Summary of Research Activities

Ana Maria Florescu

DNA is one of the most studied molecules, interesting for chemists, physicists and biologists alike, due to its ability to store the information needed for protein synthesis, but also to its complex behavior, much of which is only revealed when DNA is perturbed by interactions with proteins. DNA-protein interactions are crucial in any living cell, having a wide variety of functions, ranging from packing DNA into chromatin to gene regulation and transcription. It is still debated how proteins find their targets along DNA, what makes them bind to a particular location on the genome, and how do they open the double helix. Powerful tools for the study of these processes are theoretical modeling and computer simulations. They help understand experimental results and often complement them, by allowing scientists to make predictions and design experiments.

My research is focused on developing such models for DNA and DNA-protein interactions (through computer simulations and analytical calculations) and using them to better understand and quantify the physical properties of DNA and the forces acting between DNA and proteins.

During my PhD thesis I have been developing and studying coarse-grained models for the description of DNA denaturation and its applications in experiments for the separation of genomes and for the binding of transcription factors to prokaryotic DNA. In my postdoctoral work I am studying nucleosome repositioning by ATP-driven molecular motors and, recently, I have extended my interest to a different area of biophysics, meiosis, and in particular kinetochore capture by microtubules.

1 Doctoral work

1.1 Transcription factor target search

The main part of my PhD thesis consisted in the study of non-specific protein-DNA interactions and of the means by which transcription factors find their targets on DNA. The debate on how proteins connect to DNA started with the experiments of Riggs, Bourgeois and Cohn[1], who measured the lac repressor-operator interaction kinetics and reported an association rate of about $10^{10} \text{M}^{-1}\text{s}^{-1}$. This is one to two orders of magnitude larger than the rate that was generally assumed for the speed limit of protein-DNA association, if it were to be a purely diffusive process. However, this rate was measured in a buffer whose ionic strength was much smaller than the physiological values. This association rate decreased and became closer to that of diffusion driven reactions once the experiments were repeated at higher salinity. Riggs et al therefore concluded that, at low salinities, protein-DNA association must be sped up by electrostatic attraction between the negative charges on DNA phosphates and positive charges on the protein. Surprisingly enough, most works performed since that time ignore this conclusion and are basically aimed at proposing
mechanisms that would enable DNA-protein association to be much faster than normal diffusion[2]-[6]. The generally accepted theory, which is confirmed by recent single molecule experiments, is that a protein connects to any site on DNA through random collisions (non-specific binding), and then searches for its specific site by sliding along the DNA sequence. It then detaches and diffuses again in the buffer if it has not found its target after a certain amount of time. This alternation of sliding and diffusion through the buffer is known as "facilitated diffusion" (Figure 1).

Most models for the description of protein-DNA interactions that were developed until now are analytical models of facilitated diffusion[2]-[5]. They have as ingredients three-dimensional and one-dimensional diffusion coefficients, as well as non-specific association and dissociation rates estimated a priori. They also make suppositions about the probability for sliding, and about how many base pairs a protein scans during a single sliding event (sliding length). These assumptions are then used to estimate the expressions of various quantities of interest, such as the association rate and the total time required to find the target, as a function of a set of well-defined geometric parameters, such as the sequence length and the cell’s volume. However, no coarse-grained molecular model adapted for molecular dynamics simulations has been proposed for the description of this phenomenon. During my thesis work I developed such a model, based on a description of double stranded DNA, which is inspired from polymer physics[7] that uses one spherical bead to model fifteen base pairs. The DNA chain is free to move in three dimensions through the cytoplasm and hydrodynamic and electrostatic interactions are also taken into account.

By studying a system composed of a cell containing a protein and several DNA segments, I showed that this model successfully reproduces some of the observed properties of real systems and the predictions of previous analytical models, like the alternation between three-dimensional diffusion and one-dimensional sliding of the protein along the DNA sequence. Even though this dynamical model indeed displays facilitated diffusion, it also shows that its existence does not necessarily imply that sampling of DNA by proteins happens at rates much larger than the diffusion limit. The most important result of my work is certainly that facilitated diffusion cannot be faster than normal diffusion by a factor larger than two, which is substantially smaller than what is sometimes believed (A. M. Florescu and M. Joyeux, Journal of Chemical Physics, 130, 015103, 2009, article selected for Virtual Journal of Biological Physics Research, Volume 17, issue 2).

I used two models for describing the protein: in the first one the protein is composed by a single rigid bead with a charge placed at its center, while the second one assumes that the protein is composed of thirteen beads connected by springs. In the second case, I investigated the association dynamics of both spherical and linear proteins, in order to study the influence of their geometry on the speed of the search process. I also investigated how other physical properties of the protein, like the charge distribution, elasticity, and presence and position of a search site, affect the DNA sampling process (A. M. Florescu and M. Joyeux, Journal of Chemical Physics, 131, 135102, 2009 - article selected for Virtual Journal of Biological Physics Research, Volume 18, issue 6).

Finally, I studied in more detail whether the results obtained with the model I developed are in real contradiction with those of previous analytical models, and tried to give a clear and well-founded point of view on protein-DNA association kinetics. By computing the acceleration of targeting due to facilitated diffusion using both types of models, I showed that, for experimental values of one-dimensional and three-dimensional diffusion coefficients, such a search strategy is most often even less efficient than normal diffusion. I have proven that, in fact, when using realistic
parameters, both types of models agree on the issue that facilitated diffusion cannot be much faster than normal diffusion (A. M. Florescu and M. Joyeux, Journal of Physical Chemistry A, 114, 9662, 2010).

To conclude this part, I should mention that, at about the same time my first article on this topic was published, in the biochemistry community it was also argued that facilitated diffusion cannot speed up significantly the targeting process. Indeed, Stephen Halford published at the end of 2009 an article entitled "An end to 40 Years of mistakes in DNA-Protein Association Kinetics?". In this article he contradicts the theory that proteins bind to DNA at rates that surpass the diffusion limit, stating that there is "no known example of a protein binding to a DNA site at a rate above the diffusion limit"[8]. He also points out the fact that Riggs, Bourgeois and Cohn did show that association rates decrease as the ionic strength of the solution increases, and that these results, although validated by subsequent experiments[9, 10], have been overlooked for years - and sometimes still are. The results obtained during my thesis work therefore come as a confirmation of these statements, and point out that, indeed, there may have been some longstanding misinterpretations in protein-DNA association kinetics.

1.2 DNA denaturation and applications in experimental biology

The work on protein DNA interactions has increased my interest in various models for DNA, so I have also implemented and improved coarse-grained DNA models for the description of DNA melting. DNA melting (which is also called denaturation) is the process by which hydrogen bonds between the basis that compose DNA break and the two strands separate. These bonds are much weaker than the covalent bonds that form in the rest of the molecule so melting does not affect the primary structure of DNA and is reversible. It can be thermal, when DNA is heated, mechanical, when a force (exerted for example by proteins) pulls apart the two strands, or chemical, when DNA is immersed in a solvent that dis-favors the formation of hydrogen bonds.

The simplest models for DNA denaturation are Ising-type (or "statistical") models that describe a base pair by two possible states: open and closed. These models incorporate the effect of both stacking and pairing interactions. The algorithms that describe melting using such statistical models have the advantage that they give quite accurate results in a very short computer time[11]. Several freeware programs that compute the fraction of DNA open base pairs as a function of temperature (melting curves) are now available. These programs are very useful for biological
applications, like genome separation, primer design, or PCR.

**1.2.1 Simulation of gene separation experiments**

In collaboration with Bénédicte Lafay (Laboratoire Ampère, Ecole Centrale de Lyon) I have built a program embedding such a code to simulate gene separation experiments through two-dimensional electrophoresis display performed in her group. This technique, which is used daily in most biology laboratories, consists in separating DNA fragments in two steps, first according to their size and then according to their sequence composition. The first step uses traditional slab electrophoresis, for example in agarose or polyacrylamide gels. Collisions between DNA and the gel reduce the mobility of DNA fragments, so that the gel acts as a sieve and the electrophoretic mobility becomes size-dependent, with smaller molecules generally going faster than larger ones. In the second dimension, fragments of identical lengths are separated on the basis of their sequence composition, thanks to a gradient of either temperature (TGGE : temperature gradient gel electrophoresis) or concentration of a chemical denaturant present in the buffer, e.g., a mixture of urea and formamide (DGGE : denaturing gradient gel electrophoresis), both methods being closely related. The effective volume of denaturated regions of DNA being larger than that of double-stranded ones, the mobility of a given fragment decreases as the number of open base pairs increases. Since AT-rich regions open at lower temperatures or denaturant concentration than GC-rich ones, GC-rich fragments usually move farther than AT-rich ones.

However, this method is still essentially empirical and simulations have only been used to a very limited extent to plan experiments and analyze results. In particular, it has been shown only recently in Bénédicte Lafay’s group [12] that separation of DNA fragments in 2D display experiments can be predicted with satisfying precision using a model that combines step-by-step integration of the equations of motion of each fragment and the use of the open source program MeltSim to estimate the number of open base pairs at each step of the DGGE phase. The method was validated by predicting the outcome of the separation of 40 sequences. However, in reference [12] the computed absolute final positions of the DNA sequences were not used, but only their positions relative to two reference segments, because the absolute positions were wrong by more than 1 cm (that is, several tens of percents of the total displacement). The fact that the errors in computed absolute final positions are so large is worrying in itself, because it unambiguously indicates that something is wrong with the model. Moreover, using relative positions is quite dangerous, since an eventual error in the coordinates of one of the reference sequences implies that the positions of all other sequences will be wrongly predicted.

I have improved this model, to describe the *absolute coordinates* of the sequences after both phases of the separation with a precision that is better than that of the experiments, and I have proposed a simpler set of equations to describe the mobilities of the sequences in the gel. (A.M. Florescu, B. Lafay and M. Joyeux, Electrophoresis, 30, 3649, 2009). I have investigated how the mobility of a partly denatured DNA sequence in a gel can be described in an accurate manner, and how certain parameters (temperature of the gel, salinity of the buffer and denaturant concentration) influence these type of experiments. I have also modified a relation which connects thermal denaturation to chemical denaturation, to take into account the actual temperature of the gel. This work showed that simulations can be established as a powerful tool to optimize experimental conditions without having to perform a large number of preliminary experiments and to estimate
whether 2D DNA display is suitable to identify a particular mutation or a genetic difference that is expected to exist between two DNA sequences.

However, DNA denaturation is not only relevant to electrophoresis, its study can provide a lot of useful information about the interactions between the building blocks of DNA and also on the forces that proteins have to exert when opening DNA in process like the initiation of transcription.

1.2.2 Modeling the DNA denaturation phase transition

From the point of view of statistical physics, DNA melting can be described as a phase transition between the open and the closed state. In order to understand in more detail this transition one has to also use dynamical models, which are based on explicit expressions of the energy of the DNA sequence in terms of continuous coordinates and velocities. Such models rely therefore uniquely on a microscopic description of the system (pair interaction parameters) and can describe the whole dynamics of DNA, from small fluctuations to the large oscillations that appear close to denaturation.

During the period I spent in Marc Joyeux’s group I have worked on the study of various models of DNA denaturation: I re-parametrized a dynamical model for DNA that was previously developed in Marc Joyeux’s group[13] and obtained a better agreement with experimental results that were not taken into account up to now: the critical force needed to keep two DNA strands separated and the sequence size dependence of the critical temperature. Then, I showed that the results obtained with the improved model agree well with those obtained from statistical models. Finally, I checked that the critical properties of the dynamical model remain qualitatively similar to those predicted with the original set of parameters and that DNA denaturation looks like a first-order phase transition in a broad temperature interval. Our results also show that, however, there necessarily exists, very close to the critical temperature, a crossover to another regime. The exact nature of the melting dynamics in this second regime still has to be elucidated(M. Joyeux and A.M. Florescu, Journal of Physics: Condensed Matter, 21, 034101, 2009).

Moreover, I have adapted a coarse-grained model (three sites per nucleotide) that was previously parametrized for the description of DNA oligosequences [14] to describe unzipping of longer
sequences. I have shown that this model predicts (i) critical temperatures for long homogeneous sequences that are in good agreement with both experimental ones and those obtained from statistical models, (ii) a realistic step-like denaturation behaviour for long inhomogeneous sequences, and (iii) critical forces at ambient temperature of the order of 10 pN, close to measured values. It furthermore supports the conclusion that the thermal denaturation of long homogeneous sequences corresponds to a first-order phase transition. In contrast, it yields a critical exponent for the critical force that is higher than the ones predicted by other models and therefore suggests that the scaling law relating the characteristic exponents for specific heat and critical force that was derived for one-dimensional models no longer holds for this more complex description.

This model is both geometrically and energetically realistic, in the sense that the helical structure and the grooves, where most proteins bind, are satisfactorily reproduced (Figure 2), while the energy and force required to break a base pair lie in the expected range. It therefore represents a promising tool for studying the dynamics of DNA-protein interactions at larger time scales than those accessible through all-atom molecular dynamics simulations. This model has the advantage that it can be used to describe both DNA denaturation and DNA-protein interactions (being a bead spring model, the DNA chain can move in three dimensions, which is not possible in the models generally used of studying denaturation, and one can also include hydrodynamic interactions if a better description of the surrounding solvent is needed) (A.M. Florescu and M. Joyeux, Journal of Chemical Physics, 135, 085105, 2011 - selected for Virtual Journal of Biological Physics Research, Volume 22, issue 5). I am still pursuing a part of this work as a collaboration with prof. Ioan Andricioaei’s group at UC Irvine (manuscript in preparation).

2 Postdoctoral work

I am currently a fellow of the Max Planck-CNRS Post-Doctoral program for System Biology, which gives me the opportunity to run my personal research program hosted by the Max Planck Institute for the Physics of Complex Systems in Dresden (the group of Pr. Dr. Frank Jülicher- main host group) and collaborating with the Interdisciplinary Research Institute in Lille (the group of Dr. Ralf Blossey - minor host group). My current work focuses on applying models from statistical and polymer physics to the simulation of certain biological processes. I am currently working on several projects, some of them involving collaborations that I have started in the last two years, also with groups outside my host institutions.

As a natural extension to my work on DNA-protein interactions I became interested in the activity of chromatin remodeling factors. They are protein complexes that appear in eukaryotic cells, where DNA is compacted in chromatin, and its accessibility for various proteins needs to be regulated. Chromatin remodeling factors do this by using energy obtained from ATP hydrolysis to modify the structure and dynamics of chromatin. I am studying their mechanism and in particular how certain aspects of their functioning and regulation trigger nucleosome organization along genomes. To this aim, I am interacting with Ralf Blossey’s group at the Interdisciplinary Institute in Lille and Geeta Narlikar’s group at University of California in San Francisco and using some of their experimental data in the development of theoretical models.

My most recent project concerns a different field of cell biology: meiosis. During meiosis, when the genetic material is separated between the two daughter cells, it is pulled apart by microtubules.
In collaboration with experimentalists from Iva Tolic-Norrelykke’s group at the Max Planck Institute for Cell Biology and Genetics in Dresden we are trying to understand the mechanisms by which microtubules attach to chromosomes via kinetochores during cell division.

2.1 Kinetic proofreading and nucleosome positioning

In the coding regions of eukaryotes, nucleosomes are aligned regularly relative to transcription start sites. The factors that determine this alignment (nucleosome positioning) have been studied intensively in the recent years, and current descriptions go along three major trends (which are, however, not mutually exclusive):

1. **Sequence dependence positioning of nucleosomes.** It is believed that the positioning of nucleosomes along DNA is also influenced by the nucleotide dependence of DNA bending rigidity, and certain nucleotide sequences have been showed to favor nucleosome formation[15].

2. **Statistical positioning along DNA.** The nucleosome array is considered to act similar to a confined one dimensional fluid or Tonks gas[16, 17]. The spacing of nucleosome is induced by the cooperative effect of two factors: the nucleosome density and the presence of a potential energy barrier, which can be, for example, another protein bound on DNA.

3. **Positioning induced by chromatin remodeling factors.** Experiments performed on yeast show that genome organisation is less pronounced in the absence of ATP, stressing thus the role played by active chromatin remodeling factors[18, 19]. The activity of some chromatin remodeling factors has been shown to be coordinated by two factors: the acetylation of the histone tails[20] and the length of internucleosomal DNA (specifically for one class of chromatin remodeling factors, ISWI/ACF)[21]. ACF’s capability to sense DNA length allows it to influence the organization of chromatin by contributing to the formation of regularly spaced arrays, which are a mark of transcriptionally silent regions. Also, their sensitivity to histone tails permits the selective silencing of genes, driven by post-translational modifications. The presence of these control mechanisms has led to the suggestion by the Blossey and Narlikar groups of a kinetic proofreading mechanism for chromatin remodeling[22, 23, 24]. In this scenario, a series of equilibrium and non-equilibrium (ATP-driven) steps, leads to a very low error rate in the remodeling outcome, and thus avoid the silencing of genes whose product is still needed.

In order to further demonstrate this feature and to study its consequences on the dynamics of larger arrays of nucleosomes, in collaboration with Ralf Blossey and Helmut Schiessel, I have built a simple model for the motion of the nucleosome-remodeler complex that is suitable for Monte Carlo simulations. The model is calibrated with the experimental data for the length dependence of the speed of remodeling by ACF obtained in Geeta Narlikar’s laboratory[21].

I have studied an array of nucleosomes with fixed boundary conditions (which can be assumed equivalent to the energy barrier formed by, for example, a transcription start site) using two types of displacements of nucleosomes: (i) with rates that are length dependent, and (ii) with constant rates. Firstly, I have computed the probability that a given site on the genome is part of a nucleosome (occupancy), and shown that, due to the length dependence, the nucleosomes appear to be better positioned, because patterns in site occupancy along the sequence are better defined. To further quantify this length-dependence I have also computed the distribution of linker lengths, showing that indeed the activity of ACF favors regularly spaced nucleosomes and studied the dynamics of arrays with varying nucleosome densities. Our work is one of the first models that includes the
Figure 3: Site occupancy of a DNA sequence computed with our model. It shows the probability that every site on the sequence is covered by a nucleosome, in black for chromatin remodeling-driven positioning and in red for statistical positioning. Coordinates along the DNA sequence are normalized to the total size of the genome. The green line shows the probabilities computed using the Kornberg model for statistical positioning. [16]

features of directionality and proofreading in the context of chromatin remodeling. We have proved that appearance of an intrinsic long range scaling for inter-nucleosomal DNA length (A. M. Florescu, H. Schiessel and R. Blossey, Physical Review Letters, 109, 118103, 2012).

2.2 Bending fluctuations in semiflexible polymers

I am currently studying fluctuations in the bending of semiflexible polymers and their dependence on the polymer’s length. The aim of this project is to understand how certain motor proteins have the ability to sense length. One possible explanation for the length dependence of the rate by which the chromatin remodeling factor ACF displaces nucleosomes could be that the magnitude of the local fluctuations in the bending of internucleosomal DNA depends on the length of the chain and affects the binding or the activity of the translocation domain. We have developed a model to quantify such length dependent fluctuations for short polymers under external forces. We have furthermore confirmed our model by using Monte Carlo and molecular dynamics simulations, and studied how the fluctuation spectrum and the modes of vibration of the polymer are changing with its length and with the external force applied at one or both ends. We are now evaluating the consequences of such variations on the motility of a molecular motor interacting with the polymer, and study how could it use energy obtained from ATP hydrolysis to increase its sensitivity to length. Once complete, this model will be tested through comparison with the experimental results of Yang et al[21], who have measured the rates of nucleosome remodeling by the chromatin remodeling factor ACF as a function of the length of internucleosomal DNA length (collaboration with Kuni Iwasa and Frank Jülicher).
2.3 Kinetochore capture in meiosis

When a cell divides it splits the genetic material equally between the two daughter cells. During cell division chromosomes group to form the spindle, which will attach to microtubules that will then pull it apart, separating the chromosomes for the daughter cells. Microtubules attach to the spindle through kinetochores, which are protein complexes that form on chromosomes. However, how microtubules find kinetochores and bind to them is still not clear. Researchers in Iva Tolic-Norelykke’s group have shown using in vitro experiments and mathematical modeling that one potential search mechanism is by microtubules pivoting around the spindle pole [25]. We are currently extending this work to understand how growth and shrinkage of microtubules affects this process and whether this mechanism is still valid in vivo. I am simulating the swiping motion of microtubules around the spindle pole, taking into account their limited lifetime and the presence of dynamic instabilities and computing the microtubule-kinetochore binding rates (collaboration with Geo Cojoc, Iva Tolic-Norelykke, Nenad Pavin and Frank Jülicher).

References
