

Ecole Doctorale Carnot

Institut Carnot de Bourgogne

Département Nanosciences, Equipe OSNC, UMR CNRS 5209

ANNÉE 2011

THÈSE

Pour obtenir le grade de Docteur de l'Université de Bourgogne

Discipline : Sciences

Spécialité : Physique – Nanosciences et Nanotechnologies

Caractérisation et modélisation de structures nucléiques auto-assemblées fonctionnalisables

Soutenue par

Aude LAISNÉ

à Dijon

Le 15 décembre 2011 devant la commission d'examen composée de :

Pr	MINARD P.	Université d'Orsay, Paris 11	Rapporteur
Dr.	MILHIET P.E.	CBS CNRS Montpellier 1	Rapporteur
Pr.	DEREUX A.	Université de Bourgogne Dijon	Président du jury
Dr.	LEROY J.L.	CNRS Gif-sur-Yvette	Examineur
Dr.	CANVA M.	CNRS/ Institut d'Optique Palaiseau	Examineur
Dr.	POMPON D.	CNRS/ INSA Toulouse	Co-directeur de thèse
Pr.	LESNIEWSKA E.	Université de Bourgogne Dijon	Directeur de thèse

REMERCIEMENTS

Je remercie vivement Eric Lesniewska, directeur officiel, qui a réussi au pied levé à mettre en route la machinerie administrative afin de m'inscrire dans ma 1^{ère} année de thèse un beau jour du mois de mai de l'année 2010, sûrement inspiré par le dicton : « en mai fais ce qu'il te plaît ! » Alors ce fut fait ! Merci de m'avoir fait confiance et de m'avoir donné cette chance. Merci de ne pas avoir été effrayé par la supervision lointaine de cette thèse satellitaire et encore merci pour ton accueil et ta disponibilité lors de mes venues à Dijon.

Je remercie Denis Pompon, co-directeur de ma thèse, qui m'a accueillie dans son laboratoire sur un premier contrat de 3 mois un 1^{er} avril en l'an 2005 : un poisson, c'était le bouquet !!! Il n'y avait pas tellement cru, au début...mais trois mois, le risque était faible et l'argent prêt à s'engager ! Aux trois mois succédèrent six autres, puis un contrat ANR sur trois ans : WOW !!! Le grand luxe...décidément je m'installais ! Et encore un an et puis quelques mois, et le grand pas fut fait : je me lançais en thèse !

Merci Denis de m'avoir suivie de près pendant toutes ces années avec un encadrement aux petits oignons dans cette atmosphère si particulière où tous les jours j'apprenais ! Ta disponibilité, tes inlassables conseils et explications, ainsi que ton foisonnement d'idées qui déferlent perpétuellement...que de ski nautique...y' faut suivre...Alors vraiment, un immense merci du fond du cœur pour tous ces moments partagés et pour ton aide !

Je remercie Alain Dereux de m'avoir donné la possibilité de réaliser cette thèse sous la tutelle de son laboratoire, et de faire partie de mon jury.

Je remercie particulièrement Philippe Minard et Pierre-Emmanuel Milhiet d'avoir accepté de rapporter ma thèse malgré un programme bien chargé. Merci également à Jean-Louis et Michael Canva pour faire partie du jury.

Je remercie Frédéric Boccard pour avoir également laissé cette thèse prendre son cours alors qu'elle ne rentrait pas dans les clous d'un système bien normalisé.

Je remercie Hans Jauslin directeur de l'école doctorale Carnot pour sa compréhension et sa bienveillance vis-à-vis de la lancée et de la vie de cette thèse atypique. J'apprécie les aménagements qu'il m'a accordés pour remplir le quota des enseignements prodigués par la formation doctorale.

Je remercie Jean-Louis Leroy de m'avoir acceptée dans son antre du motif-i à partager avec lui les expériences et les remue-méninges (et pas ménage...) auxquels ont donné lieu ces fameuses structures supramoléculaires de motif-i dont les mécanismes d'association restent énigmatiques... Franchement, j'ai beaucoup apprécié interagir avec toi sur ce sujet passionnant. Merci aussi pour tes fêtes à Antony très sympas, et à quand ma visite du Louvre ? Même si j'ai eu la chance d'y aller lors du congrès AFM BioMed à Paris en août 2011, au sein d'un groupe de 22 personnes, en tentant de suivre et d'entendre notre guide

énergique nous retracer, *en anglais*, l'histoire du palais royal... Cette promenade nécessite une visite complémentaire, sans nul doute !

Je remercie Michael Canva, qui fut coordinateur du projet Nanobioforme... quel peps ! Projet grâce auquel j'ai pu être embauchée pendant 3 ans. Merci à toi. Merci pour le travail réalisé en collaboration qui a été très enrichissant. Je remercie toutes les personnes de l'équipe, d'avant et de maintenant, toujours avenantes, dynamiques et prêtes à interagir, je pense à Gisèle Roger, Alain Aide, Fabrice Bardin, Alain Bellemain, Jolanda Spadavecchia, Jérôme Hottin, Aurélien Duval, Mohammed Nakkach, et Julien Moreau. Je pense également à la lampe frontale qui m'a été offerte spécialement (merci Tati Miche) pour travailler dans les labos de l'institut d'optique. Planquée derrière une cloison, je pipetais les solutions et réalisais les dilutions à injecter dans les cuves des bancs optiques « à ciel ouvert » qui nécessitaient l'obscurité la plus totale.

Je remercie tous les membres de l'équipe ex-Gif-sur-Yvette qui au jour de ma soutenance de thèse sera totalement toulousaine à l'INSA. Merci à toi Gilles Truan (de son « petit » nom qui vous met en confiance) pour ta présence, tes conseils de qualité, ton avis, ton aide aussi, très précieuse et ta disponibilité. Ce fut un plaisir de te côtoyer durant ces quelques années. J'ai tout apprécié et t'en remercie profondément. Merci également à Philippe Urban, ses petites dépêches du matin, la bonne odeur de son incontournable café, et le ravissement de ses friandises, ma foi, bienvenues quand les méninges tournaient à vide. Merci à Jacqueline Loeper, une figure pleine de vie de l'équipe LIPM de Gif, qui ne passe forcément pas inaperçue dans les lieux qu'elle occupe... Pleine d'enthousiasme et dynamique elle sait exprimer son point de vue.

Je remercie également tous les membres de l'équipe d'Eric Lesniewska, permanents et étudiants dont j'ai également pu apprécier l'accueil lors de mes venues à Dijon. Merci à Eric Bourillot, et merci à David et Maxime pour les images et films réalisés en AFM conventionnel et à haute vitesse.

Je remercie tous les ex-étudiants thésards désormais en poste ou en post-doc, Damjana, Marcella, Laetitia, Cube, Benoît, Louise et les autres qui ont dû suffisamment m'inspirer pour que je saute moi-même dans le bain... comme quoi, y'a pas d'âge pour en manger... Merci à tous pour tous les moments partagés au labo ou à l'extérieur, comme cette escapade en Slovénie grâce à toi Damjana, que de souvenirs inoubliables... merci !

Merci à Florence et Fataneh, les dernières filles à rester au laboratoire avant le grand déménagement dans le sud, pour votre douce et bonne humeur qui réchauffent le cœur. Merci aussi pour l'accueil à Nîmes très apprécié et félicitations pour les bébés... même si Mathias a déjà fêté sa 1^{ère} année ! Et bienvenue à Diana.

C'est avec grand plaisir que j'ai partagé ce bureau, de manière permanente avec toi Laure, et de manière passagère avec les nombreux étudiants que tu as encadrés ou avec le

chercheur Kyril Turpaev. Merci Laure de ton amitié, ta présence, ton écoute et ton regard extérieur m'ont toujours été d'un réel soutien.

Je remercie toutes les personnes du bât 23B, de la RMN à haut-champ (et non AUCHAN, dans une petite anecdote que me racontait Gilles...) et de la plateforme. Je pense à Christina, Alda, Karine, Annie, Eric et Eric, Karima, Sophie aussi en pleine rédaction de thèse, ...et tous ceux que je n'ai pas cités mais bien croisés pendant ces années.

Et enfin un grand merci à ceux qui croient en moi et qui me soutiennent ! Je pense à ma famille proche Jacques, Suzanne, Amaï, Lees, et éloignée (ils sont vraiment trop nombreux pour les citer mais pas suffisamment pour les porter tous dans mon cœur) et à mes amis dont l'amitié me renforce. Merci maman pour toute l'aide précieuse apportée, notamment en prenant soin de Serena, toujours au bon moment. Merci à toi Franck pour ton soutien et ton amour incommensurable (définition : si grand qu'on ne peut le mesurer), c'est doux de pouvoir compter sur toi. Merci à Agnès pour tous les petits mots doux et encouragements. Merci à Antoine et Béatrice, pour votre accueil et merci aussi de ne pas avoir été surpris lors de mon annonce de démarrer (et finir) cette thèse. Merci à Michèle pour la relève avec Serena qui me dépanna bien des fois. Merci à Christophe, le père de Serena, pour entretenir de bonnes relations malgré notre séparation.

Je dédie cette thèse à Serena pour qu'elle sache que dans la vie tout est possible... Je dédie aussi ce travail à ma mère, qui est admirablement toujours présente aux moments clés (sans en oublier d'autres...) et pourquoi pas à mon père, mais lui va dire que « c'est des c...ries tout ça ! ».

Je finirai par « remercier » ou plutôt citer Sénèque : « Ce n'est pas parce que les choses sont difficiles que nous n'osons pas, c'est parce que nous n'osons pas qu'elles sont difficiles » que j'ai eu le loisir de lire sur un poster dans les toilettes, chez ma sœur Amaï qui habite près d'Annecy, où on distinguait un homme miniature s'attaquer à une immense paroi rocheuse de belle couleur ocre !

Et merci au ciel !

TABLE DES MATIERES

INTRODUCTION	3
PRESENTATION DU TRAVAIL DE THESE	7
CONCEPTION DE NANO-OBJETS A BASE D'ACIDES NUCLEIQUES ET DE PROTEINES	9
1. AUTO-ASSEMBLAGE.....	10
1.1. Auto-assemblage des protéines.....	12
<i>1.1.1 Mécanismes d'assemblages des protéines</i>	<i>12</i>
1.1.1.1 Notion de module structural	14
1.1.1.2 Une combinatoire infinie d'un nombre limité de blocs moléculaires.	14
1.1.1.3 Assemblages supramoléculaires	17
1.1.1.3.1 Protéines globulaires.....	17
1.1.1.3.2 Protéines fibreuses.....	18
<i>1.1.2 L'auto-assemblage comme outil de construction moléculaire.....</i>	<i>20</i>
1.2. Auto-assemblage des acides nucléiques	22
<i>1.2.1 Un peu d'histoire.....</i>	<i>22</i>
<i>1.2.2 Les différentes formes naturelles d'ADN</i>	<i>23</i>
1.2.2.1 ADN B.....	23
1.2.2.2 ADN A.....	24
1.2.2.3 ADN Z.....	24
1.2.2.1 ADN cruciforme et ADN en épingle à cheveux	24
1.2.2.2 ADN-H ou ADN triplex	25
1.2.2.3 Quadruple hélice d'ADN : les quadruplex de G	25
1.2.2.4 Quadruple hélice d'ADN : les motifs-i.....	26
<i>1.2.3 Formes artificielles associant acides nucléiques et composants peptidiques.....</i>	<i>27</i>
1.2.3.1 PNA	27
1.2.3.2 PDNA	28
<i>1.2.4 Utilisation des différentes formes d'ADN pour la fabrication de nanostructures.</i>	<i>29</i>
1.2.4.1 Stratégies d'auto-assemblage des acides nucléiques	29
1.2.4.1.1 Auto-assemblage direct ou parallèle.....	30
1.2.4.1.2 Auto-assemblage séquentiel	31
1.2.4.1.3 Auto-assemblage dirigé par nucléation	32

1.2.4.1.4 Auto-assemblage algorithmique	33
1.2.4.2 Principaux blocs moléculaires utilisés en structuration d'ADN	35
1.2.4.2.1 ADN linéaires.....	35
1.2.4.2.2 ADN branchés.....	37
1.2.4.2.3 Hélices multiples	42
1.2.4.2.4 Autres structures.....	43
1.2.4.3 Maîtrise des dimensions	46
1.2.4.4 Avantages et limitations des différentes stratégies d'auto-assemblage	47
1.2.4.5 Approche locale d'auto-assemblage versus à longue distance	49
1.2.4.6 Approche mixte pour l'auto-assemblage de nanostructures en 3D.....	49
<i>1.2.5 Les auto-assemblages hybrides.....</i>	<i>51</i>
1.2.5.1 Nécessité d'associer structure et fonctionnalité	51
1.2.5.1.1 ADN et composés inorganiques : fonctionnalisation et guide.	52
1.2.5.1.2 ADN et protéines.....	52
1.2.5.2 Interfaces et applications	60
1.2.5.2.1 ADN –Surface	60
<i>1.2.6 Assemblages irréversibles, réversibles et dynamiques</i>	<i>61</i>
2. PARAMETRES PHYSICO-CHIMIQUES CRITIQUES POUR L'ASSEMBLAGE DES ACIDES NUCLEIQUES	63
2.1. Mécanismes d'hybridation et stabilité de l'ADN B.....	63
2.1.1 <i>Température de fusion.....</i>	<i>63</i>
2.1.2 <i>Calcul de la température de fusion</i>	<i>64</i>
2.1.3 <i>Paramètres influençant la température de fusion.....</i>	<i>64</i>
2.1.4 <i>Cinétique d'hybridation</i>	<i>65</i>
2.2. Modèles de description des polymères semi-flexibles.....	67
2.2.1 <i>La longueur de persistance pour quantifier la rigidité d'un polymère.....</i>	<i>67</i>
2.2.2 <i>Cyclisation : étude du facteur J pour déterminer la longueur de persistance.....</i>	<i>68</i>
3. OUTILS DE CARACTERISATION DES NANOSTRUCTURES	70
3.1. Molécule unique & population	70
3.2. Domaines spatio-temporels des différentes techniques d'analyse.....	70
3.3. Microscopie électronique.....	71
3.4. Microscopie à force atomique (AFM)	72
3.4.1 <i>Principe du microscope AFM conventionnel.....</i>	<i>72</i>

3.4.2 Modes d'utilisation de l'AFM	73
3.4.2.1 Mode contact	73
3.4.2.2 Mode résonnant	74
3.4.3 Développement du microscope à haute vitesse	75
3.4.4 Quels substrats pour déposer les échantillons ?	77
3.4.5 Avantages et inconvénients	78
3.5. Microscopie optique Confocale	79
3.5.1 Principe du microscope confocal	79
3.5.2 Avantages et inconvénients	80
3.5.3 Fluorophores	80
3.6. Résonance plasmonique de surface (SPR)	82
3.6.1 Principe	82
3.6.2 Avantages et inconvénients de la SPR	83
MODELISATION DES ASSEMBLAGES MONODIMENSIONNELS IMPLIQUANT DES BRIQUES	
D'ADN.....	87
1. PRESENTATION DE L'ARTICLE 1.	88
2. APPENDICE 1 : ARTICLE 1	91
3. DONNEES COMPLEMENTAIRES	93
3.1. Analyse de l'état de polymérisation par électrophorèse	93
3.1.1 Méthodes	93
3.1.1 Résultats des analyses sur gel	95
3.2. Analyse de l'état de polymérisation par microscopie de force atomique (AFM)	99
3.2.1 Méthodes	99
3.2.2 Résultats	101
4. DISCUSSIONS COMPLEMENTAIRES	102
4.1. Les différentes techniques de caractérisation : apports et limites	102
4.1.1 Les gels d'électrophorèse	102
4.1.2 La microscopie de force atomique	103
4.1.3 La technique de cinétique rapide	104
4.2. Mécanismes de terminaison dans les assemblages illimités	104
4.2.1 Mécanismes dirigés par l'ajout de terminateurs	104
4.2.2 Mécanismes d'invasion de brins- espèces branchées	104

4.2.3 Mécanismes de cyclisation.....	105
4.3. Calcul des paramètres sur la base des simulations et leur limitation.....	106
4.4. Limites de l’approche combinatoire dans le cas des modèles illimités.	107
4.5. Réversibilité des réactions d’assemblage.....	108
4.6. Maîtrise des stœchiométries.....	110
4.7. Conclusion	110
LES STRUCTURES HYBRIDES PDNA (PROTÉINE-DNA)	113
1. POURQUOI ASSOCIER PROTEINES ET ACIDES NUCLEIQUES?	113
2. AVANTAGES ET INCONVENIENTS DES METHODES DE COUPLAGE PROTEINE–ADN	113
3. LES PDNA.....	114
3.1. Présentation des articles 2 et 3 : caractérisations des PDNA.....	116
3.2. Appendice 2 : article 2	119
3.3. Appendice 3 : article 3	121
3.4. Données complémentaires	123
3.4.1 Analyse par microscopie de force atomique (AFM)	123
3.4.1.1 AFM à l’air	123
3.4.1.2 AFM en phase liquide.....	124
3.4.2 Echantillons de PDNA utilisés	125
3.4.2.1 Structures définies de PDNA.....	125
3.4.2.2 Polymères de PDNA.....	125
3.4.2.3 Utilisation d’un brin matrice d’ADN simple brin obtenu par clonage.....	126
3.4.3 Résultats des caractérisations par AFM	128
3.4.3.1 AFM à l’air	128
3.4.3.2 AFM à haute vitesse (HS-AFM).....	132
3.5. Discussions complémentaires	143
4. CONCLUSION	145
AUTO-ASSEMBLAGES MULTIPLEXES SUR PUCE.....	149
1. PRESENTATION DE L’ARTICLE 4 : “MODELING AND VALIDATION OF MULTIPLEX PROTEO-NUCLEIC SELF-ASSEMBLIES MONITORED BY SURFACE PLASMON RESONANCE IMAGERY” ..	149
2. APPENDICE 5 : ARTICLE 5.....	153
3. DISCUSSION COMPLEMENTAIRE DE L’ARTICLE 4.....	155
3.1. Avantages et limitations de la SPRi.....	155
4. CONCLUSION	159

MOTIF-I, UNE STRUCTURE PARTICULIERE DE L'ADN CONTROLABLE PAR LE pH.....	163
1. PRESENTATION DE L'ARTICLE :.....	165
2. APPENDICE 5 : ARTICLE 5.....	169
3. DONNEES COMPLEMENTAIRES	171
3.1. Caractérisation par microscopie de force atomique.....	171
3.1.1 Observations des structures supramoléculaires de C_4AC_7	172
3.1.2 Observation des structures supramoléculaires de C_7TC_4	173
3.1.3 Observations de structures supramoléculaires de C_7TC_4 après tri en taille par tamisage moléculaire dans des tampons de pH croissant:	175
3.2. Discussions complémentaires	178
3.2.1 Auto-assemblage de motifs-i sous contrôle d'un substrat.....	179
3.2.2 Rôle du résidu central	180
4. CONCLUSION	181
CONCLUSIONS GENERALES ET PERSPECTIVES	183
ANNEXES	187
APPENDICE 5 : ARTICLE 6	189
BIBLIOGRAPHIE	191

TABLE DES ILLUSTRATIONS

Figure 1 : Graphe de la biologie de synthèse. [14].....	4
Figure 2 : Représentation des classifications SCOP et CATH.....	16
Figure 3 : Structures secondaires et supersecondaires.	16
Figure 4 : Auto-assemblage des protéines globulaires : α -lactalbumine et lysozyme.	18
Figure 5 : Structures secondaires et quaternaires du collagène, et auto-assemblage de la tubuline en microtubules.	19
Figure 6 : Quelques motifs de repliement des protéines utilisés dans la conception <i>de novo</i> . 20	
Figure 7 : Les différentes formes d'ADN et caractérisation de l'ADN B.....	23
Figure 8 : ADN cruciforme et ADN triplex.	25
Figure 9 : Structures d'ADN G et d'ADN I.	26
Figure 10 : Structures artificielles : PNA et PDNA.	28
Figure 11 : Formes d'ADN origami.	31
Figure 12 : Un exemple d'auto-assemblage séquentiel.....	32
Figure 13 : Exemple d'auto-assemblage nucléé.....	33
Figure 14 : Auto-assemblage algorithmique.	34
Figure 15 : Motifs d'ADN utilisés dans les auto-assemblages.	36
Figure 16 : Autres motifs 4x4 et symétriques.	40
Figure 17 : Images AFM de nanostructures à base de quadruplex de G et de motifs-i.	43
Figure 18 : Structures à base d'ARN.	44
Figure 19 : Assemblages d'ADN liés aux nanoparticules d'or.....	45
Figure 20 : Différents motifs protéiques interagissant avec l'ADN.....	45
Figure 21 : Schéma d'auto-assemblage séquentiel d'ADN polyèdres, et caractérisation de tétraèdres.	50
Figure 22 : Conception d'une boîte d'ADN par méthode séquentielle.....	51
Figure 23 : Stratégies de couplage d'un ADN avec une protéine d'intérêt (POI).	53
Figure 24 : Deux exemples d'interactions utilisées en couplage non covalent.....	55
Figure 25 : Température de fusion.	64
Figure 26 : Evènements de nucléation et de propagation.....	66
Figure 27 : Détermination du facteur j pour un fragment d'ADN avec de courtes extrémités collantes.....	69
Figure 28 : Schéma de principe du microscope à force atomique en mode contact intermittent [342].	73
Figure 29 : Résultante schématique des forces physiques [344].....	73
Figure 30 : Courbes de force. [346]	75
Figure 31 : Système d'asservissement du High-Speed AFM [342].....	76
Figure 32 : Images SEM d'un petit levier avec une pointe en forme de bec d'oiseau.....	76
Figure 33 : Structure en feuillettes du mica muscovite.....	77

Figure 34 : Schéma de principe du microscope confocal à balayage laser.	80
Figure 35 : Caractéristiques des boîtes quantiques.	81
Figure 36 : Principe de la résonance plasmonique de surface.....	82
Figure 37 : Schéma de polymérisation de briques d'ADN simple brin en présence de briques de terminaisons.....	93
Figure 38 : Marqueurs de taille.	94
Figure 39 : Caractérisation des polymères d'ADN par électrophorèse.....	96
Figure 40 : Simulations de la migration sur gel d'agarose de la copolymérisation des quatre briques d'ADN en présence d'une concentration variable du mélange de terminateurs.	97
Figure 41 : Méthode d'analyse des images AFM	100
Figure 42 : distribution de la concentration relative des objets en fonction de leur taille	101
Figure 43 : distribution de la masse relative des objets en fonction de leur taille.	101
Figure 44 : Schéma de cyclisation à partir de 4 séquences d'ADN.	106
Figure 45 : structure moléculaire et propriétés des PDNA.	115
Figure 46 : Assemblage des PDNA en solution.....	115
Figure 47 : Chimies de surface développées sur mica.	124
Figure 48 : Fonctionnalisations mono- et bis-NTA.	124
Figure 49 : Structures définies de PDNA.....	125
Figure 50 : Copolymères de PDNA ordonnés et aléatoires.	126
Figure 51 : Carte et caractéristiques du vecteur pCR II (Invitrogen).....	127
Figure 52 : Hybrides PDNA-fragment d'ADN de 500 pb.	128
Figure 53 : Schémas des copolymères de PDNA ordonnés et aléatoires.....	128
Figure 54 : Images AFM des copolymères de PDNA incubés à 22°C, déposés sur mica époxy-silane bis-NTA.	129
Figure 55 : Images AFM des copolymères de PDNA aléatoires incubés à 22°C, déposés sur mica époxy-silane mono-NTA.	130
Figure 56 : Images AFM de copolymères ordonnés de PDNA post-incubés à 45 et 55°C. ...	131
Figure 57 : Schémas du pentamère de PDNA et du contrôle ADN.	132
Figure 58 : Images HS-AFM du pentamère de PDNA.	133
Figure 59 : Images HS-AFM du pentamère de PDNA.	134
Figure 60 : Images HS-AFM du contrôle du pentamère ADN.	134
Figure 61 : Schémas des copolymères de PDNA ordonnés à partir de quatre blocs.	135
Figure 62 : Images HS-AFM des copolymères de PDNA ordonnés, impliquant 4 blocs de construction.	136
Figure 63 : Sections d'image de copolymères de PDNA.....	136
Figure 64 : Images HS-AFM des copolymères de PDNA.	137
Figure 65 : Schémas de l'hybride 5 PDNA sur le brin ADN matrice et son contrôle.	138
Figure 66 : Images HS-AFM de PDNA hybridés sur un brin d'ADN matrice.....	138
Figure 67 : Images HS-AFM du contrôle du brin d'ADN matrice sans protéine.	139
Figure 68 : Images HS-AFM illustrant l'hétérogénéité des mouvements des objets.....	141

Figure 69 : schéma illustrant les propriétés dynamiques hétérogènes des ADN et protéines sur mica.	142
Figure 70 : Schéma des hybridations possibles entre échantillons PDNA 1.	143
Figure 71 : Biopolymère dextran.	150
Figure 72 : Potentiel électrostatique du cytochrome b5.	157
Figure 73 : Prismes microstructurés.	158
Figure 74 : schéma de fluidique de la cellule SPRI.	159
Figure 75 : structure particulière de l'ADN : le motif-i.	163
Figure 76 : Représentations schématiques de motifs-i.	164
Figure 77 : Vue schématique du principe de formation des fils i à base de la séquence C ₇ . .	165
Figure 78 : Mécanisme postulé de l'assemblage de motifs-i C ₇ XC ₄ en structures supramoléculaires.	166
Figure 79 : Image AFM de C ₇ AC ₄	173
Figure 80 : Images AFM de C ₇ GC ₄	173
Figure 81 : images AFM de structures de C ₇ TC ₄	174
Figure 82 : Images AFM de structures de C ₇ TC ₄ fractionnées sur tamisage moléculaire.	177
Figure 83 : schémas hypothétiques de jonction à trois voies.	180
Figure 84 : schéma de concept de nanomachine protéonucléique à fonctionnalités multiples.	184

ABBREVIATIONS

ADN	Acide désoxyribonucléique
AFM	Atomic Force microscopy
AGAT	Protéine O ₆ -alkylguanine-ADN-alkyletransférase
APBS	Adaptive Poisson-Boltzmann Solver
ARN	Acide ribonucléique
CATH	Class, Architecture, Topology, Homologous superfamily
CCD	Charged couple device (Dispositif à transfert de charge)
CDC	Chimie dynamique constitutionnelle
CDP	Réactif chimiluminescent (1,2-dioxetane) pour détecter l'alcaline phosphatase
DAE	Double-cross Antiparallel Even (motif d'ADN à double croisement, antiparallèle impair)
DAO	Double-cross Antiparallel Odd (motif d'ADN à double croisement, antiparallèle pair)
DC	Direct Current
DDX	Double double crossover (motif d'ADN à double-double croisement)
DI	Diamètre interne
DMT	Derjaguin, Muller and Toporov
DNA	Deoxyribonucleic acid
DPE	Double-cross Parallel Even (motif d'ADN à double croisement, parallèle impair)
DPON	Double-cross Antiparallel Odd Narrow (motif d'ADN à double croisement, antiparallèle pair étroit)
DPOW	Double-cross Antiparallel Odd Wide (motif d'ADN à double croisement, antiparallèle pair large)
DRX	Diffraction des rayons X
DX	Double cross (Motif d'ADN double croisement)
EDTA	Acide 2,2',2'',2'''-(éthane-1,2-diyl)dinitrilo)tétracétique (acide éthylène diamine tétracétique)
EPL	Expressed protein ligation
EPS	Epoxy silane
FJC	Freely jointed chain
GPC	Gel Permeation Chromatography
GPTS	3-glycido-xypropyl-trimethoxysilane
GTG	NuSieve® GTG® Agarose
GTP	Guanosine triphosphate
HPLC	High-performance liquid chromatography (Chromatographie en phase liquide à haute performance)
JX	Juxtapose without crossover
LAS	Luminescent image analyzer
LMW	Low Molecular Weight
LYS	Lysozyme
MES	Tris, 2-morpholinoethane sulfonate

N_A	Constante d'Avogadro
NA	Number aperture
NHS	N-hydroxysuccinimide
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NTA	Nitrilotriacetic acid
OH	Hydroxyle
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction (Réaction en chaîne par polymérase)
PDB	Protein Data Bank
PDNA	Protein-deoxyribonucleic acid
PEG	Polyéthylène glycol
PNA	Peptide nucleic acids
POI	Protein of Interest
PX	Paranemic
QD	Quantum dot (boîtes quantiques aussi appelées Nano-cristaux semi-conducteurs)
QNM	Quantitative Nano-mechanical property Mapping
RH	Relative humidity
RMN	Résonance magnétique nucléaire
RMS	Root mean square
RNA	Ribonucleic acid
SAM	Self Assembly Monolayer
SCOP	Structural classification of proteins
SELEX	Systematic Evolution of Ligands by EXponential enrichment (évolution systématique de ligands par enrichissement exponentiel)
SEM	Scanning electron microscope
SH	Sulfhydryl group.
SNAP	Soluble NSF attachment protein, (NSF= N-Ethylmaleimide Sensitive Fusion)
SPIE	Society of Photo-Optical Instrumentation Engineers
SPR	Surface plasmon resonance
SPRI	Surface plasmon resonance imaging
STM	Scanning Tunneling Microscope
TAM	Tile Assembly Model
TBE	Tris/Borate/EDTA
TE	Tris/EDTA
TEM	Microscopie électronique en transmission
TM	Transverse magnétique
TX	Triple crossover (Motif ADN à triple croisement)
UV	Ultraviolet
VHH	Variable heavy-chain of antibody
WC	Liaisons Watson-Crick
WLC	Worm-like-chain

Introduction

INTRODUCTION

La biologie de synthèse est un domaine émergent en recherche qui combine science et ingénierie. Elle implique une panoplie d'approches, de méthodologies et de disciplines et un bon nombre de définitions existe. Elles ont en commun le dessin et la construction de systèmes et fonctions biologiques qui n'existent pas dans la nature. C'est aussi la reconstruction ou la transposition de systèmes existants afin de comprendre leurs mécanismes sous-jacents [1].

La création de nano-objets artificiels, multifonctions, capables de s'interfacer avec des supports organiques ou inorganiques constitue un défi, tant au niveau des difficultés technologiques à surmonter qu'au niveau des applications potentielles en nano-médecine, nanoélectronique, biocapteurs, nanomatériaux. Leur réalisation aborde les problèmes de liaison entre le vivant et non-vivant, l'implémentation de fonctions, la traçabilité ainsi que les difficultés de convergence de taille entre les différents objets. Le développement de ces objets fait appel aux propriétés relatives de différentes disciplines telles que la biologie, la physique, la chimie et l'électronique et s'inscrit dans le champ de la biologie de synthèse.

Le terme de biologie synthétique est apparu en 1912, avec la publication par le biologiste français S. Leduc de l'ouvrage « La biologie synthétique », dans lequel il affirme que pour tester la validité de nos connaissances en biologie, la fabrication, ou « synthèse », doit succéder à l'analyse [2]. Dans les années 1970, le biologiste moléculaire W. Szybalski introduisit ce terme lors de la découverte des enzymes de restriction, ouvrant la voie à la biologie de synthèse en tant que bio-ingénierie [3]. De même, en 1980, B. Hobom l'utilise pour décrire des bactéries génétiquement modifiées par ingénierie, c'est-à-dire des organismes biologiques modifiés par synthèse [4].

En 2000, E. Kool décrit une approche chimique de la biologie de synthèse avec la création de parties synthétiques qui fonctionnent dans des systèmes biologiques [5]. Plus largement dans ce sens, le terme fait référence aux efforts de redessiner les systèmes vivants, approche qui n'est pas étrangère à l'exobiologie [6-8]. Ceci constitue aussi une extension de la « chimie biomimétique » dans laquelle la synthèse organique est utilisée pour créer des molécules artificielles qui reproduisent le comportement de composants biologiques, tels les enzymes [9]. Toutefois, la biologie de synthèse a une portée plus large en ce sens qu'elle cherche à recréer dans des systèmes chimiques synthétiques les propriétés des systèmes biologiques, incluant l'hérédité et l'évolution [10]. Les scientifiques espèrent au travers de la biologie de synthèse comprendre les mécanismes sous-jacents à la biologie « naturelle ».

Plus récemment, l'ingénierie a donné un autre sens à la biologie de synthèse qui est vue comme une technologie. Les scientifiques cherchent à extraire des systèmes vivants des pièces interchangeables qui pourraient être testées, validées en tant qu'unités de construction, et réassemblées pour créer des dispositifs qui pourraient, ou non, avoir des analogues dans les

systèmes vivants [11]. Dans ce cas, les pièces sont d'origine biologique mais leur assemblage est synthétique. Un exemple de ces technologies est la re-conception du système de sécrétion de Type III utilisé par *Salmonella typhimurium* pour sécréter les protéines de soie d'araignée à la place de ses propres protéines naturelles infectieuses [12]. Dans le domaine d'ingénierie, D. Endy a développé les concepts de standardisation des composants, de découplage des problèmes et d'abstraction de l'information qui tentent d'être apportés à la biologie [13].

Une représentation réalisée par le groupe de D.N. Woolfson (figure 1) donne une bonne vision d'ensemble sur la biologie de synthèse avec ses différents points de départ. E.H.C. Bromley *et al.* y décrivent un « espace de biologie de synthèse » qui présente différents chemins d'accès récemment explorés selon le niveau de complexité des composants étudiés et le niveau de divergence de la nature [14].

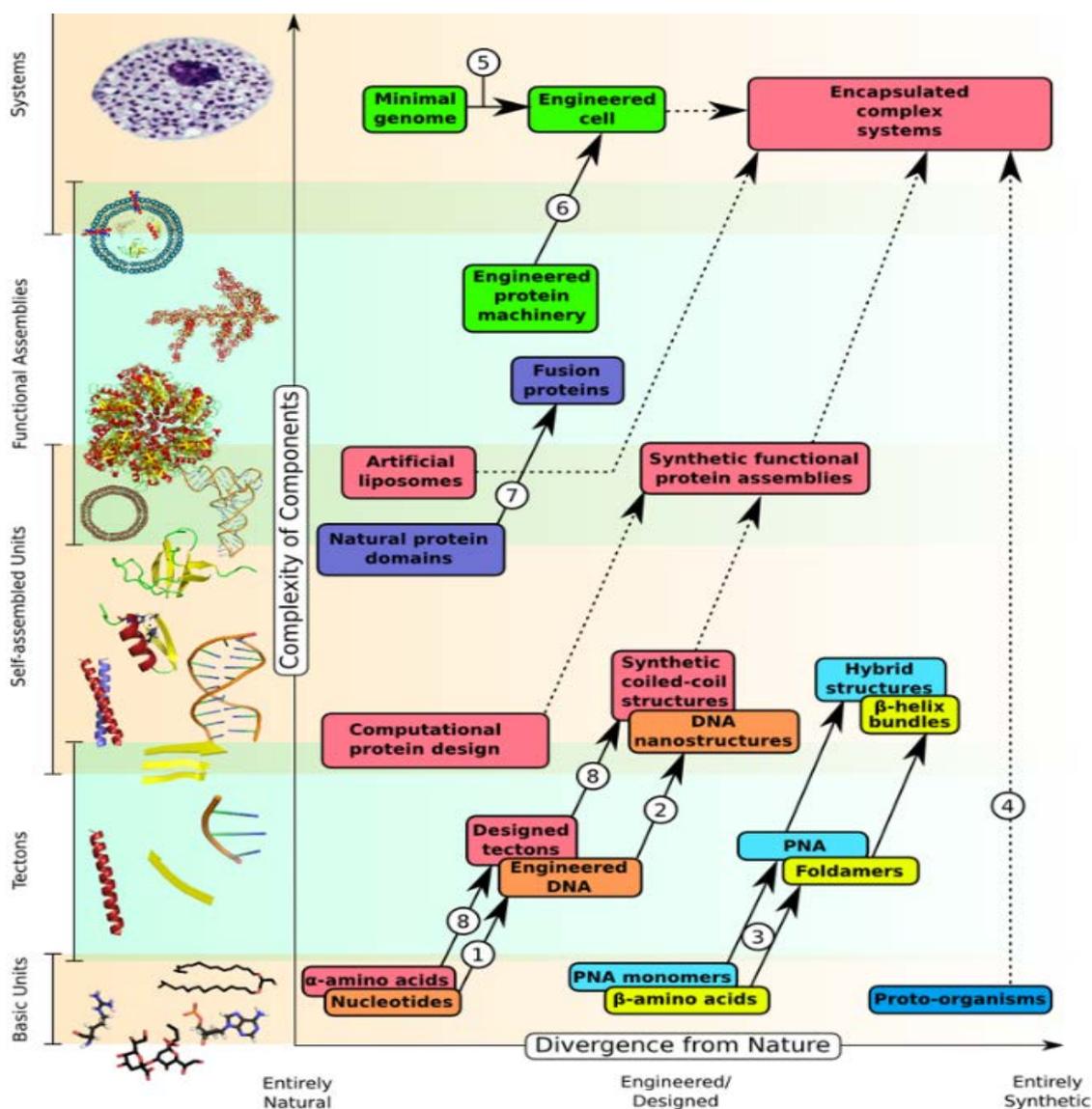


Figure 1 : Graphe de la biologie de synthèse. [14]

Les unités de base des molécules naturelles sont illustrées en bas à gauche du graphique, et selon notre capacité de synthèse de ces unités de base, elles évoluent le long de l'axe horizontal. Un défi de la biologie de synthèse est d'accroître le répertoire des chimies qui peuvent être utilisées de manière fiable et efficace dans les constructions. L'axe vertical illustre le niveau de complexité des composants allant des unités de bases qui s'assemblent en tectons¹, pour aboutir à des auto-assemblages fonctionnalisés ou au dernier degré de complexité : les systèmes. Par exemple, la voie 1 présente la construction d'ADN modifié, utilisé pour produire de nouvelles nanostructures (voie 2). Parallèlement, la voie 8 montre l'utilisation de la synthèse de peptides pour créer des briques de construction programmables qui peuvent à nouveau s'assembler en des composants protéiques fonctionnels.

Un autre point de départ de la biologie de synthèse est illustrée (voie 3) par le développement d'acides aminés et de paires de bases non standards qui peuvent être assemblés en des foldamers ou analogues d'ADN. Encore plus loin sur cet axe synthétique (voie 4), différentes études ont pour objectif de produire des briques de constructions complètement nouvelles capables de s'auto-répliquer [16].

Un autre point de départ de la biologie de synthèse dans l'échelle de complexité concerne les voies 5, 6 et 7. Au niveau cellulaire, des génomes entiers peuvent être synthétisés et transplantés dans des hôtes [17], au niveau des assemblages fonctionnels, des modifications peuvent apporter de nouvelles fonctions aux organismes [18], de même qu'au niveau inférieur, les protéines mutées ou fusion peuvent créer des assemblages fonctionnels synthétiques [19].

Ainsi, la biologie de synthèse est une science multidisciplinaire qui vise en particulier :

- l'ingénierie des systèmes biologiques avec la tentative de synthétiser des composants biologiques qui peuvent être assemblés pour créer des systèmes qui se comportent de manière prédictible ;
- la reconstitution de la vie avec la tentative de créer des systèmes vivants à partir de zéro et ensuite, de les doter de fonctions nouvelles et originales. Cette reconstitution du vivant afin d'imiter ses fonctions est un moyen d'appréhender les lois fondamentales de la vie ;
- la création d'une vie alternative avec l'utilisation de molécules synthétiques dans les systèmes vivants [1].

¹ Le terme tecton emprunté à la chimie supramoléculaire décrit des composants moléculaires souvent construits par polymérisation d'unités de bases et programmés pour se replier ou s'assembler en des structures 3D [15] M. Simard, D. Su, J.D. Wuest, Use of hydrogen-bonds to control molecular aggregation-self-assembly of 3-dimensional networks with large chambers, *J. Am. Chem. Soc.*, 113 (1991) 4696-4698.

PRESENTATION DU TRAVAIL DE THESE

L'objectif du travail de thèse est l'étude des mécanismes d'assemblage de nanostructures hybrides associant acides nucléiques et protéines au sein d'une même structure covalente (PDNA). Le couplage covalent de ces deux composants permet d'allier les propriétés d'auto-organisation de l'ADN aux propriétés fonctionnelles des protéines. Les PDNA développés associent le domaine redox du cytochrome b5 humain à une séquence d'ADN simple brin. Le cytochrome b5 démontre une forte absorption optique dans le visible qui dépend de son état d'oxydoréduction. Les structures développées sont par ailleurs taguées en C-terminal du domaine peptidique par une queue poly-histidine facilitant leur purification ou l'interaction avec un support.

Une première partie vise à comparer les propriétés d'assemblage des PDNA à celles des ADN correspondants. Nous avons développé un modèle d'assemblage de polymères linéaire à partir de briques de construction. L'assemblage de ces briques peut conduire soit à des produits finaux définis, comme dans le cas de la synthèse de gène, soit à une distribution de polymères de tailles aléatoires. Les processus combinatoires qui régissent ces assemblages ont été implémentés dans un algorithme permettant de décrire quantitativement les populations d'espèces générées au cours de la réaction de polymérisation. Les données expérimentales correspondantes ont été utilisées pour définir les paramètres du modèle en caractérisant les processus physico-chimiques qui contrôlent ces assemblages. La complexité des populations ne permettant pas de traiter le problème individuellement, la stratégie a été de définir les paramètres critiques sur des petits assemblages élémentaires et de déduire les règles d'extrapolation aux grandes structures, par ajustement des profils expérimentaux et simulés. L'approche globale s'est accompagnée d'une approche à un niveau moléculaire, au travers de l'observation et de la caractérisation des éléments par microscopie de force atomique (AFM).

La deuxième partie est consacrée à la synthèse et à la caractérisation des briques élémentaires de PDNA, ainsi qu'à l'étude de leurs assemblages en comparaison des ADN correspondants. Elle met en œuvre les outils développés dans la première partie. L'objectif est alors d'analyser l'influence des domaines protéiques sur l'assemblage des structures. La stabilité thermique des PDNA a été mesurée afin de déterminer les limites d'utilisation des complexes et leur polyvalence. Les propriétés de migration ont été analysées sur différents gels de séparation (agarose, polyacrylamide), et par tamisage moléculaire permettant d'analyser les phénomènes physico-chimiques impliqués dans les complexes protéine-ADN. Les vitesses d'hybridation des différents blocs PDNA ont été déterminées et comparées aux blocs analogues d'ADN. Les caractérisations ont été réalisées aux échelles globales macroscopiques des populations mais aussi aux échelles de la molécule unique, grâce à la microscopie de force atomique à haute vitesse (HS-AFM). Cette dernière technique permet également d'analyser la dynamique des objets sur une surface.

Ces deux premières parties sont complétées par l'analyse des mêmes mécanismes d'auto-assemblage non plus en solution mais au voisinage de surfaces et au sein d'une matrice poreuse. Ces conditions prévalent lors du multiplexage de l'assemblage des nanostructures sur puce et nous a amenés à analyser à la fois, l'influence d'une matrice poreuse de type dextran et des conditions hydrodynamiques au sein ou au voisinage de cette couche. La modélisation d'un milieu multicouche a été utilisée pour mettre en évidence les relations entre les données expérimentales obtenues par la technique de résonance des plasmons de surface en imagerie et les paramètres connus en solution.

La dernière partie du mémoire met en œuvre l'étude de solutions d'assemblage alternatives permettant de s'affranchir des problèmes de pièges cinétiques rencontrés lorsque les assemblages sont réalisés sous contrôle cinétique, loin de l'équilibre thermodynamique. Ceci a impliqué l'étude d'une structure particulière de l'ADN, le motif-i, qui présente des propriétés d'auto-assemblage contrôlables, non plus par la température mais, par modification du pH. Une approche physico-chimique détaillée a été réalisée à un niveau macroscopique afin de déterminer les paramètres impliqués dans les mécanismes d'association et de dissociation, réversibles dans des conditions douces, des motifs-i en structures supramoléculaires. Cette approche macroscopique, tout en permettant de décrire avec précision les premières étapes du mécanisme d'assemblage, n'a pas permis d'en aborder les étapes ultérieures, en particulier, au niveau du mécanisme de formation et de la structure des espèces de large taille. Afin d'accéder à cette information, nous avons à nouveau fait appel à la microscopie de force atomique, qui a conduit à des résultats surprenants et inexplicables sur la base des approches biochimiques et qui, même, viennent en contradiction avec les concepts et évidences structurales sur les motifs de base. Ceci nous a conduit à envisager que la structuration à grande échelle de cette classe d'ADN pourrait impliquer des mécanismes et des structures différentes de celles démontrées à petite échelle.

Durant le cours de ces travaux, j'ai aussi contribué pour la partie modèle biologique au développement et à la validation d'instruments, en particulier à l'implémentation d'un banc optique de mesure d'anisotropie de résonance plasmonique. Cet outil nous a permis de caractériser la formation de structures anisotropes, orientées soit par l'influence d'un champ électrique, soit de manière hydrodynamique. Ce travail réalisé en collaboration avec l'équipe de Michael Canva au Laboratoire Charles Fabry, à l'Institut d'Optique Graduate School (Université Paris Sud, CNRS, Campus Polytechnique à Palaiseau) a été valorisé dans un article. Il ne fera pas l'objet d'une description détaillée dans le cadre de ce mémoire mais est présenté en annexe.

CONCEPTION DE NANO-OBJETS A BASE D'ACIDES NUCLEIQUES ET DE PROTEINES

La biologie de synthèse, science multidisciplinaire, comprend deux grands courants [10, 20]. L'approche biomimétique qui crée des nano-objets bio-inspirés et l'approche biotechnologique qui utilise des dispositifs ou machines fabriquées par le vivant pour réaliser des objets hybrides. En effet, le courant biomimétique initie un processus naturel dont il oriente l'évolution afin de comprendre les mécanismes, de dégager les principes et d'inventer un analogue en vue d'applications [21-23]. Le courant d'ingénierie de la biologie utilise la nature comme une boîte à outils dans laquelle il emprunte des fonctionnalités qu'il veut standardiser, découpler et abstraire dans le but de les reprogrammer à ses fins, et donc de les maîtriser totalement en s'extrayant de la complexité et de l'évolution de la nature [13, 24]. Dans les deux courants, le vivant est « molécularisé » et perçu comme une collection de dispositifs fonctionnels.

Un des domaines de la biologie de synthèse correspond aux bio-nanotechnologies qui implique des objets avec des structures et des dispositifs à l'échelle nanométrique, c'est-à-dire entre 1 et 100 nanomètres et des outils qui permettent de manipuler, caractériser et comprendre les systèmes à ces échelles [25].

Les deux approches "top-down" et "bottom-up" [26] des nanotechnologies, aussi appelées voie descendante et voie ascendante, s'appliquent à la biologie de synthèse.

La démarche « top-down » des nanotechnologies consiste à réduire des dispositifs milli- ou micrométriques jusqu'à atteindre l'échelle nanométrique. Toutefois ces approches présentent l'inconvénient de dépendre d'équipements coûteux et ce, d'autant plus que la taille des motifs à réaliser est petite et de nécessiter des conditions d'utilisation particulières voire extrêmes telles que des températures cryogéniques, des pressions très basses ou l'accès à des salles blanches.

L'équivalent en biologie de synthèse consiste, à partir de systèmes vivants naturels préexistants, à les modifier par ingénierie dans des buts déterminés. Ainsi, dans le domaine de la conception cellulaire, les biologistes moléculaires et ingénieurs cherchent à assembler et construire de nouveaux chemins génétiques pour modifier les phénotypes de systèmes cellulaires vivants [27].

La démarche « bottom-up », à l'inverse de l'approche descendante, vise à organiser de nouvelles structures originales à partir d'éléments de taille nanométrique tels les atomes ou les composants biologiques soit par auto-assemblage, soit par croissance organisée. En biologie de synthèse, cette approche consiste à utiliser ou à redessiner les molécules de base de la vie, telles que les acides nucléiques [28], les acides aminés et les carbohydrates (unités

de base de la figure 1), pour créer des supra-structures déterminées avant de les insérer dans des systèmes vivants [5]. Dans le dernier cas, les nouveaux composants biologiques standardisés, connectés de façon appropriée au sein des organismes vivants, sont conçus pour ajouter un comportement désiré ou une fonction originale.

Le principe d'auto-assemblage qui est une organisation spontanée de la matière par un ensemble de forces attractives et répulsives permet de contourner les difficultés d'utilisation des propriétés physico-chimiques de la matière pour associer correctement toutes les unités [29, 30]. L'auto-assemblage biologique constitue une approche parallèle économique ne requérant ni équipements coûteux, ni conditions extrêmes, c'est pourquoi elle a largement inspiré les approches biomimétiques « bottom-up » pour le développement de composants nanométriques [31-33]. C'est une des stratégies importantes utilisée en biologie dans le développement de structures complexes fonctionnelles [34].

1. Auto-assemblage.

L'auto-assemblage se définit par la formation de structures ayant un haut degré d'organisation à partir des unités élémentaires elles-mêmes, sans intervention externe [35, 36]. C'est un mécanisme d'agrégation où l'assemblage est gouverné par la recherche du minimum d'énergie libre, avec l'intervention de liaisons non covalentes, comme les liaisons hydrogène, les interactions hydrophobes, les liaisons de van der Waals, les interactions électrostatiques, et parfois des coordinations métalliques. Dans les systèmes développés par J.M. Lehn, des liaisons covalentes réversibles peuvent aussi être mises en œuvre dans le processus d'auto-assemblage en plus des interactions non-covalentes, définissant une chimie dynamique dans la constitution de ses objets aussi bien au niveau moléculaire que supramoléculaire [37].

On peut distinguer l'auto-organisation intramoléculaire (par exemple repliement d'un domaine protéique) de l'auto-assemblage intermoléculaire qui implique plusieurs acteurs de même nature ou non (protéines, acides nucléiques, membranes, etc.). Dans le premier cas, une seule molécule est concernée et sa capacité d'adopter une structure dans l'espace va lui conférer des propriétés spécifiques. Dans le dernier cas, l'auto-assemblage s'effectue entre différents éléments. Ce sont les propriétés de reconnaissance entre les différentes structures qui vont permettre de former des supra-structures.

Le concept d'auto-assemblage est né dans les années 1960 à partir des études réalisées sur le virus de la mosaïque du tabac et de l'enzyme ribonucléase [38, 39]. L'expérience de reconstitution du virus a montré que toutes les informations nécessaires à son assemblage étaient présentes dans les parties constituantes, et l'expérience de renaturation de la protéine a mis en évidence que l'information de son repliement réside dans la séquence d'acide aminé avec une structure 3D régie par l'énergie libre de Gibbs minimale. Ainsi, les études sur le virus ont mis en évidence les qualités du mécanisme d'auto-assemblage suivantes :

- la minimisation de la quantité d'information génétique nécessaire à la création de supra-structures en utilisant une ou peu de sous-unités répétitives ;
- le contrôle de l'assemblage et du désassemblage des structures par de multiples liaisons non covalentes qui peuvent être modulées en réponse à des changements environnementaux ;
- L'autocontrôle des erreurs d'assemblage du fait de sous-unités défectueuses incapables de se lier à la structure globale ;
- L'efficacité de construction à partir de l'assemblage de briques plutôt que par construction directe [40-42].

Ceci est le cas d'un auto-assemblage strict qui implique un processus réversible conduisant à une structure stable. Toutefois, toutes les structures biologiques ne s'auto-organisent pas d'elles-mêmes. Certaines nécessitent des facteurs extérieurs pour faciliter ou diriger l'assemblage, ainsi que divers mécanismes de régulation des auto-assemblages. J.S. Lindsey propose une classification des différents procédés d'auto-assemblage [43] :

- les auto-assemblages stricts où toute l'information est contenue dans les sous-unités ou précurseurs moléculaires ; aucun facteur ou énergie n'est nécessaire à l'auto-assemblage. Les exemples classiques incluent le virus de la mosaïque du tabac, les oligomères d'ADN, les chélates de métal.
- les auto-assemblages irréversibles où le matériel de départ doit passer directement aux produits finaux sans former de défauts puisque la réparation est impossible. Les réactions de liaisons irréversibles conduisent à un produit sous contrôle cinétique, qui peut être différent du produit qui serait accumulé sous contrôle thermodynamique.
- les auto-assemblages qui dépendent de la modification d'un précurseur où des étapes additionnelles sont nécessaires. Un exemple est la liaison du GTP à la tubuline, la préparant ainsi à la polymérisation en microtubules.
- les auto-assemblages impliquant une modification ultérieure où le produit d'assemblage nécessite une modification, souvent due à un agent externe, pour acquérir sa compétence fonctionnelle ou sa stabilité définitive. On peut citer l'ensemble des modifications post-traductionnelles des protéines. Les post-modifications peuvent rendre irréversibles des processus d'auto-assemblages réversibles en verrouillant la structure assemblée.
- les assemblages assistés où un agent externe accompagne le mécanisme sans faire partie de la structure finale. C'est le cas des protéines chaperonnes qui inhibent la précipitation d'intermédiaires repliés de façon incorrecte ou partielle.
- les assemblages dirigés où un modèle temporaire participe à l'assemblage en tant qu'élément structurel sans apparaître dans le produit final. L'élément externe peut jouer un rôle thermodynamique en stabilisant l'association de sous-unités ou/et cinétique en guidant l'assemblage vers une réaction particulière.
- les assemblages utilisant une combinaison des procédés cités ci-dessus. Des alternances séquentielles d'auto-assemblage et de modifications irréversibles

conduisent à un procédé hautement contrôlé tels que dans l'auto-assemblage des bactériophages [44]. Ce sont des mécanismes d'auto-assemblages hiérarchisés.

Deux formes d'auto-assemblages se distinguent dans cette classification : les auto-assemblages parallèles où l'ensemble des éléments constitutifs sont présents dès le départ et les auto-assemblages séquentiels dans lesquels les éléments consécutifs sont amenés avec une séquence temporelle.

1.1. Auto-assemblage des protéines

La plupart des protéines, qui sont des polymères d'acides aminés présentant des groupes hydrophiles, hydrophobes et amphiphiles, se replient spontanément avec une précision atomique pour former des structures organisées en 3 dimensions allant de 1 à 100 nm. Elles présentent une diversité élevée de propriétés structurales, catalytiques, signalétiques et de reconnaissance.

La richesse de fonctionnalité des protéines provient des forces d'interactions réversibles telles que les interactions électrostatiques, les liaisons hydrogènes, les forces de van der Waals et les forces de solvation. Ces interactions sont très dynamiques et souvent fragiles du fait de leur relativement faible énergie, variant de moins de 5 kJ.mol⁻¹ pour les forces de van der Waals à environ 5-40 kJ.mol⁻¹ pour les liaisons hydrogènes, comparée aux liaisons covalentes qui peuvent avoir des énergies allant jusqu'à 400 kJ.mol⁻¹ [45]. Toutefois, si les liaisons non covalentes ne sont pas fortes de manière isolées, prises dans leur ensemble, elles gouvernent la conformation structurale de toutes les macromolécules biologiques et régissent leur interaction avec les autres molécules [46, 47]. Ces liaisons permettent aussi de casser plus facilement des formes intermédiaires incorrectes sous l'influence de facteurs externes (température, pH...), et ainsi d'éviter des pièges cinétiques. Cette dynamique est absente dans les auto-assemblages réalisés en conditions irréversibles qui impliquent une force élevée des interactions, et nécessitent de contrôler le taux d'erreur, pour aboutir à la structure cible [48, 49].

1.1.1 Mécanismes d'assemblages des protéines

La compréhension de l'auto-assemblage de structures à partir de sous-unités est capitale pour la nano-fabrication et parmi le nombre varié d'exemples d'auto-assemblage présents dans la nature, les protéines sont d'un intérêt particulier en terme d'organisation supramoléculaire [50]. En effet, les protéines se replient spontanément pour prendre leur conformation native. Ceci implique que leur structure primaire dicte leur structure tridimensionnelle. C.B. Anfinsen défend l'hypothèse «thermodynamique» selon laquelle les interactions déterminées par la séquence mènent à la conformation de plus basse énergie [38]. Le paradoxe de Levinthal illustre le fait que si la protéine échantillonnait de manière séquentielle toutes les conformations possibles pour atteindre sa configuration native correcte, il lui faudrait un temps bien supérieur aux échelles du repliement qui s'étendent de la

picoseconde à la milliseconde, parfois plus [51]. Cyrus Levinthal propose la théorie du paysage énergétique, dans laquelle une série de mécanismes séquentiels accompagne l'évolution du repliement vers l'état natif d'une augmentation de la stabilité conformationnelle. Toutefois, le paysage énergétique ou entonnoir de repliement peut être plus ou moins complexe avec des chemins de repliement multiples et une cinétique de repliement complexe. Il importe donc d'étudier des systèmes simples aussi bien que des protéines dont le repliement est plus complexe afin de comprendre les différents mécanismes de repliement au niveau atomique.

Six principaux modèles de repliement ont été proposés :

- Le modèle de nucléation – condensation où une étape de nucléation est suivie par une propagation rapide de la structure. Ce modèle prend en compte le caractère coopératif du repliement [52, 53].
- Le modèle de diffusion – collision où les hélices alpha et feuilletts bêta qui joueraient le rôle de centres de nucléation, se forment sur la chaîne polypeptidique puis diffusent et coalescent. On aboutit à des microstructures natives [54].
- Le modèle de repliement séquentiel et hiérarchique dans lequel plusieurs segments de structures sont formés et assemblés à différents niveaux suivant un chemin du repliement unique [55]. Schulz [56] a proposé que ce repliement soit hiérarchique : la nucléation est suivie par la formation de structures super - secondaires puis par celle des domaines.
- Le modèle modulable du repliement où les domaines d'une protéine sont les unités de repliement [57]. Ils se replient indépendamment et des intermédiaires du repliement (les modules structuraux) sont formés et s'assemblent pour aboutir à la structure native [58].
- Le modèle d'effondrement hydrophobe où la première étape du repliement est conditionnée par les groupements hydrophobes qui se rassemblent au cœur de la protéine. S'ensuit une phase plus lente où se forment de courts segments de structures secondaires. Cet effondrement conduit à la stabilisation de la structure native [59].
- Le modèle de la fermeture éclair hydrophobe dans lequel la formation de segments de structures secondaires est simultanée avec l'effondrement hydrophobe [60].

Aujourd'hui, l'exploration des mécanismes de repliement des protéines est réalisée par simulation de la dynamique moléculaire, devenue possible avec la puissance de calcul des ordinateurs. Les simulations prennent en considération le solvant et les atomes de la protéine. Ces études, permettant d'analyser en détail le paysage énergétique du repliement, constituent un outil prometteur pour améliorer la prédiction *ab initio* de la structure tridimensionnelle d'une protéine [61, 62]. En effet, à ce jour, même si certains facteurs favorisant le repliement des chaînes peptidiques sont connus, les règles établies ne permettent pas de prévoir complètement la structure des protéines repliées.

Baker *et al.* se démarquent dans le domaine de la conception de nouvelles protéines *ab initio*. Ils ont notamment développé une stratégie algorithmique (Rosetta) capable d'optimiser la séquence par rapport à l'énergie libre, associant mutagenèse *in silico* et prédiction de structure [63].

1.1.1.1 Notion de module structural

Lors de l'étude des protéines dans les années 1960, la complexité des séquences protéiques et des structures qui en émergent, posent le problème de la classification [64]. L'idée de domaines modulaires naît de deux observations : la confirmation de la constitution des protéines en hélices et en brins comme prédits par L. Pauling et R. Corey [65, 66] d'une part et le fait que des protéines différentes peuvent présenter des structures similaires d'autre part.

Les domaines modulaires représentent le niveau d'organisation minimum des protéines. Ils sont souvent construits à partir de plusieurs motifs ou structures secondaires, chacun composé d'hélices ou de brins, avec les motifs $\alpha\alpha$, $\beta\beta$, $\alpha\beta\alpha$, ces derniers plutôt instables individuellement, s'agrègent pour protéger les résidus hydrophobes. Les domaines modulaires ont tendance à être stables et à se replier tout seul, ce qui a permis de les isoler de protéines plus larges et de les caractériser par cristallographie. Les domaines modulaires sont structuraux mais ils sont également fonctionnels, leur combinaison de différentes façons permettant de construire des protéines différentes [67]. De cette manière, ils peuvent être présentés comme des briques de constructions telles des legos moléculaires. En biologie de synthèse, les chercheurs portent de manière croissante leur attention sur la recombinaison et la fusion de domaines entiers de la protéine par rapport à la manipulation de résidus au sein des protéines globulaires [68-71].

1.1.1.2 Une combinatoire infinie d'un nombre limité de blocs moléculaires.

Les approches systématiques en biologie structurale et l'observation générale ont mis en évidence que les protéines sont constituées d'un nombre limité de modules. Ce sont les unités fondamentales de la structure et de la fonction d'une protéine. Même s'il existe environ 100000 protéines différentes exprimées dans les systèmes eucaryotes, il y a beaucoup moins de domaines, motifs structuraux et repliements différents. Ceci est en partie dû à l'évolution, avec des gènes ou parties de gènes qui peuvent être doublés ou déplacés au sein du génome. Ainsi, un domaine protéique peut être transféré d'une protéine à l'autre, donnant alors à la protéine une nouvelle fonction. Le nombre de motifs protéiques différents étant limité, c'est la combinatoire de l'utilisation de ces motifs qui va créer la diversité des protéines observées chez les organismes vivants [72].

Les domaines qui sont formés par l'assemblage de différents motifs en des unités compactes, locales et semi-indépendantes représentent la structure tertiaire des protéines.

Celle-ci peut être divisée en quatre classes selon les structures secondaires des domaines [73] :

- les domaines α avec un noyau constitué exclusivement d'hélices α ,
- les domaines β avec un noyau comprenant des feuillets β anti-parallèles,
- les domaines $\alpha+\beta$ qui comprennent les motifs α et β ,
- les domaines α/β constitués d'une combinaison des motifs $\beta\alpha\beta$ qui forment des feuillets β antiparallèles entourés d'hélices α amphiphiles.

Les protéines peuvent être regroupées par familles d'après les domaines qu'elles comportent. Deux bases de données SCOP (Structural Classification Of Proteins) et CATH (Class, Architecture, Topology, Homologous superfamily) sont très utilisées dans la classification des domaines protéiques en fonction de leurs structures. Toutes deux organisent les repliements de façon hiérarchique. Elles ont été développées avec l'intention de mettre en évidence les relations phylogénétiques entre protéines. Elles reposent sur l'observation selon laquelle la structure est généralement mieux conservée que la séquence du fait des pressions évolutives exercées pour le maintien de la fonction [74].

SCOP classe les repliements selon quatre niveaux hiérarchiques [75] Figure 2A: (i) les familles (identité de séquence), (ii) les superfamilles (similarité de structure et de fonction), (iii) les repliements (structures secondaires) et (iv) les classes de domaines (tout-alpha, tout-Beta, Alpha/Beta, Alpha + Beta).

CATH organise également les domaines en fonction de leurs similitudes de séquence puis de structure, avec cinq niveaux de hiérarchisation [76] Figure 2B: (i) la classe (composition en structures secondaires et leur arrangement), (ii) l'architecture (arrangement des structures secondaires), (iii) la topologie (structure générale du cœur du domaine et connectivité des éléments de structures secondaires), (iv) les superfamilles homologues (domaines pouvant partager un ancêtre commun), (v) les familles de séquences (domaines présentant des similarités de séquences significatives).

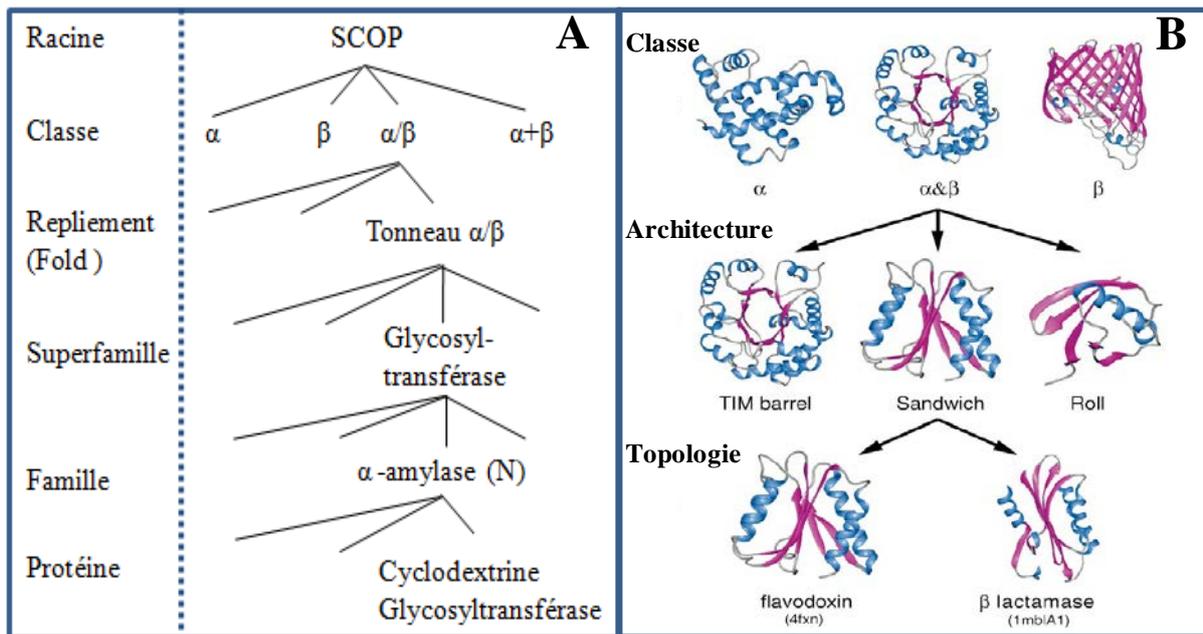


Figure 2 : Représentation des classifications SCOP et CATH. A, Hiérarchisation SCOP illustrée par un exemple [77], B, Présentation des trois premiers niveaux Classe, Architecture, Topologie de la classification CATH. Les hélices sont représentées en bleus, les feuilletts en magenta et les boucles en gris [76].

Les deux grandes structures secondaires constituant la structure générale des protéines sont l'hélice alpha et le feuillet bêta [78]. La combinaison de ces structures aboutit à des motifs structuraux (ou structures supersecondaires) telles que l'épingle à cheveux, la clé grecque [79], les unités $\beta\alpha\beta$, le motif hélice-boucle-hélice, le méandre β , le tonneau α/β présentés sur la Figure 3. Chaque motif structural est caractérisé par un type particulier d'interactions hélice-hélice [80].

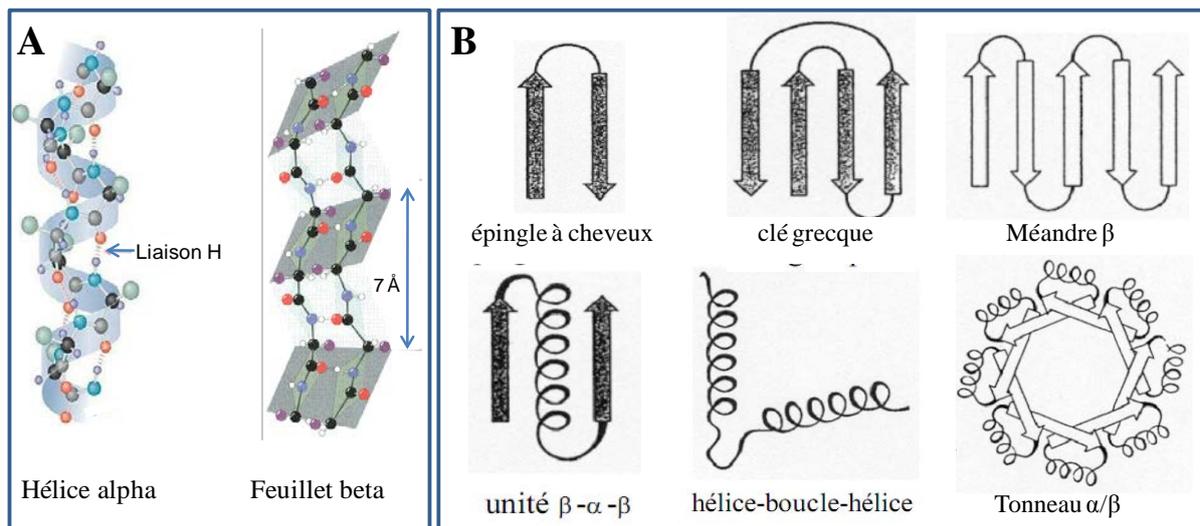


Figure 3 : Structures secondaires et supersecondaires. A, représentation de l'hélice alpha à gauche et du feuillet beta à droite [81, 82]. B, les hélice alpha et feuillet beta s'organisent spatialement en structures supersecondaires [83].

1.1.1.3 Assemblages supramoléculaires

Les composants moléculaires des cellules peuvent former des structures unidimensionnelles (fibres), bidimensionnelles (membranes) et tridimensionnelles (cristaux) [35]. Ces structures peuvent posséder différentes fonctions : mécaniques tels que les filaments du cytosquelette, enzymatiques ou informationnelles. Les structures supramoléculaires comprennent des protéines globulaires ou fibreuses. La formation de structures supramoléculaires peut suivre des processus :

- d'agrégation, réorganisation d'agrégats (exemple de la formation de microsphères entre l'alpha-lactalbumine et le lysozyme),
- de diffusion, telle que l'élongation des filaments de dimères de myosine musculaire qui est une réaction bi-moléculaire limitée par la diffusion.

1.1.1.3.1 Protéines globulaires.

Les protéines "globulaires" ont une structure tridimensionnelle assimilable à une sphère. Elles possèdent un "cœur" hydrophobe et une surface hydrophile, l'ensemble adoptant une structure compacte, dense. La stabilité des ponts disulfures est dictée par la stabilité de la structure de la protéine. Ils se forment plus facilement à l'intérieur des protéines globulaires, parce qu'ils nécessitent des conditions oxydantes et que le cytoplasme est généralement réducteur à cause de la présence de glutathione. À l'abri des molécules d'eau qui les hydrateraient, les acides aminés chargés enfouis dans les protéines globulaires peuvent former des ponts de sel entre eux. Les acides aminés hydrophobes peuvent aussi plus facilement s'y empiler pour former des régions dépourvues d'eau.

Un exemple de l'auto-assemblage des protéines globulaires est celui de l' α -lactalbumine et le lysozyme qui interagissent pour former un hétérodimère, lesquels peuvent s'assembler en nanosphères de 20 à 50 nm, lesquelles s'agrègent et s'organisent en chapelets qui se réorganisent pour former une microsphère de 3 à 4 μm (Figure 4) [84].

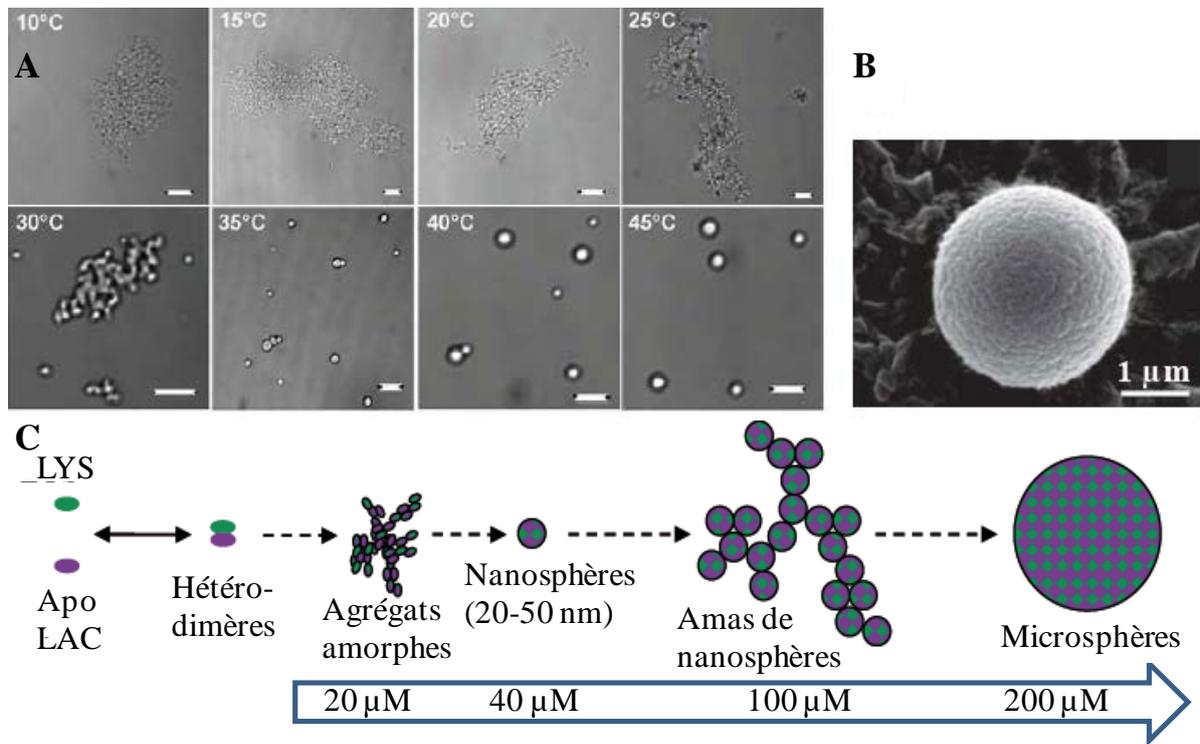


Figure 4 : Auto-assemblage des protéines globulaires : α -lactalbumine et lysozyme.

A, objets supramoléculaires résultant de l'assemblage du lysozyme et de l'apo α -lactalbumine en fonction de la température. Observations par microscopie optique, la barre d'échelle correspond à 5 μm (10-25°C) et 3 μm (30-45°C) [85]. B : Microsphères formées à 45°C observées par microscopie électronique à balayage [84]. C, Mécanisme hypothétique d'assemblage de l'apo α -lactalbumine (apoLAC) et du lysozyme (LYS) en microsphères selon des observations réalisées par microscopie électronique à transmission à différentes concentrations totales en protéines après incubation 30 min à 45°C de mélanges équimolaires [86].

1.1.1.3.2 Protéines fibreuses

Les protéines fibreuses, aussi appelées protéines structurales, constituent le principal matériau de construction chez les Vertébrés. Les fibres protéiques sont constituées de la répétition d'éléments simples qui s'associent en "fils". Linéaires, elles sont insolubles dans l'eau et d'une grande stabilité.

On peut distinguer deux types de structures fibreuses :

- celles dont la conformation est imposée par le mode d'organisation,
- celles créées par auto-assemblage modulaire.

Le premier cas peut être illustré par le collagène dont la forme allongée est due à l'assemblage d'une triple hélice formée de trois chaînes polypeptidiques. Les micro-fibrilles qui sont les unités structurales s'agrègent de façon telle que la structure brute possède une force et une élasticité spécifique. Le deuxième cas peut être illustré par les microtubules d'actine ou de tubuline qui par assemblage modulaire se polymérisent et se dépolymérisent par ajout ou suppression d'unités élémentaires (Figure 5).

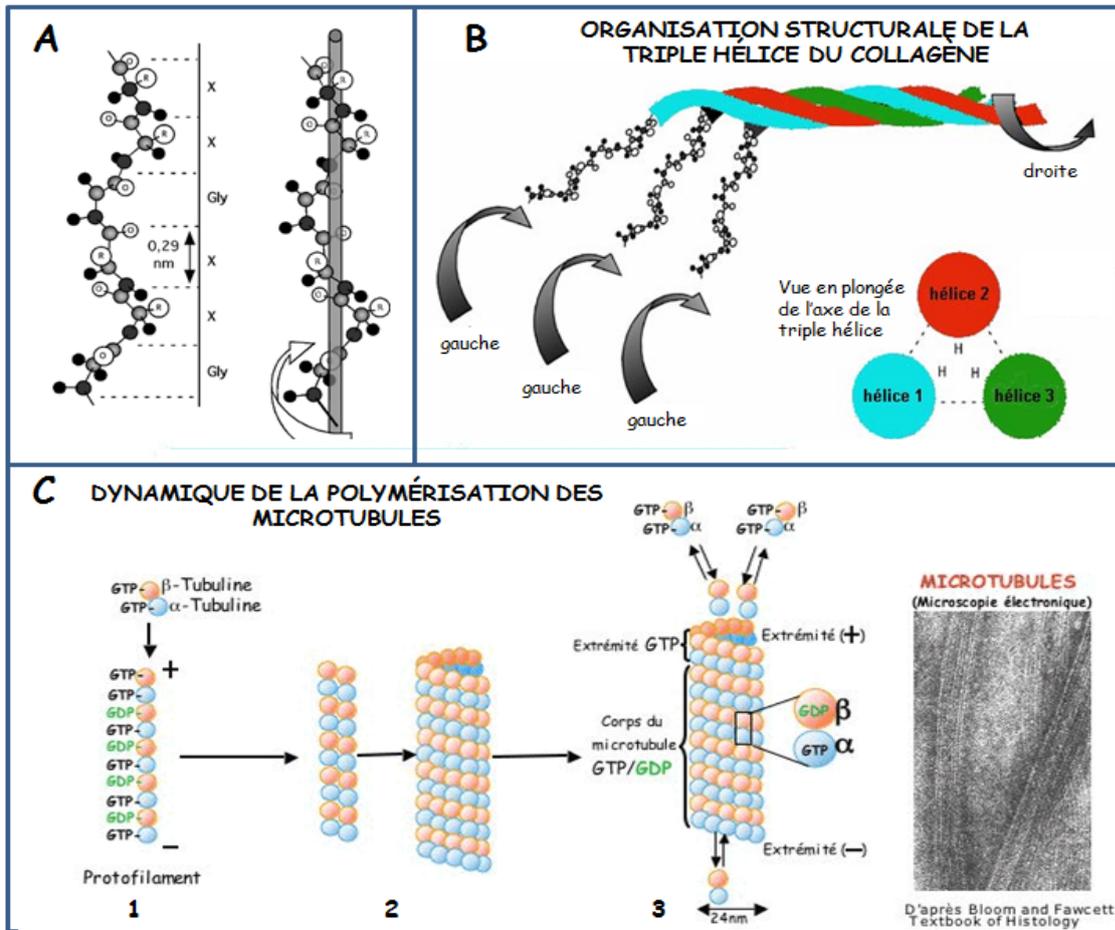


Figure 5 : Structures secondaires et quaternaires du collagène, et auto-assemblage de la tubuline en microtubules.

A, l'hélice de collagène est de pas gauche avec une glycine présente tous les trois acides aminés; l'hélice est aussi très riche en proline et en hydroxyproline. A l'inverse d'une hélice α , il n'y a pas de ponts hydrogènes stabilisants entre les résidus d'une même hélice de collagène. L'hélice de collagène s'élève de 0,29nm par résidu et elle compte un peu moins de trois résidus par tour alors que l'hélice α ne s'élève que de 0,15nm par résidu et compte 3,6 résidus par tour. B, la structure quaternaire du collagène est une microfibrille constituée de trois sous-unités, les trois hélices tournant l'une autour de l'autre en une super-hélice orientée dans le sens de la main droite [87]. C, auto-assemblage en structure d'ordre supérieur de la tubuline en microtubules. Les microtubules sont des structures polaires comme l'actine des microfilaments avec une extrémité (+) à croissance rapide dirigée vers la périphérie de la cellule et une extrémité (-) qui est associée au centrosome. Le centrosome est un complexe protéique situé près du noyau et il est constitué de deux centrioles eux-mêmes constitués de tubuline α , β , γ , δ et ϵ . (1) Assemblage du protofilament, (2) association latérale de protofilaments, (3) élongation du microtubule [88].

Les microtubules sont constitués de dimères de tubuline α et β . Chaque monomère peut s'associer au GTP mais seule la tubuline β peut échanger son GTP. C'est une protéine globulaire comme l'actine des microfilaments, à la différence des filaments intermédiaires. L'assemblage des dimères de tubuline en une structure microtubulaire se fait en plusieurs étapes : (i) la polymérisation de dimères de tubuline α et β (chargées de GTP) qui s'associent tête bêche pour former un protofilament, (ii) la formation d'un fragment de microtubule par association latérale de 10 à 15 protofilaments et repliement du feuillet pour donner une structure rigide, (iii) l'élongation du microtubule par polymérisation (ajout de dimères) à l'extrémité (+). Les microtubules sont des structures dynamiques. Dans une cellule, il y a en permanence et à vitesse variable (quelques secondes ou quelques minutes) plusieurs centaines de microtubules en cours de polymérisation et de dépolymérisation.

1.1.2 L'auto-assemblage comme outil de construction moléculaire

Dans le contexte d'ingénierie de nanostructures, les propriétés d'auto-assemblage des protéines sont largement exploitées utilisant un certain nombre de motifs à différentes échelles. Celles-ci vont des unités primaires de quelques nm, en passant par les assemblages secondaires (tubes, hélices, feuillets, ruban...) de quelques nm à μm , jusqu'aux assemblages tertiaires (doubles hélices, fibres, gels...) supérieurs au micromètre [89-93].

De nombreuses nanostructures supramoléculaires ont été réalisées à base de peptides qui, chimiquement versatiles, sont très adaptables et capables d'adopter différentes structures secondaires selon les acides aminés qu'ils contiennent [94]. Ces derniers apportent selon leur chaîne latérale des propriétés de chiralité, des dipôles électrostatiques, des groupes ionisables, des donneurs et accepteurs de liaisons hydrogènes, des ligands de liaison au métal et des domaines hydrophobes [95].

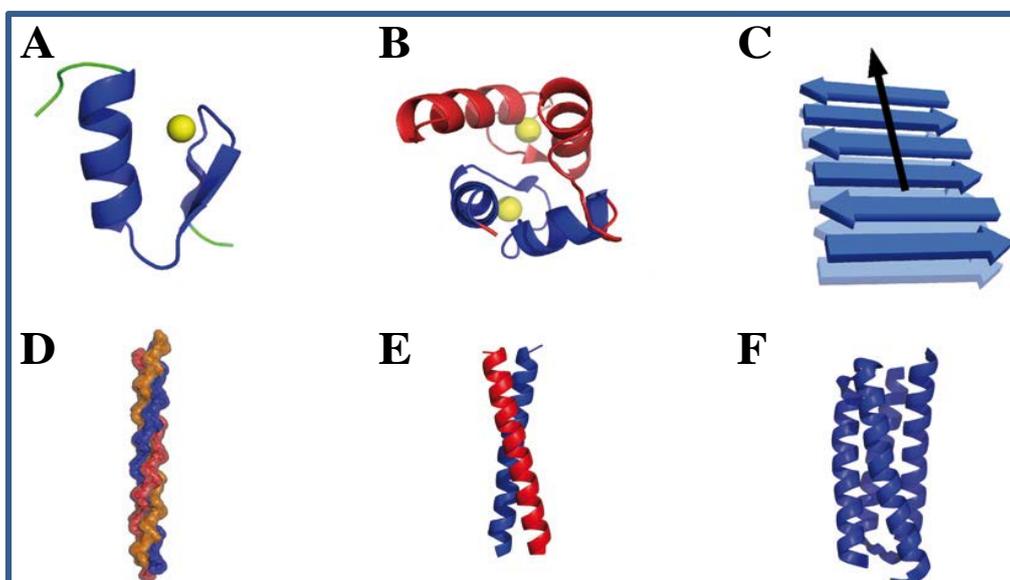


Figure 6 : Quelques motifs de repliement des protéines utilisés dans la conception *de novo*.

- A, structure doigt-zinc type C2H2 en coordination avec un ion zinc (PDN code 1ARD),
- B, deux domaines hélice-boucle-hélice covalents en coordination avec des ions calcium (PDB code 1B1G),
- C, dessin de l'assemblage en feuillets beta de fibrilles type amyloïdes, dont l'axe est pointée par le flèche noire,
- D, trois polypeptides associés pour former une triple hélice de collagène (PDN code 1K6F),
- E, le domaine de super-hélices (PDN code 2ZTA),
- F, faisceau à quatre hélices (PDN code 1ROP). Figure extraite de [96].

Les domaines autonomes tels que les doigts de zinc (Figure 6A) contiennent l'information nécessaire pour plier les peptides en des unités auto-assemblées fonctionnelles, distinctes, compactes [97]. Les domaines hélice-boucle-hélice (Figure 6B), proches des motifs en doigt zinc, car petits (30 acides aminés), nécessitant un ion métallique pour se replier et présents dans un grand nombre de protéines, sont intéressants du point de vue conception du fait de leur capacité d'oligomérisation. Actuellement les scientifiques cherchent à créer des biomatériaux à partir de ces structures [96] même si ce champ d'investigation est largement dominé par l'assemblage des peptides formant des feuillets beta en des configurations de type amyloïde

(Figure 6C) [98]. Plusieurs groupes ont réussi à réaliser des peptides qui s'assemblent en triple hélice de collagène [99, 100] et en fibres de taille allant du nano au micromètre [101]. Les super-hélices (Figure 6E) dont les relations séquence-structures sont les plus comprises sont très utilisées dans la conception de nanostructures avec des unités fonctionnelles [102]. De plus, leurs règles d'assemblage apparaissent robustes et transférables à différents contextes de super-enroulement. De la même manière les motifs de faisceaux de quatre hélices (Figure 6F) sont utilisés dans l'objectif d'apporter de nouvelles fonctions [103].

Par ailleurs, les stratégies d'ingénierie des protéines qui manipulent l'organisation des domaines et modules, en mimant la recombinaison non homologue, sont hautement efficaces pour la production de fonctions complexes et sophistiquées qui interviennent dans la reconnaissance et régulation moléculaire [104]. De nombreux exemples de conception et de production de protéines artificielles bio-inspirées illustrent l'intérêt de cette approche [105, 106]. Des protéines avec différentes symétries d'oligomérisation ont été fusionnées pour créer une gamme d'assemblages supramoléculaires incluant des cages [107], des filaments [108], des réseaux [109] et des cercles [110].

Les peptides sont également utilisés dans la création de fonctions nouvelles en reconstituant un processus évolutif. La méthode de banques combinatoires de peptides à la surface de bactériophages (phage-display) génère des populations de molécules de 10^7 à 10^{10} objets, capables de reconnaissance moléculaire de produits non biologiques, grâce à l'utilisation de plusieurs cycles d'amplification-sélection [111, 112].

En conclusion, même si les protéines représentent des briques moléculaires qui ont le plus de potentialités en termes de structures ou de catalyse, il reste actuellement difficile, malgré quelques exemples [63, 113] de créer des structures ab initio du fait d'une trop grande complexité. Nous savons maîtriser des assemblages en utilisant des protéines modulaires qui ont déjà leur fonctionnalité mais fabriquer des structures supramoléculaires qui présentent des propriétés spécifiques à partir d'éléments primaires reste un défi réalisé dans un nombre limité de cas précis [113]. C'est la raison pour laquelle les scientifiques se sont intéressés à la molécule d'ADN qui, du fait de sa taille nanométrique, de ses propriétés mécaniques stables et de sa reconnaissance spécifique hautement prédictible, représente une brique de construction moléculaire de choix en nanotechnologie.

1.2. Auto-assemblage des acides nucléiques

1.2.1 Un peu d'histoire

Si l'existence des acides nucléiques est connue depuis 1868, découverts par Miescher [114], c'est seulement en 1944 que leur rôle en tant que support et vecteur de l'information génétique fut démontré [115]. La structure de l'ADN sous sa forme B a été élucidée en 1953 par J.D. Watson et F.H.C. Crick [116, 117]. S'appuyant sur les photographies de diffraction aux rayons X de M. Wilkins et R Franklin, ils proposent un modèle en double hélice avec les deux brins reliés par des bases complémentaires A - T et G - C, en proportion équivalente selon les règles de E. Chargaff [118]. Watson et Crick décrivent la structure de l'ADN dans sa forme la plus usuelle, l'ADN B mais il existe également d'autres formes d'ADN selon la séquence et la concentration ionique des milieux qui seront décrites plus loin (Tableau 1). Ces ADN forment la structure en double hélice la plus répandue, toutefois des structures en triple et en quadruple hélices existent aussi. Nous allons d'abord présenter les trois conformations d'ADN formant une double hélice puis les autres structures d'ADN.

L'ADN adopte diverses conformations, selon son état d'hydratation. La **forme B**, physiologique, est hydratée et représente la structure générale de l'ADN dans les conditions des cellules vivantes. La **forme A**, déshydratée, a été observée dans certaines spores bactériennes, en réponse à la dessiccation du milieu mais également au niveau des hybrides ADN/ARN lors de la transcription. Cette forme, initiée par l'ARN qui n'adopte qu'une conformation A, est transitoire puisqu'à la fin de la transcription, lorsque l'ARN s'est détaché, l'ADN reprend sa conformation B. *In vitro*, lorsque la teneur en eau de la préparation d'ADN est réduite, l'ADN évolue réversiblement de la conformation B vers A. La **forme Z**, également une forme alternative de l'ADN B, a été observée lors de la transcription des gènes, au site d'initiation de la transcription, près des promoteurs des gènes activement transcrits. Pendant la transcription, l'ARN polymérase induit un super-enroulement négatif en amont du site de transcription qui favorise la formation d'ADN Z. Une fonction de l'ADN Z serait d'absorber les contraintes du super-enroulement négatif et de favoriser l'interaction des bases avec les protéines régulatrices. En fin de transcription, la topoisomérase relaxe l'ADN en conformation B.

Il existe encore d'autres formes transitoires issues de l'ADN B, telles que les ADN triplex et les jonctions Holliday qui sont des structures branchées. La **jonction de Holliday** tire son nom de Robin Holliday, qui l'a proposée en 1964 pour expliquer un type particulier d'échange d'information génétique observé chez des levures : la recombinaison homologue [119]. L'**ADN H**, ainsi nommé pour les appariements non-canoniques des acides nucléiques proposés par Karst Hoogsteen en 1963 [120], pourrait avoir un rôle dans la régulation fonctionnelle de l'expression des gènes ainsi que sur les ARN tel que la répression de la transcription. L'ADN H génère la formation de triple hélice d'ADN illustrée sur la Figure 8.

En plus de ces formes classiques, il existe des formes particulières et stables d'ADN telles que l'ADN G4 et le motif-i. L'ADN G4 observé près des promoteurs des gènes et au niveau des télomères [121-123] aurait un rôle de type régulateur dans la méiose et dans le phénomène de recombinaison. Cette structure étudiée de manière détaillée *in vitro*, se forme dans des conditions de salinité et de pH physiologiques [124]. Il a été montré que de nombreuses protéines étaient capables de fixer, stabiliser, ou au contraire dérouler ces structures *in vitro* [125]. Le motif-i est constitué de séquences riches en cytosines. Ces séquences sont présentes dans les télomères, mais à ce jour, aucun rôle fonctionnel de régulation n'a été démontré. Il a été trouvé que les séquences des extrémités télomériques peuvent adopter la configuration du motif-i [126, 127].

1.2.2 Les différentes formes naturelles d'ADN

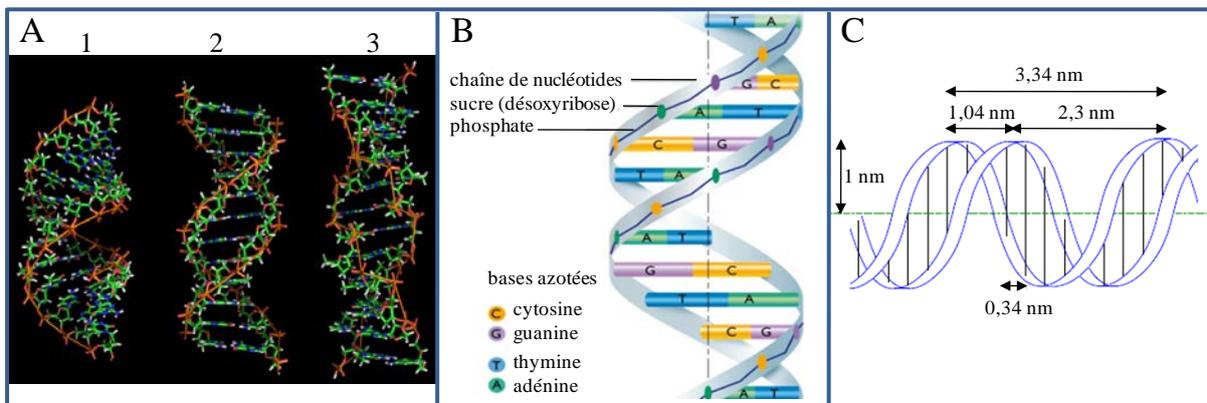


Figure 7 : Les différentes formes d'ADN et caractérisation de l'ADN B.
A, structures A, B et Z de l'ADN présentées respectivement aux positions 1, 2 et 3.
B, schéma de la double hélice d'ADN B avec la présentation des bases nucléotidiques appariées.
C, caractéristiques géométriques de la double hélice d'ADN B [122, 128]

1.2.2.1 ADN B

L'ADN B est constitué de deux brins antiparallèles qui forment une double-hélice (Figure 7). Les brins d'ADN sont reliés entre eux par les appariements entre les quatre bases azotées adénine, thymine, guanine et cytosine. La structure en double hélice est stabilisée par les liaisons hydrogène entre les paires de bases mais aussi par les interactions d'empilement de type π - π qui existent entre les cycles aromatiques des bases. Les bases, plus hydrophobes, sont à l'intérieur de l'hélice et les phosphates, hydrophiles, sont exposés à l'extérieur et peuvent interagir avec le solvant.

Le plan des bases est perpendiculaire à l'axe de l'hélice alors que le plan des résidus de sucre lui est parallèle. La rotation entre deux paires de base est de 36° . La double hélice droite a un pas de 3,4 nm, chaque tour étant constitué de 10 nucléotides sur un brin. Son organisation fait apparaître un grand et un petit sillon.

1.2.2.2 ADN A

Observé en premier lors de la formation de cristaux, l'ADN A est également une hélice droite mais elle présente une forme beaucoup plus condensée (Figure 7). Chacun des tours comprend 11 paires de bases, elle a un pas de 2,3 nm. Le plan des paires de bases n'est plus perpendiculaire à l'axe de l'hélice, mais forme un angle d'environ 75° avec celui-ci [129, 130]. Il en résulte un déplacement de l'axe de l'hélice qui ne passe plus par le centre de l'appariement des bases, mais à l'intérieur du grand sillon. Ceci induit une augmentation du diamètre de l'hélice qui passe d'environ 20 Å pour l'ADN en forme B à environ 26 Å pour la forme A [131].

1.2.2.3 ADN Z

L'ADN Z est une forme alternative de l'ADN B qui concerne les séquences de pyrimidines et purines alternées et particulièrement les séquences GC alternées qui, à haute force ionique, transitent en ADN Z, ce phénomène étant sous contrôle électrostatique. L'ADN Z est une double hélice gauche avec un seul sillon, en forme de zigzags (d'où son nom) dessinés par le squelette phosphates/pentoses. Le pas de l'hélice est un peu plus faible que celui de la forme B (30° au lieu de 36°) avec 12 paires de bases par tour. Les sucres sont alternativement 3'-endo syn et 5'endo anti.

Tableau 1 : Comparaison de quelques caractéristiques des trois formes géométriques A, B et Z de l'ADN.

	forme A	forme B	forme Z
Sens de l'hélice	droite	droite	gauche
grand sillon / petit sillon	grand / écrasé	grand / petit	sillons équivalents
Paires de bases par tour d'hélice	11	10,5	12
Distance d'empilement	2,3 Å	3,3 Å	3,8 Å
Pas par tour de l'hélice	28,2 Å	33,2 Å	45,6 Å
Tour moyen de l'hélice	+18 °	+16 °	0 °
Diamètre	26 Å	20 Å	18 Å

Sources: [122, 132, 133]

1.2.2.1 ADN cruciforme et ADN en épingle à cheveux

Une jonction de Holliday est un complexe formé de quatre brins d'ADN qui conduit à deux doubles hélices d'ADN recombinées. D. N. Gopaul *et al.* [134] ont déterminé la structure cristallographique de 2 jonctions Holliday en 1998 (voir Figure 8A, B).

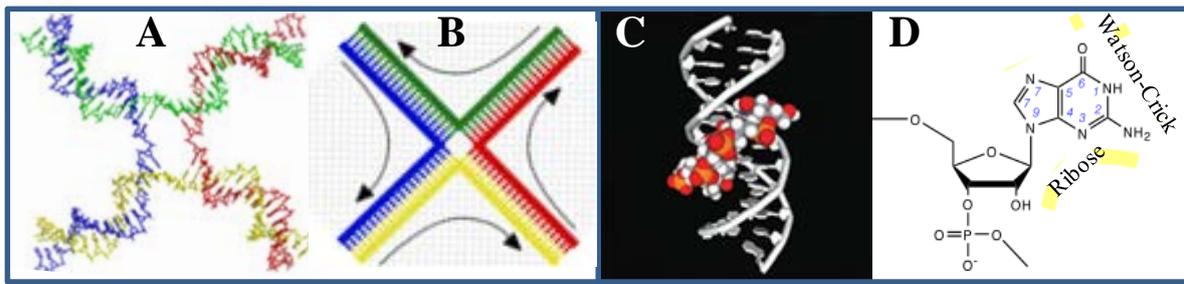


Figure 8 : ADN cruciforme et ADN triplex.

A, Structure moléculaire d'une jonction Holliday (fichier PDB, 3CRX), B, représentation de la jonction Holliday [135]. C, ADN triplex formé à partir d'un duplex hélicoïdal de 14 paires de bases (dessin en tubes) et d'un ADN simple brin de six bases remplissant l'espace dans le grand sillon (fichier pdb 1bwg) [136]. D, présentation des faces Watson-Crick et Hoogsteen de la guanosine [137].

Les jonctions Holliday sont des formes transitoires de l'ADN B, lorsque deux brins homologues coupés, se croisent et se rapparient. Des répétitions inversées de segments ADN polypurine/polypyrimidine formant des palindromes peuvent également produire des structures cruciformes ou en épingle à cheveux, par appariement intra-brin.

1.2.2.2 ADN-H ou ADN triplex

La structure ADN triplex est formée quand une molécule d'ADN simple brin vient s'apparier dans le grand sillon d'une double hélice d'ADN [138, 139]. On obtient alors un ADN triple brins plus un simple brin (Figure 8C). Ceci est rendu possible par l'appariement de Hoogsteen qui est un type de paire de base rare, observé dans les acides nucléiques. Expérimentalement, ces appariements ont été observés pour la première fois dans la structure cristallographique de l'ARN de transfert, en 1974 [140, 141].

Les liaisons de Hoogsteen impliquent une face de la purine différente de la face Watson-Crick et en particulier les positions N7 et le groupement oxo ou amino en position 6 comme illustré Figure 8D.

Comme l'appariement Hoogsteen utilise une face de la purine qui est indépendante de la face Watson-Crick, il est possible à une purine d'effectuer simultanément les deux types d'interactions. Ceci permet de former des structures à trois brins, dont deux forment une hélice canonique. Le troisième brin vient se loger dans le grand sillon de cette hélice et forme des paires Hoogsteen avec le brin d'ADN ou d'ARN contenant des purines. Cette structure ne peut se former que dans les régions où le duplexe contient une séquence de plusieurs purines consécutives.

1.2.2.3 Quadruple hélice d'ADN : les quadruplex de G

Les quadruplex de G sont des structures à base d'acides nucléiques formés par des séquences riches en guanines. Ces structures multibrins sont maintenues ensemble par des plateaux coplanaires de quatre guanines, les quartets de G, appariées par des liaisons hydrogène de type Hoogsteen (Figure 9A, B). Les quadruplex de G formés à partir d'un bloc

d'au moins deux tétrades de guanine empilées, peuvent être constitués à partir de un, deux ou quatre oligonucléotides d'ADN (ou d'ARN) [142]. Ils présentent une grande variété de topologies qui sont la conséquence de combinaisons variables de l'orientation des oligonucléotides, leur longueur et leur séquence. Les structures G-quadruplex ont été caractérisées pour la première fois en 1962 par diffraction de rayon X [143].

En plus des liaisons hydrogènes et des forces d'empilement entre les bases, les quadruplex sont stabilisés par la présence d'un cation positionné au sein de la cavité entre les niveaux de tétrades qui interagissent avec les huit guanines environnantes. La stabilité du quartet est très différente selon le cation monovalent considéré. La capacité de stabilisation des ions se classe de la façon suivante, $K^+ > Na^+ > Li^+ \approx Cs^+$ [144]. Elle est liée à l'enthalpie d'hydratation des ions, c'est-à-dire l'énergie nécessaire pour détacher les molécules d'eau d'un ion.

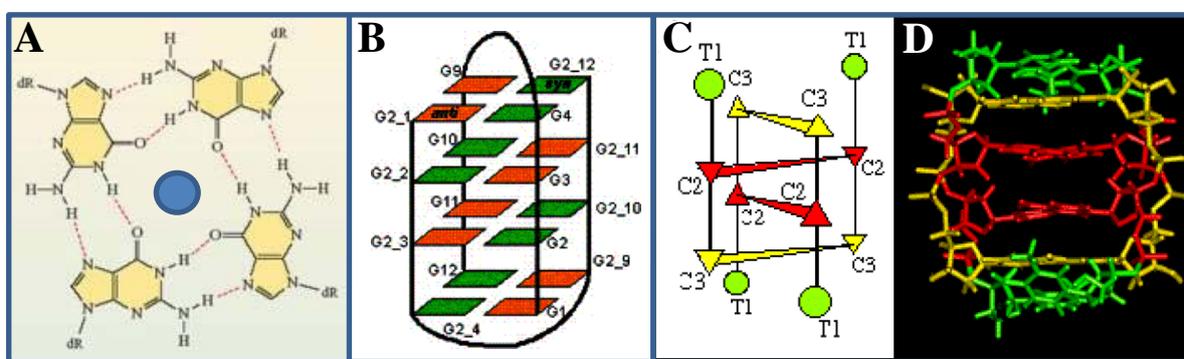


Figure 9 : Structures d'ADN G et d'ADN I.

La grande différence entre les quadruplex hélices de quartets de G et de motif-i est l'appariement coplanaire de quatre plateaux de guanines dans le premier cas (B) par rapport à l'intercalation des paires C.C+, dans le deuxième cas (C).

A, structure d'un quartet de Guanine, aussi appelé tétrade, qui est un plateau aromatique carré de quatre guanines appariées par des liaisons de Hoogsteen. B, schéma montrant le repliement dans le quadruplex intermoléculaire formé par deux molécules d(GGGGTTTTGGGG) [145]. Les couleurs réfèrent aux conformations *anti* (rouge) et *syn* (vert) des guanines. C, schéma du motif I d(TCC)₄, et D, structure RMN du tétramère d(TCC)₄ [146].

1.2.2.4 Quadruple hélice d'ADN : les motifs-i

Ce sont les propriétés extraordinaires d'échange des protons qui ont attiré l'attention sur la structure du motif-i, résolue sur le modèle de la séquence TC₅ en 1993 [147]. C'est une structure tétramérique formée de deux duplex orientés tête-bêche, entièrement intercalés, chacun des brins parallèles étant connecté par les paires de bases cytosine héli-protonées (Figure 9 C, D). L'intercalation systématique des bases distingue l'ADN I de toutes les structures d'ADN décrites.

Il a été démontré que le pH de demi-formation du motif-i est égal au pK de la cytosine +/- (2/log K_{dis}), K_{dis} étant la constante de dissociation de la base [146]. Pour cette raison, le motif-i peut être stable à des pH supérieurs au pK de la cytosine. La stabilité du motif-i dépend de la base extérieure qui peut protéger la structure de l'effilochage par les bords.

Les motifs-i diffèrent de l'ADN B par :

- l'appariement des bases,
- l'orientation parallèle des brins des deux duplex,
- l'inter-digitation des duplex pour former un quadruplex [148].
- La distance d'empilement des bases d'une valeur de 6,1 Å est supérieure à l'ADN B,
- la rotation de l'hélice de l'ADN I peut varier de 10 à 25° (~30° pour l'ADN B).
- La durée de vie d'une paire C-C⁺ peut varier de la minute à plusieurs heures tandis que les paires A-T et G-C présentent une durée de vie de l'ordre de millisecondes [149].
- La température de fusion du tétramère C₅ est de 67°C alors que celle du dimère C₅G₅ est de 55°C (à la concentration de 12 mM).

1.2.3 Formes artificielles associant acides nucléiques et composants peptidiques

Les PNA pour « peptide nucleic acids » et les PDNA pour « protein-deoxyribonucleic acid » sont des structures synthétiques hybrides covalentes. Une distinction notable est l'enchevêtrement des parties protéiques et nucléiques au sein du squelette de la structure pour les PNA tandis que ces parties sont distinctes dans les PDNA avec la partie protéique structurée et fonctionnelle liée de manière covalente à la partie ADN simple-brin.

1.2.3.1 PNA

Les PNA ont été développés par l'équipe de Nielsen en 1991 [150]. Ce sont des molécules de synthèse, analogues des acides nucléiques, dans lesquelles le squelette phosphodiester est remplacé par une chaîne pseudo-peptidique sur laquelle les bases puriques et pyrimidiques sont fixées par liaison carboxylique. Du fait de cette structure, les molécules de PNA ne sont pas chargées et s'apparient aux séquences d'acides nucléiques avec une forte affinité et spécificité. De par la nature atypique de leur squelette, les PNA présentent une remarquable stabilité et ne sont pas dégradés par les protéases et les nucléases.

L'appariement des PNA s'effectue par complémentarité de séquences selon le modèle classique de liaisons hydrogènes de Watson et Crick, mais contrairement aux acides nucléiques, les PNA peuvent s'apparier de manière parallèle ou antiparallèle et s'hybrider aux acides nucléiques simples brins ou doubles brins. Par ailleurs, les PNA riches en bases pyrimidiques ont la capacité d'envahir la double hélice et de s'apparier à leurs séquences complémentaires selon différents schémas pouvant aboutir à la formation de structures en triple hélice (triplex). Dans de tels complexes (Figure 10A), les bases pyrimidiques des PNA forment des liaisons de type Hoogsteen avec les bases puriques de l'ADN, en parallèle à l'appariement classique des bases. Ce phénomène d'invasion de la double hélice conduit au déplacement d'un des deux brins d'ADN et à la formation d'une boucle d'ADN simple brin.

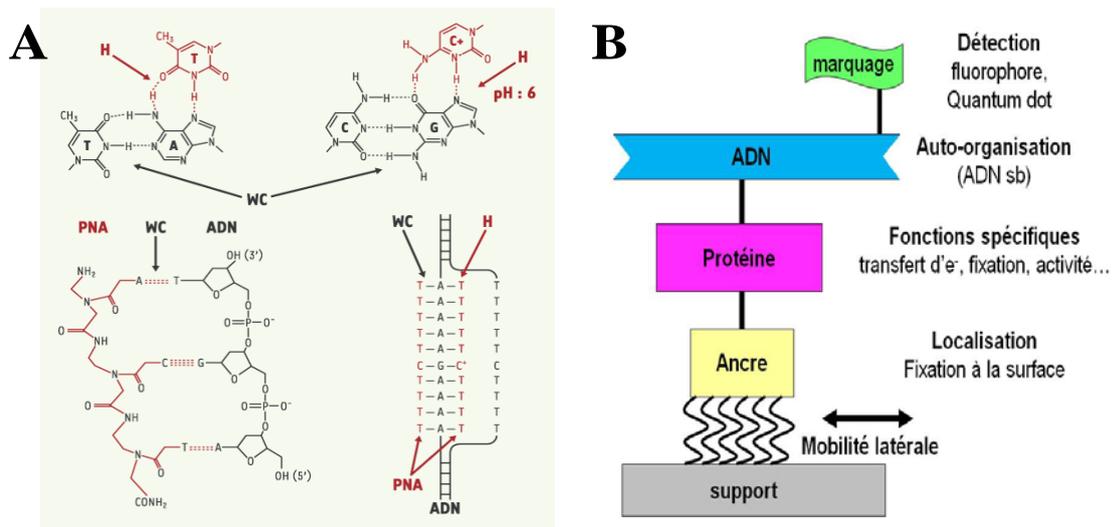


Figure 10 : Structures artificielles : PNA et PDNA.

A, structure du complexe PNA-ADN, L'hybridation du PNA aux acides nucléiques obéit aux règles classiques de l'appariement des bases : l'adénine est reconnue par la thymine et la guanine par la cytosine. Les liaisons hydrogène de type Watson-Crick sont impliquées dans la formation du duplex. Les bases pyrimidiques du PNA peuvent former des liaisons hydrogène de type Hoogsteen avec les bases puriques de l'ADN. Le PNA forme un complexe PNA-ADN-PNA avec le brin polypurique et induit le déplacement de l'autre brin de l'ADN. L'un des PNA du triplex forme des liaisons Watson-Crick (WC) avec l'ADN, l'autre des liaisons Hoogsteen (H) [151]. B, schéma fonctionnel du complexe protéo-nucléique : PDNA. Cette structure hybride constitue un bloc fonctionnel qui peut inclure les propriétés d'organisation des acides nucléiques, les propriétés fonctionnelles des protéines de même que des propriétés de détection en utilisant des marqueurs mais aussi des propriétés de reconnaissance du support. Celles-ci permettent d'ancrer les objets sur un substrat et d'accéder à la dynamique des objets dans le cas d'une mobilité latérale [152].

Depuis leur mise au point, les PNA ont été incorporés dans de nombreux protocoles d'études et de diagnostics génétiques pour la recherche de mutations, la coupure dirigée de l'ADN ou la thérapie génique.

1.2.3.2 PDNA

Les complexes protéo-nucléiques PDNA (protéine-DNA) sont des molécules de synthèse qui contiennent une protéine liée de manière covalente, par l'intermédiaire d'un adaptateur chimique, à un acide nucléique. De tels complexes ont été développés par l'équipe de D. Pompon en 2002, à partir de la protéine du cytochrome b5 humain qui a été modifiée par ingénierie et couplée par un adaptateur chimique bi-fonctionnel à un acide nucléique simple brin [153, 154]. La Figure 10B présente le schéma fonctionnel d'un PDNA avec les composantes ADN et protéine. Sur un même objet, sont présentes les propriétés d'organisation apportées par le domaine nucléique, avec un potentiel marquage qui permette le suivi et la caractérisation des structures ainsi que les propriétés fonctionnelles de la protéine qui peut autoriser l'interaction avec la surface. Selon la fonctionnalisation des surfaces, de type membrane supportée ou champ d'ions, les complexes PDNA présentent une mobilité latérale qui donne accès à l'information dynamique du système.

Le complexe PDNA étant un système modulaire, diverses protéines présentant différentes fonctions catalytiques, de reconnaissance et/ou de transport peuvent être couplées de manière covalente à un acide nucléique, pas forcément simple brin, lequel peut être engagé dans n'importe quelle structure selon les partenaires. De ce fait, les PDNA sont différents des PNA puisqu'ils comprennent des domaines protéiques structurés fonctionnels.

1.2.4 Utilisation des différentes formes d'ADN pour la fabrication de nanostructures.

La diversité structurale des ADN naturels ayant été présentée, nous allons voir comment ils ont inspiré la synthèse d'ADN artificiels en tant que briques de construction en nanotechnologie, sachant que les points forts de l'ADN pour l'auto-assemblage résident dans ses propriétés de reconnaissance spécifique et prédictible ainsi que dans sa géométrie. En effet, pour créer des structures complexes par auto-assemblage, il est nécessaire d'appréhender les mécanismes d'interactions complexes entre les unités de bases qui doivent être spécifiques mais également de géométrie bien définie.

1.2.4.1 Stratégies d'auto-assemblage des acides nucléiques

Les principes qui gouvernent l'auto-assemblage ont été développés par les chimistes [155, 156]. En 1978, J. M. Lehn introduit le terme de chimie supramoléculaire qui réfère aux agrégats de molécules ordonnées maintenues par des interactions faibles telles que les liaisons métal-ligand, les liaisons hydrogène, les forces de van der Waals. Les notions de réversibilité et de dynamique sont caractéristiques de ces systèmes.

Dans les systèmes organiques, la génération et le processus d'auto-organisation supramoléculaire reposent sur des programmes qui prennent en compte l'information structurale contenue dans les molécules constituantes ainsi que les interactions intermoléculaires mises en jeu. L'auto-assemblage utilise en particulier des séquences de sites donneur ou accepteur de liaison hydrogène, ce qui permet d'induire la formation d'architectures supramoléculaires variées mono-, di- ou tridimensionnelles [157]. Les polymères dynamiques, capables de modifier leur composition par échange et remaniement des unités, peuvent subir l'adaptation et l'évolution dirigée sous l'effet de déclencheurs externes chimiques ou physiques [158].

Parallèlement, au début des années 1980, le chercheur cristallographe N. C. Seeman inventait le concept de nanotechnologie à base d'ADN qui s'appuie sur les propriétés de reconnaissance moléculaire pour créer des structures par auto-assemblage. Frustré par le hasard et les conditions hypothétiques de cristallisation de certaines molécules, il est inspiré par une gravure de M. C. Escher et réalise que des réseaux d'ADN en 3D pourraient être utilisés pour orienter les molécules cibles, et donc simplifier les études cristallographiques en éliminant les difficultés d'obtention de cristaux purs [159].

C'est ainsi qu'en 1991, le groupe de N.C. Seeman publie la synthèse du premier objet nanométrique en 3D, un cube fait d'ADN [160]. Toutefois la flexibilité de ces objets ne permet pas de former des réseaux 3D, ce qui amena les chercheurs à développer des motifs plus rigides : les molécules DX qui permettent de créer des réseaux en 2D [161]. Ces briques ont également été utilisées en calcul algorithmique en considérant un ensemble de molécules d'ADN comme des briques de Wang [162]. Ceci constitue une des approches d'auto-assemblage.

On distingue généralement 3 catégories de méthodes d'auto-assemblage :

- l'approche simple en une seule étape, aussi appelée approche parallèle où les briques sont mélangées ensemble pour former l'assemblage final par reconnaissance des séquences,
- l'auto-assemblage séquentiel où des briques s'associent hiérarchiquement, des groupes pouvant être combinés ultérieurement,
- l'auto-assemblage nucléé où un élément est utilisé comme point de nucléation ou comme gabarit, sur lequel viennent s'assembler les motifs complémentaires pour former des formes complexes.

En réalité, c'est souvent une combinaison de ces approches qui est utilisée dans la fabrication de nanostructures complexes. Notamment l'approche algorithmique, présentée plus loin, peut être réalisée soit de manière parallèle, séquentielle ou nucléée, si bien que la classification n'est pas aisée.

1.2.4.1.1 Auto-assemblage direct ou parallèle

La stratégie parallèle se caractérise par une seule étape, réalisée en général dans des conditions réversibles ou dans des conditions de contrôle cinétique particulières permettant d'éviter les pièges cinétiques. Dans cette approche, une collection de briques individuelles est conçue, chacune avec des extrémités cohésives spécifiques qui vont mener la brique à une position unique lors de l'assemblage. Les brins de toutes les briques sont mélangés et hybridés en une seule étape, les séquences des brins étant conçues pour s'associer de manière prédictible avec leur partenaire uniquement.

Plusieurs assemblages de réseaux 2D ont été réalisés en utilisant cette approche, telles que les formes de rayures à partir de briques DX et TX avec l'utilisation de deux à quatre briques différentes [163, 164]. L'avantage de cette méthode est qu'elle ne requiert pas d'encodage de forme au sein des brins d'ADN, car la forme 2D est générée *de novo* par les relations de voisinage encodées dans la collection des briques d'ADN. Ce système étant gouverné par la séquence des briques d'ADN qui vont s'hybrider, leur conception est cruciale tant pour les structures secondaires que pour les parties collantes qui doivent permettre d'orienter les superstructures. Dans ces conditions, la stœchiométrie constitue un point crucial, car, si elle n'est pas respectée, elle génère des structures incomplètes ou incorrectes.

Lors de la réalisation de formes complexes, cette approche nécessite un contrôle délicat des phénomènes physiques tels que la vitesse de croissance et des erreurs et ce d'autant plus que le nombre de briques de différentes sortes est élevé.

La technique d'auto-assemblage appelée Origami s'est inspirée du concept d'origami issu de l'art traditionnel japonais, qui consiste à plier du papier pour lui donner des formes diverses, de poupées, d'animaux... P.W. Rothemund a appliqué ce concept sur l'ADN qui se prête bien aux pliages en mettant en œuvre ses propriétés de reconnaissance et d'assemblage pour créer des structures [165, 166]. Le principe consiste à plier l'ADN génomique simple brin du virus M13mp18, d'environ 7000 bases, à l'aide de multiples agrafes qui sont de courts oligonucléotides. L'ajout en excès des agrafes dans un tampon riche en magnésium et l'utilisation d'une température de 90°C suivie d'un lent refroidissement, aboutissent à la réalisation de diverses structures de formes définies (Figure 11).

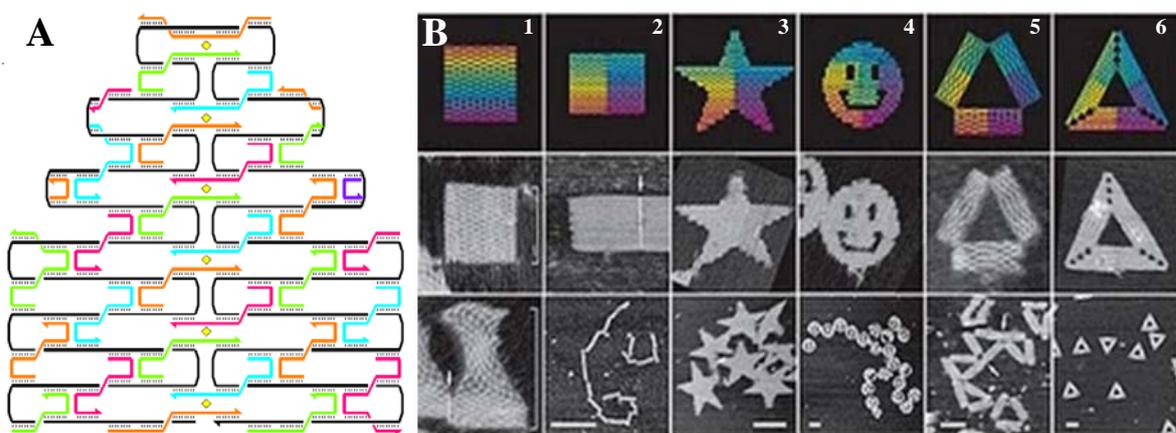


Figure 11 : Formes d'ADN origami.

A, schéma de principe du repliement de l'ADN circulaire simple brin (noir) à l'aide de nombreuses courtes agrafes nucléotidiques (en couleur). B, ligne du haut, diagramme de couleurs montrant le repliement des hélices. La couleur indique l'indice de la paire de base le long du chemin de pliage, la couleur rouge indiquant la 1ère base et la couleur violette la 7000ème. Lignes du bas, images AFM. Toutes les images sans barres d'échelle sont de même taille (165 nm x 165 nm). Les barres d'échelle de la dernière ligne indiquent 1 µm (2) et 100 nm (de 3 à 6) [165].

1.2.4.1.2 Auto-assemblage séquentiel

La stratégie séquentielle est une approche multi-étapes utilisant les propriétés spécifiques de reconnaissance des briques moléculaires pour créer des structures par un processus ordonné. En auto-assemblage hiérarchique, dans un premier temps, un petit nombre d'unités ou de blocs de construction sont assemblés. Ces blocs de construction moléculaires sont ensuite mélangés séquentiellement pour aboutir à la construction de structures plus grandes et complexes.

Un bloc d'ancrage attaché à un support peut être utilisé pour laver l'excès de réactif à chaque étape. Dans l'approche multi-étapes, le nombre de briques ou blocs moléculaires

distincts peut être réduit en réutilisant plusieurs fois les briques alternativement. Des étapes de purification peuvent être nécessaires pour augmenter le rendement de structures correctes. Cette approche pas à pas présente l'inconvénient d'être coûteuse en temps.

L'approche séquentielle a été utilisée par de nombreuses équipes pour réaliser des structures nanométriques complexes. Par exemple, Yan et ses collaborateurs ont fabriqué des briques avec des extrémités cohésives asymétriques qui, mélangées, s'auto-assemblent pour donner des structures symétriques de taille finie (Figure 12) [167]. La première étape consiste à assembler les briques symétriques Figure 12A. Les briques symétriques formées seront ensuite assemblées pour réaliser le dessin en réseau de taille finie 5x5 présenté sur la Figure 12B.

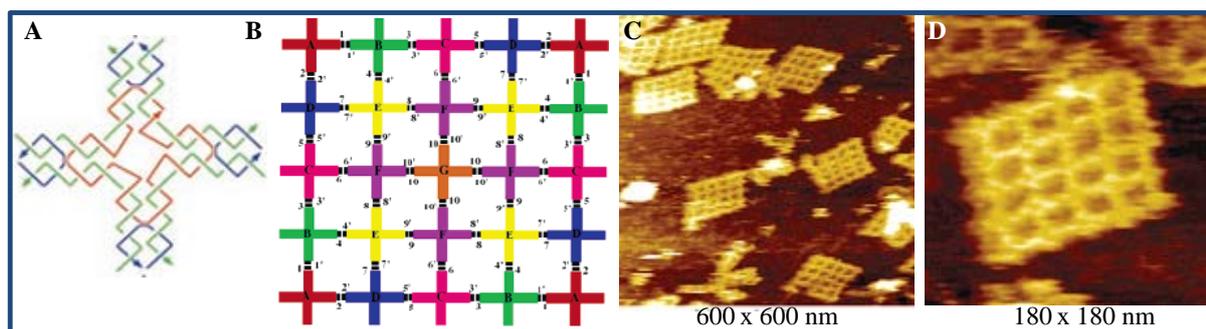


Figure 12 : Un exemple d'auto-assemblage séquentiel.

A, brique d'ADN symétrique à bout francs. B, Conception d'un réseau de taille 5x5 à partir de la brique d'ADN présentée en A. Pour former le réseau à 25 briques, un total de 7 briques différentes est nécessaire pour les 10 paires d'extrémités cohésives. C et D, images AFM [167].

1.2.4.1.3 Auto-assemblage dirigé par nucléation

L'approche d'auto-assemblage dirigé par nucléation utilise un brin d'ADN pré-assemblé qui sert de gabarit. Le brin d'ADN contient l'information de forme de la structure, par exemple une boucle ADN. Les oligonucléotides spécifiques s'assemblent autour de ce brin matrice pour former les formes désirées en 1D ou 2D. La technique d'assemblage dirigé par nucléation présentée en 1999 par T.H. LaBean. [168], est utilisée pour la première fois avec un brin gabarit qui encode une information de forme en 2003 [169]. Le brin d'ADN gabarit, construit par ligation avec de courts oligonucléotides synthétiques, contient l'information de forme en code-barres représentée par la présence ou l'absence de boucles en épingle d'ADN qui dépassent du réseau d'ADN (Figure 13).

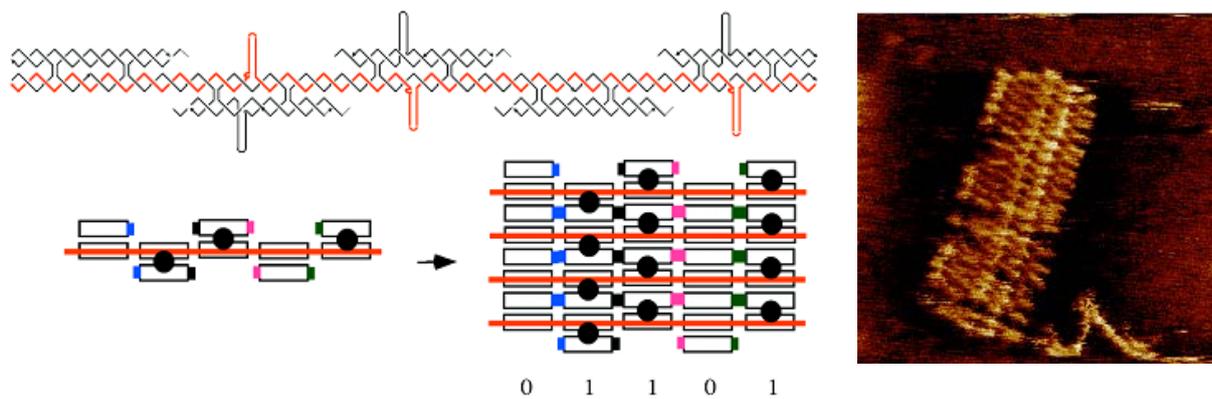


Figure 13 : Exemple d'auto-assemblage nucléé.

Réseaux de code barre. Structure du brin d'une couche à code barre de type 01101 imagée par AFM (domaine de 400 nm). Le brin rouge constitue le gabarit nécessaire à l'auto-assemblage des briques[168]

Une forme particulière d'auto-assemblage dirigé par nucléation pourrait également être celle des ADN origami car l'ADN simple brin circulaire est utilisé en tant que gabarit pour former des structures de configurations diverses. Toutefois selon la définition du terme nucléation, qui dans son sens premier indique un point d'ancrage, et non le repliement d'un ADN simple brin circulaire, la méthode originale d'ADN origami a déjà été classée dans les techniques d'auto-assemblage parallèle puisqu'elle s'effectue en une seule étape. On retrouve ici le recoupement des différentes méthodes.

1.2.4.1.4 Auto-assemblage algorithmique

L'auto-assemblage algorithmique consiste à programmer des briques avec des domaines d'association spécifiques qui vont s'assembler de façon coopérative pour former des structures complexes selon la règle algorithmique.

En 1998, E. Winfree crée une abstraction des systèmes d'auto-assemblage nommée *modèle d'assemblage de briques* : TAM (pour Tile Assembly Model) [164]. Le modèle TAM, inspiré des briques de Wang, est basé sur les molécules d'ADN définies de telle sorte qu'elles se comportent comme des briques carrées avec de la « colle » sur les côtés qui leur permettent de se lier selon les colles correspondantes. De nombreuses formes variables asymétriques ont été créées in vitro par assemblage de briques, telles que les triangles de Sierpinski (Figure 14) et les formes de comptage binaire [161, 170].

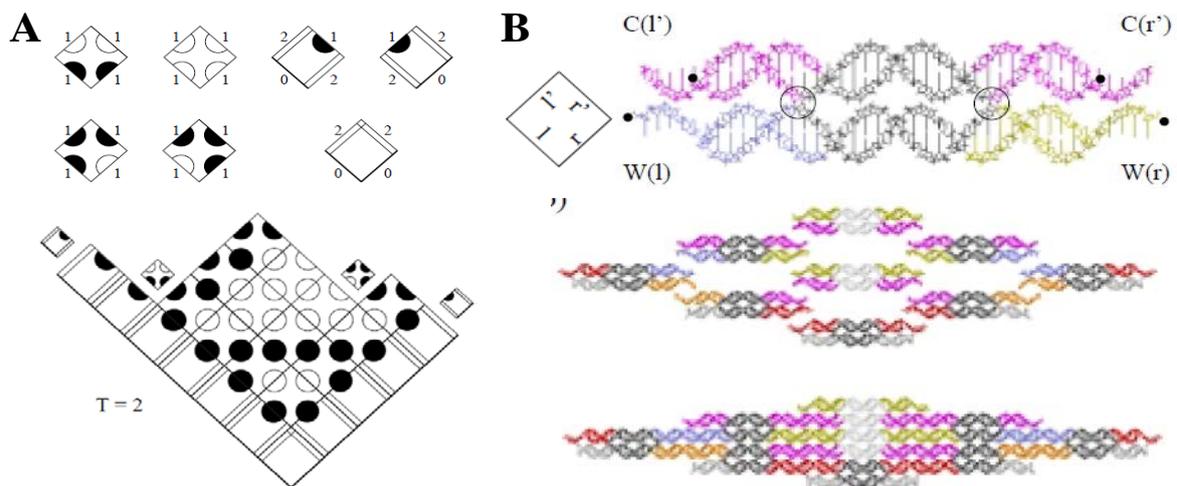


Figure 14 : Auto-assemblage algorithmique.

A, Les 7 briques de base de Sierpinski avec la force des côtés annotée et représenté graphiquement. Les petites briques indiquent les additions possibles à l'agrégat. A la « température » $T=2$, une possibilité unique existe pour chaque site aboutissant à la formation d'une structure unique. B, Représentation ADN des briques de Wang. A gauche une brique moléculaire de Wang (un double croisement) représentant les règles $(l, r) \rightarrow (l', r')$. La molécule consiste en une région structurale interne et 4 bras double brin terminés par des extrémités simple brin collantes. C et W indiquent Crick et Watson, l et r indiquent gauche et droite [164].

La première version du modèle d'assemblage de briques TAM : aTAM pour « abstract Tile Assembly Model » fait abstraction des caractéristiques physiques du modèle. Les assemblages de structures démarrent à partir de structures graine « seed » prédéfinies et croissent en superstructures finies ou infinies qui représentent des formes préétablies ou des sorties de calcul [171]. Dans ce modèle, l'assemblage croît sans erreur aussi longtemps qu'une brique peut s'ajouter au cristal.

Une deuxième version fut proposée pour rendre compte de la réversibilité des réactions biochimiques des molécules d'ADN : le *modèle d'assemblage des briques cinétique* kTAM pour « kineticTile Assembly Model ». Ce modèle prend en considération les concentrations moléculaires, les vitesses de réaction (association et dissociation) des briques et d'autres variables physiques. C'est le premier modèle à prendre en compte le taux d'erreur d'un assemblage. Les simulations indiquent qu'une diminution optimale de la température et des concentrations de brique atténue le taux d'erreur mais ceci au prix d'un ralentissement important de la vitesse de croissance, c'est pourquoi les scientifiques ont cherché à développer des techniques de diminution d'erreur qui permettent de conserver une bonne vitesse de croissance.

Dès lors, de nombreux modèles ont été proposés pour diminuer les erreurs d'assemblage qui sont de trois sortes : les erreurs de « croissance », de « facette » et de « nucléation ». Les deux premières erreurs se produisent sur le front de croissance d'un assemblage existant tandis que les erreurs de nucléation concernent l'initiation parasite d'assemblages [172]. Ces erreurs se produisent du fait que les briques forment des liaisons

faibles avec l'assemblage, elles doivent se dissocier pour qu'un processus d'assemblage correct puisse être réalisé, mais elles peuvent être stabilisées par l'arrivée d'une autre brique qui va consolider son lien à l'assemblage.

Ainsi, E. Winfree et R. Bekbolatov [173] et J.H. Reif [174] proposent des modèles de brique de correction « proof reading tile model » et des modèles de briques tolérants aux erreurs « error resilient tile model » qui traitent les problèmes de diminution d'erreurs de croissance au cours de l'assemblage tandis que Chen propose le modèle de correction serpentin « snaked proof reading tile model » qui aborde aussi les erreurs de nucléation [175]. Dans ce domaine, Fujibayashi *et al.* font intervenir de nouveaux motifs structuraux appelés brins et briques de protection qui protègent les extrémités cohésives exposées des briques pour minimiser les interactions parasites. Les briques sur le front de croissance sont déprotégées, permettant une croissance correcte à une vitesse supérieure à celle de la croissance parasite [172].

L'auto-assemblage algorithmique peut constituer un mode d'auto-assemblage qui utilise les approches directes, séquentielles ou nucléées selon la conception des structures.

1.2.4.2 Principaux blocs moléculaires utilisés en structuration d'ADN

Plusieurs revues décrivent les motifs d'ADN les plus utilisés en nanotechnologie [176-182].

1.2.4.2.1 ADN linéaires

Dans les stratégies architecturales, les molécules d'ADN B double brins linéaires sont essentiellement destinées à la réalisation de nano-fils conducteurs, même si la conductivité de l'ADN nu est un sujet controversé [183-186]. Quoique l'ADN nu présente un comportement de transfert d'électrons insuffisant pour l'ingénierie nanoélectronique, il constitue un support idéal pour organiser des matériaux hautement conducteurs, tels que les métaux. Par exemple, Lee [187] a décrit une nouvelle forme d'ADN, appelée ADN-M, où le proton imino des paires de base ADN est remplacé par un ion métallique Zn^{2+} , Ni^{2+} ou Co^{2+} illustré en Figure 15A, B.

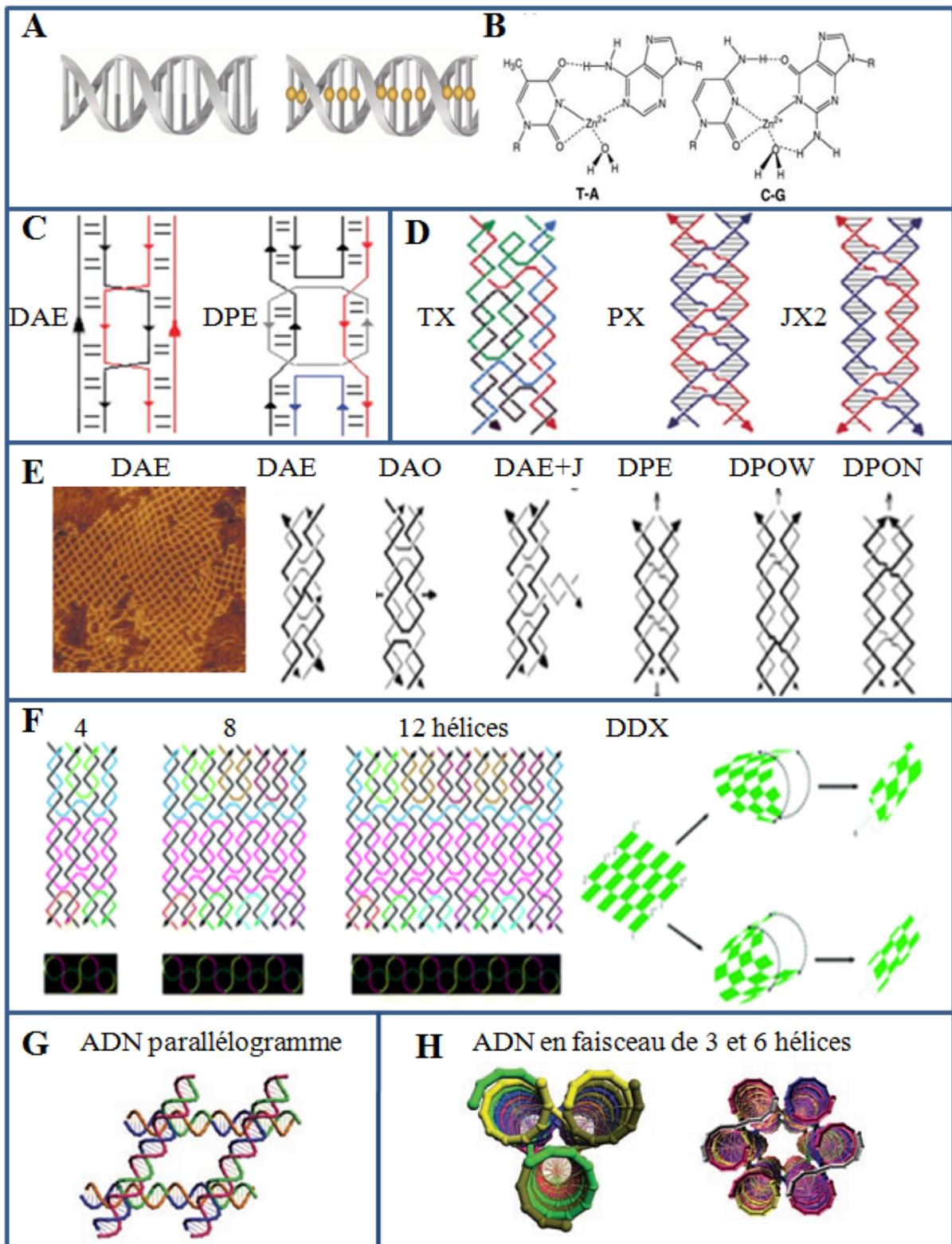


Figure 15 : Motifs d'ADN utilisés dans les auto-assemblages.

A, ADN linéaire, vue schématique de la double hélice d'ADN et de l'ADN-M, où les ions métalliques sont fixés et empilés entre les paires de bases de l'ADN [188]. B, structure proposée de paires de bases d'ADN-M [189]. C, structures branchées, jonctions à double croisements (DX) contenant un nombre pair de tours d'hélices entre les deux points de jonction. À gauche, DPE construit à partir de quatre brins et à droite, DAE construit à partir de cinq brins [190]. D, représentation schématiques des briques triples croisements (TX) et multiple croisements (PX et JX2) [191]. E, représentation schématique des différentes topologies qui peuvent être formées par des jonctions double croisement (DX) [192, 193]. F, motifs 4, 8, 12 hélices [194]. G, motif ADN parallélogramme formé par l'union de quatre jonctions Holliday en parallèle (195). H, motif d'ADN en faisceau à 3 hélices, à gauche [195] et faisceau d'ADN à 6 hélices, à droite [196].

L'ADN double brin peut également être utilisé en conjonction avec d'autres motifs ADN plus complexes et servir de contrôle de longueur et d'orientation au cours de l'auto-assemblage pas à pas de superstructures [197].

1.2.4.2.2 ADN branchés

Les molécules d'ADN branchées issues de l'ADN B, telles que les jonctions Holliday, sont des motifs intéressants pour organiser et créer des structures de manière prévisible en mettant en œuvre les propriétés spécifiques des bases d'ADN. Elles présentent toutefois l'inconvénient d'être trop flexibles au niveau des jonctions, c'est pourquoi les molécules à plusieurs croisements ont été créées afin de rigidifier les structures [161].

1.2.4.2.2.1 Motifs DX.

L'utilisation de l'ADN double croisement (DX) est un élément clé dans l'approche rationnelle développée par N.C. Seeman [190] pour la construction de nouveaux biomatériaux. Ce sont des objets géométriques individuels qui peuvent être inclus dans des constructions plus larges de structure périodique [190, 198, 199]. Les molécules DX consistent en deux hélices doubles brins côte à côte liées par deux points de croisements (Figure 15C). Il en résulte des différences topologiques selon la polarité des brins qui ont subi ou non un échange réciproque.

Sur la Figure 15E sont illustrés les différents motifs DX développés par Fu et Seeman [192] et leur nomenclature qui se décline à partir de trois lettres selon les caractéristiques géométriques. La première lettre D désigne la structure double-croisement, la deuxième lettre (P) parallèle ou (A) antiparallèle définit l'orientation de la jonction. La troisième lettre précise le nombre de demi-tours d'hélices entre chaque point de croisement qui est pair (E, pour even) ou impair (O, pour odd). A ceci s'ajoutent les lettres N (narrow) pour étroit et W (wide) pour large, qui dénotent si le demi-tour d'hélice au point de croisement correspond au grand ou au petit sillon.

Dans les isomères de motifs DX, on distingue les domaines hélicoïdaux antiparallèles (DAO, DAO et DAE+J) où les croisements se produisent entre des brins de polarité opposée et les domaines parallèles (DPE, DPOW et DPON) où les croisements se produisent entre des brins de même polarité. Ainsi, il y a environ une différence d'un demi-tour dans le phasage entre les domaines hélicoïdaux quand on compare les espèces antiparallèles et parallèles. Dans les premières espèces, l'axe du squelette est perpendiculaire aux axes de l'hélice tandis que dans les dernières espèces, l'axe du squelette est parallèle aux axes de l'hélice. Sur gels polyacrylamide natifs, le comportement des molécules parallèles n'est pas stable ; soit elles se dissocient, soit elles forment des multimères tandis que les molécules antiparallèles présentent une bande unique [192]. C'est la raison pour laquelle, un plus grand intérêt s'est porté sur les

molécules DX antiparallèles DAE et DAO, et ce particulièrement dans les systèmes nanotechnologiques.

Par ailleurs, la molécule DAE est la seule qui contienne un brin central cyclique. En effet, dans les jonctions DAE, trois des cinq brins sont impliqués dans les croisements et les deux autres sont fixés à une double hélice. Dans les trois brins, deux sont complémentaires aux terminaisons des brins fixes tandis que le troisième circulaire, est complémentaire de la partie centrale de la jonction (Figure 15E). Ce dernier brin peut également projeter ses extrémités hors du cadre de la jonction et permettre de positionner d'autres briques moléculaires d'intérêt. Ceci constitue la topologie appelée DAE+J.

1.2.4.2.2.2 Motifs TX, PX et JX2.

Les motifs TX [191] sont issus des structures DX avec trois hélices distinctes au lieu de deux, ce qui double les points de croisement. Ces motifs peuvent également être générés en utilisant quatre brins, avec chaque brin impliqué dans deux événements de croisement comme présentée dans la Figure 15D. Ces structures présentent une plus grande rigidité que les jonctions branchées simple ou double croisement.

Les motifs PX résultent d'un échange réciproque entre deux brins des deux hélices aux mêmes points de croisements. Ils présentent de multiples points de croisement à chaque tour d'hélice, où les brins de même polarité provoquent des croisements parallèles.

Les motifs JX2 sont similaires aux motifs PX hormis que l'échange réciproque s'établit tous les deux points adjacents [178].

Ainsi l'ensemble de ces briques peuvent être dotées d'extrémités cohésives à des endroits clé afin de permettre un assemblage plus complexe. Tous ces motifs sont utilisés pour construire des assemblages linéaires, des réseaux en 2D et des tubes d'ADN [191, 200].

1.2.4.2.2.3 Motifs DDX

Les motifs double-double-croisement (DDX) consistent en quatre hélices d'ADN connectées par six échanges réciproques [201]. Ke *et al.* étudient des motifs à 4, 8 et 12 hélices (Figure 15F) et montrent qu'ils peuvent s'auto-assembler en réseaux 2D et/ou en tubes avec des cavités rectangulaires de taille de variable. La conformation des structures en tube ou en réseau dépend des paramètres d'anisotropie des motifs, de contraintes de connexion et de niveau d'ondulation des motifs [194].

1.2.4.2.2.4 Motifs d'ADN parallélogramme

Les motifs d'ADN parallélogramme développés en 1999 par l'équipe de N. C. Seeman sont des structures analogues aux jonctions Holliday mais en forme de losange [202]. Ce motif provient de la fusion de quatre jonctions et consiste en quatre hélices de six tours, deux au-dessus et deux au-dessous avec les points de branchements qui définissent les sommets

séparés par quatre tours d'hélice chacun Figure 15G. Ces motifs s'auto-assemblent en réseaux 2D avec des cavités en forme de losange. Ils permettent de mesurer les angles de torsion entre les bras des jonctions branchées.

1.2.4.2.2.5 Motifs en faisceau de 3 ou 6 hélices

Les motifs en faisceau d'hélices sont des objets 3D en eux-mêmes qui peuvent s'auto-assembler en des structures 1D ou 2D et potentiellement en réseaux 3D (Figure 15H).

Le faisceau à 3 hélices consiste en 3 domaines doubles hélices d'ADN connectés par 6 jonctions de croisement immobiles de telle sorte que les axes des hélices ne sont pas coplanaires. Le motif 3HB (three-helix bundle) présente une section transversale triangulaire avec une hélice étendue dans le sillon formé par les deux autres [195]. L'assemblage de six hélices d'ADN produit un faisceau hexagonal qui contient un trou en son centre. Ce motif, appelé 6HB (6 helix-bundle) est constitué à partir de motifs DX séparés par des paires d'oligonucléotides avec ou sans extrémités cohésives.

L'ensemble des motifs présentés montre la versatilité et la programmabilité des auto-assemblages à base d'ADN. Il est à noter que la règle générale dans la conception de ces motifs est la minimisation de la symétrie des séquences dans les structures branchées afin d'éviter tous les appariements indésirables entre les brins participants et la mobilité des jonctions.

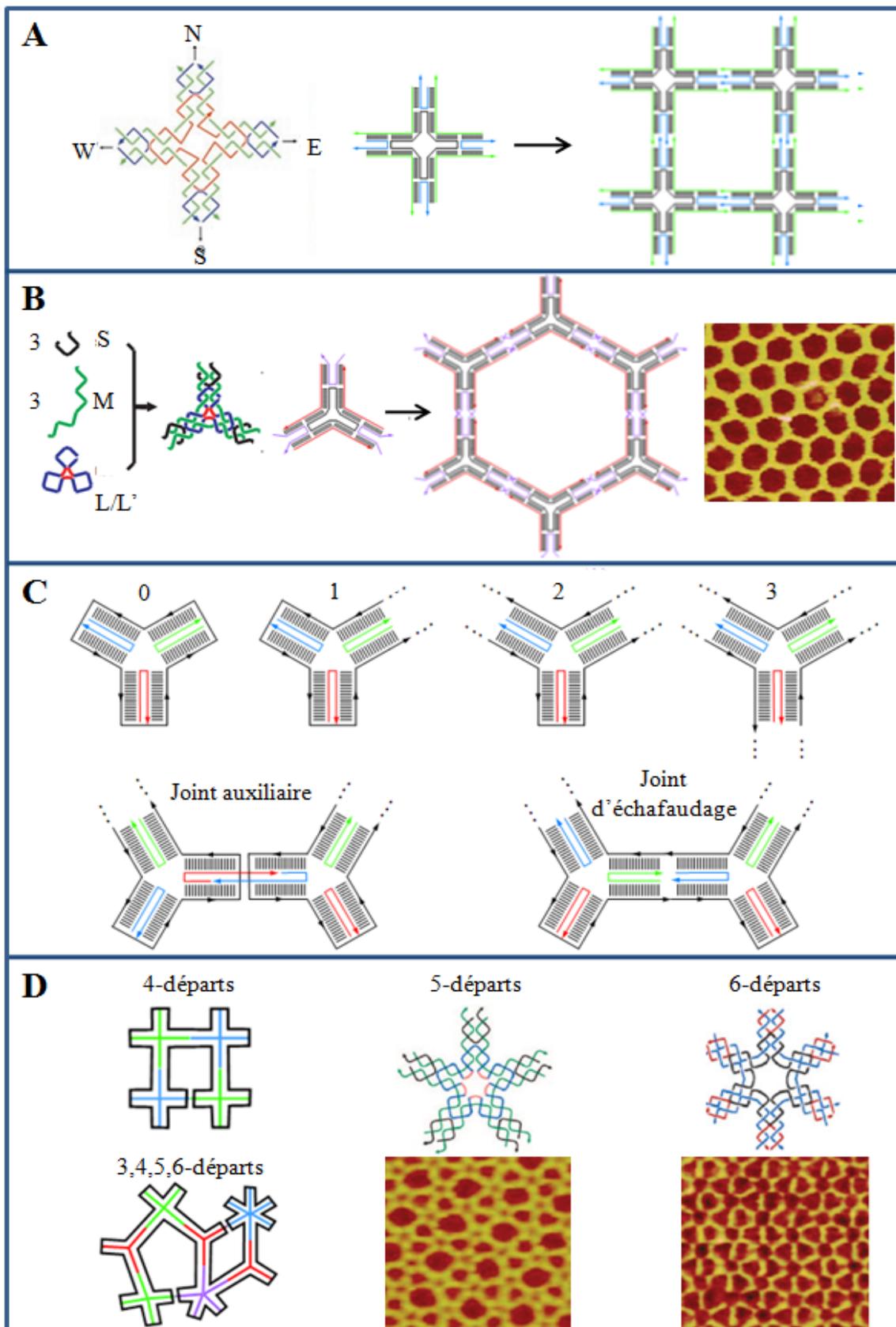


Figure 16 : Autres motifs 4x4 et symétriques.

A, Structure des brins du motif 4x4 [203] et représentation schématique de leur assemblage [166]. B, trois types d'ADN simple brin différents s'assemblent pas à pas en motif three-point-star [166, 204]. C, haut, classification des ADN 3-départs selon le nombre « d'extrémités ouvertes » du brin matrice ; C bas, deux structures générées lors de la jonction par les ADN agrafes (joint auxiliaire « helper join », à gauche) et par l'ADN matrice (joint du gabarit « scaffold join », à droite [166]). D, Différents motifs à k-point de départs permettent de créer des origamis polygonaux. Présentation de quelques images AFM issues de ces motifs [193].

1.2.4.2.2.1 Motifs 4x4

Le groupe de T.H. LaBean [203] a développé le motif 4x4 qui présente un rapport d'aspect carré et qui peut s'auto-assembler en des nano-rubans ou en des réseaux nano-grilles 2D. Ce motif contient quatre jonctions d'ADN branchés de quatre bras pointant dans les quatre directions (N, S, E, W) pour nord, sud, est, ouest dans le plan du motif (voir Figure 16A). Ainsi, dans les motifs 4x4, ce sont deux hélices d'ADN et non une seule (comme dans les motifs DX) qui constituent les bras du motif. Le motif 4x4 est composé de 9 oligonucléotides, et l'un des brins (en rouge sur la figure) participe à toutes les jonctions avec des boucles T4 qui connectent les jonctions adjacentes et qui, en diminuant les interactions d'empilement entre les jonctions adjacentes, favorisent l'orientation perpendiculaire des 4 bras. Prises individuellement, ces jonctions branchées sont plutôt flexibles, mais la contrainte par trois jonctions additionnelles au sein du motif et par une jonction de voisinage en fait des structures suffisamment rigides qui peuvent être utilisées en tant que briques moléculaires dans de plus larges superstructures.

1.2.4.2.2.2 Motifs symétriques : les three-point-stars

Le groupe de C. Mao a introduit le concept de « symétrie de séquence » dans le but de diminuer le nombre d'ADN simples brins nécessaires à la construction d'une superstructure, diminuant ainsi le coût de fabrication et le taux d'erreurs expérimentales [205]. La symétrie géométrique parfaite des motifs annule toutes les distorsions imprévisibles dans les nanostructures et permet leur croissance à des tailles plus grandes allant jusqu'à quelques millimètres [206]. Le motif trois-points-de-départ contient une symétrie rotationnelle de trois-repléments (three-fold rotational symmetry). Sept brins participent à sa formation (Figure 16B): un long brin central répétitif (bleu-rouge, brin L ou L'=long), trois brins moyens (en vert, brin M=medium) et trois brins périphériques courts identiques (noir, brin S=short). Au centre du motif, 3 boucles simple brins sont colorées en rouge. La flexibilité de ce motif peut être ajustée à volonté par la variation de la longueur de la boucle, plus elles sont longues plus la structure est flexible. Les extrémités de chaque branche du motif portent des extrémités cohésives qui permettent d'assembler ces motifs en des superstructures telles que des polyèdres.

1.2.4.2.2.3 Motifs multi-bras

Les motifs multi-bras proposés par P.W. Rothmund [166] sont similaires aux motifs 4x4 et 3-point-star présentés car ils utilisent également deux doubles hélices par bras mais ils en diffèrent par les points de croisements qui sont à la jonction entre les bras et non au milieu de chaque bras (Figure 16C et D). Ces motifs peuvent être utilisés dans le contexte du repliement des ADN origami pour créer des réseaux arbitraire de forme polygonale.

1.2.4.2.2.4 Structures « agrafes »

Les structures de type agrafes sont des oligonucléotides simple brin utilisées dans la technique d'auto-assemblage des ADN origami. Différentes agrafes sont utilisées [165, 170], courtes ou plus longues, de structures croisées ou rectilignes selon les hélices qu'elles couplent et les formes à construire. Les motifs utilisés sont ceux précédemment présentés (DX, PX, JX). Le plus souvent, les agrafes sont conçus pour se fixer à trois hélices adjacentes soit avec une géométrie de forme S ou Z. La longueur utilisée est typiquement de 32 nucléotides pour un espacement de 1,5 tour entre deux croisements ou de 52 nucléotides pour un espacement de 2,5 tours. Les 16 nucléotides centraux se fixent à une hélice tandis que les 8 extrêmes se fixent aux hélices adjacentes. Certaines agrafes sont modifiées soit pour éviter l'agrégation due aux paires de bases hydrophobes qui se situent sur les tranches des ADN origami (en ajoutant des boucles T4), soit pour donner une forme à la surface avec des hauteurs locales différentes. Les motifs multi-bras présentés (Figure 16D) constituent des structures agrafes particulières.

1.2.4.2.3 Hélices multiples

1.2.4.2.3.1 Quadruplex de G

Différentes équipes ont observé l'auto-assemblage spontané de courts oligonucléotides riches en guanine en de long fils moléculaires en présence d'un cation approprié [144, 207, 208]. De longs polymères de quartet G peuvent ainsi former des structures en forme de fils appelés fils-G, avec une largeur de 2,3 nm et des longueurs de plusieurs centaines de nanomètres (Figure 17a et b). Les tétrades de G4 ont attiré l'attention en nanotechnologie en tant que briques de construction de nano-fils pour la nanoélectronique. En effet, les fils G4 présentent une plus grande longueur de persistance et une stabilité dans des conditions variées (température, DNase). De plus, la guanine étant la base avec le potentiel d'ionisation le plus bas, une proportion élevée de guanine augmente la probabilité de transfert de charge au travers des fils G [209]. Des nanostructures continues composées de fils G mono-moléculaires contenant des centaines de tétrades ont été fabriquées par Borokov *et al.* [210]. La polarisabilité de ces structures est supérieure à celle des ADN double brins, mais elle a encore été augmentée par la fabrication de fils G tétra-moléculaires, à partir de 4 brins parallèles en utilisant la reconnaissance avidine-biotine [211, 212]. Ainsi la résistance élevée des quartets G tétra-moléculaires à des déformations mécaniques, alliée à leur polarisabilité électrique font de ces structures des outils prometteurs en nanoélectronique.

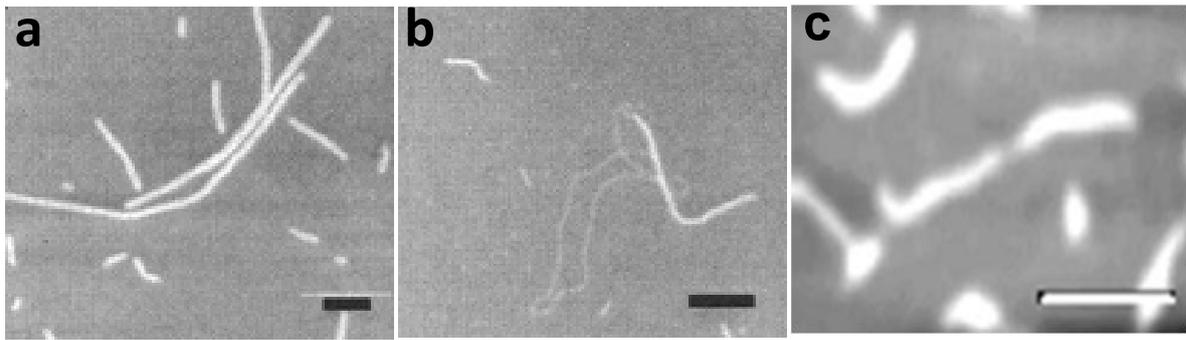


Figure 17 : Images AFM de nanostructures à base de quadruplex de G et de motifs-i.

Images AFM de la croissance de fils-G en présence de spermidine. La hauteur moyenne des fils est de 2,4 nm, leur longueur moyenne de 10 à plus de 1000 nm. Les échelles sont de 50 (a) et 100 nm (b) avec la hauteur apparente d'un ADN plasmidique de 0,7 nm en (b) [208]. C, Image AFM de fils-i formés à partir de d(C₇), barre : 0,5 μm [213]

1.2.4.2.3.2 Motifs-i

Pour la première fois en 2007, les motifs-i ont été utilisés en tant que briques de construction en nanotechnologie d'ADN structural par l'équipe de Y. Krishnan [213]. En s'appuyant sur les propriétés d'auto-assemblage liées à la symétrie du motif-i, les chercheurs ont fabriqué des fils de 2 nm de largeur et de longueurs atteignant jusqu'à 3 μm (Figure 17c).

1.2.4.2.4 Autres structures

1.2.4.2.4.1 Structures à base d'acides ribonucléiques

Au-delà de la topologie des boucles et des hélices composées de paires de base standard, l'ARN peut adopter une structure tertiaire comme une protéine. Celle-ci dépend des repliements des bases dans l'espace et de la formation des liaisons hydrogènes. Selon les appariements de type Watson et Crick (A&U, G&C), de type bancal « Wobble » (G&U), de type Hoogsten (A triple liaison C), ou les interactions entre le ribose et le phosphate voisin, on obtient des structures en double hélice, pseudo-hélice, triple hélice ou boucle variable. La conformation tridimensionnelle est maintenue par les liaisons hydrogènes coplanaires entre les bases, des liaisons hydrophobes, ou par des ensembles de boucles. C'est le cas typique du ribosome, dans lequel les ARN ribosomiaux repliés sur eux-mêmes, forment une structure tridimensionnelle compacte. La structure 3D d'un ribosome bactérien a donné lieu au Prix Nobel en 2009 [214-216] tandis que celle d'un ribosome eucaryote a été résolue avec une précision de 0.415 nm en 2010 (Figure 18A) [217].

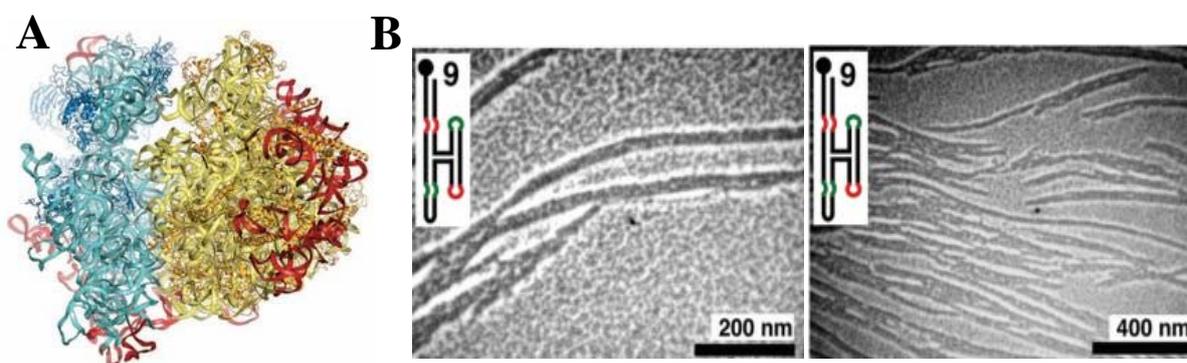


Figure 18 : Structures à base d'ARN.

A, Structure du ribosome eucaryote obtenue par diffraction aux rayons X (DRX), à une faible résolution. La sous-unité 40 S apparaît en bleu et la 60 S en jaune (les protéines sont plus foncées que les ARN) [217]. Images de microscopie électronique en transmission (TEM) d'assemblage supramoléculaires d'ARN [218].

Un autre exemple est celui des ribozymes qui sont des molécules d'ARN anti-sens ayant une activité catalytique, classés selon la nature de leur séquence et leur structure tridimensionnelle [219].

Les ARN sont des biomolécules modulaires largement composées de sous-unités structurales ou motifs. Ces motifs structuraux, via des interactions tertiaires non covalentes, génèrent des structures tridimensionnelles fonctionnelles [220, 221]. Les ARN stables peuvent être définis comme des architectures 3D modulaires et hiérarchiques [222, 223]. La construction d'architectures supramoléculaires à base d'ARN artificiels, prenant avantage sur la connaissance des règles de repliement et d'auto-assemblage de ces molécules, a été développée dans plusieurs laboratoires et s'appuie sur la création de programmes de conception (RNA tectonics) et de bases de données [224-227]. L. Nasalean *et al.* ont présenté un exemple de contrôle d'auto-assemblage d'ARN qui forme des filaments comparables à ceux de l'actine illustré sur la Figure 18B [218].

1.2.4.2.4.2 Structuration autour d'un métal

Différents travaux rapportent le développement d'auto-assemblage de nanoparticules d'or par l'ADN [228]. La modification chimique des oligonucléotides alliée aux propriétés de reconnaissance et de stabilité de l'ADN ont permis d'assembler réversiblement des nanoparticules d'or en agrégats macroscopiques 3D [229]. On peut également citer l'organisation de nano-cristaux d'or en des structures spatialement définies. Ainsi des dimères (parallèle et antiparallèle) et des trimères de nano-cristaux ont été fabriqués, avec un positionnement précis des particules contrôlé par le choix de la séquence ADN [230]. En combinant des particules d'or de deux diamètres différents (9 et 31 nm), Mucic *et al.* ont généré un assemblage binaire de nanoparticules illustré sur la Figure 19 [231].

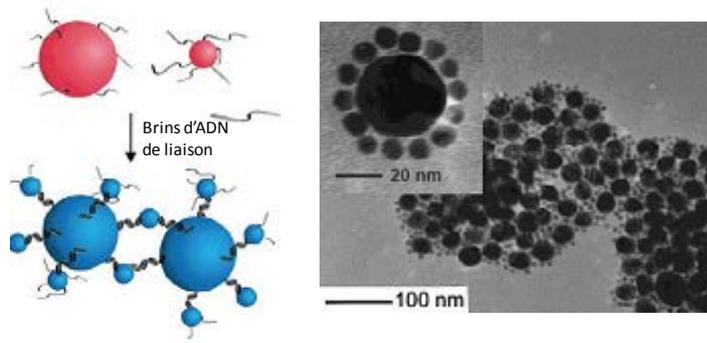


Figure 19 : Assemblages d'ADN liés aux nanoparticules d'or.

Image (TEM) montrant le réseau binaire des nanoparticules fonctionnalisées avec en insert une structure satellite [231].

Il existe également des associations avec les métaux pour les acides nucléiques et les protéines que l'on retrouve dans la famille des motifs de fixation à l'ADN.

1.2.4.2.4.3 Structures des complexes protéines-ADN

Les motifs de fixation à l'ADN, présents dans les facteurs de transcription de nombreux procaryotes et eucaryotes, sont utilisés dans la fabrication artificielle d'assemblages à l'échelle nanométrique. En effet, les règles reproductibles qui lient les séquences de ces motifs à leur structure 3D en font de potentielles briques moléculaires. R.Y. Sweeney *et al.* ont par exemple utilisé les interactions de la glissière à leucine pour diriger l'auto-assemblage spécifique de particules de bactériophage [232].

Quelques motifs de fixation à l'ADN sont présentés sur la Figure 20 :

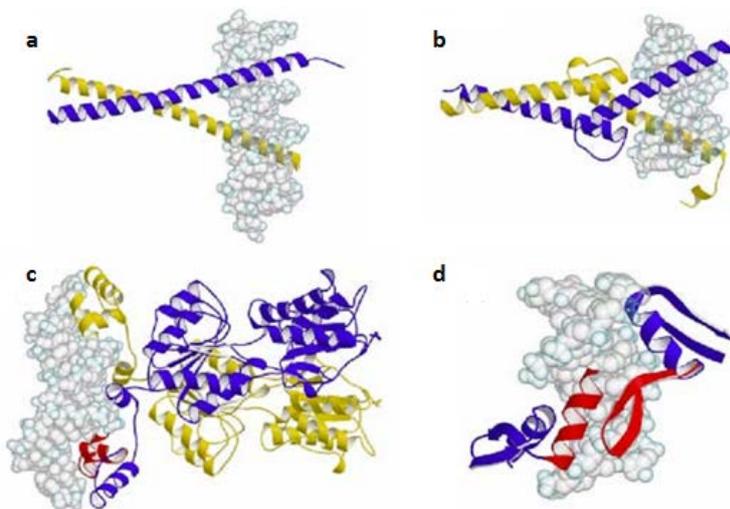


Figure 20 : Différents motifs protéiques interagissant avec l'ADN.

A, fermeture éclair à leucine, b, hélice-boucle-hélice, c, hélice-tour-hélice, d, doigt de zinc [233].

- a) La glissière à leucine [234] possède deux régions : une région constituée par deux hélices alpha, face à face, riches en leucine, qui interagissent par des liaisons hydrophobes et une autre région riche en charges positives, qui se fixe sur les groupes phosphate de l'ADN [235].

- b) Les motifs hélice-boucle-hélice sont similaires aux glissières à leucine excepté que la dimérisation des domaines de ces protéines consiste en deux hélices séparées par une boucle. Ces protéines sont constituées par un domaine riche en acides aminés basiques qui interagit avec l'ADN.
- c) Les motifs hélice-tour-hélice consistent en trois, parfois quatre, hélices alpha reliées entre elles par un coude. Le repliement tridimensionnel de ce domaine assure une forme qui permet une interaction spatiale avec la double hélice d'ADN. Les protéines qui contiennent ce motif agissent à l'état dimérique [236].
- d) Les motifs en doigt de zinc sont constitués de boucles dans lesquelles une hélice α et un feuillet β sont associés à un ion zinc. Les quatre acides aminés qui permettent la liaison du Zn^{2+} sont des cystéines ou des histidines. Elles forment une structure en carré qui délimite un espace contenant l'ion. La région protéique située entre deux doigts se lie à l'ADN et confère la spécificité de l'action de ces facteurs de transcription.

1.2.4.3 Maîtrise des dimensions

Deux classifications des dimensions des nanostructures peuvent être envisagées selon que l'on considère les propriétés géométriques ou les critères topologiques.

Si l'on considère les **propriétés géométriques**, les nanostructures sont classées en une, deux ou trois dimensions selon le motif impliqué. On peut citer en **structures 1D** la formation de longues structures linéaires générée par l'assemblage entre les extrémités cohésives des motifs d'ADN DX (double croisements) [202, 237, 238]. Les **structures 2D** impliquent les maillages ou réseaux réalisés par d'autres motifs d'ADN tels que des triangles ou des faisceaux hélicoïdaux qui, selon le nombre de sous-unités complémentaires, peut générer des auto-assemblages périodiques ou aperiodiques, également étudiés par l'approche algorithmique. Les **structures 3D** concernent les objets en volume tels que la construction du cube d'ADN [182], ou encore le réseau 3D régulier d'ADN développé par Seeman pour faciliter la cristallisation des macromolécules [190].

Toutefois, si l'on considère des **critères topologiques**, cette classification est totalement réorganisée car l'approche topologique consiste à décrire les structures sur la façon dont sont connectés les points de branchements [239].

Les **topologies à zéro-dimension** concernent la construction d'objets séparés tels que le cube ADN, l'octaèdre tronqué ou les anneaux de Borromean qui ne sont pas conçus pour interagir avec d'autres structures. Sont également inclus dans cette catégorie la formation d'octaèdre d'ADN par la technique des origami [240] ainsi que la plupart des nano-moteurs [241] même si les structures elles-mêmes peuvent être en 3D.

Les **topologies à une dimension** impliquent les réseaux linéaires où l'assemblage de géométrie définie est réalisé par interaction entre motifs. On peut reprendre ici l'exemple des Conception de nano-objets à base d'acides nucléiques et de protéines

réseaux linéaires réalisés à partir des motifs d'ADN DX ou encore la formation de réseaux périodiques ou apériodiques.

Dans **les topologies à deux dimensions**, les motifs ADN cités ci-dessus qui permettent d'augmenter le nombre d'interactions entre les briques moléculaires sont également utilisés. Les briques d'ADN sont conçues pour s'auto-assembler en réseaux périodiques à deux dimensions. Par ailleurs, les réseaux à deux dimensions qui peuvent contenir des briques modifiées chimiquement permettant un ancrage et ou un positionnement sur la surface présentent des applications potentielles en électronique moléculaire et bio-détection [179, 229].

Les topologies à trois dimensions : on revient ici à l'idée originale de N. C. Seeman, en 1982, d'utiliser un réseau 3D régulier d'ADN comme guide. C'est en 2004 qu'un réseau continu d'ADN en 3D est formé à partir de l'auto-assemblage de séquences de 13 bases. La structure consiste en des couches empilées d'hélices parallèles avec des couches adjacentes liées par des appariements non canoniques [242]. Le même dessin pourrait être utilisé pour générer des structures avec des dimensions de réseau variables, et servir comme un prototype de tamisage ou de structure d'ADN matrice pour l'inclusion d'autres molécules.

La dimension du temps. L'effort exercé pour contrôler les procédés d'auto-assemblage des structures d'ADN dans le maximum de dimensions s'applique également au temps. L'objectif est d'arriver à des mouvements contrôlés de la matière à l'échelle moléculaire, qui peuvent résulter en un changement de forme ou de taille de la construction. Des nano-dispositifs à base d'ADN ont été conçus pour changer de forme en étant activés par des mécanismes variés tels que des petites molécules, des protéines et même d'autres molécules d'ADN [243, 244]. L'obtention de mouvement de ces constructions repose sur différents principes incluant les transitions de conformation de l'ADN [245-247], les équilibres de déplacement de brins [248, 249] et la fixation des protéines [250].

1.2.4.4 Avantages et limitations des différentes stratégies d'auto-assemblage

La stratégie d'auto-assemblage parallèle, où toutes les briques élémentaires sont mélangées en une fois, présente l'avantage d'être une technique rapide mais son succès repose sur un important travail en amont de conception des briques pour éviter la formation de structures non désirées. La combinatoire sera d'autant plus élevée que le nombre de briques sera grand, elle peut constituer une limite en termes de diversité des briques à utiliser sachant que chacune doit interagir de manière spécifique pour conduire à la structure finale. Expérimentalement, les paramètres de concentration des briques et de vitesse d'association dépendant eux-mêmes des conditions d'assemblage sont critiques. L'absence d'étapes intermédiaires de purification constitue un avantage certain en termes de rendements des structures finales.

La stratégie séquentielle présente l'avantage d'obtenir des nanostructures purifiées à homogénéité. Toutefois, réalisée en plusieurs étapes, cette technique est coûteuse en temps. Selon les procédés de purification utilisés (ancrage sur une surface ou en solution), elle peut présenter de faibles rendements. De cette manière, une quantité considérable de briques de construction peut être nécessaire pour obtenir les structures désirées et réaliser diverses expériences de caractérisation, d'interaction ou d'activité des objets.

Lors de la conception par modélisation d'assemblage séquentiel, la difficulté d'induire l'assemblage provient de la combinatoire complexe nécessaire pour réaliser des objets irréguliers mais de forme bien définie ou des assemblages non périodiques. Plusieurs solutions sont proposées pour contourner les difficultés combinatoires, telles que la conception des tuiles, l'algorithme d'auto-assemblage et le protocole d'assemblage [168, 191].

Une variété importante de formes cohésives peut être générée. Différentes formes de codage au sein de la molécule peuvent impliquer aussi bien les hybridations de type Watson-Crick que la géométrie de la molécule. Le système peut également être guidé ou favorablement contraint par des complexes intermédiaires et des briques de protection peuvent empêcher la formation de complexes indésirables.

Enfin, le protocole d'assemblage peut inclure l'ajout séquentiel d'enzyme ou la présence d'un support pour réaliser l'assemblage séquentiel.

Dans la méthode Origami, de nombreuses règles critiques telles que la pureté des brins, la stricte stœchiométrie entre les brins dans les auto-assemblages à base de briques ADN ne prennent pas effet. En effet, lors de défauts d'appariement des agrafes, le mécanisme de « déplacement de brin » ou « invasion de brin » permet de corriger les erreurs en déplaçant les agrafes incorrectes. De plus, la conception des nanostructures ne reposant pas sur l'hybridation entre agrafes, leur concentration en excès dans la solution n'est pas problématique, et, lors de l'imagerie des structures sur support leur plus faible affinité du fait de leur petite taille permet de laver les agrafes libres tandis que l'ADN matrice beaucoup plus long adhère à la surface. Cette technique en une seule étape est peu coûteuse en temps et produit de bons rendements [251].

Les nanostructures d'ADN origami permettent de réaliser des formes complexes en deux et trois dimensions à façon, toutefois l'étape décisive reste leur interconnexion avec des éléments fonctionnels. Dans le cas des protéines, cette approche pose le problème de compatibilité du fait de l'utilisation de températures dénaturantes lors du protocole de fabrication des nanostructures.

Une autre limitation de l'approche Origami est la taille des nanostructures qui n'excède pas 100 nm du fait de la longueur de l'ADN matrice constitué de 7000 à 8000 bases. Pour remédier à ce problème, les chercheurs plient l'ADN matrice non plus à partir de courts

oligonucléotides appelées agrafes mais par des briques carrées, elles-mêmes conçues par la méthode Origami. La preuve de concept a été donnée par la production de briques de 5x5 et 7x8 briques d'arrangement qui ont permis d'atteindre la taille de 200 nm [252].

1.2.4.5 Approche locale d'auto-assemblage versus à longue distance

Lorsque les systèmes d'auto-assemblage sont basés sur l'utilisation de briques individuelles, on peut distinguer deux conceptions de réalisation. Le cas d'une approche globale où l'objectif est de caractériser le comportement de la population de molécules dans son ensemble et le cas d'une approche locale où ce sont les mécanismes d'association de quelques briques élémentaires qui sont ciblés. Cette dernière approche a donné lieu à de nombreuses publications qui analysent les règles de constructions des briques de base et développent de nombreux modèles afin de contrôler les interactions locales [164, 168, 170, 173]. Ainsi les assemblages gouvernés à un niveau local présentent des structures internes bien déterminées, toutefois leur forme à longue distance n'est pas contrôlée et les outils pour l'analyser assez peu développés. La conformation finale des assemblages n'étant pas forcément définie, ces systèmes ne sont « finis » que lorsque les briques à associer ne sont plus disponibles. L'approche globale, quant à elle, s'intéresse aux structures abouties des briques auto-assemblées. La technique des origamis pourrait constituer un exemple de méthode globale puisqu'elle permet de générer n'importe quelle forme géométrique finale à partir d'un long ADN simple brin plié par l'hybridation avec de courtes agrafes d'ADN [165, 251]. Toutefois, cette technique d'assemblage par nucléation ne repose pas sur l'association de briques élémentaires. L'approche globale d'auto-assemblages réalisés à partir de briques, sans utilisation de gabarit, a été peu abordée dans la littérature.

1.2.4.6 Approche mixte pour l'auto-assemblage de nanostructures en 3D

L'approche mixte qui repose sur la combinaison des approches séquentielle et parallèle est largement utilisée dans la réalisation de structures en 3D. Deux exemples sont cités, qui associent des étapes où tous les éléments sont mis ensemble avec des étapes séquentielles pouvant impliquer la purification d'un produit avant de finaliser l'assemblage ou la fonctionnalisation.

He *et al.* construisent de façon hiérarchique un polyèdre [204] en utilisant l'approche origami avec trois sortes d'ADN simple brin qui ont été assemblés progressivement en motifs symétriques appelés « three-point-star » pour aboutir à la formation du polyèdre en une étape. Les structures finales des polyèdres sont déterminées par la longueur de la boucle (de 3 à 5 bases) qui accroît la flexibilité des briques favorisant les petites structures plutôt que de larges réseaux 2D, et par la concentration d'ADN qui, plus elle est élevée, favorise les larges assemblages tels que les cristaux 2D (Figure 21).

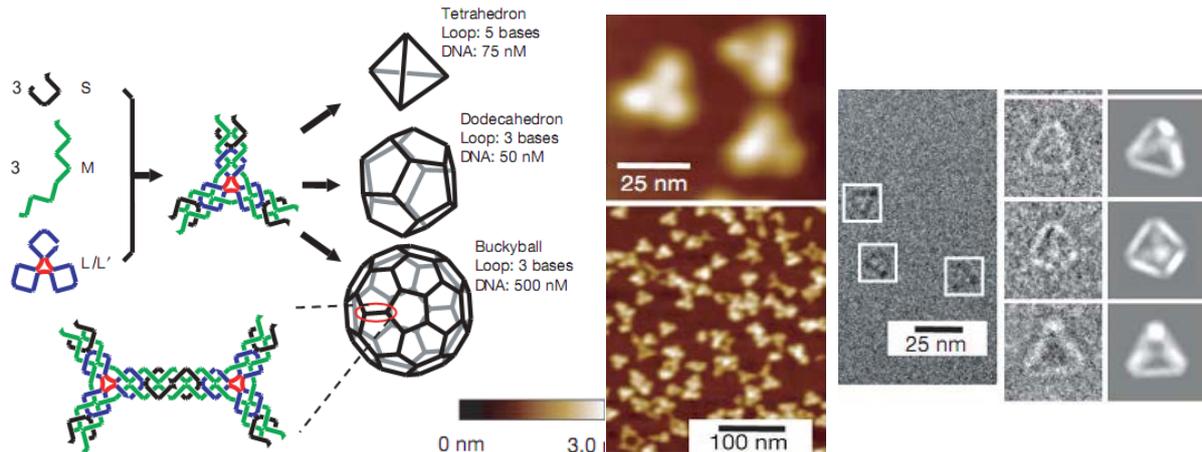


Figure 21 : Schéma d'auto-assemblage séquentiel d'ADN polyèdres, et caractérisation de tétraèdres.

A gauche, schéma d'auto-assemblage de polyèdres d'ADN avec trois sortes d'ADN simple brin qui s'assemblent pas à pas en des motifs « trois points de départ » pour s'assembler ensuite en une seule fois en polyèdres. Trois boucles simple-brin (en rouge) sont au centre du complexe. Les structures finales du polyèdre sont déterminées par la longueur de la boucle et la concentration d'ADN. Au centre, image AFM du tétraèdre d'ADN et à droite, une image représentative de cryomicroscopie électronique (cryoME), les boîtes blanches montrent les particules d'ADN. Quelques images brutes de cryoME des structures 3D du tétraèdre d'ADN et les projections correspondantes reconstruites à partir des images prises dans différentes orientations [204].

De la même façon, E.S. Andersen *et al.* [253] ont fabriqué une boîte d'ADN de $42 \times 36 \times 36 \text{ nm}^3$, qui peut être ouverte par l'ajout d'ADN clé. La première étape, par la méthode origami, a consisté à réaliser les 6 faces du cube à partir de l'ADN génomique simple brin M13 comme illustré sur la Figure 22a, b, c. Ces faces ont été séquentiellement agencées en un cube par l'ajout d'agrafes de liaison entre les faces. La structure en boîte assemblée, purifiée et fonctionnalisée par la présence de deux ADN duplexe avec des extensions, le suivi de l'ouverture du couvercle a été contrôlé par fluorescence avec l'ajout de deux ADN complémentaires qui par invasion de brins ont déplacé les oligonucléotides et ouvert la boîte. Les différentes étapes de construction ont été caractérisées par imagerie AFM (Figure 22d, e, f).

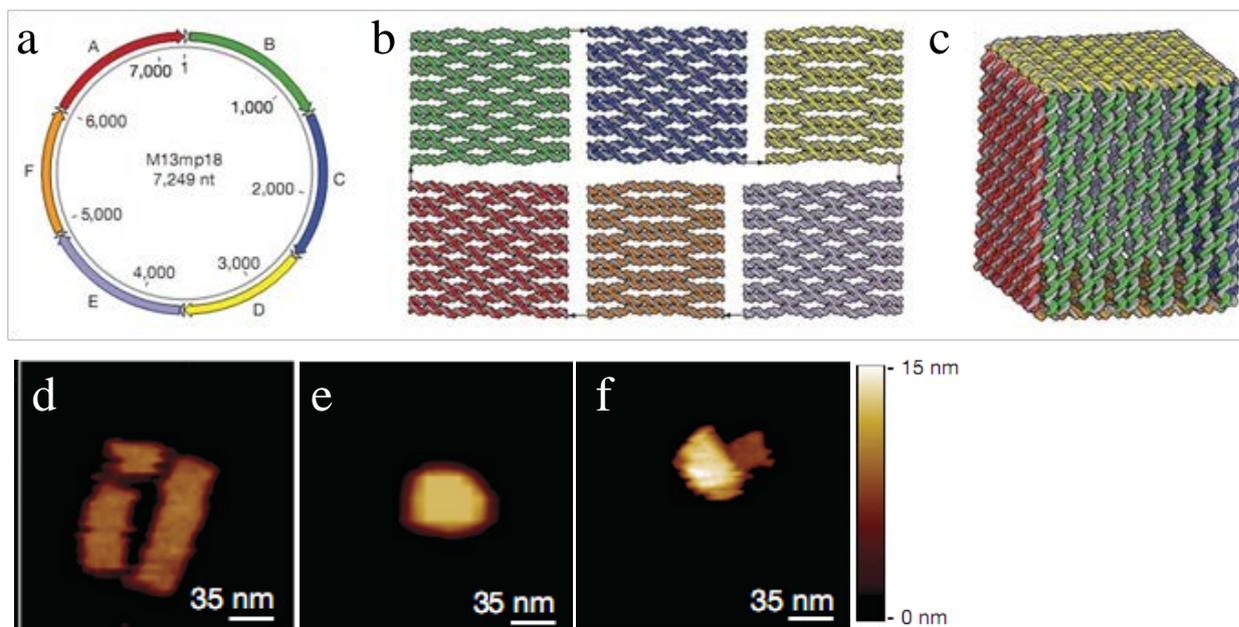


Figure 22 : Conception d'une boîte d'ADN par méthode séquentielle.

En haut, a, carte de la séquence de l'ADN simple brin du bactériophage M13 avec les régions utilisées pour plier les six feuilles d'ADN (A à F). b et c, modèles moléculaires des six feuilles d'ADN de structure plate et cubique respectivement. En bas, images AFM des structures origami en 2D et 3D. d, échantillon avec les six feuilles d'ADN pliées le long du squelette M13. e, échantillon avec la boîte formée par la liaison des côtés des feuilles d'ADN. f, échantillon avec le couvercle de la boîte ouvert [253].

1.2.5 Les auto-assemblages hybrides

Les auto-assemblages hybrides résultent de la combinaison de différentes sources de matériaux organiques, inorganiques et/ou biologiques. Parmi les matériaux bio-hybrides nano-structurés, on peut citer les exemples les plus significatifs qui résultent de la combinaison de polymères naturels tels que les polysaccharides, polyesters, acides nucléiques, polypeptides, protéine fibreuse et globuleuse, les enzymes avec des substrats inorganiques tels que la silice et les phyllosilicates, les hydroxydes double lamellaires, les phosphates et les oxydes métalliques [254]. L'incidence des matériaux hybrides s'étend des domaines de la médecine régénérative jusqu'aux nouveaux matériaux présentant des propriétés fonctionnelles et structurales améliorées. Dans ce paragraphe, nous nous intéressons particulièrement aux assemblages hybrides à base d'ADN et de protéines.

1.2.5.1 Nécessité d'associer structure et fonctionnalité

Le premier problème abordé par les auto-assemblages hybrides est l'apport de la fonctionnalité aux éléments structuraux. En effet, les différentes méthodes d'auto-assemblage des acides nucléiques sont efficaces pour construire des nanostructures à façon de formes complexes et variées mais elles apportent peu d'informations en termes de fonctionnalité. Inversement, les éléments porteurs d'informations ou de fonction de catalyse, d'actuation, de transport, de propriétés optoélectroniques, tels que les protéines ne sont pas forcément adaptés pour créer des structures de manière définie. D'où la nécessité de développer des techniques

efficaces d'intégration d'éléments fonctionnels au sein des structures. Les scientifiques cherchent à développer des structures hybrides qui soient à la fois géométriquement définies et qui constituent des plateformes multifonctions.

1.2.5.1.1 ADN et composés inorganiques : fonctionnalisation et guide.

Des **nanoparticules métalliques** ont été organisées en des structures spatialement définies par l'intermédiaire d'ADN complémentaires greffés sur les nanoparticules [229, 231].

L'ADN est utilisé comme stabilisateur et modèle pour assembler des **nanoparticules semi-conductrices** CdS en utilisant les interactions électrostatiques entre les surfaces cationiques modifiées des nanoparticules et les phosphates de l'ADN [255].

La croissance de **nano-fils métalliques** a été dirigée par l'ADN [183] mais aussi par des peptides [256]; des **nanotubes de carbones** ont été fonctionnalisés aussi bien par des ADN que des PNA (peptide nucleic acid) [257]. Ceci permet d'associer les propriétés conductrices des composés inorganiques avec les propriétés d'assemblages des ADN.

1.2.5.1.2 ADN et protéines

De nombreuses méthodes de couplage chimique entre les oligomères d'ADN synthétiques et les protéines sont utilisées pour la fabrication de structures hybrides. Elles reposent soit sur la stratégie de bio-conjugaison non covalente ou covalente dont quelques exemples sont illustrés sur la Figure 23 [258, 259]. Dans le premier cas, il est nécessaire d'utiliser des mécanismes de reconnaissance protéines-ADN spécifiques et facilement ajustables, dans le deuxième cas, les briques préformées permettent de créer une complexité illimitée.

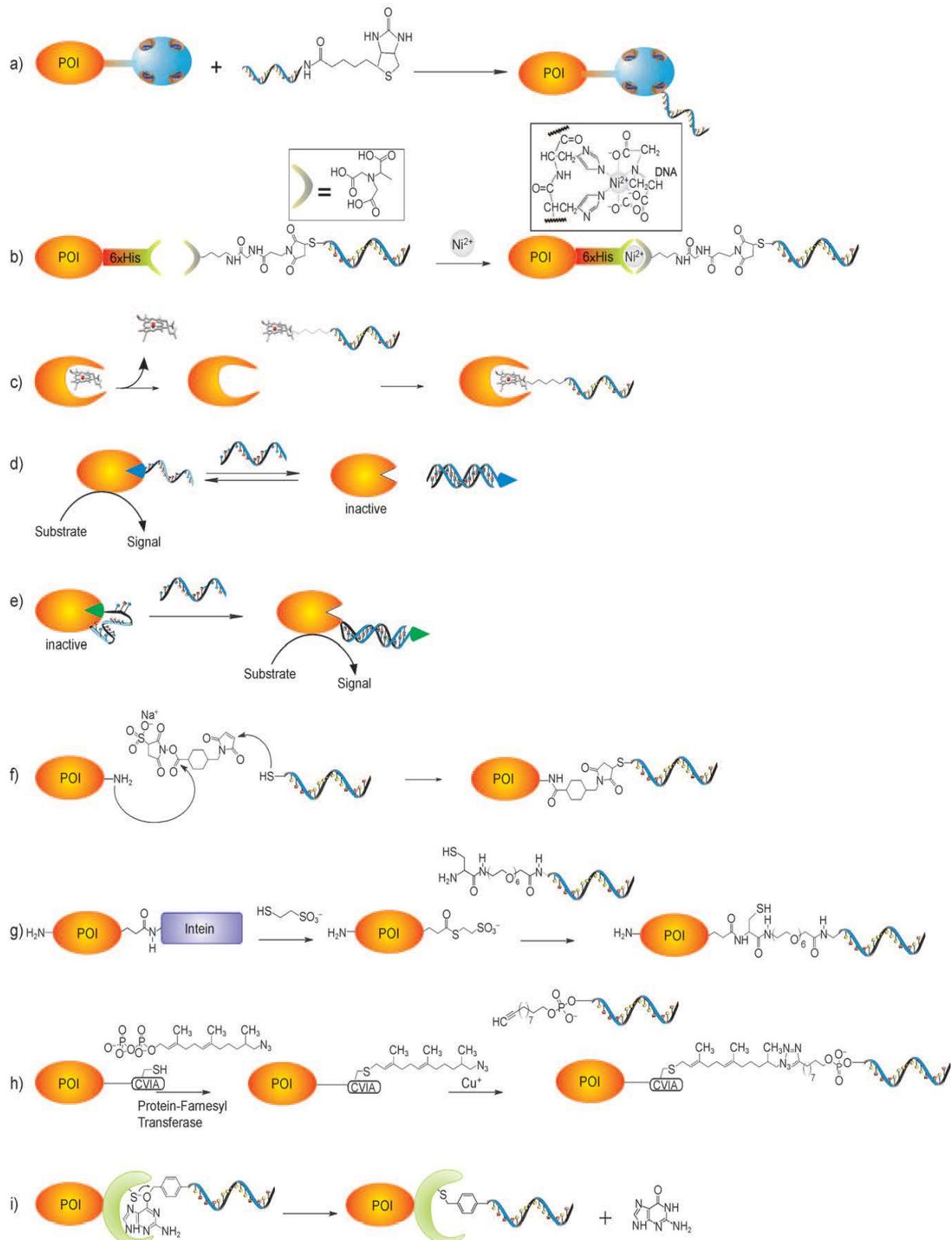


Figure 23 : Stratégies de couplage d'un ADN avec une protéine d'intérêt (POI).

a) couplage biotine-(strept)avidine ; b) conjugaison Ni-NTA-His₆ ; c) Reconstitution d'apoenzymes avec des cofacteurs qui peuvent d) activer ou e) désactiver une activité enzymatique ; le couplage covalent est souvent réalisé par l'intermédiaires d'adaptateurs bi-fonctionnels (f : le sulfoSMCC). Une alternative consiste à former des protéines fusion en utilisant une intéine (g), une modification enzymatique par l'enzyme farnesyltransférase (h), ou l'application de l'O⁶-alkylguanine-DNA-alkyltransférase (une protéine de la réparation d'ADN) (snap-tag, i) en tant que groupe réactif fusionné à la protéine d'intérêt par modification génétique [259].

1.2.5.1.2.1 Couplage non-covalent

L'interaction biotine-(strept)avidine

L'affinité élevée de la biotine avec la streptavidine ($K_a \sim 10^{13} \text{ M}^{-1}$) a été très largement exploitée dans les méthodes de couplage non covalent aussi bien dans le développement de réactifs pour des tests bio-analytiques que pour la nanofabrication (Figure 23a). Ce sont soit les ADN qui sont biotinylés et la protéine conjuguée à l'avidine par voie chimique ou par ingénierie génétique [260], soit l'inverse, les oligonucléotides marqués à l'avidine et la protéine biotinylée [261]. L'affinité est telle que même si la technique est non covalente, elle ne permet pas la dissociation du complexe, à moins d'envisager des méthodes de compétition avec de la biotine libre ou l'utilisation d'une streptavidine monomérique générée par ingénierie [262].

La nature tétramérique de la streptavidine qui lui confère cette haute affinité pose toutefois des problèmes pour contrôler la stœchiométrie lors des couplages protéine –ADN, ce qui constitue un obstacle à la conception de dispositifs nanométriques et d'organisation des structures définies. Par ailleurs la liaison streptavidine plutôt volumineuse peut aboutir à des soucis d'encombrement lors du couplage [259].

L'interaction Ni-NTA-6Histidines

L'acide nitrilotriacétique est un agent chélateur, qui par réaction chimique de complexation, se lie avec les ions métalliques et forme des complexes hydrosolubles. Dans l'interaction Ni-NTA-His, l'ion Ni^{2+} possède six liens de coordination dont quatre sont sollicités par le NTA et deux interagissent avec les atomes d'azote du cycle de la chaîne latérale de deux résidus histidines (Figures 23b et 24A). Les protéines recombinantes avec une étiquette histidine incorporée soit en N- soit en C-terminal sont purifiées par chromatographie d'affinité sur des résines fonctionnalisées NTA et chargées en ions Nickel.

Toutefois l'affinité du mono-NTA étant de l'ordre du μM , l'utilisation de cette interaction trop réversible s'est trouvée limitée dans le domaine de la nano-fabrication. Pour pallier ce problème, A. Tuberfield *et al.* ont utilisé des groupements de NTA par deux (bis-NTA) ou trois (tris-NTA), associés aux oligonucléotides, et les constantes de dissociation étant de 120 et 6 nM, respectivement [263], ils ont pu fabriquer de nombreux complexes protéines-DNA différents pour fabriquer des nanostructures. Ce couplage, site spécifique, a l'avantage d'être réversible par l'ajout d'un agent chélatant qui lie le nickel [264].

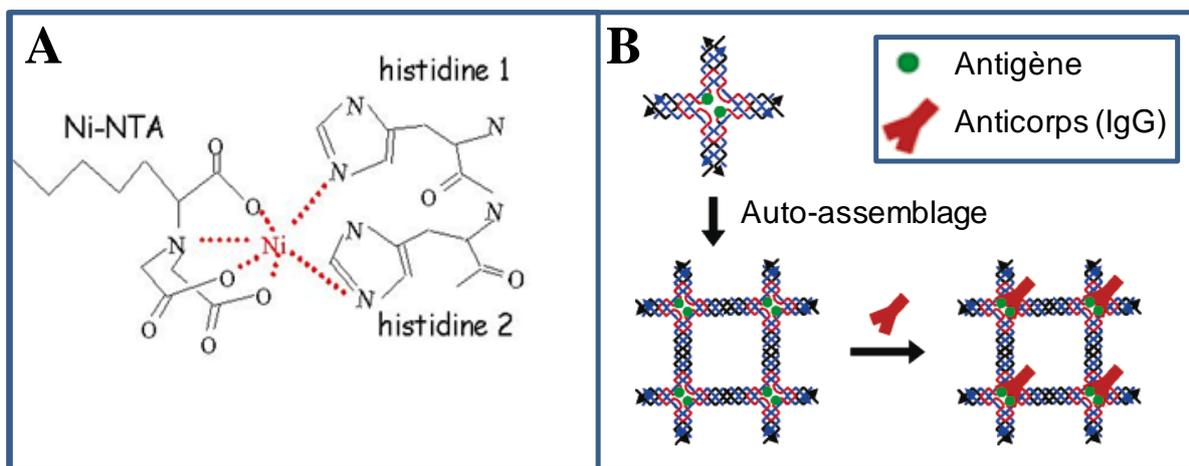


Figure 24 : Deux exemples d'interactions utilisées en couplage non covalent.

A, schéma du complexe nickel-acide nitrilotriacétique chélatant deux histidines. B, schéma d'auto-assemblage par ADN de réseaux d'anticorps (IgG) en deux dimensions. Figure extraite de [265].

L'interaction anticorps-haptène

En plus des ligands biotine et NTA, d'autres étiquettes peuvent être attachées aux oligonucléotides au cours de leur synthèse telles que la fluorescéine, qui peut être utilisée comme un haptène de liaison aux anticorps immunoglobulines anti-fluorescein [266].

L'haptène est une substance de faible masse molaire atomique incapable par elle-même de promouvoir une réaction immunitaire, mais capable de réagir avec des anticorps préformés. Cette interaction plutôt utilisée dans les essais diagnostiques, a permis, en nanofabrication, la formation de réseau d'IgG dirigée par ADN (Figure 24B) [265].

Les aptamères

Les aptamères sont des courts acides nucléiques simple-brin (ADN ou ARN) capables de fixer un ligand spécifique avec une haute affinité. Ils peuvent se plier en des structures 3D pour former des poches de fixation nécessaires à la reconnaissance spécifique et à la liaison d'une molécule cible. Les aptamères peuvent être utilisés dans des applications biotechnologiques, diagnostiques ou thérapeutiques selon la cible contre laquelle ils sont dirigés. En plusieurs jours, on peut produire et identifier *in vitro* ces acides nucléiques à partir de banques combinatoires par le processus SELEX : évolution systématique de ligands par enrichissements exponentiels [267].

Du fait de leur affinité élevée, plusieurs systèmes de capteurs à base d'aptamères ont été réalisés pour la détection d'ions métalliques, de petites molécules, de protéines, et d'acides nucléiques ; la fluorescence, la colorimétrie et l'électrochimie étant les modes de détection les plus communs de ces systèmes [268]. Toutefois, ils ont également été mis à profit dans la fonctionnalisation de réseaux à deux dimensions, ainsi capables de capturer les protéines cibles de manière sélective [269]. Il est à noter que le nombre d'aptamères présentant une

spécificité contre une protéine cible reste limité de nos jours, c'est pourquoi la plupart des études est basée sur l'aptamère antithrombine.

La reconstitution d'apoenzymes

Une approche semi-synthétique de conjugaison à l'ADN a été développée pour les enzymes contenant un cofacteur organique non diffusible qui se situe dans le site actif et est essentiel à l'activité catalytique de l'enzyme. Les dérivés de porphyrine et de flavine constituent les principaux exemples de cofacteurs. Extraits de l'enzyme native, ils peuvent être remplacés par des cofacteurs artificiels du cofacteur natif. Cette méthode est un moyen de générer des hybrides enzyme-ADN dès lors que des acides nucléiques sont couplés au cofacteur (Figure 23c). Réalisés par Fruk et Niemeyer avec un hème couplé à l'ADN rétro-induit dans les protéines d'apo-myoglobine [270] et apo-horse peroxydase [271], les complexes se sont montrés totalement fonctionnels et capables, par hybridation des oligonucléotides, d'ancrage sur des surfaces propices aux applications de capteurs et biocatalyse [272]. Des dispositifs de logique moléculaire ont été réalisés en utilisant l'hybridation des oligonucléotides comme mécanisme d'inhibition ou d'activation de l'enzyme selon que la formation de l'ADN double brin extrait le cofacteur du site actif ou la molécule inhibitrice (Figures d et e, respectivement) [273].

1.2.5.1.2.2 Couplage covalent

Les méthodes de liaison covalente des protéines aux acides nucléiques incluent les pontages chimiques des oligonucléotides aux résidus lysine ou cystéine des protéines, la protéine-ligation exprimée, les méthodes chimio-enzymatiques et l'utilisation de photo-aptamères. Les méthodes de synthèse des complexes peptide-oligonucléotides [274] ne seront pas présentées.

La bio-conjugaison : utilisation des groupes fonctionnels des protéines

La recherche en chimie de bio-conjugaison s'intéresse au développement de méthodes efficaces pour la modification et le couplage de protéines et d'autres molécules biologiques avec un degré élevé de spécificité *in vivo* et *in vitro*. Il n'existe pas de technique idéale, adaptée à tous les cas. On distingue les méthodes de fonctionnalisation des protéines directe et indirecte, cette dernière faisant intervenir un adaptateur bi-fonctionnel lié séquentiellement aux deux composants à coupler.

Les fonctionnalités des protéines les plus utilisées pour la modification chimique spécifique sont l'amine, disponible en position α -N terminale (α -N) du squelette, ainsi que les chaînes latérales d'acides aminés naturels tels que la cystéine (Cys, fonction thiol), la lysine (Lys, fonction amine), la tyrosine (Tyr, fonction phénol) et la glutamine (Gln, fonction amide).

La disponibilité des fonctions ciblées sur la protéine est souvent déterminante dans la voie à envisager, notamment l'exposition de la cible par rapport à la surface de la protéine. Dans le cas où plusieurs chaînes latérales d'acides aminés sont exposées, il devient difficile de contrôler parfaitement la réaction de conjugaison, et donc d'arriver à ce qu'une seule liaison entre la protéine et la cible soit créée. Par ailleurs, la modification doit avoir peu, ou pas, d'effet sur la conformation et la fonction de la protéine en termes d'activité catalytique et spécificité. Une fonctionnalisation sélective nécessite des informations détaillées de structure tridimensionnelle qui peuvent être obtenues par diffraction des rayons X et/ou par spectroscopie par résonance magnétique nucléaire (RMN).

- Les groupements amines

La fonction ϵ -amine de la lysine et la fonction amine α -N terminale sont d'excellents nucléophiles capables de réagir avec bon nombre de groupes fonctionnels électrophiles tels que les aldéhydes et les acides carboxyliques activés.

La réaction entre la lysine et/ou la fonction amine α -N terminale d'une part et les aldéhydes d'autre part conduit à la formation réversible dans l'eau d'une base de Schiff qui peut être réduite en présence de cyanoborohydrure de sodium (NaBH_3CN) pour former un lien amine secondaire. Cette méthode présente l'inconvénient de nécessiter de deux étapes et de la synthèse de l'aldéhyde, souvent compliquée.

La méthode d'activation des acides carboxyliques à l'aide de N-hydroxysuccinimide (NHS) pour réagir avec la lysine ou une fonction amine terminale est très utilisée, toutefois elle a l'inconvénient de présenter une sélectivité faible de la réaction et ce d'autant plus que le nombre de lysines à la surface de la protéine est élevé. Ce procédé de marquage ADN souvent aléatoire aboutit à la formation de complexes hétérogènes avec une stœchiométrie et une régio-spécificité aléatoire [275].

- Les groupements cystéines

La fonction thiol de la cystéine possède une nucléophilie faible, mais elle peut être utilisée dans des réactions de couplage redox. La cystéine est un acide aminé relativement rare et la chimie thiol offre des chemins simples pour attacher à un site spécifique des oligonucléotides [276].

La cystéine peut être modifiée soit par alkylation en utilisant des composés α -halocarbonylé ou des composés maléimide, soit en formant des ponts disulfure. Ces derniers réversibles en présence de réducteurs tels que le dithiothreitol, génèrent des résidus thiol qui peuvent être couplés à des adaptateurs bi-fonctionnels terminés par un groupement pyridyldithiol. La charge globale de la protéine est maintenue lors de la modification des protéines par la cystéine.

- Les groupements tyrosine et glutamine

La modification de la tyrosine est un autre moyen de fonctionnaliser spécifiquement les protéines par des méthodes plus ou moins douces, à pH légèrement basique (>8,5) ou quasi neutre (6,5) [277], tandis que la modification de la glutamine utilise une enzyme transglutaminase (TGase) [278]. Cette enzyme catalyse l'amidation entre la glutamine et différentes amines primaires, permettant d'introduire une cible synthétique dès l'instant qu'elle possède une amine primaire aliphatique exposée.

Même si l'utilisation des groupements fonctionnels des protéines a permis de générer avec succès le couplage covalent de complexes protéine-ADN, cette approche reste dans bien des cas peu régio-spécifique et mal caractérisée en stœchiométrie. C'est pourquoi, des méthodes de couplage orthogonales à la variété des groupements au sein des protéines natives ont été développées [276].

La chimie bio-orthogonale

Les réactions bio-orthogonales sont des réactions qui n'interfèrent pas avec les procédés biologiques. L'objectif originel de cette technique est d'introduire des groupements chimiques non naturels et non perturbants dans les cibles biologiques afin d'étudier les systèmes vivants. L'incorporation de ces groupes non natifs réalisée avec des vitesses rapides et de manière quantitative dans des conditions physiologiques présente aussi l'avantage d'être hautement chimio-sélective [279, 280].

Les réactions les plus connues sont : la ligation de Staudinger des composés modifiés azide et phosphine, la condensation des cétones/aldéhydes avec des nucléophiles N et les réactions azide et alcynes. On peut citer l'exemple d'oligonucléotides modifiés par une phosphine capables de réagir sélectivement avec des glycoprotéines à la surface de cellules « Human Embryon Kidney » marquées avec des groupes azide [281]. Les cellules marquées par les ADN ainsi modifiés ont pu être positionnées sur des surfaces ou leurs propriétés d'adhésion étudiées par la technique de microscopie de force atomique [282].

La chimie click

La chimie click permet de générer des structures originales en utilisant des réactions rapides et efficaces entre différents « blocs » donnant accès à des architectures moléculaires diverses. Une réaction peut être considérée « chimie click » quand elle répond aux critères suivants : modularité, stéréo-sélectivité, pureté et rendement élevés, insensibilité à l'eau et à l'oxygène, absence de groupements protecteurs et pas de purification chromatographique. Les principales réactions click consistent à former des liaisons carbone-hétéroatome énergétiquement très favorables [283].

Les réactions nucléophiles d'ouverture de cycle d'époxydes, d'aziridines, d'ions épisulfonium ou aziridinium permettent d'atteindre une grande diversité de composés avec une régio-sélectivité contrôlée.

Les réactions de cycloaddition de Huisgen de type 1,3 dipolaire catalysée par le cuivre (I) permettent d'obtenir de manière rapide, reproductible et quantitative une grande diversité d'hétérocyles à cinq membres [284]. Cette méthode a été utilisée récemment pour préparer des complexes ADN-protéine en combinaison avec une enzyme de modification (farnesyl-transférase) ; mais aussi pour complexer l'enzyme estérase marquée par un analogue de puromycine portant un groupe azide avec un oligonucléotide modifié en 5' par un alcyné [285]. Par ailleurs, la possibilité actuelle de modifier les oligonucléotides avec des alcynes internes offre de nouvelles opportunités de conjugaison avec les extrémités 5' et 3' libres [286].

Une autre méthode enzymatique utilise la protéine O₆-alkylguanine-ADN-alkyletransférase (AGAT) qui transfère le groupe alkyle des bases ADN et du squelette sucre-phosphate à un résidu cystéine, s'inactivant elle-même (Figure 23i). Aussi appelée « SNAP-tag », cette protéine d'auto-étiquetage développée par Johnsson *et al.*, peut également être fusionnée à une protéine d'intérêt par modification génétique puis traitée par des oligonucléotides benzylguanine [287].

La technique EPL

La technique EPL («expressed protein ligation») aussi appelée ligation des protéines exprimées a été originellement développée pour la synthèse de protéines [288]. Elle consiste à synthétiser des fragments internes de précurseurs de protéines capables de catalyser leur excision et de se réarranger induisant de nouvelles protéines fonctionnelles.

Les protéines recombinantes contenant un thioester C-terminal se couplent de manière spontanée et sélective au N-terminal des cystéines dans des conditions douces [289]. Pour réaliser ce complexe protéine-ADN, la protéine d'intérêt est modifiée génétiquement avec un domaine intéine. Exprimée dans *E. coli*, cette protéine fusion est fixée sur une colonne d'affinité, puis clivée par l'adjonction de composés thiolés, générant une protéine avec un groupe thioester en C-terminal qui pourra être couplée aux oligonucléotides précédemment conjugués à une cystéine. Récemment, Tokeda *et al.* ont développé une nouvelle méthode basée sur les intéines pour le marquage, non plus en C-, mais en N-terminal des protéines via la ligation chimique native [290].

La méthode photochimique

La méthode photochimique, qui consiste en la photo-activation des nucléobases aromatiques, peut aboutir à des liens covalents avec les électrons voisins des chaînes latérales des acides aminés dans une protéine liée. Le rendement de pontage est faible mais peut être

amélioré en utilisant la 5-bromodeoxyuridine (BrdU) plus photoréactive à la place de la thymidine des molécules d'ADN [276].

1.2.5.2 Interfaces et applications

Le deuxième problème traité par les auto-assemblages hybrides est d'une part l'interfaçage avec le vivant et, d'autre part la communication avec l'environnement extérieur. Deux cas se présentent pour que l'échange d'informations puisse se réaliser : soit les nanostructures sont liées à un support, soit elles fournissent elles-mêmes l'information en étant marquées par des fluorophores. Les exemples de couplage cités ci-dessous à base d'ADN ont également été réalisés avec des peptides et protéines.

1.2.5.2.1 ADN –Surface

Différents schémas ont été utilisés pour ancrer les molécules d'ADN sur une surface. Les assemblages d'ADN permettent soit de positionner des structures formées en solution soit d'être en mesure de suivre les interactions moléculaires à la surface dans le cadre de biopuces ou encore d'augmenter le signal quand elles sont couplées avec des nanoparticules d'or [291]. Les techniques covalentes entre les sulfures et l'or ont été largement utilisées, mais il est également possible de s'appuyer sur les systèmes de reconnaissance biotine–avidine [230]. Des assemblages de nanostructures d'ADN contrôlés ont également été réalisés sur des surfaces de mica et de silicone silanisés pour des futurs dispositifs moléculaires [292]. Des surfaces homogènes ont permis l'auto-assemblage d'ADN de par leur confinement qui réduit la flexibilité des motifs d'ADN et favorise la croissance des cristaux d'ADN 2D [293].

La surface peut elle-même constituer un guide pour les structures si elle subit des modifications de propriétés, soit par fonctionnalisation différentielle, soit par structuration. L'ingénierie des surfaces associe à l'idée conventionnelle de fonctionnalisation les technologies de micro- et nano-structuration. Celles-ci peuvent générer des caractéristiques d'ordre mécanique, optique, thermique, chimique ou encore fluïdique différentes de la surface d'origine. La conception de nouvelles puces d'ADN constitue un exemple d'application de l'ingénierie de surface [294].

Les vésicules constituent des supports courbés qui peuvent être fonctionnalisés, ce qui permet d'étudier l'influence de la géométrie du support sur les interactions moléculaires. Le marquage par fluorophores ou boîtes quantiques (QD pour « quantum dot ») permet également une détection par microscopie. Les vésicules sont des nanomatériaux qui présentent des fonctions et propriétés uniques dues à leur petite taille. Elles peuvent être associées à la notion de transport, permettant de délivrer des médicaments dans des tissus ou cellules cibles [295].

1.2.6 Assemblages irréversibles, réversibles et dynamiques

La réversibilité est au cœur du problème de la fabrication des auto-assemblages, soulignant le compromis nécessaire entre stabilité et dynamique/modularité. La notion d'irréversibilité/réversibilité des auto-assemblages fait appel à deux mécanismes de contrôle différents: les mécanismes à l'équilibre thermodynamiques et les mécanismes sous contrôle cinétique hors équilibre. La notion de dynamique des systèmes en terme d'adaptation est apparue avec la chimie dynamique constitutionnelle (CDC) qui est une extension de la chimie supramoléculaire impliquant aussi bien des liaisons non covalentes que des liaisons covalentes réversibles [155].

Les assemblages irréversibles correspondent à des structures ordonnées statiques qui, une fois faites, ne peuvent plus être ajustées ou reconfigurées. Elles ne peuvent pas réaliser de fonctions différentes après changements des paramètres extérieurs. Les assemblages irréversibles peuvent aussi bien être issus de structures réalisées à l'équilibre thermodynamique et « gelées » dans leur état final en s'éloignant de l'équilibre, que de structures générés sous contrôle cinétique.

Les ADN origamis constituent un exemple du premier cas. Ils sont créés en utilisant des hautes températures (90-95°C) qui dénaturent toutes les structures secondaires et hybridations non spécifiques puis sont refroidis lentement, sur 2h ou plus, jusqu'à température ambiante. Cette étape permet aux molécules de s'organiser dans la configuration optimale par rapport à l'énergie minimale. Le contrôle thermodynamique des assemblages présente les avantages :

- d'obtenir des nanostructures correctes avec un faible taux d'erreur,
- de concevoir plus facilement des structures complexes du fait de la prédictibilité de reconnaissance de l'ADN,
- d'éviter les réactions incomplètes.

Le choix des séquences se fait à l'aide de logiciels informatiques qui calculent l'énergie libre globale minimale pour la structure à réaliser, sans prendre en considération les facteurs cinétiques. Cette simplification a donné des synthèses d'auto-assemblage statiques réussis avec des architectures nano-structurées complexes [161, 166, 170, 191].

Toutefois la nécessité d'utiliser des températures élevées ou des conditions de pH non physiologiques ou une étape de ligation, est peu compatible avec l'utilisation d'objets biologiques fragiles, tels que les protéines. L'application du mode d'assemblage sous contrôle thermodynamique se restreint aux acides nucléiques seuls. La décoration par les protéines est alors souvent réalisée séquentiellement, après l'assemblage de la partie nucléique, s'appuyant sur des systèmes de reconnaissance spécifiques.

Les structures conçues sous contrôle cinétique sont tributaires des vitesses des réactions impliquées. Ce mode d'assemblage présente l'avantage d'utiliser des conditions

douces compatibles avec les protéines. Les inconvénients résident dans les possibles pièges cinétiques. Les erreurs d'assemblage qui seraient corrigées dans un processus réversible, peuvent être stables, dans des processus sous contrôle cinétique, et générer des structures abortives. Un travail important de modélisation, se révélant complexe du fait de la combinatoire élevée et des multiples paramètres, est alors nécessaire à la conception des briques à hybrider dans de telles conditions.

Les structures auto-assemblées dynamiques sont capables de s'associer et de se dissocier dans les conditions expérimentales. On distingue les systèmes à l'équilibre thermodynamique et les systèmes dynamiques en équilibre cinétique, dont le maintien de la structure est un processus dispersif qui consomme de l'énergie.

Dans la nature, les organisations biologiques reposent souvent sur des systèmes dynamiques réversibles mais hors équilibre. Ces systèmes peuvent avoir un mode d'organisation et des fonctions modifiables par des facteurs externes. Ils peuvent s'auto-réparer et, dans certains cas, s'auto-répliquer. Pour ce faire, les systèmes biologiques s'auto-assemblent dans des états métastables dépendants d'une source d'énergie externe. Habituellement, ceci implique l'action synchrone de multiples agents ou interactions. Certaines sont destinées à assembler les composants ensemble, d'autres à les séparer [296]. Dans ces systèmes dispersifs, les assemblages macromoléculaires loin d'atteindre un équilibre, alternent sans fin entre différentes configurations se construisant et se déconstruisant. C'est le cas de l'auto-assemblage de la tubuline en microtubules, alternant entre une élongation lente et un rétrécissement rapide contrôlés par l'hydrolyse du GTP [297].

Des systèmes initialement assemblés loin de l'équilibre sous contrôle cinétique peuvent être remaniés par l'ajout de briques ou le changement des paramètres environnementaux [48, 298, 299] mais aussi par des phénomènes processifs. Si on prend l'exemple du phénomène d'invasion de brin, une « micro-réversibilité » existe même à des températures très éloignées de la température de fusion. L'énergie mise en œuvre pour avancer d'une base est faible, permettant d'observer un phénomène localement réversible.

On peut également obtenir des assemblages réversibles en combinant les approches thermodynamiques et cinétiques [300, 301]. La compétition entre les hybridations d'ADN intra- et intermoléculaires contrôle les vitesses de réaction. Les structures intramoléculaires adoptées à basse température par les ADN simples brins ralentissent les hybridations intermoléculaires qui peuvent être accélérées soit par des élévations de température, soit par une augmentation de la concentration locale des molécules dans des périodes de temps adéquats.

Dynamique et évolution : la réalisation d'**auto-assemblages dynamiques** qui seraient la base de matériaux adaptables, auto-réplicatifs ou auto-réparateurs constitue un enjeu qui

nécessite de connaître les principes gouvernant l'auto-assemblage loin de l'équilibre ou des règles heuristiques qui permettent de concevoir de tels systèmes [296].

J. M. Lehn *et al.* ont été des pionniers dans le développement de systèmes dynamiques au sens adaptatif du terme. Ils développent la notion de chimie dynamique constitutionnelle [37], qui s'appuie sur la plasticité des composants moléculaires pour sélectionner, dans l'ensemble des molécules en recombinaison permanente, celles qui présentent le meilleur comportement d'adaptation à diverses contraintes extérieures. Cette chimie adaptative cherche à comprendre et à utiliser les processus d'évolution chimique [302].

2. Paramètres physico-chimiques critiques pour l'assemblage des acides nucléiques

Trois classes de paramètres interviennent dans l'auto-assemblage des acides nucléiques :

- (i) l'hybridation et la façon dont on peut contrôler la stabilité d'un duplex,
- (ii) les aspects cinétiques qui réfèrent à la dynamique de reconnaissance de deux ADN, aux procédés de nucléation-propagation,
- (iii) la rigidité qui inclut les notions de longueur de persistance et cyclisation.

2.1. Mécanismes d'hybridation et stabilité de l'ADN B

2.1.1 Température de fusion

L'hybridation est régie par la température de fusion aussi appelée température de demi-dénaturation (T_m) des acides nucléiques. C'est la température à laquelle 50 % des doubles hélices sont dissociées. La méthode la plus simple pour suivre la dénaturation à partir de l'état natif consiste à déterminer l'évolution du spectre d'absorbance des UV par la solution d'ADN.

Pendant la dénaturation, la séparation des deux chaînes entraîne un accroissement d'absorbance d'environ 27%, c'est l'effet d'hyperchromicité qui provient de la rupture des interactions entre les bases [303]. A 260 nm, l'absorbance de l'ADN simple brin est supérieure à celle de l'ADN double brin car les bases qui ne sont plus ordonnées dans des plans parallèles ne se masquent plus les unes les autres. Le point d'inflexion de la courbe en « S » correspond au T_m (Figure 25).

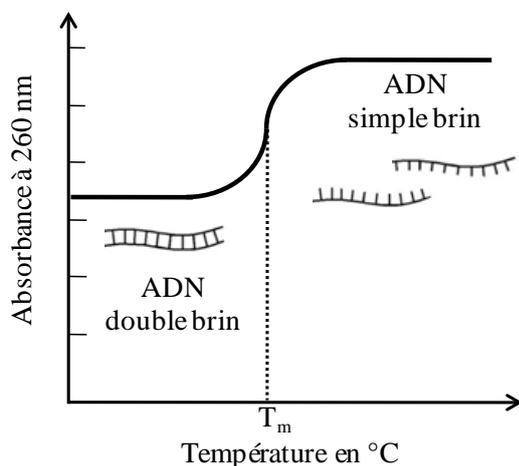


Figure 25 : Température de fusion.

Détermination de la température de fusion par la mesure d'absorbance d'une solution d'ADN double brin à différentes température et mise en évidence du caractère coopératif de la réaction [304].

2.1.2 Calcul de la température de fusion

Différentes formulations phénoménologiques ont été proposées pour calculer la température de fusion d'une chaîne d'ADN, seule la méthode du plus proche voisin, la plus couramment utilisée, est présentée ci-après. Cette méthode basée sur l'étude thermodynamique des courbes de dénaturation des double-brins d'ADN, fait intervenir les paramètres d'entropie (ΔS°), d'enthalpie (ΔH°) et d'énergie libre (ΔG) selon la formule :

$$T_m (\text{°C}) = \Delta H^\circ / (\Delta S^\circ + R \ln[\text{oligo}]) + f(Na^+) - 273.15, \text{ avec}$$

R la constante des gaz parfaits ($1,98 \text{ cal K}^{-1} \text{ mole}^{-1}$),

[oligo], la concentration molaire de l'oligonucléotide,

et la constante -273.15 qui convertit la température Kelvin en degré Celsius.

Les paramètres de calcul les plus précis ont été obtenus à partir des publications de J. Santa Lucia et al. [305-308].

La température de fusion dépend également des concentrations en sels mono- et divalents (Na^+) et (Mg^{2+}) du solvant. Le logiciel d'analyse des oligonucléotides utilisé (<http://eu.idtdna.com/Analyzer/Applications/Instructions>) emploie la fonction de correction pour les ions Na^+ et Mg^{2+} extraite des travaux de R. Owczarzy *et al.* [310, 311].

2.1.3 Paramètres influençant la température de fusion

La stabilité de la double hélice est caractérisée par la température de fusion qui dépend principalement de (i) la longueur de la séquence, (ii) la composition en bases, (iii) la concentration en sels de la solution, mais également (iv) l'ordre de succession des bases, (v) du nombre de bases complémentaires et (vi) de la concentration en monobrins.

Plus la séquence est longue, plus l'énergie requise pour dénaturer l'hélice est grande donc plus la structure est stable. L'effet de longueur est important pour les séquences courtes (<50 bases), de même que l'ordre des bases qui est pris en compte dans la méthode du plus

proche voisin. La température de fusion croît linéairement avec le nombre de paires G-C pour des fragments supérieurs à 200 paires de bases.

La concentration en sels de la solution joue un rôle important sur la température de fusion. Plus la concentration des cations monovalents est élevée (1M), plus les charges négatives des groupements phosphate de l'ADN sont masquées minimisant les forces de répulsion des deux brins d'ADN. Ceci favorise l'hybridation des séquences complémentaires et augmente la température de fusion. Les valeurs de pH inférieures à 5 et supérieures à 9 peuvent ioniser les bases et détruire les liaisons hydrogènes, affectant la stabilité de la double hélice.

La concentration d'ADN a un effet sur l'hybridation entre deux brins d'ADN qui est conditionnée par la présence des brins face à face, phénomène aléatoire régi par la fréquence des rencontres des molécules, elle-même régie par le coefficient de diffusion.

2.1.4 Cinétique d'hybridation

Trois phénomènes sont impliqués dans les vitesses d'hybridation : (i) la fréquence de rencontre des molécules, (ii) l'étape de nucléation et (iii) l'étape de propagation.

- (i) La fréquence de collision des molécules est influencée par le coefficient de diffusion D , exprimé en $\text{cm}^2 \cdot \text{s}^{-1}$ qui dépend de la vitesse de déplacement d'une molécule dans un milieu donné. La diffusion des molécules est conditionnée par la taille et la forme de la molécule, par la viscosité du milieu et par la température selon l'équation Einstein-Stokes suivante :

$$D = \frac{k_B T}{6\pi\eta r}$$

- Où k_b est la constante de Boltzmann qui relie l'énergie au niveau de la particule individuelle avec la température observée à un niveau collectif ($\left(\frac{R}{N_A}\right)$, R étant la constante des gaz parfaits et N_A la constante d'Avogadro)
 - T la température absolue,
 - η la viscosité,
 - et r le rayon équivalent d'une sphère de la particule.
- (ii) L'étape de nucléation qui est un procédé multimoléculaire au cours duquel le brin établit les contacts initiaux productifs, constitue l'étape limitant la renaturation de l'ADN. La réaction de second ordre est décrite différemment selon que l'on considère la rencontre aléatoire des brins complémentaires ou la longueur et complexité de la séquence.

Dans le premier cas, la renaturation de l'ADN est souvent considérée comme ayant une vitesse inversement proportionnelle à la complexité du génome.

Dans le deuxième cas, ce sont des mesures expérimentales réalisées par J. G. Wetmur et N. Davidson, lors de leurs premières investigations dans les années 60, qui ont permis de caractériser les cinétiques de renaturation de l'ADN [312]. L'équation reliant la constante de vitesse de nucléation k'_N et la constante de vitesse observée k_2 est la suivante :

$$k_2 = k'_N L^{0,5} / N$$

où L et N sont la longueur et la complexité de la séquence d'ADN, respectivement. L'origine et la valeur de l'exposant $\alpha=0,5$ sont largement discutées dans la littérature, la valeur de 0,51 ayant été récemment estimée par Sikorav *et al.* à partir de l'analyse de données expérimentales d'ADN de différents organismes dans des conditions de sel et de température différentes [313].

(iii) L'étape de propagation, dans laquelle la chaîne double brin est étendue, est un procédé pseudo-mono-moléculaire (Figure 26). Formellement, plus d'un brin est impliqué mais la formation d'une base paire fermée adjacente à une paire de base préformée est un procédé mono-moléculaire analogue à l'extension d'une épingle à cheveux.

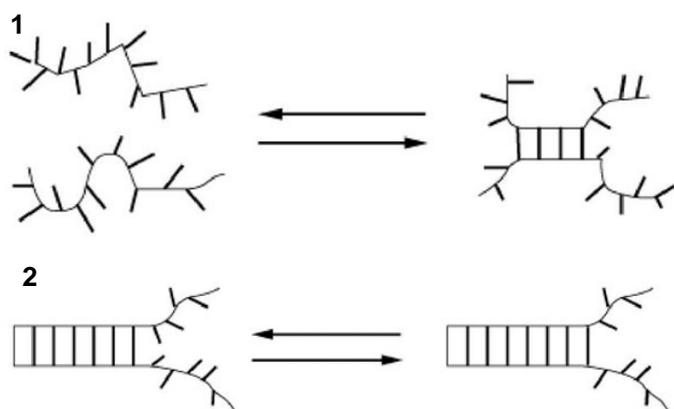


Figure 26 : Evènements de nucléation et de propagation.

1, évènement de nucléation, le nombre de brins interagissant change. Le procédé est bi-moléculaire et dépend entre autres de la concentration des brins. 2, évènement de propagation. Le procédé est pseudo-mono-moléculaire car le nombre de brins dans le complexe est invariable.

Au cours de l'hybridation ou renaturation, la probabilité d'une complémentarité exacte des bases dans le complexe initial est très faible. A l'inverse, de petits segments de duplex sont nucléés, et peuvent croître dans le cas où les séquences le permettent. Si l'énergie thermique est suffisante (c.à.d. la température suffisamment élevée), les régions nucléées se formeront et se dissocieront tant que le système cherche l'état de plus basse énergie. Si l'énergie thermique est insuffisante, les complexes improductifs domineront. Ainsi, l'hybridation est optimisée par l'incubation à la température de fusion, suivie d'un lent refroidissement qui permet la réorganisation des régions duplex, pour optimiser l'alignement de séquence.

Pour les fragments longs, la recherche des complémentarités domine la cinétique d'hybridation alors que pour les oligonucléotides courts, c'est la fréquence de rencontre des

brins complémentaires qui va dominer la formation du complexe. Dès que la nucléation se produit, l'appariement des bases restantes est rapide.

La formation de duplex pour les oligonucléotides courts peut être entravée par la formation de structures intramoléculaires, laquelle est favorisée à basse concentration et à faible force ionique. Les deux mécanismes de formation de structures intra- et intermoléculaires (duplex) peuvent donc être distingués expérimentalement par la dépendance de concentration.

Cas particulier de la vitesse d'hybridation sur une surface : Lorsque les molécules présentent des interactions avec le support, les vitesses d'hybridation sont réduites. Des résultats différents sont rapportés :

- Y. Gao *et al.* concluent que les vitesses d'hybridation sur les surfaces sont proportionnelles à celles en solution avec un facteur de diminution de 20 à 40, selon les séquences considérées et leurs structures secondaires [314].
- M. M. Sekar *et al.* observent une situation très contrastée entre les séquences aléatoires ou présentant peu de structures secondaires et les séquences comportant des structures secondaires connues. La vitesse d'hybridation diminue de trois ordres de grandeurs pour les premières alors qu'elle diminue faiblement pour les deuxièmes passant de $3 \times 10^3 \text{ M}^{-1} \cdot \text{s}^{-1}$ en solution à $2,5 \times 10^2 \text{ M}^{-1} \cdot \text{s}^{-1}$ en phase solide [315].

2.2. Modèles de description des polymères semi-flexibles

Deux modèles sont couramment utilisés pour décrire les polymères semi-flexibles et leur réponse à une contrainte extérieure : le modèle de la chaîne à articulations libres (FJC pour « freely jointed chain ») [316, 317] et celui de la chaîne à courbure continue (WLC pour « worm-like-chain ») [318-320]. Dans le premier modèle, le polymère est une succession de N segments indépendants de longueur b_k (segments de Kuhn) dont les orientations ne sont pas corrélées en l'absence de force. La rigidité de la molécule est prise en compte dans son ensemble sans tenir compte des corrélations d'orientation à petite échelle. C'est pourquoi, on utilise plutôt le modèle WLC proposé par O. Kratky et G. Porod [320], dans lequel le polymère est considéré comme un fil ayant une courbure continue. Il introduit la notion de rigidité de courbure.

2.2.1 La longueur de persistance pour quantifier la rigidité d'un polymère

La longueur de persistance est un paramètre important qui caractérise la flexibilité des polymères. Plus la longueur de persistance est grande, plus la molécule est rigide. L_p représente la longueur de la chaîne sur laquelle la direction des segments est corrélée. Si on définit un angle Θ entre le vecteur tangent au polymère à une position zéro et le vecteur tangent à la distance L loin de la position, il a été montré que la valeur attendue du cosinus de l'angle Θ décroît de manière exponentielle avec la distance :

$$\langle \cos^N \Theta \rangle = e^{-\left(\frac{L}{P}\right)}$$

avec P la longueur de persistance et les crochets indiquant la moyenne de l'ensemble des positions de départ.

La longueur de persistance ne dépend pas de la longueur totale du polymère. C'est une propriété intrinsèque du polymère dans un milieu donné. Le modèle WLC permet de bien décrire les chaînes longues de l'ADN qui est un polymère semi-flexible, présentant une certaine rigidité conférée par sa structure en double hélice et une mobilité soumise aux forces de répulsions électrostatiques des charges de ses phosphates. Une longueur de persistance de 50 nm soit 150 bases a été déterminée pour l'ADN double brin plus rigide que l'ADN simple-brin qui présente une longueur de persistance de 0,8 à 3 nm, laquelle dépend fortement de la force ionique de la solution [321, 322]. La longueur de persistance augmente quand la force ionique diminue, en effet les charges portées par le squelette sucre-phosphate sont moins écrantées par les ions.

Les premières mesures de la flexibilité de l'ADN ont été réalisées sur des solutions d'ADN par diffusion dynamique et statique de la lumière [323], analyse de la vitesse de sédimentation [324] ou viscosimétrie.

B. Tinland *et al.* [325] ont déterminé la longueur de persistance de fragments d'**ADN simple brin** de différentes tailles par la mesure du coefficient de diffusion. En négligeant les effets de volume d'exclusion, la longueur de persistance se décrit comme la somme de la longueur de persistance seule qui résulte de la rigidité intrinsèque de la chaîne et de la contribution électrostatique qui dépend de la force ionique. Ils concluent que l'ADN simple brin présente une longueur de persistance totale d'environ 4 nm pour une force ionique de 10^{-2} M en sels, cette valeur pouvant descendre à 0.8 nm pour des forces ioniques supérieures à 0.1M.

La théorie associée aux approches citées plus haut implique des approximations dans les modèles hydrodynamiques qui peuvent induire des erreurs d'interprétation. D'autres méthodes de mesure de la longueur de l'ADN ont été utilisées dont la méthode de cyclisation de l'ADN catalysée par des ligases.

2.2.2 Cyclisation : étude du facteur J pour déterminer la longueur de persistance

Le facteur j qui représente l'efficacité de cyclisation des fragments d'ADN [326] a été largement étudié pour déterminer la longueur de persistance d'ADN double-brin et double brin segmentés avec des parties simple brin. La cyclisation est la possibilité pour une molécule d'ADN de former une boucle en joignant ses deux extrémités [327-329].

Le facteur j est la concentration efficace à laquelle une extrémité d'un fragment se trouve à proximité de l'autre. Sa valeur peut être déterminée en présence d'une ligase, qui va créer un lien covalent entre les deux extrémités. Ainsi, les fragments forment des monomères

circulaires, des dimères linéaires et circulaires, des trimères... Les produits ligués sont séparés sur gel d'électrophorèse afin de mesurer les proportions relatives des monomères circulaires $C(t)$, et des dimères linéaires et circulaires $D(t)$ (Figure 27). La valeur du facteur j est calculée :

$$j = 2M_0 \lim_{t \rightarrow 0} \frac{C(t)}{D(t)}$$

avec M_0 la concentration initiale du fragment et t le temps de réaction [330]. Pour réaliser l'extrapolation de $C(t)/D(t)$ au temps zéro, le ratio doit être mesuré pour quelques valeurs de t . Une quantification typique d'une telle expérience est montrée sur la Figure 27C.

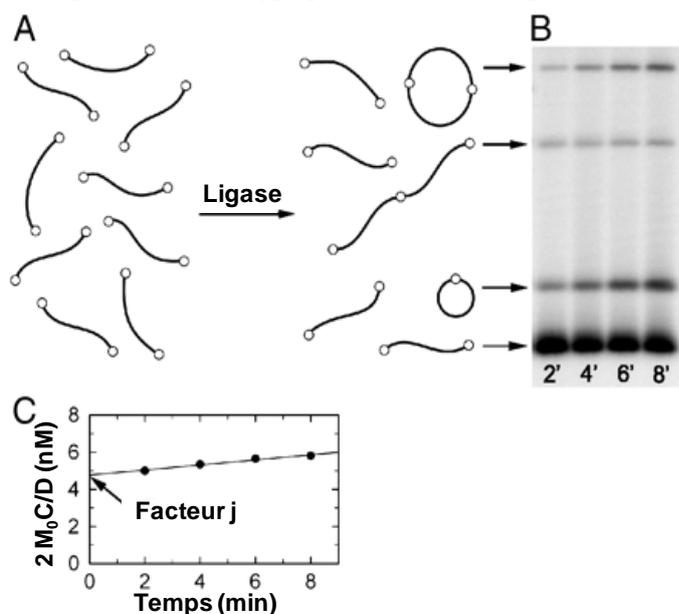


Figure 27 : Détermination du facteur j pour un fragment d'ADN avec de courtes extrémités collantes.

A, diagramme de la ligation des fragments. Les bouts collants sont représentés par des petits ronds ouverts à l'extrémité des fragments. B, séparation des produits ligués par électrophorèse sur gel d'agarose. C, extrapolation de $2 M_0 C(t) / D(t)$ au temps zéro ligation pour obtenir la valeur du facteur j . Figure extraite de [331].

En mesurant la vitesse de cyclisation de l'ADN en présence d'une ligase pour des molécules d'ADN de différentes tailles, D. Shore *et al.* [332] ont mis en évidence la flexibilité de brins courts d'ADN. Cette méthode a l'avantage de ne pas utiliser de modèle hydrodynamique mais doit être menée à très faible concentration pour éviter la formation d'agrégats.

Le facteur j est également lié aux paramètres de répétition du tour d'hélice (γ) et de rigidité de la torsion de la double hélice (C). Les propriétés de conformation des fragments d'ADN double brin linéaires sont bien décrites par le modèle du ver WLC [333, 334] et présentent les valeurs : $a=48,5$ nm, $\gamma=10,5$ pb/tour d'hélice et $C=2,4 \times 10^{-19}$ erg x cm corroborées par de nombreux travaux [327, 330, 335]. La forte contrainte de torsion pour les courts fragments d'ADN double brin implique des oscillations du facteur j à chaque tour d'hélice dont la valeur dépend de la séquence d'ADN. M. Vologoskaia *et al.* [334] ont mesuré le facteur j de courts fragments d'ADN **double brin** couvrant une période d'oscillation afin de déterminer leur longueur de persistance. Ils ont trouvé la valeur de 48 nm (+/- 1) en accord

avec la plupart des publications, toutefois ils ont montré que la contribution de la courbure intrinsèque était 20 fois plus petite que celle des fluctuations thermiques pour les brins courts (200- 250 pb).

Q. Du *et al.* ont étudié **des ADN double brin avec des parties simple brin** [327, 336] afin de remédier aux limites de l'approche de cyclisation (oscillations du facteur j). Ils ont ainsi déterminé les courbures intrinsèques et induites de l'ADN en présence d'enzymes de restriction spécifiques. La longueur de persistance des ADN double brin mesurée est également de 48,5 nm tandis que pour les parties simple brin cette valeur est 13 fois plus petite (3,7 nm).

P. Furrer *et al.* [337] ont observé un effet opposé des contre ions sur la longueur de persistance des fragments d'ADN doubles brins complets et doubles brins avec un phosphate en moins tous les 20 nucléotides. La structure « complète » présente une longueur de persistance de 80 nm (du fait de sa séquence spécifique) qui décroît jusqu'à 54 nm lors d'une augmentation de la force ionique, tandis que la structure « endommagée » qui présente une valeur de 22 nm à faible force ionique double à 44 nm lors de l'augmentation de la force ionique.

3. Outils de caractérisation des nanostructures

3.1. Molécule unique & population

On distingue deux approches selon qu'elles concernent une population de molécules ou une molécule unique. L'approche globale consiste à étudier les phénomènes physiques ou biologiques d'une population de molécules et révèle l'état moyen dans lequel se trouvent ces molécules. Ce sont des mesures d'ensemble qui permettent de déterminer une valeur moyenne sans accès à la distribution des états. Quant à l'approche molécule unique, elle permet d'accéder à la distribution d'état des molécules. Elle est adaptée à l'étude des phénomènes biologiques pour lesquels il existe des sous-populations de molécules avec des activités, des constantes de réactions ou des dynamiques distinctes. Cette approche a l'avantage de fournir une image précise de la géométrie des objets mais la technologie d'observation elle-même peut introduire un biais du fait des interactions avec l'objet.

3.2. Domaines spatio-temporels des différentes techniques d'analyse

Quelques caractéristiques des différentes techniques de microscopie électronique à balayage, microscopie électronique en transmission, microscopie confocale et microscopie de force atomique vont être présentées ci-après. Les informations fournies par ces techniques seront complémentaires sans être comparables du fait des échelles d'observation très différentes et du mode spécifique de préparation des échantillons [338].

La microscopie électronique à balayage présente une résolution allant de 300 à 5 nm avec un facteur de grossissement jusqu'à ~2000, les tensions de travail variant de 10 à 30 kV. Le champ d'analyse peut couvrir ~25 μm^2 . Les échantillons sont analysés sous vide.

La résolution de la microscopie électronique en transmission va de 300 nm à quelques angströms, avec des tensions de travail de 100 kV et plus. Le facteur de grossissement est d'environ 10000 et le champ d'analyse d'environ 100 nm. On distingue la cryomicroscopie de la microscopie électronique en transmission, qui permet d'imager des objets en trois dimensions du fait du piégeage des molécules dans la glace (voir plus loin).

Les résolutions limites d'un microscope confocal sont approximativement de 150 nm dans le plan (x,y) et de 400 nm dans l'épaisseur z de l'échantillon. La résolution latérale en microscopie confocale est améliorée d'environ 30% par rapport à la microscopie conventionnelle dont la résolution est donnée par la loi d'Abbe :

$$\delta_{xy} = 0.61\lambda/N.A.$$

avec λ la longueur d'onde en nm et N.A. l'ouverture numérique de l'objectif. En considérant $\lambda = 488$ nm et NA=1.4, on obtient une résolution latérale de 160 nm au lieu de 210 nm en microscopie conventionnelle [339]. La résolution axiale δ_z est donnée par

$$\delta_z = 1,4 n\lambda/N.A.^2$$

avec n l'indice de réfraction du milieu (1.33 pour l'eau et 1.5 pour l'huile), soit des valeurs de l'ordre de 400 à 600 nm pour un objectif d'ouverture numérique 1.4 à $\lambda = 488$ nm.

La microscopie de force atomique présente une résolution **latérale** et verticale de 0.1 nm avec un facteur de grossissement de ~100000 et permet d'imager des surfaces de 100 nm^2 à 10 μm^2 . L'observation peut être réalisée sous atmosphère contrôlée ou non, en milieu liquide ou sous pression.

3.3. Microscopie électronique

La microscopie électronique en transmission (TEM), fondée sur un contraste électronique, permet de visualiser des particules isolées (ADN, protéines ou complexes) avec une très bonne résolution spatiale (quelques nm) mais ne permet pas d'appréhender directement les aspects dynamiques. Les molécules sont immobilisées sur un substrat mince (épaisseur entre ~5 nm et ~100 nm). L'observation se fait dans une colonne optique électronique sous vide. La visualisation de l'ADN et des protéines nécessite un traitement particulier pour obtenir un bon contraste électronique : en général, soit un ombrage métallique de l'échantillon, soit le marquage par un sel de métal lourd.

En cryomicroscopie électronique en transmission (cryoTEM) les molécules ne sont plus étalées sur une surface 2D mais restent dans un espace 3D : elles sont incluses dans un film d'eau vitrifiée (glace amorphe) et observées à la température de l'azote liquide [340]. Aucun agent contrastant ou ombrage métallique n'est utilisé : le contraste de l'image est

uniquement lié à la densité de la matière observée. Dans le cas des molécules biologiques, ce contraste très faible rend difficile la distinction entre l'objet étudié et le bruit de fond de l'image. Mais la technique a l'avantage de donner accès à la conformation tridimensionnelle des molécules piégées dans la glace. La reconstruction 3D se fait à partir du traitement informatique des images 2D obtenues dans le microscope : différentes projections de l'objet, enregistrées sous différents angles d'inclinaison (tilt) en fonction de l'orientation imposée à l'échantillon dans l'appareil, sont accumulées et combinées. L'utilisation de la cryoTEM reste cependant particulièrement exigeante, en raison notamment de l'étape de congélation dans la glace amorphe, délicate à maîtriser techniquement, et du traitement d'images important à réaliser pour extraire les données structurales 3D [341].

3.4. Microscopie à force atomique (AFM)

La microscopie à force atomique (AFM -Atomic Force Microscopy) est une microscopie qui permet de visualiser la topographie de la surface d'un échantillon sans préparation spécifique.

Suite aux travaux précurseurs de Young qui a développé un nouvel instrument d'étude de surface (le « topographiner »), G. Binnig et H. Rohrer atteignent, en 1982, la résolution atomique en inventant le microscope à effet tunnel (STM – Scanning Tunneling Microscope). Ce microscope fut le précurseur des microscopes à sonde locale qui reposent sur la détection d'une interaction entre une pointe de dimension nanométrique et une surface. C'est ainsi qu'en 1986, G. Binnig et C.F. Quate (13) ont inventé le microscope à force atomique.

Aujourd'hui il existe deux types de microscopes AFM : les microscopes dits conventionnels et les microscopes à haute-vitesse. Les microscopes conventionnels collectent une image type sur une période de 25 à 512 s ou plus, pour les microscopes récents avec une résolution de 512x512 à 1024x1024 points, ce qui est bien plus lent que les processus macromoléculaires. A l'inverse, les microscopes à haute-vitesse font des acquisitions de quelques dizaines de ms/image, permettant d'analyser la dynamique des objets et leurs interactions [342, 343].

3.4.1 Principe du microscope AFM conventionnel

Le principe de fonctionnement du microscope à force atomique repose sur la détection des forces qui s'exercent entre une pointe attachée à l'extrémité d'un levier souple et la surface d'un échantillon qui est déplacé en vis-à-vis à l'aide d'un tube piézoélectrique. Le tube piézoélectrique, nommé scanner, autorise un balayage bidimensionnel (X:Y), et un asservissement de la position verticale Z de la pointe AFM par rapport à la surface de l'échantillon (Figure 28).

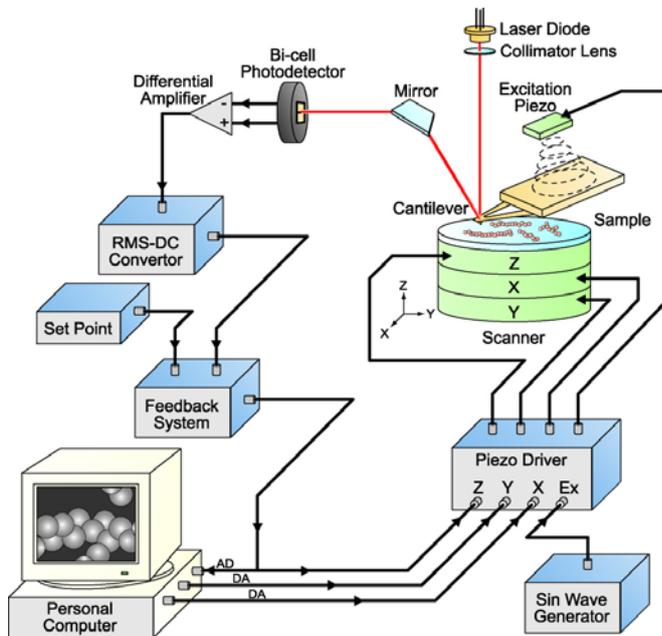


Figure 28 : Schéma de principe du microscope à force atomique en mode contact intermittent [342].

3.4.2 Modes d'utilisation de l'AFM

Il existe plusieurs modes d'utilisation de l'AFM que l'on peut regrouper en deux catégories:

- Le mode contact, statique où les principales forces d'interaction entre la pointe et la surface sont des forces répulsives de très courte portée (quelques nm au maximum).
- Le mode résonnant, dynamique où l'on discerne la modulation d'amplitude et la modulation de fréquence.

La Figure 29 illustre les forces impliquées dans les différents modes AFM.

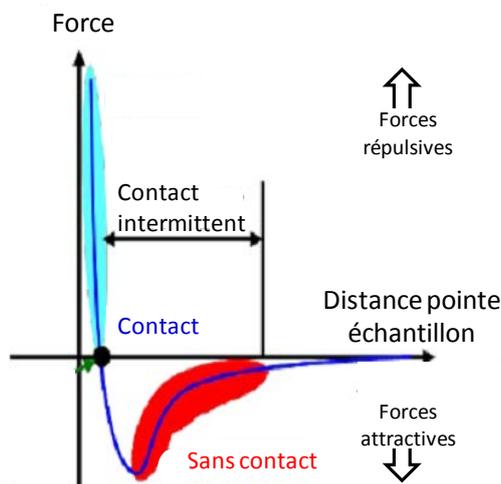


Figure 29 : Résultante schématique des forces physiques [344]

3.4.2.1 Mode contact

Dans le mode contact, les forces de répulsion, de friction et d'adhésion donnent accès aux informations de morphologie, de rugosité des échantillons. Ce mode permet d'obtenir la meilleure résolution. La résultante des forces adhésives type van der Waals (capillarité et

électrostatique) et de friction peut endommager la pointe et l'échantillon, lorsqu'on travaille sur des matériaux fragiles. Pour pallier à cet inconvénient, des modes résonants ont été développés dès 1992.

3.4.2.2 Mode résonnant

Dans le mode contact intermittent, la pointe ne vient que périodiquement en contact avec l'échantillon et les forces de friction sont ainsi évitées. C'est un mode de modulation d'amplitude qui consiste à faire vibrer le levier à sa fréquence propre de résonance (~10 kHz en solution, à 450 kHz à l'air), avec une faible amplitude (1 à 20 nm) afin que la pointe traverse la couche de contamination habituellement présente sur toute surface analysée à l'air. Ce mode est approprié à l'analyse des polymères et des échantillons biologiques. Le contrôle des interactions avec la surface de l'échantillon est suffisamment sensible pour observer les objets sans les détériorer. L'énergie échangée au contact est proportionnelle au carré de l'amplitude.

En mode contact intermittent, un mode de contraste de phase rend compte des propriétés de surface différentes avec un contraste caractéristique de l'adhésion et/ou de l'élasticité, selon la nature de l'échantillon.

Il existe également le mode sans contact qui permet de mesurer la topographie d'un échantillon avec un minimum ou aucun contact entre la pointe et la surface de l'échantillon. Les forces attractives de van der Waals entre la pointe de la sonde et l'échantillon induisent des variations de phase ou de l'amplitude de vibration du levier qui sont détectées. Cet asservissement sur la phase n'est pas correctement implémenté sur les dispositifs commerciaux.

Un nouveau mode, appelé Peak Force QNM (Quantitative Nano-mechanical property Mapping), a été développé (Figure 30). Il est basé sur la mesure d'interaction (courbe de force) à une fréquence de 2 kHz (voir 4 kHz dans la version High Speed) en tout point de balayage de l'échantillon. L'asservissement se fait sur la force minimale d'interaction. La vitesse de balayage est légèrement réduite par ce mode d'acquisition en continu de la force d'interaction. Mais ce mode présente l'avantage de décrire les propriétés d'élasticité (module de Young calculé selon le modèle DMT (Derjaguin, Muller and Toporov, [345])), d'adhésion, de dissipation.

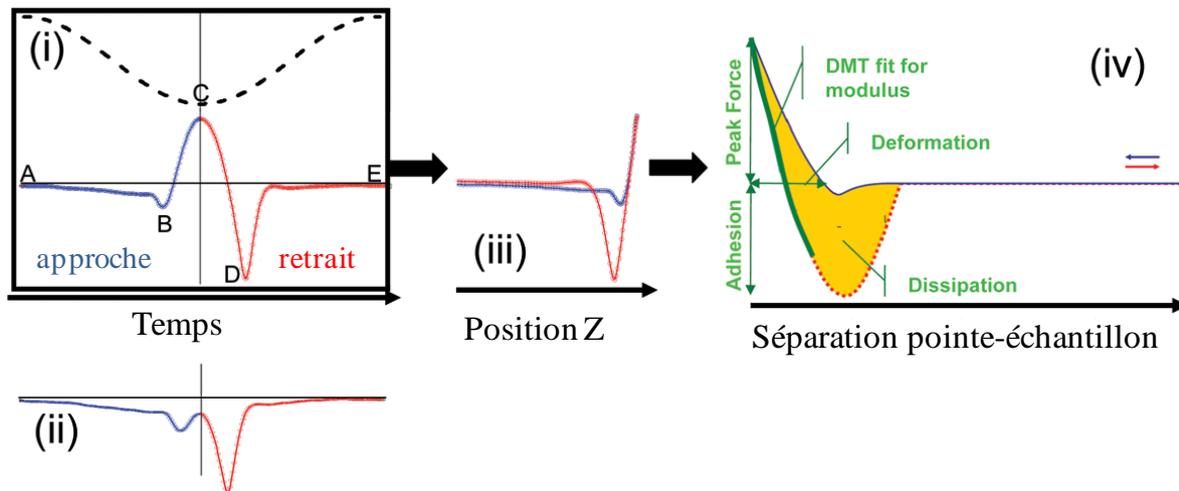


Figure 30 : Courbes de force. [346]

- (i) Graphe de la force et position en Z du piezo en fonction du temps, qui inclue (B) la mise en contact, (C) Peak Force= force maximale et (D) l'adhésion.
- (ii) Graphe de la force en fonction du temps avec une petite Peak Force.
- (iii) La courbe de force traditionnelle élimine la variable temps, en déterminant la force en fonction de la position en Z du piezo.
- (iv) Pour modéliser / ajuster les courbes, il est plus utile de déterminer la force en fonction de la séparation, avec la séparation calculée à partir de la position du piezo et la déflexion du levier.

3.4.3 Développement du microscope à haute vitesse

La motivation pour le développement de l'AFM à haute vitesse (HS-AFM) est née du besoin d'analyser plus extensivement la structure et la dynamique des objets biologiques étudiés. Différents groupes se sont mobilisés sur la question, C.F. Quate *et al.* ont développé des leviers doués d'auto-détection et d'auto-actionnement ; P. Hansma *et al.* ont tenté une augmentation de la vitesse par la réduction de la taille des leviers pour une meilleure résolution ; le groupe de M.J. Miles, à Bristol, a travaillé à l'affranchissement du système d'asservissement. Sans boucle de rétroaction, la vitesse d'acquisition par image a été diminuée à 1 ms.

Le microscope à haute-vitesse utilisé au cours du travail de thèse est un prototype développé sur le modèle de l'équipe du professeur Toshio Ando, du laboratoire de Biophysique de l'université de Kanazawa (Japon). Le bioHS-AFM se matérialise (Figure 31):

- en augmentant la bande passante de la boucle d'asservissement,
- en enregistrant les variations de tensions appliquées sur l'électrode verticale au cours de la ligne d'acquisition précédente comme pré-positionnement vertical de la prochaine ligne d'acquisition (anticipation des variations topographiques),
- en utilisant un scanner sous forme de tripode où chacun des axes X, Y et Z est découplé,
- en faisant travailler les éléments piézoélectriques X et Y sous contrainte,
- en plaçant en contre-réaction un scanner Z' de même masse au scanner Z2 support d'échantillon,

- en développant un système d'amortissement actif des vibrations du scanner,
- en développant un dispositif d'approche progressif (parachutisme),
- en développant un mode d'oscillation haute fréquence d'asservissement en amplitude ou en phase (non disponible commercialement) et un mode contact.
- en réduisant les forces d'interactions,
- en améliorant la vitesse de balayage.

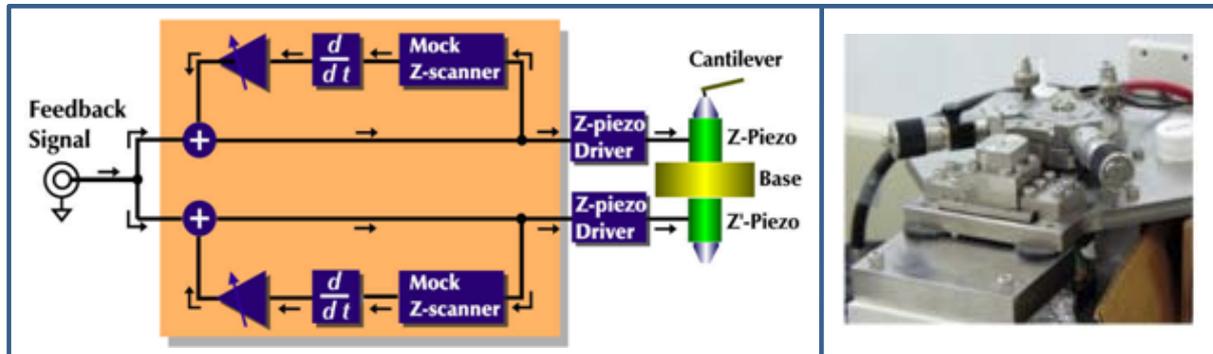


Figure 31 : Système d'asservissement du High-Speed AFM [342]

C'est ainsi qu'en 2001, le développement d'une première génération de microscope à haute vitesse a été initiée [343]. Partant du principe du microscope conventionnel, les dispositifs suivants ont été développés :

- (i) un scanner à haute-vitesse se déplaçant rapidement sous contrainte, contrebalancé verticalement en Z,
- (ii) des leviers miniaturisés de longueur 40-10 μm , de largeur 1 μm avec des pointes de longueur 2 μm obtenues par croissance sous faisceau d'électrons (Figure 32),
- (iii) un système optique de détection compatible avec la taille des leviers (spot laser réduit à 1 μm de diamètre),
- (iv) un convertisseur rapide amplitude-DC (44 kHz) qui donne le signal de RMS (Root Mean Square) [342].

Le microscope HS-AFM ainsi conçu a permis d'imager pour la première fois le mouvement processif de molécules de myosine V sur un filament d'actine, dont la vitesse est d'environ 1 $\mu\text{m/s}$ [347].

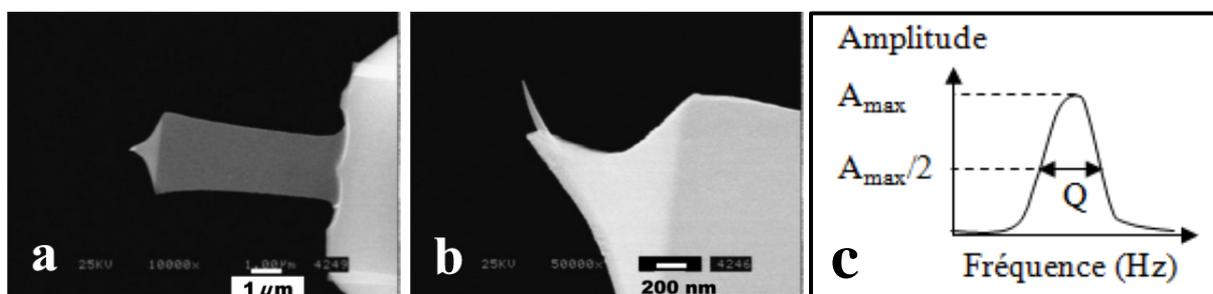


Figure 32 : Images SEM d'un petit levier avec une pointe en forme de bec d'oiseau.

A, vue du dessus, b, image grossie de la pointe. L'aiguille qui dépasse est une pointe déposée par faisceau d'électrons [342]. c, schéma de présentation du facteur de qualité Q.

Au bout de 10 ans, une deuxième génération d'appareils a été développée pour diminuer les interactions pointes-échantillons résultant d'une force de contact trop élevée et une bande passante insuffisante. Ainsi, des leviers plus petits (longueur 10^{-6} μm) possédant une fréquence de résonance supérieure (1 MHz) aux précédents ont été fabriqués.

La bande passante de la boucle de contrôle dynamique de l'oscillation du levier [348] a été augmentée. Il a été aussi développé une détection synchrone ultra rapide (20 kHz à 10 MHz). Les micro-leviers présentent des forces de résonance de l'ordre de 500 kHz – 1.2 MHz et un facteur de qualité de 2.5, en solution, qui correspond à la largeur à mi-hauteur du spectre de résonance (Figure 32c). D'autres améliorations furent apportées par la simplification du scanner à haute-vitesse, limitant le nombre de pics de résonance et augmentant sa rigidité [349]. Ces derniers développements ont permis d'imager de manière peu invasive des échantillons sans détériorer la vitesse d'imagerie.

3.4.4 Quels substrats pour déposer les échantillons ?

Tout support peut *a priori* être utilisé pour déposer les échantillons, toutefois la résolution nanométrique de la technique AFM obtenue par les interactions pointe - surface imposent de fortes contraintes de planéité et de faible rugosité du substrat [341]. De fait, la surface de prédilection de dépôt des molécules est en général le mica muscovite, qui a une structure cristallographique en feuillets de silicates reliés par des couches d'ions potassium. Après clivage avec un papier adhésif, il est facile d'obtenir une surface propre et plane à l'échelle atomique bien adaptée à l'imagerie de molécules individuelles comme l'ADN et les protéines. La séparation au niveau du plan des ions potassium (K^+) inter-foliaires implique que la charge de la surface de mica clivée est globalement négative (Figure 33) [341].

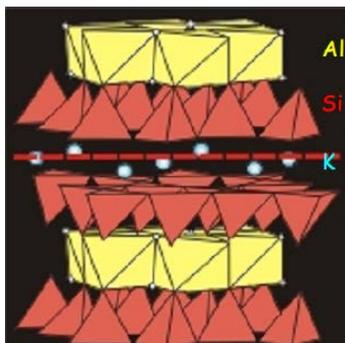


Figure 33 : Structure en feuillets du mica muscovite.
Formule : $\text{KA}_2[\text{AlSi}_3]\text{O}_{10}(\text{OH})_2$ [341]

Les ions divalents couramment utilisés pour fixer l'ADN sur le mica clivé sont le chlorure de magnésium et le chlorure de nickel. Toutefois, d'autres cations, y compris de valence plus élevée, peuvent être employés. La force d'absorption de l'ADN et des protéines sur le mica dépend principalement de la charge de la surface et des concentrations en ions dans la solution de dépôt, mais également de la taille des molécules. Ainsi dans une population de molécules dispersées, les petites molécules ayant tendance à être décrochées de la surface, soit par lavage, soit par immersion, selon les conditions d'observation à l'air ambiant

ou en liquide, ne seront pas imagées. Le dépôt des échantillons sur le substrat peut donc constituer un biais dans l'analyse de la distribution d'une population de molécules de tailles hétérogènes.

Le traitement du mica par des cations après clivage ne constitue pas forcément une méthode d'accroche satisfaisante des échantillons. Dans ce cas, il peut être nécessaire de développer une chimie de surface sur le mica telle que la silanisation [350] afin d'introduire de nouveaux groupements de fixation tels que les groupements d'acide nitrilotriacétique liés au nickel. Toutefois, l'état de surface et la rugosité après fonctionnalisation peuvent être altérés en comparaison du support nu. Aussi, un compromis doit être fait entre surface plane ne présentant pas de spécificité d'ancrage et surface fonctionnalisée qui constitue un processus délicat pouvant aboutir à des surfaces irrégulières avec des défauts localisés ou des domaines disparates.

Le substrat peut également poser un problème lors de l'analyse de la dynamique des objets en milieu liquide car les objets doivent être suffisamment ancrés sur la surface pour être imagés sans être totalement immobilisés. Ceci pose le même problème de chimie de surface présenté ci-dessus avec un compromis à trouver entre une forte immobilisation des objets qui permet d'étudier leur structure et une plus faible immobilisation qui permet d'accéder à leur dynamique.

D'autres supports tels que l'or, le graphite, l'oxyde de titane sous forme anatase, le silicium, la silice peuvent être utilisés, sachant que le point critique réside dans l'état de surface du support, sa rugosité. En effet, la morphologie du substrat nu doit être éloignée de la dimension des objets biologiques déposés même si le nouveau mode Peak Force QNM permet de discerner des objets par leur module d'élasticité. Le dépôt des échantillons peut nécessiter le développement de chimie de surface pour fixer les échantillons avant de les imager.

3.4.5 Avantages et inconvénients

L'analyse AFM présente les avantages d'être statique ou dynamique selon que les échantillons sont totalement immobilisés et observés à sec ou ancrés tout en restant mobiles en milieu liquide. Dans ce dernier cas, le développement du microscope AFM haute-vitesse qui présente des résolutions temporelles de l'ordre de dizaines de ms, comparées à quelques secondes (voire quelques minutes) pour un microscope AFM classique, permet le suivi en temps réel des interactions et ce, sans marquage des échantillons.

Le minimum de préparation de l'échantillon, la faible concentration et le volume réduit du dépôt (1 μ l) contribuent à son utilisation dans le cas d'échantillons fragiles et en faible quantité.

Toutefois, la technique elle-même peut introduire un biais du fait de l'observation de l'échantillon sur des surfaces, qui ne reflète pas forcément la réalité de la dynamique des structures en solution, du fait des perturbations liées à la surface et à sa fonctionnalisation.

De plus, il semble nécessaire d'associer une cartographie chimique des objets observés. Des évolutions récentes tendent à coupler la microscopie à force atomique à la microscopie de fluorescence ou à la spectroscopie Raman sachant que la résolution spatiale reste limitée par le critère de Rayleigh.

3.5. Microscopie optique Confocale

3.5.1 Principe du microscope confocal

L'invention du microscope confocal est attribuée à M. Minsky qui présenta un prototype en 1955, et déposa un brevet en 1961 [351]. Ce n'est toutefois que dans la fin des années 1980 que des modèles commerciaux sont apparus, rendant cette technique accessible à de nombreux laboratoires.

Un microscope confocal est un microscope optique qui a la propriété de réaliser des images de très faible profondeur de champ (environ 400 nm) appelées « sections optiques ». En positionnant le plan focal de l'objectif à différents niveaux de profondeur dans l'échantillon, il est possible de réaliser des séries d'images à partir desquelles on peut obtenir une représentation tridimensionnelle de l'objet. Le microscope confocal fonctionne en lumière réfléchie ou en fluorescence. La plupart du temps, on utilise un laser comme source de lumière. Le rayon laser, concentré par une lentille, balaie la surface, tandis qu'un trou d'aiguille (pinhole) placé devant le détecteur et centré au niveau arrière de l'objectif rejette la lumière parasite provenant des points situés hors du plan focal. Le nom « confocal » provient de la mise en correspondance de trois points : la source lumineuse ponctuelle, le point illuminé au foyer avant de l'objectif et le sténopé situé au foyer arrière. Le principe de la microscopie confocale est résumé sur la Figure 34.

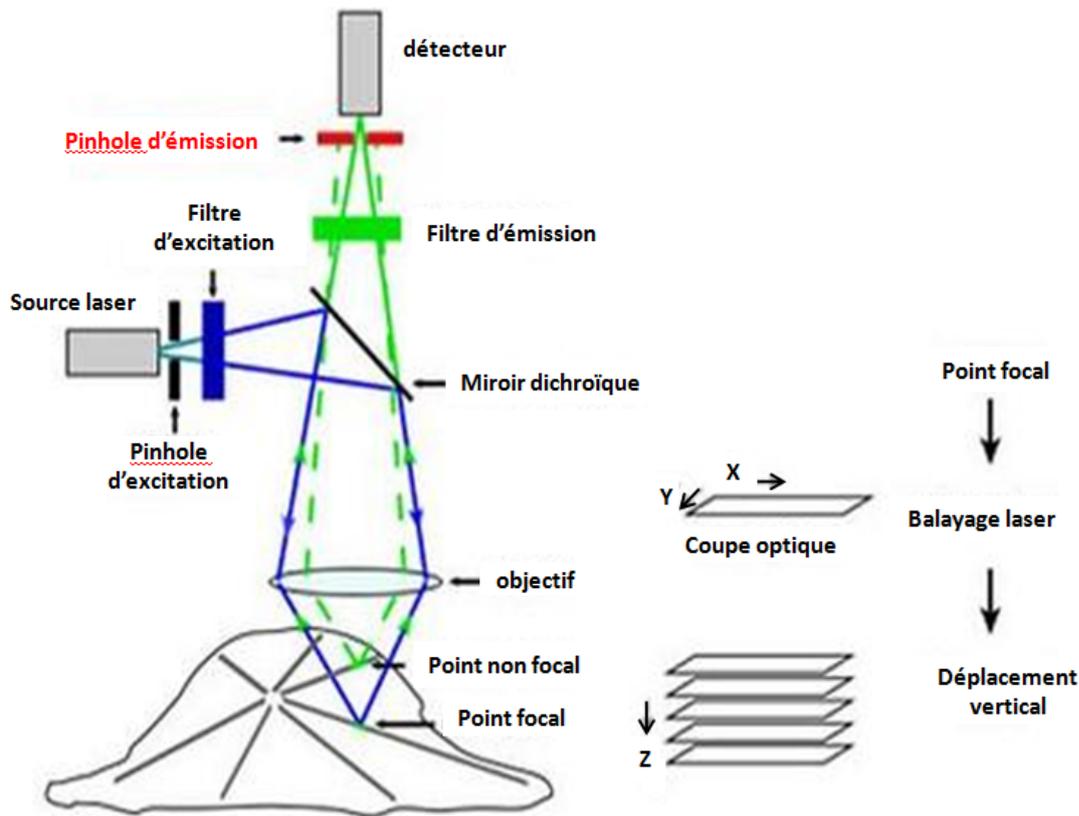


Figure 34 : Schéma de principe du microscope confocal à balayage laser.

La lumière provenant de l'échantillon, et des plans situés au-dessus et au-dessous du plan focal est éliminée par le diaphragme (pinhole) de sortie. Faculté de Jussieu (<http://www.futura-sciences.com>)

3.5.2 Avantages et inconvénients

La microscopie confocale permet d'obtenir des images sans flou parasite du fait de l'élimination du signal fluorescent provenant d'autres plans. Sa résolution en hauteur améliore le rapport signal/bruit. Elle donne accès à la reconstruction 3D rapide à partir de l'acquisition de séries de sections optiques.

La microscopie confocale a l'inconvénient de présenter une résolution temporelle limitée par la vitesse de balayage. Par ailleurs, les réflexions du signal sur le support limite la sensibilité pour des applications molécule unique. Elle nécessite une étape de marquage par des fluorophores ou des boîtes quantiques qui peuvent affecter les phénomènes étudiés.

3.5.3 Fluorophores

On distingue l'utilisation des fluorophores organiques tels que la fluorescéine des boîtes quantiques [352].

- Les fluorophores organiques

Les fluorophores organiques, de petite taille, permettent un marquage efficace avec une perturbation généralement limitée de l'activité des biomolécules marquées mais ils présentent une diminution rapide du signal de fluorescence (photo-dégradation) qui limite le temps d'observation des biomolécules individuelles de quelques secondes pour les

fluorophores traditionnels jusqu'à l'ordre de la minute pour les systèmes dendrimériques. C'est pourquoi les scientifiques ont cherché de nouveaux fluorophores plus photo-stables.

- Les nano-cristaux semi-conducteurs (QD)

Les boîtes quantiques sont des cristaux semi-conducteurs fluorescents de forme souvent sphérique dont la taille varie de 2 à 10 nm. Ils présentent des propriétés luminescentes différentes selon leur dimension par effet de confinement quantique. Plus le confinement est important, c'est-à-dire plus la taille du QD est petite, plus l'émission tend vers les ultraviolets (UV). Il est possible de synthétiser des QDs émettant de 400 à 2000 nm (Figure 35).

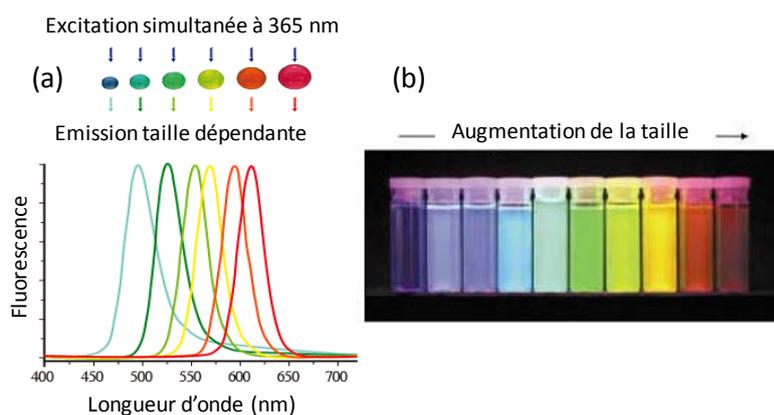


Figure 35 : Caractéristiques des boîtes quantiques.

(a) graphique représentant la variation de couleur d'un nano-cristal en fonction de la longueur d'onde émise [353]. (b) 10 solutions de nanoparticules de CdSe/ZnS de couleur d'émission distinctes, excitées sous une lampe éclairant dans le proche U.V. De gauche à droite, les maxima d'émission sont situés à 443, 473, 481, 500, 518, 543, 565, 587, 610, et 655 nm [354].

Les QDs présentent un large spectre d'absorption de l'UV jusqu'à l'infrarouge, un spectre d'émission symétrique et étroit (de 20-70 nm de largeur à mi-hauteur) (Figure 35), contrairement aux fluorophores conventionnels qui ont un spectre d'absorption étroit et un spectre d'émission large.

Les avantages des QDs par rapport aux fluorophores organiques sont leur photostabilité ($\sim 10^2$ à 10^4 fois plus élevée) et leur flux de photons admissibles de l'ordre de 20 fois plus grand malgré un rendement quantique inférieur ou similaire, et ce du fait de leur très fort coefficient d'absorption et de leur photostabilité qui permet de les irradier à des niveaux plus élevés [355]. Néanmoins, le scintillement de la fluorescence des nano-cristaux, pouvant nuire au suivi des molécules dans le temps, représente le principal obstacle dans l'utilisation très large de ces nanostructures, même si diverses stratégies peuvent aujourd'hui être utilisées pour le réduire [356]. Par ailleurs, les QDs doivent être fonctionnalisés avant couplage aux molécules d'intérêt, ce qui implique la maîtrise des procédés chimiques.

3.6. Résonance plasmonique de surface (SPR)

Historiquement, Wood découvre le phénomène optique de résonance de plasmons de surface en 1902. Il remarque de fines bandes sombres dans le spectre diffracté d'une source continue de lumière blanche. L'interprétation théorique de cette anomalie est donnée en 1941 par Fano qui l'associe aux ondes de surface, les plasmons de surface, supportées par le réseau. En 1968, A. Otto puis E. Kretschmann et H. Raether développent deux configurations qui permettent d'exciter les plasmons de surface en utilisant la réflexion totale atténuée. Suite à ces travaux, l'intérêt pour les plasmons de surface a considérablement augmenté, en particulier pour caractériser les films minces. Pour la première fois en 1983, C. Nylander et B. Liedberg, utilisent la SPR (« Surface Plasmon Resonance ») pour la détection des gaz et des biomolécules. Cette technologie a été commercialisée au début des années 90 par la société Biacore International, première entreprise spécialisée dans ce domaine.

3.6.1 Principe

La résonance plasmonique de surface est une technique de mesure en temps réel et sans marquage des interactions moléculaires par le suivi de la variation de l'indice de réfraction. La SPR met en œuvre un phénomène de résonance entre les électrons libres d'un métal et une onde (Figure 36). Ceci induit une réduction de réflectivité du fait du transfert d'énergie entre le faisceau incident et l'onde de surface qui se propage à l'interface. Le principe de mesure par SPR est basé sur la variation de la courbe de résonance. Le signal SPR exprimé en réflectivité est une fonction de l'indice de réfraction local liée à la quantité en masse et à la nature de l'échantillon sur la surface. Les interactions à la surface peuvent être mesurées soit par la variation de l'angle de résonance à une longueur d'onde fixe, soit par la variation de minimum de longueur d'onde de résonance à angle d'incidence fixe, soit par la dépendance de la réflectivité à angle et longueur d'onde fixés.

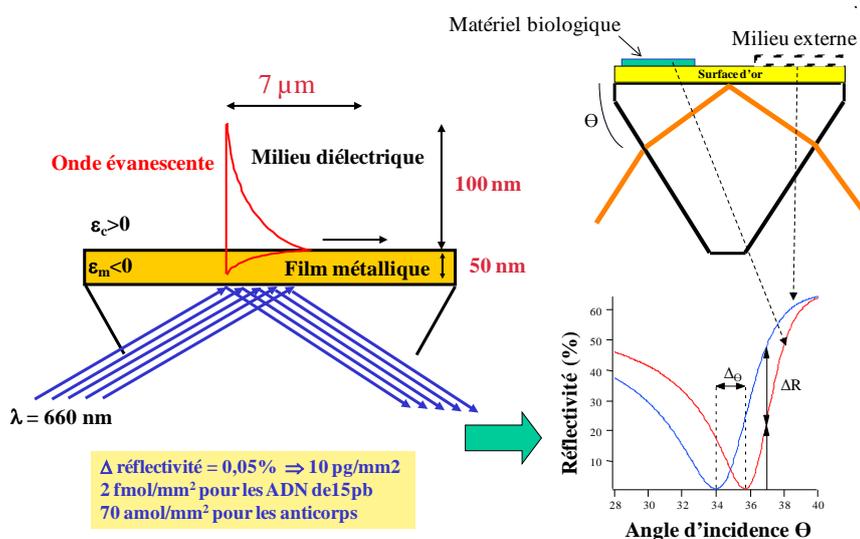


Figure 36 : Principe de la résonance plasmonique de surface.

Imagerie SPR

Le principe de mesure par imagerie SPR est basé sur le suivi de la variation de la réflectivité à un angle d'incidence fixe. A chaque expérience, l'angle d'incidence est positionné à une valeur où la pente de la courbe de réflectivité est maximale. Une caméra CCD analyse toute la surface du biocapteur et enregistre l'évolution de la réflectivité en fonction du temps à travers une succession d'images. La réponse du système SPR est donnée en pourcentage de réflectivité qui est le rapport entre la réflectivité en TM (transverse magnétique) et en TE (transverse électrique) des points de l'image. La mesure SPR nécessitant la fixation des objets sur la surface, différentes chimie de surface ont été développées (décrites ultérieurement).

3.6.2 Avantages et inconvénients de la SPR

La SPR, en tant que technique d'analyse en temps réel, donne accès à des mesures dynamiques et spatiales des molécules. Elle permet également de mesurer l'anisotropie moyenne, c'est-à-dire l'orientation de la population d'objets.

La SPR dans sa configuration classique démontre une sensibilité de détection moindre que les techniques de fluorescence, toutefois plusieurs publications présentent le développement de microscopes SPR capables de détecter des nanoparticules isolées de quelques dizaines de nanomètres. On peut citer le microscope par plasmons de surface localisés capables de différencier des nanoparticules ayant des diamètres de 10 à 200 nm, le montage imageur spectral qui permet d'analyser un objet de 30 nm de côté [357], ou encore de microscope à plasmons de surface à deux canaux et à polarisation radiale qui atteint des résolutions nanométriques [358, 359]. Dans tous les cas, il est nécessaire de développer des chimies de surface adéquates qui permettent la fixation de molécules sur les surfaces. Celles-ci font l'objet de nombreuses publications qui ciblent la mise au point de chimies de surface permettant : (i) le contrôle de l'orientation des molécules et une stabilité dans le temps des interactions support-molécule, (ii) un faible bruit de fond et une bonne résistance aux interactions non spécifiques, et enfin (iii) la minimisation de la perturbation en 3D des molécules favorisant une bonne capacité de couplage [360-364].

On peut décrire trois classes de techniques de fixation des molécules sur la surface qui vont aboutir soit à des monocouches uniformes sur l'or pour les méthodes de fixation directe et d'électro-polymérisation, soit à des multicouches assemblées dans la méthode de fixation indirecte :

- La fixation directe permet d'orienter les molécules présentant une extrémité thiol sous forme de monocouche sur le support, du fait de la forte affinité des groupements -SH avec l'or. Les interactions non spécifiques, le problème d'orientation et d'encombrement stérique sont limités par l'exposition du substrat à des thiols aliphatiques, ce qui permet d'optimiser le rendement des réactions biologiques [365].

- La fixation par polymérisation électrique de monomères capables de donner un polymère conducteur auquel seront couplées les molécules biologiques permet également de réaliser une couche uniforme à la surface du polymère. On peut citer le cas du polypyrrole dont l'épaisseur peut être contrôlée par des variations d'application du potentiel lors de l'électro-polymérisation. Le couplage des molécules biologiques aux monomères de pyrrole nécessitant de travailler avec une quantité minimale de matériel, cette technique ne peut être utilisée dans le cas d'échantillons précieux disponibles en quantité limitée, telles que des protéines ou structures synthétiques construites in vitro.
- La fixation indirecte repose sur un assemblage de plusieurs couches, le dernier étage présentant les groupements fonctionnels qui permettent la fixation des molécules soit par affinité (système avidine-biotine), soit par liaisons covalentes (maleimide-thiol, esterNHS-amine, thiocyanates- amine), soit par liaisons électrostatiques[365]. Selon la nature de la dernière couche de fixation, les molécules pourront être, soit ancrées et organisées à sa surface soit organisées en trois dimensions dans le volume de la matrice. Un polymère en 3D type dextran, présente une densité de fonctionnalisation supérieure à une monocouche qui peut toutefois induire une hétérogénéité d'accessibilité. Cette chimie est largement répandue, même si la porosité du polymère épais implique des interactions plus ou moins loin de la surface avec des cibles plus ou moins accessibles, car elle présente l'avantage de permettre la régénération des puces [360].

Une méthode alternative aux chimies sur or consiste à déposer un film fin de silice sur la surface d'or et d'ancrer les molécules par les techniques de fixation des silanes [366, 367].

D'autres couches métalliques (Ag, Cu) ou oxydes (TiO₂) peuvent être utilisées en plasmonique, sachant que le cuivre permettra l'interfaçage avec l'électronique.

En conclusion, l'auto-assemblage est une notion clé dans la conception de nanostructures hybrides à base d'acides nucléiques et de protéines. Diverses stratégies d'auto-assemblage peuvent être considérées, à l'équilibre thermodynamique ou loin de cet équilibre, notamment lorsque des éléments fragiles telles que les protéines sont impliquées. Les assemblages peuvent être parallèles ou séquentiels, selon que tous les composants sont mélangés ou ajoutés hiérarchiquement. Dans tous les cas, la connaissance des paramètres physico-chimiques critiques pour l'assemblage des acides nucléiques constitue une étape nécessaire à la réalisation des nanostructures. Finalement, c'est l'utilisation de techniques, aussi bien à l'échelle de la molécule unique qu'à l'échelle globale de la population, qui permet de caractériser les nanostructures développées et d'étudier leur dynamique par exemple au travers de la microscopie de force atomique à haute vitesse.

1^{ère} partie

MODELISATION DES ASSEMBLAGES MONODIMENSIONNELS IMPLIQUANT DES BRIQUES D'ADN.

Notre objectif est de réaliser des assemblages fonctionnels à partir de structures synthétiques hybrides constituées de protéines et d'acides nucléiques. Cela nécessite de comprendre les propriétés des PDNA dont la partie nucléique confère les propriétés d'auto-organisation et de les comparer aux ADN seuls correspondants. Pour ce faire, nous avons choisi un modèle d'assemblage linéaire et développé les outils nécessaires à la caractérisation du comportement d'assemblage des ADN ainsi que ceux permettant de comparer les données expérimentales et simulées.

Deux aspects doivent être pris en compte : (i) la réalisation des auto-assemblages dans des conditions douces compatibles avec la présence de protéines, (ii) le contrôle des assemblages avec une description globale des populations.

Dans la littérature, les systèmes d'assemblages basés sur les briques d'ADN sont pilotés via des interactions locales qui régissent les règles de construction, toutefois le contrôle de la distribution à longue distance est mal ou pas défini. Ce problème général assez peu abordé nécessite le développement d'outils.

Dans l'approche classique d'assemblage des briques d'ADN où des températures proches du point de fusion des oligonucléotides sont utilisées afin d'éviter la formation de structures illicites [175, 368, 369], de nombreux modèles ont été développés pour minimiser les taux d'erreur [370, 371]. Ces modèles traitent les erreurs de croissance lors des assemblages [173, 174], ou bien les erreurs de nucléation en limitant les réactions parasites par l'introduction de briques de protection [172] ou encore en réalisant des assemblages séquentiels [372]. Ces modèles qui constituent une approche fine et locale des mécanismes contrôlant l'addition ou la discrimination des briques individuelles dans l'assemblage, n'abordent pas le contrôle à longue distance des assemblages.

Dans l'approche cinétique, différents mécanismes peuvent être mis en œuvre tels que des processus auto-catalytiques ou l'utilisation de barrière cinétique à la nucléation impliquant le phénomène d'invasion de brins [298, 368, 373]. Dans une population contenant un nombre élevé des différents états intermédiaires, des outils computationnels sont requis ainsi que des méthodes de détermination des paramètres adaptés à l'explosion combinatoire du nombre des réactions possibles. Il est nécessaire de gérer de manière analytique des populations pas forcément définies et d'accéder aux paramètres qui contrôlent les réactions d'assemblage, tels que les constantes des vitesses d'hybridation individuelles. Pour ce faire, des outils de modélisation doivent être développés.

Lors des expériences d'assemblages monodimensionnels de briques d'ADN, nous avons été confrontés au problème du contrôle des assemblages avec l'accumulation apparente de structures intermédiaires, d'où la nécessité de modéliser les mécanismes d'assemblage afin de mieux interpréter nos données. Le besoin d'outils adaptés à notre démarche globale nous a amené à développer un modèle linéaire afin de gérer une combinatoire 1D plus facile à intégrer qu'une explosion combinatoire de systèmes branchés. Ce modèle se devait suffisamment simple pour pouvoir interpréter les expériences.

Deux types d'assemblages peuvent être considérés : ceux conduisant à un produit défini et unique et ceux conduisant à des collections finies ou infinies de produits possibles. Dans le premier cas où l'assemblage est convergent, le nombre d'espèces intermédiaires est défini et la combinatoire d'assemblage limitée. En biologie moléculaire, ce cas se rencontre en particulier dans la construction de gènes ou de bibliothèques de gènes synthétiques. Dans le domaine des nanotechnologies, un exemple correspond à l'assemblage de structures de forme déterminée telle que les nanostructures fabriquées par la technique des origamis. Le second cas correspond à des systèmes ouverts qui correspondent à une distribution illimitée de produits possibles, et par conséquent à une combinatoire infinie. C'est le cas par exemple de la formation de polymères en chimie organique ou de celle des assemblages récurrents, dendritiques ou fractals conduisant à des structures définies à courte échelle mais non définies à longue distance. Cette situation, malgré sa complexité apparente, nous est apparue plus facile à utiliser pour caractériser les règles générales d'évaluation des paramètres cinétiques pour les intermédiaires de réaction. En effet, un nombre limité de briques élémentaires étant impliquée, la détermination expérimentale des paramètres primaires est facilitée.

1. Présentation de l'article 1.

“Design and experimental validation of a generic model for combinatorial assembly of DNA tiles into 1D-structures.”

La connaissance fine des mécanismes d'auto-assemblage de briques d'ADN menant à des structures 1-D est critique pour la maîtrise de l'ingénierie de nanostructures, mais elle constitue également un enjeu pour l'optimisation de la synthèse de gènes ou de la création de banques combinatoires. De nombreuses stratégies de synthèse de gène s'appuyant sur des approches de biologie moléculaire ou computationnelles ont été décrites [374-378] ; elles visent à limiter les erreurs d'assemblage des séquences et l'incorporation d'oligonucléotides défectueux (du fait de défauts de synthèse) au sein des constructions de gène, et ce dans les meilleurs délais. Si l'on considère le cas de la construction de banques combinatoires, il est important d'aller au-delà des approches actuelles reposant largement sur des méthodes empiriques d'essai-erreur par le développement de modèles prédictifs [379, 380].

Malgré la disponibilité de nombreux outils permettant une description à une échelle locale de ces systèmes, des outils de modélisation globale permettant d'appréhender dans leur ensemble les populations de molécules en cours d'assemblage sont encore peu développés. C'est pourquoi nous nous sommes intéressés à concevoir un algorithme visant particulièrement la modélisation quantitative de ces mécanismes. Notre approche s'applique aussi bien aux problèmes posés par la génération de nanostructures par auto-assemblage, qu'à la synthèse de gènes ou de bibliothèques en biologie moléculaire. Ces développements sont particulièrement importants dans le cas de la formation de structures hybrides faisant intervenir des molécules fragiles telles que les protéines, qui pourraient être dégradées par les conditions de haute température utilisées couramment pour l'assemblage des acides nucléiques en nanostructures.

Dans cet article, nous avons réussi à modéliser quantitativement l'auto-assemblage sous contrôle cinétique de briques d'ADN en des structures monodimensionnelles. L'algorithme et les méthodes de détermination des paramètres développés s'appliquent aussi bien à des assemblages qui mènent à des produits finaux définis qu'à une distribution de copolymères. La combinaison des approches de simulation et expérimentales a permis de caractériser les facteurs contrôlant la structure et la dynamique des populations de molécules résultantes.

Le processus combinatoire décrivant l'assemblage des briques a été implémenté dans un algorithme générique qui permet une description quantitative des espèces significatives de la population s'accumulant au cours des réactions de polymérisation. L'assemblage et la caractérisation par les approches optiques et d'électrophorèse des blocs copolymères issus de l'auto-assemblage ont été utilisées pour définir et valider des règles génériques permettant de déduire les paramètres du modèle.

Les facteurs contrôlant la structure et la dynamique des populations d'ADN oligomères ont été mis en évidence pour des assemblages menant ou non à des produits finaux déterminés. Les paramètres primaires ont été déterminés expérimentalement par mesure des effets hypochromiques en cinétique rapide. L'ajustement des simulations aux profils expérimentaux a permis de définir les règles de calcul des paramètres secondaires, qui prennent en compte les contextes macro- et micro-environnementaux des étapes d'hybridation individuelles. Dans le cas des copolymères, des simulations précises du profil expérimental ont été réalisées pour la formation d'assemblages linéaires ou circulaires.

Ensemble, la longueur des espèces et la structure des régions d'ADN qui encadrent les sites d'hybridation constituent des paramètres critiques pour lesquels les règles de calcul ont été définies. L'approche computationnelle s'est montrée efficace pour prédire quantitativement les paramètres affectant l'évolution en fonction du temps et la distribution des produits d'accumulation pour une gamme importante de modèles expérimentaux.

La procédure computationnelle développée (décrite dans l'article) et les stratégies d'évaluation des paramètres s'appliquent à une large gamme d'assemblages à base de briques d'ADN mais de manière plus globale à tout type d'assemblage résultant en des structures monodimensionnelles et faisant intervenir des briques moléculaires qui présentent une reconnaissance spécifique.

2. Appendice 1 : Article 1

“Design and experimental validation of a generic model for combinatorial assembly of DNA tiles into 1D-structures.”

Laisné A., Lesniewska E. & Pompon D.

*Biochimica et Biophysica Acta (BBA) - General Subjects, (2011),
1810(6): 603-611*



Design and experimental validation of a generic model for combinatorial assembly of DNA tiles into 1D-structures

Aude Laisne^a, Eric Lesniewska^b, Denis Pompon^{a,*}

^a Centre de Génétique Moléculaire, CNRS, Avenue de la Terrasse, F91190 Gif-sur-Yvette, France

^b Institut Carnot Bourgogne, UMR CNRS 5209, Université de Bourgogne, F-21078 Dijon, France

ARTICLE INFO

Article history:

Received 4 December 2010

Received in revised form 1 March 2011

Accepted 2 March 2011

Available online 22 March 2011

Keywords:

Combinatorial assembly
Deoxyribonucleic acid tile
System biology
Modeling
Kinetic control
Gene synthesis

ABSTRACT

Background: Quantitative modeling of the self-assembly of DNA tiles leading either to defined end-products or distribution of biopolymers is of practical importance for biotechnology and synthetic biology.

Methods: The combinatorial process describing tile assembly was implemented into a generic algorithm allowing quantitative description of the population of significant species accumulating during the reaction course. Experimental formation and characterization by optical and electrophoresis approaches of copolymers resulting from the self-assembly of a limited number of half-complementary tiles were used to define and validate generic rules allowing definition of model parameters.

Results: Factors controlling the structure and the dynamic of the oligomer population were evidenced for assemblies leading or not to defined end-products. Primary parameters were experimentally determined using rapid mixing experiments. Adjustment of simulations to experimental profiles allowed definition of generic rules for calculation of secondary parameters that take into account macro- and microenvironment of individual hybridization steps. In the case of copolymers, accurate simulation of experimental profiles was achieved for formation of linear assemblies.

Conclusions: Overall length of species and structure of the DNA regions flanking the hybridization sites are critical parameters for which calculation rules were defined. The computational approach quantitatively predicted the parameters affecting time-course and distribution of accumulating products for different experimental designs.

General significance: The computational and parameter evaluation procedures designed for the assembly of DNA tiles into large 1D-structures are more generally applicable for the construction of non-DNA polymers by extremities-specific recognition of molecular blocks.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Linear assembly of DNA tiles is mainly involved in the domain of gene synthesis [1,2] and combinatorial library constructions [3]. A plethora of synthesis options that relies on basic molecular biology techniques and that aims to guarantee successful synthesis in cost-effective time and conditions have been described [4–6]. However, no modeling approach was proposed to describe the combinatorial structure of mono-dimensional DNA assemblies at the level of the time dependence of the distribution of all possible sub-assemblies. Controlled tuning of the process is particularly critical to optimize directed evolution based protein engineering. Substitution of the

current trial-and-error approaches by modeling would thus constitute a significant advantage [7]. Moreover, new emerging technologies, like the construction of networks of interacting peptidic domains [8,9] using self-assembly of synthetic nucleoprotein structures, are also demanding for modeling of linear assemblies [10,11]. When single step assembly of a preformed DNA–protein covalent structure is targeted, reversible conditions requiring high temperature must be avoided, owing to the limited thermal stabilities of proteins. Ligation or polymerase based approaches remain possible but are also problematic in these geometrically crowded systems. For these reasons, assembly processes kinetically controlled at low temperature appear interesting but require quantitative modeling to limit kinetic traps. Two experimental settings can be considered: one, typical for gene or library synthesis which involves a finite combinatorial allowing full description of all intermediate species, and the other, corresponds to unlimited block-copolymers for which an open combinatorial structure has to be managed. The two cases can be modeled but formation of repetitive copolymers involving a limited number of

Abbreviations: ss, single-stranded; ds, double-stranded

* Corresponding author at: CGM, CNRS, bât 23B, Avenue de la Terrasse, 91190 Gif-sur-Yvette, France. Tel.: +33 169823680; fax: +33 169823682.

E-mail address: pompon@cgm.cnrs-gif.fr (D. Pompon).

tiles constitutes a simpler model when investigation of general rules for parameter determination is targeted. However, such model raises the problem of the combinatorial explosion of possible reactions during the formation of unlimited copolymers. Evaluation of individual rate constants represented a challenge that was approached by the determination of a core of experimental rate constants and the definition of rules for the evaluation of the whole model parameters from the seeding set. The influence of parameters such as the length dependence of ss-DNA and the structure of flanking sequences on the hybridization rates were taken into account. Indeed, the length dependence of hybridization rates between ss-DNA was investigated [12–15] but few were reported in the case of species of variable lengths assembled through hybridization of ss-DNA segments of constant length. The present paper addresses the problem of modeling end-recognition driven self-assembly of tiles involving very large or even unlimited number of intermediate species. It specifically develops ways allowing determining critical model parameters for DNA assemblies using combination of experimental approaches and extrapolation rules. The core part of the modeling is however widely independent of the chemical nature of the building blocks and can equally apply to a large range of biopolymers. This approach aims to fill a gap between current modeling approaches either targeting to describe at a local level the interaction between tiles [16,17] or, in contrast, focusing on the global organization [18] of resulting supra-molecular assemblies.

2. Materials and methods

2.1. DNA-tile hybridization and gel analysis

Oligonucleotides were purified as described in the supplemental protocol 1 in Appendix A. Approximately 3.5 μl of the 4 polymerizing DNA-tiles ($\sim 1 \mu\text{M}$ final) in phosphate buffered saline (PBS) were separately deposited on the wall of a tube. The reaction was initiated by centrifugation and vortexing. Incubation was performed at 24 °C or 42 °C for variable times. For gel analysis, reactions were quenched by addition of a premix of the four terminating DNA-tiles ($\sim 1.5 \mu\text{M}$ final) and the reaction continued for 1 h before deposit on 3.5% Nusieve GTG agarose gel that was run in 0.5 \times TBE at 50 V and 4 °C for 2 h. Images of ethidium bromide stained gels were submitted to trapezoidal deformation to correct heterogeneity of DNA migrations. Data from different exposure times were concatenated to increase dynamic range of signal without saturation.

2.2. DNA hybridization kinetic analysis by rapid-mixing experiments

Half-complementary ss-DNA-tiles were loaded at the same concentration ($\sim 0.8 \mu\text{M}$ in PBS buffer) in the two stopped-flow syringes. After rapid mixing at 24 °C, 400–1600 spectra (40 to 1240 s) were recorded and hypochromic changes at 260 nm were fitted to exponential or hyperbolic laws to estimate initial rates and amplitudes of reactions. Second order rate constants were calculated from the ratio between the initial slope and the extrapolated total amplitude of the traces. For self-assembly of 4 ss-DNA tiles to form polymers, an equimolar mix of tiles (Fig. 3, model 2) was loaded into one syringe in PBS buffer, set to pH 11.8 with 20 mM NaOH. The second syringe was loaded with PBS supplemented by 20 mM HCl in order that the 1:1 (vol./vol.) mix in the observation chamber exhibits a pH of 7.4. Hybridization traces were recorded and rate constants calculated as previously. Pre-formed DNA-tile hybrids (Fig. 3, model 3) were obtained by incubation for 30 min at 24 °C of an equimolar mix of half-complementary tiles in PBS buffer pH 7.4. Two different types of hybrids were loaded into separated stopped-flow syringes, and after rapid mixing, hybridization traces were recorded and rate constants calculated as previously.

2.3. Experimental data processing and visualization of simulation output procedures

Expected electrophoresis motilities of species and their relative staining responses were computed based on experimental calibrations with ss- and ds-species of known structures. Molecular marker lines were scanned and peak shapes modeled using polynomial modified Gaussian distributions. The calibration curves (log plots) were fitted on polynomial laws. Experimental profiles were first corrected for inhomogeneous migrations between lanes using a non-linear local expansion/compression algorithm then were processed to improve band resolution and reduce noise: gel profiles were deconvoluted using peak shapes of molecular markers degraded about 2.1-fold by increasing their half-width. This was necessary to take into account the broader electrophoretic migration of segmented polymers. Finally, experimental and simulated data were reconstructed using the same virtual resolution typical of ds-DNA.

3. Simulation procedures

The flow chart for the simulation of the 1D self-assembly of DNA tiles is represented on Fig. 1. This computational process was specifically designed to simulate assembly of bifunctional tiles by base-pairing but can as well apply to the polymerization of any kind of linear block-polymers provided of limited adaptations. The algorithm comprises two main complementary sections: one, depicted on the left panel, aims to generate the collection of possible species which can be formed by random association of members and descendant of a seeding set of tiles, and the other (right panel) uses the generated data to automatically build up and solve differential equations describing the system. Depending on the structure of the seeding set, two types of situations can be encountered giving rise, respectively, to finite and infinite series. When combinatorial convergence is possible, all species fulfilling the association rules are generated. In contrast, in the case of non-converging sets (for example, formation of unlimited block-copolymer), only a part of possible species is generated. In such a case, a critical point is that the combinatorial procedure will always generate species which mainly contribute to the observed population in given experimental conditions.

3.1. Combinatorial generation of species and of formation paths

As described in Fig. 1, left panel, the initial seeding set is first transferred into an accumulator which stores both species description and their possible formation paths (step 1). The algorithm then enters an iterative process in which all possible couples of species present in the accumulator are tested for fulfilling criteria for a coupling process (step 2) involving either one, the other, or both of their extremities (cyclization). Allowed couplings are defined by heuristic rules, which, in the case of DNA tiles, rely on base-pairing stability but which could differ, for example, corresponding to chemical reactivity in case of some types of polymers. Species formed as the result of a successful coupling event (step 3) are compared to the content of accumulator. Novel species are created and stored. For already existing ones, the new formation path is added to the species descriptor. Non-converging combinatorial processes have to be stopped after a specific number of cycles, thus truncating species generation. During the last cycle, storage process differs: paths for existing species are registered in the accumulator while new generated species are stored in a sandbox. The concentration of species accumulated within the sandbox must remain neglectable as compared to total concentration of species at the final time simulated. Otherwise, calculation must be restarted using a larger number of combinatorial cycles or the simulation stopped at a shorter time to decrease the sandbox content.

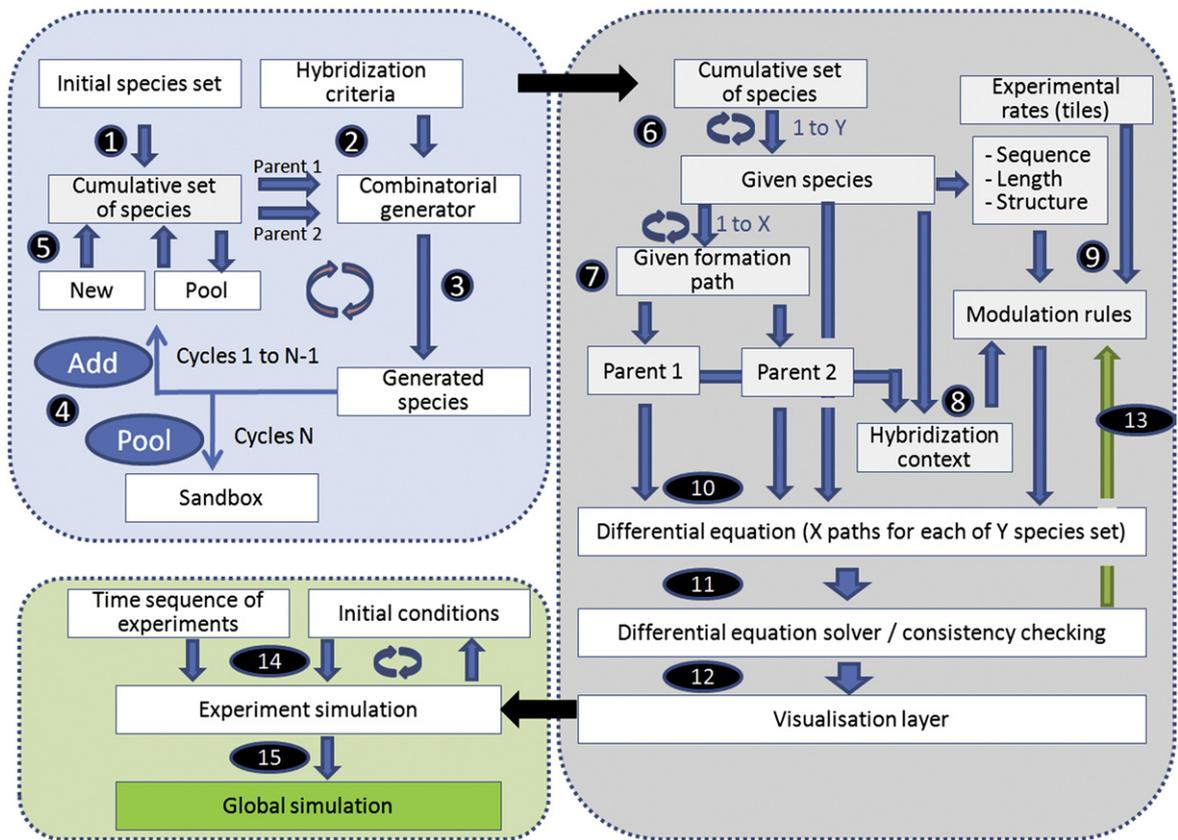


Fig. 1. Flow chart for the simulation 1D-self-assembly of DNA tiles. The top-left panel corresponds to the multi-cycle combinatorial generation of models, species and reactions that is used to generate the set of differential equations. This set is used in the right panel to calculate the time dependant evolution of the species from given initial conditions. In the bottom-left panel, successive initial conditions can be cumulated to simulate a complex multi-step experiment.

3.2. Differential equation computation and parameter evaluations

Computation starts from the previously generated contents of accumulator and sandbox (step 6 in Fig. 1, right). For each species in the accumulator, all forming paths are recovered. For a given path, parent species are first taken (step 7) and their structural descriptors compared to the ones of the daughter species. A class defining context for the event is deduced from the structures (ss- or ds-DNA, attached protein domain, etc.) surrounding the hybridizing segments of each parent (step 8). The resulting class, the sequence of hybridizing segments, the length of parent and daughter fragments, and the computed J-factors in case of formation of cyclic species are taken into account to evaluate a realistic rate constant for the corresponding coupling event (step 9). The process is repeated for all possible forming paths of a given species and for all species present in the accumulator. All coupling reactions are balanced by corresponding consumption of parents of any formed species. Formation of species included in the sandbox is similarly processed. However, for species belonging to sandbox, individual structural descriptors are not stored and their concentrations are cumulated. Accordingly, reactions in which sandbox species are involved as parents can no more be evaluated, leading to systematic errors which remain acceptable only if the overall sandbox content remains very low as compared to concentration of all other species. It is possible to check that this is always doable provided that the number of combinatorial cycles to a particular system and simulation duration be adapted.

Concerning reverse reactions (i.e., splitting of a parent species into two daughter species), the encoding process is basically the same than for forward reactions. Individual rate constants can be deduced from the calculated rates for corresponding forward reactions and from the equilibrium constants evaluated by standard thermodynamic methods [19]. However, in our experimental conditions, hybridiza-

tions are mainly irreversible and reverse reactions can be neglected. An optional feature of the algorithm involves a feedback loop to correct for unusual dead-end events (for example, formation of triplex nodes) using an empirical law (step 13). All together, steps 1–10 give rise to a set of differential equations (step 11) which are solved by classical algorithms (Runge–Kutta–Fehlberg ODE Solver). Due to the very large number of species and structures generated, output of simulations cannot be directly compared to experimental data. To solve the problem, additional routines are in charge to reconstruct synthetic profiles (electrophoresis gels or column) from simulation outputs and from experimentally derived peak shapes and calibration curves taking into account DNA structures and detection system responses (step 12). The whole algorithm can be repeated using different settings (bottom-left panel) to simulate multi-step experiments in which final conditions from one setting are used as initial conditions for another setting.

3.3. Coupling criteria and rules for kinetic rate calculations

Hybridization and diffusion rate constants were automatically generated for all possible reactions using experimental rates determined for hybridization between short ss-tiles featuring the same sequences. Generic rules are used to correct these values for the hybridization context and parent sizes. For that, an equivalent length was calculated for each parent using a weighted sum of the size of its ds- and ss-segments. A first evaluation of the hybridization rate constant was obtained as $k = \frac{(k_1 + k_2)}{2}$ with $k_1 = k_{\text{ref}} \times \left(\frac{L_{\text{ref}}}{L_1}\right)^{0.58}$ similarly for k_2 . L_1 and L_2 are the equivalent lengths for the two species involved in the hybridization and L_{ref} and k_{ref} are experimentally determined values. These equations are semi-empirical but chosen based on previous works [12–15,19–21] assuming constant

hybridization lengths (19–20 bases) but variable total length of parents. A second correction was applied to take into account the hybridization context. In the present work, modulation is only based on the presence of ds-segments flanking or not the hybridizing ss-segments. Rates of formation of cyclic species were evaluated using a J-factor table (sub-table PT2 in supplementary data) derived from literature and the equivalent length of species.

3.4. Algorithm implementation

Two companion softwares, PolyB5 (model building and simulation) and DeconvMagic (visualization routines) have been written in C++ object-oriented language. They are interactively controlled by windowed visual interfaces implemented using the Microsoft Visual Studio® development environment. All modeling parameters including alternate models and rules for rate constant calculations can be user modified (see Supplemental Tables S1 and S2). Complexity of models is solely limited by hardware capability, which is rarely limiting for system encompassing a finite number of species. Models involving unlimited polymers can be computing time demanding. As an example, eight combinatorial cycles with a model involving four different polymerizing tiles and four terminating tiles allow accurate calculation (3 digit accuracy) of the concentrations of all possible assemblies (640 species) smaller than 64 U (2496 bp). In such conditions, up to 128 alternate reactions can participate to the synthesis of a given polymer and species larger than 64 tiles per strand resulting from the 33,088 truncated reaction paths are found in the sandbox.

4. Results

Computational approaches (Fig. 1) were developed for the simulation of the combinatorial assembly of DNA tiles into defined or unlimited polymers. The algorithm described in the simulation procedure section allows to generate and to take into consideration all possible intermediate species when the assembly converges to one or

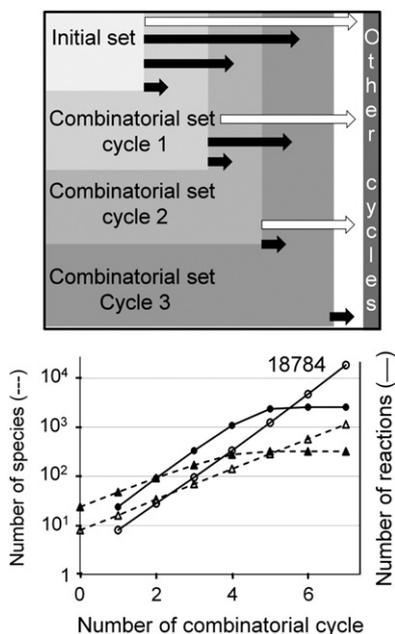


Fig. 2. Combinatorial expansion of species and reactions during model generation. Top, *in silico* combinatorial generation of species and associate reactions. Arrows illustrate how each combinatorial cycle is feeding species generation for the following cycles. Bottom, evolution of the number of species (dotted lines) and reactions (solid lines) accumulated after a given number of combinatorial cycles for converging (solid symbols) and unlimited (open symbols) experimental systems. The parameter settings (P1, P2) for corresponding experimental models 1 and 2 are described in the supplementary data (Table S3).

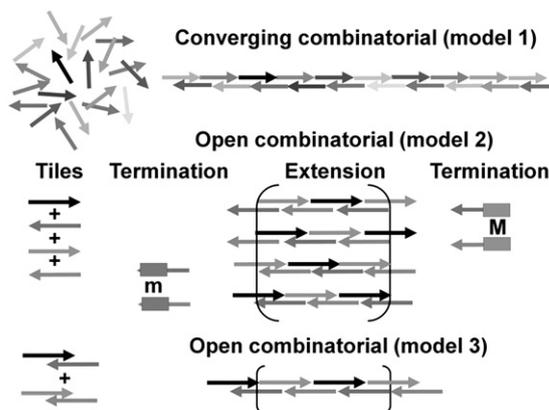


Fig. 3. Models for 1D combinatorial assembly. Model 1, converging assembly of ss-DNA-tiles (arrows) as typically found during gene synthesis. Model 2, combinatorial assembly of ss-DNA tiles in the absence or the presence of terminating tiles (m and M) able to cap extending chains. Model 3, unlimited assembly of pre-formed tile-dimers.

a limited number of end-products. In contrast, when unlimited DNA polymer extension can occur, only the subset of relevant intermediates is generated. In such a case, combinatorial truncation is performed in a way that guarantees accurate time course simulation for this subset of species, whose size depends on total polymerization duration. This is permitted by the use of a multi-cycle combinatorial procedure to generate the sets of species (Fig. 2, top). During combinatorial cycles, the number of generated species and of associated reactions increases exponentially, prior reaching a limit only in the case of converging assemblies (Fig. 2, bottom). Practically, up to a million of parallel reactions can be managed on a regular desktop computer.

4.1. Theoretical time course of species concentrations during converging and open combinatorial assemblies

Three experimental models were considered (Fig. 3). Model 1 is typical for single or combinatorial gene synthesis and is characterized by the possible formation of single or a limited number of end-products as a result of tile assembly. Models 2 and 3 correspond to an open tile assembly, which can lead to an infinite number of products. In model 2, two ss-DNA tiles (AA' and BB'; Table 1) can be polymerized by hybridization with half-complementary tiles (ab' and ba'). This results in segmented DNA polymers featuring AA'–BB' sequence repeats. The polymer length can be controlled by terminating tiles (Ma', Mb', am, bm) which introduce non-hybridizing extensions at the extremities. Alternatively, in model 3, pairs of tiles can be pre-assembled before initiating the polymerization.

Model 1 was simulated using equimolar concentration of tiles and arbitrary but realistic hybridization rate constants (see below). Predicted size distribution of intermediate species leading to the unique final assembly product is given as function of incubation time

Table 1

DNA-tile sequences and matches. The AA'-tile can hybridize with the half-complementary ab'- and ba'-tiles by matching A/a and A'/a' sequence segments, respectively. The BB'-tile can hybridize to half parts of ab'- and ba'-tiles.

Tiles	Sequences (5'–3')	Bases
A–A'	CCGTAGGGTCTCCATTTTGA–AGCGAAGTCACACTGTACA	39
B–B'	GTCCATATGCAAGTCAGCGT–CAGTGTACAGTCACGTGAC	39
a–b'	TCAAAATGGAGACCCCTACGG–CTGACGTGACTGTACACTG	39
b–a'	ACGCTGACTTGCCATATGGAC–TGTACAGTGTGACTTCGCT	39
a–m	TCAAAATGGAGACCCCTACGG–GTAAAACGACGGCCAG	36
b–m	ACGCTGACTTGCCATATGGAC–GTAAAACGACGGCCAG	36
M–a'	CAGGAAACAGCTATGAC–TGTACAGTGTGACTTCGCT	36
M–b'	CAGGAAACAGCTATGAC–CTGACGTGACTGTACACTG	36

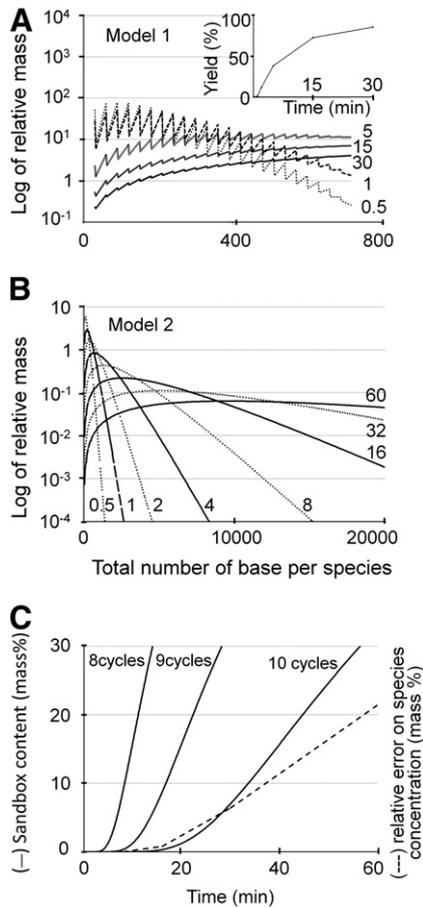


Fig. 4. Distribution of species sizes during assembly and validations. A, simulated size distribution of species for model 1 (gene synthesis) for different assembly duration (on the right in min). Sizes are given as total nucleotide number. Insert gives yield (mol %) for the fully assembled gene sequence (752 nucleotides) as function of incubation time. B, simulated size distribution of species for model 2 (unlimited polymer) for different assembly duration. C, Validation of simulations in the case of model 2. Sandbox content (mass %) relative to total species concentration in models generated using 8 to 10 combinatorial cycles (solid lines). Estimated average relative error (mass %) on simulated species concentration for a 9 cycles model as function of final simulation time (dotted line).

in Fig. 4A. Accumulation of end-product (25 tiles, 375 bp) is predicted to reach a plateau after 30 min which corresponds to about 150-fold the half-time for hybridization of 2 individual tiles. The construction of a combinatorial library of genes was simulated using the same model but assuming alternate variant tiles at three sequence positions (Supplemental Fig. S2). The figure illustrates that distribution of intermediates and variant end-products can be predicted as function of incubation time. When the open model 2 was simulated (Fig. 4B), size distribution of polymer species versus time rapidly diverged from the figure observed with model 1. No end-product is accumulating and an increasingly flat distribution of species size is observed with longer incubation time. This raises the question of the limit of validity of the truncated model when incubation times are large enough to allow formation of long polymers which cannot be taken into account in the model. This problem is addressed on Fig. 4C (model 2) which gives the content of the sandbox (non-considered species) and of the estimated relative error on calculations as a function of incubation time and of combinatorial cycles. The content of the sandbox, which is typically around 15 mass% for 20 min and 9 cycles, corresponds to a simulation error on the concentration of considered species of about 2 mol%. This content appears to be a good indicator for the number of combinatorial cycles to be used to simulate long polymerization reactions.

4.2. Analysis of critical parameters expected to affect polymer size and structure in open combinatorial systems

The domain of applications of linear assembly of DNA tiles mainly concerns gene and combinatorial library synthesis. However, such systems are generally too diverse and complex to be able to easily define the general rules governing the modulation of the large number of individual rate constants required for simulation. In order to analyze these factors, we focused on simpler models 2 and 3, which involve a limited number of types of primary hybridization events. We first attempted to identify critical parameters in such a system and particularly questioned the potential role of the formation of cyclic species and of the length dependence of hybridization rates.

Model 2 was first considered using an equimolar mixture ($1 \mu\text{M}$) of 4 polymerizing tiles and identical rate constants ($4 \mu\text{M}^{-1} \text{min}^{-1}$) for all primary hybridization reactions. In this model, only species with an even number of units can cyclize. Fig. 5A illustrates that such event is expected to significantly affect the distribution of non cyclic-species. The effect critically depends on the J-factor function, which defines the concentration at which intra-molecular and inter-molecular hybridizations between cohesive extremities exhibit the same rates. In the presented simulation, J-factor values (sub-table PT2 in supplementary data) were chosen based on the reported length-dependence of ligation rates for ds-DNA models [22–26] but might need some adaptation in the case of block copolymers including ss-nicks [22,27–29]. A second critical parameter is the modulation of hybridization rates by the length of species (Fig. 5B). In model 2, hybridizing segments of constant lengths are associated with ds-DNA parts of variable lengths. When a dependence of hybridization rate proportional to the inverse of power 0.6 of species length [15] was introduced in the simulation, the figure was markedly changed, indicating that this parameter has to be taken into account. The effect of additional factors, like relative tile concentrations, is illustrated on Supplemental Fig. S3. Slight unbalance of the molar ratio between tiles marginally affected profiles for short polymerization times but caused oscillations as a function of length for longer times.

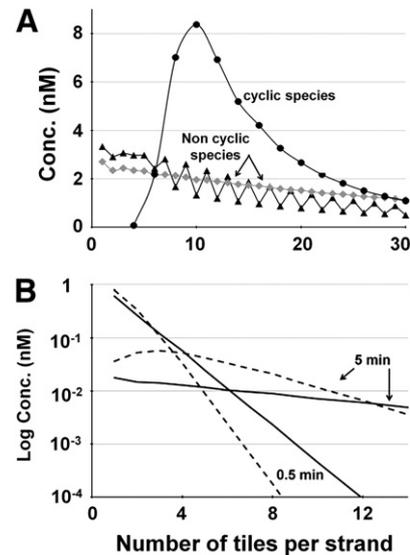


Fig. 5. Effect of the formation of cyclic species and of length dependence of rate constant on species distribution. A, Model 2 (Fig. 2) was simulated without considering formation of cyclic species (grey line) or taking them into account (black lines). The concentrations of polymers encompassing the same number of linking tiles were plotted versus size for a polymerization time of 15 min. B, simulations of the influence of the length dependence of the hybridization rate constants in Model 2. A modulation as the inverse of the power 0.58 of length was used (dotted lines) or not (solid lines). All model parameters are given in Table S3.

Table 2
Determination of hybridization rate constants by rapid mixing experiments.

Seq. 1	Seq. 2	Hybridization	Orientation ^a	Hypochromic factor (%)	Rate constant ($\mu\text{M}^{-1} \text{min}^{-1}$)
AA'	ab'	A/a	5'-5'	7.7	4.4
AA'	am	A/a	5'-5'	6.7	4.5
AA'	ba'	A'/a'	3'-3'	6.4	7.0
AA'	Ma'	A'/a'	3'-3'	9.7	4.2
BB'	ba'	B/b	5'-5'	6.8	4.3
BB'	bm	B/b	5'-5'	7.0	6.4
BB'	ab'	B'/b'	3'-3'	5.6	4.1
BB'	Mb'	B'/b'	3'-3'	5.5	2.8
AA' + ba'	BB' + ab'	A/a and B/b	Both	10.9	6.0
AA' + BB' + ba' + ab'	Buffer ^b	All	Both	17.9	3.9

Hybridization rates were determined at 24 °C in PBS buffer pH 7.4. The rates were deduced from hypochromic change at 260 nm upon hybridization of DNA in triplicate stopped-flow experiments. Relative errors on rate constant calculations are estimated to be about of 15%. Absolute error on hypochromic factor was estimated to be less than 0.5%.

^a Extremities of tiles that interact.

^b The four DNA-tiles in alkaline buffer were mixed with a neutralizing buffer (see [Materials and methods](#)).

4.3. Evaluation of model parameters based on experimental data

The observed second order rate constants for tile hybridization ranged from 2.8 to 7.0 $\mu\text{M}^{-1} \text{min}^{-1}$ depending on sequences ([Table 2](#)) consistently with described values for fully complementary ss-DNAs of 20–25 bases [[30,31](#)]. Interestingly, rates for hybridizations involving the same ss-segments but different non-hybridizing exten-

sions differ by up to 70%. [Table 2](#) also illustrates that the presence of ds-flanking regions increases by ~1.58-fold the hybridization rates as compared to ss-flanking regions. Concerning associated hypochromic factors (hf), their values depend on the length of continuous ds-segments. Hybridization of two ss-tiles on half of their length gave rise to an hf ranging from 5.5% to 9.7% of the ss-DNA absorbance, well-comparing with a theoretical value [[32](#)] of ~9%. For the formation of fully ds-structures from pre-formed tile duplexes or from ss-tiles, hf values of 11% and 18% were observed, well-comparing with theoretical values of 9% and 18%, respectively.

Two experiments were performed to test for a potential influence on hybridization rates of the presence of ds-regions flanking the ss-sequence. The first involved the initiation by a large pH shift of the polymerization of four ss-tiles (model 2) whereas the second results from the mixing of two types of preformed tile duplexes (model 3). Long copolymers constitute the end-products in the two cases. Reactions were monitored using the hybridization related hypochromic effect (hf), which was found, as expected, ~2-fold smaller with model 3 than with model 2 ([Fig. 6A](#), curves E2 and E1). Hypochromic effects (17.1%–18.2% and 10.4%–11.4%) deduced from simulations well-compared with experimental values of 17.9% and 10.9%.

Good simulations of model 2 experiments can be achieved ([Fig. 6B](#)), whatever a modulation of hybridization rate constants by the structure (ds versus ss) of flanking regions is taken or not into account. This

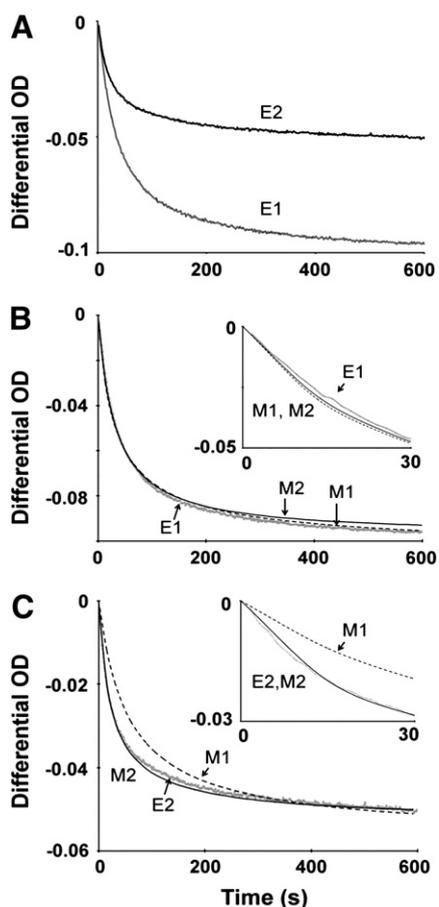


Fig. 6. Hypochromic effects resulting from hybridization were monitored during rapid-mixing experiments and simulated. A, experimental traces for hybridization of four ss-tiles (model 2, trace E1) and of hybrid tiles (model 3, trace E2). B, simulations of E1 trace from panel A using model 2 (SP5 parameters with hf=17.1%, curve M1) and SP6 parameters with hf=18.2%, (curve M2). C, simulations of E2 trace from panel A using model 3 (SP7 parameters with hf=11.4%, curve M1) and SP8 parameters with hf=10.4% (curve M2). Factor hf (hypochromic factor) represents the ratio between the absorbance change at hybridization complexation and the initial absorbance. All simulation parameters are given in [Table S3](#).

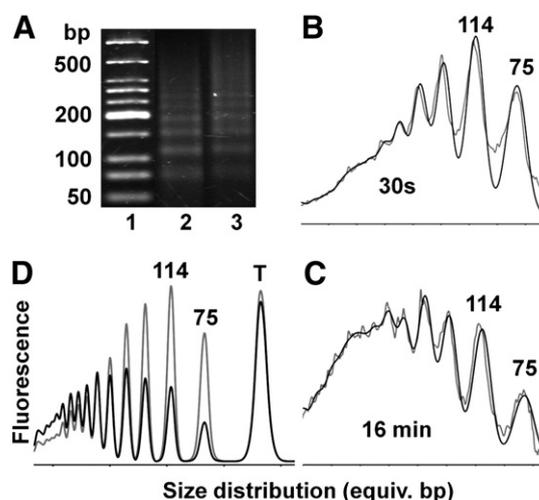


Fig. 7. Deconvolution of agarose gel profile of polymers. A, products resulting from hybridization of four polymerizing tiles and four terminating tiles (Model 2) were loaded on agarose gel and stained. Lane 1, MW markers. Lanes 2 and 3, polymers extended for 30 s and 16 min. B, lane 2 was scanned and the trace deconvoluted (grey & black traces). C, same for lane 3. D, comparison of deconvoluted and reconstructed experimental profiles for 30 s and 16 min polymerization times (grey & black lines) using improved peak shapes typical of the migration of ds-DNA.

is expected for a model where the first step involves hybridizations only between ss-tiles. In contrast, simulation of model 3, which involves only hybridizations between tile duplexes, clearly illustrated that rate constants have to be adjusted to take into account the presence of ds-flanking regions (Fig. 6C). In all simulations, an inverse power law (exponent, -0.58) for the dependence of rate constants on the length of reacting species was used. Failure to incorporate this modulation did not allow good fitting (data not shown). The J-factor function [23] used in simulation marginally affected fitting since experimental concentrations were chosen one order of magnitude higher than larger J-factors.

4.4. Experimental validation of the simulated distribution of species

Size distribution of polymers was analyzed using a variant of model 2 in which 4 terminating tiles were used to quench assembly process. Agarose and polyacrylamide gel electrophoresis were used to monitor polymer size distribution. However, the migration behavior of block-segmented DNA was found “blurred” compared with regular ds-DNA, as previously observed [27,33]. In addition, band overlaps between polymers of similar lengths contribute to the generation of an electrophoretic pattern constituted of diffuse bands imprinted on a continuum of signal. To solve these difficulties, a reconstruction approach was developed to compare experimental and simulated data. This was performed through modeling (shapes and retention times) of the analytical system response using a set of ds-DNA references. The model was used for deconvolution of experimental data (positive least square regression) and reconstruction of experimental and simulated data using the same virtual gel resolution. Polymers extension was performed for variable durations (0.5 to 16 min) before quenching by addition of an excess of terminating tiles. Species encompassing up to 10 tiles (426 bases) per strand can be reasonably quantified

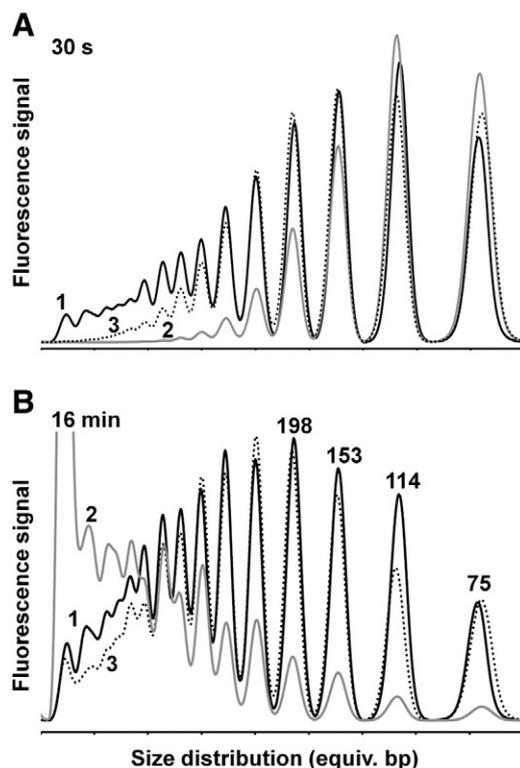


Fig. 8. Simulations of gel profiles of polymers. Curves from Fig. 7D are compared with corresponding simulated gel profiles. A, curve 1, processed experimental trace (30 s); curves 2 and 3, simulations using model 2 without or with the context modulation of rate constants, respectively B, curve 1, processed experimental trace (16 min); curves 2 and 3, simulations using model 2 without and with strand invasion mechanism, respectively. Parameters are described in Table S3.

(Fig. 7). For comparison, synthetic gel profiles were reconstructed from simulations including the quenching step. Considering short polymerization times (for example, 0.5 min; Fig. 8A), simulation (curve 2) without taking into account the modulation of hybridization rates by flanking regions diverged from experimental results (curve 1). In contrast, when this modulation was taken into account (curve 3), accurate fit was achieved. However, for longer polymerization times (larger than 2 or 3 min), experimental course significantly diverged from model prediction (Fig. 8B, curve 2). Unexpected experimental limitation of polymers elongation was observed on gels with very little further profile evolution even after 24 h of incubation. It is unclear whether the observation results from disruption of large segmented structures in electrophoresis conditions or from topological enlacement of extending species. However, this last hypothesis could be supported by AFM imaging, which evidenced the formation of mesh-like structures when polymerization time is increased (Supplemental Fig. S4). This phenomenon could be empirically introduced into the simulations by decreasing all hybridization rates proportionally to the accumulation of large polymers, thus allowing recovering realistic profiles (Fig. 8B, curve 3).

5. Discussion and conclusion

Global modeling of combinatorial assembly of DNA tiles into defined or unlimited polymers was developed. The time course distribution of copolymers was simulated in both finite and open combinatorial systems pointing out main critical parameters. Tile-based assembly of synthetic genes or combinatorial libraries is a widely used approach in biotechnology. The development of predictive models is particularly important to optimize experimental conditions for the design of combinatorial libraries used in directed evolution approaches. In the case of formation of unlimited polymers, the model is able to address the problem of the combinatorial expansion of species population, opening routes to experimental validations of parameters. A key point of the computational approach is to use an iterative combinatorial algorithm to generate only species which can accumulate in significant amounts in given experimental conditions, thus leading to manageable differential equation sets. The infinite number of non-considered species is pooled in a sandbox, whose content can be examined to secure the truncated models for given experimental conditions. The number of combinatorial cycles used for model generation can be adjusted accordingly. The algorithm can indifferently simulate irreversible or reversible hybridizations. In the later case, reverse (dissociation) reaction rates are evaluated from given association rates, length and environment dependence laws, and predicted thermodynamic parameters calculated with the classical closest neighbor approach [34,35]. However, the experimental system developed in this report only involves hybridization in nearly irreversible conditions. We checked that, in our experimental conditions, this situation is prevalent up to 50–52 °C.

A critical point is to find ways allowing evaluation from a limited number of experimental data of the large range of rate constants involved in simulations. This makes the process particularly complex in the case of gene or library synthesis, which involves a large number of different tiles. Even if recent microfluidic optical technologies could open the way to automate high-throughput measurement of primary hybridization rates [36–39], we focused on a simpler model of copolymer involving a limited set of tiles. The developed algorithm includes suitable routines to automatically classify and identify for each reaction relevant hybridization contexts. These hybridization contexts are used to evaluate model parameters as a function of accessible experimental rates and of generic rules taking into account the size of hybridizing species and the structure of the regions flanking the hybridizing sequences. The study particularly unveiled that the presence of flanking ds-sequence significantly increases the rate of hybridization of attached ds-ss-regions. The dependence of the

hybridization rates on the length of interacting species was as well investigated and unveiled an invert dependence on the square root of the overall length for hybridizing segment of constant lengths. Although such law was never previously reported for segmented structures, it appears consistent with theoretical considerations [12,19]. This contrasts with a length dependence for the hybridization rates of two fully ss-DNAs, which is roughly proportional to the square root of the length [12,14]. The inverse square root law is similar to the reported law for the variation with length of DNA species diffusion coefficients. More precisely, ds-DNA longer than a few thousand base pair have been shown to form self-avoiding random coils with a diffusion coefficient $D \sim L^{-0.6}$, but shorter ds-DNA are not long enough to form non-draining coils. Ds-DNA down to 370 bp have been shown to follow $D \sim L^{-0.7}$ [15]. Shorter ds-DNA would increasingly approach the “rigid rod” limit, ultimately probably having $D \sim \log(L/d)/L$, where the d is the diameter of the rod. In contrast, ss-DNA would probably be expected to act as a freely-jointed chain forming a random coil with $D \sim L^{-0.6}$.

Quantitative comparison of the time dependence of the experimental distribution of polymers size with simulations was performed using a reconstruction approach. The procedure involves the renormalization of gel electrophoresis figures by deconvolution in parallel with the reconstruction of realistic synthetic gel profiles from the complex outputs of simulation. This offers a unique way to compare complex experimental and simulated distribution of species, thus facilitating parameter validations. Accurate simulations of species profiles were obtained for short polymerization times allowing algorithm and parameter validation. The divergence between experimental and predicted length distributions for long polymerization times highlighted an unexpected behavior of the experimental system or potential artifact when analyses of large segmented structures were performed using gel (agarose or acrylamide) electrophoresis. AFM imaging revealed the formation of unpredicted mesh-like structures which are not taken into consideration into the modeling of linear assemblies. While observation could result from mechanism like strand invasion phenomenon [40–43], alternate hypothesis of enlacement mechanism can be considered as well. Interestingly enough, simulations evidenced that the distribution of linear species observed for long polymerization times are identical to the ones predicted in the model for shorter times. The phenomenon is thus formally equivalent to a global and parallel decrease of all hybridization rates.

In conclusion, the presented approach demonstrated that it is possible to quantitatively simulate tile-based assemblies leading to the formation of unlimited DNA block copolymers and by extension to defined end-products. This is particularly the case for combinatorial assembly of alternate tiles to generate synthetic diversity, a critical point in a tight control of the proportion of the generated alternate sequences. Modeling based on the presented algorithm would clearly open a route for optimized design taking into account specific hybridization properties of alternate tiles. The approach is generic enough to be transposed to the synthesis in reversible conditions.

Acknowledgements

We thank Dr. Gilles Truan for careful proofreading of the paper and for fruitful discussions. This work was supported by grant ANR-05-NANO-021-04 from the Agence Nationale de la Recherche, France.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.bbagen.2011.03.002.

References

- [1] A.S. Xiong, R.H. Peng, J. Zhuang, F. Gao, Y. Li, Z.M. Cheng, Q.H. Yao, Chemical gene synthesis: strategies, softwares, error corrections, and applications, *FEMS Microbiol. Rev.* 32 (2008) 522–540.
- [2] J.R. TerMaat, E. Pienaar, S.E. Whitney, T.G. Mamedov, A. Subramanian, Gene synthesis by integrated polymerase chain assembly and PCR amplification using a high-speed thermocycler, *J. Microbiol. Methods* 79 (2009) 295–300.
- [3] C. Neylon, Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution, *Nucleic Acids Res.* 32 (2004) 1448–1459.
- [4] M.J. Czar, J.C. Anderson, J.S. Bader, J. Peccoud, Gene synthesis demystified, *Trends Biotechnol.* 27 (2009) 63–72.
- [5] A.S. Xiong, R.H. Peng, J. Zhuang, J.G. Liu, F. Gao, J.M. Chen, Z.M. Cheng, Q.H. Yao, Non-polymerase-cycling-assembly-based chemical gene synthesis: strategies, methods, and progress, *Biotechnol. Adv.* 26 (2008) 121–134.
- [6] V. Stein, F. Hollfelder, An efficient method to assemble linear DNA templates for in vitro screening and selection systems, *Nucleic Acids Res.* 37 (2009) e122.
- [7] D. Liu, H. Jiang, K. Chen, R. Ji, A new approach to design virtual combinatorial library with genetic algorithm based on 3D grid property, *J. Chem. Inf. Comput. Sci.* 38 (1998) 233–242.
- [8] A.Y. Koyfman, S.N. Magonov, N.O. Reich, Self-assembly of DNA arrays into multilayer stacks, *Langmuir* 25 (2009) 1091–1096.
- [9] S.H. Park, C. Pistol, S.J. Ahn, J.H. Reif, A.R. Lebeck, C. Dwyer, T.H. LaBean, Finite-size, fully addressable DNA tile lattices formed by hierarchical assembly procedures, *Angew. Chem. Int. Ed Engl.* 45 (2006) 735–739.
- [10] W. Boireau, A.C. Duncan, D. Pompon, Bioengineering and characterization of DNA–protein assemblies floating on supported membranes, *Methods Mol. Biol.* 300 (2005) 349–368.
- [11] D. Pompon, A. Laisne, PDNA as building blocks for membrane-guided self-assemblies, *Biochem. Soc. Trans.* 35 (2007) 495–497.
- [12] R.M. Robertson, S. Laib, D.E. Smith, Diffusion of isolated DNA molecules: dependence on length and topology, *Proc. Natl Acad. Sci. U.S.A.* 103 (2006) 7310–7314.
- [13] D.E. Smith, T.T. Perkins, S. Chu, Dynamical scaling of DNA diffusion coefficients, *Macromolecules* 29 (1996) 1372–1373.
- [14] J.G. Wetmur, N. Davidson, Kinetics of renaturation of DNA, *J. Mol. Biol.* 31 (1968) 349–370.
- [15] S.S. Sorlie, R. Pecora, A dynamic light scattering study of four DNA restriction fragments, *Macromolecules* 23 (1990) 487–497.
- [16] K. Fujibayashi, D. Zhang, E. Winfree, S. Murata, Error suppression mechanisms for DNA tile self-assembly and their simulation, *Natural Computing: an International Journal* 8 (2009) 589–612.
- [17] P.W. Rothmund, N. Papadakis, E. Winfree, Algorithmic self-assembly of DNA Sierpinski triangles, *PLoS Biol.* 2 (2004) e424.
- [18] M. Muksitul Haque, A. Kalyanaraman, A. Dhingra, N. Abu-Lail, K. Graybeal, DNAjig: a new approach for building DNA nanostructures, *IEEE International Conference on Bioinformatics and Biomedicine* (2009) 379–383.
- [19] J.L. Sikorav, H. Orland, A. Braslau, Mechanism of thermal renaturation and hybridization of nucleic acids: Kramers’ process and universality in Watson–Crick base pairing, *J. Phys. Chem. B* 113 (2009) 3715–3725.
- [20] J.G. Wetmur, Excluded volume effects on the rate of renaturation of DNA, *Biopolymers* 10 (1971) 601–613.
- [21] J.G. Wetmur, Hybridization and renaturation kinetics of nucleic acids, *Annu. Rev. Biophys. Bioeng.* 5 (1976) 337–361.
- [22] Q. Du, M. Vologodskaja, H. Kuhn, M. Frank-Kamenetskii, A. Vologodskii, Gapped DNA and cyclization of short DNA fragments, *Biophys. J.* 88 (2005) 4137–4145.
- [23] J.A. Schellman, S.C. Harvey, Static contributions to the persistence length of DNA and dynamic contributions to DNA curvature, *Biophys. Chem.* 55 (1995) 95–114.
- [24] Y. Zhang, D.M. Crothers, High-throughput approach for detection of DNA bending and flexibility based on cyclization, *Proc. Natl Acad. Sci. U.S.A.* 100 (2003) 3161–3166.
- [25] Y. Zhang, D.M. Crothers, Statistical mechanics of sequence-dependent circular DNA and its application for DNA cyclization, *Biophys. J.* 84 (2003) 136–153.
- [26] Y. Zhang, A.E. McEwen, D.M. Crothers, S.D. Levene, Statistical-mechanical theory of DNA looping, *Biophys. J.* 90 (2006) 1903–1912.
- [27] P. Furrer, J. Bednar, A.Z. Stasiak, V. Katritch, D. Michoud, A. Stasiak, J. Dubochet, Opposite effect of counterions on the persistence length of nicked and non-nicked DNA, *J. Mol. Biol.* 266 (1997) 711–721.
- [28] B. Tinland, A. Pluen, J. Sturm, G. Weill, Persistence length of single-stranded DNA, *Macromolecules* 30 (1997) 5763–5765.
- [29] M. Vologodskaja, A. Vologodskii, Contribution of the intrinsic curvature to measured DNA persistence length, *J. Mol. Biol.* 317 (2002) 205–213.
- [30] Y. Gao, L.K. Wolf, R.M. Georgiadis, Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison, *Nucleic Acids Res.* 34 (2006) 3370–3377.
- [31] M.M. Sekar, W. Bloch, P.M. St John, Comparative study of sequence-dependent hybridization kinetics in solution and on microspheres, *Nucleic Acids Res.* 33 (2005) 366–375.
- [32] A.V. Tataurov, Y. You, R. Owczarzy, Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids, *Biophys. Chem.* 133 (2008) 66–70.
- [33] C. Gaillard, L.S. Shlyakhtenko, Y.L. Lyubchenko, F. Strauss, Structural analysis of hemicatenated DNA loops, *BMC Struct. Biol.* 2 (2002) 7.
- [34] J. SantaLucia Jr., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, *Proc. Natl Acad. Sci. U.S.A.* 95 (1998) 1460–1465.

- [35] J.G. Wetmur, DNA probes: applications of the principles of nucleic acid hybridization, *Crit. Rev. Biochem. Mol. Biol.* 26 (1991) 227–259.
- [36] A. Hine, X. Chen, M. Hughes, K. Zhou, E. Davies, K. Sugden, I. Bennion, L. Zhang, Optical fibre-based detection of DNA hybridization, *Biochem. Soc. Trans.* 37 (2009) 445–449.
- [37] G. Jiang, R. Chen, H. Yan, Q. Ouyang, A new method of preparing fiber-optic DNA biosensor and its array for gene detection, *Sci. China C Life Sci.* 44 (2001) 33–39.
- [38] D. Walt, Fibre optic microarrays, *Chem. Soc. Rev.* 39 (2010) 38–50.
- [39] D.R. Walt, Fiber optic array biosensors, *Biotechniques* 41 (2006) 529–535.
- [40] S.W. Choi, A. Kano, A. Maruyama, Activation of DNA strand exchange by cationic comb-type copolymers: effect of cationic moieties of the copolymers, *Nucleic Acids Res.* 36 (2008) 342–351.
- [41] K. Frykholm, B. Norden, F. Westerlund, Mechanism of DNA strand exchange at liposome surfaces investigated using mismatched DNA, *Langmuir* 25 (2009) 1606–1611.
- [42] K. Tajima, W.J. Kim, T. Akaike, A. Maruyama, Acceleration of DNA strand exchange reaction by cationic polymers, *Nucleic Acids Res. Suppl.* (2002) 265–266.
- [43] D. Lubrich, S.J. Green, A.J. Turberfield, Kinetically controlled self-assembly of DNA oligomers, *J. Am. Chem. Soc.* 131 (2009) 2422–2423.

Appendix A for:

Design and validation of a global combinatorial model for 1D-assembly of DNA-tiles

Aude Laisne^a, Eric Lesniewska^b, and Denis Pompon^{a*}

^a Centre de Génétique Moléculaire. CNRS. Avenue de la Terrasse, F91190 Gif-sur-Yvette, France, ^b Institut Carnot Bourgogne UMR CNRS 5209, Université de Bourgogne, F-21078 Dijon, France.

Corresponding Author:

Denis POMPON. pompon@cgm.cnrs-gif.fr Phone: +33169823680; fax: +33169823682

Supplemental protocol 1: DNA-tile purification.

Superior quality oligonucleotides obtained from Eurogentec were further purified by ion-pairing reverse phase chromatography to achieve purity better than 95 %. Oligonucleotides (0.1 and 5 nmoles for analytical and preparative run respectively) were loaded onto a C18 reversed-phase column (4.6 x 100 mm 3.5 μ m Waters XTerra MS C18 column) equilibrated with 0.1 M triethylammonium acetate pH 7.4 in water. Elution was performed at 1 ml/min using a linear gradient between the aqueous buffer and acetonitrile (0 to 21.4 % by vol. in 10 min). For preparative runs, collected peaks were evaporated under vacuum without heating (concentrator 5301, Eppendorf). Purified oligonucleotides were suspended in milliQ water and stored frozen (-20°C) until use. Residual salts were removed when required by gel filtration on Sephadex G25 (Pharmacia) before use. All oligonucleotides used in constructions exhibited a single symmetrical peak in analytical runs.

Supplemental protocol 2: AFM microscopy.

AA'-, BB'-, ba'- and ab'-tiles (1 μ M each) were polymerized at 24°C for 24 h in the presence of the four terminating tiles at a concentration of 0.2 μ M. Samples were dialyzed against water using dialysis unit before deposition on MgCl₂ pretreated mica. Imaging was performed in aqueous media in contact or oscillating contact mode using oxide sharpened silicon probes exhibiting spring constants between 0.01 and 0.06 N/m. The atomic force microscope used was a Nanoscope III (Veeco, Santa Barbara, CA) with low noise performance.

Table S1: Description of parameters for simulations.

Abbreviation	Type/ Unit	
LDHRC	Flag	Set when the length dependence of hybridization rate constants was used ^a .
RL	base	Reference length for which LDHRC correction is equal to 1.
JF	Option	Type of J-factor table. Options: None (no cyclization); Standard (see table PT2); Custom defined.
USI	Flag	Flag for the presence of non-linear species ^b .
SIC	μM	Critical concentration for the formation of non-linear species ^b .

^a For each reacting species, an equivalent length was first calculated by adding the length of the ds-part and the length of the ss-part multiplied by the SSEH factor. Length corrected rate constant was calculated as

$$K = \frac{K_a + K_b}{2} \quad \text{with} \quad K_a = K_a^{RL} \times \left(\frac{R_l}{L_a} \right)^{0.58} \quad \text{and} \quad K_b = K_b^{RL} \times \left(\frac{R_l}{L_b} \right)^{0.58}$$

K_a^{RL} is the hybridization rate constant for complementary ss-DNA of equivalent length R_l having the same sequences than the hybridizing parts of reacting species. L_a and L_b are the equivalent lengths of reacting species.

^b Correction of simulation for the formation of non-linear structures was performed by substituting real concentrations of species in differential equations by the “free” concentrations calculated as follows: first total concentration c_t of all species with equivalent ds-lengths equal or larger to 78 bp was calculated. Free concentration of each species was calculated as $c_{\text{free}} = c_{\text{real}} / (1 + c_t / \text{SIC})$.

Table S2: Description of parameters for visualization of synthetic gels.

Abbreviation	Type/ Unit	
SSEH	Ratio	Ratio between the equivalent and real lengths of ss-DNA.
DSEQ	Flag	When set, sizes of polymers are calculated using the DSDS ratio.
DSSS	Ratio	Ratio between the lengths of ds- and ss-DNA featuring the same electrophoretic mobility.
DSDS	Ratio	Ratio between the lengths of a ds-DNA and of a block copolymer featuring the same electrophoretic mobility.
DSCS	Ratio	Ratio between the electrophoretic mobilities of linear and circular ds-DNAs.
SSCOL	Ratio	Ratio between the fluorescence signals of ethidium bromide stained ds- and ss-DNA of equal lengths.

Table S3: parameter sets for simulation.

Figure	Curve or Panel	Model	Tile number	Simulation parameters (SP)	Visualization parameters (VP)
2	P1	1	25	SP 1	-
2	P2	2	8	SP 2	-
4	A	1	25	SP 1	-
4	B	2	4	SP 2	-
SF1	A	1	25	SP1	-
SF1	B	1	25	SP 3	-
SF1	C	1	25	SP 4	-
SF2	all	2	4	SP 2*	-
SF3	all	1	30	SP 12	-
6	A (E1)	2	4	experimental	experimental
6	A (E2)	3	2	experimental	experimental
6	B (E1)	2	4	experimental	experimental
6	B (M1)	2	4	SP 5	-
6	B (M2)	2	4	SP 6	-
6	C (E2)	3	2	experimental	experimental
6	C (M1)	3	2	SP 7	-
6	C (M2)	3	2	SP 8	-
8	A (1)	2	8	experimental	experimental
8	A (2)	2	8	SP 9	VP
8	A (3)	2	8	SP 10	VP
8	B (1)	2	8	experimental	experimental
8	B (2)	2	8	SP 9	VP
8	B (3)	2	8	SP 11	VP

SP 2*: SP2 rate constants but see legend of supplemental Figure S2 for tile concentrations
See sub-tables SP, PT1, PT2, PT3 and VP for table references.

Table S3 (continuation): sub-table SP

	Concentration (μM)	Rate constant ($\mu\text{M}^{-1} \cdot \text{min}^{-1}$)	J-Factor table	Length dependant rate modulation	Formation of non-linear structures
SP1	1 μM (all tiles)	5 (all const.)	no	no	no
SP2	1 μM (all tiles)	4 (all const.)	no	no	no
SP 3	Note 1	5 (all const.)	no	no	no
SP 4	Note 2	5 (all const.)	no	no	no
SP 5	0.39 μM (all)	PT 1, set 1	PT 2	Yes (RL =19)	no
SP 6	0.39 μM (all)	PT 1, set 2	PT 2	Yes (RL =19)	no
SP 7	0.35 μM (all)	PT 1, set 1	PT 2	Yes (RL =19)	no
SP 8	0.35 μM (all)	PT 1, set 2	PT 2	Yes (RL =19)	no
SP 9	PT 3	PT 1, set 1	PT 2	Yes (RL =19)	no

SP 10	PT 3	PT 1, set 2	PT 2	Yes (RL =19)	no
SP 11	PT 3	PT 1, set 1	PT 2	Yes (RL =19)	yes (SIC= 0.18 μM)
SP 12	Note 3	Note 4	no	Yes (RL =19)	no

Note 1: all tiles 1 μM except forward F5 tile: 1.2 μM

Note 2: Randomized equimolar stoichiometries (1 +/- 0.05 μM) of all tiles F1-F12 and R0-R12 were used.

Note 3: Randomized equimolar stoichiometries (1 +/- 0.05 μM) of non variant tiles F1-F3, F6, F8-F12, R0-R12 and variant tiles F4, F4v1, F5, F5v1 (0.5 +/-0.025) and F7, F7v1, F7v2, F7v3 (0.25 +/- 0.0125) were used.

Note 4: Hybridization rate constants between the non-variant parts of tiles were taken as 4 $\mu\text{M}^{-1} \cdot \text{min}^{-1}$. This rate was taken as 1 $\mu\text{M}^{-1} \cdot \text{min}^{-1}$ for hybridization between variant and non-variant parts of tiles.

Table S3 (continuation): sub-table PT 1

Rate constants ($\mu\text{M}^{-1} \cdot \text{min}^{-1}$)	Set 1	Set 2
A'/a'	7.0	7.0
A/a	4.4	4.4
B'/b'	4.1	4.1
B/b	4.3	4.3
A'/a' ds-ds DNA	7.0	11.1
A/a ds-ds DNA	4.4	6.9
B'/b' ds-ds DNA	4.1	6.5
B/b ds-ds DNA	4.3	6.8

Table S3 (continuation): sub-table PT 2

Size (equivalent db)	J-factor (nM)
≤ 149	0
150	0.05
200	1
250	8
300	20
350	40
400	70
600	80
1000	70
≥ 2000	60

Standard J-factor values used in simulations were calculated by logarithmic interpolation of the following table. Equivalent sizes were calculated as the sum of the lengths of the ds-parts and half of the lengths of ss-parts.

Table S3 (continuation): sub-table PT 3

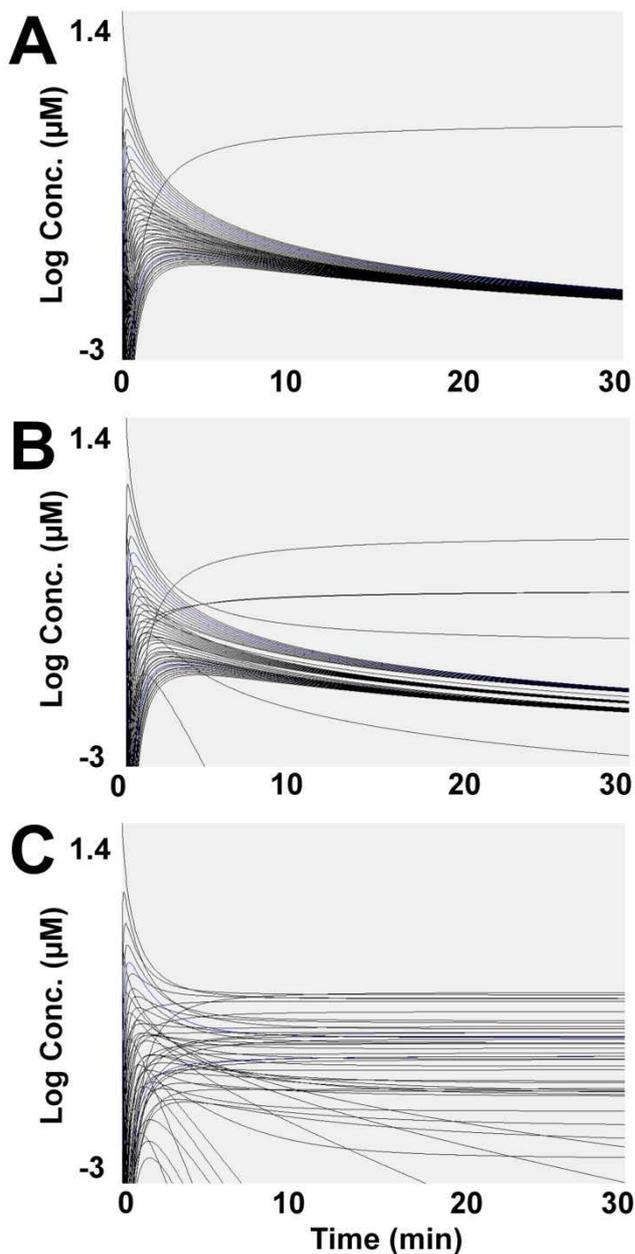
	Concentrations (μM)
Polymerizing tiles	
AA'	0.96
BB'	0.92
ab'	0.95
ba'	0.85
Terminating tiles	
Ma'	1.54
Mb'	1.5
am	1.74
bm	1.38

Table S3 (continuation): sub-table VP

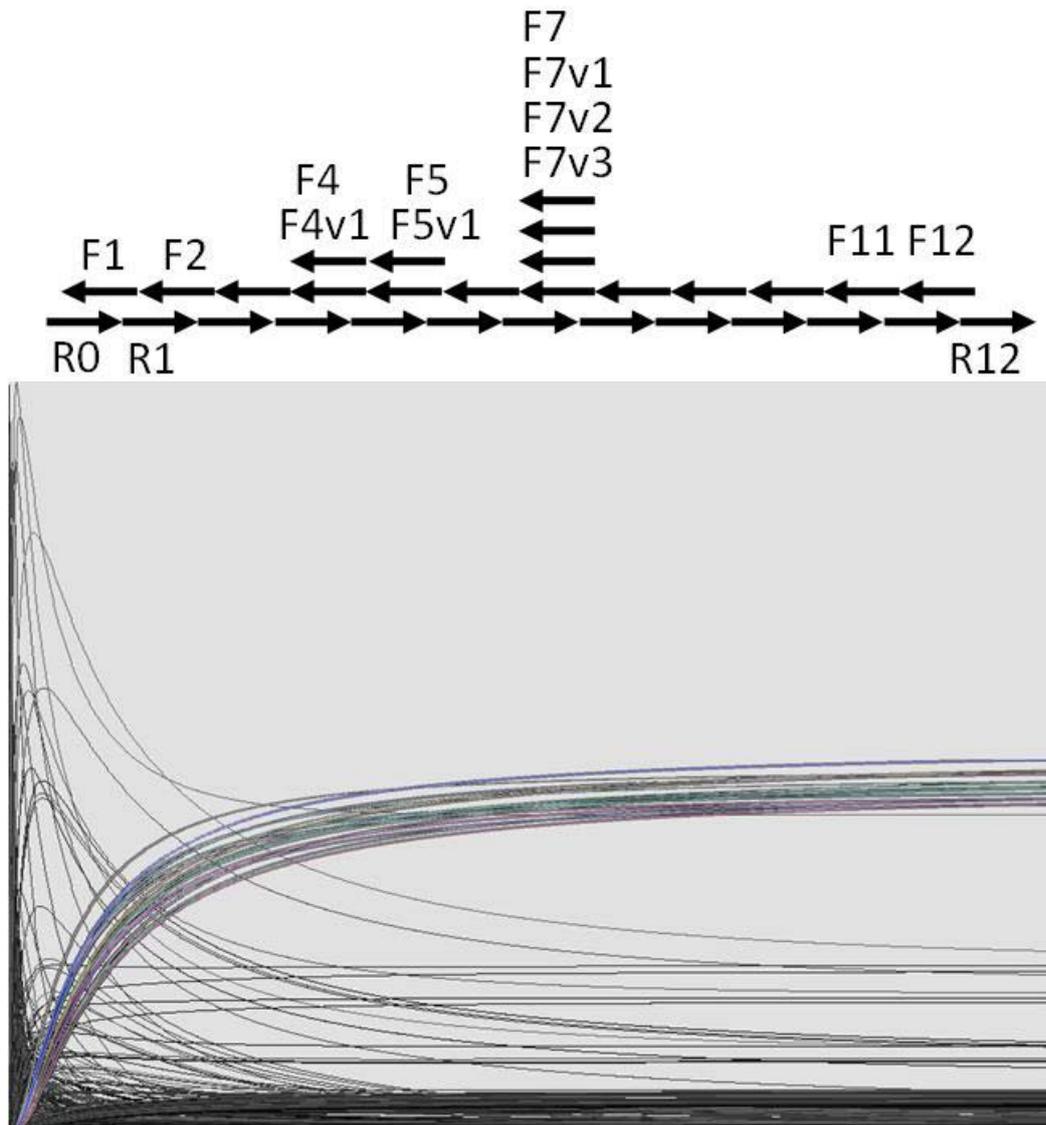
Abbreviation of parameter	
SSEH	0.5
DSEQ	yes
DSSS	0.83
DSDS	1.05
DSCS	1.3

See Table S2 for parameter definition.

Supplemental Figure S1. Simulated time course of the species concentration for model 1 (gene synthesis) using different tile stoichiometry. A, Equimolar stoichiometries ($1 \mu\text{M}$) of all tiles were used. Hybridization rate constants between ss-tiles were taken as $5 \mu\text{M}^{-1} \cdot \text{min}^{-1}$. B, Same conditions than in panel A excepted that concentration for the forward tile 5th was $1.2 \mu\text{M}$. C, Same conditions than in panel A excepted that concentration of tiles are randomly distributed between 0.8 and $1.2 \mu\text{M}$. The trace for the fully assembled gene sequence is given in bold. Other parameters are given in the supplementary data.

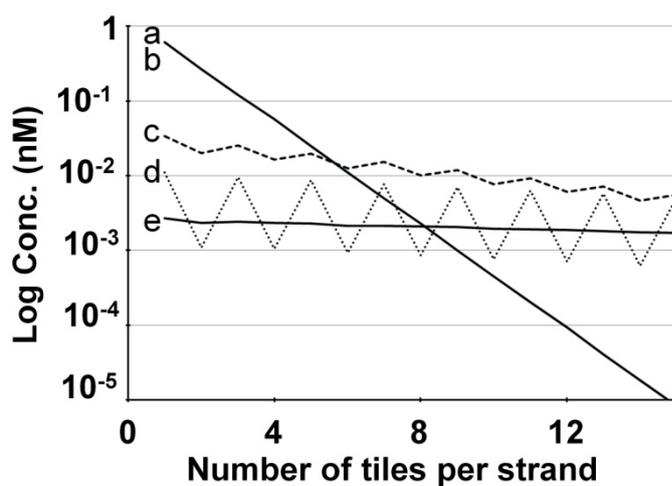


Supplemental Figure S2. Simulated time course of the species concentration for model 1 (gene library synthesis). Randomized equimolar stoichiometries ($1 \pm 0.05 \mu\text{M}$) of non variant tiles F1-F3, F6, F8-F12, R0-R12 and variant tiles F4, F4v1, F5, F5v1 (0.5 ± 0.025) and F7, F7v1, F7v2, F7v3 (0.25 ± 0.0125) were used. Hybridization rate constants between the non-variant parts of tiles were taken as $4 \mu\text{M}^{-1} \cdot \text{min}^{-1}$. This rate was taken as $1 \mu\text{M}^{-1} \cdot \text{min}^{-1}$ for hybridization between variant and non-variant parts of tiles. Parameters are given in Table S3. The 16 possible variant end-products are shown in bold colored lines.

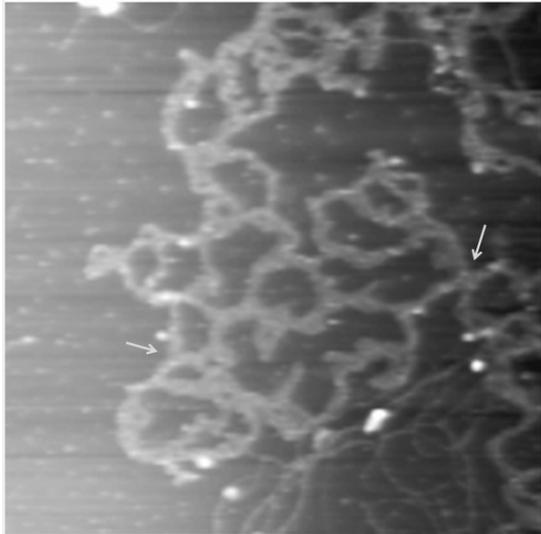


Full x scale: 300 min; full y scale: $0.08 \mu\text{M}$.

Supplemental Figure S3. Simulation of the polymerization of tiles in Model 2. Non-stoichiometric (b, c, d curves) and stoichiometric conditions (a and e) are compared. Species concentrations are shown as a function of polymerization duration and species size. Curves a, e and b, c, d correspond to a 1:1:1:1 and 1.05:0.95:1:1 molar ratio of tiles, respectively. Incubation time were 0.5, 4 and 15 min for curves (a, b), c and (d, e), respectively. Simulation parameters are given in Table S3.



Supplemental Figure S4. AFM imaging of mesh-like structures formed upon DNA assembly using model 2. Atomic force microscopy imaging of species resulting from the copolymerization of four tiles ($1\ \mu\text{M}$ each) in the presence of terminating tiles ($0.2\ \mu\text{M}$ each).



300 x 300 nm

3. Données complémentaires

La question de la caractérisation de la distribution de taille des polymères dans l'article a été principalement traitée par des méthodes à base de gels d'électrophorèse. Nous présentons ici les expériences complémentaires de caractérisation sur gel, en cinétique rapide et en compétition ainsi que les données recueillies par la technique de microscopie de force atomique. Cette dernière approche de caractérisation au niveau de la molécule unique, apporte des informations complémentaires sur la structure des polymères et leur distribution de taille.

3.1. Analyse de l'état de polymérisation par électrophorèse

3.1.1 Méthodes

Système expérimental : Des collections de polymères sont formées par l'hybridation de quatre ADN simple-brin de 39 bases mélangés en conditions stœchiométriques (1 μ M) dans du tampon 10 mM phosphate, 125 mM NaCl pH 7,4, à 24 ou à 42 °C. La polymérisation est contrôlée par l'ajout de concentrations variables d'un mélange équimolaire de quatre séquences de terminaison capables de bloquer l'élongation des polymères (Figure 37).

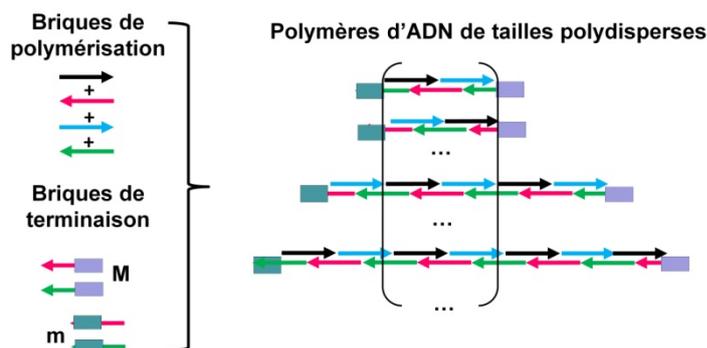


Figure 37 : Schéma de polymérisation de briques d'ADN simple brin en présence de briques de terminaisons. Les ADN simple brin sont indiqués par des flèches, les ADN briques de terminaison sont représentées par des flèches avec bloc. Les parenthèses encadrent l'unité de polymère la plus simple se formant à partir des quatre briques.

Gels : Des gels (agarose et acrylamide) ont été réalisés pour caractériser la réaction de polymérisation des briques d'ADN à des temps, des concentrations et températures différentes.

Les gels d'agarose (Nusieve GTG) à 3 et 3,5% dans un tampon Tris-borate-EDTA 1X permettent de discriminer des acides nucléiques double-brin de 70 à 300 paires de bases et de 100 à 450 pb respectivement. Un marqueur de taille appelé Low Molecular Weight DNA Ladder a été utilisé pour référence. Les gels polyacrylamide 4-12% (Bis-Tris gel, Nupage) ont également été utilisés, les échantillons d'ADN étant révélés par l'intercalant de bromure d'éthidium avec les deux marqueurs en référence : LMW et Mark 12 (Invitrogen), ce dernier étant coloré ultérieurement par le bleu de Coomassie (Figure 38).

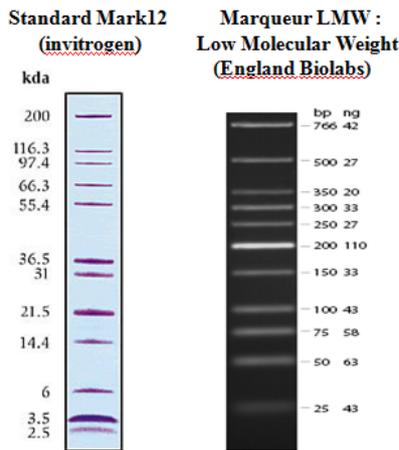


Figure 38 : Marqueurs de taille.

A gauche, le marqueur de taille Mark 12 réfère est une référence de la taille des protéines en kda. Il est révélatrice par coloration au bleu de Coomassie ; A droite, le marqueur LMW est une référence de taille des ADN double brin, révélatrice en bromure d'éthidium.

Ligation des échantillons : Une ligation enzymatique a été réalisée sur des aliquotes (1 μ l) de chaque échantillon polymérisé incubé en présence de T4 DNA ligase dans le tampon de ligation selon les indications du fournisseur. Un protocole de ligation chimique utilisant le bromure de cyanogène dans du tampon MES (Tris, 2-morpholinoethane sulfonate) 250 mM supplémenté avec 20 mM de $MgCl_2$ a également été employé [381].

Amplification des échantillons par réaction en chaîne à la polymérase (PCR) : Les échantillons ligués dilués 100 fois ont été amplifiés par PCR à l'aide des amorces M13R et M13F semi-complémentaires des briques de terminaison M13R-b', M13R-a', a-m13f et b-m13f des polymères, dans un volume de réaction de 50 μ l en utilisant la Taq DNA polymérase. 30 cycles ont été réalisés, comprenant l'étape de dénaturation à 94 °C pendant 30 secondes, l'étape de température d'hybridation à 55 °C pendant 30 secondes et l'étape d'extension à 72 °C pendant 1 minute. Un cycle final d'extension à 72 °C pendant 7 minutes permet de terminer les chaînes en cours d'élongation.

Méthode de caractérisation indirecte par détection chimio-luminescente sur les échantillons amplifiés par PCR : Les échantillons ligués ont été amplifiés de manière linéaire en utilisant 5 cycles de PCR (94°C pendant 30s, 55°C pendant 2 min et 72°C pendant 5 min) avec 0.5 U d'enzyme en présence d'une seule amorce biotynilée (utilisation d'une méthode de détection par chimioluminescence, kit Bright-Start BioDetect, Ambion). Du mélange de polymères correspondant à 10-100 fmoles de briques a été utilisé comme matrice pour l'amplification avec un rapport molaire entre la matrice et l'amorce de 200. Les produits amplifiés ont été mis à sec puis dissous dans une solution formamide : eau (98 :2 vol/vol) contenant 10 mM d'EDTA avec comme marqueur du xylène cyanol. Les échantillons ont été chargés sur des gels de séparation dénaturants en polyacrylamide 4% contenant une concentration élevée d'urée (7M) qui permet de résoudre sur gel d'acrylamide de courts (< 500 bases) fragments d'ADN simple brin. L'électrophorèse a été réalisée pendant 1 heure à

une puissance constante de 10W. Les échantillons sont ensuite électro-transférés sur une membrane de nylon chargée positivement, laquelle subit différentes étapes de lavage avant incubation avec une streptavidine alcaline phosphatase. Celle-ci conjuguée à l'oligonucléotide biotynilé des échantillons sera révélée par un réactif chimio-luminescent (CDP-star) qui permet une détection rapide et sensible de molécules en produisant de la lumière visible. L'émission de lumière a été enregistrée en utilisant un système d'imagerie avec une caméra CCD refroidie (LAS-3000 imager).

Simulation : Le logiciel présenté dans l'article a été utilisé afin de simuler diverses conditions de polymérisation. Il a permis, entre autres, d'analyser l'influence de l'ajout de briques de terminaison lors de l'hybridation des briques de construction. Les profils de distribution de taille des polymères attendus générés apparaissent sous forme de scans directement comparables avec les données expérimentales.

3.1.1 Résultats des analyses sur gel

Les résultats d'analyse sur gels des cinétiques d'assemblage des briques d'ADN sont cohérents avec les simulations pour des temps courts de polymérisation, mais s'en écartent pour des temps d'incubation plus élevés (au-delà d'une minute). Les données sur gels restent également en adéquation avec les données de cinétique rapide pour les faibles degrés de polymérisation. Lorsque de longs polymères d'ADN segmentés sont attendus, on assiste au contraire à une divergence à la fois entre données expérimentales (gel et cinétique rapide) et entre données expérimentales (gels) et simulations. Parallèlement, la microscopie de force atomique met en évidence la formation de supra-structures non linéaires qui ne sont pas prises en compte par la modélisation. Les questions qui se posent alors sont, d'une part de comprendre quels mécanismes sont impliqués dans la formation de structures non linéaires, d'autre part d'évaluer les domaines de pertinences des différentes méthodes expérimentales mises en œuvre.

Limites des techniques d'analyse sur gel des profils de polymères.

- Lors des analyses, les profils sur gel d'agarose apparaissent flous, avec quelques bandes discrètes qui se discernent au sein d'un *continuum* diffus. On peut exclure un problème de qualité des gels puisque les marqueurs de taille, ou des produits de PCR ayant des distributions de taille similaires aux polymères, présentent des bandes bien nettes et peu ou pas de continuum. Le caractère diffus des gels est donc intrinsèque aux espèces déposées qui apparaissent comme disperses au niveau de leurs propriétés de migration alors que leurs tailles sont obligatoirement des multiples de la taille des briques de base. Ceci est vrai aussi bien pour les gels d'agarose que pour les gels acrylamide (Figure 39A et B) et suggère de fortes interactions entre espèces, ou entre espèces et gel, lors de la migration. En effet, les polymères engendrés à partir des briques d'ADN sont des polymères segmentés toutes les 19-20 bases, dont les extrémités simple brin libres sont capables d'interagir. Ceci est vrai, y

compris en présence des briques de terminaison, qui présentent elles-mêmes une extension simple brin libre, non complémentaire du groupe de briques de polymérisation.

- Nous n'avons pas observé sur les gels de formes très lourdes alors que l'on s'attendrait (simple logique et simulations) à des polymères de très grande taille pour des temps longs d'incubation. A l'inverse, du matériel de relativement grande taille est observé même en présence de terminateurs (ligne 4 des gels A et B) qui auraient dû limiter la polymérisation à quelques bandes n'excédant pas 300 paires de bases. Les simulations réalisées en présence de quantités croissantes de briques de terminaison présentent une distribution de taille plus étroite, constituée majoritairement de polymères courts, à mesure que la concentration des terminateurs est élevée (Figure 40).

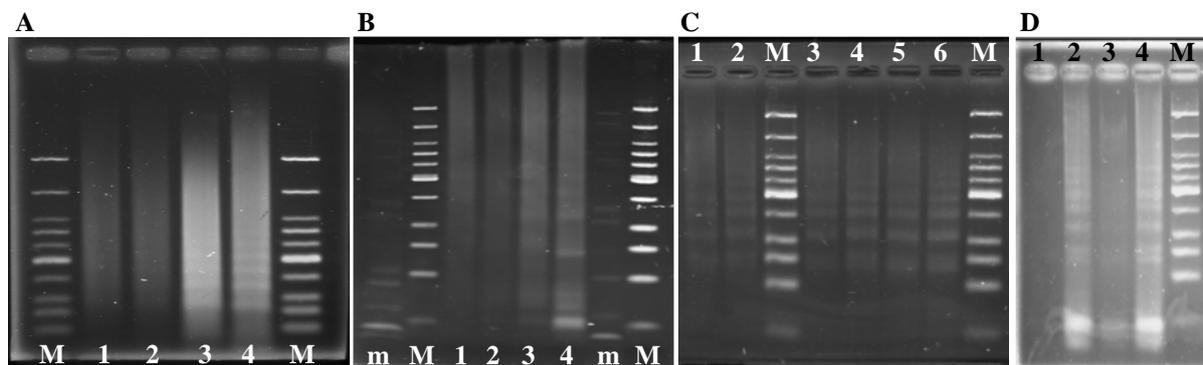


Figure 39 : Caractérisation des polymères d'ADN par électrophorèse.

A et B, comparaison sur gel d'agarose nusieve 3,5 % TBE 0.5X (A) et gel acrylamide 4-12% Bis-Tris (B) des mêmes échantillons d'ADN polymérisés à partir des quatre briques de construction en absence ou en présence de briques de terminaison. 1, produits d'hybridation sans briques de terminaison. 2, produits d'hybridation avec 1/10^{ème} de chaque brique de terminaison. 3 produits d'hybridation avec 1/3 de chaque brique de terminaison. 4, produits d'hybridation avec 1/2 de chaque brique de terminaison. C, caractérisation sur gel d'agarose de briques d'ADN polymérisées en présence de 1/5 (lignes 1, 2), 1/3 (lignes 3, 4) et 1/2 (lignes 5, 6) de briques de terminaison. Chaque 1^{ère} ligne correspond à l'échantillon ayant subi une étape de ligation (lignes 1, 3, 5) et chaque 2^{ème} ligne correspond à l'échantillon non ligué (lignes 2, 4, 6). Un facteur de dilution d'environ 4 est présent entre les échantillons ligués et non ligués. D, comparaison sur gel d'agarose 3,5 % de l'assemblage des quatre briques de construction d'ADN polymérisées à 24 et 42 °C. Les lignes 1 et 3 correspondent aux échantillons ligués à 24 et 42 °C, respectivement ; les lignes 2 et 4 correspondent aux échantillons non ligués à 24 et 42 °C, respectivement.

Simulation de la migration sur gel d'agarose des produits de polymérisation des quatre briques d'ADN en présence d'une concentration croissante des briques de terminaison.

Afin de comparer les résultats obtenus sur gels avec les résultats théoriques attendus, nous avons simulé la polymérisation des briques d'ADN dans des conditions identiques aux conditions expérimentales. Le logiciel développé au laboratoire permet de reconstituer les données simulées sous forme de profil d'intensité de fluorescence tels que présenté en Figure 40. Ceci est rendu possible par la modélisation du profil de migration des ADN simple et double brin, à partir de la migration de marqueurs connus, sur les gels. Ainsi, le logiciel génère des figures d'électrophorèse virtuelles réalistes pouvant être comparées aux profils expérimentaux. Cette comparaison permet de mettre facilement en évidence les divergences.

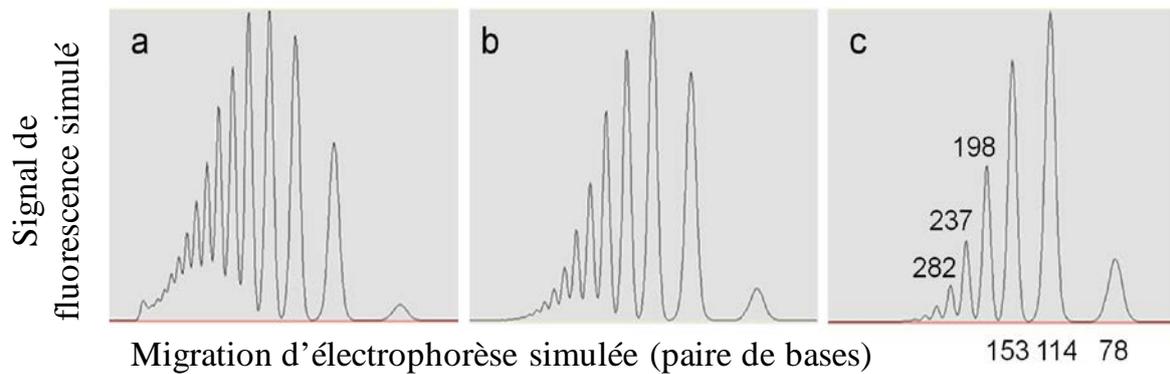


Figure 40 : Simulations de la migration sur gel d'agarose de la copolymérisation des quatre briques d'ADN en présence d'une concentration variable du mélange de terminateurs.

La concentration des briques de construction équivalente à 0.96-0.92-0.85-0.95 μM a été choisie en adéquation avec les valeurs expérimentales. La concentration des briques de terminaison va croissante avec a, 0.2 μM , b, 0.3 μM et c, 0.5 μM chaque terminateur. Les simulations prennent en compte la diffusion des molécules en fonction de leur longueur.

Interférence potentielle des cassures et des segments simple brin avec la migration sur les gels

H. Kuhn *et al.* ont montré que la mobilité de polymères d'ADN présentant des cassures dans le squelette tous les 23 paires de bases était réduite, empêchant une comparaison directe avec les marqueurs de poids moléculaire [382]. Dans notre cas, les calibrations n'ont pas montré d'écart significatif. Toutefois, les copolymères d'ADN, segmentés environ tous les 6 nanomètres, présentent vraisemblablement une longueur de persistance intermédiaire entre celle des ADN simple brin de quelques nm et celle des ADN double brin d'environ 50 nm [325, 334, 337]. Ceci peut être à l'origine de conformations multiples et de problèmes topologiques (emmêlement entre structures ou dans les mailles du gel).

Pour tenter de résoudre cette question, une étape de ligation a été introduite sur une partie des échantillons après leur polymérisation. Les résultats obtenus après ligation enzymatique n'ont pas été concluants, car des profils similaires ont été observés pour les assemblages ligués et non ligués (Figure 39C). Une première hypothèse est que la ligation n'a pas fonctionné du fait des contraintes topologiques des structures. Ces contraintes géométriques introduiraient des restrictions de rotation au niveau des cassures simple brin qui empêcheraient un positionnement correct des extrémités donneur et accepteur dans la ligation. Il est à noter que dans le cas d'assemblages de structures hybrides impliquant des protéines, des contraintes stériques pourront s'ajouter au phénomène précédent. Afin de contourner les difficultés rencontrées avec la ligation enzymatique, nous nous sommes appuyés sur les travaux de différentes équipes [381, 383-388] pour tenter de réaliser une ligation chimique catalysée par du bromure de cyanogène. Dans nos conditions expérimentales, la ligation chimique n'a pas, non plus, permis d'obtenir de meilleurs profils de gels. N.G. Dolinnaya *et al.* précisent que le degré d'exposition des groupes phosphate activés avec l'agent couplant (Bromure de cyanogène ou carbodiimidazole) est cruciale pour l'efficacité de la réaction,

laquelle dépend essentiellement de la structure de la double hélice au niveau de la coupure simple brin.

L'autre hypothèse met plutôt en cause la présence de segments simple brin pouvant interagir entre eux de manière non spécifique. La résolution de ce problème est complexe car les extensions simples brins jouent un rôle essentiel dans le processus de polymérisation. Ces extensions pourraient être bloquées en fin de réaction par hybridation avec des séquences complémentaires mais au prix d'étapes supplémentaires rendant le modèle très complexe et sans que l'effet soit garanti dans le cas de formation d'espèces réticulées.

Une approche alternative a alors été menée pour supprimer les extensions simple brin par amplification PCR afin d'obtenir des fragments double brin non segmentés qui migreraient correctement. Pour ce faire, les copolymères terminés par les briques M13R-b', M13R-a', a-m13f et b-m13f qui incluent les amorces d'amplification ont été amplifiés. Après dépôt sur gel, des bandes discrètes ont été observées. Néanmoins les résultats n'ont pu être exploités car des expériences contrôle ont mis en évidence des phénomènes de recombinaison entre les séquences répétées du polymère lors des cycles de PCR. En effet, après avoir découpé trois bandes de différentes tailles des produits de PCR sur le gel (150-200, 300-350 et 766 paires de bases), nous avons à nouveau amplifié séparément les trois sources de matériel. Après dépôt sur gels, les profils obtenus étaient similaires avec toute la gamme des bandes discrètes alors que les matrices étaient de taille homogène et différente.

Une dernière approche a alors consisté à réduire les concentrations utilisées lors de l'analyse en basculant sur des gels de polyacrylamide permettant d'utiliser un marquage froid avec amplification enzymatique plus sensible que la fluorescence. Toutefois, cette approche indirecte s'est révélée difficile à utiliser de manière quantitative même si la résolution des gels est apparue meilleure.

Présence potentielle de pièges cinétiques

Afin de répondre à la question d'un potentiel blocage lors de la polymérisation des briques d'ADN, nous avons réalisé une expérience en l'absence de briques de terminaison à 24 et 42 °C. Cette dernière température doit en effet permettre de débloquent d'éventuelles formes abortives et favoriser l'élongation des polymères. Les échantillons ligués et non ligués ont à nouveau été comparés aux deux températures (Figure 39D). Toutefois, les profils similaires observés dans toutes les conditions ne permettent pas de conclure. D'autres hypothèses ont alors été considérées, comme la formation de structures en nœuds pouvant résulter soit de phénomènes d'invasion de brin, soit de mécanisme de double nucléation. En effet, aux concentrations relativement élevées utilisées, les événements de nucléation pourraient s'avérer relativement rapides en comparaison des phénomènes de propagation, lorsque ces derniers sont ralentis par des structures secondaires relativement stables au niveau des réactants. Dans ces conditions, on ne peut exclure l'existence de double initiation

d'hybridation générant *de facto* une structure branchée qui ne se résoudra pas forcément par migration de brins. Une expérience de compétition entre des ADN seuls et des ADN couplés à des protéines pour une même cible avait été réalisée initialement afin d'estimer l'influence de la partie protéique sur les vitesses d'hybridation. Par la suite, ce même paramètre a été analysé différemment par une approche directe de mesure séparée des vitesses d'hybridation des ADN et des PDNA. De manière surprenante, des divergences excédant largement le domaine d'erreur ont été observées entre les deux approches. En fait, la première approche mesure des rapports de vitesse à l'issue de la compétition DNA/PDNA sans prendre en compte les événements d'initiation et de vitesse de propagation. La seconde approche privilégie, au contraire, ces deux mécanismes mais est imprécise sur les événements tardifs possibles (toujours pour des raisons d'incertitude sur la valeur exacte de l'hypochromicité). L'existence de divergences pointe sur la possibilité d'événements possibles de multi-nucléations, dont la résolution thermodynamique par migration de brins conduirait à des rapports apparents de vitesse différents d'une résolution sous contrôle purement cinétique (rapport de constantes de vitesse).

En conclusion, les analyses sur gels ne sont en accord avec les simulations ou les données de cinétique rapide que pour un degré polymérisation limité et donc en l'absence de longs polymères. Le problème semble multifactoriel, résultant à la fois d'artefacts analytiques liés à l'interaction des espèces au sein de la population et de formation de structures anormales jouant un rôle de terminaison dans le processus d'élongation. Les techniques de caractérisation sur gel ne permettant pas d'aller plus loin et les approches spectroscopiques par mélange rapide manquant de précision dans les stades tardifs de polymérisation, nous sommes tournés vers une caractérisation structurale par microscopie de force atomique.

3.2. Analyse de l'état de polymérisation par microscopie de force atomique (AFM)

3.2.1 Méthodes

L'observation nécessite le dépôt des échantillons sur une surface de mica préalablement traitée avec des cations, qui conditionnent l'adhésion des échantillons sur la surface. Une incubation de quelques minutes est suivie d'un lavage extensif à l'eau afin de retirer l'excès de matériel non adsorbé et surtout les sels.

Dans le cas d'échantillons d'ADN, la force ionique joue un rôle important aussi bien au niveau de la stabilité des molécules qu'au niveau de leur interaction avec la surface. La force ionique doit être la plus faible possible afin de minimiser la présence de sels qui cristallisent au cours du séchage formant une gangue autour des objets, ce qui altère l'observation, toutefois, elle doit être suffisante pour ne pas déstructurer les échantillons. Une étape de dialyse, contre de l'eau ou des sels d'acétate d'ammonium (10 mM, pH7) qui sont

volatils, permet de dessaler les échantillons préalablement préparés dans des tampons contenant 100 à 300 mM de sels.

Nous avons déposé sur mica différents échantillons afin de comparer la distribution de taille des copolymères d'ADN assemblés à partir des quatre briques de polymérisation (1 μM) en l'absence et en présence de 0.2, 0.33 et 0.5 μM (chaque) du mélange de briques de terminaison. L'assemblage des ADN a été réalisé de la même façon que présentée dans la section (méthodes du paragraphe 3.1) afin d'analyser les mêmes populations d'échantillons par les deux techniques. Le mica clivé a été traité par une solution contenant 1 mM de chlorure de nickel ou de chlorure de magnésium. Après quelques minutes d'incubation, les substrats rincés à l'eau milliQ puis séchés sont prêts pour le dépôt des échantillons eux-mêmes incubés une dizaine de minutes avant rinçage à l'eau et séchage pour l'observation à l'air.

Seules quelques images exploitables ont pu être obtenues à partir des copolymères d'ADN assemblés en présence de 0.5 μM de terminateurs, nous permettant d'évaluer l'effet d'une étape de ligation. La distribution de taille a été analysée en utilisant le logiciel « Image j » qui permet de mesurer la taille de particules. Ainsi, les objets sur les images AFM ont d'abord été retracés au crayon sur un calque afin de prendre en compte uniquement leur longueur et de s'affranchir de la gangue des sels gainant les structures (Figure 41). L'image correspondante a été convertie par seuillage puis la distribution de surface des objets calculée a été convertie dans Excel en classes de longueurs après calibration. Les classes de tailles utilisées correspondent à un pas de 12 nm équivalent à la taille des briques d'ADN de 39 bases. Les graphes ont été normalisés entre expériences pour compenser le nombre d'échantillons observés qui varie d'une expérience à l'autre.

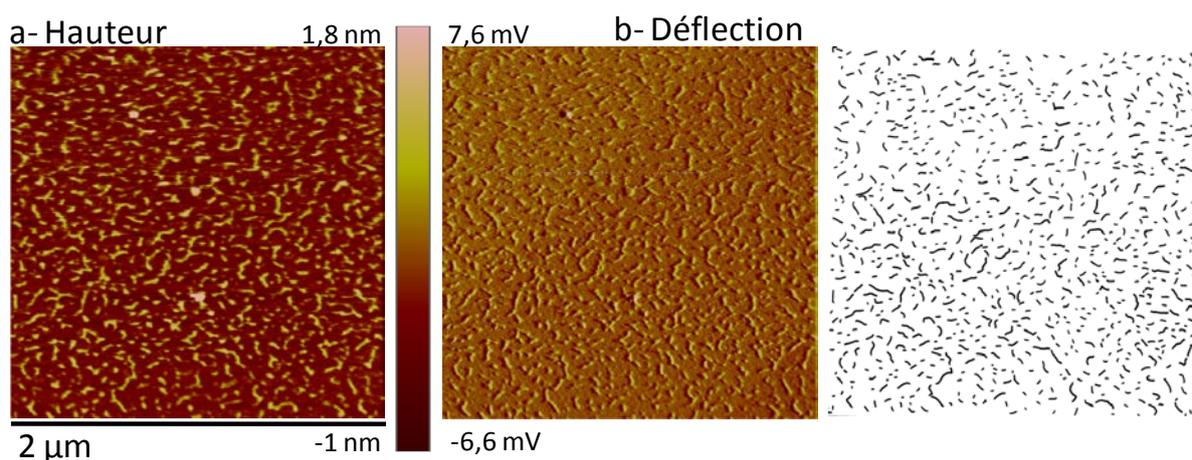


Figure 41 : Méthode d'analyse des images AFM
a, image AFM en hauteur, b, image en déflexion, c, seuillage de l'image a.

3.2.2 Résultats

Les images AFM obtenues illustrent soit des structures d'ADN isolées (ADN 1 μM et terminateurs 0.5 μM) soit des réseaux continus (ADN 1 μM et terminateurs 0.2 μM , figure S4 de l'article). Nous n'avons pas observé d'espèces cycliques par cette technique, confirmant les résultats obtenus sur gels. Les images AFM ont été utilisées pour calculer la distribution de taille des copolymères d'ADN assemblés en présence de séquences de terminaison non ligués et ligués. La distribution obtenue sur la Figure 42 est équivalente à une distribution de concentration molaire (nombre de molécules/surface).

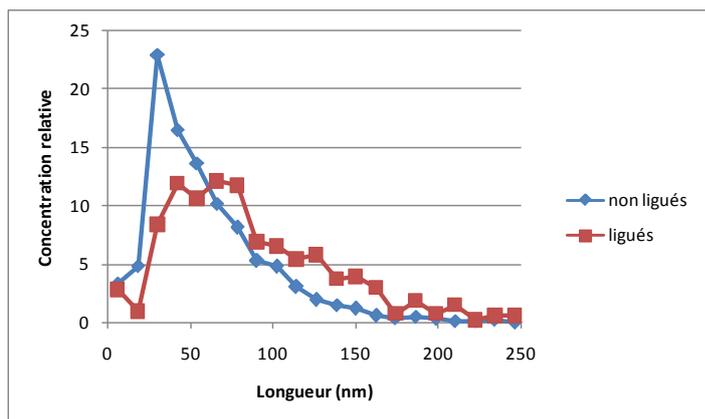


Figure 42 : distribution de la concentration relative des objets en fonction de leur taille

En multipliant les données de ce premier graphe par le nombre de paires de bases, on obtient une représentation en masse qui est équivalente à ce qu'on observerait sur gel après coloration au bromure d'éthidium.

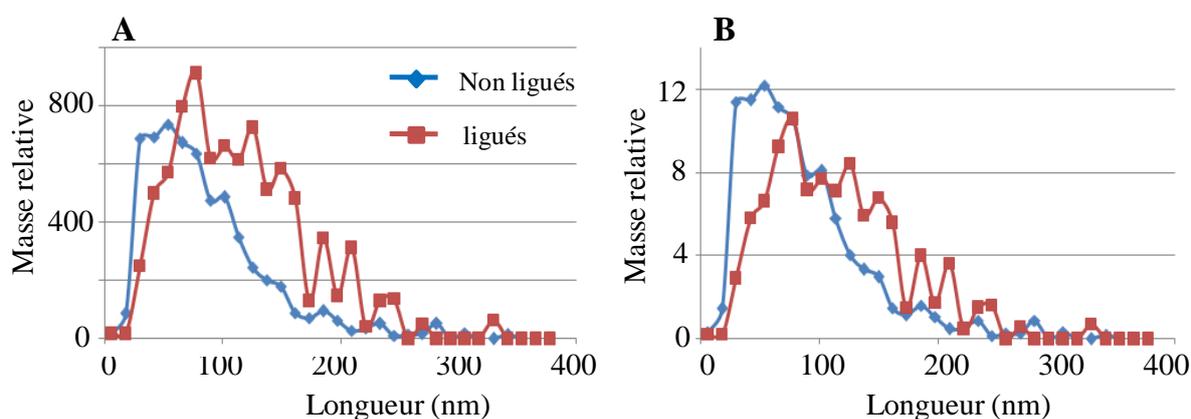


Figure 43 : distribution de la masse relative des objets en fonction de leur taille. A, Normalisation sur le nombre d'objets, B, normalisation sur la masse.

Un léger déplacement de la distribution de taille vers des longueurs plus élevées après ligation est observé même si le plus faible échantillonnage crée un bruit statistique sur les échantillons ligués (une seule image au lieu de trois). Il est à noter que le nombre d'objets par image est limité afin qu'ils n'interagissent pas ensemble et qu'il n'y ait pas de superposition. La deuxième limitation étant la taille de l'image qui présente une résolution d'autant plus

faible que sa taille est grande. L'application du test de Student indique une différence non significative entre la distribution de taille des échantillons non ligés et ligés. L'observation d'un léger décalage des courbes statistiquement non significatif confirme que la ligation n'a pas eu d'effet majeur.

En conclusion, l'AFM confirme l'absence d'espèces cycliques et apporte des informations complémentaires sur la distribution de taille des polymères observés à sec. Elle permet de caractériser des fragments de petites tailles en présence d'une grande concentration de terminateurs (1:2 de rapport molaire) et des réseaux comportant des points de branchements inattendus pour une concentration des terminateurs plus faible (rapport molaire de 1:5 par rapport aux briques (voir figure S4 de l'article)). Lorsque des grands polymères sont imagés, ils apparaissent sous forme de réseaux et non comme des structures indépendantes, semblant ainsi confirmer l'existence de terminaisons anormales. Néanmoins il convient d'être prudent avec cette observation. En effet, l'étape de dépôt sur la matrice peut elle-même créer des artefacts, soit du fait de phénomènes de projection dimensionnelle, soit en réorganisant les structures en solutions par des mécanismes de transfert de brin au niveau de la surface dans les conditions de faibles forces ioniques du lavage qui déstabilisent les hybrides (répulsion électrostatique entre phosphates) et de haute densité surfacique.

4. Discussions complémentaires

Deux classes de problèmes ont été rencontrées : (i) les problèmes liés aux artefacts d'observation dus aux contraintes de la technique utilisée, (ii) les problèmes intrinsèques au modèle biologique étudié.

4.1. Les différentes techniques de caractérisation : apports et limites

4.1.1 Les gels d'électrophorèse

La méthode de gels d'électrophorèse constitue une approche globale sur la distribution en taille d'une population d'échantillons donnés. Elle a soulevé différentes questions.

Disruption possible des structures : Les polymères d'ADN double brin segmentés tous les 6 nm, de nature probablement plus flexible que des fragments d'ADN double brin, sont-ils cassés lors de la migration dans les gels ? On peut effectivement penser que les conditions de force ionique faible et le pH élevé des tampons utilisés au cours des migrations puissent déstabiliser les structures.

Interactions entre molécules : De nombreux éléments suggèrent que les problèmes de caractérisation des assemblages sont dus aux interactions des polymères longs avec les structures plus courtes. Les extensions simple brin aux extrémités des polymères, quelque soit leur taille, sont certainement responsables d'interactions entre structures.

Charge des gels/ conformation des molécules : D'importantes concentrations d'ADN ont été utilisées afin de pouvoir révéler les polymères de toute taille, y compris les plus courts, le matériel d'ADN devant se répartir en plusieurs bandes. Toutefois, les problèmes de résolution ont également été observés dans les gels d'acrylamide qui nécessitent moins de matériel.

4.1.2 La microscopie de force atomique

Des échantillons auto-assemblés dans différentes conditions de concentrations de briques et de terminateurs ont été déposés sur du mica clivé fonctionnalisé ou non, et sur de la silice fonctionnalisée sur silicium. Nous n'avons pas pu obtenir d'images exploitables pour tous les échantillons ce qui souligne le caractère délicat de la méthode.

Force ionique : La préparation des échantillons est cruciale, en particulier la force ionique doit être suffisamment faible pour ne pas former d'amas salins qui interfèrent lors de l'imagerie et être en même temps suffisamment élevée pour ne pas déstabiliser les hybridations et les interactions ioniques avec le support. En ce qui concerne les micas, leur composition peut être variable d'un échantillon à l'autre et, pour un même mica, le clivage peut générer des zones de composition ionique différente au niveau de la surface. La chimie de surface appliquée peut apporter d'autres sources d'hétérogénéité.

Altération mécanique des structures : L'AFM ne pose pas les mêmes contraintes que les gels d'électrophorèse. Par exemple, le stress mécanique, même s'il n'est pas nul du fait de l'immobilisation des structures sur une surface, est bien moindre que dans les migrations sur gels. En observation AFM à sec, on rencontre plutôt des problèmes d'agrégation des structures sur la surface ainsi que des problèmes de sels qui cristallisent en enrobant les molécules d'ADN.

Echantillonnage surface-solution : Nous observons le matériel immobilisé sur la surface par les interactions ioniques entre les charges négatives des phosphates de l'ADN et les cations divalents, mais en quoi les structures immobilisées représentent-elles la composition en solution ? Un exemple de biais pourrait être le lavage : plus ou moins extensif, il peut éliminer une proportion plus ou moins importante des structures les plus courtes qui présentent une plus faible interaction avec le support.

Limite d'échantillonnage des images : Des compromis doivent être faits entre la taille de l'image, la résolution et le nombre d'objets sur la surface. Plus la taille de l'image est grande moins elle est résolutive mais plus on peut observer un nombre élevé d'objets, lesquels pourront présenter de grandes tailles. La taille des images peut constituer un biais d'observation, les grands objets pouvant être évincés de l'observation. La densité d'objets sur la surface constitue également un facteur limitant. Elle doit être suffisamment grande afin de permettre une analyse statistique sans être excessive afin que les objets restent isolés sur la surface. La superposition des objets empêcherait toute analyse de leur distribution de taille.

4.1.3 La technique de cinétique rapide

La cinétique rapide mesure un signal optique qui rend compte de l'empilement des bases au cours de l'hybridation. Cette mesure recouvre aussi bien les assemblages linéaires, les structures secondaires que les agrégats, sans permettre de discriminer les différents phénomènes.

Les données de cinétique rapide sont en excellent accord avec les données issues des simulations, y compris aux temps longs, mais divergent des données sur gels. Toutefois, la connaissance exacte des effets hypochromiques conditionnent la fiabilité de l'analyse des courbes aux temps longs. Le point délicat est la définition de l'état final qui va déterminer en retour le coefficient d'hypochromicité. L'ajustement effectué est compatible avec les limites théoriques mais reste un ajustement qui peut masquer de faibles divergences dans les phases tardives de la réaction.

4.2. Mécanismes de terminaison dans les assemblages illimités

Quels mécanismes pourraient expliquer un éventuel arrêt de polymérisation? Nous pouvons distinguer les mécanismes de terminaison volontaires par l'ajout de séquences terminatrices et les mécanismes involontaires, dus aux propriétés des structures, tels que les phénomènes de cyclisation et de branchement.

4.2.1 Mécanismes dirigés par l'ajout de terminateurs

La copolymérisation de quatre briques d'ADN semi-complémentaires génère une population de molécules dont l'étendue n'est limitée que par la disponibilité des briques de construction ou par des phénomènes de terminaison (structures branchées ou circulaires). Afin de contrôler le degré de polymérisation des assemblages, nous avons introduit dans le système des briques de terminaison capables de générer des extrémités non hybridables. Le degré de polymérisation est alors défini par le rapport molaire entre les briques qui polymérisent et les briques de terminaison.

L'arrêt de polymérisation étant initié par l'ajout d'excès de terminateurs, ceux-ci pourraient par déplacement de brins envahir les structures et les emprisonner. Certaines images AFM présentant des réseaux d'ADN corroboreraient cette hypothèse, mais pour la confirmer, il est nécessaire d'accumuler des données sur des échantillons en absence et présence de différentes concentrations de terminateurs et d'effectuer une analyse statistique sur la distribution en taille.

4.2.2 Mécanismes d'invasion de brins- espèces branchées

Le mécanisme d'invasion de brins peut intervenir lorsque des extrémités simple brin de polymères interagissent avec des ADN double brin comportant des cassures simple brin.

L'invasion de brin a été décrite comme un processus lent en absence de catalyseur [389], mais dans notre modèle expérimental de polymère segmenté, le nombre de points d'initiation potentiels pour l'invasion est très accru et ceci pourrait expliquer la formation de réseaux.

Pour tester cette hypothèse, nous avons réalisé une réaction de polymérisation à partir de briques d'ADN non marqué puis ajouté des ADN modifiés sur leur extrémité 3'. Nous avons suivi par gel filtration l'expérience qui dans le cas d'une invasion de brins aurait montré la formation de structures marquées par l'ADN modifié. Toutefois, aucun phénomène d'échange de brin n'a été observé dans les échelles de temps et dans des proportions compatibles avec les observations AFM. Cependant, cette expérience ne reproduit que partiellement les conditions de compétition qui prévalent durant l'assemblage. En effet, le mécanisme d'échange de brin utilisé avec succès par D. Lubrich *et al.*, lors d'assemblages d'oligomères d'ADN sous contrôle cinétique [390], mettait en œuvre des séquences d'amorçage qui n'existent pas dans nos constructions. Le rôle de ces séquences est d'amorcer le transfert de brins en fournissant un site de nucléation, situation qui existe dans la formation de nos polymères par assemblage de briques mais qui n'est pas reproduite par l'expérience menée ci-dessus.

4.2.3 Mécanismes de cyclisation

La cyclisation est la possibilité pour une molécule d'ADN de former un cercle en joignant ses deux extrémités avec une probabilité de rencontre d'autant plus élevée que la molécule est flexible. Le facteur j représente l'efficacité de cyclisation des fragments d'ADN, sa valeur dépendant également de l'orientation de torsion des extrémités du fragment.

Dans notre étude, les concentrations de briques moléculaires utilisées au cours des assemblages se situant à un ordre de grandeur au-dessus des valeurs du facteur j , elles favorisent les hybridations intermoléculaires par rapport aux intramoléculaires, limitant la formation d'espèces cycliques. Les valeurs du facteur j , utilisées dans les simulations, ont été choisies à partir de la littérature sur les vitesses de ligation en fonction de la longueur pour des ADN double brin [391-395], toutefois, étant donnée la présence de cassures simple brin dans les copolymères d'ADN ces valeurs pourraient être modifiées [325, 334, 337, 391].

La structure des briques d'ADN implique que, seules des espèces comportant un nombre pair d'unités de répétition (Figure 44) peuvent générer des formes cycliques. La formation d'espèces cycliques au cours de l'assemblage doit perturber le profil de distribution des formes linéaires provoquant une oscillation de concentration entre espèces comportant un nombre d'éléments pairs et impairs.

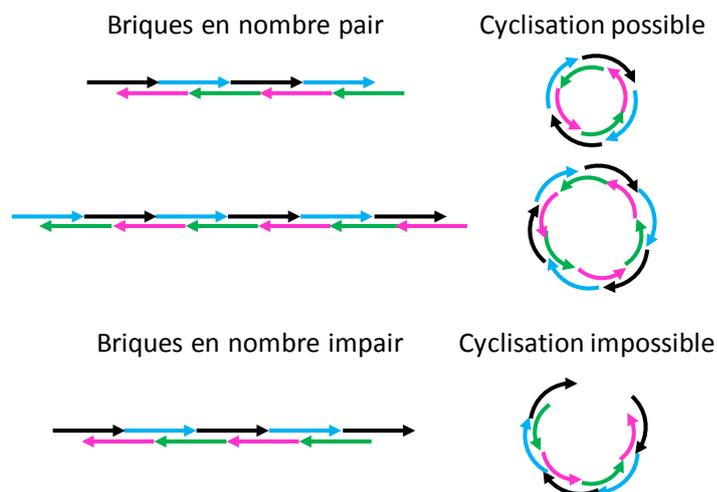


Figure 44 : Schéma de cyclisation à partir de 4 séquences d'ADN.

Le schéma a été réalisé à partir de notre modèle expérimental constitué de quatre séquences d'ADN simple brin de 39 bases semi-complémentaires.

Pour autant, nous n'avons pas observé ce phénomène sur les gels, même si la qualité limitée des migrations (déjà discutée) ne permet pas d'exclure que des formes cycliques aient échappées à l'observation.

L'absence de quantité importantes de formes cyclisées est imprévue, car on s'attend à ce que la présence de cassures simple brin au sein de fragments d'ADN double brin augmente l'efficacité de cyclisation pour des longueurs avoisinant les 200 pb [327]. En estimant une distance inter-nucléotide de 0,6 nm [396], ceci impliquerait une longueur de persistance de 6,5 nm supérieure à celles de 1 à 3 nm couramment observées pour des ADN simple brin [396-399]. Dans notre modèle expérimental où un seul nucléotide est responsable de la cassure simple brin au sein de l'ADN double brin, H. Guo et T.D. Tullius [400] ont montré que cette cassure induit une courbure anisotropique, c'est-à-dire une cassure directionnelle dans la double hélice. Il est donc possible que la géométrie induite par les cassures simple brin sur la double hélice empêche les extrémités de se rencontrer dans une situation de torsion favorable à la cyclisation malgré une plus grande flexibilité du polymère.

4.3. Calcul des paramètres sur la base des simulations et leur limitation

Nous avons défini des modèles qui prennent en compte la formation d'espèces cycliques, la modulation des vitesses d'hybridation selon la longueur des molécules et selon le contexte (simple ou double brin des séquences adjacentes au site d'hybridation). Les règles empiriques inspirées de différents travaux [314, 315, 325, 327, 334, 336, 337, 392-394] sont implémentées par l'algorithme et permettent d'obtenir des simulations en accord avec les résultats expérimentaux pour les cinétiques rapides. Toutefois, le domaine de validité de ces règles peut être discuté car s'il était compatible avec le modèle expérimental que nous avons utilisé, dans quelle mesure pouvons-nous l'appliquer à d'autres modèles ?

Par exemple, les lois de variation des vitesses d'hybridation en fonction de la longueur sont encore matière à débat, sachant que deux mécanismes opposés sont impliqués. L'un est la probabilité de nucléation qui est d'autant plus élevée que la longueur des segments simple brin correspondants est grande, l'autre est lié aux facteurs d'entropie qui affectent les vitesses avec une loi inverse. Comme discuté dans l'article, la dépendance des vitesses d'hybridation varie pour les ADN double brin selon leur taille avec un coefficient de diffusion $D \sim L^{-0.6}$ pour quelques milliers de paire de bases à $D \sim L^{-0.7}$ pour quelques centaines de paires de bases [401]. Des ADN double brin plus courts s'approcheraient de la limite de rigidité avec un coefficient $D \sim \log(L/d)/L$, où d est le diamètre de l'axe. Les ADN simple brin se comporteraient comme des chaînes libres formant des cycles aléatoires selon un coefficient de diffusion $D \sim L^{-0.6}$. Dans notre modèle qui inclut des ADN double brin segmentés par des cassures simple brin, la valeur en puissance inverse de 0.58 de la longueur totale convient et permet de simuler les expériences. Cette valeur serait à ajuster pour d'autres modèles expérimentaux impliquant des longueurs différentes.

On a essayé de voir s'il était possible de prédire *ab initio* les vitesses d'hybridation des briques d'ADN à partir des caractéristiques de séquences. On pourrait s'attendre à ce que les réactions d'hybridation soient d'autant plus lentes que l'énergie des structures secondaires intramoléculaires des réactants est élevée. Toutefois, lors de l'analyse, nous n'avons pas trouvé de corrélation générale entre la modulation des vitesses d'hybridation et les structures secondaires (voir Tableau Annexe 1). Seule, la formation de structures secondaires intramoléculaires ($\Delta G=1$ à 5 kcal/mol) affectant les extrémités des séquences, corrèle au moins partiellement avec les observations. Faute de prédiction, il serait important d'automatiser la détermination expérimentale des vitesses d'hybridation. Des techniques, telles que la détection d'hybridation par fibre optique, pourraient constituer des approches intéressantes [402-405].

4.4. Limites de l'approche combinatoire dans le cas des modèles illimités.

Dans le cas des systèmes ouverts qui génèrent une distribution illimitée de produits possibles, et par conséquent une combinatoire infinie, la modélisation de l'assemblage est rendue possible en considérant une méthode de troncation qui arrête la génération des espèces. Cette méthode s'appuie sur le fait que lors du processus combinatoire, les cycles successifs engendrent des espèces dont les concentrations décroissent exponentiellement. Ainsi au cours du dernier cycle avant la troncation, les nouvelles espèces générées sont placées dans une « boîte à sable » dont le contenu permet de valider l'ensemble du procédé. En effet, la concentration des espèces accumulées dans cette boîte doit rester très inférieure à la concentration de toutes les espèces considérées dans la simulation, autrement il est nécessaire d'augmenter le nombre de cycles combinatoires ou bien de réduire le temps de simulation pour diminuer le contenu de la boîte et être dans des conditions de calcul valides.

Les erreurs de calcul issues des approximations réalisées dans le processus combinatoire tronqué impliquent de définir les limites d'utilisation de l'algorithme et ce en fonction du modèle utilisé. Il existe un lien entre le modèle et les conditions expérimentales, qui rend unique chaque système.

Les limites computationnelles du système peuvent être atteintes pour deux raisons : la première est le temps de calcul qui croît d'autant plus que la combinatoire est élevée, la deuxième concerne la précision du calcul qui est critique lorsque l'on atteint un nombre de l'ordre du million d'espèces. Les erreurs d'arrondis et les approximations de l'algorithme d'intégration peuvent générer des instabilités numériques constituant aussi une limitation à la complexité des systèmes qui peuvent être analysés. Par exemple, un même modèle qui peut être simulé sans difficultés particulières dans les conditions où les réactions d'hybridation sont irréversibles peut devenir extrêmement instable numériquement lorsque les conditions deviennent réversibles au voisinage de la température de fusion des hybrides.

Afin de dépasser ces limites, il sera nécessaire d'échapper à une logique purement combinatoire décrivant l'ensemble des espèces pour arriver à une approche prédictive permettant de regrouper les espèces selon leur comportement global et de prédire leur évolution par des approches heuristiques. Par exemple, nous avons démontré que dans le cas d'un assemblage illimité, la distribution des concentrations des espèces suit toujours une fonction exponentielle de leur taille moyenne et que seul le coefficient exponentiel dépend du temps de réaction. On pourrait, par exemple, imaginer calculer uniquement la distribution d'un nombre limité d'espèces appartenant à un cycle combinatoire donné et en déduire une approximation du comportement de toutes les autres espèces appartenant à ce même cycle.

4.5. Réversibilité des réactions d'assemblage.

Dans notre étude, nous avons réalisé l'assemblage des briques d'ADN dans des conditions irréversibles, dans l'optique de pouvoir incorporer des molécules facilement dénaturables telles que les protéines. Toutefois, il serait intéressant de travailler dans des conditions plus proches des conditions réversibles afin d'éviter les pièges cinétiques ou la formation de structures abortives branchées. Dans le cas des assemblages testés, la température de fusion des oligonucléotides est voisine de 58°C. Des conditions optimales d'assemblage se situent typiquement d'après la littérature 5 à 7°C en dessous des températures de fusion (T_m) [406]. Nous avons pu vérifier que notre approche algorithmique et son implémentation logicielle étaient capables de gérer de telles conditions. Toutefois, ces conditions impliquent de travailler avec des objets thermostables. Dans le cadre de structures hybrides fragiles, il est possible d'approcher ce point de fonctionnement en réduisant la T_m par une diminution de la force ionique ou l'ajout de composants chimiques déstabilisant l'hybridation.

Dans ce cas, les constantes de dissociation sont obtenues à partir des constantes directes et de la constante d'équilibre déterminée par la méthode des plus proches voisins [308]. L'implémentation actuelle ne permet pas une correction automatique des constantes d'hybridation directes selon les températures. Le problème est relativement complexe dans la mesure où les vitesses d'hybridation dépendent non seulement, de paramètres relativement prédictibles comme les fréquences de collision ou les probabilités de nucléation, mais aussi de manière critique de l'évolution en fonction de la température de la structure secondaire des réactants. De plus, à des températures proches du point de fusion, la structure des complexes de nucléation évolue très rapidement rendant la prise en compte de ce facteur très difficile. En conclusion, la modélisation de l'assemblage dans des conditions réversibles est possible à condition de rester à des températures légèrement inférieures au point de fusion. Par ailleurs, les concentrations en magnésium ne sont pas prises en considération alors qu'elles-mêmes influencent significativement la cinétique et la stabilité des hybridations.

Dans le cas du modèle discuté dans l'article, les simulations montrent que pour des températures de 7° en dessous du T_m , les profils de distribution des espèces restent très semblables à ceux observés à des températures nettement inférieures où les réactions sont complètement irréversibles. Néanmoins d'un point de vue expérimental, les températures se rapprochant du T_m sont de nature à limiter les réactions parasites qui ne sont pas prises en compte dans les simulations. Ce seuil de 7°C semble donc être relativement optimal en permettant l'utilisation des simulations tout en constituant de bonnes conditions expérimentales. En effet, la simulation de modèles réversibles est beaucoup plus sensible aux phénomènes d'instabilité numérique, du fait de la propagation réverse des erreurs d'arrondi et d'approximation algorithmique au voisinage des conditions limites du système d'équation (instabilités liées à la génération accidentelle de concentrations négatives sans sens physique). Ce problème ne se rencontre pas dans la modélisation de conditions irréversibles, car les fonctions de covariance des concentrations sont dominées par les espèces les plus abondantes rendant le système peu sensible à ce type d'instabilité. La comparaison de simulations réalisées à partir d'un même modèle expérimental en conditions réversibles et irréversibles montre un temps de calcul nettement plus important dans ce dernier cas du fait des processus automatiques assurant la stabilité numérique de l'intégration.

En conclusion, d'un côté, il serait préférable de s'approcher de la réversibilité afin de résoudre un certain nombre de problèmes tels que les pièges cinétiques, d'un autre côté, ces conditions posent des problèmes en termes de modélisation.

4.6. Maîtrise des stœchiométries

Les simulations réalisées sur les deux systèmes d'assemblage convergent ou illimité/ouvert mettent en évidence deux comportements différents vis-à-vis de la stœchiométrie des réactants. Si celle-ci influence modérément un système où le produit final est unique et défini, son effet est crucial dans le cas de l'assemblage des copolymères illimités et ce d'autant plus que le temps de polymérisation est long. La simulation d'un faible déséquilibre dans le rapport molaire entre les briques met alors en évidence la présence d'oscillations de la concentration des polymères en fonction de leur longueur. Ceci est dû au fait que lors de la polymérisation, les polymères les plus courts sont utilisés en égales proportions, ce qui mène à une rapide amplification d'un quelconque déséquilibre initial et à la génération de mécanismes de terminaison par défaut de certains intermédiaires. L'élongation de la chaîne est atteinte de manière différentielle générant des oscillations. Ce phénomène, *a priori* gênant, pourrait, à l'inverse être exploité comme une approche intéressante pour l'amplification de signaux dans le cadre des biopuces à ADN mettant en œuvre des réactions compétitives.

L'algorithme nous a permis de mettre en évidence l'effet crucial des stœchiométries pour l'auto-assemblage des briques en système ouvert. Il soulève le problème expérimental de mélanger des molécules en conditions stœchiométriques, ce qui est loin d'être trivial compte tenu des incertitudes de mesure et de détermination précise des coefficients d'extinction. Une façon de résoudre le problème avec les séquences d'ADN simple brin serait de partir d'un pré-assemblage réalisé en conditions réversibles suivi d'un désassemblage d'une structure purifiée, garantissant ainsi la stœchiométrie. Cette approche n'a évidemment qu'un intérêt théorique, en permettant de valider les paramètres d'un modèle.

4.7. Conclusion

L'étude de polymérisation de briques d'ADN en conditions d'assemblage contrôlées par cinétique loin de l'équilibre thermodynamique a été réalisée. Elle a donné lieu au développement d'un algorithme permettant de simuler la distribution des populations d'espèces issues de la combinatoire générée lors de l'assemblage. Les paramètres contrôlant la distribution des espèces ont été déterminés à partir des données expérimentales. Les simulations réalisées en conditions irréversibles adaptées à notre modèle expérimental peuvent également être effectuées en conditions réversibles. L'algorithme s'applique à tout assemblage linéaire qui implique la reconnaissance entre les extrémités de briques. Ainsi, la nature des briques pourrait être changée, du moment que les règles de la reconnaissance moléculaire sont définies selon les caractéristiques des briques.

2^{ème} partie

LES STRUCTURES HYBRIDES PDNA (PROTÉINE-DNA)

1. Pourquoi associer protéines et acides nucléiques?

La nature a élaboré de nombreuses nano-machines fonctionnelles comprenant des protéines, des acides nucléiques et d'autres molécules afin de mener des tâches complexes dans les systèmes biologiques. La plupart de ces nano-machines sont réalisées par l'auto-assemblage de divers composants, en particulier de protéines et d'acides nucléiques. Le ribosome, par exemple, s'auto-assemble spontanément, de façon séquentielle, à partir de plus de 50 protéines [407]. L'approche biomimétique s'inspire de tels modèles pour tenter de générer des dispositifs artificiels de dimensions nanométriques (5 à 100 nm) difficilement abordables par les approches de micro-structuration conventionnelles ou chimiques [408].

Dans la recherche de cette fabrication d'objets nanométriques artificiels à partir de briques moléculaires, les oligonucléotides d'ADN synthétiques constituent un matériau de construction intéressant du fait de leurs propriétés d'auto-organisation prédictibles par appariement spécifique entre les bases A-T et C-G, qui permettent la conception *de novo* de n'importe quelle structure [238, 409]. A l'inverse, pour les protéines dont le repliement repose sur de multiples et complexes interactions moléculaires, la conception *de novo* reste un défi qui ne fut relevé que dans quelques cas spécifiques [96, 103]. Toutefois, les protéines soumises à des milliards d'années d'évolution présentent une large gamme de fonctions très spécifiques, bien supérieure aux fonctions limitées de reconnaissance et de catalyse des acides nucléiques. Ainsi il est apparu intéressant de combiner les propriétés des ces deux composants dans des complexes protéine-ADN qui présentent aussi bien les propriétés d'organisation spatiale qu'une diversité de fonctions.

2. Avantages et inconvénients des méthodes de couplage protéine-ADN

Les méthodes de couplage protéine-ADN ont été présentées dans la partie introductive. Elles sont divisées en deux catégories selon que les couplages sont non covalents ou covalents.

Les couplages non covalents présentent l'avantage d'être plus faciles à mettre en œuvre, toutefois leur caractérisation reste délicate du fait de la possible dissociation des composants ou encore de la difficulté de contrôler leur stoechiométrie, pour la reconnaissance avidine-biotine par exemple.

Dans de rares cas comme la reconnaissance avidine-biotine, l'affinité est telle que le complexe est stable. Dans les autres cas, les effets de changements de facteurs externes tels que la température, les pH et forces ioniques des tampons, les dilutions ou les ajouts de molécules compétitrices, sont un frein à la construction de nanostructures multiplexées. En

effet, les structures peuvent nécessiter plusieurs étapes d'assemblage et seront d'autant plus difficiles à réaliser que les briques de construction sont dissociables. Différents états, hybrides en solution vont générer des nanostructures poly-disperses.

A l'inverse, les couplages covalents permettent d'éviter la dissociation des structures hybrides. L'association des composés étant stable au cours du temps, la caractérisation des structures hybrides est plus facile à réaliser. Le couplage covalent s'avère particulièrement utile pour les systèmes multiplexés tels que ceux utilisés pour la détection simultanée de plusieurs analytes en utilisant différents complexes protéines-ADN [276].

Toutefois, la fabrication de tels objets peut s'avérer longue ou complexe. L'activité de la partie protéine peut être altérée au cours des processus d'assemblage, soit du fait des altérations de séquence nécessaires au couplage, soit du fait des conditions physico-chimiques requises pour la réaction d'assemblage. De nombreuses étapes de purification peuvent s'avérer nécessaires selon les techniques utilisées réduisant le rendement. Un autre problème à prendre en considération est la stœchiométrie et la régio-sélectivité des couplages ADN-protéines qui peuvent être difficiles à résoudre.

3. Les PDNA

Afin de générer des structures univoques, nous avons choisi de coupler un résidu cystéine rendu unique sur la protéine avec un acide nucléique modifié en utilisant un adaptateur synthétique bi-fonctionnel qui réagit avec le groupement amine d'un côté et le groupement thiol de la cystéine de l'autre. Les PDNA réalisés dans ce travail de thèse, associent :

- une molécule d'ADN simple brin présentant une modification amine avec
- un domaine protéique issu du cytochrome b5 modifié par ingénierie pour le rendre soluble et y incorporer une cystéine unique.

De ce fait, les PDNA synthétisés, présentent à la fois les propriétés d'organisation spatiale de l'ADN, les propriétés d'absorbance à 423 nm, de transfert d'électrons du cytochrome b5, et de reconnaissance du support du tag (Figure 45).

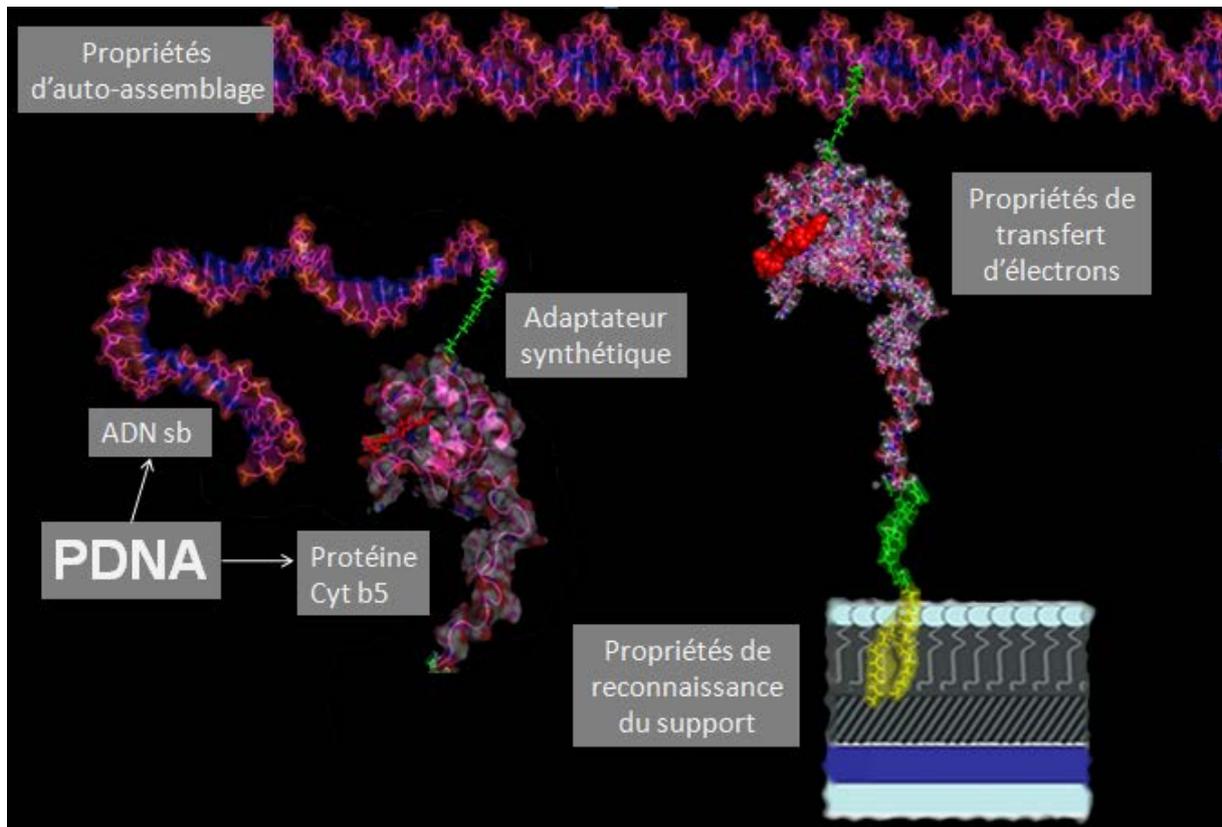


Figure 45 : structure moléculaire et propriétés des PDNA.

Les PDNA associent un ADN simple brin 39-mers au domaine soluble du cytochrome b5 modifié par ingénierie au travers d'un adaptateur synthétique bi-fonctionnel de 15Å de long. Les PDNA présentent les propriétés organisationnelles des ADN et les propriétés fonctionnelles du domaine protéique.

Les PDNA peuvent être assemblés pour former des structures définies ou des structures polymérisées selon les blocs de construction utilisés (Figure 46).

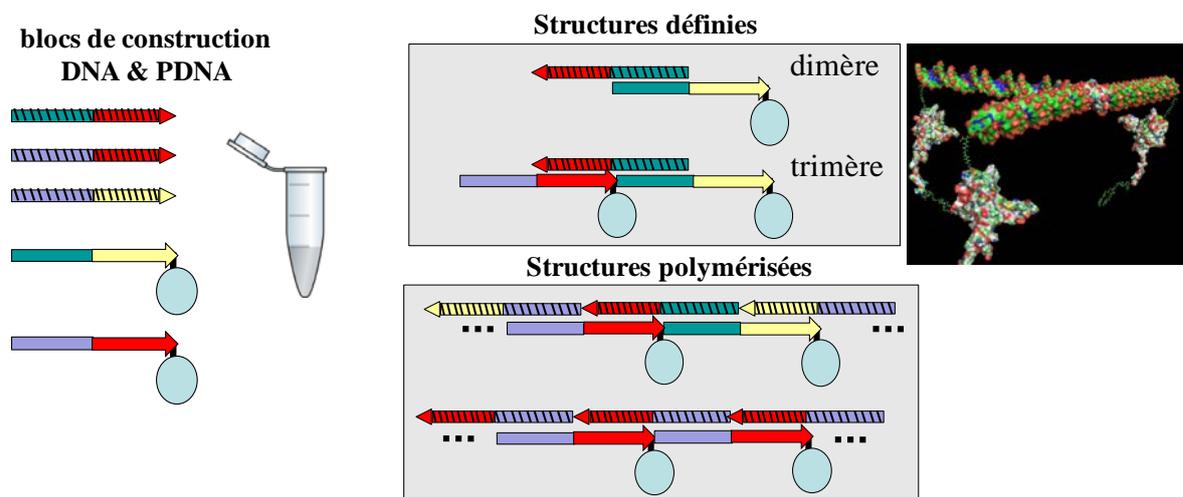


Figure 46 : Assemblage des PDNA en solution.

Les blocs de construction ADN et PDNA sont mélangés pour former soit des structures définies, soit des structures polymérisées ou encore des structures branchées.

La conception et le développement des nano-objets hybrides supposent des conditions d'assemblage douces compatibles avec chaque composant. Nous avons choisi une stratégie

d'assemblage à basse température dans des conditions loin de l'équilibre, favorables à l'intégrité des protéines.

Pour cela, nous nous sommes attachés à caractériser les propriétés physico-chimiques des PDNA ainsi que leurs propriétés d'auto-assemblage en comparaison des ADN seuls. Ces études ont été réalisées en considérant les différents environnements :

- (3D) pour les assemblages en solution,
- (2D) pour les assemblages contraints sur une surface.

Les PDNA constituent une approche originale qui pose des questions importantes aussi bien dans le domaine de la nanofabrication qu'au niveau de la compréhension des complexes protéo-nucléiques naturels.

La connaissance des paramètres impliqués dans les mécanismes d'auto-assemblage, est une étape nécessaire pour les contrôler et développer des nanostructures multifonctionnelles. Parallèlement, les études des propriétés physico-chimiques des PDNA permettent de mieux comprendre et caractériser les propriétés de structures naturelles.

3.1. Présentation des articles 2 et 3 : caractérisations des PDNA

La structure et la caractérisation de PDNA ainsi que leurs mécanismes d'assemblage sont décrits dans deux articles, un présentant le concept, l'autre détaillant les propriétés.

Un premier (article 2), s'appuie sur trois stratégies d'assemblage :

- La polymérisation en solution homogène (Figure 46),
- La polymérisation sur membrane supportée qui met en œuvre les propriétés de reconnaissance du support par la partie protéique,
- La polymérisation par élongation d'une matrice fixée sur du dextran, possible grâce à un ancrage par un ADN biotinylé qui sert d'amorce.

L'analyse par gels d'électrophorèse a montré que l'assemblage en solution se révèle adapté pour la construction de structures définies, de type dimère ou trimère, mais devient rapidement inefficace pour former des polymères longs. Même dans le cas des structures courtes, une étape supplémentaire sur gel filtration est néanmoins nécessaire pour obtenir des échantillons homogènes.

L'imagerie de résonance plasmonique a permis de suivre en temps réel l'auto-assemblage des PDNA :

- Soit lorsqu'une matrice poreuse fonctionnalisée autorise l'ancrage d'une amorce d'ADN biotinylée sur laquelle les briques de PDNA et d'ADN vont pouvoir s'hybrider pour élonger la construction. Cette technique séquentielle qui inclut des étapes de lavage entre chaque ajout donne accès à des données qualitatives et quantitatives (ces dernières seront traitées dans l'article 4) sur les mécanismes d'hybridations des blocs.

- Soit lorsqu'une membrane supportée est reconnue par le domaine protéine qui ancre alors les structures à sa surface. La qualité et l'efficacité de l'assemblage dans de telles conditions sont alors directement influencées par la planéité et l'homogénéité du substrat. L'approche nécessite un contrôle fin de la saturation de la surface. Une valeur trop élevée de PDNA peut gêner l'élongation des polymères, tandis qu'une valeur trop faible va limiter la qualité du signal.

Ces approches globales ont été complétées par des approches sur molécules uniques. Outre l'AFM, la microscopie confocale a été utilisée pour caractériser des PDNA auxquels des boîtes quantiques avaient été couplées par la reconnaissance streptavidine biotine.

Un deuxième article (article 3), nettement plus complet, reprend en détail les caractérisations engagées dans l'article 2 et utilise les outils de modélisation développés pour la caractérisation des assemblages linéaires d'ADN (article 1) afin de comparer PDNA et ADN. En particulier :

- la stabilité thermique des PDNA,
- les propriétés d'hybridation des PDNA et l'influence de la protéine,
- les propriétés électro-phorétiques et chromatographiques,
- les propriétés hydrodynamiques.

Les cinétiques d'hybridation des structures PDNA courtes et des polymères plus longs ont été caractérisées par la technique de mélange rapide. Les gels d'électrophorèse et l'AFM ont été utilisés comme approches complémentaires pour analyser la distribution de taille des polymères longs. Les problèmes de polymérisation des structures longues, mis en évidence avec les ADN dans l'article 1, ont aussi été observés avec les PDNA.

Nous avons étendu l'analyse de la polymérisation des PDNA aux approches sur membranes supportées. Ces expériences nous ont permis de montrer la plus grande stabilité d'ancrage des complexes Ni-NTA lorsque les dimères de PDNA sont préformés. En effet, la constante de dissociation des mono-complexes est de l'ordre du micro-molaire contre la centaine de nM pour les dimères [259].

Dans ce même article, nous avons également employé la microscopie de force atomique à haute vitesse pour résoudre à l'échelle nanométrique la structure des objets ancrés sur le substrat et suivre leur dynamique. Nous avons ainsi pu montrer l'existence de la formation de concatemers transitoires en équilibre dynamique, et analyser les problèmes que pose l'imagerie dynamique à échelle moléculaire d'objets hétérogènes, en termes de dynamique structurale.

PDNA as building blocks for membrane-guided self-assemblies

Pompon D, Laisné A.

Biochemical Society Transactions (2007), 35(3):495-497

PDNA as building blocks for membrane-guided self-assemblies

D. Pompon¹ and A. Laisné

Centre de Génétique Moléculaire, Centre National de la Recherche Scientifique, UPR2167, Avenue de la Terrasse, 91190 Gif-sur-Yvette, France

Abstract

Different semi-synthetic PDNAs (protein–DNA complexes), which encompass a protein core engineered from the cytochrome *b*₅ scaffold, an embedded tuneable redox cofactor, a synthetic linker and a large oligonucleotide, were designed, synthesized and purified to homogeneity. These building blocks can be reversibly attached to Ni-DOGS {1,2-dioleoyl-*sn*-glycero-3-[*N*(5-amino-1-carboxypentyl)iminodiacetic acid]succinyl}-doped supported membranes through a metal chelate bridge with the protein part and be polymerized in a fully controllable manner using a solid-phase synthesis strategy and a stepwise addition of suitable complementary oligonucleotides. The resulting structures could recreate a large range of regular distribution of patterned redox and absorbing centres separated by fully tuneable distances and geometry. Kinetic parameters for the self-assembly of building blocks were determined using SPRI (surface plasmon resonance imagery). Structures of resulting nano-objects were characterized using gel electrophoresis and single molecule approaches following decoration of assemblies with quantum dots.

Objectives

Construction of optically and electronically active nano-objects featuring highly tuneable sizes and shapes in the 10–500 nm range is a challenge in the development of the use of biological structures as building blocks for functional nano-structured devices. An additional constraint for practical use is the requirement that nano-objects can be easily incorporated into inorganic nano-structured material such as silicon-based devices, supporting electrical or optical signal transduction. The present study describes the construction of PDNA (protein–DNA complex)-based nanostructures possessing such properties. Objects were designed to be associated to metal and dielectric containing support surfaces constituting an anisotropic environment controlling positioning and orientation of the self-assembled biological structures.

Building blocks structure and construction

PDNAs are molecular building blocks that encompass a stable globular protein core embedding a tuneable redox cofactor derived from haemin, a 39 bp oligonucleotide that holds the specific recognition properties and a membrane-association domain derived from modified phospholipid. The different parts were linked together, in a unique way, using synthetic spacers as described in Figure 1. Constructed PDNAs are fully defined objects at the molecular level. Their three-dimensional structures were computed using the experimental NMR-based three-dimensional structure for

the cytochrome *b*₅ protein part and Molecular Dynamics for model refinement. Structure of the DNA part following hybridization was computed assuming the B-form of the double-stranded DNA. Shortly, the hydrophobic membrane anchor of native human cytochrome *b*₅ was removed by genetic engineering to obtain a soluble form. A tag with the -NGHHHH-CO₂H sequence was added to the truncated C-terminus of the recombinant protein and a unique cysteine residue was introduced by directed mutagenesis on a surface loop [1]. The thiol-extremity was coupled through a synthetic bifunctional linker with different types of oligonucleotides bearing an amino-group-terminal extension at their 3'-end [2]. Resulting PDNAs lacking the synthetic lipid extension were synthesized and purified to homogeneity at milligram levels [3] and their homogeneity carefully checked by gel DNA and protein electrophoresis and by HPLC coupled with MS. Natural haemin cofactor can be easily replaced by analogues with modified porphyrin rings or metal to modulate redox and optical properties. The lipid part was finally associated to the construct through a nickel-chelate involving the His₄ part of the protein and the amino-triacetate part of the modified phospholipid. In such a PDNA, the lipid–protein link can be cleaved by nickel chelation with histidine or EDTA and the nucleic acid–protein link by reduction with DTT (dithiothreitol).

Self-assembly of building blocks into nano-objects

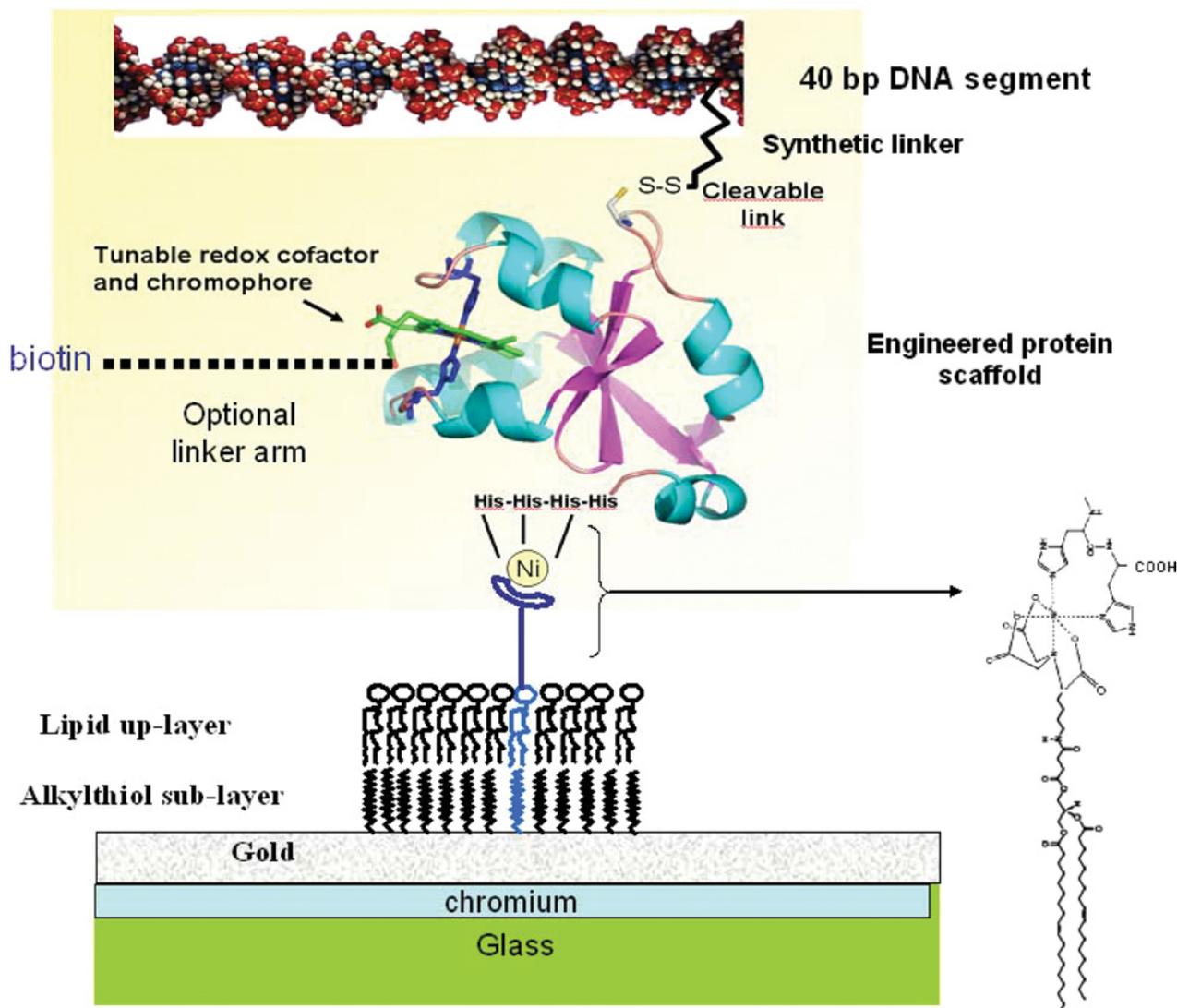
PDNAs building blocks can be assembled into large nano-objects using overlapping complementary oligonucleotides (Figure 2). Three self-assembly strategies were developed. The first was used to construct extended building blocks composed of two complementary oligonucleotides overlapping the 5'- and 3'-end of the PDNA respectively. This gives rise

Key words: atomic force microscopy (AFM), cytochrome, nano-object, protein–DNA complex, self-assembly, surface plasmon resonance.

Abbreviations used: AFM, atomic force microscopy; CM, confocal microscopy; DTT, dithiothreitol; SPRI, surface plasmon resonance imagery; PDNA, protein–DNA complex; QD, quantum dot.

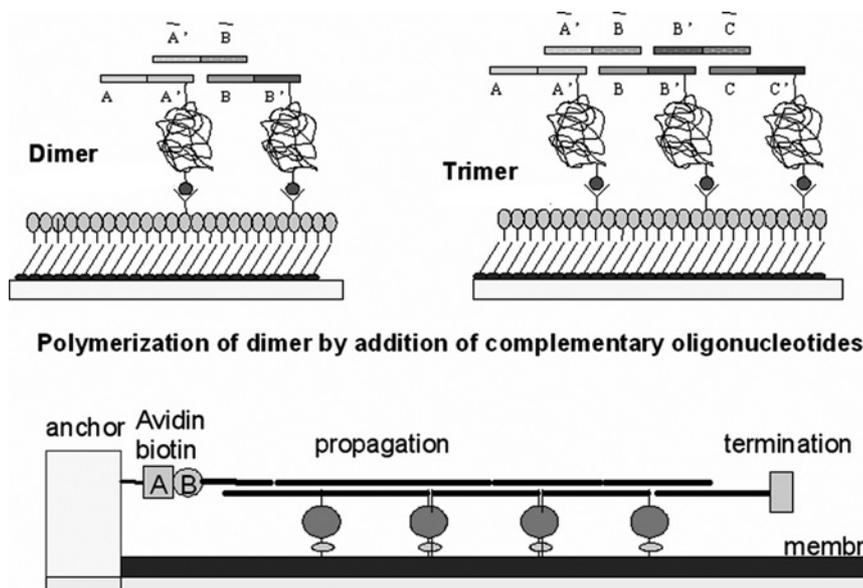
¹To whom correspondence should be addressed (email pompon@cgm.cnrs-gif.fr).

Figure 1 | Structure of synthetic protein-DNA building block (PDNA)



to molecular blocks encompassing a PDNA and two single-stranded DNA extensions that can be used as recognition elements to build extended structures. These simple objects were built by stoichiometric hybridization of the three complementary components in solution followed by purification of the formed trimer. Quality of constructs was checked by gel electrophoresis. For construction of larger objects, simple mixing of solutions of the different building blocks generally led to poor results because of the multiplicity of possible combinations and the sequence ambiguity when several competing hybridizations can occur. A more efficient approach was a solid-phase synthesis strategy consisting of linking a biotinylated priming sequence to a solid support, which can be a gold surface covered with a polymer such as dextran or micro-beads modified by covalently bound streptavidin. This primer is used to initiate a multistep sequential elongation of the construct using a suitable washing step between each addition of a new PDNA or oligonucleotide block. Use of a magnetic-bead support is convenient but does not allow

monitoring of individual hybridization steps. Variable hybridization rates of building blocks very rapidly led in such a case to uncontrolled elongation generating a high proportion of wrong structures. In contrast, running the construction on a suitable porous polymer attached to a gold surface allowed real-time monitoring of reactions by SPRI (surface plasmon resonance imaging). Use of a cleavable linker between the primer and the matrix allowed recovery of full assemblies. The third approach consisted of directly building, step by step, the objects on to a supported membrane, with or without a primer sequence attached to specific surface nano-structures. In such a case, PDNAs are bound to the membrane through their Ni-DOGS {1,2-dioleoyl-*sn*-glycero-3-[*N*(5-amino-1-carboxypentyl)iminodiacetic acid]succinyl} lipid extension. This last approach is more direct but elongation efficiency and thus quality of final constructs are highly dependent on the microrugosity and homogeneity of the support. Different building modes can be also combined; for example, building blocks lacking the lipid part can be pre-assembled in

Figure 2 | Self-assembly of PDNAs into larger objects

solution into nano-objects before binding and elongation in the membrane-bound mode.

Characterization of nano-objects

Developed nano-objects can be characterized at different levels. Use of SPRI monitoring during the object construction is a particularly useful tool, permitting both quantification of building-block addition and kinetic analysis of the different steps. The size of final constructs can be evaluated following electrophoresis on agarose gel as for regular DNA. To allow accurate evaluation of sizes, the link between the DNA and the protein/lipid part has to be cleaved prior to analysis with a reducing agent such as DTT or mercaptoethanol. Large objects including five or more building blocks were found to be relatively fragile and could be stabilized before electrophoresis by enzymatic ligation. Additional characterization of assemblies can be performed at a single-molecule level by decoration of the nano-object extremities with complementary oligonucleotides linked to QDs (quantum dots). QDs with different colours can be used at extremities. Reading was easily performed using epifluorescence microscopy with a colour video camera or better by visualization of an individual object using CM (confocal microscopy). Single-molecule emission spectra can be recorded and used for identification and the size of large construct evaluated when exceeding 150 nm. The structure of individual objects can be also monitored by AFM (atomic force microscopy) following binding on Ni-NTA (Ni²⁺-nitrilotriacetate)-modified thermal silica held on silicon.

Concluding remarks

PDNAs are very versatile molecular building blocks with tuneable electron transfer, absorption and fluorescence properties. Redox PDNAs can be assembled in a fully controllable manner in nanometric objects of variable geometries with fully tuneable sizes ranging from ten to several hundreds of nanometres. The unique approach of sequential assembly on a supported membrane allows controlled building of large objects without substrate defect interference. Both global (gel electrophoresis, SPRI) and single-molecule (CM and AFM) real-time characterizations of assemblies have been successfully developed.

We thank Dr Gilles Truan (Centre National de la Recherche Scientifique, UPR2167) for the help in molecular modelling and communication of human cytochrome *b₅* co-ordinates. The present study is part of the NANOBIOFORME project supported by the French 'Agence National de la Recherche' PNANO (Programme National en Nanosciences et Nanotechnologies).

References

- Boireau, W., Bombard, S., Sari, M.A. and Pompon, D. (2002) *Biotechnol. Bioeng.* **77**, 225–231
- Boireau, W., Zeeh, J.C., Puig, P.E. and Pompon, D. (2005) *Biosens. Bioelectron.* **20**, 1631–1637
- Boireau, W., Duncan, A.C. and Pompon, D. (2005) *Methods Mol. Biol.* **300**, 349–368

Received 2 January 2007

**Self-Assembly Properties and Dynamics of
Synthetic Proteo-Nucleic Building Blocks
in Solution and on Surfaces**

Laisné A, Ewald M, Ando T, Lesniewska E & Pompon D.

Bioconjugate Chemistry (2011),22 (9):1824-1834

Self-Assembly Properties and Dynamics of Synthetic Proteo–Nucleic Building Blocks in Solution and on Surfaces

Aude Laisne,[†] Maxime Ewald,[‡] Toshio Ando,[§] Eric Lesniewska,[‡] and Denis Pompon^{*,†}

[†]Centre de Génétique Moléculaire, CNRS, UPR3404, Avenue de la Terrasse, F91190 Gif-sur-Yvette, France

[‡]Institut Carnot Bourgogne, UMR CNRS 5209, University of Bourgogne, F-21078 Dijon, France

[§]Department of Physics, Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan

S Supporting Information

ABSTRACT: Synthetic proteo–nucleic structures (PDNAs) encompassing a single-stranded DNA sequence covalently attached to a redox protein domain able to interact with surface or matrix were designed and characterized. They constitute versatile building blocks alternative to regular DNA for creating scaffolds with optical, electrical, or catalytic properties. PDNAs self-assemble in the presence of complementary oligonucleotides, to form a network of protein domains linked by double-stranded DNA segments. Electrophoretic and hydrodynamic behaviors of PDNAs and corresponding DNA were compared under electrophoresis and gel filtration conditions. Hybridization rates between small and large assemblies were characterized by rapid-mixing experiments. Results showed that the protein part significantly contributes to hydrodynamic behaviors of structures but marginally affects the conformation and hybridization properties of the nucleic domain. PDNA metal-mediated complexes with nitriloacetate-modified phospholipids can diffuse and interact at the surface of vesicles or supported membranes. Surface plasmon resonance analysis of membrane–PDNA interactions indicated that two protein units are required to allow stable surface association and that surface occupancy constrains assembly sizes. High-speed atomic force microscopy illustrated rapid lateral diffusion of assemblies on mica, revealing transient association between noncomplementary PDNA extremities and frequent trapping by surface defects. Regularly organized protein domains were visualized using a larger DNA framework.

INTRODUCTION

The use of a synthetic DNA scaffold to build mono- or multidimensional nano-objects with fully predictable shapes rapidly developed as an important field of synthetic biology.^{1–7} However, the limited range of functionalities that can be assembled using only nucleic acid tiles raised the question of the position specific incorporation of protein block in such scaffolds.⁸ One way involved the DNA binding domain or the recognition of modified bases (for example biotin) by specific protein domains (avidin, antibodies, etc.).^{8–11} Reciprocally, alternate approaches profit from the recognition of specific protein domains by aptamer structures embedded into the DNA scaffold.¹² However, these two methods suffer strong limitations because they require fusion of a large protein recognition domain to functional parts of interest or restrict the approach to structures for which suitable aptamers can be created. Protein–DNA recognition based on chemically mediated approaches (such as NTA–metal complexes) can also be considered, but either gives rise to reversible association or lacks specificity.^{13,14}

PDNAs (protein–DNA structure), which are artificial structures that consist of a protein domain covalently linked to a single-stranded DNA (ssDNA) part through synthetic structures, constitute an attractive alternative.^{15,16} The protein part of PDNA can be chosen to carry various types of functions such as surface interaction, binding, enzymatic, optical, or electrical properties, potentially creating nanostructures that can be used as biosensors^{17,18} or active molecular devices. PDNA tiles can

self-assemble in solution or be attached to a matrix in the presence of complementary ssDNA.¹⁹ However, little is known about the stability and physicochemical and structural features of such objects or the ways they interact together compared to regular DNA tiles. This is particularly critical for defining methods adapted for specifically incorporating these building blocks into DNA scaffolds like origami⁶ or other DNA-based supramolecular assemblies.²⁰

This paper extends synthesis approaches for such PDNA building blocks and analyzes their physicochemical and assembly properties in comparison to their corresponding pure nucleotidic parts, at both global and single-molecule levels.

EXPERIMENTAL PROCEDURES

Synthesis of the PDNA Building Block. Oligonucleotides (Eurogentec) presented in section 1 of the Supporting Information were purified by ion-pairing reverse phase chromatography as previously described.²¹ PDNAs were prepared by hemisynthesis from engineered human cytochrome *b₅* as described in sections 2 and 3 of the Supporting Information and previous work.^{16,17,19} Briefly, a 3'- or internally amino linker-modified oligonucleotide was covalently attached through a disulfide bond to a modified cytochrome *b₅* (*b₅*) containing a unique Cys

Received: May 2, 2011

Revised: July 8, 2011

residue and a His₄ C-terminal tag. During the synthesis, the amino linker of the ssDNA was coupled to a second linker terminated by a thiopyridine-protected thiol using NHS chemistry. A disulfide bond was finally formed by reacting the activated linker end with the unique Cys residue of the protein. The resulting products were purified to homogeneity by a combination of ion exchange and affinity chromatography, taking advantage of the protein His tag.

PDNA Thermal Stability. Experiments were performed in 20 mM Na/K phosphate buffer (pH 7.4) using PDNA or *b*₅ concentrations in the 10 μM range. Thermal denaturation was monitored on a Lambda 2 (Perkin-Elmer) spectrophotometer using increasing temperature steps of 4 °C. Spectra were recorded between 260 and 600 nm immediately after each temperature stabilization at a rate of 120 nm/min. Denaturation curves for native, engineered *b*₅, and PDNA were calculated from the changes in the 413 nm absorbance.

Phospholipid Vesicle Preparation. 1,2-Dimyristoyl-*rac*-glycero-3-phosphocholine (DMPC) was purchased from Sigma-Aldrich and 1,2-dioleoyl-*sn*-glycero-3-{[*N*-(5-amino-1-carboxypentyl)iminodiacetic acid]succinyl} (nickel salted) (DOGS-NTA) from Avanti Polar Lipids. A chloroform solution containing 1 mg of a DMPC/nickel DOGS-NTA mixture (90:10 molar ratio) was evaporated into a glass vial under a nitrogen flow, and the residue was dried under vacuum for 2 h. Phospholipids were hydrated overnight at 4 °C with 1.4 mL of 1 × PBS buffer (Sigma-Aldrich), and the mixture was sonicated until a clear solution containing a total phospholipid concentration of 1 mM was obtained. The vesicles were extruded 19 times through polycarbonate membranes with 100 nm pores (Liposofast, Avestin).

Synthesis of Defined PDNA Assemblies. A detailed description is given in section 8 of the Supporting Information. Briefly, hybridizations of DNA and PDNA tiles were performed at 24 °C in 1 × PBS buffer using a slight molar excess of DNA compared to PDNA. The constructs were purified from excess DNA either by association of IMAC and Sephadex G25/G50 chromatography or by HPLC size exclusion chromatography. The first method involved binding to an IMAC column (0.5 cm × 1.5 cm) and elution with a buffer supplemented with 30 mM histidine. EDTA (final concentration of 10 mM) was added to the eluate prior it being loaded onto a Sephadex G25 gel filtration column (0.5 cm × 1.5 cm) equilibrated with 1 × PBS buffer. Alternatively, purification was performed by size exclusion chromatography (BioSuite 250, 4 μm UHR SEC, 4.6 mm × 300 mm, Waters) at a rate of 0.3 mL/min with 0.25 M NaCl and 0.15 M NaKPO₄ buffer (pH 6.8). DNA Ladders N3233 (Biolabs) were used for calibration.

Analysis of DNA and PDNA Hybridization Kinetics by Rapid-Mixing Experiments. Hybridization of simple DNA and PDNA tiles and preformed dimers was analyzed in rapid-mixing experiments by monitoring the hypochromic effect associated with the formation of double-stranded structures. Kinetic constants were calculated by fitting of the traces as described previously.²¹

Gel Filtration and Electrophoresis Methods. HPLC gel filtration analysis was performed on a Waters BioSuite 250 column (4 μm UHR SEC, 4.6 mm × 300 mm) at a rate of 0.3 mL/min with 0.25 M NaCl and 0.15 M NaKPO₄ buffer (pH 6.8). Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) analysis was performed using 4 to 12% NuPAGE Novex Bis-Tris polyacrylamide gels run at 24 °C and 120 V. Protein parts of PDNA were revealed by fast Coomassie

blue staining and calibrated using the Mark12 Unstained standard from Invitrogen. Agarose gel analysis was performed using 3.5% NuSieve GTG Agarose from Tebu-bio in 1 × TBE run at 50 V and 4 °C. DNA Ladders N3233 (Biolabs) were used for calibration. Agarose gel electrophoresis patterns were deconvoluted using an optimized expansion factor of 1.5 and band quantified as described previously for the analysis of regular DNA polymers.²¹

Surface Plasmon Resonance (SPR) Methods. SPR experiments were performed on a SPRi-Lab⁺ array system (Genoptics) using 1 × PBS as running buffer at a flow rate of 230 μL/min and 24 °C. Gold layer-covered prisms were cleaned by gentle sonication in a hot acid Pyranha solution (equal volume of 30% hydrogen peroxide and pure sulfuric acid), abundantly rinsed with Milli-Q water, and stored in pure HPLC grade acetonitrile until they were used. Experiments were performed following phospholipid vesicle fusion either on bare gold (lipid bilayer) or on a self-assembly monolayer of alkyl chains (lipid hemilayer). For that, the Pyranha-cleaned prism was immersed in a freshly prepared 1 mM solution of 1-octadecanethiol in ethanol (Sigma Aldrich) for 18 h and then rinsed with ethanol and water. Freshly extruded phospholipid vesicles (1 mM into 1 × PBS buffer) were fused onto the treated substrate by circulation in the SPR cell at a flow rate of 50 μL/min for 30 min at 24 °C. An excess of adsorbed vesicles was removed by washing for 10 min with pure water (hypo-osmotic choc) followed by equilibration with 1 × PBS buffer.

High-Speed Atomic Force Microscopy (HS-AFM). A laboratory-built high-speed atomic force microscope described previously was used to visualize the structure.^{22–24} All observations were performed in oscillating contact mode in liquid. Small cantilevers designed for HS-AFM with a spring constant of 0.1–0.2 N/m, a resonant frequency of 0.8–1.2 MHz, and a quality factor of ~2 in buffer solution were used.²⁵ The cantilever tips were grown by electron beam deposition (SEM Jeol 6500) and further sharpened by oxygen plasma etching (Plasma cleaner Harrick). The cantilever's free oscillation peak-to-peak amplitude (*A*₀) was set to 4–5 nm, while the feedback set point during imaging was set to ~0.9 *A*₀. The samples were diluted (2–10-fold) in an observation buffer (OB) containing 10 mM MgCl₂. A drop (~2 μL) was deposited on freshly cleaved green muscovite mica, and the mica was incubated at room temperature for 5 min. Subsequently, the mica surface was immersed in a liquid chamber filled with 60–100 μL of OB and subjected to HS-AFM. Diluted samples were used within a day (storage at 4 °C).

RESULTS

Three types of PDNAs including an engineered cytochrome *b*₅ redox domain and differing by the sequence of their nucleotide parts and by the grafting position on the sequence of the protein domain (Figure 1, right) were prepared by hemisynthesis and purified to homogeneity (sections 2 and 3 of the Supporting Information). In these structures, the protein C-terminal His tag plays an important role in permitting the formation of a metal chelate linkage between the PDNAs and Ni-NTA groups attached to either a polymer matrix, modified phospholipids, or functionalized surfaces. Lateral diffusion of anchored PDNAs remains possible either by stepwise release and rebinding of the metal chelate or, in the case of vesicles or a supported membrane, through lipid diffusion. Hybridization of the PDNA single-stranded part to complementary nucleic acid generates a more

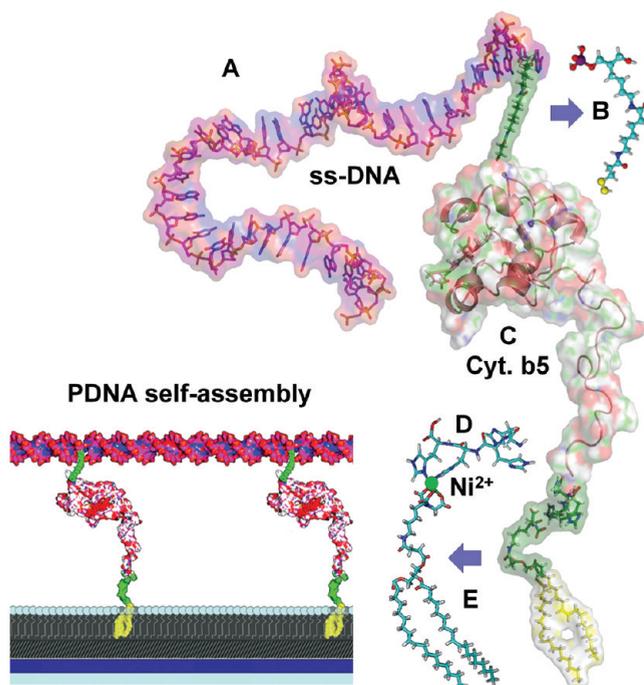


Figure 1. Structures of the PDNA building block, including a b_5 core. The structure is composed of a 39-base ssDNA part (A) attached either by its 3'-end (as illustrated) or by an internal amino C6 thymidine (not shown) to a synthetic linker (B) coupled to cytochrome b_5 (C) by the thiolate extremity of a cysteine at position 24. The His₄ C-terminal end (D) of the protein interacts with the nickel ion of an 18:1 DOGS-NTA-modified lipid (E). The structure at the bottom left shows PDNA blocks inserted into a dsDNA structure. The structure is floating on a supported membrane composed of a phospholipid monolayer covering a self-assembled octadecanethiol monolayer on gold.

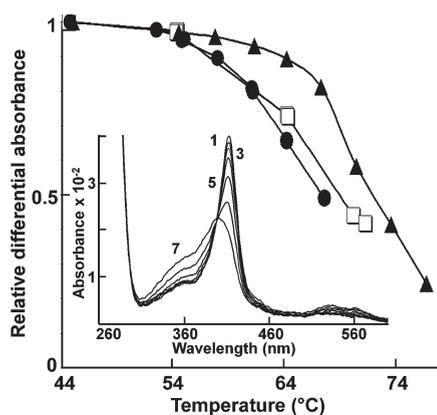


Figure 2. Thermal stabilities of b_5 folding in wild-type, Cys-modified, and corresponding PDNA structures. Protein denaturation was monitored by observation of the spectral changes of the bound heme cofactor at 413 nm. Denaturation curves were compared for native (\blacktriangle), Cys-modified b_5 (\square), and PDNA (\bullet) forms. Typical spectral changes following incubation of a PDNA at temperatures increasing from 44 to 68 °C in steps of 4 °C are illustrated in inset curves 1 (44 °C) to 7 (68 °C).

194 rigid double-stranded DNA (dsDNA) structure (Figure 1,
195 left) that, depending on sequences, could self-organize into a
196 wide range of nucleo–proteic networks. Such structures can

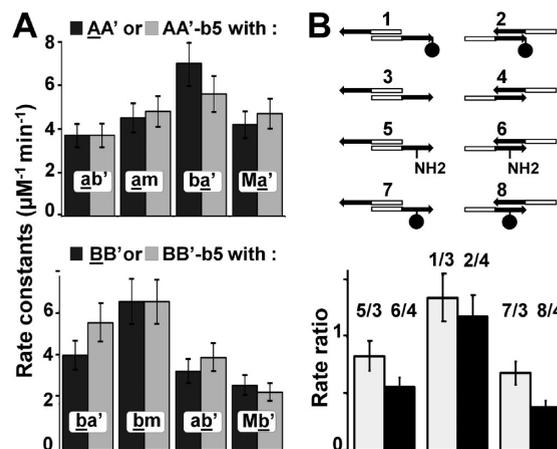


Figure 3. Comparative rate constants for DNA and PDNA hybridizations. (A) Hybridizing oligonucleotides are indicated inside bars (matching segments underlined). Dark and light gray boxes refer to hybridization of DNA and corresponding PDNA to half-cDNA, respectively. Second-order rate constants were deduced from measurements of the hypochromic effect in rapid-mixing experiments. (B) Ratios among hybridization rates for normal, base-modified, and protein-coupled oligonucleotides. These ratios are defined by a scheme in which the circle stands for the protein domain and the NH₂ label for the internal base modification. White and black bars correspond to hybridizations involving the distal and proximal segments of oligonucleotides, respectively.

alternatively be made by substituting the cytochrome b_5 domain
with any other protein of interest provided that a single grafting
point can be engineered at the surface.

Stability and Hybridization Properties of the PDNA Building Block in Solution. The thermal stability of the PDNA protein domain under conditions required for hybridization reactions is critical for the maintenance of the functional integrity and particularly redox properties. This was analyzed in solution by monitoring the temperature-dependent disruption of the cytochrome b_5 absorption spectra, which tightly depend on the iron coordination by two histidine residues of the protein. Results indicated that the reversible denaturation of the protein part of the PDNA is observed only above 55 ± 2 °C, a value identical to that found for the DNA-free engineered protein and slightly lower (2–3 °C) that that for the native protein (Figure 2). Oxidoreduction properties of the cytochrome b_5 domain in PDNA structures were found to be conserved. An identical transition in the absorption spectra, which is characteristic of the redox change of the histidine-coordinated iron porphyrin, was observed upon electron transfer in the native protein and on PDNA structures. Thus, the presence of the ssDNA part has no significant effect on the stability and redox properties of the protein domain. The thermal stability of the protein is sufficiently high to allow hybridization to proceed under classical conditions (55 °C) for oligonucleotides.

Reciprocally, the effect of the protein domain on the hybridization properties of the DNA domain was investigated for structures in which covalent grafting was performed either internally or at the 3'-end of the oligonucleotidic part. Monitoring of hypochromic effects was used to analyze the relative rates of hybridization of the different PDNAs with complementary ssDNA sequences. To discriminate between hydrodynamic and protein-related microenvironment effects, hybridization properties were analyzed on both the proximal and the distal

197
198
199
200
201
202
203
204
205
206
207
208
209
210
211 F2
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230

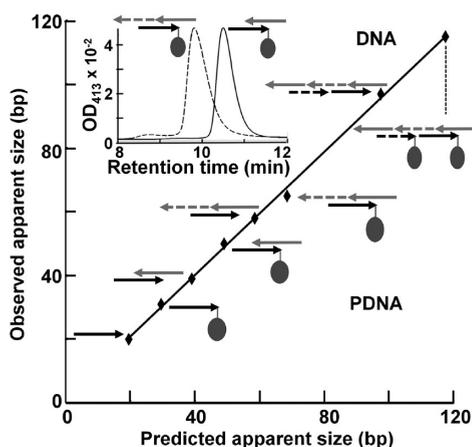


Figure 4. Size exclusion chromatography characterization of short PDNA assemblies. Structures depicted were analyzed on a HPLC gel filtration column (detection at 413 and 254 nm) and apparent molecular masses calculated using dsDNA as molecular mass standards. Apparent molecular masses (expressed in base pairs) were predicted using the formula molecular mass (base pairs) = $10P + 0.5S + D$, where P , S , and D stand for the number of protein unit(s) (●) and ss- and dsDNA bases (arrows), respectively. The inset shows experimental peak shapes for illustrated constructs.

sides (3'- and 5'-halves, respectively) of the nucleic part of PDNAs. In addition, the potential influence of the nature (ds- vs ssDNA) of the region flanking the hybridization site was investigated.

During hybridization of PDNAs, protein absorption at 260 nm represents less than 10% of the nucleic acid absorbance and does not contribute to spectral changes. Second-order hybridization rate constants ranging from 1.9 to $7 \mu\text{M}^{-1} \text{min}^{-1}$ were observed (Figure 3), while the lengths of the hybridizing segments were almost identical (19–20 bases) in all experiments. Comparison of hybridization rates for DNA and 3'-end-grafted PDNA with an identical sequence indicated that protein-dependent modulation of rates is limited with regard to the effects related to the sequence of hybridizing segments (Figure 3A). Moreover, analysis of the relative rates of hybridization events occurring in the distal and proximal parts of PDNA confirmed a hardly significant influence of the protein domain when grafted at the 3'-end of the nucleic part (labels 1/3 and 2/4 in Figure 3B). In contrast, when the grafting was performed internally in the nucleic part, a more significant effect on rates was observed, which resulted both from the presence of a modified base in the hybridizing segment (labels 5/3 and 6/4 in Figure 3B) and the attachment of the protein domain (labels 7/3 and 8/4). In both cases, distal grafting has a limited influence compared to proximal grafting, suggesting that the effect relates more directly to protein–DNA interaction rather than to changes in the global hydrodynamic properties in solution.

Electrophoretic and Chromatographic Properties of PDNA Assemblies of Increasing Sizes. To further investigate the relations between hydrodynamic behaviors and structures, different types of PDNA assemblies were built. The relative contributions of protein and nucleic domains were analyzed using a series of constructs built from hybridization-driven assembly of two PDNA tiles (or corresponding DNA controls), namely, AA'-b5 and BB'-b5 (section 1 of the Supporting Information). The AA' PDNA building block was designed to hybridize with the 5'-half of ab' and the 3'-half of ba' cDNA,

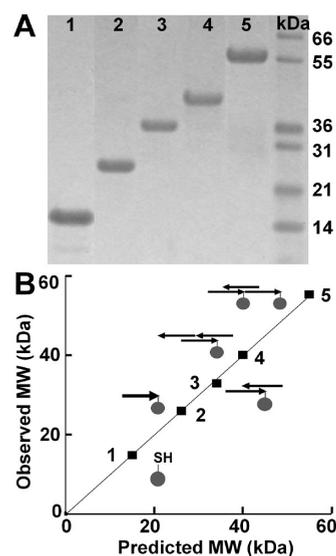


Figure 5. SDS–PAGE analysis of short PDNA assemblies. (A) Experimental gel, in which the numbering of lanes corresponds to the structure depicted in the bottom panel. (B) Relation between the observed molecular mass (calculated on the basis of a protein standard) and those predicted using the formula molecular mass (kilodaltons) = $15P + 0.282S + 0.359D$, where P , S , and D stand for the number of protein unit(s) (●) and ss- and dsDNA bases (arrows), respectively. Structures 2–4 are derived from AA'-b5 PDNAs, and structure 5 combines AA'-b5 and BB'-b5 PDNAs.

matching their a and a' parts, respectively. Similarly, the BB' PDNA block can hybridize to ba' and ab' DNA tiles by matching their b and b' parts, respectively. The complementary tiles were assembled by single-step or sequential (purification of intermediates) hybridizations at low (25–45 °C) temperatures followed by final purification to homogeneity by combinations of Ni-NTA affinity and gel filtration chromatography to resolve PDNA assemblies from excess DNA. Corresponding structures and HPLC gel filtration profiles are illustrated in Figure 4. Individual constructs migrate during gel filtration as Gaussian-shaped bands, allowing easy determination of their relative mobility under the high-salt buffer conditions required to prevent interaction with the silica matrix of columns. Interestingly, DNA and protein structural elements contributed in a perfect additive way to the apparent size calculated on the basis of dsDNA standards. The individual contributions of the ssDNA (0.5 bp/bp), dsDNA (1 bp/bp), and protein (equivalent to 10 bp/unit) parts appeared to be independent of the assemblies considered. Consequently, the plot of the apparent sizes versus calculated values for eight different constructs is linear. The 10 bp/unit equivalent contribution of the protein part corresponds to an apparent molecular mass of 6.6 kDa, a value significantly smaller than its actual size of 13 kDa, which would correspond to smaller hydrodynamic radii for proteins compared to those of dsDNA standards with identical masses. However, this is not consistent with predicted hydrodynamic radii^{26,27} that are similar (1.9–2 nm) for a dsDNA of 13 kDa (20 bp) and for the b₅ core (13 kDa). The difference can be related to the flux-dependent hydrodynamic stretching that is expected to occur and affect differentially globular and linear structures in the tightly packed HPLC gel filtration column.

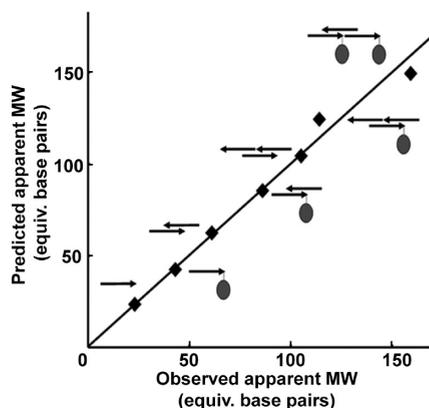


Figure 6. Agarose gel analysis of the apparent molecular mass of PDNA and DNA assemblies. Structures depicted on the figure were resolved by 3.5% agarose gel electrophoresis in $1 \times$ TBE. Calibration was performed with dsDNA standards. Molecular masses were predicted using the formula molecular mass (base pairs) = $25P + 0.6S + D$ for DNA–DNA constructs and the formula molecular mass (base pairs) = $25P + 1.55S + D$ for PDNA–DNA constructs, where P , S , and D stand for the number of protein units, ssDNA bases, and dsDNA base pairs, respectively.

To gain additional insight, the electrophoretic mobility of the different constructs was evaluated under SDS–PAGE conditions (Figure 5). Analysis illustrated that coupling of the DNA parts to the protein induces a strong shift in mobility. The apparent molecular mass increases from ~ 15 kDa for the b_5 core alone to 27 kDa for the corresponding single-stranded PDNA, the difference of 12 kDa being consistent with the molecular mass (~ 13 kDa) of the DNA part. This result contrasts with gel filtration data and could appear surprising as the protein and DNA parts exhibit very different permanent charges. However, protein in SDS–PAGE is expected to be denatured into an extended and rather rigid form and negatively charged by the interacting SDS.²⁸ SDS-denatured protein and dsDNA have thus some common conformational features that de facto result in identical electrophoretic mobilities for similar masses. In contrast, the contribution to the apparent molecular mass of the ssDNA part was consistently found to be $\sim 50\%$ larger than that of the dsDNA part for equivalent masses with a lower persistence length of ssDNA compared to that of dsDNA.

The conformation of PDNA-based structures in solution was analyzed using non-denaturing agarose gel electrophoresis (Figure 6). An additive contribution to the apparent size (expressed as base pair equivalents) of PDNA assembly components was observed as described previously with ssDNA, dsDNA, and protein parts contributing to ~ 0.6 , 1, and 25 bp equivalents, respectively (section 4 of the Supporting Information). The 16.5 kDa protein contribution is close to the 15 kDa value determined by SDS–PAGE, while calibration was performed using dsDNA standards. This feature is the reciprocal of the situation found for DNA calibrated on a protein standard using SDS–PAGE. It suggests that the folded b_5 core has electrophoretic mobility similar to that of dsDNA with the same mass in agarose gels. Likewise, the relative apparent masses of ss- and dsDNA are similar in the two electrophoretic systems. However, analysis of the electrophoretic mobilities of DNA and PDNA polymers evidenced that the contribution of ssDNA to the apparent molecular mass differs in PDNA (1.55–1.78 bp/base) and DNA (0.60–0.69 bp/base) assemblies (section 4 of the

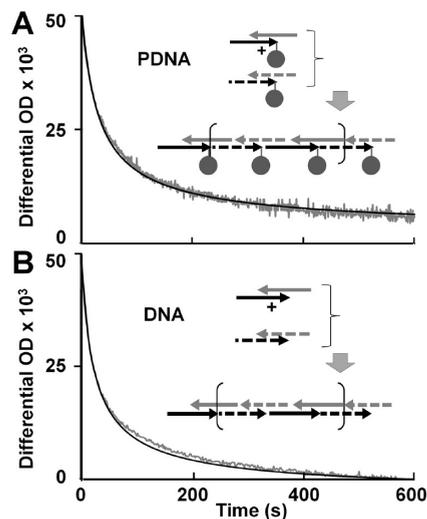


Figure 7. Rapid-mixing analysis of the formation of tile concatemers. (A) Experimental (gray line) and simulated (black line) kinetic courses for the copolymerization of the depicted preformed PDNA–DNA hybrids. (B) Same data for the hybridization of corresponding DNA–DNA hybrids. Simulations were performed using rate constants from Figure 3A and parameters from section 6 of the Supporting Information.

Supporting Information). The effect is identical with one or several protein units per construct and suggests that the presence of protein affects the contribution to mobility of linked ssDNA but not of dsDNA (section 4 of the Supporting Information). This phenomenon was not observed during gel filtration where the protein is folded but an electrical field is absent or during SDS–PAGE where the protein domain is unfolded.

Hybridization Properties of Large PDNA-Based Structures. While hybridization properties of defined small PDNA and corresponding DNA structures appeared to be very similar, we questioned the behavior of larger assemblies in which protein content could significantly impact geometry and steric hindrance causing topological restriction to base pairing. The model used was the formation of block copolymers involving regular repeats of two types of PDNAs. When AA' and BB' PDNA were mixed with half-complementary tiles (ab' and ba'), a polymer featuring AA' – BB' PDNA repeats can be formed. The reaction was first monitored using the hypochromic effect resulting from the hybridization of single-stranded sticky ends. Panels A and B of Figure 7 compare the reaction courses for PDNA–DNA and DNA–DNA polymerization, respectively. Individual rate constants described in Figure 3 were used to simulate the reaction courses using the same modeling approaches²¹ for PDNA and DNA assemblies. This model assumes that hybridization rates of large polymer fragments decrease with a power of -0.57 of species length for DNA.^{21,29} For PDNA, an equivalent length was calculated as the sum of the real size of the nucleic part and of a contribution of the protein equivalent to 10 dsDNA bases per protein unit. This value was set to be identical to the protein contribution evidenced in gel filtration experiments. The figure illustrates that experimental courses were well-simulated with this same model for DNA- and PDNA-based experiments. This indicates that, for large linear assemblies, no significant additional factor, which could differentiate PDNA from DNA, needs to be taken into account to describe the global behavior.

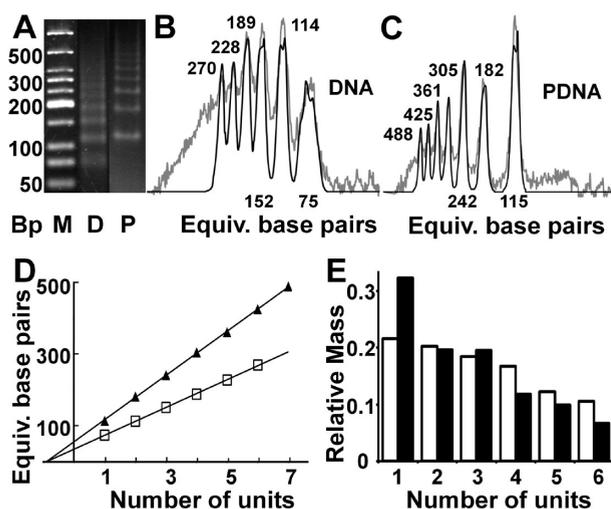


Figure 8. Size distribution of DNA and PDNA copolymers. (A) Agarose gel electrophoresis of DNA and PDNA polymer distribution after polymerization for 10 min of the four reactant tiles at $\sim 1 \mu\text{M}$ each and quenching with $\sim 1.5 \mu\text{M}$ terminator tiles (section 6 of the Supporting Information). (B and C) Scans of the experimental gel profiles (gray lines) and their deconvolution (black lines) for DNA and PDNA, respectively. (D) Plot of the dsDNA equivalent size of PDNA (\blacktriangle) and DNA (\square) polymers vs the number of units. Slopes correspond to 67 and 40.5 bp/unit for PDNA and DNA, respectively. (E) Quantification as equivalent masses of bands from panels B (white bars) and C (black bars) as a function of polymer size. The error in band quantification depends on methods but does not exceed a relative error of 10–15%.

For further comparison, the size distribution of polymer species was analyzed by gel electrophoresis and profiles were deconvoluted as previously described²¹ to calculate the electrophoretic mobility and for quantification (Figure 8A–C). However, in this case, terminator tiles (namely, am, Ma', bm, and Mb') able to cap the free ssDNA parts (A, A', B, and B', respectively) of polymer extremities have to be added to quench the polymerization and allow gel electrophoresis analysis. The quenching process was incorporated into the simulation model (sections 5 and 6 of the Supporting Information).²¹ For DNA and PDNA polymers, the apparent size was found to be proportional to the number of repeats with an additional contribution of single-stranded quencher ends (Figure 8D). As expected, the size step between bands appeared to be significantly larger for PDNA (~ 65 bp equivalent) than for DNA (40 bp equivalent) because of the contribution of protein domains to PDNA. However, relative distributions of polymer concentration obtained in DNA–DNA and PDNA–DNA experiments were found to be similar (Figure 8E) and consistent with simulations. At this point, we can conclude that DNA and PDNA recognition in solution is fully driven by the nucleic acid part with almost undetectable interference of the protein domain, except for hydrodynamic effects associated with size differences.

Self-Assembly of PDNAs on Vesicles and Supported Membranes. The protein domain of PDNA carries a C-terminal His₄ tag that can form a complex with nickel, cobalt, or copper divalent ions, mediating association through a surface and thus constraining self-assembly mechanisms. While the affinity of short His tag sequences for the nickel-nitriloacetate linker is sufficient for tight binding on affinity columns, the lifetime of such a complex is fairly short, generally a few seconds.^{14,30–32}

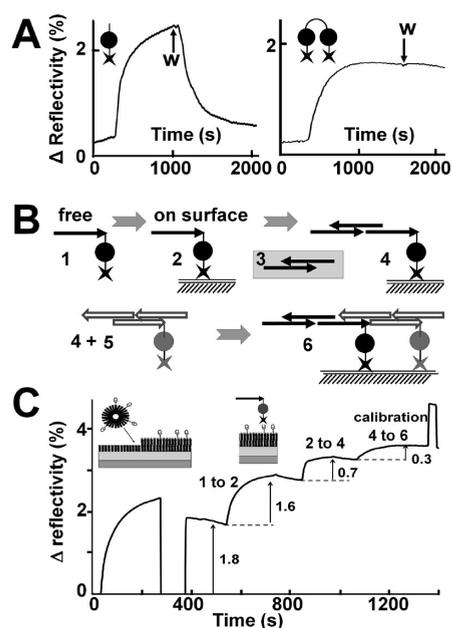


Figure 9. SPR analysis of the interaction of PDNA assemblies and of their b_5 component with phospholipids. (A) Binding kinetics of the b_5 component of PDNAs in the monomeric state (left) or following protein dimerization through a disulfide bond (right). The W label denotes the start of the washing phase. (B) Step by step self-assembly of PDNAs on a membrane surface. Free PDNA (1), supported membrane-bound PDNA (2), free DNA–tile dimer (3), coupling product of steps 2 and 3 (4), free hybrid (5) between one PDNA and two DNA tiles, and coupling product of steps 4 and 5 (6). (C) SPR sensorgram corresponding to panel B. The numbers correspond to steps depicted in panel B. The differential reflectivity for each step is indicated by the arrows.

Binding of the PDNA protein part to the phospholipid bilayer supplemented with DOGS-Ni was evaluated using 100 nm calibrated liposomes. The protein and phospholipid saturation curves are illustrated in section 7 of the Supporting Information. Half-saturation for the monomeric b_5 domain used in PDNA construction was found to be $\sim 5 \mu\text{M}$, and the binding capacity was clearly proportional to the total phospholipid concentration, leading to an average binding surface of $\sim 50 \text{ nm}^2$ for a single b_5 unit.

Assembly of PDNA on a membrane was further investigated using a DOGS-NTA-Ni-doped hemimembrane supported on gold, which allows monitoring of assembly by surface plasmon resonance (SPR). Figure 9A illustrates that the binding of the protein part of PDNA to the supported membrane is efficient but rapidly reversible with an affinity in the micromolar range, confirming the previous observation with vesicles. However, when the protein parts form a dimeric structure, a stable association is observed with a dissociation rate hardly measurable (hour range). This is due to a rebinding process in the diffusion limit layer immediately surrounding the surface when at least two linkages exist between the object and the surface. Such a situation typically occurs when two PDNA tiles become linked through hybridization with a cDNA tile. This principle is used in the experiment described in Figure 9B and is illustrated by the SPR sensorgram in Figure 9C. Formation of a supported membrane hemilayer is evidenced by the fusion of DOGS-Ni-supplemented phospholipid vesicles onto alkane thiol-modified gold. Binding of PDNA to the Ni-NTA lipid head (steps 1 and 2) appeared to be

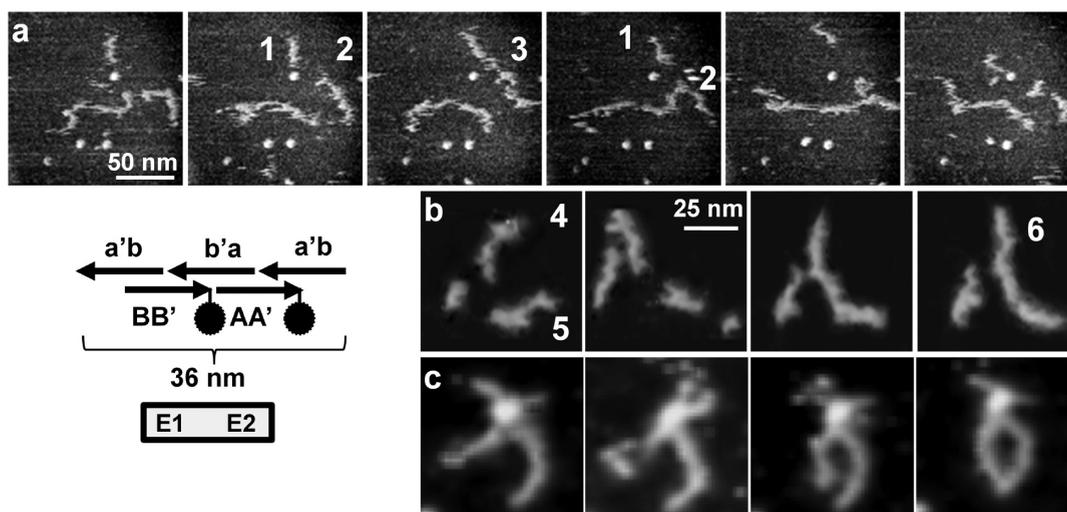


Figure 10. High-speed AFM analysis of the dynamics of PDNA structures. (a and b) Dynamic AFM imaging (10 frames/s) evidencing the formation of transient concatemers (species 1 and 2 giving 3 and 4 and 5 giving 6) of the construct described in the scheme. (c) Dynamics of two constructs anchored onto a nanoparticle. Images from lane a were untreated, while nonrelevant background noise was removed for the sake of readability in lanes b and c.

efficient, and the trace clearly illustrates that the bound PDNA can be easily hybridized (steps 2–4) with the complementary extremity of a DNA–DNA half-hybrid that acts as a linker between two PDNAs (steps 4–6) finally stabilizing the whole construct on the surface. However, the addition of mass following hybridization appeared to be significantly below the expected stoichiometry, and the defect increased with a larger number of assembly steps. This could be explained by the fact that binding of additional PDNA tiles to the surface is sterically impaired by tiles already associated on the surface. Reducing the effect would need careful control of surface saturation during assembly initiation. However, this would suppose working under conditions far from surface saturation, which would in turn dramatically reduce the magnitude of the SPR signal, making multiple-step assemblies difficult to monitor.

AFM Analysis of the Dynamics and Interactions of Short PDNA Assemblies. A construct encompassing two PDNA tiles (scheme in Figure 10) was prepared and adsorbed onto freshly cleaved mica in the presence of magnesium. High-speed AFM imaging evidenced lateral mobility with large changes in position (tenths of a nanometer) and shapes of objects within the 100 ms time lapse between two scans (Figure 10a). An overlay of consecutive images separated by a 100 ms time lapse on a $300 \text{ nm} \times 300 \text{ nm}$ scale illustrates the presence in the population of lower-mobility segments, while other regions of the molecules were rapidly laterally diffusing (image sequence available in section 9 of the Supporting Information). Surface defects, such as nanoparticles a few nanometers in diameter, can very efficiently trap diffusing PDNA structures, anchoring them onto a specific point of the surface when the remaining parts are still stochastically moving around leading to tweezerslike structures (Figure 10a, lane 3). Interestingly, two or more constructs can frequently interact together, resulting in the apparent fusion of objects to give unstable longer species (Figure 10a, lane 2). The events appeared to be mostly conservative with respect to length (daughter species has a length close to the sum of the parents), a feature expected for classical base pairing. However, the construct exhibits single-stranded extremities that are not complementary and cannot give rise to a stable duplex. Consistently, the

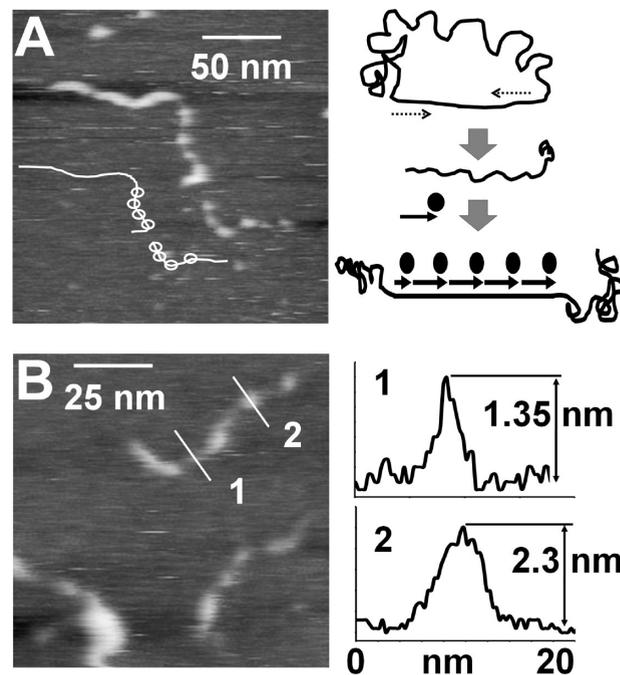


Figure 11. Structures of immobilized PDNA assemblies. (A) High-speed AFM imaging (raw data) of two assemblies and a schematic description of the image. (B) Three other objects are illustrated at a higher resolution (raw data), and section profiles of the DNA (1) and protein (2) parts are given. Constructs encompassing five PDNA units hybridized to a large single-stranded DNA were prepared as outlined in the right part of the scheme (see section 8 of the Supporting Information for details).

species appeared to be monodisperse in gel electrophoresis. However, formation during gel filtration of shoulder peaks with molecular masses greater than that of the main species suggested weak interactions of construct extremities, consistently with predicted ΔG values of -7.8 (b/b), -7.5 (a'/a'), and -3.6 kcal/mol (b/a') for the formation of intermolecular hydrogen

bonds in a salty solution (section 10 of the Supporting Information). Such energies are sufficient to create short life (few seconds) but easily observable concatemers under dynamic AFM conditions. Statistical analysis of the length distribution of the concatemer (one to eight repeats) was performed on 106 objects, and the resulting profiles were found to be consistent with a model in which single-stranded extremities of the species interact together on the surface with ΔG values of -1.8 and -0.8 kcal/mol, respectively (section 11 of the Supporting Information). The observation was found to be identical upon imaging of DNA and PDNA of identical nucleotidic structures. As a control, concatenation events were never observed with a dsDNA (513 bp) that does not include single-stranded extremities.

Structural Analysis of Large PDNA Structures. It could appear surprising that the protein domains were hardly visible under previously used experimental conditions. However, the linker attaching DNA and protein domains in PDNA is sufficiently flexible to allow fairly different dynamics for the two components. Under previously used conditions, the scanning rate probably remained too slow compared to the dynamics of the protein domain to capture the 2 nm globular structural details. To limit the overall dynamics, larger objects were designed, including a 513-base single-stranded DNA scaffold on the middle of which up to five PDNA tiles can be hybridized (Figure 11A). As previously reported,³³ ssDNA segments generate tightly folded regions strongly interacting with the substrate. Under such conditions, a string of protein domains (thickness of 2.3 nm) regularly placed (10–15 nm) along the DNA filament (thickness of 1.35 nm) can clearly be imaged by high-speed AFM. Control involving dsDNA with same sequence illustrates a filament with a homogeneous thickness of ~ 1.3 nm (not shown). This illustrates that PDNA building blocks can be easily used to decorate a DNA scaffold.

DISCUSSION

Covalent coupling of protein and DNA domains into a single building block is a versatile alternative to noncovalent complexes for building proteo–nucleic structures.^{34–37} Different PDNAs encompassing a cytochrome b_5 protein domain linked to a ssDNA were designed, synthesized, and purified to homogeneity. The C-terminal His tag on the protein part offers a way to create more specific and localized interactions with substrates than DNA alone that nonspecifically interact with ion-doped material like Ni^{2+} -treated mica.³³ The requirement to build PDNAs is the availability of a unique grafting position in the protein sequence. Use of a surface cysteine residue is a method of choice, but alternate solutions include the use of a rare or non-natural amino acid like a selenocysteine or an *N*-acetylphenylalanine,³⁸ which can be selectively grafted using adapted chemistry. A generic approach could also involve grafting of the surface cysteine opposite the CDR3 binding region onto small (14 kDa) VHH single-chain antibodies. Such antibodies could be used in turn to form a noncovalent high-affinity complex with any protein of interest.³⁹

This work specifically aimed to compare the physicochemical and self-assembly properties of DNA and PDNA in different environments. Defining and controlling parameters implied in self-assembly mechanisms of such hybrid synthetic structures will permit us not only to use them as modular building blocks for multifunctional nanostructures but also to improve our understanding of properties of natural nucleo–proteic complexes.

A contribution of the protein domain to the hydrodynamic properties of PDNAs was clearly evidenced by gel filtration and electrophoresis in SDS–PAGE and nondenaturing agarose gels. In all cases, the protein contribution appeared to be additive with respect to the one of the nucleic acid parts provided to use an equivalent DNA length that differs depending on the analysis method considered. During gel filtration, the 13 kDa protein domain migrates like a solvated dsDNA of 6.6 kDa. This contrasts with an equivalent mass of 16–17 kDa observed in nondenaturing agarose gel electrophoresis calibrated using dsDNA. This last value is consistent with a mass of 15 kDa for the protein domain in SDS–PAGE following calibration with proteins. This is in agreement with a common extended conformation of the reference molecules (DNA and protein) used for calibration in an agarose gel and during SDS–PAGE, respectively. During HPLC gel filtration, hydrodynamic constraints resulting from flow are expected to favor a rather extended conformation of the DNA used for size calibration. In contrast, the protein domain remains globular, which tends to decrease its relative apparent size. Reciprocally, during agarose gel electrophoresis, the much lower charge of the protein domain compared to that of DNA decreases the relative protein mobility, thus compensating for the more compact structure of the protein domain. The strictly additive contributions of double-stranded nucleic and protein domains to hydrodynamic behaviors for a variety of PDNA assemblies suggested that conformational interactions between the two components are likely weak under all tested conditions and do not affect their individual shapes.

In contrast to pure nucleic structures that are generally assembled under conditions close to the melting temperature to avoid formation of dead-end parasitic hybrids,^{6,20} proteo–nucleic structures must be preferentially assembled in a kinetically controlled manner,^{40–42} far below the melting temperature to prevent thermal denaturation of protein domains. Hybridization rates are thus critical parameters of PDNAs that were analyzed by monitoring hypochromic effects associated with hybridization in rapid-mixing experiments. While the length of hybridizing segments is homogeneous for all considered tiles, observed hybridization rates significantly differ over a 4-fold range. This dispersion does not simply correlate with obvious sequence features and for the same hybridizing segment depends on the nature of flanking regions. The reason is likely related to the formation of possible secondary structures of ssDNA parts, some of them having predicted stability sufficiently large (8–14 kcal/mol) to significantly interfere. In contrast, comparison of the hybridization rates of DNA and corresponding PDNA does not evidence a significant effect of the protein domain when grafted on the oligonucleotide 3'-end. However, a moderate decrease in hybridization rates was observed for internal grafting proximal to the hybridizing segment. In this case, base modification by the linker and protein presence exhibited cumulative effects. When hybridization between larger polymers was considered, no additional effect was observed compared to the situation with short building blocks, suggesting the absence of long-range interaction between PDNAs when formation of linear polymer was considered.

Altogether, our results indicate that DNA and PDNA tiles are mostly equivalent at the level of their hybridization properties and that PDNA can easily substitute for DNA in most DNA-based supramolecular assemblies, provided that the folding can sterically accommodate the protein parts. A potential limitation remains the thermal stability of the protein domain under the

fairly high-temperature conditions required to prevent formation of a dead-end product by spurious hybridization.^{6,43} However, analysis of the thermal stability of our constructs evidenced that linkage to the nucleic acid part has no significant detrimental influence on protein stability leading, in our case, to constructs stable up to 55 °C, a value that is compatible with current hybridization procedures. Thus, any reasonably stable protein can be substituted for our cytochrome *b*₅ domain, including recognition protein-like single-chain antibodies of the VHH family. In the case of very unstable protein domains, an alternative assembly procedure under kinetic control remains possible using, for example, DNA hairpin or loop motifs as programmable units.^{41,44} A combination of kinetic and thermodynamic control can also be used to direct self-assembly processes by engineering intra- and interparticle DNA hybridization.^{5,45}

PDNAs potentially offer many advantages when self-assemblies on surfaces are targeted. The protein domain can be used, for example, for the recognition of surface chemical or physical features using material specific affinity domains.⁴⁶ This also applies to redox active proteins (like cytochrome *b*₅), for which electron transfer properties at surface interfaces can be modulated by the conformation of the DNA scaffold, thus constituting biosensors. Analysis by rapid-scan AFM clearly evidenced lateral diffusion properties of PDNA assemblies on surfaces. SPR experiments demonstrated that such structures can be sequentially built by stepwise extension from a matrix-bound primer¹⁹ or by self-organization of tiles attached to membranes (this work). However, the progressive yield decrease during constructions resulted either from surface saturation or, as shown by HS-AFM, from the presence of surface defects able to anchor laterally diffusing PDNA molecules. A possible turnaround to build defect-free surfaces would be the use of phospholipid vesicles as assembly frame. This work illustrates that the fairly limited stability of the attachment of PDNA to NTA-modified phospholipids is a limitation, which could be somewhat overcome when several protein modules are bound to the same vesicles. Alternatively, multiple-head NTA constructs can be used to increase affinity.^{13,14,30–32} However, the fairly rigid structure of large PDNA–DNA supramolecular assemblies was found to conflict with the strong curvature of small (50–100 nm) vesicles and more importantly can reciprocally induce an unwanted phase transition of the support phospholipid structures, leading to uncontrolled organization.

Formation of transient end-to-end concatemers was evidenced by HS-AFM upon lateral diffusion of PDNA assemblies on the surface. Similar behavior was observed when DNA was substituted with PDNA under the same experimental conditions. The phenomenon was clearly related to the presence of ssDNA extensions on constructs that can form short stretches of paired bases (four to six bases) and was not observed with fully dsDNA. Interaction energies deduced from modeling of the concatemer distribution appeared to be ~4-fold smaller than calculated values in solution, which is consistent with the much lower salt concentration used under AFM imaging conditions. Formation of these transient structures while they are reversible competes with true hybrid formation during assemblies slowing the process on surfaces. The phenomena could be enhanced by the presence of surface defects that efficiently trap diffusing PDNA assemblies and by the observed topological cluttering on the surface (AFM and SPR experiments). These three effects combine in reducing the self-assembly efficiency of surface-bound objects.

While high-speed AFM opens a new dimension for the analysis at the single-molecule level of nucleic acid assemblies, difficulties could arise from the relative values of scan rate and object dynamics. Our data evidence that the small *b*₅ domain (13 kDa) was not visible when attached to highly dynamic structures, whereas it was clearly imaged when bound to much larger DNA exhibiting a lower mobility. Thus, it seems that some compromise has to be found between resolution and dynamic analyses, particularly when very small objects are involved. While dynamic AFM observation of a fairly small protein (calmodulin, 17 kDa) bound to a streptavidin crystal was recently reported,⁴⁷ our results go further than previous reports⁴⁸ by visualizing for the first time a small protein domain (13 kDa) linked to isolated DNA molecules in solution.

■ ASSOCIATED CONTENT

S Supporting Information. DNA tile sequences and matches (section 1), schematic for the synthesis of protein–DNA (PDNA) structures (section 2), synthesis of PDNA structures (section 3), contribution of building blocks to the apparent size in gel filtration and electrophoresis (section 4), experimental and simulated profiles for PDNA copolymers (section 5), parameter sets for simulations (section 6), cytochrome *b*₅ and phospholipid concentration dependence for the formation of protein–lipid complexes (section 7), purification of the DNA and PDNA assemblies described here (section 8), high-speed AFM image sequences (section 9), thermodynamic parameters (section 10), and statistical analysis of the formation of concatemers (section 11). This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Telephone: +33 567048806. Fax: +33 561559400. E-mail: denis.pompon@insa-toulouse.fr. Present address: LISBP/INSA, UMR INSA/CNRS 5504, 135 Avenue de Rangueil, 31077 Toulouse cedex 4, France.

Funding Sources

We acknowledge the financial support of this work by Grants ANR-05-NANO-021-04, ANR-08-NANO-010-01, and ANR-08-PCV08_343399 from the Programme Nanosciences et Nanotechnologies (PNANO) and Programme interdisciplinaire en physique et chimie du vivant (PCV) actions of the National Research Agency (ANR), France.

■ ACKNOWLEDGMENT

We thank Dr. Takayuki Uchihashi (Physics Laboratory, Kanazawa University), Drs. Ignacio Casuso and Simon Schauring (Institute Curie, Paris, France), Dr. Pierre-Emmanuel Milhiet (CBS, Montpellier, France) for technical support, and Dr. Gilles Truan for proofreading the manuscript.

■ REFERENCES

- (1) Brucala, M., Zuccheri, G., and Samori, B. (2006) Mastering the complexity of DNA nanostructures. *Trends Biotechnol.* 24, 235–243.
- (2) He, Y., Tian, Y., Chen, Y., Deng, Z., Ribbe, A. E., and Mao, C. (2005) Sequence Symmetry as a Tool for Designing DNA Nanostructures. *Angew. Chem., Int. Ed.* 44, 6694–6696.

- (3) Ke, Y., Liu, Y., Zhang, J., and Yan, H. (2006) A study of DNA tube formation mechanisms using 4-, 8-, and 12-helix DNA nanostructures. *J. Am. Chem. Soc.* 128, 4414–4421.
- (4) Liu, Y., Ke, Y., and Yan, H. (2005) Self-assembly of symmetric finite-size DNA nanoarrays. *J. Am. Chem. Soc.* 127, 17140–17141.
- (5) Rotello, V. M. (2009) DNA nanotechnology: Hot and sticky or cold and aloof. *Nat. Mater.* 8, 539–540.
- (6) Rothmund, P. W. (2006) Folding DNA to create nanoscale shapes and patterns. *Nature* 440, 297–302.
- (7) Seeman, N. C. (1998) DNA nanotechnology: Novel DNA constructions. *Annu. Rev. Biophys. Biomol. Struct.* 27, 225–248.
- (8) Teller, C., and Willner, I. (2010) Organizing protein-DNA hybrids as nanostructures with programmed functionalities. *Trends Biotechnol.* 28, 619–628.
- (9) Yan, H., Park, S. H., Finkelstein, G., Reif, J. H., and LaBean, T. H. (2003) DNA-templated self-assembly of protein arrays and highly conductive nanowires. *Science* 301, 1882–1884.
- (10) Park, S. H., Pistol, C., Ahn, S. J., Reif, J. H., Lebeck, A. R., Dwyer, C., and LaBean, T. H. (2005) Finite-Size, Fully Addressable DNA Tile Lattices Formed by Hierarchical Assembly Procedures. *Angew. Chem., Int. Ed.* 45, 735–739.
- (11) Beyer, S., Nickels, P., and Simmel, F. C. (2005) Periodic DNA nanotemplates synthesized by rolling circle amplification. *Nano Lett.* 5, 719–722.
- (12) Liu, Y., Lin, C., Li, H., and Yan, H. (2005) Aptamer-directed self-assembly of protein arrays on a DNA nanostructure. *Angew. Chem., Int. Ed.* 44, 4333–4338.
- (13) Conti, M., Falini, G., and Samori, B. (2000) How Strong Is the Coordination Bond between a Histidine Tag and Ni - Nitrilotriacetate? An Experiment of Mechanochemistry on Single Molecules. *Angew. Chem., Int. Ed.* 39, 215–218.
- (14) Lata, S., Reichel, A., Brock, R., Tampe, R., and Pehler, J. (2005) High-affinity adaptors for switchable recognition of histidine-tagged proteins. *J. Am. Chem. Soc.* 127, 10205–10215.
- (15) Boireau, W., Bombard, S., Sari, M. A., and Pompon, D. (2002) Bioengineering and characterization of DNA-protein assemblies floating on supported membranes. *Biotechnol. Bioeng.* 77, 225–231.
- (16) Boireau, W., Duncan, A. C., and Pompon, D. (2005) Bioengineering and characterization of DNA-protein assemblies floating on supported membranes. *Methods Mol. Biol.* 300, 349–368.
- (17) Boireau, W., Zeeh, J. C., Puig, P. E., and Pompon, D. (2005) Unique supramolecular assembly of a redox protein with nucleic acids onto hybrid bilayer: Towards a dynamic DNA chip. *Biosens. Bioelectron.* 20, 1631–1637.
- (18) Berthier, A., Elie-Caille, C., Lesniewska, E., Delage-Mourroux, R., and Boireau, W. (2008) Nanobioengineering and Characterization of a Novel Estrogen Receptor Biosensor. *Sensors* 8, 4413–4428.
- (19) Pompon, D., and Laisne, A. (2007) PDNA as building blocks for membrane-guided self-assemblies. *Biochem. Soc. Trans.* 35, 495–497.
- (20) Rothmund, P. W., Papadakis, N., and Winfree, E. (2004) Algorithmic self-assembly of DNA Sierpinski triangles. *PLoS Biol.* 2, e424.
- (21) Laisne, A., Lesniewska, E., and Pompon, D. (2011) Design and experimental validation of a generic model for combinatorial assembly of DNA tiles into 1D-structures. *Biochim. Biophys. Acta* 1810, 603–611.
- (22) Ando, T., Kodera, N., Takai, E., Maruyama, D., Saito, K., and Toda, A. (2001) A high-speed atomic force microscope for studying biological macromolecules. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12468–12472.
- (23) Ando, T., Uchihashi, T., Kodera, N., Yamamoto, D., Miyagi, A., Taniguchi, M., and Yamashita, H. (2008) High-speed AFM and nanovisualization of biomolecular processes. *Pfluegers Arch.* 456, 211–225.
- (24) Ando, T., Uchihashi, T., Kodera, N., Yamamoto, D., Taniguchi, M., Miyagi, A., and Yamashita, H. (2007) High-speed atomic force microscopy for observing dynamic biomolecular processes. *J. Mol. Recognit.* 20, 448–458.
- (25) Kitazawa, M., Shiotani, K., and Toda, A. (2003) Batch Fabrication of Sharpened Silicon Nitride Tips. *Jpn. J. Appl. Phys.* 42, 4844–4847.
- (26) Garcia De La Torre, J., Huertas, M. L., and Carrasco, B. (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* 78, 719–730.
- (27) Robert, C. H. (1995) Estimating friction coefficients of mixed globular/chain molecules, such as protein/DNA complexes. *Biophys. J.* 69, 840–848.
- (28) Reynolds, J. A., and Tanford, C. (1970) The gross conformation of protein-sodium dodecyl sulfate complexes. *J. Biol. Chem.* 245, 5161–5165.
- (29) Sorlie, S. S., and Pecora, R. (1990) A dynamic light scattering study of four DNA restriction fragments. *Macromolecules* 23, 487–497.
- (30) Huang, Z., Hwang, P., Watson, D. S., Cao, L., and Szoka, F. C., Jr. (2009) Tris-nitrilotriacetic acids of subnanomolar affinity toward hexahistidine tagged molecules. *Bioconjugate Chem.* 20, 1667–1672.
- (31) Huang, Z., Park, J. I., Watson, D. S., Hwang, P., and Szoka, F. C., Jr. (2006) Facile synthesis of multivalent nitrilotriacetic acid (NTA) and NTA conjugates for analytical and drug delivery applications. *Bioconjugate Chem.* 17, 1592–1600.
- (32) van Broekhoven, C. L., and Altin, J. G. (2005) The novel chelator lipid 3-(nitrilotriacetic acid)-ditetradecylamine (NTA(3)-DTDA) promotes stable binding of His-tagged proteins to liposomal membranes: Potent anti-tumor responses induced by simultaneously targeting antigen, cytokine and costimulatory signals to T cells. *Biochim. Biophys. Acta* 1716, 104–116.
- (33) Maeda, Y., Matsumoto, T., and Kawai, T. (1999) Observation of single- and double-stranded DNA using non-contact atomic force microscopy. *Appl. Surf. Sci.* 140, 400–405.
- (34) Endo, M., and Sugiyama, H. (2009) Chemical approaches to DNA nanotechnology. *ChemBioChem* 10, 2420–2443.
- (35) Niemeyer, C. M. (2001) Semi-synthetic nucleic acid-protein conjugates: Applications in life sciences and nanobiotechnology. *J. Biotechnol.* 82, 47–66.
- (36) Niemeyer, C. M. (2007) Functional devices from DNA and proteins. *Nano Today* 2, 42–52.
- (37) Niemeyer, C. M. (2010) Semisynthetic DNA-protein conjugates for biosensing and nanofabrication. *Angew. Chem., Int. Ed.* 49, 1200–1216.
- (38) Liu, C. C., and Schultz, P. G. (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* 79, 413–444.
- (39) Kastelic, D., Frkovic-Grazio, S., Baty, D., Truan, G., Komel, R., and Pompon, D. (2009) A single-step procedure of recombinant library construction for the selection of efficiently produced llama VH binders directed against cancer markers. *J. Immunol. Methods* 350, 54–62.
- (40) Lubrich, D., Green, S. J., and Turberfield, A. J. (2009) Kinetically controlled self-assembly of DNA oligomers. *J. Am. Chem. Soc.* 131, 2422–2423.
- (41) Shih, W. (2008) Biomolecular assembly: Dynamic DNA. *Nat. Mater.* 7, 98–100.
- (42) Wieland, M., Benz, A., Haar, J., Halder, K., and Hartig, J. S. (2010) Small molecule-triggered assembly of DNA nanoarchitectures. *Chem. Commun.* 46, 1866–1868.
- (43) Winfree, E., Liu, F., Wenzler, L. A., and Seeman, N. C. (1998) Design and self-assembly of two-dimensional DNA crystals. *Nature* 394, 539–544.
- (44) Yin, P., Choi, H. M., Calvert, C. R., and Pierce, N. A. (2008) Programming biomolecular self-assembly pathways. *Nature* 451, 318–322.
- (45) Leunissen, M. E., Dreyfus, R., Cheong, F. C., Grier, D. G., Sha, R., Seeman, N. C., and Chaikin, P. M. (2009) Switchable self-protected attractions in DNA-functionalized colloids. *Nat. Mater.* 8, 590–595.
- (46) Seeman, N. C. (2002) Key experimental approaches in DNA nanotechnology. *Current Protocols in Nucleic Acid Chemistry*, Chapter 12, Unit 12, p 1, Wiley, New York.
- (47) Yamamoto, D., Nagura, N., Omote, S., Taniguchi, M., and Ando, T. (2009) Streptavidin 2D crystal substrates for visualizing biomolecular processes by atomic force microscopy. *Biophys. J.* 97, 2358–2367.
- (48) Lyubchenko, Y. L., and Shlyakhtenko, L. S. (2009) AFM for analysis of structure and dynamics of DNA and protein-DNA complexes. *Methods* 47, 206–213.

Supporting Information for:

Self-assembly properties and dynamic of synthetic proteo-nucleic building blocks in solution and on surfaces

Aude LAISNE[†], Maxime EWALD[‡], Toshio ANDO[§], Eric LESNIEWSKA[‡], and Denis POMPON^{†*}

[†] Centre de Génétique Moléculaire. CNRS.UPR3404, Avenue de la Terrasse, F91190 Gif-sur-Yvette, France.

[‡] Institut Carnot Bourgogne UMR CNRS 5209, University of Bourgogne, F-21078 Dijon, France.

[§] Department of Physics, Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan.

***Corresponding Author:** Denis POMPON

*To whom correspondence should be addressed. Tel: +33 169823683; Fax: +33 169823682;
Email: pompon@cgm.cnrs-gif.fr

Present address:

Centre de Génétique Moléculaire, CNRS.UPR3404, Avenue de la Terrasse, F91190 Gif-sur-Yvette, France.

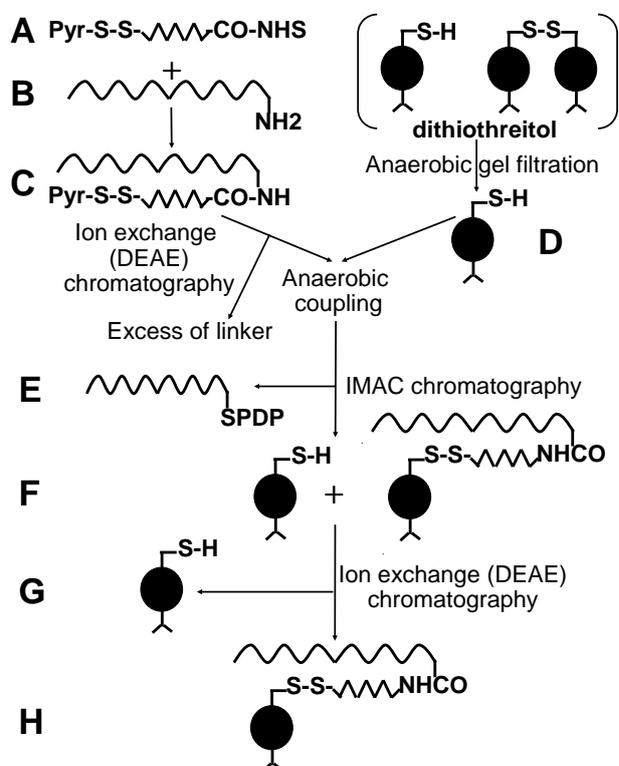
Section 1. DNA tile sequences and matches.

Tiles	Sequence (5' - 3')	Bases
A---A'	CCGTAGGGTCTCCATTTTGA---AGCGAAGTCACACTGTACA	39
B---B'	GTCCATATGCAAGTCAGCGT---CAGTGTACAGT <u>C</u> ACGTCAG	39
a---b'	TCAAAATGGAGACCCTACGG---CTGACGTGACTGTACACTG	39
b---a'	ACGCTGACTTGCATATGGAC---TGTACAGTGTGACTTCGCT	39
a---m	TCAAAATGGAGACCCTACGG---GTAAAACGACGGCCAG	36
b---m	ACGCTGACTTGCATATGGAC---GTAAAACGACGGCCAG	36
M---a'	CAGGAAACAGCTATGAC---TGTACAGTGTGACTTCGCT	36
M---b'	CAGGAAACAGCTATGAC---CTGACGTGACTGTACACTG	36

Matching tile moieties are $\begin{matrix} \text{A} \\ \rightleftharpoons \\ \text{a} \end{matrix}$ $\begin{matrix} \text{A}' \\ \rightleftharpoons \\ \text{a}' \end{matrix}$ $\begin{matrix} \text{B} \\ \rightleftharpoons \\ \text{b} \end{matrix}$ $\begin{matrix} \text{B}' \\ \rightleftharpoons \\ \text{b}' \end{matrix}$

BB' sequence was synthesized in three versions: without, with a 3' amino-linker and with an internally amino-linker coupled thymidine base (underlined). In the case of the AA' sequence, only the unmodified and 3' amino-linker versions were used. The amino-linker modifications were used for protein derivatization in PDNAs.

Section 2. Schematic for the synthesis of protein-DNA (PDNA) structures.



Section 3. Synthesis of protein-DNA (PDNA) structures.

Genetic Engineering of Cytochrome b5: A human microsomal cytochrome b5 cDNA (GenBank accession number L39792) has been previously engineered to substitute the 26 C-terminal residues constituting the hydrophobic membrane anchor tail by the -NGHHHH-COOH sequence (16) encoding an His-tag. Engineered cDNA was subcloned into pCR[®]2.1-TOPO[®] cloning (Invitrogen) and the *SphI*-*Bam*H1 fragment was ligated into the corresponding sites of digested pUHE25-2 expression vector giving pUHE-[Hb5-(HIS₄)C-term]. This construction includes an additional Leu codon inserted immediately after the normal Met initiation codon. Site directed mutagenesis of plasmid was performed with QuickChange II XL Site-Directed Mutagenesis kit (Stratagene) to introduce a Ser-24-Cys aa-sequence change. This brought a unique solvent accessible thiol-containing residue at the surface of the b5 globular domain. Mutation has been checked by sequencing resulting in the pUHE-[Hb5-24C-(HIS₄) C-term] expression plasmid.

Protein expression and purification: XL1-Blue *E. coli* cells transformed by pUHE-[Hb5-24C-(HIS₄)C-term] were transferred in 50 mL liquid Luria-Bertani containing 100 µg/mL ampicillin and grown over-night at 37°C, 160 rpm. 45 mL of this culture were transferred into 1 L of Terrific Broth (TB) containing 100 µg/mL ampicillin and bacteria were grown at 28°C for 24 hrs. 1 mL aliquotes of the culture were stored at -80°C in 1X Hogness. Protein expression was induced with 0.5 mM IPTG and 0.5 mM δ-ALA added with 500 mL of TB and occurred at 28°C for 24 hrs with shaking at 160 rpm. Cells were harvested after 10 minutes 8500 x g centrifugation at 4°C, washed 2 times with 50 mL 20 mM Na/K phosphate, 0.5 M NaCl, pH 6.5 in the presence of 1 mM PMSF, and suspended into the same buffer with anti-protease cocktail. Cells lysis was performed by repeated sonication following treatment with 1.25 g/L lysozyme, 50 mg/L ribonuclease A and 50 mg/L DNase I. After removal of cell debris by centrifugation at 18500 g for 10 min, the supernatant was applied onto a 1.5 x 6 cm column of imminodiacetate immobilised on agarose (Sigma) equilibrated successively with 0.05 M NiCl₂ in 0.1 M KAc, 0.4 M KCl adjusted to pH4.5, with HCl, washed with the same buffer without NiCl₂ and finally equilibrated with 50 mM Na/K phosphate buffer pH 7.4 (buffer A). For elution, column was washed with buffer A containing increasing NaCl concentrations (0.1 M to 0.5 M) until OD₂₈₀ of the eluate was lower than 0.3. The protein was eluted with 30 mM L-histidine solution in buffer A containing 1 mM PMSF. 5 mM EDTA was added to fractions containing the protein before loading onto a DEAE-Sepharcel column (1.5 x 9 cm) equilibrated with buffer A. The column was washed with buffer A supplemented

with 0.1 M NaCl. The protein was eluted with buffer A containing 0.35 M NaCl. Fractions were pooled and dialysed against 1 L of 20 mM Na/K phosphate, 1 mM DTT pH 6.5. Eluate was concentrated using a Vivaspin 10000 MWCO device (Vivascience) before loading on a gel filtration column (Superdex 75 prep grade Amersham Biosciences) 1.6 x 50 cm equilibrated with 20 mM Na/K phosphate pH 6.5 and eluted at 2 ml/min. Concentration of the purified protein was determined by spectrophotometry using $\epsilon_{412\text{nm}}=118 \text{ mM}^{-1} \cdot \text{cm}^{-1}$.

Linker coupling to ss-DNA: the SPDP reagents (Pierce) are heterobifunctional cross-linkers (SI Section 2A). Succinimidyl 6-[3'-(2-pyridyldithio) propionamido]hexanoate (LC-SPDP) is a 15.6 Å long spacer arm which can form, on one side an amid bond with the 3' -amino-modified oligonucleotide and, on the other side a disulfide bond with the engineered protein : Hb5-24C-(His₄)Cterm. Poorly soluble in aqueous buffers, LC-SPDP was first dissolved in DMSO (90 mM) before all dilution. The 39 mers 3'-amino-modified ss-DNA (SI Section 2B) were coupled to the bifunctional linker in 50 mM Na/K phosphate buffer pH 8, using 15 mM LC-SPDP for a DNA concentration ranging between 50 µM to 100 µM. The reaction occurred at room temperature. The coupling efficiency was checked by ion-pairing reverse phase chromatography on a Waters XTerra MS C18 column (3.5 µM, 4.6 x 100 mm) equilibrated with 0.1 M triethylammonium acetate pH 7.4 in water and eluted at 1 mL/min using a linear gradient between the aqueous buffer and acetonitrile (0 to 21.4% by vol. in 10 min). Coupled ssDNA (SI Section 2C) were purified from excess of reagent and of hydrolysis subproducts by chloroform extractions prior loading onto an 0.6 x 2 cm anion exchange column (DEAE Sephacel) equilibrated with 50 mM Na/K phosphate pH 7.4 (buffer A). After washing with 0.25 M NaCl in buffer A, DNA was eluted with 1 M NaCl in buffer A. DNAs were desalted against water using Float-A-Lyzer (Fisher) Ø5 mm, MWCO 3,5 K and concentrated by speed vacuum without heating (concentrator 5301, Eppendorf). Concentrations of coupled oligonucleotides were calculated from their absorption spectra and were stored at -20 °C.

Hb5-24C(HIS)₄ coupling with LC-SPDP-DNA: To prevent a high level of spontaneous dimerisation of Hb5-24C(HIS)₄, the protein was reduced and the monomer separated from DTT excess by gel filtration on G25 sephadex column equilibrated in 10 mM Tris-HCl pH 8 1mM EDTA under anaerobic conditions (SI Section 2D). LC-SPDP-oligonucleotides were coupled for 24 h at room temperature with 1.5 molar excess of reduced engineered b5 in a hermetic flask filled with argon. Coupling products were analyzed by ion pairing reverse phase HPLC.

Protein-DNA complex purification: Combined chromatographies were used to separate the complex from non-coupled DNA and non-coupled engineered b5. The reaction mixtures were loaded separately onto a 0.5 x 1.5 cm nickel chelate column equilibrated with buffer A. Column was washed with buffer A to remove uncoupled oligonucleotides (SI Section 2E) and pyridine-thiol subproducts until A_{260} absorbance was no more detectable. The his-tagged proteins were eluted by a solution of 30 mM histidine in buffer A. DNA-coupled and non-coupled protein fractions (SI Section 2F) were pooled based on the 413 nm absorption spectra and were treated with 5 mM EDTA in order to disrupt potential nickel-proteine complexes. Fractions were applied onto a 0.5 x 3 cm DE52 anion exchange column equilibrated with buffer A. Following a wash step with 0.35 M NaCl in buffer A to elute protein which is not coupled to oligonucleotide (SI Section 2G), the PDNA complex (H) was finally eluted with 1 M NaCl in Buffer A. The concentrations of PDNA complexes were determined by spectrophotometry at 260 and 413 nm and the purity checked by reverse phase HPLC for chemical contaminants and by reducing and non-reducing SDS-PAGE gels for protein contaminants. Both controls demonstrated the absence of any detectable DNA, protein or chemical contamination.

Section 4. Contribution of building blocks to apparent size in gel filtration and electrophoresis.

Methods	Unit	ss-DNA ¹ (base)	ss-DNA ² (base)	ds-DNA (bp)	Prot (unit)
Gel filtration ^a	kDa (bp)	0.30 (0.50)	0.30 (0.50)	0.70 (1.00)	6.6 (10)
SDS-PAGE ^b	kDa (bp)	0.28 (0.46)	0.28 (0.46)	0.36 (0.51)	15.0 (25)
Agarose gel ^{a,c}	kDa (bp)	0.36 (0.60)	0.94 (1.55)	0.70 (1.00)	16.5 (25)
Agarose gel ^{a,d}	kDa (bp)	0.42 (0.69)	1.08 (1.78)	0.73 (1.04)	17.5 (26.5)

Coefficients were calculated from the least square adjustment of calculated to experimental values.

¹Contribution per base for m and M ss-DNA extensions in protein free constructs

²Contribution per base for m and M ss-DNA extensions in protein containing constructs

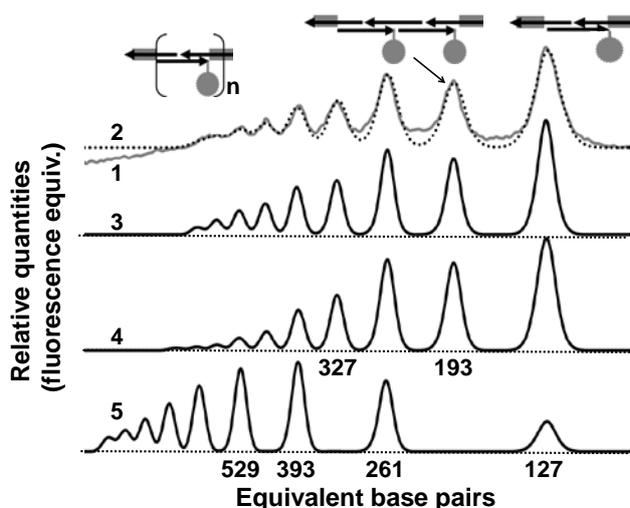
^aCalibrated with ds-DNA standards assuming an average of 330 Daltons per base.

^bCalibrated with protein standards. ^cSimple assemblies of PDNA and DNA from Figure 6.

^dData from Figure 8.

Section 5. Experimental and simulated profiles for PDNA copolymers.

Scan (line 1 on the next Figure) of the experimental PDNA gel profile from Figure 8A and its deconvolution using an optimized expansion factor of 1.5 (line 2) and regular resolution for ds-DNA (line 3). Simulations using parameters defined in SI Section 6 with (line 4) or without (line 5) taking into account the SIC value. Regular ds-DNA gel resolution was used for reconstruction.



Section 6. Parameter sets for simulations.

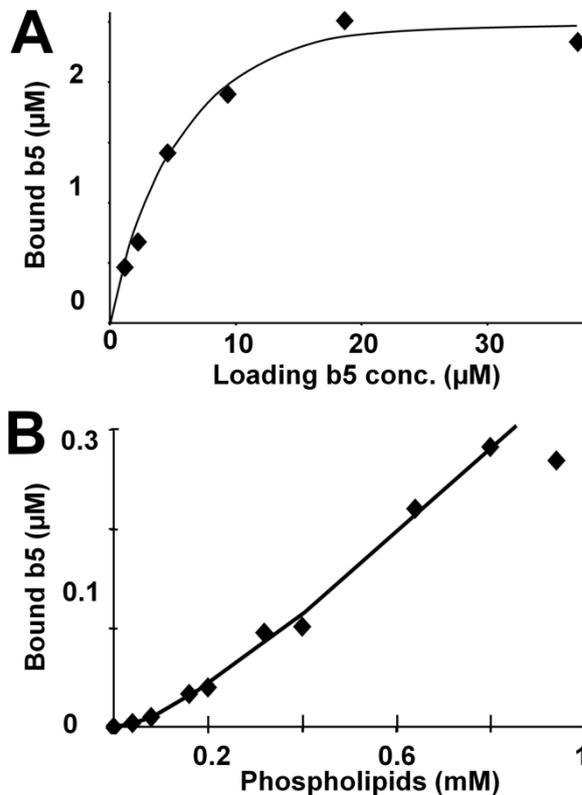
All simulations were performed and parameters defined as described in (21) and take into account the environment and length modulation of rate constants (with $RL = 19$ bases), and the J-Factor table PT2 of the same paper. For rapid-mixing experiments (Figure 7) final concentrations used were $0.28 \mu\text{M}$ of $[AA'PDNA-ba'DNA]$ and $[BB'PDNA-ab'DNA]$ hybrids (panel A) and $0.35 \mu\text{M}$ of $[AA'-ba'DNA]$ and $[BB'-ab'DNA]$ hybrids (panel B). For simulation of agarose gel profile in SI Section 5, concentrations were $AA'b5$ (0.7), $BB'b5$ (1.1), ab' (0.95), ba' (0.85), Ma' (1.54), Mb' (1.5), am (1.74), bm (1.38) μM . Rate constants used for simulation are given in the following table. Others parameters, for rapid-mixing experiments simulation, were hypochromic factors (Hf) of 11.3 and 10.4 % for PDNA and DNA, respectively. For the simulation of agarose gel profiles visualization parameters were $SSEH = 0.5$, $DSEQ$: yes, $DSSS = 1$, $DSDS = 1.05$ and $DSCS = 1$. A value of $0.09 \mu\text{M}$ for SIC parameter was only used for curve 4 in SI Section 5.

Rate constants ($\mu\text{M}^{-1} \cdot \text{min}^{-1}$)	PDNA	DNA	PDNA	DNA
	Type 1		Type 2	
A'/a'	5.6	7.0	8.8	11.1
A/a	3.7	3.7	5.8	5.8
B'/b'	4	3.4	6.3	5.4
B/b	5.5	4.1	8.7	6.5

Type 1 corresponds to the hybridization of ss-regions flanked by ss-regions. Type 2 refers to the hybridization of the ss-regions flanked by ds-regions.

Section 7. Cytochrome b5 and phospholipid concentration dependence for the formation of protein-lipid complexes.

A, saturation curve for the association of cytochrome b5 to liposomes using a total phospholipid concentration of 700 μM and variable b5 concentrations. B, same experiment using a constant b5 concentration of 10 μM and variable phospholipid concentrations. Freshly extruded (DOPC, DMPC, DOGS, 45: 45: 10, by mol.) phospholipids (0.7 mM) were



incubated for 10 min at 24°C with different concentrations of cysteine-containing protein (1.2, 2.3, 4.6, 9.2, 18.5 and 37 μM). The mix (70 μL) was diluted to 400 μL with 1 x PBS buffer, before loading onto a nickel-charged IMAC column (0.5 x 1 cm). The column was washed with 2 x 400 μL buffer. Spectra were recorded on each eluted fraction and the amount of unbound material calculated by deconvolution of the cytochrome b5 spectra. The procedure was repeated using variable concentrations of phospholipids and a fixed concentration of cytochrome b5.

Section 8. Purification of the DNA and PDNA assemblies described in the paper

Bi-block construction (AA'/ba', BB'/ab', AA'-b5/ba', BB'-b5/ab'). Corresponding tiles (PDNA or/and DNA) 1-20 μ M (depending on experiment) were hybridized in 1x PBS buffer for 1h at 42°C, then centrifugated prior purification by HPLC gel filtration performed at 0.3 ml/ min in 0.25 M NaCl, 0.15 M NaKPO₄ buffer at pH 6.8 using a 4 μ m UHR SEC, 4.6 x 300 mm column (Waters). When PDNA and DNA were hybridized together a slight molar excess (1.2 x) of the DNA tile was used to improve yield.

Tri-block construction (AA'/ba'/ab', BB'/ba'/ab', AA'/BB'/ab', AA'-b5/ba'/ab', BB'-b5/ba'/ab', AA'-b5/BB'-b5/ab'). The procedure was formally the same than for diblock synthesis except that three tiles were used.

Penta-block construction (AA'/BB'/ba'/ab'/ba', AA'-b5/BB'-b5/ba'/ab'/ba'): suitable HPLC purified di-block and tri-block (see previously) were mixed and let to hybridize at 42°C for 60 minutes prior purification by HPLC gel filtration.

Framework bound PDNA (Figure 11): The [Mb'-ba'-ab'-ba'-ab'-reverse of m] sequence block was cloned on the M13 minus strand orientation (see SI Section 1 for letter codes) in the PCR cloning site of pCR2.1 (Invitrogen) using standard molecular biology techniques. A 513 bp fragment of the plasmid was amplified by 30 cycles of unsymmetrical PCR using a 10:1 molar ratio of primers 5'-CTGTTGGGAAGGGCGATCG-3' and 5'-ACAGCTATGACCATGATTACGCC-3', respectively. The PCR product was purified from protein by phenol-chloroform extraction and ethanol precipitation. Following quantification, the PCR product was heated for 3 min at 100°C and frozen in liquid nitrogen. 1.5-stoichiometry of 3'-end grafted AA'-b5 and internally grafted BB'-b5 PDNAs (taking into account the presence of 2 binding sites for AA' and 3 for BB' on the matrix) were added to the frozen tube and let to warm to room temperature. Following incubation for 30 min at 42°C, resulting hybrids were purified by HPLC size exclusion chromatography. While asymmetric PCR mostly generates ss-matrix, the construct might contain small amount of re-hybridized double stranded matrix contaminating the expected PDNA-framework hybrid due to the impossibility to discriminate large DNA structures by gel filtration.

3.4. Données complémentaires

3.4.1 Analyse par microscopie de force atomique (AFM)

3.4.1.1 AFM à l'air

Equipement : le microscope utilisé est le Nanoscope Multimode 8 avec un scanner présentant une plage de balayage maximum latérale de 15 μm . Les pointes spécifiques des conditions à l'air sont des sondes en nitrure de silicium (référence Bruker : Scan Asyst ou DNP-S). Les leviers présentent une constante de raideur variant de 0.3 à 0.4N/m et une fréquence de résonance $f=20\text{-}30$ kHz.

Les conditions d'imagerie enregistrées par des sondes de température et d'humidité relative (Farnell ST-321) étaient variables, allant de $\sim 22^\circ\text{C} \pm 0,4$ avec 33,5% $\pm 0,4$ d'hygrométrie pour les jours les plus humides à $28^\circ\text{C} \pm 0,2$ avec une humidité relative inférieure à 5% pour les jours les plus secs.

Dépôt des échantillons : Les échantillons ont été déposés sur deux micas différents : le mica « japonais » et le mica « dijonnais » en utilisant un protocole classique : clivage du mica; traitement aux cations, rinçage à l'eau milliQ de l'excès d'ions; dépôt des échantillons incubés quelques minutes à quelques dizaines de minutes à 4°C ou température ambiante selon les échantillons et leur stabilité; rinçage de l'excès de matériel adsorbé et des sels; séchage avant observation. La difficulté pour obtenir des images exploitables avec le protocole de préparation précédent nous a amenés à développer des chimies de surface sur le mica à base d'époxy-silane ou d'amino-silane dont les principes sont présentés Figure 47.

Silanisation avec un époxysilane :

La chimie époxysilane (EPS) implique la modification de surface par la molécule 3-glycido-triméthoxysilane (GPTS), qui réagit avec les fonctions hydroxyles de la surface. Des liaisons Si-O-Si, qui se forment entre les chaînes adjacentes, stabilisent la couche (Figure 47-1). La séquence suivante a été mise en œuvre : Clivage du mica; incubation 20 min (20°C) sous azote avec 5% de saturation d'eau; incubation 14 heures (95°C) sous azote saturée par de la vapeur de toluène et de GPTS (1ml toluène + 0.1 ml de GPTS) ; balayage de 15 min (20°C) par de l'azote pur et 1 heure (20°C) par de l'air à 35% humidité relative.

Fonctionnalisations Mono-NTA et bis-NTA :

Le bis-NTA a été synthétisé à partir du $N\alpha, N\alpha$ -Bis(carboxyméthyl)-L-lysine hydrate (mono-NTA). Les fonctionnalisations au mono-NTA ont été réalisées en incubant les supports de mica préalablement fonctionnalisés avec l'EPS comme décrit précédemment. Une solution de

500 mM de mono-NTA amine dans un tampon aqueux carbonate de sodium pH 9, durant 1 h à 20°C. Les surfaces ont été lavées à l'eau milliQ, traitées avec une solution de 0.1 M de NiCl₂ pendant 3 min, puis à nouveau lavées à l'eau afin d'éliminer l'excès d'ions. Les fonctionnalisations au bis-NTA ont été réalisées en suivant les étapes du protocole précédent, mais avec une solution de 10 mM de bis-NTA amine.

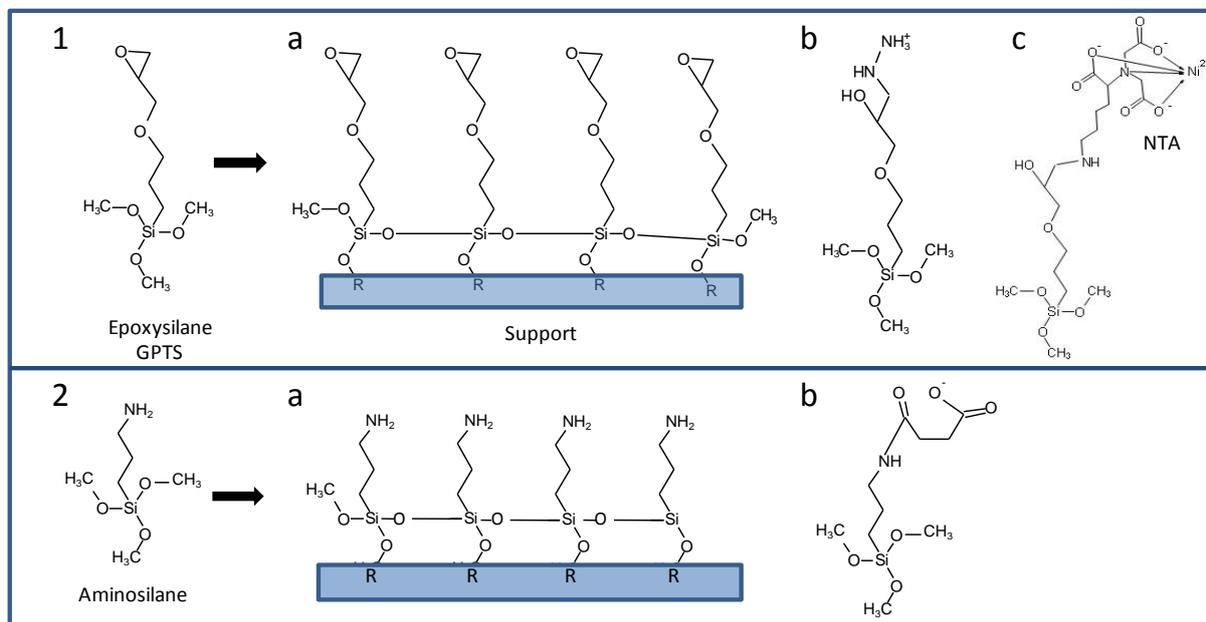


Figure 47 : Chimies de surface développées sur mica.

1, Fonctionnalisation à base d'époxy-silane : a, silanisation du support, b, époxy-silane hydrazine, c, acide nitrilotriacétique (NTA). 2, fonctionnalisation à base d'amino-silane : a, silanisation du support, b, chimie carboxylate.

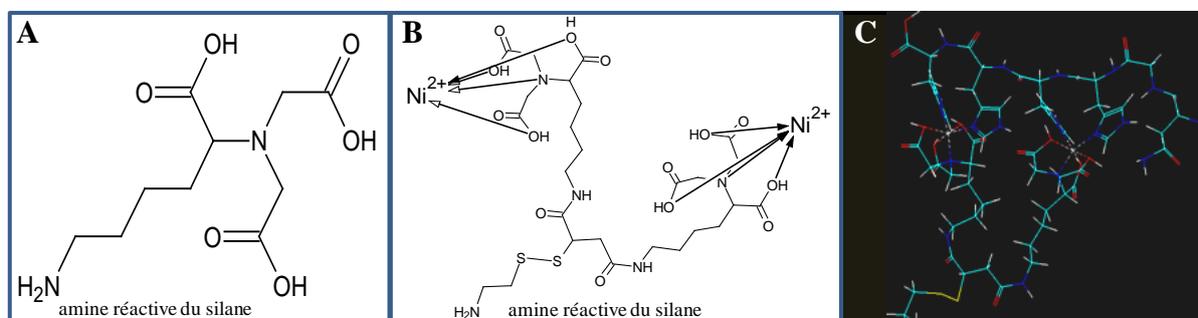


Figure 48 : Fonctionnalisations mono- et bis-NTA. A, mono-NTA. B et C, complexe bis-NTA-Ni-His

3.4.1.2 AFM en phase liquide

Les expériences en phase liquide ont été réalisées sur un microscope AFM à haute vitesse qui est un prototype dont la technologie repose sur celle de l'équipe japonaise du Pr. T. Ando (Laboratoire de Biophysique, Université de Kanazawa, Japon). Les leviers Olympus BL-AC10EGS-A2 ont été utilisés, avec des sur-pointes, de 2 à 3 μm, déposées par faisceau d'électrons par (SEM) microscopie électronique à balayage (JEOL 6500) au laboratoire à Dijon. Les forces d'interaction en mode contact intermittent étaient de l'ordre de 30 pN en

considérant un facteur de qualité Q égal à 2, une amplitude d'oscillation Δd de 0.2 nm et une constante de raideur K de 200 pN/m : $F \sim kx\Delta d/Q$.

Des leviers Nanoworld USNMCB présentant une longueur de $7\mu\text{m}$, une largeur de $3\mu\text{m}$, une épaisseur de 100nm, et une constante de raideur plus élevée de 0,6 N/m, ont aussi été utilisés, mais, dans ce cas, les forces d'interactions voisines de 600 pN/m ; ne sont plus négligeables.

Les échantillons ont été imagés soit dans 2 mM de MgCl_2 dilué dans de l'eau bi-distillée soit dans un tampon 10 mM Tris (pH 7.5), 10 mM MgCl_2 , 10 mM NaCl.

3.4.2 Echantillons de PDNA utilisés

3.4.2.1 Structures définies de PDNA

Des structures définies (dimère et trimères) ont été formées puis purifiées par gel filtration ou sur colonne échangeuse d'ions avant d'être hybridées ensemble (Figure 49). Les échantillons (pentamères-PDNA) et (pentamère-contrôle) proviennent respectivement de l'hybridation des dimères et des trimères.

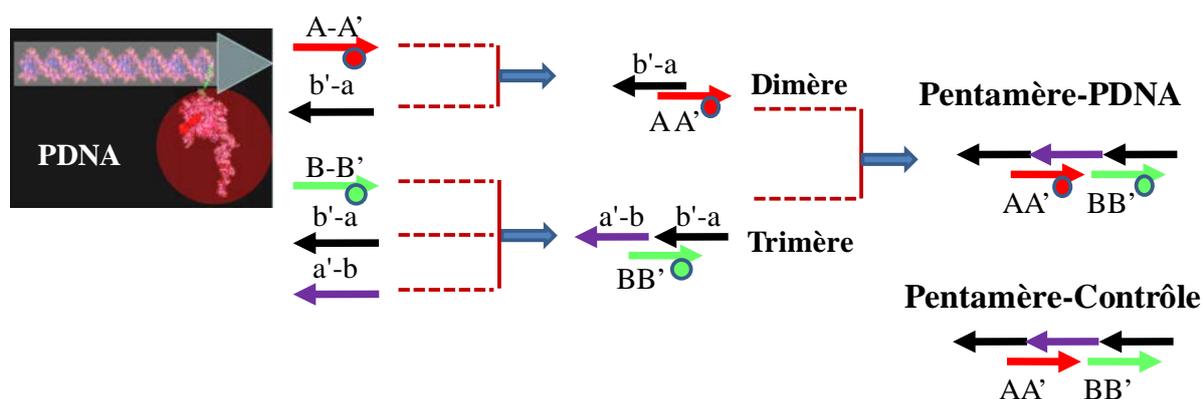


Figure 49 : Structures définies de PDNA.

Le PDNA est schématisé comme indiqué dans l'encart en haut à gauche, avec une flèche qui représente l'ADN simple brin et le cercle qui représente le domaine protéique. Les séquences d'ADN sont semi-complémentaires et permettent de former des dimères ou des trimères selon les blocs de construction utilisés. Les dimères et trimères sont purifiés avant d'être à nouveau hybridés entre eux pour donner la structure pentamérique (à droite). La pentamère contrôle (en bas à droite) constitué de cinq blocs d'ADN, ne présente aucune protéine.

3.4.2.2 Polymères de PDNA

Trois types de polymères ont été construits en utilisant 2, 4 ou 5 blocs de construction (Figure 50). Les polymères obtenus sont soit des polymères ordonnés de séquences connues, soit des polymères aléatoires.

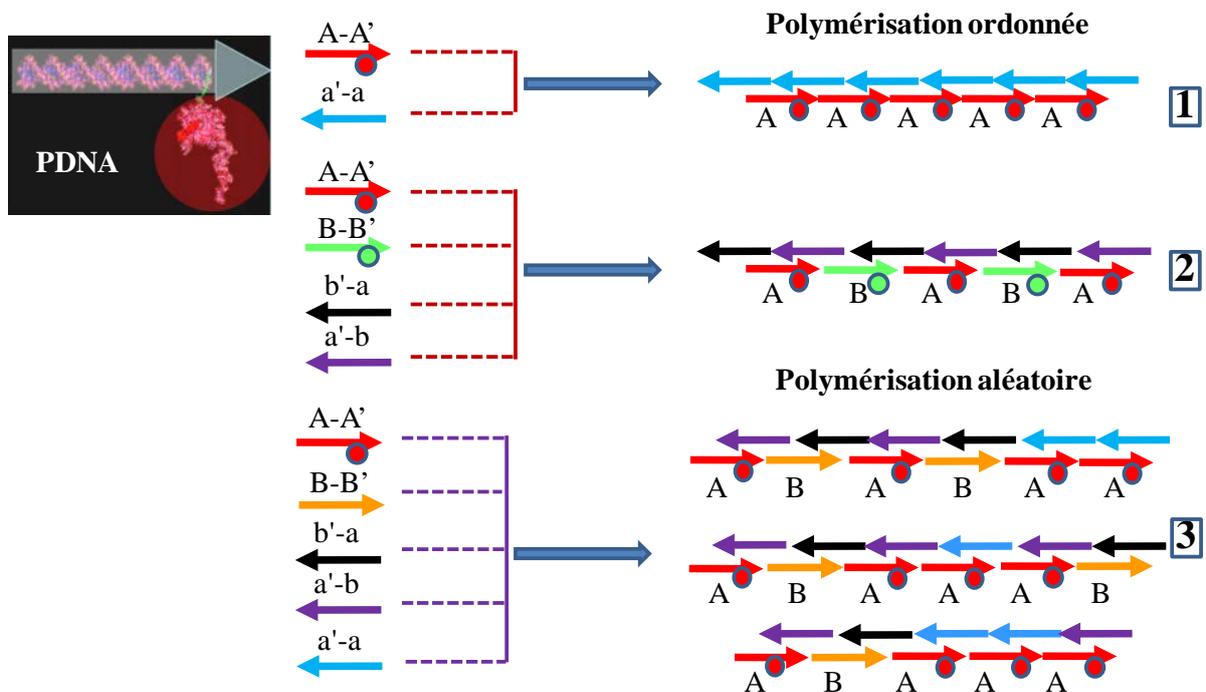
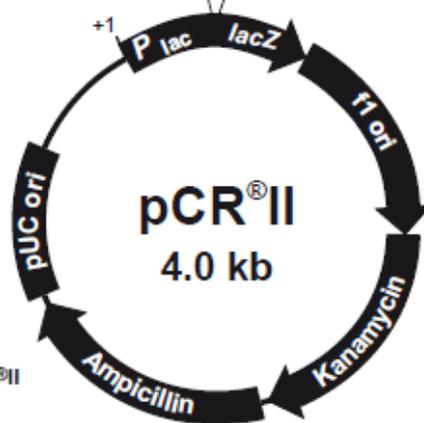
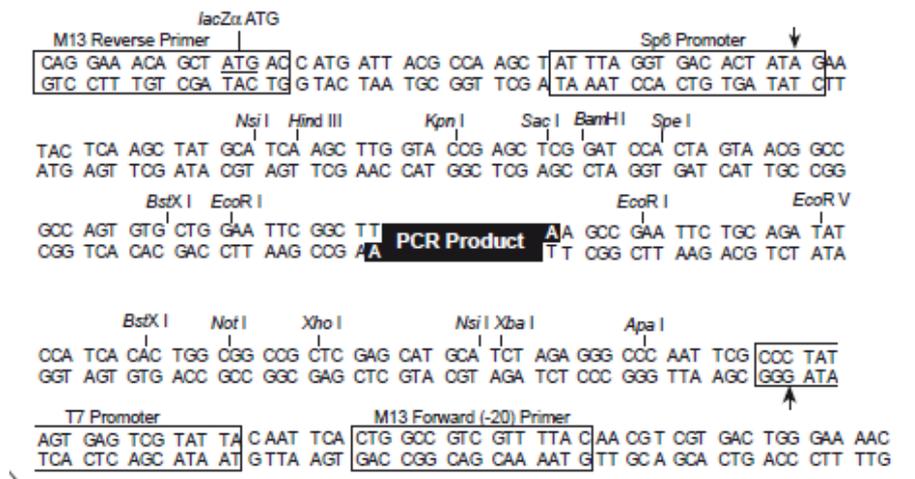


Figure 50 : Copolymères de PDNA ordonnés et aléatoires.

1, La polymérisation de deux briques semi-complémentaires (PDNA=AA'b5 et ADN=aa') génère une association ordonnée de séquence A. 2, La polymérisation de quatre briques semi-complémentaires (2 PDNA=AA'b5 et BB'b5, 2ADN=ab' et ba') génère des polymères ordonnés de séquence AB. 3, La polymérisation de cinq briques (2 PDNA et 3 ADN) donne naissance à des copolymères aléatoires divers.

3.4.2.3 Utilisation d'un brin matrice d'ADN simple brin obtenu par clonage.

L'objectif est ici d'obtenir des assemblages relativement longs de structures totalement définies. Pour cela, un brin matrice est créé par construction de biologie moléculaire et clonage ou par synthèse chimique. Le bloc de séquence [Mb'-ba'-ab'ba'-ab'-complémentaire de m] a été cloné sous forme double brin dans le vecteur pCR II (Invitrogen) dont la carte est présentée sur la Figure 51. Ce vecteur permet la génération de simple brin grâce à son origine de réplcation f1 (spécifique du phage M13 monobrin).



Comments for pCR[®]II
3971 nucleotides

LacZα gene: bases 1-587
M13 Reverse priming site: bases 205-221
Sp6 promoter: bases 239-256
T7 promoter: bases 404-423
M13 (-20) Forward priming site: bases 431-446
f1 origin: bases 588-1025
Kanamycin resistance ORF: bases 1359-2153
Ampicillin resistance ORF: bases 2171-3031
pUC origin: bases 3176-3849

Figure 51 : Carte et caractéristiques du vecteur pCR II (Invitrogen).
http://www.genomex.com/vector_maps/pcritopo_map.pdf

Un fragment de 513 paires de bases a été amplifié par PCR asymétrique de 30 cycles en utilisant les amorces 5'-CTGTTGGGAAGGGCGATCG-3' et 5'-ACAGCTATGACCATGATTACGCC-3' en proportion molaire de 10 :1. Le produit de PCR a été purifié par extraction au phénol-chloroforme puis précipitation à l'éthanol. Après quantification, les produits de PCR ont été dénaturés 3 min. à 100°C et directement congelés dans de l'azote liquide. Les PDNA AA'b5 et BB'b5 ont été ajoutés au tube congelé qui a été ramené à température ambiante sous agitation puis incubé pendant 30 min. à 42°C. Les produits hybrides ont été purifiés par gel filtration sur HPLC, permettant d'éliminer les PDNA non hybridés. Même si la PCR asymétrique génère majoritairement une matrice simple brin, il est possible que certains fragments d'ADN se soient hybridés entre eux contaminant la préparation d'assemblages ADN-PDNA. Les échantillons attendus sont schématisés sur la Figure 52.

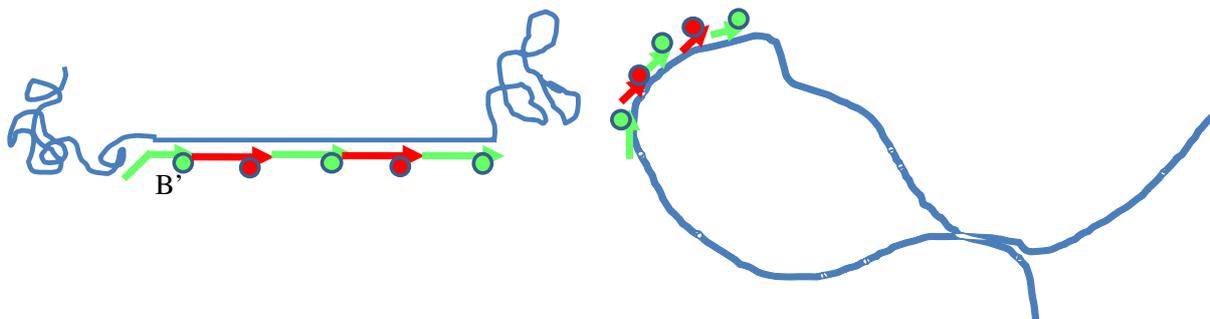


Figure 52 : Hybrides PDNA-fragment d'ADN de 500 pb.
Un long fragment d'ADN d'environ 500 bases, présente une séquence nucléotidique sur laquelle maximum 5 PDNA peuvent s'hybrider (Echantillon 6).

3.4.3 Résultats des caractérisations par AFM

3.4.3.1 AFM à l'air

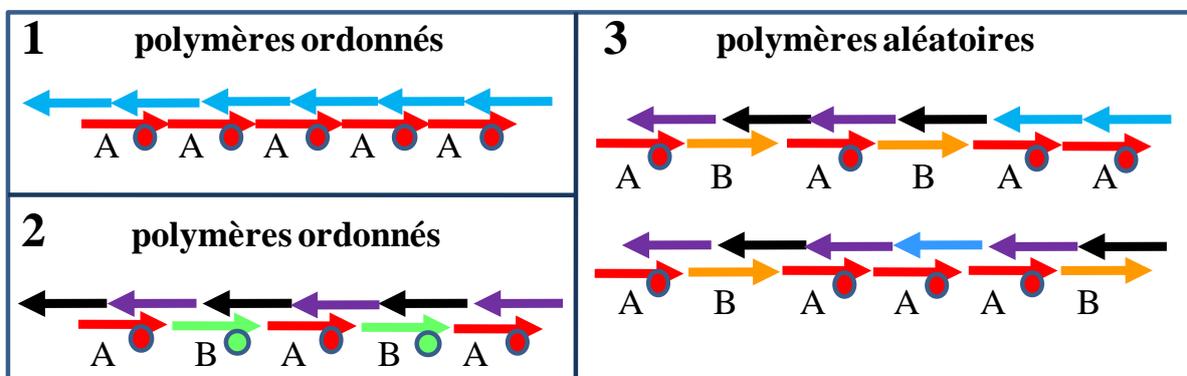


Figure 53 : Schémas des copolymères de PDNA ordonnés et aléatoires.

1, copolymères de PDNA ordonnés faisant intervenir deux blocs de construction (AA'b5 et aa'). 2, copolymères de PDNA ordonnés réalisés à partir de quatre blocs de construction (AA'b5, BB'b5, ab' et ba'). 3, copolymères aléatoires réalisés à partir de cinq blocs de construction (AA'b5, BB', ab', ba' et aa').

Les copolymères de PDNA de la Figure 53 ont été polymérisés (0.5 μ M dans 10 mM PBS) à 22°C avant d'être déposés sur différents supports de mica fonctionnalisés.

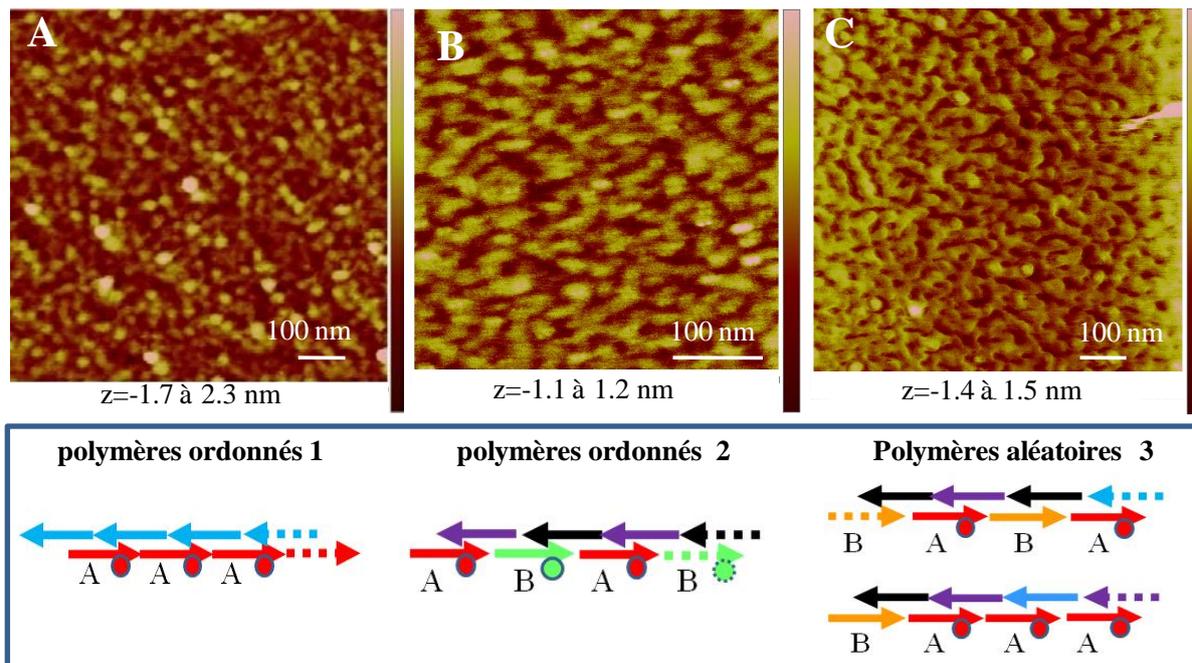


Figure 54 : Images AFM des copolymères de PDNA incubés à 22°C, déposés sur mica époxy-silane bis-NTA.
 A, copolymères de PDNA ordonnés 1, réalisés à partir de deux blocs de construction.
 B, copolymères de PDNA ordonnés 2, réalisés à partir de quatre blocs de construction.
 C, copolymères de PDNA aléatoires 3, réalisés à partir de cinq blocs de construction.

Les images AFM de la Figure 54, montrent une densité élevée d'objets sur les supports de mica qui rend difficile la distinction des copolymères de PDNA. On remarque des alignements de sphères sur l'image A qui pourraient correspondre aux protéines. Aucun filament d'ADN n'est visible. L'orientation des objets pourrait être due à un peignage des échantillons lors de l'étape de séchage par flux d'air avant observation. Les images B et C sont difficilement interprétables du fait de la saturation de la surface. Une gangue de sels recouvre les objets (image C). La disparité des qualités d'images peut provenir de différents facteurs tels que la chimie de surface multi-étapes non contrôlée, le protocole de dépôt des échantillons ou encore la nature des échantillons déposés.

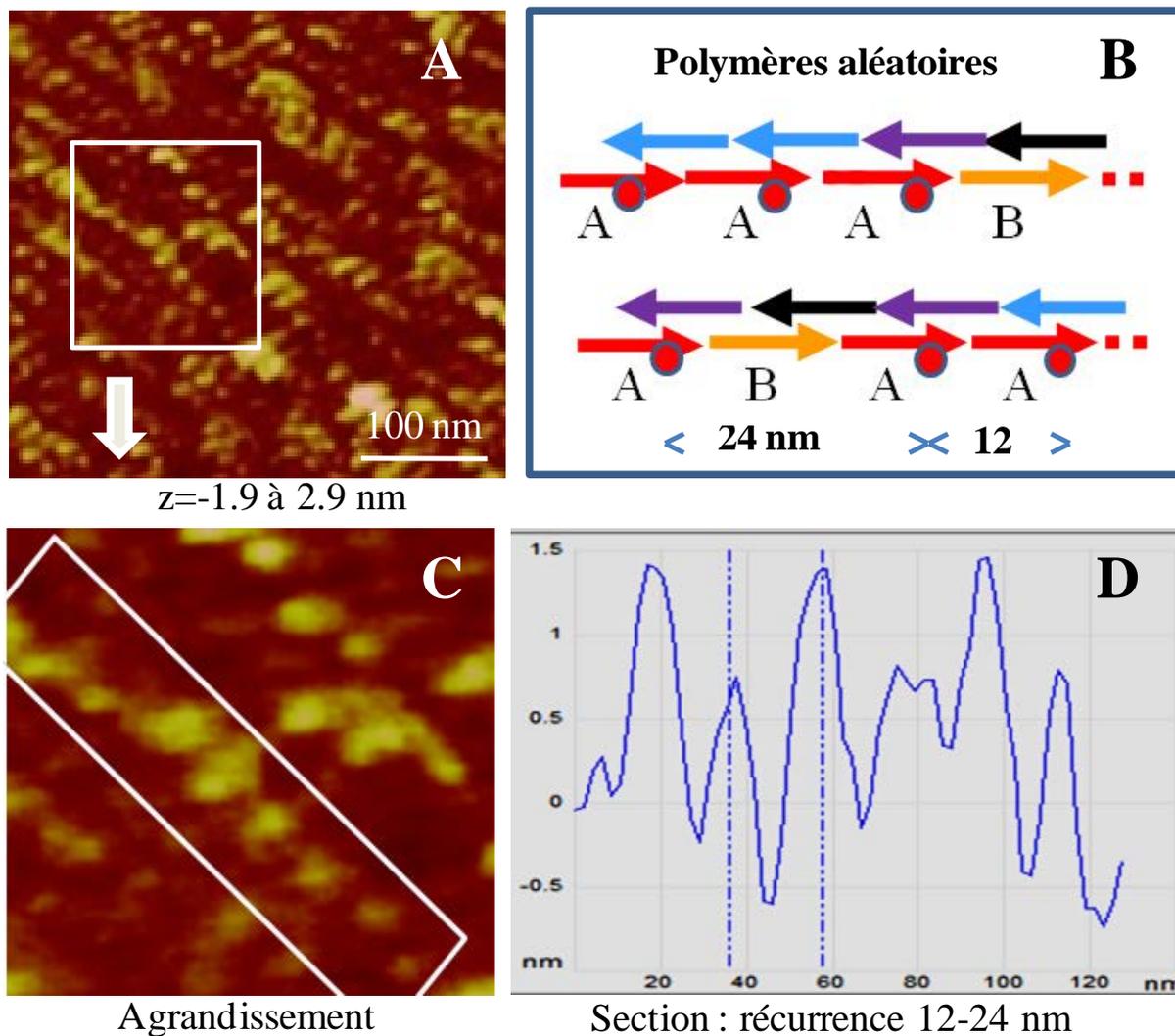


Figure 55 : Images AFM des copolymères de PDNA aléatoires incubés à 22°C, déposés sur mica époxy-silane mono-NTA.

A, copolymères de PDNA aléatoires incubés à 22°C. B, schéma des copolymères de PDNA aléatoires. C, Agrandissement correspondant au cadre blanc de l'image A. D, section correspondant au cadre blanc de l'image C.

Les images de la Figure 55 qui correspondent aux copolymères de PDNA aléatoires incubés à 22°C déposés sur mica époxy-silane mono-NTA, illustrent également des alignements, observés pour les copolymères de PDNA ordonnés (Figure 54A). Un agrandissement (image C) et une section (image D) montrent un pas de 12 à 24 nm entre deux « sphères ». Ceci est compatible avec les distances attendues égales à 12, ou à un multiple de 12 nm, dans la conception d'un copolymère aléatoire, dont le schéma est présenté en Figure 55B.

Une autre expérience a consisté à incuber les copolymères polymérisés à 22°C à des températures de 45 puis 55°C pendant 10 minutes pour chaque température.

phénomènes rappellent les observations similaires faites sur les polymères à base d'ADN seuls (partie 1).

La caractérisation à l'air des copolymères de PDNA n'a pas permis d'observer simultanément les ADN et les protéines. Ces résultats sont inattendus puisque le couplage covalent de la partie ADN au domaine protéique garantit la présence des deux composants dans les PDNA.

3.4.3.2 AFM à haute vitesse (HS-AFM)

Structure de pentamère-PDNA associant trois blocs d'ADN et deux blocs de PDNA

Nous commençons par les échantillons les plus simples, les pentamères de PDNA, qui impliquent deux blocs de PDNA et trois blocs d'ADN, tandis que le pentamère contrôle implique cinq blocs d'ADN comme présenté sur la Figure 57. Ces objets présentent une longueur maximale de 36 nm avec 78 paires de bases en double hélice encadrées par 20 bases d'ADN simple brin, à chaque extrémité. Les contrôles correspondants sans protéine, ont également été fabriqués et imagés par HS-AFM afin d'être en mesure de comparer les objets sans ou avec protéines. Les domaines protéiques de 13 kDa, correspondent à un domaine globulaire de 2 nm de diamètre incluant la partie hémique prolongée en C-terminal par une extension peu ou pas structurée. Dans la forme la plus étirée de cette dernière partie, la dimension de la protéine peut atteindre environ 4 nm.

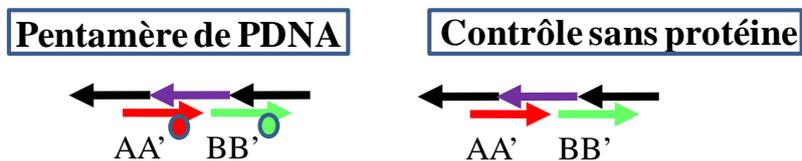


Figure 57 : Schémas du pentamère de PDNA et du contrôle ADN.

Le pentamère implique deux blocs de PDNA et trois blocs d'ADN, le contrôle n'implique que des ADN.

Images HS-AFM du pentamère de PDNA

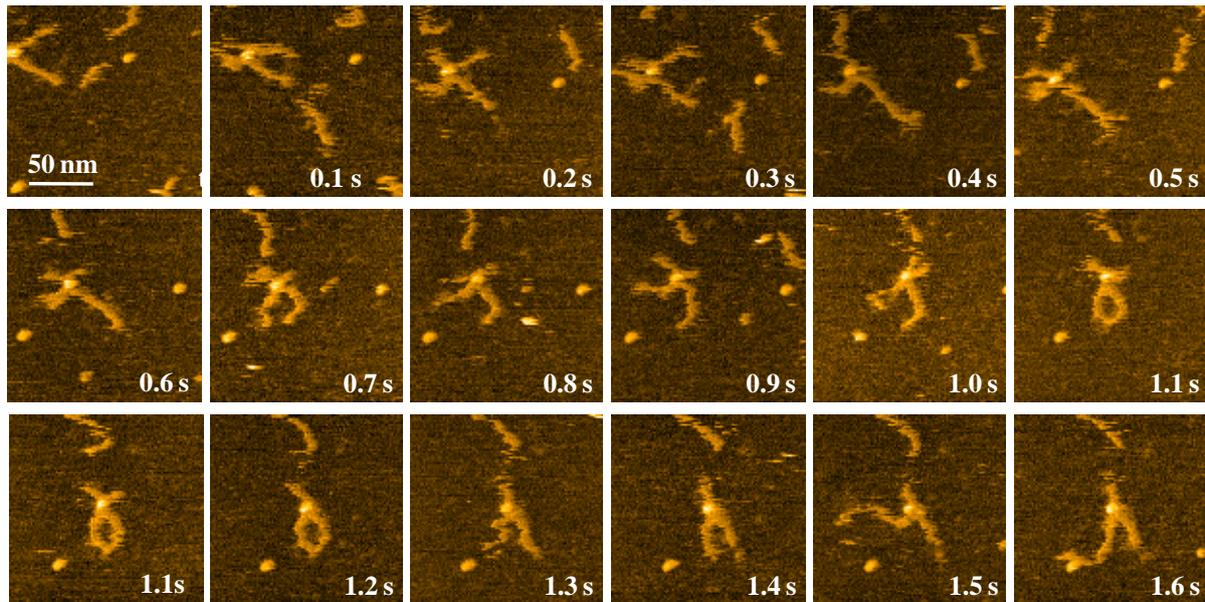
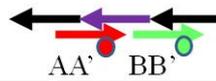


Figure 58 : Images HS-AFM du pentamère de PDNA.

La vitesse de balayage est de 10 images par seconde. Les objets sont très mobiles sur la surface. On observe l'ancrage des molécules sur des nanoparticules. Les molécules forment une structure dynamique telle une pince, qui s'ouvre et se ferme alternativement.

Les images (Figure 58) du pentamère de PDNA montrent la grande mobilité des objets. Des molécules de différentes tailles révèlent des interactions entre les extrémités simple brin des éléments de 36 nm. On observe des nanoparticules (boules) qui pourraient correspondre à des amas d'impuretés capables de piéger les objets. Les molécules ancrées sont capables de mouvoir leurs extrémités libres qui peuvent interagir entre elles pour former une structure en forme de pince, ou encore interagir avec des molécules avoisinantes non piégées par une nanoparticule. Nous n'observons pas sur ces images la présence des protéines alors que la liaison covalente, de la molécule d'ADN au domaine protéique par un adaptateur synthétique, garantit la présence des deux composants du PDNA.

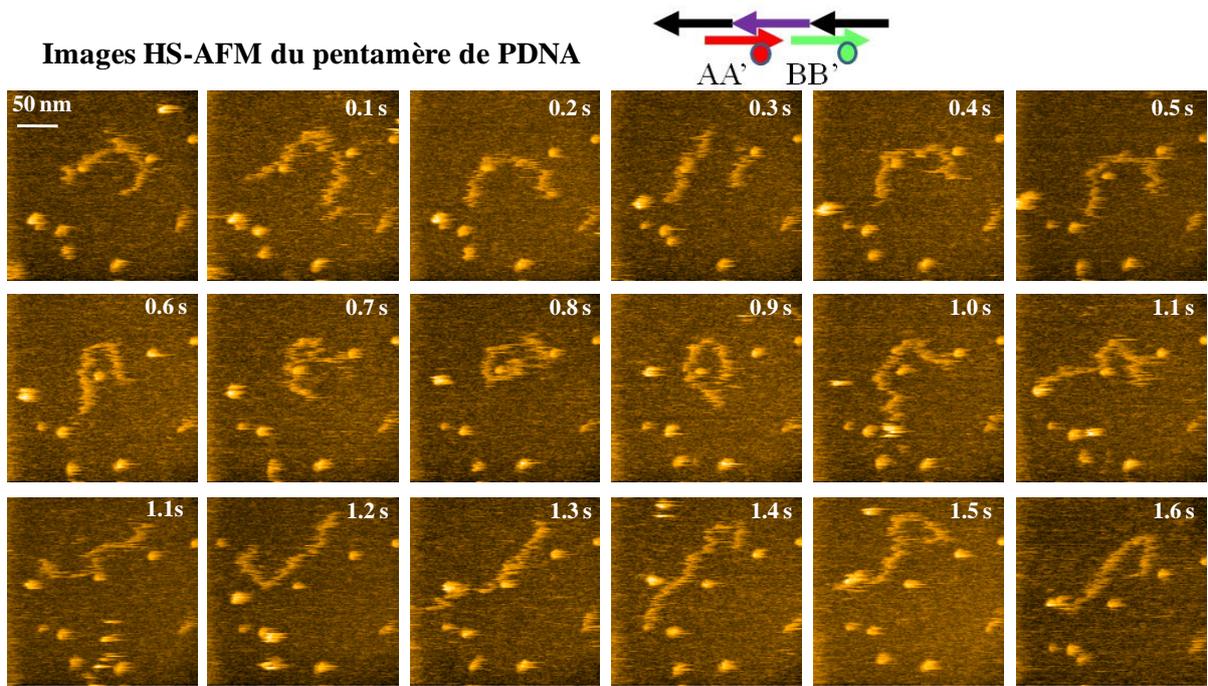


Figure 59 : Images HS-AFM du pentamère de PDNA.
 La dynamique des interactions entre molécules élémentaires est illustrée sur cette séquence d'images. Les objets s'associent et se dissocient de façon transitoire.

Les images de la Figure 59 illustrent les associations et dissociations transitoires des molécules sur la surface, donnant des tailles d'objets différentes selon l'image considérée. A nouveau, de nombreuses nanoparticules sont observées, mais cette fois, les molécules ne sont apparemment pas bloquées par les particules, semblant même tourner autour. Comme précédemment, nous ne discriminons pas la présence attendue du domaine protéique tous les 12 nm sur les brins d'ADN.

Contrôle ADN du pentamère

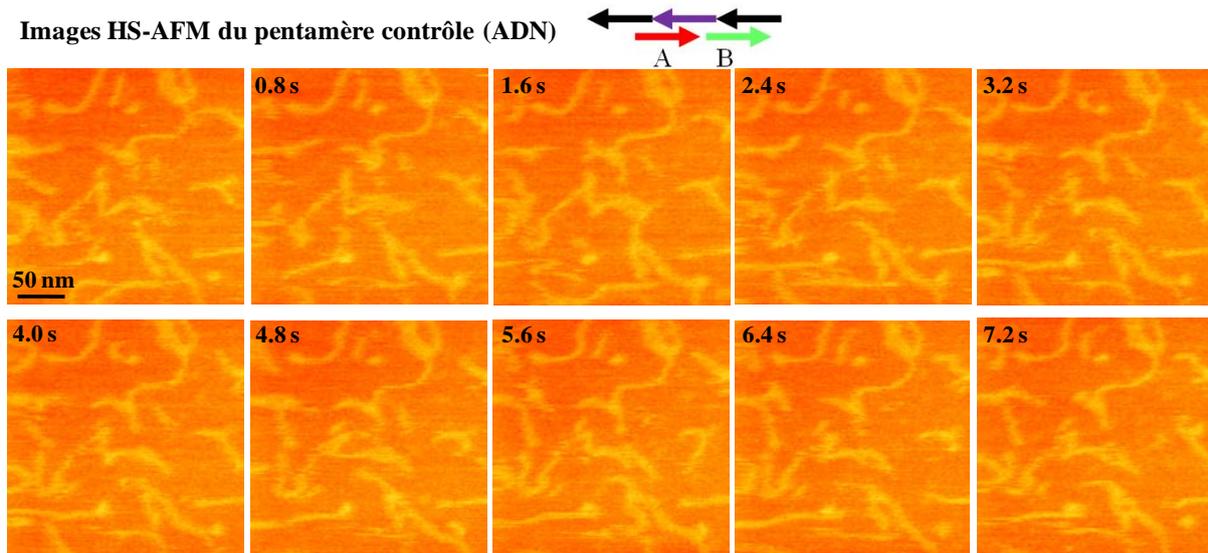


Figure 60 : Images HS-AFM du contrôle du pentamère ADN.
 Séquence d'images ADN avec 800 ms entre deux images successives.

La séquence d'images de la Figure 60 concerne le contrôle ADN de l'échantillon 1. La densité des objets est un peu élevée, si bien qu'il est difficile de discriminer les objets élémentaires des objets en interactions par les extrémités simples brins ou par encombrement de la surface. Malgré cela, un comportement semblable peut être décrit, avec certaines molécules qui s'associent et se dissocient transitoirement. Les nanoparticules sont également observées dans cet échantillon sans protéine, ce qui confirmerait que les nanoparticules sont issues du traitement de la surface, notamment des impuretés provenant des solutions de chlorure de nickel qui sont filtrées avec un seuil de 0.22 μm .

Il ressort de l'étude du pentamère de PDNA et de son contrôle par HS-AFM, les points suivants :

- On observe la dynamique des objets,
- Les molécules interagissent entre elles par leurs extrémités simple brin, générant une dispersion de taille des objets à la place de la longueur définie, attendue à 36 nm,
- On ne voit que la partie dynamique des ADN mais pas celle des protéines,
- Les interactions semblent similaires entre les échantillons PDNA et DNA.

L'absence de visualisation des protéines par HS-AFM peut être due, soit à une mobilité de la protéine relativement élevée par rapport à la vitesse du balayage, soit à l'absence d'interaction entre la protéine et la surface ne permettant pas à la pointe de la sonde.

Etant donnée la difficulté rencontrée pour observer les protéines, nous avons construit des échantillons présentant d'autres organisations afin de les comparer et de les discriminer par imagerie HS-AFM. La polymérisation de quatre briques de construction (2 PDNA et 2 ADN) aboutit à la génération de copolymères de PDNA ordonnés dont la distribution des longueurs est aléatoire. Néanmoins, ces structures présentent toutes, sous une forme étendue, une périodicité de 12 nm entre protéines (Figure 61).

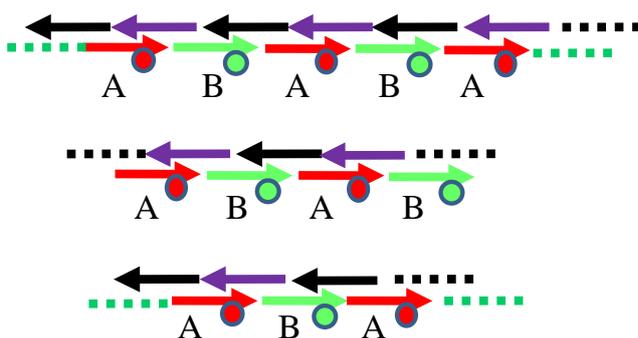


Figure 61 : Schémas des copolymères de PDNA ordonnés à partir de quatre blocs.

La longueur des copolymères varie avec le degré de polymérisation des quatre blocs de construction (2 PDNA : AA'b5 et BB'b5 et 2 blocs d'ADN : ab' et ba').

Copolymère de PDNA ordonné à partir de quatre blocs de construction.

Images HS-AFM des copolymères de PDNA

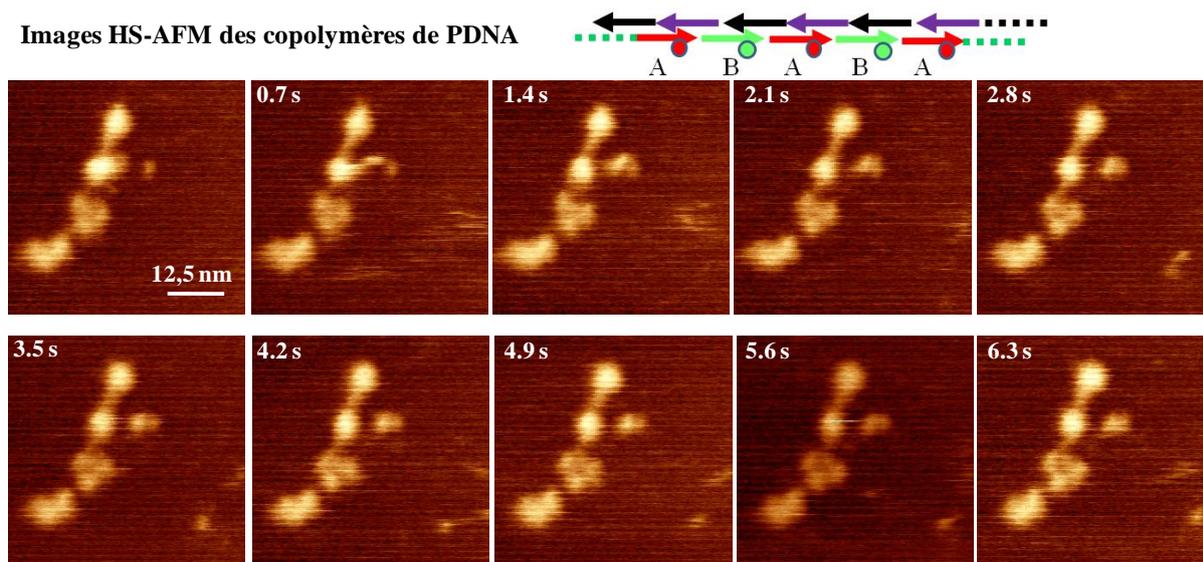


Figure 62 : Images HS-AFM des copolymères de PDNA ordonnés, impliquant 4 blocs de construction.

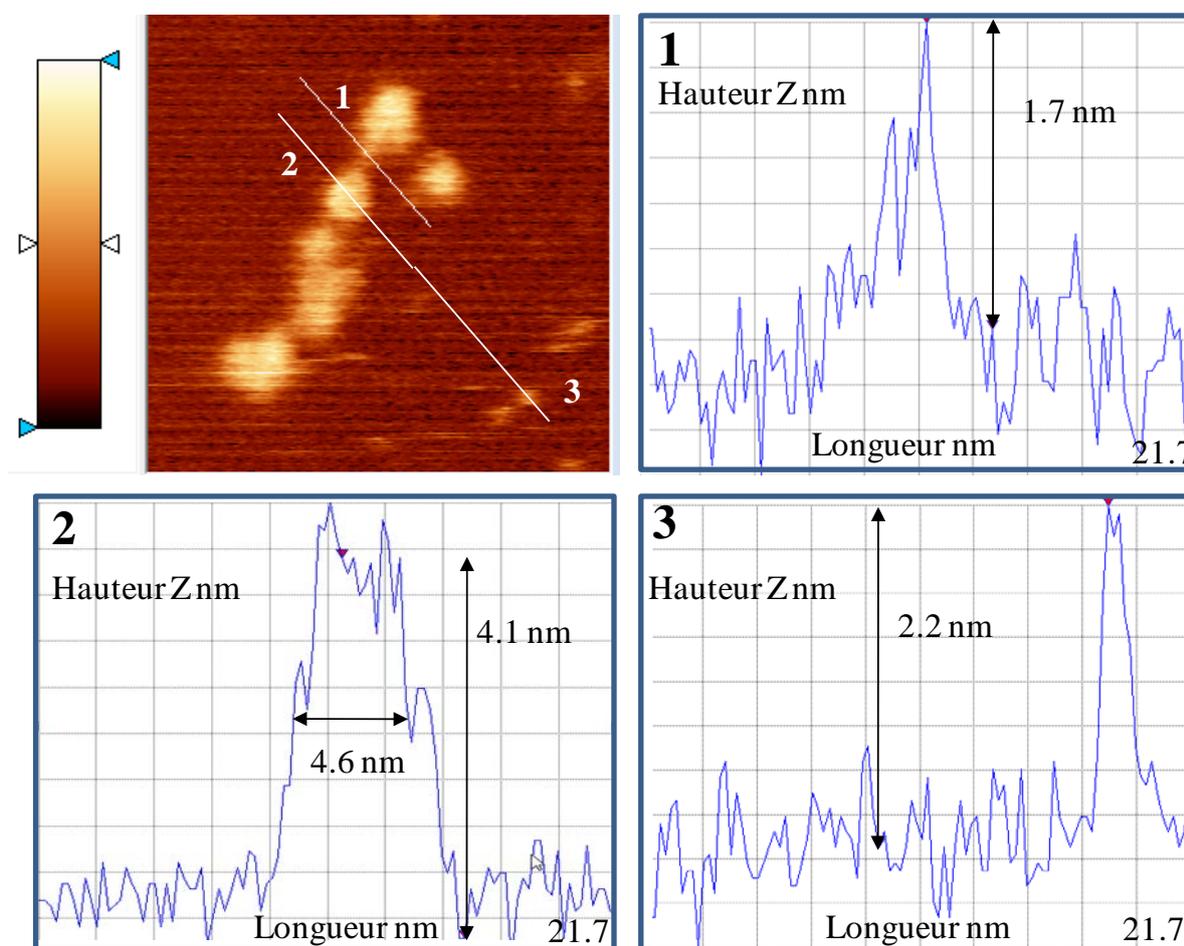


Figure 63 : Sections d'image de copolymères de PDNA.

1, section de hauteur dans le filament d'ADN entre deux sphères en position 1 sur l'image HS-AFM. 2, section dans une sphère en position 2 sur l'image HS-AFM. 3, Section d'un objet en position 3, dont la taille avoisinant 2 nm correspondrait à la hauteur attendue du domaine protéique.

Images HS-AFM des copolymères de PDNA

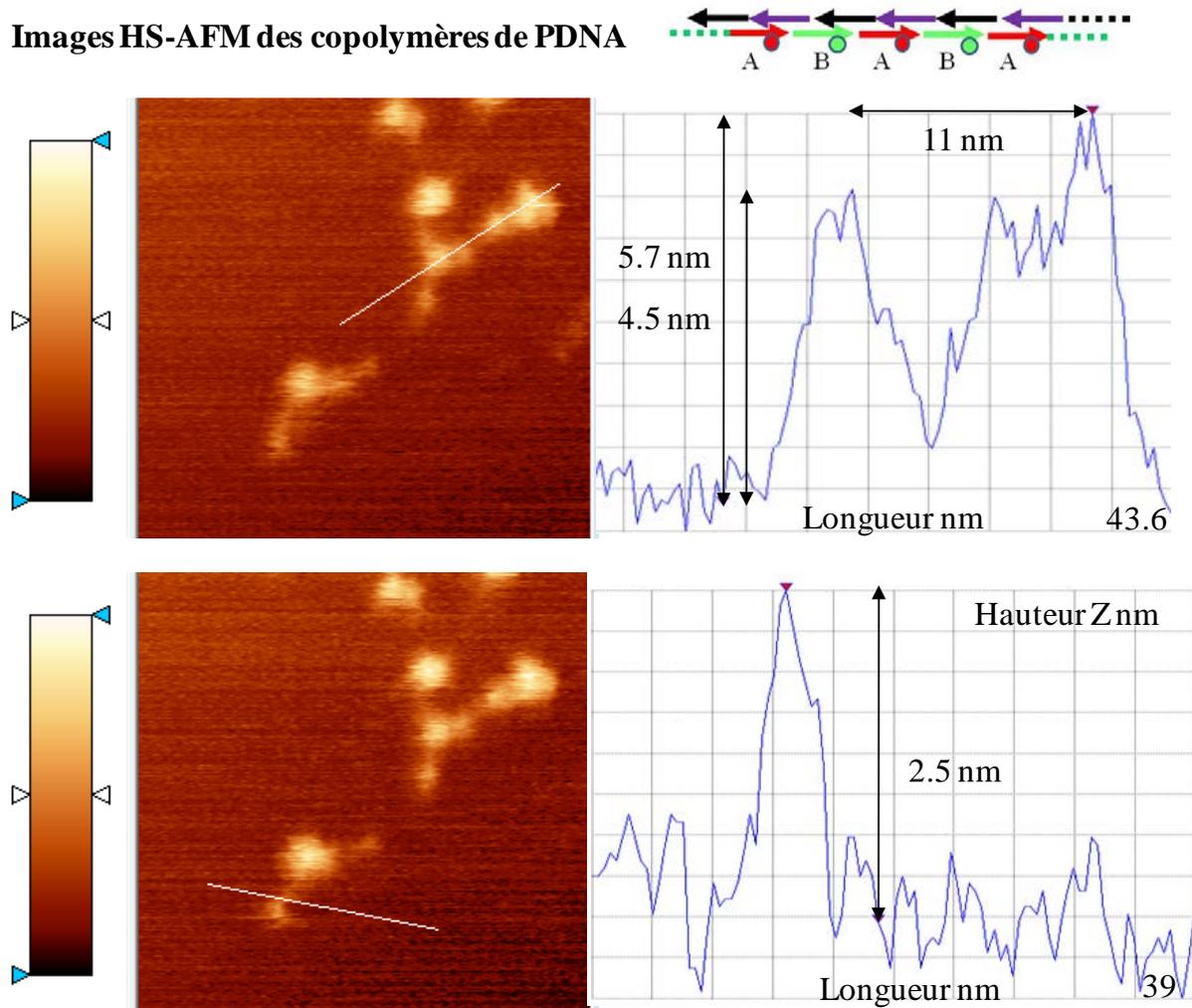


Figure 64 : Images HS-AFM des copolymères de PDNA.

Les lignes supérieures et inférieures présentent une image des copolymères dans laquelle une section indiquée par le trait blanc a été réalisée. Les profils de distance des sections sont présentés sur la partie droite de la figure.

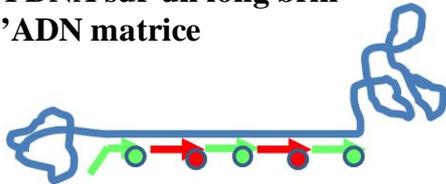
Les images révèlent des structures très différentes du pentamère de PDNA avec plusieurs « sphères » qui semblent reliées par un filament. Celui-ci, modérément visible selon les images, présente une hauteur de 1,7 nm qui est compatible avec la hauteur attendue pour des ADN double brin (Figure 63).

On peut noter que la hauteur des sphères est supérieure à celle attendue pour la partie globulaire de la protéine (~4 nm et plus au lieu de 2) sur la Figure 63 et sur la Figure 64. Néanmoins, l'extension C-terminale non structurée peut dans une forme étendue atteindre des dimensions compatibles. Des objets très dynamiques qui ont l'air de se détacher de l'ensemble et de s'y ancrer transitoirement montrent des hauteurs (2 nm) compatibles avec la taille attendue du domaine protéique, toutefois, on n'observe pas clairement de brin d'ADN (Figure 62). Il est délicat de conclure car l'ensemble des travaux montre que la visualisation simultanée n'est pas toujours observée.

Pour lever l'incertitude provenant de la taille aléatoire des polymères, nous avons conçu d'autres échantillons constitués d'un brin d'ADN de taille connue, qui autorise l'hybridation de 5 PDNA maximum. Le double avantage de cette construction est sa taille définie (~500 bp soit ~130 nm) et le nombre fini de PDNA pouvant se fixer.

Ces échantillons ainsi que leur contrôle purement ADN sont présentés sur la Figure 65.

5 PDNA sur un long brin d'ADN matrice



Contrôle ADN

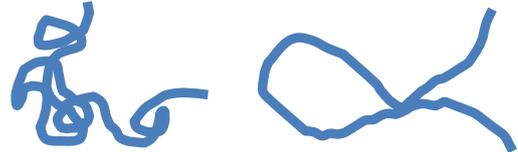


Figure 65 : Schémas de l'hybride 5 PDNA sur le brin ADN matrice et son contrôle.

Le monobrin d'ADN obtenu par clonage autorise la fixation de maximum cinq blocs de PDNA.

Images HS-AFM de PDNA sur un brin matriciel d'ADN

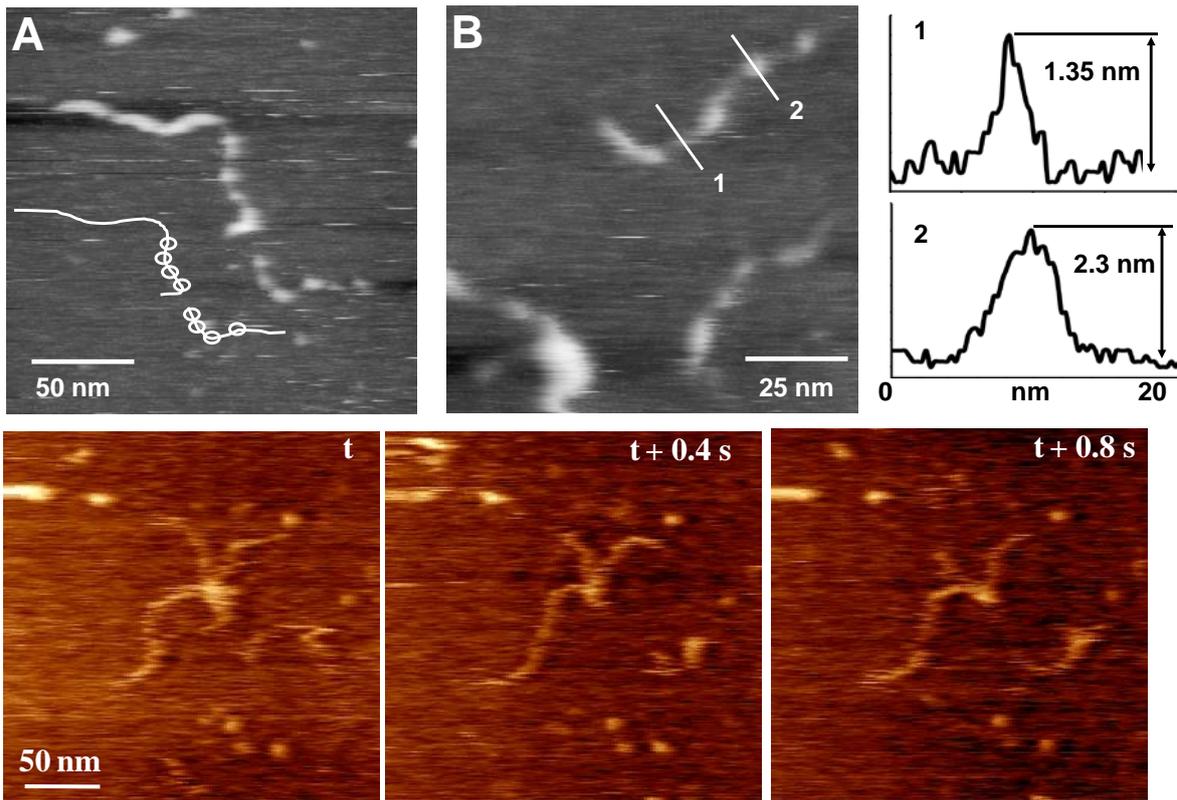
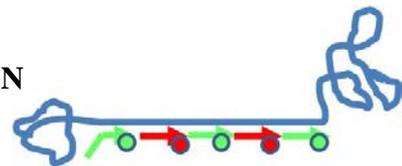


Figure 66 : Images HS-AFM de PDNA hybridés sur un brin d'ADN matrice.

A, image montrant le brin d'ADN avec les domaines protéiques. B, image montrant d'autres objets hybrides PDNA-brin d'ADN matrice. Des sections numérotées 1 et 2 sont indiquées par les traits sur l'image. Le profil de section est présenté à droite de l'image B. Ligne inférieure, trois images consécutives du même échantillon.

Les images A et B de la Figure 66, permettent d’observer, de manière claire, des domaines protéiques régulièrement espacés sur le filament d’ADN. Les sections réalisées montrent des hauteurs d’environ 1,4 et 2,3 nm, qui sont compatibles avec les hauteurs attendues pour l’ADN et le domaine globulaire du cytochrome b5. Les protéines sont espacées d’environ 10 à 12 nm sur le fragment d’ADN de grande taille immobile sur cette série. La ligne inférieure des images de la Figure 66, obtenues à partir du même échantillon mais sur un autre dépôt, ne permettent plus de distinguer les protéines, démontrant une non reproductibilité des conditions d’imagerie. En particulier, les films montrent une beaucoup plus grande mobilité des objets dans la seconde série.

Contrôle purement ADN de la construction précédente

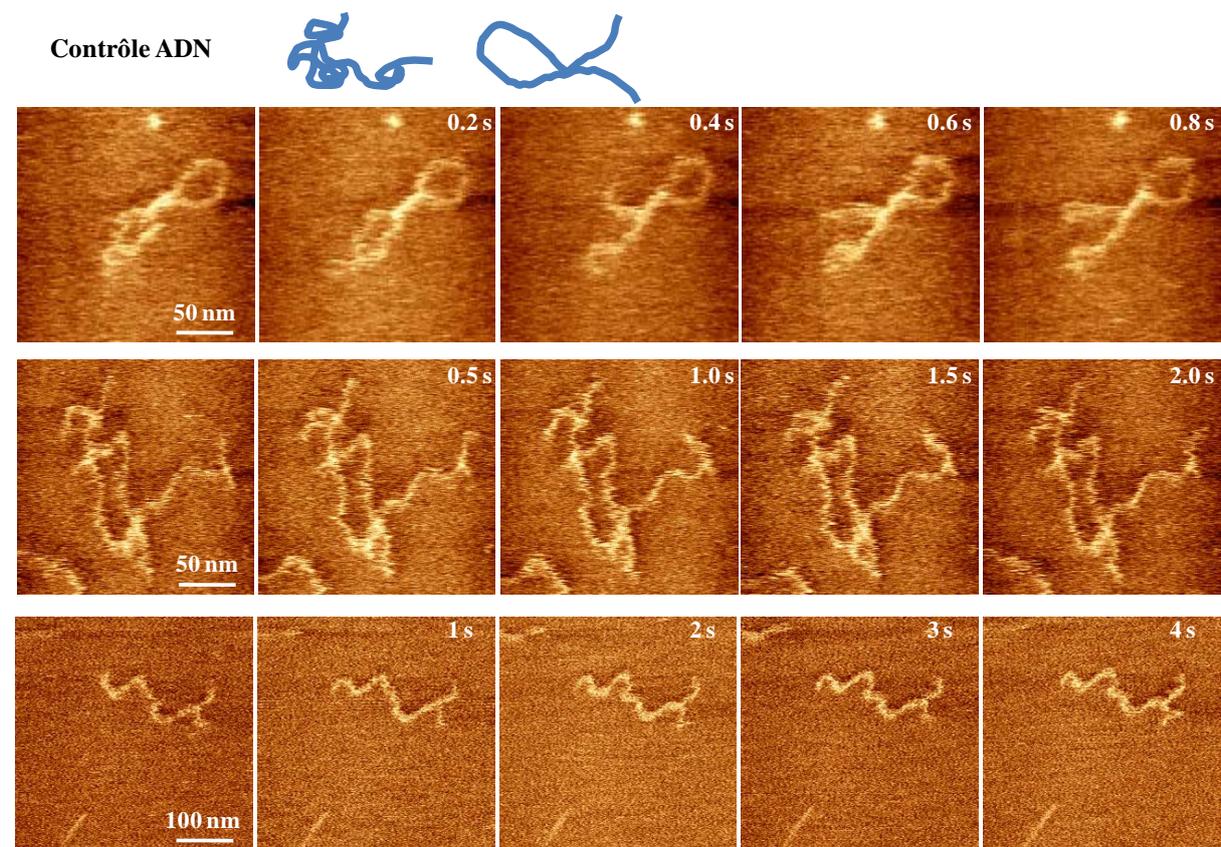


Figure 67 : Images HS-AFM du contrôle du brin d’ADN matrice sans protéine.

Les images (Figure 67) de l’échantillon contrôle montrent les filaments d’ADN réguliers dans différentes configurations, soit repliée du fait d’une partie de la séquence complémentaire, soit allongée en forme de zigzag. Conformément à ce qu’on attend, on n’observe pas la présence de surbrillance régulière. Dans cette série d’images, les objets sont relativement immobiles rappelant la situation rencontrée dans le panel supérieure de la Figure 65 où les domaines protéiques étaient visibles.

Conclusion sur les observations des échantillons par HS-AFM

Reproductibilité des conditions d'imagerie

Malgré l'association covalente entre les domaines protéiques et nucléiques au sein des PDNA, la plupart des images d'AFM dynamique ne conduit qu'à la visualisation de la partie ADN. Ceci est le cas pour le pentamère de PDNA, où l'on observe une grande mobilité de la partie nucléique sans distinguer les protéines, mais ceci est vrai aussi pour certaines séquences d'images des hybrides PDNA-brin d'ADN matrice, où selon les conditions d'imagerie, on voit, ou non, des objets qui pourraient correspondre aux protéines. A l'inverse, d'autres images (Figure 62 à 66) révèlent la présence simultanée des protéines et de l'ADN.

Ces différences peuvent être expliquées soit par les conditions d'imagerie ou une variabilité de la surface du mica (clivages, dopage ionique). En fait, les domaines protéiques et nucléiques peuvent interagir très différemment avec le support et la pointe : l'ADN est chargé négativement lorsqu'il est nu mais les ions divalents neutralisent partiellement cette charge selon leur densité à la surface. La protéine est beaucoup plus faiblement chargée et n'interagit probablement significativement avec les ions nickel qu'au niveau de son étiquette C-terminale. Quand le complexe PDNA s'attache à la surface et si la densité d'ions divalents est suffisante, il est probable que l'interaction implique majoritairement la partie ADN. Dans ces conditions, les domaines protéiques peuvent rester libres en solution « flottant » au dessus de l'ADN du fait de la flexibilité du lien protéine-ADN. Ces domaines peuvent alors interagir seulement faiblement avec la pointe et conserver une grande dynamique même si la partie ADN est relativement immobilisée. Dans ces conditions, ils peuvent être « invisibles ». Il est possible d'imaginer à l'inverse, que dans d'autres conditions, ce soit les étiquettes des protéines qui interagissent principalement avec le substrat conduisant à les immobiliser et à les rendre visibles alors que l'ADN est rejeté en solution, posé de manière lâche sur les piliers de protéines, et devient invisible. La dynamique des objets sur la surface au sein d'une série d'images est illustrée sur la Figure 68.

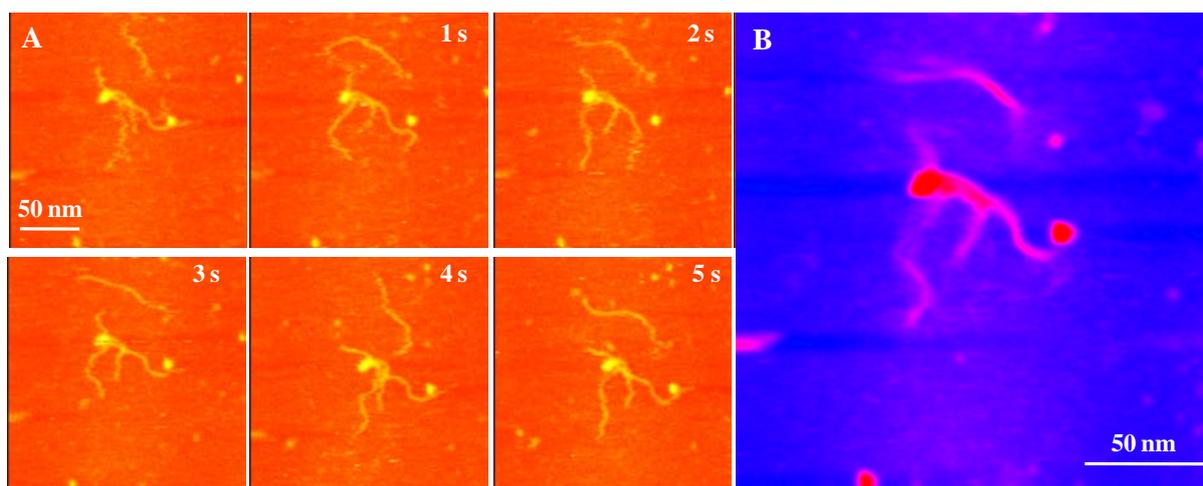


Figure 68 : Images HS-AFM illustrant l'hétérogénéité des mouvements des objets. A, images successives à 1 seconde d'intervalle montrant la dynamique des objets. B, codage couleur de l'intensité moyenne (calculée sur 154 images consécutives) révélant les parties statiques et mobiles des objets sur la surface.

Les mouvements inhomogènes sont probablement dus à deux phénomènes. L'un, visible sur les images successives, est lié à la présence de nanoparticule sur la surface qui peut ancrer les polymères sur des points fixes. L'autre, peu visible sur les images dynamiques, peut être facilement perçu lorsque l'on moyenne de nombreuses images consécutives. Le codage couleur de l'intensité moyenne révèle des segments immobiles qui apparaissent en rose intense quand les segments plus mobiles apparaissent comme des aires diffuses en rose clair. L'explication la plus simple repose sur l'hétérogénéité des propriétés de la surface de mica, soit en concentration ionique, soit au niveau de la teneur ionique et des défauts du feuillet cristallin du mica.

Afin de s'affranchir de ces phénomènes, il faudrait envisager une expérience dans laquelle, on observe en même temps les échantillons assemblés d'ADN seuls et de PDNA, chacun marqué différemment. Par exemple, il faudrait marquer les échantillons ADN seuls à la biotine, permettant une décoration à l'avidine, qui serait absente des PDNA. On pourrait également imaginer les échantillons avant puis après ajout d'anti-b5, qui décorerait la protéine. La conception d'une expérience où l'on est capable de distinguer les deux objets imagés sur le même support constitue une solution pour résoudre les problèmes de manque de reproductibilité rencontrés.

Propriétés dynamiques hétérogènes entre domaine protéique et nucléique des PDNA

Deux conditions semblent être nécessaires, mais pas forcément suffisantes, pour observer simultanément les domaines protéiques et nucléiques. L'ADN doit être suffisamment immobilisé par interaction avec la surface, et la partie protéique ne doit pas être trop mobile par rapport à l'ADN pour pouvoir être captée dans l'interaction pointe-échantillon.

La Figure 69 illustre la structure et la dynamique des objets hybrides sur la surface.

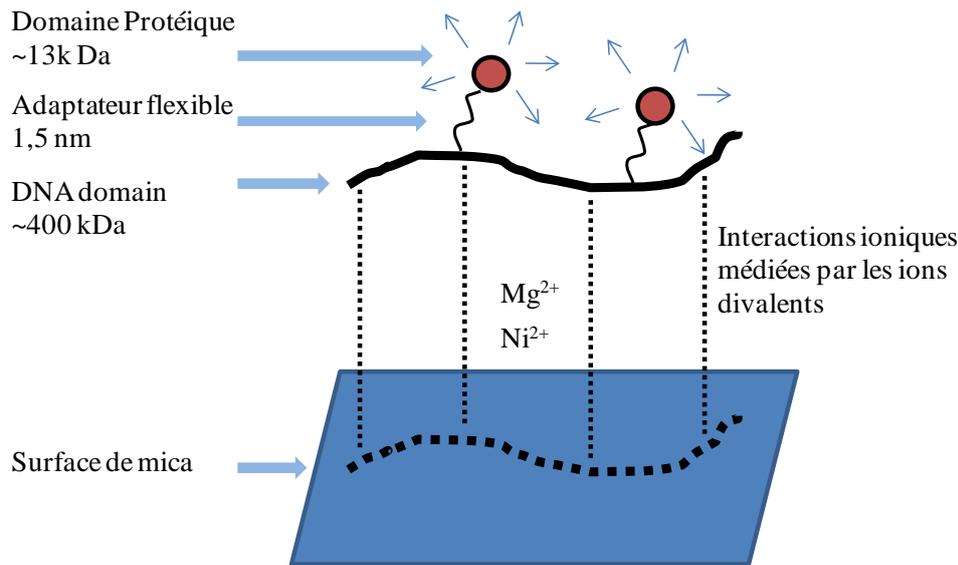


Figure 69 : schéma illustrant les propriétés dynamiques hétérogènes des ADN et protéines sur mica.

Les groupements phosphate des ADN, qui permettent d'ancrer les molécules, créent une interaction d'autant plus forte que la longueur des fragments est grande, réduisant leur dynamique au dessus d'une taille critique.

Les protéines, de petite taille (13 Kda), attachées aux séquences d'ADN par un adaptateur flexible ne sont pas forcément en interaction avec la surface. En effet, l'adaptateur synthétique de 1,5 nm autorise (si on considère la distance du centre de la partie globulaire du cytochrome b5 à l'ADN) la protéine à parcourir un cercle d'environ 3 nm de rayon. Ceci revient à ce que la protéine puisse occuper aléatoirement toutes les positions sur une surface d'environ 30 nm². En absence d'interaction spécifique avec le substrat, la dynamique de la protéine est alors la somme de sa dynamique intrinsèque et de celle de l'ADN.

Les propriétés dynamiques hétérogènes des objets hybrides impliquent donc des conditions différentes d'imagerie avec un compromis à trouver entre l'observation de la structure des objets et leur dynamique. L'analyse structurale nécessite des conditions d'immobilisation fortes qui aboutissent à l'observation des deux composants uniquement dans le cas où la protéine est également captée par la pointe. Ces conditions ne sont pas compatibles avec une analyse dynamique qui requiert des interactions avec la surface beaucoup plus faibles.

Vitesse relative balayage-échantillon

Les problématiques de résolution des structures mobiles soulèvent la question de la vitesse de balayage. Celle-ci doit être supérieure à la vitesse de déplacement des objets pour être en mesure de les imager sans artefacts. La faible reproductibilité de visualisation des protéines pourrait aussi s'expliquer par une mobilité des protéines trop élevée pour être

détectée aux vitesses de balayages utilisées (de l'ordre de 160 ms jusqu'à la seconde entre deux images) pour une surface de 256 x 256 pixels. Une expérience visant à diminuer la mobilité des protéines par exemple en augmentant la viscosité du milieu pourrait constituer une solution. Toutefois, cela impliquerait des contraintes difficiles à gérer avec l'appareillage actuel, la viscosité du milieu ayant une influence sur l'amortissement de l'oscillation de la pointe.

On pourrait également concevoir une expérience où l'ancrage des molécules sur la surface est assuré par la partie protéique comme dans le cas des membranes supportées. Néanmoins la faible durée de vie des complexes Ni-NTA imposerait de développer une approche covalente liant la partie C-terminale de la protéine au phospholipide modifié. La mobilité des structures serait alors définie par la diffusion des phospholipides au sein de la membrane, ce qui pourrait être contrôlé par des variations de température induisant une transition de phase vers une forme cristalline figée. Toutefois la réalisation de membranes supportées reste une technique délicate du fait de leur sensibilité à l'air qui déstructure la double couche lipidique.

3.5. Discussions complémentaires

L'AFM à haute vitesse s'est montré un outil très intéressant pour étudier les interactions faibles entre objets sur une surface. Notamment, nous avons pu analyser la dynamique des échantillons de pentamère de PDNA qui interagissent entre eux par l'intermédiaire des courtes extensions simple brin de 20 bases présentes à leurs extrémités. Alors que ces extensions ne sont pas complémentaires, elles peuvent néanmoins former de courts segments d'appariement de bases avec des énergies d'interaction en solution prédites à quelques kcal/mol (Figure 70). Celles-ci, qui sont insuffisantes pour former des complexes stables, sont malgré tout assez grandes pour initier des interactions significatives capables de générer la formation de complexes transitoires entre les polymères, au travers d'un attachement entre extrémités.

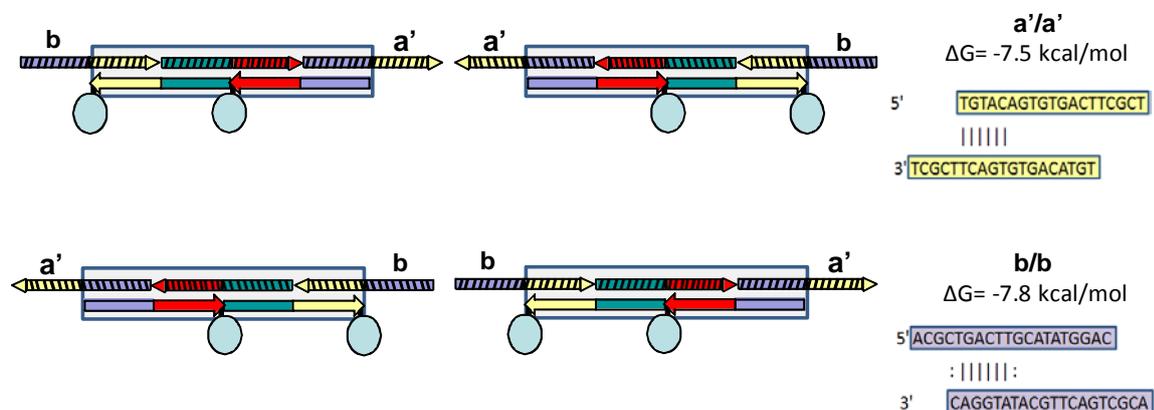


Figure 70 : Schéma des hybridations possibles entre échantillons PDNA 1. Le schéma présente les extensions simple brin dépassant b et a' ainsi que les énergies d'interaction prédites en solution entre les séquences b/b et a'/a'.

Les énergies d'interactions entre blocs sur la surface ont été caractérisées par analyse de la distribution des longueurs des polymères formés par les associations transitoires entre blocs de PDNA. Pour ce faire, une population d'une centaine d'objets a été analysée sur des images à faible résolution et la distribution de longueur a été mesurée en utilisant le logiciel Image J. La distribution expérimentale des longueurs de polymères a été simulée en utilisant un modèle qui prend en considération les interactions entre les extrémités similaires des monomères. Différentes énergies d'interactions ont été testées jusqu'à ce que la distribution simulée corresponde à la distribution expérimentale. Les énergies d'interactions déduites se sont révélées beaucoup plus basses que celles prédites en solution. Ceci peut être expliqué par les conditions expérimentales utilisées en HS-AFM où les forces ioniques sont plus faibles qu'en solution et vont aboutir à une augmentation de la répulsion électrostatique entre les groupements phosphate de l'ADN.

Les conditions d'imagerie imposées par la technique HS-AFM soulèvent la question de la pertinence des événements observés. En effet, les phénomènes observés dans des conditions de concentration de sels différentes des conditions physiologiques, ne seront pas forcément représentatifs de la réalité physiologique. Par ailleurs, le fait d'observer des systèmes biologiques sur une surface implique une mobilité des objets contrainte en 2D. Cette contrainte peut également introduire un biais d'interprétation des données du fait des conditions d'observation. La projection dimensionnelle d'objets créés en solution sur une surface peut aboutir à des défauts lors du placage des objets constituant un frein potentiel à l'utilisation cette technique pour l'analyse d'assemblages de grande taille.

L'HS-AFM peut s'appliquer à l'étude des mécanismes biologiques dans la limite où la dynamique intrinsèque des objets ou événements est inférieure à la vitesse de balayage du microscope. Ceci peut être atteint en mettant en place des conditions d'expérience qui réduisent la mobilité intrinsèque du système biologique par rapport à la vitesse de balayage. Toutefois, les problèmes d'analyse rencontrés dépendent fortement du matériel observé. Plusieurs cas peuvent être considérés et présenter des problématiques différentes :

- (i) les réseaux organisés, tels que les cristaux à 2D, où les objets au sein de la matrice sont relativement immobilisés. On peut citer l'exemple de la caractérisation AFM à haute résolution d'un cristal 2D de protéines membranaires (les aquaporines) [410], et de l'imagerie de leurs changements de conformation lors de l'addition de substrats ou d'inhibiteurs dans le tampon d'observation [411].
- (ii) les moteurs moléculaires où les molécules uniques, plus ou moins dynamiques, se déplacent sur un axe, ajoutant une dimension supplémentaire. C'est le cas de la myosine V marchant le long du filament d'actine [412]. Ces objets relativement gros (~100 kDa la chaîne lourde) présentent une forte interaction pointe-échantillon et une plus petite dynamique intrinsèque que de petits objets.

- (iii) l'étude d'objets isolés de petite taille (~10 kDa) pose plus de difficultés de détection du fait d'une plus faible interaction pointe-échantillon et d'une plus faible traînée hydrodynamique (résistance de frottement et de forme).

La taille des objets ainsi que leur mode d'ancrage constituent les points clé pour suivre les états différentiels des molécules. Plus la masse des objets est petite, plus la difficulté de détection est grande.

En conclusion, l'HS-AFM constitue un outil d'investigation des mécanismes et systèmes biologiques très prometteur du fait de développements techniques tels que l'acquisition sub-seconde, et la résolution spatiale, qui font toujours l'objet d'évolution. Le développement de supports d'observation est également en plein essor, avec la construction de supports biocompatibles, ou de supports nano-structurés...

4. Conclusion

Le couplage covalent des domaines protéiques et nucléiques en un bloc de construction unique constitue une alternative intéressante aux complexes non covalents utilisés dans la construction de structures hybrides. Le concept de PDNA constitue une solution adaptée pour les procédés d'assemblage multi-étapes utilisés en nanofabrication. Ces complexes synthétiques permettent de résoudre le problème de la fragilité des protéines qui impose des conditions d'assemblage douces. La caractérisation physico-chimique des PDNA fournit des informations précieuses qui peuvent aider à mieux comprendre le comportement de complexes nucléoprotéiques naturels. L'influence de la protéine dans les mécanismes d'hybridation et sur les propriétés hydrodynamiques des complexes a été mesurée. Bien que l'influence de la protéine soit significative, elle ne nuit pas à l'assemblage des structures, les PDNA exhibant des propriétés d'hybridation similaires aux ADN. Les PDNA synthétisés font intervenir le domaine protéique du cytochrome b5 et une séquence d'ADN simple brin. Les PDNA étant des systèmes modulaires, ils autorisent, *a priori*, l'association de toute protéine à tout acide nucléique. La technologie PDNA dépend toutefois de la possibilité de réaliser un couplage spécifique et régio-sélectif entre un résidu de la protéine et une séquence d'acides nucléiques. Cette nécessité et les multiples étapes de purification associées peuvent constituer un frein (néanmoins contournable) à son utilisation en nano-fabrication.

La microscopie de force atomique à haute vitesse constitue un nouvel outil de caractérisation, à l'échelle de la molécule unique, offrant une approche à la dynamique des objets biologiques. La pertinence des résultats obtenus par rapport aux comportements en solution reste cependant posée, du fait des perturbations liées aux interactions avec la surface.

3^{ème} partie

AUTO-ASSEMBLAGES MULTIPLEXES SUR PUCE

Les assemblages d'ADN et de PDNA ont été réalisés et caractérisés en solution homogènes et sur des surfaces. En nanofabrication, l'assemblage des nano-objets se fait de préférence sur une matrice solide. Ceci nous a amenés à étudier les mécanismes d'assemblage dans un milieu confiné.

Les micro- et nanotechnologies ont permis l'émergence dans les années 1990, de laboratoires sur puce, avec des applications dans le diagnostic clinique et la recherche pharmaceutique. Des systèmes, incorporant des dispositifs de micro-fluidique et autres miniaturisations, ont permis la création de micro-laboratoires capables, par exemple, de séparer des mélanges de molécules ou de synthétiser des structures complexes. La technologie des biopuces a ouvert la voie au criblage à haut débit avec l'analyse simultanée de multiples molécules et l'analyse des interactions entre composants biologiques.

Divers substrats (verre, silicium, or, mica...) présentant différentes propriétés physico-chimique de rugosité, hydrophobicité, groupements réactifs..., ont nécessité le développement de stratégies de fonctionnalisation variées. Celles-ci, selon la chimie de surface mise en œuvre, permettent la fixation des molécules, soit en deux dimensions pour des systèmes mono- et multicouches, soit en trois dimensions dans le cas des systèmes de matrice poreuse.

La fixation des composants biologiques en 2D, est adaptée au concept de puces à haute densité, et a l'avantage d'être reproductible. Néanmoins, la difficulté majeure est la fixation des molécules qui peut aboutir à leur inactivation du fait de problèmes d'orientation, d'organisation structurale ou de déshydratation. L'utilisation de matrice 3D, autorise une plus grande densité de fixation et supprime les contraintes d'interaction avec la surface. Toutefois, d'autres facteurs, tels que les barrières de diffusion, peuvent perturber les assemblages.

1. Présentation de l'article 4 : "Modeling and validation of multiplex proteo-nucleic self-assemblies monitored by Surface Plasmon Resonance Imagery"

La technique d'imagerie de résonance de plasmon de surface par (SPRi) a été choisie pour mesurer les vitesses d'assemblage des blocs d'ADN et de PDNA au sein d'une matrice poreuse. La SPR, présente l'avantage de pouvoir suivre en temps réel et sans marquage les interactions moléculaires qui se produisent au voisinage de la surface. La SPRi, du fait de l'imagerie offre des capacités de multiplexage absentes de la SPR conventionnelle. Nous l'avons déjà utilisée pour aborder l'assemblage des PDNA sur un support à deux dimensions, constitué d'une double couche lipidique ou d'une monocouche fusionnée sur une SAM (monocouche auto-assemblée, Self Assembly Monolayer). Toutefois, nous avons rencontrés plusieurs difficultés liées à :

- la rugosité du substrat d'or qui conditionne le succès et la qualité de la fusion membranaire et génère des soucis de reproductibilité,
- l'intensité du signal qui décroît, au fur et à mesure des étapes d'assemblage, du fait d'un encombrement de la surface,
- l'inhomogénéité de la surface, qui rend difficile la comparaison entre différentes zones de la puce et rend le choix des points de contrôle délicat.

L'ensemble de ces considérations nous a amenés à envisager l'utilisation d'une matrice poreuse plus facile à maîtriser. Le dextran est un polymère (α -D-glucopyranose) neutre et hydrophile, produit par la bactérie *Leuconostoc*. Les unités glucosyle sont liées majoritairement en α -(1,6) dans la chaîne principale et avec une proportion variable de branchements en α -(1,2), α -(1,3), α -(1,4) (Figure 71). Le pourcentage de ramification, leur arrangement spatiale ainsi que la masse molaire (10^5 à 10^9 Da) des polymères produits varient selon la souche productrice.

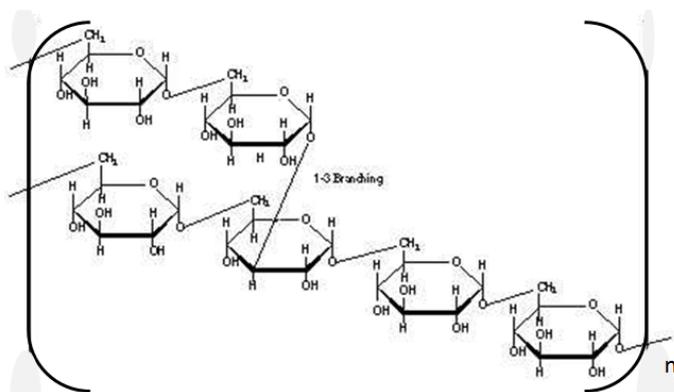


Figure 71 : Biopolymère dextran.

La fonctionnalisation d'une matrice dextran par l'avidine autorise la réalisation d'une puce multiplexée en utilisant diverses amorces d'oligonucléotides biotinylés. Dans ces conditions, nous avons pu analyser, en parallèle, différentes conditions de polymérisation des assemblages en tenant compte de la séquence, l'ordre, l'orientation et la nature des briques moléculaire injectées.

La matrice dextran n'est toutefois pas exempte de défauts, puisque sa structure peut affecter les interactions moléculaires, notamment au travers de l'accessibilité des amorces et du fait de restriction des phénomènes de diffusion. Nous avons développé un modèle qui prend en compte ces mécanismes au sein de la couche de dextran ainsi que les mécanismes de transport et de diffusion dans la phase aqueuse. Ce modèle nous a permis de comparer les résultats expérimentaux et simulés et de définir les corrections nécessaires pour prendre en compte ces phénomènes.

Nous avons également comparé différentes approches pour le calcul des vitesses d'hybridation entre les blocs DNA et PDNA. L'utilisation de ces méthodes couplée aux données de simulation, a permis d'extraire des traces expérimentales les paramètres

macroscopiques impliqués dans l'assemblage. Les méthodes d'analyse ont permis de mettre en évidence les paramètres qui étaient plus ou moins impliqués dans l'écart entre les observations expérimentales et les paramètres en solution.

L'approche SPRi multiplexée sur puce génère des données volumineuses qui ont pu être traitées rapidement, grâce au développement d'un logiciel. Celui-ci s'appuie sur des analyses statistiques permettant de normaliser, corriger, filtrer, de comparer et d'analyser plus extensivement les données, mais aussi, d'obtenir des profils automatiques des expériences réalisées. Les méthodes de classification développées sont applicables à n'importe quelle technique d'imagerie résolue dans le temps.

2. Appendice 5 : article 5

Modeling and validation of multiplex proteo-nucleic self-assemblies monitored by Surface Plasmon Resonance Imagery

Laisné A, Spadavecchia J, Moreau J, Canva M & Pompon D

Sensors and Actuators B: Chemical (2011), 160 : 1309-1315

<http://dx.doi.org/10.1016/j.snb.2011.09.067>



Modeling and validation of multiplex proteo-nucleic self-assemblies monitored by surface plasmon resonance imagery

Aude Laisné^a, Jolanda Spadavecchia^b, Julien Moreau^b, Michael Canva^b, Denis Pompon^{a,*}

^a CNRS, Centre de Génétique Moléculaire, UPR3404, Avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France

^b Laboratoire Charles Fabry (LCF), Institut d'Optique – Graduate School, Univ Paris Sud, CNRS, Campus Polytechnique, RD 128, 91127 Palaiseau Cedex, France

ARTICLE INFO

Article history:

Received 15 July 2011

Received in revised form

14 September 2011

Accepted 21 September 2011

Available online 29 September 2011

Keywords:

SPRI

Proteo-nucleic

Microenvironment

DNA-chip

Automated identification

Fitting methods

ABSTRACT

Hybridization based stepwise assembly of proteo-nucleic blocks to form hybrid linear structures was monitored by surface plasmon resonance imagery (SPRI). The multiplex approach involved the parallel extension on the chip of primers giving rise to assemblies differing by sequence, order and orientation of building blocks. Statistical analyses of reflectivity profiles following data filtering, normalization and corrections, illustrated that automated sorting of structures is feasible and that reproducible building of complex assemblies was achieved onto the dextran matrix. In order to characterize the influence of matrix environment, relations between kinetics of reflectivity changes and hybridization parameters measured into solution were investigated. A global model taking into account of transport and diffusion mechanisms into the aqueous phase and accessibility and hybridization mechanisms into the dextran layer, was set and numerically solved. The model was used to define critical assembly parameters and to derive the relations between microscopic and observable events in the experimental SPRI conditions. Different computing approaches for extracting rate constants from more or less noisy and truncated experimental traces were tested and validated. Finally the best processing methods were exploited, in association with simulation, to extract kinetic parameters. ANCOVA analysis of data was confirmatory of observations in solution concerning the influence of proteo-nucleic block orientation and the hybridization of DNA segments. However, interaction of protein domain with matrix within the dextran micro-environment was found to enhance hybridization rates of PDNAs as compared to DNA. In contrast, vicinity of hybridizing segments to dextran attachment decreased it.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Surface plasmon resonance (SPR) allows label-free real-time analysis of molecular interactions occurring in the vicinity of a surface [1–3]. In classical designs, like the Biacore system, average refractive index variations in a channel are measured by monitoring changes in the angle of resonance. SPR imagery approach (SPRI) differs in measuring at a fixed angle reflectivity changes associated with the shifts of plasmon curve [4,5]. SPRI offers the high multiplexing capabilities associated to imaging [6] but is generally found less sensitive than resonance angle monitoring [7]. Moreover, large imaging surfaces are more complex to handle from the fluidic and homogeneity points of view than the thin channels used in classical SPR [8]. Several turnarounds, such as surface nanostructuring [7,9,10], use of colloidal-gold nanoparticles [11,12] or of nanostructured digital microfluidics [13], have been developed to improve

SPRI sensitivity. However, most of commercial SPRI instruments involve uniform functionalized gold layer placed into a fairly large fluidic cell. While extensive studies were dedicated to investigate factors limiting time response in micro-channel based instruments [14–16], much less was documented concerning spatial and temporal resolution of SPRI systems [17]. Particularly, transport and diffusion mechanisms within the aqueous or polymer layer might significantly differ between the micro-channel and large fluidic cell set-ups [18]. Fluidic and matrix heterogeneities within cell can affect kinetic behaviors and interfere with automated identification of molecular interactions, thus limiting multiplexing power of imaging. Several surface chemistries were developed aiming at effective probe immobilization and avoiding non-specific binding [19–21]. Carboxy-methyl dextran matrix remained the more classical due to its high versatility and easy regeneration [19]. However, when formation of large biological assemblies is considered, it can be questioned how geometrical interactions with the matrix could interfere. The power and limitation of SPRI for the multiplex kinetic characterization of large self-assemblies were addressed using as model the multi-step hybridizations of proteo-nucleic (PDNA) building blocks. PDNAs are synthetic

* Corresponding author at: IMM-LISBP, UMR CNRS/INSA 5504, 135, Avenue de Rangueil, 31077 Toulouse Cedex 4, France. Tel.: +33 5 61 55 94 01.

E-mail address: denis.pompon@insa-toulouse.fr (D. Pompon).

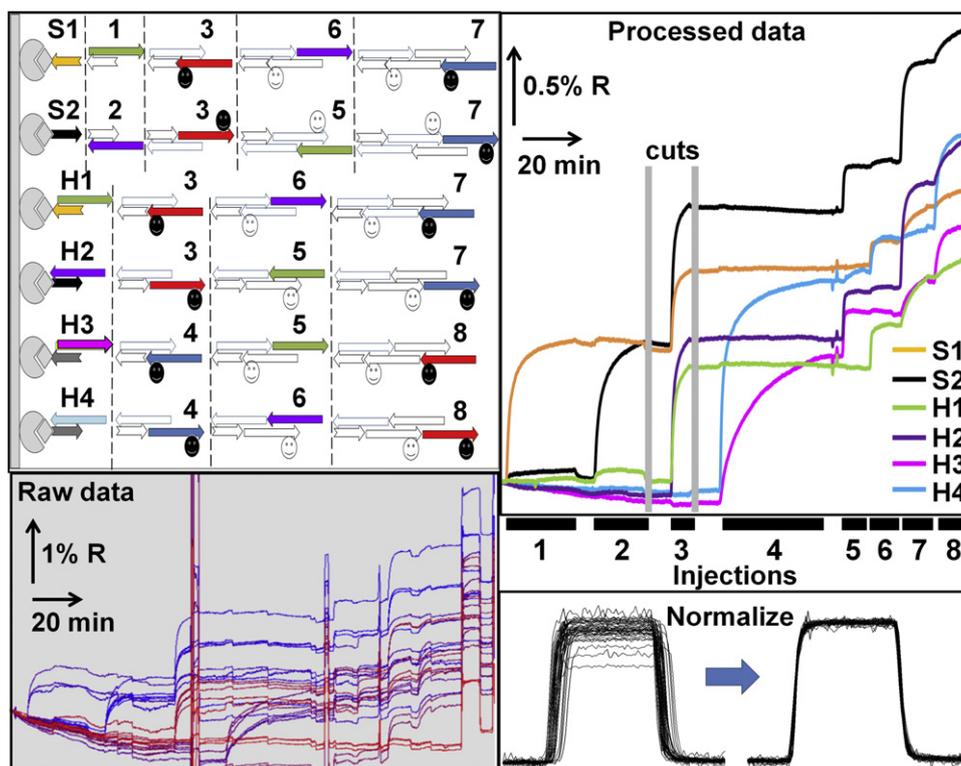


Fig. 1. SPRI experiment set-up and data processing. Upper left, flow chart of the experiment. S1, S2, H1, H2, H3 and H4 refer to the series of single and hybridized DNA probes spotted on the sensor biochip. Labels 1–8 refer to the subsequent injections. Each colored arrow corresponds to a different DNA sequence (PDNA for smile). Bottom left, raw SPR sensorgram (reflectivity) for individual spots as a function of time. Bottom right, salt jump calibration signal before (left) and after (right) normalization. Upper right, processed SPR sensorgram following normalization, background subtraction and smoothing of traces. The grey bars correspond to time cuts and curve joining to remove junky data (bubbles).

structures encompassing protein and nucleic domains in a unique covalent structure [22–24]. The protein domain can bring a wide range of functions while the nucleic domain acts similarly to simple oligonucleotide governing self-assembly mechanisms. Such assemblies feature fairly large size, multiple (protein and nucleic acids) components and were recently characterized in detail in solution [25]. They thus constitute an ideal model to address the effects of SPRI setting parameters on assembly mechanisms in supported matrix-based conditions.

2. Materials and methods

2.1. Materials, surface chemistry and spotting

Reagents and oligonucleotides are described in [Supplementary data \(SD\) 1](#) and [Table 1](#). Gold treatment, surface chemistry and spotting of DNA probes are described in [SD2](#).

2.2. SPR imaging

The surface plasmon resonance imaging (SPRI) system was described previously [20,26] and involved imaging (17.6 or 32.6 images/min) at 660 nm, with a CCD camera (1280 × 1024 pixels), a Teflon microfluidic cell (diameter 0.9 cm) with an average thickness of 125 μm and a total volume of ~10 μl driven by a flow rate of 56 μl min⁻¹. Probes ([SD Table 2](#)) were spotted in replicates onto the streptavidin functionalized dextran. The device was set-up on the SPRI apparatus and let to equilibrate with running buffer (1 × PBS) for 1 h. Eight subsequent injections of DNA and PDNA ligands were performed as described in [Fig 1](#) legend and detailed into [SD3](#) using an open circuit geometry before flow recycling after full stabilization of concentrations. Following each injection, cell was

flowed with buffer until trace stabilization. Calibration injections (1.25 × PBS) were inserted at two points in the series.

2.3. SPRI data processing

Dark-field image was subtracted from raw image sequences and TM intensities divided by corresponding TE values. Integration areas were chosen to avoid overlapping with border regions of spots. Response coefficients ($\Delta R/\Delta n$) of individual spots were normalized based on salt injections (from 1 × to 1.25 × PBS) that result into a refractive index change equivalent to 1.6 ng/mm² of protein or nucleic acids on the surface. Time-shifts associated with spot positions in the cell were corrected by modeling peak shape and calculating a suitable time offset by least square regression. Baseline correction was performed by subtracting traces from reference regions surrounding spots. For clarity, irrelevant events (bubble, electric spike, etc.) outside of kinetic traces were masked (and marked in figure) by deleting disturbed sequences and offsetting curves to reestablish continuity. Finally, polynomial curve smoothing, averaging of replicates and global offsetting were performed to facilitate trace comparisons. All described processing as well as statistical analyses and rate calculations were automated using the MagicPChip home-made software.

2.4. Modeling approaches

The SPRI cell was modeled as a biphasic system involving a mobile aqueous layer covering a static and porous dextran phase in which ligand diffusion occurred only from liquid interface. Flow rate was assumed to follow a classical laminar regime above the water–dextran interface. The SPR response coefficient ($\Delta \text{Reflectivity}/\Delta \text{Concentration}$) was considered to be

independent of the z position in the dextran layer assuming that dextran thickness is not large as compared to half-decrease length of the evanescent field. The density of binding sites within the dextran phase was supposed to be constant. However, consistency with experimental traces required modulating accessibility of binding sites in the matrix as a function of its saturation. An arbitrary but convenient function was defined as $F_{sat}(F(t)) = 1 - \sqrt{(C_{max} - C(t))/C_{max}}$ in which C_{max} and $C(t)$ are the saturating and actual concentrations of bound ligand at a given reaction time, respectively. F_{sat} stands for the ratio between the second order rate constant for hybridization in the dextran phase at a given ligand saturation and the corresponding constant in free solution. The function is normalized to 1 in the ligand free matrix and to 0 for full saturation. This describes that upon increasing saturation, possible access routes to binding sites become more crowded, thus decreasing apparent rate constant. Model parameters (SD Table 3) were adjusted to fit experimental conditions. Linear flow velocity was deduced from reflectivity change propagation within the cell following injection of high salt solution. Dextran total thickness was taken as 70 nm, flow rate $56 \mu\text{l min}^{-1}$, cell diameter 0.9 cm, average cell thickness 125 μm , and saturating concentration of ligand for dextran layer 0.3 pmole/ mm^2 . Diffusion coefficients in the aqueous phase for hybridizing DNA or PDNA were calculated from literature tables and length dependence law [27]. Default values used were 12 nm for the length of hybridizing DNA (or PDNA) and $4 \mu\text{M}^{-1} \text{min}^{-1}$ for second order rate constant for hybridization [25,27]. In the dextran phase, these coefficients were assumed to be reduced about 5-fold [28,29]. Equation solving involved model gridding (20 and 10 slices for water and dextran, respectively) in the z axis (distance from gold surface) and 5 slices of 10 μm in the y axis (distance from the spot border in the flow direction). The resulting set of differential equations was solved by finite element methods.

2.5. Methods for rate constants calculation

Three different approaches were tested to evaluate rate constants. The first is the least-square fitting based multi-exponential decomposition. Depending on experimental trace shape, one to three exponential components were taken and resulting rate constants (KFAST method) used as such or as an average value calculated from the ratio between fitted initial rate and total amplitude (IRTA method). The two other methods are two step procedures involving smoothing of experimental traces followed by alternate methods of rate calculation. Mechanism aware smoothing takes into account the relaxation nature of observed phenomenon. It involves non-negative least-square algorithm (NNLS) [30], aiming to solve multi-linear decomposition on function series by a least-square approach subject to the added constraint that coefficients do not include negative elements. Algorithm always converges requiring no cutoff parameter. We used two types of function series of exponential nature that differ by their monotonous or bimodal nature. The NNLS \ddagger series $F(t) = A_0 - A_i \times \sum_{i=1}^N \exp(-f_{\min} \times f^{(i-1)} \times (t - t_0))$ is constituted of rising exponential functions with regularly spaced rate constants. The NNLS \pm series $F(t) = A_0 + A_i \times \sum_{i=1}^N (-1)^i \times \exp(-f_{\min} \times f^{(i-1)} \times (t - t_0))$ with $f = \exp((\log(f_{\max}) - \log(f_{\min})) \div (N - 1))$ derives from the NNLS \ddagger series by switching the sign of function between each consecutive member of the series. NNLS \ddagger decomposition is formally equivalent to multi-exponential fitting with automated adjustment of the number of exponential terms. In contrast, NNLS \pm decomposition is more similar to wavelet decomposition and is formally equivalent to a Fourier transform filter with a sliding frequency window.

Provided of good complexation of experimental traces, NNLS \ddagger fitting generally allows safe estimation of slopes and amplitudes and

can be used similarly to classical multi-exponential decomposition to calculate rates by the IRTA method. In contrast, the NNLS \pm method, which performs the best within the range of experimental traces, is frequently divergent outside fitting range and thus is not recommended for evaluation of extrapolated amplitudes. In such a case, the global apparent rate method (GAR) is better suited. GAR does not require full amplitude evaluation and aims to evaluate a single exponential law that shares with experimental trace its initial rate as well as partial amplitude and slope at a second time characterized by the fact that the experimental rate reaches a given fraction of the initial one. This fraction can be chosen equal to 50%, 90% or 98% (GAR 50 to 98) depending on the available extend of experimental traces. For calculation, let take t_1 as the time where the derivative of the experimental trace (or of its fit function) $F(t)$ equals $((1 - m) \times S_{\max})$ where S is the larger derivative value (and t_0 corresponding time) of the fit function within observation interval. The apparent rate constant is then calculated as a ratio between the larger derivative of the function and the value of expression $[(F(t_1) - F(t_0)) \div m]$.

2.6. Automated sorting of hybridization behaviors

Traces associated to individual or groups of spots were automatically sorted using principal component analysis (PCA), multidimensional scaling (MDS) and hierarchical clustering (HC) analyses. Concerning PCA analysis, data were first restricted to the analysis time range and traces offset to zero at the initial time before calculation of PCA vectors and sorting by decreasing statistical significance [31]. Let define x_i, x_{i+1}, x_{i+2} as the projections on three consecutive vectors of the ranked PCA components. Values x, y and z are calculated as $x = x_i / (abs(x_i) + abs(x_{i+1}))$, $y = x_{i+1} / (abs(x_{i+1}) + abs(x_{i+2}))$, $z = abs(x_i) + abs(x_{i+1}) + abs(x_{i+2})$ and represent each spot on a 2D-map by a circle of coordinates x, y and of diameter z . For the representation given in Fig. 2, i was chosen as $i = 2$ but other choices can be more relevant depending on cases. Concerning HC, the projections on the first five principal vectors were used for hierarchical clustering and the results visualized using the TreeView software on the rectangular cladogram representation. Concerning MDS analysis, the distance matrix derived from the time profile of reflectivity for each spot was submitted to dimensional reduction [32] specifying an order of 5 for the dimensional space. To build false color images, two RGB coded images were built by concatenation in ImageJ software of differential images for injections 1-2-3 (R, G, B channels), and for injections 4-5-6 (R, G, B channels), respectively. The images were added using Photoshop following color shift of the second image to give the presented composite image.

3. Results and discussion

3.1. SPRI analysis of the sequential assembly of nucleic and proteo-nucleic structures

Multiplex SPRI analysis of the stepwise assembly of DNA and PDNA tiles to form complex linear structures was performed as described in Fig. 1. Multiple replicates of the different priming structures were spotted onto a gold-chromium sensor chip covered with an avidin functionalized dextran layer. The different priming structures, including or not double-stranded (ds) DNA spacers between the terminal biotin and the hybridizing sites, were attached to the matrix through biotin-avidin interaction. Extension of primers was performed through stepwise hybridizations of ss-DNA or PDNA blocks (19–20 bases). Tiles design was performed in order that each hybridization generates a new priming sequence. Depending on cases, extension occurred either on the 3'- or on the 5'-free extremity. To address the specific influence of the protein

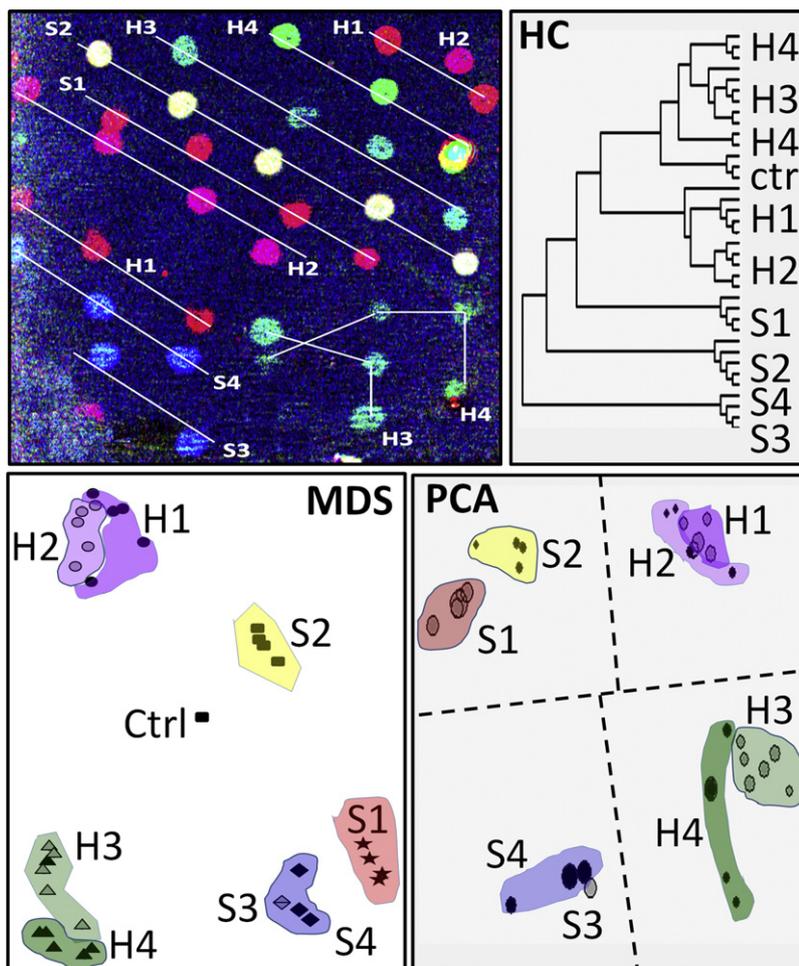


Fig. 2. Statistical analyses of SPRI data. Upper left, RGB reconstructed SPRI image summarizing the whole experiment. Reflectivity changes resulting from each injection were color coded. Thin lines are connecting spot replicates on the chip. Labels refer to Fig. 1 coding. Down left, graphical representation of multidimensional scaling analysis of SPRI traces. Colored areas enlighten replicates. Bottom right, graphical representation of principal component analysis of the same data. For easier presentation, the distances between clusters were reduced by cuts (dotted lines). Upper right, cladogram of the hierarchical clustering of PCA vector projections. Branches are individual spots on the chip.

domain on the process, assembly of DNA and PDNA tiles featuring identical nucleotidic sequences were compared. Eight sequential injections of reactants (detailed in SD3) were used to cover a wide range of situations addressing influences of matrix, chain length, tile size and protein domains on hybridization processes.

Raw SPRI describing differential reflectivities (ΔR) of spot surfaces as a function of time (Fig. 1, down-left) were submitted to several pre-processing: (i) a normalization to correct the fluctuation of the response coefficients ($\Delta R/\Delta C$) on chip surface; (ii) time shift of traces to compensate position-dependent fluidic delays; and (iii) subtraction of the signal of control regions to reduce thermal, electrical or optical drifts. Replicates averaging, polynomial smoothing, curve offsetting and deletion (marked) of accidentally disturbed time sequences were applied to generate more readable sensorgram as illustrated in Fig. 1 (top-right). Following each injection, multi-exponential shaped signals were observed for responsive spots. Considering the number of subsequent injections required to cover the full set of tested conditions, it was not manageable to allow each reaction to reach a final state. Therefore, reactions were quenched (buffer wash) when slopes of reflectivity changes reached 10% of initial values. As illustrated (Fig. 1, right), masses captured on spots during sequential injections diverged from expected stoichiometries. Experimental quenching of reactions before their completion is expected to account only for a minor part of observed divergences. However, this effect

could be under-estimated when the relation between the residual slopes of reactions and site saturation become non linear due to changes in the accessibility factor. Stoichiometry defects significantly attenuated after the first injections. This suggests that poorly accessible hybridization sites on the matrix might rapidly become non-competent for further reaction due to incomplete extension during initial steps. Remaining unhindered sites can thus be extended in later steps mostly independently of size, leading to yields approaching stoichiometry.

3.2. Automated analysis of reaction sequences associated to each spot

Image stacks corresponding to experiment of Fig. 1 were time segmented and differential images for each injection color coded (Fig. 2, top left). Replicates can be visually identified as spots of similar composite color, suggesting fairly good reproducibility of assemblies. However, automated identification on rational criteria remains a major goal for biochips. Three statistical approaches were tested. principal component analysis (PCA), which assumes that observed behaviors, can be decomposed into linear combinations of standard behaviors. Multidimensional scaling (MDS) is, in contrast, a non-linear dimensional reduction approach, while hierarchical clustering (HC) acts more as a non-metric classification method. Fig. 2 illustrates that the three approaches basically

gave rise to the similar spot classification with reliable identification of replicates independently on their position, shape or size. H1 and H2 groups are poorly differentiated by all methods, the same applying to H3 and H4 groups. H1 and H2 probes hybridize the same first and third reactants AA'b5 and BB'b5, respectively, but differ only by the distal and proximal orientation of the PDNA protein domain relative to the hybridizing site. The same situation is observed for H3 and H4 groups. This suggests that the orientation of protein domain is weakly interfering with respective hybridization reactions consistently with behaviors observed in solution [25]. S1 and S2 groups are expected to correspond to H1 and H2 groups following injections 1 and 2, respectively. S1 and S2 which correspond to constructs differing by the first injection in addition to their orientations are well differentiated by all the methods of analysis. S1 and H1 (reciprocally S2 and H2) also appear well differentiated while differing only by their behavior during the first injection. S3 and S4 groups (not presented in Fig. 1), which correspond to different constructs featuring identical behaviors during all injections are consistently not differentiated by any of the three methods. In conclusion, reliable identification of hybridization behaviors was achieved and is consistent with expected patterns.

3.3. Modeling of SPRI response of functionalized dextran biochip

Models were designed to predict effects of experimental parameters on the relations between apparent rates measured on the chip and microscopic hybridization parameters deduced from experiments in solution. This knowledge is of importance to quantitatively interpret changes in observed hybridization signal during the different steps of constructions. The global model (GM) illustrated in Fig. 3A takes into account cell properties as well as the diffusion mechanisms, density of grafting and the hybridization kinetics. It was set using experimental parameters of Figs. 1 and 2 and solved using finite element simulation approaches. The model was found to allow good fit of the rapid phases of experimental traces without parameter adjustment. However, accounting for the slow phases of traces required suitable definition of the accessibility function (see Section 2). This function describes the dependence on matrix saturation of the ratio between actual hybridization rate constants in the dextran matrix and corresponding values in solution. Satisfying results were achieved with the given empirical law. Reflectivity changes and reaction progress courses were deduced from simulation of the GM using different microscopic hybridization rates (from 0.25 to $8 \mu\text{M}^{-1} \text{min}^{-1}$), all other parameters being the experimental ones (Fig. 3B, insert). A "simplified model" (SM) was also designed which involves the same "accessibility function" but assumes the diffusion and transport phenomena to be non-limiting. The ratio between simulated reflectivity traces deduced from the two models was calculated (Fig. 3B). This ratio characterizes the influence of transport and diffusion mechanisms while minimizing effect of "accessibility function". The figure illustrates that transport and diffusion layers significantly affected only the first 10–20% of the reaction progress for actual hybridization rates lower than $0.5 \mu\text{M}^{-1} \text{min}^{-1}$ but that the differential reflectivity ratio remains low during at least 80% of the reaction progress when actual hybridization rates were in the 4 – $8 \mu\text{M}^{-1} \text{min}^{-1}$ range, making corrections necessary.

The relation between the apparent values for hybridization rate constants deduced from simulated reflectivity changes and the corresponding microscopic values in models is illustrated in Fig. 3C. For actual hybridization rates lower than $\sim 0.5 \mu\text{M}^{-1} \text{min}^{-1}$, the curve tends to a straight line with identical fitted and actual values. However, for the considered DNA and PDNA these rates are within the 4 – $8 \mu\text{M}^{-1} \text{min}^{-1}$ range [25]. Therefore in these conditions, apparent rate constants deduced from reflectivity

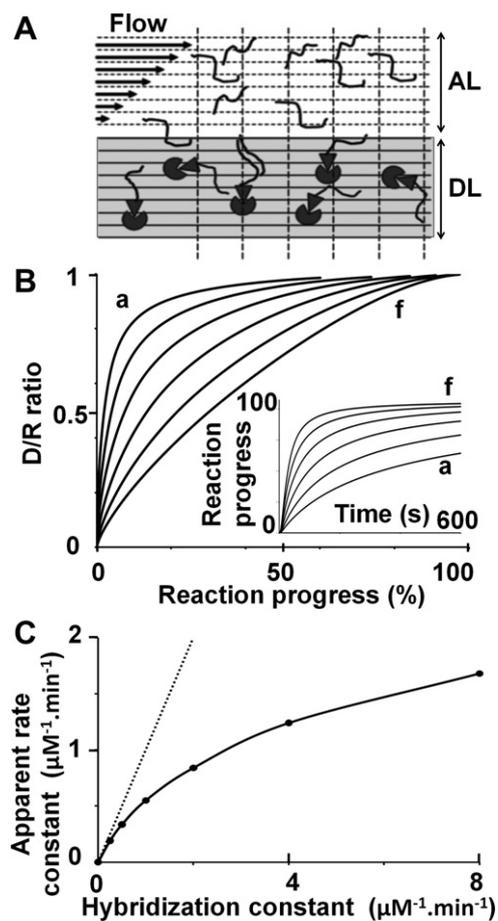


Fig. 3. SPRI modeling. (A) Schematic of the model used in simulations. AL stands for aqueous layer. Arrows symbolize flow rate profile close to the surface. DL stands for dextran layer including targets for DNA hybridization. (B) Ratios of respective simulated reflectivity traces for the full model of panel (A) and for a model assuming an infinite rate of diffusion and transport in the aqueous phase. Values are plotted as a function of reaction completion (% of saturation of the hybridization site in the dextran layer) using the same accessibility function for the two models. Insert, plot of reaction completion vs. time for the global model. Curves (a)–(f) correspond to hybridization rate constants of 0.25 , 0.5 , 1 , 2 , 4 and $8 \mu\text{M}^{-1} \text{min}^{-1}$, respectively. Other parameters are unchanged. (C) Calculated apparent first order rate constants (IRTA method in Section 2.4) using simulation global model as "experimental" data.

courses are expected to diverge from microscopic values by ~ 3 – 4 fold.

3.4. Extracting hybridization rates from experimental SPRI traces

As illustrated by simulations, minimizing effect of accessibility function required to preferentially analyze the initial part of reflectivity traces. In contrast, minimizing influences of mass transport and diffusion phenomenon would require preferentially working with the late parts of kinetics. Different methods were evaluated to define the best way to manage this problem. All approaches need first optimal noise filtering. Fig. 4A illustrates fitting error and corresponding standard deviations for a typical reflectivity curve using bimodal and monotonous NNLS filters (see Section 2 for detail). Efficient filtering but significant systematic deviations on the early and late phases of reflectivity traces were observed for monotonous NNLS and bi-exponential fitting. In contrast, bimodal NNLS filters removed noise without introducing any systematic deviation provided of suitable choice of the frequency window. Fits remained excellent even when traces significantly diverged from classical exponential aspects or exhibit lag phases. The fitted algebraic representations of reflectivity curves were used for further

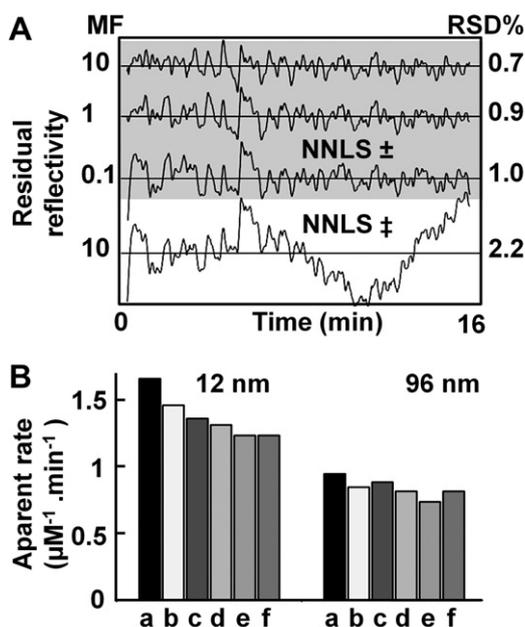


Fig. 4. Comparative methods of data filtering and rate constant calculations. (A) A noisy SPRI experimental trace was fitted using either bimodal exponential NNLS fitting (high frequency cut MF=10, 1, 0.1 min⁻¹) or a monotonous exponential fitting (MF=10 min⁻¹). Low frequency cut was always 0.01 min⁻¹. Fit residuals are illustrated and standard deviations (RSD) indicated. (B) Apparent first order rate constants were calculated using simulated reflectivity traces as data (Global Model with 12 and 96 nm long ligands). Biexponential (a and f) and NNLS (b–e) based fitting with bimodal (c and e) and monotonous (b and d) function sets. Rate constants were calculated as the KFAST (a) or the GAR50 (b and c), 98 (d and e), 100 (f) ratio methods (see Section 2.4).

evaluation of rate constants as detailed in Section 2. The classical KFAST method, which is based on predefined multi-exponential regression, was found unable to manage satisfyingly the lag phase frequently present on traces and suffers from the strong covariance of determined rate constants upon fit of multiphasic traces. A more consistent and generally applicable method to focus on the early part of traces while avoiding lag phase perturbations was defined. This uses the NNLS function to determine the maximal rate of reflectivity change and full amplitude estimation based on the time dependence of the slope of signal. When experimental trace reaches complexation, rate evaluation using the IRTA method is accurate. In the contrary, the GAR method is of interest for truncated kinetics that do not generally allow safe access to total reaction amplitudes. The calculation can be performed by taking into account a variable part (50% or 90%) or the full experimental trace and does not require trace extrapolation over experimental range.

In order to validate performance of the different methods, realistic Gaussian noise was added to computed reflectivity traces for a model simulating hybridization of 12 and 96 nm long DNA tiles on a dextran linked target (Fig. 4B). For traces reaching complexation, KFAST method gave results slightly closer to microscopic parameters used in the model than NNLS methods, while IRTA method following biexponential fit was the worst. However, when more realistic truncated kinetics were considered, NNLS methods were the sole to remain performing, while bi-exponential decomposition frequently diverged. Results also illustrated that restricting analysis to the initial part of traces (NNLS+GAR50) slightly improves results, while in a marginal way. Interestingly, values of apparent rates extracted by the best described methods (1.2–1.3 μM⁻¹ min⁻¹) nicely compared with model prediction (~1.25 μM⁻¹ min⁻¹ Fig. 3C) for the same experimental conditions. This means that combination of described methods for rate evaluation with correction curves resulting from simulation

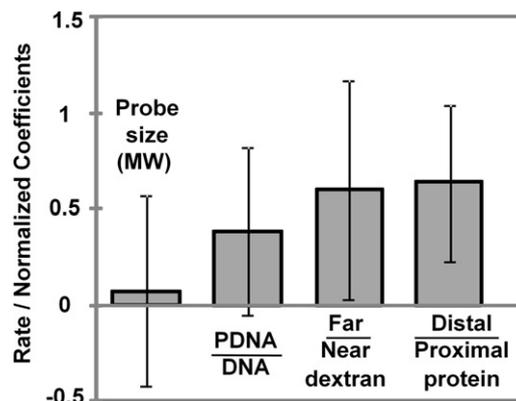


Fig. 5. Parameters modulating apparent hybridization rates on the dextran matrix. Apparent rate constants (SD Table 4) for the reactions described in Fig. 1 were calculated using the bimodal NNLS fitting and the GAR90 methods. ANCOVA analysis was performed on these rates using the MW of ligands as quantitative explicative parameters and PDNA or DNA nature (PDNA/DNA), the proximity of hybridization site to dextran matrix (far/near) and the distal/proximal orientations during hybridization as qualitative explicative parameters. Error bars correspond to the 95% confidence interval.

allow exact determination of the microscopic hybridization parameters.

3.5. Context dependent modulation of hybridization constants of DNA and PDNA building blocks during multistep assembly

Apparent rate constants for the hybridization steps of experiment described in Fig. 1 were calculated using methods described in Fig. 4 and finally converted to microscopic rate constants using simulation of model and curves of Fig. 3 (SD Table 4). Results were further analyzed for modulating factor using ANCOVA analysis. Fig. 5 illustrates that hybridization rates are not significantly influenced by the size of the dextran bound probes even if the concentration of accessible hybridization sites decreases during construct elongation. Surprisingly enough, PDNA reactants hybridize significantly faster than corresponding DNA of identical sequences, contrasting with the reported situation in solution [25]. This suggested that the PDNA protein domain might interact with the functionalized dextran matrix leading to increased local concentrations of the ligands. However, the effect appeared to be dependent on the relative orientations of the probes and interacting ligands during hybridization. Hybridization rates were found significantly higher when PDNA protein domain was distal to hybridizing segment. This observation is consistent with previous observations in solution. Finally, hybridization rates were found to be affected when the reaction occurred in vicinity of the dextran matrix as evidenced by the ANCOVA analysis and the effect of a spacer sequence. In all cases, apparent hybridization rates determined by SPRI are significantly smaller than the ones measured in solution based on hypochromic effect monitoring. However, if we take into account the correction deduced from the model presented in Fig. 3, corrected values generally appeared in close agreement with microscopic values deduced from experiment in solution. Nevertheless, corrected values for hybridization of PDNA tiles with a geometry where protein domain is distal to hybridization site were found significantly larger than expected confirming some interaction with the matrix.

4. Conclusions

Corrections and statistical analysis routines adapted to multiplex SPRI were developed allowing automated sorting of complex assemblies such as self-assembled proteo-nucleic structures.

Determination of microscopic hybridization constants and analysis of modulating parameters were performed by combining suitable data modeling and simulation. Experimental device was found to be convenient to determine hybridization rates under $0.5 \mu\text{M}^{-1} \text{min}^{-1}$ without correction and up to $8 \mu\text{M}^{-1} \text{min}^{-1}$ using microfluidic cell modeling. Functionalized dextran layer was found to marginally interfere with hybridization of pure nucleic acids compared to reaction in solution but to affect reaction rates when proteo-nucleic structures were involved. Dextran matrix prevents non-specific binding and provides high immobilization capacity for proteo-nucleic structures. However, uneven site accessibility was observed during stepwise elongation of assemblies leading to multiphasic and incomplete reactions. This raises the question to develop more adapted matrixes when quantitative solid phase synthesis is targeted. Developed data analysis and modeling tools adapted to SPRi can be of general usage to build up highly multiplex biochips with automated data analysis.

Acknowledgement

This work was supported by the grants ANR-05-NANO-021-04 from the PNANO actions of the National Research Agency (ANR), France.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.snb.2011.09.067.

References

- [1] J. Homola, S.S. Yee, G. Gauglitz, Surface plasmon resonance sensors: review, *Sens. Actuators B* 54 (1999) 3–15.
- [2] U. Jonsson, L. Fagerstam, B. Ivarsson, B. Johnsson, R. Karlsson, K. Lundh, S. Lofas, B. Persson, H. Roos, I. Ronnberg, et al., Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology, *Biotechniques* 11 (1991) 620–627.
- [3] S. Löfas, M. Malmqvist, I. Rönnerberg, E. Stenberg, B. Liedberg, I. Lundström, Bioanalysis with surface plasmon resonance, *Sens. Actuators B: Chem.* 5 (1991) 79–84.
- [4] J.M. Brockman, B.P. Nelson, R.M. Corn, Surface plasmon resonance imaging measurements of ultrathin organic films, *Annu. Rev. Phys. Chem.* 51 (2000) 41–63.
- [5] E.A. Smith, R.M. Corn, Surface plasmon resonance imaging as a tool to monitor biomolecular interactions in an array based format, *Appl. Spectrosc.* 57 (2003) 320A–332A.
- [6] C. Boozer, G. Kim, S. Cong, H. Guan, T. Londergan, Looking towards label-free biomolecular interaction analysis in a high-throughput format: a review of new surface plasmon resonance technologies, *Curr. Opin. Biotechnol.* 17 (2006) 400–405.
- [7] L. Malic, B. Cui, T. Veres, M. Tabrizian, Enhanced surface plasmon resonance imaging detection of DNA hybridization on periodic gold nanoposts, *Opt. Lett.* 32 (2007) 3092–3094.
- [8] K. Campbell, T. McGrath, S. Sjolander, T. Hanson, M. Tidare, O. Jansson, A. Moberg, M. Mooney, C. Elliott, J. Buijs, Use of a novel micro-fluidic device to create arrays for multiplex analysis of large and small molecular weight compounds by surface plasmon resonance, *Biosens. Bioelectron.* 26 (2011) 3029–3036.
- [9] K.M. Byun, S.J. Yoon, D. Kim, S.J. Kim, Experimental study of sensitivity enhancement in surface plasmon resonance biosensors by use of periodic metallic nanowires, *Opt. Lett.* 32 (2007) 1902–1904.
- [10] A. Abbas, M.J. Linman, Q. Cheng, Patterned resonance plasmonic microarrays for high-performance SPR imaging, *Anal. Chem.* 83 (2011) 3147–3152.
- [11] L. He, M.D. Musick, S.R. Nicewarner, F.G. Salinas, S.J. Benkovic, M.J. Natan, C.D. Keating, Colloidal Au-enhanced surface plasmon resonance for ultrasensitive detection of DNA hybridization, *J. Am. Chem. Soc.* 122 (2000) 9071–9077.
- [12] W. Lee, B.K. Oh, Y.W. Kim, J.W. Choi, Signal enhancement of surface plasmon resonance based on gold nanoparticle-antibody complex for immunoassay, *J. Nanosci. Nanotechnol.* 6 (2006) 3521–3525.
- [13] L. Malic, T. Veres, M. Tabrizian, Nanostructured digital microfluidics for enhanced surface plasmon resonance imaging, *Biosens. Bioelectron.* 26 (2011) 2053–2059.
- [14] D.G. Myszkka, X. He, M. Dembo, T.A. Morton, B. Goldstein, Extending the range of rate constants available from BIACORE: interpreting mass transport-influenced binding data, *Biophys. J.* 75 (1998) 583–594.
- [15] C. Wofsy, B. Goldstein, Effective rate models for receptors distributed in a layer above a surface: application to cells and Biacore, *Biophys. J.* 82 (2002) 1743–1755.
- [16] M.L. Yarmush, D.B. Patankar, D.M. Yarmush, An analysis of transport resistances in the operation of BIACORE; implications for kinetic studies of biospecific interactions, *Mol. Immunol.* 33 (1996) 1203–1214.
- [17] A.B. Dahlin, J.O. Tegenfeldt, F. Hook, Improving the instrumental resolution of sensors based on localized surface plasmon resonance, *Anal. Chem.* 78 (2006) 4416–4423.
- [18] E. Bouffartigues, H. Leh, M. Anger-Leroy, S. Rimsky, M. Buckle, Rapid coupling of surface plasmon resonance (SPR and SPRi) and protein chip based mass spectrometry for the identification of proteins in nucleoprotein interactions, *Nucleic Acids Res.* 35 (2007) e39.
- [19] I. Mannelli, L. Lecerf, M. Guerrouache, M. Goossens, M.C. Millot, M. Canva, DNA immobilisation procedures for surface plasmon resonance imaging (SPRi) based microarray systems, *Biosens. Bioelectron.* 22 (2007) 803–809.
- [20] J. Spadavecchia, J. Moreau, J. Hottin, M. Canva, New cysteamine based functionalization for biochip applications, *Sens. Actuators B* 43 (2009) 139–143.
- [21] S. Tombelli, M. Mascini, A.P. Turner, Improved procedures for immobilisation of oligonucleotides on gold-coated piezoelectric quartz crystals, *Biosens. Bioelectron.* 17 (2002) 929–936.
- [22] W. Boireau, A.C. Duncan, D. Pompon, Bioengineering and characterization of DNA-protein assemblies floating on supported membranes, *Methods Mol. Biol.* 300 (2005) 349–368.
- [23] W. Boireau, J.C. Zeeh, P.E. Puig, D. Pompon, Unique supramolecular assembly of a redox protein with nucleic acids onto hybrid bilayer: towards a dynamic DNA chip, *Biosens. Bioelectron.* 20 (2005) 1631–1637.
- [24] D. Pompon, A. Laisné, PDNA as building blocks for membrane-guided self-assemblies, *Biochem. Soc. Trans.* 35 (2007) 495–497.
- [25] A. Laisné, M. Ewald, T. Ando, E. Lesniewska, D. Pompon, Comparative physico-chemical and self-assembly properties of DNA and proteo-nucleic building blocks in solution and on surfaces, *Bioconjugate Chem.* 22 (9) (2011) 1824–1834.
- [26] J. Hottin, J. Moreau, G. Roger, J. Spadavecchia, M.C. Millot, M. Goossens, M. Canva, Plasmonic DNA: towards genetic diagnosis chips, *Plasmonics* 2 (2007) 201–215.
- [27] A. Laisné, E. Lesniewska, D. Pompon, Design and experimental validation of a generic model for combinatorial assembly of DNA tiles into 1D-structures, *Biochim. Biophys. Acta Gen. Subj.* 1810 (2011) 603–611.
- [28] A. Pluen, P.A. Netti, R.K. Jain, D.A. Berk, Diffusion of macromolecules in agarose gels: comparison of linear and globular configurations, *Biophys. J.* 77 (1999) 542–552.
- [29] N.C. Stellwagen, Electrophoresis of DNA in agarose gels, polyacrylamide gels and in free solution, *Electrophoresis* 30 (Suppl. 1) (2009) S188–S195.
- [30] C.L. Lawson, J.H. Hanson, *Solving Least Squares Problems*, Classics in Applied Mathematics ed., SIAM, Philadelphia, 1995.
- [31] F. Murtagh, Principal component analysis, in: ST-ECF/ESA/ESO, Garching-bei-Muenchen, 1986, <http://www.lthe.fr/PagePerso/ylu/program/PCAF>.
- [32] F. Murtagh, Classical multidimensional scaling, 1986, <http://www.classification-society.org/csna/mda-sw/cmds.f>.

Biographies

Aude Laisné is a Ph.D. research scholar at Institut Carnot de Bourgogne, Dijon, France in collaboration with Centre de Génétique Moléculaire, Gif-sur-Yvette. She is working on synthetic proteo-nucleic hybrid structures self-assemblies using biochemistry and single molecule approaches.

Jolanda Spadavecchia obtained her degree in Pharmaceutical Chemistry and Technology from the University of Bari, Italy following postgraduate studies at the University of Lecce. Her research interests concern the synthesis of organic compounds, sensor applications and characterization. Now, her current activity focuses on surface biomolecular interactions and gold nanoparticles.

Julien Moreau received his Ph.D. degree in Paris VI University, for studies related to the gravitational wave detector. He is assistant professor at the Institut d'Optique in Palaiseau, France. His research interests include surface plasmon resonance imaging sensors and their applications in biomolecular interactions studies as well as time-domain optical coherence tomography for biomedical studies.

Michael Canva received his Ph.D. degree in Orsay, France, for studies related to dye doped sol-gel materials for optical application. He is Director of research at CNRS heading the "Biophotonic" group at Institut d'Optique in Palaiseau, France. His current activity focuses on structured plasmonics and surface biomolecular interactions as characterized by SPRi.

Denis Pompon is Director of research at CNRS and is PI of the Molecular and Metabolic Engineering group at LISBP in Toulouse, France. Initially educated as a chemist, he extended his interest to protein and metabolic engineering and more recently to nano-biotechnology. His current research team develops multidisciplinary works in synthetic biology associating genetic, molecular biology, biochemical and biophysical approaches.

Supplementary data (SD) section for:

**Modeling and validation of multiplex proteo-nucleic self-assemblies monitored by
Surface Plasmon Resonance Imagery**

Aude Laisné^a, Jolanda Spadavecchia^b, Julien Moreau^b, Michael Canva^b, Denis Pompon^{a,*}

SD1: Materials.

11-Mercapto-1-undecanol, 1x phosphate buffer saline (PBS, Sigma P4417), bis-2-methoxyethyl ether (diglyme), epichlorohydrine, dextran (425-575 kDa), bromo-acetic acid, N-hydroxysuccinimide (NHS), N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC), streptavidin and solvents used to prepare reagent and buffer solutions were purchased from sigma Aldrich (France). Oligonucleotides and probes are described in SD Tables 1 and 2. PDNAs were synthesized by coupling covalently a 3'-amino-linker modified oligonucleotide to an engineered cytochrome b5 protein as previously described [1].

SD2: Surface chemistry onto gold surface and spotting

Glass slide (BK7, refractive index= 1.515) covered with an extra-thin layer (2.5 nm) of chromium on which a 47 nm gold film was vapor deposited was used as substrate. This gold thickness was chosen in order to maximize the SPR signal. The gold layer was cleaned by immersion and gentle sonication in extemporaneously prepared hot acidic "Piranha" solution (equal volume of 30% hydrogen peroxide and pure sulfuric acid). The surface was extensively rinsed with MilliQ water and stored in HPLC grade acetonitrile until use. The dextran chemistry procedure was performed as described [2-4] except that streptavidin (20µM) was dialyzed against 50 mM phosphate buffer pH 7.5 prior a 4 h coupling to the NHS and EDC activated surface. Excess of streptavidin was removed by washing with 1 x PBS buffer followed by water and the surface minimally air dried, before immediate spotting with biotinylated probes. DNA probes (SD Table 2) diluted at a concentration of 4 µM in 1 x PBS, 10 % glycerol solution were spotted in an 8 x 8 matrix onto slides using home-made coated glass pins and a Qiagen 8000 robot.

SD table 1: DNA sequences involved in assemblies.

Name	Sequence (5' - 3')	Bases
A-A'	CCGTAGGGTCTCCATTTTGA-AGCGAAGTCACACTGTACA	39
B-B'	GTCCATATGCAAGTCAGCGT-CAGTGTACAGTCACGTCAG	39
a-b'	TCAAATGGAGACCCTACGG-CTGACGTGACTGTACACTG	39
b-a'	ACGCTGACTTGCATATGGAC-TGTACAGTGTGACTTCGCT	39
c-b'	ACGACCATGTTGTTGGGATC-CTGACGTGACTGTACACTG	39
b-c	ACGCTGACTTGCATATGGAC-ACGACCATGTTGTTGGGATC	40
B	GTCCATATGCAAGTCAGCGT (3'-biotin)	20
B'	CAGTGTACAGTCACGTCAG (5'-biotin)	19
C	GATCCCAACAACATGGTCGT (3'-biotin)	20
C*	GATCCCAACAACATGGTCGT (5'-biotin)	20

DNA sequences A, A', B, B' and C (or C*) are inverse complementary to a, a', b, b' and c, respectively. The 5'- and 3'- biotin modifications are involved in the binding of probes to the streptavidin functionalized dextran matrix. AA' and BB' sequences were synthesized with and without the 3'-amino-linker modification which is involved in PDNA building blocks for protein attachment. Oligonucleotides (Eurogentec) were purified, when necessary, by ion-pairing reverse phase chromatography.

SD Table 2: Probes attached to the biochip by streptavidin-biotin interaction.

Probe	Structure	Orientation	Spacer*
S1	Biotin-B	free 5'-	no
S2	Biotin-B'	free 3'-	no
S3	Biotin-C	free 5'-	no
S4	Biotin-C	free 3'-	no
H1	Biotin-B/b-a'	free 5'-	yes
H2	Biotin-B'/ab'	free 3'-	yes
H3	Biotin-C/cb'	free 5'-	yes
H4	Biotin-C/bc	free 3'-	yes

*Hybridizing segment is separated to matrix dextran by double-stranded DNA region.

SD Table 3: Model parameters

Parameter	Values	Description
Rise Time	11.2 s	Salt jump rise time (SPR signal, at center of cell)
Cell thickness	125 μm	Average cell thickness
Cell section	9 mm	Cell section at cell center
Flow rate	56 $\mu\text{l}/\text{min}$	Flow rate through the cell
Ligand size	12 or 96 nm	Depends on simulations
MaxTime	600 s	Simulation time
NStep	150	Number of time steps (plotting)
NSlice	10	Number of layers (dextran phase)
NSlice	20	Number of layers (aqueous phase)
LayerSizeW	25 μm	Limit layer thickness (aqueous phase)
DLayerSize	70 nm	Dextran layer thickness
TargetC	0.3 pmole/ mm^2	Probe concentration in the dextran layer
Hc	0.25 to 8 $\mu\text{M}^{-1} \cdot \text{Min}^{-1}$	Used hybridization constants (solution)
MatxDif	0.2	Ratio of diffusion coefficient of ligand in the dextran phase compared to solution
DifC	$6/(\text{LigandSize}^{0.59})$ $\mu\text{m}^2 \cdot \text{s}$	Diffusion coefficient of ligand in solution
CellLen	1 μm	Length of a slice in the flow vector direction
CelNum	5	Number of slices in the flow vector direction

SD3: Details of injection sequence for Figure 1 experiment.

All DNA and PDNA solutions were dialyzed against running buffer to avoid artifact linked to change in salt concentrations. The following sequence of injection was used: ba' DNA (2 μ M), ab' DNA (2 μ M), AA'-b5 PDNA (0.5 μ M), BB'-b5 PDNA (0.5 μ M), ba' DNA (2 μ M), ab' DNA (2 μ M), BB'-b5 PDNA (0.5 μ M) and AA'-b5 PDNA (0.5 μ M). Therefore, upon injections 1 and 2, ba' and ab' DNA hybridized through their (b) and (b') part on spots S1 and S2, making them equivalent to H1 and H2. Upon injection 3, AA'-b5 PDNA hybridized onto S1, S2, H1 and H2 spots, whereas (BB'-b5 PDNA) on injection 4 hybridized onto H3 and H4 spots. The protein domain of PDNAs was proximal to the hybridizing site for S1, H1 and H3 spots and distal for S2, H2 and H4 spots. Upon injections 5 and 6 (ba' and ab' DNA, respectively) S2, H2, H3 and S1, H1, H4 spots were targeted, respectively. Following, injections 7 and 8 (BB'-b5 and AA'-b5 PDNAs, respectively) S1, S2, H1, H2 and H3, H4 spots were targeted, respectively. Each injection was separated from the next one by running the cell with 1x PBS buffer until signal stabilization.

SD Table 4: Apparent and corrected rate constants from SPRI.

Probes size (kDa)	6	18	42	54
Injections *	1 and 2	3 and 4	5 and 6	7 and 8
Hybridizing segment	Apparent rate constants ($\mu\text{M}^{-1} \cdot \text{min}^{-1}$)			
A	-	2.13 \pm 0.9	1.36 \pm 0.2	2.04 \pm 0.2
A'-b5	-	1.66 \pm 0.7	1.06 \pm 0.1	1.65 \pm 0.3
B	0.61 \pm 0.3	2.71 \pm 0.1	1.45 \pm 0.2	2.20 \pm 0.17
B'-b5	-	0.50 \pm 0.01	0.52 \pm 0.04	0.58 \pm 0.09
B'	0.34 \pm 0.03	-	-	-
	Corrected rate constants ($\mu\text{M}^{-1} \cdot \text{min}^{-1}$)			
A	-	>8	4.9	>8
A'-b5	-	7.8	3	7.7
B	1.20	>8	5.7	>8
B'-b5	0.51	0.85	0.9	1.1

* DNA tiles were injected in injections 1, 2, 5 and 6 while PDNA tiles were injected in injections 3, 4, 7 and 8. See Figure 1 for the flow chart of the experiment and SD3 for details of injections.

References

- [1] A. Laisne, M. Ewald, T. Ando, E. Lesniewska, D. Pompon, Comparative physico-chemical and self-assembly properties of DNA and proteo-nucleic building blocks in solution and on surfaces., *Bioconjugate Chem.*, (2011) Article ASAP.
- [2] S. Löfas, B. Johnsson, A Novel Hydrogel Matrix on Gold Surfaces in Surface Plasmon Resonance Sensors for Fast and Efficient Covalent Immobilization of Ligands, *J. Chem. Soc. Chem. Commun.*, 21 (1990) 1526-1528.
- [3] I. Mannelli, L. Lecerf, M. Guerrouache, M. Goossens, M.C. Millot, M. Canva, DNA immobilisation procedures for surface plasmon resonance imaging (SPRI) based microarray systems, *Biosens. Bioelectron.*, 22 (2007) 803-809.
- [4] S. Tombelli, M. Mascini, A.P. Turner, Improved procedures for immobilisation of oligonucleotides on gold-coated piezoelectric quartz crystals, *Biosens. Bioelectron.*, 17 (2002) 929-936.

3. Discussion complémentaire de l'article 4

3.1. Avantages et limitations de la SPRi

Les avantages de la technique de résonance des plasmons de surface (SPR) sont :

- l'absence de marquage des molécules qui simplifie la préparation des échantillons et s'affranchit des biais potentiels induits par le marquage.
- le suivi en temps réel des interactions moléculaires qui permet d'obtenir des informations dynamiques sur des complexes formés lors de l'interaction. Notamment la SPR permet de caractériser des interactions de faible affinité qui échapperaient à des mesures statiques réalisées en point final.
- L'imagerie SPR offre, en outre, la possibilité de réaliser des analyses multiplexées de plusieurs dizaines (voire centaines) d'assemblages en parallèle sur une même puce.

Les propriétés fluidiques des capteurs SPRi sont liées à la géométrie de la cellule (diamètre 0.9 cm et épaisseur de 125 μm) et diffèrent du système micro-fluidique des appareils de SPR classique, de type Biacore. L'analyse de données de SPRi nécessitent des corrections, notamment pour pouvoir comparer différentes zones de la puce.

Une difficulté est le développement d'une chimie de fixation des objets qui permet d'optimiser le signal dans des conditions où les interactions non spécifiques sont minimalisées. Cet objectif implique de contrôler la densité et l'orientation de fixation des objets tout en conservant leur activité. Les chimies de surface peuvent être classées en deux groupes correspondant à des supports en deux ou trois dimensions. Nous avons utilisé ces deux environnements pour suivre l'assemblage hiérarchique de blocs de construction ADN et PDNA, dans les différents articles (2, 3 et 4).

Supports 2D / 3D

Les supports 2D facilitent les mesures quantitatives des interactions moléculaires, du fait de l'absence de matrice limitant la diffusion, mais correspondent à des sensibilités plus faible car la densité d'objets fixés est moindre que dans un environnement en trois dimensions. Nous avons utilisé des membranes supportées qui optimisent la spécificité d'interaction, mais rend difficile le contrôle de la saturation de la surface. Une valeur trop élevée de celle-ci interfère avec les assemblages et une valeur trop faible limite la sensibilité.

Les supports 3D sont avantageux en termes de sensibilité du fait de leur capacité de fixation plus élevée pour un niveau équivalent de contrainte sur l'élongation des structures. Nous avons cependant observé une hétérogénéité d'accessibilité des amorces dans la matrice poreuse et des phénomènes de barrière de diffusion qui perturbent les stœchiométries et créent des biais dans le calcul des vitesses d'interactions. Dans ces conditions, il est nécessaire

d'appliquer des corrections sur les données expérimentales afin de compenser ces phénomènes.

Inhomogénéité des surfaces

Le support 3D telle que la matrice poreuse permet également de compenser en les masquant par une couche d'interface, l'inhomogénéité du substrat à laquelle les approches 2D sont plus sensibles. Les fluctuations de la rugosité, de l'hydrophobicité et les contaminations de surface sont des sources d'hétérogénéité qui sont déterminantes pour l'assemblage de la couche active.

- L'hydrophobicité de l'or a pour conséquence sa contamination par de nombreuses substances chimiques de l'environnement et tends à fixer les nanoparticules. C'est pourquoi une étape préalable de nettoyage est nécessaire avant toute fonctionnalisation.
- La rugosité de l'or est également une source d'inhomogénéité de la surface qui va influencer l'organisation des couches supérieures pouvant créer des micro-domaines. Des protocoles de recuit de l'or peuvent être réalisés pour tenter de générer des surfaces d'or cristallines.
- Une autre source d'inhomogénéité de l'or peut provenir de la diffusion de la couche de chrome qui est utilisée comme accroche de l'or sur le support en verre. Le dépôt d'or en l'absence du chrome a abouti à une couche très instable difficile à gérer.

Nanostructuration vs matrice poreuses

Deux méthodes sont utilisables pour augmenter la sensibilité donc la densité : la nanostructuration des surfaces pour augmenter leur surface spécifique ou la construction d'une matrice poreuse en 3D.

Dans le cas des matrices poreuses, les facteurs à prendre en compte dans le choix d'un polymère sont : (i) l'hydrophilicité, (ii) les charges, (iii) l'épaisseur dans le but d'obtenir le moins d'interactions non spécifiques. La « chimie dextran » est une fonctionnalisation largement utilisée, mais elle présente des problèmes d'hétérogénéité qui proviennent du manque de contrôle des étapes multiple de fonctionnalisation et des propriétés du polymère, telles que son épaisseur, la largeur de ses mailles, etc. Nous avons vu que le dextran, qui est un polymère hydrophile neutre, adsorbe plus les PDNA que les ADN. Ceci suggère que ces effets sont probablement dépendants des points isoélectriques des protéines considérées.

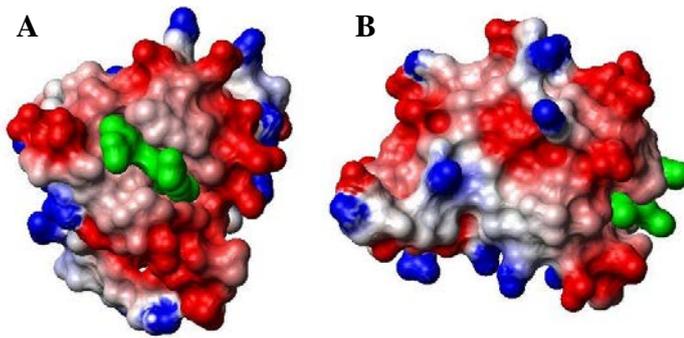


Figure 72 : Potentiel électrostatique du cytochrome b5.

A, vue du dessus par rapport à l'hème (en vert), B, vue longitudinale. Le potentiel électrostatique du cytochrome b5 a été calculé à l'aide du logiciel APBS (module Pymol), en utilisant les conditions standards de solvant (tampon PO_4 20 mM, pH 7.4). La surface (Connolly) du cytochrome b5 a été calculée en utilisant un rayon de 2,4 Å (molécule d'eau). La variation de potentiel appliquée sur la surface est représentée par un gradient allant du rouge (potentiel négatif, maximum -5 kT/e) à bleu (potentiel positif, +5 kT/e), d'après Gilles Truan.

Il est possible d'envisager d'autres polymères que les dextrans, tels que le polyéthylène glycol (PEG), qui est largement reconnu pour ses faibles propriétés d'absorption des molécules biologiques [413, 414]. Toutefois l'état de charge adéquate pour limiter les interactions varie selon la molécule considérée (lipide, acides nucléiques ou protéines), et reste à déterminer pour les objets hybrides. Les mailles du polymère et son épaisseur constituent également des facteurs critiques dans la capacité de fixation des objets et la présence de phénomènes de restriction de la diffusion.

Le développement de surfaces nano-structurées pourrait constituer une solution alternative pour augmenter la capacité d'immobilisation des assemblages sans créer d'encombrement. Par ailleurs, cette approche présente le double avantage d'augmenter la sensibilité du signal SPR du fait de l'exploitation de la modulation spatiale du champ électromagnétique à l'échelle nanométrique. La difficulté est cependant de maîtriser les procédés de structuration sur or aux dimensions nanométriques.

Fluidique

Le nombre de Reynolds ($Re = \frac{\rho dV}{\eta} = \frac{dV}{\nu}$) avec ρ la masse volumique du fluide (kg.m^{-3}), d la taille caractéristique du système (m), V une vitesse caractéristique (m.s^{-1}), η la viscosité dynamique du fluide ($\text{kg.m}^{-1}.\text{s}^{-1}$), ν la viscosité cinématique du fluide ($\text{m}^2.\text{s}^{-1}$) est une grandeur sans dimension utilisée pour appréhender les caractéristiques d'un régime d'écoulement. Dans les systèmes microfluidiques, certains phénomènes physiques deviennent prépondérants (la capillarité) quand d'autres deviennent négligeables (la gravité) par rapport aux phénomènes macroscopiques.

L'écoulement qui dépend des tailles caractéristiques du système et de la viscosité du fluide, peut être laminaire ou turbulent. Dans un système de SPR avec des micro-canaux, un écoulement laminaire (petit nombre de Reynolds) est attendu avec des lignes de courant bien dessinées et une vitesse qui dépend de la distance aux parois. Dans la configuration SPRi que

nous avons utilisée, les flux ne sont pas homogènes et les volumes relativement élevés, ce qui impose de circulariser l'échantillon pour limiter les quantités nécessaires. Des travaux présentent une méthode pour canaliser un fluide entre deux fluides « guides » de flux laminaire dans une cellule [415, 416], quand d'autres dispositifs ont été développés pour homogénéiser le contenu des microcellules, par exemple dans le cas de l'hybridation des puces à ADN [417]. Un compromis doit être trouvé entre homogénéité fluidique et volume de la cellule.

Nous avons travaillé en collaboration avec l'équipe de C. Khan Malek sur le développement de systèmes microstructurés incorporant des canaux microfluidiques (Figure 73) comme solution alternative. La conception en chapelet des détecteurs reliés par des microcanaux permet, dans ce cas, d'augmenter la dynamique de détection par capture successive des analytes dans les différents micropuits. Ce type de dispositif est multiplexable grâce à l'imagerie SPR et permet de réduire le volume d'échantillon.

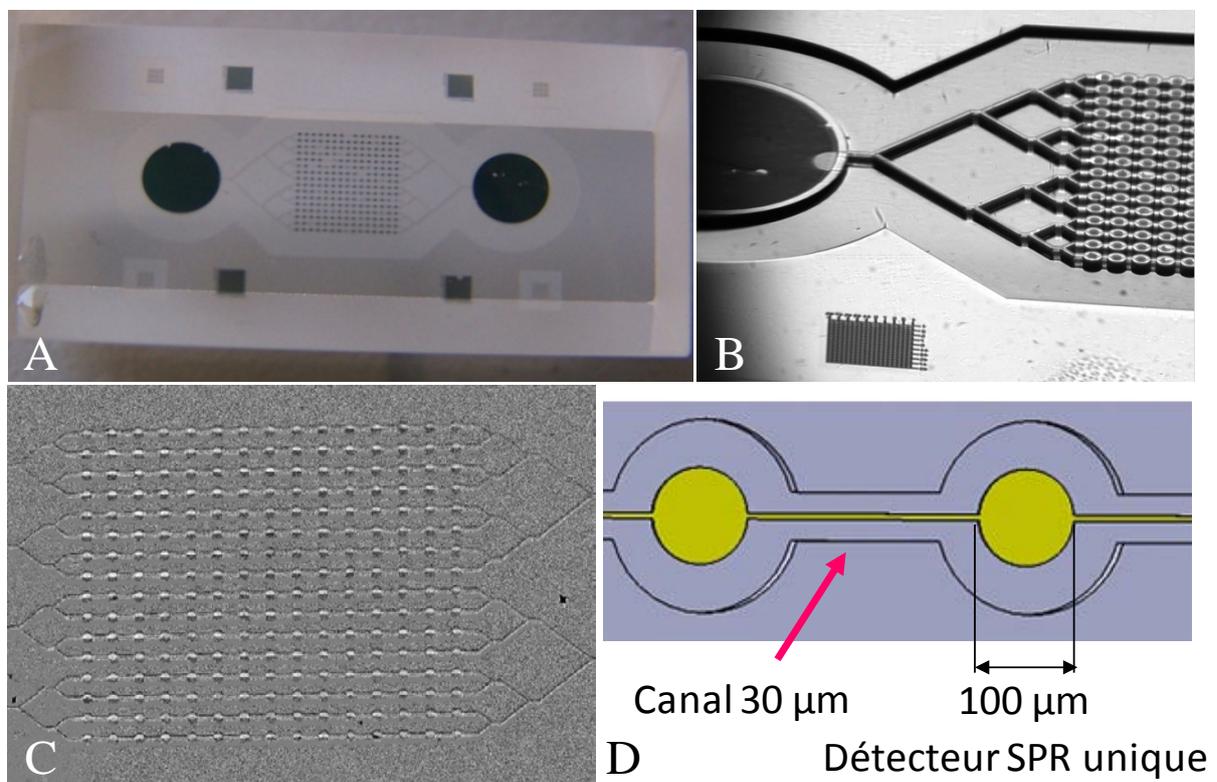


Figure 73 : Prismes microstructurés.

A, vue globale du prisme où l'on discerne les pistes d'or horizontales qui relient les quinze plots d'or. Les lignes convergent vers un cercle d'or de part et d'autre des motifs. B, grossissement du prisme. C, image plasmonique de la puce en tampon PBS. D, schéma des plots d'or reliés par le canal et leurs dimensions.

Des expériences préliminaires ont été réalisées sur ces dispositifs et néanmoins mis en évidence la difficulté à gérer les problèmes de capillarité dans ces systèmes et à éviter la formation de microbulles pouvant s'opposer au flux. L'utilisation de détergents permet partiellement de résoudre ce problème.

Limites de la modélisation

Nous avons développé un modèle (voir article 4) qui considère la cellule SPRi comme un système biphasique avec une phase aqueuse mobile qui couvre une phase poreuse, statique : la matrice de dextran. La diffusion des objets est considérée être contrôlée par des coefficients de diffusion différents dans la phase dextran et la phase aqueuse.

- Notre modèle prend en compte le dextran comme une couche homogène, ce qui n'est probablement pas vrai compte tenu que la chimie de couplage immobilise les chaînes au voisinage de la surface.
- La vitesse du flux est considérée comme nulle au sein de la matrice, ce qui représente une approximation.
- Nous considérons une fluidique laminaire homogène et avec des flux parallèles dans la phase aqueuse avec un nombre de Reynolds petit, or la géométrie de la cellule et les irrégularités de surface peuvent perturber ce régime (Figure 74).

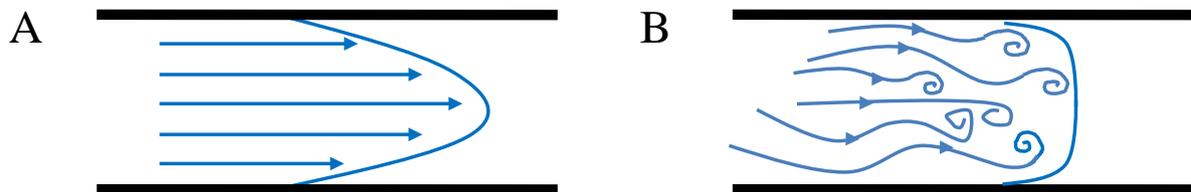


Figure 74 : schéma de fluidique de la cellule SPRi.
A, régime laminaire avec des flux parallèles, B, régime turbulent.

Malgré ces simplifications, nous obtenons des simulations de qualité avec un bon ajustement des courbes d'hybridation dans leur phase rapide. Le modèle permet de définir les limites expérimentales dans lesquelles on peut travailler sans avoir recours à des corrections. Ceci est le cas des interactions moléculaires dont les constantes de vitesse de second ordre sont inférieures à $0.5 \mu\text{M}^{-1}.\text{min}^{-1}$. Les corrections sont donc nécessaires dans nos conditions expérimentales où ces constantes atteignent des valeurs de 4 à $8 \mu\text{M}^{-1}.\text{min}^{-1}$. Le modèle permet de définir les corrections à réaliser, ainsi que les paramètres sur lesquels nous pouvons agir.

4. Conclusion

Les systèmes SPRi actuels travaillent avec des flux relativement lents qui n'empêchent pas la formation de barrière de diffusion, rendant l'approche peu adaptée à l'étude des cinétiques ayant un demi-temps caractéristique inférieur à la minute. En ce qui concerne les réactions plus rapides, les systèmes sont exploitables à condition d'apporter des corrections aux données observées. Une voie qui paraît prometteuse pour élargir les applications de la SPRi est le développement de puces microstructurées qui permettent d'améliorer le contrôle de la fluidique. La nanostructuration du support est une approche pour augmenter la sensibilité.

4^{ème} partie

MOTIF-I, UNE STRUCTURE PARTICULIERE DE L'ADN CONTROLABLE PAR LE PH

Le travail présenté dans les chapitres précédents reposait sur l'auto-assemblage de différents systèmes en utilisant les propriétés organisationnelles de l'ADN B. L'objectif de réaliser des nanostructures fonctionnelles nous a conduits à choisir une stratégie d'assemblage compatible avec la fragilité de composants biologiques tels que les protéines. L'assemblage dans des conditions irréversibles, loin de l'équilibre thermodynamique, a permis de réaliser des structures définies à base d'ADN et de PDNA sans altérer la fonction des protéines. Toutefois, lors de l'assemblage de plus longs polymères ou de structures complexes, des produits intermédiaires, non désirés, ont été générés.

Nous avons vu dans la partie introductive que la conception des assemblages sous contrôle cinétique implique des modèles complexes difficiles à généraliser. Finalement, c'est la mise en œuvre de systèmes réversibles qui permet le mieux de s'affranchir des erreurs d'assemblage. La réversibilité est atteinte dans le cas des ADN B par un cycle thermique (fusion à $\sim 90^\circ\text{C}$ puis lent refroidissement au cours duquel les structures se réorganisent). Toutefois, ces conditions sont rarement compatibles avec la stabilité des protéines, aussi il est nécessaire de trouver d'autres systèmes qui permettent de contrôler l'assemblage dans des conditions plus douces. L'utilisation de séquences riches en cytidines, dont la protonation autorise l'assemblage, a été retenue.

Génération et structure de motif-i

A pH légèrement acide, les paires $C \cdot C^+$ formées par les cytidines neutre et protonée (Figure 75A) induisent l'association des oligonucléotides riches en C en des duplex parallèles (partie B, gauche). Ces duplex hémi-protonés sont de courtes durée de vie, mais leur intercalation mutuelle, en orientation tête-bêche, conduisent à la formation de tétramères (B, droite et C) extrêmement stables, appelés motif-i [147]. La structure en motif-i présente l'originalité d'impliquer l'intercalation systématique des bases appariées. La remarquable symétrie du tétramère provient de l'association des quatre segments identiques dans la structure.

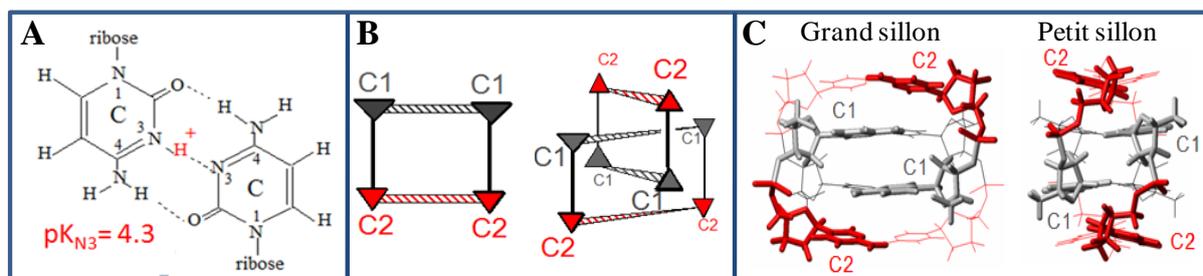


Figure 75 : structure particulière de l'ADN : le motif-i.

A, paire hémiprotonnée $C \cdot C^+$ formée par des cytidines neutre et protonée. B, gauche, représentation schématique d'un duplex parallèle hémiprotonné de CC ; B, droite, tétramère de motif-i de CC. Le tétramère est stable tandis que le

duplex est une structure de courte durée de vie en échange rapide avec le monomère CC. C, structure RMN haute résolution du motif-i [CC]₄ [418]. La vue de gauche est perpendiculaire au grand sillon et la vue de droite est perpendiculaire au petit sillon. Les résidus de l'arrière plan sont dessinés en trait fin.

Le mécanisme de formation des motifs-i a été analysé et, selon les séquences d'oligonucléotide utilisées, il peut impliquer (Figure 76) :

- quatre oligonucléotides identiques riches en C [419]. C'est le cas du tétramère de motif-i.
- deux structures en épingles à cheveux, chacune portant deux segments cytidine [420], dans le dimère de motif-i.
- un seul brin replié qui présente quatre segments de cytidine dans le monomère de motif-i [421].

La topologie de l'intercalation est sensible à la séquence des liaisons qui connectent les segments de cytidines.

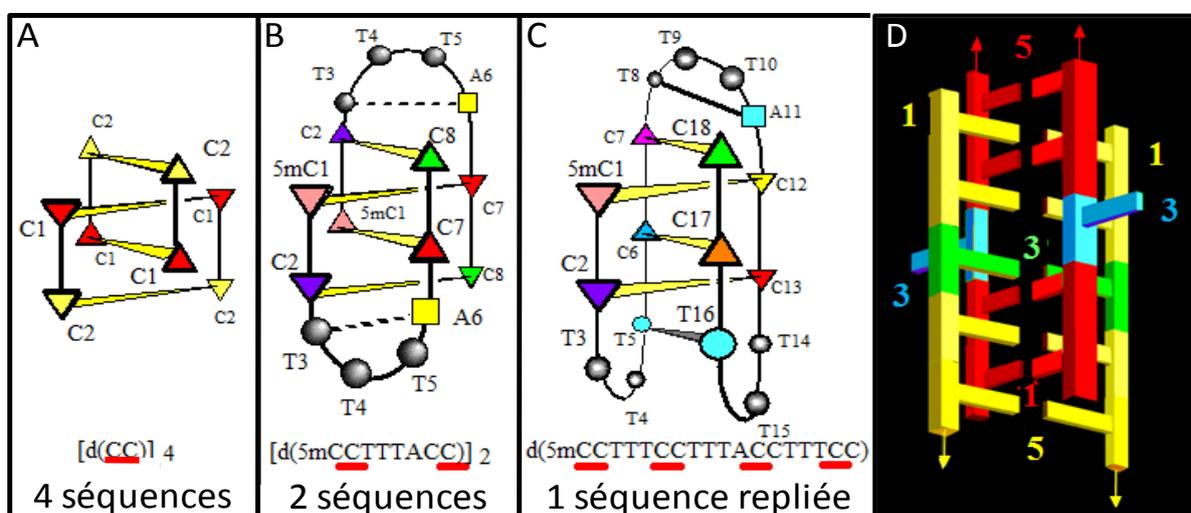


Figure 76 : Représentations schématiques de motifs-i.

A, motif-i impliquant 4 séquences (CC) [418], B, motif-i constitué de 2 séquences [420]. C, 1 seule séquence est repliée [422]. D, deuxième exemple de motif-i, un peu plus long, constitué des 4 séquences (CCTCC) [146].

Découverte et étude des propriétés d'auto-assemblages des motifs-i en nanofils.

L'équipe de Krishnan découvre, en 2007, la capacité des motifs-i à s'auto-associer pour former des nanofils [213]. A partir de motifs de C₇, les structures partiellement intercalées présentant des cytidines libres ont pu s'assembler en des structures plus stables et plus longues jusqu'à former des nano-fils pouvant atteindre des longueurs de quelques μm avec un diamètre de 2 nm [213]. Les chercheurs ont ainsi exploré le potentiel des motifs-i en tant que briques rigides pour la construction de structures à une dimension, à l'échelle nanométrique.

En 2009, l'étude cinétique du chemin de formation des tétramères de motifs-i conforte la possibilité d'assembler des motifs-i en nanofils-i [423]. L'étude a montré que plusieurs espèces, différant par la topologie d'intercalation des oligonucléotides TC_n, étaient formées à des vitesses comparables. Le principe d'élongation des motifs-i de C₇ en fils est présenté sur

la Figure 77. L'intercalation incomplète de deux duplex hémi-protonés (A) va former une structure tétramérique partiellement intercalée (B) qui peut servir de point d'ancrage pour un duplex. Le produit formé (D) est constitué de trois duplex, dont les extrémités libres peuvent, à nouveau, servir d'ancrage. Des additions ultérieures de duplex vont mener à la formation d'un fil. Les tétramères totalement intercalés (C), incapables de s'élonger, constituent des pièges cinétiques.

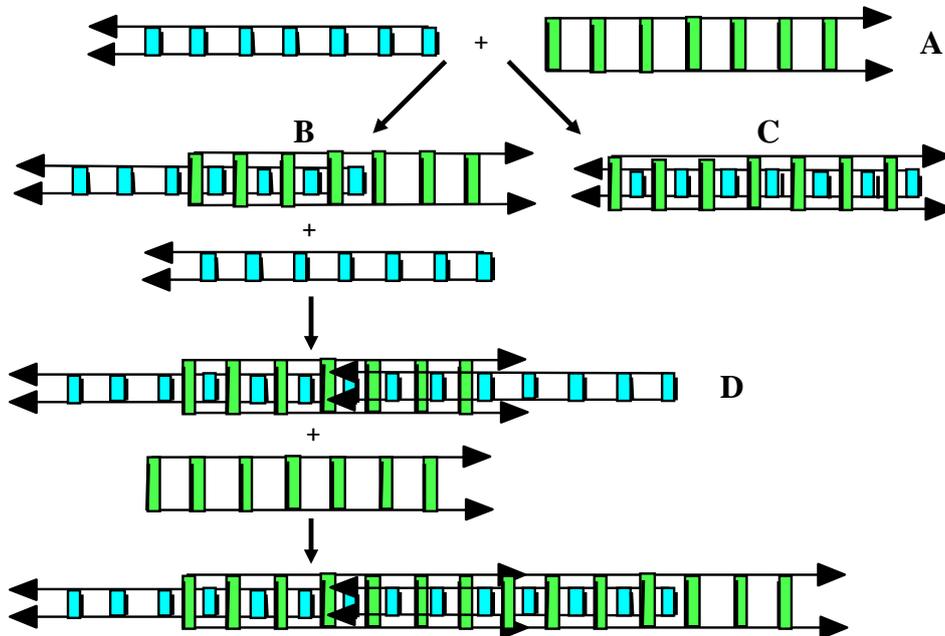


Figure 77 : Vue schématique du principe de formation des fils i à base de la séquence C_7 .

Les séquences C_7 sont capables de s'auto-assembler pour former des nano-fils, qu'en est-il des autres séquences ?

Cette question a été abordée en collaboration avec Jean-Louis Leroy, chercheur à l'ICSN (Institut de Chimie des Substances Naturelles) au CNRS de Gif-sur-Yvette. Une étude a conduit à mesurer les vitesses d'association et de dissociation des structures supramoléculaires de motifs-i. L'influence des paramètres physico-chimiques (température, pH, force ionique et concentration des oligonucléotides) a aussi été mesurée.

1. Présentation de l'article :

“ $[C_7GC_4]_4$ association into supra molecular i-motif structures”

Nous avons développé un modèle, à base de séquences C_mXC_n (m différent de n), plus adapté selon nous, à la formation de structures supramoléculaires que le modèle à base de séquences C_7 . Les oligonucléotides consistent en des segments contenant m et n cytidines, séparées par un résidu central X qui peut être une adénine, une thymidine, une guanine ou bien une autre molécule faisant un lien chimique. Les longueurs m et n des segments peuvent être identiques mais elles sont préférentiellement différentes.

Les deux hypothèses qui peuvent expliquer la lenteur de l'élongation des motifs-i en structures supramoléculaires sont :

- la formation de dimères de motif-i, qui sont des structures piège de longue durée de vie, ne pouvant s'élonger.
- la formation de tétramères de motif-i totalement intercalés, qui empêchent toute intercalation entre différents motifs.

L'idée du modèle proposé C_mXC_n est de générer des blocs de motif-i « centraux » avec des extrémités pendantes qui permettraient aux motifs de s'associer entre eux, selon le mécanisme supposé, présenté sur la Figure 78. La première étape est la formation du tétramère Te (voir schéma). Le résidu central « non-cytidine », qui ne participe pas à la formation du motif-i, devrait empêcher l'intercalation totale des segments en un unique motif-i. Le résidu central favoriserait donc la formation de tétramères symétriques, partiellement intercalés, avec les blocs de motif-i réalisés à partir des segments de régions en C les plus longues se terminant par des extrémités libres des segments les plus courts (Te). Après l'étape de nucléation, les associations successives par appariement et intercalation des extrémités des segments C libres de 2 blocs Te conduisent à la formation du bloc Te_2 , la plus courte des structures supramoléculaires. La symétrie du motif-i génère, au cours de l'assemblage de plusieurs blocs, les mêmes terminaisons de segments de cytidines non-appariées propices à l'élongation. Selon ce schéma, l'élongation des structures supramoléculaires génère des assemblages linéaires de type nano-fils ou bâtonnets plutôt rigides dont la microscopie de force atomique devrait nous permettre d'observer la distribution en taille.

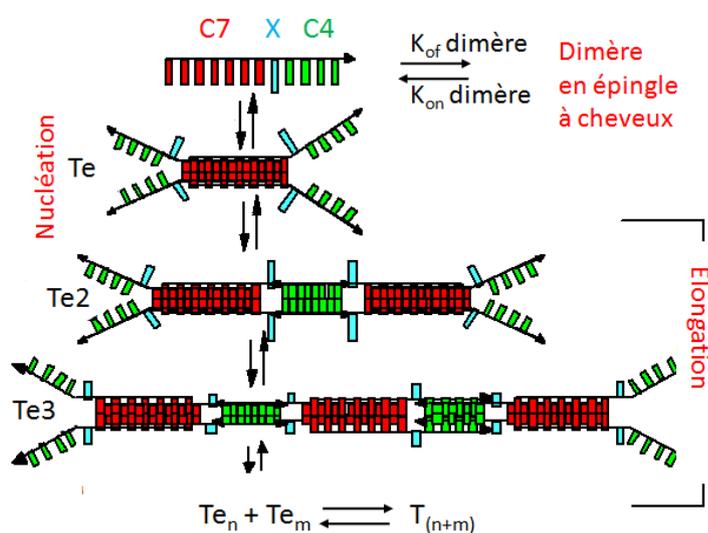


Figure 78 : Mécanisme postulé de l'assemblage de motifs-i C_7XC_4 en structures supramoléculaires.

Le tétramère Te , intercalé sur les 7 bases du segment long, qui présente le cœur de motif-i le plus stable parmi toutes les topologies d'intercalations possibles, est en équilibre avec toutes les formes de tétramères.

Les oligonucléotides $5'C_mXC_n3'$, présentés dans le Tableau 2, ont fait l'objet d'une étude préliminaire afin de déceler les structures favorables à la formation de fil et les problèmes empêchant l'élongation. L'analyse par chromatographie d'exclusion a montré que toutes les

séquences (hormis C₈TC₅) sont capables de s'auto-assembler en structures supramoléculaires. Nous avons défini arbitrairement comme structure supramoléculaire toute structure dont la taille, en chromatographie d'exclusion, est supérieure à la taille du tétramère de motif-i.

Tableau 2 : séquences de nucléotides synthétisés.

Séquences	C	m	X	C	n
	C	7	A	C	4
	C	7	T	C	4
	C	7	G	C	4
	C	7	--	C	4
	C	4	T	C	7
	C	4	G	C	7
	C	7	T	C	3
	C	6	T	C	3
	C	5	T	C	5
	C	7			
	C	8	T	C	5

Toutes les séquences, sans exception, forment des dimères de motif-i plus rapidement que le tétramère. Or les dimères participent à un piégeage cinétique. Ces dimères, d'autant plus stables que les oligonucléotides sont longs, vont avoir une influence majeure sur la génération des fils et leur vitesse de formation. Si on considère C₈TC₅, les dimères sont tellement stables qu'ils vont empêcher la formation de fils. Pour les autres oligonucléotides, les dimères les plus facilement dissociés vont accélérer la formation des fils. Ainsi, il serait préférable de choisir des oligonucléotides plus courts de type C₆TC₃ afin de minimiser le piège cinétique dû aux dimères. Toutefois, les structures supramoléculaires à base de segments courts sont moins stables que celles à base de segments longs. Aussi notre choix s'est porté sur les séquences de type C₇XC₄. Les séquences C₇GC₄ et C₇TC₄ ont été plus extensivement étudiées. L'article présente l'analyse de C₇GC₄ et les données relatives à la séquence C₇TC₄ sont présentées dans la partie « données complémentaires ».

Trois techniques de caractérisation ont été utilisées :

- la chromatographie d'exclusion en HPLC,
- la résonance magnétique nucléaire (RMN),
- la microscopie de force atomique (AFM).

Le tamisage moléculaire, qui donne accès à la distribution de taille des molécules, est la technique principale sur laquelle a reposé l'approche physico-chimique. La chromatographie d'exclusion a permis de mesurer les cinétiques de formation et de dissociation des motifs-i et les différents paramètres de température, pH, concentration et force ionique. Deux colonnes couvrant des gammes de tailles complémentaires ont été calibrées à partir d'ADN plasmidique double brin, d'ARNt, de structures de motifs-i connues

et de thymidine. La relative linéarité des courbes de calibration permet d'analyser les distributions de taille obtenues, une différence d'organisation tridimensionnelle des structures supramoléculaires pouvant toutefois apporter un biais d'interprétation.

La RMN en phase liquide permet d'accéder à la structure des motifs-i et à leur dynamique. C'est un très bon outil de caractérisation pour les motifs-i de séquences courtes. La RMN corrobore la formation de structures supramoléculaires, toutefois son utilisation est limitée par la taille des objets ainsi que leur homogénéité. Les spectres et signaux RMN sont peu résolutifs pour des objets de grande taille et/ou des populations d'objets présentant plusieurs conformations.

L'AFM à l'air apporte des informations structurales complémentaires de la RMN puisque des objets de grande taille peuvent être observés. Toutefois les conditions à l'air peuvent aboutir à des observations ne correspondant pas à la réalité des structures en solution. L'AFM présente les contraintes d'immobilisation des objets sur une surface ainsi que de diminution de la force ionique. Deux phénomènes peuvent avoir lieu lors du dépôt des échantillons sur le substrat :

- soit les structures formées en solution sont capturées sur la surface de mica,
- soit des structures peuvent se former principalement au contact de la surface et ne représentent pas l'état en solution.

Les résultats obtenus en imagerie AFM sont présentés dans la partie « données complémentaires » suivant l'article.

L'article réalisé à partir des résultats obtenus est présenté dans l'appendice 5. Le travail démontrant l'influence de la base centrale sur les assemblages a fait l'objet d'un dépôt de brevet étendu à l'international, les structures supramoléculaires de motif-i pouvant avoir des applications spécifiques en nanotechnologie et nano-médecine.

2. Appendice 5 : article 5

[C₇GC₄] association into supramolecular i-motif structures

Laisné A., Pompon D. & Leroy J-L.

Nucleic Acids Research (2010), 38(11) : 3817-3826

[C₇GC₄]₄ Association into supra molecular i-motif structures

Aude Laisné¹, Denis Pompon¹ and Jean-Louis Leroy^{2,*}

¹Laboratoire d'Ingénierie des Proteines Membranaires, Centre de Génétique Moléculaire, CNRS FRE3144, Unité associé à l'Université Pierre et Marie Curie and ²Laboratoire de Chimie et Biologie Structurale, Institut de Chimie des Substances Naturelles, Gif-sur-Yvette, 91190, France

Received December 28, 2009; Revised January 29, 2010; Accepted February 3, 2010

ABSTRACT

The self-associative properties of cytidine-rich oligonucleotides into symmetrical i-motif tetramers give to these oligonucleotides the capacity of forming supramolecular structures (sms) that have potential applications in the nanotechnology domain. In order to facilitate sms formation, oligonucleotides containing two cytidine stretches of unequal length (C_nXC_m) separated by a non-cytidine spacer were synthesized. They were designed to associate into a tetramer including an i-motif core built by intercalation of the C•C⁺ pairs of the longer C stretch with the two dangling non-intercalated strands of the shorter C stretch at each end. Gel filtration chromatography shows that the non-intercalated C-rich ends give to this structure the capacity of forming extremely stable sms. Using C₇GC₄ as a model, we find that the sms formation rate varies as the oligonucleotide concentration and increases at high temperature. Competitively with the tetramer involved in sms elongation, C_nXC_m oligonucleotides form i-motif dimers that compete with sms elongation. The dimer stability is strongly reduced when the pH is moved away from the cytidine pK. This results in an equilibrium shift towards the tetramer and in the acceleration of the sms formation rate. The chromatograms of the sms formed by C₇GC₄ indicate a broad distribution. In a 1.5 mM solution incubated at 37°C, the equilibrium distribution is centered on a molecular weight corresponding to the assembly of nine tetramers and the upper limit corresponds to 80 tetramers. The lifetime of this structure is about 4 days at 40°C, pH 4.6.

INTRODUCTION

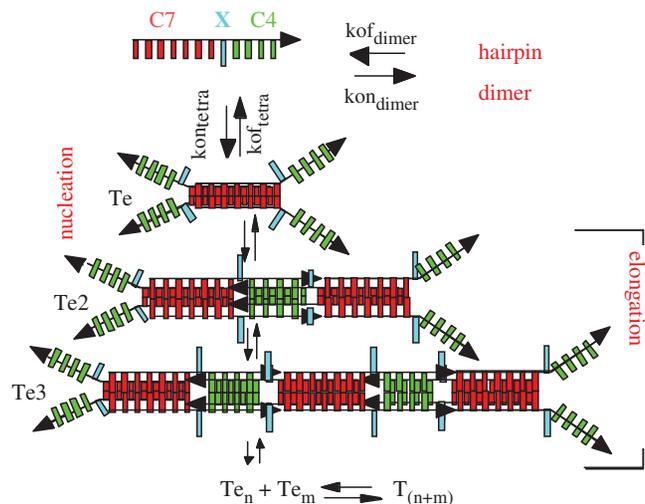
At slightly acid pH, the hemiprotonated C•C⁺ pairs formed by neutral and protonated cytidines induce the association of cytidine-rich oligonucleotides into parallel duplexes. These duplexes are short-lived elusive species, but the tetramers built by head to tail intercalation of two duplexes are extremely stable. This structural arrangement, called i-motif is exceptional in that it involves systematic base-pair intercalation (1). The i-motif structure may include four identical C-rich strands, two hairpins each caring two cytidine stretches or a folded strand caring four cytidine stretches (2,3). The remarkable symmetry of i-motif tetramers is a consequence of the equivalence of the four strands associated in the structure. Face to face intercalation of thymidine in i-motif structures is generally unfavorable (4,5) and purine intercalation seems to be sterically hindered (6,7). A recent kinetics investigation shows that i-motif tetramers with partial or full intercalation topology are formed at comparable rates and that the predominance at equilibrium of the fully intercalated species is due to its longer lifetime (8).

The self-associative properties of C-rich stretches give to i-motif tetramers the potentiality of forming supramolecular structures (sms) by mutual intercalation of the non-paired cytidines of tetramers built by partially intercalated duplexes. This property has been reported in a recent study revealing the association of C7 oligonucleotides into linear repetitive structures that could have biotechnological applications (9).

In order to understand and control the formation process of i-motif sms, we examine in the present paper how temperature, strand concentration, ionic strength and pH influence the sms formation and dissociation rates.

Oligonucleotides of the C_nXC_m series were designed to improve the assembly of i-motif tetramers into sms. They are composed of two cytidine stretches of unequal length separated by a non-cytidine residue i.e. by A, T, G or

*To whom correspondence should be addressed. Tel: +33 1 69 82 3630; Fax: +33 1 69 82 3784; Email: jean-louis.leroy@icsn.cnrs-gif.fr



Scheme 1. Postulated association pathway of C_7GC_4 into i-motif sms. The monomer is in equilibrium with one or several hairpin dimer(s) (cf. Scheme 2) and Te, the tetramer formed by full intercalation of the C_7 stretches. Association and mutual intercalation of the C_4 stretches of two Te building blocks result in the formation of the Te_2 species. The i-motif symmetry gives to the assembly of several building blocks the same elongation capacity than that of the building blocks themselves and allows the association of preformed Te_n and Te_m species into structures including $(n + m)$ building blocks.

a diol-spacer. They were synthesized with the longer C stretch either at the 5'- or the 3'-end. The non-cytidine spacer that cannot intercalate into the i-motif was intended to prevent intercalation of both cytidine stretches into a single i-motif core. It was expected that the spacer would enforce the formation of a tetramer (Te in Scheme 1) including the i-motif core built by the $C \cdot C^+$ pairs of the longer cytidine stretch and at each end the non-intercalated cytidines of the shorter stretch.

Gel filtration chromatography shows that these oligonucleotides associate into sms. The stability, the formation and dissociation rates of the sms formed by oligonucleotides differing by the C stretch length and the nature of the spacer were qualitatively compared. The relatively fast formation rate and the good stability of the sms formed by C_7GC_4 prompted us to select this sequence as a benchmark model to investigate the effect of pH, strand concentration, temperature and ionic strength on the sms formation and dissociation kinetics. The influence of the C stretch length and of the nature of non-cytidine spacer on the sms stability, formation rate and structure will be examined in separate publications.

MATERIALS AND METHODS

Oligonucleotide synthesis and sample preparation

The oligonucleotides were synthesized on a 2 or 10 μM scale, purified by chromatography on an anionic DEAE column according to procedures already described and extensively dialyzed (5). After dialysis, the oligonucleotide solutions were lyophilized and dissolved in water to make 0.5–1 mM stock solutions. The C_7GC_4 concentration

was determined using the A^{260} values, $89600 \text{ M}^{-1} \text{ cm}^{-1}$, computed according to a nearest neighbor model (10).

Formation and dissociation rate measurements of i-motif sms

Except otherwise stated, all the measurements were performed in 0.4 M NaCl, 10 mM Na acetate and 10 mM Na phosphate, a solution designed as the NAP buffer. During sms formation and dissociation, the samples, typically 50–200 μl , were incubated in a mastercycler Eppendorff[®] PCR incubator whose lid temperature was set 5°C above that of the samples in order to avoid condensation on the cap of the Eppendorff tubes.

To study the sms formation kinetics, the samples were initially melted at 100°C and rapidly cooled at the temperature of the incubator. For sms dissociation time measurements, a solution containing only sms was prepared by pooling the species eluted on a GPC 100 column with a molecular weight larger than that of a tetramer. Afterward, this solution was diluted in the NAP buffer in such way as to allow complete dissociation of the sms at equilibrium. The measurements started right after dilution.

Gel filtration chromatography

During sms formation and dissociation, the sample composition was measured at room temperature by gel filtration chromatography (11) on GPC 100 and GPC 1000 columns ($250 \times 4.6 \text{ mm I.D.}$) provided by Eprogen[®].

Aliquots took off the incubated samples were injected in the column immediately after dilution to 25 μl (the volume of the injection loop) in the NAP buffer. They were eluted with the same buffer at a flow rate of 0.4 ml min^{-1} . Systematic addition of thymidine at a final concentration of about 5 μM in the injected sample provided a reference marker on the chromatograms. The elution times ranged from 4 to 8.5 min. (Supplementary Figure S1). The elution buffer pH was generally 4.6. However, when the evolution of the sample composition during elution on the column at pH 4.6 was not negligible, the elution buffer pH was adjusted to the value of the sample pH. The GPC-100 column whose exclusion limit correspond, according to the manufacturer to an oligonucleotide containing about 300 nt was calibrated using tRNA, thymidine, C-rich non-structured oligonucleotides, C-rich i-motif monomers and tetramers. The GPC-1000 column was calibrated with double-stranded plasmids containing from 800 to 16000 nucleosides, the i-motif tetramer of $T_2C_8T_2$, tRNA and thymidine. The GPC-1000 calibration shows that oligonucleotides containing 50 and 10^4 residues are eluted, respectively, in the permeation and exclusion limits. The calibration curves displayed in Supplementary Figure S1 establishes the reasonable linearity of the measured elution times with the log of the molecular weight of reference oligonucleotides. The dispersion of the data points of the calibration curves suggests that the relative error on the estimated molecular weights is <0.5 . The sizes of the sms were characterized by the mass center, M_c , of the sms distribution. M_c was computed according to $M_c = \frac{\sum A(t) \cdot N(t)}{\sum A(t)}$,

where $A(t)$ is the absorbance of the solution and $N(t)$ the number of tetrameric units in the sms eluted on the chromatogram at time t . $N(t)$ was obtained from the calibration curve of the GPC 1000 column (Supplementary Figure S1). The dimer and tetramer stoichiometries were also routinely measured by the slope of the log–log plot of the multimer versus monomer equilibrium concentrations. In pH and temperature conditions favorable to sms formation, an opalescence characteristic of the formation of extremely large structures was observed in concentrated samples ($>1\text{ mM}$) incubated during times much longer than the sms formation time constant. Correlatively with the development of the sample opalescence, the chromatogram of these samples showed a reduction (up to 50%) of the integrated area of the eluted species indicating that a fraction of the sample composed, presumably of the largest sms, was trapped in the column. In that case, the oligonucleotide fraction retained on the column was estimated by comparison with the chromatogram of an identical melted sample and included in the sms fraction. However, in order to have a direct control of all the species formed, we generally restrained the oligonucleotide concentrations and incubation times to a range of values allowing detection of all the species on the chromatograms.

NMR method

The NMR experiments were performed using a 500 MHz Varian Inova spectrometer with the jump and return sequence for water suppression (12). The spectra were accumulated with a repetition time longer than the relaxation times to avoid saturation. The spectral intensity was multiplied by a $1/\sin$ function in order to correct the distortion introduced by the jump and return excitation (13). The spectra collected during the sms formation were normalized using the methyl proton peak of the acetate buffer.

Multimer dissociation constants

The dissociation constant of a multimer may be expressed as a function of α_{eq} , the monomer equilibrium fraction of an oligonucleotide solution at concentration $[M_0]$ by: $K_{\text{dis}} = S\alpha_{\text{eq}}^s [M_0]^{s-1} / (1 - \alpha_{\text{eq}})$ where s is the multimer stoichiometry. We will characterize the multimers stability by F_i , a parameter whose dimension is independent of the stoichiometry which is equal to the free monomer concentration for which $\alpha_{\text{eq}} = 0.5$. F_i is related to the multimer dissociation constant by: $F_i = (K_{\text{dis}}/s)^{1/(s-1)}$ and it will be designated as the reduced dissociation constant.

RESULTS

Chromatographic evidences for the formation of i-motif supramolecular structures

Immediately after melting and fast cooling, a $3\ \mu\text{M}$ C_7GC_4 sample at pH 6.2 injected into a GPC 100 column is mainly eluted as a monomer (Figure 1). The chromatograms recorded at different times during incubation at

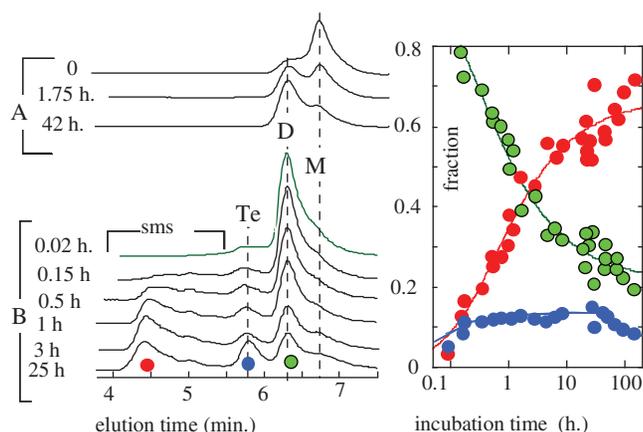


Figure 1. Association of C_7GC_4 into i-motif multimers and supra-molecular structures (sms) in the NAP buffer. The samples were initially melted. *Left panel:* GPC-100 chromatograms recorded after incubation during the times indicated. The elution times expected for a monomer (M), a dimer (D) and a tetramer (Te) are indicated. (A) the chromatogram of a $3\ \mu\text{M}$ solution, pH 6.2 injected right after melting shows a 20/80 dimer/monomer mixture. The chromatograms recorded after incubation at 20°C show that the dimer is formed with a time constant of about 1.5 h. The monomer and dimer concentrations at equilibrium correspond to the reduced dissociation constant $F_{i\text{dimer}} = 2\ \mu\text{M}$. (B) The chromatograms of a $0.15\ \text{mM}$ C_7GC_4 solution at pH 4.63 collected after melting and incubation at 50°C during the times indicated on the figure show that the dimer is slowly converted into tetramer and supra-molecular structures. *Right panel:* evolution of the sms (red), tetramer (blue) and dimer (green) fractions as a function of the incubation time at 50°C .

20°C reveal that a dimer is formed with a time constant of about 1.5 h. At equilibrium, the dimer and monomer concentrations correspond to a reduced dissociation constant $F_i = 2\ \mu\text{M}$. It has been shown that the formation rate of i-motif dimers increases with the nucleotide concentration and when the solution pH is shifted near 4.4, the value of the cytidine pK_{N3} (8). This explains that about 1 min after melting, an aliquot of a $0.15\ \text{mM}$ C_7GC_4 solution, pH 4.6 is eluted as a dimer (Figure 1). The chromatograms of samples incubated at 50°C show the appearance of two peaks, one corresponding to a tetramer and the other to unresolved heavier species that will be designed as sms.

The evolution versus time of the dimer, tetramer and sms fractions is displayed in Figure 1. After 0.5 h, the tetramer fraction reaches a maximum value corresponding to about 12% of the oligonucleotide concentration. The sms increase with a time constant of 1 h toward an equilibrium level corresponding to 65% of the oligonucleotide concentration. The dimer fraction decreases with a comparable time constant. The sms molecular weights were estimated using a GPC-1000 column. The chromatograms of aliquots taken from a $1.5\ \text{mM}$ C_7GC_4 solution incubated at 37°C , pH 4.6 are displayed in Figure 2A. The dimer and the tetramer are eluted together in the permeation volume of the GPC-1000 column. The sms half formation time is 1 h and the oligonucleotide fraction associated in sms at equilibrium is about 75% (Figure 2D). The average size of the sms increases as a function of the incubation time (Figure 2B). After

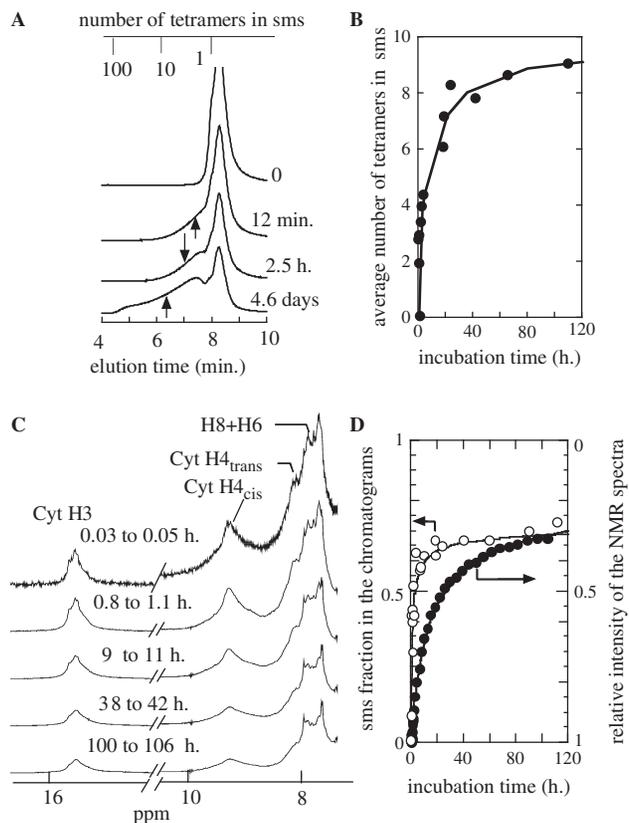
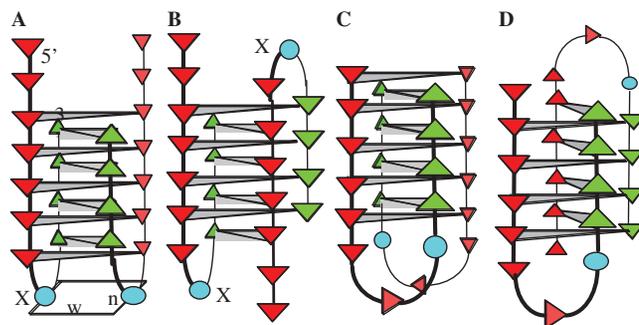


Figure 2. Association of C_7GC_4 into sms as monitored by NMR and gel exclusion chromatography on GPC 1000 column. The oligonucleotide concentration was 1.5 mM. The NMR spectra and the chromatograms were recorded after melting and incubation in the NAP buffer, pH = 4.6 at 37°C during the times indicated. (A) The tetramer and the dimer are eluted together in the GPC 1000 chromatograms. The chromatograms collected as a function of the time show the formation of sms. The upper scale shows the elution times expected for sms including the indicated number of tetrameric repeats. Note that one tetramer contains 48 nucleotides. On each chromatogram, an arrow indicates the mass center of the sms distribution. (B) The number of tetramer repeats at the center of the sms distributions is plotted as a function of the incubation time. At equilibrium, the mass center of the sms distribution corresponds to the association of nine tetramers (i.e. 432 bases). The top 15% of the sms distribution includes structures that are five times larger. (C) exchangeable and aromatic proton regions of NMR spectra collected during sms formation. The line broadening (about 40 Hz) and the reduction of the signal intensity observed as a function of the time reflect the formation of large structures. (D) oligonucleotide fractions included in sms as determined from the chromatograms of panel A (open circles) and from the intensity of the NMR peaks of the spectra of panel C (black circles). The time constant derived from the chromatography measurements, 1 hour, concerns the formation of structures heavier than a tetramer, whereas that derived from the NMR experiments, 20 h, is related to structures too large to be detected on the NMR spectra.

incubation over 4 days at 37°C, the sms exhibit an extremely broad distribution centered around 430 residues, a size corresponding to the association of nine tetramers. The elution time of the largest species, 4.5 min, is that expected for a structure including about 4000 residues, i.e. 80 tetrameric units (Figure 2A).

sms formation monitored by NMR

Right after melting and fast cooling at 0°C, the NMR spectrum of C_7GC_4 shows three broad clusters of



Scheme 2. Plausible structures of the C_7XC_4 dimer ($X = A, T$ or G). The number of intercalated $C\bullet C^+$ pairs in each structure is optimal. The narrow (n) and wide (w) grooves are indicated in structure (A). One residue is enough to make a loop across the i-motif narrow groove (6) and three residues allow loop formation across the wide groove (5). For each loop topology, the hairpins may be in parallel (A, C) or in head to tail orientation (B, D). Many other topologies can be imagined by changing the intercalation order of the $C\bullet C^+$ pairs and the number of residues in the loops. The poor resolution of the NMR spectrum of C_7GC_4 suggests that several dimers coexist.

exchangeable protons at the positions characteristic of the imino and amino protons of $C\bullet C^+$ pairs (14). The poor spectral resolution suggests the presence of multiple conformations (Supplementary Figure S2). The 10.3 ppm chemical shift and the fast exchange rate, about 40 ms, of the G imino proton establish that the guanine is not H-bonded to the cytidine N3. The number of cytidine imino protons around 15.5 ppm estimated by reference to the guanosine imino proton peak is consistent with the number of $C\bullet C^+$ pairs (seven to nine $C\bullet C^+$ pairs for two guanosine imino protons) expected for the i-motif dimers displayed in Scheme 2.

Sms formation was followed by NMR in the same experimental conditions ($[C_7GC_4] = 1.5$ mM, pH 4.6, $T = 37^\circ\text{C}$) than the gel filtration experiments described just above. At 37°C, the guanosine imino proton peak is broadened out by exchange with water (Figure 2C). The NMR spectra recorded as a function of the time after melting show a broadening of about 40 Hz and a reduction of the intensity of all the NMR peaks indicating the formation of large structures with a slow tumbling rate. The initial intensity is fully restored after heating at 100°C. At equilibrium, the reduction of the spectral intensity (~75%) is comparable to the oligonucleotide fraction found associated in sms by chromatography, but the intensity of the NMR peaks decreases with a time constant 20 times longer than that measured by chromatography (Figure 2D). The discrepancy of the time constants measured by chromatography and NMR is merely related to the difference of the species that are taken into account by each method. The time constants derived from the chromatographic investigation are related, according to the definition given to sms, to the formation of structures whose molecular weights are larger than that of a tetramer whereas the time constants derived from the NMR experiments correspond to high molecular weight structures whose proton spectra are broadened out.

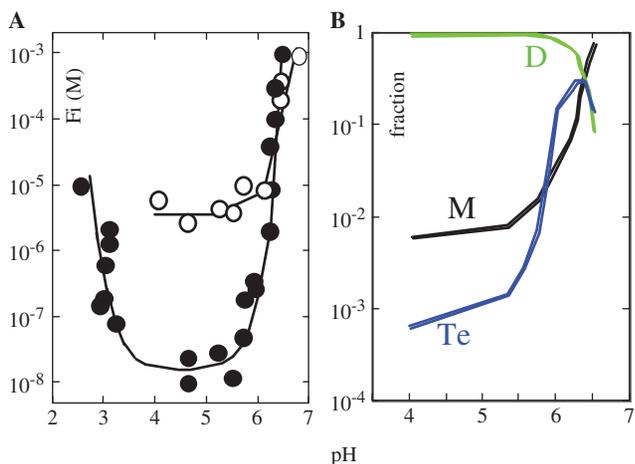


Figure 3. Effect of pH on dimer and tetramer stability at 20°C. (A) Reduced dissociation constants versus pH of the dimer(s) of C_7GC_4 (black circles) and of $[C_7T]_4$ (open circles). C_7T that do not associate into dimer or sms was used as a model to evaluate the stability of the tetramer of C_7GC_4 (Te in Scheme 1), the structure assumed to be the building block of the sms elongation pathway. (B) Computed monomer (black), dimer (green) and tetramer (blue) fractions in a 0.3 mM C_7GC_4 solution assuming that the tetramer of $[C_7GC_4]_4$ has the same reduced dissociation constant than $[C_7T]_4$. Due to the greater sensitivity of the dimer stability to pH, the tetramer fraction increases by two magnitude orders between pH 5.5 and pH 6.

Competition between dimer and tetramer formation

The concomitance of the evolution of the dimer and sms fractions (Figure 1) suggests that sms formation is controlled by dimer dissociation. In order to understand how the dimer and the tetramer of C_7GC_4 interfere with sms formation, we measured their reduced dissociation constant at different pH. The chromatograms of C_7GC_4 solutions recorded as a function of the time after melting and incubation at 20°C show the following systematic features: the first multimer formed is always a dimer. When the oligonucleotide concentration is sufficiently high, one observes that a tetramer is formed at a slower rate. Tetramer formation is always followed by the appearance of sms. The dimer reduced dissociation constant, $F_{i,dimer}$, measured from the dimer and monomer equilibrium concentrations is plotted vs. pH in Figure 3. $F_{i,dimer}$ is close to $2 \cdot 10^{-8}$ M between pH 3.5 and 5.5 and increases sharply when the solution pH is moved away from this range of values. The C_7GC_4 monomer concentration, which is controlled by the monomer–dimer equilibrium, is extremely small in the range of oligonucleotide concentrations allowing tetramer formation and for this reason the tetramer–reduced dissociation constant is experimentally inaccessible. Assuming that the tetramer of C_7GC_4 is formed by pairing and intercalation of the seven consecutive cytidines, we considered that its reduced dissociation constant should be comparable to that of any i-motif tetramer formed by intercalation of C_7 stretches and we used C_7T as a model to evaluate its reduced dissociation constant. The chromatograms of C_7T show two peaks at the positions expected for a monomer and a tetramer. The reduced dissociation constant, $F_{i,tetra}$, of $[C_7T]_4$ derived from the tetramer and monomer

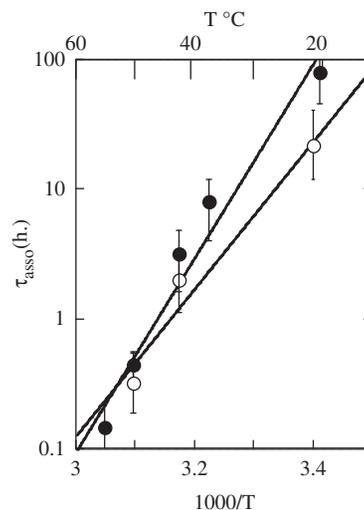


Figure 4. Sms half formation times vs. temperature of C_7GC_4 (black circles) and C_4GC_7 (open circles) 0.3 mM solutions in the NAP buffer at pH 4.6. The activation energies related to sms formation, 143 ± 40 and 110 ± 25 kJ/M are comparable.

equilibrium concentrations is displayed versus pH in Figure 3. Between pH 4 and 5.5, the tetramer reduced dissociation constant, $F_{i,tetra}$, is about 200 times larger than that of the C_7GC_4 dimer. Due to the lower sensitivity of $F_{i,tetra}$ to pH, the dimer and tetramer reduced dissociation constants are comparable around pH 6.5.

Effect of temperature on the sms formation time

We measured by chromatography on GPC 100 column the sms formation times in 0.3 mM C_7GC_4 solutions, pH 4.6 versus temperature. Figure 4 shows that the sms half formation time varies from 0.15 h at 55°C to about 90 h. at 20°C with an activation energy of 143 ± 30 kJ/M.

Effect of the oligonucleotide concentration on sms formation

Figure 5 shows that the sms half formation time decreases as the inverse of the oligonucleotide concentration. The oligonucleotide fraction associated at equilibrium in sms increases from 13% in a 28 μ M C_7GC_4 solution to about 80% in 0.6 mM solution. The oligonucleotide fraction in tetrameric structures, about 10%, depends weakly on the oligonucleotide concentration. The monomer concentration is always negligible in the concentration range explored.

Effect of pH on the sms formation time

I-motif formation involves association of neutral and protonated cytidines. For this reason, the i-motif half formation time is minimal when the product, $[f_C][f_{C^+}]$, of the neutral and protonated cytidine fractions is maximal, i.e. when the pH is equal to the cytidine pK_{N3} (8). On the contrary, the sms formation time is maximal when the pH is close to the cytidine pK_{N3} (Figure 6). It is quite puzzling to note that the effect of pH on the sms formation kinetics (shorter sms formation time when

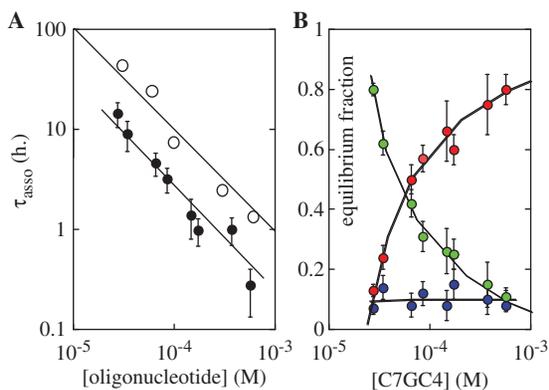


Figure 5. Effect of the oligonucleotide concentration on the sms half formation time at 42°C, pH 4.6, NAP buffer. (A) Half formation time of the sms of C₇GC₄ (black circles) and C₄TC₇ (open circles) versus the oligonucleotide concentration. The slope of the lines drawn through the data points is -1. (B) Dimer (green), tetramer (blue) and sms (red) equilibrium fractions versus the oligonucleotide concentration. The monomer fraction is negligible in this range of concentrations.

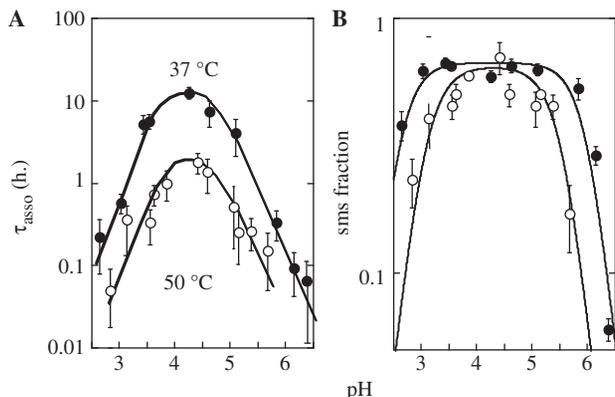


Figure 6. Effect of pH on sms formation in 0.3 mM C₇GC₄ solution at 37 (black circles) and 50°C (open circles) in the NAP buffer. (A) sms half formation time versus pH. The sms formation time is maximal when the pH is equal to the cytidine pK_{N3}. (B) C₇GC₄ fraction incorporated at equilibrium into sms at 37 and 50°C versus pH.

the pH is shifted away from 4.4) is opposite to that observed for the i-motif formation kinetics (longer i-motif formation time when the pH is shifted away from 4.4). At equilibrium, the oligonucleotide fraction incorporated in sms is maximal around pH = 4.4. The equilibrium sms fraction at pH 4.4 depends weakly on the temperature, as shown by the similarity of the sms proportions measured at 37 and 50°C.

Effect of the NaCl concentration on sms formation

Sms formation was examined in 0.3 mM C₇GC₄ solutions at 42°C, pH 4.6 as a function of the NaCl concentration in a buffer containing 10 mM sodium phosphate and 10 mM sodium acetate. The sms half formation time, 100 ± 25 h in the absence of added NaCl decreases rapidly to 3 ± 0.5 h in 0.4 M NaCl. It decreases more slowly at higher NaCl concentration and reaches about 1 h in 2 M NaCl. The C₇GC₄ fraction associated in sms structures

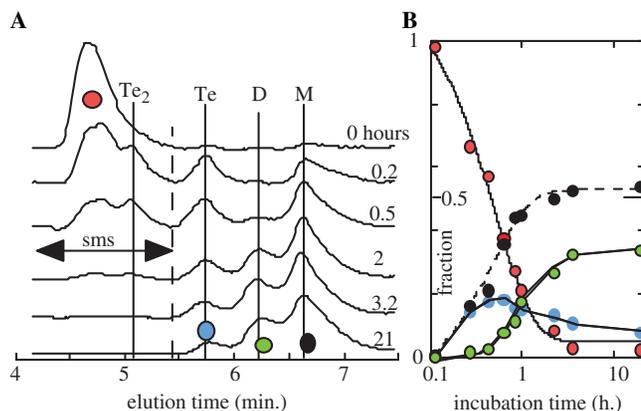


Figure 7. Dissociation of the sms of C₇GC₄ at 40°C, pH 4.63 in the NAP buffer. (A) The oligonucleotide concentration is 3 μ M. The chromatograms recorded as a function of the time show that sms dissociation first results in the formation of monomer (M), tetramer (Te) and Te₂, a species whose elution time corresponds to that expected for the assembly of two tetramers. At last the chromatograms show the formation of a dimer (D). Note that the dimer fraction increases with an initial zero slope. (B) Evolution of the Te (blue), D (green), M (black) and sms (red) fractions as the function of the time.

increases from 10% in the absence of added NaCl to 70% in 0.4 M NaCl solution and decreases slowly to 50% in 2 M NaCl.

sms dissociation kinetics

Figure 7 shows the evolution versus time of the chromatograms recorded during sms dissociation. The sms fraction decreases exponentially with a time constant of 0.46 h at 40°C, pH 4.6. At equilibrium the residual sms fraction is negligible. The chromatograms recorded during the first 30 min of the experiment show the formation of a tetramer (Te), of a species (Te₂) whose molecular weight is about twice that of a tetramer and of monomer. The monomer half formation time, 0.5 h., is close to the sms dissociation time. At last, the chromatograms show the formation of a dimer. The initial zero slope of the plot of the dimer fraction versus time suggests that the dimer is not directly generated by sms dissociation but results of monomer dimerization. The monomer and dimer equilibrium concentrations correspond to a reduced dissociation constant $F_i = 2.5 \cdot 10^{-6}$ M for the dimer at 40°C, pH 4.6.

Effect of pH and temperature on the sms dissociation time

Figure 8 shows that the sms lifetime is maximal around pH 4.4 and decreases at higher and lower pH. The activation energy related to sms dissociation, $E_{\text{ac}} = 255 \pm 20$ kJ/M, seems nearly independent of pH (Figure 8, lower panel). The sms lifetime around pH 4.4 is extraordinary long. The extrapolation of the values measured between 42 and 60°C indicates a lifetime of about 10 years at 20°C.

Other oligonucleotides

All the oligonucleotides of the C₇XC₄ family, where X is either A, T, G or an ethane-diol spacer, as well as C₄GC₇, C₄TC₇, C₆TC₃, C₇TC₃ and C₅TC₅ associate into sms. The

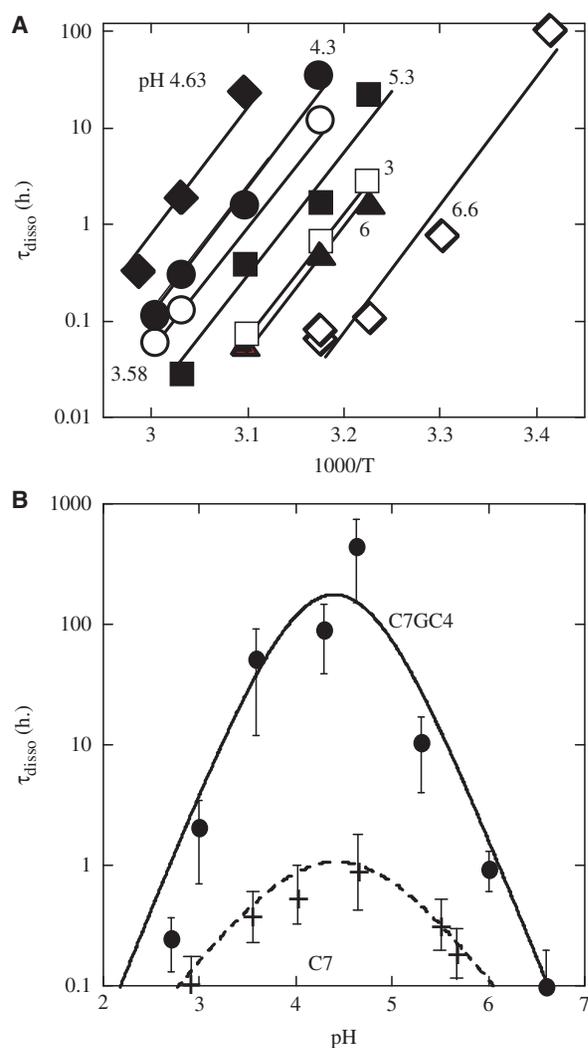


Figure 8. Effect of pH and temperature on the sms lifetimes in the NAP buffer. (A) Half dissociation time of the sms of C_7GC_4 versus temperature at pH 4.6 (black diamonds), pH 4.3 (black circles), pH 3.58 (open circles), pH 5.3 (black squares), pH 3 (open squares), pH 6 (black triangles) and pH 6.6 (open diamonds). (B) Lifetimes at 40°C versus pH of the sms of C_7GC_4 (open circles) and C_7 (crosses). The sms lifetimes are maximal when the pH is equal to the cytidine pK_{N3} and decrease symmetrically at lower or higher pH.

sms molecular weights, formation and dissociation rates and the oligonucleotide fractions associated into sms at equilibrium depend on the length of the C stretches and on the nature of the X spacer. All the C_nXC_m oligonucleotides examined form preferentially a dimer rather than a tetramer. The dimer may be the thermodynamically stable multimer, as this seems to be the case for C_7GC_4 (Figure 3) or it may be kinetically trapped during a time exceeding experimental investigations. We also observed the assembly of C_7 into sms as originally reported by Yamuna Krishnan and collaborators (9).

C_4GC_7 . The reduced dissociation constants of the dimer of C_4GC_7 is comparable to that measured for the reference C_7GC_4 oligonucleotide. The formation and dissociation

rates of the sms of C_4GC_7 (Figure 4 and Supplementary Figure S4) are very similar to the rates measured for C_7GC_4 .

C_7TC_4 , C_4TC_7 and C_5TC_5 . At pH 4.6, the reduced dissociation constants of the dimers formed by these oligonucleotides are at least 10 times smaller than that of the reference C_7GC_4 oligonucleotide. This shows that the nature of the non-cytidine spacer contributes to the dimer stability and therefore interferes with sms formation. The half association times of the sms of C_7TC_4 and C_4TC_7 are 20 and 5 times, respectively longer than that of C_7GC_4 . At 42°C, pH 4.6 the formation rate of the sms of C_5TC_5 is about 100 times slower than that measured for the reference C_7GC_4 oligonucleotide. After incubation during 5 days, in condition close to equilibrium, the chromatogram of a 5 mM C_5TC_5 solution shows that the dimer, tetramer and sms fractions are 60, 20 and 20%, respectively.

C_7 . By comparison with C_7GC_4 , the association rate of C_7 into sms is faster but the lifetimes of the sms of C_7 measured at different pH is much shorter (Figure 8).

DISCUSSION

A good knowledge of the factors that influence the sms growth is essential to control the assembly of C-rich oligonucleotides into structures with specific applications (16). The detail of the sms formation pathway is beyond the scope of this article; nevertheless some indications on the sms growing process may be derived from the experiments presented above.

In summary, we have shown that all the C_nXC_m oligonucleotides investigated associate into sms. The sms formation rates and equilibrium proportions depend on the C stretch length and on the nature of the spacer. It is interesting to note that the influence of temperature and pH on the sms formation kinetics is quite different from the effect of these factors on the i-motif formation kinetics (5). In contrast with the formation time of i-motif tetramers, which increases as the power of -2 of the monomer concentration (5), the sms formation time increases as the power of -1 (Figure 5). Moreover, we have observed that sms disruption releases short sms fragments, among which only Te_2 may be resolved on the GPC-100 column, and tetrameric structures (Te) that in turn dissociate into monomer. The similarity of the time constant for sms dissociation and monomer formation together with the exponential character of the evolution of these species indicate that tetramer dissociation is not limiting.

Competition between i-motif dimers and tetramers

A constant property of C_nXC_m oligonucleotides is their capacity to form stable i-motif dimers. It must be noticed that these dimers are certainly not parallel hemiprotonated duplexes. There is no experimental indication in the literature in favor of the existence of stable C-rich hemiprotonated duplexes in oligo-C solutions and

recent attempts to enforce their formation led to the conclusion that these species must be considered as short-lived precursors of i-motif tetramers in fast exchange with the monomer (7).

Two families of i-motif dimer have been described (2,6,17). One is formed by the parallel arrangement of two hairpins whose loops are on the same side of the i-motif core, the other by two hairpins in head to tail orientations (Scheme 2). Many conformers may coexist in each family, depending on the loop position with respect to the narrow and wide grooves, on the intercalation topology of the $C\cdot C^+$ pairs and on the loop composition. If one considers the structures displayed in Scheme 2, it is hard to imagine that they could induce or contribute to sms formation. It is therefore tempting to regard the dimer as a dead-end way in the sms formation pathway.

Several examples have been reported showing that formation of i-motif dimers is kinetically favored by comparison with tetramer formation. The thermodynamically stable multimer of $T_2C_8T_2$ (14), 5mCCTCTCC (5) and 5mCCTCTCTCC (7) are tetrameric. Nevertheless, after melting and fast cooling these oligonucleotides are kinetically trapped into hairpin dimers and it is observed that the dimer/tetramer conversion is extremely slow.

The formation time of i-motif multimers should increase in proportion with the number of the steps leading to the stable final structure. It is therefore reasonable to assume that this is the difference of the formation order of dimer and tetramer, 2 and 3, respectively (6,8), that accounts for the faster dimer formation rate.

Dissociation of $[C_nXC_m]_2$ dimers is the limiting step of sms elongation

The study of C_7TC_4 and C_4TC_7 gives indirect arguments suggesting that the dimer stability and the sms formation rate are correlated. The slow formation rate of the sms of these sequences may be a consequence of the stability of the dimers formed by these oligonucleotides. It may be also noticed about the stability of the dimers with a T spacer that the loop topology of the structure of Scheme 2C allows the formation of a cross-loop T•T pair (5) that could account for the enhanced stability of the duplexes with a thymidine spacer.

If one supposes that the competition between dimer and tetramer hinders sms growing, it is predictable that the pH and temperature conditions favorable to tetramer formation should accelerate the sms elongation. The simulations displayed in Figure 3b indicate that in response to the effect of pH on the stability of the dimer of C_7GC_4 , the tetramer fraction increases between pH 5.3 and 6.3 by more than two magnitude orders and suggests that the acceleration of the sms formation rate observed in this pH range (Figure 6) results of dimer dissociation.

Figure 4 shows that the sms formation time decreases when the temperature is raised ($E_{act} = 143$ kJ/M for the sms of C_7GC_4 , Figure 4). By contrast, the i-motif tetramer formation times increase with temperature ($E_{act} = -197$ and -306 kJ/M for $[TC_3]_4$ and $[TC_5]_4$, respectively (8). The lengthening of the formation time of i-motif tetramers at high temperature has been ascribed to the reduction of

the proportion of an intermediate dimer or trimer species (8). By analogy with the interpretation given above to the effect of pH on the sms formation rate, the acceleration of the sms formation rate at high temperature may be a consequence of the preferential dimer dissociation at high temperature.

By contrast with the formation time of i-motif tetramers that varies as the power of -2 of the monomer concentration, the sms formation time varies as the power of -1 of the concentration (Figure 5). This indicates that the sms formation kinetics is not directly controlled by the formation rate of the i-motif building blocks (Te in Scheme 1). This argues also against an sms growing process involving successive addition of monomer units to a tetrameric nucleus whose rate should be strongly dependent on the oligonucleotide concentration.

A model for sms formation

The sms formation pathway displayed in Scheme 1 is in part inspired by that proposed to account for C_7 association into i-wire structures (9). All the arguments developed above together with the similarity of the time constants for sms formation and dimer dissociation suggest that dimer dissociation is the limiting step of the sms formation pathway. According to Scheme 1, it is noteworthy that since dimer formation is faster than tetramers formation (i.e. $kon_{dimer} > kon_{tetra}$), the dimer concentration should decrease at a rate $k = kof_{dimer} \times kon_{tetra} / (kon_{dimer} + kon_{tetra})$ which is slower than the dimer dissociation rate, kof_{dimer} .

As shown in a recent study of the formation pathway of i-motif tetramers, structures with partial or full intercalation topologies are formed at comparable rates and the species with optimal intercalation topology prevails at equilibrium due to its longer lifetime. The X spacer was inserted in the C_nXC_m sequence to prevent intercalation of the $m + n$ cytidines into a single i-motif core.

The tetramer formed by intercalation of the longer C stretch (Te in Scheme 1) is expected to be the thermodynamically stable species and we postulate that it is the building block of the sms growing process. We assume that after the nucleation step, association by pairing and intercalation of the terminal C_4 stretches of two Te building blocks initiates sms formation. The i-motif symmetry gives to the assembly of several building blocks the same overhanging C_4 terminations than the building block itself. This allows sms elongation by association of preformed Te_n and Te_m assemblies into structures including $(n + m)$ building blocks (Scheme 1) or by successive association of Te blocks.

The lifetime of i-motif tetramers increases strongly with the number of intercalated $C\cdot C^+$ pairs (8). It is therefore predictable that the connecting i-motif blocks formed by the C_4 stretches should be the weak sms links. However, it is interesting to note that the lifetime of the sms of C_7GC_4 , about 100 h at 40°C, pH 4.5 (Figure 8), is about 600 times longer than that of the i-motif tetramer of TC_4 (8). This suggests that the i-motif block formed by intercalation of the C_4 stretch is stabilized in the sms by staking

interactions with the outer C•C⁺ pairs of the i-motif sections built by the C₇ moiety.

The structure of supramolecular i-motif assemblies

C•C⁺ intercalation into i-motif structures requires an extension of the helical rise at the C-C steps up to 6.3 Å, a value close to the maximum accessible without base-pair disruption (18). Due to this extreme helical stretch and compactness, the i-motif should be an unbendable structure and it is predictable that the i-motif sms should look like extended stiff structures.

We have little information at this time about the structure of supramolecular i-motif assemblies. The formation pathway proposed just above, supposes implicitly that the sms formed by C_nXC_m oligonucleotides are linear structures. Preliminary AFM investigations indicate that the structures seem markedly influenced by the nature of the X spacer. AFM images of C₇TC₄ samples deposited on mica show bent structures as long as micrometer with unexpected 120° bifurcations, those of C₄AC₇ and C₄GC₇ form rigid rods whereas the structures formed by C₇GC₄ look like aggregated shapeless pellets. It is clear that further investigations are required to give a reliable description of the supra molecular formed by C_nXC_m oligonucleotides.

NMR is obviously not appropriated to structural studies of large sms. Nevertheless, the long sms lifetimes should allow purification of tetramer assemblies short enough to be accessible to structural investigations by NMR methods. It should be specially interesting to determine the arrangement of the X spacer in the structure (Te₂, Scheme 1) formed by association of two building blocks.

It is noteworthy that regardless of their structural differences, the G-quartet (19–21) and i-motif tetramers have similar formation pathways (8,15) and associate into supramolecular structures. While it is assumed that i-motif sms are generated by the association of incompletely intercalated preformed tetrameric building blocks (9), G-wire elongation is described as a process involving association of parallel duplexes to an out-of-register G-quadruplex matrix (22).

The lifetimes of partially matched G-quadruplexes and i-motif multimers may be extremely long (6,8,23). For this reason, the formation of both tetramer families is conditioned by the evolution of species that are kinetically trapped (8,15). It is therefore predictable that as observed for the supramolecular assemblies of i-motif, the kinetic trapping of mismatched species is also a crucial limitation to the elongation of G-wire structures.

The excellent yield of automated chemical DNA synthesis together with the availability of a large variety of artificial residues with functional specificities make DNA oligonucleotides an attractive building material for sms (24–28). The DNA supramolecular assemblies based on Watson-Crick pairs or on G-quartets must be heated at melting temperature to be dissociated. In contrast, the extreme i-motif sensitivity to pH allows dissociation of i-motif sms by a mild pH change from pH 6 to 7. This property may be extremely interesting in assemblies

including standard DNA and i-motif sections by allowing a conformational change triggered by pH.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Yamuna Krishna is gratefully acknowledged for discussion and information on i-wire formation. Eric Lesniewska is acknowledged for communication of AFM images.

FUNDING

Funding for open access charge: Centre National de la Recherche Scientifique.

Conflict of interest statement. None declared.

REFERENCES

- Gehring, K., Leroy, J.-L. and Guéron, M. (1993) A tetrameric DNA structure with protonated cytidine-cytidine base pairs. *Nature*, **363**, 561–565.
- Nonin, S., Phan, A.T. and Leroy, J.-L. (1997) Solution structure and base-pair opening kinetics of the i-motif dimer of d(5mCCTTACC): a non canonical structure with possible role in chromosome stability. *Structure*, **5**, 1231–1246.
- Phan, A.T., Guéron, M. and Leroy, J.-L. (2000) The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere. *J. Mol. Biol.*, **299**, 123–144.
- Nonin, S. and Leroy, J.-L. (1996) Structure and conversion kinetics of a bi-stable DNA i-motif: broken symmetry in the [d(5mCCTCC)]₄ tetramer. *J. Mol. Biol.*, **261**, 399–414.
- Leroy, J.-L. (2004) T.T pair intercalation and duplex interconversion within i-motif tetramers. *J. Mol. Biol.*, **333**, 25–139.
- Canalia, M. and Leroy, J.-L. (2005) Structure, internal motions and association-dissociation kinetics of the i-motif dimer of d(5mCCTCACTCC). *Nucleic Acids Res.*, **33**, 5471–548.
- Canalia, M. and Leroy, J.-L. (2009) [5mCCTCTCCTCC]₄: an i-motif tetramer with intercalated T•T pairs. *J. Am. Chem. Soc.*, **131**, 12870–12871.
- Leroy, J.-L. (2009) The formation pathway of i-motif tetramers. *Nucleic Acids Res.*, **37**, 4127–4134.
- Ghodke, H.B., Krishnan, R., Vignesh, K., Kumar, P., Narayana, C. and Krishnan, Y. (2007) The I-Tetraplex building block: Rational design and controlled fabrication of robust 1D DNA scaffolds through non-Watson-Crick interactions. *Angew. Chem. Int. Ed.*, **46**, 2646–2649.
- Cantor, C.R. and Warshaw, M.M. (1970) Oligonucleotide interactions. III. circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers*, **9**, 1059–1077.
- Bhatia, D., Mehtab, S., Krishnan, R., Indi, S., Basu, A. and Krishnan, Y. (2009) Icosahedral DNA Nanocapsules by Modular Assembly. *Angew. Chem.*, **48**, 4134–4137.
- Plateau, P. and Guéron, M. (1982) Exchangeable proton NMR without base-line distortion, using strong pulse sequences. *J. Amer. Chem. Soc.*, **104**, 7310–7311.
- Guéron, M., Plateau, P. and Decors, M. (1991) Solvent signal suppression in NMR. *Prog. NMR spectr.*, **23**, 135–209.
- Leroy, J.-L., Gehring, K., Kettani, A. and Guéron, M. (1993) Acid multimers of oligo-cytidine strands: stoichiometry, base-pair characterization and proton exchange properties. *Biochemistry*, **32**, 6019–6031.
- Bardin, C. and Leroy, J.-L. (2008) The formation pathway of tetrameric G-quadruplexes. *Nucleic Acids Res.*, **36**, 477–488.

16. Davis, A.V., Yeh, R.M. and Raymond, K.N. (2002) Supramolecular assembly dynamics. *Proc. Natl Acad. Sci. USA*, **99**, 4793–4796.
17. Gallego, J., Chou, S.-H. and Reid, B. (1997) Centromeric pyrimidine strands fold into an intercalated motif by forming a double hairpin with a novel T:G:G:T tetrad: solution structure of the d(TCCCGTTTCCA) dimer. *J. Mol. Biol.*, **273**, 840–856.
18. Lebrun, A. and Lavery, R. (1996) Modelling extreme stretching of DNA. *Nucleic Acids Res.*, **24**, 2260–2267.
19. Lane, A.N., Chaires, J.B., Gray, R.D. and Trent, J.O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, **36**, 5482–5515.
20. Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K. and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
21. Davis, T. and Spada, G.P. (2007) Supramolecular architectures generated by self-assembly of guanosine derivatives. *Chem. Soc. Rev.*, **36**, 296–313.
22. Marsh, T.C., Vesenka, J. and Henderson, E. (1995) A new DNA nanostructure, the G-wire, imaged by scanning probe microscopy. *Nucleic Acids Res.*, **23**, 696–700.
23. Mergny, J.L., De Cian, A., Ghelab, A., Saccà, B. and Lacroix, L. (2005) Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res.*, **33**, 81–94.
24. Krishnan-Ghosh, Y., Dongsheng, L.D. and Balasubramanian, S. (2004) Formation of an interlocked quadruplex dimer by d(GGGT). *J. Am. Chem. Soc.*, **126**, 11009–11016.
25. Miyoshi, D., Karimata, H., Wang, Z.M., Koumoto, K. and Sugimoto, N. (2007) Artificial G-wire switch with 2,2'-bipyridine units responsive to divalent metal ions. *J. Am. Chem. Soc.*, **129**, 5919–5925.
26. Niemeyer, C.M. (2000) Self-assembled nanostructures based on DNA: towards the development of nanobiotechnology. *Curr. Opin. Chem. Biol.*, **4**, 609–618.
27. Shen, Z., Yan, H., Wang, T. and Seeman, N.C. (2004) Paranemic crossover DNA: a generalized Holliday structure with applications in nanotechnology. *J. Am. Chem. Soc.*, **126**, 1666–1674.
28. Pitchiaya, S. and Krishnan, Y. (2006) First blueprint, now bricks: DNA as construction material on the nanoscale. *Chem. Soc. Rev.*, **35**, 1111–1121.

Supplementary Material for

[C7GC4]4 Association into supra molecular i-motif structures

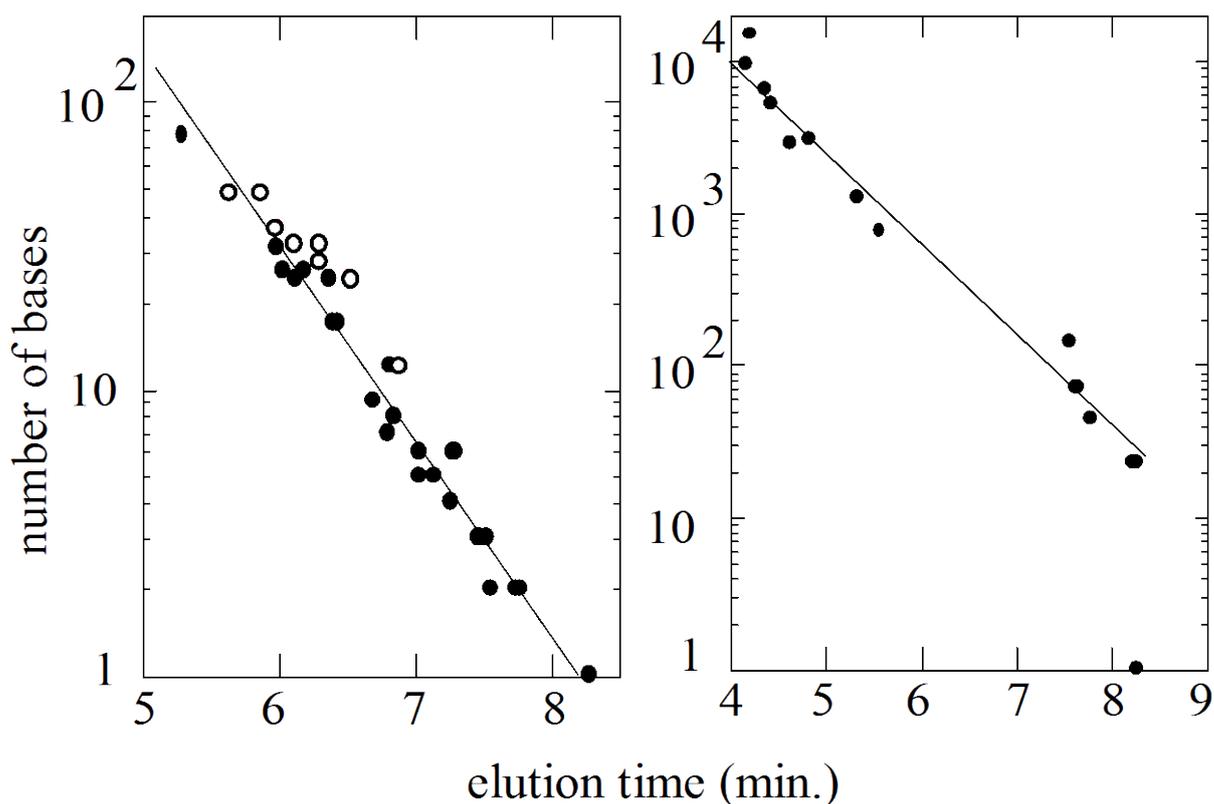


Fig. S1

Figure S1. Calibration curves of the GPC 100 (left panel) and GPC 1000 (right panel) gel filtration columns used to determine the sms molecular weight and the multimer stoichiometry of C_nXC_m i-motif assemblies. The GPC 100 column was calibrated with tRNA, thymidine, C-rich i-motif monomers, non-structured C-rich oligonucleotide (black dots) and i-motif tetramers characterized in previous NMR investigations (open dots). The data points corresponding to structured or non-structured oligonucleotide families do not show significant deviation indicating a marked influence of the oligonucleotide structure. The GPC 1000 column was calibrated with double stranded plasmids.

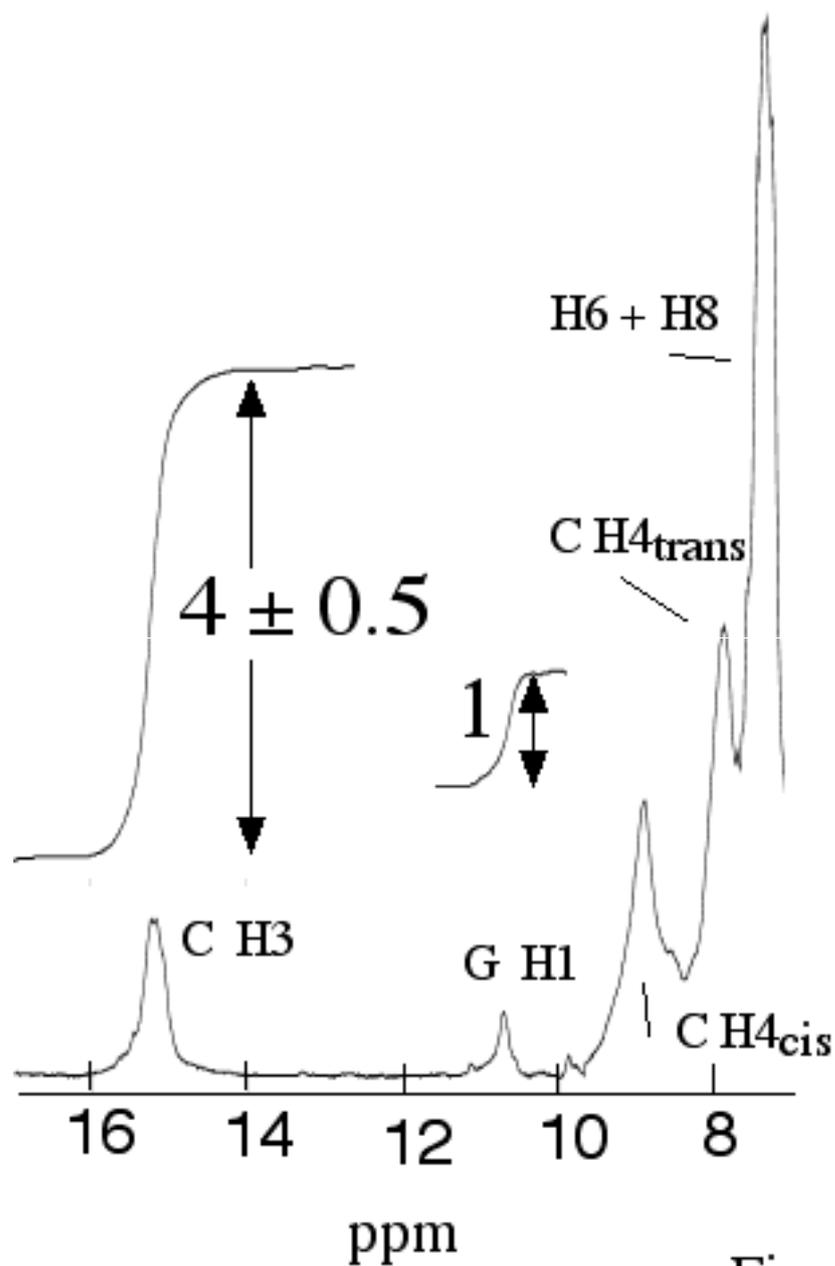


Fig. S2

Figure S2. Exchangeable and aromatic proton region of a 1.5 mM C7GC4 solution pH 4.6. The spectrum was recorded at 0°C about 2 minutes after melting and fast cooling. The area of the imino proton peak is about 4 times larger than that of the G imino proton, a value consistent with the structures displayed in scheme 2.

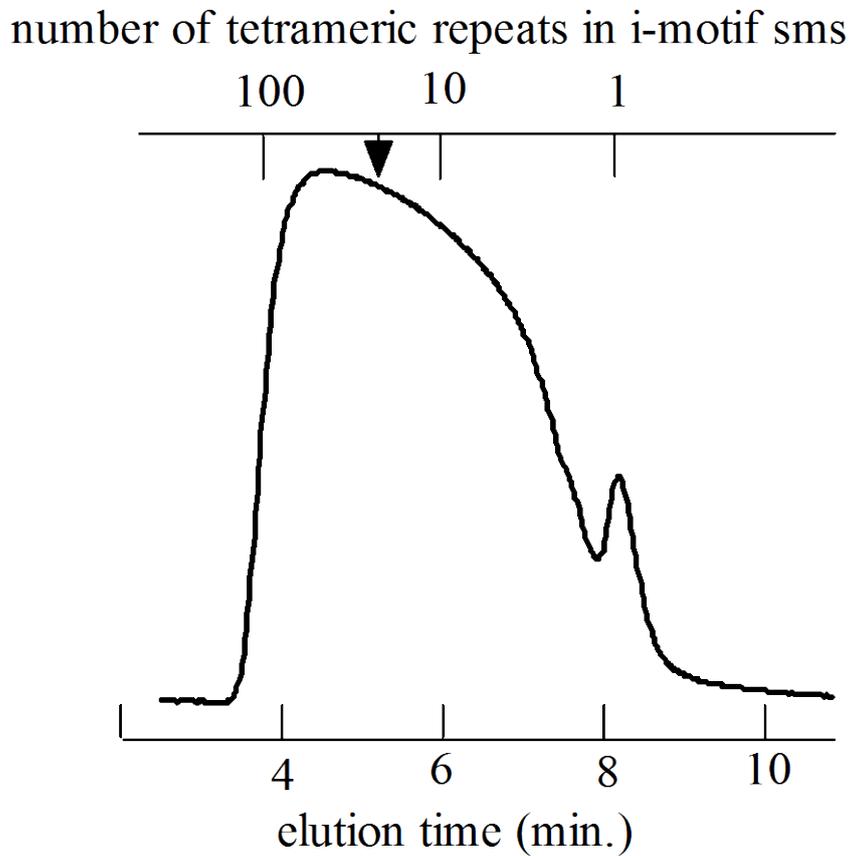


Fig. S3

Figure S3 . GPC 1000 chromatogram of a 4.4 mM C_4GC_7 solution pH 4.6 incubated at 42°C during 68 h. 90 % of oligonucleotide is associated in sms. The center of the sms distribution is indicated by an arrow. It corresponds to a molecular weight equivalent to that of 25 tetramers (i.e. 1800 résidus). About 5% of the sms include more than 90 tetramers.

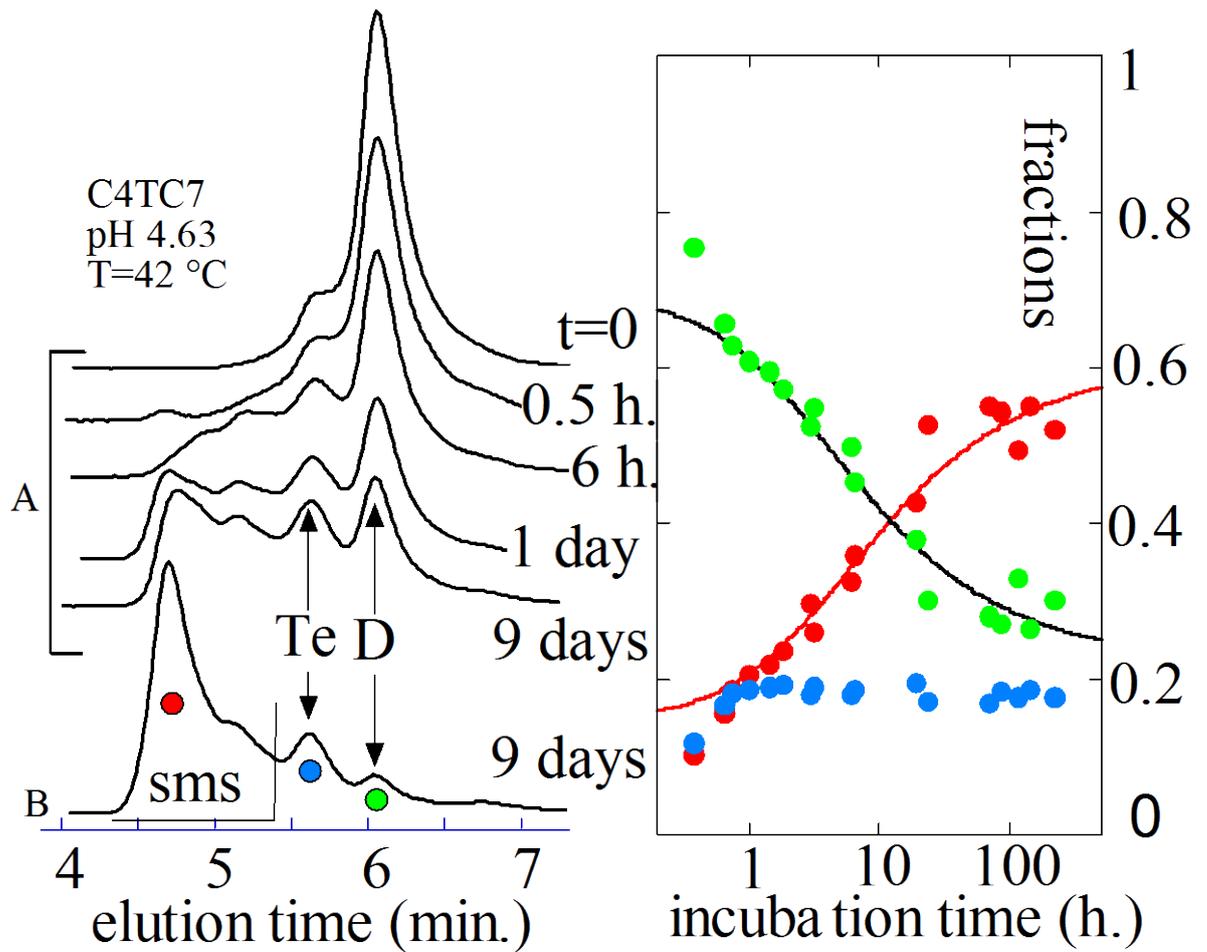


Fig. S4

Figure S4. C₄TC₇ association into sms at 42 °C, pH 4.6. **Left panel:** Chromatograms of 0.3 mM (B) and 3 mM (A) solutions recorded after incubation during the times indicated on the figure. The elution times expected for the dimer and tetramer of C₄TC₇ are labeled D and Te. **Right panel:** Evolution of the dimer (green), tetramer (blue) and sms (red) fractions as a function of the incubation time. At equilibrium, the oligonucleotide fractions associated in sms are 60 % and 80 % in 0.3 mM and 3 mM solutions respectively

L'article apporte des informations sur les facteurs clés qui influencent la formation des structures supramoléculaires de motif-i, et propose une hypothèse sur le mécanisme de formation des structures supramoléculaires. Ces connaissances sont cruciales pour contrôler l'assemblage des structures en vue d'applications spécifiques en biotechnologie.

Les séquences du Tableau 2, hormis C₈TC₅, forment des structures supramoléculaires. Les vitesses de formation et les proportions de fils à l'équilibre varient en fonction de la longueur des segments de cytidines et de la nature du résidu central, ce qui souligne le rôle crucial de ces paramètres dans les mécanismes d'élongation. La formation de dimères de motif-i, qui ne peuvent pas s'intercaler, constitue l'étape limitant la polymérisation des structures. La stabilité des dimères, qui est d'autant plus élevée que la séquence est longue, bloque l'élongation des motifs-i. La séquence C₈TC₅ constitue la limite de longueur à partir de laquelle les dimères de motifs-i plus rapidement formés que les tétramères sont suffisamment stables pour empêcher la formation de structures supramoléculaires.

La durée de vie des structures (TCCCC)₄ incluses dans les structures supramoléculaires C₇GC₄ est 600 fois plus longue (~100h à 40°C, pH4.5) que celle du tétramère de motif-i (TCCCC)₄ isolé [423]. Cette particularité pourrait s'expliquer par le fait que les blocs de C₄ intercalés au sein des structures supramoléculaires sont protégés par les bases adjacentes, avec l'énergie d'empilement qui stabilise la structure. A titre de comparaison, la durée de vie du tétramère de (TC₄)₄ présente un facteur 100 avec celle du tétramère de (TC₃)₄, alors qu'une seule base différencie les deux séquences incluses dans le motif-i. Dans le cas des structures supramoléculaires de C₇GC₄, les énergies d'interactions non covalentes, ainsi que la distribution statistique prévisible de l'étendue d'intercalation des motifs-i, paraissent des explications insuffisantes pour justifier une telle différence de durée de vie avec (TC₄)₄. Ceci soulève la question de savoir s'il existe un mécanisme unique d'élongation des structures supramoléculaires ou si plusieurs mécanismes coexistent en proportions variables selon la nature des séquences et la taille des polymères.

La structure des assemblages supramoléculaires de motif-i abordée dans l'article va être illustrée par les images obtenues en microscopie de force atomique dans la partie « Données complémentaires ».

3. Données complémentaires

3.1. Caractérisation par microscopie de force atomique.

Nous allons présenter dans cette partie l'étude structurale des assemblages supramoléculaires de motifs-i. Les formes les plus stables seront plus propices pour l'imagerie à l'air de l'auto-organisation des polymères. Selon les séquences, les échantillons qui polymérisent ont une consistance de gel, avec une distribution de taille non contrôlée. L'objectif de l'analyse par AFM est d'observer la structure et distribution des grands

polymères, qui n'ont pu être caractérisées par les méthodes de filtration sur gel et RMN employées dans l'approche physico-chimique, mais aussi d'observer les structures intermédiaires d'assemblage afin de comprendre les mécanismes de polymérisation.

L'AFM à l'air, en mode contact intermittent, réalisée sur le modèle C_mXC_n a conduit à l'observation de différentes structures selon la séquence considérée (longueur des brins m et n) et la nature de la base centrale. Les structures supramoléculaires de motifs i observées peuvent se décliner en deux groupes :

- Des amas non organisés ou des images ne montrant pas de formes définies pour C_7GC_4 . Ces mêmes échantillons ont pourtant été caractérisés en tant que structures supramoléculaires par filtration sur gel.
- Des structures organisées à longue distance pour C_7TC_4 . Différentes conditions d'expérience ont été réalisées pour cette séquence afin de mieux caractériser les structures impliquées dans les mécanismes de polymérisation.

Concernant les séquences C_6TC_3 et C_7 , aucune image n'a permis de classer les structures dans l'une des deux catégories, tandis que pour C_4AC_7 , une image a montré des structures linéaires, en formes de bâtonnets rigides.

Equipement :

Le microscope utilisé est le Nanoscope 8 Multimode avec un scanner présentant une plage de balayage maximum de 15 μm .

Les pointes de type Scan Asyst Air (Bruker) ont été utilisées. La constante de raideur était $K=0.4 \text{ N/m}$ pour une fréquence de résonance $f=70 \text{ kHz}$

3.1.1 Observations des structures supramoléculaires de C_4AC_7 .

Les séquences de type C_4AC_7 dilués dans la solution d'acide acétique 0.5 mM pH4.5 ont été déposées sur le mica prétraité puis séchées avant observation. Tous les échantillons nécessitent d'être dilués avant observation, le dépôt sur surface ne faisant qu'accentuer la concentration locale des structures qui risquent de se superposer.

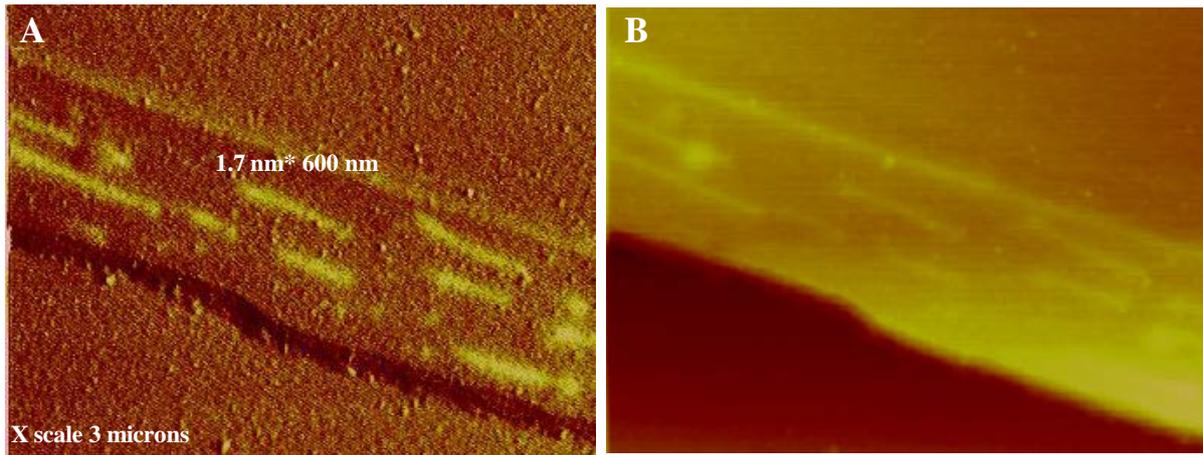


Figure 79 : Image AFM de C7AC4.

A, image de phase. B, image de hauteur avec un différentiel de 40 nm avant et après la marche du réseau cristallin.

Une seule image a pu être obtenue à partir de la séquence C₇AC₄. De longs fragments linéaires orientés par la marche cristalline du mica sont visibles sur la Figure 79. Ces structures en forme de bâtonnets rigides correspondent aux structures supramoléculaires attendues à partir des données structurales connues des motifs-i. Il est à noter que ce sont les seules structures linéaires qui ont été observées et uniquement dans ces conditions spéciales d'irrégularité de surface du mica. La structure annotée présente une longueur de 600 nm pour une hauteur de 1,7 nm relevée sur l'image de hauteur, très contrastée du fait de la marche du réseau cristallin.

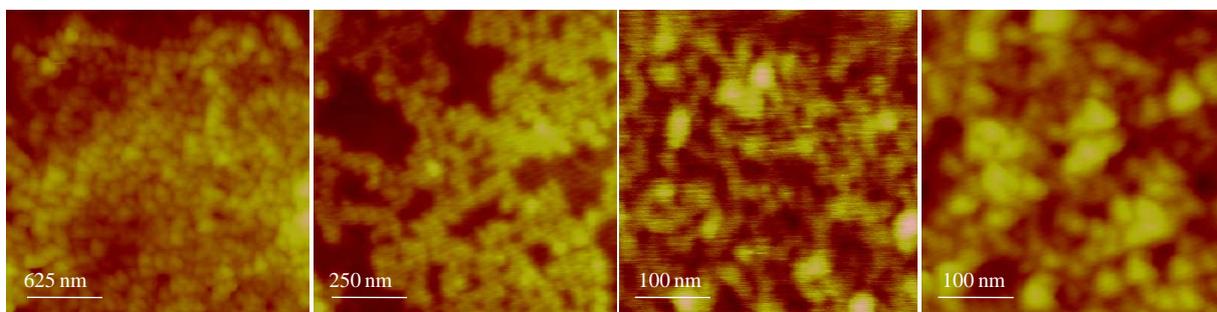


Figure 80 : Images AFM de C7GC4.

Les images obtenues à partir de la séquence C₇GC₄ ne montrent pas de structure définie, parfois en réseau, parfois en amas en forme de boules de coton.

3.1.2 Observation des structures supramoléculaires de C₇TC₄

Des solutions de C₇TC₄ de différentes concentrations à l'équilibre ont été imagées par AFM.

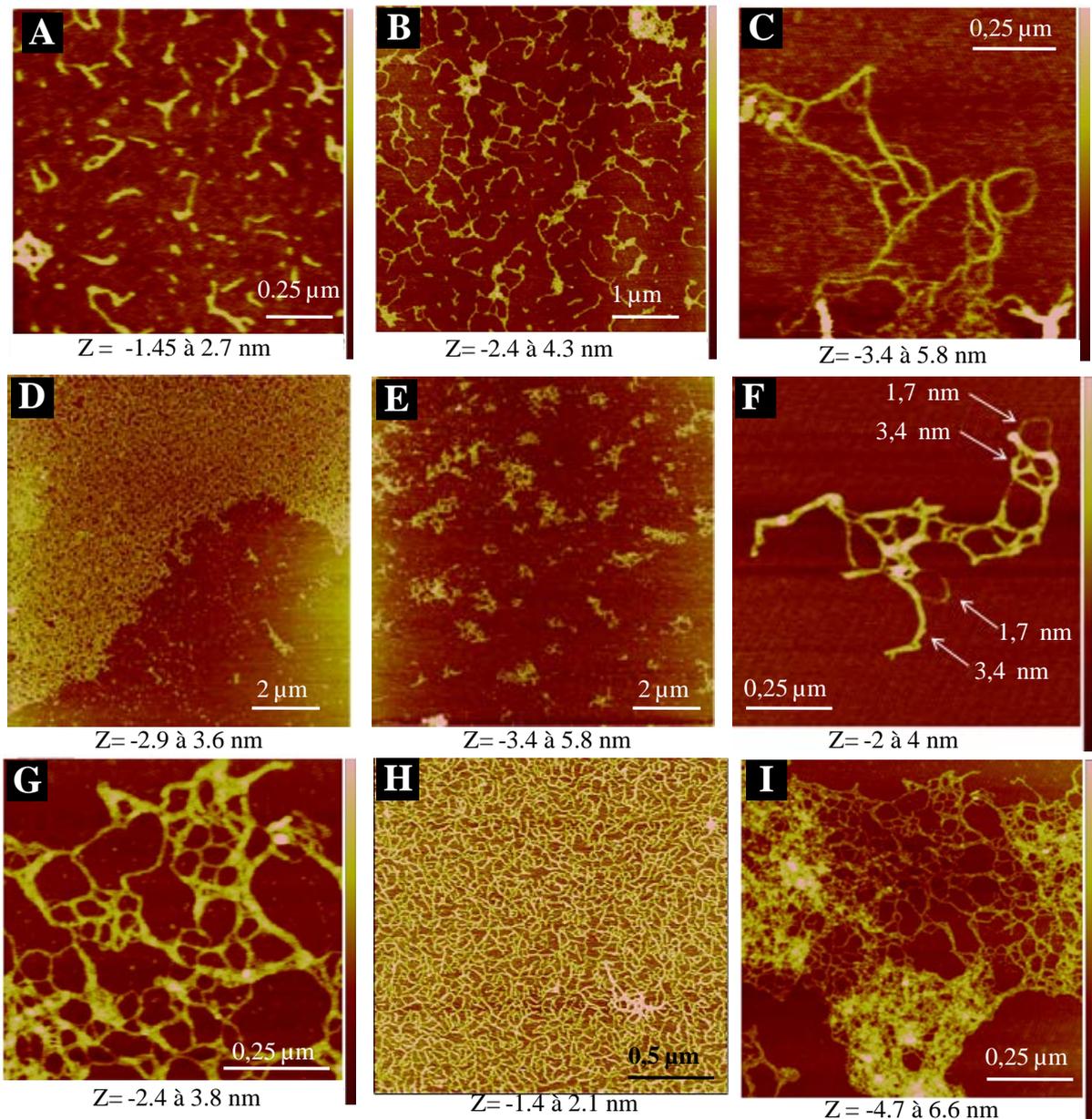


Figure 81 : images AFM de structures de C7TC4.

A, unités élémentaires. **B**, structures qui suivent le réseau cristallin. **C**, formes linéaires. **D**, marge de réseau en croissance. **E**, briques intermédiaires du réseau. **F**, structure définie avec différentes épaisseur de brins. **G**, Réseaux avec quelques superpositions et de maillage de tailles diverses. **H**, réseaux très régulier d'épaisseur régulière. **I**, réseaux imbriqués.

Les images de la Figure 81 illustrent diverses organisations des structures supramoléculaires obtenues à partir de la séquence C₇TC₄ :

- des structures élémentaires linéaires et des structures élémentaires branchées (jonction 3 voies) sur l'image A,
- des structures de type fractales, qui suivent le réseau cristallin du mica sur l'image B,
- des structures filaires longues, telles des « spaghettis », qui se branchent en quelques endroits sur l'image C,
- un réseau en croissance, où à la marge du réseau des fragments élémentaires et intermédiaires sont présents sur l'image D,

- des structures intermédiaires dans la formation des réseaux (image E),
- des structures définies dont les différentes épaisseurs suggèrent des interactions entre les brins sur l'image F du « scorpion »,
- des réseaux présentant des superpositions et des maillages de taille variable (image G),
- des réseaux réguliers, comme des filets, qui couvrent l'ensemble de la zone observée et qui sont homogènes en termes d'épaisseur (image H),
- des réseaux imbriqués hétérogènes (image I).

Nous avons l'impression d'observer des structures qui ont subi des historiques différentes. Certaines images illustrent des structures (fils) probablement formées en solution et qui auraient été plaquées sur la surface, telles que les longues structures filaires et les réseaux imbriqués et probablement certaines structures élémentaires (images C, F, G ?, I). Les structures en réseau présenteraient aussi la possibilité de se réorganiser au contact de la surface par interactions des extrémités libres, non intriquées. D'autres images (A, B, D, E, H) indiquent que les structures dynamiques se seraient formées et organisées au contact de la surface. C'est le cas des structures qui semblent s'orienter selon les mailles cristallines du mica et des grands réseaux réguliers à longue distance qui, tout en couvrant la totalité de la surface observée, présentent une homogénéité d'épaisseur sans superposition apparente dans les réseaux des images B, D, H.

Les évènements de placage de structures préformées en solution et de structures qui s'organisent au contact de la surface en étant guidée par elle, ne sont pas exclusifs. Nous pouvons observer leur cumul, ce qui rend délicat l'interprétation des images pour expliquer les phénomènes impliqués dans la polymérisation des structures. Toutes les organisations hormis les formes filaires sont totalement inattendues d'après les connaissances structurales du motif-i.

La formation de réseaux sur les surfaces et non en solution peut être la conséquence d'un ordre cinétique élevé de la réaction de couplage entre bloc avec existence d'un « seuil » de concentration qui ne serait dépassé qu'après concentration par plaquage sur la surface des monomères ou de courts polymères.

3.1.3 Observations de structures supramoléculaires de C_7TC_4 après tri en taille par tamisage moléculaire dans des tampons de pH croissant:

Afin d'identifier les différentes unités élémentaires et intermédiaires impliquées dans le mécanisme de polymérisation, nous avons séparé par taille par tamisage moléculaire des superstructures à l'équilibre de C_7TC_4 qui ont été préalablement prétraitées dans des tampons de pH croissant. Les échantillons C_7TC_4 ont été dilués soit dans un tampon à pH 4.6 (fractions J1 à J5, répertoriées dans le Tableau 3), soit dans un tampon à pH 6.8 (fractions K1, K2), soit dans un tampon à pH 7.9 (fractions L1 à L3) avant injection sur la colonne de tamisage

moléculaire GPC 1000 (Eprogen, 250 x 4.6 mm DI) à 400 µl/min en tampon 10 mM NaH₂PO₄ 400mM NaCl pH4.6.

Les fractions J1 à J5, K1-K2 et L1 à L3 correspondent respectivement à des structures de taille décroissante, comme indiqué dans le Tableau 3.

Chacune des fractions a été imagée par AFM sur mica muscovite vert prétraité au NiCl₂ [10 mM], puis rincé avec la solution d'acétate [0.5 mM - pH 4.6]. 20 µl d'échantillon C7TC4 d'une fraction sont déposés et incubés pendant 20', puis rincés avec le tampon acétate avant d'être observés.

Tableau 3 : récapitulatif des fractions de C7TC4 éluées sur colonne de filtration sur gel.

La colonne de filtration sur gel a été calibrée avec des ADN double brin, les équivalences de tailles attendues des structures sont précisées dans la colonne 3. Les temps de rétention des structures sur HPLC s'échelonnent de 4,25 à 6,5 minutes.

Dilution avant injection	Taille équiv. ADN B db paire de bases	Taille équiv. en nm 40 bp ~12 nm	Tube
pH 4.6	1260	380	J1
	912	270	J2
	602	180	J3
	323	95	J4
	169	50	J5
pH 6.8	549	165	K1
	251	75	K2
pH 7.9	501	150	L1
	316	95	L2
	208	60	L3

Seules quelques images sont présentées sur la Figure 82 pour illustrer les structures obtenues. Si celles-ci sont différentes d'une fraction à l'autre, elles ne suivent pas la distribution de taille attendue par rapport à la caractérisation en tamisage moléculaire. Nous avons obtenu quelques structures de diverses tailles (50 à 300 nm) pour une fraction où nous attendions des structures avoisinant 380 nm, et à l'inverse nous observons la formation d'un réseau avec de nombreuses superpositions pour la fraction J3 (image B) où nous attendions des structures de tailles autour de 180 nm. Les structures les plus inattendues présentant un imbroglio de brins sont observées pour la fraction L3.

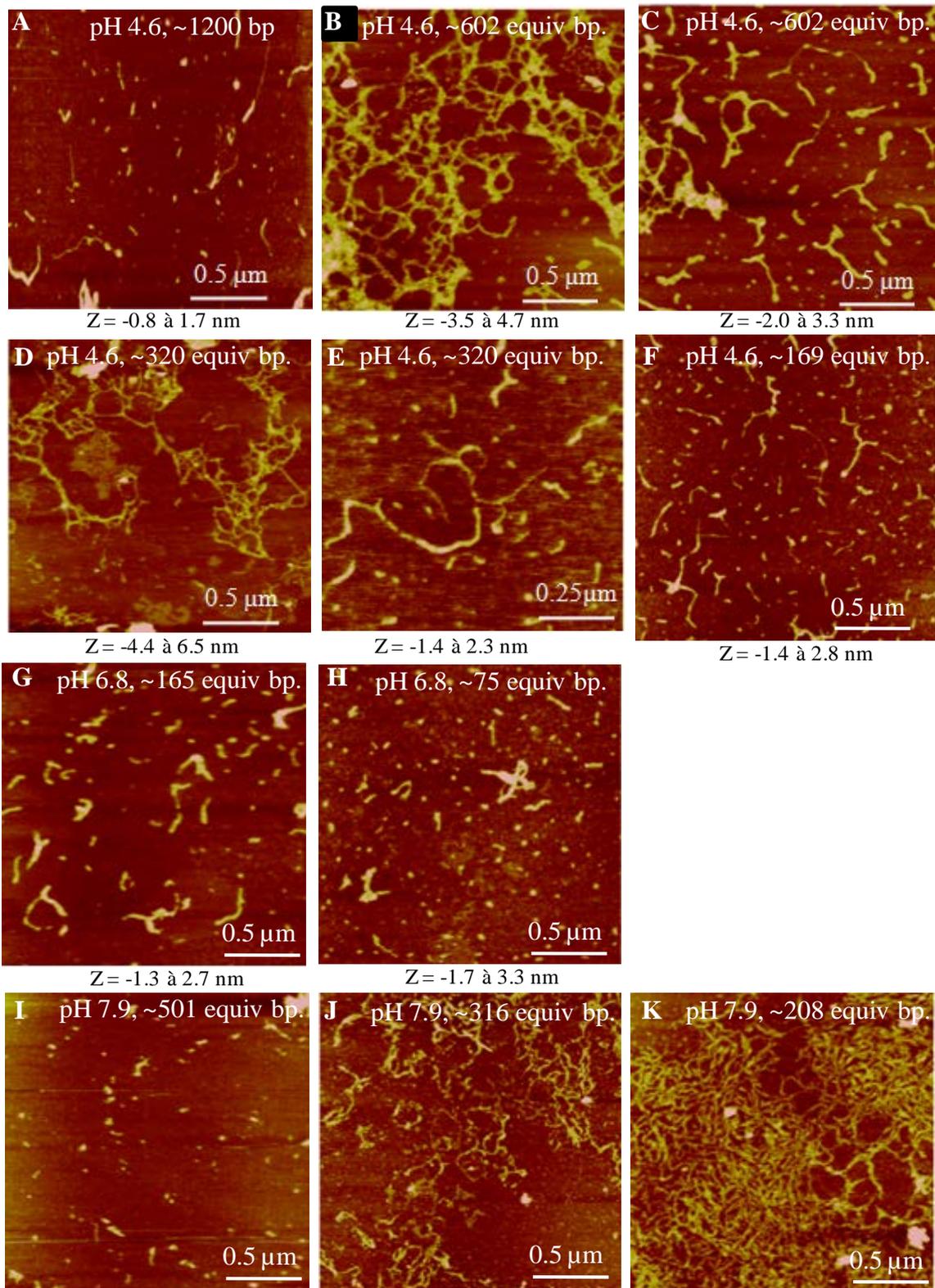


Figure 82 : Images AFM de structures de C7TC4 fractionnées sur tamisage moléculaire.

Première ligne, images issues du fractionnement de C7TC4 dilué dans le pH acide avant injection sur la colonne de tamisage en tampon à pH 4.6. A, image obtenue à partir de J1, B et C, images obtenues à partir de J3.

Deuxième ligne, suite ligne 1, C et D, images obtenues à partir de la fraction J4. E, image obtenue à partir de la fraction J5.

Troisième ligne, images issues de C7TC4 dilué dans une solution à pH 6.8 avant injection sur la colonne à pH 4.6.

F, et G, images obtenues à partir des fractions K1 et K2, respectivement.

Quatrième ligne, images issues de C7TC4 dilué dans une solution à pH 7.9 avant injection sur la colonne à pH 4.6.

H, image obtenue à partir de la fraction L2. I et J, images obtenues à partir de la fraction L3.

Il ressort de ces images :

- Des structures segmentées isolées présentant des domaines de linéarité de 25 à 200 nm avec des points de branchement plus ou moins fréquents.
- Des réseaux plus ou moins étendus formés des structures segmentées précédentes.
- Des structures enchevêtrées de hauteurs hétérogènes.

La taille des objets observés ne correspond pas aux tailles déduites de la calibration du tamisage moléculaire. Le gradient de taille attendue en fonction des fractions de tamisage n'est pas non plus observé, même s'il existe une certaine corrélation. Ceci suggère soit que la calibration des colonnes n'est pas applicable aux structures de motif-i, soit que des événements de re-polymérisation ont lieu sur la surface et que les images ne représentent donc pas l'état en solution.

Le taux de branchement semble correspondre à une fréquence d'environ 1%, par rapport au nombre de jonctions linéaires, en considérant une longueur moyenne de segment d'environ 100 nm, soit environ 160 nucléotides en considérant une distance de 6 Å entre deux bases.

Nous observons donc, à nouveau, la superposition d'objets probablement formés en solution avec des objets qui se re-polymériseraient au contact de la surface. Cette conclusion est confirmée par le fractionnement après incubation à pH élevé qui montre des structures polymériques qui auraient dû être détruites à ces pH.

Ainsi, l'AFM permettrait de distinguer trois types d'éléments :

- les éléments de petites tailles pour lesquels nous observons des structures qui s'assemblent de manière caractéristique sur le mica, du fait de leur orientation suivant les mailles cristallines.
- Des éléments de grosses tailles pour lesquels nous observons les structures réelles formées en solution puis déposées sur la surface.
- L'observation de plages de croissance des réseaux semble suggérer un mécanisme où des motifs-i élémentaires s'assembleraient en solution pour former de petites structures supramoléculaires linéaires qui se colleraient sur la surface de mica et se coupleraient pour former des réseaux.

3.2. Discussions complémentaires

D'après les caractéristiques structurales de RMN et de cristallographie, nous nous attendons à ce que le motif-i soit une structure rigide en forme de bâtonnet. En effet, les oligonucléotides riches en C sont étirés avec une distance d'empilement de 6,1 Å afin de permettre l'intercalation des bases dans le motif-i. Cet étirement maximal, au-delà duquel une cassure du brin est attendue, ne permet pas une torsion des bases comme dans la double hélice d'ADN B. L'optimisation de l'empilement par la torsion et l'inclinaison des bases de la

double hélice permet de protéger les bases du solvant, sans contraindre le sucre qui serait énergétiquement plus coûteux à positionner. Dans le motif-i, qui peut être représenté comme deux échelles dont les barreaux seraient intercalés, c'est la configuration des sucres qui est modifiée (du 3' endo en 2') pour permettre de protéger les bases du solvant.

Les images obtenues en AFM, et ce uniquement à partir de la séquence C₇TC₄, montrent des structures totalement inattendues par rapport à ce que l'on connaît du motif-i, caractérisé comme un bâtonnet rigide. Nous observons sur les images AFM:

- des branchements à trois voies,
- des doubles épaisseurs qui suggèrent l'association de tétramères accolés latéralement,
- des franges de croissance en fractales qui semblent conditionnées par le réseau cristallin du mica,
- la formation de réseaux réguliers à longue distance.

Les longs fils courbés ne sont pas forcément incompatibles avec le modèle de polymérisation des motifs-i. Les structures courbées observées ne sont pas non plus en contradiction si on considère qu'elles peuvent être un ensemble de segments linéaires articulés. Néanmoins la dualité entre observation de structures formées en solution et sur le support rend difficile les conclusions quant à la structure des polymères en solution.

Nous pouvons émettre deux hypothèses :

- soit les structures observées sont des structures à base de motif-i, et il faut trouver les mécanismes qui peuvent expliquer la formation de structures branchées.
- soit les structures observées ne sont pas des structures de motifs-i. Cela pose alors le problème de définir un modèle compatible à la fois avec des extensions linéaires et des branchements à trois voies.

3.2.1 Auto-assemblage de motifs-i sous contrôle d'un substrat

La concentration des objets au voisinage de la surface peut favoriser la polymérisation de structures élémentaires trop diluées en solution pour pouvoir interagir entre elles. Des interactions objet-substrat pourraient contraindre l'orientation des objets tout en les concentrant et favoriser la formation de jonctions trois voies, du fait de la présence de brins pendants ou de défaut dans la formation des motifs-i.

La formation des structures branchées nécessite l'ouverture du motif-i. Le rôle joué par la base centrale, qui ne participe pas au motif-i, dans la forme des structures suggère son implication dans le processus de branchement. Deux schémas de jonction à trois voies sont proposés sur la Figure 83 à partir des séquences C₇ et C₇TC₄.

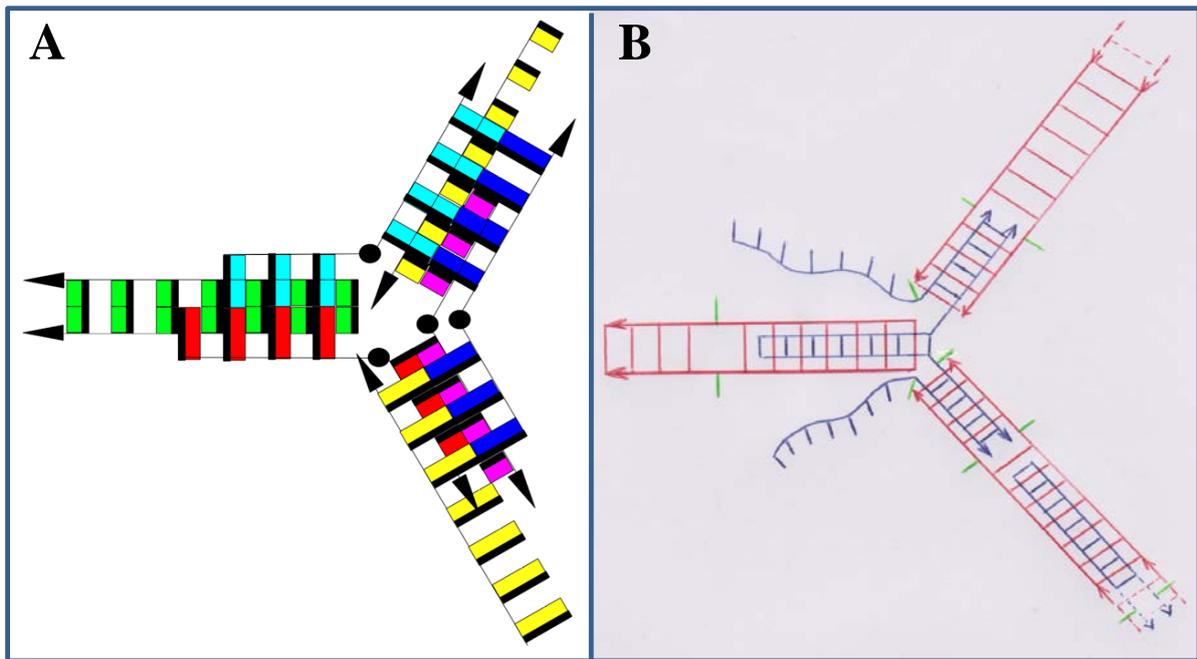


Figure 83 : schémas hypothétiques de jonction à trois voies.

A, jonction trois voies formée à partir d'oligonucléotides C_7 . Chaque brin C_7 est codé par une couleur différente, les lignes noires indiquant la face 5' des bases et les flèches l'extrémité 3' des oligonucléotides. L'association décalée des brins au sein des duplexes parallèles peut contribuer à la formation d'une grande diversité de jonctions branchées. B, jonction trois voies formée à partir d'oligonucléotides C_7TC_4 . Avec la séquence C_7TC_4 , les décalages ne sont possibles que si les résidus T sont ouverts, ce qui autorise à reproduire une jonction trois voies comme avec C_7 . On peut aussi imaginer que certains segments ne s'apparient pas, auquel cas on peut former une jonction trois voies comme présentée en B.

Dans le cas des oligonucléotides C_7 , l'assemblage des structures branchées peut entraîner des agrégats en trois dimensions et la précipitation des oligonucléotides. Dans le cas de C_7TC_4 , il est possible que du fait de leur apparente compacité structurale, l'énergie de formation des nœuds compense la perte des empilements de bases stabilisant les motifs-i. La formation de nœuds nécessite une forte courbure des chaînes d'ADN qui implique impérativement la rupture de la régularité du motif-i. Ce type de structures à trois voies est thermodynamiquement moins stable que des formes totalement appariées, mais elles peuvent être cinétiquement piégées pendant des temps très longs qui empêchent d'atteindre l'équilibre.

3.2.2 Rôle du résidu central

Nous avons observé diverses organisations de structures supramoléculaires selon la base centrale. Les structures en fils longs et peu branchés ont été observées uniquement avec les séquences C_7TC_4 et dans des conditions de concentrations et de force ionique particulières. L'image du « scorpion » révèle des hauteurs récurrentes de 1,7 nm et 3,4 nm, ce qui pourrait correspondre respectivement à la présence de 4 brins et de 8 brins d'ADN. Les structures de 3,4 nm de hauteur pourraient résulter de l'empilement parallèle de deux superstructures en motif-i. L'énergie nécessaire à cette association pourrait être médiée par une conformation ouverte des bases thymidines qui pourraient alors échanger des liaisons hydrogènes soit entre

elles, soit entre une base et le squelette en motif-i, en formant des liaisons inter-tétramères. Nous avons observé, avec la base centrale G, des amas en forme de boulettes de coton, très différentes des structures en nano-fils attendues. La base G centrale qui ne participe pas à la formation du motif-i, est forcément positionnée à l'extérieur de la structure, du fait de son encombrement qui ne permet son intercalation au sein des duplex. L'orientation de la guanine en dehors du motif-i pourrait donner lieu à des appariements aléatoires de type Watson-Crick entre les guanines et les cytidines libres de segments pendants qui conduiraient à des structures dénuées de toute régularité. Lorsque la base centrale est un A dans la structure C₇AC₄, des structures linéaires rigides et sans branchement ont été observées mais dans des conditions où une marche cristalline du mica a pu jouer un rôle de guide. L'absence de branchements étant observée uniquement dans ces conditions, nous pouvons réciproquement supposer que la présence de branchement est aussi d'une façon ou d'une autre sous le contrôle de la base centrale. Malgré les schémas présentés Figure 83, le mécanisme impliqué dans ce contrôle reste non élucidé, en particulier, il appartiendra de comprendre comment la base centrale peut influencer l'organisation des extrémités des motifs tétramériques.

4. Conclusion

Le motif-i constitue une brique de construction intéressante du fait de ses propriétés d'auto-organisation alliées à ses propriétés de réversibilité d'assemblage par modification de pH. Cette sensibilité des motifs i permet de contourner la nécessité de cycles de température pour l'assemblage sous contrôle thermodynamique de l'ADN B. Ainsi, l'utilisation du motif-i représente un outil intéressant pour notre objectif de création d'auto-assemblages fonctionnels incluant la présence de molécules fragiles telles que les protéines.

Les motifs-i s'assemblent en des structures supramoléculaires dont l'AFM a révélé diverses organisations. Néanmoins, l'assemblage de structures supramoléculaires à base de motif-i ne permet pas la création de géométries facilement contrôlable. Ils ne peuvent donc être utilisés seuls pour la nanofabrication de structures fonctionnelles. Nous pouvons envisager trois voies pour rendre leur utilisation pilotable.

Une voie consisterait à guider l'assemblage par une structuration ou un dopage de la surface. Nous pourrions nano-structurer la surface où y imprimer une fonctionnalisation chimique contrôlant l'assemblage, par greffage d'une base ou d'un composé qui interagirait avec le résidu central ou simplement par greffage ionique capable de modifier la concentration locale des briques. Une autre voie, cette fois pour contrôler l'assemblage en solution, pourrait impliquer la mise en œuvre d'un autre composant (protéine ou ADN) capable d'organiser la structuration des motifs-i. Ceci pourrait consister en des structures hybrides alliant des segments en ADN B et en motif-i. Alternativement, la fusion aux motifs-i de domaines peptidiques d'oligomérisation capables de s'assembler en structures diverses (cage, filament, membrane...) [107] peut être considérée. Une dernière voie serait de concevoir des protéines artificielles capables de reconnaître les structures en motif-i. Ces

structures pourraient être sélectionnées par « phage display », une technique de présentation à la surface de phages filamenteux qui constitue un outil puissant de sélection de domaines d'affinité. L'approche impliquerait la création et le criblage de banques combinatoires de molécules de reconnaissance telles les chaînes lourdes d'anticorps, ou de peptides ou de protéines [111].

CONCLUSIONS GENERALES ET PERSPECTIVES

L'objectif de mon travail était de concevoir des nanostructures qui utilisent les propriétés d'auto-assemblage des acides nucléiques et, permettent d'incorporer la fonctionnalité de protéines. Parmi les approches menant à la fonctionnalisation de squelette ADN, les PDNA qui, incluent sur une même molécule, un domaine nucléique et un domaine protéique par l'intermédiaire d'un adaptateur synthétique constituent un outil intéressant. En effet, le couplage covalent, faisant intervenir une cystéine unique, permet de s'affranchir des problèmes de dissociation qui peuvent être rencontrés dans les complexes protéine-ADN non covalents et d'obtenir des complexes homogènes avec une stœchiométrie et une régio-sélectivité garanties. Les PDNA, qui peuvent être purifiés à homogénéité, sont plus facilement caractérisables par diverses techniques physico-chimiques et sont utilisables dans la fabrication de nanostructures multi-étapes.

Le concept de PDNA, pouvoirs et limitations

La technique de couplage utilisée dans notre approche est restrictive car dépendante de la présence d'une cystéine unique, ce qui peut nécessiter une ingénierie, parfois délicate, lorsque les protéines comportent naturellement plusieurs de ces résidus. La restriction liée à l'utilisation du couplage par une fonction thiol peut être contournée en faisant appel soit à des techniques relativement délicates à mettre en œuvre comme le couplage par des intéines ou l'incorporation d'acides aminés non naturels [424], soit à un système plus universel où un anticorps couplé à l'ADN va reconnaître un tag spécifique attaché à la protéine d'intérêt. Dans ce cas, les PDNA pourraient inclure des mini-anticorps mono-chaîne appelés VHH (« variable heavy-chain of antibody »). Les VHH présentent les mêmes propriétés de reconnaissance que les anticorps communs (150-160 kDa), mais ils ont l'avantage d'être de petite taille (12 à 15 kDa) et d'être plus thermorésistants et stables envers les agents dénaturants. Les PDNA à base de VHH pourraient reconnaître le tag d'une protéine ou plusieurs tags si plusieurs protéines sont ciblées. Ce système permettrait de créer une boîte à outils présentant un nombre illimité de PDNA, capables de s'auto-assembler potentiellement à n'importe quelle protéine.

En effet, la sélection de domaines VHH capables de reconnaître une gamme pratiquement illimitée de cibles implique une approche par « phage-display » à partir de banques naïves naturelles ou synthétiques. L'approche est relativement universelle dans la mesure où l'introduction de la cystéine unique, nécessaire au couplage, peut être faite au niveau de la construction du squelette de la banque combinatoire et ne demandera donc pas d'autres opérations après l'étape de sélection par affinité sur la protéine cible.

Plateformes multifonctions

Les PDNA pourraient être organisés pour former des plateformes intégrées « lab on molecule » développées en vue d'applications en nanomédecine (ex. thérapie génique), en diagnostique et dans le domaine des interfaces homme-machine et des « humains augmentés ». D'un point de vue biotechnologique, elles permettraient le dessin de plateformes moléculaires pour le tri fonctionnel à haut débit utilisé pour l'ingénierie des protéines ou la découverte de médicaments. Ces plateformes pourraient impliquer différentes fonctions de reconnaissance, d'action, de visualisation et de transport organisées sur un support à base d'acides nucléiques, dont une illustration est présentée sur la Figure 84.

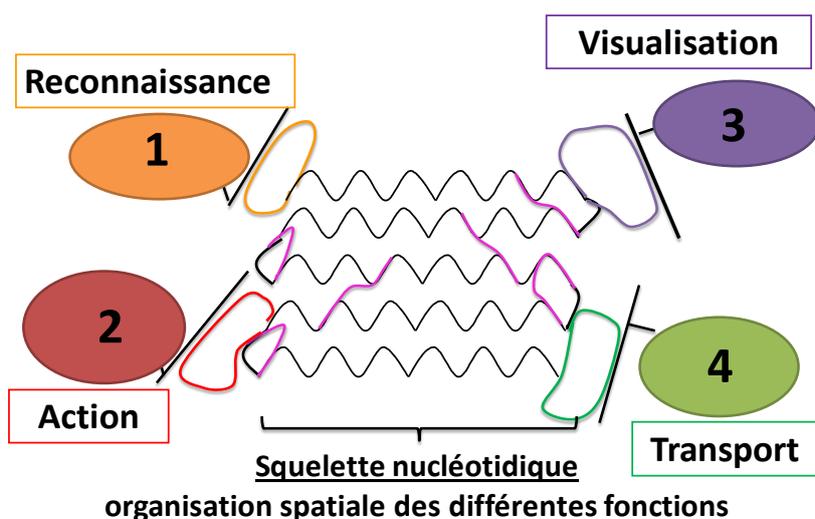


Figure 84 : schéma de concept de nanomachine protéonucléique à fonctionnalités multiples.

La connectique au sein de telles plateformes pourrait, du fait des dimensions nanométriques, être assurée par confinement diffusionnel et non plus par une microfluidique « cablée ». De plus, de tels systèmes peuvent avoir des comportements dynamiques (du fait de l'organisation spatiale des composants ou de leur déplacement sur des chemins ou circuits d'ADN) [425]. La fonction de reconnaissance de la plateforme peut concerner des séquences nucléotidiques ou des protéines en faisant intervenir des structures de doigt de zinc, des fragments d'anticorps ou des aptamères. Des actions de type catalyse, recrutement, signalisation, réparation... peuvent être ciblées. Nous pourrions concevoir des nanomachines impliquant des méganucléases capables de reconnaître des séquences d'ADN spécifiques et de les remplacer, supprimer ou modifier [426, 427] jouant, par exemple, le rôle de détecteur et d'inhibiteur de virus [428]. Couplées à des fonctions de visualisation par ancrage de composants fluorescents (boîte quantique, fluorophore) ou de composants de contraste (billes d'or) sur le squelette nucléotidique, les nanomachines pourraient être détectées et suivies dans le milieu. L'association de ces fonctions sur un même support nucléotidique nécessite d'optimiser la stratégie d'auto-assemblage.

La notion de plateforme analytique peut également être envisagée dans un système de bio-détection de type « chevet du malade ». L'approche peut aussi être utile pour aborder des

problèmes fondamentaux comme l'influence de l'organisation spatiale des différents modules de transport d'électrons. L'objectif est alors la construction de véritables « circuits moléculaires » où le chemin de transfert d'électrons se fait par sauts entre centres redox organisés de manière spatiale. Le contrôle du chemin est ainsi basé sur les distances entre centres (limitées par les effets de tunneling) et la chaîne de potentiel créée. Un exemple pourrait être la structuration de chaînes spatialement organisées de cytochromes b5 de potentiel redox rendus variables par l'utilisation de différents centres métalliques ou de porphyrines modifiées.

Approches alternatives d'assemblage

Nous avons vu que l'assemblage des structures à base d'ADN B réalisé au voisinage de l'équilibre thermodynamique est relativement simple et permet de générer, à volonté, diverses formes géométriques. Néanmoins ces conditions d'assemblage restent drastiques et les structures générées sont figées. A l'inverse, les techniques d'assemblage sous contrôle cinétique permettent d'assembler les séquences d'ADN dans des conditions plus douces pouvant présenter, de plus, un caractère dynamique par la mise en œuvre de mécanismes d'invasion de brins. Toutefois, ces systèmes sophistiqués restent complexes à maîtriser et ne permettent pas de générer n'importe quelle forme. Nous avons envisagé l'assemblage de structures à base des formes particulières de l'ADN, les motifs-i, dont la formation est dépendante du pH. Les motifs-i présentent l'avantage de générer des auto-assemblages, partiellement contrôlables par la base centrale, qui donnent des structures organisées différentes. Ce système simple, est facile à mettre en œuvre, mais les formes des structures restent limitées et la manière de coupler le motif-i à une protéine reste à définir. La solution permettant de contourner les limitations précédentes consisterait à combiner toutes les approches pour aller vers des systèmes hybrides qui associent ADN B et motif-i, les ADN B permettant alors d'organiser des structures dynamiques de motifs-i. Nous pourrions également envisager des hybrides avec des parties peptidiques qui permettraient d'organiser les ADN entre eux, telles que la glissière à leucine (leucine zipper).

Le dilemme structure-dynamique

Selon le principe d'incertitude de W. Heisenberg, on ne peut connaître simultanément la vitesse et la position de particules subatomiques, ainsi les mesures effectuées n'expriment pas des certitudes mais seulement des probabilités. De manière parallèle, la microscopie de force atomique à haute vitesse se heurte à la problématique de visualiser simultanément la dynamique réelle des objets et leur structure réelle. Cette difficulté s'accroît pour des molécules uniques hybrides impliquant divers composants moléculaires qui présentent probablement des propriétés dynamiques hétérogènes et des interactions avec la surface variables. Le choix de conditions favorables à la visualisation d'un composant n'est pas forcément adéquat pour les autres composants. Il reste délicat d'allier dynamisme et précision, et les comportements des objets sur la surface peuvent ne pas correspondre à la réalité en

solution, du fait des perturbations dues aux interactions avec la surface. La SPRi constitue une technique complémentaire qui donne l'information sur la dynamique globale d'une population d'objets. La RMN apporte des informations complémentaires de structure et de dynamique, avec l'accès aux facteurs dynamiques mais est limitée à l'étude d'objets relativement petits. La microscopie confocale apporte des informations à l'échelle de la molécule unique avec une résolution ne permettant pas d'analyser des architectures plus petites que 250 nm. La cryomicroscopie électronique donne accès aux informations structurales mais ne permet pas d'accéder à la dynamique des objets qui sont figés dans la glace.

L'ensemble de ces techniques reste nécessaire pour caractériser les nanostructures fabriquées. Parmi elles, l'AFM-HS, constitue une approche originale conciliant résolution nanométrique et dynamique.

ANNEXES

Annexe 1

Paramètres thermodynamiques des séquences d'ADN.

ΔG (kcal/mol) pour la formation d'hétéro-dimères entre les demi-séquences des briques d'ADN.

Demi-séquence	A	A'	B	B'	a	a'	b	b'	M	m
A	*	*	*	*	39	*	8	*	*	*
A'		7.5	*	10.5	*	32	8	8	*	*
B			8	*	8	8	37	*	*	*
B'				7.5	*	8	*	28	*	*
a					*	*	*	*	*	8
a'						7.5	*	10.5	*	*
b							8	*	*	*
b'								7.5	*	*
M									*	*
m										9.3

ΔG (kcal/mol) pour la formation d'hétéro- ou d'homo-dimères entre les briques complètes.

Tiles	AA'	BB'	ba'	ab'	Mb'	am	Ma'	bm
AA'	7.5	10.5	32	39	8	39	32	8.4
BB'		8	37	30	30	8	14	37
ba'			13	10.5	10.5	*	13.4	8
ab'				7.5	10.5	*	10.5	*
Mb'					8	*	10.5	*
am						11	*	9.3
Ma'							13.4	*
bm							*	9.3

Les calculs sont réalisés avec le calculateur en ligne en considérant les concentrations de 150 mM NaCl et de 1 μ M ADN (<http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer>)

Annexe 2

Appendice 5 : article 6

Polarimetric surface plasmon resonance imaging biosensor

Duval A., Laisné A., Pompon D.,
Held S., Bellemain A., Moreau J. & Canva M.

Optics Letters (2009), 34(23): 3634-3636

Polarimetric surface plasmon resonance imaging biosensor

Aurélien Duval,^{1,*} Aude Laisné,² Denis Pompon,² Sylvain Held,³ Alain Bellemain,¹ Julien Moreau,¹ and Michael Canva¹

¹Laboratoire Charles Fabry de l'Institut d'Optique, CNRS, Université Paris Sud, Campus Polytechnique RD 128, 91127 Palaiseau CEDEX, France

²Centre de Génétique Moléculaire, CNRS, Avenue de la Terrasse, 91190 Gif-sur-Yvette, France

³Institut d'Electronique Fondamentale, CNRS, Université Paris Sud, Bâtiment 220, 91405 Orsay, France

*Corresponding author: aurelien.duval@institutoptique.fr

Received September 28, 2009; accepted October 13, 2009;
posted November 2, 2009 (Doc. ID 117883); published November 19, 2009

We report the realization of a polarimetric surface plasmon resonance imaging system capable of dynamically resolving a change in the optical anisotropy of biochemical films. Anisotropies as small as 10^{-3} refractive index unit on nanometer-thick samples can be resolved. As an example, we present here the dynamical anisotropy obtained by the electrical patterning of a film consisting of a self-assembled monolayer deposited on gold, covered with a phospholipid hemimembrane. © 2009 Optical Society of America
OCIS codes: 240.6680, 310.6860, 110.5405.

Surface plasmon resonance (SPR) biosensors are being increasingly used in laboratories as sensitive biochemical transducers for several applications such as genetic diagnostics [1], kinetics of interactions analysis [2], or food analysis [3], for example. They rely on the coupling between oscillating bound-free electrons present in a metallic thin-film (the surface plasmons) and an incident TM-polarized light wave. A modification in the refractive index of the surrounding dielectric medium will change the coupling between the plasmon waves and the incident light. Monitoring the reflectivity of such a system enables the detection of minute changes in the refractive index of the surrounding medium. Surface plasmon resonance imaging (SPRI) allows parallel monitoring of multiple interactions dynamically [4].

The polarimetric surface plasmon resonance imaging (P-SPRI) system demonstrated here is based on the Kretschmann–Raether configuration [5], using two orthogonal SPRI arms both capable of capturing TM-polarized images of the biosensor's surface, following two orthogonal planes of incidence, as shown in Fig. 1. Each sensor arm can act as a completely independent SPRI system. Compared to approaches based on coupled plasmon-waveguide resonance spectroscopy [6], or more classical ellipsometric methods [7], the P-SPRI system developed is capable of dynamical measurements on label-free films.

The reflectivity variations in the image measured in each direction (x and y) of the surface of the biosensor, shown in Eq. (1), are proportional to the variations in the effective refractive indices (Δn_x and Δn_y) probed by the propagating plasmon wave, as well as the thickness (Δd) of the biological thin-film (typically 1–50 nm thick) deposited on the biosensor surface,

$$\Delta R_k \propto \Delta n_k \Delta d, \quad k = \{x, y\}. \quad (1)$$

Variations in the relative mean anisotropy Δn of the covering dielectric medium are measured by subtracting the signal of one arm from the other. In bio-

chemical systems consisting of long chains, like the self-assembled monolayer (SAM), the birefringence is given by the intrinsic anisotropy of the molecule and the order parameter of the assembly. The latter can be characterized by its azimuth ϕ and zenith θ . An (arbitrary) 1-nm-thick equivalent layer of dense birefringent rod-shaped molecules with an intrinsic anisotropy of ≈ 0.1 ($n_{iso} \approx 1.45$ [8]) tethered to gold and oriented parallel to the surface ($\theta = 90^\circ$), and one of the system's axes ($\phi = 0^\circ$) will yield a maximal relative mean effective anisotropy of $\Delta n \approx \pm 0.1$ as measured by our system. Taking for reference a layer oriented along the surface normal ($\theta = 0^\circ$), the rotation of the molecules from “up” to “laying flat” will give rise to a simulated maximal reflectivity difference $\Delta R_{aniso} = \Delta R_x - \Delta R_y$ of about $\pm 3\%$. In real applications, the signal measured will be reduced owing to smaller θ variations and an arbitrary azimuth.

For each orthogonal sensor arm, the incident light (an LED at 670 nm, FWHM of 25 nm) is first collimated, TM-polarized, and reflected off a motorized rotating mirror before being coupled to the metallic

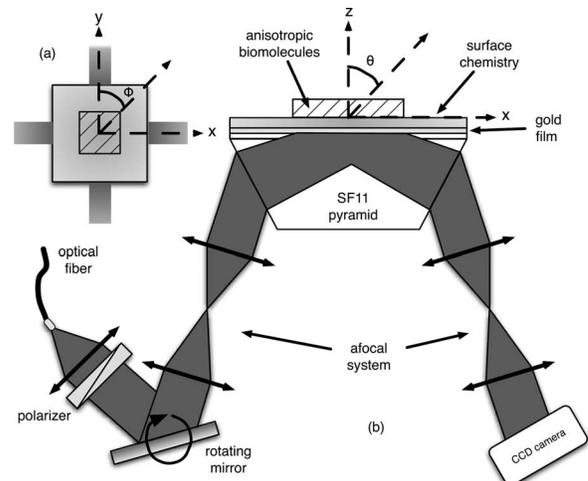


Fig. 1. Schematic representation of the P-SPRI system. (a) Top view. (b) Side view.

sensor layer through the use of an afocal system and the SF11 pyramid. On the opposite face of the pyramid, the reflected light passes through a similar afocal system and is collected by a CCD camera (Cooke Pixelfly QE). Using an afocal optical system (magnification of $1\times$) and a rotating mirror limits the angles of incidence to a $\pm 5^\circ$ region around the central angle (53° with respect to the base of the pyramid), hence limiting the range of measurable bulk refractive indices from 1.294 to 1.386. For dielectric layers ($n = 1.45$) deposited on gold and immersed in water ($n = 1.33$), it restricts the measurements to thin films (< 50 nm).

The combined reading of both arms is shot-noise limited as in more traditional SPRI systems. We determined that our P-SPRI setup yields a noise of less than $10^{-2}\%$ using spatial averaging ($400\ \mu\text{m} \times 400\ \mu\text{m}$), while keeping dynamical capabilities (1 s integration for 16 images averaged).

The anisotropy figure measured for each pixel of the resulting differential SPR images can only make sense if the P-SPRI system is calibrated so that isotropic samples do not appear as anisotropic. To demonstrate the normalization technique used, we carried out an experiment using a SF11 glass-slide covered with gold squares ($1\ \text{mm}^2$). The squares were deposited using lift-off UV lithography and vacuum evaporation techniques. They are composed of a 7-nm-thick Ti adhesion layer and a 47-nm-thick Au sensing layer. Index-matching oil was deposited between the base of the pyramid and the biosensor. The system was finally covered with a flow-cell.

At a fixed incidence angle (the maximum slope of the angular reflectivity curve), a dynamical SPRI measurement was performed simultaneously on nine gold squares (using $400\ \mu\text{m} \times 400\ \mu\text{m}$ spatial averaging) under a constant flow of water. After approximately 2 min, water was replaced with a mixture of 1% volume per volume (v/v) glycerol in water. After 6 min, water was injected and circulated until reaching a steady state. A second mixture of 1.5% (v/v) glycerol was injected after approximately 10 min.

Figures 2(a) and 2(b) show the reflectivity variations (ΔR_x and ΔR_y) induced by the injection of glycerol, averaged over nine gold squares. They correspond to bulk refractive index increases of, respectively, $\Delta n = 1.4 \times 10^{-3}$ refractive index unit (RIU) for 1% glycerol and $\Delta n = 2.1 \times 10^{-3}$ RIU for 1.5% glycerol. The resulting signal is equivalent to a thin-film deposition as in biosensing applications. Using the extended Rouard method [9], we determined that the response to 1% glycerol would give a signal similar to the deposition of about 2 nm of a dielectric material ($n = 1.45$) on the gold surface.

To account for small differences in reflectivity slopes between the two arms, reflectivity variations are normalized by the mean 1% glycerol plateau. The dynamical SPR anisotropy signal (ΔR_{aniso}) can then be determined by subtracting arms x and y signals. The result for our glycerol experiment is depicted in Fig. 2(c). As expected, no artifactual anisotropy is found outside the injection and washing regions (materialized by dashed lines in the figure). In this

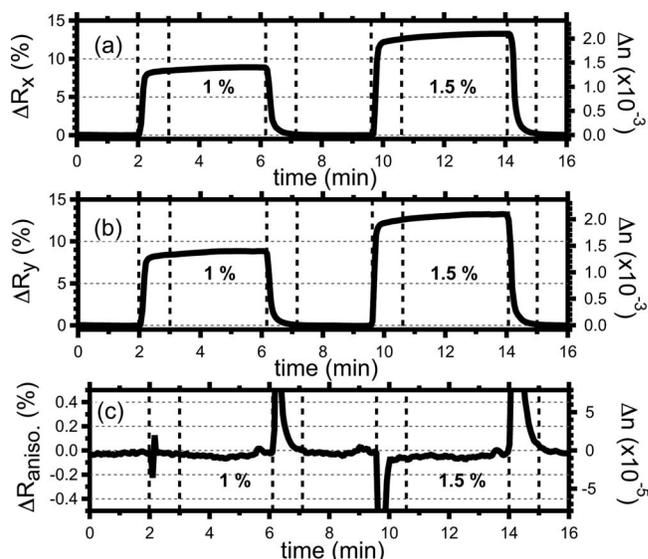


Fig. 2. Corrected mean dynamical reflectivity variations induced by the injections of 1% and 1.5% glycerol in water. (a) Arm x . (b) Arm y . (c) “Anisotropy” of the bulk liquid. Dashed lines represent fluidic transitions. Right axes, corresponding bulk Δn .

particular experiment, the residual noise is measured below $\Delta R_{aniso} = 10^{-2}\%$, which would translate into a minimal detectable bulk anisotropy of $\Delta n_{min} = 10^{-5}$ RIU or into a minimal anisotropy of 10^{-3} RIU for 1 nm of biomolecules ($n_{iso} = 1.45$) tethered onto the gold surface. Using our model of a dense layer of birefringent rod-shaped molecules with an intrinsic anisotropy of ≈ 0.1 , angular variations as small as 5° of a 1-nm-thick equivalent layer can be discriminated.

To illustrate applications of such a system, the anisotropy resulting from the patterning of a surface consisting in a SAM and a phospholipid hemimembrane was determined dynamically using our P-SPRI setup. The biochip was divided into three different regions depicted in Fig. 3(b). Regions A and B were connected to a potentiostat capable of applying an electrical potential difference between these electrodes and the buffer solution. The REF region was acting as a reference for the electric cycles, and region C was used as a control and thus isolated from the surrounding electrodes.

Prior to the experiment, a layer of alkanethiols was allowed to self-assemble on the gold surface, covering all regions of the biochip. A fusion of phospholipid liposomes was then realized on the SAM, resulting in a supported hemimembrane [10]. During the experiment, several electrical cycles were realized between the A and B regions. Each cycle consisted into a 1 min square profile, applying a potential of ± 1.5 V to the region. The mean dynamical SPR anisotropy (ΔR_{aniso}) signal of the regions was captured using the P-SPRI system and is given in Fig. 3(a), showing parallel measurements on regions A (dots) and C (line). Spatial averaging (area of $\approx 400\ \mu\text{m} \times 400\ \mu\text{m}$) was used again in order to reduce the noise, and the signal resulting from the REF region was subtracted from the curves.

It can be clearly observed that region C stays isotropic throughout the experiment with a residual av-

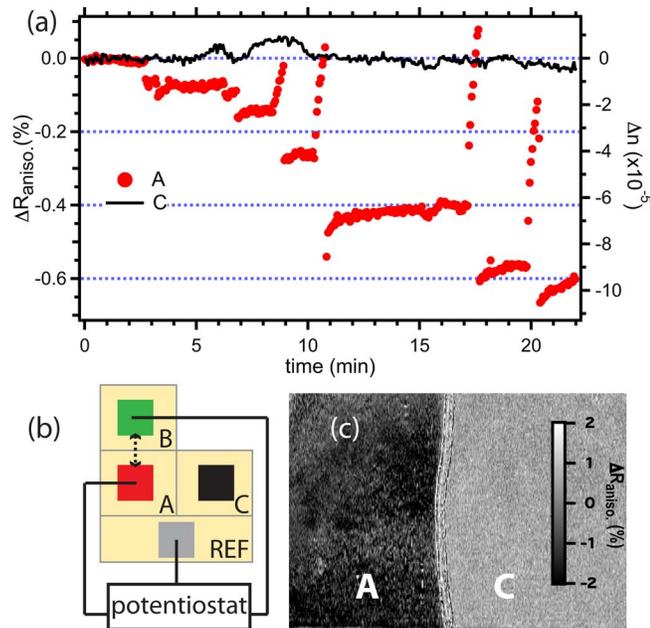


Fig. 3. (Color online) (a) Dynamical P-SPRI signal of a SAM-supported hemimembrane system. Dots, anisotropic response of electrode A. Line, response of electrode C (control). (b) Diagram of the biochip areas. (c) Image of the differential SPR anisotropy observed.

eraged SPR anisotropy of $<10^{-2}\%$ (or a bulk anisotropy of less than 10^{-5}), while region A becomes anisotropic after the first cycle of the electric potential and gains anisotropy as more cycles are applied to the region. The sharp drops observed between steady state anisotropy levels are due to the electrical cycles themselves. As the potential applied changes polarity, the change in the orientation of the surface-bound molecules is believed to be maximal. At the end of the experiment, region A exhibits an averaged SPR anisotropy of -0.6% , yielding a bulk anisotropy of $\Delta n = -9 \times 10^{-5}$ RIU or approximately 10^{-2} RIU for 1 nm of a dielectric material ($n=1.45$) deposited.

The image of Fig. 3(c) is obtained by subtracting the differential SPR images at $t=1.5$ and 13.5 min.

We can clearly observe that the region where electric potential cycles were applied gained a pattern, while the control region is still homogeneous at the end of the experiment. We believe that this patterning effect is responsible to some extent for the anisotropic signal observed.

A P-SPRI biosensor capable of resolving the dynamical anisotropy of biochemical films has been developed. The system is found to be limited by photon noise, while yielding a reading noise of less than $10^{-2}\%$, enabling the characterization of dynamical anisotropy of samples with a sensitivity on the order of 10^{-3} RIU on nanometric biological layers. A dynamical anisotropy measurement has been carried out on a biochemical film patterned by cycles of applied electrical potential, leading to a SPR signal of -0.6% or 10^{-2} RIU for 1 nm of a dielectric material ($n=1.45$) deposited.

The authors acknowledge the support of the National French Research Agency (ANR) through the National Program for Nanosciences and Nanotechnologies (PNANO) of the European Network of Excellence Photonics4Life (FP7) and the help of A. Aide on the mechanical design.

References

1. J. Hottin, J. Moreau, G. Roger, J. Spadavecchia, M.-C. Millot, and M. Goossens, *Plasmonics* **2**, 201 (2007).
2. B. P. Nelson, T. E. Grimsrud, M. R. Liles, R. M. Goodmab, and R. M. Corn, *Anal. Chem.* **73**, 1 (2001).
3. D. R. Shankaran, K. V. Gobi, and M. Miura, *Sens. Actuators B* **121**, 158 (2007).
4. M. Piliarik, H. Vaisocherov, and J. Homola, *Biosens. Bioelectron.* **20**, 2104 (2005).
5. E. Kretschmann and H. Raether, *Z. Naturforsch.* **A 23A**, 2135 (1968).
6. Z. Salamon and G. Tollin, *Biophys. J.* **80**, 1557 (2001).
7. M. Schubert, *Thin Solid Films* **313-314**, 323 (1998).
8. R. Naumann, S. M. Schiller, F. Giess, B. Grohe, K. B. Hartman, I. Kärcher, I. Köper, J. Lübber, K. Vasilev, and W. Knoll, *Langmuir* **19**, 5435 (2003).
9. P. Lecaruyer, E. Maillart, M. Canva, and J. Rolland, *Appl. Opt.* **45**, 8419 (2006).
10. W. Boireau, S. Bombard, M. Sari, and D. Pompon, *Biotechnol. Bioeng.* **77**, 225 (2002).

BIBLIOGRAPHIE

- [1] P. Chopra, A. Kamma, Engineering life through Synthetic Biology, *In Silico Biol*, 6 (2006) 401-410.
- [2] S. Leduc, *La biologie synthétique*, A. Poinat Editeur, Paris, 1912.
- [3] W. Szybalski, A. Skalka, Nobel prizes and restriction enzymes, *Gene*, 4 (1978) 181-182.
- [4] B. Hobom, Gene surgery: on the threshold of synthetic biology, *Med Klin*, 75 (1980) 834-841.
- [5] R.L. Rawis, 'Synthetic Biology' Makes Its Debut, *Chem. Eng. News*, 78 (2000) 49-53.
- [6] S.A. Benner, Redesigning life. Organic chemistry and the evolution of protein, *Chimia*, 41 (1987) 142-148.
- [7] S.A. Benner, Synthetic biology: Act natural, *Nature*, 421 (2003) 118.
- [8] J.W. Szostak, D.P. Bartel, P.L. Luisi, Synthesizing life, *Nature*, 409 (2001) 387-390.
- [9] R. Breslow, Biomimetic chemistry: biology as an inspiration, *J Biol Chem*, 284 (2009) 1337-1342.
- [10] S.A. Benner, A.M. Sismour, Synthetic biology, *Nat Rev Genet*, 6 (2005) 533-543.
- [11] W.W. Gibbs, Synthetic life, *Scientific American*, 290 (2004) 74-81.
- [12] D.M. Widmaier, D. Tullman-Ercek, E.A. Mirsky, R. Hill, S. Govindarajan, J. Minshull, C.A. Voigt, Engineering the Salmonella type III secretion system to export spider silk monomers, *Mol Syst Biol*, 5 (2009) 309.
- [13] D. Endy, Foundations for engineering biology, *Nature*, 438 (2005) 449-453.
- [14] E.H. Bromley, K. Channon, E. Moutevelis, D.N. Woolfson, Peptide and protein building blocks for synthetic biology: from programming biomolecules to self-organized biomolecular systems, *ACS Chem Biol*, 3 (2008) 38-50.
- [15] M. Simard, D. Su, J.D. Wuest, Use of hydrogen-bonds to control molecular aggregation-self-assembly of 3-dimensional networks with large chambers, *J. Am. Chem. Soc.*, 113 (1991) 4696-4698.
- [16] N. Paul, G.F. Joyce, Minimal self-replicating systems, *Curr Opin Chem Biol*, 8 (2004) 634-639.
- [17] C. Lartigue, J.I. Glass, N. Alperovich, R. Pieper, P.P. Parmar, C.A. Hutchison, 3rd, H.O. Smith, J.C. Venter, Genome transplantation in bacteria: changing one species to another, *Science (New York, N.Y.)*, 317 (2007) 632-638.
- [18] J.W. Chin, Programming and engineering biological networks, *Curr Opin Struct Biol*, 16 (2006) 551-556.
- [19] T. Pawson, R. Linding, Synthetic modular systems--reverse engineering of signal transduction, *FEBS Lett*, 579 (2005) 1808-1814.
- [20] B. Bensaude-Vincent, Biomimetic Chemistry and Synthetic Biology: A Two-way Traffic Across the Borders, *HYLE – International Journal for Philosophy of Chemistry*, 15 (2009) 31-46.
- [21] P. Ball, Life's lessons in design, *Nature*, 409 (2001) 413-416.
- [22] Y. Bar-Cohen, Biomimetics--using nature to inspire human innovation, *Bioinspir Biomim*, 1 (2006) P1-P12.
- [23] B. Bhushan, Biomimetics: lessons from nature--an overview, *Philos Transact A Math Phys Eng Sci*, 367 (2009) 1445-1486.
- [24] E. Andrianantoandro, S. Basu, D.K. Karig, R. Weiss, Synthetic biology: new engineering rules for an emerging discipline, *Mol Syst Biol*, 2 (2006) 2006 0028.
- [25] A. Maynard, Synthetic biology and nanotechnology, in: *2020 Science, A clear perspective on emerging science and technology*, 2008.
- [26] W. Merlijn van Spengen, K. Sueko, Y. Sugawara, B. Sullivan, Y. Sun, P. Swanson, N.S. Tambe, T.M. Y.C., K.L. Turner, G.W. Tyndall, P. Vettiger, G. Wei, D. Wendell, D.J. Young,

- B. Ziaie, C.A. Zorman, J. Zoval, P.K. Zysset, Springer handbook of Nanotechnology, 2nd ed., Springer, Berlin Heidelberg New York, 2007.
- [27] E.T. Kool, Synthetic biology, in, 2010.
- [28] N.C. Seeman, S.P. Lukeman, Nucleic acid nanostructures: bottom-up control of geometry on the nanoscale Reports on Progress in Physics 68 (2005) 237.
- [29] G.M. Whitesides, M. Boncheva, Beyond molecules: self-assembly of mesoscopic and macroscopic components, Proc Natl Acad Sci U S A, 99 (2002) 4769-4774.
- [30] N.C. Seeman, The challenge of structural control on the nanoscale : bottom-up self-assembly of nucleic acids in 3D, International journal of nanotechnology 2(2005) 348-370.
- [31] H. Robson Marsden, A. Kros, Self-assembly of coiled coils in synthetic biology: inspiration and progress, Angewandte Chemie (International ed, 49 (2010) 2988-3005.
- [32] V.M. Rotello, Inspiration (and perspiration) from biology, ACS Nano, 2 (2008) 4-6.
- [33] N. Huebsch, D.J. Mooney, Inspiration and application in the evolution of biomaterials, Nature, 462 (2009) 426-432.
- [34] G.M. Whitesides, Self assembly and nanotechnology, Proceedings of SPIE, the International Society for Optical Engineering 2716 (1996) 307-308.
- [35] E.D.P. De Robertis, E.M.F. De Robertis, Biologie cellulaire et moléculaire, Les Presses de l'université de Laval, Québec, 1983.
- [36] T.D. Pollard, W.C. Earnshaw, Biologie cellulaire, Elsevier SAS., Paris, 2004.
- [37] J.M. Lehn, Chimie des interactions moléculaires, in: C.d. France (Ed.) Cours et travaux du Collège de France, Collège de France, Paris, 2010, pp. 193-211.
- [38] C.B. Anfinsen, Principles that govern the folding of protein chains, Science (New York, N.Y., 181 (1973) 223-230.
- [39] H. Fraenkel-Conrat, R.C. Williams, Reconstitution of Active Tobacco Mosaic Virus from Its Inactive Protein and Nucleic Acid Components, Proc Natl Acad Sci U S A, 41 (1955) 690-698.
- [40] D.L. Caspar, Assembly and Stability of the Tobacco Mosaic Virus Particle, Adv Protein Chem, 18 (1963) 37-121.
- [41] D.L. Caspar, K. Namba, Switching in the self-assembly of tobacco mosaic virus, Adv Biophys, 26 (1990) 157-185.
- [42] A. Klug, The tobacco mosaic virus particle: structure and assembly, Philos Trans R Soc Lond B Biol Sci, 354 (1999) 531-535.
- [43] J.S. Lindsey, Self-assembly in synthetic routes to molecular devices. Biological Principles and chemical perspectives: a review, New J. Chem, 15 (1990) 153-180.
- [44] R.L. Thompson, N.S. Goel, A simulation of T4 bacteriophage assembly and operation, Biosystems, 18 (1985) 23-45.
- [45] J.W. Steed, J.L. Atwood, Supramolecular chemistry, 2nd ed., Chichester 2000.
- [46] S. Seal, Functional Nanostructures, processing, characterization, and applications, Springer Science, New York, 2008.
- [47] Y.S. Lee, Self-assembly and Nanotechnology, A Force Balance Approach, John Wiley & Sons, Inc., Hoboken, 2008.
- [48] A.W. Wilber, J.P. Doye, A.A. Louis, E.G. Noya, M.A. Miller, P. Wong, Reversible self-assembly of patchy particles into monodisperse icosahedral clusters, J Chem Phys, 127 (2007) 085106.
- [49] A. Zlotnick, Distinguishing reversible from irreversible virus capsid assembly, J Mol Biol, 366 (2007) 14-18.
- [50] S. Toksöz, M.O. Guler, Self-assembled petpidic nanostructures, NanoToday, 4 (2009) 458-469.
- [51] C. Levinthal, Are there pathways for protein folding?, J Chem Phys, 65 (1968) 44-45.
- [52] B.H. Zimm, J.K. Bragg, Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chains, J. Chem. Phys., 31 (1959) 526-535.

- [53] D.B. Wetlaufer, Nucleation, rapid folding, and globular intrachain regions in proteins, *Proc Natl Acad Sci U S A*, 70 (1973) 697-701.
- [54] M. Karplus, D.L. Weaver, Protein folding dynamics: the diffusion-collision model and experimental data., *Protein Sci*, 3 (1992) 650-668.
- [55] R.L. Baldwin, Intermediates in protein folding reactions and the mechanism of protein folding, *Annu Rev Biochem*, 44 (1975) 453-475.
- [56] G.E. Schulz, Structural rules for globular proteins. , *Angew. Chem. Int. Ed. Engl.*, 16 (1977) 23-32.
- [57] D.B. Wetlaufer, Folding of protein fragments, *Adv Protein Chem*, 34 (1981) 61-92.
- [58] C. Chothia, Principles that determine the structure of proteins, *Annu Rev Biochem*, 53 (1984) 537-572.
- [59] M. Levitt, A. Warshel, Computer simulation of protein folding, *Nature*, 253 (1975) 694-698.
- [60] K.A. Dill, K.M. Fiebig, H.S. Chan, Cooperativity in protein-folding kinetics, *Proc Natl Acad Sci U S A*, 90 (1993) 1942-1946.
- [61] E.M. Boczko, C.L. Brooks, 3rd, First-principles calculation of the folding free energy of a three-helix bundle protein, *Science (New York, N.Y)*, 269 (1995) 393-396.
- [62] A.E. Garcia, J.N. Onuchic, Folding a protein in a computer: an atomic description of the folding/unfolding of protein A, *Proc Natl Acad Sci U S A*, 100 (2003) 13898-13903.
- [63] B. Kuhlman, G. Dantas, G.C. Ireton, G. Varani, B.L. Stoddard, D. Baker, Design of a novel globular protein fold with atomic-level accuracy, *Science (New York, N.Y)*, 302 (2003) 1364-1368.
- [64] G. Cesarini, M. Gimona, M. Sudol, M. Yaffe, *Modular Protein Domains*, Wiley-VCH, Weinheim, 2005.
- [65] L. Pauling, R.B. Corey, H.R. Branson, The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain, *Proceedings of the National Academy of Sciences of the United States of America*, 37 (1951) 205-211.
- [66] L. Pauling, R.B. Corey, Configuration of polypeptide chains, *Nature*, 168 (1951) 550-551.
- [67] A.D. Moore, A.K. Bjorklund, D. Ekman, E. Bornberg-Bauer, A. Elofsson, Arrangements in the modular evolution of proteins, *Trends Biochem Sci*, 33 (2008) 444-451.
- [68] A. Heyman, Y. Barak, J. Caspi, D.B. Wilson, A. Altman, E.A. Bayer, O. Shoseyov, Multiple display of catalytic modules on a protein scaffold: nano-fabrication of enzyme particles, *J Biotechnol*, 131 (2007) 433-439.
- [69] R. McDaniel, S. Ebert-Khosla, D.A. Hopwood, C. Khosla, Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits, *Nature*, 375 (1995) 549-554.
- [70] F. Parmeggiani, R. Pellarin, A.P. Larsen, G. Varadamsetty, M.T. Stumpp, O. Zerbe, A. Caflisch, A. Pluckthun, Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core, *J Mol Biol*, 376 (2008) 1282-1304.
- [71] T. Pawson, P. Nash, Assembly of cell regulatory systems through protein interaction domains, *Science (New York, N.Y)*, 300 (2003) 445-452.
- [72] G.A. Petsko, D. Ringe, C. Sanlaville, D. Charnot-Bensimon, *Structure et fonction des protéines*, 2008.
- [73] M. Levitt, C. Chothia, Structural patterns in globular proteins, *Nature*, 261 (1976) 552-558.
- [74] J.M. Thornton, C.A. Orengo, A.E. Todd, F.M. Pearl, Protein folds, functions and evolution, *J Mol Biol*, 293 (1999) 333-342.
- [75] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, SCOP: a structural classification of proteins database for the investigation of sequences and structures, *J Mol Biol*, 247 (1995) 536-540.

- [76] C.A. Orengo, A.D. Michie, S. Jones, D.T. Jones, M.B. Swindells, J.M. Thornton, CATH-a hierarchic classification of protein domain structures, *Structure*, 5 (1997) 1093-1108.
- [77] T.J. Hubbard, B. Ailey, S.E. Brenner, A.G. Murzin, C. Chothia, SCOP: a Structural Classification of Proteins database, *Nucleic acids research*, 27 (1999) 254-256.
- [78] D. Eisenberg, The discovery of the alpha-helix and beta-sheet, the principal structural features of proteins, *Proc Natl Acad Sci U S A*, 100 (2003) 11207-11210.
- [79] J.M. Hemmingsen, K.M. Gernert, J.S. Richardson, D.C. Richardson, The tyrosine corner: a feature of most Greek key beta-barrel proteins, *Protein Sci*, 3 (1994) 1927-1937.
- [80] N. Kurochkina, Helix-helix interactions and their impact on protein motifs and assemblies, *J Theor Biol*, 264 (2010) 585-592.
- [81] cmgm.stanford.edu, Pleated beta-sheets in, 1999.
- [82] http://fr.wikipedia.org/w/index.php?title=H%C3%A9lice_alpha&action=history, Hélice alpha, in, 2011.
- [83] campbell.edu, Structure, Pauling rules for polypeptide chains, in.
- [84] D. Salvatore, Interactions à l'échelle moléculaire et mécanismes d'assemblage de deux protéines alimentaires : l' α -lactalbumine et le lysozyme., in, Université Européenne de Bretagne, Rennes, 2011.
- [85] M. Nigen, V. Le Tilly, T. Croguennec, D. Drouin-Kucma, S. Bouhallab, Molecular interaction between apo or holo alpha-lactalbumin and lysozyme: formation of heterodimers as assessed by fluorescence measurements, *Biochim Biophys Acta*, 1794 (2009) 709-715.
- [86] M. Nigen, C. Gaillard, T. Croguennec, M.N. Madec, S. Bouhallab, Dynamic and supramolecular organisation of alpha-lactalbumin/lysozyme microspheres: A microscopic study, *Biophys Chem*, 146 (2010) 30-35.
- [87] B. Leblanc, Biochimie des protéines, in, Université de Sherbrooke, 2004.
- [88] <http://www.facbio.com>, Dynamique de la polymérisation des microtubules, in, 2008.
- [89] M. Reches, E. Gazit, Molecular Self-Assembly of Peptide Nanostructures: Mechanism of Association and Potential Uses, *Current Nanoscience*, (2006) 105-111.
- [90] S. Zhang, Emerging biological materials through molecular self-assembly, *Biotechnol Adv*, 20 (2002) 321-339.
- [91] S. Zhang, D.M. Marini, W. Hwang, S. Santoso, Design of nanostructured biological materials through self-assembly of peptides and proteins, *Curr Opin Chem Biol*, 6 (2002) 865-871.
- [92] E. Gazit, Self-assembled peptide nanostructures: the design of molecular building blocks and their technological utilization, *Chem Soc Rev*, 36 (2007) 1263-1269.
- [93] C.J. Tsai, J. Zheng, D. Zanuy, N. Haspel, H. Wolfson, C. Aleman, R. Nussinov, Principles of nanostructure design with protein building blocks, *Proteins*, 68 (2007) 1-12.
- [94] D. Zanuy, R. Nussinov, C. Aleman, From peptide-based material science to protein fibrils: discipline convergence in nanobiology, *Phys Biol*, 3 (2006) S80-90.
- [95] A.G. Ogston, Chemistry of proteins, peptides, and amino acids, *Annu Rev Biochem*, 24 (1955) 181-206.
- [96] C.T. Armstrong, A.L. Boyle, E.H. Bromley, Z.N. Mahmoud, L. Smith, A.R. Thomson, D.N. Woolfson, Rational design of peptide-based building blocks for nanoscience and synthetic biology, *Faraday Discuss*, 143 (2009) 305-317; discussion 359-372.
- [97] K. Channon, E.H. Bromley, D.N. Woolfson, Synthetic biology through biomolecular design and engineering, *Current Opinion in Structural Biology*, 18 (2008) 491-498.
- [98] R.V. Ulijn, A.M. Smith, Designing peptide based nanomaterials, *Chem Soc Rev*, 37 (2008) 664-675.
- [99] B. Brodsky, G. Thiagarajan, B. Madhan, K. Kar, Triple-helical peptides: an approach to collagen conformation, stability, and self-association, *Biopolymers*, 89 (2008) 345-353.
- [100] V. Gauba, J.D. Hartgerink, Synthetic collagen heterotrimers: structural mimics of wild-type and mutant collagen type I, *J Am Chem Soc*, 130 (2008) 7509-7515.

- [101] S. Rele, Y. Song, R.P. Apkarian, Z. Qu, V.P. Conticello, E.L. Chaikof, D-periodic collagen-mimetic microfibers, *J Am Chem Soc*, 129 (2007) 14780-14787.
- [102] D.N. Woolfson, The design of coiled-coil structures and assemblies, *Adv Protein Chem*, 70 (2005) 79-112.
- [103] G. Ghirlanda, A. Osyczka, W. Liu, M. Antolovich, K.M. Smith, P.L. Dutton, A.J. Wand, W.F. DeGrado, De novo design of a D2-symmetrical protein that reproduces the diheme four-helix bundle in cytochrome bc1, *J Am Chem Soc*, 126 (2004) 8141-8147.
- [104] S. Koide, Generation of new protein functions by nonhomologous combinations and rearrangements of domains and modules, *Curr Opin Biotechnol*, 20 (2009) 398-404.
- [105] H.K. Binz, P. Amstutz, A. Pluckthun, Engineering novel binding proteins from nonimmunoglobulin domains, *Nat Biotechnol*, 23 (2005) 1257-1268.
- [106] A. Urvoas, A. Guellouz, M. Valerio-Lepiniec, M. Graille, D. Durand, D.C. Desravines, H. van Tilbeurgh, M. Desmadril, P. Minard, Design, production and molecular structure of a new family of artificial alpha-helical repeat proteins (alphaRep) based on thermostable HEAT-like repeats, *J Mol Biol*, 404 (2010) 307-327.
- [107] J.E. Padilla, C. Colovos, T.O. Yeates, Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments, *Proc Natl Acad Sci U S A*, 98 (2001) 2217-2221.
- [108] N.L. Ogiwara, G. Ghirlanda, J.W. Bryson, M. Gingery, W.F. DeGrado, D. Eisenberg, Design of three-dimensional domain-swapped dimers and fibrous oligomers, *Proc Natl Acad Sci U S A*, 98 (2001) 1404-1409.
- [109] P. Ringle, G.E. Schulz, Self-assembly of proteins into designed networks, *Science* (New York, N.Y.), 302 (2003) 106-109.
- [110] J.C. Carlson, S.S. Jena, M. Flenniken, T.F. Chou, R.A. Siegel, C.R. Wagner, Chemically controlled self-assembly of protein nanorings, *J Am Chem Soc*, 128 (2006) 7630-7638.
- [111] N.C. Seeman, A.M. Belcher, Emulating biology: building nanostructures from the bottom up, *Proc Natl Acad Sci U S A*, 99 Suppl 2 (2002) 6451-6455.
- [112] S.R. Whaley, D.S. English, E.L. Hu, P.F. Barbara, A.M. Belcher, Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly, *Nature*, 405 (2000) 665-668.
- [113] C.A. Floudas, H.K. Fung, S.R. McAllister, M. Mönnigmann, R. Rajgaria, Advances in protein structure prediction and de novo protein design: A review, *Chem Eng J*, 61 (2006) 966-988.
- [114] R. Dahm, Friedrich Miescher and the discovery of DNA, *Dev Biol*, 278 (2005) 274-288.
- [115] O.T. Avery, C.M. Macleod, M. McCarty, Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii, *J Exp Med*, 79 (1944) 137-158.
- [116] J.D. Watson, F.H. Crick, Genetical implications of the structure of deoxyribonucleic acid, *Nature*, 171 (1953) 964-967.
- [117] J.D. Watson, F.H. Crick, Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid, *Nature*, 171 (1953) 737-738.
- [118] E. Chargaff, B. Magasanik, et al., The nucleotide composition of ribonucleic acids, *J Am Chem Soc*, 71 (1949) 1513.
- [119] R. Holliday, A mechanism for gene conversion in fungi. , *Genet. Res.*, (1964) 282-304.
- [120] K. Hoogsteen, The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine *Acta Cryst.*, 16 (1963) 907-916.
- [121] P. Balagurumothy, S.K. Brahmachari, D. Mohanty, M. Bansal, V. Sasisekharan, Hairpin and parallel quartet structures for telomeric sequences, *Nucleic acids research*, 20 (1992) 4061-4067.

- [122] R.R. Sinden, DNA structure and function, in, Academic Press., 1994, pp. pp.398.
- [123] D. Sen, W. Gilbert, Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis, *Nature*, 334 (1988) 364-366.
- [124] G.N. Parkinson, M.P. Lee, S. Neidle, Crystal structure of parallel quadruplexes from human telomeric DNA, *Nature*, 417 (2002) 876-880.
- [125] J.E. Johnson, J.S. Smith, M.L. Kozak, F.B. Johnson, In vivo veritas: using yeast to probe the biological functions of G-quadruplexes, *Biochimie*, 90 (2008) 1250-1263.
- [126] S. Ahmed, A. Kintanar, E. Henderson, Human telomeric C-strand tetraplexes, *Nat Struct Biol*, 1 (1994) 83-88.
- [127] J.L. Leroy, M. Guéron, J.L. Mergny, C. Helene, Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif, *Nucleic acids research*, 22 (1994) 1600-1606.
- [128] R. Wheeler, in.
- [129] J.M. Rosenberg, N.C. Seeman, J.J. Kim, F.L. Suddath, H.B. Nicholas, A. Rich, Double helix at atomic resolution, *Nature*, 243 (1973) 150-154.
- [130] R.O. Day, N.C. Seeman, J.M. Rosenberg, A. Rich, A crystalline fragment of the double helix: the structure of the dinucleoside phosphate guanylyl-3',5'-cytidine, *Proc Natl Acad Sci U S A*, 70 (1973) 849-853.
- [131] A. Rich, D.R. Davies, A new two stranded helical structure: Polyadenylic acid and polyuridylic acid, *J. AM. Chem. Soc.*, 78 (1956) 3548-3549.
- [132] A. Rich, A. Nordheim, A.H. Wang, The chemistry and biology of left-handed Z-DNA, *Annu Rev Biochem*, 53 (1984) 791-846.
- [133] M.P. Kladde, Y. Kohwi, T. Kohwi-Shigematsu, J. Gorski, The non-B-DNA structure of d(CA/TG)_n differs from that of Z-DNA, *Proc Natl Acad Sci U S A*, 91 (1994) 1898-1902.
- [134] D.N. Gopaul, F. Guo, G.D. Van Duyne, Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination, *The EMBO journal*, 17 (1998) 4175-4187.
- [135] http://en.wikipedia.org/wiki/Holliday_junction, Holliday junction, in.
- [136] <http://openlearn.open.ac.uk>, Triplex structures, in, The Open University.
- [137] http://fr.wikipedia.org/wiki/Appariement_Hoogsteen, Appariement Hoogsteen, in.
- [138] S.M. Mirkin, M.D. Frank-Kamenetskii, H-DNA and related structures, *Annu Rev Biophys Biomol Struct*, 23 (1994) 541-576.
- [139] G. Wang, K.M. Vasquez, Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells, *Proceedings of the National Academy of Sciences of the United States of America*, 101 (2004) 13448-13453.
- [140] S.H. Kim, F.L. Suddath, G.J. Quigley, A. McPherson, J.L. Sussman, A.H. Wang, N.C. Seeman, A. Rich, Three-dimensional tertiary structure of yeast phenylalanine transfer RNA, *Science (New York, N.Y.)*, 185 (1974) 435-440.
- [141] F.L. Suddath, G.J. Quigley, A. McPherson, D. Sneden, J.J. Kim, S.H. Kim, A. Rich, Three-dimensional structure of yeast phenylalanine transfer RNA at 3.0angstroms resolution, *Nature*, 248 (1974) 20-24.
- [142] S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, S. Neidle, Quadruplex DNA: sequence, topology and structure, *Nucleic acids research*, 34 (2006) 5402-5415.
- [143] M. Gellert, M.N. Lipsett, D.R. Davies, Helix formation by guanylic acid, *Proc Natl Acad Sci U S A*, 48 (1962) 2013-2018.
- [144] E. Protozanova, R.B. Macgregor, Jr., Frayed wires: a thermally stable form of DNA with two distinct structural domains, *Biochemistry*, 35 (1996) 16638-16645.
- [145] J. Al-Nasir, O. Madhani, A detailed literature review of G-quadruplex ligand binding assays in: Kingston School of Pharmacy and Chemistry & St. Georges Hospital Medical School, London, 2009.
- [146] J.L. Leroy, S. Nonin, X. Han, A.T. Phan, M. Guéron, Switching and looping in i-motif structures, in: R.H.S. Sarma, M.H. (Ed.) *Structure, Motion, Interaction and Expression of*

- Biological Macromolecules, Adenine Press, State University of New York, Albany, 1998, pp. 49-62.
- [147] K. Gehring, J.L. Leroy, M. Gueron, A tetrameric DNA structure with protonated cytosine-cytosine base pairs, *Nature*, 363 (1993) 561-565.
- [148] J. Feigon, A new DNA quadruplex, *Curr Biol*, 3 (1993) 611-613.
- [149] Leroy J.L., Nonin S., Han X., Phan A.T., Guéron M., Switching and looping in i-motif structures, in: R.H.S. arma, M.H. (Ed.) *Structure, Motion, Interaction and Expression of Biological Macromolecules*, Adenine Press, 1998, pp. 49-62.
- [150] P.E. Nielsen, M. Egholm, R.H. Berg, O. Buchardt, Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide, *Science (New York, N.Y.)*, 254 (1991) 1497-1500.
- [151] T. Saison-Behmoaras, [One decade later, what future for PNA?], *Med Sci (Paris)*, 20 (2004) 148-150.
- [152] A. Laisne, M. Ewald, T. Ando, E. Lesniewska, D. Pompon, Self-assembly properties and dynamic of synthetic proteo-nucleic building blocks in solution and on surfaces, *Bioconjug Chem*, (2011).
- [153] W. Boireau, S. Bombard, M.A. Sari, D. Pompon, Bioengineering and characterization of DNA-protein assemblies floating on supported membranes, *Biotechnol Bioeng*, 77 (2002) 225-231.
- [154] D. Pompon, A. Laisne, PDNA as building blocks for membrane-guided self-assemblies, *Biochem Soc Trans*, 35 (2007) 495-497.
- [155] J.M. Lehn, Toward complex matter: supramolecular chemistry and self-organization, *Proc Natl Acad Sci U S A*, 99 (2002) 4763-4768.
- [156] J.M. Lehn, Toward self-organization and complex matter, *Science (New York, N.Y.)*, 295 (2002) 2400-2403.
- [157] J.M. Lehn, *Supramolecular chemistry*, *Science (New York, N.Y.)*, 260 (1993) 1762-1763.
- [158] J.M. Lehn, Dynamers: dynamic molecular and supramolecular polymers, *Progress in Polymer Science*, 30 (2005) 814-831.
- [159] L.M. Phelps, 1995 Feynman Prize in Nanotechnology Awarded For Pioneering Synthesis of 3-D DNA Objects, in: *Foresight Update* 23 1995, pp. 1.
- [160] J.H. Chen, N.C. Seeman, Synthesis from DNA of a molecule with the connectivity of a cube, *Nature*, 350 (1991) 631-633.
- [161] E. Winfree, F. Liu, L.A. Wenzler, N.C. Seeman, Design and self-assembly of two-dimensional DNA crystals, *Nature*, 394 (1998) 539-544.
- [162] H. Wang, Proving theorems by pattern recognition II, *Bell Systems Technical Journal*, 40 (1961) 1-42.
- [163] F. Liu, R. Sha, N.C. Seeman, Modifying the Surface Features of Two-Dimensional DNA Crystals, *Journal of Am. Chem. Soc.*, 121 (1999) 917-922.
- [164] E. Winfree, *Algorithmic Self-Assembly of DNA*, in, California Institute of Technology, 1998.
- [165] P.W. Rothemund, Folding DNA to create nanoscale shapes and patterns, *Nature*, 440 (2006) 297-302.
- [166] P.W. Rothemund, Scaffolded DNA origami : From generalized multi-crossovers to polygonal networks, in: *Nanotechnology: Science and Computation 2006*, pp. 3-21.
- [167] Y. Liu, Y. Ke, H. Yan, Self-assembly of symmetric finite-size DNA nanoarrays, *J Am Chem Soc*, 127 (2005) 17140-17141.
- [168] T.H. LaBean, E. Winfree, J.H. Reif, Experimental Progress in Computation by self-Assembly of DNA Tilings., in: E. Winfree, D.K. Gifford (Eds.) *DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Proceedings of the 5th DIMACS Workshop on DNA Based Computers*, Cambridge, 1999, pp. 123-140.

- [169] H. Yan, T.H. LaBean, L. Feng, J.H. Reif, Directed Nucleation Assembly of Barcode Patterned DNA Lattices, *Proc Natl Acad Sci U S A*, 100 (2003) 8103-8108.
- [170] P.W. Rothmund, N. Papadakis, E. Winfree, Algorithmic self-assembly of DNA Sierpinski triangles, *PLoS biology*, 2 (2004) e424.
- [171] M.J. Patitz, Simulation of Self-Assembly in the Abstract Tile Assembly Model with ISU TAS, in: 6th Annual Conference on Foundations of Nanoscience: Self-Assembled Architectures and Devices, Snowbird, 2009.
- [172] K. Fujibayashi, D.Y. Zhang, E. Winfree, S. Murata, Error suppression mechanisms for DNA tile self-assembly and their simulation, *Nat Comput*, 8 (2009) 589-612.
- [173] E. Winfree, R. Bekbolatov, Proofreading tile sets: error correction for algorithmic self-assembly., *DNA computing*, 9 (2004) 126-144.
- [174] J.H. Reif, S. Sahu, P. Yin, Compact error-resilient computational DNA tiling assemblies, in: C. Ferreti, G. Mauri, C. Zandron (Eds.) *DNA Computing*, Springer-Verlag, Berlin, 2005, pp. 293-307.
- [175] H.L. Chen, A. Goel, Error free self-assembly using error prone tiles, in: C. Ferreti, G. Mauri, C. Zandron (Eds.) *DNA computing*, Springer-Verlag, Berlin, 2005, pp. 1-11.
- [176] N.C. Seeman, Nanomaterials Based on DNA, *Annu Rev Biochem*, (2010).
- [177] D.E. Gilbert, J. Feigon, Multistranded DNA structures, *Curr Opin Struct Biol*, 9 (1999) 305-314.
- [178] S. Pitchiaya, Y. Krishnan, First blueprint, now bricks: DNA as construction material on the nanoscale, *Chem Soc Rev*, 35 (2006) 1111-1121.
- [179] C.M. Niemeyer, Self-assembled nanostructures based on DNA: towards the development of nanobiotechnology, *Curr Opin Chem Biol*, 4 (2000) 609-618.
- [180] T. Simonsson, G-quadruplex DNA structures--variations on a theme, *Biol Chem*, 382 (2001) 621-628.
- [181] C. Lin, Y. Liu, S. Rinker, H. Yan, DNA tile based self-assembly: building complex nanoarchitectures, *Chemphyschem*, 7 (2006) 1641-1647.
- [182] N.C. Seeman, DNA nanotechnology: novel DNA constructions, *Annu Rev Biophys Biomol Struct*, 27 (1998) 225-248.
- [183] E. Braun, Y. Eichen, U. Sivan, G. Ben-Yoseph, DNA-templated assembly and electrode attachment of a conducting silver wire, *Nature*, 391 (1998) 775-778.
- [184] H.W. Fink, C. Schonenberger, Electrical conduction through DNA molecules, *Nature*, 398 (1999) 407-410.
- [185] A.Y. Kasumov, M. Kociak, S. Gueron, B. Reulet, V.T. Volkov, D.V. Klinov, H. Bouchiat, Proximity-induced superconductivity in DNA, *Science (New York, N.Y.)*, 291 (2001) 280-282.
- [186] C.J. Murphy, M.R. Arkin, Y. Jenkins, N.D. Ghatlia, S.H. Bossmann, N.J. Turro, J.K. Barton, Long-range photoinduced electron transfer through a DNA helix, *Science (New York, N.Y.)*, 262 (1993) 1025-1029.
- [187] J.S. Lee, L.J. Latimer, R.S. Reid, A cooperative conformational change in duplex DNA induced by Zn²⁺ and other divalent metal ions, *Biochem Cell Biol*, 71 (1993) 162-168.
- [188] T.H. LaBean, H. Li, Constructing novel materials with DNA, *NanoToday*, 2 (2007) 26-35.
- [189] D.O. Wood, M.J. Dinsmore, G.A. Bare, J.S. Lee, M-DNA is stabilised in G*C tracts or by incorporation of 5-fluorouracil, *Nucleic acids research*, 30 (2002) 2244-2250.
- [190] N.C. Seeman, Nucleic acid junctions and lattices, *J Theor Biol*, 99 (1982) 237-247.
- [191] T.H. LaBean, H. Yan, J. Kopatsch, F. Liu, E. Winfree, J.H. Reif, N.C. Seeman, Construction, Analysis, Ligation, and Self-Assembly of DNA Triple Crossover Complexes, *J. Am. Chem. Soc.*, 122 (2000) 1848-1860.
- [192] T.J. Fu, N.C. Seeman, DNA double-crossover molecules, *Biochemistry*, 32 (1993) 3211-3220.

- [193] D. Yang, M.J. Campolongo, T.N. Nhi Tran, R.C. Ruiz, J.S. Kahn, D. Luo, Novel DNA materials and their applications, *Wiley Interdiscip Rev Nanomed Nanobiotechnol*, 2 (2010) 648-669.
- [194] Y. Ke, Y. Liu, J. Zhang, H. Yan, A study of DNA tube formation mechanisms using 4-, 8-, and 12-helix DNA nanostructures, *J Am Chem Soc*, 128 (2006) 4414-4421.
- [195] S.H. Park, R. Barish, H. Li, J.H. Reif, G. Finkelstein, H. Yan, T.H. Labean, Three-helix bundle DNA tiles self-assemble into 2D lattice or 1D templates for silver nanowires, *Nano letters*, 5 (2005) 693-696.
- [196] F. Mathieu, S. Liao, J. Kopatsch, T. Wang, C. Mao, N.C. Seeman, Six-helix bundles designed from DNA, *Nano letters*, 5 (2005) 661-665.
- [197] S.H. Park, G. Finkelstein, T.H. LaBean, Stepwise self-assembly of DNA tile lattices using dsDNA bridges, *J Am Chem Soc*, 130 (2008) 40-41.
- [198] N.C. Seeman, DNA engineering and its application to nanotechnology, *Trends in biotechnology*, 17 (1999) 437-443.
- [199] P. Sa-Ardyen, A.V. Vologodskii, N.C. Seeman, The flexibility of DNA double crossover molecules, *Biophys J*, 84 (2003) 3829-3837.
- [200] D. Liu, S.H. Park, J.H. Reif, T.H. LaBean, DNA nanotubes self-assembled from triple-crossover tiles as templates for conductive nanowires, *Proc Natl Acad Sci U S A*, 101 (2004) 717-722.
- [201] D. Reishus, B. Shaw, Y. Brun, N. Chelyapov, L. Adleman, Self-assembly of DNA double-double crossover complexes into high-density, doubly connected, planar structures, *J Am Chem Soc*, 127 (2005) 17590-17591.
- [202] C. Mao, W. Sun, N.C. Seeman, Designed Two-Dimensional DNA Holliday Junction Arrays Visualized by Atomic Force Microscopy, *J. Am. Chem. Soc.*, 121 (1999) 5437-5443.
- [203] H. Yan, S.H. Park, G. Finkelstein, J.H. Reif, T.H. LaBean, DNA-templated self-assembly of protein arrays and highly conductive nanowires, *Science (New York, N.Y.)*, 301 (2003) 1882-1884.
- [204] Y. He, T. Ye, M. Su, C. Zhang, A.E. Ribbe, W. Jiang, C. Mao, Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra, *Nature*, 452 (2008) 198-201.
- [205] Y. He, Y. Tian, Y. Chen, Z. Deng, A.E. Ribbe, C. Mao, Sequence Symmetry as a Tool for Designing DNA Nanostructures, *Angewandte Chemie International Edition*, 44 (2005) 6694-6696.
- [206] Y. He, Y. Chen, H. Liu, A.E. Ribbe, C. Mao, Self-assembly of hexagonal DNA two-dimensional (2D) arrays, *J Am Chem Soc*, 127 (2005) 12202-12203.
- [207] D. Sen, W. Gilbert, Novel DNA superstructures formed by telomere-like oligomers, *Biochemistry*, 31 (1992) 65-70.
- [208] T.C. Marsh, J. Vesenka, E. Henderson, A new DNA nanostructure, the G-wire, imaged by scanning probe microscopy, *Nucleic acids research*, 23 (1995) 696-700.
- [209] A. Calzolari, R. DiFelice, E. Molinari, G-quartet bio-molecular nanowires. , *Appl. Phys. Lett.*, 80 (2002) 3331-3333.
- [210] H. Cohen, T. Sapir, N. Borovok, T. Molotsky, R. Di Felice, A.B. Kotlyar, D. Porath, Polarizability of G4-DNA observed by electrostatic force microscopy measurements, *Nano letters*, 7 (2007) 981-986.
- [211] N. Borovok, N. Iram, D. Zikich, J. Ghabboun, G.I. Livshits, D. Porath, A.B. Kotlyar, Assembling of G-strands into novel tetra-molecular parallel G4-DNA nanostructures using avidin-biotin recognition, *Nucleic acids research*, 36 (2008) 5050-5060.
- [212] N. Borovok, T. Molotsky, J. Ghabboun, D. Porath, A. Kotlyar, Efficient procedure of preparation and properties of long uniform G4-DNA nanowires, *Anal Biochem*, 374 (2008) 71-78.
- [213] H.B. Ghodke, R. Krishnan, K. Vignesh, G.V. Kumar, C. Narayana, Y. Krishnan, The I-tetraplex building block: rational design and controlled fabrication of robust 1D DNA

- scaffolds through non-Watson-Crick interactions, *Angew Chem Int Ed Engl*, 46 (2007) 2646-2649.
- [214] G. Blaha, R.E. Stanley, T.A. Steitz, Formation of the first peptide bond: the structure of EF-P bound to the 70S ribosome, *Science (New York, N.Y.)*, 325 (2009) 966-970.
- [215] Y.G. Gao, M. Selmer, C.M. Dunham, A. Weixlbaumer, A.C. Kelley, V. Ramakrishnan, The structure of the ribosome with elongation factor G trapped in the posttranslocational state, *Science (New York, N.Y.)*, 326 (2009) 694-699.
- [216] T.M. Schmeing, R.M. Voorhees, A.C. Kelley, Y.G. Gao, F.V.t. Murphy, J.R. Weir, V. Ramakrishnan, The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA, *Science (New York, N.Y.)*, 326 (2009) 688-694.
- [217] A. Ben-Shem, L. Jenner, G. Yusupova, M. Yusupov, Crystal structure of the eukaryotic ribosome, *Science (New York, N.Y.)*, 330 (2010) 1203-1209.
- [218] L. Nasalean, S. Baudrey, N.B. Leontis, L. Jaeger, Controlling RNA self-assembly to form filaments, *Nucleic acids research*, 34 (2006) 1381-1392.
- [219] B.A. Bunnell, R.A. Morgan, Gene therapy for infectious diseases, *Clin Microbiol Rev*, 11 (1998) 42-56.
- [220] D.K. Hendrix, S.E. Brenner, S.R. Holbrook, RNA structural motifs: building blocks of a modular biomolecule, *Q Rev Biophys*, 38 (2005) 221-243.
- [221] N.B. Leontis, A. Lescoute, E. Westhof, The building blocks and motifs of RNA architecture, *Curr Opin Struct Biol*, 16 (2006) 279-287.
- [222] E. Westhof, B. Masquida, L. Jaeger, RNA tectonics: towards RNA design, *Fold Des*, 1 (1996) R78-88.
- [223] A. Lescoute, E. Westhof, The interaction networks of structured RNAs, *Nucleic acids research*, 34 (2006) 6587-6604.
- [224] E. Bindewald, C. Grunewald, B. Boyle, M. O'Connor, B.A. Shapiro, Computational strategies for the automated design of RNA nanoscale structures from building blocks using NanoTiler, *J Mol Graph Model*, 27 (2008) 299-308.
- [225] E. Bindewald, R. Hayes, Y.G. Yingling, W. Kasprzak, B.A. Shapiro, RNAJunction: a database of RNA junctions and kissing loops for three-dimensional structural analysis and nanodesign, *Nucleic acids research*, 36 (2008) D392-397.
- [226] L. Jaeger, Defining the syntax for self-assembling RNA tertiary architectures, *Nucleic Acids Symp Ser (Oxf)*, (2009) 83-84.
- [227] L. Jaeger, E.J. Verzemnieks, C. Geary, The UA_handle: a versatile submotif in stable RNA architectures, *Nucleic acids research*, 37 (2009) 215-230.
- [228] Y. Sun, C.H. Kiang, DNA-based Artificial Nanostructures: Fabrication, Properties, and Applications, in: Nalwa (Ed.) *Handbook of Nanostructured Biomaterials and Their Applications in Nanobiotechnology*, American Scientific Publishers, 2005, pp. 224-246.
- [229] C.A. Mirkin, R.L. Letsinger, R.C. Mucic, J.J. Storhoff, A DNA-based method for rationally assembling nanoparticles into macroscopic materials, *Nature*, 382 (1996) 607-609.
- [230] A.P. Alivisatos, K.P. Johnsson, X. Peng, T.E. Wilson, C.J. Loweth, M.P. Bruchez, Jr., P.G. Schultz, Organization of 'nanocrystal molecules' using DNA, *Nature*, 382 (1996) 609-611.
- [231] R.C. Mucic, J.J. Storhoff, C.A. Mirkin, R.L. Letsinger, DNA-Directed Synthesis of Binary Nanoparticle Network Materials, *Journal of the American Chemical Society*, 120 (1998) 12674-12675.
- [232] R.Y. Sweeney, E.Y. Park, B.L. Iverson, G. Georgiou, Assembly of multimeric phage nanostructures through leucine zipper interactions, *Biotechnol Bioeng*, 95 (2006) 539-545.
- [233] N.M. Luscombe, S.E. Austin, H.M. Berman, J.M. Thornton, An overview of the structures of protein-DNA complexes, *Genome Biol*, 1 (2000) REVIEWS001.
- [234] W.H. Landschulz, P.F. Johnson, S.L. McKnight, The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins, *Science (New York, N.Y.)*, 240 (1988) 1759-1764.

- [235] V. Guasconi, H. Yahi, S. Ait-Si-Ali, *Facteurs de Transcription*, in, 2003.
- [236] M. Catala, *Embryologie, développement précoce chez l'humain.*, Masson, Issy-les-Moulineaux, 2006.
- [237] T.J. Fu, N.C. Seeman, DNA double-crossover molecules, *Biochemistry*, 32 (1993) 3211-3220.
- [238] N.C. Seeman, DNA in a material world, *Nature*, 421 (2003) 427-431.
- [239] M. Brucalé, G. Zuccheri, B. Samori, Mastering the complexity of DNA nanostructures, *Trends Biotechnol*, 24 (2006) 235-243.
- [240] W.M. Shih, J.D. Quispe, G.F. Joyce, A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron, *Nature*, 427 (2004) 618-621.
- [241] L. Feng, S.H. Park, J.H. Reif, H. Yan, A two-state DNA lattice switched by DNA nanoactuator, *Angewandte Chemie (International ed)*, 42 (2003) 4342-4346.
- [242] P.J. Paukstelis, J. Nowakowski, J.J. Birktoft, N.C. Seeman, Crystal structure of a continuous three-dimensional DNA lattice, *Chem Biol*, 11 (2004) 1119-1126.
- [243] C.M. Niemeyer, M. Adler, Nanomechanical devices based on DNA, *Angewandte Chemie (International ed)*, 41 (2002) 3779-3783.
- [244] N.C. Seeman, From genes to machines: DNA nanomechanical devices, *Trends Biochem Sci*, 30 (2005) 119-125.
- [245] M. Brucalé, G. Zuccheri, B. Samori, The dynamic properties of an intramolecular transition from DNA duplex to cytosine-thymine motif triplex, *Org Biomol Chem*, 3 (2005) 575-577.
- [246] Y. Chen, S.H. Lee, C. Mao, A DNA nanomachine based on a duplex-triplex transition, *Angewandte Chemie (International ed)*, 43 (2004) 5335-5338.
- [247] C. Mao, W. Sun, Z. Shen, N.C. Seeman, A nanomechanical device based on the B-Z transition of DNA, *Nature*, 397 (1999) 144-146.
- [248] H. Yan, X. Zhang, Z. Shen, N.C. Seeman, A robust DNA mechanical device controlled by hybridization topology, *Nature*, 415 (2002) 62-65.
- [249] B. Yurke, A.J. Turberfield, A.P. Mills, Jr., F.C. Simmel, J.L. Neumann, A DNA-fuelled molecular machine made of DNA, *Nature*, 406 (2000) 605-608.
- [250] W. Shen, M.F. Bruist, S.D. Goodman, N.C. Seeman, A protein-driven DNA device that measures the excess binding energy of proteins that distort DNA, *Angewandte Chemie (International ed)*, 43 (2004) 4750-4752.
- [251] P.W.K. Rothemund, Design of DNA origami, in: *Proceedings of the International Conference on Computer-Aided Design (ICCAD05)*, San Jose, CA, USA, 2005, pp. 471-478.
- [252] L. Brindley, DNA origami goes large, in: *Chemistry World*, RSC, 2010.
- [253] E.S. Andersen, M. Dong, M.M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M.M. Golas, B. Sander, H. Stark, C.L. Oliveira, J.S. Pedersen, V. Birkedal, F. Besenbacher, K.V. Gothelf, J. Kjems, Self-assembly of a nanoscale DNA box with a controllable lid, *Nature*, 459 (2009) 73-76.
- [254] E. Ruitz-Hitzky, M. Darder, *Trends in Bio-Hybrid Nanostructures Materials*, *Current nanoscience*, 2 (2006) 153.
- [255] T. Torimoto, M. Yamashita, S. Kuwabata, T. Sakata, H. Mori, H. Yoneyama, Fabrication of CdS Nanoparticle Chains along DNA Double Strands *J. Phys. Chem. B*, 103 (1999) 8799-8803.
- [256] R. Djalali, Y.F. Chen, H. Matsui, Au nanocrystal growth on nanotubes controlled by conformations and charges of sequenced peptide templates, *Journal of the American Chemical Society*, 125 (2003) 5873-5879.
- [257] K.A. Williams, P.T. Veenhuizen, B.G. de la Torre, R. Eritja, C. Dekker, Nanotechnology: carbon nanotubes with DNA recognition, *Nature*, 420 (2002) 761.
- [258] C.M. Niemeyer, Functional devices from DNA and proteins, *NanoToday*, 2 (2007) 42-52.

- [259] C.M. Niemeyer, Semisynthetic DNA-protein conjugates for biosensing and nanofabrication, *Angewandte Chemie (International ed)*, 49 (2010) 1200-1216.
- [260] T. Sano, C.L. Smith, C.R. Cantor, Immuno-PCR: very sensitive antigen detection by means of specific antibody-DNA conjugates, *Science (New York, N.Y.)*, 258 (1992) 120-122.
- [261] C.M. Niemeyer, T. Sano, C.L. Smith, C.R. Cantor, Oligonucleotide-directed self-assembly of proteins: semisynthetic DNA-streptavidin hybrid molecules as connectors for the generation of macroscopic arrays and the construction of supramolecular bioconjugates, *Nucleic acids research*, 22 (1994) 5530-5539.
- [262] S.C. Wu, S.L. Wong, Engineering soluble monomeric streptavidin with reversible biotin binding capability, *J Biol Chem*, 280 (2005) 23225-23231.
- [263] M. Conti, G. Falini, B. Samori, How Strong Is the Coordination Bond between a Histidine Tag and Ni - Nitritotriacetate? An Experiment of Mechanochemistry on Single Molecules, *Angewandte Chemie (International ed)*, 39 (2000) 215-218.
- [264] R.P. Goodman, C.M. Erben, J. Malo, W.M. Ho, M.L. McKee, A.N. Kapanidis, A.J. Turberfield, A facile method for reversibly linking a recombinant protein to DNA, *Chembiochem*, 10 (2009) 1551-1557.
- [265] Y. He, Y. Tian, A.E. Ribbe, C. Mao, Antibody nanoarrays with a pitch of approximately 20 nanometers, *J Am Chem Soc*, 128 (2006) 12664-12665.
- [266] E.W. Voss, Fluorescein hapten: An immunological probe, CRC Press (Boca Raton, Fla.), 1984.
- [267] R. Stoltenburg, C. Reinemann, B. Strehlitz, SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands, *Biomol Eng*, 24 (2007) 381-403.
- [268] J. Liu, Z. Cao, Y. Lu, Functional nucleic acid sensors, *Chem Rev*, 109 (2009) 1948-1998.
- [269] R. Chhabra, J. Sharma, Y. Ke, Y. Liu, S. Rinker, S. Lindsay, H. Yan, Spatially addressable multiprotein nanoarrays templated by aptamer-tagged DNA nanoarchitectures, *J Am Chem Soc*, 129 (2007) 10304-10305.
- [270] L. Fruk, C.M. Niemeyer, Covalent heme-DNA adducts for generating a novel class of artificial heme enzymes, *Angewandte Chemie (International ed)*, 44 (2005) 2603-2606.
- [271] L. Fruk, J. Muller, C.M. Niemeyer, Kinetic analysis of semisynthetic peroxidase enzymes containing a covalent DNA-heme adduct as the cofactor, *Chemistry*, 12 (2006) 7448-7457.
- [272] L. Fruk, J. Muller, G. Weber, A. Narvaez, E. Dominguez, C.M. Niemeyer, DNA-directed immobilization of horseradish peroxidase-DNA conjugates on microelectrode arrays: towards electrochemical screening of enzyme libraries, *Chemistry*, 13 (2007) 5223-5231.
- [273] A. Saghatelian, K.M. Guckian, D.A. Thayer, M.R. Ghadiri, DNA detection and signal amplification via an engineered allosteric enzyme, *J Am Chem Soc*, 125 (2003) 344-345.
- [274] N. Venkatesan, B.H. Kim, Peptide Conjugates of Oligonucleotides: Synthesis and Applications, *Chem. Rev.*, 106 (2006) 3712-3761.
- [275] E.R. Hendrickson, T.M. Truby, R.D. Joerger, W.R. Majarian, R.C. Ebersole, High sensitivity multianalyte immunoassay using covalent DNA-labeled antibodies and polymerase chain reaction, *Nucleic acids research*, 23 (1995) 522-529.
- [276] P.M. Schaeffer, N.E. Dixon, Synthesis and Applications of Covalent Protein-DNA Conjugates, *Aust. J. Chem.*, 62 (2009) 1328-1332.
- [277] K. Li, H.D.H. Stöv, Highly crosslinked micron-range polymer microspheres by dispersion polymerization of divinylbenzene *Polymer Chemistry*, 31 (1993) 2473-2479.
- [278] J. Barruet, Polymères réactifs à base d'isocyanates bloqués: développement de méthodologies de synthèse pour la bioconjugaison, in: *Chimie et Sciences des Matériaux*, Paris, 2007.
- [279] J.A. Prescher, C.R. Bertozzi, Chemistry in living systems, *Nat Chem Biol*, 1 (2005) 13-21.

- [280] E.M. Sletten, C.R. Bertozzi, Bioorthogonal chemistry: fishing for selectivity in a sea of functionality, *Angewandte Chemie (International ed)*, 48 (2009) 6974-6998.
- [281] E.S. Douglas, R.A. Chandra, C.R. Bertozzi, R.A. Mathies, M.B. Francis, Self-assembled cellular microarrays patterned using DNA barcodes, *Lab Chip*, 7 (2007) 1442-1448.
- [282] S.C. Hsiao, A.K. Crow, W.A. Lam, C.R. Bertozzi, D.A. Fletcher, M.B. Francis, DNA-coated AFM cantilevers for the investigation of cell adhesion and the patterning of live cells, *Angewandte Chemie (International ed)*, 47 (2008) 8473-8477.
- [283] H.C. Kolb, M.G. Finn, K.B. Sharpless, Click Chemistry: Diverse Chemical Function from a Few Good Reactions, *Angewandte Chemie (International ed)*, 40 (2001) 2004-2021.
- [284] P.M. Gramlich, C.T. Wirges, A. Manetto, T. Carell, Postsynthetic DNA modification through the copper-catalyzed azide-alkyne cycloaddition reaction, *Angewandte Chemie (International ed)*, 47 (2008) 8350-8358.
- [285] M. Humenik, Y. Huang, Y. Wang, M. Sprinzl, C-terminal incorporation of bio-orthogonal azide groups into a protein and preparation of protein-oligodeoxynucleotide conjugates by Cu^I-catalyzed cycloaddition, *ChemBiochem*, 8 (2007) 1103-1106.
- [286] J. Gierlich, G.A. Burley, P.M. Gramlich, D.M. Hammond, T. Carell, Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA, *Org Lett*, 8 (2006) 3639-3642.
- [287] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, A general method for the covalent labeling of fusion proteins with small molecules in vivo, *Nat Biotechnol*, 21 (2003) 86-89.
- [288] T.W. Muir, D. Sondhi, P.A. Cole, Expressed protein ligation: a general method for protein engineering, *Proc Natl Acad Sci U S A*, 95 (1998) 6705-6710.
- [289] M. Lovrinovic, R. Seidel, R. Wacker, H. Schroeder, O. Seitz, M. Engelhard, R.S. Goody, C.M. Niemeyer, Synthesis of protein-nucleic acid conjugates by expressed protein ligation, *Chem Commun (Camb)*, (2003) 822-823.
- [290] S. Takeda, S. Tsukiji, H. Ueda, T. Nagamune, Covalent split protein fragment-DNA hybrids generated through N-terminus-specific modification of proteins by oligonucleotides, *Org Biomol Chem*, 6 (2008) 2187-2194.
- [291] T.M. Herne, M.J. Tarlov, Characterization of DNA probes immobilized on gold surfaces., *J Am Chem Soc*, 119 (1997) 8916-8920.
- [292] Z. Xiao, M. Xu, T. Ohgi, K. Sagisaka, D. Fujita, Controlled assembly of DNA nanostructures on silanized silicon and mica surfaces for future molecular devices Superlattices and Microstructures, 32 (2003) 215-220.
- [293] X. Sun, S. Hyeon Ko, C. Zhang, A.E. Ribbe, C. Mao, Surface-mediated DNA self-assembly, *J Am Chem Soc*, 131 (2009) 13248-13249.
- [294] M. Yoshino, T. Matsumura, N. Umehara, Y. Akagami, S. Aravindan, T. Ohno, Engineering surface and development of a new DNA micro array chip, *Tribology in Manufacturing Processes* 260 (2006) 274-286.
- [295] Z. Wiesman, N.B. Dom, E. Sharvit, S. Grinberg, C. Linder, E. Heldman, M. Zaccai, Novel cationic vesicle platform derived from vernonia oil for efficient delivery of DNA through plant cuticle membranes, *J Biotechnol*, 130 (2007) 85-94.
- [296] B.A. Grzybowski, E.W. Wilmer, J. Kim, B. K.P., K.J. Bishop, Self-assembly: from crystals to cells, *Soft Matter*, (2009) 1110-1128.
- [297] T. Mitchison, M. Kirschner, Dynamic instability of microtubule growth, *Nature*, 312 (1984) 237-242.
- [298] P. Yin, H.M. Choi, C.R. Calvert, N.A. Pierce, Programming biomolecular self-assembly pathways, *Nature*, 451 (2008) 318-322.
- [299] M. Wieland, A. Benz, J. Haar, K. Halder, J.S. Hartig, Small molecule-triggered assembly of DNA nanoarchitectures, *Chem Commun (Camb)*, 46 (2010) 1866-1868.

- [300] M.E. Leunissen, R. Dreyfus, F.C. Cheong, D.G. Grier, R. Sha, N.C. Seeman, P.M. Chaikin, Switchable self-protected attractions in DNA-functionalized colloids, *Nat Mater*, 8 (2009) 590-595.
- [301] V.M. Rotello, DNA nanotechnology: Hot and sticky or cold and aloof, *Nat Mater*, 8 (2009) 539-540.
- [302] S. Ulrich, Commutation morphologique et dynamique constitutionnelle, in: *Institu de science et d'ingénierie supramoléculaires*, Université Louis Pasteur de Strasbourg, Strasbourg, 2008.
- [303] D. Voet, J.G. Voet, *Biochimie*, 2004.
- [304] J.C. Kaplan, M. Delpuch, L'hybridation moléculaire in: *Biologie moléculaire et médecine*, Flammarion et Cie, France, 1993.
- [305] H.T. Allawi, J. SantaLucia, Jr., Nearest-neighbor thermodynamics of internal A.C mismatches in DNA: sequence dependence and pH effects, *Biochemistry*, 37 (1998) 9435-9444.
- [306] H.T. Allawi, J. SantaLucia, Jr., Thermodynamics of internal C.T mismatches in DNA, *Nucleic acids research*, 26 (1998) 2694-2701.
- [307] L.A. Marky, K.J. Breslauer, Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves, *Biopolymers*, 26 (1987) 1601-1620.
- [308] J. SantaLucia, Jr., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, *Proc Natl Acad Sci U S A*, 95 (1998) 1460-1465.
- [309] N. Le Novère, Melting, computing the melting temperature of nucleic acid duplex, *Bioinformatics*, 17 (2001) 1226-1227.
- [310] R. Owczarzy, B.G. Moreira, Y. You, M.A. Behlke, J.A. Walder, Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations, *Biochemistry*, 47 (2008) 5336-5353.
- [311] R. Owczarzy, Y. You, B.G. Moreira, J.A. Manthey, L. Huang, M.A. Behlke, J.A. Walder, Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures, *Biochemistry*, 43 (2004) 3537-3554.
- [312] J.G. Wetmur, N. Davidson, Kinetics of renaturation of DNA, *J Mol Biol*, 31 (1968) 349-370.
- [313] J.L. Sikorav, H. Orland, A. Braslau, Mechanism of thermal renaturation and hybridization of nucleic acids: Kramers' process and universality in Watson-Crick base pairing, *J Phys Chem B*, 113 (2009) 3715-3725.
- [314] Y. Gao, L.K. Wolf, R.M. Georgiadis, Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison, *Nucleic acids research*, 34 (2006) 3370-3377.
- [315] M.M. Sekar, W. Bloch, P.M. St John, Comparative study of sequence-dependent hybridization kinetics in solution and on microspheres, *Nucleic acids research*, 33 (2005) 366-375.
- [316] G. Glattig, R.G. Winkler, P. Reineker, Freely jointed chain with variable segment number and length, *Colloid and Polymer Science*, 273 (1995) 32-37.
- [317] G.R. Siegert, R.G. Winkler, P. Reineker, Deterministic chaos in the dynamics of a freely jointed chain with three degrees of freedom, *A journal of physical sciences*, 48 (1993) 584-594.
- [318] C. Bouchiat, M.D. Wang, J. Allemand, T. Strick, S.M. Block, V. Croquette, Estimating the persistence length of a worm-like chain molecule from force-extension measurements, *Biophys J*, 76 (1999) 409-413.
- [319] C. Bustamante, J.F. Marko, E.D. Siggia, S. Smith, Entropic elasticity of lambda-phage DNA, *Science (New York, N.Y.)*, 265 (1994) 1599-1600.
- [320] O. Kratky, G. Porod, Röntgenuntersuchung gelöster Fadenmoleküle., *Rec.Trav.Chim.*, 68 (1949) 1106-1123.
- [321] M.N. Dessinges, B. Maier, Y. Zhang, M. Peliti, D. Bensimon, V. Croquette, Stretching single stranded DNA, a model polyelectrolyte, *Phys Rev Lett*, 89 (2002) 248102.

- [322] J.B. Mills, E. Vacano, P.J. Hagerman, Flexibility of single-stranded DNA: use of gapped duplex helices to determine the persistence lengths of poly(dT) and poly(dA), *Journal of molecular biology*, 285 (1999) 245-257.
- [323] H. Einsenberg, DNA flexing, folding and function, *Acc. Chem. Res.*, 20 (1987) 276-282.
- [324] R.T. Kovacicand, K.E. Van Holde, Sedimentation of homogeneous double-stranded DNA-molecules., *Biochemistry*, 16 (1977) 1490-1498.
- [325] B. Tinland, Pluen, A., Sturm, J., Weill, G., Persistence Length of Single-Stranded DNA, *Macromolecules* 30 (1997) 5763–5765.
- [326] H. Jacobson, W.H. Stockmayer, Intramolecular reaction in polycondensation. I. Theory of linear systems., *Journal of Chemical Physics*, 18 (1950) 1600-1606.
- [327] Q. Du, M. Vologodskiaia, H. Kuhn, M. Frank-Kamenetskii, A. Vologodskii, Gapped DNA and cyclization of short DNA fragments, *Biophys J*, 88 (2005) 4137-4145.
- [328] D. Shore, R.L. Baldwin, Energetics of DNA twisting. II. Topoisomer analysis, *J Mol Biol*, 170 (1983) 983-1007.
- [329] D. Shore, R.L. Baldwin, Energetics of DNA twisting. I. Relation between twist and cyclization probability, *J Mol Biol*, 170 (1983) 957-981.
- [330] W.H. Taylor, P.J. Hagerman, Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure. II. NaCl-dependence of DNA flexibility and helical repeat, *J Mol Biol*, 212 (1990) 363-376.
- [331] S. Geggier, A. Vologodskii, Sequence dependence of DNA bending rigidity, *Proc Natl Acad Sci U S A*, 107 (2010) 15421-15426.
- [332] D. Shore, J. Langowski, R.L. Baldwin, DNA flexibility studied by covalent closure of short fragments into circles, *Proc Natl Acad Sci U S A*, 78 (1981) 4833-4837.
- [333] J. Shimada, H. Yamakawa, DNA-topoisomer analysis on the basis of the helical wormlike chain, *Biopolymers*, 23 (1984) 853-857.
- [334] M. Vologodskiaia, A. Vologodskii, Contribution of the intrinsic curvature to measured DNA persistence length, *J Mol Biol*, 317 (2002) 205-213.
- [335] P.J. Hagerman, Flexibility of DNA, *Annu Rev Biophys Biophys Chem*, 17 (1988) 265-286.
- [336] Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskiaia, A. Vologodskii, Cyclization of short DNA fragments and bending fluctuations of the double helix, *Proc Natl Acad Sci U S A*, 102 (2005) 5397-5402.
- [337] P. Furrer, J. Bednar, A.Z. Stasiak, V. Katritch, D. Michoud, A. Stasiak, J. Dubochet, Opposite effect of counterions on the persistence length of nicked and non-nicked DNA, *J Mol Biol*, 266 (1997) 711-721.
- [338] <http://www2.cndp.fr/themadoc/micro3/tableau.pdf>.
- [339] http://w3.cie.uv-tlse2.fr/fr/3-1_La%20recherche/doc/CF_Principe.pdf.
- [340] J. Frank, Single-particle imaging of macromolecules by cryo-electron microscopy, *Annu Rev Biophys Biomol Struct*, 31 (2002) 303-319.
- [341] F. Landousy, Développements méthodologiques pour la caractérisation des complexes ADN-protéines par AFM et études des interactions ADN-Ku, in: *Interfaces Physique Biologie*, Université Paris 7, Paris, 2006.
- [342] T. Ando, N. Kodera, T. Uchihashi, K. Matada, High-speed atomic force microscopy for capturing dynamic behavior of protein molecules at work, *e-J. Surf. Sci. Nanotech.*, 3 (2005) 384-392.
- [343] T. Ando, N. Kodera, E. Takai, D. Maruyama, K. Saito, A. Toda, A high-speed atomic force microscope for studying biological macromolecules, *Proc Natl Acad Sci U S A*, 98 (2001) 12468-12472.
- [344] Y. Martin, C. Williams, H. Wickramasinghe, Atomic force microscope - force mapping and profiling on a sub 100 angström scale, *J. Appl. Phys*, 61 (1987) 4723-4729.

- [345] V. Derjaguin, V.M. Muller, Y.P. Toporov, Effect of contact deformations on the adhesion of particles, *J. Colloid Interface Sci.*, 53 (1975) 314-325.
- [346] B. Pittenger, N. Erina, C. Su, Quantitative Mechanical Property Mapping at the Nanoscale with PeakForce QNM, in: Veeco application note AN128, Rev. A0, 2010, pp. 1-12.
- [347] T. Ando, N. Kodera, D. Maruyama, E. Takai, K. Saito, A. Toda, A High-Speed Atomic Force Microscope for Studying Biological Macromolecules in Action, *Jpn. J. Appl. Phys.*, 41 (2002) 4851-4856.
- [348] N. Kodera, M. Sakashita, T. Ando, Dynamic proportional-integral-differential controller for high-speed atomic force microscopy, *Rev. Sci. Instrum.*, 77 (2006) 083704.
- [349] T. Ando, T. Uchihashi, N. Kodera, D. Yamamoto, M. Taniguchi, A. Miyagi, H. Yamashita, High-speed atomic force microscopy for observing dynamic biomolecular processes, *J Mol Recognit*, 20 (2007) 448-458.
- [350] L.S. Shlyakhtenko, A.A. Gall, A. Filonov, Z. Cerovac, A. Lushnikov, Y.L. Lyubchenko, Silatrane-based surface chemistry for immobilization of DNA, protein-DNA complexes and other biological materials, *Ultramicroscopy*, 97 (2003) 279-287.
- [351] M. Minsky, *Microscopy apparatus*, in, United States, 1961.
- [352] D. Casanova, Nanoparticules d'oxyde : développements et applications comme sondes biologiques, in: *Physique*, 2008.
- [353] i.-l.-i.-i.-v. <http://nanomedecine-cancer.e-monsite.com/rubrique>, 495181.html, Nanocrystal.
- [354] M. Han, X. Gao, J.Z. Su, S. Nie, Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules, *Nat Biotechnol*, 19 (2001) 631-635.
- [355] W.C. Chan, S. Nie, Quantum dot bioconjugates for ultrasensitive nonisotopic detection, *Science (New York, N.Y.)*, 281 (1998) 2016-2018.
- [356] P. Spinicelli, B. Mahler, S. Buil, X. Quelin, B. Dubertret, J.P. Hermier, Non-blinking semiconductor colloidal quantum dots for biology, optoelectronics and quantum optics, *Chemphyschem*, 10 (2009) 879-882.
- [357] F. Bardin, A. Bellemain, G. Roger, M. Canva, Surface plasmon resonance spectro-imaging sensor for biomolecular surface interaction characterization, *Biosensors & bioelectronics*, 24 (2009) 2100-2105.
- [358] L. Berguiga, T. Roland, K. Monier, J. Elezgaray, F. Argoul, Amplitude and phase images of cellular structures with a scanning surface plasmon microscope, *Opt Express*, 19 (2011) 6571-6586.
- [359] C.H. Sung, D. Chauvat, J. Zyss, C.K. Lee, Dual-channel radially-polarized surface plasmon microscopy for simultaneous detection of fluorescence and linear scattering of nanospheres, in: *Proceedings of SPIE - The International Society for Optical Engineering, Society of Photo-Optical Instrumentation Engineers, Bellingham, WA, ETATS-UNIS, 2010.*
- [360] I. Mannelli, L. Lecerf, M. Guerrouache, M. Goossens, M.C. Millot, M. Canva, DNA immobilisation procedures for surface plasmon resonance imaging (SPRI) based microarray systems, *Biosensors & bioelectronics*, 22 (2007) 803-809.
- [361] D.Y. Petrovykh, H. Kimura-Suda, Whitman, L.J., M.J. Tarlov, Quantitative Analysis and Characterization of DNA Immobilized on Gold, *J Am Chem Soc*, 125 (2003) 5219-5226.
- [362] J. Spadavecchia, J. Moreau, J. Hottin, M. Canva, New cysteamine based functionalization for biochip applications, *New cysteamine based functionalization for biochip applications*, 143 (2009) 139-143.
- [363] H. Tanaka, M. Hanasaki, T. Isojima, H. Takeuchi, H. Kawaguchi, Enhancement of sensitivity of SPR protein microarray using a novel 3D protein immobilization, *Colloids and surfaces*, 70 (2009) 259-265.
- [364] C. Wingren, C. Steinhauer, J. Ingvarsson, E. Persson, K. Larsson, C.A. Borrebaeck, Microarrays based on affinity-tagged single-chain Fv antibodies: sensitive detection of analyte in complex proteomes, *Proteomics*, 5 (2005) 1281-1291.

- [365] N. Bassil, *Système biopuce optique en temps réel: Application au diagnostic génétique.*, in, Université Paris Sud, Orsay, 2005.
- [366] P. Boisseau, P. Houdy, M. Lahmani, *Nanoscience: Nanobiotechnology and Nanobiology*, Springer, 2009.
- [367] M. Lahmani, P. Boisseau, P. Houdy, *Les nanosciences: Tome 3, Nanobiotechnologies et nanobiologie*, Belin, 2007.
- [368] B. Jang, Y.B. Kim, F. Lombardi, Error rate reduction in DNA self-assembly by non-constant monomer concentrations and profiling, *Design, Automation & Test in Europe Conference & Exhibition*, (2007) 159.
- [369] D. Soloveichik, M. Cook, E. Winfree, Combining self-healing and proofreading in self-assembly, *Natural Computing: an international journal*, 7 (2008) 203-218.
- [370] U. Majumder, T.H. LaBean, J.H. Reif, *Activatable Tiles: Compact, Robust Programmable Assembly and Other Applications Lecture notes in computer science*, 4848 (2008) 15-25.
- [371] R. Schulman, S. Lee, N. Papadakis, E. Winfree, One Dimensional Boundaries for DNA Tile Self-Assembly *Lecture notes in computer science*, 2943 (2004) 108-125.
- [372] K. Somei, S. Kaneda, T. Fujii, S. Murata, A microfluidic device for DNA tile self-assembly, *Lecture notes in computer science*, 3892 (2006) 325-335.
- [373] R. Schulman, E. Winfree, Synthesis of crystals with a programmable kinetic barrier to nucleation, *Proc Natl Acad Sci U S A*, 104 (2007) 15236-15241.
- [374] V. Stein, F. Hollfelder, An efficient method to assemble linear DNA templates for in vitro screening and selection systems, *Nucleic acids research*, 37 (2009) e122.
- [375] J.R. TerMaat, E. Pienaar, S.E. Whitney, T.G. Mamedov, A. Subramanian, Gene synthesis by integrated polymerase chain assembly and PCR amplification using a high-speed thermocycler, *J Microbiol Methods*, 79 (2009) 295-300.
- [376] A.S. Xiong, R.H. Peng, J. Zhuang, F. Gao, Y. Li, Z.M. Cheng, Q.H. Yao, Chemical gene synthesis: strategies, softwares, error corrections, and applications, *FEMS Microbiol Rev*, 32 (2008) 522-540.
- [377] A.S. Xiong, R.H. Peng, J. Zhuang, J.G. Liu, F. Gao, J.M. Chen, Z.M. Cheng, Q.H. Yao, Non-polymerase-cycling-assembly-based chemical gene synthesis: strategies, methods, and progress, *Biotechnol Adv*, 26 (2008) 121-134.
- [378] L.S. Larsen, C.D. Wassman, G.W. Hatfield, R.H. Lathrop, Computationally Optimised DNA Assembly of synthetic genes, *Int J Bioinform Res Appl*, 4 (2008) 324-336.
- [379] D. Liu, Jiang H., Chen K., and Ji R., A New Approach to Design Virtual Combinatorial Library with Genetic Algorithm Based on 3D Grid Property, *J. Chem. Inf. Comput. Sci.*, 38 (1998) 233-242.
- [380] C. Neylon, Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution, *Nucleic acids research*, 32 (2004) 1448-1459.
- [381] Z.A. Shabarova, I.N. Merenkova, T.S. Oretskaya, N.I. Sokolova, E.A. Skripkin, E.V. Alexeyeva, A.G. Balakin, A.A. Bogdanov, Chemical ligation of DNA: the first non-enzymatic assembly of a biologically active gene, *Nucleic acids research*, 19 (1991) 4247-4251.
- [382] H. Kuhn, E. Protozanova, V.V. Demidov, Monitoring of single nicks in duplex DNA by gel electrophoretic mobility-shift assay, *Electrophoresis*, 23 (2002) 2384-2387.
- [383] N.G. Dolinnaya, O.V. Pytrauskene, Z.A. Shabarova, Probing DNA triple helix structure by chemical ligation, *FEBS Lett*, 284 (1991) 232-234.
- [384] N.G. Dolinnaya, N.I. Sokolova, D.T. Ashirbekova, Z.A. Shabarova, The use of BrCN for assembling modified DNA duplexes and DNA-RNA hybrids; comparison with water-soluble carbodiimide, *Nucleic acids research*, 19 (1991) 3067-3072.
- [385] N.G. Dolinnaya, A.V. Tsytovich, V.N. Sergeev, T.S. Oretskaya, Z.A. Shabarova, Structural and kinetic aspects of chemical reactions in DNA duplexes. *Information on DNA*

- local structure obtained from chemical ligation data, *Nucleic acids research*, 19 (1991) 3073-3080.
- [386] K.J. Luebke, P.B. Dervan, Nonenzymatic ligation of double-helical DNA by alternate-strand triple helix formation, *Nucleic acids research*, 20 (1992) 3005-3009.
- [387] N.G. Dolinnaya, N.I. Sokolova, O.I. Gryaznova, Z.A. Shabarova, Site-directed modification of DNA duplexes by chemical ligation, *Nucleic acids research*, 16 (1988) 3721-3738.
- [388] N.I. Sokolova, D.T. Ashirbekova, N.G. Dolinnaya, Z.A. Shabarova, Chemical reactions within DNA duplexes. Cyanogen bromide as an effective oligodeoxyribonucleotide coupling agent, *FEBS Lett*, 232 (1988) 153-155.
- [389] C. Cheng, S. Shuman, Recombinogenic flap ligation pathway for intrinsic repair of topoisomerase IB-induced double-strand breaks, *Mol Cell Biol*, 20 (2000) 8059-8068.
- [390] D. Lubrich, S.J. Green, A.J. Turberfield, Kinetically controlled self-assembly of DNA oligomers, *J Am Chem Soc*, 131 (2009) 2422-2423.
- [391] Q. Du, M. Vologodskaya, H. Kuhn, M. Frank-Kamenetskii, A. Vologodskii, Gapped DNA and cyclization of short DNA fragments, *Biophys J*, 88 (2005) 4137-4145.
- [392] J.A. Schellman, S.C. Harvey, Static contributions to the persistence length of DNA and dynamic contributions to DNA curvature, *Biophys Chem*, 55 (1995) 95-114.
- [393] Y. Zhang, D.M. Crothers, High-throughput approach for detection of DNA bending and flexibility based on cyclization, *Proc Natl Acad Sci U S A*, 100 (2003) 3161-3166.
- [394] Y. Zhang, D.M. Crothers, Statistical mechanics of sequence-dependent circular DNA and its application for DNA cyclization, *Biophys J*, 84 (2003) 136-153.
- [395] Y. Zhang, A.E. McEwen, D.M. Crothers, S.D. Levene, Statistical-mechanical theory of DNA looping, *Biophys J*, 90 (2006) 1903-1912.
- [396] J.B. Mills, E. Vacano, P.J. Hagerman, Flexibility of single-stranded DNA: use of gapped duplex helices to determine the persistence lengths of poly(dT) and poly(dA), *J Mol Biol*, 285 (1999) 245-257.
- [397] E.K. Achter, G. Felsenfeld, The conformation of single-strand polynucleotides in solution: sedimentation studies of apurinic acid, *Biopolymers*, 10 (1971) 1625-1634.
- [398] M.C. Murphy, I. Rasnik, W. Cheng, T.M. Lohman, T. Ha, Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy, *Biophys J*, 86 (2004) 2530-2537.
- [399] C. Rivetti, C. Walker, C. Bustamante, Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility, *J Mol Biol*, 280 (1998) 41-59.
- [400] H. Guo, T.D. Tullius, Gapped DNA is anisotropically bent, *Proceedings of the National Academy of Sciences of the United States of America*, 100 (2003) 3743-3747.
- [401] S.S. Sorlie, R. Pecora, A dynamic light scattering study of four DNA restriction fragments, *Macromolecules*, 23 (1990) 487-497.
- [402] A. Hine, X. Chen, M. Hughes, K. Zhou, E. Davies, K. Sugden, I. Bennion, L. Zhang, Optical fibre-based detection of DNA hybridization., *iochem Soc Trans.*, 37 (2009) 445-449.
- [403] G. Jiang, R. Chen, H. Yan, Q. Ouyang, A new method of preparing fiber-optic DNA biosensor and its array for gene detection, *Sci China C Life Sci*, 44 (2001) 33-39.
- [404] D. Walt, Fibre optic microarrays., *Chem Soc Rev.*, 39 (2010) 38-50.
- [405] D.R. Walt, Fiber optic array biosensors, *Biotechniques*, 41 (2006) 529, 531, 533 passim.
- [406] H. Ye, M.C. Huang, M.H. Li, J.Y. Ying, Experimental analysis of gene assembly with TopDown one-step real-time gene synthesis, *Nucleic acids research*, 37 (2009) e51.
- [407] D.M. Lilley, The ribosome functions as a ribozyme, *Chembiochem*, 2 (2001) 31-35.
- [408] C.M. Niemeyer, B. Ceyhan, DNA-Directed Functionalization of Colloidal Gold with Proteins This work was supported by Deutsche Forschungsgemeinschaft and Fonds der

- Chemischen Industrie. We thank Prof. D. Blohm for helpful discussions and generous support, *Angewandte Chemie (International ed)*, 40 (2001) 3685-3688.
- [409] C. Lin, Y. Liu, H. Yan, Designer DNA nanoarchitectures, *Biochemistry*, 48 (2009) 1663-1674.
- [410] S. Scheuring, P. Ringler, M. Borgnia, H. Stahlberg, D.J. Muller, P. Agre, A. Engel, High resolution AFM topographs of the Escherichia coli water channel aquaporin Z, *The EMBO journal*, 18 (1999) 4981-4987.
- [411] F. Rico, C. Su, S. Scheuring, Mechanical mapping of single membrane proteins at submolecular resolution, *Nano letters*, 11 (2011) 3983-3986.
- [412] N. Kodera, D. Yamamoto, R. Ishikawa, T. Ando, Video imaging of walking myosin V by high-speed atomic force microscopy, *Nature*, 468 (2010) 72-76.
- [413] E. Osterberg, K. Bergstrom, K. Holmberg, T.P. Schuman, J.A. Riggs, N.L. Burns, J.M. Van Alstine, J.M. Harris, Protein-rejecting ability of surface-bound dextran in end-on and side-on configurations: comparison to PEG, *Journal of biomedical materials research*, 29 (1995) 741-747.
- [414] M. Farrell, S. Beaudoin, Surface forces and protein adsorption on dextran- and polyethylene glycol-modified polydimethylsiloxane, *Colloids and surfaces*, 81 (2010) 468-475.
- [415] M. Malmqvist, H. Roos, S. Sjolander, M. Tidare, H. Sjodin, R. Stalberg, Method and device for laminar flow on a sensing surface, in, GE Healthcare Bio-Sciences AB, 2011.
- [416] M. Tidare, H. Roos, Sample flow positioning method and analytical system using the method, in, Biacore AB, 2007.
- [417] F. Raynal, A. Beuf, F. Plaza, J.F. Scott, P. Carrière, M. Cabrera, J.-P. Cloarec, E. Souteyrand, Towards better DNA chip hybridization using chaotic advection, *Phys. Fluids*, 19 (2007) 017112-017122.
- [418] T.E. Malliavin, K. Snoussi, J.L. Leroy, The NMR structure of [Xd(C2)]₄ investigated by molecular dynamics simulations, *Magnetic Resonance in Chemistry*, 41 (2003) 18-25.
- [419] J.L. Leroy, M. Gueron, Solution structures of the i-motif tetramers of d(TCC), d(5methylCCT) and d(T5methylCC): novel NOE connections between amino protons and sugar protons, *Structure*, 3 (1995) 101-120.
- [420] S. Nonin, A.T. Phan, J.L. Leroy, Solution structure and base pair opening kinetics of the i-motif dimer of d(5mCCTTTACC): a noncanonical structure with possible roles in chromosome stability, *Structure*, 5 (1997) 1231-1246.
- [421] A.T. Phan, M. Gueron, J.L. Leroy, The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere, *J Mol Biol*, 299 (2000) 123-144.
- [422] X. Han, J.L. Leroy, M. Gueron, An intramolecular i-motif: the solution structure and base-pair opening kinetics of d(5mCCT3CCT3ACCT3CC), *J Mol Biol*, 278 (1998) 949-965.
- [423] J.L. Leroy, The formation pathway of i-motif tetramers, *Nucleic acids research*, 37 (2009) 4127-4134.
- [424] C.J. Noren, S.J. Anthony-Cahill, M.C. Griffith, P.G. Schultz, A general method for site-specific incorporation of unnatural amino acids into proteins, *Science (New York, N.Y.)*, 244 (1989) 182-188.
- [425] K. Lund, A.J. Manzo, N. Dabby, N. Michelotti, A. Johnson-Buck, J. Nangreave, S. Taylor, R. Pei, M.N. Stojanovic, N.G. Walter, E. Winfree, H. Yan, Molecular robots guided by prescriptive landscapes, *Nature*, 465 (2010) 206-210.
- [426] S. Grizot, J.C. Epinat, S. Thomas, A. Duclert, S. Rolland, F. Paques, P. Duchateau, Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds, *Nucleic acids research*, 38 (2010) 2006-2018.
- [427] S. Grizot, J. Smith, F. Daboussi, J. Prieto, P. Redondo, N. Merino, M. Villate, S. Thomas, L. Lemaire, G. Montoya, F.J. Blanco, F. Paques, P. Duchateau, Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease, *Nucleic acids research*, 37 (2009) 5405-5419.

[428] S. Grosse, N. Huot, C. Mahiet, S. Arnould, S. Barradeau, D.L. Clerre, I. Chion-Sotinel, C. Jacqmarcq, B. Chapellier, A. Ergani, C. Desseaux, F. Cedrone, E. Conseiller, F. Paques, M. Labetoulle, J. Smith, Meganuclease-mediated Inhibition of HSV1 Infection in Cultured Cells, *Mol Ther*, 19 (2011) 694-702.

Résumé :

La thèse a pour objet l'analyse de mécanismes d'assemblage permettant la création de structures auto-assemblées à base d'ADN, fonctionnalisées par des protéines. Les PDNA sont des molécules synthétiques associant de manière covalente une séquence d'ADN et un module protéique. Basés sur le domaine hémique du cytochrome b5, les PDNA présentent des propriétés d'absorption optique dépendant de l'état d'oxydoréduction. Pour comparer leurs propriétés d'assemblage à celles des ADN correspondants, différentes briques ont été assemblées en ensembles finis ou ouverts. Un premier chapitre développe des approches expérimentales et de modélisation pour caractériser les paramètres contrôlant ces assemblages dans le cas des ADN. Un second chapitre transpose ces approches aux assemblages de PDNA afin de mettre en évidence l'influence des domaines protéiques. Dans les deux cas, les travaux sont menés à la fois à un niveau macroscopique et de molécule unique par AFM classique ou à haute vitesse. Une troisième partie vise à comprendre comment ces assemblages peuvent être multiplexés sur des puces au sein de matrices poreuses de dextran. Nous avons modélisé l'influence de la matrice sur l'assemblage et utilisé ces modèles pour définir les relations entre observables et paramètres microscopiques. Une dernière partie vise à lever les limitations (pièges d'assemblages) mises en évidence dans les réactions sous contrôle cinétique. Nous avons développé des assemblages dans des conditions douces et contrôlées par le pH impliquant des structures ADN en motifs-i. Leur mécanisme de formation et de polymérisation en suprastructures a été étudié en combinant biochimie et analyse structurale par AFM.

Mots clés : Auto-assemblage, Hybridation, Nanostructures, Protéo-nucléiques, Modélisation, High-Speed AFM, ADN non conventionnels

Abstract

The work aimed at the analysis of assembly mechanisms allowing the creation of DNA-based self-assembled structures functionalized with proteins. PDNA are synthetic molecules covalently associating a DNA strand to a protein domain. The ones based on the heme binding domain of cytochrome b5 feature optic properties that depend on oxydoreduction state. In order to compare self-assembly properties of PDNA and corresponding DNA, different building blocks were assembled into converging or open ensembles. A first part develops experimental and modeling approaches to characterize parameters controlling such assemblies of DNA blocks. A second part extends these approaches to PDNA assemblies, bringing in evidence the influence of the protein domain on nucleic parts. Studies were both performed at macroscopic and single molecule levels using conventional and high-speed AFM. A third part describes how chip multiplexing of those assemblies onto a porous dextran matrix influences behaviors. Modeling the matrix influence on assembly allowed defining relations between experimental and macroscopic parameters. A last part aims at bypassing bottlenecks related to the formation of trapped structures when reactions are kinetically driven. We developed i-motif based DNA structures that can be assembled into mild, pH controlled conditions. Their mechanism of formation and polymerization into supramolecular structures were studied combining physico-chemical and structural analyses using AFM.

Keywords: self-assembly, hybridization, nanostructures, proteo-nucleic, modeling, high-speed AFM, non conventional DNA