Antony Lesage

Docteur en Physique

5 rue Robineau
75020 Paris
+33 (0)6 24 43 24 94
lesage.antony@gmail.com
github.com/antonylsg

Thèse

titre *Inférence bayésienne des paramètres structuraux de la chromatine de la drosophile sur des images super-résolues de domaines épigénétiques*

directrice Maria Barbi

laboratoire Laboratoire de Physique Théorique de la Matière Condensée (LPTMC)

Formation

2016–2019 **Doctorat en Physique**, *Sorbonne Université*, Paris.

2013–2016 **Magistère de Physique fondamentale d'Orsay**, *Université Paris-Saclay*, Orsay, *avec mention Bien*.

2015–2016 **Master 2 en Systèmes Complexes**, *Université Paris-Saclay*, Orsay, *avec mention Bien*.

2014–2015 **Master 1 en Physique Fondamentale**, *Université Paris-Sud*, Orsay, *avec mention Assez Bien*.

2013–2014 **Licence en Physique Fondamentale**, *Université Paris-Sud*, Orsay, *avec mention Assez Bien*.

2011–2013 **Classes Préparatoires aux Grandes Écoles**, *Lycée Franklin-Roosevelt*, Reims, PCSI-PC *mention globale A*.

2009–2011 **Baccalauréat Scientifique**, *Lycée Godart Roger*, Épernay, *avec mention Bien et mention européenne allemand*.

Publications

[1] **Mai 2019** (avec Antony Lesage et al.). « Polymer coil–globule phase transition is a universal folding principle of Drosophila epigenetic domains ». In : *Epigenetics & Chromatin* 12.1. DOI : 10.1186/s13072-019-0269-6. URL : https://doi.org/10.1186/s13072-019-0269-6.

[2] **Fév. 2019** (avec Marius Socol et al.). « Contraction and Tumbling Dynamics of DNA in Shear Flows under Confinement Induced by Transverse Viscoelastic Forces ». In : *Macromolecules* 52.4, p. 1843-1852. DOI : 10.1021/acs.macromol.8b02184. URL : https://doi.org/10.1021/acs.macromol.8b02184.

WEINBERG COLLEGE OF ARTS & SCIENCES

Department of Physics and Astronomy
2145 Sheridan Road
Evanston, Illinois 60208-3112

www.physics.northwestern.edu

Judd A. and Marjorie Weinberg College of Arts and Sciences
Northwestern University

Phone 847-491-3685
Fax 847-491-9982

NORTHWESTERN UNIVERSITY

August 29, 2019

Ecole doctorale No. 564
Physique en Ile de France
Sorbonne Université
Paris, France

Report on Ph.D. thesis of Mr. Antony Lesage *Inférence bayésienne des paramètres structuraux de la chromatine de la drosophilie sur des images super-résolues de domaines épigénétiques*

I am pleased to report that this thesis, which reports quantitative modeling and analysis of the structure of chromosomes using a combination of theoretical physics from polymer statistical mechanics and analysis of experimental data from molecular biological analysis of chromosomes in cells of fruit flies, is of the highest quality, originality and importance to its field of research. The thesis is deserving of acceptance by Sorbonne Université as a doctoral thesis. I will summarize the major results and my evaluation of them, interspersing suggestions and questions (indicated by ★) throughout my report.

The aim of the thesis is twofold: first, to develop the state of the art of theoretical physics of polymers, particularly the description of the dense "collapsed" or "globule" phase. The second aim is application of the theoretical results to description and analysis of experimental data for chromosome organization in cells, published by other research groups. These two aims are brought together very nicely in the thesis, with the quality of theory at the highest levels of statistical mechanics, and with the analysis of experimental data being thorough and precise, using a state-of-the-art maximum-entropy (also called Bayesian) data modeling approach. The combination of these elements make the thesis a useful, novel, satisfying and impressive piece of scholarship in the area of genome biophysics.

Chapter 1 begins the thesis with an excellent summary of biological facts concerning genome organization and chromatin structure, introducing the reader to the genomics and chromosome organization of *Drosophila*. I have a few minor suggestions regarding this section.

★Biologists consider there to be three kingdoms of life (not two): eukaryota, bacteria and archaea (p. 1).

★I was quite intrigued by the quote from Flemming and was interested in what exactly he meant by chromatin's "refractile nature"? (p. 2) And why exactly is euchromatin more easily stained – simply because of its less dense nature? Or is there something more chemical in play?

★On p.4 it might be good to mention supposed functions of linker histone H1 as it is shown in Fig. 1.5.

★In Fig. 1.6 it would be good to show the arrows going both ways for each reaction (e.g., demethylation as well as methylation).

★On p.5 it might be good to mention some of the histone marks that work the opposite to methyl-repression and acetyl-activation, just to give the reader some idea of the complexity of the regulatory functions of histone marks (e.g., H3K4me; a table of histone marks relevant to gene regulation in *Drosophila* might be a nice thing to include). Also the role of PRC2 might be briefly touched on.

★This reviewer knows what HP1 is (although this reviewer is ignorant of whether *Drosophia* has multiple HP1 types as in the human case where we have HP1α, HP1β…), but the reviewer is ignorant of what the SU(HW) trace represents (Fig. 1.10). Also – what function are the red, blue and black domains in the figure – the same as in Fig. 1.11 (where there are five domains?). It might be good to indicate the histone markings associated with these different color-coded domains, and to indicate whether the same color codes are used through Figs. 1.14-1.16.

Chapter 2 begins development of the theoretical part of the thesis, with a simple but deep discussion of radius of gyration and how it can be decomposed into different contributions (monomer mass distribution vs. conformational contributions). This is followed by a very clear, parallel development of what this reviewer considers a rather novel formalism for a *faisceau du polymers* or a "polymer bundle", or polymers

which have highly correlated conformations so as to follow one another. Again an additive decomposition of radius of gyration into a series of collective variables (micromeres, micro-conformeres, macro-conformeres) is demonstrated, which will have utility for the description of chromosomes since they have been demonstrated to be conformationally bundled in the *Drosophila* cells of interest (Fig. 2.6).

★It might be good to note what the bundle number n is in Figs. 2.5-2.6 (it is n=2 or are more chromosomes bundled?) and also to change the n in the caption of Fig. 2.5 (where n = 1564 and 5052 refer to numbers of cells analyzed rather than the bundle number?).

★Similarly, in Fig. 2.8 it might be good to note that the bundle shown has n=4 (the number relevant to the subsequent analysis of the tetraploid cell line).

I was struck how elegantly and naturally one was led to the χ^2 distribution for the conformational size distributions for bundles thus quickly explaining the distributions observed in Fig. 2.7.

★I wanted to check that the λ and σ in Sec. 2.5 for the "Maxwellian bundle" (which could also be called a Gaussian polymer bundle) represented precisely the same quantities as those in Sec. 2.4 (I think this is the case and perhaps it is even stated but good to make very clear if true).

Chapter 3 focuses on developing physical models of polymers to a larger extent, introducing the familiar random flight and semiflexible polymer models.

★It might be good to mention that both models can be developed starting from rather simple Hamiltonians using Boltzmann statistical mechanics (for the random flight, H=0, while for the semiflexible polymer it is just the integral of curvature-squared, the simplest elastic model for bending a thin rod). In the semiflexible polymer case this approach can be used to *derive* the key correlation function preceeding Eq. 3.6 (it is stated as fact in the thesis as it stands). It would also be good to state not just that $l_K = 2L_p$ but also that $N = L/l_K$.

★The narrative then moves to discussion of "real chains" and their statistics – the radius exponents reported are of course for d=3 (Table 3.1) and. it might be nice to note the dimension dependence of ψ.

★On p. 41, given that the semiflexible polymer has been invoked, it might be good to note that the stretching regime can involve two exponents, corresponding to Pincus stretching (if self-avoidance is present) plus a later bending-fluctuation-mediated regime, and what the relevant exponents are in d=3. Notably the latter strong-stretching is (probably) not relevant to the crossovers of interest in the thesis, but it is worth mentioning its distinct physics (I think there will be a later, additional crossover from Pincus to semiflexible chain behavior?).

Chapter 4 focuses on further development of polymer theory to the description of the coil-globule transition, which will be highly relevant to description of epigenetic chromatin compaction later in the thesis. This chapter gives a rather clear summary of extant theory concerning the polymer statistics of chains subject to self-attraction.

★There appears to be inconsistency concerning the sign of the parameter ε (a familiar issue in statistical mechanics thanks to the Boltzmann distribution's minus sign in $e^{-\beta E}$). When the energy J is introduced (p. 50) we are told that the energy of contacts is $E_c = -Jm$, indicating that $J > 0$ will drive collapse (negative energy and higher probability of configurations with contacts). Then on page 52, we have the definition $\varepsilon = -\beta J$, meaning that $\beta E_c = \varepsilon m$, but now the relevant values of ε are negative (inconsistent with later usage) and inconsistent with Eq. 4.9. Perhaps the definition at the top of p.52 should be $\varepsilon = +\beta J$?(giving J and ε the same sign).

Chapter 4 ends with a key result – the free energy as a function of ε, which will be a key tool used later in the thesis. I note the statement that the $a_i(\varepsilon)$ and $b_i(\varepsilon)$ are analytic (at least sufficiently smooth?) functions of ε (just under Eq. 4.12) as this will be important to this report shortly.

Chapter 5 describes numerical simulations of the random walk with self-avoidance and self-attraction using Monte Carlo methods.

★It is worth noting that the usual Metropolis rule for $A(c',c)$ shown at the bottom of p. 59 is just one of many possibilities (sometimes it is of value to use one of the others depending on the distribution of energy differences and there can be reasons to choose other rules that satisfy detailed balance). Perhaps there is a principle that exists for choosing specific acceptance probabilities (although the Metropolis one is somewhat asymmetric perhaps it maximizes total acceptance probability which would be desirable for most simulation situations).

The description of the maximum entropy (Bayesian) approach to the fitting of parameters to describe the numerical results is well explained, and then the use of a reasonable physical model for finite-size corrections due to globule surface tension is demonstrated to improve fitting of the model to numerical results for coil size versus chain length with excellent final agreement. Given the global weak- to strong-coupling crossover nature of the results (Fig. 5.5) the numerical agreement between analytical and simulation methods is impressive.

This allows the free energy expansion parameters $a_i(\varepsilon)$ to be extracted, as a function of the attraction parameter ε. This leads to some questions regarding the theory – the $a_i(\varepsilon)$ were hypothesized to be analytic functions of ε but the final analysis yields quite singular-looking behaviors (Fig. 5.6). There appear to be singularities or at least very sharp features at the collapse transition (or theta point, near $\varepsilon = 0.29$) and at a second, weaker value of coupling ($\varepsilon = 0.17$).

★Should the earlier statement about the $a_i(\varepsilon)$ being analytic be qualified ("analytic away from phase transitions", or "piecewise continuous")? Some more explanation of the expectation for this could be added in Chapter 4, and then this would be less surprising when singular behavior of the $a_i(\varepsilon)$ appears in Chapter 5.

★It would also be great to have a more physical description of these singularities – what thermodynamic aspects of the collapse transition does, e.g., the cusp in a_3 reflect?

★What is the physics associated with the additional kink apparent in a_2 and a_4 near $\varepsilon = 0.17$? Given that these additional kinks likely reflect non-smooth behaviors in the simulation results there may be something quite interesting to examine further there. However, if the appearance of the $\varepsilon = 0.17$ kink is just a consequence of the parametrization of the free energy this should be commented on (to the reviewer this seems unlikely but the author has no doubt a deeper understanding of the situation).

Chapter 6 describes application of the theory for collapsed polymers to experimental data for *Drosophila* chromosomes, with what I would describe as great success. The fitting allows excellent description of the coil size distributions (Figs. 6.2-6.4) as well as means and medians (Fig. 6.5), using a rather limited number of easily physically understood parameters (Table 6.2). Essentially, two models are presented in Chapter 6: an extremely simple pure power law scaling model (of Boettiger *et al.*, Fig. 6.1, two parameters per curve) and the polymer physics model of this thesis (summarized by the coil size data and fits of Fig. 6.5, 4 or 6 parameters per curve). The theory of this thesis is based on a microscopic Hamiltonian and gives additional insight into underlying physics (rather than an unrestricted power-law fit). However, the theory also provides more flexibility for fitting with the additional parameters (although the fits are done differently, with maximum-entropy fitting being done to the coil size distributions in this thesis rather than the more elementary least-square minimization applied to coil sizes of Boettiger).

★What is the relative, quantitative, goodness of fit of the two models (in the standard χ^2 sense)? The theory certainly provides new statistical-mechanical (physical) insight into the experimental data, but as a model for the data is it also providing a superior description from a purely statistical perspective?

★Is the bundle number *n* relevant to the fitting or is the faisceau model introduced in Sec. 6.1.2 independent of *n* (*i.e.*, dependent only on the bundle thickness?).

The theory of this thesis gives valuable insight into what the polymer properties of chromatin are, and they are quite interesting. Euchromatin has a very weak self-attraction ($\varepsilon \approx 0.15$) and a quite short segment length (≈ 20 nm, 0.4 kb, just 2 to 3 nucleosomes) while heterochromatin has sufficient self-attraction to collapse ($\varepsilon \approx 0.36$) and a bulkier/longer segment (20 to 60 nm, 1 to 3 kb, 5 to 15 nucleosomes, very roughly). As the author points out, determination of polymer properties of chromatin has been problematic (for many reasons) and this new polymer-statistical-mechanics-based estimate including compaction effects will be of great interest. As mentioned in the thesis, the rather short (2-3 nucleosome) segment length appears consistent with recent methods for imaging individual nucleosomes using electron microscopy and live-cell-microscopy methodologies.

★The author also rightly points out past estimates of the segment length that have been much longer/larger (*e.g.*,work of Bystricky *et al.* giving 400 nm, that of Dekker *et al.* giving more than 100 nm, discussed on p. 85). With the new theory in hand, does the author have any insight into why those prior studies concluded such large values (400 nm seems completely off-scale).

★Similarly, it might be worth explaining why Boettiger *et al.* arrived at $\varepsilon \approx 3.5$ which is tenfold larger than the value found here (the discussion on p. 86 could have one or two sentences added).

★It would be very nice to include in Ch. 6 a figure showing images of typical domains of a few different sizes (say 10 kb, 50 kb, 200 kb and 500 kb) for the parameters arrived at for active, inactive and repressed domains, just to give the reader an idea of what kinds of structures the final model produces.

In conclusion I found the thesis to be of the highest quality and originality, presenting a synthesis of new ideas and results in polymer theory along with applications to analysis of state-of-the-art experimental data from cell and molecular biology. The results will be of broad interest to workers I biophysics and molecular/cell biology, particularly to those interested in organization of genomes. Evidence for this is that portions of the thesis have already appeared in a published peer-reviewed paper led by the author of this thesis; I understand that at least two more journal papers are forthcoming. The changes I have outlined are all suggested; the thesis as it stands is acceptable. Of course I hope the author finds my comments and suggestions useful and incorporates them into the final accepted version.

Yours truly,

John F. Marko
Professor of Physics and Astronomy
Professor of Molecular Biosciences
Northwestern University
john-marko@northwestern.edu

Centre de Biochimie Structurale
29 rue de Navacelles, Montpellier, France
Tel/Fax: +33 4 67 41 79 07/13
Email: diego.cattoni@cbs.cnrs.fr

Montpellier, le 4 septembre 2019

Rapport Thèse de Antony LESAGE pour obtenir le grade de Docteur de Sorbonne Université

L'objectif de la thèse est de développer un modèle basé sur des principes de la physique de polymères capable de décrire les propriétés physiques et statistiques de l'ADN en fonction de son état épigénétique. Pour valider son modèle, Mr Lesage s'est servi des résultats obtenus précédemment par des méthodes d'imagerie de super résolution chez la Drosophile.

La découverte récente de l'existence d'un nouveau niveau d'organisation dans la chromatine chez les eucaryotes a bousculé notre regard du rôle de la structure 3D dans la régulation des processus nucléaires. Le champ d'étude de l'organisation de la chromatine est en ébullition actuellement, dont la création des modèles physiques capables d'interpréter les nouvelles données est impérative. Les méthodes développées dans la thèse permettent d'analyser des résultats d'imagerie à l'échelle de la molécule unique et d'avoir accès à des paramètres structuraux et énergétiques très raffinés. Les travaux de thèse de Mr. Lesage sont donc très pertinents et d'un intérêt majeur pour la communauté physique et biologique en général.

Le manuscrit est très bien rédigé, clair et pédagogique. Il s'organise en sept chapitres dont une introduction et une conclusion, quatre chapitres de présentation de la méthodologie et des outils développés ainsi qu'un chapitre où le modèle proposé est confronté avec des résultats d'imagerie de super résolution. L'introduction est concise et décrit brièvement les aspects essentiels de l'organisation de la chromatine et présente les données expérimentales qui seront analysées par la suite. Les chapitres suivants présentent d'une façon très complète tous les aspects de la physique des polymères et comment, en partant des principes fondamentaux, le modèle est construit pas à pas pour prédire les distributions du rayon de giration dans des conditions épigénétiques différentes. Les hypothèses et considérations pour chaque étape de la construction du modèle sont clairement décrites et justifiées. La version initial du modele est presente elegamment dans une équation capable d'estimer l'énergie libre d'une marche auto-évitante attractive contenant tout la phénoménologie de la transition coil-globule, un aspect proposé par la communauté spécialisé dans la chromatine comme acteur essentiel de la régulation transcriptionnel. Ensuite le modèle est testé par des simulations numériques en employant

Centre de Biochimie Structurale
29 rue de Navacelles, Montpellier, France
Tel/Fax: +33 4 67 41 79 07/13
Email: diego.cattoni@cbs.cnrs.fr

l'algorithme de Metropolis implémenté par Mr Lesage. Cette stratégie a permis de démontrer, après une analyse exhaustive, que la tension de surface, surtout dans les conformations globulaires, joue un rôle très important et doit être inclus dans la version finale du modèle.

Les résultats du modèle complet permettent d'avoir accès à des paramètres physiques et thermodynamiques de difficile estimation par d'autres stratégies, telle que la longueur de persistance de l'ADN *in vivo* et l'énergie d'interaction entre monomers. Intéressement, l'ajustement du modèle aux données expérimentales montre que les aspects structuraux des domaines actifs et réprimés sont extrêmement similaires et que la différence entre eux est née de l'énergie d'interaction, ce qui faciliterait une rapide transition entre états de la chromatine. A noter aussi que, pour la première fois à ma connaissance, le modèle de l'organisation spatiale de la chromatine contemple la polyploïdie chromosomique chez la Drosophile, donnant donc une valeur ajoutée à la signification biologique des prédictions. Finalement, les résultats de l'ensemble permettent à l'auteur de conférer un rôle clé aux marques épigénétiques des histones en régulant l'attraction entre nucléosomes et en conséquence l'architecture globale de la chromatine.

Le coeur du travail de thèse a déjà porté ses fruits et vient d'être publié dans un journal spécialisé reconnu internationalement et Mr Lesage est co-auteur aussi dans un deuxième travail. Au delà des travaux publiés, la stratégie de modélisation développée par Mr Lesage fait partie d'un projet récemment financé par l'ANR pour étudier l'association entre la dynamique structurale de la chromatine et la régulation transcriptionnelle chez les eucaryotes.

En conclusion, **je donne mon accord sans réserve pour la soutenance de thèse par Antony Lesage.**

Diego I. CATTONI
PhD, HDR
Chargé de recherche au Institut National de la Santé et de la Recherche Médicale

Rapport de soutenance de thèse – Antony Lesage

M. Antony Lesage a présenté avec aisance, de façon très didactique et synthétique, l'ensemble de ses travaux de thèse, portant sur un sujet complexe de physique théorique à l'interface avec la biologie et incluant des développements originaux. Cet exposé, s'appuyant sur une iconographie soignée, a montré l'ampleur du travail accompli et la qualité des résultats obtenus, ainsi que la très bonne maitrise du candidat de son sujet de thèse. Antony Lesage a en effet apporté une contribution significative dans le domaine de l'étude de la chromatine, notamment en développant une approche originale et élégante d'inférence de paramètres basée sur la physique statistique de faisceaux de polymères. L'application de cette méthode à des données expérimentales de microscopie lui a permis de montrer que la fibre de chromatine in vivo est très flexible et dans un état proche de la transition coil-globule. Ces résultats auront, à n'en pas douter, de très fortes implications dans le domaine.

Au cours de cet exposé et de la discussion qui s'en est suivie, le jury a particulièrement apprécié la précision du discours du candidat, ses explications pédagogiques et son sens critique, montrant une excellente capacité d'analyse. L'ensemble de la soutenance a démontré la forte maturité de M. Antony Lesage et sa capacité à aborder des notions biologiques complexes, dans un contexte multidisciplinaire, dans le but de combiner données théoriques et expérimentales. Le jury a également souligné la curiosité et l'opiniâtreté dont a fait preuve Antony Lesage pour mener à bien un projet ambitieux, ainsi que la rigueur exemplaire de son travail.

En considération de ces éléments, le jury s'est accordé pour saluer le travail remarquable accompli par M. Antony Lesage, et par conséquent, lui délivrer à l'unanimité le grade de docteur de Sorbonne Université. Le jury a en outre tenu à lui adresser ses félicitations orales pour la qualité du travail réalisé.