A liquid crystal continuum modeling framework for icosahedra bacteriophage viruses is developed and tested. The main assumptions of the model are the chromonic columnar hexagonal structure of confined DNA, the high resistance to bending and the phase transition from solid to fluid-like states as the concentration of DNA in the capsid decreases during infection. The model predicts osmotic pressure inside the capsid and the ejection force of the DNA as well as the size of the isotropic volume at the center of the capsid. Extensions of the model are discussed.

Double-stranded (ds)DNA bacteriophages are of renewed interest due to their use in medicine and biotechnology. Icosahedral bacteriophages consist of a protein capsid with icosahedral symmetry whose assembly is followed by the packing, by a molecular motor, of a single naked dsDNA molecule. The DNA molecule inside the viral capsid is found under extreme concentration and osmotic pressure. At the time of infection the DNA is released by a mechanism that suggests a phase transition, possibly into a ‘liquid-like’ state. Both processes, packing and releasing of the genome, are highly dependent on how the DNA molecule folds inside the viral capsid; however our understanding of this folding remains very limited.

The concentration of the DNA molecule inside the viral capsid is between 200 and 800 mg/ml and the estimated osmotic pressure ranges between 40 and 60 atmospheres. Three factors contribute to the excess pressure found inside the viral capsid: the decrease in entropy associated with the confinement imposed by the capsid, the high resistance of the DNA molecule to bending beyond its persistence length and the self-repulsion of the DNA molecule. Experimental and theoretical studies acquired over the last 30 years have shown that under such conditions the DNA molecule forms a columnar hexagonal liquid crystal phase. In particular, liquid crystalline phases in bacteriophages were first proposed, with an explicit reference to hexagonal packing made in and since then, consistent data have been accumulating.

A number of theoretical models, based on experimental and/or simulation data, have been proposed to describe the folding of the DNA molecule inside the bacteriophage capsid with only few attempts to describe the DNA molecule in a liquid crystalline phase. These attempts however were based on energy fields originally proposed for modeling DNA molecules in free solution and do not provide adequate description of DNA inside the phage capsid.

In this article, we use cryo-electron microscopy (cryo-EM) data, the hexagonal chromonic liquid crystal structure of the packed DNA, and the continuum theory of liquid crystals to build a new model of DNA folding inside the viral capsid. The final packed structure of the DNA molecule corresponds to an energy minimizing configuration of the energy proposed below.

Our model assumes that the hydrated DNA fills the entire volume of the capsid hence two relevant parameters are the molar c and volume cv concentration of DNA. We take the customary point of view that the DNA molecule is a semiflexible elastic polymer characterized by the persistence length Lp which is about the size of the radius of the capsid. Since Cryo-EM data for most bacteriophages present multilayered spooling-like configurations on the outer layers of the packed genome, we assume that the hexagonal chromonic structure provides a geometric scaffolding that sustains the trajectory of the DNA molecule. The decreasing available volume during DNA packing induces an extreme bending on the DNA molecule whose bending resistance prevents the hexagonal ordering from completely filling the capsid, so an inner core of disordered/isotropic DNA is assumed to form.
and protein core, the size and shape of the disordered region, the volume \( V \) of the capsid and the DNA density graphs that allow us to obtain the number, \( M \), of concentric layers \([31,37]\); (2) ordering of the DNA molecule at the boundary as promoted by the capsid \([12,21,22,36]\); (3) DNA effective diameter \( d \) and genome length \( L \); (4) pressure measurements as well as speeds of DNA ejection \([8,11,39–41]\).

To make our model precise, we assume that the capsid corresponds to a bounded, discretely axisymmetric region \( B \in \mathbb{R}^3 \) with \( \partial B \) representing the piecewise smooth, faceted, viral capsid. We let \( l_0 > 0 \) denote the length of the axis that connects the location of the connector with its antipodal site in the viral capsid. Let \( \Omega_0 \subset B \) denote the isotropic region of the capsid, also taken to be axisymmetric with respect to \( l_0 \) and \( \Omega := B \setminus \Omega_0 \), nonempty, denote the region occupied by the hexagonal chromonic liquid crystal phase. A piecewise smooth curve \( r = r(s), s \in [0, L] \), describes the axis of the DNA. Note that, in general, the size of \( \Omega_0 \) is an unknown of the problem.

The model of hexagonal chromonic liquid crystals that we present corresponds to that of de Gennes in the case of small distortions \([42\text{, sec.7.1}]\). A triple of linearly independent unit vectors \( \mathbf{n}, \mathbf{m}, \mathbf{p} \) represents the uniaxial nematic director, along the direction tangent to the axis of the DNA molecule, and two local directions of ordering, respectively. The vectors \( \mathbf{m} \) and \( \mathbf{p} \) correspond to the lattice vectors of the columnar phase and account for the meridian and parallel arrangements. We will assume that the three vectors are almost mutually perpendicular in the sense made precise by the energy described below. In an energy minimizing configuration, these directions will be determined from their corresponding boundary values, that is, the filament organization in the contact with the capsid. Two complex valued functions \( \psi = \rho e^{i\varphi} \) and \( \gamma = \rho e^{i\psi} \) account for the density of ordered material and describe the space filling conformation. Specifically, \( \rho \geq 0 \) gives the packing density of the layered molecules (\( \rho = 0 \) corresponds to disordered DNA) along the two preferred directions. The real number \( q \) corresponds to the frequency of the layers with \( d = 2\pi/q \) representing the the effective diameter of the DNA and interlayer distance.

The chromonic structure is described by a discrete 2-

\[
F_{\text{chr}} = F_N(\nabla \mathbf{n}, \mathbf{n}) + F_N(\nabla \mathbf{m}, \mathbf{m}) + F_N(\nabla \mathbf{p}, \mathbf{p}) + C_d(\nabla \psi - iq \psi \mathbf{m})^2 + |\nabla \gamma - iq \gamma \mathbf{p}|^2 + F_{\text{lux}}(\mathbf{n}, \mathbf{p}, \mathbf{m})
\]

\[
F_N(\nabla \mathbf{n}, \mathbf{n}) = k_1(\nabla \cdot \mathbf{n})^2 + k_2(\mathbf{n} \cdot \nabla \times \mathbf{n})^2 + k_3|\mathbf{n} \times (\nabla \times \mathbf{n})|^2 + (k_2 + k_4)(\text{tr}(\nabla \mathbf{n})^2 - (\nabla \cdot \mathbf{n})^2)
\]

\[
F_{\text{lux}}(\mathbf{n}, \mathbf{m}, \mathbf{p}) = A(\mathbf{m} \cdot \mathbf{n})^2 + B(\mathbf{m} \cdot \mathbf{p})^2 + C(\mathbf{n} \cdot \mathbf{p})^2.
\]

Not all the terms in the energy are relevant to the continuum theory. Assuming that \( k_1 \) and \( k_2 \) diverge near the effective diameter of the DNA and interlayer distance.

\[
\{S^l_m\}_{l=0}^M : \omega(x) = id, \quad \{S^p_m\}_{j=0}^P : \vartheta(x) = jd\quad (1)
\]

where \( M \) and \( P = \lfloor \frac{L}{d} \rfloor \) (with brackets indicating the integer immediately below the value of the quotient) are nonzero positive integers. The first family corresponds to spheres concentric with the capsid while the second are the planes defined by \( z = \text{constant} \). As part of satisfying the boundary conditions of the problem, we take \( S^0_m = \partial B \), with \( j = 0 \) corresponding to the parallel surface \( z = \frac{l_0}{2} \). These choices correspond to a spooling direction starting on the lower part of the capsid surface. The family of intersecting curves \( \mathcal{C}_{ij} = S^i_m \cap S^j_p \), \( 0 \leq i \leq M, 0 \leq j \leq P \), give the location of the axis of the DNA molecule in the capsid and its ordering determines the direction of DNA spooling in the capsid. The final outcome, however, is independent of the ordering choices.

We propose that the total energy is given by the sum of the ordered chromonic phase plus the disordered isotropic core \( \Omega_0 \) and a surface energy penalizing the interface between the two:

\[
E = \int_{\Omega} F_{\text{chr}} \, dx + \nu \text{Vol}(\Omega_0) + \sigma \text{Area}(\partial \Omega_0), \quad \text{with } (2)
\]
the phase transition to the hexagonal phase leads to a pure bending distortion at the limit (42, sect. 7.1: 43, page 315). The bending modulus is taken to be that of a semiflexible polymer in confinement, $k = K_B T L_p r^{-1}$. I is the geometric moment of inertia with respect to the axis. The value of I depends on the distance of the DNA molecule to the axis of the capsid. For simplicity, we associate the Oseen-Frank energy function $F_N(\nabla n, n)$ to the director fields $m$ and $p$, expressing, in particular, high resistance to bending, to ensure a solid-like packing in the ordered region. This corresponds to the assumption that the filament properties also determine the geometry of packing. As for isotropic liquid crystals, the energy density $\nu$ of the isotropic phase is assumed to be a function of the molar concentration $c$. Specifically, we use the approximate expressions derived by Onsager for isotropic liquid crystals, that in large concentration regimes (44), take the form: $\nu = K_B T \ln(c^2 - \frac{4}{35} + e^{-2}) - 1$. The surface energy density $\sigma = K_B T \frac{0.257}{4\pi}$ is based on the Onsager’s theory and derived in (45) for such regimes. (Approximations appropriate to low concentration regimes can also be found in the aforementioned references). The quantity $C_d$ corresponds to the compressibility modulus. The volume concentration $c_v$, which measures how much of the capsid volume is occupied by DNA, together with the observation that the genome tends to fill the entire capsid (46) leads us to assume that $C_d = C_d(c_v)$ satisfying $C_d(0) = 0$, $C_d'(c_v) > 0$ and $\lim_{c_v \to 1^-} C_d(c_v) = +\infty$, the latter enforcing an idealized perfect packing structure at the limit of high concentrations. Note that these assumptions on $C_d$ encode the phase transition solid to liquid-like behavior since the contribution of the chromonic phase will decrease as $c_v$ decreases at the time of infection. The energy (5) allows for the relaxation of the orthogonality constraints between pairs of vectors $n, m, p$. The positive constants $A, B, C$ are then taken to be larger than the maximum dimensionless parameter of the energy. The combination of the energy terms (4) and (5) allows for the relaxation of the orthogonality constraints between pairs of vectors $n, m, p$. The positive constants $A, B, C$ are then taken to be larger than the maximum dimensionless parameter of the energy. The combination of the energy terms (4) and (5) allows for the relaxation of the orthogonality constraints between pairs of vectors $n, m, p$. The positive constants $A, B, C$ are then taken to be larger than the maximum dimensionless parameter of the energy.
TABLE I. Physical measurements of four different bacteriophages. The symbol $L_P$ denotes the persistence length of a DNA chain of length $L$, effective diameter $d$, molar concentration $c$ in a sphere-like capsid of radius $R$ with a measured radius $r_0$ of the disordered core. T4 [50, 51]; T5 [36]; T7 [21]; $\epsilon_{15}$ [37].

| Virus | $L_P$ | $d$ | $c$ | $L$ (nm) | $R$ (nm) | $r_0/R$ |
|-------|-------|----|----|---------|---------|--------|
| T4    | 1.32  | 0.06 | 21.37 | 55047.6  | 40.0    | 0.5500 |
| T5    | 1.39  | 0.07 | 17.85 | 39423.8  | 42.0    | 0.4286 |
| T7    | 2.03  | 0.10 | 18.17 | 12932.0  | 26.05   | 0.5889 |
| $\epsilon_{15}$ | 1.9 | 0.09 | 13.98 | 12846.0  | 28.37   | 0.5735 |

TABLE II. Predicted size of the disordered phase and osmotic pressure for different bacteriophages. The table lists the measured and predicted radii of the isotropic for several viruses. The values are scaled by the corresponding capsid radii as listed in Table I. The percentage error is the difference between the measured and predicted size of the isotropic phase, divided by the measured core size. In the calculation for T5 and $\epsilon_{15}$, the expressions of the material parameters $\sigma$ [41] and $\sigma$ [55] correspond to the approximation in the low concentration regime. The evaluations for the other entries in the table are taken at the high concentration limits [54–56]. The last column lists the pressure calculated near the capsid boundary.

| Virus | Measured Core Size | Predicted Core Size | Error | P( atm) |
|-------|-------------------|-------------------|-------|--------|
| T4    | 0.5500            | 0.5348            | 2.76% | 28.70  |
| T5    | 0.4286            | 0.4268            | 0.40% | 30.17  |
| T7    | 0.5889            | 0.5712            | 3.00% | 44.02  |
| $\epsilon_{15}$ | 0.5735 | 0.569 | 0.63% | 40.41  |

* The authors acknowledge the discussions and advice from Dr. Oleg Lavrentovich. Arsuaga, Calderer and Vazquez wish also to acknowledge the support from the grants DMS/NIGMS R01 GM109457, nfs-dmref 1435372 and DMS1057284, respectively.

1. I. U. Haq, W. N. Chaudhry, M. N. Akhtar, S. Andleeb, and I. Qadri, Virology journal 9, 9 (2012).
2. A. Sulakvelidze, Z. Alavizde, and J. G. Morris, Antimicrobial agents and chemotherapy 45, 649 (2001).
3. D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante, Nature 413, 748 (2001).
4. A. Leforestier and F. Livolant, Journal of molecular biology 396, 384 (2010).
5. T. Liu, U. Sae-Ueng, D. Li, G. C. Lander, X. Zuo, B. Jönsson, D. Rau, I. Shefer, and A. Evilevitch, Proceedings of the National Academy of Sciences 111, 14675 (2014).
6. U. Sae-Ueng, D. Li, X. Zuo, J. B. Huffman, F. L. Homa, D. Rau, and A. Evilevitch, Nature chemical biology 10, 861 (2014).
7. E. Kellenberger, E. Carlemalm, J. Sechaud, A. Ryter, and G. De Haller, in Bacterial chromatin (Springer, 1986) pp. 11–25.
8. A. Evilevitch, L. Lavelle, C. M. Knobler, E. Raspaud, and W. M. Gelbart, Proceedings of the National Academy of Sciences 100, 9292 (2003).
9. M. Jeembaeva, M. Castelnovo, F. Larsson, and A. Evilevitch, Journal of molecular biology 381, 310 (2008).
10. S. C. Riemer and V. A. Bloomfield, Biopolymers 17, 785 (1978).
11. A. Leforestier and F. Livolant, Proceedings of the National Academy of Sciences 106, 9157 (2009).
12. J. Lepault, J. Dubochet, W. Baschong, and E. Kellenberger, The EMBO journal 6, 1507 (1987).
13. D. Reith, P. Cifra, A. Stasiak, and P. Vrnau, Nucleic acid research 40, 5129 (2012).
14. R. L. Rill, Proceedings of the National Academy of Sciences 83, 342 (1986).
15. T. E. Strzelecka, M. W. Davidson, and R. L. Rill, Nature...
331, 457 (1988).
[16] H.-S. Park, S.-W. Kang, L. Tortora, Y. Nastishin, D. Finotello, S. Kumar, and O. D. Lavrentovich, The Journal of Physical Chemistry B 112, 16307 (2008).
[17] F. Livolant, Physica A: Statistical Mechanics and its Applications 176, 117 (1991).
[18] A. Leforestier and F. Livolant, Biophysical journal 65, 56 (1993).
[19] A. Leforestier, S. Brasiles, M. De Frutos, E. Raspaud, L. Letellier, P. Tavares, and F. Livolant, Journal of molecular biology 384, 730 (2008).
[20] D. Marenduzzo, E. Orlandini, A. Stasiak, L. Tubiana, C. Micheletti, et al., Proceedings of the National Academy of Sciences 106, 22269 (2009).
[21] M. E. Cerritelli, N. Cheng, A. H. Rosenberg, C. E. McPherson, F. P. Booy, and A. C. Steven, Cell 91, 271 (1997).
[22] W. C. Earnshaw and S. R. Casjens, Cell 21, 319 (1980).
[23] N. V. Hud, Biophysical journal 69, 1355 (1995).
[24] A. S. Petrov, M. B. Boz, and S. C. Harvey, Journal of structural biology 160, 241 (2007).
[25] P. Serwer, S. J. Hayes, and R. H. Watson, Journal of molecular biology 223, 999 (1992).
[26] J. Arsuaga and Y. Diao, Computational and Mathematical Methods in Medicine 9, 303 (2008).
[27] J. Arsuaga, R. K.-Z. Tan, M. Vazquez, S. C. Harvey, et al., Biophysical chemistry 101, 475 (2002).
[28] J. Arsuaga, M. Vazquez, S. Trigueros, J. Roca, et al., Proceedings of the National Academy of Sciences 99, 5373 (2002).
[29] J. Arsuaga, M. Vazquez, P. McGuirk, S. Trigueros, J. Roca, et al., Proceedings of the National Academy of Sciences of the United States of America 102, 9165 (2005).
[30] D. Marenduzzo, C. Micheletti, E. Orlandini, et al., Proceedings of the National Academy of Sciences 110, 20081 (2013).
[31] G. C. Rollins, A. S. Petrov, and S. C. Harvey, Biophysical journal 94, L38 (2008).
[32] A. J. Spakowitz and Z.-G. Wang, Biophysical journal 88, 3912 (2005).
[33] E. Kellenberger, E. Carlemark, J. Schaud, A. Ryter, and G. De Haller, Bacterial chromatin 1, 11 (1986).
[34] L. R. Comoli, A. J. Spakowitz, C. E. Siegerist, P. J. Jardine, S. Grimes, D. L. Anderson, C. Bustamante, and K. H. Downing, Virology 371, 267 (2008).