



**HAL**  
open science

# Auto-assemblage de la protéine bactérienne Hfq, actrice du métabolisme de l'ARN : rôle structural du domaine C-terminal

Antoine Malabirade

► **To cite this version:**

Antoine Malabirade. Auto-assemblage de la protéine bactérienne Hfq, actrice du métabolisme de l'ARN : rôle structural du domaine C-terminal. Biochimie [q-bio.BM]. Université Paris Saclay (COMUE), 2017. Français. NNT : 2017SACLS234 . tel-01661404

**HAL Id: tel-01661404**

**<https://theses.hal.science/tel-01661404>**

Submitted on 11 Dec 2017

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

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

# Auto-assemblage de la protéine bactérienne Hfq, actrice du métabolisme de l'ARN : rôle structural du domaine C-terminal.

Thèse de doctorat de l'Université Paris-Saclay  
Préparée à l'Université Paris-Sud

École doctorale n°571 Sciences Chimiques :  
Molécules, Matériaux, Instrumentation et Biosystèmes (2MIB)  
Spécialité de doctorat : Chimie

Thèse présentée et soutenue à Saclay, le 6 octobre 2017, par

**M. Antoine Malabirade**

Composition du Jury :

<b>Mme Nara Figueroa-Bossi</b> Directrice de recherche, CNRS UMR 9198, Institut de Biologie Intégrative de la Cellule Université Paris-Saclay	Présidente
<b>M. Marc Boudvillain</b> Directeur de recherche, CNRS UPR 4301, Centre de biophysique moléculaire	Rapporteur
<b>M. Carlos Óscar Sánchez Sorzano</b> Professeur, Université San Pablo CEU Bioengineering Laboratory	Rapporteur
<b>Mme Andrea Gomez-Zavaglia</b> Professeure, Université nationale de La Plata CIDCA	Examinatrice
<b>Mme Véronique Arluison</b> Maître de conférences, Université Paris Diderot UMR 12, Laboratoire Léon Brillouin Université Paris-Saclay	Directrice de thèse





*A mes grands-parents.*

# Remerciements

Ce travail de thèse a été effectué au sein du Laboratoire Léon Brillouin, Unité Mixte de Recherche n°12 CEA-CNRS, localisé au sein du centre du Commissariat à l'Énergie Atomique et aux Énergies Alternatives de Saclay. Pour ce faire, j'ai bénéficié d'un financement sous contrat doctoral par l'Université Paris-Sud ainsi que des moyens matériels mis à disposition par le LLB. Je tiens à remercier Christiane Alba-Simionesco, directrice du LLB, ainsi que Jean-Paul Visticot et Éric Eliot, tous deux directeurs adjoints, pour leur accueil et leur soutien.

Je souhaite évidemment adresser mes profonds remerciements à Véronique Arluison, ma directrice de thèse. Grâce à toi, j'ai pu expérimenter de nombreuses techniques passionnantes, rencontrer des opportunités professionnelles extraordinaires en France et bien plus loin. Ton soutien m'a été précieux tout au long de cette thèse, dans mes recherches comme dans mes enseignements. J'adresse aussi mes remerciements sincères à M. Marc Boudvillain (DR) et M. Carlos Óscar Sánchez Sorzano (Professeur) pour avoir accepté d'être rapporteurs de ma thèse, ainsi que Mme Nara Figueroa-Bossi (DR) et Mme Andrea Gomez-Zavaglia (Professeure) pour avoir accepté d'être examinatrices.

Je remercie les très nombreux collègues et amis avec qui j'ai travaillé et vécu mes journées (ou parfois mes nuits) dans les différents laboratoires, notamment :

- Au LLB d'abord (et parfois en dehors), le fameux troisième étage et l'équipe Matière Molle, Biophysique et Nanomatériaux, ainsi que les nombreux stagiaires et doctorants, post-doctorants, passés et présents, qui se reconnaîtront, et ont tous montré leur gentillesse dans des moments parfois difficiles ...mais aussi dans des moments bien plus détendus ! L'ambiance de travail a toujours été agréable et c'est grâce à vous tous. Un merci particulier à Marc-Antoine Thual et Marguerite Leang, la synchronisation du contrat doctoral ça rapproche toujours ! *Tusind tak to you Mikkel Hartmann Jensen for helping me to procrastinate very efficiently, something not easy at all !* Merci aussi à David Partouche pour l'aide quotidienne au labo, ainsi que les nombreux échanges scientifiques (et non scientifiques !) pertinents que nous avons pu avoir ;
- À l'Institut Curie d'Orsay, l'ensemble de l'équipe de Sergio Marco et notamment Sylvain Trépout et son usine à café sans laquelle rien n'aurait été possible, un énorme merci pour ton aide en cryo-TEM. Merci aussi à Liliane Mouawad pour ses remarques judicieuses ;
- À l'*Instituto de Catálisis y Petroleoquímica* de l'Université Autonome de Madrid, *muchas gracias a Marisela Velez y Ileana Marquez por su ayuda en AFM y su buen humor. ¡ Nunca serán olvidados los desayunos, los churros y el jamón ibérico tampoco !*

- À la *National University of Singapore*, thank you very much Johan Van der Maarel, Jiang Kai and Fan Liu for your warm welcome in the lab, your help in nanofluidics and your kindness. Wonderful hikes, and this chilli crab was a blast !
- À l'Université Paris Diderot où j'ai enseigné avec plaisir pendant trois années : merci aux collègues qui m'ont aidé à faire de ces cours une réussite (je l'espère !), notamment Josiane Wantyghem et Sophie Lefèvre pour les TD et TP de métabolisme ainsi que Florent Busi pour les TP d'enzymologie.
- Et puis aussi : Frank Wien à DISCO (SOLEIL) ; Alexandre Dazzi, Ariane Deniset, Jérémie Mathurin au LCP (Paris-Sud) ; Jean-François Berret au MSC (Université Paris Diderot) ; Frédéric Gobeaux au LIONS (CEA) ; Nara Figueroa Bossi, Lionello Bossi, Benoît Guillemardet et Mathilde Ratel ainsi que Jérôme Seguin à l'I2BC (CEA-CNRS) ; Christophe Lavelle au Museum National d'Histoire Naturelle de Paris ; Géraldine Carrot au LICSEN (CEA)...

Je voudrais aussi remercier ici Gilles Truan, directeur de recherches au LISBP de Toulouse qui à travers ma participation au concours iGEM m'a convaincu de poursuivre ma route en thèse ; de même pour Thomas Lautier, Claire Moulis, et Claude Maranges, tous trois au LISBP et à l'INSA de Toulouse.

Au-delà de mes journées de travail, il y a évidemment tout le reste de ces presque quatre années parisiennes (sans vouloir vexer les parisiens qui me lisent : vu des Pyrénées, Antony ou Saclay c'est aussi Paris ; c'est ça le pouvoir de séparation : l'ensemble est dans le même brouillard). Merci à tous pour les innombrables activités « extra-thésiennes » essentielles à ma santé mentale et ma santé tout court, où se sont mélangés amis d'enfance, de lycée, d'école... et nouveaux venus. Sans tous les citer mais ils se reconnaîtront : Audrey et Loïc, Marion, Rémi@GuldenDraak@ENS, Tristan@Mépris, Jimmy@Davae, Stéphane@SxB (la Science avance), Caro, Simon@PowerslaveXI (*tack min vän*), Quentin@Re4cube@ASN et Sylvain@Ertemel, Benjamin@BenJ et Florian@NaiRolF et l'ensemble des fous de rails tordus (iSpeed>Katun), Julien@Juju@Padanouille, Simon@Todor et les Bleus du 91 et à côté, Ronan@Roro, Baptiste@Ashes... Sans oublier les Toulousains et Vicquois (ou assimilés) Maximilien@VieuxFrère, Matthieu@Narzhùl, Pierre@VéritéOuVérité, Marie, Youyou, Émilie et plus généralement la team PMF, la 5yn4p5@INSA au complet *of course*, ainsi que mes compagnons iGEMers Sarah, Manon et les autres GB47... Et énormément d'autres encore qui par leur présence petite ou grande ont contribué à cette aventure.

Le doctorat est évidemment plus qu'un simple contrat de travail, et la recherche ne s'arrête pas à la porte du laboratoire, pour le plus grand malheur de certains et le grand bonheur d'autres. Son accomplissement n'aurait tout simplement pas été possible sans le soutien indéfectible de mes parents, de mon frère et de toute ma famille, qu'ils soient dans le Sud, le Nord ou dans l'Ouest, toujours positifs et présents depuis bien avant cette étape. Je m'en vais vers l'Est, ainsi la boucle est bouclée ! Merci à vous, et *adishatz a tots* !

# Abstract

RNA-based regulations of gene expression allow quick and versatile responses from cells to changing environmental conditions. However, these regulations are often protein-mediated. The bacterial protein Hfq is one of the most studied RNA-based regulation partner. Found in various prokaryotes, it is an important virulence factor involved in many cellular processes. Hfq's structure resembles a torus, formed by multiple  $\beta$ -sheets. Apart from this N-terminal region (NTR), a supplemental C-terminal region (CTR) with variable lengths and sequences may exist in some species. In *E. coli*, this specific region measures around 30 residues and is predicted as intrinsically disordered. Few studies focused on Hfq-CTR until recently.

This work highlights new potential roles for Hfq-CTR. First, this region is able to self-interact and forms amyloid fibers, which explains the self-assembled Hfq superstructures observed *in vivo*. Second, the protein can bind and efficiently condense DNA *in vitro*, strengthening the suggested role of Hfq in shaping the bacterial chromosome. This compaction is fully dependent on the CTR which is responsible for DNA bridging. Third, the CTR also gives to Hfq the ability to self-assemble on a lipid bilayer, explaining its membrane-localized fraction observed *in vivo*. The subsequent membrane reorganization might facilitate the release of RNAs in the extracellular medium, with potential implications on bacterial communication and interaction with surrounding cells and environment.

**Note to non-French speakers:**

*As most of the results are published in English peer-reviewed journals, non-French readers can directly refer to the articles included in the appendix of this manuscript.*

# Résumé

La régulation de l'expression génique par des ARNs permet une réponse rapide et polyvalente des cellules à des changements environnementaux. Cependant, elle nécessite souvent des partenaires protéiques. La protéine bactérienne Hfq en est un bon exemple. Facteur de virulence, elle est présente chez une variété de procaryotes et intervient dans nombre de circuits de régulation. Structurellement, Hfq adopte un repliement caractéristique, le repliement Sm. Ainsi, Les feuilletts  $\beta$  qui la constituent se regroupent et forment un hexamère toroïdal. Outre cette région N-terminale, il existe aussi parfois une région C-terminale (CTR) de séquence et de longueur variables. Chez *E. coli*, cette région comprend une trentaine de résidus et est prédite comme non-structurée. Jusqu'à présent, son rôle n'a été que peu étudié.

Ce travail de thèse met en lumière de nouvelles pistes quant à la fonction du CTR. Nous avons constaté sa capacité à former des fibres amyloïdes, expliquant la formation de structures auto-assemblées *in vivo*. De plus, la protéine est capable de lier l'ADN et de le condenser fortement *in vitro*. Cette compaction est complètement dépendante de la présence du CTR, qui permet de ponter les brins d'ADN. Ce résultat suggère une nouvelle fonction de Hfq dans la structuration du chromosome. Enfin, nous avons démontré que ce domaine permet aussi à Hfq de s'assembler à la surface d'une bicouche lipidique, expliquant sa localisation membranaire. La désorganisation de la membrane qui en résulte pourrait permettre le passage d'ARNs dans le milieu extracellulaire, avec d'importantes implications sur la capacité de la bactérie à interagir avec ses voisines et son environnement.

# Liste des Illustrations

Figure 1 : Représentation du cytoplasme de <i>E. coli</i> au cours d'une simulation Brownienne, avec 1000 macromolécules dans un cube de 80 nm de côté.....	23 -
Figure 2 : Les composants cellulaires bactériens peuvent être localisés de façon hétérogène.....	25 -
Figure 3 : Comparaison de trois voies de répression d'un gène unique.....	27 -
Figure 4 : Régulation de l'expression de métabolites par les riboswitches .....	29 -
Figure 5 : Régulation de la transcription par l'ARN 6S.....	30 -
Figure 6 : Le mécanisme de défense bactérien CRISPR/Cas9 contre les infections phagiques.....	32 -
Figure 7 : Régulation négative de la traduction par l'action de petits ARN bactériens.....	33 -
Figure 8 : Schéma consensuel du procédé de dégradation des ARNm chez <i>E. coli</i> .....	34 -
Figure 9 : Carte schématique présentant les sous-domaines d'interaction du domaine C-terminal de l'endoribonucléase RNaseE, composant majeur du dégradosome chez <i>E. coli</i> .....	36 -
Figure 10 : Représentation schématique tridimensionnelle d'un modèle du dégradosome de <i>E. coli</i> interagissant avec la membrane via un tétramère de RNaseE.....	36 -
Figure 11 : Organisation cellulaire en forme d'hélice sous-membranaire de la RNaseE.....	37 -
Figure 12 : Structures filamenteuses formées par RhlB <i>in vitro</i> en présence de Ca <sup>2+</sup> et d'ATP.....	38 -
Figure 13 : Images des structures organisées de diverses protéines impliquées dans le métabolisme de l'ARN chez <i>E. coli</i> .....	39 -
Figure 14 : Localisation membranaire dynamique de la RNaseE chez <i>E. coli</i> vue en microscopie haute résolution (TIRFm).....	39 -
Figure 15 : Schéma résumant la variété et la présence des protéines Sm dans les trois règnes du vivant.....	41 -
Figure 16 : Sélection de structures cristallines obtenues pour divers assemblages de type Sm/Lsm.....	42 -
Figure 17 : Alignement multiple de séquences protéiques correspondant à la partie N-terminale (tore) de Hfq chez différentes espèces bactériennes.....	44 -
Figure 18 : Alignement multiple de séquences protéiques correspondant à la partie C-terminale (queue flexible) de Hfq chez différentes espèces bactériennes .....	45 -
Figure 19 : Représentation graphique de la conservation des séquences protéiques de Hfq chez <i>Escherichia coli</i> , <i>Salmonella enterica</i> , <i>Shigella sonnei</i> , <i>Cronobacter</i>	



<i>sakazakii</i> , <i>Yersinia pestis</i> , et <i>Serratia plymuthica</i> , espèces bactériennes membres du groupe des <i>Enterobacteriales</i> .....	46 -
Figure 20 : Structure des 72 premiers acides-aminés composant la partie N-terminale de la protéine Hfq chez <i>E. coli</i> , d'après l'entrée PDB 1HK9.....	46 -
Figure 21 : Taux de désordre et taux d'hydrophobie le long de la séquence primaire de Hfq .....	47 -
Figure 22 : Hfq intervient de différentes façons auprès des ARN bactériens .-	51 -
Figure 23 : Les sARN sont divisés en deux classes en fonction de leur type d'interaction avec Hfq.....	54 -
Figure 24 : Localisation de Hfq par cryo-TEM avec un étiquetage à la metallothionéine.....	57 -
Figure 25 : Auto-assemblage de Hfq.....	58 -
Figure 26 : Particularités de la composition en acides-aminés des IDPs .....	62 -
Figure 27 : Fractions relatives de domaines protéiques connus, domaines protéiques inconnus et domaines protéiques non-structurés dans 10 organismes modèles, déterminés par le système prédictif DICHOT .....	63 -
Figure 28 : Représentation schématique de la variété d'états conformationnels que peut adopter une chaîne polypeptidique après sa biosynthèse ainsi que les transitions possibles entre ces différents états .....	66 -
Figure 29 : Formation progressive de fibres amyloïdes.....	67 -
Figure 30 : Fibres amyloïdes : une morphologie commune basée sur une structure cross- $\beta$ , des arrangements variables.....	69 -
Figure 31 : Mécanisme de sécrétion et d'assemblage du curli amyloïde chez <i>E. coli</i> .....	71 -
Figure 32 : Une hypothèse du processus d'agrégation de la région C-terminale de Hfq observé <i>in vitro</i> .....	75 -
Figure 33 : Modèles d'auto-assemblage de Hfq entière.....	76 -
Figure 34 : Le chromosome bactérien est organisé en sous-domaines topologiques regroupant des <i>loci</i> spatialement proches et pouvant interagir.....	79 -
Figure 35 : Influence de la topologie du chromosome bactérien sur la transcription... et vice-versa .....	80 -
Figure 36 : Expression de différentes NAPs selon les phases de culture.....	81 -
Figure 37 : Dissociation facilitée de Fis du chromosome bactérien observée <i>in vitro</i> .....	82 -
Figure 38 : Quelques Protéines Associées au Nucléoïde parmi les plus étudiées et les plus abondantes dans <i>E. coli</i> .....	83 -

Figure 39 : Les protéines H-NS filamentent et pontent l'ADN en se regroupant à l'échelle du chromosome.....	87 -
Figure 40 : Le système MukBEF chez <i>E. coli</i> permet de rapprocher des régions chromosomiques éloignées .....	88 -
Figure 41 : Expériences de nanofluidique.....	90 -
Figure 42 : Modèle du mécanisme de compaction de l'ADN par Hfq.....	94 -
Figure 43 : Comparaison de la compaction d'un ADN de phage T4 contraint dans un nano-canal de 200 x 300 nm de section par Hfq, HU et H-NS.....	95 -
Figure 44 : Organisation de l'enveloppe bactérienne de <i>E. coli</i> .....	99 -
Figure 45 : Formule chimique développée et structure 3D des trois lipides les plus abondants dans la membrane interne de <i>E. coli</i> .....	102 -
Figure 46 : Localisation des cardiolipines dans la membrane interne de <i>E. coli</i> ...-	104 -
Figure 47 : Evolution de la représentation des membranes biologiques.....	104 -
Figure 48 : Diverses fonctions de Hfq dans lesquelles le domaine C-terminal est ou pourrait être impliqué.....	112 -

# Liste des Tableaux

Tableau 1 : Comparaison des concentrations seuils ( $\mu\text{M}$ ) de différentes macromolécules à partir desquelles la condensation de l'ADN est observée.....- 97 -

Tableau 2 : Composition de *E. coli* en phospholipides pour différents taux de croissance et conditions de culture.....- 102 -

# Liste des Abréviations

3'-UTR : 3'-*Untranslated Transcribed Region*, région 3' transcrite et non-traduite.

5'-UTR : 5'- *Untranslated Transcribed Region*, région 5' transcrite et non-traduite.

ADN : Acide DésoxyriboNucléique.

AFM : *Atomic Force Microscopy*, microscopie de force atomique.

ARN : Acide RiboNucléique.

ARN pol : ARN polymérase.

ARNm : Acide RiboNucléique messenger.

ARNnc : Acide RiboNucléique non-codant.

ARNr : Acide RiboNucléique ribosomique.

ARNt : Acide RiboNucléique de transfert.

CL : CardioLipine.

CRISPR : *Clustered Regularly Interspaced Short Palindromic Repeats*, courtes répétitions palindromiques groupées et régulièrement espacées.

CTR : *Carboxyl-Terminus Region*, région carboxyl-terminale.

EPE : *E. coli Polar Extract lipids*, lipides polaires extraits de *E. coli*.

FTIR : *Fourier Transformed InfraRed spectroscopy*, spectroscopie infrarouge à transformée de Fourier

IDP : *Intrinsically Disordered Protein*, protéine intrinsèquement désordonnée.

IM : *Inner Membrane*, membrane interne.

LPS : LipoPolySaccharide.

NAP : *Nucleoid Associated Protein*, protéine associée au nucléoïde.

NTR : *amiNo-Terminus Region*, région amino-terminale.

OM : *Outer Membrane*, membrane externe.

OMV : *Outer Membrane Vesicle*, vésicule membranaire externe.

PAM : *Protospacer Adjacent Motif*, séquence adjacente au proto-espaceur.

PCR : *Polymerase Chain Reaction*, réaction en chaîne par polymérase.

PDB : *Protein Data Bank*, banque de données des structures protéiques.

PE : Phosphatidyl-Éthanolamine.

PG : Phosphatidyl-Glycérol.

PS : Phosphatidyl-Sérine.

RBS : *Ribosome Binding Site*, site de fixation du ribosome.

RMN : Résonance Magnétique Nucléaire.

sARN : petit Acide RiboNucléique.

SRCD : *Synchrotron Radiation Circular Dichroism*, dichroïsme circulaire avec rayonnement synchrotron.

STORM : *STochastic Optical Reconstruction Microscopy*, microscopie optique par reconstruction stochastique.

TEM : *Transmission Electron Microscopy*, microscopie électronique par transmission.

TIRFm : *Total Internal Reflexion Fluorescence microscopy*, microscopie de fluorescence par réflexion totale interne.

# Sommaire



Remerciements.....	3 -
Abstract.....	6 -
Résumé.....	7 -
Liste des Illustrations.....	8 -
Liste des Tableaux.....	12 -
Liste des Abréviations.....	14 -
Sommaire.....	17 -
<b>Chapitre I : Introduction.....</b>	<b>21 -</b>
1    Préambule. Compartimentation cellulaire : une membrane ou rien ? .....	22 -
1.1    Une piscine ? En plein mois d'août, éventuellement.....	22 -
1.2    Un sac bien rangé.....	23 -
2    Contextualisation de la présente étude.....	26 -
2.1    Régulation transcriptionnelle, post-transcriptionnelle et traductionnelle par les ARN non-codants bactériens. ....	26 -
2.1.1    Un exemple de régulation « en cis » : les riboswitches. ....	28 -
2.1.2    Un exemple de régulation « en trans » : l'ARN 6S.....	29 -
2.1.3    Les ARN de type CRISPR.....	31 -
2.1.4    Les petits ARN (sARN) .....	32 -
2.2    Dégradation des ARN messagers chez <i>E. coli</i> .....	33 -
2.2.1    Voie de dégradation canonique.....	34 -
2.2.2    Assemblage des acteurs de la dégradation des ARNm en structures organisées.....	35 -
3    La protéine Hfq.....	41 -
3.1    Hfq : Phylogénie et structure. ....	41 -
3.1.1    La famille Sm/Lsm.....	41 -
3.1.2    Les protéines Hfq bactériennes.....	43 -
3.1.3    Hfq chez <i>E. coli</i> .....	46 -
3.2    Hfq, un régulateur pléiotropique. ....	47 -
3.2.1    Expression génique et abondance.....	48 -
3.2.2    Socle de l'homéostasie. ....	48 -
3.3    Hfq, chaperonne ARN riborégulatrice.....	49 -
3.3.1    Considérations fonctionnelles : modes d'actions de Hfq. ....	49 -

3.3.2	Considérations structurales.....	- 52 -
3.4	Article A : The <i>Escherichia coli</i> Hfq protein: an unattended DNA- transactions regulator.....	- 55 -
3.5	Interactions protéine-protéine.....	- 55 -
3.6	Localisation cellulaire.....	- 56 -
4	But de la présente étude.....	- 59 -
<b>Chapitre II : Agrégation amyloïde du domaine carboxyl terminal de la protéine Hfq.</b> .....		<b>- 60 -</b>
1	Contextualisation : les domaines non-structurés et les amyloïdes.....	- 61 -
1.1	Les domaines intrinsèquement désordonnés.....	- 61 -
1.2	Les structures de type cross- $\beta$ -amyloïde.....	- 64 -
1.2.1	Les amyloïdes sont une des structures possibles dans l'espace conformationnel accessible aux protéines.....	- 65 -
1.2.2	Mécanisme de formation.....	- 67 -
1.2.3	Caractéristiques structurales générales des fibres amyloïdes.....	- 68 -
1.2.4	L'émergence du concept d'amyloïde fonctionnel.....	- 70 -
2	Article B : New insight into the structure and function of Hfq carboxyl terminus.....	- 73 -
3	Article C : Techniques to analyse sRNA protein cofactor self-assembly <i>in vitro</i> . .....	- 74 -
4	Discussion.....	- 75 -
<b>Chapitre III : Interaction de la protéine Hfq avec l'ADN et contribution de son domaine carboxyl terminal.</b> .....		<b>- 77 -</b>
1	Contextualisation : compaction de l'ADN et organisation spatiale du génome chez <i>E. coli</i> .....	- 78 -
1.1	Les domaines topologiques.....	- 78 -
1.2	Les Protéines Associées au Nucléotide (NAPs) ou « protéines architectes ». - 80 -	
1.2.1	Protéines architectes : affinité, abondance.....	- 81 -
1.2.2	Quelques exemples de NAPs.....	- 83 -
1.2.3	Rôle dans la réparation de l'ADN.....	- 88 -
2	Article D : Effects of Hfq on the conformation and compaction of DNA..	- 90 -
3	Article E : Compaction and condensation of DNA mediated by the amyloidogenic C-terminal domain of Hfq.....	- 92 -

4	Discussion. ....	- 94 -
4.1	Mécanisme de la condensation par Hfq.....	- 94 -
4.2	Comparaison avec d'autres macromolécules. ....	- 95 -
<b>Chapitre IV : Agrégation de la protéine Hfq sur une membrane lipidique. ...</b>		<b>- 98 -</b>
1	Contextualisation : composition des enveloppes bactériennes. ....	- 99 -
1.1	La membrane externe.....	- 100 -
1.2	La paroi en peptidoglycane. ....	- 100 -
1.3	Le périplasme. ....	- 101 -
1.4	La membrane interne. ....	- 101 -
2	Article F : Membrane association of the bacterial riboregulator Hfq and functional perspectives.....	- 105 -
3	Discussion. ....	- 107 -
<b>Conclusion .....</b>		<b>- 109 -</b>
<b>Bibliographie.....</b>		<b>- 114 -</b>
<b>Annexes .....</b>		<b>- 129 -</b>
	Annexe A : .....	- 130 -
	Annexe B : .....	- 137 -
	Annexe C : .....	- 147 -
	Annexe D : .....	- 176 -
	Annexe E : .....	- 187 -
	Annexe F : .....	- 198 -

# Chapitre I :

# Introduction

# 1 Préambule. Compartimentation cellulaire : une membrane ou rien ?

Les bactéries sont parmi les organismes les plus présents sur Terre. Au cours de plusieurs centaines de millions d'années d'évolution, elles ont conquis presque chaque niche biologique disponible et développé une variété de voies métaboliques, de cycles cellulaires et de morphologies surprenantes. Avec une pression de sélection continue favorisant les taux de reproduction élevés, elles ont adopté des architectures élémentaires et des génomes compacts qui maximisent les portions codantes, par rapport aux organismes eucaryotes. Cette organisation simple a fréquemment été regardée et considérée comme le signe d'un stade cellulaire primitif. Historiquement, les caractéristiques plus complexes comme la compartimentation, les processus de transport intracellulaires, la présence d'un cytosquelette, la régulation spatiale de la transcription, la localisation des protéines étaient associées aux cellules eucaryotes. Pourtant, cette vision simpliste fut de plus en plus difficile à concilier avec la caractérisation croissante de nombreux processus complexes découverts chez les bactéries. Par exemple, la régulation de la forme des cellules et celle de la division cellulaire impliquent forcément une régulation spatio-temporelle d'une précision remarquable [1].

## 1.1 Une piscine ? En plein mois d'août, éventuellement.

Cette idée de complexité, devenue évidente, est encore renforcée lorsqu'on réalise l'énorme densité de molécules dans une bactérie, et tous les effets que cela implique sur la diffusion et la rencontre des différents partenaires réactionnels ou structuraux. La vision souvent erronée d'une piscine cytoplasmique remplie de quelques nageurs protéiques peut mener à des raisonnements s'éloignant fortement de la réalité biologique. Des travaux précurseurs, visant à combler un vide entre la structure d'une molécule et celle à l'échelle supra-cellulaire, comme ceux de Goodsell, ont permis de faire émerger le modèle déjà plus adéquat d'un « sac de macromolécules », à nuancer comme nous le verrons plus loin [2]. Des tentatives récentes s'appuyant sur des données structurales expérimentales, des coefficients diffusifs provenant de caractérisations *in vivo* et des données de protéomique poursuivent d'ailleurs ce travail de représentation d'un environnement intracellulaire plus proche de la réalité, dont un extrait est représenté sur la Figure 1 [3], [4].

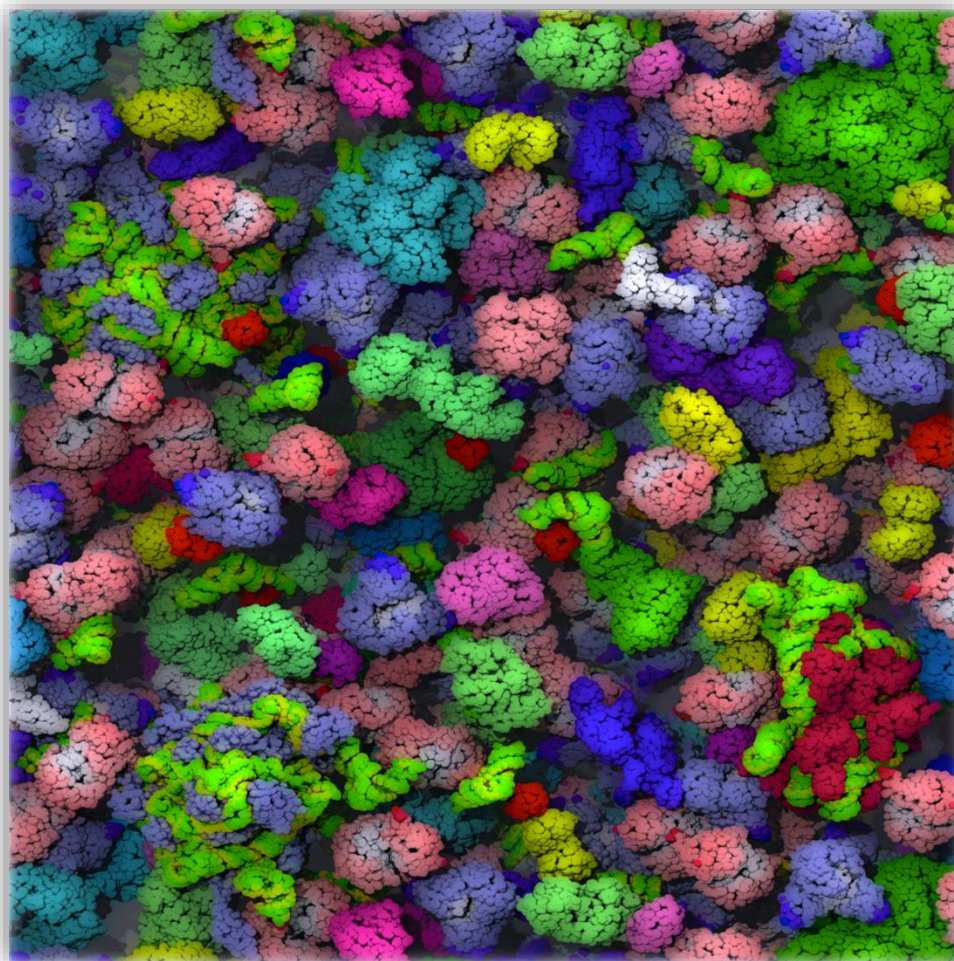


Figure 1 : Représentation du cytoplasme de *E. coli* au cours d'une simulation Brownienne, avec 1000 macromolécules dans un cube de 80 nm de côté. La concentration en matériel est de  $275 \text{ g.L}^{-1}$ , légèrement plus faible que les conditions physiologiques, qui sont de 300 à  $340 \text{ g.L}^{-1}$ . Les structures des différentes macromolécules représentées (protéines et ARN) sont issues de la PDB et sont parmi les plus abondantes selon les données expérimentales. L'ARN est représenté en jaune et en vert. L'ADN, les ARNm, les lipides et le glycérol sont omis.

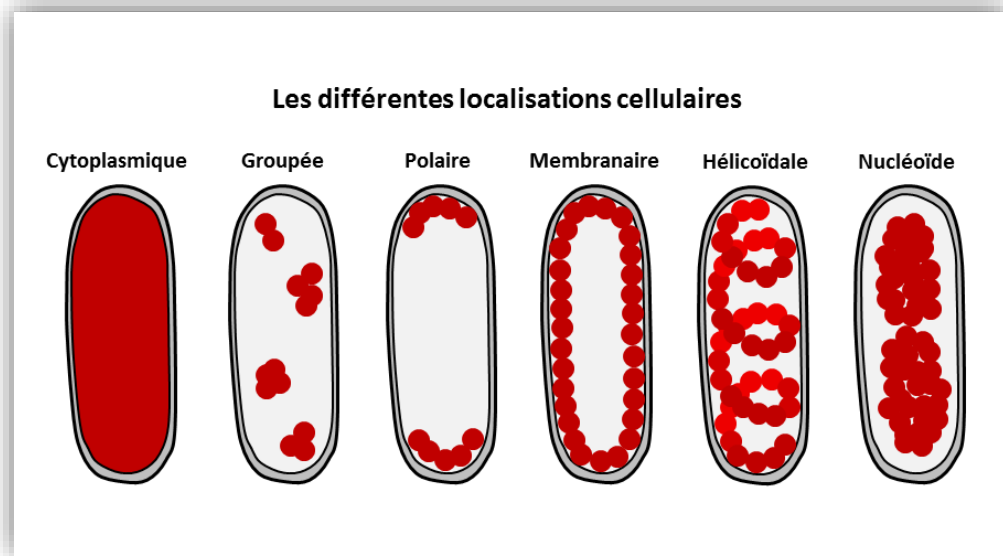
Reproduit d'après [4].

## 1.2 Un sac bien rangé.

Malgré tout, il existe quelques procaryotes faisant exception en étant sous-compartmentés par des membranes, comme *Gemmata obscuriglobus* ou bien *Streptomyces* sous certaines phases de culture [5], [6]. Mais même si les procaryotes sont généralement exempts d'organites et moins compartimentés que les cellules eucaryotes, le cytoplasme bactérien est loin d'être comparable à une soupe où les macromolécules sont mélangées aléatoirement. Beaucoup de protéines et d'acides nucléiques sont organisés spatialement, et temporellement. Malgré la forte densité en matériel biologique du cytoplasme et de la membrane cellulaire, la mobilité des molécules

est relativement haute tant qu'elles n'ont pas d'affinité significative pour les autres composants : par exemple, la diffusion d'une GFP (*Green Fluorescent Protein*) pôle à pôle chez *E. coli* est estimée à 0,5 seconde [7]. Mais même dans le cas de molécules diffusant librement, ce déplacement peut être limitant pour les processus physiologiques : des résultats de plus en plus nombreux accréditent le fait que l'encombrement est utilisé comme moyen de confinement en regroupant des protéines à l'emplacement où elles doivent accomplir leur fonction biologique, ou bien au contraire à l'opposé de cet emplacement afin de les empêcher d'agir. Cette ségrégation spatiale, plus versatile et dynamique qu'une séparation membranaire, s'accorde parfaitement avec un temps de division rapide, qui souvent fait partie du cycle cellulaire procaryote et notamment bactérien.

Parmi ces dispositions spatiales hétérogènes, on peut citer l'organisation du chromosome, le cytosquelette bactérien, la répartition des ribosomes, des lipides, et même la localisation des ARN [8]–[10]. Des assemblages ou superstructures formées d'un ou plusieurs acteurs principalement protéiques sont documentés dans de nombreuses espèces, avec des tailles allant de quelques dizaines de nanomètres jusqu'au micron [11]. On peut citer l'alignement du peptidoglycane chez *E. coli*, les magnétosomes de *Magnetospirillum*, les moteurs flagellaires, la ségrégation des copies plasmidiques par les filaments analogues de l'actine ParM, le positionnement complexe du septum de division par l'action du système MinCDE-FtsZ de *E. coli* ou le fonctionnement analogue chez *C. crescentus* [1], [12]. On peut classer ces différentes localisations cellulaires par catégorie comme sur la **Figure 2**, en distinguant les répartitions cytoplasmiques homogènes, hétérogènes regroupées (ex : ARN polymérase, H-NS, dans ces cas regroupées dans le nucléoïde), polaires (ex : cardiolipine, MinD), membranaires (ex : chaîne respiratoire), en grands ensembles hélicoïdaux ou au niveau du nucléoïde de manière diffuse (ex : HU) [13]–[17].



**Figure 2 : Les composants cellulaires bactériens peuvent être localisés de façon hétérogène. Il peut s'agir de protéines, d'acides nucléiques ou encore de lipides.**

Ces types de localisations ne sont que quelques exemples, et il existe des macromolécules pouvant en cumuler plusieurs à la fois ou bien passer de l'une à l'autre en fonction des phases de croissance et des conditions environnementales. Les techniques expérimentales utilisées pour déterminer la localisation d'une macromolécule peuvent aussi donner des résultats variables en produisant certains artefacts ou en étant plus ou moins précises : impact d'un ajout de protéine étiquette, limite de diffusion des anticorps en immuno-marquage, type de microscopie, protocoles d'extraction...



## 2 Contextualisation de la présente étude.

**L**a théorie fondamentale de la biologie moléculaire telle que formulée par F. Crick en 1958<sup>a</sup> a souvent été simplifiée en l’assertion suivante : l’ADN est transcrit en ARN, l’ARN est traduit en protéines, les protéines nous font tels que nous sommes («*DNA makes RNA, RNA makes proteins, proteins make us.*») [18]. La biosynthèse des protéines découle d’une séquence de processus cellulaires centraux : transcription de l’ADN en ARN messager, puis sa traduction en protéines. Ainsi, la régulation de la concentration en protéines peut avoir lieu à de nombreuses étapes et via de nombreux mécanismes. On peut citer dans l’ordre chronologique : nombre de copies de la séquence d’ADN codante et accessibilité spatiale à ces séquences par la machinerie transcriptionnelle, régulation de la transcription et de sa terminaison, temps de vie des transcrits, régulation de la traduction, taux de dégradation des protéines et/ou export extracellulaire. Cette partie vise à présenter quelques-uns de ces mécanismes centraux, notamment ceux impliquant l’ARN, qui seront en relation partielle ou directe avec la suite de l’étude portant sur la protéine bactérienne Hfq.

### 2.1 Régulation transcriptionnelle, post-transcriptionnelle et traductionnelle par les ARN non-codants bactériens.

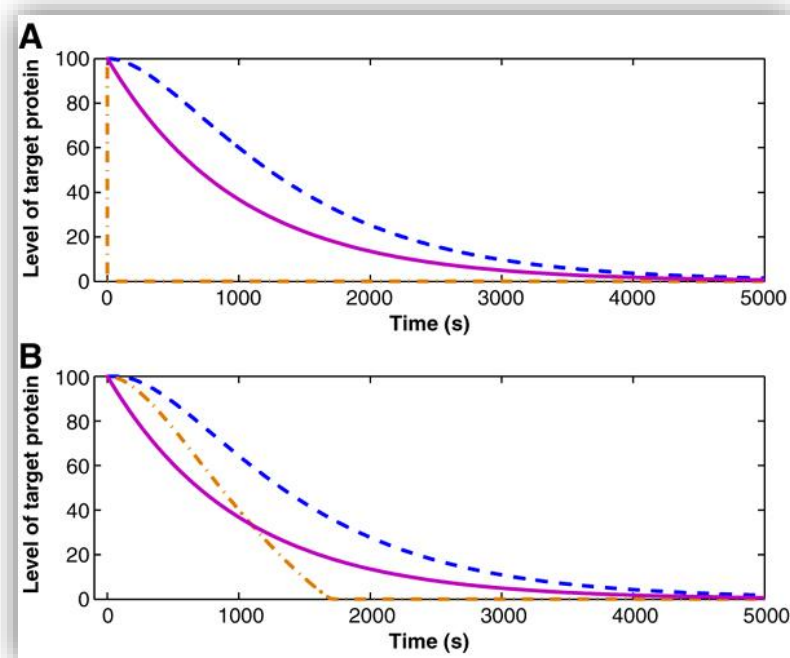
Le modèle d’une vie primitive basée sur l’ARN postule que dans les premières formes de vie, l’ensemble du stockage de l’information et l’ensemble des composants fonctionnels étaient composés par de l’ARN. Puis, l’émergence de l’ADN et des protéines auraient permis un stockage plus robuste et des fonctionnalités plus diverses et élaborées [19]. De fait, l’ARN a d’abord été relégué au stade d’intermédiaire, simple support du transfert de l’information depuis l’ADN (ARN<sub>m</sub>) et assistant de la synthèse des protéines (ARN<sub>t</sub> et ARN<sub>r</sub>). Mais depuis une trentaine d’années, il est devenu de plus en plus évident que les rôles de l’ARN dans le vivant sont bien plus vastes qu’une implication directe dans la synthèse protéique. Notamment, la découverte que 96% des ARN transcrits ne sont pas traduits en protéines a permis de lever le voile sur de nombreux mécanismes de régulation impliquant ces ARN. Génériquement appelés ARN non-codants (ARN<sub>nc</sub>), ils regroupent les ARN<sub>t</sub>, les ARN<sub>r</sub> (ainsi que leurs précurseurs) et les petits ARN régulateurs, ces derniers ne représentant que 1% du total. Ainsi, les premiers ARN régulateurs ont été découverts chez les eucaryotes dans les années 1990, ouvrant la voie à la myriade de systèmes que nous connaissons aujourd’hui. Certains ARN ayant des fonctions catalytiques ont aussi été caractérisés, formant des ribozymes en s’associant avec des

---

<sup>a</sup> “*The Central Dogma. This states that once ‘information’ has passed into protein it cannot get out again. In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible. Information means here the precise determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein.*”

séquences protéiques (par exemple, le brin d'ARN M1 de la RNaseP catalyse la maturation des ARNt [20]).

Parmi la palette de régulations bactériennes possibles contrôlant la synthèse et la concentration d'une protéine dans une cellule, la régulation par les ARN est sans doute celle qui a été étudiée le plus tardivement. Pourtant, il s'agit d'un pivot majeur de l'homéostasie bactérienne : comparée aux régulations transcriptionnelles et post-traductionnelles par dégradation, la régulation post-transcriptionnelle par les ARN est la plus rapide si les espèces régulatrices doivent être synthétisés (**Figure 3**) [21]. En première approximation, cela est justifiable par le fait que la transcription d'un petit ARN est bien plus rapide que la traduction d'une protéine régulatrice. L'aspect énergétique est aussi à considérer, puisque la synthèse et la dégradation d'un ARN régulateur coûtent moins d'énergie à la cellule que celles d'une protéine.



**Figure 3 : Comparaison de trois voies de répression d'un gène unique.** Evolution de la concentration en protéine codée (en nombre de molécules) en fonction du temps dans le cas d'une régulation transcriptionnelle par un répresseur (bleu), d'une régulation post-transcriptionnelle par un sARN (rose), et d'une régulation post-traductionnelle par une interaction protéine-protéine altérant la fonction de la cible (jaune). **A.** La protéine cible et son régulateur sont déjà présents dans la cellule à  $t=0$ . La régulation la plus rapide est celle effectuée via une interaction protéine régulatrice-protéine cible. **B.** La protéine cible est présente alors que le régulateur est produit en réponse à un stimulus extérieur ayant lieu à  $t=0$ . La régulation la plus rapide juste après le stimulus est alors la régulation par les petits ARN.

Les paramètres déterministes de la simulation (vitesses de synthèse et dégradation des différents acteurs) ont été choisis d'après des données expérimentales chez *E. coli* et le modèle donne des résultats identiques pour un panel de vitesses observées *in vivo*. Reproduit d'après

[21].

Ce moyen de régulation est probablement sous-estimé. En effet, la transcription permissive est très courante comme découvert récemment grâce à l'amélioration des techniques de cartographie transcriptionnelle, notamment l'immuno-précipitation de la chromatine suivie par un séquençage effectué sur l'ARN polymérase ou les facteurs  $\sigma$  (*Chromatine Immunoprecipitation Sequencing, ChIP-Seq*), le séquençage différentiel de l'ARN (*dRNA-seq*), ou encore le séquençage de l'ARN en cellule unique (*single-cell RNA-seq*). De très nombreuses régions d'initiation transcriptionnelles sont ainsi retrouvées à des endroits inattendus, notamment sur des portions non codantes du chromosome. Longtemps assimilés à du bruit transcriptionnel, les modèles décrivant ces événements comme de la « transcription envahissante » (*pervasive transcription*) ayant de vraies fonctions régulatrices sont de plus en plus supportés, rejetant l'idée qu'il s'agit de simples artefacts [22].

Dans cette partie, nous allons faire un rapide tour d'horizon de certains mécanismes régulateurs impliquant les ARN. On distingue des régulations agissant « en *cis* », où le régulateur ARN est sur la même molécule que sa cible, et des régulations « en *trans* », où le régulateur et sa cible sont deux molécules distinctes.

### 2.1.1 Un exemple de régulation « en *cis* » : les riboswitches.

Les riboswitches sont des séquences régulatrices d'un ARN messager capables de fixer un ligand et qui vont influencer directement ou indirectement la synthèse de la protéine correspondante. Ce type de régulateur, fréquent chez les bactéries, est présent dans les voies de synthèses de certains acides aminés, des bases azotées ou bien de vitamines et cofacteurs (revue des mécanismes dans [23]). La concentration intracellulaire d'un métabolite influence directement sa synthèse.

Les séquences se trouvent généralement dans la région 5' non traduite des ARNm. C'est la liaison (ou non-liaison) avec un ou des ligands qui détermine le type de régulation, positive ou négative. La région 5'-UTR adopte un repliement caractéristique en forme de poche à ligand (*ligand binding pocket*). Aussi appelé aptamère, ce repliement est évidemment sensible aux conditions environnementales (pH, force ionique, température), ce qui peut en faire un outil de réponse au stress (cas des thermosenseurs). Le principe général du couplage est basé sur la reconnaissance de surfaces complémentaires entre le ligand et l'aptamère. De façon analogue à une reconnaissance enzyme-substrat, c'est la formation de ponts hydrogène, d'interactions électrostatiques et d'interactions hydrophobes qui favorise le couplage. La fixation du ligand entraîne un changement conformationnel d'une zone adjacente de l'ARNm, appelée plateforme d'expression. Ce changement a pour conséquence la libération ou la séquestration de la zone de fixation du ribosome, et donc influence sur l'initiation de la traduction (**Figure 4A**). Plus rare, il existe aussi des riboswitches qui régulent la terminaison de la transcription en adoptant

une structure terminatrice Rho-indépendante lorsque le ligand est lié à l'aptamère (Figure 4B).

Le mode de fonctionnement de la régulation est propre à chaque riboswitch mais généralement on assiste à un mécanisme de rétro-inhibition en présence du ligand. Chez *E. coli*, c'est le cas par exemple de la voie de synthèse de la FMN (flavine mononucléotide) [24].

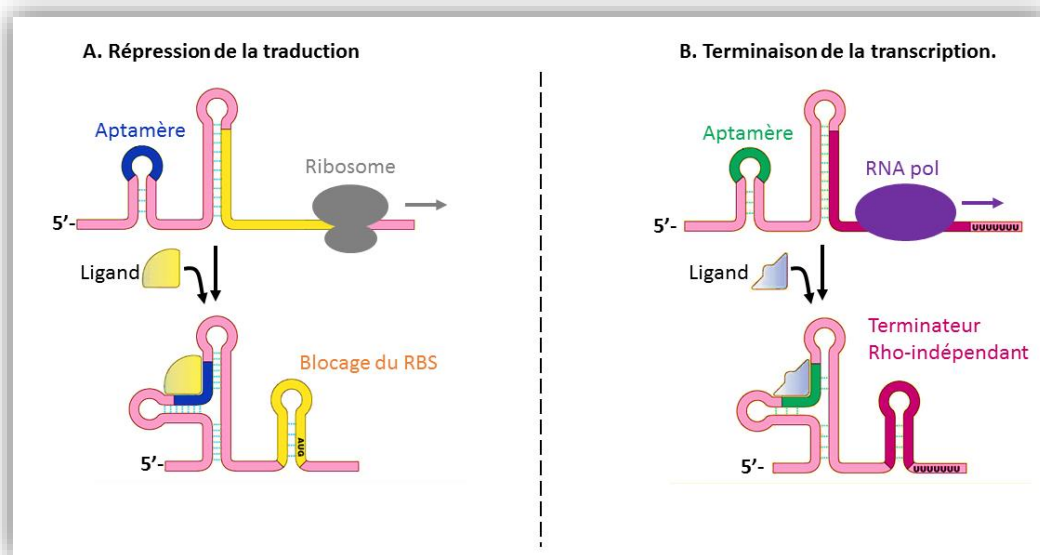


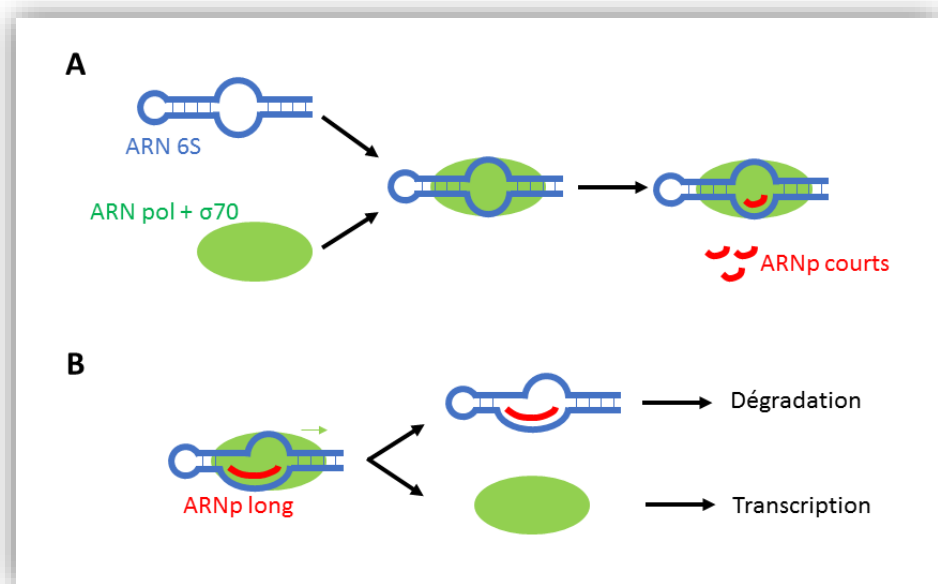
Figure 4 : Régulation de l'expression de métabolites par les riboswitches. A. Répression de la traduction par blocage du RBS à la suite de la fixation du ligand à l'aptamère ARN. B. Terminaison de la transcription par la formation d'un terminateur successive à la fixation du ligand.

Afin d'éviter des réponses géniques non désirées, les aptamères doivent être capables de différencier des composés affiliés avec un grand taux de discrimination. Par exemple le riboswitch purine est capable de différencier une adénine d'une guanine avec un facteur d'affinité de 20000 entre les deux [25]. De la même façon, le riboswitch de la lysine différencie la lysine de l'ornithine, deux acides aminés différents d'un groupe méthylène dans leur chaîne secondaire, avec un facteur de sensibilité de 5000 [26]. En théorie un aptamère est susceptible de différencier deux énantiomères, ce qui en fait des éléments régulateurs très sélectifs.

### 2.1.2 Un exemple de régulation « en *trans* » : l'ARN 6S.

L'ARN 6S est un ARNnc bactérien très abondant, parmi les premiers à avoir été identifié et séquencé [27]. Cependant, sa fonction n'a été déterminée que récemment : il permet chez de nombreuses espèces bactériennes de réguler la transcription d'un nombre conséquent de gènes. Chez *E. coli*, d'une longueur de 184 nucléotides, il se replie en épingle à cheveux avec une zone ouverte en son centre. Le mécanisme de régulation est

le suivant : l'ARN polymérase reconnaît la zone ouverte comme un promoteur ouvert de transcription, et l'ARN 6S se fixe au niveau des sous-unités  $\beta$ - $\beta'$  de l'ARN polymérase et du facteur de transcription « de ménage »  $\sigma_{70}$  [28]. Cela empêche le complexe ARN polymérase- $\sigma_{70}$  de procéder à la transcription normale des gènes contrôlés par ce facteur de transcription (**Figure 5A**). Ce dernier est exprimé en conditions normales de croissance. Lorsque la bactérie entre en phase stationnaire, elle produit un facteur de transcription particulier appelé RpoS ou  $\sigma_{38}$ . L'ARN 6S ne possède qu'une affinité très faible pour RpoS, et la polymérase contenant ce facteur peut alors librement reconnaître les promoteurs de transcriptions  $\sigma_{38}$  et exprimer les gènes concernés. Ainsi, l'ARN 6S réprime l'expression des gènes possédant un promoteur reconnu par  $\sigma_{70}$ , et sa régulation est basée sur une interaction directe protéine-ARN [29]. Comme son expression augmente en début de phase stationnaire, il semble avoir un rôle global dans la transition entre les phases exponentielle et stationnaire. Cependant, il pourrait aussi intervenir dans d'autres adaptations aux conditions environnementales, comme les variations de pH ou le stress oxydatif [30].



**Figure 5 : Régulation de la transcription par l'ARN 6S.** **A.** Lorsque la bactérie passe en phase de croissance stationnaire, l'ARN 6S est fortement exprimé et se lie à l'ARN polymérase complexée au facteur  $\sigma_{70}$ . La transcription de l'ADN est réprimée, et la polymérase devient ARN-dépendante, se servant de l'ARN 6S comme matrice. Cependant, la faible disponibilité en ribonucléotides ne permet que la synthèse de très courts ARN appelés ARNp. **B.** Lorsque les conditions de croissance redeviennent favorables, la concentration en ribonucléotides augmente et les ARNp synthétisés sont plus longs. Ils entraînent un changement conformationnel, relâchant la polymérase qui peut aller transcrire l'ADN normalement. Adapté d'après [28].

De façon surprenante, la fixation de l'ARN 6S à l'ARN polymérase en fait une polymérase ARN-dépendante. Elle procède ainsi à la synthèse de petits ARN complémentaires à l'ARN 6S appelés ARNp (« ARN produits »). Lorsque la

concentration en ribonucléotides est faible, les transcrits sont avortés. Lorsqu'elle augmente, c'est-à-dire que la disponibilité en nutriments augmente et que la bactérie sort de la phase stationnaire, les ARNp atteignent des longueurs plus importantes, de 12 à 14 nucléotides. Cela a pour conséquence de défaire le complexe 6S-ARNpol, et la polymérase est à nouveau disponible pour transcrire les gènes contrôlés par  $\sigma_{70}$  (**Figure 5B**). L'ARN 6S encode donc lui-même un régulateur *cis*-sARN (voir §2.1.4) de son activité répressive [31]. L'ensemble du mécanisme permet une régulation très rapide.

### 2.1.3 Les ARN de type CRISPR.

Le système CRISPR (pour *Clustered Regularly Interspaced Short Palindromic Repeats*) est présent dans la plupart des archées et environ 40% des bactéries connues [32]. Il s'agit d'un système de défense essentiel ciblant les ADN étrangers, différent de l'interférence par les ARNmi chez les eucaryotes.

Une cassette CRISPR consiste en des séquences identiques répétées de 23 à 47 paires de bases (séquences PAM pour *Protospacer Adjacent Motif*) alternées avec des séquences de taille identiques mais toutes différentes (*spacers*). Ces dernières sont en fait des portions d'ADN étrangers déjà rencontrés par la bactérie, comme des bactériophages ou des plasmides, ce qui en fait une mémoire immunitaire bactérienne.

Lors d'une nouvelle entrée d'un ADN étranger dans la cellule, le locus CRISPR est transcrit et un opéron spécial codant pour des protéines appelées Cas est exprimé. L'ARN CRISPR, appelé ARNcr, est mûri puis associé à une protéine Cas, formant un complexe ribonucléoprotéique. Enfin, l'interférence se fait par reconnaissance de l'ADN étranger, grâce à l'ARNcr qui lui est complémentaire. Une fois apparié, la protéine Cas coupe l'ADN, entraînant sa dégradation.

Les gènes *cas* varient considérablement selon les espèces, et on distingue 2 classes et de nombreux types de systèmes CRISPR/Cas, qui diffèrent notamment par la maturation des ARNcr ou le mécanisme de coupure de l'ADN étranger [33]. La **Figure 6** représente le système CRISPR/Cas9 de classe 2, type II mais c'est le système Cascade de classe 1, type I qui est présent chez *E. coli* [33], [34].

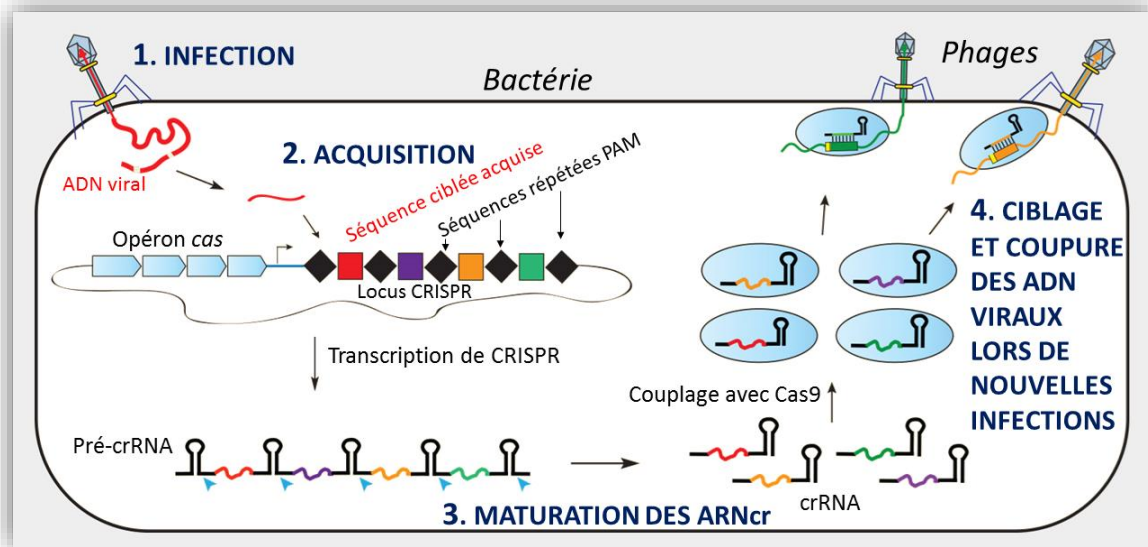


Figure 6 : Le mécanisme de défense bactérien CRISPR/Cas9 contre les infections phagiques. Lors de précédentes infections, la bactérie a stocké de petites portions d'ADN viral dans son génome, séparées par des séquences répétées PAM. La transcription de ces séquences et leur couplage au système protéique Cas permet le ciblage et la coupure des ADN viraux lors d'une nouvelle infection.

#### 2.1.4 Les petits ARN (sARN)

La régulation par les sARN (s pour *small*, ces ARN étant généralement courts) est un mécanisme central du contrôle de l'expression des gènes chez les bactéries. En effet, chez certaines espèces bactériennes, ce type de régulation touche au moins 40% des protéines [35]. La régulation prend des forces aussi variées que l'interférence ou l'atténuation transcriptionnelle, l'inhibition ou l'activation traductionnelle, ou encore le déclenchement de la dégradation des ARNm [36]. Cette régulation est de type « *trans* » puisque le petit ARN et son ARNm cible sont deux entités distinctes.

Néanmoins, à l'intérieur du groupe des sARN, on distingue deux types de petits ARN régulateurs, dont le mécanisme d'action canonique sur la traduction d'un ARNm est représenté sur la Figure 7 [37]. Les *cis*-sARN, aussi simplement appelés « ARN anti-sens », sont transcrits sur le brin d'ADN complémentaire au gène qu'ils régulent. Ainsi, leur séquence est généralement parfaitement appariée à leur ARNm cible, et ils en sont à proximité directe. Les *trans*-sARN, eux, sont transcrits à un *locus* différent et parfois très éloigné de leur cible. Ils régulent de nombreux gènes et leur séquence régulatrice n'est pas parfaitement complémentaire au transcrit ciblé. Ils nécessitent donc souvent un cofacteur protéique afin de les protéger de la dégradation des ribonucléases d'une part, et de faciliter leur appariement au messager d'autre part. Bien que de tailles et de structures très variables, on retrouve trois éléments communs chez les *trans*-sARN : une zone d'appariement avec leurs cibles d'environ 10 nucléotides, une zone variable



d'interaction avec la protéine chaperonne, et un terminateur de transcription Rho-indépendant (tige-boucle suivie d'une zone poly-U) [38]. La protéine la plus connue remplissant cette fonction est Hfq, objet de la présente thèse.

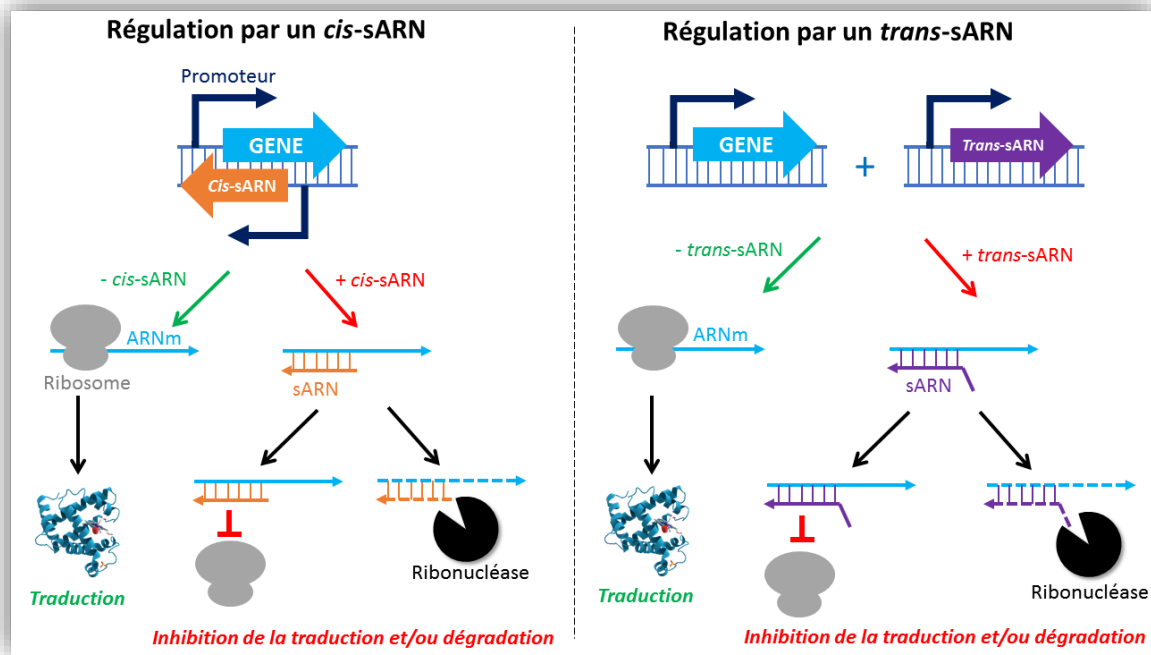


Figure 7 : Régulation négative de la traduction par l'action de petits ARN bactériens. On distingue les petits ARN en *cis*, transcrits directement depuis le brin complémentaire du gène qu'ils régulent, et les petits ARN en *trans*, transcrits ailleurs dans le génome. Les seconds ont une séquence moins complémentaire à leur cible et peuvent nécessiter une protéine facilitant leur appariement, Hfq. La régulation se fait dans la majorité des cas par blocage de la traduction et/ou dégradation du transcrit.

Une autre protéine remplissant un rôle analogue a été découverte dans des travaux récents, par *Grad-seq* chez *S. enterica* (gradient de glycérol des ARN puis séquençage à haut débit des fractions) [39]. Il s'agit de ProQ, retrouvée associée à une classe de sARN distincte de ceux de Hfq. Une protéine homologue dont la structure a tout juste été résolue est présente chez *E. Coli*, et le domaine d'interaction avec les ARN, appelé FinO, se retrouve chez plusieurs autres espèces bactériennes [40]. Ainsi, ProQ pourrait constituer une nouvelle branche de la régulation post-transcriptionnelle par les *trans*-sARN et ces résultats lèvent le voile sur un réseau de régulations plus complexe qu'initialement proposé [41].

## 2.2 Dégradation des ARN messagers chez *E. coli*.

Par définition, la dégradation des messagers influe leur turnover, et donc participe à la régulation de l'expression génique. Chez *E. coli*, le temps de vie chimique des ARN messagers (ARNm) est d'environ 7 à 8 minutes [42], même si celui-ci est dépendant de



l'ARN concerné et peut être modulé suivant les conditions de croissance et les situations de stress auxquelles la bactérie est soumise. La modulation est faite grâce aux différents partenaires qui interagissent avec les transcrits, partenaires protéiques ou nucléiques qui peuvent suivant le cas protéger ou exposer l'ARN aux enzymes de dégradation.

### 2.2.1 Voie de dégradation canonique.

Chez *E. coli*, le schéma de dégradation des ARN messagers est généralement le suivant (Figure 8).

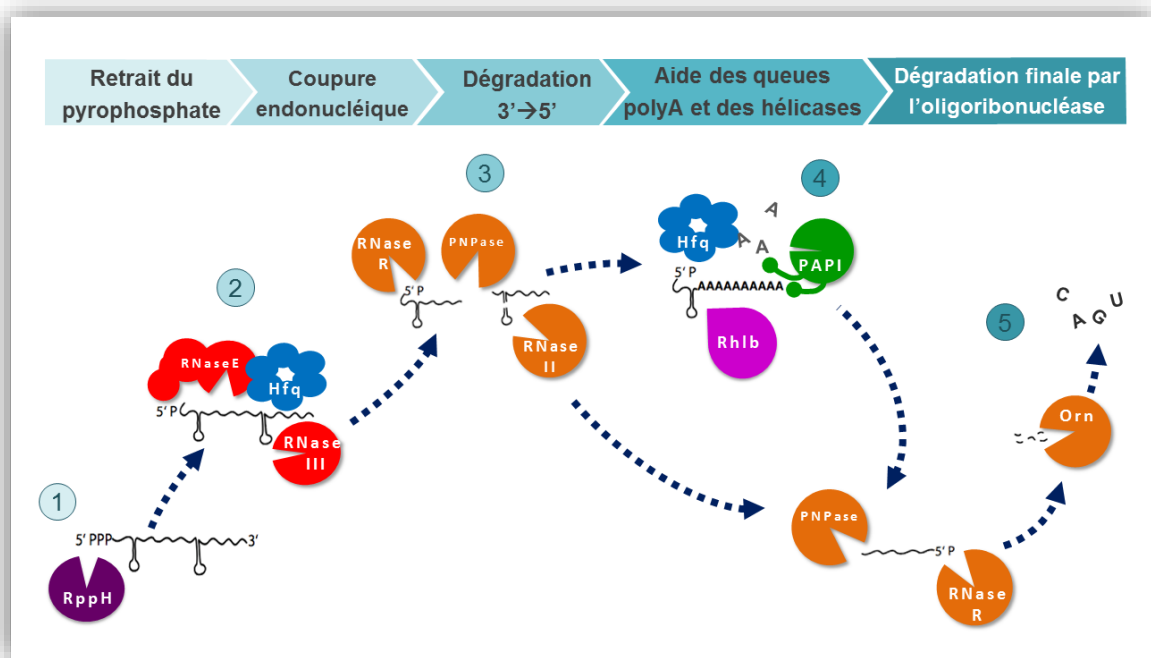


Figure 8 : Schéma consensuel du procédé de dégradation des ARNm chez *E. coli*.

Tout d'abord, une endoribonucléase essentielle, la RNaseE, vient couper l'ARN, ce qui le destine à être détruit. Bien que la RNaseE interagisse avec les ARN possédant une extrémité 5' triphosphate, elle coupe plus facilement ceux qui sont 5' monophosphate [43]. Il existe deux voies favorisant la formation d'une extrémité 5' monophosphate : (i) soit une pyrophosphohydrolase, appelée RppH, vient enlever les groupements triphosphates en 5' [44] (ii) soit la RNaseE génère elle-même une nouvelle extrémité 5' monophosphate lors d'une coupure interne initiale. Il faut noter qu'une autre endonucléase, la RNaseIII, peut aussi jouer un rôle similaire à celui de la RNaseE. Les spécificités des deux endonucléases sont cependant différentes puisque la RNaseE coupe plutôt du simple brin, alors que la RNaseIII coupe du double brin (donc dans une tige-boucle interne). Dans tous les cas, les fragments générés peuvent être à nouveau coupés par les endonucléases ou bien être dégradés par des exoribonucléases (RNaseII, PNPase) dans la direction 3'→5'. Une enzyme telle que la PolyA polymérase PAPI peut aussi favoriser cette dégradation exonucléolytique en ajoutant une queue d'adénines (polyA)

en 3' ce qui permet de déstabiliser des structures en tige boucle [45] ; de même qu'une hélicase comme RhlB peut détruire des structures secondaires afin de favoriser l'action des exonucléases. Enfin, les oligoribonucléases finissent de dégrader les fragments des transcrits en nucléotides.

Certaines des enzymes listées sur la **Figure 8** ont aussi d'autres rôles dans le métabolisme bactérien. Par exemple, la RNaseE intervient dans la maturation de certains ARN en coupant des précurseurs d'ARN ayant des structures secondaires, notamment lors de la maturation du ribosome, expliquant son caractère essentiel [46].

## 2.2.2 Assemblage des acteurs de la dégradation des ARNm en structures organisées.

### → Le dégradosome bactérien.

Dans beaucoup d'espèces bactériennes, la RNaseE (ou son homologue RNaseY chez *B. subtilis*) n'est pas isolée, et on observe la formation d'un complexe appelé dégradosome, qui rassemble différents acteurs enzymatiques reliés à la dégradation des ARN. Les composants de ce complexe multi-enzymatique coopèrent pour maturer ou détruire les ARN [47].

Chez *E. coli*, le cœur de ce complexe est ainsi organisé autour de la RNaseE et plus particulièrement le long de son domaine C-terminal (**Figure 9**). Cette endoribonucléase de taille importante (1061 acides-aminés) sert de support aux autres composants du dégradosome. On y retrouve une hélicase de type DEAD-box, RhlB (*RNA Helicase B*), une exoribonucléase phosphate-dépendante, la PNPase (PolyNucléotide Phosphorylase), ainsi que l'é nolase. Le rôle précis de cette dernière dans ce complexe reste encore mystérieux, même si on la pense impliquée dans la dégradation d'ARN spécifiques comme celui codant pour le transporteur du glucose [48]. Ce type d'association est par ailleurs retrouvé dans des dégradosomes d'autres espèces et présente des similarités avec des enzymes chez les archées ou les eucaryotes [49]. La présence de l'hélicase RhlB est d'autant plus importante que les autres enzymes ne sont actives que sur des ARN simple brin, rendant obligatoire le dépliement des structures secondaires [50]. Dans certains cas, il semblerait que RhlB puisse être remplacée par une autre hélicase, CsdA [51]. Deux hélicases supplémentaires, SrmB et RhlE, sont aussi capables de se fixer à la RNaseE et de remplacer RhlB lors de tests de dégradations d'ARN *in vitro* [52].

Le domaine C-terminal de la RNaseE semble dépourvu de structure tri-dimensionnelle définie. Cependant, l'association de l'enzyme avec les autres composants du dégradosome à travers les différents domaines d'interaction pourrait faire émerger des microdomaines structurés [53].

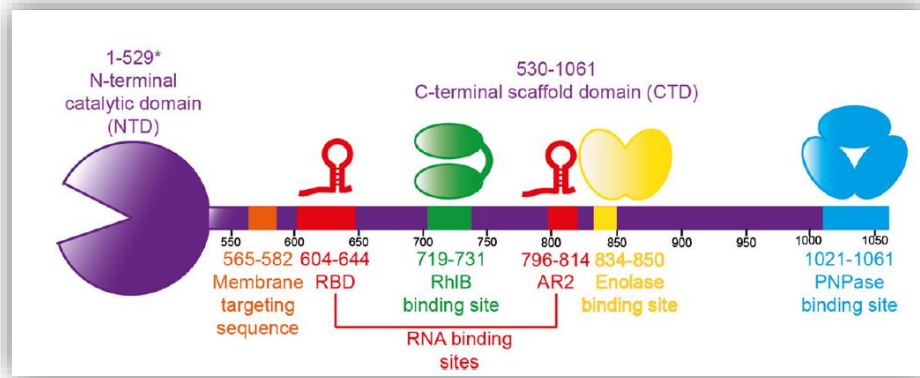


Figure 9 : Carte schématique présentant les sous-domaines d'interaction du domaine C-terminal de l'endoribonucléase RNaseE, composant majeur du dégradosome chez *E. coli*. La région catalytique N-terminale ne représente que la première moitié de la protéine. Tiré de [54].

Une autre zone clé de la RNaseE est située aux alentours du 570<sup>ème</sup> acide aminé : il s'agit d'une hélice amphipatique à l'origine de la fixation de l'enzyme à la membrane interne de la bactérie [55]. De telles hélices permettant une fixation membranaire existent chez d'autres protéines bactériennes comme MreB (homologue de l'actine), FtsA et MinD (impliquées dans la division cellulaire) [56],[57]. La face hydrophobe de l'hélice s'enfonce dans la couche lipidique de la membrane, laissant la face hydrophile et chargée interagir avec l'environnement cytoplasmique de la cellule et les têtes polaires des phospholipides.

L'ensemble du dégradosome peut être représenté suivant le modèle tridimensionnel de la Figure 10, qui met en lumière un tétramère de RNaseE interagissant avec ses partenaires et la membrane interne de *E. coli*. Les premières observations en microscopie électronique suggèrent une taille d'environ 30 nm [54].

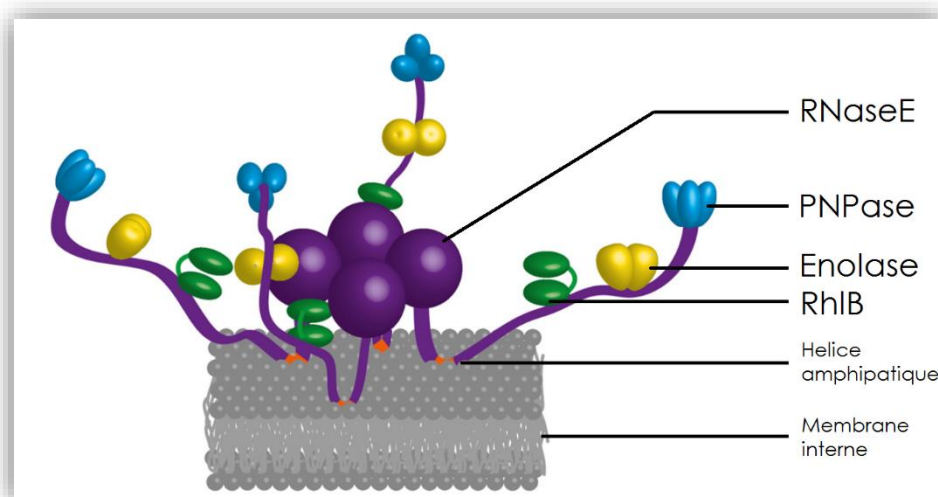
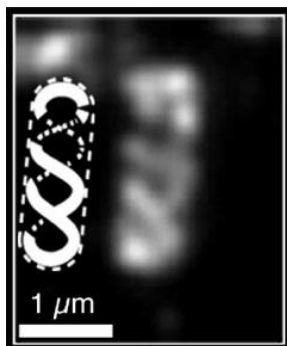


Figure 10 : Représentation schématique tridimensionnelle d'un modèle du dégradosome de *E. coli* interagissant avec la membrane via un tétramère de RNaseE. Adapté de [54].

→ Les structures sous-membranaires hélicoïdales.

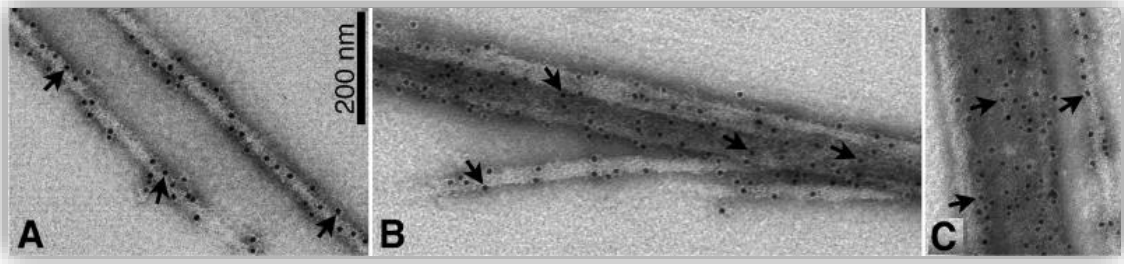
En parallèle des observations précédentes, d'autres travaux ont révélé une organisation structurée des différents acteurs de la dégradation des ARN à plus grande échelle. Il a ainsi été rapporté que la RNaseE et les autres composants du dégradosome sont organisés sous la forme de structures hélicoïdales qui s'enroulent le long des bactéries, structures qui ont été observées par immunofluorescence (**Figure 11**) [58]. Ces structures ressemblent à celles formées par les protéines du cytosquelette bactérien, composé notamment de MreB (homologues de l'actine), Ftsz (homologue de la tubuline) et MinD [9]. Cependant, elles sont indépendantes de ce cytosquelette car elles se maintiennent en son absence.



**Figure 11 : Organisation cellulaire en forme d'hélice sous-membranaire de la RNaseE.**  
Microscopie par immunofluorescence de RNaseE étiquetées avec de l'hémagglutinine (RNaseE:HA) en utilisant des anticorps primaires monoclonaux anti-HA. Tiré de [59].

Il a été démontré que la RNaseE et l'hélicase RhlB sont deux protéines capables de former des structures hélicoïdales même en l'absence des autres constituants du dégradosome. En cela elles diffèrent de l'énolase et de la PNPase qui se servent de la RNaseE comme support. La PNPase garde tout de même la capacité à former ces structures en l'absence de RNaseE, mais uniquement en présence de l'hélicase avec laquelle elle interagit également. Cela n'est pas le cas de l'énolase, dont l'organisation est totalement dépendante de la RNaseE pour former les structures [58]. Des interactions RNaseE-RNaseE et RhlB-RhlB ont aussi été rapportées, pouvant expliquer la formation de polymères à grande échelle [59].

Ces données sont à rapprocher d'autres travaux qui ont démontré la capacité de RhlB à s'assembler *in vitro* en long filaments en présence d'ions calcium  $\text{Ca}^{2+}$  et d'ATP, celui-ci n'étant cependant pas hydrolysé lors de l'assemblage [60]. Le nombre total de copies de RhlB *in vivo* a été estimé autour de 1350 par cellule. Il permettrait de former des structures de 8 à 10  $\mu\text{m}$  de long, cohérentes avec celles observées *in vivo* enroulées le long des bactéries (pour une taille de l'hélicase d'environ 70 Å d'après les données de cristallisation). Ces fibres observées possédaient trois types de structures : des fibres uniques et séparées, des faisceaux fusionnés, et de larges plaques (**Figure 12**).



**Figure 12 : Structures filamenteuses formées par RhIB *in vitro* en présence de  $\text{Ca}^{2+}$  et d'ATP.** Microscopie électronique en transmission avec coloration négative à l'acétate d'uranyle et marquage par anticorps fonctionnalisés avec bille d'or (flèches noires) dirigé contre RhIB. **A.** Filaments distincts. **B.** Filaments en faisceaux. **C.** Large feuillet (gauche) et filament (droite). D'après [60].

Plus récemment, il a été observé par immunofluorescence qu'en plus des enzymes qui composent le dégradosome, plusieurs autres protéines associées au métabolisme des ARN chez *E. coli* forment le même type de structures hélicoïdales sous-membranaires. Des expériences ont permis l'obtention d'images, que l'on peut voir **Figure 13**, soit grâce à des anticorps dirigés contre les protéines, soit grâce à des anticorps dirigés contre de courtes étiquettes FLAG (DYKDDDDK) ou HA (YPYDVPDYA). En effet, les protéines de fusion avec une étiquette fluorescente de type GFP (*Green Fluorescent Protein*) occasionnent assez souvent la perte de la localisation ou de la fonction des protéines.

Formant le même type de structures, en plus des composants du dégradosome canonique de *E. coli* vu précédemment on trouve : la PolyA Polymérase (PAP), l'oligoribonucléase Orn, les ribonucléases RNaseII et RNaseIII, la chaperone Hfq et enfin deux régulateurs de l'activité de la RNaseE, RraA et RraB. Toutes ces protéines montrent le même type de structure qui semble se déployer sous la membrane interne, contrairement aux images obtenues pour des protéines cytoplasmiques ou membranaires intrinsèques qui montrent respectivement une fluorescence diffuse dans le cytoplasme ou bien localisée à la périphérie de la cellule [61].

Il a été avancé que les différentes protéines nécessitaient la présence d'un ARN matrice pour former les complexes observés, mais cette hypothèse a été invalidée par l'observation des mêmes structures dans des bactéries traitées à la rifampicine, qui inhibe la transcription. De même, ces assemblages ne sont pas dépendants de la position du nucléoïde. Elles ne sont pas non plus reliées au cytosquelette bactérien, car elles sont présentes même si celui-ci est affecté génétiquement [61].

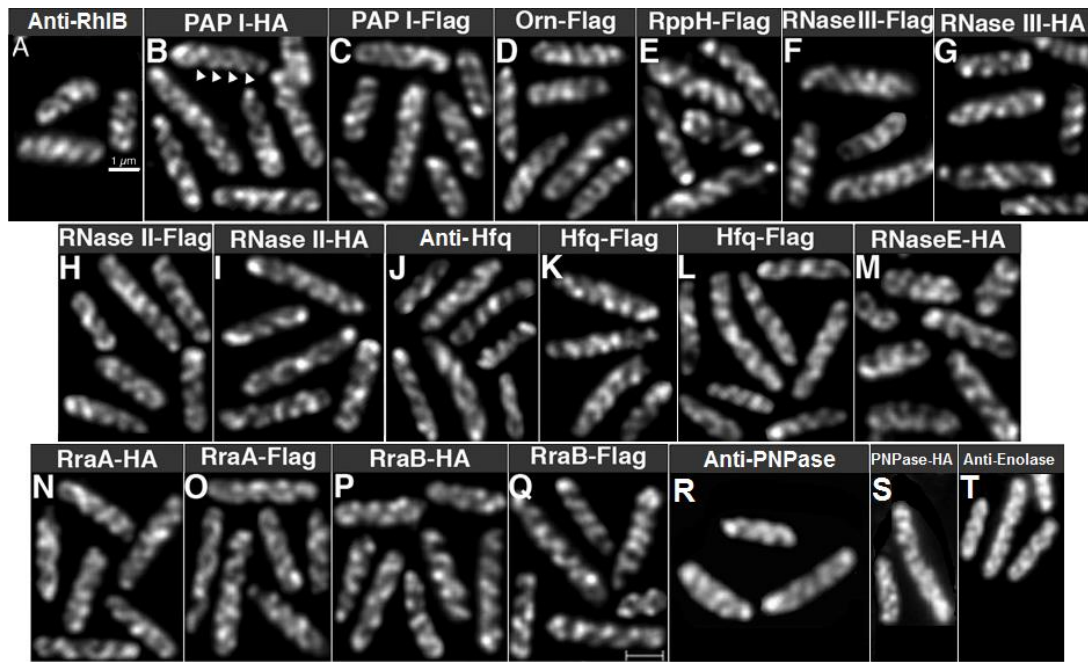


Figure 13 : Images des structures organisées de diverses protéines impliquées dans le métabolisme de l'ARN chez *E. coli*. Microscopie en immunofluorescence ; les anticorps primaires sont soit dirigés contre les protéines, soit contre de courtes étiquettes FLAG ou HA. Adapté de [61].

Une autre étude en TIRFm (*Total Internal Reflexion Fluorescence microscopy*) de la localisation de la RNaseE dans des bactéries vivantes montre une localisation membranaire très dynamique de la RNaseE, organisée en *loci* qui se déplacent, mais pas de structure hélicoïdale (Figure 14) [62]. Ce résultat n'est cependant pas contradictoire avec le précédent, puisque les travaux en immunofluorescence étaient à la limite résolutive de la microscopie conventionnelle. De plus, la fusion de la RNaseE avec des protéines fluorescentes peut déstabiliser ou inhiber leur auto-assemblage, comme c'est le cas pour la chaperonne Hfq ou de nombreuses autres protéines.

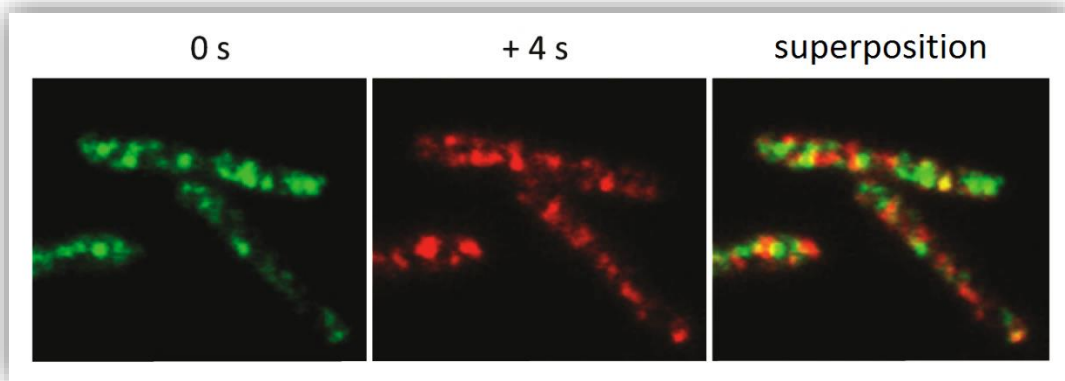


Figure 14 : Localisation membranaire dynamique de la RNaseE chez *E. coli* vue en microscopie haute résolution (TIRFm). Les bactéries vivantes immobilisées sur de l'agarose expriment une version de la RNaseE étiquetée avec une YFP (*Yellow Fluorescent Protein*). Les images sont

colorisées artificiellement pour mettre en relief les différences de localisations à  $t = 0s$  et  $t = 4s$ .  
Tiré de [62].

Les réseaux d'interactions entre ces différentes protéines, le mode d'assemblage, l'organisation spatiale ainsi que le rôle fonctionnel de ces structures sont encore inconnus pour la plupart et restent à déterminer. Une fonction possible consisterait en une séparation spatiale de la synthèse et de la dégradation des ARNm, ce qui pourrait être comparé à la compartimentation observée chez les eucaryotes. Elle permettrait d'éviter la destruction d'un ARNm juste après sa transcription. La co-localisation des différentes protéines impliquées dans le métabolisme de l'ARN dans les mêmes structures pourrait aussi avoir un avantage en permettant de rapprocher des enzymes qui agissent simultanément ou séquentiellement. Leur action concertée serait facilitée par cette proximité spatiale entre les différents acteurs, évitant à l'ARN de devoir diffuser aléatoirement dans le cytoplasme bactérien d'un enzyme à l'autre.

Ceci est à rapprocher des distributions hétérogènes des ARNm dans la cellule, observées notamment par microscopie de haute résolution : les ARNm codant pour des protéines de la membrane interne sont principalement retrouvés proches de la membrane, contrairement aux ARNm codant pour des protéines cytoplasmiques ou périsplasmiques. Cette localisation est due à une traduction et une insertion membranaire simultanées de la protéine naissante, alors que ces ARN ont un temps de vie plus court que la normale. Il apparaît que cette stabilité plus faible est justement la conséquence de la localisation membranaire du dégradosome bactérien, puisque le temps de vie de ces transcrits est rallongé lorsque l'ancrage membranaire de la RNaseE est supprimé [63].



### 3 La protéine Hfq.

**H**fq a été identifiée dans les années 1970 comme étant un facteur indispensable à la réplication d'un virus infectant les bactéries : le bactériophage Q $\beta$  [64]. Celui-ci a donné son nom à Hfq, pour Host factor Q $\beta$ . Ce n'est que bien plus tard que son rôle plus large dans le métabolisme des ARN a été mis à jour. Nous allons d'abord faire un rapide tour d'horizon des homologues de cette protéine avant de s'intéresser plus précisément à la version de *E. coli*, sujet de la présente thèse.

#### 3.1 Hfq : Phylogénie et structure.

##### 3.1.1 La famille Sm/Lsm.

Il apparaît que Hfq, par la relative homologie de sa séquence protéique et de sa structure, fait pleinement partie de la grande famille des protéines Sm/Lsm (pour *Sm-like*) [65]. Cette famille de protéines remarquablement conservée au niveau phylogénétique dans tous les règnes du vivant regroupe cependant des fonctions très diverses. On y retrouve les protéines Sm des archées (SmAPs) ainsi que les protéines Sm et Lsm des eucaryotes qui sont impliquées dans des fonctions telles l'épissage, le « décoiffage » ou la dégradation des ARN messagers, mais aussi la maintenance des télomères [66]–[68] (Figure 15).

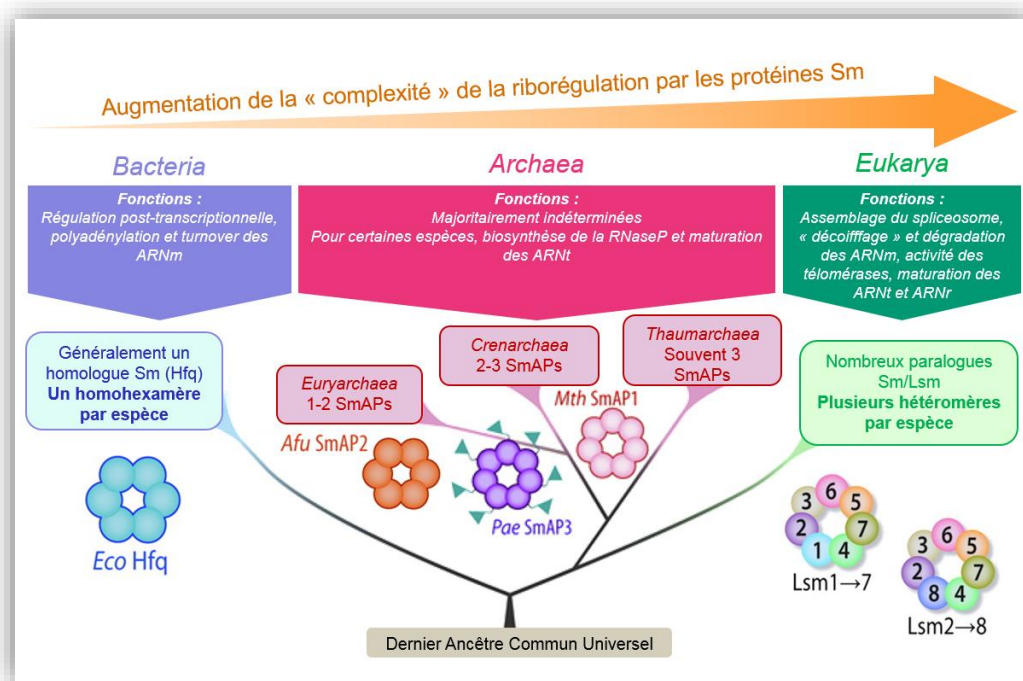


Figure 15 : Schéma résumant la variété et la présence des protéines Sm dans les trois règnes du vivant. Reproduit et adapté d'après [69].



Si la structure en anneau est sans conteste le caractère distinctif de cette famille, il faut noter que les protéines archées et eucaryotes sont principalement des homo/hétéro-heptamères alors que les versions bactériennes sont systématiquement homo-hexamériques [70] (Figure 16). La structure secondaire d'un monomère est composée de cinq feuillets  $\beta$ , précédés par une hélice  $\alpha$ . Les trois premiers feuillets  $\beta$  forment le motif Sm1 dont la séquence est fortement conservée, alors que le motif Sm2 qui englobe les deux feuillets suivants diverge chez les bactéries. Ce dernier est responsable de l'assemblage des différents monomères entre eux par des liaisons polaires [71]. Ainsi, les hélices se retrouvent exposées sur la même face de l'anneau, appelée « surface proximale ». Par opposition, la seconde face des multi-mères est appelée « surface distale ».

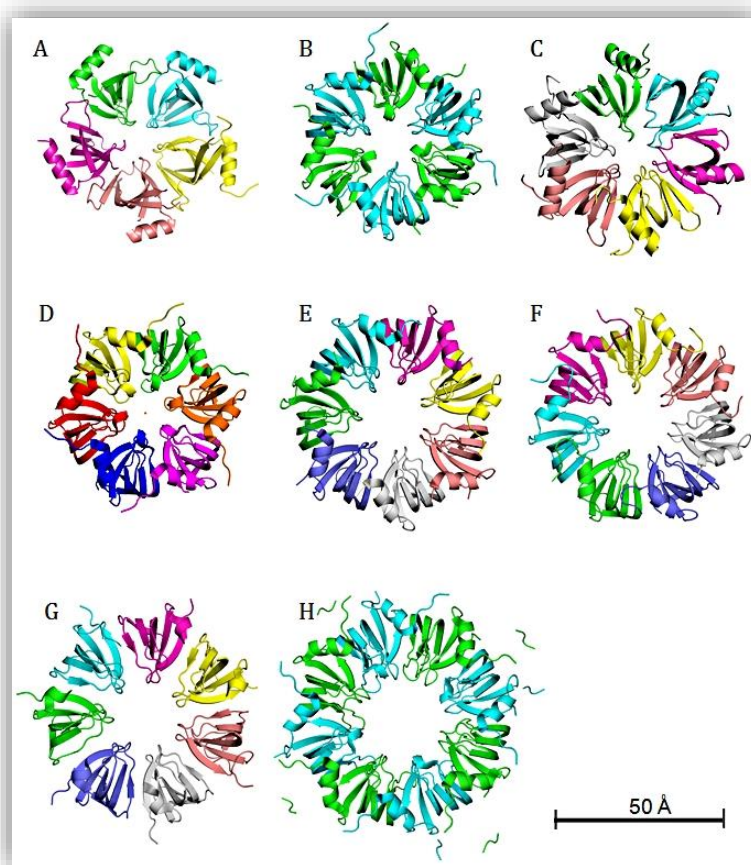


Figure 16 : Sélection de structures cristallines obtenues pour divers assemblages de type Sm/Lsm. A. Pentamère du cyanophage ECX21941 (PDB 3BY7). B. Hexamère de Lsm5 chez *C. parvum* (PDB 3PGG). C. Hexamère de Hfq chez *S. aureus* (PDB 1KQ1). D. Hexamère de Sm2 chez *A. fulgidus* Sm2 (PDB 1LJO). E. Heptamère de Sm1 chez *A. fulgidus* (PDB 1I4K). G. Heptamère de Lsm chez *M. thermoautotrophicum* (PDB 1I81). F. Heptamère de Sm-F chez *S. cerevisiae* (PDB 1N9R). H. Octamère de Lsm3 chez *S. cerevisiae* (PDB 3BW1). Reproduit d'après [70].

Les protéines Sm constituent très probablement l'une des plus anciennes super-famille de protéines [72]. Elles pourraient permettre de mieux comprendre la transition depuis les formes de vie primordiales, dont le modèle actuel pose l'ARN à la fois comme support de l'information et comme agent fonctionnel, aux formes de vie contemporaines, ayant développé l'ADN pour le support de l'information et les protéines pour l'accomplissement des fonctions, catalytiques notamment.

La diversité des oligomères Sm est certainement apparue par une série de duplications et de divergences des différentes copies, un mécanisme permettant l'apparition de paralogues différents capables de s'oligomériser. Cela expliquerait le développement des hétéromères chez les eucaryotes depuis les homo-oligomères bactériens et eucaryotes. Cependant, le transfert génétique horizontal pourrait aussi avoir joué un rôle dans la multiplication des différentes versions [73]. Alors qu'au premier abord le passage de l'homo-oligomérisation à l'hétéro-oligomérisation pourrait passer pour une dépense énergétique superflue, il faut noter qu'il permet de complexifier les fonctions de la protéine ainsi que sa régulation, qui peut s'effectuer de façon plus fine *via* les différentes étapes pour parvenir à la formation du complexe quaternaire.

Ainsi, en passant des bactéries aux archées puis aux eucaryotes, on note une augmentation du nombre des protéines homologues de la famille Sm : alors qu'il n'en existe généralement qu'une copie chez les bactéries (même s'il a récemment été trouvé une seconde version chez certaines espèces, nommée *hfq2*), on passe à 2 ou 3 chez les archées, toujours des homo-oligomères. Chez les eucaryotes, le nombre augmente encore pour atteindre 15 chez *Homo sapiens* par exemple.

### 3.1.2 Les protéines Hfq bactériennes.

Homo-hexamériques et longues d'environ 70-100 acides-aminés par monomère, les protéines Sm bactériennes possèdent une séquence très bien conservée en ce qui concerne la partie torique, c'est-à-dire le domaine N-terminal (**Figure 17**). Celui-ci a pu être cristallisé sous sa forme hexamérique, par exemple chez *S. aureus*, *E. coli*, *P. aeruginosa* et possède bien le repliement caractéristique des protéines Sm [74]–[76]. Par opposition, le domaine C-terminal, appelé CTR dans la suite du présent manuscrit, est lui d'une longueur très variable et bien moins conservé, presque totalement absent chez *Bacillus* (**Figure 18**). Il existe néanmoins quelques exemples extraordinairement longs telle la protéine putative de *Sphingomonas assacharolytica* (160 acides-aminés) et celle de *Acinetobacter baylyi* (172 acides-aminés). Cette dernière version se distingue par un CTR très riche en glycines, et il a été montré que celui-ci est bien exprimé et fait partie intégrante de la protéine fonctionnelle *in vivo* [77]. A côté de ces exemples extrêmes, l'ordre des *Enterobacteriales* dont fait partie *E. coli* (espèces représentatives en rouge **Figure 17** et **Figure 18**) a conservé un CTR relativement long, d'une trentaine d'acides-aminés et riche en résidus polaires (**Figure 19**).

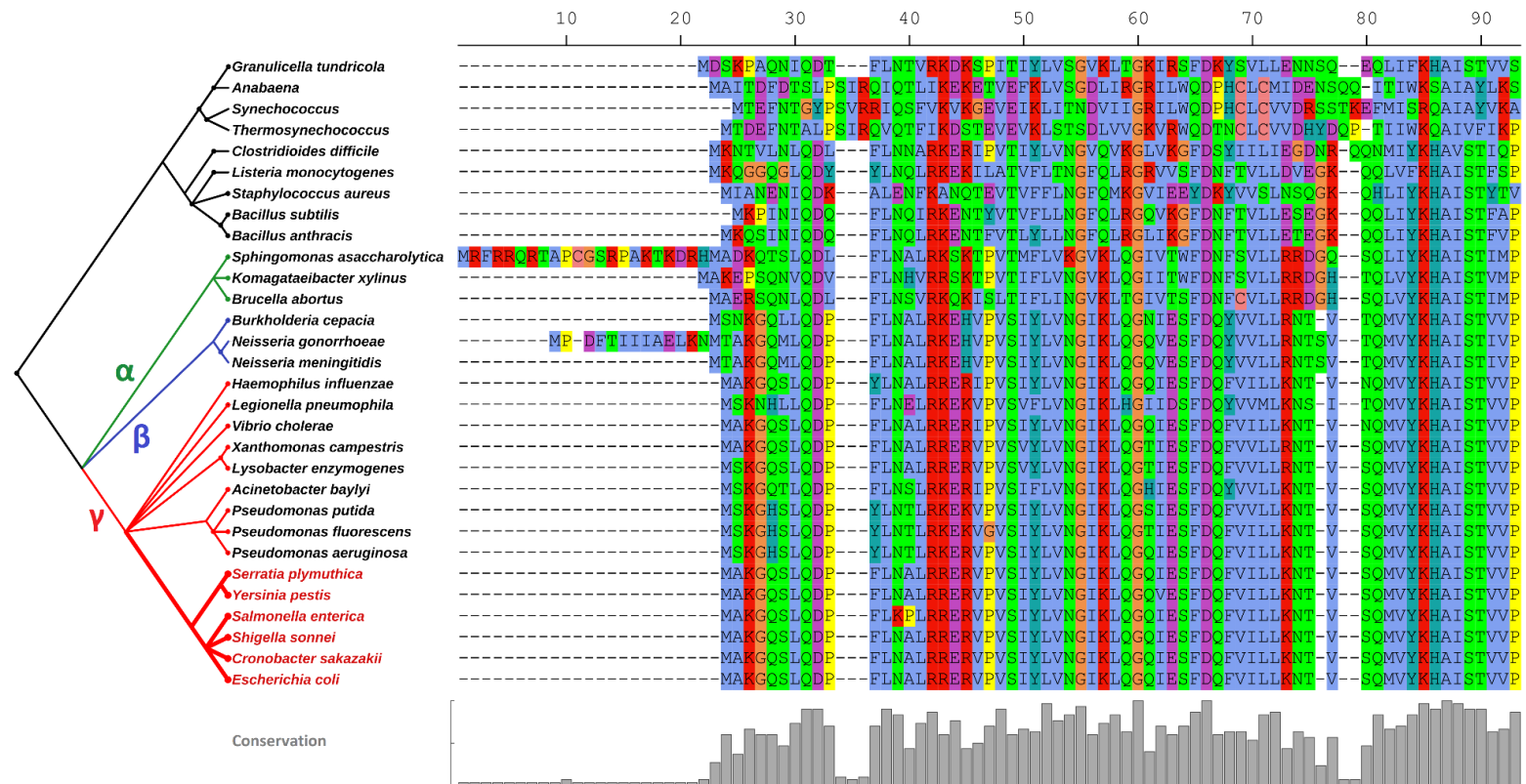


Figure 17 : Alignement multiple de séquences protéiques correspondant à la partie N-terminale (tore) de Hfq chez différentes espèces bactériennes. La partie gauche présente l'arbre phylogénétique des différents taxons comparés, construit grâce à [78]. Branches vertes : classe des *Alphaproteobacteria*. Branches bleues : classe des *Betaproteobacteria*. Branches rouges : classe des *Gammaproteobacteria*. Espèces en rouge : groupe des *Enterobacteriales*. Droite : alignement des séquences de Hfq correspondantes obtenues sur UniProt, effectué avec Clustal Omega. Code couleur des acides aminés : hydrophobes en bleu pastel, chargés positivement en rouge, chargés négativement en violet, polaires en vert, cycliques en bleu turquoise, glycines en orange, prolines en jaune, cystéines en rose.

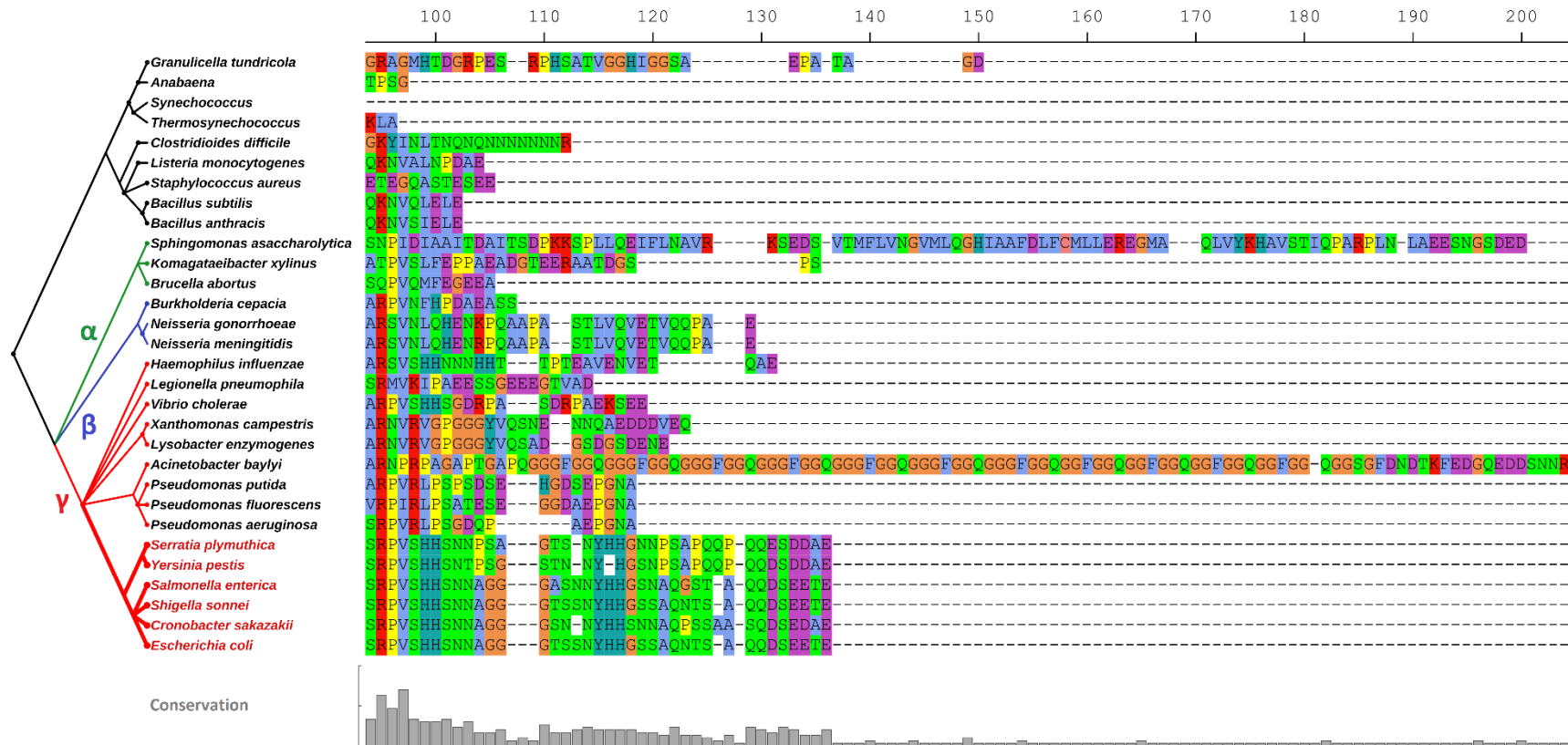


Figure 18 : Alignement multiple de séquences protéiques correspondant à la partie C-terminale (queue flexible) de Hfq chez différentes espèces bactériennes. La partie gauche présente l'arbre phylogénétique des différents taxons comparés, construit grâce à [78]. Branches vertes : classe des *Alphaproteobacteria*. Branches bleues : classe des *Betaproteobacteria*. Branches rouges : classe des *Gammaproteobacteria*. Espèces en rouge : groupe des *Enterobacteriales*. Droite : alignement des séquences de Hfq correspondantes obtenues sur UniProt, effectué avec Clustal Omega. Code couleur des acides aminés : hydrophobes en bleu pastel, chargés positivement en rouge, chargés négativement en violet, polaires en vert, cycliques en bleu turquoise, glycines en orange, prolines en jaune, cystéines en rose.

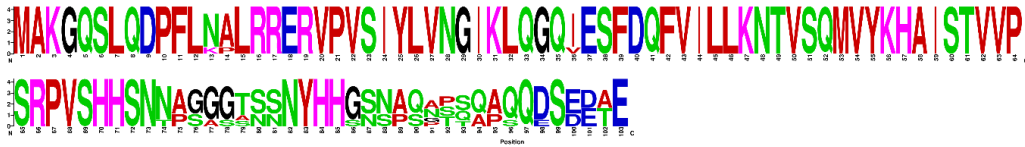


Figure 19 : Représentation graphique de la conservation des séquences protéiques de Hfq chez *Escherichia coli*, *Salmonella enterica*, *Shigella sonnei*, *Cronobacter sakazakii*, *Yersinia pestis*, et *Serratia plymuthica*, espèces bactériennes membres du groupe des *Enterobacteriales*. La hauteur relative de chaque lettre représente sa fréquence à la position spécifiée, et la hauteur totale des lettres à une position représente son degré de conservation mesuré en bits d'information dans les séquences alignées. Code couleur des acides-aminés : hydrophobes en rouge, chargés positivement en rose, chargés négativement en bleu, polaires en vert et glycines en noir. Schéma construit via [79].

### 3.1.3 Hfq chez *E. coli*.

Chez *E. coli*, Hfq est donc une protéine homo-hexamérique de 66 kDa. Chaque monomère adopte le repliement caractéristique des protéines Sm, c'est-à-dire l'enchaînement d'une hélice  $\alpha$  puis de cinq feuillets  $\beta$  (Figure 20). Le tout est suivi d'une queue C-terminale d'une trentaine d'acides-aminés. Même si Hfq existe sous forme monomérique *in vivo*, cette forme est fortement minoritaire et la protéine n'est active que sous forme hexamérique [80].

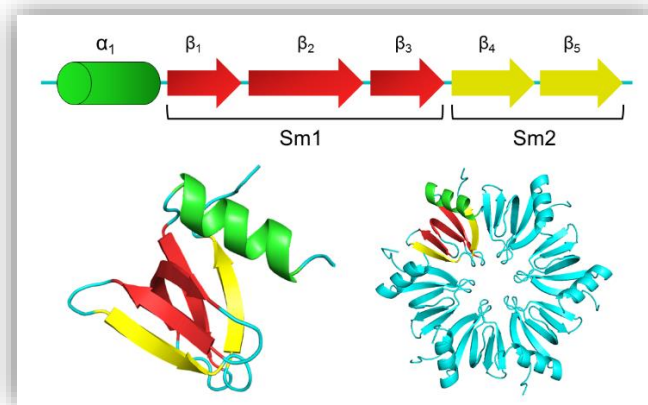


Figure 20 : Structure des 72 premiers acides-aminés composant la partie N-terminale de la protéine Hfq chez *E. coli*, d'après l'entrée PDB 1HK9. Haut : séquence linéaire d'un monomère montrant l'enchaînement de l'hélice et des 5 feuillets caractéristiques du repliement de type Sm. Bas-gauche : structure 3D correspondante. Bas-droite : hexamère complet.

La séquence C-terminale, bien conservée chez les *Enterobacteriales*, n'a pourtant pas eu de fonction précise attribuée pendant longtemps. Cependant, il a été démontré que la stabilité du complexe hexamérique est diminuée lorsque les queues CTR sont absentes [81].

Du point de vue structural, le CTR est principalement polaire, flexible et prédit comme désordonné par la majorité des outils d'analyse (Figure 21). Cette prédiction est confirmée par des données de spectroscopies en solution, qui permettent en plus de démontrer que les queues CTR ont tendance à s'étendre loin du tore et ont la faculté d'interagir avec les ARN de grande taille lorsqu'ils dépassent de l'hexamère [82]. Les données de cristallographie obtenues sur la protéine entière après de nombreux essais infructueux dus à la flexibilité et la polarité des queues CTR les montrent proches du tore, même si la dernière portion du domaine est projetée vers le solvant, ce qui n'est guère surprenant [83].

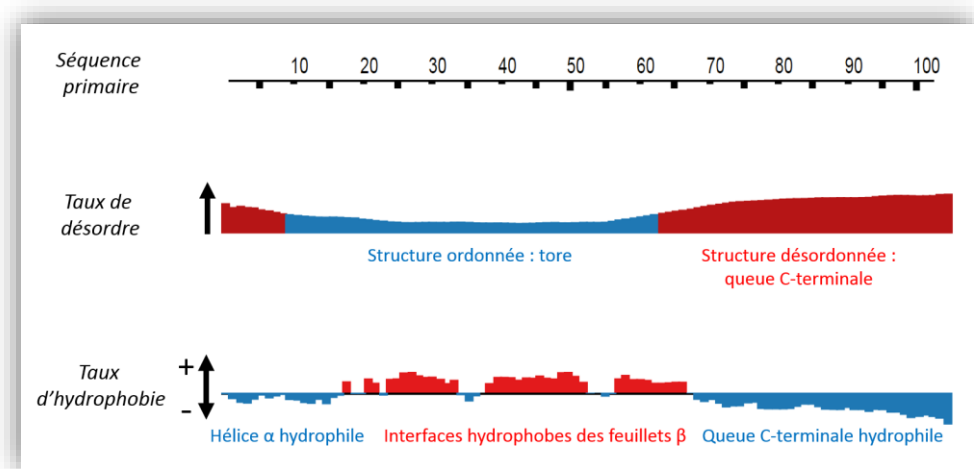


Figure 21 : Taux de désordre et taux d'hydrophobie le long de la séquence primaire de Hfq. Données collectées sur UniProt (P0A6X3). Le désordre est calculé selon [84] et l'hydrophobie provient de la somme des score d'hydrophobie des acides aminés selon [85] sur une fenêtre glissante de 15 résidus.

### 3.2 Hfq, un régulateur pléiotropique.



### 3.2.1 Expression génique et abondance.

Le gène codant pour Hfq possède plusieurs voies de régulation, transcriptionnelles et traductionnelles. Il fait partie de l'opéron *amiB-mutL-miaA-hfq-hflX-hflK-hflC* qui comporte plusieurs promoteurs. Certains sont reconnus par la sous-unité  $\sigma_{70}$  qui est le facteur de transcription primaire chez *E. coli*, mais *hfq* comporte aussi un promoteur reconnu par le facteur  $\sigma_{32}$ , exprimé en cas de choc thermique. Cela démontre l'importance de Hfq en cas de stress, la protéine étant essentielle pour la croissance et la survie des bactéries à plus de 45°C [86]. La production de Hfq est aussi contrôlée par une boucle de rétro-régulation négative : la protéine se lie à son propre ARNm et limite sa traduction en séquestrant le RBS [87]. De plus, la protéine CsrA réprime aussi l'expression de *hfq* par un mécanisme semblable [88]. Malgré tout, Hfq est fortement synthétisée lors de l'ensemble des phases de culture, d'une manière comparable à celle des ribosomes. Avec un nombre de copies par cellule entre 6000 et 10000 hexamères, la concentration intracellulaire maximum atteinte est d'environ 15  $\mu\text{M}$ . Les mesures d'abondance ne font pas encore consensus et varient selon les souches et conditions de culture, même si l'augmentation de la production en phase stationnaire est communément admise [89], [90].

### 3.2.2 Socle de l'homéostasie.

Hfq est au centre d'une myriade de régulations, en faisant un des acteurs incontournables de l'homéostasie bactérienne [91]. En effet, les mutants *hfq* de *E. coli* présentent de nombreux phénotypes particuliers, dont les suivants : taux de croissance et rendements en biomasse réduits, osmosensitivité accrue, allongement des cellules, hypersensibilité aux rayonnements ultraviolets, augmentation des taux d'oxydation des sources carbonées, diminution du surenroulement négatif des plasmides en phase stationnaire, modification de la composition de la membrane... Une étude de transcriptomique chez *S. enterica* par pyroséquençage à haut débit montre par exemple que Hfq est impliquée dans l'expression de 20% des gènes [92]. Chez *E. coli*, il a été montré que les bactéries mutées, généralement plus sensibles aux stress et aux changements d'environnement, étaient moins virulentes : la délétion de Hfq entraîne une diminution de la capacité des bactéries à coloniser et survivre dans l'hôte infecté [93].

Une partie de ces phénotypes relatifs aux stress sont liés à la régulation de *rpoS* par Hfq, gène codant pour le facteur  $\sigma_{38}$  qui est le régulateur central de la phase stationnaire et plus généralement de la réponse aux stress [86], [94]. En l'absence de la protéine, la réponse de la bactérie est donc déficiente.

### 3.3 Hfq, chaperonne ARN riborégulatrice.

#### 3.3.1 Considérations fonctionnelles : modes d'actions de Hfq.

Hfq est connue pour intervenir à différents stades dans le processus de régulation post-transcriptionnelle, principalement à travers le *chaperonnage* de l'hybridation imparfaite entre trans-sARN et ARNm. Cette non-complémentarité entre sARN et ARNm signifie que beaucoup des sARN ne s'apparient pas à leur ARNm cible en l'absence de Hfq ; ils sont dits « Hfq-dépendants ». Comme expliqué précédemment, les sARN sont très hétérogènes en tailles et en structures secondaires, et sont impliqués dans divers processus incluant le *quorum sensing*, la réponse à divers stress, la virulence... Ces interactions Hfq-sRNA-ARNm peuvent avoir toute une variété d'effets, dont certains exemples sont illustrés dans la **Figure 22** [95].

→ Hfq peut inhiber la traduction d'un ARNm en permettant la fixation d'un sRNA en 5' du transcrit, cachant le RBS et empêchant la fixation du ribosome (A) : c'est le cas des couples sARN-ARNm *MicA/ompA* (protéine de la membrane externe, [96]), *SgrS/ptsG* (transporteur du glucose-phosphate, [97]), *RprA/csgD* (formation du biofilm, [98]) ou encore *RyhB/sodB* (disponibilité en fer, [99]).

→ Elle peut aussi, à l'opposé, permettre de libérer un RBS bloqué par une structure secondaire et déclencher la traduction de l'ARNm, même si cela est moins fréquent (B) : c'est le cas de *DsrA/rpoS* (choc thermique froid, [100]), *RprA/rpoS* (choc osmotique et formation du biofilm, [101]).

→ A un autre niveau, Hfq permet de stabiliser des sARN en les fixant, évitant l'action des ribonucléases qui les dégraderaient avant qu'ils n'atteignent leur cible (C) [102].



→ Elle intervient aussi dans la dégradation des ARNm en permettant la fixation de sARN sur un ARNm cible qui déclenchera l'action de la RNaseE (D). Cette dégradation est dite « passive » car elle est simplement due à l'arrêt de la traduction (les ribosomes cachent les sites de clivage par la RNaseE lorsque l'ARNm est traduit) [103]. Une dégradation « active » peut aussi arriver lorsque Hfq, interagissant avec la RNaseE, fait directement intervenir l'enzyme sur le transcrit.

→ Hfq peut aussi se fixer directement sur l'ARNm et recruter la PolyA Polymérase, entraînant la dégradation du messager par les exonucléases 3'→5'(E) [104].

→ L'hybridation entre un sARN et son ARNm pendant sa transcription permet enfin d'empêcher la terminaison prématurée Rho-dépendante en évitant la fixation de Rho : c'est le cas des trois sARN DsrA, ArcZ, et RprA [105] (F). Ici il s'agit donc de régulation transcriptionnelle plus que traductionnelle.

Hfq est donc capable d'interagir avec de très nombreux ARN. Mais combien ? Une récente étude chez *S. Typhimurium* permet d'avoir une idée du *targetome* de Hfq [106]. La recherche des partenaires ARN de Hfq a été effectuée par CLIP-Seq (liaison protéine-ARN par UV et immuno-précipitation, puis séquençage haut débit) et a permis de trouver de nouveaux candidats sARN. Les résultats annotés donnent 553 ARNm et 126 sARN interagissant avec Hfq. Les zones d'interaction étaient majoritairement : pour les ARNm, les zones 5'-UTR ou 3'-UTR avec une préférence pour les terminateurs de transcriptions riches en uracile ; et pour les sARN leur zone 3' pour la même raison. Cela est en accord avec les résultats expérimentaux précédents et avec l'hypothèse d'un enchaînement où les ARN viennent sur Hfq dans un premier temps, avant de s'apparier dans un second temps, ce qui nécessite que l'interaction avec Hfq rapproche les zones d'interaction ARN-ARN et facilite la formation du duplex. Une étude du même ordre a aussi été conduite chez *E. coli*, parvenant aux mêmes conclusions [107].

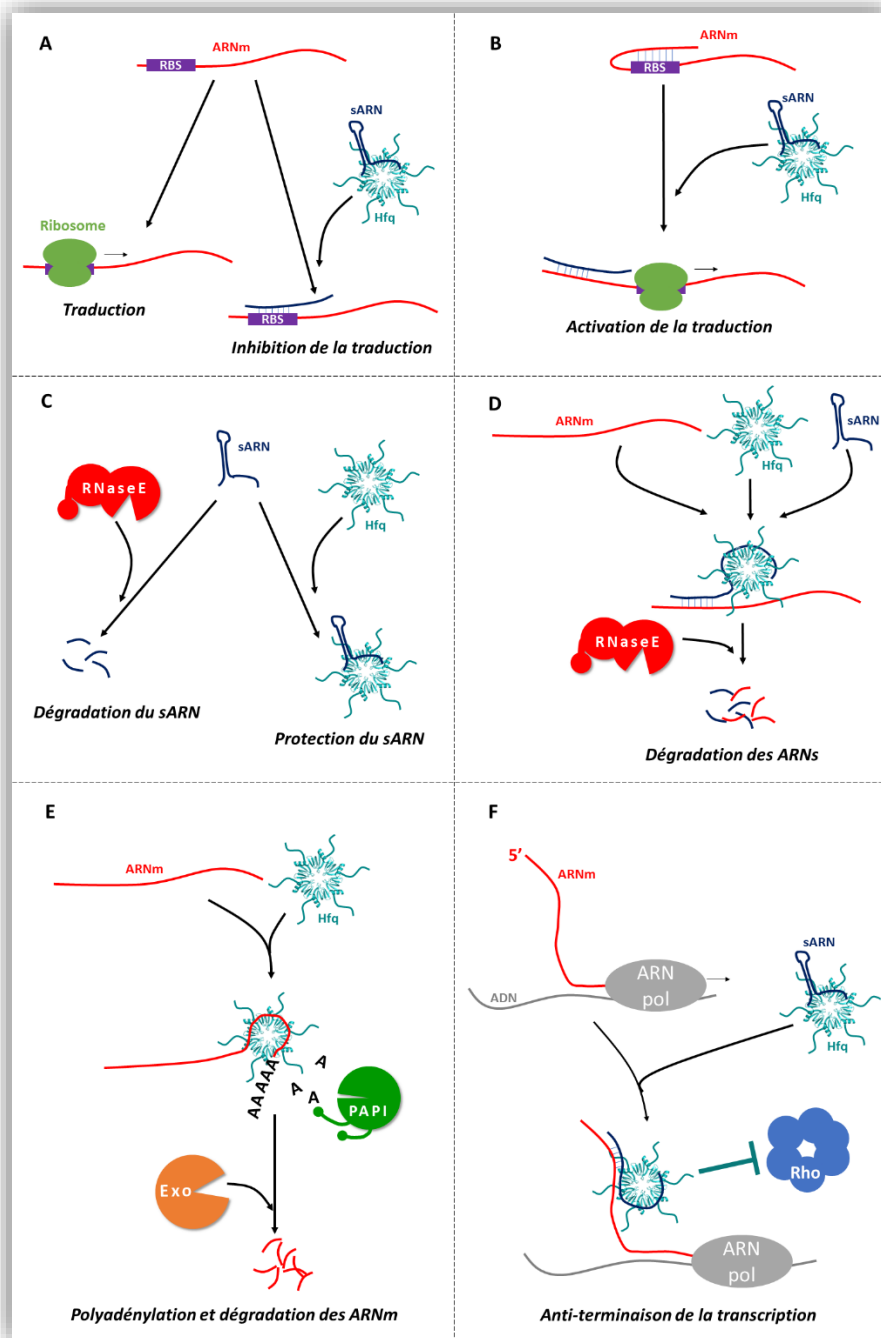


Figure 22 : Hfq intervient de différentes façons auprès des ARN bactériens. **A.** Inhibition de la traduction par blocage du RBS. **B.** Activation de la traduction par libération du RBS. **C.** Protection des sARN de l'action des ribonucléases. **D.** Induction de la dégradation de certains sARN et de leur ARNm par la RNaseE. **E.** Stimulation de la polyadénylation d'ARNm et de leur dégradation par des exonucléases. **F.** Anti-terminaison de la transcription.

Ce grand nombre de partenaires ribonucléiques pose question sur la manière dont Hfq les discrimine, les niveaux d'expression des ARN et leur affinité pour Hfq étant très variables. Cette affinité est parfois relativement faible, ce qui est par exemple le cas de plusieurs ARNm cibles de RyhB tels que *sodA*, *nuoA* ou *fdoG*, gènes impliqués dans l'utilisation de la ressource en fer. Comment la chaperonne peut-elle assurer ses fonctions de régulation dans ces conditions ? Une étude a pu en partie répondre à la question [108]. Les auteurs ont démontré qu'une autre protéine, une (p)ppGpp synthétase exprimée en cas de baisse de la disponibilité en acides-aminés appelée RelA, favorise la multimérisation de Hfq, ce qui augmente la disponibilité en chaperonne pour fixer les ARN et permettrait d'effectuer la régulation via RyhB, impossible autrement. Le mécanisme d'action de RelA sur Hfq est cependant encore inconnu.

### 3.3.2 Considérations structurales.

La rencontre sARN-Hfq-ARNm suit généralement les 3 étapes suivantes : 1. L'interaction d'un ARN avec Hfq entraîne l'ouverture de ses structures secondaires ; 2. Ce dépliement permet son appariement avec l'ARN partenaire ; 3. La formation du duplex ARN-ARN entraîne la libération d'Hfq, qui devient disponible pour un nouveau cycle. L'existence de complexes Hfq-sARN-ARNm est donc plutôt transitoire, contrairement à ce qui avait été initialement proposé. Cela permet aussi d'expliquer le paradoxe « haute affinité Hfq-ARN, haut turnover » : les sARN doivent alterner rapidement sur Hfq pour permettre à la bactérie de répondre rapidement aux changements et assurer une riborégulation efficace [109], [110].

Les différentes parties de la structure ont chacune un rôle particulier. On distingue la face proximale (où sont les hélices  $\alpha$ ), la face distale, le pourtour de l'anneau, et le domaine flexible C-terminal. Tout comme pour les autres protéines Sm, la face proximale de Hfq est connue pour interagir avec les régions riches en uracile. Plus précisément, il s'agit de la zone proche du pore, avec une stoechiométrie d'un nucléotide par monomère. Le site d'interaction est au niveau d'un aromatique bien conservé (Phe-42 pour *E. coli*) [111]. En revanche, la face distale interagit préférentiellement avec les régions riches en adénine, et particulièrement le motif Adénine-Purine-indéterminé (noté A.R.N.), avec trois nucléotides et donc trois sites de

fixation par monomère [112]. La zone est plus éloignée du pore central que pour la face proximale permettant la fixation de séquences plus longues. Ainsi, on retrouve généralement les sARN sur la face proximale, puisqu'ils possèdent une séquence riche en uracile en 3' (terminateur Rho-indépendant) [113]. Les ARNm, qui ont souvent une zone riche en adénine dans leur région 5'-UTR et pas systématiquement de terminateur Rho-indépendant, vont sur la face distale [114], [115]. L'anneau latéral, lui, possède une triade arginine permettant la stabilisation des interactions en fixant les zones UA répétées, sans discrimination entre sARN et ARNm [116]. Des travaux récents ont suggéré que cette zone puisse servir d'aire de contact entre les deux ARN, ou bien qu'elle joue un rôle dans le cycle de fixation-libération et la discrimination des différents ARN [117].

Un modèle récent propose, en combinant les faces proximales et distale ainsi que l'anneau externe, de ranger les petits ARN en deux classes suivant leur type interaction avec Hfq et leur ARNm cible [117]. La première classe correspond à des appariements de type « proximale-distale », avec les deux partenaires chacun d'un côté de la protéine. On y retrouve DsrA par exemple. La seconde classe, elle, regroupe les sARN interagissant à la fois avec les faces distale et proximale, comme ChiX. Le messenger vient alors s'enrouler autour de l'anneau externe pour s'apparier (**Figure 23**). Des constructions chimériques astucieuses inversant les propriétés des deux classes d'ARN ont confirmé ces résultats chez *Salmonella*, même s'il reste des sARN ne correspondant à aucune des deux catégories proposées [118].

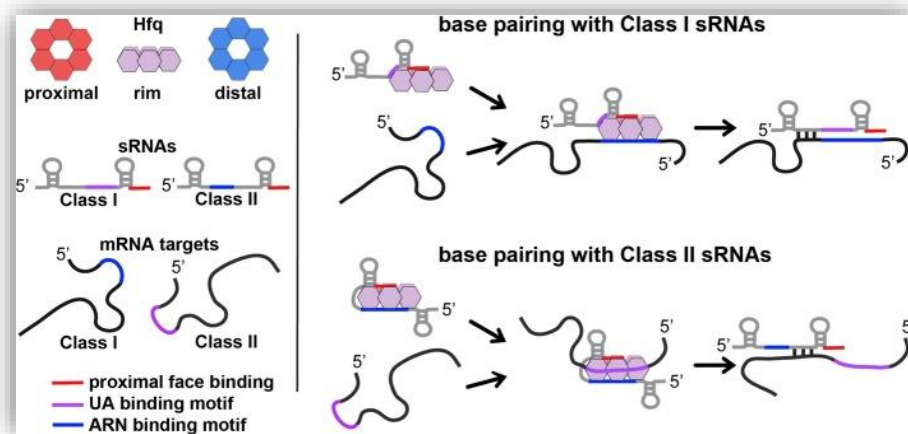


Figure 23 : Les sARN sont divisés en deux classes en fonction de leur type d'interaction avec Hfq. Classe I : le sARN fixe la face proximale avec son extrémité 3' riche en uracile, et le messager fixe la face distale avec le motif « A.R.N. ». Classe II : le sARN fixe à la fois les faces proximale et distale, et le messager s'enroule autour de l'anneau externe avec une zone riche en UA.

Concernant le domaine C-terminal, les versions tronquées de la protéine assurent encore la majorité de leurs fonctions de riborégulation [119]. La complémentation d'un mutant *hfq*- de *E. coli* par la protéine Hfq de *C. difficile* qui possède un CTR très court et polaire restaure certaines fonctions (autorégulation de Hfq, régulation négative de l'expression du transporteur *oppA*, activation de l'expression de *rpoS*) mais ne permet plus de répondre aussi efficacement à un stress phosphoglycosidique via SgrS, entraînant une baisse du taux de croissance [120]. On peut aussi citer des variants d'autres espèces possédant un CTR court ou absent et qui pourtant sont capables de substituer en partie la riborégulation effectuée par la version naturelle de Hfq chez *E. coli* : *P. aeruginosa* (réplication du phage Q $\beta$ , expression de *rpoS*, répression de *ompA*) [121] et même l'orthologue de Hfq chez l'archée *M. jannaschii* (expression de *rpoS* et régulation du métabolisme du fer via RyhB) [122].

Ces résultats suggèrent que le cœur de la riborégulation de Hfq réside principalement dans son tore. Cependant, des travaux plus récents en cinétique rapide (*stopped-flow*) ont permis la mise en évidence d'un rôle « d'écarteur » du CTR [123]. En effet, le domaine ne permet pas d'accélérer la formation du complexe sARN-ARNm mais au contraire est requis pour sa libération : là où l'hexamère facilite l'hybridation petit ARN - ARN

messager, le CTR facilite leur dissociation de la protéine, la rendant ainsi disponible pour d'autres ARN. Il permet aussi une compétition et une discrimination entre sARN de classe I et de classe II : la suppression du domaine modifie l'ordre de priorité d'interaction entre les différents sARN, qui possèdent différentes affinités, et la protéine. Ce rôle est central pour une riborégulation efficace puisque la quantité de Hfq est limitante dans la cellule comparée à celle en sARN.

### 3.4 Article A : *The Escherichia coli Hfq protein: an unattended DNA-transactions regulator.*

« La protéine Hfq de *Escherichia coli* : un régulateur sous-estimé de processus impliquant l'ADN. »

Publié dans *Frontiers in Molecular Biosciences*. Joint en **Annexe A**.

Grzegorz Cech, Agnieszka Szalewska-Pałasz, Krzysztof Kubiak, **Antoine Malabirade**, Wilfried Grange, Veronique Arluison, Grzegorz Wegrzyn.

Même si Hfq est principalement connue pour son rôle de riborégulation, la protéine interagit aussi avec l'ADN, et pourrait remplir des fonctions de régulation supplémentaires au niveau transcriptionnel. La revue qui suit examine cet aspect en discutant son mode de liaison à l'ADN, son appartenance aux NAPs et ses caractéristiques comparées à leurs principales représentantes. Plusieurs autres processus mettant en jeu de l'ADN et où Hfq est impliquée de façon indirecte sont exposés, notamment la répllication des plasmides de type ColE1 et certaines transpositions.

### 3.5 Interactions protéine-protéine.

De par sa fonction de régulateur pléiotropique, l'*interactome* de Hfq est très étendu. Ainsi, une étude effectuée par co-précipitation a relevé presque 60 partenaires protéiques, même si toutes ces interactions ne sont pas démontrées et qu'un certain nombre pourraient être indirectes [124]. L'hélicase CsdA par exemple n'a pas été retrouvée directement en complexe avec Hfq contrairement à ce qui était admis [51]. Ces interactions pourraient être parfois transitoires, labiles et même rares, et des techniques plus précises comme le *single-molecule pull-down* (dispositif appât-proie en molécule unique) pourraient permettre de mieux les appréhender [125].

Parmi les partenaires confirmés, on retrouve évidemment la RNaseE, formant un complexe direct avec Hfq et indépendant du dégradosome ainsi que des ARN [126]. On retrouve aussi des acteurs de la dégradation des ARN, comme une exoribonucléase (PNPase) et la polyA-polymérase (PAP1) [104]. De plus, Hfq s'associe à la catalase HP11 dont elle régule l'expression à travers RpoS, sans altérer sa fonction [127]. La plupart des autres partenaires potentiels sont notamment les produits des gènes régulés par Hfq, comme des protéines et transporteurs membranaires, des facteurs de transcription mais aussi la protéine *histone-like* H-NS.

### 3.6 Localisation cellulaire.

Nous avons vu que Hfq est impliquée dans des voies de régulation nombreuses et diverses. A ceci s'ajoute un grand nombre de partenaires, variés qui plus est : complexes protéiques, ADN, ARN.

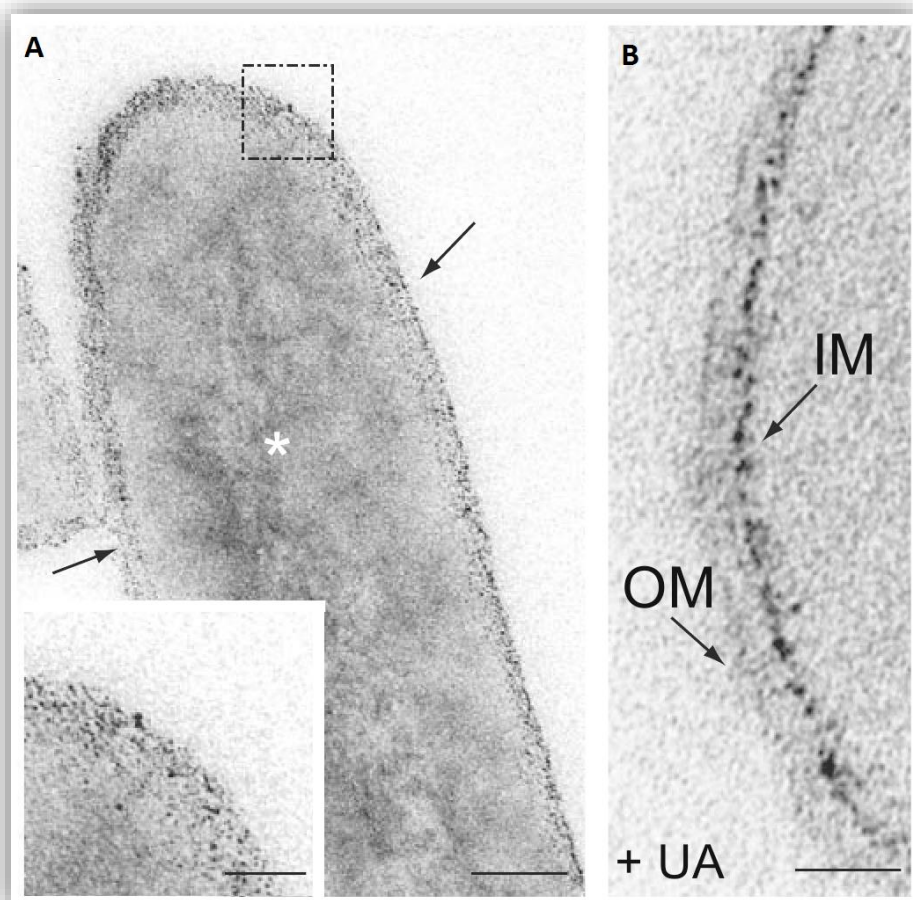
Ainsi, même si majoritairement cytoplasmique, une faible proportion (10-20%) de la chaperonne est retrouvée dans le nucléoïde bactérien [89], [128]. Sur les 10 protéines les plus associées au nucléoïde, elle est même la troisième la plus abondante sur le chromosome lors de la phase exponentielle de croissance [129].

Par ailleurs, des observations en microscopie électronique montrent une fraction importante de Hfq proche de la membrane interne, dont quelques images sont reproduites sur la **Figure 24** [130]. Elles confirment aussi la présence de Hfq dans le nucléoïde (astérisque blanc sur la **Figure 24A**). La protéine, étiquetée avec un fixateur de métal électrodense, était exprimée depuis un plasmide inductible à des concentrations comparables à celles normalement atteintes en phase stationnaire afin d'obtenir une densité suffisante.

La protéine a aussi été localisée par immunofluorescence directement à partir de son expression chromosomique. De façon surprenante, on observe une répartition hétérogène, semblant former de longues structures sous-membranaires (**Figure 25A**), semblables à celles observées de la même façon pour d'autres acteurs du métabolisme des ARN comme les composants du dégradosome. Cependant, les structures observées avec Hfq sont indépendantes de la formation du dégradosome, ainsi que de la présence d'ARN ou encore du cytosquelette bactérien [61]. La fraction présente dans le nucléoïde n'est pas visible car elle est minoritaire et cachée par la fraction

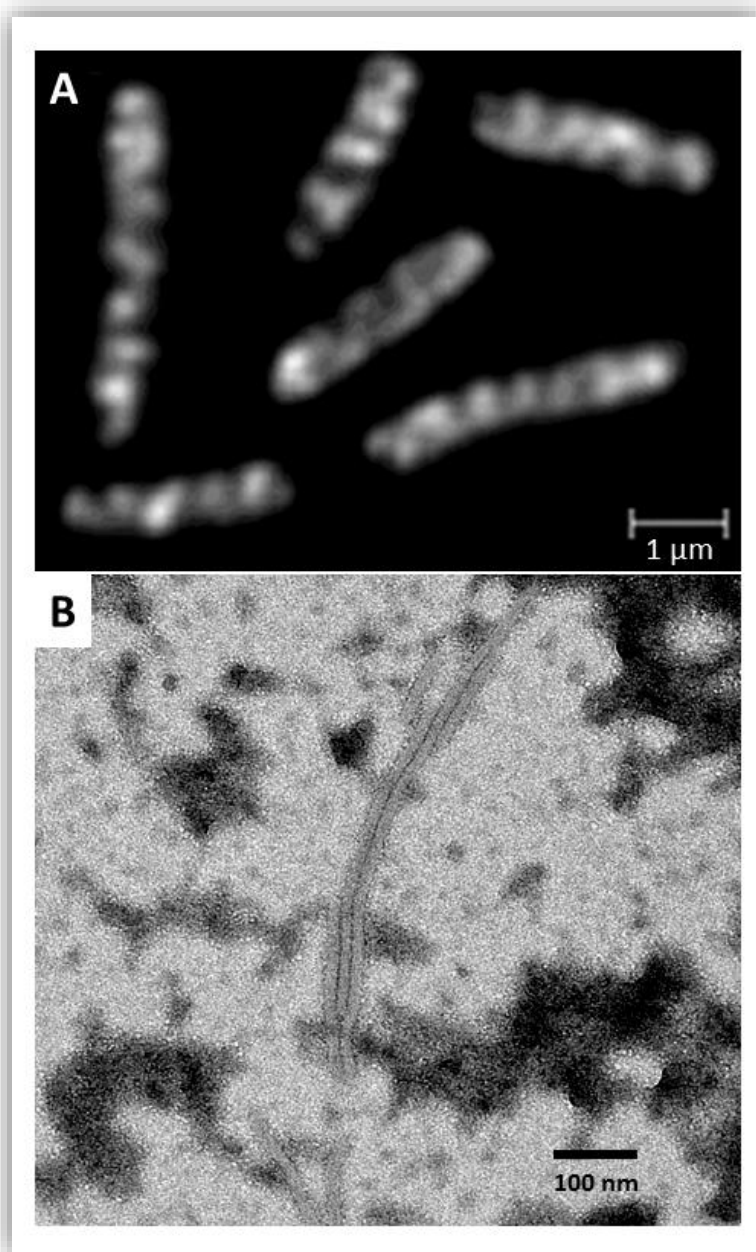
membranaire et cytoplasmique. La diffraction limitant la résolution obtenue, ces structures restent énigmatiques.

Il faut rapprocher leur existence d'une étude ayant montré la capacité de Hfq à auto-interagir et à former des fibres protéiques *in vitro* (Figure 25B), semblables à ce qui est parfois observé chez d'autres protéines Sm [131]. Même si les conditions de formation de ces assemblages sont assez éloignées des conditions physiologiques (faible force ionique), ces résultats suggèrent que Hfq pourrait s'auto-assembler proche de la membrane interne de *E. coli* pour former les structures observées par immunofluorescence.



**Figure 24 : Localisation de Hfq par cryo-TEM avec un étiquetage à la métallothionéine. A.** Une fraction importante de la protéine est présente au niveau de la membrane interne de la bactérie, ainsi que dans le nucléoïde. **B.** L'ajout d'un agent de contraste (acétate d'uranyle, UA) permet de révéler la membrane externe (OM) et ainsi de confirmer que les protéines sont proches de la membrane interne (IM). Barres d'échelle : 100 nm. Reproduit d'après [130].





**Figure 25 : Auto-assemblage de Hfq.** A. La localisation de Hfq chez *E. coli* par immunofluorescence montre une répartition hétérogène qui semble être organisée en super structures sous-membranaires. B. Microscopie électronique montrant que sous certaines conditions, Hfq est capable de s'auto-assembler *in vitro* en formant de longues structures fibrillaires. D'après [131], [132].

## 4 But de la présente étude.

**H**fq, protéine pléiotropique bactérienne, est principalement connue pour son action riborégulatrice. Mon travail de thèse s'est focalisé sur des aspects peu étudiés de la protéine, à savoir (I) son domaine C-terminal dont la fonction n'est pas comprise, et cela dans un contexte de caractérisation (II) du rôle structural de l'interaction de Hfq avec le nucléoïde et (III) de la formation des structures sous membranaires observées dans la bactérie.

Chapitre II :  
Agrégation  
amyloïde du  
domaine carboxyl  
terminal de la  
protéine Hfq.

# 1 Contextualisation : les domaines non-structurés et les amyloïdes.

## 1.1 Les domaines intrinsèquement désordonnés.

Les 15 dernières années ont permis l'émergence d'un nouveau niveau de compréhension de la relation entre structure et fonction des protéines. L'existence d'une classe entière de polypeptides et sous-domaines protéiques qui ne possèdent pas de structure rigide mais accomplissent des fonctions biologiques cruciales a été longtemps sous-estimée, malgré quelques exemples bien connus [133]. Les travaux visant à comprendre ce que les protéines nativement non-repliées avaient de si spécial ont suscité un véritable intérêt, et ont entraîné l'émergence d'un nouveau domaine scientifique à part entière, dédié aux IDPs.

Un domaine intrinsèquement désordonné (IDP, *Intrinsically Disordered Protein*, mais cela peut concerner un domaine protéique uniquement) est défini comme étant une séquence ou une sous-séquence protéique n'adoptant pas une structure tridimensionnelle préférentielle, mais plutôt une variété de conformations différentes et souvent dynamiques. Ainsi, il est difficile voire impossible de cristalliser ces domaines. En général, les séquences intrinsèquement désordonnées sont caractérisées par un faible taux d'acides aminés hydrophobes (dont Val, Leu, Ile, Met, Phe, Trp, Tyr), et une forte proportion de résidus polaires et chargés (dont Thr, Asn, Gln, His, Ser, Pro, Glu, Lys, Gly) (Figure 26) [133], [134]. Ceci leur confère une bonne capacité d'hydratation. De plus, les nombreuses charges électriques sont génératrices de désordre de par les répulsions électrostatiques existantes entre les charges de même signe. Ainsi, les IDPs ne peuvent pas suffisamment enfermer un cœur hydrophobe pour se replier sous forme globulaire. Ces domaines n'ont pas de structure secondaire régulière et sont souvent appelés « régions flexibles ». Beaucoup d'IDPs sont aussi caractérisés par une faible complexité de séquence, avec une surreprésentation de certains acides aminés. Cependant, même si des séquences de faible complexité sont souvent synonymes d'un IDP, le contraire n'est pas nécessairement vrai et il existe des IDPs avec des séquences variées.

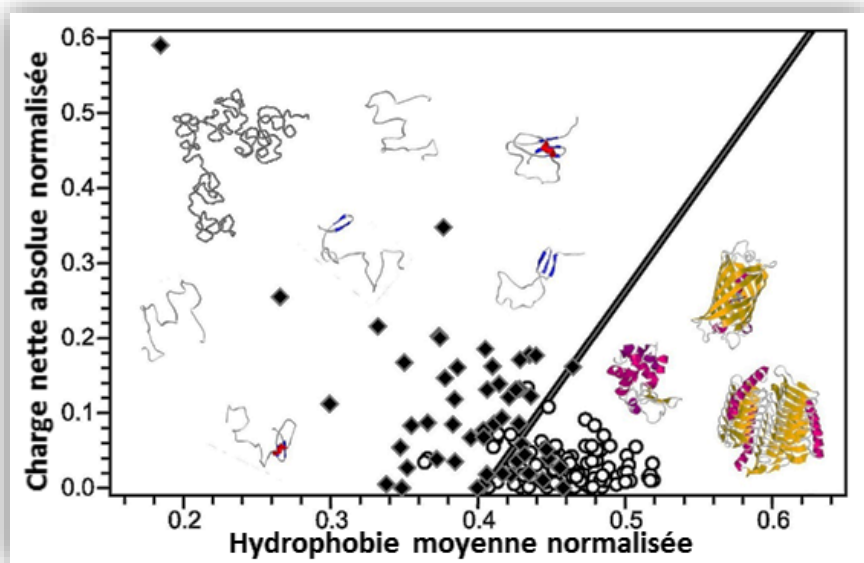
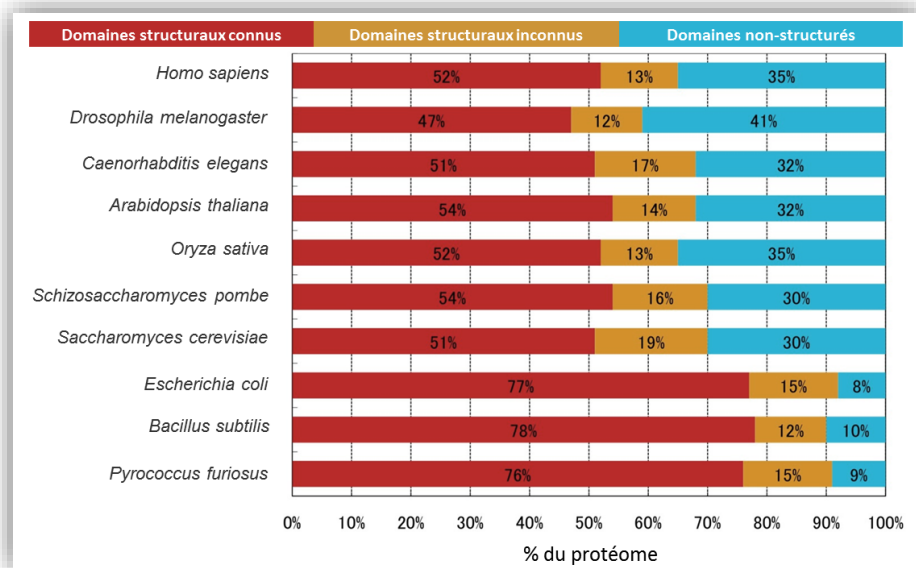


Figure 26 : Particularités de la composition en acides-aminés des IDPs. Comparaison de la charge nette moyenne et de l'hydrophobie moyenne de 275 protéines structurées (cercles) et 91 protéines nativement non-structurées (losanges). La ligne noire représente la frontière entre IDPs et protéines globulaires compactes. Quelques exemples de structures sont présentés. © 2011 Elsevier Ltd, reproduit avec autorisation de [133].

Les IDPs sont fréquents dans la plupart des protéomes. Les prédictions informatiques basées sur l'analyse des séquences primaires en acides-aminés indiquent que la fraction de protéines possédant un ou plusieurs IDPs est plus importante chez les eucaryotes que chez les archées/bactéries. Ainsi, plusieurs sources prédisent que chez les eucaryotes, environ 30-35 % des protéines contiennent des IDPs de plus de 30 résidus. Chez les archées et chez les bactéries, cette proportion tombe à 5-10 % [135], [136] (Figure 27). La fraction d'IDPs chez les procaryotes concerne plus fréquemment des régions courtes que chez les eucaryotes [136]. De plus, la proportion d'IDPs est plus grande parmi les protéines localisées au niveau nucléaire / nucléoïde, et particulièrement chez les protéines qui se lient à l'ADN dont les facteurs de transcription [136]. De même, on retrouve des IDPs en plus grande proportion chez les protéines fixant des ions métalliques ou des hèmes, ainsi que les protéines interagissant avec les membranes [133].



**Figure 27** : Fractions relatives de domaines protéiques connus, domaines protéiques inconnus et domaines protéiques non-structurés dans 10 organismes modèles, déterminés par le système prédictif DICHOT. Les domaines connus concernent des structures résolues dans PDB, les domaines non déterminés concernent des structures globulaires encore non répertoriées dans PDB mais conservées dans de nombreuses séquences et donc prédites comme structurées, les domaines non-structurés sont des IDPs. Adapté d'après [136].

Les régions protéiques non-structurées sont sujettes à des évolutions de séquence plus fréquentes que la moyenne. En effet, les IDPs sont par définition peu contraintes au niveau structural, contrairement aux régions enfouies. Il a été montré que les changements d'acides-aminés exposés au solvant sont plus fréquents que les autres [137]. En suivant cette logique, les régions non-structurées d'une protéine devraient être moins bien conservées que les régions structurées, et c'est ce qui est constaté dans la plupart des protéines étudiées [138].

Les IDPs ne sont pas dépourvus d'intérêts fonctionnels, bien au contraire. De nouveaux exemples de domaines désordonnés fonctionnels émergent constamment. Ces fonctions comprennent la régulation de la transcription et de la traduction, la transmission de signaux cellulaires, le stockage de petites molécules, et la régulation de l'auto-assemblage de complexes multiprotéiques, comme le flagelle bactérien ou les ribosomes [134]. Il a aussi été rapporté des cas d'IDPs servant de châssis pour l'interaction de

partenaires protéiques [139]. Chez les eucaryotes, les protéines ayant de nombreux partenaires possèdent d'ailleurs plus fréquemment des domaines non-structurés que la totalité des protéines [140]. Et nous avons déjà vu le cas du domaine C-terminal de la RNaseE chez *E. coli*.

De nombreux IDPs peuvent effectuer des transitions vers des états plus ordonnés ou bien se replier de façon stable en structures secondaires lorsqu'ils interagissent avec leurs cibles : le processus est appelé « repliement et interaction couplés ». Dans ce cas, le coût entropique du repliement est « payé » par l'énergie enthalpique de la liaison [134]. En effet, la liaison est généralement une contribution enthalpique spontanée et constitue donc le moteur du processus. Il s'agit d'un exemple de compensation enthalpie-entropie. Les processus de repliement et interaction couplés sont fréquemment retrouvés dans le cas de partenaires ayant une spécificité haute et une affinité relativement faible, ce qui est approprié pour des transmetteurs de signaux cellulaires par exemple, qui doivent à la fois être capable d'interagir pour transférer un signal mais aussi facilement se dissocier lorsque le signal s'éteint. De façon générale, tous les processus impliquant des interactions labiles peuvent intégrer un rôle fonctionnel des IDPs.

## 1.2 Les structures de type cross- $\beta$ -amyloïde.

Les structures amyloïdes ont d'abord été découvertes car elles sont impliquées dans un certain nombre de maladies, notamment neurodégénératives. Regroupées sous le terme « amyloïdoses », ces maladies à agrégation protéique sont provoquées par des polypeptides ou protéines s'auto-assemblant en structures partiellement ordonnées, les fibres amyloïdes, souvent insolubles et qui se déposent dans les tissus et les organes. Plus de 30 protéines ou polypeptides amyloïdogéniques différents sont associés avec des maladies humaines, dont l'amyloïdose systémique, Alzheimer, Parkinson ou encore le diabète de type 2 [141]–[143]. Ces fibres amyloïdes partagent des caractéristiques structurales communes, même s'il n'y a pas d'homologie de séquence parmi les protéines qui forment des amyloïdes *in vivo*. Ce sont des protéines linéaires, non ramifiées, et riches en feuillets  $\beta$ . Ces feuillets s'agglomèrent en couches successives pour finir par former de longues structures fibrillaires, l'axe de la fibre s'étendant perpendiculairement à la

direction des feuillets, et les liaisons hydrogènes inter-feuillets étant parallèles à cet axe. Cette conformation dite cross- $\beta$  est commune à l'ensemble des amyloïdes caractérisés jusqu'ici.

### 1.2.1 Les amyloïdes sont une des structures possibles dans l'espace conformationnel accessible aux protéines.

Les protéines amyloïdogéniques peuvent être classées en deux classes structurales : celles qui forment des structures globulaires habituelles dans leur état natif, et celles qui sont initialement intrinsèquement désordonnées. Parmi les exemples de protéines appartenant à cette dernière catégorie, on retrouve l'amyline ou IAPP (*Islet Amyloid Polypeptide*) responsable de l'amyloïdose pancréatique du diabète de type 2, le peptide A $\beta$  de la maladie d'Alzheimer, et l' $\alpha$ -synucléine, impliquée dans la maladie de Parkinson. De façon plus large, les amyloïdes sont l'un des multiples états conformationnels accessibles aux protéines. En effet, une séquence protéique peut adopter de nombreux états conformationnels dans un système vivant, entre son étape de biosynthèse par le ribosome et sa dégradation finale par protéolyse, résumés sur la **Figure 28** [142]. Ces différentes conformations impliquent une série d'équilibres complexes. Leurs paramètres thermodynamiques et cinétiques dans le système cellulaire sont déterminés par la séquence primaire en acides-aminés, ainsi que par les paramètres physico-chimiques, les interactions avec les chaperonnes moléculaires, les processus de dégradations, et d'autres mécanismes de « contrôle qualité » de la cellule [144]. Dans certaines circonstances, ces protéines ne sont pas maintenues dans un état soluble et peuvent se convertir en agrégats potentiellement non-fonctionnels et toxiques pour le système vivant.



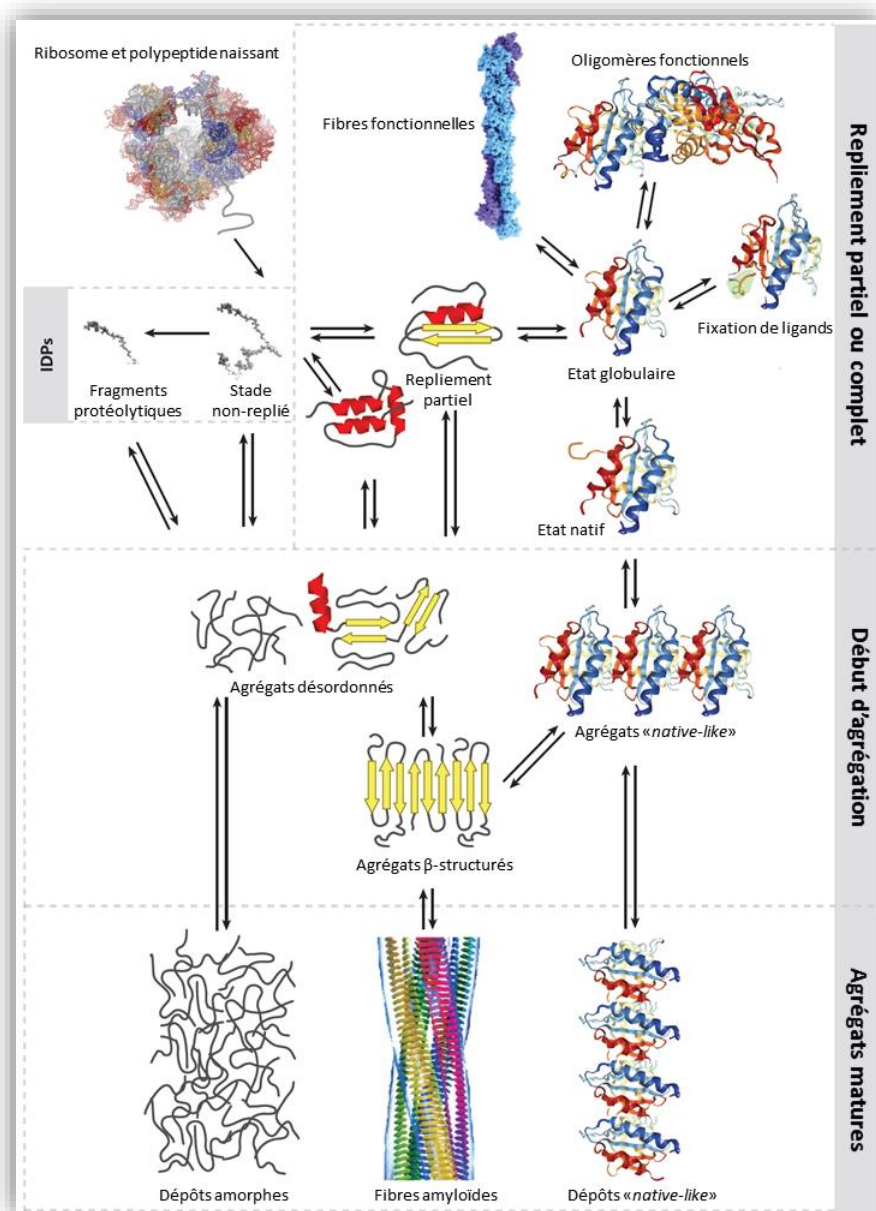


Figure 28 : Représentation schématique de la variété d'états conformationnels que peut adopter une chaîne polypeptidique après sa biosynthèse ainsi que les transitions possibles entre ces différents états.

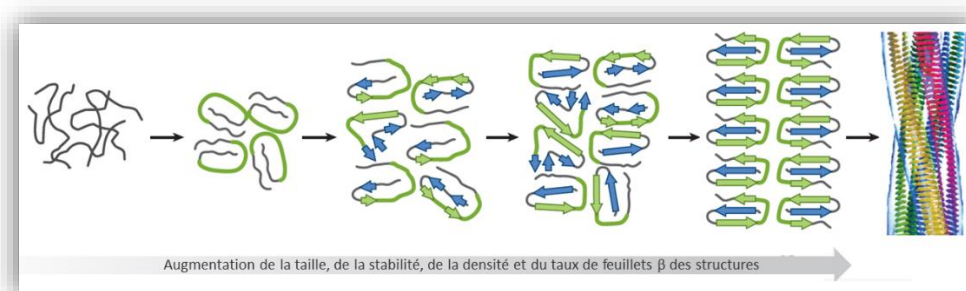
L'ensemble des conformations et leurs interconnexions sont régulés dans l'environnement cellulaire par le réseau de la « protéostasie ». Les agrégats peuvent former des fibres amyloïdes (bas, centre), des dépôts de protéines natives (bas, droite) ou bien des dépôts amorphes (bas, gauche) qui peuvent être impliqués dans des pathologies lorsqu'ils se forment de manière incontrôlée.

D'après [142].

### 1.2.2 Mécanisme de formation.

Le processus de formation des fibres est progressif, impliquant des étapes de nucléation puis de croissance des agrégats. Dans un deuxième temps, des étapes de fragmentation et de nucléation secondaire (où la surface d'une fibre en croissance peut catalyser la formation de nouveaux agrégats) peuvent se produire [145]. Les espèces qui se forment au début de la phase d'agrégation sont des petits regroupements qui gardent souvent une conformation proche de celle des monomères les composant, donnant des oligomères non-structurés, partiellement structurés ou complètement repliés [146]–[148]. Cependant cela n'est pas systématique et il arrive qu'un IDP donne une structure oligomérique très ordonnée et vice-versa. Ces premiers oligomères sont peu stables, impliquant des interactions faibles, et leur dissociation pour retourner à un état monomérique est encore possible pour un faible coût énergétique. L'étude expérimentale de ces stades métastables est par conséquent difficile. Elle est pourtant d'un grand intérêt, car ces oligomères amyloïdes ont une capacité accrue à pénétrer et déstabiliser les membranes, à interagir avec les composants extracellulaires et à provoquer une cytotoxicité, comparé aux formes monomériques et fibrillaires des mêmes protéines [149]–[151].

Lorsque l'agrégation se poursuit, ces premiers oligomères peuvent subir de nouvelles réorganisations internes résultant en des espèces plus stables, plus compactes, de taille et de contenu en feuillets  $\beta$  croissants (**Figure 29**). De façon intéressante, un certain nombre d'IDPs évoluant en amyloïdes le font en passant par un intermédiaire  $\alpha$ -hélicale [152].



**Figure 29 : Formation progressive de fibres amyloïdes.** Le schéma représente les réarrangements structuraux prenant place pendant l'oligomérisation. Les segments amyloïdogéniques initiaux de la séquence primaire sont colorés en vert, la flèche (bas) indique les paramètres structuraux en augmentation à mesure que l'agrégation se poursuit. D'après [142].

Ces oligomères  $\beta$ -structurés peuvent ensuite incorporer de nouveaux monomères ou s'assembler entre eux pour évoluer en fibres bien définies, de structure cross- $\beta$ . Alternativement, les oligomères désordonnés ou nativement repliés peuvent s'agglomérer sans réarrangement structurel majeur et former de grands dépôts amorphes ou des assemblages de protéines natives (**Figure 28**).

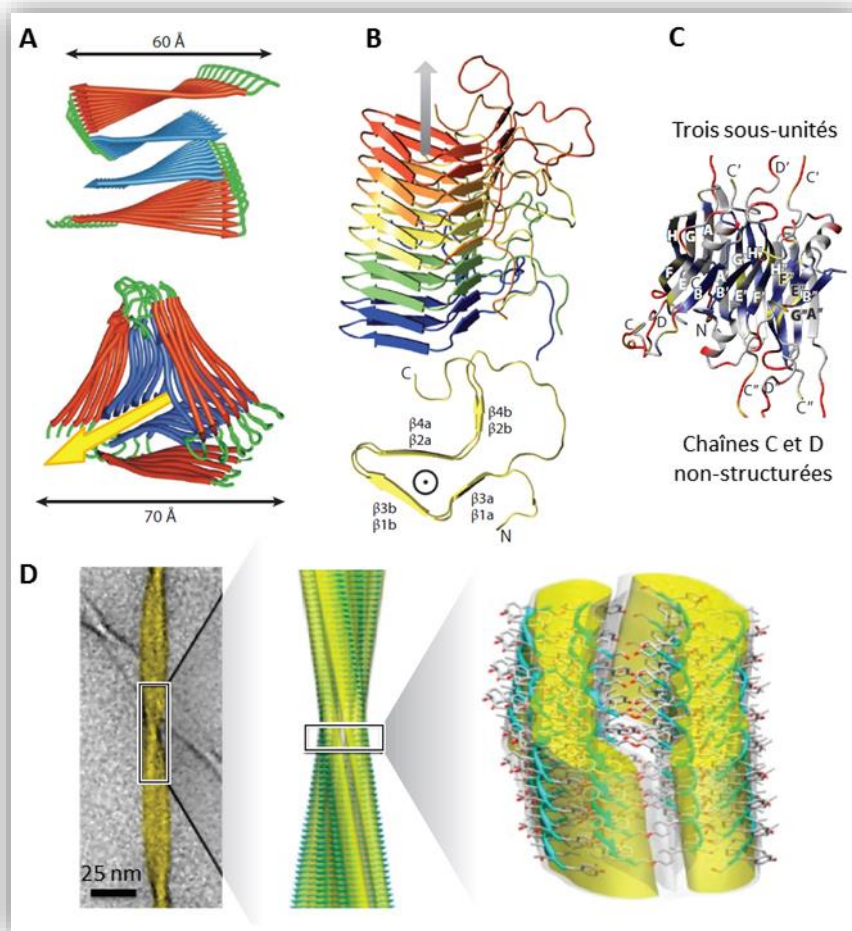
### 1.2.3 Caractéristiques structurales générales des fibres amyloïdes.

Les fibres amyloïdes font typiquement 7-13 nm de diamètre et jusqu'à plusieurs microns de long, tel qu'observé par TEM ou AFM. Elles sont souvent composées de 2 à 8 protofilaments, tournant les uns autour des autres ou bien étant associés latéralement en rubans superposés (**Figure 30**). Il existe aussi des fibres faites d'un unique filament. Les feuillets sont empilés perpendiculairement à l'axe de la fibre, comme observé par diffraction aux rayons X, RMN du solide et FTIR [153], [154]. Il arrive que les feuillets soient parallèles ou antiparallèles, mais une grande partie des structures résolues montre une conformation parallèle « *in-register* », c'est-à-dire avec les résidus identiques de chaque feuillet empilés à la même position.

L'existence de cette structure cross- $\beta$  peut aussi être confirmée par la liaison de sondes fluorescentes comme la Thioflavine-T ou le rouge de Congo, qui forment des dépôts ordonnés le long des fibres donnant des réponses spectrales spécifiques [155]. L'empilement des feuillets et le regroupement de plusieurs protofilaments donnent lieu à de nombreuses liaisons hydrogènes, responsables de la grande stabilité et résistance mécanique de la plupart des fibres amyloïdes, constatée par AFM notamment [156]. Ces trois caractéristiques (morphologie fibrillaire, structure cross- $\beta$ , réactivité spécifique des sondes) sont les piliers expérimentaux utilisés pour classer un agrégat protéique comme étant de structure amyloïde. Différentes techniques sont détaillées dans l'Article C : *Techniques to analyse sRNA protein cofactor self-assembly in vitro*.

Un aspect intrigant est la remarquable similarité de morphologies et de structures entre les différentes protéines amyloïdes étudiées, qui sont pourtant issues d'une grande variété d'organismes et de séquences primaires. Cette similarité peut être attribuée à la prépondérance de la chaîne carbonée principale, commune à tous les polypeptides, dans la formation de la

structure. Les variations observées viendraient donc des différences de longueur et de chaînes latérales des peptides, ainsi que d'autres facteurs dont les conditions physico-chimiques [157]. Ce dernier point est responsable du polymorphisme de certaines fibres qui adoptent des arrangements différents selon les conditions d'agrégation (Figure 30A).

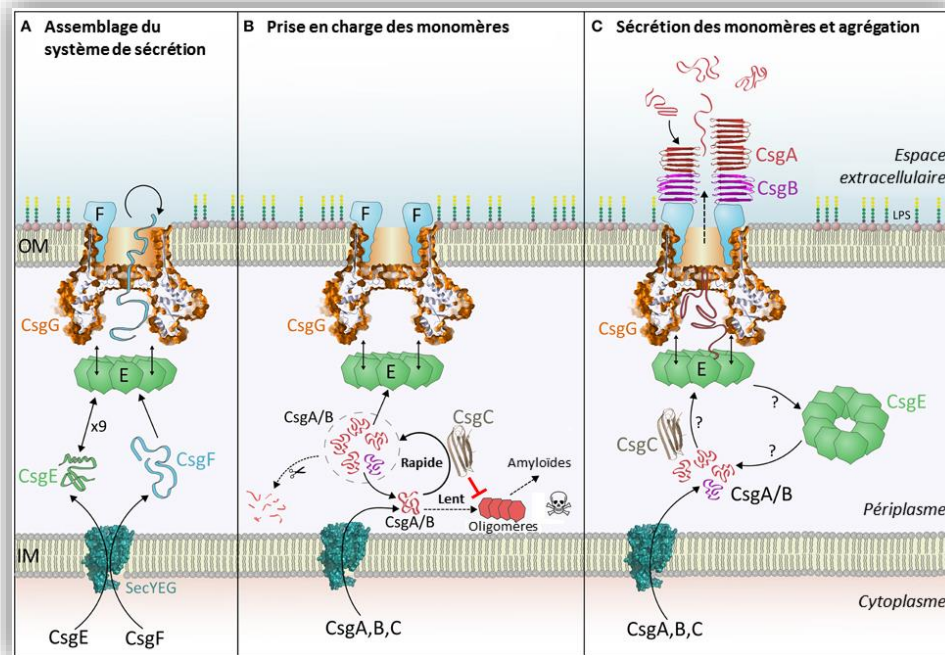


**Figure 30 : Fibres amyloïdes : une morphologie commune basée sur une structure cross- $\beta$ , des arrangements variables.** A. Deux modèles structuraux polymorphes des protofilaments formés par le peptide A $\beta$ <sub>40</sub>. Chaque couche consiste en deux (haut) ou trois (bas) molécules. B. Structure en solénoïde de la fibre formée par le domaine prion du HET de *Podospora anserina*. Chaque couleur indique une molécule, contenant 8 feuillets. C. Modèle structural des filaments « native-like » formés par la transthyréine, ayant les chaînes A-G-H et B-E-H sous forme de feuillets, composant le cœur  $\beta$  de la fibre et les chaînes C et D sous forme non-structurée s'étendant latéralement. D. Structure de la fibre à 6 protofilaments formée par le peptide correspondant aux résidus 105 à 115 de la transthyréine. Chaque couche contient 12 peptides. Adapté de [142].

#### 1.2.4 L'émergence du concept d'amyloïde fonctionnel.

Alors que la plupart des efforts de recherche étaient orientés sur les amyloïdes cytotoxiques, des travaux ont peu à peu démontré l'existence de nombreux amyloïdes utiles à la cellule. Ceux-ci sont dits « fonctionnels », et existent dans de nombreux organismes, des mammifères aux bactéries. Leurs rôles sont variés : formation du biofilm, développement de structures cellulaires externes, support structural, transfert de l'information (notamment sous forme de prions), etc. L'apport évolutif des amyloïdes doit être apprécié en regard de ses avantages : résistance élevée à la protéolyse, propriété d'auto-propagation, et un rôle de mémoire moléculaire transmissible [158].

Chez les bactéries, l'amyloïde le plus étudié concerne le développement des fibres de type curli. Ces fibres font partie de la matrice extracellulaire permettant la formation du biofilm et participent à l'adhésion de la bactérie lors de l'invasion de son hôte. Chez *E. coli*, c'est la protéine CsgA qui est à l'origine de cet assemblage, aidée de ses partenaires CsgBCEFG, transcrits par deux opérons. Lorsqu'elle est excrétée, elle se transforme en fibre amyloïde au contact de CsgB, positionné sur la membrane externe et servant de nucléateur. Le processus, complexe, est présenté **Figure 31**. On notera l'intervention élégante de CsgC dans le périplasme, permettant d'inhiber la formation prématurée des amyloïdes par CsgA et CsgB avant leur traversée de la membrane externe, ce qui serait toxique pour la bactérie le cas échéant [159]. La structure amyloïde, elle, est formée à partir d'une séquence courte répétée où les asparagines et glutamines jouent un rôle prépondérant [160]. Un système analogue a été découvert chez *Pseudomonas*, le système Fap, qui est encodé par un seul opéron *fapABCDEF* [161]. On peut aussi citer les protéines amyloïdogéniques P1, WapA et SMU\_63c qui sont impliquées dans le développement et l'architecture du biofilm chez *Streptococcus mutans* [162].



**Figure 31 : Mécanisme de sécrétion et d'assemblage du curli amyloïde chez *E. coli*.** A. CsgE et CsgG permettent la translocation de CsgF à travers la membrane externe, qui se replie et se fixe sur CsgG. B. Les monomères CsgA ou CsgB interagissent avec CsgC, qui retarde la formation spontanée d'oligomères toxiques dans le périplasm. C. Les monomères rencontrent CsgE et deviennent piégés dans la cavité périplasmique de CsgG [163]. Les sous-unités partiellement repliées traversent ensuite le pore central et sont relarguées dans le milieu extracellulaire. Une fois dehors, CsgB se fixe à CsgF et initie la nucléation de la fibre amyloïde formée par CsgA. Tiré de [159].

Un autre amyloïde découvert chez une bactérie est plus surprenant encore : il s'agit d'un domaine du régulateur de la terminaison de la transcription Rho, qui constituerait le premier prion bactérien connu à ce jour [164]. Les prions sont un groupe de protéines singulières, qui peuvent alterner entre une conformation soluble et une conformation amyloïde qui s'auto-perpétue [165]. Chez les levures, cette conversion conformationnelle entraîne des changements phénotypiques transmissibles aux cellules filles. Même si ces changements sont parfois délétères, ils peuvent aussi permettre une adaptation à de nouvelles conditions environnementales, ce qui place les prions de levure parmi les amyloïdes fonctionnels. Ces spécificités conformationnelles viennent d'une région spécifique de la séquence des prions,



appelée « domaine prion » (PrD), habituellement un IDP de basse complexité séquentielle, et capable de former des amyloïdes [166].

L'identification de ces PrDs a permis de détecter des séquences similaires dans de nombreux organismes, dont plus de 2000 candidats chez les bactéries [167]. Parmi eux, le terminateur de transcription Rho de *Clostridium botulinum*. Il a été démontré que celui-ci est capable de former des assemblages amyloïdes chez *E. coli* et que son PrD peut remplacer celui du prion de levure Sup35NM [168]. Une autre série d'expériences par la même équipe a été faite dans une souche hydride possédant un gène *rho* chimère, avec le domaine N-terminal contenant le PrD de *C. Botulinum*, et le domaine C-terminal du Rho de *E. coli*. Un rapporteur, composé d'un promoteur suivi d'un terminateur de transcription Rho-dépendant puis de *lacZ*, a permis de constater que certaines bactéries possédaient une version de Rho fonctionnelle, et d'autres non, ce constat étant basé sur l'intensité de la couleur des colonies (lorsque Rho est fonctionnel, la transcription de *lacZ* est atténuée et la couleur des bactéries cultivées sur X-gal plus claire). De plus, les nuances de couleur se propagent aux cellules filles, démontrant l'hérédité du phénotype dont est responsable le prion.

## 2 Article B : *New insight into the structure and function of Hfq carboxyl terminus.*

« Nouvelles perspectives concernant la structure et la fonction du domaine carboxyl terminal de Hfq. »

Publié dans *Bioscience Reports*. Joint en **Annexe B**.

Emilie Fortas, Federica Piccirilli, **Antoine Malabirade**, Valeria Militello, Sylvain Trépout, Sergio Marco, Aziz Taghbalout, Véronique Arluison.

Cet article présente une étude structurale du CTR de Hfq chez *E. coli*. Il démontre que ce domaine peut se structurer et qu'il est nécessaire à la répartition de la protéine en superstructures vues au §3.6 du Chapitre I.

Pour ce faire, un peptide de séquence identique aux 38 derniers acides-aminés de Hfq, prédite comme non-structurée, est synthétisé chimiquement et utilisé comme modèle. La purification de ce domaine est impossible car il se dégrade lorsqu'il est exprimé seul. Justement, une fois mis en solution, le peptide est capable de s'auto-assembler en longues fibres observées par TEM. Il se trouve que ces fibres fixent la ThioflavineT et que leur spectre infrarouge montre un épaulement caractéristique de feuilletts  $\beta$  intermoléculaires, indiquant qu'il s'agit de structures amyloïdes.

De plus, lorsqu'une version tronquée de la protéine amputée du CTR est exprimée et sa localisation observée *in vivo* par immunofluorescence, les superstructures sous-membranaires ne sont plus observées.

L'hypothèse formulée est donc que le domaine C-terminal de Hfq permet à la protéine de s'auto-assembler à travers la formation de feuilletts  $\beta$  entre les hexamères, pour former les assemblages sous-membranaires.



### 3 Article C : *Techniques to analyse sRNA protein cofactor self-assembly in vitro.*

« Techniques pour l'analyse *in vitro* de l'auto-assemblage d'un cofacteur protéique de sARN. »

Soumis à *Methods in Molecular Biology*. Joint en **Annexe C**.

David Partouche, **Antoine Malabirade**, Marisela Velez, Frank Wien, Thomas Bizien, Christophe Sandt, Ariane Deniset-Besseau, Sylvain Trepout, Sergio Marco, Véronique Arluison.

Ce chapitre de livre résume les différentes méthodes de caractérisation biophysique des fibres amyloïdes, en prenant pour exemple le domaine C-terminal de Hfq.

Les différentes techniques passées en revue sont :

- L'imagerie par microscopie de fluorescence avec un marquage à la Thioflavine-T ;
- L'imagerie par microscopie électronique (TEM), en coloration négative ou en cryo-microscopie ;
- L'imagerie par AFM en environnement sec et liquide ;
- La spectroscopie infrarouge (FTIR et ATR) ;
- Le dichroïsme circulaire avec rayonnement synchrotron (SRCD) ;
- La diffusion des rayons X aux petits angles (SAXS).

## 4 Discussion.

La région carboxyl-terminale de Hfq, prédite comme IDP, est donc capable de s'agréger et de former de longues fibres de structure cross- $\beta$ -amyloïde. C'est un exemple supplémentaire de structuration des IDPs, observée *in vitro* avec des cinétiques lentes.

En examinant la séquence du peptide utilisé pour les expériences de plus près, on peut imaginer une structuration possible de ces fibres : à pH physiologique, on distingue une zone chargée positivement, avec les quatre histidines et l'arginine, et une zone finale chargée négativement, avec l'aspartate et trois glutamates (**Figure 32**). Elles sont séparées par une zone de 11 acides-aminés (en bleu) que nous avons déterminée comme étant la séquence minimale pour la formation des fibres (voir Article F : *Membrane association of the bacterial riboregulator Hfq and functional perspectives*). Cette dernière devrait donc former les feuillets au cœur de la fibre, et les zones chargées positivement et négativement pourraient former des liaisons électrostatiques autour.

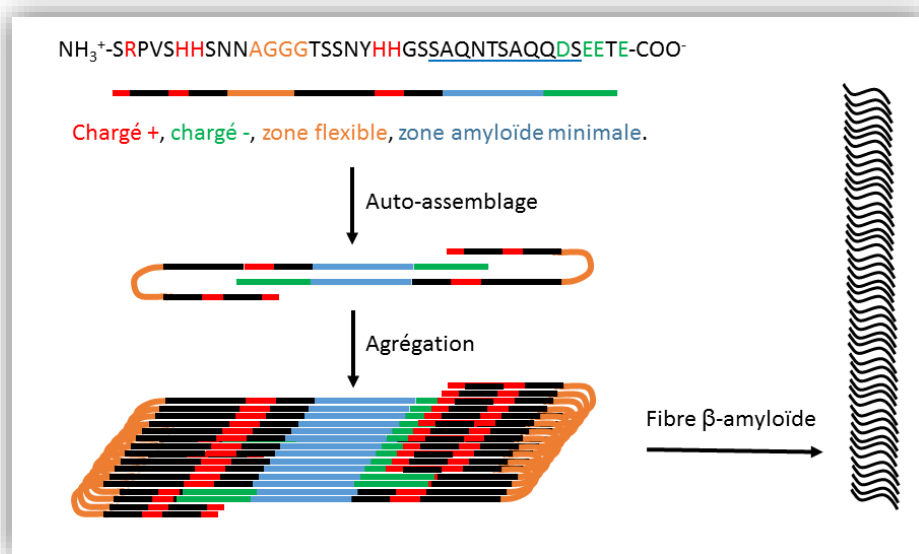
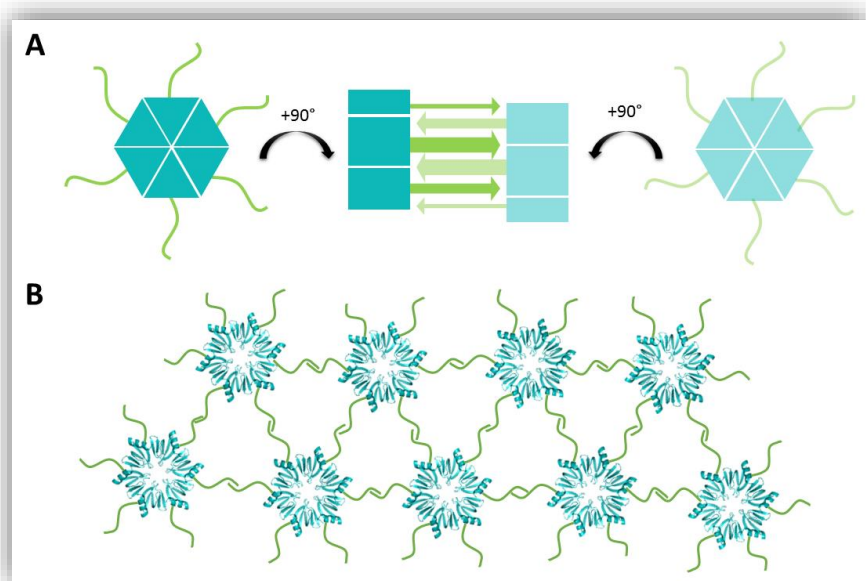


Figure 32 : Une hypothèse du processus d'agrégation de la région C-terminale de Hfq observé *in vitro*. Le modèle n'est pas à l'échelle.

L'auto-assemblage de Hfq avait déjà observé auparavant [131]. Cependant, ces nouveaux résultats suggèrent que le CTR de la protéine joue un rôle prépondérant, sinon essentiel, dans ce processus. En effet, libérées des

contraintes stériques imposées par le tore, les queues flexibles C-terminales s'empilent spontanément et forment des fibres extrêmement longues.

Une question se pose : ces empilements  $\beta$  existent-ils lorsque la protéine est entière, en présence du tore, et/ou dans la bactérie ? Les fibres formées par la protéine native pourraient contenir des portions amyloïdes, mais il paraît difficile d'envisager un empilement spontané des queues C-terminales sur plusieurs microns. La surface du NTR étant fortement chargée et encombrante, la stabilité d'une telle structure dans les conditions physiologiques est peu probable. En revanche, la formation d'un réseau protéique est possible. De tels réseaux peuvent être difficiles à observer, la nature des interactions étant faibles étant donné le caractère flexible des CTR. De plus, il est facile de perturber leur bonne formation par l'ajout d'étiquettes fluorescentes par exemple. Cependant, des formations de dodécamères de Hfq *via* leur queues C-terminales ont été observées *in vivo* avec une modulation selon les phases de croissance [169]. La fonction physiologique de tels assemblages était inconnue, et le chapitre suivant présente des résultats proposant un rôle dans la compaction de l'ADN.



**Figure 33 : Modèles d'auto-assemblage de Hfq entière.** A. Formation d'un dodécamère avec imbrication des queues C-terminales de chaque hexamère, d'après [169]. B. Formation d'un réseau protéique d'hexamères, avec possiblement des structurations en feuillets aux contacts CTR-CTR entre les protéines. Bleu : NTR ; vert : CTR. Les dimensions ne sont pas à l'échelle.

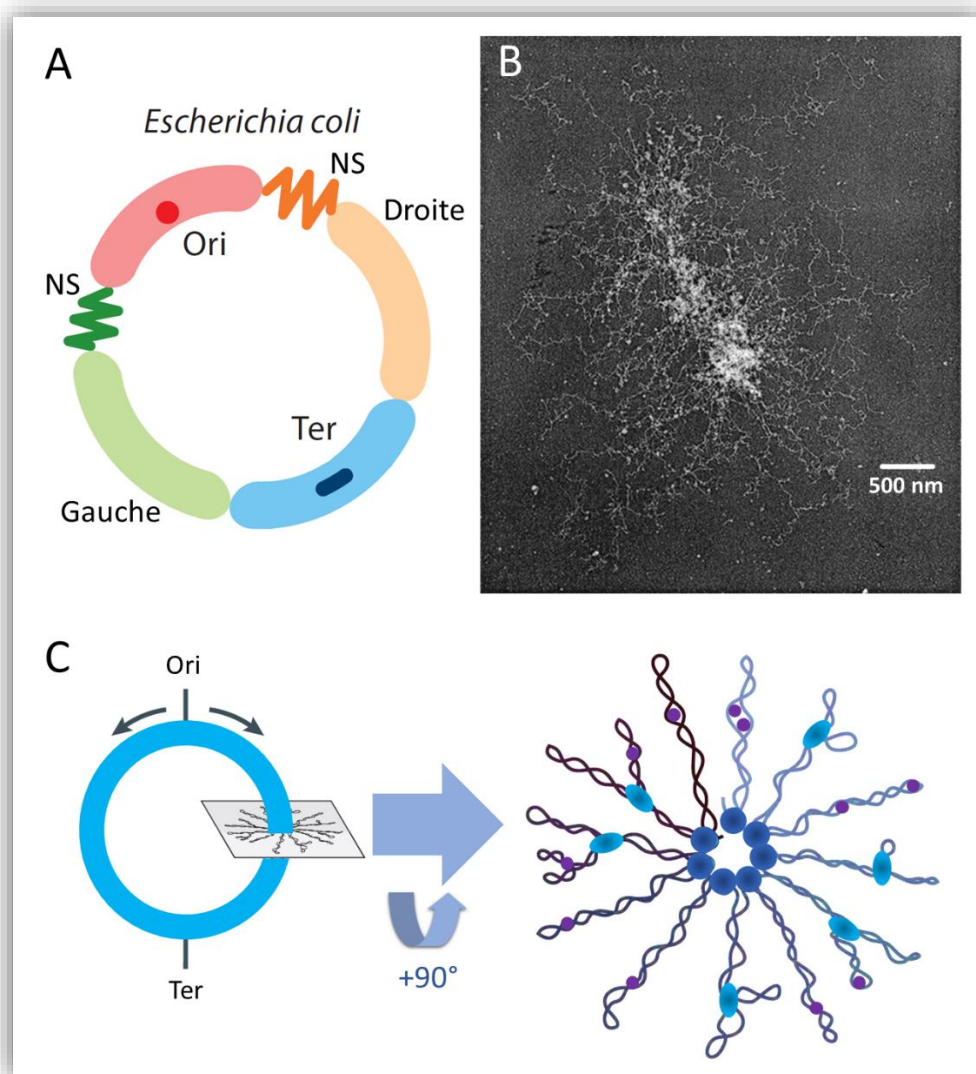
Chapitre III :  
Interaction de la  
protéine Hfq avec  
l'ADN et  
contribution de  
son domaine  
carboxyl terminal.

# 1 Contextualisation : compaction de l'ADN et organisation spatiale du génome chez *E. coli*.

L'ensemble des organismes cellulaires, procaryotes et eucaryotes, doivent dramatiquement condenser leur ADN génomique afin de le loger dans des cellules mesurant moins d'un dix-millième de sa longueur. Cependant, l'ADN doit rester accessible pour les processus essentiels tels la réplication, la réparation et la transcription. Chez les eucaryotes, l'ADN génomique est enroulé autour de protéines appelées histones qui forment des nucléosomes, associés ensuite dans des structures de plus grand ordre, le tout étant confiné dans un noyau. Les bactéries, elles, n'ont pas d'histones à proprement parler mais possèdent d'autres mécanismes pour condenser et organiser leur chromosome. Ces mécanismes incluent le surenroulement de l'ADN, l'encombrement macromoléculaire et enfin les petites Protéines Associées au Nucléoïde (NAPs). Les NAPs jouent un rôle crucial non seulement dans le maintien de la compaction du génome, mais aussi dans sa réorganisation dynamique afin de le rendre accessible pour l'expression génique.

## 1.1 Les domaines topologiques.

En se liant de façon spécifique ou non à l'ADN, les NAPs et autres complexes protéiques contribuent à la formation de domaines topologiques isolés appelés micro-domaines, d'une taille avoisinant les 10 kilobases de long, ce qui fait au total environ 400 micro-domaines par chromosome [170]. Leur existence est transitoire et l'ensemble du système est dynamique, impacté par les conditions de culture et la présence de divers complexes protéines : polymérase, gyrase, topoisomérase et NAPs influent sur le surenroulement et sa propagation ou son isolement. Ces micro-domaines sont ensuite organisés en plus grandes régions appelées macro-domaines, qui eux font environ une mégabase de long et sont au nombre de quatre (**Figure 34**) [8].



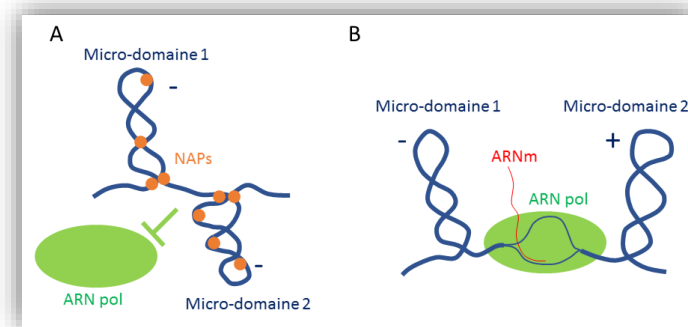
**Figure 34 : Le chromosome bactérien est organisé en sous-domaines topologiques regroupant des *loci* spatialement proches et pouvant interagir.**

**A.** Chromosome de *E. coli* schématisé découpé en quatre macro-domaines : Origine de réplication, Terminateur de réplication, Gauche et Droite. NS : Non Structuré. Tiré et modifié de [8] **B.** Chromosome de *E. coli* visualisé par microscopie électronique, obtenu par lyse douce et purifié via gradient de saccharose. Les micro-domaines surenroulés et couverts de protéines sont visibles sous formes de boucles émergeant d'un « cœur » protéique. Tiré de [170]. **C.** Vue schématique des micro-domaines, associés à des NAPs. Les segments consécutifs d'ADN chromosomique sont repliés en micro-domaines qui sont empilés les uns sur les autres et forment la superstructure circulaire du chromosome.

## 1.2 Les Protéines Associées au Nucléoïde (NAPs) ou « protéines architectes ».

Auparavant appelées « protéines *histone-like* » par analogie avec les histones eucaryotes, il est apparu que ces protéines remplissent des fonctions plus étendues et plus variées. Elles sont maintenant collectivement regroupées sous le terme de Protéines Associées au Nucléoïde, ce qui reflète plus précisément leur localisation cellulaire sans impliquer à tort qu'elles auraient une homologie structurale avec les histones. Les NAPs sont nombreuses, diverses et leurs mécanismes d'action ne sont pas tous compris. Elles possèdent la capacité de se lier à l'ADN et d'affecter sa trajectoire, c'est-à-dire la direction qu'il prend dans l'espace, en l'enroulant, le pontant, le rigidifiant ou le pliant. Par extension, beaucoup de NAPs sont capables d'influencer la transcription de manière positive ou négative, en empêchant ou en favorisant l'accès de l'ARN polymérase à la séquence codante (**Figure 35**). L'ouverture des boucles de transcription et de réplication est conditionnée au taux de surenroulement du chromosome et l'action des NAPs est donc cruciale pour le préserver [171].

De plus, les NAPs doivent contraindre l'ADN sous forme condensée. En effet, si de telles barrières topologiques n'existaient pas, une cassure simple ou double de l'ADN provoquerait un relâchement dramatique du surenroulement du chromosome, augmentant sa taille apparente de façon incontrôlable pour la bactérie [172].



**Figure 35 : Influence de la topologie du chromosome bactérien sur la transcription... et vice-versa. A.** Les domaines topologiques surenroulés et contraints par les NAPs peuvent empêcher l'accès de l'ARN polymérase aux gènes. **B.** La bulle de transcription génère un surenroulement négatif à l'arrière et un positif à l'avant de l'ARN polymérase, modifiant la topologie de l'ADN.

Inspiré de [173].

### 1.2.1 Protéines architectes : affinité, abondance.

Les affinités des NAPs pour l'ADN sont variables, avec une constante de dissociation  $K_d$  allant de 25 à 250 nM en fonction de la nature de l'ADN et de la protéine. Le classement d'une partie des NAPs par affinité décroissante pour l'ADN est le suivant : HU > IHF > Lrp > CbpB > Fis > H-NS > StpA > CbpA > IciA > Hfq/Dps [129]. Comme les contraintes stériques et dynamiques sont dépendantes de l'environnement et de la phase de croissance, les différentes NAPs ne sont pas exprimées de façon constante, permettant de contrôler finement l'expression spatio-temporelle des différentes portions du génome (**Figure 36**) [129], [174]. Ainsi, alors qu'en phase exponentielle l'ordre d'abondance des NAPs majeures est Fis > HU > H-NS > IHF, l'entrée en phase stationnaire rebat les cartes et il devient Dps > IHF > HU > H-NS > Fis, ce changement étant aussi directement visible sur la forme du nucléoïde qui devient plus compact [175].

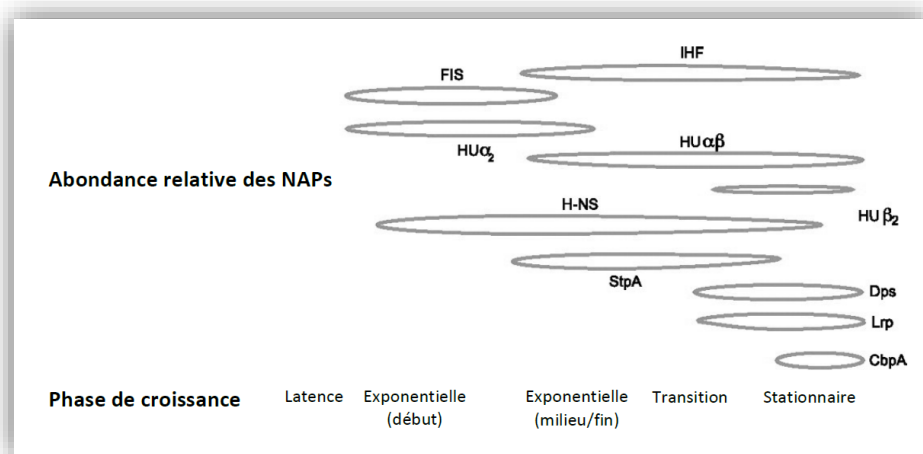
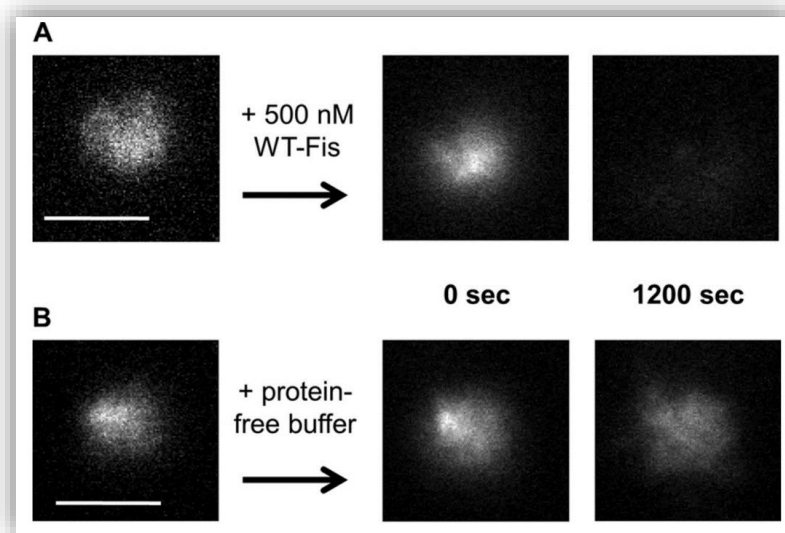


Figure 36 : Expression de différentes NAPs selon les phases de culture.  
Adapté de [174] et reproduit de [176].

Ces affinités et abondances relatives des NAPs sont à nuancer par le mode de liaison et de libération de ces protéines avec l'ADN. En effet, en ce qui concerne les interactions non spécifiques, c'est-à-dire sans séquence d'ADN cible particulière, il existe une cinétique de dissociation facilitée, mise en évidence par des expériences de pinces magnétiques *in vitro* sur Fis et HU notamment. Ce résultat a été reproduit directement sur des chromosomes extrait de cellules de *E. coli* [177], [178]. Alors que les cinétiques



d'association sont généralement bien caractérisées et résultent en des complexes protéine-ADN stables, la dissociation des NAPs est souvent considérée comme étant une propriété intrinsèque de la nature du complexe, indépendante du reste du système. Or, ces travaux montrent que la cinétique de dissociation dépend de la quantité de protéine libre, avec laquelle il existe un échange dynamique : plus la fraction protéique en solution est grande, plus l'échange est rapide et important (**Figure 37**). Plus encore, la présence de compétiteurs en solution (autres NAPs) intensifie le phénomène. L'échange augmente linéairement avec la concentration d'espèces libres en solution, jusqu'à atteindre un plateau. Ce résultat expérimental a ensuite été reproduit par simulation, attribuant le phénomène à des états intermédiaires entre l'état « libre » et l'état « lié » des NAPs, porte d'entrée à la venue d'un troisième joueur facilitant la dissociation des deux premiers partenaires [179]. Il faut noter que l'inverse est aussi constaté : Fis est capable d'échanger des brins d'ADN de façon croissante avec la concentration d'ADN libre, même si cette situation est moins susceptible de se produire *in vivo* [180].



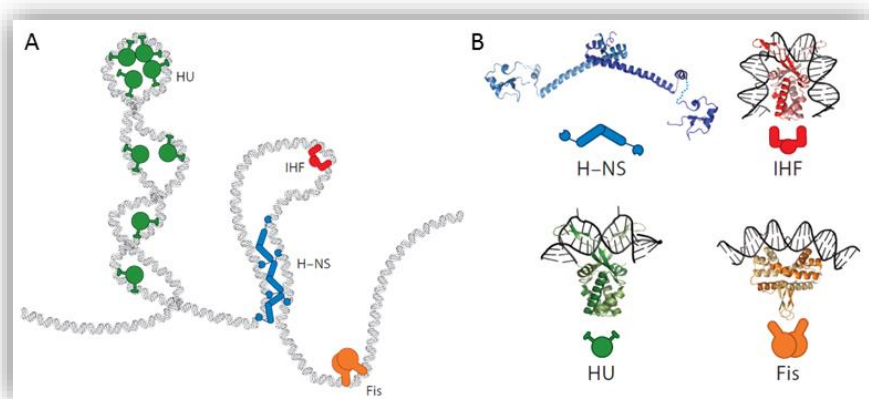
**Figure 37 : Dissociation facilitée de Fis du chromosome bactérien observée *in vitro*.** **A.** Le chromosome isolé d'une souche exprimant un variant de Fis étiqueté avec une GFP (*Green Fluorescent Protein*) est extrait et observé en microscopie de fluorescence. Puis, la solution est échangée avec un tampon contenant de la protéine purifiée native. Le nucléoïde est rincé avant de prendre une première image, puis une seconde 20 min plus tard. La fluorescence décroît, démontrant que le Fis sauvage s'est échangé avec le Fis marqué. **B.** Même expérience mais sans ajouter de Fis sauvage. La fluorescence décroît très peu : le complexe ADN-

Fis est stable si la solution ne contient pas de protéine libre. Il s'agit aussi d'un contrôle montrant que les rinçages n'affectent pas la liaison Fis-ADN, et que l'effet du photoblanchiment est négligeable. Tiré de [178].

Ainsi, les protéines en solution peuvent jouer un rôle clé dans la dissociation et le remplacement des protéines liées à la double hélice, un effet qui pourrait être déterminant pour permettre le turnover des NAPs liées au chromosome *in vivo* et ainsi contrôler la dynamique de la régulation génique.

### 1.2.2 Quelques exemples de NAPs.

Concernant les différences fonctionnelles de chacune des NAPs, ce type d'étude est rendu difficile par le fait que ce sont des régulateurs pléiotropiques pour la plupart : il est donc ardu de séparer les phénotypes directs et indirects observés lors d'une délétion. On peut néanmoins classer les NAPs en fonction du type de contrainte qu'elles exercent lors de leur fixation sur l'ADN : torsion, pontage, enroulement (**Figure 38A**). Les structures et fonctions de quelques protéines représentatives sont détaillées ci-après.



**Figure 38 : Quelques Protéines Associées au Nucléotide parmi les plus étudiées et les plus abondantes dans *E. coli*.** A. Schéma présentant les effets des NAPs sur l'ADN : les protéines dimériques HU, Fis et IHF induisent des courbures plus ou moins fortes au brin d'ADN, alors que H-NS recouvre et pontage des brins d'ADN en polymérisant le long du filament. B. Structures PDB des différentes protéines : 1P78 pour HU, 1IHF pour IHF, 3JRA pour Fis et une combinaison de 3NR7 + 1HNR pour H-NS. Reproduit et modifié depuis [8].

IHF (*Integration Host Factor*) a d'abord été identifiée pour son rôle dans la réplication du phage  $\lambda$  [181]. Elle est fortement présente, entre 12 000 et 60 000 copies par cellule en phase exponentielle ou stationnaire, respectivement [129]. C'est un hétéro-dimère formé de deux structures semblables, codées par les gènes *ihfA* et *ihfB*, qui se replient pour donner une molécule unique avec deux longs feuilletts  $\beta$  responsables de sa fixation à l'ADN. Elle possède 40% d'homologie de séquence avec HU. La structure cristalline d'IHF sur l'ADN suggère que ces bras s'insèrent dans le petit sillon de la double-hélice, et qu'une proline s'intercale entre les paires de bases. La fixation résulte en une courbure drastique du brin d'ADN, supérieure à  $120^\circ$  et jusqu'à  $180^\circ$ , ce qui facilite la formation de boucles (**Figure 38B**). L'interaction est par ailleurs inhibée par de hautes concentrations en KCl, alors qu'au contraire de faibles concentrations en KCl et la présence de  $Mg^{2+}$  favorisent non seulement la torsion mais aussi l'agrégation des brins d'ADN [182]. La fixation peut se faire soit sur un site spécifique d'une quinzaine de paires de bases avec un  $K_d$  de 0,3 à 2,3 nM, soit aléatoirement sur environ 35 paires de bases avec un  $K_d$  de 2  $\mu$ M [183], [184].

HU (*Heat-Unstable nucleoid protein*) possède une séquence fortement basique qui présente une partie homologue aux histones eucaryotes. Composée de deux sous-unités  $\alpha$  et  $\beta$  codées par *hupA* et *hupB*, on trouve la protéine sous les trois combinaisons possibles : hétéro-dimérique  $\alpha\beta$  et homo-dimérique  $\alpha_2$  ou  $\beta_2$ , cette dernière ayant une affinité moins grande pour l'ADN que les autres. Ainsi on passe de la forme  $\alpha_2$  en phase exponentielle à la forme  $\beta_2$  en phase stationnaire avec une transition entre les deux *via* les hétéro-dimères (**Figure 36**). Sa concentration atteint 30 000 copies par cellule en phase exponentielle [129]. Structurellement, la fixation de l'ADN se fait de façon analogue à IHF, avec une proline dans le petit sillon, mais n'interagit pas préférentiellement avec une séquence particulière (**Figure 38B**). Le site de fixation mesure seulement 9 paires de bases, et HU se fixe plus facilement à des portions d'ADN déformées, stabilisant des courbures de  $105^\circ$  à  $140^\circ$  [129]. De plus, contrairement à IHF, la fixation de HU est coopérative et la protéine peut former un filament rigide s'enroulant sur l'ADN à de fortes concentrations, augmentant sa longueur de persistance [185]. Un recouvrement partiel du double-brin est aussi observé à basse concentration mais nécessite aussi une faible concentration en sels, montrant

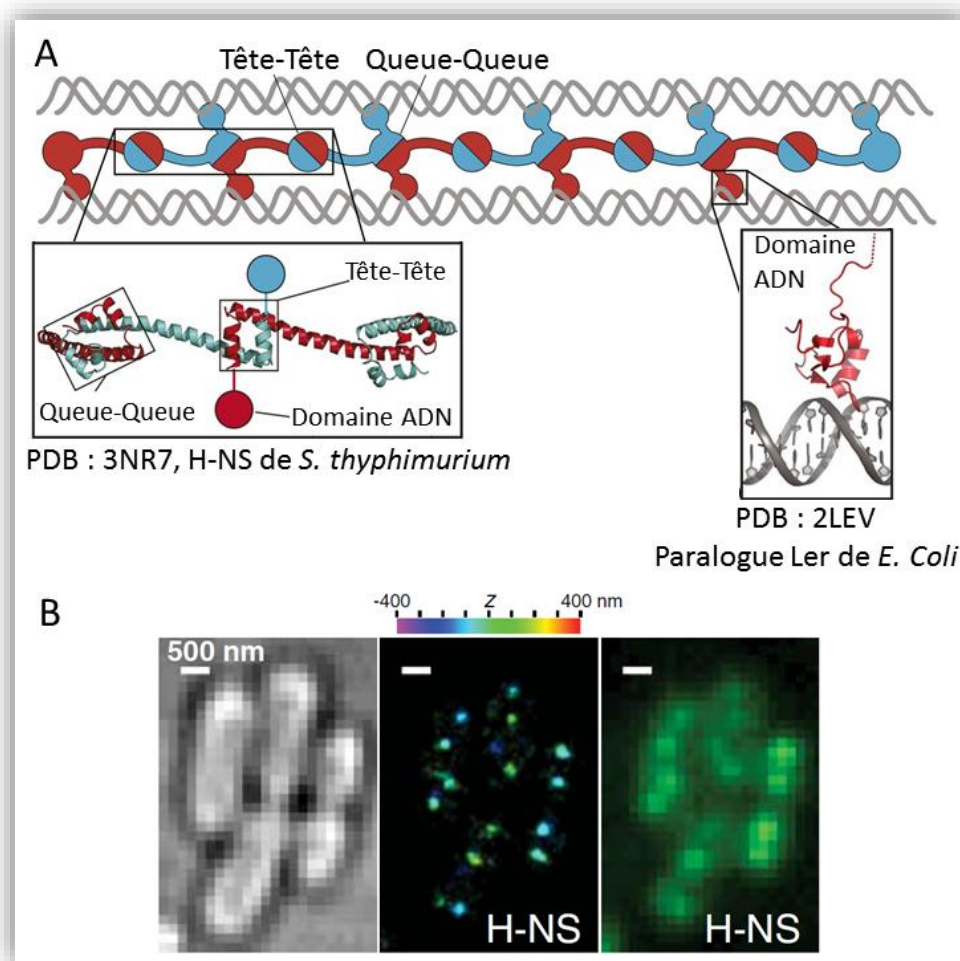
que la fixation coopérative est d'origine électrostatique. A concentration physiologique maximale, on estime à 10% la proportion du chromosome recouverte par HU [8]. Les mutants où *hupA* et *hupB* sont supprimés ont un nucléoïde plus étendu et un surenroulement moins important que la normale, démontrant le rôle de HU dans la condensation globale du génome [186]. *A contrario*, un mutant augmentant l'affinité de HU pour l'ADN possède un nucléoïde sur-condensé [187].

**Fis** (*Factor for Inversion Stimulation*) est fortement exprimée en début de phase exponentielle où on atteint 50 000 copies par cellule, puis sa concentration décroît progressivement. Structuellement, c'est un homodimère qui se fixe à une séquence consensus de 17 paires de bases riche en AT surtout retrouvée en dehors des séquences codantes et dans l'ensemble du génome, affectant la réplication et la transcription de nombreux gènes [188]. La fixation est assurée par deux bras hélice-tour-hélice se glissant dans le petit sillon et résulte en une courbure modérée du brin d'ADN d'environ 65° (**Figure 38B**), le complexe formé étant très stable [189]. En plus de ces effets directs, Fis réprime aussi la transcription des gyrases et par conséquent impacte indirectement le surenroulement du chromosome [190].

**Dps** (*DNA-binding protein from starved cells*) est comme son nom l'indique exprimée en phase stationnaire seulement. Cette protéine de type ferritine s'oligomérisse en dodécamères et vient se fixer à l'ADN en cas de stress nutritionnel ou oxydatif [191]. Elle protège l'ADN de l'attaque des UV, des radicaux libres, des irradiations ou encore du cuivre et du fer. Chargée négativement sur presque toute sa surface, le dodécamère se fixe à l'ADN via les lysines de ses domaines N-terminaux d'une façon pH-dépendante. La double hélice est compactée et recouverte de nombreux dodécamères ayant un effet protecteur, signifiant aussi qu'il existe des interactions entre les dodécamères [192]. Cette fixation coopérative, de façon intéressante, est extrêmement rapide. Lors de gels retards effectués avec Dps et de l'ADN, des complexes de grande taille sont formés et retenus dans les puits, et très peu de complexes de tailles intermédiaires sont détectés. Des images AFM permettent de confirmer l'existence de grands complexes nucléoprotéiques [193].

**H-NS** (*Histone-like Nucleoid Structuring protein*) se fixe de façon non-spécifique à des centaines de sites à travers le chromosome, avec une

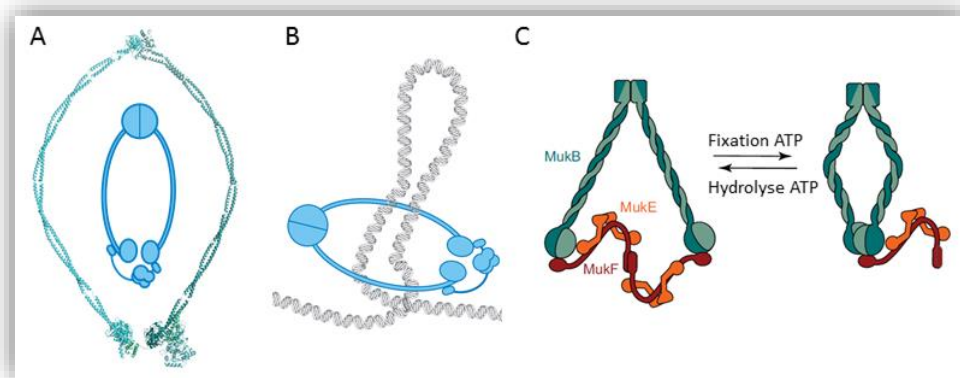
préférence pour les régions courbes et/ou riches en AT, mais ne déforme pas son site d'interaction. En revanche, sa fixation réduit la barrière énergétique pour qu'une autre protéine vienne se fixer à proximité et la protéine est capable de multimériser en formant un filament rigide sur l'ADN. Elle peut ponter différents brins d'ADN entre eux, amenant des *loci* éloignés sur la séquence primaire de l'ADN en proximité physique les uns des autres (**Figure 39**) [194]. Structuellement, H-NS est un homo-dimère dont chaque sous-unité possède un domaine d'interaction avec l'ADN au niveau C-terminal et deux domaines (dits « tête » et « queue ») permettant l'oligomérisation au niveau N-terminal, reliés par une séquence flexible (**Figure 39A**) [8]. La multimérisation se fait grâce à une structuration *coiled-coil*, où deux hélices  $\alpha$  s'enroulent l'une autour de l'autre. De façon analogue à ce qui est observé pour HU, les bactéries mutantes n'exprimant pas H-NS ont un plus faible niveau de condensation du nucléoïde que les bactéries sauvages [195]. La protéine est particulièrement utile à la cellule lors de phases de croissance rapide où la quantité d'ADN par cellule est accrue [196]. Au niveau fonctionnel, H-NS agit comme un « silenceur » et réprime la transcription de plus de 35 gènes en couvrant leur promoteur et leur séquence codante, notamment ceux nécessitant le facteur de transcription RpoS qui sont séquestrés dans une boucle d'ADN [197], [198]. Ces filaments de H-NS co-localisent dans deux larges regroupements par chromosome, de part et d'autre du nucléoïde (**Figure 39B**). La protéine est plus fortement exprimée en phase exponentielle (20 000 copies par cellule, suffisant pour couvrir 14% du chromosome en filaments) qu'en phase stationnaire, où sa concentration baisse de 40% environ [199].



**Figure 39 : Les protéines H-NS filamentent et pontent l'ADN en se regroupant à l'échelle du chromosome. A.** Schéma présentant le pontage de deux brins d'ADN impliquant des contacts protéine-protéine entre les deux domaines d'oligomérisation (Tête et Queue) et d'autres contacts ADN-protéine au niveau du domaine d'interaction avec l'ADN. Adapté de [194]. **B.** H-NS se regroupe en deux agrégats par chromosome bactérien, pontant des domaines très éloignés de l'ADN. Gauche : microscopie en fond clair. Milieu : image en super-résolution 3D STORM de H-NS fusionnée à la protéine fluorescente mEos2. Droite : image en microscopie de fluorescence conventionnelle. *From W. Wang, G.-W. Li, C. Chen, X. S. Xie, and X. Zhuang, 'Chromosome organization by a nucleoid-associated protein in live bacteria', Science, vol. 333, no. 6048, pp. 1445–1449, Sep. 2011. [199] Reprinted with permission from AAAS.*

Le **SMC** (*Structural Maintenance of Chromosomes*) est un assemblage protéique complexe en forme de pince, homologue de la condensine eucaryote.

D'un poids moléculaire d'environ 125 kDa, il est bien plus grand que les autres NAPs. Chez *E. coli*, il est composé de trois protéines, MukB, MukE et MukF. MukB, la plus grande, est un long bras d'hélices  $\alpha$  comprenant un domaine dit « charnière » à une extrémité, et un domaine ATPase à l'autre. Elle s'homodimérise et forme un grand anneau capable d'entourer des brins d'ADN (**Figure 40 A et B**). L'association de MukE et MukF permet la régulation de l'activité ATPase et le contrôle de la fermeture ou de l'ouverture de l'anneau (**Figure 40C**) [200]. Des mutations du SMC provoquent un relâchement de la condensation du nucléoïde, et chez *E. coli* particulièrement, un niveau de surenroulement du chromosome anormal [201]. L'assemblage permet donc de ponter et rapprocher des *loci* spatialement éloignés. De façon notable, le système SMC s'associe avec les régions proches de l'origine de réplication du chromosome et est requis pour le bon positionnement des deux chromosomes pendant la réplication [202].



**Figure 40 : Le système MukBEF chez *E. coli* permet de rapprocher des régions chromosomiques éloignées.** A. Structure de l'ensemble protéique d'un assemblage d'homologues bactériens de MukBEF, dérivé des structures PDB 4I98, 4I99 et 3ZGX. B. Schéma représentant le SMC pontant deux brins d'ADN. C. Mécanisme d'hydrolyse de l'ATP permettant le passage de la position ouverte à la position fermée. A et B reproduits d'après [8], C reproduit d'après [200].

### 1.2.3 Rôle dans la réparation de l'ADN.

Une autre implication fonctionnelle de certaines NAPs concerne la réparation mutagène des cassures de l'ADN. Lors d'une cassure double brin, la bactérie réagit en faisant appel au système SOS et/ou au circuit général du stress via le facteur de transcription RpoS. Ces systèmes vont exprimer des polymérases permissives afin de provoquer des mutations et

possiblement l'émergence d'une adaptation au stress rencontré. Or, il a été montré récemment que les NAPs H-NS et Fis favorisent la réparation mutagène des cassures en déclenchant le système SOS, alors que Dps la réprime [203]. HU pourrait aussi avoir une implication plus locale dans les réparations, de par son affinité plus forte pour les régions abimées de l'ADN [204].

Pour conclure, la grande variété de NAPs exerce des changements architecturaux tout aussi différents sur l'ADN : rigidification du filament, rapprochement de *loci* éloignés, pontage de brins entiers, courbures et formation de boucles, condensation... Ceux-ci ont un fort impact sur le déroulement de la vie cellulaire, en conservant le surenroulement du chromosome, en influençant la transcription des gènes, en protégeant l'ADN d'agressions physico-chimiques, et en répartissant spatialement les chromosomes naissants en vue de la division cellulaire.



## 2 Article D : *Effects of Hfq on the conformation and compaction of DNA.*

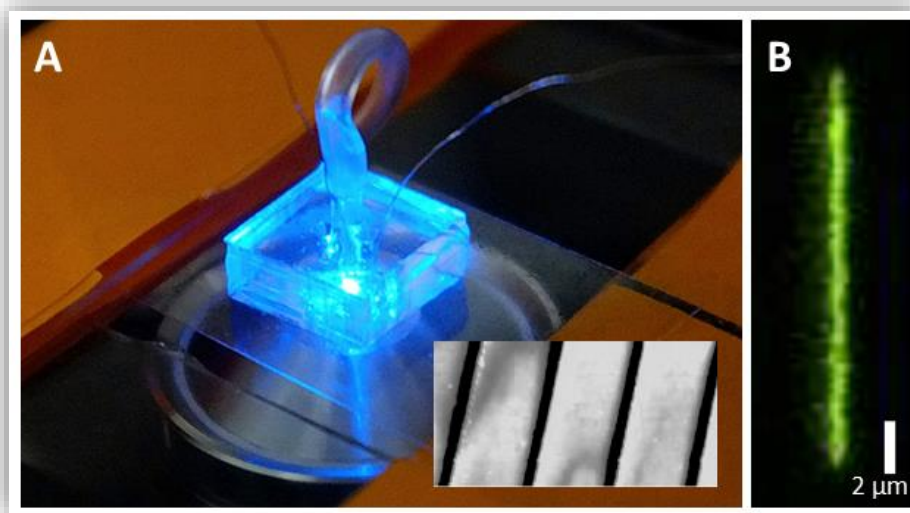
« Effets de Hfq sur la conformation et la compaction de l'ADN. »

Publié dans *Nucleic Acids Research*. Joint en **Annexe D**.

Kai Jiang, Ce Zhang, Durgarao Guttula, Fan Liu, Jeroen Van Kan, Christophe Lavelle, Krzysztof Kubiak, Antoine Malabirade, Alain Lapp, Véronique Arluison, Johan Van Der Maarel.

Cet article étudie la compaction de l'ADN en présence de Hfq.

Un dispositif original de nanofluidique est utilisé [205]. Celui-ci est composé de canaux de section rectangulaire de quelques centaines de nanomètres de côté et plusieurs dizaines de microns de long, faits d'un polymère, le PDMS (polydiméthylsiloxane) (voir **Figure 41**). Un ADN de phage T4 isolé est mis en contact avec des concentrations croissantes de Hfq dans divers tampons. Après incubation, une sonde fluorescente permettant de marquer l'ADN est ajoutée au mélange, puis celui-ci est introduit par un champ électrique dans les nano-canaux.



**Figure 41 : Expériences de nanofluidique.** A. Dispositif de nanofluidique. Les nano-canaux visibles sur le zoom en bas à droite sont dans la puce en PDMS visible au centre, à laquelle sont reliés des réservoirs contenant le mélange à observer et les électrodes pour le faire entrer dans les canaux. B. Image typique d'un ADN de phage T4 dans un nano-canal, marqué avec une sonde fluorescente de type YOYO-1.

Une fois relaxé, l'ADN est imagé par microscopie de fluorescence et la longueur apparente de celui-ci dans le canal est mesurée. Les résultats montrent que cette longueur décroît progressivement puis, au-delà d'une concentration seuil de Hfq, décroît drastiquement et résulte en un ADN très condensé. Cette condensation est aussi observée lorsque le mélange ADN-protéines n'est pas contraint dans un canal, mais le seuil de concentration en protéines est bien plus élevé (un demi ordre de grandeur environ).

Une autre expérience d'AFM permet de voir qu'en présence de la protéine, les fragments d'ADN se recouvrent et sont plus épais. L'analyse de leur longueur apparente montre aussi qu'ils sont plus courts. Ceci démontre que Hfq couvre l'ADN et change sa conformation, en le pontant notamment.

Enfin, une dernière expérience utilise la diffusion de neutrons aux petits angles (DNPA). Les paramètres structuraux qui en sont déduits permettent d'avoir une idée de la façon dont la protéine interagit avec l'ADN : ils montrent que le tore n'est pas fermement attaché à la double hélice, mais légèrement éloigné de celle-ci. Grâce à ce résultat, nous avons donc supposé que la protéine pourrait être « accrochée » à l'ADN par l'intermédiaire des queues C-terminales, et que la compaction pourrait venir de cet accrochage. C'est précisément ce dernier résultat qui va être à l'origine des travaux exposés dans l'article suivant.

### **3 Article E : *Compaction and condensation of DNA mediated by the amyloidogenic C-terminal domain of Hfq.***

« Compaction et condensation de l'ADN induites par le domaine C-terminal amyloïde de Hfq. »

Publié dans *Nucleic Acids Research*. Joint en **Annexe E**.

Antoine Malabirade, Kai Jiang, Krzysztof Kubiak, Alvaro Diaz-Mendoza, Fan Liu, Jeroen A. van Kan, Jean-François Berret, Véronique Arluison, Johan R.C. van der Maarel.

Cet article, prolongement naturel du précédent, se concentre sur la contribution du domaine C-terminal de Hfq dans la condensation de l'ADN.

Le même dispositif de nanofluidique est utilisé, avec le même ADN. Mais cette fois, l'expérience est reproduite avec le NTR ou le CTR de Hfq et non la protéine complète. Lorsque c'est la partie N-terminale de la protéine formant le tore qui est mélangée à l'ADN, aucune condensation n'est observée, tout juste une faible diminution de la longueur apparente de l'ADN dans le canal. Au contraire, lorsque c'est le peptide synthétique C-terminal qui est incubé avec l'ADN, le même comportement que pour la protéine complète est observé : l'ADN est condensé au-delà d'une valeur seuil, que ce soit dans le cas d'un environnement contraint comme les nano-canaux ou bien en environnement libre. Le CTR est donc la région clé dans le processus de compaction de l'ADN par Hfq.

Les expériences d'AFM sont aussi reproduites avec le CTR. Même si les fragments d'ADN sont aussi plus courts et plus épais en présence du peptide, c'est le cas pour des concentrations bien supérieures que dans le cas de la protéine entière.

Une analyse complémentaire est effectuée par gel retard. De l'ADN est incubé en présence de protéine entière, du NTR ou du CTR. Dans les trois cas, on observe un ou des retards du(s) à la fixation de la protéine sur l'ADN. Mais le comportement est très différent selon les cas : là où le NTR montre un retard modéré mais qui opère à faible concentration en protéine (haute affinité), le CTR piège carrément l'ADN dans les puits, formant de gros complexes qui ne peuvent pas pénétrer dans le gel. Cependant, ce comportement est observé pour de fortes concentrations uniquement (plus faible affinité). Dans le cas de la protéine native, on observe une combinaison des deux effets, avec un premier type de retard comparable au NTR puis la

formation d'agrégats dans les puits à plus forte concentration. C'est donc bien le CTR qui permet de ponter et agréger plusieurs fragments d'ADN entre eux, formant ces agrégats.

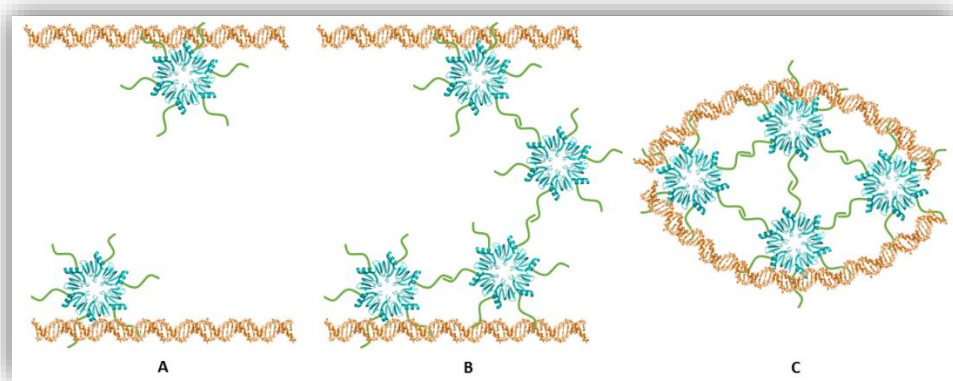
Enfin, une caractérisation thermodynamique de l'interaction CTR-ADN est faite par ITC (*Isothermal Titration Calorimetry*, Titration Calorimétrie Isotherme). Celle-ci est bien confirmée, et il s'avère qu'elle est principalement de nature enthalpique et exothermique.

L'ensemble des résultats permet de renforcer l'hypothèse de départ : le CTR de Hfq est impliqué dans la compaction et la condensation de l'ADN observée précédemment.

## 4 Discussion.

### 4.1 Mécanisme de la condensation par Hfq.

Ainsi, Hfq est capable de condenser et compacter fortement l'ADN, et le domaine C-terminal est essentiel au processus. Ceci est à relier avec le fait que cette région s'auto-assemble en plus d'interagir avec l'ADN, comme vu au chapitre précédent. Cependant, l'affinité de la protéine pour l'ADN vient principalement de son NTR : on peut donc supposer un mécanisme complexe où le NTR permet la fixation à l'ADN, et le CTR d'une part renforce cette fixation et d'autre part permet de ponter différents *loci* en interagissant avec d'autres CTR ou NTR, finissant par compacter l'ADN. Cela est confirmé par les résultats des gels retard, où l'affinité de la protéine entière est plus importante que le NTR seul. Ces gels mettent aussi en évidence la coopérativité apportée par le CTR. Elle permet une fixation facilitée de Hfq lorsqu'il existe déjà un complexe Hfq-ADN, de manière comparable à ce qui est constaté pour H-NS.



**Figure 42 : Modèle illustratif du mécanisme de compaction de l'ADN par Hfq.** A. Hfq se fixe à l'ADN à travers son tore, et s'ancre avec ses queues C-terminales, modifiant probablement la structure de la double hélice. B. La fixation d'autres protéines à l'ADN est facilitée par coopérativité, à travers les interactions CTR-CTR, possiblement sous forme d'assemblages  $\beta$  intermoléculaires. C. Si la concentration seuil en Hfq est atteinte, alors l'ADN est condensé par le réseau protéique. Bleu : structure du tore selon l'entrée PDB 1HK9. Vert : CTR (pas à l'échelle). Orange : ADN double-brin d'après l'entrée PDB 1D29.

Les autres questions soulevées concernent le mécanisme d'interaction et de compaction en lui-même. La séquence des événements est inconnue (formation d'un réseau de Hfq, puis fixation à l'ADN, ou bien le contraire). La titration calorimétrique permet d'avancer que l'interaction ADN-CTR est principalement enthalpique. Dans le cas d'une interaction électrostatique, il est constaté que la contribution entropique prédomine. Il semblerait donc que les liaisons soient de type hydrophobe ou polaire. On peut aussi imaginer que l'ADN pourrait servir de site de nucléation à l'assemblage amyloïde du CTR. Néanmoins, ce n'est qu'un indice et des expériences supplémentaires sont nécessaires pour répondre à d'autres questions structurales : Hfq est connue pour interagir préférentiellement avec les séquences riches en A-T, qu'en est-il de la contribution du CTR ? Est-ce que la fixation et le pontage effectué par Hfq et son CTR ont un effet sur la topologie de la double hélice ? Nos travaux actuels tentent de répondre à ces interrogations, en utilisant des analyses de topologie de l'ADN sur gel, des pinces magnétiques et de la PCR quantitative.

## 4.2 Comparaison avec d'autres macromolécules.

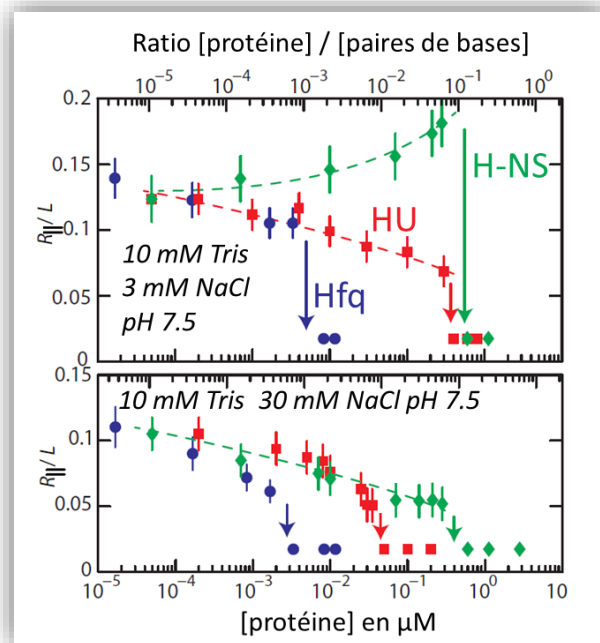


Figure 43 : Comparaison de la compaction d'un ADN de phage T4 contraint dans un nano-canal de 200 x 300 nm de section par Hfq, HU et H-NS. L'ADN

a été incubé à  $3 \text{ mg.L}^{-1}$  avec les protéines avant de l'introduire dans le canal. Les flèches symbolisent le seuil de compaction de l'ADN.  $R_{||}/L$  représente l'extension apparente mesurée de l'ADN divisée par la longueur totale de cet ADN s'il était totalement étendu (soit  $57 \mu\text{m}$  ici). Hfq est la protéine ayant la concentration seuil la plus basse. Lorsque la force ionique est faible (haut), on notera le fait que H-NS forme un filament sur l'ADN ayant pour effet de l'allonger.

De plus, il est intéressant de comparer ces valeurs seuils à celles obtenues pour d'autres types de macromolécules (**Tableau 1**). Dans un environnement contraint, même les agents d'encombrement qui ne se fixent pas particulièrement à l'ADN finissent par le condenser à forte concentration. C'est le cas des dextrans, ou bien de la BSA, qui condensent l'ADN de phage à environ  $70\text{-}80 \mu\text{M}$ , par un phénomène d'exclusion volumique. Les petites polyamines ou protéines riches en arginines comme la spermine ou la protamine, elles, ont un seuil de condensation bien plus faible. Le mécanisme de condensation est très différent, et elles servent de remplaçantes aux histones dans les cellules germinales, ou bien de stabilisateur de l'ADN. Néanmoins, encore une fois, Hfq montre une remarquable capacité de pontage, puisqu'elle est la protéine la plus efficace pour compacter l'ADN dans les mêmes conditions, comparable à la spermine. Cette efficacité peut compenser le fait qu'une faible proportion de Hfq est liée à l'ADN, et que la protéine est limitante de par son interaction avec les ARN.

Les NAPs sont présentes simultanément *in vivo* et on peut imaginer des synergies ou autres effets de leur action concertée : il serait intéressant de poursuivre cette étude directement dans la bactérie. Cependant, comme précisé au début dans la contextualisation de ce chapitre, l'effet de certaines NAPs sur l'ADN est difficile à observer au niveau cellulaire, car ces protéines sont souvent des régulateurs pléiotropiques. Hfq en est l'exemple même, avec sa myriade de mécanismes d'action et de partenaires. L'observation directe d'un effet d'une délétion de *hfq* sur le nucléoïde est donc loin d'être triviale : la protéine a par exemple un effet important dans la régulation de la division cellulaire via FtsZ, et la forme des nucléoïdes est forcément impactée. Ces expériences ont l'avantage de mettre en lumière le fort impact de la fixation de Hfq sur l'ADN en le séparant des autres effets indirects.

Tableau 1 : Comparaison des concentrations seuils ( $\mu\text{M}$ ) de différentes macromolécules à partir desquelles la condensation de l'ADN est observée. ADN de phage T4 à  $3 \text{ mg.L}^{-1}$ , nano-canal de 200 nm x 300 nm de section.

Macromolécule		Concentration seuil ( $\mu\text{M}$ )	
		10 mM Tris	10 mM Tris 3 mM NaCl
Agents d'encombrement	Dextrane (50kDa)	80	-
	Hémoglobine	80	-
	BSA	70	-
NAPs	H-NS	-	0,4
	HU	-	0,4
	<i>Hfq (entière)</i>	<i>0,03 ± 0,02</i>	<i>0,006 ± 0,003</i>
Ligands cationiques multivalents	Protamine	0,4 +/- 0,1	0,06 ± 0,02
	Co-hex	0,7 +/- 0,2	-
	Spermidine	0,4 ± 0,1	0,15 ± 0,05
	Spermine	0,045 ± 0,005	0,025 ± 0,005

Evidemment, la présence de l'ARN est aussi une donnée cruciale pour entrevoir l'effet de ce phénomène *in vivo*. Mais même si l'affinité de Hfq est bien plus grande pour l'ARN que pour l'ADN, une fraction non-négligeable de la protéine est retrouvée dans le nucléoïde. Reproduire les expériences en nano-canal en présence de compétiteurs ARN augmentera probablement la concentration seuil de compaction de l'ADN que nous observons, et il est de première importance de comprendre dans quelle mesure.

Une autre possibilité pour limiter les effets indirects produits par la délétion de Hfq *in vivo* serait d'utiliser un outil de dégradation contrôlée de la protéine, comme par exemple la protéase ClpXP [206]. La bactérie est amenée dans la phase de croissance souhaitée en présence de Hfq, puis on induit la production de la protéase qui dégrade la protéine marquée. On observe ensuite immédiatement l'effet de la disparition de Hfq sur le nucléoïde.



Chapitre IV :  
Agrégation de la  
protéine Hfq sur  
une membrane  
lipidique.

# 1 Contextualisation : composition des enveloppes bactériennes.

Contrairement aux cellules des organismes multicellulaires, la bactérie est en contact direct avec son environnement. Pour survivre, les bactéries se sont pourvues d'enveloppes sophistiquées, protectrices et sélectives, laissant passer les nutriments vers l'intérieur et les déchets vers l'extérieur.

Ces enveloppes sont constituées d'une structure à plusieurs couches servant à protéger l'organisme de l'environnement immédiat dans lequel il évolue (Figure 44B). La majorité des bactéries se répartit en deux grands groupes : les bactéries à Gram-négatif sont protégées par deux membranes séparées d'une fine couche de peptidoglycane, alors que les bactéries à Gram-positif ont une membrane unique entourée d'une couche de peptidoglycane bien plus épaisse, entremêlée avec de longs polymères anioniques. Intéressons-nous plus précisément à la structure de l'enveloppe de *E. coli*, bactérie à Gram-négatif.

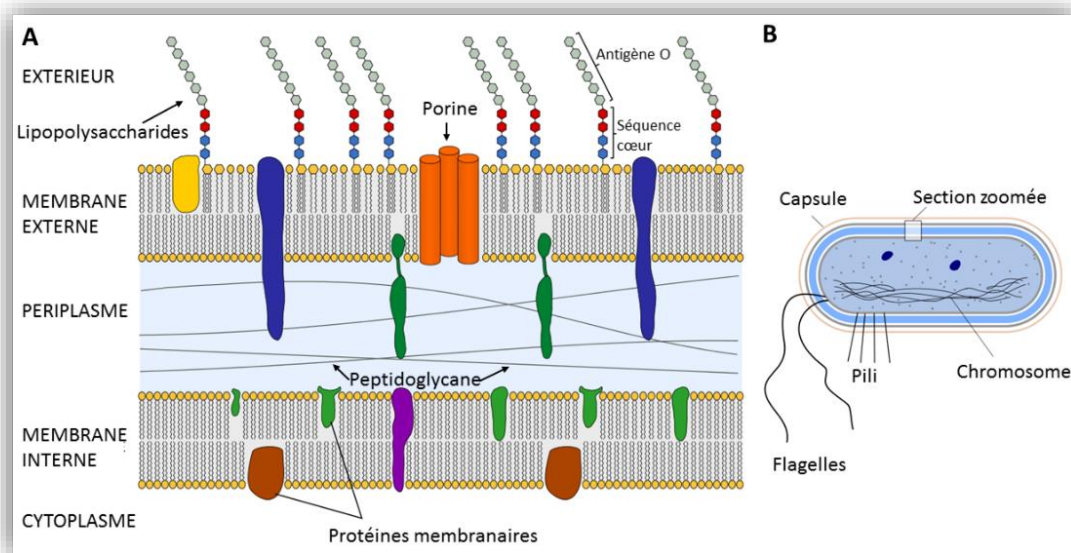


Figure 44 : Organisation de l'enveloppe bactérienne de *E. coli*. A. Coupe montrant la double membrane bactérienne de part et d'autre de la paroi cellulaire en peptidoglycane. B. Schéma de *E. coli* permettant de voir la section décrite en A. Modifié d'après [207].

## 1.1 La membrane externe.

En allant de l'extérieur vers l'intérieur, la première couche rencontrée est la membrane externe (*outer membrane*, OM), si l'on excepte une capsule qui peut être présente chez certaines souches [208]. Cette membrane externe est une caractéristique distinctive des bactéries à Gram-négatif, absente chez celles à Gram-positif. Comme les autres membranes biologiques, elle est composée d'une bicouche lipidique. Cependant, il ne s'agit pas d'une bicouche phospholipidique, mais d'une membrane asymétrique : les phospholipides sont présents mais uniquement dans le feuillet interne de la membrane. La couche externe, elle, est composée de glycolipides, principalement des lipopolysaccharides (LPS) (**Figure 44A**). Les LPS sont connus car ils sont responsables des chocs endotoxiques provoqués par les bactéries à Gram-négatif pathogènes : le système immunitaire est sensibilisé à ces molécules car elles sont un marqueur fiable d'infection. Les LPS jouent un rôle crucial dans la fonction de barrière de la membrane externe. Ils se collent fortement les uns aux autres, d'autant plus en présence de  $Mg^{2+}$ . Ce continuum est une barrière efficace contre les molécules hydrophobes. La membrane externe accueille aussi de nombreuses protéines, dont les porines. Celles-ci permettent le passage de beaucoup de biomolécules, et limitent le passage de celles supérieures à environ 700 Daltons [209]. Ces caractéristiques font de la membrane externe un outil puissant de la défense des bactéries, ce qui participe à la meilleure résistance des bactéries à Gram-négatif aux antibiotiques [210].

## 1.2 La paroi en peptidoglycane.

Les bactéries ne lysent pas lorsqu'elles sont immergées dans de l'eau distillée car elles possèdent un exosquelette rigide, capable de résister à la pression osmotique. Cet exosquelette est formé par le peptidoglycane, qui est constitué d'une répétition de disaccharides (N-actétyl-glucosamine - acide N-acétyl-muramique) liés par des chaînes pentapeptidiques [211]. La paroi en peptidoglycane, grâce à sa rigidité, détermine la forme de la bactérie. La membrane externe est accrochée à la paroi *via* des lipoprotéines comme Lpp ou la muréine (en bleu foncé sur la **Figure 44A**). Lpp est la protéine la plus abondante chez *E. coli*, avec plus de 500 000 copies par cellule [209].

De plus, d'autres protéines comme OmpA fixent le peptidoglycane à la membrane de façon non covalente (en vert foncé sur la **Figure 44A**).

### 1.3 Le périplasme.

La membrane externe et la membrane interne délimitent un compartiment cellulaire aqueux appelé le périplasme. Cet espace très dense en protéines est légèrement plus visqueux que le cytoplasme [212]. Il permet aux bactéries d'isoler certaines enzymes potentiellement dangereuses comme les RNases et les phosphatases alcalines.

### 1.4 La membrane interne.

Un des traits propres aux cellules eucaryotes est la présence d'organelles intracellulaires. Ceux-ci sont délimités par des membranes et assurent nombre de processus cellulaires essentiels : la mitochondrie participe à la production d'énergie, le réticulum endoplasmique lisse synthétise les lipides... *E. coli* ne possède pas ces organelles. Par conséquent, la totalité des fonctions associées aux membranes doit se faire au niveau de la membrane interne (*inner membrane*, IM) : production d'énergie via la chaîne respiratoire, biosynthèse des lipides, sécrétion des protéines. Cette membrane est composée d'une bicouche phospholipidique, décorée d'hélices protéiques et de divers complexes. Chez *E. coli*, elle est principalement composée de phosphatidyl-éthanolamine (PE) et de phosphatidyl-glycérol (PG), ainsi qu'une plus faible part de cardiolipine (CL) et des traces de phosphatidyl-sérine (PS). Les PG et CL sont chargés négativement, alors que les PE sont zwitterioniques (**Figure 45**).

Malgré les représentations fréquentes avec une tête doublement anionique, les CL ne portent qu'une charge négative car un proton est généralement piégé entre les deux phosphates et en neutralise un [10]. Les proportions de ces lipides varient en fonction des conditions de culture et de la phase de croissance, avec 75% de PE , 15-20% de PG et 5-10% de CL en moyenne (**Tableau 2**) [213], [214].

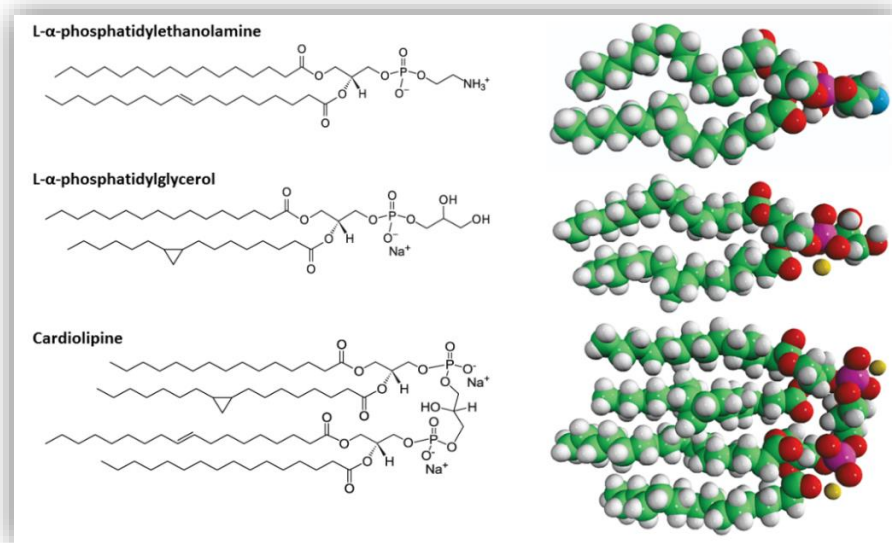


Figure 45 : Formule chimique développée et structure 3D des trois lipides les plus abondants dans la membrane interne de *E. coli*. Une seule formule est représentée à titre d'exemple. Code couleurs : Vert=C, blanc=H, rouge=O, violet=P, bleu=N, beige=Na. D'après les données de Avanti Polar Lipids, Inc.

Tableau 2 : Composition de *E. coli* en phospholipides pour différents taux de croissance et conditions de culture.  $\mu$  est le taux de croissance, ou vitesse de croissance spécifique, représentant la vitesse de croissance de la biomasse par unité de biomasse (nombre de cellules ou masse). Cette grandeur est définie par  $\frac{dN}{dt} = \mu \times N \Leftrightarrow N = N_0 \times e^{\mu t}$  où  $N$  est le nombre de bactéries au temps  $t$  et  $N_0$  est le nombre de bactéries à  $t_0$ . ND : Non Détecté. D'après [213].

Type de culture	Culture continue		Culture discontinue (fed-batch)	
	Bas (0,05 h <sup>-1</sup> )	Haut (0,6 h <sup>-1</sup> )	Bas (0,05 h <sup>-1</sup> )	Haut (0,6 h <sup>-1</sup> )
Phospholipides selon le type de tête (% mol)				
PE	70	78	77	77
PG	16	15	12	11
CL	8	7	12	12
PS	7	ND	ND	ND
Phospholipides selon le type de chaîne alkyle (% mol)				
C14:0	3	2	3	3
C16:0	53	52	44	44
C16:1	16	21	19	26
C18:1	8	14	12	18
C17:1	20	11	22	9

La fonction des différents lipides n'est pas entièrement comprise, mais les PG sont essentiels. Il semblerait que leur caractère acide soit nécessaire pour un certain nombre de processus cellulaires [215].

Les lipides insaturés, et notamment les lipides anioniques comme les PG, sont aussi utilisés par la bactérie pour rendre sa membrane plus fluide, facilitant la translocation et/ou la fixation de certaines protéines dont la surface est cationique [214]. Ils interagissent entre eux par un important réseau de liaisons hydrogènes entre les groupements hydroxyles des glycérols et les oxygènes non estérifiés des groupements phosphate [216].

La cardiolipine, elle, est essentielle pour le bon déroulement de la division cellulaire. On la retrouve en plus grande proportion au niveau des pôles de la cellule et du septum, car sa géométrie particulière avec quatre queues alkyles lui permet de former des membranes plus fortement concaves en réduisant leur énergie de tension (**Figure 46**) [15]. Là aussi, de nombreuses protéines interagissent préférentiellement avec les domaines formés par les cardiolipines. Cela influence leur localisation, comme dans le cas de DnaA, FtsY, GlpD, PssA, SecA ou MinD [10].

Enfin, les PE sont capables de former des microdomaines grâce à leur tête portant une charge positive et une charge négative, qui se lie aux têtes adjacentes.

La formation de ces différents domaines a été vérifiée *in vitro* en jouant sur la pression de surface appliquées aux membranes. Sous une variété de pressions, compatibles avec les conditions physiologiques, les lipides bactériens forment des domaines dans les membranes, ayant des propriétés physico-chimiques différentes comme leur rigidité ou leur température de transition de phase [217].

Ces domaines hétérogènes rebattent les cartes. En effet, le concept d'une membrane fluide et homogène postulé en 1972 propose un environnement comparable à un océan lipidique parsemé de quelques icebergs protéiques flottant dans deux dimensions (**Figure 47A**) [218]. Cependant, les preuves que les lipides sont séparés dans des régions particulières, appelées domaines, en fonction de leur composition chimique, s'accumulent. Et ce autant chez les eucaryotes que chez les procaryotes. Une représentation plus adéquate serait celle d'une mosaïque de petits groupes de CL, PE et PG s'assemblant d'après leurs caractéristiques physico-chimiques, possiblement maintenue ou favorisée par certaines protéines (**Figure 47B**) [219].

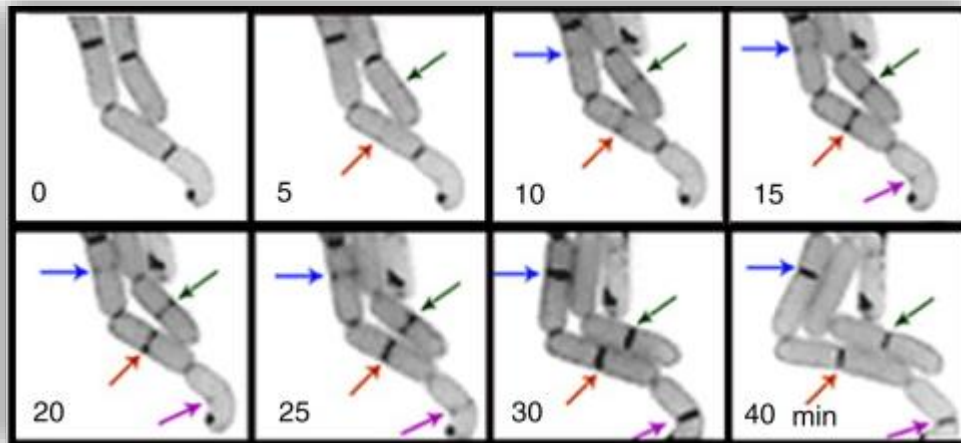


Figure 46 : Localisation des cardiolipines dans la membrane interne de *E. coli*. Les cardiolipines sont marquées par la sonde fluorescente NAO (10-N-nonyl-acridine orange). Les flèches pointent sur les domaines en formation au niveau des *septa*. Tiré de [15].

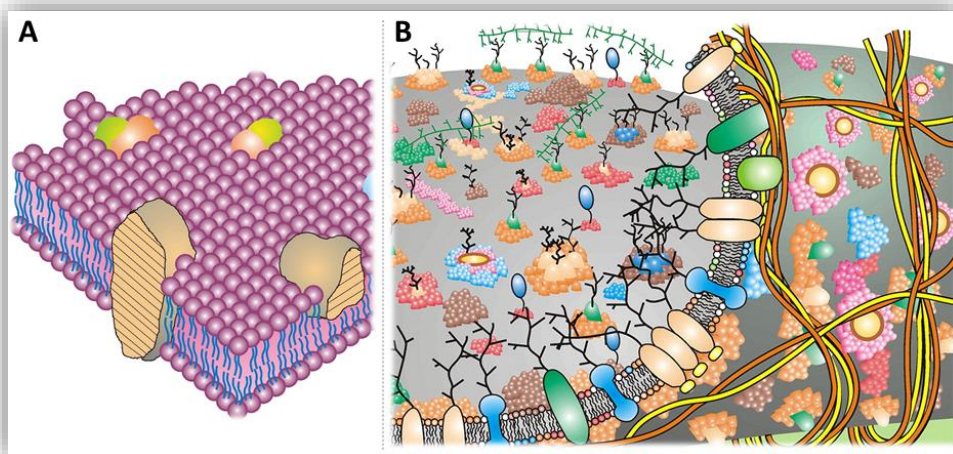


Figure 47 : Evolution de la représentation des membranes biologiques. A. Le modèle proposé en 1972 par Singer et Nicholson décrivait une membrane fluide et homogène dans laquelle des protéines diffusaient librement. B. Un modèle plus proche des connaissances accumulées ces 40 dernières années, représentant la complexité et l'hétérogénéité de la membrane plus proche d'une mosaïque. La bicouche est retournée sur la droite, révélant la couche intérieure décorée de différents domaines lipidiques, avec leurs protéines, ainsi que les filaments du cytosquelette. D'après [219].

## 2 Article F : *Membrane association of the bacterial riboregulator Hfq and functional perspectives.*

« Association membranaire du riborégulateur Hfq et perspectives fonctionnelles. »

Publié dans *Scientific Reports*. Joint en **Annexe F**.

Antoine Malabirade, Javier Morgado-Brajones, Sylvain Trépout, Frank Wien, Ileana Marquez, Jérôme Seguin, Sergio Marco, Marisela Velez, Véronique Arluison.

Cet article s'intéresse à l'association membranaire de Hfq, dont l'existence est corroborée par imagerie d'après des travaux précédents [61], [130]. Cependant, contrairement à ces études, il s'agit ici d'expériences *in vitro*. Les seules espèces en présence sont une membrane lipidique et la protéine Hfq sous différentes formes : entière, NTR seulement, CTR seulement. Les membranes modèles utilisées sont soit naturelles et de compositions comparables à la membrane interne de *E. coli* (*E. coli Polar Extract*, abrégé EPE), soit synthétiques et fortement polaires (DOPG) ou faiblement polaires (DOPC).

Une première série d'expériences permet de démontrer l'interaction entre Hfq et des liposomes par co-sédimentation. Que ce soit le NTR, le CTR ou la protéine entière, les trois sont retrouvés dans le culot après centrifugation et sédimentation des liposomes EPE.

Cette interaction est confirmée par imagerie : cryo-microscopie où l'on constate une « couche » de protéines sur la membrane des liposomes, et AFM liquide, où des « patchs » protéiques sont retrouvés sur une membrane supportée sur lame de mica. Lorsque l'incubation est prolongée, la membrane est affectée et des trous sont observés. Or, ce n'est pas le cas lorsque c'est le NTR qui est incubé avec les membranes : celui-ci n'interagit que faiblement. Cela indique donc que le CTR est responsable de ces agrégats et de la réorganisation des lipides.

Connaissant la propension du CTR à s'auto-assembler mais aussi la capacité de certains amyloïdes à interagir et modifier l'organisation des membranes lipidiques, nous avons effectué des tests afin de déterminer la zone minimale du CTR nécessaire à cet assemblage en fibres amyloïdes. Nous avons trouvé une séquence de 11 acides-aminés, appelée CTR<sub>11</sub>. Une fois cette zone déterminée, les expériences d'AFM et de TEM ont été répétées



avec un peptide synthétique ayant cette séquence. Ce peptide est capable de former des trous dans la membrane supportée en AFM, et déforme les liposomes en TEM.

De plus, l'incubation du CTR avec des liposomes induit un changement conformationnel de celui-ci, observé par dichroïsme circulaire : une partie des peptides adopte une conformation en hélice distordue au contact de la membrane.

Il semblerait donc que le CTR de Hfq soit responsable de l'agrégation et de l'ancrage de Hfq sur la membrane interne de *E. coli*.

### 3 Discussion.

Avec cette série d'expériences, l'hypothèse émise à la suite de l'observation des structures par immunofluorescence est renforcée : ces assemblages protéiques peuvent être en partie dus à l'interaction de Hfq avec la membrane interne de *E. coli*. Dans ce cas encore, le CTR semble jouer un rôle important : sans lui, cette interaction est trop faible pour être constatée par les techniques d'imagerie, même si elle est vue au niveau macroscopique par co-sédimentation. Là aussi, le CTR joue un rôle dans la formation des grands agrégats en permettant l'auto-assemblage des protéines sur la membrane.

Qualitativement, on constate une affinité bien plus grande de Hfq pour les têtes lipidiques polaires comme les PG, alors que la fixation est très faible sur les lipides non polaires comme les PC. L'affinité élevée pour l'extrait de lipides naturels de *E. coli*, qui contiennent une proportion importante de PG, permet d'émettre l'hypothèse raisonnable que la fixation de la protéine se fait principalement par des interactions polaires, au moins dans son étape déterminante. Cette affinité plus importante pour les PG et peut-être les CL pourrait aussi participer à l'hétérogénéité de la répartition de Hfq à la membrane, et à la formation des agrégats, puisque ces lipides sont organisés en domaines.

Lorsqu'on prend en compte l'effet du CTR, le système devient plus complexe. Alors qu'il semble stabiliser la protéine sur la membrane dans un premier temps, il affecte la structure de celle-ci pour des temps d'incubation plus longs. Plus particulièrement, c'est la séquence clé formant les fibres amyloïdes *in vitro* qui produit cet effet. Comme précisé dans la contextualisation de ce chapitre et dans l'article, les feuillettes  $\beta$  intermoléculaires retrouvés dans les amyloïdes sont connus pour déstabiliser les membranes lipidiques, et parfois y créer des pores. L'apparition d'une petite proportion d'hélice constatée par SRCD pourrait aussi expliquer cet ancrage et la déstabilisation de la membrane. Reproduire les expériences avec un mutant protéique purifié ne possédant pas la petite zone amyloïde permettrait de confirmer l'hypothèse de son rôle dans le processus d'agrégation. De plus, une étude plus approfondie permettant de mettre en évidence ces trous par le suivi de sondes fluorescentes par exemple, paraît tout indiquée pour compléter ce résultat.

Le rôle fonctionnel de cette agrégation et cette déstabilisation de la membrane reste mystérieux. L'étape d'agrégation sur la membrane, elle, pourrait simplement servir de stockage temporaire de la protéine. La bactérie pourrait adapter la quantité de chaperonne active en fonction des besoins. Une interaction électrostatique ou polaire, relativement faible, permettrait le relargage rapide et facile lors d'une augmentation du besoin en Hfq ou un changement des conditions physiologiques, régulant finement la disponibilité en protéine pour ses fonctions de riborégulation.

Comme il a récemment été montré que les ARN et notamment des sARN se retrouvent nombreux dans le milieu extracellulaire, une hypothèse séduisante pour expliquer la déstabilisation serait que Hfq permette d'exporter des ARN en leur faisant traverser la membrane interne [220]. Il faudrait alors tracer des ARN en présence de Hfq et en son absence afin de voir s'ils peuvent être excrétés ou non. De plus, la membrane interne n'est qu'une des barrières protégeant la bactérie, et ces ARN auraient encore à passer le peptidoglycane. Si Hfq était aussi exportée, elle pourrait protéger les ARN de la dégradation, mais d'un autre côté rendre plus difficile la traversée de la paroi. Une autre possibilité serait que l'export se fasse *via* les OMVs (*Outer Membrane Vesicles*), des petites vésicules exportées par la bactérie et contenant notamment des ARN. Nous travaillons sur une étude des ARN extracellulaires en la présence et en l'absence de Hfq pour progresser sur cette hypothèse.

Enfin, il sera aussi crucial d'évaluer l'interaction de Hfq avec une membrane en présence d'ARN. L'affinité de la protéine étant très forte pour certains sARN, le comportement observé sera probablement affecté. Notamment, le rôle du CTR dans le relargage de certains petits ARN pourrait s'avérer essentiel pour rendre la surface de la protéine accessible aux phospholipides. Reproduire le même type d'expériences que dans l'Article F avec des compétiteurs ARN serait déjà une première étape.

# Conclusion

Ces travaux ont permis d'attribuer de nouvelles fonctions possibles au domaine C-terminal de Hfq, en plus des résultats récents concernant son impact sur la fixation-libération des ARN.

→ **Le CTR permet le cyclage des ARN partenaires de Hfq.**

La quantité cellulaire de Hfq est limitante en regard de ses partenaires ARN. Pour assurer une régulation efficace, il faut donc que Hfq puisse interagir avec un grand nombre d'ARN, de façon successive, pour faciliter de nombreux appariements sARN-ARNm. Le CTR permet cela en jouant un rôle « d'écarteur » : par sa présence, la libération du complexe sARN-ARNm formé sur Hfq est facilitée et Hfq est rendue disponible pour une autre tâche. Il aide aussi à la discrimination entre les différentes paires d'ARN, notamment entre les sARN de classe I et de classe II qui se fixent différemment sur la protéine.

→ **Le CTR facilite et renforce la capacité de Hfq à s'auto-assembler.**

La capacité de Hfq à s'auto-assembler en structures de plus grand ordre était déjà connue. Nous démontrons ici que le CTR joue un rôle prépondérant dans cette structuration. Il est capable d'adopter une conformation en feuillets  $\beta$  inter-moléculaires et de former des fibres amyloïdes lorsqu'il est séparé du tore. Cela indique à *minima* une participation importante des interactions CTR-CTR dans la formation de ces super-structures, probablement par empilement des queues C-terminales en feuillets. C'est un exemple supplémentaire d'amyloïdes fonctionnels.

→ **Le CTR permet le pontage et la condensation de l'ADN bactérien.**

Hfq est une NAP remarquablement efficace quand il s'agit de compacter de l'ADN, et cela est principalement dû à son CTR. Celui-ci permet à la fois de renforcer la fixation de la protéine à la double hélice, mais surtout de ponter différents brins grâce aux interactions avec d'autres complexes ADN-Hfq existants et la formation d'un réseau.

Cette compaction a lieu même à des concentrations relativement basses (à partir de quelques nM) et pourrait vraisemblablement jouer un rôle régulateur comparable à celui d'autres NAPs dans le nucléoïde, même si seulement une faible proportion de Hfq y est localisée.

→ Le CTR facilite et renforce la fixation de Hfq sur la membrane interne de la bactérie.

Les expériences effectuées *in vitro* montrent que le CTR permet l'agrégation de la protéine à la surface d'une bicouche lipidique en adoptant une conformation particulière au contact des phospholipides, justifiant la localisation membranaire d'une fraction de la protéine constatée par imagerie. La fixation de Hfq se fait avec une préférence pour les lipides ayant une tête polaire. La formation des assemblages de Hfq à la membrane pourrait avoir un rôle de stockage de la protéine, ou bien n'être que la première étape d'un processus plus complexe.

En effet, lors de temps d'incubation prolongés, le CTR déstabilise la bicouche lipidique. Ce phénomène pourrait faciliter l'export d'ARN à travers la membrane interne, et être en partie responsable du complément en ARN retrouvé à l'extérieur des bactéries.

→ Cependant, le CTR n'est pas essentiel dans des conditions favorables de culture.

La souche MG1655 de *E. coli* délétée pour les 30 acides-aminés C-terminaux de Hfq ne montre pas de défection particulière de croissance lorsqu'elle est cultivée en conditions favorables, contrairement à lorsque c'est la totalité de la protéine qui est absente (milieu nutritif riche, température et agitation optimales). Néanmoins, les impacts physiologiques liés aux différentes fonctions du CTR pourraient exister dans d'autres environnements, notamment lors de réponses à divers stress pour lesquels la bonne fonction de Hfq est importante. Des effets ont déjà été constatés dans certaines riborégulations, qui ont lieu avec une efficacité réduite lors de l'absence du CTR. Il faudrait donc poursuivre les investigations dans des conditions de croissance particulières (phase stationnaire prolongée, limitation en fer, choc osmotique, choc thermique...).

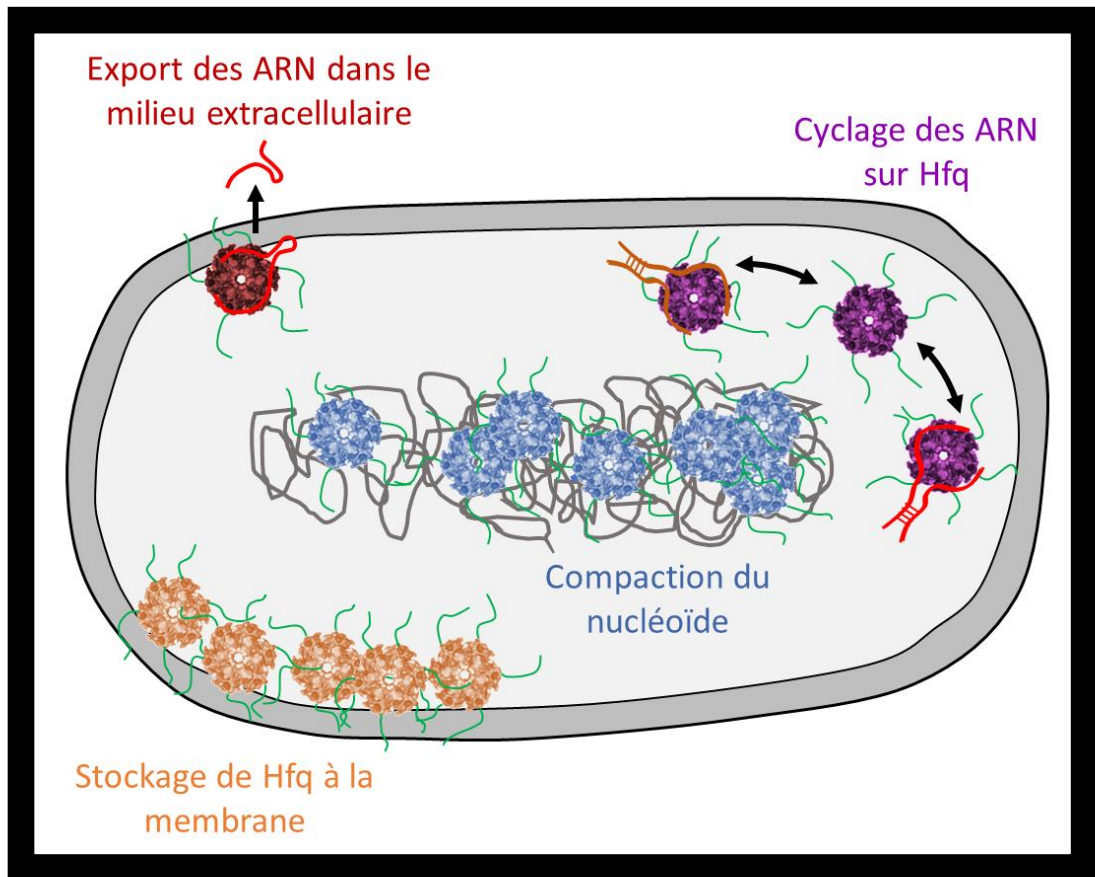


Figure 48 : Diverses fonctions de Hfq dans lesquelles le domaine C-terminal est ou pourrait être impliqué. Le tore d'Hfq (PDB 1HK9) est représenté de couleur différente selon l'aspect fonctionnel considéré. Les queues C-terminales sont représentées en vert.

Pour finir, un point crucial concerne la répartition de Hfq *in vivo*. Les différentes études donnent souvent des résultats variables et parfois contradictoires quant aux fractions membranaires, cytoplasmiques et la part associée au nucléoïde. La localisation de Hfq est difficile pour deux raisons principales :

- Son association avec les ARN fait qu'elle peut co-localiser avec ses ARN cibles. Par exemple, des ARNm impliqués dans la synthèse de protéines membranaires sont enrichis à la membrane [221]. Il est difficile de séparer une localisation « propre » de la protéine d'une association avec des ARN. Cependant, les hétérogénéités de

localisation de Hfq constatées par imagerie existent toujours en présence de rifampicine, qui bloque la transcription [61]. Et l'absence de la chaperonne ne semble pas affecter la répartition cellulaire de certains sARN [222].

- La fusion de Hfq avec une étiquette affecte très souvent sa localisation. L'ajout d'une protéine fluorescente empêche la formation des structures, notamment en perturbant l'interaction CTR-CTR ou bien carrément l'assemblage de l'hexamère. La plupart des techniques de microscopie à haute résolution sont donc inopérantes. L'ajout d'un court épitope de type FLAG ou HA peut lui conduire à un effet important sur la liaison aux acides nucléiques, de par la charge nette de ces séquences.

Ces caractéristiques expliquent que les données sur la localisation *in vivo* de Hfq sont encore insuffisantes. Il est nécessaire de poursuivre les efforts dans cette voie pour compléter les données *in vitro* qui ont été obtenues.



# Bibliographie

- [1] M. Thanbichler and L. Shapiro, 'Getting organized — how bacterial cells move proteins and DNA', *Nat. Rev. Microbiol.*, vol. 6, no. 1, pp. 28–40, Jan. 2008.
- [2] D. S. Goodsell, 'Inside a living cell', *Trends Biochem. Sci.*, vol. 16, pp. 203–206, 1991.
- [3] D. Ridgway *et al.*, 'Coarse-grained molecular simulation of diffusion and reaction kinetics in a crowded virtual cytoplasm', *Biophys. J.*, vol. 94, no. 10, pp. 3748–3759, May 2008.
- [4] S. R. McGuffee and A. H. Elcock, 'Diffusion, Crowding & Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm', *PLoS Comput. Biol.*, vol. 6, no. 3, p. e1000694, Mar. 2010.
- [5] E. Y. Gottshall, C. Seebart, J. C. Gatlin, and N. L. Ward, 'Spatially segregated transcription and translation in cells of the endomembrane-containing bacterium *Gemmata obscuriglobus*', *Proc. Natl. Acad. Sci.*, vol. 111, no. 30, pp. 11067–11072, Jul. 2014.
- [6] K. Celler, R. I. Koning, J. Willemse, A. J. Koster, and G. P. van Wezel, 'Cross-membranes orchestrate compartmentalization and morphogenesis in *Streptomyces*', *Nat. Commun.*, vol. 7, p. ncomms11836, Jun. 2016.
- [7] J. T. Mika and B. Poolman, 'Macromolecule diffusion and confinement in prokaryotic cells', *Curr. Opin. Biotechnol.*, vol. 22, no. 1, pp. 117–126, Feb. 2011.
- [8] A. Badrinarayanan, T. B. K. Le, and M. T. Laub, 'Bacterial Chromosome Organization and Segregation', *Annu. Rev. Cell Dev. Biol.*, vol. 31, no. 1, pp. 171–199, Nov. 2015.
- [9] Y.-L. Shih and L. Rothfield, 'The bacterial cytoskeleton', *Microbiol. Mol. Biol. Rev. MMBR*, vol. 70, no. 3, pp. 729–754, Sep. 2006.
- [10] K. Matsumoto, J. Kusaka, A. Nishibori, and H. Hara, 'Lipid domains in bacterial membranes', *Mol. Microbiol.*, vol. 61, no. 5, pp. 1110–1117, Sep. 2006.
- [11] M. Z. Wilson and Z. Gitai, 'Beyond the cytoskeleton: mesoscale assemblies and their function in spatial organization', *Curr. Opin. Microbiol.*, vol. 16, no. 2, pp. 177–183, Apr. 2013.
- [12] C. M. Oikonomou and G. J. Jensen, 'A new view into prokaryotic cell biology from electron cryotomography', *Nat. Rev. Microbiol.*, vol. 14, no. 4, pp. 205–220, Feb. 2016.
- [13] E. Evguenieva-Hackenberg, V. Roppelt, C. Lassek, and G. Klug, 'Subcellular localization of RNA degrading proteins and protein complexes in prokaryotes', *RNA Biol.*, vol. 8, no. 1, pp. 49–54, Jan. 2011.
- [14] X. Weng and J. Xiao, 'Spatial organization of transcription in bacterial cells', *Trends Genet.*, vol. 30, no. 7, pp. 287–297, Jul. 2014.
- [15] E. Mileykovskaya and W. Dowhan, 'Cardiolipin membrane domains in prokaryotes and eukaryotes', *Biochim. Biophys. Acta BBA - Biomembr.*, vol. 1788, no. 10, pp. 2084–2091, Oct. 2009.
- [16] W. Wang, G.-W. Li, C. Chen, X. S. Xie, and X. Zhuang, 'Chromosome Organization by a Nucleoid-Associated Protein in Live Bacteria', *Science*, vol. 333, no. 6048, pp. 1445–1449, Sep. 2011.
- [17] A. A. Buskila, S. Kannaiah, and O. Amster-Choder, 'RNA localization in bacteria', *RNA Biol.*, vol. 11, no. 8, pp. 1051–1060, Oct. 2014.
- [18] F. H. Crick, 'On protein synthesis', *Symp. Soc. Exp. Biol.*, vol. 12, pp. 138–163, 1958.

- [19] M. P. Robertson and G. F. Joyce, 'The Origins of the RNA World', *Cold Spring Harb. Perspect. Biol.*, vol. 4, no. 5, May 2012.
- [20] B. Masquida and E. Westhof, 'RNase P: at last, the key finds its lock', *RNA N. Y. N.*, vol. 17, no. 9, pp. 1615–1618, Sep. 2011.
- [21] Y. Shimoni *et al.*, 'Regulation of gene expression by small non-coding RNAs: a quantitative view', *Mol. Syst. Biol.*, vol. 3, Sep. 2007.
- [22] J. T. Wade and D. C. Grainger, 'Pervasive transcription: illuminating the dark matter of bacterial transcriptomes', *Nat. Rev. Microbiol.*, vol. 12, no. 9, pp. 647–653, 2014.
- [23] E. Lioliou, C. Romilly, P. Romby, and P. Fechter, 'RNA-mediated regulation in bacteria: from natural to artificial systems', *New Biotechnol.*, vol. 27, no. 3, pp. 222–235, Jul. 2010.
- [24] J. A. Howe *et al.*, 'Atomic resolution mechanistic studies of ribocil: A highly selective unnatural ligand mimic of the E. coli FMN riboswitch', *RNA Biol.*, vol. 13, no. 10, pp. 946–954, Aug. 2016.
- [25] A. Serganov *et al.*, 'Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs', *Chem. Biol.*, vol. 11, no. 12, pp. 1729–1741, Dec. 2004.
- [26] A. D. Garst, E. B. Porter, and R. T. Batey, 'Insights into the regulatory landscape of the lysine riboswitch', *J. Mol. Biol.*, vol. 423, no. 1, pp. 17–33, Oct. 2012.
- [27] J. Hindley, 'Fractionation of <sup>32</sup>P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting', *J. Mol. Biol.*, vol. 30, no. 1, pp. 125–136, Nov. 1967.
- [28] B. Steuten *et al.*, 'Regulation of transcription by 6S RNAs: insights from the Escherichia coli and Bacillus subtilis model systems', *RNA Biol.*, vol. 11, no. 5, pp. 508–521, 2014.
- [29] K. M. Wassarman and G. Storz, '6S RNA regulates E. coli RNA polymerase activity', *Cell*, vol. 101, no. 6, pp. 613–623, Jun. 2000.
- [30] B. Steuten, S. Schneider, and R. Wagner, '6S RNA: recent answers--future questions', *Mol. Microbiol.*, vol. 91, no. 4, pp. 641–648, Feb. 2014.
- [31] J. F. Kugel and J. A. Goodrich, 'An RNA transcriptional regulator templates its own regulatory RNA', *Nat. Chem. Biol.*, vol. 3, no. 2, pp. 89–90, Feb. 2007.
- [32] D. H. Haft, J. Selengut, E. F. Mongodin, and K. E. Nelson, 'A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes', *PLoS Comput. Biol.*, vol. 1, no. 6, p. e60, Nov. 2005.
- [33] M. L. Luo, R. T. Leenay, and C. L. Beisel, 'Current and future prospects for CRISPR-based tools in bacteria', *Biotechnol. Bioeng.*, vol. 113, no. 5, pp. 930–943, May 2016.
- [34] K. Pougach *et al.*, 'Transcription, processing and function of CRISPR cassettes in Escherichia coli: Transcription and function of E. coli CRISPR locus', *Mol. Microbiol.*, vol. 77, no. 6, pp. 1367–1379, Sep. 2010.
- [35] J. Georg and W. R. Hess, 'cis-antisense RNA, another level of gene regulation in bacteria', *Microbiol. Mol. Biol. Rev. MMBR*, vol. 75, no. 2, pp. 286–300, Jun. 2011.
- [36] M. Lybecker, I. Bilusic, and R. Raghavan, 'Pervasive transcription: detecting functional RNAs in bacteria', *Transcription*, vol. 5, no. 4, Oct. 2014.

- [37] F. Saberi, M. Kamali, A. Najafi, A. Yazdanparast, and M. M. Moghaddam, 'Natural antisense RNAs as mRNA regulatory elements in bacteria: a review on function and applications', *Cell. Mol. Biol. Lett.*, vol. 21, no. 1, Dec. 2016.
- [38] G. Storz, J. Vogel, and K. M. Wassarman, 'Regulation by Small RNAs in Bacteria: Expanding Frontiers', *Mol. Cell*, vol. 43, no. 6, pp. 880–891, Sep. 2011.
- [39] A. Smirnov *et al.*, 'Grad-seq guides the discovery of ProQ as a major small RNA-binding protein', *Proc. Natl. Acad. Sci.*, vol. 113, no. 41, pp. 11591–11596, Oct. 2016.
- [40] G. M. Gonzalez *et al.*, 'Structure of the Escherichia coli ProQ RNA-binding protein', *RNA N. Y. N*, vol. 23, no. 5, pp. 696–711, May 2017.
- [41] M. Olejniczak and G. Storz, 'ProQ/FinO-domain proteins: another ubiquitous family of RNA matchmakers?: New family of RNA matchmakers', *Mol. Microbiol.*, Apr. 2017.
- [42] D. W. Selinger, R. M. Saxena, K. J. Cheung, G. M. Church, and C. Rosenow, 'Global RNA half-life analysis in Escherichia coli reveals positional patterns of transcript degradation', *Genome Res.*, vol. 13, no. 2, pp. 216–223, Feb. 2003.
- [43] G. A. Mackie, 'Ribonuclease E is a 5'-end-dependent endonuclease', *Nature*, vol. 395, no. 6703, pp. 720–723, Oct. 1998.
- [44] A. Deana, H. Celesnik, and J. G. Belasco, 'The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal', *Nature*, vol. 451, no. 7176, pp. 355–358, Jan. 2008.
- [45] G. Joanny *et al.*, 'Polyadenylation of a functional mRNA controls gene expression in Escherichia coli', *Nucleic Acids Res.*, vol. 35, no. 8, pp. 2494–2502, 2007.
- [46] G. A. Mackie, 'RNase E: at the interface of bacterial RNA processing and decay', *Nat. Rev. Microbiol.*, vol. 11, no. 1, pp. 45–57, Dec. 2012.
- [47] M. W. Górna, A. J. Carpousis, and B. F. Luisi, 'From conformational chaos to robust regulation: the structure and function of the multi-enzyme RNA degradosome', *Q. Rev. Biophys.*, vol. 45, no. 02, pp. 105–145, May 2012.
- [48] V. Chandran and B. F. Luisi, 'Recognition of enolase in the Escherichia coli RNA degradosome', *J. Mol. Biol.*, vol. 358, no. 1, pp. 8–15, Apr. 2006.
- [49] S. W. Hardwick and B. F. Luisi, 'Rarely at rest: RNA helicases and their busy contributions to RNA degradation, regulation and quality control', *RNA Biol.*, vol. 10, no. 1, pp. 56–70, Jan. 2013.
- [50] B. Py, C. F. Higgins, H. M. Krisch, and A. J. Carpousis, 'A DEAD-box RNA helicase in the Escherichia coli RNA degradosome', *Nature*, vol. 381, no. 6578, pp. 169–172, May 1996.
- [51] A. Resch, B. Večerek, K. Palavra, and U. Bläsi, 'Requirement of the CsdA DEAD-box helicase for low temperature riboregulation of rpoS mRNA', *RNA Biol.*, vol. 7, no. 6, pp. 796–802, Dec. 2010.
- [52] V. Khemici, I. Toesca, L. Poljak, N. F. Vanzo, and A. J. Carpousis, 'The RNase E of Escherichia coli has at least two binding sites for DEAD-box RNA helicases: functional replacement of RhlB by RhlE', *Mol. Microbiol.*, vol. 54, no. 5, pp. 1422–1430, Dec. 2004.
- [53] A. J. Callaghan *et al.*, 'Studies of the RNA degradosome-organizing domain of the Escherichia coli ribonuclease RNase E', *J. Mol. Biol.*, vol. 340, no. 5, pp. 965–979, Jul. 2004.

- [54] K. J. Bandyra, M. Bouvier, A. J. Carpousis, and B. F. Luisi, 'The social fabric of the RNA degradosome', *Biochim. Biophys. Acta BBA - Gene Regul. Mech.*, vol. 1829, no. 6–7, pp. 514–522, Jun. 2013.
- [55] V. Khemici, L. Poljak, B. F. Luisi, and A. J. Carpousis, 'The RNase E of *Escherichia coli* is a membrane-binding protein', *Mol. Microbiol.*, Nov. 2008.
- [56] J. Salje, F. van den Ent, P. de Boer, and J. Löwe, 'Direct membrane binding by bacterial actin MreB', *Mol. Cell*, vol. 43, no. 3, pp. 478–487, Aug. 2011.
- [57] H. Zhou and J. Lutkenhaus, 'Membrane binding by MinD involves insertion of hydrophobic residues within the C-terminal amphipathic helix into the bilayer', *J. Bacteriol.*, vol. 185, no. 15, pp. 4326–4335, Aug. 2003.
- [58] A. Taghbalout and L. Rothfield, 'RNaseE and the other constituents of the RNA degradosome are components of the bacterial cytoskeleton', *Proc. Natl. Acad. Sci.*, vol. 104, no. 5, pp. 1667–1672, 2007.
- [59] A. Taghbalout and L. Rothfield, 'RNaseE and RNA Helicase B Play Central Roles in the Cytoskeletal Organization of the RNA Degradosome', *J. Biol. Chem.*, vol. 283, no. 20, pp. 13850–13855, May 2008.
- [60] A. Taghbalout and Q. Yang, 'Self-Assembly of the Bacterial Cytoskeleton-Associated RNA Helicase B Protein into Polymeric Filamentous Structures', *J. Bacteriol.*, vol. 192, no. 12, pp. 3222–3226, Jun. 2010.
- [61] A. Taghbalout, Q. Yang, and V. Arluison, 'The *Escherichia coli* RNA processing and degradation machinery is compartmentalized within an organized cellular network', *Biochem. J.*, vol. 458, no. 1, pp. 11–22, Feb. 2014.
- [62] H. Strahl *et al.*, 'Membrane Recognition and Dynamics of the RNA Degradosome', *PLoS Genet.*, vol. 11, no. 2, p. e1004961, Feb. 2015.
- [63] J. R. Moffitt, S. Pandey, A. N. Boettiger, S. Wang, and X. Zhuang, 'Spatial organization shapes the turnover of a bacterial transcriptome', *Elife*, vol. 5, p. e13065, 2016.
- [64] M. T. Franze de Fernandez, L. Eoyang, and J. T. August, 'Factor fraction required for the synthesis of bacteriophage Qbeta-RNA', *Nature*, vol. 219, no. 5154, pp. 588–590, Aug. 1968.
- [65] X. Sun, I. Zhulin, and R. M. Wartell, 'Predicted structure and phyletic distribution of the RNA-binding protein Hfq', *Nucleic Acids Res.*, vol. 30, no. 17, pp. 3662–3671, Sep. 2002.
- [66] C. L. Will and R. Lührmann, 'Spliceosomal UsnRNP biogenesis, structure and function', *Curr. Opin. Cell Biol.*, vol. 13, no. 3, pp. 290–301, Jun. 2001.
- [67] S. Tharun, W. He, A. E. Mayes, P. Lennertz, and others, 'Yeast Sm-like proteins function in mRNA decapping and decay', *Nature*, vol. 404, no. 6777, p. 515, 2000.
- [68] A. G. Seto, A. J. Zaug, S. G. Sobel, S. L. Wolin, and T. R. Cech, '*Saccharomyces cerevisiae* telomerase is an Sm small nuclear ribonucleoprotein particle', *Nature*, vol. 401, no. 6749, pp. 177–180, Sep. 1999.
- [69] C. Mura, P. S. Randolph, J. Patterson, and A. E. Cozen, 'Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function', *RNA Biol.*, vol. 10, no. 4, pp. 636–651, Apr. 2013.
- [70] J. M. Moll, M. Sobti, and B. C. Mabbutt, 'The Lsm Proteins: Ring Architectures for RNA Capture', in *RNA Processing*, InTech, 2011.
- [71] E. Sauer, 'Structure and RNA-binding properties of the bacterial LSm protein Hfq', *RNA Biol.*, vol. 10, no. 4, pp. 610–618, Apr. 2013.

- [72] D. G. Scofield and M. Lynch, 'Evolutionary Diversification of the Sm Family of RNA-Associated Proteins', *Mol. Biol. Evol.*, vol. 25, no. 11, pp. 2255–2267, Nov. 2008.
- [73] S. Veretnik, C. Wills, P. Youkharibache, R. E. Valas, and P. E. Bourne, 'Sm/Lsm genes provide a glimpse into the early evolution of the spliceosome', *PLoS Comput. Biol.*, vol. 5, no. 3, p. e1000315, Mar. 2009.
- [74] M. A. Schumacher, R. F. Pearson, T. Møller, P. Valentin-Hansen, and R. G. Brennan, 'Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein', *EMBO J.*, vol. 21, no. 13, pp. 3546–3556, Jul. 2002.
- [75] C. Sauter, J. Basquin, and D. Suck, 'Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*', *Nucleic Acids Res.*, vol. 31, no. 14, pp. 4091–4098, Jul. 2003.
- [76] A. Nikulin *et al.*, 'Structure of *Pseudomonas aeruginosa* Hfq protein', *Acta Crystallogr. D Biol. Crystallogr.*, vol. 61, no. Pt 2, pp. 141–146, Feb. 2005.
- [77] D. Schilling and U. Gerischer, 'The *Acinetobacter baylyi* hfq Gene Encodes a Large Protein with an Unusual C Terminus', *J. Bacteriol.*, vol. 191, no. 17, pp. 5553–5562, Sep. 2009.
- [78] I. Letunic and P. Bork, 'Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees', *Nucleic Acids Res.*, vol. 44, no. W1, pp. W242–245, Jul. 2016.
- [79] T. D. Schneider and R. M. Stephens, 'Sequence logos: a new way to display consensus sequences', *Nucleic Acids Res.*, vol. 18, no. 20, pp. 6097–6100, Oct. 1990.
- [80] S. Panja and S. A. Woodson, 'Hexamer to Monomer Equilibrium of *E. coli* Hfq in Solution and Its Impact on RNA Annealing', *J. Mol. Biol.*, vol. 417, no. 5, pp. 406–412, Apr. 2012.
- [81] V. Arluison *et al.*, 'The C-terminal domain of *Escherichia coli* Hfq increases the stability of the hexamer: The role of the C-terminal domain on Hfq', *Eur. J. Biochem.*, vol. 271, no. 7, pp. 1258–1265, Mar. 2004.
- [82] M. Beich-Frandsen *et al.*, 'Structural insights into the dynamics and function of the C-terminus of the *E. coli* RNA chaperone Hfq', *Nucleic Acids Res.*, vol. 39, no. 11, pp. 4900–4915, Jun. 2011.
- [83] M. Beich-Frandsen, B. Večerek, B. Sjöblom, U. Bläsi, and K. Djinović-Carugo, 'Structural analysis of full-length Hfq from *Escherichia coli*', *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.*, vol. 67, no. 5, pp. 536–540, May 2011.
- [84] Z. R. Yang, R. Thomson, P. McNeil, and R. M. Esnouf, 'RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins', *Bioinforma. Oxf. Engl.*, vol. 21, no. 16, pp. 3369–3376, Aug. 2005.
- [85] J. Kyte and R. F. Doolittle, 'A simple method for displaying the hydropathic character of a protein', *J. Mol. Biol.*, vol. 157, no. 1, pp. 105–132, May 1982.
- [86] H. C. Tsui, G. Feng, and M. E. Winkler, 'Negative regulation of mutS and mutH repair gene expression by the Hfq and RpoS global regulators of *Escherichia coli* K-12.', *J. Bacteriol.*, vol. 179, no. 23, pp. 7476–7487, 1997.
- [87] B. VEČEREK, I. MOLL, and U. BLÄSI, 'Translational autocontrol of the *Escherichia coli* hfq RNA chaperone gene', *RNA*, vol. 11, no. 6, pp. 976–984, Jun. 2005.

- [88] C. S. Baker, L. A. Eöry, H. Yakhnin, J. Mercante, T. Romeo, and P. Babitzke, 'CsrA Inhibits Translation Initiation of Escherichia coli hfq by Binding to a Single Site Overlapping the Shine-Dalgarno Sequence', *J. Bacteriol.*, vol. 189, no. 15, pp. 5472–5481, Aug. 2007.
- [89] M. Kajitani, A. Kato, A. Wada, Y. Inokuchi, and A. Ishihama, 'Regulation of the Escherichia coli hfq gene encoding the host factor for phage Q beta.', *J. Bacteriol.*, vol. 176, no. 2, pp. 531–534, Jan. 1994.
- [90] G. M. Cech, B. Pakuła, D. Kamrowska, G. Węgrzyn, V. Arluison, and A. Szalewska-Pałasz, 'Hfq protein deficiency in Escherichia coli affects ColE1-like but not  $\lambda$  plasmid DNA replication', *Plasmid*, vol. 73, pp. 10–15, May 2014.
- [91] H.-C. T. Tsui, H.-C. E. Leung, and M. E. Winkler, 'Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12', *Mol. Microbiol.*, vol. 13, no. 1, pp. 35–49, 1994.
- [92] A. Sittka *et al.*, 'Deep Sequencing Analysis of Small Noncoding RNA and mRNA Targets of the Global Post-Transcriptional Regulator, Hfq', *PLoS Genet.*, vol. 4, no. 8, p. e1000163, Aug. 2008.
- [93] J. R. Feliciano, A. M. Grilo, S. I. Guerreiro, S. A. Sousa, and J. H. Leitão, 'Hfq: a multifaceted RNA chaperone involved in virulence', *Future Microbiol.*, vol. 11, no. 1, pp. 137–151, 2016.
- [94] A. Muffler, D. Fischer, and R. Hengge-Aronis, 'The RNA-binding protein HF-I, known as a host factor for phage Qbeta RNA replication, is essential for rpoS translation in Escherichia coli', *Genes Dev.*, vol. 10, no. 9, pp. 1143–1151, May 1996.
- [95] S. Gottesman, 'The small RNA regulators of Escherichia coli: roles and mechanisms\*', *Annu. Rev. Microbiol.*, vol. 58, pp. 303–328, 2004.
- [96] K. I. Udekwi, F. Darfeuille, J. Vogel, J. Reimegård, E. Holmqvist, and E. G. H. Wagner, 'Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA', *Genes Dev.*, vol. 19, no. 19, pp. 2355–2366, Oct. 2005.
- [97] C. K. Vanderpool and S. Gottesman, 'Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system', *Mol. Microbiol.*, vol. 54, no. 4, pp. 1076–1089, Nov. 2004.
- [98] F. Mika *et al.*, 'Targeting of csgD by the small regulatory RNA RprA links stationary phase, biofilm formation and cell envelope stress in Escherichia coli', *Mol. Microbiol.*, vol. 84, no. 1, pp. 51–65, Apr. 2012.
- [99] E. Massé, C. K. Vanderpool, and S. Gottesman, 'Effect of RyhB Small RNA on Global Iron Use in Escherichia coli', *J. Bacteriol.*, vol. 187, no. 20, pp. 6962–6971, Oct. 2005.
- [100] N. Majdalani, C. Cunning, D. Sledjeski, T. Elliott, and S. Gottesman, 'DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 21, pp. 12462–12467, Oct. 1998.
- [101] N. Majdalani, D. Hernandez, and S. Gottesman, 'Regulation and mode of action of the second small RNA activator of RpoS translation, RprA', *Mol. Microbiol.*, vol. 46, no. 3, pp. 813–826, Nov. 2002.

- [102] I. Moll, T. Afonyushkin, O. Vytvytska, V. R. Kaberdin, and U. Bläsi, 'Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs', *RNA N. Y. N.*, vol. 9, no. 11, pp. 1308–1314, Nov. 2003.
- [103] T. Afonyushkin, B. Vecerek, I. Moll, U. Bläsi, and V. R. Kaberdin, 'Both RNase E and RNase III control the stability of *sodB* mRNA upon translational inhibition by the small regulatory RNA *RyhB*', *Nucleic Acids Res.*, vol. 33, no. 5, pp. 1678–1689, 2005.
- [104] B. K. Mohanty, V. F. Maples, and S. R. Kushner, 'The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in *Escherichia coli*', *Mol. Microbiol.*, vol. 54, no. 4, pp. 905–920, Nov. 2004.
- [105] N. Sedlyarova *et al.*, 'sRNA-Mediated Control of Transcription Termination in *E. coli*', *Cell*, vol. 167, no. 1, p. 111–121.e13, Sep. 2016.
- [106] E. Holmqvist *et al.*, 'Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo', *EMBO J.*, Apr. 2016.
- [107] S. Melamed *et al.*, 'Global Mapping of Small RNA-Target Interactions in Bacteria', *Mol. Cell*, vol. 63, no. 5, pp. 884–897, Sep. 2016.
- [108] L. Argaman, M. Elgrably-Weiss, T. Hershko, J. Vogel, and S. Altuvia, 'RelA protein stimulates the activity of *RyhB* small RNA by acting on RNA-binding protein Hfq', *Proc. Natl. Acad. Sci.*, vol. 109, no. 12, pp. 4621–4626, Mar. 2012.
- [109] A. Fender, J. Elf, K. Hampel, B. Zimmermann, and E. G. H. Wagner, 'RNAs actively cycle on the Sm-like protein Hfq', *Genes Dev.*, vol. 24, no. 23, pp. 2621–2626, Dec. 2010.
- [110] E. G. H. Wagner, 'Cycling of RNAs on Hfq', *RNA Biol.*, vol. 10, no. 4, pp. 619–626, Apr. 2013.
- [111] J. Vogel and B. F. Luisi, 'Hfq and its constellation of RNA', *Nat. Rev. Microbiol.*, vol. 9, no. 8, pp. 578–589, Aug. 2011.
- [112] T. M. Link, P. Valentin-Hansen, and R. G. Brennan, 'Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 46, pp. 19292–19297, Nov. 2009.
- [113] E. A. Lesnik, R. Sampath, H. B. Levene, T. J. Henderson, J. A. McNeil, and D. J. Ecker, 'Prediction of rho-independent transcriptional terminators in *Escherichia coli*', *Nucleic Acids Res.*, vol. 29, no. 17, pp. 3583–3594, Sep. 2001.
- [114] T. A. Geissmann and D. Touati, 'Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator', *EMBO J.*, vol. 23, no. 2, pp. 396–405, Jan. 2004.
- [115] C. C. Brescia, P. J. Mikulecky, A. L. Feig, and D. D. Sledjeski, 'Identification of the Hfq-binding site on *DsrA* RNA: Hfq binds without altering *DsrA* secondary structure', *RNA N. Y. N.*, vol. 9, no. 1, pp. 33–43, Jan. 2003.
- [116] S. Panja, D. J. Schu, and S. A. Woodson, 'Conserved arginines on the rim of Hfq catalyze base pair formation and exchange', *Nucleic Acids Res.*, vol. 41, no. 15, pp. 7536–7546, Aug. 2013.
- [117] D. J. Schu, A. Zhang, S. Gottesman, and G. Storz, 'Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition', *EMBO J.*, vol. 34, no. 20, pp. 2557–2573, Oct. 2015.
- [118] B. Guillemardet, *Rôle de la protéine chaperonne Hfq dans la réponse des ARN messagers à la régulation par les petits ARN*. Paris Saclay, 2016.



- [119] A. S. Olsen, J. Møller-Jensen, R. G. Brennan, and P. Valentin-Hansen, 'C-Terminally Truncated Derivatives of Escherichia coli Hfq Are Proficient in Riboregulation', *J. Mol. Biol.*, vol. 404, no. 2, pp. 173–182, Nov. 2010.
- [120] J. Caillet, C. Gracia, F. Fontaine, and E. Hajnsdorf, '*Clostridium difficile* Hfq can replace *Escherichia coli* Hfq for most of its function', *RNA*, vol. 20, no. 10, pp. 1567–1578, Oct. 2014.
- [121] E. Sonnleitner, I. Moll, and U. Bläsi, 'Functional replacement of the Escherichia coli hfq gene by the homologue of Pseudomonas aeruginosa', *Microbiology*, vol. 148, no. 3, pp. 883–891, 2002.
- [122] J. S. Nielsen *et al.*, 'An Hfq-like protein in archaea: Crystal structure and functional characterization of the Sm protein from Methanococcus jannaschii', *RNA*, vol. 13, no. 12, pp. 2213–2223, Sep. 2007.
- [123] A. Santiago-Frangos, K. Kavita, D. J. Schu, S. Gottesman, and S. A. Woodson, 'C-terminal domain of the RNA chaperone Hfq drives sRNA competition and release of target RNA', *Proc. Natl. Acad. Sci.*, p. 201613053, 2016.
- [124] G. Butland *et al.*, 'Interaction network containing conserved and essential protein complexes in Escherichia coli', *Nature*, vol. 433, no. 7025, pp. 531–537, Feb. 2005.
- [125] A. Jain, R. Liu, Y. K. Xiang, and T. Ha, 'Single-molecule pull-down for studying protein interactions', *Nat. Protoc.*, vol. 7, no. 3, pp. 445–452, Feb. 2012.
- [126] T. Morita, K. Maki, and H. Aiba, 'RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs', *Genes Dev.*, vol. 19, no. 18, pp. 2176–2186, 2005.
- [127] K. Yonekura, M. Watanabe, Y. Kageyama, K. Hirata, M. Yamamoto, and S. Maki-Yonekura, 'Post-transcriptional regulator Hfq binds catalase HPII: crystal structure of the complex', *PLoS One*, vol. 8, no. 11, p. e78216, 2013.
- [128] T. A. Azam, S. Hiraga, and A. Ishihama, 'Two types of localization of the DNA-binding proteins within the Escherichia coli nucleoid', *Genes Cells*, vol. 5, no. 8, pp. 613–626, 2000.
- [129] T. A. Azam, A. Iwata, A. Nishimura, S. Ueda, and A. Ishihama, 'Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid', *J. Bacteriol.*, vol. 181, no. 20, pp. 6361–6370, 1999.
- [130] E. Diestra, B. Cayrol, V. Arluison, and C. Risco, 'Cellular Electron Microscopy Imaging Reveals the Localization of the Hfq Protein Close to the Bacterial Membrane', *PLoS ONE*, vol. 4, no. 12, p. e8301, Dec. 2009.
- [131] V. Arluison *et al.*, 'Three-dimensional Structures of Fibrillar Sm Proteins: Hfq and Other Sm-like Proteins', *J. Mol. Biol.*, vol. 356, no. 1, pp. 86–96, Feb. 2006.
- [132] E. Fortas *et al.*, 'New insight into the structure and function of Hfq C-terminus', *Biosci. Rep.*, vol. 35, no. 2, 2015.
- [133] V. N. Uversky, 'Intrinsically disordered proteins from A to Z', *Int. J. Biochem. Cell Biol.*, vol. 43, no. 8, pp. 1090–1103, Aug. 2011.
- [134] H. J. Dyson and P. E. Wright, 'Intrinsically unstructured proteins and their functions', *Nat. Rev. Mol. Cell Biol.*, vol. 6, no. 3, pp. 197–208, Mar. 2005.
- [135] J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, and D. T. Jones, 'Prediction and functional analysis of native disorder in proteins from the three kingdoms of life', *J. Mol. Biol.*, vol. 337, no. 3, pp. 635–645, Mar. 2004.

- [136] S. Fukuchi, K. Hosoda, K. Homma, T. Gojobori, and K. Nishikawa, 'Binary classification of protein molecules into intrinsically disordered and ordered segments', *BMC Struct. Biol.*, vol. 11, no. 1, p. 29, 2011.
- [137] Y.-S. Lin, W.-L. Hsu, J.-K. Hwang, and W.-H. Li, 'Proportion of solvent-exposed amino acids in a protein and rate of protein evolution', *Mol. Biol. Evol.*, vol. 24, no. 4, pp. 1005–1011, Apr. 2007.
- [138] C. J. Brown *et al.*, 'Evolutionary rate heterogeneity in proteins with long disordered regions', *J. Mol. Evol.*, vol. 55, no. 1, pp. 104–110, Jul. 2002.
- [139] K. Namba, 'Roles of partly unfolded conformations in macromolecular self-assembly', *Genes Cells Devoted Mol. Cell. Mech.*, vol. 6, no. 1, pp. 1–12, Jan. 2001.
- [140] C. Haynes *et al.*, 'Intrinsic Disorder Is a Common Feature of Hub Proteins from Four Eukaryotic Interactomes', *PLoS Comput. Biol.*, vol. 2, no. 8, p. e100, 2006.
- [141] D. J. Selkoe, 'Cell biology of protein misfolding: the examples of Alzheimer's and Parkinson's diseases', *Nat. Cell Biol.*, vol. 6, no. 11, pp. 1054–1061, Nov. 2004.
- [142] F. Chiti and C. M. Dobson, 'Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade', *Annu. Rev. Biochem.*, May 2017.
- [143] P. Westermark, A. Andersson, and G. T. Westermark, 'Islet amyloid polypeptide, islet amyloid, and diabetes mellitus', *Physiol. Rev.*, vol. 91, no. 3, pp. 795–826, Jul. 2011.
- [144] C. M. Dobson, 'Protein folding and misfolding', *Nature*, vol. 426, no. 6968, pp. 884–890, Dec. 2003.
- [145] P. Arosio, T. P. J. Knowles, and S. Linse, 'On the lag phase in amyloid fibril formation', *Phys Chem Chem Phys*, vol. 17, no. 12, pp. 7606–7618, 2015.
- [146] G. Plakoutsi, F. Bemporad, M. Calamai, N. Taddei, C. M. Dobson, and F. Chiti, 'Evidence for a mechanism of amyloid formation involving molecular reorganisation within native-like precursor aggregates', *J. Mol. Biol.*, vol. 351, no. 4, pp. 910–922, Aug. 2005.
- [147] N. Carulla *et al.*, 'Experimental characterization of disordered and ordered aggregates populated during the process of amyloid fibril formation', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 19, pp. 7828–7833, May 2009.
- [148] J. Lee, E. K. Culyba, E. T. Powers, and J. W. Kelly, 'Amyloid- $\beta$  forms fibrils by nucleated conformational conversion of oligomers', *Nat. Chem. Biol.*, vol. 7, no. 9, pp. 602–609, Jul. 2011.
- [149] T. K. Karamanos, A. P. Kalverda, G. S. Thompson, and S. E. Radford, 'Mechanisms of amyloid formation revealed by solution NMR', *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 0, pp. 86–104, Aug. 2015.
- [150] R. Friedman, R. Pellarin, and A. Caflich, 'Amyloid Aggregation on Lipid Bilayers and Its Impact on Membrane Permeability', *J. Mol. Biol.*, vol. 387, no. 2, pp. 407–415, Mar. 2009.
- [151] L. Milanese *et al.*, 'Direct three-dimensional visualization of membrane disruption by amyloid fibrils', *Proc. Natl. Acad. Sci.*, vol. 109, no. 50, pp. 20455–20460, Dec. 2012.
- [152] A. Abedini and D. P. Raleigh, 'A critical assessment of the role of helical intermediates in amyloid formation by natively unfolded proteins and polypeptides', *Protein Eng. Des. Sel.*, vol. 22, no. 8, pp. 453–459, Aug. 2009.

- [153] M. Sunde and C. Blake, 'The structure of amyloid fibrils by electron microscopy and X-ray diffraction', *Adv. Protein Chem.*, vol. 50, pp. 123–159, 1997.
- [154] G. Zandomenighi, M. R. H. Krebs, M. G. McCammon, and M. Fändrich, 'FTIR reveals structural differences between native beta-sheet proteins and amyloid fibrils', *Protein Sci. Publ. Protein Soc.*, vol. 13, no. 12, pp. 3314–3321, Dec. 2004.
- [155] M. Nilsson, 'Techniques to study amyloid fibril formation in vitro', *Methods*, vol. 34, no. 1, pp. 151–160, Sep. 2004.
- [156] T. P. Knowles *et al.*, 'Role of intermolecular forces in defining material properties of protein nanofibrils', *Science*, vol. 318, no. 5858, pp. 1900–1903, Dec. 2007.
- [157] M. Fändrich and C. M. Dobson, 'The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation', *EMBO J.*, vol. 21, no. 21, pp. 5682–5690, Nov. 2002.
- [158] C. P. J. Maury, 'The emerging concept of functional amyloid', *J. Intern. Med.*, vol. 265, no. 3, pp. 329–334, Mar. 2009.
- [159] J. D. Taylor and S. J. Matthews, 'New insight into the molecular control of bacterial functional amyloids', *Front. Cell. Infect. Microbiol.*, vol. 5, Apr. 2015.
- [160] W. Hawthorne, S. Rouse, L. Sewell, and S. J. Matthews, 'Structural insights into functional amyloid inhibition in Gram -ve bacteria', *Biochem. Soc. Trans.*, vol. 44, no. 6, pp. 1643–1649, Dec. 2016.
- [161] M. S. Dueholm *et al.*, 'Functional amyloid in *Pseudomonas*', *Mol. Microbiol.*, vol. 77, no. 4, pp. 1009–1020, Aug. 2010.
- [162] R. N. Besingi *et al.*, 'Functional amyloids in *Streptococcus mutans*, their use as targets of biofilm inhibition and initial characterization of SMU\_63c', *Microbiol. Read. Engl.*, vol. 163, no. 4, pp. 488–501, Apr. 2017.
- [163] P. Goyal *et al.*, 'Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG', *Nature*, vol. 516, no. 7530, pp. 250–253, Sep. 2014.
- [164] I. Pallarès and S. Ventura, 'The Transcription Terminator Rho: A First Bacterial Prion', *Trends Microbiol.*, 2017.
- [165] J. Shorter and S. Lindquist, 'Prions as adaptive conduits of memory and inheritance', *Nat. Rev. Genet.*, vol. 6, no. 6, pp. 435–450, Jun. 2005.
- [166] D. C. Masison and R. B. Wickner, 'Prion-inducing domain of yeast Ure2p and protease resistance of Ure2p in prion-containing cells', *Science*, vol. 270, no. 5233, pp. 93–95, Oct. 1995.
- [167] V. Iglesias, N. S. de Groot, and S. Ventura, 'Computational analysis of candidate prion-like proteins in bacteria and their role', *Front. Microbiol.*, vol. 6, Oct. 2015.
- [168] A. H. Yuan and A. Hochschild, 'A bacterial global regulator forms a prion', *Science*, vol. 355, no. 6321, pp. 198–201, Jan. 2017.
- [169] T. C. Jackson and M. V. Sukhodolets, 'Functional analyses of putative PalS ( Pal indromic S elf-recognition) motifs in bacterial Hfq', *Biochem. Biophys. Res. Commun.*, vol. 486, no. 4, pp. 1048–1054, May 2017.
- [170] L. Postow, C. D. Hardy, J. Arsuaga, and N. R. Cozzarelli, 'Topological domain structure of the *Escherichia coli* chromosome', *Genes Dev.*, vol. 18, no. 14, pp. 1766–1779, 2004.
- [171] C. J. Dorman, 'DNA supercoiling and bacterial gene expression', *Sci. Prog.*, vol. 89, no. Pt 3-4, pp. 151–166, 2006.

- [172] C. D. Hardy and N. R. Cozzarelli, 'A genetic selection for supercoiling mutants of *Escherichia coli* reveals proteins implicated in chromosome structure', *Mol. Microbiol.*, vol. 57, no. 6, pp. 1636–1652, Sep. 2005.
- [173] D. Song and J. J. Loparo, 'Building bridges within the bacterial chromosome', *Trends Genet.*, vol. 31, no. 3, pp. 164–173, Mar. 2015.
- [174] P. Sobetzko, A. Travers, and G. Muskhelishvili, 'Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle', *Proc. Natl. Acad. Sci.*, vol. 109, no. 2, pp. E42–E50, Jan. 2012.
- [175] J. Kim, S. H. Yoshimura, K. Hizume, R. L. Ohniwa, A. Ishihama, and K. Takeyasu, 'Fundamental structural units of the *Escherichia coli* nucleoid revealed by atomic force microscopy', *Nucleic Acids Res.*, vol. 32, no. 6, pp. 1982–1992, 2004.
- [176] C. Cibot, 'Etude du surenroulement diffusible de l'ADN chromosomique chez la bactérie *Escherichia Coli*', phdthesis, Université Paris-Saclay, 2015.
- [177] J. S. Graham, R. C. Johnson, and J. F. Marko, 'Concentration-dependent exchange accelerates turnover of proteins bound to double-stranded DNA', *Nucleic Acids Res.*, vol. 39, no. 6, pp. 2249–2259, Mar. 2011.
- [178] N. Hadizadeh, R. C. Johnson, and J. F. Marko, 'Facilitated Dissociation of a Nucleoid Protein from the Bacterial Chromosome', *J. Bacteriol.*, vol. 198, no. 12, pp. 1735–1742, Jun. 2016.
- [179] K. Dahlke and C. E. Sing, 'Facilitated Dissociation Kinetics of Dimeric Nucleoid-Associated Proteins Follow a Universal Curve', *Biophys. J.*, vol. 112, no. 3, pp. 543–551, Feb. 2017.
- [180] R. D. Giuntoli *et al.*, 'DNA-Segment-Facilitated Dissociation of Fis and NHP6A from DNA Detected via Single-Molecule Mechanical Response', *J. Mol. Biol.*, vol. 427, no. 19, pp. 3123–3136, Sep. 2015.
- [181] N. L. Craig and H. A. Nash, 'E. coli integration host factor binds to specific sites in DNA', *Cell*, vol. 39, no. 3 Pt 2, pp. 707–716, Dec. 1984.
- [182] S. Khrapunov, M. Brenowitz, P. A. Rice, and C. E. Catalano, 'Binding then bending: A mechanism for wrapping DNA', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 51, pp. 19217–19218, Dec. 2006.
- [183] M. Engelhorn, F. Boccard, C. Murtin, P. Prentki, and J. Geiselmann, 'In vivo interaction of the *Escherichia coli* integration host factor with its specific binding sites', *Nucleic Acids Res.*, vol. 23, no. 17, pp. 2959–2965, Sep. 1995.
- [184] C. Murtin, M. Engelhorn, J. Geiselmann, and F. Boccard, 'A quantitative UV laser footprinting analysis of the interaction of IHF with specific binding sites: re-evaluation of the effective concentration of IHF in the cell', *J. Mol. Biol.*, vol. 284, no. 4, pp. 949–961, Dec. 1998.
- [185] B. Kundukad, P. Cong, J. R. C. van der Maarel, and P. S. Doyle, 'Time-dependent bending rigidity and helical twist of DNA by rearrangement of bound HU protein', *Nucleic Acids Res.*, vol. 41, no. 17, pp. 8280–8288, Sep. 2013.
- [186] M. Hammel *et al.*, 'HU multimerization shift controls nucleoid compaction', *Sci. Adv.*, vol. 2, no. 7, pp. e1600650–e1600650, Jul. 2016.
- [187] S. Kar, R. Edgar, and S. Adhya, 'Nucleoid remodeling by an altered HU protein: reorganization of the transcription program', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 45, pp. 16397–16402, Nov. 2005.
- [188] C. J. Dorman, 'Nucleoid-associated proteins and bacterial physiology', *Adv. Appl. Microbiol.*, vol. 67, pp. 47–64, 2009.

- [189] S. Stella, D. Cascio, and R. C. Johnson, 'The shape of the DNA minor groove directs binding by the DNA-bending protein Fis', *Genes Dev.*, vol. 24, no. 8, pp. 814–826, Apr. 2010.
- [190] R. Schneider, A. Travers, T. Kutateladze, and G. Muskhelishvili, 'A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*', *Mol. Microbiol.*, vol. 34, no. 5, pp. 953–964, Dec. 1999.
- [191] N. N. Vtyurina, D. Dulin, M. W. Docter, A. S. Meyer, N. H. Dekker, and E. A. Abbondanzieri, 'Hysteresis in DNA compaction by Dps is described by an Ising model', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 18, pp. 4982–4987, May 2016.
- [192] S. G. Wolf, D. Frenkiel, T. Arad, S. E. Finkel, R. Kolter, and A. Minsky, 'DNA protection by stress-induced biocrystallization', *Nature*, vol. 400, no. 6739, pp. 83–85, Jul. 1999.
- [193] P. Ceci, 'DNA condensation and self-aggregation of *Escherichia coli* Dps are coupled phenomena related to the properties of the N-terminus', *Nucleic Acids Res.*, vol. 32, no. 19, pp. 5935–5944, Oct. 2004.
- [194] M. V. Kotlajich, D. R. Hron, B. A. Boudreau, Z. Sun, Y. L. Lyubchenko, and R. Landick, 'Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria', *Elife*, vol. 4, p. e04970, 2015.
- [195] A. Kaidow, M. Wachi, J. Nakamura, J. Magae, and K. Nagai, 'Anucleate cell production by *Escherichia coli* delta hns mutant lacking a histone-like protein, H-NS.', *J. Bacteriol.*, vol. 177, no. 12, pp. 3589–3592, Jun. 1995.
- [196] E. Helgesen, S. Fossum-Raunehaug, and K. Skarstad, 'Lack of the H-NS Protein Results in Extended and Aberrantly Positioned DNA during Chromosome Replication and Segregation in *Escherichia coli*', *J. Bacteriol.*, vol. 198, no. 8, pp. 1305–1316, Apr. 2016.
- [197] M. Shin *et al.*, 'DNA looping-mediated repression by histone-like protein H-NS: specific requirement of  $E\sigma 70$  as a cofactor for looping', *Genes Dev.*, vol. 19, no. 19, pp. 2388–2398, Oct. 2005.
- [198] N. Blot, R. Mavathur, M. Geertz, A. Travers, and G. Muskhelishvili, 'Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome', *EMBO Rep.*, vol. 7, no. 7, pp. 710–715, Jul. 2006.
- [199] W. Wang, G.-W. Li, C. Chen, X. S. Xie, and X. Zhuang, 'Chromosome organization by a nucleoid-associated protein in live bacteria', *Science*, vol. 333, no. 6048, pp. 1445–1449, Sep. 2011.
- [200] S. Nolivos and D. Sherratt, 'The bacterial chromosome: architecture and action of bacterial SMC and SMC-like complexes', *FEMS Microbiol. Rev.*, vol. 38, no. 3, pp. 380–392, May 2014.
- [201] T. Weitao, K. Nordström, and S. Dasgupta, '*Escherichia coli* cell cycle control genes affect chromosome superhelicity', *EMBO Rep.*, vol. 1, no. 6, pp. 494–499, Dec. 2000.
- [202] O. Danilova, R. Reyes-Lamothe, M. Pinskaya, D. Sherratt, and C. Possoz, 'MukB colocalizes with the oriC region and is required for organization of the two *Escherichia coli* chromosome arms into separate cell halves', *Mol. Microbiol.*, vol. 65, no. 6, pp. 1485–1492, Sep. 2007.
- [203] J. M. Moore *et al.*, 'Roles of Nucleoid-Associated Proteins in Stress-Induced Mutagenic Break Repair in Starving *Escherichia coli*', *Genetics*, vol. 201, no. 4, pp. 1349–1362, Dec. 2015.

- [204] S. Li and R. Waters, 'Escherichia coli strains lacking protein HU are UV sensitive due to a role for HU in homologous recombination', *J. Bacteriol.*, vol. 180, no. 15, pp. 3750–3756, Aug. 1998.
- [205] J. A. van Kan, C. Zhang, P. Perumal Malar, and J. R. C. van der Maarel, 'High throughput fabrication of disposable nanofluidic lab-on-chip devices for single molecule studies', *Biomicrofluidics*, vol. 6, no. 3, Jul. 2012.
- [206] K. E. McGinness, T. A. Baker, and R. T. Sauer, 'Engineering controllable protein degradation', *Mol. Cell*, vol. 22, no. 5, pp. 701–707, Jun. 2006.
- [207] J. Dahl, *Gram negative cell wall*. 2008.
- [208] K. Jann and B. Jann, 'Capsules of Escherichia coli, expression and biological significance', *Can. J. Microbiol.*, vol. 38, no. 7, pp. 705–710, Jul. 1992.
- [209] T. J. Silhavy, D. Kahne, and S. Walker, 'The Bacterial Cell Envelope', *Cold Spring Harb. Perspect. Biol.*, vol. 2, no. 5, pp. a000414–a000414, May 2010.
- [210] H. Nikaido, 'Molecular Basis of Bacterial Outer Membrane Permeability Revisited', *Microbiol. Mol. Biol. Rev.*, vol. 67, no. 4, pp. 593–656, Dec. 2003.
- [211] W. Vollmer, D. Blanot, and M. A. de Pedro, 'Peptidoglycan structure and architecture', *FEMS Microbiol. Rev.*, vol. 32, no. 2, pp. 149–167, Mar. 2008.
- [212] C. W. Mullineaux, A. Nenninger, N. Ray, and C. Robinson, 'Diffusion of green fluorescent protein in three cell environments in Escherichia coli', *J. Bacteriol.*, vol. 188, no. 10, pp. 3442–3448, May 2006.
- [213] A. Shokri and G. Larsson, 'Characterisation of the Escherichia coli membrane structure and function during fedbatch cultivation', *Microb. Cell Factories*, vol. 3, no. 1, p. 9, Jul. 2004.
- [214] W. van Klompenburg and B. de Kruijff, 'The role of anionic lipids in protein insertion and translocation in bacterial membranes', *J. Membr. Biol.*, vol. 162, no. 1, pp. 1–7, Mar. 1998.
- [215] C. R. H. Raetz and W. Dowhan, 'Biosynthesis and function of phospholipids in Escherichia coli.', *J. Biol. Chem.*, vol. 265, no. 3, pp. 1235–1238, 1990.
- [216] J. M. Boggs, 'Lipid intermolecular hydrogen bonding: influence on structural organization and membrane function', *Biochim. Biophys. Acta*, vol. 906, no. 3, pp. 353–404, Oct. 1987.
- [217] Z. Zerrouk, S. Alexandre, C. Lafontaine, V. Norris, and J.-M. Valleton, 'Inner membrane lipids of Escherichia coli form domains', *Colloids Surf. B Biointerfaces*, vol. 63, no. 2, pp. 306–310, Jun. 2008.
- [218] S. J. Singer and G. L. Nicolson, 'The fluid mosaic model of the structure of cell membranes', *Science*, vol. 175, no. 4023, pp. 720–731, Feb. 1972.
- [219] G. L. Nicolson, 'The Fluid-Mosaic Model of Membrane Structure: still relevant to understanding the structure, function and dynamics of biological membranes after more than 40 years', *Biochim. Biophys. Acta*, vol. 1838, no. 6, pp. 1451–1466, Jun. 2014.
- [220] A. Ghosal *et al.*, 'The extracellular RNA complement of Escherichia coli', *MicrobiologyOpen*, Jan. 2015.
- [221] K. Nevo-Dinur, A. Nussbaum-Shochat, S. Ben-Yehuda, and O. Amster-Choder, 'Translation-independent localization of mRNA in E. coli', *Science*, vol. 331, no. 6020, pp. 1081–1084, 2011.

- [222] H. Sheng, W. T. Stauffer, R. Hussein, C. Lin, and H. N. Lim, 'Nucleoid and cytoplasmic localization of small RNAs in *Escherichia coli*', *Nucleic Acids Res.*, p. gkx023, Jan. 2017.

# Annexes



Annexe A :

*The Escherichia coli Hfq protein:  
an unattended DNA-transactions  
regulator*



# The *Escherichia Coli* Hfq Protein: An Unattended DNA-Transactions Regulator

Grzegorz M. Cech<sup>1</sup>, Agnieszka Szalewska-Pałasz<sup>1</sup>, Krzysztof Kubiak<sup>1,2,3</sup>, Antoine Malabirade<sup>2</sup>, Wilfried Grange<sup>3,4</sup>, Veronique Arluison<sup>2,4\*</sup> and Grzegorz Węgrzyn<sup>1\*</sup>

<sup>1</sup> Department of Molecular Biology, University of Gdańsk, Gdańsk, Poland, <sup>2</sup> Laboratoire Léon Brillouin, CEA, Centre National de la Recherche Scientifique, Université Paris Saclay, CEA Saclay, Gif-sur-Yvette, France, <sup>3</sup> IPCMS/Centre National de la Recherche Scientifique, Strasbourg, France, <sup>4</sup> Université Paris Diderot, UFR Science du Vivant, Paris, France

## OPEN ACCESS

### Edited by:

Manuel Espinosa,  
Spanish National Research Council,  
Spain

### Reviewed by:

Jorge Humberto Leitão,  
Universidade de Lisboa, Portugal  
Mikolaj Olejniczak,  
Adam Mickiewicz University in  
Poznań, Poland

### \*Correspondence:

Veronique Arluison  
veronique.arluison@  
univ-paris-diderot.fr  
Grzegorz Węgrzyn  
grzegorz.wegrzyn@biol.ug.edu.pl

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 27 May 2016

**Accepted:** 13 July 2016

**Published:** 28 July 2016

### Citation:

Cech GM, Szalewska-Pałasz A,  
Kubiak K, Malabirade A, Grange W,  
Arluison V and Węgrzyn G (2016) The  
*Escherichia Coli* Hfq Protein: An  
Unattended DNA-Transactions  
Regulator. *Front. Mol. Biosci.* 3:36.  
doi: 10.3389/fmolb.2016.00036

The Hfq protein was discovered in *Escherichia coli* as a host factor for bacteriophage Q $\beta$  RNA replication. Subsequent studies indicated that Hfq is a pleiotropic regulator of bacterial gene expression. The regulatory role of Hfq is ascribed mainly to its function as an RNA-chaperone, facilitating interactions between bacterial non-coding RNA and its mRNA target. Thus, it modulates mRNA translation and stability. Nevertheless, Hfq is able to interact with DNA as well. Its role in the regulation of DNA-related processes has been demonstrated. In this mini-review, it is discussed how Hfq interacts with DNA and what is the role of this protein in regulation of DNA transactions. Particularly, Hfq has been demonstrated to be involved in the control of ColE1 plasmid DNA replication, transposition, and possibly also transcription. Possible mechanisms of these Hfq-mediated regulations are described and discussed.

**Keywords:** Hfq, RNA chaperone, nucleoid associated protein, DNA replication, transposition

## INTRODUCTION

The story of the Hfq protein of *Escherichia coli* was initiated in 1968, when the roles of host factors required for replication of bacteriophage Q $\beta$  genetic material, which is RNA, have been evidenced (Franze De Fernandez et al., 1972). Subsequent studies demonstrated that there are at least two such factors (Shapiro et al., 1968). One of them was purified and identified as an RNA-binding protein (Franze De Fernandez et al., 1972). It has been named Host Factor I (HF I), as Q $\beta$  RNA synthesis *in vitro* by the phage-encoded RNA polymerase was strictly dependent on this protein (Franze De Fernandez et al., 1972). The purified protein was demonstrated to interact with single-stranded RNA, however, no binding of HF I to double-stranded RNA and to single- or double-stranded DNA was detected (Franze De Fernandez et al., 1972). The HF I name has then been replaced with Hfq (for host factor for phage Q $\beta$  replication), after cloning and sequencing the corresponding gene (Kajitani and Ishihama, 1991).

Today, the Hfq protein is known to be a major riboregulator that facilitates cellular RNA-RNA interactions. Particularly well documented is the binding of Hfq to small non-coding RNAs (sRNA) that play important roles in the regulation of gene expression at the post-transcriptional level. Enhancement of sRNA-mRNA interaction, which is facilitated by Hfq, most often inhibits translation by blocking the Shine-Dalgarno and/or start codon regions, but also affects RNA

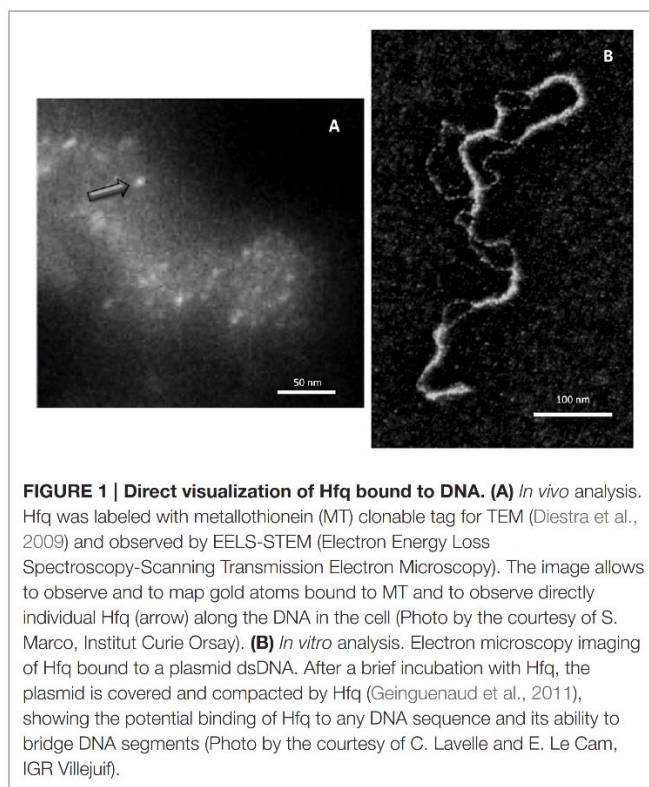
**Abbreviations:** NAP, Nucleoid Associated Protein; sRNA, Small non-coding RNA; CTR/NTR, C/N-terminal domain.

stability. This allows bacteria to adapt to their environment, especially in the case of the host infection. The multiple functions of Hfq connected to its interactions with RNA molecules have been recently reviewed in several excellent articles (Vogel and Luisi, 2011; Sobrero and Valverde, 2012; Gottesman and Storz, 2015; Updegrove et al., 2016) In this mini-review, we will focus on other Hfq activities, namely its involvement in DNA transactions.

## DIRECT INTERACTIONS BETWEEN HFQ AND DNA

Although early experiments failed to identify Hfq binding to DNA (Franze De Fernandez et al., 1972), the ability of the *hfq* gene product to interact with both supercoiled and linear plasmid DNA has been demonstrated 25 years later (Takada et al., 1997). Clearly, Hfq preferentially binds RNA molecules to DNA: while equilibrium dissociation constants ( $K_d$ ) for DNA range from nM to  $\mu$ M (Updegrove et al., 2010; Geinguenaud et al., 2011), for cellular RNA they range from tens of pM for *rpsO* polyadenylated mRNA to nM for sRNAs, such as MicA and DsrA (Folichon et al., 2003; Lease and Woodson, 2004; Fender et al., 2010). For shorter model oligonucleotides, the tightest value measured for oligoriboadenylate (rA<sub>16</sub>) was 1.4 nM, and affinity was 60 times weaker for the corresponding oligodesoxyriboadenylates (dA<sub>6</sub>) vs. oligoriboadenylates (rA<sub>6</sub>) (Link et al., 2009). Despite its apparent cellular abundance (10  $\mu$ M), Hfq low availability *in vivo* questions about its simultaneous binding to RNA and DNA (Hussein and Lim, 2011; Wagner, 2013). Nevertheless, Hfq has been shown to be one of the nucleoid-associated proteins (NAP) (Azam and Ishihama, 1999). If the presence of Hfq in the nucleoid could result from its binding to transcribed RNA, its direct binding to genomic DNA also occurs as DNA fragments are found associated with the purified protein (Updegrove et al., 2010). Note that the direct observation of Hfq in the nucleoid is possible, but difficult taking into account its abundance along the inner bacterial membrane (Azam et al., 2000; Taghbalout et al., 2014). Furthermore, Hfq nucleoid fraction represents 10–20%, while its cytoplasmic and membrane-bound fractions are about 30 and 50%, respectively (Diestra et al., 2009). This makes its observation inside the cell difficult, but its presence along the DNA *in vivo* could be confirmed by electron microscopy imaging of bacteria ultrathin-section (**Figure 1A**; Diestra et al., 2009).

*In vitro* analyses allowed measuring the  $K_d$  constants of Hfq-DNA complexes. For the complexes with single stranded dA<sub>20</sub>, the  $K_d$  was  $\sim$  200 nM, while for double stranded dA<sub>20</sub>-dT<sub>20</sub> it was  $\sim$  250 nM. This agrees with previous reports indicating that Hfq-bound DNA fragments are of curved topology (Azam and Ishihama, 1999). Moreover, low affinity was measured for dT<sub>20</sub> and dG<sub>20</sub> ( $>1$   $\mu$ M) and no complex was detected for dC<sub>20</sub> or dC<sub>20</sub>-dG<sub>20</sub> (Geinguenaud et al., 2011). On the other hand, *in vivo* analyses indicated a preferred Hfq-binding motif (<sup>A/T</sup>)T(<sup>A/G</sup>)TGCCG (Updegrove et al., 2010). The affinity of Hfq to this motif was slightly lower than to A-tracts. Intriguingly, analysis of identified Hfq-binding sequences indicated that most of them derived from genes coding for membrane proteins (Updegrove et al., 2010). The identified genes did not include any cistrons, which mRNAs were known to be regulated by Hfq,



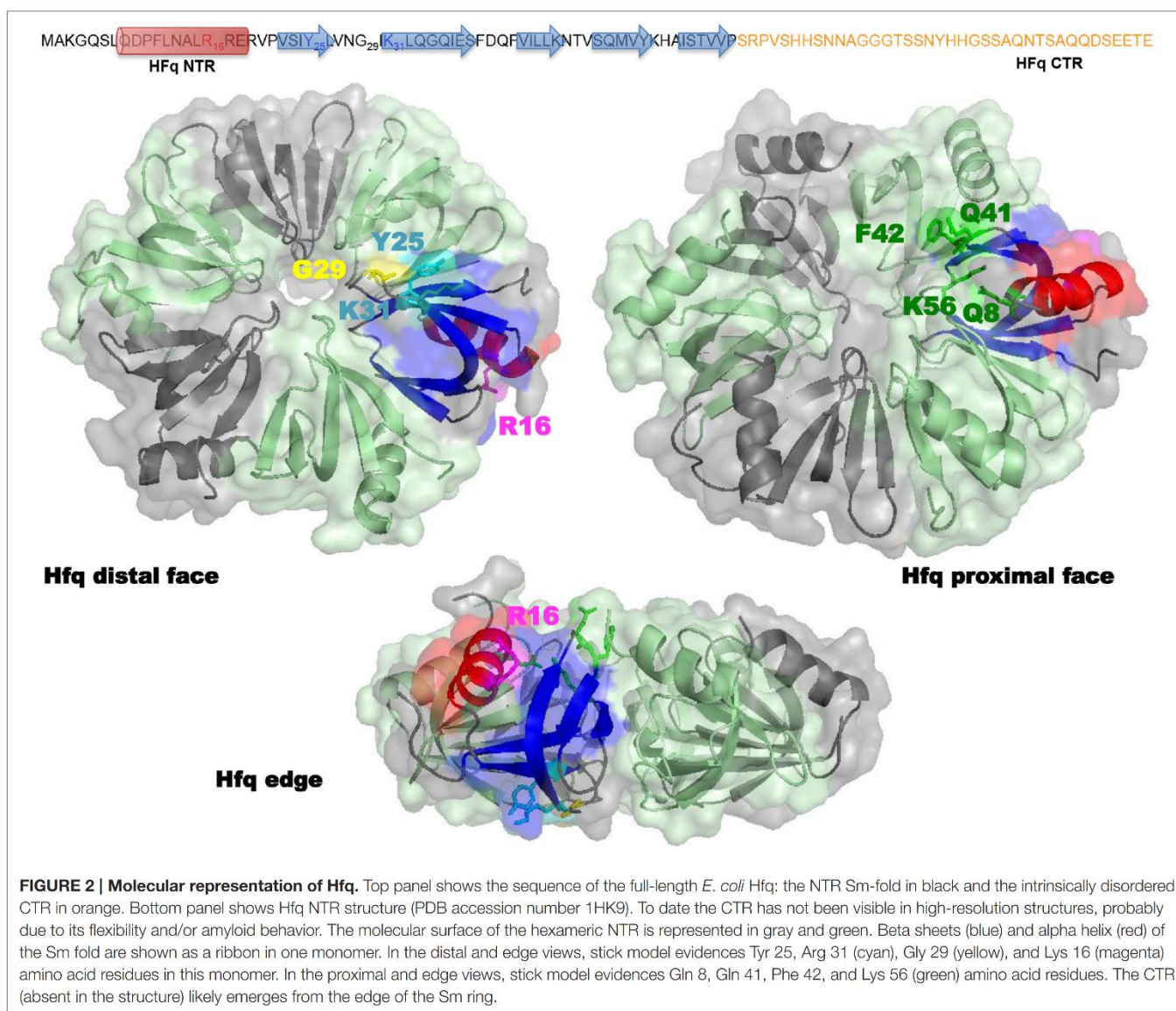
**FIGURE 1 | Direct visualization of Hfq bound to DNA. (A)** *In vivo* analysis. Hfq was labeled with metallothionein (MT) clonable tag for TEM (Diestra et al., 2009) and observed by EELS-STEM (Electron Energy Loss Spectroscopy-Scanning Transmission Electron Microscopy). The image allows to observe and to map gold atoms bound to MT and to observe directly individual Hfq (arrow) along the DNA in the cell (Photo by the courtesy of S. Marco, Institut Curie Orsay). **(B)** *In vitro* analysis. Electron microscopy imaging of Hfq bound to a plasmid dsDNA. After a brief incubation with Hfq, the plasmid is covered and compacted by Hfq (Geinguenaud et al., 2011), showing the potential binding of Hfq to any DNA sequence and its ability to bridge DNA segments (Photo by the courtesy of C. Lavelle and E. Le Cam, IGR Villejuif).

and no previously-described sRNA were encoded by these DNA fragments. This result suggests a general role for Hfq in the regulation of membrane protein expression (Guillier et al., 2006).

Despite a relative high affinity to A-tracts regions, at higher Hfq concentrations the protein interacts with DNA in a sequence-nonspecific manner, as suggested earlier (Takada et al., 1997; Azam and Ishihama, 1999). As seen by molecular imaging, Hfq binds and covers long DNA sequences (**Figure 1B**). The presence of continuous Hfq stretches, separated by naked DNA, suggests a cooperative binding mechanism by which Hfq could nucleate on high affinity sites followed by spreading along surrounding regions. This model is in agreement with previous observations showing supershifted DNA bands on gel, suggesting multiple Hfq binding to DNA when increasing Hfq concentrations (Azam and Ishihama, 1999; Updegrove et al., 2010).

Structurally, *E. coli* Hfq forms an Sm-fold in its N-terminal region ( $\sim$  65 amino acids; **Figure 2**). This fold consists of a five  $\beta$ -strands antiparallel  $\beta$ -sheet, capped by an  $\alpha$ -helix. The  $\beta$ -sheets from six monomers interact with each other to assemble in a toroidal structure with two non-equivalent faces, i.e., the proximal (on which the  $\alpha$ -helix is exposed) and distal surfaces (Link et al., 2009; **Figure 2**). It appears that the distal surface, the edge and the C-terminal region (CTR,  $\sim$  35 amino acids) of Hfq, are involved in DNA binding. Such a conclusion was made on the basis of experiments with *hfq* mutants bearing either deletions of Hfq C-terminal (CTR) domain, or point mutations, such as R16A (edge), K31A and Y25A (distal face) (Updegrove et al., 2010; **Figure 2**). While CTR is dispensable for RNA binding (Arluisson et al., 2004),





the distal surface and the edge of the protein seem to be involved in both DNA and RNA binding. Oppositely, the proximal surface seems to be involved in RNA binding only (mainly Q8, Q41, F42, K56, and H57 amino acid residues, **Figure 2**; Updegrave et al., 2010; Wang et al., 2011). This makes sense taking into account adenylate binding to the distal face, while the proximal side is more dedicated to uridine-rich sequences, absent in DNA (Link et al., 2009; Schu et al., 2015). The preference of Hfq for A-rich RNA over DNA is explained by the formation of a hydrogen bond between the ribosyl 2' hydroxyl group and the carbonyl oxygen of residue Gly 29 (Link et al., 2009; **Figure 2**). While dispensing for DNA binding, it was proposed that the C-terminal domain might anchor the Hfq protein to DNA, while the distal face of the Sm-core could be required to direct interactions with the nucleic acid.

Indeed, the involvement of Hfq CTR in DNA binding has also been recently confirmed by low resolution structure of Hfq:DNA

complex (Jiang et al., 2015). Previous electron microscopy imaging evidenced that Hfq acts by bridging together two DNA molecules (**Figure 1B**; Geinguenaud et al., 2011). Such an activity was formerly documented for the H-NS (histone-like nucleoid structuring) NAP. A model for Hfq propensity for bridging has thus been proposed to be linked to its CTR-arm. This interaction results in a change of the mechanical properties of the double helix, and in a compaction of DNA into a condensed form (Jiang et al., 2015). Recently, a new unexpected property of Hfq CTR-arm has also been evidenced. Indeed, the CTR region, which was considered as intrinsically unstructured (Vogel and Luisi, 2011), contains an amyloid sequence that allows the protein to self-assemble (Arлуison et al., 2006; Fortas et al., 2015). This new property of the CTR could explain its self-assembly and spreading of Hfq on DNA, but this needs to be investigated further.

Finally, one important question remains about the precise effect of Hfq on DNA *in vivo*. Hfq has been proposed to change DNA superhelicity, precisely plasmids purified in the absence of



Hfq are less negatively supercoiled during the stationary phase (Tsui et al., 1994). Nevertheless, this property has never been analyzed in detail. Sugar re-puckering has also been described but should be confirmed by *in vivo* analyses (Geinguenaud et al., 2011).

As for the relative amount of Hfq compared to other NAP, its abundance in actively growing cells is similar to those of the most abundant NAPs, Fis (factor for inversion simulation) or HU (heat-unstable protein). Nevertheless, Hfq quantity in the nucleoid is lower than for HU and Fis, which are present in the nucleoid only. A rough estimation indicate that Hfq represents about 5% of the total nucleoid-associated proteins (vs. 20 and 40% for Fis and HU during the exponential phase of growth, respectively, Talukder and Ishihama, 2015). It is now established that Hfq concentration increases when reaching the stationary phase (Tsui et al., 1997; Diestra et al., 2009; Cech et al., 2014), even if Dps (DNA-binding protein from starved cells) is the main nucleoid-associated protein during this phase of growth. Hfq nucleoid-bound fraction remains more or less constant during the cell cycle (Talukder and Ishihama, 2015). Furthermore, it appears that Hfq can not only interact with DNA, but also: (i) interacts with some NAP, such as H-NS (Kajitani and Ishihama, 1991), (ii) cooperates in the organization of the bacterial chromosome with other proteins, like Fis, HU, H-NS, IHF (integration host factor protein) and StpA (suppressor of *td* mutant phenotype A; Ohniwa et al., 2013) and (iii) regulates the expression of other NAPs (Lease and Belfort, 2000; Lu et al., 2016). The proteins associated to the nucleoid are usually divided into two groups depending on whether they bridge or bend DNA (Gruber, 2014). While HU, IHF and Fis belong to the bending group of NAP that causes local folding of DNA, H-NS and Hfq belong to the bridging group that organize large parts of chromosomes into isolated domains (Dorman, 2009; Geinguenaud et al., 2011).

Structurally, the nucleoid structure during the exponential growth of *E. coli* is described as a chromatin-like fibers ranging from 5 to 80 nm in diameter (Ohniwa et al., 2013). Atomic force microscopic analyses indicate that over 60% of the fiber structures enter into the "thin" category (i.e., 5–20 nm), and that this chromatin-like structure is significantly condensed upon entering into the stationary phase. The absence of some NAPs can change the relative abundance of "thin" vs. "thick" fibers (Ohniwa et al., 2013). The absence of Hfq results in an increase of both "thin" and "thick" fiber populations. None of the single deletions of NAP were described to cause a significant change in the fraction of the fibers population and more likely the lack of only one nucleoid-associated proteins can be compensated by functions of other proteins from the same group (Ohniwa et al., 2013). Indeed, whether the bridging Hfq and H-NS could replace each other is still unknown.

## INDIRECT INVOLVEMENT OF HFQ IN DNA-RELATED PROCESSES

The fact that Hfq can interact with DNA and change its properties, inspired studies on the putative role of this protein in various DNA transactions. However, to date, only two processes

were investigated in more depth, replication and transposition. Some clues tend to indicate an effect in transcription, but the role of Hfq in this process is still an open question.

## Replication

The role of Hfq in the regulation of DNA replication has been investigated to date only using the models of plasmid replicons. Employing wild-type *E. coli* strain and otherwise isogenic *hfq* mutant, the efficiency of replication of ColE1-like (pMB1—and p15A) and bacteriophage  $\lambda$ -derived plasmids was investigated. In bacteria devoid of Hfq, significant differences in plasmid amount and kinetics of plasmid DNA synthesis were observed relative to wild-type cells, but only for ColE1-like plasmids, not for  $\lambda$  replicons (Cech et al., 2014).

Levels of the Hfq protein in the wild-type strain were increased at the late exponential and early stationary phases of bacterial culture relative to the early exponential phase. Accordingly, ColE1-like plasmids replicated in the *hfq* mutant more efficiently than in the wild-type bacteria at late exponential and early stationary phases, while less efficiently at the early exponential phase (Cech et al., 2014). Thus, the differences between wild-type and *hfq* mutant hosts were the most pronounced under conditions corresponding to the highest levels of Hfq in wild-type bacteria. Interestingly, effects of the *hfq* deletion on ColE1-like plasmid replication were impaired in the absence of the *rom* gene (Cech et al., 2014). The regulation of replication of ColE1-like plasmids is based on the action of an anti-sense RNA, called RNA I, which impairs the priming reaction by binding to pre-primer RNA, called RNA II. Since *rom* codes for a protein responsible for enhancing RNA I-RNA II interactions, it was hypothesized that Hfq might either modulate these interactions or interplay with Rom to influence the negative regulation facilitated by this protein (Cech et al., 2014).

In the  $\lambda$  replicons, which are not affected by the absence of Hfq, no RNA-RNA interactions are required for the replication regulation. Although transcription plays a crucial role in the replication initiation from *ori $\lambda$* , it appears that Hfq does not modulate the transcriptional activation of this site (Cech et al., 2014).

## Transposition

Hfq influences many cellular processes at the post-transcriptional level mostly by regulating and promoting interactions between RNAs. In addition to direct interactions of Hfq with DNA, indirect effects of the protein on the DNA-related processes have been reported. These include the regulation of transposition process in several known transposon systems in bacteria: Tn10, Tn5, and IS200. In all of these systems, Hfq is a potent inhibitor of transposition as shown in experiments with *hfq*-deficient mutant strains. However, the mode of Hfq action in transposition varies among these systems (Ellis and Haniford, 2016).

For Tn10, Hfq exerts its function typically by promoting base-pairing between transposase RNA (RNA-IN) and its antisense RNA (RNA-OUT) encoded by Tn10/IS10 antisense system (Ross et al., 2013). As a result, downregulation of RNA-IN expression *in vivo* occurs. In addition to this effect, Hfq can inhibit transposase expression independently of the antisense

system. It was proposed that Hfq binds directly to RNA-IN, blocking its translation (Ross et al., 2010; Ellis and Haniford, 2016).

Hfq-mediated regulation of transcription of the gene coding for the IS200 transposase might also explain the inhibition of Tn5 transposition (see Transcription; Haniford and Ellis, 2015). Moreover, Crp (cyclic AMP receptor protein) was proposed to mediate Hfq effect on Tn5 transposition (Ross et al., 2014). This, together with the known role of Hfq in stress response processes, supports the hypothesis of linking of transposition to the physiological status of the cell.

IS200 elements, abundant in Enterobacteriaceae, transpose with a very low frequency and are limited by the expression on the transposase gene, *tnpA*. In addition, transposase expression is downregulated by antisense RNA, *art200* (Ellis et al., 2015). However, the role of Hfq in the regulation of IS200 transposition does not involve pairing between mRNA and *art200*. It was shown that Hfq binds upstream of SD sequence of *tnpA* mRNA, inhibiting ribosome binding (Ellis et al., 2015).

## Transcription

Finally, few studies signified a role for Hfq in transcription. So far, how Hfq affects transcription remains unclear. Direct effects could exist as Fourier spectroscopy (FTIR) technique allowed to observe a partial unwinding of a DNA double helix at AT-rich regions by the Hfq protein (Geinguenaud et al., 2011). This opens the discussion on a possible role for this protein in the regulation of transcription initiation. A role for Hfq in the modulation of transcription elongation has also been proposed. According to the hypothesis, interaction of Hfq with a nascent transcript would help to overcome transcription pauses in order to prevent preliminary transcript release (Le Derout et al., 2010). Independently, indirect effects could also occur. Indeed, Hfq function may also be mediated by protein-protein contact and its interaction with RNA polymerase or Rho (a transcription termination factor) have been described (Sukhodolets and Garges, 2003; Rabhi et al., 2011). Transcriptional control by Hfq still remains largely unexplored.

## REFERENCES

- Arluison, V., Folichon, M., Marco, S., Derreumaux, P., Pellegrini, O., Seguin, J., et al. (2004). The C-terminal domain of *Escherichia coli* Hfq increases the stability of the hexamer. *Eur. J. Biochem.* 271, 1258–1265. doi: 10.1111/j.1432-1033.2004.04026.x
- Arluison, V., Mura, C., Guzman, M. R., Liquier, J., Pellegrini, O., Gingery, M., et al. (2006). Three-dimensional structures of fibrillar sm proteins: Hfq and other sm-like proteins. *J. Mol. Biol.* 356, 86–96. doi: 10.1016/j.jmb.2005.11.010
- Azam, T. A., Hiraga, S., and Ishihama, A. (2000). Two types of localization of the DNA-binding proteins within the *Escherichia coli* nucleoid. *Genes Cells* 5, 613–626. doi: 10.1046/j.1365-2443.2000.00350.x
- Azam, T. A., and Ishihama, A. (1999). Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. *J. Biol. Chem.* 274, 33105–33113. doi: 10.1074/jbc.274.46.33105

## CONCLUSIONS

Although the Hfq protein has been discovered as an RNA-interacting factor, and investigated in this light for many years, it also binds DNA and affects significantly the structure of this nucleic acid. Experimental evidence that Hfq is involved directly or indirectly in different DNA transactions exists, even if molecular mechanisms of these regulatory processes are still poorly understood. Nevertheless, Hfq-mediated control of ColE1-like plasmids appears to be specific, thus, particular mechanisms can be recognized in forthcoming studies. Regulation of transposition and transcription by Hfq might be less complicated than the function of this protein in the control of DNA replication. Regulations at both the post-transcriptional and the DNA-structuring levels allow bacteria to adapt to their environment, with important consequences for its physiology (including virulence). The physiological impacts of Hfq-DNA interactions *in vivo* thus need further investigations.

## AUTHOR CONTRIBUTIONS

GC drafted a part of chapter 2, and chapters 3.1 and 3.3, and participated in preparation of the final manuscript. AS wrote chapter 3.2. KK and AM participated in the general assembly of the text. WG participated in the biophysical description of Hfq. VA drafted chapter 2, prepared figures, and participated in preparation of the final version of the manuscript. GW elaborated the concept of the manuscript, drafted chapters 1 and 4, and prepared the final version of the paper. VA and GW contributed equally to this work.

## ACKNOWLEDGMENTS

This work was supported by the National Science Center (Poland) grant no. 2012/04/M/NZ1/00067 (GW), the POLONIUM program grant no. 33547NE, the joint research projects USPC/NUS RL2015-106 (VA) and by the Institut français de Pologne (KK). We thank PICT-Ibisa for access to TEM facilities.

- Cech, G. M., Pakula, B., Kamrowska, D., Wegrzyn, G., Arluison, V., and Szalewska-Palasz, A. (2014). Hfq protein deficiency in *Escherichia coli* affects ColE1-like but not lambda plasmid DNA replication. *Plasmid* 73, 10–15. doi: 10.1016/j.plasmid.2014.04.005
- Diestra, E., Cayrol, B., Arluison, V., and Risco, C. (2009). Cellular electron microscopy imaging reveals the localization of the Hfq protein close to the bacterial membrane. *PLoS ONE* 4:e8301. doi: 10.1371/journal.pone.0008301
- Dorman, C. J. (2009). Nucleoid-associated proteins and bacterial physiology. *Adv. Appl. Microbiol.* 67, 47–64. doi: 10.1016/S0065-2164(08)01002-2
- Ellis, M. J., and Haniford, D. B. (2016). Riboregulation of bacterial and archaeal transposition. *Wiley Interdiscip. Rev. RNA* 7, 382–398. doi: 10.1002/wrna.1341
- Ellis, M. J., Trussler, R. S., and Haniford, D. B. (2015). A cis-encoded sRNA, Hfq and mRNA secondary structure act independently to suppress IS200 transposition. *Nucleic Acids Res.* 43, 6511–6527. doi: 10.1093/nar/gkv584
- Fender, A., Elf, J., Hampel, K., Zimmermann, B., and Wagner, E. G. (2010). RNAs actively cycle on the Sm-like protein Hfq. *Genes Dev.* 24, 2621–2626. doi: 10.1101/gad.591310



- Folichon, M., Arluison, V., Pellegrini, O., Huntzinger, E., Regnier, P., and Hajnsdorf, E. (2003). The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. *Nucleic Acids Res.* 31, 7302–7310. doi: 10.1093/nar/gkg915
- Fortas, E., Piccirilli, F., Malabirade, A., Militello, V., Trépout, S., Marco, S., et al. (2015). New insight into the structure and function of Hfq C-terminus. *Biosci. Rep.* 35:e00190. doi: 10.1042/BSR20140128
- Franze De Fernandez, M. T., Hayward, W. S., and August, J. T. (1972). Bacterial proteins required for replication of phage Qb ribonucleic acid. *J. Biol. Chem.* 247, 824–821.
- Geinguenaud, F., Calandrini, V., Teixeira, J., Mayer, C., Liquier, J., Lavelle, C., et al. (2011). Conformational transition of DNA bound to Hfq probed by infrared spectroscopy. *Phys. Chem. Chem. Phys.* 13, 1222–1229. doi: 10.1039/C0CP01084G
- Gottesman, S., and Storz, G. (2015). RNA reflections: converging on Hfq. *RNA* 21, 511–512. doi: 10.1261/rna.050047.115
- Gruber, S. (2014). Multilayer chromosome organization through DNA bending, bridging and extrusion. *Curr. Opin. Microbiol.* 22, 102–110. doi: 10.1016/j.mib.2014.09.018
- Guillier, M., Gottesman, S., and Storz, G. (2006). Modulating the outer membrane with small RNAs. *Genes Dev.* 20, 2338–2348. doi: 10.1101/gad.1457506
- Haniford, D. B., and Ellis, M. J. (2015). Transposons Tn10 and Tn5. *Microbiol. Spectr.* 3:MDNA3-0002-2014. doi: 10.1128/microbiolspec.mdna3-0002-2014
- Hussein, R., and Lim, H. N. (2011). Disruption of small RNA signaling caused by competition for Hfq. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1110–1115. doi: 10.1073/pnas.1010082108
- Jiang, K., Zhang, C., Guttula, D., Liu, F., Van Kan, J. A., Lavelle, C., et al. (2015). Effects of Hfq on the conformation and compaction of DNA. *Nucl. Acids Res.* 43, 4332–4341. doi: 10.1093/nar/gkv268
- Kajitani, M., and Ishihama, A. (1991). Identification and sequence determination of the host factor gene for bacteriophage Qb. *Nucl. Acids Res.* 9, 1063–1066. doi: 10.1093/nar/19.5.1063
- Lease, R. A., and Belfort, M. (2000). Riboregulation by DsrA RNA: transactions for global economy. *Mol. Microbiol.* 38, 667–672. doi: 10.1046/j.1365-2958.2000.02162.x
- Lease, R. A., and Woodson, S. A. (2004). Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. *J. Mol. Biol.* 344, 1211–1223. doi: 10.1016/j.jmb.2004.10.006
- Le Derout, J., Boni, I. V., Regnier, P., and Hajnsdorf, E. (2010). Hfq affects mRNA levels independently of degradation. *BMC Mol. Biol.* 11:17. doi: 10.1186/1471-2199-11-17
- Link, T. M., Valentin-Hansen, P., and Brennan, R. G. (2009). Structure of *Escherichia coli* Hfq bound to polyribadenylate RNA. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19292–19297. doi: 10.1073/pnas.0908744106
- Lu, P., Wang, Y., Zhang, Y., Hu, Y., Thompson, K. M., and Chen, S. (2016). RpoS-dependent sRNA RgsA regulates Fis and AcpP in *Pseudomonas aeruginosa*. *Mol. Microbiol.* doi: 10.1111/mmi.13458. [Epub ahead of print].
- Ohniwa, R. L., Muchaku, H., Saito, S., Wada, C., and Morikawa, K. (2013). Atomic force microscopy analysis of the role of major DNA-binding proteins in organization of the nucleoid in *Escherichia coli*. *PLoS ONE* 8:e72954. doi: 10.1371/journal.pone.0072954
- Rabhi, M., Espeli, O., Schwartz, A., Cayrol, B., Rahmouni, A. R., Arluison, V., et al. (2011). The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *EMBO J.* 30, 2805–2816. doi: 10.1038/emboj.2011.192
- Ross, J. A., Ellis, M. J., Hossain, S., and Haniford, D. B. (2013). Hfq restructures RNA-IN and RNA-OUT and facilitates antisense pairing in the Tn10/IS10 system. *RNA* 19, 670–684. doi: 10.1261/rna.037747.112
- Ross, J. A., Trussler, R. S., Black, M. D., McLellan, C. R., and Haniford, D. B. (2014). Tn5 transposition in *Escherichia coli* is repressed by Hfq and activated by over-expression of the small non-coding RNA SgrS. *Mob. DNA* 5, 27. doi: 10.1186/s13100-014-0027-z
- Ross, J. A., Wardle, S. J., and Haniford, D. B. (2010). Tn10/IS10 transposition is downregulated at the level of transposase expression by the RNA-binding protein Hfq. *Mol. Microbiol.* 78, 607–621. doi: 10.1111/j.1365-2958.2010.07359.x
- Schu, D. J., Zhang, A., Gottesman, S., and Storz, G. (2015). Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. *EMBO J.* 34, 2557–2573. doi: 10.15252/embj.201591569
- Shapiro, L., Franze De Fernandez, M. T., and August, J. T. (1968). Resolution of two factors required in the Q-beta-RNA polymerase reaction. *Nature* 220, 478–480. doi: 10.1038/220478a0
- Sobrero, P., and Valverde, C. (2012). The bacterial protein Hfq: much more than a mere RNA-binding factor. *Crit. Rev. Microbiol.* 38, 276–299. doi: 10.3109/1040841X.2012.664540
- Sukhodolets, M. V., and Garges, S. (2003). Interaction of *Escherichia coli* RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. *Biochemistry* 42, 8022–8034. doi: 10.1021/bi020638i
- Taghbalout, A., Yang, Q., and Arluison, V. (2014). The *Escherichia coli* RNA processing and degradation machinery is compartmentalized within an organized cellular network. *Biochem. J.* 458, 11–22. doi: 10.1042/BJ20131287
- Takada, A., Wachi, M., Kaidow, A., Takamura, M., and Nagai, K. (1997). DNA binding properties of the *hfq* gene product of *Escherichia coli*. *Biochem. Biophys. Res. Com.* 236, 576–579. doi: 10.1006/bbrc.1997.7013
- Talukder, A., and Ishihama, A. (2015). Growth phase dependent changes in the structure and protein composition of nucleoid in *Escherichia coli*. *Sci. China Life Sci.* 58, 902–911. doi: 10.1007/s11427-015-4898-0
- Tsui, H. C., Feng, G., and Winkler, M. E. (1997). Negative regulation of *mutS* and *mutH* repair gene expression by the Hfq and RpoS global regulators of *Escherichia coli* K-12. *J. Bacteriol.* 179, 7476–7487.
- Tsui, H. C., Leung, H. C., and Winkler, M. E. (1994). Characterization of broadly pleiotropic phenotypes caused by an *hfq* insertion mutation in *Escherichia coli* K-12. *Mol. Microbiol.* 13, 35–49. doi: 10.1111/j.1365-2958.1994.tb00400.x
- Updegrave, T. B., Correia, J. J., Galletto, R., Bujalowski, W., and Wartell, R. M. (2010). *E. coli* DNA associated with isolated Hfq interacts with Hfq's distal surface and C-terminal domain. *Biochim. Biophys. Acta* 1799, 588–596. doi: 10.1016/j.bbagr.2010.06.007
- Updegrave, T. B., Zhang, A., and Storz, G. (2016). Hfq: the flexible RNA matchmaker. *Curr. Opin. Microbiol.* 30, 133–138. doi: 10.1016/j.mib.2016.02.003
- Vogel, J., and Luisi, B. F. (2011). Hfq and its constellation of RNA. *Nat. Rev. Microbiol.* 9, 578–589. doi: 10.1038/nrmicro2615
- Wagner, E. G. (2013). Cycling of RNAs on Hfq. *RNA Biol.* 10, 619–626. doi: 10.4161/rna.24044
- Wang, W., Wang, L., Zou, Y., Zhang, J., Gong, Q., Wu, J., et al. (2011). Cooperation of *Escherichia coli* Hfq hexamers in DsrA binding. *Genes Dev.* 25, 2106–2117. doi: 10.1101/gad.16746011

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Cech, Szalewska-Palasz, Kubiak, Malabirade, Grange, Arluison and Węgrzyn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Annexe B :

*New insight into the structure  
and function of Hfq carboxyl  
terminus*





## OPEN ACCESS

# New insight into the structure and function of Hfq C-terminus

Emilie Fortas\*†, Federica Piccirilli‡, Antoine Malabirade\*†¶, Valeria Militello‡, Sylvain Trépout§||, Sergio Marco§||, Aziz Taghbalout¶|| and Véronique Arluison\*†\*\*<sup>1</sup>

\*Laboratoire Léon Brillouin, CEA – Centre de Saclay, 91191 Gif-sur-Yvette, France

†UMR 12 CEA/CNRS, 91191 Gif-sur-Yvette, France

‡Department of Physics and Chemistry, University of Palermo, Viale delle Scienze, Ed. 18, 90128 Palermo, Italy

§Institut Curie, Centre de Recherche, Campus Universitaire d'Orsay, bât 112, 91405 Orsay Cedex, France

||INSERM U1196, Campus Universitaire d'Orsay, bât 112, 91405 Orsay Cedex, France

¶Université Paris Sud, 91405 Orsay Cedex, France

¶¶Department of Molecular Biology and Biophysics, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, CT 06032, U.S.A.

\*\*Univ Paris Diderot, Sorbone Paris Cité, 75013 Paris, France

## Synopsis

Accumulating evidence indicates that RNA metabolism components assemble into supramolecular cellular structures to mediate functional compartmentalization within the cytoplasmic membrane of the bacterial cell. This cellular compartmentalization could play important roles in the processes of RNA degradation and maturation. These components include Hfq, the RNA chaperone protein, which is involved in the post-transcriptional control of protein synthesis mainly by the virtue of its interactions with several small regulatory ncRNAs (sRNA). The *Escherichia coli* Hfq is structurally organized into two domains. An N-terminal domain that folds as strongly bent  $\beta$ -sheets within individual protomers to assemble into a typical toroidal hexameric ring. A C-terminal flexible domain that encompasses approximately one-third of the protein seems intrinsically unstructured. RNA-binding function of Hfq mainly lies within its N-terminal core, whereas the function of the flexible domain remains controversial and largely unknown. In the present study, we demonstrate that the Hfq-C-terminal region (CTR) has an intrinsic property to self-assemble into long amyloid-like fibrillar structures *in vitro*. We show that normal localization of Hfq within membrane-associated coiled structures *in vivo* requires this C-terminal domain. This finding establishes for the first time a function for the hitherto puzzling CTR, with a plausible central role in RNA transactions.

**Key words:** amyloid fibrils, cellular compartmentalization, post-transcriptional regulation, ribonucleic acid (RNA) processing and degradation, small non-coding ribonucleic acid (RNA), sub-membrane macromolecular assembly.

Cite this article as: Bioscience Reports (2015) 35, e00190, doi:10.1042/BSR20140128

## INTRODUCTION

A bacterial cell has multiple protein assemblies, which form higher order cellular structures and play important roles during cellular division [1]. We have shown that another family of proteins, those involved in RNA metabolism, also forms cellular assemblies organized into a membrane-associated framework, whose cellular pattern resembles previously described long-range order bacterial cytoskeletal-like structures [2–5]. This cellular framework includes the RNA chaperone Hfq (Host Factor for phage Q beta RNA replication), an abundant and conserved protein found in a variety of bacteria [2,6]. One of

the best-characterized functions of Hfq is to modulate post-transcriptionally gene expression via its interaction with various nucleic acids, including regulatory non-coding RNAs (ncRNAs) [6,7]. Currently, approximately 30000 sequences of ncRNA from different species have been identified and among them, those from bacteria are in average 100 nts long, hence their name of small ncRNA (sRNA) [8,9]. The sRNA often functions by base-pairing within regions around the translation initiation signal of associated mRNA target and therefore acts on both mRNA translation and mRNA stability [10]. Annealing of the regulatory sRNA to its cognate mRNA requires Hfq [11,12]. Due to the diversity of mRNA targets of sRNA, Hfq is thus involved in diverse metabolic pathways such as sugar transport, membrane remodelling

**Abbreviations:** CTR, Hfq C-terminal region; FSD, Fourier self-deconvolution; FTIR, Fourier transform infrared spectroscopy; HfqCTR<sub>p</sub>, Hfq C-terminal peptide; IDP, intrinsically-disordered proteins; ncRNA, regulatory non-coding RNA; NTR, N-terminal region; sRNA, small non-coding RNA; ThT, thioflavin T; WT, wild-type.

<sup>1</sup> To whom correspondence should be addressed (email Veronique.Arluison@univ-paris-diderot.fr).

or quorum sensing [13–15]. Indeed, Hfq is a pleiotropic regulator and nearly a half of Hfq-bound sRNA regulates the expression of membrane proteins or proteins involved in membrane-related cellular processes [16,17]. The cellular compartmentalization of Hfq within the cell periphery could thus play a major role in regulating the expression of these membrane proteins.

The 66 kDa homo-hexameric Hfq protein belongs to Sm-like family of proteins [18]. Sm or Sm-like proteins are commonly made of approximately 100 amino acids that typically fold into a bent five-stranded antiparallel  $\beta$ -sheet capped by an N-terminal  $\alpha$ -helix. The individual  $\beta$ -strands of adjacent monomers assemble into a continuous intermolecular antiparallel  $\beta$ -sheet to form a characteristic toroid-shaped structure with a charged pore [19]. Though Hfq differs from Sm proteins in a way that Sm proteins fold into heptameric toroid [18], Hfq and Sm proteins have in common the ability to polymerize into well-ordered fibres *in vitro* [20,21].

As for Sm proteins, it has been possible to obtain 3D-structures for Hfq of various bacteria [22–25]. Nevertheless, most of Hfq 3D structures lack the C-terminal region (CTR) and include only the N-terminal region (NTR) of the protein, i.e. the Sm-like domain. This NTR (Sm) region presents two main RNA-binding sites: one on the proximal face (the N-terminal  $\alpha$ -helix containing face) with a strong specificity for U-rich RNA and the second RNA-binding site on the opposite distal face with specificity for polyA [26]. Other Hfq regions such as the lateral surface or CTR could also play a role in RNA recognition and binding [27].

The determination of a high-resolution structure for Hfq CTR that makes one-third of the size of *Escherichia coli* Hfq remains elusive [28]. The CTR seems to extend outside the Sm-core and to be intrinsically non-structured [29,30]. This domain has been shown to firm up the subunit interface of the toroidal-shaped N-terminal domain hexamer and is presumably important for the protein stability [29,31]. Hfq-mediated sRNA regulation has been primarily evidenced in bacteria containing Hfq proteins with a CTR extension. However, the role of the CTR in the Hfq-based riboregulation is controversial. Hfq that lack CTR binds some RNA and mediates efficiently the associated riboregulation [31,32]. Conversely, the lack of the CTR could affect other sRNA-mediated regulations [28,29,33,34] and some residues within the CTR have been shown to interact with several RNA species [28,35].

In the present work, we demonstrate that the CTR of Hfq has an unexpected and outstanding property, as it is responsible for the self-assembly of the protein into long fibrils. This feature is correlated to the formation of supramolecular structures by the protein close to the bacterial inner membrane *in vivo*. This finding opens new perspectives on the role of this puzzling C-terminal extension in Hfq-dependent riboregulation.

## EXPERIMENTAL

### *E. coli* strains and plasmids

MC1000  $\Delta hfq$  strain (AT165) was constructed as described previously in Taghbalout et al. [2]. Full-length and truncated Hfq pro-

teins (Hfq-NTR amino acid residues 1–72 and Hfq-CTR amino acid residues 65–102) were expressed from a plasmid under the control of the arabinose-inducible  $P_{BAD}$  promoter. This plasmid has been previously described in Sledjeski et al. [36]. The Hfq-NTR and Hfq-CTR plasmid constructs derived from this plasmid were made with QuickChange mutagenesis kit (Agilent) by replacing the seventy-third codon with a translation stop signal (for pHfq-NTR) or by deleting the sequence between the second and sixty-fourth codons (for pHfq-CTR). The plasmids were transformed into MC1000  $\Delta hfq$  strain and the expression of the proteins was induced by using a wide range of arabinose concentrations including conditions that gave normal levels of Hfq full-length expression, as previously shown by quantitative Western blot [37].

### Hfq C-terminal peptide synthesis and preparation

The peptide corresponding to the CTR domain (residues 64–102) is referred as HfqCTR<sub>p</sub> (Hfq C-terminal peptide) throughout the manuscript. The HfqCTR<sub>p</sub> (SRPVSHHSNNAGGGTSSNYH-HGSSAQNTSAQQDSEETE) was synthesized by Genosphere Biotech (France) using Fmoc chemistry and solid-support resin. After cleavage, the peptide was purified by reverse phase chromatography with an octyl carbon chain (C<sub>8</sub>)-bonded silica column and lyophilized.

HfqCTR<sub>p</sub> peptide was reconstituted in phosphate buffer 100 mM, pH 7, at 20 mg/ml. This buffer could be either hydrogenated or deuterated for FTIR (Fourier transform infrared spectroscopy) analyses. The self-assembly of HfqCTR<sub>p</sub> into large fibres is not instantaneous. The time needed to observe aggregation is dependent on the temperature and ranges from few days at 37 °C to 2 weeks at 4 °C.

### EM analyses of Hfq peptides

#### Sample preparation

A 5  $\mu$ l drop of HfqCTR<sub>p</sub> sample at 1 mg/ml in phosphate buffer 50 mM, pH7, (i.e., stock solution diluted 1:20 in water) was deposited on a glow discharged carbon-coated EM copper grid (200 mesh square grid, EMS). After 5 min, the excess of sample was blotted out using a Whatman filter paper and then the grid was washed with water three times to eliminate phosphate buffer, which would react with uranyl acetate used for negative staining. To perform negative staining, 5  $\mu$ l of uranyl acetate solution (2%) was applied on to the grid containing peptidic sample. After 30-s incubation, the excess of uranyl acetate was blotted out and then the grids were kept in a dry, dark, dust-free environment until observation with the electron microscope.

#### Sample observation

The EM grid was then mounted on to a room temperature-equilibrated holder and subsequently introduced into a JEOL 2200FS electron microscope (JEOL). Images (2048  $\times$  2048 pixels) were acquired using a ssCCD camera (Gatan UltraScan



894) at 40 000 $\times$  (nominal magnification, the corresponding pixel size was 0.32 nm).

### FTIR spectroscopy analysis of HfqCTR<sub>p</sub>

Infrared absorption measurements have been performed in the region between 4000 and 400  $\text{cm}^{-1}$  through the use of a Vertex70 interferometer coupled with a conventional IR source and a Triglycine sulfate (DTGS) detector. The HfqCTR<sub>p</sub> fibrils sample (20 mg/ml) has been loaded into a conventional liquid cell equipped with CaF<sub>2</sub> windows and a Teflon spacer 6- $\mu\text{m}$  thick. A total of 256 scans were collected and five independent measurements were averaged. Spectra were baseline corrected and background subtracted before analysis. Qualitative evaluation of secondary structure has been performed through the combination of second derivative and Fourier self-deconvolution (FSD) analyses [38]. Overlapping peaks composing the amide I band were fitted with Voigt profiles with a bandwidth varying between 15 and 25  $\text{cm}^{-1}$ .

### Confocal fluorescence microscopy analysis of HfqCTR<sub>p</sub>

HfqCTR<sub>p</sub> sample was diluted 1:100 and stained with thioflavin T (ThT), a benzothiazole salt which it is widely used to visualize and quantify the presence of amyloid-like aggregates, both *in vitro* and *in vivo* [39]. Ten microlitre aliquots of stained HfqCTR<sub>p</sub> samples were placed on microscope slides and imaged at 1024  $\times$  1024 pixel resolution using a Leica RCS SP5 confocal laser scanning microscope with a 63 $\times$  oil objective numerical aperture (NA) = 1.4 (Leica Microsystems), at a scanning frequency of 400 Hz. The laser excitation was set at 415 nm and emission detected in the range 450–500 nm.

### *In vivo* immunofluorescence microscopy

Immunofluorescence (IF) experiments were done as described previously [2,40]. Briefly, wild-type (WT) MC1000 [41], MC1000  $\Delta hfq/P_{ara-hfq}$  or MC1000  $\Delta hfq/P_{ara-hfq}^{-72}$  cells were fixed in the growth medium for 30 min at room temperature in the presence of 2% formaldehyde and 0.016% glutaraldehyde and then adsorbed on silane-coated cover slips. Cells were stained with purified rat anti-Hfq antibody [2]. Rat antiserum directed against untagged-Hfq (Eurogentech) was purified by absorption to purified Hfq bound to a PVDF membrane followed by an elution with 0.2 M glycine (pH 2) and a renaturation with 1.5 M Tris-base (pH 8.8) [2]. Alexa fluor 488-conjugated goat anti-rat secondary antibodies (Molecular Probes) were used to detect the stained Hfq cellular proteins. Images were collected using the Volocity image acquisition program (Perkin–Elmer). Images were not subjected to deconvolution and were captured at 520–550 nm using the filter set # 89006 (Chroma Technology). Cells (250–300) were analysed for each condition and the described localization patterns were present in most of the cells. Image processing was done using Volocity (Perkin–Elmer) and Photoshop (Adobe) softwares.

## RESULTS

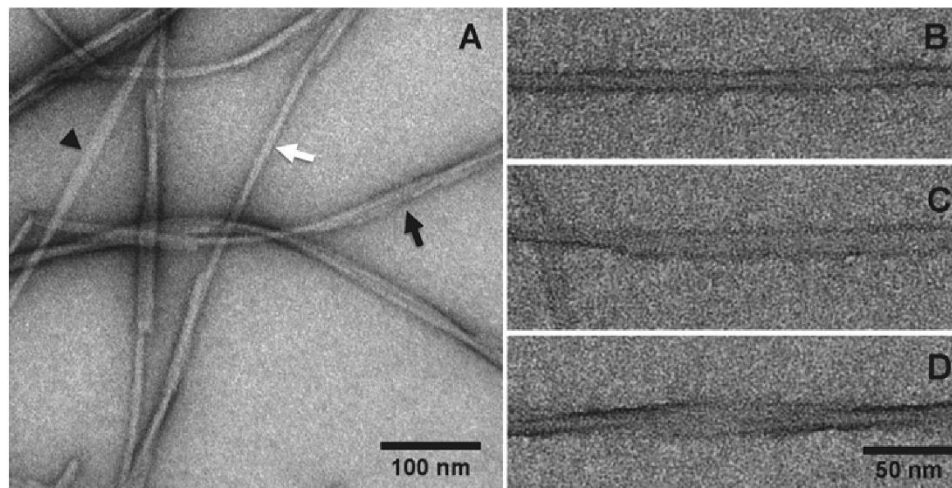
### HfqCTR<sub>p</sub> folding and self-assembly

In order to probe for the presence of secondary structure within the synthesized CTR Hfq peptide, we subjected the peptide to CD and FTIR analyses. We recorded the spectra few hours after dissolving lyophilized HfqCTR<sub>p</sub> in phosphate buffer. None of the spectra indicated the presence of secondary structure (result not shown). This result was in agreement with the high frequency of amino acids-promoting disorder (such as A, N, Q, E, G, K, H, D and S) and with an intrinsically-disordered protein (IDP) [42,43].

After few days (the time being dependent on the temperature, see ‘Experimental’), the HfqCTR<sub>p</sub> peptide solution undergoes changes. The solution became very viscous and presented small particles. To understand the nature of these changes, we subjected the peptide solution to negative staining TEM analysis. This revealed the presence of long fibrillar structures (Figure 1), indicating a tendency of C-terminal Hfq peptide to self-assemble into long-ordered structures. Similar preparation with the HfqNTR<sub>p</sub> did not form any long-ordered structures (Supplementary Figure S1). HfqCTR<sub>p</sub> fibrils revealed by negative staining EM seem to be monodisperse in width but with varying lengths (up to several micrometres). The width of the fibrillar structures as measured perpendicularly to the longitudinal axis is approximately 10 nm. TEM also evidences that these fibrils are twisted and sometimes entangled to form larger bundles of fibrils (Figure 1).

Proteins could show a high tendency to form fibrillar aggregates under certain thermodynamic conditions. The aggregates referred to as amyloid fibrils, reported for a substantial number of proteins, are elongated structures whose spine is formed by highly ordered  $\beta$ -sheets running parallel to the fibril axis [44]. Amyloid aggregation is considered as a general characteristic of proteins, being the amyloid fibril the most stable state for the polypeptide chain [45]. In particular, there are several examples of proteins forming amyloid fibrils whose native state lacks of a well-defined secondary structure or belongs rather to the IDP family [46]. The observed self-assembly into fibrillar extended structures of HfqCTR<sub>p</sub> was thus quite interesting given the fact that secondary structures was absent from freshly dissolved peptide. We probed thus for the presence of amyloid-like secondary structures in Hfq peptide after self-assembly by using FTIR spectroscopy. The sensitivity of amide vibrations to secondary structures makes FTIR an excellent probe for protein ordering, especially for  $\beta$ -sheet like structures [47]. As shown in Figure 2, FTIR absorption spectrum of HfqCTR<sub>p</sub> fibrils in the region of amide I (1720–1560  $\text{cm}^{-1}$ ) indicates the band is peaked approximately 1648  $\text{cm}^{-1}$  with a remarkable shoulder at approximately 1615  $\text{cm}^{-1}$ , suggesting that peptides in the aggregates are arranged into at least two secondary structural moieties [48,49]. The second derivative of the curve, together with FSD-enhancement technique allowed identifying the different spectral components of amide I band (Figure 2A), which were used for initializing the curve fitting procedure (Figure 2B). Five main moieties was





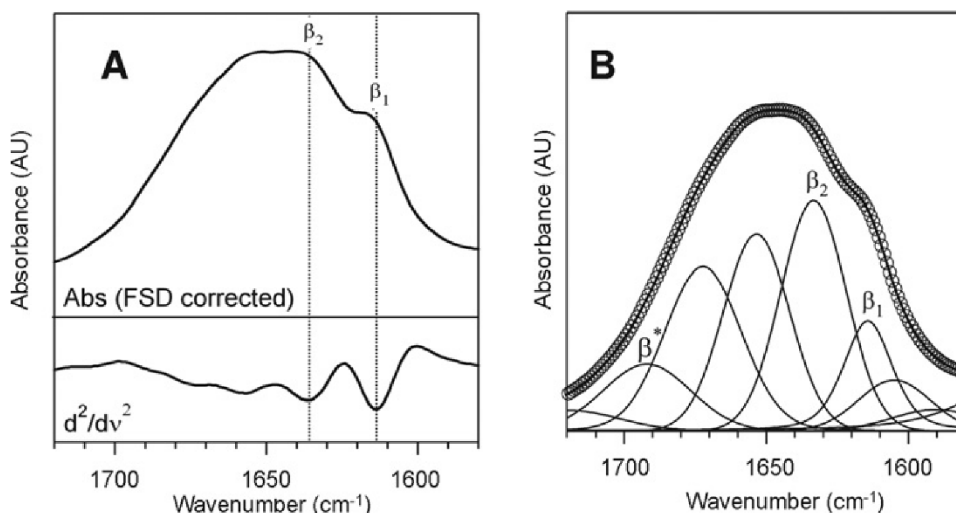
**Figure 1** TEM negative staining micrographs showing Hfq C-terminal fibres (A) General microscopy field of Hfq C-terminal sample, illustrating the heterogeneity in fibres organization: closed (white arrow), opened (arrowhead) and twisted (black arrow) fibres. (B–D) Close-up views of closed, opened and twisted fibres respectively. Scale bars are 100 nm (A) and 50 nm (B–D).

detected, absorbing at approximately  $1616$  and  $1634\text{ cm}^{-1}$ , both of them corresponding to vibrational motions of the backbone amide moieties in  $\beta$ -sheet conformations,  $1655\text{ cm}^{-1}$ , ascribed to random coil peptidic segments,  $1673\text{ cm}^{-1}$ , typical amide I frequency of turns and  $1690\text{ cm}^{-1}$  which corresponds to the amide I high frequency component of  $\beta$ -sheet vibrations, due to excitonic splitting, in case of structures arranged in an antiparallel fashion [50,51]. The results obtained from fitting procedure revealed that approximately 42% of the total amide I area (18% ascribed to the peak at approximately  $1616\text{ cm}^{-1}$  and 24% due to the one approximately  $1634\text{ cm}^{-1}$ ) is arranged in  $\beta$ -sheet conformation whereas only 21% of the peptide shows a random coil arrangement. The observation of a high content of  $\beta$ -sheets reflects the aggregated nature of the sample. In particular, the low frequency  $\beta$ -sheet moiety, at approximately  $1616\text{ cm}^{-1}$ , can be associated to multi-stranded intermolecular  $\beta$ -sheet and resembles the typical fingerprint of the cross- $\beta$  motif found in amyloid-like fibrils [52,53].

The evidence of a substantial  $\beta$ -sheet structure together with the fibrillar morphology of HfqCTR<sub>p</sub> aggregates suggests an assembly via a similar mechanism that lead to formation of amyloids. To test whether the observed fibrillar structures are similar to amyloids, we used ThT stain that is commonly used to detect amyloid structures (see ‘Experimental’ for details). Stained fibrils were imaged through confocal fluorescence microscopy [54]. As seen on Figure 3, the enhanced fluorescence of ThT confirms the presence of intermolecular  $\beta$ -sheet distributed along the fibrils. They furthermore highlight two distinct morphologies of the aggregates present in the sample: laterally packed fibrils bundles (Figures 3A and 3B) as well as clustered wires with diameters smaller than  $1\text{ }\mu\text{m}$  (Figure 3C).

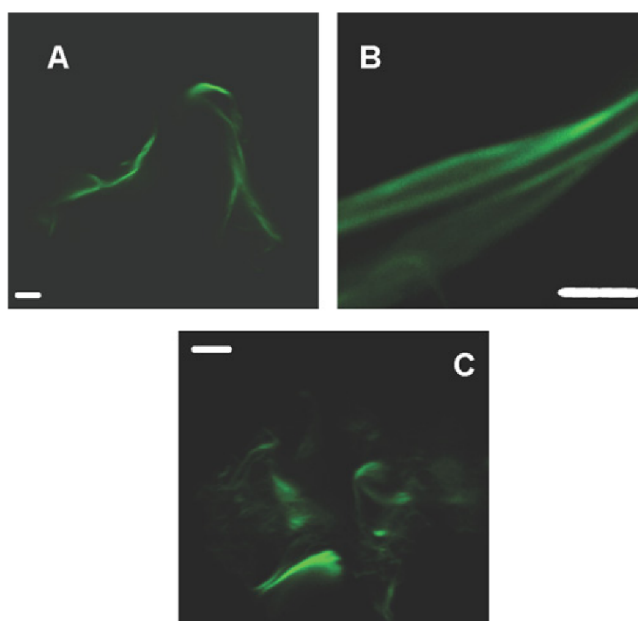
### Hfq-CTR is required for assembly of the Hfq high-order cellular structures

We previously reported and illustrated in the present study IF microscopy using WT cells (Figure 4C). Cells stained with a purified anti-Hfq antibody showed that Hfq, unlike cytoplasmic proteins, is not diffusely distributed within the cell, but is localized as repetitive extended elements oriented roughly at  $45^\circ$  angles along the axis of the cell (Figure 4C) [2]. The structures appear to coil around the cell periphery suggesting a membrane association of the Hfq cellular fibril-like structures. To determine the role of the self-assembly property of the C-terminal domain in the formation of such higher order cellular structures observed with full-length Hfq in intact WT cells, we carried IF localization studies in cells that lacked Hfq CTR. The cells expressed under the control of arabinose promoter a plasmid encoded full-length Hfq or Hfq protein without the C-terminal domain (HFq-NTR, amino acid residues 1–72) or the C-terminal domain of the protein alone (HFq-CTR, amino acid residues 65–102). Because the limit between NTR and CTR is not formally known, seven overlapping residues (residues 65–72) were included in both NTR and CTR constructs. This showed that the HFq-NTR domain, which lacks the C-terminal extension of the full-length Hfq, was uniformly distributed throughout the cytoplasm without any detectable organization (Figure 4A). In contrast, full-length Hfq expressed from the plasmid showed similar organized cellular structures observed in WT cells expressing chromosomally-encoded Hfq (Figures 4C and 4E). These localization patterns were observed over a wide range of inducer concentrations (including conditions that give normal levels of expression of Hfq full-length [37]), indicating that the observed diffuse localization is not due to under or overexpression of the protein. Taken together, the



**Figure 2 Secondary structure of fibrillar Hfq CTR from FTIR absorption spectroscopy**

(A) Amide I FSD corrected spectrum (top) and second-derivative spectrum (down) of C-terminal Hfq fibrils. Two  $\beta$ -sheet moieties,  $\beta_1$  and  $\beta_2$ , are clearly detected. (B) Experimental (circles) and fitted (lines) data of amide I band of C-terminal Hfq fibrils. Two distinct  $\beta$ -sheet structures are present in the aggregates,  $\beta_1$  and  $\beta_2$  and the presence of the high frequency component  $\beta^*$  suggests the possible presence of antiparallel moieties [51].



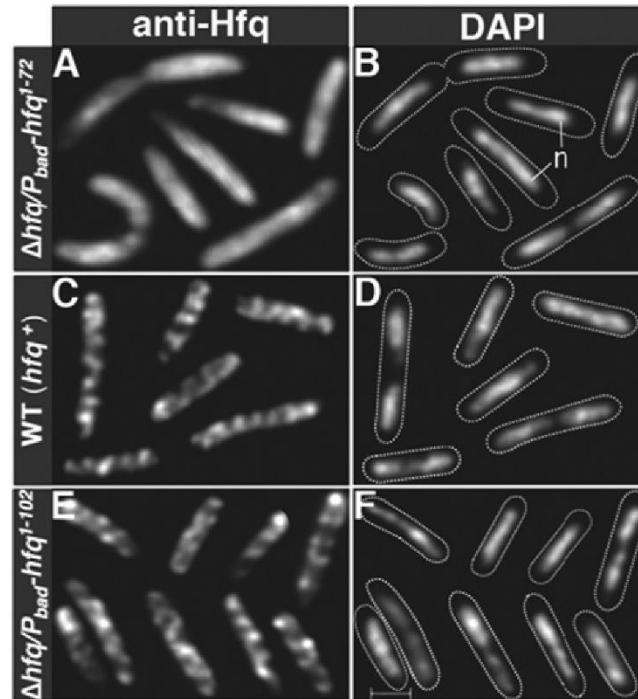
**Figure 3 Confocal microscopy fluorescence images for CTR Hfq fibrils**

Samples (Hfq-CTR peptide) were stained by ThT as indicated under 'Experimental' section. (Scale bar = 5  $\mu$ m). Distinct morphologies are present in the sample: laterally packed fibril bundles (A and B) or clustered wires (C). Scale bar = 5  $\mu$ m.

intrinsic property of the CTR to self-assemble with the fact that loss of CTR domain of Hfq was associated with loss of assembly of higher order cellular structures observed with full-length proteins, strongly suggest a direct role of the CTR domain of Hfq in the macromolecular assembly of Hfq cellular structure in the vicinity of the membrane of *E. coli* bacterium.

Attempts to test the ability of the CTR domain to assemble within long-range cellular structures in the absence of the rest of the protein were not successful due to inability to stably express the CTR domain. The CTR domain expressed alone, apart from NTR, is presumably very unstable and degraded quickly *in vivo* as no corresponding band was detected in





**Figure 4 Assembly of Hfq higher-order cellular structures requires the C-terminal domain of Hfq**

(A, B, E and F) *hfq* Null cells expressed Hfq under the control of  $P_{ara}$  promoter plasmid-encoded truncated Hfq<sup>(1-72)</sup> (A and B) or plasmid-encoded full-length Hfq (E and F). (C and D) WT cells expressed chromosomally encoded WT Hfq (C and D). Cells were doubly stained with DAPI, for nucleoid staining and by IF using a purified anti-Hfq polyclonal antibody. Cells shown in the composite micrographs represent cells selected from different areas of one field and are representative of most cells in the indicated experiments. In DAPI micrographs dotted lines show boundaries of cells and n indicates nucleoid. Scale bar = 1  $\mu$ m.

Western blot analyses of extract from cells grown in the presence of a range of arabinose concentration (result not shown). This is consistent with protein instability prediction analysis for this region [55].

## DISCUSSION

Collectively, our results evidence that Hfq CTR has a strong propensity to self-associate and to form amyloid-like fibrillar structures. The assembly of the Hfq-CTR fibrillar structures was shown by negative staining EM, thioflavin staining fluorescence microscopy and FTIR spectroscopy. The self-assembly of the full-length Hfq protein in amyloid-type fibres has been previously reported [21]. Nevertheless, the region of the protein required for the self-assembly was not clearly identified. In the present study, we show that Hfq-CTR has an intrinsic property to self-assemble independently of the rest of the protein, whereas the Hfq-NTR fails to assemble into extended structures. The present work reveals that Hfq-CTR is required for the self-assembly of Hfq into long-range order cellular structures. Furthermore, as fibrillar Hfq

polymers retains their RNA-binding properties [21], this work shed light on how Hfq CTR region could have a physiological function in directing Hfq cellular localization in order to govern its activity. Note that previous reports showed that the NTR (core Sm-like domain) alone retains most Hfq RNA-binding properties [31,32], indicating that the loss of cellular localization of Hfq that lacks the CTR is not due to failure in RNA binding.

Previous studies indicated that this Hfq CTR fragment is intrinsically disordered [30] and that IDPs often have a tendency to form amyloid-like structures [56]. Formation of amyloid-type fibrillar structure *in vitro* shows that Hfq-CTR domain belongs to this family of amyloid IDPs prone to aggregation. We show in the present study that self-assembly of the CTR region is associated with formation of different secondary structures within the CTR fragment as indicated by the changes in FTIR spectra observed after assembly of the fibrillar structures. The mechanism underlying folding and assembly of such a natively unstructured domain of Hfq remains to be elucidated. Whether interactions with other cellular components such as RNA molecules could play a role in accelerating the observed slow folding/self-assembly process is not known. Flexible IDP regions have, for instance, been shown to favour interactions with large cellular partners, like nucleic

acids and have been found in some RNA chaperones [57]. This could apply to Hfq whose CTR could play an additional role in RNA recognition [28].

Amyloid fibrils are usually the hallmark of neurodegenerative diseases, but they have also been reported in bacteria, in particular for proteins that form curli at the bacterial cell surface [58,59]. Commonly, intracellular amyloid-type protein aggregation is considered potentially toxic for bacteria [60]. Nevertheless, it has been shown that some proteins can gain new functions when present as amyloid fibrils. For example, the hydrogenase maturation factor HypF N-terminal domain in its aggregated form binds cell membrane and DNA whereas its monomeric form does not [61]. In the present study, we present a new example where a bacterial amyloid-like self-assembly plays an important role in directing assembly of Hfq into supramolecular cellular structures that includes other components of the RNA processing and degradation pathway. How self-assembly of the CTR tail of Hfq might lead to the formation of the organized cellular structures seen in intact cells remains to be elucidated. Understanding of temporal and spatial self-assembly process of Hfq will thus be paramount to understand how post-transcriptional regulation occurs as a function of the cell cycle.

#### AUTHOR CONTRIBUTION

Emilie Fortas and Antoine Malabirade performed experiments. Federica Piccirilli, Sylvain Trépoint and Aziz Taghbalout designed experiments, performed experiments, analysed the data and wrote the manuscript. Valeria Militello, Sergio Marco and Véronique Arluison designed experiments, analysed the data and wrote the manuscript.

#### ACKNOWLEDGEMENTS

We would like to thank B. Cayrol (LLB, CEA Saclay) for his help in cloning Hfq-NTD and PICT-Ibisa for providing access to JEOL 2200FS TEM. We are also indebted to J. Seguin (iBiTec-S, CEA Saclay) for his help in DC analysis, to A. Gomez Zavaglia (UNLP, Argentina) for fruitful discussions in the frame of Ecos-Sud A09B02 program.

#### FUNDING

This work was supported by FP7 DIVINOCELL [grant number Health-F3-2009-223431]; the University Paris Diderot; the National Center for Scientific Research (CNRS); the French Alternative Energies and Atomic Energy Commission (CEA); and the National Institute of Health [grant number R37 GM060632 (to A.T.)].

## REFERENCES

- de Boer, P.A. (2010) Advances in understanding *E. coli* cell fission. *Curr. Opin. Microbiol.* **13**, 730–737 [CrossRef PubMed](#)
- Taghbalout, A., Yang, Q. and Arluison, V. (2014) The *Escherichia coli* RNA processing and degradation machinery is compartmentalized within an organized cellular network. *Biochem. J.* **458**, 11–22 [CrossRef PubMed](#)
- Hoch, P.G. and Hartmann, R.K. (2014) Supramolecular membrane-associated assemblies of RNA metabolic proteins in *Escherichia coli*. *Biochem. J.* **458**, e1–e3 [CrossRef PubMed](#)
- Lu, F. and Taghbalout, A. (2013) Membrane association via an amino-terminal amphipathic helix is required for the cellular organization and function of RNase II. *J. Biol. Chem.* **288**, 7241–7251 [CrossRef PubMed](#)
- Taghbalout, A., Ma, L. and Rothfield, L. (2006) Role of MinD-membrane association in Min protein interactions. *J. Bacteriol.* **188**, 2993–3001 [CrossRef PubMed](#)
- Vogel, J. and Luisi, B.F. (2011) Hfq and its constellation of RNA. *Nat. Rev. Microbiol.* **9**, 578–589 [CrossRef PubMed](#)
- Geinguenaud, F., Calandrini, V., Teixeira, J., Mayer, C., Liquier, J., Lavelle, C. and Arluison, V. (2011) Conformational transition of DNA bound to Hfq probed by infrared spectroscopy. *Phys. Chem. Chem. Phys.* **13**, 1222–1229 [CrossRef PubMed](#)
- Gottesman, S. and Storz, G. (2011) Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring. Harb. Perspect. Biol.* **3**, a003798 [CrossRef PubMed](#)
- Storz, G., Vogel, J. and Wassarman, K.M. (2011) Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell* **43**, 880–891 [CrossRef PubMed](#)
- Aiba, H. (2007) Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr. Opin. Microbiol.* **10**, 134–139 [CrossRef PubMed](#)
- Zhang, A., Wassarman, K.M., Ortega, J., Steven, A.C. and Storz, G. (2002) The Sm-like Hfq protein increases oxyS RNA interaction with target mRNAs. *Mol. Cell* **9**, 11–22 [CrossRef PubMed](#)
- Moller, T., Franch, T., Hojrup, P., Keene, D.R., Bachinger, H.P., Brennan, R.G. and Valentin-Hansen, P. (2002) Hfq. A bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol. Cell.* **9**, 23–30 [CrossRef PubMed](#)
- Rice, J.B. and Vanderpool, C.K. (2011) The small RNA SgrS controls sugar-phosphate accumulation by regulating multiple PTS genes. *Nucleic Acids Res.* **39**, 3806–3819 [CrossRef PubMed](#)
- Guillier, M. and Gottesman, S. (2006) Remodelling of the *Escherichia coli* outer membrane by two small regulatory RNAs. *Mol. Microbiol.* **59**, 231–247 [CrossRef PubMed](#)
- Bardill, J.P., Zhao, X. and Hammer, B.K. (2011) The *Vibrio cholerae* quorum sensing response is mediated by Hfq-dependent sRNA/mRNA base pairing interactions. *Mol. Microbiol.* **80**, 1381–1394 [CrossRef PubMed](#)
- Guillier, M., Gottesman, S. and Storz, G. (2006) Modulating the outer membrane with small RNAs. *Genes Dev.* **20**, 2338–2348 [CrossRef PubMed](#)
- Vogt, S.L. and Raivio, T.L. (2014) Hfq reduces envelope stress by controlling expression of envelope-localized proteins and protein complexes in enteropathogenic *Escherichia coli*. *Mol. Microbiol.* **92**, 681–697 [CrossRef PubMed](#)
- Mura, C., Randolph, P.S., Patterson, J. and Cozen, A.E. (2013) Archaeal and eukaryotic homologs of Hfq: a structural and evolutionary perspective on Sm function. *RNA Biol.* **10**, 636–651 [CrossRef PubMed](#)
- Brennan, R.G. and Link, T.M. (2007) Hfq structure, function and ligand binding. *Curr. Opin. Microbiol.* **10**, 125–133 [CrossRef PubMed](#)
- Mura, C., Phillips, M., Kozhukhovskiy, A. and Eisenberg, D. (2003) Structure and assembly of an augmented Sm-like archaeal protein 14-mer. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4539–4544 [CrossRef PubMed](#)
- Arluison, V., Mura, C., Guzman, M.R., Liquier, J., Pellegrini, O., Gingery, M., Regnier, P. and Marco, S. (2006) Three-dimensional structures of fibrillar Sm proteins: Hfq and other Sm-like proteins. *J. Mol. Biol.* **356**, 86–96 [CrossRef PubMed](#)





- 22 Schumacher, M.A., Pearson, R.F., Moller, T., Valentin-Hansen, P and Brennan, R.G. (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J.* **21**, 3546–3556 [CrossRef PubMed](#)
- 23 Sauter, C., Basquin, J. and Suck, D. (2003) Sm-like proteins in eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*. *Nucleic Acids Res.* **31**, 4091–4098 [CrossRef PubMed](#)
- 24 Nikulin, A., Stolboushina, E., Perederina, A., Vassilieva, I., Blaesi, U., Moll, I., Kachalova, G., Yokoyama, S., Vassilyev, D., Garber, M. and Nikonov, S. (2005) Structure of *Pseudomonas aeruginosa* Hfq protein. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 141–146 [CrossRef PubMed](#)
- 25 Sauer, E. and Weichenrieder, O. (2011) Structural basis for RNA 3'-end recognition by Hfq. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 13065–13070 [CrossRef PubMed](#)
- 26 Link, T.M., Valentin-Hansen, P and Brennan, R.G. (2009) Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19292–19297 [CrossRef PubMed](#)
- 27 Sauer, E. (2013) Structure and RNA-binding properties of the bacterial LSm protein Hfq. *RNA Biol.* **10**, 610–618 [CrossRef PubMed](#)
- 28 Dimastrogiovanni, D., Frohlich, K.S., Bandyra, K.J., Bruce, H.A., Hohensee, S., Vogel, J. and Luisi, B.F. (2014) Recognition of the small regulatory RNA RydC by the bacterial Hfq protein. *eLife* **3**, doi: 10.7554/eLife.05375 [CrossRef](#)
- 29 Vincent, H.A., Henderson, C.A., Stone, C.M., Cary, P.D., Gowers, D.M., Sobott, F., Taylor, J.E. and Callaghan, A.J. (2012) The low-resolution solution structure of *Vibrio cholerae* Hfq in complex with Qrr1 sRNA. *Nucleic Acids Res.* **40**, 8698–8710 [CrossRef PubMed](#)
- 30 Beich-Frandsen, M., Vecerek, B., Konarev, P.V., Sjoblom, B., Kloiber, K., Hammerle, H., Rajkowitsch, L., Miles, A.J., Kontaxis, G., Wallace, B.A. et al. (2011) Structural insights into the dynamics and function of the C-terminus of the *E. coli* RNA chaperone Hfq. *Nucleic Acids Res.* **39**, 4900–4915 [CrossRef PubMed](#)
- 31 Arluison, V., Folichon, M., Marco, S., Derreumaux, P., Pellegrini, O., Seguin, J., Hajnsdorf, E. and Regnier, P. (2004) The C-terminal domain of *Escherichia coli* Hfq increases the stability of the hexamer. *Eur. J. Biochem.* **271**, 1258–1265 [CrossRef PubMed](#)
- 32 Olsen, A.S., Moller-Jensen, J., Brennan, R.G. and Valentin-Hansen, P. (2010) C-Terminally truncated derivatives of *Escherichia coli* Hfq are proficient in riboregulation. *J. Mol. Biol.* **404**, 173–182 [CrossRef PubMed](#)
- 33 Vecerek, B., Rajkowitsch, L., Sonnleitner, E., Schroeder, R. and Blasi, U. (2008) The C-terminal domain of *Escherichia coli* Hfq is required for regulation. *Nucleic Acids Res.* **36**, 133–143 [CrossRef PubMed](#)
- 34 Salim, N.N., Faner, M.A., Philip, J.A. and Feig, A.L. (2012) Requirement of upstream Hfq-binding (ARN)<sub>x</sub> elements in glmS and the Hfq C-terminal region for GlmS upregulation by sRNAs GlmZ and GlmY. *Nucleic Acids Res.* **40**, 8021–8032 [CrossRef PubMed](#)
- 35 Robinson, K.E., Orans, J., Kovach, A.R., Link, T.M. and Brennan, R.G. (2014) Mapping Hfq-RNA interaction surfaces using tryptophan fluorescence quenching. *Nucleic Acids Res.* **42**, 2736–2749 [CrossRef PubMed](#)
- 36 Sledjeski, D.D., Whitman, C. and Zhang, A. (2001) Hfq is necessary for regulation by the untranslated RNA DsrA. *J. Bacteriol.* **183**, 1997–2005 [CrossRef PubMed](#)
- 37 Diestra, E., Cayrol, B., Arluison, V. and Risco, C. (2009) Cellular electron microscopy imaging reveals the localization of the Hfq protein close to the bacterial membrane. *PLoS One* **4**, e8301 [CrossRef PubMed](#)
- 38 Byler, D.M. and Susi, H. (1986) Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers* **25**, 469–487 [CrossRef PubMed](#)
- 39 Reinke, A.A. and Gestwicki, J.E. (2011) Insight into amyloid structure using chemical probes. *Chem. Biol. Drug Des.* **77**, 399–411 [CrossRef PubMed](#)
- 40 Arluison, V. and Taghbalout, A. (2015) Cellular localization of RNA degradation and processing components in *Escherichia coli*. *Methods Mol. Biol.* **1259**, 87–101 [CrossRef PubMed](#)
- 41 de Boer, P.A., Crossley, R.E. and Rothfield, L.I. (1989) A division inhibitor and a topological specificity factor coded for by the minicell locus determine proper placement of the division septum in *E. coli*. *Cell* **56**, 641–649 [CrossRef PubMed](#)
- 42 Campen, A., Williams, R.M., Brown, C.J., Meng, J., Uversky, V.N. and Dunker, A.K. (2008) TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept. Lett.* **15**, 956–963 [CrossRef PubMed](#)
- 43 Oldfield, C.J. and Dunker, A.K. (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **83**, 553–584 [CrossRef PubMed](#)
- 44 Knowles, T.P., Vendruscolo, M. and Dobson, C.M. (2014) The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15**, 384–396 [CrossRef PubMed](#)
- 45 Baldwin, A.J., Knowles, T.P., Tartaglia, G.G., Fitzpatrick, A.W., Devlin, G.L., Shammass, S.L., Waudby, C.A., Mossuto, M.F., Meehan, S., Gras, S.L. et al. (2011) Metastability of native proteins and the phenomenon of amyloid formation. *J. Am. Chem. Soc.* **133**, 14160–14163 [CrossRef PubMed](#)
- 46 Uversky, V.N. (2002) What does it mean to be natively unfolded? *Eur. J. Biochem.* **269**, 2–12 [CrossRef PubMed](#)
- 47 Piccirilli, F., Mangialardo, S., Postorino, P., Baldassarre, L., Lupi, S. and Perucchi, A. (2012) Sequential dissociation of insulin amyloids probed by high pressure Fourier transform infrared spectroscopy. *Journal of neurological surgery. Soft Matter.* **8**, 11863–11870
- 48 Vetri, V., D'Amico, M., Fodera, V., Leone, M., Ponzoni, A., Sberveglieri, G. and Militello, V. (2011) Bovine serum albumin protofibril-like aggregates formation: solo but not simple mechanism. *Arch. Biochem. Biophys.* **508**, 13–24 [CrossRef PubMed](#)
- 49 Vetri, V. and Militello, V. (2005) Thermal induced conformational changes involved in the aggregation pathways of beta-lactoglobulin. *Biophys. Chem.* **113**, 83–91 [CrossRef PubMed](#)
- 50 Chirgadze, Y.N. and Nevskaya, N.A. (1976) Infrared spectra and resonance interaction of amide-I vibration of the parallel-chain pleated sheets. *Biopolymers* **15**, 627–636 [CrossRef PubMed](#)
- 51 Shivu, B., Seshadri, S., Li, J., Oberg, K.A., Uversky, V.N. and Fink, A.L. (2013) Distinct beta-sheet structure in protein aggregates determined by ATR-FTIR spectroscopy. *Biochemistry* **52**, 5176–5183 [CrossRef PubMed](#)
- 52 Szabo, Z., Klement, E., Jost, K., Zarandi, M., Soos, K. and Penke, B. (1999) An FT-IR study of the beta-amyloid conformation: standardization of aggregation grade. *Biochem. Biophys. Res. Commun.* **265**, 297–300 [CrossRef PubMed](#)
- 53 Sunde, M., Serpell, L.C., Bartlam, M., Fraser, P.E., Pepys, M.B. and Blake, C.C. (1997) Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J. Mol. Biol.* **273**, 729–739 [CrossRef PubMed](#)
- 54 D'Amico, M., Di Carlo, M.G., Groenning, M., Militello, V., Vetri, V. and Leone, M. (2012) Thioflavin T promotes Aβ(1–40) amyloid fibrils formation. *Biophys. Chem.* **3**, 1596–1601
- 55 Wilkins, M.R., Gasteiger, E., Bairoch, A., Sanchez, J.C., Williams, K.L., Appel, R.D. and Hochstrasser, D.F. (1999) Protein identification and analysis tools in the EXPASY server. *Methods Mol. Biol.* **112**, 531–552 [PubMed](#)
- 56 Zhang, T., Faraggi, E., Li, Z. and Zhou, Y. (2013) Intrinsically semi-disordered state and its role in induced folding and protein aggregation. *Cell Biochem. Biophys.* **67**, 1193–1205 [CrossRef PubMed](#)
- 57 Ivanyi-Nagy, R., Davidovic, L., Khandjian, E.W. and Darlix, J.L. (2005) Disordered RNA chaperone proteins: from functions to disease. *Cell Mol. Life Sci.* **62**, 1409–1417 [CrossRef PubMed](#)



- 58 Zhou, Y., Blanco, L.P., Smith, D.R. and Chapman, M.R. (2012) Bacterial amyloids. *Methods Mol. Biol.* **849**, 303–320 [CrossRef PubMed](#)
- 59 Zhou, Y., Smith, D., Leong, B.J., Brannstrom, K., Almqvist, F. and Chapman, M.R. (2012) Promiscuous cross-seeding between bacterial amyloids promotes interspecies biofilms. *J. Biol. Chem.* **287**, 35092–35103 [CrossRef PubMed](#)
- 60 Upadhyay, A.K., Murmu, A., Singh, A. and Panda, A.K. (2012) Kinetics of inclusion body formation and its correlation with the characteristics of protein aggregates in *Escherichia coli*. *PLoS One* **7**, e33951 [CrossRef PubMed](#)
- 61 Wang, L., Schubert, D., Sawaya, M.R., Eisenberg, D. and Riek, R. (2010) Multidimensional structure-activity relationship of a protein in its aggregated states. *Angew Chem. Int. Ed. Engl.* **49**, 3904–3908 [CrossRef PubMed](#)

---

Received 28 August 2014/2 February 2015; accepted 2 February 2015

Published as Immediate Publication 16 March 2015, doi 10.1042/BSR20140128

---

Annexe C :

*Techniques to analyse sRNA  
protein cofactor self-assembly  
in vitro*

## Techniques to analyse sRNA protein cofactor self-assembly *in vitro*

David Partouche<sup>1,2</sup>, Antoine Malabirade<sup>2</sup>, Thomas Bizien<sup>1</sup>, Marisela Velez<sup>3</sup>, Sylvain Trepout<sup>4,5</sup>, Sergio Marco<sup>4,5</sup>, Valeria Militello<sup>6</sup>, Christophe Sandt<sup>1</sup>, Frank Wien<sup>1</sup> and Véronique Arluison<sup>2,7</sup>

<sup>1</sup> Synchrotron SOLEIL, L'Orme des Merisiers Saint Aubin, 91410 Gif-sur-Yvette, France

<sup>2</sup> Laboratoire Léon Brillouin LLB, CEA, CNRS UMR12, Université Paris Saclay, CEA Saclay, 91191 Gif-sur-Yvette, France.

<sup>3</sup> Instituto de Catálisis y Petroleoquímica, CSIC, c/ Marie Curie, 2, Cantoblanco, E-28049 Madrid, Spain.

<sup>4</sup> INSERM, U1196, Université Paris Sud, Université Paris-Saclay, F-91405 Orsay, France

<sup>5</sup> Institut Curie, PSL Research University, CNRS, UMR 91873348, F-91405 Orsay, France

<sup>6</sup> Department of Physics and Chemistry, University of Palermo, Viale delle Scienze, Ed. 18, 90128 Palermo, Italy

<sup>7</sup> Université Paris Diderot-Paris 7, Sorbonne Paris Cité, 75013 Paris, FRANCE

To whom correspondence should be addressed: Veronique.Arluison@univ-paris-diderot.fr

## ABSTRACT

Post-transcriptional control of gene expression by small regulatory noncoding RNA (sRNA) needs protein accomplices to occur. Past research mainly focused on the RNA chaperone Hfq as cofactor. Nevertheless, recent studies indicated that other proteins might be involved in sRNA-based regulations. As some of these proteins have been shown to self-assemble, we describe in this chapter protocols to analyse the nano-assemblies formed. Precisely, we focus our analysis on *Escherichia coli* Hfq as a model, but the protocols presented here can be applied to analyse any polymer of proteins. This chapter thus provides a guideline to develop commonly used approaches to detect prokaryotic protein self-assembly, with a special focus on the detection of amyloidogenic polymers.

**Keywords:** protein self-assembly; noncoding RNA cofactor; functional amyloid

**Abbreviations:** AFM: Atomic Force Microscopy; CF: Curve Fitting; FTIR: Fourier Transform InfraRed spectroscopy; SAXS: Small Angle X-ray Scattering; SRCD: Synchrotron Radiation Circular Dichroism; TEM: Transmission Electron Microscopy; ThT: Thioflavin T

**Running head:** sRNA cofactor self-assembly

## 1. Introduction

In prokaryotes, the components of RNA metabolism self-assemble into complex structures, resulting in functional compartmentalisation within the cell [1-3]. The RNA maturation components include proteins involved in post-transcriptional genetic regulation such as ribonucleases (RNAses) or RNA chaperone, but also small regulatory noncoding RNAs (sRNA) [2,4,5]. The best-characterised protein involved in such pathways using regulatory sRNA is probably the RNA chaperone Hfq [6]. Nevertheless, some others have also been identified such as ProQ and the archaeal Sm proteins [7,8]. Our recent studies demonstrated that specific sequences in the *E. coli* Hfq C-terminus region (CTR) are able to polymerise, resulting in the formation of amyloid fibers [9]. Indeed, increasing evidence shows that amyloids are not only the result of protein misfolding associated with neurodegenerative diseases, but are also found in cells for useful reasons [10]. In the latter case, amyloids contribute to the physiology of the cell and are referred as “*functional*” [10]. In fact, these functional amyloids have been widely reported in different organisms, including bacteria. In prokaryotes, diverse functions have been described for these amyloids, such as a role in biofilm development [11], formation of curli and pili [12], or creation of pores in the host membrane [13]. In this chapter, we intend to provide a guideline to develop approaches that can be used to detect the fibers of sRNA protein cofactors, precisely molecular imaging, such as atomic force microscopy and electron microscopy, infrared spectroscopy, synchrotron radiation circular dichroism and small angle X-ray scattering. Our chapter will focus on Hfq and its amyloid nature as an example, but it could be applied to any self-assembling prokaryotic proteins.

## 2. Materials

### 2.1 Preparation of samples

1. Synthetic peptide: 20 mg/mL in deionized RNA-grade water (see **note 1**)
2. Purified Hfq/Sm proteins, prepared as described in the chapter 15 by Stanek and Mura. Depending on buffer used and protein concentration

reached during the purification process, the same limits apply as those described in **note 1**.

3. UV spectrophotometer suitable for the precise determination of protein concentration measurements (micro-volumes).

### ***2.2 Thioflavin staining***

1. ThT stock solution: 0.8 mg/mL in 10mM phosphate buffer pH7 containing 150mM NaCl.

2. Fluorescence microscope equipped with a digital camera and emission/excitation filter sets to permit visualization of ThT staining.

### ***2.2 AFM imaging***

1. AFM microscope with liquid cell

2. AFM tips with different force constants: 2-75 N/m and resonant frequencies in the range of 100-320 kHz for imaging in air or in the range of 0.01-0.8 N/m and resonant frequencies in the range of 100-20 kHz for imaging in solution.

4. AFM-grade mica.

5. AFM free image analysis softwaresuch as WsXM (<http://www.wsxmsolutions.com/>) and Gwyddion (<http://gwyddion.net/>).

### ***2.3 TEM imaging***

1. Cryo-transmission electron microscope equipped with a slow scan CCD camera.

2. Carbon-coated electron microscopy copper grid, CF200-Cu (200 mesh) or CF300-Cu (300 mesh), EMS.

3. Carbon-coated electron microscopy copper grid, ref 01881-F (200 mesh) or 01883-F (300 mesh), Ted Pella.

4. Uranyl acetate solution: 2% (w/v).

5. Uranyl-less solution containing Gadolinium salts, Delta Microscopies.

6. Plasma cleaner.
7. Leica EM-CPC plunge-freezing device for cryo-fixation.

#### ***2.4 FTIR spectroscopy***

1. Interferometer-based spectrophotometer coupled with an IR source.
2. For transmission measurements, liquid cell equipped with CaF<sub>2</sub> windows and a 6 µm-thick polytetrafluoroethylene (Teflon) spacer (for example Omni-Cell transmission Cell from Specac Company or Demountable liquid Cell from Piketech).
3. For reflection measurements, attenuated total reflection (ATR) sampling set-up.

#### ***2.5 SRCD spectroscopy***

1. SRCD endstation such as that of DISCO beamline at SOLEIL synchrotron [14,15].
2. Manually loaded circular demountable CaF<sub>2</sub> cells of 30 µm path length with a loading volume of 2µl [16].
3. CDtool software for data acquisition and treatment, including averaging, smoothing, subtraction of buffer baselines from sample-spectra, calibration ((+)-camphor-10-sulfonic acid CSA) and normalization [17].
4. BestSel open access software (<http://bestsel.elte.hu/>) [18].

#### ***2.6 Small Angle X-ray Scattering (SAXS)***

1. SAXS beamline such as SWING at SOLEIL.
2. Quartz capillaries of 1.5 mm diameter and 0.01 mm thickness
3. Foxtrot software  
(<http://www.synchrotronsoleil.fr/Recherche/LignesLumiere/SWING>)

### **3. Techniques**

#### ***3.1 Analysis of Hfq self-assembly by molecular imaging***

Characterization of Hfq self-assembly can be performed by microscopy techniques, including fluorescence microscopy, transmission electron microscopy (TEM) and atomic force microscopy (AFM).

### *3.1.1 Thioflavin staining*

Thioflavin T (ThT) is a benzothiazole salt commonly used to visualize amyloids. When bound to aggregated  $\beta$ -sheets, such as those found in amyloids, the dye displays an enhanced red-shifted fluorescence emission spectrum [19]. Even if the dye fluorescence is not completely specific and may bind to other structures (such as double-stranded DNA), it is however a good indicator that amyloid structures are formed in the sample. Practically, for ThT staining:

1. Dilute protein samples to 0.2 mg/mL.
2. Prepare a stock solution of ThT. This solution can be filtered and stored in the dark for one week. Before use, the stock solution is diluted in 10mM phosphate buffer pH7 containing 150 mM NaCl (dilution 1/50 = working solution); this must be done extemporaneously.
3. For a spectroscopic assay, add a 10  $\mu$ L aliquot of protein to 1 mL of working solution, incubate for 1 min and measure the fluorescence intensity with excitation set at 440 nm and emission detected in the range 450–500 nm (*see Note 2*).
4. For slide staining and visualization under a microscope, place 10  $\mu$ L aliquots of stained samples and image using a fluorescence microscope with a 60 $\times$  oil-objective (NA = 1.4). The laser excitation can be set at 415 or 450 nm and emission is detected at 482 nm.

### *3.1.2 AFM imaging*

Atomic force microscopy (AFM) has become a common tool for the analysis of protein aggregation and fibrilisation [20]. It provides height information that complements structural information from two-dimensional projection images obtained using electron microscopy (*see 3.2.3*). In essence, an AFM set up consists of a microtip attached to a cantilever held on a piezo scanner that moves along the surface of the sample with high precision over a defined distance within the micrometer range. The deflection of the



cantilever follows the surface topography and is monitored following the position of a laser spot reflecting from its surface on a photodiode. The motion of the cantilever can also be used to interrogate the mechanical properties of the sample. AFM imaging thus provides information about biological objects dimensions at nanometer resolution [21]. As amyloid fibrils are usually in the  $\mu\text{m}$  range in length with a lateral dimension between 3 and 30 nm, AFM is thus perfectly adapted to probe fibrils structure.

1. Substrate preparation. In order to study fibrillar protein structures, it is essential to first adsorb them to a flat surface. This step is crucial for obtaining high quality data. The fibers formed in solution have to be at an adequate density and firmly attached to allow the tip to scan over them. One of the surfaces that is the most frequently used for AFM studies is mica, although clean glass or annealed gold can also be used. Here we will concentrate on describing the adsorption of fibrils to mica surfaces. Mica is convenient because it is atomically flat and can be easily cleaved immediately before use to obtain a clean surface. AFM-grade mica can be easily obtained from several AFM products suppliers. The freshly cleaved mica is glued to the sample holder using, for example, double sided tape before incubating the protein sample.

2. Protein adsorption. A protein solution is incubated on the mica surface to allow for fibril adsorption to the surface (see **note 3**). The size, charge and mechanical stability of the fibers determine the details of the preparation protocol. Basically, this protocol consists of two steps: first, conditioning the surface where the fibers will be deposited and second, incubating the protein solution to allow for filament deposition. As the surface/protein interaction will be affected by the surface charge of the protein and surface, the pH of the solution has to be selected to optimize this interaction. Precisely, mica is constituted of sheets of octahedral hydroxyl-aluminium lying between two silicon tetrahedral layers. The most favourable cleavage plane is located in the layer of unbound  $\text{K}^+$  that lies in the interlayer space between neighbouring silicon layers. When immersed in water, the hydrated potassium ions can be dissociated from the mica surface, leaving negatively charged aluminium tetrahedra ( $\text{AlO}_2^-$ ) in the first outer layer [22]. Thus, the solution pH and ionic salt concentration

should be adjusted to allow for the protein to be positively charged to promote its interaction with the negatively charged mica surface.

3. Imaging in air. If the protein binding to the surface is not strong enough to allow imaging with the sample immersed in solution, the best alternative is to image in air. Incubate a small amount of diluted protein (typically 10  $\mu\text{l}$  of 5–10  $\mu\text{g}/\text{ml}$  fibrils) on the mica surface and allow to air dry at room temperature. All the fibers in the solution will now be adsorbed on the mica surface. Before imaging, the sample must be extensively washed with Milli-Q water to remove the salt from the buffer and then carefully air-dried again. AFM can operate in different modes, and the ones more conveniently used to image amyloid self-assemblies are contact mode and tapping mode, depending on the characteristics of the fibrils of interest [21]. Contact mode images are obtained when the tip is in contact with the sample, so it requires that the fibers are firmly attached to the surface and rigid enough to withstand the tip pressure as it scans along the surface without deformation or sliding. In tapping mode, the cantilever is oscillated at its resonant frequency during the scan, so that it does not drag the sample along as it moves over the surface. Resonant frequencies of the soft cantilevers used for biological materials are, typically, within the 70-10 KHz range, depending on whether the microscope is operated in air or in solution, so that the interaction time between the tip and sample is reduced dramatically to only a few microseconds. Fibers can be imaged either in tapping or in contact mode, depending on the characteristics of the sample.

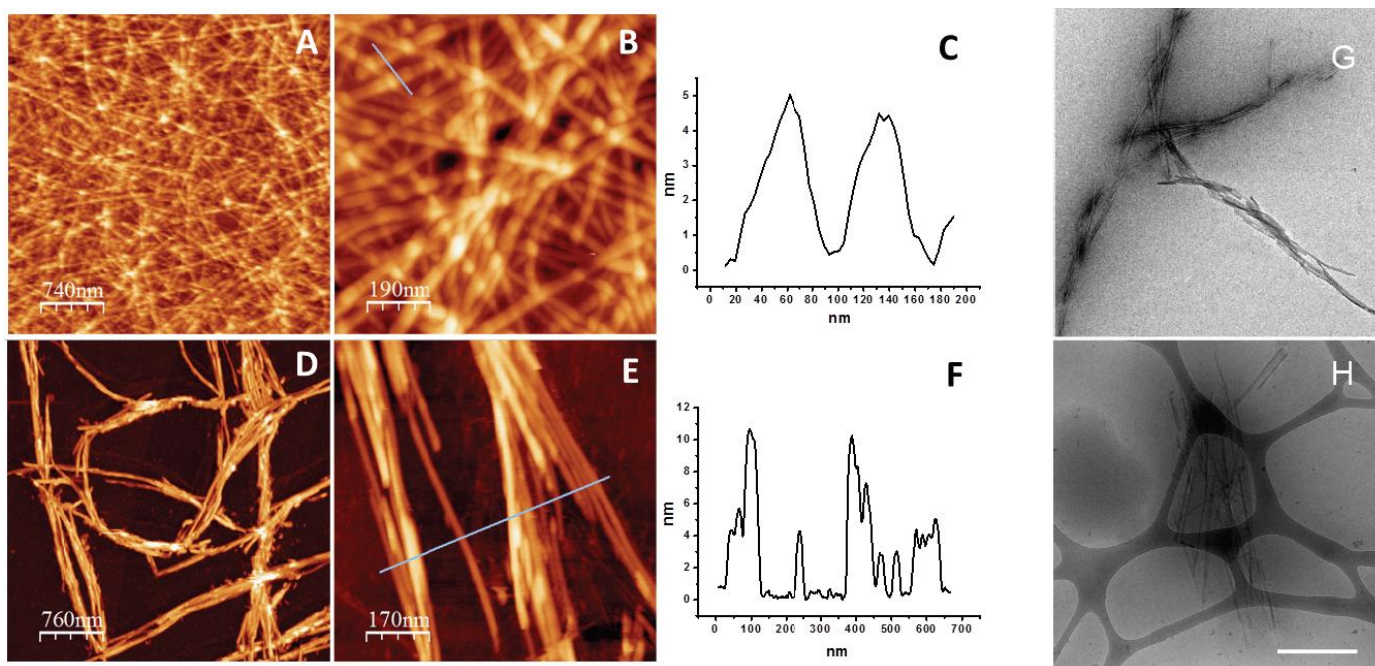
4. Imaging in solution. If the fibers are sufficiently long and their interaction is strong enough to allow for a stable adsorption to the surface, the fibers will remain firmly attached during imaging in solution. After allowing 10-60 minutes of incubation, the sample is extensively rinsed with buffer to remove excess protein. The imaging buffer is selected to optimise the interaction between the tip and the sample and could differ from the one selected for fiber adsorption to the mica. The pH or salt concentration can be adjusted again so that the tip-protein interaction is minimized to reduce damaging the fibers during imaging. Operating in solution offers the possibility of imaging fiber growth. In such a case, imaging should be done in the presence of small amounts of protein in the solution to avoid

interfering with the imaging process. Images of the same area taken at different times can be used to follow filament growth kinetics.

5. AFM imaging procedure. Once the sample is properly prepared, the AFM is operated selecting the most appropriate tip for the imaging mode. If imaging in air, stiffer tips with force constants in the range of 2-75 N/m and resonant frequencies in the range of 100-320 kHz can be more convenient to reduce the snap-to-contact region. For imaging in solution, softer tips, with force constants in the range of 0.01-0.8 N/m and resonant frequencies in the range of 100-20 kHz can be used. In brief, AFM operation includes the following steps: placing the cantilever in the holder and then both to the scanner; aligning the laser to impinge the cantilever tip; aligning the laser reflection on the center of the photodiode; placing the sample in the sample holder; if operating in tapping mode, the cantilever resonance frequency has to be selected; approaching the tip and sample. Once the tip is touching the sample, the scanning parameters: gains, scanning speed and setpoint are adjusted to obtain reproducible trace and retrace images.

6. AFM topographic analysis. AFM obtains direct topographical images of the sample surface. Each AFM provider includes software to analyse the images, but there are also free specialized softwares available to process scanning probe microscopy images (WsXM or Gwyddion).

Figure 1A-F shows Hfq peptide fibers imaged in air or in solution.



**Figure 1. Visualization of Hfq self-assembly by molecular microscopy.** Transmission electron microscopy (TEM) and atomic force microscopy (AFM) are efficient tools to investigate protein nanostructures. **A-F.** AFM images of Hfq peptide fibers in air and under liquid. Upper panels (A,B,C) are fibers imaged in air. A drop of the protein solution was allowed to dry on the mica before imaging in contact mode. Lower panels (C,D,F) show filaments observed under buffer at pH 4 in tapping mode, after incubating the fibers for five minutes on mica at pH 4 at a concentration of 0.5 mg/mL. Panels C and F shows the height profile under the blue lines shown, respectively, in panels B and E. **G-H.** In TEM experiments, 5  $\mu$ L of Hfq solution is deposited on a glow-discharged electron microscopy grid covered by a continuous carbon film for negative staining (G) or a holey carbon film for cryo-TEM (F). **G.** The sample is dried and subsequently stained with a solution of either uranyl acetate or uranyl-less (filling the protein cavities with heavy atoms). **H.** The excess of sample is blotted and subsequently frozen in ethane at liquid nitrogen temperature (embedding the sample in thin amorphous ice layers). Both methods show filaments that seem more aggregated in negative staining. *Scale bar: 500 nm.*

### 3.1.3 TEM imaging

A quick way to check the presence and structural characteristics of fibers (length, diameter, interweaving), as well as the homogeneity of the sample, is negative staining and observation by transmission electron microscopy.

#### *Protocol for Negative staining*

1. Sample preparation. Deposit a 5  $\mu$ l drop of Hfq sample at 1 mg/mL in 50 mM phosphate buffer pH7 (*i.e.* stock solution diluted in water) on a

glow-discharged carbon-coated electron microscopy copper grid (200 mesh square grid, EMS). In order to optimize the adsorption of the sample to the grid it is recommended, but not strictly required, to realize a glow-discharge (gas ionization process giving the carbon film hydrophilic).

2. Sample adsorption. After 5 minutes of interaction, blot out the excess sample using a Whatman filter paper (1-3 minutes is usually enough to have efficient adsorption of the sample onto the grid).

3. Contrasting of the sample. The adsorbed sample can then be put in contact with the contrasting agent. To perform negative staining, 5  $\mu$ l of uranyl acetate solution (2%) is applied onto the grid-containing sample. After 30 seconds incubation, the excess of uranyl acetate is blotted out and then the grids can be kept in a dry dark dust-free environment until observation at the electron microscope. Gadolinium salt (uranyl-less) is also used for negative staining. It is used instead of uranyl acetate because it is safer. However, it may create artefacts such as crystals. Store it at cold temperature and use it at room temperature to avoid salt deposition as much as possible.

The grids can be stored at room temperature for further observation on any standard transmission electron microscope.

4. Sample observation. The electron microscopy grid is then mounted onto a room temperature-equilibrated holder and subsequently introduced into an electron microscope (JEOL 2200FS, JEOL, Tokyo, Japan).

5. Image acquisition. 2k by 2k images are acquired using a Gatan Ultrascan 894 US1000 slow scan CCD Camera (Gatan, Pleasanton, CA, USA) at 40,000x nominal magnification (corresponding pixel size was 0.32 nm).

Figure 1G shows Hfq peptide fibers imaged by TEM after negative staining.

While it is true that negative staining is the tool of choice for quick characterization, it implies the dehydration of the sample, which can lead to artefactual aggregations of fibers or to structural modifications. Therefore, when high-resolution ultrastructural studies are required, samples should be observed in hydrated conditions, which is possible by cryo-transmission electron microscopy (cryo-TEM). Cryo-TEM is based on

fast-freezing of the electron microscopy grid containing a thin layer (100-200 nm or less) of the adsorbed sample by quick immersion in liquid ethane cooled-down at liquid nitrogen temperature. This can be achieved by using a lab-made or commercial apparatus. Some commercial apparatuses can be fully automatized so that repetitiveness can be easily achieved when freezing conditions have been set up.

*Protocol for cryo-electron transmission microscopy (Cryo-TEM)*

1. Sample preparation. Deposit a 5 $\mu$ l drop of Hfq suspension on a glow-discharged lacey carbon film grid (300 mesh).
2. Sample freezing. Blot the suspension through the grid for about 1 second using a Whatman filter paper to create thin suspended films in the carbon holes of the grid. Using a Leica EM-CPC, plunge the grid into liquid ethane cooled-down with liquid nitrogen. Prior to sample freezing, the grid is kept in a liquid nitrogen storage unit until observation at the electron microscope.
3. Transfer to the electron microscope. The grid with the frozen hydrated sample is transferred to a cryo-TEM for observation. During this transfer process, and during observation, the sample should remain under liquid nitrogen to prevent water crystallization, which impedes the observation of the sample. Since the observed sample is fully hydrated, the obtained images reflect the near native structure of proteins and fibers. A simple but detailed description of negative staining and cryo-TEM techniques is described in [23]. Grids are transferred into a JEOL JEM 2200FS cryo electron microscope equipped with a  $\Omega$ -energy loss filter by using a Gatan 914 cryo-transfer system.
4. Image acquisition. Hfq peptide cryo-images can be acquired using a 2k by 2k Gatan Ultrascan 894 US1000 slow scan CCD Camera (Gatan, Pleasanton, CA, USA) at 40,000x nominal magnification (corresponding pixel size was 0.32 nm) with an energy window of 20eV.

Figure 1H shows Hfq peptide fibers imaged by cryoTEM.

***3.2 Analysis of Hfq amyloid-like self-assembly secondary and super-secondary structure***

### 3.2.1 Infrared (FTIR) spectroscopy

Fourier transform infrared (FTIR) spectroscopy can be used to analyse protein secondary and super-secondary structures such as amyloid. The peptide bond is an amide group (CONH) that gives rise to seven active absorption bands in the infrared. This method can thus give information on secondary structure (three bands carry information on the secondary structure, but the Amide I band is the most used, see below) and tertiary structure (through the Amide II band). Indeed, the Amide II spectral region studied in presence of D<sub>2</sub>O solution (*i.e.* Amide II', see **note 4**) is a marker of the partial opening of the protein and thus of the tertiary structure changes, during which the hydrogens remained within the protein core undergo H-D exchange [24]. The Amide I is most often used for secondary structure elucidation since it has the smallest contribution from peptide side chain absorptions. The Amide I band absorption is mainly due to the C=O stretching of the amide group vibration but this vibration is strongly delocalised over the amide C-N and C-H bonds giving rise to a strong coupling: the trans dipole coupling. The exact frequency of the absorption peak is thus sensitive to the angle of the amide bond, which is dependent on the local structure of the protein domain.  $\alpha$ -helices,  $\beta$ -sheets,  $\beta$ -turns, random coils, and 3-10 helices are each restricted to a different set of angles as seen in a Ramachandran plot, giving rise to amide peaks located at different frequencies. This gives rises to the sensitivity of infrared spectroscopy to the secondary structure of proteins. Here, we thus focus on the infrared zone in the region of the Amide I.

1. Hfq fibril samples (from 1 to 20 mg/mL) are loaded into liquid cell equipped with CaF<sub>2</sub> windows and a 6  $\mu$ m thick spacer. This configuration allows studying a sample in solution for several hours (8 on average) and therefore allows for following the aggregation kinetics.

2. FTIR absorption measurements are performed between 400 and 4000 cm<sup>-1</sup> through the use of an interferometer-based spectrophotometer coupled with an IR source and a Triglycine sulfate (DTGS) detector. The best way to have good transmission spectra is by performing a measurement in which 256 scans are collected and at least 5 independent measurements are averaged. Spectra at a nominal resolution of 2 cm<sup>-1</sup> are baseline corrected

and background subtracted before analysis. A spectrum of the buffer must be collected using the same experimental parameters for water subtraction. Buffer subtraction is performed by an iterative procedure to remove the signal originating from the water. The association band of water between 1800 and 2400  $\text{cm}^{-1}$  is used to compute the subtraction factor. Most commercial software packages propose an integrated subtraction routine.

3. Alternatively an attenuated total reflection (ATR) sampling set-up can be used. With an ATR set-up, the infrared beam light is reflected inside an infrared-transparent, high refractive index crystal. The crystal can be single bounce or have a multiple bounce geometry to increase sensitivity. An evanescent wave is created at the crystal-solution interface, which extends into the sample with a penetration depth typically around 1  $\mu\text{m}$  (depending on the wavelength and on the crystal's refractive index). A background spectrum without sample is first recorded. A total of 64 scans are usually averaged. Then,  $\sim 5\text{-}20$   $\mu\text{l}$  of fibril sample (1 to 20  $\text{mg/mL}$ ) are deposited directly on the crystal and covered with a cap to reduce evaporation. A spectrum of the buffer must also be recorded in the same conditions for latter subtraction.

4. The analysis consists of qualitative evaluation of secondary structure through the combination of 2<sup>nd</sup>-derivative, Fourier self-deconvolution (FSD), and curve-fitting (CF) analysis [25]. FSD and the 2<sup>nd</sup>-derivative can be used to find the number and exact positions of the overlapping peaks composing the Amide I band. CF involves fitting a series of model peaks to the experimental data. The best function for peak shape in liquid samples is usually a convolution of Lorentzian and Gaussian bands, the Voigt profile. Overlapping peaks composing the Amide I band are fitted with Voigt profiles with a bandwidth varying between 15 and 25  $\text{cm}^{-1}$ . The goodness-of-fit is generally evaluated by comparing the rest with the noise. The CF, being an iterative algorithm, is pursued until the rest is less than 5 times the noise. The integration of the area of the component bands used to generate the final curve fit can provide an estimate of the relative percentages of secondary structure present. Both qualitative and quantitative analyses can be achieved with dedicated software. Most software packages such as OMNIC<sup>TM</sup> (Thermo Fisher Scientific) and Opus (Bruker) provide curve-fitting algorithms.

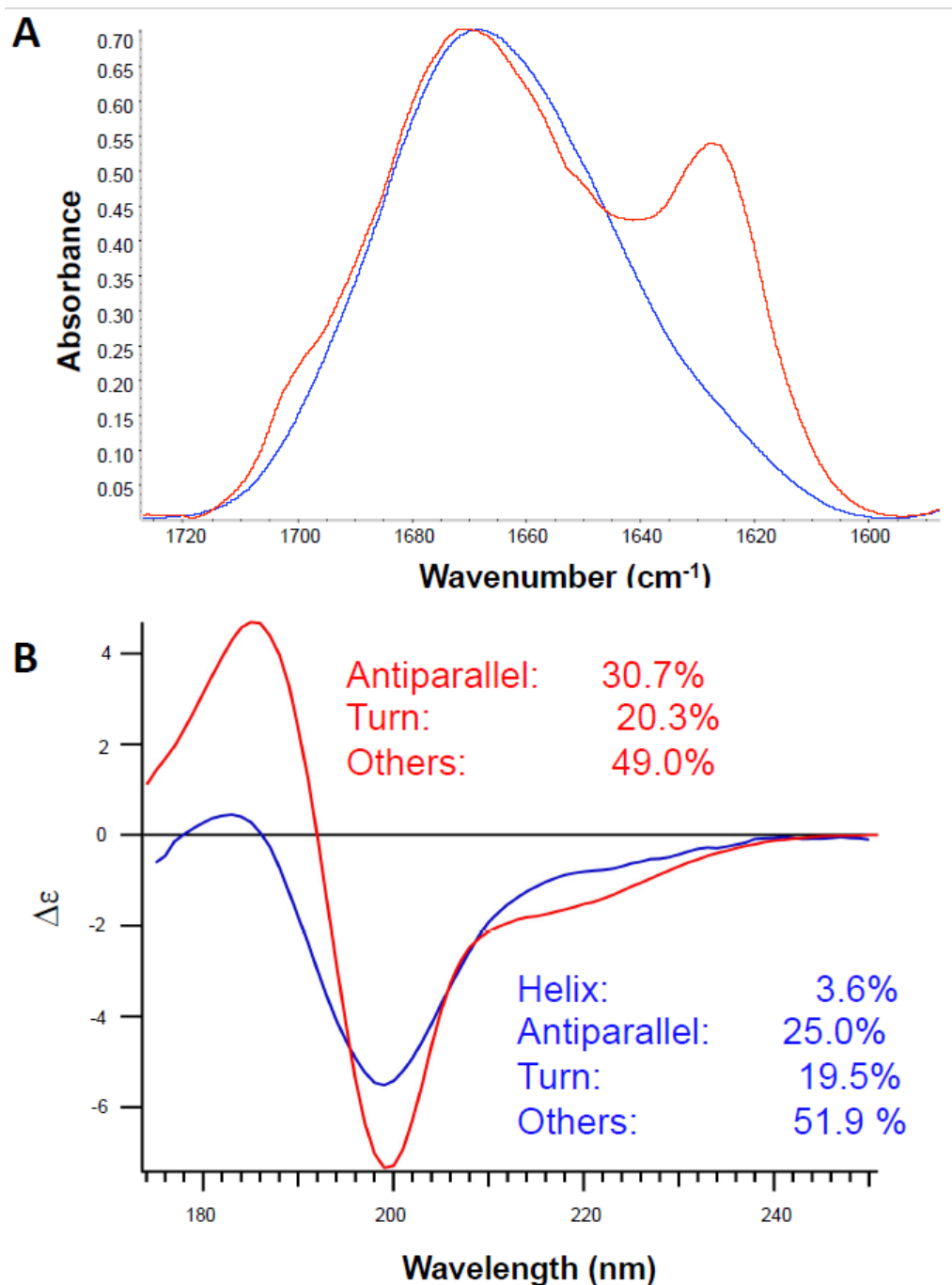


5. The assignment of a peak at a given wavenumber to a given secondary structure should be done accordingly to table 1. Briefly, a peak near 1645  $\text{cm}^{-1}$  is indicative of random coil, 1655  $\text{cm}^{-1}$  of  $\alpha$ -helix, and 1630  $\text{cm}^{-1}$  and 1690  $\text{cm}^{-1}$  of  $\beta$ -sheet (see **note 5**). Stronger hydrogen bonding results in a shift to lower wavenumbers and, therefore, amyloid fibrils often have  $\beta$ -sheet peaks below 1630  $\text{cm}^{-1}$  [26].

**Table 1:** Empirical Assignments for Amide I infrared (IR) bands characteristic of proteins [25]

Wavenumber ( $\text{cm}^{-1}$ )	Assignment
1615-1625	Intermolecular $\beta$ -sheet (in particular vibrations of strongly bound intermolecular aggregated $\beta$ -strands)
1630-1640	Intramolecular native $\beta$ -sheet
1640-1650	Disordered random coil
1650-1660	$\alpha$ -helix
1660-1695	$\beta$ -turn
1675-1695	Intermolecular $\beta$ -sheet (in particular anti-parallel aggregated $\beta$ -sheets)

Figure 2A shows Hfq peptide fibers analysed by FTIR spectroscopy.



**Figure 2. Secondary and supersecondary-structure determination.** *A.* FTIR spectra of Hfq peptide immediately after diluting in water (blue) and after incubation for a one day at 20°C (red). The peaks shown in figure 2A are due to intermolecular aggregated  $\beta$ -sheet at 1620 cm<sup>-1</sup>, to presence of native  $\beta$ -turn at 1670 cm<sup>-1</sup> and intermolecular aggregated  $\beta$ -sheet at 1690 cm<sup>-1</sup> respectively. In addition, in a lot of proteins studied, the contemporary increase of the components at about 1620 and 1680 cm<sup>-1</sup> indicates the building of anti-parallel  $\beta$ -aggregated structures; the low frequency  $\beta$ -sheet moiety at approximately 1620 cm<sup>-1</sup> are attributed to multi-stranded

intermolecular  $\beta$ -sheet and resembles the typical fingerprint of the cross- $\beta$  motif found in amyloid-like fibrils [36,37]. **B.** SRCD spectra of Hfq peptide at 20 mg/mL in water (blue) or after incubation at 20°C for a few days at 100 mg/mL to ensure the formation of fibrils (red). In this case, sonication procedure to fragment fibrils into shorter forms or their monomeric form was not necessary. Secondary structure determination by BESTSEL (inset) revealed a structural increase of 5% in  $\beta$ -sheet content for the fibrils as opposed to the monomeric Hfq peptide.

### 3.2.2 Synchrotron radiation circular dichroism SRCD

Circular Dichroism (CD) is a sensitive absorption spectroscopy technique for studying biological samples such as proteins and polynucleotides (DNA and RNA) as well as sugars. Circular left and right polarized light is differentially absorbed, e.g. by proteins due to the excitation of the n- $\pi$  and  $\pi$ - $\pi^*$  electronic transition of the peptide bonds. Additionally, the aromatic side chains produce characteristic absorptions between 210-275nm. Standardised CD spectra can be deconvoluted and used for protein secondary structure determination. Synchrotron Radiation Circular dichroism (SRCD) allows for the extension of the spectral range down to 168nm in aqueous solution and 120nm in hydrated films, with very good signal to noise ratios (typically  $\sim$ 20mdeg / 0.2mdeg). The extension of the spectral range down to 168nm has improved the information content obtainable from a SRCD spectrum. Especially for the weak CD signals of proteins with high  $\beta$ -sheet or disordered content, SRCD has improved the spectral analysis substantially [27]. Numerous algorithms exist for the estimation of the secondary structure composition from CD spectra. For the special case of  $\beta$ -sheet-rich proteins, with a broad range of protein folds, a recent publicly accessible algorithm BestSel [18], allows the determination of secondary structure contents including parallel and antiparallel  $\beta$ -sheets and potential fold recognition. The novel reference dataset with spectra that significantly differ from present reference sets, extends the information content for secondary structure determination.

1. Sample preparation. Homogenized Hfq peptide fibrils were prepared by diluted peptide in water at 20 mg/mL to 100 mg/mL. Observation of fibrils can be made directly after few days (see **note 1**) as described in 5.

Nevertheless, in some cases long and associated fibrils need to be fragmented to shorter and individual pieces to be analysed, as described below (2 to 4).

2. Seeds can be prepared by a sonication procedure [28]. The sonication strength and time had to be carefully chosen. At low force, homogenization of the fibril solution will just decrease the viscosity by breaking the fibrils into shorter pieces. A micro tip sonicator set-up in the cold room needs to be regulated to minimal strength (force, amplitude), with a 50% duty cycle (e.g. 1s pulses interrupted by 1s silence). Sample volume should be at least 100  $\mu$ l in microtubes. Sonication shall be carried out with the tubes containing sample cooled in ice-cold solution (for better contact), with the tip plunged right to the bottom of the tube without touching the tube wall. For homogenization of peptides and small proteins four cycles of five pulses with a 10 second rest between cycles produces reliable and reproducible results. During the rest period, mix well the solution containing the peptide by lifting and plunging the tip up and down. Droplet formation on the tube wall should be avoided. Gloves, mouth mask and glasses are recommended for safety.

3. SRCD spectra are taken before and after sonication allowing the spectral distinction between fibril seeds and amyloid fibrils.

4. Ultracentrifugation in Beckmann TLA-100 rotor run at 53500-96500 Relative Centrifugal Field (RCF) xg for 15-30min may be used to distinguish eventual monomers floating in the supernatant and the fibrils accumulating in the pellet. The baseline spectrum for the fibril seeds should be taken then from the supernatant. Once the sonication and centrifugation is finished, keep samples at 4°C.

5. Sample concentration, buffer composition and pathlength choices for SRCD data acquisition should be handled with care. In general concentrations should be for  $\alpha$ -helix rich proteins 1-2g/l, for  $\beta$ -sheet rich proteins 3-4g/l and for unordered structures above 5g/l in a 20  $\mu$ m pathlength cell. Loading volumes are 2 microliters in special 20  $\mu$ m CaF<sub>2</sub> cells [16] (figure 3). The choice of buffer is a function of how much chloride will be considered essential. (10-50mM NaCl allows obtaining spectra down to 185nm). Chloride is readily replaced by fluoride allowing twice the initial

chloride concentration. Ideally phosphate buffer is used being more temperature stable than Tris-Cl and being deep-UV transparent.

6. All spectra are normalized to the mean residue weight ellipticity ( $\Theta_{\text{MRW}}$ , deg.cm<sup>2</sup>.dmole<sup>-1</sup>).

7. For secondary structure determination with BestSel, normalised root-mean-square deviation (NRMSD) indicates the most accurate fit for each spectrum; values of <0.15 are considered significant.

Figure 2B shows Hfq peptide analysed by SRCD spectroscopy.

### *3.3 Characterisation of cross- $\beta$ structure*

Small Angle X-ray Scattering (SAXS) is a simple and powerful technique that directly provides structural information at scales ranging typically from 1 nm to 1  $\mu\text{m}$  [29-31]. The physical principle is the same as X-ray crystallography; it is based on elastic scattering of an incoming X-ray wave by the electrons of the sample. The resulting “signal” is a scattering pattern most frequently collected using a bi-dimensional detector positioned behind the sample and perpendicular to the incoming beam. Practically, SAXS experiment can bring information for either peptides in solutions or assembled materials made of peptides. In this section, we will focus on how to prepare and analyse amyloid peptides and particularly Hfq cross- $\beta$  structure.

#### *3.3.1 Sample preparation*

1. Different solutions of peptide are prepared (25  $\mu\text{l}$  to 100  $\mu\text{l}$ ), ranging from 0.5% w/v to the highest concentration by 2.5% steps by dissolving the peptide into an appropriate buffer. The buffer must be the same for the previously described experiments (FTIR, SRCD...) in order to compare the results; buffer containing phosphorous should be avoided and low salt concentration is better.

2. When all the solutions are ready, directly fill quartz capillaries of 1.5 mm diameter and 0.01 mm thickness. Then, capillaries are sealed using wax or nail polish. The capillaries containing the peptides are left at room temperature until cross- $\beta$  strands are obtained (this is usually determined by other methods).

3. If the cross- $\beta$  strand formation kinetics are unknown, FTIR will be a better method than SAXS to determine the kinetics of assembly.

### 3.3.2 SAXS setup and recording

The SAXS experiment can bring information for both peptides in solutions or assembled materials made of peptides. The intensity collected is proportional to the number of objects within the sample,  $N$ . It directly results from their size, shape and internal structure, given by the Fourier Transformation modulus of their electronic density, showed here for an isotropic sample:

$$I(q) = r_e^2 \sum_{p=1}^N \left\langle \left| \iiint_{V_p} (\rho_e(r) - \rho_0) e^{-i\vec{q}\cdot r} d^3 r \right|^2 \right\rangle \Omega$$

where the triple integral is performed within the volume  $V_p$  of each particle and the average performed over all possible orientations  $\Omega$ .  $r_e$  is the classical electron radius,  $\rho_e(\vec{r})$  is the electronic density at position  $r$  within the particle,  $\rho_0$  is the average electronic density of the surrounding buffer,  $q = \frac{4\pi \sin \theta}{\lambda}$  is the momentum transfer,  $\theta$  is the scattering angle, typically below a few degrees, and  $\lambda$  is the X-ray wavelength. The rotational average severely limits the structural information content of the scattering pattern for a non-assembled material obtained in a SAXS experiment. When isolated macromolecules assemble into regular structures, new correlations appear in the electronic density of the whole structure, giving rise to specific modulations in the scattering intensity. When motifs are repeated along a defined axis with a repetition distance  $d$ , the scattered photon waves undergo strong constructive interferences at  $q$  values given by,  $q = n \frac{\pi}{d}$ , where  $n$  is an integer (Bragg law). The phenomenon is called diffraction, but is not different in nature from solution scattering. When identical units, e.g. a peptide or a protein, are regularly spaced, with periodicities typically on the order of or larger than 1 nm, diffraction is observed in the Small Angle range (SAXD) [32].

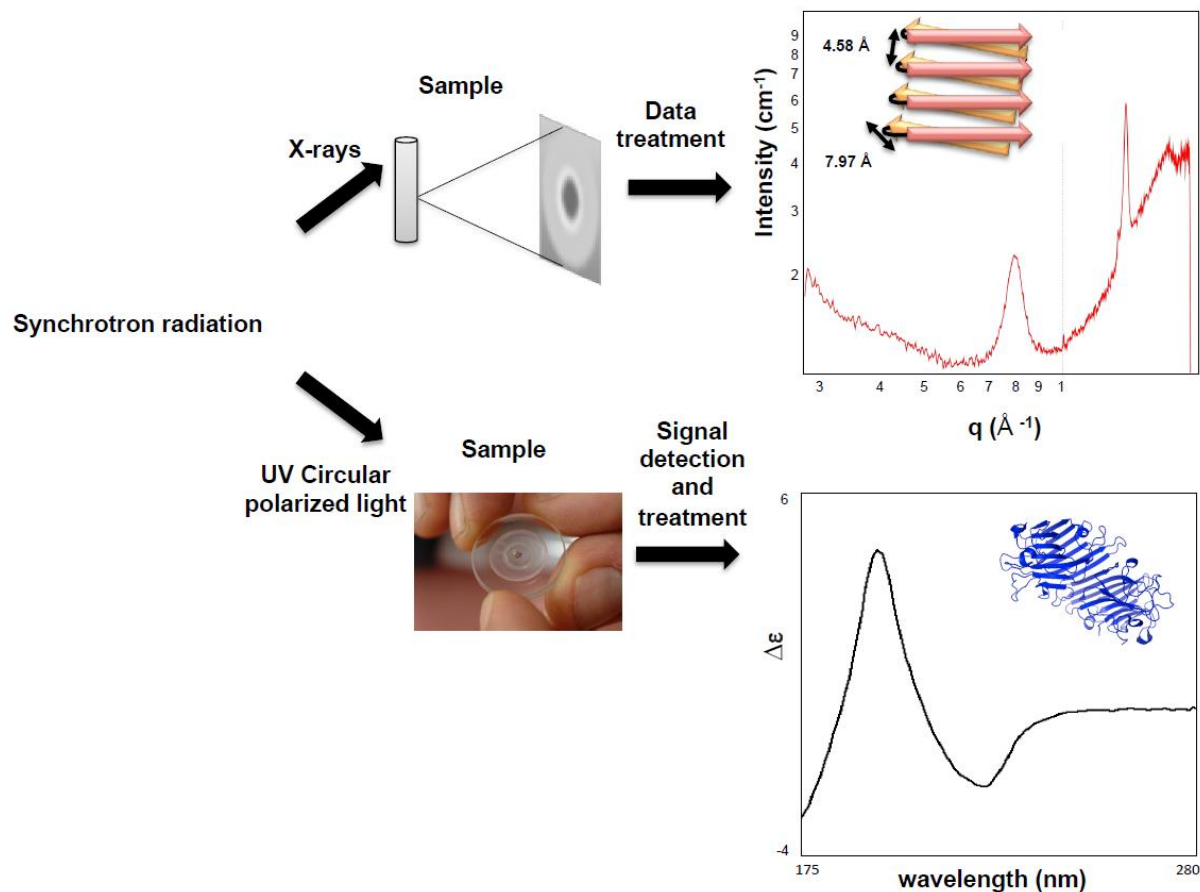
1. Cross- $\beta$  structure is investigated here. The sample-to-detector distance needs to be optimized. The typical signals resulting from  $\beta$ -sheet structure correspond to inter-distances of around 10 and 4.7 Å, the available  $q$ -range should thus be from 0.3 to 2 Å<sup>-1</sup>

2. Capillaries containing the samples and buffer (control) are mounted onto capillary holder system and positioned into the beam path.
3. Depending on the sample homogeneity, several positions within the samples capillaries can be chosen. At least 3 positions are enough: one at the top of the sample, one in the middle and the last at the bottom.
4. One important step here is the detection of radiation damage. When exposed to X-rays, organic molecules tend to form free radicals, leading to aggregation of the sample onto the capillary wall. Radiation damage does not occur for all samples. This phenomenon is observable after data treatment by an increase of the curve intensity at small  $q$ . In order to avoid this problem, samples should be exposed to X-rays and then let to rest for a few second before a new exposition at the same spot.
5. As samples on the same capillary holder can vary greatly, the exposure time must be adjusted, in order to get enough signal and to avoid saturation on the detector (depending of the beamline detector).
6. Due to radiation damage, acquisition time should not be higher than 3 seconds with at least 5 seconds of pause. It is better to make 10 acquisitions of 3 seconds than 1 of 30 seconds.
7. Record the scattering pattern with short and identifiable sample and buffer names

### *3.3.3 Data treatment*

1. When the entire scattering patterns are recorded as 2D images, radial integration is performed in order to obtain SAXS curves for each position inside all the sample and buffers. Such operation can be done using the Foxtrot software or equivalent.
2. For curves corresponding to the same position within the same sample, averaging is suggested in order to obtain better statistics. This must be done carefully in order to avoid averaging data with radiation damage.
3. As samples are analysed within the buffer, the buffer signal needs to be subtracted. Data subtraction is provided by most of the SAXS data reduction software.

4. When the resulting final curve (averaging + subtraction of buffer) is obtained, diffraction peaks corresponding to cross- $\beta$  sheets should be observed. The example of Hfq C-terminus peptide is shown on figure 3 (top), where the anisotropic reflections are indicative of a partially aligned fiber, perpendicular to the X-ray beam. The reflections at  $d=4.58\text{\AA}$  and  $d=7.97\text{\AA}$  correspond respectively to the inter-strand and inter-sheet spacing.



**Figure 3.** Use of Synchrotron radiation for the analysis of amyloids self-assemblies. **Top:** Basic sketch of data acquisition and treatment at a SAXS beamline. The sample scattering pattern recorded by a 2D detector is radially averaged to and background is subtracted to give the following SAXS curve. The CTR peptide of the Hfq protein exhibits typical diffraction peaks corresponding to a cross- $\beta$  structure sketch in overlay. **Bottom:** Data acquisition and treatment at a SRCD beamline. A rich  $\beta$ -strands containing protein is shown as an example (Concanavalin A, PDB ID 4PF5 [38]).

### Concluding remarks



The self-assembly of biological macromolecules constitutes a key process in all living organisms. We recently reported that sRNA cofactor Hfq belongs to the family of functional amyloids [9]. The various experimental approaches described in this chapter aim at further investigating the self-assembling properties of sRNA cofactors. This could indeed represent a versatile means to regulate sRNA-related processes *in vivo*. Furthermore, this would also enable the formation of mixed synthetic sRNA:protein synthetic self-assemblies, with future perspectives for nanotechnologies.

#### 4. Notes

1. The main obstacle for the examination of peptides and proteins that polymerise, and in particular those that form amyloids, is sample preparation. Indeed, the buffer used for the synthesis/purification processes and its composition (counter-ions, mainly chloride ions or TriFluoroacetic Acid TFA<sup>-</sup> ions used for peptide chemical synthesis) may affect fibrillation [33]. For this reason, batch-to-batch variability of synthetic peptides is frequently reported, resulting in a poor reproducibility of experiments in term of kinetics of self-assembling. Proper storage conditions are also important. Typically, peptides are most stable when stored lyophilized at -20 °C. Nevertheless, even under this condition, hydration occurs. This may affect the presence of pre-formed aggregates in samples. Therefore, in order to ensure comparison between various experiments, we do not recommend storing peptides and proteins for a long time and rather suggest ordering or preparing a fresh batch and performing experiments in a short timeframe.
2. Use of ThT staining is a widely used method and relatively well accepted as an indicator of the presence of amyloid fibrils. Nevertheless it has to be taken with precaution as ThT can also be an inhibitor of fibrillation [34].
3. A frequently used surface modification protocol for AFM involves fusing unilamellar lipid vesicles on glass or mica to form a supported lipid bilayer that can then be exposed to the protein fibrils. Indeed, interaction of amyloids with membrane is commonly reported and has been observed for Hfq [35].

4. D<sub>2</sub>O solutions may be used in samples to avoid the spectral overlaps between the Amide I band and strong absorption band of water at 1640 cm<sup>-1</sup>. Note that Amide I wavenumbers are lowered in D<sub>2</sub>O environment (5-10cm<sup>-1</sup>). In this case Amide I/II bands are referred to as Amide I'/II'.

5. The interpretation of an amyloid FTIR spectrum should usually begin with the examination of the primary sequence of the protein/peptide. Indeed, side chains such as asparagine and glutamine, which are very common in amyloid-forming proteins, have IR vibrations that overlap in the Amide I band.

#### ACKNOWLEDGEMENTS

This work was supported by Université Paris Diderot, CNRS, CEA and Synchrotron SOLEIL. We gratefully acknowledge help to MV from the French Embassy for their program for scientific and university cooperation. We are indebted to F. Gobeaux (CEA Saclay, Gif-sur-Yvette, France) and A. Deniset-Besseau (LCP, Université Paris-Sud, France) for many fruitful discussions. We thank Kimberly Stanek (University of Virginia) for her careful and critical reading of our manuscript.

#### 4. REFERENCES

1. Taghbalout A, Yang Q, Arluison V (2014) The Escherichia coli RNA processing and degradation machinery is compartmentalized within an organized cellular network. *Biochem J* 458:11-22.
2. Arluison V, Taghbalout A (2015) Cellular localization of RNA degradation and processing components in Escherichia coli. *Methods Mol Biol* 1259:87-101.
3. Lavelle C, Busi F, Arluison V (2015) Multiple approaches for the investigation of bacterial small regulatory RNAs self-assembly. *Methods Mol Biol* 1297:21-42.
4. Cayrol B, Geinguenaud F, Lacoste J, Busi F, Le Derout J, Pietrement O, Le Cam E, Regnier P, Lavelle C, Arluison V (2009) Auto-assembly of E. coli DsrA small noncoding RNA: Molecular characteristics and functional consequences. *RNA Biol* 6:434-445.
5. Busi F, Cayrol B, Lavelle C, LeDerout J, Pietrement O, Le Cam E, Geinguenaud F, Lacoste J, Regnier P, Arluison V (2009) Auto-assembly as a new regulatory mechanism of noncoding RNA. *Cell Cycle* 8:952-954.
6. Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. *Nat Rev Microbiol* 9:578-589.
7. Fischer S, Benz J, Spath B, Maier LK, Straub J, Granzow M, Raabe M, Urlaub H, Hoffmann J, Brutschy B, Allers T, Soppa J, Marchfelder A (2010) The archaeal Lsm protein binds to small RNAs. *J Biol Chem* 285:34429-34438.
8. Smirnov A, Wang C, Drewry LL, Vogel J (2017) Molecular mechanism of mRNA repression in trans by a ProQ-dependent small RNA. *EMBO J* 36:1029-1045.
9. Fortas E, Piccirilli F, Malabirade A, Militello V, Trepout S, Marco S, Taghbalout A, Arluison V (2015) New insight into the structure and function of Hfq C-terminus. *Biosci Rep* 35.
10. Maury CP (2009) The emerging concept of functional amyloid. *J Intern Med* 265:329-334.

11. Ostrowski A, Mehert A, Prescott A, Kiley TB, Stanley-Wall NR (2011) YuaB functions synergistically with the exopolysaccharide and TasA amyloid fibers to allow biofilm formation by *Bacillus subtilis*. *J Bacteriol* 193:4821-4831.
12. Zhou Y, Blanco LP, Smith DR, Chapman MR (2012) Bacterial amyloids. *Methods Mol Biol* 849:303-320.
13. Aguilera P, Marcoleta A, Lobos-Ruiz P, Arranz R, Valpuesta JM, Monasterio O, Lagos R (2016) Identification of Key Amino Acid Residues Modulating Intracellular and In vitro Microcin E492 Amyloid Formation. *Front Microbiol* 7:35.
14. Refregiers M, Wien F, Ta HP, Premvardhan L, Bac S, Jamme F, Rouam V, Lagarde B, Polack F, Giorgetta JL, Ricaud JP, Bordessoule M, Giuliani A (2012) DISCO synchrotron-radiation circular-dichroism endstation at SOLEIL. *J Synchrotron Radiat* 19:831-835.
15. Giuliani A, Jamme F, Rouam V, Wien F, Giorgetta JL, Lagarde B, Chubar O, Bac S, Yao I, Rey S, Herbeaux C, Marlats JL, Zerbib D, Polack F, Refregiers M (2009) DISCO: a low-energy multipurpose beamline at synchrotron SOLEIL. *J Synchrotron Radiat* 16:835-841.
16. Wien F, Wallace BA (2005) Calcium fluoride micro cells for synchrotron radiation circular dichroism spectroscopy. *Appl Spectrosc* 59:1109-1113.
17. Lees JG, Smith BR, Wien F, Miles AJ, Wallace BA (2004) CDtool-an integrated software package for circular dichroism spectroscopic data processing, analysis, and archiving. *Anal Biochem* 332:285-289.
18. Micsonai A, Wien F, Kernya L, Lee YH, Goto Y, Refregiers M, Kardos J (2015) Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc Natl Acad Sci U S A* 112:E3095-3103.
19. Reinke AA, Gestwicki JE (2011) Insight into amyloid structure using chemical probes. *Chem Biol Drug Des* 77:399-411.
20. Rhiannon GC, Christopher TG, Nicolas HV (2012) Characterization of Fiber-Forming Peptides and Proteins by Means of Atomic Force Microscopy. *Current Protein & Peptide Science* 13:232-257.

21. Dufrene YF, Ando T, Garcia R, Alsteens D, Martinez-Martin D, Engel A, Gerber C, Muller DJ (2017) Imaging modes of atomic force microscopy for application in molecular and cell biology. *Nat Nano* 12:295-307.
22. Liberelle B, Banquy X, Giasson S (2008) Stability of Silanols and Grafted Alkylsilane Monolayers on Plasma-Activated Mica Surfaces. *Langmuir : the ACS journal of surfaces and colloids* 24:3280-3288.
23. Thompson RF, Walker M, Siebert CA, Muench SP, Ranson NA (2016) An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods* 100:3-15.
24. Militello V, Casarino C, Emanuele A, Giostra A, Pullara F, Leone M (2004) Aggregation kinetics of bovine serum albumin studied by FTIR spectroscopy and light scattering. *Biophys Chem* 107:175-187.
25. Byler DM, Susi H (1986) Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers* 25:469-487.
26. Zandomenighi G, Krebs MR, McCammon MG, Fandrich M (2004) FTIR reveals structural differences between native beta-sheet proteins and amyloid fibrils. *Protein Sci* 13:3314-3321.
27. Wallace BA (2009) Protein characterisation by synchrotron radiation circular dichroism spectroscopy. *Q Rev Biophys* 42:317-370.
28. Kardos J, Micsonai A, Pal-Gabor H, Petrik E, Graf L, Kovacs J, Lee YH, Naiki H, Goto Y (2011) Reversible heat-induced dissociation of beta2-microglobulin amyloid fibrils. *Biochemistry* 50:3211-3220.
29. Guinier A (1955) Small-angle scattering of x-rays. Structure of matter series, Wiley, NY, USA.
30. Feigin LAS, Svergun DI, Taylor GW (1987) Structure analysis by small-angle X-ray and neutron scattering. Plenum Press, NY, USA.
31. Feigin O, Kratky O (1982) Small angle x-ray scattering. Academic Press, New York & London
32. Guinier A (1994) X-Ray Diffraction: In Crystals, Imperfect Crystals, and Amorphous Bodies. Courier Corporation, North Chelmsford, MA, USA.

33. Kaneko I, Tutumi S (1997) Matters Arising: Conformations of  $\beta$ -Amyloid in Solution. *J Neurochem* 68:438-439.
34. Stains CI, Mondal K, Ghosh I (2007) Molecules that target beta-amyloid. *ChemMedChem* 2:1674-1692.
35. Herrera AI, Tomich JM, Prakash O (2016) Membrane Interacting Peptides: A Review. *Curr Protein Pept Sci* 17:827-841.
36. Vetri V, D'Amico M, Fodera V, Leone M, Ponzoni A, Sberveglieri G, Militello V (2011) Bovine Serum Albumin protofibril-like aggregates formation: solo but not simple mechanism. *Arch Biochem Biophys* 508:13-24.
37. Vetri V, Militello V (2005) Thermal induced conformational changes involved in the aggregation pathways of beta-lactoglobulin. *Biophys Chem* 113:83-91.
38. Francois-Heude M, Mendez-Ardoy A, Cendret V, Lafite P, Daniellou R, Ortiz Mellet C, Garcia Fernandez JM, Moreau V, Djedaini-Pilard F (2015) Synthesis of high-mannose oligosaccharide analogues through click chemistry: true functional mimics of their natural counterparts against lectins? *Chemistry* 21:1978-1991.

Annexe D :

*Effects of Hfq on the conformation  
and compaction of DNA*

# Effects of Hfq on the conformation and compaction of DNA

Kai Jiang<sup>1,†</sup>, Ce Zhang<sup>1,†</sup>, Durgarao Guttula<sup>1</sup>, Fan Liu<sup>1</sup>, Jeroen A. van Kan<sup>1</sup>,  
Christophe Lavelle<sup>2</sup>, Krzysztof Kubiak<sup>3,4</sup>, Antoine Malabirade<sup>3</sup>, Alain Lapp<sup>3</sup>,  
Véronique Arluison<sup>3,5,\*</sup> and Johan R.C. van der Maarel<sup>1,\*</sup>

<sup>1</sup>Department of Physics, National University of Singapore, 2 Science Drive 3, 117542, Singapore, <sup>2</sup>Genomes Structure and Instability, Sorbonne Universities, National Museum of Natural History, Inserm U 1154, CNRS UMR 7196, 75005 Paris, France, <sup>3</sup>Laboratoire Léon Brillouin, UMR 12 CEA/CNRS, CEA-Saclay, Gif sur Yvette Cedex 91191, France, <sup>4</sup>Department of Molecular Biology, University of Gdansk, Wita Stwosza 59, 80-308 Gdansk, Poland and <sup>5</sup>Université Paris Diderot, Sorbonne Paris Cité, 75013 Paris, France

Received January 13, 2015; Revised March 17, 2015; Accepted March 18, 2015

## ABSTRACT

**Hfq is a bacterial pleiotropic regulator that mediates several aspects of nucleic acids metabolism. The protein notably influences translation and turnover of cellular RNAs. Although most previous contributions concentrated on Hfq's interaction with RNA, its association to DNA has also been observed *in vitro* and *in vivo*. Here, we focus on DNA-compacting properties of Hfq. Various experimental technologies, including fluorescence microscopy imaging of single DNA molecules confined inside nanofluidic channels, atomic force microscopy and small angle neutron scattering have been used to follow the assembly of Hfq on DNA. Our results show that Hfq forms a nucleoprotein complex, changes the mechanical properties of the double helix and compacts DNA into a condensed form. We propose a compaction mechanism based on protein-mediated bridging of DNA segments. The propensity for bridging is presumably related to multi-arm functionality of the Hfq hexamer, resulting from binding of the C-terminal domains to the duplex. Results are discussed in regard to previous results obtained for H-NS, with important implications for protein binding related gene regulation.**

## INTRODUCTION

Hfq is a widespread and phylogenetically conserved protein found at high concentration in bacteria. Consistent with this high level of expression, the protein is central to multiple regulatory processes. For instance, the initial description

of Hfq was associated with its involvement in the activity of bacteriophage Q $\beta$  replicase [hence its iconic name standing for host factor for phage Q $\beta$  RNA replication (1)]. However, pleiotropic functions of Hfq have been highlighted when the *hfq* gene was disrupted in *Escherichia coli*, yielding a decrease in growth rate, an increase in UV, oxidant and osmo sensitivity, as well as decreased negative supercoiling of plasmids (2). Most of these *hfq*-null phenotypes are due to the implication of Hfq in RNA regulation by small non-coding RNA (sRNA) (3). Indeed, Hfq is required to mediate sRNA stress-response. This regulatory mechanism is based on the hybridization of sRNA to its target mRNA, therefore altering mRNA expression. Most often, the hybridization of sRNA occurs in the vicinity of the ribosome binding site to prevent the initiation of translation, a process concomitant with a decrease in mRNA stability (4,5).

Similar to its eukaryotic homologs participating in pre-mRNA splicing and decay (6), Hfq is particularly important in RNA metabolism. It is structurally related to the Sm family of proteins. The amino-terminal region of Hfq forms a fold, which is also found in archaeal and eukaryotic Sm proteins (7). This N-terminal region (NTR) comprises five  $\beta$ -strands that form a bent  $\beta$ -sheet, capped by an  $\alpha$ -helix. The  $\beta$ -sheets from six monomers interact with each other to assemble into a toroidal structure. The main difference between Hfq and Sm proteins is the number of polypeptides forming the core of the torus: whereas Sm proteins consist in heptameric rings, Hfq is composed of six polypeptides and is hexameric (8). Although the mechanism by which Hfq binds RNA is not completely understood, it is now well-established that the NTR of the protein (~65 amino acids) binds RNA in the absence of the C-terminal region (CTR) (9), uridine-rich RNA sequences are bound to the proximal face, whereas purine-rich RNA sequences bind to

\*To whom correspondence should be addressed. Tel: +65 65164396; Fax: +65 67776126; Email: johanmaarel@gmail.com  
Correspondence may also be addressed to Véronique Arluison. Tel: +33 169083282; Fax: +33 169088261; Email: veronique.arluison@cea.fr

<sup>†</sup>These authors contributed equally to the paper as first authors.



the distal face of the torus (10,11). Furthermore, the hexamer exhibits a rim that has been proposed to participate in RNA binding (12).

As previously mentioned, in addition to its NTR domain forming the hexameric torus, the protein also includes a CTR domain of ~35 amino acids. This domain appears mainly non-structured (13) and its function has not yet been resolved. Indeed, the importance of the CTR tail of Hfq is a matter of controversy. On one hand, some studies have shown that tail-less Hfq is efficient for RNA binding and regulation (9,14). On the other hand, other studies have suggested that a lack of the tail reduces RNA-mediated regulation (15–18) and that some residues in the CTR interact with RNA (19). Moreover, we previously observed that Hfq is able to self-assemble and that this ability is dependent on the presence of the CTR (20). This is in accordance with the recent observation that the CTR of Hfq is flexible, a feature which could facilitate protein–protein interactions (13).

In addition to its role in post-transcriptional regulation, Hfq is capable of binding to DNA, but with a weaker affinity than to RNA (21–24). This was notably emphasized in cellular localization experiments, demonstrating that ~20% of Hfq is complexed with DNA (25,26), along with a tenth of other DNA-binding proteins within the bacterial nucleoid (23,27–28). Hfq appears to have a heterogeneous localization in the nucleoid, as opposed to the more uniform distribution of most nucleoid associated proteins such as H-NS (histone-like nucleoid structuring protein) or HU (heat-unstable nucleoid protein) (25). A few studies have shed light on the importance of Hfq in DNA metabolism, as constraining negative supercoiling *in vivo* (2) or signifying its role in transcription (29–31), replication (32) and transposition (33,34). Motivated by this unresolved role of Hfq, here, we investigate the effect of Hfq on the conformation and compaction of DNA using nanofluidics and atomic force microscopy. In parallel, information on the structural arrangement of bound Hfq on DNA is obtained with small angle neutron scattering (SANS).

A relatively new technology for the investigation of single DNA molecules and their complexes with ligands and proteins is nanofluidics. Quasi one-dimensional channels with cross-sectional diameters of one to a few hundred nanometer in combination with fluorescence microscopy can be used to study the conformation and compaction of single DNA molecules (35–37). In particular, it was shown that DNA confined inside such a nanochannel can be compacted into a condensed form for over-threshold concentrations of a neutral crowding agent (dextran) and like-charged proteins (bovine serum albumin and haemoglobin) (38–40). DNA condensation with DNA-binding proteins protamine and H-NS has also been reported (41,42). Here, bacteriophage T4-DNA molecules are pre-incubated with Hfq and brought into an array of nanochannels of various cross-sectional diameter. Information on the bending rigidity (persistence length) of the nucleoprotein complex and Hfq mediated bridging between different segments of the DNA molecule is then derived from the measurement of the stretch of well-equilibrated DNA molecules along the direction of channels, as well as the determination of the critical concentration of Hfq for the compaction of DNA into the condensed form. For reference, we also measure the size and

compaction of unconstrained DNA molecules in the bulk phase.

The results obtained with nanofluidics regarding the bending rigidity and Hfq mediated bridging are corroborated with atomic force microscopy. Atomic force microscopy is particularly useful to study protein–DNA interaction, because the complexes are directly visualized without relying on elaborate sample preparation. Here, we investigate the effect of Hfq on the persistence length from an analysis of the orientation correlation of single DNA molecules adsorbed on a silica surface without additional components to promote binding (43). Furthermore, we trace the length of the molecules along the contour to gauge a possible effect of Hfq on the rise per bp along the helical axis. In order to avoid complications related to looping, bridging and side-by-side aggregation, we use relatively short DNA fragments of 1000 bps. Hfq mediated bridging is confirmed, however, by imaging of longer DNA fragments consisting of 10 000 bps.

Nanofluidics and atomic force microscopy are not well adapted to the investigation of the structural arrangement of a protein about DNA. Typical structure-related properties are best and quantitatively inferred from scattering experiments. In the case of DNA, scattering studies have focused on the radial density profile of counterions away from the DNA axis (44–48). To the best of our knowledge, ordering of DNA-binding protein has never been investigated before by similar methods. An advantage of SANS is the possibility for variation of DNA and protein contrasts (49). In this contribution, we apply contrast variation in mixtures of H<sub>2</sub>O and D<sub>2</sub>O to obtain the scattering from the protein in solutions of persistence length (50 nm) DNA fragments. The scattering is analyzed in terms of a radial distribution of Hfq around DNA, to arrive at a consistent model of the binding of the Hfq hexamer on DNA and the formation of bridges with CTR multi-arm functionality.

Our results enable to propose a mechanism underlying the function of Hfq in coordinating DNA packaging and unveil a new role for this pleiotropic factor in the control of gene expression.

## MATERIALS AND METHODS

### Sample preparation

T4 GT7 DNA (T4-DNA, 165.65 kbp) was purchased from Nippon Gene, Tokyo and used without further purification. The integrity of T4-DNA was verified with pulsed gel electrophoresis. No fragments of ones to tens of kbps were observed. Hfq ( $M_w = 67$  kDa, hexamer) was purified from BL21(DE3)/pTE608 cells expressing untagged Hfq as previously described (50). Samples were prepared by dialyzing solutions of DNA against 10 mM Tris–HCl (T) buffer with the relevant concentration of NaCl and/or MgCl<sub>2</sub> in micro dialyzers. For one sample series, NaCl was substituted by potassium glutamate (KGlu). Solutions of Hfq in the same buffer were also prepared. The Tris–HCl concentration is 10 mM Tris adjusted with HCl to pH 7.5 (i.e. 8.1 mM Tris–Cl and 1.9 mM Tris). The ionic strength of the buffer was calculated with the Davies equation for estimating the activity coefficients of the ions and a dissociation constant  $pK = 8.08$  for Tris. Solutions of Hfq and DNA were subsequently

mixed and incubated for 12 h at 277 K. YOYO-1 fluorescence staining dye was purchased from Invitrogen, Carlsbad, CA, USA. DNA was stained with YOYO-1 with an incubation time of 12 h and an intercalation ratio of 100 bps per dye. No anti-photo bleaching agent was used.

### Electrophoretic mobility shift assay

The binding of Hfq to DNA was investigated with a gel shift assay. A 250-bp DNA fragment, including a 20 A-tract, was produced by polymerase chain reaction. The fragments (30 nM) were incubated with YOYO-1 (50 bps per dye) and various concentrations of Hfq (0.2–5  $\mu$ M) and electrophoresed on a 1.5% agarose gel. A control experiment was done without YOYO-1. The gel was stained with SYBR green I and imaged with a G:BOX system (Syngene, Cambridge, UK).

### Chip fabrication

The nanofluidic devices were fabricated by replication in polydimethylsiloxane (PDMS) of patterned master stamps (51,52). The nanochannels were made in hydrogen silsesquioxane (HSQ) resist (Dow Corning, Midland, MI, USA) using a lithography process with proton beam writing (53). An array of nanochannels is connected to two loading reservoirs through a superposing set of microchannels made in SU-8 resist with UV lithography. The heights and widths of the ridges in the master stamps were measured with atomic force microscopy (Dimension 3000, Veeco, Woodbury, NY) and scanning electron microscopy, respectively. Two stamps were made featuring nanochannels of length 60  $\mu$ m and rectangular cross-sections of 150  $\times$  250 and 200  $\times$  300 nm<sup>2</sup>, respectively. The stamp was coated with a 5-nm thick teflon layer to guarantee perfect release of the replicated PDMS chips (54). The stamps were replicated in PDMS followed by curing with a curing agent (Sylgard, Dow Corning) at 338 K for 24 h. The PDMS replica was sealed with a glass coverslip, after both substrates were plasma oxidized (Harrick, Ossining, NY).

### Nanofluidics

The pre-incubated and stained DNA molecules were loaded into one of the two reservoirs connected to the array of nanochannels. The DNA molecules were subsequently driven into the channels by electrophoresis. For this purpose, two platinum electrodes were immersed in the reservoirs and connected to a power supply with a relatively low voltage in the range 0.1–10 V (Keithley, Cleveland, OH, USA). Once the DNA molecules were localized inside the nanochannels, the electric field was switched off and the molecules were allowed to relax to their equilibrium state for at least 60 s. The stained DNA molecules were visualized with a Nikon Eclipse Ti inverted fluorescence microscope equipped with a 200 W metal halide lamp, a filter set, and a 100 $\times$  oil immersion objective. A UV light shutter controlled the exposure time. Images were collected with an electron multiplying charge coupled device (EMCCD) camera (iXon X3, Andor Technology, Belfast, UK) and the extension of the DNA molecules inside the channels was measured with IMAGEJ software (<http://rsb.info.nih.gov/ij/>). For intensity

threshold, we have used two times the signal to background noise ratio.

### Bulk phase imaging

A droplet of solution was deposited on a microscope slide and sealed with a coverslip separated by a 0.12 mm spacer. The YOYO-1 stained T4-DNA molecules were imaged with the above mentioned microscope. One minute video clips were collected with a frame rate of five frames per second. The video clips were analyzed with MATLAB (Natick, MA, USA). For each frame (time  $t$ ), the radius of gyration tensor of the imaged molecule was constructed according to  $\mathbf{S}(t) = \sum_{m,n} (\mathbf{r}_{mn} - \mathbf{r}_{cm})^2 I_{mn} / I_0$ , with  $I_{mn}$  the fluorescence intensity of pixel  $[m, n]$  at position  $\mathbf{r}_{mn}$  in the  $xy$ -plane and  $I_0$  is the total intensity of the frame. The center of mass  $\mathbf{r}_{cm}$  of the molecule was calculated according to  $\mathbf{r}_{cm} = \sum_{m,n} \mathbf{r}_{mn} I_{mn} / I_0$ . We determined the principal eigenvalues and vectors of  $\mathbf{S}$  using a singular value decomposition and with the assumption of cylindrical symmetry. We subsequently derived the length of the long axis according to  $L_{\parallel} = 12 < S_{\parallel} >$ , where  $S_{\parallel}$  represents the largest eigenvalue and the brackets denote an average over all frames.

### Atomic force microscopy

DNA fragments were purchased from Thermo Scientific (1000 bp, Waltham, MA, USA) and New England Biolabs (10 000 bp, Ipswich, MA, USA). All imaging experiments were done at room temperature in air with a Veeco Dimension 3000 atomic force microscope (Woodbury, NY). Images were acquired in the tapping mode with silicon (Si) cantilevers (spring constant of 20–100 N/m) and operated below their resonance frequency (typically 230–410 kHz). The images were flattened, and the contrast and brightness were adjusted for optimal viewing conditions. A 20- $\mu$ l droplet was spotted onto a silica surface. After 10 min to allow for DNA adsorption onto the surface, the specimens were developed by flushing them with ultra pure water followed by drying in a stream of N<sub>2</sub> gas.

### Small angle neutron scattering

DNA fragments (150 bp) were obtained by micrococcal digestion of calf thymus chromatin (55). A set of solutions was prepared by dissolving DNA and Hfq in mixtures of H<sub>2</sub>O and D<sub>2</sub>O to a concentration of 10 and 140 g/l, respectively. The solutions also contained 100 mM KCl. We applied contrast variation with four solvent compositions: 0, 40 (Hfq matched), 64 (DNA matched) and 100% D<sub>2</sub>O. Small angle neutron scattering was measured at ambient temperature with the PAXY diffractometer situated on the cold source of the Laboratoire Léon Brillouin. A wavelength of 0.6 nm with a 10% spread was selected. The scattering contributions pertaining to DNA and Hfq are then obtained by a simultaneous two-parameter fit to the data from the four solvent compositions (46,48).



## RESULTS

### Nanofluidics

T4-DNA molecules with a concentration of 3 mg of DNA/l were incubated with the relevant buffer for at least 24 h. It was checked that the DNA molecules are in the coil state before they were brought into the channels of the nanofluidic device with an electric field. After the field is switched off, the molecules relax to their equilibrium state within 60 s. We have verified that there is no further change in stretch for more than 3 h. A montage of images of single DNA molecules confined inside rectangular channels is shown in Figure 1. The equilibrated stretch in the longitudinal direction of the channel decreases with increasing concentration of Hfq. For over-threshold concentrations of Hfq, condensation of DNA into a compact form is observed. Condensed DNA is visible as a bright fluorescence spot and can easily be discerned from the extended form. We observed the same qualitative behaviour, irrespective channel cross-section and ionic composition of the buffer. There are quantitative differences however in the values of the stretch, relative decrease with increasing concentration of Hfq and critical concentration of Hfq for condensation. For buffer molarities exceeding  $\sim 40$  mM (10 mM Tris/HCl with 30 mM NaCl or KGlu), fluorescence imaging of the DNA molecules becomes problematic because of quenching and reduced binding of the dye. Furthermore, at high ionic strength the stretch becomes too short for a reliable detection of the transition from the coil to the condensed state.

We have measured the extension of the DNA molecules with two channel systems with channel cross-sections of  $150 \times 250$  and  $200 \times 300$  nm<sup>2</sup>, respectively. For each experimental condition, that is buffer composition, channel diameter and Hfq concentration, we have used a fresh PDMS replica and measured  $\sim 30$  molecules. The distribution in extension is close to Gaussian (56). An example of such a distribution is shown in Figure 1D. Fragmented DNAs can easily be discerned, because their extensions clearly fall below the values pertaining to the intact molecules. For the cut off, we have used the mean value minus two times the standard deviation. Resolution broadening can be neglected, because the optical resolution is one order of magnitude smaller than the variance. The mean relative extension  $R_{||}/L$ , that is the mean extension divided by the contour length of 57  $\mu$ m, is set out in Figure 2 as a function of Hfq concentration. With increasing concentration of Hfq, a relative decrease in extension by 10 to 60% is observed with the largest value for molecules bathed in a buffer of higher ionic strength. Due to the increased screening of electrostatic interaction, the extension decreases with increasing ionic strength. In the presence of MgCl<sub>2</sub> or KGlu, the same qualitative behavior is observed. Note that for subthreshold concentrations of Hfq the relative extensions are in the range 0.05–0.3, which implies that the DNA molecules remain coiled. Furthermore, related to the stronger confinement, an increased stretch is observed in the  $150 \times 250$  nm<sup>2</sup> channel system.

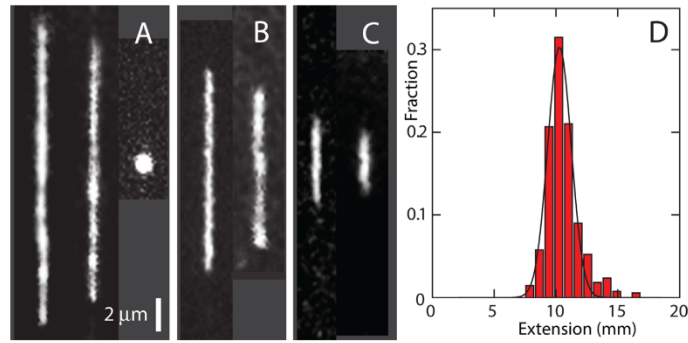
For over-threshold concentrations of Hfq, the DNA molecules compact into a condensed form. This is facilitated by the confinement inside the nanochannel, because

we did not observe condensation in the feeding microchannels and/or reservoirs of the chip. The critical concentration of Hfq for condensation decreases with increasing ionic strength and/or in the presence of a submillimolar concentration of MgCl<sub>2</sub>. Furthermore, the threshold shifts to a higher value with decreasing channel cross-sectional diameter. The highest observed critical concentration of Hfq for condensation inside  $150 \times 250$  nm<sup>2</sup> channels and at an ionic strength of  $\sim 8$  mM (T-buffer) corresponds with an Hfq to bp ratio of 1:40. The latter ratio decreases dramatically through hundreds to thousands of bps per Hfq hexamer for wider channels and with increased concentration of salts.

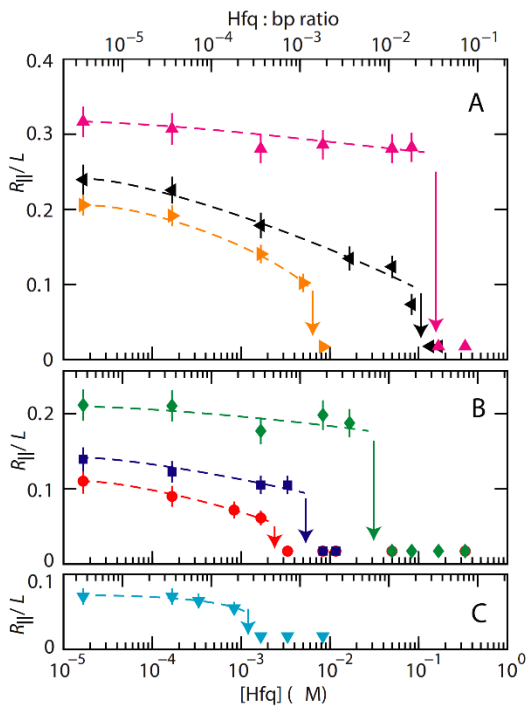
### Condensation in the bulk phase

We have also investigated how the coil-size of unconstrained DNA molecules changes with the addition of Hfq. Fluorescence microscopy experiments were done with 0.03, 0.3 and 3 mg of T4-DNA/l in T-buffer and T-buffer with 3 and 30 mM of added NaCl (0.03 mg of DNA/l only). The molecules were observed to be slightly anisotropic. We measured the radius of gyration tensor and derived the length of the long axis of the molecules. The results, as well as some characteristic images, pertaining to 0.03 mg of DNA/l are shown in Figure 3 (notice that the nanofluidics experiments were done with a loading buffer containing 3 mg of DNA/l). Qualitatively, we observe the same behavior as for the measurement of the extension in nanochannels. A difference is that in the bulk phase and minimal screening conditions, the long axis slightly increases with increasing Hfq concentration. Furthermore, in the bulk phase we observed coexistent condensed and non-condensed molecules at the critical concentration of Hfq, in agreement with a discontinuous, first order transition. In the channel system, the transition appears to be continuous (38). Quantitatively, there are differences in the scale and critical concentration of Hfq for condensation. In the bulk phase, the critical concentration depends on the concentration of DNA. The critical concentrations are 0.05, 0.08 and 0.5  $\mu$ M Hfq for 0.03, 0.3 and 3.0 mg of DNA/l, respectively (T-buffer). For the  $200 \times 300$  and  $150 \times 250$  nm<sup>2</sup> channel systems, the corresponding critical concentrations are 0.03 and 0.1  $\mu$ M Hfq, respectively (Figure 2). Condensation in the bulk phase requires, accordingly, an order of magnitude higher concentration of Hfq.

Fluorescence microscopy was done with a minimal level of staining of 100 bps per YOYO-1 molecule. For such a low staining level, there is no appreciable effect on the contour length, net charge and bending rigidity of the DNA molecules (51,57). To check whether there are complications associated with the staining, we have also done a series of experiments with 50 bps per dye. We obtained identical results for the stretch and condensation threshold. Furthermore, in an electrophoretic mobility shift assay (Supplementary Figure S1), the observed band shift is not affected by YOYO-1. Accordingly, staining of DNA with YOYO-1 in the range 50–100 bps per dye does not affect the formation of the Hfq–DNA complex.



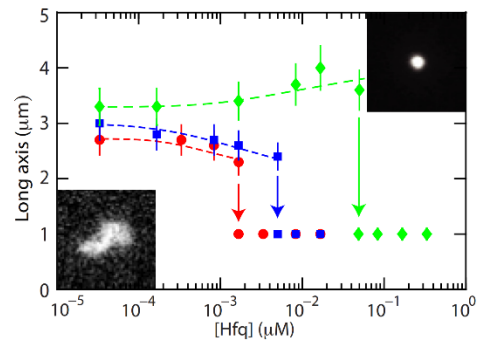
**Figure 1.** (A) Montage of fluorescence images of T4-DNA molecules inside  $150 \times 250 \text{ nm}^2$  channels and in T-buffer. From left to right  $2 \times 10^{-4}$ ,  $2 \times 10^{-3}$  and  $3.3 \times 10^{-1}$  (condensed)  $\mu\text{M}$  Hfq. (B) T4-DNA inside  $200 \times 300 \text{ nm}^2$  channels and in T-buffer with  $2 \times 10^{-4}$  (left) and  $2 \times 10^{-3}$  (right)  $\mu\text{M}$  Hfq. (C) T4-DNA inside  $200 \times 300 \text{ nm}^2$  channels and in T-buffer with 30 mM NaCl. The Hfq concentrations are  $2 \times 10^{-4}$  (left) and  $2 \times 10^{-3}$  (right)  $\mu\text{M}$ . (D) Distribution in extension of a population of 30 molecules in T-buffer with  $2 \times 10^{-3} \mu\text{M}$  Hfq. A Gaussian fit gives a mean extension  $R_{||} = 10 \pm 1 \mu\text{m}$ .



**Figure 2.** (A) Relative extension  $R_{||}/L$  of T4-DNA in T-buffer (magenta,  $\Delta$ ), T with 3 mM NaCl (black,  $\triangleleft$ ) and T with 30 mM KGLu (gold,  $\triangleright$ ) versus the concentration of Hfq (bottom axis) and Hfq to bp ratio (top axis). The molecules are inside  $150 \times 250 \text{ nm}^2$  channels. (B) As in panel (A), but for molecules inside  $200 \times 300 \text{ nm}^2$  channels. The buffers are T (green,  $\diamond$ ), T with 3 mM NaCl (blue,  $\square$ ) and T with 30 mM NaCl (red,  $\circ$ ). (C) As in panel (A), but inside  $200 \times 300 \text{ nm}^2$  channels and in T-buffer with 3 mM NaCl and 0.4 mM  $\text{MgCl}_2$  (cyan,  $\nabla$ ). The dashed curves are drawn as an aid to the eye and the arrows denote the condensation thresholds.

### Contour and persistence length

Binding of Hfq might affect the conformation of DNA through a change in mechanical properties of the duplex. To check this out, we have investigated the contour and persistence length of Hfq-coated DNA with atomic force microscopy. Linear DNAs of 10 000 and 1000 bps were used at a concentration of 3 mg/l. The DNAs were incubated

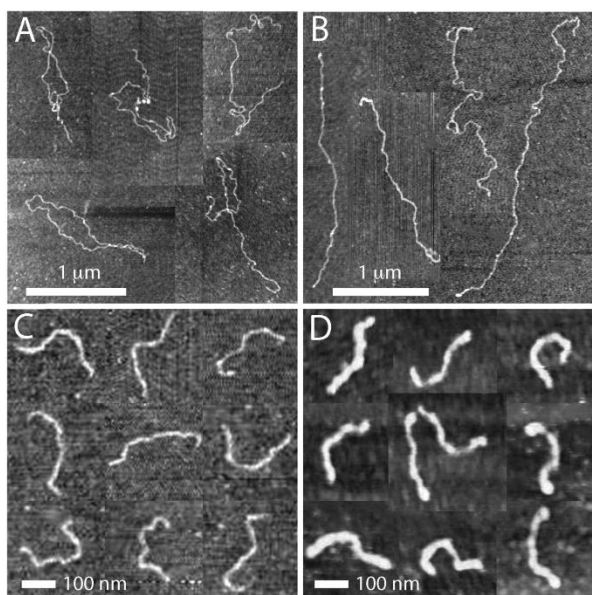


**Figure 3.** Long axis of unconstrained T4-DNA versus the concentration of Hfq. The buffers are T (green,  $\diamond$ ), T with 3 mM NaCl (blue,  $\square$ ) and T with 30 mM NaCl (red,  $\circ$ ). The dashed curves are drawn as an aid to the eye and the arrows denote the condensation thresholds. The insets show fluorescence images of a non-condensed (left) and condensed (right) T4-DNA molecule.

with 0.015, 0.30 or 0.75  $\mu\text{M}$  Hfq for 12 h, which correspond with an Hfq to bp ratio of 1:300, 1:15 and 1:6, respectively. For each sample, a droplet of 5  $\mu\text{l}$  was spotted onto a silica surface. The Hfq coated DNA molecules weakly adsorb to silica, so that no additions to the buffer are necessary to promote adhesion. Furthermore, the molecules equilibrate on the surface in a 2D conformation. Excess protein was removed by flushing the specimens with ultra pure water. Subsequently, the specimens were  $\text{N}_2$  dried. A series of typical images for increasing Hfq to bp ratio and for the two DNA molecular weights are shown in Figure 4.

For 10 000 bp DNA with an Hfq to bp ratio of 1:300 (Figure 4A), the length of the molecules measured along the contour agrees with the expected value based on molecular weight (3.4  $\mu\text{m}$ ). However, once the Hfq to bp ratio is increased to 1:15, intramolecular back folding, looping and side-by-side aggregation is observed (Figure 4B). The side-by-side aggregation is caused by Hfq mediated bridging of segments of the DNA molecule. Intramolecular aggregation complicates the analysis of the images for the determination of the contour and persistence lengths. As can be seen in panels C and D of Figure 4, shorter molecules of 1000 bps are visible as semi-flexible rods and do not exhibit





**Figure 4.** (A) Tapping mode atomic force microscopy images of DNA (10 000 bps) with Hfq to bp ratio of 1:300. (B) As in panel (A), but with Hfq to bp ratio of 1:15. (C) DNA (1000 bp) with Hfq to bp ratio of 1:300. (D) As in panel (C), but with Hfq to bp ratio of 1:6. Panels (A–D) are a montage.

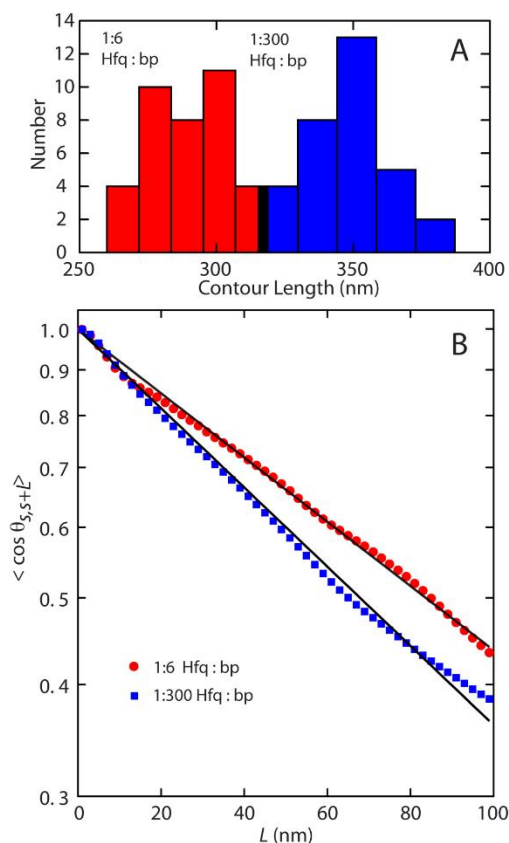
aggregation nor looping. Accordingly, we have used the latter molecules to investigate the effect of binding of Hfq to DNA on the contour and persistence length of the nucleoprotein complex.

To obtain the contour and persistence length, we traced the centerline of a population of 31–37 individual 1000 bp molecules. The distributions in length are displayed in Figure 5A. A significant reduction in contour length is observed if the Hfq to bp ratio is increased. In the case of one Hfq per 300 bps, the contour length agrees with the regular value of  $350 \pm 20$  nm for double-stranded DNA in the B-form. However, with one Hfq per six bps, the contour length is decreased and takes a value of  $270 \pm 20$  nm.

The centerlines were also used to obtain the tangent vector correlation function  $\langle \cos \theta_{s,s+L} \rangle$ , where  $\theta$  is the angle between tangent vectors at points  $s$  and  $s+L$ , by averaging  $s$  along the contour (43). For DNA molecules equilibrated in 2D conformation, this correlation follows:

$$\langle \cos \theta_{s,s+L} \rangle = \exp[-L/(2P)] \quad (1)$$

The inverse of the exponential decay constant gives hence the persistence length  $P$ . Experimental and fitted tangent correlation functions are shown in Figure 5B. With increasing Hfq to bp ratio from 1:300 to 1:6,  $P$  increases from  $49.1 \pm 0.3$  to  $60.4 \pm 0.3$  nm. The value obtained for Hfq to bp ratio of 1:300 agrees with the commonly accepted value of 50 nm for bare DNA (43). The moderate increase in persistence length for coated DNA shows that the flexibility of the nucleoprotein complex is largely preserved. The images clearly show that the DNA molecules are coated with Hfq. The averaged heights of the complexes are  $0.9 \pm 0.1$  and  $2.6 \pm 0.4$  nm for Hfq to bp ratio of 1:300 and 1:6, respectively. These values are however indicative, because the complexes are dried and spread on the silica surface.



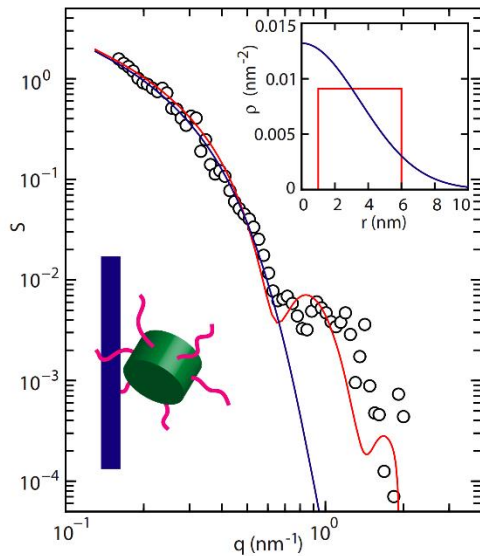
**Figure 5.** (A) Distribution in contour length of DNA (1000 bp) molecules with Hfq to bp ratio of 1:6 (red) and 1:300 (blue). (B) Orientation correlation function of the tangent vectors at a pair of points separated by distance  $L$  along the contour. The closed symbols are the experimental data obtained by averaging 31–37 molecules. The solid lines represent exponential fits.

### Cross-sectional profile

For more detailed structural information about the nucleoprotein complex, low resolution SANS experiments were done on a system of 150-bp DNA fragments and Hfq in 100 mM KCl (Hfq to bp ratio of 1:7.5). The contribution to the scattering from the protein (protein structure factor) is shown in Figure 6 as a function of momentum transfer  $q$ . Momentum transfer  $q$  is defined by the wavelength  $\lambda$  of the radiation and scattering angle  $\theta$  between the incident and scattered beam according to  $q = 4\pi/\lambda \sin(\theta/2)$ .

The protein structure factor represents the spatial Fourier transform of the amino acid density correlation function (49). Information on the structure of the nucleoprotein complex can be obtained by comparison of a coarse grained model calculation and the experimental, low resolution data. For this purpose, we consider the Hfq hexamer as the elementary scattering unit. Furthermore, we recognize that a fraction of Hfq is bound to DNA ( $N_p$  proteins per DNA molecule) and forms a cylindrical complex with a length  $L$  and radial distribution in amino acid density  $\rho(r)$ . The distribution of Hfq along the DNA molecule is assumed to be homogeneous. Any possible ordering of Hfq in register with the phosphate moieties along the DNA molecule is beyond





**Figure 6.** Protein structure factor  $S$  versus momentum transfer  $q$ . The blue and red curves represent the model calculations for a cylindrical complex with a Gaussian and shell-like radial distribution in amino acid density, respectively. The corresponding distributions are shown in the inset. The distribution in blue is Gaussian and provides a cross-sectional radius of gyration 5 nm. The distribution in red represents a shell-like protein coat with an inner and outer radius of 1 and 6 nm, respectively. The structure factor is normalized to  $N_p$  for  $q \rightarrow 0$ . The cartoon illustrates a dangling Hfq hexamer bound on DNA (not to scale).

detection in scattering from solution with isotropic orientation averaging of the complex. For a cylindrical nucleoprotein complex, the protein structure factor takes the form

$$S(q) = \frac{\pi N_p}{qL} a^2(q) \quad (qL \gg 1) \quad (2)$$

with  $a(q)$  the Hankel transformation of the radial amino acid density profile, that is

$$a(q) = \int_0^\infty 2\pi r J_0(qr) \rho(r) dr \quad (3)$$

and  $J_0$  denotes the zero order Bessel function of the first kind. Notice that the structure factor is normalized to  $N_p/L$ , that is the inverse interspersion length of bound protein along the contour. Free protein does not significantly contribute to the scattering at small angles, because of lack of long range order.

For the amino acid distribution in the radial direction away from the spine-axis of the nucleoprotein complex, we have used two different models. In the first model, we assume a Gaussian profile  $\rho(r) = \exp(-r^2/r_p^2)/(\pi r_p^2)$  with cross-sectional radius of gyration  $r_p$ . The corresponding model calculation with optimized  $r_p = 5.0 \pm 0.5$  nm is shown in Figure 6 (the radial profile is shown in the inset). The use of a Gaussian profile results in a reasonable fit of the protein structure factor in the lower range of momentum transfer, but fails to predict the characteristic shoulder observed at higher values of  $q$ . Notice that the Gaussian model ignores the depletion of protein density at the core of the complex. In the second model, the protein is assumed to form a coat surrounding DNA with an inner and

outer radius  $r_1$  and  $r_2$ , respectively. The radial amino acid distribution is, hence, constant for  $r_1 < r < r_2$  and given by  $\rho(r) \pi (r_2^2 - r_1^2) = 1$ . The shell-like radial profile and corresponding model calculation are also shown in Figure 6. The cut-off in radial density at the outer radius of the coat results in a reasonable reproduction of the shoulder. In the fit procedure, the inner radius was set to the radius of the DNA molecule, that is  $r_1 = 1$  nm. The outer radius was optimized and takes the value  $r_2 = 6.0 \pm 0.5$  nm. For both models, the contour distance between bound protein follows from the normalization and reads  $L/N_p = 10 \pm 1$  nm (i.e., 30 bps per Hfq hexamer).

## DISCUSSION

### Stretch of Hfq-incubated DNA

With increasing concentration of Hfq, the stretch of confined DNA along the direction of the channel decreases. This observation differs from what was previously reported for H-NS (42). In the case of H-NS, contraction is only observed at higher ionic strength (33 mM). At moderate ionic strength of 11 mM (T-buffer with 3 mM NaCl), the molecules elongate with increasing concentration of H-NS. The conformational response was interpreted in terms of polymerization of H-NS on DNA and H-NS mediated bridging interaction between different segments of the DNA molecule. Binding of H-NS results in an increase in persistence length to 130 nm for an H-NS dimer to bp ratio of 1:6. Bridging becomes effective only at higher ionic strength, related to increased screening of electrostatic repulsion between like-charged segments of the DNA molecule. In the case of Hfq, no elongation is observed. The absence of elongation is consistent with our atomic force microscopy results. The persistence and contour length were seen to increase and decrease, respectively, by 20% following binding of Hfq on DNA with a hexamer to bp ratio of 1:6. As shown by Monte Carlo computer simulation, these changes in contour and persistence length should result in a decrease of the stretch by  $\sim 10\%$  (42). Such a decrease in stretch agrees with the results obtained at moderate ionic strength irrespective channel cross-sectional diameter. However, at higher ionic strength, the observed decrease by 30–60% exceeds the prediction based on changes in persistence and contour length. Hfq mediated bridging becomes progressively more important and reduces the stretch at higher concentrations of salt. Bridging resulting in back folding, looping and side-by-side aggregation is also evident in the atomic force microscopy images of longer DNAs.

### Condensation of DNA by Hfq

For over-threshold concentrations of Hfq, the DNA molecules compact into a condensed form. Nanochannel facilitated condensation of DNA by neutral crowders, like-charge proteins (haemoglobin and bovine serum albumin) and H-NS has been reported before (38–39,42). The critical concentrations pertaining to Hfq and H-NS are at least two orders of magnitude lower than those for neutral crowders and like-charge proteins. Furthermore, there is a significant dependence on channel diameter, medium ionic strength



and/or the presence of Mg cations. Hfq is much more effective in condensing DNA than H-NS. For instance, for DNA molecules inside  $200 \times 300 \text{ nm}^2$  channels and in T-buffer with 3 mM NaCl, the critical concentrations are  $0.005 \mu\text{M}$  Hfq and  $0.5 \mu\text{M}$  H-NS. These values correspond with an Hfq hexamer and H-NS dimer to bp ratio of 1:1000 and 1:10, respectively. Hfq also condenses DNA in the bulk phase, but this requires an order of magnitude higher concentration. Unconstrained DNA cannot be compacted into a condensed form by H-NS in prevalent environmental conditions (42).

In view of the submicromolar concentrations of Hfq, the observed phenomena cannot be explained by macromolecular crowding. A plausible explanation for condensation is Hfq mediated bridging of segments of the DNA molecule, like multivalent polyamines bind and bridge DNA (58). For a transition to a condensed state, the segments should be (almost) parallel and juxtaposed. The confinement inside a nanochannel imposes orientation order, thereby increasing the attractive interaction resulting in the compaction into the condensed form. On the other hand, due to the linearization of the DNA molecule, the probability for segment juxtaposition is reduced. This reduction in segment juxtaposition results in a shift of the critical concentration for condensation toward higher values for smaller channel diameter. In the narrow channel and minimal screening conditions, the molecules do not condense for an Hfq-to-bp ratio less than 1:40. As in the case of contraction, condensation is facilitated by screening of the electrostatic repulsion between like-charged segments of the DNA molecule through an increase in ionic strength and/or binding of divalent magnesium.

#### Structural arrangement of bound Hfq on DNA

The moderate increase in persistence length by 20% indicates that Hfq does not polymerize on DNA and no rigid filament is formed. Insight in the arrangement of bound Hfq can be obtained from a comparison of the known dimensions of the Hfq hexamer and the structural parameters derived from the scattering experiment. The hexameric core of Hfq can be modeled as a disk-like object with a diameter and a thickness of around 7 and 3 nm, respectively (13). The outer radius of the nucleoprotein complex of 6 nm indicates that the protein is not firmly docked in either the transverse or lateral direction. If the protein binds with the flat face of the core facing the duplex, the expected radial outer dimension would be  $<4 \text{ nm}$ . In a transverse configuration, a larger outer radius exceeding 8 nm is expected. Our structural data do agree, however, with a dangling protein with some degree of freedom, possibly attached to the duplex through the C-terminal extension of the Hfq hexamer. Indeed, it has been suggested that the C-terminal tail could track along the grooves and interact with the phosphate backbone (5). Such a dangling protein would take a lateral distance along the contour of  $\sim 10 \text{ nm}$  and a radial distance away from the helical axis of the DNA molecule of  $\sim 5 \text{ nm}$ . The propensity for bridging can then be understood by binding of the protein, again possibly through one or more of the C-terminal extensions to other segment(s) of the DNA molecule. The 20% reduction in contour length following binding of Hfq

might be related to opening of AT-rich tracks of the double helix (21).

#### CONCLUSIONS

Hfq is limiting in the cell and RNA-related regulations compete with DNA binding. Nevertheless, RNA has been shown to be actively cycled *in vivo*, thus allowing effective regulation (59). Taking into account the lower affinity of Hfq for DNA compared to that of RNA and the high concentration of DNA inside the bacterial cell (bp concentration of  $\sim 20 \text{ mM}$ ), it is most likely that Hfq is also actively cycled on DNA (21,60). Inside bacteria, the estimated concentration of Hfq bound to DNA is a few hundred nanomolar (26). This concentration falls within the range of concentrations in our experiments. We obtained similar results if the chloride salt is substituted by the more physiologically relevant potassium glutamate (61). Furthermore, the cross-sectional diameters of the nanochannels of a few hundred nanometer are comparable to the size of the bacterial nucleoid. Accordingly, we believe that our results have implications for the understanding of the function of Hfq in the regulation of the conformation and compaction of DNA *in vivo*.

Our experiments show that Hfq binds on double stranded DNA, but a rigid filament like in the case of binding of H-NS on DNA is not formed. Rather the nucleoprotein complex retains its flexibility with a moderate increase in bending persistence length. Hfq has a strong propensity for bridging of DNA segments, resulting in compaction and, eventually, a collapse to a condensed state for over-threshold concentrations of Hfq. Interaction between segments of the DNA molecule, controlled by electrostatic screening and segment orientation order imposed by nanoconfinement, plays a pivotal role. These observations indicate that Hfq is involved in the architectural organization of the genome, rather than regulation of gene expression *per se*. Our scattering data show that the protein is not firmly docked, but dangles at a radial distance of  $\sim 5 \text{ nm}$  from the helical axis of the duplex. Furthermore, we observed a decrease in contour length of the nucleoprotein complex with respect to the protein-free state. A plausible structural model is that the protein binds with the C-terminal extension to AT-rich sequences, which might require opening of the duplex with a concomitant decrease in contour length. Once bound, the dangling Hfq hexamer can bind one or more other DNA segments with the other C-terminal extensions and thus form a bridge with multi-arm functionality.

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

#### ACKNOWLEDGEMENTS

We kindly acknowledge the Laboratoire Léon Brillouin for providing the neutron scattering facilities. We thank M. Gesson for her contribution at an early stage.

## FUNDING

FRC [R-144-000-336-112, R-144-000-312-112]; MOE Academic Research Fund [MOE2014-T2-1-001]. Funding for open access charge: Ministry of education.  
*Conflict of interest statement.* None declared.

## REFERENCES

1. Franze de Fernandez, M.T., Hayward, W.S. and August, J.T. (1972) Bacterial proteins required for replication of phage Qb ribonucleic acid. *J. Biol. Chem.*, **247**, 824–821.
2. Tsui, H.C., Leung, H.C. and Winkler, M.E. (1994) Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in *Escherichia coli* K-12. *Mol. Microbiol.*, **13**, 35–49.
3. Storz, G., Vogel, J. and Wassarman, K.M. (2011) Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell*, **43**, 880–891.
4. Gottesman, S., McCullen, C.A., Guillier, M., Vanderpool, C.K., Majdalani, N., Benhammou, J., Thompson, K.M., FitzGerald, P.C., Sowa, N.A. and FitzGerald, D.J. (2006) Small RNA regulators and the bacterial response to stress. *Cold Spring Harb. Symp. Quant. Biol.*, **71**, 1–11.
5. Vogel, J. and Luisi, B.F. (2011) Hfq and its constellation of RNA. *Nat. Rev. Microbiol.*, **9**, 578–589.
6. Wilusz, C.J. and Wilusz, J. (2013) Lsm proteins and Hfq: Life at the 3' end. *RNA Biol.*, **10**, 592–601.
7. Brennan, R.G. and Link, T.M. (2007) Hfq structure, function and ligand binding. *Curr. Opin. Microbiol.*, **10**, 125–133.
8. Mura, C., Randolph, P.S., Patterson, J. and Cozen, A.E. (2013) Archaeal and eukaryotic homologs of Hfq: a structural and evolutionary perspective on Sm function. *RNA Biol.*, **10**, 636–651.
9. Arluison, V., Folichon, M., Marco, S., Derreumaux, P., Pellegrini, O., Seguin, J., Hajnsdorf, E. and Regnier, P. (2004) The C-terminal domain of *Escherichia coli* Hfq increases the stability of the hexamer. *Eur. J. Biochem.*, **271**, 1258–1265.
10. Link, T.M., Valentin-Hansen, P. and Brennan, R.G. (2009) Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. *Proc. Natl. Acad. Sci. USA*, **106**, 19292–19297.
11. Schumacher, M.A., Pearson, R.F., Moller, T., Valentin-Hansen, P. and Brennan, R.G. (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq–RNA complex: a bacterial Sm-like protein. *EMBO J.*, **21**, 3546–3556.
12. Sauer, E., Schmidt, S. and Weichenrieder, O. (2012) Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 9396–9401.
13. Beich-Fransden, M., Vecerek, B., Konarev, P.V., Sjöblom, B., Kloiber, K., Hämmerle, H., Rajkowsch, L., Miles, A.J., Kontaxis, G., Wallace, B. et al. (2011) Structural insights into the dynamics and function of the C-terminus of the *E. coli* RNA chaperone Hfq. *Nucleic Acids Res.*, **39**, 4900–4915.
14. Olsen, A.S., Moller-Jensen, J., Brennan, R.G. and Valentin-Hansen, P. (2010) C-terminally truncated derivatives of *Escherichia coli* Hfq are proficient in riboregulation. *J. Mol. Biol.*, **404**, 173–182.
15. Salim, N.N., Faner, M.A., Philip, J.A. and Feig, A.L. (2012) Requirement of upstream Hfq-binding (ARN)<sub>x</sub> elements in *glmS* and the Hfq C-terminal region for *GlmS* upregulation by sRNAs *GlmZ* and *GlmY*. *Nucleic Acids Res.*, **40**, 8021–8032.
16. Vecerek, B., Rajkowsch, L., Sonnleitner, E., Schroeder, R. and Blasi, U. (2008) The C-terminal domain of *Escherichia coli* Hfq is required for regulation. *Nucleic Acids Res.*, **36**, 133–143.
17. Vincent, H.A., Henderson, C.A., Stone, C.M., Cary, P.D., Gowers, D.M., Sobott, F., Taylor, J.E. and Callaghan, A.J. (2012) The low-resolution solution structure of *Vibrio cholerae* Hfq in complex with *Qrr1* sRNA. *Nucleic Acids Res.*, **40**, 8698–8710.
18. Dimastrogiovanni, D., Fröhlich, K.S., Bandyra, K.J., Bruce, H.A., Hohensee, S., Vogel, J. and Luisi, B.F. (2014) Recognition of the small regulatory RNA *RydC* by the bacterial Hfq protein. *Elife*, **3**, e05375.
19. Robinson, K.E., Orans, J., Kovach, A.R., Link, T.M. and Brennan, R.G. (2014) Mapping Hfq–RNA interaction surfaces using tryptophan fluorescence quenching. *Nucleic Acids Res.*, **42**, 2736–2749.
20. Arluison, V., Mura, C., Guzmán, M.R., Liqueur, J., Pellegrini, O., Gingery, M., Régner, P. and Marco, S. (2006) Three-dimensional structures of fibrillar Sm proteins: Hfq and other Sm-like proteins. *J. Mol. Biol.*, **356**, 86–96.
21. Geinguenaud, F., Calandrini, V., Teixeira, J., Mayer, C., Liqueur, J., Lavelle, C. and Arluison, V. (2011) Conformational transition of DNA bound to Hfq probed by infrared spectroscopy. *Phys. Chem. Chem. Phys.*, **13**, 1222–1229.
22. Updegrove, T.B., Correia, J.J., Galletto, R., Bujalowski, W. and Wartell, R.M. (2010) *E. coli* DNA associated with isolated Hfq interacts with Hfq's distal surface and C-terminal domain. *Biochim. Biophys. Acta*, **1799**, 588–596.
23. Takada, A., Wachi, M., Kaidow, A., Takamura, M. and Nagai, K. (1997) DNA binding properties of the hfq gene product of *Escherichia coli*. *Biochem. Biophys. Res. Com.*, **236**, 576–579.
24. Ohniwa, R.L., Muchaku, H., Saito, S., Wada, C. and Morikawa, K. (2013) Atomic force microscopy analysis of the role of major DNA-binding proteins in organization of the nucleoid in *Escherichia coli*. *PLoS One*, **8**, e72954.
25. Azam, T.A., Hiraga, S. and Ishihama, A. (2000) Two types of localization of the DNA-binding proteins within the *Escherichia coli* nucleoid. *Genes Cells*, **5**, 613–626.
26. Diestra, E., Cayrol, B., Arluison, V. and Risco, C. (2009) Cellular electron microscopy imaging reveals the localization of the Hfq protein close to the bacterial membrane. *PLoS One*, **4**, e8301.
27. Azam, T.A. and Ishihama, A. (1999) Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. *J. Biol. Chem.*, **274**, 33105–33113.
28. Ishihama, A. (1999) Modulation of the nucleoid, the transcription apparatus, and the translation machinery in bacteria for stationary phase survival. *Genes Cells*, **4**, 135–143.
29. Sukhodolets, M.V. and Garges, S. (2003) Interaction of *Escherichia coli* RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. *Biochemistry*, **42**, 8022–8034.
30. Rabhi, M., Espeli, O., Schwartz, A., Cayrol, B., Rahmouni, A.R., Arluison, V. and Boudvillain, M. (2011) The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *EMBO J.*, **30**, 2805–2816.
31. Le Derout, J., Boni, I.V., Regnier, P. and Hajnsdorf, E. (2010) Hfq affects mRNA levels independently of degradation. *BMC Mol. Biol.*, **11**, 17.
32. Cech, G.M., Pakula, B., Kamrowska, D., Wegrzyn, G., Arluison, V. and Szalewska-Palasz, A. (2014) Hfq protein deficiency in *Escherichia coli* affects ColE1-like but not lambda plasmid DNA replication. *Plasmid*, **73**, 10–15.
33. Ross, J.A., Wardle, S.J. and Haniford, D.B. (2010) Tn10/IS10 transposition is downregulated at the level of transposase expression by the RNA-binding protein Hfq. *Mol. Microbiol.*, **78**, 607–621.
34. Ross, J.A., Trussler, R.S., Black, M.D., McLellan, C.R. and Haniford, D.B. (2014) Tn5 transposition in *Escherichia coli* is repressed by Hfq and activated by over-expression of the small non-coding RNA *SgrS*. *Mobile DNA*, **5**, 27.
35. Levy, S.L. and Craighead, H.G. (2010) DNA manipulation, sorting, and mapping in nanofluidic systems. *Chem. Soc. Rev.*, **3**, 1133–1152.
36. Reisner, W., Pedersen, J.N. and Austin, R.H. (2012) DNA confinement in nanochannels: physics and biological applications. *Rep. Prog. Phys.*, **75**, 106601.
37. van der Maarel, J.R.C., Zhang, C. and van Kan, J.A. (2014) A nanochannel platform for single DNA studies: from crowding, protein DNA interaction, to sequencing of genomic information. *Isr. J. Chem.*, **54**, 1573–1588.
38. Zhang, C., Shao, P.G., van Kan, J.A. and van der Maarel, J.R.C. (2009) Macromolecular crowding induced elongation and compaction of single DNA molecules confined in a nanochannel. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 16651–16656.
39. Zhang, C., Gong, Z.Y., Guttula, D., Malar, P.P., van Kan, J.A., Doyle, P.S. and van der Maarel, J.R.C. (2012) Nanofluidic compaction of DNA by like-charged protein. *J. Phys. Chem. B*, **116**, 3031–3036.
40. Jones, J.J., van der Maarel, J.R.C. and Doyle, P.S. (2011) Effect of nanochannel geometry on DNA structure in the presence of macromolecular crowding agent. *Nano Lett.*, **11**, 5047–5053.
41. Zhang, C., Jiang, K., Liu, F., Doyle, P.S., van Kan, J.A. and van der Maarel, J.R.C. (2013) A nanofluidic device for single molecule studies with in situ control of environmental solution conditions. *Lab. Chip*, **13**, 2821–2826.



42. Zhang,C., Guttula,D., Liu,F., Malar,P.P., Ng,S., Dai,L., Doyle,P.S., van Kan,J.A. and van der Maarel,J.R.C. (2013) Effect of H-NS on the elongation and compaction of single DNA molecules in a nanospace. *Soft Matter*, **9**, 9593–9601.
43. Wiggins,P.A., van der Heijden,T., Herrero,F.M., Spakowitz,A., Phillips,R., Widom,J., Dekker,C. and Nelson,P.C. (2006) High flexibility of DNA on short length scales probed by atomic force microscopy. *Nat. Nanotechnol.*, **1**, 137–141.
44. Chang,S.L., Chen,S.H., Rill,R.L. and Lin,J.S. (1990) Measurements of monovalent and divalent counterion distributions around persistence length DNA fragments in solution. *J. Phys. Chem.*, **94**, 8025–8028.
45. van der Maarel,J.R.C., Groot,L.C.A., Mandel,M., Jesse,W., Jannink,G. and Rodriguez,V. (1992) Partial and charge structure functions of monodisperse DNA fragments in salt free aqueous solution. *J. Phys. II France*, **2**, 109–122.
46. Zakharova,S.S., Egelhaaf,S.U., Bhuiyan,L.B., Outhwaite,C.W., Bratko,D. and van der Maarel,J.R.C. (1999) Multivalent ion–DNA interaction: neutron scattering estimates of polyamine distribution. *J. Chem. Phys.*, **111**, 10706–10716.
47. Andresen,K., Das,R., Park,H.Y., Smith,H., Kwok,L.W., Lamb,J.S., Kirkland,E.J., Herschlag,D., Finkelstein,K.D. and Pollack,L. (2004) Spatial distribution of competing ions around DNA in solution. *Phys. Rev. Lett.*, **93**, 248103.
48. Dai,L., Mu,Y., Nordenskiöld,L., Lapp,A. and van der Maarel,J.R.C. (2007) Charge structure and counterion distribution in hexagonal DNA liquid crystal. *Biophys. J.*, **92**, 947–958.
49. Jacrot,B. (1976) The study of biological structures by neutron scattering from solution. *Rep. Prog. Phys.*, **39**, 911–953.
50. Taghbalout,A., Yang,Q. and Arluison,V. (2014) The Escherichia coli RNA processing and degradation machinery is compartmentalized within an organized cellular network. *Biochem J.*, **458**, 11–22.
51. Zhang,C., Zhang,F., van Kan,J.A. and van der Maarel,J.R.C. (2008) Effects of electrostatic screening on the conformation of single DNA molecules confined in a nanochannel. *J. Chem. Phys.*, **128**, 225109.
52. van Kan,J.A., Zhang,C., Malar,P.P. and van der Maarel,J.R.C. (2012) High throughput fabrication of disposable nanofluidic lab-on-chip devices for single molecule studies. *Biomicrofluidics*, **6**, 36502.
53. van Kan,J.A., Bettiol,A.A. and Watt,F. (2006) Proton beam writing of three-dimensional nanostructures in hydrogen silsesquioxane. *Nano Lett.*, **6**, 579–582.
54. van Kan,J.A., Shao,P.G., Wang,Y. and Malar,P.P. (2011) Proton beam writing a platform technology for high quality three-dimensional metal mold fabrication for nanofluidic applications. *Microsyst. Technol.*, **17**, 1519–1527.
55. Wang,L.X., Ferrari,M. and Bloomfield,V.A. (1990) Large-scale preparation of mononucleosomal DNA from calf thymus for biophysical studies. *Biotechniques*, **9**, 24–27.
56. Tegenfeldt,J.O., Prinz,C., Cao,H., Chou,S., Reisner,W.W., Riehn,R. and Austin,R.H. (2004) The dynamics of genomic-length DNA molecules in 100-nm channels. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 10979–10983.
57. Johansen,F. and Jacobsen,J.P. (1998) <sup>1</sup>H NMR studies of bis-intercalation of a homodimeric oxazole yellow dye in DNA oligonucleotides. *J. Biomol. Struct. Dyn.*, **16**, 205–222.
58. Dai,L., Mu,Y., Nordenskiöld,L. and van der Maarel,J.R.C. (2008) Molecular dynamics simulation of multivalent ion mediated attraction between DNA molecules. *Phys. Rev. Lett.*, **100**, 118301.
59. Wagner,E.G. (2013) Cycling of RNAs on Hfq. *RNA Biol.*, **10**, 619–626.
60. Elowitz,M.B., Surette,M.G., Wolf,P.E., Stock,J.B. and Leibler,S. (1999) Protein mobility in the cytoplasm of Escherichia coli. *J. Bacteriol.*, **181**, 197–203.
61. Leirmo,S., Harrison,C., Cayley,D.S., Burgess,R.R. and Record,M.T. Jr (1987) Replacement of potassium chloride by potassium glutamate dramatically enhances protein-DNA interactions in vitro. *Biochemistry*, **26**, 2095–2101.

Annexe E :

*Compaction and condensation of  
DNA mediated by the C-terminal  
domain of Hfq*

# Compaction and condensation of DNA mediated by the C-terminal domain of Hfq

Antoine Malabirade<sup>1,†</sup>, Kai Jiang<sup>2,†</sup>, Krzysztof Kubiak<sup>1,3</sup>, Alvaro Diaz-Mendoza<sup>4</sup>, Fan Liu<sup>2</sup>, Jeroen A. van Kan<sup>2</sup>, Jean-François Berret<sup>4</sup>, Véronique Arluison<sup>1,4,\*</sup> and Johan R.C. van der Maarel<sup>2,\*</sup>

<sup>1</sup>Laboratoire Léon Brillouin, CEA, CNRS, Université Paris Saclay, 91191 Gif-sur-Yvette, France, <sup>2</sup>Department of Physics, National University of Singapore, 2 Science Drive 3, 117542, Singapore, <sup>3</sup>Department of Molecular Biology, University of Gdansk, Wita Stwosza 59, 80-308 Gdansk, Poland and <sup>4</sup>Université Paris Diderot, Sorbonne Paris Cité, 75013 Paris, France

Received March 18, 2017; Revised April 17, 2017; Editorial Decision May 02, 2017; Accepted May 03, 2017

## ABSTRACT

Hfq is a bacterial protein that is involved in several aspects of nucleic acids metabolism. It has been described as one of the nucleoid associated proteins shaping the bacterial chromosome, although it is better known to influence translation and turnover of cellular RNAs. Here, we explore the role of *Escherichia coli* Hfq's C-terminal domain in the compaction of double stranded DNA. Various experimental methodologies, including fluorescence microscopy imaging of single DNA molecules confined inside nanofluidic channels, atomic force microscopy, isothermal titration microcalorimetry and electrophoretic mobility assays have been used to follow the assembly of the C-terminal and N-terminal regions of Hfq on DNA. Results highlight the role of Hfq's C-terminal arms in DNA binding, change in mechanical properties of the double helix and compaction of DNA into a condensed form. The propensity for bridging and compaction of DNA by the C-terminal domain might be related to aggregation of bound protein and may have implications for protein binding related gene regulation.

## INTRODUCTION

Hfq is a phylogenetically conserved protein found at high concentration in approximately half of the sequenced bacterial genomes (1). Consistent with this abundance, the protein coordinates multiple roles in bacterial metabolism. For instance, the first description of Hfq is related to its role in bacteriophage Q $\beta$  RNA replication [Hfq stands for host factor for phage Q $\beta$ ] (2). The pleiotropic functions of Hfq

were highlighted when the *hfq* gene was disrupted in *Escherichia coli*, resulting in various phenotypes (3). Most of these *hfq*-null phenotypes are due to the role of Hfq in RNA regulation. Indeed, even if its requirement seems to be facultative in some bacteria, it is usually needed to mediate post-transcriptional stress-response with small noncoding RNA (sRNA). This regulatory mechanism is based on the hybridization of sRNA to its target mRNA. Accordingly, the majority of known sRNA interacts nearby the ribosome binding site to prevent initiation of translation and thus favors mRNA decay (4–6). Few studies have shed light on the importance of Hfq in DNA metabolism. It has been suggested that Hfq affects plasmid replication and DNA transposition (7,8). Hfq might also be involved in transcription control, such as regulation of transcription initiation and/or elongation (9,10). Hfq might also regulate transcription by protein interaction, mainly with RNA polymerase and Rho transcription termination factor (11,12).

Hfq has been described as one of the nucleoid associated proteins (NAPs) shaping the bacterial chromosome (13). The amount of Hfq in the bacterial cell is comparable to those of the most abundant NAPs, that is Fis (factor for inversion simulation) and HU (heat-unstable protein) (14). Its quantity in the nucleoid might however be less than the ones for HU and Fis, as only 10–20% has been estimated to be chromosome bound (15,16). The cellular concentration of Hfq increases when reaching the stationary phase, but the nucleoid-bound fraction remains approximately constant (7,16–18). Hfq has a significant affinity for DNA with an equilibrium dissociation constant  $K_d$  in the range from nano- to micromolar (the values of  $K_d$  pertaining to binding to RNA are three orders of magnitude smaller) (10,19). Hfq not only interacts with DNA but also with other NAPs such as H-NS (20,21). Hfq belongs to a family of NAPs that form bridges between sections of the

\*To whom correspondence should be addressed. Tel: +65 6516 4396; Fax: +65 6777 6126; Email: johanmaarel@gmail.com  
Correspondence may also be addressed to Véronique Arluison. Tel: +33 169 083 282; Fax: +33 169 088 261; Email: veronique.arluison@cea.fr  
†These authors contributed equally to the paper as first authors.



DNA molecule, thereby organizing large parts of the chromosome into domains (10,22). It has also been suggested that Hfq constrains negative supercoiling *in vivo* (3). However, the mechanism how Hfq affects DNA topology is unclear.

Structurally, the 102 amino acids Hfq from *Escherichia coli* forms an Sm-fold in its N-terminal region (NTR, 65 amino acids) (23). The NTR consists of a strongly bent, five stranded antiparallel  $\beta$ -sheet capped by an  $\alpha$ -helix. The  $\beta$ -sheets from six monomers assemble into a toroidal structure with two faces, that is a proximal (where the  $\alpha$ -helix is located) and a positively charged, distal face (23). The distal face, the edge and the C-terminal region (CTR, 35 amino acids) are involved in DNA binding. This information is based on experiments with Hfq mutants with deleted CTR or a variety of point mutations such as K31A, Y25A (distal face) or R16A (edge) (24). While the CTR is non-essential for RNA binding, the distal face and edge are involved in both DNA and RNA fixation (25). *In vitro*, the protein shows a preference for A-tracts ( $K_d = 200$  nM) (10). A-tracts are usually associated with an intrinsic curvature of the DNA molecule (13). *In vivo*, Hfq has a preferred (A/T)T(A/G)TGCCG binding motif with a curved topology (24). Besides these preferred binding sequences, Hfq also interacts with DNA in a sequence-nonspecific manner (13,26). Although the CTR region is usually considered to be intrinsically disordered, it was recently reported that it contains an amyloid sequence that allows the protein to self-assemble (4,27).

In previous work, we have focused on DNA-compacting properties of wild-type Hfq (28). Various experimental technologies, including fluorescence microscopy imaging of single DNA molecules confined inside nanofluidic channels, atomic force microscopy and small angle neutron scattering were used to follow the assembly of Hfq on DNA. It was shown that Hfq forms a nucleoprotein complex, changes the mechanical properties of the double helix and compacts DNA into a condensed form. A compaction mechanism based on protein-mediated bridging of DNA segments was proposed. The propensity for bridging was hypothesized to be related to the multi-arm functionality of the Hfq hexamer, resulting from binding of the C-terminal regions to the duplex. These previous experiments were done with full length Hfq. Here, we report similar molecular imaging experiments and electrophoretic mobility assays (EMSA) of DNA in the presence of a 38 amino acids peptide representing the C-terminal region (Hfq-CTR) as well as a truncated version of the protein composed of the 72 first amino acids, representing the N-terminal region (Hfq-NTR). Note that Hfq-NTR forms the hexameric torus, whereas Hfq-CTR represents the protruding C-terminal arms of wild type Hfq. The thermodynamics of Hfq-CTR's association with DNA is investigated using isothermal titration calorimetry (ITC). Our new results enable to firmly establish the role of the C-terminal domain in Hfq-mediated bridging and compaction of double stranded DNA.

## MATERIALS AND METHODS

### Sample preparation

His-tagged *Escherichia coli* Hfq-NTR (residues 1-72) was purified from over expressing BL21-(DE3) $\Delta$ hfq/pLATE11-hfqntr cells. The pLATE11-hfqntr expression vector was constructed by the ligation independent cloning method (Thermo Fisher). The oligonucleotides used for cloning were AGAAGGAGATATAACTATGGCTAAGGGGCAATCTTTACAAGATC (for) and GGAGATGGGAAGTCATTAAGTGTGATGAGAAACCGGGC (rev). Insertion of the His-tag was performed by site-directed mutagenesis (Q5, New England Biolabs). As the His-tag potentially modifies the properties of Hfq, a Tobacco Etch Virus (TEV) protease site (ENLYFQG) was inserted between the N-terminal His-tag and Hfq sequence. The oligonucleotides used for this double insertion were GAGAACCCTGTACTTCCAGGGCGCTAAGGGGCAATCTTTAC (for) and ACCATGATGATGATGATGATGATGCATAGTTATATCTCCTTCTGG (rev). Cells from post-induction cultures were resuspended in 20 mM Tris-HCl, pH 7.5, 0.5 M NaCl, 10% (v/v) glycerol and a protease inhibitor (Sigma) at 277 K. The suspension was sonicated and the lysed cells were cleared by centrifugation at 15 000 g for 30 min. DNase I (40 g/L) and RNase A (30 g/L) were added to the cleared lysate at 303 K. The solution was then applied to a Ni<sup>2+</sup>-NTA column (GE Healthcare). The resin was washed with 20 mM Tris-HCl, pH 7.8, 0.3 M NaCl, 20 mM imidazole and the protein was eluted with a gradient of imidazole (20–500 mM). TEV digestion was carried out according to the manufacturer's instructions (Thermo Fisher). After digestion, Hfq-NTR was purified with a Hi-Trap Q HP 1 mL column (GE Healthcare). Before injection on the anion exchange column, Hfq-NTR was diluted five times in equilibration buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 7) to reduce salt concentration. The resin was equilibrated with the same buffer and the protein was eluted with a gradient of NaCl (0–1 M). Digested His-tag, non-digested Hfq-NTR and contaminants were removed with the anion exchange column (27). Hfq-CTR (residues 64–102) was obtained from Genosphere Biotech (France). Full length Hfq has been prepared and purified as described before (28).

T4 GT7 DNA (T4-DNA, 165.65 kbp) was obtained from Nippon Gene, Tokyo and used without further purification. No DNA fragments of ones to tens of kbps were observed by pulsed gel electrophoresis. Samples were prepared by dialyzing solutions of DNA against 10 mM Tris-HCl (T) buffer and T-buffer with 30 mM potassium glutamate (KGlut). Solutions of Hfq-CTR and Hfq-NTR in the same buffer were also prepared. The Tris-HCl concentration is 10 mM Tris adjusted with HCl to pH 7.5 (that is 8.1 mM Tris-Cl and 1.9 mM Tris). The ionic strength of the buffer was calculated with the Davies equation for estimating the activity coefficients of the ions and a dissociation constant  $pK = 8.08$  for Tris. Solutions of Hfq-CTR or Hfq-NTR and DNA were mixed and incubated for 12 h at 277 K. YOYO-1 fluorescence staining dye was purchased from Invitrogen, Carlsbad, CA, USA. DNA was stained with YOYO-1 with an incubation time of 12 h and an intercalation ratio of 100 bps per dye. No anti-photo bleaching agent was used.

### Electrophoretic mobility shift assay (EMSA)

The binding of Hfq, Hfq-CTR and Hfq-NTR to DNA was investigated through the measurement of the equilibrium dissociation constant ( $K_d$ ) with a gel shift assay. A 1000 bps DNA fragment was obtained from Thermo Scientific. The fragments (fragment concentration of 15 nM) were incubated with an excess of Hfq, Hfq-CTR, and Hfq-NTR, respectively, at room temperature for 20 min in T-buffer with 50 mM NaCl, pH 7.5. Band shifts were resolved on non-denaturing gradient 4–12% polyacrylamide gel. The native gel was run for 2 h at room temperature with 40 mM Tris-Acetate, 1 mM ethylenediaminetetraacetic acid, pH 8.0 (TAE) buffer, stained with GelRed (Biotium), and imaged with a G:BOX system (Syngene, Cambridge, UK). Quantification of bands was achieved by IMAGEJ software (<http://rsb.info.nih.gov/ij/>). Depending on the protein, either a cooperative (sigmoidal) or non-cooperative (hyperbola) binding model was applied. Affinities have been estimated through the measurement of the equilibrium dissociation constant  $K_d$  in a non-cooperative model (expressed in mol/L). In case of the cooperative model, the apparent dissociation constant  $K_d^{app}$  is reported together with the Hill coefficient  $n_H$  as an indication of the extent of the cooperativity.

### Chip fabrication

The channel systems were fabricated by replication in polydimethylsiloxane (PDMS) of patterned master stamps (29,30). The nanochannel part of the stamps was made in hydrogen silsesquioxane (HSQ) resist (Dow Corning, Midland, MI, USA) using a lithography process with proton beam writing (31). An array of nanochannels is connected to two loading reservoirs through a superposing set of microchannels made in SU-8 resist with UV lithography. The heights and widths of the ridges in the master stamps were measured with atomic force microscopy (Dimension 3000, Veeco, Woodbury, NY, USA) and scanning electron microscopy, respectively. Two stamps were made featuring nanochannels of length 60  $\mu\text{m}$  and rectangular cross-sections of 150  $\times$  250 and 200  $\times$  300 nm<sup>2</sup>, respectively. The stamp was coated with a 5 nm thick teflon layer for perfect release of the replicated PDMS chips (32). The stamps were replicated in PDMS followed by curing with a curing agent (Sylgard, Dow Corning) at 338 K for 24 h. The replicas were sealed with a glass coverslip, after both substrates were plasma oxidized (Harrick, Ossining, NY, USA).

### Nanofluidics

The pre-incubated and stained DNA molecules were loaded into one of the two reservoirs of the biochip. The DNA molecules were subsequently driven into the nanochannels by electrophoresis. For this purpose, two platinum electrodes were immersed in the reservoirs and connected to a power supply with a voltage in the range 0.1–10 V (Keithley, Cleveland, OH, USA). Once the DNA molecules were localized inside the nanochannels, the electric field was switched off and the molecules were allowed to relax to their equilibrium state for at least 60 s. The stained DNA molecules were visualized with a Nikon Eclipse Ti

inverted fluorescence microscope equipped with a 200 W metal halide lamp, a filter set, and a 100 $\times$  oil immersion objective. A UV light shutter controlled the exposure time. Images were collected with an electron multiplying charge coupled device (EMCCD) camera (iXon X3, Andor Technology, Belfast, UK). One minute video clips were collected with a frame rate of five frames per second. The video clips were analyzed with MATLAB (Natick, MA, USA). For each frame (time  $t$ ), the radius of gyration tensor of the imaged molecule was constructed according to  $\mathbf{S}(t) = \sum_{m,n} n(\mathbf{r}_{mn} - \mathbf{r}_{cm})^2 I_{mn} / I_0$ , with  $I_{mn}$  the fluorescence intensity of pixel  $[m, n]$  at position  $\mathbf{r}_{mn}$  in the  $xy$ -plane and  $I_0$  is the total intensity of the frame. For intensity threshold, we have used two times the signal to background noise ratio. The centre of mass  $\mathbf{r}_{cm}$  of the molecule was calculated according to  $\mathbf{r}_{cm} = \sum_{m,n} n \mathbf{r}_{mn} I_{mn} / I_0$ . We determined the principal eigenvalues and vectors of  $\mathbf{S}$  using a singular value decomposition and with the assumption of cylindrical symmetry. We subsequently derived the extension of the molecule along the direction of the channel according to  $L_{\parallel}^2 = 12 \langle S_{\parallel} \rangle$ , where  $S_{\parallel}$  represents the largest eigenvalue and the brackets denote an average over all frames.

### Bulk phase imaging

A droplet of solution was deposited on a microscope slide and sealed with a coverslip separated by a 0.12 mm spacer. The YOYO-1 stained T4-DNA molecules were imaged with the above mentioned microscope. One minute video clips were collected with a frame rate of five frames per second. The video clips were analyzed as described above, but with the largest eigenvalue of the radius of gyration tensor corresponding to the length of the long axis of the molecule.

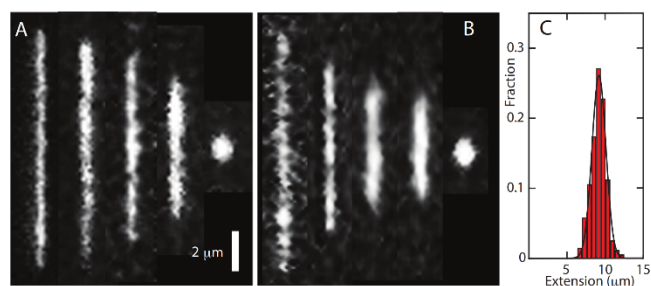
### Atomic force microscopy

DNA fragments were purchased from Thermo Scientific (1000 and 10 000 bps, Waltham, MA, USA). All imaging experiments were done at room temperature in air with a Veeco Dimension 3000 atomic force microscope (Woodbury, NY, USA). Images were acquired in the tapping mode with silicon (Si) cantilevers (spring constant of 20–100 N/m) and operated below their resonance frequency (typically 230–410 kHz). The images were flattened, and the contrast and brightness were adjusted for optimal viewing conditions. A 5  $\mu\text{L}$  droplet was spotted onto a silica surface. After 10 min to allow for DNA adsorption onto the surface, the specimens were developed by flushing them with ultra pure water followed by drying in a stream of N<sub>2</sub> gas.

### Isothermal titration calorimetry

ITC was performed using a Microcal VP-ITC calorimeter (Northampton, MA, USA) with a cell volume of 1.464 mL, working at 298 K and an agitation speed of 307 rpm. The syringe and the measuring cell were filled with degassed solutions of DNA (1000 bps) and Hfq-CTR in T-buffer (pH 7.5), respectively. Typical concentrations in the syringe were between 1 and 4 mmol of DNA bps/L, whereas the measuring cell contains the Hfq-CTR peptide dispersion at molar concentrations of 0.05–0.5 mM. All experiments were done





**Figure 1.** (A) Montage of fluorescence images of T4-DNA molecules inside  $200 \times 300 \text{ nm}^2$  channels and in T-buffer ( $\text{pH } 7.5$ ). From left to right  $1 \times 10^{-4}$ ,  $1 \times 10^{-3}$ ,  $1 \times 10^{-2}$ ,  $5 \times 10^{-2}$  and  $1 \times 10^{-1}$  (condensed)  $\mu\text{M}$  Hfq-CTR. (B) As in panel A, but for T4-DNA inside  $150 \times 250 \text{ nm}^2$  channels and in T-buffer with 30 mM KGlu. (C) Distribution in extension of a population of 30 molecules in T-buffer with  $1 \times 10^{-3} \mu\text{M}$  Hfq-CTR. A Gaussian fit gives a mean extension  $R_{||} = 9 \pm 1 \mu\text{m}$ .

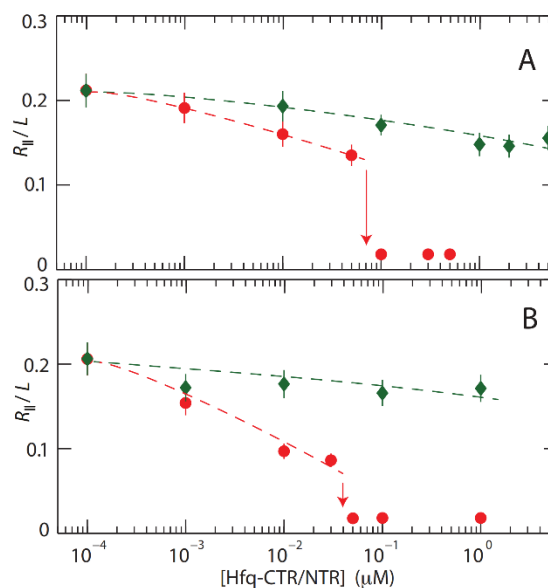
by adding DNA to the peptide. The titration experiment consisted of a preliminary injection of  $2 \mu\text{L}$ , followed by 28 injections of  $10 \mu\text{L}$  at 10 min intervals. A typical ITC experiment consists in measuring the differential power provided by the calorimeter to keep the temperature of the cell constant after each injection. In a second step, the integration over time of the ITC signal was performed and translated in terms of binding enthalpy as a function of the molar ratio  $X = [\text{DNA}]/[\text{Hfq-CTR}]$ . Control experiments corresponding to dilution of DNA with T-buffer were also done and the corresponding integrated signal was subtracted from the raw binding enthalpy.

## RESULTS

### Nanofluidics

Prior to the nanofluidic experiments, T4-DNA molecules with a concentration of 3 mg of DNA/L were incubated with the relevant buffer for at least 24 h. The buffers are 10 mM Tris/HCl (T-buffer) and T-buffer with 30 mM KGlu ( $\text{pH } 7.5$ ). The DNA molecules were subsequently driven into the channels of the nanofluidic device by electrophoresis. We have used two different devices with channel cross-sections of  $200 \times 300$  and  $150 \times 250 \text{ nm}^2$ , respectively. After switching off the electric field, the stretch of the molecules along the longitudinal direction of the channels equilibrates within 60 s. Montages of fluorescence microscopy images showing equilibrated, single T4-DNA molecules in the presence of various concentrations of the C-terminal domain (Hfq-CTR) pertaining to the two channel systems are shown in Figure 1A and B, respectively. With increasing concentration of Hfq-CTR, the stretch decreases and, eventually, the DNA molecule compacts into a condensed form. Condensed DNA is clearly discernible as a bright fluorescence spot. In the case of the N-terminal torus (Hfq-NTR), the stretch shows a moderate decrease with increasing concentration of the protein, but no condensation is observed (montages for Hfq-NTR are not shown).

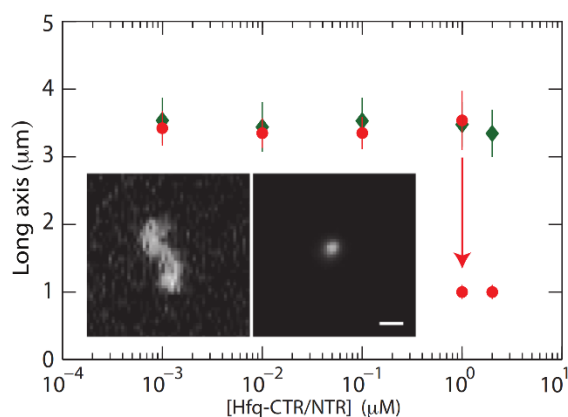
For each channel system and buffer composition, we measured the stretch of about 30 DNA molecules in a fresh PDMS replica. The observed distribution in length is close to Gaussian (an example is shown in Figure 1C). Frag-



**Figure 2.** (A) Relative extension  $R_{||}/L$  of T4-DNA inside  $200 \times 300 \text{ nm}^2$  channels and in T-buffer versus the concentration of Hfq-CTR (red,  $\circ$ ) and Hfq-NTR (green,  $\diamond$ ). (B) As in panel A, but for T4-DNA inside  $150 \times 250 \text{ nm}^2$  channels and in T-buffer with 30 mM KGlu. The dashed curves are drawn as an aid to the eye and the arrows denote the condensation thresholds.

mented DNAs with an extension less than the mean value minus two times the standard deviation were discarded. The mean relative extension  $R_{||}/L$ , that is the mean stretch divided by the contour length of the T4-DNA molecule ( $57 \mu\text{m}$ ), is set out in Figure 2 versus the concentration of the relevant protein. Note that the protein concentrations refer to the C- and N-terminal domains of Hfq in monomeric form. With increasing concentration of Hfq-CTR, the relative extension decreases. For an over-threshold concentration of Hfq-CTR, the DNA molecules compact into a condensed form. A small decrease in stretch and no condensation are observed with increasing concentration of Hfq-NTR. Qualitatively, the same behavior is observed for the two channel systems. The increase in stretch caused by stronger confinement in a channel of smaller cross-sectional diameter offsets the decrease in stretch caused by the increase in concentration of salt.

For over-threshold concentrations of Hfq-CTR, the DNA molecules compact into a condensed form. Notice that the condensation is facilitated by nano-confinement. In the feeding microchannels and reservoirs of the chip, the DNA molecules are in the non-condensed, coiled form. The critical concentration of Hfq-CTR for condensation inside  $150 \times 250 \text{ nm}^2$  channels and in T-buffer with 30 mM KGlu is  $0.07 \pm 0.02 \mu\text{M}$ . For the wider  $200 \times 300 \text{ nm}^2$  channel system and lower ionic strength (T-buffer), this critical concentration decreases somewhat to  $0.04 \pm 0.01 \mu\text{M}$ . With a base pair concentration of  $4.6 \mu\text{M}$ , the DNA molecule hence condenses with an Hfq-CTR to base pair ratio on the order of 1:100. In the case of the N-terminal torus Hfq-NTR, no condensation is observed.



**Figure 3.** Long axis of unconstrained T4-DNA in T-buffer versus the concentration of Hfq-CTR (red, ○) and Hfq-NTR (green, ◇). The dashed curves are drawn as an aid to the eye and the arrow denotes the condensation threshold (for Hfq-CTR only). The inset shows fluorescence images of a non-condensed (left) and condensed (right) T4-DNA molecule.

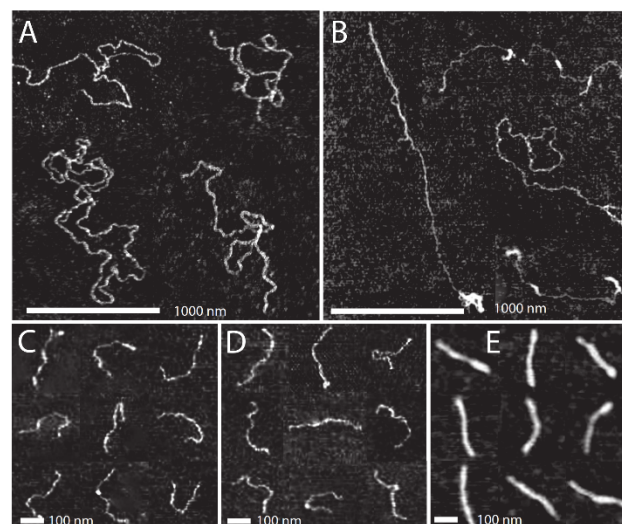
### Condensation in the bulk phase

The size of unconstrained DNA coils in the presence of various concentrations of Hfq-NTR or Hfq-CTR was investigated with fluorescence microscopy. The experiments were done with 0.03 mg of T4-DNA/L in T-buffer (*pH* 7.5). Notice that the concentration of DNA is two orders of magnitude lower than the one pertaining to the loading buffer in the nanofluidics experiments. Furthermore, the DNA molecules were stained with a minimal level of staining of 100 bps per YOYO-1 molecule. Such a low staining level ensures no appreciable effect on the contour length, net charge and bending rigidity of the DNA molecules (29,33). Unconstrained DNA coils are typically slightly anisotropic and their physical extent is best represented by the length of the long axis of the radius of gyration tensor. The measured lengths of the long axis, as well as some characteristic fluorescence images, are shown in Figure 3.

The length of the long axis pertaining to non-condensed molecules is about 3.5 μm and almost constant with increasing concentration of the peptides. For over-threshold concentrations of Hfq-CTR, the DNA molecules compact into a condensed form. A difference is the coexistence of condensed and non-condensed molecules at the critical concentration of Hfq-CTR in the bulk phase, indicating a discontinuous, first order transition. In the channel systems, the transition appears to be continuous (34). The critical concentration of Hfq-CTR for condensation is 1.0 μM. Condensation in the bulk phase hence requires more than an order of magnitude higher concentration of Hfq-CTR. In the case of Hfq-NTR, no condensation in the bulk phase is observed.

### Contour and persistence length

The nanofluidic and bulk phase results indicate interaction of Hfq-CTR with DNA, which eventually results in compaction of the DNA molecule into a condensed form. We have investigated the length of the DNA molecule along its contour as well as its persistence length in the presence

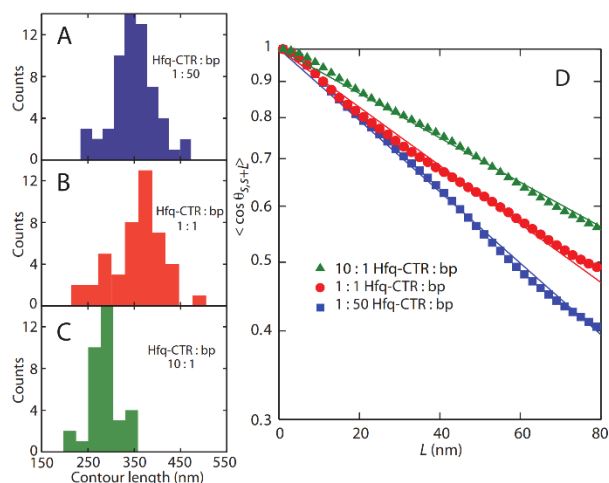


**Figure 4.** (A) Tapping mode atomic force microscopy images of DNA (10 000 bps) with Hfq-CTR to bp ratio of 1:50. (B) As in panel A, but with Hfq-CTR to bp ratio of 1:1. (C) DNA (1000 bps) with Hfq-CTR to bp ratio of 1:50. (D) As in panel C, but with Hfq-CTR to bp ratio of 1:1. (E) As in panel C, but with Hfq-CTR to bp ratio of 10:1. Panels A-E are a montage.

of Hfq-CTR with atomic force microscopy. Linear DNA molecules of 10 000 and 1000 bps were used at a concentration of 3 mg/L. The molecules were incubated with 0.1, 5 and 50 (1000 bps DNA only) μM Hfq-CTR for 12 h, which correspond with an Hfq-CTR to bp ratio of about 1:50, 1:1 and 10:1 respectively. In the presence of Hfq-CTR, no additions to the buffer are necessary to promote adhesion of the DNA molecules to silica. The weakly adsorbed molecules equilibrate on the surface in a 2D conformation. Montages of images for 10 000 bps DNA with increasing Hfq-CTR to bp ratio are shown in Figure 4A and B. The results for 1000 bps DNA are shown in Figure 4C through E. There is no adhesion of DNA to the silica surface without peptide or protein and/or in the presence of Hfq-NTR.

For 10 000 bps DNA with an Hfq-CTR to bp ratio of 1:50 (Figure 4A) the molecules are spread on the surface without evident side-by-side aggregation. However, for a larger Hfq-CTR to bp ratio of 1:1, intramolecular back folding and looping is observed (Figure 4B). The 1000 bps molecules are visible as semi-flexible rods and do not exhibit aggregation nor looping. In the presence of Hfq-CTR there is no obvious change in the averaged heights of the molecules ( $0.9 \pm 0.1$  nm). The measured values of the height are however indicative, because the complexes are dried and spread on the silica surface. To obtain the contour length, the centreline of the molecules was traced. The distributions in length are displayed in Figure 5A-C and the averaged values and standard deviations are collected in Table 1. For an Hfq-CTR to bp ratio of 1:50 and 1:1 the average value of the contour length agrees with the contour length of 1000 bps DNA in the B-form (340 nm). For a relatively large Hfq-CTR to bp ratio of 10:1, the contour length takes a value of  $285 \pm 20$  nm. The latter value is in good agreement with the contour length pertaining to the A-form (290 nm).





**Figure 5.** (A–C) Distribution in contour length of DNA (1000 bps) molecules with Hfq–CTR to bp ratio of 1:50 (blue), 1:1 (red) and 10:1 (green). (D) Orientation correlation function of the tangent vectors at a pair of points separated by distance  $L$  along the contour. The closed symbols are the experimental data obtained by averaging 34–49 molecules. The solid lines are exponential fits.

**Table 1.** Contour length  $L$  and persistence length  $P$  of 1000 bps DNA for various Hfq–CTR to base pair ratios as determined by AFM

Hfq–CTR : bp	$L$ (nm)	$P$ (nm)
1:50	$340 \pm 20$	$43.1 \pm 0.3$
1:1	$350 \pm 20$	$52.7 \pm 0.4$
10:1	$285 \pm 20$	$69.2 \pm 0.6$

The centrelines were also used to obtain the tangent vector correlation function  $\langle \cos \theta_{s,s+L} \rangle$ , where  $\theta$  is the angle between tangent vectors at points  $s$  and  $s + L$ , by averaging  $s$  along the contour (35). For DNA molecules equilibrated in 2D conformation, the correlation functions takes the form  $\langle \cos \theta_{s,s+L} \rangle = \exp[-L/(2P)]$ . The inverse of the exponential decay constant gives hence the persistence length  $P$ . Experimental and fitted tangent correlation functions are shown in Figure 5D. The fitted values of  $P$  are also collected in Table 1. With increasing Hfq–CTR to bp ratio from 1:50 to 10:1, a moderate increase in the values of  $P$  is observed. Atomic force microscopy of naked DNA adsorbed on mica mediated by  $Mg^{2+}$  has previously given  $P = 56 \pm 4$  nm (36). Accordingly, DNA's flexibility is moderately affected by the interaction with Hfq–CTR.

### DNA binding properties of Hfq protein and domains

The interactions of Hfq, Hfq–CTR and Hfq–NTR with DNA have been explored with EMSA. The gels and band shifts are shown in Figure 6. The concentrations are expressed in moles of monomeric peptide/L. For full length Hfq, a two-step band shift is observed. At lower concentrations of Hfq, the band shifts midway the well and the band pertaining to protein-free DNA. This shift corresponds with a binding mode of higher affinity (apparent dissociation constant  $K_d^{ap} = 1.5 \pm 0.3 \mu\text{M}$ ). When the concentration is increased beyond  $1 \mu\text{M}$ , the intermediate band disappears and is shifted to a migration distance close to

the loading wells. This second band shift agrees with a cooperative binding mode of lower affinity (apparent  $K_d^{ap} = 6.5 \pm 1.0 \mu\text{M}$ ) and a Hill coefficient  $n_h$  of  $3.9 \pm 0.3$ . In the case of Hfq–NTR, a single diffuse band shift is observed but no significant amount of material shifted close to the wells. Several gels were analyzed and either no or small cooperativity was observed (Hill coefficient  $n_h = 1-2$ , apparent  $K_d^{ap}$  around  $20 \mu\text{M}$ ). For Hfq–CTR, a single band shifted to a migration distance close to the wells is observed for relatively high concentrations of the peptide. The latter band shift agrees with a binding mode of relatively low affinity ( $K_d = 240 \pm 30 \mu\text{M}$ ) without cooperativity.

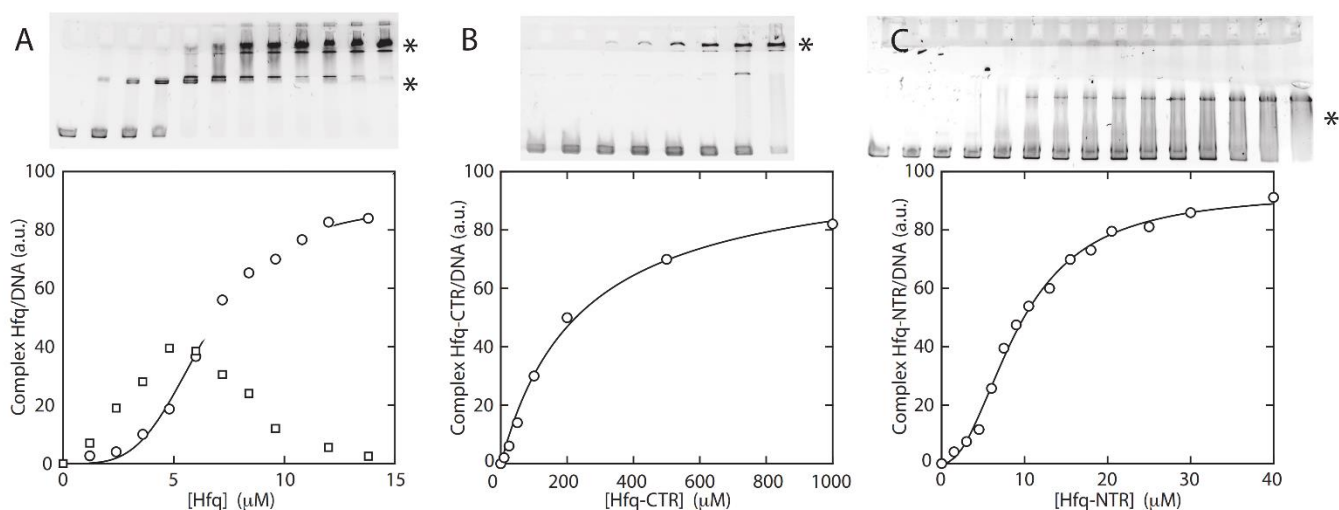
### Thermodynamics of Hfq–CTR binding

The differential power delivered by the calorimeter, corresponding to the time-stepped titration of  $0.5 \text{ mM}$  Hfq–CTR with a  $2.5 \text{ mM}$  (bp) DNA solution in T-buffer ( $\text{pH } 7.5$ ), is displayed in Figure 7. The thermogram exhibits a negative signal, which indicates that the reaction is exothermic and associated with heat release. At long times, heat exchanges close to zero mark the decrease of thermodynamic interactions between DNA and Hfq–CTR. The light gray curve in Figure 7A represents the thermogram obtained from the dilution of DNA in T-buffer. The endothermic signal remains weak and its amplitude is consistent with that of dilution of a polymer or surfactant in this range of concentrations (37,38). Binding isotherms obtained with different concentrations and molar ratios  $X$  of DNA and Hfq–CTR are displayed in Figure 7B–D. Titrations were carried out in various conditions to test the reliability of the assay as well as to explore different ranges of DNA and Hfq–CTR concentrations. The enthalpy curves exhibit a sigmoidal increase with increasing  $X$ , with initial binding values starting around  $-20 \text{ kJ mol}^{-1}$ . Until the end of the titration, the enthalpies remain negative, in accordance with an exothermic reaction.

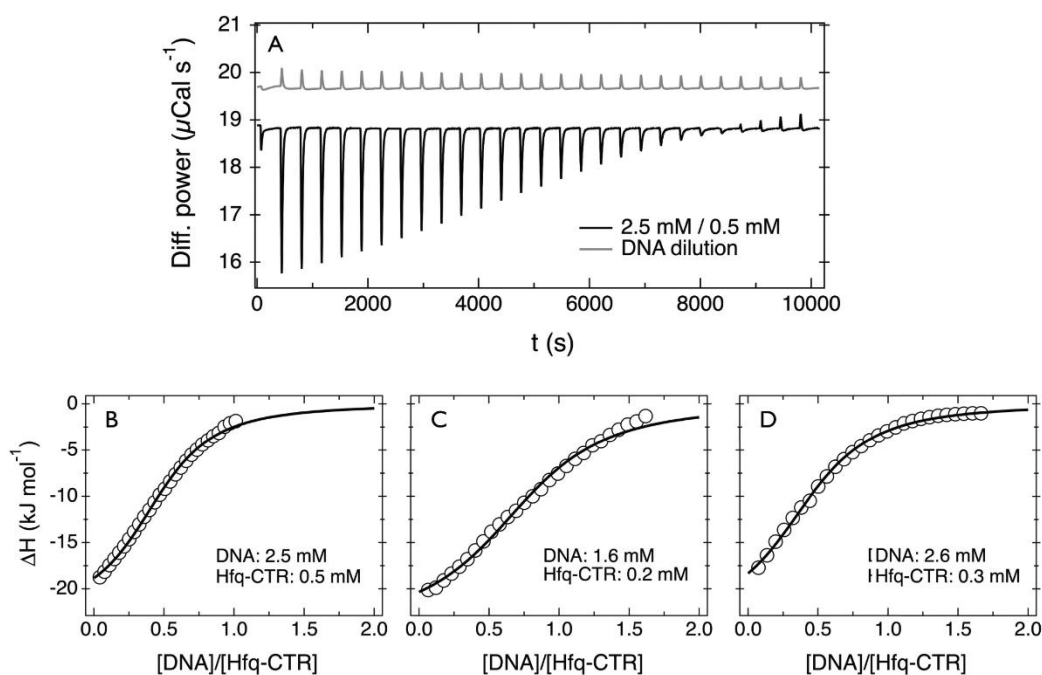
The ITC data were analyzed with the multiple non-interacting sites (MNIS) model (37,39,40). This model gives an expression for the heat exchange of the form:

$$\Delta H(X, n, r) = \frac{1}{2} \Delta H_b \left( 1 + \frac{n - X - r}{((n + X + r)^2 - 4Xn)^{1/2}} \right) \quad (1)$$

where  $\Delta H_b$  is the binding enthalpy,  $X$  the molar ratio and  $n$  the reaction stoichiometry. The coefficient  $r$  is defined as  $r = 1/(K_b[\text{Hfq–CTR}])$ , where  $K_b$  denotes the affinity binding constant. Least square calculations using Equation (1) give the continuous curves shown in Figure 7B–D. The agreement between the prediction of the MNIS model and the measured heat exchange is excellent, and the fit procedure provides a consistent set of thermodynamic data (collected in Table 2). By taking the average of the data from the three titration experiments, we obtain for the binding enthalpy  $\Delta H_b = -23.7 \pm 0.7 \text{ kJ mol}^{-1}$  and for the binding constant  $K_b = (2.3 \pm 0.7) \times 10^4 \text{ M}^{-1}$ . The averaged value of the stoichiometry parameter  $n = 0.6 \pm 0.2$ , which implies that approximately two Hfq–CTR peptides interact with one base pair. Knowing the binding enthalpy and binding constant, the changes in reaction free energy  $\Delta G$  and entropy  $\Delta S$  can be calculated using  $\Delta G = -RT \ln K_b$  and  $\Delta S = (\Delta H_b -$



**Figure 6.** *In vitro* analysis of Hfq protein–DNA binding properties. (A) Gel shift showing the binding of wild-type Hfq to DNA. (B) Binding of Hfq-CTR. (C) Binding of Hfq-NTR. The intensities of the asterisk-marked bands are set out in the graphs versus the concentration of monomeric Hfq, Hfq-CTR and Hfq-NTR, respectively. Note that a significant supershifted species is observed only in the case of wild-type Hfq. Curves represent non-linear least-squares fits of cooperative (Hfq, Hfq-NTR) and non-cooperative (Hfq-CTR) binding models.



**Figure 7.** (A) Differential power as a function of time obtained from calorimetry experiments by titrating Hfq-CTR with DNA ( $T = 298$  K). Molar concentrations are  $[DNA] = 2.5$  mM (bp) and  $[Hfq-CTR] = 0.5$  mM. The upper light-gray curve corresponds to a dilution of DNA in T-buffer. (B–D) Binding isotherms obtained in DNA/Hfq-CTR titration experiments performed in different conditions. In all three cases, the enthalpies are negative and the reactions exothermic. The continuous curves represent Equation (1) with optimized thermodynamic parameters collected in Table 2.

$\Delta G)/T$ . Again, the data obtained from the three titration assays are consistent, with averaged free energy and entropy changes of  $\Delta G = -24.8 \pm 0.8$  kJ mol<sup>-1</sup> and  $\Delta S = 3.7 \pm 0.6$  J mol<sup>-1</sup> K<sup>-1</sup>, respectively.

## DISCUSSION

Qualitatively, the same compaction behavior is observed for Hfq-CTR and full length Hfq in the channel systems

as well as the bulk phase (28). Condensation in the bulk phase requires, however, more than an order of magnitude higher concentration of protein or peptide. Similar behavior has previously been reported for another NAP (H-NS), like-charged proteins (bovine serum albumin and hemoglobin) and neutral crowders (34,41–43). This shows that the conformation, folding and condensation of DNA are not only related to classical controlling factors, such as



**Table 2.** Thermodynamic parameters obtained from the fit of Equation (1) to the binding isotherms in Figure 7

[DNA] (mM)	[Hfq-CTR] (mM)	$\Delta H_b$ (kJ mol <sup>-1</sup> )	$K_b$ (10 <sup>4</sup> M <sup>-1</sup> )	$n$	$\Delta G$ (kJ mol <sup>-1</sup> )	$\Delta S$ (mol <sup>-1</sup> K <sup>-1</sup> )
2.6	0.3	-23.9 ± 0.6	2.2 ± 0.1	0.51 ± 0.01	-24.7 ± 0.2	3 ± 3
2.5	0.5	-23.0 ± 0.5	1.7 ± 0.1	0.51 ± 0.01	-24.1 ± 0.2	4 ± 2
1.6	0.2	-24.3 ± 0.7	3.1 ± 0.3	0.81 ± 0.02	-25.6 ± 0.3	4 ± 3

$\Delta H_b$ ,  $K_b$ ,  $n$ ,  $\Delta G$  and  $\Delta S$  denote the binding enthalpy, binding constant, stoichiometry, free energy and entropy changes, respectively.

osmotic pressure, charge and ligand binding, but that the interplay with confinement is of paramount importance. The critical (monomer) concentrations of Hfq for condensation are similar to those pertaining to Hfq-CTR. The decrease in stretch in the channel systems and compaction into a condensed form indicate Hfq-CTR mediated bridging interaction between different segments of the DNA molecule. In contrast, the marginal effect of Hfq-NTR on the conformation shows that the hexameric torus is not directly involved in the compaction of DNA. Electrostatics plays a minor role in Hfq-mediated compaction, because similar behavior is observed for DNA in T-buffer and T-buffer with 30 mM KGl<sub>u</sub>.

Bridging and looping of longer DNAs by Hfq-CTR is confirmed by the atomic force microscopy images. Furthermore, the images of shorter molecules show ~15% decrease in DNA contour length and ~40% increase in persistence length following binding of Hfq-CTR. These changes in mechanical properties of the double stranded DNA molecule are similar to those previously observed for full length Hfq. However, in order to obtain this result, the Hfq-CTR to bp ratio needed to be increased ten-fold with respect to Hfq, in accordance with Hfq-CTR's relatively low affinity for binding to the double helix as shown by the ITC assay. Note that the changes in contour and persistence length have a minor effect on the extension of the DNA molecules inside the channels. The gradual shortening prior to the collapse into the condensed state is mainly caused by bridging interaction mediated by Hfq-CTR.

The electrophoretic mobility assays show two band shifts for full length Hfq. As previously reported, the high affinity dissociation constant  $K_d$  corresponds to binding of the protein to DNA (10). The second band shift with lower affinity and with a migration distance close to the loading wells indicates the formation of complexes of relatively large molecular weight through protein-mediated bridging of multiple DNA molecules. The bridges can be formed by a single protein and/or self-interactions among multiple proteins. For Hfq-CTR and Hfq-NTR, we observed single band shifts with widely different dissociation constants. In the case of Hfq-NTR, the band shift corresponds to binding of the torus with an affinity comparable to the one of full length Hfq. No aggregated complexes are observed, which agrees with the absence of bridging interaction. A surprising result is that no intermediate band shift is observed for Hfq-CTR. This might be related to a marginal effect on the electrophoretic mobility of DNA following binding of a peptide with relatively small molecular weight. However, the band shift close to the wells shows significant bridging and aggregation of DNA at relatively high concentrations of Hfq-CTR. These results indicate that full length Hfq binds on DNA by interfacing its torus, possibly through electrostatic

interaction between the phosphates and positively charged residues located at the rim and/or distal face of the torus. However, binding of the torus on the double helix by itself does not result in aggregation. For the assembly of multiple DNA molecules or segments thereof, the CTR is clearly required.

The single-molecule compaction, electrophoretic mobility and calorimetry experiments provide information about the propensity for binding and bridging of DNA by Hfq-CTR. Note that the compaction and electrophoretic mobility experiments are sensitive to Hfq-CTR-DNA interaction in an indirect way through the effect on DNA conformation and DNA-DNA aggregation, respectively. In the calorimetry experiment, binding is directly monitored through the exchange of heat. The compaction experiments show a significant effect of nano-confinement, but in the case of the bulk phase assay we obtained a critical Hfq-CTR concentration of around 1  $\mu$ M. As inferred from electrophoresis, the dissociation constant associated with DNA aggregation takes a value  $K_d = 240 \pm 30 \mu$ M. In the calorimetry experiment, the dissociation constant is the inverse of the reaction binding constant  $K_b$ , so that we obtain  $K_d = 1/K_b = 40 \pm 10 \mu$ M. This variety in affinity pertaining to different experimental approaches shows that compaction and aggregation by bridging interaction requires different degrees of saturation. A relatively small number of bridges is sufficient for intramolecular compaction, whereas for aggregation of multiple DNA molecules an excess of the bridging ligand is needed. Besides ligand binding, other environmental factors such as DNA concentration, confinement (gel matrix), crowding and buffer conditions are clearly of importance in controlling conformation and intermolecular organization.

The fitted stoichiometry ratio  $n$  implies about two bound Hfq-CTR peptides per base pair. However, not all bound peptides are necessarily anchored to the double helix. They may also be bound by peptide aggregation with concomitant less peptide-DNA interaction. Expressed in thermal energy units, the binding enthalpy, free energy and entropy contributions are around -9.5, -10 and 0.5  $k_B T$ , respectively. These values agree with a strong, enthalpy driven assembly and a marginal entropic contribution to the reaction thermodynamics. From a comparison with literature data for condensation of DNA by multivalent ions or phospholipids, it follows that the binding of Hfq-CTR on DNA is not controlled by electrostatics (44,45). In the case of electrostatic complexation, the entropic contribution is expected to be much larger than the enthalpic one, due to the release of counterions associated with ion pairing. Other calorimetric studies on complexation involving nanoparticles, polymers and/or surfactants have shown similar behavior (37,38,46). Furthermore, at physiological pH the

CTR is almost electroneutral (one basic and four acidic residues), highly flexible and contains many residues with short polar side-groups. Accordingly, other types of interaction, such as hydrogen-bonding with the relatively abundant histidine and serine residues, must be involved in the binding of the CTR on DNA.

## CONCLUSION

From the similar equilibrium dissociation constants of Hfq and Hfq-NTR, it is clear that the torus facilitates binding of full length Hfq on DNA. However, as shown by the nanofluidics and bulk phase condensation experiments, binding by itself does not result in compaction of DNA into a condensed state. For compaction, the presence of the CTR is clearly required. Binding of Hfq-CTR results in similar changes in the mechanical properties of the duplex as Hfq. However, as inferred from the moderate increase in bending rigidity, no rigid filament is formed as for another NAP H-NS (42,47,48). The decrease in contour length indicates a modification of DNA's secondary structure, that is opening of AT-rich tracts and/or a transition of the duplex from the B- to the A-form (10). Precisely, the transition from the B- to A-form is favoured by low hydration, so that the expulsion of interfacial water by protein binding affects DNA helicity.

In consideration of the present and previous results, we arrive at the following model for binding and bridging of Hfq on the double helix. With high affinity, the hexameric protein lands on the duplex through interfacing its torus, followed by anchoring through binding of one or more of its C-terminal arms. The initial binding is possibly facilitated by electrostatic interaction between the phosphates of the double helix and positively charged residues located at the rim and/or distal face of the torus. Anchoring by binding of the CTR is however strongly enthalpy driven and stabilized by non-electrostatic interactions such as hydrogen bonding with the polar residues. It could also be sequence specific, since structural modification of the duplex is involved as shown by changes in bending rigidity and contour length of the double helix. Once bound, Hfq can form a bridge through anchoring one or more of its other CTR arms on another section of the same or another DNA molecule and/or by CTR-mediated self-interactions among multiple proteins. The propensity for bridging and compaction of DNA by the C-terminal domain might be related to aggregation of bound protein and may have implications for protein binding related gene regulation.

## FUNDING

NUS-USPC collaboration grant; Singapore Ministry of Education grant [MOE2014-T2-1-001]; French Embassy in Poland Scholarship (to K. K.). Funding for open access charge: Singapore Ministry of Education.

*Conflict of interest statement.* None declared.

## REFERENCES

- Sun, X., Zhulin, I. and Wartell, R. M. (2002) Predicted structure and phylogenetic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res.*, **30**, 3662–3671.
- de Fernandez, M.T.F., Hayward, W.S. and August, J.T. (1972) Bacterial proteins required for replication of phage Q $\beta$  ribonucleic acid. *J. Biol. Chem.*, **247**, 824–831.
- Tsui, H.C.T., Leung, H. C.E. and Winkler, M.E. (1994) Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12. *Mol. Microbiol.*, **13**, 35–49.
- Vogel, J. and Luisi, B.F. (2011) Hfq and its constellation of RNA. *Nat. Rev. Microbiol.*, **9**, 578–589.
- Sobrero, P. and Valverde, C. (2012) The bacterial protein Hfq: much more than a mere RNA-binding factor. *Crit. Rev. Microbiol.*, **38**, 1–24.
- Gottesman, S. and Storz, G. (2015) RNA reflections: converging on Hfq. *RNA*, **21**, 511–512.
- Cech, G.M., Pakuła, B., Kamrowska, D., Wegrzyn, G., Arluison, V. and Szalewska-Pałasz, A. (2014) Hfq protein deficiency in Escherichia coli affects ColE1-like but not  $\lambda$  plasmid DNA replication. *Plasmid*, **73**, 10–15.
- Ellis, M.J. and Haniford, D.B. (2016) Riboregulation of bacterial and archaeal transposition. *Wiley Interdiscip. Rev. RNA*, **7**, 382–398.
- Le Derout, J., Boni, I.V., R gnier, P. and Hajsnsdorf, E. (2010) Hfq affects mRNA levels independently of degradation. *BMC Mol. Biol.*, **11**, 17.
- Geinguenaud, F., Calandrini, V., Teixeira, J., Mayer, C., Liquier, J., Lavelle, C. and Arluison, V. (2011) Conformational transition of DNA bound to Hfq probed by infrared spectroscopy. *Phys. Chem. Chem. Phys.*, **13**, 1222–1229.
- Sukhodolets, M.V. and Garges, S. (2003) Interaction of Escherichia coli RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. *Biochemistry*, **42**, 8022–8034.
- Rabhi, M., Esp li, O., Schwartz, A., Cayrol, B., Rahmouni, A.R., Arluison, V. and Boudvillain, M. (2011) The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *EMBO J.*, **30**, 2805–2816.
- Azam, T.A. and Ishihama, A. (1999) Twelve species of the nucleoid-associated protein from Escherichia coli sequence recognition specificity and DNA binding affinity. *J. Biol. Chem.*, **274**, 33105–33113.
- Azam, T.A., Iwata, A., Nishimura, A., Ueda, S. and Ishihama, A. (1999) Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *J. Bacteriol.*, **181**, 6361–6370.
- Azam, T.A., Hiraga, S. and Ishihama, A. (2000) Two types of localization of the DNA-binding proteins within the Escherichia coli nucleoid. *Genes Cells*, **5**, 613–626.
- Diestra, E., Cayrol, B., Arluison, V. and Risco, C. (2009) Cellular electron microscopy imaging reveals the localization of the Hfq protein close to the bacterial membrane. *PLoS One*, **4**, e8301.
- Tsui, H.C.T., Feng, G. and Winkler, M.E. (1997) Negative regulation of mutS and mutH repair gene expression by the Hfq and RpoS global regulators of Escherichia coli K-12. *J. Bacteriol.*, **179**, 7476–7487.
- Azam, T.A. and Ishihama, A. (2015) Growth phase dependent changes in the structure and protein composition of nucleoid in Escherichia coli. *Sci. China Life Sci.*, **58**, 902–911.
- Folichon, M., Arluison, V., Pellegrini, O., Huntzinger, E., R gnier, P. and Hajsnsdorf, E. (2003) The poly (A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. *Nucleic Acids Res.*, **31**, 7302–7310.
- Kajitani, M. and Ishihama, A. (1991) Identification and sequence determination of the host factor gene for bacteriophage Q $\beta$ . *Nucleic Acids Res.*, **19**, 1063–1066.
- Ohniwa, R.L., Muchaku, H., Saito, S., Wada, C. and Morikawa, K. (2013) Atomic force microscopy analysis of the role of major DNA-binding proteins in organization of the nucleoid in Escherichia coli. *PLoS One*, **8**, e72954.
- Dorman, C.J. (2009) Nucleoid-associated proteins and bacterial physiology. *Adv. Appl. Microbiol.*, **67**, 47–64.
- Link, T.M., Valentin-Hansen, P. and Brennan, R.G. (2009) Structure of Escherichia coli Hfq bound to polyribadenylate RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 19292–19297.
- Updegrove, T.B., Correia, J.J., Galletto, R., Bujalowski, W. and Wartell, R.M. (2010) E. coli DNA associated with isolated Hfq interacts with Hfq's distal surface and C-terminal domain. *Biochim. Biophys. Acta*, **1799**, 588–596.



25. Arluison, V., Folichon, M., Marco, S., Derreumaux, P., Pellegrini, O., Seguin, J., Hajnsdorf, E. and Regnier, P. (2004) The C-terminal domain of *Escherichia coli* Hfq increases the stability of the hexamer. *Eur. J. Biochem.*, **271**, 1258–1265.
26. Takada, A., Wachi, M., Kaidow, A., Takamura, M. and Nagai, K. (1997) DNA binding properties of the hfq gene product of *Escherichia coli*. *Biochem. Biophys. Res. Commun.*, **236**, 576–579.
27. Fortas, E., Piccirilli, F., Malabirade, A., Militello, V., Tréput, S., Marco, S., Taghbalout, A. and Arluison, V. (2015) New insight into the structure and function of Hfq C-terminus. *Biosci. Rep.*, **35**, 1–9.
28. Jiang, K., Zhang, C., Guttula, D., Liu, F., van Kan, J.A., Lavelle, C., Kubiak, K., Malabirade, A., Lapp, A., Arluison, V. et al. (2015) Effects of Hfq on the conformation and compaction of DNA. *Nucleic Acids Res.*, **43**, 4332–4341.
29. Zhang, C., Zhang, F., van Kan, J.A. and van der Maarel, J.R.C. (2008) Effects of electrostatic screening on the conformation of single DNA molecules confined in a nanochannel. *J. Chem. Phys.*, **128**, 225109.
30. van Kan, J.A., Zhang, C., Malar, P. and van der Maarel, J.R.C. (2012) High throughput fabrication of disposable nanofluidic lab-on-chip devices for single molecule studies. *Biomicrofluidics*, **6**, 036502.
31. van Kan, J.A., Bettioli, A.A. and Watt, F. (2006) Proton beam writing of three-dimensional nanostructures in hydrogen silsesquioxane. *Nano Lett.*, **6**, 579–582.
32. van Kan, J.A., Shao, P.G., Wang, Y.H. and Malar, P. (2011) Proton beam writing a platform technology for high quality three-dimensional metal mold fabrication for nanofluidic applications. *Microsyst. Technol.*, **17**, 1519–1527.
33. Johansen, F. and Jacobsen, J.P. (1998) <sup>1</sup>H NMR studies of the bis-intercalation of a homodimeric oxazole yellow dye in DNA oligonucleotides. *J. Biomol. Struct. Dyn.*, **16**, 205–222.
34. Zhang, C., Shao, P.G., van Kan, J.A. and van der Maarel, J.R.C. (2009) Macromolecular crowding induced elongation and compaction of single DNA molecules confined in a nanochannel. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 16651–16656.
35. Wiggins, P.A., van der Heijden, T., Moreno-Herrero, F., Spakowitz, A., Phillips, R., Widom, J., Dekker, C. and Nelson, P.C. (2006) High flexibility of DNA on short length scales probed by atomic force microscopy. *Nat. Nanotechnol.*, **1**, 137–141.
36. Kundukad, B., Cong, P., van der Maarel, J.R.C. and Doyle, P.S. (2013) Time-dependent bending rigidity and helical twist of DNA by rearrangement of bound HU protein. *Nucleic Acids Res.*, **41**, 8280–8288.
37. Courtois, J. and Berret, J.-F. (2010) Probing oppositely charged surfactant and copolymer interactions by isothermal titration microcalorimetry. *Langmuir*, **26**, 11750–11758.
38. Vitorazi, L., Ould-Moussa, N., Sekar, S., Fresnais, J., Loh, W., Chapel, J.-P. and Berret, J.-F. (2014) Evidence of a two-step process and pathway dependency in the thermodynamics of poly (diallyldimethylammonium chloride)/poly (sodium acrylate) complexation. *Soft Matter*, **10**, 9496–9505.
39. Herrera, I. and Winnik, M.A. (2013) Differential binding models for isothermal titration calorimetry: moving beyond the Wiseman isotherm. *J. Phys. Chem. B*, **117**, 8659–8672.
40. Wiseman, T., Williston, S., Brandts, J.F. and Lin, L.-N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.*, **179**, 131–137.
41. Zhang, C., Gong, Z., Guttula, D., Malar, P.P., van Kan, J.A., Doyle, P.S. and van der Maarel, J.R.C. (2012) Nanoionic compaction of DNA by like-charged protein. *J. Phys. Chem. B*, **116**, 3031–3036.
42. Zhang, C., Guttula, D., Liu, F., Malar, P.P., Ng, S.Y., Dai, L., Doyle, P.S., van Kan, J.A. and van der Maarel, J.R.C. (2013) Effect of H-NS on the elongation and compaction of single DNA molecules in a nanospace. *Soft Matter*, **9**, 9593–9601.
43. van der Maarel, J.R.C., Zhang, C. and van Kan, J.A. (2014) A nanochannel platform for single DNA studies: from crowding, protein DNA interaction, to sequencing of genomic information. *Isr. J. Chem.*, **54**, 1573–1588.
44. Matulis, D., Rouzina, I. and Bloomfield, V.A. (2000) Thermodynamics of DNA binding and condensation: isothermal titration calorimetry and electrostatic mechanism. *J. Mol. Biol.*, **296**, 1053–1063.
45. Kennedy, M.T., Pozharski, E.V., Rakhmanova, V.A. and MacDonald, R.C. (2000) Factors governing the assembly of cationic phospholipid-DNA complexes. *Biophys. J.*, **78**, 1620–1633.
46. Loosli, F., Vitorazi, L., Berret, J.-F. and Stoll, S. (2015) Towards a better understanding on agglomeration mechanisms and thermodynamic properties of TiO<sub>2</sub> nanoparticles interacting with natural organic matter. *Water Res.*, **80**, 139–148.
47. Amit, R., Oppenheim, A.B. and Stavans, J. (2003) Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor. *Biophys. J.*, **84**, 2467–2473.
48. van der Maarel, J.R.C., Guttula, D., Arluison, V., Egelhaaf, S.U., Grillo, I. and Forsyth, V.T. (2016) Structure of the H-NS-DNA nucleoprotein complex. *Soft Matter*, **12**, 3636–3642.

Annexe F :

*Membrane association of the  
bacterial riboregulator Hfq and  
functional perspectives*

# SCIENTIFIC REPORTS

OPEN

## Membrane association of the bacterial riboregulator Hfq and functional perspectives

Antoine Malabirade<sup>1</sup>, Javier Morgado-Brajones<sup>1,2</sup>, Sylvain Trépout<sup>3,4</sup>, Frank Wien<sup>5</sup>, Ileana Marquez<sup>2</sup>, Jérôme Seguin<sup>6</sup>, Sergio Marco<sup>3,4</sup>, Marisela Velez<sup>2</sup> & Arluison Véronique<sup>1,7</sup>

Received: 5 June 2017

Accepted: 17 August 2017

Published online: 06 September 2017

Hfq is a bacterial RNA binding protein that carries out several roles in genetic expression regulation, mainly at the post-transcriptional level. Previous studies have shown its importance in growth and virulence of bacteria. Here, we provide the direct observation of its ability to interact with membranes. This was established by co-sedimentation assay, cryo-transmission electron (cryo-TEM) and atomic force (AFM) microscopies. Furthermore, our results suggest a role for its C-terminus amyloidogenic domain in membrane disruption. Precisely, AFM images of lipid bilayers in contact with Hfq C-terminus fibrils show the emergence of holes with a size dependent on the time of interaction. Cryo-TEM observations also show that liposomes are in contact with clusters of fibrils, with occasional deformation of the vesicles and afterward the apparition of a multitude of tiny vesicles in the proximity of the fibrils, suggesting peptide-induced breakage of the liposomes. Finally, circular dichroism spectroscopy demonstrated a change in the secondary structure of Hfq C-terminus upon interaction with liposomes. Altogether, these results show an unexpected property of Hfq and suggest a possible new role for the protein, exporting sRNA outside of the bacterial cell.

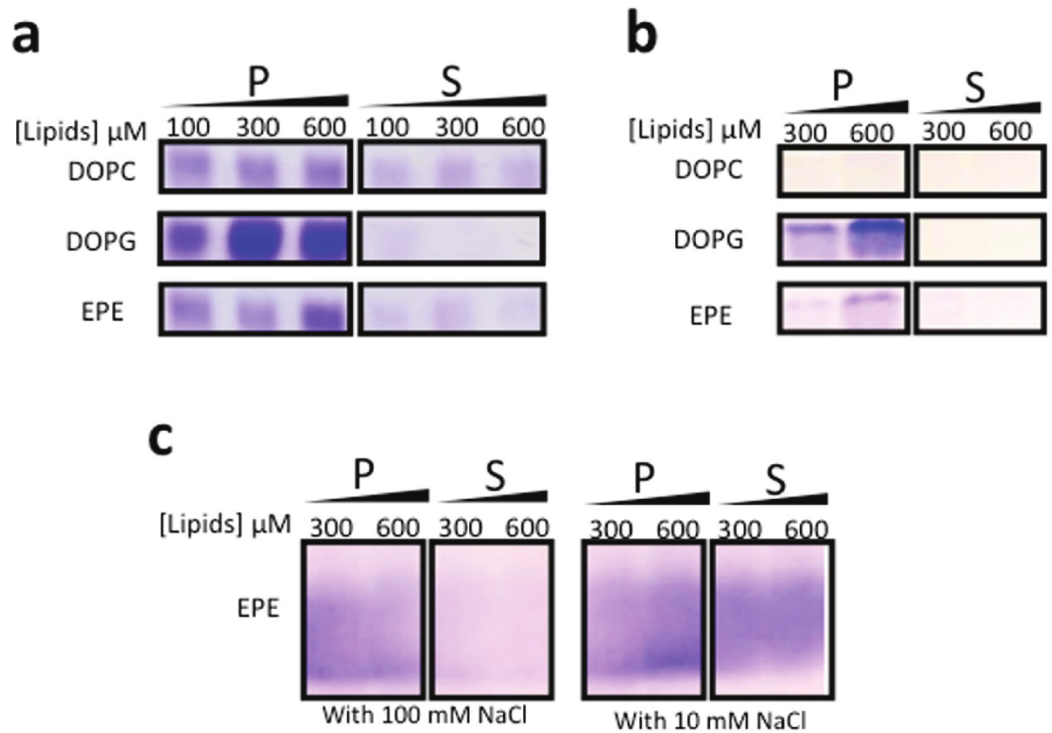
The bacterial Hfq protein was identified in *Escherichia coli* as an abundant RNA-binding protein about fifty years ago<sup>1</sup>. Today Hfq is recognized as the core component of a global post-transcriptional network that facilitates the imperfect base-pairing of small regulatory noncoding RNA (sRNA) with trans-encoded mRNA targets<sup>2,3</sup>. Consequently, Hfq induces the repression or activation of translation, with important related consequences for RNA stability<sup>4</sup>. This mode of regulation is of primary importance for the adaptation of bacteria to the changing environment, and thus for the control of cell division and for the virulence of pathogenic species<sup>5,6</sup>.

The ability of Hfq to interact with DNA has also been demonstrated and it has been described as one of the *E. coli* nucleoid associated proteins (NAP) *in vivo*<sup>7,8</sup>. Nevertheless, the fraction of Hfq within the nucleoid represents only 10–20% of the total protein concentration, while approximately 50% of the protein is found in close proximity of the membrane<sup>9,10</sup>. This makes sense as the majority of Hfq-regulated sRNA targets encode membrane proteins<sup>11</sup>.

Regarding its molecular structure, *E. coli* Hfq forms an Sm-fold in its N-terminal region (NTR, 2/3 of the 102 amino acid residues protein)<sup>12,13</sup>. This fold consists of a five-stranded antiparallel  $\beta$ -sheet capped by an N-terminal  $\alpha$ -helix. The  $\beta$ -sheets from six monomers interact with each other to assemble in a toroidal structure with two different surfaces<sup>14</sup>. It appears that the distal face and the edge of the protein are involved in both DNA and RNA binding, while the proximal face (on which the  $\alpha$ -helix is exposed) seems to be involved in RNA binding only<sup>15,16</sup>. Besides its Sm-like domain, other regions of Hfq such as the C-terminal region (CTR) also play a role in nucleic acid recognition and binding<sup>17–19</sup>. The Hfq three-dimensional structures of various bacteria have been resolved<sup>20–24</sup>, but until now all lack the CTR, so the way this Hfq region folds remains enigmatic.

<sup>1</sup>Laboratoire Léon Brillouin LLB, CEA, CNRS UMR12, Université Paris Saclay, CEA Saclay, 91191, Gif-sur-Yvette, France. <sup>2</sup>Instituto de Catálisis y Petroleoquímica, CSIC, c/ Marie Curie, 2, Cantoblanco, E-28049, Madrid, Spain. <sup>3</sup>Institut Curie, Research Center, PSL Research University, Chemistry, Modélisation and Imaging for Biology (CMIB) Bât 110-112, Centre Universitaire, 91405, Orsay, France. <sup>4</sup>INSERM U 1196, CNRS UMR 9187, Université Paris Saclay, Université Paris-Sud, Bât 110-112, Centre Universitaire, Rue Henri Becquerel, 91405, Orsay, France. <sup>5</sup>DISCO Beamline, Synchrotron SOLEIL, 91192, Gif-sur-Yvette, France. <sup>6</sup>Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91198, Gif-sur-Yvette, Cedex, France. <sup>7</sup>Université Paris Diderot, 75013, Paris, France. Correspondence and requests for materials should be addressed to A.V. (email: [veronique.arluison@univ-paris-diderot.fr](mailto:veronique.arluison@univ-paris-diderot.fr))





**Figure 1.** Co-sedimentation assay of Hfq with select lipids. SUVs in SUV buffer (10 mM Tris-HCl pH 7.5, 100 mM NaCl) were incubated with 3 μM Hfq-WT/Hfq-NTR<sub>65</sub> or 126 μM Hfq-CTR<sub>38</sub>. **(a)** Wild-type Hfq sediments with a preference for polar lipids. **(b)** The same qualitative result is obtained with truncated Hfq-NTR<sub>65</sub>. **(c)** Hfq-CTR<sub>38</sub> peptides also bind SUVs and the interaction is promoted by high ionic strength. P, Pellet fraction; S, Supernatant fraction. Gels were different for Hfq-WT, Hfq-NTR<sub>65</sub> and Hfq-CTR<sub>38</sub> peptides, but for each condition protein analysis was made on the same gel. Note that as previously described, Hfq-WT hexamers are only partially denatured by SDS-PAGE and migrate mostly as hexamers<sup>27,60,61</sup>, while truncated Hfq-NTR<sub>65</sub> is less stable and can be dissociated in the PAGE<sup>25,27</sup> (Fig. S6). Protein concentrations were chosen in order to have similar mass concentrations.

Structurally, the CTR seems to extend outside the Sm-core and appears to be intrinsically disordered<sup>25,26</sup>. This domain is also presumably important for protein stability<sup>25,27</sup>. Recently, it has been shown that the CTR region of *E. coli* Hfq contains an amyloid sequence<sup>28</sup>. This allows the Hfq protein to self-assemble, a feature common to Hfq and Sm proteins<sup>29,30</sup>.

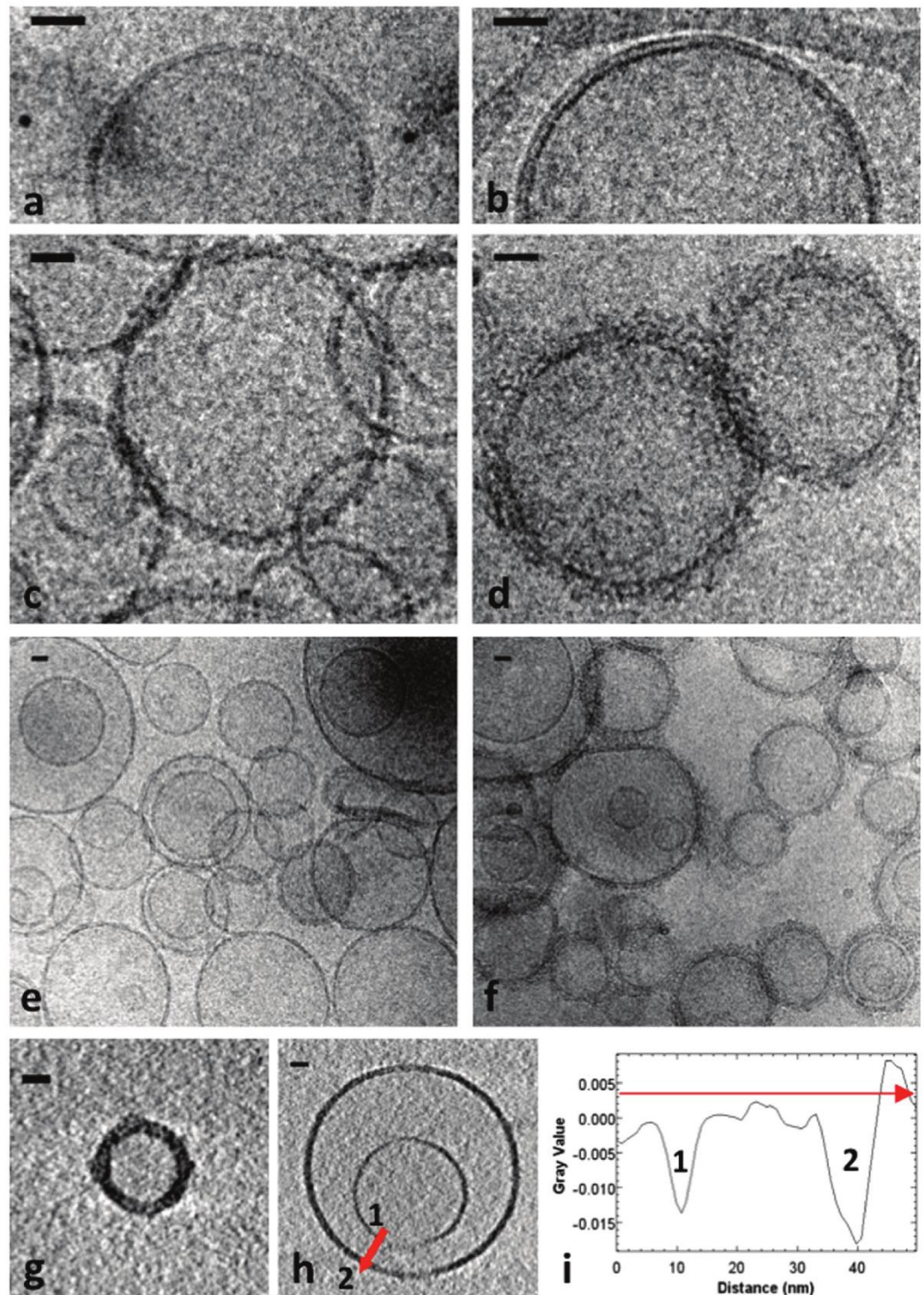
In this work, we focus our attention on a new unexplored feature of Hfq, namely its ability to bind to membranes. This finding opens new perspectives in Hfq-dependent riboregulation that will be discussed herein.

## Results

**Hfq CTR and NTR domains both interact with lipid bilayers.** In order to have a broad view of the ability of Hfq to interact with lipid bilayers, several lipids were first tested for potential interactions by co-sedimentation assays. In all cases, we observed that the protein alone at the concentration specified did not sediment in the conditions used. Results are presented in Fig. 1. The wild-type Hfq hexamer has a pronounced affinity for dioleoylphosphatidylglycerol (DOPG) polar lipids, which are present in the natural inner membrane of *E. coli* and therefore in the *E. coli* polar extract (EPE) mixture, partially explaining the good affinity for the EPE small unilamellar vesicles (SUVs) (Fig. 1a). Only a small interaction is observed with dioleoylphosphatidylcholine (DOPC), an example of less polar lipids. This result is however not surprising considering the highly hydrophilic and positively charged *E. coli* Hfq surface. A comparable behavior is obtained with truncated Hfq-NTR<sub>65</sub> missing the Hfq CTR tail (Fig. 1b).

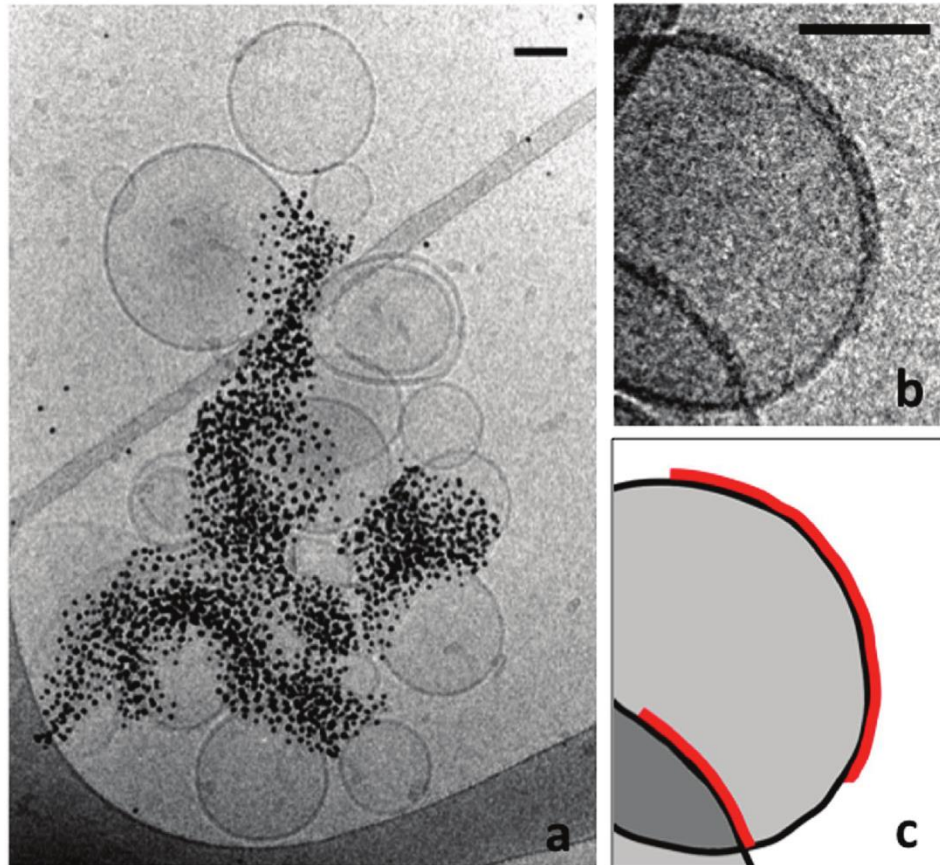
We also investigated the Hfq-CTR<sub>38</sub> propensity to interact with SUVs, knowing its tendency to aggregate into amyloid fibers<sup>28</sup> and the natural property of such a domain to interact with membranes<sup>31</sup>. Results with the 38 amino-acid peptide corresponding to Hfq CTR tail are presented in Fig. 1c, showing a significant binding of CTR peptide to EPE lipids. Note that this interaction is promoted by the presence of NaCl, and the opposite is observed for the NTR domain alone, thus suggesting a different mode of interaction.

**Clusters of Hfq are in direct contact with lipid bilayers.** As seen by Cryo-electron microscopy (cryo-TEM), SUVs formed from EPE or DOPG lipids are round-shaped with a mean diameter ~ 120 nm (Fig. 2a and b). The membrane bilayer with the expected thickness of  $4.1 \pm 0.3$  nm is clearly observed (in agreement with previous reports<sup>32</sup>). Liposomes were sparsely distributed over the grid and are mostly found isolated. Mixing SUVs with full-length WT Hfq induces SUVs clustering. The liposomes are decorated with proteins, leading to



**Figure 2.** Cryo-microscopy images and tomography of wild-type Hfq bound to SUVs. **(a)** Liposomes composed of 100% EPE lipids show a clean bilayer. **(b)** Same as **(a)** but with DOPG lipids. **(c)** EPE liposomes incubated with 10 μM Hfq. **(d)** DOPG liposomes incubated with 10 μM Hfq. **(e)** and **(f)** Same as **(c)** and **(d)** with lower magnification, respectively. **(g)** Cryo-tomography of Hfq and DOPG SUVs showing a bilayer completely covered by proteins, which eventually form aggregates visible on the surface. **(h)** Tomography of double DOPG liposomes showing that only the outer membrane (2) is decorated by Hfq, and is thicker. The inner membrane (1) has an undecorated bilayer. **(i)** Grey intensity profile is displayed according to the red arrow on image **(h)**, showing the thickness difference between liposome 1 and 2. Scale bars: 20 nm.



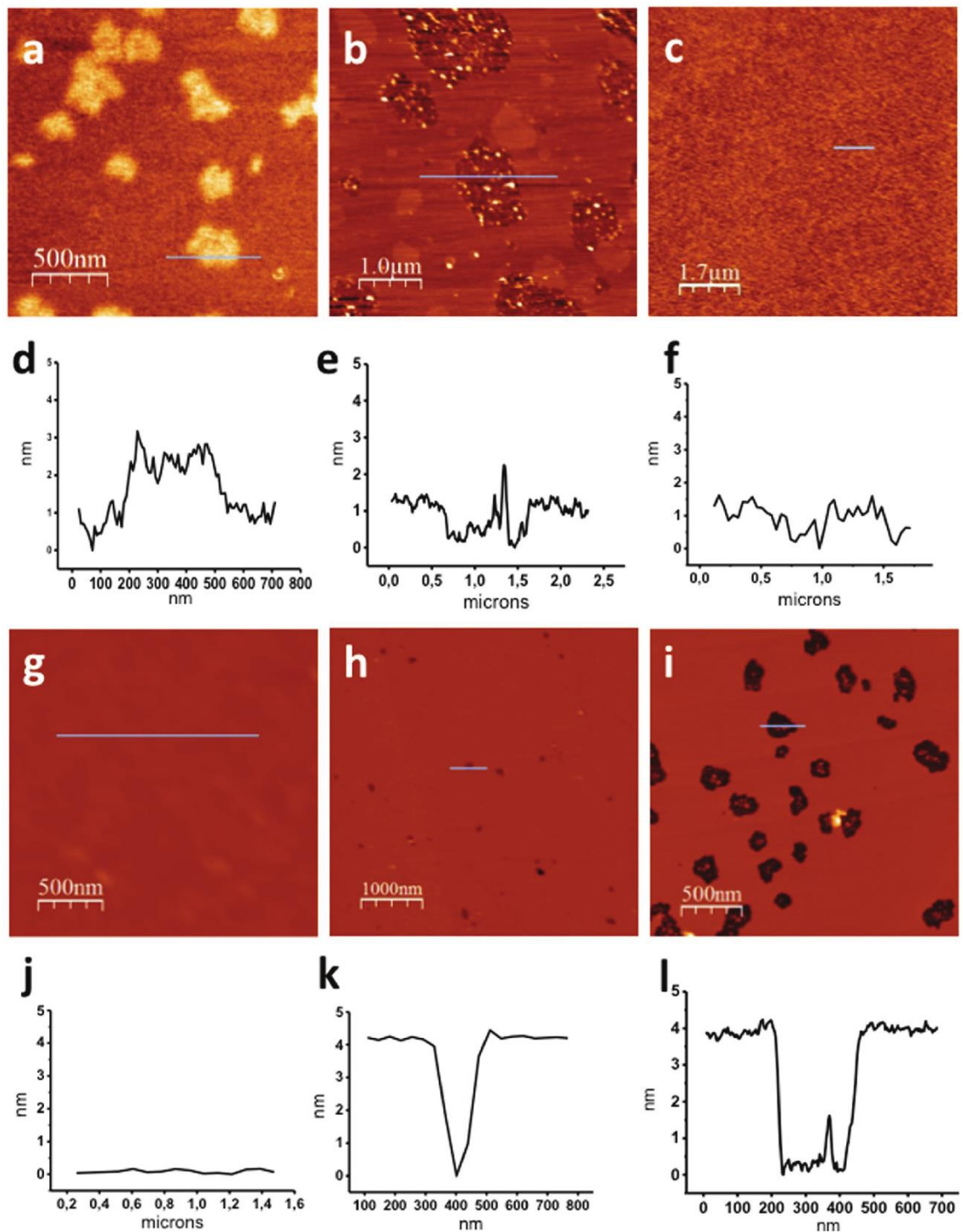


**Figure 3.** Cryo-microscopy images of Hfq bound to SUVs. **(a)** EPE SUVs incubated with 3  $\mu\text{M}$  His-tagged Hfq, 4 nM Ni-NTA-NanoGold beads were added a few seconds before freezing. **(b)** EPE SUVs incubated with 10  $\mu\text{M}$  wild-type Hfq. **(c)** Schematic drawing underlining membranes (black) and protein layers (red). Note the formation of Hfq clusters at the liposome interface, suggesting a cooperative binding of the protein to membranes. Scale bars: 50 nm.

a significant increase of membrane thickness:  $7.6 \pm 1.0$  nm or  $8.2 \pm 1.2$  nm for EPE or DOPG liposomes, respectively (Fig. 2c–f). A subtraction between the respective membrane thickness of uncoated and coated liposomes gives a layer of protein measuring approximately 3–3.5 nm for EPE. As Hfq hexamer measures about 3 nm in height, this suggests that hexamers are lying horizontally on the membrane. Note that the protein layer is not completely flat, reflecting a complex arrangement, and that the bilayer is sometimes no longer visible, suggesting that Hfq may sometimes affect the membrane surface. Furthermore, proteins cover almost all the DOPG liposomes, showing some thicker stacks, whereas the EPE SUVs are not always decorated (Fig. 2e). This is consistent with the apparent higher affinity for DOPG observed by co-sedimentation. In order to evaluate the global shape and protein coating of SUVs, we also performed cryo-tomography. 3D reconstruction gives very thick liposomes, normally shaped and fully decorated by proteins, in accordance with TEM results (Fig. 2g). Occasionally, a small SUV is trapped inside a larger one coated by Hfq (Fig. 2h), allowing clear observation of the different thicknesses; measurement of the membranes thickness along the red arrow gives approximately 3.87 nm for the inner liposome<sup>32</sup> and 6.82 nm for the outer one, in accordance with measurements obtained on projections (Fig. 2i).

To prove that Hfq promotes liposome-liposome contacts, we then used C-terminal His-tagged Hfq in the presence of  $\text{Ni}^{2+}$ -functionalized gold beads, to denote the position of the proteins (Fig. 3a). Note that in this condition, the His-tagged protein concentration is three times lower than the wild-type Hfq concentration used in Fig. 2. Control experiments showing beads and liposomes without protein (Fig. S1a) and beads weakly bound to the natural histidines of the wild-type protein (Fig. S1b) are provided as supplementary data. As seen, gold beads are not randomly coating liposomes but clustered, suggesting that proteins are organized in clusters (possibly fibers) at the liposome-liposome interfaces. This observation is consistent with the partial decoration of liposomes, indicating that Hfq protein tends to bind preferentially to protein-lipid clusters already formed (Fig. 3b and highlight in Fig. 3c). This observation may suggest that Hfq binding to membranes is cooperative.

In order to confirm this result, we then used atomic force microscopy (AFM) in solution to observe Hfq behavior on supported lipid bilayers. As expected, after a short incubation time on an EPE bilayer, proteins are distributed over the surface in small aggregates, compatible with hexamers regrouped in clusters (Fig. 4a). The aggregates have an average height of  $2.7 \pm 0.2$  nm, obtained from measuring the profile of 20 different aggregates



**Figure 4.** AFM images of wild-type Hfq at  $2\ \mu\text{M}$  (*a,b,d,e*), Hfq-NTR<sub>65</sub> at  $2\ \mu\text{M}$  (*c,f*), or Hfq-CTR<sub>11</sub> at  $176\ \mu\text{M}$  (*g-l*) in solution on an EPE supported bilayer. (*a*) After 5 min incubation, Hfq binds on the top and forms groups of several hexamers. (*b*) After 1 h incubation, the bilayer is affected and 1–2 nm deep holes appear, containing undetermined material inside. (*c*) After 1 h incubation, no trace of Hfq-NTR<sub>65</sub> is visible and the membrane is clean. *g*. Clean bilayer before adding Hfq-CTR<sub>11</sub>. *h*. After 30 min incubation, small 4 nm deep holes start to form. *i*. After 1 h incubation, holes have grown to several hundreds of nm wide, containing undetermined material inside. *d, e, f, j, k* and *l*, height profiles of grey lines on images *a, b, c, g, h* and *i*, respectively. Protein concentrations were chosen in order to have similar mass concentrations.



(Fig. S2). These are lower than previous cryo-TEM measurements, but membrane curvature is different and the measurement is still compatible with a horizontal-lying torus that could slightly enter the first lipid layer.

**Full length Hfq affects the bilayer organization, but not HfqNTR<sub>65</sub>.** Surprisingly, after longer incubation times, the membrane exhibits a drastic reorganization. Lipid domains appear, which is comprehensible when a positively charged protein interacts with anionic phospholipids present in the EPE<sup>33</sup>. However, we also observed the presence of numerous unexpected holes with a diameter around 100 nm and 1–2 nm in depth (Fig. 4b). As the membrane treated in the same condition in the absence of Hfq does not present such a reorganization on the same timescale, this indicates that Hfq affects the membrane integrity. Note that some undetermined material is still present inside the deep holes, which does not correspond to the ~4 nm thickness of the bilayer (Fig. 4e). We assume this could be Hfq or Hfq-lipids aggregates resulting from the membrane reorganization.

We then performed the same set of experiments with Hfq-NTR<sub>65</sub>. Cryo-TEM revealed that, to a lower extent than full length Hfq, Hfq-NTR<sub>65</sub> was able to regroup liposomes, but no visible protein decoration is observed on the surface (Fig. S3). Furthermore, AFM with Hfq-NTR<sub>65</sub> gives different results than with Hfq-WT, closer to a normal membrane (Fig. 4c,f). The bilayer is rougher than before being incubated with the protein, but does not undergo reorganization and holes formation even after a long incubation time. It suggests that the truncated protein adsorbs weakly on the membrane and is easily washed away when rinsing with buffer before imaging, or pushed by the AFM tip.

In order to understand which part of the protein is involved in the hole-formation observed with the full-length protein, we then investigated the effects of the amyloidogenic Hfq-CTR, which is also able to interact with EPE SUVs (Fig. 1c).

**Identification of the amyloid region in Hfq-CTR.** First, the amyloid region within CTR domain of Hfq was identified precisely. To that end, various peptides corresponding to overlapping sequences in the CTR were synthesized according to Waltz algorithm predictions<sup>34</sup> (see Table S1 for sequences). Self-assembly of these peptides was tested by cryo-TEM and positive results are summarized in Fig. S4. This allowed us to determine that a short sequence of 11 amino acids residues is responsible for the self-assembly, precisely the region SAQNTSAQQDS. We also confirmed by FTIR spectroscopy<sup>35</sup> that this small peptide self-assembles into amyloid fibrils. This 11 amino-acid residues peptide will be referred to as Hfq-CTR<sub>11</sub>.

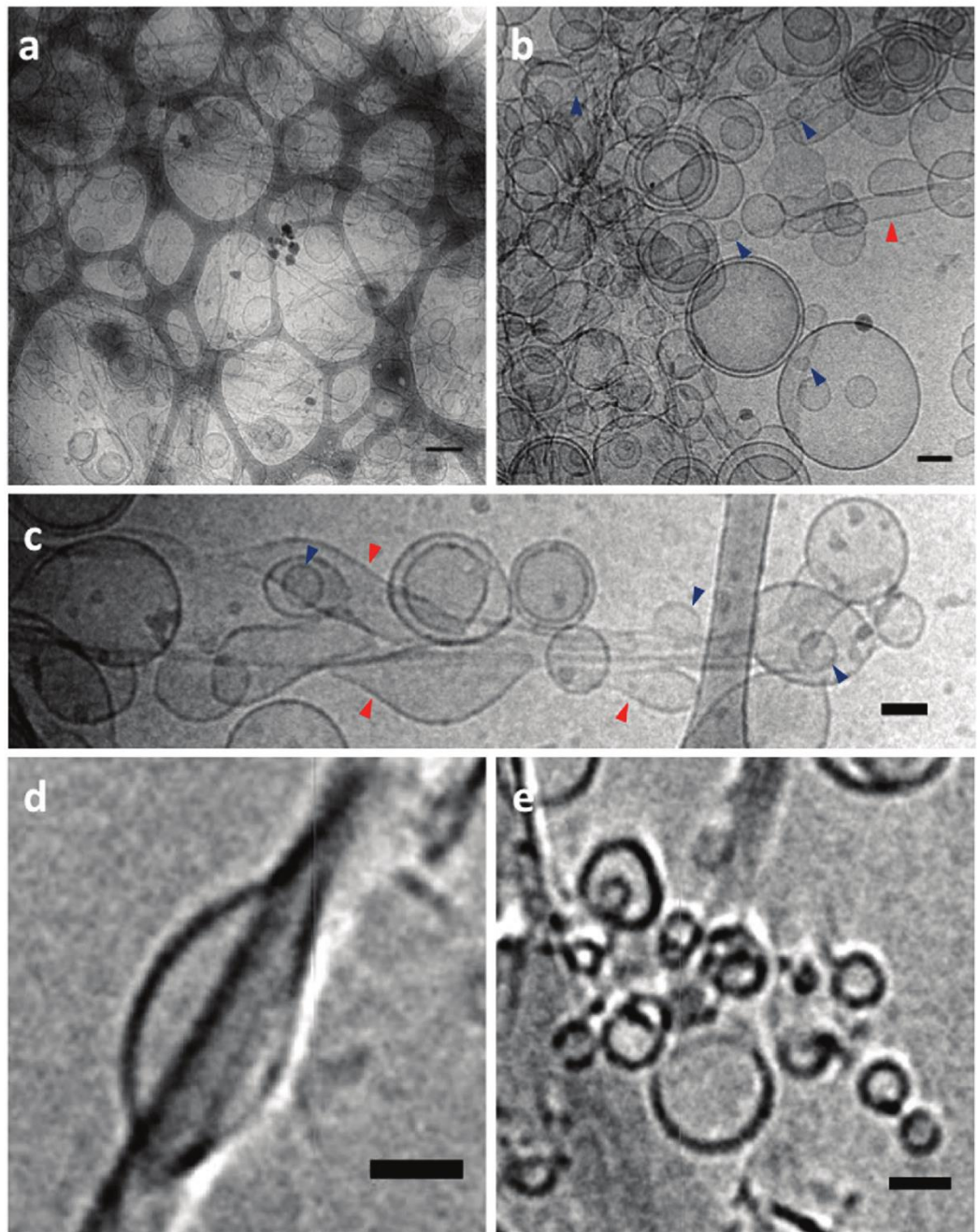
**Hfq-CTR<sub>11</sub> amyloid fibers trigger liposome breaking and membrane disruption.** The same experiments carried out previously with the native protein were repeated with Hfq-CTR<sub>11</sub>. In this case, fibrils can be clearly observed with Cryo-TEM (peptide concentration 440 μM, *i.e.* 0.5 g.L<sup>-1</sup>) (Fig. 5). Stacks of fibrils decorated with liposomes form along a specific direction, different for each cluster (Fig. 5a). Liposome deformations at the liposome-fibril contact regions are frequent (red arrows Fig. 5b and c). Furthermore, many small liposomes of 10–30 nm appear (blue arrows Fig. 5), resulting from the rupture of larger SUVs. Note that, although the cryo-TEM indicates an interaction between fibrils and liposomes, it is not possible to determine from individual projections whether overlap occurs in the same plane of the sample. Therefore, cryo-electron tomography has been used to reconstruct the 3D structure of the interaction between fibrils and liposomes (Fig. 5d and e). Tomography images directly confirm contacts between the two components. Furthermore, analysis of the liposome morphology also demonstrates that some of the liposomes are deformed by fibrils (Fig. 5d) and small 10–30 nm liposomes are also observed (Fig. 5e).

Then, the effect of Hfq-CTR<sub>11</sub> on EPE mica-supported lipid membranes was observed by AFM. Before protein incubation, they presented a roughness under 1 nm (Fig. 4g and j). After addition of Hfq-CTR<sub>11</sub>, small holes clearly appear (Fig. 4h) that grow with incubation time (Fig. 4i). The measured depth of these holes is 4 nm, which means that unlike the native protein, the Hfq-CTR<sub>11</sub> peptide removes the whole bilayer (Fig. 4k and l), a result compatible with the SUV deformation and splitting observed in TEM. Again, we assume that protein/lipid aggregates are likely present in the holes formed (Fig. 4i). But in this case no aggregated protein is found adsorbed on the intact membrane area.

**Interaction with membrane induces a conformational change in Hfq CTR.** Finally, to assess how the Hfq-CTR is interacting with SUVs, we performed Synchrotron Radiation Circular Dichroism (SRCD). The SRCD spectra were obtained for freshly dissolved Hfq-CTR<sub>38</sub> in solution and Hfq-CTR<sub>38</sub> incubated with SUVs (Fig. 6). The respective secondary structure compositions were determined using Bestsel and are reported in Table 1. As expected from our previous observations<sup>28</sup>, the peptide alone is partially unstructured or folded in β-sheets (38%). When incubated with EPE liposomes, the percentage of β-sheets decreases to 33%, while when incubated with DOPG liposomes, the percentage of β-sheets is reduced to 28% with a small increase of turns and apparition of 5.6% α-helices (classified as “distorted helix” category according to Bestsel). In the later case, it is likely that only a fraction of the peptide is interacting the membrane and that more than 5–6% helices is formed in the population bound to membrane. Note that the helices formed seem to be unstable without the interaction of the peptides with the polar phosphoglycerol lipid heads.

## Discussion

Our results demonstrate that Hfq has a propensity to interact with membranes. Importantly, the native protein used for all experiments here was exempt of post-translational modification such as lipid addition. This has been confirmed by mass spectrometry analysis (Fig. S5). Thus, such a modification is not needed for interaction of the protein with a bilayer membrane *in vitro*, even if the existence of this modification has been reported previously<sup>36</sup>.

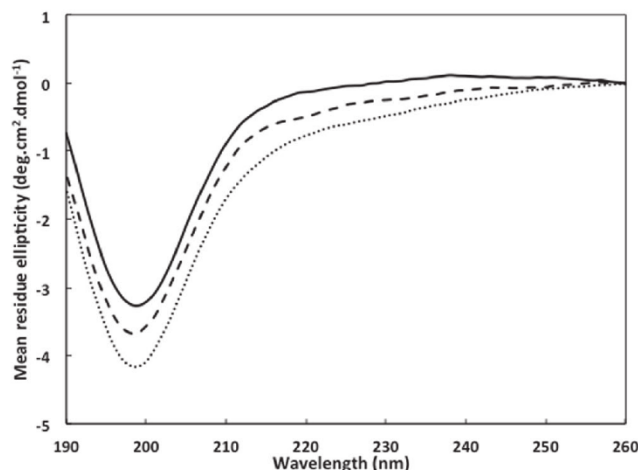


**Figure 5.** Cryo-TEM images and tomography of Hfq-CTR11 incubated with EPE SUVs. (a) Hfq-CTR<sub>11</sub> peptides aggregate to form amyloid fibrils and regroup large clusters of liposomes. (b) A high proportion of small liposomes of 10–30 nm (blue arrows) are present, possibly coming from SUVs disruption by Hfq-CTR<sub>11</sub>. (c) Multiple contacts between liposomes and fibers create SUV deformations (red arrows). (d) Liposome deformation as seen by cryo-tomography. (e) Small liposomes (10–30 nm) as seen by cryo-tomography. Scale bars: 200 nm for (a), 50 nm for (b,c) and 20 nm for (d,e).

For Hfq-membrane interaction, two distinct actions may be described. First, a relatively weak, reversible binding with the torus lying on the membrane is observed. The affinity of the native protein for polar lipids suggests that binding is mainly driven by hydrogen bonding and electrostatic interactions. The importance of weak electrostatic interactions between basic residues and acidic lipids could be considered as a mean for proteins to be transiently released in the cytoplasm<sup>37</sup> and could explain why Hfq was not identified in the membrane fraction in large-scale proteomic studies<sup>38</sup>.

The self-assembly of Hfq through its C-terminal tails at the bilayer surface also provides a possible explanation for the superstructures observed previously *in vivo* in close proximity of the inner membrane<sup>10,28,39</sup>. The C-terminal tails strengthen the protein-lipid interaction. As membrane association is lost when this CTR is





**Figure 6.** SRCD spectra of the Hfq-CTR<sub>38</sub> peptide interacting with liposomes. Hfq-CTR<sub>38</sub> alone (solid line), with EPE SUVs (dashed line) or with DOPG SUVs (dotted line). The respective contribution of buffer (10 mM sodium phosphate and 20 mM NaCl) and/or lipids were subtracted.

Secondary structure (%)	Hfq-CTR <sub>38</sub>	Hfq-CTR <sub>38</sub> + EPE	Hfq-CTR <sub>38</sub> + DOPG
α helix	0	0	5.6
Turn	14	15.8	17
Antiparallel β sheets	38.1	32.7	27.9
Others	47.9	51.4	49.4

**Table 1.** SRCD spectra analysis of Hfq-CTR<sub>38</sub> in the absence and in the presence of liposomes. Analysis was performed using the Bestsel software. The results are given as percentage.

deleted<sup>28</sup>, we assumed this part of the protein was essential for the interaction with membrane *in vivo*. Our present results with Hfq CTR region confirm this assumption. This result is also in agreement with previous works on the ability of fibrils including amyloid fibrils to interact with membranes, also in the case of peptide-forming fibers<sup>31,40,41</sup>.

The second action we have observed is a drastic change in the membrane organization with longer incubation times. This rearrangement is induced by the C-terminal amyloidogenic portion of the protein. AFM measurements suggest that the full-length protein remove one monolayer of the membrane, in agreement with the recent work of Pyne *et al.*<sup>42</sup>. Taking into account that monolayer poration has already been reported for other proteins<sup>42–44</sup>, one possibility is that similarly to Tilamin, Hfq-induced monolayer-pores may cause local membrane rupture<sup>42</sup>. Nevertheless, in the case of Tilamin, monolayer poration induces progressive removal of the outer membranes and bacterial death<sup>42</sup>. More likely in the case of Hfq, the protein induces transient pores *in vivo* and not membrane disintegration. We also confirm here the ability of amyloid regions to disrupt and distort membranes<sup>45</sup>.

Finally, the secondary structural change of Hfq-CTR<sub>38</sub> observed by SRCD suggests that the amyloid-like region could be partially converted into a α-helix, possibly a trans-membrane helix as reported earlier for other amyloids<sup>46</sup>. Another possibility would be that an amphipathic helix lying horizontally in the acidic phosphate-heads layer forms, such as that observed for RNase E or RNase II<sup>47,48</sup>. This would be in agreement with the flexibility of Hfq CTR and with its composition of both basic/polar (H, Q, N, T) and non-polar (A, Y) amino acid residues. Alternatively, a tilted helix may also form as in the case of Tilamin, resulting in its insertion in the monolayer<sup>42</sup>. These possibilities however need to be investigated further, for instance using orientated circular dichroism (OCD) to observe the orientation of the helix inside the membrane<sup>49</sup>.

In conclusion, the key finding of this work is that Hfq is able to affect biological membranes, as shown in AFM and cryo-TEM. This finding may support a possible mechanism by which Hfq could contribute to the export of RNA outside the bacteria or in the periplasmic space. To date, known exported molecules for quorum sensing were limited to small molecules. Nevertheless bacterial RNAs have been identified extracellularly<sup>50</sup>. Among those RNAs, small noncoding RNA are found. Here we propose a new possible role for Hfq in sRNA export, with important consequences for bacterial communication.

## Materials and Methods

**Protein expression and purification.** Wild type *Escherichia coli* Hfq was purified as described previously<sup>39</sup>. Hfq-NTR<sub>65</sub> (residues 1–65) was purified from over expressing BL21(DE3)Δhfq/pLATE11-hfqntr<sub>65</sub> cells. The pLATE11-hfqntr<sub>65</sub> expression vector was constructed according to the manufacturer protocol

(ThermoFisher). As the his-tag potentially modifies Hfq properties<sup>51</sup>, a sequence encoding the Tobacco Etch Virus (TEV) protease site (ENLYFQG) was inserted between the N-terminal His-tag and Hfq sequence. For the purification of Hfq-NTR<sub>65</sub>, cells from post-induction cultures (IPTG 1 mM, 3 h induction) were resuspended on ice in 20 mM Tris HCl pH 7.5 containing 0.5 M NaCl, 10% (v/v) glycerol and a protease inhibitor cocktail (Sigma). The cells were lysed by sonication and the lysate was cleared by centrifugation at 15,000 g for 30 min. DNase I (40 g. L<sup>-1</sup>) and RNase A (30 g.L<sup>-1</sup>) were added to the cleared lysate at room temperature. The solution was then applied to a His-trap column (GE Healthcare). The resin was washed with 20 mM Tris HCl pH 7.5 containing 0.3 M NaCl and 20 mM imidazole; the protein was eluted with a gradient of imidazole (20–500 mM) in the same buffer. The eluted protein was dialyzed in the appropriate buffer and TEV digestion was carried out according to the manufacturer's instructions (ThermoFisher). After digestion, Hfq-NTR<sub>65</sub> was subsequently purified with a HiTrap SP HP 1 mL column (GE Healthcare). Before injection on the cation exchange column, Hfq-NTR<sub>65</sub> was diluted five times in equilibration buffer (50 mM HEPES pH 8) to reduce salt concentration. The resin was equilibrated with the same buffer and the protein was eluted with a gradient of NaCl (0–1 M). The cleaved His-tag, non-digested Hfq-NTR<sub>65</sub>, and other remaining contaminants are eliminated with this cation exchange column.

Hfq C-terminus peptides were obtained from Proteogenix (France). Fibrils of synthetic peptides were prepared from dissolved peptide in deionized water at 0.5 g.L<sup>-1</sup>. Fibrils were usually observed after one week. Note that kinetics of assembly can be greatly affected by sample batch-to-batch variability. Several factors contribute to this variability, including the presence of salts in batches and the presence of various amounts of pre-formed aggregates in samples.

**Preparation of small unilamellar vesicles (SUVs).** Lipids purchased from Avanti Polar Lipids (Alabaster, AL, USA) were dissolved at 10 mg.mL<sup>-1</sup> in chloroform/methanol 1:1 (v/v) on ice. The lipids used were *E. coli* Polar Extract (abbreviated EPE, ref 100600P), 1,2-dioleoyl-sn-glycero-3-phosphocholine (abbreviated DOPC, ref 850375) and 1, 2-dioleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (abbreviated DOPG, ref 840475). The solvent was slowly evaporated with N<sub>2</sub>(g), forming a lipid film that was then dried for 30 min under N<sub>2</sub> flow. Lipids were hydrated in SUV buffer (10 mM Tris pH 7.5 containing 100 mM NaCl). Vesicles formed during slow stirring at room temperature over 30 min. The solution was then passed 31 times through a 0.2 μm polycarbonate filter (Avanti Mini Extruder) to obtain SUVs. Samples were kept at 4 °C and used for subsequent experiments within one week. For liposomes used in CD experiments, the protocol was identical but Tris-HCl was replaced by phosphate buffer (10 mM sodium phosphate buffer pH 7.5 with 100 mM NaCl) and using 0.1 μm polycarbonate filter for extrusion.

**Co-sedimentation assay.** Gradual concentrations of SUVs (100, 300 and 600 μM lipids) were mixed with proteins in SUV buffer with or without NaCl and left 20 min at 22 °C before being ultracentrifuged 1 h at 100 000 g at the same temperature in a Beckman TL-100 ultracentrifuge with a Beckman TLA100 rotor. 10% of total sample volume were recovered from the supernatants and pellets which were then used for loading on polyacrylamide gels for SDS-PAGE. Gels were stained with R-250 or G-250 Coomassie Brilliant Blue.

**AFM imaging in solution.** SUV suspension diluted at 0.4 g.L<sup>-1</sup> in SUV buffer supplemented with 2 mM CaCl<sub>2</sub> was incubated on a freshly-cleaved mica surface for one hour at 37 °C. The sample was then rinsed ten times with SUV buffer to remove excess SUVs. Supported bilayers were checked by AFM to confirm the extent of surface coverage and the quality of the lipid bilayer on the mica before protein incubation.

The proteins were added to the surface at a given concentration at room temperature for various times (from 5 min to several hours). Incubations were done in a sealed box to avoid evaporation. The sample was then rinsed ten times to remove unbound proteins and imaged in the same buffer. Protein concentrations were 2 μM for both native Hfq and Hfq-NTR<sub>65</sub>, and 176 μM for Hfq-CTR<sub>11</sub> peptide (1136 g/mol). Protein concentrations were chosen in order to have similar mass concentrations.

In this work, commercial Olympus rectangular silicon nitride cantilevers (RC800PSA) and two different commercial AFMs were used: an Agilent technologies 5500 microscope (Santa Clara, CA, USA) operated in the tapping mode at a resonance frequency of 15 kHz and 0.73 N.m<sup>-1</sup> spring constant cantilever and a Nanotec microscope (Madrid, Spain) operated in the jumping mode with an applied force < 100 pN<sup>52</sup>. After the sample was installed on the AFM stage, about 30 min were required for the system to reach thermal equilibrium before commencing the AFM scans. We imaged 5 to 10 positions at different resolutions to ensure consistency among the observations.

Before obtaining the height profiles from the AFM images, the tilt of the surface and the line to line noise were removed using the plane and flattening filters provided in the WSxM free software ([www.wsxmsolutions.com](http://www.wsxmsolutions.com))<sup>53</sup>. The height profiles were all obtained from line scans perpendicular to the slow scanning direction.

**Cryo-Electron microscopy (cryo-TEM).** SUVs in SUV buffer (0.2 g.L<sup>-1</sup> final concentration) and protein in the same buffer were mixed and left at room temperature for 20 min. When necessary, 5 nm Ni-NTA-NanoGold beads (NanoProbe, Yaphank, NY) were added 2 minutes before freezing at a final concentration of 4 nM in order to label His-tagged proteins. The high clustering observed in cryo-TEM images originates from the fact that Ni-NTA gold beads possess multiple Ni-NTA moieties that can bind several liposomes together. Protein concentrations were 10 μM for both Hfq-WT and Hfq-NTR<sub>65</sub>, and 440 μM for the Hfq-CTR<sub>11</sub> peptide. Protein concentrations were chosen in order to have similar mass concentrations.

5 μL of samples were then deposited on air plasma-cleaned EM grids (Lacey carbon films on 300 mesh copper grids). The excess was blotted with a filter paper and the grid was immediately plunged into a liquid ethane bath cooled with liquid nitrogen. Images were recorded on a Gatan Ultrascan camera using a 2200FS transmission



electron microscope from JEOL Ltd operated at 200 kV. High magnification images were taken over large sections at random locations, in order to avoid statistical bias. Tomographic series were taken from regions of interest covering an angular range from  $-60^\circ$  to  $+60^\circ$ . Statistics on membrane thickness were calculated from  $n = 20$  pictures.

Images were processed with ImageJ software<sup>54</sup> and tomogram reconstructions were calculated by the OS-SART method using the TomoJ plugin<sup>55</sup>. Stitching and quantitative analysis of the images were also carried out in ImageJ.

**Synchrotron Radiation Circular Dichroism (SRCD).** Measurements and data collection were carried out on DISCO beam-line at the SOLEIL Synchrotron (Gif-sur-Yvette, France)<sup>56</sup>. 0.25 mM (UV-absorption) Hfq-CTR<sub>38</sub> was mixed with 1 mM of EPE or DOPG SUVs in 10 mM sodium phosphate, 20 mM NaCl, pH 7.5. After 20 min incubation, 2–4  $\mu$ L of samples were loaded into circular demountable CaF<sub>2</sub> cells of 30 microns' path length<sup>57</sup>. Three separated data collections with fresh sample preparations were carried out to ensure consistency and repeatability. Spectral acquisitions of 1 nm steps at 1.2 integration time, between 260 and 180 nm were performed in triplicate for the samples as well as for the baselines (consisting of buffer alone or buffer + SUVs). (+)-camphor-10-sulfonic acid (CSA) was used to calibrate amplitudes and wavelength positions of the SRCD experiment. Data-analyses including averaging, baseline subtraction, smoothing, scaling and standardization were carried out with CDtool<sup>58</sup>. Secondary structure content was determined using BestSel<sup>59</sup>. Normalized root-mean-square deviation (NRMSD) indicated the most accurate fit for each spectrum; values of  $<0.15$  were considered significant.

**Data availability.** All data generated or analyzed during this study are included within this article and the Supplementary information or are available from the corresponding author upon request.

## References

- Shapiro, L., Franze de Fernandez, M. T. & August, J. T. Resolution of two factors required in the Q-beta-RNA polymerase reaction. *Nature* **220**, 478–480 (1968).
- Gottesman, S. & Storz, G. RNA reflections: converging on Hfq. *RNA* **21**, 511–512 (2015).
- Arluisson, V. & Taghbalout, A. Cellular localization of RNA degradation and processing components in Escherichia coli. *Methods Mol Biol* **1259**, 87–101 (2015).
- De Lay, N., Schu, D. J. & Gottesman, S. Bacterial small RNA-based negative regulation: Hfq and its accomplices. *J Biol Chem* **288**, 7996–8003 (2013).
- Zambrano, N. *et al.* Involvement of Hfq protein in the post-transcriptional regulation of E. coli bacterial cytoskeleton and cell division proteins. *Cell Cycle* **8**, 2470–2472 (2009).
- Gripenland, J. *et al.* RNAs: regulators of bacterial virulence. *Nat Rev Microbiol* **8**, 857–866 (2010).
- Azam, T. A. & Ishihama, A. Twelve species of the nucleoid-associated protein from Escherichia coli. Sequence recognition specificity and DNA binding affinity. *J Biol Chem* **274**, 33105–33113 (1999).
- Cech, G. M. *et al.* The Escherichia Coli Hfq Protein: An Unattended DNA-Transactions Regulator. *Front Mol Biosci* **3**, 36, doi:10.3389/fmolb.2016.00036 (2016).
- Azam, T. A., Hiraga, S. & Ishihama, A. Two types of localization of the DNA-binding proteins within the Escherichia coli nucleoid. *Genes Cells* **5**, 613–626 (2000).
- Diestra, E., Cayrol, B., Arluisson, V. & Risco, C. Cellular electron microscopy imaging reveals the localization of the Hfq protein close to the bacterial membrane. *PLoS One* **4**, e8301, doi:10.1371/journal.pone.0008301 (2009).
- Guillier, M. & Gottesman, S. Remodelling of the Escherichia coli outer membrane by two small regulatory RNAs. *Mol Microbiol* **59**, 231–247 (2006).
- Mura, C., Randolph, P. S., Patterson, J. & Cozen, A. E. Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function. *RNA Biol* **10**, 636–651 (2013).
- Wilusz, C. J. & Wilusz, J. Eukaryotic Lsm proteins: lessons from bacteria. *Nat Struct Mol Biol* **12**, 1031–1036 (2005).
- Brennan, R. G. & Link, T. M. Hfq structure, function and ligand binding. *Curr Opin Microbiol* **10**, 125–133 (2007).
- Link, T. M., Valentin-Hansen, P. & Brennan, R. G. Structure of Escherichia coli Hfq bound to polyribadenylate RNA. *Proc Natl Acad Sci USA* **106**, 19292–19297 (2009).
- Updegrave, T. B., Correia, J. J., Galletto, R., Bujalowski, W. & Wartell, R. M. E. coli DNA associated with isolated Hfq interacts with Hfq's distal surface and C-terminal domain. *Biochim Biophys Acta* **1799**, 588–596 (2010).
- Sauer, E. Structure and RNA-binding properties of the bacterial Lsm protein Hfq. *RNA Biol* **10**, 610–618 (2013).
- Jiang, K. *et al.* Effects of Hfq on the conformation and compaction of DNA. *Nucleic Acids Res* **43**, 4332–4341 (2015).
- Malabirade, A. *et al.* Compaction and condensation of DNA mediated by the C-terminal domain of Hfq. *Nucleic Acids Res*. doi:10.1093/nar/gkv268 In press (2017).
- Schumacher, M. A., Pearson, R. F., Moller, T., Valentin-Hansen, P. & Brennan, R. G. Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *Embo J* **21**, 3546–3556 (2002).
- Sauter, C., Basquin, J. & Suck, D. Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from Escherichia coli. *Nucleic Acids Res* **31**, 4091–4098 (2003).
- Nikulin, A. *et al.* Structure of Pseudomonas aeruginosa Hfq protein. *Acta Crystallogr D Biol Crystallogr* **61**, 141–146 (2005).
- Sauer, E. & Weichenrieder, O. Structural basis for RNA 3'-end recognition by Hfq. *Proc Natl Acad Sci USA* **108**, 13065–13070 (2011).
- Stanek, K. A., Patterson-West, J., Randolph, P. S. & Mura, C. Crystal structure and RNA-binding properties of an Hfq homolog from the deep-branching Aquificae: conservation of the lateral RNA-binding mode. *Acta Crystallogr D Struct Biol* **73**, 294–315 (2017).
- Vincent, H. A. *et al.* The low-resolution solution structure of Vibrio cholerae Hfq in complex with Qrr1 sRNA. *Nucleic Acids Res* **40**, 8698–8710 (2012).
- Beich-Frandsen, M. *et al.* Structural insights into the dynamics and function of the C-terminus of the E. coli RNA chaperone Hfq. *Nucleic Acids Res* (2011).
- Arluisson, V. *et al.* The C-terminal domain of Escherichia coli Hfq increases the stability of the hexamer. *Eur J Biochem* **271**, 1258–1265 (2004).
- Fortas, E. *et al.* New insight into the structure and function of Hfq C-terminus. *Biosci Rep* **35** (2015).
- Mura, C., Phillips, M., Kozhukhovsky, A. & Eisenberg, D. Structure and assembly of an augmented Sm-like archaeal protein 14-mer. *Proc Natl Acad Sci USA* **100**, 4539–4544 (2003).
- Arluisson, V. *et al.* Three-dimensional Structures of Fibrillar Sm Proteins: Hfq and Other Sm-like Proteins. *J Mol Biol* **356**, 86–96 (2006).
- Herrera, A. I., Tomich, J. M. & Prakash, O. Membrane Interacting Peptides: A Review. *Curr Protein Pept Sci* **17**, 827–841 (2016).

32. Mitra, K., Ubarretxena-Belandia, I., Taguchi, T., Warren, G. & Engelman, D. M. Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. *Proc Natl Acad Sci USA* **101**, 4083–4088 (2004).
33. Epanand, R. M. & Epanand, R. F. Domains in bacterial membranes and the action of antimicrobial agents. *Mol Biosyst* **5**, 580–587 (2009).
34. Maurer-Stroh, S. *et al.* Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Methods* **7**, 237–242 (2010).
35. Calero, M. & Gasset, M. Featuring amyloids with Fourier transform infrared and circular dichroism spectroscopies. *Methods Mol Biol* **849**, 53–68 (2012).
36. Obregon, K. A., Hoch, C. T. & Sukhodolets, M. V. Sm-like protein Hfq: Composition of the native complex, modifications, and interactions. *Biochim Biophys Acta* **1854**, 950–966 (2015).
37. Khan, H. M. *et al.* A Role for Weak Electrostatic Interactions in Peripheral Membrane Protein Binding. *Biophys J* **110**, 1367–1378 (2016).
38. Papanastasiou, M. *et al.* The Escherichia coli peripheral inner membrane proteome. *Mol Cell Proteomics* **12**, 599–610 (2013).
39. Taghbalout, A., Yang, Q. & Arluison, V. The Escherichia coli RNA processing and degradation machinery is compartmentalized within an organized cellular network. *Biochem J* **458**, 11–22 (2014).
40. Chu, H. *et al.* Human alpha-defensin 6 promotes mucosal innate immunity through self-assembled peptide nanonets. *Science* **337**, 477–481 (2012).
41. Faruqi, N. *et al.* Differentially instructive extracellular protein micro-nets. *J Am Chem Soc* **136**, 7889–7898 (2014).
42. Pyne, A. *et al.* Engineering monolayer poration for rapid exfoliation of microbial membranes. *Chem Sci* **8**, 1105–1115 (2017).
43. Haubertin, D. Y., Madaoui, H., Sanson, A., Guerois, R. & Orłowski, S. Molecular dynamics simulations of E. coli MsbA transmembrane domain: formation of a semipore structure. *Biophys J* **91**, 2517–2531 (2006).
44. Negrete, H. O., Rivers, R. L., Goughs, A. H., Colombini, M. & Zeidel, M. L. Individual leaflets of a membrane bilayer can independently regulate permeability. *J Biol Chem* **271**, 11627–11630 (1996).
45. Milanesi, L. *et al.* Direct three-dimensional visualization of membrane disruption by amyloid fibrils. *Proc Natl Acad Sci USA* **109**, 20455–20460 (2012).
46. Stroobants, K. *et al.* Amyloid-like fibrils from an alpha-helical transmembrane protein. *Biochemistry*. doi:10.1021/acs.biochem.7b00157 (2017).
47. Khemici, V., Poljak, L., Luisi, B. F. & Carpousis, A. J. The RNase E of Escherichia coli is a membrane-binding protein. *Mol Microbiol* **70**, 799–813 (2008).
48. Lu, F. & Taghbalout, A. Membrane association via an amino-terminal amphipathic helix is required for the cellular organization and function of RNase II. *J Biol Chem* **288**, 7241–7251 (2013).
49. Wu, Y., Huang, H. W. & Olah, G. A. Method of oriented circular dichroism. *Biophys J* **57**, 797–806 (1990).
50. Ghosal, A. *et al.* The extracellular RNA complement of Escherichia coli. *Microbiologyopen*. doi:10.1002/mbo3.235 (2015).
51. Jackson, T. C. & Sukhodolets, M. V. Functional analyses of putative Pals (Palindromic Self-recognition) motifs in bacterial Hfq. *Biochem Biophys Res Commun*. doi:10.1016/j.bbrc.2017.03.160 (2017).
52. de Pablo, P. J., Colchero, J., Gomez-Herrero, J. & Baro, A. Jumping mode scanning force microscopy. *Appl. Phys. Lett.* **73**, 3300–3302 (1998).
53. Horcas, I. *et al.* WSXM: A software for scanning probe microscopy and a tool for nanotechnology. *Review of scientific instruments* **78**, 013705 (2007).
54. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675 (2012).
55. Sorzano, C. O. *et al.* Marker-free image registration of electron tomography tilt-series. *BMC Bioinformatics* **10**, 124. doi:10.1186/1471-2105-10-124 (2009).
56. Refregiers, M. *et al.* DISCO synchrotron-radiation circular-dichroism endstation at SOLEIL. *J Synchrotron Radiat* **19**, 831–835 (2012).
57. Wien, F. & Wallace, B. A. Calcium fluoride micro cells for synchrotron radiation circular dichroism spectroscopy. *Appl Spectrosc* **59**, 1109–1113 (2005).
58. Lees, J. G., Smith, B. R., Wien, F., Miles, A. J. & Wallace, B. A. CDtool—an integrated software package for circular dichroism spectroscopic data processing, analysis, and archiving. *Anal Biochem* **332**, 285–289 (2004).
59. Micsonai, A. *et al.* Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc Natl Acad Sci USA* **112**, E3095–3103 (2015).
60. Carmichael, G. G., Weber, K., Niveleau, A. & Wahba, A. J. The host factor required for RNA phage Q $\beta$  RNA replication *in vitro*. *J Biol Chem* **250**, 3607–3612 (1975).
61. Rabhi, M. *et al.* The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *EMBO J* **30**, 2805–2816 (2011).

## Acknowledgements

We would like to thank PICT-Ibisa for providing access to JEOL 2200FS TEM. SRCD measurements on DISCO beamline at the SOLEIL Synchrotron were performed under proposal #20161398. This work was supported by CNRS and CEA (VA), Université Paris Saclay (AM). We gratefully acknowledge help to MV from the French Embassy for their program for scientific and university cooperation. We are indebted to F. Gobeaux (CEA Saclay, Gif-sur-Yvette, France), Liliane Mouawad (Institut Curie, Orsay, France), Marc Lemaire and Jose-Luis Vasquez-Ibar for many fruitful discussions and to Jean-Pierre Lecaer (ICSN, CNRS, Gif-sur-Yvette, France) for mass spectrometry analysis. We thank Kimberly Stanek (University of Virginia) for her careful and critical reading of our manuscript.

## Author Contributions

A.M. and V.A. conceived the original idea and designed the experiments. A.M. purified proteins. A.M. and J.S. performed co-sedimentation. A.M., J.M.B. and I.M. performed A.F.M. experiments. A.M., J.M.B., S.T. and S.M. performed T.E.M. experiments. A.M., V.A. and F.W. performed SRCD experiments. A.M. and V.A. wrote the manuscript. All authors analyzed results and commented on the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-11157-5

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

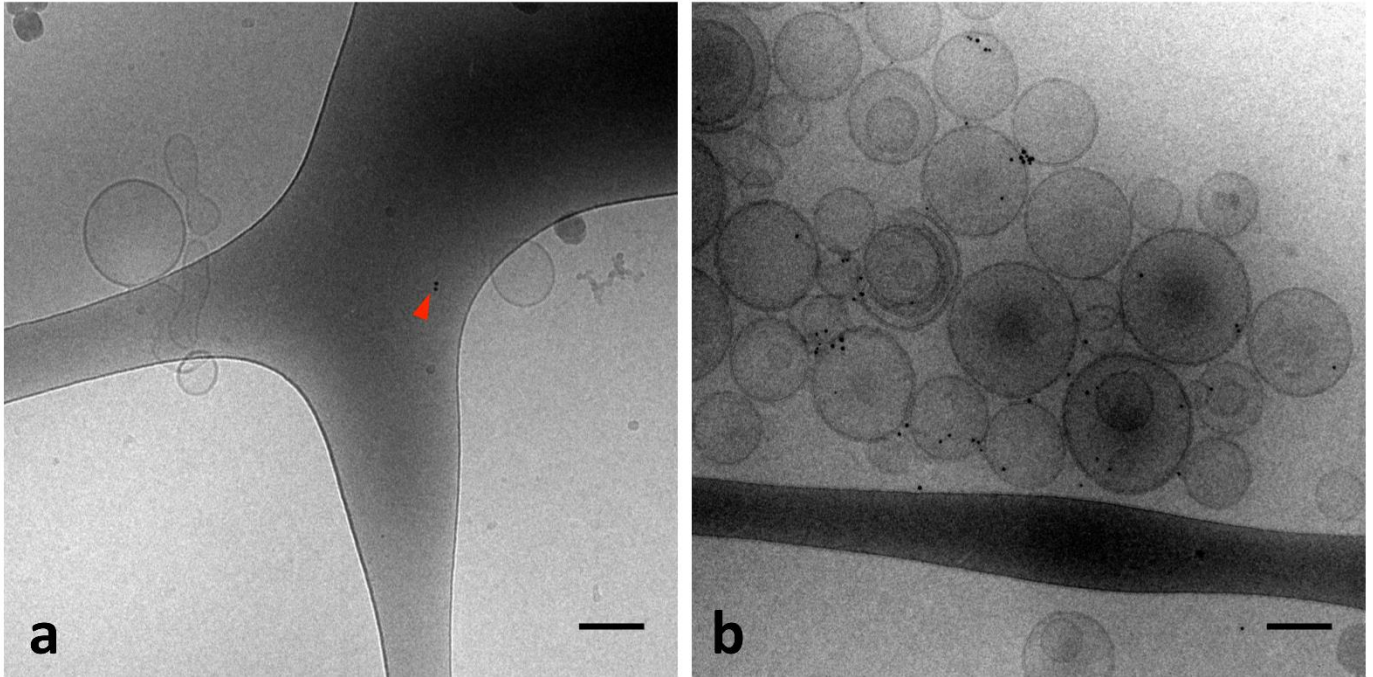
© The Author(s) 2017

# **Supplementary material**

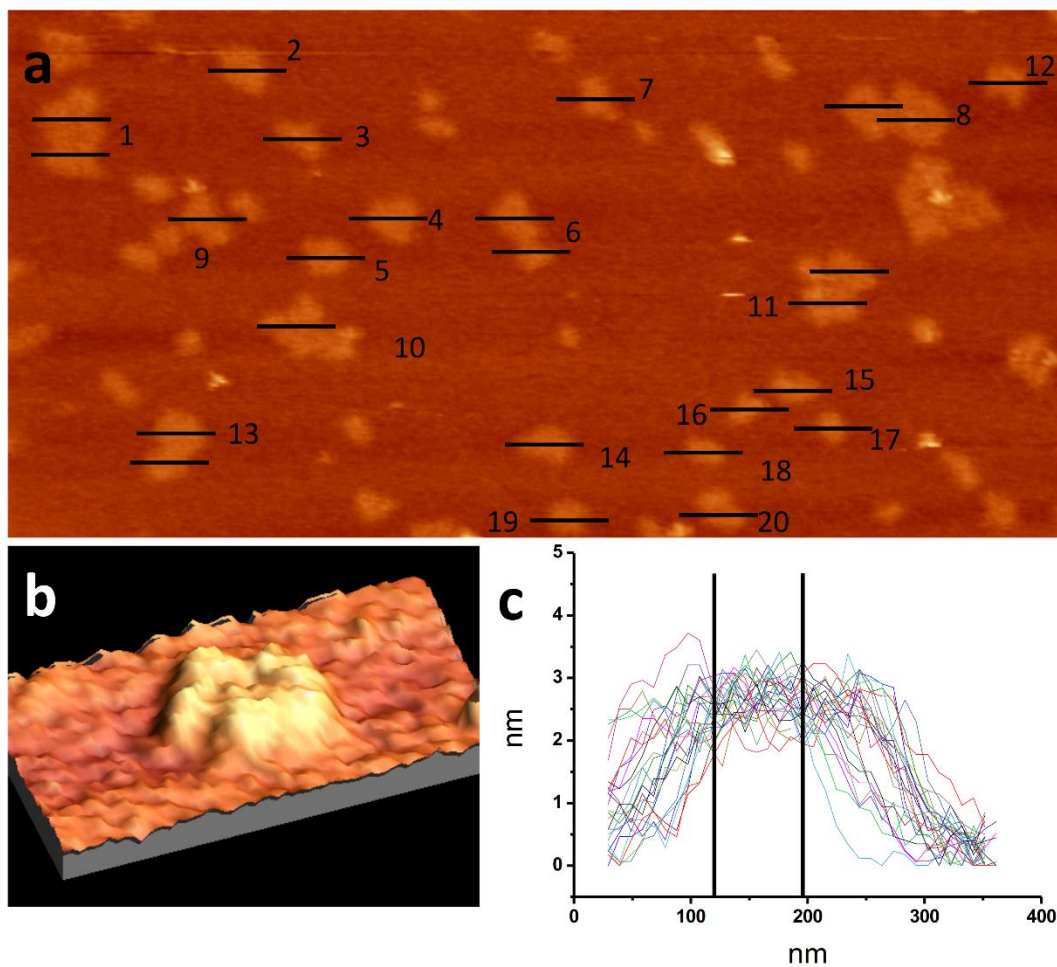
## **Membrane association of the bacterial riboregulator Hfq and functional perspectives**

Antoine Malabirade, Javier Morgado-Brajones, Sylvain Trépout, Frank Wien, Ileana Marquez, Jérôme Seguin, Sergio Marco, Marisela Velez and Véronique Arluison

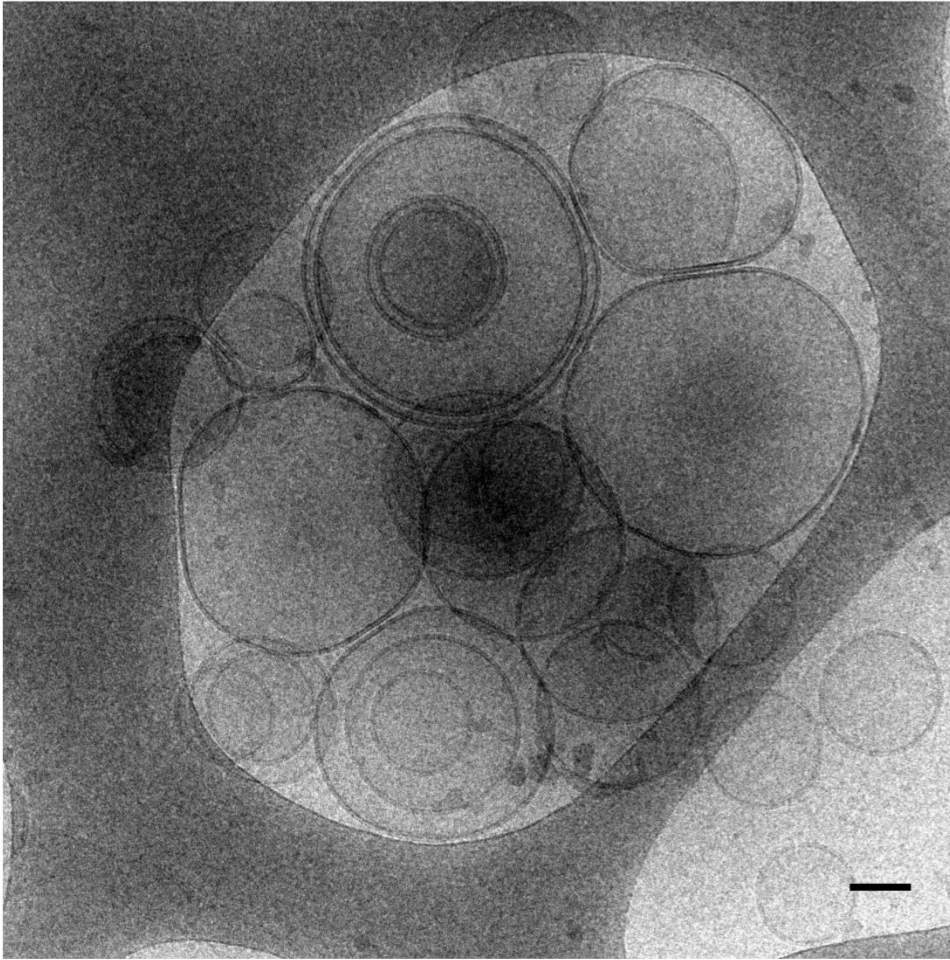




**Fig. S1:** *a.* EPE SUVs and 4nM NiNTA-NanoGold beads. The few beads found do not bind lipidic membranes and prefer to stick to the carbon film (red arrow). *b.* EPE SUVs incubated with wild-type Hfq and 4nM NiNTA-NanoGold beads. The four Histidines naturally present in the C-terminal tails of Hfq are sufficient to bind some beads, indicating that His-tag labelling does not significantly influence the protein-membrane interaction. Scale bars 100 nm.

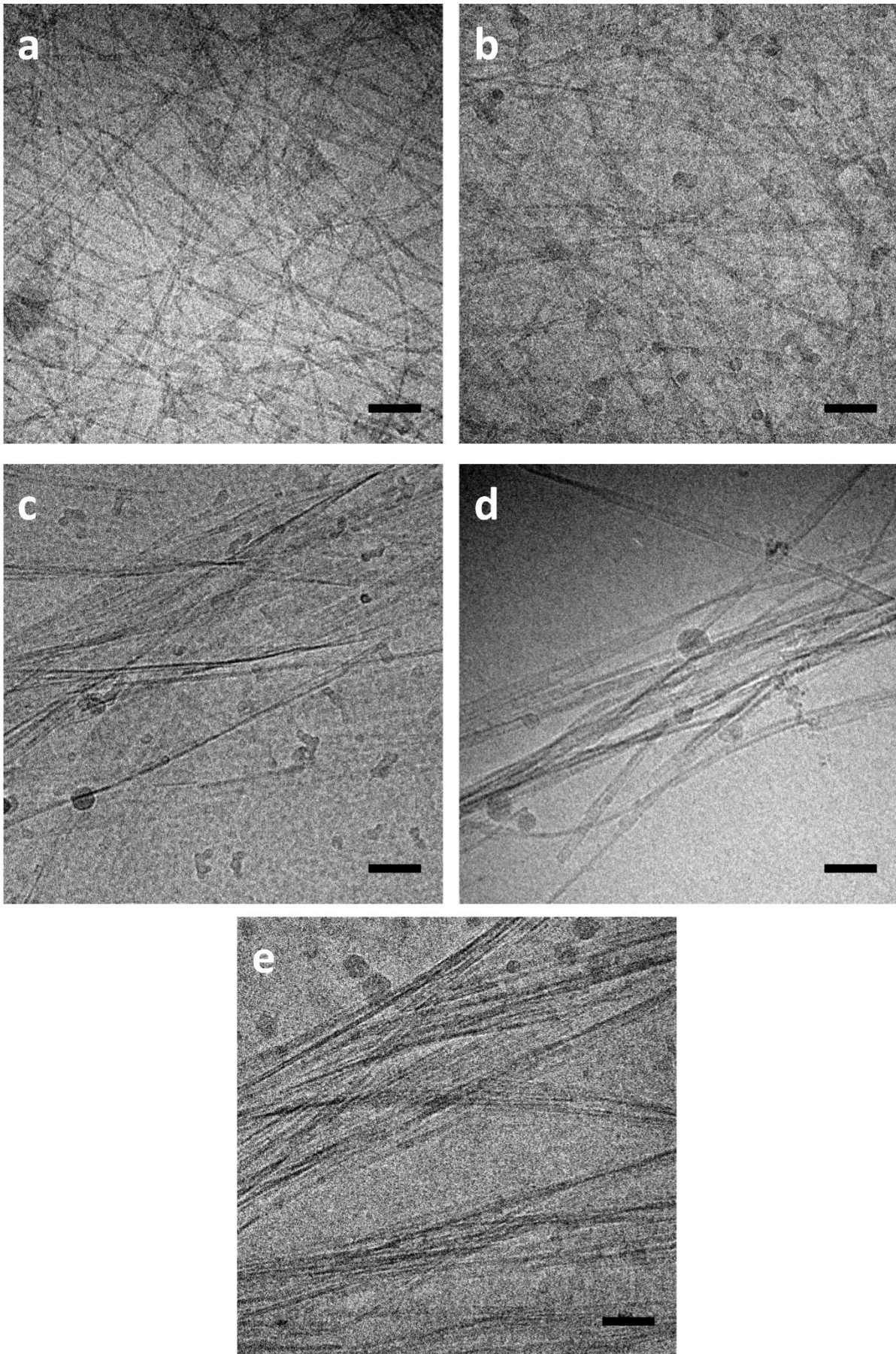


**Fig. S2: Height measurement of Hfq aggregates on an EPE lipid bilayer by liquid AFM.** **a.** Aggregates and profiles used to obtain the average height. **b.** The 3D representation of one aggregate shows that the surface is not flat. **c.** Compilation of all the profiles used. We selected the highest profile of each aggregate, which is shown as a black line in panel a, and the plateau region that was selected to average is indicated between two lines in panel c. The average height measured from 20 aggregates is  $2.7 \pm 0.2$  nm.

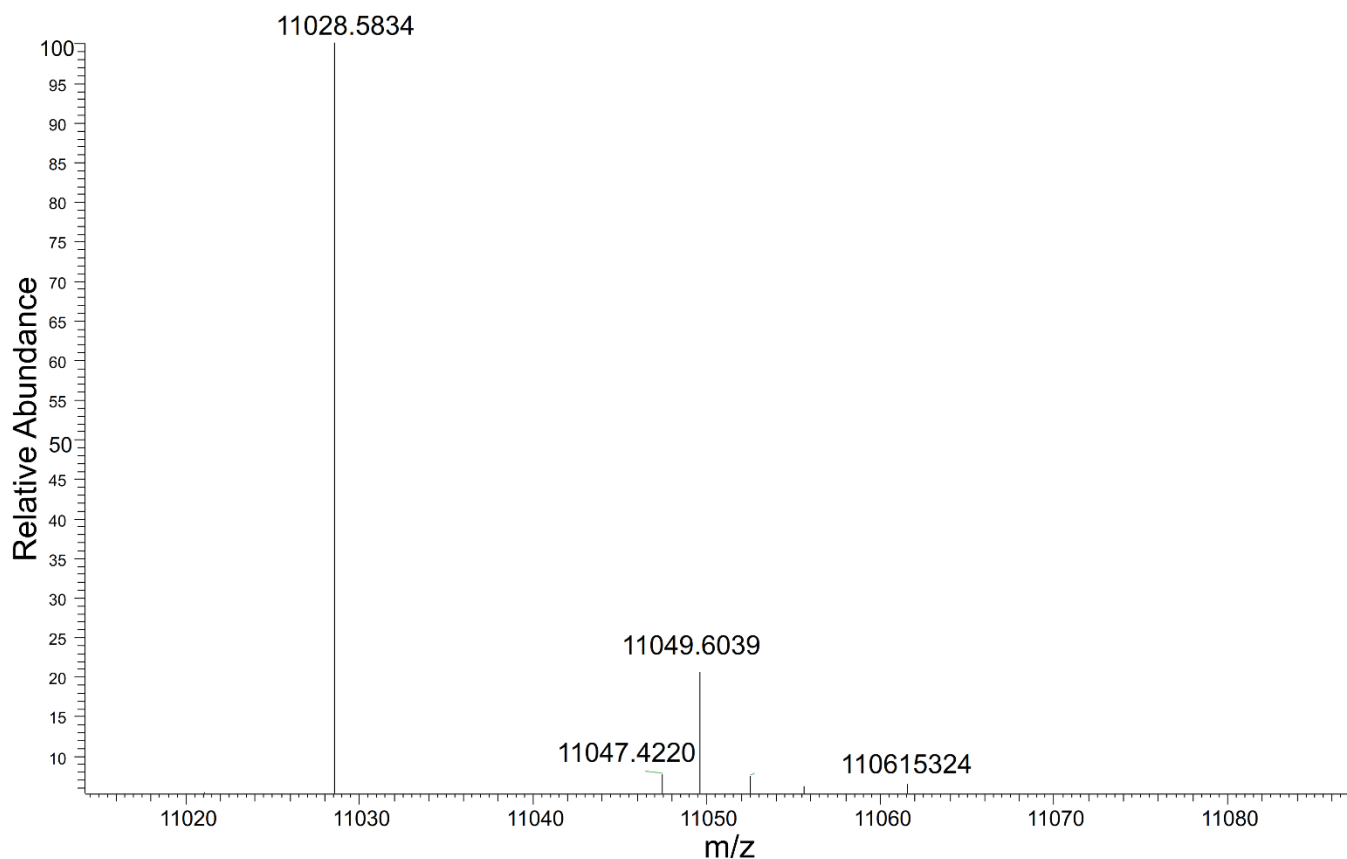


**Fig. S3: Experiment with Hfq-NTR<sub>65</sub>.** Cryo-TEM image showing that the truncated torus is able to bind and regroup some liposomes but without visible effect on the membrane thickness. Scale bar: 50 nm.



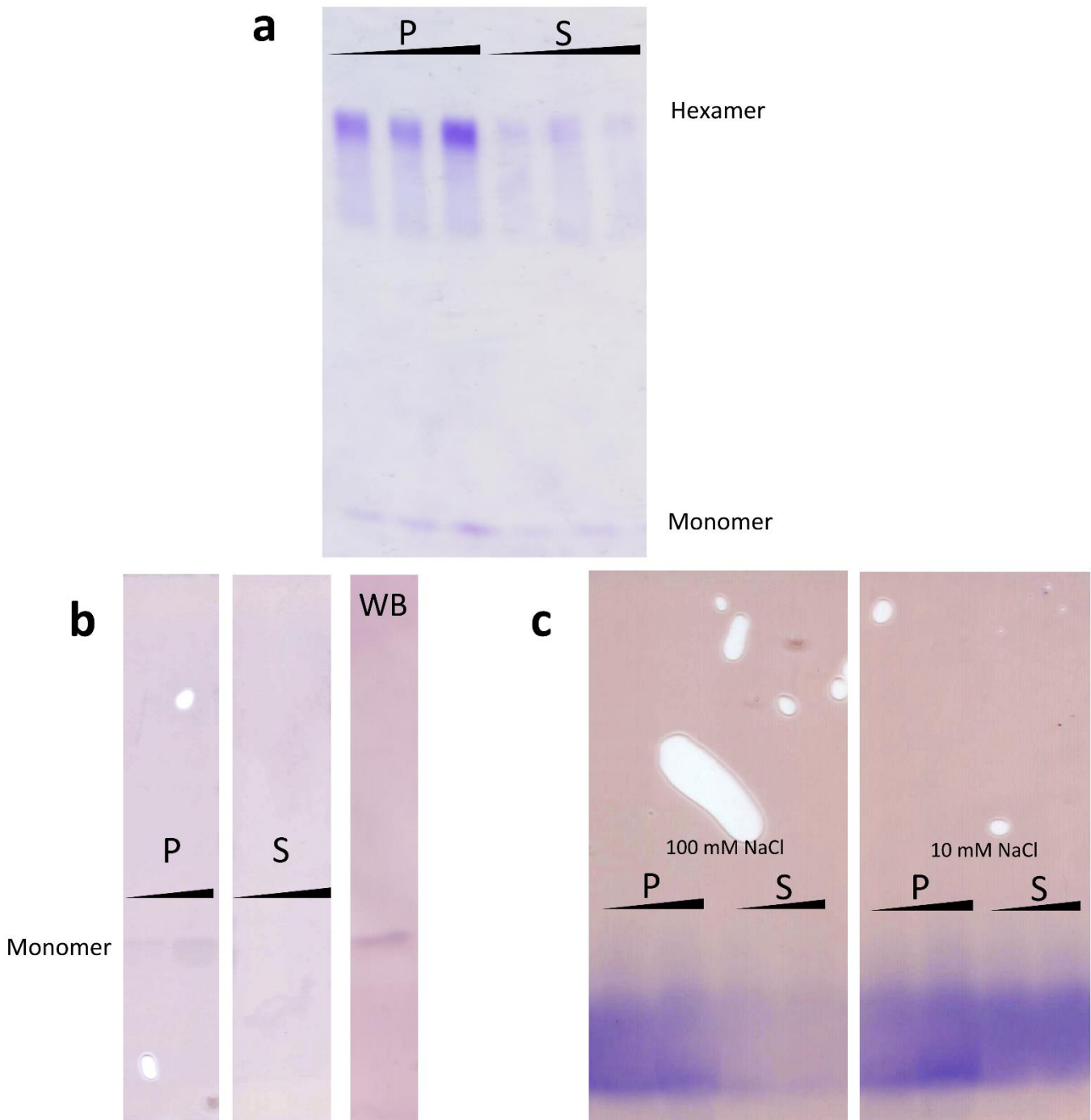


**Fig. S4: Hfq-CTR peptides successfully tested for amyloid fibers formation by Cryo-TEM. *a-e*:** Hfq-CTR<sub>38</sub>, Hfq-CTR<sub>26</sub>, Hfq-CTR<sub>17a</sub>, Hfq-CTR<sub>17b</sub>, Hfq-CTR<sub>11</sub> respectively. Concentration 0.5 g.L<sup>-1</sup>, scale bar 50 nm.



**Fig. S5: Mono-isotopic deconvoluted spectra of ESI/Orbitrap Mass-Spectrometry on purified native Hfq.** The majoritary pic at 11028.58 Da corresponds to a monomer of native Hfq (in accordance with Expasy Compute pI/Mw: 11 028,54 Da). The 11049.60 Da pic probably corresponds to Hfq+Na<sup>+</sup>. The protein solution was exchanged on a C4-reverse phase Ziptip (Merck Millipore) with 60% acetonitrile in water before electrospray.





**Fig. S6: Full SDS-PAGE for Hfq co-sedimentation assay (EPE).** **a.** Wild-type Hfq with increasing concentrations of EPE (100, 300, 600 $\mu$ M); gel percentage 8-16%. **b.** Hfq-NTR<sub>65</sub> (EPE 300, 600 $\mu$ M); gel percentage 4-20%. **c.** Hfq-CTR38 peptides (EPE 300, 600 $\mu$ M) in the presence of 100 and 10 mM NaCl; gel percentage 12%. P, Pellet fraction; S, Supernatant fraction. As previously described, WT Hfq hexamers are only partially denatured by SDS-PAGE and migrate mostly as hexamers, while truncated Hfq-NTR<sub>65</sub> is less stable and can be dissociated on the PAGE<sup>25,60,61</sup>. In **b**, a Western Blot (WB) is also provided to confirm the nature of Hfq-NTR<sub>65</sub> (rat anti-Hfq IgG).



**Sup Table S1: Hfq CTR peptides tested for amyloidogenic assembly**

Region	Amino acid sequence	Amyloid fibers
Hfq Wild	MAKGQSLQDPFLNALRRERVPVSIYLVNGIKLQGQIESFD	
Type	QFVILLKNTVSQMVKHAISTVVPSRPVSHHSNNAGGGT SSNYHHGSSAQNTSAQQDSEETE	
Hfq-CTR <sub>38</sub>	SRPVSHHSNNAGGGTSSNYHHGSSAQNTSAQQDSEETE	<b>Yes</b>
-	SRPVSHHSNNAGGGT	No
Hfq-CTR <sub>26</sub>	GGTSSNYHHGSSAQNTSAQQDSEETE	<b>Yes</b>
-	VSHHSNNAGGGTSSNYHHGS	No
-	GTSSNYHHGSSAQNTSA	No
Hfq-CTR <sub>17a</sub>	GSSAQNTSAQQDSEETE	<b>Yes</b>
Hfq-CTR <sub>17b</sub>	NYHHGSSAQNTSAQQDS	<b>Yes</b>
-	NTSAQQDSEETE	No
-	SAQQDSEETE	No
Hfq-CTR <sub>11</sub>	SAQNTSAQQDS	<b>Yes</b>

**Titre :** Auto-assemblage de la protéine bactérienne Hfq, actrice du métabolisme de l'ARN : rôle structural du domaine C-terminal.

**Mots clés :** Métabolisme de l'ARN, fibres amyloïdes, compaction de l'ADN, protéine membranaire.

**Résumé :** La régulation de l'expression génique par des ARNs permet une réponse rapide et polyvalente des cellules à des changements environnementaux. Cependant, elle nécessite souvent des partenaires protéiques. La protéine bactérienne Hfq en est un bon exemple. Facteur de virulence, elle est présente chez une variété de procaryotes et intervient dans nombre de circuits de régulation. Structurellement, Hfq adopte un repliement caractéristique, le repliement Sm. Ainsi, Les feuilletts  $\beta$  qui la constituent se regroupent et forment un hexamère toroïdal. Outre cette région N-terminale, il existe aussi parfois une région C-terminale (CTR) de séquence et de longueur variables. Chez *E. coli*, cette région comprend une trentaine de résidus et est prédite comme non-structurée. Jusqu'à présent, son rôle n'a été que peu étudié.

Ce travail de thèse met en lumière de nouvelles pistes quant à la fonction du CTR. Nous avons constaté sa capacité à former des fibres amyloïdes, expliquant la formation de structures auto-assemblées *in vivo*. De plus, la protéine est capable de lier l'ADN et de le condenser fortement *in vitro*. Cette compaction est complètement dépendante de la présence du CTR, qui permet de ponter les brins d'ADN. Ce résultat suggère une nouvelle fonction de Hfq dans la structuration du chromosome. Enfin, nous avons démontré que ce domaine permet aussi à Hfq de s'assembler à la surface d'une bicouche lipidique, expliquant sa localisation membranaire. La désorganisation de la membrane qui en résulte pourrait permettre le passage d'ARNs dans le milieu extracellulaire, avec d'importantes implications sur la capacité de la bactérie à interagir avec ses voisines et son environnement.

**Title :** Self-assembly of the bacterial protein Hfq, an actor of RNA metabolism : structural role of the C-terminal domain.

**Keywords :** RNA metabolism, amyloid fibers, DNA compaction, membrane protein.

**Abstract:** RNA-based regulations of gene expression allow quick and versatile responses from cells to changing environmental conditions. However, these regulations are often protein-mediated. The bacterial protein Hfq is one of the most studied RNA-based regulation partner. Found in various prokaryotes, it is an important virulence factor involved in many cellular processes. Hfq's structure resembles a torus, formed by multiple  $\beta$ -sheets. Apart from this N-terminal region (NTR), a supplemental C-terminal region (CTR) with variable lengths and sequences may exist in some species. In *E. coli*, this specific region measures around 30 residues and is predicted as intrinsically disordered. Few studies focused on Hfq-CTR until recently.

This work highlights new potential roles for Hfq-CTR. First, this region is able to self-interact and forms amyloid fibers which explains the self-assembled Hfq superstructures observed *in vivo*. Second, the protein can bind and efficiently condense DNA *in vitro*, strengthening the suggested role of Hfq in shaping the bacterial chromosome. This compaction is fully dependent on the CTR which is responsible for DNA bridging. Third, the CTR also gives to Hfq the ability to self-assemble on a lipid bilayer, explaining its membrane-localized fraction observed *in vivo*. The subsequent membrane reorganization might facilitate the release of RNAs in the extracellular medium, with potential implications on bacterial communication and interaction with surrounding cells and environment.

