

## Modéliser et simuler les chromosomes: propriétés physiques et fonctions biologiques

Pascal Carrivain<sup>1,2</sup>, Jean-Marc Victor<sup>1</sup> et Annick Lesne<sup>1</sup>

(1) *Laboratoire de Physique Théorique de la Matière Condensée (LPTMC) CNRS UMR 7600, Université Pierre et Marie Curie-Paris 6, 4 place Jussieu, 75252 Paris cedex 5 France*

(2) *Institut de Génétique Humaine (IGH), CNRS UPR 1142, 141 rue de la Cardonille, 34396 Montpellier cedex 5*

### Résumé

Une question centrale et encore ouverte de la biologie est de comprendre les mécanismes qui contrôlent l'expression des gènes, dans des situations aussi bien normales que pathologiques. Nous défendons ici l'idée que ces mécanismes impliquent de façon essentielle les propriétés physiques de l'ADN, du chromosome, et du niveau d'organisation intermédiaire qu'est la fibre de chromatine. Ces propriétés imposent de fortes contraintes géométriques, électrostatiques ou mécaniques qui doivent être contournées, ou au contraire utilisées, suivant les cas, par des facteurs biologiques spécifiques. Tant la mise en œuvre de cette idée que sa validation expérimentale passent par la modélisation du génome, de son organisation et de sa dynamique. Cette modélisation doit être multi-échelles et rendre compte des relations réciproques entre les caractéristiques des éléments (ici les bases de l'ADN, ou bien les sites génomiques, ou bien les boucles de chromatine suivant l'échelle envisagée) et les propriétés de leur assemblage. Elle doit aussi s'appuyer sur notre connaissance des fonctions biologiques en jeu, pour ne retenir que les ingrédients essentiels. Des prédictions quantitatives, pouvant être confrontées aux observations disponibles *in vitro* et *in vivo*, sont obtenues par des outils de simulation innovants, inspirés des jeux vidéo et adaptés au contexte biophysique. Cette approche ouvre une nouvelle voie d'étude de la régulation génétique et épigénétique des organismes vivants.

### Abstract:

A central and yet still open question in biology is to understand mechanisms that control gene expression in either normal or pathological situations. We here argue that these mechanisms involve in an essential way the physical properties of the DNA, the chromosomes, and the intermediary level of organization achieved by the chromatin fiber. These properties reflect in strong geometrical, electrostatic and mechanical constraints, that have to be circumvented or harnessed by the --- most often specific --- biological factors. Both the implementation and the validation of this idea rely on the modeling of the genome, its organization and its dynamics. This modeling has to be multi-scale and to account for the reciprocal couplings between the elements characteristics (elements being here either the DNA bases, or genomic loci, or chromatin loops according to the considered scale) and the features of their assembly. It has also to rely on our knowledge of biological functions, so as to adequately prune the details and keep only the essential ingredients. Quantitative predictions, that could be confronted to available *in vitro* or *in vivo* observations, are obtained using novel simulation, "physical engines", inspired of video games and adapted to the biophysical context. This approach opens a new direction for investigating genetic and epigenetic regulation of living organisms.

**Notice biographique:** Jean-Marc Victor et Annick Lesne sont directeurs de recherche CNRS au Laboratoire de Physique Théorique de la Matière Condensée, à Paris, où ils ont fondé en 2000 l'équipe «Modélisation multi-échelle de la matière vivante». Leurs travaux portent essentiellement sur la compréhension du rôle biologique des propriétés physiques du chromosome et de son organisation. Pascal Carrivain a fait sa thèse dans cette équipe et il est maintenant post-doctorant à l'Institut de Génétique Humaine à Montpellier.

Contact: [carrivain, victor, lesne] @lptmc.jussieu.fr

### Le contexte biologique: ADN, chromatine et processus génomiques

Nous savons depuis plus d'un siècle que le support de l'hérédité est l'ensemble des 46 molécules d'*ADN* (Acide DésoxyriboNucléique) constituant chacune l'un de nos chromosomes. Il y a un peu plus d'un demi-siècle, Watson et Crick ont établi que la structure de cette molécule est une double hélice, constituée de deux brins portant chacun une séquence de groupements moléculaires A, T, G, C (Adénine, Thymine, Guanine, Cytosine), appelés des *bases* (voir une analyse de l'article princeps de Watson et Crick à la page: <http://www.bibnum.education.fr/sciencesdelavie/biologie/une-structure-de-l%E2%80%99acide-d%C3%A9soxyribonucl%C3%A9ique-adn>). Ces bases sont appariées de façon complémentaire suivant la règle [A-T, G-C], de sorte que les deux brins sont en miroir l'un de l'autre. De façon imagée, la molécule se présente comme une échelle dont les montants seraient enroulées l'un autour de l'autre en hélice. La séquence des bases est appelée le *génom*e, et il constitue le patrimoine génétique de l'individu, présent à l'identique dans chacune de ses cellules, depuis l'œuf primordial. Les *gènes* sont des séquences contenues dans le génome, fournissant à la cellule la «formule» pour fabriquer toutes les protéines dont elle a besoin. Les rôles des différentes protéines sont variés: certaines constituent les briques de construction de l'organisme, d'autres sont des enzymes (catalyseurs), des moteurs moléculaires, ou encore des agents régulateurs sur lesquels reposent à la fois la structure de la cellule et son fonctionnement. Un gène, inscrit dans un des brins, est tout d'abord transcrit à l'intérieur du noyau en un ARN (Acide RiboNucléique) complémentaire suivant la règle [A-U, T-A, G-C, C-G] (la thymine est remplacée par l'uracyle dans les ARN). Après avoir traversé la membrane nucléaire<sup>1</sup> sous forme d'ARN messager

([http://fr.wikipedia.org/wiki/Acide\\_ribonucl%C3%A9ique\\_messager](http://fr.wikipedia.org/wiki/Acide_ribonucl%C3%A9ique_messager)) pour rejoindre le cytoplasme, il est ensuite transcrit dans une machinerie complexe appelée le ribosome (<http://fr.wikipedia.org/wiki/Ribosome>). Cette transcription obéit au *code génétique*, où chaque triplet de lettres correspond à un acide aminé. La séquence génétique est ainsi le patron d'une séquence d'acides aminés, qui une fois repliée de façon appropriée dans l'espace, constitue une protéine fonctionnelle. Mais les gènes ne couvrent pas l'ensemble du génome. D'autres séquences existent, beaucoup plus nombreuses, dites *non-codantes* car elles ne sont pas traduites en protéines. Certaines sont des sites de liaison de facteurs biochimiques spécifiques ou des complexes protéiques formant la machinerie assurant la réplication ([http://fr.wikipedia.org/wiki/R%C3%A9plication\\_de\\_l%27ADN](http://fr.wikipedia.org/wiki/R%C3%A9plication_de_l%27ADN)) ou la transcription. Les autres séquences vont également nous intéresser, car il semble maintenant clair qu'elles jouent un rôle dans l'architecture fonctionnelle du chromosome et de ce fait dans la régulation de tous les processus génomiques, comme nous allons l'expliquer. A chaque division, le génome est répliqué à l'identique pour que chaque cellule fille en contienne une copie.

Chez tous les organismes dont la cellule contient un noyau, le génome présente un remarquable niveau d'organisation intermédiaire entre l'ADN et le chromosome: la *chromatine* (<http://fr.wikipedia.org/wiki/Chromatine>). Il s'agit d'un complexe d'ADN et de protéines, les histones. Les unités de base de la chromatine, appelées les *nucléosomes*, sont formées chacune d'un cœur de huit histones autour duquel s'enroulent 146 paires de base de la molécule d'ADN. Le collier de perles ainsi formé s'organise en une super-structure complexe, appelée la fibre de chromatine. Cette fibre constitue le fil élémentaire de l'écheveau que sont nos chromosomes (Figure 1).

2-1

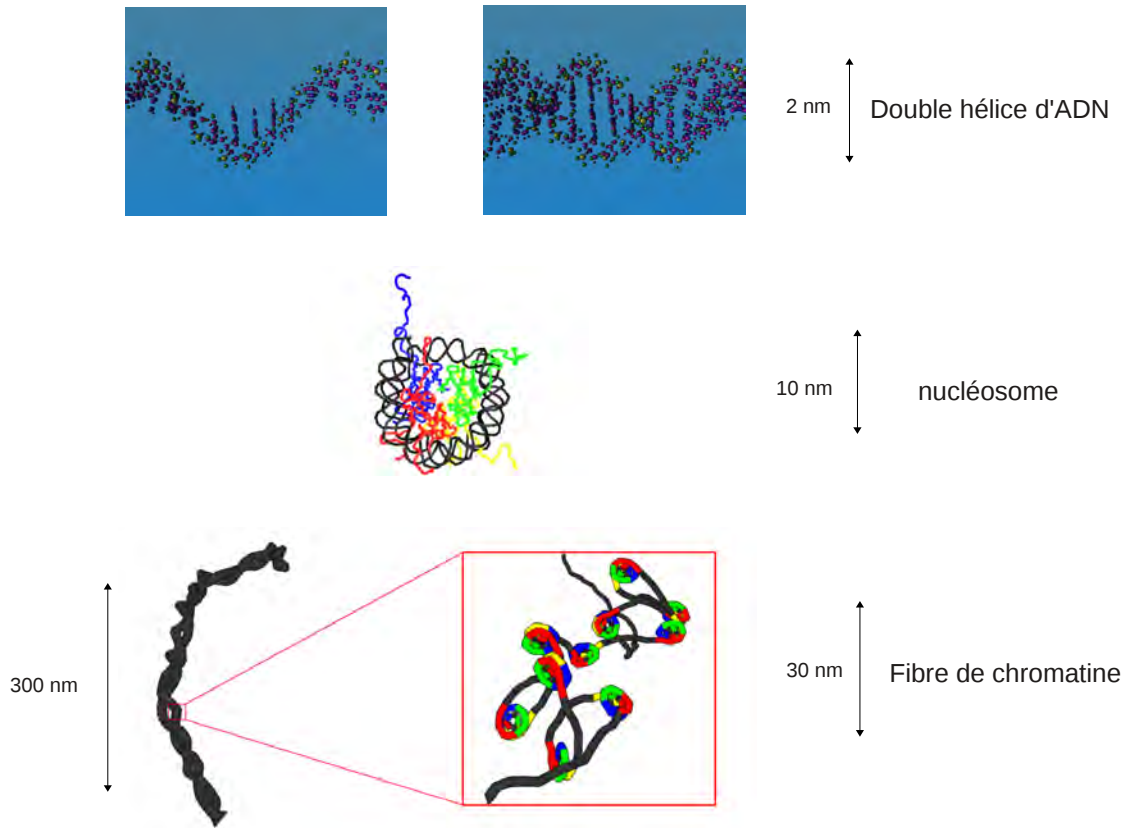


Figure 1 : structure multi-échelles des chromosomes. L'ADN (en haut à gauche : simple brin, à droite : double brin) est ici figuré dans le détail de sa structure atomique. A l'échelle supérieure, l'ADN (en noir) est enroulé autour d'un coeur de protéines (en couleurs) pour former le nucléosome. Puis les nucléosomes s'enchaînent dans la « fibre de chromatine » (en bs à droite) qui est elle-même repliée dans des superstructures telles que des plectonèmes (en bas à gauche). L'architecture globale des chromosomes n'est pas encore connue.

Ce bref résumé des fonctions biologiques du génome souligne un point important et parfois négligé: le génome doit être décrit conjointement du point de vue de l'information génétique qu'il contient, du point de vue des liaisons chimiques entre les atomes de la molécule d'ADN et des interactions biochimiques qu'elle est susceptible d'établir avec divers facteurs, et d'un point de vue physique, ce que nous allons présenter dans ce chapitre. En effet, l'ADN est un *poly-électrolyte*, c'est-à-dire un polymère chargé, ici négativement (deux charges négatives par paire de bases). Sont associées des propriétés électrostatiques: par exemple, l'ADN va se trouver entouré d'une gaine d'ions positifs neutralisant sa charge négative. La nature polymérique de l'ADN affecte également ses propriétés mécaniques: l'ADN se comporte comme un filament flexible ayant une élasticité de courbure et une élasticité de torsion. Nous allons discuter dans ce chapitre toutes les conséquences biologiques de ces propriétés.

### La place de la physique en biologie

Quelle peut être la place de la physique et des physiciens en biologie, au-delà des techniques de manipulation, de mesure et d'imagerie? La réponse va venir d'une reformulation de cette question: quelle est la place des mécanismes physiques dans le fonctionnement d'un organisme? La question que nous allons aborder est de comprendre le rôle des propriétés physiques de l'ADN et de la fibre de chromatine dans le fonctionnement biologique du chromosome, et plus particulièrement dans la régulation de l'expression des

3-2

3-3

3-5

gènes. Plus généralement, même si les interactions biochimiques (qu'elles soient spécifiques ou non) jouent un rôle établi dans l'organisation des chromosomes, il est apparu récemment que les propriétés physiques des chromosomes, en particulier leur nature polymérique, sont d'une importance tout aussi cruciale.

La première étape est de déterminer les contraintes que les lois de la physique imposent aux structures et aux processus biologiques. Il peut s'agir de contraintes électrostatiques, mécaniques, ou des contraintes plus subtiles de la mécanique statistique, prenant en compte l'agitation thermique qui règne de façon importante sinon dominante à l'intérieur de la cellule. On sait depuis plus d'un siècle qu'un grain de pollen en suspension dans l'eau est animé d'un mouvement erratique, le mouvement brownien (du nom du botaniste Robert Brown qui l'a observé pour la première fois en 1827, [http://fr.wikipedia.org/wiki/Mouvement\\_brownien](http://fr.wikipedia.org/wiki/Mouvement_brownien)). Ce mouvement est dû aux innombrables collisions du grain de pollen avec les molécules d'eau, elles-mêmes en mouvement aléatoire. Aux échelles sub-microniques, tous les degrés de liberté vont avoir une composante aléatoire, soit directement par agitation thermique comme les molécules d'eau (par exemple les vibrations atomiques), soit indirectement comme le grain de pollen (par exemple la diffusion d'une protéine à l'intérieur de la cellule). Ces mouvements aléatoires sont d'autant plus importants que la température est élevée, et ils interviennent aussi dans toutes les réactions chimiques (barrières d'activation à franchir) et les liaisons entre les molécules (affinités).

L'étude physique des systèmes biologiques permet ainsi de déterminer les structures et les dynamiques possibles. Dans un second temps, une approche intégrative du système biologique étudié, articulant les comportements possibles du point de vue physico-chimiques et les faits biologiques, s'avère nécessaire. Il va s'agir également de comprendre comment un événement biologique (par exemple la liaison d'un facteur de transcription sur l'ADN, l'assemblage d'un complexe spécifique, une réaction métabolique) a des conséquences physiques, souvent à des portées très supra-moléculaires, puis de comprendre comment ces conséquences physiques servent à leur tour à initier et coordonner de nouveaux événements biologiques. Au bout du compte, on cherche à obtenir une compréhension complète des mécanismes qui contrôlent l'expression des gènes autant au cours du fonctionnement cellulaire normal, par exemple au cours du développement de l'embryon, que dans des situations pathologiques, par exemple dans des cellules cancéreuses.

#### Propriétés physiques, fonctions biologiques

La démarche que nous venons de décrire repose sur le fait, autrefois nié par les idées de génération spontanée et d'élan vital, mais maintenant consensuel, que rien dans le vivant n'échappe aux lois de la physique. Mais alors qu'un système physique possède des *propriétés*, il faut souligner la présence dans le vivant de *fonctions* et de mécanismes de *régulation* de ces fonctions. Par exemple, notre organisme est capable de digérer des aliments, de maintenir sa température interne à 37°, d'entretenir un rythme cardiaque robuste et régulier, ou encore de réparer une coupure de la peau. Il est capable de persister à peu près identique à lui-même pendant de nombreuses années alors même que les cellules qui le constituent sont renouvelées. Il apparaît en général que les mécanismes physiques sous-jacents sont ou bien contournés ou bien asservis à la fonction biologique: leurs ingrédients et leurs paramètres sont ajustés par la sélection naturelle et la co-évolution qui s'ensuit, de façon à remplir correctement des fonctions biologiques. Ces fonctions ne sont pas forcément réalisées de façon simple et encore moins unique. La diversité des espèces est la trace de la diversité des solutions adoptées par le monde vivant de façon contingente, suivant différents compromis entre précision et coût, entre spécialisation et généralisme, entre efficacité et robustesse, pour n'en citer que quelques uns. Cette formulation, couramment utilisée en biologie, ne doit pas leurrer par son caractère apparemment finaliste. En effet, elle reflète implicitement l'évolution à l'œuvre dans le vivant, au sens darwinien: la finalité n'est pas donnée a priori comme dans un objet manufacturé, mais elle émerge a posteriori quand nous observons aujourd'hui le résultat de la sélection naturelle.

Etudier les liens entre d'une part l'architecture d'un système biologique et ses propriétés physiques et d'autre part son organisation fonctionnelle revisite la question de la relation «structure-fonction» en y incluant davantage de mécanique, d'élasticité, de physique des polymères et surtout de dynamique, toutes notions d'ordinaire peu utilisées. Cela demande des outils théoriques nouveaux, en particulier des outils de simulation, tels ceux que nous allons présenter.

### L'exemple des propriétés mécaniques de l'ADN et de la chromatine

A une échelle suffisamment grande, supérieure à celle des paires de bases, la molécule d'ADN peut être décrite comme un polymère - plus exactement un polyélectrolyte car elle est électriquement chargée (négativement) - possédant un module élastique de flexion ainsi que de torsion. Ces modules dépendent de la température et du pH, contrôlé *in vivo* par la nature et la concentration des ions présents au voisinage de l'ADN. Cette description peut être validée expérimentalement par des micromanipulations dans lesquelles on exerce une traction ou une torsion sur une molécule d'ADN unique, attachée sur une surface et greffée à son autre extrémité à une bille permettant la micromanipulation grâce à des pinces optiques<sup>2</sup> ou magnétiques (voir encadré).

La fibre de chromatine est elle-même une structure de flexibilité variable. Elle apparaît se comporter comme un ressort supra-moléculaire dans lequel l'ADN est enroulé autour des bobines d'histones le long de la fibre de chromatine comme un fil de métal dans un ressort hélicoïdal. Les propriétés élastiques de la fibre de chromatine ont également été mises en évidence expérimentalement grâce aux pinces magnétiques<sup>3</sup> et ont été analysées grâce aux outils de modélisation que nous présentons plus loin (Figure 2).

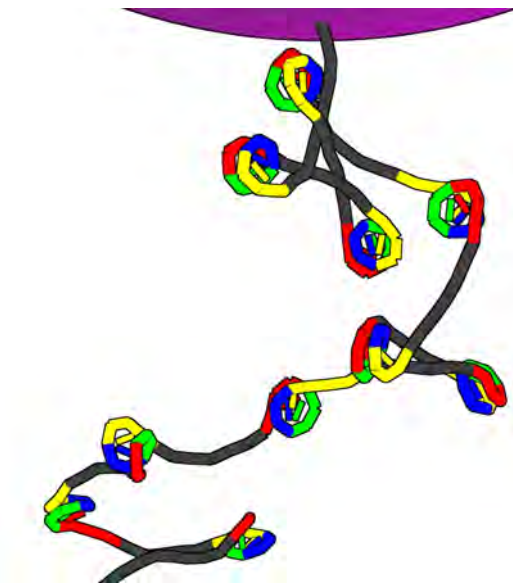
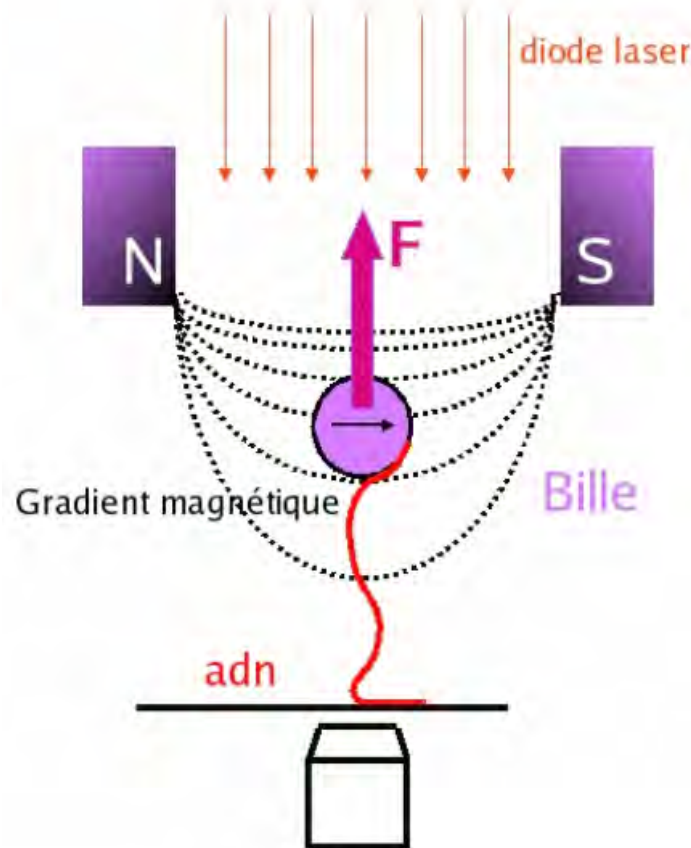


Figure 2 : Fibre de chromatine soumise à un couple de torsion dans une expérience de pinces magnétiques. Image obtenue par simulation numérique à l'aide du moteur physique ODE (<http://www.ode.org>). Voir aussi la vidéo à l'adresse : [vimeo.com/m/52072568](https://vimeo.com/m/52072568).

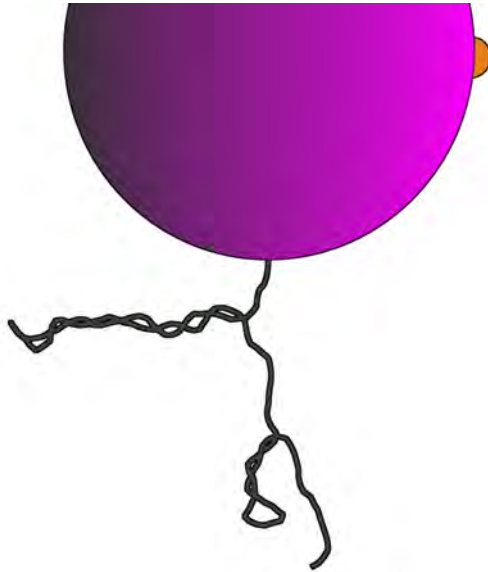
### Encadré : Pincettes magnétiques et nombre d'enlacement.

1) Dans une expérience de **pincettes magnétiques**, une bille aimantée est placée dans l'entrefer d'un électro-aimant ou d'une paire d'aimants permanents; elle est alors soumise à la fois à une force de traction due au gradient de champ magnétique existant dans l'entrefer et à un couple de torsion lorsque les aimants sont mis en rotation, provoquant à la fois l'étirement de la molécule d'ADN (attachée à la fois en haut sur la bille et en bas sur une lame de verre) et son surenroulement :



2) **nombre d'enlacement** : lorsque l'ADN est circulaire ou que ses extrémités sont fixées, typiquement parce qu'elles sont liées à des nucléosomes dont les positions sont gelées au sein d'un complexe protéique, le nombre de tours emmagasinés dans l'ADN est une quantité conservée appelée le *nombre d'enlacement*  $L_k$ . C'est aussi le nombre de fois qu'un des brins de l'ADN croise l'autre, en projection plane. On parle d'invariant topologique (quantité invariante par toute déformation continue d'un ruban ou d'un tube sans aucune coupure). On peut montrer que  $L_k$  est la somme de la torsion  $T_w$  (*twist* en anglais) subie par la molécule autour de son axe et d'une torsion  $W_r$  dans l'espace tri-dimensionnel, appelée en anglais *writhe*. Cette relation mathématique très simple  $L_k = T_w + W_r$  a une conséquence remarquable qu'on peut facilement observer sur une cordelette ou en manipulant le fil d'un ancien téléphone. Si on tord ce fil sur lui-même en imposant  $L_k$  tours et qu'on pince ensuite ses extrémités entre les doigts, on va observer des changements de la forme tri-dimensionnelle de ce fil si on change l'écartement des mains. Si les mains sont éloignées, le fil est rectiligne et les tours imposés au fil vont s'exprimer sous forme de torsion:  $T_w = L_k$ . Si au contraire on rapproche les mains, un tortillon appelé un plectonème va se former, voire plusieurs si on a imposé beaucoup de tours ( $L_k$  grand), et le surenroulement va maintenant s'exprimer sous forme de torsion tri-

dimensionnelle:  $W_r=L_k$ . Tous les intermédiaires entre ces deux extrêmes sont observés. En jouant avec le fil, on se convainc rapidement qu'il y a là un phénomène de «vases communicants» et que la contrainte imposée en tordant le fil sur lui-même est conservée: seule sa façon d'apparaître dans la forme tridimensionnelle du fil varie.



Formation de deux plectonèmes dans une expérience de pinces magnétiques. La bille (en magenta) a effectué 15 tours sur elle-même, entraînant la formation de deux plectonèmes dans la molécule d'ADN. Voir aussi la video à l'adresse : [vimeo.com/m/51918378](https://vimeo.com/m/51918378).



### A quoi servent ici la simulation et la modélisation

Il arrive que nous sachions rendre compte de certains phénomènes physiques ou biologiques par une ou plusieurs équations: la diffusion de l'encre dans l'eau, certaines étapes du métabolisme, la réponse à une traction de la molécule d'ADN. Cette étape déjà ardue n'est que la première à franchir avant d'obtenir un résultat; il faut encore résoudre ces équations. Dans la plupart des cas il n'est pas possible de trouver des solutions analytiques exactes et on ne peut que formuler des approximations de la véritable solution. Récemment divers algorithmes de résolution numérique d'équations ont vu le jour et permettent de tracer une solution graphique en fonction des paramètres du problème. Nous pouvons citer l'exemple des problèmes des patterns de Turing qui interviennent en morphogenèse ou dans les problèmes de réaction-diffusion en chimie. Ces deux problèmes sont bien définis en termes d'équations aux dérivées partielles (EDP, cf [http://fr.wikipedia.org/wiki/%C3%89quation\\_aux\\_d%C3%A9riv%C3%A9es\\_partielles](http://fr.wikipedia.org/wiki/%C3%89quation_aux_d%C3%A9riv%C3%A9es_partielles)) mais la résolution fait appel aux méthodes numériques. Il est fréquent que cette approche, relevant des mathématiques appliquées, ne soit pas suffisante.

8-1

C'est le cas pour un problème aussi complexe que l'organisation des chromosomes au sein des noyaux des cellules des organismes vivants. Si l'on part pour le chromosome d'équations décrivant la molécule d'ADN qui le constitue, non seulement le nombre d'équations à résoudre est colossal mais ces équations font intervenir trop de variables et de paramètres pour pouvoir prédire de façon robuste le comportement. De plus, ces équations, écrites dans un formalisme bien défini (équations aux dérivées partielles pour des concentrations locales) mais de validité limitée, ne peuvent prendre en compte tous les mécanismes sous-jacents, ni toutes les contraintes que le micro-environnement exerce sur le chromosome aux différentes échelles. Une solution courante est de faire intervenir des termes aléatoires, mais ils compliquent encore plus le problème et ne peuvent rendre compte de la coordination entre les différents facteurs mis en jeu, ni de l'organisation hiérarchique du chromosome et de sa dynamique.

8-2

La modélisation n'a donc pas toujours comme objectif de capturer le phénomène dans une ou plusieurs équations. Les modèles de réaction-diffusion sont en ce sens exemplaires, par leur portée limitée puisqu'ils ne décrivent que ce que peut produire une EDP, donc en particulier pas les effets d'aggrégation aléatoire ou les effets de mémoire produisant des événements rares mais cruciaux pour la biologie. La modélisation des chromosomes rend précisément illusoire de capturer dans une équation, dans un seul formalisme et à une seule échelle la richesse et la variété des mécanismes et contraintes en jeu, prenant place à tous les niveaux d'organisation. Il va s'agir au contraire d'articuler des modèles partiels, un peu comme il y a des sous-routines, des procédures ou des fonctions dans un programme. On va décrire comment on passe d'une description tous-atomes de l'ADN à une description effective en termes d'empilement de paires de base puis comme un filament élastique, comment on passe d'un modèle de nucléosome au niveau des résidus à un solide articulé, comment on passe de leur assemblage à une fibre élastique etc. Un autre intérêt majeur que présente ce lien ---via la modélisation--- entre les différents niveaux, est qu'il permet indirectement d'assembler ou de confronter des résultats expérimentaux obtenus par des techniques différentes à des échelles différentes.

8-4

Mais il faut aller plus loin que cette démarche qui procède des niveaux microscopiques vers les niveaux macroscopiques. La modélisation multi-échelles va procéder en articulant deux directions opposées: comment les comportements collectifs mésoscopiques ou macroscopiques émergent des éléments en interaction et peuvent être décrits par quelques paramètres effectifs, et comment les structures et processus à grande échelles imposent des contraintes sur les éléments, leurs propriétés et leurs interactions. En pratique la modélisation multi-échelles procède en articulant les modèles obtenus à chaque échelle dans le formalisme adapté, pour intégrer les données obtenues aux différents niveaux d'organisation du chromosome.

La simulation qui intervient pour implémenter, explicitement, certaines étapes de ce schéma multi-échelle, permet aussi de faire de la phénoménologie. C'est en ce sens qu'on parle d'expérience numérique. Le fait que les simulations commencent à se substituer aux expériences qu'on ne peut pas faire est dû au fait qu'on peut simuler des expériences dans les mêmes conditions et en particulier à la même échelle (avec le même nombre de particules), mais en contrôlant indépendamment chaque paramètre et en pouvant enregistrer les variations de n'importe quelle grandeur observable, aussi bien locale que globale (par exemple aussi bien la position d'une particule ou leur densité moyenne). Les simulations de milieux granulaires en sont un



9-1  
excellent exemple. Et la simulation permet ensuite de construire un modèle (modèle explicatif), comme des expériences permettent d'échafauder une théorie. On peut tester les hypothèses du modèle en faisant tourner des simulations et non en résolvant des équations. Cette vision phénoménologique de la simulation s'applique bien sûr aux chromosomes.

Les modèles *explicatifs* dont il est question dans ce texte, conçus pour dégager des mécanismes dominants et faire la preuve de notre bonne compréhension qualitative, sont fondamentalement différents des modèles *prédictifs* utilisés par exemple en ingénierie ou pour guider la prise de décision. Ces derniers sont des modèles empiriques, à grand nombre de paramètres, et statistiques au sens où ils ne reposent en général pas sur la dynamique réelle du système, mais plutôt sur un modèle statistique (par exemple une superposition d'oscillations, ou ce qu'on appelle des modèles auto-régressifs qui expriment la variable au temps présent en fonction de ses valeurs passées) qui reproduit à court ou moyen terme le comportement réel. La construction d'un modèle explicatif est cruciale : le physicien Eugene Wigner aimait à dire : « L'ordinateur a compris, mais moi aussi j'aimerais comprendre ! ».

### Une modélisation multi-échelle et fonctionnelle

9-2  
Différents niveaux de modélisation de l'ADN ou de la fibre de chromatine sont possibles. La description la plus naturelle peut se faire à l'aide d'un modèle où nous incluons la totalité des atomes. La simulation du nombre colossal de ces unités élémentaires peut occuper un grand nombre de nos plus puissants ordinateurs pour un temps assez long. Tout cela est sans compter avec la complexité des champs de forces ([https://fr.wikipedia.org/wiki/Champ\\_de\\_force\\_%28chimie%29](https://fr.wikipedia.org/wiki/Champ_de_force_%28chimie%29)) qui régissent le monde à l'échelle atomique et que nous devons paramétrer finement pour ne pas obtenir des résultats aberrants à la fin de la simulation. La validité des résultats n'étant pas assurée du fait de ce manque de robustesse de la simulation, il est intéressant de construire des modèles « gros-grain » où l'utilisateur maîtrise plus facilement la modélisation. Il s'agit en réalité de contrôler les résultats aux différentes échelles impliquées, et d'intégrer les connaissances, en particulier expérimentales, disponibles à ces différentes échelles. Nous cherchons donc à articuler des modèles de physique des polymères ([https://fr.wikipedia.org/wiki/Polym%C3%A8re\\_en\\_solution](https://fr.wikipedia.org/wiki/Polym%C3%A8re_en_solution)) pour décrire les mouvements aléatoires de la fibre, des modèles issus de la théorie de l'élasticité pour apprécier les contraintes mécaniques ressenties tant au niveau de la fibre que de l'ADN, et des modèles atomiques pour comprendre comment sont contrôlées les affinités entre l'ADN et les protéines qui s'y lient. Un enjeu est par exemple de comprendre comment les déformations de l'ADN se propagent le long de l'ADN lorsque celui-ci est inclus dans une fibre de chromatine, et de ce fait est mécaniquement et topologiquement contraint. Cette approche multi-échelles est fonctionnelle, au sens où elle guidée par notre connaissance de la fonction biologique effectivement réalisée.

9-3  
9-5  
Le travail du physicien est par exemple de comprendre quels facteurs peuvent modifier la conformation de la fibre et ses propriétés mécaniques. Ces modifications vont d'une part modifier les mouvements et les fluctuations des boucles de chromatine, mais aussi, en cascade, modifier les contraintes mécaniques ressenties par l'ADN. On voit ici apparaître une « boucle causale » traversant les niveaux d'organisation<sup>4</sup>. Ses médiateurs sont à la fois les facteurs biochimiques et les propriétés mécaniques et électrostatiques : les déformations induites par la liaison de protéines ou les modifications de certains groupements chimiques prenant place au niveau de l'ADN ou des histones vont affecter la structure de la fibre; en retour les modifications de la fibre vont changer le contexte conditionnant toutes les interactions prenant place au niveau de l'ADN ou des nucléosomes (Figure 3).

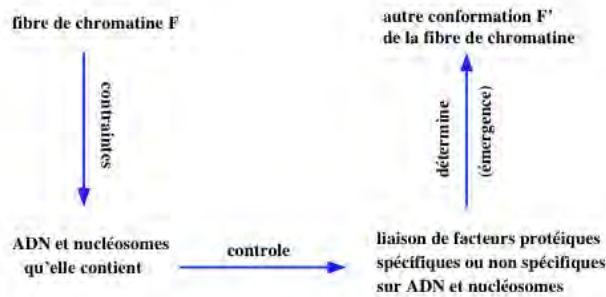


Figure 3 : « boule causale » de la fibre de chromatine. Les contraintes mécaniques et topologiques qui s'exercent sur la fibre contrôlent la liaison de facteurs protéiques spécifiques ou non spécifiques sur l'ADN ou sur les histones. Une fois liés, ces facteurs modifient la structure de la fibre de chromatine et ses propriétés mécaniques et topologiques, qui à leur tour vont contrôler la liaison de nouveaux facteurs protéiques, et ainsi de suite.

### Des outils de simulation dédiés, inspirés des jeux vidéo

Le premier essai pour construire un modèle « gros-grain » de l'ADN, rendant compte de ses propriétés élastiques (voir plus haut), fut de remplacer un certain nombre de paires de bases par un ressort qui mime les effets entropiques<sup>5</sup>. Une molécule d'ADN est alors représentée par une chaîne de ressorts plongée dans un bain thermique et soumise à des contraintes extérieures, et dont nous savons simuler la dynamique. Le bain thermique est une technique de simulation permettant de reproduire l'effet de la température sur la molécule d'ADN. Mais là encore nous devons faire attention au choix de la raideur de nos ressorts. Pour des ressorts trop mous les paires de bases subissent une élongation bien trop importante et pour des ressorts trop raides ce sont les forces de rappel qui deviennent trop importantes et posent des problèmes numériques. (Un dernier problème, plus technique, de ce type de simulation est que l'ergodicité n'est en général pas respectée : le système reste piégé dans un sous-ensemble de l'espace des configurations possibles ; de plus la thermalisation du système n'est pas réalisée : différentes parties du système se comportent comme si elles étaient à des températures différentes, sans équilibration entre elles.)

Nous serions tentés d'utiliser un ressort de raideur infinie mais d'autres problèmes numériques nous en empêchent. Les physiciens ont alors tenté une nouvelle approche basée sur les contraintes dites « holonomes » ([https://fr.wikipedia.org/wiki/Contrainte\\_holonyme](https://fr.wikipedia.org/wiki/Contrainte_holonyme)), c'est-à-dire réduisant le nombre de degrés de liberté, en imposant par exemple des contraintes sur les distances entre les atomes liés entre eux. L'idée est de construire des liaisons rigides n'impliquant pas de forces de magnitudes infinies. Ce nouveau type d'algorithme, dont le prototype historique s'appelle « SHAKE », permet de simuler une chaîne de billes articulées, dans laquelle chaque paire de billes consécutives est toujours séparée par la même distance au cours de la simulation. Il nous manque encore un ingrédient. L'espace entre deux billes consécutives est vide et les billes sont considérées comme ponctuelles. Pour pallier ce problème nous sommes allés chercher les algorithmes de corps rigides sous contraintes développés par les mécaniciens et les roboticiens pour les jeux vidéo. Ils régissent comment les personnages se déplacent dans les trois dimensions de l'espace, avec l'objectif de n'obtenir que des mouvements réalistes, non seulement en apparence, mais respectant précisément les lois de la physique. Nous avons adapté ces « moteurs de jeu » pour les utiliser avec un bain thermique, c'est-à-dire pour des objets microscopiques où les forces aléatoires générées par l'agitation thermique des molécules environnantes jouent un rôle important ([http://tel.archives-ouvertes.fr/docs/00/80/22/04/PDF/these\\_Pascal\\_Carrivain.pdf](http://tel.archives-ouvertes.fr/docs/00/80/22/04/PDF/these_Pascal_Carrivain.pdf)) et <http://www.cnrs.fr/fr/pdf/jdc/272/index.html#/12/>). Rappelons que l'algorithme « SHAKE » ne permet que de mimer la longueur d'un certain nombre de paires de bases et non le rayon de la double hélice d'ADN. Avec un moteur de jeu nous pouvons utiliser un système de cylindres articulés par des joints et le rayon des

10-1

10-3

10-4

10-5

11-1  
cylindres correspond au rayon cristallographique de l'ADN (ou plus généralement à un rayon effectif tenant compte des interactions électrostatiques). Notre but est de comprendre les transitions conformationnelles de l'ADN (par exemple la formation de plectonèmes), ce qui contrôle le régime de fluctuations conformationnelles de la fibre de chromatine, et quelles sont les parts respectives des mécanismes purement physiques et des facteurs biologiques spécifiques dans la dynamique fonctionnelle du chromosome.

11-2  
11-4  
Parmi les résultats récents, citons l'élucidation de l'organisation fonctionnelle des chromosomes de la levure, pour lesquels on commence à disposer d'observations in vivo, à la fois structurales<sup>6</sup> et dynamiques<sup>7</sup>. La modélisation physique montre que l'essentiel du fonctionnement peut être assuré par une architecture en brosses de polymères<sup>8</sup>. L'encombrement stérique des chromosomes, tous attachés (par leur centromère) à une même structure appelée SPB (de l'anglais « Spindle Pole Body »), conduit en effet à une organisation en brosse, dans laquelle les chromosomes sont tous grossièrement parallèles, de telle sorte qu'ils ne s'emmêlent jamais. Chez les eucaryotes supérieurs, par exemple chez la drosophile et plus encore chez l'homme, la taille du génome n'est plus compatible avec une simple structure en brosse. L'élucidation de l'organisation nucléaire des eucaryotes supérieurs constitue aujourd'hui l'un des défis majeurs de la biologie. Les approches de physique des polymères, dont la pertinence a été démontrée dans le cas de la levure, ne manqueront pas d'y figurer en bonne place.

### Le point de vue des biologistes

Plus généralement, à quoi sert la modélisation physique dans un contexte biologique? D'après notre collaborateur Thierry Forné (IGMM, Montpellier), les modèles doivent être bâtis à partir des données, et non abstraitement par le théoricien. Ce faisant, non seulement la validité du modèle se trouve augmentée, mais surtout, le modèle va pouvoir être utilisé pour donner du sens aux données, en les replaçant dans un schéma mécaniste explicatif (modèle explicatif). En effet, les données in vivo sont souvent obtenues par des techniques indirectes, sans contrôle individuel des différents paramètres, et un modèle est alors nécessaire pour identifier les différents phénomènes en jeu et la part que chacun joue dans ce qui a été observé, autrement dit pour réellement comprendre (au sens de mettre ensemble, articuler) ce qu'on a observé, qualitativement mais aussi quantitativement. Plus encore, les approches multi-échelles, où il existe des méthodes (homogénéisation, réduction dimensionnelle) pour relier les formalismes et descriptions à différentes échelles, vont indirectement permettre de relier les différentes échelles d'observation.

11-5  
L'enjeu va même beaucoup plus loin que l'utilisation d'outils physiques pour comprendre un système biologique: il s'agit d'une part de comprendre les propriétés physiques non spécifiques, qui régissent les structures et dynamiques globales. Il s'agit d'autre part d'identifier les facteurs locaux spécifiques qui créent la fonction. Mais les deux facettes sont indissociables: la physique ne peut expliquer la fonction biologique, et la connaissance des facteurs biologiques est incomplète sans celle des contraintes et mécanismes physiques qui contrôlent et coordonnent leur action.

On parle souvent de modélisation mathématique. A notre avis il y a là une confusion. Les mathématiques sont extrêmement puissantes pour analyser les propriétés d'un modèle. Mais la construction du modèle, c'est-à-dire l'étape d'abstraction, de passage d'une réalité complexe à une formulation mathématique, échappe aux mathématiques, au sens où elle ne peut être fondée sur des arguments mathématiques. Elle relève davantage de la physique, dont la démarche depuis toujours est de capturer dans des représentations simples des phénomènes naturels. Et elle s'accompagne de précautions qui elles aussi ne relèvent pas de la pratique mathématique: la détermination du domaine de validité d'un modèle, et sa mise en relation avec des modèles plus fins ou plus grossiers.

La modélisation physique est enfin un formidable moyen de communication entre disciplines. Elle propose une nouvelle façon de se représenter les systèmes biologiques et leurs fonctions et par-là même de faciliter le dialogue, non seulement entre biologistes et physiciens, mais aussi entre biologistes eux-mêmes. La modélisation physique élague un grand nombre de détails qui ne sont pertinents que pour une espèce donnée ou une technique donnée d'observation, et fournit un schéma beaucoup plus clair et fédérateur aux différentes communautés scientifiques. Un modèle est d'abord une manière *opérationnelle* de se représenter le système qu'on étudie.

## Questions ouvertes

La modélisation et la simulation multi-échelles du génome que nous avons présentées ici permettent d'étudier les mécanismes physiques de la régulation épigénétique<sup>9</sup> physiologique ou pathologique (notamment dans les cancers où on commence à envisager des thérapies épigénétiques). Une question ouverte est la possibilité d'une composante épigénétique dans l'hérédité. Une autre est de discriminer le rôle des interactions spécifiques et celui des interactions non spécifiques dans la régulation des processus génomiques. Une architecture suppose des interactions spécifiques, qui vont faire le tri parmi des interactions non spécifiques et les rendre ainsi adaptées au fonctionnement biologique (une métaphore serait un immeuble qui structurellement permet ou non l'accès aux personnes handicapées). L'idée que nous défendons est celle d'un contrôle physique global et non spécifique, permettant ou non à des facteurs biologiques spécifiques d'effectuer une action localisée. Ce scénario a déjà été validé chez la levure. Il reste maintenant à l'explorer chez les organismes supérieurs et en particulier chez l'homme.

12-1

12-2

---

<sup>1</sup> Nous ne parlerons ici que des organismes, dits « eucaryotes », dont les cellules contiennent un noyau. Ils incluent tous les organismes pluricellulaires et quelques monocellulaires comme les levures. Ils s'opposent aux bactéries, dites « procaryotes », qui ne contiennent pas de noyau et dont le génome a une organisation et un fonctionnement plus simple.

<sup>2</sup> Dans les pinces optiques, la réfraction d'un faisceau lumineux par une bille transparente produit une force résultante sur la bille en modifiant la quantité de mouvement des photons qui composent le faisceau, de telle sorte que la bille est entraînée vers une position stable à proximité du foyer du faisceau, là où l'intensité lumineuse est maximale ; plus d'infos dans l'article MEDECINE/SCIENCES 2003 ; 19 : 364-7 à la page [http://biologie.univ-mrs.fr/upload/p204/Coirault\\_\\_MedSc03\\_\\_Pinces\\_Optiques\\_en\\_Bio\\_et\\_en\\_Medecine.pdf](http://biologie.univ-mrs.fr/upload/p204/Coirault__MedSc03__Pinces_Optiques_en_Bio_et_en_Medecine.pdf)

<sup>3</sup> Voir par exemple le communiqué de presse du CNRS « quand les chromosomes font leur stretching » à l'adresse <http://www2.cnrs.fr/presse/communiqued/855.htm>

<sup>4</sup> Voir A. Lesne, Acta Biotheoretica 61(1):3-19. doi: 10.1007/s10441-013-9170-z

<sup>5</sup> Concernant l'interaction que ressentent les deux bases d'une paire, on parle d'effet entropique lorsque ce qui favorise une paire de positions, par rapport à une autre, est le nombre de configurations atomiques qui lui correspondent; plus ce nombre est grand, plus la paire sera statistiquement favorisée, ce qu'on reproduit à l'aide d'une force effective équivalente.

<sup>6</sup> Ces observations structurales sont aujourd'hui obtenues par les techniques dites de capture conformationnelle des chromosomes, cf Duan et al 2010, Nature 465, 363–367 doi:10.1038/nature08973

<sup>7</sup> Hajjouli/Mathon et al 2013, Genome Research, sous presse.

<sup>8</sup> Wong et al 2012, Current Biology 22(20):1881-90. doi: 10.1016/j.cub.2012.07.069

<sup>9</sup> On parle d'épigénétique pour les mécanismes liés au génome mais qui impliquent d'autres facteurs que les séquences génétiques, par exemple des modifications chimiques locales telles la méthylation de l'ADN ou l'acétylation des histones.

# Notes

---

2-1 Pas sur que les refs a wikipedia soient necessaires...

3-2 Peut on lier cela à des differences de types de modeles?( c est le theme)

3-3 Une reponse?

3-5 Peut etre: "mecanismes modelises par la physique". Enun sens tous les mecanismes enbiologie sont ontologiquement des mecanismes physiques ( ou en tout cas beaucoup le pensent).

4-1 Idem

4-4 Lien discutable. La generationspontanee a aussi servi unmaterialisme hardcore

4-6 Peut etre trouver une formulation plus neutre. La fonction est parfois concue comme definie par la slection naturelle, alors que cette formulationpresuppose qu elle preexiste a la selection

8-1 Idem.

Eventuellement parler ici d equations de reaction-diffusion?

8-2 Expliquer ou exemplifier

8-4 En termes atomique?

9-1 Peut etre preciser. L idee semble etre que les simulatinons permettent a la fois un modele phenomenologique ( comme on le fait a partir d une experience) puis construire un modele explicatif. Si c est ca ca pourrait etre plus clair. Et aussi: est ce que ce dernier modele est un modele de la simulation? Ou bien du phenomene vise par la simulation? Il y a la une question peutetre delicate (ou peut etre vide, si jai mal compris le texte)

9-2 Est ce que par exemple les modeles qui donnent des scenarios du climat dans 50 and, ou les modeles predisant des distributions d espece, sont ce genre de modeles predictifs? Ils pourraient alors etre donnes en exemple

9-3 Pas sur de comprendre. Le lien " manque de robustesse" et champs de forces atomiques pourrait etre developpè

9-5 Developper eventuellement

10-1 E quoi le defaut d ergodicité est il un probleme?

Et: cette parenthese ne peut elle pas aller en note?

10-3 Meme reserve sur les refs wikipedia

10-4 Reference a l article, auteurs, date, etc

10-5 Note

11-1 Definition?

# Notes

---

11-2 Commeny le definiti on?

11-4 Des figures seraient bienvenues!

11-5 Ppurquoi? Y a t il une reference implicite a la selection naturelle? A l histoire des systemes consideres?

12-1 Peut etre en dire plus, ou sinon enlever cette mention car c est un sujet complexe ettres debattu..

12-2 La notion de specficite pourrait etre definie car elle est importante dans tout le texte.