Lecture 09: Model of Dimerization of Single-Stranded DNA

At very high temperatures double stranded DNA can split up into two single strands. When the temperature is lowered again, the two single strands re-associated to form double stranded DNA. In these notes we use the methods we have developed to this point to model how the average number of double-stranded DNA in a collection of single-stranded DNA varies as a function of temperature. The model is an approximate one, but it does well to get us in the ball park of the melting temperature of DNA.

1 Self-assembly in biology

Viruses mostly consist of containers made from protein macromolecules which enclose highly packed genetic material. What is particularly interesting about such a structure is that, once all the component protein parts exist, the structure appears to assemble spontaneously within a cell, without any external input of work. That is, once once brings all of the individual component proteins of the shell together, then not only do these components manage to find one another and combine into a large structure, they also seem able to distinguish correct from incorrect contacts and *correctly* assemble into the larger structure. It appears that the molecular biology of such virus proteins is so finely turned that the structures essentially make themselves!

There are many other self-assembled systems in the cell. Multiple proteins can come together spontaneously to assemble into a larger conglomerate protein. Phospholipids can come together to form (depending on their type) cellular membranes or spherical shells. The ubiquity of such processes suggests that **self-assembly** is a motif of small scale biological systems, and the fact that these processes occur spontaneously suggests that they should be comprehensible through statistical physics.

We try to build such an understanding by considering a simpler and somewhat biologically artificial process of individual subunits coming together in pairs[1]. The genetic material of living organisms is made up of DNA which itself is comprised of two strands of nucleic acids bound together in the form of a helix. When DNA is taken out of the cell, it can be heated up to very high temperatures, temperatures at which the DNA separates into its two single-stranded components. If we were to cool the new collection of single-stranded DNA back to lower temperatures, the strands would reassemble into their correct double stranded DNA form. Written as a reaction equation this process is

$$\text{ssDNA}_1 + \text{ssDNA}_2 \rightleftharpoons \text{dsDNA}_{12} \tag{1}$$

where ssDNA = single-stranded DNA and dsDNA = double-stranded DNA. The left and right arrows, $\rightleftharpoons$, indicate that the double-stranded DNA can dissociate into two single-stranded DNA, or the two single-stranded DNA can combine to form a double-stranded DNA. This process is sufficiently simple that we should be able to model some aspects of it using the statistical physics we have developed so far. Understanding the equilibrium thermodynamics of the above process will be the objective of these current notes. In particular we would like to find a way to determine, as a function of temperature, the fraction of DNA strands which exist in their double stranded form versus their single stranded form.

Framing Questions

How can we use equilibrium statistical physics to determine how the proportion of strands in dsDNA (relative to the total number of strands) in the reaction Eq.(1) varies

[1]This process is biologically artificial because there is no system in the cell in which single-stranded DNA combines with its complement to form double stranded DNA. But this process can well model what happens during various steps of a **polymerase chain reaction**.

with temperature?

2 Beginnings of the Model

(a) (b)

Figure 2: (a) The basic reaction depicting two single strands of DNA combining to form double-stranded DNA or double-stranded DNA dissociating into a pair of single-stranded DNA. (b) The system of interest–a collection of single stranded DNA and double stranded DNA confined to a volume V and temperature T. At thermal equilibrium the strands are variously combining and dissociating and have macroscopic properties that can be computed from equilibrium statistical physics.

Before we get to the mathematical formalism underpinning the reaction Eq.(1) (depicted in Fig. **??**), we make some qualitative assumptions which render our model less general but more soluble. The individual strands which make up double-stranded DNA are not in general identical. Rather they are **complementary** to one another, each one consisting of a specific sequence of DNA base pairs that uniquely binds to the specific, but generally different, sequence of the other. This complementarity is defined on a per base-pair level, where the base pairs A and T are complementary and base pairs G and C are complementary. A single strand of DNA is defined with a starting end and a finishing end such that a sample sequence is written as[2]

$$\text{start-A-T-A-G-T-C-C-end.} \tag{2}$$

For double-stranded DNA, the "start" end of one single-stranded DNA is adjacent to the "end" end of the other. Thus, the sequence complementary to Eq.(2) would be

$$\text{start-G-G-A-C-T-A-T-end,} \tag{3}$$

and the full double-stranded DNA would be depicted as

$$\begin{array}{c}\text{start-A-T-A-G-T-C-C-end}\\ \quad|\ |\ |\ |\ |\ |\ |\ |\\ \text{end-T-A-T-C-A-G-G-start}\end{array}$$

where the vertical lines stand for hydrogen bonds between the DNA base pairs. Thus, for each double-stranded DNA, there are two different sequences defining its individual single strands.

This reality is too complicated for us in these notes. So instead of having to contend with many different sequences we will consider a collection of single-stranded DNA all of which have the same sequence. In order for such identical sequences to be able to form double-stranded DNA, they must be complementary to themselves (i.e., **self-complementary**). An example, of such a self-complementary sequence is

$$\text{start-G-C-T-A-T-A-G-C-end.} \tag{4}$$

[2]The starting and ending ends are termed the "5-prime" and "3-prime" ends, respectively, of the DNA strand.

and the double stranded DNA it forms is

start-G-C-T-A-T-A-G-C-end
| | | | | | | | |
end-C-G-A-T-A-T-C-G-start .

Also, when a real heating and cooling experiment takes place it is usually in a solution of water in which the strands of DNA act as flexible chains. However, we will approximate the system as consisting of point-particles in an ideal gas. Physicists routinely approximate non-particle like objects as point particles (this is the essence of the ideal gas approximation), and ideal gases are good approximations for water based solutions consisting of very few particles. So, neither of these approximations are flagrant ones and we can still capture much of the the physical properties of DNA dimerization with them.

In summary, we will study the reaction Eq.(1) under three simplifications

- All single-strands are identical and self-complementary
- Single-strands will be approximated as point particles
- Collection of single and double strands will be approximated as an ideal gas

With these assumptions, the system of interest (single-stranded DNA combining in water to form double-stranded DNA) can be reduced to a system of identical point particles which can combine to form particle pairs. From this fact we can begin constructing a model from which we can derive a partition function.

3 Modeling the Dissociation and Association of DNA strands

Our system is at a temperature T and is contained within a volume V (See Fig. 1b). Let us say that each single-stranded DNA has a mass m_0. Thus the double-stranded DNA has a mass $2m_0$. Now, in typical reactions when two subunits combine to form a larger unit, energy is released. This released energy is termed the **binding energy** of the larger unit, and it is typically expressed as the larger unit having an energy less than the two units alone. Let us say the binding energy of the two single-strands of DNA is $-E_0$ for $E_0 > 0$. Because the single-stranded DNA are all identical to one another, this binding energy characterizes each combination of single-stranded DNA into double-stranded DNA.

Other than the way two single-strands of DNA can come together to form double-stranded DNA, we assume that individual DNA molecules (single-stranded or double-stranded), do not interact with one another. This assumption is in essence a claim to remain faithful to the ideal gas approximation. With these preliminaries we can begin to investigate the equilibrium statistical physics of this system. The way we will undertake this investigation is by computing the partition function for this system and then using the partition function to find physical quantities (like the average number of DNA strands in their double-stranded form).

3.1 Computing the Partition Function

To compute the partition of the ssDNA + ssDNA $\rightleftharpoons$ dsDNA system we will make extensive use of our aforementioned approximations. First, all partition functions consist of summations over some state space where the summation is weighted by a Boltzmann factor multiplied by a degeneracy factor. The energy that enters the Boltzmann factor of this system comes from two sources: the binding energy of two single-stranded DNA fragments existing as double-stranded DNA and the kinetic energy of the individual single-stranded DNA and double-stranded DNA particles.

Say we have $2N$ total single-stranded DNA fragments in our system, where this counting includes fragments that are binded together in the form of double-stranded DNA. Now let us say there are k double-stranded DNA in the system. That is, $2k$ single-stranded fragments have formed pairs and have become

double-stranded DNA. Thus, in this system of $2N$ total single-stranded DNA fragments in the system, only $2N - 2k$ are unpaired. Recalling that the double-stranded DNA has mass $2m_0$ and that the single stranded DNA has mass m_0, we can write the total energy of the system with k double-stranded DNA fragments as

$$E_{\{\mathbf{p}_i\},\{\mathbf{p}'_j\},k} = \sum_{i=1}^{k} \frac{\mathbf{p}_i^2}{2(2m_0)} + \sum_{j=1}^{2N-2k} \frac{\mathbf{p}_j'^2}{2m_0} - E_0 k, \tag{5}$$

where $\mathbf{p}_i$ is the momentum of double-stranded fragment i, and $\mathbf{p}'_j$ is the momentum of single-stranded fragment j. To compute the partition function corresponding to the energy given by Eq.(5), we have to determine the space of microstates we will be summing over. Since there are $2N$ total single-stranded DNA fragments (existing alone or as part of some double-stranded DNA), and there are two fragments in each double-stranded DNA, then the number k of double-stranded DNA can run from 0 to N. Therefore, somewhat schematically, the partition function for this system can be written as

$$Z = \sum_{k=0}^{N} \sum_{\{\text{all } \mathbf{p} \text{ and } \mathbf{q} \text{ of ssDNAs}\}} \sum_{\{\text{all } \mathbf{p} \text{ and } \mathbf{q} \text{ of dsDNAs}\}} e^{-\beta E_{\{\mathbf{p}_i\},\{\mathbf{p}'_j\},k}} \tag{6}$$

In Eq.(6), we are summing over all possible numbers of double-stranded DNA and also summing over the various momentum (denoted $\mathbf{p}$) and position (denoted $\mathbf{q}$) states of the single-stranded and double-stranded DNA. To further reduce this result, we note that the form of the energy in Eq.(5) allows us to factor the Boltzmann factor in Eq.(6) into three parts. Namely, the partition function becomes

$$Z = \sum_{k=0}^{N} e^{-\beta E_0 k} \sum_{\{\text{all } \mathbf{p} \text{ and } \mathbf{q} \text{ of ssDNAs}\}} \exp\left(-\beta \sum_{j=1}^{2N-2k} \frac{\mathbf{p}_j'^2}{2m_0}\right) \sum_{\{\text{all } \mathbf{p} \text{ and } \mathbf{q} \text{ of dsDNAs}\}} \exp\left(-\beta \sum_{i=1}^{k} \frac{\mathbf{p}_i^2}{4m_0}\right). \tag{7}$$

From here, we can compute the summations in sequence. We note that we have previously computed summations over position and momentum states of the kind shown in Eq.(7). One thing we noted back then was that such summations had to be converted into integrals because position and momentum are continuous variables. In particular, for the ideal gas with N particles of mass m in a volume V and at a temperature T, we found

$$\begin{aligned} Z_{\text{ideal gas}} &= \sum_{\{\text{all } \mathbf{p} \text{ and } \mathbf{q} \text{ of ssDNAs}\}} \exp\left(-\beta \sum_{k=1}^{N} \frac{\mathbf{p}_k^2}{2m}\right) \\ &= \frac{1}{N! h^{3N}} \int_V d^3\mathbf{q}_1 \int_{\text{all } \mathbf{p}} d^3\mathbf{p}_1 \cdots \int_V d^3\mathbf{q}_N \int_{\text{all } \mathbf{p}} d^3\mathbf{p}_N \, \exp\left(-\beta \sum_{k=1}^{N} \frac{\mathbf{p}_k^2}{2m}\right) \\ &= \frac{1}{N!} \left(\frac{V}{\lambda^3}\right)^N; \quad \lambda = \frac{h}{\sqrt{2\pi m k_B T}}. \end{aligned} \tag{8}$$

We can use this result to compute the right-most summations over states in Eq.(7). Namely, for k identical double-stranded DNA particles of mass $2m_0$ in a volume V, we have

$$\sum_{\{\text{all } \mathbf{p} \text{ and } \mathbf{q} \text{ of dsDNAs}\}} \exp\left(-\beta \sum_{i=1}^{k} \frac{\mathbf{p}_i^2}{4m_0}\right) = \frac{1}{k!} \left(\frac{V}{(\lambda_0/\sqrt{2})^3}\right)^k, \tag{9}$$

and for $2N - 2k$ identical single-stranded DNA particles of mass m_0 in a volume V, we have

$$\sum_{\{\text{all } \mathbf{p} \text{ and } \mathbf{q} \text{ of ssDNAs}\}} \exp\left(-\beta \sum_{i=1}^{2N-2k} \frac{\mathbf{p}_j'^2}{2m_0}\right) = \frac{1}{(2N-2j)!}\left(\frac{V}{\lambda_0^3}\right)^{2N-2k}, \tag{10}$$

where we defined

$$\lambda_0 = \frac{h}{\sqrt{2\pi m_0 k_B T}}. \tag{11}$$

Inserting these results into the partition function Eq.(7), we obtain

$$\begin{aligned} Z &= \sum_{k=0}^{N} e^{\beta E_0 k} \frac{1}{k!}\left(\frac{V}{(\lambda_0/\sqrt{2})^3}\right)^k \frac{1}{(2N-2k)!}\left(\frac{V}{\lambda_0^3}\right)^{2N-2k} \\ &= \left(\frac{V}{\lambda_0^3}\right)^{2N} \sum_{k=0}^{N} \frac{1}{k!}\frac{1}{(2N-2k)!}\left(\frac{2\sqrt{2}\lambda_0^3}{V} e^{\beta E_0}\right)^k. \end{aligned} \tag{12}$$

At this point, we have reduced our partition function from a schematic summation over states to a mathematically precise summation over k, the number of double-stranded DNA in the system. From here, it is possible to further reduce this discrete summation to an integral. To do so, we first note that the factors of inverse factorials can be written in terms of a binomial coefficient:

$$\begin{aligned} \frac{1}{k!}\frac{1}{(2N-2k)!} &= \frac{(2k)!}{2^k k!}\frac{(2N)!}{(2k)!(2N-2k)!}\frac{2^k}{(2N)!} \\ &= (2k-1)!!\binom{2N}{2k}\frac{2^k}{(2N)!}, \end{aligned} \tag{13}$$

where we used the definition of the double factorial

$$(2k-1)!! \equiv \frac{(2k)!}{2^k k!}, \tag{14}$$

in the final equality of Eq.(13). Conceptually, the double factorial $(2k-1)!!$ represents the number of ways to create unique collections of pairs of $2k$ different people. With Eq.(13), the partition function for the system can be written as

$$Z = \frac{1}{(2N)!}\left(\frac{V}{\lambda_0^3}\right)^{2N} \sum_{k=0}^{N} \binom{2N}{2k}(2k-1)!!\left(\frac{4\sqrt{2}\lambda_0^3}{V} e^{\beta E_0}\right)^k. \tag{15}$$

From here we will use the definition of the Gamma function with half-integer arguments to write the double factorial $(2k-1)!!$ in Eq.(14) as an integral. Similar to the way we can use integration by parts to show that $N! = \Gamma(N+1) = \int_0^\infty dx\, e^{-x} x^N$, we can show

$$(2k-1)!! = \frac{2^k}{\sqrt{\pi}}\Gamma(k+1/2) = \frac{2^k}{\sqrt{\pi}}\int_0^\infty dt\, e^{-t} t^{k-1/2}. \tag{16}$$

Inserting this identity into the summation, we obtain

$$\begin{aligned} Z &= \frac{1}{(2N)!\sqrt{\pi}}\left(\frac{V}{\lambda_0^3}\right)^{2N} \sum_{k=0}^{N} \int_0^\infty dt\, e^{-t} t^{k-1/2} 2^k \binom{2N}{2k}\left(\frac{4\sqrt{2}\lambda_0^3}{V} e^{\beta E_0}\right)^k \\ &= \frac{1}{(2N)!\sqrt{\pi}}\left(\frac{V}{\lambda_0^3}\right)^{2N} \int_0^\infty dt\, \frac{e^{-t}}{\sqrt{t}} \sum_{k=0}^{N} \binom{2N}{2k}\left(t\,\frac{8\sqrt{2}\lambda_0^3}{V} e^{\beta E_0}\right)^k \end{aligned} \tag{17}$$

To compute the final summation, we apply the binomial theorem in a slightly modified form. We note that, by the binomial theorem,

$$\sum_{\ell=0}^{2N} \binom{2N}{\ell} (-\sqrt{x})^\ell = (1-\sqrt{x})^{2N}, \qquad \sum_{\ell=0}^{2N} \binom{2N}{\ell} (\sqrt{x})^\ell = (1+\sqrt{x})^{2N}. \tag{18}$$

Adding the two summations in Eq.(18), we obtain

$$\begin{aligned}(1-\sqrt{x})^{2N} + (1+\sqrt{x})^{2N} &= \sum_{\ell=0}^{2N} \binom{2N}{\ell} (-\sqrt{x})^\ell + \sum_{\ell=0}^{2N} \binom{2N}{\ell} (\sqrt{x})^\ell \\ &= \sum_{\ell=0}^{2N} (\sqrt{x})^\ell \left[1 + (-1)^\ell\right] \\ &= 2\sum_{k=0}^{N} \binom{2N}{2k} x^k, \end{aligned} \tag{19}$$

which implies

$$\sum_{k=0}^{N} \binom{2N}{2k} x^k = \frac{1}{2}\left[(1-\sqrt{x})^{2N} + (1+\sqrt{x})^{2N}\right]. \tag{20}$$

Defining

$$\eta \equiv \left(\frac{8\sqrt{2}\lambda_0^3}{V} e^{\beta E_0}\right)^{1/2}, \tag{21}$$

we find that the final form of the partition function for our system of single-stranded and double-stranded DNA is

$$\boxed{Z = \frac{1}{(2N)!\sqrt{\pi}} \left(\frac{V}{\lambda_0^3}\right)^{2N} \int_0^\infty dt\, \frac{e^{-t}}{\sqrt{t}} \left[(1+\eta\sqrt{t})^{2N} + (1-\eta\sqrt{t})^{2N}\right]} \tag{22}$$

Now this seemingly complicated final result is actually amenable to simplification, and with these simplifications we can reduce the integral to a form from which we can extract information like how the average number of double-stranded DNA varies with the temperature of the system. Given that k in Eq.(12) represents the number of double-stranded DNA in the system, the average of k is given by

$$\begin{aligned}\langle k \rangle &= \frac{1}{Z}\left(\frac{V}{\lambda_0^3}\right)^{2N} \sum_{k=0}^{N} k\, \frac{1}{k!}\frac{1}{(2N-2k)!}\left(\frac{2\sqrt{2}\lambda_0^3}{V} e^{\beta E_0}\right)^k \\ &= \frac{\partial}{\partial(\beta E_0)} \ln Z. \end{aligned} \tag{23}$$

With the average number of double-stranded DNA defined in terms of a partial derivative of the natural logarithm of the partition function, evaluating the partition function should allow us to determine the average number of double stranded DNA as a function of temperature.

4 Computing $\langle k \rangle$ for the system of DNA

In the previous section, we found the partition function for our system of identical single-stranded DNA which can combine to form double stranded DNA. In this section, we will use Laplace's method to derive a temperature-dependent formula for $\langle k \rangle$ the average number of double-stranded DNA in the system.

We begin with our final partition function Eq.(22). By using integration by parts, we can write Eq.(22) in a form more amenable to Laplace's method. Taking $u = e^{-t}$ and

$$v = \frac{2}{\eta(1+2N)} \left[\left(1+\eta\sqrt{t}\right)^{2N+1} - \left(1-\eta\sqrt{t}\right)^{2N+1} \right], \tag{24}$$

we find Eq.(22) is equivalent to

$$Z = \frac{1}{\eta(2N+1)!\sqrt{\pi}} \left(\frac{V}{\lambda_0^3}\right)^{2N} \int_0^\infty dt\, e^{-t} \left[\left(1+\eta\sqrt{t}\right)^{2N+1} - \left(1-\eta\sqrt{t}\right)^{2N+1} \right]. \tag{25}$$

Now, we would like to compute $\langle k \rangle$, the mean number of double-stranded DNA in the system, as a function of temperature. To do so, we begin with the definition

$$\langle k \rangle = \frac{\partial}{\partial(\beta E_0)} \ln Z, \tag{26}$$

and we use Laplace's method to estimate the free energy F defined according to

$$Z = \int_0^\infty dt\, \exp(-\beta F(t)) \simeq \exp(-\beta F(\bar{t})), \tag{27}$$

modulo a thermodynamically irrelevant prefactor and where $\bar{t}$ is the value of t where $F(t)$ has a local minimum. From Eq.(27), $F(t)$ is defined as

$$\beta F(t) = t - \ln\left[\left(1+\eta\sqrt{t}\right)^{2N+1} - \left(1-\eta\sqrt{t}\right)^{2N+1} \right] - \ln\eta + \beta F_0(N,V,\beta), \tag{28}$$

where $F_0(N,V,\beta)$ contain all the terms which do not depend on βE_0. For $N \gg 1$, and $x \geq 0$, we have $(1-x)^N \ll (1+x)^{N+1} \simeq (1+x)^N$, and thus this $F(t)$ can be further approximated as

$$\beta F(t) \simeq t - 2N \ln\left(1+\eta\sqrt{t}\right) - \ln\eta + \beta F_0(N,V,\beta). \tag{29}$$

Computing the value of t at which this quantity has a critical point, we find the condition

$$1 = \frac{N\eta}{\sqrt{\bar{t}} + \eta\bar{t}}, \tag{30}$$

which is stable for

$$0 < \frac{N\eta}{\left(\sqrt{\bar{t}} + \eta\bar{t}\right)^2} \left(\frac{1}{2\sqrt{\bar{t}}} + \eta\right), \tag{31}$$

that is, for all positive values of $\bar{t}$. Since the domain of t is $(0,\infty)$, Eq.(31) is automatically satisfied for any valid solution of Eq.(30). Before we solve the condition Eq.(30), we use Eq.(26) and Eq.(27) to determine $\langle k \rangle$ in terms of $\bar{t}$. Doing so, we find

$$\begin{aligned}
\langle k \rangle &\simeq -\frac{\partial}{\partial(\beta E_0)} \beta F(\bar{t}) \\
&= \left[-1 + \frac{N\eta}{\sqrt{\bar{t}} + \eta\bar{t}} \right] \frac{\partial \bar{t}}{\partial(\beta E_0)} + \frac{2N\eta\sqrt{\bar{t}}}{1+\eta\sqrt{\bar{t}}} \cdot \frac{1}{2} + \frac{1}{2} \\
&= \bar{t} \cdot \frac{N\eta}{\sqrt{\bar{t}} + \eta\bar{t}} + \frac{1}{2}
\end{aligned}$$

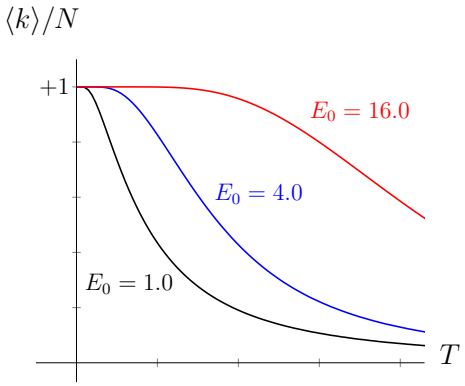

Figure 3: Plot of $\langle k \rangle / N$ versus T for various choices of E_0 (in units of $k_B T$). We see that as we increase the binding energy of double-stranded DNA, the fraction of DNA strands in their double-stranded form remains close to 1 for even high temperatures. The melting temperature is taken to be the temperature at which half of the DNA strands are in their double-stranded form. Thus we see that as the binding energy increases, so too does the melting temperature.

$$= \bar{t} + \frac{1}{2}, \tag{32}$$

where we used $\eta = e^{\beta E_0/2} \cdots$ in the second line and Eq.(30) in the final equality. We can now return to Eq.(30). Defining $s \equiv \sqrt{\bar{t}}$, we find the quadratic equation

$$s^2 + \frac{1}{\eta} s - N = 0, \tag{33}$$

which has the solutions

$$s_\pm = \frac{-1/\eta \pm \sqrt{1/\eta^2 + 4N}}{2} = \frac{1}{2\eta} \left[-1 \pm \sqrt{1 + 4N\eta^2} \right]. \tag{34}$$

We will chose the positive root in order to ensure that $\sqrt{\bar{t}}$ is positive. Doing so, we find

$$\begin{aligned} t_+ &= \frac{1}{4\eta^2} \left[1 - \sqrt{1 + 4N\eta^2} \right]^2 \\ &= \frac{1}{4\eta^2} \left[2 + 4N\eta^2 - 2\sqrt{1 + 4N\eta^2} \right] \\ &= N + \frac{1}{2\eta^2} \left[1 - \sqrt{1 + 4N\eta^2} \right]. \end{aligned} \tag{35}$$

With Eq.(21), and Eq.(32), we find that the average number of pairs in the system as a function of temperature is

$$\langle k \rangle \simeq N + \frac{V e^{-\beta E_0}}{16\sqrt{2}\,\lambda_0^3} \left[1 - \sqrt{1 + \frac{32\sqrt{2}\,\lambda_0^3 N}{V} e^{\beta E_0}} \right] + \mathcal{O}(1), \tag{36}$$

where we subsumed the factor of $1/2$ into $\mathcal{O}(1)$. Or, introducing molarity/number density through the definition

$$n \equiv N/V, \qquad \text{[Molarity or Number density]} \tag{37}$$

the fraction of the single-stranded DNA which exist in their double-stranded form is given by

$$\boxed{\langle k\rangle/N \simeq 1 - \frac{e^{-\beta E_0}}{16\sqrt{2}\,\lambda_0^3 n}\left[\sqrt{1+32\sqrt{2}\,\lambda_0^3 n\, e^{\beta E_0}}-1\right] + \mathcal{O}(N^{-1}),} \tag{38}$$

which provides an answer to the framing question of these notes.

4.1 Melting Temperature of DNA

To obtain a qualitative sense of the temperature dependence of Eq.(38) we plot it in Fig. 3 for various choices of E_0. From the plot we observe two things: For a single curve, we see that as we increase temperature, the average fraction of double-stranded DNA decreases. This makes sense since as we raise the temperature we are essentially "melting" the system, making the higher entropy state of all dissociated DNA strands more thermodynamically favorable. Second, we see from comparing all three curves that as the binding energy E_0 increases, the fraction of DNA strands in their double-stranded form increases for fixed temperature. These two results are qualitative predictions that can be made more precise through quantitative arguments.

For example, from Eq.(38) we can compute $T_{1/2}$, the temperature at which half of the strands of DNA are in their double-stranded form (i.e., $\langle k\rangle/N = 1/2$). Setting the right hand side of Eq.(38) to $1/2$, we can derive

$$k_B T_{1/2} = \frac{2E_0}{3}\left[W_0\left(\frac{4\pi m_0 E_0}{3h^2\left(4n\sqrt{2}\right)^{2/3}}\right)\right]^{-1}. \qquad \text{[Melting temperature of DNA]} \tag{39}$$

where h is Planck's constant, m_0 is the mass of a single-strand of DNA, and $W_0(x)$ is the principal branch of the **Lambert-W function**
defined by the condition

$$W_0(xe^x) = x. \qquad \text{[Lambert-}W\text{ function]}. \tag{40}$$

As an example calculation, we could see what this result predicts for the dissociation of DNA strands. For a strand of DNA with 20 base-pairs[3], we have

$$E_0 = (20 \pm 3)\text{ kcal/mol} = (1.4 \pm .21)\times 10^{-22}\text{ kJ} \qquad \text{and} \qquad m_0 = 9.7\times 10^{-24}\text{ kg}, \tag{41}$$

as the binding energy and the mass of a ssDNA, respectively. The former value can be found from the OligoCalc tool **(http://biotools.nubic.northwestern.edu/OligoCalc.html)** and the latter value can be found by noting that a base pair (which is *two* nucleotides) has a mass of about 650 Daltons. We will assume these strands exist in a 50 micromolar (i.e., 10^{-3} mol/m^3) solution:

$$n = 50.0\times 10^{-3}\text{ mol/m}^3 = 3.01\times 10^{22}\text{ m}^{-3} \tag{42}$$

We then find that $T_{1/2}$ has the value

$$T_{1/2} = (352 \pm 50)\text{ K} \qquad \text{(For 20 base-pair strand of DNA)}, \tag{43}$$

Which is reassuringly of the correct order or magnitude of the correct result ≈ 315 K found using *OligoCalc*.

5 Final Remarks

With Eq.(43) we have come to a partial end of the process of theoretical modeling. We abstractly considered the properties of a system of interest; built up a physical model to encompass said system; and then used the physical model to make a prediction[4]. Although the model we considered might have seemed involved, it is actually relatively simple relative to the models we could have created. More complicated versions

[3] In what follows, the exact sequence of base pairs does not change the rough estimates of E_0 or m_0.
[4] The final step would involve testing our prediction.

would have included the differing geometry between the single-stranded DNA and double-stranded DNA, the vibrational energy of the double-stranded DNA, or single-strands which were not identical. However, even with all of our simplifications, we managed to get the ball-park estimate of melting temperature of double-stranded DNA. Such is the power of the general models of statistical physics.

Resource: Introduction to Statistical Physics

Mobolaji Williams