
Etude théorique de la translocation de biomolécules à travers un nanopore.

Abstract : *Projet bibliographique, Timothée Ménais, M 2R EP, 2012.*

Durant mon stage, j'étudierai d'un point de vue analytique et numérique la translocation (passage) de biomolécules (principalement de l'ADN simple brin) à travers un nanopore de graphène, au sein du Groupe Théorie, SPrAM, INAC du CEA de Grenoble. Intéressante d'un point de vue fondamental, cette étude peut également avoir des retombées importantes pour la compréhension de phénomènes biologiques et des applications techniques notamment pour le séquençage de génomes. Des outils théoriques de physique statistique et numériques de dynamique moléculaire seront mis en œuvre afin de comprendre le rôle et l'importance de la structure du polymère et des forces appliquées sur le temps nécessaire à la translocation. De récentes expériences utilisant des pores dans une mono-couche de graphène ont produit des résultats encourageants qui nous poussent à concentrer notre étude sur la compréhension et la modélisation de ces nouveaux pores.

Mots clés : *ADN, Graphène, Dynamique moléculaire, Physique statistique.*

Stage encadré par :

Dr Arnaud Buhot

arnaud.buhot@cea.fr / tel : ++33 438 78 38 68

Groupe Théorie

Structures et Propriétés d'Architectures Moléculaires

Institut Nanoscience et Cryogénie

17 rue des Martyrs 38054 Grenoble cedex 9 France

http://inac.cea.fr/spram/Phoce/Vie_des_labos/Ast/ast_groupe.php?id_groupe=397

Introduction

La translocation de biomolécules à travers un nanopore est un domaine scientifique très actif tant sur le plan expérimental que sur le plan théorique. Ceci est principalement dû aux applications potentielles en biotechnologies et en médecine, notamment le séquençage rapide et économique envisagé lors de la translocation de brins d'ADN.

L'aspect appliqué est important mais ce problème soulève aussi des questions fondamentales très intéressantes sur le comportement de polymères quand ils sont confinés.

De récentes expériences de translocation mettent en œuvre des nanopores creusés dans des couches monoatomique de graphène. Le graphène, isolé pour la première fois en 2004, est un cristal aromatique bidimensionnel de carbone dont l'empilement de plusieurs plans forme le graphite. Ces expériences contournent les problèmes d'épaisseur inhérents aux autres types de pores.

Le but de mon stage est de modéliser la translocation d'ADN simple brin à travers un nanopore de graphène en caractérisant le temps de translocation, l'influence des propriétés du graphène et discuter de l'éventuelle possibilité de séquençage.

Notre étude théorique de la translocation de biomolécules à travers un nanopore s'articulera autour d'un premier chapitre présentant les outils analytiques de physique statistique mis à contribution, d'un second chapitre portant sur la modélisation numérique de dynamique moléculaire, et d'un dernier chapitre traitant de résultats expérimentaux et de l'introduction du graphène.

Table des matières

1 Outils analytiques	3
1.1 Polymères idéaux et non idéaux	3
1.2 Éléments dynamiques	4
1.3 Translocation	5
2 Dynamique moléculaire	7
2.1 Le système biologique	7
2.2 La modélisation	8
3 Expériences et utilisation du graphène	10
3.1 Courant de translocation et graphène	10
3.2 Vers la modélisation du graphène	10
4 Conclusion	11

1 Outils analytiques

Dans cette partie nous allons développer les outils analytiques de physique statistique nécessaires pour aborder le comportement des polymères et ce notamment lors de la translocation.

1.1 Polymères idéaux et non idéaux

Nous allons dans un premier temps présenter un modèle naïf de polymère idéal présenté dans la thèse de Henk Vocks [?] afin d'avoir une idée générale de la situation et de reposer quelques bases de physique statistique.

Le polymère évolue sur un réseau périodique carré de paramètre λ . La tête du polymère est placée sur un des nœuds du réseau. Les monomères consécutifs sont placés sur un des z ($z = 4$ pour un réseau carré) sites plus proches voisins, le polymère réalise alors une marche aléatoire (Figure 1).

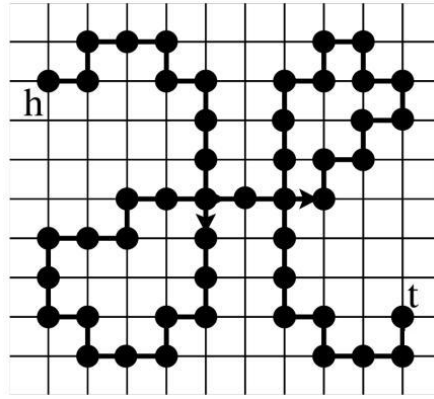


FIGURE 1 – Réseau carré montrant la marche aléatoire du polymère.

On déduit facilement que pour un degré de polymérisation $N + 1$, il y a $Z_{ideal} = z^N$ polymères distincts possibles. Rappelons des résultats concernant la marche aléatoire : soit \mathbf{r} le vecteur position du marcheur par rapport à l'origine, $\langle \mathbf{r} \rangle = 0$ et $\langle \mathbf{r}^2 \rangle = d^2 t$ avec d la taille d'un pas et t le nombre de pas effectués. Ici $t = N$ d'où une taille moyenne du polymère $R_0 = \lambda N^{\frac{1}{2}}$.

Afin de trouver l'énergie libre correspondant à chaque conformation, on établit une équation différentielle sur la probabilité, $P(\mathbf{r}, N)$, d'un monomère d'être en \mathbf{r} , car en introduisant \mathbf{b}_i les z vecteur plus proches de la position \mathbf{r} :

$$P(\mathbf{r}, N) = \frac{1}{z} \sum_{i=1}^z (P(\mathbf{r} - \mathbf{b}_i, N - 1)) \quad (1)$$

L'équation (1) est linéarisée, l'équation différentielle obtenue est résolue et l'énergie libre est donc :

$$F(r) = -K_B T \ln(P(r, N)) = F(0) + \frac{3K_B T r^2}{2R_0^2} \quad (2)$$

Avec K_B la constante de Boltzmann et T la température thermodynamique (voir la thèse de Vocks [?]).

Bien entendu ce modèle est simpliste, un vrai polymère ne peut en aucun s'intersecter lui même, les monomères peuvent être différents, ou encore des forces peuvent être appliquées.

Le lecteur intéressé trouvera un exemple de polymère non idéal développé dans la thèse de Vocks [?], exemple incluant la condition de non intersection (self avoiding walk) ainsi que des interactions répulsives entre monomères pour déduire une forme différente de l'énergie libre ainsi qu'une taille dépendant maintenant de l'exposant de Fleury (N^ν avec $\nu = \frac{3}{5}$) et non plus de $N^{\frac{1}{2}}$.

1.2 Éléments dynamiques

Nous allons maintenant proposer (comme dans la thèse de Vocks [?]) un modèle permettant d'aborder la dynamique d'une chaîne. Il s'agit du modèle de Rouse (Figure 2).

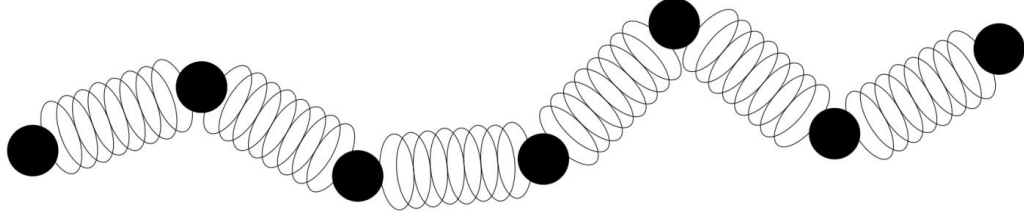


FIGURE 2 – Modélisation du polymère par une chaîne d'oscillateurs.

Le polymère est modélisé par une chaîne d'oscillateurs (de type billes/ressorts). Physiquement, le ressort ne représente pas une liaison entre monomères mais plutôt une partie du polymère assez longue pour que la statistique gaussienne puisse lui être appliquée. De l'équation (2) obtenue précédemment on déduit l'énergie élastique du ressort connectant les billes n et $n + 1$ séparées par une distance d'équilibre moyenne λ :

$$F_{n,n+1} = \frac{3K_B T}{2} \frac{(\mathbf{r}_{n+1} - \mathbf{r}_n)^2}{\lambda^2} \quad (3)$$

L'énergie élastique totale du polymère est alors la somme des contributions séparées :

$$F_{tot} = \sum_{n=0}^{N-1} F_{n,n+1} \quad (4)$$

En dérivant cette énergie, on obtient la force élastique exercée sur une bille. Le polymère évolue dans un solvant, il y aura donc une force de frottements visqueux et de plus le mouvement brownien va aussi devoir être pris en compte. L'équation du mouvement est donc la suivante :

$$m \frac{d^2 \mathbf{r}_n}{dt^2} = - \frac{\partial F_{tot}}{\partial \mathbf{r}_n} - \epsilon \frac{d \mathbf{r}_n}{dt} + \epsilon \mathbf{g}_n \quad (5)$$

Avec ϵ le coefficient de frottement et $\epsilon \mathbf{g}_n$ la contribution brownienne, une force de moyenne nulle et de variance $2K_B T$.

Comme nous nous situons à une échelle nanométrique, le nombre de Reynolds¹ de l'écoulement va être très faible, ce qui permet de négliger les termes inertiels et d'obtenir ainsi l'équation de Langevin, résultat de l'équilibre des forces avec le Stokes drag :

$$\frac{d \mathbf{r}_n}{dt} = - \frac{1}{\epsilon} \frac{\partial F_{tot}}{\partial \mathbf{r}_n} + \mathbf{g}_n \quad (6)$$

Cette équation est résolue en faisant passer n de valeurs discrètes à des valeurs continues (détails dans la thèse de Vocks [?]). Il faut principalement retenir que le déplacement carré moyen du centre de masse :

$$\langle (\mathbf{r}_{CM}(t) - \mathbf{r}_{CM}(0))^2 \rangle = \frac{6K_B T}{N\epsilon} t = 6D_R t \quad (7)$$

1. Le nombre de Reynolds est le rapport des échelles de temps caractéristiques inertielles et visqueuses

Et le fait que le coefficient de diffusion D_R soit proportionnel à N^{-1} . Pour un monomère, le déplacement carré moyen à temps long est la somme de celui du centre de masse et d'une constante multipliée par N , à temps court il croît en $t^{\frac{1}{2}}$.

Il est intéressant de noter que le temps de relaxation nécessaire au centre de masse pour parcourir sa longueur est alors proportionnel à N^2 (ou $N^{1+2\nu}$ dans le cas non idéal). Ce qui n'est pas observé expérimentalement à cause des interactions hydrodynamiques.

1.3 Translocation

La translocation d'un polymère est son passage d'un compartiment cloisonné à un autre par un pore les reliant. La molécule passe d'un coté appelé cis à l'autre appelé trans en deux temps. Le premier temps consiste en ce qu'une extrémité de la chaîne atteigne l'entrée du pore. Cette probabilité et le temps qui lui est associé peuvent être estimés analytiquement et ne présentent pas de défi [?]. Le second temps représente la translocation en elle même (voir Figure 3), il s'agit du temps pendant lequel le polymère occupe le pore.

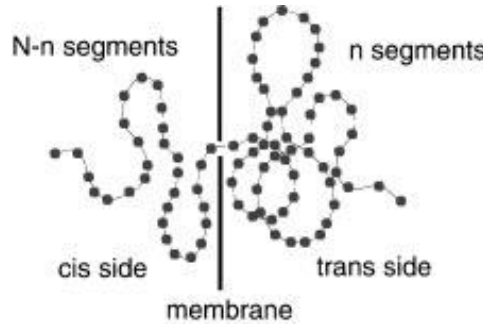


FIGURE 3 – Un polymère en cours de translocation.

Ce second temps, appelé temps de translocation (τ) est très important puisqu'il représente dans beaucoup de cas une des seules mesures expérimentales accessibles.

La translocation peut se dérouler dans différents contextes. Elle peut être due uniquement aux fluctuations thermiques, on parle alors de translocation non biaisée, ou être pilotée par une force extérieure, on parle alors de translocation forcée (driven translocation). Les forces extérieures peuvent être de natures variées, on citera notamment l'utilisation de champs électriques (électrophorèse), de gradients de potentiels chimiques, de flux imposé sur le solvant, ou encore l'emploi de pinces optiques.

Le temps de translocation dépend alors fortement de la taille du polymère et de la force appliquée, d'une façon générale on a :

$$\tau = N^\alpha f^{-\delta} \quad (8)$$

Il faut alors déterminer α et δ .

La première étape consiste à déterminer la forme de la barrière énergétique à franchir au cours de la translocation. Deux démonstrations intéressantes sont présentées respectivement dans la thèse de Vocks [?] et dans un article de Sung et Park [?].

Dans le premier exemple, Vock invoque la modification de la fonction de partition due aux contraintes topologiques imposées par la présence du pore et de la cloison :

$Z_{nonideal} = \tilde{z}^N N^{\gamma-1}$ devient $Z_{nonideal}^{cloison} = \tilde{z}^N N^{\gamma_1-1}$ La fonction de partition totale étant alors le produit

$Z_{nonideal}^{cloison}(n)Z_{nonideal}^{cloison}(N-n)$ avec n le nombre de monomères du coté trans.

L'énergie libre alors dérivée est de la forme :

$$F(N, n) = K_B T [(1 - \gamma_1) \ln[n(N - n)] - N \ln(\tilde{z})] \quad (9)$$

Dans leur article, Sung et Park [?] s'affranchissent de la contrainte du réseau et dérivent une forme similaire uniquement avec des probabilités (le terme $N \ln(\tilde{z})$ est alors défini comme une constante puisqu'il modifie la hauteur de la barrière énergétique mais pas sa forme vis à vis de la coordonnée réactionnelle).

A cette énergie libre, appelée barrière entropique (la barrière est due uniquement aux restrictions topologiques et n'a pas de contribution enthalpique, d'où le terme entropique) peuvent s'ajouter des potentiels dus aux forces appliquées, la Figure 4 présente la forme de la barrière avec différents potentiels ajoutés :

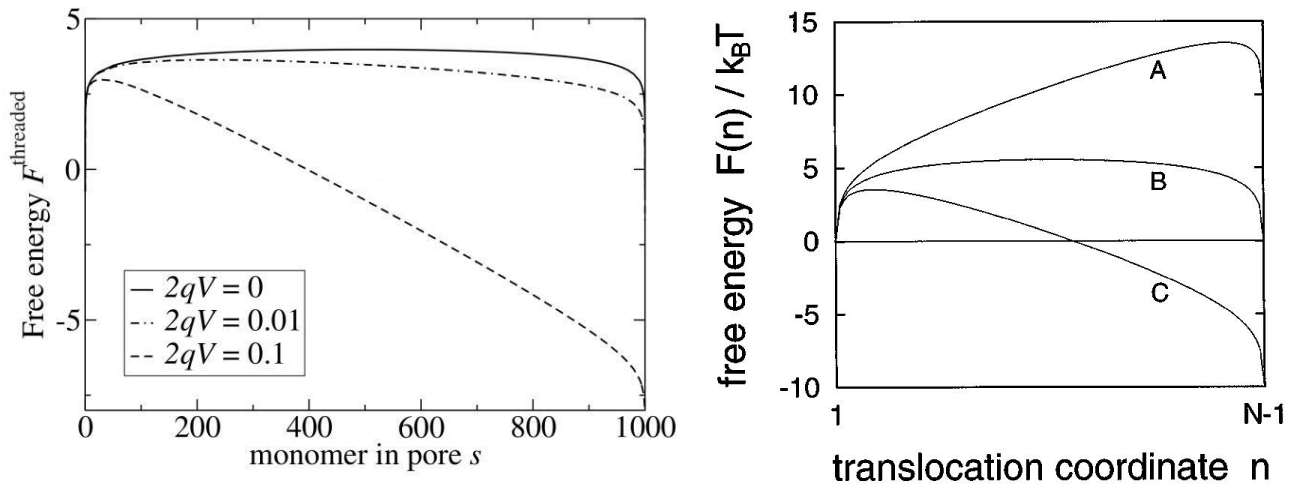


FIGURE 4 – A gauche : Influence d'un potentiel électrique sur la barrière énergétique [?]. A droite : utilisation d'une différence de potentiel chimique (A et C) [?].

Une fois les contraintes énergétiques connues, la translocation est traitée comme un processus de diffusion aléatoire gouverné par l'équation de Fokker-Planck. Il s'agit d'une équation différentielle sur les probabilités, le développement des calculs est réalisé dans l'article de Sung et Park [?] ainsi que dans l'article de review de Milchev [?]. Pour en comprendre vraiment l'origine, il est conseillé de se plonger dans un ouvrage complet de physique statistique.

On retiendra le résultat général pour la translocation non biaisée : τ est proportionnel à $N^{1+2\nu}$. Les valeurs relatives de potentiel introduit par une force et des fluctuations thermiques (de l'ordre de $K_B T$) va influencer sur les valeurs de α et δ [?, ?].

Il est également intéressant de noter l'influence des interactions hydrodynamiques présentées dans la review de Milchev [?]. La dernière partie de l'article de Sung et Park [?] présente l'influence purement entropique des protéines chaperones lors de la translocation.

On notera également l'originalité d'un article publié par Arnold J Storm and co-authors [?] qui présente deux forces (champ électrique et Stokes drag) opposées lors de la translocation.

2 Dynamique moléculaire

Dans le cadre de mon projet de stage, nous envisageons une approche assez différente des simulations classiques sur réseau avec maillage comme c'est souvent le cas en mécanique des fluides ou dans certains travaux sur les polymères comme dans la thèse de Vocks [?]. Nous proposons de modéliser les différentes parties du polymère, les champs de force et intégrer numériquement les équations du mouvement, c'est ce qu'on appelle faire de la dynamique moléculaire.

2.1 Le système biologique

Il est primordiale afin de décrire au mieux la réalité physique, de comprendre ce qui constitue le système étudié : l'ADN.

L'ADN est le support de l'information génétique du vivant. Il s'agit d'un bio-polymère constitué d'une séquence de quatre monomères différents. Ces monomères, appelés nucléotides sont constitués de phosphate, de sucre et d'une base azotée, seul élément distinct entre nucléotides : l'adénine, la cytosine, la guanine et la thymine (l'uracile présente uniquement en substitut de la thymine dans l'ARN ne sera pas abordée ici). Il se trouve dans les cellules, contenu généralement dans le noyau. La translocation de l'ADN est le mécanisme impliqué dans nombre d'infections virales et pourrait être utilisé pour la thérapie génique.

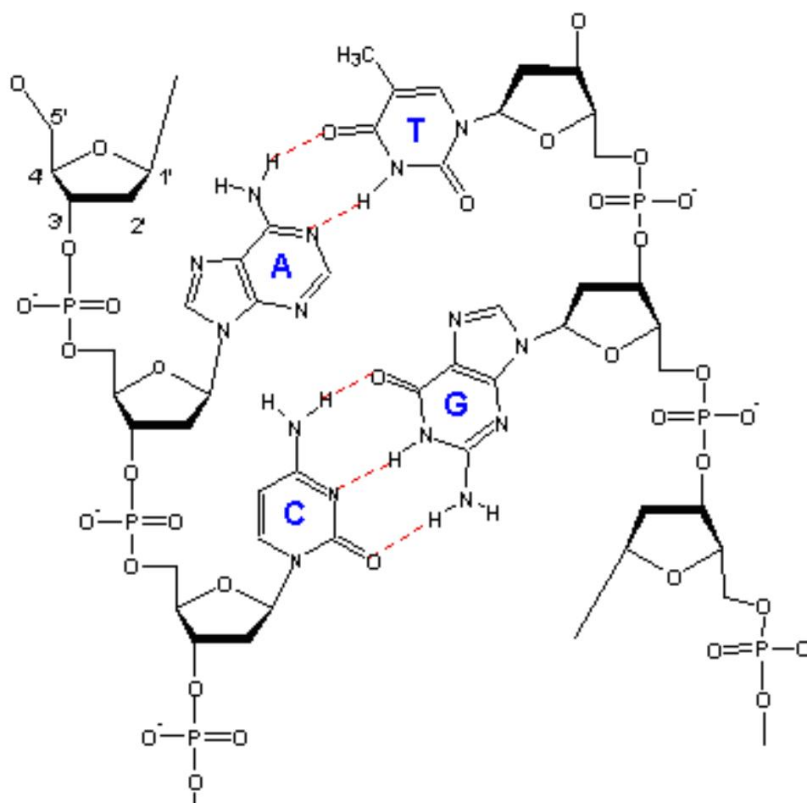


FIGURE 5 – Structure chimique de l'ADN. Le squelette est composé d'une succession de phosphates et de sucres, à chaque sucre est liée une base azotée. Les bases azotées peuvent interagir par liaisons hydrogènes.

La structure chimique de l'ADN est aujourd'hui parfaitement connue (Figure 5), la nature des liaisons et interactions ne sont plus mystérieuses. Numériquement, il est impossible de faire converger les calculs en modélisant chaque atome et chaque interaction, la difficulté de modélisation sera de réduire le modèle sans exagérations pour que les calculs soient réalisables et conservent un sens physique.

2.2 La modélisation

Afin de créer un modèle numérique fiable et robuste, il convient de décomposer chaque monomère en sous système. Il semble naturel de faire une décomposition en sucre, phosphate et base azotée comme l'on fait Margaret C Linak and co-authors [?]. Leur résultats étant fiables, nous envisageons de nous inspirer fortement de la manière dont ils ont modélisé l'ADN simple brin (Figure 6).

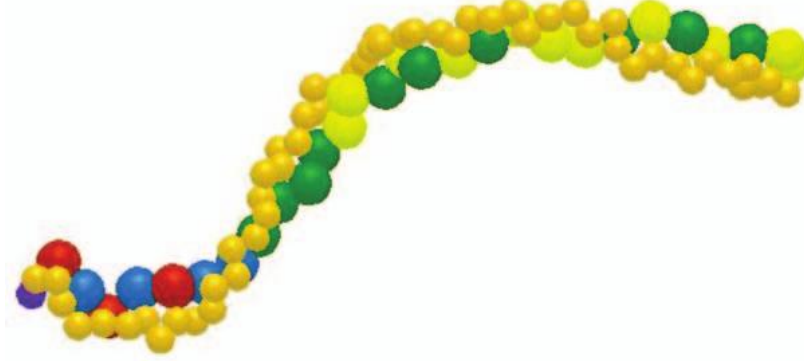


FIGURE 6 – Chaîne de nucléotides avec trois composants : sucres, phosphates et bases azotées[?].

Une fois ce découpage préparé, l'étape suivante consiste à définir tous les potentiels d'interaction entre tous les éléments.

Dans leur papier, Margaret C Linak and co-authors [?], définissent plusieurs types de potentiels à partir d'étalons de distance (σ , R_0 et γ) et d'énergie (ϵ) :

Les interactions non spécifiques :

Il s'agit d'une part des interactions de type volume exclu, modélisées par un potentiel de Lennard-Jones :

$$U_{EV}(r_{ij}) = 4\epsilon\left[\left(\frac{\gamma}{r_{ij}}\right)^{12} - \left(\frac{\gamma}{r_{ij}}\right)^6\right] + \epsilon \quad (10)$$

r_{ij} étant la distance entre les billes i et j , ce potentiel empêche l'interpénétration à faible distance et est équivalent aux interactions de Van der Waals à longues distances.

Et d'autre part de la modélisation des liaisons covalentes par un potentiel harmonique modifié :

$$U_{MH}(r_{ij}) = -15\epsilon\left(\frac{R_0}{\sigma}\right)^2 \ln\left[1 - \left(\frac{r_{ij}}{R_0}\right)^2\right] \quad (11)$$

Modèle de liaison de type ressort un peu plus poussé.

La rigidité du squelette du polymère :

La rigidité angulaire du polymère est traduite par un potentiel fonction de la torsion :

$$U_{BB}(\phi) = 12\epsilon(1 + \cos(\phi))^2 \quad (12)$$

ϕ étant l'angle formé localement par le squelette, au repos $\phi = \pi$ et $U_{BB} = 0$.

Les interactions entre bases :

Les bases interagissent entre elles en formant des liaisons hydrogènes, il y en a plusieurs types définies dans l'article [?] :

$$U_k(r_{ij}, \theta) = -\delta_k^{ij} f_k(\theta) \epsilon [\exp(20 \frac{r_{ij}}{\sigma} - 30) + 1]^{-1} \quad (13)$$

Avec $f_k(\theta)$ une fonction du type d'interaction k et d'un angle et δ_k^{ij} représente la force d'un lien de type k entre i et j .

La Figure 7 récapitule toutes les interactions :

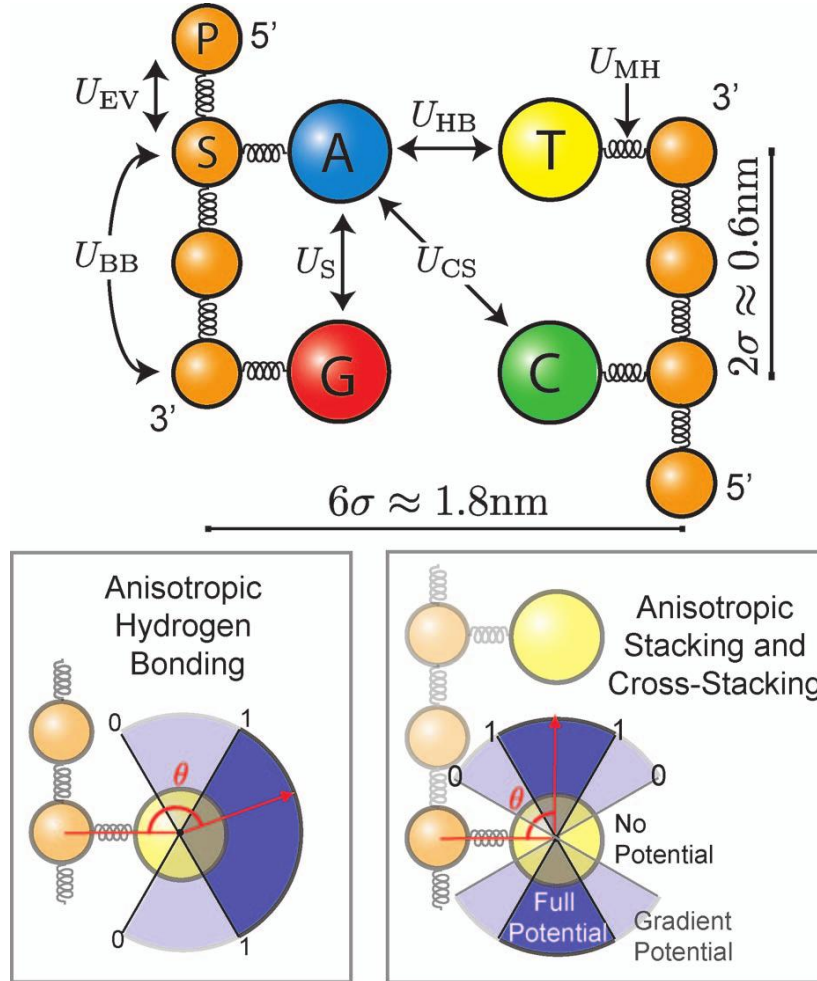


FIGURE 7 – Les différentes interactions modélisées

Une fois ces potentiels calculés, ils sont sommés et dérivés afin d'être introduits dans l'équation de Langevin (7) qui est ensuite numériquement résolue.

3 Expériences et utilisation du graphène

Dans cette dernière partie, nous allons présenter le contexte expérimental dans lequel notre étude va se situer, les changements apportés par le graphène et la façon dont nous allons les modéliser.

3.1 Courant de translocation et graphène

La translocation d'ADN n'est pas un phénomène nouveau, et l'idée même de pouvoir s'en servir à des fins de séquençage a plus d'une quinzaine d'années [?]. L'idée est de mesurer le courant ionique à travers le pore lors de la translocation. Ce courant dépend de l'occupation ou non et de la nature de l'occupant du pore.

Dans les travaux précédents utilisant des bio-pores [?], la taille du pore était un obstacle majeur. En effet, l'épaisseur du pore ayant la taille d'une centaine de paire de bases (environ 30 nm), le courant mesuré ne permettait pas de distinguer les séquences les unes des autres. Néanmoins la nature et la conformation de l'objet en translocation pouvaient être déterminées (protéine, ADN, ARN ...).

La récente utilisation du graphène permet de surmonter ce problème d'épaisseur puisqu'une mono-couche de graphène ne mesure que 0,3 nm.

La Figure 8 représente une vue d'artiste d'un système de séquençage utilisant le graphène :

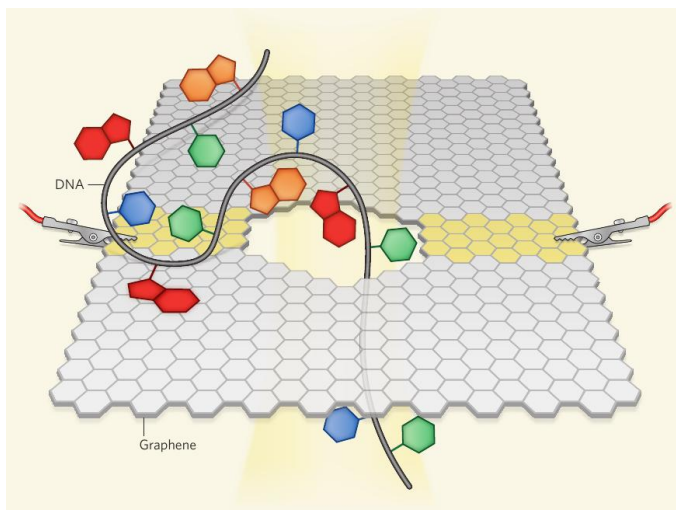


FIGURE 8 – Vers l'utilisation du graphène pour séquencer l'ADN en mesurant le courant de translocation.

3.2 Vers la modélisation du graphène

Nous venons de voir que le graphène présente un potentiel intéressant afin de séquencer l'ADN, il faut maintenant s'interroger sur la façon dont nous allons le modéliser.

La première partie consiste à modéliser le graphène en lui même, faudra-t-il modéliser chaque atome, chaque cycle de six atomes, ou des ensembles de cycles ? Puisque la flexibilité du plan de graphène sera un élément clé, cette étape ne doit pas être survolée.

La seconde partie de la modélisation va consister en la description des interactions, au sein du graphène et avec les différents éléments du polymère. Il faudra notamment prendre en compte les interactions hydrophobe, d'éventuelles interactions électrostatiques, voir même des interactions hydrodynamiques. Toutes ces interactions vont demander beaucoup de ressources informatiques et il faudra donc étudier leurs importances relatives afin de ne pas surcharger le modèle (une réflexion aura lieu sur la nécessité de prendre en compte les liaisons hydrogènes par exemple). On pourra également s'inspirer de travaux récents [?].

4 Conclusion

Nous avons vu, dans la première partie, le comportement théorique d'un polymère et ce notamment lors de la translocation à travers un pore. Les outils analytiques développés vont permettre d'aider au développement et d'avoir un point de vue critique sur les résultats que nous obtiendrons avec une modélisation de type dynamique moléculaire introduite en seconde partie. Nous espérons que nos résultats permettront de relever les défis envisageables grâce aux récentes avancées expérimentales présentées en dernière section.

Nous nous attacherons particulièrement à déterminer les paramètres influant sur le temps de translocation, à étudier l'influence du graphène, notamment la flexibilité du plan et à discuter d'une éventuelle possibilité de séquencer l'ADN.