The handle http://hdl.handle.net/1887/37026 holds various files of this Leiden University dissertation.

Author: Schram, Raoul Diederick
Title: Enumeration and simulation of lattice polymers as models for compact biological macromolecules
Issue Date: 2015-12-15
Chapter 1

Introduction

If only our world was two-dimensional. Or if we had been given just one extra spatial dimension, beyond our three-dimensional world. Polymer physics would be a very different field. We would have analytic expressions for the relevant critical exponents. Instead, the three-dimensional case is often the hardest one theoretically, but then again also the most interesting.

Polymer physics was in its full bloom during the late 60’s and 70’s, when de Gennes [1] and Edwards [2] spearheaded its development. In the last decade a new surge of interest has revitalized polymer physics. This renewed interest can be attributed for a large part to advancements in experimental techniques for studies on biopolymers such as proteins and DNA.

Biopolymers are, more often than not, very complex and exhibit a variety of structures at different length scales. We can ask ourselves how comparatively simple polymer models can say anything about a system such as the human chromosome, which is so complex that we are hardly scratching the surface of our understanding of its mechanisms. Then perhaps a counter-question is appropriate. How can we ever hope to fundamentally understand a complex system such as the human chromosome, if we do not even understand the simple models representing it?

Perhaps both viewpoints are valid, in that we can neither understand complex biopolymer systems using only simple polymer physics, nor can we understand them without. Thus, as a theorist we are given a choice: how complex are the models that we decide to study? In this thesis, we study relatively simple models. The advantage is that the
results presented are more general (fundamental if you will). The downside is that direct applications to biological systems are hardly ever possible.

An interesting case to explore this idea of experimentally versus fundamentally driven approach is the story of the fractal globule. It started as a crumpled globule in a paper by Grosberg et al. [3] in the late 80’s, where they suggested that DNA is folded in a fractal way, such that pieces of DNA close together along the chain are close together spatially as well. It was suggested that this organization of the DNA allows better accessibility for transcription due to a lack of entanglements. At this point in time experiments were technically not advanced enough to probe either of these statements. Thus, in the early stages studies were driven from a more fundamental theoretical point of view.

About a decade later, things started to improve on the experimental side with a new method called Fluorescence In Situ Hybridization (FISH) [4]. This was one of the first methods that could probe large scale chromatin organization \textit{in vivo}. As exciting as this prospect was, it was unfortunately marred by large sample to sample variation, which made it impossible to draw any firm conclusions on the subject of the existence or non-existence of the fractal globule.

Then an experimental breakthrough came with the advent of a new technique called Chromatin Conformation Capture (CCC) [5]. It rapidly evolved in various iterations such as 4C [6] and Hi-C [7]. Interestingly, whereas the FISH experiments directly probed the spatial distances of segments of the chromatin, these new CCC experiments would instead probe whether two segments anywhere on the chromatin had a high or low likelihood of being spatially near. They called this the contact probability. Now, in an interesting twist, the theory of the fractal/crumpled globule in itself does not predict the behavior of the contact probability without further assumptions, see Appendix A for a derivation.

This is where the tide seemed to turn in favor of a more experimentally driven approach. Experiments showed an interesting feature at the very large chromatin length scales: for slightly more than a decade in genomic distance, the contact probability decays as a $-1$ power law”. This spurred a new surge of interest in polymer models, that would show this power law. Many options were suggested by different studies, including the interdigitating fractal globule, crosslinked globules and ring polymers. This became a hotly debated subject, and I truly mean \textit{hotly} debated. During talks regarding these models, the same question came up with a definite regularity: “Does your model observe the $-1$ power law?”. The answer was almost invariably positive.
Among these models that were suggested, one model stood out from the crowd. It is a model that was studied before the CCC experiments were performed, and was, at least originally, a subject of study that was driven by fundamental curiosity. The nonconcatenated ring polymer model was first studied in depth by Cates and Deutsch in 1986 [46]. In this model, ring polymers are configured topologically, such that they are separable without breaking the chain. The rings themselves are topologically equivalent to a trivial loop (without knots). Cates and Deutsch conjectured that the extension of these rings would asymptotically grow slower with length than linear polymers. The difference being due to the topological constraints from neighboring ring polymers, which could only be resolved by breaking the chain.

Recent work suggests that in fact rings are as compact as the fractal globule, and have the same fractal dimension. This is a curious find indeed, when you consider that the difference with a linear polymer is only a single bond. It shows the impact that topological constraints can have on the equilibrium properties of polymers. What happened in those 25 odd years, that enabled us to revisit a problem and push the boundaries of our understanding?

For a large part it was the increase in compute power that enabled us to do longer simulations on larger systems. Due to major manufacturing improvements, year-over-year, the available compute power has increased exponentially over the years. If a problem was not solvable, one could wait several years to obtain the computational improvements needed. Unfortunately, this rapid increase in compute power is slowing down, especially if we consider the performance of a single core. We are simply reaching the limit of how fast a transistor can switch with the current state-of-the-art manufacturing technology.

Regardless of improving manufacturing capabilities, there are several other ways (as physicists) to slash computational barriers. Two of those are:

- Improved algorithms
- More effective use of available resources [parallelism]

Significant improvements of algorithms in statistical physics are often hard to come by. Often we are limited to small improvements of implementation such as reducing the memory footprint and better cache usage. These kinds of improvements are useful, but limited. There are other types of problems that are more suitable for algorithmic improvements. One type of these are exact enumeration problems. In these problems, we
count the exact number of configurations and depending on the problem, we are able to obtain other observables exactly as well. Algorithms for solving these questions can be vastly superior in efficiency to a brute-force approach. The running time complexity of practically any exact enumeration algorithm is exponential in the system size, i.e. $O(\exp(\alpha N))$, but clever algorithms can reduce the exponent $\alpha$, thus creating exponential speed-ups in system size. One such algorithm (exact enumeration of Hamiltonian walks) is presented in this thesis.

The other improvement, parallelism, is already an important component in many compute intensive calculations. This is only set to increase in the coming years. It is often much cheaper to manufacture three processing units than one processing unit that is three times as fast. This is also the reason that we have jumped from having only a single core per CPU towards 12 or even more cores. A trend that tends even more in the direction of parallelism is the usage of the GPU (Graphics Processing Unit). This compute device was originally intended to just show (video game) graphics on the screen, but since graphics is very compute intensive, with many of the same instructions on different objects (pixels), it is also very good at number crunching in general. At a very basic level, the GPU consists of thousands of small processing units that are slow on their own, but very fast when their compute power is combined. Thus, harnessing this compute power for statistical physics is a goal very much worth pursuing.

The organization of this thesis is as follows: Chapter 2 discusses the fractal globule, and investigates its dynamical properties. It is the chapter most closely linked to experimental results. The following chapter is of a more technical nature and introduces a very fast implementation of a Monte Carlo polymer simulation model on the GPU. It also introduces an implementation on the CPU that is able to use multiple cores. Chapter 4 makes use of this implementation to simulate a melt of rings. This system is analyzed using Rouse modes, which prove to be a very useful tool in analyzing both the dynamic as well as static properties. An exact enumeration algorithm for Hamiltonian walks is presented in Chapter 5, and we find the number of them to be significantly higher than previous studies predicted. Additionally, a well known algorithm for other polymer systems (PERM) is used to simulate systems up to a larger size, with some unexpected results.