Calculable Sequence Dependent Quantities for DNA

We saw last time that the statistical mechanical model pointed out the importance of the partition function

$$Z = Z(T) = \sum_i e^{-G(i)/RT},$$

where i runs over the possible states of our system – we are taking into account duplex denaturation at a *nt* position, and twisting of the two DNA strands about each other when there is a run of open Bp's – and $G(i)$ is the free energy from superhelical stress in the i configuration; R is Boltzmann's constant and T is the absolute temperature. The expectation of the random variable

$$n_x = \begin{cases} 1 & \text{if position } x \text{ open} \\ 0 & \text{if } x \text{ is closed} \end{cases}$$

There are at least two ways to measure the likelihood for denaturation of a given position

x in a DNA chain (due to suprehelical stresses). We could calculate the probability that x is open. This would be a function of x, expressed as the expectation of n_x, i.e.,

$$p(x) = E(n_x) = \frac{\sum_i n_x(i) e^{-G(i)/RT}}{Z(T)}.$$

A more sensitive measure is the *incremental energy* for denaturation at the position x, call it $G(x)$. It is the difference of

$$\begin{aligned} \bar{G} &= E(G(i)) \\ &= \textstyle\sum_i G(i) e^{-G(i)/RT} / Z(T), \quad \text{and} \\ \bar{G}(x) &= \frac{\sum_i n_x(i) G(i) e^{-G(i)/RT}}{E(n_x)}. \end{aligned}$$

So,

$$G(x) = \bar{G}(x) - \bar{G}.$$

A SIDD *profile* is the graph of $G(x)$ against x.