Utilizing Statistical Mechanics and Thermodynamics In Determining Nucleosome Positioning along DNA

Christian Rose*
UCLA Summer REU
Department of Physics and Astronomy
University of California, Los Angeles,
Los Angeles, CA 90095 USA
(Dated: October 8, 2006)

Various methods have been employed to determine the density and spacing of nucleosomes along DNA via thermodynamic means. Building from previous models this paper focuses on the use of J.K. Percus' mathematical solution to the Equilibrium State of a Classical Fluid of Hard Rods in an External Field.

THEORY AND BACKGROUND

Each cell in the human body contains over one meter of DNA. It has long been understood that various compaction methods must be employed to keep the DNA localized in the nucleus of the cell. One such compaction method is through the use of nucleosomes, a structure consisting of 1.75 left-hand turns of DNA around 8 core histone proteins in a super helix.

It has been shown that the positioning of these nucleosomes is non-random and exhibits some sequence specificity in the interaction between the histone proteins and the nucleic acid base pairs. The purpose of much of the previous work has been to determine whether the spacing is purely statistical, obeying the laws of thermodynamics in its positioning or whether it is the specificity of the base pair sequencing that determines spacing.

EARLY APPROACHES

We begin our research into nucleosome positioning by taking everything we know of the system into account. If we are to believe a purely statistical mechanics and thermodynamics approach, then we must build an appropriate model.

First, we consider the system of nucleosomes along DNA and determine that the dyadic axis will be our x position from the origin. We may then say that a nucleosome may be positioned anywhere from the origin to x. Therefore the next nucleosome may be positioned from x to x', the third may be positioned from x' to x", and so on. Since two nucleosomes cannot overlap we say that:

$$x_2 - x_1 > a \tag{1}$$

Where a is the length of a nucleosome, or the distance between dyadic axes. Since the particles are indistinguishable, we know that the partition function will have the form:

$$Z = \frac{1}{N!} Z_1^N \tag{2}$$

And Z will be found by summing the probablility that each particle will be found over some distance along the DNA suth that:

$$Z = \int_0^L dx_1 \int_0^L dx_2 ... \tag{3}$$

But we know that each nucleosome cannot go all the way to L, but only L-na, and we can also substitute x = a + y. Proceeding accordingly we find:

$$Z = \int_0^{L-an} dy_1 \int_0^{L-a(n-1)} dy_2 ... \int_0^{L-a(n-n)} dy_n \tag{4}$$

And remembering that L - a(n - 1) = y1 we obtain:

$$Z = \int_0^{L-an} dy_1 \int_0^{y_1} dy_2 ... \int_0^{n-1} dy_n \tag{5}$$

Finally, knowing that the volume of an n-dimensional hyperspace is:

$$(L - na)^n = \int_0^{L-na} dy_1 \int_0^{L-na} dy_2 ... \int_0^{L-na} dy_n \tag{6}$$

So,

$$Z((n)(n-1)(n-2)...) = (L - na)^n \tag{7}$$

$$Z = \frac{(L - na)^n}{n!} \tag{8}$$

Now that we have obtained an expression for the partition function, we may use this information for the basis to the grand partition function. We may set up the grand partition function in the form:

$$\mathcal{Q}(x) = \sum_{N<La} \frac{L - an)^N}{N! a^N} e^{\beta \mu N} \tag{9}$$

Where β and μ are the traditional thermodynamic properties for inverse temperature and chemical potential. In this case we know that Sterling's approximation may be used to find a more discrete value for $\mathcal{Q}(x)$.

DENSITY

The purpose of determining $\mathcal{Q}(x)$ is so that it may be used to find the density of states. We can see that the sum in Eq. 9 will be dominated by the N value for which the summand is a maximum and therefore we may utilize the method of steepest descent to find that term. Then rearranging terms and expressing in terms of the density $\rho=\frac{N}{L}$:

$$\exp[L(\rho \ln \frac{1-a\rho}{a\rho} + \rho + \rho\beta\mu)] \tag{10}$$

Then, taking the derivative in order to maximize,

$$\ln \frac{1-a\rho}{\rho} - \frac{a\rho}{1-a\rho} + \beta\mu = 0 \tag{11}$$

By introducing the new variable $h = \frac{a\rho}{1-a\rho}$ and solving, we find that the grand partition function becomes:

$$\frac{\exp[L(L(\rho \ln \frac{1-a\rho}{a\rho} + \rho + \rho\beta\mu))]}{1+h} \tag{12}$$

With ρ replaced by it's value at the extremum. In this case, the full form of Sterling's plays an important role in the solution to $\mathcal{Q}(x)$.

Dependence on position from the end of an interval

To determine the density, we must first fix the position of one of the rods at a position x_0 and then calculate the grand partition function for all of the other rods consistent with the location of x_0. The density, $\gamma(x)$, is then given by the equation:

$$\gamma(x) = \frac{1}{a}\frac{\mathcal{Q}(x-a)\mathcal{Q}(L-x)}{\mathcal{Q}(L)}e^{\beta\mu} \tag{13}$$

Where a is still the length of a rod and $e^{\beta\mu}$ represents the fact that every rod carries an $e^{\beta\mu}$. Now, solving for any arbitrary x we find that:

$$\gamma(x) = \sum_{N+1<\frac{x}{a}} \frac{(x-(N+1)a)^N}{a^N N!} e^{(N+1)\beta\mu} e^{-xF_0} \tag{14}$$

This equation holds for any value of a or $\beta\mu$ but in this case we can see how the density becomes indeterminable deep in the bulk when we consider $a = 1$ and $\beta\mu$=3.5.

UTILIZING PERCUS' EQUATIONS

According to J.K. Percus, we know that the general solution to the equilibrium state a fluid of one-dimensional hard cores (as we model our system) is:

$$\beta\mu(x)+\ln\rho(x)-\ln z = \ln[1-\int_{x-a}^{x}\rho(\omega)d\omega]-\int_{x}^{x+a}\frac{\rho(z)}{1-\int_{z-a}^{z}\rho(\omega)d\omega}dz \tag{15}$$

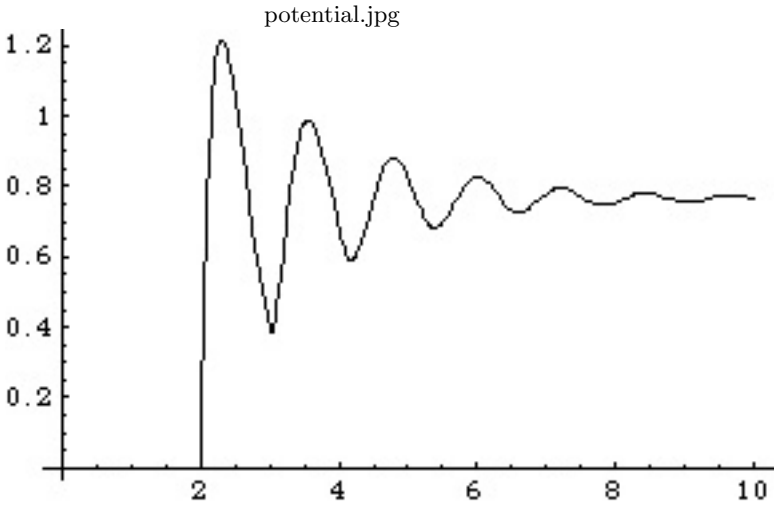

FIG. 1: Density of rods as given by (**??**) with $\beta\mu = 3.5$.

However, this solution does not help us to find much of anything so we must determine a way to solve the equation for the variables that we would like to control. In order to do so, we followed the method provided by Vanderlick et. al. to th Percus equations, separating the equation into two unknown variables, h and x.

$$\frac{dh(x)}{dx} = -h(x)[\beta\frac{d\phi(x)}{dx} + h(x) - h(x-\sigma)] \tag{16}$$

$$\frac{dl(x)}{dx} = -h(x)l(x) - h(x+\sigma)l(x+\sigma) \tag{17}$$

The key to solving these equations was to recast them into a solvable form. After making substitutions and rearranging the limits of integration we find that,

$$h(x) = \frac{e^{-\beta\phi(x)} + \int_{x0}^{x} h(x'-\sigma)dx'}{\frac{e^{-\beta\phi(x)}}{h(x_0)} + \int_{x_0}^{x} e^{-\beta\phi(x'') + \int_{x_0}^{x''} h(x'-\sigma)dx'}dx''} \tag{18}$$

Once we have $h(x)$ we can solve for $l(x)$, and, using a similar method as previously described we find that,

$$l(x) = l(x_0)e^{\int_{x_0}^{x} h(x')dx'} - \int_{x_0}^{x} e^{-\int_{x}^{x''} h(x')dx'} h(x''+\sigma)l(x''+\sigma)dx'' \tag{19}$$

We can solve the equation for $h(x)$ by looking at successive intervals on the x axis. Then, once we have a solution for $h(x)$ the solution for **??** gives a way of generating $l(x)$ in successive intervals in the opposite direction from how we found $h(x)$. This way we can still find a density for nucleosomes at the opposite end of the DNA strand.

This method for solving for the density was then translated to Mathematica so that a distribution for the density of nucleosomes could be determined. In Mathematica, performing numerous calculation over an interval length of our choosing, we could graph the density and

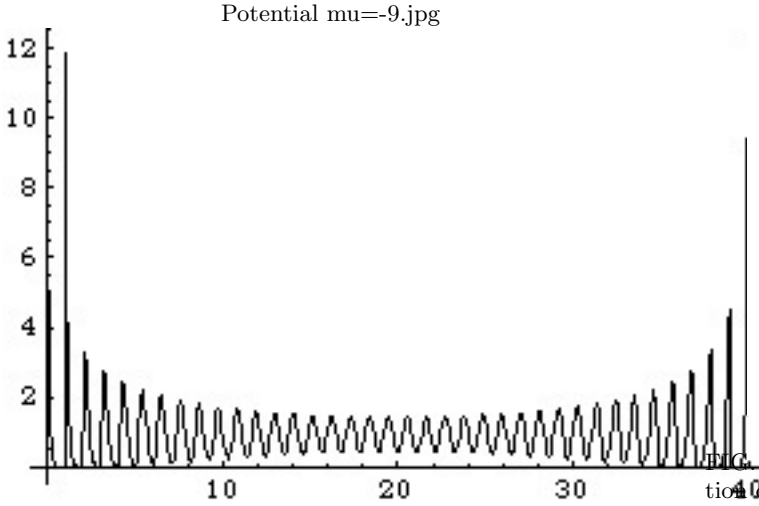

FIG. 2: Density of Nucleosomes as given by Percus using Venderlick method with $\beta\mu = -9$.

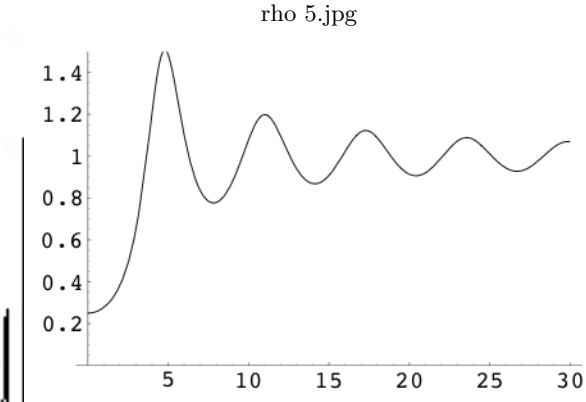

FIG. 3: The solution for $B/\rho_0\beta A$ when $\sigma\rho_0 = 0.5$ as a function of $k\sigma$.

see that it is, in fact, oscillatory with the amplitude of the oscillations evening out toward the bulk area of the interval.

Sinusoidal Potential

Since determining the probability density of nucleosomes at locations along DNA was straightforward with respect to a constant potential we next looked into the case of a sinusoidal external potential acting on the nucleosomes. In a natural system like that of nucleosomes in a cell it is more likely that the external potential will oscillate and, thus, it is important to make sure that our method holds up for oscillating potentials.

In this case I replaced the constant potential directly in Mathematica and then observed the resulting density distribution for different amplitudes of the potential as well as a range of values for $\sin \pi x$. An example of the probability density deep within the bulk is shown in the following figure.

Insert figure Potential with Sin(1/4 pi).jpg Probability density of nucleosomes deep within the bulk from a sinusoidal potential.

Once we know that our method works for a sinusoidal potential, the next step is to analyze the response of the density to a sinusoidal potential in order to understand the effects of the potential on the positioning of nucleosomes.

Here we assume that the potential is small so that we can make use of linear response approximations. First, we rewrite the Percus equation with variables consistent with the Vanderlick method

$$\beta\mu(x)+\ln(\sigma\rho(x)-\beta\mu = \ln[1-\int_{x-\sigma}^{x}\rho(x')dx']-\int_{x}^{x+\sigma}\frac{\rho(x')}{1-\int_{x'-\sigma}^{x'}\rho(x'')d} \tag{20}$$

Next, we assume that when there is no sinusoidal potential we can neglect $\mu(x)$ and $\rho(x) \to \rho_0$, so Eq. **??** becomes

$$\ln\sigma\rho_0 - \beta\mu = \ln(1-\sigma\rho_0) - \frac{\sigma\rho_0}{1-\sigma\rho_0} \tag{21}$$

Which has the implicit solution $\sigma\rho_0 = \frac{h_0}{1+h_0}$, where

$$h_0 = e^{\beta\mu - h_0} \tag{22}$$

The is the same equation for the density deep in the interior of the hard-rod gas that we calculated earlier. We can now expand the density using

$$u(x) = Ae^{ikx} \tag{23}$$

$$\rho(x) = \rho_0 + Be^{ikx} \tag{24}$$

By plugging these back into Eq. **??** and then solving, we find that

$$\frac{B}{\rho_0} = -\frac{\beta A}{1+2\frac{\sigma\rho_0}{1-\sigma\rho_0}\frac{\sin k\sigma}{k\sigma}+2(\frac{\sigma\rho_0}{1-\sigma\rho_0})^2\frac{1-\cos k\sigma}{(k\sigma)^2}} \tag{25}$$

Plotting this solution for different values of $\sigma\rho_0$ we see that for $\sigma\rho_0 = 0.5$ we get

Or if we use $\sigma\rho_0 = 0.9$ we can see that

Finally, if we plot the smallest value of $k\sigma$ at which there is a maximum in the response as a function of $\sigma\rho_0$

It is important to note that the maximum occurs when $\sigma\rho_0 = 1$.

rho 9.jpg

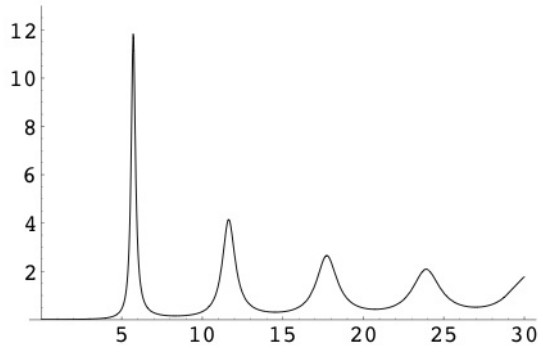

FIG. 4: The solution for $B/\rho_0\beta A$ when $\sigma\rho_0 = 0.9$ as a function of $k\sigma$.

linear fig 3.jpg

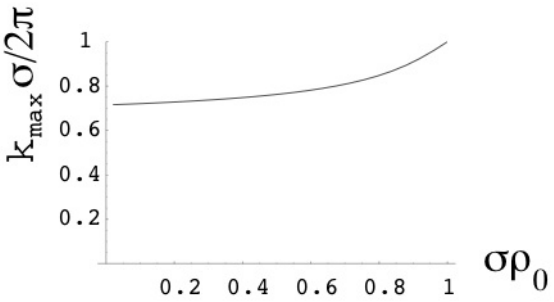

FIG. 5: $k_max\sigma/2\pi$ vs. $\sigma\rho_0$.

FUTURE WORK

After completing the Mathematica program that worked to our liking for both constant and sinusoidal external potentials, the final step is to move on to discrete data for the external potential of the chemical potential of yeast DNA base pairs. In doing so there is little change that must be made to the method that we have already developed. In Eqs. **??** and **??** the integral is replaced with sum for the discrete values of the potential at particular base pairs in reference to their distance from the beginning of the DNA strand.

ACKNOWLEDGMENTS

I would like to thank Professor Joseph Rudnick for his tireless work with the Mathematica files as well as his countless hours offered toward helping me to understand the proper methods for solving physics problems by understanding the mathematical principles that determine their solution. I would also like to thank Professor Robijn Bruinsma for his ability to clarify and illuminate the various ways in which physics may be applied to the solution of biological phenomena. Finally, I would like to thank the NSF for funding this research project which has ultimately lead me to explore areas of biophysics and myself that I did not know existed.

* chrose@vassar.edu