

Université de Montréal

**Exploiter la coopérativité d'assemblages supramoléculaires d'ADN  
pour contrôler la plage dynamique d'interrupteurs moléculaires**

*Par*

Dominic Lauzon

Département de chimie

Faculté des arts et des sciences

Thèse présentée en vue de l'obtention du grade de  
Philosophiae Doctor (Ph.D.) en chimie

Avril 2022

© Dominic Lauzon, 2022

*Cette thèse intitulée*

**Exploiter la coopérativité d'assemblages de l'ADN pour programmer la plage dynamique  
d'interrupteurs moléculaires**

*Présentée par*

**Dominic Lauzon**

*A été évaluée par un jury composé des personnes suivantes*

**Karen Waldron**

Président-rapporteur

**Alexis Vallée-Bélisle**

Directeur de recherche

**Joelle Pelletier**

Membre du jury

**Maarten Merx**

Examineur externe

## Résumé

L'autoassemblage de diverses biomolécules pour former des complexes moléculaires est à la base de la machinerie cellulaire et des processus biologiques qui s'y rattachent. Il est typiquement considéré qu'un assemblage de plusieurs protéines offre des avantages régulatifs comparativement à une structure protéique similaire construite avec une ou un nombre inférieur de composantes. Ces assemblages offrent, par exemple, la possibilité de contrôler l'activité d'un complexe grâce à la dépendance directe de l'assemblage sur la concentration de ces composantes. De plus, la coopérativité d'interaction entre ces diverses composantes ouvre la voie vers l'obtention de nouveaux mécanismes de régulation. Toutefois, les avantages et les inconvénients directement reliés au nombre de composantes impliquées dans un assemblage ne sont pas totalement bien compris puisque les protéines ont évolué et ont divergé suivant des millions d'années d'évolution. L'objectif principal de cette thèse est d'abord de créer un modèle moléculaire simplifié permettant de mieux comprendre les avantages coopératifs des autoassemblages biologiques pour ensuite s'en inspirer afin de mettre au point de nouveaux mécanismes moléculaires permettant d'optimiser la plage dynamique d'interrupteurs moléculaires autoassemblés. En même temps, il sera possible de mettre en lumière certains avantages évolutifs qui ont poussé les protéines à acquérir plus de composantes moléculaires.

Tout d'abord, la création d'assemblages moléculaires fut effectuée en fragmentant une structure unimoléculaire en plusieurs fragments qui pourront, grâce à leurs interactions, reformer la structure originale. Grâce à une nanostructure simple d'ADN, *c.-à-d.* une jonction à trois branches, il fut possible d'étudier directement l'impact du nombre de composantes sur la fonctionnalité et la régulation d'assemblages multimériques. Il fut observé, malgré l'association plus lente d'un assemblage de trois composantes, que ce même assemblage s'associe de manière plus coopérative tout en permettant la création de nouveaux mécanismes de régulation (*p. ex.* plage dynamique étendue, auto-inhibition et minuterie moléculaire). Ce système simplifié d'ADN a donc permis de conclure que la fragmentation d'une nanostructure en plusieurs composantes est une méthode simple permettant d'optimiser un nanosystème artificiel ou naturel.

Ensuite, une autre méthode de création d'assemblages moléculaires fut étudiée. Celle-ci consiste à fusionner des domaines interagissant par le biais d'un espaceur. Dans une telle stratégie, l'espaceur est appelé à jouer un rôle important dans les propriétés de l'assemblage. Ainsi, en utilisant le même modèle d'ADN à trois composantes, il fut en effet observé que les propriétés de l'espaceur (*p. ex.* sa longueur, sa composition ou sa nature chimique) affectent grandement les propriétés d'assemblage d'un système à trois composantes (*p. ex.* sa stabilité, son niveau de coopérativité ou sa plage dynamique d'assemblage). En effectuant une étude thermodynamique approfondie sur divers assemblages trimériques d'ADN, il fut découvert qu'un espaceur optimal stabilise l'association des diverses composantes en créant une structure plus compacte où les espaceurs se cachent au cœur de la jonction. Il fut aussi démontré qu'en optimisant l'espaceur, il est possible de programmer précisément la plage dynamique d'un assemblage moléculaire à trois composantes.

Finalement, ces découvertes sur les avantages d'un assemblage à trois composantes ont permis la création d'une nouvelle stratégie afin d'optimiser la plage dynamique d'interrupteurs moléculaires. À l'inverse des activateurs allostériques classiques qui altèrent la force d'interaction d'un ligand, *c.-à-d.* le  $K_D$ , en modifiant la conformation de l'interrupteur, un activateur multivalent permet de programmer précisément la plage dynamique de l'interrupteur en exploitant une nouvelle surface d'interaction grâce à la formation d'un assemblage à trois composantes. Cette nouvelle stratégie d'optimisation des interrupteurs moléculaires fut validée grâce à une tige-boucle d'ADN servant comme balise moléculaire. Cette preuve de concept permet de démontrer la viabilité des assemblages moléculaires pour conceptualiser de nouvelles nanotechnologies avec une plage dynamique optimisée. Il est donc possible d'imaginer que les assemblages moléculaires auront un impact immédiat dans divers domaines de la nanotechnologie comme en diagnostic médical, en délivrance contrôlée de médicaments ou en imagerie moléculaire.

**Mots-clés :** Nanotechnologie d'ADN, autoassemblage, coopérativité, plage dynamique, interrupteurs moléculaires, allostérie, effet chélate, mécanismes de régulation, évolution.



## Abstract

The self-assembly of various biomolecules to form molecular complexes is at the basis of the cellular machinery and their related biological processes. It is typically thought that an assembly of several proteins provides regulatory advantages compared to a similar protein built with one or fewer molecular components. These molecular assemblies offer, for example, the possibility to control their activity through the direct dependency of the assembly on the concentration of its components. Moreover, the cooperativity of interaction between their multiple components opens the door to acquiring novel regulation mechanisms. However, the advantages and disadvantages directly related to the number of components involved in an assembly are not totally understood since proteins have evolved and diverged over millions of years of evolution. The main objective of this thesis is to first create a simplified molecular model that will enable to better understand the cooperative advantages of biological self-assemblies. Then, inspired by these new understandings, novel molecular mechanisms will be developed to enable the optimization of the dynamic range of self-assembled molecular switches. Meanwhile, it will be possible to highlight some advantages that have pushed proteins to acquire more molecular components.

The creation of molecular assemblies was demonstrated by fragmenting a nanostructure into multiple fragments which, through their intermolecular interactions, reassemble into the original structure. Using a simple DNA-based nanostructure, *i.e.*, a three-way junction, it was possible to directly study the impact of the number of components on the functionality and regulation of multimeric assemblies. It was found that despite the slower assembly rate of a three-component assembly, this same assembly undergoes a more cooperative assembly enabling the creation of new regulatory mechanisms (*e.g.*, extended dynamic range, self-inhibition and molecular timers). This simplified DNA-based system has therefore made it possible to conclude that fragmenting a nanostructure into multiple components is a simple method to optimize an artificial or a natural nanosystem.

Next, another method to create molecular assemblies was studied. This method consists in fusing interacting domains through a linker. In this strategy, the linker will play an important

role in dictating the properties of the assembly. Therefore, by using the same three-component DNA-based model, it has been observed that the chemical properties of the linker (*e.g.*, its length, its composition, or its chemical nature) considerably affect the assembly properties of a three-component system (*e.g.*, its stability, its level of cooperativity, or its dynamic range). Through an exhaustive thermodynamic study on various trimeric DNA-based assemblies, it was determined that the optimal linker stabilizes the association of all components by creating a more compact assembly where the linkers are buried within the core of the junction. It was also demonstrated that the optimization of the linkers allows to precisely program the dynamic range of the assembly.

Finally, these discoveries on the advantages of a three-component assembly have enabled the creation of a new design strategy to optimize the dynamic range of molecular switches. In contrast to the classic allosteric activator which alters the affinity of a ligand (*i.e.*, the  $K_D$ ) by changing the conformation of the switch, a multivalent activator enables to precisely program the dynamic range of a switch by exploiting a new interacting interface through the formation of a three-component assembly. This new strategy to optimize molecular switches was validated using a DNA-based molecular beacon. This proof of concept demonstrates the viability of molecular assemblies to design novel nanotechnologies with optimized dynamic range. It is possible to imagine that these molecular assemblies could have a direct impact on multiple fields of nanotechnology including medical diagnostics, controlled drug delivery and molecular imaging.

**Keywords:** DNA nanotechnology, self-assembly, cooperativity, dynamic range, molecular switches, allostery, chelate effect, regulation mechanisms, evolution.

# Table des matières

Résumé.....	iii
Abstract.....	v
Table des matières.....	vii
Liste des tableaux.....	xi
Liste des figures.....	xii
Liste des sigles et abréviations.....	xviii
Remerciements.....	xxi
Chapitre 1 – Introduction.....	1
1.1 L’autoassemblage moléculaire.....	1
1.1.1 Les nanomachines naturelles.....	2
1.1.2 La reconnaissance moléculaire : l’énergie chimique de liaison.....	3
1.1.3 Les avantages des autoassemblages protéiques multimériques.....	5
1.1.4 L’évolution vers des systèmes protéiques multimériques.....	8
1.2 Autoassemblage d’interrupteurs moléculaires dynamiques.....	10
1.2.1 Les interrupteurs moléculaires autoassemblés.....	11
1.2.2 Caractérisation d’une courbe de liaison : l’exemple d’un récepteur et son ligand.....	13
1.2.3 Le mécanisme de déplacement de population.....	16
1.2.4 La coopérativité des assemblages moléculaires.....	18
1.2.4.1 L’allostérie : une question de distance.....	19
1.2.4.2 L’effet chélate : une question de proximité.....	21
1.2.4.3 Cycle du double mutant : contribution de l’allostérie et de l’effet chélate.....	24
1.3 L’ADN : un modèle moléculaire pour étudier la régulation d’assemblages moléculaires.....	25

1.3.1 Les différentes structures de l'ADN .....	26
1.3.2 Prédire la stabilité des nanostructures d'ADN .....	28
1.3.3 Mesurer la stabilité des nanostructures d'ADN .....	30
1.3.4 Création d'autoassemblages moléculaires dynamiques à base d'ADN. ....	33
1.4 La biochimie synthétique : comprendre les mécanismes naturels de régulations à partir de modèles moléculaires simplifiés. ....	36
1.4.1 Le mécanisme de déplacement de population .....	37
1.4.2 Le mécanisme d'allostérie et de coopérativité .....	38
1.4.3 Le mécanisme de séquestration moléculaire et d'ultrasensibilité .....	40
1.4.4 Les mécanismes de l'ajustement induit et de la sélection conformationnelle .....	42
1.5 Objectif de la thèse .....	44
1.6 Contenu de la thèse .....	46
1.7 Références.....	49
Chapitre 2 – Functional advantages of building nanosystems using multiple molecular components.....	62
2.1 Abstract .....	62
2.2 Introduction.....	63
2.3 Results .....	64
2.4 Discussion et conclusions.....	77
2.5 Materials and Methods.....	80
2.6 Acknowledgements.....	88
2.7 Supporting Figures .....	89
2.8 References.....	112
Chapitre 3 – Programming the cooperativity of multimeric assemblies by tuning their compactness via optimal linker connection.....	120

3.1 Abstract .....	120
3.2 Introduction.....	121
3.3 Results .....	122
3.4 Discussion et conclusions .....	135
3.5 Materials and Methods.....	137
3.6 Acknowledgements .....	141
3.7 Supporting Figures .....	142
3.8 References.....	160
Chapitre 4 – Multivalent activators: Bioinspired effectors to tune the affinity and dynamic range of molecular switches .....	166
4.1 Abstract .....	166
4.2 Introduction.....	167
4.3 Results .....	168
4.4 Discussion et conclusions .....	175
4.5 Materials and Methods.....	176
4.6 Acknowledgements .....	180
4.7 Supporting Figures .....	181
4.8 References.....	187
Chapitre 5 – Conclusion.....	191
5.1 Conclusions générales .....	191
5.1.1 Création d’assemblages moléculaires par la fragmentation et la fusion .....	191
5.1.2 Avantages et inconvénients des assemblages moléculaires biologiques .....	194
5.1.3 Les interrupteurs moléculaires multivalents .....	198
5.2 Perspectives .....	201

5.2.1	Conceptualisation de biocapteurs .....	201
5.2.2	Transporteurs moléculaires et systèmes de délivrance de médicaments coopératifs à base d'ADN .....	203
5.2.3	Système hors équilibre : l'essence moléculaire .....	204
5.2.4	Matériaux à base d'ADN .....	205
5.2.5	L'évolution vers des assemblages multimériques .....	205
5.3	Références.....	207
Annexes	.....	213

## Liste des tableaux

Table S2.1. – Thermodynamic and kinetic parameters estimated from the chemical denaturation of the one-, two- and three-components system by urea denaturation, thermal denaturation and NUPACK. ....	95
Table S2.2. – Dissociation constants of the dimer BC. ....	96
Table S3.1. – Thermodynamic dissociation constants, allosteric parameters ( $\alpha_B$ and $\alpha_C$ ) and configurational cooperativity parameters ( $\beta$ and EM) of all three-way junctions. ....	148
Table S3.2. – Contribution of allostery ( $\alpha_B$ and $\alpha_C$ ) and the configurational cooperativity ( $\beta$ and EM) to the overall stability of the assembly. ....	148
Table S3.3. – The ionic strength has almost a negligible impact on the $\Delta H_{Ass}$ (in kcal·mol <sup>-1</sup> ) and the $-T\Delta S_{Ass}$ (in kcal·mol <sup>-1</sup> ) of both binding domains. ....	150
Table S3.4. – The $\Delta H_{Ass}$ (in kcal·mol <sup>-1</sup> ) and the $-T\Delta S_{Ass}$ (in kcal·mol <sup>-1</sup> ) of the 3-way junction assemblies are strongly dependent on the ionic strength (see Fig. 3.4b). ....	151
Table S3.5. – Thermal and urea denaturation parameters of polythymine linker systems. ....	154
Table S3.6. – Thermal and urea denaturation parameters of PEG linker systems. ....	154
Table S4.1 – The impact of an allosteric activator and a multivalent activator on the switching equilibrium of a DNA-based switch. ....	184
Table S4.2 – Kinetic parameters ( $k_{on}$ and $k_{off}$ ) of an allosterically and multivalently activated switches. ....	186

## Liste des figures

Figure 1.1 – La cellule et ses nanomachines. ....	3
Figure 1.2 – Schématisation des avantages des autoassemblages protéiques multimériques. ....	6
Figure 1.3 – Schématisation de certains mécanismes d'évolution menant à la formation de systèmes protéiques multimériques. ....	9
Figure 1.4 – Courbe de liaison entre un récepteur et son ligand. ....	15
Figure 1.5 – Schématisation du mécanisme de déplacement de population. ....	17
Figure 1.6 – Schématisation des mécanismes moléculaires menant à de la coopérativité. ....	21
Figure 1.7 – Schématisation du cycle du double mutant. ....	25
Figure 1.8 – Différentes structures pouvant être adoptées par l'ADN. ....	26
Figure 1.9 – Méthodes expérimentales pour mesurer l'énergie de nanostructures d'ADN. ....	31
Figure 1.10 – Stratégies pour construire des autoassemblages d'ADN dynamiques. ....	35
Figure 1.11 – Biochimie synthétique : les mécanismes de déplacement de population et de l'allostérie. ....	38
Figure 1.12 – Biochimie synthétique : le mécanisme de séquestration et les mécanismes cinétiques de l'ajustement induit et de la sélection conformationnelle. ....	42
Figure 2.1 – Thermodynamic and kinetic characterization of a nanosystem built with one (black, 1c), two (blue, 2c), or three (green, 3c) molecular components. ....	65
Figure 2.2 – The assembly of a nanosystem becomes more programmable (regulable) as its number of components increases. ....	68
Figure 2.3 – Specific mutations on trimeric assemblies are more impactful than on their dimeric counterparts and can lead to novel regulatory mechanisms. ....	71
Figure 2.4 – Programming the assembly of trimeric nanosystems over time by exploiting dimeric kinetic traps. ....	73
Figure 2.5 – Programming novel regulatory mechanisms in a complex catalytic nanosystem through fragmentation. ....	76
Figure S2.1 – Nupack analysis of our DNA based nanostructures. ....	89
Figure S2.2 – Urea denaturation profiles. ....	90



Figure S2.3 – The 1-component urea denaturation profile in Fig. S2.2 was fitted by using the fluorescence behaviour of the unfolded state determined on a less stable 1c\* systems ( $F^{\circ}_{Diss}$ ,  $\sigma_{Ass}$ ).  
.....91

Figure S2.4 – m-values and  $\Delta G^{\circ}_{Ass}$  acquired using: A) urea equilibrium unfolding experiments and, B) assembly/disassembly kinetic experiments using urea concentration jumps. ....91

Figure S2.5 – Thermal denaturation analyses of the 1c, 2c and 3s systems. ....92

Figure S2.6 – Kinetic analysis (chevron plot) of the 1c, 2c, and 3c systems. ....93

Figure S2.7 – Kinetic traces for the association and dissociation of the one-component system at different urea concentrations. ....93

Figure S2.8 – Kinetic traces for the association and dissociation of the two-component system at different urea concentrations. ....94

Figure S2.9 – Kinetic traces for the association and dissociation of the three-component system at different urea concentrations. ....94

Figure S2.10 – Estimation of the  $K_D$  between component B and component C. ....96

Figure S2.11 – Understanding the broad response coming with of a system made of three components.....97

Figure S2.12 – The assembly behaviour of the 2c and 3c systems (Fig. 2.2) are well predicted by numerical simulations. ....98

Figure S2.13 – We can tune the  $[A]_{50\%}$  and the dynamic range of the 3c system by simply changing the concentration ratio between the component B and C. ....99

Figure S2.14 – We can also tune the  $[A]_{50\%}$  and the dynamic range of the 3c system by simply changing the affinity between the components via a temperature change. ....99

Figure S2.15 – We can tune the  $[A]_{50\%}$  and the dynamic range of the 3c system by simply performing mutations that reduce the affinity between the components. ....99

Figure S2.16 – Numerical simulations of the assembly profile of a theoretical four-component system (4c system). ....100

Figure S2.17 – Effect of the thymine spacer length between each arm on the 3-way junction stability. ....101

Figure S2.18 – Numerical simulations of the effect of destabilizing only $\Delta G_{Tri}$ on the assembly of a 3-component system.....	101
Figure S2.19 – Native PAGE titration. ....	102
Figure S2.20 – Rate constant measurement of dimer association used for the numerical simulations. ....	102
Figure S2.21 – Rate constant measurement of the binding of the third component on a preorganized dimer used for the numerical simulations.....	103
Figure S2.22 – Numerical simulation of the time-dependent assembly of a 3c system. ....	103
Figure S2.23 – Estimation of the quenching efficiency at 37°C of the dimer and trimer by melting curve analysis. ....	104
Figure S2.24 – Raw kinetic traces of time-dependent programmed assembly (Fig. 2.4B).....	104
Figure S2.25 – Mutational kinetic analysis.....	105
Figure S2.26 – Numerical simulations of the kinetic assembly pathways of a theoretical four-component system (4c system). ....	106
Figure S2.27 – Estimation of the $K_D$ between the DNAzyme (NaA43) and a non-cleavable substrate. ....	107
Figure S2.28 – Activity of the cut DNAzyme. ....	107
Figure S2.29 – Production of an AND gate due to the inactivity of the sequestered dimers.....	108
Figure S2.30 – Raw kinetic traces of A) Fig. 2.5A, <i>left</i> B) Fig. 2.5A, <i>right</i> C) Fig. 2.5B, <i>left</i> and D) Fig. 2.5B, <i>right</i> . ....	108
Figure S2.31 – Raw kinetic traces of A) Fig. 2.5C, <i>left</i> and B) Fig. 2.5C, <i>right</i> . ....	109
Figure S2.32 – Design of a variant DNAzyme for the implementation of the controller strand. ....	110
Figure S2.33 – Adding a third component. ....	111
Figure 3.1 – Programming molecular assemblies by connecting interacting domains via different linker lengths.....	123
Figure 3.2 – Assessing the contributions of allostery and configurational cooperativity using a double mutant cycle analysis. ....	125

Figure 3.3 – Configurational cooperativity ( $\beta$ ) is the main determinant of the difference in assembly stability ( $\Delta G_{Ass}$ ) observed among all molecular assemblies built with various linker lengths. ....	126
Figure 3.4 – High stability arises from increases in enthalpy and level of compaction.....	130
Figure 3.5 – Programming the assembly profile of trimeric assemblies by tuning stability via various linker lengths. ....	134
Figure S3.1 – Thermal denaturation of the trimeric assembly (ABC) built with variable linker length (0T to 8T). ....	142
Figure S3.2 – The trimeric assembly of nanosystems built with variable linker length (0T to 8T). ....	143
Figure S3.3 – The variation of $\Delta G_{Ass}$ when increasing the length of the linker is directly related to the final step of the assembly ( $\Delta G_{Tri}$ ). ....	144
Figure S3.4 – Binding curves for $K_{A1}$ and $K_{A2}$ measurements.....	144
Figure S3.5 – Binding curves for $K_{B1}$ and $K_{B2}$ measurements.....	145
Figure S3.6 – Binding curves for $K_{C1}$ and $K_{C2}$ measurements.....	146
Figure S3.7 – Thermal denaturation of the $K_D$ equilibrium reveals the major contribution of the chelate cooperativity.....	147
Figure S3.8 – Our different assemblies display an enthalpy-entropy compensation.....	149
Figure S3.9 – Thermal denaturation of the binding domains at different salt concentrations...150	
Figure S3.10 – Thermal denaturation of the trimeric assembly (ABC) built with variable linker length (1T to 8T) at different salt concentrations.....	151
Figure S3.11 – Urea denaturation of the $K_D$ equilibrium reveals the role of the linker in the capacity of the system to bury more surfaces in the final assembly. ....	152
Figure S3.12 – Thermal and urea denaturation of the $K_D$ equilibrium of 3-way junction built using PEG linkers.....	153
Figure S3.13 – Thermal denaturation of the trimeric assembly (ABC) built by varying one linker length from 0T to 8T (strand A) while keeping the length of the linker on strand B and C at 2T. ....	155

Figure S3.14 – Changing only the linker on strand A while keeping the two others at 2T provides a straightforward way to program the level of compaction of the assembly. ....156

Figure S3.15 – The trimeric assembly of nanosystems built with two 2T linkers and one variable linker (0T to 24T). ....157

Figure S3.16 – The trimeric assembly of nanosystems built with 2T linkers at variable temperature (35°C to 45°C). ....158

Figure S3.17 – The trimeric assembly of nanosystems built with 8T linkers at variable temperature (31°C to 41°C). ....159

Figure 4.1 – Engineering bioinspired multivalent activators for controlling molecular switches. ....168

Figure 4.2 – A DNA molecular switch modulated via an allosteric or a multivalent mechanism. ....170

Figure 4.3 – Multivalent activators enable programming both the affinity ( $K_D^{APP}$ ) and the dynamic range of the molecular switch by introducing an additional interacting interface. ....171

Figure 4.4 – Fine tuning multivalent activation mechanism by varying the concentration of the switch-activator complex. ....172

Figure 4.5 – Programming the kinetics of activation and deactivation of molecular switches through allosteric and multivalent activation mechanisms.....174

Figure S4.1 – A multivalent activator provides multiple interacting pathways.....181

Figure S4.2 – Comparison between the original switch from Plaxco et al.<sup>3</sup> and our multivalent switch. ....182

Figure S4.3 – Switching equilibrium ( $K_S$ ) of the inactivated, allosterically activated and multivalently activated DNA-base switch. ....182

Figure S4.4 – Stability of the DNA-based switch measured by urea denaturation. ....183

Figure S4.5 – The impact of an allosteric activator and a multivalent activator on switching equilibrium of a DNA-based switch.....183

Figure S4.6 – The extent to which multivalent activators can modulate the  $K_D^{APP}$  and the dynamic range is also concentration dependent. ....184

Figure S4.7 – A molecular switch can also be programmed by simply changing the concentration of multivalent activators. ....185

Figure S4.8 – Determination of the activation kinetic ( $k_{on}$ ) of an allosterically and multivalently activated switches. ....185

Figure S4.9 – Determination of the deactivation kinetic rate constant ( $k_{off}$ ) of an allosterically and multivalently activated switches. ....186

Figure 5.1 – Le futur des assemblages moléculaires multimériques. ....200

## Liste des sigles et abréviations

ADN : Acide désoxyribonucléique (*DNA, Deoxyribonucleic acid*)

ARN : Acide ribonucléique (*RNA, Ribonucleic acid*)

ATP : Adénosine triphosphate (*Adenosine triphosphate*)

BHQ-1 : Black Hole Quencher<sup>®</sup>-1

bp : Paire de bases (*base pairs*)

DMC : Cycle du double mutant (*Double mutant cycle*)

DR : Plage dynamique (*Dynamic range*)

EDTA : Acide éthylènediaminetétraacétique (*Ethylenediaminetetraacetic acid*)

EM : Concentration effective molaire (*Effective molarity*)

FAM : Fluorescéine (*fluorescein*)

FRET : Transfert d'énergie par résonance de type Förster (*Förster resonance energy transfer*)

$K_A$  : Constante d'équilibre d'association (*Association equilibrium constant*)

$K_D$  : Constante d'équilibre de dissociation (*Dissociation equilibrium constant*)

$K_D^{APP}$  : Constante d'équilibre de dissociation apparente (*Apparent dissociation equilibrium constant*)

$K_{Intra}$  : Constante d'équilibre intramoléculaire (*Intramolecular equilibrium constant*)

$k_{Dim}$  : Constante de vitesse de formation de dimère (*dimeric formation rate constant*)

$K_M$  : Constante de Michaelis (*Michaelis constant*)

$k_{off}$  : Constante de vitesse de dissociation (*Dissociation rate constant*)

$k_{on}$  : Constante de vitesse d'association (*Association rate constant*)

$k_{\text{Tri}}$  : Constante de vitesse de formation du trimère (*trimeric formation rate constant*)

$K_S$  : Constante d'équilibre d'interconversion (*Switching equilibrium constant*)

i.e. : id est (c'est-à-dire, *that is*)

KNF : Modèle séquentiel de Koshland–Nemethy–Filmer

MWC : Modèle concerté de Monod–Wyman–Changeux

$n_H$  : Coefficient de Hill (*Hill factor*)

nt : nucléotide (*nucleotide*)

PEG : Polyéthylène glycol (*Polyethylene glycol*)

RCPG : Récepteurs couplés aux protéines G

SELEX : Évolution systématique de ligands par enrichissement exponentiel (*Systematic evolution of ligands by exponential enrichment*)

$t_{1/2}$  : Demi-vie (*half-life*)

$T_m$  : Température de fusion (*Melting temperature*)

vs : versus (*versus*)

$\Delta\text{ASA}$  : Variation de la surface accessible (*Variation in accessible surface area*)

*« Rien ne sert de courir ; il faut partir à point. » - Jean de La Fontaine*



## Remerciements

Tout d'abord, je tiens à remercier mon directeur de recherche Pr Alexis Vallée-Bélisle. Merci de m'avoir accueilli dans ton laboratoire, de m'avoir soutenu durant ces six années au doctorat, mais surtout d'avoir cru en moi et de m'avoir donné la pleine liberté dans mes recherches. Je suis aussi reconnaissant pour les nombreuses heures passées à écrire ensemble de la science pour ensuite nous relire et tout simplement l'effacer en entier pour réécrire presque la même chose, et ce, encore et encore! De nombreuses heures qui ont, parfois, atteint les limites de ma patience, mais qui ont considérablement changé ma façon de percevoir la science et de la communiquer au grand public. Moi qui ai toujours détesté écrire des textes et présenter oralement, je me retrouve maintenant plus outillé que jamais!

J'aimerais aussi remercier Audrey Laventure qui, à la suite de mon premier cours de laboratoire au baccalauréat, m'a offert l'opportunité de réaliser un stage de recherche avec elle. Merci de m'avoir transmis ton grand sens de l'organisation qui me fut très utile durant tout mon doctorat, mais surtout merci de m'avoir transmis ta passion pour la science et la recherche. De te voir maintenant comme professeure au département de chimie de l'UdeM me fait réaliser que tout est possible lorsqu'on le veut! Un grand merci au Pr Christian Pellerin et au Pr Michel Lafleur d'avoir éveillé ma curiosité scientifique en m'accueillant comme stagiaire dans leur laboratoire respectif à l'été 2014 et à l'été 2015. J'aimerais aussi remercier Jean-François St-Amant avec qui j'ai partagé l'enseignement du cours de Chimie Générale pendant plus de 5 ans et qui a su confirmer ma curiosité envers l'enseignement. Merci aussi à mes enseignants du secondaire, Yves Corbel, Denis Bouchard, Dominic Renaud, Dominic Benoît et Sébastien Huot ainsi que mes enseignants du cégep, Karine Gendron, Éric Lefebvre, Johannie Gravel et Christian Côté qui m'ont fait découvrir et aimer les sciences de la nature. Sans vous tous, je n'aurais certainement pas poursuivi une carrière en chimie.

J'aimerais aussi remercier tous les membres du laboratoire qui ont fait de ces six dernières années des années exceptionnelles. Merci Arnaud Desrosiers, Carl Prévost-Tremblay, Simon Diallo-Blais, Laurianne Phan, Marie-Elaine Bérubé, Elizabeth Maurice-Elder, Cassandra Milien, Raphaëlle Zicat-Cloutier, Laurence Robillard, Pierrot-Baptiste Lemée-Jolicoeur et Dr David

Charbonneau. *Thank you* Dr Scott Harroun, Eleanor Campbell et Alison Bateman. *Grazie* Erica Del Grosso. *Xièxiè* Dr Guichi Zhu et Xiaomeng Wang. *Spasibo* Tatiana Lyalina. *Ďakujem* Katarína Nemčková. *Dankjewel* Koen van Putten. *Dhan'yavāda* Dr Bal Ram Adhikari. *Dhanyavaad* Dr. Nagarjun Narayanaswamy. Mention honorable à notre agente de recherche Liliana Pedro, sans toi le laboratoire ne serait pas pareil!

J'aimerais aussi prendre le temps de remercier les membres de ma famille. Mes parents, Gilberte Grenier et Sylvain Lauzon, qui ont toujours cru en moi et qui m'ont transmis les outils nécessaires pour passer au travers de mon doctorat même s'ils ne comprennent toujours rien à ce que je fais. Merci à mon frère et mon ami le plus proche, Dany Lauzon, qui est toujours là lorsque j'ai besoin d'aide même s'il ne manque pas une occasion pour rire de moi et de me ridiculiser. Ton sens de l'humour et ta joie de vivre sont contagieux! À ma bien-aimée, Laurence Forget-Lamarre, avec qui je partage les dix dernières années, tu as su me soutenir dans les moments les plus difficiles. Sans toi, je ne serais pas où je suis présentement. À nos enfants poilues, Alice et Estel, avec qui nous avons la chance de partager nos vies. Comment se passer de ces deux petites bêtes qui ensoleillent chacune de nos journées, parfois des anges, parfois de vrais démons. Vous trois avez été d'un soutien important durant ces deux dernières années de pandémie. Merci à tous mes amis qui avec les tournois de poker, les pratiques de Kinball, les vendredis bières et les soirées « *gamages* » ont su embellir ma vie et continueront de le faire.

Je suis aussi reconnaissant envers les organismes subventionnaires incluant le Conseil de Recherches en Sciences Naturelles et en Génie du Canada, le Fonds de Recherche du Québec - Nature et Technologie et l'Université de Montréal qui ont soutenu financièrement mes études, me permettant ainsi de me concentrer pleinement sur mes projets de recherches. Merci aussi au département de chimie de l'Université de Montréal et au regroupement PROTÉO pour l'obtention de bourse de congrès qui m'ont permis, respectivement, de présenter mes résultats de recherche au 3<sup>e</sup> *Functional DNA Nanotechnology Workshop* en Italie ainsi qu'au *Pacifichem 2021* à Hawaii. Eh, je veux dire en direct de mon salon (pandémie oblige!).

Finalement, à tous ceux et celles que j'ai oubliés et qui m'ont soutenu de près ou de loin :  
Merci!

# Chapitre 1 – Introduction

Les assemblages moléculaires résultant des interactions et de l'autoassemblage entre diverses biomolécules sont à la base de la majorité des processus biologiques et de la nanomachinerie cellulaire<sup>1, 2</sup>. Le ribosome, un complexe moléculaire responsable de la synthèse des protéines, par exemple, est constitué d'unités d'acides ribonucléiques (ARN) responsables de la biocatalyse, mais aussi de plusieurs dizaines de protéines ayant des rôles auxiliaires à la synthèse. En contraste, il est proposé qu'avant même d'être régulées par de tels complexes, les réactions chimiques importantes à la vie étaient catalysées par de simples brins d'ARN unimoléculaires provenant d'un « monde à ARN »<sup>3</sup>. Il y a donc eu des événements évolutifs majeurs qui ont poussé ces systèmes biologiques à acquérir plus de composantes moléculaires. Nous savons depuis quelques années, par exemple, que plus de 55% des protéines caractérisées jusqu'à présent sont constituées de plus d'une sous-unité et permettent, en comparatif à leur équivalent unimoléculaire, un gain d'économie génétique, de régulation et de coopérativité entre les diverses sous-unités<sup>4, 5</sup>. Il demeure toutefois difficile de déterminer les avantages thermodynamiques et cinétiques fondamentaux ayant poussé certains de ces systèmes à former des assemblages moléculaires tandis que d'autres sont demeurés sous leurs formes monomériques. En effet, suivant des millions d'années d'évolution, ces systèmes monomériques et polymériques ont divergé les uns des autres pour former des entités uniques et incomparables<sup>5-7</sup>. Avec l'utilisation de systèmes moléculaires plus simples comme l'ADN, il sera possible de reproduire artificiellement des événements évolutifs menant à des systèmes multimériques de sorte à mieux étudier les avantages thermodynamiques, cinétiques, régulateurs et peut-être même évolutifs de la formation d'assemblages moléculaires. Une meilleure compréhension de ces systèmes multimériques permettra aussi d'exploiter leurs propriétés d'assemblage pour développer une nouvelle génération d'interrupteurs moléculaires coopératifs.

## 1.1 L'autoassemblage moléculaire

Les nanomachines macromoléculaires, ces autoassemblages de molécules biologiques, sont à la base même des différentes tâches essentielles à la vie. Dans cette section, nous verrons

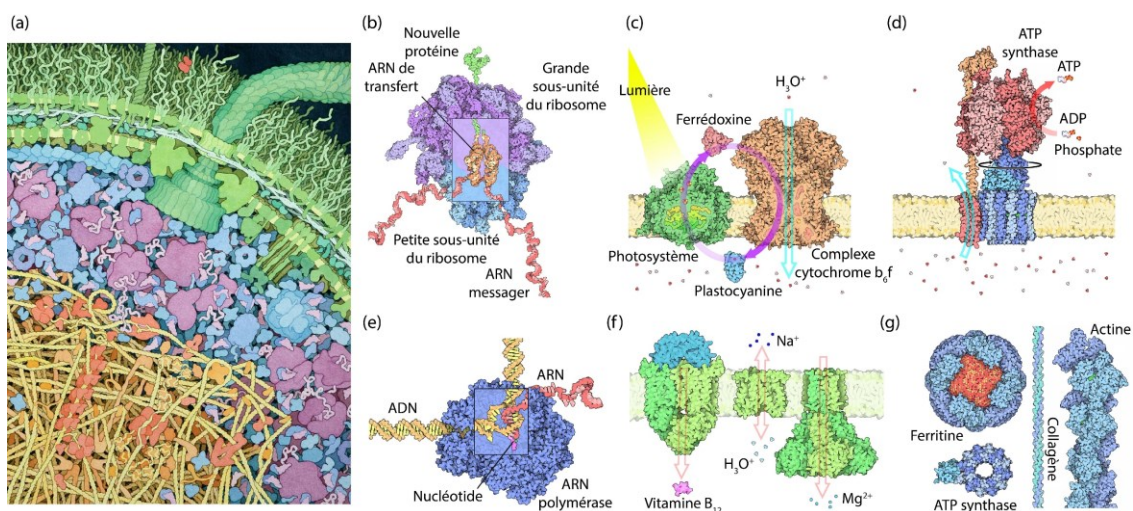
comment les interactions intermoléculaires entre diverses molécules apportent l'énergie nécessaire au maintien et à la fonction d'un autoassemblage, tout en décrivant certaines voies évolutives ayant permis la formation de ces assemblages biologiques. La compréhension des forces motrices derrière ces autoassemblages biologiques et des mécanismes naturels d'évolution qui y sont reliés fut à la base des récents progrès au niveau de la synthèse *de novo* d'assemblages moléculaires fonctionnels en laboratoire<sup>8</sup>.

### **1.1.1 Les nanomachines naturelles**

Afin d'assurer le bon fonctionnement de l'organisme, les cellules font appel à une myriade de petites machines nanométriques accomplissant diverses tâches nécessaires au maintien de l'homéostasie (Fig. 1.1a). Ces nanomachines moléculaires, par définition, consistent en une unité simple ou un assemblage de plusieurs composantes nanométriques qui, suite à un stimulus (*p. ex.* lumière, pH, température, pression, interaction moléculaire, etc.), produisent un mouvement quasi-mécanique<sup>9</sup>. Les nanomachines naturelles, selon le même principe, utilisent l'énergie disponible autour d'elles pour accomplir leurs tâches. Par exemple, le centre réactionnel de la photosynthèse exploite l'énergie lumineuse afin de synthétiser les molécules organiques du vivants<sup>10</sup>. Un autre exemple marquant est l'enzyme adénosine triphosphate (ATP) synthase qui tire profit de l'énergie provenant d'un gradient de protons pour générer un mouvement de rotation qui dicte les changements de conformation nécessaires à la catalyse<sup>11</sup>.

Les nanomachines naturelles peuvent se retrouver sous différentes formes (Fig. 1.1b-g)<sup>12</sup>. Par exemple, les premières images qui nous viennent à l'esprit lorsqu'on entend parler de nanomachines biologiques sont probablement les protéines. Il suffit de penser à l'exemple classique de la kinésine; cette protéine qui semble marcher le long des microtubules pour transporter différents cargos comme des camions sur une autoroute. Les enzymes, ces protéines ayant une activité catalytique, sont aussi souvent utilisées pour représenter la puissance des molécules biologiques pour accomplir plusieurs fonctions dans notre corps, de vrais petits ouvriers. Néanmoins, il ne faut pas oublier que les polymères d'acides ribonucléiques (ARN) ou d'acides désoxyribonucléiques (ADN) peuvent aussi agir comme nanomachine. Le ribosome, par exemple, est un assemblage de protéines structurelles et d'ARN catalytiques qui contient toute

la machinerie moléculaire permettant de lire l'information génétique et ainsi synthétiser toutes les protéines de notre corps; une vraie usine à nanomachines! D'un autre côté, les lipides s'autoassemblent de manière réversible pour créer des compartiments dans lesquels des biomolécules, incluant d'autres lipides, peuvent mieux accomplir leurs fonctions loin de toutes interférences<sup>13</sup>. Les glucides, quant à eux, sont une source d'énergie indispensable aux cellules tout en étant à la base de plusieurs événements de reconnaissance entre diverses biomolécules<sup>14</sup>. Au final, la capacité de toutes ces différentes nanomachines et biomolécules à accomplir diverses tâches est remarquable et a inspiré le développement de plusieurs nanomachines artificielles trouvant leurs applications, par exemple, en diagnostic médical<sup>15</sup>, en système de délivrance de médicaments<sup>16</sup>, en traitement et entreposage d'information<sup>17</sup> et en nanorobotiques<sup>18</sup>, pour ne nommer que ceux-là.



**Figure 1.1 – La cellule et ses nanomachines.** (a) Représentation artistique de la complexité moléculaire au niveau cellulaire. (b) Le ribosome : l'usine à nouvelle protéine. (c) La nanomachinerie moléculaire responsable de la génération d'un gradient de protons à partir de l'énergie lumineuse. (d) L'ATP synthase : un moteur rotatif alimenté par un gradient de protons. (e) L'ARN polymérase : une nanomachine qui lit l'information génétique. (f) Nanomachines responsables du transport de molécules. (g) Nanomachineries moléculaires autoassemblées. Figures adaptées et traduites de la référence 12, © 2009 Springer Science + Business Media.

### 1.1.2 La reconnaissance moléculaire : l'énergie chimique de liaison

Le repliement d'une longue chaîne polymérique pour former une nanomachine fonctionnelle est régi par plusieurs forces intermoléculaires faibles qui, une fois additionnées ensemble, permettent la stabilisation d'une structure tridimensionnelle bien précise. Ces

interactions comprennent l'effet hydrophobe, la formation de ponts hydrogènes, les interactions électrostatiques et les interactions Van der Waals<sup>19</sup>. Chez les protéines, l'effet hydrophobe est la force motrice principale derrière leur repliement. En effet, la nature hydrophobe de certains acides aminés favorise la formation d'un cœur hydrophobe à l'intérieur de la protéine repliée tout en favorisant l'exposition des acides aminés hydrophiles au solvant aqueux<sup>20</sup>. À l'instar des protéines, l'ADN et l'ARN se replient typiquement suivant un patron de ponts hydrogènes bien précis, mais il ne faut toutefois pas négliger la contribution hydrophobe provenant de l'empilement des bases azotées<sup>21</sup>. Les acides gras, ces lipides amphiphiles, s'autoassemblent aussi suivant l'effet hydrophobe où les groupements hydrophobes des queues apolaires évitent les contacts avec l'eau au profit de la tête polaire<sup>13</sup>. Outre l'effet hydrophobe et les ponts hydrogènes, il existe aussi d'autres interactions qui permettent de maintenir l'intégrité de ces structures repliées naturelles. Par exemple, les interactions entre les acides aminés chargés (*p. ex.* la lysine et l'acide aspartique) ou les interactions des groupements phosphates chargés négativement de l'ADN avec des cations en solution sont deux exemples simples de l'impact des forces électrostatiques dans la stabilisation d'une structure tridimensionnelle. De plus, certaines protéines ou certains brins d'ADN vont même former des liaisons covalentes (*p. ex.* pont disulfure) ou des liaisons métalliques (*p. ex.* G-quadruplex ou les doigts de zinc) afin de stabiliser leur structure<sup>22</sup>. De plus, en s'inspirant de ces différentes forces intermoléculaires, il est possible de conceptualiser des molécules synthétiques pouvant se replier, nommés foldamères, de l'anglais « foldamer »<sup>23</sup>.

Les forces moléculaires qui régissent l'association entre différentes composantes moléculaires ne sont pas si différentes de celles impliquées dans le repliement d'un long biopolymère en une structure distincte. Par exemple, l'interface entre différentes protéines formant des complexes est typiquement enrichie en acides aminés hydrophobes tandis que les acides aminés polaires ou chargés se retrouvent à la surface de ces complexes protéiques<sup>20</sup>. Les forces derrière l'association de plusieurs brins d'ADN sont, quant à elles, similaires à celles que l'on retrouve dans un simple brin replié. Néanmoins, plusieurs brins d'ADN sélectionnés artificiellement adoptent des structures non canoniques et interagissent avec des partenaires de

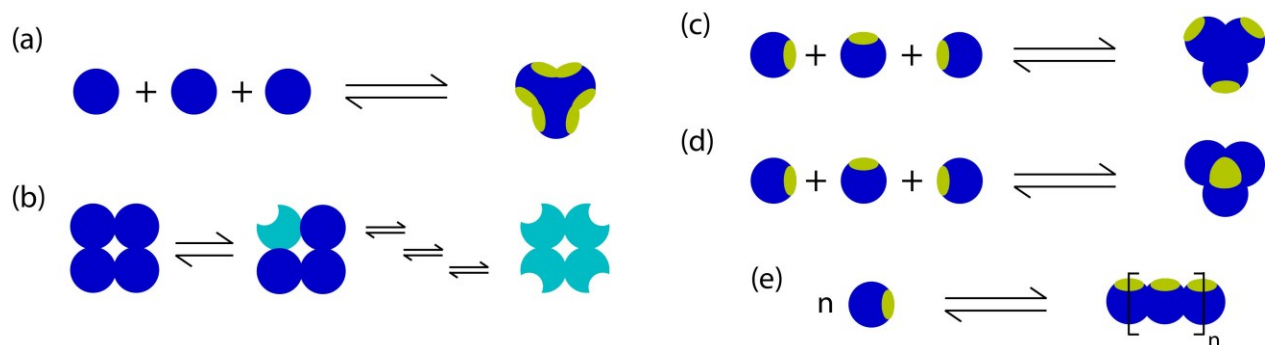
natures chimiques différentes<sup>24</sup>. Ainsi, la nature des forces moléculaires en présence n'est pas toujours aussi bien comprise (voir section 1.3).

Outre la formation d'assemblages moléculaires, les interactions réversibles entre un hôte et un ligand (*p. ex.* une enzyme et son substrat) jouent un rôle central dans divers processus biologiques en permettant la production d'un mouvement quasi-mécanique par le biais de l'énergie chimique de liaison. Les enzymes, ces nanomachines catalytiques du vivant, exploitent constamment l'énergie provenant de ces interactions où, en général, une simple fraction de l'énergie chimique provenant de la liaison du substrat à l'enzyme est suffisante pour accélérer la vitesse de la réaction chimique en abaissant la barrière d'activation<sup>25</sup>. Par exemple, l'apport énergétique provenant de la liaison chimique peut induire un mouvement quasi mécanique de l'enzyme qui changera sa conformation et positionnera donc tous les groupements chimiques impliqués dans la réaction les uns à côté des autres. Néanmoins, plusieurs de ces nanomachines catalytiques ont besoin d'une source d'énergie supplémentaire afin d'effectuer leur travail. Pour ce faire, ces nanomachines vont coupler une réaction exergonique, *c.-à-d.* une réaction chimique qui libère de l'énergie, avec la réaction chimique ou le travail qu'elles effectuent. Par exemple, la pompe sodium-potassium utilise l'énergie provenant de l'hydrolyse de l'ATP pour changer de conformation et ainsi transporter le sodium et le potassium à contresens de leurs gradients de concentration<sup>26</sup>. Même si l'utilisation de molécules externes comme source d'énergie devient de plus en plus populaire dans la conceptualisation de nanomachines artificielles<sup>27</sup>, nous focaliserons, dans le cadre de cette thèse, sur les liaisons chimiques non covalentes entre plusieurs molécules; et comment cette énergie peut être exploitée pour créer des interrupteurs moléculaires autoassemblés (voir section 1.2).

### **1.1.3 Les avantages des autoassemblages protéiques multimériques**

Dans les systèmes biologiques, les protéines agissent rarement de manière isolée. Plutôt, elles interagissent avec d'autres biomolécules, souvent d'autres protéines, afin de susciter une réponse biologique adéquate<sup>7</sup>. En effet, plus de 55% des protéines caractérisées jusqu'à présent comprennent plus qu'une sous-unité protéique<sup>5</sup>. Sous la forme la plus simple, la formation d'un complexe protéique à plusieurs composantes permet une activation ou une désactivation de sa

fonction; une conséquence directe de la dépendance de l'assemblage envers la concentration de ces constituants. Plus précisément, l'association de plusieurs sous-unités protéiques peut favoriser la formation d'un nouveau site actif à l'interface de ces mêmes sous-unités, permettre une régulation allostérique grâce à une communication par le biais de ces différentes sous-unités (*p. ex.* l'hémoglobine), augmenter la concentration locale de sites actifs en joignant plusieurs sous-unités possédant le même site actif ensemble, cacher un site actif ou bien favoriser leur polymérisation non-covalente (Fig. 1.2)<sup>4</sup>. De plus, lorsque vient le temps de créer de gros systèmes protéiques stables, il est préférable de le faire via l'association de plusieurs petites sous-unités pour éviter les problèmes de repliements d'une longue chaîne peptidique<sup>28</sup>. Par ailleurs, un plus grand nombre de composantes dans un assemblage permet de réduire considérablement la surface de protéines exposée au solvant<sup>22, 29</sup>. Cela a pour conséquence de minimiser la contribution entropique du solvant et ainsi produire des systèmes protéiques plus stables, car la stabilité des protéines est une fine équilibre entre le coût entropique provenant de la solvataion et des restrictions conformationnelles, et le gain enthalpique provenant de plusieurs interactions faibles<sup>29, 30</sup>.



**Figure 1.2 – Schématisation des avantages des autoassemblages protéiques multimériques.** (a) Formation de nouveaux sites actifs (en jaune) aux interfaces des composantes moléculaires (en bleu). (b) Régulation allostérique. (c) Augmentation de la concentration locale en sites actifs. (d) Inhibition par l'enterrement des sites actifs. (e) Création de larges assemblages moléculaires par polymérisation non covalente.

Un autre avantage d'avoir plusieurs composantes se retrouve d'un point de vue synthétique où avoir plusieurs petites sous-unités protéiques permet de réduire le nombre d'erreurs lors de la traduction<sup>22</sup>. Il fut estimé que le taux de mutations qui change la nature de l'acide aminé, est d'environ  $5 \times 10^{-4}$  erreur par codon<sup>31</sup>. Ainsi, une protéine de 2000 acides aminés



devrait typiquement avoir un acide aminé muté résultant possiblement en une protéine non fonctionnelle. Toutefois, si cette même protéine est constituée de quatre sous-unités de 500 acides aminés, chaque sous-unité possède une chance sur quatre d'avoir une mutation. Cela donnera donc trois sous-unités fonctionnelles qui pourront quand même s'autoassembler correctement. Évidemment, en extrapolant ce simple calcul à des dizaines de milliers d'acides aminés, il est possible de comprendre pourquoi, par exemple, les filaments d'actine proviennent de l'autoassemblage de plusieurs petites sous-unités d'actine d'environ 375 acides aminés au lieu d'une seule immense chaîne peptidique<sup>32</sup>. De plus, les complexes protéiques homomériques, *c.-à-d.* un complexe protéique construit avec la même sous-unité protéique, permettent une économie considérable du code génétique en réduisant la longueur du gène codant pour l'assemblage protéique<sup>33</sup>. D'un point de vue de la chimie de synthèse, les peptides synthétiques<sup>34</sup> et les polymères d'acides nucléiques<sup>35</sup> sont limités à une longueur maximale d'environ 50 acides aminés et 100 acides nucléiques respectivement, ce qui limite grandement la synthèse chimique de protéines ou d'ADN/ARN unimoléculaires fonctionnels, mais permet toutefois la création d'autoassemblages fonctionnels<sup>36, 37</sup>.

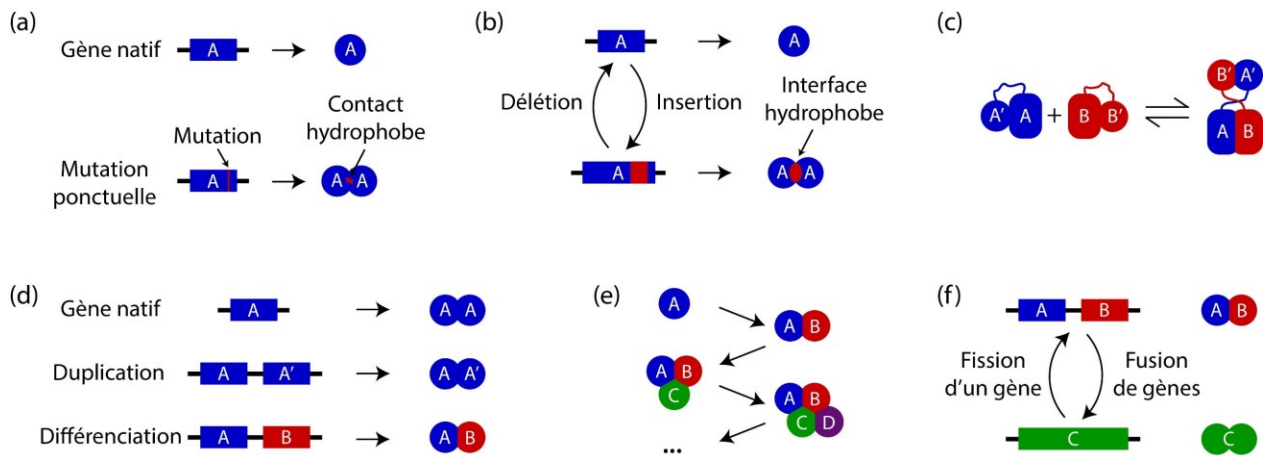
Néanmoins, les systèmes unimoléculaires ont aussi certains avantages par rapport aux complexes multimériques. Typiquement, les protéines unimoléculaires sont de plus petite taille que les complexes protéiques. Cela est dû au fait qu'il est difficile de former un petit assemblage moléculaire comprenant plusieurs composantes protéiques, car la surface et la force d'interaction entre les nombreuses composantes seraient trop faibles pour maintenir l'intégrité de la structure, surtout dans des conditions aussi drastiques que l'environnement encombré des cellules<sup>22</sup>. De plus, l'association entre les composantes de l'assemblage est dépendante de la concentration des composantes rendant ainsi les assemblages moléculaires instables à faible concentration<sup>4, 27</sup>. En contraste, l'activité des protéines unimoléculaires est indépendante de la concentration faisant en sorte que la majorité des systèmes protéiques fonctionnant à faible concentration sont monomériques<sup>22</sup>. De plus, en étant plus petite, la diffusion dans la cellule vers un site actif est plus rapide, ce qui explique pourquoi typiquement les protéines de transfert d'électrons, les hormones peptidiques, les hydrolases et plusieurs toxines sont unimoléculaires<sup>22</sup>.

Malgré tout, il demeure plusieurs points d'interrogation sur les avantages d'un système unimoléculaire comparé à un système identique construit avec plus qu'une composante. En étudiant des protéines de différentes espèces accomplissant des rôles similaires, mais ayant un niveau différent de multimérisation, aucune corrélation évidente ne fut observée entre le nombre de composantes et la performance d'une protéine. En effet, il existe plusieurs exemples où une protéine ayant moins de composantes surperforme comparativement à une protéine similaire ayant plus de composantes et *vice-versa*<sup>5</sup>. En revanche, ces protéines qui proviennent de différentes espèces ont évolué sous des pressions sélectives différentes faisant en sorte qu'il est difficile de les comparer uniquement du point de vue du nombre de composantes et de leurs structures<sup>6, 7</sup>. De plus, il existe un grand biais structurel dans les bases de données au niveau des protéines monomériques et homomériques, principalement à cause de leur facilité à cristalliser comparativement aux protéines hétéromériques<sup>1</sup>. Cependant, les récentes avancées en microscopie électronique permettent d'envisager une augmentation du nombre de structures hétéromériques, facilitant ainsi leur comparaison avec des structures homomériques ou monomériques similaires. Cinétiquement, il semble en général y avoir une relation directe entre la grosseur d'une protéine et sa vitesse de repliement<sup>28</sup>. Toutefois, si deux protéines de la même taille diffèrent uniquement par leur nombre de composantes (*p. ex.* un monomère et un dimère), moins d'évidences semblent pointer vers l'un ou l'autre comme étant plus rapides. En effet, le repliement d'un monomère est purement intramoléculaire tandis que les composantes d'un dimère, en étant deux fois plus petites, devraient se replier plus rapidement, mais leur vitesse d'association est quant à elle dépendante de la concentration des deux composantes rendant ainsi la comparaison plus difficile. En somme, la création d'un modèle moléculaire simplifié permettant uniquement d'étudier l'influence du nombre de composantes devrait grandement faciliter la réponse à ces questions.

#### **1.1.4 L'évolution vers des systèmes protéiques multimériques**

Il existe plusieurs voies évolutives permettant à une protéine d'acquérir de nouveaux partenaires d'interaction<sup>1, 6</sup>. La substitution d'acides aminés hydrophiles à la surface d'une protéine par des acides aminés hydrophobes par le biais de mutation ponctuelle est une méthode simple pour favoriser la formation d'une interface hydrophobe entre deux monomères (Fig. 1.3a).

Il fut démontré, par exemple, que changer seulement un acide aminé dans la séquence d'une protéine permet d'augmenter artificiellement le niveau de multimérisation de certaines protéines<sup>38</sup>. Outre la mutation ponctuelle, l'insertion et la délétion d'une portion d'acides aminés à la surface d'une protéine ou aux interfaces d'un complexe protéique peuvent aussi permettre de favoriser ou défavoriser la formation d'un assemblage (Fig. 1.3b)<sup>39</sup>. Similairement aux mutations ponctuelles, les acides aminés que l'on retrouve dans ces nouveaux domaines sont principalement hydrophobes ou aromatiques. Toutefois, comparativement à l'interface d'origine, ces domaines ont un léger biais envers des acides aminés polaires et chargés, ce qui semble illustrer un rôle important des interactions électrostatiques et polaires aux interfaces de ces assemblages. De plus, une protéine peut aussi être constituée de plusieurs sous-domaines qui peuvent s'échanger avec ceux d'autres monomères (Fig. 1.3c)<sup>40</sup>. Ainsi, en échangeant leurs domaines, ces monomères peuvent former des complexes dimériques, mais aussi permettre l'agrégation de la protéine en de larges complexes.



**Figure 1.3 – Schématisation de certains mécanismes d'évolution menant à la formation de systèmes protéiques multimériques.** (a) Mutations ponctuelles. (b) Insertion et délétion de domaines. (c) Échange de domaines. (d) Duplication et différenciation. (e) Évolution séquentielle. (f) Fission d'un gène et fusion de gènes.

Un autre mécanisme évolutif commun permettant l'acquisition de nouvelles sous-unités dans un autoassemblage protéique implique la duplication de gènes (Fig. 1.3d)<sup>1</sup>. Si, par exemple, un dimère homomérique voit le gène qui code pour ses deux sous-unités être dupliqué, il est possible que l'un de ces deux gènes, suivant plusieurs mutations, se différencie. Ainsi, cela donne deux gènes codant pour deux séquences protéiques différentes résultant en la formation d'un

dimère hétéromérique où l'activité de ce dimère peut maintenant être régulée via ses deux différents monomères. Malgré les évidences envers ce mécanisme évolutif, plusieurs complexes hétéromériques chez les bactéries, les archées et les eucaryotes semblent simplement avoir évolué par l'acquisition séquentielle de partenaires d'interaction, *c.-à-d.* un partenaire à la fois (Fig. 1.3e)<sup>2</sup>. En effet, il est peu probable qu'une simple protéine, par exemple, ait acquis trois nouveaux partenaires d'interactions différents simultanément pour ainsi former un tétramère hétéromérique. L'acquisition de nouveaux partenaires d'interaction peut aussi se faire par la fission ou la fusion d'un gène (Fig. 1.3f)<sup>1</sup>. Dans le cas de la fission, un gène se retrouve fragmenté en deux, produisant ainsi deux protéines distinctes pouvant interagir pour reformer la structure d'origine. Inversement, deux fragments de gène peuvent se combiner pour réduire le nombre de composantes dans un assemblage. Fait intéressant, les événements de fusion sont quatre fois plus communs que les événements de fission<sup>41</sup>.

Tous ces différents mécanismes évolutifs démontrent à quel point l'acquisition de nouveaux partenaires d'interaction est, au niveau évolutif, un processus fastidieux qui implique souvent plusieurs étapes et plusieurs milliers d'années d'évolution. Néanmoins, cela offre des pistes d'inspiration aux ingénieurs moléculaires qui aspirent à créer des assemblages comprenant plusieurs molécules. En effet, nous verrons plus tard comment les spécialistes en nanotechnologie d'ADN sont parvenus, en s'inspirant des protéines et de leur évolution, à créer des autoassemblages moléculaires dynamiques à base d'ADN (section 1.3.4). De plus, nous explorerons dans cette thèse comment la fission d'une nanostructure monomérique (chapitre 2) et la fusion de domaines interagissant (chapitre 3) permettent la création d'assemblages trimériques avec des propriétés d'assemblage améliorées et programmables.

## **1.2 Autoassemblage d'interrupteurs moléculaires dynamiques**

Autant que les interactions entre diverses composantes d'un autoassemblage biologique soient importantes pour susciter une réponse biologique appropriée, ces mêmes interactions peuvent être exploités dans la conceptualisation d'autoassemblages artificiels et dynamiques. Par exemple, en exploitant le changement d'état accompagné par l'interaction entre deux ou plusieurs molécules, il est possible de mettre au point des interrupteurs moléculaires

autoassemblés utiles pour la création de biocapteurs, pour le développement de systèmes de délivrance de médicaments, pour l'imagerie moléculaire ou pour le théranostique<sup>24, 42</sup>. Grâce à un système simple d'interrupteur composé d'un récepteur moléculaire et de son ligand, nous allons voir les concepts derrière la création d'un interrupteur moléculaire et comment appliquer ces mêmes principes pour la création d'interrupteurs moléculaires autoassemblés.

### **1.2.1 Les interrupteurs moléculaires autoassemblés**

La capacité de certains systèmes moléculaires à alterner entre deux états distincts est cruciale à la régulation et au maintien de certaines fonctions biologiques. En étant capables de sonder et d'interagir avec leur environnement, les cellules peuvent répondre adéquatement aux perturbations par une réponse biologique appropriée. Les récepteurs couplés aux protéines G (RCPG), par exemple, sont des récepteurs transmembranaires responsables de la communication entre le milieu intra- et extracellulaire<sup>43</sup>. Cette famille de récepteurs peut interagir avec une grande variété d'agonistes se trouvant dans le milieu extracellulaire incluant des ions, des petites molécules organiques et même des protéines. L'interaction entre le récepteur et l'agoniste induit un changement de conformation à travers la membrane entraînant la liaison, dans le milieu intracellulaire, de la sous-unité  $\alpha$  de la protéine G induisant ainsi une voie de signalisation en aval pour répondre au stimulus<sup>44</sup>.

Cette capacité à tirer profit de l'énergie d'interaction entre un récepteur et son ligand (*p. ex.* l'agoniste) pour créer un mouvement quasi-mécanique est primordiale dans la conceptualisation de systèmes dynamiques. Ce mouvement est produit lors du changement réversible entre deux conformations du récepteur. Il est important que chacune de ces deux conformations soit stable puisque cette bistabilité assure que la présence du ligand favorise l'interconversion entre ces deux conformations<sup>42</sup>. Puisque l'activité du récepteur est typiquement corrélée seulement à l'une de ces deux conformations, on décrit souvent ces systèmes comme étant des interrupteurs moléculaires puisqu'ils alternent entre un état inactif (« OFF ») et un état actif (« ON »). Inspiré par ces systèmes biologiques tels que les RCPGs, une multitude d'interrupteurs moléculaires artificiels ont été mis au point ces dernières années. Des facteurs de transcription, par exemple, furent évolués artificiellement par évolution dirigée pour générer des

mutants ayant une spécificité d'interaction différente de la protéine native<sup>45, 46</sup>. Les riborégulateurs, de l'anglais « riboswitch », sont aussi couramment utilisés pour leur capacité à interagir avec divers ligands qui, une fois liés, induisent un changement de conformation de la plateforme d'expression génique<sup>47</sup>. Similairement, des brins d'ADN ou d'ARN peuvent être artificiellement sélectionnés pour leur capacité à changer de conformation lorsqu'ils interagissent avec une molécule donnée<sup>48</sup>, voir même à catalyser spécifiquement une réaction chimique<sup>49</sup>. Diverses molécules organiques furent aussi synthétisées pour leur capacité à changer de conformation comme des azobenzènes<sup>50</sup>, des rotaxanes<sup>51</sup>, des éthers couronnes<sup>52</sup> ou bien des polymères sensibles<sup>53</sup>.

Ces interrupteurs moléculaires peuvent interagir avec une multitude de différents stimulus pouvant prendre différentes formes allant de variations physiques (température, pression, lumière, etc.), à des variations chimiques (pH, ions métalliques, molécules organiques, etc.) ainsi que des variations biochimiques (lipides, saccharides, protéines, etc.)<sup>24, 54, 55</sup>. Toutefois, pour obtenir un interrupteur fonctionnel, il faut que le changement de conformation provenant de la détection de ce stimulus permette la production d'un signal détectable et mesurable. Biologiquement parlant, la détection du stimulus peut activer des canaux ioniques<sup>56</sup>, activer la dégradation de protéines<sup>57</sup>, induire l'isomérisation de molécules dans les photorécepteurs de l'œil<sup>58</sup> ou bien enclencher des réactions enzymatiques<sup>59</sup>. Du point de vue de l'ingénierie moléculaire, les interrupteurs moléculaires sont principalement utilisés dans la création de biocapteurs et de systèmes de délivrance de médicaments. Dans le premier cas, l'interaction d'un analyte avec un interrupteur induit un changement au niveau de ce dernier permettant la production d'un signal quantifiable proportionnel à la concentration de l'analyte par le biais, par exemple, d'un changement de propriétés optiques, de fluorescence, de propriétés plasmoniques ou bien d'activité électrochimique de l'interrupteur<sup>24, 60</sup>. Dans le second cas, l'interaction entre un stimulus biologique (*p. ex.* un gradient de pH, un anticorps, une protéine membranaire, etc.) et un interrupteur permet, via un changement de conformation, la libération d'un médicament encapsulé à l'intérieur de ce dernier<sup>61</sup>. La transduction d'un stimulus en un mouvement quasi mécanique par l'intermédiaire d'un interrupteur moléculaire est donc simplement limitée par

l'imaginaire d'une personne à conceptualiser un récepteur moléculaire dynamique pour sa cible d'intérêt<sup>24</sup>.

Les assemblages moléculaires peuvent aussi agir comme interrupteur moléculaire où la présence d'un stimulus induit l'assemblage de toutes les composantes du système en une structure bien définie. La présence d'un stimulus peut aussi, inversement, déstabiliser cette structure, libérant toutes ces composantes moléculaires. Il existe donc deux états bien définis pour un système autoassemblé, soit l'état dissocié (« OFF ») et l'état assemblé (« ON »). Certains peptides amphiphiles, par exemple, peuvent s'autoassembler spontanément en fibrille lorsque le pH devient acide et, inversement, se dissocier en milieu neutre ou basique<sup>62</sup>. Divers nanomatériaux autoassemblés ont aussi, similairement, été mis au point pour leurs propriétés réversibles sous l'influence de stimulus externes incluant des assemblages supramoléculaires<sup>63</sup>, des nanoparticules<sup>64</sup> et des polymères auto-cicatrisants<sup>65</sup>. Nous verrons aussi plus tard comment l'ADN peut être utilisé afin de créer des autoassemblages dynamiques et réversibles pouvant servir comme interrupteurs moléculaires (voir section 1.3.1). Dans la présente thèse, nous nous attarderons uniquement à la conceptualisation d'autoassemblages à base d'ADN afin de programmer et optimiser la plage dynamique d'interrupteurs moléculaires.

### **1.2.2 Caractérisation d'une courbe de liaison : l'exemple d'un récepteur et son ligand**

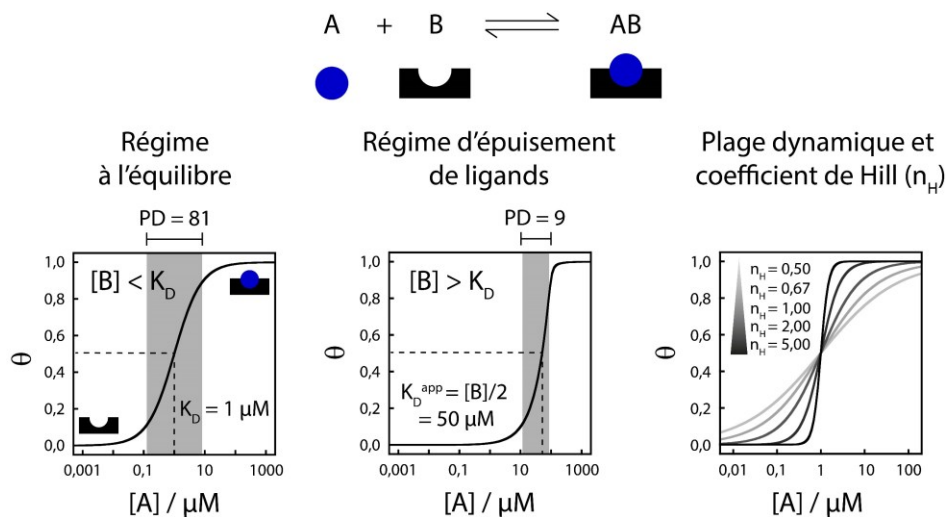
L'interaction entre un récepteur moléculaire et son ligand se caractérise par l'équilibre chimique suivant :  $A + B \rightleftharpoons AB$  où A est le ligand, B est le récepteur moléculaire et AB est le complexe formé par l'interaction de ces deux molécules (Fig. 1.4). Cela se traduit par la constante d'équilibre d'association suivante :  $K_A = [AB]/[A][B]$ . Typiquement, une telle réaction à l'équilibre est de deuxième ordre, *c.-à-d.* que la vitesse de formation du complexe dépend autant de la concentration du récepteur moléculaire que du ligand<sup>66</sup>. Expérimentalement, l'on mesure l'interaction d'un récepteur pour son ligand lors d'une expérience de titrage où le ligand est ajouté à une quantité fixe de récepteur jusqu'à la saturation de ce dernier. Autant les chimistes préfèrent exprimer les équilibres chimiques en termes d'association, les biochimistes préfèrent exprimer les équilibres chimiques en termes de dissociation :  $K_D = [A][B]/[AB]$ , ce qui correspond

à l'inverse de la constante d'équilibre d'association ( $K_D = 1/K_A$ ). La raison principale derrière cette préférence est qu'en exprimant l'équilibre en termes de dissociation, la constante d'équilibre de dissociation ( $K_D$ ), maintenant exprimée en unité de concentration, peut facilement être déterminée lors du titrage. En effet, visuellement, le  $K_D$  correspond à la concentration de ligand nécessaire pour lier 50% du récepteur (Fig. 1.4, *gauche*). Toutefois, pour que cette observation visuelle demeure vraie, il faut s'assurer que la concentration de récepteur demeure inférieure au  $K_D$  de la réaction d'équilibre lors du titrage ( $[B] < K_D$ ). Si, à l'inverse, la concentration de récepteur est plus grande que le  $K_D$  ( $[B] > K_D$ ), alors le système passe d'un régime de liaison à l'équilibre vers un régime d'épuisement de ligand, de l'anglais « ligand-depletion regime »<sup>67, 68</sup>. Dans un tel régime, il n'y a pas d'équilibre apparent et le ligand lie le récepteur au fur et à mesure que du ligand est rajouté dans la solution lors du titrage. Cela se traduit par un régime de liaison 1:1 où la concentration de ligand nécessaire pour lier 50% du récepteur n'est plus égale au  $K_D$ , mais plutôt à la moitié de la concentration de récepteurs (Fig. 1.4, *centre*). Malgré tout, cette valeur est représentée comme la constante d'équilibre de dissociation apparente ( $K_D^{app}$ ) du système puisqu'elle représente quand même la concentration de ligand menant à 50% de liaison, mais ne représente pas la vraie valeur de la constante d'équilibre.

Dans un régime de liaison à l'équilibre, la courbe de formation du complexe (AB) normalisée par le récepteur (B) se traduit par l'équation suivante :  $\theta = [AB]_{eq}/[B]_T = [A]_T/([A]_T + K_D)$ , qui représente le pourcentage de récepteur lié ( $\theta$ ) lorsqu'on ajoute du ligand (A) lors du titrage<sup>69</sup>. Il y a deux paramètres importants qui permettent de décrire cette courbe de liaison. Premièrement le  $K_D$  qui, comme mentionné, correspond au point milieu de la courbe ou la concentration à laquelle 50% du récepteur est lié. Deuxièmement, il y a la plage dynamique (PD) qui correspond à l'étendue de la courbe, *c.-à-d.* la fenêtre de concentrations sur laquelle la transition survient. En général, la plage dynamique se calcule par le ratio entre la concentration de ligand menant à 90% de liaison et la concentration de ligand menant à 10% de liaison ( $PD = [A]_{90\%}/[A]_{10\%}$ ). Pour un équilibre classique entre un récepteur et son ligand, la plage dynamique est de 81 (Fig. 1.4, *gauche*). Toutefois, la plage dynamique peut parfois être plus étroite ou plus large. Ce comportement fut, par exemple, observé il y a plus de 100 ans lors de la liaison de l'oxygène à l'hémoglobine. Dans ce cas précis, la plage dynamique était beaucoup plus faible que



81 et un nouveau modèle de liaison devenait nécessaire pour interpréter ces résultats. Ainsi, A. V. Hill proposa une équation empirique permettant d'analyser ce type de courbe coopératif. L'équation est similaire à celle de la courbe de liaison, toutefois, un exposant est ajouté à chacun des termes de l'équation :  $\theta = [A]_T^{n_H} / ([A]_T^{n_H} + K_D^{n_H})$  où  $n_H$  est le coefficient de Hill<sup>70</sup>. Selon ce modèle, la valeur de  $n_H$  devrait correspondre au nombre de sites de liaison. Toutefois, cela est uniquement vrai s'il existe une coopérativité parfaite entre le premier événement de liaison et les liaisons subséquentes<sup>70</sup>. Donc, expérimentalement, il est rare que le coefficient de Hill corresponde précisément au nombre de sites de liaison. Malgré tout, cette équation est couramment utilisée de manière empirique pour déterminer le niveau de sensibilité lors de la liaison de ligands, car la plage dynamique et le coefficient de Hill sont directement reliés selon :  $PD = 81^{1/n_H}$  (Fig. 1.4, *droite*). Par exemple, dans un régime d'épuisement de ligands, la plage dynamique est plus faible (PD = 9) et se traduit par un coefficient de Hill de 2 (Fig. 1.4, *centre*).



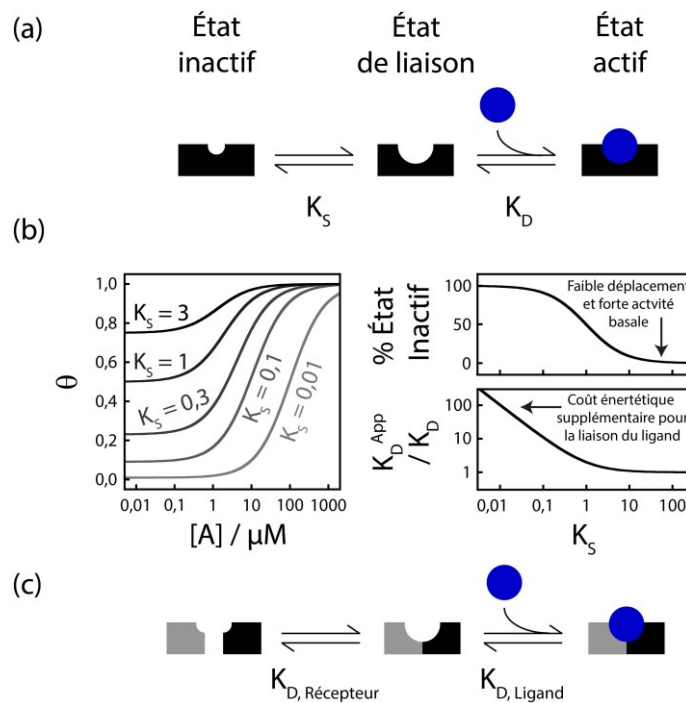
**Figure 1.4 – Courbe de liaison entre un récepteur et son ligand.** L'équilibre d'interaction entre un récepteur (B) et son ligand (A) peut se caractériser avec une expérience de titrage où diverses quantités de ligand sont ajoutées à une quantité fixe de récepteur. *Gauche*. Dans un régime à l'équilibre, *c.-à-d.* que la concentration de récepteur est inférieure au  $K_D$ , la transition sigmoïdale est centrée sur le  $K_D$  de la réaction et possède une plage dynamique (PD) de 81. Dans cet exemple,  $K_D = 1 \mu\text{M}$  et  $[B] = 0,01 \mu\text{M}$ . *Centre*. Dans un régime d'épuisement de ligands, *c.-à-d.* que la concentration de récepteur est supérieure au  $K_D$ , la transition sigmoïdale survient à plus haute concentration et ne permet pas de déterminer le  $K_D$  de la réaction d'équilibre. Dans ce cas, la transition est centrée sur la moitié de la concentration de récepteur et possède une plage dynamique beaucoup plus petite (PD = 9). Dans cet exemple,  $K_D = 1 \mu\text{M}$  et  $[B] = 100 \mu\text{M}$ . *Droite*. La coopérativité de liaison peut se caractériser par l'équation de Hill où le coefficient de Hill ( $n_H$ ) est directement relié à la plage dynamique de la transition.

L'équilibre chimique démontré précédemment entre un récepteur et son ligand sert comme un exemple simple pour illustrer les principes de base derrière l'analyse d'un équilibre chimique. Toutefois, dans le cadre de cette thèse, les systèmes à l'étude impliqueront plus que deux molécules. Plus spécifiquement, les assemblages moléculaires à l'étude sont composés de trois molécules selon l'équilibre chimique suivant :  $A + B + C \rightleftharpoons ABC$ . Les courbes de titrage seront donc réalisées en titrant la composante A de l'assemblage à des concentrations fixes des composantes B et C. Cet assemblage sera suivi principalement par spectroscopie de fluorescence où l'induction de l'assemblage par l'ajout de la composante A modifiera la fluorescence globale du système. Plus précisément, un extincteur de fluorescence sera chimiquement attaché à l'une des composantes (*p. ex.* A) qui, à la suite de l'assemblage, se trouvera près d'un fluorophore attaché à l'autre composante (*p. ex.* B ou C) entraînant ainsi une diminution de la fluorescence par transfert d'énergie (FRET) entre le fluorophore et l'extincteur. Les courbes obtenues par cette titration seront analysées en utilisant l'équation empirique de Hill pour quantifier le point milieu de la courbe et la plage dynamique.

### 1.2.3 Le mécanisme de déplacement de population

Bien qu'un interrupteur se retrouve principalement dans l'un de ces deux états (inactif ou actif), il est typiquement considéré que les interrupteurs moléculaires suivent un mécanisme où l'interaction entre l'interrupteur et son stimulus moléculaire (*p. ex.* un ligand) suivent un déplacement de population faisant intervenir trois états<sup>71-73</sup>. Plus spécifiquement, ce modèle fait intervenir deux équilibres chimiques entre trois états distincts, soit l'état inactif, l'état de liaison et l'état actif (Fig. 1.5a). En absence de ligand, l'interrupteur se retrouve majoritairement dans l'état inactif avec une faible fraction de la population d'interrupteurs qui adopte une conformation optimale pour la liaison du ligand, soit l'état de liaison. L'équilibre chimique entre ces deux états est défini par la constante d'équilibre d'interconversion  $K_S$ , soit le ratio entre la quantité d'interrupteur dans l'état de liaison et la quantité d'interrupteur dans l'état inactif. En présence de ligand, un déplacement de population survient où l'interaction entre le ligand et l'interrupteur ( $K_D$ ) stabilise l'état actif de l'interrupteur. Pour obtenir un bon déplacement de population de l'état inactif vers l'état actif, il est important que le  $K_S$  demeure inférieur à 0,1<sup>71</sup>. Cela assure que l'interrupteur se retrouve majoritaire dans l'état inactif permettant, ainsi, un

important changement de population en présence du ligand et un faible niveau d'activation de base de l'interrupteur. Toutefois, en surstabilisant l'interrupteur dans son état inactif ( $K_S \ll 0,1$ ), le ligand devra payer un prix énergétique plus important se traduisant par une augmentation de la concentration de ligands nécessaire pour induire l'activation de l'interrupteur ( $K_D^{App}$ ) par rapport à l'affinité intrinsèque entre ce ligand et l'interrupteur ( $K_D$ ) (Fig. 1.5b). Il est donc primordial de bien sélectionner l'interrupteur moléculaire pour tenir compte de cette augmentation apparente du  $K_D$ .



**Figure 1.5 – Schématisation du mécanisme de déplacement de population.** (a) Le mécanisme de déplacement de population fait intervenir trois états soit : 1) l'état inactif où le récepteur possède une affinité réduite envers le ligand, 2) l'état de liaison où le récepteur affiche une affinité augmentée envers le ligand et 3) l'état actif où le ligand est lié au récepteur. (b) Lorsque le récepteur est plus stable dans son état inactif ( $K_S < 1$ ), le déplacement de population est maximisé, mais survient au détriment d'une augmentation significative de la concentration nécessaire de ligand pour activer le récepteur ( $K_D^{App} > K_D$ ). Lorsque le récepteur est plus stable dans son état de liaison ( $K_S > 1$ ), l'amplitude du déplacement de population est considérablement réduite tout en ayant une forte activité basale du récepteur. Toutefois, le  $K_D$  demeure inchangé. (c) Le mécanisme de déplacement de population peut aussi être adapté pour la formation d'assemblages moléculaires à plusieurs composantes où un ligand permet d'induire un déplacement de population de l'état non assemblé vers l'état assemblé.

Toutefois, il est probable que l'interrupteur choisi n'ait pas une affinité envers le ligand qui soit dans la plage dynamique désirée. Ainsi, il est nécessaire d'optimiser cette plage

dynamique en stabilisant ou déstabilisant l'interrupteur afin de déplacer la plage dynamique vers des concentrations plus faibles ou plus élevées. De plus, il pourrait aussi être nécessaire de programmer la sensibilité de l'interrupteur, pour détecter le ligand sur une plage dynamique plus fine ou étendue afin de mieux représenter la plage de concentration du ligand. Nous verrons plus tard comment exploiter divers mécanismes biologiques afin de programmer précisément la plage dynamique d'un interrupteur (voir section 1.4). Malgré tout, une méthode simple pour programmer la plage dynamique consiste, comme mentionnée précédemment, à optimiser le  $K_S$ <sup>71</sup>. Toutefois, en stabilisant l'état non liant, il est uniquement possible de déplacer la plage dynamique vers des concentrations plus élevées (Fig. 1.5b). Afin de réduire la plage dynamique vers de basses concentrations, il est généralement nécessaire de procéder à une série de mutations et de modifications au niveau de l'interrupteur afin d'augmenter son affinité pour le ligand<sup>74-76</sup>. Nous verrons, toutefois, qu'il existe des mécanismes moléculaires employés par la Nature permettant d'éviter ce long processus d'optimisation de la plage dynamique (voir section 1.4). La sensibilité, *c.-à-d.* comment sensible est l'activation de l'interrupteur face à des variations importantes en concentration de ligand, peut être programmée en optimisant la coopérativité d'interaction entre l'interrupteur et le ligand. Nous verrons, dans le paragraphe suivant, comment exploiter cette coopérativité d'interaction et ainsi l'utiliser pour optimiser l'interaction entre diverses molécules. Finalement, même si le mécanisme de déplacement de population fut représenté par l'interaction entre un seul récepteur et son ligand, il est aussi possible de représenter ce même mécanisme dans le cadre d'un assemblage moléculaire où le récepteur est constitué de plus d'une molécule (Fig. 1.5c). Par exemple, le récepteur peut être constitué de deux composantes qui ont une faible affinité pour le ligand (état inactif), mais qui peuvent interagir afin de former un assemblage ayant une affinité accrue pour le ligand (état de liaison). La différence avec le récepteur unique est que maintenant l'équilibre unimoléculaire d'interconversion ( $K_S$ ) devient un équilibre de dimérisation entre deux composantes ( $K_{D, \text{récepteur}}$ ).

#### **1.2.4 La coopérativité des assemblages moléculaires**

L'autoassemblage de diverses molécules entre elles pour former un système moléculaire plus complexe nécessite l'intervention clé d'interactions coopératives permettant le lien entre les composantes moléculaires du système et ces propriétés<sup>77</sup>. Cette coopérativité provient d'un effet

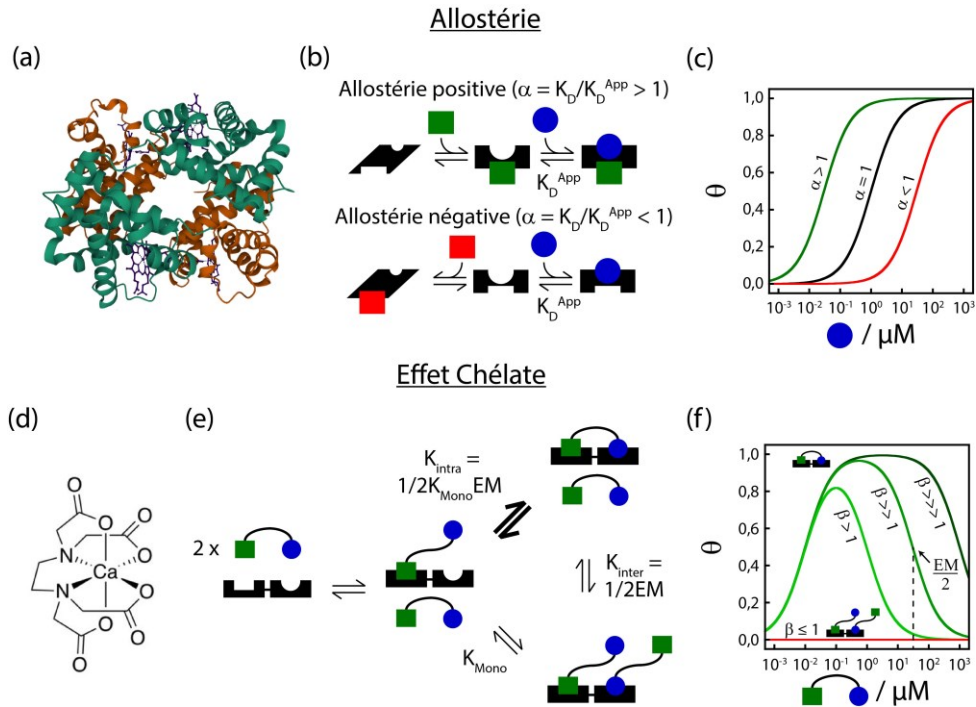
additif entre deux ou plusieurs interactions moléculaires qui, dans leur ensemble, possèdent un comportement différent de celui attendu par la somme de chacune des interactions individuelles<sup>77-79</sup>. Cette coopérativité peut provenir de divers phénomènes moléculaires rendant ainsi sa définition complexe. Un biochimiste entendra probablement parler pour la première fois de coopérativité lors de l'étude du fonctionnement de l'hémoglobine avec laquelle l'interaction d'une première molécule de O<sub>2</sub> augmentera la force d'interaction de la seconde, et ce, jusqu'à ce que les quatre sites soient occupés. À l'inverse, un chimiste entendra plutôt parler de coopérativité dans le contexte de complexation d'un ligand multivalent avec un métal (*p. ex.* la complexation du calcium par l'éthylènediaminetétraacétique) où la première liaison de coordination facilitera les prochaines liaisons. Dans ces deux cas classiques, le principe reste le même, c'est-à-dire qu'une première interaction aura un effet positif ou négatif sur la suivante et ainsi de suite. Cela peut provenir, par exemple, d'un effet hydrophobe accentué, d'une répulsion ou d'une attraction électrostatique plus prononcée, de l'apparition d'un nouveau site d'interaction ou simplement d'une orientation plus précise de ponts hydrogènes, de ponts salins ou d'empilements aromatiques<sup>78</sup>. Il existe deux principaux types de mécanismes permettant d'introduire de la coopérativité dans un système moléculaire, soit l'allostérie et l'effet chélate (aussi connue sous les termes de coopérativité configurationnel ou de préorganisation)<sup>77-80</sup>. Il est à noter que ces deux mécanismes ne sont pas mutuellement exclusifs et qu'ils peuvent être tous les deux présents dans le même nanosystème<sup>81, 82</sup>.

#### **1.2.4.1 L'allostérie : une question de distance**

Le mécanisme d'allostérie survient lorsque l'interaction d'un premier ligand stabilise une conformation différente du récepteur modifiant ainsi l'affinité du second ligand pour le récepteur<sup>77</sup>. C'est le type de coopérativité qui intervient, par exemple, lors de la liaison des quatre unités d'oxygène avec l'hémoglobine (Fig. 1.6a). Il est généralement commun d'interpréter le comportement des protéines allostériques en suivant le modèle concerté de Monod–Wyman–Changeux (MWC) ou le modèle séquentiel de Koshland–Nemethy–Filmer (KNF)<sup>83-85</sup>. Dans ces deux modèles, les sous-unités d'une protéine peuvent adopter deux états conformationnels différents nommés T (faible affinité envers le ligand) et R (forte affinité envers le ligand). Dans le modèle MWC, toutes les sous-unités existent simultanément dans le même état conformationnel

et c'est l'interaction avec les ligands qui déplace l'équilibre entre ces deux états augmentant ainsi l'affinité pour un prochain ligand. À l'inverse, dans le modèle KNF, le ligand induit le changement de conformation entre l'état T et R uniquement à la sous-unité qu'il liera, augmentant ainsi l'affinité. Toutefois, des observations récentes ont démontré qu'une augmentation d'affinité n'est pas toujours accompagnée d'un changement conformationnel. En effet, la communication entre deux sites de liaison éloignés les uns des autres peut se produire grâce à un changement de propriétés dynamiques du récepteur<sup>86</sup>, par l'intermédiaire de régions désordonnées<sup>87</sup>, une modification de densité électronique<sup>88</sup> ou par une modification du paysage énergétique<sup>89</sup>. Nonobstant les différents mécanismes derrière l'apparition d'un comportement allostérique, l'interaction entre l'effecteur allostérique et le récepteur se traduit toujours par une diminution (inhibiteur allostérique) ou une augmentation (activateur allostérique) apparente de l'affinité du ligand pour le récepteur.

Pour mieux illustrer cette modification de l'affinité par un mécanisme allostérique, considérons le cas simple d'un récepteur contenant deux sites de liaison; l'un pour le ligand et l'autre pour l'effecteur allostérique (Fig. 1.6b)<sup>24, 90</sup>. L'impact d'un effecteur allostérique sur la liaison du ligand se caractérise par rapport à l'affinité intrinsèque du ligand pour le récepteur ( $K_D$ , courbe noire). Toutefois, si l'effecteur allostérique augmente l'affinité pour le ligand en stabilisant l'état de liaison, la courbe de liaison du ligand se déplacera vers de plus faibles concentrations ( $K_D^{App} < K_D$ ) résultant en une coopérativité positive (Fig. 1.6c, courbe verte). Inversement, si l'effecteur allostérique diminue l'affinité pour le ligand en stabilisant, par exemple, l'état inactif du récepteur, la courbe de liaison du ligand se déplacera vers de plus grandes concentrations ( $K_D^{App} > K_D$ ) résultant en une coopérativité négative (Fig. 1.6c, courbe rouge). Quantitativement, on définit l'impact d'un effecteur allostérique par le paramètre  $\alpha$ , qui correspond au ratio entre l'affinité en présence de l'effecteur allostérique et l'affinité intrinsèque, chacun exprimé en constante d'association ( $\alpha = K_A^{App} / K_A$  ou  $\alpha = K_D / K_D^{App}$ , si exprimé en constante de dissociation). Dans ce cas, une valeur de 1 indique l'absence de coopérativité, une valeur inférieure à 1 indique une coopérativité négative et une valeur supérieure à 1 indique une coopérativité positive. Nous discuterons aussi, plus tard, de l'allostérie dans une perspective du mécanisme de déplacement de population (voir section 1.4.2).



**Figure 1.6 – Schématisation des mécanismes moléculaires menant à de la coopérativité.** (a) Le mécanisme d’allostérie survient lorsque le premier événement de liaison augmente (ou diminue) la force d’interaction des événements de liaison subséquents comme c’est le cas avec la liaison de l’oxygène aux quatre sites de liaison de l’hémoglobine. (b) L’allostérie peut être positive ( $\alpha > 1$ ), négative ( $\alpha < 1$ ) ou nulle ( $\alpha = 1$ ). (c) Une allostérie positive (courbe verte) se traduit par une diminution du  $K_D^{App}$  tandis qu’une allostérie négative (courbe rouge) se traduit par une augmentation du  $K_D^{App}$ . (d) Le mécanisme de l’effet chélate survient avec des ligands multivalents où le premier événement de liaison rapproche dans l’espace les événements de liaison subséquents comme c’est le cas avec la complexation du calcium par l’EDTA. (e) À la suite du premier événement de liaison entre un ligand bivalent et un récepteur bivalent, *c.-à-d.* la liaison du carré vert, le second site de liaison du récepteur peut être occupé par le même ligand suivant une réaction intramoléculaire (équilibre supérieur en gras) ou peut être occupé par un autre ligand bivalent (équilibre inférieur). (f) Un effet chélate positif permet la formation du complexe fermé sur une plus grande plage dynamique de ligand (courbes vertes). Un effet chélate négatif, en comparaison, défavorise la formation du complexe fermé (courbe rouge). La structure de l’hémoglobine provient de la « Protein Data Bank » (rcsb.org, PDB ID: 1A3N)<sup>91</sup>.

### 1.2.4.2 L’effet chélate : une question de proximité

Le mécanisme d’effet chélate survient lorsque la première interaction d’un ligand bidentate avec un récepteur rapproche dans l’espace le second site de liaison augmentant ainsi la force d’interaction<sup>77</sup>. C’est le type de coopérativité qui intervient, par exemple, lors de la liaison de l’EDTA avec le calcium (Fig. 1.6d). À l’instar de l’allostérie, le phénomène de l’effet chélate implique des ligands polyvalents, *c.-à-d.* des ligands qui possèdent plusieurs sites de liaison avec

le récepteur. Il est généralement commun d'interpréter l'effet chélate par une augmentation de la concentration molaire effective (EM, de l'anglais « *effective molarity* »). En d'autres mots, lorsqu'une première interaction entre le ligand bidentate et le récepteur se produit, le second site de liaison du ligand se retrouve maintenant à proximité du récepteur faisant en sorte que ce récepteur s'expose à une « concentration plus élevée » du ligand dans son environnement moléculaire. Cette préorganisation du ligand permet de réduire le coût entropique de la seconde liaison puisque la première liaison a déjà payé le coût énergétique en réduisant le nombre de configurations non-productives que le ligand et le récepteur peuvent adopter<sup>77, 82, 92</sup>. À cause de cette réduction du coût entropique accompagné par une contraction de la structure qui augmente le gain enthalpique de liaison<sup>93</sup>, l'affinité globale du ligand bivalent envers le récepteur se retrouve augmentée.

Pour mieux illustrer cette modification de l'affinité par l'effet chélate, considérons le cas simple d'un récepteur moléculaire contenant deux sites de liaison pour un ligand bidentate (Fig. 1.6e). À la différence de l'allostérie, les deux sites de liaison du ligand sont maintenant interconnectés ensemble par l'entremise d'un espaceur. Ainsi, il y a une possibilité que le récepteur puisse lier deux différents ligands bivalents et ainsi multimériser au lieu de former le système cyclique composé d'un seul ligand bivalent et d'un seul récepteur. Quantitativement, on définit la tendance du système à fermer le cycle, et donc l'impact de l'effet chélate, par l'équilibre intramoléculaire de fermeture du cycle (Fig. 1.6e, équilibre supérieur). Toutefois, la liaison entre un récepteur monovalent et un ligand monovalent est de nature intermoléculaire et est donc définie par sa propre constante d'équilibre d'association ( $K_{\text{Mono}}$ ) qui ne représente malheureusement pas la nature intramoléculaire de l'interaction menant à la fermeture du cycle<sup>80, 88</sup>. Ainsi, il est important d'introduire un facteur de conversion pour convertir cette constante d'équilibre intermoléculaire en une constante d'équilibre intramoléculaire. Ce paramètre est la concentration molaire effective (EM), exprimée en  $\text{mol}\cdot\text{L}^{-1}$ , et fut introduit la première fois pour décrire la catalyse de réaction intramoléculaire<sup>94</sup>.

Afin de décrire la coopérativité de l'effet chélate avec un facteur similaire au paramètre  $\alpha$  de l'allostérie, il fut proposé d'utiliser la constante d'équilibre de la réaction intramoléculaire ( $K_{\text{intra}}$ ) soit  $\beta$  où  $\beta = K_{\text{Mono}}\cdot\text{EM}$  (Fig. 1.6e, équilibre supérieur)<sup>79</sup>. À noter que le facteur statistique



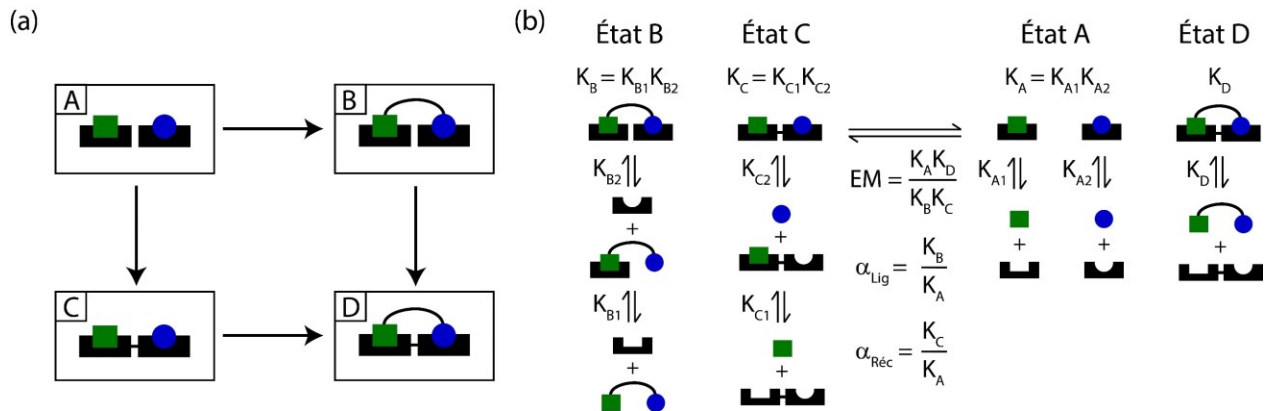
de cyclisation  $\frac{1}{2}$  est omis dans la définition de  $\beta$ , car ce dernier se retrouve imbriqué indirectement dans la valeur de EM. Toutefois, il fut aussi démontré qu'à l'instar de l'allostérie, la coopérativité provenant de l'effet chélate est dépendante de la concentration du ligand bivalent où une forte concentration de ligand peut empêcher la formation du complexe fermé pour favoriser la formation d'un complexe comprenant deux ligands pour un récepteur (Fig. 1.6e, équilibre inférieur)<sup>82</sup>. Ainsi, le paramètre  $\beta$  ne permettant pas de capturer cette dépendance de la concentration du ligand, un autre paramètre fut introduit soit  $\beta'$  où  $\beta' = EM/2[\text{Ligand}]$  et correspond à la constante d'équilibre intermoléculaire ( $K_{\text{inter}}$ ) entre le complexe fermé et le complexe doublement lié, le tout normalisé par la concentration de ligand<sup>82</sup>. Dans les deux cas, ces paramètres permettent de rationnellement quantifier la coopérativité provenant de l'effet chélate où une valeur supérieure à 1 indique une coopérativité positive ( $\beta > 1$  ou  $\beta' > 1$ ) et une valeur inférieure à 1 indique une coopérativité négative ( $\beta < 1$  ou  $\beta' < 1$ ). Puisque les deux paramètres sont définis différemment, il est possible que l'un soit positif et l'autre négatif (*p. ex.*  $\beta < 1$  et  $\beta' > 1$ ). Toutefois, aux meilleures des connaissances, une telle situation n'a pas encore été observée<sup>80</sup>. Ainsi, l'un ou l'autre des paramètres peut être utilisé afin de décrire l'effet chélate. Ici, nous utiliserons le paramètre  $\beta$  au lieu du paramètre  $\beta'$ , car ce dernier nécessite que la concentration de ligand soit en large excès par rapport au récepteur, ce qui ne sera pas toujours le cas lors de nos expériences de titrage.

Concrètement, l'impact d'un effet chélate neutre ( $\beta = 1$ ) ou négatif ( $\beta < 1$ ) se traduit par l'incapacité du système à former le complexe fermé (Fig. 1.6f, courbe rouge). C'est seulement en présence d'un effet chélate positif ( $\beta > 1$ ) que le complexe fermé devient favorable. Néanmoins, une augmentation importante de la concentration de ligands bivalents permet une compétition entre la formation du complexe fermé (1:1) et la formation du complexe doublement lié (2:1) qui se traduit par une diminution de la concentration de complexe fermé à haute concentration de ligands. Le seuil de concentration de ligands où le système doublement lié devient majoritaire est directement relié à la concentration molaire effectivement selon :  $[\text{Ligand}] = EM/2$  (Fig. 1.6f, ligne pointillée)<sup>79</sup>. En augmentant la concentration molaire effective, *c.-à-d.* que  $\beta$  devient plus grand, le complexe fermé restera stable sur une plus grande plage dynamique (Fig. 1.6f, courbe verte). Il devient donc important de mesurer cette concentration molaire effective afin de comprendre

si la formation du complexe fermé sera favorisée. Malheureusement, il est impossible de mesurer directement la concentration molaire effective<sup>88</sup>. Pour ce faire, il faut donc passer par un cycle thermodynamique qui permettra de disséquer uniquement la contribution de l'effet chélate, car on ne peut pas négliger le fait que le premier événement de liaison du ligand bivalent puisse altérer, par un effet allostérique, le second événement de liaison<sup>80</sup>.

#### **1.2.4.3 Cycle du double mutant : contribution de l'allostérie et de l'effet chélate**

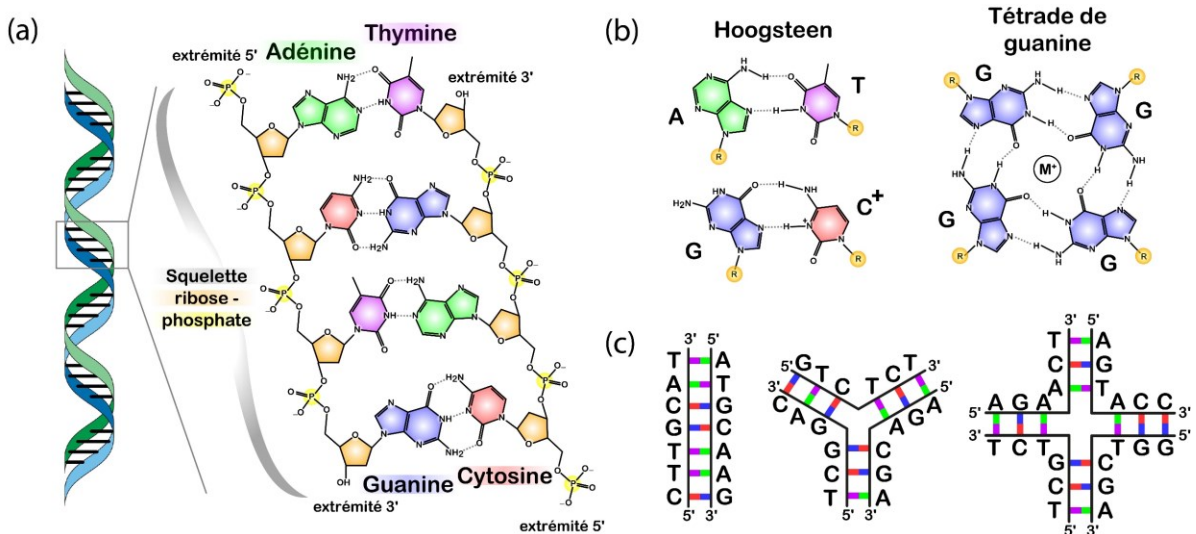
Le cycle du double mutant introduit par W. P. Jencks en 1981 est une méthode expérimentale permettant d'analyser les forces responsables de la liaison d'une molécule bivalente à une protéine<sup>95</sup> et fut ultérieurement adapté pour disséquer la contribution énergétique provenant de l'allostérie et de l'effet chélate (Fig. 1.7a)<sup>96</sup>. Pour ce faire, deux mutations sont introduites dans le système divalent à l'étude. La première mutation enlève l'espaceur qui relie les deux sites de liaison du ligand tandis que la seconde mutation enlève l'espaceur qui relie les deux sites de liaison du récepteur. Le cycle thermodynamique commence en analysant le système où les deux mutations sont appliquées, *c.-à-d.* l'interaction des ligands monovalents envers leurs récepteurs monovalents respectifs (état A, Fig. 1.7b). Ensuite, on introduit l'espaceur entre les deux ligands monovalents pour déterminer la contribution allostérique de la première liaison du récepteur sur la deuxième ( $\alpha_{\text{Lig}}$ ) grâce à la communication maintenant rendue possible par le biais de l'espaceur du ligand divalent (état B, Fig. 1.7b). Similairement, on introduit l'espaceur entre les deux récepteurs monovalents pour déterminer la contribution allostérique de la première liaison du ligand sur la deuxième ( $\alpha_{\text{Rec}}$ ) grâce à la communication maintenant rendue possible par le biais de l'espaceur du récepteur divalent (état C, Fig. 1.7b). Finalement, on introduit les deux espaceurs afin de mesurer l'énergie globale du système bivalent (état D, Fig. 1.7b). La magie derrière ce cycle thermodynamique provient du fait que chaque composante moléculaire du cycle se retrouve de chaque côté de l'équilibre global du cycle (Fig. 1.7b). Par conséquent, l'équilibre global du cycle permet directement de mesurer la concentration molaire effective (EM) et l'effet chélate ( $\beta$ ) tout en mesurant les contributions allostériques ( $\alpha_B$  et  $\alpha_C$ ) grâce aux états B et C. De plus, ces mesures sont possibles en mesurant simplement quatre constantes d'équilibre qui décrivent chacune les quatre états du cycle.



**Figure 1.7 – Schématisation du cycle du double mutant.** (a) *État A*. Les deux sites de liaison des ligands sont mesurés indépendamment et serviront de référence pour mesurer l’impact de la formation d’un ligand ou d’un récepteur bivalent. *État B*. Un espaceur est introduit entre les deux ligands pour former un ligand bivalent. *État C*. Un espaceur est introduit entre les deux récepteurs pour former un récepteur bivalent. *État D*. Les deux espaceurs sont introduits au système. (b) La contribution allostérique de l’espaceur du ligand ( $\alpha_{Lig}$ ) et la contribution allostérique de l’espaceur du récepteur ( $\alpha_{Rec}$ ) peuvent être déterminées respectivement en mesurant l’effet de l’espaceur sur la liaison successive des deux récepteurs libres ( $K_{B1}$  et  $K_{B2}$ ) et des deux ligands libres ( $K_{C1}$  et  $K_{C2}$ ), et en comparant le tout au système sans espaceurs ( $K_{A1}$  et  $K_{A2}$ ). La contribution de l’effet chélate qui est proportionnel à EM est déterminée en considérant l’équilibre global du cycle, soit la tendance du système à former le complexe bivalent fermé ( $K_D$ ). À noter que toutes les constantes d’équilibre sont définies en termes de constantes d’association.

### 1.3 L’ADN : un modèle moléculaire pour étudier la régulation d’assemblages moléculaires.

Afin de mieux comprendre la complexité des mécanismes permettant de réguler l’activité des protéines naturelles, il est impératif de trouver des systèmes moléculaires plus simples. L’ADN est un biopolymère de choix en raison de sa simplicité de synthèse, sa capacité à former des structures spécifiques bien définies ainsi que la facilité à prédire la stabilité de ses structures finales. Durant les dernières décennies, l’ADN a démontré sa pertinence pour recréer artificiellement des mécanismes de régulation observés chez les récepteurs protéiques tels l’allostérie<sup>90</sup>, la séquestration moléculaire<sup>97</sup> et la coopérativité négative<sup>98</sup>. L’ADN s’impose donc comme une alternative intéressante pour étudier expérimentalement les mécanismes de régulation coopératifs derrière les interactions des protéines et des biomolécules, mais aussi pour mettre en lumière les avantages évolutifs de ces assemblages moléculaires.



**Figure 1.8 – Différentes structures pouvant être adoptées par l'ADN.** (a) L'acide désoxyribonucléique (ADN) est un biopolymère pouvant adopter une structure de double hélice antiparallèle. Ce biopolymère est constitué de monomères comprenant un groupement phosphate, un désoxyribose et l'une des quatre bases azotées naturelles soit : l'adénosine, la thymine, la guanine ou la cytosine. Ces quatre différents monomères peuvent interagir suivant un patron de ponts hydrogènes nommé Watson-Crick. (b) L'ADN peut aussi adopter des structures non-canoniques comprenant les interactions Hoogsteen (gauche) et les tétrades de guanine (droite). (c) L'ADN peut aussi former des jonctions plus grandes qu'une simple double hélice (gauche) telles une jonction à trois branches (centre) ou une jonction à quatre branches (droite). Les structures d'ADN furent adaptées de la figure originale de M. P. Ball.**1.3.1 Les différentes structures de l'ADN**

L'ADN est un biopolymère où deux brins peuvent s'assembler pour former une double hélice antiparallèle, *c.-à-d.* que leurs extrémités 3'/5' sont opposées (voir Fig. 1.8a pour la structure chimique de ce biopolymère). Les bases azotées qui composent l'ADN sont catégorisées en deux familles selon la taille de leur hétérocycle : les pyrimidines (hétérocycle à six atomes comprenant la thymine et la cytosine) et les purines (bishétérocycle à 9 atomes comprenant l'adénosine et la guanine). De plus, chaque base azotée possède des groupements accepteurs-donneurs (A-D) de ponts hydrogène de sorte que le couple adénosine/thymine (DA/AD) et le couple cytosine/guanine (DAA/ADD) interagissent spécifiquement ensemble (Fig. 1.8b). Puisque les charges négatives du phosphate sont exposées au solvant, l'hybridation de l'ADN en double hélice est fortement favorisée par la présence de cations (Na<sup>+</sup>, K<sup>+</sup>, Mg<sup>2+</sup>, etc.). La formation de la double hélice se produit généralement entre deux brins d'ADN distincts, mais la double hélice

peut aussi se former de façon intramoléculaire entre deux sections du même biopolymère menant ainsi à la formation d'une tige-boucle par exemple.

Outre la capacité de l'ADN à former une double hélice, l'ADN peut aussi former diverses structures grâce à l'appariement non-canonique des bases azotées (Fig. 1.8c). Les interactions Hoogsteen, qui se basent sur les ponts hydrogènes plus faibles entre les azotes N7 des purines et N3 des pyrimidines, permettent la création d'une grande variété de structures<sup>99</sup>. Le triplex, par exemple, se forme lorsqu'un troisième brin d'ADN s'intercale dans le sillon majeur d'une double hélice grâce à des interactions Hoogsteen avec un brin riche en purines<sup>100</sup>. Les tétrades de guanine se forment aussi suivant ce patron d'interactions. Les cytosines, quant à elles, peuvent aussi former des motifs-*i* en milieu acide suivant la protonation de l'azote N3 (pKa = 4.5) rendant ainsi son assemblage sensible au pH. La protonation spécifique de la cytosine permet aussi de former des triplex sensible au pH grâce à l'interaction d'une cytosine protonée avec une guanine. Il est aussi possible de sélectionner artificiellement, grâce à la méthode d'évolution systématique de ligands par enrichissement exponentiel (SELEX)<sup>101</sup>, des brins d'ADN pouvant interagir avec diverses molécules comme de petites molécules organiques, des protéines et même des cellules<sup>24</sup>. Ces brins d'ADN se replient généralement sur eux-mêmes grâce aux interactions Watson-Crick, mais interagissent avec leur cible suivant diverses interactions comme des ponts hydrogènes, des interactions électrostatiques, des interactions hydrophobes, des interactions  $\pi$ - $\pi$  ou des interactions faibles de type Van der Waals<sup>102</sup>. Certains mésappariements de bases dans une double hélice peuvent aussi interagir spécifiquement avec des cations métalliques<sup>103</sup>. Les mésappariements T-T ou C-C, par exemple, permettent respectivement la liaison du mercure ( $\text{Hg}^{2+}$ ) et de l'argent ( $\text{Ag}^+$ ).

L'ADN peut aussi former des structures plus complexes que la double hélice impliquant l'assemblage de plusieurs brins d'ADN. En s'inspirant de la grande prédictibilité des interactions Watson-Crick, Seeman proposa en 1982 les règles de base de conceptualisation pour la construction de structure d'ADN ramifié<sup>104</sup> ouvrant ainsi une nouvelle ère pour les nanotechnologies à base d'ADN<sup>105</sup>. Ces ADN ramifiés artificiels s'inspirent en grande partie d'importantes découvertes en biologie. Il fut proposé, à l'époque, que la réparation des doubles brins d'ADN, par exemple, fasse intervenir la présence d'une jonction à quatre brins d'ADN

nommé jonction de Holliday en l'honneur du scientifique ayant proposé ce mécanisme<sup>106</sup>. Pour simplifier la description des ADN ramifiés, prenons pour exemple la structure ramifiée la plus simple pouvant être construite, soit la jonction à trois branches (Fig. 1.8d). L'assemblage d'une telle jonction peut être réalisé de deux manières : séquentiellement ou tout-en-un<sup>105</sup>. La méthode séquentielle implique premièrement une incubation de deux brins ensemble pour favoriser leurs interactions, puis le troisième brin est ajouté pour fermer la jonction. Dans la méthode tout-en-un, tous les brins sont mélangés, puis un recuit est réalisé afin de favoriser l'assemblage de la jonction. Même si cette jonction à trois branches est structurellement simple, on peut facilement envisager des problèmes d'assemblage en augmentant le nombre de brins (*p. ex.* jonction à quatre, cinq ou six branches)<sup>105</sup>. Ainsi, la construction de structures plus imposantes fait souvent intervenir l'assemblage de plusieurs petites jonctions à l'aide d'extrémités collantes complémentaires à l'extrémité de plusieurs autres jonctions. Ainsi, il est possible de construire un tétraèdre formé de jonctions à trois branches<sup>107</sup> ou bien de construire un cube avec des jonctions à quatre branches<sup>108</sup>. De plus, il est possible de prendre un long brin d'ADN viral (~ 7000 nucléotides) et d'y ajouter des petites attaches d'ADN afin de replier, comme un origami, ce long brin en une structure d'ADN bien précise<sup>109</sup>. Afin de faciliter la construction de ces structures d'ADN de plus en plus imposantes, des logiciels furent développés afin de prédire l'appariement des différents brins d'ADN et d'identifier la séquence de ces derniers<sup>110-112</sup>.

### 1.3.2 Prédire la stabilité des nanostructures d'ADN

La double hélice d'ADN doit sa forte stabilité en milieux aqueux principalement aux ponts hydrogènes entre les bases azotées, mais aussi à la stabilisation hydrophobe provenant de l'empilement des cycles aromatiques des bases à l'intérieur de l'hélice<sup>21</sup>. Ces deux forces sont directement reliées aux bases azotées faisant en sorte que la stabilité de la double hélice augmente proportionnellement avec le nombre de paires de bases impliquées. De plus, ces forces d'interactions vont changer selon la composition du brin d'ADN. Plusieurs mesures expérimentales exhaustives sur différentes combinaisons de brins d'ADN ont permis de développer un modèle unifié du plus proche voisin qui stipule que la stabilité d'une paire de bases dépend de l'identité et de l'orientation des paires de bases voisines<sup>113</sup>. Ce modèle, par exemple, permet d'estimer l'énergie d'interaction entre deux brins<sup>114</sup>. Ainsi, une substitution d'une paire

de bases AT par une paire de bases GC permet de stabiliser la double hélice par environ 0,6 kcal·mol<sup>-1</sup> (à 37°C et 1 M de NaCl). Bien sûr, cette substitution aura un effet moins important si elle se retrouve à l'extrémité de la double hélice. De plus, cela a permis ces dernières années de mettre au point des programmes informatiques comme NUPACK<sup>110</sup>, Mfold<sup>111</sup> et IDT SciTools<sup>112</sup> pouvant prédire la structure secondaire d'un ou plusieurs brins d'ADN tout en estimant la stabilité de cette structure. En connaissant la stabilité des nanostructures d'ADN, il est possible de déterminer la quantité d'ADN formant la structure versus la quantité d'ADN libre suivant la relation entre la stabilité de la nanostructure d'ADN, *c.-à-d.* son énergie libre de Gibbs ( $\Delta G$ ), et la constante d'équilibre du système (Eq. 1.1 et Eq. 1.2). Dans ces équations,  $K_D$  représente la constante d'équilibre de dissociation en mol·L<sup>-1</sup>,  $\Delta G$  représente l'énergie libre de Gibbs en kcal·mol<sup>-1</sup>, T représente la température en Kelvin, R représente la constante universelle des gaz parfaits (1,987 cal·mol<sup>-1</sup>·K<sup>-1</sup>),  $[AB]_{eq}$  représente la concentration de duplex à l'équilibre alors que  $[A]_T$  et  $[B]_T$  représentent les concentrations initiales des brins.

$$\Delta G = RT \ln(K_D) \quad (\text{Eq. 1.1})$$

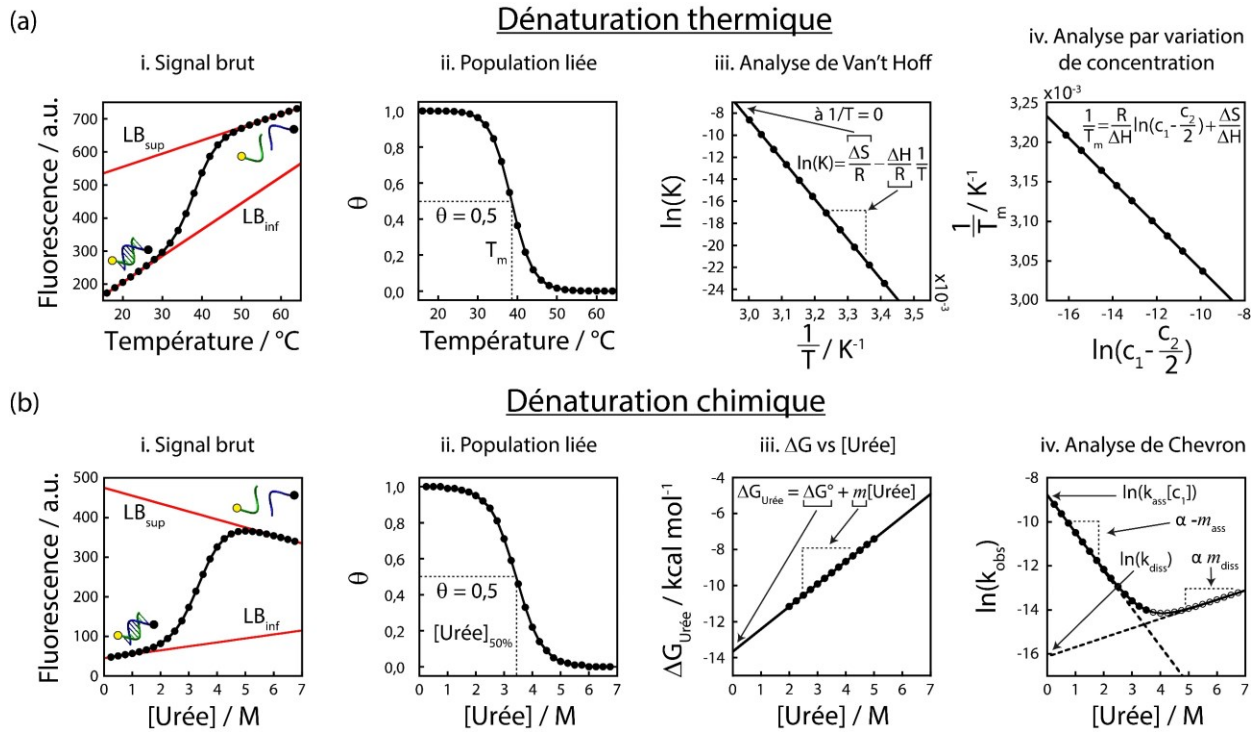
$$[AB]_{eq} = \frac{[A]_T + [B]_T + K_D - \sqrt{([A]_T + [B]_T + K_D)^2 - 4[A]_T[B]_T}}{2} \quad (\text{Eq. 1.2})$$

Malgré tout, il demeure difficile de prédire la stabilité des nanostructures d'ADN impliquant des interactions non-canoniques. En effet, les modèles utilisés par NUPACK, Mfold et IDT SciTools permettent uniquement d'estimer l'énergie libre basée sur des interactions Watson-Crick des doubles hélices<sup>24</sup>. De plus, au niveau des assemblages moléculaires, seulement NUPACK permet de déterminer la stabilité d'interaction de plusieurs brins d'ADN (> 2 brins). Ainsi, ces programmes n'arrivent pas à estimer les contributions énergétiques d'interactions non-canoniques comme l'impact d'un cœur hydrophobe dans une jonction, l'effet de torsion dans une structure rigide comme un tétraèdre ou bien l'énergie d'une tétrade de guanines par exemple. Ainsi, la stabilité des nanostructures d'ADN non canoniques doit être mesurée expérimentalement soit par dénaturation thermique<sup>115</sup> ou par dénaturation par un agent chaotropique tel l'urée<sup>116</sup>.

### 1.3.3 Mesurer la stabilité des nanostructures d'ADN

Sous l'effet d'une augmentation de température, une nanostructure d'ADN se dissocie en ses divers constituants. Principalement, cette dissociation est mesurée en suivant une variation de propriétés optiques de la nanostructure d'ADN (*p. ex.* absorbance, fluorescence ou ellipticité). Avant tout, l'analyse se base en émettant l'hypothèse que la dénaturation se produit en deux états, *c.-à-d.* d'un état assemblé vers un état désassemblé. Pour démontrer cette analyse, prenons l'exemple de la dénaturation d'un simple duplex d'ADN (Fig. 1.9a). Initialement, 100% de l'ADN se retrouve sous forme de duplex et au fur et à mesure que l'on augmente la température, l'ADN se dissociera en deux simples brins. Cette transition sigmoïdale est caractérisée par la température de fusion ( $T_m$ ), soit la température où 50% de l'ADN se retrouve en simples brins. Par contre, la majorité du temps, les propriétés optiques de l'ADN changeront, souvent linéairement, en fonction de la température<sup>117</sup>. Ainsi, il est important de tenir compte de cette dépendance et corriger la ligne de base supérieure ( $LB_{sup}$ ) et inférieure ( $LB_{inf}$ ) (Fig. 1.9a, *i*). À la suite de cette correction, nous obtenons l'évolution de la population du duplex ( $\theta$ ) en fonction de la température où la  $T_m$  correspond à la température lorsque  $\theta = 0,5$  (Fig. 1.9a, *ii*). Il existe deux méthodes pour calculer la stabilité du duplex à partir de cette courbe<sup>118</sup>. La première est la méthode de van't Hoff qui consiste à calculer la constante d'équilibre du système à toutes les températures sachant la population en duplex ( $\theta$ ) et la population de brins libres ( $1 - \theta$ ). Un graphique du logarithme naturel de la constante d'équilibre en fonction de l'inverse de la température (en Kelvin) permet ainsi de calculer le  $\Delta H$  et le  $\Delta S$  de l'hybridation du duplex d'ADN et donc de calculer l'énergie libre de Gibbs suivant la relation  $\Delta G = \Delta H - T\Delta S$  (Fig. 1.9a, *iii*). L'autre méthode consiste à mesurer la  $T_m$  à différentes concentrations de brins d'ADN et d'utiliser la relation linéaire entre le logarithme naturel de la concentration des brins d'ADN et entre l'inverse de la  $T_m$  (en Kelvin) pour calculer le  $\Delta H$  et le  $\Delta S$  de la réaction d'hybridation (Fig. 1.9a, *iv*). À noter que si la dénaturation d'une nanostructure d'ADN se produit aussi en deux états, alors la même analyse s'appliquera. L'avantage principal de cette technique est qu'elle permet de mesurer les variations d'enthalpie et d'entropie associées à la transition étudiée, amenant des informations intéressantes supplémentaires sur les forces thermodynamiques derrière l'assemblage de la nanostructure d'ADN.





**Figure 1.9 – Méthodes expérimentales pour mesurer l'énergie de nanostructures d'ADN.** (a) Dénaturation thermique. *i.* Représentation typique du signal brut d'une dénaturation thermique d'un duplex d'ADN obtenu par spectroscopie de fluorescence. *ii.* Normalisation du signal brut. *iii.* Détermination des paramètres thermodynamiques par l'analyse de Van't Hoff. *iv.* Détermination des paramètres thermodynamiques par variation de concentrations. (b) Dénaturation chimique. *i.* Représentation typique du signal brut d'une dénaturation d'un duplex d'ADN en présence d'urée obtenu par spectroscopie de fluorescence. *ii.* Normalisation du signal brut. *iii.* Détermination des paramètres thermodynamiques par régression linéaire. *iv.* Détermination des paramètres thermodynamiques par cinétique en utilisant la courbe de chevron. Les différentes courbes furent calculées provenant d'un duplex d'ADN ayant un  $\Delta H = -79 \text{ kcal}\cdot\text{mol}^{-1}$ ,  $\Delta S = -199 \text{ cal}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ ,  $c_1 = 100 \text{ nM}$ ,  $c_2 = 1 \text{ nM}$ ,  $m = 1,25 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ ,  $k_{\text{ass}} = 1,54 \times 10^3 \text{ M}^{-1}\cdot\text{s}^{-1}$ ,  $k_{\text{diss}} = 1,00 \times 10^{-7} \text{ s}^{-1}$ ,  $m_{\text{ass}} = 1,00 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$  et  $m_{\text{diss}} = 0,25 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ .

Similairement, il est possible de mesurer l'énergie d'une nanostructure d'ADN en la dissociant sous l'effet d'agents chaotropiques comme l'urée, le chlorure de guanidium ou le formamide<sup>119, 120</sup>. Ces agents chaotropiques dénaturent l'ADN principalement en perturbant les ponts hydrogènes et la sphère d'hydratation de la nanostructure. Tout comme la dénaturation thermique, la dénaturation d'un duplex par un agent chaotropique implique aussi l'hypothèse que la dénaturation survient en deux états (Fig. 1.9b). De plus, les dénaturants affectent aussi les propriétés optiques de l'ADN nécessitant donc une correction de ligne de base. À l'instar de la dénaturation thermique où la température déstabilise l'énergie libre en défavorisant le terme entropique ( $\Delta G = \Delta H - T\Delta S$ ), les dénaturants chaotropiques affectent directement l'énergie libre

selon  $\Delta G_{\text{urée}} = \Delta G^\circ - m[\text{urée}]$  où  $\Delta G_{\text{urée}}$  et  $\Delta G^\circ$  sont les énergies libres avec et sans urée en  $\text{kcal}\cdot\text{mol}^{-1}$ ,  $[\text{urée}]$  est la concentration d'urée en  $\text{mol}\cdot\text{L}^{-1}$  et  $m$  est le « *m-value* » en  $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ , un paramètre qui caractérise l'influence de l'urée sur l'énergie libre et qui est proportionnel avec la quantité de surface exposée au solvant lors de la dénaturation<sup>121</sup>. La courbe de dénaturation peut donc directement être analysée en utilisant la bonne équation (Fig. 1.9b, *i*) ou bien l'énergie libre peut être calculée à chaque concentration d'urée sachant la population de duplex ( $\theta$ ) et la population de brin libre ( $1 - \theta$ ) (Fig. 1.9b, *ii*) et tracé en fonction de la concentration d'urée (Fig. 1.9b, *iii*). L'extrapolation de cette régression linéaire donne accès à l'énergie libre sans urée et la pente correspond à la valeur du paramètre  $m$ . L'avantage de cette technique, par rapport à la dénaturation thermique, est qu'elle offre de l'information, grâce à la valeur de  $m$ , sur l'état de compaction de la nanostructure d'ADN et de son accessibilité au solvant.

De plus, une dénaturation similaire par un agent chaotropique peut être réalisée cinétiquement en mesurant la vitesse d'association et de dissociation de la nanostructure à différentes concentrations d'urée. Cette méthode est communément nommée une courbe de chevron, de l'anglais « chevron plot », qui fait référence à la forme en « v » de la courbe. Avec cette méthode de caractérisation, l'énergie libre est calculée en mesurant la constante d'équilibre obtenue par le ratio entre la vitesse de dissociation et d'association en absence d'urée ( $K_{\text{eq}} = k^\circ_{\text{diss}} / k^\circ_{\text{ass}}$ ). Ces vitesses sont extrapolées à partir de la courbe de chevron par la régression linéaire du domaine associatif et dissociatif de la courbe (Fig. 1.9b, *iv*). Tout comme avec la dénaturation à l'équilibre, les vitesses d'association et de dissociation dépendent linéairement de la concentration d'urée et cette dépendance est caractérisée par les valeurs  $m_{\text{ass}}$  et  $m_{\text{diss}}$ . Dans ces deux dénaturations, à l'équilibre et par cinétique, les valeurs de  $\Delta G^\circ$  doivent être les mêmes et la somme des deux valeurs de  $m$  cinétiques ( $m_{\text{ass}}$  et  $m_{\text{diss}}$ ) doit être égale à la valeur de  $m$  mesuré à l'équilibre. Si ces valeurs ne sont pas égales, alors cela confirme que l'hypothèse que la transition se fait en deux états n'est pas bonne et que probablement un état intermédiaire est présent lors de la dénaturation<sup>122</sup>. De plus, le bêta de Tanford ( $\beta_T$ ), un paramètre qui correspond au ratio entre  $m_{\text{ass}}$  et le  $m$  à l'équilibre ( $\beta_T = m_{\text{ass}} / m$ ), permet d'obtenir de l'information sur l'accessibilité au solvant de l'état de transition par rapport à l'état natif non assemblé. En d'autres mots, une faible valeur de  $\beta_T$  stipule que l'atteinte de l'état de transition implique une faible portion de la surface

de l'ADN tandis qu'une grande valeur de  $\beta_T$  stipule que l'état de transition se rapproche structurellement de l'état assemblé de la nanostructure d'ADN. En somme, la dénaturation à l'urée permet de calculer l'énergie libre d'une nanostructure d'ADN tout en offrant d'importantes informations structurelles sur l'accessibilité au solvant de l'état assemblé, mais aussi sur l'état de transition menant à l'autoassemblage des brins qui la compose.

#### **1.3.4 Création d'autoassemblages moléculaires dynamiques à base d'ADN.**

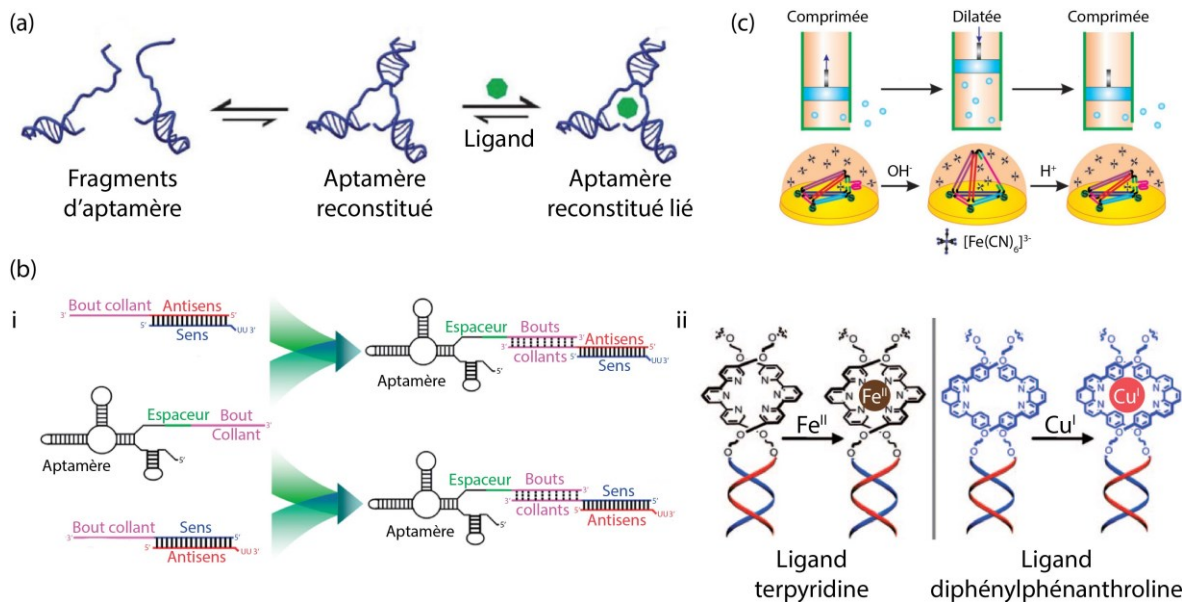
Tout comme les protéines, il est possible de construire des autoassemblages multimériques dynamiques à base d'ADN. Comme mentionné précédemment (section 1.1.3), les protéines ont évolué pour former des autoassemblages via, par exemple, des événements de fission, de fusion ou d'insertion. Nous verrons comment ces trois mêmes stratégies peuvent, similairement, être utilisées afin de construire des autoassemblages d'ADN (Fig. 1.10).

Une méthode simple pour créer des autoassemblages d'ADN dynamiques est de simplement couper une structure d'ADN statique ou dynamique en plusieurs fragments où un stimulus permettra d'induire l'assemblage des fragments. Par exemple, en coupant en deux un aptamère (Fig. 1.10a), *c.-à-d.* un oligonucléotide synthétique pouvant lier un ligand ou catalyser une réaction chimique avec ce dernier<sup>123</sup>, il est possible de créer des autoassemblages, car l'interaction du ligand avec les fragments fournira l'énergie nécessaire pour reconstituer l'aptamère. Puisque les aptamères sont sélectionnés pour lier préférentiellement leur ligand, ils sont souvent plus stables dans leur état de liaison ce qui limite grandement l'amplitude du déplacement de population (voir section 1.2.3). En coupant la structure en deux, le système fragmenté se retrouvera majoritairement dans son état inactif, ce qui assure maintenant que la présence du ligand induira un plus grand déplacement de population lors que la reconstitution de l'aptamère. En se reposant sur ce principe, plusieurs aptamères furent coupés en plus petits fragments afin de créer des systèmes de détection en sandwich; une approche qui fut directement inspirée de la version en sandwich de la méthode immuno-enzymatique ELISA<sup>124</sup>. La première étape dans la création de tels systèmes est d'identifier un site de clivage dans la séquence de l'aptamère. Comme mentionné précédemment, la structure tertiaire des aptamères n'est pas toujours connue, ce qui limite grandement l'identification rationnelle d'un site de coupe.

Même si la prédiction de la structure secondaire peut remédier à ce problème, le manque d'information sur l'identité des nucléotides impliqués dans la liaison du ligand empêche aussi l'identification rationnelle d'un site de coupe. En effet, une coupe au niveau des nucléotides responsables de la liaison pourrait grandement affecter la sélectivité et la spécificité de l'interaction avec le ligand<sup>125</sup>. Pour ce faire, les sites de coupes sont souvent identifiés en testant une grande variété d'aptamères ayant été coupée à différentes positions. Toutefois, pour les aptamères où la structure tertiaire avec le ligand est bien connue, l'identification d'un site de coupe est beaucoup plus simple. Par exemple, l'aptamère d'ATP forme une tige-boucle où la liaison avec le ligand se produit avec des nucléotides formant deux tétrades de guanine au centre de la tige<sup>126</sup>. Ainsi, il est possible de couper l'extrémité de la tige-boucle de l'aptamère sans affecter la formation des tétrades lors de la dimérisation.<sup>127, 128</sup> Similairement, les aptamères formant des jonctions unimoléculaires à trois branches comme l'aptamère de cocaïne peuvent facilement être coupés au niveau de l'une de leur boucle<sup>128</sup>. Malgré cela, parmi la centaine d'aptamères qui ont été découverts à ce jour, seulement quelques-uns (< 10) furent divisés avec succès en plusieurs fragments<sup>124, 125</sup>.

Une alternative à la fragmentation pour créer des assemblages moléculaires dynamiques consiste au phénomène inverse où deux domaines d'interactions sont fusionnés par l'entremise d'un espaceur. La combinaison la plus simple imaginable avec de l'ADN est de fusionner deux brins d'ADN formant des duplex ensemble pour ainsi former des jonctions d'ADN de différentes tailles. La formation de ces assemblages moléculaires devient donc dépendante des brins d'ADN qui la compose rendant ainsi l'assemblage dynamique, car l'ajout de l'un des brins de l'assemblage aux autres brins permet d'induire la formation de la jonction. Bien qu'il soit facile de produire des jonctions avec des brins d'ADN, il est aussi possible de fusionner des aptamères ou des éléments de reconnaissance avec des brins d'ADN ou d'ARN<sup>129</sup>. Dans cet exemple qui exploite la chimie de l'ARN, un aptamère d'ARN fut attaché, grâce à un espaceur, à un bout collant d'ARN permettant l'attachement non-covalente à un petit brin d'ARN interférent fusionné à la séquence complémentaire du bout collant (Fig. 1.10b, *i*)<sup>130</sup>. Cela permet ainsi la formation d'un dimère pouvant exploiter la capacité de reconnaissance moléculaire de l'aptamère d'ARN et la capacité d'inhibition de l'ARN interférent. Il est aussi possible de créer des conjugués ADN-

protéine afin de former des nanoassemblages, une stratégie souvent utilisée pour la délivrance de médicament et la création de biocapteurs<sup>131</sup>. Finalement, la capacité de complexation de métaux avec des ligands organiques peut aussi être exploitée afin de créer des assemblages d'ADN métal-dépendant<sup>132, 133</sup>. Par exemple, en attachant de manière covalente des ligands terpyridines ou des ligands diphenylphénanthrolines à des brins d'ADN, il est possible de former des assemblages d'ADN dépendants respectivement de la présence de Fe(II) ou de Cu(I) (Fig. 1.10b, ii).



**Figure 1.10 – Stratégies pour construire des autoassemblages d'ADN dynamiques.** (a) Diviser une nanostructure en plusieurs fragments (*p. ex.* un aptamère). Figure adaptée et traduite de la référence 24, © 2018 The Royal Society of Chemistry. (b) Combiner différents éléments de liaison. *i* Joindre des segments d'ARN/ADN par l'utilisation de bout collant, *c.-à-d.* une extrémité d'ARN/ADN pouvant lier spécifiquement un autre brin d'ARN/ADN. Exemple avec un système entièrement construit d'ARN. Figure adaptée et traduite de la référence 130, © 2009 Oxford University Press. *ii* Utilisation de ligands organiques pour induire l'assemblage d'ADN en présence de métaux. Figure adaptée et traduite de la référence 133, © 2009 WILEY-VCH Verlag GmbH & Co. KGaA. (c) Insertion d'un motif-*i* dynamique dans une arête d'un tétraèdre d'ADN. Figure adaptée et traduite de la référence 135, © 2019 WILEY-VCH Verlag GmbH & Co. KGaA.

À l'instar de la méthode précédente où deux séquences sont combinées, il est aussi possible d'introduire du dynamisme dans un autoassemblage d'ADN statique (*c.-à-d.* qui ne change pas de conformation) en incorporant un segment d'ADN dynamique (*c.-à-d.* qui peut adopter différentes conformations) à l'intérieur même de l'auto-assemblage<sup>134</sup>. Les structures non canoniques d'ADN et les aptamères sont des candidats intéressants pour accomplir une telle

stratégie, car leurs interactions diffèrent typiquement considérablement des interactions Watson-Crick que l'on retrouve principalement dans les autoassemblages statiques d'ADN. Une méthode simple consiste à exploiter la différence de taille entre la forme simple brin et la forme repliée de ces différentes nanostructures dynamiques d'ADN. Par exemple, une séquence d'un motif-*i* fut incorporée dans les arêtes d'un tétraèdre<sup>135</sup> et d'un prisme à base triangulaire<sup>136</sup> d'ADN. Puisque la longueur du brin d'ADN étendu diffère de celle du motif replié (5,5 nm vs 1,6 nm)<sup>135</sup>, la présence de protons permet la contraction de ces deux nanostructures où le premier agit comme une nanopompe (Fig. 1.10c). Similairement, des chercheurs ont introduit des séquences de tétrades de guanine sensibles au potassium sur une seule des faces d'une nanotige d'ADN. Ainsi, sans potassium la tige est bien droite, mais la présence de potassium induit la formation de tétrades de guanines qui sont beaucoup plus petites causant ainsi une tension sur l'une des faces de la tige ce qui la force à plier en forme de « U ». À l'inverse de la stratégie précédente, il est aussi possible d'ajouter une séquence d'ADN déjà repliée sur elle-même dans une nanostructure qui pourra s'allonger en présence de son ligand, ce qui dilatera la nanostructure. C'est le cas, par exemple, avec un autre tétraèdre d'ADN où une tige-boucle fut incorporée dans l'une des arêtes<sup>137</sup>. En présence d'un micro-ARN complémentaire, la tige-boucle s'ouvre pour former un duplex rigide, ce qui allonge l'un des côtés du tétraèdre. Il est aussi possible, de la même manière, d'insérer des séquences d'aptamères<sup>138</sup> ou des mésappariements<sup>139</sup> dans des nanostructures d'ADN afin de rendre la dynamique d'assemblage ligand-dépendante.

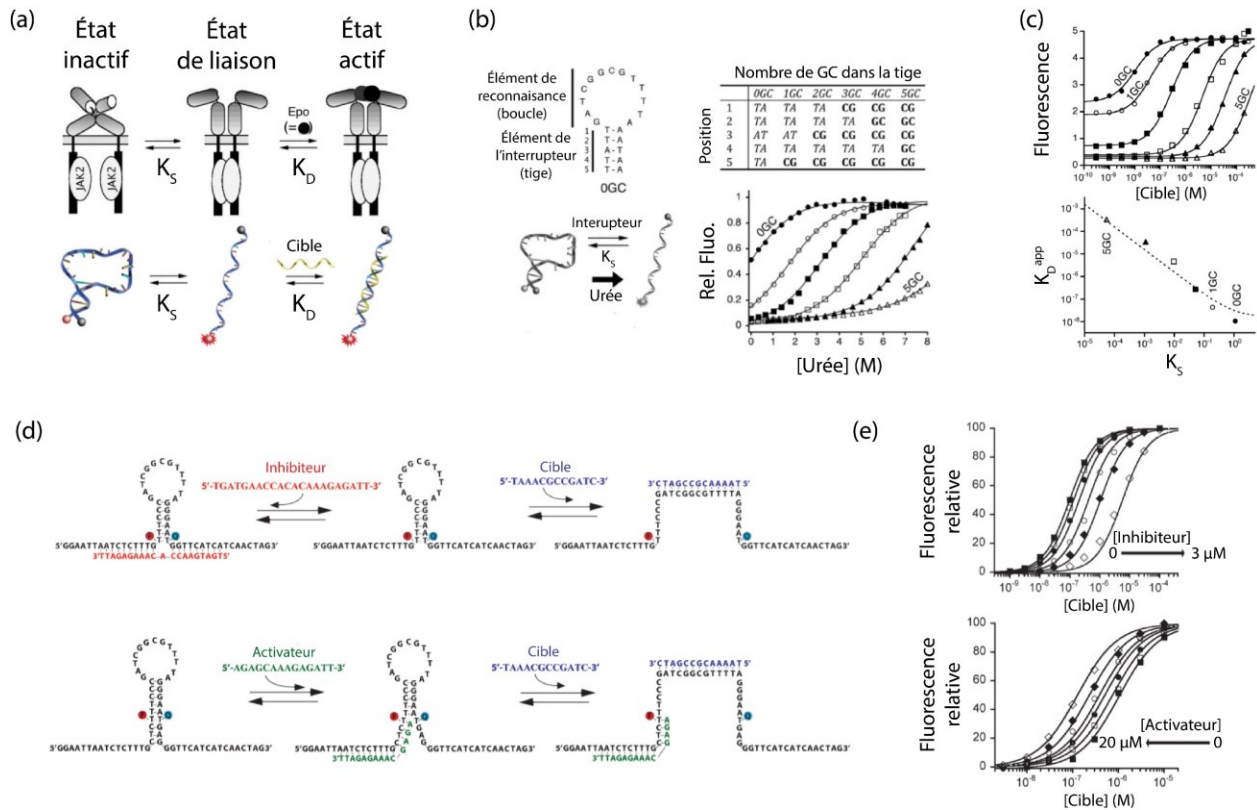
## **1.4 La biochimie synthétique : comprendre les mécanismes naturels de régulations à partir de modèles moléculaires simplifiés.**

Outre l'utilisation de l'ADN pour créer de nouvelles nanotechnologies<sup>37</sup>, la simplicité de conceptualisation des autoassemblages d'ADN offre une opportunité pour recréer et analyser des mécanismes naturels de régulation. L'étude de ces mécanismes grâce à des protéines naturelles n'est pas si facile que l'on peut le croire. En effet, il est difficile de faire varier uniquement un paramètre thermodynamique, cinétique ou structural d'une protéine avec des mutations sans affecter indirectement plusieurs phénomènes biochimiques souvent mutuellement interconnectés (*p. ex.* la stabilité, la coopérativité, la force d'interaction avec un ligand, la plage

dynamique, la vitesse d'assemblage ou l'efficacité du déplacement de population)<sup>140</sup>. Toutefois, l'ADN permet la création de modèles moléculaires facilitant l'étude rationnelle de l'influence de ces paramètres thermodynamiques et cinétiques sur les mécanismes biochimiques naturels. Ici, nous verrons donc comment des modèles simples d'ADN ont permis de mieux comprendre le mécanisme de déplacement de population, d'allostérie, de coopérativité, de séquestration moléculaire, d'ultrasensibilité ainsi que les mécanismes cinétiques de l'ajustement induit et de la sélection conformationnelle.

#### 1.4.1 Le mécanisme de déplacement de population

Des chercheurs ont démontré les bases thermodynamiques derrière le mécanisme de déplacement de population en exploitant la simplicité structurale de l'ADN (Fig. 1.11a)<sup>71</sup>. En effet, ils ont créé différents interrupteurs moléculaires à base de tige-boucles d'ADN ayant tous une longueur identique, mais des stabilités différentes grâce à des mutations successives de paires de bases AT par des paires de bases GC dans le duplex de la tige (Fig. 1.11b). De plus, tous ces interrupteurs s'ouvrent en présence du même ligand : un brin d'ADN qui interagit avec la boucle. Ils ont donc créé six tiges-boucles contenant entre 0GC et 5GC et qui possèdent des constantes d'équilibre d'interconversion variant entre  $K_S = 1,1 \pm 0,2$  (0GC) et  $K_S = (5,5 \pm 0,7) \times 10^{-5}$  (5GC). Avec ces différentes tiges-boucles, ils ont validé l'importance de la stabilité de l'interrupteur ( $K_S$ ) sur l'interaction entre l'interrupteur et son ligand. Par exemple, un interrupteur très stable, comme la tige-boucle contenant 5GC, impose une pénalité énergétique importante sur la liaison du ligand se traduisant ainsi par une augmentation de la constante d'affinité apparente du ligand ( $K_D^{app} > K_D$ ) (Fig. 1.11c). Inversement, un interrupteur peu stable, comme la tige-boucle contenant 0GC, n'affecte pas significativement l'affinité du ligand ( $K_D^{app} \approx K_D$ ), mais diminue de façon importante l'amplitude du déplacement de population, car 50% de l'interrupteur existe déjà dans une conformation ouverte active. Ainsi, cette étude avec de simples tiges-boucles d'ADN a démontré quantitativement pourquoi le  $K_S$  d'un interrupteur moléculaire biologique doit être près de 0,1 afin de minimiser la pénalité énergétique sur la liaison d'un ligand tout en maximisant l'amplitude du déplacement de population.



**Figure 1.11 – Biochimie synthétique : les mécanismes de déplacement de population et de l'allostérie.** (a) Avec une simple tige-boucle d'ADN, il est possible de mieux comprendre le mécanisme de déplacement de population. (b) En changeant la composition de la tige, il est possible de changer la stabilité de l'interrupteur ( $K_S$ ). (c) Ce système a permis de valider expérimentalement les équations démontrant la relation entre la stabilité de l'interrupteur ( $K_S$ ) sur l'interaction entre l'interrupteur et son ligand ( $K_D^{app}$ ). Les panneaux a-c furent adaptés et traduits de la référence 71, © 2009 National Academy of Sciences. (d) En ajoutant un second brin d'ADN qui stabilise l'état inactif (inhibiteur) ou qui stabilise l'état de liaison (activateur), il fut possible de mieux comprendre le mécanisme d'allostérie dans une perspective du déplacement de population. (e) Un inhibiteur allostérique diminue l'affinité du ligand pour la boucle ( $K_D^{app} > K_D$ ) tandis qu'un activateur allostérique augmente l'affinité du ligand pour la boucle ( $K_D^{app} < K_D$ ). Les panneaux d-e furent adaptés et traduits de la référence 90 © 2012 American Chemical Society.

### 1.4.2 Le mécanisme d'allostérie et de coopérativité

Suivant cette étude sur le mécanisme de déplacement de population, le même groupe de recherche a voulu déterminer l'impact d'un effecteur allostérique sur la liaison d'un ligand et de son interrupteur moléculaire<sup>90</sup>. Pour ce faire, ils ont utilisé le même système d'ADN constitué d'une tige-boucle et d'un ligand d'ADN pouvant lier la boucle. Toutefois, ils ont conceptualisé deux nouveaux brins d'ADN pouvant soit stabiliser l'état inactif de l'interrupteur (inhibiteur



allostérique) ou stabiliser l'état de liaison de l'interrupteur (activateur allostérique) (Fig. 1.11d). En d'autres mots, ces deux effecteurs allostériques affectent directement l'équilibre d'interconversion de l'interrupteur ( $K_S$ ) se traduisant ainsi par une modification de l'interaction entre l'interrupteur et son ligand. Grâce à ce système d'ADN, ces chercheurs ont validé l'ensemble des équations mathématiques qui décrivent l'effet d'un effecteur allostérique. Par exemple, ils ont démontré qu'un effecteur allostérique qui stabilise l'état inactif d'un interrupteur ( $\downarrow K_S$ ) entraîne un coût énergétique plus important pour le ligand se traduisant par une diminution de l'affinité du ligand ( $K_D^{app} > K_D$ ) (Fig. 1.11e, *haut*). À l'inverse, un effecteur allostérique qui stabilise l'état de liaison de l'interrupteur ( $\uparrow K_S$ ) entraîne une réduction du coût énergétique pour le ligand se traduisant par une augmentation de l'affinité du ligand ( $K_D^{app} < K_D$ ) (Fig. 1.11e, *bas*).

La coopérativité résultant d'un mécanisme allostérique fut aussi étudiée avec une tige-boucle d'ADN<sup>141</sup>. Toutefois, à la différence du système précédent, les séquences du ligand et de l'effecteur allostérique sont identiques et lient seulement la moitié de la boucle. Ainsi, l'ouverture de la tige-boucle sera facilitée par la liaison d'un premier ligand qui agira comme un activateur allostérique et qui stabilisera l'état liant de l'interrupteur augmentant ainsi l'affinité pour un second ligand identique. Ces travaux ont démontré, par exemple, que l'ouverture de la tige-boucle devient plus coopérative lorsqu'on augmente l'affinité du ligand, car une plus grande population de l'interrupteur se déplacera vers l'état liant à la suite de la liaison d'un premier ligand. En d'autres mots, plus la différence d'énergie entre le premier et le deuxième événement de liaison est grande, plus le système sera coopératif. Pour cette raison, il est donc possible d'augmenter la coopérativité en stabilisant l'interrupteur, *c.-à-d.* en diminuant le  $K_S$ . En effet, l'ouverture d'une tige-boucle comprenant 6GC ( $K_S = 0,0025$ ) est plus coopérative ( $n_H = 1,94 \pm 0,17$ ) que celle comprenant 5GC ( $K_S = 0,14$ ,  $n_H = 1,43 \pm 0,13$ ), car le coût énergétique de la première liaison devient plus important.

Ces deux travaux sur l'allostérie et la coopérativité démontrent comment une nanostructure simple d'ADN comme une tige-boucle permet de décrire et valider les fondements thermodynamiques derrière un mécanisme omniprésent chez les systèmes biologiques naturels. De plus, ces nouvelles compréhensions permettent aussi de mieux conceptualiser les nanotechnologies de demain par la création d'interrupteurs moléculaires optimisés. Par exemple,

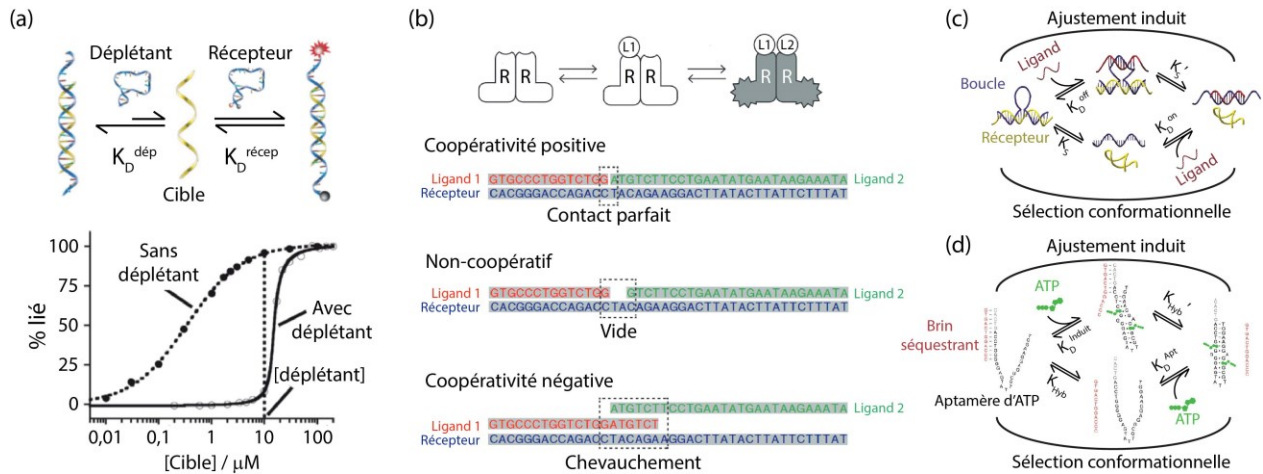
grâce à l’allostérie et la coopérativité, il est possible de créer rationnellement des interrupteurs moléculaires pour la délivrance spécifique de médicament<sup>142, 143</sup>, la détection d’un analyte sur une plage dynamique bien précise<sup>144-146</sup> ou pour l’imagerie moléculaire<sup>147</sup>.

### 1.4.3 Le mécanisme de séquestration moléculaire et d’ultrasensibilité

Les réseaux de régulation de gènes exploitent la présence de répresseurs afin de séquestrer un gène et ainsi obtenir une réponse ultrasensible. Par exemple, des simulations numériques ont permis de mettre en lumière les équations et les principes d’équilibres de base derrière le comportement d’un réseau de gènes sous l’influence de cette séquestration moléculaire<sup>148</sup>. Toutefois, ces équations n’ont pas été entièrement validées avec des systèmes moléculaires, mais ce comportement fut toutefois observé par le même groupe de recherche avec un circuit génétique synthétique<sup>149</sup>. Pour valider ces équations, un système d’ADN fut mis au point<sup>97</sup>. Ce système est constitué d’un ligand d’ADN pouvant interagir avec une tige-boucle servant de récepteur, mais qui peut aussi se faire séquestrer par une autre tige-boucle d’ADN servant de déplétant empêchant ainsi la liaison entre le ligand et le récepteur (Fig. 1.12a). En modifiant le contenu en GC de la tige, divers récepteurs et déplétants ont été créés pour avoir différentes affinités envers le ligand, ce qui permet de valider indépendamment l’influence du  $K_D$  du déplétant ( $K_D^{\text{dép}}$ ) et du  $K_D$  du récepteur ( $K_D^{\text{récep}}$ ). Premièrement, ce système a permis de vérifier que l’ultrasensibilité est obtenue uniquement si la concentration de déplétant est supérieure à l’affinité du ligand pour le récepteur ( $[\text{dép}] / K_D^{\text{récep}} > 1$ ). En d’autres mots, il faut suffisamment de déplétant pour que le ligand lie préférentiellement le déplétant au lieu du récepteur. Deuxièmement, ce système confirme que le paramètre clé de l’ultrasensibilité, tel que démontré par les simulations, est le ratio  $[\text{dép}] / K_D^{\text{dép}}$ , car une augmentation de la concentration du déplétant ou une diminution de  $K_D^{\text{dép}}$  facilite dans les deux cas la séquestration du ligand avec le déplétant. Finalement, ce système démontre que varier  $K_D^{\text{récep}}$  n’affecte pas l’ultrasensibilité, ce qui n’avait pas été démontré avant avec les simulations.

Similairement à ce mécanisme, un autre système d’ADN fut mis au point afin de caractériser le mécanisme de séquestration moléculaire dans un récepteur bivalent<sup>98</sup>. À l’instar de la situation précédente où le ligand peut lier soit le déplétant ou le récepteur, maintenant le

récepteur contient deux sites de liaison pour deux ligands différents où l'un des deux peut agir comme un déplétant (Fig. 1.12b, *haut*). Dans ce cas précis, la liaison des deux ligands est nécessaire pour l'activation du récepteur. Ainsi, si la liaison d'un premier ligand empêche la liaison du second par un effet de coopérativité négatif, il y aura séquestration par le premier ligand empêchant ainsi l'activation du récepteur. Inversement, par un effet de coopérativité positive, le premier ligand pourrait faciliter la liaison du second ligand et ainsi promouvoir l'activation du récepteur. Pour valider ce mécanisme impliquant un récepteur bivalent, un système simple d'ADN a été conçu où deux simples brins d'ADN agissent comme ligands (l'un avec une faible affinité et l'autre avec une grande affinité) et compétitionnent pour l'hybridation sur un autre simple brin d'ADN agissant comme récepteur (Fig. 1.12b, *bas*). Une coopérativité négative est obtenue lorsque les deux ligands d'ADN s'imbriquent l'un sur l'autre lors de l'hybridation et une coopérativité positive est obtenue lorsque la complémentarité des deux ligands avec le récepteur est parfaite. Ce modèle d'ADN, aussi appuyé par des simulations numériques, a permis de confirmer que l'ultrasensibilité est obtenue seulement en présence de coopérativité négative. À première vue, cette conclusion est contre-intuitive, car on associe la plupart du temps une augmentation de la sensibilité avec une coopérativité positive (*p. ex.* la faible plage dynamique de l'hémoglobine associée à son allostérie positive)<sup>77, 78</sup>. Toutefois, ce modèle d'ADN illustre bien comment la présence de coopérativité négative favorise la séquestration du récepteur sous forme de deux complexes dimériques (1:1) avec les deux ligands empêchant ainsi la formation d'un récepteur doublement lié (1:1:1). Une réponse ultrasensible est obtenue lorsque la concentration de ligands dépasse l'effet de séquestration du récepteur et favorise donc la formation d'un complexe 1:1:1 avec les deux ligands. En somme, l'utilisation de systèmes simples d'ADN a permis de confirmer expérimentalement le mécanisme de séquestration moléculaire ouvrant ainsi la porte, par exemple, au développement de capteurs électrochimiques<sup>150</sup>, à la régulation de facteur de transcription<sup>151</sup> et à une meilleure compréhension des voies de signalisations impliquant des protéines multimériques<sup>152</sup>.



**Figure 1.12 – Biochimie synthétique : le mécanisme de séquestration et les mécanismes cinétiques de l'ajustement induit et de la sélection conformationnelle.** (a) Utilisation d'une tige-boucle d'ADN pour comprendre le mécanisme de séquestration moléculaire. La figure fut adaptée et traduite de la référence 97, © 2011 Ricci et al. (b) Utilisation de simples duplexes d'ADN pour étudier le rôle de la coopérativité négative dans la séquestration d'un récepteur bivalent. La figure fut adaptée et traduite de la référence 98, © 2016 American Association for the Advancement of Science. (c) Utilisation d'une tige-boucle d'ADN pour étudier les mécanismes cinétiques de l'ajustement induit et de la sélection conformationnelle. La figure fut adaptée de la référence 153, © 2016 Carl Prévost-Tremblay. (d) Étude des mécanismes cinétiques de l'ajustement induit et de la sélection conformationnelle sur la liaison d'un ligand avec son aptamère séquestré. La figure fut adaptée et traduite de la référence 157, © 2017 The Royal Society of Chemistry.

#### 1.4.4 Les mécanismes de l'ajustement induit et de la sélection conformationnelle

Il existe deux voies mécanistiques décrivant la liaison d'un ligand à son récepteur soit 1) la voie de l'ajustement induit où la liaison du ligand induit un changement de conformation du récepteur qui accommode le ligand et 2) la voie de la sélection conformationnelle où le récepteur change spontanément de conformation permettant ainsi la liaison du ligand. Puisque les biomolécules naturelles possèdent une diversité structurale et thermodynamique trop grande pour les comparer efficacement, des chercheurs ont proposé un modèle d'interrupteurs moléculaires à base d'ADN<sup>153</sup>. Ce modèle comprend deux brins d'ADN complémentaires (jaune et bleu) avec le brin bleu ayant une boucle non appariée qui peut interagir avec le brin rouge qui agit comme le ligand (Fig. 1.12c). En changeant la composition des nucléotides du duplex formé entre le brin jaune et le brin bleu ainsi que les nucléotides entre le brin bleu et le brin rouge, il est possible de moduler respectivement la stabilité de l'interrupteur ( $K_S$ ) et la constante de dissociation du ligand ( $K_D$ ). La force de ce modèle réside dans la boucle du brin bleu où une

variation de sa longueur permet une comparaison directe de l'impact cinétique des deux mécanismes sur un système quasi identique. En effet, une longue boucle (13 nt) permet au ligand d'interagir avec l'interrupteur ce qui induit la dissociation du brin jaune (Fig. 1.12c, voie supérieure, ajustement induit). Inversement, une petite boucle (1 nt) ne permet pas cette préassociation avec le ligand et donc le ligand doit attendre la dissociation spontanée du brin jaune avant de lier le brin bleu (Fig. 1.12c, voie inférieure, sélection conformationnelle). Grâce à ce modèle simplifié, les chercheurs ont découvert que pour le même système (*p. ex.* le même interrupteur moléculaire) l'ajustement induit permet l'obtention de vitesses d'activation et de désactivation plus rapides comparativement au mécanisme par sélection conformationnelle<sup>153</sup>. De plus, ces résultats concordent avec plusieurs autres études théoriques sur ces deux mécanismes de liaison<sup>154, 155</sup>.

Indépendamment, un autre groupe de recherche a démontré les conséquences cinétiques de ces deux mécanismes sur la liaison d'un ligand avec un aptamère séquestré par un brin d'ADN<sup>156, 157</sup>. Comme mentionné précédemment, les aptamères sont généralement plus stables dans leur état de liaison. Pour maximiser le déplacement de population, une stratégie consiste à ajouter un brin séquestrant venant stabiliser l'état non liant de l'aptamère (Fig. 1.12d). Toutefois, une telle stratégie peut avoir un impact important sur la cinétique et la thermodynamique de liaison entre l'aptamère et son ligand. Par exemple, il fut démontré que la stratégie d'inhibition allostérique employée par Porchetta *et al.* (sections 1.2.4.1 et 1.4.2)<sup>144</sup> suit un mécanisme de sélection conformationnelle où le ligand (la cocaïne) doit attendre la dissociation spontanée de l'inhibiteur d'ADN avant de pouvoir lier l'aptamère<sup>157</sup>. Cela se traduit par une diminution de  $K_D$  apparent de liaison entre la cocaïne et l'aptamère. En contraste, si un tel système suivait un mécanisme par l'ajustement induit, le  $K_D$  apparent de liaison entre un ligand et son aptamère devrait être très similaire au  $K_D$  en absence d'inhibiteur<sup>156</sup>. En utilisant l'aptamère d'ATP, les chercheurs ont démontré que la présence de l'inhibiteur (ou brin séquestrant) près du site de liaison du ligand permet de réduire l'importance du mécanisme de sélection conformationnelle et de maximiser l'ajustement induit, car la présence de l'inhibiteur près des nucléotides importants pour la liaison du ligand aiderait à prédisposer la poche de liaison du ligand<sup>156</sup>. De plus, ils ont démontré que l'implémentation de l'ajustement induit en prédisposant l'inhibiteur

près du site de liaison fonctionne aussi avec des aptamères d'ARN, l'aptamère de thrombine et un riborégulateur, mais qu'elle ne fonctionne pas avec l'aptamère de cocaïne employé par Porchetta *et al.*, qui lui, emprunte uniquement un mécanisme de sélection conformationnelle, peu importe l'emplacement de l'inhibiteur<sup>156</sup>. Le fait que certains aptamères empruntent la voie de l'ajustement induit tandis que d'autres, la voie de la sélection conformationnelle, démontrent l'importance de développer des modèles simplifiés afin de mieux comprendre ces deux mécanismes d'un point de vue structurel, thermodynamique et cinétique.

## 1.5 Objectif de la thèse

Le but de la présente thèse consiste à mettre au point un nouveau mécanisme moléculaire permettant de programmer la plage dynamique d'interrupteurs moléculaires. En observant les mécanismes actuels, peu de mécanismes permettent d'augmenter ou de diminuer simultanément la plage dynamique. Généralement, les mécanismes basés sur l'allostérie ou la séquestration moléculaire permettent de diminuer la plage dynamique d'un interrupteur<sup>97, 98, 141-145</sup> tandis que les plages dynamiques étendues sont obtenues en combinant plusieurs différents interrupteurs<sup>144, 145, 150</sup>. Idéalement, il serait intéressant de mettre au point un mécanisme moléculaire permettant autant de diminuer que d'augmenter la plage dynamique d'un interrupteur afin que ce dernier accomode la plage de concentrations du ligand. En s'inspirant des complexes protéiques multimériques, il est envisageable de créer des interrupteurs moléculaires à base d'ADN comprenant plusieurs composantes afin d'exploiter la coopérativité d'interaction entre les diverses composantes de l'assemblage moléculaire. Il serait donc possible, avec une coopérativité positive, de créer des assemblages coopératifs et, avec une coopérativité négative, de créer des assemblages avec une plage dynamique étendue.

Le premier objectif consiste à mettre au point un modèle simple d'ADN à plusieurs composantes et d'étudier comment un tel assemblage permettrait d'obtenir des assemblages coopératifs ou étendus. La première stratégie qui sera employée pour former un tel assemblage consistera à fragmenter une structure unimoléculaire en plusieurs fragments qui pourront, par le biais de leurs interactions, reformer la structure originale (voir l'exemple de la Fig. 1.10a ci-haut). Évidemment, ce ne sont pas tous les systèmes qui peuvent résister à une fragmentation de leur

structure et il sera donc important de vérifier l'impact de la fragmentation sur la stabilité, la cinétique et l'intégrité structurale d'un système fragmenté. Il est donc probable que certaines mutations ou modifications rationnelles de la structure soient nécessaires afin de retrouver un système similaire au système non fragmenté. Une seconde stratégie qui sera employée pour former un assemblage moléculaire consiste à fusionner des domaines d'interaction par le biais d'un espaceur qui pourront interagir spécifiquement avec leur partenaire pour ainsi former un assemblage multimérique (voir l'exemple de la Fig. 1.10b ci-haut). Dans une telle stratégie, il sera important de bien comprendre le rôle de l'espaceur sur la coopérativité de liaison entre les domaines d'interactions fusionnés ensemble. Les informations tirées de ces deux stratégies d'assemblage serviront de base afin de mettre au point une première génération d'interrupteurs moléculaires multivalents.

Le second objectif consiste à utiliser les découvertes obtenues sur un système simplifié d'ADN et généraliser ces conclusions aux systèmes protéiques et leur évolution. Sachant que la majorité (> 55%) des protéines existent sous forme de complexes, il devrait y avoir certains avantages thermodynamiques, cinétiques ou fonctionnels derrière l'évolution de protéines monomériques vers des complexes dimériques, trimériques ou multimériques. Ainsi, il est possible que durant leur évolution, certaines protéines aient acquis des avantages fonctionnels similaires à ceux qui seront observés avec un système simplifié à base d'ADN. De plus, la fission et la fusion des gènes sont deux mécanismes employés durant l'évolution afin de créer des assemblages moléculaires, ce qui s'apparente aux stratégies qui seront employées pour créer des assemblages moléculaires d'ADN, soit la fragmentation d'une structure unimoléculaire en divers fragments et la fusion de domaines interagissant par le biais d'un espaceur. Ce modèle d'ADN sera donc une alternative moléculaire intéressante pour étudier expérimentalement la coopérativité d'assemblage des complexes protéiques tout en mettant en lumière certains avantages évolutifs.

Finalement, le dernier objectif consiste à conceptualiser rationnellement des interrupteurs moléculaires pouvant exploiter la coopérativité d'interaction d'assemblages moléculaires afin de programmer leur plage dynamique. Pour réaliser cet objectif, il est important de choisir un modèle pertinent avec un certain niveau de complexité afin de démontrer

l'universalité de la stratégie de création d'assemblages moléculaires. Pour ce faire, une ADNzyme, *c.-à-d.* un brin d'ADN avec une activité catalytique, sera artificiellement évoluée d'un système à deux composantes vers un système à trois composantes. De plus, l'activité de liaison d'une balise moléculaire à base d'ADN, de l'anglais « *molecular beacon* », sera modifiée en un système à trois composantes en ajoutant un nouveau brin d'ADN pouvant interagir avec la balise et son ligand, agissant ainsi comme un inhibiteur ou un activateur. Ainsi, cet effecteur multivalent permettra d'optimiser la plage dynamique de la balise moléculaire. Ce point est très important, car il permet de démontrer la viabilité de cette stratégie pour conceptualiser de nouvelles nanotechnologies avec une plage dynamique programmable pouvant avoir un impact immédiat dans les domaines des biocapteurs, de la délivrance de médicaments, de l'imagerie moléculaire, du théranostique et de l'informatique moléculaire<sup>24, 42</sup>.

## 1.6 Contenu de la thèse

Cette thèse est divisée en cinq chapitres comprenant les chapitres 2 et 3 qui discutent de deux méthodes inspirées de la Nature pour créer des assemblages moléculaires, soit la fragmentation et la fusion de domaines interagissant. Le chapitre 4, de son côté, relate comment exploiter la coopérativité d'interaction entre les diverses composantes d'un assemblage moléculaire afin d'optimiser la plage dynamique d'un interrupteur moléculaire. Finalement, le dernier chapitre reprend les grandes lignes de discussion des chapitres 2, 3 et 4 tout en ouvrant sur les perspectives de l'utilisation de simples assemblages moléculaires pour le développement de nouvelles nanotechnologies tout en offrant des pièces d'information sur l'évolution de protéines monomériques vers des complexes protéiques.

Plus précisément, le chapitre 2 explore les avantages et les inconvénients provenant de la fragmentation d'une structure unimoléculaire pour créer des assemblages moléculaires comprenant deux ou trois composantes. Ce chapitre illustrera le modèle simplifié d'ADN employé pour découvrir les conséquences de la fragmentation. Cet article publié dans la revue *Nature Chemistry* (Lauzon, D. & Vallée-Bélisle, A., Functional advantages of building nanosystems using multiple molecular components, *Nat. Chem.*, **2023**)<sup>158</sup> démontre, par le biais d'un modèle d'ADN et de simulations numériques, les nombreux avantages d'un système à trois composantes



permettant de moduler les propriétés d'assemblage d'un nanosystème. L'article explore aussi, d'un point de vue évolutif, les avantages régulateurs provenant de la fragmentation d'une protéine en plusieurs composantes et comment cela pourrait permettre aux nanomachines de la vie d'acquies de nouveaux mécanismes de régulation.

Le chapitre 3 explore une autre stratégie de conceptualisation permettant de créer des assemblages moléculaires à partir de domaines interagissants. Ce chapitre illustre comment divers domaines d'interaction peuvent être fusionnés grâce à un espaceur pour permettre leur interaction et ainsi former des assemblages moléculaires comprenant plusieurs composantes moléculaires. Cet article soumis à *Nature Chemical Biology* suivant l'acceptation du premier article (Lauzon, D. & Vallée-Bélisle, A., Programming the cooperativity of multimeric assemblies by tuning their compactness via optimal linker connection, *Nat. Chem. Biol.*, **2022**, soumis) démontre comment l'optimisation de l'espaceur non structuré entre les domaines interagissants permet la formation de nanostructures autoassemblées compactes et coopératives. Ainsi, en optimisant le niveau de coopérativité par le biais de l'espaceur, il est possible de programmer la plage dynamique d'un assemblage moléculaire. Cet article explore aussi, avec le même modèle simplifié d'ADN, comment cette optimisation de la coopérativité par le biais d'un espaceur non structuré pourrait permettre de mieux comprendre, par exemple, l'évolution des domaines intrinsèquement désordonnés que l'on observe chez plusieurs complexes protéiques.

Le chapitre 4 explore l'utilisation d'assemblage moléculaire à trois composantes afin d'optimiser la plage dynamique d'un interrupteur moléculaire à base d'ADN. Ainsi, un activateur multivalent, *c.-à-d.* une molécule qui augmente le nombre de sites d'interactions, lorsque lié à l'interrupteur moléculaire, permet d'augmenter ou de réduire la plage dynamique de l'interrupteur moléculaire qu'il régule. Cet article en revision par les pairs dans le *Journal of the American Chemical Society* (Lauzon, D. & Vallée-Bélisle, A., Multivalent activators: Bioinspired effectors to tune the affinity and cooperativity of molecular switches, *J. Am. Chem. Soc.*, **2022**, In revision) démontre comment les différents avantages des assemblages moléculaires à trois composantes observés dans les deux chapitres précédents peuvent facilement être incorporés dans des interrupteurs moléculaires telle une balise moléculaire. Cet article discute aussi des

avantages et des inconvénients thermodynamiques et cinétiques de cette stratégie par rapport aux autres stratégies existantes pour optimiser la plage dynamique d'un interrupteur moléculaire.

Finalement, le dernier chapitre regroupe les conclusions des trois chapitres précédents sous une perspective commune de création d'assemblages moléculaires coopératifs permettant l'optimisation de la plage dynamique d'un assemblage moléculaire multimérique. Ce chapitre envisage aussi comment les règles de conceptualisation démontrées dans ces différents projets peuvent avoir un impact important sur le développement de nanomachines à plusieurs composantes pouvant servir pour développer une nouvelle génération de biocapteurs et de systèmes de délivrance de médicament. De plus, ce chapitre explore, d'un point de vue spéculatif, les nombreux avantages évolutifs qui peuvent avoir été obtenus lorsqu'une protéine a évolué pour acquérir d'autres partenaires d'interactions que ce soit par fragmentation ou par fusion.

Outre ces travaux, j'ai eu l'opportunité de travailler et collaborer sur de nombreux projets de recherches durant mon doctorat qui ne se retrouveront malheureusement pas dans cette thèse. Tout récemment, les travaux du laboratoire sur l'étude des enzymes par l'entremise d'une nanoantenne fluorescente d'ADN furent publiés dans le journal *Nature Methods*<sup>159</sup>, travaux où j'ai développé un code permettant l'analyse des traces cinétiques mesurées par la nanoantenne d'ADN. Similairement, j'ai mis au point des codes afin de simuler le comportement et la réponse d'un biocapteur sous contrôle cinétique afin de mieux comprendre le niveau de programmabilité de ce nouveau mécanisme (manuscrits en préparation)<sup>160</sup>. J'ai aussi collaboré avec des collègues en conceptualisant et synthétisant des brins d'ADN modifiés pour la mise au point de systèmes de délivrance de médicaments, dont l'un est en préparation<sup>161</sup>, et les deux autres publiés respectivement dans le journal *Nature Communications*<sup>162</sup> et le journal *PNAS*<sup>163</sup>. À cela se rajoute une revue de littérature dans *Nanoscale* sur les interrupteurs moléculaires à base d'ADN où j'ai écrit la section portant sur la thermodynamique derrière la conceptualisation d'interrupteurs moléculaires<sup>24</sup> ainsi qu'un chapitre de livre que j'ai chapeauté sur le même sujet, mais dans une optique de création d'ordinateurs moléculaires<sup>164</sup>. Finalement, d'autres collaborations sont encore en cours comme l'exploration de la spectrométrie de masse en conditions natives pour étudier les assemblages moléculaires à base d'ADN (collaboration avec le groupe du Pr Vicki Wisowski, *Ohio State University*).

## 1.7 Références

1. Marsh, J. A.; Teichmann, S. A., Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **2015**, *84* (1), 551-575.
2. Marsh, J. A.; Teichmann, S. A., Protein Flexibility Facilitates Quaternary Structure Assembly and Evolution. *PLoS Biol.* **2014**, *12* (5), e1001870.
3. Bokov, K.; Steinberg, S. V., A hierarchical model for evolution of 23S ribosomal RNA. *Nature* **2009**, *457* (7232), 977-980.
4. Marianayagam, N. J.; Sunde, M.; Matthews, J. M., The power of two: protein dimerization in biology. *Trends Biochem. Sci.* **2004**, *29* (11), 618-625.
5. Lynch, M., The evolution of multimeric protein assemblages. *Mol. Biol. Evol.* **2012**, *29* (5), 1353-1366.
6. Hashimoto, K.; Nishi, H.; Bryant, S.; Panchenko, A. R., Caught in self-interaction: evolutionary and functional mechanisms of protein homooligomerization. *Phys. Biol.* **2011**, *8* (3), 035007.
7. Hagner, K.; Setayeshgar, S.; Lynch, M., Stochastic protein multimerization, activity, and fitness. *Phys. Rev. E* **2018**, *98* (6), 062401.
8. Korendovych, I. V.; DeGrado, W. F., De novo protein design, a retrospective. *Q. Rev. Biophys.* **2020**, *53*, e3.
9. Ballardini, R.; Balzani, V.; Credi, A.; Gandolfi, M. T.; Venturi, M., Artificial Molecular-Level Machines: Which Energy To Make Them Work? *Acc. Chem. Res.* **2001**, *34* (6), 445-455.
10. Balzani, V.; Credi, A.; Venturi, M., Light powered molecular machines. *Chem. Soc. Rev.* **2009**, *38* (6), 1542-1550.
11. Boyer, P. D., A Research Journey with ATP Synthase. *J. Biol. Chem.* **2002**, *277* (42), 39045-39061.
12. Goodsell, D. S., *The Machinery of Life*. 2nd ed.; Springer Science + Business Media: 2009.
13. Olzmann, J. A.; Carvalho, P., Dynamics and functions of lipid droplets. *Nat. Rev. Mol. Cell Biol.* **2019**, *20* (3), 137-155.
14. Bucior, I.; Burger, M. M., Carbohydrate-carbohydrate interactions in cell recognition. *Curr. Opin. Struct. Biol.* **2004**, *14* (5), 631-637.

15. Vallée-Bélisle, A.; Plaxco, K. W., Structure-switching biosensors: inspired by Nature. *Curr. Opin. Struct. Biol.* **2010**, *20* (4), 518-526.
16. Farokhzad, O. C.; Langer, R., Impact of Nanotechnology on Drug Delivery. *ACS Nano* **2009**, *3* (1), 16-20.
17. Szaciłowski, K., Digital Information Processing in Molecular Systems. *Chem. Rev.* **2008**, *108* (9), 3481-3548.
18. Murata, S.; Konagaya, A.; Kobayashi, S.; Saito, H.; Hagiya, M., Molecular Robotics: A New Paradigm for Artifacts. *New Gener. Comput.* **2013**, *31* (1), 27-45.
19. Newberry, R. W.; Raines, R. T., Secondary Forces in Protein Folding. *ACS Chem. Biol.* **2019**, *14* (8), 1677-1686.
20. Tsai, C.-J.; Lin, S. L.; Wolfson, H. J.; Nussinov, R., Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect. *Protein Sci.* **1997**, *6* (1), 53-64.
21. Feng, B.; Sosa, R. P.; Mårtensson, A. K. F.; Jiang, K.; Tong, A.; Dorfman, K. D.; Takahashi, M.; Lincoln, P.; Bustamante, C. J.; Westerlund, F.; Nordén, B., Hydrophobic catalysis and a potential biological role of DNA unstacking induced by environment effects. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (35), 17169.
22. Goodsell, D. S.; Olson, A. J., Structural Symmetry and Protein Function. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29* (1), 105-153.
23. Rinaldi, S., The Diverse World of Foldamers: Endless Possibilities of Self-Assembly. *Molecules* **2020**, *25* (14).
24. Harroun, S. G.; Prévost-Tremblay, C.; Lauzon, D.; Desrosiers, A.; Wang, X.; Pedro, L.; Vallée-Bélisle, A., Programmable DNA switches and their applications. *Nanoscale* **2018**, *10* (10), 4607-4641.
25. Drexler, K. E., Molecular Nanomachines: Physical Principles and Implementation Strategies. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23* (1), 377-405.
26. Holmgren, M.; Wagg, J.; Bezanilla, F.; Rakowski, R. F.; De Weer, P.; Gadsby, D. C., Three distinct and sequential steps in the release of sodium ions by the Na<sup>+</sup>/K<sup>+</sup>-ATPase. *Nature* **2000**, *403* (6772), 898-901.

27. Mattia, E.; Otto, S., Supramolecular systems chemistry. *Nat. Nanotechnol.* **2015**, *10* (2), 111-119.
28. Rumfeldt, J. A. O.; Galvagnion, C.; Vassall, K. A.; Meiering, E. M., Conformational stability and folding mechanisms of dimeric proteins. *Prog. Biophys. Mol. Biol.* **2008**, *98* (1), 61-84.
29. Miller, S.; Lesk, A. M.; Janin, J.; Chothia, C., The accessible surface area and stability of oligomeric proteins. *Nature* **1987**, *328* (6133), 834-836.
30. Lumry, R.; Rajender, S., Enthalpy–entropy compensation phenomena in water solutions of proteins and small molecules: A ubiquitous property of water. *Biopolymers* **1970**, *9* (10), 1125-1227.
31. Parker, J., Errors and alternatives in reading the universal genetic code. *Microbiol. Rev.* **1989**, *53* (3), 273-298.
32. Bremer, A.; Aebi, U., The structure of the F-actin filament and the actin molecule. *Curr. Opin. Cell Biol.* **1992**, *4* (1), 20-26.
33. Gwyther, R. E. A.; Jones, D. D.; Worthy, H. L., Better together: building protein oligomers naturally and by design. *Biochem. Soc. Trans.* **2019**, *47* (6), 1773-1780.
34. Behrendt, R.; White, P.; Offer, J., Advances in Fmoc solid-phase peptide synthesis. *J. Pept. Sci.* **2016**, *22* (1), 4-27.
35. Hughes, R. A.; Ellington, A. D., Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology. *Cold Spring Harbor Perspect. Biol.* **2017**, *9* (1), a023812.
36. Levin, A.; Hakala, T. A.; Schnaider, L.; Bernardes, G. J. L.; Gazit, E.; Knowles, T. P. J., Biomimetic peptide self-assembly for functional materials. *Nat. Rev. Chem.* **2020**, *4* (11), 615-634.
37. Seeman, N. C.; Sleiman, H. F., DNA nanotechnology. *Nat. Rev. Mater.* **2017**, *3* (1), 17068.
38. Grueninger, D.; Treiber, N.; Ziegler Mathias, O. P.; Koetter Jochen, W. A.; Schulze, M.-S.; Schulz Georg, E., Designed Protein-Protein Association. *Science* **2008**, *319* (5860), 206-209.
39. Hashimoto, K.; Panchenko, A. R., Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (47), 20352.
40. Mascarenhas, N. M.; Gosavi, S., Understanding protein domain-swapping using structure-based models of protein folding. *Prog. Biophys. Mol. Biol.* **2017**, *128*, 113-120.

41. Kummerfeld, S. K.; Teichmann, S. A., Relative rates of gene fusion and fission in multi-domain proteins. *Trends Genet.* **2005**, *21* (1), 25-30.
42. Feringa, B. L., The Art of Building Small: From Molecular Switches to Molecular Motors. *J. Org. Chem.* **2007**, *72* (18), 6635-6652.
43. Venkatakrisnan, A. J.; Deupi, X.; Lebon, G.; Tate, C. G.; Schertler, G. F.; Babu, M. M., Molecular signatures of G-protein-coupled receptors. *Nature* **2013**, *494* (7436), 185-194.
44. Weis, W. I.; Kobilka, B. K., The Molecular Basis of G Protein–Coupled Receptor Activation. *Annu. Rev. Biochem.* **2014**, *87* (1), 897-919.
45. Dougherty, M. J.; Arnold, F. H., Directed evolution: new parts and optimized function. *Curr. Opin. Biotechnol.* **2009**, *20* (4), 486-491.
46. Collins, C. H.; Leadbetter, J. R.; Arnold, F. H., Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR. *Nat. Biotechnol.* **2006**, *24* (6), 708-712.
47. Serganov, A.; Nudler, E., A Decade of Riboswitches. *Cell* **2013**, *152* (1), 17-24.
48. Wang, T.; Chen, C.; Larcher, L. M.; Barrero, R. A.; Veedu, R. N., Three decades of nucleic acid aptamer technologies: Lessons learned, progress and opportunities on aptamer development. *Biotechnol. Adv.* **2019**, *37* (1), 28-50.
49. Willner, I.; Shlyahovsky, B.; Zayats, M.; Willner, B., DNAzymes for sensing, nanobiotechnology and logic gate applications. *Chem. Soc. Rev.* **2008**, *37* (6), 1153-1165.
50. Beharry, A. A.; Woolley, G. A., Azobenzene photoswitches for biomolecules. *Chem. Soc. Rev.* **2011**, *40* (8), 4422-4437.
51. Xue, M.; Yang, Y.; Chi, X.; Yan, X.; Huang, F., Development of Pseudorotaxanes and Rotaxanes: From Synthesis to Stimuli-Responsive Motions to Applications. *Chem. Rev.* **2015**, *115* (15), 7398-7501.
52. Liu, Z.; Nalluri, S. K. M.; Stoddart, J. F., Surveying macrocyclic chemistry: from flexible crown ethers to rigid cyclophanes. *Chem. Soc. Rev.* **2017**, *46* (9), 2459-2478.
53. Wei, M.; Gao, Y.; Li, X.; Serpe, M. J., Stimuli-responsive polymers and their applications. *Polym. Chem.* **2017**, *8* (1), 127-143.

54. Harris, J. D.; Moran, M. J.; Aprahamian, I., New molecular switch architectures. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (38), 9414.
55. Kay, E. R.; Leigh, D. A.; Zerbetto, F., Synthetic Molecular Motors and Mechanical Machines. *Angew. Chem. Int. Ed.* **2007**, *46* (1-2), 72-191.
56. Saimi, Y.; Kung, C., Calmodulin as an Ion Channel Subunit. *Annu. Rev. Physiol.* **2002**, *64* (1), 289-311.
57. Ng, A. H.; Nguyen, T. H.; Gómez-Schiavon, M.; Dods, G.; Langan, R. A.; Boyken, S. E.; Samson, J. A.; Waldburger, L. M.; Dueber, J. E.; Baker, D.; El-Samad, H., Modular and tunable biological feedback control using a de novo protein switch. *Nature* **2019**, *572* (7768), 265-269.
58. van der Horst, M. A.; Hellingwerf, K. J., Photoreceptor Proteins, "Star Actors of Modern Times": A Review of the Functional Dynamics in the Structure of Representative Members of Six Different Photoreceptor Families. *Acc. Chem. Res.* **2004**, *37* (1), 13-20.
59. Claaßen, C.; Gerlach, T.; Rother, D., Stimulus-Responsive Regulation of Enzyme Activity for One-Step and Multi-Step Syntheses. *Adv. Synth. Catal.* **2019**, *361* (11), 2387-2401.
60. Plaxco, K. W.; Soh, H. T., Switch-based biosensors: a new approach towards real-time, in vivo molecular detection. *Trends Biotechnol.* **2011**, *29* (1), 1-5.
61. Mura, S.; Nicolas, J.; Couvreur, P., Stimuli-responsive nanocarriers for drug delivery. *Nat. Mater.* **2013**, *12* (11), 991-1003.
62. Hamley, I. W., Self-assembly of amphiphilic peptides. *Soft Matter* **2011**, *7* (9), 4122-4138.
63. Kakuta, T.; Yamagishi, T.-a.; Ogoshi, T., Stimuli-Responsive Supramolecular Assemblies Constructed from Pillar[n]arenes. *Acc. Chem. Res.* **2018**, *51* (7), 1656-1666.
64. Grzelczak, M.; Liz-Marzán, L. M.; Klajn, R., Stimuli-responsive self-assembly of nanoparticles. *Chem. Soc. Rev.* **2019**, *48* (5), 1342-1361.
65. Wang, S.; Urban, M. W., Self-healing polymers. *Nat. Rev. Mater.* **2020**, *5* (8), 562-583.
66. Feldman, H. A., Mathematical theory of complex ligand-binding systems at equilibrium: Some methods for parameter fitting. *Anal. Biochem.* **1972**, *48* (2), 317-338.
67. Jarmoskaite, I.; AlSadhan, I.; Vaidyanathan, P. P.; Herschlag, D., How to measure and evaluate binding affinities. *eLife* **2020**, *9*, e57264.

68. Esteban Fernández de Ávila, B.; Watkins, H. M.; Pingarrón, J. M.; Plaxco, K. W.; Palleschi, G.; Ricci, F., Determinants of the Detection Limit and Specificity of Surface-Based Biosensors. *Anal. Chem.* **2013**, *85* (14), 6593-6597.
69. Bisswanger, H., Multiple Equilibria. In *Enzyme Kinetics: Principles and Methods*, 2nd ed.; Bisswanger, H., Ed. WILEY-VCH Verlag GmbH & Co. KGaA: Weinheim, 2008; pp 7-58.
70. Weiss, J. N., The Hill equation revisited: uses and misuses. *FASEB J.* **1997**, *11* (11), 835-841.
71. Vallée-Bélisle, A.; Ricci, F.; Plaxco, K. W., Thermodynamic basis for the optimization of binding-induced biomolecular switches and structure-switching biosensors. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (33), 13802.
72. Tsai, C.-J.; Ma, B.; Nussinov, R., Folding and binding cascades: Shifts in energy landscapes. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (18), 9970.
73. Cui, Q.; Karplus, M., Allostery and cooperativity revisited. *Protein Sci.* **2008**, *17* (8), 1295-1307.
74. Kohn, J. E.; Plaxco, K. W., Engineering a signal transduction mechanism for protein-based biosensors. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (31), 10841.
75. Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z., Intrinsically disordered protein. *J. Mol. Graphics Modell.* **2001**, *19* (1), 26-59.
76. Vallée-Bélisle, A.; Ricci, F.; Plaxco, K. W., Engineering Biosensors with Extended, Narrowed, or Arbitrarily Edited Dynamic Range. *J. Am. Chem. Soc.* **2012**, *134* (6), 2876-2879.
77. Whitty, A., Cooperativity and biological complexity. *Nat. Chem. Biol.* **2008**, *4* (8), 435-439.
78. Williamson, J. R., Cooperativity in macromolecular assembly. *Nat. Chem. Biol.* **2008**, *4* (8), 458-465.
79. Hunter, C. A.; Anderson, H. L., What is Cooperativity? *Angew. Chem. Int. Ed.* **2009**, *48* (41), 7488-7499.
80. von Krbek, L. K. S.; Schalley, C. A.; Thordarson, P., Assessing cooperativity in supramolecular systems. *Chem. Soc. Rev.* **2017**, *46* (9), 2622-2637.
81. Li, Y.; Wang, Y.; Huang, G.; Gao, J., Cooperativity Principles in Self-Assembled Nanomedicine. *Chem. Rev.* **2018**, *118* (11), 5359-5391.



82. Ercolani, G.; Schiaffino, L., Allosteric, Chelate, and Interannular Cooperativity: A Mise au Point. *Angew. Chem. Int. Ed.* **2011**, *50* (8), 1762-1768.
83. Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N., Structural and Energetic Basis of Allostery. *Annu. Rev. Biophys.* **2012**, *41* (1), 585-609.
84. Koshland, D. E.; Némethy, G.; Filmer, D., Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits\*. *Biochemistry* **1966**, *5* (1), 365-385.
85. Monod, J.; Wyman, J.; Changeux, J.-P., On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **1965**, *12* (1), 88-118.
86. Hilser, V. J.; Thompson, E. B., Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (20), 8311.
87. Ferreon, A. C. M.; Ferreon, J. C.; Wright, P. E.; Deniz, A. A., Modulation of allostery by protein intrinsic disorder. *Nature* **2013**, *498* (7454), 390-394.
88. von Krbeek, L. K. S.; Achazi, A. J.; Solleder, M.; Weber, M.; Paulus, B.; Schalley, C. A., Allosteric and Chelate Cooperativity in Divalent Crown Ether/Ammonium Complexes with Strong Binding Enhancement. *Chem. Eur. J.* **2016**, *22* (43), 15475-15484.
89. Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J., The ensemble nature of allostery. *Nature* **2014**, *508* (7496), 331-339.
90. Ricci, F.; Vallée-Bélisle, A.; Porchetta, A.; Plaxco, K. W., Rational Design of Allosteric Inhibitors and Activators Using the Population-Shift Model: In Vitro Validation and Application to an Artificial Biosensor. *J. Am. Chem. Soc.* **2012**, *134* (37), 15177-15180.
91. Tame, J. R.; Vallone, B., The structures of deoxy human haemoglobin and the mutant Hb Tyr $\alpha$ 42His at 120 K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2000**, *56* (Pt 7), 805-11.
92. Page, M. I.; Jencks, W. P., Entropic Contributions to Rate Accelerations in Enzymic and Intramolecular Reactions and the Chelate Effect. *Proc. Natl. Acad. Sci. U. S. A.* **1971**, *68* (8), 1678.
93. Calderone, C. T.; Williams, D. H., An Enthalpic Component in Cooperativity: The Relationship between Enthalpy, Entropy, and Noncovalent Structure in Weak Associations. *J. Am. Chem. Soc.* **2001**, *123* (26), 6262-6267.
94. Cacciapaglia, R.; Di Stefano, S.; Mandolini, L., Effective Molarities in Supramolecular Catalysis of Two-Substrate Reactions. *Acc. Chem. Res.* **2004**, *37* (2), 113-122.

95. Jencks, W. P., On the attribution and additivity of binding energies. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78* (7), 4046.
96. Camara-Campos, A.; Musumeci, D.; Hunter, C. A.; Turega, S., Chemical Double Mutant Cycles for the Quantification of Cooperativity in H-Bonded Complexes. *J. Am. Chem. Soc.* **2009**, *131* (51), 18518-18524.
97. Ricci, F.; Vallée-Bélisle, A.; Plaxco, K. W., High-Precision, In Vitro Validation of the Sequestration Mechanism for Generating Ultrasensitive Dose-Response Curves in Regulatory Networks. *PLoS Comput. Biol.* **2011**, *7* (10), e1002171.
98. Ha, S. H.; Ferrell, J. E., Thresholds and ultrasensitivity from negative cooperativity. *Science* **2016**, *352* (6288), 990-993.
99. Xu, Y.; McSally, J.; Andricioaei, I.; Al-Hashimi, H. M., Modulation of Hoogsteen dynamics on DNA recognition. *Nat. Commun.* **2018**, *9* (1), 1473.
100. Chandrasekaran, A. R.; Rusling, D. A., Triplex-forming oligonucleotides: a third strand for DNA nanotechnology. *Nucleic Acids Res.* **2018**, *46* (3), 1021-1037.
101. Tuerk, C.; Gold, L., Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science* **1990**, *249* (4968), 505-510.
102. Cai, S.; Yan, J.; Xiong, H.; Liu, Y.; Peng, D.; Liu, Z., Investigations on the interface of nucleic acid aptamers and binding targets. *Analyst* **2018**, *143* (22), 5317-5338.
103. Farzin, L.; Shamsipur, M.; Sheibani, S., A review: Aptamer-based analytical strategies using the nanomaterials for environmental and human monitoring of toxic heavy metals. *Talanta* **2017**, *174*, 619-627.
104. Seeman, N. C., Nucleic acid junctions and lattices. *J. Theor. Biol.* **1982**, *99* (2), 237-247.
105. Dong, Y.; Yao, C.; Zhu, Y.; Yang, L.; Luo, D.; Yang, D., DNA Functional Materials Assembled from Branched DNA: Design, Synthesis, and Applications. *Chem. Rev.* **2020**, *120* (17), 9420-9481.
106. Holliday, R., A mechanism for gene conversion in fungi. *Genet. Res.* **1964**, *5* (2), 282-304.
107. Sadowski, J. P.; Calvert, C. R.; Zhang, D. Y.; Pierce, N. A.; Yin, P., Developmental Self-Assembly of a DNA Tetrahedron. *ACS Nano* **2014**, *8* (4), 3251-3259.
108. Chen, J.; Seeman, N. C., Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **1991**, *350* (6319), 631-633.

109. Dey, S.; Fan, C.; Gothelf, K. V.; Li, J.; Lin, C.; Liu, L.; Liu, N.; Nijenhuis, M. A. D.; Saccà, B.; Simmel, F. C.; Yan, H.; Zhan, P., DNA origami. *Nat. Rev. Methods Primers* **2021**, *1* (1), 13.
110. Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A., NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **2011**, *32* (1), 170-173.
111. Zuker, M., Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, *31* (13), 3406-3415.
112. Owczarzy, R.; Tataurov, A. V.; Wu, Y.; Manthey, J. A.; McQuisten, K. A.; Almagbrazi, H. G.; Pedersen, K. F.; Lin, Y.; Garretson, J.; McEntaggart, N. O.; Sailor, C. A.; Dawson, R. B.; Peek, A. S., IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Res.* **2008**, *36* (suppl\_2), W163-W169.
113. SantaLucia, J., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (4), 1460.
114. SantaLucia, J.; Hicks, D., The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33* (1), 415-440.
115. Saccà, B.; Meyer, R.; Feldkamp, U.; Schroeder, H.; Niemeyer, C. M., High-Throughput, Real-Time Monitoring of the Self-Assembly of DNA Nanostructures by FRET Spectroscopy. *Angew. Chem. Int. Ed.* **2008**, *47* (11), 2135-2137.
116. Idili, A.; Ricci, F.; Vallée-Bélisle, A., Determining the folding and binding free energy of DNA-based nanodevices and nanoswitches using urea titration curves. *Nucleic Acids Res.* **2017**, *45* (13), 7571-7580.
117. Owczarzy, R., Melting temperatures of nucleic acids: Discrepancies in analysis. *Biophys. Chem.* **2005**, *117* (3), 207-215.
118. You, Y.; Tataurov, A. V.; Owczarzy, R., Measuring thermodynamic details of DNA hybridization using fluorescence. *Biopolymers* **2011**, *95* (7), 472-486.
119. Hutton, J. R., Renaturation Kinetics and thermal stability of DNA in aqueous solutions of formamide and urea. *Nucleic Acids Res.* **1977**, *4* (10), 3537-3555.

120. Ramakrishnan, S.; Krainer, G.; Grundmeier, G.; Schlierf, M.; Keller, A., Structural stability of DNA origami nanostructures in the presence of chaotropic agents. *Nanoscale* **2016**, *8* (19), 10398-10405.
121. Shelton, V. M.; Sosnick, T. R.; Pan, T., Applicability of Urea in the Thermodynamic Analysis of Secondary and Tertiary RNA Folding. *Biochemistry* **1999**, *38* (51), 16831-16839.
122. Lawrence, C.; Vallée-Bélisle, A.; Pfeil, S. H.; de Mornay, D.; Lipman, E. A.; Plaxco, K. W., A comparison of the folding kinetics of a small, artificially selected DNA aptamer with those of equivalently simple naturally occurring proteins. *Protein Sci.* **2014**, *23* (1), 56-66.
123. Toulmé, J.-J.; Giegé, R., Les aptamères : des ligands et des catalyseurs oligonucléotidiques obtenus par sélection in vitro. *Med Sci (Paris)* **1998**, *14*, 155-166.
124. Chen, A.; Yan, M.; Yang, S., Split aptamers and their applications in sandwich aptasensors. *TrAC, Trends Anal. Chem.* **2016**, *80*, 581-593.
125. Debais, M.; Lelievre, A.; Smietana, M.; Müller, S., Splitting aptamers and nucleic acid enzymes for the development of advanced biosensors. *Nucleic Acids Res.* **2020**, *48* (7), 3400-3422.
126. Feigon, J.; Dieckmann, T.; Smith, F. W., Aptamer structures from A to  $\zeta$ . *Chem. Biol.* **1996**, *3* (8), 611-617.
127. Huizenga, D. E.; Szostak, J. W., A DNA Aptamer That Binds Adenosine and ATP. *Biochemistry* **1995**, *34* (2), 656-665.
128. Stojanovic, M. N.; de Prada, P.; Landry, D. W., Fluorescent Sensors Based on Aptamer Self-Assembly. *J. Am. Chem. Soc.* **2000**, *122* (46), 11547-11548.
129. Zhou, J.; Rossi, J. J., Cell-type-specific, Aptamer-functionalized Agents for Targeted Disease Therapy. *Mol. Ther. Nucleic Acids* **2014**, *3*, e169.
130. Zhou, J.; Swiderski, P.; Li, H.; Zhang, J.; Neff, C. P.; Akkina, R.; Rossi, J. J., Selection, characterization and application of new RNA HIV gp 120 aptamers for facile delivery of Dicer substrate siRNAs into HIV infected cells. *Nucleic Acids Res.* **2009**, *37* (9), 3094-3109.
131. Zhao, D.; Kong, Y.; Zhao, S.; Xing, H., Engineering Functional DNA-Protein Conjugates for Biosensing, Biomedical, and Nanoassembly Applications. *Top. Curr. Chem.* **2020**, *378*, 41.
132. McLaughlin, C. K.; Hamblin, G. D.; Sleiman, H. F., Supramolecular DNA assembly. *Chem. Soc. Rev.* **2011**, *40* (12), 5647-5656.

133. Yang, H.; Rys, A. Z.; McLaughlin, C. K.; Sleiman, H. F., Templated Ligand Environments for the Selective Incorporation of Different Metals into DNA. *Angew. Chem. Int. Ed.* **2009**, *48* (52), 9919-9923.
134. He, S.; Ge, Z.; Zuo, X.; Fan, C.; Mao, X., Dynamic regulation of DNA nanostructures by noncanonical nucleic acids. *NPG Asia Mater.* **2021**, *13* (1), 42.
135. Zhu, D.; Pei, H.; Yao, G.; Wang, L.; Su, S.; Chao, J.; Wang, L.; Aldalbahi, A.; Song, S.; Shi, J.; Hu, J.; Fan, C.; Zuo, X., A Surface-Confined Proton-Driven DNA Pump Using a Dynamic 3D DNA Scaffold. *Adv. Mater.* **2016**, *28* (32), 6860-6865.
136. Zhou, Y.-J.; Wan, Y.-H.; Nie, C.-P.; Zhang, J.; Chen, T.-T.; Chu, X., Molecular Switching of a Self-Assembled 3D DNA Nanomachine for Spatiotemporal pH Mapping in Living Cells. *Anal. Chem.* **2019**, *91* (16), 10366-10370.
137. Zhou, Z.; Sohn, Y. S.; Nechushtai, R.; Willner, I., DNA Tetrahedra Modules as Versatile Optical Sensing Platforms for Multiplexed Analysis of miRNAs, Endonucleases, and Aptamer–Ligand Complexes. *ACS Nano* **2020**, *14* (7), 9021-9031.
138. Douglas Shawn, M.; Bachelet, I.; Church George, M., A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science* **2012**, *335* (6070), 831-834.
139. Dohno, C.; Atsumi, H.; Nakatani, K., Ligand inducible assembly of a DNA tetrahedron. *Chem. Commun.* **2011**, *47* (12), 3499-3501.
140. Tokuriki, N.; Tawfik, D. S., Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **2009**, *19* (5), 596-604.
141. Simon, A. J.; Vallée-Bélisle, A.; Ricci, F.; Watkins, H. M.; Plaxco, K. W., Using the Population-Shift Mechanism to Rationally Introduce “Hill-type” Cooperativity into a Normally Non-Cooperative Receptor. *Angew. Chem. Int. Ed.* **2014**, *53* (36), 9471-9475.
142. Mariottini, D.; Idili, A.; Vallée-Bélisle, A.; Plaxco, K. W.; Ricci, F., A DNA Nanodevice That Loads and Releases a Cargo with Hemoglobin-Like Allosteric Control and Cooperativity. *Nano Lett.* **2017**, *17* (5), 3225-3230.
143. Desrosiers, A.; Vallée-Bélisle, A., Nature-inspired DNA switches: applications in medicine. *Nanomedicine* **2017**, *12* (3), 175-179.

144. Porchetta, A.; Vallée-Bélisle, A.; Plaxco, K. W.; Ricci, F., Using Distal-Site Mutations and Allosteric Inhibition To Tune, Extend, and Narrow the Useful Dynamic Range of Aptamer-Based Sensors. *J. Am. Chem. Soc.* **2012**, *134* (51), 20601-20604.
145. Porchetta, A.; Vallée-Bélisle, A.; Plaxco, K. W.; Ricci, F., Allosterically Tunable, DNA-Based Switches Triggered by Heavy Metals. *J. Am. Chem. Soc.* **2013**, *135* (36), 13238-13241.
146. Liu, L.; Yuan, M.; Jin, Y.; Zhou, G.; Li, T.; Li, L.; Peng, H.; Chen, W., Tunable Dual-Effector Allostery System for Nucleic Acid Analysis with Enhanced Sensitivity and an Extended Dynamic Range. *Anal. Chem.* **2021**, *93* (23), 8170-8177.
147. Li, H.; Zhou, X.; Yao, D.; Liang, H., pH-Responsive spherical nucleic acid for intracellular lysosome imaging and an effective drug delivery system. *Chem. Commun.* **2018**, *54* (28), 3520-3523.
148. Buchler, N. E.; Louis, M., Molecular Titration and Ultrasensitivity in Regulatory Networks. *J. Mol. Biol.* **2008**, *384* (5), 1106-1119.
149. Buchler, N. E.; Cross, F. R., Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol. Syst. Biol.* **2009**, *5* (1), 272.
150. Kang, D.; Vallée-Bélisle, A.; Porchetta, A.; Plaxco, K. W.; Ricci, F., Re-engineering Electrochemical Biosensors To Narrow or Extend Their Useful Dynamic Range. *Angew. Chem. Int. Ed.* **2012**, *51* (27), 6717-6721.
151. Brewster, Robert C.; Weinert, Franz M.; Garcia, Hernan G.; Song, D.; Rydenfelt, M.; Phillips, R., The Transcription Factor Titration Effect Dictates Level of Gene Expression. *Cell* **2014**, *156* (6), 1312-1323.
152. Antebi, Y. E.; Nandagopal, N.; Elowitz, M. B., An operational view of intercellular signaling pathways. *Curr. Opin. Syst. Biol.* **2017**, *1*, 16-24.
153. Prévost-Tremblay, C. Recréer et comprendre les mécanismes de liaison des biomolécules à l'aide d'un modèle d'ADN. Mémoire de maîtrise, Université de Montréal, Montréal, QC, 2017.
154. Weikl, T. R.; von Deuster, C., Selected-fit versus induced-fit protein binding: Kinetic differences and mutational analysis. *Proteins: Struct., Funct., Bioinf.* **2009**, *75* (1), 104-110.
155. Tummino, P. J.; Copeland, R. A., Residence Time of Receptor–Ligand Complexes and Its Effect on Biological Function. *Biochemistry* **2008**, *47* (20), 5481-5492.

156. Munzar, J. D.; Ng, A.; Juncker, D., Comprehensive profiling of the ligand binding landscapes of duplexed aptamer families reveals widespread induced fit. *Nat. Commun.* **2018**, *9* (1), 343.
157. Munzar, J. D.; Ng, A.; Corrado, M.; Juncker, D., Complementary oligonucleotides regulate induced fit ligand binding in duplexed aptamers. *Chem. Sci.* **2017**, *8* (3), 2251-2256.
158. Lauzon, D.; Vallée-Bélisle, A., Functional advantages of building nanosystems using multiple molecular components. *Nat. Chem.* **2023**.
159. Harroun, S. G.; Lauzon, D.; Ebert, M. C. C. J. C.; Desrosiers, A.; Wang, X.; Vallée-Bélisle, A., Monitoring protein conformational changes using fluorescent nanoantennas. *Nat. Methods* **2022**, *19* (1), 71-80.
160. Zhu, G.; Lauzon, D.; Prévost-Tremblay, C.; Adhikari, B. R.; Vallée-Bélisle, A., Rapid, one-step molecular detection in a drop of blood using kinetically programmed constitutional dynamic chemistry. *Nat. Chem.* **2023**, Manuscript in preparation.
161. Desrosiers, A.; Hassine, S.; Lauzon, D.; Vallée-Bélisle, A., Application of molecular buffers in vitro and with the in vivo murine malaria disease model *Plasmodium chabaudii*. **2023**, Manuscript in preparation.
162. Desrosiers, A.; Derbali, R. M.; Hassine, S.; Berdugo, J.; Long, V.; Lauzon, D.; De Guire, V.; Fiset, C.; DesGroseillers, L.; Leblond Chain, J.; Vallée-Bélisle, A., Programmable self-regulated molecular buffers for precise sustained drug delivery. *Nat. Commun.* **2022**, *13* (1), 6504.
163. Phan, H. T.; Lauzon, D.; Vallée-Bélisle, A.; Angioletti-Uberti, S.; Leblond Chain, J.; Giasson, S., Bimodal brush-functionalized nanoparticles selective to receptor surface density. *Proc. Natl. Acad. Sci. U. S. A.* **2023**, *120* (3), e2208377120.
164. Lauzon, D.; Zhu, G.; Vallée-Bélisle, A., Engineering DNA Switches for DNA Computing Applications. In *DNA- and RNA-Based Computing Systems*, Katz, E., Ed. WILEY-VCH GmbH: Weinheim, Germany, 2021; pp 105-124.

# Chapitre 2 – Functional advantages of building nanosystems using multiple molecular components\*

## 2.1 Abstract

What are the advantages and disadvantages of building nanosystems using one, two or multiple molecular components? More than 55% of all the natural nanomachines in living organisms are multimeric and likely exploit the self-assembly of their components to provide functional benefits. However, these specific benefits remain relatively unexplored at the thermodynamic, kinetic, and functional levels. To explore this question, we herein use theory and a simple DNA-based model that was fragmented into two and three molecular components to obtain the same exact nanostructures built with different number of components. We first found that despite its slower assembly rate, a system built with three components may undergo a more cooperative assembly transition from less preorganized components, which facilitate the creation of novel functionalities. Using simple variations of its components (*e.g.*, concentration, ratio, and mutations), we also found that trimeric nanosystems display a much higher level of programmability than their dimeric counterparts by being able to assemble with various levels of cooperativity, self-inhibition, and time-dependent properties. Using these findings, we show how two simple strategies (*e.g.*, fragmenting and adding components) can be employed to efficiently program the regulatory function of a more complex, artificially selected RNA-cleaving catalytic nanosystem. These molecular assembly strategies have been employed by natural biosystems and are expected to provide nanoengineers and protein engineers with the thermodynamic and kinetics framework to develop and optimize the next generation of self-regulated nanotechnologies.

---

\* Article publié dans le journal *Nature Chemistry*: Lauzon, D.; Vallée-Bélisle, A. *Nat. Chem.* 2023. D.L. et A.V.B. ont conjointement conceptualisé les expériences. D.L. a réalisé toutes les expériences et les simulations numériques. D.L. et A.V.B. ont conjointement analysé et interprété les résultats. D.L. a écrit la première version du manuscrit. D.L. et A.V.B. ont conjointement écrit la version finale. A. V. B. a supervisé la réalisation du projet.



## 2.2 Introduction

Finely regulated self-assembled molecular systems, or nanosystems, are central to life and are increasingly important in nanotechnology<sup>1, 2</sup>. In living organisms, nanosystems have evolved to respond precisely to specific variations in stimuli such as temperature, pressure, light, pH, osmolarity, small molecules, and macromolecules<sup>3-5</sup>. These nanosystems typically self-assemble via the formation of multiple noncovalent interactions, through either intramolecular folding or the intermolecular association of two or more molecular components<sup>6, 7</sup>. Of all the proteins characterized so far from living organisms, more than 55% of these are multimeric and likely exploit molecular assembly to provide functional benefits<sup>7, 8</sup>. The tetrameric protein haemoglobin represents a good example of functional adaptation through multimerization. While the monomeric myoglobin is better suited for oxygen storage due to its higher affinity, the tetrameric haemoglobin transports four O<sub>2</sub> molecules using a highly cooperative load-and-release mechanism that can be allosterically controlled through pH and other molecular effectors<sup>9, 10</sup>. Inspired by such sophisticated nanosystems, chemists and engineers aspire to develop similar self-regulated systems for various nanotechnological applications, including biosensing, drug delivery and chemical computing<sup>11-13</sup>. However, the advantages or limitations of building such self-regulated molecular systems using either one or multiple molecular components is often overlooked by biologists, engineers, and chemists.

Potential advantages and limitations of building self-regulated molecular systems using multiple molecular components have been proposed. These hypotheses, however, remain quite challenging to test by comparing natural monomeric and multimeric proteins given their differentiation throughout evolution<sup>10, 14, 15</sup>. An obvious advantage of unimolecular systems is their fast assembly due to typically fast intramolecular folding rates<sup>16</sup>. However, large unimolecular systems increase the complexity inherent to their covalent synthesis (multistep, error rate, low yield, etc.) and are more likely to misfold, thereby reducing their activity and even leading to harmful misfolded intermediates<sup>17</sup>. Proteins made from longer polypeptide chains, for example, typically require the assistance of complex molecular chaperones to limit misfolding during assembly<sup>17</sup>. This problem can be avoided, in principle, by reducing the sizes and complexities of these large individual components by employing several smaller self-assembling

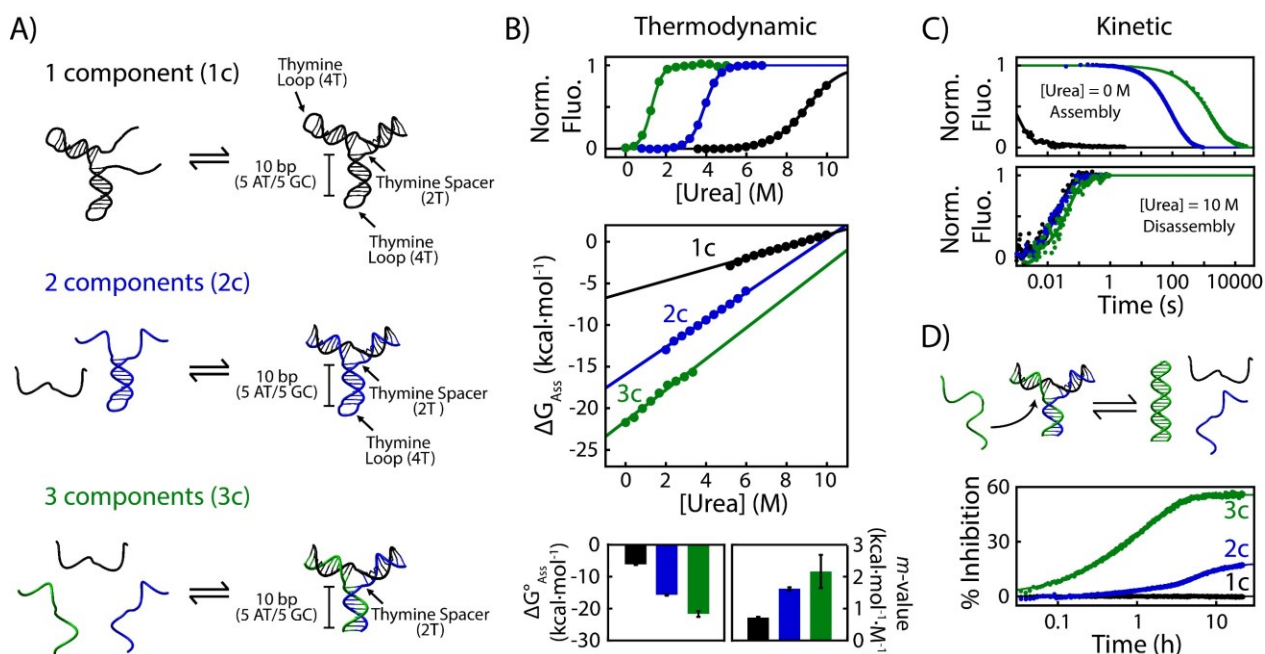
components<sup>18</sup>. Having more molecular components could also increase the regulatory potential of assembly by rendering its function dependent on the concentration of its components<sup>18-21</sup>, but can also lead to unwanted kinetic traps, thus slowing down its assembly<sup>21</sup>. These challenges are also typically observed when assembling artificial nanosystems made of DNA, RNA, and proteins, which require long and optimized assembly protocols with often complex temperature annealing or dialysis strategies<sup>22-24</sup>.

One strategy to understand the functional advantages of building nanosystems using one, or multiple molecular components could consist in exploring the effect of fragmenting a functional unimolecular structure into multiple components. In the last decades, researchers have found that natural proteins and artificial nucleic acid-based systems can readily support fragmentation in two or multiple components (*e.g.*, protein-fragment complementary assay and split aptamers)<sup>25-28</sup>. However, to date, it remains unclear whether fragmenting a self-assembled system into more components may allow to artificially evolve and, therefore, improve its functionality. Here, we explored this hypothesis by studying the thermodynamic and kinetic impact of fragmenting a self-assembled nanosystem into multiple components that reassemble into the same exact nanostructure. Using theoretical simulations and experiments on two nucleic acid model systems, we demonstrated how self-assembled nanosystems can be readily programmed to acquire complex regulation mechanisms through a simple fragmentation strategy.

## 2.3 Results

To explore whether fragmenting a self-assembled nanosystem into more components could allow to program novel regulatory functionalities, we designed a simple DNA three-way junction system that contains programmable interacting surfaces (Fig. 2.1A). We utilized DNA to recreate a multicomponent biological assembly instead of proteins due to the high predictability and programmability of its interactions (affinity and kinetic)<sup>29, 30</sup>, its ease of synthesis<sup>31</sup>, its chemical modifications (*e.g.*, the addition of fluorophores and quenchers)<sup>32</sup>, and its ability to create specific, well-defined and highly soluble structures<sup>22</sup>. The three-way junction, for example, is a widely occurring nucleic acid motif in all kingdoms of life (*e.g.*, rRNA, riboswitches, artificial

ribozymes, and DNAzymes) and an important building block in DNA nanotechnology<sup>33-36</sup>. This motif contains three arms, each of which are separated by a short 2-thymine spacer. Each arm is made from 10 base pairs (5 AT/5 GC) and has a predicted free energy of -10 kcal/mol ( $K_D \sim 100$  nM at 37°C)<sup>37</sup>, which enables assembly within a range of concentrations ideally suited for fluorescence measurements.



**Figure 2.1 – Thermodynamic and kinetic characterization of a nanosystem built with one (black, 1c), two (blue, 2c), or three (green, 3c) molecular components.** A) A simple DNA-based self-assembled “three-way junction” nanosystem containing three arms of 10 base pairs (5 AT/5 GC). The 2c and 3c systems were obtained through fragmentation by removing either one (2c), or two (3c) thymine loops. All assemblies were monitored using fluorescently labelled DNA strands (yellow circle represents the fluorophore moiety FAM and black circle represents the quencher moiety BHQ-1), and the data were normalized accordingly (see Methods). B) Thermodynamic analysis of nanosystem assembly using urea equilibrium denaturation curves (see also Fig. S2.5 for the thermal denaturation curves). The free energy of assembly ( $\Delta G^{\circ}_{Ass}$ ) and the  $m$ -values, which reflect the amount of surface area exposed to solvent upon disassembly, increase with the number of components (see Fig. S2.2-S2.3 for raw data and Table S2.1 for the values). The lower  $m$ -values and energies of assembly of the 1c and 2c systems are attributable to preorganized structures (hairpins) that remain in the disassemble state (see panel A and Fig. S2.4). C) Increasing the number of components reduces the rate of assembly (*top*) but does not affect the rate of disassembly at 10 M urea (*bottom*). Of note, the assembly parameters ( $\Delta G^{\circ}_{Ass}$  and  $m$ -values) extracted from urea kinetic traces (Fig. S2.6) are within the experimental error of the values determined by equilibrium experiments (see Table S2.1). D) Increasing the number of components reduces the presence of preorganized structures, which facilitates the creation of regulation mechanism via competitive inhibitors.

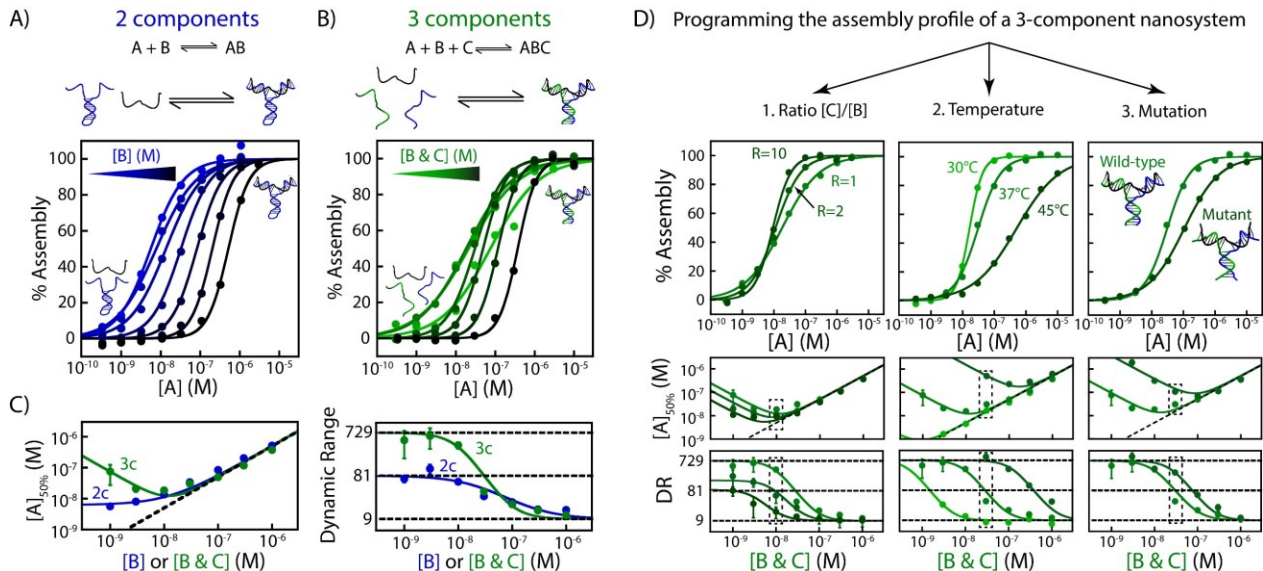
We first characterized the assembly thermodynamics of three-way junctions built with one, two or three components by determining the difference in energy between their assembled and disassembled conformations,  $\Delta G_{\text{Ass}}$ , using chemical and thermal denaturation procedures. Urea denaturation curves typically provide two important parameters: 1) an estimation of  $\Delta G_{\text{Ass}}$  in the absence of urea ( $\Delta G^{\circ}_{\text{Ass}}$ ) and 2) the  $m$ -value, which correlates with the amount of surface area exposed to solvent upon disassembly (see Fig. 2.1B, *top*)<sup>38, 39</sup>. The one-component system (1c) displays a free energy of assembly of  $-6.4 \pm 0.4$  kcal/mol, which remains significantly lower than the predicted free energy of assembly of the complete system ( $-22.9$  kcal/mol, Fig. S2.1). This suggests that part of the disassembled state of the 1c system may remain assembled even at high urea concentrations. To verify this hypothesis, we characterized the isolated hairpins and found that they indeed remained assembled even at urea concentrations above 10 M (Fig. S2.2-S2.3). The small  $\Delta G^{\circ}_{\text{Ass}}$  measured for the 1c nanosystem is indeed consistent with the value predicted ( $-5.7$  kcal/mol) for a simple closure of the three-way junction with preorganized hairpins (Fig. 2.1A, Fig. S2.1). We then cut the 1c system into two components by removing one of the hairpin loops (2c). Its  $\Delta G^{\circ}_{\text{Ass}}$  ( $-15.8 \pm 0.2$  kcal/mol) and  $m$ -value ( $1.61 \pm 0.05$  kcal/mol·M) were found to be approximately twice the amount measured for the 1c system (Fig. 2.1B, *blue curve*, Table S2.1). This indicates that the 2c system forms approximately two times more interactions while burying two times more surface than the 1c system during its assembly (Fig. S2.4). A similar trend is also observed when fragmenting the nanosystem in three components (3c) with its assembly transition involving nearly three times more interactions ( $\Delta G^{\circ}_{\text{Ass}} = -21.8 \pm 0.9$  kcal/mol·M) and burying three times more surface ( $m$ -value =  $2.1 \pm 0.5$  kcal/mol·M) than the 1c system (Fig. 2.1B, *green curves*). Thermal denaturation experiments (Fig. S2.5) also support these observations and further reveal that the smaller  $\Delta G^{\circ}_{\text{Ass}}$  of the 1c system is attributable to reductions in both enthalpy ( $54 \pm 9\%$ ) and entropy ( $49 \pm 10\%$ ) compared to those of the 3c system (see Table S2.1). A structural analysis of these systems using the NUPACK software<sup>40</sup> yielded results that were consistent with these experimentally measured free energies and this hypothesis (Fig. S2.1).

We then determined the assembly kinetics of these nanosystems and found, as expected, that their rates of assembly decreased drastically when the number of components was increased

(Fig. 2.1C, *top*). For example, while 50% of the 1c system folds within 1 ms in absence of urea ( $t_{1/2} = 0.5 \pm 0.1$  ms), the 2c ( $t_{1/2} = 32 \pm 9$  s) and 3c ( $t_{1/2} = 1287 \pm 378$  s) systems assemble five and seven orders of magnitude more slowly at similar concentration, respectively (Fig. 2.1C, *top*). This is consistent with the fact that unimolecular reactions are typically faster than multimolecular reactions and that many interactions are already formed in the 1c and 2c nanosystems (*i.e.*, preorganized hairpins) before the observed assembly transition takes place. We also compared the disassembly rate of all the three-way junctions at 10 M urea and found that they all disassemble at a similar rate ( $t_{1/2} = 24 \pm 10$  ms, Fig. 2.1C, *bottom*). This suggests that fragmentation did not significantly impact the disassembly mechanism and half-life of each nanosystem. As noted above, however, the disassembly of the 1c and 2c systems remains only partial since their hairpins remain organized above 10 M urea (see Fig. S2.6 for more details). The  $\Delta G^{\circ}_{\text{Ass}}$  and  $m$ -values estimated from the assembly and disassembly kinetics further support the presence of preorganized structure given that they are within the experimental error of the values determined by equilibrium experiments (see Table S2.1 and Fig. S2.6-S2.9)<sup>41</sup>.

Taken together, these thermodynamic and kinetics results suggest that fragmenting a nanosystem into multiple components may help reduce the fraction of preorganized structures in the individual components, thus ensuring that the assembly transition maximizes the number of newly formed interactions. More specifically, our results demonstrate that the 1c DNA three-way junction assembles rapidly in a sequential manner via the preorganization of two hairpins followed by a closure of the last arm while the slower assembly of the 3c system displays a more global cooperative assembly, with all its interactions being formed simultaneously in a one-step process. Besides the variation in assembly rate, we further explored whether the presence of preorganized structures in the disassembly state of a nanosystem could impact the regulation of its assembly. Since preorganized structures reduce the availability of some interactions, we hypothesized that they could also prevent, or enable, the interaction or binding with other external molecules. To highlight this feature, we designed a complementary DNA strand that binds specifically both hairpin domains (Fig. 2.1D). We found, indeed, that the 1c system cannot be inhibited by this classic complementary DNA inhibitor due to the low accessibility of the nucleotides locked in the preorganized hairpins. In contrast, the 2c and 3c systems are

increasingly more sensitive to the inhibitor due to having more surfaces exposed in their disassembled states (higher  $m$ -value), thus displaying a greater potential for the design of regulatory mechanisms.



**Figure 2.2 – The assembly of a nanosystem becomes more programmable (regulable) as its number of components increases.** A) Programming the assembly ( $[A]_{50\%}$  and dynamic range) of the 2c and B) 3c systems by increasing  $[A]$ . Assemblies are performed at different fixed concentrations of strand B (2c) or strands B and C (3c). C) When the concentration of strand B increases, the  $[A]_{50\%}$  of the 2c system also increases from 5.6 nM to 516 nM (left), while its dynamic range decreases from 81- to 9-fold (right). When the concentrations of both strands B and C are similarly increased, the  $[A]_{50\%}$  of the 3c system displays a more complex relationship, changing from 92 nM to 18 nM to 372 nM (left), and its dynamic range decreases from 729- to 9-fold (right). D) The assembly profile of a three-component nanosystem can be further programmed by 1) increasing the concentration ratio between components B and C, 2) changing the temperature, or 3) introducing destabilizing mutations in the duplex arms. The  $[A]_{50\%}$  and dynamic range in the dotted box are derived from the experimental curves represented above (see Fig. S2.13 to Fig. S2.15 for the rest of the data). All dose-response curves are fitted using the Hill equation, while the  $[A]_{50\%}$  and dynamic range are fitted using equations derived from the 2c and 3c models (see supplementary information). The experimental conditions were optimized to capture the entire assembly process over the experimentally allowed five orders of magnitude of  $[A]$ . The error bars represent the error of the non-linear fitting.

We then determined the impact of varying the number of nanosystem components on their assembly mechanisms. The assembly mechanism of unimolecular nanosystems, such as the folding of DNA<sup>39</sup> or protein<sup>42</sup> structures, can take place in milliseconds to hours and is generally hard to program since it mostly depends on the complexity of the structure<sup>43</sup>. Some proteins start folding co-translationally as soon as they reach the ribosomal exit tunnel while other unimolecular

systems remain under kinetic control<sup>44-47</sup>. Unimolecular nanosystems, nonetheless, require to fold into their native structure to fulfil their functions<sup>46</sup> and their activity generally varies linearly with their concentration (*i.e.*, higher yield of synthesis leads to higher activity). Typical strategies to regulate these systems generally require the help of external molecules, such as allosteric effectors that will reduce (inhibitor) or enhance (activator) its activity<sup>48</sup>. In contrast, the assembly of the dimeric 2c system can be regulated by tuning the concentration of one of its components (herein called A), resulting in a classic dose-response behaviour (see Fig. 2.2A) with two programmable parameters: 1) the midpoint or  $[A]_{50\%}$ , *i.e.*, the concentration at which 50% of the system is assembled, and 2) the cooperativity of the response or dynamic range (DR), *i.e.*, the broadness of the transition, defined as the change in [A] required to provide a change in response from 10% to 90% ( $DR = [A]_{90\%}/[A]_{10\%}$ ). For example, when the 2c system assembles in the presence of a low concentration of one of its components (*e.g.*,  $[B] = 1 \text{ nM} < K_D^{AB}$ , the dissociation constant of AB, see Fig. 2.2A, *lightest blue curve*), its  $[A]_{50\%}$  remains constant and equal to the dissociation constant of the system ( $[A]_{50\%} = K_D^{AB} = 5.6 \pm 0.5 \text{ nM}$ ). In such cases, the dynamic range of the assembly remains at approximately 81-fold ( $DR = 69 \pm 12$ -fold). On the other hand, in a saturation regime, *i.e.*, when the concentration of component B exceeds  $K_D$ ,  $[A]_{50\%}$  increases linearly with [B] ( $0.5 \times [B]$ ), indicating a 1:1 binding regime (see Fig. 2.2C, *left, blue curve*)<sup>49</sup>. In such situations, the observed transition becomes more “cooperative”, and the dynamic range is reduced by up to 9-fold (Fig. 2.2C, *right, blue curve*).

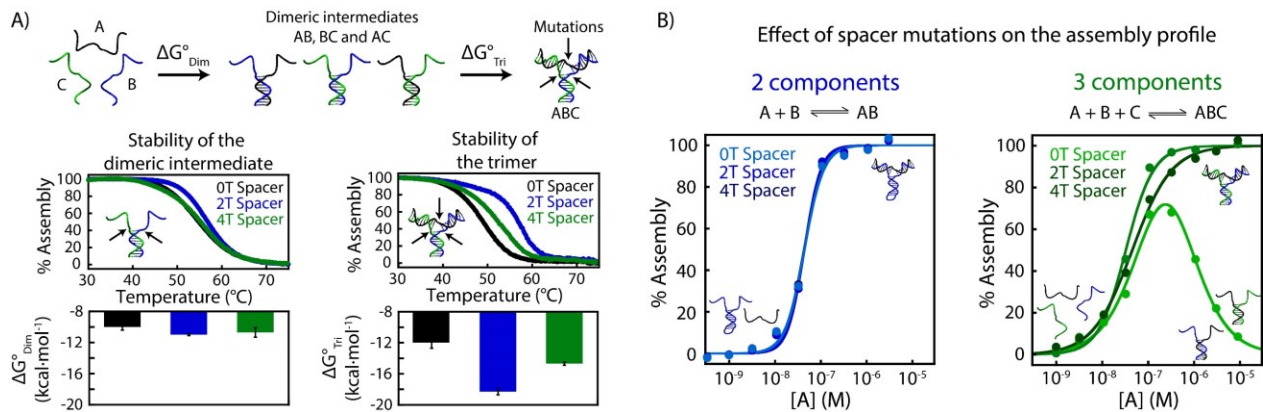
The assembly of the trimeric 3c system displays even more programmability by enabling assembly over a wider range of concentrations. For instance, varying the concentration of components B and C at levels lower or higher than  $K_D^{BC}$  ( $41 \pm 23 \text{ nM}$ , see Fig. S2.10) allows for tuning of their preorganization level (*i.e.*, [BC]). For example, when the B and C concentrations are higher than  $K_D^{BC}$ , the dimer preforms in a state similar to that of the 2c system, and the 3c system consequently behaves similarly to the 2c system (*i.e.*, a cooperative dynamic range of 9-fold with a  $[A]_{50\%}$  equals  $0.5 \times [B]$ ) (Fig. 2.2A-B, *darker curves*, and Fig. 2.2C). In contrast, when using B and C at concentrations lower than  $K_D^{BC}$ , the system is not preorganized (*i.e.*, B and C remain dissociated), and the assembly displays a much higher  $[A]_{50\%}$  and a broader dynamic range since the assembly proceeds directly from monomers to trimer without forming dimeric

intermediates (see Fig. S2.11). For example, at 1 nM of B and 1 nM of C (Fig. 2.2B, *lightest green curve*), the 3c systems display a  $[A]_{50\%}$  of  $76 \pm 48$  nM (vs  $5.6 \pm 0.5$  nM for 2c) with a broad anticooperative dynamic range near 729-fold (vs 81-fold for 2c) (Fig. 2.2C, *green curves*). These assembly behaviours of the 2c and 3c systems are also well modelled by numerical simulations (see Fig. S2.12).

Including more components in the assembly process helps optimize the assembly profile and therefore conveniently tunes the  $[A]_{50\%}$  and the dynamic range. For example, simply increasing the ratio between components B and C ( $R = [C]/[B]$ ) shifts the assembly profile from anticooperative behaviour to a typical 81-fold dynamic range (*e.g.*, from  $DR = 383 \pm 62$  to  $DR = 57 \pm 25$ ) without affecting  $[A]_{50\%}$  (Fig. 2.2D, *left*). This provides a useful strategy to specifically program either  $[A]_{50\%}$  or the dynamic range independently. Alternatively, the assembly profile of a 3c system can be optimized by tuning the component B and C preorganization levels (*i.e.*,  $[BC]$ ), which can be readily achieved by changing the temperature (Fig. 2.2D, *middle*) or via specific mutations (Fig. 2.2D, *right*). For example, increasing the temperature to 45°C destabilizes the preorganized BC dimer by one order of magnitude ( $K_D^{BC}$  increases from  $41 \pm 23$  nM to  $547 \pm 254$  nM, see Fig. S2.10), thereby shifting  $[A]_{50\%}$  and the dynamic range profiles to the same extent. Decreasing the temperature to 30°C has the opposite effect ( $K_D^{BC}$  from  $41 \pm 23$  nM to  $3.8 \pm 2.4$  nM, see Fig. S2.10). Mutations can also be used to achieve a similar level of programmability. For example, when destabilizing the duplex hairpin by inserting one mismatch at each extremity ( $K_D^{BC}$  from  $41 \pm 23$  nM to  $3.3 \pm 2.3$   $\mu$ M, see Fig. S2.10), the  $[A]_{50\%}$  and the dynamic range profiles also shift to higher concentrations. Of note, the high preorganization of components B and C can help mitigate the effects of perturbing mutations. For example, when using B and C at 1000 nM, the wild-type, and the mutant show similar  $[A]_{50\%}$  values ( $372 \pm 22$  vs  $349 \pm 42$  nM) and dynamic ranges ( $10 \pm 1$  vs  $10 \pm 2$ ). We also used our numerical simulations to get insights into a system that can be potentially divided into more than three molecular components. For example, when having four molecular components, the  $[A]_{50\%}$  of the nanosystem follows a steeper bell-shaped trend than the three-component system while the dynamic range shifts from 9, at high concentration of components, to a much broader dynamic range of 6561 at low concentration of components (see Fig. S2.16 for more details). Overall, these results demonstrate that the process



of assembling nanostructures made from more than two components can be programmed to have substantially more diverse regulatory profiles than structurally similar one-, or two-component systems.



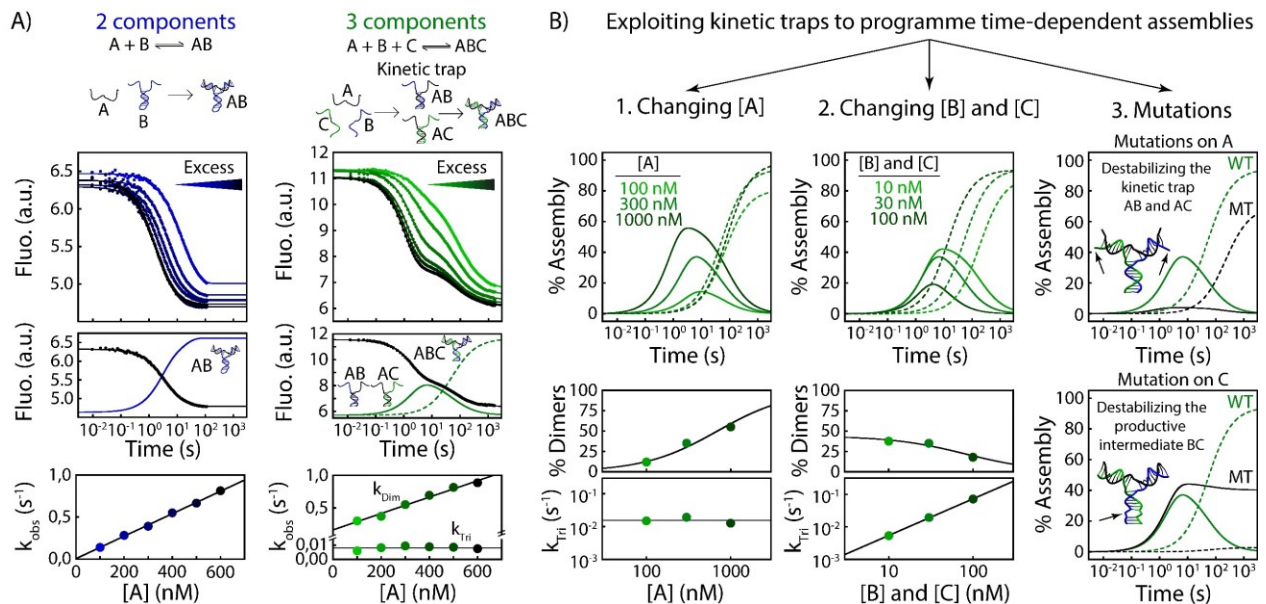
**Figure 2.3 – Specific mutations on trimeric assemblies are more impactful than on their dimeric counterparts and can lead to novel regulatory mechanisms.** A) *Top*. The assembly of a trimeric nanosystem is achieved via dimeric intermediates. To program the assembly profile of a trimeric nanosystem, we introduced insertion/deletion mutations in all spacers connecting the duplex arms using either 0, 2, or 4 thymines (0T, 2T, 4T). *Bottom*. Thermal denaturations confirm that these mutations specifically affect the trimeric stability ( $\Delta G^{\circ}_{\text{Tri}}$ ) without destabilizing the dimeric intermediates ( $\Delta G^{\circ}_{\text{Dim}}$ ). B) *Left*. The assembly of the 2c system is insensitive to spacer length mutations. *Right*. In contrast, the same mutations in the 3c system result in new assembly profiles. For example, a low  $\Delta G^{\circ}_{\text{Tri}}$  (0T) creates a self-inhibited trimeric system, which disassembles into two dimers (AB and AC) at a higher [A]. The assemblies were performed using 100 nM of component B and C. The inhibition of trimeric assembly into dimers was confirmed by polyacrylamide gel electrophoresis (see Fig. S2.19).

Specific mutations can have an even more drastic impact on the assembly profile of a three-component system. While the assembly of a two-component system consists of the simple dimerization of two monomers, that of a three-component system involves the assembly of multiple dimeric intermediates before the formation of the trimeric structure (Fig. 2.3A, *top*). We next explored whether mutations that specifically affect the stability of the trimeric three-way junction ( $\Delta G^{\circ}_{\text{Tri}}$ ), without affecting the stability of the dimeric intermediates ( $\Delta G^{\circ}_{\text{Dim}}$ ), could provide a novel avenue for diversifying functionality. We introduced simple insertion/deletion mutations at the junction between the duplex arms by either decreasing the spacer length from two thymines (2T, wild-type) to none (0T, deletion) or by increasing it to four thymines (4T, insertion)<sup>50,51</sup>. Using thermal denaturation (Fig. 2.3A, *top* and Fig. S2.17), we confirmed that these mutations do not affect the stability of the dimeric intermediates ( $\Delta G^{\circ}_{\text{Dim}}$  of approximately -10.6

$\pm 0.4$  kcal/mol) but decrease the trimeric stability by  $3.6 \pm 0.6$  kcal/mol (4T) and  $6.3 \pm 1.1$  kcal/mol (0T). These mutations drastically impact the assembly profile of the 3c system despite having no effect on the assembly of the structurally similar 2c system (Fig. 2.3B, *left*). While the most stable 2T trimeric system assembles in a highly cooperative manner ( $DR = 21 \pm 5$ ), the mildly destabilized 4T system displays less cooperative assembly ( $DR = 73 \pm 18$ ), whereas the most destabilized 0T system assembles via a novel self-inhibited profile (“none-all-none” mechanism)<sup>52</sup> (Fig. 2.3B, *right*). In this self-inhibitory profile, the percentage of assembled trimer increases from 10% to 70% when [A] is changed from 5 nM to 200 nM and then decreases to 10% when [A] is further increased to 10  $\mu$ M ( $DR = 2000$ -fold, see also how to program this self-inhibitory dynamic range in Fig. S2.18). This mechanism occurs when the stability of the trimeric system is similar to that of the dimeric intermediate (*e.g.*, for the 0T system,  $\Delta G^\circ_{\text{Tri}} - \Delta G^\circ_{\text{Dim}} = -2.0 \pm 1.1$  kcal/mol). In such cases, the formation of dimeric intermediates becomes thermodynamically favoured at higher concentrations of [A] (see Fig. S2.19 for proof of the mechanism). In contrast, the same mutations have no effect on the assembly profile of the 2c system because they do not sufficiently destabilize the 2c system and its  $K_D$  to affect its assembly profile. Taken together, these results demonstrate that in addition to permitting novel regulation mechanisms, trimeric assemblies are also much more sensitive to small perturbations or mutations than their dimeric counterparts.

While the regulatory mechanisms discussed above all take place at equilibrium, the assembly of nanosystems with multiple components can also be controlled by kinetics<sup>21, 53, 54</sup>. As mentioned previously, unimolecular systems (1c) typically fold rapidly unless misfolding occurs<sup>16</sup>. Increasing the number of components to two can easily program the assembly rate by exploiting the law of mass action. For example, increasing the concentration of A from 100 to 600 nM with a small concentration of B (10 nM) increases the observed assembly rate ( $k_{\text{obs}}$ ) by 6-fold (Fig. 2.4A, *left*). For the 3c system, however, mixing an excess of component A (100 to 600 nM) with smaller concentrations of B and C (10 nM) traps the latter into the nonfunctional dimeric intermediates AB and AC (Fig. 2.4A, *right*). Dissociation of these dimers is then required to enable trimer formation by the slow association of the formerly sequestered components B and C, thus resulting in biphasic kinetics. We confirmed that the fastest phase,  $k_{\text{Dim}}$ , represents the formation of dimers AB and AC given their linear dependency on the concentration of the monomer A. In contrast, the

slowest phase,  $k_{Tri}$ , represents the formation of the trimer and is rate-limited by the formation of the productive dimer BC, thus explaining its insensitivity to the concentration of component A. This kinetically controlled mechanism of assembly provides interesting time-dependent formation/dissociation profiles for both the dimeric intermediates and the final trimeric assembly.



**Figure 2.4 – Programming the assembly of trimeric nanosystems over time by exploiting dimeric kinetic traps.** A) *Two components*: At a low concentration of monomer B (10 nM), the addition of an excess of monomer A (100 to 600 nM) results in increasingly faster single-phase kinetics. *Three components*: At low concentrations of monomers B and C (10 nM, mainly unbound), the addition of an excess of monomer A (100 to 600 nM) results in slower biphasic kinetics due to the sequestration of monomers B and C into the dimeric intermediates AB and AC (green solid line). Trimer formation (green dashed line) then proceeds through a slow component-exchange mechanism limited by the availability of free components B and C. B) *Left*. Tuning the concentration of A can program the rate of formation (activation) and the percentage of formed dimer without affecting the rate of trimeric assembly ([B] = [C] = 30 nM). *Middle*. Increasing the concentrations of [B] and [C] and thereby increasing their level of preorganization can increase the rate of trimer assembly and decrease the percentage of transiently formed dimers ([A] = 300 nM). *Right*: Mutations specifically destabilizing the dimeric intermediates AB and AC (*top*) or the productive BC intermediate (*bottom*) prevent dimeric intermediate and trimer formation, respectively ([A] = 300 nM and [B] = [C] = 30 nM). See Fig. S2.23 to Fig. S2.25 for panel B raw data.

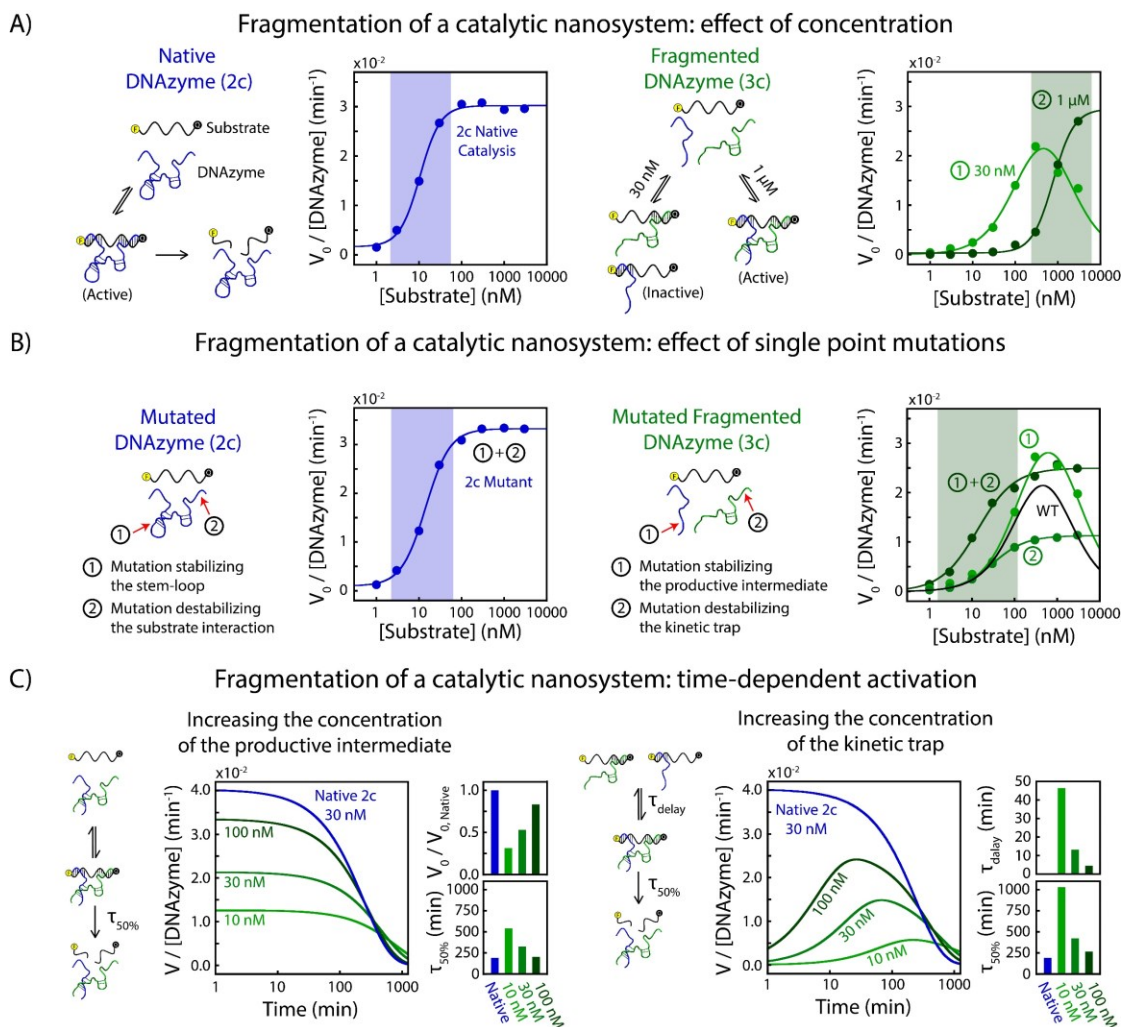
The kinetic analysis (Fig. 2.4A) and numerical simulations (Fig. S2.20 to Fig. S2.22) of trimeric assembly through the formation of dimeric kinetic traps provide strategies for precisely programming the assembly rate of nanosystems. For instance, simply increasing [A] can increase

the rate of formation (activation) and the percentage of the transient intermediates AB and AC without affecting the trimer assembly rate (Fig. 2.4B, *left*). In contrast, increasing the concentrations of [B] and [C] and therefore their level of preorganization can increase the trimer assembly rate (Fig. 2.4B, *middle*). Notably, increasing the rate of trimer formation also decreases the percentages of the transiently formed dimeric intermediates AB and AC, and specific mutations have similar effects (Fig. 2.4B, *right* and Fig. S2.25). For example, a mutation specifically destabilizing the transient intermediates AB and AC prevents the formation of these dimers without substantially affecting the rate of trimer formation, and  $k_{\text{Tri}}$  is reduced by only 5-fold (Fig. 2.4B, *right, top*). On the other hand, a mutation specifically destabilizing the productive BC intermediate traps the nanosystem into the AB and AC dimers and prevents trimer formation (Fig. 2.4B, *right, bottom*). Similar trends and regulatory opportunities were also observed for nanosystems built with more than three molecular components. Numerical simulations on a nanosystem built with four components, for example, reveal more assembly pathways leading to a higher variety of programmable time-dependant activation/deactivation profiles (see Fig. S2.26 for more details). Having more components, however, can also increase the opportunities to form unwanted and harmful intermediate assemblies. All these results (experiments and simulations) exemplify that building a nanosystem with multiple molecular components can generate programmable time-dependent nanosystems by exploiting the law of mass action and simple kinetic traps. Such time-dependent nanosystems are present in various biochemical processes, such as signal transduction, protein synthesis, and the cell cycle, and will likely have promising applications in future self-regulated nanotechnologies<sup>55-57</sup>.

In this study, our three-way junction served as a convenient synthetic toolkit to quantitatively evaluate the impact of fragmentation and the number of components on the functionality of a nanosystem, and the extent to which this impact can be harnessed to create novel regulatory mechanisms. To test the generality and predictability of these findings, we used these rules to program the catalytic activity of NaA43, a sodium-specific RNA-cleaving DNAzyme previously used as a sensor to monitor the sodium concentration inside cells<sup>58</sup>. We first measured the apparent activity of NaA43 and revealed an apparent  $K_M$  of  $11 \pm 1$  nM, corresponding to approximately half of the DNAzyme concentration used in our assay (30 nM), and a dynamic range

of  $11 \pm 4$  (Fig. 2.5A, *left*). We also estimated  $K_D^{\text{DNAzyme}}$ , the dissociation constant between the substrate and the DNAzyme, to be in the fM range (Fig. S2.27), which confirms why this DNAzyme/substrate system operates in a saturation regime ( $[\text{DNAzyme}] > K_D^{\text{DNAzyme}}$ , see also Fig. 2.5A, *left*). We then artificially evolved this catalytic nanosystem by splitting the DNAzyme in two halves at the extremity of its loop (Fig. 2.5A, *right*). This modification still yields a functional DNAzyme despite a  $36 \pm 2\%$  reduction in catalytic activity (Fig. S2.28). We first explored the effect of DNAzyme concentration on the activity of this fragmented DNAzyme. At a high concentration of DNAzyme components ( $1 \mu\text{M}$ ), which ensures high preorganization,  $[A]_{50\%}$  remains near half the concentration of the DNAzyme ( $778 \pm 53 \text{ nM}$ ) with a “cooperative” dynamic range ( $\text{DR} = 12 \pm 3$ ). In contrast, as predicted by our model (Fig. 2.4B), using a low concentration of DNAzyme components (*i.e.*,  $30 \text{ nM}$ ) produces a kinetic trap, which can be used to program a substrate inhibition regulatory mechanism (*i.e.*, “none-all-none” regulation), where both DNAzyme components become sequestered into nonfunctional dimers at high substrate concentrations (see Fig. S2.29).

The function of this fragmented catalytic system can be further optimized using single point mutations. For example, when implementing a stabilizing mutation in the loop of the fragmented DNAzyme that increase its preorganization level ( $A \rightarrow C$ , blue strand Fig. 2.5B), we increased the activity level by  $25 \pm 1\%$  while keeping the self-inhibition mechanism. In contrast, when implementing a mutation that destabilizes the kinetic trap ( $G \rightarrow C$ , green strand Fig. 2.5B), we disrupt the self-inhibition mechanism at the cost of a reduced activity ( $53 \pm 2\%$  slower). When combining these two mutations we obtained a nanosystem that displays both high activity and no self-inhibition mechanism. Interestingly, these two point mutations alone were able to shift the  $K_M$  value of the trimeric system back to the value of the original uncut DNAzyme ( $14 \pm 2 \text{ nM}$  versus  $11 \pm 1 \text{ nM}$ ) while displaying a broader dynamic range ( $78 \pm 33$  versus  $12 \pm 3$ ). Interestingly, none of these mutations impacted the catalytic activity of the native 2c system (Fig. 2.5B, *left*). These experiments illustrate how fragmented nanosystems are more affected by single point mutations or chemical modifications and, therefore, are more programmable compared to their unfragmented counterparts.



**Figure 2.5 – Programming novel regulatory mechanisms in a complex catalytic nanosystem through fragmentation.** A) *Left.* The NaA43 cleaving DNAzyme<sup>58</sup> displays a  $[A]_{50\%}$  of  $11 \pm 1$  nM at 30 nM DNAzyme and a dynamic range of  $11 \pm 1$ , consistent with a saturation regime. *Right.* Creating a trimeric assembly by fragmenting the DNAzyme can create a kinetic trap at a low concentration of DNAzyme (30 nM), leading to high sequestration of dimer intermediates and lower activity. However, preorganization of the productive dimer using a higher concentration of its component (1  $\mu$ M) results in the system displaying a normal saturation regime profile ( $[A]_{50\%} = 778 \pm 53$  nM and DR =  $12 \pm 3$ ). B) Single point mutations that do not affect the catalytic activity of the native 2c system (*left*) display a drastic impact on the catalytic activity of the 3c system (*right*). A mutation stabilizing the productive dimer (mutation 1) enables to increase the catalytic activity while a mutation destabilizing the dimeric trap (mutation 2) enables to get rid of the self-inhibition at the cost of a reduced catalytic activity. C) Programming the catalytic activity of this trimeric nanosystem over time by exploiting dimeric kinetic traps. *Left.* Increasing the concentration of the productive intermediate of the 3c system enables the programming of the initial rate of the reaction. *Right.* Increasing the concentration of the kinetic trap components enables to delay the activation of the catalytic function. The catalytic rates were obtained using the derivatives of the fitted kinetic traces of product generation. See Fig. S2.30 for panel A and B raw data and Fig. S2.31 for panel C raw data.

Fragmentation of this catalytic system into a three-component nanosystem also enables us to program its activity over time by either increasing the amount of productive dimer or by favouring the kinetic trap. The activity of the fragmented DNAzyme, for example, can be easily tuned by increasing the concentration of its components (*i.e.*, favouring the productive dimer, Fig. 2.5C, *left*). At a high level of the productive dimer, this catalytic 3c system shows a fast kinetic profile that behaves similarly to the 2c system. Its activity can be reduced by decreasing its preorganization level using a lower concentration of DNAzyme components (*light green*). The catalytic activity of the nanosystem can also be delayed by increasing the concentration of the dimeric trap (Fig. 2.5C, *right*). By sequestering both DNAzyme components in a complex with the substrate (*i.e.*, each DNAzyme fragment is first put in contact with the substrate), we can delay the catalytic activity by up to 46 minutes. Such kinetically programmed functional assemblies are also often employed in biological assembly. For example, compartmentalization can be used to organize the stepwise interaction of components to avoid misassembly<sup>59</sup> while kinetic traps are employed to delay the recruitment of protein units to control specific functions over time<sup>60</sup>. It is also interesting to note that a similar kinetic profile and programmability level can be achieved by increasing the number of components through the addition of an extra DNA strand<sup>61</sup> (Fig. S2.32 and Fig. S2.33). Although designing an extra interacting component for DNA assemblies remains relatively easy to realize, this strategy, however, remains much more challenging to perform with proteins compared to fragmentation, especially via random natural evolution mechanisms. Overall, these results exemplify the simplicity and efficiency of the fragmentation strategy to program and optimize the function of complex nanosystems by increasing the number of components involved in their assembly.

## 2.4 Discussion et conclusions

Here, we demonstrate how the functionalities of nanosystems can simply be optimized by dividing their structures into multiple components. We first demonstrate that despite its slower assembly rate, our model three-component system undergoes a more global and cooperative assembly transition with fewer preorganized structures, which facilitate the creation of novel functionalities. For example, smaller, less preorganized components display more potential to interact with other molecular effectors, demonstrating novel avenues for regulating their

assembly (*e.g.*, complementary DNA inhibitor, Fig. 2.1D). Another advantage of systems with more components, like the three-component nanosystem, is their ability to be programmed to permit assembly using both a “cooperative” and an anticooperative process (Fig. 2.2). We also showed that three-component system, in contrast to a two-component system, are more sensitive to mutations and can be readily tuned to exhibit self-inhibition mechanisms (Fig. 2.3) as well as time-dependent activation/deactivation mechanisms (Fig. 2.4). All these complex regulatory profiles (at equilibrium or over time) are readily accessible by simply fragmenting an existing self-assembled nanosystem into multiple components. We further demonstrated and validated the simplicity and usefulness of this strategy by artificially evolving the functionality of a more complex catalytic nanosystem that possess an RNA-cleaving activity (Fig. 2.5).

The fragmentation strategy can be easily implemented using nucleic acid-based nanosystems and could also be employed to create novel functionalities in proteins given their similar self-assembled characteristics<sup>39, 62</sup>. Experts in DNA/RNA nanotechnology, for example, have exploited split aptamers to improve the properties of DNA/RNA-based sensors and therapeutics<sup>26, 63</sup>. Similarly, proteins can also sustain fragmentation into multiple molecular components by folding back into their native conformation<sup>27, 28, 64-67</sup>. Indeed, multiple protein-fragment complementation assays (PCA) have been developed using such artificial dimeric proteins, and even trimeric protein,<sup>66</sup> to study and control biological systems<sup>28</sup>. Additionally, computational design strategies have been recently developed to identify the cutting site in proteins and to optimize the reconstitution of various fragments through specific mutations<sup>27, 67, 68</sup>. Although these studies demonstrate the flexibility of self-assembled protein nanosystems to sustain fragmentation, no studies, to the best of our knowledge, have highlighted, or even suggested that a fragmented nanosystem displays more ability to acquire novel regulation mechanisms. We believe that this fragmentation strategy, which simply consists in cutting an already existing nanosystem into multiple fragments, is also conceptually easier to performed than designing, from scratch, additional components that can interact and regulate the function of a nanosystem. Given the apparent simplicity of fragmenting a nanosystem, we anticipate that this strategy may represent an efficient alternative over engineering allosteric regulatory mechanisms to program and create novel regulatory functions in self-assembled systems<sup>69</sup>.



Engineering complex self-regulatory mechanisms using multiple components provides a programmable and quantitative chemical strategy to develop and optimize the function of any self-assembled nanosystems with applications ranging from biosensing to chemical computing and drug delivery. For example, current strategies to narrow the dynamic range of sensors typically employ allosteric mechanisms<sup>70, 71</sup>, while strategies to extend the dynamic range combine two or multiple sensors with different affinities<sup>72-76</sup>. In contrast, here we illustrated how a DNA sensing nanosystem can be programmed with a narrow (9-fold) or an extended (729-fold) dynamic range by simply building it with three components. This ability to tune the dynamic range is also useful for programming and optimizing the responses of molecular logic gates. For example, a narrow dynamic range creates a more efficient all-or-none response, while a self-inhibited nanosystem (Fig. 2.3B, *right*) produces a “bandpass” filter (*i.e.*, none-all-none response), a regulatory mechanism observed in many cellular functions<sup>77</sup>. A three-component system may also help maintain drug concentrations within a specific therapeutic window. Nature, for example, employs various substrate inhibition strategies to maintain the levels of crucial product metabolites despite large variations in substrate concentration<sup>78, 79</sup>. A simple three-component nanosystem may be able to control and maintain the level of an active drug following its activation through a regulated catalytic system (Fig. 2.5A-B). Finally, a three-component system with programmed kinetic traps enables the time-specific activation and deactivation of various active biomolecules, leading to flexible and custom disease treatment strategies (Fig. 2.5C)<sup>80, 81</sup>.

In addition to providing new strategies to optimize complex self-regulated nanosystems, we believe that the thermodynamic and kinetic principles described herein will contribute to better understand the advantages of natural protein complexes. Indeed, it has been demonstrated that natural evolution has exploited fragmentation strategies to evolve the function of its various nanosystems<sup>82-84</sup>. Our findings now demonstrate that, in addition to being regulated through different independent promoters (*e.g.* an “AND” logic gate)<sup>85</sup>, protein assemblies can additionally be regulated through the creation and optimization of various mechanisms (*e.g.* cooperativity, anti-cooperativity, self-inhibition, and molecular timer). Given that more than 55% of all proteins in living organisms are multimeric<sup>8</sup>, it will be interesting to explore whether the functional gains of these protein complexes have simply emerged from the

advantages derived from being built with more components<sup>86</sup>. Answering this question remains challenging, however, given that multimeric proteins have evolved and diverged over billions of years making their direct comparison quite difficult<sup>8,86,87</sup>. In perspective, the ability of fragmented nanosystems to self-assemble through a myriad a weak intermolecular force and acquire new functionalities remains surprising, if not amazing. This may well represent a unique feature of self-assembled nanosystems since such strategy is unlikely to be useful on macroscale human-designed technologies, where components are typically held together via screws, glue or by welding.

## 2.5 Materials and Methods

### Chemicals.

Anhydrous acetonitrile, activation reagents (0.25 M ETT in ACN), oxidizer (0.1 M I2 in H2O/Pyridine/THF), deblock solution (3% TCA in DCM), Cap A mix (Ac2O/Pyridine/THF) and Cap B mix (16% Melm in THF) were purchased from ChemGenes Corporation, Wilmington. dA-CE phosphoramidite (n-bz), dC-CE phosphoramidite (n-bz), dG-CE phosphoramidite (n-ibu), dT-CE phosphoramidite and 2'-TBDMS rA-CE phosphoramidite (n-bz) were purchased from ChemGenes Corporation, Wilmington. Pre-pack 1000Å CPG synthesis columns of dA(Bz), dC(Bz), dG(DMF), dT and BHQ-1 were purchased from Biosearch Technologies, Novato. Fluorescein T amidite was purchased from Biosearch Technologies, Novato. MicroPure II columns for purification were purchased from Biosearch Technologies, Novato. Ammonium hydroxide solution (ACS, 28.0-30.0%) and acetic acid (ACS, ≥99.7%) were purchased from Sigma Aldrich. Triethylamine was purchased from Fisher Scientific. Trifluoroacetic acid (99%) and triethylamine·3HF were purchased from Alfa Aesar. Ammonium bicarbonate and DMSO were purchased from BioShop. Acrylamide, bis-acrylamide, TEMED and ammonium persulfate were purchased from Bioshop. Milli-Q water was prepared in the lab when needed. All buffer solutions were prepared in the lab.

### Oligonucleotides Synthesis.

DNA/RNA synthesis is performed in the lab using standard phosphoramidite chemistry with the DNA/RNA synthesizer H-6 (K&A Laborgeraete, Schaafheim). Cleavage from the solid

support and deprotection of the bases are done overnight (16h) at 60°C in 28.0-30.0% aqueous ammonia. Purification of DNA strand is done by solid-phase extraction using the P-8 purifier from K&A Laborgeraete, Schaafheim. Purification protocol follows the K&A Laborgeraete's method. For strand containing a ribonucleotide, a second deprotection step is required prior to purification. Solutions are dried and the deprotection of the 2'-TBDMS of the ribonucleotide is done for 1h30 at 60°C in 250 µL of TEA·3HF cocktail (1.5 mL DMSO, 750 µL TEA and 1 mL TEA·3HF). The mixture is cooled down to room temperature then quenched with 1 mL of 1.5 M ammonium bicarbonate. Solutions are passed through a buffer exchange column (NAPTM-10 column Sephadex™ G-25 DNA grade) prior to purification by HPLC (1260 infinity, Agilent Technologies) with a reversed-phase column (XBridge® Oligonucleotide BEH C18 OBDTM Prep column, 130Å, 2.5 µm, 10 x 50 mm). Purification protocol follows Waters' recommendation. After purification, all DNA/RNA strands are dried with a Genevac™ miVac centrifugal concentrator, resuspended in Milli-Q water and quantified by UV-Vis (Cary 60, Agilent Technologies). Absorption coefficients ( $\epsilon$ ) are estimated from the DNA sequence using the IDT OligoAnalyzer tool.

## **Sequence.**

### 1c system

Strand ABC with 2T spacers and 4T loops: 5'- GTA GTT CGA GTT CTT GAC GTA CTT TTG TAC GTC AAG TTG TAC ATC AGC TTT TGC TGA TGT ACT TCT CGA ACT AC -FAM-3'

### 2c system

Strand A with a 2T spacer: 5'- GCT GAT GTA CTT CTC GAA CTA C -(FAM)-3'

Strand BC with 2T spacers: 5'- GTA GTT CGA GTT CTT GAC GTA CTT TTG TAC GTC AAG TTG TAC ATC AGC -(BHQ1)-3'

### 3c system

Strand A with a 0T spacer: 5'- GCT GAT GTA CCT CGA ACT AC -BHQ1-3'

Strand A with a 2T spacer: 5'- GCT GAT GTA CTT CTC GAA CTA C -BHQ1-3'

Strand A with a 4T spacer: 5'- GCT GAT GTA CTT TTC TCG AAC TAC -BHQ1-3'

Strand B with a 0T spacer: 5'- GTA GTT CGA GCT(*FAM*) TGA CGT AC -3'

Strand B with a 2T spacer: 5'- GTA GTT CGA GTT(*FAM*) CTT GAC GTA C -3'

Strand B with a 4T spacer: 5'- GTA GTT CGA GTT(*FAM*) TTC TTG ACG TAC -3'

Strand C with a 0T spacer: 5'- GTA CGT CAA GGT ACA TCA GC -3'

Strand C with a 2T spacer: 5'- GTA CGT CAA GTT GTA CAT CAG C -3'

Strand C with a 4T spacer: 5'- GTA CGT CAA GTT TTG TAC ATC AGC -3'

Strand A with a 2T spacer, mutant #1: 5'- CCT GAT GTA CTT CTC GAA CTA G -*BHQ1*-3'

Strand A with a 2T spacer, mutant #2: 5'- GCT GAA GTA CTT CTC GTA CTA C -*BHQ1*-3'

Strand C with a 2T spacer, mutant #1: 5'- CTA CGT CAA GTT GTA CAT CAG C -3'

Strand C with a 2T spacer, mutant #2: 5'- GTA GGA CAA GTT GTA CAT CAG C -3'

### DNAzyme

Substrate: 5'-*T*(*FAM*)- CTC TAT CTA TrAG GAA GTA CCG CCG C -*BHQ1*-3'

Native DNAzyme (NaA43): 5'- GCG GCG GTA CCA GGT CAA AGG TGG GTG AGG GGA CGC CAA GAG  
TCC CCG CGG TTA GAT AGA G -3'

Mutant DNAzyme (NaA43): 5'- GCC GCG GTA CCA GGT CAA AGG TGG GTG AGG GGA CGC CAA  
GCG TCC CCG CGG TTA GAT AGA G -3'

Cut DNAzyme #1: 5'- GCG GCG GTA CCA GGT CAA AGG TGG GTG AGG GGA CGC C-3'

Cut DNAzyme #1, mutant: 5'- GCC GCG GTA CCA GGT CAA AGG TGG GTG AGG GGA CGC C-3'

Cut DNAzyme #2: 5'- AAG AGT CCC CGC GGT TAG ATA GAG-3'

Cut DNAzyme #2, mutant: 5'- AAG CGT CCC CGC GGT TAG ATA GAG -3'

In bold and underlined are the modifications relative to the native strand. *FAM* corresponds to a fluorescein, *T*(*FAM*) corresponds to a thymidine conjugated with a fluorescein, *BHQ1* corresponds to a black hole quencher 1 and *rA* corresponds to an adenosine.

## Fluorescence experiments.

*Urea Titration curves.* Urea titration curves were performed following a method developed by our lab<sup>38</sup>. We started with a 900  $\mu\text{L}$  solution of the DNA-based system of interest in 10 M urea buffered solution (10 mM  $\text{NaH}_2\text{PO}_4$ , 40 mM NaCl, pH = 7.00). We then sequentially diluted this solution with a buffered solution containing the same concentration of DNA-based system but without urea. Each sample was equilibrated for 2 minutes before recording their fluorescence (Cary Eclipse, Agilent Technologies). Unimolecular system titration ( $A_{\text{Diss}} \rightleftharpoons A_{\text{Ass}}$ ) was performed at a concentration of 10 nM of fluorescent DNA strand and fitted using Eq. 2.1. Bimolecular system titration ( $A + B \rightleftharpoons AB$ ) was performed at a concentration of 10 nM of fluorescent DNA strand and 100 nM of the quencher DNA strand (10-fold excess) and fitted using Eq. 2.2. Trimolecular system titration ( $A + B + C \rightleftharpoons ABC$ ) was performed at a concentration of 10 nM of fluorescent DNA strand and 100 nM of the quencher and unlabelled DNA strands (10-fold excess) and fitted using Eq. 2.3. In each equation,  $F$  is the fluorescence signal,  $F^{\circ}_{\text{Ass}}$  and  $F^{\circ}_{\text{Diss}}$  are the intrinsic fluorescence of the assembled and disassembled states,  $\sigma_{\text{Ass}}$  and  $\sigma_{\text{Diss}}$  are the urea dependency of the fluorescence signal of the assembled and disassembled states,  $[U]$  is the concentration of urea,  $\Delta G^{\circ}_{\text{Ass}}$  is the Gibbs free energy of assembly in absence of urea,  $[D_Q]$  and  $[D_{Un}]$  are the concentration of the quencher and unlabelled DNA strands (which are in 10-time excess),  $T$  is the temperature and  $R$  is the gas constant.

$$F = \frac{(F^{\circ}_{\text{Ass}} + \sigma_{\text{Ass}}[U]) + (F^{\circ}_{\text{Diss}} + \sigma_{\text{Diss}}[U]) \cdot e^{-\frac{(\Delta G^{\circ}_{\text{Ass}} - m[U])}{RT}}}{\left(1 + e^{-\frac{\Delta G^{\circ}_{\text{Ass}} - m[U]}{RT}}\right)} \quad (\text{Eq. 2.1})$$

$$F = \frac{[D_Q] \cdot (F^{\circ}_{\text{Ass}} + \sigma_{\text{Ass}}[U]) + (F^{\circ}_{\text{Diss}} + \sigma_{\text{Diss}}[U]) \cdot e^{-\frac{(\Delta G^{\circ}_{\text{Ass}} - m[U])}{RT}}}{\left([D_Q] + e^{-\frac{\Delta G^{\circ}_{\text{Ass}} - m[U]}{RT}}\right)} \quad (\text{Eq. 2.2})$$

$$F = \frac{[D_Q] \cdot [D_{Un}] \cdot (F^{\circ}_{\text{Ass}} + \sigma_{\text{Ass}}[U]) + (F^{\circ}_{\text{Diss}} + \sigma_{\text{Diss}}[U]) \cdot e^{-\frac{(\Delta G^{\circ}_{\text{Ass}} - m[U])}{RT}}}{\left([D_Q] \cdot [D_{Un}] + e^{-\frac{\Delta G^{\circ}_{\text{Ass}} - m[U]}{RT}}\right)} \quad (\text{Eq. 2.3})$$

*Temperature melting curves.* Thermodynamic constants measured using temperature melting curves were performed based on the work of Owczarzy et al.<sup>88</sup>. The melting curves of all

complexes are measured at different concentrations using a fluorimeter (Cary Eclipse, Agilent Technologies). All samples are heated to 90°C at 10°C/min with a 2 minutes hold then cool down to 20°C at 10°C/min with a 10 minutes hold prior the recording of the melting curve from 20°C to 90°C at 0.8°C/min. The  $T_m$  is extracted from the derivative of the melting curve (dF/dT), which displays a "bell shape" curve. A Gaussian distribution fitting is used to evaluate the maximum of this dF/dT plot. The thermodynamic parameters  $\Delta H$  and  $\Delta S$  are measures by linear regression (Eq. 2.4) and enables the calculation of  $\Delta G$ . For all experiments, the concentration of the monomer with a fluorophore (called A) was kept constant at 1  $\mu\text{M}$  for  $\Delta G_{\text{Dim}}$  measurement and at 0.1  $\mu\text{M}$  for  $\Delta G_{\text{Tri}}$  measurement, while the concentration of the other monomer containing the quencher (called B) was varied between 5 and 200  $\mu\text{M}$  for  $\Delta G_{\text{Dim}}$  measurement and between 0.5 and 20  $\mu\text{M}$  for  $\Delta G_{\text{Tri}}$  measurement. In the equation,  $T_m$  is the melting temperature,  $\Delta H$  is the change in enthalpy,  $\Delta S$  is the change in entropy,  $[A]_T$  is the concentration of strand A and  $[B]_T$  is the concentration of strand B.

$$\frac{1}{T_m} = \frac{R}{\Delta H} \ln \left( [B]_T - \frac{[A]_T}{2} \right) + \frac{\Delta S}{\Delta H} \quad (\text{Eq. 2.4})$$

*Binding Curve.* Appropriate dilutions are made such as the volume of the probe solution (strand containing the fluorophore) is always 900  $\mu\text{L}$ . The probe solution contains either one strand (for the 2c system, called B) or two strands (for the 3c system, called B and C). To this solution, a small volume of titrant (strand with a quencher, called A) is added and the fluorescence is recorded after 20 min of equilibration with a fluorimeter (Cary Eclipse, Agilent Technologies). The total volume of titrant added to the probe solution is kept below 5% to avoid high dilution of the probe. We can then assume that the concentration of the probe remains approximately constant over the titration. The titration curve is then fitted using the Hill equation (Eq. 2.5) where  $K_{\text{obs}}$  is the observed dissociation constant (or the concentration at which 50% of the probe is bound, also called  $[A]_{50\%}$ ),  $n_H$  is the Hill factor (an indicator of the cooperativity enabling the calculation of the dynamic range<sup>70</sup>),  $F$  is the fluorescence signal,  $Bsl$  is the baseline fluorescence and  $Amp$  is the amplitude, or change in fluorescence upon binding of the strand A.

$$F = Bsl + Amp \frac{[A]^{n_H}}{K_{obs}^{n_H} + [A]^{n_H}} \quad (\text{Eq. 2.5})$$

*Kinetic experiments.* For fast kinetic (< 30 min), appropriate dilutions are made, and solutions of A and B are rapidly mixed using a stopped-flow instrument coupled with a fluorimeter (SX20, Applied Photophysics). Slower kinetics (> 30 min) are recorded using a standard fluorimeter (Cary Eclipse, Agilent Technologies). A small volume of solution A is added with a pipette inside a cuvette containing solution B and manually mixed using the same pipette. Oil is deposited on top of the solution to avoid water evaporation over the long kinetic measurement. DNA hybridization follows a second-order kinetic:  $A + B \rightleftharpoons AB$  where A and B are single-stranded DNA and AB is the duplex formed from these two strands<sup>89</sup>. Pseudo-first order kinetics are achieved by using at least a 10-fold excess of strand A. The kinetic traces are then fitted using an exponential equation (Eq. 2.6) where  $k_{obs}$  is the first-order observed rate constant. The relationship of  $k_{obs}$  with the concentration of A is linear enabling the measurement of the association rate constant ( $k_{ass}$ ) and the dissociation rate constant ( $k_{diss}$ ). Second-order kinetics are achieved when using equimolar concentrations of A and B ( $[A]_0 = [B]_0$ ). This system acts like a dimerization reaction (i.g.  $2A \rightleftharpoons A_2$ ) and can thus be fitted using Eq. 2.7 where  $k_{obs}$  is now the second-order observed rate constant.

$$F = Bsl + Amp \cdot e^{-k_{obs}t} \quad (\text{Eq. 2.6})$$

$$F = Bsl + Amp \cdot \frac{1}{1 + 2[A]_0 k_{obs}t} \quad (\text{Eq. 2.7})$$

The trimeric association follows this reaction:  $A + B + C \rightleftharpoons ABC$  where A, B, and C are single-stranded DNA and ABC is the three-way junction formed from these strands. When measuring the assembly, we use at least a 10-fold excess of molecule A in buffer that is rapidly mixed using a stopped-flow instrument coupled with a fluorimeter (SX20, Applied Photophysics) with a solution of B and C in water (to avoid pre-association of B and C). Data are fitted using a combination of a pseudo-first-order kinetic and a second-order kinetic (Eq. 2.8).

$$F = Bsl + Amp_1 \cdot e^{-k_{obs,2}t} + Amp_2 \cdot \frac{1}{1 + 2[A]_0 k_{obs,1}t} \quad (\text{Eq. 2.8})$$

In all equation,  $t$  is the time,  $F$  is the fluorescence signal,  $Bsl$  is the baseline fluorescence,  $Amp$  is the amplitude, or change in fluorescence upon binding and  $[A]_0$  is the concentration of strand A.

### **Native PAGE experiments.**

Appropriate dilutions of unlabelled DNA solutions were made such as the concentration of strands B and C is kept at 1  $\mu$ M and the concentration of strand A is changed from 30 nM to 100  $\mu$ M. Solutions are then mixed in a 5:1 ratio with the 6x loading buffer (2.5 mg/mL bromothymol blue, 2.5 mg/mL xylene cyanol FF and 30% glycerol in water). A 15% polyacrylamide gel is hand cast following Bio-Rad protocol and incubated in the running buffer (0.5x TBE buffer containing 5 mM of  $MgCl_2$ ) for 1h. 10  $\mu$ L of samples are run for 90 min at 120 V using the Mini-PROTEAN Tetra cell electrophoresis unit (Bio-Rad) and the Bio-Rad PowerPac Basic power supply. Gels are stained with a 0.5x solution of GelRed™ (Biotium) for 10 minutes and are analyzed on an imaging system (ChemiDoc™ XRS+, Bio-Rad). The integration of band intensity is then performed to evaluate the amount of assembled DNA-based systems.

### **2c and 3c equilibrium binding curve models.**

#### 2c equilibrium binding curve model

The 2-component system is assumed to proceed through a single-site binding mechanism following this equilibrium:  $A + B \rightleftharpoons AB$  where A is the titrant molecule and B is the probe molecule (fixed concentration). The dissociation equilibrium constant ( $K_D$ ) of this system is described by Eq. 2.9. The observed dissociation constant ( $K_{obs}$ ) is defined by the titrant concentration needed to bind half of the probe ( $[A]_{50\%}$ ). The concentration of AB at 50% thus corresponds to half the concentration of probe ( $[AB]_{50\%} = 0.5[B]_T$ ) and can be inserted into Eq. 2.9 to give Eq. 2.10. By isolating the concentration of titrant needed to achieve 50% of formed AB, we can get the equation for  $K_{obs}$  which corresponds to Eq. 2.11. The dynamic range (DR) is defined by the ratio of titrant concentration needed to bind 90% and 10% of the probe ( $DR = [A]_{90\%}/[A]_{10\%}$ ). We use the same calculation method as for  $K_{obs}$  and get the equation for the DR which corresponds to Eq. 2.12.



$$K_D = \frac{[A]_{eq}[B]_{eq}}{[AB]_{eq}} \quad (\text{Eq. 2.9})$$

$$K_D = \frac{([A]_{50\%} - [AB]_{50\%}) \cdot ([B]_T - [AB]_{50\%})}{[AB]_{50\%}}; [AB]_{50\%} = 0,5[B]_T \quad (\text{Eq. 2.10})$$

$$K_{obs} = [A]_{50\%} = K_D + \frac{[B]_T}{2} \quad (\text{Eq. 2.11})$$

$$R = \frac{[A]_{90\%}}{[A]_{10\%}} = 81 \frac{(K_D + 0.1[B]_T)}{(K_D + 0.9[B]_T)} \quad (\text{Eq. 2.12})$$

### 3c equilibrium binding curve model

The 3-component assembly model is assumed to proceed through the formation of a dimer by following this equilibrium:  $A + B + C \rightleftharpoons A + BC \rightleftharpoons ABC$  where A is the titrant molecule and B and C are the probe molecules (fixed concentrations). The dissociation equilibrium constant for the formation of the dimer ( $K_{D,D}$ ) and the dissociation equilibrium constant for the formation of the trimer ( $K_{D,T}$ ) are described by Eq. 2.13. The observed dissociation constant ( $K_{obs}$ ) is defined by the titrant concentration needed to bind half of the probes ( $[A]_{50\%}$ ). The concentration of ABC at 50% thus correspond to half the concentration of the probes ( $[ABC]_{50\%} = 0.5[B]_T = 0.5[C]_T$ ) if using equimolar concentration of probes. We obtain the equation providing  $K_{obs}$  (Eq. 2.16) by isolating  $[BC]_{50\%}$  in the  $K_{D,D}$  equation using the quadratic equation (Eq. 2.15) and by introducing it into the  $K_{D,T}$  equation (Eq. 2.14). The dynamic range (DR) is defined by the ratio of titrant concentration needed to bind 90% and 10% of the probes ( $DR = [A]_{90\%}/[A]_{10\%}$ ). We use the same calculation method as for  $K_{obs}$  and get the equation for the DR which corresponds to Eq. 2.18 using Eq. 2.17. When using a non-equimolar condition between  $[B]_T$  and  $[C]_T$ , the equation of  $[BC]_{50\%}$  must be changed following Eq. 2.19 by introducing the parameter R which is the ratio of concentration between B and C ( $R = [C]_T / [B]_T$ ).

$$K_{D,D} = \frac{[B]_{eq}[C]_{eq}}{[BC]_{eq}}; K_{D,T} = \frac{[A]_{eq}[BC]_{eq}}{[ABC]_{eq}} \quad (\text{Eq. 2.13})$$

$$K_{D,D} = \frac{([B]_T - [BC]_{50\%} - [ABC]_{50\%}) \cdot ([C]_T - [BC]_{50\%} - [ABC]_{50\%})}{[BC]_{50\%}}; \quad (\text{Eq. 2.14})$$

$$K_{D,T} = \frac{([A]_{50\%} - [ABC]_{50\%}) \cdot [BC]_{50\%}}{[ABC]_{50\%}} ; [ABC]_{50\%} = 0,5[B]_T = 0,5[C]_T$$

$$[BC]_{50\%} = \frac{K_{D,D} + [B]_T - \sqrt{K_{D,D}^2 + 2[B]_T K_{D,D}}}{2} \quad (\text{Eq. 2.15})$$

$$RK_{obs} = [A]_{50\%} = \frac{[B]_T}{2} \left( 1 + \frac{K_{D,T}}{[BC]_{50\%}} \right) \quad (\text{Eq. 2.16})$$

$$[BC]_{90\%} = \frac{K_{D,D} - 2[B]_T(0.9 - 1) - \sqrt{K_{D,D}^2 + 4[B]_T K_{D,D}(1 - 0.9)}}{2} \quad (\text{Eq. 2.17})$$

$$[BC]_{10\%} = \frac{K_{D,D} - 2[B]_T(0.1 - 1) - \sqrt{K_{D,D}^2 + 4[B]_T K_{D,D}(1 - 0.1)}}{2}$$

$$DR = \frac{[A]_{90\%}}{[A]_{10\%}} = 9 \frac{\left( 1 + \frac{K_{D,T}}{[BC]_{90\%}} \right)}{\left( 1 + \frac{K_{D,T}}{[BC]_{10\%}} \right)} \quad (\text{Eq. 2.18})$$

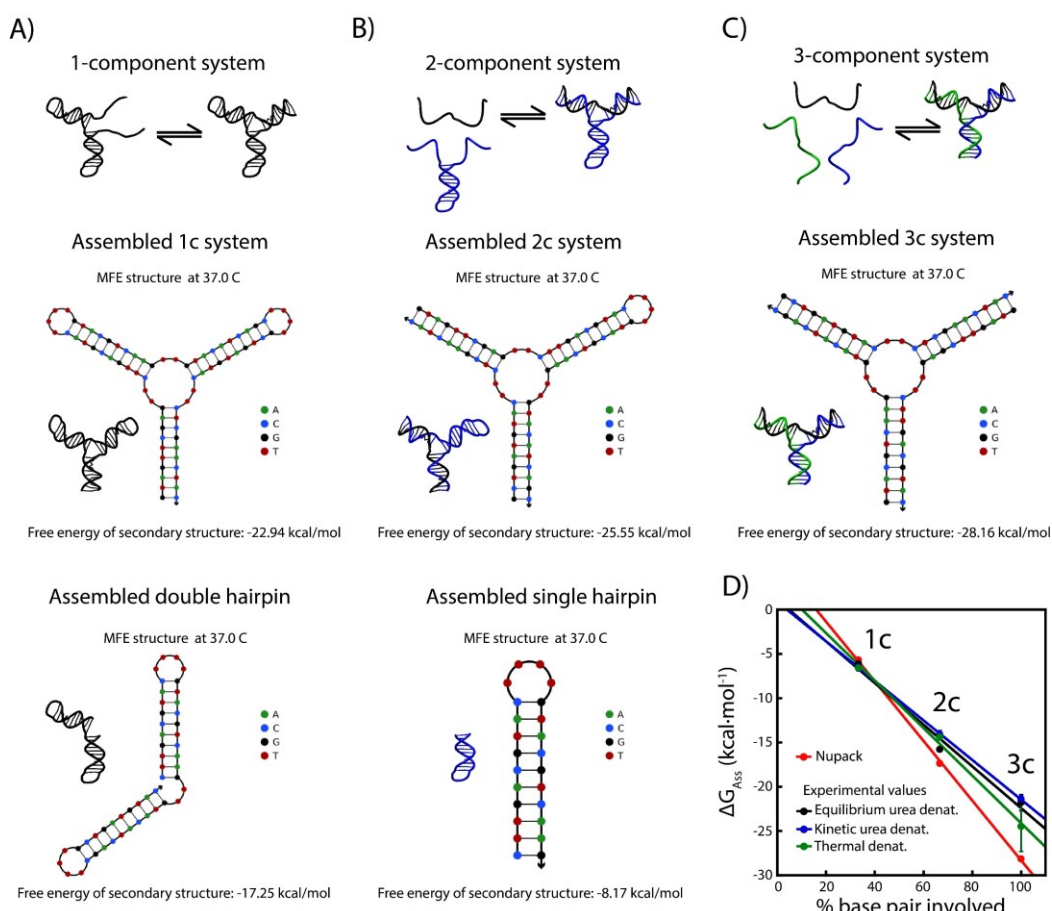
$$[BC] = \frac{K_{D,D} - [B]_T(2x - 1 - R) - \sqrt{([B]_T(2x - 1 - R) - K_{D,D})^2 + 4[B]_T(1 - x)(R - x)}}{2} \quad (\text{Eq. 2.19})$$

where  $x = 0.9, 0.5$  or  $0.1$  for  $[BC]_{90\%}, [BC]_{50\%}$  and  $[BC]_{10\%}$  respectively

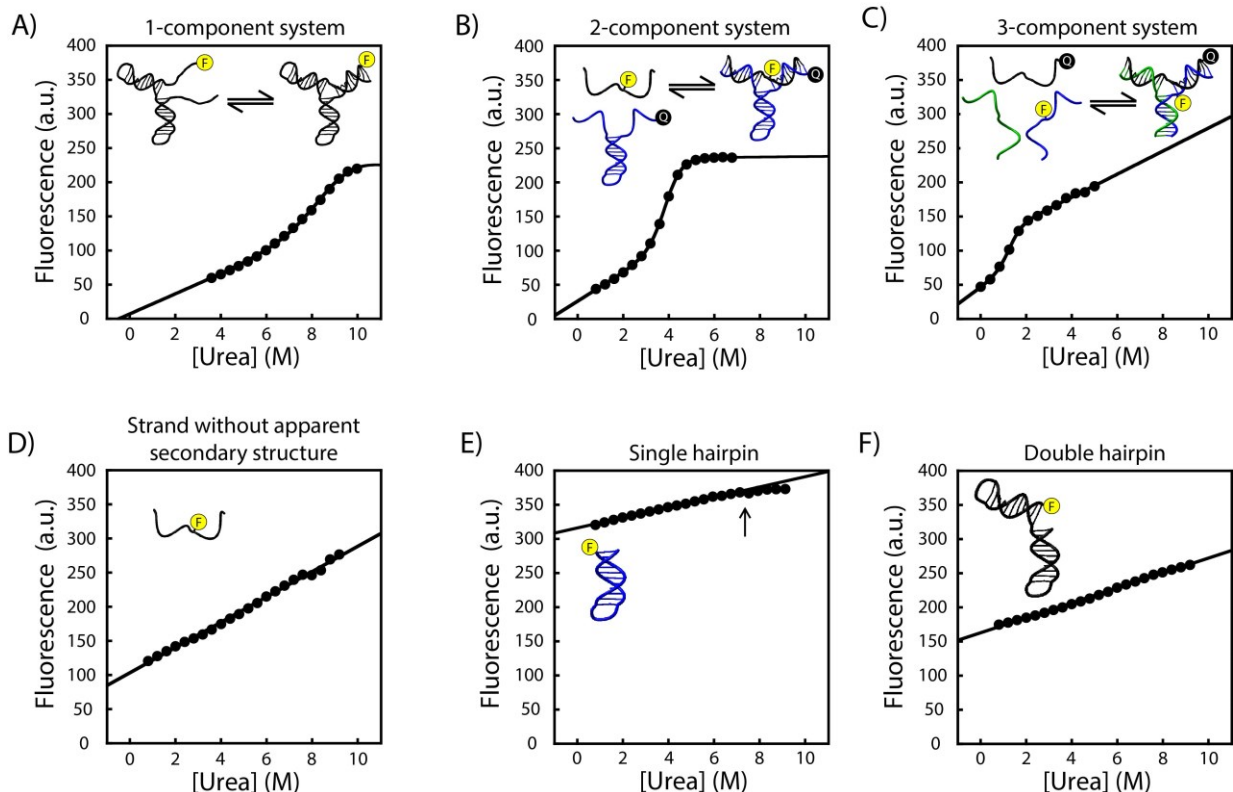
## 2.6 Acknowledgements

This research was conducted through the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants (RGPIN-2020-06975) (A.V.-B). A.V.-B. is Canada Research Chair in Bioengineering and Bionanotechnology, Tier II. D. L. acknowledges a Canada graduate scholarship master (CGS M) from NSERC and a 3<sup>rd</sup> cycle scholarship from the Fonds de recherche du Québec – Nature et technologies (FRQNT). The authors would like to thank L. Pedro, K. Nemčková and S. G. Harroun for the helpful discussions and comments on the paper.

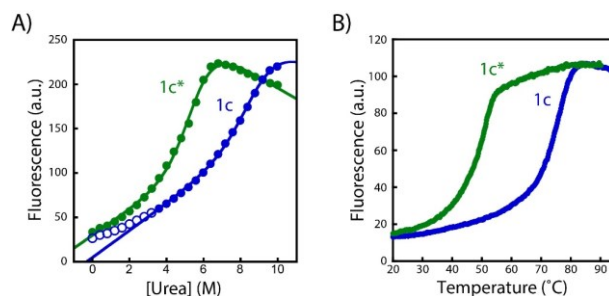
## 2.7 Supporting Figures



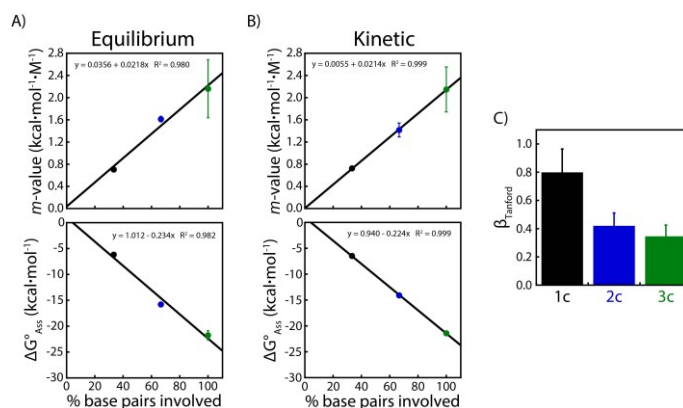
**Figure S2.1 – Nupack analysis of our DNA based nanostructures.** The 1c system displays a predicted free energy ( $\Delta G$ ) of -22.94 kcal/mol while its preorganized loops, which cannot be unfolded using urea (Fig. S2.2), show a predicted  $\Delta G$  of -17.25 kcal/mol. The subtraction of both leads to a free energy of assembly ( $\Delta G_{Ass}$ ) of -5.69 kcal/mol. B) The 2c system displays a predicted  $\Delta G$  of -25.55 kcal/mol while its preorganized loop has a predicted  $\Delta G$  of -8.17 kcal/mol. The subtraction of both leads to a  $\Delta G_{Ass}$  of -17.38 kcal/mol. C) The 3c system displays a  $\Delta G$  of -28.16 kcal/mol. D) All  $\Delta G$  predicted by NUPACK are in good agreement with the experimentally derived values and correlate with the number of base pairs involved in the transition. Of note, NUPACK seems to overestimate the  $\Delta G_{Ass}$  of the 2c system and the 3c system. This is not unusual as similar discrepancies were observed for the urea denaturation of DNA-DNA complexes<sup>38</sup>.



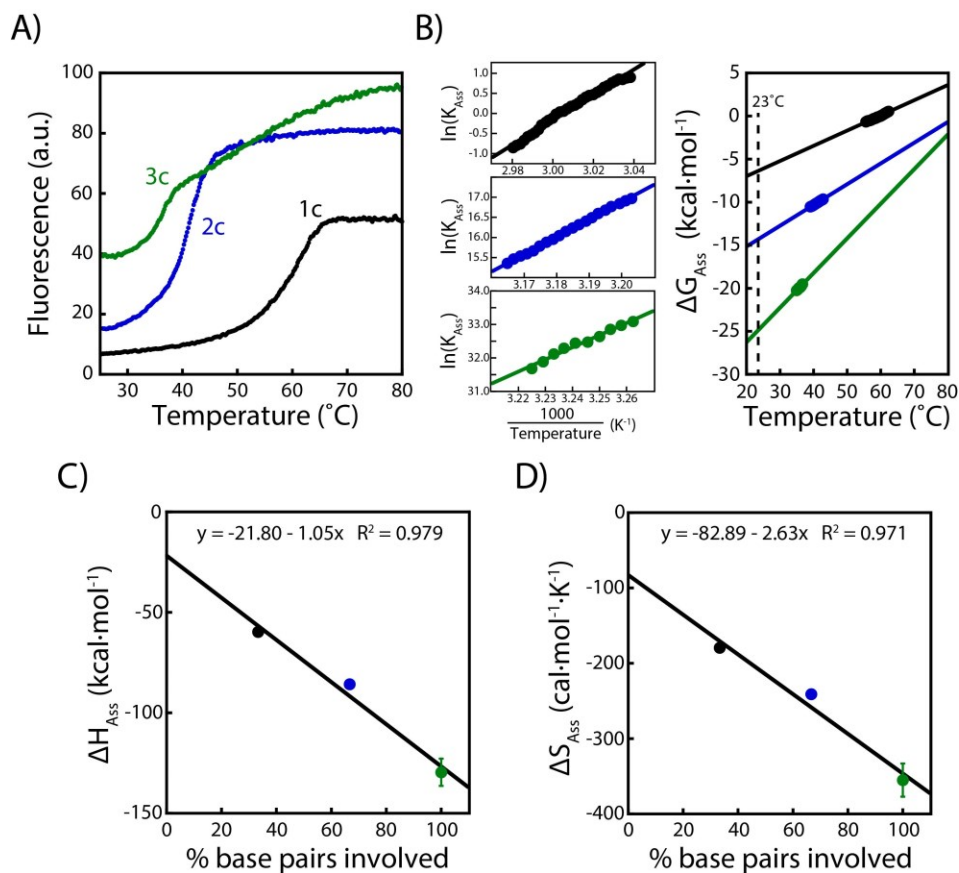
**Figure S2.2 – Urea denaturation profiles.** A) 1-component (see also Fig. S2.3), B) 2-component, and C) 3-component systems. D) Urea affects linearly the fluorescence of the FAM molecule, which is also observed for the denaturation of the 1c, 2c, and 3c systems. E-F) No visible change in fluorescence is observed when denaturing the single hairpin and the double hairpin present in the 2c and 1c systems, suggesting that the hairpin remains folded up to 10 M urea. Of note, a little transition starting around 8 M urea (see arrow) is observed in the denaturation of the single hairpin. However, this does not affect the analysis of the dissociation transition of the 2-component system that takes place at 4 M urea. Data are fitted using Eq. 2.1 for panel A, Eq. 2.2 for panel B, Eq. 2.3 for panel C and linear regression for panel D-F. All experiments are performed at 23°C in PBS buffer (10 mM NaH<sub>2</sub>PO<sub>4</sub>, 40 mM NaCl, pH = 7.00). The fluorophore FAM is represented by a yellow circle and the quencher BHQ-1 is represented by a black circle.



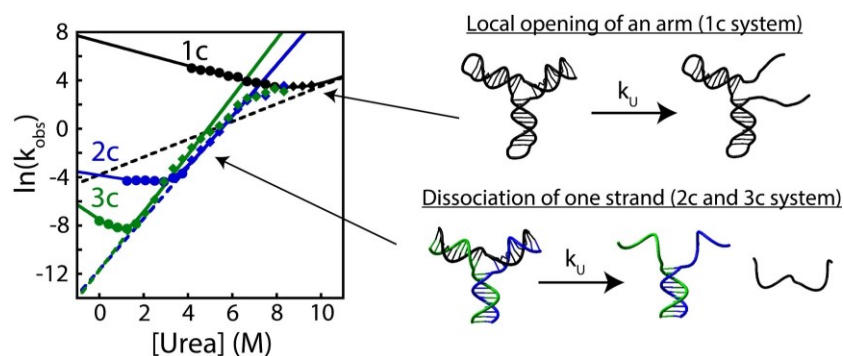
**Figure S2.3 – The 1-component urea denaturation profile in Fig. S2.2 was fitted by using the fluorescence behaviour of the unfolded state determined on a less stable 1c\* systems ( $F^{\circ}_{\text{Diss}}$ ,  $\sigma_{\text{Ass}}$ ).** A) Urea unfolding of the 1c systems (blue) is not fully achievable even at 10 M urea making it hard to estimate the fluorescence behaviour of the unfolded state. We found that removing the NaCl from the buffer decreases the thermal stability ( $T_m$ ) of the 1c system by up to 25°C while keeping the 1c system still assembled at 23°C (panel B). This enables us to estimate the fluorescence of the disassembled state at 0 M urea ( $F^{\circ}_{\text{Diss}} = 316$  a.u.) and to estimate the linear fluorescence dependency of the assembled state ( $\sigma_{\text{Ass}} = 14.5$  a.u./M). We use these two values to facilitate the fitting in the native condition (blue) using Eq. 2.1. Native experiments (in blue) are performed at 23°C in PBS buffer (10 mM  $\text{NaH}_2\text{PO}_4$ , 40 mM NaCl, pH = 7.00).



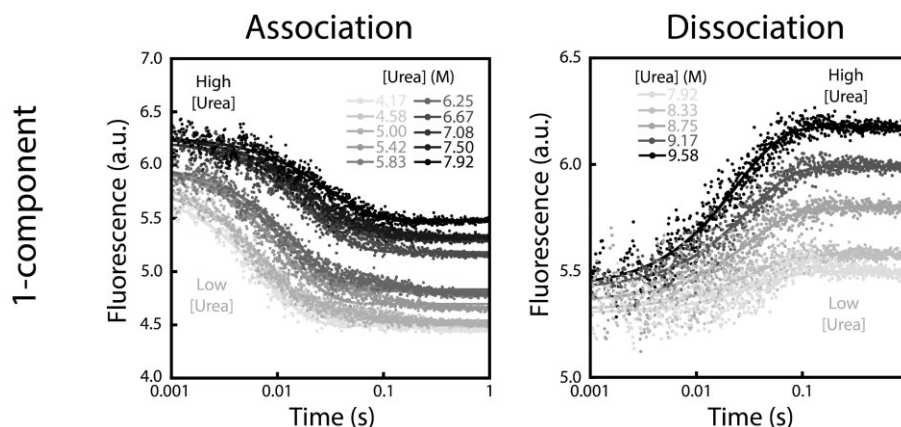
**Figure S2.4 –  $m$ -values and  $\Delta G^{\circ}_{\text{Ass}}$  acquired using: A) urea equilibrium unfolding experiments and, B) assembly/disassembly kinetic experiments using urea concentration jumps.**  $m$ -values and  $\Delta G^{\circ}_{\text{Ass}}$  are proportional to the number of base pairs that are broken/formed as the 3-way junction assembles/disassembles. Given that the hairpins remain folded in the disassembled state (see Fig. S2.1 to Fig. S2.3), only 33% and 67% of the base pairs are predicted to be involved in the transition of the 1-component system and the 2-component system, respectively. C) The  $\beta_{\text{Tanford}}$ , a measure of the level of compaction of the transition state, can be determined by comparing the effect of denaturant on the unfolding rate ( $m_U$ ) with its effect on the unfolding equilibrium ( $m$ ) (see Fig. S2.6 to Fig. S2.9 for kinetic measurements). To facilitate the comparison between the transition states, we employed the  $m$ -value of the fully disassembled 3-component system ( $m = 2.15 \pm 0.40$  kcal·mol $^{-1}$ ·M $^{-1}$ ) and assumed that all systems have the same level of compaction in the assembled state. As expected, we found that the transition state of the 1c system is much closer to the assembled state compared to the 2c and 3c systems. Black = 1c system, blue = 2c system, green = 3c system. Error bars represent the error from the non-linear regression.



**Figure S2.5 – Thermal denaturation analyses of the 1c, 2c and 3s systems.** A) Thermal denaturation profiles of the 1-component (black), 2-component (blue) and 3-component systems (green). B) *Left*: Van't Hoff analysis of the thermal denaturation curves enables the extraction of the thermodynamic parameters  $\Delta H_{\text{Ass}}$  and  $\Delta S_{\text{Ass}}$  (see panels C and D, respectively). *Right*: extrapolation of the  $\Delta G_{\text{Ass}}$  at 23°C (right side) are in good agreement with the values obtained using urea denaturation. C-D) As observed for the  $m$ -values and  $\Delta G^{\circ}_{\text{Ass}}$  (see Fig. S2.4),  $\Delta H_{\text{Ass}}$  and  $\Delta S_{\text{Ass}}$  are also linearly dependant on the number of base pairs broken/formed in the assembly/disassembly transition.  $\Delta H_{\text{Ass}} = -60 \pm 1 \text{ kcal}\cdot\text{mol}^{-1}$ ,  $-85.8 \pm 0.8 \text{ kcal}\cdot\text{mol}^{-1}$ , and  $-130 \pm 7 \text{ kcal}\cdot\text{mol}^{-1}$ , for the 1c, 2c, and 3c system, respectively.  $\Delta S_{\text{Ass}} = -179 \pm 4 \text{ cal}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ ,  $-241 \pm 3 \text{ cal}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ , and  $-355 \pm 22 \text{ cal}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$  for the 1c, 2c, and 3c system, respectively. All experiments are performed in PBS buffer (10 mM  $\text{NaH}_2\text{PO}_4$ , 40 mM NaCl, pH = 7.00). Error bars represent the error from the non-linear regression.

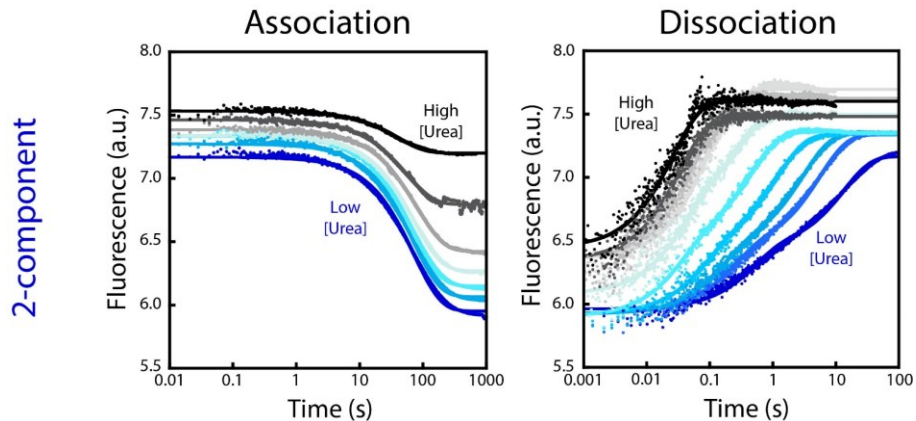


**Figure S2.6 – Kinetic analysis (chevron plot) of the 1c, 2c, and 3c systems.** A) The kinetics of the association/folding (circles) and dissociation/unfolding (diamonds) of all nanosystems under various urea concentrations (chevron plot) reveal a two-state association/dissociation mechanism around their denaturation transition (see Table S2.1). The 2c and 3c systems display similar unfolding transitions ( $k_{U,2c} = 8.6 \pm 3.1 \times 10^{-6} \text{ s}^{-1}$  and  $k_{U,3c} = 8.3 \pm 3.2 \times 10^{-6} \text{ s}^{-1}$ ;  $m_{U,2c} = 1.24 \pm 0.04 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$  and  $m_{U,3c} = 1.41 \pm 0.06 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ ) suggesting that they dissociate via the same mechanism (dissociation of a complete strand). In contrast, the 1c system displays an unfolding rate that is 3000-time faster ( $k_{U,1c} = 2.2 \pm 0.9 \times 10^{-2} \text{ s}^{-1}$ ) with a 3-time smaller urea dependency ( $m_{U,1c} = 0.43 \pm 0.01 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ ). This is consistent with the smaller local unfolding expected to take place for the 1c system, the unfolding slope ( $m_U$ ) is proportional to the surface area made accessible upon dissociation/unfolding. The slower dissociation kinetics of the 2c and 3c systems, extrapolated in absence of urea, is also consistent with the fact their dissociation requires the disruption of 20 base pairs compared to the disruption of 10 base pairs for the 1c system (see cartoon). Interestingly, the dissociation mechanism for the 2c and 3c systems becomes non-linear and similar to the 1c system at a high concentration of urea ( $> 7\text{M}$ ) suggesting that their disassembly becomes only rate limited by the opening of one arm. For raw kinetic traces see Fig. S2.7 to Fig. S2.9 below.

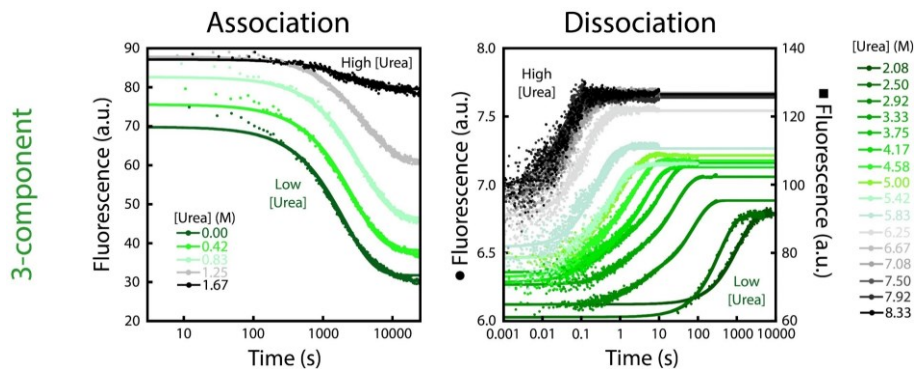


**Figure S2.7 – Kinetic traces for the association and dissociation of the one-component system at different urea concentrations.** All data were fitted using a single exponential function (Eq. 2.6). All experiments are performed in PBS buffer (10 mM  $\text{NaH}_2\text{PO}_4$ , 40 mM NaCl, pH = 7.00).





**Figure S2.8 – Kinetic traces for the association and dissociation of the two-component system at different urea concentrations.** All association data were fitted using a single exponential function whereas dissociation data were fitted using a single exponential function or a double exponential function (Eq. 2.6). At the midpoint of denaturation ( $\sim 4$  M), the association rates converge with the slowest transition detected during the dissociation process (applied only for dissociation curves near 4 M). The slowest phase of the double exponential was therefore employed for the analysis. All experiments were performed in PBS buffer (10 mM  $\text{NaH}_2\text{PO}_4$ , 40 mM NaCl, pH = 7.00).



**Figure S2.9 – Kinetic traces for the association and dissociation of the three-component system at different urea concentrations.** All association data were fitted using a single exponential function whereas all dissociation data were fitted using a single exponential function or a double exponential function (Eq. 2.6). At the midpoint of denaturation ( $\sim 2$  M), the association rates converge with the slowest transition detected during the dissociation process (applied only for dissociation curves near 2 M). The slowest phase was therefore employed for the analysis. All experiments were performed in PBS buffer (10 mM  $\text{NaH}_2\text{PO}_4$ , 40 mM NaCl, pH = 7.00).



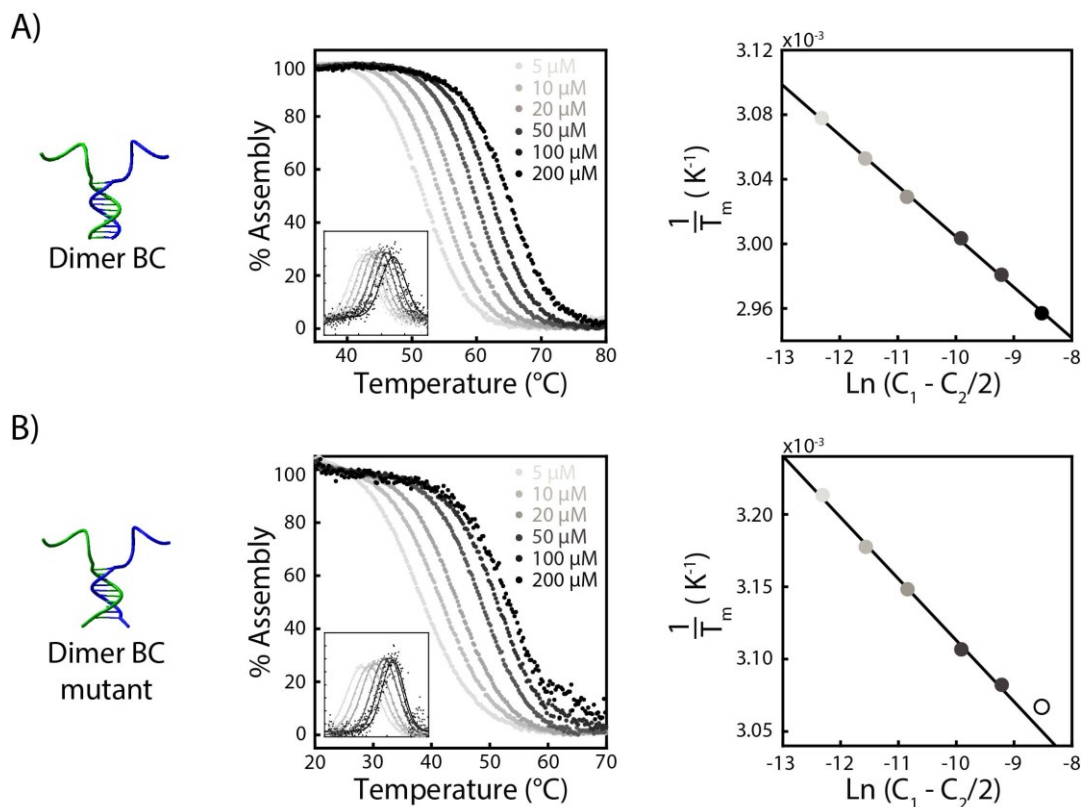
**Table S2.1. – Thermodynamic and kinetic parameters estimated from the chemical denaturation of the one-, two- and three-components system by urea denaturation, thermal denaturation and NUPACK.**

	Urea denaturation at equilibrium (Fig. 2.1B, Fig. S2.4A)			Urea kinetics denaturation (Fig. S2.6, Fig. S2.4B)			Melting curve analyses (Fig. S2.5)			NUPACK Prediction (Fig. S2.1)		
	1c	2c	3c	1c	2c	3c	1c	2c	3c	1c	2c	3c
$\Delta G^{\circ}_{Ass}$ (kcal·mol <sup>-1</sup> )	-6.4 ± 0.4	-15.8 ± 0.2	-21.8 ± 0.9	-6.5 ± 0.4 <sup>a</sup>	-14.1 ± 0.4 <sup>a</sup>	-21.4 ± 0.4 <sup>a</sup>	-6.6 ± 0.2	-14.4 ± 0.3	-24.5 ± 2.8	-5.69	-17.38	-28.16
$m_F$ (kcal·mol <sup>-1</sup> ·M <sup>-1</sup> )				0.29 ± 0.02	0.17 ± 0.08	0.74 ± 0.34						
$m_U$ (kcal·mol <sup>-1</sup> ·M <sup>-1</sup> )				0.43 ± 0.01	1.24 ± 0.04	1.41 ± 0.06						
$m$ (kcal·mol <sup>-1</sup> ·M <sup>-1</sup> )	0.71 ± 0.05	1.61 ± 0.05	2.1 ± 0.5	0.72 ± 0.03 <sup>b</sup>	1.42 ± 0.12 <sup>b</sup>	2.15 ± 0.40 <sup>b</sup>						
$\beta_{Tanford}$				0.80 ± 0.17 <sup>c</sup>	0.42 ± 0.09 <sup>c</sup>	0.34 ± 0.08						
$k_F$				(1.4 ± 0.2) × 10 <sup>3</sup> s <sup>-1</sup>	(2.2 ± 0.6) × 10 <sup>5</sup> s <sup>-1</sup> ·M <sup>-1</sup>	(5.3 ± 1.5) × 10 <sup>10</sup> s <sup>-1</sup> ·M <sup>-2</sup>						
$k_U$ (s <sup>-1</sup> )				(2.2 ± 0.9) × 10 <sup>-2</sup>	(8.6 ± 3.1) × 10 <sup>-6</sup>	(8.3 ± 3.2) × 10 <sup>-6</sup>						
$\Delta H_{Ass}$ (kcal·mol <sup>-1</sup> )							-60 ± 1	-86 ± 1	-130 ± 7			
$\Delta S_{Ass}$ (cal·mol <sup>-1</sup> ·K <sup>-1</sup> )							-179 ± 4	-241 ± 3	-355 ± 22			

<sup>a</sup> The kinetic  $\Delta G^{\circ}_{Ass}$  were determined from the equilibrium constant calculated using  $k_F$  and  $k_U$ .

<sup>b</sup> The kinetic  $m$ -values were determined using  $m_F + m_U$ .

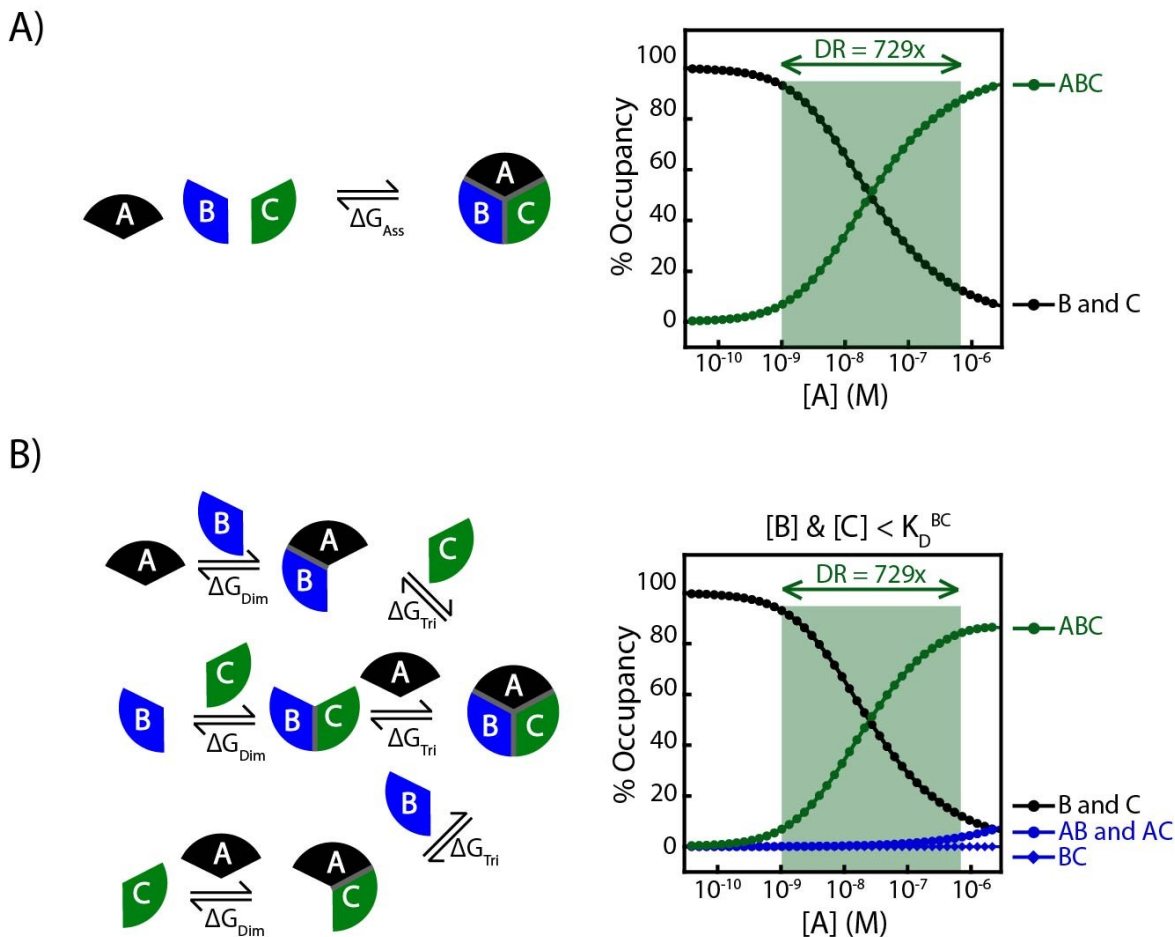
<sup>c</sup> The level of compaction of the transition state was compared to the fully disassembled state of the 3c system assuming that all systems have the same level of compaction in the assembled state.



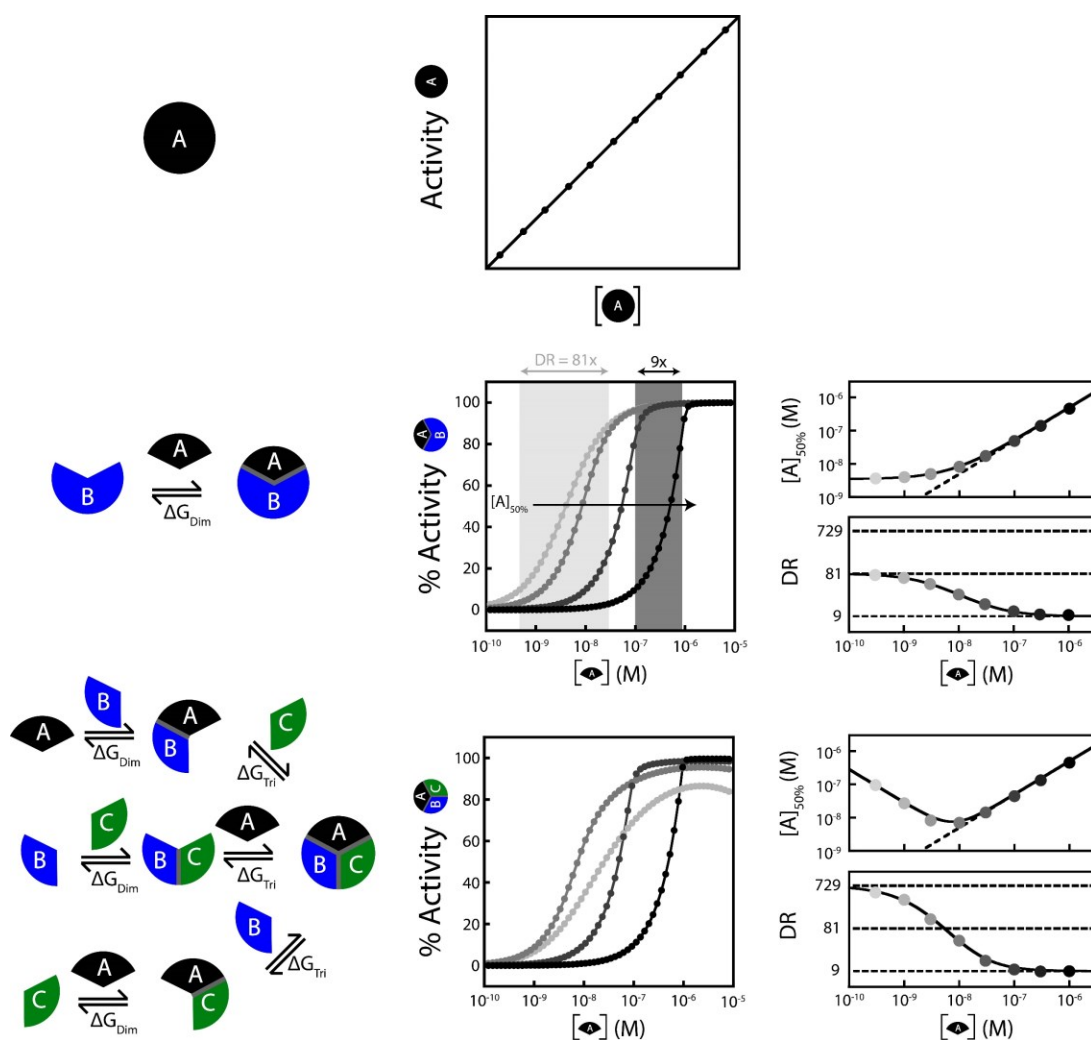
**Figure S2.10 – Estimation of the  $K_D$  between component B and component C.** A) Melting curve analysis using 1  $\mu\text{M}$  of the fluorescent component B and an excess of component C (5 to 200  $\mu\text{M}$ ). B) Melting curve analysis using 1  $\mu\text{M}$  of the fluorescent component B and an excess of mutated component C (5 to 200  $\mu\text{M}$ ). The inset represents the derivative of the melting curve ( $dF/dT$ ) fitted with a Gaussian distribution, which has been used to extract the  $T_m$  that is then fitted using Eq. 2.4 to extract the thermodynamic parameters. All experiments have been done in PBS buffer (50 mM  $\text{Na}_2\text{HPO}_4$ , 400 mM  $\text{NaCl}$ , pH = 7.00).

**Table S2.2. – Dissociation constants of the dimer BC.**

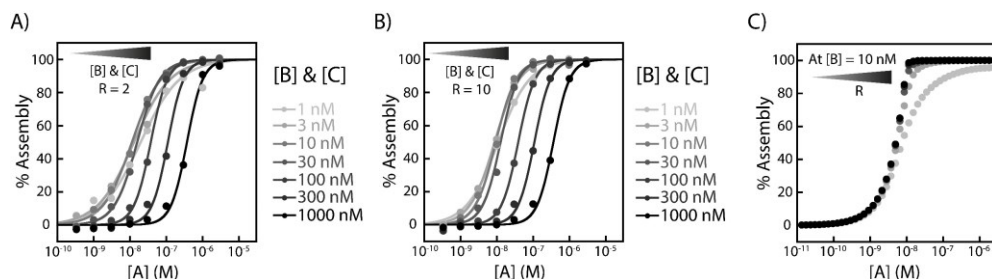
System	Temperature ( $^{\circ}\text{C}$ )	$\Delta G^{BC}$ ( $\text{kcal}\cdot\text{mol}^{-1}$ )	$K_D^{BC}$
Dimer BC	30	$-11.7 \pm 0.4$	$3.8 \pm 2.4$ nM
	37	$-10.5 \pm 0.3$	$41 \pm 23$ nM
	45	$-9.1 \pm 0.3$	$547 \pm 254$ nM
Dimer BC, mutant	37	$-7.8 \pm 0.4$	$3.3 \pm 2.3$ $\mu\text{M}$



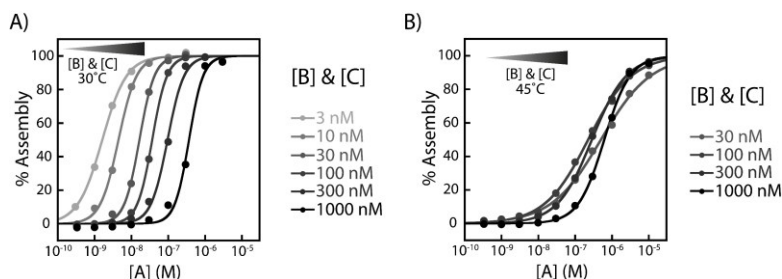
**Figure S2.11 – Understanding the broad response coming with of a system made of three components.** A) Numerical simulation of a trimeric nanosystem that assembles directly from monomer to a trimer displays a broad transition with a dynamic range of 729. B) Similarly, when using a concentration of B and C below  $K_D^{BC}$ , the dimeric intermediate BC is not pre-assembled, and the trimer assembles also directly from monomers showing a broad transition with a dynamic range 729. Of note, at high excess of A ( $> 10^{-6}$  M), the trimeric assembly diminishes, and the dimeric intermediates (AB and AC) increases. Numerical simulations are done using MATLAB® with a dimeric stability of  $-8$  kcal/mol ( $\Delta G_{Dim}$ ), a trimeric stability of  $-16$  kcal/mol ( $\Delta G_{Tri}$ ) and an overall assembly stability of  $-24$  kcal/mol ( $\Delta G_{Ass} = \Delta G_{Dim} + \Delta G_{Tri}$ ). The concentration of B and C are set at 1 nM and the concentration of A is varied from 0.03 nM to 3000 nM.



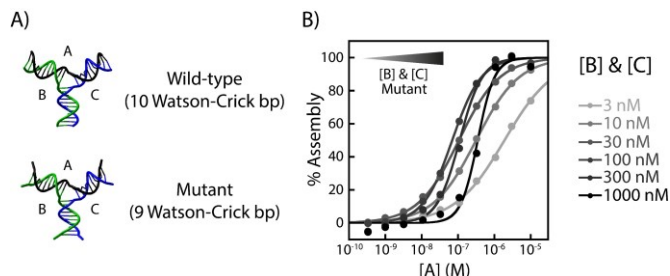
**Figure S2.12 – The assembly behaviour of the 2c and 3c systems (Fig. 2.2) are well predicted by numerical simulations.** *Top.* Because the one-component system already folds into its active conformation, its activity remains linear with its concentration. *Middle.* The assembly of the two-components system induced by an increase of component A can be triggered at different  $[A]_{50\%}$  by increasing the concentration of the limiting component (here B). While  $[A]_{50\%}$  shifts towards higher concentration, the dynamic range (DR) shifts from 81-fold to 9-fold. *Bottom.* The assembly of the three-components system induced by an increase of component A can also be triggered at different  $[A]_{50\%}$  by increasing the concentration of the limiting components (here B and C). Interestingly, the  $[A]_{50\%}$  shifts from high to low concentration and shifts back to high concentration, while the dynamic range (DR) shifts from 729-fold to 9-fold. Of note, at a low concentration of components B and C, the overall yield of the 3-component system decreases in the presence of large excess of component A. This is because the system favours the formation of the two dimeric intermediates (here AB and AC) instead of the trimer (see also Fig. 2.3 and Fig. S2.18). Numerical simulations are done using MATLAB<sup>®</sup> with a dimeric stability of  $-8$  kcal/mol ( $\Delta G_{\text{Dim}}$ ) and a trimeric stability of  $-16$  kcal/mol ( $\Delta G_{\text{Tri}}$ ) leading to an overall assembly stability of  $-24$  kcal/mol ( $\Delta G_{\text{Ass}} = \Delta G_{\text{Dim}} + \Delta G_{\text{Tri}}$ ).



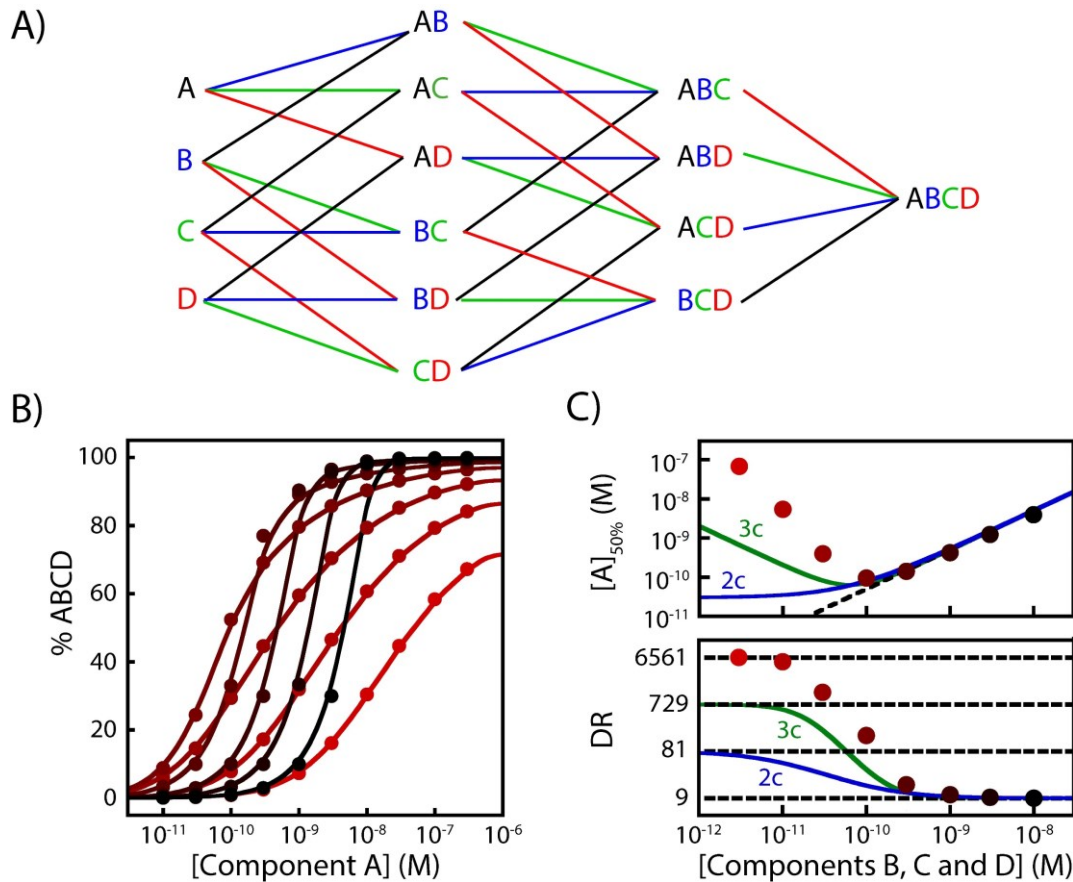
**Figure S2.13 – We can tune the  $[A]_{50\%}$  and the dynamic range of the 3c system by simply changing the concentration ratio between the component B and C.** Binding curve experiments at a fixed  $[C]/[B]$  ratio of A)  $R = 2$  and B)  $R = 10$ . C) Numerical simulations of the thermodynamic model also support the experimental observations illustrated in Fig. 2.2D, *left*. All experiments are done in PBS buffer (50 mM  $\text{Na}_2\text{HPO}_4$ , 400 mM NaCl, pH = 7.00) at 37°C. Binding curves are fitted using the Hill equation (Eq. 2.5),  $[A]_{50\%}$  curves are fitted using Eq. 2.16 with the use of Eq. 2.19 and dynamic range curves are fitted using Eq. 2.18 with the use of Eq. 2.19.



**Figure S2.14 – We can also tune the  $[A]_{50\%}$  and the dynamic range of the 3c system by simply changing the affinity between the components via a temperature change.** Binding curve experiments at A) 30°C and B) 45°C.  $[A]_{50\%}$  and dynamic range profiles are shifted towards lower concentration when the temperature is decreased (30°C, stabilizing both  $\Delta G_{\text{Dim}}$  and  $\Delta G_{\text{Tri}}$ ) and are shifted towards higher concentration when the temperature is increased (45°C, destabilizing both  $\Delta G_{\text{Dim}}$  and  $\Delta G_{\text{Tri}}$ ). All experiments are done in PBS buffer (50 mM  $\text{Na}_2\text{HPO}_4$ , 400 mM NaCl, pH = 7.00). Binding curves are fitted using the Hill equation (Eq. 2.5),  $[A]_{50\%}$  curves are fitted using Eq. 2.16 and dynamic range curves are fitted using Eq. 2.18.

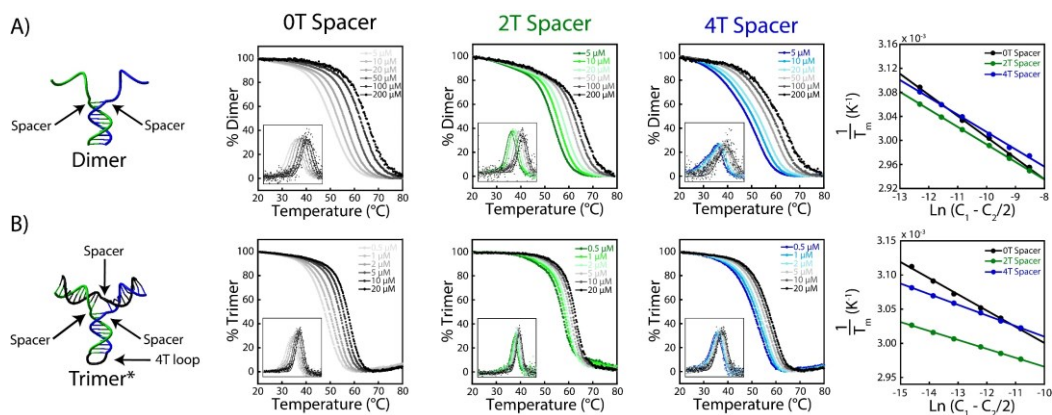


**Figure S2.15 – We can tune the  $[A]_{50\%}$  and the dynamic range of the 3c system by simply performing mutations that reduce the affinity between the components.** A) We performed mutations at the extremities of each arm destabilizing both  $\Delta G_{\text{Dim}}$  and  $\Delta G_{\text{Tri}}$ . B) Binding curve experiments with a mutated 3-way junction containing 9 Watson-Crick base pairs. All experiments are done in PBS buffer (50 mM  $\text{Na}_2\text{HPO}_4$ , 400 mM NaCl, pH = 7.00) at 37°C. Binding curves are fitted using the Hill equation (Eq. 2.5),  $[A]_{50\%}$  curves are fitted using Eq. 2.16 and dynamic range curves are fitted using Eq. 2.18.

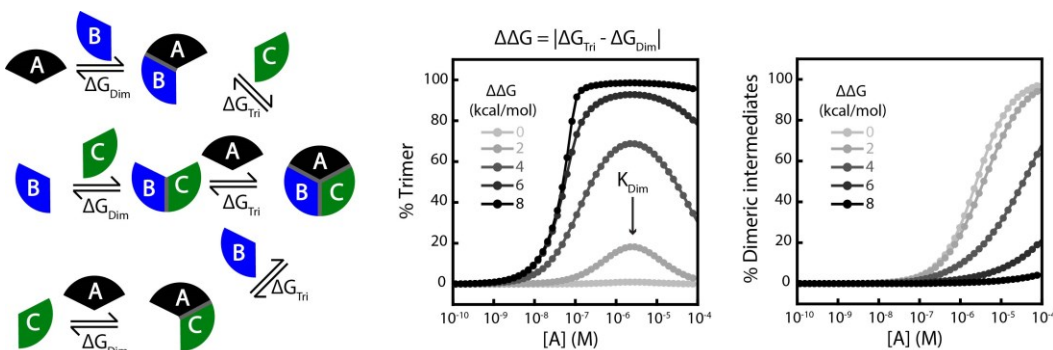


**Figure S2.16 – Numerical simulations of the assembly profile of a theoretical four-component system (4c system).** In contrast to a 3c system, a 4c system involves six different dimeric intermediates, four trimeric intermediates and one final tetrameric assembly. B). We simulated the assembly of a four-components nanosystem at different concentration of components B, C and D while varying the concentration of component A from  $10^{-11}$  M to  $10^{-6}$  M. The assembly profile of a 4c system display a similar tendency than a 3c system. C) The variation of the  $[A]_{50\%}$ , however, is much steeper at a lower concentration than a 3c system (green curve), or a 2c system (blue curve). Furthermore, the dynamic range varies from 9-fold to 6561-fold (circles) instead of 9-fold to 729-fold and 9-fold to 81-fold for the 3c (green curve) and 2c system (blue curve), respectively. Of note, at large excess of component A, the overall yield of the 4c system decreases like observed with the 3c system. Numerical simulations are done using MATLAB<sup>®</sup> with a dimeric stability of -8 kcal/mol ( $\Delta G_{\text{Dim}}$ ), a trimeric stability of -16 kcal/mol ( $\Delta G_{\text{Tri}}$ ), a tetrameric stability of -20 kcal/mol ( $\Delta G_{\text{Tetra}}$ ) and an overall assembly stability of -44 kcal/mol ( $\Delta G_{\text{Ass}} = \Delta G_{\text{Dim}} + \Delta G_{\text{Tri}} + \Delta G_{\text{Tetra}}$ ).

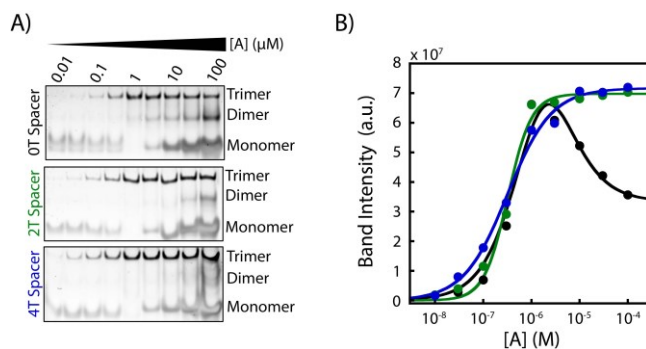




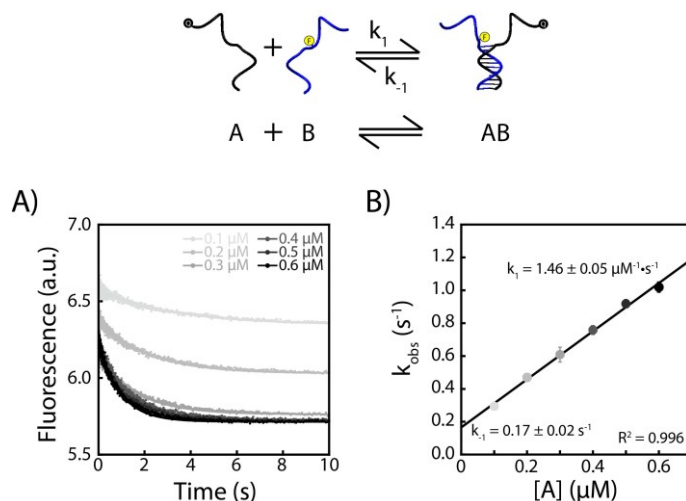
**Figure S2.17 – Effect of the thymine spacer length between each arm on the 3-way junction stability.** A) Melting curve analysis of the dimeric equilibrium reveals that, as expected, the spacer length (0T, 2T, and 4T) does not significantly change the dimeric affinity ( $\Delta G_{\text{Dim}}$ ). B) Melting curve analysis of the trimeric equilibrium reveals a strong dependence of the spacer length on the trimeric affinity ( $\Delta G_{\text{Tri}}$ ). \*Of note, to measure the trimeric affinity,  $\Delta G_{\text{Tri}}$ , and to exclude the dimeric energy, we “locked” the dimer into a unimolecular system using a 4T loop (black loop at the bottom of the blue-green arm). The DNA-hairpin, therefore, mimics the dimeric structure. Inset of A) and B) represent the derivative of the melting curve ( $dF/dT$ ) fitted with a Gaussian distribution, which has been used to extract the  $T_m$  that are then fitted using Eq. 2.4 to extract thermodynamic parameters. All experiments have been done in PBS buffer (50 mM  $\text{Na}_2\text{HPO}_4$ , 400 mM NaCl, pH = 7.00).



**Figure S2.18 – Numerical simulations of the effect of destabilizing only  $\Delta G_{\text{Tri}}$  on the assembly of a 3-component system.** When  $\Delta G_{\text{Tri}}$  and  $\Delta G_{\text{Dim}}$  are similar ( $\Delta\Delta G_{\text{Tri-Dim}} = 0 \text{ kcal}\cdot\text{mol}^{-1}$ ) the assembly of the 3-component system is not favoured (lightest gray). Increasing the value of  $\Delta\Delta G_{\text{Tri-Dim}}$  leads to a better yield of assembly of the 3-component system. However, when component A becomes more concentrated than the dissociation constant of the dimeric intermediates ( $K_{\text{Dim}}$ ), the assembly of the 3-component system becomes less favourable and the dimeric intermediates AB and AC form instead. A higher  $\Delta\Delta G_{\text{Tri-Dim}}$  enables to minimize this effect. Numerical simulations are done using MATLAB®. All  $\Delta G_{\text{Dim}}$  are fixed at  $-8 \text{ kcal}\cdot\text{mol}^{-1}$  and  $\Delta G_{\text{Tri}}$  is varied between  $-8 \text{ kcal/mol}$  and  $-16 \text{ kcal}\cdot\text{mol}^{-1}$ . Temperature is fixed at  $37^\circ\text{C}$  (310.15 K), the concentration of components B and C are fixed at 100 nM and the concentration of component A is varied from 0.1 nM to 100  $\mu\text{M}$ .

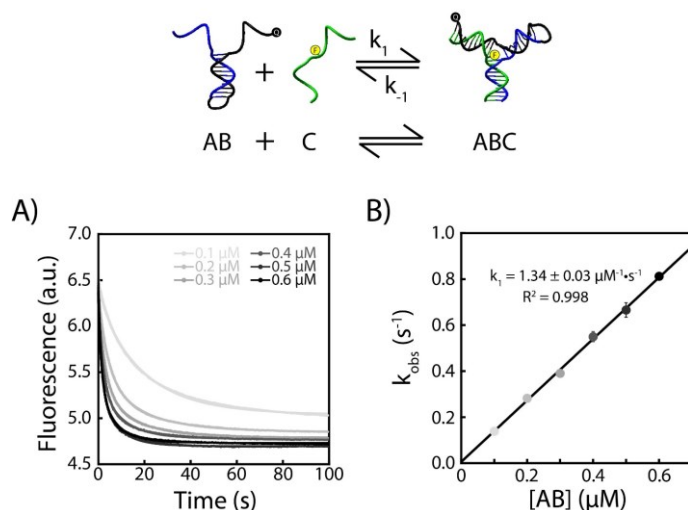


**Figure S2.19 – Native PAGE titration.** A-B) Titrations analyzed by native polyacrylamide gel electrophoresis (native PAGE) support the dimers sequestration mechanism. Like our fluorescence titration experiments (Fig. 2.3B, *right*), the 2T system displays more cooperativity ( $DR = 12 \pm 5$ ) than the 4T system ( $DR = 100 \pm 47$ ). Furthermore, the 0T system (black) shows a decrease in assembly when the concentration of component A is larger than the concentration of components B and C ( $1 \mu\text{M}$ ) (see the reduction of trimer band and increase of dimer band, dimeric intermediates AB and AC). This phenomenon is also observed for the 2T and 4T systems, although requiring a much higher concentration of component A. This result is in good agreement with our numerical simulations (see Fig. S2.18) Black = 0T spacer, Green = 2T spacer and Blue = 4T spacer. \*Of note, this experiment has been done at room temperature ( $\sim 23^\circ\text{C}$ ) whereas the fluorescence titration has been done at  $37^\circ\text{C}$ . Also, because PAGE is a less sensitive technique compared to fluorescence titration, this experiment has been performed at  $1 \mu\text{M}$  of components B and C compared to the  $100 \text{ nM}$  used in fluorescence titration. Nonetheless, both experiments demonstrate the same tendencies and are in good agreement with our numerical simulations. Binding curves of 2T and 4T spacer are fitted using the Hill equation (Eq. 2.5) while the binding curve of the 0T spacer is fitted using a double dose-response curve.

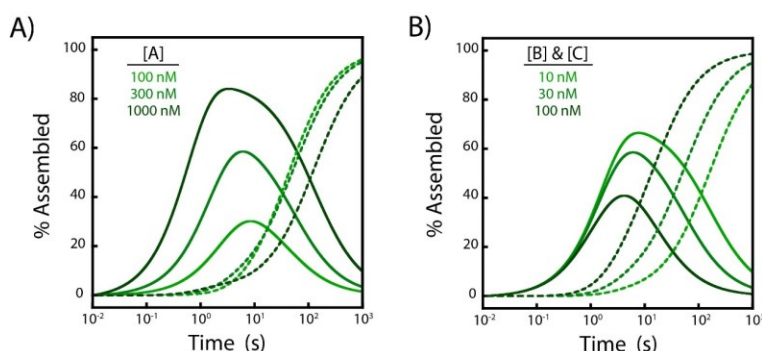


**Figure S2.20 – Rate constant measurement of dimer association used for the numerical simulations.** A) Raw kinetic traces of dimer association using  $0.01 \mu\text{M}$  of component C with a least a 10-fold excess of component A ( $0.1$  to  $0.6 \mu\text{M}$ ). B) The association rate constant ( $k_1$ ) is found to be  $1.46 \pm 0.05 \mu\text{M}^{-1}\cdot\text{s}^{-1}$  and the dissociation rate constant ( $k^{-1}$ ) is found to be  $0.17 \pm 0.02 \text{ s}^{-1}$ . Data on panel A are fitted using a single exponential (Eq. 2.6) All experiments are done in triplicate at  $37^\circ\text{C}$  in PBS buffer ( $50 \text{ mM Na}_2\text{HPO}_4$ ,  $400 \text{ mM NaCl}$ ,  $\text{pH} = 7.00$ ).

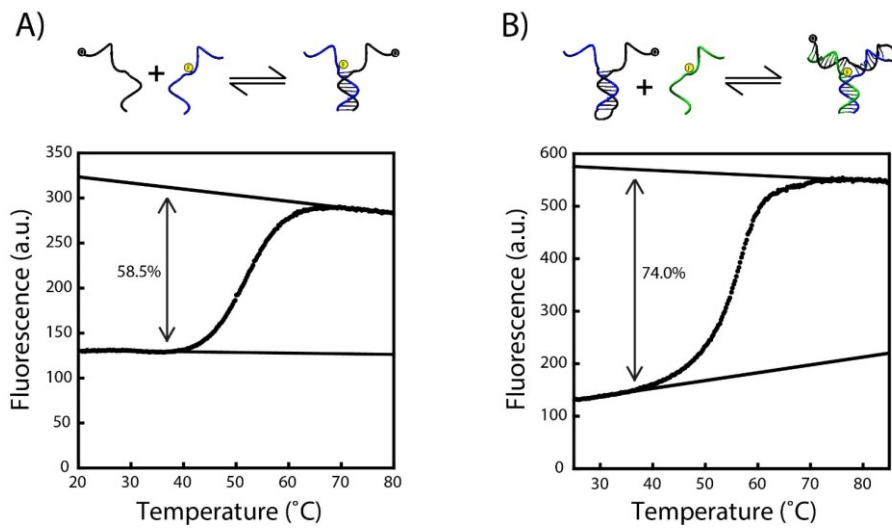




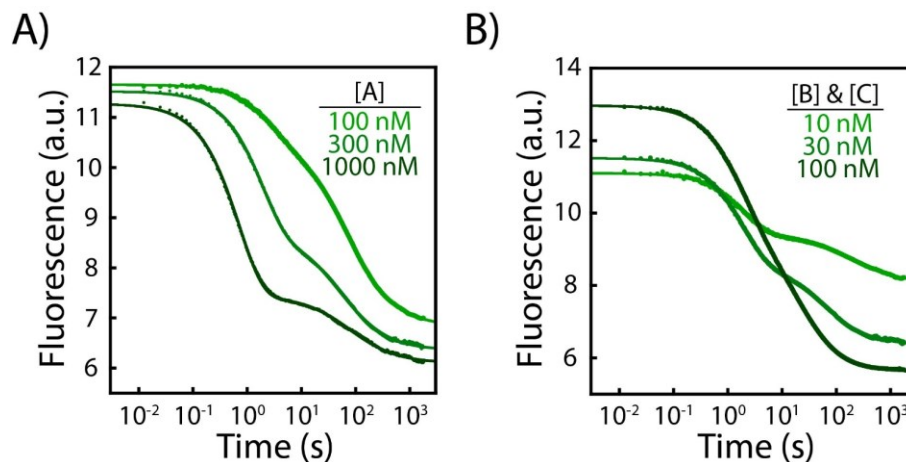
**Figure S2.21 – Rate constant measurement of the binding of the third component on a preorganized dimer used for the numerical simulations.** A) Raw kinetic traces of trimer association using 0.01  $\mu\text{M}$  of component B with a least a 10-fold excess of preorganized dimer AC (0.1 to 0.6  $\mu\text{M}$ ). B) The association rate constant ( $k_1$ ) is found to be  $1.34 \pm 0.01 \mu\text{M}^{-1}\cdot\text{s}^{-1}$  and is similar to the association rate constant of the dimer (see Fig. S2.20). The dissociation rate constant is too small to be extrapolated precisely from this experiment and is thus estimated from the  $K_D$  ( $\Delta G = -18.3 \text{ kcal/mol}$  at  $37^\circ\text{C}$ ) (see Fig. S2.17). The dissociation rate constant ( $k_{-1}$ ) is estimated to be  $1.70 \times 10^{-7} \text{ s}^{-1}$ . Data on panel A are fitted using a single exponential (Eq. 2.6). All experiments are done in triplicate at  $37^\circ\text{C}$  in PBS buffer (50 mM  $\text{Na}_2\text{HPO}_4$ , 400 mM NaCl, pH = 7.00).



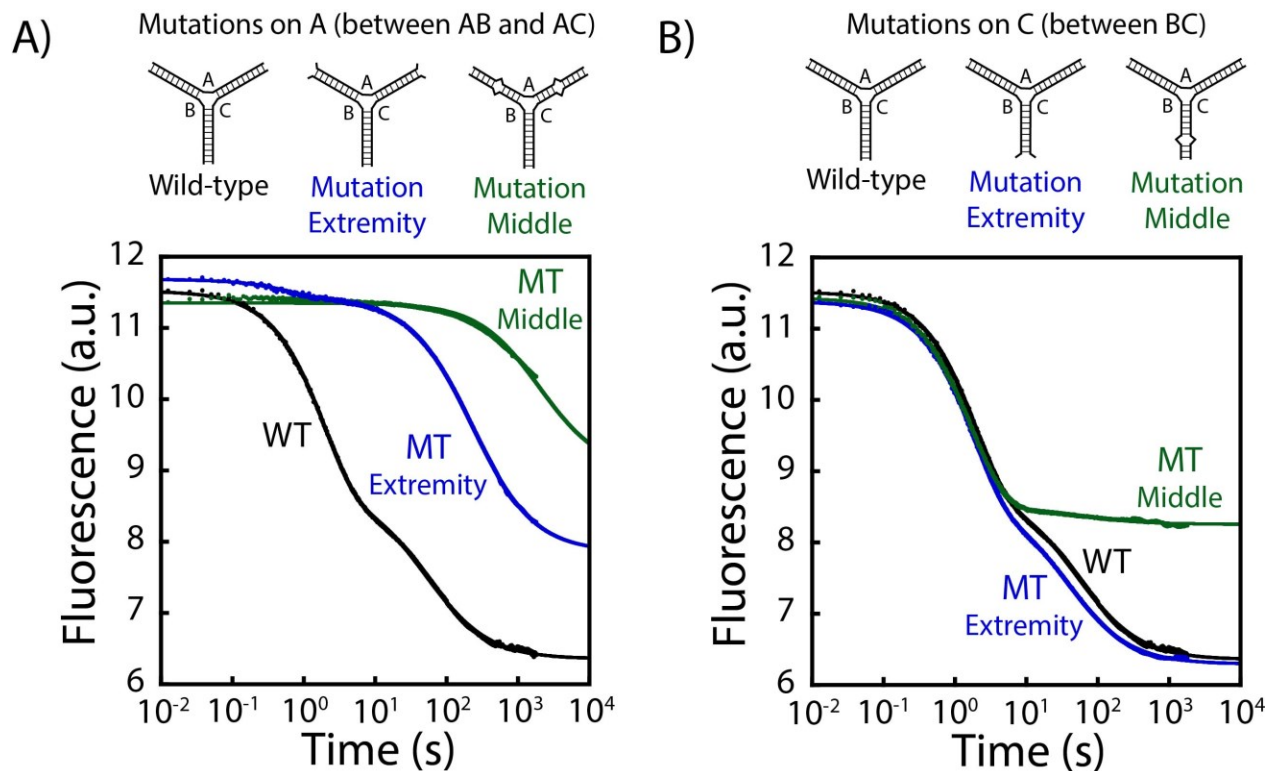
**Figure S2.22 – Numerical simulation of the time-dependent assembly of a 3c system.** Using the rate constants measured from Fig. S2.20 and Fig. S2.21, we can predict the kinetic traces of assembly of the dimer (plain line) and the trimer (dashed line) using numerical simulation. This enables us to test different concentration conditions before experimentation. A) Changing [A] only affects dimer assembly while B) changing [B] and [C] affects the formation of the trimer. This illustrates that the limiting step of trimer assembly, in these conditions, is the formation of the productive dimer BC. These results agree with our experimental data (Fig. 2.4B in main text and Fig. S2.24 and Fig. S2.25). Numerical simulations are performed using MATLAB®.



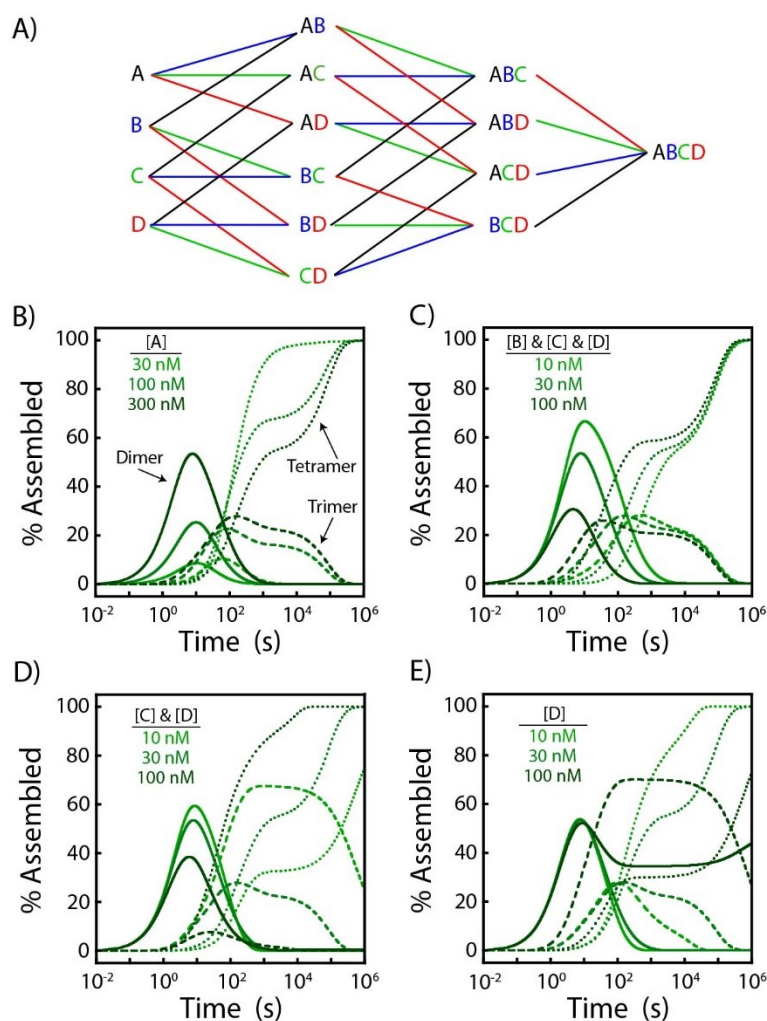
**Figure S2.23 – Estimation of the quenching efficiency at 37°C of the dimer and trimer by melting curve analysis.**



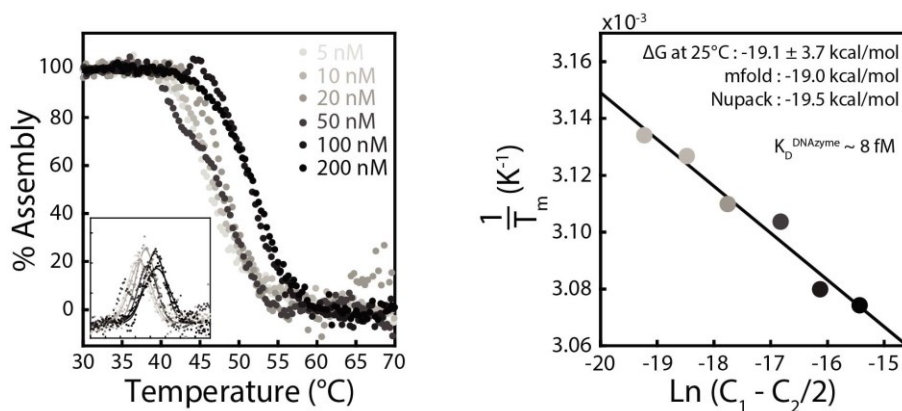
**Figure S2.24 – Raw kinetic traces of time-dependent programmed assembly (Fig. 2.4B).** A) Kinetic traces with [B] and [C] fixed at 30 nM. B) Kinetic traces with [A] at 300 nM. In each case, [A] is kept in excess (more than 3-fold). Data are fitted using a combination of a pseudo-first-order reaction (fast phase) and a second-order reaction (slow phase) (Eq. 2.7). Knowing the amplitude of each phase, the efficiency of quenching (see Fig. S2.23) and the percentage of trimeric assembly (see Fig. 2.2), the data are converted from fluorescence to concentration. Experiments are performed in PBS buffer (50mM  $\text{Na}_2\text{HPO}_4$ , 400 mM NaCl, pH=7.00) at 37°C.



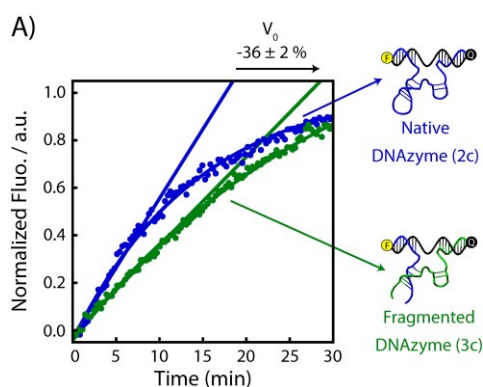
**Figure S2.25 – Mutational kinetic analysis.** In an excess of component A, trimer formation proceeds through a slow strand exchange mechanism limited by the association of the sequestered components B and C. According to “the rule of seven” in Watson-Crick base-pairing<sup>29</sup>, a mutation at the extremity of a 10bp duplex should not affect its kinetics of assembly but should affect its kinetics of disassembly and therefore its affinity. In contrast, an internal mutation is expected to drastically affect both its kinetics of assembly and its affinity. A) A mutation on component A at the extremity drastically reduces the amount of the dimeric intermediates AB and AC and reduces the amount of formed trimer. However, the trimer assembly still proceeds at a similar rate than the wild-type (5-fold slower). A mutation in the middle of component A prevent the formation of the dimeric intermediates AB and AC, but still enables the formation of the productive dimer BC, therefore, leading to some trimer being formed at a much slower rate (40-fold slower). B) Mutation between the duplex B and C at the extremity does not affect the trimeric assembly. However, a mutation inside this duplex renders the assembly of the trimer impossible indicating that the formation of the productive dimer BC is key for driving trimer assembly in an excess of component A.



**Figure S2.26 – Numerical simulations of the kinetic assembly pathways of a theoretical four-component system (4c system).** Using similar rate constants than the trimeric assembly simulation (Fig. S2.22), we simulated the assembly of a four components nanosystem at 30 nM of components B, C and D with a 10-fold excess of component A of 300 nM. In contrast with a 3c system, a 4c system involves six different dimeric intermediate, four trimeric intermediates and one final trimeric assembly. B) An excess of component A (from 30 to 300 nM) leads to a two-step assembly profile involving dimeric and trimeric traps. C) Increasing the concentration of B, C and D simultaneously leads to a faster first step of assembly (related to the dimeric trap) while not affecting the slower second step of assembly (related to the trimeric trap). D) Increasing the concentration of components C and D only disfavours the trimeric intermediates and leads to a faster, almost one-step assembly. E) Increasing only the component D leads to a higher number of trimeric intermediates thus leading to a slower two-step assembly. At large, a four-component nanosystem enables the programming of diverse assembly pathways but also can lead to harmful intermediates that slow down considerably the assembly.

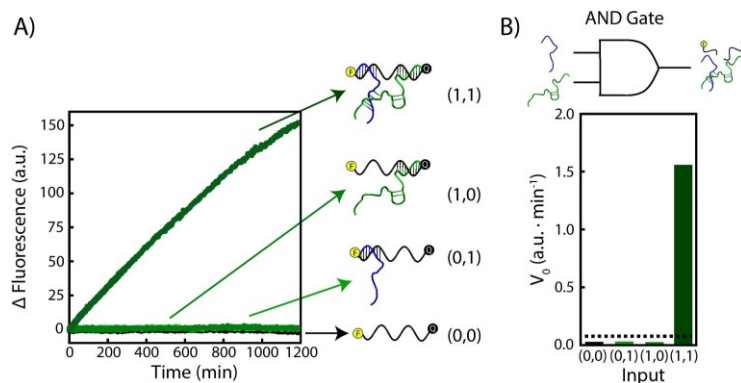


**Figure S2.27 – Estimation of the  $K_D$  between the DNAzyme (NaA43) and a non-cleavable substrate.** Melting curve analysis using 1 nM of the fluorescent substrate and an excess of DNAzyme (5 to 200 nM) reveals that the  $\Delta G$  is  $-19.1 \pm 3.7$  kcal/mol which gives a  $K_D$  in the fM range ( $\sim 8$  fM). This value is within the experimental error of the value expected from mfold and Nupack. The inset represents the derivative of the melting curve ( $dF/dT$ ) fitted with a Gaussian distribution, which has been used to extract the  $T_m$  that is then fitted using Eq. 2.4 to extract thermodynamic parameters. All experiments have been done in PBS buffer (50 mM  $\text{Na}_2\text{HPO}_4$ , 400 mM NaCl, pH = 7.00).

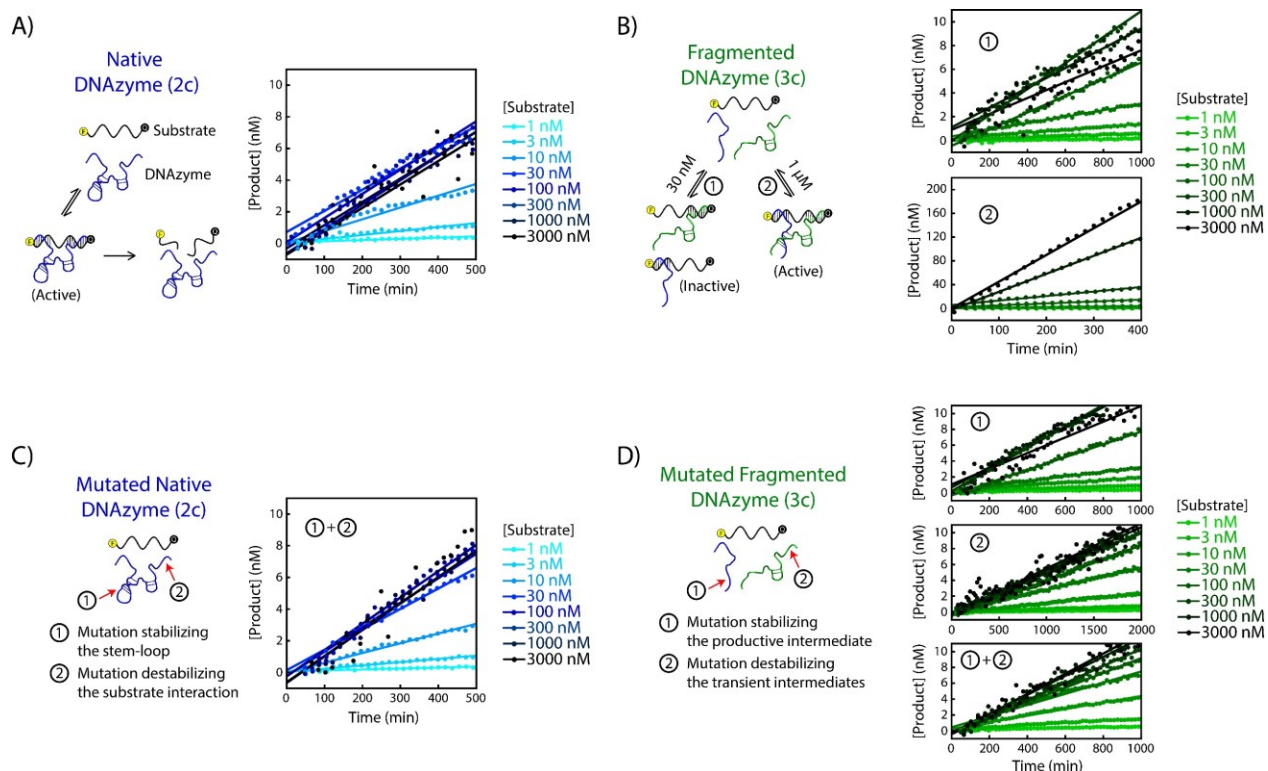


**Figure S2.28 – Activity of the cut DNAzyme.** To explore the effect of the number of components on the regulation of a DNAzyme, we engineered a dimeric DNAzyme by fragmenting the loop far from the catalytic site, thus creating a three-component system (two DNAzyme components plus the substrate). This modification still provides a functional DNAzyme despite a  $36 \pm 2\%$  reduction in catalytic activity (initial rate,  $V_0$ ). Both experiments have been done in PBS buffer (50 mM  $\text{Na}_2\text{HPO}_4$ , 400 mM NaCl, pH = 7.00) at  $25^\circ\text{C}$  using 30 nM of DNAzyme (native and fragmented) with a 10-fold excess of substrates (300 nM).

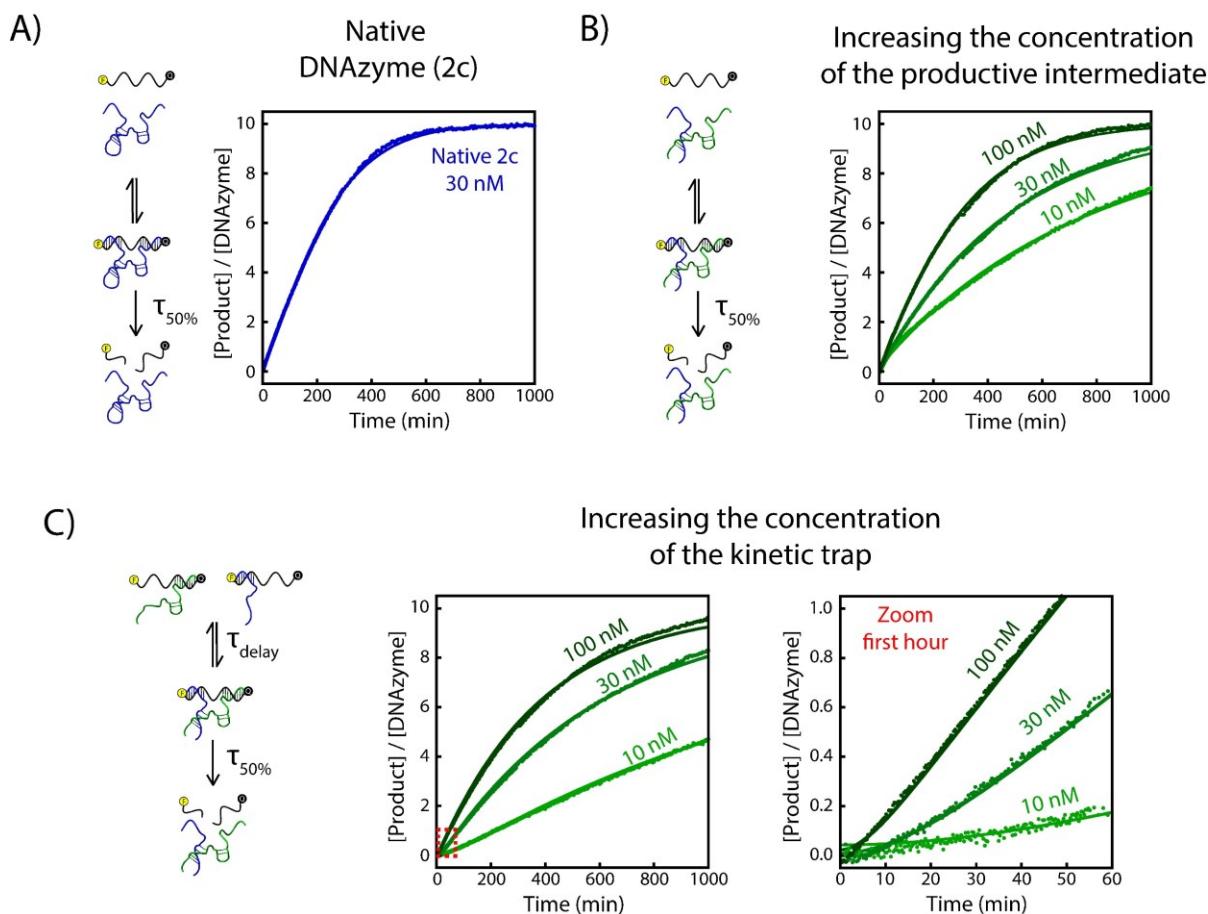




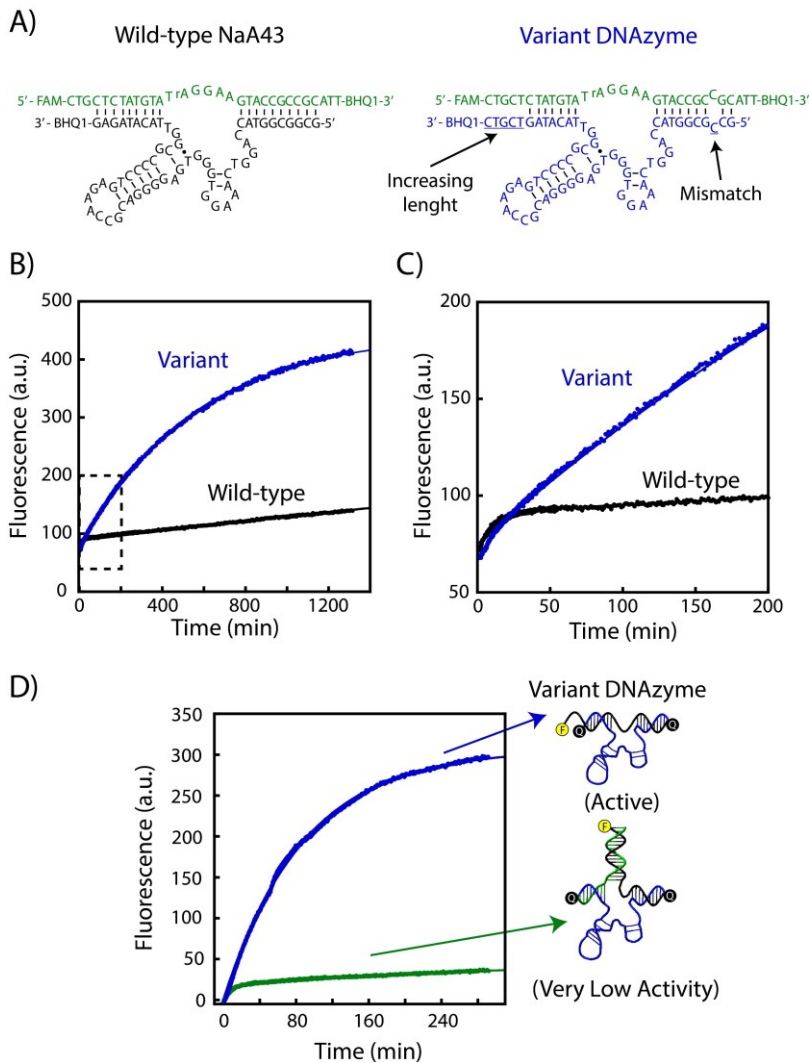
**Figure S2.29 – Production of an AND gate due to the inactivity of the sequestered dimers.** A) Kinetic traces of the cut DNAzyme system (10 nM) with a 10-fold excess of substrates (100 nM). B) A logic response is obtained where the presence of both fragmented components of the DNAzyme is required to retrieve the native activity. Experiments have been done in PBS buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 400 mM NaCl, pH = 7.00) at 25°C using the double mutant system (one destabilizing mutation in the arm and one stabilizing mutation in the stem-loop section).



**Figure S2.30 – Raw kinetic traces of A) Fig. 2.5A, left B) Fig. 2.5A, right C) Fig. 2.5B, left and D) Fig. 2.5B, right.** From these curves, we extract the initial rate ( $V_0$ ) which is then plotted against the concentration of substrates used for each experiment. Fluorescence data were converted to concentration values using a calibration curve (data not shown). All experiments have been done in PBS buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 400 mM NaCl, pH = 7.00) at 25°C.



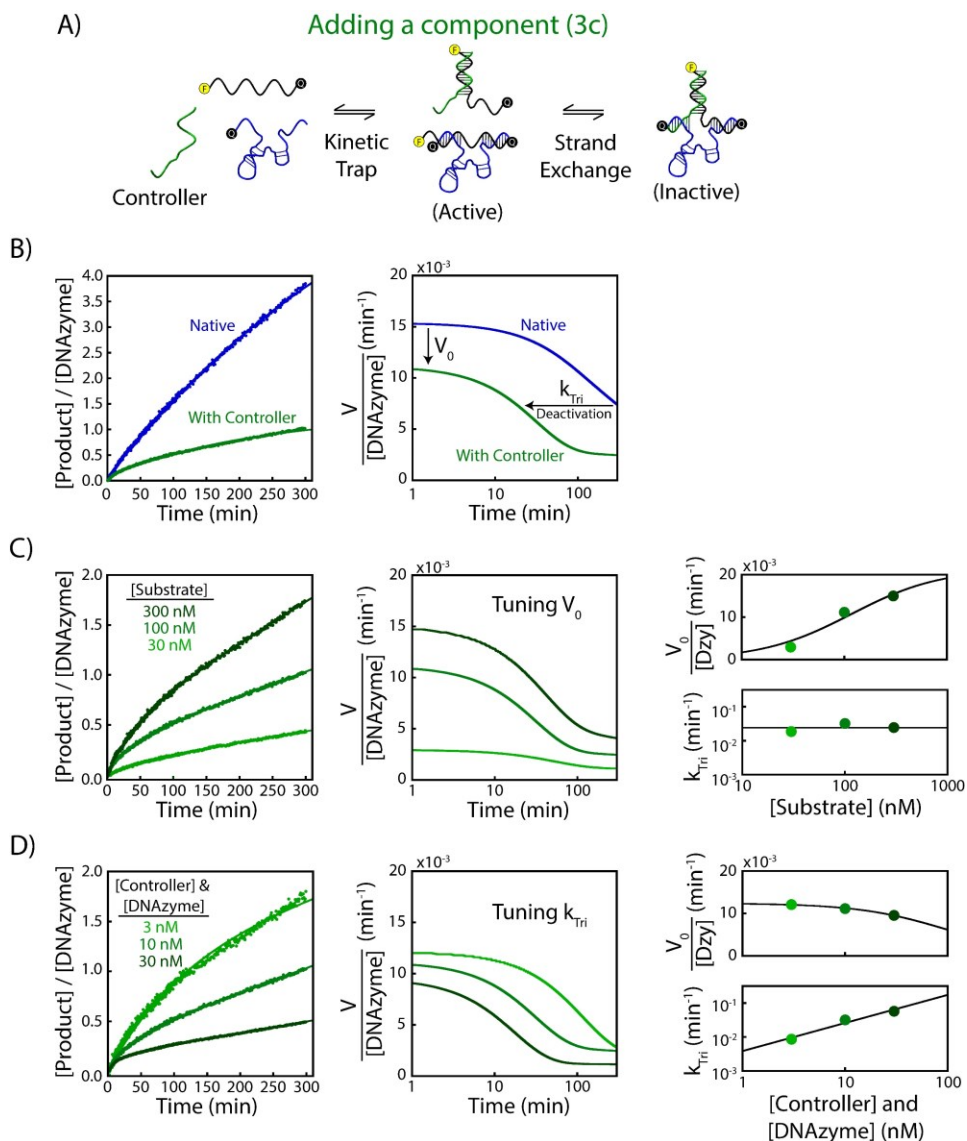
**Figure S2.31 – Raw kinetic traces of A) Fig. 2.5C, left and B) Fig. 2.5C, right.** Kinetic traces of the product generation in panels A and B are fitted using an exponential function (Eq. 2.6) while the kinetic traces of the product generation in panel C are fitted using a delayed exponential function<sup>90</sup>. All kinetic traces are normalized by the concentration of DNAzyme components used in the experiment. The derivatives of these fitting (*i.e.*, the rate of the reaction) are used in the main figure. Fluorescence data were converted to concentration values using a calibration curve (data not shown). All experiments are done at the concentration of DNAzyme components indicated in the figure with a 10-fold excess of substrates (100, 300 or 1000 nM). All experiments have been done in PBS buffer (50 mM  $Na_2HPO_4$ , 400 mM NaCl, pH = 7.00) at 25°C.



**Figure S2.32 – Design of a variant DNAzyme for the implementation of the controller strand.**

A) The native NaA43 DNAzyme contains a very stable 8 GC base pairs arm (the right one) that slows down the dissociation of the product. We thus introduced a mismatch on the third nucleotide of the native DNAzyme starting from the 5' extremity to increase the dissociation rate constant<sup>29</sup>. We also increased the length of the left arm to provide a stronger binding site for the controller strand. B-C) Introducing a mismatch does not change the initial rate of the reaction but prevents the DNAzyme from being poisoned by the stable product. Therefore, the variant DNAzyme is not rate-limited by the dissociation of the product. D) The formation of the stable DNAzyme-substrate-controller complex leads to an inactive trimer. Data are fitted with a double exponential equation (Eq. 2.6). Experiments were done in PBS buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 400 mM NaCl, pH = 7.00) at 25°C. Controller sequence: 5'- GAC GAC TAT GTA TGT TAC ATA GAG CAG -3'.





**Figure S2.33 – Adding a third component.** A) A trimeric assembly can also be created by introducing a new component, called the controller, that can interact with both the native DNAzyme and the substrate. B) This controller strand can modulate the level of activity and the deactivation time via the formation of an inactive trimer. C) In the presence of a 10 nM controller and DNAzyme, increasing the concentration of substrate increases the formation of the active dimer and the catalytic rate without significantly affecting the trimeric deactivation rate  $k_{\text{Tri}}$ . D) At 100 nM substrate, increasing the concentration of controller and DNAzyme increases the rate of trimer formation and thus the rate of DNAzyme deactivation. Kinetic traces (left panel) are fitted using a single exponential (native) or a double exponential (with controller) to extract rate constants of dimeric and trimeric formation (*i.e.*, deactivation of the DNAzyme) (Eq. 2.6). The derivatives (middle panel) are plotted to better show the variation in initial rates and deactivation times. Fluorescence data were converted to the concentration of product using a calibration curve (data not shown) and normalized by the concentration of DNAzyme. All experiments have been done in PBS buffer (50 mM  $\text{Na}_2\text{HPO}_4$ , 400 mM NaCl, pH = 7.00) at 25°C.

## 2.8 References

1. Grzybowski, B. A.; Huck, W. T. S., The nanotechnology of life-inspired systems. *Nat. Nanotechnol.* **2016**, *11*, 585.
2. Murphy, K. G.; Bloom, S. R., Gut hormones and the regulation of energy homeostasis. *Nature* **2006**, *444* (7121), 854-859.
3. Cameron, D. E.; Bashor, C. J.; Collins, J. J., A brief history of synthetic biology. *Nat. Rev. Microbiol.* **2014**, *12*, 381.
4. Benner, S. A.; Sismour, A. M., Synthetic biology. *Nat. Rev. Genet.* **2005**, *6*, 533.
5. Camilli, A.; Bassler, B. L., Bacterial Small-Molecule Signaling Pathways. *Science* **2006**, *311* (5764), 1113.
6. Dill, K. A.; MacCallum, J. L., The Protein-Folding Problem, 50 Years On. *Science* **2012**, *338* (6110), 1042.
7. Marsh, J. A.; Teichmann, S. A., Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **2015**, *84* (1), 551-575.
8. Lynch, M., The evolution of multimeric protein assemblages. *Mol. Biol. Evol.* **2012**, *29* (5), 1353-1366.
9. Thomas, C.; Lumb, A. B., Physiology of haemoglobin. *Contin. Educ. Anaesth. Crit. Care Pain.* **2012**, *12* (5), 251-256.
10. Pillai, A. S.; Chandler, S. A.; Liu, Y.; Signore, A. V.; Cortez-Romero, C. R.; Benesch, J. L. P.; Laganowsky, A.; Storz, J. F.; Hochberg, G. K. A.; Thornton, J. W., Origin of complexity in haemoglobin evolution. *Nature* **2020**, *581* (7809), 480-485.
11. Erbas-Cakmak, S.; Kolemen, S.; Sedgwick, A. C.; Gunnlaugsson, T.; James, T. D.; Yoon, J.; Akkaya, E. U., Molecular logic gates: the past, present and future. *Chem. Soc. Rev.* **2018**, *47* (7), 2228-2248.
12. Galés, C.; Rebois, R. V.; Hogue, M.; Trieu, P.; Breit, A.; Hébert, T. E.; Bouvier, M., Real-time monitoring of receptor and G-protein interactions in living cells. *Nat. Methods* **2005**, *2* (3), 177-184.
13. Pakulska, M. M.; Miersch, S.; Shoichet, M. S., Designer protein delivery: From natural to engineered affinity-controlled release systems. *Science* **2016**, *351* (6279), aac4750.

14. Kinch, L. N.; Grishin, N. V., Evolution of protein structures and functions. *Curr. Opin. Struct. Biol.* **2002**, *12* (3), 400-408.
15. Nooren, I. M. A.; Thornton, J. M., Diversity of protein–protein interactions. *EMBO J.* **2003**, *22* (14), 3486-3492.
16. Plaxco, K. W.; Simons, K. T.; Ruczinski, I.; Baker, D., Topology, Stability, Sequence, and Length: Defining the Determinants of Two-State Protein Folding Kinetics. *Biochemistry* **2000**, *39* (37), 11177-11183.
17. Grantcharova, V.; Alm, E. J.; Baker, D.; Horwich, A. L., Mechanisms of protein folding. *Curr. Opin. Struct. Biol.* **2001**, *11* (1), 70-82.
18. Marianayagam, N. J.; Sunde, M.; Matthews, J. M., The power of two: protein dimerization in biology. *Trends Biochem. Sci.* **2004**, *29* (11), 618-625.
19. Whitty, A., Cooperativity and biological complexity. *Nat. Chem. Biol.* **2008**, *4* (8), 435-439.
20. Williamson, J. R., Cooperativity in macromolecular assembly. *Nat. Chem. Biol.* **2008**, *4* (8), 458-465.
21. Mattia, E.; Otto, S., Supramolecular systems chemistry. *Nat. Nanotechnol.* **2015**, *10* (2), 111-119.
22. Seeman, N. C.; Sleiman, H. F., DNA nanotechnology. *Nat. Rev. Mater.* **2017**, *3* (1), 17068.
23. Li, M.; Zheng, M.; Wu, S.; Tian, C.; Liu, D.; Weizmann, Y.; Jiang, W.; Wang, G.; Mao, C., In vivo production of RNA nanostructures via programmed folding of single-stranded RNAs. *Nat. Commun.* **2018**, *9* (1), 2196.
24. Gradišar, H.; Božič, S.; Doles, T.; Vengust, D.; Hafner-Bratkovič, I.; Mertelj, A.; Webb, B.; Šali, A.; Klavžar, S.; Jerala, R., Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nat. Chem. Biol.* **2013**, *9* (6), 362-366.
25. Shekhawat, S. S.; Ghosh, I., Split-protein systems: beyond binary protein–protein interactions. *Curr. Opin. Chem. Biol.* **2011**, *15* (6), 789-797.
26. Debais, M.; Lelievre, A.; Smietana, M.; Müller, S., Splitting aptamers and nucleic acid enzymes for the development of advanced biosensors. *Nucleic Acids Res.* **2020**, *48* (7), 3400-3422.

27. Dolberg, T. B.; Meger, A. T.; Boucher, J. D.; Corcoran, W. K.; Schauer, E. E.; Prybutok, A. N.; Raman, S.; Leonard, J. N., Computation-guided optimization of split protein systems. *Nat. Chem. Biol.* **2021**, *17* (5), 531-539.
28. Michnick, S. W.; Ear, P. H.; Manderson, E. N.; Remy, I.; Stefan, E., Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nat. Rev. Drug Discovery* **2007**, *6* (7), 569-582.
29. Cisse, I. I.; Kim, H.; Ha, T., A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nat. Struct. Mol. Biol.* **2012**, *19* (6), 623-627.
30. Vallée-Bélisle, A.; Ricci, F.; Plaxco, K. W., Thermodynamic basis for the optimization of binding-induced biomolecular switches and structure-switching biosensors. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (33), 13802.
31. Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J., The ensemble nature of allostery. *Nature* **2014**, *508* (7496), 331-339.
32. Madsen, M.; Gothelf, K. V., Chemistries for DNA Nanotechnology. *Chem. Rev.* **2019**, *119* (10), 6384-6458.
33. Sadowski, J. P.; Calvert, C. R.; Zhang, D. Y.; Pierce, N. A.; Yin, P., Developmental Self-Assembly of a DNA Tetrahedron. *ACS Nano* **2014**, *8* (4), 3251-3259.
34. Hao, C.; Li, X.; Tian, C.; Jiang, W.; Wang, G.; Mao, C., Construction of RNA nanocages by re-engineering the packaging RNA of Phi29 bacteriophage. *Nat. Commun.* **2014**, *5*, 3890.
35. Silverman, S. K., Catalytic DNA: Scope, Applications, and Biochemistry of Deoxyribozymes. *Trends Biochem. Sci.* **2016**, *41* (7), 595-609.
36. Peña, M. d. I.; Dufour, D.; Gallego, J., Three-way RNA junctions with remote tertiary contacts: A recurrent and highly versatile fold. *RNA* **2009**, *15* (11), 1949-1964.
37. Zuker, M., Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, *31* (13), 3406-3415.
38. Idili, A.; Ricci, F.; Vallée-Bélisle, A., Determining the folding and binding free energy of DNA-based nanodevices and nanoswitches using urea titration curves. *Nucleic Acids Res.* **2017**, *45* (13), 7571-7580.

39. Lawrence, C.; Vallée-Bélisle, A.; Pfeil, S. H.; de Mornay, D.; Lipman, E. A.; Plaxco, K. W., A comparison of the folding kinetics of a small, artificially selected DNA aptamer with those of equivalently simple naturally occurring proteins. *Protein Sci.* **2014**, *23* (1), 56-66.
40. Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A., NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **2011**, *32* (1), 170-173.
41. Jackson, S. E.; Fersht, A. R., Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **1991**, *30* (43), 10428-10435.
42. Maxwell, K. L.; Wildes, D.; Zarrine-Afsar, A.; De Los Rios, M. A.; Brown, A. G.; Friel, C. T.; Hedberg, L.; Horng, J.-C.; Bona, D.; Miller, E. J.; Vallée-Bélisle, A.; Main, E. R. G.; Bemporad, F.; Qiu, L.; Teilum, K.; Vu, N.-D.; Edwards, A. M.; Ruczinski, I.; Poulsen, F. M.; Kragelund, B. B.; Michnick, S. W.; Chiti, F.; Bai, Y.; Hagen, S. J.; Serrano, L.; Oliveberg, M.; Raleigh, D. P.; Wittung-Stafshede, P.; Radford, S. E.; Jackson, S. E.; Sosnick, T. R.; Marqusee, S.; Davidson, A. R.; Plaxco, K. W., Protein folding: Defining a “standard” set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci.* **2005**, *14* (3), 602-616.
43. Plaxco, K. W.; Simons, K. T.; Baker, D., Contact order, transition state placement and the refolding rates of single domain proteins<sup>11</sup>Edited by P. E. Wright. *J. Mol. Biol.* **1998**, *277* (4), 985-994.
44. Sanchez-Ruiz, J. M., Protein kinetic stability. *Biophys. Chem.* **2010**, *148* (1), 1-15.
45. Baskakov, I. V.; Legname, G.; Prusiner, S. B.; Cohen, F. E., Folding of Prion Protein to Its Native  $\alpha$ -Helical Conformation Is under Kinetic Control\*. *J. Biol. Chem.* **2001**, *276* (23), 19687-19690.
46. Thommen, M.; Holtkamp, W.; Rodnina, M. V., Co-translational protein folding: progress and methods. *Curr. Opin. Struct. Biol.* **2017**, *42*, 83-89.
47. Waudby, C. A.; Dobson, C. M.; Christodoulou, J., Nature and Regulation of Protein Folding on the Ribosome. *Trends Biochem. Sci.* **2019**, *44* (11), 914-926.
48. Wells, J. A.; McClendon, C. L., Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature* **2007**, *450* (7172), 1001-1009.
49. Hulme, E. C.; Trevethick, M. A., Ligand binding assays at equilibrium: validation and interpretation. *Br. J. Pharmacol.* **2010**, *161* (6), 1219-1237.

50. Tian, L.; Heyduk, T., Bivalent Ligands with Long Nanometer-Scale Flexible Linkers. *Biochemistry* **2009**, *48* (2), 264-275.
51. Lescoute, A.; Westhof, E., Topology of three-way junctions in folded RNAs. *RNA* **2006**, *12* (1), 83-93.
52. Ercolani, G.; Schiaffino, L., Allosteric, Chelate, and Interannular Cooperativity: A Mise au Point. *Angew. Chem. Int. Ed.* **2011**, *50* (8), 1762-1768.
53. Korevaar, P. A.; George, S. J.; Markvoort, A. J.; Smulders, M. M. J.; Hilbers, P. A. J.; Schenning, A. P. H. J.; De Greef, T. F. A.; Meijer, E. W., Pathway complexity in supramolecular polymerization. *Nature* **2012**, *481*, 492.
54. Wang, J.; Liu, K.; Xing, R.; Yan, X., Peptide self-assembly: thermodynamics and kinetics. *Chem. Soc. Rev.* **2016**, *45* (20), 5589-5604.
55. Li, G.; Zhang, X. C., GTP Hydrolysis Mechanism of Ras-like GTPases. *J. Mol. Biol.* **2004**, *340* (5), 921-932.
56. Lu, K. P.; Finn, G.; Lee, T. H.; Nicholson, L. K., Prolyl cis-trans isomerization as a molecular timer. *Nat. Chem. Biol.* **2007**, *3* (10), 619-629.
57. Hou, W.-S.; Van Parijs, L., A Bcl-2-dependent molecular timer regulates the lifespan and immunogenicity of dendritic cells. *Nat. Immunol.* **2004**, *5* (6), 583-589.
58. Torabi, S.-F.; Wu, P.; McGhee, C. E.; Chen, L.; Hwang, K.; Zheng, N.; Cheng, J.; Lu, Y., In vitro selection of a sodium-specific DNAzyme and its application in intracellular sensing. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (19), 5903-5908.
59. Maity, T. S.; Weeks, K. M., A Threefold RNA-Protein Interface in the Signal Recognition Particle Gates Native Complex Assembly. *J. Mol. Biol.* **2007**, *369* (2), 512-524.
60. Yi, J.; Balagopalan, L.; Nguyen, T.; McIntire, K. M.; Samelson, L. E., TCR microclusters form spatially segregated domains and sequentially assemble in calcium-dependent kinetic steps. *Nat. Commun.* **2019**, *10* (1), 277.
61. Hunter, C. A.; Anderson, H. L., What is Cooperativity? *Angew. Chem. Int. Ed.* **2009**, *48* (41), 7488-7499.
62. Robustelli, P.; Piana, S.; Shaw, D. E., Mechanism of Coupled Folding-upon-Binding of an Intrinsically Disordered Protein. *J. Am. Chem. Soc.* **2020**, *142* (25), 11092-11101.

63. Chen, A.; Yan, M.; Yang, S., Split aptamers and their applications in sandwich aptasensors. *TrAC, Trends Anal. Chem.* **2016**, *80*, 581-593.
64. Müller, J.; Johnsson, N., Split-Ubiquitin and the Split-Protein Sensors: Chessman for the Endgame. *ChemBioChem* **2008**, *9* (13), 2029-2038.
65. Shiba, K.; Schimmel, P., Functional assembly of a randomly cleaved protein. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (5), 1880.
66. Cabantous, S.; Nguyen, H. B.; Pedelacq, J.-D.; Koraïchi, F.; Chaudhary, A.; Ganguly, K.; Lockard, M. A.; Favre, G.; Terwilliger, T. C.; Waldo, G. S., A New Protein-Protein Interaction Sensor Based on Tripartite Split-GFP Association. *Sci. Rep.* **2013**, *3* (1), 2854.
67. Rohl, C. A.; Strauss, C. E. M.; Misura, K. M. S.; Baker, D., Protein Structure Prediction Using Rosetta. In *Methods in Enzymology*, Academic Press: 2004; Vol. 383, pp 66-93.
68. Dagliyan, O.; Krokhotin, A.; Ozkan-Dagliyan, I.; Deiters, A.; Der, C. J.; Hahn, K. M.; Dokholyan, N. V., Computational design of chemogenetic and optogenetic split proteins. *Nat. Commun.* **2018**, *9* (1), 4042.
69. Ricci, F.; Vallée-Bélisle, A.; Porchetta, A.; Plaxco, K. W., Rational Design of Allosteric Inhibitors and Activators Using the Population-Shift Model: In Vitro Validation and Application to an Artificial Biosensor. *J. Am. Chem. Soc.* **2012**, *134* (37), 15177-15180.
70. Simon, A. J.; Vallée-Bélisle, A.; Ricci, F.; Plaxco, K. W., Intrinsic disorder as a generalizable strategy for the rational design of highly responsive, allosterically cooperative receptors. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (42), 15048-15053.
71. Simon, A. J.; Vallée-Bélisle, A.; Ricci, F.; Watkins, H. M.; Plaxco, K. W., Using the Population-Shift Mechanism to Rationally Introduce "Hill-type" Cooperativity into a Normally Non-Cooperative Receptor. *Angew. Chem. Int. Ed.* **2014**, *53* (36), 9471-9475.
72. Kang, D.; Vallée-Bélisle, A.; Porchetta, A.; Plaxco, K. W.; Ricci, F., Re-engineering Electrochemical Biosensors To Narrow or Extend Their Useful Dynamic Range. *Angew. Chem. Int. Ed.* **2012**, *51* (27), 6717-6721.
73. Porchetta, A.; Vallée-Bélisle, A.; Plaxco, K. W.; Ricci, F., Allosterically Tunable, DNA-Based Switches Triggered by Heavy Metals. *J. Am. Chem. Soc.* **2013**, *135* (36), 13238-13241.

74. Harroun, S. G.; Prévost-Tremblay, C.; Lauzon, D.; Desrosiers, A.; Wang, X.; Pedro, L.; Vallée-Bélisle, A., Programmable DNA switches and their applications. *Nanoscale* **2018**, *10* (10), 4607-4641.
75. Vallée-Bélisle, A.; Ricci, F.; Plaxco, K. W., Engineering Biosensors with Extended, Narrowed, or Arbitrarily Edited Dynamic Range. *J. Am. Chem. Soc.* **2012**, *134* (6), 2876-2879.
76. Ricci, F.; Vallée-Bélisle, A.; Simon, A. J.; Porchetta, A.; Plaxco, K. W., Using Nature's "Tricks" To Rationally Tune the Binding Properties of Biomolecular Receptors. *Acc. Chem. Res.* **2016**, *49* (9), 1884-1892.
77. Gao, X. J.; Chong, L. S.; Kim, M. S.; Elowitz, M. B., Programmable protein circuits in living cells. *Science* **2018**, *361* (6408), 1252.
78. Best, J. A.; Nijhout, H. F.; Reed, M. C., Homeostatic mechanisms in dopamine synthesis and release: a mathematical model. *Theor. Biol. Med. Modell.* **2009**, *6* (1), 21.
79. Reed, M.; Best, J.; Golubitsky, M.; Stewart, I.; Nijhout, H. F., Analysis of Homeostatic Mechanisms in Biochemical Networks. *Bull. Math. Biol.* **2017**, *79* (11), 2534-2557.
80. Patra, J. K.; Das, G.; Fraceto, L. F.; Campos, E. V. R.; Rodriguez-Torres, M. d. P.; Acosta-Torres, L. S.; Diaz-Torres, L. A.; Grillo, R.; Swamy, M. K.; Sharma, S.; Habtemariam, S.; Shin, H.-S., Nano based drug delivery systems: recent developments and future prospects. *J. Nanobiotechnol.* **2018**, *16* (1), 71.
81. Lu, H.; Wang, J.; Wang, T.; Zhong, J.; Bao, Y.; Hao, H., Recent progress on nanostructures for drug delivery applications. *J. Nanomater.* **2016**, 2016.
82. Pasek, S.; Risler, J.-L.; Brézellec, P., Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. *Bioinformatics* **2006**, *22* (12), 1418-1423.
83. Snel, B.; Bork, P.; Huynen, M., Genome evolution: gene fusion versus gene fission. *Trends Genet.* **2000**, *16* (1), 9-11.
84. Kuriyan, J.; Eisenberg, D., The origin of protein interactions and allostery in colocalization. *Nature* **2007**, *450* (7172), 983-990.
85. Pandey, N.; Nobles, C. L.; Zechiedrich, L.; Maresso, A. W.; Silberg, J. J., Combining Random Gene Fission and Rational Gene Fusion To Discover Near-Infrared Fluorescent Protein Fragments That Report on Protein-Protein Interactions. *ACS Synth. Biol.* **2015**, *4* (5), 615-624.



86. Hagner, K.; Setayeshgar, S.; Lynch, M., Stochastic protein multimerization, activity, and fitness. *Phys. Rev. E* **2018**, *98* (6), 062401.
87. Hashimoto, K.; Nishi, H.; Bryant, S.; Panchenko, A. R., Caught in self-interaction: evolutionary and functional mechanisms of protein homooligomerization. *Phys. Biol.* **2011**, *8* (3), 035007.
88. You, Y.; Tataurov, A. V.; Owczarzy, R., Measuring thermodynamic details of DNA hybridization using fluorescence. *Biopolymers* **2011**, *95* (7), 472-486.
89. Galau, G. A.; Britten, R. J.; Davidson, E. H., Studies on nucleic acid reassociation kinetics: rate of hybridization of excess RNA with DNA, compared to the rate of DNA renaturation. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74* (3), 1020.
90. Guan, S.; Price, J. C.; Ghaemmaghami, S.; Prusiner, S. B.; Burlingame, A. L., Compartment Modeling for Mammalian Protein Turnover Studies by Stable Isotope Metabolic Labeling. *Anal. Chem.* **2012**, *84* (9), 4014-4021.

# Chapitre 3 – Programming the cooperativity of multimeric assemblies by tuning their compactness via optimal linker connection \*

## 3.1 Abstract

Most functional nanosystems in living organisms are constructed using multimeric assemblies that provide multiple advantages over their monomeric counterparts. A simple strategy to create molecular assemblies through evolution and employed by scientists consists of fusing interacting domains to promote their complexation. Despite the importance and utility of this strategy, little is known about the programmable nature of this strategy and its ability to optimize the function of multimeric assemblies. Here, we used synthetic DNA assemblies and double-cycle analysis to explore the thermodynamic basis of trimeric assemblies and strategies to rationally program their assembly mechanism. As expected, we found that the cooperativity of the assembly of our trimeric model is largely attributable to configurational cooperativity (or chelate effect) and not allostery due to the introduction of linkers between interacting domains that increase their effective molarities. Interestingly, we also found that the cooperativity of assembly is well modelled as a function of the stability of the trimeric assembly, which can be readily tuned by simply varying the connection between the interacting domains. Finally, we found that maximal cooperativity is obtained with an optimal linker length that maximizes the compactness of the assembly via favourable enthalpic contributions. These results exemplify how molecular assemblies can be efficiently built by fusing interacting domains and optimizing their connectivity. They also provide a rational basis for understanding the thermodynamic and mechanistic forces driving the evolution of multimeric biological complexes.

---

\* Article soumis au journal *Nature Chemical Biology*: Lauzon, D.; Vallée-Bélisle, A. *Nat. Chem. Biol.* (Submitted for publication). D.L. a conceptualisé et réalisé toutes les expériences et les simulations numériques. D.L. et A.V.B. ont conjointement analysé et interprété les résultats. D.L. a écrit la première version du manuscrit. D.L. et A.V.B. ont conjointement écrit la version finale. A. V. B. a supervisé la réalisation du projet.

## 3.2 Introduction

Nanomachines in living organisms are at the basis of life<sup>1</sup> and are inspiring the development of many self-assembled nanotechnologies with numerous applications<sup>2-4</sup>. The majority of natural biomolecular nanomachines consist of multiple molecular components that need to self-assemble to trigger specific functions and cellular responses<sup>4-7</sup>. Molecular assemblies have been shown to provide both structural and functional advantages. These advantages include the creation<sup>8</sup> or stabilization<sup>9</sup> of active sites, as well as the creation of new self-regulated mechanisms through allostery<sup>5</sup> and/or configurational cooperativity<sup>10</sup>. One strategy employed in nature to create new functional assemblies consists of fusing modular interacting domains to promote their complexation<sup>11-13</sup>. This strategy has also been employed extensively by researchers to create novel molecular assemblies by simply linking binding partners consisting of proteins<sup>14, 15</sup>, peptides<sup>16</sup>, DNA<sup>17, 18</sup>, RNA<sup>19, 20</sup> and even organic molecules<sup>21</sup>. For example, various pH-induced molecular systems have been created by assembling pH-sensitive components, such as triplex-DNA<sup>22</sup>, i-motif<sup>23</sup>, organic polymers<sup>24</sup> or proteins<sup>25</sup>, into other specific functional units<sup>26-28</sup>.

Although the impact of fusing interacting domains to create self-assembled systems has been previously explored, most of these examples remain focused on the optimization of monomeric entities or dimeric host-guest systems. Numerous researchers have demonstrated that tuning the length of the linker between two interacting domains may allow programming of the binding efficiency, specificity, and level of cooperativity of these monomeric<sup>17, 29-31</sup> and host-guest nanosystems<sup>18, 21, 32</sup>. For example, Lim et al. exploited the fusion of multiple interaction domains to program the level of cooperativity of the output response<sup>29</sup>. In contrast, Ricci et al. exploited the entropic nature of unstructured linkers to tune the input/output response of unimolecular pH-dependent switches<sup>17</sup> and dimeric clamp-like DNA-based nanodevices<sup>18</sup>. Similarly, Fesik et al. demonstrated how optimal linkers inserted between two ligands can considerably improve the host-guest affinity between this new bidentate ligand and a specific protein receptor<sup>32</sup>. Despite these advances, however, little is known about how multimeric assemblies consisting of more than two components can be rationally programmed or optimized (*e.g.*, stability, compactness, cooperativity, functionality). Since nanosystems involving multiple components have proven to provide more functional advantages over their monomeric

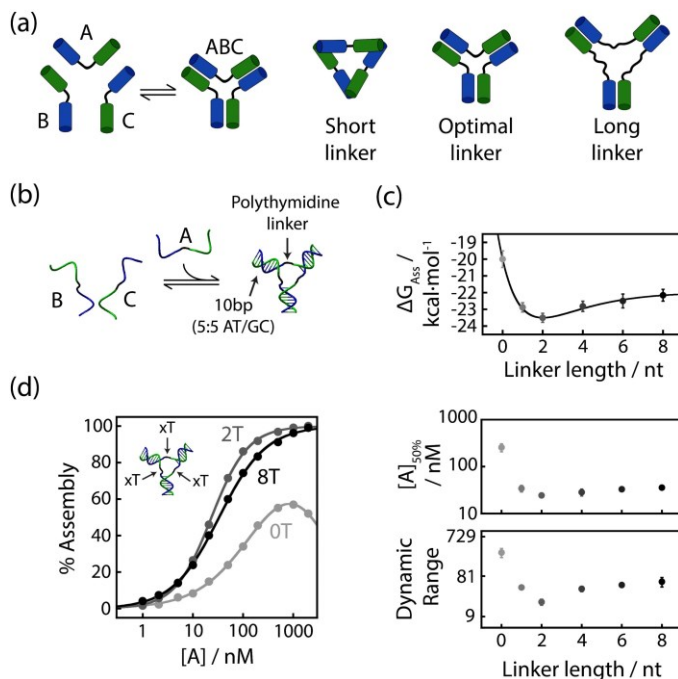
counterparts<sup>5</sup> and are becoming increasingly important in the development of future nanotechnologies<sup>2-4</sup>, a better understanding of how these systems can be tuned or optimized is needed.

Here, we used synthetic DNA assemblies to explore the thermodynamic basis of trimeric assemblies and strategies to efficiently evolve their function (Fig. 3.1a). This model DNA system was designed by fusing modular interacting domains to promote their complexation. We employed a double mutant cycle (DMC) analysis to determine whether varying the connectivity between the interacting domains affects their assembly mechanism via allosteric cooperativity or configurational cooperativity (also known as the chelate effect)<sup>10, 33, 34</sup>. Using unfolding experiments, we also determined the origin of the thermodynamic contribution of the linkers (*i.e.*, entropic and/or enthalpic) to the cooperativity of assembly. In agreement with our mathematical simulations, our results suggest that the fine-tuning of the linker properties is an efficient approach to program the cooperativity of multimeric nanosystems simply built from the fusion of interacting domains.

### 3.3 Results

To understand the implication of fusing interacting domains in the design of cooperative molecular assembly, we designed a simple three-component assembly that takes advantage of the high programmability, complementarity and versatility of DNA–DNA interactions (Fig. 3.1b)<sup>35-37</sup>. We selected a three-way junction as our model since each duplex-forming domain can be easily fused together using unstructured linkers. Furthermore, these interacting domains can be designed to be complementary and selective for one another while containing the same proportions of nucleobases, thereby ensuring similar stability for all interacting domains. More specifically, all our interacting domains contain 10 nucleobases (5 AT/5 GC) connected by an unstructured polythymidine linker, ranging from zero (0T) to eight thymidines (8T). We first determined the free energy ( $\Delta G_{\text{Ass}}$ ) of our trimeric assembly and found that the linker length, as demonstrated in our previous study<sup>38</sup>, had a drastic effect on the stability of all assemblies (Fig. 3.1c). For example, while a 2T linker provided the most stable trimeric assembly ( $\Delta G_{\text{Ass}} = -23.5 \pm 0.2 \text{ kcal}\cdot\text{mol}^{-1}$ ), shorter and longer linkers led to less stable assemblies ( $\Delta G_{\text{Ass},0\text{T}} = -20.0 \pm 0.5$

kcal·mol<sup>-1</sup> and  $\Delta G_{\text{Ass},8\text{T}} = -22.2 \pm 0.3$  kcal·mol<sup>-1</sup>). This decrease in stability had an immediate consequence on the assembly profile of these nanosystems. For instance, when forming the assembly by increasing the concentration of strand A to a fixed concentration of strands B and C (30 nM), we found a direct correlation between both  $\Delta G_{\text{Ass}}$  and  $[A]_{50\%}$ , the midpoint of the transition, and the dynamic range (DR), the concentration window over which the assembly occupancy shifts from 10% to 90%. We found that both  $[A]_{50\%}$  and DR can be minimized by more than one order of magnitude when using the energetically optimal 2T linker ( $[A]_{50\%} = 24 \pm 1$  nM and DR =  $34 \pm 5$ ). These results suggested that the linker properties, *e.g.*, its length, can provide an easy path to optimize the cooperativity level of the molecular assembly and even prevent its assembly (see the self-inhibition of the OT system).



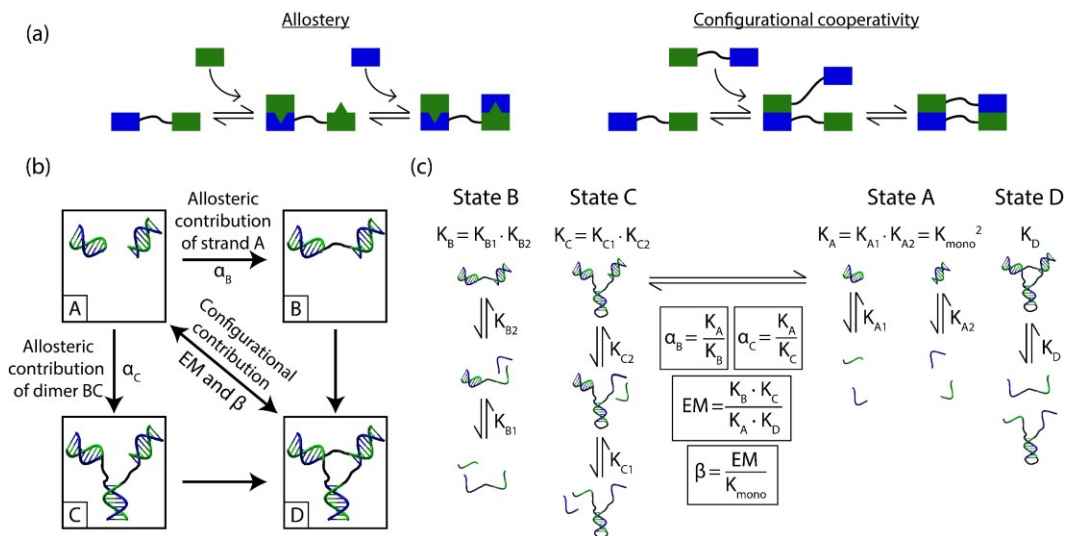
**Figure 3.1 – Programming molecular assemblies by connecting interacting domains via different linker lengths.** (a) A trimeric assembly can be engineered through the assembly of three components built by linking two interacting domains (blue and green) together using a linker (black). We explored the effect of the linker length on the efficiency of such model assembly, with shorter linkers expected to prevent some interaction within the interacting domains due to steric hindrance and longer linkers may entropically compromise the assembly. (b) Our model system consists of a DNA three-way junction with polythymidine sequences of different lengths (0, 1, 2, 4, 6, and 8) linking the interacting domains (10 nucleotides) together. Of note, binding domains were specifically designed to favour trimeric assembly over monomers or dimers (see DNA sequences in Supplementary Information). (c) We found that the stability of the assembly ( $\Delta G_{\text{Ass}}$ ) depends on the linker length, with the maximal stability reached using a 2T linker. (d) The

stability of the assembly is also correlated with its affinity and level of cooperativity. For example,  $[A]_{50\%}$ , the midpoint of the assembly transition, and the dynamic range (DR), the concentration window over which the assembly occupancy shifts from 10% to 90%, have the same dependence on linker length as  $\Delta G_{\text{Ass}}$ . Assemblies were tested in triplicate using 30 nM of strands B and C with increasing concentrations of A in PBS (50 mM  $\text{Na}_2\text{HPO}_4$ , 100 mM NaCl, pH = 7.00) at 37°C. See Fig. S3.1 for raw data leading to  $\Delta G_{\text{Ass}}$  and Fig. S3.2 for raw binding curve data.

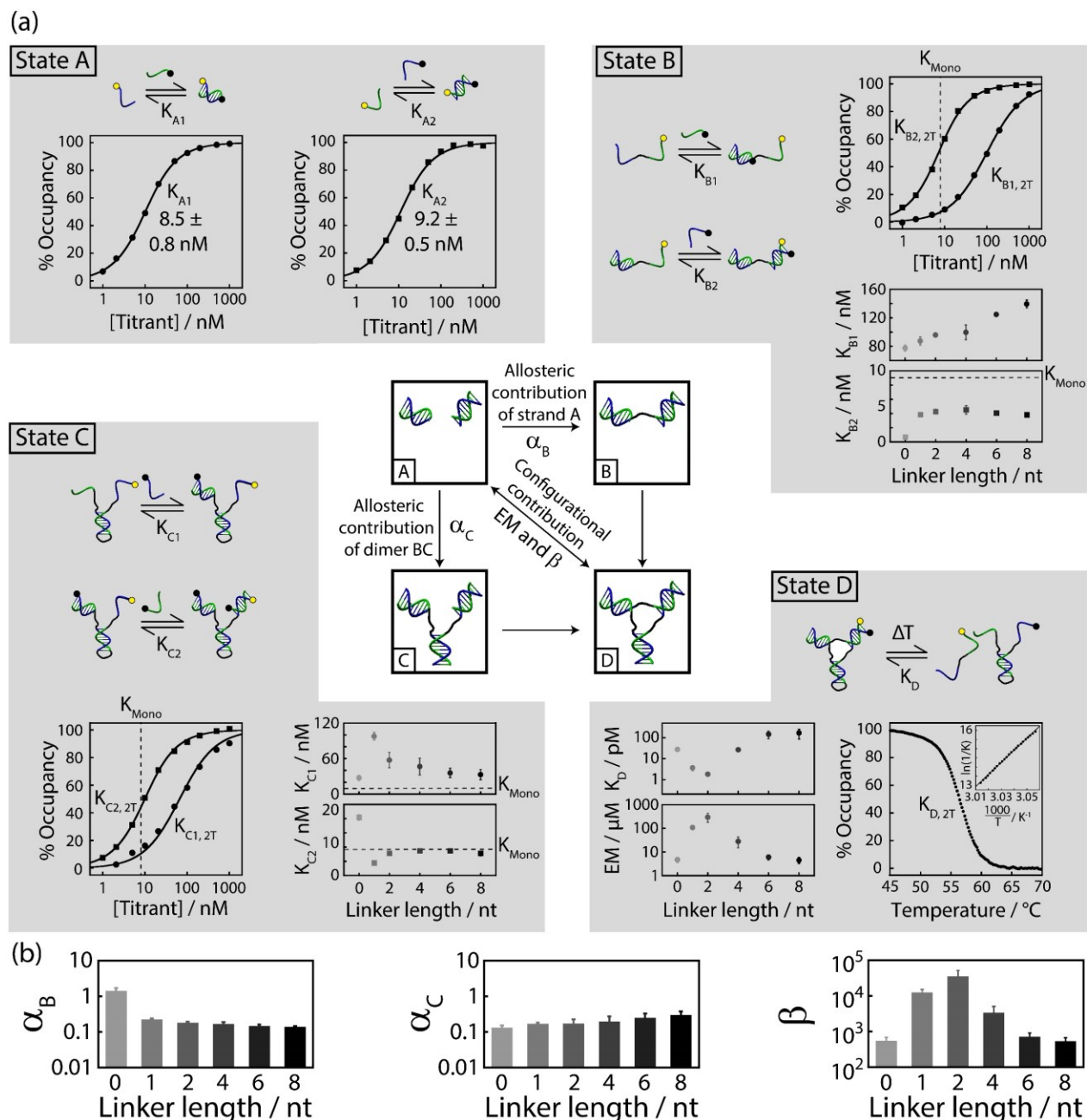
Two mechanisms are typically involved in cooperative molecular assembly: allostery and configurational cooperativity (Fig. 3.2a)<sup>10, 33, 34</sup>. In the allostery mechanism, the binding of a domain to its interacting partner is altered by the binding of an additional interacting domain (Fig. 3.2a, *left*). This communication between both interacting domains is typically considered to occur through a conformational change, but it can also be mediated through changes in dynamic properties<sup>10</sup>. The binding of oxygen to haemoglobin, for instance, represents a well-known example of allostery where the binding of the first oxygen molecule to haemoglobin increases the affinity of the three other binding sites through a global dynamic conformational change<sup>39</sup>. In the configurational cooperativity mechanism, the binding of a domain to its binding partner is enhanced by an increase in effective molarity cause by its attachment to a second interacting domain nearby (Fig. 3.2a, *right*). Such cooperative binding is often exploited in drug design where, for example, two drugs are chemically linked together to increase the strength of interaction with a protein receptor containing two close binding sites<sup>40</sup>. While we hypothesized that the enhanced stability observed when using the 2T linker may be linked to an optimal configurational cooperativity, we cannot rule out that the formation of the first DNA double helix may affect, through allostery, the affinity of the second and third double helices.

We explored the relative contributions of allostery and configurational cooperativity to the assembly of our system by performing a double mutant cycle (DMC) analysis<sup>33</sup>. This type of cycle is often used to dissect the specific energetic contribution of a mutation, a motif or a structure from all the interactions typically present in a system<sup>21, 33, 41</sup>. Here, we simplified the DMC analysis of our DNA assemblies to the binding on the last strand (*i.e.*, binding of strand A to the dimer BC) given that all DNA interacting domains are energetically equivalent and that the overall stability of the assembly ( $\Delta G_{\text{Ass}}$ ) is directly proportional to this last assembly step (Fig. S3.3). DMC analysis (Fig. 3.2b) began with assessing the intrinsic affinity of each individual

interacting domain between strand A and its respective binding partners in the preformed dimer BC (state A). We then introduced the linker on the strand A (state B) and on the preformed dimer BC (state C) to address the allosteric contributions of these linkers to the subsequent binding of both interacting domains ( $\alpha_B$  and  $\alpha_C$ , respectively). Finally, we introduced both linkers into the system (state D) to address the impact of bringing one interacting domain near another through configurational cooperativity ( $\beta$ ). For all these cooperative parameters ( $\alpha_B$ ,  $\alpha_C$  and  $\beta$ ), a value higher than unity indicates a positive cooperative effect, while values lower than unity indicate a negative cooperative effect.



**Figure 3.2 – Assessing the contributions of allostery and configurational cooperativity using a double mutant cycle analysis.** (a) In allostery, the first binding event induces a change in the state of the receptor that improves the affinity for the second binding event. In configurational cooperativity, the first binding event improves the affinity of the second binding event through an increase in effective molarity. (b) We designed a double mutant cycle (DMC) to assess the contribution of each mechanism. The cycle starts by first measuring the intrinsic affinity of both interacting domains (state A). Then, we introduced the linkers on strand A (state B) and on strand BC (state C) to address the allosteric contribution of the linker after subsequent binding of both domains ( $\alpha_B$  and  $\alpha_C$ , respectively). Finally, we added all linkers to the system (state D) to address the impact of bringing one interacting domain near another through an increase in effective molarity (EM). (c) The allosteric cooperativity ( $\alpha_B$  and  $\alpha_C$ ) and the configurational cooperativity ( $\beta$ ) can be estimated by measuring all the experimentally accessible equilibrium constants of each state ( $K_A$ ,  $K_B$ ,  $K_C$  and  $K_D$ ). All equilibrium constants presented here are defined as dissociation constants. Notably, for the  $K_C$  and  $K_D$  equilibrium, strands B (green) and C (blue) were covalently linked together using a 4T loop to ensure that they remained in a dimeric state and did not disassemble during the analysis.



**Figure 3.3 – Configurational cooperativity ( $\beta$ ) is the main determinant of the difference in assembly stability ( $\Delta G_{Ass}$ ) observed among all molecular assemblies built with various linker lengths.** *State A.* Characterization of the intrinsic affinity of the two interacting domains without linkers ( $K_{A1}$  and  $K_{A2}$ ). *State B.* Characterization of the allosteric contribution of linking together the two interacting domains of strand A on the successive binding of its respective binding partners on strand B ( $K_{B1}$ ) and strand C ( $K_{B2}$ ). *State C.* Characterization of the allosteric contribution of the linkers on the preformed dimer BC to the successive binding of its two respective binding partners ( $K_{C1}$  and  $K_{C2}$ ) on strand A. *State D.* Characterization of the configurational cooperativity contribution of both linkers on the final assembly measured by thermal denaturation ( $K_D$ ). (b) The DMC analysis reveals that the allosteric contribution of linking two interacting domains remains mainly negligible ( $0.1 < \alpha_B < 1$  and  $0.1 < \alpha_C < 1$ ). Configurational



cooperativity ( $\beta$ ), in contrast, remains the main determinant of the stability improvement of the assembly ( $10^2 < \beta < 10^5$ ), with a contribution up to 5 orders of magnitude higher than the allosteric contributions. Refer to Fig. 3.2 for the calculation of each parameter ( $\alpha_B$ ,  $\alpha_C$ , EM and  $\beta$ ). For clarity, only the binding curves and the melting curves of the 2T system are shown; see Fig. S3.4 ( $K_{A1}$ - $K_{A2}$ ), Fig. S3.5 ( $K_{B1}$ - $K_{B2}$ ), Fig. S3.6 ( $K_{C1}$ - $K_{C2}$ ) and Fig. S3.7 ( $K_D$ ) for the full dataset. Each binding constant corresponds to the average and standard deviation of three independent binding curves (see Table S3.1 for values). All DNA strands are chemically labelled with either a fluorophore moiety (FAM, yellow circle) or a quencher moiety (BHQ-1, black circle). All equilibrium constants are reported as dissociation constants. Experiments were performed at 37°C in PBS (50 mM  $\text{Na}_2\text{HPO}_4$ , 100 mM NaCl, pH=7.00).

All thermodynamic parameters of the cycle can be directly measured by determining the dissociation constants involved in each state of the cycle (see equations in Fig. 3.2c). We started the cycle by measuring the intrinsic binding affinity corresponding to the binding affinity of a monovalent interacting domain without the presence of the linkers,  $K_{\text{mono}}$ , *i.e.*, the 10 bp duplex containing 5 AT and 5 GC. We found, as expected, that the two interacting domains involved in the closing of the DNA assembly display similar affinities of  $8.5 \pm 0.8$  nM ( $K_{D1}$ ) and  $9.2 \pm 0.5$  nM ( $K_{D2}$ ) (Fig. 3.3a, State A), thus giving a  $K_{\text{mono}}$  of  $8.8 \pm 0.7$  nM.

We then characterized the allosteric contribution of linking the two interacting domains of strand A to the successive binding of its respective binding partners on strand B ( $K_{B1}$ ) and strand C ( $K_{B2}$ ) (Fig. 3.3b, State B). For example, we found that the first binding event ( $K_{B1}$ ) of the 2T system displayed a higher binding constant of  $96 \pm 2$  nM compared to  $K_{\text{mono}}$ , corresponding to a destabilization of  $1.5 \pm 0.1$  kcal·mol<sup>-1</sup>. Increased destabilization was also observed when employing longer linkers with  $K_{B1}$  varying from  $77 \pm 4$  nM (0T) to  $139 \pm 6$  nM (8T). This increase in  $K_{B1}$  is likely attributable to additional charge repulsion and steric hindrance created by the presence of additional nucleotides in the vicinity of the interacting domain. In contrast, the second binding event ( $K_{B2}$ ) of the 2T system displayed a binding constant similar to  $K_{\text{mono}}$  ( $4.3 \pm 0.3$  nM), suggesting that the double helix minimizes charge repulsion and/or improves the accessibility of the second interacting domain. Interestingly, all linkers displayed a binding constant of approximately  $4.1 \pm 0.3$  nM (1T to 8T), except for the 0T system which showed a much lower binding constant of  $0.7 \pm 0.1$  nM. This is likely attributable to potential base-stacking interactions between the nucleotides at the extremities of both duplexes<sup>42</sup>. Overall, these stabilizing or destabilizing effects are not large enough to significantly impact the stability of the

assembly. Indeed, the allosteric contribution quantified by  $\alpha_B$  remains slightly below unity ( $0.1 < \alpha_B < 3$ ) for all assemblies (Fig. 3.3b, *left*). From these  $\alpha_B$  values, we calculated that the allosteric contribution of the linker on strand A destabilizes the assembly of the three-way junction by  $0.93 \pm 0.05$  kcal·mol<sup>-1</sup> (1T) to  $1.22 \pm 0.02$  kcal·mol<sup>-1</sup> (8T), while the absence of a linker (0T) has a negligible contribution of  $-0.2 \pm 0.1$  kcal·mol<sup>-1</sup>.

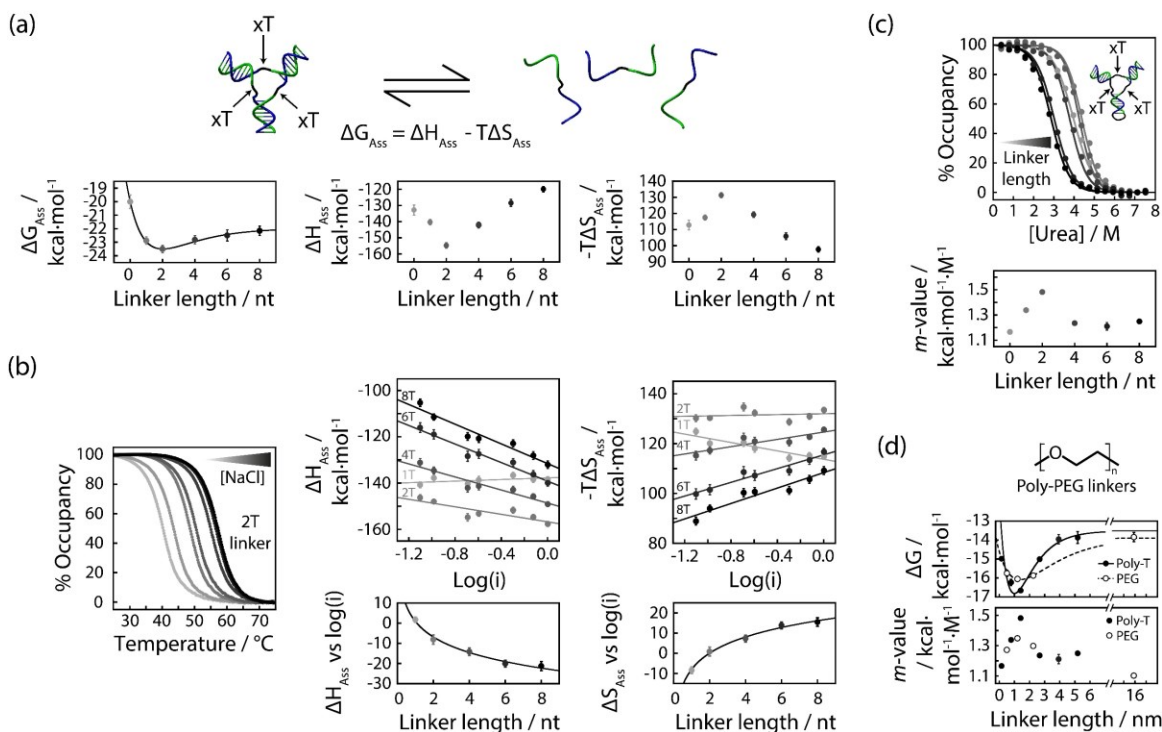
We then characterized the allosteric contributions of the linkers on the preformed dimer BC to the successive binding of its two respective binding partners ( $K_{C1}$  and  $K_{C2}$ ) on strand A (Fig. 3.3c, State C). For example, we found that the first binding event ( $K_{C1}$ ) of the 2T system displayed a higher binding constant of  $58 \pm 13$  nM than  $K_{\text{mono}}$ , corresponding to a destabilization of  $1.2 \pm 0.2$  kcal·mol<sup>-1</sup> for the hybridization of the first DNA binding domain. Similarly, all systems containing thymidine linkers (1T to 8T) displayed binding constants higher than  $K_{\text{mono}}$  starting from  $98 \pm 6$  nM (1T) and plateauing at approximately  $33 \pm 8$  nM (8T). This decrease in  $K_{C1}$  may be attributable to the increased distance between the interacting domain and the bulky negatively charged stem loop of the preformed dimer. In contrast, the absence of linkers (0T) resulted in a more stable system with a binding constant of  $28 \pm 4$  nM ( $K_{C1,0T}$ ) which suggests potential base stacking between the two duplexes<sup>42</sup>. In contrast, the second binding event ( $K_{C2}$ ), which takes place near two double helices, displayed binding constants similar to  $K_{\text{mono}}$  for linkers ranging from 2T to 8T ( $7.8 \pm 0.4$  nM). Interestingly, the shorter linkers (0T and 1T) behaved in the opposite manner to their first binding events. While the first binding event of the 0T system displayed a higher affinity than its second binding event ( $K_{C1,0T} < K_{C2,0T}$ ), the opposite trend was observed for the 1T system ( $K_{C2,1T} < K_{C1,1T}$ ). This discrepancy, however, cancels out and provides allosteric factors displaying small negative contributions ( $0.1 < \alpha_C < 1$ ), with  $\alpha_C$  remaining below unity with an upwards trend when the length of the linker was increased (Fig. 3.3c, *middle*). The allosteric impact of linking the blue and green domains together, therefore, destabilizes the assembly of the three-way junction by  $1.24 \pm 0.09$  kcal·mol<sup>-1</sup> (0T) to  $0.7 \pm 0.2$  kcal·mol<sup>-1</sup> (8T). Overall, we found that both allosteric factors ( $\alpha_B$  and  $\alpha_C$ ) remain too small to rationalize the large impact of the linker length on the stability of the assembly. This suggests that the addition of these linkers did not induce a change in conformation significant enough to affect binding affinities. Indeed, the two allosteric

contributions ( $\alpha_B$  and  $\alpha_C$ ) contribute less than 5% of the stability of the assembly (Table S3.1-S3.2).

Given that the allosteric contributions of the linkers were found to be negligible, we determined the contribution of configurational cooperativity coming from bringing two interacting domains into such close proximity. To do so, we determined the effective molarity (EM) of all systems by measuring the affinity of strand A towards the preformed dimer BC containing its respective binding partners ( $K_D$ ) using thermal denaturation (Fig. 3.3a, State D). We found that  $K_D$  showed a bell-shaped distribution, with the 2T system being the most stable ( $1.8 \pm 0.2$  pM) by two orders of magnitude compared to the less stable 8T system ( $170 \pm 85$  pM). This translates into an effective molarity of  $0.3 \pm 0.1$  mM for the 2T system compared to  $5 \pm 1$   $\mu$ M for the 8T system. We then calculated the contributions of the configurational cooperativity ( $\beta$ ) and found that it follows a similar trend to  $\Delta G_{Ass}$  (Fig. 3.1c), varying between  $(5 \pm 2) \times 10^2$  (8T) and  $(4 \pm 2) \times 10^4$  (2T) (Fig. 3.3c, *right*). In contrast to the allosteric contributions, which have negligible effects on the assembly ( $0.1 < \alpha_B$  and  $\alpha_C < 1$ ), the configurational cooperativity contribution  $\beta$  was found to be 2 to 5 orders of magnitude larger ( $500 < \beta < 40\,000$ ). This represents, for example, a configurational cooperativity contribution of  $7.4 \pm 0.2$  kcal·mol<sup>-1</sup> for the 2T system and  $4.4 \pm 0.2$  kcal·mol<sup>-1</sup> for the less stable 8T system (Table S3.1-S3.2). Overall, these results indicate that configurational cooperativity is the major determinant in stabilizing our DNA-based assemblies.

Having identified configurational cooperativity as the main determinant of cooperative assembly of our DNA-based model, we then explored whether this effect arises more from an enthalpic or an entropic contribution. To do so, we performed thermal denaturation of our assemblies and determined their  $\Delta H_{Ass}$  and  $\Delta S_{Ass}$  (Fig. 3.4a). We found that the variation in enthalpy of our systems correlated well with the variation in free energy, with the 2T linker system displaying both the strongest  $\Delta H_{Ass}$  ( $-155 \pm 2$  kcal·mol<sup>-1</sup>) and  $\Delta G_{Ass}$  ( $-23.5 \pm 0.3$  kcal·mol<sup>-1</sup>). In contrast, the entropic contribution displayed an opposite correlation with  $\Delta G_{Ass}$ . For example, the assembly of the most stable 2T system was found to be the most entropically unfavourable ( $131 \pm 2$  kcal·mol<sup>-1</sup>). Since the interacting domains remain constant among our different assemblies, these enthalpic and entropic variations likely originate from the different linkers and their impact

on the conformation of the junction. As observed in other systems, we also observed enthalpy-entropy compensation among our different assemblies (Fig. S3.8)<sup>43, 44</sup>. This suggests that the enthalpically favourable interactions arising from varying linker length also produce an unfavourable entropic contribution, likely via mobility restriction.



**Figure 3.4 – High stability arises from increases in enthalpy and level of compaction.**  $\Delta H_{Ass}$  and  $\Delta S_{Ass}$  measured through thermal denaturation (Fig. S3.10) showed that the contribution of the linkers is not purely entropic when the length of the linker is increased. (b) To facilitate assembly, we increased the ionic strength to overcome the electrostatic repulsion between the negatively charged polythymidine linkers. For longer linkers (> 2T), we found that higher ionic strength led to stronger enthalpic contributions and less favourable entropic contributions. In contrast, we found that both the enthalpic and entropic contributions of the 1T system decreased with increasing ionic strength, suggesting that this assembly displays fewer interactions and becomes less ordered at high salt concentrations. Interestingly, the  $\Delta H_{Ass}$  and the  $\Delta S_{Ass}$  of the interacting domains remain relatively unchanged with increasing ionic strength (Fig. S3.9, Table S3.3), suggesting that the variations the variation observed in the trimeric assemblies is likely attributable to the structural variation taking place upon the formation of the three-way junction. (c) Urea denaturation (Fig. S3.11) showed that the higher stability of the 2T system comes from its ability to bury more surfaces upon assembly, *i.e.*, a larger  $m$ -value<sup>45, 46</sup>. (d) A less bulky linker, such as PEG (Fig. S3.12), also led to trends and observations similar to those of the polythymidine linkers. See Tables S3.3-S3.6 for the full dataset and all values. All thermal denaturation experiments were performed in PBS buffer with varying NaCl concentrations, and all urea denaturation experiments were performed at 37°C in PBS (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 100 mM NaCl, pH = 7.00).

We decided to further explore the nature of the enthalpic contribution. Since the linkers are negatively charged and increasing their length may lead to further charge repulsion, we hypothesized that increasing the ionic strength may improve stability by minimizing charge repulsion. Using thermal denaturation, we measured the  $\Delta H_{\text{Ass}}$  and the  $\Delta S_{\text{Ass}}$  of all our systems containing linkers (1T to 8T) at various sodium chloride concentrations (Fig. 3.4b). We first measured the impact of ionic strength on the interacting domains in the absence of the linker and found that their  $\Delta H$  and  $\Delta S$  of assembly remained relatively unchanged (Fig. S3.9). This implies that any variation measured on the trimeric assemblies is likely attributable to the structural variation taking place upon the formation of the three-way junction. We thus determined the  $\Delta H$  and  $\Delta S$  of our trimeric assemblies and found that increasing the ionic strength of the solvent generally increase both  $\Delta H_{\text{Ass}}$  and  $\Delta S_{\text{Ass}}$  (Fig. 3.4b). As expected, increasing the ionic strength resulted in a higher  $\Delta H_{\text{Ass}}$ , likely by reducing charge repulsion between the linkers and the DNA-binding domains. The observed increase in  $\Delta S_{\text{Ass}}$  with ionic strength, on the other hand, suggests that the three-way junction becomes more ordered as charge repulsion decreases. Interestingly, however, we noted that despite having the most favourable  $\Delta H_{\text{Ass}}$ , the optimal 2T system displays only a small decrease in  $\Delta H_{\text{Ass}}$  with increasing ionic strength and almost no variation in  $\Delta S_{\text{Ass}}$ . This result suggests that the conformation of the 2T junction probably remains relatively unchanged at various ionic strength, perhaps due to an optimal compaction of the junction. In comparison, junctions made with longer linkers displayed increased  $\Delta S_{\text{Ass}}$  with increasing ionic strength, suggesting that the assembly becomes more ordered at high salt concentrations. In contrast, we found that both the enthalpic and entropic contributions of the 1T decreased with increasing ionic strength, suggesting that this assembly displays fewer interactions and becomes less ordered at high salt concentrations. Given that the linker length for the 1T system is too short for optimal stability, one hypothesis explaining this behaviour may be that increasing the salt concentration stabilizes the formation of an optimal “2T-like” junction leading to the dissociation of one base pair in the interacting domains (9 bp vs. 10 bp). Overall, these results suggest that the 2T optimizes the the stability of the trimeric assembly by optimizing its  $\Delta H_{\text{Ass}}$ .

We wanted to further explore whether the conformation of the 2T junction, which has the lowest enthalpic energy and which remains relatively unchanged at various ionic strength

optimizes its intramolecular interactions. For this purpose, we measured the variation in accessible surface area ( $\Delta$ ASA), and thus the level of compaction of each assembly by performing urea denaturation (Fig. 3.4c)<sup>47</sup>. This can be done by determining the  $m$ -value of each unfolding transition, a parameter that measures how  $\Delta G$  varies with the urea concentration, which also correlates with the accessible surface area. Surprisingly, we found that the 2T system buries 17% more surface area during its assembly ( $m$ -value =  $1.481 \pm 0.009$  kcal·mol<sup>-1</sup>·M<sup>-1</sup>, Fig. 3.4c, *left*), than the 4T, 6T and 8T systems (average  $m$ -value =  $1.23 \pm 0.02$  kcal·mol<sup>-1</sup>·M<sup>-1</sup>)<sup>45, 46</sup>. This suggests that the 2-thymidine linker creates a more compact junction that maximizes solvent exclusion upon assembly<sup>48, 49</sup>. In contrast, assemblies with longer linker (> 2T) displayed lower  $m$ -values, suggesting that their junction accommodate more solvent molecules. This interpretation is also consistent with the ionic strength experiments, water-filled junctions (4T, 6T, 8T) were more sensitive to salt concentrations (Fig. 3.4b). In contrast, we observed that the 0T and 1T assemblies displayed lower  $m$ -values, with 0T displaying the least compact conformation. This suggests that the possibility that these assemblies are not able to form the same number of interactions as the 2T system due to steric hindrance between the interacting domains at the branchpoint<sup>50</sup>. Overall, these results provide strong evidence that the extra enthalpic contribution at the origin of the high cooperativity of the 2T assembly is due to additional interactions at the junction providing optimal packing between the interacting domains.

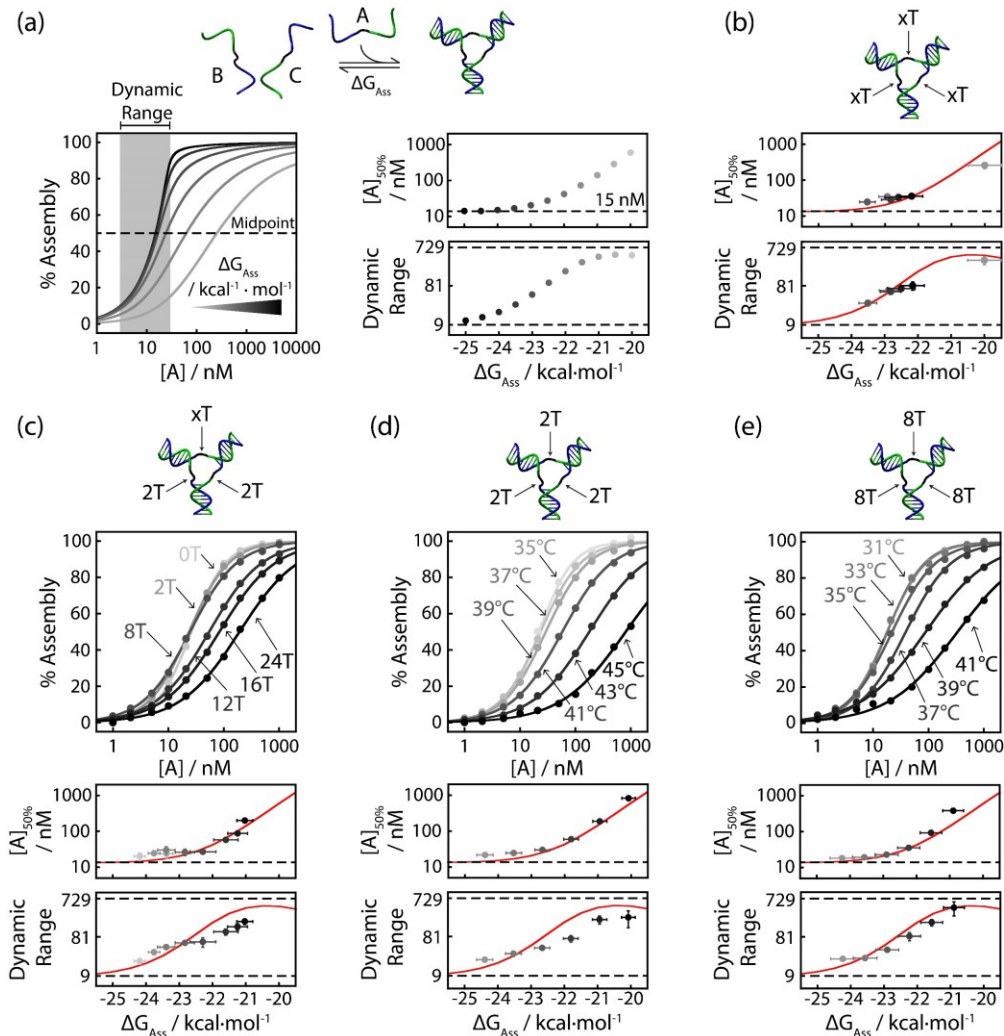
Similar conclusions were also reached by employing PEG linkers instead of polythymidine linkers to link the interacting domains. The same thermal and urea denaturation experiments revealed that the stability of the PEG linker systems followed the same bell-shaped behaviour as the poly-thymidine linker systems, with the optimal PEG system being slightly less stable ( $-16.1 \pm 0.1$  kcal·mol<sup>-1</sup>) than its thymidine-linker counterpart ( $-16.8 \pm 0.2$  kcal·mol<sup>-1</sup>) (Fig. 3.4d). Similarly, we found that the stability of all PEG systems correlated well with their level of compaction (Tables S3.5-S3.6). Interestingly, we also found that the PEG linker system requires a slightly longer linker for optimal stability ( $1.34 \pm 0.03$  nm) compared to its polythymidine counterpart ( $1.00 \pm 0.07$  nm). The improved stability provided by polythymidine linkers compared to PEG linkers may be rationalized by stronger intermolecular interactions between the thymidine linkers (e.g., H-bonding, and base stacking). In combination, these results suggest that the linker length

is an important determinant in creating compact junctions in molecular assemblies. This also suggests that the cooperativity level of molecular assemblies could be easily tuned or evolved through simple optimization of the length of a linker connecting two interacting domains.

Optimizing the cooperativity of complex assembly by varying the length of a linker between two interacting domains is a relatively simple and straightforward strategy to evolve a specific function. To demonstrate the impact of this strategy, we performed numerical simulations that confirmed that the level of cooperativity of a trimeric assembly, as measured by the dynamic range of assembly, can be varied by up to two orders of magnitude by simply tuning the stability of assembly (Fig. 3.5a). These numerical simulations revealed two important trends. On one hand, the most stable assembly ( $\Delta G_{\text{Ass}} = -25 \text{ kcal}\cdot\text{mol}^{-1}$ , darker curves Fig. 3.5a) was triggered at lower component concentrations ( $A_{50\%} = 15 \text{ nM}$ ) and displays the narrowest assembly profile (DR = 9). On the other hand, an assembly destabilized by only  $5 \text{ kcal}\cdot\text{mol}^{-1}$  ( $\Delta G_{\text{Ass}} = -20 \text{ kcal}\cdot\text{mol}^{-1}$ , lighter curves Fig. 3.5a) was triggered at a 40-fold higher component concentration ( $[A]_{50\%} = 592 \text{ nM}$ ) and displayed the most extended assembly profile (DR = 729). Of note, below a certain threshold ( $\Delta G_{\text{Ass}} < -20 \text{ kcal}\cdot\text{mol}^{-1}$ ), assembly became less favourable and exhibited self-inhibition through the formation of dimeric intermediates (see, for example, the OT system in Fig. 3.1d).

We validated these simulations by tuning the linker length and therefore the level of compaction of the assembly and confirmed the relationships between the stability of the assembly ( $\Delta G_{\text{Ass}}$ ), the  $[A]_{50\%}$  and the dynamic range (Fig. 3.5b). For example, a compact, highly stable assembly, such as the 2T system, assembled cooperatively (DR =  $34 \pm 5$ ) at low concentrations ( $[A]_{50\%} = 24 \pm 1 \text{ nM}$ ). In contrast, the less compact OT system assembled with a broader dynamic range (DR =  $358 \pm 75$ ) at much higher concentrations ( $[A]_{50\%} = 257 \pm 49 \text{ nM}$ ). In these examples, all the linkers were changed simultaneously, but the variation of only one linker (Fig. 3.5c) also enabled precise control over both the  $[A]_{50\%}$  and the dynamic range via a slight variation in the level of compaction, and stability of the assembly (Fig. S3.13 and Fig. S3.14). We further demonstrated the links between  $\Delta G_{\text{Ass}}$ ,  $[A]_{50\%}$  and the dynamic range by changing the temperature. Using the 2T (Fig. 3.5d) and 8T (Fig. 3.5e) linker assemblies, we demonstrated that a  $10^\circ\text{C}$  variation provided a programming ability ( $[A]_{50\%}$  and dynamic range) similar to that of

linker variation (nearly two orders of magnitude change). These results exemplify the simplicity with which the assembly profile of molecular assemblies can be optimized and evolved by simply shortening (deletion mutation) or increasing (insertion mutation) the linkers between their interacting domains, thereby adjusting the level of compaction of the assembly.



**Figure 3.5 – Programming the assembly profile of trimeric assemblies by tuning stability via various linker lengths.** (a) Numerical simulations of the assembly show that the stability of the assembly ( $\Delta G_{Ass}$ ) dictates its  $[A]_{50\%}$  and its dynamic range. (b) When changing all linkers simultaneously (see Fig. 3.1c-d), a good relationship between  $\Delta G_{Ass}$  and the assembly properties was observed. (c) A more precise way to program the stability of the assembly consisted of changing only one linker while keeping the two others at the optimal length of 2T (see Fig. S3.13 and Fig. S3.14). (d-e) The stability of the assemblies (here, 2T and 8T systems) can also be programmed by changing the temperature. All binding experiments were performed in triplicate in PBS (50 mM  $\text{Na}_2\text{HPO}_4$ , 100 mM NaCl, pH = 7.00) at 37°C if not mentioned otherwise. For all raw binding curves, see Fig. S3.15 (panel c), Fig. S3.16 (panel d) and Fig. S3.17 (panel e). Simulations were performed by numerically solving the system of equations representing the



molecular assembly thermodynamic equilibrium scheme using MATLAB (see SI). The assembly was triggered by adding component A, at concentrations from 1 to 1000 nM, to a fixed concentration (30 nM) of components B and C.

### 3.4 Discussion et conclusions

Here, we have demonstrated how cooperative molecular assemblies can be readily evolved by optimizing their level of compaction by tuning the connection between the interacting domains. We have demonstrated this by exploiting the high programmability and versatility of DNA–DNA interactions to create a simple trimeric “three-way junction” model system. More specifically, we have first shown that optimal cooperativity of assembly can be reached through a small, precise linker length (2T) that offers optimal stability (Fig. 3.1). Using a double mutant cycle analysis, we determined that the optimal connection between the interacting domains of this 2T system was mainly attributable to configurational cooperativity ( $\beta > 1000$ ) and displayed negative allosteric contributions ( $0.1 < \alpha_B$  and  $\alpha_C < 1$ ) (Fig. 3.2 and 3.3). By further exploring the thermodynamic basis of our cooperative trimeric assemblies, we demonstrated that the high cooperativity was not simply due to the minimization of the entropic contribution during assembly but instead was achieved through a linker that maximizes the enthalpic contribution. Indeed, we found that the 2T system produces the largest enthalpic contribution due to its ability to bury larger interacting surfaces at the interface of all components (Fig. 3.4). This suggests that the optimal linker helps to form favorable intermolecular interactions at the centre of the junction, thus replacing water molecules and forming a more compact assembly. Finally, we showed that simple variation in the linker length between two interacting domains in a trimeric assembly enables easy programming of its assembly profile ( $[A]_{50\%}$  and DR) (Fig. 3.5).

Engineering molecular assemblies by tuning the length of the linkers between interacting domains provides a simple and programmable chemical strategy to optimize the cooperativity of assembly of any self-assembled nanosystem. Here, we have demonstrated this concept using a DNA-based model system through simple variation of the length of an unstructured polythymidine linker. The universality of this concept was further suggested by obtaining a similar conclusion when the unstructured polythymidine linker was replaced with a more flexible, less bulky PEG linker. Furthermore, all assembly profiles with different levels of cooperativity were

also well modelled as a simple function of their stability, demonstrating the quantitative nature of this strategy for programming the  $[A]_{50\%}$  and the dynamic range of the assembly. Although we demonstrated this strategy using nucleic acids and PEG molecules, we believe that our conclusion also applies to other biosystems, including proteins. Indeed, scientists have also shown that fusing protein interacting domains using intrinsically disordered linkers can create a variety of molecular assemblies with different stabilities<sup>15, 51-53</sup>. For example, by changing the length of a flexible peptide linker between two DHFR proteins, researchers observed different levels of oligomerization<sup>11</sup>. Varying the linker length in fused proteins<sup>54</sup> or fused ligands<sup>55</sup> was also shown to improve affinity through the avidity effect. More specifically, PROTAC drugs could benefit from linker optimization that maximizes compaction and cooperative PPI interactions to promote the assembly of the ternary complex, thus preventing self-inhibition and the Hook effect<sup>56</sup>. We therefore anticipate that this linker optimization strategy could become an important strategy to engineer de novo protein assemblies with improved cooperativity and regulatory functions.

The ability to narrow or broaden the dynamic range of molecular functions remains important both for living organisms and for the development of various biotechnologies. One well-known alternative strategy to program the dynamic range of functional nanosystems consists of creating allosteric systems<sup>57, 58</sup>. One limitation of allosteric strategies, however, remains its design complexity, since it typically requires the creation of energetically interconnected binding sites into single- or multidomain proteins<sup>59, 60</sup>. In contrast, we have shown here that multiple interacting domains can form in a cooperatively self-assembled nanosystem through a simple optimization of the linkers connecting the interacting domains. This strategy seems to have occurred extensively during evolution to create functional protein complexes through, for example, gene fusion events where the genes coding for two interaction domains are fused into one open reading frame<sup>61, 62</sup>. During evolution, these fused complexes can often acquire long interdomain linkers that can play a key role in their function<sup>63, 64</sup>. For example, direct evolutionary evidence of linker optimization was recently observed in the Cdk inhibitor p21. The linker in this protein was found to be optimal for both its stability and inhibition efficiency ( $IC_{50}$ ), while elongation or truncation of the linker almost always resulted in less stable assemblies with lower inhibition activities and even in some cases a reduction in cooperativity<sup>65</sup>. In summary, we

believe that our findings can provide protein engineers and nanoengineers with a thermodynamic framework to create highly cooperative self-assembled nanosystems by linking interacting domains while providing a new understanding of the thermodynamic and mechanistic forces that could have driven the evolution of multidomain protein complexes.

### **3.5 Materials and Methods**

#### **Chemicals.**

All ancillary reagents for the DNA synthesizer were purchased from ChemGenes Corporation, Wilmington: Cap A mix (Ac<sub>2</sub>O/Pyridine/THF), Cap B (16% Melm in THF), activation reagents (0.25 M ETT in ACN), oxidizer (0.1 M I<sub>2</sub> in H<sub>2</sub>O/Pyridine/THF), DMT removal reagent (3% TCA in DCM) and anhydrous acetonitrile. All four DNA phosphoramidites (A, T, G and C) were purchased from ChemGenes Corporation, Wilmington: dT-CE, dG-CE (n-ibu), dA-CE (n-bz) and dC-CE (n-bz). Pre-pack 1000Å 1µmol Suc-CPG synthesis columns were purchased from Biosearch Technologies, Navato: dT, dG(dmf), dA(bz), dC(bz), BHQ-1 and fluorescein. Labelled DNA phosphoramidites were purchased from Biosearch Technologies, Navato: 6-CarboxyFluorescein-Aminohexyl Amidite (6-FAM). PEG phosphoramidites were purchased from ChemGenes Corporation, Wilmington: DMT-ethane-Diol (01PEG), DMT-triethyloxy-Glycol (03PEG), DMT Hexaethylene Glycol (06PEG) and Polyethyleneglycol 2000 CED (2kPEG) phosphoramidite. Solid-phase extraction columns (MicroPure II) were purchased from Biosearch Technologies, Novato. Triethylamine was purchased from Fisher Scientific while acetic acid (ACS, ≥99.7%) and Ammonium hydroxide solution (ACS, 28.0-30.0%) were purchased from Sigma Aldrich. Milli-Q water was prepared in the lab when needed. All buffer solutions were prepared in the lab using anhydrous Na<sub>2</sub>HPO<sub>4</sub>, NaCl and NaOH purchased from Bioshop Canada In., Burlington. HCl 25% was purchased from VWR Canada, Mississauga.

#### **Oligonucleotides Synthesis.**

We performed the synthesis in the lab using standard phosphoramidite chemistry with the DNA/RNA synthesizer H-6 (K&A Laborgeraete, Schaafheim). The deprotection of the bases and the cleavage from the solid support were performed in fresh 28.0-30.0% aqueous ammonia

at 60°C overnight (16h). Purification of unlabelled DNA strands were performed by solid-phase extraction using the P-8 purifier (K&A Laborgeraete, Schaafheim) while DNA strands containing labelled (FAM and BHQ-1) were purified by HPLC (1260 infinity, Agilent Technologies) with a reversed-phase column (XBridge® Oligonucleotide BEH C18 OBDTM Prep column, 130Å, 2.5 µm, 10 x 50 mm) using a gradient of anhydrous ACN and 0.1 M TEAA buffer. After their purifications, all DNA strands were dried with a Genevac™ miVac centrifugal concentrator. The DNA pellets were then resuspended in Milli-Q water and quantified by UV-Vis (NanoDrop™ 2000c, Thermo Scientific). Their concentrations were calculated from their absorbances at 260 nm using the absorption coefficients ( $\epsilon$ ) estimated from their sequences using the IDT OligoAnalyzer tool.

### **DNA sequences.**

#### Trimeric 3-way junction

Strand A: 5'-GTA GTT CGA G(T)<sub>X</sub>C TTG ACG TAC-3'

Strand B: 5'-FAM-GCT GAT GTA C(T)<sub>X</sub>C TCG AAC TAC-3'

Strand C: 5'- GTA CGT CAA G(T)<sub>X</sub>G TAC ATC AGC-BHQ1-3'

with X = 0, 1, 2, 4, 6 or 8

#### Strands for the double mutant cycle

Binding domain #1 BHQ (state A,  $K_{A1}$ ): 5'- GCT GAT GTA C-BHQ1-3'

Binding domain #2 FAM (state A,  $K_{A1}$ ): 5'- GTA CAT CAG C-FAM-3'

Binding domain #2 BHQ (state B,  $K_{B2}$ ): 5'- GTA CAT CAG C-BHQ1-3'

Binding domain #3 BHQ (state A, State C,  $K_{A2}$ ,  $K_{C1}$ ): 5'- CTC GAA CTA C-BHQ1-3'

Binding domain #3 Unlabelled (State C,  $K_{C2}$ ): 5'- CTC GAA CTA C-3'

Binding domain #4 FAM (state A,  $K_{A2}$ ): 5'- GTA GTT CGA G-FAM-3'

Binding domain #4 BHQ (state B,  $K_{B1}$ ): 5'- GTA GTT CGA G-BHQ1-3'

Binding domain #4 Unlabelled (state B,  $K_{B2}$ ): 5'- GTA GTT CGA G-3'

Target strand with different linkers (State B, State D,  $K_{B1}$ ,  $K_{B2}$ ,  $K_D$ ): 5'-FAM-GCT GAT GTA C(T)<sub>X</sub>C  
TCG AAC TAC-3'

Stem-loop dimer with different linkers (State C, State D,  $K_{C1}$ ,  $K_{C2}$ ,  $K_D$ ): 5'-GTA GTT CGA G(T)<sub>X</sub>C TTG  
ACG TAC TTT TGT ACG TCA AG(T)<sub>X</sub> GTA CAT CAG C -BHQ1-3'

### PEG system

Target strand: 5'-FAM-GCT GAT GTA C(PEG)<sub>X</sub>C TCG AAC TAC-3'

Stem-loop dimer strand: 5'-GTA GTT CGA G(PEG)<sub>X</sub>C TTG ACG TAC TTT TGT ACG TCA AG(PEG)<sub>X</sub> GTA  
CAT CAG C -BHQ1-3'

with  $X = 1, 3, 6$  or  $2k$ .

Underlined nucleotides represent the linker section. *FAM* corresponds to a fluorescein moiety while *BHQ1* corresponds to a black hole quencher<sup>®</sup>-1 moiety. *PEG* corresponds to a polyethylene glycol moiety. Of note, the  $2k$  *PEG* linker corresponds to a linker with a molecular weight of  $2k$ .

### **Fluorescence experiments.**

Binding Curves. In a quartz cuvette, appropriate dilutions of the probe solution (*i.e.*, the strand containing the fluorophore) are made to obtain a final volume of  $900 \mu\text{L}$ . The probe solution contains either one or two strands depending on whether it is a dimeric ( $A + B \leftrightarrow AB$ ) or a trimeric titration ( $A + B + C \leftrightarrow ABC$ ). To this solution, a small volume of the titrant is added, and the mixture is left to equilibrate for 10 minutes. When the equilibrium is reached (*i.e.*, the signal is not changing over time), the fluorescence signal is recorded ( $496/520 \text{ nm}$ ) with a fluorimeter (Cary Eclipse, Agilent Technologies). To avoid a high dilution of the probe solution, the total volume of added titrant is kept below 5% of the probe solution volume. Therefore, we can then assume that the probe concentration stays approximately constant during the titration. The titration curves are fitted with the quadratic equation (Eq. 3.1) for a dimeric titration or with the Hill equation (Eq. 3.2) for a cooperative titration. In both equations,  $K_{D,obs}$  is the observed dissociation constant,  $F$  is the fluorescence signal,  $Bsl$  is the baseline signal,  $Amp$  is the amplitude change of the fluorescent when the probe is fully bound by the titrant,  $[P]$  is the concentration of

the probe, [T] is the concentration of the titrant and nH is the Hill factor which is related to the dynamic range by  $DR = 81^{nH}$ .

$$F = Bsl + \frac{Amp}{2} \left( [T] + [P] + K_{D,obs} - \sqrt{([T] + [P] + K_{D,obs})^2 + 4[P][T]} \right) \quad (\text{Eq. 3.1})$$

$$F = Bsl + Amp \frac{[T]^{nH}}{K_{D,obs}^{nH} + [T]^{nH}} \quad (\text{Eq 3.2})$$

Thermal denaturation curves. All melting curves were obtained by measuring the variation in fluorescence at different temperatures using a fluorimeter coupled with a Peltier temperature controller (Cary Eclipse, Agilent Technologies). All samples are first heated to 90°C at 5°C/min and cool down to 10°C at 5°C/min to anneal the sample. The sample is left at 10°C for 5 minutes prior to the recording of the melting curve from 10°C to 90°C at 0.4°C/min. The melting curves are then normalized by an upper and lower linear baseline. The thermodynamic parameters  $\Delta H_{Ass}$  and  $\Delta S_{Ass}$  were obtained from the Van't Hoff analysis (Eq. 3.3). The  $\Delta G_{Ass}$  is calculated from these two values:  $\Delta G_{Ass} = \Delta H_{Ass} - T\Delta S_{Ass}$ .

$$\ln(K_{Ass}) = \frac{\Delta H_{Ass}}{R} \frac{1}{T} + \frac{\Delta S_{Ass}}{R} \quad (\text{Eq. 3.3})$$

Urea denaturation curves. Denaturation by urea was performed following a method developed by our lab<sup>47</sup>. A 900  $\mu$ L solution of the DNA-based system of interest in 10 M urea buffered solution is first made. We then diluted this solution with a solution containing the same concentration of the DNA-based system in pure buffer. Prior to the recording of the fluorescence signal (Cary Eclipse, Agilent Technologies), all samples were equilibrated for two minutes. The bimolecular denaturation ( $AB \rightleftharpoons A + B$ ) is performed by keeping one of the two components in a 10-time excess compared to the other one. The denaturation curve is then fitted using Eq. 3.4<sup>47</sup>. In the equation, [D] is the concentration of the DNA strand in excess, [U] is the urea concentration, F is the fluorescence signal,  $F^{\circ}_{Ass}$  and  $F^{\circ}_{Diss}$  are the fluorescence signals of the assembled and disassembled state,  $\sigma_{Ass}$  and  $\sigma_{Diss}$  correspond to the influence of urea on the fluorescence signal of the assembled and disassembled state,  $\Delta G^{\circ}_{Ass}$  is the Gibbs free energy of the system in absence of urea, T is the temperature and R is the gas constant and m is the *m*-value which corresponds

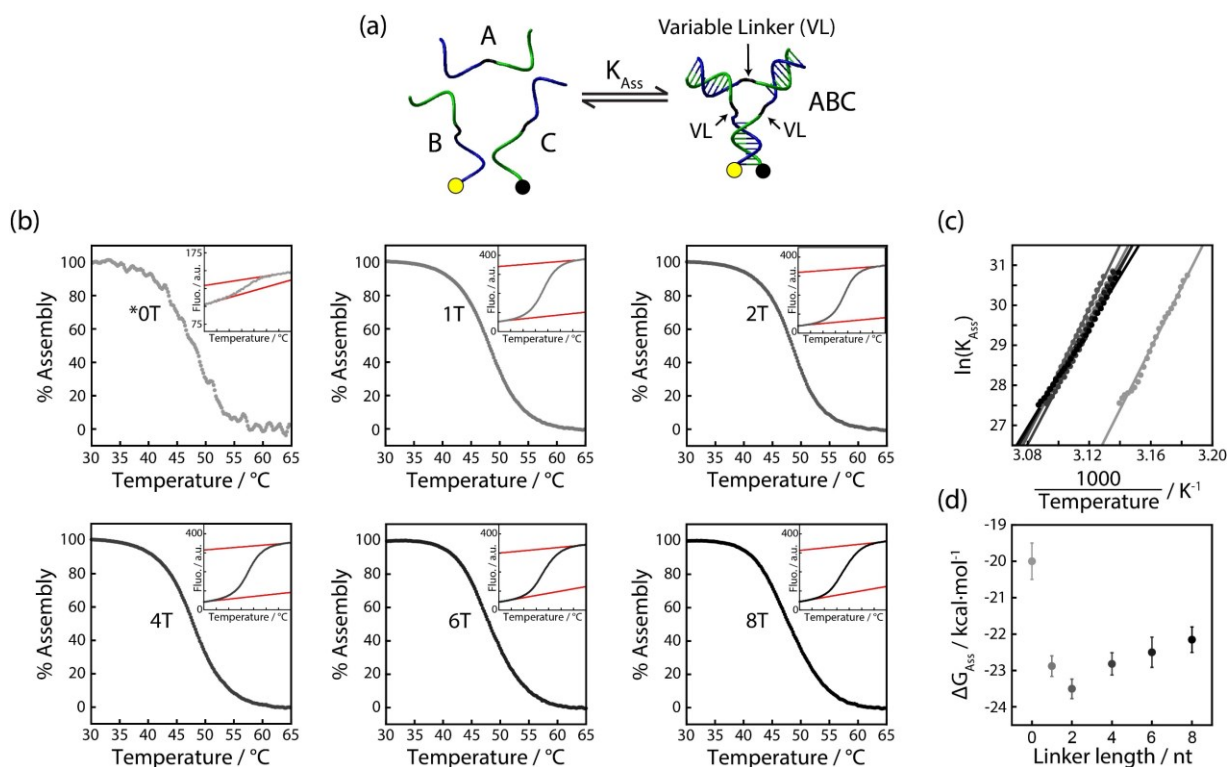
to the influence of urea on the  $\Delta G^{\circ}_{Ass}$  and correlates with the variation in accessible surface area ( $\Delta ASA$ )<sup>47</sup>.

$$F = \frac{[D] \cdot (F^{\circ}_{Ass} + \sigma_{Ass}[U]) + (F^{\circ}_{Diss} + \sigma_{Diss}[U]) \cdot e^{-\frac{(\Delta G^{\circ}_{Ass} - m[U])}{RT}}}{\left( [D] + e^{-\frac{\Delta G^{\circ}_{Ass} - m[U]}{RT}} \right)} \quad (\text{Eq. 3.4})$$

### 3.6 Acknowledgements

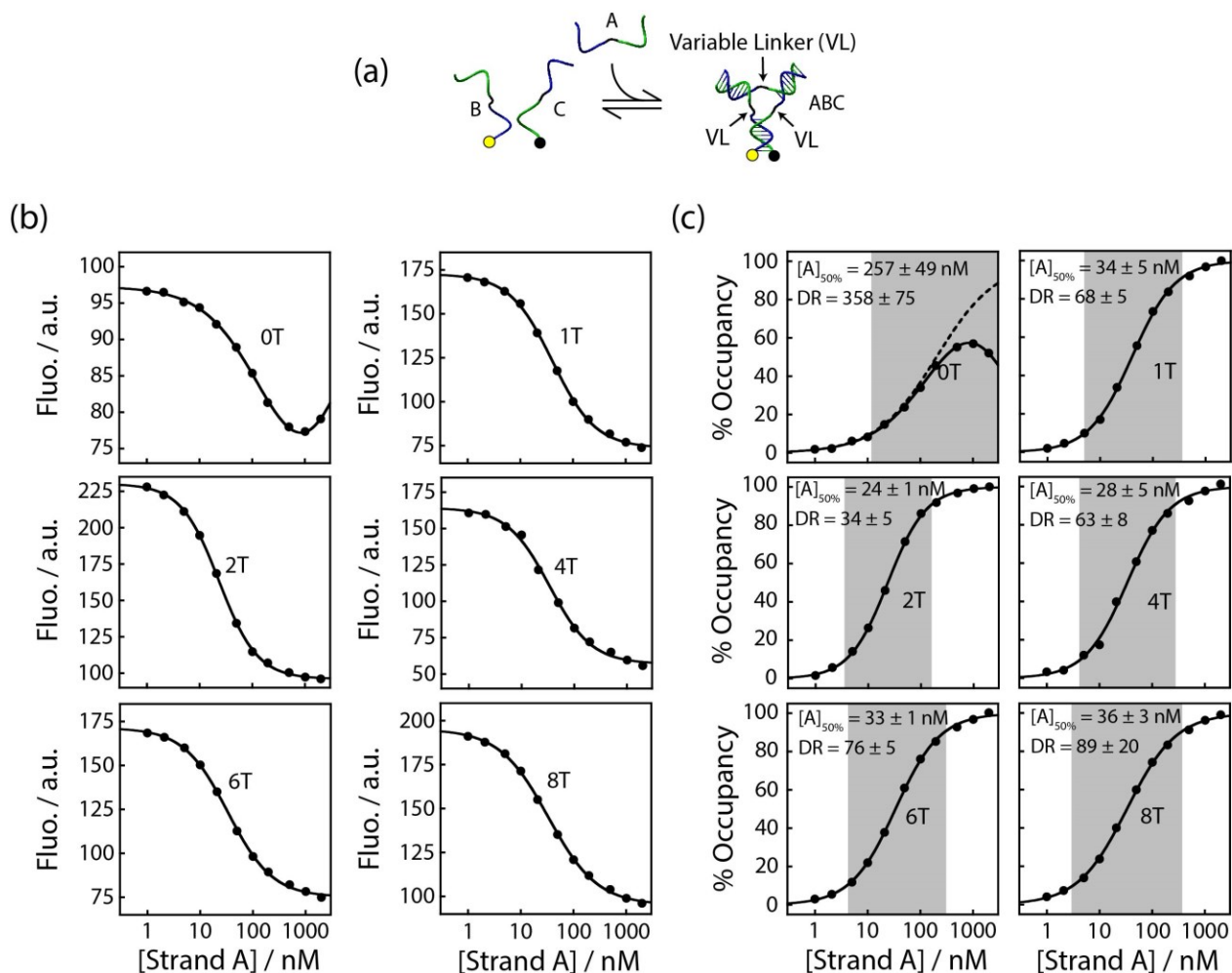
This research was conducted through the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants (RGPIN-2020-06975) (A.V.-B). A.V.-B. is Canada Research Chair in Bioengineering and Bionanotechnology, Tier II. D. L. acknowledges a 3rd cycle scholarship from the Fonds de recherche du Québec – Nature et technologies (FRQNT).

### 3.7 Supporting Figures



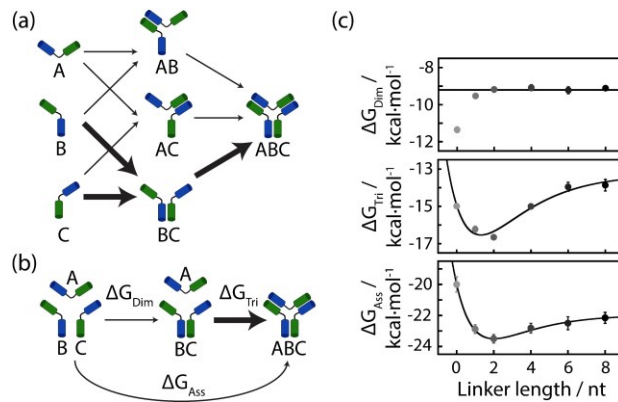
**Figure S3.1 – Thermal denaturation of the trimeric assembly (ABC) built with variable linker length (0T to 8T).** (a) Stability analysis of the trimeric assembly ABC. The yellow circle represents a FAM moiety at the 5' extremity while a black circle represents a BHQ-1 moiety at the 3' extremity. \*For the 0T system, due to a higher stability of the dimers (see Fig. S3.3), the FAM moiety was moved at the 3' extremity to prevent stacking interaction. This led to a higher background with a smaller variation in fluorescence. (b) Thermal denaturation at 1  $\mu$ M of each strand. Raw data are represented in inset and are normalized by two linear regressions corresponding to the lower and upper baselines (red lines). (c) Van't Hoff analysis of the thermal denaturation curves. (d) The 2T system is the most stable assembly while the 0T system is the less stable ( $\Delta G_{Ass,0T} = -20.0 \pm 0.5$ ,  $\Delta G_{Ass,1T} = -22.9 \pm 0.3$ ,  $\Delta G_{Ass,2T} = -23.5 \pm 0.3$ ,  $\Delta G_{Ass,4T} = -22.8 \pm 0.3$ ,  $\Delta G_{Ass,6T} = -22.5 \pm 0.4$  and  $\Delta G_{Ass,8T} = -22.2 \pm 0.3$  kcal·mol<sup>-1</sup>). The error bars represent the error calculated from the Van't Hoff linear regression. All experiments were performed in PBS (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 100 mM NaCl, pH=7.00).  $K_{Ass}$  represents the association constant of the trimeric assembly.



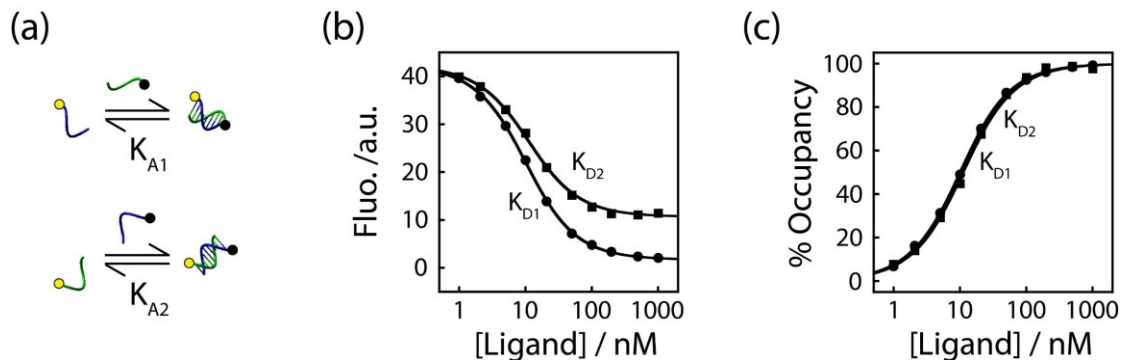


**Figure S3.2 – The trimeric assembly of nanosystems built with variable linker length (0T to 8T).**

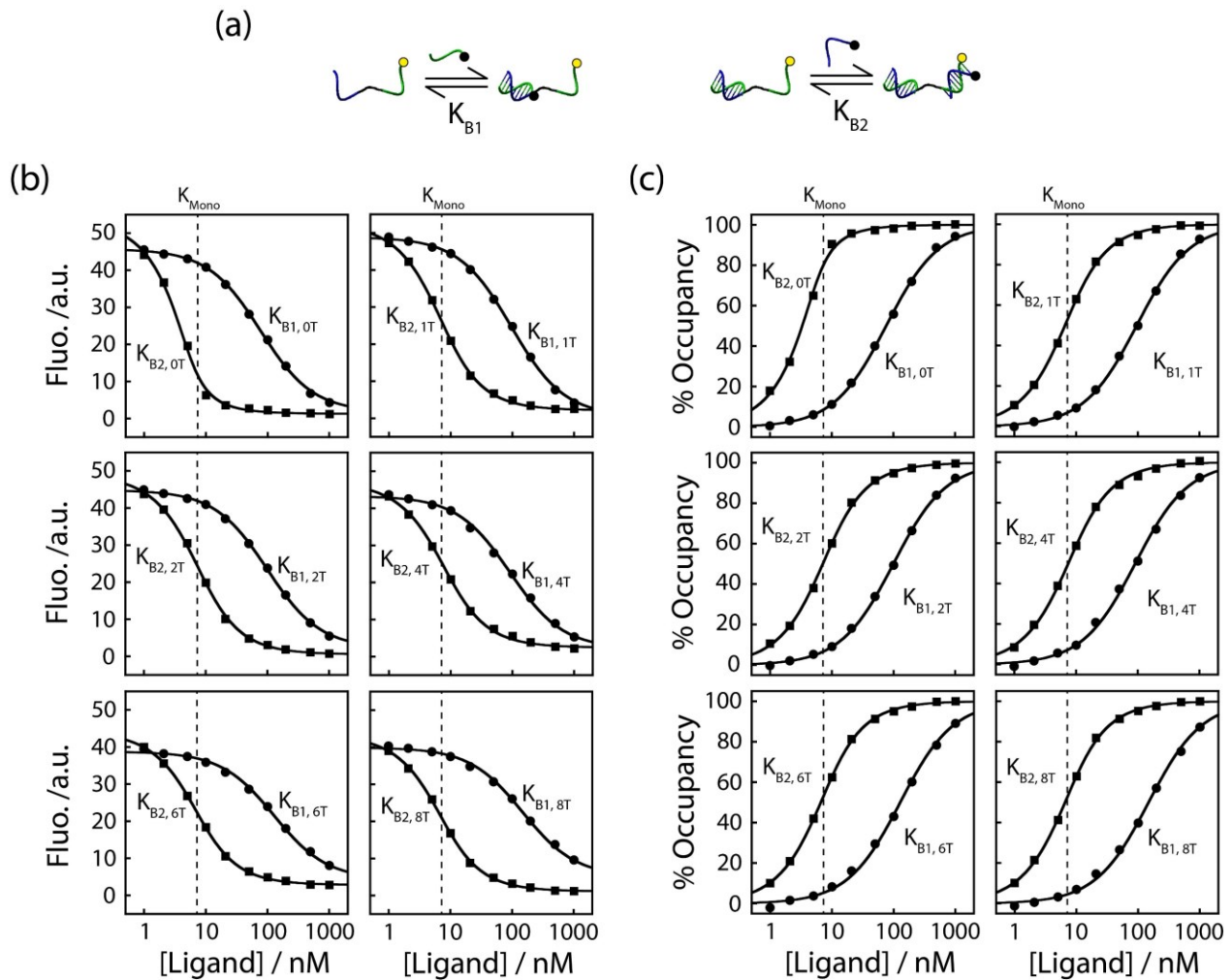
(a) All assemblies were performed in triplicate (only one data set is shown). The trimeric assemblies were performed using 30 nM of fluorescently labelled strand B and 30 nM of strand C labelled with a quencher. The assembly is induced by adding an increasing amount of unlabelled strand A. Yellow circle represents a FAM moiety at the 5' extremity while a black circle represents a BHQ-1 moiety at the 3' extremity. (b) Raw fluorescence binding curves. Of note, the 0T system displays a strong initial quenching (100 a.u. vs 200 a.u.) and a self-inhibition behaviour at higher concentration of strand A because of its lower trimeric stability (Fig. S3.1) and a higher stability of its dimeric intermediates (Fig. S3.3). (c) Normalized binding curves. The  $[A]_{50\%}$  and the dynamic range (DR) of all assemblies represent the average and standard deviation of three independent experiments. The gray rectangles illustrate the dynamic range of the assemblies. The experiments were performed at 37°C in PBS (50 mM  $\text{Na}_2\text{HPO}_4$ , 100 mM NaCl, pH=7.00).



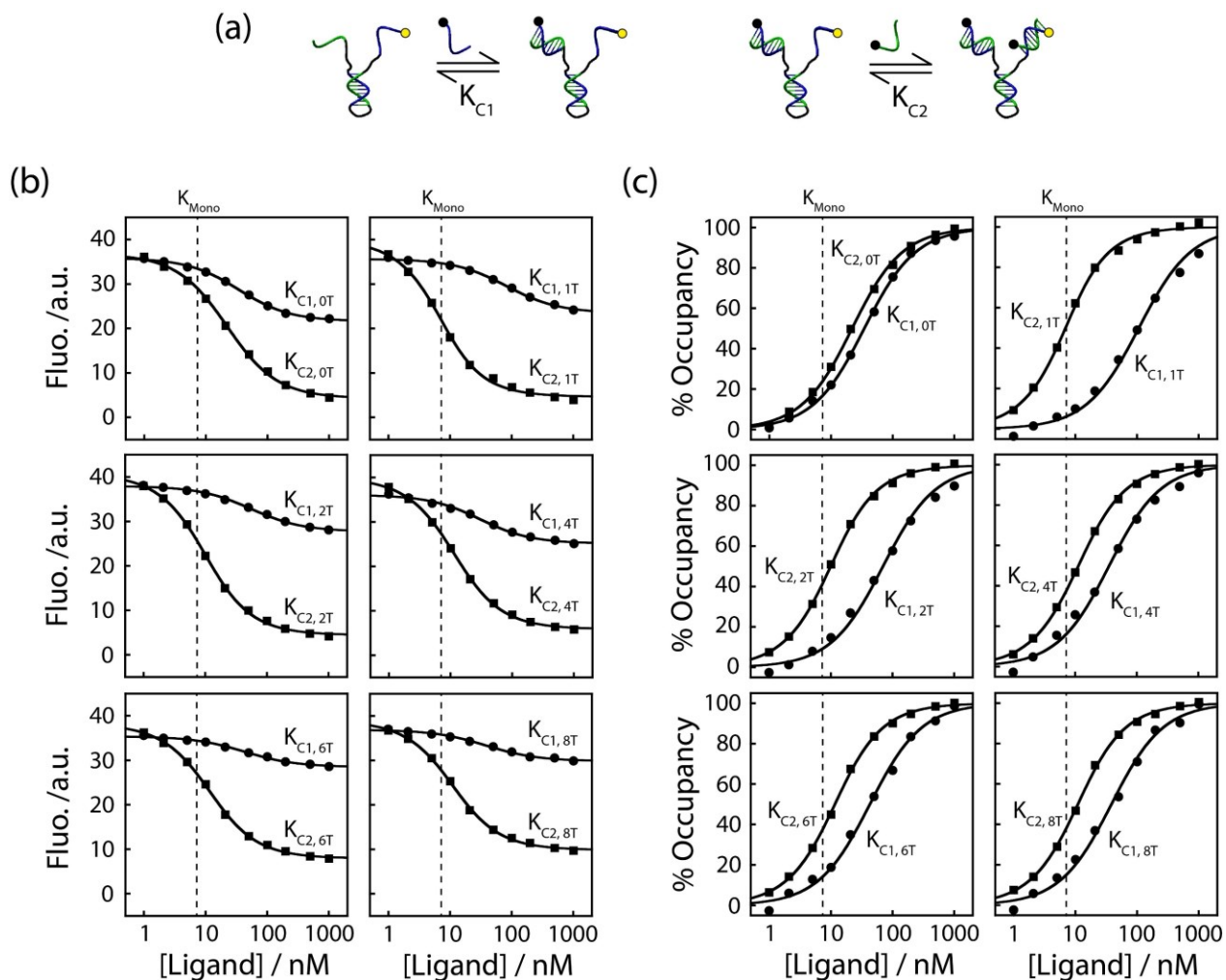
**Figure S3.3 – The variation of  $\Delta G_{Ass}$  when increasing the length of the linker is directly related to the final step if the assembly ( $\Delta G_{Tri}$ ).** A trimeric assembly can be assembled from three different path involving three different dimeric intermediates. (b) Since we designed all binding domains to be equivalent (10nt involving 5 AT and 5 GC), we will focus simply on the analysis of one path consisting of the association of strand B and C into a dimeric intermediate followed by the binding of strand A to form the trimer. (c) The stability of the dimer ( $\Delta G_{Dim}$ ) is independent of the linker length between 1T and 8T. In contrast, the linker length has a direct impact on the last step of the assembly ( $\Delta G_{Tri}$ ) which correlates well with the tendency observed with the overall stability of the system ( $\Delta G_{Ass}$ ). Of note, the 0T system behaves quite differently than the system containing linkers by displaying a higher dimeric stability. Since there is no linker to separate the two binding domains of strand B and C, some interactions can happen between the binding domains specific to strand A. This results in a double helix of 12nt instead of 10nt, as estimated from Mfold. All stabilities were measured in PBS (50 mM  $\text{Na}_2\text{HPO}_4$ , 100 mM NaCl, pH=7.00) using thermal denaturation (see Fig. S3.1 for an example). The error bars represent the error calculated from the Van't Hoff linear regression.



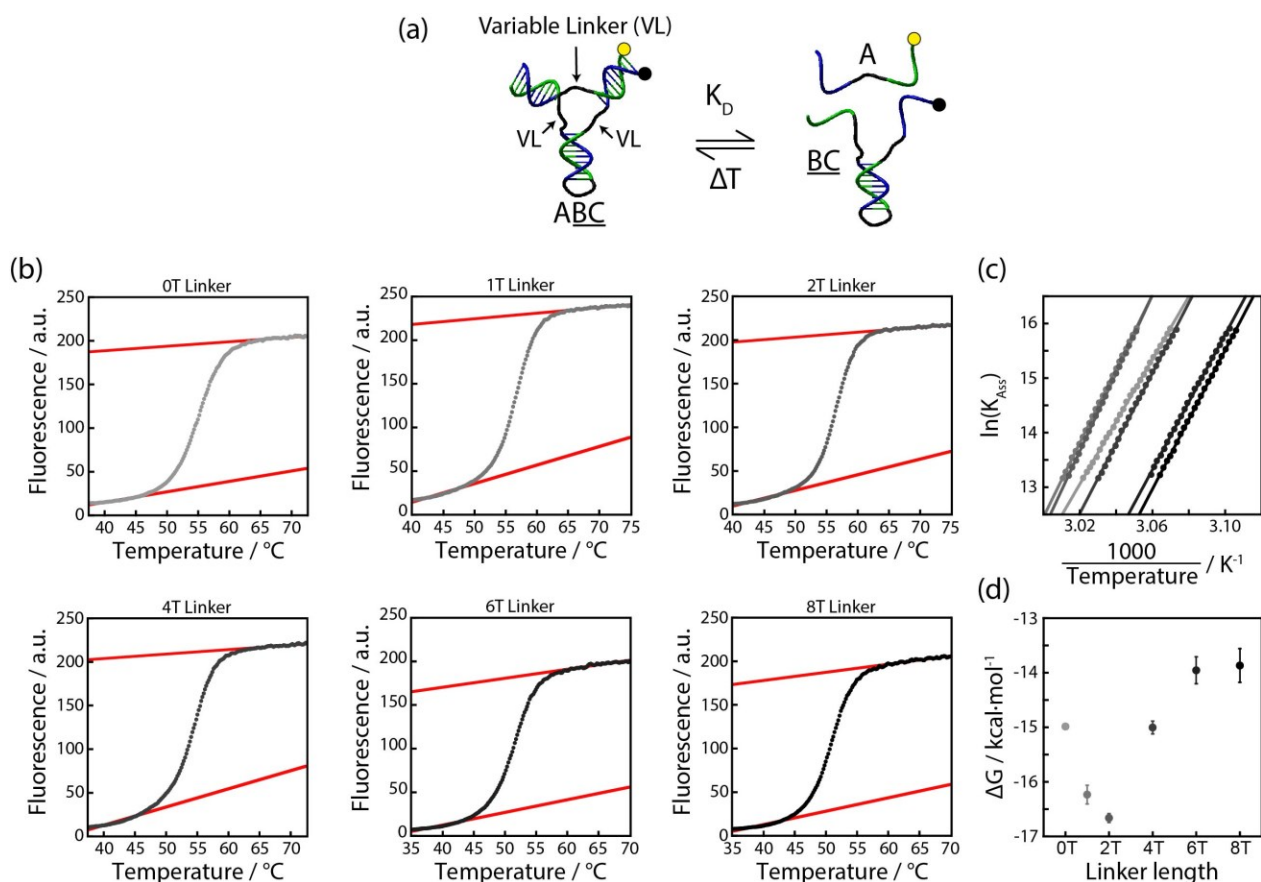
**Figure S3.4 – Binding curves for  $K_{A1}$  and  $K_{A2}$  measurements.** (a) All binding curves were performed in triplicate (only one data set is shown).  $K_{A1}$  and  $K_{A2}$  were measurement using 5 nM of fluorescently labelled receptors and by adding different amounts of the ligand labelled with a quencher. Yellow circle represents a FAM moiety at the 5' extremity while a black circle represents a BHQ-1 moiety at the 3' extremity. (b) Raw fluorescence binding curves. (c) Normalized binding curves. The experiments were performed at 37°C in PBS (50 mM  $\text{Na}_2\text{HPO}_4$ , 100 mM NaCl, pH=7.00).



**Figure S3.5 – Binding curves for  $K_{B1}$  and  $K_{B2}$  measurements.** (a) All binding curves were performed in triplicate (only one data set is shown). All  $K_{B1}$  were measured using 5 nM of fluorescently labelled receptors and by adding an increasing amount of the first ligand labelled with a quencher. All  $K_{B2}$  were measured using 5 nM of fluorescently labelled receptors and 5  $\mu$ M of the first ligand unlabelled ensuring that the receptor is fully bound when adding an increasing amount of the second ligand labelled with a quencher. Yellow circle represents a FAM moiety at the 5' extremity while a black circle represents a BHQ-1 moiety at the 3' extremity. (b) Raw fluorescence binding curves. (c) Normalized binding curves. The experiments were performed at 37°C in PBS (50 mM  $\text{Na}_2\text{HPO}_4$ , 100 mM NaCl, pH=7.00).



**Figure S3.6 – Binding curves for  $K_{C1}$  and  $K_{C2}$  measurements.** (a) All binding curves were performed in triplicate (only one data set is shown). All  $K_{C1}$  were measured using 5 nM of fluorescently labelled receptors (stem loop) and by adding different amounts of the first ligand labelled with a quencher. All  $K_{C2}$  were measured using 5 nM of fluorescently labelled receptors (stem loop) and 5  $\mu$ M of the first ligand labelled with a quencher ensuring that the receptor is fully bound when adding an increasing amount of the second ligand labelled with a quencher. Yellow circle represents a FAM moiety at the 5' extremity while a black circle represents a BHQ-1 moiety at the 3' extremity. (b) Raw fluorescence binding curves. (c) Normalized binding curves. The experiments were performed at 37°C in PBS (50 mM  $\text{Na}_2\text{HPO}_4$ , 100 mM NaCl, pH=7.00).



**Figure S3.7 – Thermal denaturation of the  $K_D$  equilibrium reveals the major contribution of the chelate cooperativity.** (a) Thermal denaturation of strand A bound to the preformed dimer BC ( $K_D$ ) using variable linkers (VL) ranging from 0T to 8T. (b) Thermal denaturation curves of poly-thymine linker system. Data are normalized through the linear regression of the upper baseline ( $\sim 15^\circ\text{C}$  above the  $T_m$ ) and the lower baseline ( $\sim 15^\circ\text{C}$  below the  $T_m$ ). The baselines are shown in red. (c) Van't Hoff analysis of the thermal denaturation curves. (d) The 2T system shows the highest stability while the 8T system shows the lowest stability:  $\Delta G_{0T} = -14.99 \pm 0.04$ ,  $\Delta G_{1T} = -16.2 \pm 0.2$ ,  $\Delta G_{2T} = -16.66 \pm 0.08$ ,  $\Delta G_{4T} = -15.0 \pm 0.1$ ,  $\Delta G_{6T} = -14.0 \pm 0.2$  and  $\Delta G_{8T} = -13.9 \pm 0.3$  kcal·mol<sup>-1</sup>. The  $\Delta G$  and their standard deviations were calculated from three independent measurements. The huge impact of the linker on the stability of  $K_D$  highlights the major contribution of the chelate cooperativity ( $\beta$ ). We used 50 nM of fluorescently labelled strand A (FAM at the 5' extremity) and 500 nM of preformed dimers BC containing a quencher (BHQ-1 at the 3' extremity). Experiments were performed in PBS (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 100 mM NaCl, pH=7.00).

**Table S3.1. – Thermodynamic dissociation constants, allosteric parameters ( $\alpha_B$  and  $\alpha_C$ ) and configurational cooperativity parameters ( $\beta$  and EM) of all three-way junctions.**

	0T	1T	2T	4T	6T	8T
$K_{A1} / \text{nM}$	8.5 ± 0.8					
$K_{A2} / \text{nM}$	9.2 ± 0.5					
$K_{\text{Mono}} / \text{nM}$	8.8 ± 0.7					
$K_{B1} / \text{nM}$	77 ± 4	87 ± 6	96 ± 2	100 ± 11	125 ± 2	139 ± 6
$K_{B2} / \text{nM}$	0.7 ± 0.1	3.8 ± 0.2	4.3 ± 0.3	4.5 ± 0.6	4.1 ± 0.3	3.8 ± 0.2
$K_{C1} / \text{nM}$	28 ± 4	98 ± 6	58 ± 13	47 ± 14	36 ± 8	33 ± 8
$K_{C2} / \text{nM}$	20 ± 1	4.4 ± 0.3	7.8 ± 0.4	8.7 ± 0.3	8.7 ± 0.5	7.8 ± 0.8
$K_D / \text{pM}$	28 ± 2	3.6 ± 1.0	1.8 ± 0.2	27 ± 5	147 ± 59	170 ± 85
$\alpha_B$	1.4 ± 0.3	0.22 ± 0.02	0.18 ± 0.01	0.17 ± 0.02	0.15 ± 0.01	0.14 ± 0.01
$\alpha_C$	0.13 ± 0.02	0.17 ± 0.01	0.17 ± 0.05	0.20 ± 0.08	0.25 ± 0.08	0.30 ± 0.08
EM / $\mu\text{M}$	14.7 ± 3.7	548 ± 99	1403 ± 479	95 ± 42	14.7 ± 3.8	11.1 ± 4.1
$\beta$	(1.7 ± 0.4) × 10 <sup>3</sup>	(64 ± 11) × 10 <sup>3</sup>	(165 ± 63) × 10 <sup>3</sup>	(11 ± 5) × 10 <sup>3</sup>	(1.7 ± 0.4) × 10 <sup>3</sup>	(1.3 ± 0.4) × 10 <sup>3</sup>

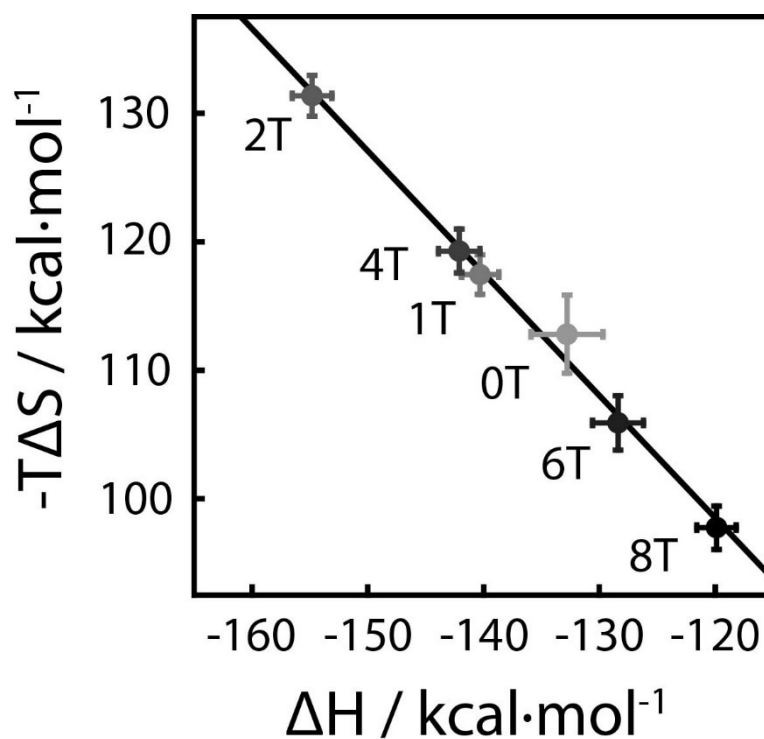
\*All values and standard deviations were calculated from three independent measurements.

**Table S3.2. – Contribution of allostery ( $\alpha_B$  and  $\alpha_C$ ) and the configurational cooperativity ( $\beta$  and EM) to the overall stability of the assembly.**

	0T	1T	2T	4T	6T	8T
$\Delta G_{\text{Ass}}^* / \text{kcal}\cdot\text{mol}^{-1}$	-20.0 ± 0.5	-22.9 ± 0.3	-23.5 ± 0.3	-22.8 ± 0.3	-22.5 ± 0.4	-22.2 ± 0.3
$\Delta G\alpha_B / \text{kcal}\cdot\text{mol}^{-1}$	-0.2 ± 0.1	0.93 ± 0.05	1.05 ± 0.04	1.11 ± 0.08	1.19 ± 0.06	1.22 ± 0.02
% $\Delta G\alpha_B^{**}$	1.1 ± 0.5 %	-4.1 ± 0.3 %	-4.5 ± 0.2 %	-4.8 ± 0.4 %	-5.2 ± 0.4 %	-5.5 ± 0.2 %
$\Delta G\alpha_C / \text{kcal}\cdot\text{mol}^{-1}$	1.24 ± 0.09	1.09 ± 0.05	1.1 ± 0.2	1.0 ± 0.2	0.9 ± 0.2	0.7 ± 0.2
% $\Delta G\alpha_C^{**}$	-6.0 ± 0.6 %	-4.8 ± 0.3 %	-4.6 ± 0.9 %	-4 ± 1 %	-4 ± 1 %	-3.3 ± 0.8 %
$\Delta G\beta / \text{kcal}\cdot\text{mol}^{-1}$	-4.6 ± 0.2	-6.8 ± 0.1	-7.4 ± 0.2	-5.7 ± 0.3	-4.6 ± 0.2	-4.4 ± 0.2
% $\Delta G\beta^{**}$	22 ± 1 %	30 ± 1 %	32 ± 1 %	26 ± 2 %	21 ± 1 %	20 ± 1 %

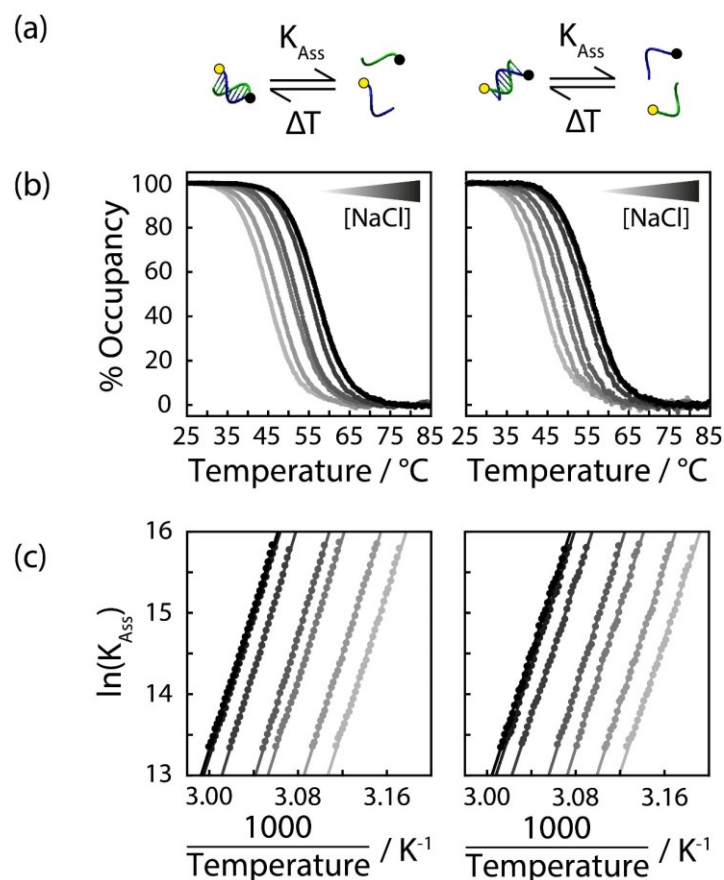
\* All  $\Delta G_{\text{Ass}}$  were calculated from Fig. S3.1.

\*\* A negative sign indicates a negative contribution towards the  $\Delta G_{\text{Ass}}$ .



**Figure S3.8 – Our different assemblies display an enthalpy-entropy compensation.** This likely suggests that the gain of interaction at the center of the junction comes at the expense of a mobility restriction. All experiments were performed in PBS (50 mM  $\text{Na}_2\text{HPO}_4$ , 100 mM NaCl, pH=7.00).  $\Delta H$  and  $-T\Delta S$  (at 37°C) were measured from the melting in Fig. S3.1.





**Figure S3.9 – Thermal denaturation of the binding domains at different salt concentrations.** (a) Thermodynamic analysis of the binding domains. The yellow circle represents a FAM moiety at the 5' extremity while a black circle represents a BHQ-1 moiety at the 3' extremity. (b) Normalized thermal denaturation at 1  $\mu\text{M}$  of each strand. See Fig. S3.1 for the normalization procedure. (c) Van't Hoff analysis of the thermal denaturation curves. All experiments were performed in PBS (50 mM  $\text{Na}_2\text{HPO}_4$ , 100 mM NaCl, pH=7.00).  $K_{\text{Ass}}$  represents the association constant of the trimeric assembly. These experiments were not performed in triplicate.

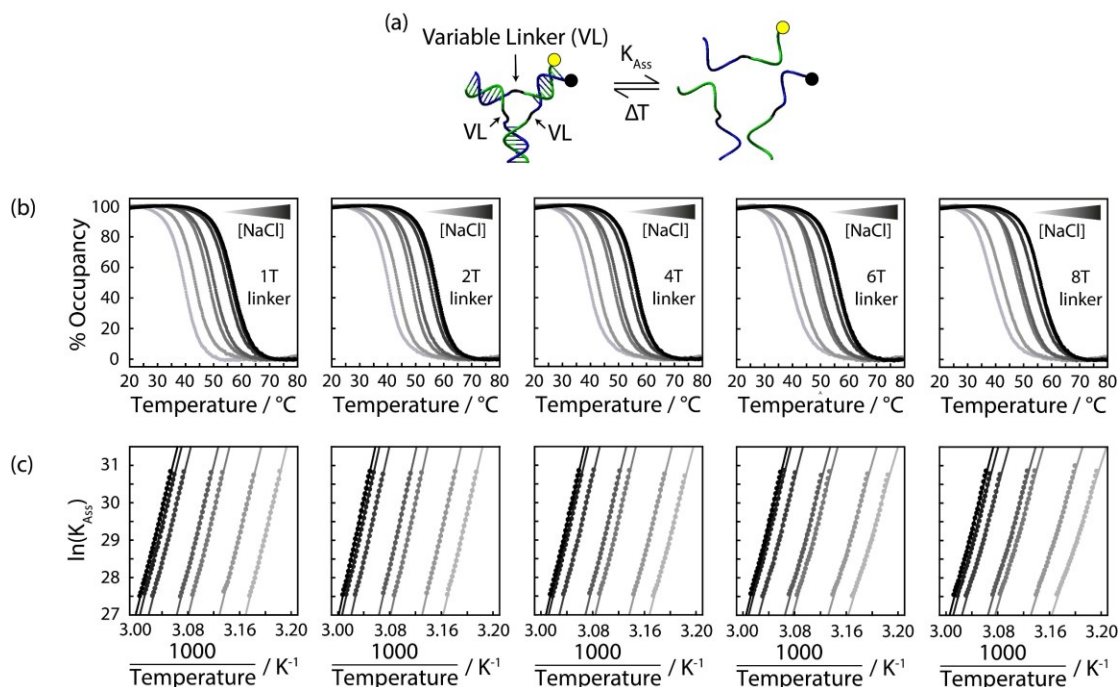
**Table S3.3. – The ionic strength has almost a negligible impact on the  $\Delta H_{\text{Ass}}$  (in  $\text{kcal}\cdot\text{mol}^{-1}$ ) and the  $-\text{T}\Delta S_{\text{Ass}}$  (in  $\text{kcal}\cdot\text{mol}^{-1}$ ) of both binding domains.**

[Na <sup>+</sup> ]	Left arm		Right arm	
	$\Delta H_{\text{Ass}}$	$-\text{T}\Delta S_{\text{Ass}}$	$\Delta H_{\text{Ass}}$	$-\text{T}\Delta S_{\text{Ass}}$
50	$-84.2 \pm 0.7$	$75.0 \pm 0.7$	$-81.7 \pm 0.8$	$71.0 \pm 0.8$
100	$-86.3 \pm 0.5$	$76.5 \pm 0.5$	$-83.8 \pm 0.7$	$72.5 \pm 0.7$
200	$-86.4 \pm 0.7$	$75.7 \pm 0.7$	$-86.4 \pm 0.7$	$74.4 \pm 0.6$
250	$-89.2 \pm 0.5$	$78.1 \pm 0.4$	$-86.6 \pm 0.7$	$74.1 \pm 0.6$
500	$-89.4 \pm 0.5$	$77.4 \pm 0.5$	$-82.1 \pm 0.6$	$69.0 \pm 0.5$
750	$-86.2 \pm 0.5$	$74.0 \pm 0.5$	$-84.1 \pm 0.7$	$70.4 \pm 0.6$
1000	$-85.9 \pm 0.5$	$73.6 \pm 0.5$	$-84.1 \pm 0.5$	$70.4 \pm 0.5$
Average	$-87 \pm 2$	$76 \pm 2$	$-84 \pm 2$	$72 \pm 2$

\*Buffers from 100 to 1000 mM Na<sup>+</sup> are made with 50 mM  $\text{Na}_2\text{HPO}_4$  with added NaCl to reach the final concentration of Na<sup>+</sup>. The buffer at 50 mM Na<sup>+</sup> is made with 50 mM  $\text{Na}_2\text{HPO}_4$ . All buffers are at pH = 7.00.

\*\* Error bars correspond to the error calculated from the Van't Hoff linear regression (see Fig. S3.9c)





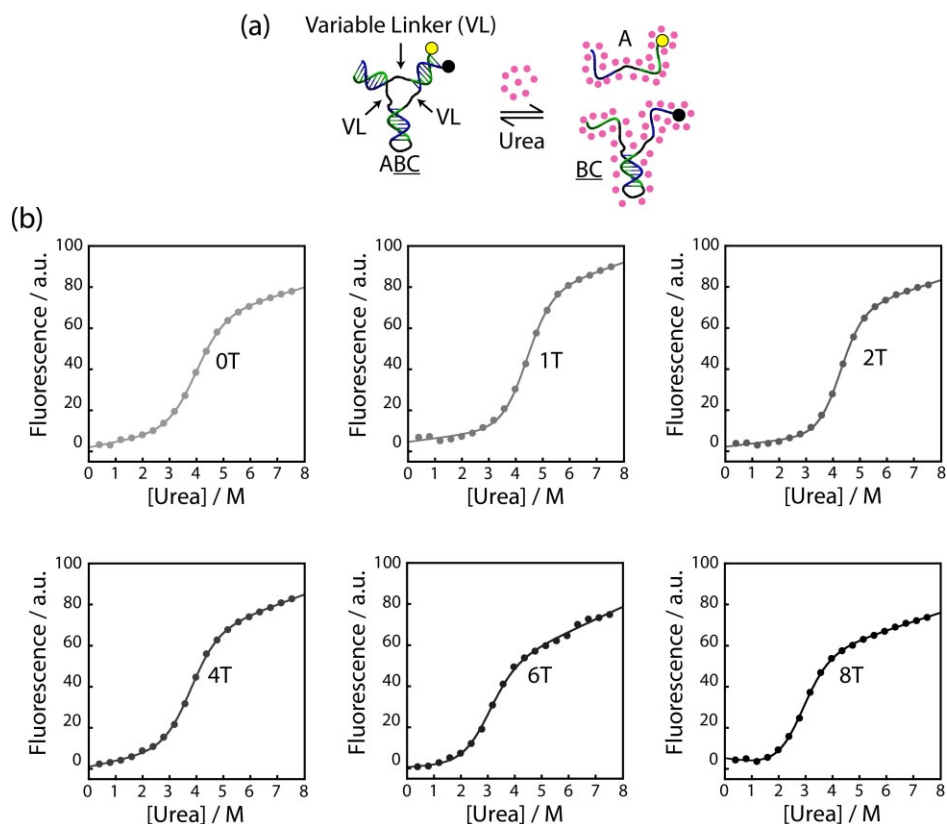
**Figure S3.10 – Thermal denaturation of the trimeric assembly (ABC) built with variable linker length (1T to 8T) at different salt concentrations.** (a) Stability analysis of the trimeric assembly ABC. The yellow circle represents a FAM moiety at the 5' extremity while a black circle represents a BHQ-1 moiety at the 3' extremity. (b) Normalized thermal denaturation at 1  $\mu$ M of each strand. See Fig. S3.1 for the normalization procedure. (c) Van't Hoff analysis of the thermal denaturation curves. All experiments were performed in PBS (50 mM  $\text{Na}_2\text{HPO}_4$ , 100 mM NaCl, pH=7.00).  $K_{\text{Ass}}$  represents the association constant of assembly. These experiments were not performed in triplicate.

**Table S3.4. – The  $\Delta H_{\text{Ass}}$  (in  $\text{kcal}\cdot\text{mol}^{-1}$ ) and the  $-\Delta S_{\text{Ass}}$  (in  $\text{kcal}\cdot\text{mol}^{-1}$ ) of the 3-way junction assemblies are strongly dependent on the ionic strength (see Fig. 3.4b).**

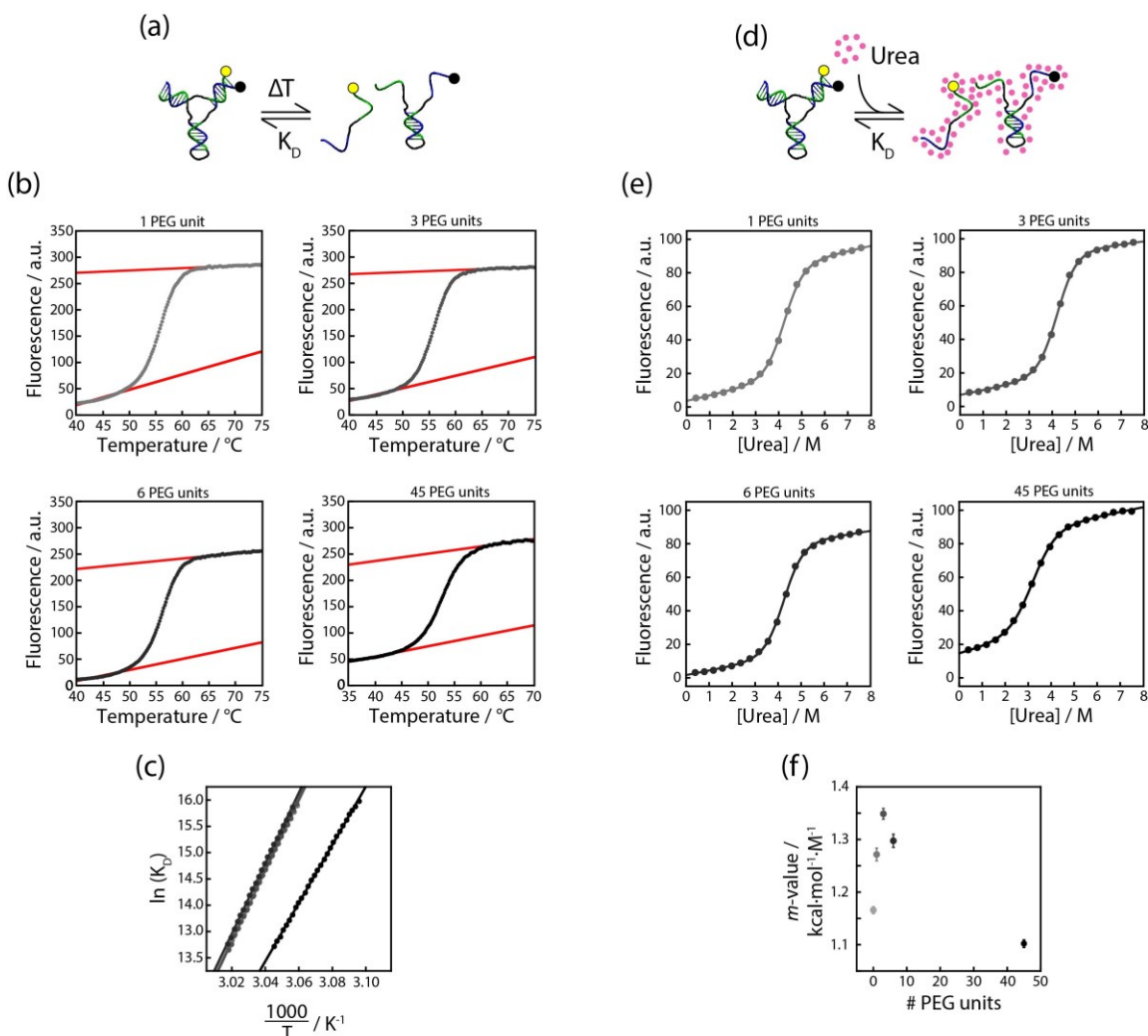
[Na <sup>+</sup> ]	1T		2T		4T		6T		8T	
	$\Delta H_{\text{Ass}}$	$-\Delta S_{\text{Ass}}$	$\Delta H_{\text{Ass}}$	$-\Delta S_{\text{Ass}}$	$\Delta H_{\text{Ass}}$	$-\Delta S_{\text{Ass}}$	$\Delta H_{\text{Ass}}$	$-\Delta S_{\text{Ass}}$	$\Delta H_{\text{Ass}}$	$-\Delta S_{\text{Ass}}$
50	-141 $\pm 1$	125 $\pm 1$	-146 $\pm 1$	130 $\pm 1$	-131 $\pm 2$	115 $\pm 2$	-116 $\pm 2$	100 $\pm 2$	-105 $\pm 2$	89 $\pm 2$
100	-138 $\pm 2$	120 $\pm 2$	-148 $\pm 1$	130 $\pm 1$	-135 $\pm 2$	117 $\pm 1$	-119 $\pm 2$	102 $\pm 2$	-112 $\pm 1$	94 $\pm 1$
200	-140 $\pm 2$	120 $\pm 2$	-155 $\pm 2$	135 $\pm 2$	-142 $\pm 2$	122 $\pm 2$	-128 $\pm 2$	109 $\pm 2$	-120 $\pm 2$	100 $\pm 2$
250	-139 $\pm 2$	118 $\pm 2$	-153 $\pm 1$	132 $\pm 1$	-141 $\pm 2$	121 $\pm 2$	-127 $\pm 2$	107 $\pm 2$	-121 $\pm 2$	101 $\pm 2$
500	-137 $\pm 1$	114 $\pm 1$	-152 $\pm 1$	129 $\pm 1$	-143 $\pm 2$	121 $\pm 2$	-131 $\pm 2$	109 $\pm 2$	-123 $\pm 2$	101 $\pm 2$
750	-138 $\pm 1$	115 $\pm 1$	-155 $\pm 1$	131 $\pm 1$	-146 $\pm 1$	123 $\pm 1$	-136 $\pm 2$	113 $\pm 2$	-128 $\pm 2$	106 $\pm 2$
1000	-138 $\pm 1$	115 $\pm 1$	-158 $\pm 1$	133 $\pm 1$	-149 $\pm 1$	126 $\pm 1$	-140 $\pm 2$	117 $\pm 1$	-132 $\pm 2$	109 $\pm 2$

\*Buffers from 100 to 1000 mM Na<sup>+</sup> are made with 50 mM  $\text{Na}_2\text{HPO}_4$  with added NaCl to reach the final concentration of Na<sup>+</sup>. The buffer at 50 mM Na<sup>+</sup> is made with 50 mM  $\text{Na}_2\text{HPO}_4$ . All buffers are at pH = 7.00.

\*\* Error bars correspond to the error calculated from the Van't Hoff linear regression (see Fig. S3.10c)



**Figure S3.11 – Urea denaturation of the  $K_D$  equilibrium reveals the role of the linker in the capacity of the system to bury more surfaces in the final assembly.** (a) Urea denaturation of the  $K_D$  equilibrium. (b) Raw data of the urea denaturation used to extract the  $m$ -value, a parameter proportional to accessible surface area ( $\Delta ASA$ )<sup>47, 66</sup>. The analysis of the urea denaturation curves reveals the highest  $m$ -value for the 2T system and the lowest  $m$ -value for the 0T system:  $m_{0T} = 1.166 \pm 0.007$ ,  $m_{1T} = 1.34 \pm 0.01$ ,  $m_{2T} = 1.481 \pm 0.009$ ,  $m_{4T} = 1.235 \pm 0.009$ ,  $m_{6T} = 1.21 \pm 0.03$ , and  $m_{8T} = 1.25 \pm 0.01$  kcal·mol<sup>-1</sup>·M<sup>-1</sup>. The error bars represent the error from the nonlinear regression<sup>47</sup>. All urea denaturation experiments were performed by diluting a 8 M urea solution of 50 nM DNA nanosystem (50 nM strand A, 500 nM strand BC) in PBS (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 100 mM NaCl, pH = 7.00) with a similar solution containing no urea (see A. Idili et al. for the detail procedure)<sup>47</sup>. All experiments were performed at 37°C.



**Figure S3.12 – Thermal and urea denaturation of the  $K_D$  equilibrium of 3-way junction built using PEG linkers.** (a) Thermal denaturation of strand A bound to the preformed dimer BC ( $K_D$ ) using variable PEG linkers (VL) ranging from 1, 3, 6 and 45 units. (b) Thermal denaturation curves of PEG linker system. Data are normalized through the linear regression of the upper baseline ( $\sim 15^\circ\text{C}$  above the  $T_m$ ) and the lower baseline ( $\sim 15^\circ\text{C}$  below the  $T_m$ ). The baselines are shown in red. (c) Van't Hoff analysis of the thermal denaturation curves to extract  $\Delta G K_D$  in Fig. 3.4d:  $\Delta G_{1\text{PEG}} = -15.75 \pm 0.06$ ,  $\Delta G_{3\text{PEG}} = -16.1 \pm 0.1$ ,  $\Delta G_{6\text{PEG}} = -15.9 \pm 0.1$ , and  $\Delta G_{45\text{PEG}} = -13.8 \pm 0.1$  kcal·mol $^{-1}$ . The  $\Delta G$  and their standard deviations are calculated from three independent measurements. (d) Urea denaturation of the  $K_D$  equilibrium gives information about the accessible surface area upon assembly ( $\Delta\text{ASA}$ )<sup>47, 66</sup>. (e) Urea denaturation used to extract the  $m$ -values; a parameter proportional to accessible surface area ( $\Delta\text{ASA}$ ). (f) The analysis of the urea denaturation curves reveals the highest  $m$ -value for the 3-PEG system:  $m_{1\text{PEG}} = 1.27 \pm 0.01$ ,  $m_{3\text{PEG}} = 1.35 \pm 0.01$ ,  $m_{6\text{PEG}} = 1.30 \pm 0.01$ , and  $m_{45\text{PEG}} = 1.102 \pm 0.007$  kcal·mol $^{-1}$ ·M $^{-1}$ . The error bars represent the error from the nonlinear regression<sup>47</sup>. All experiments were performed in PBS (50 mM Na $_2$ HPO $_4$ , 100 mM NaCl, pH=7.00) following the same procedure as polythymine linkers.

**Table S3.5. – Thermal and urea denaturation parameters of polythymine linker systems.**

	<b>0T</b>	<b>1T</b>	<b>2T</b>	<b>4T</b>	<b>6T</b>	<b>8T</b>
<b>K<sub>D</sub> / pM</b>	28 ± 2	3.6 ± 1.0	1.8 ± 0.2	27 ± 5	147 ± 59	170 ± 85
<b>ΔG K<sub>D</sub> / kcal·mol<sup>-1</sup></b>	-14.99 ± 0.04	-16.2 ± 0.2	-16.66 ± 0.08	-15.0 ± 0.1	-14.0 ± 0.2	-13.9 ± 0.3
<b><i>m</i>-value / kcal·mol<sup>-1</sup>·M<sup>-1</sup></b>	1.166 ± 0.007	1.34 ± 0.01	1.481 ± 0.009	1.235 ± 0.009	1.21 ± 0.03	1.25 ± 0.01
<b>ΔG K<sub>D</sub> max / kcal·mol<sup>-1</sup></b>	-16.8 ± 0.2					
<b>Optimal length / nm</b>	1.00 ± 0.07					
<b>Optimal length / nt</b>	1.35 ± 0.09					

\*Thermal denaturation values (K<sub>D</sub> and ΔG K<sub>D</sub>) and their standard deviations were calculated from three independent measurements.

\*\*Urea denaturation values (*m*-values) and their standard deviations were calculated from the fit of one individual experiment. In the fitting process, the ΔG obtained from the thermal denaturation were used.

\*\*\* The maximal ΔG K<sub>D</sub> and the optimal length was estimated from the Morse potential fitting.

**Table S3.6. – Thermal and urea denaturation parameters of PEG linker systems.**

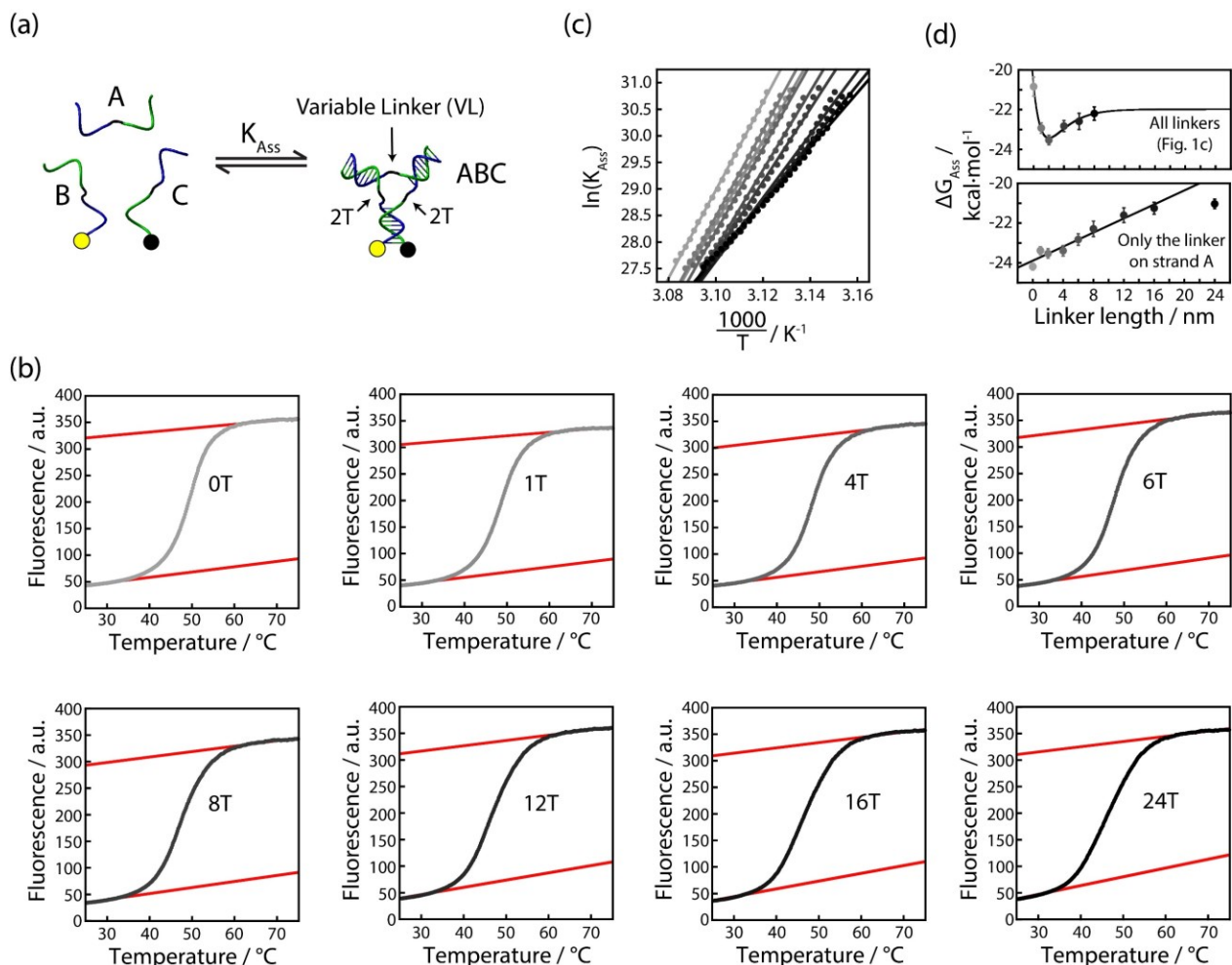
	<b>0PEG</b>	<b>1PEG</b>	<b>3PEG</b>	<b>6PEG</b>	<b>45PEG</b>
<b>K<sub>D</sub> / pM</b>	28 ± 2	8 ± 2	4.9 ± 0.7	7 ± 1	176 ± 70
<b>ΔG K<sub>D</sub> / kcal·mol<sup>-1</sup></b>	-14.99 ± 0.1	-15.75 ± 0.06	-16.1 ± 0.1	-15.9 ± 0.1	-13.8 ± 0.1
<b><i>m</i>-value / kcal·mol<sup>-1</sup>·M<sup>-1</sup></b>	1.166 ± 0.007	1.27 ± 0.01	1.35 ± 0.01	1.30 ± 0.01	1.102 ± 0.007
<b>ΔG K<sub>D</sub> max / kcal·mol<sup>-1</sup></b>	-16.1 ± 0.1				
<b>Optimal length / nm</b>	1.34 ± 0.03				
<b>Optimal length / PEG units</b>	3.4 ± 0.1				

\*Thermal denaturation values (K<sub>D</sub> and ΔG K<sub>D</sub>) and their standard deviations are calculated from three independent measurements.

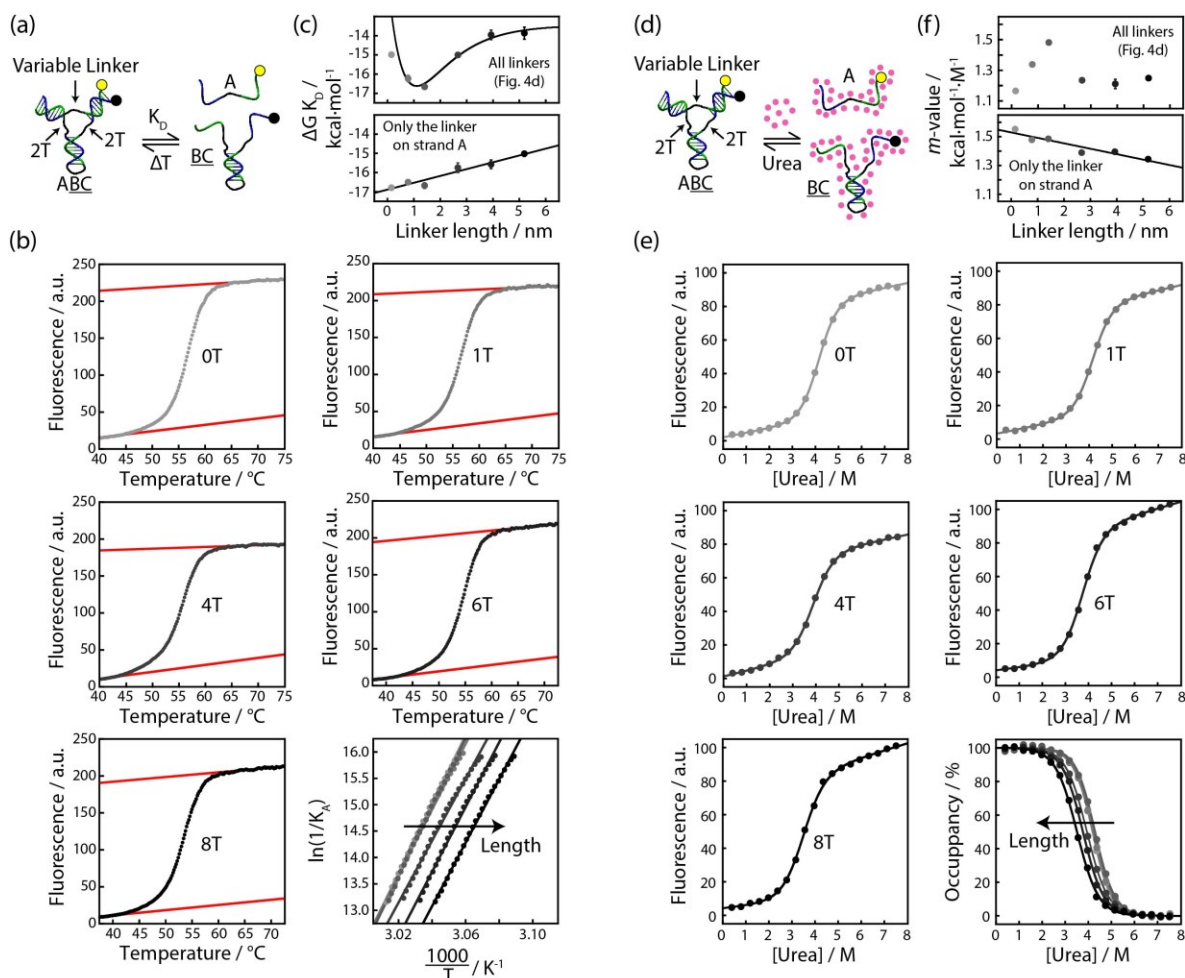
\*\*Urea denaturation values (*m*-values) and their standard deviations are calculated from the fit of one individual experiment. In the fitting process, the ΔG obtained from the thermal denaturation were used.

\*\*\* The maximal ΔG K<sub>D</sub> and the optimal length was estimated from the Morse potential fitting.

\*\*\*\*0PEG is the same system as the OT system.

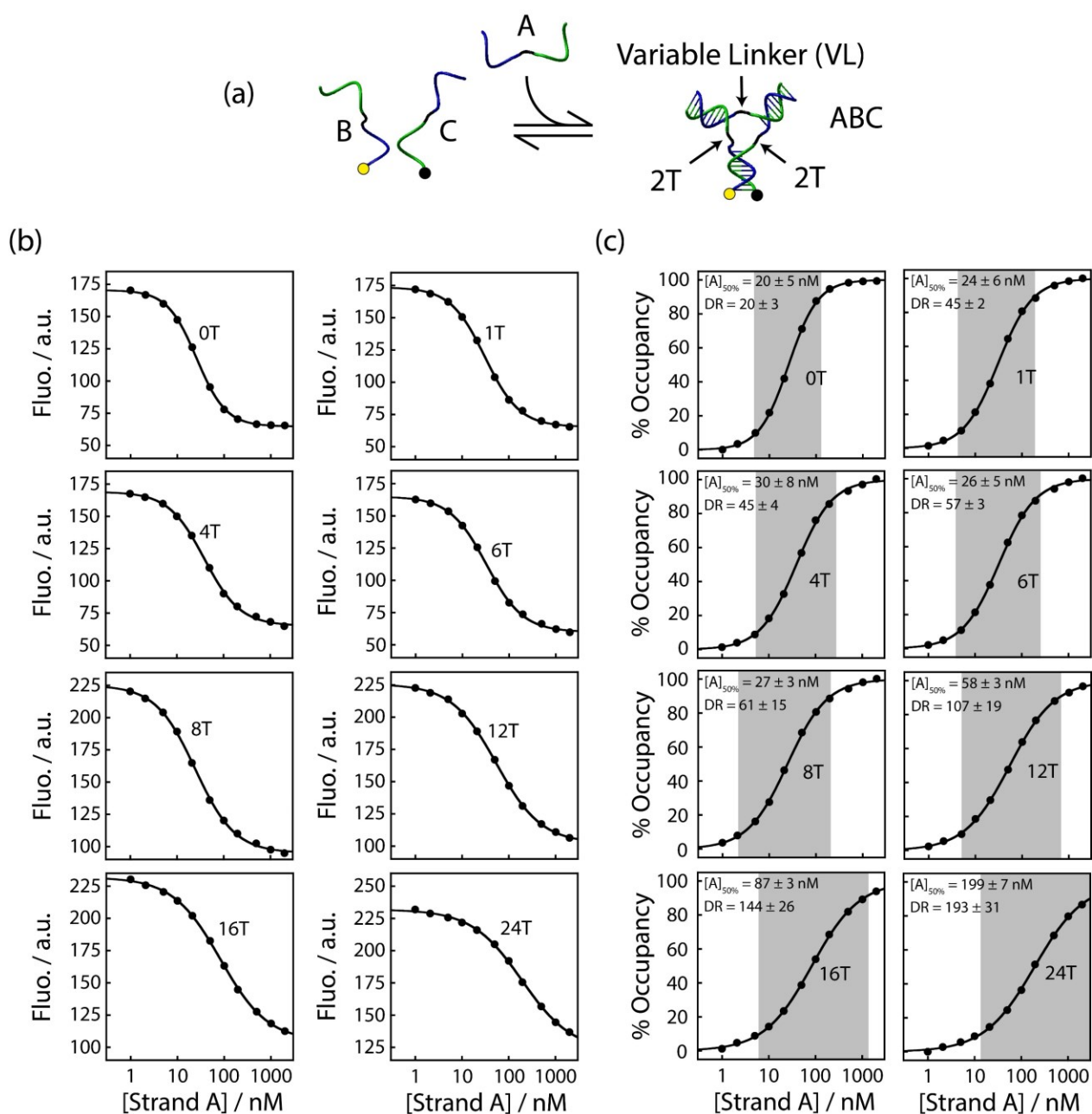


**Figure S3.13 – Thermal denaturation of the trimeric assembly (ABC) built by varying one linker length from 0T to 8T (strand A) while keeping the length of the linker on strand B and C at 2T.** (a) Stability analysis of the trimeric assembly ABC. The yellow circle represents a FAM moiety at the 5' extremity while a black circle represents a BHQ-1 moiety at the 3' extremity. (b) Thermal denaturation at 1  $\mu$ M of each strand. Raw data are normalized by two linear regressions corresponding to the lower and upper baselines (red lines). (c) Van't Hoff analysis of the thermal denaturation curves. (d) The stability of each system varies linearly when increasing the linker length ( $\Delta G_{Ass,0T} = -24.2 \pm 0.2$ ,  $\Delta G_{Ass,1T} = -23.4 \pm 0.2$ ,  $\Delta G_{Ass,2T} = -23.5 \pm 0.2$ ,  $\Delta G_{Ass,4T} = -23.4 \pm 0.3$ ,  $\Delta G_{Ass,6T} = -22.8 \pm 0.3$ ,  $\Delta G_{Ass,8T} = -22.3 \pm 0.4$ ,  $\Delta G_{Ass,12T} = -21.6 \pm 0.4$ ,  $\Delta G_{Ass,16T} = -21.3 \pm 0.3$  and  $\Delta G_{Ass,24T} = -21.0 \pm 0.2$  kcal·mol<sup>-1</sup>). The error bars represent the error from the linear regression of the Van't Hoff analysis. Of note, the stability starts plateauing when employing longer linkers (see 24T linker). All experiments were performed in PBS (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 100 mM NaCl, pH=7.00).  $K_{Ass}$  represents the association constant of the trimeric assembly.

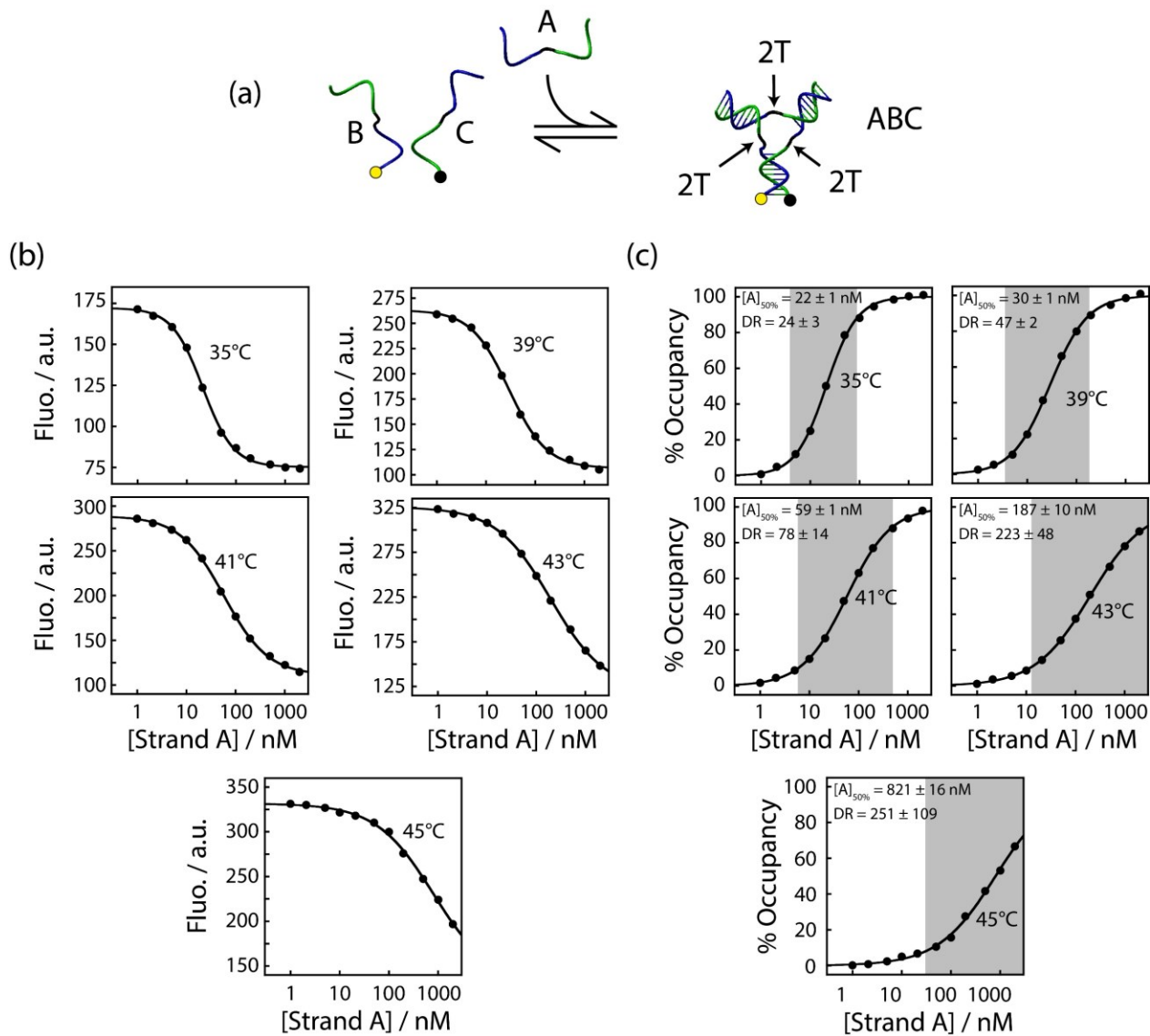


**Figure S3.14 – Changing only the linker on strand A while keeping the two others at 2T provides a straightforward way to program the level of compaction of the assembly.** Thermal denaturation of the  $K_D$  equilibrium with variable linkers on strand A (0T to 8T) while keeping the BC dimer with 2T linkers. (b) Thermal denaturation curves of poly-thymine linker system. Data are normalized through the linear regression of the upper baseline ( $\sim 15^{\circ}\text{C}$  above the  $T_m$ ) and the lower baseline ( $\sim 15^{\circ}\text{C}$  below the  $T_m$ ). The baselines are shown in red. Van't Hoff analysis of the thermal denaturation curves are shown in the lower right panel. (c) The stability decreases when increasing the linker length:  $\Delta G_{0T} = -16.8 \pm 0.2$ ,  $\Delta G_{1T} = -16.5 \pm 0.2$ ,  $\Delta G_{2T} = -16.66 \pm 0.08$ ,  $\Delta G_{4T} = -15.8 \pm 0.3$ ,  $\Delta G_{6T} = -15.6 \pm 0.2$  and  $\Delta G_{8T} = -15.0 \pm 0.2$  kcal $\cdot\text{mol}^{-1}$ . The error bars represent the error from the linear regression of the Van't Hoff analysis. (d) Urea denaturation of the same equilibrium. (e) Raw data of the urea denaturation used to extract the  $m$ -values. (f) The  $m$ -value decreases when increasing the linker length:  $m_{0T} = 1.552 \pm 0.006$ ,  $m_{1T} = 1.478 \pm 0.007$ ,  $m_{2T} = 1.481 \pm 0.009$ ,  $m_{4T} = 1.389 \pm 0.009$ ,  $m_{6T} = 1.395 \pm 0.015$ , and  $m_{8T} = 1.343 \pm 0.013$  kcal $\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ . The error bars represent the error from the nonlinear regression<sup>47</sup>. For thermal and urea denaturation, we used 50 nM of fluorescently labelled strand A (FAM at the 5' extremity) and a 10-fold excess (500 nM) of preformed dimers BC containing a quencher (BHQ-1 at the 3' extremity). All denaturation experiments were performed in PBS (50 mM  $\text{Na}_2\text{HPO}_4$ , 100 mM NaCl, pH=7.00). Urea denaturation experiments were performed at  $37^{\circ}\text{C}$ .



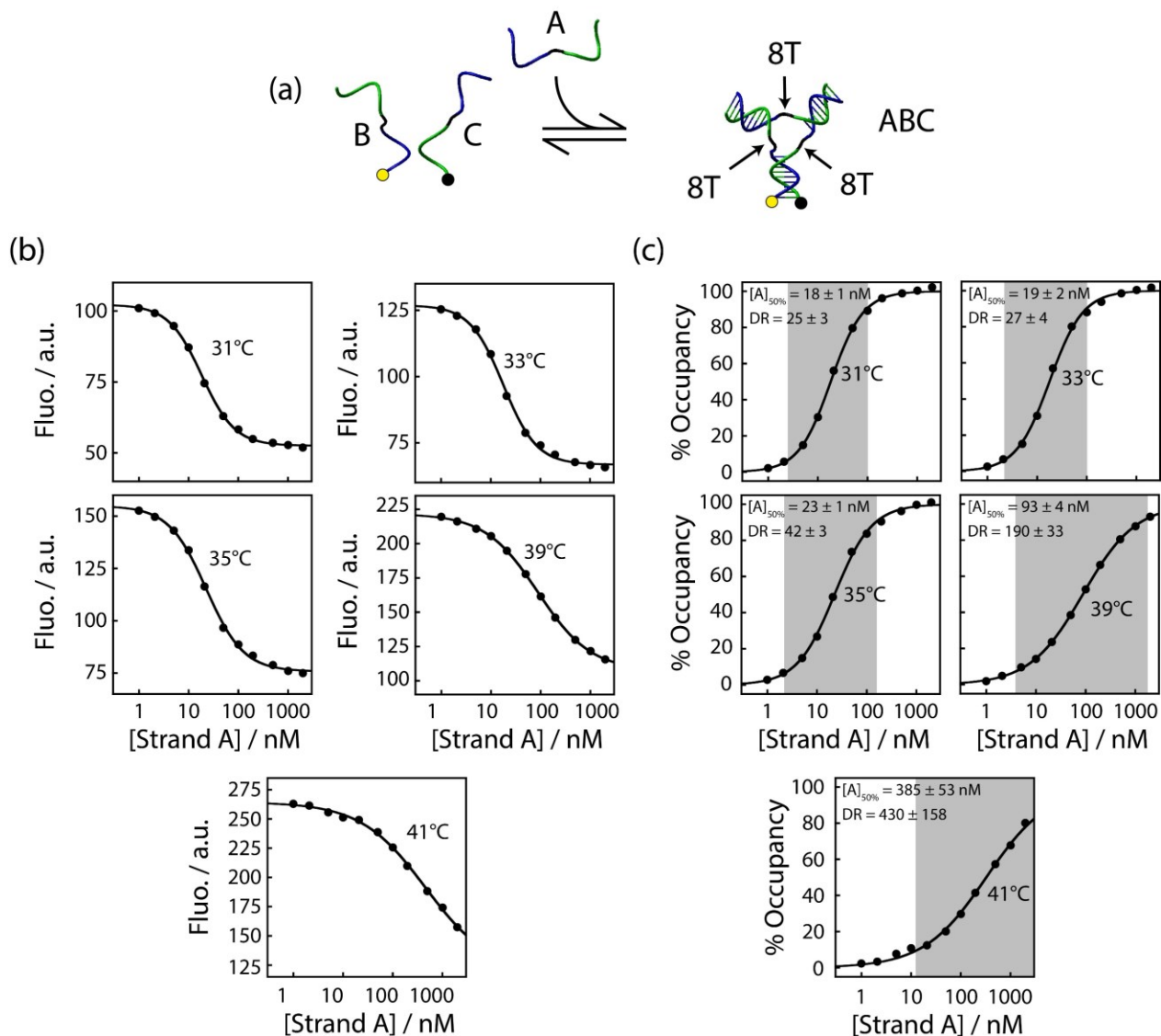


**Figure S3.15 – The trimeric assembly of nanosystems built with two 2T linkers and one variable linker (0T to 24T).** All assemblies were performed in triplicate (only one data set is shown). The trimeric assemblies were performed using 30 nM of fluorescently labelled strand B and 30 nM of strand C labelled with a quencher. The assembly is induced by adding an increasing amount of unlabelled strand A. Yellow circle represents a FAM moiety at the 5' extremity while a black circle represents a BHQ-1 moiety at the 3' extremity. (b) Raw fluorescence binding curves. (c) Normalized binding curves. The  $[A]_{50\%}$  and the dynamic range (DR) of all assemblies represent the average and standard deviation of the triplicate. The gray rectangles illustrate the dynamic range of the assemblies. The data for the 2T-2T-2T system can be found in Fig. S3.2. The experiments were performed at 37°C in PBS (50 mM  $\text{Na}_2\text{HPO}_4$ , 100 mM NaCl, pH=7.00).



**Figure S3.16 – The trimeric assembly of nanosystems built with 2T linkers at variable temperature (35°C to 45°C).** All assemblies were performed in triplicate (only one data set is shown). The trimeric assemblies were performed using 30 nM of fluorescently labelled strand B and 30 nM of strand C labelled with a quencher. The assembly is induced by adding an increasing amount of unlabelled strand A. Yellow circle represents a FAM moiety at the 5' extremity while a black circle represents a BHQ-1 moiety at the 3' extremity. (b) Raw fluorescence binding curves. (c) Normalized binding curves. The  $[A]_{50\%}$  and the dynamic range (DR) of all assemblies represent the average and standard deviation of the triplicate. The gray rectangles illustrate the dynamic range of the assemblies. The data for the 2T-2T-2T at 37°C system can be found in Fig. S3.2. The experiments were performed in PBS (50 mM  $\text{Na}_2\text{HPO}_4$ , 100 mM NaCl, pH=7.00).





**Figure S3.17 – The trimeric assembly of nanosystems built with 8T linkers at variable temperature (31°C to 41°C).** All assemblies were performed in triplicate (only one data set is shown). The trimeric assemblies were performed using 30 nM of fluorescently labelled strand B and 30 nM of strand C labelled with a quencher. The assembly is induced by adding an increasing amount of unlabelled strand A. Yellow circle represents a FAM moiety at the 5' extremity while a black circle represents a BHQ-1 moiety at the 3' extremity. (b) Raw fluorescence binding curves. (c) Normalized binding curves. The  $[A]_{50\%}$  and the dynamic range (DR) of all assemblies represent the average and standard deviation of the triplicate. The gray rectangles illustrate the dynamic range of the assemblies. The data for the 8T-8T-8T at 37°C system can be found in Fig. S3.2. The experiments were performed in PBS (50 mM  $\text{Na}_2\text{HPO}_4$ , 100 mM NaCl, pH=7.00).

### 3.8 References

1. Lynch, M., The evolution of multimeric protein assemblages. *Mol. Biol. Evol.* **2012**, *29* (5), 1353-1366.
2. Mattia, E.; Otto, S., Supramolecular systems chemistry. *Nat. Nanotechnol.* **2015**, *10* (2), 111-119.
3. Seeman, N. C.; Sleiman, H. F., DNA nanotechnology. *Nat. Rev. Mater.* **2017**, *3* (1), 17068.
4. Raymond, D. M.; Nilsson, B. L., Multicomponent peptide assemblies. *Chem. Soc. Rev.* **2018**, *47* (10), 3659-3720.
5. Marianayagam, N. J.; Sunde, M.; Matthews, J. M., The power of two: protein dimerization in biology. *Trends Biochem. Sci.* **2004**, *29* (11), 618-625.
6. Hagner, K.; Setayeshgar, S.; Lynch, M., Stochastic protein multimerization, activity, and fitness. *Phys. Rev. E* **2018**, *98* (6), 062401.
7. Bou-Nader, C.; Zhang, J., Structural Insights into RNA Dimerization: Motifs, Interfaces and Functions. *Molecules* **2020**, *25* (12).
8. Ali, M. H.; Imperiali, B., Protein oligomerization: How and why. *Bioorg. Med. Chem.* **2005**, *13* (17), 5013-5020.
9. Griffin, M. D. W.; Dobson, R. C. J.; Pearce, F. G.; Antonio, L.; Whitten, A. E.; Liew, C. K.; Mackay, J. P.; Trehwella, J.; Jameson, G. B.; Perugini, M. A.; Gerrard, J. A., Evolution of Quaternary Structure in a Homotetrameric Enzyme. *J. Mol. Biol.* **2008**, *380* (4), 691-703.
10. Whitty, A., Cooperativity and biological complexity. *Nat. Chem. Biol.* **2008**, *4* (8), 435-439.
11. Carlson, J. C. T.; Jena, S. S.; Flenniken, M.; Chou, T.-f.; Siegel, R. A.; Wagner, C. R., Chemically Controlled Self-Assembly of Protein Nanorings. *J. Am. Chem. Soc.* **2006**, *128* (23), 7630-7638.
12. van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; Kim, P. M.; Kriwacki, R. W.; Oldfield, C. J.; Pappu, R. V.; Tompa, P.; Uversky, V. N.; Wright, P. E.; Babu, M. M., Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **2014**, *114* (13), 6589-6631.
13. Lim, W. A., Designing customized cell signalling circuits. *Nat. Rev. Mol. Cell Biol.* **2010**, *11* (6), 393-403.

14. Bai, Y.; Luo, Q.; Liu, J., Protein self-assembly via supramolecular strategies. *Chem. Soc. Rev.* **2016**, *45* (10), 2756-2767.
15. Chen, X.; Zaro, J. L.; Shen, W.-C., Fusion protein linkers: Property, design and functionality. *Adv. Drug Delivery Rev.* **2013**, *65* (10), 1357-1369.
16. Wang, Y.; Cheetham, A. G.; Angacian, G.; Su, H.; Xie, L.; Cui, H., Peptide–drug conjugates as effective prodrug strategies for targeted delivery. *Adv. Drug Delivery Rev.* **2017**, *110-111*, 112-126.
17. Mariottini, D.; Idili, A.; Nijenhuis, M. A. D.; Ercolani, G.; Ricci, F., Entropy-Based Rational Modulation of the pKa of a Synthetic pH-Dependent Nanoswitch. *J. Am. Chem. Soc.* **2019**, *141* (29), 11367-11371.
18. Mariottini, D.; Idili, A.; Nijenhuis, M. A. D.; de Greef, T. F. A.; Ricci, F., DNA-Based Nanodevices Controlled by Purely Entropic Linker Domains. *J. Am. Chem. Soc.* **2018**, *140* (44), 14725-14734.
19. Ma, M. Y. X.; McCallum, K.; Climie, S. C.; Kuperman, R.; Lin, W. C.; Sumner-Smith, M.; Barnett, R. W., Design and synthesis of RNA miniduplexes via a synthetic linker approach. 2. Generation of covalently closed, double-stranded cyclic HIV-1 TAR RNA analogs with high Tat-binding affinity. *Nucleic Acids Res.* **1993**, *21* (11), 2585-2589.
20. Zhou, J.; Swiderski, P.; Li, H.; Zhang, J.; Neff, C. P.; Akkina, R.; Rossi, J. J., Selection, characterization and application of new RNA HIV gp 120 aptamers for facile delivery of Dicer substrate siRNAs into HIV infected cells. *Nucleic Acids Res.* **2009**, *37* (9), 3094-3109.
21. von Krbek, L. K. S.; Achazi, A. J.; Solleder, M.; Weber, M.; Paulus, B.; Schalley, C. A., Allosteric and Chelate Cooperativity in Divalent Crown Ether/Ammonium Complexes with Strong Binding Enhancement. *Chem. Eur. J.* **2016**, *22* (43), 15475-15484.
22. Idili, A.; Vallée-Bélisle, A.; Ricci, F., Programmable pH-Triggered DNA Nanoswitches. *J. Am. Chem. Soc.* **2014**, *136* (16), 5836-5839.
23. Dembska, A.; Bielecka, P.; Juskowiak, B., pH-Sensing fluorescence oligonucleotide probes based on an i-motif scaffold: a review. *Anal. Methods* **2017**, *9* (43), 6092-6106.
24. Kocak, G.; Tuncer, C.; Bütün, V., pH-Responsive polymers. *Polym. Chem.* **2017**, *8* (1), 144-176.

25. Askarieh, G.; Hedhammar, M.; Nordling, K.; Saenz, A.; Casals, C.; Rising, A.; Johansson, J.; Knight, S. D., Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. *Nature* **2010**, *465* (7295), 236-238.
26. Thompson, I. A. P.; Zheng, L.; Eisenstein, M.; Soh, H. T., Rational design of aptamer switches with programmable pH response. *Nat. Commun.* **2020**, *11* (1), 2946.
27. Porchetta, A.; Idili, A.; Vallée-Bélisle, A.; Ricci, F., General Strategy to Introduce pH-Induced Allostery in DNA-Based Receptors to Achieve Controlled Release of Ligands. *Nano Lett.* **2015**, *15* (7), 4467-4471.
28. Voutilainen, S.; Paananen, A.; Lille, M.; Linder, M. B., Modular protein architectures for pH-dependent interactions and switchable assembly of nanocellulose. *Int. J. Biol. Macromol.* **2019**, *137*, 270-276.
29. Dueber, J. E.; Mirsky, E. A.; Lim, W. A., Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nat. Biotechnol.* **2007**, *25* (6), 660-662.
30. Simon, A. J.; Vallée-Bélisle, A.; Ricci, F.; Plaxco, K. W., Intrinsic disorder as a generalizable strategy for the rational design of highly responsive, allosterically cooperative receptors. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (42), 15048-15053.
31. Gokhale, R. S.; Khosla, C., Role of linkers in communication between protein modules. *Curr. Opin. Chem. Biol.* **2000**, *4* (1), 22-27.
32. Shuker Suzanne, B.; Hajduk Philip, J.; Meadows Robert, P.; Fesik Stephen, W., Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996**, *274* (5292), 1531-1534.
33. von Krbek, L. K. S.; Schalley, C. A.; Thordarson, P., Assessing cooperativity in supramolecular systems. *Chem. Soc. Rev.* **2017**, *46* (9), 2622-2637.
34. Williamson, J. R., Cooperativity in macromolecular assembly. *Nat. Chem. Biol.* **2008**, *4* (8), 458-465.
35. Harroun, S. G.; Prévost-Tremblay, C.; Lauzon, D.; Desrosiers, A.; Wang, X.; Pedro, L.; Vallée-Bélisle, A., Programmable DNA switches and their applications. *Nanoscale* **2018**, *10* (10), 4607-4641.
36. Tang, Y.; Ge, B.; Sen, D.; Yu, H.-Z., Functional DNA switches: rational design and electrochemical signaling. *Chem. Soc. Rev.* **2014**, *43* (2), 518-529.

37. Wang, F.; Liu, X.; Willner, I., DNA Switches: From Principles to Applications. *Angew. Chem. Int. Ed.* **2015**, *54* (4), 1098-1129.
38. Lauzon, D.; Vallée-Bélisle, A., Functional advantages of building nanosystems using multiple molecular components. *Nat. Chem.* **2023**.
39. Yuan, Y.; Tam, M. F.; Simplaceanu, V.; Ho, C., New Look at Hemoglobin Allostery. *Chem. Rev.* **2015**, *115* (4), 1702-1724.
40. Portoghese, P. S., Bivalent ligands and the message-address concept in the design of selective opioid receptor antagonists. *Trends Pharmacol. Sci.* **1989**, *10* (6), 230-235.
41. Cockroft, S. L.; Hunter, C. A., Chemical double-mutant cycles: dissecting non-covalent interactions. *Chem. Soc. Rev.* **2007**, *36* (2), 172-188.
42. Ha, S. H.; Ferrell, J. E., Thresholds and ultrasensitivity from negative cooperativity. *Science* **2016**, *352* (6288), 990-993.
43. Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q., Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *Int. J. Mol. Sci.* **2016**, *17* (2).
44. Chodera, J. D.; Mobley, D. L., Entropy-Enthalpy Compensation: Role and Ramifications in Biomolecular Ligand Recognition and Design. *Annu. Rev. Biophys.* **2013**, *42* (1), 121-142.
45. Myers, J. K.; Nick Pace, C.; Martin Scholtz, J., Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **1995**, *4* (10), 2138-2148.
46. Shelton, V. M.; Sosnick, T. R.; Pan, T., Applicability of Urea in the Thermodynamic Analysis of Secondary and Tertiary RNA Folding. *Biochemistry* **1999**, *38* (51), 16831-16839.
47. Idili, A.; Ricci, F.; Vallée-Bélisle, A., Determining the folding and binding free energy of DNA-based nanodevices and nanoswitches using urea titration curves. *Nucleic Acids Res.* **2017**, *45* (13), 7571-7580.
48. Muhuri, S.; Mimura, K.; Miyoshi, D.; Sugimoto, N., Stabilization of Three-Way Junctions of DNA under Molecular Crowding Conditions. *J. Am. Chem. Soc.* **2009**, *131* (26), 9268-9280.
49. Nakano, S.-i.; Miyoshi, D.; Sugimoto, N., Effects of Molecular Crowding on the Structures, Interactions, and Functions of Nucleic Acids. *Chem. Rev.* **2014**, *114* (5), 2733-2758.

50. Sabir, T.; Toulmin, A.; Ma, L.; Jones, A. C.; McGlynn, P.; Schröder, G. F.; Magennis, S. W., Branchpoint Expansion in a Fully Complementary Three-Way DNA Junction. *J. Am. Chem. Soc.* **2012**, *134* (14), 6280-6285.
51. Klein, J. S.; Jiang, S.; Galimidi, R. P.; Keeffe, J. R.; Bjorkman, P. J., Design and characterization of structured protein linkers with differing flexibilities. *Protein Eng., Des. Sel.* **2014**, *27* (10), 325-330.
52. Sørensen Charlotte, S.; Kjaergaard, M., Effective concentrations enforced by intrinsically disordered linkers are governed by polymer physics. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (46), 23124-23131.
53. van Rosmalen, M.; Krom, M.; Merkx, M., Tuning the Flexibility of Glycine-Serine Linkers To Allow Rational Design of Multidomain Proteins. *Biochemistry* **2017**, *56* (50), 6565-6574.
54. Sørensen, C. S.; Jendroszek, A.; Kjaergaard, M., Linker Dependence of Avidity in Multivalent Interactions Between Disordered Proteins. *J. Mol. Biol.* **2019**, *431* (24), 4784-4795.
55. Mack, E. T.; Snyder, P. W.; Perez-Castillejos, R.; Bilgiçer, B.; Moustakas, D. T.; Butte, M. J.; Whitesides, G. M., Dependence of Avidity on Linker Length for a Bivalent Ligand–Bivalent Receptor Model System. *J. Am. Chem. Soc.* **2012**, *134* (1), 333-345.
56. Cecchini, C.; Pannilunghi, S.; Tardy, S.; Scapozza, L., From Conception to Development: Investigating PROTACs Features for Improved Cell Permeability and Successful Protein Degradation. *Front. Chem.* **2021**, *9*.
57. Ricci, F.; Vallée-Bélisle, A.; Simon, A. J.; Porchetta, A.; Plaxco, K. W., Using Nature’s “Tricks” To Rationally Tune the Binding Properties of Biomolecular Receptors. *Acc. Chem. Res.* **2016**, *49* (9), 1884-1892.
58. Ricci, F.; Vallée-Bélisle, A.; Porchetta, A.; Plaxco, K. W., Rational Design of Allosteric Inhibitors and Activators Using the Population-Shift Model: In Vitro Validation and Application to an Artificial Biosensor. *J. Am. Chem. Soc.* **2012**, *134* (37), 15177-15180.
59. Churchfield, L. A.; Medina-Morales, A.; Brodin, J. D.; Perez, A.; Tezcan, F. A., De Novo Design of an Allosteric Metalloprotein Assembly with Strained Disulfide Bonds. *J. Am. Chem. Soc.* **2016**, *138* (40), 13163-13166.

60. Pirro, F.; Schmidt, N.; Lincoff, J.; Widel Zachary, X.; Polizzi Nicholas, F.; Liu, L.; Therien Michael, J.; Grabe, M.; Chino, M.; Lombardi, A.; DeGrado William, F., Allosteric cooperation in a de novo-designed two-domain protein. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (52), 33246-33253.
61. Marsh, J. A.; Teichmann, S. A., Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **2015**, *84* (1), 551-575.
62. Björklund, Å. K.; Ekman, D.; Light, S.; Frey-Skött, J.; Elofsson, A., Domain Rearrangements in Protein Evolution. *J. Mol. Biol.* **2005**, *353* (4), 911-923.
63. Perica, T.; Marsh, Joseph A.; Sousa, Filipa L.; Natan, E.; Colwell, Lucy J.; Ahnert, Sebastian E.; Teichmann, Sarah A., The emergence of protein complexes: quaternary structure, dynamics and allostery. *Biochem. Soc. Trans.* **2012**, *40* (3), 475-491.
64. Bashton, M.; Chothia, C., The geometry of domain combination in proteins<sup>11</sup>Edited by J. Thornton. *J. Mol. Biol.* **2002**, *315* (4), 927-939.
65. Wang, Y.; Fisher, J. C.; Mathew, R.; Ou, L.; Otieno, S.; Sublet, J.; Xiao, L.; Chen, J.; Roussel, M. F.; Kriwacki, R. W., Intrinsic disorder mediates the diverse regulatory functions of the Cdk inhibitor p21. *Nat. Chem. Biol.* **2011**, *7* (4), 214-221.
66. Lawrence, C.; Vallée-Bélisle, A.; Pfeil, S. H.; de Mornay, D.; Lipman, E. A.; Plaxco, K. W., A comparison of the folding kinetics of a small, artificially selected DNA aptamer with those of equivalently simple naturally occurring proteins. *Protein Sci.* **2014**, *23* (1), 56-66.

# Chapitre 4 – Multivalent activators: Bioinspired effectors to tune the affinity and dynamic range of molecular switches\*

## 4.1 Abstract

Multivalency can be rationally implemented into molecular switches to program their activation in response to specific concentrations of two molecular inputs. As a proof-of-concept to understand the thermodynamic basis and design principle of this bioinspired mechanism, we have engineered a highly programmable DNA-based switch that can be triggered by a DNA multivalent activator that binds both the switch and the target molecules. Through precise design of the activator binding interface, we show that the affinity, dynamic range, and activated half-life of the molecular switch can be programmed with even more flexibility than classic allosteric activation strategies. The simplicity by which the activation properties of a molecular switch can be rationally tuned through the formation of a multivalent assembly suggests that it may find many applications in biosensing, drug delivery, synthetic biology, and molecular computation, fields where the control over the transduction of binding events into a specific output is key.

---

\* Article (communication) en évaluation par les pairs au *Journal of the American Chemical Society*: Lauzon, D.; Vallée-Bélisle, A. *J. Am. Chem. Soc.* (Manuscript in Revision). D.L. a conceptualisé et réalisé toutes les expériences. D.L. et A.V.B. ont conjointement analysé et interprété les résultats. D.L. a écrit la première version du manuscrit. D.L. et A.V.B. ont conjointement écrit la version finale. A. V. B. a supervisé la réalisation du projet.



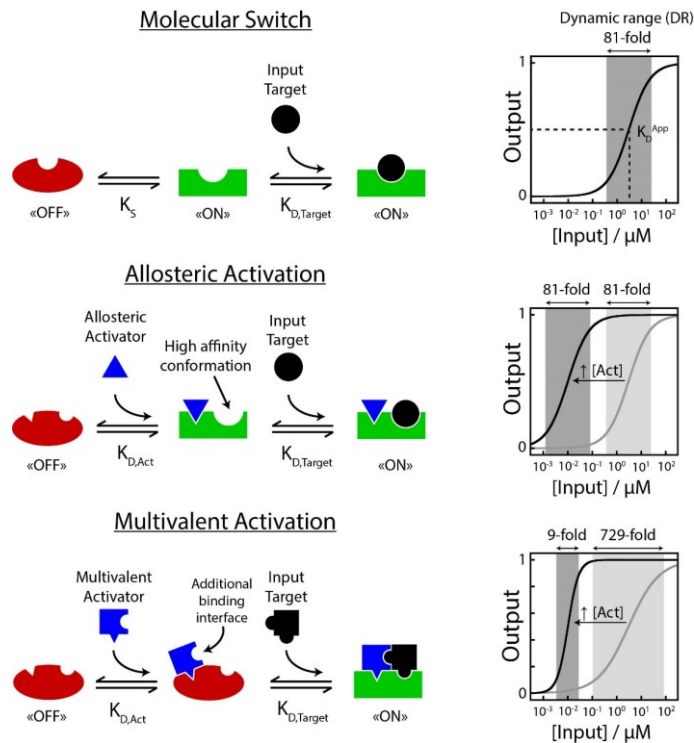
## 4.2 Introduction

Through billions of years of evolution, cells have developed a myriad of finely regulated nanomachines that monitor variations in their surroundings. To respond to these molecular changes, cells rely on molecular switches that exploit different mechanisms to detect and integrate these specific chemical or physical inputs (*e.g.*, temperature, pressure, pH, small molecules, and proteins) into relevant biological activity (Fig. 4.1, *top*)<sup>1</sup>. Allosteric activation, for example, is a mechanism by which a biomolecular switch becomes more sensitive to the presence of a molecular input (target) in the presence of an activator molecule that favors a higher affinity conformation of the switch (Fig. 4.1, *middle*)<sup>2</sup>. A better thermodynamic understanding of this mechanism has recently provided a variety of strategies to optimize the binding affinity<sup>3</sup>, sensitivity<sup>4, 5</sup>, and specificity<sup>6</sup> of artificial nanosystems employed in biosensing<sup>7-11</sup> and drug delivery<sup>11-15</sup>.

The multivalent activation is another useful mechanism by which a biomolecular switch becomes more sensitive to the presence of a molecular input (target) in the presence of an activator molecule that introduces an additional binding interface (Fig. 4.1, *bottom*)<sup>16-18</sup>. Although multivalency has been exploited extensively in nature (*e.g.*, the ribosome, antibodies, or cell surface receptors)<sup>19-21</sup>, little is known about how this mechanism enables programming the affinity or cooperativity of molecular switches. Here, we demonstrate how multivalency, by means of a so-called multivalent activator, allows programming the apparent affinity and dynamic range of molecular switches with even more versatility and simplicity than when employing allosteric activators (Fig. S4.1).

To highlight the potential of multivalent activation mechanisms to regulate molecular switches, we exploited the high programmability and versatility of DNA–DNA interactions. DNA offers the advantage of being easy to synthesize and modify with chemical moieties (*e.g.*, fluorophores, quenchers, redox elements, and photoactive elements)<sup>22, 23</sup> while having highly predictable interactions enabling the construction of precise, well-defined nanostructures<sup>24</sup>. For these reasons, DNA-based systems have been used to establish the design rules behind structure

switching receptors<sup>25</sup> and to program molecular switches through allostery<sup>3</sup> and ultrasensitivity<sup>26</sup>,  
27.



**Figure 4.1 – Engineering bioinspired multivalent activators for controlling molecular switches.** *Top.* Target-binding activation of a molecular switch through a structure-switching mechanism<sup>25</sup>. The target binds the high-affinity “ON-state” shifting the switch equilibrium proportionally between the “OFF-” and “ON-state”. *Middle.* A molecular switch regulated via a classic allosteric activation mechanism<sup>2</sup>, where an allosteric activator binds and favors the “ON-state” of the molecular switch, thus improving its affinity for the target (see arrow)<sup>3</sup>. *Bottom.* A molecular switch regulated via a multivalent activation mechanism where a multivalent activator binds and introduces an additional interacting interface near the molecular switch, thus improving its affinity for the target (see arrow). In addition to providing a means by which the apparent  $K_D$  for the target can be programmed toward lower concentration, this so-called multivalent activator should in principle enable the programming of the switch dynamic range (see Fig. S4.1).

### 4.3 Results

To understand the thermodynamic and kinetic basis of the multivalent activation mechanism, we re-engineered a DNA-based switch designed by Plaxco et al. that has been previously employed to recreate the allosteric activation mechanism (Fig. S4.2)<sup>3</sup>. This model switch that supports both mechanisms (i.e., allostery and multivalency) will further enable efficient comparison between the performance of both activation strategies. The DNA-based

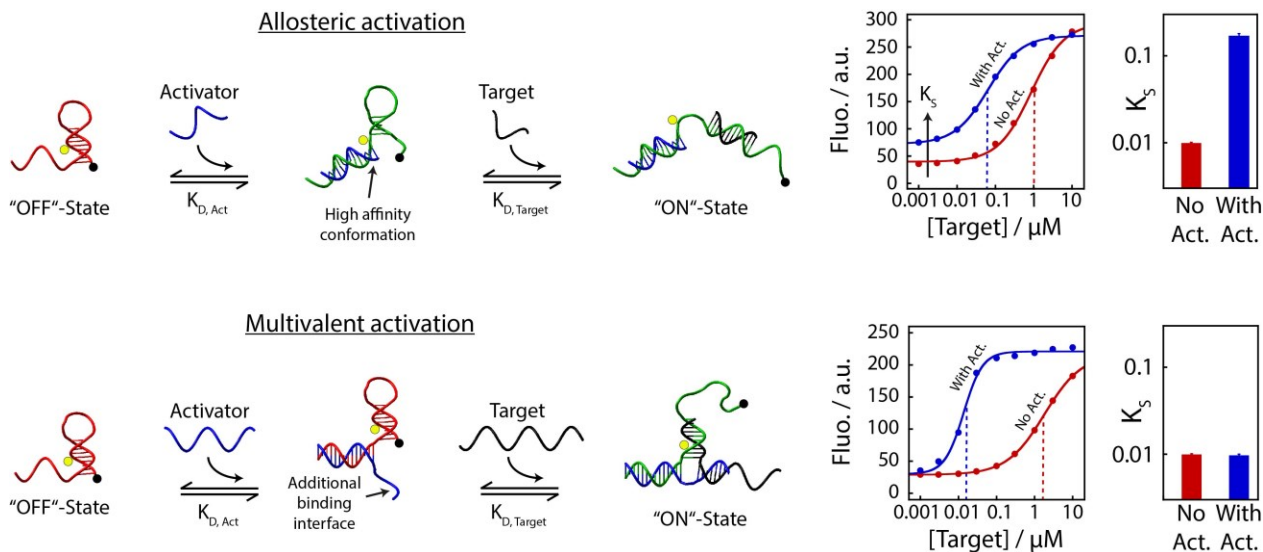
switch consists of a fluorescently labeled stem–loop<sup>28</sup> that contains a fluorophore and a quencher located next to each other when the switch conformation is in the “OFF state” in the absence of a target. The addition of a DNA target strand complementary to the loop of this switch disrupts the stem and separates the fluorophore and the quencher thus providing a fluorescent signal, which can then be quantified.

In the case of the allosteric activation mechanism, the activators were designed to invade between 2 and 6 base pairs in the stem by interacting with the “tail” added near the stem (Fig. 4.2, *top*). This invasion by the activator contributes to destabilizing the “OFF-state” of the stem–loop and to increasing the switching equilibrium  $K_S$  (Fig. S4.3), thus favoring the high-affinity state by reducing the energetic penalty involved with the binding of the target to the molecular beacon<sup>25</sup>. To recreate a multivalent activator mechanism, we designed multivalent activators to interact with the “tail” of the switch. However, instead of favoring the “ON-state” of the switch, the multivalent activator simply provides an additional interacting interface for the target (Fig. 2, *bottom*). This new interacting interface increases the affinity of the target for the switch through the formation of a three-way junction assembly<sup>29</sup>.

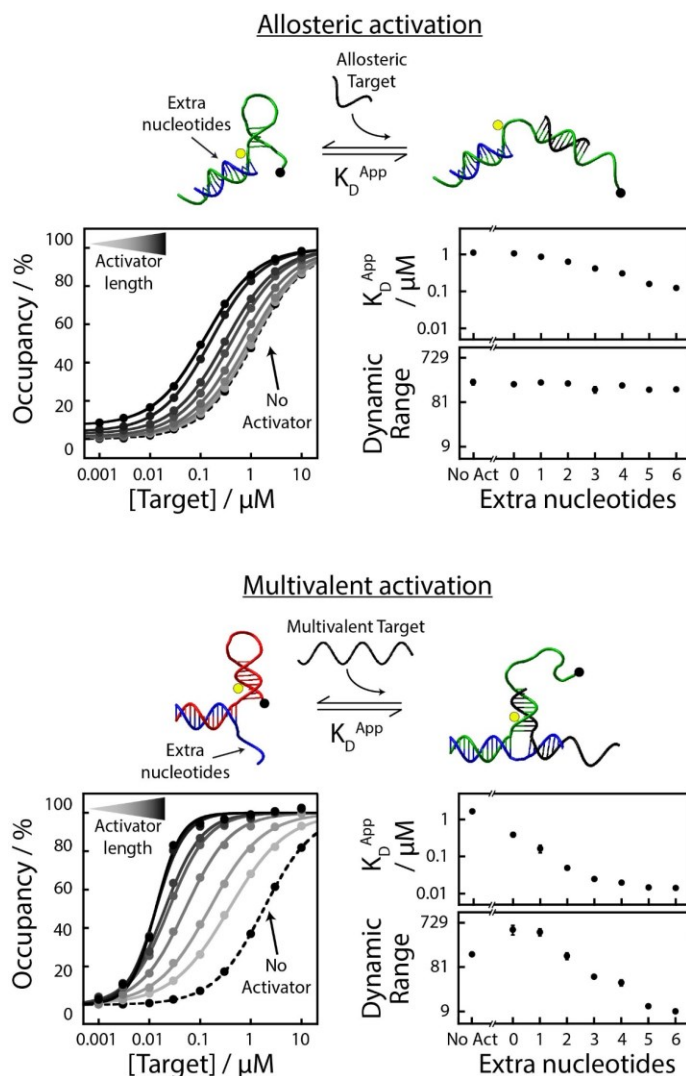
We first validated the impact of the different activators on the equilibrium of the switch (Fig. 4.2 and Fig. S4.3). To do so, we characterized the switching equilibrium of our DNA-based switch using urea denaturation procedures (Fig. S4.4)<sup>30</sup>. We measured a  $\Delta G$  of  $3.00 \pm 0.06$  kcal·mol<sup>-1</sup>, which led to a switching constant ( $K_S$ ) of  $0.0089 \pm 0.0008$ , meaning that 99% of the switch remained in the “OFF-state” in the absence of activator or target. As expected, upon binding to the switch, an allosteric activator alters the switching equilibrium of the switch from  $K_S = 0.0089 \pm 0.0008$  to  $0.17 \pm 0.01$  as confirmed by the increase in the initial fluorescence of the switch<sup>3, 25</sup>. Favoring the equilibrium toward the “ON-state” (only 85% of the switch remains in the “OFF-state”) enables to improve the apparent affinity of the target by 10-fold (from  $1.14 \pm 0.03$   $\mu$ M to  $68 \pm 4$  nM).

In contrast, we confirmed that the multivalent activator does not alter the switching equilibrium. Indeed, this latter remains unchanged ( $K_S = 0.0092 \pm 0.0009$ ) since no impact on the initial background signal of the switch was observed. We also found that the additional interacting

interface provided by the activator also improved the apparent affinity of the target by up to 100-fold (from  $1.7 \pm 0.1 \mu\text{M}$  to  $14.2 \pm 0.3 \text{ nM}$ ). We further looked at the impact of longer allosteric and multivalent activators and we confirmed that no matter the length of the additional interacting interface, multivalent activators never affected the switching equilibrium (Fig. S4.5). Interestingly, we also noted that multivalent activators drastically affected the cooperativity, or dynamic range, of the molecular switch. Indeed, while the dynamic range of an allosterically activated switch remains unchanged, a multivalently activated switch displays a narrower dynamic range (from  $\text{DR} = 153 \pm 10$  to  $\text{DR} = 10 \pm 1$ ). Finally, we also observed that in the absence of activator, the switch displays a similar affinity for both targets employed in the different strategies (red binding curves,  $K_{D,\text{Target}} = 1.14 \pm 0.03 \mu\text{M}$  and  $1.7 \pm 0.1 \mu\text{M}$ , respectively).



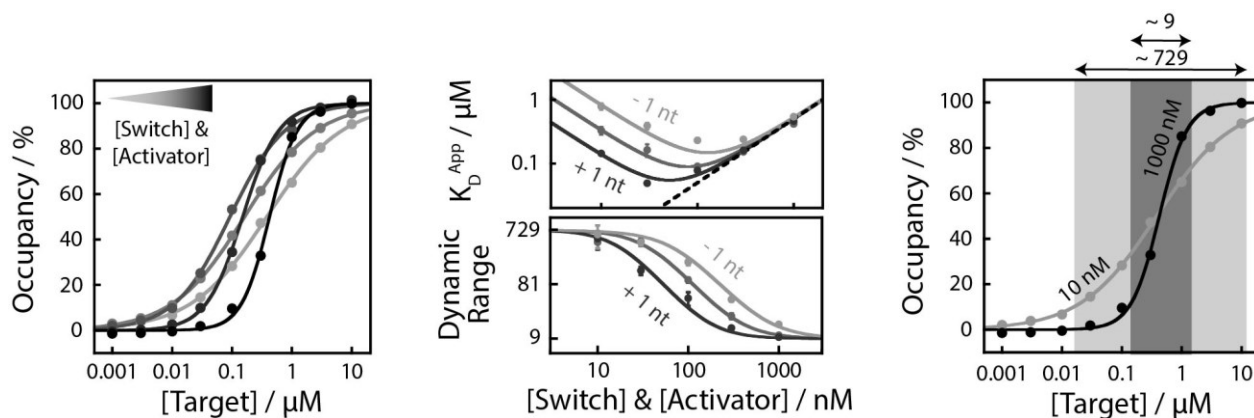
**Figure 4.2 – A DNA molecular switch modulated via an allosteric or a multivalent mechanism.** *Top.* A DNA- switch controlled via an allosteric activator (blue strand)<sup>3</sup>. In this model proposed by Plaxco et al., the DNA activator invades and destabilizes the stem of a fluorescent DNA switch thus favoring the equilibrium ( $K_S$ ) toward the “ON state” and improving target affinity (see arrow and Fig. S4.3). The increased “ON-state” population is confirmed by an increase in the fluorescence background of the switch (blue curve). *Bottom.* A DNA switch controlled by a multivalent activator (blue strand). Upon binding to the switch, this DNA activator introduces an additional hybridization interface that increase the affinity of the target (see arrow). In contrast to the allosteric activator, the multivalent activator does not alter the switching equilibrium (similar background fluorescence in the absence and presence of activator, see also Fig. S4.3). The yellow circle represents a FAM moiety chemically attached to a thymine base, while the black circle represents BHQ-1 attached at the 3’ extremity. Both were introduced as a labeling strategy to determine the conformational state of the switch. See Fig. S4.4 and Fig. S4.5 to see for the calculation of  $K_S$ .  $K_S$  were calculated from three independent experiments.



**Figure 4.3 – Multivalent activators enable programming both the affinity ( $K_D^{App}$ ) and the dynamic range of the molecular switch by introducing an additional interacting interface. *Top.* The allosteric activator enables programming the affinity of the switch for its target ( $K_D^{App}$ ) but has no impact on the dynamic range of the switch. *Bottom.* In addition to enabling a much stronger increase in the apparent affinity (up to one order of magnitude more), multivalent activators also enable tuning the dynamic range of the switch (from 9- to 729-fold of target concentration) by simply increasing the length of the new interacting interface. All binding curves are performed in triplicate.**

Having determined that the multivalent activator behaved as expected, we then explored its programming ability (Fig. 4.3). To do so, we designed activators of different lengths to increase the size of the new interacting interface (from 7 to 13 nucleotides), which in turn increases the affinity of the target toward the switch-activator complex. In the absence of an activator, the switch displayed a  $K_D$  of  $1.7 \pm 0.2 \mu\text{M}$  with a dynamic range of  $153 \pm 10$  ( $n_H = 0.87 \pm 0.01$ ). Adding

a multivalent activator that contains a minimal interacting interface of 7 nt produced a switch that opened at 4 times less target concentration ( $K_D^{App} = 0.39 \pm 0.05 \mu\text{M}$ ) with an extended dynamic range reaching up to 729-fold (DR =  $514 \pm 118$ ). When increasing the length of this activator by up to 6 extra nucleotides (from 7 nt to 13 nt), we observed a 100-fold decrease in  $K_D^{App}$  until reaching the ligand-depletion limit ( $K_D^{App} = 15 \text{ nM} = \frac{1}{2} [\text{switch}]$ )<sup>31, 32</sup>. Furthermore, we found that the dynamic range of the switch bound to the longest activator (13 nt) is now reduced to a “cooperative-like” 9-fold dynamic range, a behavior also expected from a molecular system under a ligand-depletion regime. In contrast, activating the same switch using the allosteric mechanism did not affect the dynamic range while only reducing the  $K_D^{App}$  of the switch by 10-fold. As demonstrated in previous studies, the extent to which an allosteric activator can improve the binding affinity of a molecular switch is limited by the affinity of the target for the “OFF-” and “ON-state” of the switch<sup>3</sup>. In contrast, the extent to which a multivalent activator can improve the binding affinity of a molecular switch is directly proportional to the additional energy provided by the new interacting interface which, in theory, is *ad infinitum*.



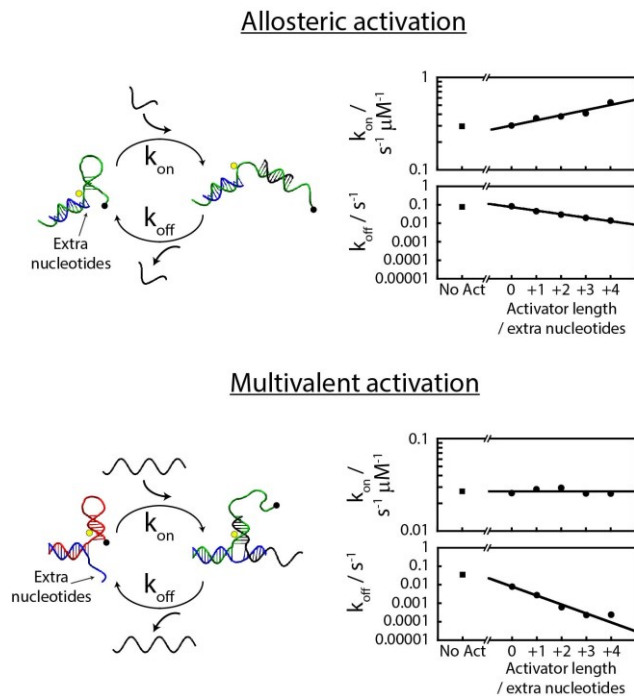
**Figure 4.4 – Fine tuning multivalent activation mechanism by varying the concentration of the switch-activator complex.** *Left.* Using the activator providing 8 additional nucleotides, we found that varying altogether the concentration of the switch and the activator provides a way to program the level of preorganization and therefore the  $K_D^{App}$  and dynamic range. *Middle.* By doing so, the  $K_D^{App}$  varies following a bell-shaped tendency, whereas the dynamic range varies following a sigmoidal tendency between 729 and 9. Increasing (+ 1 nt, Fig. S4.6, *left*) or decreasing (- 1 nt, Fig. S4.6, *right*) the number of nucleotides in the additional interacting interface also enables programming these two concentration-dependent profiles. *Right.* Multivalent activators also provide a means by which the dynamic range can be programmed independently from its  $K_D^{App}$ . All binding curves are performed in triplicate.

The activation of a molecular switch via multivalent activators can also be programmed by simply varying the concentration of the switch-activator complex. Here, we used an activator that provides a short 8-nucleotide interacting interface and demonstrated that an increase in the concentration of activated switches enables programming both its  $K_D^{App}$  and its dynamic range (Fig. 4.4, *left* and Fig. S4.6). At low concentrations of the switch-activator complex (*e.g.*, 10 nM), the switch is not bound to the activator, and trimeric assembly occurs only at higher target concentrations ( $K_D^{App} = 0.33 \pm 0.05 \mu\text{M}$ ) in an anti-cooperative manner (extended dynamic range,  $DR \sim 729$ ). As we increased the concentration of the switch-activator complex (*e.g.*, 100 nM), the binding of the target occurred at a lower concentration ( $K_D^{App} = 0.090 \pm 0.002 \mu\text{M}$ ) with a reduced dynamic range ( $DR = 96 \pm 9$ ). At higher concentrations of the switch-activator complex (*e.g.*, 1000 nM), the ligand depletion regime pushes back the  $K_D^{App}$  to higher values (*e.g.*,  $0.417 \pm 0.004 \mu\text{M}$ ), corresponding to nearly half the concentration of the switch being used (*i.e.*, 500 nM). As expected from the ligand-depletion regime, such conditions provide a narrower, cooperative-like, switch opening ( $DR \sim 9$ ).

Shorter and longer activators also enable programming the activation profile of the switch (Fig. 4.4, *middle*). For instance, a weaker multivalent activator (- 1 nt) displaces the concentration-dependent profile of  $K_D^{App}$  and the dynamic range toward higher concentrations while a stronger multivalent activator (+ 1 nt) displaces them toward lower concentrations. A multivalent activator also enables independent programming of the dynamic range from its  $K_D^{App}$  (Fig. 4.4, *right*). For example, a 100-fold increase in concentration (from 10 nM to 1000 nM) led to activation that occurred at a similar concentration ( $K_D^{App} = 0.33 \pm 0.05 \mu\text{M}$  vs.  $0.417 \pm 0.004 \mu\text{M}$ , respectively), but with 40-fold variation in the dynamic range ( $DR = 446 \pm 117$  vs.  $10.4 \pm 0.4$ , respectively). Furthermore, the activation of the switch can also be performed by only changing the concentration of activator while keeping the concentration of the switch constant (Fig. S4.7).

Multivalent activators also possess a remarkable ability to program the half-life of the “ON state” of a switch when compared to their allosteric counterpart (Fig. 4.5). For example, when invading the stem and favoring the high-affinity “ON-state”, the allosteric activator enables slight acceleration of the binding of the target ( $k_{on}$ ) from  $0.30 \pm 0.02$  to  $0.536 \pm 0.007 \text{ s}^{-1} \cdot \mu\text{M}^{-1}$  while also slowing down its dissociation rate ( $k_{off}$ ) from  $0.078 \pm 0.008$  to  $0.0140 \pm 0.0008 \text{ s}^{-1}$ , both by less

than one order of magnitude. In the case of the multivalent activator, however, the additional interface does not alter the conformational state of the switch, and thus, the binding of the target remains rate-limited by its invasion into the stem (all  $k_{on} = 0.027 \pm 0.002 \text{ s}^{-1} \cdot \mu\text{M}^{-1}$ ). The additional interface, however, considerably impacts the dissociation of the activated switch complex since more base-pairs need to be broken for dissociation to occur and  $k_{off}$  decreases by up to two orders of magnitude from  $0.035 \pm 0.002$  to  $(2.4 \pm 0.6) \times 10^{-4} \text{ s}^{-1}$ . Therefore, while an allosterically activated switch displays association and dissociation rates linked to its switching equilibrium ( $K_S$ ), a multivalently activated switch only sees its deactivation rate impacted, which provides a way to precisely program the half-life of the activated state of the switch without changing its rate of activation.



**Figure 4.5 – Programming the kinetics of activation and deactivation of molecular switches through allosteric and multivalent activation mechanisms.** *Top.* Increasing the activator length (from 14 nt to 18 nt), and thus the invasion of the stem, slightly accelerates the activation of the switch by increasing the fraction of the high affinity “ON-state”. Increasing the activator length also slightly decreases the deactivation rate of the switch likely by preventing spontaneous stem-loop formation that can trigger target dissociation. *Bottom.* Increasing the multivalent activator length (from 23 nt to 27 nt) provides a larger and more energetic binding interface for the target. Although this longer binding interface has no impact on the activation kinetics of the switch ( $k_{on}$ ), it allows tuning the half-life of the “ON state” by up to two orders of magnitude due to the additional base-pairs that need to be broken during the deactivation of the switch. All kinetics traces were performed in triplicate (see Fig. S4.8 and Fig. S4.9 for raw data).



## 4.4 Discussion et conclusions

Here, we have exploited multivalency to develop a new class of activators to program the affinity and dynamic range of molecular switches. Compared to the classic allosteric mechanism<sup>3, 5, 8, 10</sup>, multivalent activators create a novel interacting interface for the target that provides a means to design switches with a narrow dynamic range ( $DR < 81$ ) and switches displaying an extended dynamic range ( $DR > 81$ ), a feature that was previously only obtainable by combining multiple switches together<sup>33, 34</sup>. Multivalent activators also enable programming the dynamic range independently from its apparent affinity, leading to precise control over the activation threshold and the sensitivity of molecular switches. Furthermore, multivalent activators provide a way to rationally tune the half-life of the “ON state” of the switch ( $k_{off}$ ) while not affecting its activation rate ( $k_{on}$ ), which can still be optimized through an allosteric activation mechanism if needed. Ultimately, the ability to rationally control the properties of molecular switches through the formation of a multivalent assembly can find many applications in DNA-based nanotechnology to improve the binding efficiency of DNA switches, to modulate the drug-release properties of DNA cargo or to program DNA-based computation<sup>34</sup>.

Such a multivalent mechanism may have broad repercussions in the design of *de novo* protein-based assembly where such assemblies, where such assemblies are often designed through a modular approach of linking multivalent interacting domains together<sup>18</sup>. In addition, we also believe that a better understanding and comparison of the allostery and multivalency mechanisms provide new insights to better understand the evolution of biomolecular receptors and switches and why some have selected multivalency over allostery (e.g., Notch-1 transcriptional complex, the assembly of the 30S ribosome, and  $\beta$ -interferon enhanceosome)<sup>35-37</sup>.

## 4.5 Materials and Methods

### Chemicals.

All ancillary reagents used in our DNA synthesizer which includes anhydrous acetonitrile, oxidizer solution (0.1 M I<sub>2</sub> in H<sub>2</sub>O/Pyridine/THF), activation reagents (0.25 M ETT in ACN), the Cap A mix (Ac<sub>2</sub>O/Pyridine/THF), the Cap B mix (16% MeIm in THF) and DMT removal reagent (3% TCA in DCM) were purchased from ChemGenes Corporation, Wilmington. All unmodified DNA phosphoramidites (G, C, A and T) were purchased from ChemGenes Corporation, Wilmington in the form of dG-CE (n-ibu), dC-CE (n-bz), dA-CE (n-bz) and dT-CE. Synthesis columns were purchased from Biosearch Technologies, Navato in the form of pre-pack 1000Å 1 µmol Suc-CPG synthesis columns: dT, dG(dmf), dA(bz), dC(bz). The 3'-BHQ1 CGP was purchased from the same supplier in a 200 nmol scale format. Modified DNA phosphoramidites were purchased from Biosearch Technologies, Navato: Fluorescein-dT CE-Phosphoramidite (dT-FAM). MicroPure II solid-phase extraction columns were purchased from Biosearch Technologies, Novato. Amonium hydroxide solution (ACS, 28.0-30.0%) and acetic acid (ACS, ≥99.7%) were purchased from Sigma Aldrich. Triethylamine (reagent grade) was purchased from Fisher Scientific. All buffer solutions were prepared in the lab using Milli-Q water which was also prepared in the lab when needed. Na<sub>2</sub>HPO<sub>4</sub>, NaOH and NaCl were purchased from Bioshop Canada In., Burlington. HCl 25% was purchased from VWR Canada, Mississauga.

### Oligonucleotides Synthesis.

All oligonucleotides were synthesized in the lab using standard phosphoramidite chemistry. The syntheses were performed automatically on a DNA/RNA synthesizer H-6 (K&A Laborgeraete, Schaafheim). Once synthesized, the DNA sequences were cleaved from the solid support and deprotected in fresh 28.0-30.0% aqueous ammonia at 60°C overnight (16h). The purification of unlabelled oligonucleotides was performed by solid-phase extraction using the P-8 purifier (K&A Laborgeraete, Schaafheim) while labelled oligonucleotides containing a dT-FAM and a BHQ-1 were purified by reversed-phase chromatography using a gradient of anhydrous ACN and 0.1 M TEAA buffer (HPLC: 1260 infinity, Agilent Technologies) (column: XBridge®Oligonucleotide BEH C18 OBDTM Prep column, 130Å, 2.5 µm, 10 x 50 mm). All

oligonucleotides, after their purification, were dried with a Genevac™ miVac centrifugal concentrator and resuspended in Milli-Q water prior to their quantification by UV-Vis spectrophotometry (Cary 100 Bio, Varian). The concentrations of all oligonucleotides were calculated from their absorbances at 260 nm using the absorption coefficients ( $\epsilon$ ) estimated from their sequences using the IDT OligoAnalyzer tool.

## DNA sequences.

### Multivalent system

Switch: 5'- CCT TAA TCT CTT TGC TC(dT-FAM) TTC CCG ATC GGC GTT TTA GGG AAT GAG-BHQ1 -3'

Multivalent target: 5'- CGA TCG GGA AAG AGTICG TCG GTT CCT CTC -3'

Activator #1 (23nt): 5'- **ACC GAC GTT** CAA AGA GAT TAA GG -3'

Activator #2 (24nt): 5'- **AAC CGA CGT ICA** AAG AGA TTA AGG -3'

Activator #3 (25nt): 5'- **GAA CCG ACG TTC** AAA GAG ATT AAG G -3'

Activator #4 (26nt): 5'- **GGA ACC GAC GTT** CAA AGA GAT TAA GG -3'

Activator #5 (27nt): 5'- **AGG AAC CGA CGT ICA** AAG AGA TTA AGG -3'

Activator #6 (28nt): 5'- **GAG GAA CCG ACG TTC** AAA GAG ATT AAG G -3'

Activator #7 (29nt): 5'- **AGA GGA ACC GAC GTT** CAA AGA GAT TAA GG -3'

**Bold:** The interacting interface with the target and the activator

Underline: Linker between both binding domains

### Allosteric system

Switch: same as the multivalent system

Target (loop target): 5'- TAA AAC GCC GAT C -3'

Activator #1 (14nt): 5'- AGC AAA GAG ATT AA -3'

Activator #2 (15nt): 5'- GAG CAA AGA GAT TAA -3'

Activator #3 (16nt): 5'- AGA GCA AAG AGA TTA A -3'

Activator #4 (17nt): 5'- AAG AGC AAA GAG ATT AA-3'

Activator #5 (18nt): 5'- AAA GAG CAA AGA GAT TAA-3'

Activator #6 (12nt): 5'- AGA GCA AAG AGA -3'

Activator #7 (13nt): 5'- AGA GCA AAG AGA T -3'

Activator #8 (14nt): 5'- AGA GCA AAG AGA TT -3'

Activator #9 (15nt): 5'- AGA GCA AAG AGA TTA -3'

Activator #10 (16nt): same as activator #3

Activator #11 (17nt): 5'- AGA GCA AAG AGA TTA AG -3'

Activator #12 (18nt): 5'- AGA GCA AAG AGA TTA AGG -3'

Underline: Domain that invades the stem

Activators from #1 to #5 have longer domain to invade the stem while activators from #6 to #12 have longer domain to bind with the tail of the switch.

dT-FAM corresponds to a thymidine modified with a fluorescein moiety while *BHQ1* corresponds to a black hole quencher 1 moiety.

### **Fluorescence experiments.**

Binding Curves. In a quartz cuvette of 1.5 mL, appropriate dilutions of the DNA switch (with and without activator) were made to obtain a final volume of 900  $\mu$ L at the concentration specified in the figures. To this solution, a small volume of the target was added, and the solution was mixed and left to equilibrate for 20 minutes. When the signal was not changing over time anymore (*i.e.*, the equilibrium is reached), the fluorescence signal was recorded (ex./em. = 496/520 nm) with a fluorimeter (Cary Eclipse, Agilent Technologies). The total volume of targets added to the switch solution was kept below 5% (< 950 mL) to avoid a high dilution of the switch solution. Therefore, we assumed that the concentration of the switch stayed approximately

constant during the titration. The titration curves were fitted using the Hill equation (Eq. 4.1) where  $F$  is the fluorescence signal,  $Bsl$  is the baseline signal,  $Amp$  is the amplitude in the change of fluorescence when the target is fully bound by the switch,  $[T]$  is the concentration of the target,  $K_{D,obs}$  is the observed dissociation constant and  $n_H$  is the Hill factor which is related to the dynamic range by  $DR = 81^{1/n_H}$ .

$$F = Bsl + Amp \frac{[T]^{n_H}}{K_{D,obs}^{n_H} + [T]^{n_H}} \quad (\text{Eq. 3.1})$$

Chemical denaturation curves. The stability of the switch was measured by urea denaturation following a method developed by our lab<sup>30</sup>. Briefly, appropriate dilutions (900  $\mu\text{L}$ ) of the DNA switch was made in 10 M urea buffered solution. The solution is then diluted with a solution containing the same concentration of the DNA switch in pure buffer. After the dilution, the sample was left to equilibrate for two minutes prior to the recording of the fluorescence spectra (ex./em. = 496/520 nm) (Cary Eclipse, Agilent Technologies). The denaturation curve was fitted using Eq. 4.2 where  $[U]$  is the concentration of urea,  $F$  is the fluorescence signal,  $F^{\circ}_{Closed}$  and  $F^{\circ}_{Open}$  are the fluorescence signals of the closed and open state of the switch,  $\sigma_{Closed}$  and  $\sigma_{Open}$  correspond to the influence of urea on the fluorescence signal of these same states,  $\Delta G^{\circ}_{Sw}$  is the Gibbs free energy of the switch in absence of urea,  $R$  is the gas constant,  $T$  is the temperature and  $m$  is the  $m$ -value which corresponds to the influence of urea on the  $\Delta G^{\circ}_{Ass}$ .

$$F = \frac{(F^{\circ}_{Closed} + \sigma_{Closed}[U]) + (F^{\circ}_{Open} + \sigma_{Open}[U]) \cdot e^{-\frac{(\Delta G^{\circ}_{Sw} - m[U])}{RT}}}{\left(1 + e^{-\frac{\Delta G^{\circ}_{Sw} - m[U]}{RT}}\right)} \quad (\text{Eq. 4.2})$$

Kinetic traces. The activation rate of allosterically and multivalently activated switches were measured by rapidly mixing the activated switch solution (switch + activator) to the target solution by using a stopped-flow instrument coupled with a fluorimeter (SX20, Applied Photophysics). The final concentration of the switch and activator solution was 100 nM while the final concentration of the target solution was varied from 1  $\mu\text{M}$  to 6  $\mu\text{M}$ . The kinetic traces were fitted using an exponential equation (Eq. 4.3) to determine the observed rate constant which was then fitted according to Eq. 4.4 to obtain the activation rate constant ( $k_{on}$ ). Since  $k_{off}$  was too small

to be accurately extrapolated from that fit,  $k_{off}$  was determined using dilution experiments. In these experiments, dissociation is triggered by diluting a 10  $\mu\text{M}$  solution of the “ON-state” switch (switch + activator + target) directly in pure buffer in a 1:1000 ratio leading to a final concentration of 10 nM. The dissociation traces are recorded using a standard fluorimeter (ex./em. = 496/520 nm) (Cary Eclipse, Agilent Technologies). Oil was deposited on top of the cuvette to avoid evaporation of the solution throughout the course of this long dissociation. Dissociation traces were fitted using Eq. 3 where  $k_{off}$  replaces  $k_{obs}$ .

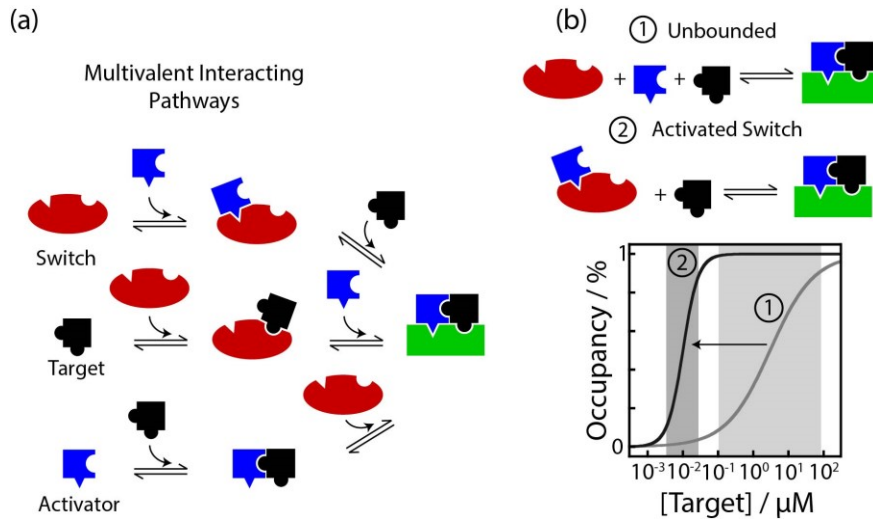
$$F = Bsl + Ape^{-k_{obs}t} \quad (\text{Eq. 4.3})$$

$$k_{obs} = k_{on}[Target] + k_{off} \quad (\text{Eq. 4.4})$$

## 4.6 Acknowledgements

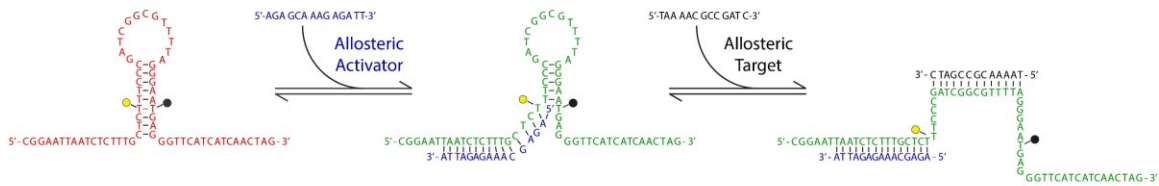
This research was conducted through the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants (RGPIN-2020-06975) (A.V.-B). A.V.-B. is Canada Research Chair in Bioengineering and Bionanotechnology, Tier II. D. L. acknowledges a 3rd cycle scholarship from the Fonds de recherche du Québec – Nature et technologies (FRQNT).

## 4.7 Supporting Figures

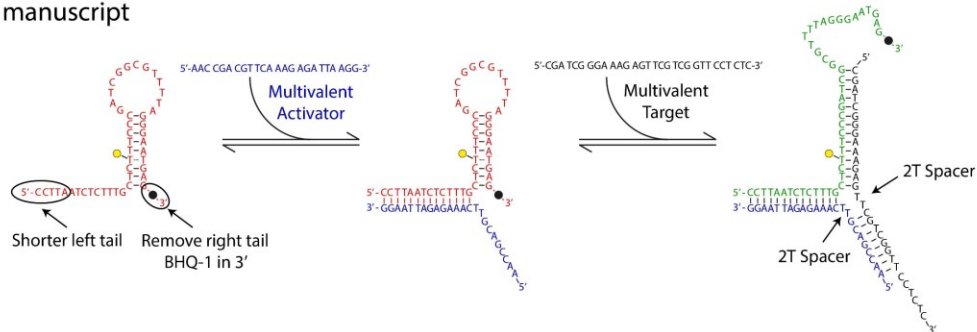


**Figure S4.1 – A multivalent activator provides multiple interacting pathways.** The activation of a switch (from red to green) by a multivalent activator (blue) can happen through three different paths involving three different dimeric intermediates (*i.e.*, the concentration of all components is below the  $K_D$  of all dimers), the switch is activated by directly forming a trimeric assembly which displays an extended dynamic range (DR = 729) happening at high concentration<sup>29</sup>. In contrast, when the activator is already bound to the switch (*i.e.*, the concentration of activator is above the  $K_D$  of the dimer), a larger binding interface is created which enable the activation of the switch to happen at lower concentration while displaying a narrower dynamic range (DR = 9). A narrower dynamic range is achieved when the concentration of switch is above the  $K_D$  of the target towards that new larger binding interface, thus leading to a ligand-depletion regime<sup>31, 32</sup>. Simulations are based on our previous reported trimeric binding model<sup>29</sup>.

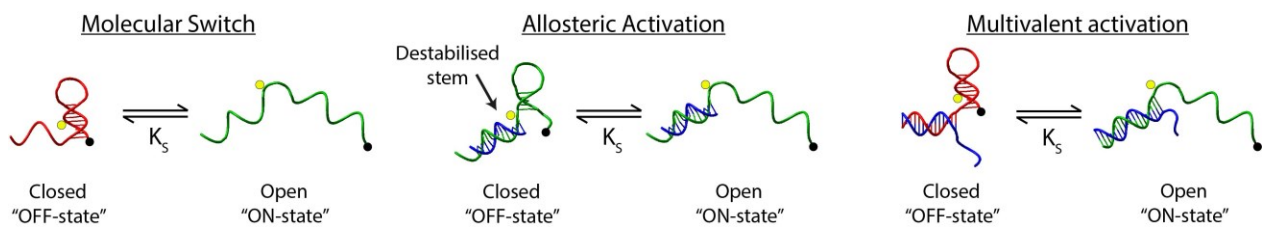
(a) K. W. Plaxco et al.



(b) This manuscript

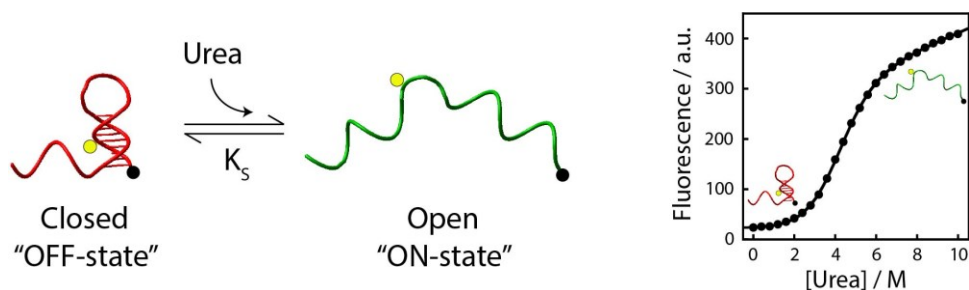


**Figure S4.2 – Comparison between the original switch from Plaxco et al.<sup>3</sup> and our multivalent switch.** (a) The switch used to validate the allosteric mechanism consists in a molecular beacon containing two tails, one of which can bind to an allosteric activator that can invade the stem of the molecular beacon (see ref. 3). This invasion disrupts some interaction rendering the binding with the target in the loop more efficient. (b) Our multivalent switch is based on the same design. However, both tails were modified. The left tail was shortened to avoid self-complementarity with the target while the right tail was removed. The multivalent activator was designed to fully hybridize with the left tail and to provide a new interacting interface for the target. By interacting with the activator, the target can open the stem and induce the formation of a three-way junction. To ensure the formation of a stable three-way junction, spacers of two thymidine were introduced on the activator and target strands.

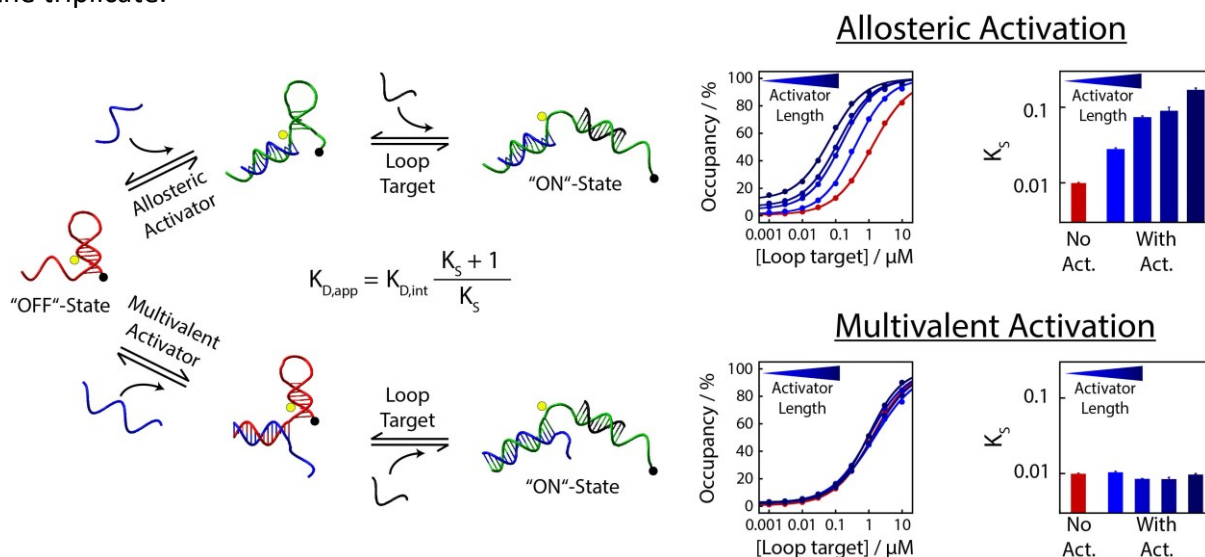


**Figure S4.3 – Switching equilibrium ( $K_s$ ) of the inactivated, allosterically activated and multivalently activated DNA-base switch.** *Left.* An inactivated switch is in equilibrium between its closed “OFF-state” conformation and its open “ON-state” conformation. This equilibrium is defined by the switching equilibrium  $K_s$  where  $K_s = [\text{Open}]/[\text{Closed}]$ . *Middle.* In an allosterically activated switch, the activator destabilises the stem, thus reducing the stability of the closed “OFF-state” conformation, as illustrated with the stem-loop being slightly open in its “OFF-state”. *Right.* In a multivalently activated switch, the activator does not affect the conformation of the switch and the switch should, in principle, display the same switching equilibrium ( $K_s$ ). See Fig. S4.4 for the  $K_s$  of the switch in absence of activator and Fig. S4.5 for the  $K_s$  in presence of activators.





**Figure S4.4 – Stability of the DNA-based switch measured by urea denaturation.** Using a method developed in the lab<sup>30</sup>, we measured the switching equilibrium ( $K_S$ ) between the closed and open state of our DNA-based molecular switch. We measured a  $\Delta G$  of  $3.00 \pm 0.06$  kcal·mol<sup>-1</sup> which corresponds to a switching constant of  $0.0089 \pm 0.0008$  ( $K_S = [\text{Open}]/[\text{Closed}]$ ). This implies that more than 99% of the switch is in the closed “OFF-state” conformation. The denaturation of the switch was performed in triplicate in PBS buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 150 mM NaCl, pH = 7.00) at 45°C. The curves were fitted using Eq. 4.2. The error corresponds to the standard deviation of the triplicate.

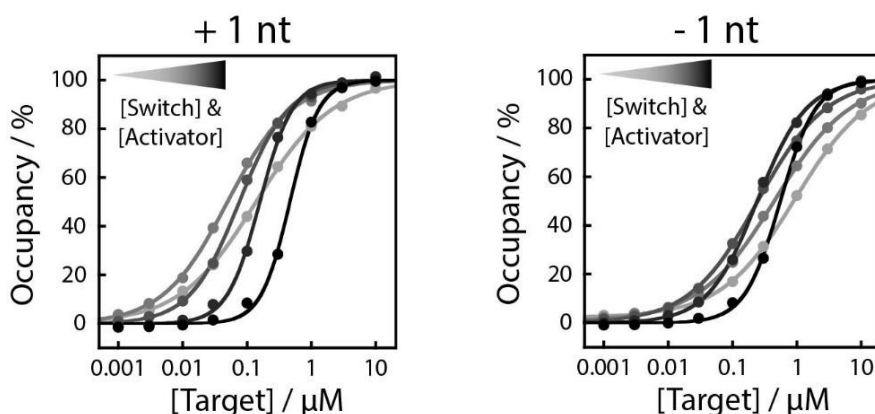


**Figure S4.5 – The impact of an allosteric activator and a multivalent activator on switching equilibrium of a DNA-based switch.** *Top.* The activator was designed by Plaxco *et al.* to invade the stem and therefore destabilise the switch (*i.e.*, increase  $K_S$ )<sup>3</sup>. Indeed, when adding a target that binds at the loop of the DNA switch, the apparent affinity of the target ( $K_{D,app}$ ) is shifted towards lower concentration, resulting in higher  $K_S$ . *Bottom.* Here, we have designed a multivalent activator that does not invade the stem but rather introduce a new interaction interface. To confirm that this new interacting interface does not impact the switching equilibrium and thus its stability (*i.e.*,  $K_S$ ), we performed the same experiments where we add a target that bind at the loop of the switch. We found that, the apparent affinity of the target remains unchanged, suggesting that all the multivalent activator does not affect the  $K_S$  of the DNA switch. All binding curves were performed in triplicate in PBS buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 150 mM NaCl, pH = 7.00) at 45°C using [switch] = 30 nM and [activator] = 30 nM. The  $K_S$  was measured in Fig. S4.4 ( $K_S = 0.0089 \pm 0.0008$ ). The intrinsic affinity of the target ( $K_{D,int}$ ) was measured in the absence of activators and was calculated following the equation derived by Plaxco *et al.* ( $K_{D,int} = 9.9 \pm 0.2$  nM)<sup>3</sup>. The  $K_S$  in presence of activator was calculated in the same way.

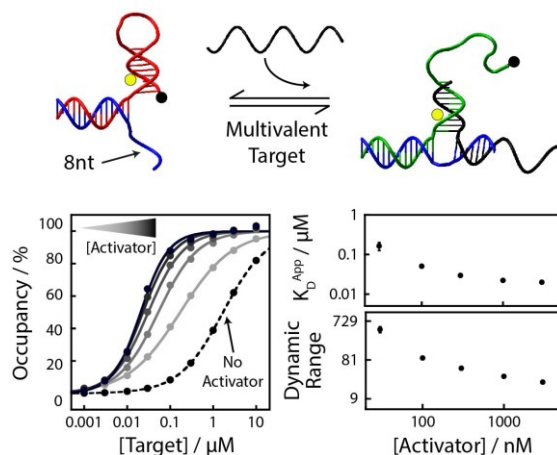
**Table S4.1 – The impact of an allosteric activator and a multivalent activator on the switching equilibrium of a DNA-based switch.**

Allosteric Activation				Multivalent Activation			
# Act.	$K_{D,obs} / \mu\text{M}$	$K_{D,int} / \text{nM}$	$K_S$	# Act.	$K_{D,obs} / \mu\text{M}$	$K_{D,int} / \text{nM}$	$K_S$
None	$1.14 \pm 0.03$	$9.9 \pm 0.2$	$0.0089 \pm 0.0008$	None	$1.14 \pm 0.03$	$9.9 \pm 0.2$	$0.0089 \pm 0.0008$
#2	$0.362 \pm 0.007$	$9.9 \pm 0.2$	$0.0282 \pm 0.0006$	#2	$0.97 \pm 0.03$	$9.9 \pm 0.2$	$0.0104 \pm 0.0003$
#3	$0.143 \pm 0.005$	$9.9 \pm 0.2$	$0.074 \pm 0.003$	#3	$1.18 \pm 0.02$	$9.9 \pm 0.2$	$0.0085 \pm 0.0001$
#4	$0.115 \pm 0.007$	$9.9 \pm 0.2$	$0.095 \pm 0.007$	#4	$1.20 \pm 0.07$	$9.9 \pm 0.2$	$0.0084 \pm 0.0005$
#5	$0.068 \pm 0.004$	$9.9 \pm 0.2$	$0.17 \pm 0.01$	#5	$1.04 \pm 0.03$	$9.9 \pm 0.2$	$0.0097 \pm 0.0003$

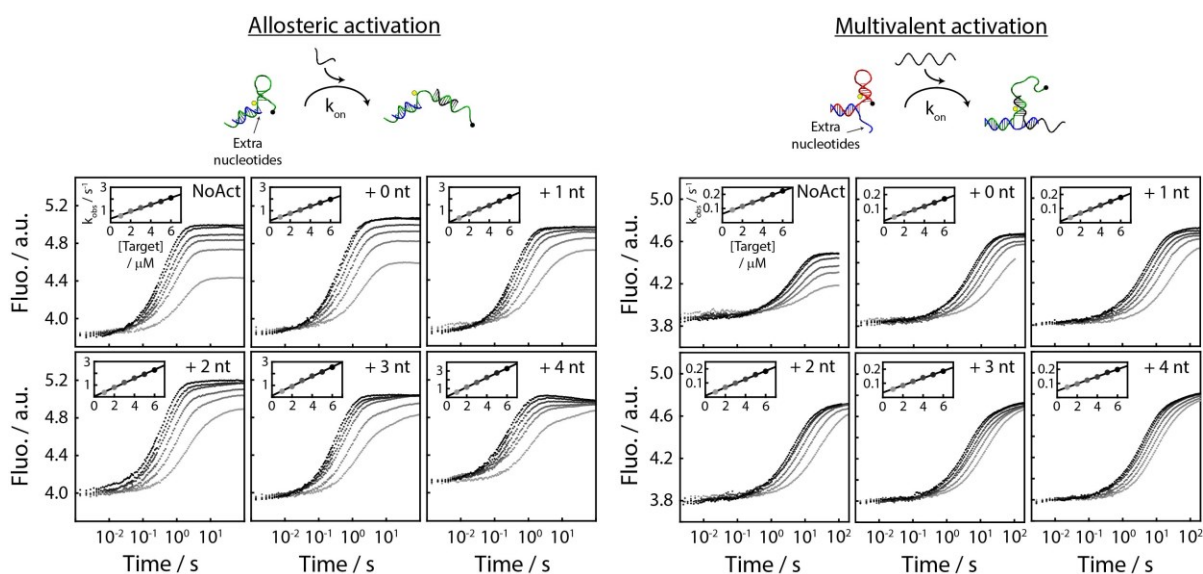
\* Binding constant values ( $K_{D,obs}$ ,  $K_{D,int}$ , and  $K_S$ ) and their standard deviations are calculated from three independent measurements.



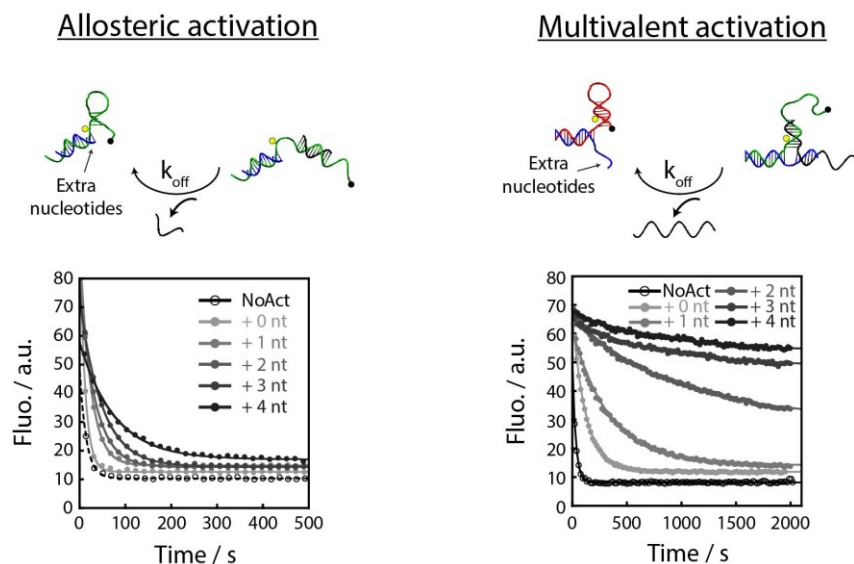
**Figure S4.6 – The extent to which multivalent activators can modulate the  $K_D^{APP}$  and the dynamic range is also concentration dependent.** *Left.* Binding curves of an activated switch using a multivalent activator providing an additional interacting interface of 9 nucleotides (9 nt, +1 nt). *Right* Binding curves of an activated switch using a multivalent activator providing an additional interacting interface of 7 nucleotides (7 nt, -1 nt). All binding curves were performed in triplicate in PBS buffer (50 mM  $\text{Na}_2\text{HPO}_4$ , 150 mM NaCl, pH = 7.00) at 45°C. The binding curves were fitted using the Hill equation (Eq. 4.1).



**Figure S4.7 – A molecular switch can also be programmed by simply changing the concentration of multivalent activators.** Using the multivalent activator providing an additional interacting interface of 8 nucleotides (8 nt), the  $K_D^{App}$  and the dynamic range can be decreased when increasing the concentration of activators. By doing so, for example, the  $K_D^{App}$  can be decreased from  $0.16 \pm 0.04 \mu\text{M}$  ([activator] = 30 nM) to  $0.020 \pm 0.001 \mu\text{M}$  ([activator] = 3000 nM) while the dynamic range can be shifted from  $457 \pm 76$  ([activator] = 30 nM) to  $23 \pm 2$  ([activator] = 3000 nM). All binding curves were performed in triplicate in PBS buffer (50 mM  $\text{Na}_2\text{HPO}_4$ , 150 mM NaCl, pH = 7.00) at 45°C. The binding curves were fitted using the Hill equation. The error corresponds to the standard deviation of the triplicate.



**Figure S4.8 – Determination of the activation kinetic ( $k_{on}$ ) of an allosterically and multivalently activated switches.** The  $k_{on}$  is measured by rapidly mixing the target (from 1  $\mu\text{M}$  to 6  $\mu\text{M}$ ) to a 100 nM switch solution containing, or not, the activator (also at 100 nM). The  $k_{on}$  is measured as the dependency of the observed rate constant ( $k_{obs}$ ) towards the concentration of targets:  $k_{obs} = k_{on}[\text{Target}] + k_{off}$ . Since the  $k_{off}$  were too small to be extrapolated accurately from these set of experiments, they were determined in a dilution experiment (see Fig. S4.9). All association experiments were performed in triplicate in PBS buffer (50 mM  $\text{Na}_2\text{HPO}_4$ , 150 mM NaCl, pH = 7.00) at 45°C. Values in Table S4.2.



**Figure S4.9 – Determination of the deactivation kinetic rate constant ( $k_{off}$ ) of an allosterically and multivalently activated switches.** The  $k_{off}$  is measured by diluting a 10  $\mu\text{M}$  solution of the activated switch (switch + activator + target) directly in buffer in a 1:1000 ratio (*i.e.*, 0.900 mL of DNA solution in 900 mL buffer). All dissociation experiments were performed in triplicate in PBS buffer (50 mM  $\text{Na}_2\text{HPO}_4$ , 150 mM NaCl, pH = 7.00) at 45°C. Values in Table S2.

**Table S4.2 – Kinetic parameters ( $k_{on}$  and  $k_{off}$ ) of an allosterically and multivalently activated switches.**

	Allosteric Activation		Multivalent Activation	
	$k_{on} / \text{s}^{-1} \cdot \mu\text{M}^{-1}$	$k_{off} / \text{s}^{-1}$	$k_{on} / \text{s}^{-1} \cdot \mu\text{M}^{-1}$	$k_{off} / \text{s}^{-1}$
No Act	$0.30 \pm 0.02$	$0.078 \pm 0.008$	$0.0269 \pm 0.0005$	$0.035 \pm 0.002$
+ 0 nt	$0.301 \pm 0.003$	$0.082 \pm 0.004$	$0.0257 \pm 0.0002$	$0.0078 \pm 0.0002$
+ 1 nt	$0.361 \pm 0.005$	$0.045 \pm 0.002$	$0.0283 \pm 0.0003$	$0.00274 \pm 0.00009$
+ 2 nt	$0.378 \pm 0.009$	$0.0294 \pm 0.0001$	$0.0293 \pm 0.0006$	$0.00061 \pm 0.00001$
+ 3 nt	$0.42 \pm 0.02$	$0.020 \pm 0.002$	$0.0255 \pm 0.0006$	$0.00023 \pm 0.00004$
+ 4 nt	$0.536 \pm 0.007$	$0.0140 \pm 0.0008$	$0.0254 \pm 0.0005$	$0.00024 \pm 0.00006$

\* Kinetic rate constant values ( $k_{on}$  and  $k_{off}$ ) and their standard deviations are calculated from three independent measurements.

## 4.8 References

1. Nel, A. E.; Mädler, L.; Velegol, D.; Xia, T.; Hoek, E. M. V.; Somasundaran, P.; Klaessig, F.; Castranova, V.; Thompson, M., Understanding biophysicochemical interactions at the nano–bio interface. *Nat. Mater.* **2009**, *8* (7), 543-557.
2. Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J., The ensemble nature of allostery. *Nature* **2014**, *508* (7496), 331-339.
3. Ricci, F.; Vallée-Bélisle, A.; Porchetta, A.; Plaxco, K. W., Rational Design of Allosteric Inhibitors and Activators Using the Population-Shift Model: In Vitro Validation and Application to an Artificial Biosensor. *J. Am. Chem. Soc.* **2012**, *134* (37), 15177-15180.
4. Simon, A. J.; Vallée-Bélisle, A.; Ricci, F.; Plaxco, K. W., Intrinsic disorder as a generalizable strategy for the rational design of highly responsive, allosterically cooperative receptors. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (42), 15048-15053.
5. Simon, A. J.; Vallée-Bélisle, A.; Ricci, F.; Watkins, H. M.; Plaxco, K. W., Using the Population-Shift Mechanism to Rationally Introduce “Hill-type” Cooperativity into a Normally Non-Cooperative Receptor. *Angew. Chem. Int. Ed.* **2014**, *53* (36), 9471-9475.
6. Bissonnette, S.; Del Grosso, E.; Simon, A. J.; Plaxco, K. W.; Ricci, F.; Vallée-Bélisle, A., Optimizing the Specificity Window of Biomolecular Receptors Using Structure-Switching and Allostery. *ACS Sensors* **2020**, *5* (7), 1937-1942.
7. Munzar, J. D.; Ng, A.; Juncker, D., Duplexed aptamers: history, design, theory, and application to biosensing. *Chem. Soc. Rev.* **2019**, *48* (5), 1390-1419.
8. Porchetta, A.; Vallée-Bélisle, A.; Plaxco, K. W.; Ricci, F., Allosterically Tunable, DNA-Based Switches Triggered by Heavy Metals. *J. Am. Chem. Soc.* **2013**, *135* (36), 13238-13241.
9. Rossetti, M.; Porchetta, A., Allosterically regulated DNA-based switches: From design to bioanalytical applications. *Anal. Chim. Acta* **2018**, *1012*, 30-41.
10. Porchetta, A.; Vallée-Bélisle, A.; Plaxco, K. W.; Ricci, F., Using Distal-Site Mutations and Allosteric Inhibition To Tune, Extend, and Narrow the Useful Dynamic Range of Aptamer-Based Sensors. *J. Am. Chem. Soc.* **2012**, *134* (51), 20601-20604.
11. Wodak, S. J.; Paci, E.; Dokholyan, N. V.; Berezovsky, I. N.; Horovitz, A.; Li, J.; Hilser, V. J.; Bahar, I.; Karanicolas, J.; Stock, G.; Hamm, P.; Stote, R. H.; Eberhardt, J.; Chebaro, Y.; Dejaegere, A.;

Cecchini, M.; Changeux, J.-P.; Bolhuis, P. G.; Vreede, J.; Faccioli, P.; Orioli, S.; Ravasio, R.; Yan, L.; Brito, C.; Wyart, M.; Gkeka, P.; Rivalta, I.; Palermo, G.; McCammon, J. A.; Panecka-Hofman, J.; Wade, R. C.; Di Pizio, A.; Niv, M. Y.; Nussinov, R.; Tsai, C.-J.; Jang, H.; Padhorny, D.; Kozakov, D.; McLeish, T., Allostery in Its Many Disguises: From Theory to Applications. *Structure* **2019**, *27* (4), 566-578.

12. Mariottini, D.; Idili, A.; Vallée-Bélisle, A.; Plaxco, K. W.; Ricci, F., A DNA Nanodevice That Loads and Releases a Cargo with Hemoglobin-Like Allosteric Control and Cooperativity. *Nano Lett.* **2017**, *17* (5), 3225-3230.

13. Rossetti, M.; Ranallo, S.; Idili, A.; Palleschi, G.; Porchetta, A.; Ricci, F., Allosteric DNA nanoswitches for controlled release of a molecular cargo triggered by biological inputs. *Chem. Sci.* **2017**, *8* (2), 914-920.

14. Del Grosso, E.; Ragazzon, G.; Prins, L. J.; Ricci, F., Fuel-Responsive Allosteric DNA-Based Aptamers for the Transient Release of ATP and Cocaine. *Angew. Chem. Int. Ed.* **2019**, *58* (17), 5582-5586.

15. Porchetta, A.; Idili, A.; Vallée-Bélisle, A.; Ricci, F., General Strategy to Introduce pH-Induced Allostery in DNA-Based Receptors to Achieve Controlled Release of Ligands. *Nano Lett.* **2015**, *15* (7), 4467-4471.

16. Whitty, A., Cooperativity and biological complexity. *Nat. Chem. Biol.* **2008**, *4* (8), 435-439.

17. Williamson, J. R., Cooperativity in macromolecular assembly. *Nat. Chem. Biol.* **2008**, *4* (8), 458-465.

18. Bai, Y.; Luo, Q.; Liu, J., Protein self-assembly via supramolecular strategies. *Chem. Soc. Rev.* **2016**, *45* (10), 2756-2767.

19. Hanebuth, M. A.; Kityk, R.; Fries, S. J.; Jain, A.; Kriel, A.; Albanese, V.; Frickey, T.; Peter, C.; Mayer, M. P.; Frydman, J.; Deuerling, E., Multivalent contacts of the Hsp70 Ssb contribute to its architecture on ribosomes and nascent chain interaction. *Nat. Commun.* **2016**, *7* (1), 13695.

20. Errington Wesley, J.; Bruncsics, B.; Sarkar Casim, A., Mechanisms of noncanonical binding dynamics in multivalent protein–protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (51), 25659-25667.

21. Cuesta, Á. M.; Sainz-Pastor, N.; Bonet, J.; Oliva, B.; Álvarez-Vallina, L., Multivalent antibodies: when design surpasses evolution. *Trends Biotechnol.* **2010**, *28* (7), 355-362.
22. Hughes, R. A.; Ellington, A. D., Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology. *Cold Spring Harbor Perspect. Biol.* **2017**, *9* (1), a023812.
23. Madsen, M.; Gothelf, K. V., Chemistries for DNA Nanotechnology. *Chem. Rev.* **2019**, *119* (10), 6384-6458.
24. Seeman, N. C.; Sleiman, H. F., DNA nanotechnology. *Nat. Rev. Mater.* **2017**, *3* (1), 17068.
25. Vallée-Bélisle, A.; Ricci, F.; Plaxco, K. W., Thermodynamic basis for the optimization of binding-induced biomolecular switches and structure-switching biosensors. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (33), 13802.
26. Ricci, F.; Vallée-Bélisle, A.; Plaxco, K. W., High-Precision, In Vitro Validation of the Sequestration Mechanism for Generating Ultrasensitive Dose-Response Curves in Regulatory Networks. *PLoS Comput. Biol.* **2011**, *7* (10), e1002171.
27. Ha, S. H.; Ferrell, J. E., Thresholds and ultrasensitivity from negative cooperativity. *Science* **2016**, *352* (6288), 990-993.
28. Tyagi, S.; Kramer, F. R., Molecular Beacons: Probes that Fluoresce upon Hybridization. *Nat. Biotechnol.* **1996**, *14* (3), 303-308.
29. Lauzon, D.; Vallée-Bélisle, A., Functional advantages of building nanosystems using multiple molecular components. *Nat. Chem.* **2023**.
30. Idili, A.; Ricci, F.; Vallée-Bélisle, A., Determining the folding and binding free energy of DNA-based nanodevices and nanoswitches using urea titration curves. *Nucleic Acids Res.* **2017**, *45* (13), 7571-7580.
31. Hulme, E. C.; Trevethick, M. A., Ligand binding assays at equilibrium: validation and interpretation. *Br. J. Pharmacol.* **2010**, *161* (6), 1219-1237.
32. Esteban Fernández de Ávila, B.; Watkins, H. M.; Pingarrón, J. M.; Plaxco, K. W.; Palleschi, G.; Ricci, F., Determinants of the Detection Limit and Specificity of Surface-Based Biosensors. *Anal. Chem.* **2013**, *85* (14), 6593-6597.

33. Ricci, F.; Vallée-Bélisle, A.; Simon, A. J.; Porchetta, A.; Plaxco, K. W., Using Nature's "Tricks" To Rationally Tune the Binding Properties of Biomolecular Receptors. *Acc. Chem. Res.* **2016**, *49* (9), 1884-1892.
34. Harroun, S. G.; Prévost-Tremblay, C.; Lauzon, D.; Desrosiers, A.; Wang, X.; Pedro, L.; Vallée-Bélisle, A., Programmable DNA switches and their applications. *Nanoscale* **2018**, *10* (10), 4607-4641.
35. Del Bianco, C.; Aster, J. C.; Blacklow, S. C., Mutational and Energetic Studies of Notch1 Transcription Complexes. *J. Mol. Biol.* **2008**, *376* (1), 131-140.
36. Recht, M. I.; Williamson, J. R., RNA Tertiary Structure and Cooperative Assembly of a Large Ribonucleoprotein Complex. *J. Mol. Biol.* **2004**, *344* (2), 395-407.
37. Courey, A. J., Cooperativity in transcriptional control. *Curr. Biol.* **2001**, *11* (7), R250-R252.



# Chapitre 5 – Conclusion

## 5.1 Conclusions générales

Le but principal de cette thèse consistait à développer un nouveau mécanisme moléculaire pour programmer la plage dynamique d'interrupteurs moléculaires permettant d'optimiser leurs propriétés dans diverses applications. Pour y arriver, ce projet de recherche a été divisé en trois objectifs distincts qui ont chacun été abordés à différents niveaux dans les trois chapitres précédents. Le premier objectif consistait à s'inspirer de la nature pour mettre au point un modèle multimérique simple à base d'ADN pour mieux étudier l'importance des complexes biologiques. Deux stratégies furent employées, soit la fragmentation et la fusion. Ces modèles ont servi, en deuxième lieu, à tirer des conclusions sur les avantages thermodynamiques, cinétiques et fonctionnels des assemblages moléculaires permettant ainsi de mieux comprendre le rôle et la fonction des systèmes protéiques multimériques, et potentiellement tirer des informations pertinentes quant à leur évolution. Finalement, ces connaissances furent utilisées afin de mettre au point une nouvelle génération d'interrupteurs moléculaires basés sur des assemblages multimériques. Ce présent chapitre vise donc à faire un retour sur les conclusions générales des trois précédents chapitres tout en permettant de mettre en lumière comment ils répondent aux objectifs de cette thèse. Finalement, différentes avenues ou applications de recherches découlant de ces travaux seront suggérées et approfondies.

### 5.1.1 Création d'assemblages moléculaires par la fragmentation et la fusion

Dans le chapitre 2, il fut présenté qu'un assemblage moléculaire peut facilement être obtenu en coupant un système unimoléculaire en deux ou trois fragments qui peuvent interagir pour reformer la structure d'origine. À l'opposé, dans le chapitre 3, il fut démontré comment un assemblage trimérique peut être construit en fusionnant des partenaires d'interactions par le biais d'un espaceur qui dicte les propriétés finales de l'assemblage. Ces deux méthodes, directement inspirées de méthodes évolutives<sup>1-3</sup>, permettent l'obtention du même assemblage

trimérique, soit une jonction d'ADN à trois branches. Ici, l'ADN a permis la création d'une structure moléculaire trimérique bien définie, spécifique, et très soluble<sup>4</sup>. Contrairement à l'ADN, il aurait été difficile d'entrevoir un système protéique ayant un niveau de programmabilité similaire ou supérieur à l'ADN tout en permettant un contrôle rationnel sur la relation structure-activité des différents assemblages (*p. ex.* un monomère versus un dimère versus un trimère)<sup>5</sup>. De plus, évolutivement parlant, les protéines naturelles ont évolué à la suite de millions d'années, rendant difficile leur comparaison sachant qu'ils ont certainement expérimenté des pressions évolutives différentes afin d'accomplir des fonctions différentes<sup>6-8</sup>. Ainsi, par le biais d'un modèle simplifié tel que l'ADN, il fut possible de mieux comprendre les règles de conceptualisation menant à la création d'assemblages moléculaires multimériques, *c.-à-d.* un trimère. De ce système, il en ressort que la création d'assemblages moléculaires fonctionnels est fondamentalement très simple (*c.-à-d.* couper une structure existante ou combiner des domaines de liaison), mais regorge néanmoins d'un certain niveau de subtilité dans leur mise en pratique.

La fragmentation d'une nanostructure, plus précisément, peut avoir un impact considérable sur les propriétés intrinsèques du système non fractionné d'origine. En effet, dans le chapitre 2, il fut démontré qu'une simple fragmentation d'un ADNzyme en deux fragments a eu pour conséquence de réduire sa vitesse maximale d'environ un tiers. Même si le système fractionné peut se réassembler et fonctionner convenablement, deux mutations furent nécessaires afin de retrouver une activité similaire au système non fractionné. Cette observation n'est pas atypique, car plusieurs chercheurs ont démontré que certains systèmes moléculaires perdent de leur activité lorsque fractionnés et qu'une optimisation post-fragmentation est souvent nécessaire pour retrouver leur activité<sup>9, 10</sup>. Pour cette raison, plusieurs efforts ont été réalisés très récemment afin de faciliter la recherche de sites de coupes potentielles qui minimisent la perte d'activité dans des systèmes fragmentés à base d'ADN<sup>11</sup> ou de protéines<sup>12</sup>. Cependant, ce qui contraste avec ces recherches est que le nombre de mutations nécessaire pour retrouver un comportement natif est très faible, soit simplement deux mutations et ce pour un système fragmenté en trois composantes. Par exemple, pour maximiser la reconstitution d'une peroxydase fragmentée en deux, des chercheurs ont réalisé quatre rondes d'évolution dirigée menant ainsi à une peroxydase ayant neuf mutations par rapport au système natif<sup>13</sup>, une valeur

typique en évolution dirigée<sup>14</sup>. Il est envisageable que le nombre de mutations nécessaire pour retrouver une reconstitution maximale dans un système fragmenté en trois composantes soit encore plus grand. Des avancements récents en chimie computationnel permettent d'envisager une identification rationnel de sites de mutations permettant d'augmenter les performances d'un système protéique fragmenté<sup>10</sup>. Nos résultats suggèrent, néanmoins, qu'à un certain niveau les systèmes unimoléculaires peuvent facilement résister à une fragmentation de leur structure et permettre la création d'assemblages multimériques comprenant plusieurs composantes. À cet égard, les systèmes unimoléculaires à base d'acides nucléiques<sup>9</sup> ou de protéines<sup>15,16</sup> sont souvent fragmentés en deux composantes et utilisés dans des essais moléculaires de détection d'interactions ou de biocapteurs. Ici, il fut démontré comment une structure unimoléculaire peut, sans problème majeur, être divisée en trois fragments permettant d'éliminer la présence de structures préorganisées et ainsi maximiser le nombre de nouvelles interactions lors de l'assemblage. Le tout a de nombreuses conséquences importantes sur la régulation et la fonctionnalité de ces systèmes fragmentés (voir section 5.1.2). Outre la fragmentation, la création d'un assemblage trimérique fut aussi étudiée dans le chapitre 2 par l'ajout d'une troisième composante dans un système dimérique. Un tel système a permis d'obtenir un contrôle sur la régulation et la fonctionnalité qui est similaire à un système fragmenté.

La fusion de domaines d'interaction, en contraste, nécessite une optimisation précise des propriétés de l'espaceur afin de bien contrôler l'activité de l'assemblage ainsi construit. Dans le chapitre 3, il fut démontré comment différents domaines d'interaction peuvent être combinés par le biais d'un espaceur pour créer des composantes moléculaires pouvant s'autoassembler en un trimère. Dans ce cas de figure, une modification de la composition de l'espaceur entre les domaines d'interaction permet de facilement programmer la stabilité de l'assemblage modifiant du fait même son niveau de régulation et sa fonctionnalité (voir section 5.1.2). Grâce à l'étude thermodynamique exhaustive de différentes jonctions à trois branches, il fut démontré que la stabilité et la coopérativité de ces systèmes trimériques sont directement reliées à leurs capacités à enfouir une importante portion de leur surface au centre de l'interface créée par les domaines d'interaction. De plus, la formation de cette jonction au cœur de l'assemblage est favorisée enthalpiquement par des interactions espaceur-espaceur, limitant l'accessibilité au solvant. En

effet, la contribution enthalpique de l'assemblage fut augmentée significativement en ajoutant plus de cations de sodium dans la solution, ce qui a pour effet de minimiser la répulsion électrostatique entre les groupements phosphates des espaceurs d'acides nucléiques. Toutefois, la contribution entropique est demeurée la même suggérant que le niveau de compaction de l'assemblage demeure inchangé. De plus, les mêmes tendances furent obtenues en changeant les espaceurs de polythymidine par des espaceurs de polyéthylène glycol (PEG), suggérant que ce phénomène de maximisation de la compaction n'est pas unique à un espaceur d'ADN. Malgré tous, les espaceurs PEG n'ont pas fourni des assemblages aussi stables que les systèmes d'ADN ce qui accentue l'importance des interactions espaceur-espaceur au cœur de la jonction (*p. ex.* pont-H, empilement  $\pi$ - $\pi$ , interaction électrostatique, etc.). À ce titre, des espaceurs introduisant de fortes interactions espaceur-espaceur au cœur de la jonction comme des espaceurs hydrophobes<sup>17</sup> ou des espaceurs chélatants de métaux<sup>18</sup> permettent une forte stabilisation des jonctions à trois branches d'ADN. Par exemple, l'utilisation d'espaceurs 1,2-dodécanediol a permis d'augmenter de 25°C la  $T_m$  d'une jonction à trois branches tandis que des espaceurs chélatants bipyridines ont permis d'augmenter la  $T_m$  de 27°C en présence de  $Ni^{2+}$ , les deux augmentations définies par rapport à une jonction similaire sans espaceur. Cela contraste avec l'espaceur optimal 2T qui a simplement permis d'augmenter la  $T_m$  d'environ 10°C par rapport au système 0T, supportant ainsi l'importance des interactions fortes espaceur-espaceur au cœur d'une jonction. À la lumière de nos résultats, il est plausible d'envisager que ces systèmes d'espaceur (*c.-à-d.* 1,2-dodécanediol et bipyridine) sont stabilisés par une augmentation du niveau de compaction de la jonction menant à des interactions espaceur-espaceur accentuées.

### 5.1.2 Avantages et inconvénients des assemblages moléculaires biologiques

L'utilisation de l'ADN afin d'élucider les avantages et les inconvénients d'un assemblage moléculaire fut aussi motivée par le haut niveau de programmabilité des paramètres thermodynamiques et cinétiques dictant ces interactions<sup>19-22</sup>. Par exemple, une simple substitution d'une paire de bases (AT  $\rightarrow$  GC) permet une stabilisation d'un duplex d'ADN de 0,6 kcal·mol<sup>-1</sup>,<sup>21</sup> tandis que la vitesse d'assemblage ou de désassemblage d'un duplex peut facilement être programmée par le nombre de paires de bases ou la présence de mésappariement<sup>22</sup>. L'ADN

a donc offert un moyen par laquelle il fut possible d'analyser précisément l'impact de certains paramètres thermodynamiques et cinétiques, mais aussi l'effet du nombre de composantes sur la fonction d'un assemblage. Dans le chapitre 2, il fut démontré que la fragmentation d'une biomolécule en trois composantes permet un contrôle plus important sur son activité par rapport à un système similaire fragmenté en deux composantes ou un système non fragmenté. Du même coup, dans le chapitre 3, il fut démontré comment la formation d'un assemblage trimérique offre un énorme contrôle quant à la programmation de la plage dynamique d'un assemblage moléculaire.

L'un des nombreux avantages d'un système multimérique par rapport à un système unimoléculaire ou dimérique structurellement identique provient de sa capacité à maximiser le nombre de nouvelles interactions lors de son assemblage. Par exemple, il fut démontré dans le chapitre 2 qu'un système trimérique, en ayant plus de possibilités d'interaction, peut facilement être régulé par un inhibiteur tandis qu'un système dimérique et monomérique sont moins affectés par ce même inhibiteur, car leurs interactions sont indisponibles dû à la présence de structures préorganisées. De plus, en limitant la présence de structures préorganisées, un système trimérique augmente, d'une certaine manière, l'exposition au solvant de ces composantes, ce qui ouvre la porte à la création de systèmes moléculaires plus solubles. Par exemple, un rare exemple de système fragmenté trimérique illustre comment la fragmentation d'une protéine fluorescente verte (GFP) en trois composantes a permis de résoudre un problème de solubilités lorsque ces composantes sont fusionnées à d'autres protéines lors d'essai PCA<sup>23</sup>. Toutefois, cette affirmation n'est pas totalement généralisable aux protéines qui dépendent de la création d'un cœur hydrophobe pour se replier<sup>24</sup>. Ainsi, une mauvaise fragmentation d'une protéine pourrait exposer une plus grande surface hydrophobe au solvant favorisant ainsi une perte de solubilité en milieux aqueux et potentiellement l'agrégation de ces fragments<sup>25</sup>.

Un système trimérique permet un niveau de programmation plus diversifié de la plage dynamique d'un assemblage moléculaire, allant de petites à grandes plages dynamiques vers des seuils d'activation survenant à de faibles ou à de grandes concentrations. Tandis qu'un système dimérique peut voir sa plage dynamique variée entre 9 et 81, il fut démontré qu'un système trimérique permet l'obtention d'un assemblage plus étendu avec des plages dynamiques

pouvant varier entre 9 et 729. Fait intéressant, les simulations numériques ont démontré qu'un système tétramérique pourrait théoriquement élargir la plage jusqu'à 6561, soit  $9^4$ . Ainsi, il semblerait que la limite maximale de l'étendue de la plage dynamique soit limitée par le nombre de composantes de l'assemblage, variant entre  $9^1$  et  $9^n$  où  $n$  est le nombre de composantes dans l'assemblage. De plus, le seuil d'activation, *c.-à-d.* la concentration à laquelle 50% du système moléculaire est assemblé, peut aussi être programmé sur différentes plages de concentration. Plus précisément, il est possible de programmer indépendamment le seuil d'activation et la plage dynamique, ce qui permet de créer des systèmes moléculaires s'assemblant autour de la même concentration, mais avec des niveaux de coopérativité différents. Il est aussi possible, comme démontré dans les chapitres 2 et 3, de précisément programmer la plage dynamique et le seuil d'activation d'un assemblage trimérique via des mutations dans l'espaceur ou dans les domaines de liaison ce qui démontre la nature quantitative de cette stratégie.

Les systèmes trimériques ouvrent aussi la voie à de nouveaux mécanismes de régulation. Par exemple, dans les chapitres 2 et 3, il fut démontré qu'un assemblage moléculaire trimérique ayant un espaceur trop petit s'assemble suivant un profil d'auto-inhibition où à forte concentration de l'une de ces composantes, la formation d'intermédiaires dimériques est favorisée au détriment de l'assemblage trimérique. Ce phénomène est principalement attribuable à l'effet de l'espaceur sur la stabilité de l'assemblage où un petit espaceur réduit considérablement la stabilité du trimère sans affecter la stabilité des intermédiaires dimériques. Cela est une conséquence directe de l'effet chélate qui dicte qu'à un certain seuil de concentration (*c.-à-d.*  $EM/[Ligand]$ ) le système favorisera la formation d'un système ouvert (deux dimères) au lieu d'un système fermé (trimère)<sup>26</sup>. En effet, il fut mesuré qu'un espaceur de 0T réduit considérablement la concentration effective molaire par rapport au système optimal comprenant des espaceurs 2T, *c.-à-d.* 100 fois plus petit. À titre comparatif, les mêmes variations de l'espaceur dans un système dimérique n'ont pas permis d'obtenir le même niveau de régulation, ce qui illustre comment un système à plusieurs composantes offre un avantage régulateur par rapport à un système structurellement identique construit avec moins de composantes. De plus, ce phénomène d'auto-inhibition n'est pas inconnu des nanomachines naturelles puisque plusieurs enzymes sont inhibées par leur propre substrat, ce qui permet de

maintenir l'homéostasie<sup>27-29</sup>. Nous verrons plus tard, comment ce phénomène d'auto-inhibition pourrait être utilisé dans la conceptualisation de systèmes supramoléculaires fonctionnant loin de l'équilibre (voir section 5.2.3)<sup>30</sup>.

Les assemblages moléculaires à plusieurs composantes ouvrent aussi la porte à la création de systèmes sous contrôle cinétique où l'interaction entre les diverses composantes de l'assemblage peut être réalisée à différent temps en exploitant les pièges cinétiques. Simplement, un système unimoléculaire est typiquement régulé dans le temps par sa vitesse de repliement intramoléculaire, ce qui implique qu'il est très difficile d'optimiser la vitesse du repliement sans modifier le système ou utiliser des molécules externes<sup>31,32</sup>. Par contre, l'assemblage d'un système comprenant plusieurs composantes est régulé dans le temps par la vitesse intermoléculaire qui dicte l'interaction entre chacune des composantes<sup>30</sup>. Cela implique donc que l'assemblage est dépendant de la concentration de chacune des composantes, ce qui permet de contrôler son assemblage dans le temps, comme une horloge moléculaire. Dans le chapitre 2, il fut démontré qu'il est possible de programmer précisément la vitesse de formation des intermédiaires dimériques ainsi que la formation d'un assemblage trimérique simplement en variant la concentration des composantes. En comparaison, un système à deux composantes permet simplement de programmer la vitesse de formation de l'assemblage final tandis que des simulations numériques ont révélé qu'un système à quatre composantes permet un contrôle temporel encore plus important sur les différentes voies d'assemblage. Malgré tout, une augmentation du nombre de composantes peut aussi augmenter la possibilité de former des assemblages intermédiaires indésirables et nuisibles au bon fonctionnement du système. Ce contrôle temporel peut aussi être exploité pour contrôler l'activité d'un système enzymatique comme démontré avec une ADNzyme où l'activité d'une ADNzyme à trois composantes, à l'instar d'une ADNzyme à deux composantes, peut être programmé pour s'arrêter à un temps spécifique ou être retardé pendant un temps précis avant d'être activé.

Les systèmes trimériques sont, malgré tous leurs avantages, très vulnérables aux mutations. En effet, dans les chapitres 2 et 3, plusieurs mutations ponctuelles, de mutations d'insertions ou de mutations de délétions furent réalisés sur des assemblages trimériques et dimériques. Puisque qu'il fut démontré qu'un système à plusieurs composantes maximise le

nombre de nouvelles interactions par rapport à un système comprenant moins de composantes, il est normal qu'une mutation quelque-conque proche des interactions importantes à l'assemblage ait un effet plus important sur un trimère qu'un dimère. En effet, toutes mutations ayant eu un impact important sur un assemblage trimérique n'eurent aucun effet sur un équivalent dimérique, autant avec une jonction à trois branches modèle qu'avec une ADNzyme. Cela illustre d'une certaine manière que les systèmes à moindre nombre de composantes permettent un maintien de l'intégrité structurelle et de la fonction du système moléculaire. D'une autre côté, ces nombreux résultats démontrent que les systèmes moléculaires à plusieurs composantes peuvent être perçus comme étant plus facilement « évolutables », *c.-à-d.* qu'ils peuvent acquérir plus facilement de nouvelles fonctions à la suite d'évènements qui affectent leurs structures et leurs composantes (*p. ex.* l'auto-inhibition et les horloges moléculaires discutées précédemment). Toutefois, rien ne garantit qu'une telle mutation mène à un gain de fonction puisqu'il est possible que cette mutation soit néfaste pour le système et mène vers un cul-de-sac évolutif, voir tue simplement l'activité du système moléculaire.

### 5.1.3 Les interrupteurs moléculaires multivalents

Les jonctions à trois branches utilisées dans les chapitres 2 et 3 ont servi de modèles simplifiés pour comprendre et étudier les avantages coopératifs d'un assemblage moléculaire comprenant plus que deux composantes. Dans le chapitre 4, ces nouvelles connaissances furent employées afin de mettre au point une nouvelle catégorie d'effecteurs moléculaires permettant de rationnellement programmer la plage dynamique d'un interrupteur moléculaire. Typiquement, les interrupteurs moléculaires fonctionnent en suivant un mécanisme de déplacement de population (section 1.2.3), où l'interrupteur moléculaire est préférablement plus stable dans son état inactif que dans son état de liaison ( $K_s$  typiquement  $< 0.1$ )<sup>20</sup>. C'est la présence du ligand qui favorisera donc le déplacement vers l'état actif. Suivant ce modèle, l'ajout d'un effecteur moléculaire qui stabilise l'état de liaison ou qui stabilise l'état inactif aura respectivement pour effet d'augmenter (activation) ou de réduire (inhibition) l'affinité apparente du ligand pour l'interrupteur<sup>33</sup>. Ce type d'effecteur fonctionne donc en changeant l'équilibre d'interconversion de l'interrupteur entre son état inactif et son état de liaison et pour cette raison on les considère comme des effecteurs allostériques. Cependant, un tel effecteur ne permet pas

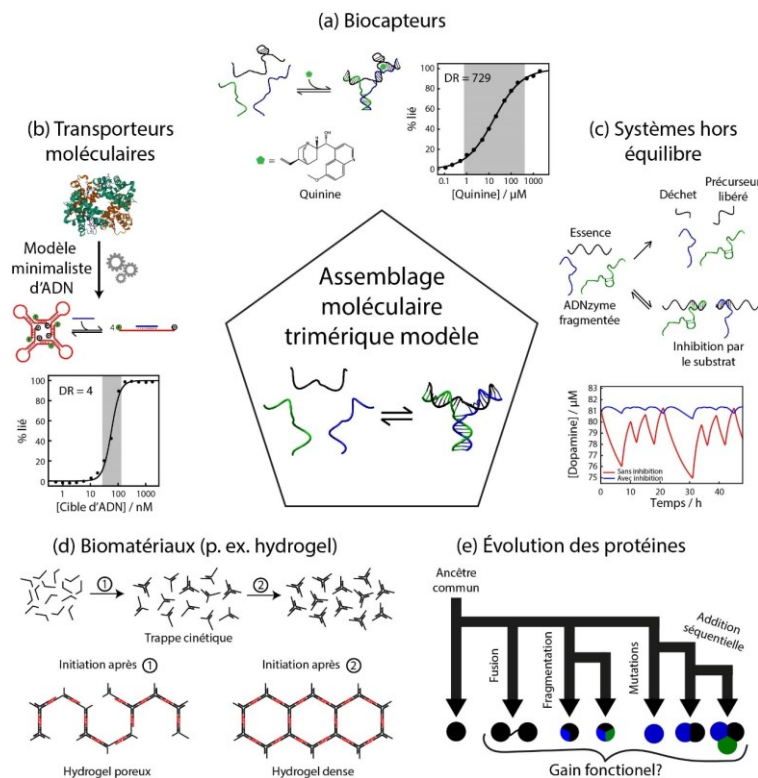


de programmer la plage dynamique d'un interrupteur afin d'obtenir des interrupteurs sensibles ou des interrupteurs avec une plage dynamique étendue. Comme démontré dans le chapitre 4, un activateur multivalent favorisant la formation d'un assemblage trimérique permet un contrôle rationnel sur la plage dynamique d'un interrupteur.

À l'instar d'un activateur allostérique qui fonctionne en changeant la conformation de l'interrupteur, un activateur qui favorise la formation d'un assemblage trimérique fonctionne simplement en ajoutant une nouvelle interface d'interaction pour le ligand. Pour mieux comprendre ce nouveau mécanisme de régulation, le même interrupteur d'ADN ayant servi à démystifier les conséquences d'un activateur allostérique fut employé<sup>33</sup>. Il fut démontré, comme envisagé, qu'un activateur allostérique fonctionne en changeant l'équilibre d'interconversion ( $K_S$ ) de l'interrupteur. En contraste, un effecteur qui favorise la formation d'un assemblage trimérique n'affecte pas l'équilibre d'interconversion, mais plutôt ajoute une interface d'interaction supplémentaire. Ainsi, l'activateur possède deux surfaces d'interaction, une pour l'interrupteur et une autre pour le ligand, d'où son nom d'activateur multivalent. L'avantage d'un tel mécanisme est qu'il ne nécessite pas la création d'un activateur qui doit modifier la conformation de l'interrupteur, ce qui n'est pas toujours évident d'un point de vue de sa conceptualisation puisque cela nécessite une optimisation précise de la différence d'énergie entre les deux conformations de l'interrupteur<sup>19</sup>. Ici, un activateur multivalent n'affecte pas l'interrupteur, qui lui, demeure inchangé à la suite de la liaison de l'activateur. Il vient plutôt se lier à l'interrupteur afin d'élargir la surface d'interaction entre l'interrupteur et le ligand. Ainsi, ce système est très modulaire puisque chaque composante de l'assemblage trimérique est composée de deux domaines d'interaction interagissant avec l'un des deux domaines des autres composantes. Ainsi, chaque domaine peut être conceptualisé indépendamment les uns des autres, puis simplement connecté l'un avec l'autre par un espaceur, comme démontré dans le chapitre 3.

Une activation multivalente permet de programmer facilement et indépendamment l'affinité apparente du ligand et la plage dynamique d'un interrupteur moléculaire. Au contraire d'un activateur allostérique qui ne permet pas de programmer la plage dynamique, un activateur multivalent permet de programmer la réponse d'un interrupteur afin d'obtenir une réponse sensible ( $PD = 9$ ) ou étendue ( $PD = 729$ ). Cela est très important, par exemple, pour la

conceptualisation d'interrupteurs moléculaires servant à la création de biocapteur. En effet, autant il existe plusieurs stratégies pour rendre un interrupteur plus sensible en réduisant sa plage dynamique<sup>34</sup>, il existe très peu de stratégies moléculaires permettant d'élargir la plage dynamique et ainsi détecter et quantifier la présence d'un analyte se trouvant à des concentrations très variables<sup>35</sup>. Typiquement, la méthode employée pour élargir la plage dynamique consiste à utiliser divers interrupteurs moléculaires ayant chacun un seuil d'activation différent<sup>36-39</sup>. Cependant, cette méthode nécessite d'optimiser l'affinité des différents interrupteurs de sorte que chaque interrupteur possède une différence d'affinité d'environ 30-fois les uns des autres, ce qui permet de maximiser la linéarité de la courbe dose-réponse logarithmique<sup>38, 39</sup>. Cela contraste avec la stratégie d'activateur multivalent qui ne nécessite aucune optimisation au niveau de l'interrupteur, mais simplement l'ajout modulaire de domaines d'interaction supplémentaires.



**Figure 5.1 – Le futur des assemblages moléculaires multimériques.** (a) Conceptualisation de biocapteurs avec plages dynamiques optimisés. Résultat préliminaire d'un assemblage trimérique d'ADN comprenant un aptamère de quinine. (b) Mise au point de transporteurs moléculaires coopératifs et de systèmes de délivrances de médicaments. Résultat préliminaire d'un autoassemblage tétramérique ultrasensible analogue à l'hémoglobine comprenant quatre tiges-boucles. (c) Développement de système hors équilibre. (d) Les règles thermodynamiques et

cinétiques découvertes par le biais d'un système autoassemblé trimérique modèle ouvrent la porte à l'optimisation des propriétés de biomatériaux (*p. ex.* un hydrogel). (e) Des modèles simplifiés d'ADN permettront de répondre à d'autres questions fondamentales sur l'évolution des protéines. La structure de l'hémoglobine provient de la « Protein Data Bank » (rcsb.org, PDB ID: 1A3N)<sup>40</sup>. Le panneau c fut adapté et traduit de la référence 27, © 2010 WILEY Periodicals, Inc. Les résultats préliminaires furent obtenus par moi-même.

## 5.2 Perspectives

Cette thèse démontre comment une meilleure compréhension des mécanismes de régulation naturels, *c.-à-d.* des assemblages moléculaires coopératifs, ouvre la porte à la conceptualisation de nouvelles stratégies pour optimiser les performances de nanomachines artificielles. À cet effet, les nombreux avantages des assemblages moléculaires permettent d'envisager plusieurs avenues où de tels systèmes autoassemblés auront possiblement un impact positif. Les assemblages moléculaires multimériques s'ajoutent donc au « coffre à outils » des chimistes, des biochimistes et des ingénieurs moléculaires qui cherchent à mettre au point les nouvelles nanotechnologies de demain.

### 5.2.1 Conceptualisation de biocapteurs

Les assemblages moléculaires à plusieurs composantes permettent la programmation de la plage dynamique de biocapteurs. Les biocapteurs consistent en des éléments de biorecognition qui peuvent traduire un événement de liaison en un signal détectable et mesurable<sup>41</sup>. Grâce à l'utilisation d'interrupteurs moléculaires, il est possible de mettre au point des systèmes moléculaires pouvant détecter spécifiquement un analyte en changeant de conformation ou d'état, ce qui induit la production d'un signal quantifiable proportionnel à la concentration de l'analyte. Toutefois, dans la conceptualisation de ces dispositifs de détection moléculaires, il est parfois nécessaire d'optimiser divers paramètres entourant l'élément de biorecognition et son mécanisme de reconnaissance (*p. ex.*, sa sensibilité, sa spécificité, sa plage dynamique, son temps de réponse, etc.)<sup>42</sup>. À cet égard, l'utilisation d'assemblages moléculaires offrirait un moyen par lequel il serait possible de programmer la plage dynamique d'un biocapteur indépendamment de la spécificité du ligand puisque les composantes supplémentaires n'affecteront pas la conformation ou l'état du récepteur. De plus, par la présence de pièges cinétiques, il sera possible de programmer le temps d'assemblage, donc du

même coup le temps de réponse d'un biocapteur multimérique. Cela facilitera, par exemple, la création de systèmes séquentiels où des événements de purification, de mélange, de liaison ou de détection surviendront à différents moments dans divers compartiments<sup>43</sup>. Cependant, l'assemblage de plusieurs composantes est typiquement plus lent qu'un équivalent unimoléculaire ou dimérique ce qui peut mener à des temps de réponse au-delà de 5 minutes, qui est typiquement considéré comme le seuil critique pour la commercialisation de biocapteurs. Une optimisation de la cinétique d'assemblage sera donc primordiale au bon fonctionnement d'un biocapteur multimérique.

Les assemblages d'ADN multimériques peuvent facilement être modifiés pour inclure des séquences d'aptamères, *c.-à-d.* des brins d'ADN pouvant interagir spécifiquement avec un ligand. Par exemple, l'un des brins d'un assemblage trimérique peut être remplacé par la séquence précise d'un aptamère qui, en absence de son ligand, permettra la formation de l'assemblage et, inversement, en présence de son ligand empêchera la formation de l'assemblage. Une autre stratégie consiste à exploiter la fragmentation d'un aptamère afin d'incorporer les deux fragments dans l'un des bras d'une jonction à trois branches (Fig. 5.1a). Une telle stratégie, comme illustrée avec l'aptamère de quinine, permet la création d'un biocapteur fluorimétrique ayant une plage dynamique étendue d'environ 729 (*c.-à-d.*,  $9^3$ ) (Fig. 5.1a, résultat préliminaire). Évidemment, plusieurs questions demeurent en suspens quant à la viabilité de ces assemblages moléculaires à base d'ADN pour la création, par exemple, de biocapteurs électrochimiques ou plasmoniques qui impliqueraient des phénomènes d'hybridation d'ADN à la surface d'une électrode ou d'une puce (*p. ex.* encombrement stérique, temps de diffusion vers la surface, temps de réponse, rendement d'assemblage, etc.). À noter que divers biocapteurs impliquant des jonctions à trois branches d'ADN ont démontré des temps de réponse d'environ 30 minutes, illustrant, comme mentionné précédemment, qu'une optimisation de la vitesse d'assemblage est primordiale pour ce type d'architecture<sup>44-46</sup>. Toutefois, ces mêmes biocapteurs ont démontré qu'un assemblage moléculaire permet de réduire significativement le signal de fond en évitant la préassociation des composantes puisque leurs affinités intrinsèques les unes pour les autres sont plus faibles. Ainsi, l'association de toutes les composantes est nécessaire pour favoriser l'assemblage. Finalement, ces découvertes sur les assemblages moléculaires ne s'appliquent pas

uniquement aux systèmes à base d'ADN et pourraient, similairement, être exploitées dans la conceptualisation et l'optimisation de biocapteurs à bases de peptides<sup>47</sup>, de protéines<sup>48</sup> ou de molécules organiques<sup>49</sup>.

## **5.2.2 Transporteurs moléculaires et systèmes de délivrance de médicaments coopératifs à base d'ADN**

Grâce à son habilité à lier diverses biomolécules, l'ADN est souvent utilisé dans la conceptualisation de systèmes permettant une libération contrôlée de médicaments. Par exemple, *Ricci et al.* ont développé de simples interrupteurs d'ADN pouvant libérer un cargo moléculaire en présence d'ADN<sup>50</sup>, de variation de pH<sup>51</sup> ou de facteurs de transcription<sup>52</sup>. À plus grande échelle, un interrupteur d'ADN fut introduit par *Church et al.* dans un origami d'ADN permettant l'encapsulation d'une plus grande quantité de médicaments ainsi qu'une libération contrôlée près de la surface de cellules cancéreuses<sup>53</sup>. Dans tous ces exemples, un paramètre important au bon fonctionnement de ces systèmes est la création de systèmes « tout ou rien », *c.-à-d.* des systèmes pouvant répondre de manière ultrasensible à la présence d'un marqueur de maladie (*p. ex.* pH, petites molécules, protéines membranaires, cellules cancéreuses, etc.) et libérer l'entièreté du cargo moléculaire<sup>54</sup>.

Les assemblages moléculaires offrent un moyen par lequel de tels systèmes pourraient être construits<sup>55</sup>. En effet, il existe une corrélation entre le nombre de composantes et la coopérativité maximale pouvant être obtenue d'un assemblage. Ainsi, en augmentant le nombre de composantes au-delà des trois composantes employées dans cette thèse, il est envisageable d'augmenter considérablement le niveau de coopérativité. Pour mettre en lumière cette affirmation, un nouveau système modèle devra être mis au point où l'assemblage moléculaire à plusieurs composantes servira de transporteur moléculaire pour un médicament quelconque. À cet effet, le modèle suivant est proposé où divers interrupteurs moléculaires sont combinés pour former une jonction à quatre branches (Fig. 5.1b, *haut*). Les règles thermodynamiques et cinétiques découvertes dans cette thèse laissent entrevoir un rendement d'assemblage optimal en employant des espaceurs de 2T (Fig. 5.1b, *bas*, résultat préliminaire). De plus, il ne faut pas négliger l'apport coopératif provenant de l'allostérie puisque chaque interrupteur lie la même

molécule, *c.-à-d.* un brin d'ADN ou une molécule quelconque par l'entremise d'un aptamère. Ainsi, un tel modèle, puisqu'il ressemble structurellement à l'hémoglobine, pourrait aussi servir à mieux comprendre les modèles MWC et KNF d'un point de vue d'un interrupteur moléculaire et de l'ingénierie moléculaire.

### 5.2.3 Système hors équilibre : l'essence moléculaire

La présente thèse illustre comment un assemblage moléculaire peut être programmé à l'équilibre ou dans le temps par l'exploitation de pièges cinétiques. Le futur des nanotechnologies à base d'ADN, ou de tout autres systèmes supramoléculaires, réside dans la capacité à créer des systèmes moléculaires pouvant fonctionner loin de l'équilibre comme c'est le cas pour la nanomachinerie biologique<sup>30</sup>. De tels systèmes fonctionnent grâce à un apport constant d'énergie sous la forme, par exemple, de lumière, de protons ou de molécules qui permettent de pousser le système dans un état dissipatif régulé par l'apport et la dissipation de cette source énergie<sup>56</sup>. Ainsi, si cette source d'énergie est coupée, le système se dirigera vers un état thermodynamique minimal ou sera piégé cinétiquement dans un minimum local en route vers le minimum global. Un mécanisme naturel afin de maintenir cette source d'énergie chimique constante consiste à exploiter le phénomène d'auto-inhibition. En effet, un système auto-inhibé peut servir de convertisseur permettant de transformer une réserve de molécules à concentration variable en une source d'énergie moléculaire ayant une concentration fixe et constante<sup>27-29</sup>. Par exemple, le maintien de la dopamine, un neurotransmetteur responsable du « bonheur », est directement contrôlé par la tyrosine hydroxylase qui est auto-inhibée par son propre substrat, la tyrosine. Cela permet, suivant un apport élevé en tyrosine après les repas, de maintenir un niveau constant de la L-dopa, le précurseur à la synthèse de la dopamine (Fig. 5.1c)<sup>29</sup>. Ainsi, il serait envisageable de mettre au point un module d'ADN ou d'ADNzyme permettant d'obtenir un comportement similaire. L'un pourrait imaginer qu'une ADNzyme auto-inhibée pourrait réguler de manière constante la concentration de son produit qui, lui, servirait à alimenter à flux constant différentes nanomachines d'ADN<sup>57-59</sup>. Cette approche modulaire s'inscrit dans la lignée de la robotique moléculaire, plus précisément de la robotique d'ADN<sup>60-62</sup>.

## 5.2.4 Matériaux à base d'ADN

Le contrôle cinétique offert par l'exploitation des pièges cinétiques ouvre aussi la porte à l'optimisation des propriétés de matériaux à bases d'ADN. L'ADN, par exemple, peut former des hydrogels grâce à une polymérisation non covalente de jonctions d'ADN multivalentes grâce à l'utilisation de bout collant et d'un brin d'ADN agissant comme réticulateur<sup>63</sup>. De plus, le type de jonction employé (*p. ex.* une jonction à trois, quatre ou cinq branches) offre différentes propriétés d'encapsulation de médicament et de dégradation selon la porosité de leurs structures<sup>64</sup>. Afin de programmer le niveau de porosité du matériel, il est envisageable d'exploiter les pièges cinétiques afin de contrôler la nucléation d'un hydrogel (Fig. 5.1d). Un hydrogel parfait serait donc obtenu si, initialement, tous les brins d'ADN adoptent une jonction à trois branches avant d'initier la polymérisation avec l'ajout de l'agent réticulant. Toutefois, si la jonction est sous contrôle cinétique et que la majorité des brins sont emprisonnés dans un état intermédiaire dimérique, il serait possible d'obtenir des pores plus gros à la suite de l'initiation grâce à la formation d'une matrice moins réticulée. Cela pourrait donc ouvrir la porte à la création d'hydrogel stable pouvant encapsuler des biomolécules de différentes tailles ou bien permettre plus facilement leur diffusion à travers du biomatériau.

## 5.2.5 L'évolution vers des assemblages multimériques

Finalement, les résultats obtenus par l'étude d'un système modèle à base d'ADN ouvrent la porte à plusieurs questions fondamentales sur l'évolution des biomolécules et l'acquisition de complexité (Fig. 5.1e). Puisque près du deux tiers de toutes les protéines existent sous formes oligomériques, *c.-à-d.* qu'elles sont constituées de plus d'une unité protéique, il doit certainement avoir des avantages fonctionnels à être construit à partir de plusieurs composantes moléculaires<sup>8</sup>. Ici, la fragmentation d'une structure unimoléculaire en divers fragments a justement permis l'acquisition de nouvelles fonctionnalités. Est-ce que les protéines ont évolué suivant de tels évènements de fragmentation? Cette question, aussi simple soit-elle, demeure compliquée à répondre. Malgré tout, en étudiant les gènes codant pour diverses protéines, il fut observé qu'une fraction de protéines ont évolué à la suite d'un évènement de fission de gènes<sup>1-3</sup>. Toutefois, ces études, qui ont été principalement réalisées au tournant dans les années 2000, se concentraient uniquement sur les mécanismes permettant à des protéines monomériques

d'évoluer vers des systèmes oligomériques et ne se concentraient donc pas sur leurs fonctionnalités. À cet effet, il était généralement bien accepté par la communauté scientifique que l'évolution vers des systèmes multimériques devrait inévitablement être associée à des avantages évolutifs clairs ayant permis leur persistance dans le temps<sup>65</sup>. Néanmoins, de plus récentes études démontrent que l'évolution vers des systèmes multimériques serait un processus non adaptatif et purement stochastique<sup>8, 66</sup>. Ces études avancent aussi que les avantages de l'oligomérisation ne sont pas toujours présents, mais dépendent plutôt du contexte évolutif et des propriétés individuelles de la protéine<sup>6</sup>. Cela expliquerait donc pourquoi une protéine accomplissant une fonction spécifique pourrait être monomérique chez une espèce et multimérique chez une autre puisque leur environnement n'était pas identique. Ainsi, elles ont donc évolué indépendamment vers leurs fonctions respectives. Malgré tout, afin de mieux comprendre leur évolution, de meilleures techniques expérimentales devront être mises au point afin d'étudier plusieurs groupes de protéines multimériques simultanément au lieu de simplement se concentrer sur certains systèmes individuels comme c'est le cas présentement<sup>7</sup>. Il sera plus qu'intéressant de suivre comment les conclusions de cette thèse affecteront les hypothèses futures des biochimistes en ce qui concerne l'évolution naturelle des protéines.

Sous un autre angle, nos expériences quotidiennes nous démontrent clairement qu'un système composé de plusieurs composantes est typiquement plus « fonctionnel » qu'un système en ayant moins. Les technologies autour de nous illustrent bien comment l'ajout de plusieurs « gadgets » à un système existant permet une diversification et une amélioration de sa fonction. Il suffit de penser aux machines d'aujourd'hui qui, grâce à une myriade de caméras, de capteurs et d'ordinateurs, peuvent détecter les variations dans leurs environnements pour prendre des décisions par elles-mêmes. Autant ces ajouts rendent leur fabrication plus difficile et coûteuse, leur complexité rapporte sur le plan de leur versatilité, leur fidélité et leur fonctionnalité. Cette approche séquentielle où l'on ajoute des composantes fonctionnelles à une machine déjà fonctionnelle est toutefois très différente des méthodes évolutives employées dans cette thèse. Ici, par exemple, une évolution fonctionnelle fut obtenue par le biais d'une fragmentation d'une nanostructure existante où l'autoassemblage de la nanostructure est possible au moyen de plusieurs interactions intermoléculaires non covalentes. Il est plutôt difficile d'imaginer



l'évolution d'une machine macroscopique découlant d'événements de fragmentation sachant que ces composantes sont typiquement fixées ensemble de manière « covalente ». Ce contraste permet d'entrevoir que l'évolution fractionnelle d'une nanostructure moléculaire est probablement une caractéristique unique aux systèmes nanométriques et permet donc d'imaginer de nouvelles voies évolutives au niveau nanométrique qui sont fondamentalement différentes des méthodes d'ingénieries typiquement employées dans un monde macroscopique.

### 5.3 Références

1. Pasek, S.; Risler, J.-L.; Brézellec, P., Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. *Bioinformatics* **2006**, *22* (12), 1418-1423.
2. Marsh, J. A.; Teichmann, S. A., Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **2015**, *84* (1), 551-575.
3. Kummerfeld, S. K.; Teichmann, S. A., Relative rates of gene fusion and fission in multi-domain proteins. *Trends Genet.* **2005**, *21* (1), 25-30.
4. Seeman, N. C.; Sleiman, H. F., DNA nanotechnology. *Nat. Rev. Mater.* **2017**, *3* (1), 17068.
5. Aloy, P.; Pichaud, M.; Russell, R. B., Protein complexes: structure prediction challenges for the 21st century. *Curr. Opin. Struct. Biol.* **2005**, *15* (1), 15-22.
6. Hagner, K.; Setayeshgar, S.; Lynch, M., Stochastic protein multimerization, activity, and fitness. *Phys. Rev. E* **2018**, *98* (6), 062401.
7. Hashimoto, K.; Nishi, H.; Bryant, S.; Panchenko, A. R., Caught in self-interaction: evolutionary and functional mechanisms of protein homooligomerization. *Phys. Biol.* **2011**, *8* (3), 035007.
8. Lynch, M., The evolution of multimeric protein assemblages. *Mol. Biol. Evol.* **2012**, *29* (5), 1353-1366.
9. Debais, M.; Lelievre, A.; Smietana, M.; Müller, S., Splitting aptamers and nucleic acid enzymes for the development of advanced biosensors. *Nucleic Acids Res.* **2020**, *48* (7), 3400-3422.
10. Dolberg, T. B.; Meger, A. T.; Boucher, J. D.; Corcoran, W. K.; Schauer, E. E.; Prybutok, A. N.; Raman, S.; Leonard, J. N., Computation-guided optimization of split protein systems. *Nat. Chem. Biol.* **2021**, *17* (5), 531-539.

11. Wang, R.; Zhang, Q.; Zhang, Y.; Shi, H.; Nguyen, K. T.; Zhou, X., Unconventional Split Aptamers Cleaved at Functionally Essential Sites Preserve Biorecognition Capability. *Anal. Chem.* **2019**, *91* (24), 15811-15817.
12. Dagliyan, O.; Krokhotin, A.; Ozkan-Dagliyan, I.; Deiters, A.; Der, C. J.; Hahn, K. M.; Dokholyan, N. V., Computational design of chemogenetic and optogenetic split proteins. *Nat. Commun.* **2018**, *9* (1), 4042.
13. Han, Y.; Branon, T. C.; Martell, J. D.; Boassa, D.; Shechner, D.; Ellisman, M. H.; Ting, A., Directed Evolution of Split APEX2 Peroxidase. *Directed Evolution of Split APEX2 Peroxidase* **2019**, *14* (4), 619-635.
14. Romero, P. A.; Arnold, F. H., Exploring protein fitness landscapes by directed evolution. *Exploring protein fitness landscapes by directed evolution* **2009**, *10* (12), 866-876.
15. Michnick, S. W.; Ear, P. H.; Manderson, E. N.; Remy, I.; Stefan, E., Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nat. Rev. Drug Discovery* **2007**, *6* (7), 569-582.
16. Shekhawat, S. S.; Ghosh, I., Split-protein systems: beyond binary protein–protein interactions. *Curr. Opin. Chem. Biol.* **2011**, *15* (6), 789-797.
17. Laing, B. M.; Juliano, R. L., DNA Three-Way Junctions Stabilized by Hydrophobic Interactions for Creation of Functional Nanostructures. *ChemBioChem* **2015**, *16* (9), 1284-1287.
18. Takezawa, Y.; Yoneda, S.; Duprey, J.-L. H. A.; Nakama, T.; Shionoya, M., Metal-responsive structural transformation between artificial DNA duplexes and three-way junctions. *Chem. Sci.* **2016**, *7* (5), 3006-3010.
19. Harroun, S. G.; Prévost-Tremblay, C.; Lauzon, D.; Desrosiers, A.; Wang, X.; Pedro, L.; Vallée-Bélisle, A., Programmable DNA switches and their applications. *Nanoscale* **2018**, *10* (10), 4607-4641.
20. Vallée-Bélisle, A.; Ricci, F.; Plaxco, K. W., Thermodynamic basis for the optimization of binding-induced biomolecular switches and structure-switching biosensors. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (33), 13802.
21. SantaLucia, J.; Hicks, D., The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33* (1), 415-440.

22. Cisse, I. I.; Kim, H.; Ha, T., A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nat. Struct. Mol. Biol.* **2012**, *19* (6), 623-627.
23. Cabantous, S.; Nguyen, H. B.; Pedelacq, J.-D.; Koraïchi, F.; Chaudhary, A.; Ganguly, K.; Lockard, M. A.; Favre, G.; Terwilliger, T. C.; Waldo, G. S., A New Protein-Protein Interaction Sensor Based on Tripartite Split-GFP Association. *Sci. Rep.* **2013**, *3* (1), 2854.
24. Tsai, C.-J.; Lin, S. L.; Wolfson, H. J.; Nussinov, R., Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect. *Protein Sci.* **1997**, *6* (1), 53-64.
25. Marianayagam, N. J.; Sunde, M.; Matthews, J. M., The power of two: protein dimerization in biology. *Trends Biochem. Sci.* **2004**, *29* (11), 618-625.
26. Hunter, C. A.; Anderson, H. L., What is Cooperativity? *Angew. Chem. Int. Ed.* **2009**, *48* (41), 7488-7499.
27. Best, J. A.; Nijhout, H. F.; Reed, M. C., Homeostatic mechanisms in dopamine synthesis and release: a mathematical model. *Theor. Biol. Med. Modell.* **2009**, *6* (1), 21.
28. Reed, M.; Best, J.; Golubitsky, M.; Stewart, I.; Nijhout, H. F., Analysis of Homeostatic Mechanisms in Biochemical Networks. *Bull. Math. Biol.* **2017**, *79* (11), 2534-2557.
29. Reed, M. C.; Lieb, A.; Nijhout, H. F., The biological significance of substrate inhibition: A mechanism with diverse functions. *BioEssays* **2010**, *32* (5), 422-429.
30. Mattia, E.; Otto, S., Supramolecular systems chemistry. *Nat. Nanotechnol.* **2015**, *10* (2), 111-119.
31. Sanchez-Ruiz, J. M., Protein kinetic stability. *Biophys. Chem.* **2010**, *148* (1), 1-15.
32. Kim, Y. E.; Hipp, M. S.; Bracher, A.; Hayer-Hartl, M.; Ulrich Hartl, F., Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annu. Rev. Biochem.* **2013**, *82* (1), 323-355.
33. Ricci, F.; Vallée-Bélisle, A.; Porchetta, A.; Plaxco, K. W., Rational Design of Allosteric Inhibitors and Activators Using the Population-Shift Model: In Vitro Validation and Application to an Artificial Biosensor. *J. Am. Chem. Soc.* **2012**, *134* (37), 15177-15180.
34. Ricci, F.; Vallée-Bélisle, A.; Simon, A. J.; Porchetta, A.; Plaxco, K. W., Using Nature's "Tricks" To Rationally Tune the Binding Properties of Biomolecular Receptors. *Acc. Chem. Res.* **2016**, *49* (9), 1884-1892.

35. Labib, M.; Sargent, E. H.; Kelley, S. O., Electrochemical Methods for the Analysis of Clinically Relevant Biomolecules. *Chem. Rev.* **2016**, *116* (16), 9001-9090.
36. Porchetta, A.; Vallée-Bélisle, A.; Plaxco, K. W.; Ricci, F., Using Distal-Site Mutations and Allosteric Inhibition To Tune, Extend, and Narrow the Useful Dynamic Range of Aptamer-Based Sensors. *J. Am. Chem. Soc.* **2012**, *134* (51), 20601-20604.
37. Porchetta, A.; Vallée-Bélisle, A.; Plaxco, K. W.; Ricci, F., Allosterically Tunable, DNA-Based Switches Triggered by Heavy Metals. *J. Am. Chem. Soc.* **2013**, *135* (36), 13238-13241.
38. Kang, D.; Vallée-Bélisle, A.; Porchetta, A.; Plaxco, K. W.; Ricci, F., Re-engineering Electrochemical Biosensors To Narrow or Extend Their Useful Dynamic Range. *Angew. Chem. Int. Ed.* **2012**, *51* (27), 6717-6721.
39. Vallée-Bélisle, A.; Ricci, F.; Plaxco, K. W., Engineering Biosensors with Extended, Narrowed, or Arbitrarily Edited Dynamic Range. *J. Am. Chem. Soc.* **2012**, *134* (6), 2876-2879.
40. Tame, J. R.; Vallone, B., The structures of deoxy human haemoglobin and the mutant Hb Tyr $\alpha$ 42His at 120 K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2000**, *56* (Pt 7), 805-11.
41. Turner, A. P. F., Biosensors: sense and sensibility. *Chem. Soc. Rev.* **2013**, *42* (8), 3184-3196.
42. Dincer, C.; Bruch, R.; Costa-Rama, E.; Fernández-Abedul, M. T.; Merkoçi, A.; Manz, A.; Urban, G. A.; Güder, F., Disposable Sensors in Diagnostics, Food, and Environmental Monitoring. *Adv. Mater.* **2019**, *31* (30), 1806739.
43. Derkus, B., Applying the miniaturization technologies for biosensor design. *Biosens. Bioelectron.* **2016**, *79*, 901-913.
44. Feng, L.; Zhao, C.; Xiao, Y.; Wu, L.; Ren, J.; Qu, X., Electrochemical DNA three-way junction based sensor for distinguishing chiral metallo-supramolecular complexes. *Chem. Commun.* **2012**, *48* (55), 6900-6902.
45. Hu, Z.; Zhao, B.; Miao, P.; Hou, X.; Xing, F.; Chen, Y.; Feng, L., Three-way junction DNA based electrochemical biosensor for microRNAs detection with distinguishable locked nucleic acid recognition and redox cycling signal amplification. *J. Electroanal. Chem.* **2021**, *880*, 114861.
46. Cui, L.; Wu, J.; Li, M.; Ju, H., Highly sensitive electrochemical detection of mercury (II) via single ion-induced three-way junction of DNA. *Electrochem. Commun.* **2015**, *59*, 77-80.
47. Liu, Q.; Wang, J.; Boyd, B. J., Peptide-based biosensors. *Talanta* **2015**, *136*, 114-127.

48. Ibraheem, A.; Campbell, R. E., Designs and applications of fluorescent protein-based biosensors. *Curr. Opin. Chem. Biol.* **2010**, *14* (1), 30-36.
49. Lim, S.; Kuang, Y.; Ardoña, H. A. M., Evolution of Supramolecular Systems Towards Next-Generation Biosensors. *The future of supramolecular biosensors* **2021**, *9*.
50. Mariottini, D.; Idili, A.; Vallée-Bélisle, A.; Plaxco, K. W.; Ricci, F., A DNA Nanodevice That Loads and Releases a Cargo with Hemoglobin-Like Allosteric Control and Cooperativity. *Nano Lett.* **2017**, *17* (5), 3225-3230.
51. Porchetta, A.; Idili, A.; Vallée-Bélisle, A.; Ricci, F., General Strategy to Introduce pH-Induced Allostery in DNA-Based Receptors to Achieve Controlled Release of Ligands. *Nano Lett.* **2015**, *15* (7), 4467-4471.
52. Rossetti, M.; Ranallo, S.; Idili, A.; Palleschi, G.; Porchetta, A.; Ricci, F., Allosteric DNA nanoswitches for controlled release of a molecular cargo triggered by biological inputs. *Chem. Sci.* **2017**, *8* (2), 914-920.
53. Douglas Shawn, M.; Bachelet, I.; Church George, M., A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science* **2012**, *335* (6070), 831-834.
54. Desrosiers, A.; Vallée-Bélisle, A., Nature-inspired DNA switches: applications in medicine. *Nanomedicine* **2017**, *12* (3), 175-179.
55. Vargason, A. M.; Anselmo, A. C.; Mitragotri, S., The evolution of commercial drug delivery technologies. *Nat. Biomed. Eng.* **2021**, *5* (9), 951-967.
56. Ragazzon, G.; Prins, L. J., Energy consumption in chemical fuel-driven self-assembly. *Nat. Nanotechnol.* **2018**, *13* (10), 882-889.
57. Del Grosso, E.; Ragazzon, G.; Prins, L. J.; Ricci, F., Fuel-Responsive Allosteric DNA-Based Aptamers for the Transient Release of ATP and Cocaine. *Angew. Chem. Int. Ed.* **2019**, *58* (17), 5582-5586.
58. Peng, P.; Shi, L.; Wang, H.; Li, T., A DNA nanoswitch-controlled reversible nanosensor. *Nucleic Acids Res.* **2017**, *45* (2), 541-546.
59. Ramezani, H.; Dietz, H., Building machines with DNA molecules. *Nat. Rev. Genet.* **2020**, *21* (1), 5-26.

60. Thubagere Anupama, J.; Li, W.; Johnson Robert, F.; Chen, Z.; Doroudi, S.; Lee Yae, L.; Izatt, G.; Wittman, S.; Srinivas, N.; Woods, D.; Winfree, E.; Qian, L., A cargo-sorting DNA robot. *Science* **2017**, *357* (6356), eaan6558.
61. Kopperger, E.; List, J.; Madhira, S.; Rothfischer, F.; Lamb Don, C.; Simmel Friedrich, C., A self-assembled nanoscale robotic arm controlled by electric fields. *Science* **2018**, *359* (6373), 296-301.
62. Liu, H.; Wang, J.; Song, S.; Fan, C.; Gothelf, K. V., A DNA-based system for selecting and displaying the combined result of two input variables. *Nat. Commun.* **2015**, *6* (1), 10089.
63. Bush, J.; Hu, C.-H.; Veneziano, R., Mechanical Properties of DNA Hydrogels: Towards Highly Programmable Biomaterials. *Appl. Sci.* **2021**, *11* (4).
64. Um, S. H.; Lee, J. B.; Park, N.; Kwon, S. Y.; Umbach, C. C.; Luo, D., Enzyme-catalysed assembly of DNA hydrogel. *Nat. Mater.* **2006**, *5* (10), 797-801.
65. Gaber, A.; Pavšič, M., Modeling and Structure Determination of Homo-Oligomeric Proteins: An Overview of Challenges and Current Approaches. *Int. J. Mol. Sci.* **2021**, *22* (16).
66. Lynch, M., Evolutionary diversification of the multimeric states of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (30), E2821-E2828.

## Annexes

### MATLAB® simulations.

#### Thermodynamic Simulations

The assembly of the trimer can be described using six equilibriums (see Fig. S2.18). We described this system using the seven species (A, B, C, AB, AC, BC, and ABC), the 3 equations of mass conservation, the 3 equilibrium equations for the formation of each possible dimer (AB, AC, and BC), and the equation for the global assembly of the trimer (ABC). This system of 7 variables and 7 equations can be numerically solved using MATLAB® by providing the initial concentration of each component ( $[A]_T$ ,  $[B]_T$  and  $[C]_T$ ) and the related  $\Delta G$  of each equilibrium ( $\Delta G_{Dim,AB}$ ,  $\Delta G_{Dim,AC}$ ,  $\Delta G_{Dim,BC}$  and  $\Delta G_G$ ). There is no need to describe all the  $\Delta G_{Tri}$  because they are related to the  $\Delta G_{Dim}$  and  $\Delta G_G$  by this relation:  $\Delta G_G = \Delta G_{Dim} + \Delta G_{Tri}$ .

$$[A]_T = [A]_{eq} + [AB]_{eq} + [AC]_{eq} + [ABC]_{eq}$$

$$[B]_T = [B]_{eq} + [AB]_{eq} + [BC]_{eq} + [ABC]_{eq}$$

$$[C]_T = [C]_{eq} + [BC]_{eq} + [AC]_{eq} + [ABC]_{eq}$$

$$K_{Dim,AB} = \frac{[A]_{eq} \cdot [B]_{eq}}{[AB]_{eq}}; \Delta G_{Dim,AB} = -RT \ln(K_{AB})$$

$$K_{Dim,AC} = \frac{[A]_{eq} \cdot [C]_{eq}}{[AC]_{eq}}; \Delta G_{Dim,AC} = -RT \ln(K_{AC})$$

$$K_{Dim,BC} = \frac{[B]_{eq} \cdot [C]_{eq}}{[BC]_{eq}}; \Delta G_{Dim,BC} = -RT \ln(K_{BC})$$

$$K_G = \frac{[A]_{eq} \cdot [B]_{eq} \cdot [C]_{eq}}{[ABC]_{eq}}; \Delta G_G = -RT \ln(K_{ABC})$$

## Kinetic Simulations

The assembly of the trimer can be described by six equilibriums (see Fig. S2.18). For the simulations, we assumed that all dimeric equilibriums (Fig. S2.20) and trimeric equilibriums (Fig. S2.21) have identical rate constants, which have been calculated previously. With these equilibriums, we derived all the rate equations for A, B, C, AB, AC, BC, and ABC. We then solve these equations using Euler's method with MATLAB®.

$$\frac{dA}{dt} = -k_1[A][B] + k_{-1}[AB] - k_1[A][C] + k_{-1}[AC] - k_2[A][BC] + \frac{1}{3}k_{-2}[ABC]$$

$$\frac{dB}{dt} = -k_1[A][B] + k_{-1}[AB] - k_1[B][C] + k_{-1}[BC] - k_2[B][AC] + \frac{1}{3}k_{-2}[ABC]$$

$$\frac{dC}{dt} = -k_1[B][C] + k_{-1}[BC] - k_1[A][C] + k_{-1}[AC] - k_2[C][AB] + \frac{1}{3}k_{-2}[ABC]$$

$$\frac{dAB}{dt} = k_1[A][B] - k_{-1}[AB] - k_2[C][AB] + \frac{1}{3}k_{-2}[ABC]$$

$$\frac{dAC}{dt} = k_1[A][C] - k_{-1}[AC] - k_2[B][AC] + \frac{1}{3}k_{-2}[ABC]$$

$$\frac{dBC}{dt} = k_1[B][C] - k_{-1}[BC] - k_2[A][BC] + \frac{1}{3}k_{-2}[ABC]$$

$$\frac{dABC}{dt} = k_2[A][BC] + k_2[B][AC] + k_2[C][AB] - k_{-2}[ABC]$$

\* Of note, a statistical factor, 1/3, is used for the dissociation of ABC since this dissociation can occur in three different ways.