

The Poland-Scheraga Model

8.334 Final Project

Simon Lieu

May 14, 2015

Abstract

The Poland-Scheraga (PS) Model is one of the most celebrated applications of statistical physics to biology. By relying on basic laws of statistical phase transitions the PS model describes the denaturation of DNA strands as a function of temperature according to a minimal set of assumptions. The basic model makes the prediction of a second-order phase transition in the order parameter of the system (i.e. the fraction of the monomers which are bound to each other) in 3 dimensions, yet experimental observations suggest a first order transition. Thus an active area of research in the past few decades has involved complicating the PS model in such a way as to obtain the observed first-order phase transition. In this review, we describe the basic PS model and its successes/failures; then we extend the model in a novel direction, examining the effects of a term that suppresses long bound chains; finally we review a recent paper which describes how incorporating “excluded volume effects” into the model results in a first-order transition (in agreement with experiment).

1 Description of the Model

1.1 Introduction

DNA is composed of two strands which, at low temperatures, form the famous double-helix structure with G-C and A-T bonds binding the monomers together. However at higher temperatures the bonds begin to break and “denaturated loops” start to form. Thus at a given temperature, the two strands may either be bound or unbound at a given site (as depicted in the image below). If the monomers are unbound, they are able to form loops that traverse space but come back to the same point where they become bound again. The bound state has low energy but no entropy. The loop state has high energy but carries a lot of entropy. The order parameter in this transition is the fraction of sites which are form loops which is clearly 0 at low temperatures and 1 at high temperatures



Thus one may view the phase transition of DNA denaturation as the competition of energy vs. entropy, similar to the Ising model transition.

1.2 Predictions of the Basic Model

In this section we closely follow Prof. Kardar’s 8.592 lecture notes [3]. We begin examining this model by attempting to find the partition function of the system. The probability of a given configuration of bound sites and loops (i.e. l_1 bound sites, loop of length l_2 , l_3 bound sites, etc.) is:

$$p(l_1, l_2, l_3, \dots) = \frac{B(l_1)L(l_2)B(l_3)\dots}{Z} \quad (1)$$

where $B(l)$ is the probability weight of a bound strand of length l , $L(l)$ is the probability weight of a loop of length l , and Z is the full partition function (summing over all possible configurations). As mentioned in the previous section, we set the bound sites to have lower energy than the loops. Thus

$$B(l) = e^{-\beta\epsilon l} \equiv w^l \quad (2)$$

The weight of the loops is more complicated and involves solving the problem of how many different random walks of length l are there between two points separated a distance \mathbf{r} . Here we cite the result:

$$L(l) = \frac{g^l}{l^c} \quad (3)$$

where $c = d/2$ and g is a non-universal constant that controls the critical temperature of the phase transition.

It is easiest to find the partition function in the grand canonical ensemble by introducing a “fugacity”

that ensures the average length is L . Our grand partition function is:

$$\mathcal{Z}(z) = \sum_{L=1}^{\infty} z^L Z(L) \quad (4)$$

Now we can split the grand partition function calculation up into a product of independent sums:

$$\mathcal{Z}(z) = \left(\sum_{l_1=1}^{\infty} z^{l_1} B(l_1) \right) \left(\sum_{l_2=1}^{\infty} z^{l_2} L(l_2) \right) \dots \quad (5)$$

From the previous paragraph we can calculate explicit expressions for $B(z)$ and $L(z)$:

$$B(z) = \sum_{l=1}^{\infty} (zw)^l = \frac{zw}{1-zw} \quad (6)$$

$$L(z) = \sum_{l=1}^{\infty} \frac{(zg)^l}{l^c} \equiv f_c^+(zg) \quad (7)$$

where f_c^+ is known as the polylog function.

Now we must sum over all possible combination bound sites and loops between the two ends of the DNA:

$$\mathcal{Z}(z) = B(z) + B(z)L(z)B(z) + \dots \quad (8)$$

which is just a geometric series:

$$\mathcal{Z}(z) = \frac{B(z)}{1-L(z)B(z)} = \frac{1}{(zw)^{-1} - 1 - f_c^+(zg)} \quad (9)$$

Thus we may now find the expectation value of the length of the DNA strand and the average number of bound sites:

$$\langle L \rangle = z \frac{\partial}{\partial z} \log \mathcal{Z}(z) = \frac{\frac{1}{zw} + f_{c-1}^+(zg)}{\frac{1}{zw} - 1 - f_c^+(zg)} \quad (10)$$

$$\langle N_B \rangle = w \frac{\partial}{\partial w} \log \mathcal{Z}(z) = \frac{\frac{1}{zw}}{\frac{1}{zw} - 1 - f_c^+(zg)} \quad (11)$$

where we have used the fact that $z \frac{df_c^+(zg)}{dz} = f_{c-1}^+(zg)$. We take the ratio of these two quantities to obtain the order parameter:

$$\Theta = \frac{\langle N_B \rangle}{\langle L \rangle} = \frac{1}{1 + zw f_{c-1}^+(zg)} \quad (12)$$

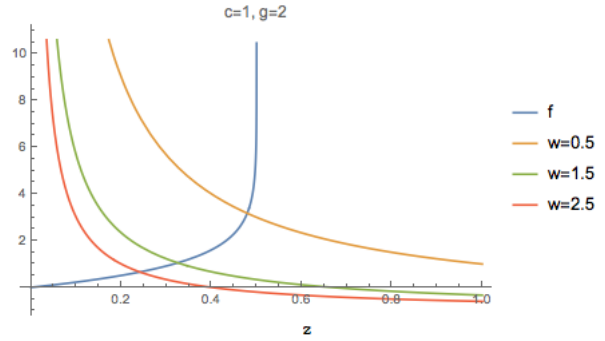
In the thermodynamic limit both L and N_B tend to infinity. This implies we can solve for z by setting $\frac{1}{zw} - 1 - f_c^+(zg) = 0$, or

$$f_c^+(zg) = \frac{1}{zw} - 1 \quad (13)$$

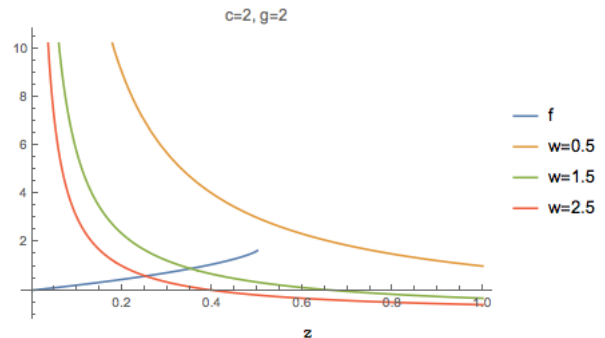
Clearly the value of c must be specified in order to solve for z and consequently plot the order parameter

as a function of the temperature T (which w depends on in the above expression).

For $c < 1$ the left and right hand side of Eqn 13 is plotted in the image below. As one can see from the graph, as w decreases (i.e. the temperature increases) the intersection of the two curves moves closer and closer to the value of $z = 1/g$; however for any finite temperature the equilibrium z value is less than $1/g$. The denominator in Eqn 12 remains finite so long as $zg < 1$ so for $c < 1$ it is clear the order parameter Θ varies smoothly between 0 and 1.

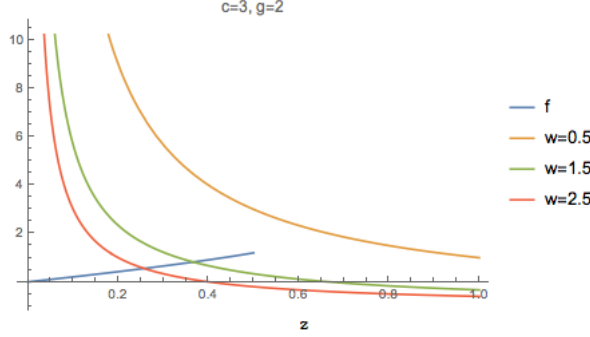


For $1 \leq c \leq 2$ the left and right hand side of Eqn 13 is plotted in the image below. The graph of f abruptly terminates at $gz = 1$ which implies there is a critical value for w below which z is fixed at $1/g$. From the previous paragraph we know that for $1 \leq c \leq 2$, f_{c-1}^+ diverges as its argument goes to 1. Since Eqn 12 has $1 \leq c \leq 2$ in the denominator, Θ assumes a value of 0 until the temperature reaches $w(T_c) = w_c$. Once the temperature reaches T_c , any drop in temperature will smoothly increase the value of the order parameter until all monomers are bound at $T = 0$. Thus for $1 \leq c \leq 2$ there is a second order transition in Θ .



For $c > 2$ the left and right hand side of Eqn 13 are again plotted below. Once again the graph of f abruptly terminates at $gz = 1$ which implies there is a critical value for w below which z is fixed at $1/g$. However for $c > 2$ we know from the previous paragraph that for values of z slightly below $1/g$, the value

of $f_{c-1}^+(zg)$ jumps from ∞ to a finite value abruptly. Thus the denominator of Eqn 12 jumps from ∞ to a finite value, resulting in a first order phase transition in Θ .



2 Extension of the Model

One new avenue that has yet to be investigated is what happens to the PS model if one includes a term that algebraically suppresses long chains of bound segments. We do this by altering the weight of a bound segment $B(l) = w^l \rightarrow w^l/l^x$. One might expect such a term if there were some external influence in the medium in which the DNA propagates that encourages the break-up of long bound segments. Intuitively one would expect that adding such a term would smooth out the transition, since such a term would penalize long loops even at low temperatures.

We proceed to follow the steps of the previous section. The weight of $B(z)$ is now another polylog function: $B(z) = f_x^+(zw)$. Carrying through the analysis we find:

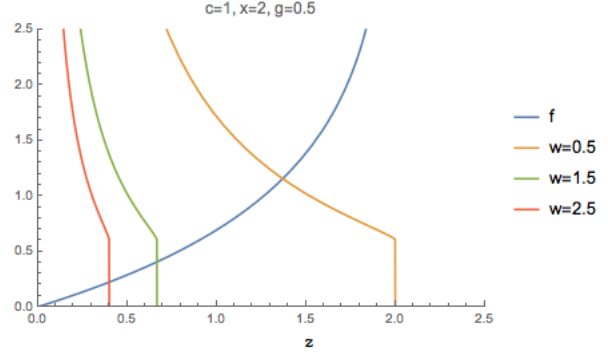
$$\langle L \rangle = \frac{\frac{f_{x-1}^+(zw)}{f_x^+(zw)^2} + f_{c-1}^+(zg)}{f_x^+(zw)^{-1} - f_c^+(zg)} \quad (14)$$

$$\langle N_B \rangle = \frac{\frac{f_{x-1}^+(zw)}{f_x^+(zw)^2}}{f_x^+(zw)^{-1} - f_c^+(zg)} \quad (15)$$

$$\Theta = \frac{f_{x-1}^+(zw)}{f_{x-1}^+(zw) + f_x^+(zw)^2 f_{c-1}^+(zg)} \quad (16)$$

Once again we take the thermodynamic limit by setting the denominator to zero and solving for z , i.e. $f_x^+(zw)^{-1} = f_c^+(zg)$. The value of z is set determined by setting c and x .

Let's examine the following graph plotting the interaction of $f_x^+(zw)^{-1}$ and $f_c^+(zg)$, where $c = 1$, $x = 2$, $g = 0.5$



From the diagram we can deduce a couple of features: 1) for temperatures above a value $w(T_F)$ (such as the $w = 0.5$ graph above), the intersection of the two graphs does not lie on the vertical part of f_x and consequently z varies smoothly with the temperature 2) for temperatures such that $w(T) > w(T_F)$ the equilibrium z value is set to $1/w$ (the location of the kinks in the picture above). When $z = 1/w$ then the order parameter Θ is fixed at 1 (since $f_{x-1}^+(1) = \infty$ regardless of x , and both numerator and denominator have this term). Thus below a certain ‘‘frozen temperature’’ Θ is held at 1. As one can clearly see from the graph, the order parameter varies smoothly (for $c = 1$) from the value of 1 at $w(T_F)$ to 0, since the factor of $f_{c-1}^+(zg)$ in the denominator makes Θ go to 0 as z tends to $1/g$.

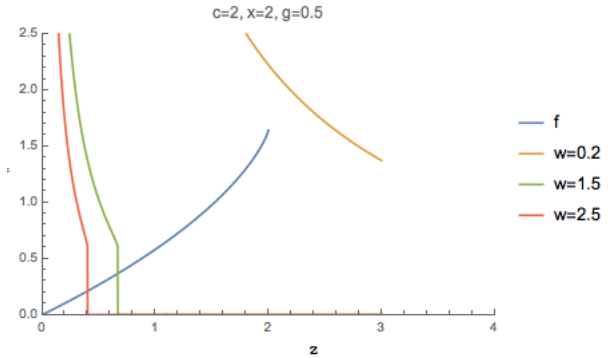
The ‘‘frozen temperature’’ is set by the following criterion:

$$f_x^+(zw(T_F))^{-1} = f_c^+(zg) \Big|_{z=\frac{1}{w(T_F)}} \quad (17)$$

Or, rewriting it in terms of the Riemann-zeta function:

$$f_c^+ \left(\frac{g}{wT_F} \right) = \zeta_x^{-1} \quad (18)$$

For $c > 2$. Consider the following graph (parameters labeled):



Once again we find that the second order phase transition is unchanged by inclusion of the long-loop

penalty. Also we find a “frozen temperature” below which the order parameter is fixed to 1.

We summarize the two main findings of including a term that suppresses long chains:

- (a) There are two special temperatures: the critical temperature which is where the order parameter begins to be non-zero, and the “frozen” temperature below which the order parameter is always 1.
- (b) The critical temperature (where the order parameter is first non zero) is effected by the inclusion of this long-loop penalty, however the types of transitions (i.e. first/second order) are not.

3 Excluded Volume Effects

From the previous sections we have seen that although the PS model is successful in predicting a phase transition for the DNA denaturation, it does not predict the type of discontinuous phase transition that is observed experimentally via gel electrophoresis (as depicted below) [4].

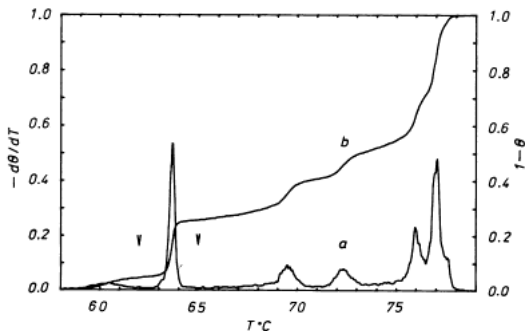
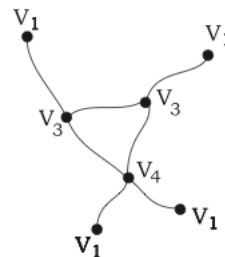


Figure 1 Curve (b) in the figure above shows how the order parameter (here defined as $1 - \Theta$ in our notation) changes as a function of temperature. Curve (a) is the derivative of curve (b). Note the delta function-like peak in the curve of (a) near 64 C and a subsequent jump in the order parameter, implying a first-order phase transition.

Although there are many ways to alter the aforementioned model (such as the one presented in the previous section), the most promising avenue to obtain a first-order transition stems from including self-avoidance. Up to now we have assumed that for the denaturated loops execute random walks; strictly speaking this is an approximation, since the DNA strands are not allowed to fold back onto themselves. The self-avoiding random walk is a much more difficult problem. The number of configurations scales as: $\Gamma \sim s^L L^{\gamma-1}$, a similar expression to Eqn 3, where

s is a non-universal constant and γ is a universal exponent that depends on the number of dimensions of the random walk. The result of this analysis is that the value of c goes from 1.5 to 1.8 in three dimensions, thus making the transition sharper (i.e. more closely resembling a first-order transition).

In this section we briefly review highlights of a recent paper by Kafri, Mukamel, and Peliti which incorporated self-avoidance not only for individual loops, but also between the rest of the (bound) chain and the loops [1,2]. The end result is that the value of c changes yet again to 2.1, resulting in a first-order transition that agrees with experiment.



For a general topology (as depicted above) composed of N chains of lengths l_1, l_2, \dots, l_N such that $L = \sum_i l_i$, the expression for the scaling of number of configurations is given by

$$\Gamma_G \sim s^L L^{\gamma_G-1} g\left(\frac{l_1}{L}, \frac{l_2}{L}, \dots, \frac{l_N}{L}\right) \quad (19)$$

$$\gamma_G = 1 - \nu d \mathcal{L} + \sum_{k \geq 1} n_k \sigma_k \quad (20)$$

where g is a scaling function that is smooth as long as its arguments are finite, ν is the radius of gyration exponent, \mathcal{L} is the number of loops, d is the number of dimensions, n_k is the number of vertices with k edges, and $\sigma_k = (2 - k)(9k + 2)/64$ for $d = 2$.



Taking the results of the previous paragraph for granted, we may now specifically examine the case of a single loop embedded in a chain (depicted above). This topology takes into account the interaction of a given loop with the rest of the chain (neglecting its structure). We are interested in finding the number of configurations in the limit that the loop size is much smaller than the rest of the chain, i.e. $l \ll L$. From the results of the previous paragraph:

$$\Gamma \sim s^{L+l} (L + l)^{\gamma-1} g(l/L) \quad (21)$$

$$\gamma = 1 - d\nu + 2\sigma_1 + 2\sigma_3 \quad (22)$$

where we have used that fact that there are two vertices with 1 edge and 2 vertices with 3 edges, and only one loop. Now we impose the condition that in the limit as $l/L \rightarrow 0$, Γ must be equivalent to the expression for the general self avoiding walk of length L between two vertices: $s^L L^{\gamma'-1}$ for $\gamma' = 1 + 2\sigma_1$. This allows us to find the asymptotic value of the scaling function g . By matching these two expressions, we find $g(x) \sim x^{\gamma-\gamma'}$ in the limit $x \ll 1$. In this limit we find:

$$\Gamma \sim s^{L+l} L^{\gamma-1} \left(\frac{l}{L}\right)^{\gamma-\gamma'} = s^{L+l} L^{\gamma'-1} l^{\gamma-\gamma'} \quad (23)$$

Thus we have found that the total number of configurations decomposes into a product of the loop configurations and the rest of the chain. By incorporating interacting loops into our analysis we have found the effective exponent to be:

$$c = \gamma' - \gamma = d\nu - 2\sigma_3 \quad (24)$$

In two dimensions, $d = 2, \sigma_3 = -29/64$, and $\nu = 3/4$. Thus $c = 2 + 13/32 > 2$ and we have demonstrated that by including the interaction of a loop with the rest of the chain, a first order phase transition occurs in two dimensions—in agreement with experiment.

4 Conclusion

In this short review paper we have summarized the main assumptions, approximations, and results of the well-known Poland-Scheraga model. After demonstrating its failure to predict a first order phase transition in three-dimensions, we presented a novel complication to the model which involved suppressing long-loops by incurring an algebraic energy cost. After demonstrating that inclusion of this energy cost was not sufficient to change the nature of the DNA denaturation transition, we followed through the derivation of Kafri, Mukamel, and Peliti to demonstrate that excluded volume effects could be used to demonstrate a first order phase transition for the two-dimensional problem.

5 References

[1] Duplantier, Phys. Rev. Lett. 57, 941 (1986); J. Stat. Phys. 54, 581 (1989).

[2] Kafri, Y., D. Mukamel, and L. Peliti. "Melting and Unzipping of DNA." The European Physical Journal B - Condensed Matter 27.1 (2002): 135-46. Web.

[3] Kardar, Mehran. "Statistical Physics in Biology." MA, Cambridge. 15 May 2015. Web..

[4] Lyamichev, V.i., I.g. Panyutin, and Yu.l. Lyubchenko. "Gel Electrophoresis of Partially Denatured DNA. Retardation Effect: Its Analysis and Application." Nucl Acids Res Nucleic Acids Research 10.15 (1982): 4813-826. Web.