6 Experimental Approaches to DNA structure and dynamics

6.1 Introduction

In the case of DNA of a few hundreds of base-pairs, we have seen in the previous lectures the basics of the Cosserat theory of rods which constitutes the abstract representation of the DNA. In principle, this abstract representation can be used for any type of physical situation where an elastic rod is considered. To apply this model in a specific context (as DNA for instance), we need to extract from experimental data on DNA both:

- the reference state, or in other words the $\hat{\boldsymbol{u}}_i(\boldsymbol{s})$ and $\hat{\boldsymbol{v}}_i(\boldsymbol{s})$
- the parameters K and A in the constitutive relations, which relate how the strains u and v are affected by the forces n and moments m via

$$\boldsymbol{m}_i = \boldsymbol{K}_i(\boldsymbol{u}_i - \hat{\boldsymbol{u}}_i), \qquad \boldsymbol{n}_i = \boldsymbol{A}_i(\boldsymbol{v}_i - \hat{\boldsymbol{v}}_i).$$

Among the various experimental methods available to study DNA structure, we're going to explore in more details two of them today, one which gives high resolution data (positions of atoms in a short molecule of a few base pairs), X-Ray crystallography, and one which is used routinely in biochemistry labs, gel electrophoresis, which can be used to extract low-resolution data. These two methods, among others, were involved in the discovery of DNA bending, i.e. the fact that the *reference state* of B-DNA is not straight in general.
This discovery is about 20 years old now, and has shed a new light on the role of DNA structure on its function. Several examples of this role have been reported in the literature, and all functions of DNA (transcription, replication, recombination and packaging) are concerned by DNA bending.

6.2 X-Ray Diffraction and Crystallography

Early Developments.
The most important method, at least from a historical point of view, has been the analysis of DNA structure by X-ray diffraction. This is the method which gave rise to the discovery of the DNA double helix by Watson and Crick in 1953. They did not make experiments themselves though, but benefited from the results they saw in Rosalind Franklin's lab, where she had collected X-ray diffractograms from natural DNA fibers and she observed

that when her fibers were kept dry, the pattern looked different than when they were wet. She named the wet case the B form, and the dry one the A form. These fiber forms are the ancestors of what we today call the A and B form DNA structures. Later work on fibers showed that the DNA structure was extremely polymorphic, i.e. each fiber preparation could produce different diffraction patterns depending on its salt and water content. In the 1980s, it became clear from these fibers studies that the DNA double helix was more complex that the canonical model proposed by Watson and Crick for the B form, Arnott and coworkers for the A form, and Alexander Rich for the Z form DNA. Yet although X-ray pictures of a fiber sample can show well enough that the fiber forms discovered then (A, B, C, C', D, E) are distinct, they do not yield sufficient information to determine the detailed three-dimensional structure of those different forms.
During this period, chemists learned how to synthesize DNA chemically in large amounts, and how to purify it so that one could grow crystals of a particular given short DNA sequence. These crystals would not grow unless the preparation is pure, which means that these short DNA fragments, named *oligomers*, all have identical basepair and atomic compositions. The first structure to be solved by this method was the very short oligomer ATAT by Viswamitra in 1978. It proved to be very disappointing because the molecule did not form a double-helix. What would you try as the next sequence? If you remember that AT base pairs are easier to open than GC base pairs, you would do CGCG. That's exactly what several labs tried during the years 1979-1980, and there the surprise was that the double helices produced by these GC rich oligomers were not right but left handed... However, earlier low resolution experimental data (circular dichroism) already suggested that depending on the salt concentration, the alternating GC rich fragments may be either left of right handed, but only a few crystallographers had taken them seriously. Today, the obstacles to longer and purer oligomers have been surmounted, and these last 15 years have seen a tremendous development of X-Ray crystallographic studies of biomolecules. The method.
First the crystallographer order the oligomer to a company or another lab. For DNA protein complexes, s(he) also needs to prepare large amount of the desired protein in chemically pure form, usually by cloning a gene for the protein into a bacteria, and then using the bacterial expression machinery to produce the protein by feeding the bacteria with amino acids and ATP. The next step is to grow a crystal, and this is certainly a very chancy part.

See figure 4.

Crystal from AGCATGCT combined with the antibiotic nogalamycin.

Each crystal is about 1mm long.

In about two weeks, you may endup with a 1mm long crystal suitable to be exposed into an X-ray beam, and if the crystal is well ordered, an X-Ray photograph can be taken. Any large crystal is made from millions of identical DNA molecules, closed packed in space. The geometical locations of spots of the diffractogram all correspond to a given constructive interference of X-rays, and tell us what sort of array the molecules have formed. The first task is to determine which of the 65 possible symmetries the crystal has chosen. And after that, one can start by measuring the relative intensities of the spots depending on the different orientations fo the crystal in the beam (actually, several photographs are taken at different angles). Today, this procedure is automated, and high resolution can be obtained by high power X-ray beams. The final task is to translate the relative intensities into a model of the atomic structure. Each non-hydrogen atom (carbon, oxygen, nitrogen, phosphorus) can be located to an accuracy of about 0.1-0.2 **r**Ain 3D if this last job of mapping the electron-density map onto a 3D molecular model is done properly. But this last job is the hardest in the procedure, and requires heavy mathematics centered on the Fourier Transform theory.

See figure 5.

Even in the best case though, some parts of the molecular assembly are more flexible and floppy within the crystal. Actually, some biological activity may even be observed without disassembling the crystal (one replication step can be performed in the crystalline stage). Also, due to crystal packing forces, some *intermolecular contacts*, or contacts between different molecules in the crystal, tend to influence the structure of the individual molecules building the crystal. These packing forces are very sensitive to the environment in which the crystal is grown. These two problems explain why the best resolution for a biomacromolecule usually does not exceed 2 **r**A. Anyway, these 3D models are regarded as broadly representative of the structure in solution, *on average,* for if the structure in solution was too different than the one in the crystal, the molecule would never have crystallised! These 3D models form the whole underpinning for the science of molecular biology; and this is why we have explained how they are derived in more detail. Compilation of X-Ray crystallographic data, in terms of one dinucleotide model parameters (three rotations, three translations). In principal, these 6 parameters are the ones that define the $\hat{\boldsymbol{u}}_i(\boldsymbol{s})$ and $\hat{\boldsymbol{v}}_i(\boldsymbol{s})$.

Step	N	Twist(°)	Tilt(°)	Roll(°)	Shift(rA)	Slide(**r**A)	Rise(**r**A)
AA	**81**	**35.5(±3.6)**	**−0.8(±3.1)**	**0.1(±3.8)**	**0.0(±0.39)**	**−0.14(±0.30)**	**3.28(±0.17**
AG							
GA							
GG							
AC							
AT							
GC							
CA							
CG							
TA							

Parameters taken from Gorin, A. A., Zhurkin, V. B. & Olson, W. K. (1995)

The standard deviations about each average may serve as sequence-dependant flexibilities, or once averaged properly they can be assigned to the K and A parameters of our rod model.

6.3 Gel Electrophoresis

This is a method routinely used by biochemists to separate DNA fragments according to their shapes. First let us say some words about the gel itself. A gel can be considered as a 3D array of tiny, randomly oriented fibers. A typical agarose gel (resembling very much the jello dessert, with no color though) is usually around 0.2-4 agarose (a molecule similar to gelatin), the rest being water. An electrical field is applied to force the charged polyelectrolytes (DNA, and most proteins, are charged polymers) to move through the gel. Polyacrylamide is another gel very commonly used in DNA gel electrophoresis.

See figure 6.

The characteristic interfiber spacing, sometimes called the mean pore size, is very different in these two types of gels: in typical agarose gels, the pore sizes vary from 400 to 40 nm, whereas in polyacrylamide gels, this quantity varies from 1 to 8 nm
The characteristic length that measures the stiffness of the DNA is called the persistence length, P, normally taken to be about 50 nm. Therefore, P is smaller or similar to the mean pore size of agarose gels, but many times larger than the interfiber spacing in polyacrylamide gels. DNA molecules of defined size generally behave in a very predictable way when run on agarose

or acrylamide gels. Shorter molecules migrate more rapidly than larger molecules. Although the migration in such a gel is not yet completely elucidated, the best model available so far for migration is based on the ability of the DNA to reptate or snake through the gel matrix. Occasionally, a DNA band of known length does not run at its expected position on an acrylamide gel. This occasional behaviour has been given the name of *anomalous migration*, and has become the experimental evidence of DNA bending. For DNA fragments of the same size, a permanently (or intrinsically) bent fragment will take more time to snake through the gel than a non-bent DNA fragment (this does not occur in agarose gels because the pore size is larger). One of the first examples of such an anomalous migration was shown for a 414 base pair fragment of kinetoplast DNA from *Crithidia fasciculata* (Marini et al, 1983) . Studied by electron microscopy, this fragment clearly showed *intrinsic curvature.* In other words, its equilibrium shape can be considered as stably bent.

See figure 7.

6.3.1 The circulary permuted DNA

One year later, Wu and Crothers (1984) designed a very clever experiment to map the site of permanent bending. It is referred to as the *circularly permuted assay.* The idea is to clone a bent fragment from a kinetoplast DNA as a dimer of two 241 identical DNA fragments. By cutting the dimer at different sites with specific restriction enzymes, and chosing enzymes that cut only once in each 241-bp fragment, one would produce permutated fragments of the same length (241 bp). Although all molecules have the same length, contain the same sequence, and incorporate essentially the same bent, they do not migrate at the same speed in the gel. The interpretation of this result is that the bends are located at different location along the circularly permuted family of molecules, and this creates different end-to-end distances for the DNA fragment depending on the bent location. For relative mobilities involving molecules of the same length, this end-to-end distance is important in determining gel mobility. The molecule with the shortest end-to-end distance will migrate most slowly. Plotted as a function of the position of the restriction cutting site used to generate the family, the distance in migration allows to find the center of the bent.

See figure 8.

The DNA sequence found at the bent center carries 5 runs of 4 or 5

As ($\boldsymbol{A_{4-5}}$) which, in each case, are phased by approximately 10 bp. These blocs of A are now given the name of A-tract, and the key to produce major bends in DNA resides in the correct phasing of these A-tracts.

See figure 9.

Historically, this experiment gave rise to two bending models named: the *junction model* and the *wedge model.*The basic idea about the junction model is that bending occurs at the junction between B-DNA and A-tracts, whose structure is different from regular B-DNA. The wedge model on the contrary attributes the global bent to the summation of small bents in each AA step.

See figure 10.

The shortest circle based only on phased A-tracts (21 bp repeats with two A tracts each, one of 5 and one of 6 A) is 126 bps. First note that this is well below persistence length (150bp), so that a random DNA sequence of 126 bp would have almost no chance to circularize. Thus efficient circularization is a serious indication for stable DNA bending of 360 degrees. As we have 54 AA steps in the 126 bp circle, we could calculate in the wedge model frame how much an AA step introduces of local bending. DNA sequences that do not contain runs of As can also be bent. The bends observed in DNA lacking phased A-tracts are usually not as large as A-tract bends. These sequences have not been studied as intensively as A-tracts. DNA bending can also occur when DNA is wrapped around a protein core, like in the nucleosome; in this situation, 80 bp of DNA are strained into a circle. The phasing in protein binding can be shown to play a critical role, as in the case of the arabinose system which was described in the introduction. Following the same line of experiments, researchers designed databases of presumably bent sequences and studied their relative gel anomalous migrations by statistical fitting with their dinucleotide (or more) models. One historical example is given by the Bolshoy-Trifonov work (1987) where a list of 57 DNA repeated sequences (most of them containing phased A-tracts) were analysed and from which an equilibrium wedge angle table of angles was deduced.

Step	N	Twist(°)	Tilt(°)	Roll(°)
AA	57	35.6	3.2	-6.5
AG				
GA				
GG				
AC				
AT				
GC				
CA				
CG				
TA				

Parameters taken from Bolshoy and Trifonov, 1987. You may wonder why the AA step average roll devised from the statistical survey of crystallographic data is so small compared to the roll angle mentioned here. One reason is that X-Ray sequences only rarely are A-tracts, whereas the computation based on gel anomaly data was focusing on A-tracts. This is one reason to suspect that dinucleotide models are not appropriate once A-tracts are present, because a AA step has different characteristics in a regular sequence as compared to an A-tract (where the double helix form is now called B'). The discovery of DNA bending, as well as its role in biology, has revealed that DNA sequence not only serves as a support for the *genetic code* (triplet code), but implies the simultaneous existence, on the same molecule, of a different (more structural) code which influences the DNA functions.

6.4 Cyclization Probability

The theory of gel electrophoresis is not yet sufficiently well developed to enable calculation of absolute values for curvature from the mobility anomaly. However, there is another experimental procedure which has a more rigorous theoretical basis, namely the rate of cyclization of DNA (Shore and Baldwin 1983). The idea here is to study how slight differences in sequence (either length or basepair content) affect the probability of covalently closing the DNA circle versus the probability of covalently ligating two fragments forming a dimer. The DNA fragments with *overhangs* (single stranded ends complementary to each other) are mixed in a tube with an enzyme (T4 DNA ligase) which ligates nicks in DNA strands. The ligase-trapping of apposed DNA ends do not directly measure the probabilities of cyclization, but rather the rate constants $(\boldsymbol{k_c}, \boldsymbol{k_d})$ for cyclization and dimerization (the bonding of one molecule's $\mathbf{5'}$ end to another molecule's $\mathbf{3'}$ end to form a DNA dimer). These in turn can be related to the equilibrium constants $(\boldsymbol{K_c}, \boldsymbol{K_d})$ for cyclization and dimerization by

$$\boldsymbol{k_c / k_d = K_c / K_d}.$$

This quantity is defined to be the Jacobsen-Stockmayer factor $\boldsymbol{J}$:

$$\boldsymbol{J \equiv K_c / K_d}.$$

Experimentally, $\boldsymbol{J}$ is measured by evaluating (densitometric measurement) the intensity of the bands corresponding to the dimer and the circular species in the same gel. $\boldsymbol{J}$ depends on the concentration of one end of a DNA molecule near the other end, the angular orientation between the two ends, and the twist angle between the two ends, which is a function of the DNA length and helix screw. Standard statistical mechanical expressions for the equilibrium constants $\boldsymbol{K_c}$ and $\boldsymbol{K_d}$ give:

$$\boldsymbol{J = \exp\left[-\frac{\Delta G_c^0 - \Delta G_d^0}{RT}\right]},$$

where $\boldsymbol{\Delta G_c^0}$ and $\boldsymbol{\Delta G_d^0}$ are the standard molar free energy changes in the cyclization and dimerization reactions. Hence, the quantity

$$\boldsymbol{\Delta G_{exp} \equiv \Delta G_c^0 - \Delta G_d^0 = \Delta H ...... T \Delta S_c^0 - \left(\Delta H_d^0 - T\Delta S_d^0\right)}.$$

can be experimentally determined. We focus first on the enthalpy contribution. The term $\boldsymbol{\Delta H_d^0}$ is the enthalpy for the new chemical bonds at the $\mathbf{5'}$ -$\mathbf{3'}$ connection. The term $\boldsymbol{\Delta H_c^0}$ contains the same enthalpy of the $\mathbf{5'}$ - $\mathbf{3'}$ connection, and also the enthalpy change due to the rearranged molecular shape in cyclization. The enthalpy from the new bonds will cancel in the difference $\boldsymbol{\Delta H_{exp} \equiv \Delta H_c^0 - \Delta H_d^0}$ to leave only the enthalpy change due to the rearranged cyclized shape; it is exactly this enthalpy change $\boldsymbol{\Delta H_{exp}}$ which should be well-approximated by the strain energies in the rod model. Accordingly, we can compare the experimentally determined $\boldsymbol{J}$ factors with strain energies computed in our continuum rod model, for different ratios $\boldsymbol{K_3/K_1}$. A set of 11 molecules was chosen to cover the range of unstressed shapes from the C to the S typical reference states.

See figure 11.

According to phasing rules that we discussed earlier, from these molecules the A17 and T15 molecules are more or less C-shaped, whereas the A09, T09 and A11 molecules were approximately S-shaped in their unstressed shape (reference state). The figure represents the strain energies for each sequence. The table below the experimentally measured free energy differences for them.

See figure 12.

There are two important points to consider in making this comparison. First, the experimentally-determined $\boldsymbol{J}$ factors, since they involve *free* energies, will contain entropy contributions not computed in the continuum model. Second, experimental difficulties imply that *relative* $\boldsymbol{J}$ factors among a set of molecules are generally more reliable than the absolutey $\boldsymbol{J}$ factors.

6.5 The biology of DNA bending

As an example, transcription regulation is generally performed via subtle mechanisms, implying repressors, promoters, enhancers, three names for DNA fragments which are not coding for the protein but interact with transcription factors and the RNA polymerase (which reads DNA to produce a messenger RNA) to regulate the level of transcription of this protein.

See figure 13.

(A) The regulatory elements of the arabinose system operon include the $ara\boldsymbol{O_2}$ site at the left of the operon, the $ara\boldsymbol{O_1}$ site at about -120 bp, and an $ara\boldsymbol{I_1}$ -$ara\boldsymbol{I_2}$ site at around -30 bp from the transcription initiation site (RNA polymerase starts transcription at position +1). (B) A repression loop formed by an AraC dimer binding to both $ara\boldsymbol{O_2}$ and $ara\boldsymbol{I_1}$. (C) A transcriptionally active configuration formed by an AraC dimer binding to $ara\boldsymbol{I_1}$ -$ara\boldsymbol{I_2}$ sites. The release by AraC of the $ara\boldsymbol{O_2}$ site is mediated by the binding of the inducer arabinose.