

La modélisation des acides nucléiques

Sommaire

1	La structure des acides nucléiques	4
1.1	Briques élémentaires des acides nucléiques	4
1.2	Structure tridimensionnelle	6
1.3	Méthodes de détermination	11
1.4	Nouvelles perspectives	19
2	Les acides nucléiques et le modèle de tige élastique	21
2.1	Modèles de Physique des polymères	21
2.2	Essais de caractérisation mécanique des acides nucléiques	25
2.3	Structures tridimensionnelles à partir des tiges élastiques	31
3	L'approche BCE	35
3.1	Une approche multi-échelle et hiérarchique	35
3.2	Exemple : détermination d'une structure en épingle à cheveux	36
3.3	Vers une généralisation de la méthode	40
4	Objectifs de la thèse	40
4.1	Pour la modélisation des acides nucléiques	40
4.2	Sur une perspective plus large	41

1 La structure des acides nucléiques

Les acides nucléiques sont des molécules biologiques au cœur des mécanismes de la vie, situées dans les cellules des organismes procaryotes et eucaryotes ainsi que dans les capsides des virus. L'acide désoxyribonucléique (ADN) est le support de l'information génétique utilisée pour le développement, le fonctionnement et la reproduction des êtres vivants. L'acide ribonucléique (ARN) participe à l'expression de cette information, c'est-à-dire à la construction de protéines à partir du modèle codé dans l'ADN. Certains virus, cependant, ne possèdent pas d'ADN et stockent leur matériel génétique directement sous forme d'ARN.

Connaître précisément la structure de ces molécules est indispensable pour comprendre leur fonctionnement, et lutter contre les maladies génétiques ou le développement de virus comme le VIH. Or, les acides nucléiques présentent une grande variété structurale. Leur composition chimique ainsi que leur formule développée sont bien identifiées et seront présentées en sous-section 1.1. Toutefois, leur structure tridimensionnelle, *i.e.* l'arrangement et le positionnement des atomes dans l'espace, est moins complètement déterminée. Elle est l'objet d'une diversité significative énoncée en sous-section 1.2. Quoique largement étudiée avec différentes méthodes prometteuses, la détermination de la structure 3D des acides nucléiques présente toujours des enjeux, qui seront discutés en sous-section 1.3.

1.1 Briques élémentaires des acides nucléiques

Les acides nucléiques sont des polymères, c'est-à-dire de longues chaînes moléculaires constituées par la répétition de nombreuses sous-unités appelées nucléotides. La plus longue molécule d'ADN du corps humain en contient environ 220 millions (Gregory *et al.* 2006), pour une longueur rectiligne supérieure à 7 cm. Chaque nucléotide se compose de trois parties : un phosphate PO_4^{3-} , un sucre à cinq atomes de carbone $C_5H_{10}O_5$ (pentose) lié au phosphate par une liaison ester, et une base azotée liée au sucre par une liaison glycosidique (Figure 1.1a). L'ADN diffère de l'ARN par l'absence d'un atome d'oxygène en liaison avec le carbone $C2'$ du sucre, qui lui vaut le préfixe "désoxy".

Les bases azotées sont les éléments fondamentaux du code génétique. Elles existent dans l'ADN sous quatre formes différentes : l'Adénine (A), la Guanine (G), la Cytosine (C) et la Thymine (T). On les retrouve également dans l'ARN, à l'exception de la Thymine qui est remplacée par l'Uracile (U). C'est l'ordre d'enchaînement de ces bases, ou *séquence*, qui code l'information génétique. C'est pourquoi elles sont souvent qualifiées de "briques élémentaires" des acides nucléiques. Les bases de l'ADN sont regroupées en "paquets", les *gènes*, chacun spécialisé dans le codage d'une ou plusieurs protéines. Chez les eucaryotes, les gènes sont en grande partie constitués d'introns, des séquences d'ADN ne codant généralement pas de protéine mais utilisées pour la régulation, l'organisation et la maintenance du génome.

L'enchaînement des nucléotides de l'ADN et de l'ARN se produit par l'intermédiaire des phosphates, qui forment une liaison ester avec le carbone $C3'$ d'un sucre et une autre liaison ester avec le carbone $C5'$ du sucre suivant (liaison phosphodiester). Cela constitue la chaîne *sucre-phosphate*, aussi appelée squelette de la molécule, sur laquelle les bases azotées sont accrochées. Cette chaîne est directionnelle : elle a une extrémité $5'$ et une extrémité $3'$, dont les atomes $C5'$ et,

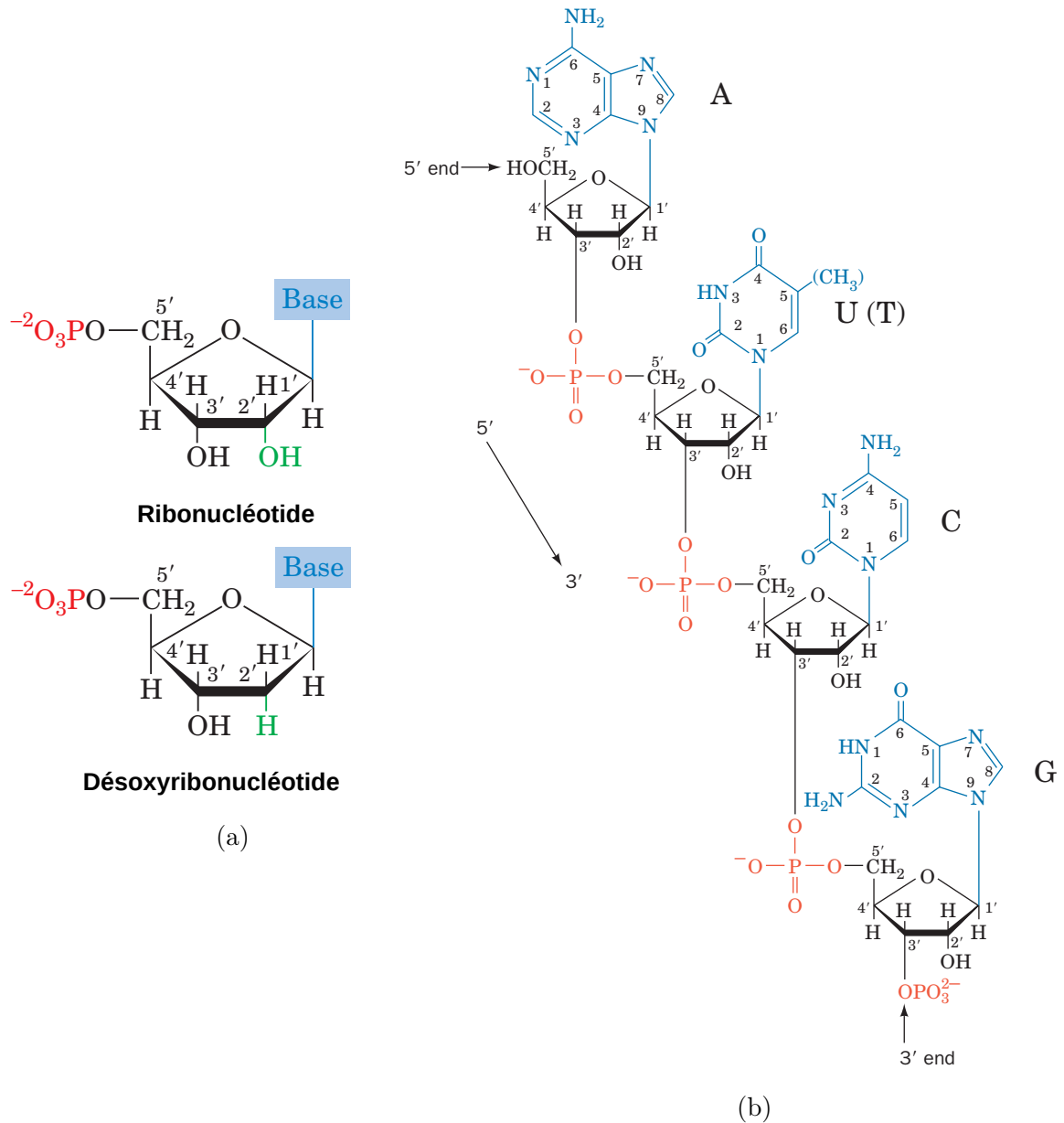


Figure 1.1. Formule développée des acides nucléiques : (a) nucléotides d'ARN (ribonucléotide) et d'ADN (désoxyribonucléotide) et (b) exemple d'une chaîne d'ARN de séquence AUCG (ou d'une chaîne d'ADN de séquence ATCG avec les informations entre parenthèses). Schémas de Voet et Voet (2011).

respectivement, $\text{C}3'$ ne sont liés à aucun nucléotide. Par convention, on écrit un acide nucléique dans le sens $5' \rightarrow 3'$. La Figure 1.1b représente un exemple d'enchaînement des quatre bases A, U (T pour l'ADN), C et G. Les phosphates sont à l'origine de l'acidité du squelette, ainsi que de sa charge négative.

L'expression des gènes, autrement dit la construction de protéines à partir de "morceaux de séquence d'ADN", est généralement décrite chez les procaryotes et les eucaryotes par une procédure schématique appelée le *dogme central*. Une enzyme, l'ARN polymérase, reconnaît le début

du gène, le promoteur, et se fixe dessus. Elle parcourt ensuite le gène et fabrique l'ARN dit *messenger* : cette étape est appelée la *transcription*. L'ARN messenger transporte alors l'information génétique jusqu'au ribosome, un important complexe d'ARN (dit *ribosomique*) et de protéines. Le ribosome parcourt l'ARN messenger, et une protéine est construite par l'enchaînement d'acides aminés apportés par l'ARN *de transfert* : c'est la *traduction*.

1.2 Structure tridimensionnelle

La structure tridimensionnelle, ou *conformation*, est l'arrangement et le positionnement des atomes dans l'espace tridimensionnel. Elle présente, dans le cas des acides nucléiques, une telle diversité qu'il est encore difficile de la caractériser entièrement. Elle varie notamment en fonction de la séquence de la molécule et des contraintes qui lui sont imposées par son environnement physico-chimique.

1.2.1 Structure 3D de l'ADN

Double-hélice

Watson et Crick (1953) ont découvert et caractérisé la conformation d'acide nucléique qui est probablement la plus connue : la double-hélice d'ADN-B. Elle se présente sous la forme de deux brins de polynucléotides dont les chaînes sucre-phosphate s'enroulent autour d'un axe commun en formant des hélices droites d'environ 2 nm de diamètre et 3.3 nm de pas (Figure 1.2a). Ces

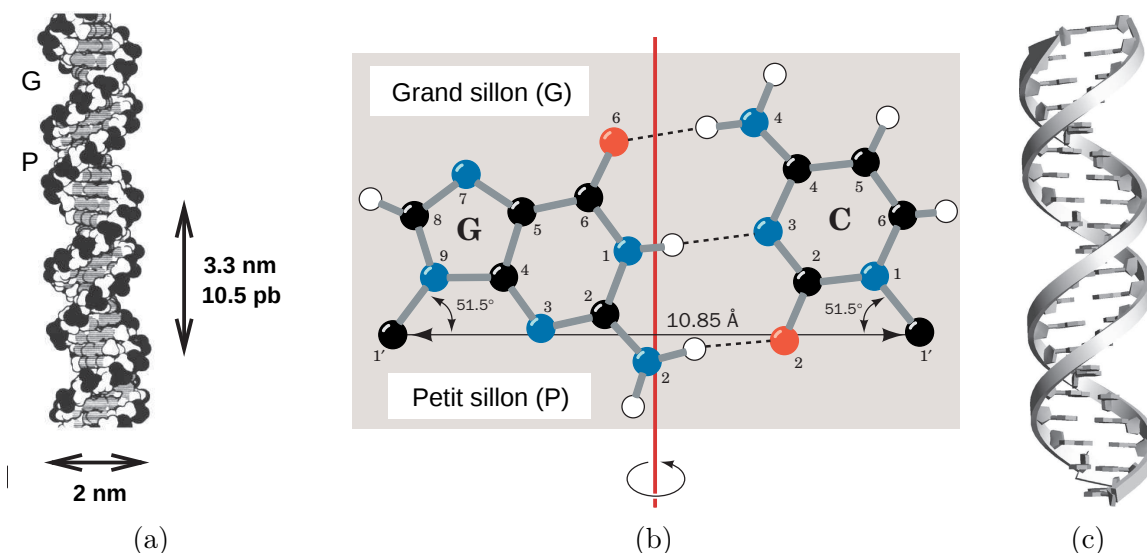


Figure 1.2. Double-hélice d'ADN-B. (a) Représentation de van der Waals, image de **Goodsell (1992)**. (b) Appariement G-C en représentation boule-bâton, schéma de **Voet et Voet (2011)**. (c) Représentation en ruban, image de **Neidle (2008)**.

deux brins sont antiparallèles : ils évoluent en sens opposés, et l'extrémité 3' de l'un est en face de l'extrémité 5' de l'autre. Ils diffèrent par une symétrie pseudo-dyadique : le squelette de l'un

peut être obtenu à partir de celui de l'autre par une rotation de 180 degrés autour d'un axe (judicieusement choisi) perpendiculaire à l'axe des hélices. Enfin, ils sont accrochés entre eux par des liaisons hydrogène entre leurs bases azotées. Les bases A et T s'apparient en formant deux liaisons hydrogène, et les bases C et G font de même avec trois liaisons hydrogène.

Les paires de bases ainsi constituées sont presque planes et perpendiculaires à l'axe des hélices. Elles sont empilées les unes sur les autres avec un décalage angulaire quasi constant d'environ 36 degrés. La double-hélice possède deux sillons extérieurs qui s'enroulent entre les chaînes sucre-phosphates. Ces sillons sont de tailles inégales : le petit sillon correspond aux bords des paires de bases où le carbone $C1'$ ressort, tandis que le grand sillon correspond au bord opposé (Figure 1.2b). Pour visualiser plus facilement la structure de cette double-hélice, il est courant de représenter le squelette sucre-phosphate par un ruban qui décrit sa trajectoire et les sucres et bases par des blocs (Figure 1.2c).

Cependant, l'ADN-B n'est pas la seule structure en double-hélice. Il en existe d'autres, comme l'ADN-A qui présente une différence de taille plus marquée entre les sillons, et des paires de bases inclinées de 19 degrés par rapport au plan perpendiculaire à l'axe des hélices. Un autre exemple, l'ADN-Z, fait intervenir des hélices gauches avec des sillons de tailles équivalentes.

Autres conformations

Bien que la double-hélice soit probablement la structure la plus fréquente de l'ADN, il en existe beaucoup d'autres, qui jouent des rôles fondamentaux dans divers processus biologiques.

Premièrement, l'ADN simple-brin, constitué d'une seule chaîne de nucléotides, apparaît comme structure transitoire lors de processus tels que la transcription, la réplication (division cellulaire) ou la réparation. Certains virus possèdent uniquement cette structure d'ADN, généralement sous forme circulaire, *i.e.* avec les extrémités qui se rejoignent.

Ensuite, l'ADN peut générer des structures tige-boucle, dites en épingle à cheveux (Figure 1.3a). Il s'agit de repliements d'un seul brin sur lui-même, avec une partie en double-hélice (la tige) refermée par une partie simple-brin (la boucle) comportant quelques bases non appariées ou mésappariées. Pour se former, ces structures requièrent la présence de séquences *inversées répétées* (ou séquences *palindromiques*), c'est-à-dire répétées à deux endroits de la molécule avec des sens opposés. Les structures en épingle à cheveux sont essentielles dans de nombreux processus biologiques, notamment la réplication de l'ADN et l'expression des gènes. Dans certains cas, elles peuvent perturber la transmission du matériel génétique et induire des maladies telles que la dystrophie musculaire de Duchenne ou la chorée de Huntington. Elles ont cependant des applications thérapeutiques intéressantes, notamment pour l'inhibition de la production de protéines du VIH (Boiziau *et al.* 1999) et la répression des facteurs de croissance de la DMLA au moyen d'aptamères. Plusieurs structures d'épingles à cheveux sont possibles, notamment liées au positionnement des bases de la boucle (Figure 1.3b).

Pour l'ADN double-brin, d'autres appariements entre bases que ceux définis par Watson et Crick peuvent exister : on parle alors de mésappariement. Il peut s'agir de liaisons hydrogène faisant intervenir des atomes différents, comme dans l'appariement décrit par Hoogsteen (1963) entre l'adénine et la thymine. Des appariements G·A, G·T, A·A, G·G, T·T, C·C, T·C ou A·C

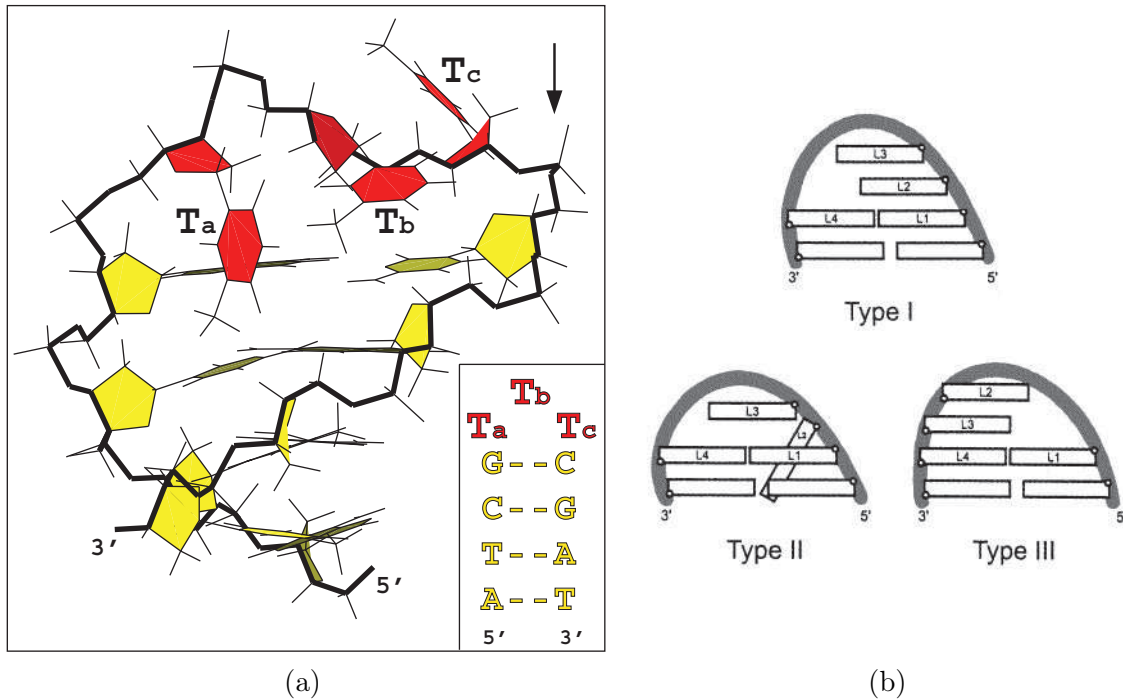


Figure 1.3. Structures en épingle à cheveux d'ADN. (a) Séquence ATCGTTTCGAT, figure de Pakleza (2002). (b) Les trois types de repliement d'une structure avec quatre bases dans la boucle : empilement des bases d'un côté ou de l'autre de la boucle (types I et III) ou positionnement de la deuxième base dans le petit sillon de la tige (type II). Figure de van Dongen *et al.* (1996).

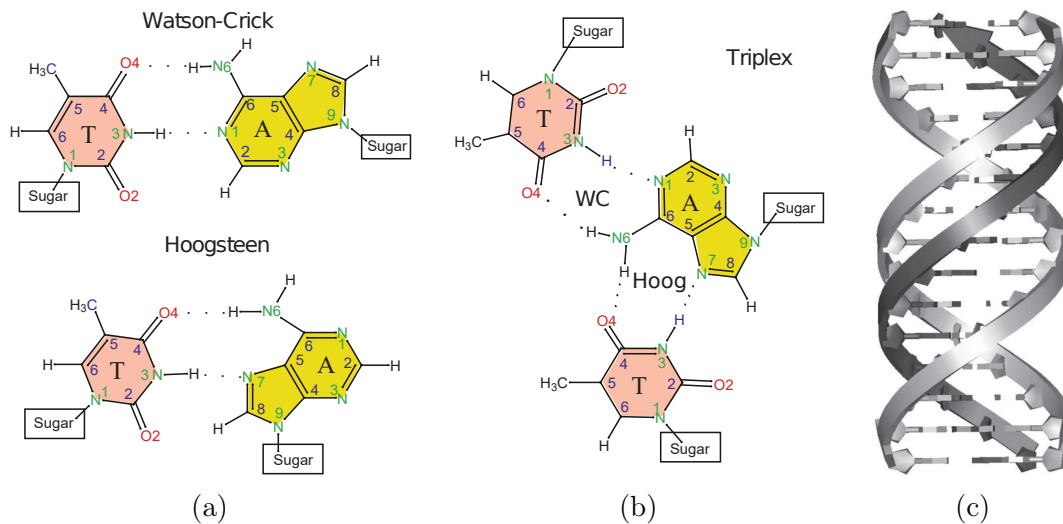


Figure 1.4. Triple-hélice d'ADN. (a) Comparaison entre un appariement A·T de type Watson-Crick et un mésappariement de type Hoogsteen ; schéma de Schlick (2010). (b) Appariements T·A·T dans un triplex, schéma de Schlick (2010). (c) Représentation d'un triplex en ruban, image de Neidle (2008).

peuvent également se produire. Dans chaque cas, les bases sont soit dans leur conformation normale (anti), soit dans leur conformation retournée (syn). Enfin, il arrive parfois qu'une des deux

bases soit absente (site abasique). Ces structures non canoniques déforment l'ADN double-brin et le déstabilisent. Elles peuvent conduire à des mutations génétiques ainsi qu'à une altération du code génétique. Les organismes disposent néanmoins de mécanismes pour les réparer.

Des triples-hélices d'ADN, ou *triplexes*, peuvent également se former (Figure 1.4). Il s'agit d'empilements de groupes de trois bases, avec un appariement Watson-Crick et un mésappariement (de type Hoogsteen par exemple). Elles apparaissent notamment lorsque de petites molécules d'ADN simple-brin, les oligonucléotides, se fixent sur une double-hélice d'ADN. Ces oligonucléotides, généralement obtenus synthétiquement, peuvent être utilisés pour inhiber l'expression de certains gènes (Grigoriev *et al.* 1993). Cela a déjà été effectué de nombreuses fois *in vitro*, notamment pour ralentir l'activité de l'oncogène *c-myc* qui favorise les cancers (Catapano *et al.* 2000) ainsi que celle des séquences *polypurine tracts* (PPT) nécessaires à la réplication du VIH (Faria *et al.* 2000).

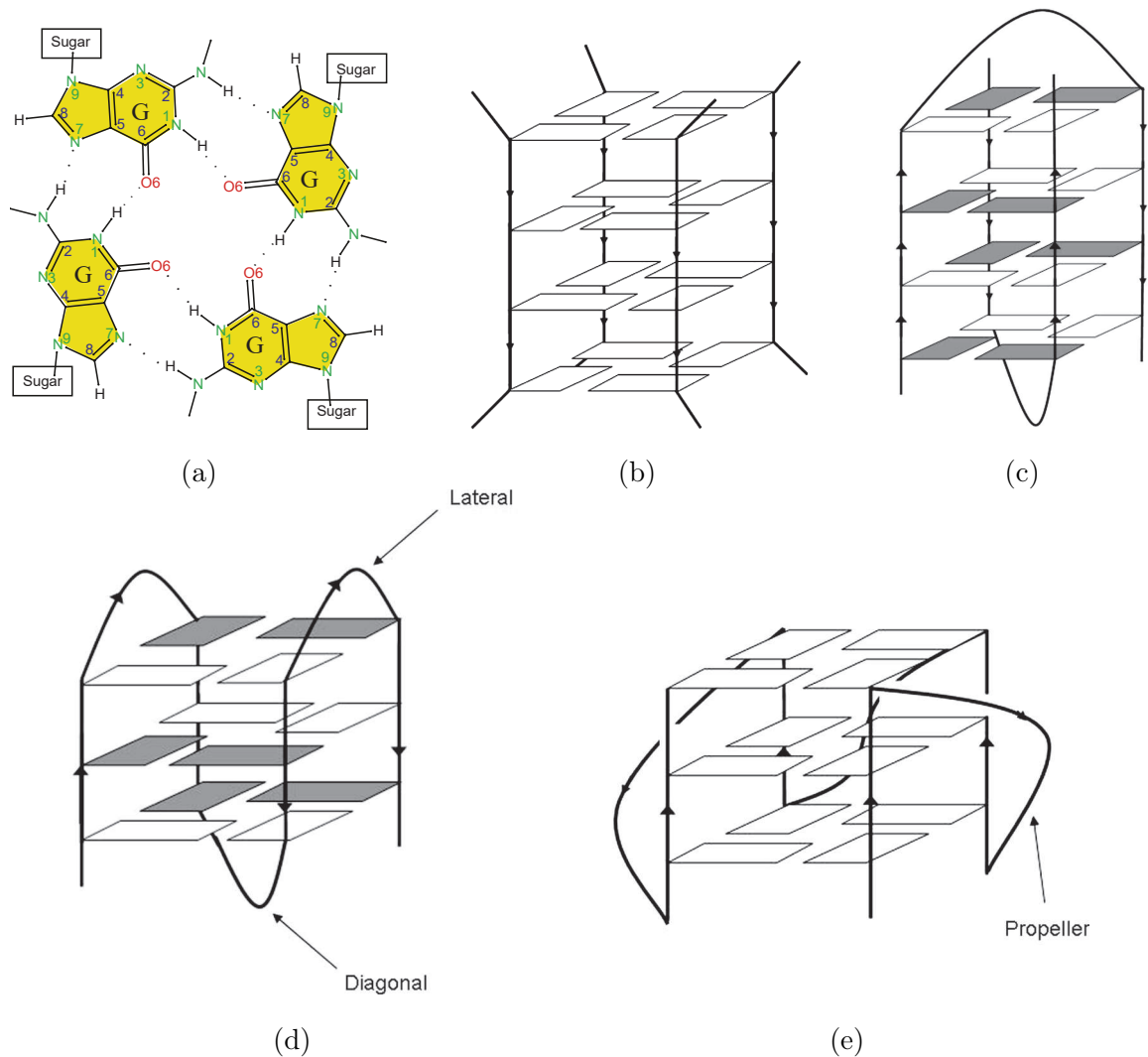


Figure 1.5. Quadruplexes d'ADN. (a) Tétrade de quatre guanines, schéma de Schlick (2010). (b-e) Représentation des différents repliements possibles d'un quadruplexe : quatre brins parallèles en (b), deux brins antiparallèles en (c), un seul brin avec des boucles latérale et diagonale en (d) et propeller en (e) ; schémas de Neidle (2008).

Enfin, il existe des structures d'ADN avec quatre brins, ou *quadruplexes*, constituées par l'empilement de plateaux de quatre guanines (tétrades) reliées entre elles par huit liaisons hydrogène (Figure 1.5a). Chez les eucaryotes, ces structures apparaissent aux extrémités des chromosomes pour les protéger. D'autre part, certains quadruplexes ont des propriétés anticancérigènes remarquables, en inhibant l'expression du gène STAT3 (Jing *et al.* 2006). Les quadruplexes existent sous de nombreuses formes (Figure 1.5b-e). Ils peuvent être générés par l'association de quatre brins d'ADN parallèles, de deux brins antiparallèles ou par le repliement d'un seul brin en formant des boucles latérales, diagonales ou "propeller".

1.2.2 Structure 3D de l'ARN

Les conformations mentionnées au paragraphe précédent existent aussi pour l'ARN (l'exemple d'une double-hélice A est donné Figure 1.6a). Toutefois, ce dernier est plus souvent sous forme d'un seul brin que l'ADN, et pour cette raison il présente en pratique une variété structurale encore plus grande. En effet, lorsque la molécule est longue, les nombreuses possibilités de repliement donnent des éléments topologiques variés avec des régions simple-brin, double-brin et différents types de boucles (Figure 1.6b). Notamment, des hernies se produisent sur les double-brin d'ARN lorsqu'un ou plusieurs nucléotides non-appariés ressortent de l'un des deux brins vers l'extérieur de la double-hélice. D'autre part, des bases non appariées sur chacun des deux brins de la molécule peuvent produire des boucles dites *internes*. Des jonctions ou *boucles multiples* se constituent également pour relier plusieurs brins entre eux. Enfin, des pseudonœuds apparaissent lorsque les nucléotides d'une boucle en épingle à cheveux s'apparient avec un autre brin en formant une structure tige-boucle supplémentaire.

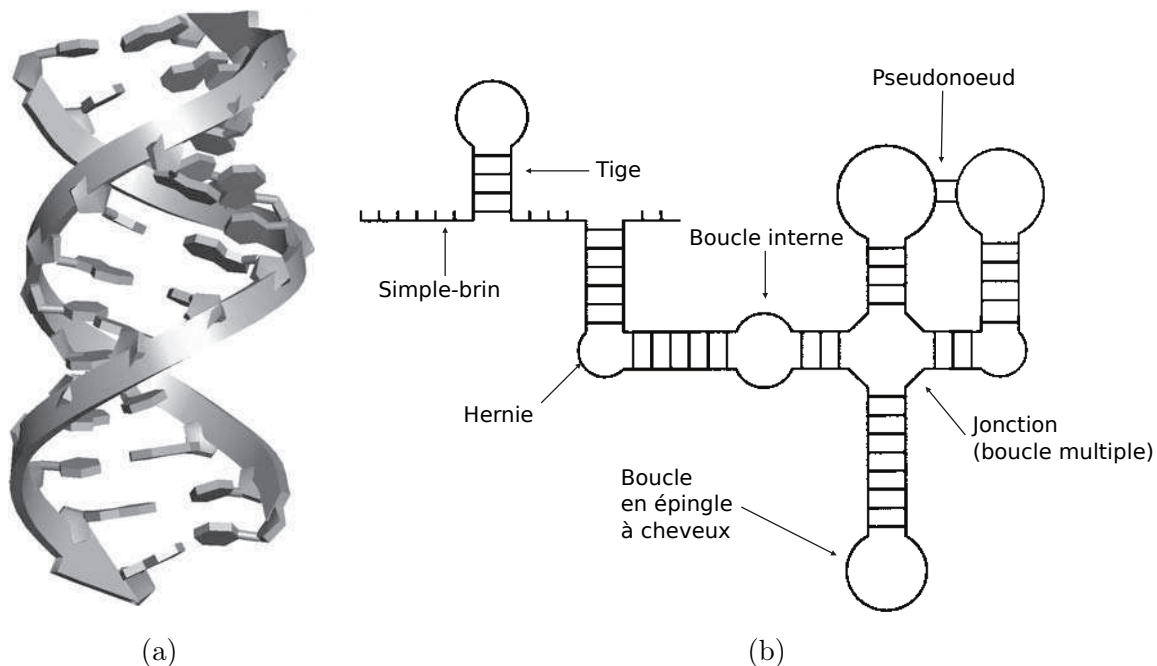


Figure 1.6. Variété structurale de l'ARN. (a) Double-hélice d'ARN-A, image de Neidle (2008). (b) Graphe des différents repliements possibles de l'ARN; figure modifiée à partir de Wuchty *et al.* (1999).

1.2.3 Diversité structurale des acides nucléiques

L'inventaire qui précède, sur les structures tridimensionnelles des acides nucléiques, n'est pas exhaustif. Il donne uniquement un aperçu de la grande diversité de ces structures. En théorie, cette diversité est supérieure à celle des protéines, dans le sens où la majeure partie de l'ADN des eucaryotes est non codante et intervient dans la régulation, l'organisation et la maintenance du génome. Or, très peu de structures d'acides nucléiques ont été déterminées en comparaison des protéines. La *Protein Data Bank* (PDB), principale source de données de biologie structurale, contient plus de 120 000 protéines contre environ 3000 acides nucléiques (Figure 1.7).

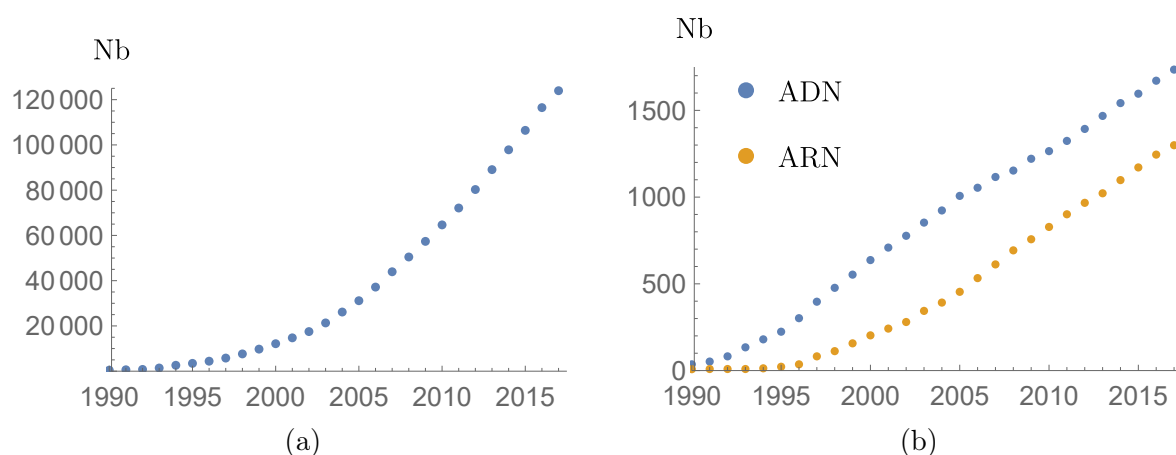


Figure 1.7. Tracé de l'évolution du nombre de structures déposées dans la PDB depuis 1990. (a) Croissance du nombre de protéines déposées, d'abord quadratique puis linéaire. (b) Croissance du nombre d'acides nucléiques déposés, plutôt linéaire donc inférieure à celle des protéines.

Il reste donc beaucoup d'études à mener sur l'ADN et l'ARN, dont la progression risque d'être freinée par les enjeux liés à la détermination de leurs conformations.

1.3 Méthodes de détermination

En effet, les conformations des acides nucléiques impliquent des échelles tellement petites qu'il est difficile de les déterminer complètement et précisément de façon expérimentale. Les méthodes théoriques posent quant à elles des difficultés liées au temps de calcul des simulations numériques.

1.3.1 Coordonnées cartésiennes et coordonnées internes

Avant de présenter ces méthodes expérimentales et théoriques, il est utile de préciser que la conformation d'une molécule peut être décrite de deux façons différentes : par les coordonnées cartésiennes des positions des atomes ou par un système de coordonnées dites *internes* (Figure 1.8). Trois types de coordonnées internes sont généralement définies : la longueur de liaison, l'angle de valence (angle entre deux liaisons consécutives) et l'angle de torsion (dans un enchaînement d'atomes A-B-C-D, il s'agit de l'angle entre le plan contenant les atomes A-B-C et

celui contenant les atomes B-C-D). Les coordonnées internes sont parfois préférées aux coordonnées cartésiennes, notamment parce que les longueurs de liaison et les angles de valence varient relativement peu.

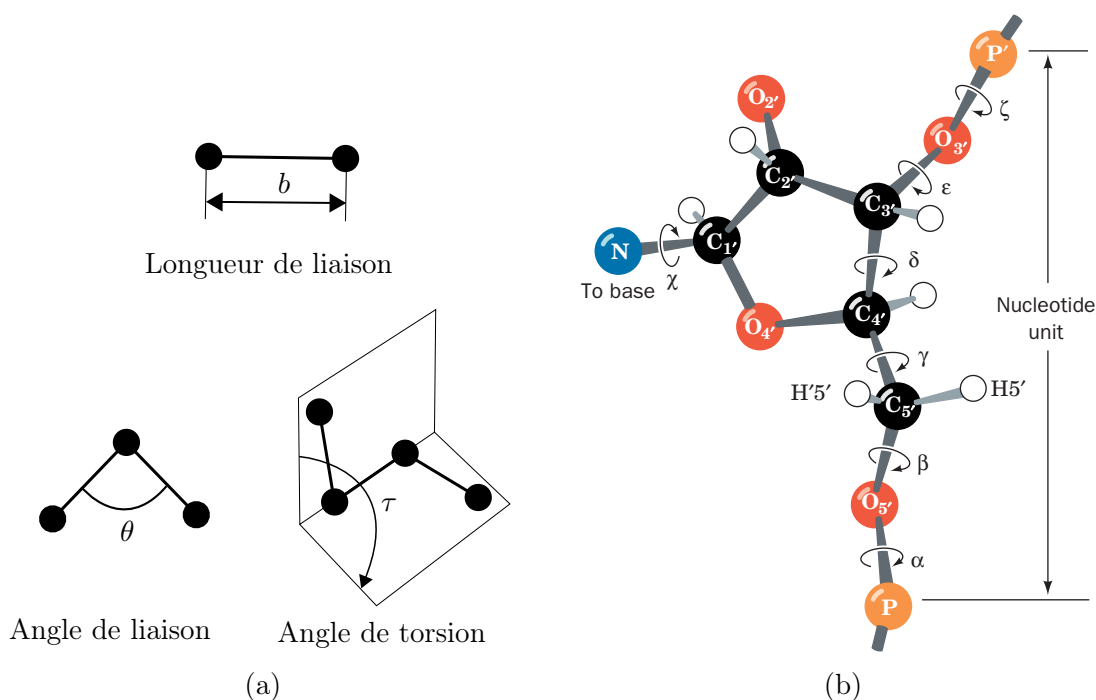


Figure 1.8. Coordonnées internes. (a) Définition. (b) Paramétrisation des angles de torsion d'un nucléotide, schéma de Voet et Voet (2011).

1.3.2 Techniques expérimentales

Les trois principales méthodes utilisées pour déterminer expérimentalement la conformation des acides nucléiques sont la cristallographie aux rayons X, la spectroscopie par résonance magnétique nucléaire (RMN) et la cryo-microscopie électronique (cryo-ME).

Cristallographie

Utilisée depuis le début du XX^e siècle, la cristallographie consiste à analyser le diagramme de diffraction obtenu en dirigeant un faisceau de rayons X sur la molécule cristallisée (Figure 1.9a). Les rayons X ont une longueur d'onde de l'ordre des liaisons atomiques (environ 1.5 Å). Lorsqu'ils rencontrent la couche électronique d'un atome de la molécule, leur trajectoire est déviée. Après avoir traversé le cristal, chaque rayon vient former une tâche sur une surface photosensible, dont la position donne l'angle de diffraction. Par une transformée de Fourier, une image tridimensionnelle de la densité électronique de la molécule est obtenue à partir de l'amplitude, la longueur d'onde et la phase de chaque rayon diffracté. Cette image donne la position 3D de chaque atome. La résolution de la méthode est fonction de la longueur d'onde λ des rayons X et de l'angle maximal θ de diffraction : elle vaut $\lambda/(2\sin\theta)$ d'après la loi de Bragg. Pour obtenir la conformation d'une

molécule à l'échelle atomique, elle doit se situer autour de 1.0 à 1.2 Å (Figure 1.9b-c). Ces dernières années, une telle résolution est devenue largement accessible avec la mise à disposition des synchrotrons pour les études de biologie structurale : il est même possible de descendre à 0.8 Å. Cependant, l'inconvénient principal de cette technique est que le réseau cristallin impose à la molécule des contraintes non naturelles, susceptibles de modifier sa structure (Dickerson *et al.* 1994).

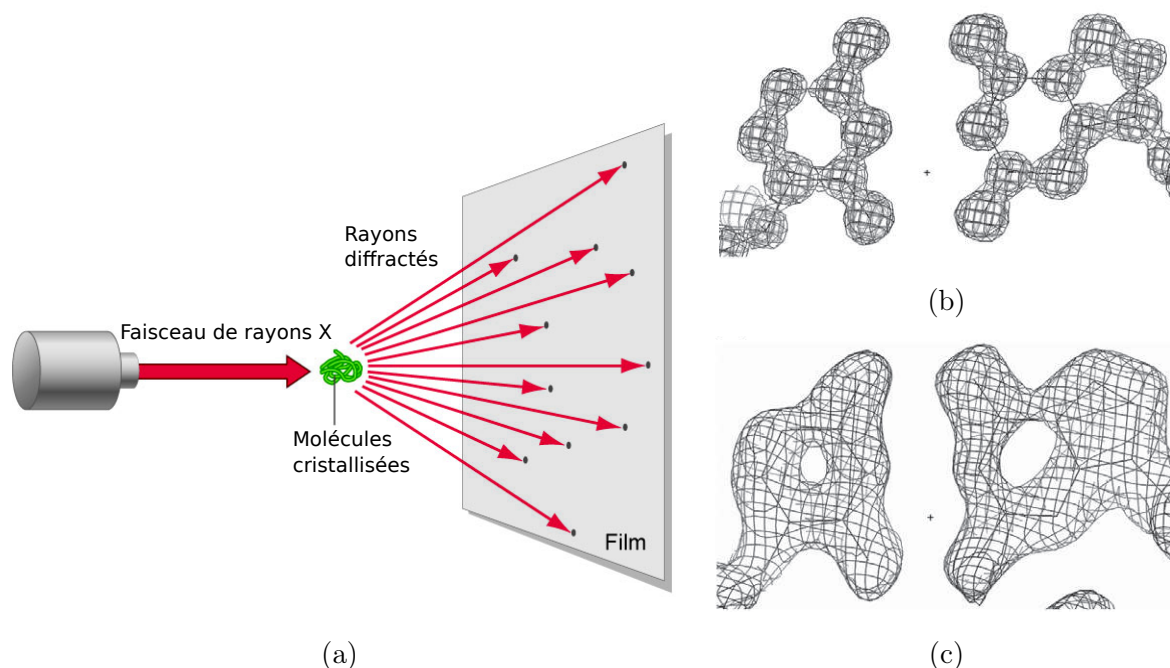


Figure 1.9. Cristallographie aux rayons X. (a) Description schématique de la technique. (b-c) Densités électroniques calculées dans le plan d'une paire de base C-G, avec une résolution de 0.9 Å en (b) et 2 Å en (c) ; images de Neidle (2008).

Spectroscopie RMN

L'application de la spectroscopie RMN à la détermination de la structure des molécules est apparue dans les années 60. Il s'agit de placer une solution de la molécule étudiée dans un champ magnétique homogène puissant et de la soumettre à des impulsions d'ondes électromagnétiques haute fréquence. Soumis à de telles excitations, les noyaux atomiques possédant un spin magnétique (notamment celui de l'Hydrogène) se comportent comme des aimants et entrent en rotation en absorbant de l'énergie électromagnétique. Après chaque impulsion, ils reviennent à l'équilibre en émettant des ondes électromagnétiques. Comme chaque atome émet à une fréquence légèrement différente, fonction de l'arrangement spatial avec les atomes voisins, l'analyse des spectres de Fourier de ces ondes aboutit à la composition chimique de la molécule. Les informations spatiales sont ensuite obtenues en étudiant les interactions entre les spins magnétiques, souvent en exerçant plusieurs impulsions consécutives. Ainsi, le couplage scalaire, qui se produit par l'intermédiaire des liaisons chimiques, donne une partie des angles de torsion de la chaîne sucre-phosphate via l'équation de Karplus. D'autre part, l'effet Overhauser nucléaire, qui décrit l'interaction entre deux spins à travers l'espace, fournit une mesure des distances interatomiques (limitée à 5 ou 6 Å). Finalement, la spectroscopie RMN n'aboutit pas à la conformation de la

molécule, mais à un ensemble de contraintes géométriques qui sont déposées dans la *Biological Magnetic Resonance Data Bank* (BMRB).

Pour déterminer la structure 3D, il faut encore restreindre l'espace conformationnel en ajoutant des informations (longueurs et angles de liaison, contraintes éventuelles sur certains angles de torsion) puis en appliquant un algorithme de géométrie de distance. Enfin, une vingtaine de propositions de structures sont dérivées après quelques traitements théoriques supplémentaires : souvent, un algorithme de recuit simulé, 10 ps de dynamique moléculaire et une minimisation d'énergie. Ces calculs sont implémentés dans des programmes spécialisés comme DYANA (Güntert *et al.* 1997), devenu CYANA (Güntert 2004).

Contrairement à la cristallographie, la spectroscopie RMN ne nécessite pas de cristalliser la molécule, ce qui constitue un avantage significatif. Cependant, la taille des structures RMN est limitée par le nombre de pics observables sur le spectre. De plus, les données fournies sont incomplètes : la méthode est sous-déterminée et requiert par conséquent l'utilisation de modèles (Santini *et al.* 2009).

Cryo-microscopie électronique (cryo-EM)

Enfin, depuis les années 80, la cryo-microscopie électronique ne cesse de progresser et devient une méthode prometteuse pour la détermination de structures d'acides nucléiques et de protéines de très grandes tailles. Les images 2D d'un microscope électronique en transmission, prises avec différents angles de vue, sont analysées, classées puis combinées pour reconstruire la structure 3D de la molécule. L'échantillon est congelé à très basse température pour réduire au maximum l'endommagement provoqué par le faisceau d'électrons, mais il n'est pas cristallisé : il reste par conséquent dans un environnement physico-chimique proche des conditions naturelles. Malgré ces précautions, le débit d'électrons doit être limité pour minimiser davantage l'endommagement. En conséquence, les images produites présentent un bruit important et doivent être filtrées. Finalement, la carte de densité 3D ("enveloppe" de la molécule) est obtenue en accumulant plusieurs dizaines de milliers d'images. A partir de cette dernière, en ajoutant des informations issues de cristallographie, de spectroscopie RMN ou d'autres mesures de cryo-EM, la conformation 3D de la molécule est calculée. Avec cette technique, il est possible d'étudier de très gros complexes moléculaires comme le ribosome. Cependant, elle est encore limitée à des résolutions nettement inférieures à la cristallographie, au mieux 3 ou 4 Å (Henderson 2015).

Enjeux de la détermination expérimentale des acides nucléiques

Environ 90 % des molécules de la PDB ont été obtenues par cristallographie, contre 9 % pour la RMN et 1 % pour la cryo-microscopie. Cela soulève la question de l'impact des modifications conformationnelles liées à la cristallisation des échantillons. D'autre part, l'obtention de résultats en RMN requiert l'utilisation de modèles. Ainsi, la modélisation des acides nucléiques est indispensable pour compléter les études expérimentales.

1.3.3 Modélisation moléculaire

Mécanique moléculaire quantique

La mécanique quantique est *a priori* la discipline la plus appropriée pour modéliser les molécules à l'échelle atomique et subatomique. Les mouvements de chaque particule (atome ou électron) de masse m , de position $\mathbf{r} = x \mathbf{e}_x + y \mathbf{e}_y + z \mathbf{e}_z$ et d'énergie totale E_{part} sont décrits par l'équation de Schrödinger stationnaire :

$$\hat{H} \Psi(\mathbf{r}) = E_{\text{part}} \Psi(\mathbf{r}) \quad (1.1)$$

où l'opérateur Hamiltonien $\hat{H}$ est défini par

$$\hat{H} = -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + V_{\text{part}}(\mathbf{r}).$$

La fonction d'onde $\Psi(\mathbf{r})$, à valeurs complexes, caractérise le mouvement de la particule. En particulier, la quantité $|\Psi(\mathbf{r})|^2$ est la densité de probabilité de la position $\mathbf{r}$. Le scalaire $\hbar$ désigne la constante de Planck divisée par 2π , et la fonction $V_{\text{part}}(\mathbf{r})$ est l'énergie potentielle de la particule à la position $\mathbf{r}$.

L'approximation de Born-Oppenheimer, qui considère les noyaux atomiques comme fixes, est souvent utilisée. Ainsi, la résolution de l'équation de Schrödinger aboutit à la distribution de densité spatiale des électrons des atomes de la molécule. Avec le théorème de Hellmann-Feynman, il est ensuite possible d'obtenir l'ensemble des forces présentes dans la molécule. Toutefois, les calculs en mécanique quantique sont très coûteux et par conséquent, limités quant au nombre d'atomes de la molécule.

Mécanique moléculaire classique

En mécanique moléculaire "classique", les mouvements des électrons sont ignorés et la molécule est représentée par un ensemble de masses centrées sur les noyaux des atomes. Ces masses sont connectées entre elles par des liaisons généralement considérées comme des ressorts (Figure 1.10a). La conformation est alors obtenue en minimisant l'énergie potentielle $V(X)$ du système, qui peut être exprimée en fonction de la liste X des positions atomiques comme la somme de quatre termes :

$$V(X) = V_{\text{liaisons}}(X) + V_{\text{angles}}(X) + V_{\text{torsions}}(X) + V_{\text{distance}}(X). \quad (1.2)$$

Les deux premiers termes décrivent les coûts énergétiques liés à l'écart des longueurs de liaisons b_i et des angles de liaisons θ_i avec leurs valeurs d'équilibre $b_{i,0}$ et $\theta_{i,0}$. Ils sont généralement modélisés par des potentiels harmoniques :

$$V_{\text{liaisons}}(X) = \sum_{\text{liaisons } i} \frac{K_{1i}}{2} (b_i - b_{i,0})^2, \quad V_{\text{angles}}(X) = \sum_{\text{angles } i} \frac{K_{2i}}{2} (\theta_i - \theta_{i,0})^2. \quad (1.3)$$

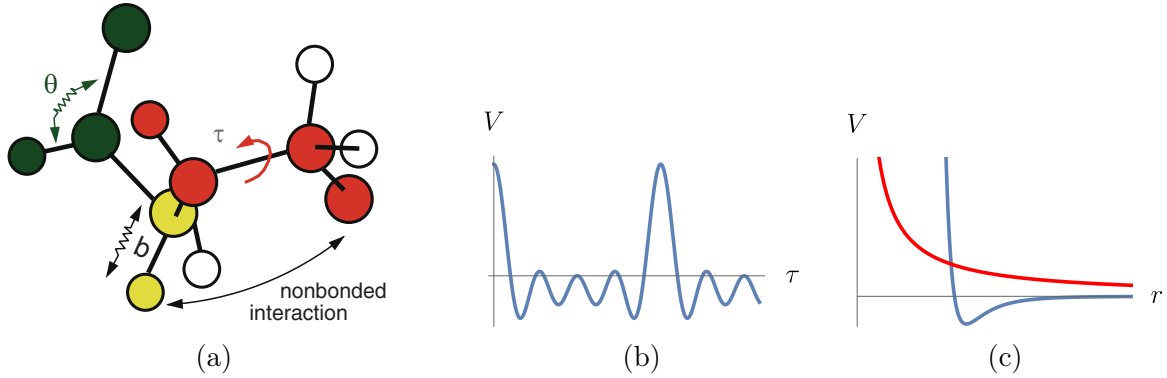


Figure 1.10. Mécanique moléculaire. (a) Paramétrisation d'une molécule : des potentiels harmoniques sont associés aux longueurs et angles de liaisons ; schéma de [Schlick \(2010\)](#). (b) Allure du potentiel associé aux angles de torsion τ . (c) Allure du potentiel de Lennard-Jones (bleu) et du potentiel électrostatique (rouge) utilisés pour décrire les interactions entre deux atomes à une distance r .

Le troisième terme décrit la variation énergétique due à la rotation des liaisons d'angles de torsion τ_i . Il est souvent représenté par une somme de potentiels sinusoïdaux (Figure 1.10b) :

$$V_{\text{torsions}}(X) = \sum_{\text{torsions } i} \sum_n \frac{V_{ni}}{2} [1 + \cos(n\tau_i - \gamma_i)] . \quad (1.4)$$

Enfin, le quatrième terme décrit les interactions entre atomes non liés chimiquement. Il est généralement exprimé comme la somme de deux contributions (Figure 1.10c) : (i) le potentiel de Lennard-Jones, qui regroupe les forces répulsives dues à l'interaction des nuages électroniques (principe d'exclusion de Pauli) et les forces attractives de van der Waals, et (ii) le potentiel des forces électrostatiques, soit

$$V_{\text{distance}}(X) = \sum_{i,j,i < j} V_{\text{LJ}}(r_{ij}) + V_{\text{Elec}}(r_{ij}) \quad (1.5)$$

avec

$$V_{\text{LJ}}(r_{ij}) = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} , \quad V_{\text{Elec}}(r_{ij}) = \frac{q_i q_j}{4\pi \epsilon_0 \epsilon(r_{ij}) r_{ij}} \quad (1.6)$$

où r_{ij} représente la distance entre les atomes i et j , ϵ_0 la permittivité du vide et $\epsilon(r_{ij})$ la permittivité relative.

La donnée de la fonction $V(X)$ et des valeurs de ses paramètres K_{1i} , $b_{i,0}$, K_{2i} , $\theta_{i,0}$, V_{ni} , γ_i , A_{ij} et B_{ij} en fonction des atomes, types d'atomes et interactions entre atomes est appelée *champ de forces*. De nombreuses propositions de champs de forces existent et se distinguent de plusieurs façons. D'abord, les fonctions utilisées pour définir l'énergie potentielle peuvent être différentes. Elles sont choisies selon un compromis entre précision et rapidité de calcul. Ensuite, les valeurs des paramètres peuvent également varier. Généralement, elles dépendent de l'agencement des atomes et de l'environnement de la molécule : certains champs de forces tiennent compte de cette dépendance. La conformation peut être paramétrée en coordonnées cartésiennes ou en coordonnées internes. Parfois, les longueurs et angles de liaisons sont fixés et seuls les angles de torsion sont considérés comme inconnus.

Bien que réduit par rapport à la mécanique moléculaire quantique, le temps de calcul augmente rapidement avec le nombre d'atomes et doit être surveillé. Une conformation implique en principe $3N - 6$ degrés de liberté, où N est le nombre d'atomes (six degrés de liberté sont retranchés pour décompter les mouvements de corps rigide). Le nombre de liaisons, d'angles de liaisons et d'angles de torsion varie linéairement avec N , tandis que le nombre d'interactions à distance varie quadratiquement. En conséquence, ces dernières font généralement l'objet de traitements spécifiques pour réduire la complexité du calcul : rayon de troncature, sommation d'Ewald, méthode du maillage particulier d'Ewald et méthode multipolaire rapide (Schlick 2010).

Modes normaux

Bien qu'utile pour déterminer des conformations "moyennes", la minimisation d'énergie ne prend pas en compte les effets dynamiques. Or, les molécules sont constamment en vibration sous l'action de la température. Tout mode vibratoire d'une molécule de N atomes peut être décrit comme la superposition de $3N - 6$ oscillations fondamentales appelées *modes normaux*. Sans résoudre complètement les équations du principe fondamental de la dynamique, la description des modes normaux (fréquence propre, direction, amplitude et sens) apporte des informations sur la mobilité des atomes et la flexibilité structurale de la molécule, souvent utilisées pour raffiner des résultats expérimentaux (Ma 2005). A titre d'exemple, la Figure 1.11 représente les modes normaux d'une molécule d'eau.

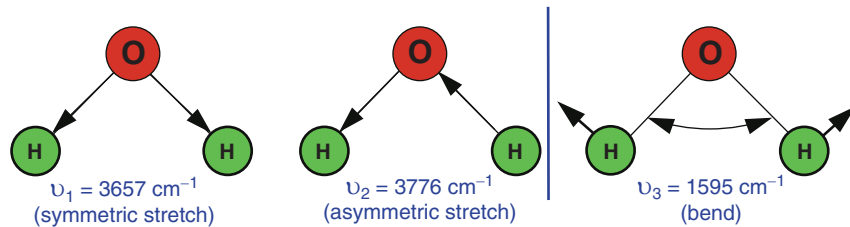


Figure 1.11. Modes normaux d'une molécule d'eau. Figure de Schlick (2010).

Une telle description peut être obtenue à partir de la matrice Hessienne. En notant $X = \{x_i\}$ l'ensemble des $3N$ coordonnées donnant la position 3D de chaque atome et $V(X)$ l'énergie potentielle de la molécule (définie par le champ de force), cette matrice s'écrit $\mathbf{V}'' = (\partial^2 V / \partial x_i \partial x_j)_{ij}$. Une matrice $\mathbf{K}$ de rigidité par unité de masse peut alors être définie (Leach 2001) par

$$\mathbf{K} = \mathbf{M}^{-1/2} \mathbf{V}'' \mathbf{M}^{-1/2}, \quad (1.7)$$

où $\mathbf{M}$ est une matrice diagonale de taille $3N \times 3N$ contenant les masses des atomes répétées trois fois chacune ($M_{11} = m_1, M_{22} = m_1, M_{33} = m_1, M_{44} = m_2$, etc.). Les fréquences des différents modes normaux sont calculées à partir des valeurs propres de la matrice $\mathbf{K}$, et les directions, amplitudes et sens à partir des vecteurs propres. A noter que six des valeurs propres sont forcément nulles et ne décrivent aucun mode normal, puisque le nombre effectif de degrés de liberté est $3N - 6$.

Simulation de Monte Carlo

Une autre façon de considérer l'agitation thermique est de générer des conformations à partir de changements aléatoires des positions des atomes. C'est ce qui est proposé dans les simulations de Monte Carlo, qui sont historiquement les premiers calculs de modélisation moléculaire. D'après la loi de Boltzmann, la probabilité d'une conformation X d'énergie totale $E(X)$ est proportionnelle à $\exp[-E(X)/(k_B T)]$, où k_B est la constante de Boltzmann et T la température. Cette propriété est utilisée dans un algorithme proposé par [Metropolis et al. \(1953\)](#) pour explorer l'espace conformationnel d'une molécule, et peut entre autres servir à minimiser l'énergie potentielle. A chaque itération, une nouvelle conformation est calculée en changeant aléatoirement les coordonnées cartésiennes d'une particule choisie aléatoirement. Si la variation d'énergie potentielle ΔV impliquée par ce changement est négative, la nouvelle conformation est acceptée. Si au contraire ΔV est positive, la nouvelle conformation est acceptée avec une probabilité $\exp[-\Delta V/(k_B T)]$. En pratique, dans le cas $\Delta V > 0$ on choisit aléatoirement un nombre ξ entre 0 et 1, et on accepte la nouvelle conformation seulement si

$$\xi = \text{rand}(0, 1) \leq \exp[-\Delta V/(k_B T)]. \quad (1.8)$$

Cette méthode minimise l'énergie avec un coût de calcul raisonnable en se basant sur une modélisation statistique du mouvement Brownien, mais ne tient pas compte des équations de la dynamique.

Dynamique moléculaire

Malgré l'utilité incontestable des autres approches, la modélisation des conformations moléculaires et de leur évolution dans le temps n'est accessible qu'au travers de la dynamique moléculaire. Cette méthode consiste à résoudre le système d'équations différentielles résultant de l'application de la deuxième loi de Newton sur chaque atome (ou groupe d'atomes) de la molécule :

$$\mathbf{M} \frac{d^2 \mathbf{X}}{dt^2}(t) = \mathbf{F}(\mathbf{X}(t)) = -\nabla V(\mathbf{X}(t)) + \mathbf{F}_{\text{add}}(t), \quad (1.9)$$

où $\mathbf{M}$ est la même matrice de masse $3N \times 3N$ que pour l'équation (1.7) et $\mathbf{X} = \{x_i\}$ est la liste des $3N$ coordonnées cartésiennes de chaque atome. Le vecteur $\mathbf{F}$ est la liste des $3N$ composantes des forces appliquées sur chaque atome. Il se décompose entre l'opposé du gradient de l'énergie potentielle et un terme additif (éventuel) $\mathbf{F}_{\text{add}}(t)$ qui représente l'action de l'environnement de la molécule. Chaque composante $i \in \{1, \dots, 3N\}$ du gradient s'écrit

$$\nabla V(\mathbf{X})_i = \frac{\partial V}{\partial x_i}(\mathbf{X}). \quad (1.10)$$

Généralement, l'énergie potentielle est déterminée grâce à un champ de force.

L'équation (1.9) est résolue par intégration numérique. L'algorithme de Verlet ([Verlet 1967](#)) est sans doute le plus fréquemment utilisé :

$$\mathbf{X}(t + \delta t) = 2\mathbf{X}(t) - \mathbf{X}(t - \delta t) + \delta t^2 \mathbf{M}^{-1} \mathbf{F}(\mathbf{X}(t)) + O(\delta t^4). \quad (1.11)$$

Toutefois, de nombreuses variantes (Leapfrog, Verlet vitesse, Beeman) ainsi que d'autres schémas d'intégration, comme les méthodes prédicteur-correcteur, peuvent également être utilisés.

Quelque soit la méthode d'intégration, les simulations de dynamique moléculaire doivent être démarrées en spécifiant des coordonnées et vitesses initiales de chaque atome. Généralement, les coordonnées des positions initiales sont obtenues à partir de résultats expérimentaux, suivis de minimisations d'énergie. Les vitesses initiales sont quant à elles générées théoriquement à partir d'une distribution Gaussienne.

A la difficulté de l'initialisation (et de l'étude de son influence sur les trajectoires obtenues) s'ajoute la complexité du calcul. Cette dernière est due au nombre important de degrés de liberté, et surtout à la nécessité d'utiliser un pas de temps très faible, au plus 1 fs, pour décrire les mouvements à haute fréquence avec une résolution convenable. Ainsi, un million de pas doivent être effectués pour couvrir 1 ns de simulation. La plupart des simulations effectuées sont par conséquent limitées à des durées allant de la nanoseconde à la microseconde. Il y a également une limitation sur la taille de la molécule. Or, les mécanismes biologiques peuvent impliquer des durées d'ordres de grandeur importants, jusqu'à la seconde (juxtaposition de sites de l'ADN surenroulé), l'heure ou même davantage. Les échelles de longueurs peuvent également être considérables, jusqu'à des centaines d'Angstroms.

Plateformes de modélisation moléculaire

Les modèles qui viennent d'être présentés sont implémentés dans les trois principales plateformes de modélisation moléculaire : CHARMM (Brooks *et al.* 2009), Amber (Salomon-Ferrer *et al.* 2013) et GROMACS (Hess *et al.* 2008). Les logiciels VMD (Humphrey *et al.* 1996) et Chimera (Pettersen *et al.* 2004) sont très largement utilisés pour visualiser et analyser les résultats. Pour illustrer les enjeux liés au temps de calcul, les performances de la cinquième version de GROMACS sont tracées dans la Figure 1.12 en termes de nanosecondes simulables par jour, en fonction du nombre d'atomes et du calculateur utilisé.

1.4 Nouvelles perspectives

Des développements présentés en sous-section 1.3, on déduit qu'il n'existe actuellement pas de technique entièrement expérimentale pour déterminer les structures d'acides nucléiques en solution, ni de méthode entièrement théorique pour calculer les conformations en fonction du temps à l'échelle atomique. En effet, tandis que la cristallographie peut s'accompagner de modifications structurales de l'échantillon, la RMN et la microscopie électronique ne fournissent généralement que des contraintes géométriques, complétées par des modèles théoriques. D'autre part, les simulations de dynamique moléculaire sont limitées à des durées et des déplacements relativement courts, et nécessitent par conséquent des données expérimentales ou phylogénétiques pour être initialisées. Les calculs plus simples, comme la minimisation d'énergie, les modes normaux ou la méthode de Monte Carlo, sont utiles mais ne donnent pas d'information sur la dynamique de la molécule.

Face à ces difficultés, plusieurs approches prometteuses sont développées. Tout d'abord, la résolution des techniques de microscopie électronique pour la détermination conformationnelle des molécules ne cesse d'être améliorée (Henderson 2015), ce qui pourrait aboutir à une méthode entièrement expérimentale. Au laboratoire national Lawrence-Berkeley en Californie, des atomes

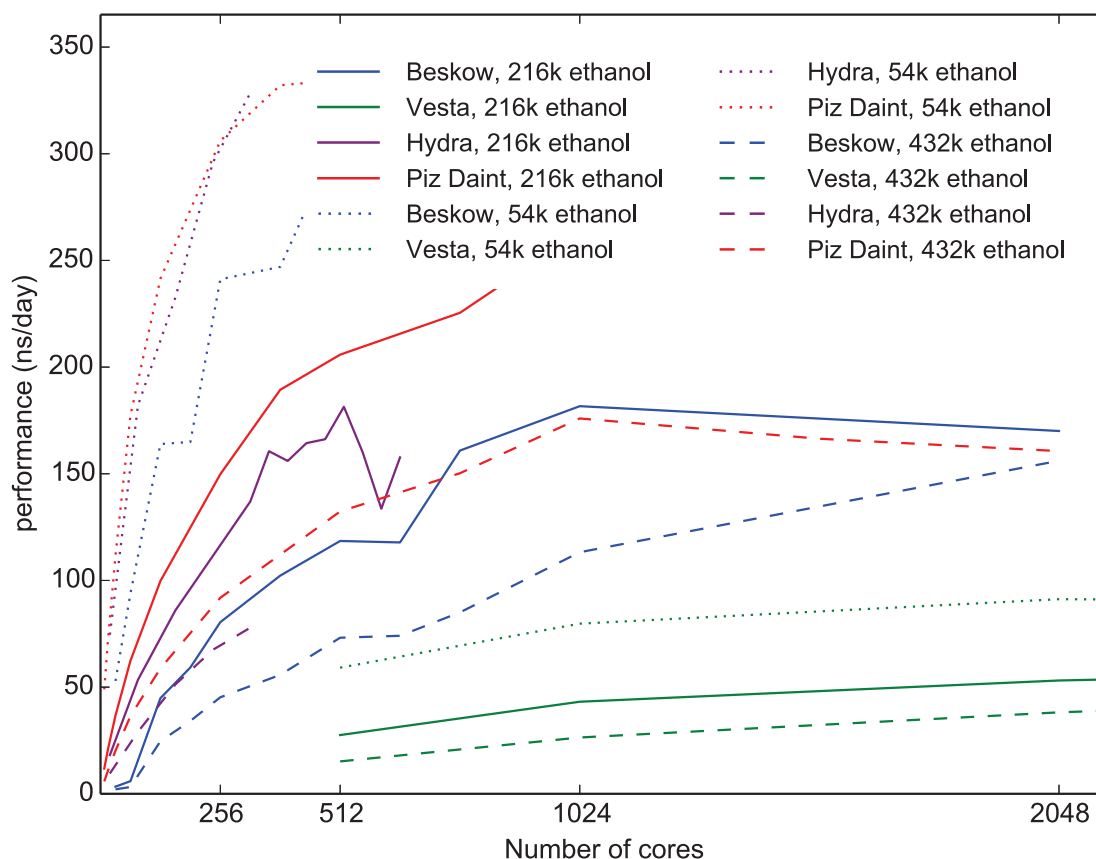


Figure 1.12. Performances de GROMACS 5.0 (en nanosecondes simulées par jour) en fonction du nombre de cœurs utilisés, pour des solutions d'éthanol dans de l'eau SPC/E, en utilisant le champ de force "OPLS All-Atom" (Jorgensen *et al.* 1996). Différentes tailles de molécules sont comparées (54k, 216k et 432k particules) ainsi que plusieurs calculateurs (Beskow, Vesta, Hydra et PizDaint). Figure diffusée sur le site internet de GROMACS (Science for Life Laboratory).

de Silicium séparés de 0.78 Å ont pu être observés (O'Keefe *et al.* 2001) : cela implique cependant des débits d'électrons encore trop importants pour être utilisé sur les biopolymères.

Pour réduire le temps de calcul des simulations numériques, différentes stratégies sont proposées. L'une d'entre elles consiste à utiliser des plateformes de calcul distribué. Ainsi, le code de dynamique moléculaire NAMD, développé au *Center for Macromolecular Modeling and Bioinformatics* de l'université de l'Illinois, peut simuler des molécules de millions d'atomes en parallélisant plus de 500 000 cœurs (Phillips *et al.* 2005, 2014). Un autre exemple est le supercalculateur Anton, construit par l'entreprise de D. E. Shaw à New-York, qui regroupe 512 processeurs microcodés pour la dynamique moléculaire. Ce dernier a une performance de plusieurs microsecondes simulées par jour pour des millions d'atomes (Shaw *et al.* 2008, 2014). Avec de telles technologies, il serait possible de modéliser des déplacements suffisamment importants pour s'affranchir de données expérimentales. Cela demande toutefois de s'assurer que les champs de force utilisés sont corrects à ces échelles de temps.

Une approche plus accessible pour disposer d'une puissance de calcul significative consiste à solliciter des volontaires pour mettre à disposition les ressources non utilisées de leurs ordi-

nateurs personnels. Par exemple, la plateforme *Berkeley Open Infrastructure for Network Computing* (BOINC) réunit des centaines de milliers d'utilisateurs pour une puissance de calcul de 11.16 PFLOPS. Cette plateforme est notamment utilisée par le projet Rosetta@home pour la modélisation des protéines, dont la version interactive, Foldit, rencontre un franc succès. Les utilisateurs de ce jeu vidéo ont réussi à obtenir des structures de bonne qualité sur des problèmes non résolus (Khatib *et al.* 2011).

Plutôt que d'augmenter la puissance de calcul, d'autres chercheurs s'intéressent au développement de modèles moins coûteux. Des modèles gros grains comme oxDNA, où les acides nucléiques sont représentés par des chaînes de nucléotides rigides, sont proposés pour modéliser efficacement des molécules de plusieurs milliers de nucléotides (Snodin *et al.* 2015, Yagyu *et al.* 2017). Dans le même ordre d'idée, Cuesta-Lopez *et al.* (2005) puis Buyukdagli *et al.* (2006) proposent de représenter un brin d'acide nucléique par un système masses-ressorts dans un bain thermique, où chaque masse représente un nucléotide et se déplace suivant seulement deux degrés de liberté. Au Laboratoire d'Analyse et d'Architecture des Systèmes (LAAS) à Toulouse, des algorithmes de *motion planning* issus de la robotique sont utilisés pour explorer et filtrer l'espace conformationnel des protéines. Implémentés dans le logiciel Move3D, ces algorithmes sont capables de modéliser des mouvements de grande amplitude, ce qui n'est pas toujours faisable en dynamique moléculaire (Cortés *et al.* 2005).

Le modèle qui motive cette thèse consiste à représenter la chaîne sucre-phosphate des acides nucléiques par une tige élastique (aussi appelée *poutre élastique* dans ce document) en grandes déformations. Il est basé sur des observations théoriques et expérimentales qui ont conduit de nombreux chercheurs à comparer le comportement mécanique de l'ADN double-brin à celui des tiges élastiques.

2 Les acides nucléiques et le modèle de tige élastique

Ce rapprochement entre acides nucléiques et tiges élastiques apparaît presque naturellement à partir des modèles de Physique des polymères. Ces modèles, présentés en sous-section 2.1, n'ont pas pour objectif de déterminer la conformation des molécules, mais plutôt d'étudier leurs propriétés mécaniques. Depuis les années 90, de nombreux essais de caractérisation mécanique de l'ADN double-brin et, parfois, simple-brin ont confirmé ce comportement élastique. Une présentation non exhaustive en est faite en sous-section 2.2. De ces observations, plusieurs auteurs ont eu l'idée d'utiliser des courbes continues et mêmes des tiges élastiques pour analyser ou déterminer les structures tridimensionnelles de molécules d'ADN. Cela est discuté en sous-section 2.3.

2.1 Modèles de Physique des polymères

En Physique des polymères (Cantor et Schimmel 1980, Rubinstein et Colby 2003), le squelette de la molécule est modélisé par une succession de $n + 1$ atomes (ou groupements d'atomes) en mouvement aléatoire, avec des liaisons d'égale longueur l (Figure 1.13a). Cette représentation simplifiée est suffisante pour obtenir des informations sur les propriétés géométriques et mécaniques des acides nucléiques, notamment des relations de comportement force-extension.

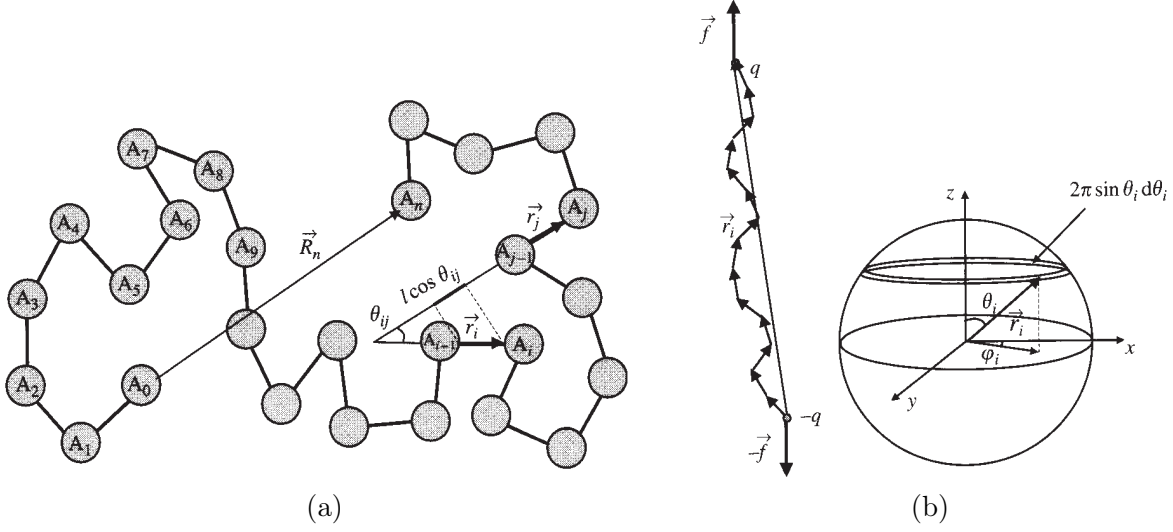


Figure 1.13. Modèle de Physique des polymères. (a) La molécule est représentée par un enchaînement de $n + 1$ atomes avec des liaisons d'égale longueur l . (b) Chaîne moléculaire soumise à une force de traction $\vec{f}$. Figures de [Rubinstein et Colby \(2003\)](#).

La distance bout-à-bout $\mathbf{R}$ est définie par

$$\mathbf{R} = \sum_{i=1}^n \mathbf{r}_i, \quad (1.12)$$

où $\mathbf{r}_i$ représente le vecteur de la liaison i , dirigé de l'atome A_{i-1} vers l'atome A_i .

La longueur totale de la chaîne complètement étirée est appelée longueur de contour L_c . L'agitation thermique provoque un mouvement aléatoire des atomes qui conduit à un recroquevillement de la chaîne. En conséquence, la valeur moyenne $\langle R \rangle$ de la distance bout-à-bout est toujours inférieure à la longueur de contour :

$$\langle R \rangle < L_c = nl. \quad (1.13)$$

2.1.1 Modèle de la chaîne librement jointe ou Freely Jointed Chain (FJC)

En première approximation, les forces d'interaction entre atomes sont négligées et les directions des liaisons sont considérées comme décorréées, c'est-à-dire indépendantes les unes des autres : c'est le modèle de la chaîne librement jointe, en anglais *Freely Jointed Chain* (FJC).

Relation force-extension

Considérons une telle chaîne soumise à une force de traction $\mathbf{F}$ suivant un axe z . On note θ_i et φ_i les coordonnées sphériques de $\mathbf{r}_i$ (voir Figure 1.13b). L'énergie E de la chaîne est égale au travail de la force :

$$E = -\mathbf{F} \cdot \mathbf{R} = -FR_z = -Fl \sum_{i=1}^n \cos(\theta_i). \quad (1.14)$$

Or, la probabilité d'une conformation donnée est proportionnelle au facteur de Boltzmann $\exp[-E/(k_B T)]$. On appelle fonction de partition Z la somme infinie des facteurs de Boltzmann sur l'ensemble des conformations possibles :

$$Z = \int_0^{2\pi} \int_0^\pi \exp\left(\frac{Fl}{k_B T} \sum_{i=1}^n \cos \theta_i\right) \prod_{i=1}^n \sin \theta_i d\theta_i d\varphi_i = \left[\frac{4\pi \sinh[Fl/(k_B T)]}{Fl/(k_B T)} \right]^n. \quad (1.15)$$

A partir de cette fonction, la plupart des variables thermodynamiques peuvent être calculées. En particulier, l'énergie libre G de Gibbs est donnée par

$$G = -k_B T \ln(Z), \quad (1.16)$$

et la valeur moyenne $\langle R \rangle$ de la distance bout-à-bout est obtenue en fonction de la force F par

$$\langle R \rangle = -\frac{\partial G}{\partial f} = nl \left[\coth\left(\frac{Fl}{k_B T}\right) - \frac{k_B T}{Fl} \right]. \quad (1.17)$$

La relation force-extension (1.17) traduit l'élasticité de la chaîne, d'origine purement entropique : la réduction du nombre de conformations accessibles, résultant de l'étirement, engendre un coût entropique qui se traduit par une force de résistance.

Le modèle FJC exprime relativement bien le comportement de chaînes moléculaires peu rigides comme certains ADNs et ARNs simple-brin. Toutefois, de nombreuses structures d'acides nucléiques, comme l'ADN double-brin, ont une rigidité importante liée à l'empilement des bases et aux différentes interactions entre atomes. Cette rigidité en flexion n'est pas représentée par le modèle FJC, à cause de l'absence de corrélation entre les directions des liaisons.

Longueur de persistance

Pour évaluer la rigidité d'un polymère, une grandeur fondamentale est la longueur de persistance L_p . Il s'agit de la valeur moyenne de la projection de la distance bout-à-bout sur la première liaison de la chaîne, calculée à la limite $n \rightarrow +\infty$:

$$L_p = \lim_{n \rightarrow \infty} \left\langle \frac{\mathbf{r}_1}{l} \cdot \sum_{i=1}^n \mathbf{r}_i \right\rangle. \quad (1.18)$$

Cette grandeur peut être vue comme la longueur au bout de laquelle les corrélations d'orientation des liaisons sont perdues, à cause des fluctuations thermiques. Plus la longueur de persistance est importante, plus la rigidité en flexion de la chaîne est grande. Elle est évaluée à environ 50 nm pour l'ADN double-brin dans les conditions physiologiques (Bustamante *et al.* 2003).

2.1.2 Modèle du ver ou Worm Like Chain (WLC)

Pour tenir compte de la rigidité en flexion, Kratky et Porod (1949) ont proposé un modèle où les longueurs de liaison et les angles de liaison sont tous égaux et indépendants du temps. Seuls les angles de torsion τ_i varient, et toutes les valeurs possibles de $-\pi \leq \tau_i \leq \pi$ sont considérées équiprobables. D'autre part, la valeur θ des angles de déviation entre deux liaisons consécutives est considérée comme très petite. Pour avoir une valeur finie de longueur de persistance, cela implique que la longueur l des liaisons tend vers zéro. Ce modèle est appelé modèle du ver, en anglais *Worm-Like Chain* (WLC).

Energie de flexion

Plutôt que de considérer les angles de liaison égaux et constants, une variante très utilisée du WLC consiste à associer un coût énergétique V à la flexion du polymère (Strick *et al.* 2000, Marko 2005) :

$$V = -\frac{K_0}{l} \sum_{i=1}^{n-1} \mathbf{t}_i \cdot \mathbf{t}_{i+1} , \quad (1.19)$$

où K_0 est la rigidité en flexion exprimée en Nm^2 , et $\mathbf{t}_i = \mathbf{r}_i/l$ est le vecteur unitaire donnant la direction de la liaison i .

La relation

$$\mathbf{t}_i \cdot \mathbf{t}_{i+1} = 1 - \frac{(\mathbf{t}_{i+1} - \mathbf{t}_i)^2}{2} \quad (1.20)$$

implique que cette énergie de flexion vaut, à une constante près :

$$V = \frac{K_0}{l} \sum_{i=1}^{n-1} \frac{(\mathbf{t}_{i+1} - \mathbf{t}_i)^2}{2} = \frac{1}{2} K_0 \sum_{i=1}^{n-1} l \left(\frac{\mathbf{t}_{i+1} - \mathbf{t}_i}{l} \right)^2 . \quad (1.21)$$

Le passage à la limite $l \rightarrow 0$ donne une courbe continue de longueur L , dont l'énergie de flexion est identique à celle d'une tige élastique inextensible :

$$V = \frac{1}{2} K_0 \int_0^L \left\| \frac{d\mathbf{t}}{ds} \right\|^2 ds. \quad (1.22)$$

Ainsi, la prise en compte de la rigidité en flexion conduit à un modèle que ne diffère des tiges élastiques que par l'absence de rigidité en torsion.

D'autre part, Landau et Lifschitz ont montré (Landau et Lifshitz 1969) que la rigidité en flexion K_0 est directement liée à la longueur de persistance par la relation :

$$K_0 = k_B T L_p . \quad (1.23)$$

Relation force-extension

Lorsque la chaîne est en traction, son énergie totale est obtenue en ajoutant le travail de la force F à l'énergie de flexion :

$$E = \int_0^L \left(\frac{1}{2} K_0 \left\| \frac{d\mathbf{t}}{ds} \right\|^2 - F \cdot \mathbf{t} \right) ds. \quad (1.24)$$

Le calcul de la fonction de partition et de l'énergie libre pour obtenir une relation force-extension a été résolu numériquement par Marko et Siggia (Bustamante *et al.* 1994, Marko et Siggia 1995) à partir d'une analogie avec un problème de mécanique quantique (dipôle tournant dans un champ électrique). Ils ont également obtenu une formule analytique approchée :

$$F = \frac{k_B T}{L_p} \left[\frac{1}{4(1 - z/L)^2} - \frac{1}{4} + \frac{z}{L} \right], \quad (1.25)$$

où z est l'extension de la chaîne. Cette formule a ensuite été raffinée par Bouchiat *et al.* (1999).

2.1.3 Modèle de tige élastique ou Rod Like Chain (RLC)

Plusieurs essais de caractérisation de l'ADN double-brin en torsion, présentés en sous-section 2.2, ont conduit à l'utilisation d'un modèle de tige élastique. Ce dernier, souvent appelé *Rod Like Chain* (RLC), conduit à une énergie de déformation

$$V = \frac{1}{2}K_0 \int_0^L \left\| \frac{dt}{ds} \right\|^2 ds + \frac{1}{2}K_3 \int_0^L \Omega_t(s)^2 ds \quad (1.26)$$

où K_3 est la rigidité en torsion exprimée en Nm^2 et $\Omega_t(s)$ est la densité de torsion physique.

2.2 Essais de caractérisation mécanique des acides nucléiques

De nombreux résultats expérimentaux ont mis en évidence l'élasticité de l'ADN et ont été correctement interprétés par les modèles continus WLC et RLC. Quelques-uns sont présentés dans cette sous-section pour montrer la proximité du comportement des acides nucléiques avec celui des tiges élastiques.

2.2.1 Traction de l'ADN double-brin

Smith *et al.* (1992) ont attaché chimiquement des molécules uniques d'ADN double-brin (48.5 kpb) à une lamelle en verre sur une extrémité et à une bille micrométrique sur l'autre extrémité. Ils ont ensuite sollicité ces molécules en traction en déplaçant la bille avec des forces magnétiques et hydrodynamiques connues, puis observé et mesuré l'évolution de la position de la bille par microscopie optique. Les courbes force-extension ainsi obtenues ont été analysées avec la participation de Marko et Siggia (Bustamante *et al.* 1994, Marko et Siggia 1995) : la figure 1.14a est un de leurs résultats. Le modèle FJC décrit par l'équation (1.17) s'écarte des résultats expérimentaux parce qu'il ne tient pas compte de la rigidité en flexion de la double-hélice. Au contraire, le modèle WLC décrit par l'équation (1.25) se superpose bien aux données et représente donc correctement le comportement de l'ADN double-brin en traction. Toutefois, WLC devient incorrect lorsque des forces relativement importantes sont appliquées (supérieures à 10 pN), pour lesquelles les liaisons chimiques entre atomes sont étirées et la double-hélice se déforme. Ce régime se traduit par une contribution enthalpique de l'élasticité en surcroît de la part entropique. Il donne lieu à une évolution linéaire de l'extension en fonction de la force et peut être décrit par un modèle WLC extensible, *i.e.* auquel une élasticité linéaire en traction a été ajoutée (Odijk 1995).

Des courbes de traction sur une plage de force plus large, jusque'à 80 pN, ont été obtenues en utilisant des pinces optiques pour mesurer la force appliquée (Smith *et al.* 1996, Wang *et al.* 1997). Ainsi, le modèle WLC extensible a pu être validé jusqu'à des tensions de 45 pN environ. Au-delà de cette limite, l'élasticité de l'ADN double-brin redevient non-linéaire, puis une sur-extension importante est observée à partir d'environ 65 pN. La non-linéarité sur la plage [45 pN, 65 pN] pourrait être due au couplage traction-torsion induit par la géométrie hélicoïdale de l'ADN et démontré expérimentalement par Gore *et al.* (2006). La sur-extension après 65 pN proviendrait quant à elle d'une rupture progressive des interactions entre bases (van Mameren *et al.* 2009) ou

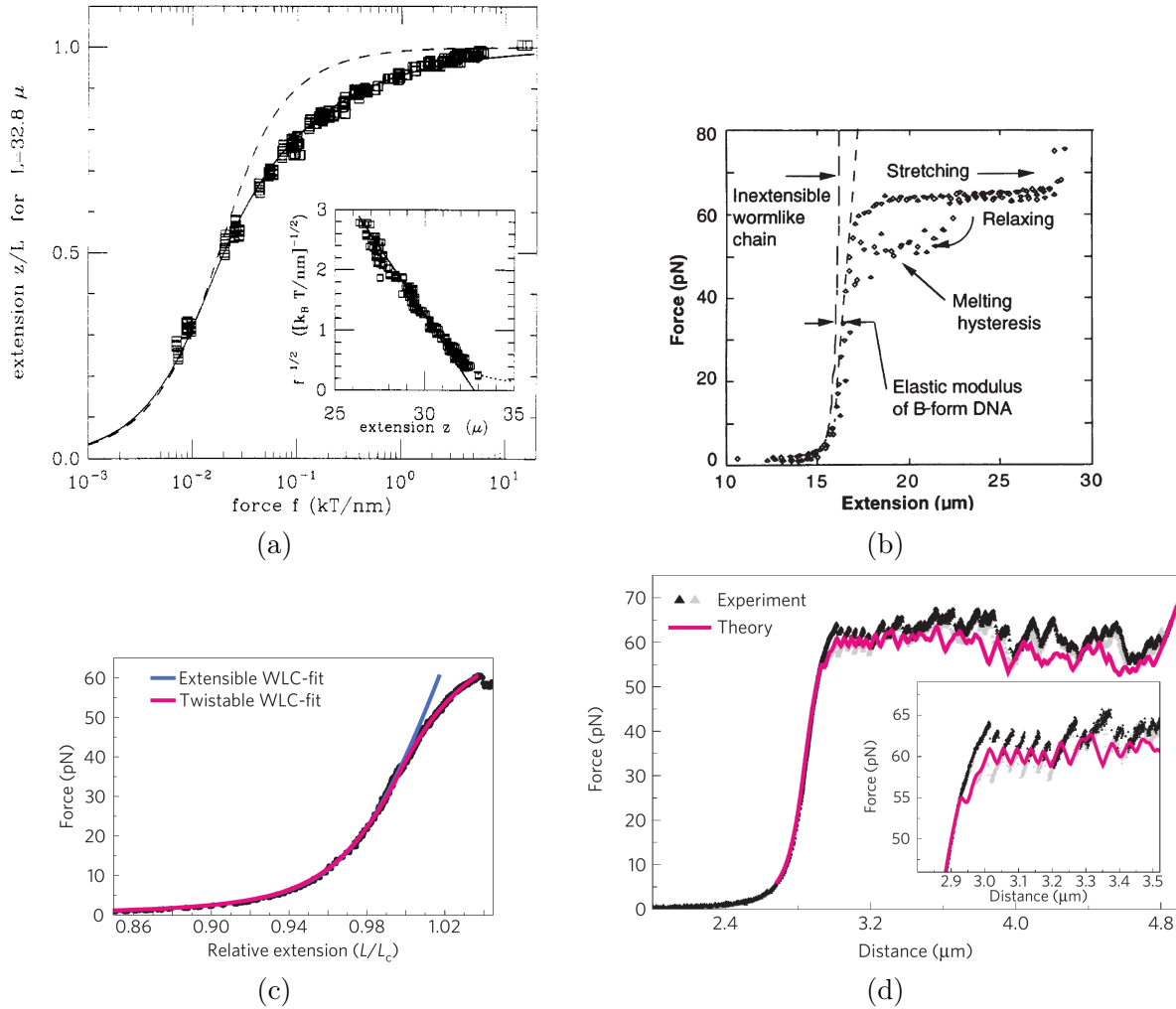


Figure 1.14. Essais de traction de molécules uniques d'ADN double-brin. (a) Courbe expérimentale force-extension de [Smith *et al.* \(1992\)](#), superposée aux modèles FJC (tirets) et WLC (ligne continue) par [Marko et Siggia \(1995\)](#). Encart : racine carré de la force en fonction de l'extension, pour des valeurs de force importantes; le modèle WLC extensible (pointillés) se superpose mieux aux données expérimentales que le WLC "classique" (ligne continue). (b) Courbe expérimentale force-extension de [Smith *et al.* \(1996\)](#). (c-d) Courbe expérimentale force-extension de [Gross *et al.* \(2011\)](#), superposée aux modèles WLC (bleu) et twistable WLC (rouge).

d'une transition de phase ([Albrecht *et al.* 2008](#), [Danilowicz *et al.* 2009](#)). Un raffinement du WLC en "twistable WLC" (tWLC) a été proposé par [Gross *et al.* \(2011\)](#) pour prendre en compte le couplage traction-torsion, et se superpose correctement avec les données expérimentales jusqu'à la sur-extension, liée à la perte progressive des interactions entre bases. Les auteurs ont intégré à leur modèle cette dynamique, qui dépend de la séquence de la molécule.

Ces différentes études montrent la pertinence du modèle continu WLC, très proche des tiges élastiques, sur une plage de forces relativement importante (sous réserve de raffinements). Elles évaluent également autour de 65 pN la limite du comportement élastique, bien que cette valeur soit dépendante de la séquence.

2.2.2 Torsion de l'ADN double-brin

Des expériences assez similaires peuvent être réalisées pour obtenir des courbes de torsion de l'ADN double-brin. Cette fois-ci, la microbille n'est pas seulement positionnée mais aussi mise en rotation, généralement par un champ magnétique. Si la molécule est attachée sur la bille avec plusieurs points de contacts sur les deux brins, cette rotation engendre une contrainte en torsion de la double-hélice. [Bryant et al. \(2003\)](#) ont proposé de vérifier la linéarité de l'élasticité en torsion de molécules (22.9 kpb) étirées à 15 pN au moins, ce qui garantit en principe l'absence de courbure de la chaîne. Les courbes obtenues sont très bruitées en raison de l'agitation thermique (Figure 1.15a), mais la moyenne d'un grand nombre de mesures identifie une plage assez importante d'élasticité quasiment linéaire (Figure 1.15b). Des valeurs de couples délimitant la zone d'élasticité sont proposées, ainsi qu'une valeur de rigidité de torsion autour de 420 pN nm². Ces résultats confortent évidemment l'utilisation du modèle de poutre élastique (RLC) plutôt que le WLC, qui ne prévoit pas de rigidité en torsion.

Lorsque des molécules d'ADN étirées à des forces moins importantes (typiquement, autour de 1 pN) sont sollicitées en torsion, elles se surenroulent et forment des plectonèmes (Figure 1.15c). Pour étudier ces géométries, le *link* Lk est très souvent utilisé : il s'agit d'une mesure du nombre de tours d'hélice, *i.e.* du nombre de tours effectués par un brin autour de l'axe central de la double-hélice. D'après le théorème de Călugăreanu-White-Fuller, cette quantité se décompose comme suit :

$$Lk = Tw + Wr. \quad (1.27)$$

Le *twist* Tw est le nombre de tours d'hélice lorsque l'axe de la molécule a une torsion géométrique nulle, c'est-à-dire une géométrie plane. C'est la contribution au link Lk de l'empilement des bases ainsi que des contraintes de torsion. Lorsque la géométrie de l'axe devient tridimensionnelle, le nombre de tours d'hélice est altéré par la contorsion de la chaîne. Le *writhe* Wr est alors utilisé pour comptabiliser la contribution de cette contorsion au link. Il peut être défini comme le nombre moyen d'auto-croisements de la double-hélice lorsqu'on la regarde sous tous les angles de vue possibles. La formation de plectonèmes est due à un transfert de Tw en Wr , autrement dit à une augmentation de la part de Wr dans le link. A noter que le link est normalement défini pour des chaînes fermées, mais peut également être étendu à des chaînes ouvertes moyennant certaines précautions ([Neukirch 2004](#)).

[Strick et al. \(1996\)](#) ont obtenu expérimentalement des courbes de l'extension en fonction du degré de surenroulement σ :

$$\sigma = \frac{\Delta Lk}{Lk_0}, \quad (1.28)$$

où ΔLk est la variation de link induite par la mise en rotation de la microbille, et Lk_0 la valeur du link intrinsèque à la molécule. Ces courbes sont reportées sur la Figure 1.15d pour différentes valeurs de force. L'apparition d'un plectonème est visible dans les cas $F = 1.3$ pN et $F = 0.3$ pN par une décroissance de l'extension avec σ .

Plusieurs travaux théoriques sont parvenus à modéliser de tels résultats en représentant l'ADN double-brin par une poutre élastique ([Moroz et Nelson 1997](#), [Haijun et al. 1999](#), [Bouchiat et Mezard 2000](#)). [Neukirch \(2004\)](#) propose même une étude où les fluctuations thermiques sont ignorées et simule finement, en tenant compte de l'auto-contact, la partie de la courbe correspondant à la formation du plectonème. Il obtient par ailleurs une estimation de la taille du plectonème et de

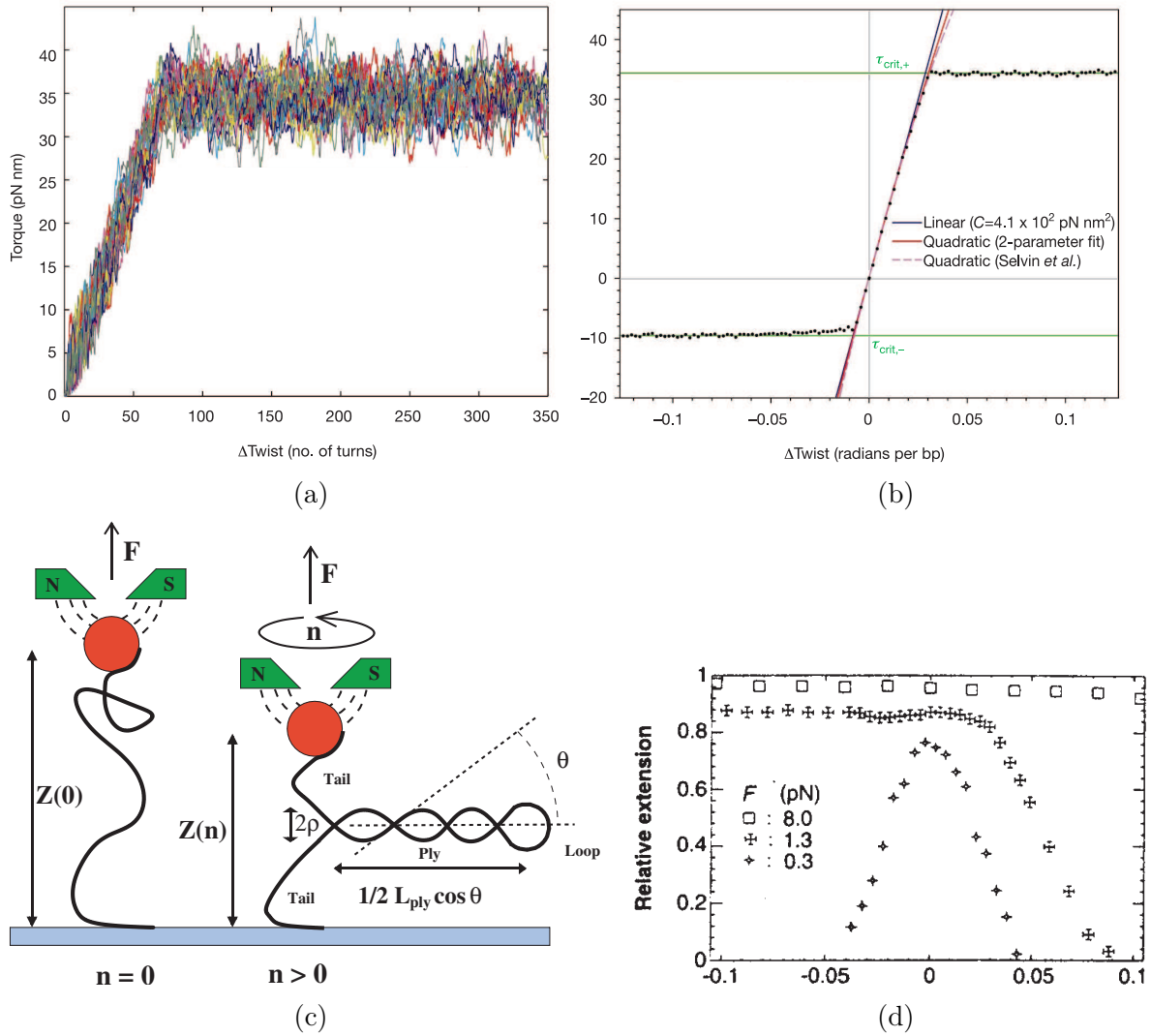


Figure 1.15. Essais de torsion de molécules uniques d'ADN double-brin. (a-b) Courbes de torsion de Bryant *et al.* (2003) : résultats expérimentaux en (a) et courbes moyennées en (b). (c) Formation d'un plectonème, schéma de Neukirch (2004). (d) Courbe expérimentale de l'extension en fonction du degré de surenroulement, obtenue par Strick *et al.* (1996).

la rigidité en torsion de la molécule. Enfin, Marko et Neukirch (2012) ont développé un modèle basé sur la théorie des poutres élastiques pour prévoir le nombre de boucles et de plectonèmes formés lorsque ΔLk augmente, et leur influence sur le comportement de la chaîne (notamment, sur l'énergie libre et l'extension).

2.2.3 Cyclisation de l'ADN double-brin

La cyclisation de l'ADN double-brin par la ligase est une méthode physico-chimique efficace pour évaluer sa rigidité en flexion. Historiquement, elle a été étudiée avant les expériences de traction et de torsion sur molécules uniques. La ligase est une enzyme qui catalyse la formation

d'une liaison phosphodiester entre deux segments d'ADN. Dans une solution contenant une concentration donnée de la molécule testée, l'action de la ligase débouche sur deux réactions : la jonction des deux extrémités d'un segment d'ADN ou cyclisation (constante d'équilibre K_C), et la jonction de deux segments pour former un dimère (constante d'équilibre K_D). La probabilité de cyclisation peut être évaluée par le facteur de Jacobson-Stockmayer $J = K_C/K_D$. Plus ce facteur est grand, plus la molécule a une probabilité importante d'être cyclisée et donc, moindre est sa rigidité en flexion. Expérimentalement, la valeur de J peut être obtenue à partir du rapport entre la quantité de monomères cyclisés et la quantité de dimères, déterminé par électrophorèse (Taylor et Hagerman 1990, Geggier et Vologodskii 2010). Les essais de cyclisation consistent à tracer des courbes expérimentales d'évolution du facteur J en fonction de la longueur de contour des fragments d'ADN utilisés.

A nouveau, le modèle WLC parvient à reproduire théoriquement de telles courbes. Cependant, un écart important a été constaté entre les valeurs expérimentales et les prévisions du WLC pour des fragments d'ADN de faible longueur : le modèle surestimerait la rigidité de la molécule (Cloutier et Widom 2004, Shroff *et al.* 2005). Pour résoudre ce problème, plusieurs auteurs ont proposé des raffinements qui tiennent compte des non-linéarités et singularités de la chaîne, notamment les cassures (changements brusques de direction) pouvant apparaître en certains points (Zhang et Crothers 2003, Wiggins *et al.* 2005, Yan *et al.* 2005, Ranjith *et al.* 2005). En particulier, Du *et al.* (2005) ajoutent au WLC la modélisation des cassures, mais aussi l'influence de la rigidité en torsion de la double hélice (Shore et Baldwin 1983). Celle-ci est à l'origine des oscillations du facteur J avec la longueur de contour, de pseudo-période égale au pas de la double hélice, soit 10.5 pb. Ils obtiennent une bonne superposition avec leurs données expérimentales (Figure 1.16a). De leur côté, après validation du WLC pour de grandes longueurs de contour, Wiggins et Nelson (2006) proposent le modèle de chaîne subélastique (SEC) qui diffère du WLC seulement sur les longueurs où ce dernier échoue. Ils obtiennent une courbe théorique en accord avec les résultats expérimentaux de Cloutier et Widom (2004) (Figure 1.16b).

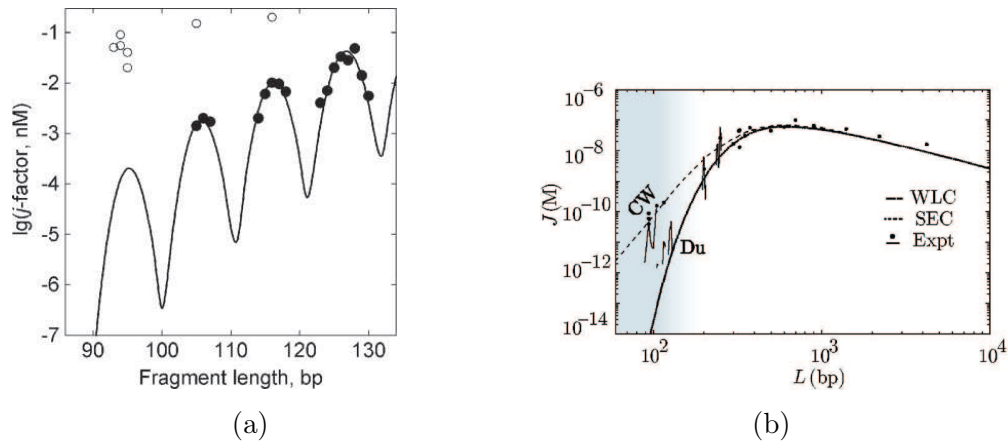


Figure 1.16. Essais de cyclisation de l'ADN double-brin : évolution du facteur J en fonction de la longueur de contour. (a) Valeurs expérimentales de Du *et al.* (2005) (cercles pleins), superposées à leur modèle WLC "amélioré" (prise en compte des cassures et de la rigidité en torsion). (b) Valeurs expérimentales de Cloutier et Widom (2004) (CW) et de Du *et al.* (2005) (Du), superposées au SEC (tirets) et au WLC (ligne continue) ; figure de Wiggins et Nelson (2006).

Ainsi, les tiges élastiques constituent un bon modèle de base pour la cyclisation de l'ADN

double-brin. Leur domaine de validité doit toutefois être défini, et des raffinements sont nécessaires pour tenir compte de la complexité de la chaîne. L'utilisation de ce modèle de tige pour obtenir des probabilités de cyclisation peut déboucher sur des calculs très efficaces et applicables à un grand nombre de cas (Manning *et al.* 1996, Cotta-Ramusino et Maddocks 2010).

2.2.4 Elasticité de l'ADN simple-brin

Contrairement au double-brin, très peu d'essais mécaniques ont été réalisés sur l'ADN simple-brin. Ce dernier est considéré comme très flexible et par conséquent, plus proche du modèle FJC que des tiges élastiques. Tinland *et al.* (1997) ont évalué sa longueur de persistance à environ 4 nm, soit plus de dix fois moins que la double hélice. Leur mesure est basée sur les coefficients de diffusion en électrophorèse sur gel, liés à la longueur de persistance par la relation de Stokes-Einstein.

Cependant, Smith *et al.* (1996) ont effectué un essai de traction d'un brin unique d'ADN, et la courbe obtenue (Figure 1.17a) montre une extension importante, sans rupture de la chaîne.

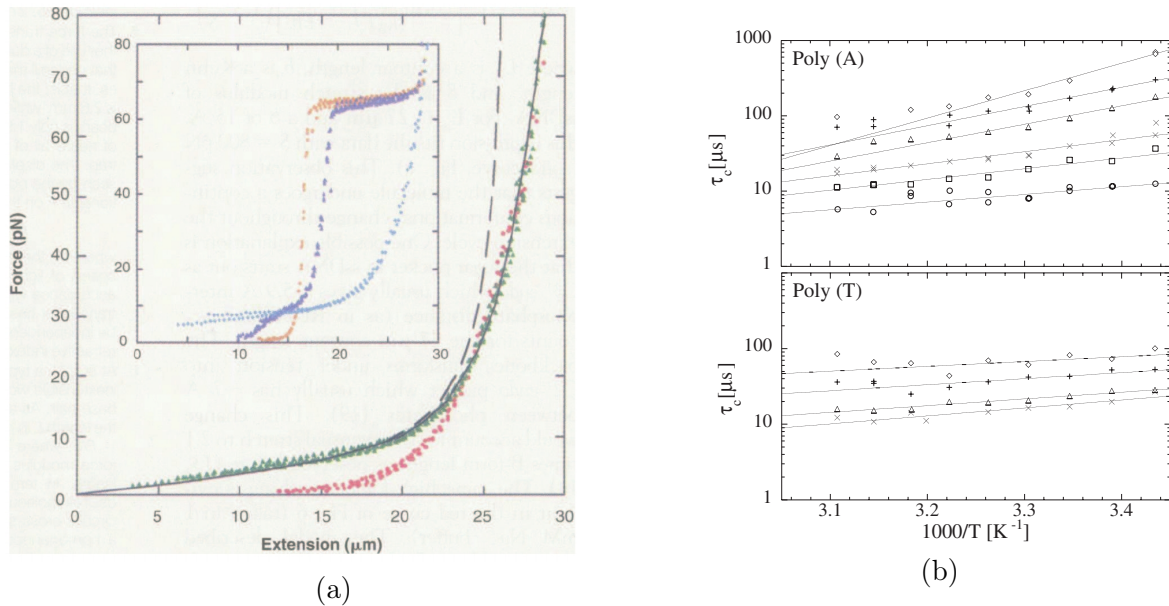


Figure 1.17. Elasticité de l'ADN simple-brin. (a) Courbes expérimentales de traction, obtenues par Smith *et al.* (1996), pour deux molécules simple-brin différentes (vert et rouge). Superposition des modèles FJC (tirets) et FJC extensible (ligne continue). Encart : courbes force-extension expérimentales de trois molécules, une double-brin (orange), une partiellement double-brin et partiellement simple-brin (violet) et une simple-brin (bleu ciel). (b) Evolution du temps de fermeture de structures en épingle à cheveux poly(A) et poly(T) en fonction de la température, pour différentes longueurs de la boucle (en nombre de bases) : $\circ = 8$, $\square = 10$, $\times = 12$, $\triangle = 16$, $+$ = 21, $\diamond = 30$. Figure de Goddard *et al.* (2000).

Pour être correctement modélisé, ce comportement nécessite d'ajouter une extensibilité au FJC. Cela met en évidence l'élasticité de l'ADN simple-brin. Pour des longueurs de contour proches de la longueur de persistance, on peut alors s'attendre à une rigidité significative de la molécule, qui serait correctement représentée par un modèle de tige élastique. D'autre part, Goddard *et al.*

(2000) ont observé que la longueur de persistance de l'ADN simple-brin est fortement dépendante de sa séquence, et peut être suffisamment importante pour remettre en cause le FJC. Pour ce faire, ils ont comparé les temps de fermeture des structures en épingle à cheveux poly(T) et poly(A). Leur résultat (Figure 1.17b) suggère que les boucles poly(A) sont beaucoup plus rigides, à cause de l'empilement des bases. Enfin, le modèle WLC, qui prévoit une rigidité en flexion, a parfois été utilisé sur l'ADN simple-brin. Il estime correctement la fraction de paires de bases cassées lors de la fusion de boucles en épingle à cheveux (Kuznetsov *et al.* 2001), ou encore l'efficacité FRET (Fluorescence Resonance Energy Transfer) d'oligonucléotides (Murphy *et al.* 2004).

Ainsi, malgré la flexibilité importante de l'ADN simple-brin, il existe probablement des cas où celui-ci est mieux décrit par une tige élastique que par une chaîne librement jointe (FJC).

2.3 Structures tridimensionnelles à partir des tiges élastiques

Les études présentées dans les deux sous-sections précédentes ont montré la validité d'un modèle continu pour représenter le comportement mécanique des acides nucléiques, et plus particulièrement, la validité du modèle de tige élastique. À partir de ces observations, on peut se demander si ces modèles sont aussi efficaces pour visualiser, décrire et prévoir la structure tridimensionnelle des acides nucléiques. À ce sujet, de nombreux travaux offrent des perspectives intéressantes.

2.3.1 Les rubans : outils de visualisation et d'analyse

Pour simplifier la visualisation des acides nucléiques, il est courant de représenter le squelette de la chaîne moléculaire par un ruban, et les bases par des blocs parallélépipédiques (Figure 1.18a). Ces descriptions schématisées contiennent une bonne partie des informations nécessaires à l'analyse, la classification et la comparaison des acides nucléiques. *3DNA* (Lu *et Olson* 2003, 2008) est l'un des programmes d'analyse les plus utilisés. Il considère les polynucléotides comme des successions de paires de bases dont l'agencement est caractérisé par un ensemble de paramètres : six par paire pour positionner et orienter relativement les bases, et six pour positionner et orienter relativement deux paires consécutives. Ce programme s'inscrit dans la continuité des conventions de Cambridge (Diekmann 1989) et de Tsukuba (Olson *et al.* 2001), où des normes ont été définies pour la description de la géométrie des ANs, en particulier des arrangements des bases et des paires de bases. Dans la même lignée, les programmes *Curves* (Lavery *et Sklenar* 1988) et *Curves+* (Lavery *et al.* 2009) ajoutent une description globale où les paires de bases sont positionnées par rapport à un axe curviligne (via un repère local) autour duquel la chaîne sucre-phosphate, décrite par une courbe, s'enroule (Figure 1.18b). Cet axe curviligne s'avère particulièrement utile pour déterminer efficacement la valeur et la direction de la courbure de la double-hélice (Pasi *et al.* 2017).

L'utilisation de trajectoires de tiges élastiques pour représenter soit l'axe curviligne de la double-hélice, soit la double-hélice elle-même, pourrait enrichir cette description en ajoutant des informations comme les sollicitations mécaniques, l'énergie et la stabilité de la molécule. Il reste cependant à savoir si de telles trajectoires élastiques présentent une structure géométrique suffisamment proche de celle observée pour les acides nucléiques. Si tel est le cas, il est alors

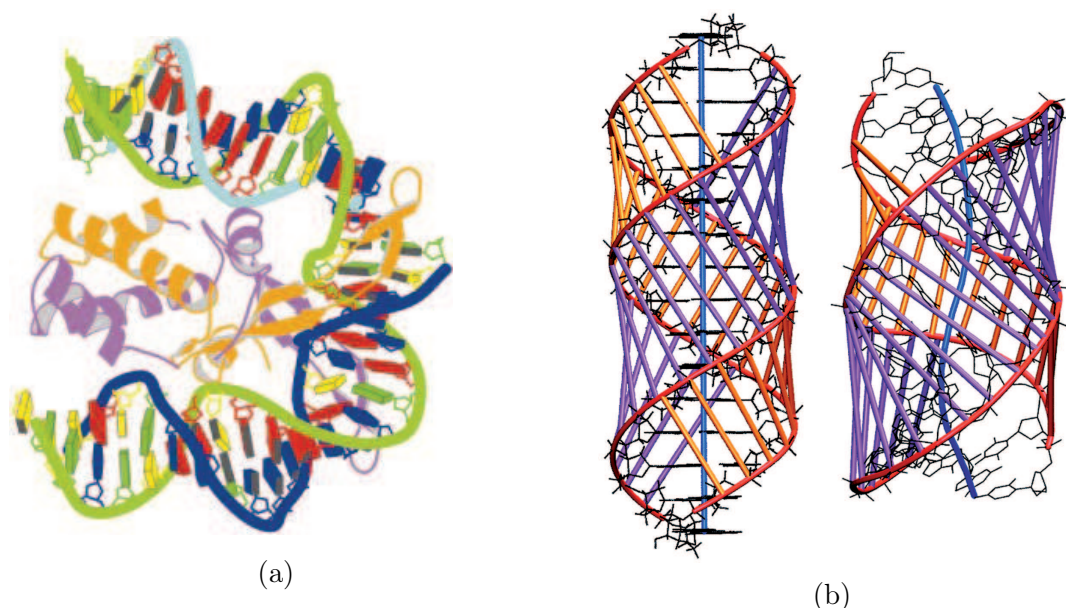


Figure 1.18. Outils de visualisation et d'analyse des acides nucléiques. (a) Représentation schématique d'un complexe IHF (Integration Host Factor) - ADN : les squelettes des molécules sont figurés par des rubans, et les chaînes latérales par des blocs. Figure de [Lu et Olson \(2003\)](#). (b) Molécule d'ADN double-brin CTGCTATAAAAGGCTG, en conformation ADN-B (gauche) et après fixation d'une protéine TBP (droite), visualisée avec le programme *Curves+*. Les chaînes sucre-phosphate sont représentées par des courbes (en rouge) qui s'enroulent autour d'un axe curviligne (en bleu). Les largeurs des sillons sont matérialisées par des vecteurs (violets pour le grand sillon, et oranges pour le petit sillon). Figure de [Lavery et al. \(2009\)](#).

possible de décrire la chaîne moléculaire par une tige élastique et même, de proposer des méthodes pour déterminer la conformation à partir de cette tige. Cela constituerait une démarche inverse à celle qui est habituellement employée pour générer des rubans à partir de la position 3D des atomes.

2.3.2 Description de l'ADN double-brin par des tiges élastiques

De nombreux chercheurs se sont intéressés aux géométries de tiges élastiques pour étudier les conformations de l'ADN double-brin ([Schlick 1995](#), [Olson 1996](#)). Dans ces travaux, la tige représente l'ensemble de la double-hélice. C'est dans ce contexte que des équations analytiques des géométries élastiques ont été obtenues, en considérant une courbure intrinsèque nulle, c'est-à-dire des tiges rectilignes en l'absence de contraintes mécaniques ([Benham 1979](#), [Tobias et al. 1994](#), [Shi et Hearst 1994](#), [Shi et al. 1995](#)). [Tobias et Olson \(1993\)](#) ont utilisé la théorie de l'élasticité des tiges pour étudier l'influence de la courbure intrinsèque sur les conformations d'équilibre de l'ADN surenroulé. [Manning et al. \(1996\)](#) proposent un modèle tige qui tient compte de l'influence de la séquence, en lissant les données issues du modèle discret d'angles de coins (Figure 1.19a) de [Bolshoy et al. \(1991\)](#). Ils déterminent ainsi la courbure intrinsèque locale de la tige, ainsi que les valeurs des rigidités en flexion et en torsion. La même équipe ([Furrer et al. 2000](#)) utilise ce modèle pour analyser la relation entre la courbure intrinsèque de l'ADN

circulaire et l'existence de plusieurs équilibres stables de même link (Figure 1.19b). Swigon *et al.* (1998) déterminent les configurations d'équilibre de nucléosomes en minimisant l'énergie d'une tige élastique (Figure 1.19c), et modélisent l'auto-contact pour investiguer minutieusement les plasmides d'ADN (Coleman et Swigon 2000). Enfin, Hearst et Shi (1995) présentent un tracé 3D d'une tige en superhélice représentant la chromatine (Figure 1.19d).

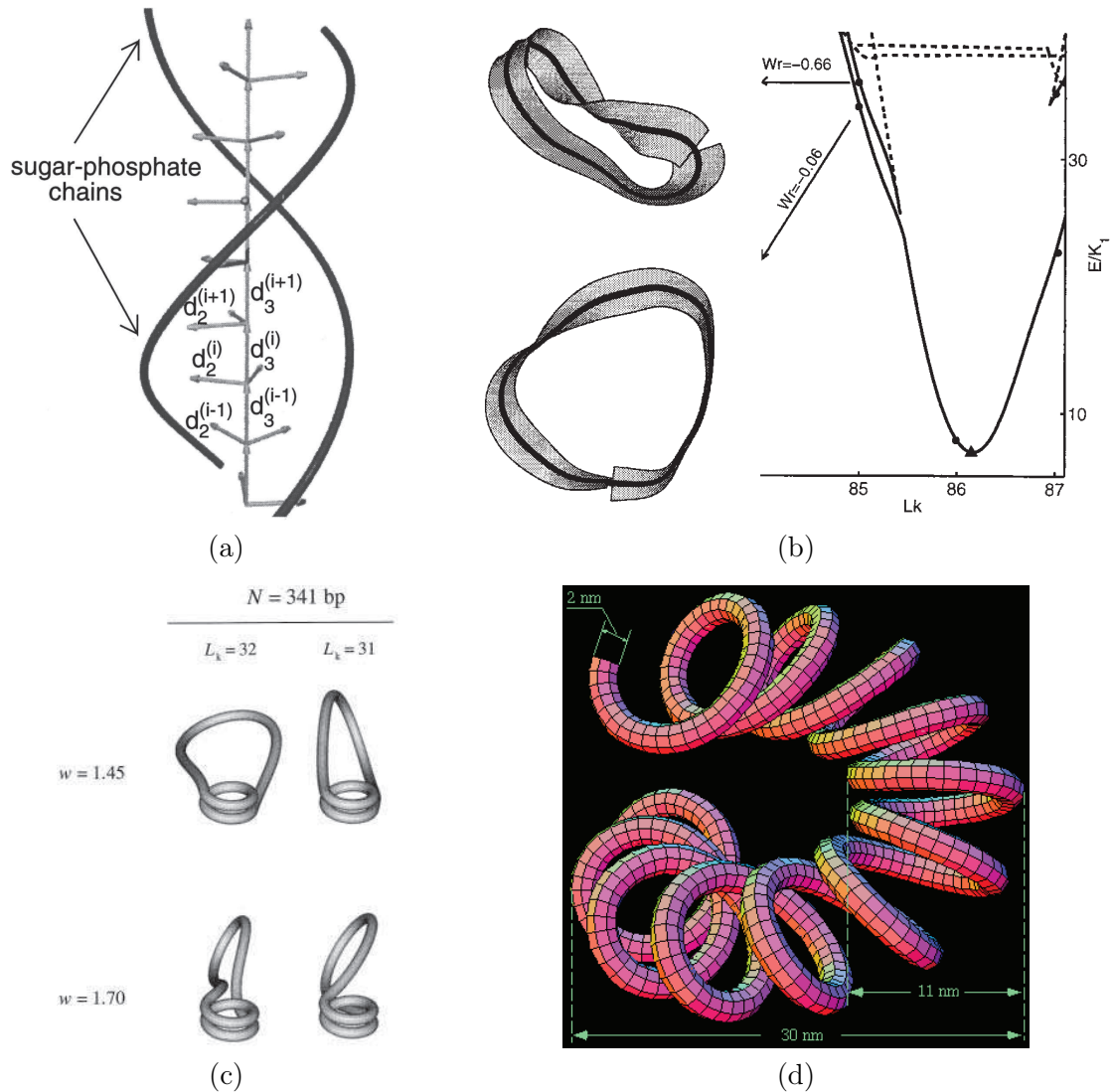


Figure 1.19. Modélisation de l'ADN par des poutres élastiques. (a) Modèle discret coin-angle de Bolshoy *et al.* (1991). Chaque repère a son origine au centre d'une paire de base, son axe $\mathbf{d}_3$ qui pointe vers la paire de base suivante, et son axe $\mathbf{d}_1$ qui pointe vers le centre du grand sillon. (b) Deux conformations de même link d'un ADN double-brin circulaire (900 paires de bases), représentées par une tige élastique dont la trajectoire a été obtenue par Furrer *et al.* (2000) à partir du modèle de Manning *et al.* (1996). (c) Modèles de nucléosomes de 341 paires de bases, décrits avec une tige élastique par Swigon *et al.* (1998). Différentes conformations de même link ou de même writhe sont mises en évidence. (d) Tige élastique en superhélice représentant la chromatine, obtenue par Hearst et Shi (1995).

Un modèle multi-échelle basé sur l'élasticité des tiges (Figure 1.20) a été développé pour

déterminer la conformation de l'opéron lactose de la bactérie *E. Coli* (Balaeff *et al.* 1999, 2006). Pour empêcher la transcription de cette séquence d'ADN, une protéine (le répresseur *lac*) isole le promoteur du reste de la chaîne en se fixant sur deux sites et en courbant la double-hélice. Selon les cas, ce mécanisme produit une boucle de 76 ou de 385 paires de bases. Les auteurs ont proposé de représenter cette boucle par une tige élastique dont le repère local indique l'orientation des paires de bases. La protéine, simulée en dynamique moléculaire, impose à la tige des conditions aux limites en position et en orientation. Ces conditions d'encastrement sont utilisées pour calculer la géométrie de la tige. Cela se traduit par un problème aux limites, résolu par continuation numérique : deux solutions sont obtenues, et distinguées selon un critère énergétique. Les forces exercées par la boucle sur la protéine aux points d'encastrement sont calculées via la théorie de l'élasticité, puis utilisées en dynamique moléculaire pour mettre à jour la structure 3D de la protéine et donc, obtenir des nouvelles conditions aux limites de la tige. La courbure et la torsion intrinsèques de l'ADN, son anisotropie ainsi que les forces électrostatiques dues aux charges négatives des phosphates sont prises en compte. D'après Balaeff *et al.* (2006), le fait d'ajouter le modèle de tige élastique à la dynamique moléculaire conduit vers une structure potentiellement plus proche de la réalité, avec un temps de calcul moindre.

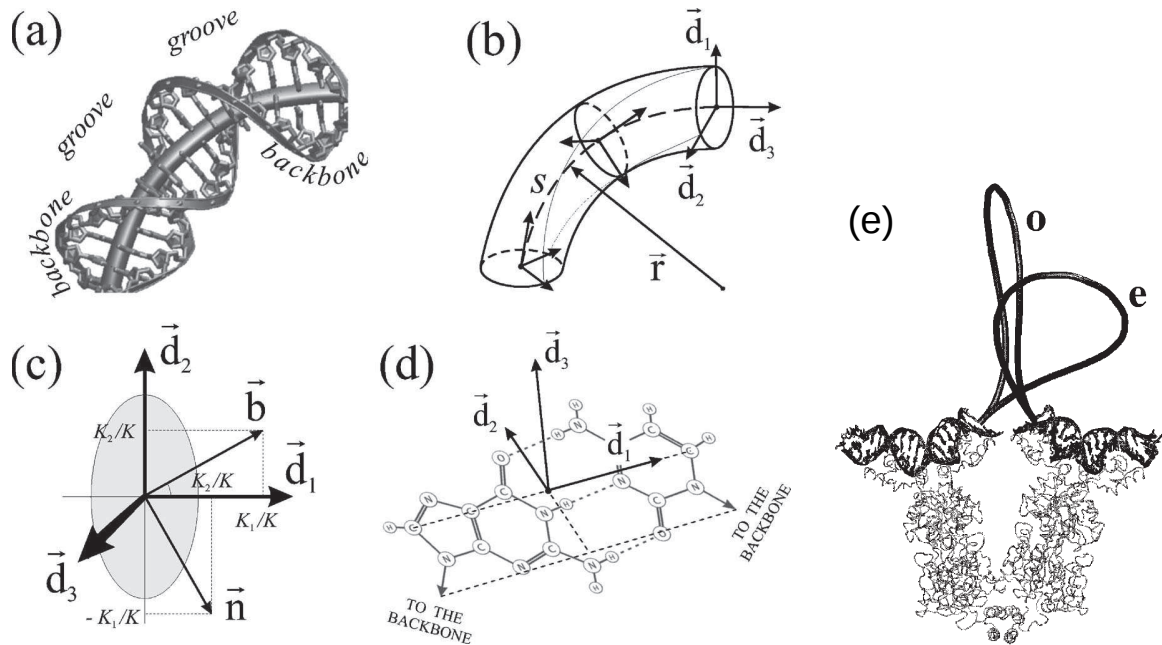


Figure 1.20. Modèle multi-échelle proposé par Balaeff *et al.* pour déterminer la conformation de l'opéron lactose de la bactérie *E. Coli*, déformé par le répresseur *lac*. (a) La double-hélice est représentée par une poutre élastique. (b) Cette poutre élastique est caractérisée par sa trajectoire $\mathbf{r}(s)$ et son repère local $(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3)$. (c) Le repère local est en rotation par rapport au repère de Frenet autour de la tangente $\mathbf{t} = \mathbf{d}_3$. (d) Le repère local est utilisé pour représenter les paires de bases. Les vecteurs unitaires sont tournés de $\pi/2$ autour de $\mathbf{d}_3$ par rapport à la convention de Tsukuba (Olson *et al.* 2001). (e) La géométrie de la poutre est calculée en résolvant le problème aux limites imposé par la protéine : deux solutions sont obtenues par continuation numérique, puis distinguées selon un critère énergétique. Figures de Balaeff *et al.* (1999, 2006).

Ainsi, la démarche de modélisation moléculaire proposée par Balaeff *et al.* se traduit en un problème de mécanique des milieux continus : imposer des conditions d'encastrement (position

et orientation relatives des extrémités) à une tige élastique. Le même problème est apparu dans l'approche *Biopolymer Chain Elasticity* (BCE) développée au Laboratoire Jean Perrin (LJP), et appliquée sur les boucles d'acides nucléiques simple-brin en épingle à cheveux. La résolution efficace de ce problème pour toutes conditions d'encastrement est l'un des principaux objectifs de la thèse, parce qu'elle semble indispensable à l'implémentation de modèles robustes fondés sur la théorie de l'élasticité.

3 L'approche BCE

Biopolymer Chain Elasticity (BCE) est une approche de modélisation moléculaire qui postule que les chaînes de biopolymères se comportent comme des poutres élastiques. Elle utilise donc la théorie de l'élasticité, avec différents niveaux d'approximation, pour calculer les conformations de ces molécules. Une conséquence fondamentale de cette approche, présentée en sous-section 3.1, est de proposer une représentation multi-échelle et hiérarchique de la molécule qui décompose efficacement le calcul de sa structure tridimensionnelle. BCE a été appliquée avec succès aux boucles d'ADN et d'ARN en épingle à cheveux (Pakleza et Cognet 2003, Santini *et al.* 2003, 2009, Baouendi *et al.* 2012), en suivant trois étapes expliquées en sous-section 3.2. Toutefois, des problèmes de mécanique doivent être résolus pour généraliser la méthode à tous types de structures : cela est discuté en sous-section 3.3. L'objectif principal de la thèse est de fournir les outils nécessaires à la résolution de ces problèmes.

3.1 Une approche multi-échelle et hiérarchique

BCE est une approche multi-échelle, c'est-à-dire qu'elle ne représente pas la molécule de la même façon selon l'échelle de longueur considérée. A une échelle bien choisie (de l'ordre de la longueur de persistance) dite *globale*, la molécule est supposée suffisamment longue pour être représentée par un modèle continu, et en même temps suffisamment rigide pour que ce modèle soit une tige élastique en grande déformation. Pour des longueurs plus importantes, la molécule est regardée comme un assemblage de plusieurs tiges élastiques. Dans ce cas, les liaisons entre éléments peuvent être de différentes natures selon le problème traité (continuité géométrique ou discontinuité angulaire), mais s'accompagnent systématiquement de discontinuités en efforts et/ou en moments. Pour des longueurs plus faibles (généralement, quelques résidus) dites *intermédiaires*, la chaîne est manipulée comme un ensemble de blocs rigides d'atomes (résidus ou parties de résidus). Enfin, à l'échelle *locale* (quelques atomes), la position 3D de chaque atome est ajustée par minimisation d'énergie. Ainsi, le niveau de détail est adapté à l'échelle, ce qui réduit le nombre de degrés de liberté du modèle géométrique.

BCE est aussi une approche hiérarchique, c'est-à-dire qu'elle établit un ordre de priorité entre les différentes échelles. Dans les approches de modélisation habituelles, une seule fonction d'énergie est utilisée pour décrire en même temps les interactions fortes, comme les liaisons covalentes, et les interactions faibles telles les liaisons hydrogène. Dans le cas de déformations importantes de la molécule à l'échelle globale, il est alors fréquent d'aboutir à des contraintes mécaniques élevées à l'échelle locale, qu'il faut ensuite relâcher progressivement. Avec l'approche BCE, ce problème est résolu en décomposant la modélisation en plusieurs étapes. D'abord, la molécule est déformée à l'échelle globale en utilisant le modèle de tige élastique. Ensuite, les positions relatives

des bases sont optimisées à l'aide du modèle de blocs rigides d'atomes. En dernier, les positions atomiques locales sont corrigées. Cet ordre hiérarchique des échelles considérées (globale $\rightarrow$ intermédiaire $\rightarrow$ locale) offre une façon plus naturelle de déformer les molécules. D'un point de vue géométrique, il est alors plus facile de contrôler la conformation globale sans engendrer de modifications irréalistes des coordonnées internes.

Étant donnée cette idée fondamentale d'une approche multi-échelle et hiérarchique, plusieurs types de simulations peuvent être implémentées : mécanique/dynamique moléculaire, modes normaux, Monte Carlo, etc. A l'échelle globale, une tige élastique peut être utilisée pour représenter soit l'ensemble des deux brins d'un acide nucléique, soit une seule chaîne sucre-phosphate. Ce dernier cas nous intéresse particulièrement car il a été très peu proposé dans la littérature, essentiellement parce que les chaînes simple-brin sont considérées comme trop flexibles pour être modélisées par une tige élastique. Toutefois, les résultats obtenus avec BCE pour des boucles en épingle à cheveux montrent que le modèle de tige élastique appliqué à l'ADN ou l'ARN simple-brin peut conduire à des structures remarquablement bonnes.

3.2 Exemple : détermination d'une structure en épingle à cheveux

D'excellentes superpositions entre résultats théoriques et expérimentaux ont en effet été obtenues pour des structures en épingle à cheveux d'acides nucléiques. La méthode utilisée pour déterminer ces structures 3D a évolué au cours des années, et est présentée ici dans sa version la plus récente. Elle se décompose en trois étapes : (i) la construction d'une structure de départ à l'aide de la théorie de l'élasticité, (ii) l'optimisation de la conformation par rotation des bases (considérées comme des blocs rigides) et (iii) l'obtention de la structure finale par minimisation d'énergie à l'échelle locale (atomique).

3.2.1 Construction d'une structure de départ

Tout d'abord, les positions 3D des atomes de la tige en double-hélice (d'ADN-B ou d'ADN-A) sont calculées à partir des résultats de [Chandrasekaran et Arnott \(1989\)](#) pour les plateaux de bases, et de dynamique moléculaire pour les angles de torsion de la chaîne sucre-phosphate. Les hélices obtenues sont tracées dans un repère cartésien orthonormé direct, choisi suivant la convention de Cambridge ([Diekmann 1989](#)) pour le dernier plateau de paire de bases de la tige. Ensuite, une courbe en hélice $\{\mathbf{r}_H(s)\}$ est superposée à chaque brin de la structure 3D de la double-hélice (Figure 1.21a) :

$$\mathbf{r}_H(s) = \begin{cases} x_H(s) = \rho_0 \cos(s/l_t + \varphi) \\ y_H(s) = \rho_0 \sin(s/l_t + \varphi) \\ z_H(s) = p_z/(2\pi) (s/l_t) \end{cases} \quad (1.29)$$

où

ρ_0 est le rayon de l'hélice,

p_z est le pas de l'hélice,

φ est l'angle de déphasage de l'hélice,

$l_t = \sqrt{\rho_0^2 + (p_z/2\pi)^2}$ est la longueur d'un tour d'hélice divisée par 2π .

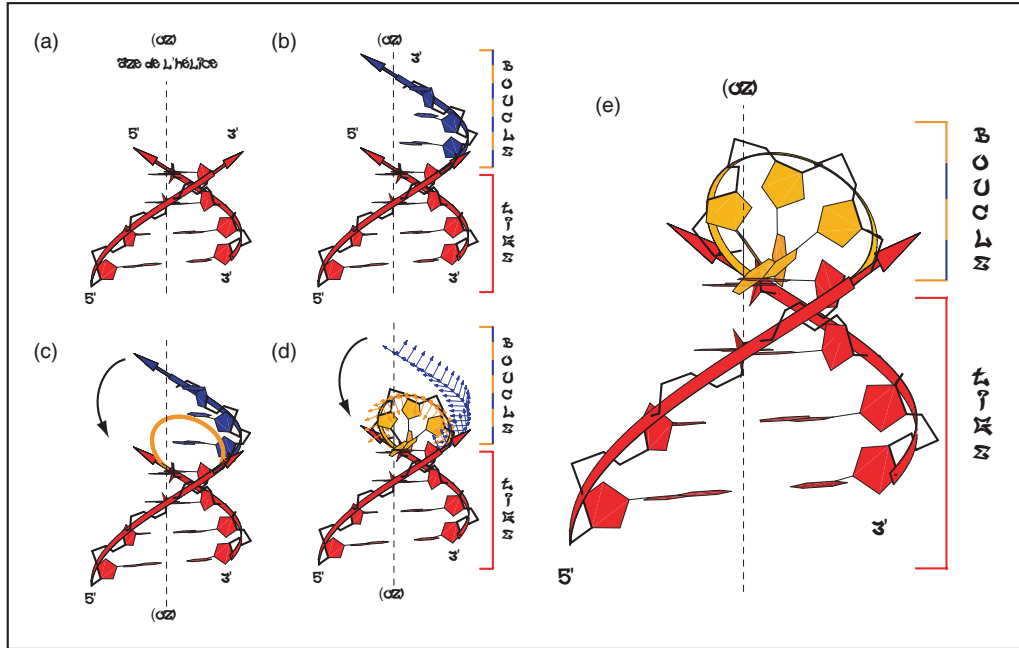


Figure 1.21. Construction de la structure de départ. (a) Une poutre élastique est associée à chaque brin de la tige en double-hélice. (b) Un simple-brin en hélice régulière (en bleu) est généré avec la séquence de la boucle. (c) La géométrie de la poutre élastique décrivant la boucle est calculée en résolvant le problème aux limites imposé par la tige. (c-e) La conformation de la tige. (e) La conformation de la boucle est obtenue en "repliant" le simple-brin sur la poutre élastique. Figure de [Santini \(2005\)](#).

Les valeurs des paramètres ρ_0 , p_z et φ sont ajustées numériquement de façon à minimiser la distance entre la courbe en hélice et les atomes C5', C4', C3', O3' et P du squelette du brin considéré.

Les extrémités des courbes associées aux deux hélices de la tige définissent des conditions aux limites pour la poutre élastique décrivant la boucle. Plus précisément, elles définissent la position et l'orientation relatives des tangentes extrêmes. Toutefois, elles ne donnent pas d'information sur la torsion physique de la poutre, *i.e.* l'orientation des sections autour de la tangente. Pour y remédier, on fait l'hypothèse que cette information est donnée par le repère de Frenet des hélices. Il serait cependant judicieux d'utiliser les atomes des chaînes latérales pour définir plus précisément cette torsion physique. D'autre part, les courbes en hélice ne donnent pas d'information sur l'état de contrainte et d'énergie de la tige double-brin. Pour une modélisation complète de cette dernière, il est donc nécessaire de déterminer les poutres élastiques représentant les deux brins d'acide nucléique, c'est-à-dire possédant les mêmes trajectoires que les courbes en hélice et dont la torsion physique est cohérente avec l'orientation des chaînes latérales.

Ces poutres donnent les conditions d'encastrement complètes de la boucle, exprimées comme la position et l'orientation relatives des deux repères attachés aux sections extrêmes. La longueur de la boucle étant évaluée à partir de sa séquence, la géométrie de la poutre qui la représente est calculée en résolvant le problème aux limites ainsi posé avec la théorie de l'élasticité des poutres.

Une fois que la poutre élastique de la boucle est déterminée, il s'agit d'obtenir la position 3D de chacun de ses atomes. Pour ce faire, un simple brin en hélice régulière est généré avec la

séquence de la boucle (en bleu sur la Figure 1.21b), puis une courbe en hélice lui est associée comme précédemment pour la tige, et la poutre élastique correspondante est déterminée. Ensuite, les atomes de ce brin sont regroupés en blocs rigides : on passe à une modélisation à l'échelle intermédiaire. Les blocs sont définis sur la Figure 1.22a. A chaque bloc, on associe un repère attaché à une section de la poutre élastique. L'abscisse curviligne de ce repère local est donnée par la projection sur la poutre élastique d'un des atomes du bloc, appelé atome pivot. Dans ce repère, les coordonnées des atomes du bloc sont constantes et peuvent être entièrement déterminées. Pour aboutir à la structure 3D de la boucle, il suffit de définir les repères de chaque bloc sur la poutre élastique qui la représente, en utilisant les mêmes abscisses curvilignes que pour le simple brin d'hélice régulière généré. Ces repères sont alors utilisés pour obtenir la position de chaque atome. En résumé, l'opération consiste à "replier" un simple-brin sur la poutre élastique de la boucle (Figure 1.21c-e).

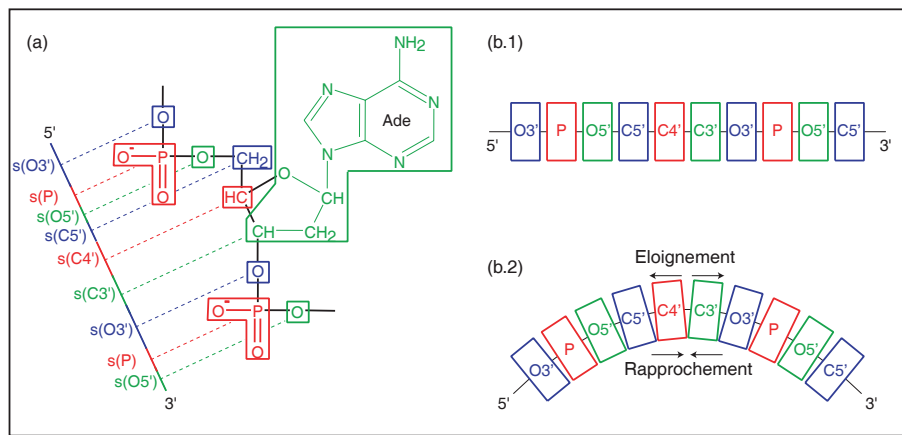


Figure 1.22. Modélisation à l'échelle intermédiaire. (a) Définition des blocs rigides d'atomes. (b) Influence de la déformation globale de la molécule sur les positions relatives des blocs. Figure de Santini (2005).

Comme illustré sur la Figure 1.22b, ce "repliement" modifie les positions relatives entre blocs adjacents. Toutefois, les valeurs des coordonnées internes (longueurs et angles de liaisons, angles de torsion) sont globalement préservées par rapport à celles de l'hélice avant repliement (Santini 2005). Ainsi, la structure obtenue grâce à la théorie de l'élasticité constitue un bon point de départ, à partir duquel des raffinements aux échelles intermédiaire et locale peuvent être effectués.

3.2.2 Optimisation de la conformation par rotation des bases

A l'issu de l'opération de repliement, les orientations des bases de la boucle ne sont généralement pas correctes, car elles ne satisfont pas les contraintes d'encombrement stérique. Elles doivent donc être corrigées. Pour ce faire, trois degrés de liberté (Ω , χ , Θ_{empil}) par base sont définis (Figure 1.23). L'angle Ω caractérise l'orientation du sucre et de la base autour de la tangente à la trajectoire de la poutre élastique. Il s'applique aux blocs C3' et C4', autour des tangentes aux abscisses curvilignes $s(\text{C3}')$ et $s(\text{C4}')$ respectivement. L'angle χ caractérise l'orientation de la base autour de la liaison glycosidique (C1'-N9 pour les purines et C1'-N1 pour les pyrimidines). Enfin, l'angle de redressement d'empilement Θ_{empil} caractérise l'orientation du sucre et de la

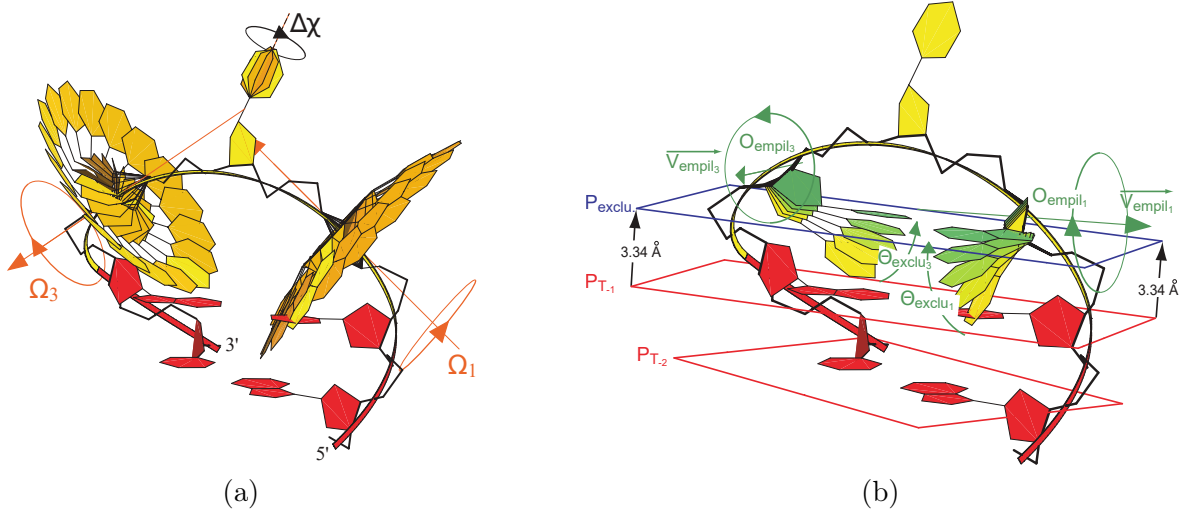


Figure 1.23. Degrés de liberté utilisés pour ajuster l'orientation des bases. (a) Angles Ω et χ de rotation autour de la tangente et de la liaison glycosidique respectivement. (b) Angle Θ_{empil} de redressement d'empilement. Figures de Santini (2005).

base autour de l'axe (O_{empil} , $\mathbf{V}_{\text{empil}}$) défini par

$$s(O_{\text{empil}}) = \frac{s(C5') + s(O3')}{2} , \quad (1.30)$$

$$\mathbf{V}_{\text{empil}} = \mathbf{V}_{\text{ideal}} \times \mathbf{V}_{\text{glyco}} , \quad (1.31)$$

où

O_{empil} est un point sur la trajectoire de la poutre élastique,

$\mathbf{V}_{\text{ideal}}$ est le vecteur unitaire normal au plan qu'aurait décrit, dans le cas d'une hélice, une base supplémentaire empilée sur le dernier plateau de la tige,

$\mathbf{V}_{\text{glyco}}$ est le vecteur unitaire donnant la direction de la liaison glycosidique.

Les valeurs des angles Ω et χ sont déterminées à partir des données structurales dont on souhaite tenir compte (RMN, PDB, etc.). L'angle de redressement d'empilement n'est en principe utilisé que pour les bases extrémales de la boucle, dans le cas où elles se ré-apparient. Il est choisi à la valeur $\Theta_{\text{empil}} = \Theta_{\text{exclu}}$ qui place les atomes de la base redressée à une distance moyenne de 3.34 Å du plan de la dernière paire de bases de la tige.

Ces trois rotations ne modifient pas la trajectoire de la boucle. En revanche, la rotation d'angle Ω induit une discontinuité des angles de torsion qui n'est pas toujours réaliste. Par conséquent, elle est généralement suivie d'un traitement qui répartit continûment la variation des angles de torsion.

3.2.3 Minimisation d'énergie

La conformation obtenue après optimisation par rotation des bases est déjà très proche d'un minimum global d'énergie. Pour l'améliorer, une courte minimisation d'énergie est effectuée avec les méthodes classiques de mécanique moléculaire à l'échelle atomique. Cette relaxation finale ajuste les positions 3D des atomes en modifiant très peu la structure globale mise en place lors des manipulations aux échelles supérieures. Enfin, des dynamiques moléculaires peuvent être démarrées à partir de la structure obtenue. Les simulations testées ont mis en évidence la bonne stabilité structurale des épingles à cheveux élaborées avec l'approche BCE, ainsi que l'accord qui en résulte avec les données de la RMN.

3.3 Vers une généralisation de la méthode

En fin de compte, la méthode utilisée pour calculer des épingles à cheveux avec BCE pose deux problèmes de mécanique des poutres élastiques :

- (i) superposer une poutre élastique à chaque brin de la tige en double-hélice,
- (ii) déterminer la géométrie de la poutre décrivant la boucle, en résolvant le problème aux limites posé par la double-hélice.

Le problème (i) se pose car les courbes en hélice associées aux brins d'acide nucléique par un algorithme de minimisation de distance ne donnent pas l'orientation des sections, ni le chargement mécanique de la structure. Il faut donc déterminer les poutres élastiques ayant les mêmes rayon, pas et déphasage que ces hélices, et dont la torsion physique est cohérente avec l'orientation des chaînes latérales. Une fois ce premier problème résolu, les conditions d'encastrement de la boucle sur la tige sont définies et donc, le problème (ii) est posé.

Pour généraliser cette démarche de résolution de conformation à tous types de structures, on souhaite résoudre efficacement ces deux problèmes. C'est l'un des objectifs de la thèse.

4 Objectifs de la thèse

La motivation principale de la thèse réside dans l'idée que l'utilisation de la mécanique des tiges élastiques peut résoudre certaines des difficultés rencontrées lors de la modélisation des acides nucléiques. Cela se décline en plusieurs objectifs détaillés en sous-section 4.1. Toutefois, le comportement des tiges élastiques est suffisamment riche pour décrire une grande variété de problèmes. De fait, nous souhaitons également exploiter les développements de cette thèse dans différentes thématiques de biophysique, mécanique et robotique. Ces objectifs à plus long terme sont mentionnés dans la sous-section 4.2.

4.1 Pour la modélisation des acides nucléiques

La richesse structurale des acides nucléiques étant importante, il est nécessaire de classer les différentes conformations existantes, c'est-à-dire de définir des propriétés géométriques ou méca-

niques pour les comparer. Cela ne peut se faire qu'en proposant des modèles simplifiés, comme les empilements de bases de 3DNA ou les courbes de Curves+. Le modèle de poutre élastique nous semble particulièrement approprié pour cette tâche, car il donne un grand nombre d'informations (trajectoire, orientation des sections donc des chaînes latérales, chargement mécanique, énergie) à partir de peu de paramètres. Ainsi, un premier objectif de la thèse est de proposer une classification des géométries de poutres élastiques à l'équilibre, à savoir :

- repérer toutes les géométries particulières, potentiellement assimilables à des structures d'acides nucléiques connues (notamment, les hélices),
- dégager des propriétés géométriques simples des poutres élastiques,
- montrer que la richesse structurelle de ces poutres est suffisante pour représenter une très grande partie des conformations d'acides nucléiques.

Cette classification doit fournir une vision globale du problème de repliement des poutres élastiques, avec l'objectif de représenter schématiquement l'espace conformationnel des acides nucléiques. Elle est étudiée dans les chapitres 2 et 3.

En plus d'un outil de classification, l'approche BCE a pour objectif de déterminer les conformations 3D des molécules. Il a été vu en sous-section 3.3 que cela nécessite la résolution de problèmes de mécanique des tiges élastiques. En particulier, il s'agit de résoudre le problème aux limites pour imposer des conditions d'encastrement, *i.e.* des positions et orientations relatives des sections extrémales. Cela est l'objet du chapitre 4.

4.2 Sur une perspective plus large

Les poutres élastiques peuvent représenter une grande variété de problèmes à diverses échelles de longueur.

En modélisation moléculaire, leur application peut certainement s'étendre à d'autres biopolymères que les acides nucléiques. En particulier, ces modèles décrivent correctement le comportement mécanique des protéines (Lapidus *et al.* 2002, Storm *et al.* 2005) et des fibres bactériennes (Guhados *et al.* 2005). D'autre part, le développement de l'origami depuis 2006 (Rothemund 2006, Douglas *et al.* 2009) a abouti à la construction d'architectures nanométriques d'ADN avec une précision atomique. Des nanosystèmes et même des nanorobots d'ADN ont été élaborés, et pourraient être utilisés dans diverses tâches médicales comme l'acheminement de médicaments contre le cancer (Jiang *et al.* 2012, Amir *et al.* 2014). La modélisation de ces nouveaux systèmes est nécessaire pour améliorer leur conception, et peut faire intervenir l'élasticité des poutres (Kauert *et al.* 2011, Kim *et al.* 2011). Dans la même idée, les nanotubes de carbone jouent un rôle prépondérant dans la construction de NEMS (Nano-Electro-Mechanical Systems) puisqu'ils pourraient servir à la conception de capteurs, systèmes de stockage ou de conversion d'énergie, sondes, etc. (Baughman *et al.* 2002). A nouveau, les poutres élastiques sont utilisées pour les modéliser (Aydogdu 2009).

Le succès des NEMS s'accompagne de diverses problématiques, qui doivent être étudiées avec des outils adaptés à cette échelle où les phénomènes physico-chimiques à modéliser sont très particuliers. En réponse à ce besoin, l'équipe NANO-D de l'Inria de Grenoble développe SAMSON (Software for Adaptive Modeling and Simulation Of Nanosystems), une plateforme pour la modélisation des nanosystèmes (NANO-D). SAMSON fonctionne avec une architecture

ouverte, de façon à ce que chaque utilisateur puisse étendre le programme et l'adapter à ses besoins en développant de nouveaux modules. Ainsi, un module utilisant la méthode BCE pour la modélisation moléculaire ou, plus spécifiquement, la conception de nanosystèmes biologiques est envisagé. A l'ISIR, une simulation interactive de la déformation des molécules avec retour d'effort a été développée dans SAMSON (Bolopion *et al.* 2010, 2011). L'association entre l'approche BCE et l'algorithme de contrôle utilisé dans cette simulation pourrait offrir une immersion encore plus intuitive dans le milieu moléculaire.

A une échelle un peu plus grande, le développement des MEMS (Micro-Electro-Mechanical Systems) peut conduire à divers problèmes traités avec la théorie de l'élasticité, comme l'influence de la capillarité sur les micro-leviers (Neukirch *et al.* 2007). Les mouvements oscillatoires de poutres élastiques peuvent également être utilisés pour représenter les modes de locomotions de systèmes ou d'animaux à des échelles très différentes : microrobots (Oulmas *et al.* 2017), cellules (Qian *et al.* 2008), serpents (Dal Corso *et al.* 2017), lamproies, anguilles et serpents aquatiques (McMillen et Holmes 2006) par exemple. A l'échelle macroscopique, la théorie de l'élasticité intervient dans de nombreux problèmes, comme la simulation du système vibrissal de mammifères (Goss et Chaouki 2016, Claverie *et al.* 2016), l'insertion de cathéters flexibles en radiologie (Kratchman *et al.* 2017), la conception de ressorts non-linéaires (Fang et Chen 2013), la prédiction du mouvement des cheveux (Bertails *et al.* 2006) ou encore le guidage de câbles sous-marins (Goyal *et al.* 2008, Wicks *et al.* 2008).

En robotique, un nouveau thème émerge autour des structures continûment déformables, qui offrent de nouvelles possibilités par rapport aux chaînes articulées (Majidi 2014). Ce thème en plein essor, désigné par "robotique souple" (soft robotics), s'intéresse à la conception de robots avec des matériaux aux propriétés élastiques et rhéologiques proches des organes et des tissus biologiques : fluides, gels et élastomères. Ces nouveaux robots seront plus adaptés pour interagir avec l'être humain en toute sécurité, et évoluer dans des environnements surcontraints. Ce contexte offre de nouvelles applications au modèle de poutre élastique en grandes déformations.

En fournissant une classification des géométries de poutres élastiques, ainsi qu'un nouveau moyen de les calculer à partir de contraintes géométriques, nous espérons apporter une contribution dans quelques-unes de ces thématiques, en plus de la modélisation des acides nucléiques.