Biophysical Journal, Volume 96

Supporting Material

Energy landscape for DNA rotation and sliding through a phage portal

Jeremiah Nummela and Ioan Andricioaei

SUPPLEMENTARY MATERIAL
Energy landscape for DNA rotation and sliding through a phage portal

Jeremiah Nummela and Ioan Andricioaei
Department of Chemistry
University of California, Irvine, CA 92697

Computational Methods

Generation of a closed 12-mer structure

The available crystal structure for the portal protein is of a homo-13-mer which forms and crystallizes out of solution (1). Structural information is missing for 28 residues at the N-terminus, 36 residues at the C-terminus, and residues 170 to 238 in the middle of each subunit. The N-terminal residues, as well as the missing segment in the middle of the protein are on the outer surface of the pore, and should have a minimal effect on the protein-DNA interaction. The backbone of 30 of the missing residues from the middle of the protein was identified from the crystal structure data, and modeled as polyalanine, but the actual identity of the residues is unknown. They were included in the simulation as alanines, and are contained in the set residues with constrained alpha carbons.

In order to produce a model of the 12-mer found to assemble along with the procapsid, a short, iterative, guided molecular dynamics (2) simulation was followed by all atom fitting of the resulting structure to match the cryo-EM image of the 12-mer assembled into the connector particle. Molecular dynamics simulations were carried out using the CHARMM27 force field (3, 4). One subunit of the 13-mer was deleted. An harmonic RMSD constraint was then enforced on the two monomers flanking the gap. The two monomers directly opposite in the portal were used as the target structure to serve as a model for the protein-protein interface between gp6 subunits. An offset was used so that the target RMSD was equal to the current RMSD minus 5 Å or 0 Å whichever was greater, and the force constant for the constraint was 100 kcal/mol/Å^2. Every one hundred steps the target structure was updated to reflect changes in the curvature of the portal as it closes to a 12-mer, and the RMSD goal is recalculated. A $4r$ dielectric constant was used in the modeling of non-bonding interactions during biased molecular dynamics, and was replaced by explicit water solvation in the free energy calculations (see below). The SHAKE algorithm was used so that the trajectory could be propagated with a time step of 1.5 fs. A Langevin thermostat with frictional coefficient of 5.0 ps^{-1} for heavy atoms was used to maintain the system at 300 K. After 50 ps of simulation time the new protein-protein interface has formed, modeled after the interface directly opposite in the portal.

Normal mode flexible fitting

Normal mode flexible fitting (5) was carried out using the all atom structure of the closed 12-mer. Using the first 20 normal modes the fitting converged to a correlation coefficient of 0.268 after 18 iterations. The inclusion of 30 modes failed to substantially increase the correlation coefficient, indicating satisfactory convergence. The relatively low correlation is not a sign of poor fit, but is simply due to the presence of the full connector particle in the EM data, to which only the gp6 12-mer was fit.

Double stranded B-DNA, 48 base pairs in length, was generated with the program NAB (6). The sequence used was a repetition of CGAT. The DNA was centered in the pore of the portal protein structure obtained by NMFF, and protrudes about 20 Å out of each side of the pore. Explicit solvent (TIP3P (7) water model) was added in a cylindrical region with a radius of 25 Å, and a length of 210 Å, completely solvating the DNA, the interior of the pore, and the protein-DNA interface. Ninety four sodium ions were added, neutralizing the negative charge of the DNA backbone, but not neutralizing the overall charge on the system. This concentration of ions exceeds bulk average physiological concentrations, but is expected to occur locally due to the large negative charge of the system and is necessary to prevent large distortions in the DNA structure. The exterior of the protein, which would typically interact with other capsid proteins, is left unsolvated. This is done to minimize the size of the simulation, and because the nature of the interaction between the portal protein and the rest of the capsid is unclear (this is the site where the ATPase ring would bind). However, the alpha carbons of residues at the outer periphery have been harmonically constrained with a force constant of 1 kcal/mol/Å^2 to the location they occupy after the NMFF procedure in order to maintain the structure observed in the EM experiment.

Equilibration of the model

Minimization and equilibration of the system were performed with NAMD (8) using the CHARMM27 force field. All heavy atoms of the protein and nucleic acid were constrained during the initial equilibration with a force constant of 20 kcal/mol/Å^2. The constraint was gradually released over the course of 1 ns, although the outer periphery of the protein remained harmonically restrained with a force constant of 1 kcal/mol/Å^2. A Langevin thermostat was used to maintain a temperature of 300 K, and SHAKE was used to allow a 2 fs time step. Following an initial equilibration of 5 ns, protein residues with all of their atoms falling outside the cylinder of radius 25 Å used to constrain the solvent were fixed to average position, thereby removing the need to recalculate fixed-fixed interactions and thus reducing the computational complexity. The system was equilibrated for eight more nanoseconds with the outer atoms fixed. The final protein structure had the turbine-like shape (see Fig. 1 in the main text) common amid different bacteriophages and herpesviruses (9). In the final 800 ps a harmonic RMSD restraint was placed on the DNA to prevent it from deform-

ing significantly from an ideal helix. Although the DNA is stable during the equilibration when sufficient positive ions are added, the nature of the umbrella sampling restraints used can cause deformation of the helix in the absence of this RMSD restraint.

The fitting procedure and the subsequent results are similar to that from the analysis performed in the original report of a first structural model of the pore (1). Our final structure was derived by allowing relaxation through normal mode motion and molecular dynamics. The structure reported in the original report was generated by a fitting which involved moving sub-domains of the protein as rigid bodies, and maintaining an imposed twelve-fold symmetry. The most significant discrepancy is that the width of the channel at its narrowest point is narrower in the rigid model than in ours by approximately 1 Å. In an other difference, unlike in the structural report, we did not use an imposed twelve fold symmetry; this was important to allow proper equilibration of monomer interfaces and protein-DNA contacts. The use of normal modes and molecular dynamics equilibration instead of the rigid body fitting in the original structural report was important so that we produce a relaxed structure that was a good minimum energy model for the Hamiltonian used to compute the free energy.

Umbrella sampling

Umbrella sampling (10) in combination with a multidimensional weighted histogram analysis method (11) was used. Umbrella sampling was performed to calculate a two dimensional potential of mean force relevant to the potential role of portal protein rotation in the import of DNA. The calculation is performed according to the description of Roux (12), using the WHAM software package written by Alan Grossfield (13); periodicity was implemented by wrapping around a torus in the two coordinates. The two coordinates chosen were

$$q = \frac{1}{n}\sum_i (z_i - z_{0i}), \tag{1}$$

where n is the number of heavy atoms in the DNA, z_i is the z coordinate of the ith heavy atom, and z_{0i} is the z coordinate of the ith heavy atom of an idealized DNA helix centered in the pore, and

$$\phi = \frac{1}{n}\sum_i (\theta_i - \theta_{0i}), \tag{2}$$

where θ_i is the angle made with the x-axis in the xy plane containing it by the ith heavy atom. Note that the portal protein has been centered so that translocation occurs along the z-axis in the positive direction. The coordinate q summarizes translation of the DNA through the portal, while ϕ summarizes rotation of the DNA relative to the portal. Harmonic restraining potentials of $V_q = \frac{1}{2}k_q(q_0 - q)^2$, and $V_\phi = \frac{1}{2}k_\phi(\phi_0 - \phi)^2$ were adopted for sampling windows centered at (q_0, ϕ_0) grid points, with force constants k_q and k_ϕ as described below.

Because the portal is composed of 12 identical subunits, the potential of mean force will be periodic after the DNA is rotated only 30 degrees about the z-axis. Additionally, along the coordinate q, there is an approximate periodicity after the import of one base pair (3.4 Å), followed by a rotation of 6 degrees. This approximate periodicity is broken only by inhomogeneity in the base pair sequence. To take advantage of these periodicities the sampling range was taken to be -1.7 Å to 1.7 Å along the coordinate q, and -15° to 15° along the coordinate ϕ. Initial state for the sampling of each window were generated by restarting the equilibration trajectory with large constraint force constants of $k_q = 1000$ kcal/Å^2 and $k_\phi = 50000$ kcal/rad^2. One hundred and twenty windows were sampled with 12 even increments along the coordinate q and 10 along the coordinate ϕ. This grid constituted the unit cell (a "tile") for the periodic free energy surface presented in Fig. 2 of the main paper. The restraining force constants were relaxed to $k_q = 100$ kcal/Å^2 and $k_\phi = 5000$ kcal/rad^2, and in each window the system was allowed to equilibrate for 200 ps prior to sampling. Each window involved sampling for 2 ns, during which the reaction coordinate was recorded every picosecond for processing with a two-dimensional WHAM implementation (13).

Regarding the choice of ionic species in the simulation, a discussion on the role of various ions is appropriate. Because of its high negative charge, the cost of packing of the DNA polymer is expected to depend on the degree of electrostatic repulsion, which in turn will be modulated by the presence in solution of cations, which will effectively screen the electrostatic repulsion and which will alter the hydration spine of DNA (and some of which, those of larger charge, may even condense DNA).

A study by Fuller et al. (14) showed that Mg^{2+} affected DNA import in a complex fashion, so ejection may be similarly modulated by Mg^{2+}. It is formally possible that Mg^{2+} screening can lower a component of our computed free energy surface: the electrostatic interaction between DNA and the inner portal protein. On the other hand however, the larger hydration radius of Mg^{2+} and the resulting larger effective radius of solvated DNA (possibly with Mg^{2+} bound) can have the opposite effect in imposing additional "friction" and hindrance for DNA sliding, thus raising back the barriers that screening may decrease.

As of which of the features of the free energy surface would be affected if Mg^{2+} ions were included in the simulation, it is possible that the computed barrier heights may change (for both translation and rotation concomitantly). However, because the two factors through which the presence of Mg^{2+} can affect the heights (see above) are in opposite directions, it is difficult to estimate by how much, or if the two effects don't actually cancel out. Also, the zig-zag mechanism is likely to still hold in be presence of Mg^{2+} because it is unlikely that Mg^{2+} can alter significantly (by many tens of kcal/mol) just the rotation and not the translation. Hence, while the presence of Mg^{2+} may have minor quantitative effects in modulating the interaction, we expect the overall qualitative picture presented in the Discussion section of the main text, including the zig-zag model and the prediction of lack of overall rotation, to still hold.

References

[1] A.A. Lebedev, M.H. Krause, A.L. Isidro, A.A. Vagin, E.V. Orlova, J. Turner, E.J. Dodson, P. Tavares, and A.A. Antson. Structural framework for DNA translocation via the viral portal protein. *The EMBO Journal*, 26:1984–1994, 2007.

[2] E. Paci and M. Karplus. Forced unfolding of fibronectin type 3 modules: An analysis by biased molecular dynamics simulations. *Journal of Molecular Biology*, 288:441–459, 1999.

[3] B.R. Brooks, R.E. Bruccoleri, D.J. Olafson, D.J. States, S. Swaminathan, and M. Karplus. Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4:187–217, 1983.

[4] A.D. MacKerel Jr., C.L. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus. *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*, volume 1 of *The Encyclopedia of Computational Chemistry*, pages 271–277. John Wiley & Sons: Chichester, 1998.

[5] F. Tama, O. Miyashita, and C.L. Brooks III. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *Journal of Structural Biology*, 147(3):315–326, 2004.

[6] T. Macke. *Nab, a language for molecular manipulation.* PhD thesis, PhD Thesis, The Scripps Research Institute, 1996.

[7] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

[8] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with NAMD. *J. Comput. Chem*, 26(16):1781–1802, 2005.

[9] J.M. Valpuesta and J.L. Carrascosa. Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. *Quarterly Reviews in Biophysics*, 27:107–155, 1994.

[10] G. M. Torrie and J. P. Valleau. Non-physical sampling distributions in Monte-Carlo free-energy estimation - Umbrella sampling. *Journal of Computational Physics*, 23:187–199, 1977.

[11] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. Multidimensional free-energy calculations using the weighted histogram analysis method. *Journal of Computational Chemistry*, 16:1339–1350, Nov 1995.

[12] B. Roux. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications*, 91(1-3):275–282, 1995.

[13] A. Grossfield. source code from http://membrane.urmc.rochester.edu/index.html.

[14] D. N. Fuller, J. P. Rickgauer, P. J. Jardine, S. Grimes, D. L. Anderson, and D. E. Smith. Ionic effects on viral DNA packaging and portal motor function in bacteriophage ϕ29. *Proc. Natl. Acad. Sci. U.S.A.*, 104:11245–11250, 2007.