CONFORMATIONAL STATISTICS

The packaging of 3 meters of DNA in a eukaryotic cell obviously involves the folding of the molecule back on itself. How the polynucleotide bends is thus important to placement in the cell. Moreover, many regulatory functions involve bringing together protein binding sites that are well separated from each other-a process that is thought to require DNA bending. Because of what we have already learned about the sequence specific factors that influence DNA conformation of DNA, we would also like to know something about how base sequence may affect the folding of DNA upon itself.

Before we discuss the effect of base sequence on DNA deformability, we have to have an understanding of how we would begin to measure flexibility and how we would characterize the movements of a DNA chain.

Native DNAs from organisms have very high molecular weights-the MW of the *E coli* chromosome is ~3 $X10^9$. As we know, DNA is made of segments, in the simplest case these segments can be thought of as the base paired nucleotide. Since these segments have finite size, it is clear that many conformations of the chain in solution will be excluded, because of the prohibition of physical overlap of two segments in the same volume of space. Outside of these rather obvious considerations, nucleic acid polymer flexibility can be modeled as existing between two limiting case of chain deformability-the completely rigid rod and the completely flexible random coil. In fact DNA can be considered as existing in either of these two regimes, depending on the number of segments that one includes in the analysis of DNA deformation. As we shall see, short stretches of DNA can behave as rigid rods while enormous DNA lengths can assume the characteristics of completely flexible chains. In actual fact, DNA of intermediate length (number of segments) is best modeled between these two limiting cases, as the so-called worm-like coil, which is of intermediate flexibility.

Average dimensions.

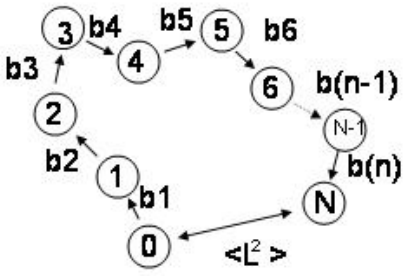
A solution of flexible polymers contains molecules that may have a different conformation at any given instant. Further, due to the random bombardment of the polymer molecules by solvent (Brownian motion), the conformation of a particular polymer molecule will change with time. This suggests that we cannot specify a single dimensional parameter to characterize flexible polymers in solution, but must instead deal with average dimensions. One of the most commonly used average dimensions is that of the mean square end-to-end distance $\langle \mathbf{L}^2 \rangle$. We have a polymer of N repeating units, each connected by bond vectors of length **b**. A vector **L** is drawn between the beginning and end of the chain, which equals the sum of the individual bond vectors $\mathbf{b}_i$. Therefore

$$\mathbf{L} = \sum_{i=1}^{N} \mathbf{b}_i$$

Thus,

$$\langle \mathbf{L}^2 \rangle = \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbf{b}_i \cdot \mathbf{b}_j$$

This all means that the root mean square distance (L) between point i and point j is the sum of the vectors that lie between them.



Now that we have defined our measurements somewhat, we shall proceed to apply these considerations to particular models of polymer chains.

Rigid Rod.

This is relatively easy to understand, since the rod is inflexible, no averaging over internal conformations is required. Thus, the distance between any two points is equal to

$$L_{ij}= b*i-j$$

where all bonds have constant length b and the distance between two points at the ends of the chain is

$$L=b*N$$

where N is the number of segments and b is the length of the bond that connects the segments.

Flexible Linear Chains

At the opposite extreme from the rigid rod, let us imagine a completely flexible chain containing N segments each separated by bonds of length b. We can imagine that the bonds are connected by universal joints that allow completely free rotation. The conformation of the polymer is therefore that of a random walk in which successive steps are completely uncorrelated in direction. The quantity which must be ascertained for the polymer is its distribution function; that is the probability W(L,N)dL that after N steps the end of the chain will be at distance L and L+dL from the origin.
Chains that conform to this distribution function are often called gaussian chains because of the gaussian character of the population of chain conformations
Thus the mean end to end distance of a completely flexible chain is proportional to the first power of N

$$<L^2>=Nb^2$$

as opposed to the N^2 for the rigid rod.

The concept of a freely jointed chain is obviously an inaccurate representation of a natural polymer. In the case of a nucleic acid, the joints connecting the segments are not universal joints because of steric considerations and there are fixed bond angles (e.g. geometry imposed by tetrahedral carbons). Thus, both N and b for a real polymer are modified to effective N and b. The values for N_e are less than actual N and b_e smaller than b. The distribution function for the effective chain is just that for the freely jointed chain with N replaced by N_e and b with b_e. The mean square end to end distance of this chain is thus,

$$<L^2> = b_e^2 N_e$$

Worm-like chains.

Even with the concept of statistically equivalent (N_e-type) chains, the gaussian distribution requirements are such that N_e must be substantially greater than 1, although $N_e << N$. For sufficiently short chains or relatively stiff chains, this requirement may not be satisfied. In fact native DNA is stiff enough that gaussian statistics are inapplicable for many purposes. This situation thus demands an even more elaborate treatment. Since stiff chains can be envisioned to bend only gradually and smoothly, somewhat like a worm, hence the term wormlike chain

In order to characterize the stiffness of a polymer chain, we imagine that the polymer is laying along the y-axis of a 3-D coordinate reference frame. The chain consists of N segments, connected by bonds with length b. We imagine that the first step is taken along the positive z axis. We then ask, what is the average projection <z> of the chain at N steps along chain at the z-axis? It is obvious that if the rod were completely stiff, <z> would have its maximum value of N*b; for a freely jointed chain, <z>=b since steps beyond the

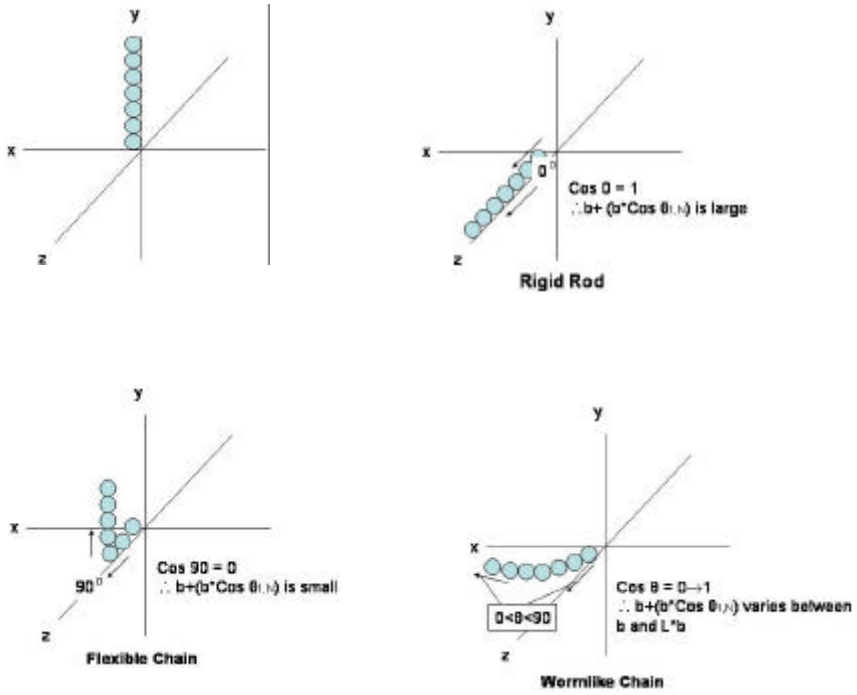
first the first would have equal probability to be taken in the (+) or (-) z direction. Thus, polymers of intermediate stiffness will have intermediate values of <z>. The average projection must consider the projections of all succeeding bonds on the first, thus,

$$\langle z \rangle = b + b\langle \cos \theta_{1,2} \rangle + b \langle \cos \theta_{1,3} \rangle + \ldots b\langle \cos \theta_{1,N} \rangle$$

for the chain with free rotation

$$\langle z \rangle = b^* \, 1-\cos^N \theta / 1- \cos \theta$$

if the cos ? is close to zero (i.e. angle of 90^o) <z> = b/(1-cos ?), however if ? is 0^o, true for very stiff chains, cos ?=1 then <z> becomes very large.



There is, therefore a continuum, in which both segment length b and the bond angle ? are taken to the infinitesimal limit in such a way that the quantity

$$a = b/1- \cos \theta$$

remains finite. The quantity a is called the persistence length of the worm like chain and represents the average extension along the z-axis of an indefinite length polymer. In other words the persistence length is that number of segments N where the chain behaves as a rigid rod. In fact for a wormlike chain, the effective segment length b_e is twice its persistence length. Thus for a short wormlike chain, the limiting behavior is a rigid rod.

Thus for considering DNA, the persistence length in units of distance is a measure of its stiffness. The longer the persistence length, the more stiff the chain.

With all of this stuff in mind let us take a look at how to apply it to real DNA

In the Shore et al. (1981), they wished to determine experimentally the degree of DNA flexibility and whether the model of a worm like coil held true for DNA of all lengths or were other considerations needed. To do this, they measure the rate of circularizing several linear pieces of DNA of different lengths and compare the results to that predicted from the model. The ring closure probability can be understood as the effective concentration of one end in the vicinity of the other. That is if they are so far away the concentration of one in near the other is low and closure probability decreases, as they are brought together, the effective concentration increases, as does the closure rate. In all cases they compare the rate of closure of the linear DNA into a circle, and normalize this to the probability of joining two ends on separate pieces of DNA. The relative increase is the benefit from being on the same piece of DNA.

In fact the effective concentration of ends (which is proportional to the average mean square distance $<L^2>$ and thus to the number of base pairs) is not the only thing that limits joining of the two ends. If the DNA is particularly short, the relative orientations of the two ends are correlated. For joining to occur, the number of bases has to be close to 10, i.e., the number of bases/10.5 should be an integer.
The data shown in **FIGURE 43** is in fact in quite good agreement with the worm like coil representation of DNA >500 base pairs and end to end correlation do not have an effect until below this value. The fact they do indicates that the twisting motions of DNA are energetically restricted. This leaves open the possibility of using this technique to measure the energetics of twisting.

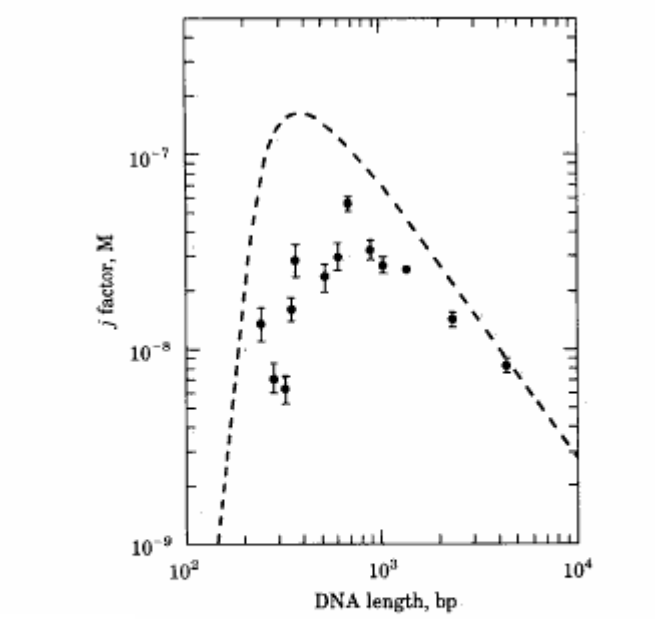


Figure 43

j factor versus DNA length, determined at 20°C. Fragment lengths were 242, 288, 318, 345, 366, 504, 611, 670, 880, 1015, 1361, 2302, and 4361 bp; their origin is described in Table 1. A 126-bp fragment is not shown on this graph because its *j* factor is at least 100-fold lower than that of the 1361-bp reference fragment. The dashed curve is the angle-independent ring closure probability calculated from equation 62 of Yamakawa and Stockmayer (6) for a persistence length of 500 Å.