus configuration in the physiological regime is such that the total charge on the encapsidated ssRNA is comparable to the charge on the capsid, being equal as the salt concentration decreases[63,90,116]. This is relatively easy to understand. When the charges on the capsid and on the ssRNA molecule are equal (but of the opposite signs), the ssRNA completely screens the protein charges, so that the salt ions almost need not to redistribute at all, especially when the ssRNA and the capsid can be brought in close contact. In all other cases, there is an effective, remaining charge, that the salt ions must screen by rearranging, increasing the total electrostatic energy of the system in this way. This simple argument is complicated by the fact that the ssRNA consists of *connected* charges - it is a polyelectrolyte molecule. This line of reasoning applies only to the total charge on the ssRNA and the capsid. When each nucleotide is assumed to carry a fixed charge, this also fixes the total length of the ssRNA molecule. There are studies, however, which predict that not all ssRNA bases carry an elementary charge[63]. This depends on the dissociation equilibrium and charge regulation[117], but it does not importantly influence simple energy estimates to be presented below. Similar arguments can be used also for encapsidated polyelectrolyte cargo different from ssRNA[118,119].

The ssRNA configuration inside the capsid as obtained from the full polyelectrolyte PB theory that contain also the polyelectrolyte entropy can be approximated by two concentric spherical shells with opposite charge, $\sigma \equiv \sigma_2 \approx -\sigma_1$, separated by $\sim a$ (see Fig. 4). Thus, Eq. (9) should be of use to estimate the free energy of such a configuration. The important difference of the ssRNA virus with respect to the dielectric slab model of the capsid is that the space between the ssRNA and the proteins is permeable to salt and water. However, when $a \lesssim \kappa^{-1}$, the salt induced screening of ssRNA-capsid protein electrostatic interaction is incomplete, and the free energy of salt ion distribution in between the ssRNA and the capsid is small compared to the analogous contributions in the rest of the space.

Thus, a simplest estimate of the electrostatic complexation free energy (ΔF_C) of a ssRNA virus in the DH approximation can be obtained from Eq. (9), using $\varepsilon_p \approx \varepsilon$ and $\kappa\delta \sim 1$ (in physiological conditions) and subtracting the electrostatic

self-energy of the capsid, Eq. 7. This yields

$$\Delta F_C \approx F_{DH}(-\sigma,\sigma,a) - F_{DH}(0,\sigma,0) \approx \frac{\pi\sigma^2 R^2}{\epsilon_0\epsilon}\left(fa - \kappa^{-1}\right),$$
(15)

where f is a numerical factor between 4/3 and 2 (it depends on κ and δ). Note here that the sign of this free energy difference depends on the competition between two length scales: a and κ^{-1}.

The above result is not very accurate in high salt concentrations, when $\kappa^{-1} < a$ and does not include the non-electrostatic self-repulsion contribution of the polyelectrolyte which can be shown to be smaller then the leading electrostatic contribution[90]. Using the "typical" virus parameters and $a = 0.5$ nm, we obtain $\Delta F_C \sim 0$, since $fa - \kappa^{-1} \sim 0$. A more detailed calculation[111] gives ΔF_C which is negative and about a quarter of the free energy of an empty capsid (at physiological conditions) $F_{DH}(0,\sigma,0)$, consistent with lower bound on f. This signifies that spontaneous encapsidation of ssRNA is a delicate process that may even be suppressed in thermodynamical equilibrium, so that only *empty* capsids form. Indeed, formation of empty capsids and free ssRNA was theoreticaly predicted in the regime of high ionic concentrations, but also when the charge on the ssRNA is larger than the charge on proteins by a factor of ~ 2 or more[111]. In physiological conditions, and when the charge on the ssRNA is about the same as the charge on the proteins (but of the opposite sign), it was found that the complexation free energy is negative so that ssRNA encapsidation takes place.

4.1.2 Details of the energetics and conformation of ss-RNA inside viral shells.

A more detailed physical model of an ssRNA virus should produce the ssRNA distribution *as a result* of a free energy minimization. Different types of models of the ssRNA virus packing have been examined some of them emphasizing the discrete nature of the ssRNA[116], and some, as we have seen in the previous section, representing the ssRNA in the continuum limit via the density field $\rho(\mathbf{r})$ describing the spatial distribution of RNA "monomers"[63,111,120]. Both approaches have their advantages and limitations and they necessarily simplify the physics of ssRNA to some manageable model. Going into details of these studies is beyond the scope of this review.

It is of interest, however, to discuss the typical spatial distributions that are obtained in these models. On panel b) of Fig. 4 we show the ssRNA monomer density for several values of its length. The results are from Ref.[111]. One sees that, indeed, the ssRNA molecule occupies a shell closely separated from the capsid interior surface. As the ssRNA molecule becomes longer, and its charge larger from the capsid charge, the interior or the capsid gradually fills up. This effect is more pronounced for higher concentrations of (mono-valent) salt. The localization of ssRNA in a shell close to the capsid was also

found in Ref.[116], in a Brownian dynamics study of ssRNA (generic polymer) virus assembly where the electrostatic interactions were modeled via effective pairwise potentials[121], and in molecular dynamics simulation of assembled Satellite Tobacco Mosaic virus in Ref.[122].
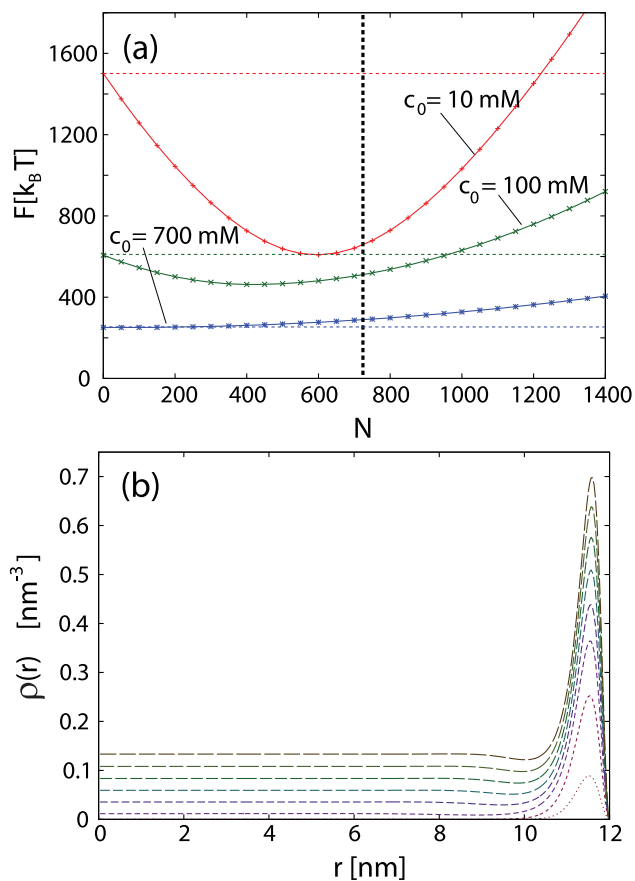

Fig. 4 Panel a): The free energies of the capsid-ssRNA complexes as a function of the number of RNA bases and for three different monovalent salt concentration, as indicated. The free energy of the *empty* capsid is obtained when $N = 0$. Panel b): The ssRNA ($a = 0.5$ nm) concentration profile in a capsid of radius $R = 12$ nm and surface charge density $\sigma = 0.4\ e_0/nm^2$ for $c_0=100$ mM. The curves displayed correspond to $N =$100, 300, 500, 700, 900, 1100, 1300, and 1500 bases. The lines are styled so that the length of their dashes is proportional to N. See Ref.[111] for details.

It was found that the energetics of ssRNA viruses also depends on the details of spatial distribution of the capsid charge, $e_0\rho_p(r)$, in particular on its delocalization on the capsid protein tails that protrude into virus interior[111] (see Fig. 5) [§].

---

§ Here we see the importance of the electrostatic interactions that is directly reflected in the shape of a virus. Furthermore, they probably underlie and influence the evolutionary pathway of a particular virus, whose evolutionary fitness is to a certain extent determined by the electrostatic interactions that govern its (dis)assembly.
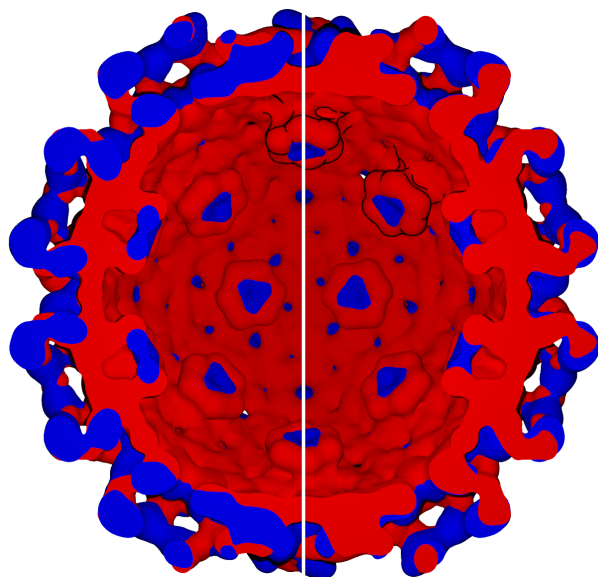
Fig. 5 One half of the cucumber mosaic (ssRNA) virus capsid (strain FNY). The image was constructed from RCSB Protein Databank entry 1F15 in the same manner as the one panel c) of Fig. 2. One can observe "buttons" of positive charge density formed by the tails of capsid proteins protruding into the interior of the virus.

The viruses that have this feature were found to be more stable with respect to changes in the ionic concentration, so that the length of the ssRNA that is encapsidated in the thermodynamically optimal conditions does not vary significantly with the salt concentration[63,111]. Figure 6 shows the free energies of capsid-ssRNA complexes (panel a) and the profiles of the ssRNA density (panel b) obtained in the model that, in contrast to results shown in Fig. 4, accounts for the delocalized capsid charge, i.e. the effects of protein tails. The free energies do not contain the part related to attractive protein-protein interaction, so that the empty capsid electrostatic free energies (to be compared with e.g. Eq. (7)) can be read out from the figures as the free energies of the complex in the limit when number of ssRNA monomers goes to zero. One sees that the optimal encapsidated ssRNA length (positions of the minima in the complex free energy curves as a function of the ssRNA length[¶] in the case of infinitely thin capsid) is practically fixed at the position where the total ssRNA charge equals the capsid charge (vertical thick dashed line in panel a) of Fig. 6), almost irrespectively of the concentration of monovalent salt. One also sees that the thickness of the ssRNA shell becomes influenced by the length of the charged protein N-tails, in addition to its dependence on the screening length and monomer size (compare Figs. 4 and 6).

The optimal encapsidated ssRNA length was found to vary

[¶] The free energies at these points are estimated by Eq.(15).

much more in the case of infinitely thin capsid (panel a) of Fig. 4), so that at salinities above about 700 mM, the thermodynamically optimal ssRNA length drops down to zero and the model predicts formation of empty viral shells only[111]. In sufficiently low monovalent salt concentrations (10 mM - 100 mM), it was found that viruses containing the ssRNA that has about two times more charges than the capsid, are still able to self-assemble (i.e. that their total free energy is smaller than the free energy of assembled empty capsids and free-floating ssRNA, see Fig. 6).
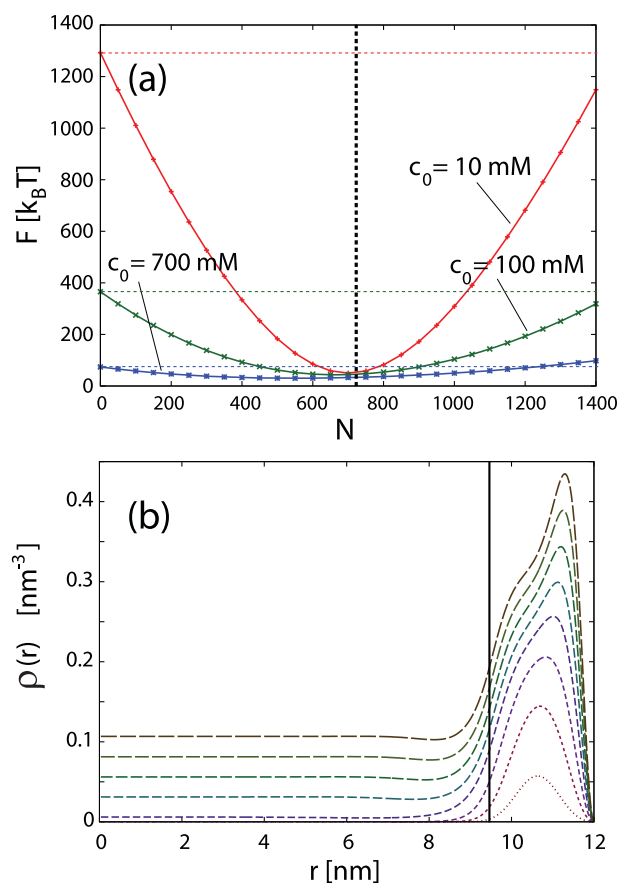

Fig. 6 Panel a): The free energy of the capsid-ssRNA complex as a function of the total number of ssRNA bases. The capsid radius is R=12 nm. The capsid charge is distributed in a spherical shell of 2.5 nm thickness. Panel b): The ssRNA density profiles for N =100, 300, 500, 700, 900, 1100, 1300, and 1500. A vertical line at r=9.5 nm indicates the position of the N-termini of the capsid proteins, i.e. the interior surface of the charged protein shell.

This was confirmed in two quite different studies, one emphasizing the continuum aspects of the ssRNA[111] (the total charge on the ssRNA allowed thermodynamically and found in this study is about twice the charge on the capsid), and the other its discrete features[116] (there the total charge on the ssRNA allowed thermodynamically is about three times the

charge on the capsid). This means that it is possible that the viruses carry effective charge, although the most stable state of an ssRNA virus is the one where the charges on the ssRNA and the capsid are about the same, i.e. when the effective charge is close to zero - this was found both in continuum approaches in Refs. [63,111] and in the approach emphasizing ssRNA discreteness in Ref. [116].

All the studies discussed thus far agree that the ssRNA packages into a shell in the immediate vicinity of the inner capsid wall. There is also an interesting question of its ordering [123]. Being very flexible, ssRNA is much more difficult to order even in a confined space then the much stiffer DNA [124,125]. It seems likely that the spatial inhomogeneity of the capsid charge, obeying the icosahedral symmetry of the capsid, may be responsible for ordering of the ssRNA molecule [120] that has been observed in several experiments [108,109], but also in simulations in Ref. [121]. It is also of interest to note that the ssRNA may locally form double stranded RNA hairpins, increasing the binding energy in this way, both via the increase of packing density and occurrence of inter-base hydrogen bonding.

The calculations we presented do not take into account the inextensibility of ssRNA, i.e. the fact that the RNA monomers are linked together by chemical bonds that can not be stretched beyond their chemical size [126]. In cases of sufficiently long and flexible ssRNA, this fact should not influence the energy significantly, however, it may be of importance for smaller ssRNA lengths. In such cases, not all monomer density distributions predicted by continuum calculations can be realized by connected monomers, and one has to explicitly account for inextensibility as was done in Ref. [127], by introducing the maximal extensibility constraint in the free energy functional.

5 Energies of bacteriophages

The formation of bacteriophages in an infected bacterium proceeds by the formation of empty protein capsids first. The DNA is then inserted into a (pro)capsid [14], an entropically and enthalpically unfavorable process [128] that can be accomplished only via ATP-driven molecular motor [129]. 2 to 2.5 base pairs of DNA are packaged by one ATP molecule and there appear to be strong electrostatic interactions between the DNA phosphate backbone and the positively charged internal molecular motor wall [130]. It is not entirely clear whether there are also significant electrostatic capsid protein-DNA interaction as we have seen in ssRNA viruses ‖ but is an experimentally ob-

served fact that DNA molecule in the bacteriophages is nearly homogeneously distributed in their interior, so that its density is quite uniform [131,132]. This is an important difference in charge distribution with respect to the ssRNA viruses, and the reason for quite different energetics of the two types of viruses. The region of DNA close to the capsid shows an onion-like ordering, while the DNA in the capsid center is typically less ordered [133]. The outer layers most likely have a toroidal, spool-like geometry [124], but there are other suggestions proposing a liquid crystal [134], a folded toroid [135,136], a folded coaxial spool [137] and the spiral fold [138] configurations.

The physical description of the ordered packing of the dsDNA inside the bacteriophage capsid is usually captured in one of the variants of the *inverse spool* model first invoked in the study of the T2 bacteriophage [139–141] and formalized by Grosberg et al. [142]. The inverse spool model has been subsequently refined by Odijk and Gelbart and coworkers [143–150] and is based on the decomposition of the dsDNA total energy within the viral capsid into an interaction term and a bending term which leads to a reasonable description of the genome ejection process [151,152]. Apart from the detailed simulation approaches to the DNA packing within the capsid [153–155], all theoretical work is based on assumptions regarding the form of the curvature energy of the DNA forced to reside within the confines of the capsid, as well as the interactions among the highly charged and hydrated DNA segments packed at high densities within the capsid.

Elastic curvature energy appears to be the lesser of the two unknowns, though some very recent work might point to the contrary [156]. It is proportional to the square of local DNA curvature and in fact follows from the Euler-Kirchhoff model of an elastic filament. Though this model contains some subtle features due to the strong interhelical forces between the segments of the molecule [157], it nevertheless appears to be a consistent description of DNA [158] on mesoscopic scales [156]. The parameters of the Euler-Kirchhoffian model of DNA, such as its persistence length, are well established and have been measured by a variety of methods with satisfactory consensus among the results [159].

5.1 Electrostatic interactions and energies of bacteriophages

As the DNA molecule is highly charged (two elementary charges per base pair), its insertion in the bacteriophage interior results in a buildup of repulsive electrostatic force and energy [160]. The electrostatic energy of a rough model where the DNA molecule is homogeneously distributed in the capsid interior can be estimated in the DH framework using a technique similar to that presented for derivation of Eq. (7). In the

---

‖ The bacteriophages do not enter the bacteria as complete particles, rather, they insert their DNA through the bacterial membrane, leaving the capsid on the outside. Strong protein-DNA interactions in bacteriophages may be disadvantageous with respect to the insertion of the DNA, i.e. they may induce "sticking" of the DNA to the interior of the capsid and thus prevent its entry through the bacterial membrane.

limit $\kappa R \gg 1$, one obtains

$$\lim_{\kappa R \gg 1} F_{DH} = \frac{2\rho^2 R^3 \pi}{3\kappa^2 \varepsilon_0 \varepsilon}, \qquad (16)$$

where ρ is the volume density of the DNA charge (2 elementary negative charges per base pair). The electrostatic energy can now be estimated in the case of λ bacteriophage, whose radius is $R \approx 30$ nm, and whose DNA contains N_{BP}=41500 base pairs. The density is $\rho \sim 0.7\ e_0/\text{nm}^3$, and the electrostatic energy in physiological conditions $F_{DH} \sim 3\ 10^5\ k_B T$. This is at least an order of magnitude larger from the repulsive electrostatic energy of ssRNA viruses, although the protein-DNA interaction in bacteriophages may reduce this energy somewhat (the attractive protein-DNA contribution scales with R^2, however, so it becomes less relevant for larger bacteriophages). The total free energy balance upon assembly can hardly be made negative by the attractive protein-protein interactions. It then follows that the bacteriophages can not self assemble ** and they must therefore use other means of assembly.

Electrostatics accounts for only a part of the interaction energy between DNA segments. In fact these interactions can be measured directly in osmotic stress experiments[161] and can be deconvoluted into a longer ranged electrostatic contribution[162] and a shorter ranged hydration component[163]. Both of them have been quantified in terms of magnitudes and decay lengths[164]. The various formulations of the inverse spool model mostly differ in terms of the exact form of the interaction potential. While some are based entirely on theoretical polyelectrolyte models[143,144,165], others are based on semiempirical chemical potential expressions[146,147,149,150]. The best strategy would be of course to use directly the measured osmotic pressure from the osmotic stress experiments as an input for the formulation of the theory[132].

5.2 Elastic energy of DNA packing

A simple calculation we used to estimate the electrostatic energy of the bacteriophage DNA, Eq. (16), does not take into account the intrinsic stiffness of the dsDNA and the estimate obtained would be the same were the DNA base pairs completely disconnected from one another, i.e. packed homogeneously within the capsid as a cloud of charged monomers. The fact that DNA is confined and thus substantially coiled brings its bending energy clearly into focus. The bending elasticity of DNA can be accounted for within the Euler-Kirchhoff

---

** The attractive energy of protein-protein interactions scales with the total length of protein contacts (buried surface area), i.e. with R^2. In case of ssRNA viruses, the repulsive electrostatic energy, for given surface charge density scales also with R^2, which suggests that ssRNA viruses of similar surface charge densities but different radii (i.e. made of similar proteins, but of different T numbers) can all rely on attractive protein-protein interactions in order to self assemble. This is very unlike the case of bacteriophages where the electrostatic energy of DNA, for given packing density scales with R^3.

elastic energy[166,167] and introduces an additional length scale, $\mathcal{L}_P$, to the problem, referred to as the persistence length[159]. The persistence length represents the length scale over which the direction of the DNA is correlated and was measured to be about 50 nm in physiological conditions[168,169]. One should note here that this number includes also an electrostatic contribution to the DNA elasticity, i.e. the fact that charges on an elastic string also contribute to its elasticity[168,170], as they want to remain as distant as possible, favoring thus flat DNA strand conformations. One can describe this as an electrostatic renormalization of the "bare" (chemical) DNA elasticity, i.e. the elasticity that the DNA strand would have were its bases uncharged[157,171,172]. The bare persistence length, $\mathcal{L}_P$ is thus smaller than 50 nm.

To estimate the elastic contribution to the DNA packing, one can use techniques similar to those presented in Refs.[173,174]. The elastic energy of packing can be expressed in different equivalent continuum forms[132]. In what follows we use the form

$$U_{elastic} = \frac{\mathcal{L}_P L_{BP}}{2} k_B T \int d^3 r \frac{\rho_{BP}(r)}{R^2(r)}, \qquad (17)$$

where $\mathcal{L}_P$ is the persistence length of DNA, $L_{BP} = 0.34$ nm is the spacing between the DNA base pairs, $\rho_{BP}(r)$ is the DNA base-pair density (the charge density is thus $\rho(r) = 2e_0\rho_{BP}(r)$) at a radial distance r from the axis of DNA packing, and $R^2(r)$ is the squared radius of curvature of DNA, the assumption being that the configuration of DNA orientational order has cylindrical symmetry.

The elasticity of DNA should induce an inhomogeneity in the DNA packing density, $\rho_{BP}(r)$, which would be constant were the electrostatic energy the only part of the total free energy. Looking at Eq. 17, one sees that the elastic contribution diverges as the radii of curvature approach to zero. Assuming that the DNA is packed in a bacteriophage capsids in a spool-like manner, the regions which are most disfavored with respect to elastic energy are those close to the spooling axis, where the radius of curvature, R, is small, decreasing to zero on the axis. A reasonable guess of $\rho(r)$ (or variational ansatz) that includes the effects of elasticity would be a homogeneous density of DNA outside a cylinder of radius R_0, that drops to zero inside the cylindrical void, i.e. the zone of DNA exclusion due to elastic effects[174,175]. This approach is amenable to analysis, as it allows one to combine the elastic and electrostatic energies of a DNA distribution with exclusion radius R_0, and to find the radius, requiring that the total energy be minimal. The result in the case when $\kappa R \gg 1$, $R_0 \ll R$, $N_{BP} \gg 1$, $R = 30$ nm (bacteriophage λ), and $\mathcal{L}_P = 50$ nm (maximal estimate for bare DNA persistence length) is[62]

$$R_0 \sim \frac{10R}{\sqrt{N_{BP}}}. \qquad (18)$$

This equation indicates small effects of elasticity on the distribution of packed DNA. For fully packed bacteriophage λ, $N_{BP} = 41500$, and the exclusion radius amounts to $R_0 \approx 0.05R$. This is smaller (1.5 nm) than the diameter of the DNA strand (~ 2.5 nm) which means that the exclusion void, if it exist, is determined by the discrete nature of the DNA molecule, and the minimum possible length of a DNA kink. Note also that the value of the persistence length that we used in the calculations above (50 nm) includes some of the effects of electrostatic interactions that have already been included in Eq. (16), so that the void radius is expected to be even smaller.

It is also possible that nonlinear elastic effects and knotting[35,176] may be of importance for such a highly bent DNA configuration[177]. The relative contribution of elastic energy to the total energy of packing in this model scales with N_{BP} as $\ln N_{BP}/N_{BP}$. In the case of fully packed bacteriophage, it amounts to less than 5 % of total energy[62]. A similar estimate has also been obtained on the basis of experimental data in Ref.[30]. Smallness of elastic contribution to the total energy has been discussed in Ref.[132], using different techniques that do not attempt to estimate the electrostatic contribution to the DNA packing but rather utilize the data from bulk DNA osmotic pressure experiments[162] in order to obtain the required information. The region of depleted DNA density is not seen in experiments[131,178]. Rather, a more disordered phase of DNA is seen around the center of the capsid, which does not have an "onion-like" type ordering characteristic for DNA packed close to the capsid interior.

The contribution of DNA elasticity to bacteriophage packing energy and the inhomogeneities in the density of the encapsidated DNA are both quite small when the DNA is in a non-condensed state[132,179], i.e. when the interactions between the DNA segments are repulsive. In fact this is always the case either when the counterions in the bathing solution are monovalent and/or there are no crowding agents - such as polyethylene-glycol - present in the bathing solution. Monovalent counterions can change the strength of intersegment DNA interactions but they can not change their character, i.e. they can not turn repulsive interactions into attractive interactions[162].

Monovalent salt ions in general renormalize the bare phosphate backbone charge along DNA as well as determine the range of electrostatic interaction via their screening properties. The electrostatics of the DNA[180,181] becomes complicated in the presence of strongly charged counterions, basic proteins, crowding agents, the DNA confinement, etc.[182] . One of the salient features of these complicated interactions is that adding even small amounts of polyvalent counterions (spermine, spermidine, CoHex, ...) changes the nature of the electrostatic interactions so that they can turn attractive even between nominally equally charged DNA segments where one would expect repulsions[30,183,184]. The same effect can be observed also in the bacteriophage with a partially ejected genome[131] where

addition of spermine *condenses* the remaining DNA into a toroidal spool completely contained within the bacteriophage capsid. In this case the contribution of the elastic energy to the equilibrium free energy is non-negligible being in fact essential in order to counteract the attractive interactions that would tend to compact DNA into a disordered globule.

6 Osmotic pressure in viruses

Since solvent can equilibrate across the viral capsid the mechanical pressure acting on the capsid wall equals the osmotic pressure of the solution enclosed within the capsid[185,186]. This osmotic pressure has various contributions that differ in the case of flexible ssRNA viruses as opposed to the stiff dsDNA bacteriophages due to the nature of the polymer elasticity. In the former case we have purely entropic elasticity stemming simply from the connectivity of a flexible chain, while in the latter case we have enthalpic elasticity of a semiflexible Eulerian filament very different in form and magnitude[142].

The differences in molecular flexibility of the encapsidated nucleic acid component furthermore engender also qualitative differences in the ordering of the genome. In the case of ssRNA its flexibility implies, in the simplest case, a completely disordered polymer solution of homogeneous density as can be observed experimentally with e.g. hyaluronic acid[187], while in the case of dsDNA at high concentrations its stiffness leads to an orientationally and positionally ordered mesophase[188,189]. This idealized dichotomy ceases to be valid in the context of viruses at special conditions when ssRNA can assume a more ordered hypotope-bound configuration reflecting the icosahedral symmetry of the inner capsid wall[123]. The packing of dsDNA within bacteriophages seems to be governed by the stiffness of DNA that leads to the formation of local nematic alignments[134] whose orientational order is constrained by the bacteriophage capsid. The orientational ordering of dsDNA within the bacteriophage can thus be seen as a type of constrained liquid crystalline ordering that can be analyzed by a local thermodynamic mesoscopic theory[124,125,190].

In a thermodynamic equilibrium osmotic pressure at every point within the capsid has to be the same and can be set by external osmoticants such as PEG[191,192]. For various models of the viral core comprised of the nucleic acid component and the bathing ionic solution, the total osmotic pressure can be decomposed into separate terms stemming from the different components of the bathing solution. Within the mean-field approximation of the ionic component of the bathing solution it follows that its osmotic pressure is given by the van't Hoff ideal osmotic pressure of all the ionic components[42] at the hypotopic wall of the capsid. As for the nucleic acid component it is given by an additive term due to either the entropic elasticity of a flexible chain[102] or enthalpic elasticity of a semiflexible chain[132].
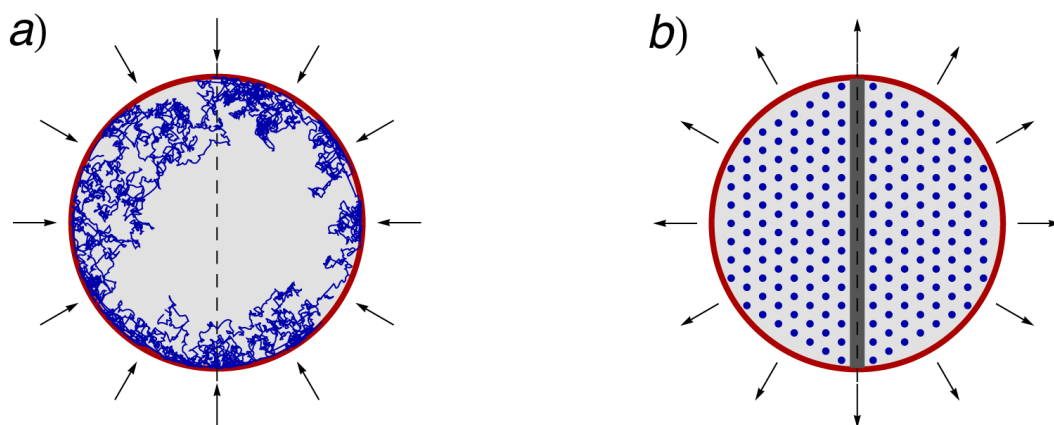
Fig. 7 Pressures sketch for an ssRNA [Panel a)] vs dsDNA virus [Panel b)]. Genome is shown in blue, capsid in red, arrows indicate the direction of pressure. Cross section through the capsid. In the ssRNA case the genome is disordered but shows a distinct boundary layer close to the hypotope of the virus with a depleted region in the middle. In the dsDNA case the genome density within the capsid is (almost) uniform, showing pronounced orientational and positional ordering.

Specifically, the case of flexible polyelectrolyte chain(s) such as ssRNA in an ionic solution the total osmotic pressure decouples into two contributions[102]: the first one is the osmotic pressure of ions inside the capsid due to electrostatic interactions between the ions themselves, as well as between the ions and the polyelectrolyte and the capsid charges; within the PB frame it equals the ideal van't Hoff expression for all the components of the solution, except the polyelectrolyte chain, evaluated at the inner surface of the capsid wall; the second one takes into account the connectedness of the polyelectrolyte chain and its interaction with the mean electrostatic field. While the first term is positive, the second one can be of either sign. When it becomes negative we refer to it as the *polyelectrolyte bridging* contribution[107]. Neither of the two terms is in general dominant and the overall sign is due to a subtle interplay of electrolyte and polyelectrolyte properties of multicomponent solution within the virus capsid[102,106,107].

For the stiff dsDNA polyelectrolyte a different decomposition can be derived[132]. In that case the osmotic pressure again decouples into two terms: the first term now corresponds to all the self-interactions between the segments of the encapsidated DNA molecule[162], whereas the second one is the contribution of the elastic bending deformation of the ordered DNA mesophase imposed by the capsid confinement[124,193]. While the second term is strictly positive, since any deformation of the orientation of the DNA molecule increases its free energy, the first one can be of either sign depending on the nature of interactions between DNA molecules[162]. In principle it contains *all* interactions between DNA molecules that could be either electrostatic or non-electrostatic in nature. Comparing the measured interaction term with the curvature term one can make a general conclusion that the latter is important only in the immediate vicinity of the central axis of DNA packing symmetry, consistent with the model calculations of bacteriophage energy presented in the previous section.

We now estimate the magnitude of the osmotic pressure in both types of viruses by making a very restrictive assumption that the interactions are strictly electrostatic in origin. This allows us to use simple scaling arguments in the derivation.

6.1 Osmotic pressure in self-assembled ssRNA viruses

The scaling form of the polyelectrolyte osmotic pressure, p_{PE}, acting in ssRNA viruses can be obtained from equation Eq. (15) if one assumes that electrostatic interactions can be described on the DH level and that confined ssRNA distribution can be described as a dense hypotopal layer of thickness a and a depleted core. One first needs to express the equation for the virus free energy Eq. 15 within the DH approximation in terms of the total charge on the capsid (Q) and on the ssRNA ($\sim -Q$),

$$F_{DH} \approx F_{DH}(-\sigma, \sigma, a) = \frac{fQ^2a}{16R^2\pi\varepsilon_0\varepsilon}. \qquad (19)$$

The osmotic pressure of the polyelectrolyte (PE) chain can then be calculated from the appropriate derivative of the free energy at constant number of surface charges,

$$p_{PE} = -\frac{1}{4R^2\pi}\left.\frac{\partial F_{DH}}{\partial R}\right|_Q. \qquad (20)$$

The sign here corresponds to the standard definition with pressure being positive, i.e. repulsive, for free energy that decays with radius. This gives

$$p_{PE} \approx \frac{fQ^2a}{32R^5\pi^2\varepsilon_0\varepsilon}. \quad (21)$$

For a "typical" ssRNA virus, this gives $p_{PE} \sim 10$ atm, acting to increase the capsid radius (outward). There is, however a component of pressure missing in the above evaluation. This is the component related to attractive protein-protein interactions in the capsid.

Namely, there is an outward pressure acting in the *empty capsid* also, since in this case the charges on the capsid decrease their energy by separating. On the DH level this purely electrostatic (ES) component of osmotic pressure is then given by

$$p_{ES}^{empty} = -\frac{1}{4R^2\pi}\left.\frac{\partial F_{DH}^{empty}}{\partial R}\right|_Q = -\frac{Q^2}{32R^5\pi^2\varepsilon_0\varepsilon\kappa}, \quad (22)$$

where F_{DH}^{empty} is the free energy of an empty capsid given in Eq. (7). If we assume that the empty capsid of an ssRNA virus in question can self-assemble, we are led to propose that the repulsive electrostatic pressure in *empty capsids* is exactly, or to a good proportion canceled by pressure arising from the attractive interactions. Were the proteins infinitesimally small particles, the two pressures should exactly cancel in the assembled empty capsid.

If we include the attractive protein-protein interactions in the calculation of total pressure, it is then given by

$$p = p_{PE} + p_{ES} \approx \frac{Q^2}{32R^5\pi^2\varepsilon_0\varepsilon}\left(fa - \kappa^{-1}\right). \quad (23)$$

Since $fa \sim 1$ nm, and $\kappa^{-1} \sim 1$ nm, we see that the total pressure in ssRNA virus *almost vanishes*, since the different contributions nearly perfectly cancel.

In more detailed calculations of osmotic pressure [111], it was found that the pressure in ssRNA virus can be either positive or negative, depending on the total charge on the virus and its relation to other parameters in the problem, see Fig. 8, but it was found to be slightly negative (half an atmosphere) at the point of optimal assembly, i.e. at the point where the length of the ssRNA is such to minimize the complexation free energy functional.

The negative values of osmotic pressure can be understood by analogy with charged planar surfaces with an oppositely charged polyelectrolyte chain in between and are due to bridging configurations of the chain [107]. One could then connect the negative values of the osmotic pressure in the case discussed here, with a similar mechanism. The ssRNA enclosed within the capsid would thus bridge the space between neighboring
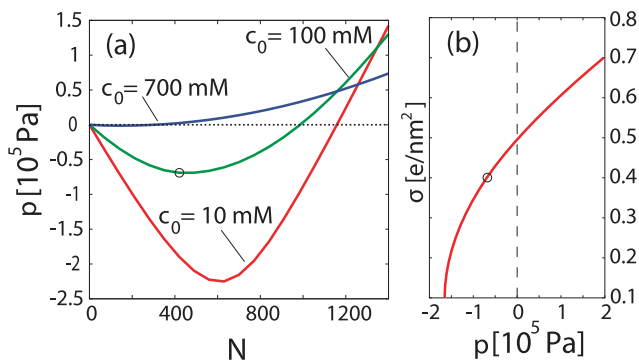

Fig. 8 Panel a): Osmotic pressure acting on the viral capsid as a function of the number of monomers. Panel b): Osmotic pressure for fixed number of monomers $N = 420$, $c_0 = 100$ mM [the point denoted by a circle in panel a)] as a function of the capsid charge density, σ. The capsid is assumed to be infinitely thin, with radius of $R=12$ nm. Other parameters of the calculation are the same as in Fig. 4 (adapted from Ref. [111]).

sections along the curved capsid and induce attractive interactions between them. Summing these attractions along the total surface of the capsid would give rise to an overall negative osmotic pressure of the polyelectrolyte. This interpretation is certainly corroborated by the characteristic density profile of the polyelectrolyte concentration in the vicinity of the capsid wall, which is in fact very similar to the bimodal polyelectrolyte concentration profile observed in the case of planar polyelectrolyte confinement [102,106].

6.2 Osmotic pressure in bacteriophages

In the case of the dsDNA bacteriophage the osmotic pressure due to DNA self-interaction is in general much larger than the one due to the DNA elasticity, which we can thus safely ignore. Assuming furthermore, that the only interaction between DNA molecules is electrostatic we can estimate its magnitude with scaling arguments. There is nevertheless large variation in the magnitude of the osmotic pressure for various virus types even if it is mostly of electrostatic nature [194].

Osmotic pressure arising from the dense packing of DNA in bacteriophages can be obtained by examining how the electrostatic energy of encapsidated DNA changes upon the increase of the capsid radius. The energy should be expressed in terms of the total number of DNA base pairs which is conserved in

the process of infinitesimal increase of the radius, i.e.

$$p \approx -\frac{1}{4R^2\pi} \left.\frac{\partial F_{DH}}{\partial R}\right|_{N_{BP}}. \tag{24}$$

Using Eq. (16) in the limit of $\kappa R \gg 1$ this yields

$$p = \frac{9N_{BP}^2 e_0^2}{8\pi^2 R^6 \kappa^2 \varepsilon_0 \varepsilon}, \tag{25}$$

which to the lowest order coincides with the estimates based on the Donnan potential [144,165,195]. Osmotic pressure is thus positive corresponding to a repulsive force acting on the capsid wall trying to increase its radius. The above equation can be evaluated for bacteriophage λ and it gives $p \sim 100$ atm, which is somewhat higher than measured in experiments [160]. On one hand, this is the pressure that the bacteriophage capsid needs to withstand, and on the other this is the pressure of the DNA coiled osmotic spring piled up against the inner surface of the capsid ready to release its chemical and mechanical energy through the portal complex on docking onto a bacterial wall [36].

For real bacteriophages at a dsDNA concentration corresponding to ~ 2.7 nm inter DNA spacing, which amounts to ~ 500mg/ml, the osmotic pressure is given not only by the electrostatic interaction between DNA molecules [196] but also by the hydration interactions stemming from the ordered vicinal layers of water close to the DNA surface [162].

6.3 Osmotic pressure: ssRNA viruses vs. dsDNA bacteriophages

Osmotic pressure of the two types of viruses [191] differs crucially in its magnitude as well as sign. While for ssRNA it is usually small and could be negative, it is strictly positive and large for bacteriophages. Negative osmotic pressures are ubiquitous in polymer shells that are impenetrable to osmoticants [197–199]. The difference of osmotic pressures also mirrors the very different assembly paths of the two types of viruses: spontaneous self-assembly at almost vanishing osmotic pressure difference vs. non-equilibrium (active) packing mechanisms where DNA encapsidation has to fight an enormous positive osmotic pressure inside the capsid [200]. These salient features of osmotic pressure carry on even into more realistic models of interactions between the nucleic acid component of the viral core [162].

The scaling forms of electrostatic interaction contribution to osmotic pressure of ssRNA and dsDNA viruses also convey another interesting contrast between the two. While in the case of the dsDNA bacteriophages the electrostatic interaction energy stems largely from the interactions acting between the segments of nucleic acid component within the volume of the virus [201], in the case of the ssRNA viruses the most important contribution comes from the interaction between the ssRNA

genome and the hypotopic surface of the capsid. This contrast in the interaction mechanisms shows up eventually as the R^{-6} scaling for DNA viruses vs. the R^{-5} scaling for the ssRNA viruses in the simple picture introduced above.

7 Electrostatic effects in viruses for highly charged counterions

Many bacteriophages require Mg^{2+}, Zn^{2+} and Ca^{2+} ions in order to attach to bacterial membrane [14]. Spermine (Spm^{4+}) and other multiply charged polyamines have been reported in a variety of viruses [202]. Arginine rich protamines are essential for sperm head condensation and DNA stabilization in spermatogenesis [203,204] and the effect of long polycations such as poly-L-lysine, poly-L-arginine and linear and branched polyethylene imine on interactions between DNA molecules have been studied in detail [205,206].

The electrostatic interactions mediated by the polyvalent counterions can not in general be discussed within the mean-field *ansatz* and their presence has consequences that do not have their counterpart in the monovalent case [196]. We now understand that polyvalent counterions even in mM concentrations mediate a different type of electrostatic interactions than the monovalent counterions [40,41]. They can drastically modify either interactions between the nucleic acid component of the virion, or the interaction between the capsomere subunits of the capsid. In both cases their effect could most probably be ascribed to a diminished repulsion or indeed net attraction between the similarly charged macromolecular subunits [207]. In the context of virus electrostatics these effects have not been analyzed yet and can be of several types that we address separately.

7.1 Specific binding of polyvalent counterions

Ions even of the same net charge have many properties which are specific for each ion type [47]. It is well known that their interaction with the DNA surface is not governed by the charge only but is a complicated combination of their size, hydration and polarization properties [208]. Ion specific effects are quite important [183,184] but are notoriously difficult to understand [209,210]. Specifically in the context of DNA interactions the nature of the ions sets the magnitude of the repulsive electrostatic interaction in the case of monovalent counterions [211], as well as the emergence of counterion mediated attractions leading to the onset of DNA condensation [212]. At present the best one can do is to include the ion specific effects on some kind of a phenomenological level as is the case when dealing with counterion adsorption along DNA [181].

Ionic specificity becomes particularly important when analyzing the details of the DNA ejection mechanisms [213] or capsid protein assembly and stability [214]. Capsid proteins often

contain Angstrom-sized "voids", holes, i.e. three-dimensional empty spaces which act as geometrical sieves, enabling entry of ions of small enough size[78]. When the void is surrounded with a particular distribution of protein charges, the void may act also as an electrostatic sieve, permitting entry and localization of only sufficiently charged ions of proper effective size. In such a localized configuration of counterions and the protein, even quantum effects may be important, which may further differentiate between specific electronic structure of different ions. All this may result in a specificity of counterion binding. Note that the mechanism discussed gives rise to a *localization* of the counterions, an effect that is not included in the PB and DH approaches. This results in effective *binding* of the counterion in the holes formed by charged protein groups[81], an effect that in the context of protein interactions is usually known as *coordinate bonding* and involves di- and trivalent ions of Fe, Zn, Co, Ca, Mg, as well as other metallic ions[78].

The role of such ions may be important both for establishment of the capsid protein fold (i.e. the functional configuration of the capsid protein or proteins in general[215]), and also for the binding of proteins to form a capsid. In this respect, the presence of these ions may act as a switch, yielding an effective attraction between the proteins when the (multivalent) specific counterions are added in required (small) concentrations. Attractions of this type can be invoked also in other contexts of the nano-scale interactions[216]. Thus, these effects that can be quite difficult to quantify can be considered as contributing to the attractive component of the protein-protein interactions, and the approach presented still serves to establish the importance of monovalent ions in the repulsive electrostatic interactions between the proteins.

7.2 "Dressed counterion" approximation: nonspecific effects of polyvalent counterions

As we already stated electrostatic interactions in the presence of polyvalent counterions or highly charged macroions can not be properly understood on the mean-field level. Sometimes even the sign of the interaction predicted by the mean-field theory is wrong. While this shortcoming of the PB theory has been known in the general colloidal context since the mid-eighties[217,218], it made its debut in the study of DNA interaction only about 10 years later[219,220]. The salient features, not the ion specific effects, of the polyvalent counterion mediated interactions can be formulated analytically based on the *strong coupling* electrostatics that has been pioneered by Rouzina and Bloomfield[221], elaborated later by Shklovskii et al.[222], Levin et al.[223], and brought into final form by Netz et al.[41,224]. An important feature of the strong coupling regime is that the counterion mediated interactions between nominally equally charged macroions can become attractive and can in-

deed cause condensation of DNA[225,226].

As the bathing solution of viruses often requires the presence of polyvalent counterions, the question remains of the nature of the polyvalent counterion effect in the electrostatics of viruses, interactions between DNAs notwithstanding. We are *not* addressing the specific binding effects[181,227] but the *non-specific* universal features of the polyvalent counterions. These effects have not been analyzed in detail yet, but there are some features that can be addressed without delving too deeply into the theory of strongly charged Coulomb systems. Usually, the aqueous solution of viruses contains a mixture of monovalent as well as polyvalent salts at relatively high and low concentrations, respectively. An intermediate approximation termed the *dressed counterion* (DC) approximation allows for an approximate treatment of all the components in this highly asymmetric multicomponent electrolyte system[228,229]. It starts with the observation that the concentration of polyvalent counterions is usually small and thus can be dealt with on the lowest order virial expansion level, whereas the monovalent salt ions can still be treated on the mean-field DH level.

Within the DC approximation the electrostatic free energy of a charged shell is composed of two parts: the first one is the DH free energy that we already evaluated before, Eq. (7), and is due to the electrostatic interaction between the charges on the capsid and the monovalent salt ions of the bathing ionic solution. The second part is due to the presence of a small concentration of polyvalent counterions interacting via the DH screened potential. In the limit of small concentrations and highly charged counterions the contribution of the polyvalent counterions can be well approximated by the first order virial expansion corresponding to a single counterion interacting with all the charges in the system, i.e. the monovalent salt ion charges as well as the fixed charges on the capsid[228,229]. The grand canonical potential can thus be written

$$
\begin{aligned}
\beta\Phi &= \beta F_{DH} - \beta F_{DC}^{(1)} = \\
&= \beta F_{DH} - n_{DC} \int_V \exp\left(-\beta e_0 v \phi_{DH}(\mathbf{r})\right) \mathrm{d}^3\mathbf{r}, \quad (26)
\end{aligned}
$$

where n_{DC} is the bulk concentration of the added v-valent counterions. In a typical case of virus assembly, the solution contains relatively large concentration of monovalent salt (~ 100 mM) and by comparison a small concentration of v-valent ions (~ 1 mM). The electrostatic potential ϕ_{DH} is again the solution of the Debye-Hückel equation for the given system.

The DC part of the total electrostatic free energy contribution $F_{DC}^{(1)}$ scales quite differently with the capsid surface charge density, σ, and the DH screening length, κ, than the mean-field contribution F_{DH} [Eq. (7)]. This is also the case when the DC calculation is performed for a sphere homogeneously filled with charge, which is a simple model of a bacteriophage

[with the free energy in the DH limit given by Eq. (16)]. Since the v-valent ions and the bulk/surface charges are usually of opposite signs, the DC part of the total free energy acts to lower the DH contribution, with the possibility of eventually completely taking over – if this happens, the total electrostatic free energy changes sign. We demonstrate this effect on a model of a bacteriophage and summarize our results in Fig. 9.
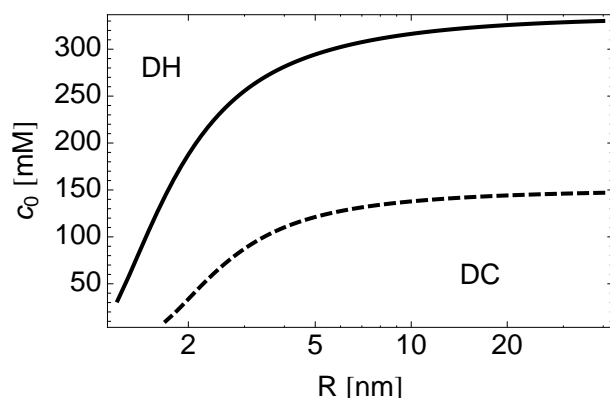

Fig. 9 The lines separating the positive (above) and negative (below) values of the total free energy of a model bacteriophage with $\rho = -0.7\ e_0/\text{nm}^3$, obtained in the dressed counterion approximation. Above the lines the DH term dominates, whereas below the lines the DC term takes over. The bacteriophage is assumed to be in a bathing solution that contains 1 mM of v-valent salt, in addition to c_0 of mono-valent salt. Full line for $v = 4$ and dashed for $v = 2$.

The boundary line shows where the change of the sign of the free energy happens, i.e. the region below the two curves is where the DH approximation that neglects the non-mean-field effects of v-valent counterions may fail. At higher monovalent salt concentrations or smaller bacteriophage radii the DC part is much smaller than the leading DH term, meaning that the behavior of the total free energy is almost completely determined by it. Below the boundary line, the DC term becomes the dominant one, with the total free energy increasing very rapidly with lower salt concentrations or larger radii. Increasing the bulk charge in general means increasing the regime where the DC part dominates. One should, however, exercise some care in interpreting these results. The dressed counterion approach is tailored for description of multivalent counterions with $v \gg 1$ and vanishing concentration, so that it may overestimate the effects pertaining to 2-valent or even 4-valent counterions. Yet, the model results that we presented do indicate that the strong electrostatic interactions engendered by polyvalent counterions may be important for viruses, especially in the conditions of poor screening by the bathing monovalent salt.

The osmotic pressure of the virion that can be deduced from

the above discussion of the free energy behavior can become less repulsive or indeed shows an emerging attractive, i.e. negative, component as a function of the volume density of the encapsidated DNA charge, ρ. This component has the same origin as the attraction observed between two charged macromolecules mediated by polyvalent counterions [40] and is due to counterion correlations in the strong electrostatic field of capsid charges. It thus acts to stabilize the capsid and diminish the osmotic pressure acting within.

8 Relevance of physical insight for architecture of real viruses and their "lifecycle"

We have seen that the balance of charges and the electrostatic energy is important for the virus stability, its assembly and function. This physical fact restricts the space of virus variability, i.e. the types of proteins encoded by its genome. Of course, not all proteins encoded by the genome will be good candidates for a capsid - some of them will not be able to form a capsid for sterical reasons. Yet, some of them may easily form empty capsids in the conditions of thermodynamical equilibrium, but due to inadequate total charge, or its spatial distribution, cannot encapsidate the RNA molecule. This depends on the amino acid content of the virus protein, and the spatial distribution of charge-carrying amino acids. Thus, physical reasons encoded in the thermodynamics and free energy of the assembled virus importantly reduce the landscape of possible virus mutants. This is especially important in self-assembling ssRNA viruses. In this case the total charge on the ssRNA and in the capsid need to be in a well defined relation [63,111,116] in order to prevent the assembly of empty capsids in thermodynamical equilibrium. The stabilizing effects of the nucleic acid in this case transpires also through the negative values of the overall osmotic pressure dur to the polyelectrolyte bridging effects of ssRNA. The length of ssRNA may thus act as a regulator of the capsid size, gathering sufficiently large number of proteins in order to screen its charge, i.e. to bring the total charge of the assembled virus within the borders enabling a spontaneous assembly (Brownian dynamics studies of ssRNA (generic polymer) virus assembly which account for the electrostatic interactions via model pairwise interactions can be found in Ref. [121]). The electrostatic interactions between the ssRNA and the proteins can thus be viewed as the reason for the characteristic size of the virus [111,230] ††.

The calculations we presented predict that there is a polyelectrolyte depleted region in the center of typical ssRNA viruses. One may wonder whether this depleted region can be partially or completely eliminated in viruses of smaller radii.

---

†† A smilar effect has been experimentally found [231] and theoretically discussed [127] in context of assembly of virus proteins around functionalized (negatively charged) nanoscopic golden cores.

Such viruses may indeed be viable, but note that they would need to carry more charge on the capsid proteins. Imagine, for example, the evolution of an ssRNA virus whereby its proteins change and instead of e.g. $T = 3$ capsid they form a $T = 1$ capsid [*]. Presuming that the ssRNA is not much shortened in this drastic evolutionary event and that it carries similar charge as before, the proteins would need to carry three times more charge in order to pack the ssRNA with the same efficiency as before (here we neglect the change in entropic component of the confined ssRNA molecule). It seems reasonable that it would be more difficult to reconcile the stability of the fold of such highly charged proteins with the requirement that the proteins, once folded, interact attractively with their neighbors in a formed capsid, i.e. that they expose required amino-acids in contact regions [78]. The electrostatics of individual proteins, the proteins in capsid and in contact with ssRNA may thus conspire to impose restrictions on virus size.

The total charge on the DNA, i.e. its total length, seems to be a decisive factor concerning the stability of bacteriophages. Mutant phages with genomes longer or shorter (deletion mutants) from the wild type virus were studied experimentally (see Ref. [233] and references therein). Intriguingly, it was found that the deletion mutants were more resistant to changes in the environment, particularly to heat shock. The mutant phages with genomes longer from the wild type variant were, on the other hand, found to be highly unstable [233]. The consideration of electrostatic forces can explain these experimental findings, as the internal pressure in the shorter genome (less charged) mutants will be smaller. This may lead to improved mechanical stability of such mutants, but this does not necessarily mean that such mutants have evolutionary advantage. Viruses need to be sufficiently stable physically, but they must, in addition, *(i)* contain all the information needed for their replication, and *(ii)* dissamble efficiently in appropriate conditions. One may speculate that the longer genomes may offer an evolutionary advantage as they enable the coding of additional proteins, which may be used for different purposes, e.g. as scaffolding proteins [6,234] enabling a more precise (and possibly quicker) assembly of viruses in infected cells. On the other hand, the need for packing the additional "information" may render the virus unstable. The virus evolution is thus directly influenced by physical constraints on the molecular and nano-scale levels. This is of course true for evolution of all organisms, but it becomes particularly transparent in the case of simple macromolecular assemblies such as viruses.

It was found experimentally that the size and mass of bacteriophage capsids are highly correlated with the genome size [233]. This can again be explained from the electrostatic considerations, as the phages with longer genomes need larger capsids in order to reduce the electrostatic repulsion in the functional (filled) virus. Interestingly, the same type of correlation does not hold for ssRNA viruses [63,233], as the electrostatic interactions there scale differently with capsid radius, and the total interior capsid charge is a decisive factor that determines the ssRNA length [63,111] (see Figs. 4 and 6). The observed differences in "architectures" of ssRNA and bacteriophage viruses are also a consequence of their "lifestyles", in particular the fact that many ssRNA viruses self-assemble, in contrast to bacteriophages which do not.

9 Conclusions

We have shown that the size and architecture of viruses are importantly constrained by the (nonspecific) electrostatic interactions acting between the constituents of a virus, i.e. its proteins and its genome molecule. A rough understanding of these *physical* constraints can be grasped from simple mean-field expression for the electrostatic contribution to virus energy (Eqs. (7) and (16)) and pressure (Eqs. (23) and (25)). A more detailed and quantitative argumentation would also have to take into account that electrostatic interactions are only part of the whole picture and that e.g. ion specific effects together with hydration interactions put finer details on the conclusions reached above. Unfortunately these effects are much less understood and do not allow for simple quantitative estimates as is the case with electrostatic interactions.

Viruses are information encoding machines and most microbiological/virological studies of viruses concentrate on this aspect of viruses (see e.g. Ref. [235]). Indeed, the viral genome must encode all the proteins required for the life-cycle of a virus and missing in the cellular environment. The typical length of ssRNA genome is often discussed in terms of replication fidelity [2,236]. The rate of mutation of ssRNA viruses [237] is about 10^{-4} to 10^{-5} errors per nucleotide per round of copying which means that viruses in the daughter generation typically contain one mutation when compared to the parental generation [2,236]. This replication error rate is often considered as the factor restricting the genome length that can be faithfully maintained. We have shown that the genome length is also a *physically* important quantity that should be examined in context of the size and charge distribution of virus capsid (phenotype). This means that the genome length is constrained additionally by the physical interactions conveyed by the genome molecule when packed with proteins it encodes. We find here an interesting interplay of non-physical (non-material) and physical/structural aspects of the information encoded in a genome molecule that must be tuned properly in a functional virus.

---

[*] Some viruses are *dimorphic* when assembled without the genome molecule. To observed the dimorphism a modification of the capsid protein is sometimes required. C-termini modified (shortened) capsid proteins of hepatitis B virus have been experimentally observed to form $T = 3$ and $T = 4$ capsids [232]. The percentage of $T = 4$ capsid was found to vary, depending on the amount of shortening of the C-termini [232] (see also Ref. [73]).

Acknowledgments

References

1　R. J. Parsons, M. Breitbart, M. W. Lomas and C. A. Carlson, *ISME Journal*, 2011.

2　E. C. Holmes, *The Evolution and Emergence of RNA Viruses*, Oxford University Press, Oxford, 1st edn, 2009.

3　D. H. Duckworth and P. A. Gulig, *Biodrugs*, 2002, **16**, 57–62.

4　J. M. Inal, *Archivum Immunologiae et Therapiae Experimentalis*, 2003, **51**, 237244.

5　R. Podgornik, D. Harries, J. DeRouchey, H. H. Strey and V. A. Parsegian, *in Gene Therapy: Therapeutic Mechanisms and Strategies, N. Smyth Templeton, Marcel Dekker, New York, Third Edition*, 2008, 443–484.

6　T. S. Baker, N. H. Olson, and S. D. Fuller, *Microbiol. Mol. Biol. Rev.*, 1999, **63**, 862–922.

7　C. Xiao, Y. G. Kuznetsov, S. Sun, S. L. Hafenstein, V. A. Kostyuchenko, P. R. Chipman, M. Suzan-Monti, D. Raoult, A. McPherson and M. G. Rossmann, *PLoS Biology*, 2011, **7**, 0958–0966.

8　V. S. Reddy, P. Natarajan, B. Okerberg, K. Li, K. V. Damodaran, R. T. Morton, C. L. B. III and J. E. Johnson, *Journal of Virology*, 2001, **75**, 11943–11947.

9　E. C. Pollard, *The Physics of Viruses*, Academic Press, New York, 1st edn, 1953.

10　*Emerging Topics in Physical Virology*, ed. P. G. Stockley and R. Twarock, Imperial College Press, London, 1st edn, 2010.

11　W. H. Roos, R. Bruinsma and G. J. L. Wuite, *Nature physics*, 2010, **6**, 733–743.

12　A. Zlotnick, *PNAS*, 2004, **101**, 15549–15550.

13　J. Cello, A. Paul and E. Wimmer, *Science*, 2002, **297**, 1016–1018.

14　A. J. Cann, *Principles of Molecular Virology*, Elsevier, London, 4th edn, 2005.

15　A. Evilevitch and M. Castelnovo, *Physics of Viral Infectivity: Matching Genome Length with Capsid Size, in Emerging Topics in Physical Virology*, Imperial College Press, London, 1st edn, 2010.

16　S. Tzlil, M. Deserno, W. M. Gelbart and A. Ben-Shaul, *Biophysical Journal*, 2004, **86**, 2037–2048.

17　M. Deserno and T. Bickel, *Europhys. Letts.*, 2003, **62**, 767.

18　L. YANG, H. LIANG, T. E. ANGELINI, J. BUTLER, R. CORIDAN, J. X. TANG and G. C. L. WONG, *Nature Materials*, 2004, **3**, 615–619.

19　A. Arkhipov, P. L. Freddolino and K. Schulten, *Structure*, 2006, **14**, 1767–1777.

20　R. Zandi, P. van der Schoot, D. Reguera, W. Kegel and H. Reiss, *Biophysical Journal*, 2006, **90**, 1939–1948.

21　N. F. Steinmetz, *Nanomedicine: Nanotechnology, Biology and Medicine*, 2010, **6**, 634–641.

22　N. F. Steinmetz, V. Hong, E. D. Spoerke, P. Lu, K. Breitenkamp, M. G. Finn and M. Manchester, *JACS*, 2009, **131**, 1709317095.

23　B. Millman, T. Irving, B. Nickel and M. Loosley-Millman, *Biophys. J.*, 1984, **45**, 551–556.

24　J. Bernal and I. Fankuchen, *J.Gen.Physiol.*, 1941, **25**, 111–165.

25　G. Oster, *J.Gen.Physiol.*, 1950, **33**, 445–473.

26　E. J. W. Verwey and J. T. G. Overbeek, *Theory of the stability of lyophobic colloids*, Elsevier, Amsterdam, 1st edn, 1948.

27　V. Parsegian, *Ph.D. Thesis, Harvard*, 1965.

28　R. Konecny, J. Trylska, F. Tama, D. Zhang, N. A. Baker, I. Charles L. Brooks and J. A. McCammon, *Biopolymers*, 2006, **82**, 106–120.

29　P. A. Shields and S. R. Farrah, *Applied and Environmental Microbiology*, 1983, **45**, 526–531.

30　X. Qiu, D. Rau, V. Parsegian, L. Fang, C. Knobler and W. Gelbart, *Phys. Rev. Lett.*, 2011, **106**, 028102–1–028102–4.

31　T. Jiang, Z.-G. Wang and J. Wu, *Biophys. J.*, 2009, **96**, 3065–3073.

32　D. F. Evans and H. Wennerstrm, *The colloidal domain: where physics, chemistry, biology, and technology meet (Advances in Interfacial Engineering)*, Wiley-VCH, New York, 1st edn, 1999.

33　D. Marenduzzo and C. Micheletti, *Journal of Molecular Biology*, 2003, **330**, 485–492.

34　I. Ali, D. Marenduzzo and J. M. Yeomans, *Phys. Rev. Lett.*, 2006, **96**, 208102–1–208102–4.

35　D. Marenduzzo, E. Orlandini, A. Stasiak, D. W. Sumners, L. Tubiana and C. Micheletti, *PNAS*, 2009, **106**, 22269–22274.

36　W. M. Gelbart and C. M. Knobler, *Physics Today*, 2008, **61**, 42–54.

37　A. Šiber, *arXiv:0711.3527*, 2007.

38　A. V. Dobrynin and M. Rubinstein, *Prog. Polym. Sci.*, 2005, **30**, 1049–118.

39　R. R. Netz and D. Andelman, *Physics Reports*, 2003, **380**, 1–95.

40　A. Naji, M. Kanduč, R. R. Netz and R. Podgornik, *Understanding Soft Condensed Matter via Modeling and Computation Eds. W.-B. Hu & A.-C. Shi, Series in Soft Condensed Matter Edited by David Andelman and Günter Reiter, World Scientific, Singapore*, 2010.

41　H. Boroudjerdi, Y. W. Kim, A. Naji, R. R. Netza, X. Schlagberger and A. Serr, *Phys. Rep.*, 2005, **416**, 129–199.

42　D. Andelman, *Handbook of Biological Physics: Structure and Dynamics of Membranes, edited by Lipowsky R. and Sackmann E., (Elsevier Science B.V., Amsterdam)*, 1995, **1B**,.

43　C. Holm, P. K. ekicheff and E. Rudolf Podgornik, *Electrostatic Effects in Soft Matter and Biophysics*, Kluwer, Dordrecht, 1st edn, 2001.

44　W. C. Poon and D. Andelman, *Soft Condensed Matter Physics in Molecular and Cell Biology*, Taylor & Francis, New York, 1st edn, 2006.

45　E. in soft matter, *J. Phys.: Condens. Matter*, 2009, **21**, 113102.

46　D. Ben-Yaakov, D. Andelman and R. Podgornik, *J. Chem. Phys.*, 2011, **134**, 0747051–1–0747051–12.

47　D. Ben-Yaakov, D. Andelman, R. Podgornik and D. Harries, *Current Opinion in Colloid & Interface Science*, 2011.

48　F. M. Hughson, *Current Biology*, 1997, **7**, R565–R569.

49　F. H. Crick and J. D. Watson, *Nature*, 1956, **177**, 473–475.

50　R. Zandi, D. Reguera, R. F. Bruinsma, W. M. Gelbart and J. Rudnick, *PNAS*, 2004, **101**, 15556–15560.

51　P. Ceres and A. Zlotnick, *Biochemistry*, 2002, **41**, 11525–11531.

52　J. Lidmar, L. Mirny and D. R. Nelson, *Phys. Rev. E*, 2003, **68**, 051910–1–051910–10.

53　A. Šiber, *Phys. Rev. E*, 2006, **73**, 061915–1–061915–10.

54　A. Luque, R. Zandi and D. Reguera, *PNAS*, 2010, **107**, 5323–5328.

55　A. Fokine, P. Chipman, P. Leiman, V. Mesyanzhinov, V. Rao and M. Rossmann, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 6003–6008.

56　A. Levandovsky and R. Zandi, *Phys. Rev. Lett.*, 2009, **102**, 198102 –4.

57　J. Benjamin, B. K. Ganser-Pornillos, W. F. Tivol, W. I. Sundquist and G. J. Jensen, *J. Mol. Biol.*, 2005, **346**, 577–588.

58　V. L. Lorman and S. B. Rochal, *Phys. Rev. Lett.*, 2007, **98**, 185502–1–185502–4.

59　V. L. Lorman and S. B. Rochal, *Phys. Rev. B*, 2008, **77**, 224109–12.

60　C.-K. Limn and P. Roy, *J. Virol.*, 2003, **77**, 11114–11124.

61　A. L. Shaw, S. K. Samal, K. Subramanian and B. V. Prasad, *Structure*, 1996, **4**, 957–967.

62 A. L. Božič, A. Šiber and R. Podgornik, *in preparation*, 2011.

63 V. A. Belyi and M. Muthukumar, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 17174–17178.

64 J. Blinn, *ACM Transactions on Graphics*, 1982, **1**, 235–256.

65 W. K. Kegel and P. van der Schoot, *Biophys. J.*, 2004, **86**, 3905–3913.

66 W.-Y. Ching, *(Personal communication)*.

67 F. Pichierri, *arXiv:1107.5853v1*, 2011.

68 T. Dokland, R. A. Bernal, A. Burch, S. Pletnev, B. A. Fane and M. G. Rossmann, *J. Mol. Biol.*, 1999, **288**, 595–608.

69 J. Bancroft, *Advances in Virus Research*, 1970, **16**, 99–134.

70 M. F. Hagan and D. Chandler, *Biophysical Journal*, 2006, **91**, 42–54.

71 H. Fraenkel-Conrat and R. Williams, *Proc. Natl. Acad. Sci. USA*, 1955, **41**, 690–698.

72 R. F. Bruinsma, W. M. Gelbart, D. Reguera, J. Rudnick and R. Zandi, *Phys. Rev. Lett.*, 2003, **90**, 248101–1–248101–4.

73 A. Šiber and A. Majdandžić, *Phys. Rev. E*, 2009, **80**, 021910–1–021910–7.

74 V. Parsegian, *Annu. Rev. Biophys. Bioeng.*, 1973, **2**, 221–255.

75 D. Leckband and J. Israelachvili, *Quarterly Reviews of Biophysics*, 2001, **34**, 105–267.

76 M. Daune, *Molecular Biophysics: Structures in motion*, Oxford University Press, Oxford, 1st edn, 1999.

77 J. Israelachvili, *Intermolecular & Surface Forces*, Academic Press, Amsterdam, 2nd edn, 1991.

78 A. V. Finkelstein and O. B. Ptitsyn, *Protein Physics*, Academic Press, London, 1st edn, 2002.

79 V. A. Bloomfield, D. M. Crothers, I. Tinoco, J. E. Hearst, D. E. Wemmer, P. A. Killman and D. H. Turner, *Nucleic Acids: Structures, Properties, and Functions*, University Science Books, Herndon, 1st edn, 2000.

80 S. Karlin and V. Brendel, *Proc. Natl. Acad. Sci. U.S.A.*, 1988, **85**, 9396–9400.

81 F. Oosawa, *Polyelectrolytes*, Marcel Dekker, Inc., New York, 1st edn, 1971.

82 R. R. Novoa, G. Calderita, R. Arranz, J. Fontana, H. Granzow and C. Risco, *Biol. Cell*, 2005, **97**, 147–172.

83 D. Ben-Yaakov and D. Andelman, *Physica A*, 2010, **389**, 2956–2961.

84 S. Safran, *Statistical Thermodynamics Of Surfaces, Interfaces, And Membranes*, Cornell University Press, Westview Press, 1st edn, 2003.

85 R. Kjellander and R. Ramirez, *Journal of Physics: Condensed Matter*, 2008, **20**, 494209–16.

86 N. V. Churaev, B. Derjaguin and V. Muller, *Surface forces*, Springer, Berlin, 1st edn, 1987.

87 A. Lošdorfer Božič, A. Šiber and R. Podgornik, *Phys. Rev. E*, 2011, **83**, 041916–1–041916–13.

88 G. Rose, A. Geselowitz, G. Lesser, R. Lee and M. Zehfus, *Science*, 1985, **229**, 834–838.

89 W. K. Kegel and P. van der Schoot, *Biophys. J.*, 2006, **91**, 1501–1512.

90 A. Šiber and R. Podgornik, *Phys. Rev. E*, 2007, **76**, 061906–1–061906–10.

91 K. W. Adolph and P. J. G. Butler, *J. Mol. Biol.*, 1974, 327–341.

92 L. Lavelle, M. Gingery, M. Phillips, W. M. Gelbart, C. M. Knobler, R. D. Cadena-Nava, J. R. Vega-Acosta, L. A. Pinedo-Torres and J. Ruiz-Garcia, *J. Phys. Chem. B*, 2009, **113**, 3813–3819.

93 D. G. Isom, C. A. Castaneda, B. R. Cannon, P. D. Velu and B. G.-M. E., *Proc. Acad. Natl. Sci.*, 2010, **107**, 1609616100.

94 M. Harms, C. Castaneda, J. Schlessman, G. Sue, D. Isom, B. Cannon and B. Garcia-Moreno, *J. Mol. Biol.*, 2009, **389**, 34–47.

95 W.-Y. Ching and P. Rulis, *Electronic Structure Method for Complex Materials:The orthogonalized linear combination of atomic orbitals*.

96 C. J. Marzec and L. A. Day, *Biophys. J*, 1993, **65**, 2559–2577.

97 J. Bancroft, E. Hiebert and C. Bracker, *Virology*, 1969, **39**, 924–930.

98 P.-G. de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca, 1st edn, 1979.

99 M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics*, Clarendon Press, Oxford, 1st edn, 1999.

100 P.-G. de Gennes, *Rep. Prog. Phys*, 1969, **32**, 187–205.

101 M. Muthukumar, *Journal of Chemical Physics*, 1987, **86**, 7230–7236.

102 R. Podgornik, *J. Phys. Chem.*, 1992, **96**, 884–896.

103 L. R. G. Treloar, *The physics of rubber elasticity*, Oxford University Press, USA, Oxford, 3rd edn, 2005.

104 M. Rubinstein and R. H. Colby, *Polymer Physics*, Oxford University Press, USA, Oxford, 1st edn, 2003.

105 D. Swigon, *The mathematics of DNA structure, mechanics and dynamics*, Springer Science + Business Media, Berlin, 1st edn, 2009.

106 R. Podgornik, *Journal of Polymer Science Part B: Polymer Physics*, 2004, **42**, 3539–3556.

107 M. Ličer and R. Podgornik, *J. Phys.: Condens. Matter*, 2010, **22**, 414102–414109.

108 L. Tang, K. N. Johnson, L. A. Ball, T. Lin, M. Yeager and J. E. Johnson, *Nature Struct. Biol.*, 2001, **8**, 77–83.

109 M. Tihova, K. A. Dryden, T. vy L. Le, S. C. Harvey, J. E. Johnson, M. Yeager and A. Schneemann, *J. Virol.*, 2004, **78**, 2897–2905.

110 C. E. Woodward and B. Jansson, *Chem. Phys.*, 1991, **155**, 207–219.

111 A. Šiber and R. Podgornik, *Phys. Rev. E*, 2008, **78**, 051915–1–051915–9.

112 T. Akesson, C. Woodward and B. Jonsson, *J. Chem. Phys.*, 1989, **91**, 2461–2469.

113 M. Turesson, T. Akesson and J. Forsman, *Langmuir*, 2007, **23**, 9555–9558.

114 I. Borukhov, D. Andelman and H. Orland, *J. Phys. Chem. B*, 1999, **103**, 5042–5057.

115 T. Abraham, D. Christendat, Z. Xu, J. Masliyah, J. Gohy and R. Jerome, *AIChE J.*, 2004, **50**, 261326.

116 D. G. Angelescu, R. Bruinsma and P. Linse, *Phys. Rev. E*, 2006, **73**, 041921–1–041921–17.

117 B. Ninham and V. Parsegian, *J. Theor. Biol.*, 1971, **31**, 405–428.

118 Y. Hu, R. Zandi, A. Anavitarte, C. M. Knobler and W. M. Gelbart, *Biophysical Journal*, 2008, **94**, 1428–1436.

119 M. F. Hagan, *J. Chem. Phys.*, 2009, **130**, 114902–17.

120 P. van der Schoot and R. Bruinsma, *Phys. Rev. E*, 2005, **71**, 061928–1–061928–12.

121 O. M. Elrad and M. F. Hagan, *Phys. Biol.*, 2010, **7**, 1–17.

122 P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson and K. Schulten, *Structure*, 2006, **14**, 437–449.

123 R. Bruinsma, *Eur. Phys. J. E*, 2006, **19**, 303–310.

124 D. Svenšek, G. Veble and R. Podgornik, *Phys. Rev. E*, 2010, **82**, 011708–1–011708–14.

125 H. Shin and G. M. Grason, *arXiv*, 2011, 1107.4406v1.

126 P. J. Flory, *Principles of polymer chemistry*, Cornell University Press, Ithaca, 1st edn, 1953.

127 A. Šiber, R. Zandi and R. Podgornik, *Phys. Rev. E*, 2010, **81**, 051919–1–051919–11.

128 D. Marenduzzo and C. Micheletti, *Journal of Molecular Biology*, 2003, **330**, 485–492.

129 P. Jing, F. Haque, D. Shu, C. Montemagno and P. Guo, *Nano Lett.*, 2011.

130 A. Guasch, J. Pous, G. Ibarra, F. Gomis-Ruth, J. Valpuesta, N. Sousa, J. Carrascosa and M. Coll, *J. Mol. Biol.*, 2002, **315**, 663676.

131 A. Leforestier, A. Šiber, F. Livolant and R. Podgornik, *Biophys. J.*, 2011, **100**, 2209–2216.

132 A. Šiber, M. Dragar, V. A. Parsegian and R. Podgornik, *Eur. Phys. J. E*, 2008, **26**, 317–325.

133 L. R. Comolli, A. J. Spakowitz, C. E. Siegerist, P. J. Jardine, S. Grimes, D. L. Anderson, B. Bustamante and K. H. Downing, *Virology*, 2008, **371**, 267–277.

134 J. Lepault, J. Dubochet, W. Baschong and E. Kellenberger, *EMBO J.*, 1987, **6**, 1507–1512.

135 N. Hud and K. Downing, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 14925–14930.

136 C. Forrey and M. Muthukumar, *Biophysical Journal*, 2006, **91**, 25–41.

137 P. Serwer, *J. Mol. Biol.*, 1986, **190**, 509–512.

138 L. Black, W. Newcomb, J. Boring and J. Brown, *Proc. Natl. Acad. Sci. USA*, 1985, **82**, 7960–7964.

139 R. KILKSON and M. F. MAESTRE, *Nature*, 1962, **195**, 494 – 495.

140 A. Cole and R. Langley, *Biophysical Journal*, 1963, **3**, 189–197.

141 S. M. KLMENKO, T. I. TIKCHONENKO and V. M. ANDREEV, *J. Mol. Biol.*, 1967, **23**, 523–533.

142 I. Gabashvili, A. Y. Grosberg, D. Kuznetsov and G. Mrevlishvili, *Biofizika*, 1991, **36**, 780–787.

143 T. Odijk, *Biophys. J.*, 1998, **75**, 1223.

144 T. Odijk, *Phil. Trans. Roy. Soc. A*, 2004, **362**, 1497–1517.

145 J. Kindt, S. Tzlil, A. Ben-Shaul and W. M. Gelbart, *Proc. Natl. Acad. Sci.*, 2001, **98**, 13671.

146 P. K. Purohit, J. Kondev and R. Phillips, *Proc. Natl. Acad. Sci.*, 2003, **100**, 3173.

147 P. K. Purohit, M. M. Inamdar, P. D. Grayson, T. M. Squires, J. Kondev and R. Phillips, *Biophys. J.*, 2005, **88**, 851.

148 S. Tzlil, J. T. Kindt, W. M. Gelbart and A. Ben-Shaul, *Biophys. J.*, 2003, **84**, 1616–1627.

149 W. S. Klug, M. T. Feldmann and M. Ortiz, *Comput. Mech.*, 2005, **35**, 146.

150 W. S. Klug and M. Ortiz, *J. Mech. Phys. Sol.*, 2003, **51**, 1815.

151 A. Evilevitch, L. Lavelle, C. M. Knobler, E. Raspaud and W. M. Gelbart, *Emerging Topics in Physics Virology*, 2003, **100**, 9292.

152 P. Grayson, A. Evilevitch, M. M. Inamdar, P. Purohit, C. M. Knobler, R. Phillips and W. M. Gelbart, *Biophys. J.*, 2005, **88**, 230.

153 J. Arsuaga, R. K.-Z. Tan, M. Vazquez, D. W. Sumners and S. C. Harvey, *Biophys. Chem.*, 2002, **101**, 475.

154 A. J. Spakowitz and Z.-G. Wang, *Biophys. J.*, 2005, **88**,.

155 D. Marenduzzo, *Computational and Mathematical Methods in Medicine*, 2008, **9**, 317–325.

156 P. A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker and P. C. Nelson, *Nature Nanotechnology*, 2006, **1**, 137.

157 R. Podgornik, P. Hansen and V. Parsegian, *J. Chem. Phys.*, 1999, **113**, 9343–9350.

158 M. C. Williams and I. Rouzina, *Curr. Op. Struc. Biol.*, 2002, **12**, 330.

159 P. J. Hagerman, *Ann. Rev. Biophys. Biophys. Chem.*, 1988, **17**, 265–286.

160 D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson and C. Bustamante, *Nature*, 2001, **413**, 748–752.

161 V. A. Parsegian, R. P. Rand and N. L. F. et al., *Methods in Enzymology*, 1986, **127**, 400.

162 H. H. Strey, R. Podgornik, D. C. Rau and V. Parsegian, *Curr. Opin. Struc. Biol.*, 1998, **8**, 309–313.

163 S. Leikin, V. A. Parsegian and D. C. Rau, *Ann. Rev. Phys. Chem.*, 1993, **44**, 369.

164 H. H. Strey, V. A. Parsegian and R. Podgornik, *Phys. Rev. E*, 1999, **59**, 999.

165 T. Odijk and F. Slok, *J. Phys. Chem. B*, 2003, **107**, 8074–8077.

166 J. F. Marko and E. D. Siggia, *Macromolecules*, 1995, **28**, 8759–8770.

167 S. Neukirch, *Phys. Rev. Lett.*, 2004, **93**, 198107–1–198107–4.

168 C. G. Baumann, S. B. Smith, V. A. Bloomfield and C. Bustamante, *Proc. Natl. Acad. Sci.*, 1997, **94**, 6185–6190.

169 A. Peterlin, *Nature*, 1953, **171**, 259–260.

170 R. Zandi, J. Rudnick and R. Golestanian, *Phys. Rev. E*, 2003, **67**, 061805–1–061805–10.

171 T. Odijk, *J. Polym. Sci., Polym. Phys. Ed.*, 1977, **15**, 477.

172 J. Skolnick and M. Fixman, *Macromolecules*, 1977, **10**, 948.

173 J. Ubbink and U. Odijk, *Europhys. Lett.*, 1996, **33**, 353–358.

174 P. K. Purohit, M. M. Inamdar, P. D. Grayson, T. M. Squires, J. Kondev and R. Phillips, *Biophys. J.*, 2005, **88**, 851–866.

175 T. Odijk, *Phil. Trans. R. Soc. Lond. A*, 2004, **362**, 1497–1517.

176 J. C. LaMarque, T. vy L. Le and S. C. Harvey, *Biopolymers*, 2004, **73**, 348–355.

177 P. A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker and P. C. Nelson, *Nature Nanotechnology*, 2006, **1**, 137–141.

178 L. R. Comolli, A. J. Spakowitz, C. E. Siegerist, P. J. Jardine, S. Grimes, D. L. Anderson, C. Bustamante and K. H. Downing, *Virology*, 2008, **371**, 267–277.

179 A. Petrov and S. C. Harvey, *Biophys. J.*, 2008, **95**, 497–502.

180 A. G. Cherstvy, *Phys. Chem. Chem. Phys.*, 2011, **13**, 9942–9968.

181 A. A. Kornyshev, D. Lee, S. Leikin and A. Wynveen, *Reviews of Modern Physics*, 2007, **179**, 943–996.

182 D. Marenduzzo, C. Micheletti and E. Orlandini, *J. Phys.: Condens. Matter*, 2010, **22**, 283102.

183 D. C. Rau, and V. A. Parsegian, *Biophys. J.*, 1992, **61**, 260–271.

184 D. Rau and V.A.Parsegian, *Biophys. J.*, 1992, **61**, 246–259.

185 D. Panja and I. J. Molineux, *Phys. Biol.*, 2010, **9**, 045006–15.

186 I. J. Molineux, *Virology*, 2006, **344**, 221 229.

187 R. M. Peitzsch and W. F. Reed, *Biopolymers*, 1992, **32**, 219–238.

188 R. Podgornik, H. H. Strey, K. Gawrisch, D. C.Rau, A. Rupprecht and V. Parsegian, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 4261–4266.

189 R. Podgornik, H. Strey and V. Parsegian, *Curr. Opin. Coll. and Interf. Sci.*, 1998, **3**, 534–539.

190 I. M. Kulić, D. Andrienko and M. Deserno, *Europhys. Lett.*, 2004, **67**, 418.

191 E. Nurmemmedova, M. Castelnovo, C. Catalano and A. Evilevitch, *Quart. Rev. Biophys.*, 2007, **40**, 327–56.

192 J. A. Cohen, R. Podgornik, P. L. Hansen and V. A. Parsegian, *J. Phys. Chem. B*, 2009, **113**, 37093714.

193 H. Shin and G. M. Grason, *(Preprint)*, 2011.

194 D. N. Fuller, J. P. Rickgauer, P. J. Jardine, S. Grimes, D. L. Anderson and D. E. Smith, *PNAS*, 2007, **104**, 1124511250.

195 A. Cordova, M. Deserno, W. M. Gelbart and A. Ben-Shaul, *Biophysical Journal*, 2003, **85**, 70–74.

196 W. M. Gelbart, R. F. Bruinsma, P. A. Pincus and V. A. Parsegian, *Physics Today*, 2000, **53**, 38–44.

197 G. B. Sukhorukov, M. Brumen, E. Donath and H. M. ohwald, *J. Phys. Chem. B*, 1999, **103**, 6434–6440.

198 P. Zupanovic, D. Kuic, M. Brumen, A. Fajmut and D. Juretic, *arXiv:0907.2874v1*, 2009.

199 O. I. Vinogradova, O. V. Lebedeva and B.-S. Kim, *Annu. Rev. Mater. Res.*, 2006, **36**, 143–178.

200 A. Evilevitch, L. Fang, A. Yoffe, M. Castelnovo, D. C. Rau, V. A. Parsegian, W. M. Gelbart and C. M. Knobler, *Biophysical Journal*, 2008, **94**, 1110–1120.

201 R. Podgornik, *in Electrostatis effects in soft matter and biophysics, C. Holm, P. Kekicheff and R. Podgornik Eds., NATO Science Series II - Mathematics, Physics and Chemistry*, 2001, **46**, 409–433.

202 J. Carter and V. Saunders, *Virology: Principles and Applications*, Wiley, Chichester, 1st edn, 2007.

203 R. Balhorn, *Genome Biol.*, 2007, **8**, 227.1–227.8.

204 J. DeRouchey and D. Rau, *(submitted)*, 2011.

205 J. DeRouchey, R. Netz and J. R. adler, *Eur. Phys. J. E*, 2005, **16**, 1728.

206 J. DeRouchey, V. A. Parsegian and D. C. Rau, *Biophys. J.*, 2010, **99**, 2608–2615.

207 X. Qiu, K. Andresen, L. W. Kwok, J. S. Lamb, H. Y. Park and L. Pollack,

Phys. Rev. Lett., 2007, **99**, 038104–1–038104–4.

208 N. Hud and J. Plavec, *Biopolymers*, 2003, **69**, 144–158.

209 W. Kunz, P. L. Nostro and B. Ninham, *Curr. Op. Coll. Interf. Sci.*, 2004, **9**, 1–18.

210 *Specifc ion effects*, ed. W. Kunz, World Scientic, Singapore, 1st edn, 2010.

211 R. Podgornik, D. Rau and V. Parsegian, *Biophys. J.*, 1994, **66**, 962–971.

212 D. Rau and C. Stanley, *Curr. Op. Coll. Interf. Sci.*, 2011.

213 A. Evilevitch, *Journal of Physical Chemistry B*, 2006, **110**, 22261–22265.

214 A. Evilevitch, W. H. Roos, I. L. Ivanovska, M. Jeembaeva, B. Jonsson and G. J. L. Wuite, *Journal of Molecular Biology*, 2011, **405**, 18–23.

215 A. Pey, D. Rodriguez-Larrea, J. Gavira, B. G.-M. E. and J. Sanchez-Ruiz, *J. Am. Chem. Soc.*, 2010, **132**, 1219–1219.

216 R. H. French, V. A. Parsegian, R. Podgornik and et al., *Reviews Of Modern Physics*, 2010, **82**, 1887–1944.

217 L. Gulbrand, B. Jonsson, H. Wennerstrom and P. Linse, *J. Chem. Phys.*, 1984, **82**, 2221–2228.

218 M. L. Bret and B. H. Zimm, *Biopolymers*, 1984, **23**, 287–312.

219 N. Gronbech-Jensen, R. J. Mashl, R. F. Bruinsma and W. M. Gelbart, *Phys. Rev. Lett.*, 1997, **78**, 2477–2480.

220 A. P. Lyubartsev and L. Nordenskioeld, *J. Phys. Chem.*, 1995, **99**, 10373–10382.

221 I. Rouzina and V. A. Bloomfield, *J. Phys. Chem.*, 1996, **100**, 9977.

222 A. Y. Grosberg, T. T. Nguyen and B. I. Shklovskii, *Rev. Mod. Phys.*, 2002, **74**, 329–345.

223 Y. Levin, *Rep. Prog. Phys.*, 2002, **65**, 1577.

224 A. Naji, S. Jungblut, A. G. Moreira and R. R. Netz, *Physica A*, 2005, **352**, 131.

225 M. O. Khan, S. M. Mel'nikov and B. Jonsson, *Macromolecules*, 1999, **32**, 8836–8840.

226 P. Hansen, R. Podgornik, D. Svensek and V. Parsegian, *Phys. Rev. E*, 1999, **60**, 1956–1966.

227 S. Lee, C. V. Tran and T. T. Nguyen, *J. Chem. Phys*, 2011, **134**, 125104–1–125104–7.

228 M. Kanduč, A. Naji, J. Forsman and R. Podgornik, *J. Chem. Phys.*, 2010, **132**, 124701–15.

229 M. Kanduč, A. Naji, J. Forsman and R. Podgornik, *Phys. Rev. E*, 2011, **84**, 011502–1–011502–9.

230 R. Zandi and P. van der Schoot, *Biophys. J.*, 2009, **96**, 9–20.

231 J. Sun, C. DuFort, M.-C. Daniel, A. Murali, C. Chen, K. Gopinath, B. Stein, M. De, V. M. Rotello, A. Holzenburg, C. C. Kao and B. Dragnea, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 1354–1359.

232 A. Zlotnick, N. Cheng, J. F. Conway, F. P. Booy, A. C. Steven, S. J. Stahl and P. T. Wingfield, *Biochemistry*, 1996, **35**, 7412–7421.

233 M. D. Paepe and F. Taddei, *PLoS Biology*, 2006, **4**, 1248–1256.

234 P. A. Thuman-Commike, B. Greene, J. A. Malinski, J. King and W. Chiu, *Biophys. J.*, 1998, **74**, 559–568.

235 A. H. Reid, T. G. Fanning, J. V. Hultin and J. K. Taubenberger, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 1651–1656.

236 S. C. Manrubia and E. Lázaro, *Physics of Life Reviews*, 2006, **3**, 65–92.

237 J. W. Drake and J. J. Holland, *Proc. Natl. Acad. Sci.*, 1999, **96**, 13910–13913.