



**HAL**  
open science

# Étude et analyse de l'ADN pour l'auto-assemblage microscopique

Ahlem Abbaci

► **To cite this version:**

Ahlem Abbaci. Étude et analyse de l'ADN pour l'auto-assemblage microscopique. Automatique / Robotique. Université Pierre et Marie Curie - Paris VI, 2010. Français. NNT: . tel-00647240v2

**HAL Id: tel-00647240**

**<https://theses.hal.science/tel-00647240v2>**

Submitted on 30 Jan 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE

présentée à

L'UNIVERSITÉ PIERRE ET MARIE CURIE

par

Ahlem ABBACI

pour obtenir le diplôme de

DOCTORAT DE L'UNIVERSITÉ  
PIERRE ET MARIE CURIE

Spécialité : Robotique

---

Étude et analyse de l'ADN pour l'auto-assemblage  
microscopique

---

Soutenue le 06 juillet 2010

## JURY

M.	J. COGNET	Pr. à l'Université Pierre et Marie Curie	Président
M.	J. DEMONGEOT	Pr. à l'Université de Grenoble Directeur du laboratoire TIMC-IMAG	Rapporteur
M.	A. FERREIRA	Pr. à l'École Nationale Supérieure d'Ingénieurs de Bourges	Rapporteur
M.	Y. BELLOUARD	Pr. à l'Université d'Eindhoven	Examineur
M.	A. CUENAT	Dr. au Laboratoire National de Physique de Londres (NPL)	Examineur
M.	S. HALIYO	Co-Directeur de thèse MC. à l'Université Pierre et Marie Curie	Examineur
M.	S. REGNIER	Directeur de thèse Pr. à l'Université Pierre et Marie Curie	Examineur



---

# REMERCIEMENTS

---

Je tiens à remercier mon directeur de thèse, le professeur Stéphane REGNIER, pour la confiance et la liberté qu'il m'a accordé, puis pour m'avoir guider, conseiller et encourager dans cette entreprise. Je remercie M. Sinan HALIYO, mon encadrant, pour la confiance et le temps qu'il m'a accordé.

Je remercie M. Yves BELLOUARD, M. Alexandre CUENAT et son équipe, ainsi que toute l'équipe du projet GOLEM pour leurs conseils et les différentes rencontres enrichissantes sur les plans scientifique et humain.

Je tiens aussi à remercier M. Florian KROHS de l'équipe du laboratoire AMIR à Oldenburg, le Professeur ANGULO-MORA, ainsi que M. Alexandre CUENAT et son équipe du NPL à un second titre, sans qui je n'aurais pu réaliser l'ensemble des travaux expérimentaux.

Je remercie particulièrement le professeur Jean COGNET, pour les longues et enrichissantes discussions, pour ses conseils et son expertise sur la molécule d'ADN, puis pour m'avoir fait l'honneur d'être le président de mon jury.

Je remercie également le Professeur Jacques DEMONGEOT pour m'avoir accueilli en stage au sein de son laboratoire pendant mes deux années de master, stage qui m'a fait découvrir l'exercice passionnant de la recherche, puis pour m'avoir fait l'honneur de rapporter ce travail de thèse.

Je remercie le professeur Antoine FERREIRA, pour m'avoir fait l'honneur de rapporter ce document et pour ses questions et conseils.

Mes remerciements vont aussi à Mme. Françoise GIROUD pour m'avoir initié à l'exercice de la recherche, puis pour m'avoir encourager dans ce sens.

Je tiens aussi à remercier mes amis et collègues de l'ancien LRP (Laboratoire de Robotique de Paris) et du nouvel institut ISIR pour toutes les expériences de "thésards" partagées.

Parce qu'une thèse est aussi une expérience humaine, à Houyam, Louiza, Zohra et ... tous

mes amis qui m'ont accompagné tout au long de cette expérience, merci pour votre soutien et encouragements.

A Nazim, merci.

A Hatem et Imad, mes frères, et à ma petite sœur Abir, merci pour votre soutien et encouragements. Merci d'avoir été présents.

A mes parents, sans qui ce cheminement n'aurait été possible. Ils ont été la source d'énergie, d'inspiration et d'encouragements. Merci.

Je dédie ce travail à M. Mohamed ABBACI et Mme Khédidja ABBACI, mes parents.

---

# TABLE DES MATIÈRES

---

Table des matières	i
Table des figures	v
Introduction	1
<b>1 Problématique de l'auto-assemblage microscopique par processus biologiques</b>	<b>3</b>
1 Le micro-assemblage dans le projet Golem . . . . .	4
2 L'auto-assemblage microscopique . . . . .	7
2.1 Propriétés de l'auto-assemblage . . . . .	7
2.2 Etat de l'art . . . . .	8
2.2.1 L'auto-assemblage microscopique <i>inorganique</i> . . . . .	8
2.2.2 L'auto-assemblage microscopique par processus biologiques . . .	11
2.3 Choix d'un processus biologique . . . . .	16
3 L'ADN pour l'auto-assemblage . . . . .	16
3.1 Structure de l'ADN . . . . .	16
3.2 L'hybridation de l'ADN . . . . .	19
3.2.1 La Spécificité . . . . .	19
3.2.2 La Stabilité . . . . .	20
3.3 Les paramètres de contrôle de l'hybridation . . . . .	20
3.3.1 Paramètres extrinsèques . . . . .	20
3.3.1.1 La Température . . . . .	20
3.3.1.2 La Composition du milieu . . . . .	21
3.3.2 Paramètres intrinsèques . . . . .	21
4 Problématique de la thèse . . . . .	22
4.1 Caractérisation des forces d'interaction de l'ADN . . . . .	22
4.2 Plan de la thèse . . . . .	23
<b>2 Etude des paramètres de contrôle de la stabilité</b>	<b>25</b>
1 Approche individuelle . . . . .	27
1.1 Caractérisation de la stabilité . . . . .	27
1.2 Problématique de l'estimation de la force . . . . .	28

1.3	Discussion . . . . .	29
1.3.1	L'élasticité de l'ADN . . . . .	30
1.3.2	La densité de répartition des brins d'ADN . . . . .	31
1.4	Conclusion . . . . .	31
2	Approche populationnelle . . . . .	32
2.1	Quantification de la stabilité . . . . .	32
2.2	Modèle du plus proche voisin . . . . .	33
2.2.1	Algorithme du modèle du plus proche voisin . . . . .	33
2.2.2	Les paramètres intrinsèques . . . . .	34
2.2.3	Les paramètres environnementaux . . . . .	37
2.3	Estimation de la force d'interaction à partir de $\Delta G$ . . . . .	39
2.4	Récapitulatif . . . . .	40
3	Conclusion . . . . .	41
<b>3</b>	<b>Caractérisation expérimentale de la stabilité</b>	<b>43</b>
1	L'AFM, les variables et les paramètres de mesure . . . . .	44
1.1	Principe général de l'AFM . . . . .	44
1.2	L'AFM pour les interactions biomoléculaires . . . . .	45
1.3	Les variables de mesure et les paramètres expérimentaux . . . . .	46
1.3.1	La force de rupture . . . . .	47
1.3.2	Taux de chargement . . . . .	49
1.3.3	Conclusion . . . . .	51
2	Les protocoles expérimentaux pour la quantification de la stabilité . . . . .	51
3	Extraction des données significatives par analyse statistique . . . . .	54
3.1	Méthode statistique pour l'identification des expériences significatives . . . . .	54
3.2	Identification des données exploitables . . . . .	57
3.3	Moyennes et distributions des forces de ruptures . . . . .	60
3.4	Conclusion . . . . .	63
4	Analyse énergétique des données expérimentales . . . . .	63
4.1	Variabilité de configuration et force de rupture . . . . .	63
4.2	Etude des événements locaux de rupture . . . . .	66
4.3	Discussion des résultats expérimentaux . . . . .	68
5	Conclusions . . . . .	71
<b>4</b>	<b>Etude et caractérisation expérimentale de la spécificité</b>	<b>73</b>
1	Etat de l'art des méthodes de génération de séquences . . . . .	73
2	Génération de séquence pour l'A.A microscopique . . . . .	74
2.1	Principe de la génération de séquence . . . . .	75
2.2	Algorithme de génération de séquence . . . . .	76
3	Analyse théorique d'une séquence générée . . . . .	77
4	Analyse expérimentale . . . . .	80
4.1	Description des expériences . . . . .	81
4.2	Analyse énergétique des événements de rupture locaux . . . . .	83
4.3	Analyse expérimentale de la spécificité . . . . .	83
4.3.1	Etude de la spécificité . . . . .	85
4.3.2	Influence de la vitesse sur la spécificité et la stabilité . . . . .	86
4.3.3	Influence du temps d'attente . . . . .	89
4.4	Conclusion . . . . .	92

---

5	Conclusions . . . . .	93
	<b>Conclusion et perspectives</b>	<b>95</b>
	<b>ANNEXE A</b>	<b>99</b>
1	Les séquences $S_1$ et $S_2$ . . . . .	99
2	Les séquences $S_3$ et $S_4$ . . . . .	99
	<b>ANNEXE B</b>	<b>101</b>
	<b>ANNEXE C</b>	<b>103</b>
	<b>Liste des symboles</b>	<b>105</b>
	<b>Liste des abréviations</b>	<b>107</b>
	<b>Liste des publications</b>	<b>109</b>
	<b>Bibliographie</b>	<b>111</b>





---

## TABLE DES FIGURES

---

1.1	Micro-assemblage sériel. Transport, transfert et positionnement robotisés de micro-objets [Golem 06]. . . . .	4
1.2	Auto-assemblage microscopique par processus biologiques. Exemple d'une manipulation parallèle sans contact de micro-objets en milieu aqueux [Golem 06]. . .	6
1.3	Structure 3D du complexe Cd1d-C6ph chez la souris [Schiefner 09]. En violet :Cd1d et en bleu : C6ph. . . . .	8
1.4	Auto-assemblage hétérogène de Stauth et al. [Stauth 06]. Les microcomposants sont introduits dans un environnement liquide et transportés par mouvement fluïdique. L'auto-assemblage a lieu dès que les microcomposants rencontrent leurs formes complémentaires. Ils s'attachent ensuite par forces capillaires résultantes de l'alliage métallique déposé sur le substrat. . . . .	10
1.5	Nanostructure d'ADN [Rothemund 06]. Echelle 100nm. Sont illustrés successivement pour chaque structure, le modèle et l'image AFM correspondante. . . . .	12
1.6	Nanostructures polyhedrales [Yu He 08]. Images AFM et Cryo-EM. Les boîtes blanches indiquent les nanostructures. . . . .	12
1.7	Moteur moléculaire protéique composé de myosine et de kinésine [Hess 04]. . . .	13
1.8	Les velcros biologiques : l'hybridation de l'ADN/ADN, protéine/ADN et protéine/protéine. . . . .	13
1.9	Auto-assemblage utilisant l'ADN et les interactions avidin/biotin sur la surface en or des deux particules (Gauche). (Milieu) Image MEB des particules en silicium micro-fabriquée dans [McNally 03]. (Droite) Principe de l'auto-assemblage par ADN et avidin/biotin. . . . .	15
1.10	Auto-assemblage spécifique et réversible [Valignat 05]. (A) Des microparticules recouvertes d'oligonucléotides : les oligonucléotides utilisés sont synthétisés de telle sorte que la partie attachée à la bille est de type double brin, et la partie libre est de type simple brin. Les parties simples brins des oligonucléotides sur les billes rouges et sur les billes vertes sont complémentaires et donc spécifiquement liées. (B) Assemblage spécifique, les billes rouges sont liées aux billes vertes. (C) En stabilisant la température à 23C, l'assemblage est stable. (D) En augmentant la température à 50C, les billes sont complètement dispersées. . . . .	15

1.11	Représentation de la composition chimique d'une molécule d'ADN. Assemblage de deux brins orientés 3'-5' antiparallèles (TGAC) et (GTCA). L'orientation est donnée par les carbones 3' et 5' du désoxyribose. . . . .	17
1.12	Structure 3D d'une double hélice d'ADN ; brin A (cgcganacgcc), brin B (ggcg-tatcgcg) [Beger 98]. . . . .	18
1.13	Structures secondaires imparfaites : mismatch [Dimitrov 04] . . . . .	21
2.1	Principe de l'A.A microscopique basé sur l'ADN. . . . .	26
2.2	Axe hypothétique de la double hélice. . . . .	28
2.3	Evolution de la force d'interaction estimée par extrapolation de l'énergie d'interaction en fonction de la longueur de séquence dans l'approche individuelle. . . . .	30
2.4	Modélisation par segmentation de l'hybridation dans les puces à ADN dans [Jayaraman 06]. Les brins A et B sont complémentaires. . . . .	31
2.5	Principe des puces à ADN. Hybridation entre deux populations de brins complémentaires A et B. . . . .	33
2.6	Algorithme pour le calcul de $\Delta G$ . Une fenêtre de couple de paires de bases glisse le long de la représentation des deux séquences complémentaires hybridées. Ces séquences présentent 5 couples de paires de bases. . . . .	34
2.7	Tableau des paramètres standards du modèle du plus voisin pour l'ADN [SantaLucia 98]. . . . .	35
2.8	Les couples de paires significatifs. a. les couples sont semblables, un des deux couples est pris en compte dans les paramètres numériques. b. Les deux couples sont différents. Les deux couples sont pris en compte dans les paramètres numériques. . . . .	36
2.9	Exemple d'une estimation de $\Delta G_{37}$ pour une séquence CGTTGA [SantaLucia 98].	36
2.10	Evolution de $\Delta G$ pour une séquence de 120 bases en fonction de $[Na^+]$ pour plusieurs valeurs fixes de $T$ . . . . .	38
2.11	Evolution de $\Delta G$ pour une séquence de 120 bases en fonction de $T$ pour plusieurs valeurs fixes de $[Na^+]$ . . . . .	38
2.12	Energie d'interaction $\Delta G$ en fonction de la longueur de séquence pour les séquences minimale (droite $D_{AT}$ ) et maximale (droite $D_{GC}$ ). . . . .	40
3.1	Principe général de l'AFM. . . . .	44
3.2	Utilisation de l'AFM pour la caractérisation des surfaces. A gauche, une courbe A/R pour un substrat en silicium dans un milieu liquide. La force d'interaction est représentée en fonction du déplacement de la poutre. La phase approche est représentée en bleu et la phase retrait en rouge. A droite, la topographie d'une surface en silicium. Ces images (courbe et topographie) ont été obtenues en collaboration avec le laboratoire AMIR de l'université de Oldenburg. . . . .	44
3.3	Principe d'une mesure des forces d'interaction biomoléculaires par l'AFM. a. Formation du complexe ligand-récepteur Calmoduline-skMLCK (CaM-skMLCK). Cette protéine est composée de deux domaines DomN (rouge) et DomC (bleu), l'activation par des ions $Ca^{2+}$ (en gris) permet la formation du complexe entre skMLCK (vert) et CaM. b. La poutre est fonctionnalisée avec CaM et le substrat avec le ligand. En gris, des molécules intermédiaires de fonctionnalisation (domaines de filamine) [Junker 09]. . . . .	46

3.4	Courbe A/R avec plusieurs événements de rupture obtenues en collaboration avec le NPL pour un substrat en silicium fonctionnalisé avec la séquence S1. Les sauts 1 et 2 sont des événements de rupture. Dans l'encadré rouge, La courbe publiée par [Strunz 99]. . . . .	47
3.5	Protocole expérimental de Sattin et al. [Sattin 04] pour la mesure des forces d'interaction entre différents brins d'ADN dans l'axe de la double hélice. 4 séquences sont utilisées. Les deux séquences TIP et MATCH sont complémentaires. La séquence MIS1 est identique à MATCH avec la modification du dixième nucléotide. La séquence MIS2 est identique à MIS1 avec la modification du onzième nucléotide. L'introduction de ces modifications conduit les auteurs à un premier type de mesures spécifiques, entre les séquences [TIP et MATCH], et à deux types de mesures non spécifiques, entre les séquences [TIP et MIS1] et [TIP et MIS2]. . . . .	48
3.6	Synthèse des mesures AFM des interactions entre des brins d'ADN dans l'axe de la double hélice : conditions expérimentales. Liner : type de molécule utilisé pour la fonctionnalisation. . . . .	49
3.7	Synthèse des mesures AFM des interactions entre des brins d'ADN dans l'axe de la double hélice : résultats. . . . .	50
3.8	Scénario d'un processus d'A/R lors d'une mesure AFM de l'interaction entre les brins d'ADN complémentaires fixés aux surfaces de la pointe et du substrat. . . . .	52
3.9	Taux de chargement pour chaque expérience de mesure d'interaction inter brins d'ADN sur la séquence S1. . . . .	52
3.10	Courbe approche/retrait avec deux événements de rupture correspondant à la première itération de l'expérience Exp1. . . . .	53
3.11	Courbes de variances des données entières. $V_\lambda$ en fonction de $\lambda$ . . . . .	56
3.12	Comparaison des courbes de variance des données entières et des données coupées pour les expériences exploitables. . . . .	58
3.13	Comparaison des courbes de variance des données entières et des données coupées pour les expériences non exploitables. . . . .	59
3.14	Les Forces de rupture extraites pour les expériences exploitables (Exp1, Exp2 et Exp6) et pour les données exploitables (Exp3, Exp4 et Exp5). L'écart entre la moyenne (en rouge) et la médiane (en vert) donne une indication sur la représentativité de la valeur moyenne pour chaque expérience. . . . .	61
3.15	Histogrammes des forces de rupture. . . . .	62
3.16	Différents types de configuration pour une même séquence. a. Configuration "Perfect Match". b. Combinaisons de deux configurations : boîte 1. Configuration <i>hétéro-complémentaire</i> , boîte 2 : Configuration en <i>épingle à cheveux</i> . . . . .	64
3.17	Les configurations MM de type <i>épingle à cheveux</i> . A gauche, 3 <i>épingles à cheveux</i> sont formées. A droite, carte des configurations MM de type <i>épingle à cheveux</i> d'après mFold. Les deux axes représentent les indices des bases de la séquence étudiée. 14 configurations sont stables (Analyse Hairpin sur OligoAnalyzer 3.1). . . . .	65
3.18	Extraction de la force de rupture $Fr$ et de la variation de déplacement $d$ . . . . .	66
3.19	Diagramme des énergies libre de Gibbs calculée d'après le modèle du plus proche voisin en fonction du nombre de la longueur des segments hybridés. Les énergies des configurations PM et <i>hétéro-complémentaires</i> de la séquence $S_1$ , l'une des séquences utilisées dans le projet Golem (Annexe A) sont représentées sur ce diagramme. Ces énergies sont comprises dans l'espace entre les deux droites $D_{AT}$ et $D_{GC}$ . . . . .	68

3.20	Energies $\Delta G_r(Kcal/mol)$ des événements locaux de rupture des 198 premières courbes A/R pour les expériences Exp1 ( $v = 1\mu m/s$ ), Exp2 ( $v = 0.1\mu m/s$ ) et Exp6 ( $v = 1\mu m/s$ , scan). . . . .	69
3.21	Analyse énergétique des événements locaux de rupture des 198 premières courbes A/R pour les expériences Exp1 ( $1\mu m/s$ ), Exp2 ( $0.1\mu m/s$ ) et Exp6 ( $1\mu m/s$ , scan). La droite $D_{AT}$ de la séquence minimale en bleu et la droite $D_{GC}$ de la séquence minimale en rouge. . . . .	69
3.22	Répartition des événements locaux de rupture sur les catégories 1, 2 et 3 des 198 premières courbes A/R pour les expériences Exp1 ( $1\mu m/s$ ), Exp2 ( $0.1\mu m/s$ ) et Exp6 ( $1\mu m/s$ , scan). . . . .	70
3.23	Histogrammes des énergies de rupture $\Delta G_r(Kcal/mol)$ des 198 premières courbes A/R pour les expériences Exp1 ( $1\mu m/s$ ), Exp2 ( $0.1\mu m/s$ ) et Exp6 ( $1\mu m/s$ , scan). . . . .	70
4.1	Scénarii des configurations possibles entre deux brins d'ADN complémentaires. . . . .	75
4.2	Schéma représentatif de l'algorithme de génération de séquence. . . . .	78
4.3	Comparaison des configurations en <i>épingle à cheveux</i> des séquences $S_1$ et $S_3$ . . . . .	79
4.4	Comparaison des énergies des configurations <i>hétéro-complémentaires</i> et " <i>Perfect match</i> " stables des séquences $S_1$ et $S_3$ . . . . .	80
4.5	Répartition des événements de rupture locaux sur les trois catégories de la méthode d'analyse énergétique. . . . .	83
4.6	Courbes A/R . . . . .	84
4.7	Pourcentage des courbes en fonction du nombre de paliers . . . . .	85
4.8	Etude de la spécificité. Comparaison de la distribution des forces de rupture mesurées pour les expériences Exp10 et Exp14. . . . .	86
4.9	Etude de la spécificité. Comparaison de la distribution des distances de rupture mesurées pour les expériences Exp10 et Exp14. . . . .	87
4.10	Comparaison de la distribution des forces de rupture mesurées pour les expériences Exp10 et Exp16. . . . .	88
4.11	Comparaison de la distribution des distances de rupture mesurées pour les expériences Exp10 et Exp16. . . . .	89
4.12	Etude du temps d'attente pour $v = 0.1\mu m/s$ . Comparaison des distributions des distances de rupture mesurées pour les expériences Exp10( $T_a = 0s$ ) et Exp13( $T_a = 60s$ ). . . . .	90
4.13	Etude du temps d'attente pour $v = 0.05\mu m/s$ . Comparaison des distributions des distances de rupture mesurées pour les expériences Exp16( $T_a = 0s$ ) et Exp17( $T_a = 60s$ ). . . . .	90
4.14	Etude du temps d'attente pour $v = 0.1\mu m/s$ . Comparaison des distributions des forces de rupture mesurées pour les expériences Exp10( $T_a = 0s$ ) et Exp13( $T_a = 60s$ ). . . . .	91
4.15	Etude du temps d'attente pour $v = 0.1\mu m/s$ . Comparaison des forces de rupture moyennes par groupe pour les expériences Exp10( $T_a = 0s$ ) et Exp13( $T_a = 60s$ ). . . . .	91
4.16	Etude du temps d'attente pour $v = 0.05\mu m/s$ . Comparaison des distributions des forces de rupture mesurées pour les expériences Exp16( $T_a = 0s$ ) et Exp17( $T_a = 60s$ ). . . . .	92
4.17	Etude du temps d'attente pour $v = 0.05\mu m/s$ . Comparaison des forces de rupture moyennes par groupe pour les expériences Exp16( $T_a = 0s$ ) et Exp17( $T_a = 60s$ ). . . . .	93

---

# INTRODUCTION GÉNÉRALE

---

L'assemblage d'objets microscopiques est actuellement un verrou majeur au cœur de domaines aussi importants et variés que les nanosciences, les MEMs ou la médecine. Le projet européen GOLEM s'inscrit dans une démarche novatrice pour résoudre cette problématique et propose d'utiliser des processus biologiques pour assembler des micro-objets de façon parallèle. Dans le cadre de ce projet, cette thèse aborde l'étude et l'analyse des méthodes d'assemblage bioinspirées pour assembler de façon réversible et contrôlée ces objets. L'approche proposée est multidisciplinaire, puisqu'il s'agit d'exploiter des connaissances issues de la biologie dans un contexte microrobotique. Le rapprochement entre ces deux domaines est une étape importante pour les développements futurs en micro et nanotechnologies et ouvre des perspectives intéressantes.

Le choix de processus biologiques pour assurer l'assemblage de façon sélective et contrôlable entre les objets est crucial dans ce travail, et dépendent de deux propriétés fondamentales identifiées : la stabilité et la spécificité de l'interaction biologique. La stabilité concerne la capacité des composants à garder un état assemblé, sur une période de temps donnée et dans un environnement déstructuré. La spécificité concerne la reconnaissance mutuelle entre les composants à assembler. L'intégration de ces propriétés dans un processus d'assemblage rend envisageable la conception d'une méthode d'auto-assemblage, massive et parallèle, où les différents composants, formant un système complexe, se positionnent et s'assemblent dans un processus stochastique sans intervention individuelle. L'intégration de ces propriétés biologiques sur la matière inorganique est abordée en plusieurs étapes :

- le choix d'un processus biologique, dont la stabilité et la spécificité sont intrinsèques et contrôlables,
- l'étude théorique et expérimentale de la stabilité et la spécificité de cette adhésion biologique sur des composants fonctionnalisés,
- et la détermination des paramètres de contrôle de l'assemblage et leur optimisation.

Ainsi, cette thèse cherche dans un premier temps à analyser les différents travaux dans le domaine et principalement l'auto-assemblage aux échelles micro et nanoscopiques. Cet état de l'art multidisciplinaire ouvre sur le choix de l'ADN avec ses propriétés particulières, capables

de répondre aux besoins de l'auto-assemblage. Les propriétés de stabilité et de spécificité du processus d'hybridation d'ADN sont ensuite étudiées.

La stabilité est abordée selon deux angles, l'échelle individuelle et l'échelle populationnelle. L'échelle individuelle s'intéresse à l'interaction entre deux brins d'ADN simple avec une approche empruntée aux travaux en modélisation moléculaire. L'échelle d'étude est ensuite élargie à deux populations de brins complémentaires, fixées sur des surfaces à assembler. Cette configuration géométriquement contrainte est étudiée avec une approche thermodynamique au regard des différents paramètres extrinsèques comme la température et la concentration saline du milieu.

Cette thèse aborde aussi la validation expérimentale des modèles développés. Dans un premier temps, un modèle est proposé pour coupler les résultats exprimés en termes d'énergie et les mesures expérimentales en termes de force. La première campagne expérimentale effectuée à Londres au NPL cherche à déterminer la stabilité de l'hybridation. Des outils méthodologiques d'analyse statique et énergétique sont mis en place pour l'appréhender et la quantifier. La méthode énergétique est originale et couple des outils de simulation, de prédiction et d'analyse. Elle montre ainsi que la stabilité et la spécificité sont intrinsèquement liées par la composition de la séquence en bases azotées des brins d'ADN. Il apparaît donc que celle-ci peut être optimisée pour produire des configurations plus favorables statistiquement.

Un algorithme original de génération de séquences, basé sur le caractère programmable de l'ADN, est alors proposé pour disposer de couples de brin d'ADN dont les propriétés de stabilités et de spécificités sont optimisées. Une validation expérimentale est ensuite entreprise en collaboration avec le laboratoire AMIR à Oldenburg en Allemagne. Ces expériences valident l'approche de conception de séquences et met en évidence l'influence de paramètres clefs comme la vitesse d'approche et le temps d'attente sur l'assemblage.

Le premier chapitre est ainsi dédié à un état de l'art en auto-assemblage et couvre plusieurs domaines. Il vise à proposer une approche synthétique des différents thèmes de recherche abordés dans ce travail. Le deuxième chapitre aborde l'étude des propriétés de stabilité et de spécificité à travers une modélisation multi-échelle et les paramètres de contrôle et environnementaux. Le troisième chapitre s'intéresse à la validation expérimentale et propose des méthodes d'analyse statistique et énergétique originales pour comprendre et classifier les interactions en termes de stabilité et de spécificité. Le dernier chapitre s'attache à définir l'outil logiciel développé pour concevoir des séquences spécifiques optimisées du point de vue de l'auto-assemblage. Une seconde validation expérimentale est entreprise et montre l'intérêt de cette approche, l'optimisation des propriétés fondamentales de stabilité et de spécificité pour résoudre le problème de l'auto-assemblage d'objets microscopiques à partir de l'ADN.

# PROBLÉMATIQUE DE L'AUTO-ASSEMBLAGE MICROSCOPIQUE PAR PROCESSUS BIOLOGIQUES

---

Cette thèse s'inscrit dans la dynamique scientifique du projet européen Golem (FP6) [Golem 06]. Son objectif principal est d'étudier la faisabilité de l'auto-assemblage par processus biologiques pour des objets de dimensions microscopiques. Golem réunit ainsi onze équipes de recherche à travers toute l'Europe autour de ce nouveau concept d'assemblage.

Ce travail de thèse s'inscrit dans le cadre de ce projet, et plus particulièrement dans la définition et l'analyse des paramètres des processus biologiques pour l'auto-assemblage à l'échelle microscopique. Ainsi, la problématique couvre le choix de l'entité biologique pour le collage des micro-objets, l'étude des paramètres intrinsèques, l'analyse des paramètres pour contrôler le processus et enfin la caractérisation expérimentale.

Ce chapitre s'attache à présenter la problématique de l'auto-assemblage microscopique par processus biologiques. Le contexte et les difficultés du micro-assemblage en général et l'utilisation de l'auto-assemblage sont présentés dans la première partie de ce chapitre. Avant d'exposer un panorama des travaux majeurs en auto-assemblage général et plus spécifiquement dans l'auto-assemblage par processus biologiques, quelques définitions et outils d'analyse sont proposés dans la deuxième partie.

La contribution de mon travail de thèse repose sur l'hypothèse centrale du choix de l'ADN pour la réalisation du collage des micro-objets. Son utilisation dans ce contexte précis est justifiée en troisième partie de ce chapitre. Une étude des propriétés de l'ADN au regard de la problématique de l'auto-assemblage microscopique est aussi introduite.

Dans la quatrième partie, il s'agit de mettre en avant la nécessité de comprendre théoriquement et expérimentalement les interactions ADN-ADN. Le choix de la microscopie à force



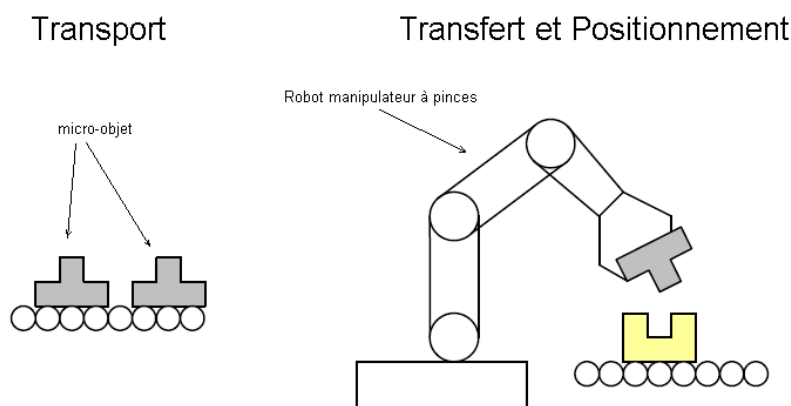
atomique est proposé pour ces expériences. Enfin, le plan de travail de cette thèse est exposé.

## 1 Le micro-assemblage dans le projet Golem

L'assemblage de systèmes intégrés miniaturisés est une procédure complexe à cause des difficultés rencontrées lors de la manipulation de composants de taille inférieure au millimètre. De nombreux domaines comme les MEMS, la chirurgie, les télécommunications ou l'industrie automobile utilisent des systèmes intégrant des microcomposants issus des techniques d'assemblage conventionnelles. Par exemple, dans l'industrie électronique, la manipulation de composants microscopiques est une étape complexe et difficile [Golem 06].

Malgré les réelles avancées dans le domaine du micro-assemblage, l'assembler d'un *grand nombre de micro-composants en parallèle avec une hétérogénéité de matière, de géométrie et de fonction* n'est pas possible à partir des connaissances et les technologies actuelles [Morris 05]. A ces échelles et particulièrement en dessous de  $100\mu m$ , les techniques conventionnelles ne peuvent plus être utilisées [Régnier 08]. Deux techniques d'assemblage fondamentales et complémentaires existent.

La première technique est l'assemblage "sériel". Ce terme définit la manipulation d'objets en série par une succession d'étapes de saisie, de transfert et de positionnement [Gauthier 07] illustrées à la figure 1.1 [Golem 06]. Les systèmes formés par cette technique trouvent leur limite inférieure au voisinage de  $10^{-4}m$ . La seconde technique est la synthèse chimique de macromolécules à partir d'atomes. A ces échelles, les systèmes formés trouvent leurs limite supérieure au voisinage de  $10^{-8}m$  [Golem 06].



**Figure 1.1** - *Micro-assemblage sériel. Transport, transfert et positionnement robotisés de micro-objets [Golem 06].*

Ces deux techniques présentent de sérieuses lacunes et l'assemblage microscopique entre  $10^{-8}m$  et  $10^{-4}m$  reste un véritable champ de recherche [Lambert 06]. La problématique du micro-assemblage est très complexe du fait d'un grand nombre de difficultés apportées par le changement d'échelle. Cette dernière se trouve à la frontière de deux espaces classiques : l'espace

macroscopique régi par la physique newtonienne et l'espace atomique régi par la physique quantique [Gauthier 07].

La différence majeure avec l'échelle macroscopique est due à la prépondérance des forces surfaciques sur les forces volumiques. En effet, pour une longueur  $L$  dans l'intervalle  $[0 ; 1]$ , la valeur de  $L^3$  étant plus petite que celle de  $L^2$ , les forces volumiques deviennent négligeables devant les forces surfaciques. Celles-ci modifient considérablement les interactions physiques à l'échelle microscopique.

Trois grandes classes de forces d'adhésion paramétrées par la distance d'interaction apparaissent et doivent être prises en considération [Isr 91] :

- Les forces de van der Waals. Elles s'expriment par exemple pour une interaction entre une sphère de rayon  $R$  et un plan par :

$$F_{vdw}(D) = \frac{AR}{6D^2}$$

avec  $A$  la constante de Hamaker et  $D$  la distance d'interaction.

- Les forces électrostatiques, forces de Coulomb classiques qui dépendent des charges acquises par les surfaces :

$$F_{elec}(D) = \frac{q^2}{4\pi\epsilon_0 D^2}$$

avec  $q$  la charge électrique et  $\epsilon_0$  la permittivité de l'air.

- Les forces capillaires, dont l'existence est déterminée par les conditions d'humidité de l'environnement :

$$F_{elec}(D) = \frac{4\pi R\gamma_l \cos(\theta)}{1 + \frac{D}{d}}$$

avec  $\gamma_l$  la tension de surface du liquide,  $d$  la hauteur du ménisque et  $\theta$  l'angle de contact.

Ces forces sont non linéaires et imprédictibles car certains paramètres sont mal connus comme la constante de Hamaker, voire inconnus comme les charges électriques présentes sur les matériaux. L'ensemble de ces forces rendent l'assemblage microscopique complexe tant sur le plan technique que sur le plan de la compréhension et de la modélisation [Alvo 10].

Pour répondre à l'ensemble de ces limitations propres à l'échelle microscopique, il existe aujourd'hui une alternative, l'auto-assemblage parallèle. Ce dernier met en jeu des lois d'interaction spécifiques pour l'auto-positionnement des micro-composants, basées sur des caractéristiques particulières, telles que la géométrie ou les matériaux de fabrication. Ainsi, le processus d'auto-assemblage doit être réalisé dans un environnement particulier pour alimenter ces interactions spécifiques. Plusieurs exemples d'auto-assemblage existent avec différents types d'interactions testés expérimentalement : forces capillaires [Stauth 06] [Zheng 04], interactions chimiques [Xiong 03], champs de force électrique [O'Riordan 04a] et champs magnétique [Huang 06].

Malgré l'intérêt grandissant pour l'auto-assemblage, cette technique présente encore des inconvénients majeurs notamment pour le développement industriel. Les inconvénients les plus

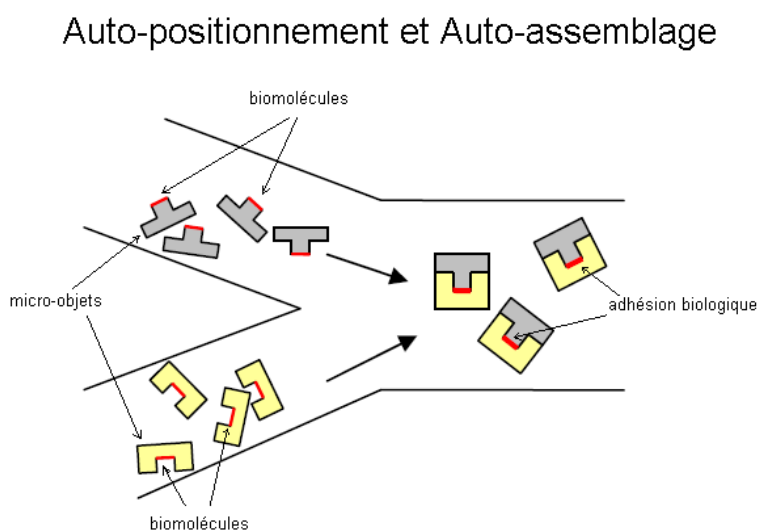
importants sont sans doute les nombreuses contraintes concernant la géométrie ou l'incompatibilité des matériaux utilisés [Morris 05].

Dans le domaine du vivant, l'auto-assemblage est un processus trivial et incontournable. Utiliser des processus existants dans la biologie est une solution originale pour minimiser les coûts énergétiques, et exploiter de nouvelles stratégies. Celles-ci peuvent faire intervenir des géométries complexes jusqu'ici incompatibles.

Dans le cadre du projet européen Golem, une solution est proposée pour répondre à l'ensemble de ces limitations rencontrées à l'échelle microscopique. Il s'agit d'utiliser des processus biologiques pour la réalisation d'un assemblage sélectif des composants. Ainsi, l'objectif général du projet est d'utiliser des biomolécules fixées à la surface des composants pour réaliser l'adhésion stable et spécifique dans un processus d'auto-assemblage microscopique en parallèle. Un tel processus est illustré à la figure 1.2 [Golem 06] : deux composants différents sont couverts sur leurs surfaces par des molécules biologiques complémentaires. La reconnaissance entre ces molécules assure l'assemblage des composants au cours d'un processus stochastique.

Dans cet objectif, plusieurs facteurs sont pris en compte :

- La fixation des biomolécules à la surface des micro-objets. Cette procédure est appelée *fonctionnalisation des surfaces*.
- Le choix des molécules biologiques adéquates.
- La caractérisation des interactions biomoléculaires mises en jeu.
- L'optimisation de la force de guidage pour augmenter la probabilité de rencontre entre les micro-objets. Ce facteur fait l'objet d'un travail important de Golem notamment dans les travaux de Vela et al. [Vela Saavedra 09].



**Figure 1.2** - *Auto-assemblage microscopique par processus biologiques. Exemple d'une manipulation parallèle sans contact de micro-objets en milieu aqueux [Golem 06].*

Ce travail de thèse s'intéresse particulièrement aux questions du choix des molécules biologiques adéquates, de l'estimation et de l'analyse des forces d'interaction mises en jeu. Cette étude est au cœur du développement du projet Golem car elle s'inscrit dans :

- la simulation du processus d'auto-assemblage,
- la conception du prototype et l'étude des paramètres physiques mis en jeu,
- le contrôle du processus d'auto-assemblage.

La partie suivante propose une définition de l'auto-assemblage et de ses propriétés fondamentales. Un état de l'art des travaux existants à l'échelle microscopique en auto-assemblage inorganique et en auto-assemblage par processus biologiques est ensuite détaillé.

## 2 L'auto-assemblage microscopique

### 2.1 Propriétés de l'auto-assemblage

En général, l'auto-assemblage, défini comme "la réalisation de l'ordre spontané dans un système" par [Boncheva 05], est un processus complexe du point de vue de la modélisation [Rothmund 00]. Dans le but de comprendre, de caractériser et de quantifier ce processus, il est nécessaire d'identifier ses propriétés fondamentales : la *force de guidage*, la *stabilité* et la *spécificité*.

- **Force de guidage :**

C'est une force externe appliquée au système pour augmenter les probabilités de rencontre entre les objets. A l'échelle moléculaire, la force de guidage est généralement due à l'agitation thermique et au mouvement brownien. A l'échelle microscopique, plusieurs approches sont envisageables comme le contrôle du flux fluidique, la diélectrophorèse...

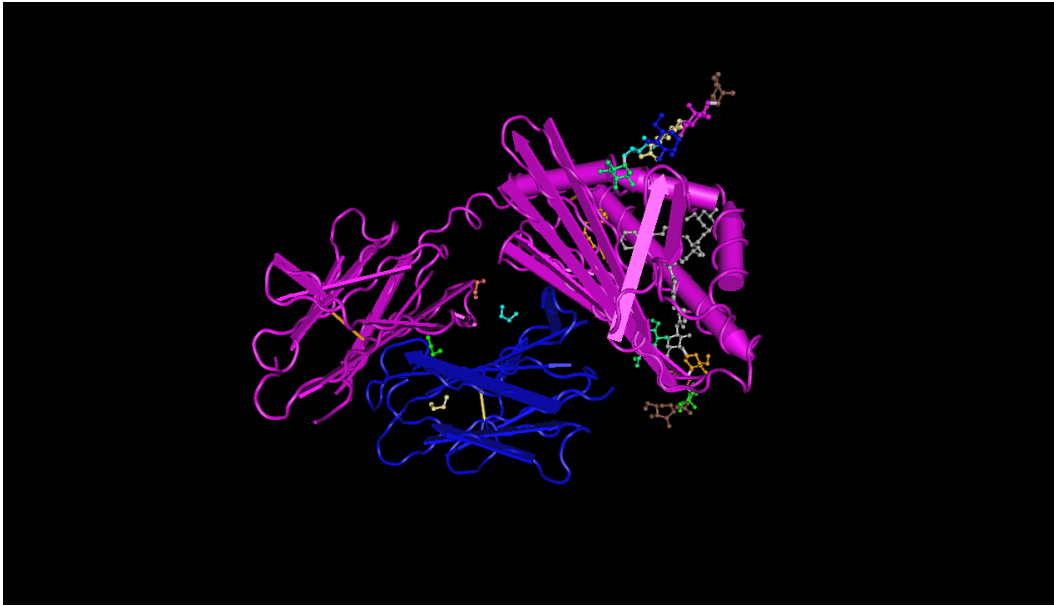
- **Stabilité :**

Il s'agit de la capacité d'un système d'objets à rester dans un état d'équilibre, en présence de forces extérieures. Cet état correspond naturellement à un minimum d'énergie potentielle. Dans le cas d'un auto-assemblage microscopique, il s'agit du potentiel des forces magnétiques, capillaires ou biologiques, en fonction de la technique utilisée.

- **Spécificité :**

La spécificité désigne le caractère *sélectif* du processus d'auto-assemblage. Il définit la reconnaissance mutuelle des objets avec un caractère particulier. Dans le cas général, et à l'échelle microscopique, la spécificité est souvent réalisée à partir de la complémentarité géométrique [Zheng 04].

Afin d'illustrer un processus d'auto-assemblage en particulier, la description d'un exemple biomoléculaire est détaillé ici. Ce type d'auto-assemblage met en jeu des forces avec des valeurs de l'ordre de l'énergie thermique entre un grand nombre de molécules. Ces forces sont principalement dues à des interactions ou des liaisons réversibles non-covalentes. L'assemblage résulte en la formation de *complexes*, à partir de molécules dont les conformations sont complémentaires. La figure 1.3 illustre un exemple de structure tridimensionnelle, obtenue par cristallographie, d'un complexe biomoléculaire formé de deux molécules : Cd1d et C6ph de souris. La formation d'un tel complexe nécessite un très grand nombre de molécules afin d'augmenter la probabilité de rencontre à partir de l'agitation thermique.



**Figure 1.3** - Structure 3D du complexe Cd1d-C6ph chez la souris [Schiefner 09]. En violet : Cd1d et en bleu : C6ph.

La réversibilité des liaisons chimiques formées dans les complexes implique l'ajustement en cas de non complémentarité parfaite entre les molécules. Les interactions mises en jeu conduisent le système vers un état plus stable (un état d'énergie plus bas) et se produisent avec un taux d'erreurs fini en raison du grand nombre de molécules en interaction. Pour conduire le système d'un minimum d'énergie local à un minimum global, ou l'auto-assemblage parfait, un apport d'énergie externe sous la forme de chaleur est nécessaire. Dans cet exemple, *la force de guidage* est principalement due au mouvement brownien; *la stabilité* est réalisée avec des ponts hydrogène ou disulfure et *la spécificité* est due aux complémentarités des conformations (pour le docking moléculaire) ou au codage génétique dans le cas de l'ADN. Ce processus d'auto-assemblage biomoléculaire réalise des complexes à partir de molécules dont la taille est mille fois plus petite que la taille des composants assemblés dans un auto-assemblage microscopique.

Les techniques adaptées à l'échelle microscopique peuvent être classées en deux types : l'auto-assemblage microscopique inorganique, ne faisant intervenir aucun processus biologique par opposition à l'auto-assemblage microscopique avec processus biologique. Ces deux classes d'auto-assemblage sont successivement détaillées dans la suite.

## 2.2 Etat de l'art

### 2.2.1 L'auto-assemblage microscopique *inorganique*

L'auto-assemblage inorganique désigne toute procédure d'auto-assemblage ne faisant intervenir aucun matériel biologique. Dans le tableau 1.1, les méthodes classiques sont classées selon les 3 éléments définis dans l'analyse : *la force de guidage*, *la stabilité* et *la spécificité*. Cette classification est basée sur l'état de l'art publié par Morris et al. en 2005 [Morris 05]. Ainsi, plusieurs principes peuvent être utilisés pour réaliser chacun des trois éléments définissant un processus d'auto-assemblage.

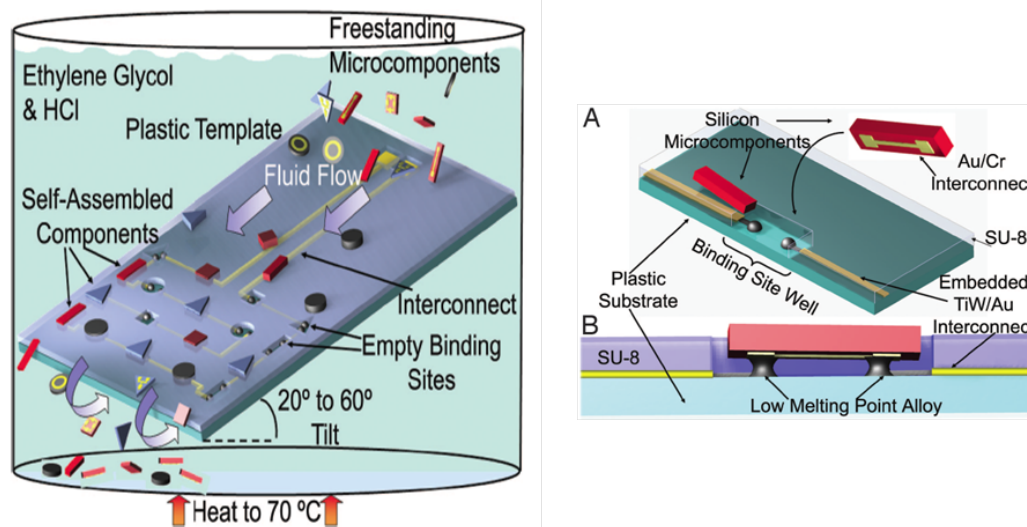
Principe	Forces de guidage	Spécificité	Stabilité	Difficultés
2D fluïdique, gravitation	Gravitation	Formes géométriques	Van der Waals + gravité	
2D fluïdique, capillaire	Capillaire	Formes géométriques	Forces capillaire (minimisation de l'énergie de surface interfaciale)	Précision d'alignement
2D Electrostatique	Electrostatique	Electrostatique	Electrostatique	Contrôle de charge sur les composants
3D	Hydrophobique/hydrophilique Capillaire	Formes géométriques	Forces capillaires	Très faible taux d'assemblage

**Tableau 1.1** - *Tableau récapitulatif des méthodes existantes d'auto-assemblage.*

Pour la force de guidage, les principes les plus réponsus sont les forces gravitationnelles, capillaires, électrostatiques, ou les propriétés hydrophobe/hydrophile des matériaux. D'autres procédés peuvent aussi être utilisés comme les champs de forces électriques ou magnétiques. L'assemblage *parallèle* n'est pas réalisé dans ces différentes techniques puisque les composants sont introduits dans le milieu séquentiellement. Les travaux de O'Riordan et al. [O'Riordan 04b] illustrent ce problème. Il s'agit de l'assemblage de composants de l'ordre de  $50\mu m$  de diamètre spécifiquement localisés, transportés, et déposés sur un substrat en silicium à partir d'un champ électrique. Les composants sont introduits séquentiellement dans le milieu par type de forme géométrique afin de limiter le nombre d'erreurs. Le caractère parallèle de l'auto-assemblage n'est ainsi pas réalisé.

Un second problème concerne la spécificité. En effet, dans les processus d'auto-assemblage 2D les plus aboutis, avec la gravitation ou les forces capillaires comme force de guidage, le principe le plus utilisé est la reconnaissance de formes géométriques [Zheng 04]. Lorsque ce principe est appliqué à un processus 3D, le taux d'assemblage devient beaucoup plus faible et une perte de rentabilité apparaît. L'évolution du 2D au 3D multiplie la complexité du processus de reconnaissance entre deux composants par le nombre de leurs faces. La probabilité de rencontre est donc amoindrie. En 2004, Zheng et al. [Zheng 04] réalisent l'auto-assemblage de microcomposants semi-conducteurs de l'ordre de  $200\mu m$ , où les composants sont introduits séquentiellement dans l'environnement liquide d'assemblage. Ils constatent que le taux d'assemblage reste assez faible.

Dans un processus 2D, un taux d'assemblage nettement plus élevé est publié par Stauth et al. en 2006 [Stauth 06] pour des composants du même ordre de dimension. Dans ce processus un assemblage à 97% de FETs (silicon field-effect transistors) est obtenu sur un substrat en plastique avec un taux de 24 000 composants, de taille caractéristique  $100\mu m$ , par heure. Ces travaux sont illustrés sur la figure 1.4. La force de guidage est réalisée par des forces gravitationnelles, et la stabilité par des forces capillaires. L'apparition de ces forces est réalisée à partir d'un alliage à bas point de fusion situé sur les sites de liaison du substrat et d'un bloc métallique positionné sur les sites de liaison des composants. Quand la forme d'un microcomposant correspond, avec un contrôle de la température de fusion de l'alliage des sites de liaison, le contact entre ces derniers et le bloc métallique positionné sur la surface du microcomposant est établi. Après l'achèvement du processus d'A.A, la température est abaissée pour solidifier l'alliage et établir les connexions mécaniques et électriques entre le micro-composant et le substrat.



**Figure 1.4** - *Auto-assemblage hétérogène de Stauth et al. [Stauth 06]. Les microcomposants sont introduits dans un environnement liquide et transportés par mouvement fluidique. L'auto-assemblage a lieu dès que les microcomposants rencontrent leurs formes complémentaires. Ils s'attachent ensuite par forces capillaires résultantes de l'alliage métallique déposé sur le substrat.*

Les travaux remarquables de Stauth et al. présentent néanmoins un inconvénient majeur. En effet, ce processus 2D est difficilement réalisable en trois dimensions. Puisque l'adhésion est homogène à tous les micro-composants, pour réaliser un assemblage spécifique, une complexité géométrique très élevée serait nécessaire.

Un autre inconvénient s'ajoute au problème de spécificité géométrique : le choix limité du matériau utilisé. En effet, l'or ou le silicium restent les matières les plus répandues en raison de leurs propriétés physicochimiques.

L'industrialisation des techniques d'auto-assemblage existantes nécessite l'optimisation du taux d'assemblage. Ce taux dépend des propriétés matérielles des micro-composants. Cette dépendance est un réel handicap pour l'optimisation des taux d'assemblage. En effet, comme la spécificité et la stabilité sont des paramètres interdépendants, lorsque ces deux propriétés dépendent des caractéristiques intrinsèques aux composants, les systèmes formés sont très limités en terme de *géométrie*, de *matière* et de *fonction*, et les taux d'assemblage restent très faibles. Une solution serait d'optimiser la spécificité et la stabilité du processus à travers les propriétés intrinsèques au processus local d'adhésion.

La biologie semble proposer des solutions potentielles concernant la stabilité et la spécificité. En effet, dans le domaine du vivant, l'auto-assemblage est un processus trivial et récurrent [Glade 04]. Plusieurs processus de ce type ont été étudiés en pharmacologie et en biochimie et sont aujourd'hui clairement identifiés, tels que l'adhésion cellulaire et les processus de signalisation cellulaire. De plus, les processus d'auto-assemblage biomoléculaires sont connus pour être optimisés à un niveau d'énergie minimum. Par conséquent, utiliser des entités biologiques pour remplacer la manipulation semble être un bon compromis. L'exploitation directe des processus d'auto-assemblage biologiques peut répondre aux défauts engendrés par l'auto-assemblage microscopique actuel. Il s'agit en effet de phénomènes intrinsèquement spécifiques, stables et parallèles. Finalement les inconvénients connus dans les processus actuels peuvent être résolus, notamment pour l'hétérogénéité de matière et de formes géométriques.

### 2.2.2 L'auto-assemblage microscopique par processus biologiques

Cet auto-assemblage consiste à utiliser des molécules biologiques capables de former des complexes à partir de deux types de biomolécules (cf. section 2.1 page 8). Elles sont fixées à la surface des microcomposants destinés à l'assemblage. Les principes de stabilité et de spécificité biomoléculaires propres à la nature des molécules utilisées sont alors mis en jeu dans un contexte populationnel avec plusieurs milliards de molécules fixées en interaction.

Ces processus d'auto-assemblage biomoléculaires sont largement utilisés dans l'auto-assemblage nanoscopique. Par la transformation orientée des complexes biomoléculaires, ces processus engendrent des réalisations comme les moteurs moléculaires et d'autres structures nanométriques [Mann 09]. Pour parvenir à ces nouvelles structures, les fonctions premières sont détournées de leur fonctionnement usuel et les conformations spatiales des molécules originelles sont exploitées.

La fabrication, les propriétés et les applications de nanostructures à base d'ADN sont décrites dans un état de l'art par R. Bashir en 2001 [Bashir 01] puis par Young Sun et Ching-Hwa Kiang en 2005 [You 05]. Le principe est de contrôler la configuration 3D de la molécule en jouant sur ses propriétés thermodynamiques et géométriques. Une liste non-exhaustive des applications possibles de l'ADN en nanotechnologies est proposée :

- Nano-structure d'ADN pure [Seeman 83, Shih 04, Rothermund 06](figure 1.5), et les travaux les plus récents dans ce domaine sont des structures polyhedrale [Yu He 08].
- Assemblage de nano-particules de métal à base d'ADN de Mirkin et al. [Mirkin 96].
- Synthèse de semi-conducteurs en utilisant l'ADN de Coffey et al. [Coffey 92].

Les nanostructures à base de protéines sont naturellement présentes dans la nature. Les premiers prototypes [Hess 04] font l'objet de nombreuses études. Par exemple le moteur protéique 1.7 composé de myosine et de kinésine est un actionneur et transporteur in-vivo. Plus généralement, sur les nano-structures hybrides [Mann 09, Hamdi 09], il apparaît que la *coordination*, l'*organisation* et la *transformation* des nano-objets inorganiques sont des tâches réalisées à partir de biomolécules. Au sein de ces nano-objets hybrides (organiques-inorganiques), des niveaux d'organisation élevés peuvent être atteints. D'autres exemples d'auto-assemblage nanoscopique sont proposés sur les figures 1.5 et 1.6. L'ADN apparaît comme la molécule la plus exploitée dans ce domaine. En effet, les nanostructures artificielles à base d'ADN sont plus évoluées que celles basées sur d'autres complexes biomoléculaires [Hess 04]. Ces propriétés potentielles des biomolécules sont ainsi une solution potentielle pertinente pour l'auto-assemblage microscopiques.

L'ensemble de ces processus biomoléculaires exploités dans l'auto-assemblage nanoscopique peuvent être utilisés dans l'auto-assemblage microscopique. En effet, collées à la surface des micro-composants, ces biomolécules pourraient réaliser la spécificité et la stabilité nécessaire à l'auto-assemblage microscopique. Dans la suite de ce document, ces biomolécules sont qualifiées de "*velcro biologique*" puisque de façon similaire à une adhésion velcro, il s'agit d'une population de molécules fixées aux surfaces des composants qui adhèrent les unes aux autres de façon spécifique. Il existe dans la Nature plusieurs processus pouvant servir de colle biologique pour l'auto-assemblage micro et nanoscopique.

A l'échelle *microscopique* le choix du velcro adéquat doit prendre en compte l'ordre de grandeur des forces d'interaction et de la spécificité propres. Ces processus peuvent être classés



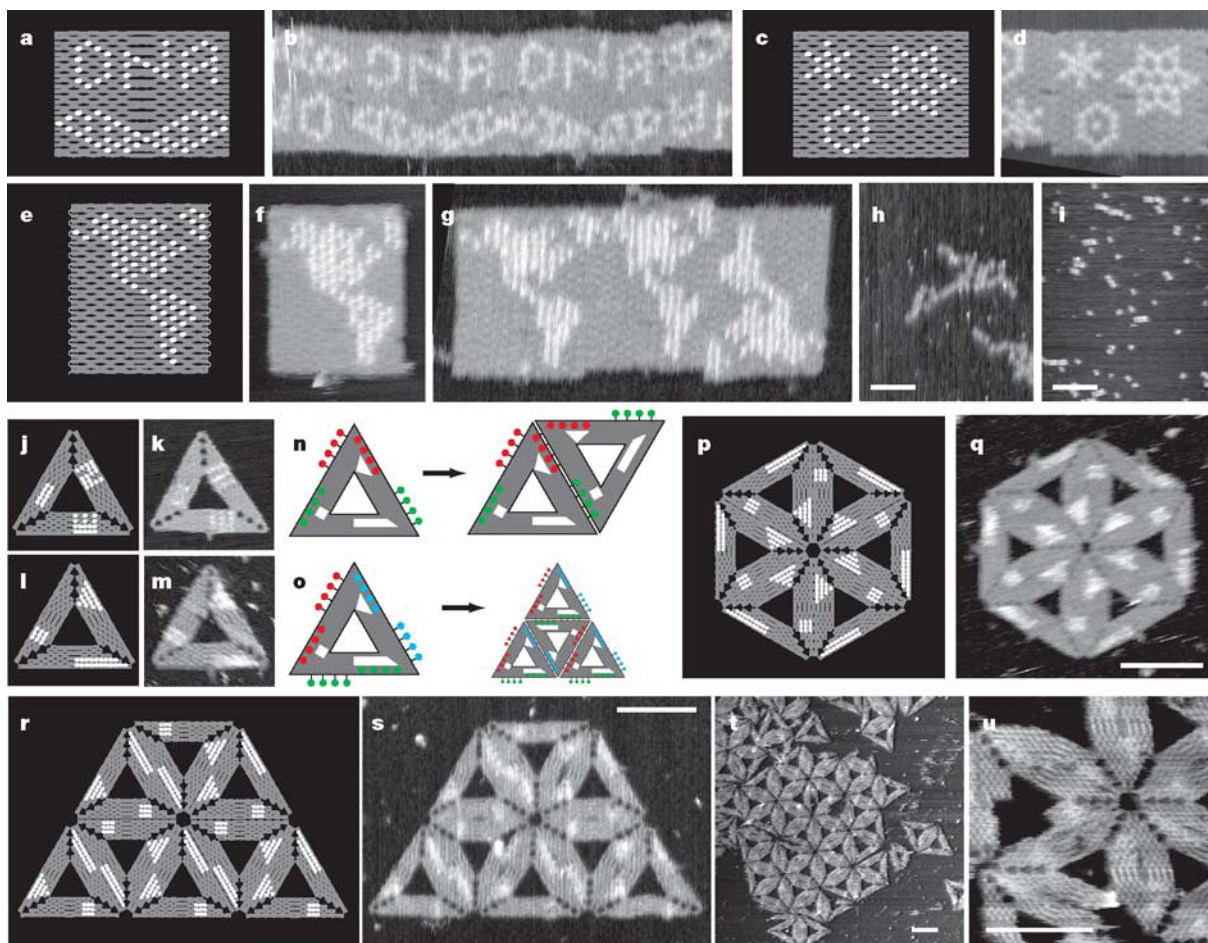


Figure 1.5 - Nanostructure d'ADN [Rothemund 06]. Echelle 100nm. Sont illustrés successivement pour chaque structure, le modèle et l'image AFM correspondante.

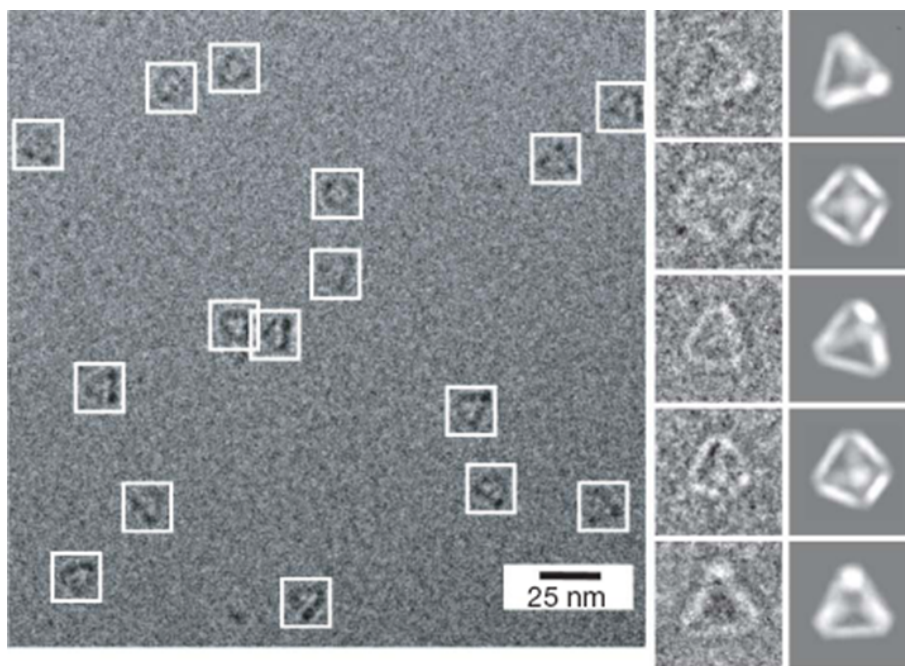


Figure 1.6 - Nanostructures polyhedrales [Yu He 08]. Images AFM et Cryo-EM. Les boites blanches indiquent les nanostructures.

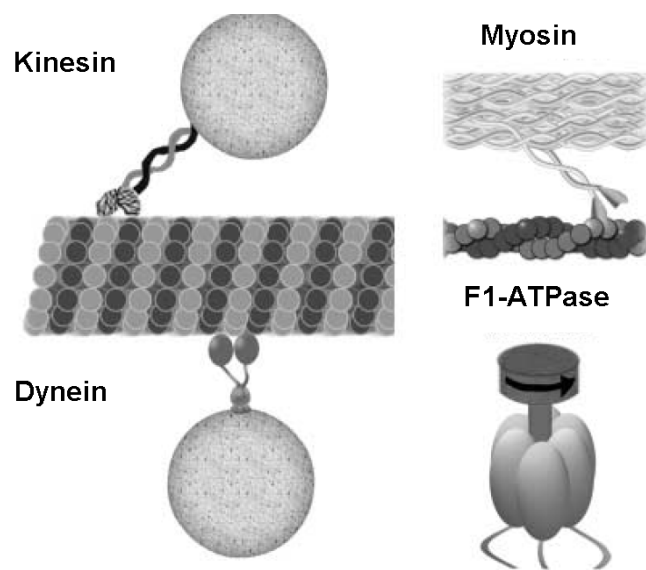


Figure 1.7 - Moteur moléculaire protéique composé de myosine et de kinésine [Hess 04].

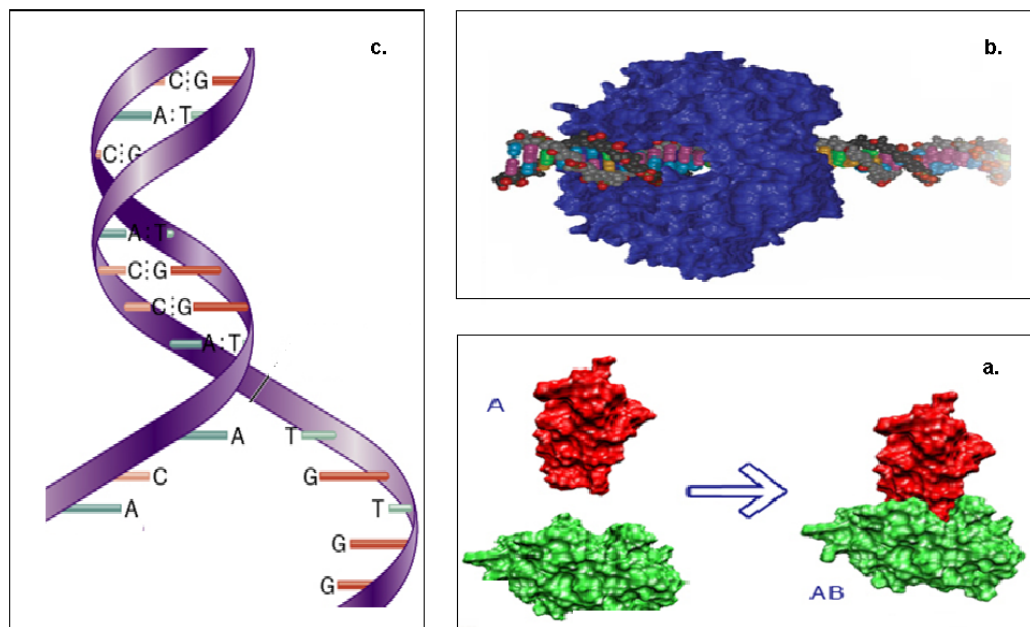


Figure 1.8 - Les velcros biologiques : l'hybridation de l'ADN/ADN, protéine/ADN et protéine/protéine.

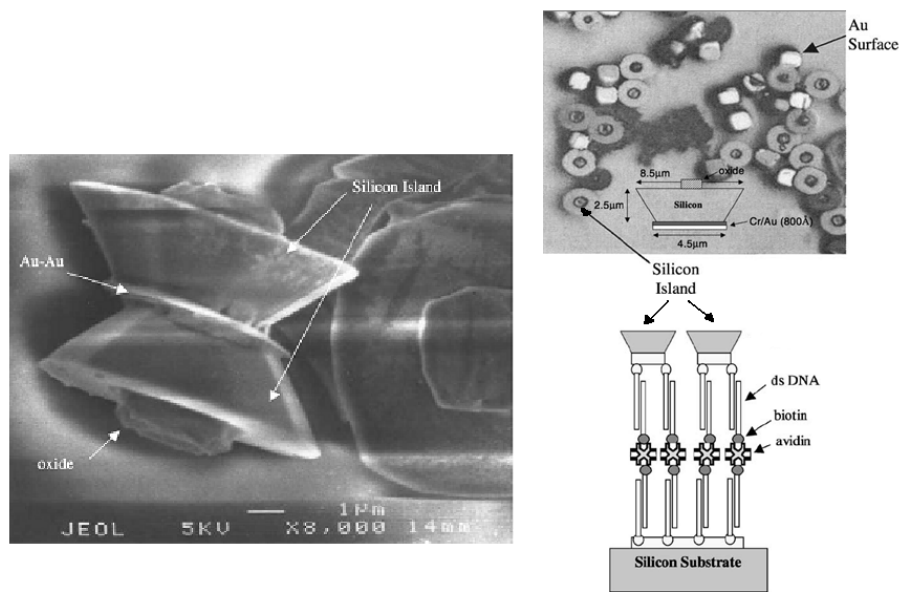
en trois types (figure 1.8) :

1. Processus d'auto-assemblage protéine/protéine (figure 1.8.a) : ce processus est aussi appelé "docking moléculaire". Les forces d'interaction intervenant dans ce processus pour réaliser la stabilité sont importantes et dépendent de la conformation 3D de chaque protéine. L'adhésion cellulaire par exemple, à la base de la construction des tissus organiques, est basée sur ce processus. Dans notre contexte la réalisation de la spécificité consisterait à utiliser autant de complexes protéiques différents que de micro-composants. L'utilisation de ces complexes nécessiterait donc la caractérisation de la stabilité pour chaque entité biologique utilisée et la mise au point de protocoles particulièrement étudiés pour chaque biomolécule. Les coûts de cette caractérisation et de la fonctionnalisation sont des inconvénients majeurs.
2. Processus d'auto-assemblage protéine/ADN (figure 1.8.b) : il existe un nombre presque illimité de protéines capables de se lier à l'ADN. L'ADN Polymérase, l'une de ces protéines, est connue pour son rôle dans la réparation de l'ADN. Le principal inconvénient est le coût de la caractérisation de la stabilité de chaque entité biologique utilisée.
3. Processus d'auto-assemblage ADN/ADN (figure 1.8.c) : ou *hybridation de l'ADN*. Cet auto-assemblage biomoléculaire sert de principe de base dans la technologie des puces à ADN [Carlson 06]. Le protocole de fixation, appelé *fonctionnalisation*, est aujourd'hui maîtrisé même si la densité des molécules effectivement fixées reste inconnue. Cette problématique reste importante puisqu'il n'existe pas aujourd'hui de technique fiable de quantification des biomolécules pour ce type de protocole. Néanmoins, la maîtrise des protocoles de fonctionnalisation et de synthèse est un avantage certain pour l'utilisation de ce processus. En effet, il est possible de synthétiser des molécules d'ADN artificielles répondant à des critères d'adhésion et de spécificité précis. Ces molécules d'ADN sont capables de se reconnaître spécifiquement entre elles et sont synthétisées avec le même protocole. De plus, la spécificité peut être réalisée avec des molécules différentes. La richesse et la diversité des travaux montre l'abondance des connaissances sur l'auto-assemblage à base d'ADN. Ceux-ci sont basés sur les propriétés programmables et intrinsèques de la molécule d'ADN.

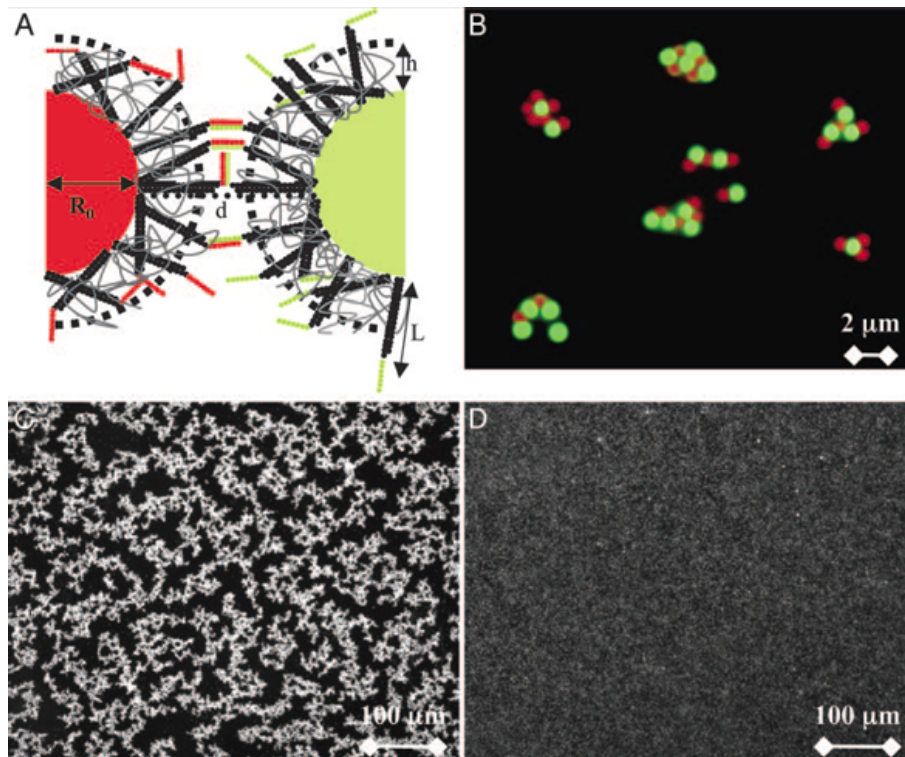
L'hybridation de l'ADN semble une solution prometteuse pour l'auto-assemblage biomoléculaire programmable. De même, de nombreuses combinaisons entre complexes protéiques et nucléiques peuvent aussi être envisagées pour réaliser le velcro biologique.

A cette échelle microscopique, deux travaux majeurs et pionniers sont à retenir. En 2003, McNally et al. [McNally 03] réalisent l'auto-assemblage de microparticules de l'ordre de 5 à 10 micromètres en utilisant le principe du docking moléculaire (ligand/récepteur), du complexe protéique biotin-avidin. Les microparticules en silicium, chrome et or (voir figure 1.9), sont fabriquées par photolithographie. De même, Valignat et al. [Valignat 05] réalisent un assemblage spécifique à partir de l'hybridation entre les molécules d'ADN simple brins fixées à la surface des microparticules en polystyrène d'un micromètre de diamètre. Ces particules sont transportées et approchées les unes des autres en utilisant un dispositif de manipulation sans contact, des pinces optiques.

Les travaux réalisés en auto-assemblage microscopique par processus biologiques ne s'appuient sur aucune mesure ni sur aucune caractérisation mécanique ni physique. L'auto-assemblage par processus biologiques est un processus dont les composées principales sont à étudier et à caractériser. Il est nécessaire d'identifier les limites physiques et/ou mécaniques pour une exploitation générale de ce processus. En effet, la stabilité et la spécificité doivent être caractérisées



**Figure 1.9** - Auto-assemblage utilisant l'ADN et les interactions avidin/biotin sur la surface en or des deux particules (Gauche). (Milieu) Image MEB des particules en silicium micro-fabriquée dans [McNally 03]. (Droite) Principe de l'auto-assemblage par ADN et avidin/biotin.



**Figure 1.10** - Auto-assemblage spécifique et réversible [Valignat 05]. (A) Des microparticules recouvertes d'oligonucléotides : les oligonucléotides utilisés sont synthétisés de telle sorte que la partie attachée à la bille est de type double brin, et la partie libre est de type simple brin. Les parties simples brins des oligonucléotides sur les billes rouges et sur les billes vertes sont complémentaires et donc spécifiquement liées. (B) Assemblage spécifique, les billes rouges sont liées aux billes vertes. (C) En stabilisant la température à 23C, l'assemblage est stable. (D) En augmentant la température à 50C, les billes sont complètement dispersées.

afin d'identifier les paramètres de contrôle et d'optimiser le processus pour réaliser des systèmes 3D assemblés en parallèle.

Pour un auto-assemblage microscopique par processus biologiques il est important de distinguer deux échelles concernant la force de guidage. D'une part, à *une échelle locale*, la force de guidage augmente la probabilité de rencontre entre les entités biomoléculaires collées à la surface des composants. D'autre part, à *une échelle globale*, la force de guidage augmente la probabilité de rencontre entre les composants à assembler. En raison de la taille des composants mille fois plus grande que la taille des molécules, il n'est pas possible de guider les micro-composants par une agitation thermique. La force de guidage globale est optimisée dans le cadre du projet Golem notamment dans les travaux de Emir Vella et al. [Vela Saavedra 09]. Dans ce travail de thèse, il s'agit d'étudier particulièrement la stabilité et la spécificité locales.

## 2.3 Choix d'un processus biologique

Il semble impératif que l'auto-assemblage développé dans ce travail réponde aux exigences suivantes :

- un taux d'assemblage important,
- une indépendance par rapport à la géométrie des pièces,
- un processus d'assemblage parallèle.

Ces trois points ne sont réalisables que si la spécificité et la stabilité dépendent uniquement du processus d'auto-assemblage biomoléculaire choisi. L'optimisation de ces deux propriétés est nécessaire.

L'ADN présente un grand potentiel en termes de spécificité (caractère programmable) et de stabilité (grandeur des forces de liaison mises en jeu). L'ADN semble donc être le candidat idéal pour l'optimisation de la spécificité et la stabilité de l'auto-assemblage microscopique. Son étude est proposée dans la section suivante.

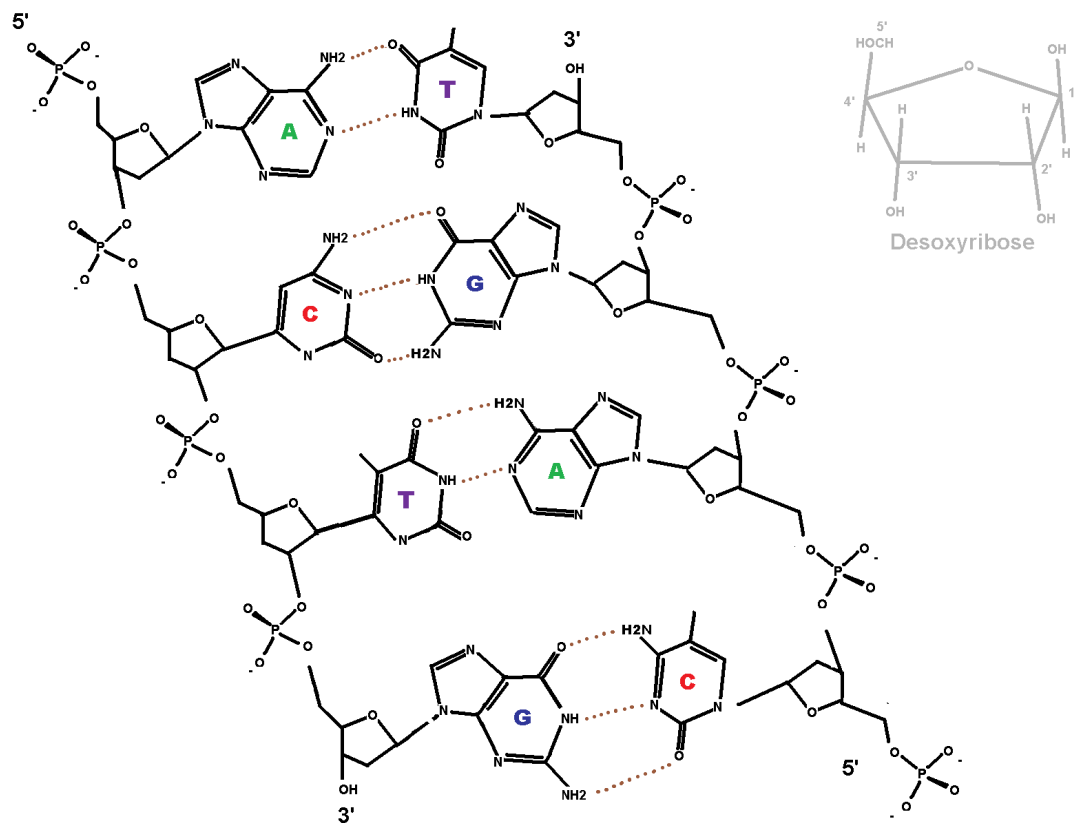
# 3 L'ADN pour l'auto-assemblage

L'utilisation de l'ADN suppose l'utilisation du processus biomoléculaire d'auto-assemblage propre à cette molécule. Ce processus est appelé hybridation. L'objectif de cette partie est de décrire l'ADN et les propriétés mises en jeu lors des phénomènes d'adhésion. Cette étude n'est pas exhaustive mais elle pose les fondements nécessaires à son application. De plus, quelques exemples d'utilisation sont donnés pour illustrer les possibilités de réalisation.

## 3.1 Structure de l'ADN

L'ADN se présente sous la forme d'un assemblage en double hélice composée de deux chaînes nucléiques complémentaires appelées *brins*. Les petites unités formant ces chaînes sont des

*nucléotides*, dont la composition est un groupement phosphate lié à un sucre (désoxyribose), lui-même lié à une base azotée qui, pour l'ADN, peut être de quatre types : Adénine (A), Guanine (G), Cytosine (C), et Thymine (T). L'adénine et la guanine sont des *purines*, tandis que la cytosine et la thymine sont des *pyrimidines*. Par abus de langage, le terme *base* est souvent utilisé pour désigner un nucléotide. L'agencement des bases sur une même chaîne est orienté. Celui-ci est effectué à l'aide de l'orientation de leur sucre (figure 1.11). Ainsi, le phosphate relie le carbone 3 du premier sucre au carbone 5 du sucre suivant, qui à son tour associe son carbone 3, à travers un phosphate, au carbone 5 du sucre suivant, et ainsi de suite, de proche en proche. Cette propriété détermine une chaîne de nucléotides orientée de l'extrémité 3' à l'extrémité 5'.

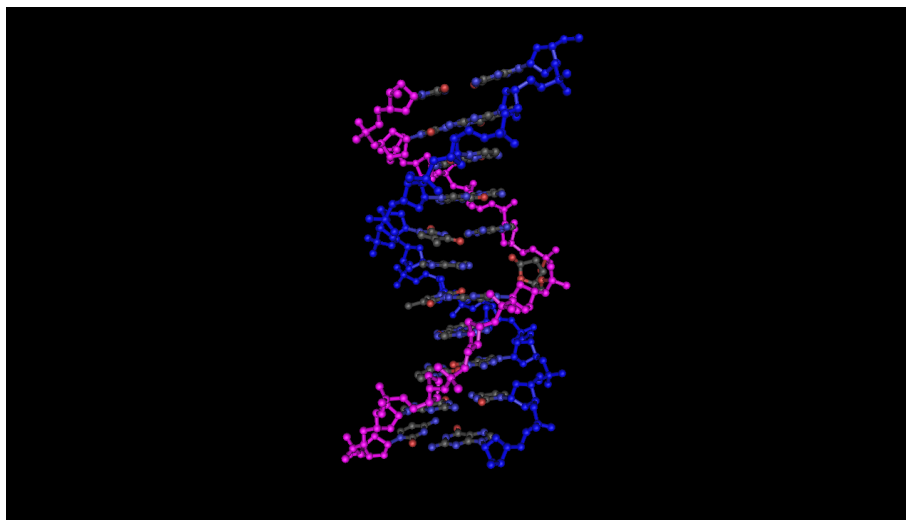


**Figure 1.11** - Représentation de la composition chimique d'une molécule d'ADN. Assemblage de deux brins orientés 3'-5' antiparallèles (TGAC) et (GTCA). L'orientation est donnée par les carbones 3' et 5' du désoxyribose.

La complémentarité des deux brins est basée sur le principe de complémentarité de Watson-Crick (WC) [Watson 53]. Celui-ci associe à chaque base une base complémentaire. Ainsi, à chaque base A d'un brin est associée une base T du brin complémentaire, et inversement. Et à chaque base C d'un brin est associée une base G du brin complémentaire, et inversement. De plus, l'assemblage en double hélice nécessite une orientation opposée des deux brins complémentaires, et la double hélice est donc dite *antiparallèle*.

Une molécule d'ADN peut être décrite sous plusieurs formats. La structure primaire consiste en la composition orientée en bases azotées de l'un des deux brins (voir figure 1.12). On parle alors de *séquence en bases azotées* ou *séquence nucléotidique*. La structure secondaire représente la conformation 3D de la molécule.

Plusieurs formes sont connues [Phan 06]. Pour la forme B, et une concentration de  $1M$  en  $[Na^+]$ , le diamètre de la molécule est de  $2nm$  et le tour d'une hélice est d'environ  $0.34nm$ . Pour cette même forme, la *longueur de persistance*, propriété importante de l'ADN, est d'environ 11 nucléotides pour un brin [Williams 01]. La notion de longueur de persistance dépend de la salinité du milieu et est primordiale pour l'élasticité de la molécule.



**Figure 1.12** - Structure 3D d'une double hélice d'ADN; brin A (cgcganacgcc), brin B (ggcg-tatcgcg) [Beger 98].

L'ensemble des propriétés et des termes associés pour qualifier la structure de l'ADN dans ce document sont résumés dans le tableau suivant.

Propriétés	Définition
Séquence	Composition orientée en bases azotées.
Orientation antiparallèle	La double hélice est formée de 2 brins <b>opposés</b> (figure 1.11) : $3'-5' / 5'-3'$ .
Structure secondaire	Conformation 3D d'une molécule.
Nucléotide	Base azotée, elles sont de 4 types A, T, G et C.
Purine	Les bases : A et G.
Pyrimidine	Les bases : T et C.
Complémentarité WC	Les deux couples : A-T et G-C.
Chaîne latérale	Chaîne composée des phosphates et désoxyriboses de chaque nucléotide d'un brin.
Longueur de persistance	Segment de 11 nucléotides en forme B.
Axe-db	Axe hypothétique de la double hélice.
PM	Configuration " <i>Perfect-Match</i> ".
MM	Configuration " <i>Mis-Match</i> ".

**Tableau 1.2** - Tableau récapitulatif des propriétés et des termes qualificatifs de la structure de l'ADN.

## 3.2 L'hybridation de l'ADN

Dans son contexte d'origine, l'ADN joue un rôle majeur dans la cellule vivante. Dans les cellules eucaryotes, l'ADN est contenu dans le noyau et une petite partie se trouve dans la matrice des mitochondries ainsi que dans les chloroplastes pour les cellules végétales. Dans les cellules procaryotes, l'ADN est contenu dans le cytoplasme. Certains virus possèdent également de l'ADN dans leur capsid.

*In-vivo*, l'ADN intervient dans plusieurs phénomènes, notamment dans la réplication du matériel génétique de la cellule. Pendant ce processus, la cellule se divise dans le but de régénérer ou spécialiser des tissus. A partir de chaque brin de la molécule, un brin complémentaire est synthétisé, nucléotide par nucléotide. Ainsi, chaque nouvelle molécule est identique à la molécule d'ADN initiale. Un grand nombre d'enzymes interviennent dans le mécanisme moléculaire de la duplication de l'ADN formant le complexe enzymatique de réplication. Cette propriété nécessite un apport considérable d'énergie pour faciliter la rencontre entre le nucléotide du brin initial et le nouveau nucléotide du brin formé.

Aujourd'hui, il est possible de synthétiser et d'utiliser l'ADN en dehors de son contexte *in vivo*. Par exemple, l'utilisation de l'ADN dans des conditions expérimentales (« *in vitro* ») est au cœur de la technologie des bio-puces à ADN [Erickson 03]. Cette technique de laboratoire est basée sur le principe de l'hybridation et a pour objectif de savoir si un gène est représenté dans une cellule. Ainsi, en s'inspirant du phénomène de réplication de l'ADN et de manière similaire à la réplication, l'hybridation met en jeu la complémentarité de Watson-Crick pour associer deux brins d'ADN complémentaires. Cependant, pour l'hybridation, il ne s'agit pas de construire des brins complémentaires, mais d'auto-assembler deux brins d'ADN entiers de séquences complémentaires. Cette association nécessite un apport d'énergie facilitant la rencontre entre les brins et l'assemblage a seulement lieu si les paires de bases sont complémentaires sur un même plan, en établissant des liaisons hydrogène (H).

Comme tout processus d'auto-assemblage, l'hybridation présente les 3 éléments clés précédemment définis : la force de guidage, la spécificité et la stabilité. La force de guidage est réalisée par le mouvement stochastique brownien dû à l'apport d'énergie par agitation thermique. Pour les deux derniers éléments, ils sont dus, d'une part à la reconnaissance moléculaire entre les paires de bases pour la spécificité, et d'autre part, aux liaisons hydrogène pour la stabilité.

### 3.2.1 La Spécificité

Le caractère spécifique de la molécule d'ADN consiste en l'unicité de la composition d'un brin, appelée *la séquence*. En effet, connaissant la composition d'un brin, il est facile d'en déduire le brin complémentaire. Cependant, plusieurs exceptions sont possibles. En effet, deux brins non complémentaires peuvent s'assembler partiellement. Ces exceptions sont appelées "mismatch", notées MM, par opposition au "perfect match", notées PM, pour l'hybridation parfaite en double hélice [Zuker 03].

Le principe de complémentarité fait intervenir des liaisons H entre les bases complémentaires. Elles sont au nombre de 2 entre A et T, et 3 entre G et C. Ces liaisons H se forment spontanément à la rencontre des deux bases complémentaires : une purine et une pyrimidine (cf. section 1.3.1)



La liaison H est une liaison chimique non covalente, de type dipôle-dipôle (figure 1.11). Une telle liaison nécessite la présence d'un donneur de liaison hydrogène et d'un accepteur. Lorsqu'une liaison hydrogène s'établit, les deux hétéroatomes, nucléotides, se trouvent dans le même plan à une distance d'environ  $0.2nm$ . La configuration spatiale des nucléotides est donc capitale à la construction de ces ponts (figure 1.12).

### 3.2.2 La Stabilité

La liaison H est de basse intensité, vingt fois plus faible que la liaison covalente classique, sa valeur est d'environ  $0.5 Kcal$ . Elle relie des molécules en impliquant un atome d'hydrogène. Son origine est à 90% électrostatique [Martin 99]. Bien que de la même famille que les forces de van der Waals, les liaisons hydrogène s'en distinguent par leur intensité : leur énergie de liaison est environ dix fois supérieure à ces dernières. La stabilité du processus d'hybridation est construite par l'ensemble des ponts hydrogène de chaque paire de bases complémentaires assemblées.

La spécificité ainsi que la stabilité sont donc le résultat du principe de complémentarité de Watson-Crick. Ce sont des caractères simultanés et propres à la molécule d'ADN. L'optimisation et le contrôle de l'A.A microscopique nécessite l'optimisation de l'A.A biomoléculaire. L'hybridation est contrôlée par des facteurs environnementaux tels que la température, ou propres à la molécule, telle que sa séquence. Ces paramètres de contrôle sont détaillés dans la suite.

## 3.3 Les paramètres de contrôle de l'hybridation

L'étude de l'hybridation nécessite une considération toute particulière des différents paramètres environnementaux. En effet, ces derniers peuvent influencer le comportement de l'ADN. Ces paramètres sont aujourd'hui connus [SantaLucia 98]. Il s'agit principalement de la *température* et de la *composition saline* du solvant. La séquence en bases azotées et sa longueur influencent aussi le comportement de la molécule. Ces deux derniers paramètres sont *intrinsèques* par opposition aux deux premiers *extrinsèques*.

### 3.3.1 Paramètres extrinsèques

#### 3.3.1.1 La Température

Les liaisons hydrogène sont sensibles à la température. En effet, à moins de zéro degré, les liaisons hydrogène sont sous forme cristalline, et à  $100^{\circ}C$ , toutes les liaisons hydrogène sont rompues. La température contrôle la formation des liaisons H lors de l'hybridation. Elle est donc un paramètre clé.

### 3.3.1.2 La Composition du milieu

Sur chaque brin de la double hélice et pour chacun des nucléotides, le groupement phosphate est dirigé vers l'extérieur de la double hélice [Watson 53]. Chaque brin est donc chargé négativement. Dans un milieu aqueux, ces charges négatives sont équilibrées par la présence de cations monovalents tels que  $Na^+$  ou  $K^+$ . Les bases azotées sont à l'intérieur de l'hélice et constituent la partie hydrophobe de la molécule. La concentration du milieu en cation agit sur la configuration spatiale de la chaîne nucléique [Sae 83]. Ainsi, la composition du milieu aqueux joue un rôle très important sur l'équilibre et la stabilité de la double hélice et donc sur le processus d'hybridation. La concentration ionique peut être considérée comme un paramètre de contrôle.

### 3.3.2 Paramètres intrinsèques

La composition en bases azotées [Matthias Rief 99], la longueur de séquence et le pourcentage des nucléotides G et C de la molécule d'ADN influencent le processus d'hybridation. Le pourcentage en nucléotides G et C est un indice paramétrable de la force de liaison entre deux brins complémentaires (3 liaisons H). Ce pourcentage peut influencer la spécificité du processus d'hybridation. Ainsi, si la séquence comporte deux segments complémentaires sur le même brin avec un très haut pourcentage en GC, par exemple, une épingle à cheveux peut être formée (figure 1.13).

La concentration en oligonucléotides influence aussi la stabilité et la spécificité du processus puisqu'elle détermine la probabilité de rencontre entre deux brins.

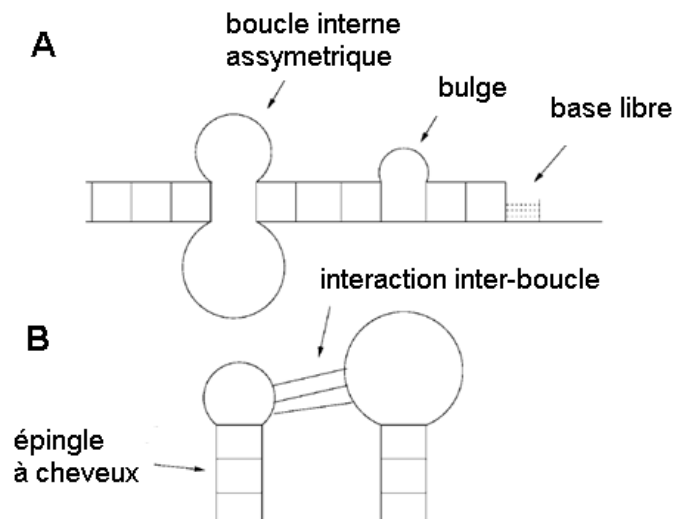


Figure 1.13 - Structures secondaires imparfaites : mismatch [Dimitrov 04]

L'ADN semble être le candidat idéal pour la réalisation d'une colle biologique à la fois *stable* et *spécifique* dédiées à l'auto-assemblage. Néanmoins, l'utilisation de ce complexe biologique à l'échelle microscopique pour assembler des composants de taille 1000 fois supérieure au diamètre de la molécule est complexe.

## 4 Problématique de la thèse

Un récapitulatif des besoins en auto-assemblage microscopique est détaillé dans les parties 1.1 et 1.2 de ce chapitre. Il apparaît que, parmi les trois éléments définissant un processus d'auto-assemblage, la stabilité et la spécificité globales peuvent être optimisées en utilisant l'ADN comme "velcro". A l'échelle nanoscopique, l'utilisation potentielle de la molécule dans l'auto-assemblage est largement exploitée et démontrée. La caractérisation de l'assemblage biologique à l'échelle microscopique reste néanmoins un large champ d'investigation.

Concernant la stabilité, les forces d'interaction propres à l'ADN doivent être caractérisées. Plusieurs techniques de caractérisation des interactions biomoléculaires ont été proposées ces vingt dernières années. Celles-ci sont principalement basées sur des principes physiques ou thermodynamiques et sont par exemple, les rayons X, la cristallographie, ou encore la spectroscopie à résonance magnétique.

### 4.1 Caractérisation des forces d'interaction de l'ADN

Dans le cas de la caractérisation des interactions des acides désoxyribonucléiques, plusieurs techniques existent. La technique de pression osmotique est la première utilisée pour mesurer l'interaction non spécifique de condensation de l'ADN (il s'agit des interactions entre doubles hélices) [Rau 84]. Pour les interactions moléculaires plus fines, dans lesquelles l'adhésion et l'orientation interviennent, comme la reconnaissance moléculaire, une mesure directe est nécessaire. Ainsi, des techniques comme les pinces magnétiques [Smith 92] ou optiques [Perkins 94] sont utilisées pour mesurer les propriétés individuelles d'élongation d'une double hélice de longueur  $10 \mu m$ . D'autres types de propriétés sont explorées comme l'ouverture d'une double hélice par déroulement. En utilisant la technique des micro-aiguilles souples, l'équipe de Roulet et al. [Essevaz-Roulet 97] obtient des forces allant de 10 à  $15 pN$ . En 2002, Bockelmann et al. [Bockelmann 02] confirme les résultats obtenus par ses prédécesseurs en utilisant une autre technique, le piégeage optique par interférométrie. Ils apportent un élément nouveau concernant la dépendance de la force d'interaction à la teneur en GC des séquences nucléotidiques.

Ainsi, plusieurs études ont été menées pour caractériser l'ouverture par déroulement des deux brins et les interactions non spécifiques entre double hélices. Les informations issues de ces différents travaux sont d'une grande importance puisque la compréhension de la réponse de la molécule d'ADN à une perturbation mécanique externe est facilitée. Toutefois, la séparation des brins par une force appliquée dans le sens de l'axe hypothétique de la double hélice reste un phénomène inexploré. De plus, dans notre cas d'étude, le comportement collectif des brins d'ADN collés sur une surface prend toute son importance. Cet état constitue une difficulté supplémentaire à la caractérisation de la stabilité. En effet, il s'agit d'un système moléculaire de nature stochastique contenant des centaines de milliers de molécules. Par conséquent, la caractérisation des forces d'interaction nécessite l'utilisation d'une plateforme instrumentale dédiée et mesurant les interactions engendrées par une force appliquée dans le sens perpendiculaire au plan des surfaces.

Parmi les instruments de mesure des forces surfaciques existants, le microscope à force atomique (AFM) est l'instrument le plus approprié, en raison de sa sensibilité dans la mesure

d'effort, de l'ordre de la centaine de piconewtons. Les conditions environnementales de l'auto-assemblage microscopique peuvent être reproduites lors de la séparation de deux objets fonctionnalisés, telles que les forces de Van der Waals, électrostatiques, la traction hydrodynamique, les vitesses d'approche retrait, et l'application d'une force externe. Cet outil est choisi pour l'ensemble des expériences menées dans ce document.

## 4.2 Plan de la thèse

L'objectif de ce travail de thèse est de répondre à des besoins particuliers du projet Golem pour la réalisation de l'auto-assemblage microscopique par processus biologiques. La forte interaction scientifique entre les différentes équipes de recherche du projet a guidé ce travail de façon à répondre à différentes questions. En effet, d'une part, la conception de la chambre d'assemblage, coordonnée par l'EPFL (Ecole Polytechnique Fédérale de Lausanne) a nécessité l'identification et l'étude des paramètres de contrôle environnementaux. D'autre part, la modélisation et la simulation du processus d'auto-assemblage global, coordonnées par l'université de Stuttgart, ont suscités l'estimation théorique et expérimentale des forces d'interaction biomoléculaires propre au "velcro biologique". Ainsi, les questions du choix de la molécule biologique adéquate, de l'estimation des forces en jeu, et de l'identification et l'étude des paramètres de contrôle sont traitées. Dans ce document, ces problématiques sont étudiées et quelques réponses sont apportées, notamment pour l'estimation des forces d'interaction mises en jeu.

Au cours de ce premier chapitre, l'ADN est identifié comme une solution originale pour la problématique de l'auto-assemblage microscopique par processus biologiques. Son utilisation pour la réalisation du velcro biologique entre les micro-objets est étudiée. Dans le deuxième chapitre, deux approches sont explorées pour approximer les forces mises en jeu dans l'interaction entre les objets. La première s'intéresse au niveau individuel au comportement d'un brin d'ADN. La seconde considère un niveau global à partir d'une approche thermodynamique. Ces deux études sont comparées pour évaluer la validité de chacune vis-à-vis du système étudié, des expériences envisagées et de l'influence des paramètres de contrôle.

La caractérisation expérimentale d'un velcro composé de deux populations de brins d'ADN complémentaires fait l'objet du troisième chapitre de ce document. Des expériences, réalisées à l'aide d'un AFM commercial, rendu disponible au sein du laboratoire national de Londres (NPL), sont analysés. La stabilité du velcro et ses paramètres de contrôle sont examinés. Pour atteindre cet objectif, une méthode originale d'analyse des données couplée à une approche énergétique est proposée.

L'élément central de ce travail de thèse est le choix de l'ADN. Ce choix repose en partie sur le caractère programmable de la molécule. Par conséquent, l'étude de la programmabilité du velcro est nécessaire. Un algorithme dédié à la génération de séquences optimales et spécifiquement dédié pour la problématique de l'auto-assemblage est proposé dans le quatrième chapitre. Pour valider cette approche, des expériences obtenues par la microscopie à force atomique au laboratoire AMIR de l'université d'Oldenburg (Allemagne), sont analysées. Dans ces expériences réalisées sur une séquence générée à l'aide de l'algorithme proposé, un nouveau paramètre temporel est introduit et son influence sur la stabilité et la spécificité globale du velcro est détaillée.

Enfin, les conclusions et perspectives de ce travail sont discutées dans la dernière partie de ce document.

---

# ETUDE DES PARAMÈTRES DE CONTRÔLE DE LA STABILITÉ

---

Au chapitre précédent, la stabilité a été identifiée comme une propriété fondamentale de l'A.A microscopique à base d'ADN. A l'échelle du velcro, cette propriété est locale et consiste en la capacité des deux populations de brins d'ADN à rester dans l'état hybridé. A l'échelle des micro-composants, cette propriété est globale et repose sur son aptitude à rester dans l'état assemblé. La stabilité est donc une propriété à la fois locale et globale, mais son optimisation peut être réalisée d'un point de vue local. Ce chapitre s'attache à étudier cette propriété de ce point de vue.

La prédiction et le contrôle de cette propriété sont des éléments majeurs, notamment pour les partenaires européens du projet Golem. Ainsi, il s'agit de fournir des valeurs de paramètres par exemple pour la température, la salinité du milieu d'assemblage, les forces d'interaction, d'une part pour la conception de la chambre d'assemblage coordonnée par l'EPFL (École Polytechnique Fédérale de Lausanne), et d'autre part pour le simulateur global de l'A.A des micro-objets coordonné par l'université de Stuttgart. Ces informations dépendent de la séquence utilisée pour construire le velcro, et le choix de la séquence résulte de l'intensité de l'interaction désirée. L'étude de la stabilité nécessite ainsi la mise en expression et la quantification de cette propriété en fonction des paramètres de contrôle.

La stabilité étant de nature biomoléculaire, il s'agit d'exprimer cette propriété en terme d'*énergie d'interaction*. Plusieurs approches existent pour prédire ce type d'énergie [Brand-sdal 03]. Ainsi, l'objectif de ce chapitre est de déterminer l'approche appropriée pour notre cas d'étude en particulier.

Dans cette étude, l'adhésion spécifique entre les micro-objets est réalisée par un velcro biologique constitué de deux *populations* de brins d'ADN complémentaires. Une première population est utilisée pour fonctionnaliser une surface d'un micro-composant. La seconde population, dont les brins sont complémentaires à la première, est employée pour fonctionnaliser une surface d'un deuxième micro-composant. Une illustration de ce système est proposée figure 2.1. Dans cha-

cune des populations, les brins d'ADN sont identiques et aucune interaction n'est supposée possible entre les brins d'une même population.

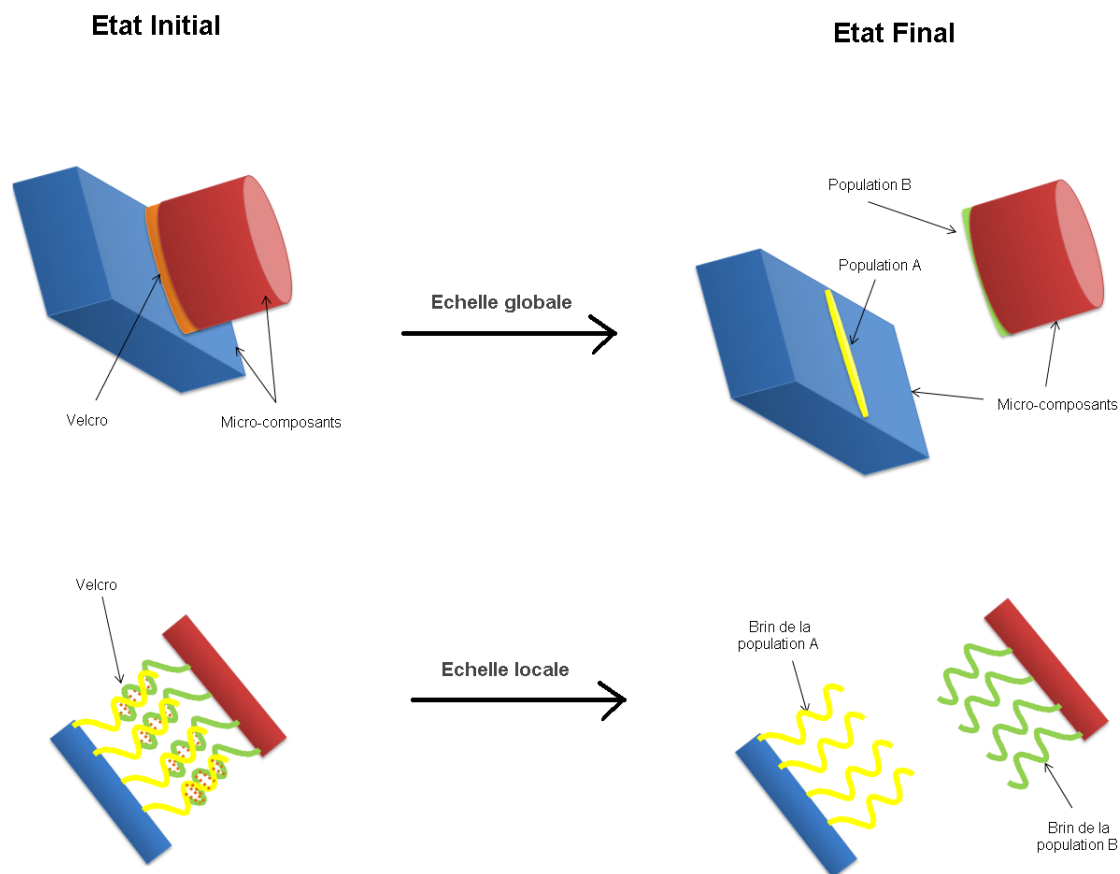


Figure 2.1 - Principe de l'A.A. microscopique basé sur l'ADN.

Au début de ce travail de modélisation, au sein du laboratoire, des outils de modélisation des forces d'interaction biomoléculaires étaient disponibles. Les travaux de thèse de Bruno Daunay [Daunay 07] sur le "*docking moléculaire*" et la disponibilité du logiciel de modélisation moléculaire MOE ont conduit ce travail vers une approche dans laquelle la contribution de chaque brin d'ADN du velcro est considérée d'un point de vue individuel. Dans la première partie de ce chapitre, il s'agit de détailler cette **approche individuelle**. Ainsi, le comportement des brins d'ADN pendant l'hybridation est évalué et les besoins propres de notre problématique sont soulignés.

Une deuxième approche est proposée dans la deuxième partie de ce chapitre. Cette dernière répond aux difficultés rencontrées dans l'approche individuelle et aux besoins particuliers de la problématique de l'A.A. Cette approche est appelée **approche populationnelle**, puisqu'il s'agit d'analyser l'ADN du point de vue de sa fonction de velcro dans son ensemble. Le système étudié est alors un ensemble statistique et thermodynamique.

Les deux approches sont comparées pour évaluer la validité de chacune vis-à-vis du système étudié, de l'étude expérimentale envisagée et de l'influence des paramètres de contrôle.

# 1 Approche individuelle

La caractérisation de la stabilité à l'échelle individuelle en fonction des paramètres de contrôle est au cœur de cette partie. Il s'agit dans le premier paragraphe d'introduire les outils méthodologiques de calcul, issus de la modélisation moléculaire, utilisés pour quantifier la stabilité à cette échelle.

Pour corrélérer les mesures expérimentales définies dans le chapitre 3, la conversion de l'énergie d'interaction en force d'interaction est primordiale. Cette problématique et la méthode d'estimation de la force à partir de l'énergie sont l'objet du second paragraphe.

Enfin, dans le troisième paragraphe, l'approche individuelle est analysée puis replacée dans le contexte de la stabilité de l'A.A microscopique à base d'ADN.

## 1.1 Caractérisation de la stabilité

A l'échelle moléculaire, plusieurs méthodes existent pour l'estimation de l'énergie d'interaction, comme par exemple la modélisation moléculaire ou la modélisation de Monte carlo [Brandstadal 03]. En modélisation moléculaire, l'estimation de l'énergie est basée sur l'équation de Schrödinger.

$$H\Psi = E\Psi \quad (2.1)$$

$$E = E_c + E_p \quad (2.2)$$

avec  $H$  l'opérateur hamiltonien,  $E$  l'énergie composée de l'énergie cinétique  $E_c$  et potentielle  $E_p$  et  $\Psi$  la fonction d'onde. L'énergie d'interaction  $V$  qui caractérise l'interaction entre deux brins d'ADN est une composante de l'énergie potentielle  $E_p$ . Cette énergie dépend de la conformation et fondamentalement de la structure très particulière des deux brins en interaction.

Une estimation quantique de l'énergie qui prend en compte l'ensemble des électrons de chaque atome constituant chacun des brins est une méthode très couteuse en temps de calcul. Une évaluation par champs de forces est souvent préférable. L'estimation de l'énergie d'un atome, basée sur la notion de champ de forces, ignore les mouvements d'électrons et calcule l'énergie de l'atome comme une fonction de la position du noyau, en utilisant une méthode d'approximation, comme par exemple la méthode de Born-Oppenheimer [Leach 01].

Un champ de forces est un ensemble de paramètres numériques déterminés grâce à des techniques expérimentales telles que la cristallographie. Ces paramètres sont affectés aux différentes composantes de l'énergie selon un modèle physique simple des liaisons chimiques. L'hypothèse de ce modèle est qu'une liaison atome-atome est un segment liant les deux noyaux des atomes. Ce dernier peut être étiré, plié ou encore tordu.

L'énergie potentielle d'un atome est la somme de plusieurs termes :

- $E_{str}$  énergie d'extension de la liaison,
- $E_{tor}$  énergie de torsion,
- $E_{vdw}$  énergie de Van der Waals,
- $E_{ele}$  énergie électrostatique,
- $E_{sol}$  énergie de solvation.
- ...etc.



Une fois les paramètres de chaque type d'énergie fixés, celles-ci sont affectées à une expression analytique de l'énergie d'un atome selon l'équation suivante :

$$E_p(x) = E_{str}(x) + E_{tor}(x) + E_{vdw}(x) + E_{ele}(x) + E_{sol}(x) + \dots etc. \quad (2.3)$$

avec  $x$  le vecteur position de l'atome considéré.

Les composantes de l'énergie potentielle  $E_p(x)$  peuvent être classées en deux composantes principales, l'**énergie de liaison** et l'**énergie de non-liaison**. L'énergie de liaison est la somme de toutes les formes d'énergie faisant intervenir des liaisons covalentes. Elles prennent pour la plupart une forme quadratique. L'énergie de non-liaison est la somme de toutes les énergies induites par des liaisons non covalentes comme les liaisons de type Van der Waals et électrostatique. L'énergie d'interaction  $V$  correspond à l'énergie de non-liaison.

Afin de disposer de la puissance de calcul nécessaire à la modélisation moléculaire, le logiciel MOE (Molecular Operating Environment) a été utilisé. Ce logiciel dispose d'une interface conviviale et d'un générateur de structures moléculaires, parmi celles-ci la structure atomique des acides nucléiques. Ainsi, à partir d'une séquence en bases azotées, MOE génère automatiquement la structure 3D atomique de la double hélice correspondante et calcule la valeur de la variation d'énergie en fonction du champs de force choisit (Amber, CHARMM ...) [Guv 08].

Dans le but de comparer les valeurs d'énergie d'interaction théorique, calculées par MOE, et les valeurs de force d'interaction expérimentales, obtenues par l'AFM (cf. chapitre 3), il est primordial d'estimer une force d'interaction théorique.

## 1.2 Problématique de l'estimation de la force

A l'échelle globale (figure 2.1) la stabilité est caractérisée par l'aptitude des micro-composants, soumis à un processus d'A.A à base d'ADN, à rester dans l'état assemblé. Caractériser la stabilité globale revient donc à caractériser le processus de séparation des deux micro-composants. A l'échelle locale, échelle du velcro, la stabilité est définie par la quantité d'énergie fournie par le système moléculaire pour maintenir l'état hybridé des brins d'ADN complémentaires. En conséquence, une considération particulière est portée sur *l'axe hypothétique de la double hélice* (figure 2.2), puisque la séparation des micro-composants, donc des deux brins hybridés, est réalisée dans cet axe, perpendiculaire aux surfaces fonctionnalisées.

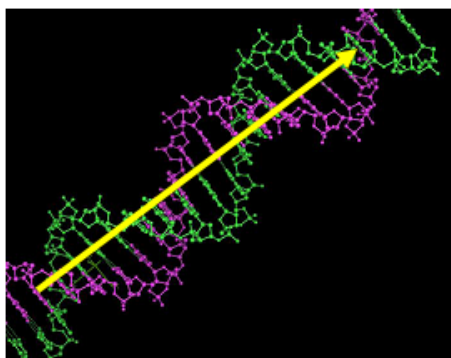


Figure 2.2 - Axe hypothétique de la double hélice.

La force peut être estimée à partir de l'étude de la variation d'énergie sur une variation de déplacement. Dans notre cas d'étude, celui-ci est supposé s'effectuer sur l'axe de la double hélice. D'un point de vue analytique, cette force exercée sur l'axe  $\vec{x}$  s'écrit :

$$\sum_{ext} F = -\frac{dV}{dx} \quad (2.4)$$

Comme l'expression analytique de l'énergie d'interaction est inaccessible à travers le logiciel MOE, une méthode alternative est nécessaire. La méthode utilisée pour l'estimation de la force à partir de l'énergie consiste à approximer l'énergie d'interaction par une fonction analytique [Daunay 07] à partir d'un ensemble de points. La dérivation de cette fonction donne ensuite une estimation de la force d'interaction.

L'expression analytique de l'énergie d'interaction pour un couple de brins d'ADN est évaluée à l'aide d'une extrapolation d'un ensemble de valeurs autour d'une position fixe. Ainsi, en fixant la position de tous les atomes d'un premier brin, et en déplaçant le brin complémentaire aléatoirement d'une distance  $\delta$  du premier brin selon l'axe de la double hélice, un ensemble de points décrivant l'espace énergétique de l'interaction est obtenu. L'extrapolation de cet espace énergétique est réalisée par l'équation suivante :

$$E(p, p_e, \delta) = \frac{1}{2}(p - p_e) + \delta \quad (2.5)$$

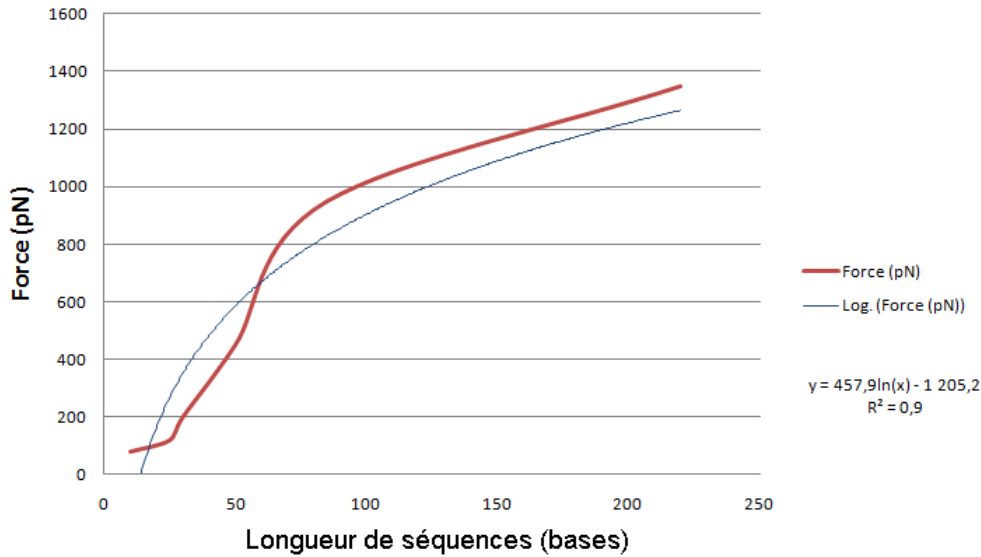
avec  $p$  la position du centre de masse de la molécule et  $p_e$  la position aléatoire au voisinage de  $p$ , relativement au déplacement  $\delta$ . La dérivation de cette expression relativement au déplacement donne une estimation directe de la force nécessaire pour conduire le système d'un état lié à un état non lié.

A l'aide du langage SVL propre au logiciel MOE, un programme est développé pour réaliser cette extrapolation. Ainsi, la force d'interaction est estimée pour des couples de brins de séquences en bases azotées arbitraires et de longueurs différentes variant de 10 à 220 bases. La courbe figure 2.3 illustre l'évolution des forces estimées en fonction de la longueur des séquences.

L'une des séquences utilisées dans le projet Golem est longue de 75 bases. Pour cette même longueur, la force d'interaction entre deux brins d'ADN complémentaires est estimée à  $800pN$ . Cette valeur est la première estimation de la force d'interaction entre deux brins d'ADN complémentaires transmise à nos collaborateurs du projet Golem.

### 1.3 Discussion

A partir de cette approche, une première estimation de la force d'interaction entre deux brins d'ADN complémentaires lors d'une hybridation est obtenue. Cette méthode est cependant simplifiée puisqu'elle ne considère pas, par exemple, la rigidité des molécules. En effet, les positions relatives des atomes de chaque brin sont fixes. Le centre de masse de la molécule est donc l'élément central de l'approximation de la fonction d'énergie. En réalité, les brins d'ADN



**Figure 2.3** - Evolution de la force d'interaction estimée par extrapolation de l'énergie d'interaction en fonction de la longueur de séquence dans l'approche individuelle.

sont élastiques, et la prise en compte de cette caractéristique est possible à l'aide de la méthode introduite dans le paragraphe suivant.

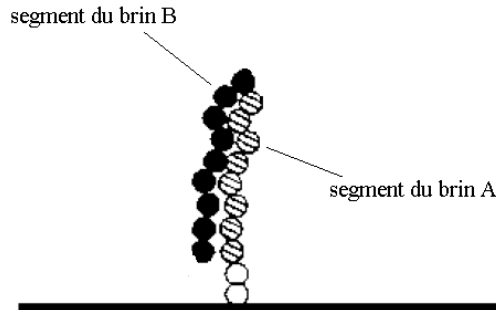
### 1.3.1 L'élasticité de l'ADN

La structure remarquable de la molécule d'ADN est à l'origine de l'une de ses propriétés les plus importantes : l'élasticité. Cette structure est basée sur la complémentarité WC réalisée avec des liaisons hydrogène. La structure chimique (cf. figure 1.11) des bases azotées et celle des chaînes latérales confèrent à la molécule la capacité de construire des liaisons hydrogène entre les bases complémentaires de chaque brin. Pour construire ces liaisons entre une base purique et une base pyrimidique, il est nécessaire que les plans respectifs des bases complémentaires coïncident. Cette structure remarquable est à l'origine de la propriété élastique, propriété primordiale pour le processus d'hybridation. En effet, la molécule explore avec son élasticité l'ensemble des configurations spatiales possibles à la recherche de la configuration la plus stable. Cette caractéristique de l'ADN fait l'objet de plusieurs travaux de recherche où la molécule d'ADN est identifiée à un câble élastique [Poland 66, Neukirch 04]. Dans cette approche individuelle, une étude simplifiée du problème est privilégiée.

Afin de prendre en compte cette propriété élastique dans l'approche individuelle, un moyen est de segmenter chaque brin en une succession de segments. Cette idée s'inspire des travaux de Jayaraman et al. [Jayaraman 06], dans lesquels une modélisation du processus d'hybridation dans les puces à ADN est proposée. En effet, l'ADN peut être considéré comme une chaîne articulée dont les segments sont caractérisés par une longueur  $l_s$ . Cette longueur correspond à la plus petite unité rigide de la chaîne<sup>1</sup>. Dans cette modélisation, l'un des brins en interaction est fixé à un substrat par l'une de ses deux extrémités, l'autre brin est libre en solution (figure

1. Les auteurs [Jayaraman 06] utilisent une segmentation des brins avec des segments d'une longueur de 11 bases. Cette valeur correspond à la longueur de persistance de l'ADN simple brin dans une concentration saline de 1M [Williams 01].

2.4). Pour chacun d'eux, la position du centre de masse de chaque segment change aléatoirement autour de leurs articulations.



**Figure 2.4** - Modélisation par segmentation de l'hybridation dans les puces à ADN dans [Jayaraman 06]. Les brins A et B sont complémentaires.

L'application de cette solution à la méthode d'approximation de l'énergie consisterait à réaliser cette approximation pour chaque segment. Afin d'obtenir la force d'interaction entre les deux brins, l'intégration, dans un modèle physique représentant une chaîne articulée, des forces d'interaction estimées pour chaque couple de segments appartenant respectivement à chacun des brins, rendrait l'approche individuelle plus réaliste. Néanmoins, plusieurs inconvénients persistent dans cette approche, parmi eux, la densité de répartition.

### 1.3.2 La densité de répartition des brins d'ADN

Le velcro à base d'ADN est constitué d'un nombre important de brins fixés à la surface des micro-composants. Des méthodes pour la détermination de la densité de répartition des brins fixés sur les surfaces existent comme par exemple la fluorescence [Nom 07]. Ces méthodes nécessitent des équipements spécialisés et doivent faire l'objet d'une étude à part entière. Dans ce travail de thèse il n'a pas été possible de traiter cette question. En effet, nous ne disposons, ni des équipements, ni des connaissances ou du temps nécessaires pour analyser cette problématique en particulier.

En conséquence, même s'il est possible de résoudre le problème de l'élasticité à l'aide du modèle de la chaîne articulée, l'approche individuelle reste complexe et inappropriée à notre cas d'étude.

## 1.4 Conclusion

L'extrapolation du profil énergétique de l'interaction entre deux brins d'ADN a donné une première estimation de la force d'interaction. Cette approximation est basée sur l'hypothèse d'une molécule rigide et ne prend pas compte du comportement élastique de la molécule. La modélisation en chaîne articulée semble être une solution alternative pour l'étude de ce comportement.

Plusieurs inconvénients rendent, cependant, l'approche individuelle inappropriée et difficile à mettre en œuvre dans le contexte du velcro à base d'ADN. En effet, en plus du paramètre *densité de répartition* inconnu, l'énergie potentielle estimée par MOE dépend du champ de forces utilisé. L'énergie potentielle est estimée sur la base des paramètres numériques spécifiques à un champ de forces choisi au préalable. La valeur de la force déduite n'est donc pas absolue. De plus, les paramètres environnementaux tels que l'hydratation et la présence de cations, ainsi que le comportement populationnel des brins d'ADN sont trop complexes pour être étudié d'un point de vue moléculaire.

L'approche individuelle est donc une approche centrée sur le comportement individuel de chaque brin où le comportement populationnel de l'ensemble des brins en interaction peut être difficilement pris en compte. Il est donc nécessaire d'explorer une approche différente.

## 2 Approche populationnelle

L'approche populationnelle prend en compte à la fois le comportement global du système moléculaire et le comportement individuel de chaque paire de brins. La quantification de la stabilité est réalisée à partir d'outils issus de la thermodynamique. Ces derniers donnent une quantification de la différence d'énergie du système dans sa globalité ainsi que le degré de désordre dû aux différentes configurations adoptées par les paires de brins pendant l'hybridation. Dans le premier paragraphe, il s'agit d'identifier plus précisément le système étudié et les outils nécessaires à la quantification de la stabilité. Parmi les modèles existants, le modèle du plus proche voisin est choisi. Ce modèle et l'ensemble de ses paramètres sont présentés dans le deuxième paragraphe. Enfin, la méthode d'estimation de la force d'interaction à partir de l'énergie d'interaction caractéristique est proposée dans le dernier paragraphe.

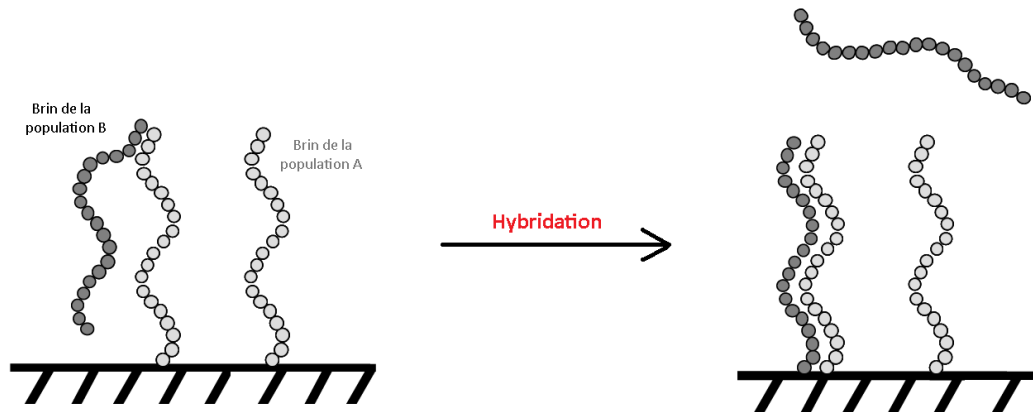
### 2.1 Quantification de la stabilité

Du point de vue de la thermodynamique, le système étudié est défini comme un système ouvert. Deux états sont ainsi étudiés. L'*état initial* correspond à l'état *hybridé* du système moléculaire, et l'*état final* correspond à l'état *dés-hybridé* du système moléculaire (figure 2.1). La transition de l'état initial à l'état final est caractérisée par une différence d'énergie qui témoigne de la capacité du système à rester dans son état assemblé. Elle correspond à la différence d'énergie libre de la liaison globale.

L'énergie libre s'écrit en fonction de deux composantes : la fonction d'enthalpie du système  $\Delta H$ , et l'entropie  $\Delta S$ . La variation d'énergie libre pour un système à pression et à température constante  $T$  est :

$$\Delta G = \Delta H - T\Delta S \quad (2.6)$$

La quantité d'énergie nécessaire à la transition est propre au système moléculaire constitué de deux populations de brins complémentaires. Ce dernier peut être comparé au système moléculaire d'une bio-puce à ADN.



**Figure 2.5** - Principe des puces à ADN. Hybridation entre deux populations de brins complémentaires A et B.

Dans cette technologie, l'hybridation a lieu entre des brins libres en solution et des brins fixés à une surface, généralement en verre, comme illustré sur la figure 2.5. La caractérisation de ce processus est possible à partir de l'estimation de la différence de l'énergie libre de liaison. Plusieurs modèles dédiés à l'estimation de cette fonction d'état en fonction des paramètres intrinsèques et extrinsèques de l'ADN existent. Le modèle du plus proche voisin est l'un des modèles les plus utilisés dans ce domaine [Glazer 06].

## 2.2 Modèle du plus proche voisin

Depuis une dizaine d'années, le développement très rapide de la technique des puces à ADN engendre un intérêt très particulier pour le phénomène de l'hybridation de l'ADN du point de vue de la modélisation. Aussi, plusieurs méthodologies pour l'estimation de l'énergie libre de liaison sont disponibles dans la littérature [Brandsdal 03]. Le modèle le plus utilisé est le modèle du plus proche voisin [SantaLucia 98]. L'algorithme de base et les paramètres de ce modèle sont présentés dans les paragraphes suivants. L'influence des paramètres environnementaux sur  $\Delta G$  est détaillée dans le dernier paragraphe.

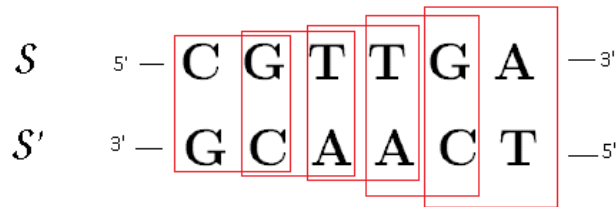
### 2.2.1 Algorithme du modèle du plus proche voisin

Dans cette approche, la variation d'énergie libre  $\Delta G$  est la fonction d'état caractéristique de la stabilité du velcro dont la séquence en bases azotées est caractéristique. L'estimation de cette variation d'énergie nécessite la connaissance de la séquence du velcro. En effet, l'hypothèse centrale de ce modèle est l'estimation de l'énergie libre d'une paire de bases en fonction des énergies de ses deux paires de bases voisines.

Avant d'aborder la méthode de calcul en détail, il est primordial de définir la notion de **couple de paires de bases**. Par exemple, soient deux séquences complémentaires, S orientée 5'-3', et S' orientée 3'-5' comme illustré sur la figure 2.6. Avec une longueur de 6 bases, S

comporte 5 couples de bases : CG , GT, TT, TG, GA. De la même manière S' comporte les 5 couples de bases suivants : GC, CA, AA, AC, CT. En conséquence, la double hélice formée par l'hybridation parfaite de S et S' comporte 5 couples de paires de bases : CG/GC , GT/CA, TT/AA, TG/AC, GA/CT.

Dans l'état initial, une paire de brins est dans l'état hybridé (figure 2.1). Dans le cas d'une hybridation "*perfect match*" (PM), ce modèle suppose que la stabilité d'une paire de bases, constituée de deux nucléotides appartenant respectivement aux deux brins de l'hélice, dépend de l'identité de ces deux nucléotides en particulier, et de leurs voisins respectifs. Ainsi, pour une séquence donnée, il est possible d'imaginer une fenêtre d'un couple de paires de bases glissant le long des deux séquences des brins complémentaires (figure 2.6).



**Figure 2.6** - Algorithme pour le calcul de  $\Delta G$ . Une fenêtre de couple de paires de bases glisse le long de la représentation des deux séquences complémentaires hybridées. Ces séquences présentent 5 couples de paires de bases.

D'après le modèle du plus proche voisin pour les biopolymères, la différence d'énergie libre est donnée par l'expression suivante [SantaLucia 98] :

$$\Delta G = \sum_i \Delta G(i) \quad (2.7)$$

avec  $\Delta G(i)$  l'énergie libre du  $i$ ème couple de paires de bases.

### 2.2.2 Les paramètres intrinsèques

Les paramètres de ce modèle sont estimés dans plusieurs travaux et normalisés dans le papier [SantaLucia 98]. Il s'agit des variations d'entropie  $\Delta S$  et d'enthalpie  $\Delta H$  pour chaque couple de paires de bases possibles. Les valeurs de ces paramètres sont déterminés pour  $T = 37^\circ\text{C}$  et pour une concentration ionique de  $[Na^+] = 1M$ . Ces valeurs sont données dans la figure 2.7. Ainsi, la variation d'énergie libre d'un couple, notée  $\Delta G_{37}(i)$  pour  $T = 37^\circ\text{C}$  et  $[Na^+] = 1M$  est estimée à partir de l'équation 2.8.

La variation d'énergie libre pour une séquence donnée, notée  $\Delta G_{37}$  pour les mêmes valeurs de température et de concentration ionique, est estimée à l'aide de l'équation 2.9.

Sequence	$\Delta H^\circ$ kcal/mol	$\Delta S^\circ$ cal/k-mol
AA/TT	-7.9	-22.2
AT/TA	-7.2	-20.4
TA/AT	-7.2	-21.3
CA/GT	-8.5	-22.7
GT/CA	-8.4	-22.4
CT/GA	-7.8	-21.0
GA/CT	-8.2	-22.2
CG/GC	-10.6	-27.2
GC/CG	-9.8	-24.4
GG/CC	-8.0	-19.9
Init. w/term. G·C	0.1	-2.8
Init. w/term. A·T	2.3	4.1
Symmetry correction	0	-1.4

**Figure 2.7** - Tableau des paramètres standards du modèle du plus voisin pour l'ADN [SantaLucia 98].

$$\Delta G_{37}(i) = \Delta H_{37}(i) - T\Delta S_{37}(i) \quad (2.8)$$

$$\Delta G_{37} = \sum_i \Delta G_{37}(i) \quad (2.9)$$

avec  $\Delta S_{37}(i)$ , et  $\Delta H_{37}(i)$  l'entropie et l'enthalpie des couples (figure 2.7).

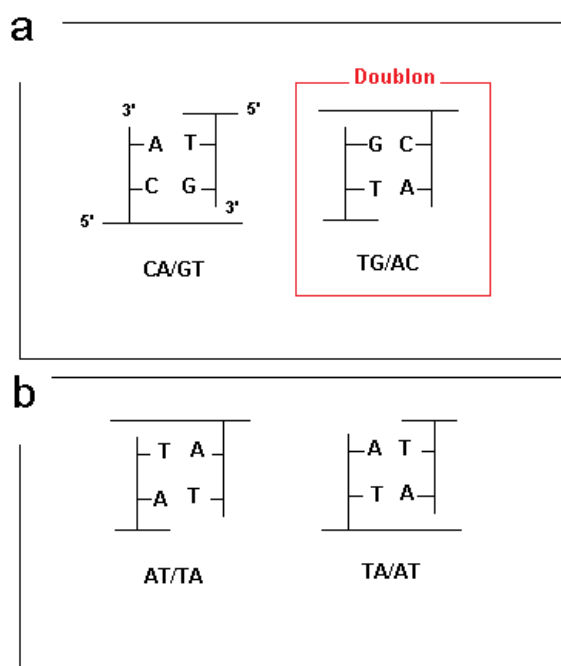
Seize couples peuvent être construits à partir des quatre bases azotées A, T, C et G. Les paramètres  $\Delta S_{37}$  et  $\Delta H_{37}$  sont estimés uniquement pour dix couples (figure 2.7). Cette caractéristique est due à la redondance des couples de paires de bases.

Par exemple, sur la figure 2.8.a, sont illustrés deux couples de paires de bases : CA/GT et TG/AC appartenant respectivement à deux brins du velcro. Ces deux couples sont semblables et représentent la même entité. Pour détecter plus facilement la similitude entre les couples, un moyen est de repérer l'orientation et l'identité des bases fixées au substrat. Si les bases fixées sont identiques pour la même orientation, alors il s'agit d'un même couple. Dans cet exemple, les bases fixées en 5' sont identiques : C et T pour les deux couples. Dans le tableau 2.1, le couple TG/AC est donc un doublon et ses paramètres numériques sont identiques à ceux de CA/GT.

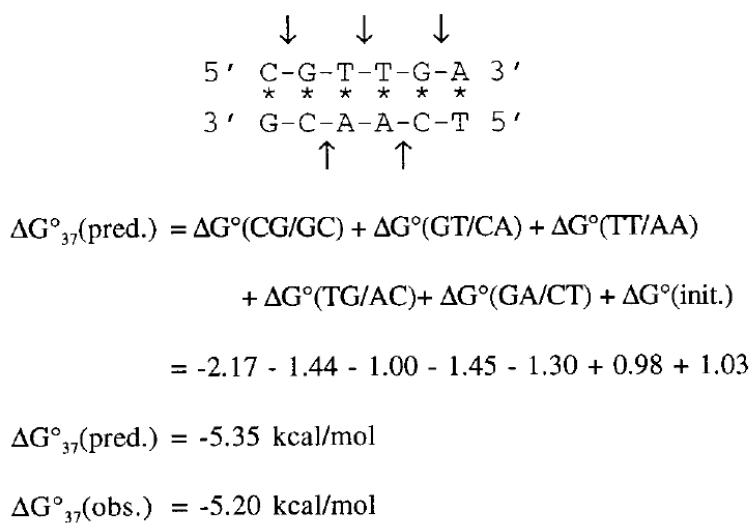
Sur la figure 2.8.b, sont représentés deux autres couples à première vue identiques. Or, le couple TA/AT est différent du couple AT/TA. Dans le couple TA/AT les bases fixées sont T et T en 5' alors que pour le couple AT/TA les bases fixées sont A et A en 5'. Cette différence rend les deux couples différents dans le tableau 2.1.

L'estimation de  $\Delta G_{37}$  pour une séquence donnée est illustrée à titre d'exemple sur la figure 2.9. Les paramètres "Init" et "Symmetry" représentés sur la figure 2.7 et dans l'exemple illustré sur la figure 2.9 sont des paramètres correcteurs. Par souci de simplification, dans cette approche populationnelle, même si ces paramètres affinent l'estimation de  $\Delta G$ , ils ne sont pas utilisés dans la suite de ce document.





**Figure 2.8** - Les couples de paires significatifs. a. les couples sont semblables, un des deux couples est pris en compte dans les paramètres numériques. b. Les deux couples sont différents. Les deux couples sont pris en compte dans les paramètres numériques.



**Figure 2.9** - Exemple d'une estimation de  $\Delta G_{37}$  pour une séquence CGTTGA [SantaLucia 98].

	A	T	C	G
A	<b>AA</b>	<b>AT</b>	AC	AG
T	<b>TA</b>	TT	TC	TG
C	<b>CA</b>	<b>CT</b>	CC	<b>CG</b>
G	<b>GA</b>	<b>GT</b>	<b>GC</b>	<b>GG</b>

**Tableau 2.1** - *Combinatoire des paires de bases. Les paires de bases sont orientées 5'-3'. Le couple de paires de bases est obtenu en complétant le couple par son complémentaire. Exemple : **TA** couple de base, **TA/AT** couple de paires de bases.*

La différence d'énergie libre  $\Delta G_{37}$  pour une séquence en bases azotées peut être estimée pour une température fixe de 37°C et une concentration saline fixe de  $[Na^+] = 1M$ . La généralisation pour différentes températures et différentes concentrations est donnée dans la suite.

### 2.2.3 Les paramètres environnementaux

L'estimation de la différence d'énergie libre pour une concentration ionique en sodium  $[Na^+]$  quelconque est donnée par l'équation 2.10 [SantaLucia 98] :

$$\Delta G_{37}(Na^+) = \Delta G_{37} - 0.114 \frac{N}{2} \ln[Na^+] \quad (2.10)$$

avec  $\Delta G_{37}$  la variation d'énergie calculée pour  $T = 37^\circ C$  et  $[Na^+] = 1M$ .  $N$  est le nombre de phosphates sur un brins de la double hélice et peut être facilement comptabilisé puisqu'il existe un groupement phosphate intercalé entre chaque deux nucléotides d'une séquence ( $\frac{N}{2} = n - 1$ , avec  $n$  la longueur de la séquence). L'influence de la concentration ionique est ainsi liée à la longueur de la séquence.

Afin d'obtenir une expression générale de  $\Delta G$  en fonction des deux paramètres environnementaux  $T$  et  $[Na^+]$ , l'équation 2.11 est déduite à partir de l'équation 2.10.

$$\Delta G(T, Na^+) = \sum_i (\Delta H_{37}(i) - T \Delta S_{37}(i)) - 0.114 \frac{N}{2} \ln[Na^+] \quad (2.11)$$

avec  $i$  l'indice du ième couple de bases sur la séquence. Cette équation implique que les deux paramètres  $T$  et  $[Na^+]$  sont indépendants. L'étude de l'influence de la température et de la concentration ionique est donc possible avec cette seule équation (2.11). Les figures 2.10 et 2.11 illustrent respectivement l'influence de la température et de la concentration ionique sur la stabilité quantifiée pour une séquence de 120 bases.

D'une part, d'après la figure 2.10, lorsque  $T$  croît, alors  $\Delta G$  croît aussi. Par conséquent, plus la température  $T$  est faible, plus la stabilité est grande. D'autre part, d'après la figure 2.11, lorsque  $[Na^+]$  croît, alors  $\Delta G$  décroît. Par conséquent, plus la concentration  $[Na^+]$  est forte, plus la stabilité est grande.

Afin d'aboutir à une équation prenant compte à la fois des paramètres intrinsèques et extrinsèques, une dernière simplification est possible en remplaçant  $N$  par  $n - 1$  dans l'équation

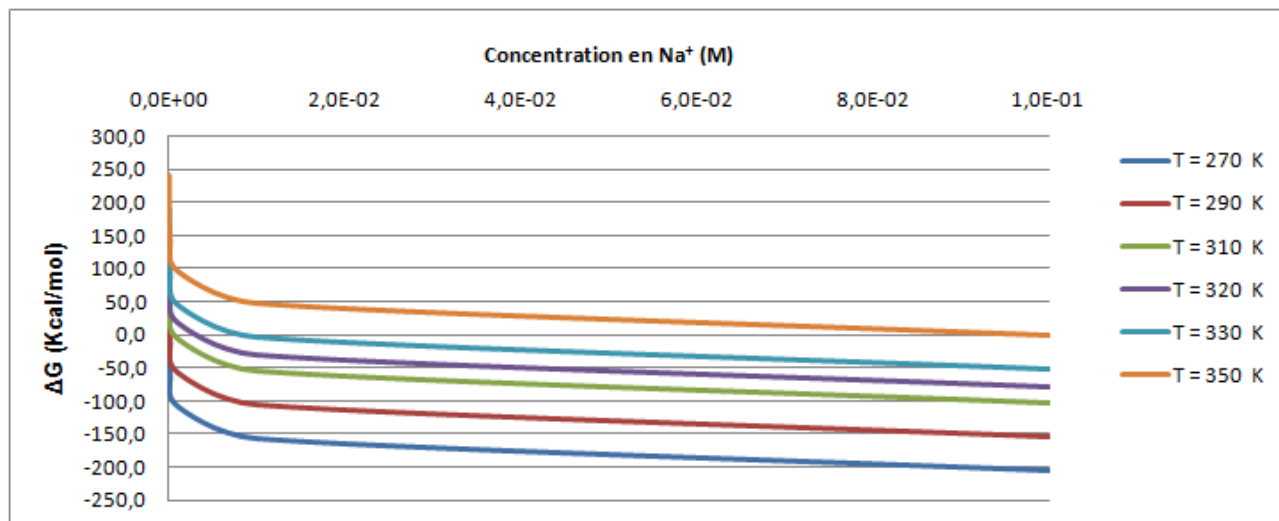


Figure 2.10 - Evolution de  $\Delta G$  pour une séquence de 120 bases en fonction de  $[Na^+]$  pour plusieurs valeurs fixes de  $T$ .

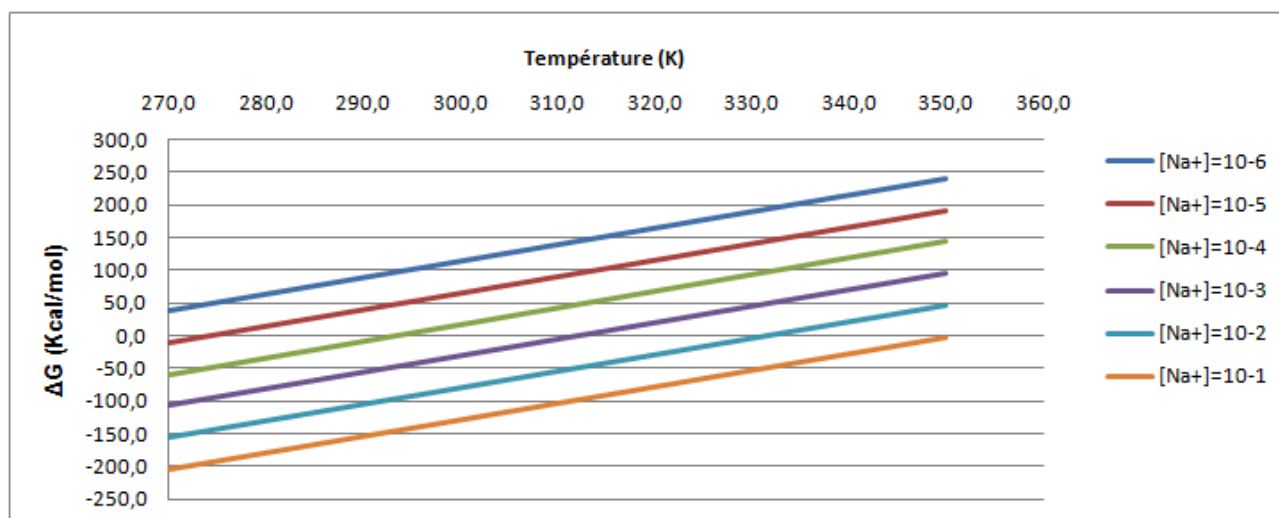


Figure 2.11 - Evolution de  $\Delta G$  pour une séquence de 120 bases en fonction de  $T$  pour plusieurs valeurs fixes de  $[Na^+]$ .

2.11 :

$$\Delta G(T, Na^+) = \sum_i (\Delta H_{37}(i) - T\Delta S_{37}(i)) - 0.114(n-1)\ln[Na^+] \quad (2.12)$$

Cette équation caractérise la transition entre l'état initial et l'état final. Convertie en force, cette valeur énergétique correspond à la force d'interaction mise en jeu lors de l'hybridation.

### 2.3 Estimation de la force d'interaction à partir de $\Delta G$

Pour corrélérer les valeurs énergétiques  $\Delta G$  calculées à partir du modèle du plus proche voisin avec les expériences, l'estimation de la force d'interaction  $F$  doit être effectuée. En effet, dans le cadre du projet Golem, une quantification de l'interaction en terme de force, même approximative, entre deux populations de brins d'ADN est pertinente pour les partenaires du projet.

La conversion de  $\Delta G$  en  $F$  peut s'exprimer de façon approximée selon l'équation 2.13.

$$F = -\frac{d\Delta G}{dx} \quad (2.13)$$

$F$  est la somme des forces instantanées extérieures exercées sur le système, et  $x$  le déplacement selon l'axe de la double hélice (figure 2.2). Puisqu'il s'agit d'une variation d'énergie et non d'une différentielle, l'estimation de la force donne une valeur moyenne et non une valeur instantanée. La force d'interaction moyenne  $\bar{F}$  est obtenue par :

$$\bar{F} = -\frac{\Delta G}{\Delta x} \quad (2.14)$$

Pour une hybridation de type "*Perfect Match*" avec un appariement de toutes les bases le long de la double hélice, le déplacement  $\Delta x$  peut être approximé par la longueur de la séquence. En effet, la différence d'énergie calculée pour une hybridation de ce type correspond à la séparation de  $n$  bases azotées (avec  $n$  le nombre de bases de la séquence).  $\Delta x$  peut donc être approximé selon l'équation 2.15.

$$\Delta x = n d_{base} \quad (2.15)$$

avec  $d_{base}$  la distance entre deux bases consécutives d'un brin. Par exemple,  $d_{base} = 0.34nm$  pour une concentration saline de  $1M$ .

Pour la séquence  $S_1$ , une des séquences proposées dans le projet Golem, pour une hybridation PM,  $\Delta G = -164.42Kcal/mol$  et  $\bar{F}$  est estimée à  $44pN$ .

Dans le cas d'une hybridation de type "*mis-match*"<sup>2</sup> (MM), les structures tridimensionnelles complexes empêchent une estimation simple du déplacement. Par simplification,  $\Delta x$  est assimilé à la longueur des segments hybridés de type PM et la valeur de  $\Delta G$  est encadrée par deux valeurs caractéristiques :  $\Delta G_{min}$  et  $\Delta G_{max}$ . Ces deux énergies sont les variations d'énergie libre de la séquence minimale  $((AT)_{n/2})$ , et de la séquence maximale  $((GC)_{n/2})$  de longueur  $n$ . Ainsi, le

---

2. Chapitre 1, paragraphe 3.3.2.

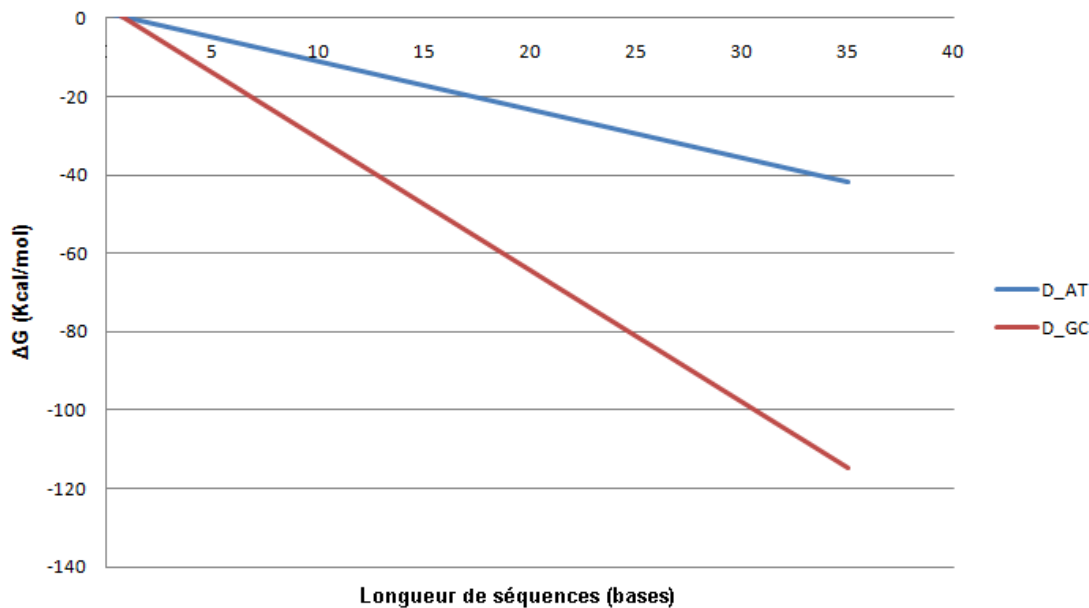
couple (A,T) dont l'énergie est minimale en valeur absolue, et le couple (G,C) dont l'énergie est maximale (figure 2.7), définissent, d'une part une séquence construite d'une suite de (A,T) avec une énergie plus faible qu'une séquence construite aléatoirement pour la même longueur. D'autre part, de façon similaire, une séquence construite d'une suite de (G,C) a une énergie plus élevée qu'une séquence construite aléatoirement pour la même longueur.

Les énergies des séquences minimale ( $(AT)_{n/2}$ ) et maximale ( $(GC)_{n/2}$ ), respectivement  $\Delta G_{min}$  et  $\Delta G_{max}$  sont calculées pour des valeurs différentes de  $n$  à partir l'équation 2.12. Les deux équations 2.16 et 2.17, obtenues par régression linéaire, expriment ces énergies en fonction de  $n$  pour  $T = 37^\circ\text{C}$  et une concentration saline à  $1M$ .

$$\Delta G_{min} = 1.23n - 1.15 \quad (2.16)$$

$$\Delta G_{max} = 3.37n - 3.14 \quad (2.17)$$

Les représentations graphiques des énergies d'interaction en fonction de la longueur de séquence(en bases), pour ces deux séquences sont les droites  $D_{AT}$  et  $D_{GC}$  (figure 3.19).



**Figure 2.12** - Energie d'interaction  $\Delta G$  en fonction de la longueur de séquence pour les séquences minimale (droite  $D_{AT}$ ) et maximale (droite  $D_{GC}$ ).

En raison de la linéarité de ces deux équations, leur dérivation donne deux valeurs constantes pour les forces d'interaction. Ces deux valeurs sont respectivement, la valeur moyenne minimale et la valeur moyenne maximale de la force d'interaction entre deux brins d'ADN pour des configurations de type PM. Finalement, la force d'interaction est dans l'intervalle  $[25; 69]pN$ .

## 2.4 Récapitulatif

Dans cette approche populationnelle, la prise en compte du comportement individuel et populationnel de l'ADN dans la quantification de la stabilité est possible. Cette quantification

est basée sur une estimation énergétique dont les deux composantes principales sont l'enthalpie et l'entropie. Les paramètres de contrôle sont facilement accessibles et la conversion de l'énergie en force est réalisée. Cette conversion donne une valeur moyenne de la force d'interaction pour des configurations de type PM. Pour une séquence de longueur 75 bases, dans cette approche, la force moyenne est estimée à  $44pN$ .

La prise en compte des configurations de type MM est complexe et impliquerait une estimation complexe du déplacement propre à la force d'interaction. Pour répondre à cette difficulté, un encadrement de la force d'interaction est donné à travers les définitions des deux séquences caractéristiques, la séquence minimale et la séquence maximale.

### 3 Conclusion

Dans ce chapitre, deux approches sont proposées pour approximer les forces mises en jeu dans l'interaction entre les objets. La première s'intéresse au comportement d'un couple de brins d'ADN au niveau individuel. Elle utilise des outils de la modélisation moléculaires. La seconde considère un niveau global à partir d'une approche thermodynamique. Ces deux outils sont appliqués sur des séquences utilisées dans le cadre du projet Golem.

La force d'interaction pour une séquence de longueur de 75 bases est estimée à  $800pN$  à partir de l'approche individuelle. Pour plusieurs raisons, l'accessibilité des paramètres de contrôle et l'élasticité de la molécule, cette approche est trop simplifiée.

L'approche populationnelle donne un accès direct aux paramètres intrinsèques et extrinsèques de l'ADN. Dans cette approche, la force d'interaction est estimée approximativement à  $44pN$  (pour l'une des séquences utilisées dans Golem, la séquence  $S_1$  de longueur 75 bases). Cette estimation est valable pour une hybridation complète des bases le long de la double hélice formée par les deux brins complémentaires. Cette valeur est retenue dans une première approximation et est comparée à des mesures expérimentales dans la suite de ce document.



# CARACTÉRISATION EXPÉRIMENTALE DE LA STABILITÉ

---

Ce chapitre s'intéresse à la stabilité d'un velcro composé de deux populations de brins d'ADN complémentaires. Des expériences sont réalisées à l'aide d'un AFM commercial au sein du laboratoire national de Londres (NPL). L'objectif principal de ce travail expérimental est de quantifier la stabilité du velcro et d'étudier ses paramètres de contrôle. Pour atteindre ce dessein, une méthode d'analyse des données couplée à une approche énergétique est proposée.

La première partie de ce chapitre, est une introduction aux connaissances de base nécessaires à la compréhension des données expérimentales obtenues. Ainsi, l'utilisation de l'AFM pour la mesure des interactions biomoléculaires est introduite. Les questions techniques de calibration et de détection optique ne sont pas abordées, puisque l'intérêt premier de ce travail est l'exploitation des données expérimentales. Enfin, les différents paramètres et variables de caractérisation des interactions sont identifiés à travers un travail bibliographique.

La deuxième partie de ce chapitre est dédiée aux expériences réalisées en collaboration avec le NPL et l'université d'Edinburgh. Une analyse statistique pour l'identification des données significatives est proposée dans la troisième partie. Enfin, dans la quatrième et dernière partie, une nouvelle méthode d'analyse énergétique des données est introduite. Cette analyse a pour objectif de se doter d'un outil d'évaluation des forces d'interaction en tenant compte des spécifications des brins d'ADN.



# 1 L'AFM, les variables et les paramètres de mesure

## 1.1 Principe général de l'AFM

Un microscope à force atomique est constitué d'une poutre de dimension longitudinale d'une centaine de micromètres, munie d'une pointe en interaction avec une surface. Un détecteur mesure le déplacement de la poutre et une boucle de contrôle régule les paramètres comme la distance pointe-surface ou la force appliquée [Butt 05](figure 3.1).

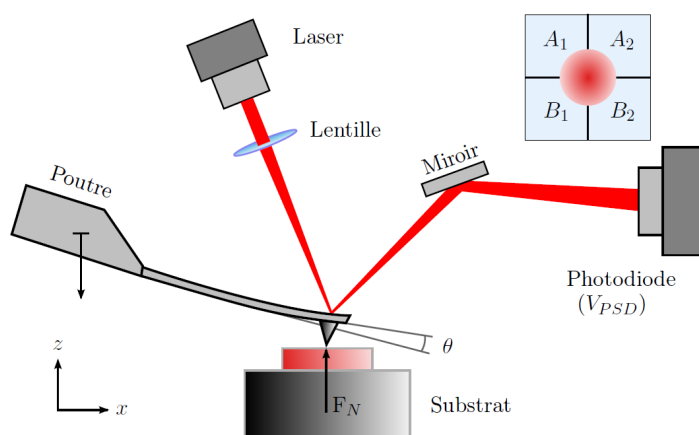


Figure 3.1 - Principe général de l'AFM.

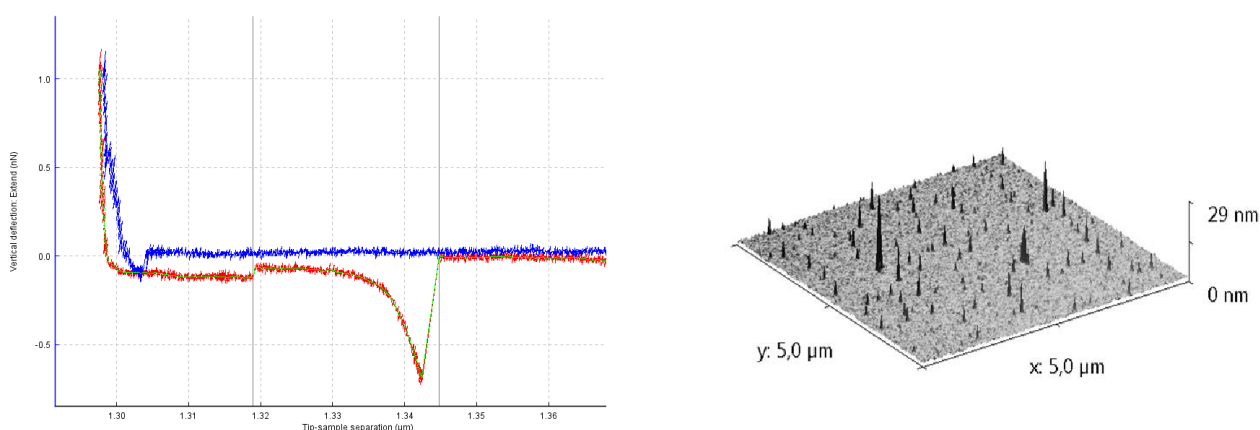


Figure 3.2 - Utilisation de l'AFM pour la caractérisation des surfaces. À gauche, une courbe  $A/R$  pour un substrat en silicium dans un milieu liquide. La force d'interaction est représentée en fonction du déplacement de la poutre. La phase approche est représentée en bleu et la phase retrait en rouge. À droite, la topographie d'une surface en silicium. Ces images (courbe et topographie) ont été obtenues en collaboration avec le laboratoire AMIR de l'université de Oldenburg.

Inventé par Gerd K. Binnig, prix Nobel de physique [Binnig 86], le principe de l'AFM est de balayer la surface par la pointe de la poutre et de reproduire la topographie de la surface (figure 3.2). Celle-ci est obtenue en mesurant et représentant la déflexion de la poutre en fonction de sa position.

L'AFM peut aussi être utilisé pour mesurer des forces d'interaction nécessaires pour la détermination du comportement physique d'objets à des échelles en dessous du micromètre. L'exploitation de ces mesures et l'analyse des phénomènes physiques à l'échelle microscopique sont des outils majeurs pour la manipulation ou l'assemblage de micro-objets.

Dans ce contexte expérimental, l'AFM est détourné de sa fonction originale pour mesurer les forces d'interaction entre la pointe de la poutre et la surface du micro-objet. Connaissant la raideur de la poutre, notée  $k$ , les efforts peuvent être déduits. La figure 3.2 illustre une mesure AFM entre la pointe d'une poutre et un substrat en silicium. La courbe est appelée "*courbe approche/retrait*" (A/R) en raison du processus d'approche et de retrait de la poutre (ou du substrat) lors de la mesure.

La raideur  $k$  caractérise à la fois les dimensions, la forme géométrique et la nature du matériau utilisé dans la fabrication de la poutre. En raison des difficultés rencontrées par les constructeurs pour contrôler la largeur des poutres AFM durant les processus de microfabrication, cette constante fournie par les constructeurs est souvent incorrecte [Giroto 07]. La calibration de la poutre est donc une étape importante lors d'une expérience de mesure AFM. Plusieurs techniques de calibration existent [Xie 08], et sont généralement disponibles à travers les logiciels d'utilisation pour les AFM commerciaux, tels que XEI et JPK.

## 1.2 L'AFM pour les interactions biomoléculaires

Depuis une dizaine d'années, cette technologie est utilisée par une communauté d'expérimentateurs biophysiciens pour la caractérisation des forces d'interaction inter et intramoléculaires. Les avantages de l'AFM pour ce type d'utilisation sont la possibilité de réaliser des mesures en milieu aqueux [Marti 87], et de modifier le préhenseur chimiquement pour mesurer des forces spécifiques [Butt 05]. Ainsi, étant donné un complexe A-B, les molécules de type A collées sur la pointe de la poutre, et les molécules de type B collées sur la surface, il s'agit d'utiliser le principe d'A/R de l'AFM afin d'étudier les liaisons biochimiques des complexes dans leur environnement originel. La figure 3.3 illustre ce processus de mesure pour le complexe protéique CaM-skMLCK [Junker 09].

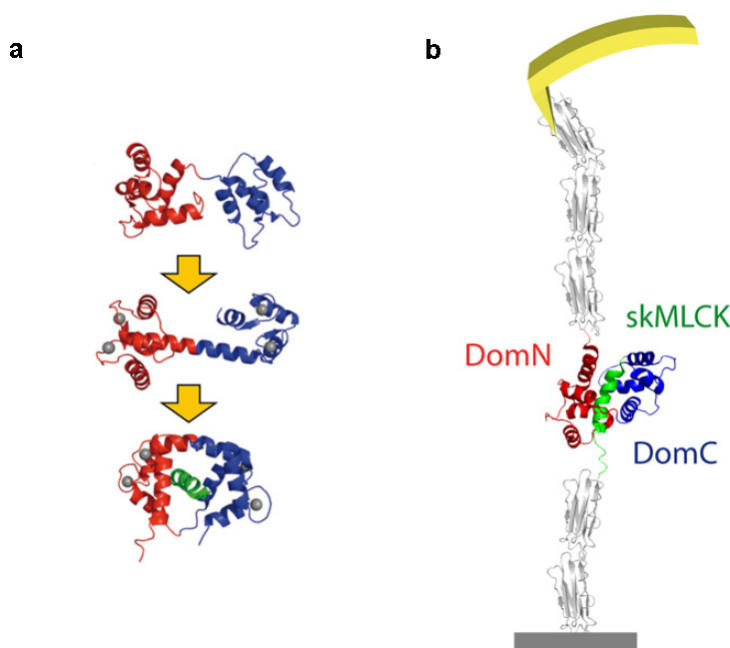
La fixation des biomolécules sur les différentes surfaces est un processus de fonctionnalisation qui répond aux mêmes principes de fonctionnalisation des lames en verre dans les bio-puces à ADN par exemple [Leunissen 10].

Les premières mesures AFM de forces d'interaction biomoléculaires datent de 1994 [Lee 94, Florin 94]. C'est d'abord les travaux sur le complexe Biotin-avidin [Florin 94] qui ont suscité le plus d'intérêt et ont contribué à l'avancée dans les méthodes d'interprétation des mesures d'interactions biochimiques par AFM [Evans 99]. Les travaux sur l'ADN [Lee 94] ont été ralentis en raison de la difficulté d'interprétation face à la *structure spatiotemporelle complexe* de la molécule. En effet, au cours de l'hybridation, la formation de structure biomoléculaires

complexes et inattendues appelées "*mis-matches*" (MM) est en grande partie responsable de ce comportement complexe.

Afin d'orienter nos expériences de mesures AFM, il est important de préciser les méthodes d'interprétation, le type d'information exploitable à partir de ce type d'expérience, et les paramètres d'influence.

Dans le paragraphe suivant, les variables et paramètres expérimentaux caractéristiques des mesures des forces d'interaction entre des brins d'ADN dans l'axe de la double hélice sont introduites à travers une analyse bibliographique.



**Figure 3.3** - Principe d'une mesure des forces d'interaction biomoléculaires par l'AFM. a. Formation du complexe ligand-récepteur Calmoduline-skMLCK (CaM-skMLCK). Cette protéine est composée de deux domaines DomN (rouge) et DomC (bleu), l'activation par des ions  $Ca^{2+}$  (en gris) permet la formation du complexe entre skMLCK (vert) et CaM. b. La poutre est fonctionnalisée avec CaM et le substrat avec le ligand. En gris, des molécules intermédiaires de fonctionnalisation (domaines de filamine) [Junker 09].

### 1.3 Les variables de mesure et les paramètres expérimentaux

Une liaison biomoléculaire est caractérisée par son intensité, aussi appelée énergie d'activation. Celle-ci est la quantité d'énergie nécessaire pour initier ou rompre une liaison. C'est aussi le travail fourni par une force extérieure en déplacement dans le sens physique de la liaison. Cette force est appelée **force de rupture**, notée ( $F_r$ ) et peut se définir comme la force nécessaire et suffisante pour rompre la barrière énergétique.

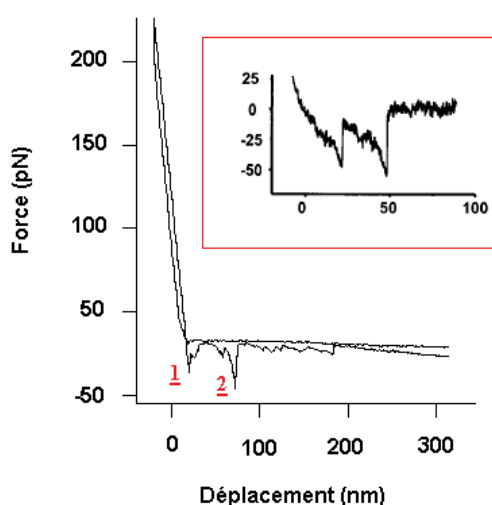
Cette force de rupture peut être obtenue à partir d'une courbe A/R d'une mesure AFM. S'agissant particulièrement de la molécule d'ADN, l'obtention de cette force de rupture n'est pas simple. Le premier paragraphe de cette partie est dédié à la définition de cette variable,

à travers un travail bibliographique, ainsi qu'à l'identification des méthodes conventionnelles pour son obtention.

Un paramètre expérimental influence directement la force de rupture. Il s'agit du **taux de chargement** noté  $T_c$ . Ce paramètre traduit l'impact de l'énergie chargée par la poutre sur le système moléculaire, et sa prise en compte dans nos expériences est primordiale. La définition du taux de chargement est proposée dans le deuxième paragraphe de cette partie. Il s'agit de montrer son importance et la méthode conventionnelle utilisée pour étudier son influence.

### 1.3.1 La force de rupture

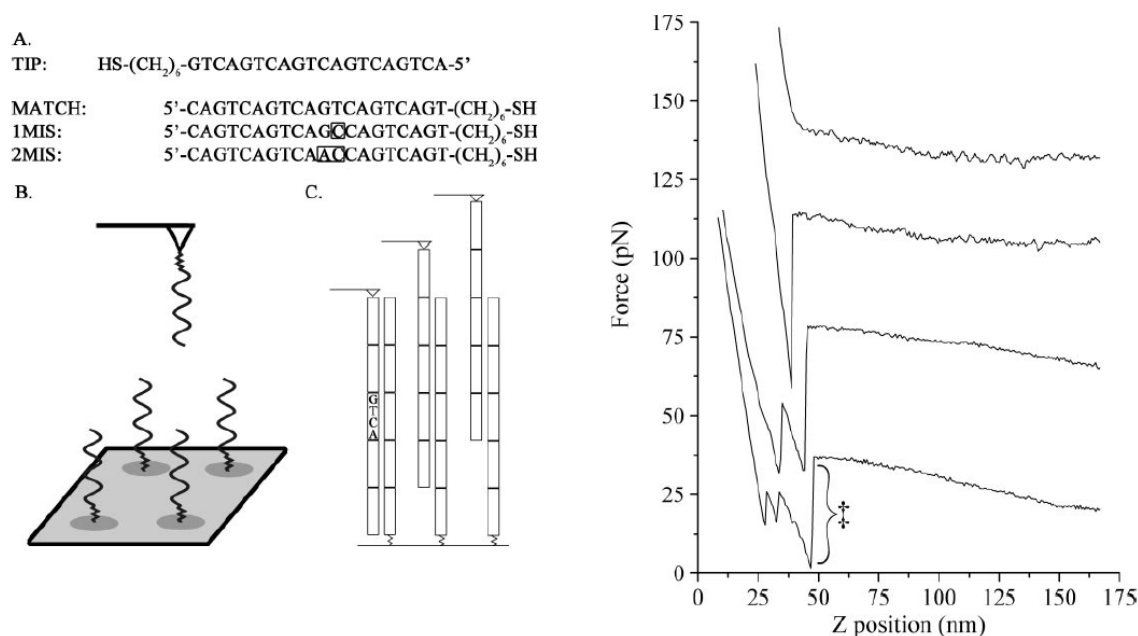
Pendant une expérience de mesure AFM pour la caractérisation d'une interaction entre deux surfaces fonctionnalisées, la force mesurée quantifie l'intensité de l'interaction entre les molécules fixées sur chacune des surfaces. Dans le cas précis d'une fonctionnalisation avec des brins d'ADN, sur la courbe A/R d'une mesure des forces d'interaction en jeu dans l'axe de la double hélice, plusieurs *sauts* correspondant à des *événements de rupture* différents peuvent apparaître. La représentation graphique de cette mesure prend la forme d'une courbe A/R de la même allure que la courbe représentée sur la figure 3.4. Ces sauts peuvent être dus à des couples de brins hybridés différents, ou à des interactions entre certains segments d'un même couple de brins, ou encore à l'élasticité de la molécule. Il est donc difficile d'interpréter une courbe A/R lorsqu'il s'agit d'une mesure sur l'ADN. Dans la plupart des travaux rencontrés dans la littérature, l'hypothèse d'une interaction individuelle entre deux brins d'ADN conduit les auteurs à interpréter les données récoltées dans le sens d'une molécule unique [Lee 94, Florin 94, Strunz 99]. Un état de l'art non exhaustif des travaux majeurs est résumé dans les figures 3.6 et 3.7. Dans l'ensemble de ces travaux, cette hypothèse rend l'interprétation des données plus simple. Dans notre analyse, et dans la continuité des travaux de nos prédécesseurs, cette hypothèse est maintenue.



**Figure 3.4** - Courbe A/R avec plusieurs événements de rupture obtenues en collaboration avec le NPL pour un substrat en silicium fonctionnalisé avec la séquence S1. Les sauts 1 et 2 sont des événements de rupture. Dans l'encadré rouge, La courbe publiée par [Strunz 99].

Les premiers travaux de mesure des forces d'interaction entre des brins d'ADN dans l'axe

de la double hélice sont réalisés sur une séquence d'une longueur de 20 bases<sup>1</sup> par Lee et al. [Lee 94]. Les forces mesurées sont comprises entre 1520 et 830 pN. Une seconde étude réalisée sur la même séquence par Sattin et al. [Sattin 04] révèle des mesures très différentes avec des valeurs comprises entre 10 et 80 pN. Les deux équipes montrent néanmoins une grande variabilité des mesures. Les analyses statistiques proposées dans les deux études tentent de mettre en relation la variabilité observée avec les différentes configurations adoptées par les brins d'ADN lors de l'hybridation. Lee et al. [Lee 94] présentent leurs mesures AFM sous la forme de trois classes statistiques en correspondance avec trois configurations possibles<sup>2</sup>. Les classes statistiques représentent les valeurs les plus probables pour la force de rupture. Sattin et al. [Sattin 04] montrent que le nombre de classes statistiques augmente avec le nombre de répétitions des mesures. Ainsi, les auteurs répètent les mesures avec la même poutre 1500 fois sur différents plots fonctionnalisés avec des séquences différentes (figure 3.5). Ce protocole rend alors possible la comparaison entre les mesures des différentes interactions. Les auteurs montrent ainsi la nécessité de disposer d'un grand nombre de mesures pour les analyses statistiques de la variabilité.



**Figure 3.5** - Protocole expérimental de Sattin et al. [Sattin 04] pour la mesure des forces d'interaction entre différents brins d'ADN dans l'axe de la double hélice. 4 séquences sont utilisées. Les deux séquences TIP et MATCH sont complémentaires. La séquence MIS1 est identique à MATCH avec la modification du dixième nucléotide. La séquence MIS2 est identique à MIS1 avec la modification du onzième nucléotide. L'introduction de ces modifications conduit les auteurs à un premier type de mesures spécifiques, entre les séquences [TIP et MATCH], et à deux types de mesures non spécifiques, entre les séquences [TIP et MIS1] et [TIP et MIS2].

Deux autres travaux majeurs sont retenus, ceux de Strunz et al. [Strunz 99] et Morfill et al. [Morfill 07]. Ces deux équipes réalisent des mesures AFM sur des brins d'ADN avec des séquences différentes mais une même longueur de 20 bases. Ils obtiennent des résultats du même ordre de grandeur que Sattin et al. [Sattin 04]. Les forces de rupture sont comprises entre 20 et 80 pN. Dans ces travaux, la variabilité observée est mise en relation avec les propriétés structurales des brins lors de l'hybridation, notamment l'élasticité et la transition B-S de la

1. La séquence en bases azotées est ((CAGT)<sub>5</sub>)

2. Ces configurations correspondent à 20, 16 et 12 bases azotées hybridées consécutivement.

double hélice. Ainsi, une molécule d'ADN, longue de plusieurs milliers de bases azotées, peut s'étendre jusqu'à 1.7 fois sa propre longueur et peut rester stable pour une force inférieure à  $65pN$  [Morfill 07]. Pour une force entre 65 et  $130pN$  la double hélice se trouve sous la forme S. Une transition B-S est alors observée sur un palier d'une longueur de  $6nm$ . Lorsque la séquence est plus courte, 20 bases, la transition n'apparaît pas. L'observation de cette transition est donc importante puisque la séparation des deux brins d'une configuration hybridée se passe pendant ou après cette transition.

Les résultats obtenus par Strunz et al. [Strunz 99], Morfill et al. [Morfill 07] et Sattin et al. [Sattin 04] sont du même ordre de grandeur à la différence des travaux de Lee et al. [Lee 94]. Ces trois premiers travaux ont en commun un paramètre expérimental du même ordre de grandeur. Ce paramètre est le taux de chargement, il pourrait expliquer cette homogénéité. Dans le paragraphe suivant, l'influence de ce paramètre expérimental est introduite.

Publication	Conditions expérimentales										
	$T_c$ N/s	$v$ $\mu m/s$	$k$ N/m	Types de surface		Protocole					
				Substrat	Poutre	Solution NaCl M	Linker	Séquence		Mode de mesure	
								séquence 5'-3'	Longueur bases	Temps d'attente s	Nombre de répétitions
Lee et al. 1994	?	$[10^{-4}; 10^{-3}]$	?	verre	verre	0.1	Thiol	(ACTG)5	20	0	?
Strunz et al. 1999	$4 \cdot 10^{-9}$	$10^{-1}$	8 poutres $[0.012; 0.017]$	verre	Nitru de silicium	PH=7.3 ?	PEG	a b c d	10 20 30	0	500 par mesure
Sattin et al. 2004	?	$25 \cdot 10^{-3}$	? Une seule poutre pour toutes les mesures	Or	Or	PH=7.4 ?	Thiol	Match TIP MIS1 MIS2	20	0	1500 par mesure
Morfill et al. 2007	$1.8 \cdot 10^{-9}$	1	$[0.007; 0.020]$ $[0.030; 0.040]$	verre	verre	$150 \cdot 10^{-3}$	PEG	A	20	0	350
	B							30	860		

**Figure 3.6** - Synthèse des mesures AFM des interactions entre des brins d'ADN dans l'axe de la double hélice : conditions expérimentales. Liner : type de molécule utilisé pour la fonctionnalisation.

### 1.3.2 Taux de chargement

L'énergie apportée à un système biomoléculaire lors d'un processus A/R d'une mesure AFM, influence directement la réponse physique du système étudié. Cette énergie s'exprime en fonction de la vitesse d'A/R notée  $v$ , et de la raideur de la poutre  $k$ . Le produit de ces deux paramètres est appelé taux de chargement, noté  $T_c$ . La force extérieure exercée sur le système moléculaire en fonction du temps peut être considérée à travers ce paramètre.

La dépendance de la force de rupture au logarithme du taux de chargement est proposée

Publication	Fr pN	Méthode d'analyse et interprétation	Dépendance à Tc
Lee et al. 1994	[ 830 ; 1520 ]	<b>Statistique :</b> Correspondance entre 3 classes et 3 configurations stables.	
Strunz et al. 1999	[ 20 ; 50 ]	<b>Fr dépend :</b> <ul style="list-style-type: none"> <li>• <math>v</math></li> <li>• Taux de chargement</li> <li>• Séquence</li> </ul> Mise en équation de $dr$ en fonction du nombre de bases $n$ .	<b>x</b>
Sattin et al. 2004	[ 10 ; 80 ]	<b>Statistique :</b> Contribution de : <ul style="list-style-type: none"> <li>• Liaisons WC</li> <li>• Empilement des bases</li> </ul> <b>Etude comparative avec Lee et al. 1994</b>	
Morfill et al. 2007	[20 ; 65[ Fr le plus probable : 54  [50 ; 80 ] Fr le plus probable : 65	<b>Etude de la transition B-S pour les oligos courts.</b> <ul style="list-style-type: none"> <li>• 6 nm : plateau de la transition B-S pour la séquence B.</li> <li>• La transition n'est pas observable sur la séquence A : le taux de chargement est insuffisant.</li> <li>• Dépendance de <b>Fr</b> et <b>dr</b> :               <ul style="list-style-type: none"> <li>○ Taux de chargement</li> <li>○ Séquence</li> </ul> </li> <li>• Utilisation de Bell-Evans pour l'analyse et l'estimation des probabilités</li> </ul>	<b>x</b>

**Figure 3.7** - Synthèse des mesures AFM des interactions entre des brins d'ADN dans l'axe de la double hélice : résultats.

dans les travaux de Evans et al. [Evans 97] à travers la théorie de la spectroscopie à force dynamique (DFS), et prouvée expérimentalement pour plusieurs complexes protéiques [Janovjak 05, Odrowaz 06, Liu 08], ainsi que pour l'ADN [Strunz 99, Morfill 07]. La théorie de Kramers [Kramers 40] et les travaux de Bell [Bell 78] sont à l'origine de l'étude de la dépendance des interactions biomoléculaires à la force exercée par la poutre en fonction du temps dans la théorie de la DFS [Evans 97]. Dans cette théorie, une liaison biomoléculaire est caractérisée par un *temps de vie* (ou de relaxation). Cette notion est introduite par Bell en 1978 dans le cadre de la caractérisation de la séparation mécanique de deux cellules induite par la séparation des complexes protéiques membranaires. Ce temps varie en fonction de la force appliquée et du temps d'application de cette force. Evans et al. enrichissent le modèle de Bell de façon à tenir compte de la variabilité de la force appliquée au système moléculaire due à l'environnement liquide.

Concernant l'ADN, les mesures de Strunz et al. [Strunz 99], de Sattin et al. [Sattin 04] et de Morfill et al. [Morfill 07] sont réalisées avec des taux de chargement du même ordre de grandeur ( $10^{-9}N/s$ ). Ainsi, les trois équipes obtiennent des forces de rupture du même ordre de grandeur. En conséquence, la comparaison de mesures AFM nécessite de maintenir ce paramètre constant. La considération du paramètre taux de chargement est donc primordiale dans l'analyse des mesures des forces de rupture propre à l'ADN.

### 1.3.3 Conclusion

La caractérisation de la stabilité d'un velcro à base d'ADN à l'aide d'un AFM consiste à mesurer les forces de rupture caractéristiques des événements de rupture. La variabilité dans les mesures est montrée par plusieurs travaux expérimentaux [Lee 94, Strunz 99, Sattin 04, Morfill 07]. Un certain nombre de répétitions est ainsi nécessaire pour la réalisation de traitements statistiques.

La comparaison des données impose des paramètres expérimentaux constants. En effet, la force de rupture dépend non seulement de la séquence en bases azotées étudiée et de sa longueur, mais aussi d'un paramètre expérimental important, le taux de chargement. Par ailleurs, la caractérisation de la force de rupture pour des valeurs de taux de chargement différentes apporterait des connaissances majeures pour notre problématique de l'assemblage microscopique, notamment en ce qui concerne la vitesse de guidage des micro-composants lors de l'assemblage<sup>3</sup>.

Enfin, la structure élastique et complexe de l'ADN, et la correspondance entre le nombre de classes statistiques expérimentales et le nombre de configurations possibles pour un couple de brins lors d'une hybridation, semblent être à l'origine de la variabilité observée. Néanmoins, le lien entre ces éléments n'est pas clairement identifié dans la littérature.

Dans la partie suivante, les protocoles expérimentaux utilisés dans nos expériences sont brièvement exposés. Il s'agit dans ces expériences de caractériser en particulier la séquence en bases azotées proposée dans le projet Golem, pour plusieurs valeurs de paramètres expérimentaux, notamment pour le taux de chargement.

## 2 Les protocoles expérimentaux pour la quantification de la stabilité

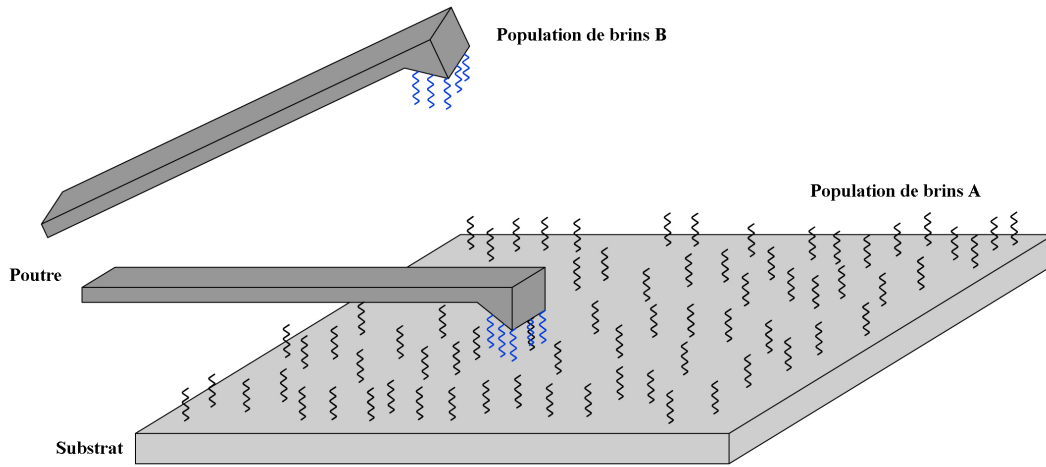
Un ensemble d'expériences de mesures AFM a été réalisé pour la caractérisation de la stabilité de la séquence  $S_1$  proposée dans le projet Golem (ANNEXE A). Ainsi, la poutre et le substrat sont fonctionnalisés avec deux populations de brins complémentaires de séquence  $S_1$  et  $S_2$  (figure 3.8). Dans chaque expérience, un paramètre expérimental varie et le processus d'A/R est successivement répété 256 fois. Durant ces expériences d'A/R, la force en fonction de la distance est enregistrée à des taux de chargements connus. La dissociation des deux surfaces pendant la phase retrait est induite par la séparation des brins des structures hybridées formées durant la phase approche de l'expérience.

Deux protocoles de fonctionnalisation sont utilisés. Le premier protocole consiste à fixer les brins d'ADN sur les surfaces (ANNEXE B). Le second protocole (ANNEXE C) est identique au premier avec une étape supplémentaire qui consiste à fixer des molécules d'éthylamine ( $\text{CH}_3\text{-CH}_2\text{-NH}_2$ ) pour bloquer les surfaces non fonctionnalisées. L'objectif de ce second protocole est d'éviter des forces d'interaction non spécifiques engendrées par le contact direct entre la surface de la poutre et celle du substrat.

---

3. Chapitre 1, paragraphe 2.1.





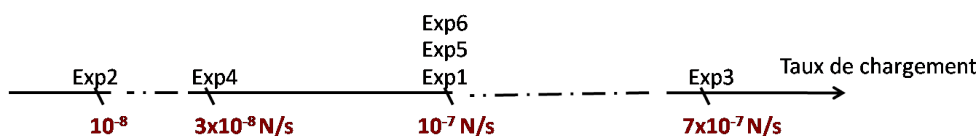
**Figure 3.8** - Scénario d'un processus d'A/R lors d'une mesure AFM de l'interaction entre les brins d'ADN complémentaires fixés aux surfaces de la pointe et du substrat.

Deux modes de mesure sont proposés. Un premier mode appelé *ponctuel*, consiste à répéter plusieurs fois le processus A/R sur un même point du substrat. Un deuxième mode appelé *scan*, consiste à répéter le processus A/R en balayant la surface du substrat. Le mode ponctuel expérimente la répétabilité de la mesure sur un même système moléculaire. Le mode scan compare la réponse de plusieurs systèmes moléculaires supposés identiques.

Différents paramètres expérimentaux sont étudiés durant ces mesures (figure 3.8) :

- Constante de raideur de la poutre  $k$  :  $0.1\text{N/m}$ ,  $0.03\text{N/m}$  et vitesse d'A/R  $v$  :  $0.1\mu\text{m/s}$ ,  $1\mu\text{m/s}$ ,  $5\mu\text{m/s}$ . Quatre taux de chargement ont donc été utilisés :  $10^{-8}\text{N/s}$ ,  $3 \cdot 10^{-8}\text{N/s}$ ,  $10^{-7}\text{N/s}$ ,  $7 \cdot 10^{-7}\text{N/s}$ .
- Protocole de préparation de l'échantillon, B : Annexe X1, et NB : Annexe X2.
- Mode de mesure :
  - mode ponctuel : 256 itérations sur le même point,
  - mode scan : balayage d'une surface de  $1\mu\text{m} \times 1\mu\text{m}$  sur 1024 points ( $32 \times 32$ ).

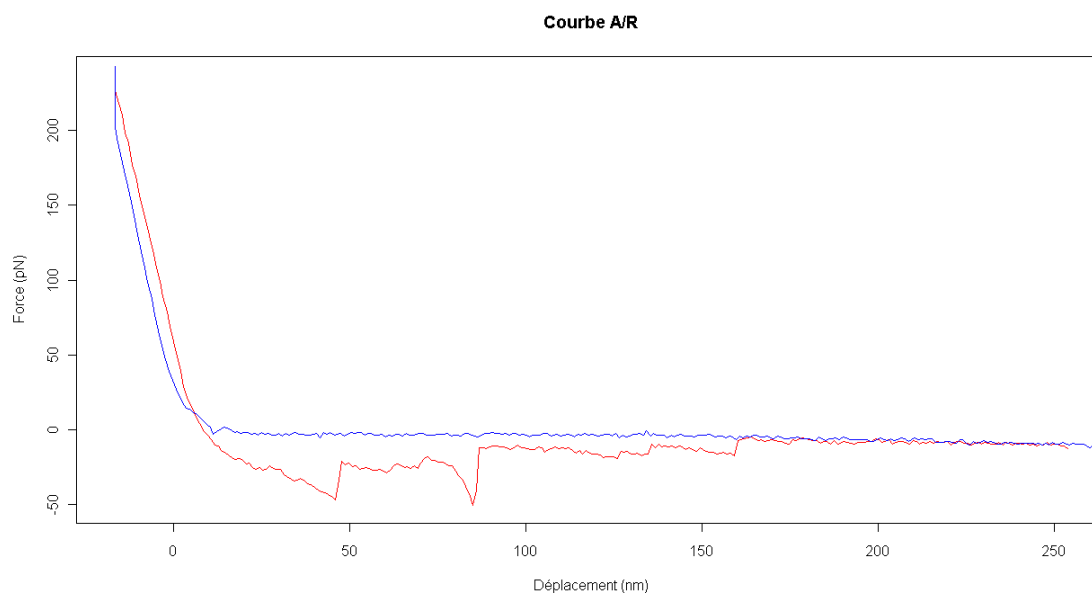
Le tableau ci-dessous est un récapitulatif de l'ensemble des expériences menées (tableau 3.1). Le taux de chargement pour chaque expérience est représenté à la figure 3.9.



**Figure 3.9** - Taux de chargement pour chaque expérience de mesure d'interaction inter brins d'ADN sur la séquence S1.

	$v = 0.1\mu\text{m}/\text{s}$	$v = 1\mu\text{m}/\text{s}$	$v = 5\mu\text{m}/\text{s}$
$k = 0.1\text{N}/\text{s}$	Exp2 (B)	Exp1 (B) Exp5 (NB) Exp6 (B) (s)	Exp3 (B)
$k = 0.03\text{N}/\text{s}$		Exp4 (B)	

**Tableau 3.1** - Tableau récapitulatif des expériences de mesures d'interaction inter brins d'ADN sur la séquence S1 classées en fonction de  $k$  (raideur de la poutre) et  $v$  (vitesse d'A/R). (B) : protocole de fonctionnalisation bloquant, (NB) : protocole de fonctionnalisation non bloquant, (s) : mode de mesure scan.



**Figure 3.10** - Courbe approche/retrait avec deux événements de rupture correspondant à la première itération de l'expérience Exp1.

A titre d'exemple, la courbe de la figure 3.10 illustre la mesure AFM obtenue pour la première itération de l'expérience Exp1 ( $v = 1\mu\text{m/s}$ ,  $k = 0.1\text{N/m}$ , B).

Pour l'estimation des forces de rupture, seules les expériences dont les mesures correspondant à des réponses moléculaires sont retenues à travers l'analyse statistique proposée dans la partie suivante.

### 3 Extraction des données significatives par analyse statistique

La distinction des expériences pour lesquelles un bruit de mesure trop important empêche l'exploitation des données est primordiale. Ce bruit peut être dû au matériel expérimental utilisé, ou à des paramètres expérimentaux (vitesse, raideur de poutre, protocole de fonctionnalisation) inadaptés pour la mesure d'une réponse moléculaire.

Dans le but d'identifier les expériences pendant lesquelles les réponses moléculaires sont enregistrées, les données expérimentales obtenues sont analysées à l'aide d'un indicateur statistique quantitatif. Ce dernier est utilisé pour qualifier chaque itération de chaque expérience. Les outils statistiques utilisés sont des outils simples, tels que la moyenne, la médiane, la variance et l'histogramme. L'environnement et le langage statistique open source **R** sont utilisés dans toutes les analyses statistiques qui suivent.

#### 3.1 Méthode statistique pour l'identification des expériences significatives

Le principe de cette analyse est de comparer les allures des courbes A/R. La comparaison est réalisée à l'aide d'une mesure de la dispersion du nuage de points de chaque courbe A/R. Le calcul de cette dispersion consiste en une variance. Le formalisme de ce calcul est détaillé dans ce qui suit.

Chaque expérience  $e$  est considérée comme une série statistique temporelle où chaque processus approche/retrait,  $\lambda$ , est lui-même une série statistique temporelle de mesures  $\varepsilon$ . Une courbe A/R peut être représentée par la variable  $\Gamma$  :

$$[\Gamma_\lambda]_e = [(f_\varepsilon, n_\varepsilon)_\lambda]_e \quad (3.1)$$

avec  $f_\varepsilon$  une force enregistrée correspondant à un point de la courbe approche/retrait  $\lambda$  de l'expérience  $e$ .  $n_\varepsilon$  est l'occurrence de  $f_\varepsilon$ . La variance calculée pour la variable  $[\Gamma_\lambda]_e$ , notée  $[V_\lambda]_e$ , décrit la dispersion des points.

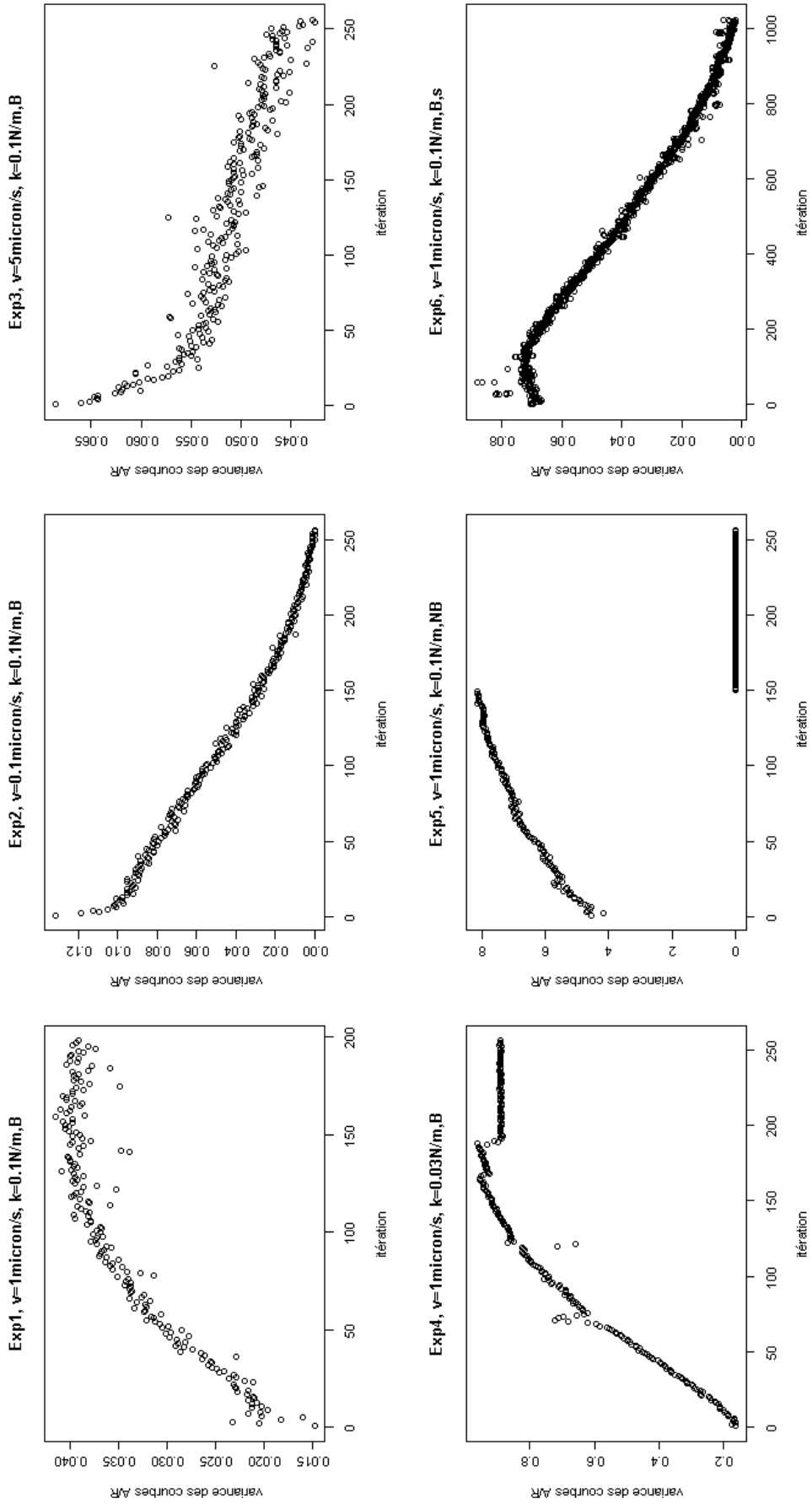
En représentant  $V_\lambda$  en fonction de  $\lambda$ , pour chaque expérience  $e$ , les profils des courbes A/R sont comparés. Cette représentation est une *courbe de variance*. La répétabilité du processus A/R est donc analysée à travers ces courbes pour chaque expérience.

Les courbes de variance représentées sur la figure 3.1 sont monotones. Leur sens de variation semblent dépendre de la vitesse. Celles-ci sont croissantes pour  $v = 1\mu\text{m/s}$  (Exp1, Exp4, Exp5), et décroissantes pour  $v = 0.1\mu\text{m/s}$  (Exp2) et  $5\mu\text{m/s}$  (Exp3). Pourtant, l'expérience Exp6 ne

répond pas à cette hypothèse. Ici, la monotonie des courbes de variance est une information plus pertinente que le sens de variation des fonctions représentées par ces courbes. En effet, cette monotonie renseigne la dépendance des événements successifs.

Dans l'hypothèse d'une dépendance due à des limites inhérentes du protocole de mesure (manipulation de l'AFM, matériel expérimental ou valeurs de paramètres expérimentaux inadaptés), le retrait de la partie *approche* des courbes A/R devrait distinguer les données significatives.

Figure 3.11 - Courbes de variances de données entières.  $V_\lambda$  en fonction de  $\lambda$ .



### 3.2 Identification des données exploitables

Dans l'hypothèse d'une dépendance due au système de translation, donc à l'AFM, le simple retrait de la partie approche des courbes A/R devrait être suffisant pour faire disparaître la monotonie sur les courbes de variance. Dans ce cas, l'expérience correspondante est considérée comme exploitable. Ainsi, un nouvel ensemble de données est considéré. Dans cet ensemble, il est admis que la partie approche des courbes A/R ne dépend que de l'AFM. Cet ensemble est appelé *données coupées*, par opposition aux *données entières* comprenant les courbes A/R entières. La comparaison des courbes de variance des deux ensembles de données conduit à l'identification des données exploitables.

La monotonie disparaît sur les courbes de variance correspondant aux expériences Exp1, Exp2 et Exp6. Lors de ces expériences, un seul type de poutre est utilisé :  $k = 0.1N/m$  et deux vitesses :  $v = 1\mu m/s$  et  $0.1\mu m/s$ . Pour cet ensemble d'expériences, il est certain que la monotonie des courbes entières provient du protocole de mesure, puisque le retrait de la partie approche suffit à isoler le phénomène de dépendance (figure 3.12).

La monotonie est persistante sur les courbes de variance correspondant aux expériences Exp3, Exp4, Exp5 (figure 3.13). Ces données sont donc considérées comme des données non exploitables puisqu'il n'est pas possible de déterminer la cause de cette monotonie. Néanmoins, plusieurs phénomènes peuvent être à l'origine de la monotonie persistante. Pour Exp5, le protocole de fonctionnalisation peut être mis en cause. En effet, des forces non spécifiques ont pu être enregistrées et ainsi altérer les mesures. Pour Exp4, une raideur de poutre faible ( $0.03N/m$ ) peut être mise en cause. En effet, ce type de poutre peut induire un bruit important dans les milieux aqueux, en considérant les forces hydrodynamiques par exemple. Enfin, dans Exp3, la vitesse élevée ( $5\mu m/s$ ) peut aussi être à l'origine de la monotonie.

Pour les données significatives, les paramètres expérimentaux correspondent à une vitesse d'A/R inférieure à  $0.1\mu m/s$ , une raideur de poutre supérieure à  $0.03N/m$ , un taux de chargement inférieur à  $10^{-7} N/s$ , et enfin, le protocole de fonctionnalisation utilisé comprend une étape de fixation des molécules bloquantes telle que l'éthylamine.

Les trois expériences significatives sont exploitées dans le paragraphe suivant pour estimer la force de rupture.

Paramètres de mesure	Calibration
$v$	$[0 ; 1 ] \mu m/s$
$k$	$]0.03 ; 0.1 [ N/m$
Protocole de fonctionnalisation	Protocole B
Taux de chargement	$[0 ; 10^{-7} ] N/s$

**Tableau 3.2** - Tableau récapitulatif des paramètres de mesure.

Figure 3.12 - Comparaison des courbes de variance de données entières et des données coupées pour les expériences exploitables.

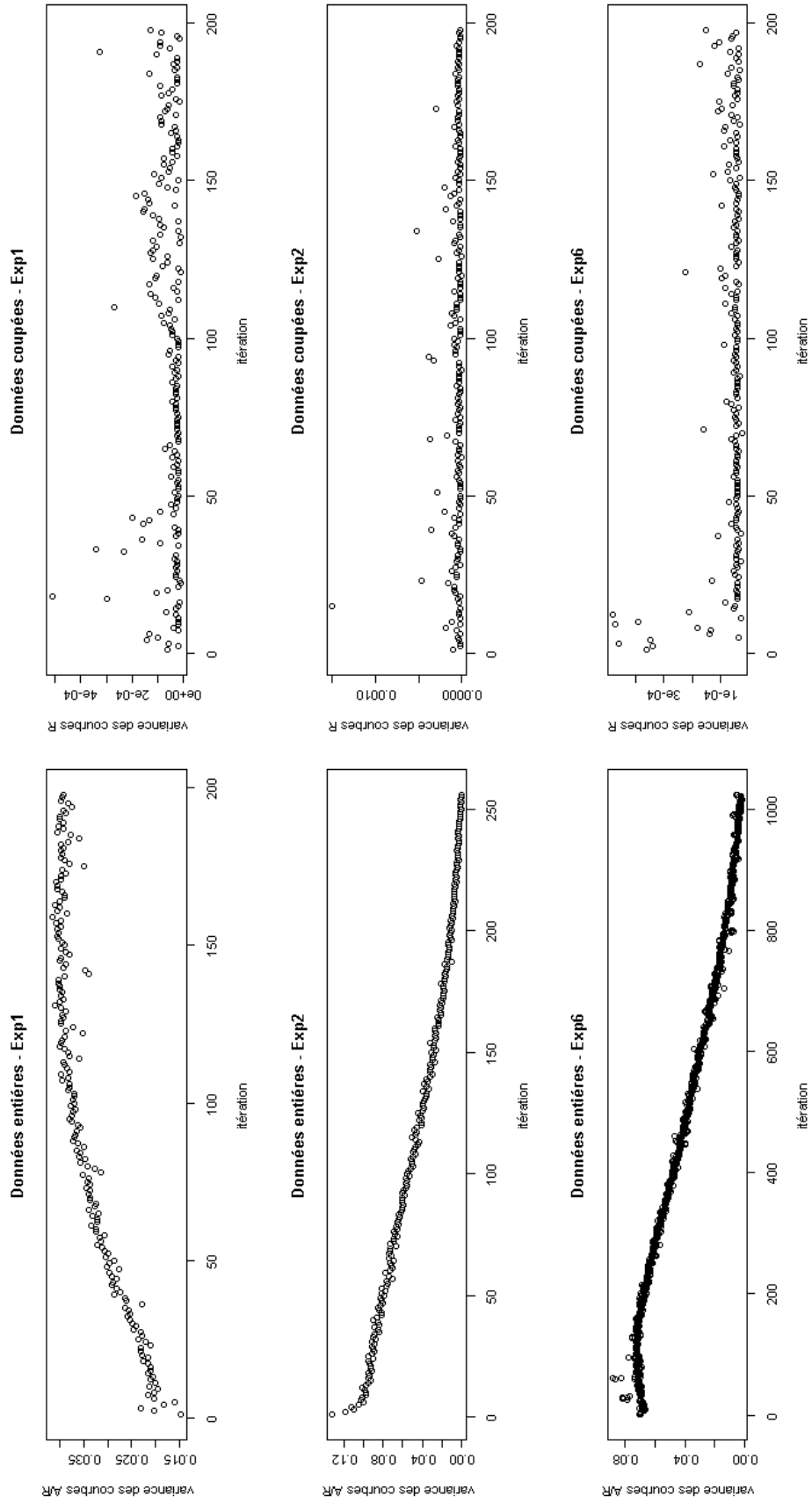
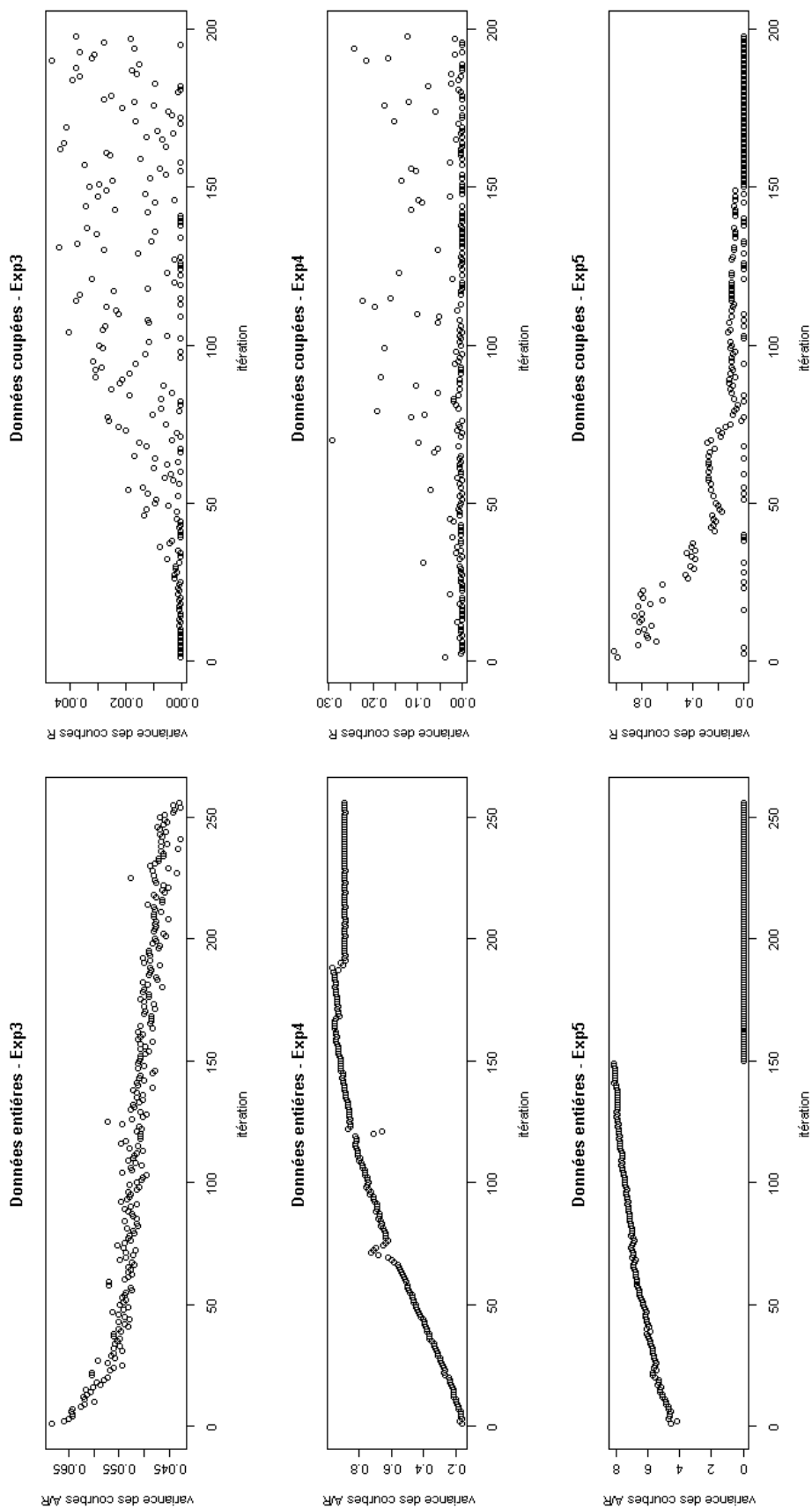


Figure 3.13 - Comparaison des courbes de variance de données entières et des données coupées pour les expériences non exploitables.





### 3.3 Moyennes et distributions des forces de ruptures

Le système biomoléculaire étudié est supposé être composé uniquement de deux brins d'ADN en interaction<sup>4</sup>. L'objectif de notre étude est de caractériser la rupture définitive des liaisons. Pour une mesure donnée, la force de rupture  $F_r$  correspond donc au maximum des sauts enregistrés sur la courbe A/R.

Sur la figure 3.14 sont représentées les forces de rupture  $F_r$  extraites à partir des mesures pour chaque expérience. Une distribution homogène est observée sur les données exploitables. En effet, les écarts entre la moyenne (en rouge) et la médiane (en vert) restent très faibles. Les valeurs enregistrées pour la force de rupture sont inférieures à  $150pN$  avec des taux importants : 99% pour Exp1, 98% pour Exp2 et 99% pour Exp6. La valeur moyenne est donc représentative sur ces échantillons et est égale à  $43pN$  pour Exp1,  $38pN$  pour Exp2 et  $36pN$  pour Exp6.

Les histogrammes représentés sur la figure 3.15, montrent le comportement variable des données pour chaque expérience. La classification est réalisée avec un découpage constant de  $10pN$ . Pour les données exploitables, entre 13 et 15 classes statistiques sont observées. Ces histogrammes montrent la variabilité suspectée dans l'étude bibliographique de la force de rupture.

---

4. Chapitre 3. paragraphe 1.3.1.

**Figure 3.14** - Les Forces de rupture extraites pour les expériences exploitables (Exp1, Exp2 et Exp6) et pour les données exploitables (Exp3, Exp4 et Exp5). L'écart entre la moyenne (en rouge) et la médiane (en vert) donne une indication sur la représentativité de la valeur moyenne pour chaque expérience.

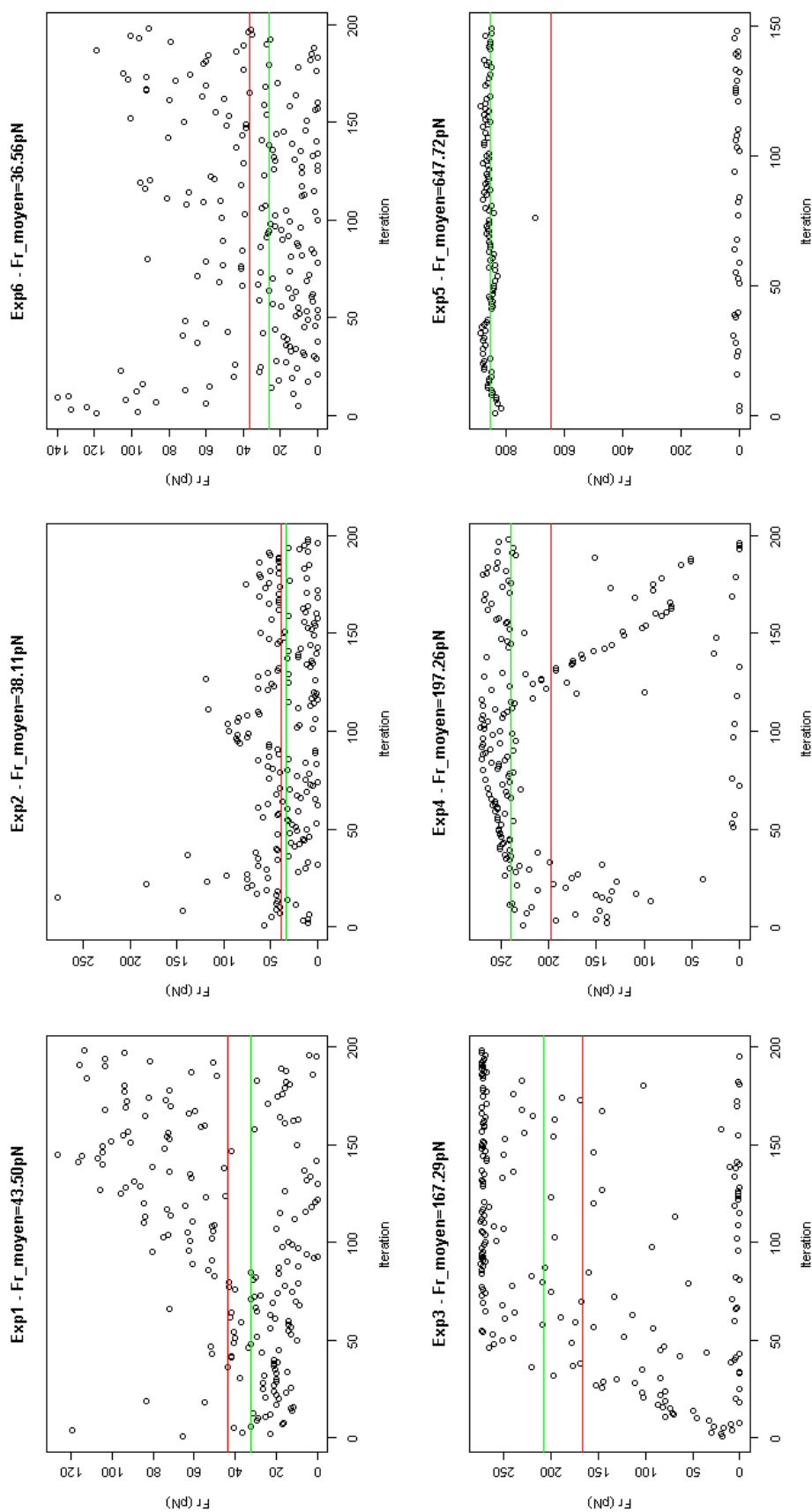
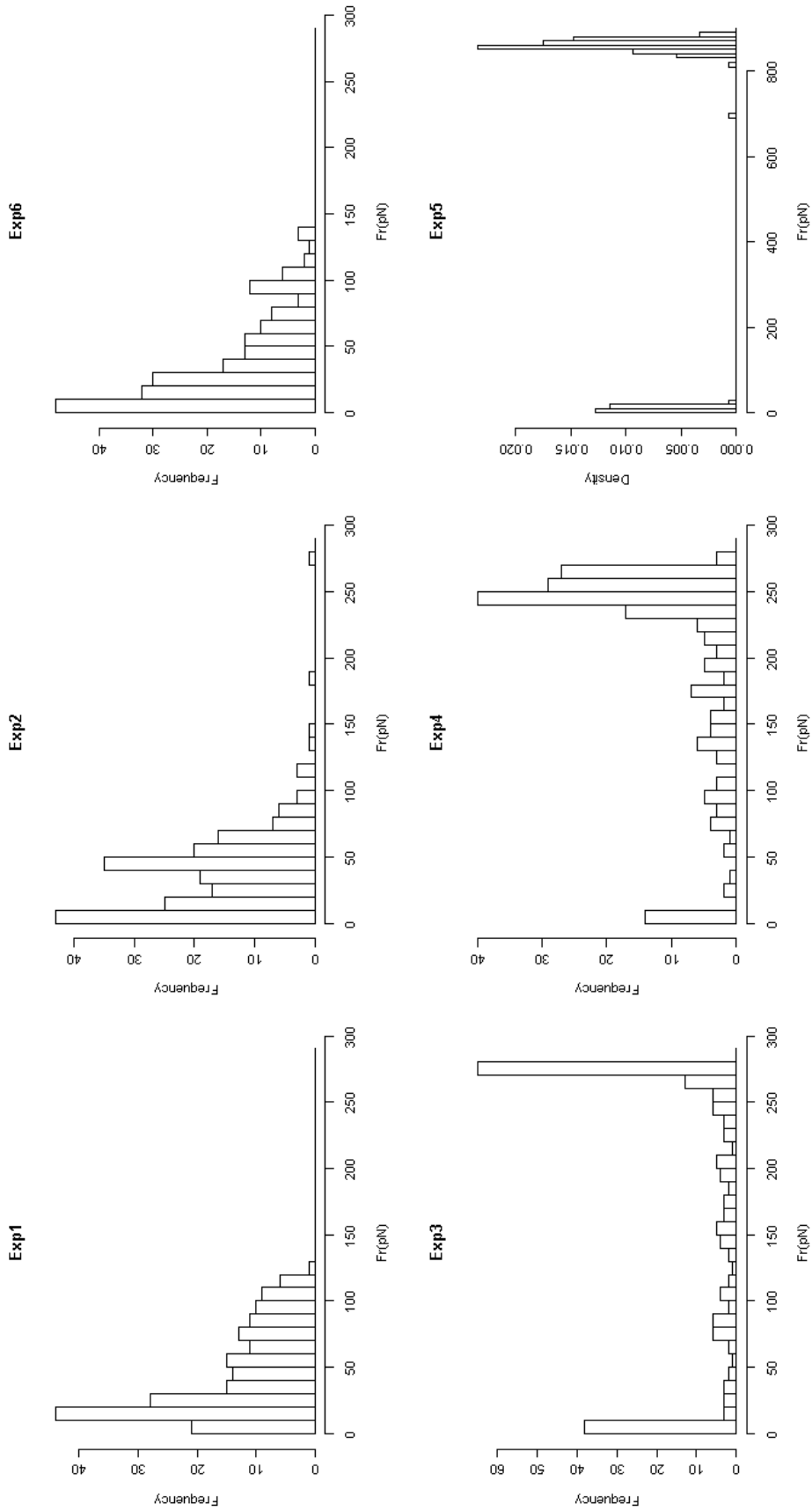


Figure 3.15 - Histogrammes des forces de rupture.



### 3.4 Conclusion

Pour la séquence  $S_1$ , la force de rupture est estimée dans l'intervalle  $[36pN, 44pN]$  à partir des données significatives identifiées à l'aide de la méthode des courbes de variance.

La variabilité décrite dans la littérature [Sattin 04, Morfill 07] apparaît sur nos mesures, et rend complexe la qualification de la stabilité de la séquence  $S_1$  pour la réalisation d'une adhésion lors d'un assemblage de micro-objets. L'utilisation d'une séquence donnée dans notre problématique de l'A.A microscopique, nécessite le contrôle de cette variabilité. Une méthode par classification énergétique est proposée dans la prochaine partie afin d'identifier dans quelle mesure il est possible de contrôler cette variabilité.

## 4 Analyse énergétique des données expérimentales

Dans la partie précédente, la caractérisation de la stabilité de la séquence  $S_1$  à travers l'estimation des valeurs moyennes des forces de rupture montre un caractère très variable (figure 3.14). Ce caractère s'exprime à travers le nombre de *classes statistiques* observées sur les histogrammes des forces de ruptures mesurées (figure 3.15). La difficulté dans l'utilisation de l'ADN réside dans le contrôle de cette variabilité. En effet, les origines de ce comportement se situent à différentes échelles : atomique, moléculaire et populationnelle. Il est donc difficile de réduire ou de contrôler directement la variabilité observée.

Dans cette partie, une méthode d'analyse énergétique est proposée. A travers cette méthode, il s'agit de s'affranchir de cette variabilité, et de proposer un outil pour l'évaluation de la stabilité d'une séquence en bases azotées. Cette méthode est basée sur une classification énergétique des types de réponses moléculaires pour un couple de brins d'ADN en fonction de ses configurations possibles.

Dans le premier paragraphe, les différents types de configurations pour une séquence en bases azotées sont introduits. L'étude énergétique des événements de rupture d'une mesure AFM en fonction des types de configuration possibles est proposée dans le deuxième paragraphe. Enfin, l'analyse des données expérimentales par cette méthode est détaillée.

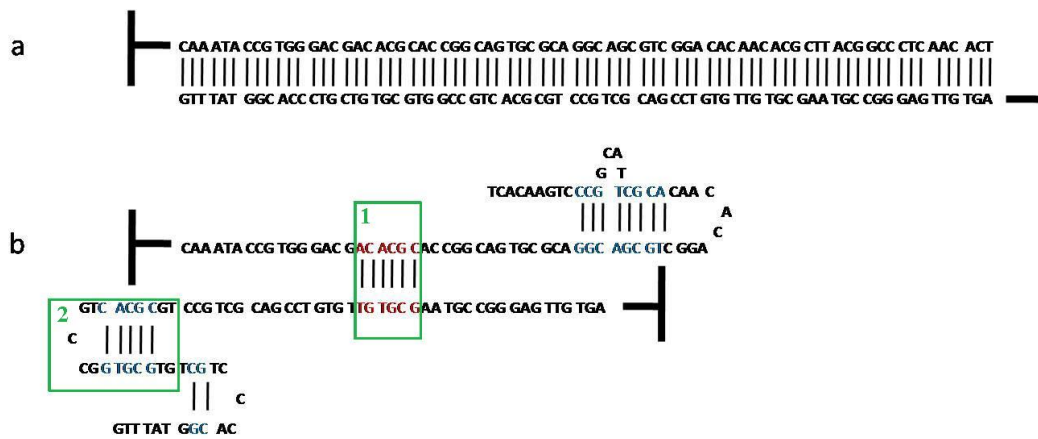
### 4.1 Variabilité de configuration et force de rupture

Pour une séquence d'ADN donnée, lors d'un processus d'hybridation, plusieurs configurations sont possibles. Cette caractéristique est principalement due aux appariements WC<sup>5</sup> possibles entre des bases d'indices différents. Ces configurations sont conventionnellement classées en 2 catégories : "*Perfect Match*" (PM) pour un appariement optimal sous la forme d'une double hélice (figure 3.16.a), et "*Mis Match*" (MM) pour un appariement non optimal entre des bases d'indices différents (figure 3.16.b. boîtes 1 et 2).

Dans la catégorie MM, quatre types sont distingués :

---

5. Chapitre 1, tableau 1.2 : Complémentarité WC.



**Figure 3.16** - Différents types de configuration pour une même séquence. a. Configuration "Perfect Match". b. Combinaisons de deux configurations : boîte 1. Configuration hétéro-complémentaire, boîte 2 : Configuration en épingle à cheveux.

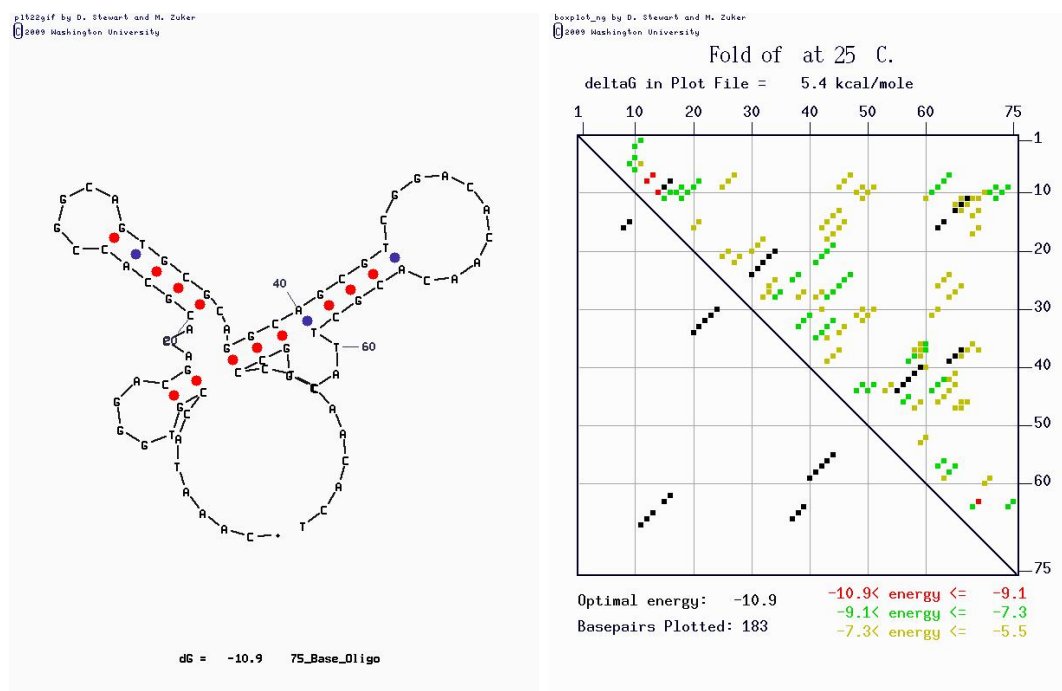
- *hétéro-complémentaire* : hybridation de quelques segments complémentaires ne possédant pas le même indice et appartenant respectivement à deux brins avec la même séquence en bases azotées (figure 3.16.b.1).
- *auto-complémentaire* : hybridation de quelques segments complémentaires ne possédant pas le même indice et appartenant au même brin.
- *épingle à cheveux* : configurations auto-complémentaires pour un même brin (figure 3.16.b.2).
- *"mis-match" complexe* : combinaisons des configurations *épingle à cheveux* et *hétéro-complémentaire* (3.16.b).

Des méthodes algorithmiques de prédiction de ces configurations existent. Elles consistent à répertorier tous les appariements WC possibles entre les bases de deux brins d'ADN. L'algorithme le plus répandu est **Fold** de Zuker et al. [Zuker 03]. Le logiciel OligoAnalyzer 3.1, basé sur Fold et disponible sur la plateforme de IDT SciTools<sup>6</sup>, a été utilisé pour générer les différents types de configuration pour les séquences étudiées. Par exemple, pour la séquence  $S_1$  utilisée dans les mesures AFM détaillées dans la partie précédente, 14 configurations *épingle à cheveux* (figure 3.17), et 118 configurations *hétéro-complémentaire* sont répertoriées.

Chaque configuration est caractérisée par une valeur énergétique  $\Delta G$ , calculée en fonction des bases appariées selon le modèle du plus proche voisin (cf. chapitre 2). Les configurations retenues sont celles pour lesquelles les segments appariés comportent un nombre de bases supérieur à 3. Ce critère impose que le  $\Delta G$  des configurations soit approximativement supérieur à  $4Kcal/mol$ . Ainsi, pour une séquence en bases azotées donnée, les valeurs qui caractérisent chaque configuration constituent un ensemble de valeurs énergétique théorique caractéristiques de la séquence. Cet ensemble énergétique théorique est noté  $\Omega_{configs}$ .

Dans l'hypothèse d'une correspondance simple entre l'ensemble des valeurs énergétiques théoriques  $\Omega_{configs}$  et le nombre de classes statistiques observées sur l'histogramme des forces de ruptures mesurées, la variabilité obtenue pourrait être expliquée. Or, Sattin et al. [Sattin 04] démontrent que le nombre de classes statistiques observées sur les histogrammes des forces de rupture mesurées augmente avec le nombre de mesures. Une correspondance simple ne peut

6. <http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/>



**Figure 3.17** - Les configurations MM de type épingle à cheveux. A gauche, 3 épingles à cheveux sont formées. A droite, carte des configurations MM de type épingle à cheveux d'après mFold. Les deux axes représentent les indices des bases de la séquence étudiée. 14 configurations sont stables (Analyse Hairpin sur OligoAnalyzer 3.1).

donc être établi entre les mesures, et les différentes configurations caractérisées par leur valeurs énergétiques. En effet, certains paramètres physiques ne sont pas pris en compte dans l'approche thermodynamique comme par exemple, la vitesse du chargement mécanique par la poutre.

Ainsi, pendant une mesure AFM, les deux populations de brins d'ADN sont soumises à des phénomènes physiques complexes. L'appariement entre les bases complémentaires des deux brins en interaction a lieu pendant la **phase approche**. Le type de configuration adopté par les deux brins pendant cette phase dépend de la séquence en bases azotées des brins. Elle dépend aussi des paramètres environnementaux (la température et la concentration saline) et des paramètres expérimentaux (la vitesse, le taux de chargement, et le temps accordé aux deux brins pour s'apparier).

Pendant la **phase retrait**, la réponse moléculaire du complexe formé par les deux brins est mesurée. Cette réponse dépend de la configuration adoptée pendant la phase approche, des paramètres environnementaux ainsi que des paramètres expérimentaux.

Ainsi, même si pour une séquence donnée, le nombre de configurations est déterminé, la relation entre la réponse moléculaire au chargement mécanique de la poutre et le nombre de ces configurations n'est pas directe. L'identification de cette relation dépend de la détermination de deux informations : le nombre de liaisons WC formées pendant la phase approche et le nombre de liaisons WC rompues pendant la phase retrait. Or, pendant les expériences AFM, il est impossible d'obtenir directement ces deux informations. La force de rupture  $F_r$  mesurée ne contient pas ces informations et ne suffit naturellement pas à décrire pleinement le phénomène étudié.

Afin d'évaluer la stabilité de l'assemblage à travers des expériences AFM, il est nécessaire d'extraire des informations complémentaires des courbes A/R.

## 4.2 Etude des événements locaux de rupture

A partir de chaque événement local de rupture d'une courbe A/R (figure 3.18), une force de rupture  $F_r$  et une distance de rupture, notée  $d_r$  sont extraites. Leur produit est la variation d'énergie expérimentale<sup>7</sup> notée  $\Delta G_r$ . Cette dernière donne aussi une mesure de la stabilité de l'assemblage.

La distance de rupture correspond au déplacement effectué par la poutre (donc par une extrémité d'un brin) pendant un événement local de rupture. La distance de rupture n'est pas directement représentée sur les courbes A/R. En effet, le déplacement représenté sur ces courbes est celui de la base de la poutre et non de la pointe. Pour obtenir la distance de rupture il faut soustraire du déplacement mesuré  $d$  la flexion de la poutre  $z_p$ , selon l'équation 3.2 [Butt 05].

$$d_r = d - z_p \quad (3.2)$$

$$z_p = F_r/k \quad (3.3)$$

avec  $k$  la raideur de la poutre et  $d$  le déplacement effectivement mesuré lors de l'expérience.

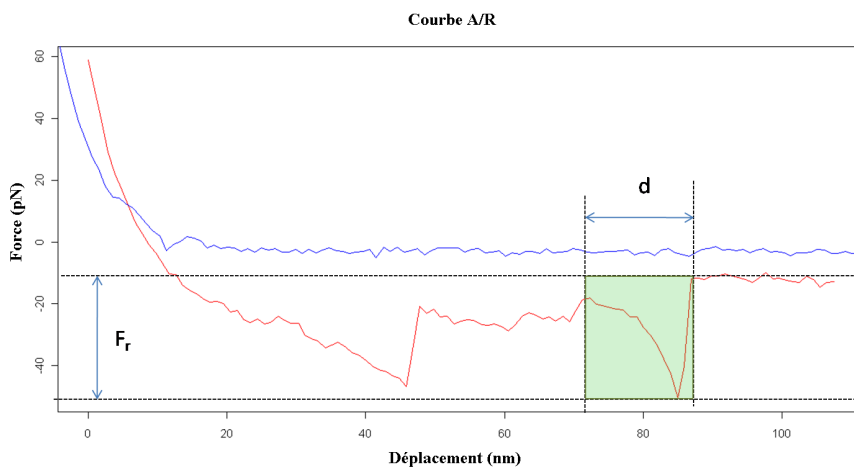


Figure 3.18 - Extraction de la force de rupture  $F_r$  et de la variation de déplacement  $d$ .

La variabilité observée sur les mesures des forces de rupture est naturellement observée sur les énergies expérimentales  $\Delta G_r$  estimées. La classification de ces valeurs énergétiques nécessite la comparaison des résultats expérimentaux à une référence théorique. Le choix de l'ensemble des valeurs énergétiques théoriques  $\Omega_{configs}$ , caractéristiques de la séquence étudiée, comme référence théorique rend possible cette comparaison. Ainsi, une énergie expérimentale  $\Delta G_r$  est comparée à une énergie théorique  $\Delta G$  (appartenant à l'ensemble  $\Omega_{configs}$ ) pour un même nombre de bases appariées. La connaissance du nombre de bases n'est pas suffisante. En effet, le calcul de  $\Delta G$  nécessite la connaissance de l'identité de la séquence en bases azotées. Or,

7. La constante de conversion d'une énergie exprimée en  $Kcal/mol$  en force exprimée en  $pN$  est  $C = 0.397$ , avec la distance  $d_r$  en  $nm$ .

lors d'une mesure AFM, il est impossible de déterminer la portion hybridée de la séquence. L'identité de la séquence qui correspond à  $\Delta G_r$  reste donc une inconnue.

Néanmoins, il est possible d'encadrer la valeur de  $\Delta G$  pour  $n$  bases appariées consécutivement par les énergies des séquences minimale,  $\Delta G_{min}^n$ , et maximale,  $\Delta G_{max}^n$ . Le calcul de ces deux énergies est possible à partir des équations 3.4 et 3.5 définies au chapitre précédent (page 40).

$$\Delta G_{min}^n = 1.2n - 1.1 \quad (3.4)$$

$$\Delta G_{max}^n = 3.4n - 3.1 \quad (3.5)$$

L'énergie libre  $\Delta G$  d'une séquence de  $n$  bases hybridées selon une configuration de type *PM* ou de type *hétéro-complémentaire*, notée  $\Delta G^n$  est comprise dans l'intervalle  $[\Delta G_{min}^n, \Delta G_{max}^n]$ .

La comparaison de  $\Delta G_r$ , énergie caractéristique d'un événement de rupture avec une distance de rupture  $d_r$ , à  $\Delta G^n$  nécessite une hypothèse sur le nombre de bases séparées (ou le nombre de liaisons WC brisées) pendant la rupture. En effet, pour un déplacement  $d_r$ , on peut supposer qu'au moins  $n$  bases sont séparées, avec  $n = d_r/d_{base}$  et  $d_{base}$  la distance entre deux bases consécutives d'un brin<sup>8</sup>. La variation d'énergie  $\Delta G_r$  en fonction de  $n$  est donc notée  $\Delta G_r^n$ . Cette hypothèse n'écarte pas la possibilité que le nombre de bases séparées soit supérieur à  $n$  et constitue une première approche pour l'étude comparative de  $\Delta G_r^n$  et  $\Delta G^n$ . L'intervalle d'encadrement de  $\Delta G^n$  choisi comme référence de comparaison et de classification des  $\Delta G_r^n$  est illustré sur la figure 3.19. Ce graphique représente les droites  $D_{AT}$  et  $D_{GC}$  des séquences minimale et maximale qui encadrent  $\Delta G^n$ .

Une  $\Delta G_r$  correspondant à un événement local de rupture d'une distance de rupture  $d_r$  peut apparaître dans l'une des 3 catégories suivantes :

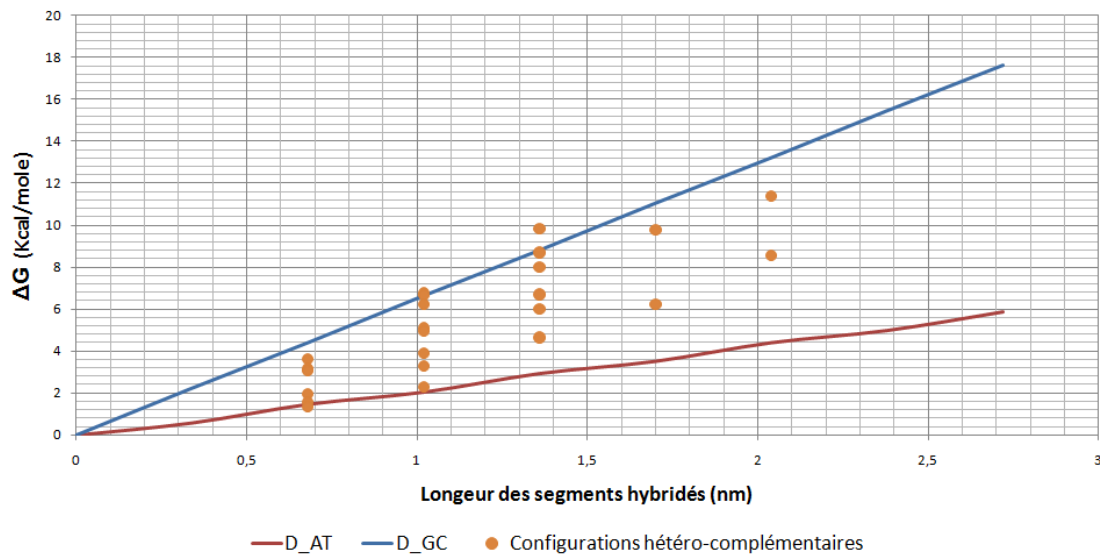
1.  $\Delta G_r^n > \Delta G_{max}^n$  : la stabilité est optimale, une variation d'énergie plus grande que celle prévue par l'approche théorique est observée.
2.  $\Delta G_r^n \in [\Delta G_{min}^n, \Delta G_{max}^n]$  : la stabilité est celle prévue par l'approche théorique. La considération de ces événements de rupture dans la stabilité de l'assemblage dépend de l'amplitude et la fréquence des mesures énergétiques  $\Delta G_r^n$ .
3.  $\Delta G_r^n < \Delta G_{min}^n$  : l'énergie des événements de ruptures correspondants est plus faible que celle prévue par l'approche théorique.

La superposition des valeurs de  $\Delta G_r^n$  à ce diagramme est une méthode de classification des événements de rupture pour une évaluation de la stabilité d'un assemblage. Ainsi, pour les événements de la catégorie 1, la variation d'énergie supérieure aux prédictions théoriques peut être due à une sous estimation de la distance de rupture prévue par l'hypothèse  $d_r = n d_{base}$ . Un nombre de liaisons rompues supérieur à  $n$ , ou une configuration géométrique rigide de type "*mis-match*" complexe comme la configuration illustrée à la figure 3.16.b. peuvent être à l'origine de cette sous estimation. Les événements de la catégorie 1 sont donc avantageux pour la stabilité.

La catégorie 3, à l'inverse de la catégorie 1, correspond aux événements de faible énergie. Cette situation est possible dans les cas où la rupture est comparable à la propagation d'une fissure (l'ouverture de quelques liaisons est suffisante pour séparer les deux brins). Ce type d'événement implique une configuration géométrique avec une ouverture similaire à une fermeture éclair.

8. En fonction de la salinité du milieu, une forme particulière est adoptée par l'ADN. Par exemple pour une concentration 1M, la forme adoptée est B. Pour cette forme,  $d_{base} = 0.34$  nm.





**Figure 3.19** - Diagramme des énergies libre de Gibbs calculée d'après le modèle du plus proche voisin en fonction du nombre de la longueur des segments hybridés. Les énergies des configurations PM et hétéro-complémentaires de la séquence  $S_1$ , l'une des séquences utilisées dans le projet Golem (Annexe A) sont représentées sur ce diagramme. Ces énergies sont comprises dans l'espace entre les deux droites  $D_{AT}$  et  $D_{GC}$ .

Cette méthode de classification donne une indication qualitative de la stabilité d'un assemblage par l'ADN en s'affranchissant de la variabilité du comportement de la molécule. Dans le paragraphe suivant, sont analysées par cette approche les données issues des 3 expériences sélectionnées comme significatives dans la partie dédiée à l'analyse statistique des données expérimentales.

### 4.3 Discussion des résultats expérimentaux

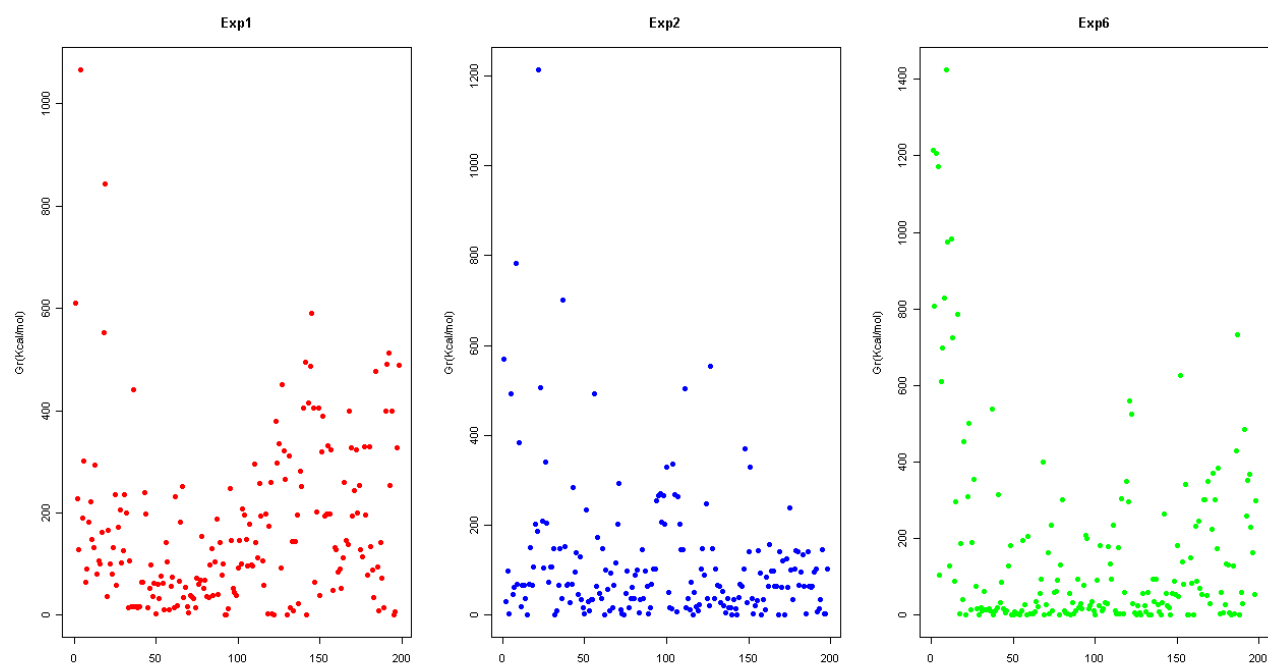
Les expériences significatives Exp1 ( $1\mu m/s$ ), Exp2 ( $0.1\mu m/s$ ) et Exp6 ( $1\mu m/s$ , scan) sont analysées afin d'extraire à partir des événements de rupture locaux les valeurs de  $\Delta G_r$  représentées à la figure 3.20.

La représentation graphique de  $\Delta G_r$  en fonction de  $d_r$  des événements de rupture locaux des 198 premières courbes A/R est illustrée sur la figure 3.21. 53% des points pour Exp1 et 58% des points pour Exp2 sont corrélés avec une parabole, dont l'équation peut s'exprimer sous la forme d'une énergie potentielle ou élastique :

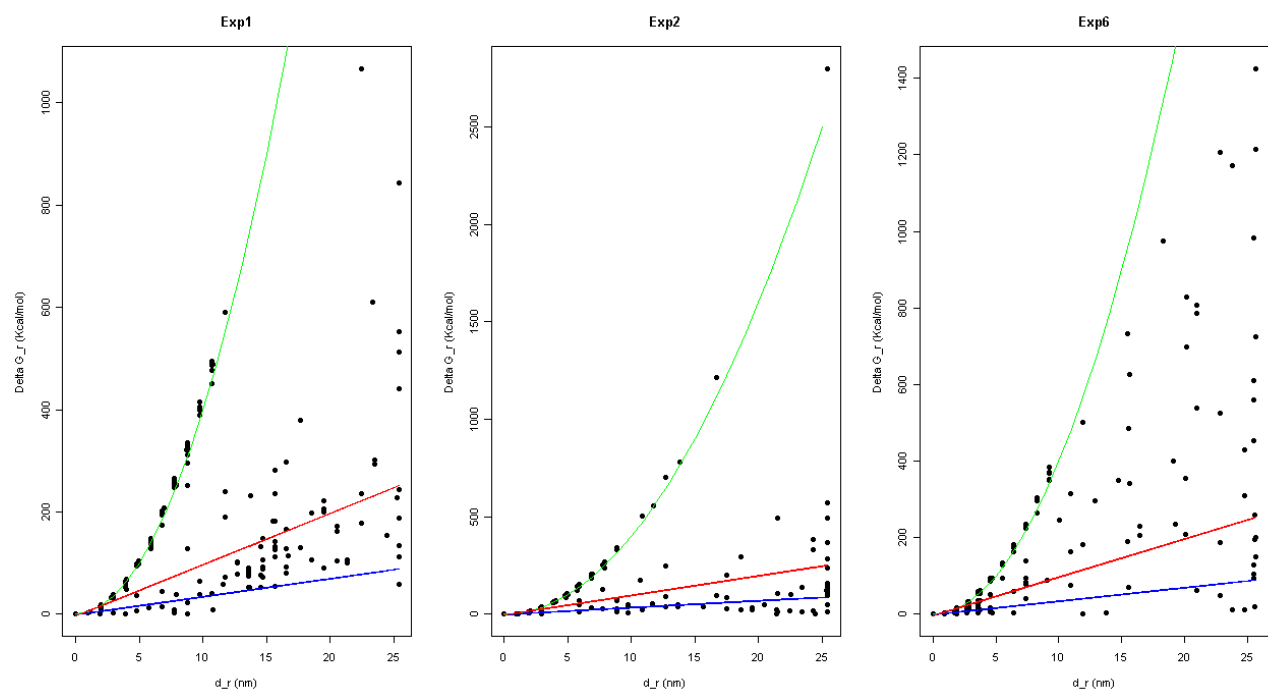
$$E_{parabole} = \frac{1}{2} k_p d_r^2 \quad (3.6)$$

La valeur de  $k_p$ , déterminée par une régression numérique est estimée à 0.02 N/m. L'origine de ce comportement peut provenir du comportement intrinsèque de la molécule ou de la méthode de mesure de l'AFM.

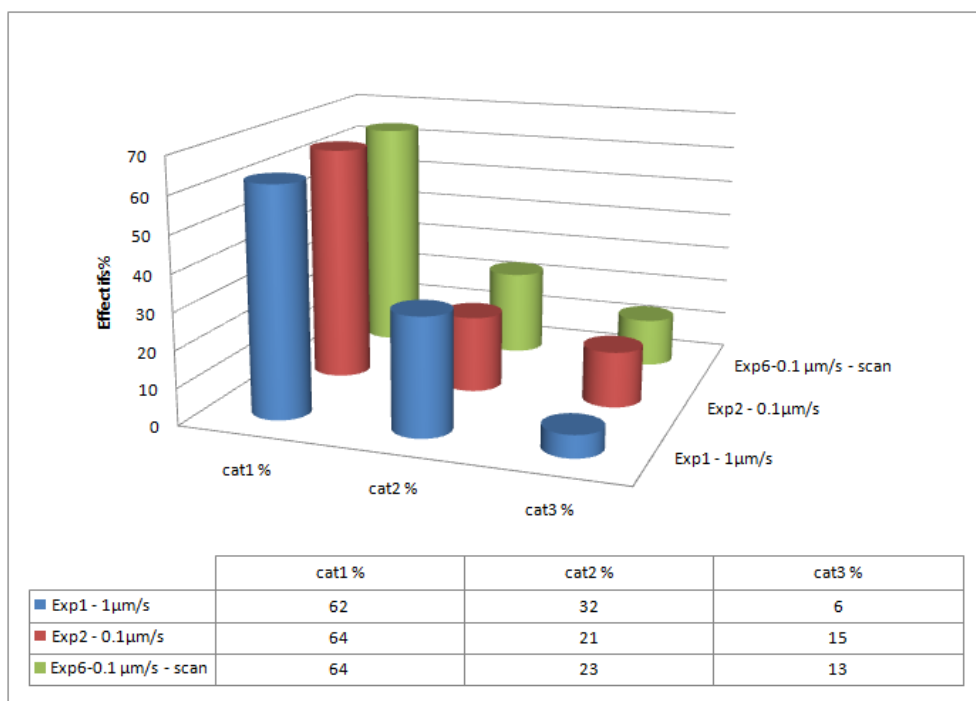
La superposition des points ( $\Delta G_r$ ,  $d_r$ ) montre une répartition non homogène sur les 3 différentes catégories. Une analyse numérique évalue l'effectif de chaque catégorie pour chaque expérience. Le résultat de cette analyse est représenté à la figure 3.22.



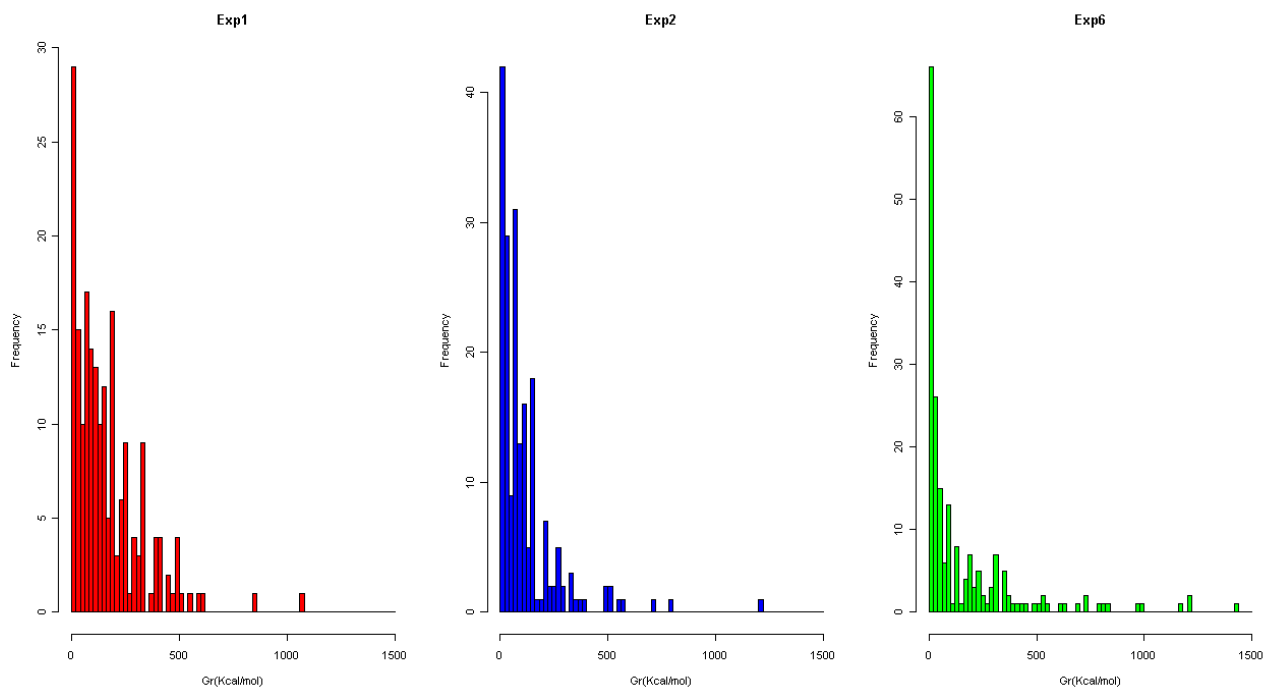
**Figure 3.20** - Energies  $\Delta G_r$  (Kcal/mol) des événements locaux de rupture des 198 premières courbes A/R pour les expériences *Exp1* ( $v = 1\mu\text{m/s}$ ), *Exp2* ( $v = 0.1\mu\text{m/s}$ ) et *Exp6* ( $v = 1\mu\text{m/s}$ , scan).



**Figure 3.21** - Analyse énergétique des événements locaux de rupture des 198 premières courbes A/R pour les expériences *Exp1* ( $1\mu\text{m/s}$ ), *Exp2* ( $0.1\mu\text{m/s}$ ) et *Exp6* ( $1\mu\text{m/s}$ , scan). La droite  $D_{AT}$  de la séquence minimale en bleu et la droite  $D_{GC}$  de la séquence minimale en rouge.



**Figure 3.22** - Répartition des événements locaux de rupture sur les catégories 1, 2 et 3 des 198 premières courbes A/R pour les expériences Exp1 ( $1\mu\text{m/s}$ ), Exp2 ( $0.1\mu\text{m/s}$ ) et Exp6 ( $1\mu\text{m/s}$ , scan).



**Figure 3.23** - Histogrammes des énergies de rupture  $\Delta G_r$  (Kcal/mol) des 198 premières courbes A/R pour les expériences Exp1 ( $1\mu\text{m/s}$ ), Exp2 ( $0.1\mu\text{m/s}$ ) et Exp6 ( $1\mu\text{m/s}$ , scan).

L'étude des effectifs des 3 catégories sur les expériences donnent des résultats similaires avec une distribution à 60% pour la catégorie 1 (cat 1) respectivement (62%, 64% et 64% respectivement pour Exp1, Exp2 et Exp6). La répartition sur les autres catégories est de 20% à 30% pour la catégorie 2 et de 6% à 13% pour la catégorie 3.

A travers ces 3 expériences réalisées sur la séquences  $S_1$  dans des conditions expérimentales proches, la consistance des résultats obtenus montre que la stabilité de la séquence étudiée est optimale à 60%.

Comme mentionné plus haut, la séquence  $S_1$  présente plus de 150 possibilités de configurations de type MM. Ces configurations impliquent un petit nombre de bases appariées. Les histogrammes de la figure 3.23 montrent aussi qu'un grand nombre des  $\Delta G_r^n$  relevés correspondent à des énergies faibles. Les événements locaux de rupture qui correspondent à ces énergies appartiennent à la catégorie 3 définie plus haut. L'élimination des événements correspondant à cette catégorie optimiserait la stabilité du velcro à base d'ADN et reviendrait à utiliser des séquences d'ADN pour lesquelles les configurations de faible énergie sont peu probables.

## 5 Conclusions

La caractérisation expérimentale de la stabilité du velcro constitué des brins d'ADN de séquence  $S_1$  est réalisée à travers des expériences de mesures AFM. Les données exploitables sont identifiées à l'aide d'une analyse statistique. A partir de ces données, une valeur moyenne de la force de rupture est estimée à  $39 \pm 4$  pN. Rapportée à la surface d'interaction<sup>9</sup>, elle est estimée à  $4nN/\mu m^2$ .

La variabilité observée sur les mesures est difficile à interpréter et rend la qualification de la stabilité du velcro vis à vis de l'assemblage complexe. Dans le but de s'affranchir de cette variabilité, et obtenir une indication quantitative sur la qualité du velcro, une méthode d'analyse énergétique des événements locaux de rupture est proposée. Celle-ci, appliquée aux mesures AFM réalisées sur la séquence  $S_1$ , montre la profusion des configurations de faible énergie avec un nombre de bases hybridées petit. L'élimination de ce type de configuration revient à optimiser la séquence en bases azotées et est proposée dans le chapitre suivant.

---

9. D'après les constructeurs, la pointe de la poutre peut être assimilée à une sphère de diamètre compris entre 20 et 60 nm.



---

# ETUDE ET CARACTÉRISATION EXPÉRIMENTALE DE LA SPÉCIFICITÉ

---

La spécificité désigne le caractère sélectif du processus d'auto-assemblage. Cette propriété caractérise le degré de "reconnaissance sélective" mutuelle des micro-composants. Elle est obtenue à partir de la complémentarité WC entre les brins d'ADN des deux populations fixées à leurs surfaces. Cette sélectivité peut être codée par la séquence en bases azotées des brins d'ADN pour réaliser l'adhésion. Pour notre problématique, il s'agit de proposer un algorithme de génération de séquence optimale pour réaliser cette propriété.

L'étude de la stabilité de la séquence  $S_1$  a démontré la nécessité de réduire les possibilités de formation de configurations de type "*Mis-Match*" (MM). La minimisation de ce type de configuration implique l'utilisation d'une séquence optimisée dont la configuration la plus probable est de type "*Perfect-Match*" (PM).

L'objectif de ce chapitre est ainsi de décrire, dans la première partie l'algorithme pour produire cette séquence puis dans une seconde partie les expériences obtenues par la microcopie à force atomique pour valider cette approche.

## 1 Etat de l'art des méthodes de génération de séquences

Pour un brin d'ADN donné, la séquence en bases azotées est codée de façon structurale par la séquence de son brin complémentaire. La base de ce codage est la complémentarité WC entre les bases G et C d'une part, et A et T d'autre part. Ainsi, le caractère spécifique de la molécule d'ADN consiste en l'unicité de cette composition. L'assemblage biomoléculaire spécifique en double hélice peut donc être utilisé pour mémoriser l'information et cette unicité.

Plusieurs méthodes algorithmiques et techniques expérimentales développées autour du do-

maine de la programmation moléculaire de l'ADN existent. Cette nouvelle discipline est commune à deux domaines de recherche, les ordinateurs à ADN [Adleman 94] et les nanostructures nucléiques [Seeman 83]. Le principe de base est analogue à la programmation machine des ordinateurs électroniques. Il s'agit en effet, d'une programmation visant à établir des principes de conception systématiques pour des systèmes moléculaires avec des propriétés désirées. Appliqué à l'ADN, l'objectif principal est de concevoir des séquences singulières et dissemblables. Ces dernières doivent répondre à des conditions particulières qui peuvent concerner les propriétés thermodynamiques. Par exemple, il peut s'agir pour notre problème d'éviter des structures stables de type MM. Cet objectif peut être atteint à travers la combinatoire de l'ADN. La difficulté principale concerne néanmoins le compromis entre l'optimisation de la stabilité et celle de la spécificité.

<b>Approches</b>	<b>Auteurs</b>
Recherche exhaustive	[Hartemink 98]
Recherche aléatoire	[Penchovsky 03]
Stratégie Template-map	[Frutos 97] , [Arita 02]
Graphes	[Feldkamp 02]
Méthodes stochastiques	[Tanaka 01]
Programmation dynamique	[Marathe 99]
Méthodes bio-inspirées	[Deaton 02]
Algorithmes évolutionnaires	[Shin 05]

**Tableau 4.1** - *Etat de l'art des algorithmes de génération de séquence [Feldkamp 02].*

Les algorithmes développés dans cette discipline sont nombreux et très variés. Le tableau 4.1 est un récapitulatif non exhaustif de l'état de l'art des algorithmes de génération de séquences. Deux grandes familles d'algorithmes répondent à cette problématique. La première optimise ces conditions pendant la construction de la séquence. Dans la deuxième famille, il s'agit de générer des séquences aléatoirement et de vérifier à posteriori, les conditions en utilisant un "filtre". Ce dernier est un ensemble de conditions prédéfinies pour conserver les séquences désirées. Dans les deux types d'algorithmes, les conditions sont définies pour des exigences de stabilité et spécificité en particulier.

## 2 Génération de séquence pour l'A.A microscopique

L'approche la plus adaptée à la combinatoire de notre problème est l'approche de type "filtre". Celle-ci sera au coeur du développement de séquences spécifiques dédiées à l'A.A microscopique à base d'ADN. Ce filtre peut être défini comme un ensemble de règles syntaxiques avec comme objectif principal, la réduction du nombre de MM lors du processus d'hybridation. Les définitions de ces règles sont exposées dans le premier paragraphe puis l'algorithme de génération de séquence est détaillé.

## 2.1 Principe de la génération de séquence

Lors de l'hybridation d'un brin fixé avec son complémentaire, plusieurs configurations sont possibles. Celles-ci dépendent essentiellement de la séquence en bases azotées des brins et sont répertoriées sur la figure 4.1. La solution proposée pour optimiser la stabilité est d'augmenter la probabilité de formation d'une configuration de type PM ou de réduire le nombre de configurations de type MM de faible énergie. Le cas A correspond à une hybridation parfaite de type PM entre les brins  $X$  et  $Y$ . Les cas sur la représentation de la figure 4.1, B,C,D,E,F et G représentent les configurations de type MM non désirées.

Dans la suite du document, la séquence « mère » est fixée sur le substrat par son extrémité 5'. Elle est indexée de 1 à  $n$ , où  $n$  est la longueur du brin en bases azotées dans le sens 5'-3'.

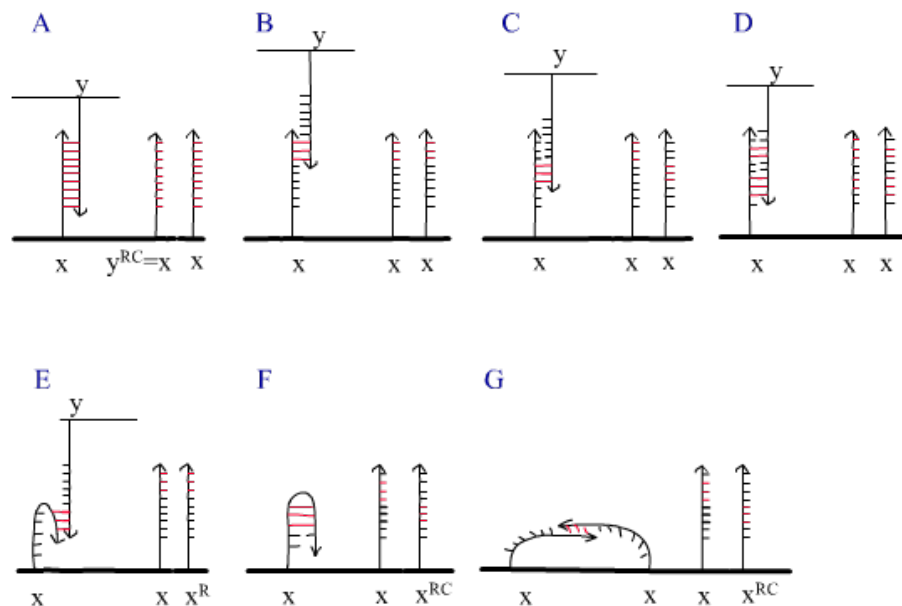


Figure 4.1 - Scénarii des configurations possibles entre deux brins d'ADN complémentaires.

Le filtre recherché pour répondre à cette problématique doit éliminer ces configurations indésirables. Les conditions qui constituent ce filtre, consistent en la comparaison des couples de segments de longueur  $\beta$  appartenant respectivement à chacun des brins en interaction. En informatique théorique, c'est une mesure de la différence entre deux mots de même longueur d'un langage donné. Pour comparer deux mots de même longueur  $m1$  et  $m2$ , la fonction  $H(m1, m2)$  est définie comme le nombre de positions pour lesquelles les symboles sont différents. Soit les deux mots  $m1 = ATCGAAA$  et  $m2 = ATGGAAT$  alors  $H(m1, m2) = 2$ .

L'utilisation de cette distance implique 6 comparaisons. Pour réduire le nombre de comparaisons, une optimisation est possible, en considérant les définitions de la complémentaire, notée d'un exposant C, et de la séquence inverse notée d'un exposant R (reverse).

### Séquence complémentaire :

Pour chaque symbole A (respectivement C) de la séquence mère, un symbole T (respectivement G) est équivalent sur la séquence complémentaire et inversement. Le brin obtenu est orienté 3'-5' pour la même indexation. Dans le cadre des expérimentations avec l'AFM, la séquence complémentaire est fixée sur la pointe de la poutre AFM par son extrémité 5'.



**Séquence inverse (reverse) :**

La séquence inversée est obtenue en changeant l'indexation d'une séquence de 5'-3' à 3'-5'.

**Propriété :**

La séquence complémentaire inverse d'une séquence  $Y$ , complémentaire de  $X$ , est  $X$  elle-même.

$$Y^{RC} = X \quad (4.1)$$

Cette propriété réduit le nombre de comparaisons entre les segments de brins correspondant aux 6 cas de MM de la figure 4.1. Ainsi, en considérant les segments  $x_i$  de longueur  $\beta$  d'une séquence  $X$ , la comparaison de ces segments aux segments  $y_j$  de la séquence complémentaire  $Y$  revient à comparer les segments  $x$  de la séquence étudiée à eux même avec des indices  $i$  et  $j$  différents.

Les 6 cas de MM : B, C, D, E, F et G de la figure 4.1 peuvent être résumés dans les 3 conditions suivantes :

- Pour éviter les cas B, C et D, chacun des deux segments  $x_i$  et  $x_j$  de la séquence produite doivent être différents.
- Pour éviter les cas F et G, chacun des deux segments  $x_i^R$  et  $x_j^C$  de la séquence produite doivent être différents.
- Pour éviter le cas E, chacun des deux segments  $x_i$  et  $x_j^R$  de la séquence produite doivent être différents.

avec  $i$  et  $j$  les indices différents des segments sur la séquence étudiée.

Ces comparaisons peuvent s'exprimer à travers la distance de Hamming. Les trois conditions précédentes s'écrivent alors sous la forme suivante :

- $H(x_i, x_j) > \alpha$
- $H(x_i^R, x_j^C) > \alpha$
- $H(x_i, x_j^R) > \alpha$

Le paramètre  $\alpha$  est une distance de Hamming minimale fixée par l'utilisateur. Il assure une différence minimale de  $\alpha$  bases entre les segments. Les 3 règles ci-dessus définissent le filtre de l'algorithme de génération de séquence détaillé dans le paragraphe suivant.

## 2.2 Algorithme de génération de séquence

L'algorithme de génération de séquences spécifiques a pour objectif de générer une séquence dont la configuration la plus probable est un PM. Pour optimiser la stabilité, cette séquence doit aussi posséder une valeur de  $\Delta G$  fixée. Cet algorithme est composé de deux étapes. La première étape consiste en la génération aléatoire d'une séquence en bases azotées et la seconde vérifie la spécificité de la séquence à travers les 3 règles définies.

Le filtre défini vérifie uniquement la spécificité de la séquence. La stabilité est caractérisée par deux paramètres  $\Delta G$ , la variation d'énergie libre de Gibbs, et  $l$  la longueur de la séquence inconnue. L'énergie  $\Delta G$  est calculée à partir du modèle du plus proche voisin dans l'approche populationnelle étudiée au chapitre 2. Ces deux paramètres sont liés. En effet, son calcul nécessite la connaissance de la séquence en bases azotées et dépend donc indirectement de la longueur  $l$  exprimée en bases azotées. Cette dernière étant inconnue, il est nécessaire de fixer l'un des deux paramètres pour notre algorithme. En fixant une valeur maximale de l'énergie, notée  $\Delta G_m$ , telle que l'énergie libre de la séquence inconnue soit inférieure à ce seuil, un encadrement de  $l$  peut aussi être proposé. Le choix d'une valeur pour  $l$  impose alors à l'algorithme la production d'une séquence de cette longueur.

L'intervalle estimé est de la forme  $[l_{max}; l_{min}]$  dont les bornes sont les longueurs des séquences maximale et minimale respectivement. En effet, comme il s'agit d'une configuration hybridée de type PM, l'énergie de la séquence est comprise dans l'intervalle défini par les énergies des deux séquences caractéristiques. Ces séquences ainsi que les expressions analytiques de leurs énergies en fonction de la longueur en bases azotées ont été étudiées au chapitre 2 et 3. D'après ces expressions (équations 3.5 et 3.4)  $l_{max}$  et  $l_{min}$  sont estimées à l'aide des équations 4.2 et 4.3 :

$$l_{max} = \frac{\Delta G_m + 3.1}{3.4} \quad (4.2)$$

$$l_{min} = \frac{\Delta G_m + 1.1}{1.2} \quad (4.3)$$

La séquence aléatoire produite par l'algorithme est donc de longueur  $l$  et d'énergie comprise dans l'intervalle  $[\Delta G_{min}^l, \Delta G_{max}^l]$  d'après les équations 3.4 et 3.5, avec  $\Delta G_{max}^l = \Delta G_m$ .

La deuxième étape est la vérification de la spécificité de la séquence à travers le filtre défini par les 3 règles. Celle-ci nécessite la détermination de deux paramètres :  $\alpha$  et  $\beta$ . Le paramètre  $\alpha$  est la distance de Hamming minimale pour la définition des 3 règles, et  $\beta$  est la longueur des segments à comparer en nombre de bases. Un schéma représentatif de cet algorithme est illustré dans la figure 4.2.

### 3 Analyse théorique d'une séquence générée

L'objectif de l'algorithme est de proposer des séquences pour lesquelles les configurations définies dans le filtre sont peu probables. La probabilité de formation des configurations est donnée à travers la valeur du  $\Delta G$  correspondant. Ainsi, les configurations dont la variation d'énergie de Gibbs est inférieure à  $4 \text{ Kcal/mol}$  sont considérées comme peu probables [Lee 94]. Quatre bases correspondent à cette énergie et le paramètre  $\alpha$  est donc initialisé à cette valeur.

A l'aide de la fonction de Hamming, des segments de longueur  $\beta$  sont comparés. Cette longueur peut être initialisée avec des valeurs proches de celle de persistance de l'ADN simple brin, estimée à 11 bases (pour une forme B de l'ADN) [Williams 01]. Cette longueur caractérise la plus petite unité d'un brin assimilable à un segment rigide<sup>1</sup>. Les séquences étant générées aléatoirement, l'optimisation du temps de calcul conduit à une sous estimation de cette longueur,

---

1. Chapitre 1, paragraphe 1.3.1

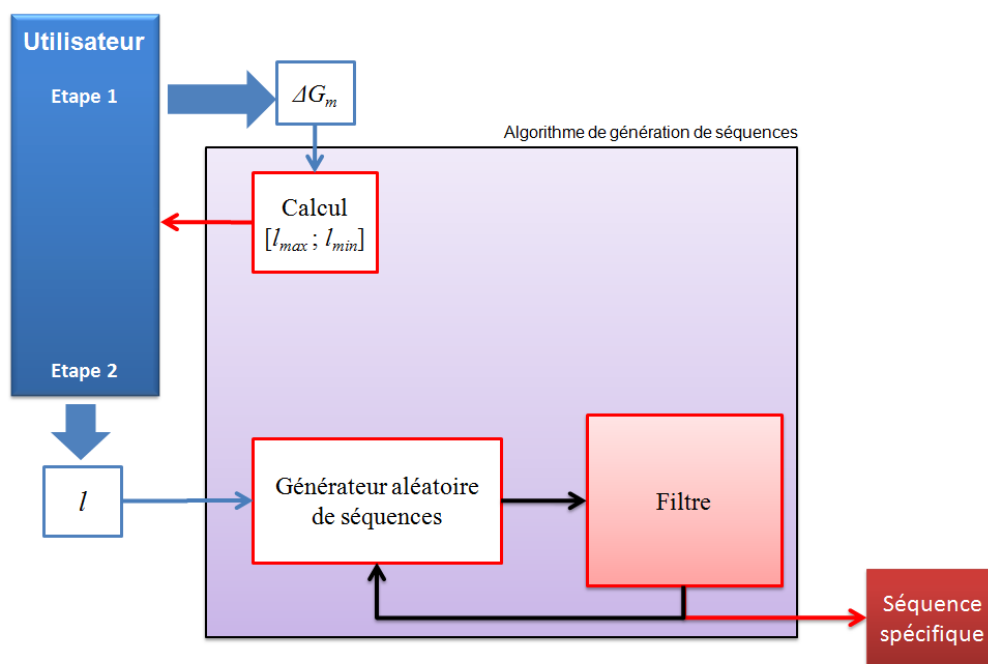


Figure 4.2 - Schéma représentatif de l'algorithme de génération de séquence.

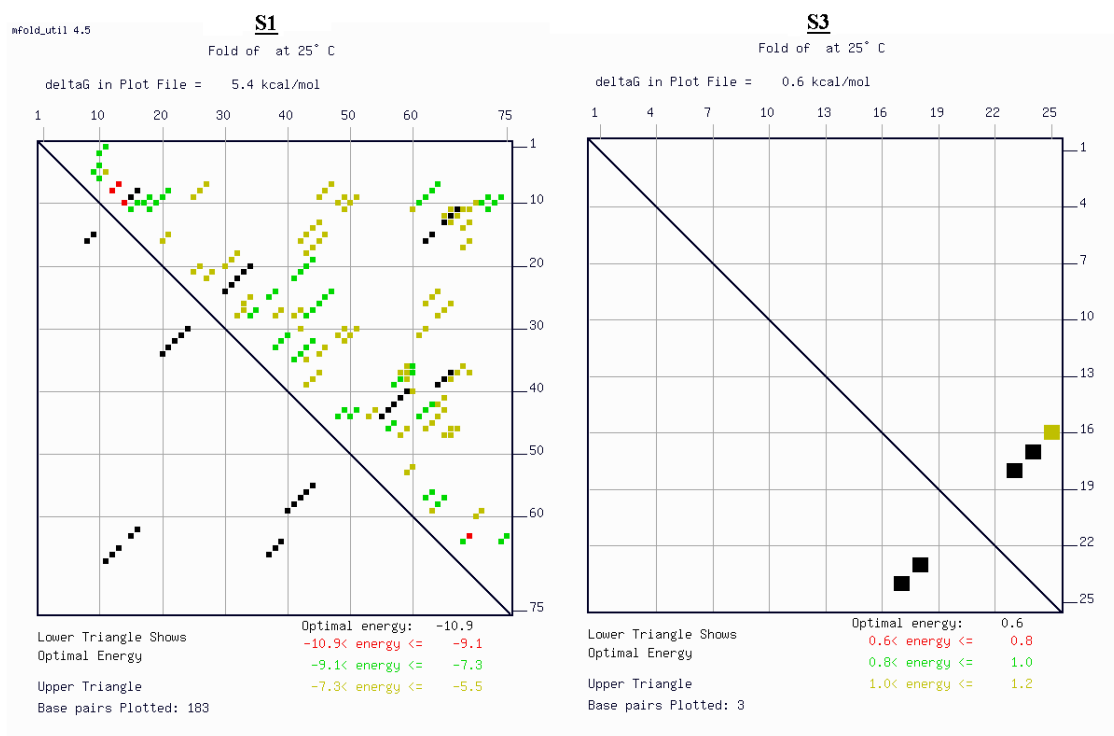
initialisée ainsi à 8 bases. Les choix de l'énergie et de la longueur de la séquence sont fixés en regard de la séquence spécifique recherchée. En effet, une meilleure adhésion est probable avec une énergie de configuration PM plus faible.

La probabilité de formation de MM augmente avec la longueur de séquence. Des séquences de longueurs faibles sont donc privilégiées. D'une part, par comparaison à la séquence  $S_1$  étudiée au chapitre 3, la longueur des séquences générées doit être inférieure. D'autre part, pour des raisons de détection, la caractérisation expérimentale des séquences impose un choix de longueur au minimum supérieure à  $10nm$ . Une séquence exemple de longueur 30 bases, notée  $S_3$  est ainsi produite à l'aide de cet algorithme. Les séquences en bases azotées de  $S_3$  et de sa séquence complémentaire  $S_4$  sont données dans l'Annexe A.

A travers une analyse par le logiciel *mFold*, une comparaison des configurations possibles entre la séquence spécifique  $S_3$  et l'ancienne  $S_1$  est donnée dans le tableau 4.2. Une configuration instable de type *épingle à cheveux* est possible pour la séquence  $S_3$  alors que les configurations de ce type sont stables dans leur ensemble pour la séquence  $S_1$  (figure 4.3). Pour les configurations de type *hétéro-complémentaire*, aucune configuration stable n'est possible pour la séquence proposée.

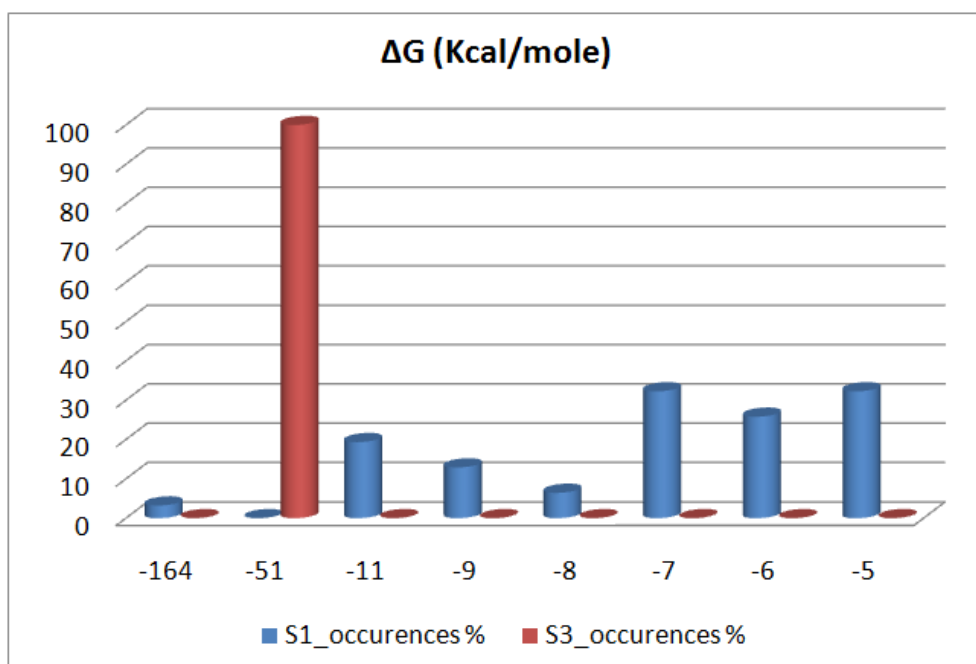
Avec une probabilité optimale pour la configuration PM, la séquence  $S_3$  semble plus spécifique que la séquence  $S_1$  initialement testée dans le projet Golem.

Afin de valider la spécificité de cette séquence, des expériences sont menées. La stabilité mesurée est analysée à l'aide de la méthode énergétique proposée au chapitre 3.

Figure 4.3 - Comparaison des configurations en épingle à cheveux des séquences  $S_1$  et  $S_3$ .

Propriétés	$S_1$	$S_3$
Hétéro-complémentaire	118	41
Épingle à cheveux	14	1
Hétéro-complémentaire (énergie $>4Kcal/mol$ )	27	0
Épingle à cheveux (énergie $>4Kcal/mol$ )	14	0
$\Delta G$ (PM)	164Kcal/mol	51 Kcal/mol

Tableau 4.2 - Tableau comparatif de  $S_1$  et  $S_3$ .



**Figure 4.4** - Comparaison des énergies des configurations hétéro-complémentaires et "Perfect match" stables des séquences  $S_1$  et  $S_3$ .

## 4 Analyse expérimentale

Des expériences sont réalisées dans le but de caractériser la spécificité d'une séquence produite par l'algorithme. Celles-ci sont effectuées à l'aide d'un AFM commercial du laboratoire AMIR de l'université d'Oldenburg (Allemagne). Les substrats utilisés possèdent les mêmes propriétés et sont toujours en silicium avec le même protocole (Annexe C).

Cette partie s'attache à caractériser les séquences  $S_3$  et sa complémentaire  $S_4$ . Une étude particulière est menée pour la spécificité à partir des forces de ruptures mesurées et de leur comparaison avec celles du chapitre 3. Ainsi, pour l'interaction spécifique, une distribution homogène des forces est attendue puisqu'une seule configuration stable est possible alors que pour les interaction non spécifique, une distribution non homogène est attendue.

Pour chaque expérience réalisée, le processus A/R est répété 10 fois sur des points différents du substrat, la répétabilité du processus étant déjà démontrée dans le précédent chapitre. Les expériences réalisées à Londres sont différentes des expériences détaillées dans ce présent chapitre car le matériel utilisé (AFM, poutres, température ambiante, bruit...) est dissemblable. L'objectif apparaît aussi dissemblable puisque ces expérimentations visent à montrer la spécificité de la nouvelle séquence et l'influence de deux paramètres expérimentaux la vitesse et le temps d'attente.

Le premier paragraphe rappelle les éléments constitutifs de ces expériences. La méthode énergétique est ensuite utilisée pour analyser la stabilité. L'influence des paramètres fondamentaux comme la vitesse est aussi étudiée. Enfin, un nouveau paramètre est introduit, le temps d'attente. Il correspond à un temps mesuré en secondes, introduit entre les deux étapes approche et retrait d'une mesure AFM. L'introduction de ce paramètre dans la caractérisation des interactions entre des brins d'ADN est un moyen d'étudier l'influence du temps sur la stabilité du

velcro spécifique.

## 4.1 Description des expériences

Sept expériences sont menées<sup>2</sup> : Exp1, Exp5, Exp10, Exp14, Exp16, Exp17, Exp13. Les paramètres expérimentaux de chaque expérience sont résumés dans le tableau 4.1

Quelques spécificités sont à noter :

- les expériences de contrôle Exp1 (sans fonctionnalisation) et Exp5 (poutre fonctionnalisée avec  $S_3$  et substrat non fonctionnalisé (NF)) sont réalisées à titre comparatif.
- les expériences Exp10, Exp16, Exp17 et Exp13 s'intéressent à l'interaction spécifique entre les séquences complémentaires  $S_3$  et  $S_4$ . Les différents paramètres expérimentaux, vitesse  $v$  et temps d'attente  $T_a$  varient.
- l'expérience Exp14 étudie les interactions d'une hybridation non spécifique entre  $S_1$  et  $S_3$ .
- Pour l'ensemble de ces expériences, le taux de chargement  $T_c$  est compris entre  $10^{-9}$  et  $10^{-8}$  N/s. Toutes les poutres appartiennent à un même lot.

---

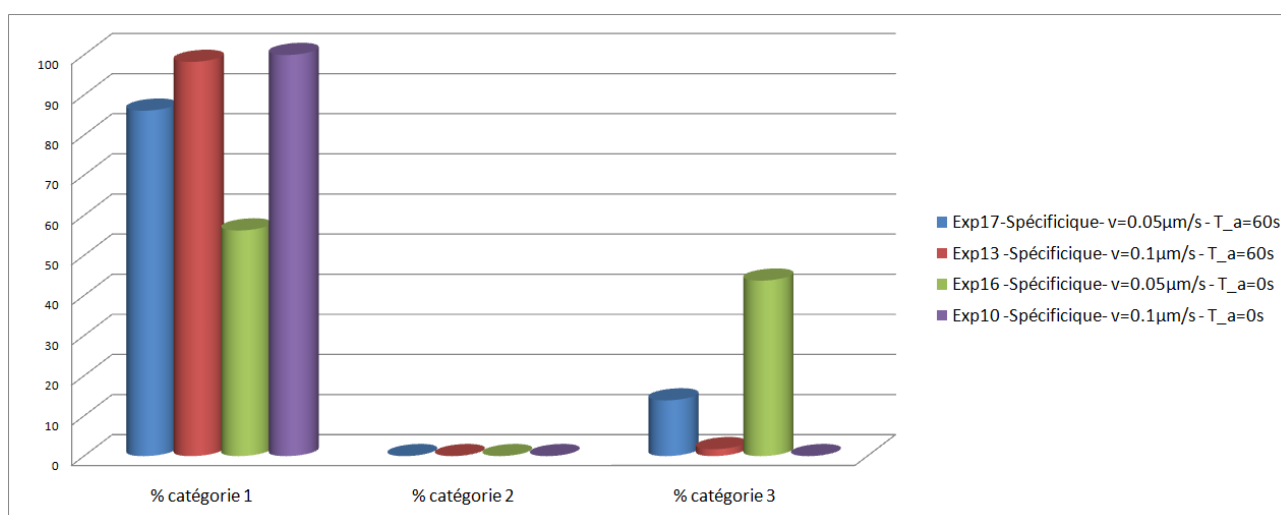
2. La classification originale est conservée pour corrélérer les graphiques issus des logiciels de traitement des données.

Tableau 4.3 - Tableau des paramètres expérimentaux pour l'étude de la spécificité.

Expérience	Exp1	Exp5	Exp14	Exp10	Exp16	Exp17	Exp13
Poutre	P	P	P-S3	P-S4	P-S4	P-S4	P-S4
Substrat	sub	sub-S3	sub-S1	sub-S3	sub-S3	sub-S3	sub-S3
Vitesse ( $\mu\text{m/s}$ )	0.1	0.1	0.1	0.1	0.05	0.05	0.1
Distance ( $\mu\text{m}$ )	1	1	1	1	1	1	1
Nb de répétitions	10	10	20	20	10	10	10
Nb de points	1	5	12	8	20	20	16
Milieux	PBS (1M)	PBS (1M)	PBS (1M)	PBS (1M)	PBS (1M)	PBS (1M)	PBS (1M)
k (N/m)	0.28	0.26	0.21	0.043	0.18	0.18	0.14
Tc (N/s)	$3 \cdot 10^{-8}$	$3 \cdot 10^{-8}$	$2 \cdot 10^{-8}$	$4 \cdot 10^{-9}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-8}$	$1.4 \cdot 10^{-8}$
Ta (s)	0	0	0	0	0	60	60

## 4.2 Analyse énergétique des événements de rupture locaux

Les forces de rupture et distance de rupture extraites à partir des expériences spécifiques sont utilisées pour évaluer l'énergie  $\Delta G_r$ . Ces valeurs énergétiques sont analysées à l'aide de la méthode d'analyse énergétique des événements de rupture locaux proposée dans le chapitre précédent. La répartition des points ( $\Delta G_r$  ;  $d_r$ ) pour chaque expérience sur les 3 catégories définies est effectuée et le résultat est représenté à la figure 4.5.



**Figure 4.5** - Répartition des événements de rupture locaux sur les trois catégories de la méthode d'analyse énergétique.

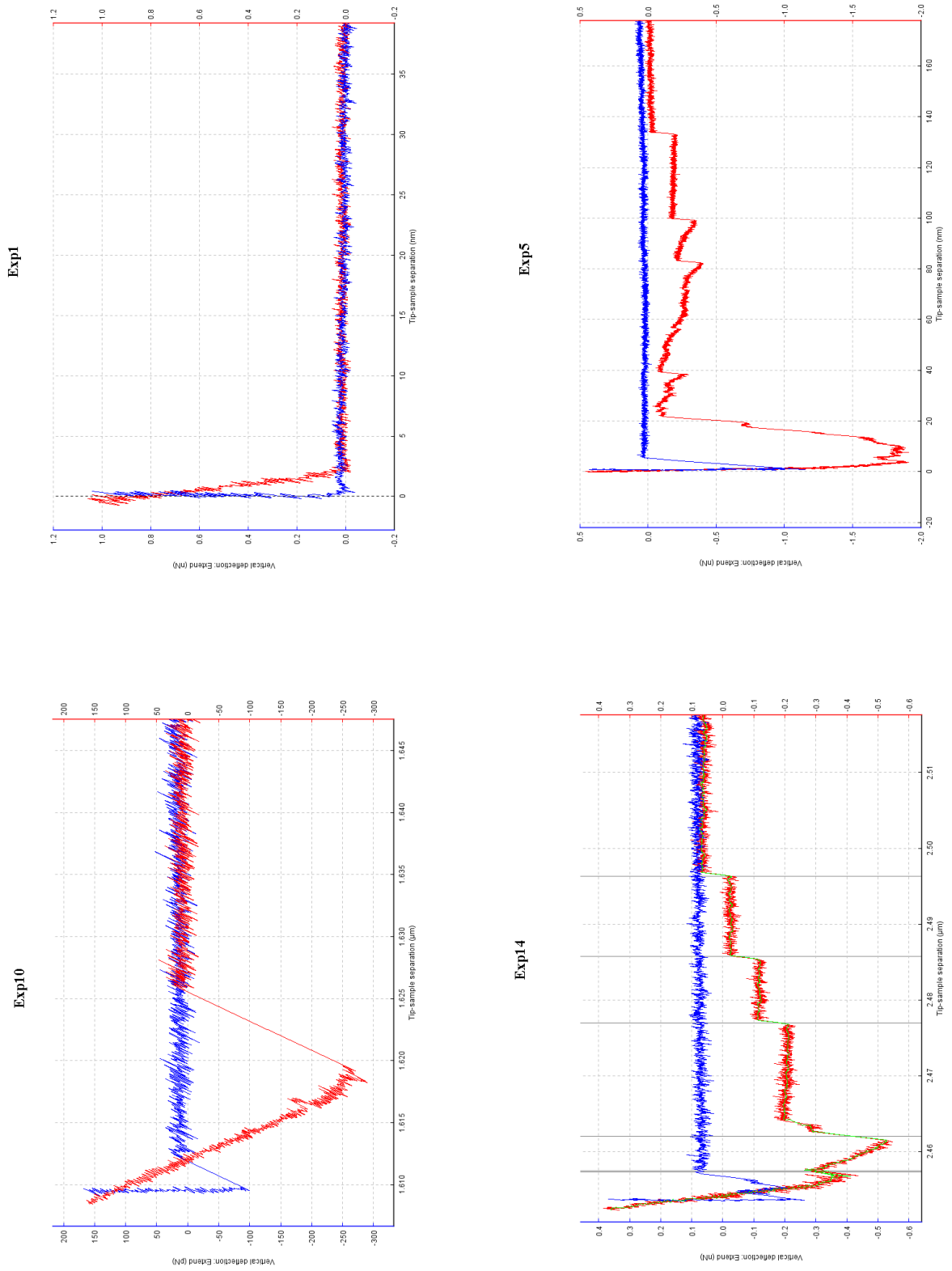
Pour la séquence optimisée, une majorité de points est positionnée dans la catégorie 1, optimale en terme de stabilité. Néanmoins, il n'est pas possible de conclure définitivement puisque le nombre de points est insuffisant pour un calcul précis statistique. Dans la suite de cette partie, la spécificité de la séquence générée est étudiée à travers une comparaison des forces et distances de rupture extraites des différentes expériences.

## 4.3 Analyse expérimentale de la spécificité

Les expériences Exp1 et Exp5 sont des expériences de contrôle des mesures, la première s'intéressant à l'interaction entre deux substrats non fonctionnalisés et la seconde étudiant une poutre non fonctionnalisée et un substrat fonctionnalisé avec  $S_3$ . Ces deux campagnes de mesures sont particulières puisque l'interaction mesurée pour la première est nulle et la seconde fait apparaître des paliers irréguliers pour des valeurs supérieures à  $2000pN$  (figure 4.6).



Figure 4.6 - Courbes A/R



Par ailleurs, un phénomène de paliers réguliers est observé sur toutes les expériences ( mis à part pour Exp10, réalisée avec une vitesse  $v = 0.1 \mu m/s$ ). Pour l'expérience non spécifique Exp14, le phénomène de paliers est quasi-systématique (figure 4.7). Le nombre majoritaire de paliers pour cette expérience est de 4 soit 60%.

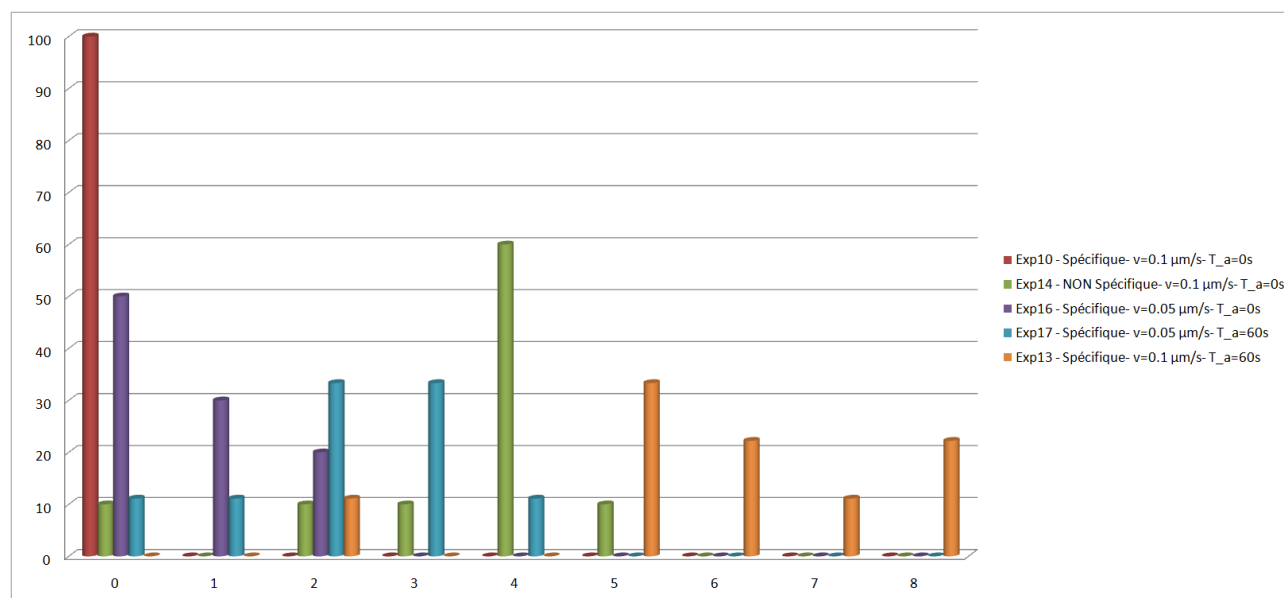


Figure 4.7 - Pourcentage des courbes en fonction du nombre de paliers

Pour Exp16 réalisée avec une vitesse  $v = 0.05 \mu m/s$ , 50% des courbes ne présentent aucun palier et les 50% complémentaires présentent entre 1 et 2 paliers. Pour Exp17, une majorité des courbes enregistrées (60%) présente entre 2 et 3 paliers. Cette expérience est réalisée dans les mêmes conditions que l'expérience Exp16 mais avec un temps d'attente de 60 secondes. Pour Exp13, le phénomène de paliers est systématique et une majorité des courbes (80%) comporte de 5 à 8 paliers.

L'apparition de ces paliers lors de la variation de la vitesse ou du temps d'attente implique la classification des données de chaque expérience par groupe en fonction des paliers apparues. L'étude de l'homogénéité des valeurs de force de rupture  $F_r$  et distance de rupture  $d_r$  en fonction de ces groupes est un moyen d'analyser les phénomènes caractéristiques de chaque expérience.

Ainsi, pour Exp14, 5 groupes sont distingués et notés respectivement  $Exp14\_pic1$ ,  $Exp14\_pic2$ ,  $Exp14\_pic3$ ,  $Exp14\_pic4$ ,  $Exp14\_pic5$ . De façon similaire, pour Exp16, 3 groupes, pour Exp13, 6 et pour Exp17, 4 groupes sont respectivement considérés.

#### 4.3.1 Etude de la spécificité

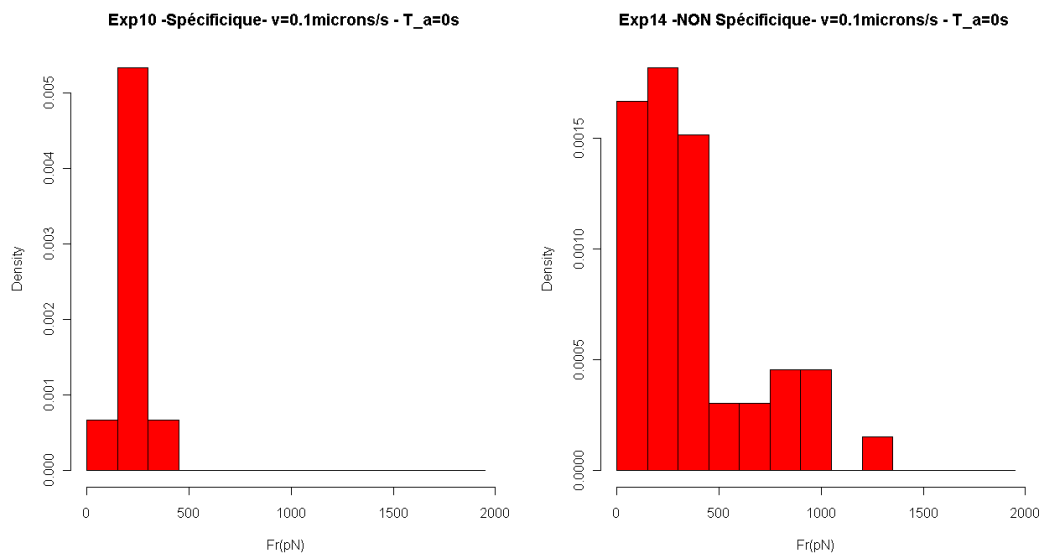
L'analyse de la spécificité se base sur la comparaison des forces de rupture et distances de rupture caractéristiques pour des interactions spécifiques d'une part, et non spécifiques d'autre part. Ainsi, les expériences réalisées dans les mêmes conditions environnementales (tableau 4.1), Exp10 et Exp14, sont comparées dans la suite.

Les forces de rupture mesurées dans Exp14 sont plus élevées que celle de Exp10. Les valeurs

les plus élevées correspondent à la force enregistrée lors du premier saut. Cette force correspond au groupe *Exp14\_pic1* dont la force de rupture moyenne est de  $793pN$ . Pour les deux groupes *Exp14\_pic2*, *Exp14\_pic3*, les forces de rupture moyennes sont de  $378pN$  et  $251pN$  respectivement. Ces valeurs restent plus élevées que la valeur moyenne estimée à  $219pN$  pour Exp10. Les forces de rupture de Exp10, sont naturellement plus faibles, puisque la séquence  $S_3$  est plus petite  $S_1$ , et que dans Exp14, des configurations complexes se mettent en place lors de l'hybridation. Ainsi, il ne s'agit pas de comparer les valeurs obtenues, mais les distributions des forces de rupture et distance de rupture pour chaque expérience.

Pour la distance de rupture, la valeur moyenne est  $2nm$  pour Exp10, et de  $7nm$  pour Exp14 (avec des valeurs moyennes pour les 3 premiers groupes,  $2nm$  pour *Exp14\_pic1*,  $4nm$  pour *Exp14\_pic2* et  $7nm$  pour *Exp14\_pic3*). La longueur de la séquence  $S_3$  est de  $10nm$ . Lorsque la distance de rupture mesurée est supérieure à cette valeur, l'interaction mesurée sur le palier correspond à un couple de brins en interaction différent du premier. Cependant, lorsque la distance de rupture est inférieure à cette valeur, il n'est pas possible de déduire le nombre de couples de brins en interaction. En conséquence, les interactions mesurées dans Exp14 peuvent correspondre indifféremment à un ou plusieurs couples de brins  $S_3-S_1$ .

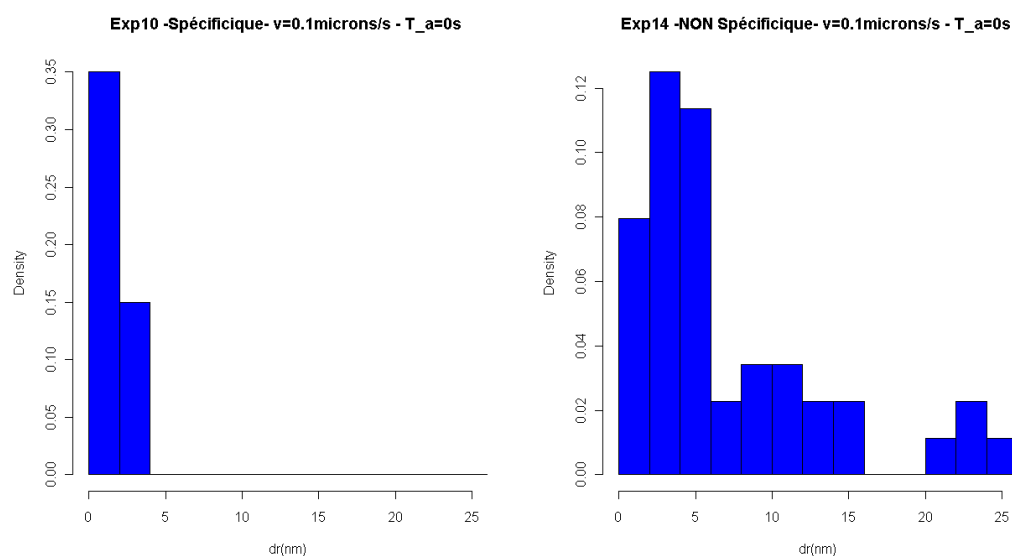
Les histogrammes des figures 4.9 et 4.8 montrent l'homogénéité des mesures de Exp10. L'écart type évalué pour la force de rupture (tableau 4.4) et la distance de rupture (tableau 4.5) montre plusieurs propriétés. La valeur moyenne pour Exp10 est significative et les mesures de Exp14 sont nettement plus hétérogènes. Même si les forces mesurées sont plus faibles dans l'expérience spécifique, ces deux expériences montrent que la séquence  $S_3$  est bien spécifique avec la répétabilité du phénomène.



**Figure 4.8** - Etude de la spécificité. Comparaison de la distribution des forces de rupture mesurées pour les expériences Exp10 et Exp14.

### 4.3.2 Influence de la vitesse sur la spécificité et la stabilité

Dans le but d'étudier l'influence de la vitesse sur la réponse moléculaire de  $S_3$  et  $S_4$ , l'expérience Exp16 est réalisée dans les mêmes conditions environnementales que Exp10 (Tableau



**Figure 4.9** - *Etude de la spécificité. Comparaison de la distribution des distances de rupture mesurées pour les expériences Exp10 et Exp14.*

Expérience	Ecart type ( $F_r$ )
Exp10	59
Exp14_pic1	244
Exp14_pic2	136
Exp14_pic3	93

**Tableau 4.4** - *Etude de la spécificité. Tableau comparatif des écarts types pour la variable  $F_r$  des expériences Exp10 et Exp14.*

Expérience	Ecart type ( $d_r$ )
Exp10	1.1
Exp14_pic1	1.6
Exp14_pic2	2.3
Exp14_pic3	5.7

**Tableau 4.5** - *Etude de la spécificité. Tableau comparatif des écarts types pour la variable  $d_r$  des expériences Exp10 et Exp14.*

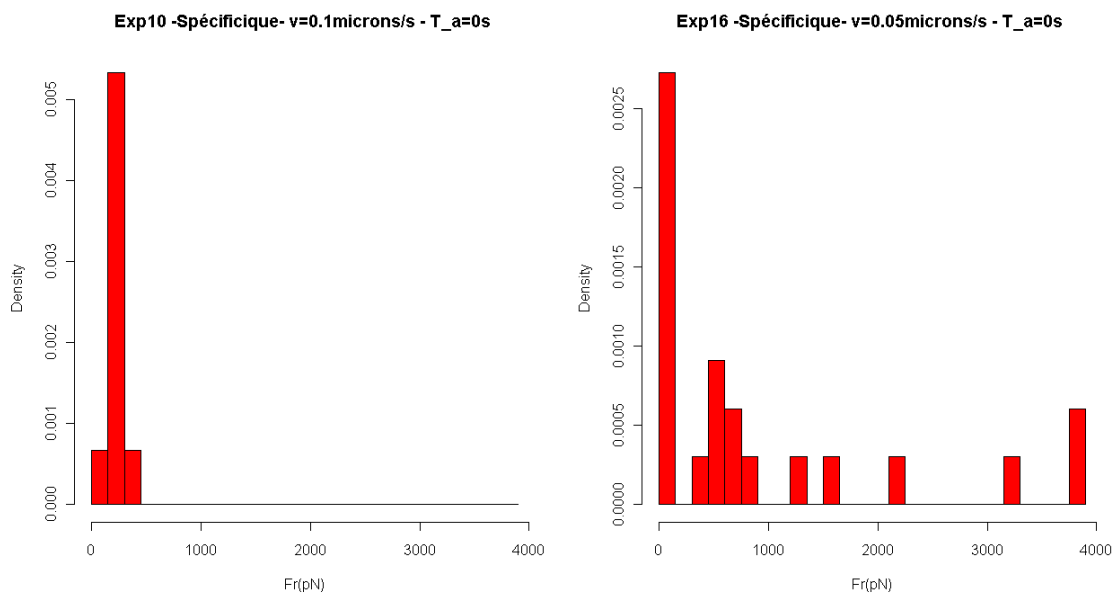
4.1) mais avec une vitesse plus faible  $0.05\mu\text{m/s}$  et une raideur de poutre plus grande pour garantir un taux de chargement du même ordre de grandeur. Cette analyse comparative prend en compte les distributions des forces de rupture et des distances de rupture ainsi que leur amplitude.

Concernant la force de rupture, les moyennes calculées pour les différents groupes de Exp16 sont largement supérieures à la moyenne évaluée pour Exp10. De plus, les moyennes de  $F_r$  pour les différents groupes de Exp16 sont décroissantes :  $1510\text{pN}$ ,  $456\text{pN}$  et  $81\text{pN}$ . Une interaction plus forte est donc mesurée pour Exp16.

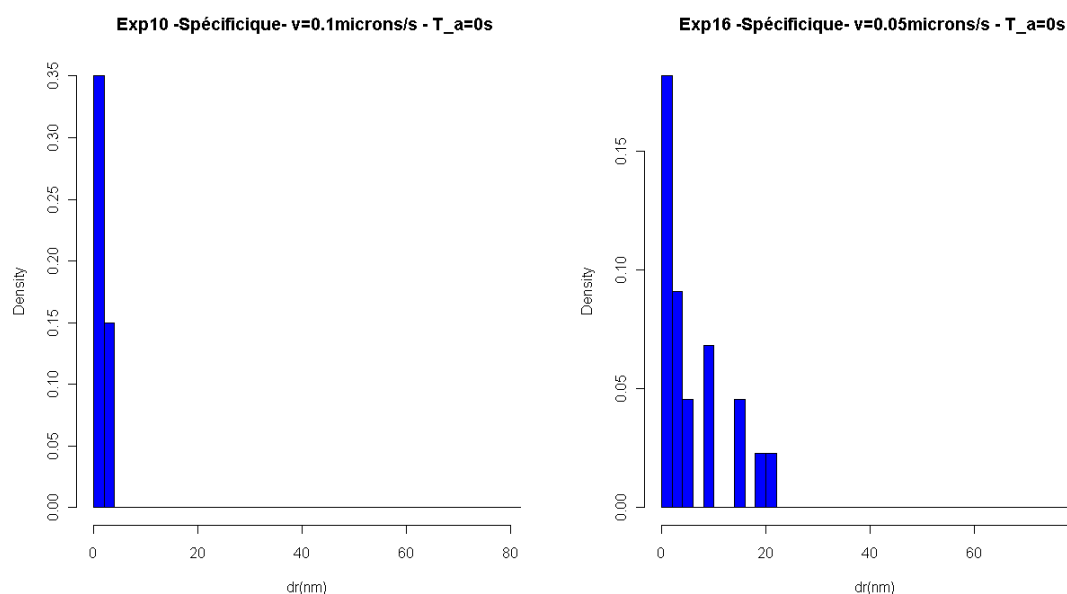
Pour la distance de rupture, les valeurs mesurées sont largement supérieures et hétérogènes dans l'expérience Exp16 (figure 4.11). Les moyennes des différents groupes sont 4, 8 et 9 nm. Celles-ci sont croissantes et sont proches de la valeur caractéristique  $10\text{nm}$ . Il semble que les interactions mesurées pour chaque saut pour Exp16 correspondent à un couple de brins en interaction différent.

L'histogramme de la distribution de la force de rupture pour l'expérience à vitesse faible, Exp16, montre une répartition en trois groupes principaux centrés sur des valeurs proches des valeurs moyennes des groupes (figure 4.10). Il est possible que la décroissance de ces forces de rupture soit due à un détachement successif des brins hybridés pendant la phase approche.

Une vitesse faible améliore la stabilité globale du velcro en faisant intervenir un nombre de brins plus important. En effet, avec une vitesse plus faible, l'analyse montre que les brins d'ADN semblent s'hybrider plus souvent avec une adhésion plus forte.



**Figure 4.10** - Comparaison de la distribution des forces de rupture mesurées pour les expériences Exp10 et Exp16.



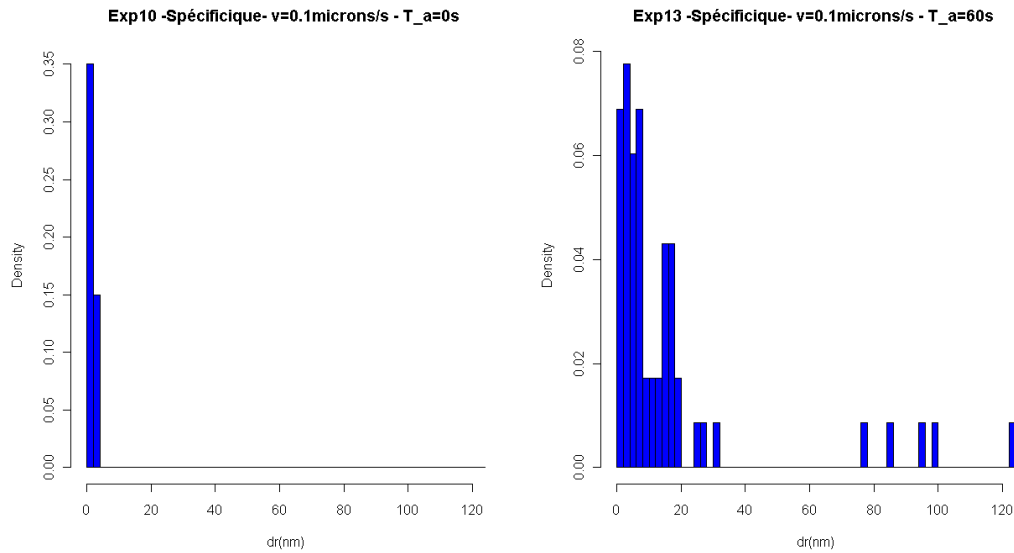
**Figure 4.11** - Comparaison de la distribution des distances de rupture mesurées pour les expériences Exp10 et Exp16.

### 4.3.3 Influence du temps d'attente

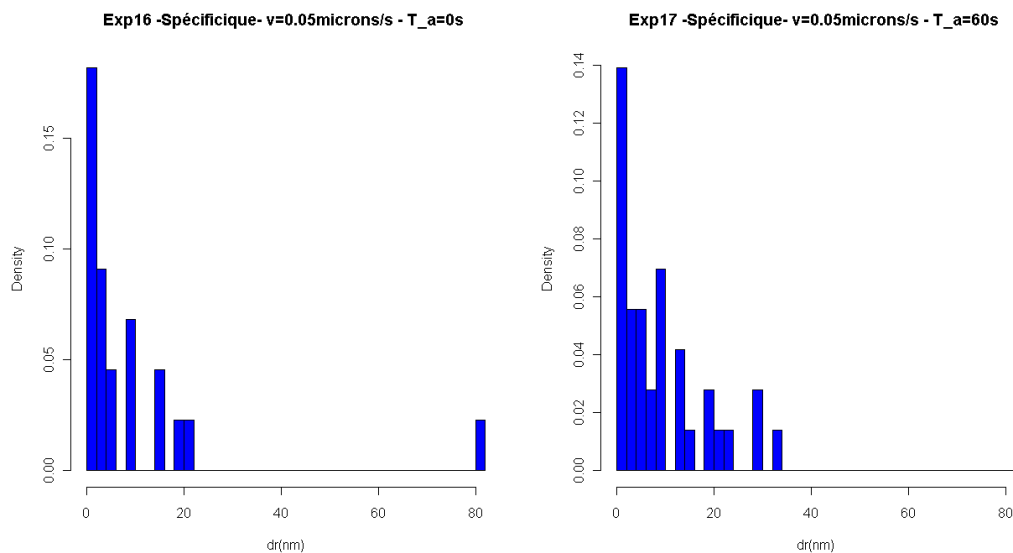
Dans le paragraphe précédent, les expériences ont montré que la vitesse influence la stabilité globale du velcro en faisant intervenir un nombre de brins plus important. Nous pouvons donc supposer que l'hybridation nécessite un certain temps pour la réalisation de l'assemblage entre les brins d'ADN. Dans le but d'étudier l'influence du temps sur l'interaction  $S_3$ - $S_4$ , les expériences Exp13 et Exp17 ont été réalisées. Il s'agit dans ces expériences d'introduire un temps mesuré en secondes appelé temps d'attente  $T_a$  entre les deux étapes approche et retrait d'une mesure AFM. Dans ce paragraphe, il ne s'agit pas de déterminer le temps nécessaire à l'hybridation, mais de relever si l'influence de ce paramètre est caractéristique.

Cette étude est menée à travers deux comparaisons. La première comparaison a lieu entre Exp10 et Exp13. Ces expériences sont réalisées à la même vitesse  $v = 0.1 \mu\text{m/s}$ . La seconde comparaison a lieu entre Exp16 et Exp17. Ces expériences sont réalisées à la même vitesse  $v = 0.01 \mu\text{m/s}$ . Les interactions mesurées dans les expériences avec l'introduction du temps d'attente de 60s entre les phases approche et retrait, Exp 13 et Exp17 correspondent aussi à des couples de brins d'ADN différents en raison des valeurs élevées mesurées pour la distance d'interaction (figures 4.12 et 4.13).

Pour une vitesse  $v = 0.1 \mu\text{m/s}$ , et un temps d'attente de 60s, le nombre de paliers devient très élevé, et les valeurs de  $F_r$  mesurées beaucoup plus grandes que celles mesurées pour Exp10 avec un temps d'attente nul. De plus, les valeurs mesurées pour les groupes *Exp13\_pic4*, *Exp13\_pic5* et *Exp13\_pic6* sont comprises entre 425 et 540 pN et restent plus élevées que la valeur moyenne estimée pour Exp10 à 219 pN. L'introduction d'un temps d'attente à cette vitesse, implique donc un nombre de brins plus important dans l'adhésion entre les deux populations de brins complémentaires. Il s'agit ici d'une séparation progressive des brins hybridés pendant la phase approche.



**Figure 4.12** - Etude du temps d'attente pour  $v = 0.1 \mu\text{m/s}$ . Comparaison des distributions des distances de rupture mesurées pour les expériences  $\text{Exp10}(T_a=0\text{s})$  et  $\text{Exp13}(T_a=60\text{s})$ .



**Figure 4.13** - Etude du temps d'attente pour  $v = 0.05 \mu\text{m/s}$ . Comparaison des distributions des distances de rupture mesurées pour les expériences  $\text{Exp16}(T_a=0\text{s})$  et  $\text{Exp17}(T_a=60\text{s})$ .

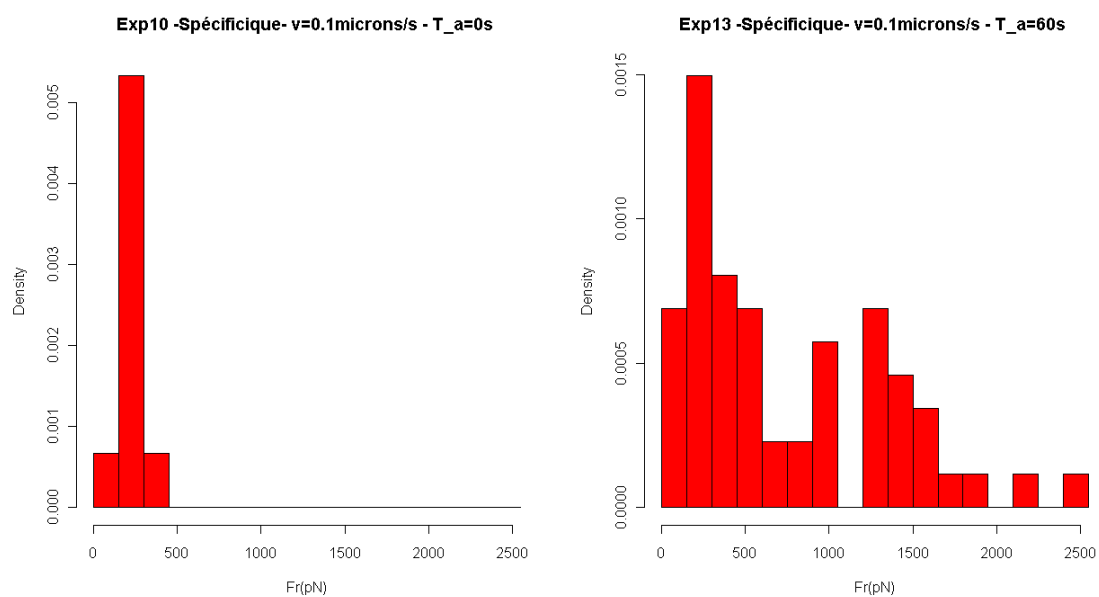


Figure 4.14 - Etude du temps d'attente pour  $v = 0.1 \mu\text{m/s}$ . Comparaison des distributions des forces de rupture mesurées pour les expériences  $\text{Exp10}(T_a=0\text{s})$  et  $\text{Exp13}(T_a=60\text{s})$ .

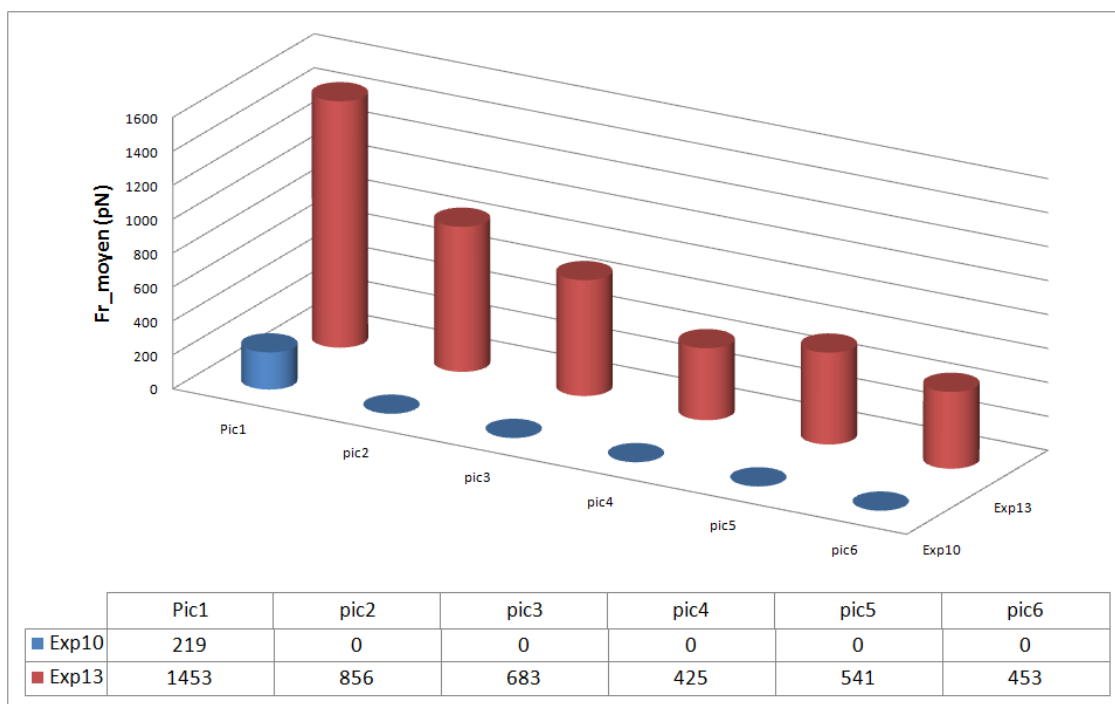


Figure 4.15 - Etude du temps d'attente pour  $v = 0.1 \mu\text{m/s}$ . Comparaison des forces de rupture moyennes par groupe pour les expériences  $\text{Exp10}(T_a=0\text{s})$  et  $\text{Exp13}(T_a=60\text{s})$ .



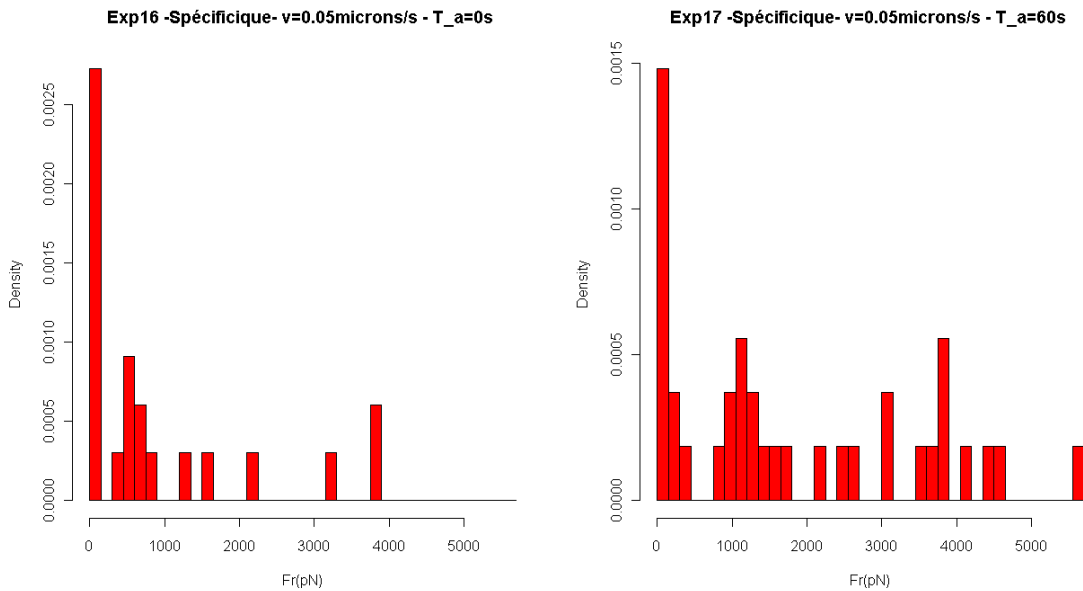
Pour une vitesse  $v = 0.05 \mu\text{m/s}$ , et un temps d'attente de 60s, dans l'expérience Exp17, globalement le nombre de paliers ne varie pas comparativement à l'expérience Exp16 avec un temps d'attente nul, mais les valeurs de  $F_r$  mesurées sont beaucoup plus grandes lorsque le temps d'attente est introduit (figure 4.16). Ces valeurs restent en moyenne plus élevées dans l'expérience avec le temps d'attente pour chaque groupe (figure 4.17). Ici, encore, l'introduction d'un temps d'attente à une vitesse constante, implique un nombre de brins plus important dans l'adhésion entre les deux populations de brins complémentaires. Il s'agit ici d'une séparation progressive des brins hybridés pendant la phase approche.

Enfin, la comparaison entre l'expérience Exp17 et Exp13 pour lesquelles le temps d'attente est de 60s, mais la vitesse est plus faible pour Exp17, montre que le nombre de paliers est moins important dans l'expérience à vitesse faible, mais les valeurs de  $F_r$  sont nettement plus élevées.

### Conclusion :

Le temps d'attente influence considérablement la stabilité globale du velcro. En effet, le nombre de couples de brins en interaction augmente et l'intensité des forces de rupture mesurées sont optimisés par l'introduction de ce paramètre.

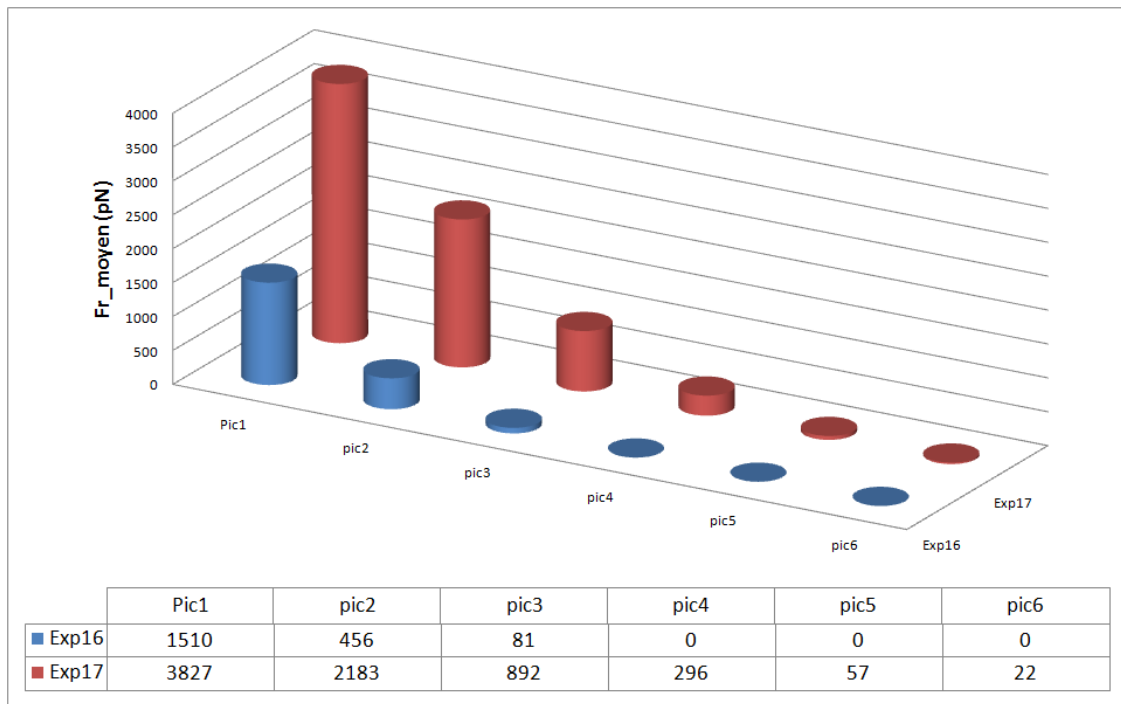
Néanmoins, l'influence de ce paramètre dépend de la vitesse d'A/R. En effet, lorsque la vitesse est importante ( $0.1 \mu\text{m/s}$ ), la stabilité du velcro est optimisée grâce à l'augmentation du nombre de brins en interaction, et lorsque la vitesse est plus faible ( $0.05 \mu\text{m/s}$ ), la stabilité du velcro est optimisée grâce à l'augmentation de l'intensité des interactions.



**Figure 4.16** - Etude du temps d'attente pour  $v = 0.05 \mu\text{m/s}$ . Comparaison des distributions des forces de rupture mesurées pour les expériences Exp16 ( $T_a = 0\text{s}$ ) et Exp17 ( $T_a = 60\text{s}$ ).

## 4.4 Conclusion

Comparativement aux expériences non spécifiques, la distribution des mesures pour l'expérience spécifique est homogène. La variation de l'interaction est faible et la spécificité de la



**Figure 4.17** - Etude du temps d'attente pour  $v = 0.05 \mu\text{m/s}$ . Comparaison des forces de rupture moyennes par groupe pour les expériences  $\text{Exp16}(T_a = 0\text{s})$  et  $\text{Exp17}(T_a = 60\text{s})$ .

séquence  $S_3$  est prouvée par cette campagne de mesures.

La variation des paramètres *vitesse* et *temps d'attente* influence la stabilité globale du velcro. Pour une vitesse faible, un nombre d'interactions plus important est observé à travers le phénomène de paliers. Ces interactions correspondent à des couples de brins différents. De plus, l'intensité des interactions est largement supérieure. La stabilité globale du velcro est optimisée pour une vitesse plus faible. Un autre paraît fondamental et directement corrélé à ce dernier, le temps d'attente. Ainsi, l'introduction d'un temps d'attente améliore aussi l'adhésion des systèmes mis en regard lors des expériences.

L'étude de l'influence des paramètres vitesse et temps d'attente est réalisée à travers des comparaisons des amplitudes et des distributions pour les forces et les distances de rupture. La réponse homogène de la séquence  $S_3$  due à sa spécificité a contribué à simplifier cette analyse et à ouvrir de nouvelles voies d'analyse.

## 5 Conclusions

Pour répondre à la problématique de la spécificité, un algorithme de génération de séquences est proposé dans ce chapitre. Cet algorithme est le cœur d'un outil logiciel de conception de séquences spécifiques. Un exemple est ensuite proposé pour illustrer notre propos. Cette nouvelle séquence, notée  $S_3$ , est caractérisée expérimentalement à partir des outils d'analyse mis en place dans le chapitre 3 de ce document.

L'analyse montre que cette séquence est plus intéressante du point de vue de sa spécificité

pour résoudre le problème de l'auto assemblage. Elle met aussi en avant deux paramètres clef, la vitesse d'approche/retrait et le temps d'attente. Des hypothèses sont ensuite formulées pour établir à la fois la séquence la plus pertinente et des valeurs des paramètres optimisant la potentialité de la séquence.

Ce chapitre ouvre de nouvelles perspectives pour le choix de l'ADN comme velcro entre deux micro-objets à assembler. Des outils d'études et d'analyse expérimentale sont mis en place pour proposer des nouvelles solutions à ce problème, une séquence adaptée et des paramètres de contrôle optimisés. Cette problématique de l'auto-assemblage bioinspiré par l'ADN semble ainsi prometteuse scientifiquement.

---

## CONCLUSION ET PERSPECTIVES

---

L'auto-assemblage des composants à l'échelle microscopique est une approche novatrice pour l'usage généralisé et l'industrialisation de systèmes émergents basés sur les micro- et nanotechnologies. Le projet européen GOLEM constitue un premier pas dans cette direction en explorant les phénomènes d'origine biomoléculaire pouvant être exploités pour un assemblage massif, parallèle et programmable.

Les travaux multidisciplinaires présentés ici s'inscrivent dans les objectifs de ce projet et visent particulièrement le choix d'un processus biologique, son étude théorique et sa validation expérimentale.

Un état de l'art présenté dans le premier chapitre distingue dans la littérature en auto-assemblage plusieurs approches, en fonction de l'échelle, micro ou nanoscopique, et des principes de bases, inorganique et biologique. Notamment, il apparaît que ce dernier est exclusivement utilisé à l'échelle nanoscopique et qu'il existe plusieurs processus exploitables à l'échelle microscopiques. Parmi ces processus, l'hybridation de l'ADN semble être une solution idéale car elle fournit, à travers la séquence en bases azotées des brins, la stabilité et la spécificité nécessaires à tout auto-assemblage.

Le deuxième chapitre propose une étude théorique plus approfondie de l'hybridation de l'ADN, à travers une approche multi-échelle. Des outils de modélisation de la dynamique moléculaire pour des brins individuels et une approche thermodynamique pour une population de brins sur une surface fonctionnalisée sont décrits. Ce chapitre met en place aussi une approche mécanique, pour relier des informations exprimées en termes d'énergie à des résultats de mesures par la microscopie à force atomique (AFM) exprimés en termes de force.

Le troisième chapitre est dédié à l'analyse expérimentale de la stabilité d'une adhésion par hybridation. Les expériences réalisées en collaboration avec le laboratoire NPL à Londres sont décrites et analysées. Notamment, une analyse statistique définit les données significatives en fonction des paramètres expérimentaux et établit une valeur de  $39 \pm 4 pN$  comme force de rup-

ture moyenne entre deux brins hybridés. Néanmoins, la grande variabilité des données récoltées rend difficile la classification et la quantification du comportement de l'ADN, par rapport aux propriétés de stabilité et de spécificité. Une méthode de classification basée sur la comparaison des mesures expérimentales et du modèle thermodynamique est alors proposée pour quantifier la stabilité d'un assemblage par ADN. Cette analyse fait apparaître clairement la corrélation entre la spécificité et la stabilité à travers la séquence en bases azotées. Une optimisation de ces deux propriétés nécessite donc un travail dédié à la conception de séquence d'ADN.

Le quatrième chapitre met en place une méthodologie pour concevoir des séquences optimisées pour l'auto-assemblage. L'idée directrice est de diminuer la probabilité des mésappariements possibles et d'assurer une hybridation en double hélice parfaite. Un état de l'art des méthodes de conception de séquence est présenté pour enfin proposer un algorithme de génération de séquence dédié. Cet algorithme prend en entrée une valeur d'énergie d'interaction désirée, calcule la longueur correspondante et la séquence.

Une séquence générée par cette méthode est ensuite analysée afin de valider la faible probabilité des mésappariements. Une campagne de mesures AFM est ensuite effectuée en collaboration cette fois avec le laboratoire AMIR en Allemagne. Ces expériences démontrent une meilleure stabilité et spécificité de la séquence conçue et valident l'approche proposée dans cette thèse.

Les perspectives de ce travail sont nombreuses et dépassent le cadre de l'auto-assemblage microscopique. L'intégration des fonctions biologiques sur des micro et nanocomposants ouvre une nouvelle voie pour l'implémentation des propriétés d'actionnement, de détection et d'interaction contrôlés. Ces fonctions existantes dans la nature sont compliqués à reproduire en utilisant uniquement de la matière inorganique, alors qu'elles font partie des processus fondamentaux de la biologie. Par exemple, des nanorobots pour des applications in-vivo pourraient être dotés de capacités d'interaction biologique contrôlée pour transporter des médicaments ou effectuer des diagnostics.





---

# ANNEXE A

---

## Les séquences en bases azotées

Les terminaisons 5' des séquences étudiées sont modifiées chimiquement avec un groupement Amine C6 179.16. Les séquences complémentaires  $S_1$  et  $S_2$  sont proposées dans le projet Golem. Les séquences complémentaires  $S_3$  et  $S_4$  sont générées par l'algorithme de génération de séquences.

### 1 Les séquences $S_1$ et $S_2$

Longueur	75 bases, soit 25.4 nm pour une double hélice en forme B.
$T_m$	75.9°C.
$S_1$ (5'-3')	CAA ATA CCG TGG GAC GAC ACG CAC CGG CAG TGC GCA GGC AGC GTC GGA CAC AAC ACG CTT ACG GCC CTC AAC ACT
$S_2$ (5'-3')	AGT GTT GAG GGC CGT AAG CGT GTT GTG TCC GAC GCT GCC TGC GCA CTG CCG GTG CGT GTC GTC CCA CGG TAT TTG

### 2 Les séquences $S_3$ et $S_4$

Longueur	30 bases, soit 10.2 nm pour une double hélice en forme B.
$T_m$	55.2°C.
$S_3$ (5'-3')	TTA ACA ATC CTT CTC TAT TCA AAT ACC ACC
$S_4$ (5'-3')	GGT GGT ATT TGA ATA GAG AAG GAT TGT TAA





---

# ANNEXE B

---

## Protocole de fonctionnalisation *non-bloquant* : NB

Les substrats et les poutres utilisées sont en silicium.

- Préparer une solution [35g NaOH, 140ml H<sub>2</sub>O and 240 ml EtOH]. Ajouter NaOH dans l'eau progressivement. Lorsque la totalité de NaOH est dissoute, ajouter EtOH. Une grande précaution est requise pour la manipulation de cette solution !
- Déposer les poutres et les substrats dans la solution pendant 2 heures.
- Rincer une fois dans l'eau, et deux fois dans EtOH.
- Déposer les poutres et les substrats dans 1% GOPS dans 95% EtOH, 5% eau pendant une heure.
- Rincer trois fois dans EtOH.
- Les brins d'ADN doivent être dilués dans une solution PBS avec une concentration final de 20  $\mu M$ . Laisser cette solution 30 min à température ambiante.
- Déposer des volumes de 20  $\mu L$  de cette solution sur les poutres et les substrats. Déposer la préparation dans une chambre humide pendant 30 minutes.
- Rincer une fois avec une solution PBS, et deux fois dans l'eau.



---

# ANNEXE C

---

## Protocole de fonctionnalisation *bloquant* : B

Les substrats et les poutres utilisées sont en silicium.

- Préparer une solution [35g NaOH, 140ml H<sub>2</sub>O and 240 ml EtOH]. Ajouter NaOH dans l'eau progressivement. Lorsque la totalité de NaOH est dissoute, ajouter EtOH. Une grande précaution est requise pour la manipulation de cette solution !
- Déposer les poutres et les substrats dans la solution pendant 2 heures.
- Rincer une fois dans l'eau, et deux fois dans EtOH.
- Déposer les poutres et les substrats dans 1% GOPS dans 95% EtOH, 5% eau pendant une heure.
- Rincer trois fois dans EtOH.
- Les brins d'ADN doivent être dilués dans une solution PBS avec une concentration final de 20  $\mu$ M. Laisser cette solution 30 min à température ambiante.
- Déposer des volumes de 20  $\mu$ L de cette solution sur les poutres et les substrats. Déposer la
- préparation dans une chambre humide pendant 30 minutes.
- Rincer une fois avec une solution PBS, et deux fois dans l'eau.
- Préparer la solution bloquante avec 50 mM Ethanolamine, 0.1% Tris buffer.
- Déposer les poutres et les substrats dans la solution bloquante 15 minutes à température 50 degC.
- Rincer trois fois dans l'eau.



---

## LISTE DES SYMBOLES

---

$\alpha$	distance de Hamming minimale estimée en nombre de bases azotées.
$\beta$	longueur d'un segment d'un brin d'ADN dans l'algorithme de génération de séquences estimée en nombre de bases azotées.
$d$	Déplacement de la poutre pendant une mesure AFM.
$d_{base}$	Distance entre deux Bases consécutives d'un brin.
$d_r$	Distance de Rupture.
$\Delta G$	énergie libre de Gibbs estimé en <i>Kcal/mol</i> .
$\Delta G_{max}^n$	énergie de Gibbs de la séquence maximale $(GC)_{n/2}$ estimé en <i>Kcal/mol</i> .
$\Delta G_{min}^n$	énergie de Gibbs de la séquence minimale $(AT)_{n/2}$ estimé en <i>Kcal/mol</i> .
$\Delta G_r$	énergie expérimentale estimée à partir de $F_r$ et $d_r$ .
$\Delta H$	enthalpie libre estimée en <i>Kcal/mol</i> .
$\Delta S$	entropie estimée en <i>Kcal/mol</i> .
$\Delta x$	Déplacement sur l'axe de la double hélice.
$E$	Energie d'un système moléculaire.
$E_c$	Energie Cinétique.
$E_p$	Energie Potentielle.
$\varepsilon$	point d'une courbe approche/retrait
$F$	Force d'interaction.
$F_r$	Force de Rupture.
$f_\varepsilon$	Force mesurée au point $\varepsilon$ d'une courbe approche/retrait.
$H$	opérateur Hamiltonien.
$H(m_1, m_1)$	distance de Hamming entre deux mots $m_1$ et $m_2$ .
$k$	raideur d'une poutre AFM estimée en <i>N/m</i> .
$l$	Longueur d'un brin d'ADN estimée en nombre de bases azotées.

---

$l_s$	Longueur d'un Segment sur un brins d'ADN estimée en nombre de bases azotées.
$n$	Nombre de bases azotées d'une séquence.
$N$	Nombre de phosphates sur un brin d'ADN.
$[Na^+]$	concentration saline en ion sodium.
$p$	vecteur Position du centre de masse d'une molécule.
$p_e$	vecteur Position aléatoire autour du centre de masse d'une molécule.
$\Psi$	fonction d'onde.
$T$	Température.
$T_a$	Temps d'Attente estimé en secondes.
$T_m$	Température de fusion de l'ADN simple brins dans les puces à ADN.
$T_c$	Taux de Chargement estimé en $N/s$ .
$v$	Vitesse d'approche/retrait pendant une mesure AFM.
$V$	énergie d'interaction d'un système moléculaire.
$[V_\lambda]_e$	Variance d'une courbe approche/retrait $\lambda$ d'une expérience $e$ .
$z_p$	flexion de la poutre pendant une mesure AFM.

---

## LISTE DES ABRÉVIATIONS

---

A/R	Approche/Retrait.
B	protocole de fonctionnalisation Bloquant.
DFS	Spectroscopie à Force Dynamique.
<i>e</i>	Expérience.
MM	configuration "Mis-Match".
NB	protocole de fonctionnalisation Non Bloquant.
$\Omega_{configs}$	ensemble des valeurs de $\Delta G$ pour les configurations possibles pour un couple de brins d'ADN.
PM	configuration "Perfect Match".
s	mode de mesure AFM par balayage.
$\lambda$	courbe approche/retrait.





---

---

## LISTE DES PUBLICATIONS

---

Abbaci, A. and Haliyo, S. and Régnier, S. (2008). DNA as building block for self-assembly of micro-components.

IEEE Second International Conference on Quantum, Nano and Micro Technologies. Pages 28-33.

Abbaci, A and Daunay, B. and Haliyo, S. and Régnier, S. and Boyd, R. and Cuenat, A. (2008). Characterization of DNA Bio-bonds for Meso-Scale Self-assembly.

30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'08).

Daunay, B. and Abbaci, A. and Micaelli, A. and Régnier, S. (2006). The wave variables, a solution for stable haptic feedback in molecular docking simulations.

IEEE CISSE / IETA 2006, International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering. Pages 67-73.



---

## BIBLIOGRAPHIE

---

- [Adleman 94] **L.M. Adleman.** *Molecular computation of solutions to combinatorial problems.* Science, 1994.
- [Alvo 10] **S. Alvo, P. Lambert, M. Gauthier et Régnier S.** *A van der Waals Force Based Adhesion Model for Micromanipulation.* Journal of Adhesion Science and Technology, 2010.
- [Arita 02] **Masanori Arita.** *"DNA Sequence Design Using Templates."*, 2002.
- [Bashir 01] **R. Bashir.** *Invited Review, DNA-mediated artificial nanobiostructures : state of the art and future directions.* Superlattices and Microstructures, vol. 29, n° 1, pages 1–16, 2001.
- [Beger 98] **Richard D. Beger et Philip H. Bolton.** *Structures of Apurinic and Apyrimidinic Sites in Duplex DNAs.* J. Biol. Chem., vol. 273, pages 15565–15573, 1998.
- [Bell 78] **GI Bell.** *Models for the specific adhesion of cells to cells.* Science, vol. 200, n° 4342, pages 618–627, 1978.
- [Binnig 86] **G. Binnig, C. F. Quate et Ch. Gerber.** *Atomic Force Microscope.* Phys. Rev. Lett., vol. 56, n° 9, pages 930–933, Mar 1986.
- [Bockelmann 02] **U. Bockelmann, Ph. Thomen, B. Essevaz-Roulet, V. Viasnoff et F. Heslot.** *Unzipping DNA with Optical Tweezers : High Sequence Sensitivity and Force Flips.* Biophysical Journal, 2002.
- [Boncheva 05] **M. Boncheva et G. M. Whitesides.** *Making Things by Self-Assembly.* MRS Bulletin, vol. 30, pages 736–742, 2005.
- [Brandsdal 03] **Bo. Brandsdal, I. Feierberg Vb. Luzhkov F. Osterberg M. Almlöf et J. Aqvist.** *Free energy calculations and ligand binding.* Adv. Protein Chem, vol. 66, page 123158, 2003.

- [Butt 05] **H. Butt, B. Cappella et M. Kappl.** *Force measurements with the atomic force microscope : Technique, interpretation and applications.* Surface Science Reports, vol. 59, n° 1-6, pages 1 – 152, 2005.
- [Carlon 06] **Enrico Carlon et Thomas Heim.** *Thermodynamics of RNA/DNA hybridization in high-density oligonucleotide microarrays.* Physica A : Statistical Mechanics and its Applications, vol. 362, n° 2, pages 433 – 449, 2006.
- [Coffer 92] **J L Coffer, S R Bigham, R F Pinizzotto et H Yang.** *Characterization of quantum-confined CdS nanocrystallites stabilized by deoxyribonucleic acid (DNA).* Nanotechnology, vol. 3, n° 2, pages 69–76, 1992.
- [Daunay 07] **B. Daunay, A. Micaelli et S. Régnier.** *Energy-Field Reconstruction for Haptics-Based Molecular Docking Using Energy Minimization Processes.* In Actes de IROS'07 : IEEE International Conference on Intelligent Robots and Systems, San Diego, USA, Novembre 2007.
- [Deaton 02] **R. Deaton, J. Chen, H. Bi et J. A. Rose.** *A software tool for generating noncrosshybridization libraries of DNA oligonucleotides.* In 8th Int. Workshop DNA Based Comput, page 252261, 2002.
- [Dimitrov 04] **Roumen A. Dimitrov et Michael Zuker.** *Prediction of Hybridization and Melting for Double-Stranded Nucleic Acids.* Biophysical Journal, vol. 87, n° 1, pages 215 – 226, 2004.
- [Erickson 03] **David Erickson, Dongqing Li et Ulrich J.Krull.** *Modeling of DNA hybridization kinetics for spatially resolved biochips.* Analytical Biochemistry, vol. 317, pages 186–200, 2003.
- [Essevaz-Roulet 97] **B. Essevaz-Roulet, U. Bockelmann et F. Heslot.** *Mechanical separation of the complementary strands of DNA.* Proceedings of the National Academy of Sciences of the United States of America, vol. 94, n° 22, pages 11935–11940, 1997.
- [Evans 97] **E. Evans et K. Ritchie.** *Dynamic strength of molecular adhesion bonds.* Biophysical Journal, vol. 72, n° 4, pages 1541 – 1555, 1997.
- [Evans 99] **Evan Evans et Ken Ritchie.** *Strength of a Weak Bond Connecting Flexible Polymer Chains.* Biophysical Journal, vol. 76, n° 5, pages 2439 – 2447, 1999.
- [Feldkamp 02] **Udo Feldkamp, Sam Saghafi, Wolfgang Banzhaf et Hilmar Rauhe.** *Dna computing, volume 2340, chapitre DNASequencesGenerator : A Program for the Construction of DNA Sequences, pages 23–32.* SpringerLink, 2002.
- [Florin 94] **EL Florin, VT Moy et HE Gaub.** *Adhesion forces between individual ligand-receptor pairs.* Science, vol. 264, n° 5157, pages 415–417, 1994.
- [Frutos 97] **A. G. Frutos, A. J. Thiel, A. E. Condon, L. M. Smith et R. M. Corn.** *"DNA computing at surfaces : Four base mismatch word design".* In 3rd DIMACS Workshop DNA Based Comput., page 238, 1997.

- [Gauthier 07] **M. Gauthier et S. Régnier.** *La manipulation aux échelles microscopiques.* In Actes des Journées Nationales de Recherche en Robotique, JNRR'07, Strasbourg, France, 2007.
- [Giroto 07] **Maxime Giroto.** *"Plate-forme d'étude de la Mécanotransduction cellulaire".* Thèse de doctorat, Université Pierre et Marie Curie, 2007.
- [Glade 04] **Nicolas Glade, Jacques Demongeot et James Tabony.** *Microtubule self-organisation by reaction-diffusion processes causes collective transport and organisation of cellular particles.* BMC Cell Biology, vol. 5, n° 1, page 23, 2004.
- [Glazer 06] **Marc Glazer, Jacqueline A. Fidanza, Glenn H. McGall, Mark O. Trulson, Jon E. Forman, Audrey Suseno et Curtis W. Frank.** *Kinetics of oligonucleotide hybridization to photolithographically patterned DNA arrays.* Analytical Biochemistry, vol. 358, n° 2, pages 225 – 238, 2006.
- [Golem 06] **Projet européen Golem.** <http://www.golem-project.eu/team.htm>. Rapport technique, 2006.
- [Guv 08] "molecular modeling of proteins", chapitre "Comparison of Protein Force Fields for Molecular Dynamics Simulations". Springer, 2008.
- [Hamdi 09] **Mustapha Hamdi et Antoine Ferreira.** *Multiscale Design and Modeling of Protein-based Nanomechanisms for Nanorobotics.* I. J. Robotic Res., vol. 28, n° 4, pages 436–449, 2009.
- [Hartemink 98] **Alexander J. Hartemink, Er J. Hartemink, David K. Gifford et Julia Khodor.** *Automated Constraint-Based Nucleotide Sequence Selection for DNA Computation.* In 4th Int. Meeting on DNA-Based Computing, pages 227–235, 1998.
- [Hess 04] **H. Hess, D. George et V. V. Bachand.** *Powering Nanodevices with Biomolecular Motors.* Chemistry - A European Journal, 2004.
- [Huang 06] **H. Huang, J. N. Anker, K. Wang et R. Kopelman.** *Magnetically Assisted and Accelerated Self-Assembly of Strawberry-like Nano/Microparticles.* J. Phys. Chem. B, vol. 110, n° 40, pages 19929–19934, 2006.
- [Isr 91] Intermolecular and surface forces. 1991.
- [Janovjak 05] **H. Janovjak, K.T. Sapra et D.J. Müller.** *Complex stability of single proteins explored by forced unfolding experiments.* Biophysical Journal, 2005.
- [Jayaraman 06] **Arthi Jayaraman, Carol K. Hall et Jan Genzer.** *Computer Simulation Study of Molecular Recognition in Model DNA Microarrays.* Biophysical Journal, vol. 91, pages 2227–2236, 2006.
- [Junker 09] **J. P. Junker et M. Rief.** *Single-molecule force spectroscopy distinguishes target binding modes of calmodulin.* Biological Sciences - Biophysics and Computational Biology, vol. 106, pages 14361–14366, 2009.

- [Kramers 40] **H. Kramers.** *Brownian motion in a field of force and the diffusion model of chemical reactions.* Physica, vol. 7, n° 4, pages 284–304, April 1940.
- [Lambert 06] **P. Lambert et S. Régnier.** *Surface and contact forces models within the framework of microassembly.* International Journal of Micromechatronics, vol. 3, n° 2, pages 123–157, 2006.
- [Leach 01] **Andrew Leach.** *Molecular modelling : Principles and applications* (2nd edition). Prentice Hall, 2 édition, April 2001.
- [Lee 94] **GU Lee, LA Chrisey et RJ Colton.** *Direct measurement of the forces between complementary strands of DNA.* Science, vol. 266, n° 5186, pages 771–773, 1994.
- [Leunissen 10] **M. E. Leunissen, R. Dreyfus, Sha, R. C. Nadrian et Paul M. Chaikin Seeman.** *Quantitative Study of the Association Thermodynamics and Kinetics of DNA-Coated Particles for Different Functionalization Schemes.* Journal of the American Chemical Society, 2010.
- [Liu 08] **W. Liu, Vedrana Montana, Vladimir Parpura et U. Mohideen.** *Comparative Energy Measurements in Single Molecule Interactions.* Biophysical Journal, vol. 95, n° 1, pages 419–425, juillet 2008.
- [Mann 09] **Stephen Mann.** *Self-assembly and transformation of hybrid nano-objects and nanostructures under equilibrium and non-equilibrium conditions.* Nature Materials, 2009.
- [Marathe 99] **A. Marathe, A. E. Condon et R. M. Corn.** *"On combinatorial DNA word design".* In 5th DIMACS Workshop DNA Based Comput., page 7589, 1999.
- [Marti 87] **O. Marti, B. Drake et P. K. Hansma.** *Atomic force microscopy of liquid-covered surfaces : Atomic resolution images.* Applied Physics Letters, vol. 51, n° 7, pages 484–486, 1987.
- [Martin 99] **T.W. Martin et Zygmunt S. Derewenda.** *The name is bond H bond.* Nature Structural Biology, vol. 6, pages 403–406, 1999.
- [Matthias Rief 99] **Hermann E. Gaub Matthias Rief Hauke Clausen-Schaumann.** *Sequence-dependent mechanics of single DNA molecules.* Nature Structural Biology, 1999.
- [McNally 03] **H. McNally, M. Pingle, S. W. Lee, D. Guo, D. E. Bergstrom et R. Bashir.** *Self-assembly of micro- and nano-scale particles using bio-inspired events.* Applied Surface Science, vol. 214, n° 1-4, pages 109 – 119, 2003.
- [Mirkin 96] **Chad A. Mirkin, Robert L. Letsinger, Robert C. Mucic et James J. Storhoff.** *A DNA-based method for rationally assembling nanoparticles into macroscopic materials.* Nature, vol. 382, pages 607–609, 1996.
- [Morfill 07] **Julia Morfill, Ferdinand Kühner, Kerstin Blank, Robert A. Luga-maier, Julia Sedlmair et Hermann E. Gaub.** *B-S Transition in Short*

- Oligonucleotides*. Biophysical Journal, vol. 93, n° 7, pages 2400 – 2409, 2007.
- [Morris 05] **Stauth S.A.-Parviz B.A. Morris C.J.** *Self-assembly for microscale and nanoscale packaging : steps toward self-packaging*. Advanced Packaging, IEEE Transactions on, vol. 28, n° 4, pages 600 – 611, Nov. 2005.
- [Neukirch 04] **S. Neukirch**. *Extracting DNA twist rigidity from experimental supercoiling data*. Physical Review Letters, vol. 93, page 19, 2004.
- [Nom 07] "current pharmaceutical biotechnology ", chapitre "Direct Quantification of Gene Expression Using Fluorescence Correlation Spectroscopy ". Bentham Science Publishers, 2007.
- [Odrowaz 06] **P.M. Odrowaz, P. Czuba, M. Targosz, K. Burda et M. Szymonski**. *Dynamic force measurements of avidin-biotin and streptavidin-biotin interactions using AFM*. Acta Biochim Pol, 2006.
- [O’Riordan 04a] **A. O’Riordan, P. Delaney et G. Redmond**. *Field Configured Assembly : Programmed Manipulation and Self-assembly at the Mesoscale*. Nano Letters, vol. 4, n° 5, pages 761–765, 2004.
- [O’Riordan 04b] **A. O’Riordan, P. Delaney et G. Redmond**. *Field Configured Assembly : Programmed Manipulation and Self-assembly at the Mesoscale*. Nano Letters, vol. 4, n° 5, pages 761–765, 2004.
- [Penchovsky 03] **R. Penchovsky et J. Ackermann**. *DNA library design for molecular computation*. Journal of Computational Biology, vol. 10, page 215229, 2003.
- [Perkins 94] **TT Perkins, SR Quake, DE Smith et S Chu**. *Relaxation of a single DNA molecule observed by optical microscopy*. Science, vol. 264, n° 5160, pages 822–826, 1994.
- [Phan 06] **Anh Tuân Phan, Vitaly Kuryavyi et Dinshaw J Patel**. *DNA architecture : from G to Z*. Current Opinion in Structural Biology, vol. 16, n° 3, pages 288 – 298, 2006. Nucleic acids/Sequences and topology - Anna Marie Pyle and Jonathan Widom/Nick V Grishin and Sarah A Teichmann.
- [Poland 66] **D. Poland et H. A. Scheraga**. *Phase Transitions in One Dimension and the HelixCoil Transition in Polyamino Acids*. J. Chem. Phys, vol. 45, page 1456, 1966.
- [Rau 84] **D C Rau, B Lee et V A Parsegian**. *Measurement of the repulsive force between polyelectrolyte molecules in ionic solution : hydration forces between parallel DNA double helices*. Proceedings of the National Academy of Sciences of the United States of America, vol. 81, n° 9, pages 2621–2625, 1984.
- [Régnier 08] **S. Régnier et N. Chaillet**. *La microrobotique - applications à la micro-manipulation*. 2008.
- [Rothmund 00] **Paul W. K. Rothmund et Erik Winfree**. *The program-size complexity of self-assembled squares (extended abstract)*. In STOC '00 : Proceedings of



- the thirty-second annual ACM symposium on Theory of computing, pages 459–468, New York, NY, USA, 2000. ACM.
- [Rothmund 06] **Paul W. K. Rothmund.** *Folding DNA to create nanoscale shapes and patterns.* Nature, vol. 440, pages 297–302, 2006.
- [Rothermund 06] **P. W. K. Rothermund.** *Folding DNA to create nanoscale shapes and patterns.* Nature, vol. 440, pages 297–302, 2006.
- [Sae 83] Principles of nucleic acid structure. Springer, 1983.
- [SantaLucia 98] **Jr. SantaLucia.** *A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics.* PNAS, vol. 95, pages 1460–1465, 1998.
- [Sattin 04] **B.D. Sattin, A.E. Pelling et Goh C.M.** *DNA base pair resolution by single molecule force spectroscopy.* Nucleic Acids Research, vol. 32, n° 16, pages 4876–4883, 2004.
- [Schiefner 09] **André Schiefner, Masakazu Fujio, Douglass Wu, Chi-Huey Wong et Ian A. Wilson.** *Structural Evaluation of Potent NKT Cell Agonists : Implications for Design of Novel Stimulatory Ligands.* Journal of Molecular Biology, vol. 394, n° 1, pages 71 – 82, 2009.
- [Seeman 83] **N.C. Seeman et N.R. Kallenbach.** *Design of immobile nucleic acid junctions.* Biophysical Journal, vol. 44, n° 2, pages 201–209, novembre 1983.
- [Shih 04] **W. M. Shih, J. D. Quispe et G. F. Joyce.** *1.7-Kilobase single stranded DNA that folds into a nanoscale octahedron.* Nature, vol. 427, pages 618–621, 2004.
- [Shin 05] **S. Shin, I. Lee, D. Kim et B. Zhang.** *Multi-objective evolutionary optimization of DNA sequences for reliable DNA computing.* IEEE Transactions on Evolutionary Computation, vol. 9, pages 143–158, 2005.
- [Smith 92] **SB Smith, L Finzi et C Bustamante.** *Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads.* Science, vol. 258, n° 5085, pages 1122–1126, 1992.
- [Stauth 06] **Sean A. Stauth et Babak A. Parviz.** *Self-assembled single-crystal silicon circuits on plastic.* Proceedings of the National Academy of Sciences, vol. 103, n° 38, pages 13922–13927, 2006.
- [Strunz 99] **Torsten Strunz, Krisztina Oroszlan, Rolf Schäfer et Hans-Joachim Güntherodt.** *Dynamic force spectroscopy of single DNA molecules.* Proceedings of the National Academy of Sciences of the United States of America, vol. 96, n° 20, pages 11277–11282, 1999.
- [Tanaka 01] **M. Tanaka F.and Nakatsugawa, M. Yamamoto, T. Shiba et A. Ohuchi.** *"Developing support system for sequence design in DNA computing".* In 7th Int. Workshop DNA Based Comput., page 340349, 2001.

- [Valignat 05] **Marie-Pierre Valignat, Olivier Theodoly, John C. Crocker, William B. Russel et Paul M. Chaikin.** *Reversible self-assembly and directed assembly of DNA-linked micrometer-sized colloids.* Proceedings of the National Academy of Sciences of the United States of America, vol. 102, n° 12, pages 4225–4229, 2005.
- [Vela Saavedra 09] **E. Vela Saavedra, M. Hafez et S. Régnier.** *Laser-Induced Thermo-capillary Convection for Mesoscale Manipulation.* International Journal of Optomechatronics, vol. 3, n° 4, pages 289–302, 2009.
- [Watson 53] **J. D. Watson et F. H. C. CRICK.** *Genetical Implications of the Structure of Deoxyribonucleic Acid.* Nature, 1953.
- [Williams 01] **J. R. Wenner I. Rouzina Williams M. C. et V. A. Bloomfield.** *Effect of pH on the overstretching transition of double-stranded DNA : evidence of force-induced DNA melting.* Biophys. J., vol. 80, page 874881, 2001.
- [Xie 08] **H. Xie, J. Vitard, S. Haliyo et S. Régnier.** *Calibration and Nonlinearity Compensation for Force Application in AFM based Nanomanipulation.* In 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2008.
- [Xiong 03] **X. Xiong, Y. Hanein, J. Fang, Y. Wang, W. Wang, D. T. Schwartzand et K. F. Böhringer.** *Controlled Multibatch Self-Assembly of Micro-devices.* Journal of Microelectromechanical Systems, vol. 12, n° 2, pages 117–127, 2003.
- [You 05] Handbook of nanostructured biomaterials and their applications, chapitre DNA-Based Artificial Nanostructures : Fabrication, Properties, and Applications. American Scientific Publishers, Stevenson Ranch, CA,, 2005.
- [Yu He 08] **Min Su Chuan Zhang Alexander E. Ribbe Wen Jiang Yu He Tao Ye et Chengde Mao.** *Hierarchical self-assembly of DNA into symmetric supramolecular polyhedral.* Nature, vol. 452, pages 198–201, 2008.
- [Zheng 04] **W. Zheng, P. Buhlmann et H. O. Jacobs.** *Sequential shape-and-solder-directed self-assembly of functional microsystems.* PNAS, vol. 101, n° 35, pages 12814–12817, 2004.
- [Zuker 03] **M. Zuker.** *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Res, vol. 31, pages 3406–15, 2003.