



**Dynamique du biopolymère ADN  
en suivi de particule unique :  
simulations et expériences  
dans le cadre du mécanisme de transposition**

Laboratoire de Physique Théorique (UMR CNRS/UPS 5152)  
118 route de Narbonne, 31400 Toulouse

Institut de Pharmacologie et de Biologie Structurale (UMR CNRS/UPS 5089)  
205 route de Narbonne, 31077 Toulouse

Julien Baglio

Stage de recherche M2 Physique théorique de Paris  
*sous la direction de Manoel Manghi et Nicolas Destainville*

14 janvier 2008 - 29 février 2008

---

## Remerciements

Je voudrais ici remercier tout d'abord le laboratoire de physique théorique pour m'avoir si bien accueilli en son sein. Je remercie bien évidemment Manoel Manghi et Nicolas Destainville, mes tuteurs, qui m'ont guidé tout au long de ce stage et qui ont répondu avec gentillesse à toutes mes questions. Je remercie aussi Laurence Salomé de l'IPBS pour m'avoir accueilli dans son équipe pour les expériences et bien sûr je remercie Catherine Tardin, de l'IPBS qui m'a guidé dans leurs réalisations et m'a fait découvrir à quoi pouvait bien ressembler une expérience de biophysique.

Je tiens aussi à remercier les gens que j'ai rencontrés par ailleurs au LPT, et qui ont rendu mon séjour à Toulouse le plus agréable possible aussi bien pour leurs discussions de physique qui m'ont appris beaucoup que pour les moments extérieurs au laboratoire : Arnaud Ralko, Andreas Abendschein, Fabien Trouselet, Julie Douady, Clément Touya, Thomas Portet, Ludovic Arnaud, Gaspard Bousquet.

Enfin je remercie mon amie Rayna Stamboliyska pour ses corrections plus qu'utiles sur l'introduction biologique de ce rapport.

## **Abstract**

This report focuses on the first steps of the transposition mechanism within DNA. Roughly, it is the ability of some parts of a DNA chain to migrate within the same chain or another DNA chain. We examine the physical properties and the dynamics of the very first step consisting in the formation of a loop between the two IRs sites of the moving DNA part called a transposon.

We will show by simulations that the TPM technique used in experiments is actually able to visualise this process. Several significant experimental results on DNA without loop are presented, showing the basics of the technique and its awaited improvements.

## **Résumé**

Le but de ce stage est l'étude du mécanisme de transposition dans l'ADN, qui est de manière générale le mécanisme assurant la migration de bouts d'ADN, appelés transposons, d'une partie à une autre d'une même molécule ou sur une autre molécule d'ADN. Nous examinons les propriétés physique et la dynamique de la toute première étape qui consiste en la formation d'une boucle sur l'ADN entre les deux sites d'activation (IRs) du transposon.

Nous montrons via les simulations que la méthode dite TPM utilisée dans les expériences permet bien de suivre ce processus, et nous présenterons des résultats expérimentaux obtenus durant le stage montrant les bases de cette technique ainsi que les améliorations utiles pour la suite des recherches.

# Table des matières

<b>1</b>	<b>Introduction, contexte biologique du stage</b>	<b>1</b>
1.1	La molécule ADN . . . . .	1
1.2	La transposition de l'ADN . . . . .	2
<b>2</b>	<b>Méthode expérimentale</b>	<b>3</b>
2.1	Présentation de la TPM . . . . .	3
2.2	Application à notre étude, protocole utilisé . . . . .	4
2.3	Résultats antérieurs . . . . .	6
<b>3</b>	<b>Modèle théorique</b>	<b>8</b>
3.1	Modèle de ver (worm like chain) . . . . .	8
3.2	Modélisation via les forces . . . . .	9
3.3	Modélisation de l'attraction entre IRs . . . . .	10
<b>4</b>	<b>Résultats des simulations et résultats expérimentaux</b>	<b>11</b>
4.1	Résultats des simulations sur l'ADN de 2080 bp . . . . .	11
4.2	Résultats sur l'ADN de 400 bp . . . . .	14
	<b>Conclusion</b>	<b>19</b>

# 1 Introduction, contexte biologique du stage

## 1.1 La molécule ADN

L'ADN, ou acide désoxyribonucléique, est une molécule présente dans tous les organismes vivants et le support de l'information génétique. C'est aussi le siège de la synthèse protéique. La découverte de sa structure en double hélice date de 1953, suite aux travaux de Rosalind Franklin, James Watson et Francis Crick [9]. Ils montrèrent ce résultat par une étude cristallographique aux rayons X, ce qui fut confirmé par la suite et qui est une structure se retrouvant aussi bien chez les organismes eucaryotes que chez les organismes procaryotes.

L'ADN est une chaîne polymère dont l'unité de base, ou monomère, est constituée d'un sucre - un pentose -, d'une base azotée et d'un acide phosphorique. Les bases constitutives de l'ADN sont au nombre de 4 : l'adénine (A), la thymine (T), la guanine (G) et la cytosine (C). L'ADN est structuré naturellement en deux brins formant une hélice, et l'appariement des paires de bases suit la règle dite de complémentarité : A avec T, C avec G, ceci étant dû au nombre de liaisons hydrogène entre ces paires de bases : 2 pour la liaison A-T, 3 pour la liaison C-G.

La figure 1 est un schéma de cette molécule :

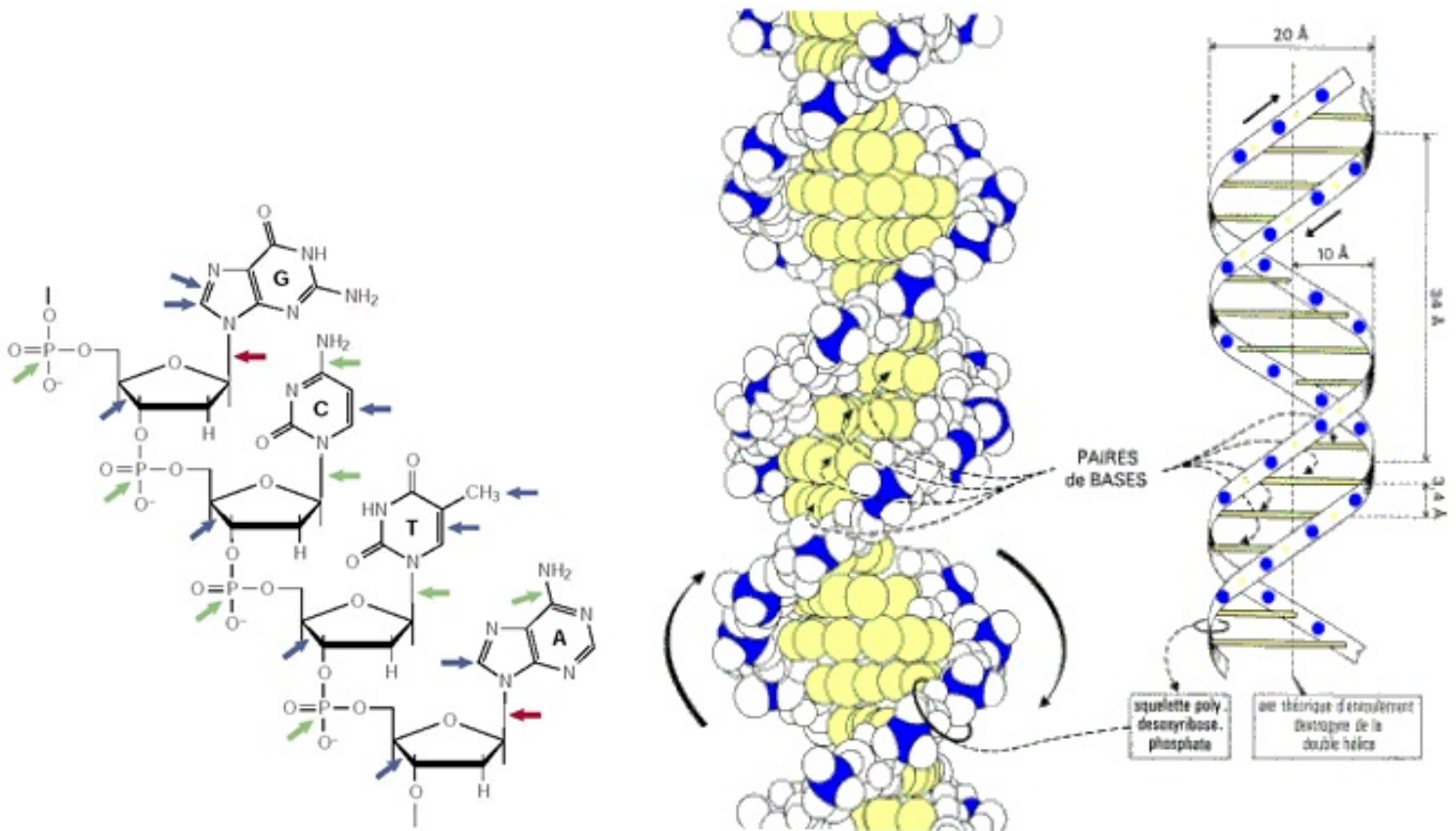


FIG. 1 – L'ADN : sa structure chimique avec le cycle pentose à gauche, une représentation atomique montrant la conformation hélicoïdale au centre, une représentation schématique des paires de bases à droite.

Il est utile de savoir que l'ADN, au sein d'une cellule, possède plusieurs conformations. Notamment, la topologie de l'ADN peut être variée, et il est courant que l'hélice soit sur-enroulée : il y a alors apparition de supertours ce qui permet une plus grande compactification de l'ADN. Ceci explique aussi l'existence d'enzymes spécifiques appelées topoisomérases qui vont jouer un rôle lorsque il sera nécessaire d'accéder à l'information génétique, lors des processus de réplication et de transcription.

Lors de notre étude expérimentale, nous utilisons un double-brin d'ADN sans supertours, ce qui n'est donc pas la configuration la plus courante *in vivo*. De plus, l'ADN *in vitro* est en solution donc le solvant n'est pas le même qu'*in vivo*, et la concentration en sel est différente.

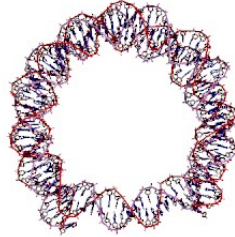


FIG. 2 – Sur-enroulement de l'ADN.

La taille de l'ADN se compte habituellement en paire de bases (bp). Une paire de base a une taille caractéristique d'environ 0,34 nm (cf. fig 1) ; les ADNs utilisés dans ce travail de recherche sont entre 400 bp et 2100 bp, ce qui correspond à des tailles caractéristiques entre 136 nm et 707 nm : nous sommes dans le domaine nanométrique.

L'ADN est une molécule possédant une certaine rigidité, principalement due à la double hélice. On va donc introduire une grandeur appelée *longueur de persistance* qui est telle qu'au-delà de cette longueur on commence à voir apparaître la flexibilité de l'ADN, en-deçà l'ADN est essentiellement rigide. Cette longueur est de l'ordre de 150 bp.

## 1.2 La transposition de l'ADN

Il existe dans l'ADN des séquences capables de se détacher et de migrer vers une autre portion de la même molécule d'ADN ou vers une autre molécule. On les appelle transposons, et le mécanisme décrivant cette action est le mécanisme de transposition dans l'ADN. Découvert par Barbara McClintock [4], ce qui lui a valu le prix Nobel en 1983, ce mécanisme joue un rôle majeur dans la diversité génétique du fait de la capacité des transposons à "sauter" (pour information, 45 % du génome humain est constitués d'éléments transposables). Ce mécanisme est surtout actif lors des divisions cellulaires. Plusieurs expériences penchent en faveur d'un rôle prédominant de ces éléments transposables dans l'évolution des espèces, par la création de nouveaux gènes.

Le mécanisme peut se décrire à l'aide du schéma suivant :

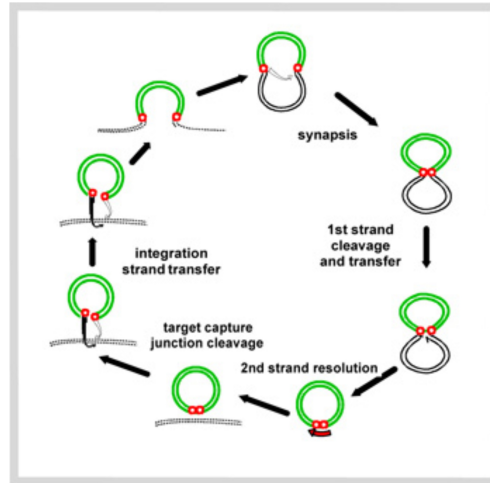


FIG. 3 – Schéma de principe du mécanisme de transposition dans l'ADN : formation, régulée par une enzyme appelée transposase, d'une boucle ; clivage et transfert de cette boucle sur un autre brin d'ADN ou une autre partie de la même molécule ; capture et intégration de cette boucle, puis débouclage.

L'enzyme régulatrice est appelée transposase ; dans les conditions *in vivo*, une deuxième enzyme appelée intégrase détecte le site initiateur et "aide" ainsi la transposase à agir. Dans nos expériences qui sont *in vitro*, seule la transposase joue un rôle : elle se fixe sur un des sites initiateurs appelé IR (on aura IRL et IRR pour Left et Right), puis initie la boucle. Ensuite il y a coupure du brin, puis migration vers un autre endroit de l'ADN ou un autre brin d'ADN, puis la transposase initie le recollement du nouveau brin.

Le but de notre recherche est de comprendre la première étape, qui est la formation de la boucle, et de suivre son évolution dynamique. Notamment, deux questions se posent, encore non complètement résolues :

- comment la transposase trouve-t-elle le site initiateur, par exploration 1D sur l'ADN ou par exploration 3D dans l'environnement du brin ?
- A-t-on fixation d'une transposase sur un IR qui ensuite attire le second IR, ou fixation de deux transposases sur chacun des IRs et interaction entre les deux protéines, suivi de la formation d'un dimère ?

Nous présenterons d'abord la méthode expérimentale utilisée pour visualiser le processus dynamique, puis nous donnerons les éléments du modèle théorique employé afin de comprendre la physique du processus. Enfin dans une dernière partie seront présentés les résultats expérimentaux obtenus au cours de ce stage afin d'améliorer la technique notamment ainsi que les résultats des simulations théoriques éclairant l'apport de la TPM à cette étude de la transposition.

## 2 Méthode expérimentale

### 2.1 Présentation de la TPM

Afin de suivre de manière dynamique l'étape de formation de la boucle, nous avons utilisé la technique dite TPM pour tethered particle motion, ou mouvement d'une particule attachée.

Dans cette approche expérimentale, le double-brin d'ADN est fixé sur une plaque de verre et à son autre extrémité est attachée une microbille fluorescente. Le mouvement brownien de cette microbille est alors suivi au microscope, équipé d'une vidéo d'acquisition. L'amplitude des fluctuations du mouvement de la bille dans le plan transverse à l'ADN croît avec la longueur effective de l'ADN : si une boucle se forme, ces fluctuations seront moins importantes, ce qui se verra directement au suivi par caméra. Cette approche a été appliquée pour la première fois à l'étude de l'élongation de la transcription par une ARN-polymérase unique [7].

Nous mesurons ainsi notamment la distance bout-à-bout de l'ADN, c'est-à-dire la distance dans le plan longitudinal entre le point d'attache de l'ADN et le point de fixation de la bille.

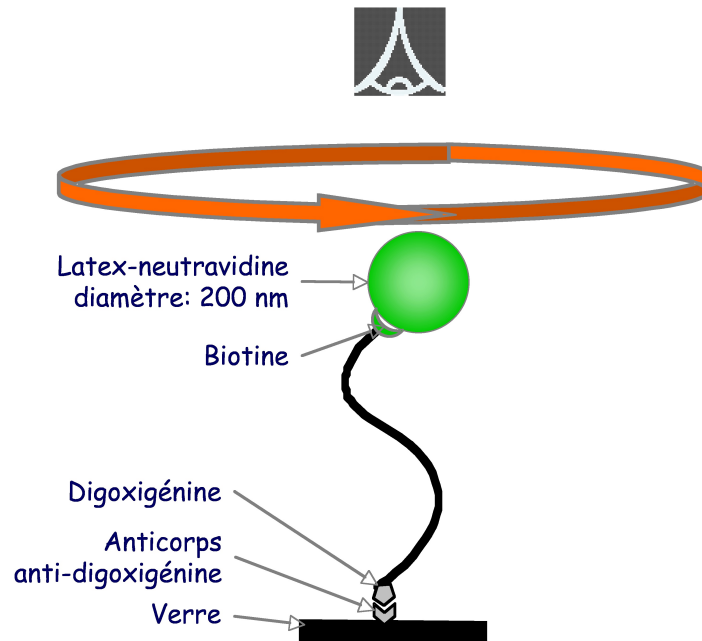


FIG. 4 – Schéma de principe de la méthode TPM.

Dans les expériences, nous mesurons directement la trajectoire d'une microbille unique, avec un temps d'acquisition de 40 ms dans un premier temps, et de 5 ms dans un second temps afin d'améliorer la technique et de s'affranchir des soucis de moyennage [1]. Il a été montré auparavant [5] que la relation entre le mouvement brownien et la taille effective de l'ADN était pratiquement linéaire pour des échantillons allant jusqu'à 2400 bp.

Il est nécessaire que l'échantillon baigne dans des protéines non actives permettant d'éviter que l'ADN ne se colle complètement sur la surface, sans compter les éléments permettant une fixation du brin sur la plaque de verre.

## 2.2 Application à notre étude, protocole utilisé

Dans le cadre de notre étude, nous avons donc utilisé la méthode TPM afin de suivre les premières étapes de la transposition de la séquence IS911 d'*E. coli*. Voici le schéma de principe :



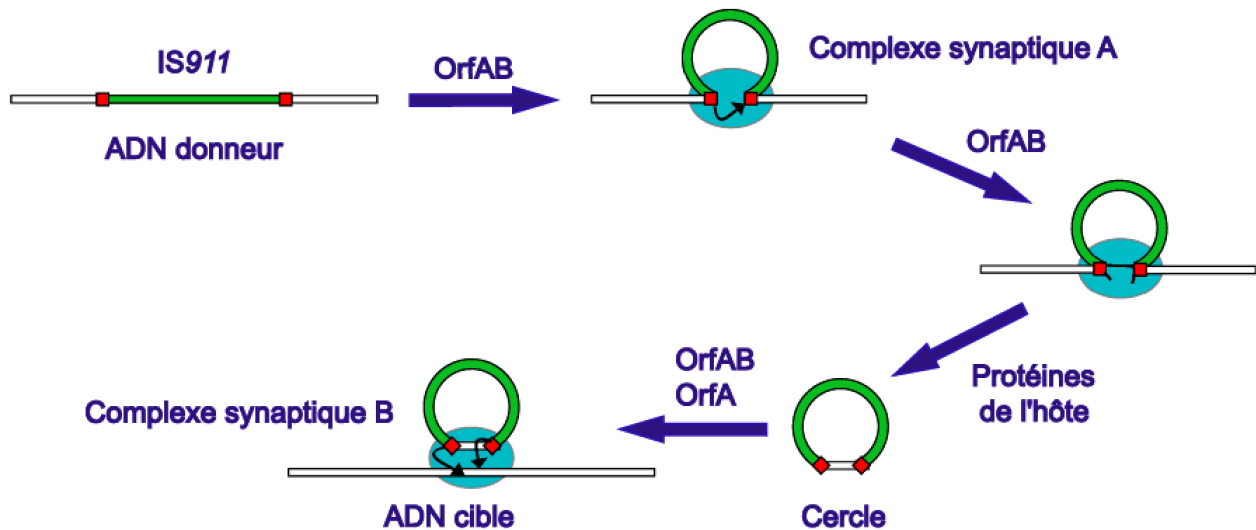


FIG. 5 – Schéma de principe de la transposition de *IS911* par la transposase OrfAB.

Dans la figure 5, l'étape que nous suivons dans cette étude est la toute première qui est la formation du complexe synaptique A, avec action de la protéine OrfAB sur l'*IS911* et la formation de la boucle en conséquence.

Pour cela nous utilisons la TPM, selon l'idée suivante :

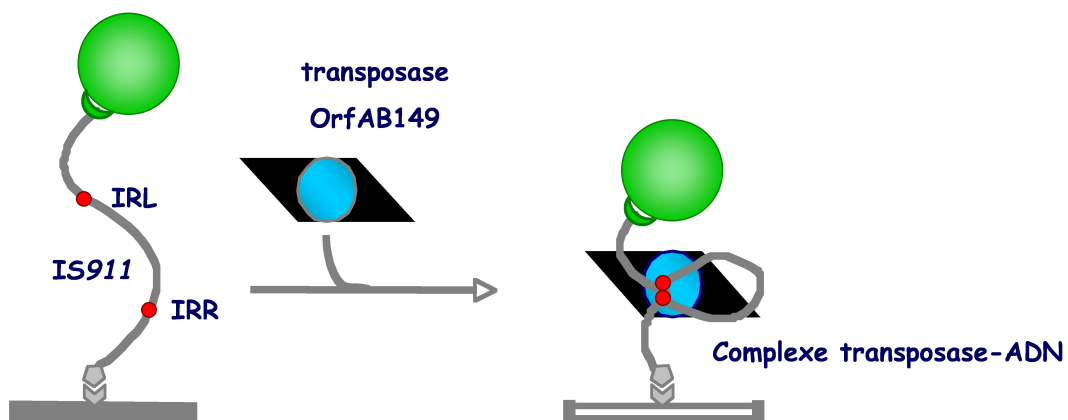


FIG. 6 – Diminution effective de la taille de l'ADN en présence de la transposase OrfAB : à gauche l'ADN non bouclé, à droite l'ADN bouclé et complexé, dont la taille effective, ie sans compter la boucle, est réduite ; les fluctuations observées de la bille sont donc différentes.

La formation du complexe et donc de la boucle induit une réduction de taille effective de l'échantillon étudié. Cette réduction de taille effective doit donc pouvoir s'observer dans les variations de fluctuations de la bille attachée.

Voici le protocole utilisé pour la préparation des échantillons :

### Jour 1, fonctionnalisation des lames

1. Laver lames et lamelles à l'alcool pur puis sonication (ultrasons) pendant 5 minutes. Faire sécher à l'étuve 110 °C.
2. Construire la chambre d'observation (lamelle sur lame avec scotch double face 100  $\mu\text{m}$ ).

3. Incuber les chambres avec 1 volume chambre d'anti-DIG à  $20 \mu\text{g.mL}^{-1}$  en dilution 10x (exemple :  $90 \mu\text{L PBS} + 10 \mu\text{L anti-DIG à } 200 \mu\text{g.mL}^{-1}$ ). L'anti-DIG est une protéine fonctionnelle accolée à l'ADN.
4. Rincer la chambre avec  $10V_{\text{chambre}}$  de PBS, et rajouter un TI (tampon d'intégration) de caséine à  $1 \text{ mg.mL}^{-1}$  durant la nuit, afin de passiver complètement la surface et de ne pas laisser à l'ADN la possibilité de migrer.

### Jour 2, préparation des lamelles

1. Sécher la chambre.
2. Incuber les chambres avec  $2V_{\text{chambre}}$  d'ADN dilué dans le TI ( $1,4.10^{-10} \text{ M}$ ) pendant 1h et soniquer les particules diluées 10x dans TI.
3. Rincer la chambre avec  $10V_{\text{chambre}}$  de TI.
4. Incuber les chambres avec les particules à 0,3% dans le TI ( $145 \mu\text{L TI} + 5 \mu\text{L de solution de particules diluées } 10\text{x}$ ) pendant 20 minutes.
5. Rincer la chambre avec  $10V_{\text{chambre}}$  de TI.
6. Injecter  $2V_{\text{chambre}}$  de TI contenant la protéine ( $0,1 \mu\text{M}$ ).

Dans l'ensemble d'expériences effectuées sur 4 jours, nous avons  $V_{\text{chambre}} = 50\mu\text{L}$ .

Comme dit plus haut dans le protocole, l'intérêt majeur de rajouter des protéines annexes telles que la caséine est de passiver la surface, et donc d'empêcher le collage de l'ADN sur la plaque de verre. La lame est ensuite mise sous un microscope équipé d'une caméra, et éclairée par un laser, comme on peut le voir sur la figure 7. Toute ces manipulations ont fait partie intégrante du stage.

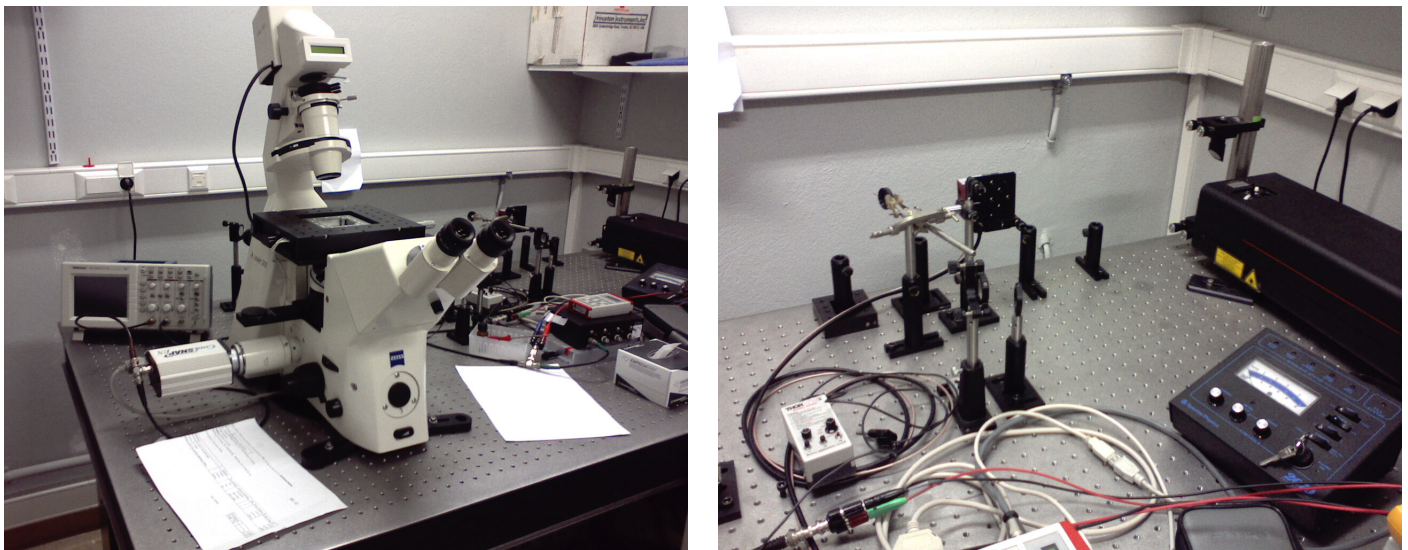


FIG. 7 – Microscope utilisé (à gauche) et alignement laser et lentille pour amener le faisceau laser sur l'échantillon (à droite).

### 2.3 Résultats antérieurs

Avant de discuter des résultats expérimentaux obtenus durant le stage, qui concernent notamment les améliorations désirées de la technique, il est bon de présenter un aperçu des précédents

résultats expérimentaux obtenus lors de la thèse de Noelle Pouget dans [6].

L'ADN utilisé dans ces expériences était un ADN double-brin d'une taille d'environ 2000 bp. La distance entre les deux IRs était d'environ 1200 bp, et cela sur 6 échantillons d'ADN différents. Voici tout particulièrement la séquence pour l'ADN de 2080 bp, utilisé en simulation :

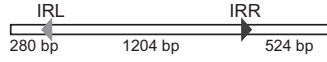


FIG. 8 – Position des IRs sur le brin d'ADN de 2080 bp. La longueur effective après bouclage est de 804 bp.

Lors de l'analyse des données collectées, on ne tient pas compte de toutes les mesures où plus d'une particule s'est attachée à un brin d'ADN, de même pour les mesures où une même particule s'est attachée à plusieurs brins. Toutes ces mesures sont discriminées en temps réel, le premier type de biais étant déterminé à l'oeil en visualisant plusieurs particules à la caméra, le second étant déterminé par la sphéricité de la distribution spatiale des positions de la particule.

Le résultat essentiel est visualisé par la figure 9 suivante :

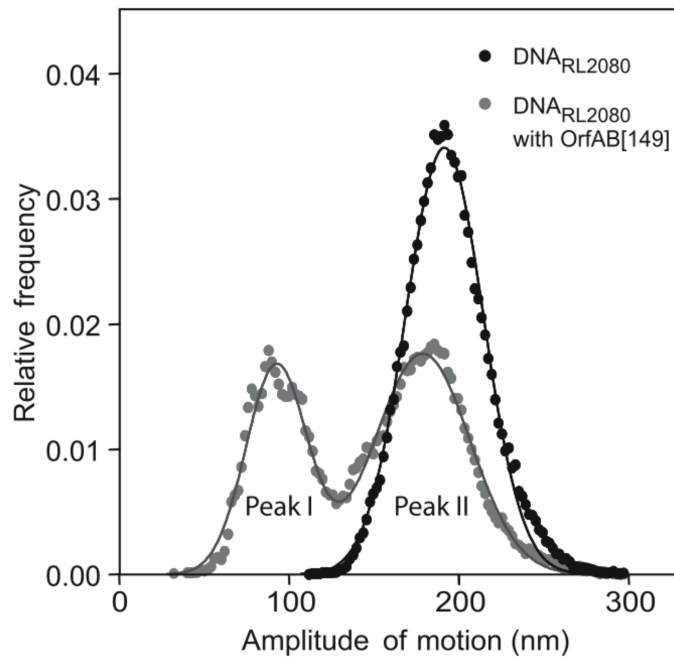


FIG. 9 – Distribution des positions de la bille par rapport au point d'attache, résultats expérimentaux obtenus par [6] pour un temps d'acquisition de 40 ms.

On arrive donc très clairement à distinguer les deux états de la molécule à l'aide de la TPM. Les résultats de l'étude [6] ont montré que la formation de la boucle est relativement rapide (moins d'une seconde) et qu'une fois formés, les complexes enzyme-ADN bouclé sont stables. Il est bon de noter qu'une étude a montré que les temps donnés par les données brutes étaient des temps apparents, l'article [1] montre comment remonter aux temps réels à partir des données expérimentales.

Nous souhaitons dans notre étude reprendre la méthode TPM avec un temps d'acquisition plus court d'une part, à 5 ms au lieu de 40 ms comme ci-dessus et qui induit un processus de moyennage

sur ces 40 ms, et simuler de manière théorique l'ADN afin de comprendre la dynamique de la boucle d'autre part.

### 3 Modèle théorique

Pour modéliser notre double-brin d'ADN dans l'expérience TPM, on va faire des hypothèses simplificatrices et utiliser des modèles statistiques de polymères.

Le parti pris de faire des simulations en lieu et place de calculs analytiques s'explique par la prise en compte des effets de volume exclu ; ces effets jouent un rôle non-négligeable dans la dynamique de la formation de la boucle, or il est impossible de calculer explicitement ces effets par une démarche purement analytique. Il est donc nécessaire de faire des simulations pour être au plus près des expériences. La description détaillée des méthodes de simulations se trouve dans [3].

#### 3.1 Modèle de ver (worm like chain)

Le double-brin d'ADN est un polymère doté d'une grande rigidité au niveau de la paire de base, ceci étant dû principalement à la structure en double-hélice. Pour rendre compte de cet effet, le modèle le plus courant de l'ADN est le modèle du ver (worm like chain ou WMC), qui est un modèle continu de polymère.

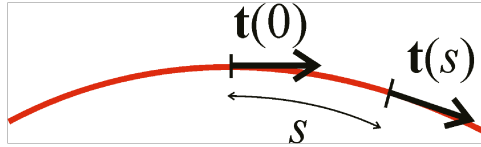


FIG. 10 – Modèle WMC,  $\vec{t}(s)$  désigne le vecteur tangent à la chaîne,  $s$  est l'abscisse curviligne (ie la position le long de la chaîne).

Sur la figure 10, le vecteur  $\vec{t}(s)$  est le paramètre principal le long de l'ADN et désigne le vecteur tangent à la position  $s$ . Dans le modèle WMC, les vecteurs  $\vec{t}(s)$  sont corrélés :

$$\langle \vec{t}(s) \cdot \vec{t}(s') \rangle = e^{-\frac{|s-s'|}{l_p}} \quad (1)$$

et l'on fait apparaître la longueur de corrélation  $l_p$  qui est la longueur de persistance de l'ADN, rendant compte de sa rigidité à courte distance. Cette longueur fondamentale, qui dépend de la concentration en sel, vaut  $l_p \simeq 150$  bp pour des concentration mM, c'est-à-dire 50 nm.

Dans notre simulation, nous introduisons alors un potentiel modélisant cette rigidité locale et nous discrétisons le modèle. La grandeur  $a$  représente la longueur d'un monomère,  $T$  est la température ambiante.

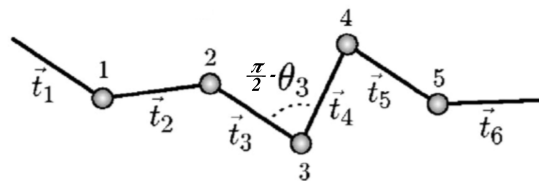


FIG. 11 – Paramètres du modèle de ver discrétisé.

Ici  $\theta_i$  représente l'angle entre deux liens successifs,  $\vec{t}_i = \vec{r}_{i+1} - \vec{r}_i$ .

Nous introduisons dans le modèle le potentiel suivant :

$$\tilde{U}_{WLC} = \sum_{i=0}^{N-1} \left( \frac{\tilde{\varepsilon}}{2} (1 - \cos \theta_i) + \frac{\tilde{\gamma}}{4} (|\vec{t}_i| - 2)^2 \right) \quad (2)$$

où  $N$  est le nombre de monomères du modèle, et où toutes les quantités sont adimensionnées :

$$\vec{r} = \frac{\vec{r}}{a}; \tilde{\gamma} = \frac{\gamma a}{k_B T}; \tilde{\varepsilon} = \frac{\varepsilon}{a k_B T}; \tilde{U} = \frac{U}{k_B T}$$

Le premier terme dans l'équation 2 est le terme spécifique du modèle de ver, et représente la rigidité locale, puisque c'est un terme qui a tendance à aligner les vecteurs tangents  $\vec{t}_i$ . Le second terme est un terme d'étirement longitudinal de la chaîne, modélisé via un modèle classique d'oscillateur harmonique. Dans notre implémentation, les modules d'étirement  $\tilde{\gamma}$  et de courbure  $\tilde{\varepsilon}$  sont reliés par la relation suivante, valable pour un cylindre élastique isotrope

$$\tilde{\gamma} = 4\tilde{\varepsilon} \quad (3)$$

Les références [2][3] nous donnent la relation fondamentale suivante entre les paramètres du modèles et les paramètres physiques :

$$\frac{L}{l_p} = \frac{2N}{\tilde{\varepsilon}} \quad (4)$$

Nous pourrions ainsi déterminer le paramètre  $\tilde{\varepsilon}$  de manière réaliste.

### 3.2 Modélisation via les forces

Dans notre simulation, nous effectuons un calcul de forces. Nous modélisons la chaîne d'ADN par le potentiel WMC décrit ci-dessus, nous rajoutons les effets de volume exclu ainsi que le mouvement brownien des monomères.

Les effets de volume exclu sont des effets stériques, dûs au fait que chaque monomère occupe un certain volume de l'espace inaccessible aux autres monomères. C'est donc à la fois un effet entropique et un effet type "sphère dures".

Nous utilisons un potentiel classique de Lennard-Jones afin d'en tenir compte dans le calcul des forces, actif dès lors que  $|\vec{r}_i - \vec{r}_j| < 2a$  :

$$\tilde{U}_{LJ} = \sum_{i < j} \left( \left( \frac{2}{|\vec{r}_i - \vec{r}_j|} \right)^{12} - 2 \left( \frac{2}{|\vec{r}_i - \vec{r}_j|} \right)^6 \right) \quad (5)$$

De plus, le double-brin d'ADN est attaché à la surface de la plaque de verre. Nous restreignons ainsi le mouvement des monomères à être dans le demi-espace supérieur  $z > 0$ , ce qui se réalise [2] par la condition

$$r_z < a \rightarrow r_{z_{refl}} = 2a - r_z$$

Pour la bille attachée à l'ADN, nous changeons ces conditions en utilisant  $\tilde{R}$  à la place de  $\tilde{r}$ .

La prise en compte du mouvement brownien se fait à l'aide d'une équation de Langevin :

$$\frac{d\vec{r}}{dt} = \mu_0 \left( -\mathbf{grad}_{\vec{r}} U + \vec{F}_i^{ext} \right) + \vec{\zeta}_i(t) \quad (6)$$

où  $\mu_0$  est la mobilité d'une sphère de rayon  $a$  dans un solvant de viscosité  $\eta$ ,  $\mu_0 = \frac{1}{6\pi\eta a}$ . Pour l'eau par exemple,  $\eta \simeq 10^{-3}$  Pa.s .

Le terme  $\vec{\zeta}_i(t)$  est le terme de force aléatoire dit de Langevin :

$$\langle \vec{\zeta}_i(t) \rangle = 0; \langle \vec{\zeta}_i(t) \cdot \vec{\zeta}_j(t') \rangle = 6k_B T \mu_0 \delta_{ij} \delta(t - t'). \quad (7)$$

Les autres effets hydrodynamiques sont négligés, notamment ceux entre monomères.

Cette équation est ensuite discrétisée en temps et renormalisée :

$$\tilde{r}_i(n+1) = \tilde{r}_i(n) + \tilde{\mu} \left( -\mathbf{grad}_{\tilde{r}_i} \tilde{U} + \tilde{F}_i^{ext} \right) + \sqrt{6\tilde{\mu}} \tilde{\zeta}_i(n) \quad (8)$$

où l'on a

$$n = \frac{t}{\Delta t}; \tilde{\mu} = \Delta t \frac{k_B T}{a^2} \mu_0$$

$\tilde{\zeta}_i$  est une fonction renormalisée à 1, donc variant entre -1 et 1, et sera implémentée via un générateur pseudo-aléatoire.

### 3.3 Modélisation de l'attraction entre IRs

Dans un premier temps, nous avons modélisé l'action des protéines comme une attraction effective entre les deux IRs, en terme de champ moyen qui tient compte de l'énergie libre entre protéines ainsi que de l'énergie du complexe OrfAB+IRs.

Pour modéliser l'attraction entre les deux sites d'activation du transposon (les IRs), nous utilisons un potentiel de Morse dont les deux paramètres sont la profondeur  $D$  et la portée  $\alpha^{-1}$  :

$$U_M(d) = D (1 - e^{-\alpha d})^2 \quad (9)$$

La distance  $d$  représente la distance euclidienne entre les deux IRs. Dans la figure 12 ci-dessous le potentiel utilisé est représenté :

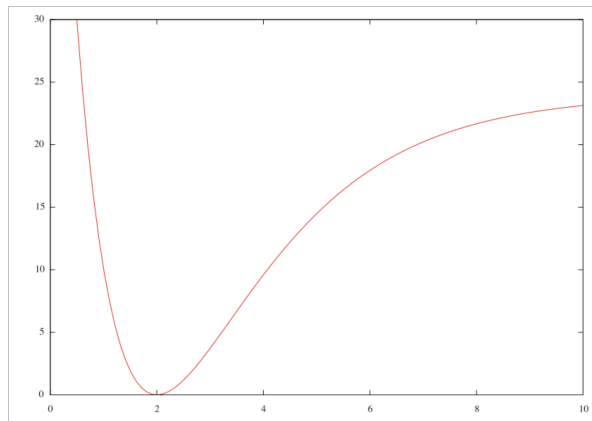


FIG. 12 – Potentiel de Morse décalé pour tenir compte des effets de volume exclu.

On a décalé l'origine pour tenir compte des effets de volume exclu. Le choix de ce potentiel permet de tenir compte de la possible ouverture de la boucle, à cause des fluctuations thermiques. De ce potentiel est dérivée l'expression de la force entre les deux IRs.

Afin de savoir si la boucle peut tenir une fois formée, il faut comparer l'effet de ce potentiel à l'énergie de rigidité. Ceci est fait dans [2] et donne la condition

$$D > 2\pi^2 \frac{\varepsilon}{S} \quad (10)$$

où  $\varepsilon$  est le module de courbure introduit dans la modélisation de la WMC,  $S$  est la longueur de la boucle formée.

Il faudra tenir compte de cette contrainte lors du choix des paramètres lors de la simulation.

## 4 Résultats des simulations et résultats expérimentaux

Avant d'aborder les résultats des simulations théoriques et des manipulations expérimentales, il est bon de donner ici quelques outils importants pour interpréter les données issues des simulations.

Le programme utilisé pour les simulations repose majoritairement sur un code développé dans [2]. Quelques retouches ont été effectuées durant le stage, mais essentiellement il a été question de manipulation des paramètres essentiels du programme. Deux formules sont essentielles pour remonter des données brutes aux données physiques, puisque tous les résultats sont donnés en unité de  $a$ , qui est la longueur d'un monomère simulé selon le modèle théorique développé dans la partie précédente.

D'une part nous utilisons la formule

$$a_{eff} = 0,34 \frac{L_{chain}(bp)}{L_{simu}} \text{ (nm)} \quad (11)$$

afin d'avoir la valeur numérique du  $a$  simulé, où  $L_{chain}$  désigne la longueur de l'échantillon simulé, en bp, et  $L_{simu}$  désigne la longueur simulée qui est donc deux fois le nombre de monomères ; d'autre part nous avons

$$\Delta t = 4,7 \cdot 10^{15} a_{eff}^3 \text{ s pour une mobilité réduite } \tilde{\mu} = 10^{-3}. \quad (12)$$

### 4.1 Résultats des simulations sur l'ADN de 2080 bp

Les simulations sur l'ADN à 2080 bp se confrontent aux expériences réalisées par Noelle Pouget *et al.* [6]. Le but est ici de valider d'un point de vue théorique la méthode expérimentale d'une part, et de s'approcher le plus possible des résultats expérimentaux d'autre part par un bon choix de paramètres.

Il faut noter que le temps de simulation est assez élevé, ce qui limite le nombre de simulations possibles sur le temps imparti au stage : les simulations durent environ 2 jours, pour un temps réel de 2 s. Afin d'avoir une statistique, nous avons effectué une série de 10 simulations où l'on a changé le générateur pseudo-aléatoire utilisé dans l'équation de Langevin, ceci afin de simuler les variations obtenues au sein d'une prise de données expérimentales.

Au début nous avons eu quelques soucis avec la simulation du potentiel de Morse, notamment la boucle ne se formait pas. Nous avons donc remplacé ce potentiel par un potentiel harmonique, puisque ce qui nous intéressait était l'état bouclé et non la possibilité que la boucle ne soit pas stable. Nous reviendrons plus tard sur le potentiel de Morse. Voici les résultats obtenus dans cette optique :

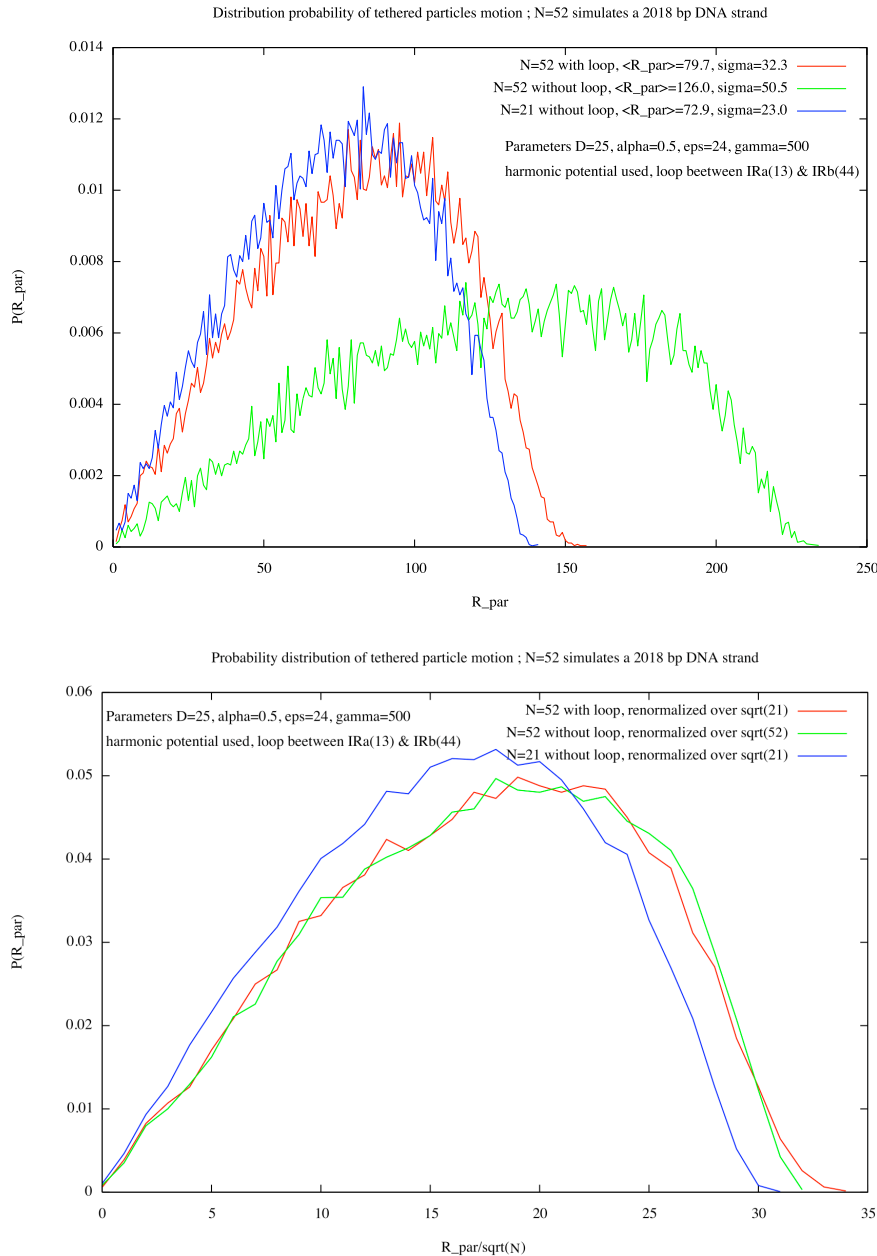


FIG. 13 – Distribution théorique des positions de la bille avec un ADN de longueur 2080 bp - en haut les données brutes, en bas les données ramenées à la longueur effective de l'ADN. Ici le potentiel harmonique a été utilisé pour le bouclage.

La simulation avec 21 monomères concerne un brin d'ADN de longueur 804 bp (qui est la longueur effective d'un brin d'ADN 2080 bouclé entre les deux IRs). On voit donc immédiatement avec le graphe du haut de la figure 13 que l'on peut distinguer l'état bouclé de l'état non bouclé, et que l'état bouclé a un comportement très similaire à un état non bouclé, mais de même longueur effective. Le graphe du bas de la figure 13 permet de valider cette approche, puisque on voit que les courbes bleues et rouges se superposent, ce qui discrimine bien de la distribution bleue qui concerne l'ADN test non bouclé.

De plus, l'écart que l'on observe sur le graphe du haut de la figure 13 entre les courbes rouge et bleue est un effet entropique dû à la présence de la boucle, qui limite le nombre de configurations de la chaîne, ce qui permet de mettre en évidence la présence de la boucle, au-delà de la simple réduction de longueur effective.



On a ensuite calculé le temps de corrélation, via la formule habituelle

$$C(s) = \frac{1}{T - s + 1} \sum_{t=0}^{T-s} R(t)R(t + s) \quad (13)$$

où  $R(t)$  désigne la position de la bille à l'instant  $t$ . On fait un ajustement exponentiel de la corrélation  $C(s)$  qui nous donne un temps caractéristique  $\tau$  qui est le temps de corrélation. On obtient typiquement des graphes de ce type :

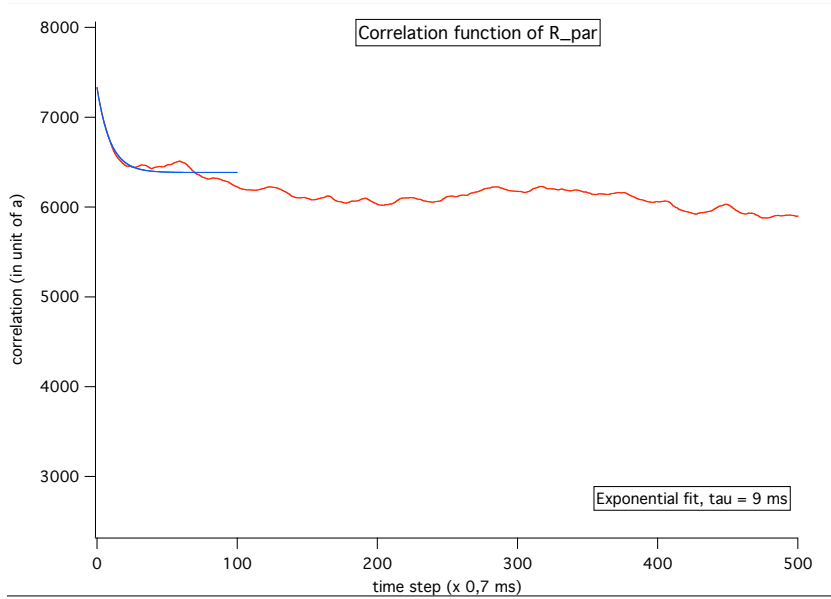


FIG. 14 – Corrélation temporelle typique pour la position de la bille au cours d’une simulation de l’ADN 2080 bp.

On obtient un temps de corrélation moyen sur l’ensemble des simulations  $\tau = (7,0 \pm 1,5)$  ms. Il est à noter que ce temps est assez petit par rapport aux expériences (cf section suivante), ceci est possiblement dû au fait que les effets hydrodynamiques sont globalement négligés lors des simulations, notamment l’effet de la paroi de verre sur les frottements visqueux. Ceci peut ralentir les effets diffusifs.

Par la suite, le problème de code en ce qui concerne le potentiel de Morse a été résolu, il s’est avéré qu’il était dû à une combinaison erronée des paramètres. Une série de 10 nouvelles simulations avec un potentiel de Morse a été réalisée afin de tracer le graphe de la distribution de probabilité pour l’ADN 2080 bouclé, ce qui a donné le résultat de la figure 15 suivante :

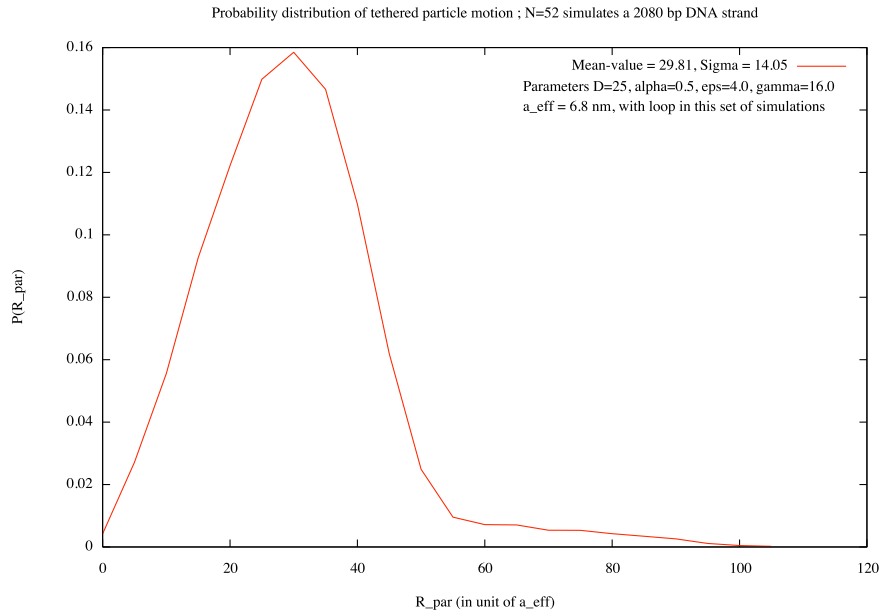


FIG. 15 – Distribution théorique des positions de la bille avec un ADN de longueur 2080 bp bouclé, avec un potentiel de Morse pour l’attraction entre IRs.

La queue à gauche est possiblement due au fait que de rares ouvertures de la boucle sont possibles, nous négligeons l’effet de l’état initial la boucle se formant très vite.

Les simulations ont été montées en vidéo, afin de suivre la dynamique de formation de la boucle. Celle-ci se forme assez rapidement, en accord avec les expériences. Voici un instantané donnant une idée de ce que l’on observe dans de telles vidéos :

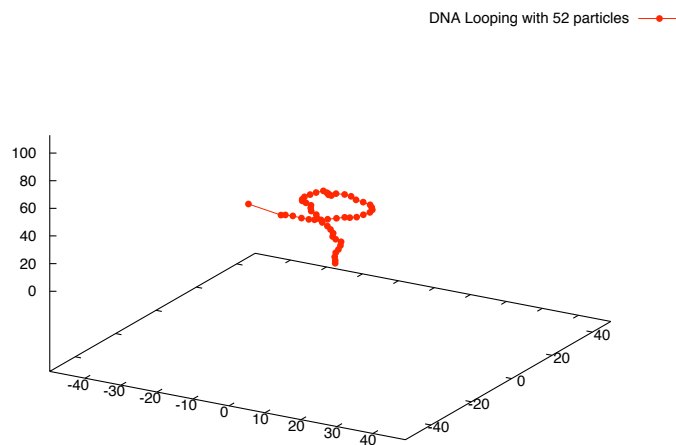


FIG. 16 – Instantané d’une simulation sur l’ADN de 2080 bp qui se boucle, avec un potentiel réaliste de Morse.

## 4.2 Résultats sur l’ADN de 400 bp

L’ADN à 400 bp concerne les échantillons utilisés durant les expériences réalisées au cours du stage. Deux séries d’expériences différentes ont été réalisées : l’une avec des temps d’exposition de 40 ms, la seconde avec des temps d’exposition de 5 ms obtenus par un cache périodique du laser à

l'aide d'un dispositif mécanique. Nous font remarquer ici que tous les graphes comportent la mention "DNA 500 bp", car en effet à l'origine l'expérience devait se dérouler avec des ADN d'une longueur de 500 bp, ce n'est qu'après avoir obtenu et analysé les résultats expérimentaux que les biologistes de notre équipe ont remarqué que l'ADN préparé était en fait d'une longueur de 400 bp.

Voici ce que l'on peut observer comme distribution typique sur le plan transverse en ce qui concerne le relevé des positions de la bille, ramenées à la position du point d'accroche de l'ADN sur la plaque de verre :

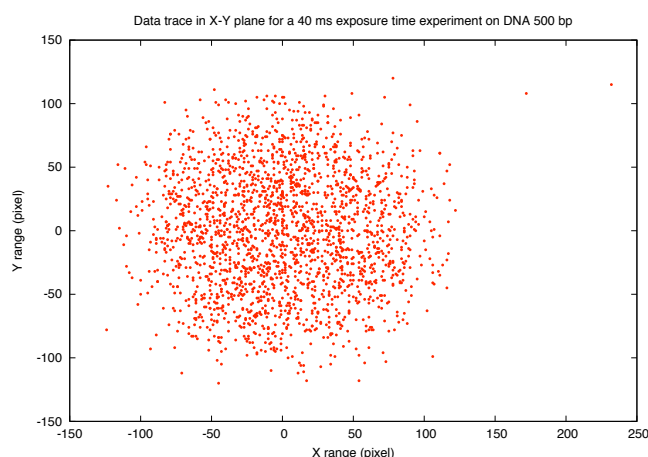


FIG. 17 – Distribution spatiale des positions de la bille accrochée à l'échantillon d'ADN pour une prise de mesure ; les distances sont en pixels.

Ainsi, cette prise de mesure a été gardée à l'analyse car elle correspond bien au suivi d'un seul double-brin d'ADN par une seule bille, comme en témoigne la relative bonne isotropie de la tâche.

Voici d'abord les résultats de l'analyse des deux séries de mesure à 40 ms sur les deux jours d'expériences :

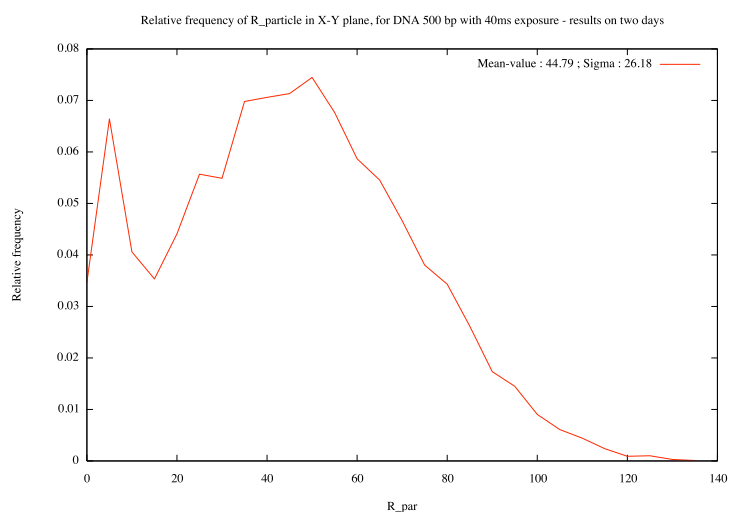


FIG. 18 – Distribution expérimentale des positions de la bille avec un temps d'exposition de 40 ms, collecte des données sur 2 jours.

On peut observer une distribution qui serait gaussienne sans le second pic à gauche, et qui serait conforme à ce que l'on attendrait pour des échantillons sans protéines, cf figure 15. On a donc procédé à une séparation des deux prises de mesures des 29 et 30 janvier 2008, ce que montre la figure 19 :

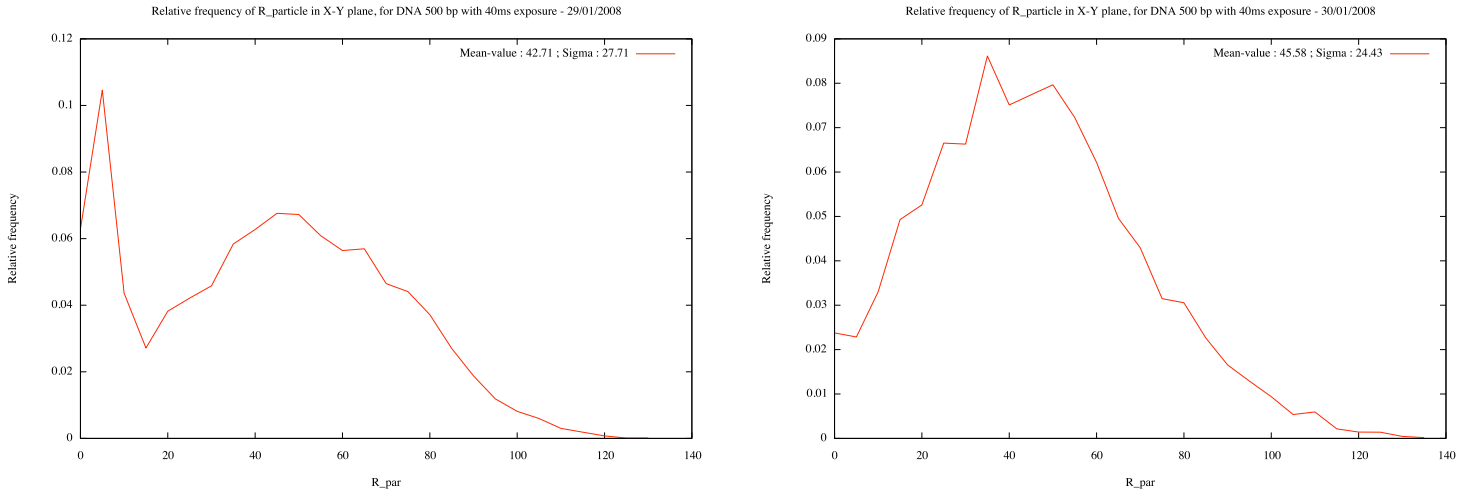


FIG. 19 – Distribution expérimentale des positions de la bille avec un temps d'exposition de 40 ms. Expérience du 29 janvier (à gauche) et du 30 janvier (à droite).

On voit donc que le pic supplémentaire est dû à la première expérience. Afin d'en savoir plus sur la distribution, voici un ajustement des données du 29 janvier, réalisé avec une fonction di-gaussienne :

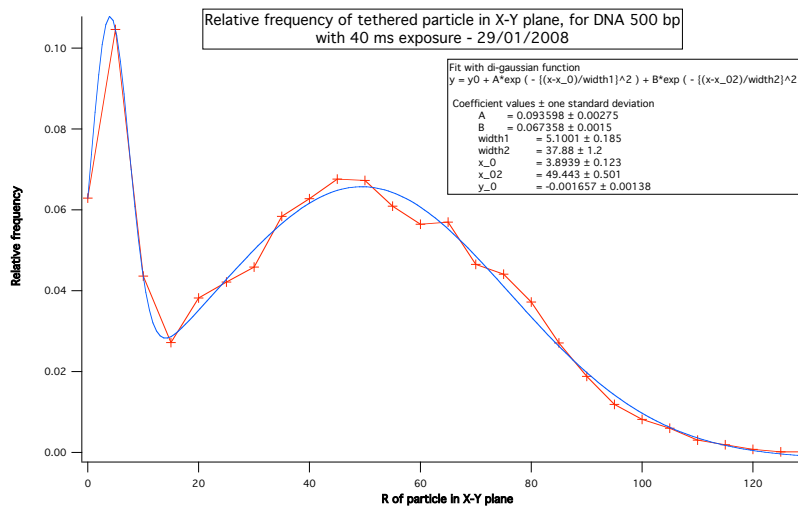


FIG. 20 – Ajustement des données expérimentales du 29 janvier 2008.

L'ajustement réalisé dans la figure 20 est correct, et permet de donner la valeur moyenne  $x_{02} = R_{par} = 49,4$  nm et son écart-type  $width2 = \sigma = 37,9$  nm. Une explication possible de ce premier pic dans les données serait une adsorption de certains échantillons sur la plaque de verre support de l'échantillon d'ADN. Ainsi, on observerait une plus grande fréquence des petites amplitudes de la bille, puisque celle-ci étant adsorbée elle ne bouge pratiquement plus. De récentes expériences réalisées après la fin du stage semblent confirmer cette hypothèse.

En ce qui concerne la série de mesures avec un temps d'exposition de 5 ms, nous obtenons les résultats suivants :

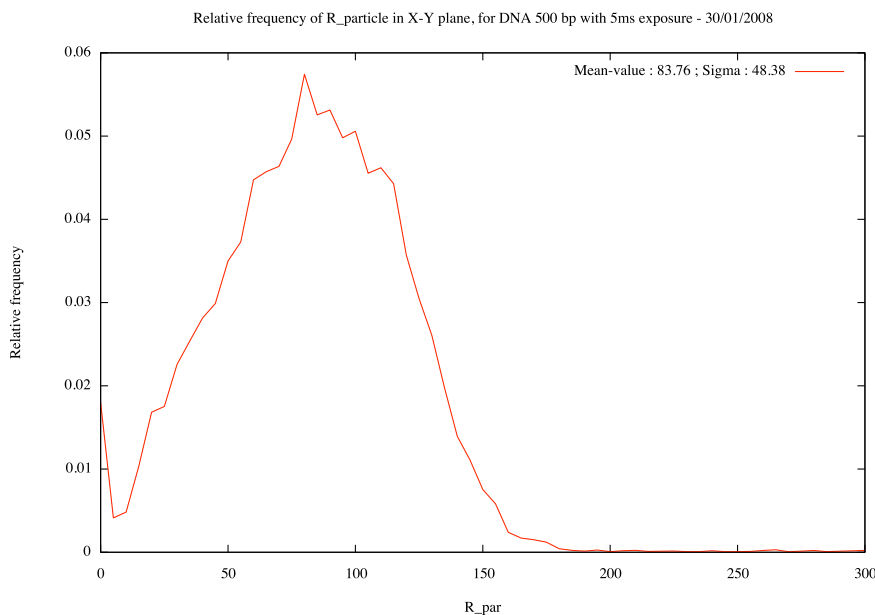


FIG. 21 – Distribution expérimentale des positions de la bille avec un temps d'exposition de 5 ms.

Le nombre de données correctes lors de cette prise de données est de 14, ce qui est le même ordre de grandeur que les deux séries combinées de 40 ms. Il est bon de noter que les expériences réalisées le second jour, concernant la deuxième série de 40 ms et la série de 5 ms, ont été bien meilleures que l'expérience réalisée la veille, ceci étant possiblement dû à une légère modification du protocole expérimental concernant la caséine.

Voici, pour comparaison, le résultat d'une simulation théorique, dont les paramètres sont indiqués sur la figure 22 :

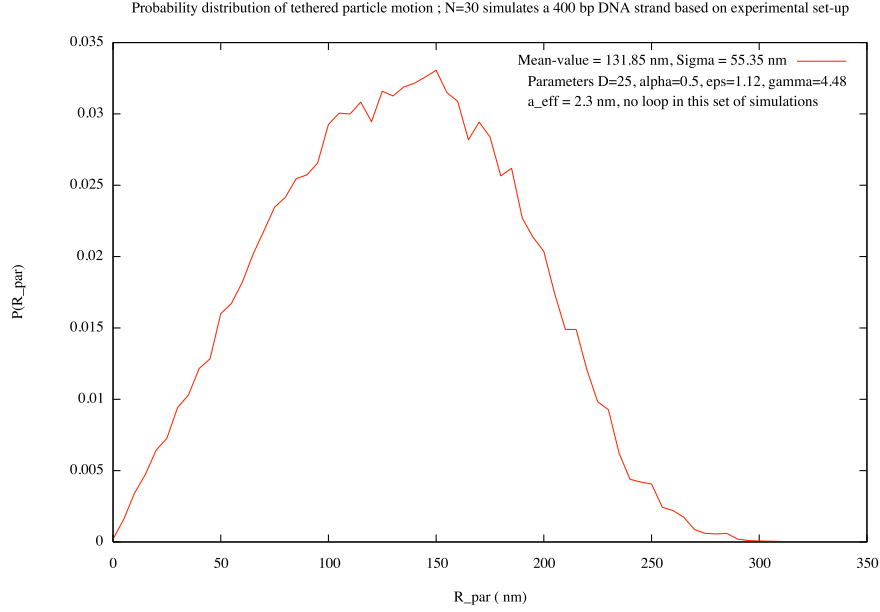


FIG. 22 – Distribution théorique des positions de la bille avec un ADN de longueur 400 bp.

Ces deux graphes ont la même forme qualitative, ce qui est significatif en ce qui concerne l'accord théorie-expérience. De plus, nous avons ici un résultat significatif quant à la faisabilité de l'expérience avec des temps de 5 ms, puisque nous obtenons le même type de courbe expérimentale, et l'écart observé entre la position moyenne (en pixel) entre la série à 40 ms et la série à 5 ms est celui que l'on attendait. De plus la comparaison de  $\langle R_{par} \rangle = 132$  nm avec le rayon de la bille  $R = 200$  nm est qualitativement bonne, il reste à améliorer l'aspect quantitatif. Par contre le décalage entre les deux moyennes reste à expliquer.

Nous avons obtenu les temps de corrélation suivants pour les deux séries d'expériences :

TAB. 1 – Temps ergodiques (ie de corrélation) des expériences

Expérience	Temps moyen $\tau$ (ms)	Écart-type $\sqrt{\sigma_\tau^2}$ (ms)
Série de 40 ms du 29 janvier 2008	14,2	3,8
Série 40 ms du 30 janvier 2008	16,3	2,1
Combinaison des deux séries de 40 ms	15,4	3,2
Série de 5 ms	9,9	1,4

Ces temps sont comparables à ceux obtenus pour la simulation théorique de l'ADN à 400 bp avec les paramètres expérimentaux :  $\tau_{theorique} = (17.6 \pm 3.5)$  ms, ce qui est un signe de bon accord qualitatif entre le modèle théorie et les expérience. Plusieurs points restent cependant à expliquer, tels que la différence d'extension entre les figures 21 et 22 ou bien le fait que les temps de corrélation, bien que comparables, sont néanmoins différents.

## Conclusion

Ce stage, qui s'est déroulé du 14 janvier au 20 février, a été riche d'enseignements. Tout d'abord il a été l'occasion pour moi d'aborder la biophysique, et de découvrir une technique expérimentale relativement nouvelle permettant de faire des expériences dynamiques en molécule unique sur les objets biologiques. J'ai aussi pu mettre en pratique mes connaissances en physique statistique pour pouvoir simuler de manière théorique les phénomènes physiques mis en jeu dans le mécanisme étudié, celui de la transposition de l'ADN, un mécanisme majeur pour la plasticité du génome.

Quels sont les résultats acquis ? En ce qui concerne la partie purement expérimentale, les expériences réalisées sur des brins d'ADN d'une taille de 400 bp sont très prometteuses quant à la possibilité de passer à une acquisition de 5 ms, ce qui permettrait d'avoir de meilleurs résultats dynamiques, puisque le moyennage d'un point expérimental ne se ferait plus que sur une fenêtre de 5 ms, au lieu de 40 ms précédemment.

Les simulations ont permis de valider la méthode TPM, et ont montré que l'on observait bien un effet de la boucle via les courbes de distribution de la position de la bille. Quelques soucis de programmation - soucis corrigés en fin de stage - n'ont pas pu permettre d'ajuster correctement au cours du stage les paramètres du potentiel de Morse afin de reproduire au mieux les expériences.

Il reste à simuler avec un bain de protéines, afin de se rapprocher le plus possible des conditions expérimentales et notamment de comprendre si l'interaction entre IRs est due à une seule protéine ou si elle passe par l'interaction entre deux transposases, chacune fixées sur un des IRs. En ce qui concerne le volet expérimental, des expériences sont en cours de réalisation pour voir le comportement de l'ADN en présence de transposases, avec des fenêtres d'acquisition de 5 ms.

## Références

- [1] Nicolas Destainville and Laurence Salomé, *Quantification and Correction of Systematic Errors Due to Detector Time-Averaging in Single-Molecule Tracking Experiments*, Biophysical Journal (2006), L17–L19.
- [2] Dennis Gotta, *Brownian dynamics simulation of tethered particle motion experiments : DNA loop formation*, Rapport de stage L3 Erasmus (2007).
- [3] Manoel Manghi, Xavier Schlagberger, Yong-Woon Kim, and Roland R. Netz, *Hydrodynamic effects in driven soft matter*, Soft Matter **2** (2006), 653–668.
- [4] Barbara McClintock, *The origin and behavior of mutable loci in maize.*, Proc Natl Acad Sci U.S.A. (1950), 344–355.
- [5] Noelle Pouget, Cynthia Dennis, Catherine Turlan, Mikhail Grigoriev, Michael Chandler, and Laurence Salomé, *Single-particle tracking for DNA tether length monitoring*, Nucleic Acids Research **32**, No. **9** (2004), 73–80.
- [6] Noelle Pouget, Catherine Turlan, Nicolas Destainville, Laurence Salomé, and Michael Chandler, *IS911 transpososome assembly as analysed by tethered particle motion*, Nucleic Acids Research **34**, No. **16** (2006), 4313–4323.
- [7] D. A. Schafer, J. Gelles, M. P. Sheetz, and R. Landick, *Transcription by single molecules of RNA polymerase observed by light microscopy*, Nature **352** (1991), 444–448.
- [8] Catherine Tardin, Philippe Rousseau, Olivier Walisko, Nicolas Destainville, Michael Chandler, and Laurence Salomé, *Insights in the transposition mechanism of the bacterial insertion sequence IS911 revealed by Tethered Particle Motion*, Biophysical Journal **92** (2008), 272.
- [9] James Watson and Francis Crick, *Molecular structure of nucleic acids ; a structure for deoxyribose nucleic acid*, Nature **171** (**4356**) (1953), 737–738.