The DNA Model

We want to give a brief description of the mechanical model used for duplex destabilization as it depends on primary DNA sequence. The overall approach is via statistical mechanics, where the first main task is to determine the Boltzmann distribution over the set of all possible states which the system could be in. For us the system means the possible states of a DNA molecule with respect to its destabilization (melting). So, we have a DNA molecule of length N, and each of its positions can be either open or closed (Watson-Crick H-bonds intact or broken). This gives us 2^N states. There will be further continuous parameters when we take into account the twisting of the two strands of an open stretch of the helix around one another. Boltzmann describes the probability of being in state i of the system as

$$p_i = \frac{e^{-G_i/RT}}{Z},$$

where

 G_i = the free energy of state i,

$$R = Boltzmann's constant,$$

what Prof Meiners called k_B ,

$$T = absolute temperature, i.e., in degrees Kelvin,$$

Z = the **partition function**, which guarantees we have a prob. distribution,

so that

$$Z = \sum_{i} e^{-G_i/RT}.$$

When we add continuous parameters, we will have to integrate out over those as well in the partition function. How do we describe G_i ?

There is a specific linking number α which the helix will have over the N bp length. Introduce a variable of the states n_i which is 1 if the state is open and 0 if it is closed, at the *i*-th position in the chain. Partition the forces acting on this segment as coming from three sources.

1. The separation energy

Suppose there are r open stretches in the duplex. Then the separation component of the free energy is given by

$$G_s = ar + \sum_{j=1}^N b_j n_j,$$

where the b_i are the energies needed to denature the j-th base pair. This depends on the identity of the base pair and its (nearest ?) neighbors, and is the reason this analysis might apply to gene regulation. These have been well measured experimentally. The summand ar means that it costs an amount a of energy to open a run of denatured bp's: this is because a great deal of the stabilization of the DNA helix comes from the near parallel arrangement of the near planar configurations of the rings of the bases on either strand. (This could be partially due to the alignment of magnetic moments of such rings or from packing effects coming from the ambient water, DB

doesn't know the details here.) a also has been measured experimentally in a variety of conditions.

2. The twist energy

This is where the continuous parameters start entering. The total twist \mathcal{T} within the separated regions is given by

$$\mathcal{T} = \sum_{j=1}^{N} \frac{n_j \tau_j}{2\pi},$$

where τ_j is the *local helicity* (radians per unit base pair length) at position j. The energy contribution is just a "circular spring" model, with Hooke's Law free energy:

$$G_t = C \times \frac{1}{2} \sum_j n_j \tau_j^2.$$

The constant of proportionality C is the *effective torsional stiffness*, and has been measured experimentally.

3. Residual linking

This is where we take into account the energy which has been stored in the global supercoiling of the helix. Define

$$n=\sum n_j,$$

the total number of open bp's. Opening these bp's should reduce the stress on the DNA by an amount equal to n/A, where A = 10.4 is the helicity of unstressed DNA in the standard B-form. There will still be torsional stresses unless you have cancelled out exactly the global torsional stress, i.e., unless you have hit exactly $n = -\alpha/A$. The free strands will also respond by twisting when denatured. Set α_r , the residual linking difference, equal to the component of α which is not met by the two previous deformations. We get

$$\alpha = -\frac{n}{A} + \mathcal{T} + \alpha_r = constant.$$

It turns out that α_r can be measured experimentally, so that we don't have to describe it more fully here. Finally, the residual contribution to the free energy is given by

$$G_r = K \frac{\alpha_r^2}{2},$$

another Hooke's law, where the constant of proportionality has been measured.

Returning to the main argument, we get

$$G(\mathsf{state}) = Gs + G_t + G_r,$$

where now specifying a state means specifying the n_j 's and the τ_j 's at each j where nj = 1. (Note that r can be calculated from the n_j 's.) It is noteworthy that the partition function, and many useful expressions like it can be computed by a dynamic program, using the socalled *energy transfer method* due to the physicists who studied the Ising model right after WW II, but that it requires arbitrary precision arithmetic in its computation because of instabilities coming from cancellations of large numbers.