**Electrostatics of single-stranded genome binding in viruses**

Vladimir A. Belyi, M. Muthukumar ~ UMass, Amherst

**Virus as a physical object**

Physically, virus particles consist of a viral genome surrounded by a protein capsid. The genome can be single stranded or double stranded, composed of RNA or DNA, and stored in one or more polynucleotide chains. The surrounding capsid is assembled by the association of the repeat units of similar or identical proteins.

**Setting up the Model for the Virus Interior**

There are two major components that need to be treated independently:

- **Viral genome**
  - General properties:
    - ssRNA or ssDNA
    - Typical size: 1000-12000 nucleotides
  - approximate genome by a flexible polyelectrolyte

- **Capsid peptide arms**
  - General properties:
    - Typical arm length: 30-130 residues
    - Typical arm charge: 10-30 e
    - Typical spacing: 1.5-3 nm
  - approximate peptide arms by a polyelectrolyte brush

**Breaking up the mutual interactions between genome/peptide arms/ions**

**Interaction Hamiltonian**

\[ H = \frac{k_B T}{2} \sum \frac{\langle \xi_i \rangle^2}{\langle \xi_i \rangle^2} + \sum \frac{\langle \xi_o \rangle^2}{\langle \xi_o \rangle^2} + \frac{k_B T}{2} \sum \frac{\langle \xi_i \rangle^2}{\langle \xi_i \rangle^2} + \frac{k_B T}{2} \sum \frac{\langle \xi_o \rangle^2}{\langle \xi_o \rangle^2} \]

where \( \xi_i \) and \( \xi_o \) are the fluctuations of the genome and peptide arms, respectively.

At the saddle point, the fluctuating field coincides with the electrostatic potential, and the partition function breaks into separate components for genome, peptide arms, ions, and field.

**At this point we introduce fluctuating field**

\( \omega \), and perform Hubbard-Stratonovich transformation:

\[ Z = \int D\xi_o \exp \left[ -\beta H \right] \]

\[ = \prod \int D\xi_o \exp \left[ \frac{1}{\beta} \int (\nabla \phi)^2 d\tau \right] \]

\[ = \prod \int D\xi_o \exp \left[ \frac{1}{\beta} \int (\nabla \phi)^2 d\tau \right] \]

**Virus Genome**

Genome of the Semliki Forest Virus

**Satellite Tobacco Mosaic Virus**

[Images from VIPER - http://mmtsb.scripps.edu/viper/]

There are several scenarios of genome packing in viruses:

- Bacteriophages package genome (dsDNA) using protein motor
- Some ssRNA viruses directly bind genome to the capsid proteins
- Other ssRNA viruses bind genome using highly charged peptide arms

This is the type of viruses considered in this work.

\[ Z = Z_g|\phi|Z_o|\phi|Z_i|\phi| \exp \left[ -\beta \epsilon_o \int (\nabla \phi)^2 d\tau \right] \]
Mathematical treatment of the capsid peptide arms:

Peptide arms are densely grafted, so that space constraint leads to their extension. Free energy can then be written as

$$\beta F = \frac{1}{2} \frac{d}{dr} \left( \frac{d}{dr} r^2 \right) + q_{\text{ext}} \Phi (r)$$

Minimization with constraint on peptide arm monodispersity leads to parabolic electrostatic potential

$$\Phi (r) = \frac{\pi^2}{6} \frac{r^2}{\beta}$$

This potential is robust – its shape does not depend on ions or genome. Both ions and genome “live” in the parabolic potential of the peptide brush!

($q_{\text{ext}}$, $N_{\text{arm}}$ – segment charge, segment size, segment count for brush chain)

Mathematical treatment of the genome

Statistics of the genome, submerged into external field, is best described via Green’s function $G$

$$\frac{\partial}{\partial r} \left( \frac{\partial}{\partial r} G (r, r', n) \right) + q_{\text{RNA}} \Phi (r) = 0$$

Use ground-state approximation:

$$\frac{\partial}{\partial r} \left( \frac{\partial}{\partial r} \left( G (r, r', n) \right) \right) + q_{\text{RNA}} \Phi (r) = \delta (r - r')$$

Finally, we put the two results together:

Peptide arms enforce parabolic shape of the electrostatic potential. Genome “lives” in this potential

$$\psi (z) \propto z^2 \exp (-z^2 / c_{\text{max}}^2)$$

The problem thus maps onto Quantum Harmonic Oscillator. Predicted nucleotide density profile

$$\rho \propto \psi (z) \propto z^2 \exp (-z^2 / c_{\text{max}}^2)$$

Comparison with known viruses

- Nucleotides are expected to occupy a single spherical shell
- There’s a gap expected between nucleotides and capsid

CryoEM data

Is Genome Length Random in Itself?

Some published data on RNA viruses

<table>
<thead>
<tr>
<th>Virus Name</th>
<th>Number of Capid Protein</th>
<th>Net charge per particle</th>
<th>Genome size per particle</th>
</tr>
</thead>
<tbody>
<tr>
<td>dsDNA virus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fiDNA virus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCV virus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV virus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSV virus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MHV virus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDV virus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSV virus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIV virus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLE virus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VSV virus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZMV virus</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Some published data on RNA viruses

Conclusions

- Have analytically treatable model of genome - peptide arm interactions inside RNA viruses
- Can predict placement and density of genomic nucleotides
- Proposed model explains observed gap between virus capsid and genome
- Can predict the length of viral genome based on the net charge of capsid peptide arms