Solitons and Physics of the Lysogenic to Lytic Transition in Enterobacteria Lambda Phage

Andrei Krokhotin\textsuperscript{1} and Antti J. Niemi\textsuperscript{2,1}

\textsuperscript{1}Department of Physics and Astronomy, Uppsala University, P.O. Box 803, S-75108, Uppsala, Sweden
\textsuperscript{2}Laboratoire de Mathématiques et Physique Théorique CNRS UMR 6083, Fédération Denis Poisson, Université de Tours, Parc de Grandmont, F37200, Tours, France

The lambda phage is a paradigm temperate bacteriophage. Its lysogenic and lytic life cycles echo competition between the DNA binding CI and CRO proteins. Here we address the Physics of this transition in terms of an energy function that portrays the backbone as a multi-soliton configuration. The precision of the individual solitons far exceeds the B-factor accuracy of the experimentally determined protein conformations giving us confidence to conclude that three of the four loops are each composites of two closely located solitons. The only exception is the repressive DNA binding turn, it is the sole single soliton configuration of the backbone. When we compare the solitons with the Protein Data Bank we find that the one preceding the DNA recognition helix is unique to the CI protein, prompting us to conclude that the lysogenic to lytic transition is due to a saddle-node bifurcation involving a soliton-antisoliton annihilation that removes the first loop.

The CI repressor protein and its lytic counterpart, the CRO protein of the Escherichia coli binding $\lambda$ phage, are among the most extensively studied proteins in molecular biology \cite{1,2}. They display a highly intriguing biological behavior by controlling the transitions between the lysogenic and lytic phases, that has been detailed in numerous molecular biology textbooks and review articles. At a qualitative, mechanistic level the transition between the lysogenic and lytic phases is quite well understood \cite{1,2}. But at a quantitative level we do not yet understand the physical principle that triggers the transition. Since lysogeny is an important example of gene control by repressors, a quantitative Physics based explanation should have wide biophysical interest and applicability.

In this Letter we search for a Physics based explanation for the transition between lysogenic and lytic phases in the $\lambda$ phage. For this we scrutinize the fine structure of the folded CI repressor protein, a homo-dimer with 92 residues in each of the two monomers \cite{1}. The protein binds to DNA with a helix-turn-helix motif that is located between the residue sites 33-51. Full crystallographic information is available in Protein Data Bank (PDB) under code 1LMB.

We construct the 1LMB backbone conformation explicitly in terms of solitons that emerge as classical solutions to the following energy function \cite{3-6},

\[
E = - \sum_{i=1}^{N-1} 2 \kappa_{i+1} \kappa_i + \sum_{i=1}^{N} \{ 2 \kappa_i^2 + c \cdot (\kappa_i^2 - m^2)^2 \} + \sum_{i=1}^{N} \{ b \kappa_i^2 \tau_i^2 + d \tau_i + e \tau_i^2 \}
\]  

(1)

The summation extends over all residues with $\kappa_i \in [-\pi, \pi]$ mod(2$\pi$) the bond angle along the lattice that is formed by the central $C_{\alpha}$ carbons, and $\tau_i \in [-\pi, \pi]$ mod(2$\pi$) the ensuing torsion angle. The parameters $(c, m, b, d, e)$ are all global and specific to a given motif. Once these angles are known, we can use the discrete Frenet equation to reconstruct the protein backbone as a piecewise linear polygonal chain.

We emphasize that the energy function (1) does not purport to explain the details of the atomary level mechanisms that fold the CI protein. Rather, it enables us to examine the properties of the folded CI in terms of universal physical arguments. Curiously, (1) has the functional form of the discretized Landau-Ginzburg free energy, that similarly describes the Physics of superconductivity \cite{7}. In a continuum limit the first two terms of (1) combine into derivative of curvature that plays the rôle of Cooper pair density in the Landau-Ginzburg theory. The third term is the symmetry breaking potential. The fourth term has its origin in spontaneous symmetry breaking, its presence leads to the notorious Meißner effect in superconductivity \cite{7}. The fifth term which is absent in the standard Landau-Ginzburg free energy, is the Chern-Simons term that gives the protein backbone its chirality. Finally, the last term is like the Proca mass of a supercurrent. In fact, in (1) we have included \textit{exactly} all those terms that are consistent with general principles of universality and gauge invariance \cite{3}.

We start by introducing the classical equations of motion for (1). We first eliminate $\tau_i$ in terms of the bond angles

\[
\frac{\partial E}{\partial \kappa_i} = 2b\kappa_i^2 \tau_i + 2c \kappa_i + d = 0 \Rightarrow \tau_i[\kappa_i] = -\frac{1}{2} \frac{d}{e + b\kappa_i^2} \equiv -\frac{1}{2} \frac{1}{\frac{e}{d} + \frac{b}{d} \kappa_i^2}
\]  

(2)
Consequently the torsion angles are determined entirely by the bond angles and the two parameter ratios. When we substitute (2) to the equation for $\kappa_i$, we arrive at

$$\kappa_{i+1} - 2\kappa_i + \kappa_{i-1} = U'[\kappa_i] \kappa_i = \frac{dU[\kappa]}{d\kappa_i^2} \kappa_i \quad (i = 1, \ldots, N)$$

(with $\kappa_0 = \kappa_{N+1} = 0$) where

$$U[\kappa] = -\frac{d}{2} \cdot \tau[\kappa] - 2cm^2 \cdot \kappa^2 + c \cdot \kappa^4$$

Since the torsion angle depends only on the parameter ratios $\frac{e}{d}$ and $\frac{b}{d}$, if we scale $d$, $e$ and $b$ equally the profile of $\tau[\kappa]$ remains intact. In the limit where $d$ becomes vanishingly small the first term in (4) can then be safely removed, and the equation reduces to the ubiquitous spontaneously broken discrete nonlinear Schrödinger equation [8]. Thus, in the $d \to 0$ limit the solution approaches the soliton profile of the ensuing continuum equation [9].

$$\kappa_i = (-1)^{r+1} \frac{m_{r1} \cdot e^{c_{r1}(i-s_r)} - m_{r2} \cdot e^{-c_{r2}(i-s_r)}}{e^{c_{r1}(i-s_r)} + e^{-c_{r2}(i-s_r)}}$$

Here $r$ labels the different helix-loop-helix motifs of 1LMB, and $(c_{r1}, c_{r2}, m_{r1}, m_{r2}, s_r)$ are specific to the motif; the parameter $s_r$ that is absent in (1) specifies the location of the $r$th loop. The parameters $(c_{r1}, c_{r2})$ characterize the length of the loop, and $(m_{r1}, m_{r2})$ together with the ratios $(\frac{e}{d}, \frac{b}{d})_r$ determine the global character of the helices and strands that are adjacent to the loops. Remarkably this leaves us with no other loop specific parameters besides the $c_r$ that determine the length of the loops, and $s_r$ that determine their positions.

We propose that as such the classical soliton profiles are duly describing the $C_{\alpha}$ lattice only in the limit where thermal fluctuations vanish. But even near zero temperature the protein remains subject to residual zero-point fluctuations. It is difficult to estimate and even harder to accurately calculate the amplitude of these zero-point fluctuations. As a consequence, in order to get a realistic order of magnitude estimate we have inspected the distribution of the B-factors that characterize experimental uncertainties, for all PDB structures where the crystallographic measurements have been made at temperatures less than 50K. The result is displayed in Figure 1. From it we conclude that for the $C_{\alpha}$ carbons the zero point fluctuations have an amplitude somewhere in the vicinity of the lower bound which is around 0.15 Å. Consequently we describe the estimated range of zero point fluctuations around our classical soliton profiles by dressing them with a tubular dominion that has a radius of 0.15 Å.

FIG. 1: The number of entries in PDB with temperature below 50K vs. Debye-Waller fluctuation distance.
In Table 1 we provide the parameter values for (2) and (5), computed from the PDB data of 1LMB using a Monte
Carlo fitting algorithm. In Figure 2 we compare the ensuing solitons with the folded 1LMB: The solitons describe the structural motifs of 1LMB with a precision that is substantially better than the experimental accuracy determined by B-factors, even when we account for the 0.15 Å estimate of the solitons zero point fluctuations.

With the aid of our high accuracy solitons we conclude that in 1LMB there are a total of seven α-helices and one β-strand. But two of the α-helices and the sole β-strand are so short that until now they have been interpreted as parts of loops. They become exposed only by the high accuracy of our construction. This refinement of the consensus interpretation has important fully testable repercussions to the CI protein that allow us to address the Physics of the lysogenic to lytic transition:

The only motif where our soliton picture identifies a loop as a single isolated soliton is the DNA binding one. All of the remaining three putative loops consist of a soliton-antisoliton pair, with the solitons separated from each other either by a very short α-helix in case of the residues (23, 33) and (69, 90), or by a very short β-strand in case of residues (51, 61). This interpretation reveals itself only when we scrutinize the fine details of the \((\kappa_i, \tau_i)\) spectrum in terms of our solitons. As an example, in Figure 3 we display the putative first helix-loop-helix motif and for comparison we display the corresponding structure in the CRO protein with PDB code 2OVG. Our refined interpretation is palpable, in the case of CI the motif is clearly a bound state of two solitons while in the case of CRO we have a single isolated soliton:

The parameter \(s_r\) in (5) determines the center of soliton i.e. the position of the inflection point in the ensuing space curve where the interpolated bond angle in Figure 3 vanishes. An isolated inflection point such as the one in the right hand side of Figure 3 (2OVG) is topologically stable in the sense that it can not be created nor removed by any continuous local deformation. For a given finite length curve, an individual inflection point i.e. a soliton can be made or deleted only by transporting it through one of the end points of the curve. On the other hand, a pair of inflection points i.e. a soliton-antisoliton pair such as the one in the left hand side of Figure 3 (1LMB) is not topologically stable but can be created or removed locally by a saddle-node bifurcation that brings the two inflection points together.

A comparison between the CI and CRO soliton profiles in Figure 3 then proposes the following experimentally fully testable mechanism for the lysogenic-lytic transition: Under lysogenic conditions where the CI protein prevails, the soliton-antisoliton pairs of the CI protein that are located immediately prior and after the DNA binding domain are relatively stable. But when there is a change in the environmental conditions that excites phonon fluctuations along the protein chain, such as raise in temperature or UV radiation, either of these soliton-antisoliton pairs can discharge by a saddle-node bifurcation. This bifurcation disturbs the structure of the immediately adjacent DNA binding motif to the extent that the protein looses its capability to maintain the lysogenic phase. Since each of the corresponding motifs in the CRO protein are topologically stable single solitons they are insensitive to local phonon excitations, and the lytic phase takes over.

We note that the shoulder of the short α-helix in the last loop is anchored by the presence of a proline at site 78. Consequently in the first approximation we can safely exclude a bifurcation instability from occurring in the putative helix-loop-helix motif between the residues (69, 90).
We still need to conclude which of the two motifs of CI that are adjacent to the DNA binding domain is the one that looses its stability in the proposed bifurcation transition. Unfortunately, it appears that a full answer must wait until computational methods have reached sufficient maturity [9]. However, to provide the probable answer we have performed a statistical analysis on the occurrence of our seven solitons in all PDB proteins. In Table 2 we list

| Soliton | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---------|---|---|---|---|---|---|---|
| Sites   | (20,28) | (27,36) | (36,46) | (50,58) | (55,63) | (66,75) | (74,82) |
| Matches | 9601 | 4 | 810 | 159 | 1552 | 1342 | 406 |

the number of matches that each of these soliton has when we search PDB for configurations that deviate from the given soliton by an overall RMSD distance less than 0.5 Å. We have chosen this cut-off value since it is representative of the Debye-Waller fluctuation distance in the experimental 1LMB data. The remarkable observation is that for the second soliton in the loop preceding the DNA recognition helix, the only matching structures are located in the different PDB entries of the λ phage CI protein itself. This absence of the second soliton in PDB strongly proposes that the ensuing loop must be unstable when the protein is in any other in vivo environment. Thus the most probable source of the lysogenic to lytic transition is the saddle-node bifurcation that takes place in the first loop and makes its soliton-antisoliton pair to annihilate each other. The bifurcation causes the ensuing structure to act like a crowbar that lifts the recognition helix from its place in the DNA groove. We note that as such, it is obvious from (1) that in isolation the first soliton-antisoliton pair has an energy which is higher than that of the ground state i.e. an α-helix. But a more detailed molecular dynamics simulation needs to be performed to confirm our proposal.

All of the other solitons are ubiquitous in PDB and except for the turn that participates directly to the regulatory process their biophysical role remains to be clarified.

Finally, our soliton interpretation reveals the following fully testable pattern for the folding pathways of 1LMB: The functionally pertinent DNA binding loop is an isolated soliton, while all the remaining three loop structures consist of a soliton-antisoliton pair. Since an isolated soliton is topologically stable while soliton-antisoliton pairs are not, the DNA binding loop must be created very early, presumably during translation. The initial configuration for the folding process is then a single soliton state. During the folding process the remaining three motifs are created by local phononic fluctuations, as soliton-antisoliton pairs. Due to the presence of the proline at site 78, the ensuing motif probably emerges very early.

A.N. thanks H. Frauenfelder and G. Petsko for communications and J. Áqvist for discussions.

[1] M. Ptashne, A Genetic Switch, Third Edition, Phage Lambda Revisited. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2004)
[2] M.E. Gottesman and R.A. Weisberg, Microbiol. Mol. Biol. Rev. 68 796 (2004)
[3] A.J. Niemi, Phys. Rev. D67 106004 (2003)
[4] U.H. Danielsson, M. Lundgren and A.J. Niemi, Phys. Rev. E82 021910 (2010)
[5] M. Chernodub, S. Hu and A.J. Niemi, Phys. Rev. E82 011916 (2010)
[6] N. Molkenthin, S. Hu and A.J. Niemi, Phys. Rev. Lett. 106 078102 (2011)
[7] P.G. de Gennes, Superconductivity of Metals and Alloys, Perseus Books, 2nd Revised Edition (1995)
[8] P.G. Kevrekidis, The Discrete Nonlinear Schrödinger Equation: Mathematical Analysis, Numerical Computations and Physical Perspectives Springer-Verlag, Berlin (2009)
[9] P.L. Freddolino, C.B. Harrison, Y. Liu and Y. Schulten, Nature Phys. 6, 751 (2010)