

ÉCOLE DOCTORALE SANTE, SCIENCES BIOLOGIQUES ET CHIMIE DU VIVANT
CENTRE DE BIOPHYSIQUE MOLÉCULAIRE, CNRS Orléans

THÈSE présentée par :
Justine LARGILLIERE

soutenue le : **10 juillet 2020**

pour obtenir le grade de : **Docteur de l'Université d'Orléans**
Discipline/ Spécialité : Chimie organique, minérale, industrielle

**Architecture moléculaire et dynamique
de protéines histone-like de bactérie et
d'archée**

THÈSE dirigée par :
Céline LANDON

CR, CBM, Orléans

Et co-encadrée par :
Karine LOTH

MCU, Université d'Orléans, Orléans

RAPPORTEURS :
Jean-Pierre SIMORRE
Ewen LESCOP

DR, IBS, Grenoble
CR, ICSN, Gif-sur-Yvette

JURY :
Marc BOUDVILLAIN
Isabelle LANDRIEU
Céline LANDON
Karine LOTH
Ewen LESCOP
Jean-Pierre SIMORRE
Bertrand CASTAING

DR, CBM, Orléans, Président du jury
DR, UGSF, Lille
CR, CBM, Orléans
MCU, Université d'Orléans, Orléans
CR, ICSN, Gif-sur-Yvette
DR, IBS, Grenoble
DR, CBM, Orléans

Remerciements

Ces travaux de thèse ont été réalisés au Centre de Biophysique Moléculaire d'Orléans (CBM, CNRS-UPR 4301) au sein du groupe thématique « RMN des biomolécules ». Ils ont été financés par une bourse de la Région Centre-Val de Loire. Je remercie le Dr Eva Jakab Toth de m'avoir accueillie au sein du laboratoire.

Je remercie profondément le Dr Ewen Lescop et le Dr Jean-Pierre Simorre d'avoir accepté d'être les rapporteurs de ce manuscrit. Je remercie sincèrement le Dr Marc Boudvillain et Dr Isabelle Landrieu d'avoir accepté de participer à l'examen de mes travaux.

Je remercie toutes les personnes qui ont participé à ce travail :

Je tiens à remercier Agnès Delmas pour m'avoir accueillie au sein du groupe RMN et pour les nombreux conseils pendant mon séjour dans l'équipe. Nous sommes arrivées dans le groupe à la même période et je tiens à te remercier pour tous les conseils, pour avoir toujours été là quand j'avais des questions et pour ta façon d'être qui m'ont aidé à m'affirmer et à murir pendant ces 3 ans.

Je tiens à remercier Céline Landon, ma directrice de thèse, pour m'avoir permis de réaliser ce travail de thèse, pour m'avoir fait confiance et pour son soutien pendant ces trois ans. Je te remercie pour tes conseils, ton écoute, tes nombreux encouragements et pour m'avoir appris à reconnaître ma valeur.

Je tiens à remercier Karine Loth pour son encadrement et ses enseignements tout au long de la thèse. Merci de m'avoir fait confiance pour mener à bien ce projet mais aussi pour m'avoir laissé squatter ton bureau pendant ces trois ans. Merci pour tout ce que tu m'as appris et pour l'autonomie que j'ai pu acquérir. Merci aussi pour les conversations à base de mangas, de dramas, et de super-héros. Ça fait une sacrée liste de livres/séries/films à voir.

Je remercie Françoise Keita Paquet pour son aide et ses connaissances sur la protéine MC1. Merci également pour l'initiation au yoga et pour toutes les conversations que nous avons partagées allant des différents chakras à l'histoire du labo et jusqu'à l'étude structurale d'oligonucléotides par RMN.

Je remercie Bertrand Castaing de m'avoir accueillie dans son équipe pour la préparation des protéines et les études structurales de MC1. Merci pour toutes les connaissances que tu m'as transmises sur les protéines chaperonnes et en particulier sur HU. Merci pour ta bonne humeur et pour tous tes conseils bienveillants. Merci de m'avoir laissé refaire la déco du labo et pour la représentation personnifiée de HU que j'aurai mise à la fin de toutes mes présentations orales.

Je remercie Franck Coste pour m'avoir initiée à la cristallographie, aux études de SAXS et à la préparation de protéines. Je garde un excellent souvenir des expéditions au synchrotron et tiens à te remercier pour tous tes conseils avisés. Merci également pour l'intégration au groupe des « cassos » et pour les innombrables « alors, on en est où ? » qui m'ont fait sursauter pendant 3 ans. Il n'y a pas à dire, bosser avec toi ça fortifie le caractère !

Un merci tout particulier à Virginie Gaudon pour m'avoir formée à la préparation et purification de protéines. Tu m'as appris que « oui ! On peut ajuster un volume directement en éprouvette de 2L » alors que jusqu'alors je ne jurais que par les fioles jaugées. Merci pour tous tes conseils à la paillasse, pour les nombreux

Remerciements

fou-rires et l'ambiance musicale dans le labo. Toi qui es devenue une très bonne amie, merci pour ton soutien durant ces trois années et pour tous ces moments que l'on a partagé.

Je remercie Stéphane Goffinont pour m'avoir initiée aux gels retard et pour avoir fourni le plasmide de MC1t. Merci de m'avoir partagé ton goût pour la bonne nourriture et pour avoir été un fidèle testeur de mes expérimentations culinaires.

Je remercie Françoise Culard pour tous ses conseils et pour ses connaissances sur la protéine MC1. On ne se sera pas côtoyé longtemps mais je te remercie d'avoir pris le temps de répondre aux emails quand j'avais des questions.

Je remercie Guillaume Gabant pour les analyses de pureté sur mes échantillons et pour ses conseils. Merci pour ta gentillesse et ta bonne humeur (et également pour les batailles d'eau... rafraichissantes).

Je remercie Rémy Le Meur, dont j'aurai entendu parler tout au long de la thèse comme s'il était toujours là. Merci de m'avoir préparé le terrain pour réaliser ce travail de thèse et pour m'avoir toujours pris le temps de me répondre quand j'avais des questions.

Je tiens également à remercier Alexandre, Coline, Matéo, Mylène, Lisa, Emeline et Michael, mon groupe EDIFICE. C'était un vrai challenge de vous faire comprendre et de vous intégrer au projet de thèse. Au final je suis très fière de tous vos progrès et j'ai beaucoup apprécié les échanges que nous avons eu au fil des séances.

Il faudrait mettre les noms de tout le monde et je ne voudrais pas oublier quelqu'un... Je tiens à remercier chaleureusement l'ensemble des membres du CBM anciens et actuels. Merci pour votre bienveillance et pour les nombreuses conversations que l'on a pu avoir dans les couloirs, autour d'un café, ou d'un repas.

Un grand merci à l'ensemble des membres de l'équipe AMV et tout particulièrement aux personnes des groupes « RMN des biomolécules » et « Réparation de l'ADN: Structure, Fonction et Dynamique » anciens et actuels : Agnès Delmas, Céline Landon, Françoise Keita-Paquet, Karine Loth, Marylène Bertrand-Urbaniak, Laurence Jouvensal, Hervé Meudal, Françoise Culard, Abdennour Braka, Bertrand Castaing, Franck Coste, Stéphane Goffinont, Martine Guérin, Virginie Gaudon, Norbert Garnier, Julien Cros et Alan Wacquiez. Vous avez tous été d'un grand soutien pour aller jusqu'au bout de cette thèse mais également forts de conseils culinaires et touristiques. J'ai pris beaucoup de plaisir à venir travailler tous les jours avec vous.

Je remercie mes sauveurs du quotidien : Hervé et Yannick, pour avoir fait en sorte que mon pc ne rende pas l'âme avant la fin de la thèse, et Justo et Martial, pour vous être régulièrement battus contre la climatisation du bureau et ainsi m'éviter de porter des pulls en pleine canicule.

Je remercie tous ceux qui m'auront accompagnée pour faire trempette à la piscine histoire de garder la forme : Hervé, Virginie, Julien, Agnès, Céline, Bertrand, Norbert, Stéphane, Mohammed, Hélène, Béatrice et Cédric.

Je remercie les joueurs de ping-pong et de tennis de table du mercredi soir pour m'avoir appris ce sport et à renvoyer les balles (avec plus ou moins de succès).

Remerciements

Un grand merci à tous les amis pour les bons moments partagés au labo et en dehors :

Julien, meilleur ami, confident et plus grand soutien tout au long de cette thèse, tu auras supporté toutes mes humeurs les bonnes comme les mauvaises. Tu auras subi des attaques à la pissette et de multiples tags sur ta paillasse en ayant la politesse de toujours me rendre la pareille. Merci pour ta bonne humeur et pour avoir fait en sorte qu'en trois ans le quotidien ne soit jamais une routine au labo comme à la maison.

Mohammed ou comme je t'appelle affectueusement : « mon pote de galère » ! Nous sommes arrivés en même temps et nous irons jusqu'au bout de nos thèses. Merci pour ton soutien, pour tes blagues, pour tes bricks délicieuses, pour être régulièrement venu me taxer des friandises et pour tous ces excellents moments passés au labo et en dehors.

Saïda pour ton sourire ravageur et ton sens de la répartie, Skander pour toutes tes idées de plans foireux (il faudrait plus d'une vie pour tout faire), Florian pour l'organisation des dits plans foireux et pour ton âme d'artiste, Vinni, Lauren et Camille pour les soirées filles et pour tous les bons moments passés ensemble, Alan pour ta jovialité et pour ton côté grand-frère, Cédric pour toutes tes histoires et pour tes pâtisseries, Anton pour nous avoir appris qu'aucun alcool ne se dilue, Isabelle pour ta bonne humeur et pour toutes tes attentions gourmandes pendant que je rédigeais, Maxence et Hugo pour toutes les sorties et les soirées (vous savez toujours comment mettre l'ambiance), Abdenmour pour être venu en soutien quand je faisais mes cinétiques le week-end, la bande des normands : Alexis, Mika, Dr Picard, Dr DSJ, Thomas, Laure, Lia, Thibaud, Marine, Alicia, Florian, Valentin, Clémence, Loïc et Gildas pour votre soutien à distance et pour toutes les grosses marrades passées présentes et à venir ! Merci également à Daouda, Patrick, Mamar, Mehdi, Caroline, Elodie, Amandine, Nadira, Chaïma, Florence, Oliver, Geoffrey, Kévin, Keyron et les autres avec qui j'ai pu partager de très bons moments. Merci pour les petites pauses sucrées, pour les soirées et les sorties ADSO, pour les pauses en bord de Loire, pour les batailles d'eau ou de neige, pour la collection de photos-dossier et pour les plans foireux du week-end. Vous avez tous été d'un grand soutien et c'est certain, ces trois années à Orléans n'auraient pas été aussi joyeuses et pleines de souvenirs sans vous tous.

Pour finir, un grand merci à ma famille ! Vous êtes toujours à mes côtés même quand géographiquement vous êtes loin. Vous m'encouragez et me soutenez depuis le tout début, même si vous ne comprenez pas forcément ce que je fais. Merci de m'avoir fait confiance et de m'avoir aidé à aller jusqu'au bout de cette thèse.

A ma famille de sang,

A ma famille de cœur,

Et à toi Ju, qui m'a fait découvrir une nouvelle façon d'aimer,

Je vous dédie cette thèse et vous remercie de faire que chaque jour est un bonheur partagé.

Sommaire

Liste des abréviations	5
Liste des figures et des tableaux	9
Introduction générale	15
Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine <i>EcHU</i>	23
Chapitre 1. HU, une protéine au mécanisme d'échange complexe, essentielle à la croissance et à la survie des Bactéries	25
1.1. La protéine HU	25
1.1.1. Découverte et principales caractéristiques de la protéine	25
1.1.2. HU chez <i>E.coli</i>	30
1.1.3. Structure de <i>EcHU</i>	33
1.1.4. Interaction de <i>EcHU</i> avec différents types d'ADN	33
1.1.5. Propriétés physico-chimiques de <i>EcHU</i>	34
1.2. Mise en évidence d'une conformation intermédiaire d'<i>EcHU</i>	35
1.2.1. Premières descriptions structurales de l'intermédiaire	35
1.2.2. Première caractérisation de l'échange entre les conformations, par RMN	38
1.2.3. Proposition d'un modèle de la conformation intermédiaire combinant les données RMN et MD	39
1.3. L'hypothèse d'une forme tétramérique transitoire dans la formation de l'hétérodimère	40
1.3.1. Hypothèse sur le mécanisme de formation de l'hétérodimère	40
1.3.2. L'oligomère transitoire	41
1.3.3. Mise en évidence de formes tétramériques	41
1.4. Conclusion : <i>EcHU</i>, un système d'étude de choix, mais un système d'étude complexe	43
Chapitre 2. Matériel et méthodes pour la description des <i>EcHU</i>	45
2.1. Du plasmide à l'échantillon de protéine pure	45
2.1.1. Surexpression de <i>EcHU</i> non marquée ou marquée (^{15}N , $^{15}\text{N}^{13}\text{C}$)	46
2.1.2. Purification des protéines $\text{HU}\alpha_2$ et $\text{HU}\beta_2$	47

2.1.3. Préparation des échantillons pour les différentes études par RMN	49
2.2. Détermination de la structure spatiale d'une protéine.....	50
2.2.1. Première étape : Attribution.....	52
2.2.2. Deuxième étape : Détermination des contraintes RMN.....	55
2.2.3. Troisième étape : Modélisation moléculaire et calcul de structure.....	56
2.3. Les principaux processus d'échange accessibles en RMN	58
2.3.1. Processus d'échange rapide (ps-ns).....	59
2.3.2. Processus d'échange intermédiaire (μs-ms)	60
2.3.3. Processus d'échange lent (>ms).....	61
2.4. Détermination de la cinétique globale d'un système biochimique par RMN	63
2.4.1. Etude de la réaction globale.....	64
2.4.2. Caractérisation de la vitesse de réaction globale	66
Chapitre 3. Etape clé pour l'échange de chaînes : le dimère intermédiaire.....	69
3.1. N₂/I₂ : un équilibre thermodynamique très dépendant de la température	69
3.2. Caractérisation structurale de la conformation native à basse température	73
3.2.1. Attribution et prédiction des éléments de structure secondaire.....	73
3.2.2. Structure de la conformation native en solution.....	75
3.2.3. Caractérisation de la dynamique interne	80
3.3. Caractérisation structurale de la conformation native à haute température de l'hétérodimère.....	82
3.3.1. Caractérisation des résonances.....	82
3.3.2. Structure de la conformation native à haute température	83
3.4. Caractérisation structurale de la conformation intermédiaire.....	86
3.4.1. Attribution de l'hétérodimère dans sa conformation intermédiaire	87
3.4.2. Structure de l'intermédiaire à haute température.....	90
3.5. Conclusion	94
Chapitre 4. Etude de la cinétique globale de la réaction.....	97
4.1. Titration de l'échange de chaînes.....	98
4.2. Etude cinétique de l'échange de chaînes	102
4.3. Caractérisation cinétique des étapes intermédiaires.....	107

4.4. Conclusion : la formation du tétramère transitoire comme étape limitante de la réaction	109
--	-----

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1 111

Chapitre 1. Caractérisation d'une courbure en V de l'ADN induite par MC1 113

1.1. MC1 : une protéine uniquement présente chez les Archées	113
1.1.1. Les Archées en bref.....	113
1.1.2. Les principales protéines architecturales d'Archées.....	118
1.1.3. La protéine MC1	123
1.2. Complexe ADN ^{23bp} /MC1	130
1.2.1. Matériel et méthodes.....	132
1.2.2. Perturbations de déplacement chimique	133
1.2.1. Relaxation Dispersion Carr-Purcell-Meiboom-Gill	135
1.3. Conclusion	139

Chapitre 2. Vers une caractérisation aux rayons X du complexe ADN^{23bp}/MC1 141

2.1. Matériel et méthodes	141
2.1.1. Préparation des échantillons de complexe	141
2.1.2. Cristallographie aux rayons X de protéines.....	151
2.1.3. Diffusion de rayons X aux petits angles (SAXS).....	155
2.2. Optimisation du complexe pour l'étude	157
2.2.1. Surexpression et purification de la protéine MC1	157
2.2.2. Surexpression et purification de la protéine MC1t	157
2.2.1. Complexation avec l'ADN ^{23bp}	165
2.3. Vers une description atomique du complexe	167
2.3.1. Optimisation de la cristallogenèse.....	167
2.3.2. Tests de diffraction des rayons X au synchrotron.....	171
2.4. Vers une description de l'enveloppe spatiale du complexe	172
2.4.1. Essais de diffusion SAXS.....	172
2.4.2. Calculs d'enveloppe.....	175
2.5. Conclusion : vers une détermination électronique de la structure	179

Sommaire

Conclusion générale.....181
Références bibliographiques..... 185
Annexes..... 199

Liste des abréviations

Abréviations générales

1D.....	unidimensionnel
2D.....	bidimensionnel
20CM.....	20 CarboxyMethyl
3D.....	tridimensionnel
4D.....	quadridimensionnel
6HIS.....	étiquette de six résidus Histidine
6HIS-MC1t.....	protéine MC1t fusionnée à l'étiquette « 6HIS »
AcA54.....	mélange Acrylamide 5 % et Agarose 4 %
ACN.....	ACétoNitrile
AcOH.....	acide acétique
ADN.....	Acide DésoxyriboNucléique
Alba.....	Acetylation lowers binding affinity
Ampi.....	Ampicilline
ARIA.....	Ambiguous Restraints for Iterative Assignment
ARN.....	Acide RiboNucléique
CC1.....	Crenarchaeal Chromatin protein 1
CD.....	dichroïsme circulaire (Circular Dichroism)
CNS.....	système de cristallographie et de RMN (Crystallography and NMR System)
CRP.....	protéine réceptrice de l'AMP cyclique (Cyclic AMP Receptor Protein)
CS.....	déplacement chimique (Chemical Shift)
CSI.....	index de déplacement chimique (Chemical Shift Index)
CSP.....	perturbation de déplacement chimique (Chemical Shift Perturbation)
CV.....	volume de colonne (Column Volume)
D ₂ O.....	eau deutérée
DANGLE.....	Dihedral ANGles from Global Likelihood Estimates
DBD.....	domaine de liaison à l'ADN (DNA-Binding-Domain)
DPANN.....	Diapherotrites-Parvarchaeota-Aenigmarchaeota-Nanoarchaeota-Nanohaloarchaeota
DSC.....	calorimétrie différentielle à balayage (Differential Scanning Calorimetry)
DTT.....	DiThioThréitol
EB.....	Extrait Brut
EcHU.....	protéine HU d' <i>Escherichia coli</i>
EcHUi ou i.....	chaîne protéique i de EcHU
EcHU α_2 ou $\alpha\alpha$	protéine homodimérique HU α_2 d' <i>E. coli</i>
EcHU $\alpha\beta$ ou $\alpha\beta$	protéine hétérodimérique HU $\alpha\beta$ d' <i>E. coli</i>
EcHU β_2 ou $\beta\beta$	protéine homodimérique HU β_2 d' <i>E. coli</i>
EDTA.....	EthylèneDiamineTétraAcétique
EXSY.....	spectroscopie d'échange ou échange ZZ (Exchange Spectroscopy)
Fis.....	facteur d'inversion-stimulation (Factor for inversion stimulation)
HADDOCK.....	High Ambiguity Driven protein-protein DOCKing
Hepes.....	acide 4-(2-Hydroxyéthyl)-1-pipérazine éthane sulfonique
HMf[A/B].....	Histone [A/B] de <i>Methanothermus fervidus</i>
HMG.....	groupe à haute mobilité (High Mobility Group)
HMQC.....	Heteronuclear Multiple Quantum Coherence
H-NS.....	Histone-like Nucleoid Structuring protein
HSQC.....	Heteronuclear Single Quantum Coherence
HTH.....	hélice-tour-hélice (Helix-Turn-Helix)
HU.....	Histone-like U93
IHF.....	facteur d'intégration de l'hôte (Integration Host Factor)
IMAC.....	chromatographie d'affinité sur ions métalliques immobilisés (Immobilized Metal Affinity Chromatography)

Liste des abréviations

IPTG	IsoPropyl β -D-1-ThioGalactopyranoside
LAECA	Last Archaeo-Eukaryotic Common Ancestor
LB (ou LB Broth)	Luria Bertani Broth
MC1	Methanogen Chromosomal protein 1
MC1t	Glycine Sérine Histidine Méthionine - Methanogen Chromosomal protein 1 (= GSHM-MC1)
MD	dynamique moléculaire (Molecular Dynamics)
MonoS	Monodisperse bead Sulfomethyl
MPD	2-Méthyl-2,4-PentaneDiol
MS	spectrométrie de masse (Mass Spectrometry)
MW	masse molaire (Molecular Weight)
NAP	protéine associée au nucléoïde (Nucleoid Associated Protein)
NBD	domaine de liaison au nucléosome (Nucleosomal Binding Domain)
NESSY	NMR rELaxation diSpersion SpectroscopY
NM	Non Marqué
nOe (ou NOE)	effet Overhauser nucléaire (nuclear Overhauser effect)
NOESY	Nuclear Overhauser SpectroscopY
OVN	pendant la nuit (OverNight)
pdb	banque de données sur les protéines (protein data bank)
PEG	PolyEthylène Glycol
pET	plasmid Expression vector under T7 control
PMSF	fluorure de phénylméthylsulfonyl (PhenylMethylSulfonyl Fluoride)
RDC	couplage résiduel dipolaire (Residual Dipolar Coupling)
RD CPMG (ou RD ^{15}N -CPMG)	Relaxation Dispersion Carr-Purcell-Meiboom-Gill
RMN	Résonance Magnétique Nucléaire
RMN-TR	RMN en Temps Réel
RMSD	déviation de la racine de la moyenne des carrés (Root Mean Square Deviation)
RX	Rayons X
SA	Sulfate d'Ammonium
SAXS	diffusion de rayons X aux petits angles (Small Angle X-rays Scattering)
SDS-PAGE	Sodium Dodecyl Sulfate - PolyAcrylamide Gel Electrophoresis
SELEX	Systematic Evolution of Ligands by EXponential enrichment
SOFAST	band-Selective Optimized-Flip-Angle Short-Transient
SP-HP	SulphoPropyl Sepharose High Performance
SP-FF	SulphoPropyl Sepharose Fast Flow
TACK	Thaumarchaeota-Aigarchaeota-Crenarchaeota-Korarchaeota
TCEP	Tris(2-CarboxyEthyl)Phosphine
TFAM	facteur de transcription mitochondrial A (Mitochondrial Transcription Factor A)
Thr	thrombine
TOCSY	TOTal CORrelated SpectroscopY
Tp	tampon
Tris	Tris(hydroxyméthyl)aminométhane
UV	UltraViolet
viol	violation
β mercpto	β -mercaptoéthanol

Symboles

^1H	hydrogène 1
$^3J_{\text{IS}}$	constante de couplage entre les atomes I et S séparés par 3 liaisons covalentes
^{13}C	carbone 13
^{15}N	azote 15
^{32}P	phosphore 32
α_i	hélice <i>alpha</i> n°i
A_{Xnm}	Absorbance à X nm
B_0	champs magnétique principal
β_i	brin <i>beta</i> n°i

Liste des abréviations

bp		paire de bases (base pair)
C_i		complexe D/MC1t
d		distance
D		état dénaturé ou déstructuré de la protéine HU
D_i		duplex d'ADN de brin inférieur n°i
D_{max}		dimension maximale
$\Delta_r G^0$		enthalpie libre standard de réaction
ξ		avancement chimique
E_a		énergie d'activation
E_i		énergie i
I (ou I(T), I(t), I(q))		intensité
I_2		état intermédiaire de la protéine HU
INF_i		brin inférieur n°i d'ADN
K		constante d'équilibre
K_D		constante de dissociation
k_{ex}		constante d'échange
k_i		constante de vitesse de i
λ		longueur d'onde
LP_i		boucle n°i (Loop n°i)
n°		numéro
N_2		état natif de la protéine HU
η_{NH}		effet Overhauser nucléaire 1H - ^{15}N
p_i		population i
pH		potentiel hydrogène
pl		point Isoélectrique
q		vecteur de diffusion
R		constante des gaz parfaits
R_1		vitesse de relaxation longitudinale ($=1/T_1$)
R_2 ou R_2^0		vitesse de relaxation transversale ($=1/T_2$)
$R_{2,eff}$		vitesse d'échange effective
R_{ex}		vitesse d'échange
R_g		Rayon de giration
SUP		brin supérieur d'ADN
θ_i		angle i
t		temps
T		temps de mélange
T°		température
T_1		temps de relaxation longitudinale ($=1/R_1$)
$t_{1/2}$		temps de demi-vie
T_2		temps de relaxation transversale ($=1/R_2$)
τ		taux d'avancement
τ_{CPMG}		délai CPMG
τ_{ex}		durée de vie réduite de l'échange
T_m		température de transition (melting temperature)
T_{relax}		délai de relaxation
V		volume
ν_{CPMG}		fréquence d'impulsion CPMG
ν_r		vitesse de réaction
ω (ou δ)		déplacement chimique
ω_0		fréquence de Larmor

Unités

%		pourcent
°C		degré Celcius

Liste des abréviations

Å	angström (10^{-10} m)
cal	calorie
Da (ou g.mol^{-1})	dalton (ou gramme par mole)
g	gramme
h	heure
Hz	hertz
K	Kelvin
L	litre
m	mètre
M (ou mol.L^{-1})	molaire (ou mole par litre)
min	minute
mol	mole
ppm	partie par millions
rpm	rotation par minute
s	seconde
UA_{Xnm}	unité d'absorbance à X nm

Abréviations des acides aminés

A – Ala – Alanine	I – Ile – Isoleucine	R – Arg – Arginine
C – Cys – Cystéine	K – Lys – Lysine	S – Ser – Sérine
D – Asp – Aspartate	L – Leu – Leucine	T – Thr – Thréonine
E – Glu – Glutamate	M – Met – Méthionine	V – Val – Valine
F – Phe – Phénylalanine	N – Asn – Asparagine	W – Trp – Tryptophane
G – Gly – Glycine	P – Pro – Proline	Y – Tyr – Tyrosine
H – His – Histidine	Q – Gln – Glutamine	

Abréviations des nucléosides

A – Ade – Adénine	C – Cyt – Cytosine	G – Gua – Guanine	T – Thy – Thymine
-------------------	--------------------	-------------------	-------------------

Les structures primaires et la nomenclature (IUPAC) des acides aminés et des nucléosides sont présentées en **Annexe 1**.

Liste des figures et des tableaux

Liste des figures

Figure 1 : Arbre phylogénique des trois grands domaines du Vivant.....	17
Figure 2 : Schéma des différents types de compaction de l'ADN par les protéines architecturales	18
Figure 3 : Structure tridimensionnelle et diagramme de topologie des protéines architecturales majeurs du Vivant.....	19
Figure 4 : Clichés de microscopie électronique de molécules d'ADN λ seules et complexées avec la protéine HU.....	26
Figure 5 : Structure tridimensionnelle des protéines de la superfamille HU/IHF	27
Figure 6 : Modèles de structure de HU complexé à un ADN.....	28
Figure 7 : Schéma simplifié présentant les gènes hupA et hupB.....	31
Figure 8 : Prédominance des dimères de EchU au cours du cycle bactérien.....	32
Figure 9 : Séquence en acides aminés et éléments de structure secondaire des chaînes protéiques EchU α et EchU β	33
Figure 10 : Affinité des protéines EchU $\alpha\beta$, EchU α_2 et EchU β_2 à différentes structures d'ADNs synthétiques (K_D)	34
Figure 11 : Schéma du dépliement de EchU basé sur les données de CD et de DSC	37
Figure 12 : Suivi de l'évolution au cours du temps des éléments de structure secondaire de chaque résidu dans EchU β_2	38
Figure 13 : Modèle structural de l'intermédiaire déduit des expériences RMN et de simulations « gros grains »	40
Figure 14 : Modèle du tétramère transitoire et schéma du mécanisme de formation de l'hétérodimère à partir des structures cristallographiques et des modèles « gros-grain »...	41
Figure 15 : Spectre de masse en conditions natives de suivi de la formation de l'hétérodimère	42
Figure 16 : Schéma du mécanisme complet d'échange de chaînes de la protéine HU d'E. coli	44
Figure 17 : Schéma de préparation des échantillons de protéines HU.....	45
Figure 18 : Etapes de la détermination de structure spatiale de protéine par RMN.....	50
Figure 19 : Illustration de la stratégie d'attribution du squelette peptidique étendu aux corrélations β	52
Figure 20 : Distribution moyenne des déplacements chimiques ^1H et ^{13}C des acides aminés	53
Figure 21 : Schéma simplifié du processus d'attribution et de calcul de structures itératif de ARIA.....	57

Figure 22 : L'effet de l'échange chimique entre deux états, A et B, sur des spectres RMN. ...	58
Figure 23 : Echelle de temps observable en RMN des mouvements biomoléculaires	59
Figure 24 : Relaxation de spin	60
Figure 25 : Expérience de Relaxation Dispersion CPMG	61
Figure 26 : Spectroscopie d'échange, EXSY ou échange ZZ.....	62
Figure 27 : Types de réaction chimique	64
Figure 28 : Schéma du montage pour les essais de titrage de EchU β_2 par EchU α_2	66
Figure 29 : Suivi de l'évolution des populations N ₂ /I ₂ avec la température sur chacun des dimères EchU par des ¹⁵ N-HSQC.....	70
Figure 30 : Suivi des intensités normalisées du résidu 90 des dimères de EchU à différentes températures	71
Figure 31 : Prédiction des éléments de structure secondaire des chaînes EchU α et EchU β de l'hétérodimère natif, à 20 °C	74
Figure 32 : Superposition des 20 meilleurs modèles de structure de la conformation native de EchU	76
Figure 33 : Mise en évidence des trois zones d'interactions polaires inter chaînes.....	78
Figure 34 : Poches hydrophobes des dimères de EchU.....	79
Figure 35 : Relaxation ¹⁵ N R ₁ , R ₂ et NOE du squelette peptidique des dimères de EchU	81
Figure 36 : Prédiction des éléments de structure secondaire des chaînes EchU α et EchU β pour la conformation native de l'hétérodimère à 20 °C et à 37 °C	83
Figure 37 : Superposition des 20 meilleurs modèles de la conformation native de EchU $\alpha\beta$ à 37 °C.....	84
Figure 38 : Superposition des deux meilleurs modèles structuraux de l'hétérodimère natif à basse et haute température.....	85
Figure 39 : Cavité hydrophobe de EchU $\alpha\beta$ dans sa conformation native à 37 °C	86
Figure 40 : Spectres ¹⁵ N-HSQC des chaînes α et β de EchU $\alpha\beta$ à 20 °C et à 37 °C	88
Figure 41 : Perturbation des déplacements chimiques entre la conformation native et intermédiaire de EchU $\alpha\beta$ à 37 °C.....	89
Figure 42 : Perturbation des déplacements chimiques entre les conformations natives et intermédiaires des homodimères	89
Figure 43 : Prédiction des éléments de structure secondaire entre la conformation native (N ₂) et intermédiaire (I ₂) de l'hétérodimère	90
Figure 44 : Structure tridimensionnelle de l'hétérodimère EchU $\alpha\beta$ dans sa conformation intermédiaire, à 37 °C	92
Figure 45 : Cavité hydrophobe de EchU $\alpha\beta$ dans sa conformation intermédiaire à 37 °C	93
Figure 46: Position des résidus considérés comme des déterminants structuraux de la stabilité de la forme native	94
Figure 47 : Equation-bilan du mécanisme global de formation de l'hétérodimère.....	97

Figure 48 : Schéma du mécanisme complet d'échange de chaînes de la protéine HU d'E. coli	97
Figure 49 : Etude cinétique du système de EchU par titrage à l'équilibre	98
Figure 50 : ^{15}N-HSQC de EchUβ_2:EchUα_2 au ratio 1:0,1 du titrage	99
Figure 51 : Suivi de la formation de l'hétérodimère lors du titrage de EchUβ_2 par EchUα_2 selon deux méthodes	100
Figure 52 : Suivi de la formation de l'hétérodimère et de la disparition de l'homodimère titré lors du titrage de EchUβ_2 par EchUα_2	101
Figure 53 : Etude cinétique du système de EchU par suivi cinétique de la réaction globale	102
Figure 54 : Suivi cinétique de la formation de l'hétérodimère de EchU à 20 °C	105
Figure 55 : Vérification de la loi d'Arrhénius à partir des données cinétiques obtenues par RMN et spectrométrie de masse en conditions natives	106
Figure 56 : Etude cinétique du système de EchU par la caractérisation des réactions intermédiaires, en particulier l'échange N_2/I_2	107
Figure 57 : Diversité des Archées et de leurs biotopes	113
Figure 58 : Evolution de la classification des Archées dans l'Arbre phylogénique du Vivant	115
Figure 59 : Différents arbres phylogénétiques résultant de l'inclusion du superphylum des Asgards	118
Figure 60 : Distribution des protéines de structure majeures du nucléoïde archéen	118
Figure 61 : Structure tridimensionnelle de HMfB$_2$	119
Figure 62 : Alignement de séquences de HU d'Archées	120
Figure 63 : Structure tridimensionnelle d'Alba par cristallographie aux rayons X	121
Figure 64 : Structure tridimensionnelle des protéines Sul7 et de Cren7 complexées à un ADN par cristallographie aux rayons X	122
Figure 65 : Séquences primaires et secondaires de Sul7 et Cren7	122
Figure 66 : Images de microscopie électronique de complexes ADN$^{176\text{bp}}$/MC1	124
Figure 67 : Structure de la protéine MC1 de M. thermophila en solution	124
Figure 68 : Alignement de la séquence de MC1 de Methanosarcina thermophila (HMC1_METTE) avec des protéines MC1 des classes Methanomicrobia et Halobacteria d'Archée	125
Figure 69 : Structure de protéines chromatiniennes d'Archées, différentes des histones et des homologues d'Alba	126
Figure 70 : Potentiel électrostatique de la protéine MC1	126
Figure 71 : Surface d'interaction de MC1 avec l'ADN$^{15\text{bp}}$	127
Figure 72 : Séquence d'ADN$^{15\text{bp}}$ et ses propriétés d'interaction	128
Figure 73 : Modèle de l'interaction de MC1 avec un ADN$^{15\text{bp}}$ obtenue par HADDOCK	129

Liste des figures et des tableaux

Figure 74 : Détermination structurale du complexe ADN ^{15bp} /MC1 par RMN.	130
Figure 75 : séquences d'ADN d'interaction préférentielle des protéines TFAM et MC1	131
Figure 76 : Modèle de l'interaction de MC1 avec un ADN ^{23bp}	131
Figure 77 : Spectre 700 MHz ¹⁵ N-HSQC de la protéine MC1 marquée ¹⁵ N	134
Figure 78 : Spectre 700 MHz ¹⁵ N-HSQC de la protéine MC1 marquée ¹⁵ N et complexée à un ADN de 23bp	134
Figure 79 : Courbes de l'évolution du signal CPMG des résidus Phe19 et Lys86 de la protéine MC1 en solution.	135
Figure 80 : Courbes de l'évolution du signal CPMG des résidus Arg71, Glu77 et Ile79 de la protéine MC1 dans le complexe à 15bp.....	136
Figure 81 : Courbe de l'évolution du signal CPMG des résidus Glu15 et Gly17 de la protéine MC1 dans le complexe ADN ^{23bp} /MC1.	137
Figure 82 : Interaction dynamique de MC1 avec un ADN ^{23bp}	138
Figure 83 : Schéma des principales phases pour la préparation d'échantillons de protéine.	141
Figure 84 : Carte du plasmide pET-28b(+).	142
Figure 85 : Régulation de la transcription de l'opéron lactose.	143
Figure 86 : Construction de la protéine MC1t fusionnée avec une étiquette « 6HIS » et exprimée par le gène d'intérêt.	143
Figure 87 : Principales étapes pour la surexpression de protéines.	143
Figure 88 : Schéma des différentes étapes de purification de la protéine MC1t.	146
Figure 89 : Illustration des différents duplex D _i d'ADN utilisés pour la caractérisation aux rayons X.	149
Figure 90 : Schéma des différentes étapes de préparation des complexes d'ADN ^{23bp} avec la protéine MC1t	149
Figure 91 : Diagramme de phase à deux dimensions d'une macromolécule en fonction de sa propre concentration et de la concentration d'un agent cristallisant	151
Figure 92 : Illustration de cristallogenèse par la méthode d'échange de vapeur.....	152
Figure 93 : Cliché de diffraction aux rayons X d'un cristal de lysozyme.	153
Figure 94 : Schéma simplifié du dispositif de diffraction des rayons X.....	154
Figure 95 : Intensités de diffusion et fonctions de distribution de la distance de formes géométriques simples	156
Figure 96 : Boîtes de culture cellulaire pour la surproduction de 6HIS-MC1t	158
Figure 97 : Evolution des cultures de bactéries au cours de la surproduction de 6HIS-MC1t	159
Figure 98 : Analyse de l'élution sur colonne IMAC	160
Figure 99 : Analyse de l'élution de la colonne HS20	162
Figure 100 : Clivage de l'étiquette « 6HIS » à la thrombine.....	163

Liste des figures et des tableaux

Figure 101 : Analyse de l'élution de la colonne S75	164
Figure 102 : Spectre de masse déconvolué de la protéine MC1t pure	165
Figure 103 : Chromatogramme de l'élution de SUP_{RX}^{23bp}	166
Figure 104 : Gouttes issues de tests de criblage sur les complexes D_i/15N-MC1	168
Figure 105 : Cristaux du complexe D₃/15N-MC1 en conditions 150 mM d'acétate de calcium, 100 mM de cacodylate de sodium pH 6, 17 % de PEG 8k et 5 % de glycérol	169
Figure 106 : Cristaux du complexe D₃/MC1t (ou C₃) dans les conditions COMPAS 1.2 (0,5 M KCl, 12 % PEG 8k, 10 % glycérol)	170
Figure 107 : Optimisation de la cristallogénèse sur le complexe C₃ à 20 °C	171
Figure 108 : Cliché de diffraction par rayons X d'un cristal de complexe C₃	172
Figure 109 : Superposition des courbes de diffusion SAXS de MC1t, de D₁ et de C₁	173
Figure 110 : Superposition des tracés de Kratky de MC1t, de D₁ et de C₁	175
Figure 111 : Enveloppe et courbe de diffusion générés par DAMMIN	176
Figure 112 : Courbes de diffusion générés par CRY SOL	177
Figure 113 : Enveloppe et courbe de diffusion générés par MONSA sur le complexe C₁	178

Liste des tableaux

Tableau 1 : Paramètres thermodynamiques des dimères de EchU déterminés par calorimétrie (DSC)	35
Tableau 2 : Eléments de structuration secondaire des dimères calculées par dichroïsme circulaire	36
Tableau 3 : Composition des solutions et des tampons pour la surproduction de protéines HU dans E. coli	46
Tableau 4 : Composition des tampons pour la purification de protéines HU de E. coli	48
Tableau 5 : Nombre et type de contraintes utilisées dans les calculs de structure des dimères de EchU	56
Tableau 6 : Tableau d'avancement théorique et volumes des différents ajouts du premier titrage de l'homodimère EchUβ₂ par l'homodimère EchUα₂	65
Tableau 7 : Description des suivis cinétiques de la formation de l'hétérodimère	67
Tableau 8 : Caractérisation des populations N₂/I₂ d'EchU en température par RMN et par DSC	72
Tableau 9 : Attribution des dimères de EchU dans leur conformation native, à 20 °C pour EchUα₂ et EchUαβ, à 15 °C pour EchUβ₂	74
Tableau 10 : Contraintes RMN et statistiques structurales de la conformation native de EchUα₂ à 20 °C, de EchUβ₂ à 15 °C et de l'hétérodimère à 20 °C	75
Tableau 11 : Attribution de l'hétérodimère de EchU dans sa conformation native, à 37 °C ...	82

Tableau 12 : Contraintes RMN et statistiques structurales de la conformation native de l'hétérodimère à 37 °C.	83
Tableau 13 : Attribution de l'hétérodimère de EchU dans sa conformation intermédiaire à 37 °C.	87
Tableau 14 : Contraintes RMN et statistiques structurales de la conformation intermédiaire de l'hétérodimère à 37 °C.	91
Tableau 15 : Tableau d'avancement de la formation d'hétérodimère en conditions stœchiométriques.	103
Tableau 16 : Tableau d'avancement de la formation d'hétérodimère en conditions non stœchiométriques.	104
Tableau 17 : Récapitulatif des différents suivis cinétiques et des valeurs de constante de vitesse d'échange obtenues.	106
Tableau 18 : Constantes de vitesse d'échange N_2/I_2 de l'homodimère EchUα_2 pour différentes températures.	108
Tableau 19 : Matrice des coefficients d'association (S_{AB}) entre différents organismes des trois domaines du Vivant	114
Tableau 20 : Constantes d'échange issues des expériences de Relaxation Dispersion ^{15}N-CPMG sur le complexe ADN^{23bp}/MC1.	137
Tableau 21 : Paramètres des données SAXS des échantillons analysés	174

Introduction générale

Introduction générale

Depuis les années 90, le Vivant peut se représenter sous la forme d'un arbre phylogénétique à trois branches ^[1]. Ces branches correspondent aux trois grands domaines du vivant : **les Eucaryotes**, **les Bactéries** et **les Archées** (Figure 1). Chacun de ces domaines, possède un patrimoine génétique différent qui leur est indispensable de conserver et de protéger. Le support de ce patrimoine génétique est l'**Acide DésoxyriboNucléique (ADN)**. Chez les Eucaryotes, on le retrouve dans le noyau des cellules, dans la lumière des mitochondries et dans les chloroplastes. Chez les Bactéries et les Archées, on le retrouve directement dans le cytoplasme sous la forme d'un agglomérat du matériel génétique chromosomique de forme irrégulière, **le nucléoïde**, et également sous la forme de petites molécules d'ADN circulaires appelées plasmides.

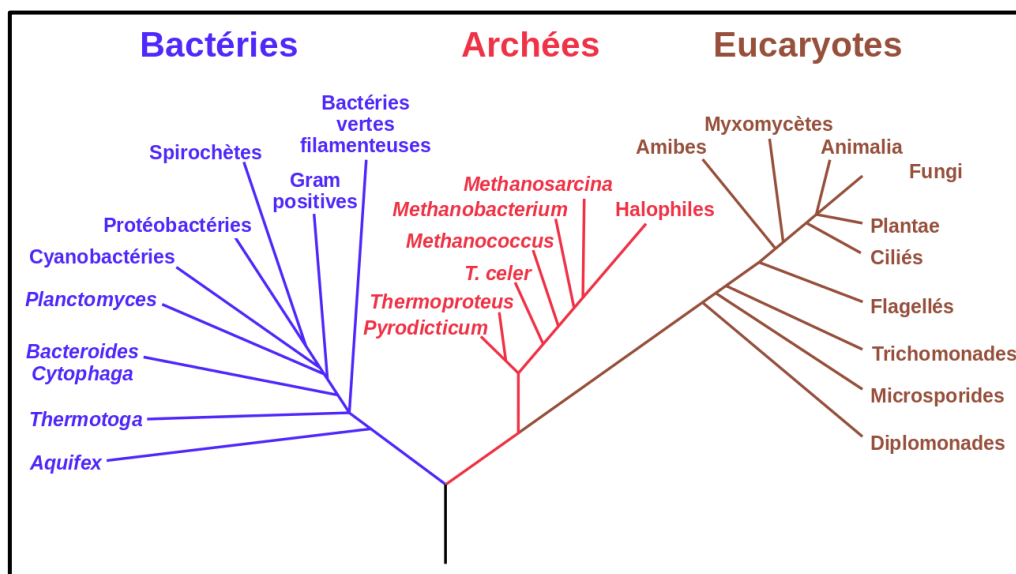


Figure 1 : **Arbre phylogénétique des trois grands domaines du Vivant** (adapté de Eric Gaba, Wikimedia Commons) ^[2]. Sont représentés en marron les Eucaryotes, en bleu les Bactéries et en rouge les Archées.

L'ADN possède une **structuration dynamique** dont l'organisation varie dans le temps et en fonction des conditions (température, pH, etc.). La modulation de cette dynamique est liée à différentes actions mises en place au sein de la cellule pour préserver son patrimoine génétique. L'une de ces actions est la modification de la géométrie de l'ADN par divers modes de **compaction / décompaction** ^[3]. Les premiers acteurs de cette dynamique sont des enzymes de topologie : **les topoisomérases**. Elles sont présentes dans les trois domaines du Vivant ^[4,5] et agissent en insérant ou en retirant des supertours sur la double hélice d'ADN, par coupure temporaire simple ou double brin. Les seconds acteurs responsables de la dynamique sont **les protéines architecturales**, également appelées **protéines chaperonnes de l'ADN**. Elles compactent ce dernier selon trois modes (Figure 2) : **en le pontant** (« bridgers »), **en l'enroulant** (« wrappers ») et/ou **en le courbant** (« benders »). Ces acteurs interagissent avec l'ADN de façon dynamique et **réversible** et les différentes actions exercées par ceux-ci permettent de **réguler des fonctions biologiques** comme la transcription ou la réplication de l'information génétique ^[6]. Ainsi,

de nombreuses études ont été et sont menées sur le **décryptage des systèmes protéiques interagissant avec l'ADN** dans les trois domaines du Vivant [7,8]. La connaissance détaillée de chacun des acteurs de la machinerie génétique et de leur fonctionnement propre permet de mieux comprendre l'ensemble des mécanismes cellulaires afin de pouvoir potentiellement les moduler ou les imiter. C'est dans ce cadre que s'inscrit le travail de thèse présenté dans ce manuscrit. Il est mené sur deux systèmes de protéines associées au nucléoïde : HU (protéine bactérienne, *E.coli*) et MC1 (protéine d'Archée, *M. thermophila*).

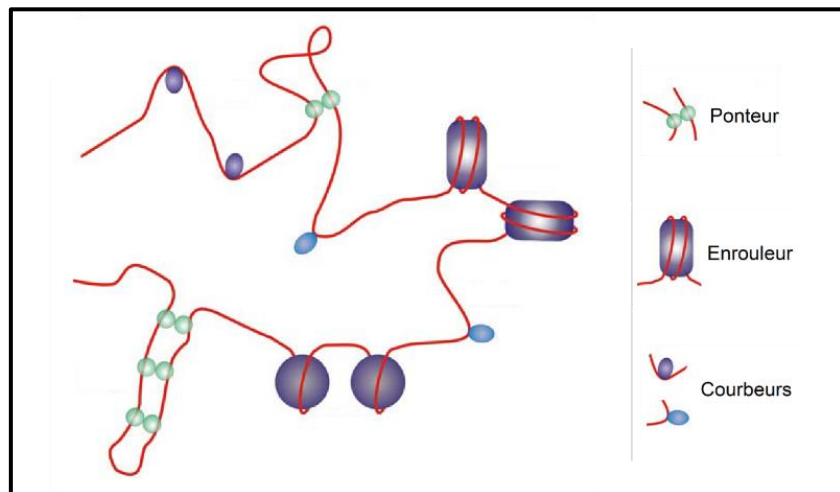


Figure 2 : Schéma des différents types de compaction de l'ADN par les protéines architecturales (adapté de Luijsterburg et al., 2008) [3].

Chez les Eucaryotes, on retrouve principalement deux types de protéines de structure : les histones [9,10] et les protéines HMG [11–14] (Figure 3). Les **histones** sont majoritairement des enrouleurs d'ADN. Il en existe 5 sous-unités différentes H1, H2A, H2B, H3 et H4 qui possèdent un motif caractéristique de trois hélices α . La sous-unité H1, une histone de liaison, reste sous forme monomérique pour interagir avec l'ADN. A l'inverse, les sous-unités H2A, H2B, H3 et H4 sont sous forme dimérique en solution (H2A-H2B et H3-H4) et s'assemblent en octamère pour interagir avec l'ADN. L'octamère, formé de deux copies de chaque sous-unité, enroule l'ADN de façon non spécifique. Cet enroulement, combiné à l'action de liaison de H1, compacte l'ensemble en un nucléosome rigide. Les protéines **HMG** (pour « High Mobility Group ») sont des courbeurs d'ADN structurées, comme les histones, en trois hélices α . Elles représentent trois familles de protéines (HMGA, HMGB, HMGN) possédant chacune un motif fonctionnel propre. Le motif fonctionnel permet une interaction avec des ADN spécifiques. Il correspond à une boîte-HMG (HMG-box) pour la famille HMGB, à une succession de trois crochets AT (AT hook) pour la famille HMGA, et à un domaine de liaison au nucléosome (NBD) pour la famille HMGN. L'ensemble des protéines HMG interagit avec l'ADN sous forme monomérique et de manière compétitive vis-à-vis de l'histone H1.

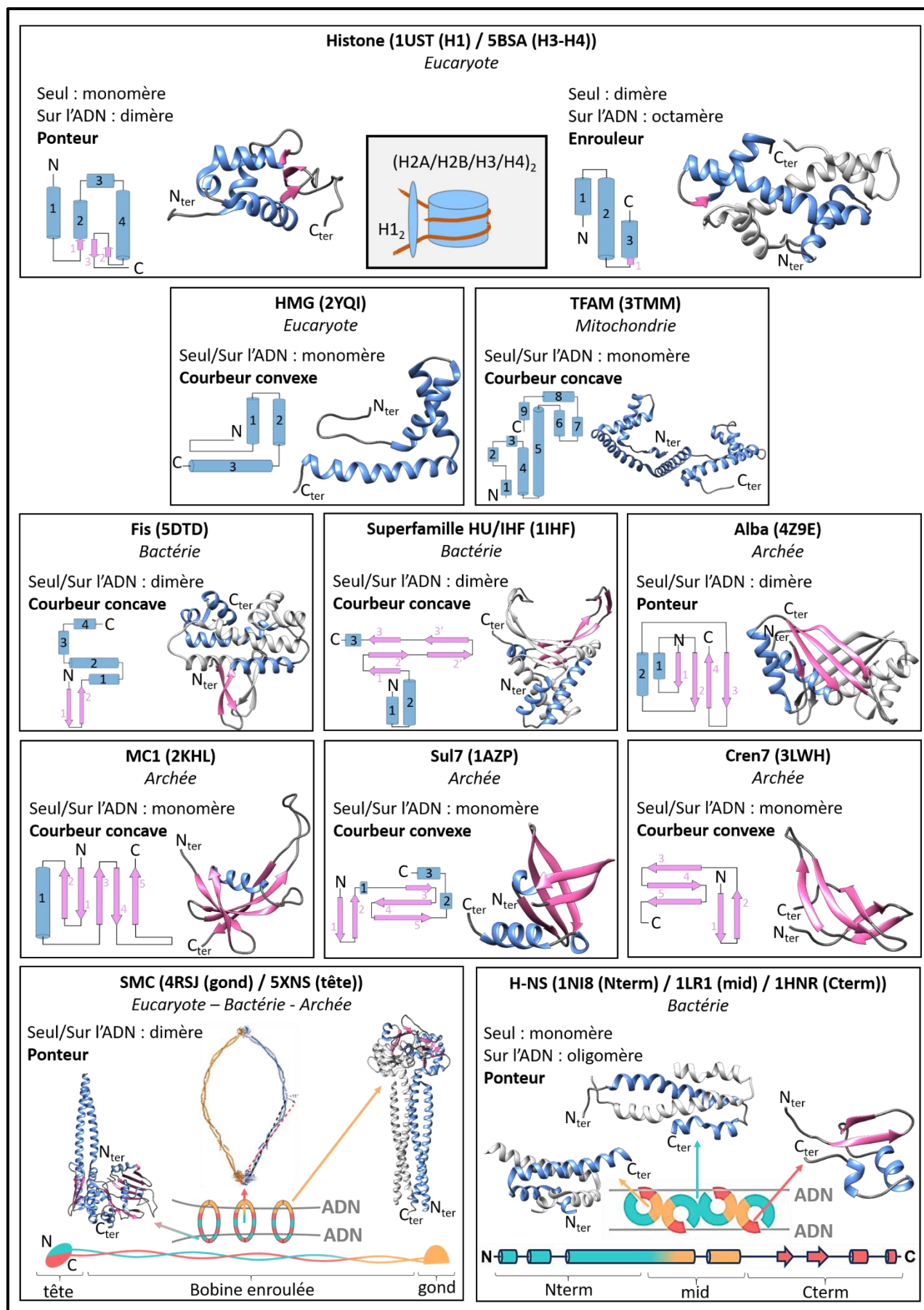


Figure 3 : Structure tridimensionnelle et diagramme de topologie des protéines architecturales majeurs du Vivant.

Chez les Bactéries, on ne retrouve pas d'histones ^[15]. La compaction de l'ADN est réalisée par des **Protéines Associées au Nucléoïde (NAPs)**. Ces protéines sont en général petites (de l'ordre de 5-15 kDa par sous-unité), abondantes ($>10^4$ protéines/cellule), souvent multimériques et présentent un caractère basique avec des points isoélectriques (pI) supérieurs à 8. Les NAPs possèdent un mode principal de compaction de l'ADN qui n'est pas toujours unique. Par exemple, parmi les protéines les plus étudiées : HU (pour « Histone-like U93 ») est principalement un courbeur d'ADN sans spécificité de séquence qui peut également enrouler ce dernier ; IHF (pour « Integration Host Factor »), un homologue structural de HU et membre de la même superfamille, courbe l'ADN en reconnaissant une séquence spécifique de nucléotides ; Fis (pour « Factor of inversion stimulation ») courbe également l'ADN et intervient dans de nombreux processus cellulaires principalement en lien avec la transcription du génome ; et H-NS (pour « Histone-like Nucleoid Structuring protein ») est un ponteur pouvant notamment contraindre des surenroulements sur l'ADN ^[7,16,17] (Figure 3). Chacune de ces protéines interagit avec l'ADN principalement sous forme dimérique, mais également à des niveaux d'oligomérisation plus grands pour certaines d'entre elles.

Chez les Archées, au moins deux types de protéines chromatinienne sont présentes pour chaque espèce présentant un mode de compaction principal différent. Les NAPs majoritaires d'Archées, indépendamment des espèces, sont : les histones, Alba, HU, MC1, Sul7, Cren7 et CC1 ^[18–20] (Figure 3). Les **histones** d'Archées sont très proches des histones d'Eucaryotes, tant par leur structure que par leur rôle. La protéine **Alba** (pour « Acetylation lowers binding affinity ») est le ponteur d'ADN majeur des Archées. On la retrouve chez toutes les Archées à l'exception de deux classes : Methanomicrobia et Halobacteria. Pour ces deux classes, la protéine **MC1** (pour « Methanogen Chromosomal protein 1 ») est présente en tant que courbeur d'ADN. **Sul7**, **Cren7** et **CC1** sont également des courbeurs d'ADN et sont retrouvées chez les espèces d'Archées dépourvues d'histones.

La **première partie** de ce manuscrit concerne la protéine **HU d'*Escherichia coli* (EchU)** et l'étude de son mécanisme d'échange de chaînes en trois étapes : passage d'une forme native dimérique à une forme intermédiaire dimérique, association de deux formes intermédiaires pour former un tétramère transitoire, dissociation du tétramère pour former un nouveau dimère natif. **L'objectif du présent travail est de décrire complètement le mécanisme d'échange de chaînes de cette protéine dimérique tant d'un point de vue structural que cinétique.**

Le **chapitre 1** expose l'état des connaissances scientifiques sur les protéines HU, en particulier celles d'*E. coli*.

Le **chapitre 2** présente l'ensemble des méthodes et le matériel utilisés pour l'étude des protéines EchU dans ce projet.

Le **chapitre 3** expose l'étude structurale et dynamique de la première étape du mécanisme d'échange de chaînes : le passage du dimère natif à un dimère dit intermédiaire.

Le **chapitre 4** présente l'étude cinétique du mécanisme d'échange de chaînes et la caractérisation de toutes les constantes d'échange.

La **seconde partie** de ce manuscrit concerne la protéine **MC1 de *Methanosarcina Thermophila*** et en particulier son interaction d'un point de vue dynamique et structural avec l'ADN. Il a été montré que la protéine induisait de fortes courbures sur des mini cercles d'ADN double-brin avec lesquels elle interagissait ^[21]. Pour mieux caractériser cette interaction, une étude complète a été menée, au sein de notre laboratoire, sur MC1 en complexe avec un ADN de 15 paires de bases (ADN^{15bp}) ^[22–25]. Pour aller plus loin dans l'analyse des interactions entre la protéine et l'ADN, un nouveau complexe, ADN^{23bp}/MC1, avec un ADN plus long a été étudié. **Cette étude n'a pu être réalisée que grâce à la mise en place d'une approche multidisciplinaire.**

Le **chapitre 1** présente le domaine des Archées et expose l'état des connaissances scientifiques sur la protéine MC1 ainsi que la description des complexes ADN/MC1 en solution par Résonance Magnétique Nucléaire.

Le **chapitre 2** correspond à la caractérisation structurale du complexe ADN^{23bp}/MC1 par l'utilisation de rayons X.

Partie 1. Vers une description complète
du mécanisme d'échange de chaînes de
la protéine *EcHU*

Chapitre 1. HU, une protéine au mécanisme d'échange complexe, essentielle à la croissance et à la survie des Bactéries

1.1. La protéine HU

1.1.1. Découverte et principales caractéristiques de la protéine

Une histone-like bactérienne

HU a été caractérisée pour la première fois en 1975 à partir d'un extrait soluble d'*Escherichia coli*, par Josette Rouvière-Yaniv et François Gros [26]. Les auteurs ont isolé cette protéine par chromatographie d'affinité sur colonne de cellulose avec greffage d'ADN (DNA-cellulose chromatography) [27], mettant ainsi en lumière l'une des principales caractéristiques de HU à savoir sa **capacité à interagir avec l'ADN**. Ils ont montré qu'il s'agit d'une **protéine dimérique de petite taille** (Masse Moléculaire : MW ~20 kDa) par deux techniques d'analyse complémentaires. La chromatographie d'exclusion stérique a permis de déterminer la valeur de la masse moléculaire de la protéine en condition native, soit 20 kDa. L'électrophorèse sur gel en condition dénaturante a, quant à elle, révélé une masse moléculaire deux fois plus faible que précédemment montrant que la protéine est dimérique à l'état natif. Au cours de cette étude, les auteurs ont quantifié HU dans les bactéries et ont montré que cette dernière était **très abondante** dans la cellule (au moins 10^4 protéines par bactérie). Ils ont déterminé sa composition en acides aminés par la méthode du biuret [28]. Il s'est avéré que HU est une protéine **riche en lysines et en alanines**, et dépourvue de cystéines. Cette méthode a été complétée par d'autres analyses, comme par exemple, des mesures d'absorbance à 280 nm pour quantifier les résidus aromatiques. Ces analyses ont montré que HU est dépourvue de tryptophane et de tyrosine. Composée à 19 % de résidus basiques (Arg, Lys) et à 18 % de résidus acides (Asp, Glu), la protéine est globalement **basique** (pI ~9,6).

Au niveau de ses fonctions, HU **stimule la transcription de l'ADN du bactériophage λ** et fixe ce dernier de façon non spécifique en formant de nombreuses boucles, des structures en épingle à cheveux et des pelotes (Figure 4) [27]. HU est classée parmi les **protéines dites de « type histone »** (histone-like) pour ses propriétés physico-chimiques (petites protéines basiques) et fonctionnelles (compaction de l'ADN) proches des histones eucaryotes.

Depuis 1975, HU a été caractérisée chez de nombreuses autres espèces et est à présent largement connue des biochimistes et des biologistes. C'est une **Protéine Associée au Nucléoïde (NAP)**, **essentielle à la croissance et à la vie des bactéries**. Elle est présente chez toutes les Bactéries (~ 3.10^4 dimères/cellule en moyenne), et chez les Euryarchées du genre *Thermoplasma* [29]. Cette protéine architecturale remplit de nombreuses fonctions en lien avec le métabolisme de l'ADN au

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

sein du nucléoïde. En plus de son rôle dans la transcription, elle est impliquée dans de nombreux autres processus comme la réplication, la réparation et la réponse à divers stress de l'ADN comme celui provoqué par une irradiation gamma par exemple [30–34].

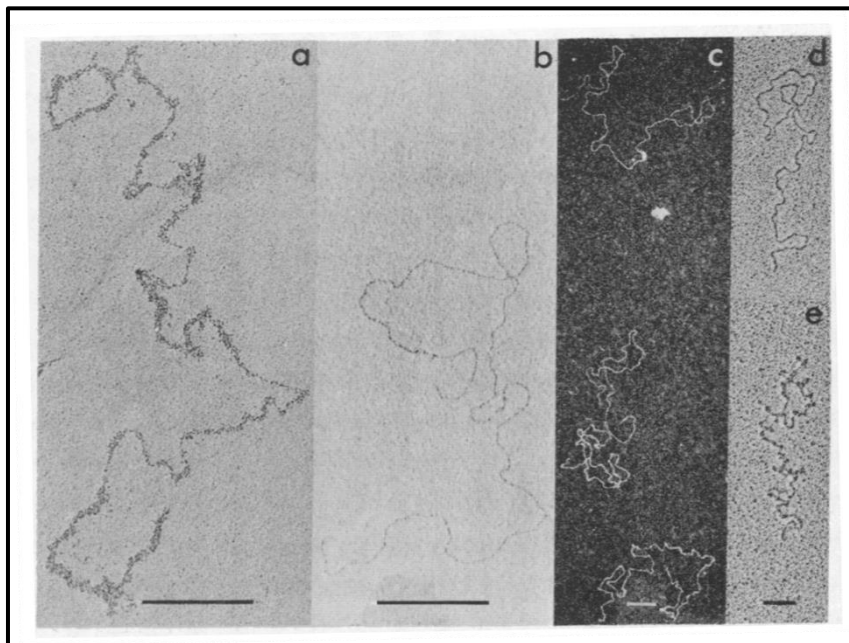


Figure 4 : Clichés de microscopie électronique de molécules d'ADN λ seules et complexées avec la protéine HU (issues de Rouvière-Yaniv et Gros, 1975) [26]. a,e) ADN complexé à HU ; b,d) ADN seul ; c) mélange d'ADN seul et d'ADN complexé à HU. L'échelle indique 100 nm.

La superfamille HU/IHF

HU appartient à la **superfamille des protéines HU/IHF** [35,36]. Cette superfamille se divise en trois groupes : les protéines HU, les protéines IHF_A et IHF_B qui partagent une identité de séquence significative [37,38]. Au-delà d'une **homologie séquentielle**, les protéines de cette superfamille partagent une **structuration spatiale particulièrement conservée** [37]. Ces protéines d'environ 10 kDa par sous-unité sont basiques et sont des **dimères**. Les protéines HU sont généralement des homodimères alors que les protéines IHF sont toujours des hétérodimères IHF $\alpha\beta$. Seules les HU d'entérobactérie et du genre *Vibrio* existent sous forme hétérodimérique HU $\alpha\beta$, c'est-à-dire qu'elles sont encodées par 2 gènes codant respectivement pour les chaînes *alpha* (HU α) et *beta* (HU β). Au niveau de la structure du dimère, chaque sous-unité protéique est composée de trois hélices α (α_1 , α_2 et α_3) et de 5 brins β (β_1 , β_2 , β_2' , β_3 et β_3') selon l'enchaînement : α_1 -t- α_2 - β_1 - β_2 - β_2' - β_3 - β_3' - α_3 (Figure 5) [39,40]. Le motif « hélice-tour-hélice », α -t- α (ou **HTH**), est régulièrement retrouvé chez les protéines d'architecture de l'ADN comme motif d'interaction protéine-protéine. Au sein du dimère, ce double motif forme, avec les trois brins antiparallèles β_1 β_2 et β_3 des deux feuilletts β formant la « selle », le **cœur hydrophobe** de la protéine. Ce cœur, ou corps, hydrophobe est **stabilisé** par des liaisons hydrogène, par des interactions hydrophobes et/ou par la formation de ponts salins inter et intra chaînes. Ces interactions sont discutées dans le Chapitre 3 lors de l'analyse des structures des

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

dimères natifs et intermédiaires. Les hélices α_3 terminales participent également à la stabilité de ce cœur en interagissant avec le haut de l'hélice α_2 de la chaîne complémentaire. Les brins antiparallèles β_2 ' et β_3 ' définissent, quant à eux et pour chaque unité protéique, un motif **en épingle à cheveux (β hairpin) au bout d'un « bras » flexible**. Cette partie de la protéine, de par sa flexibilité, n'est pas toujours visible dans les structures décrites par cristallographie des rayons X (RX) mais est essentielle pour la liaison de la protéine à l'ADN.

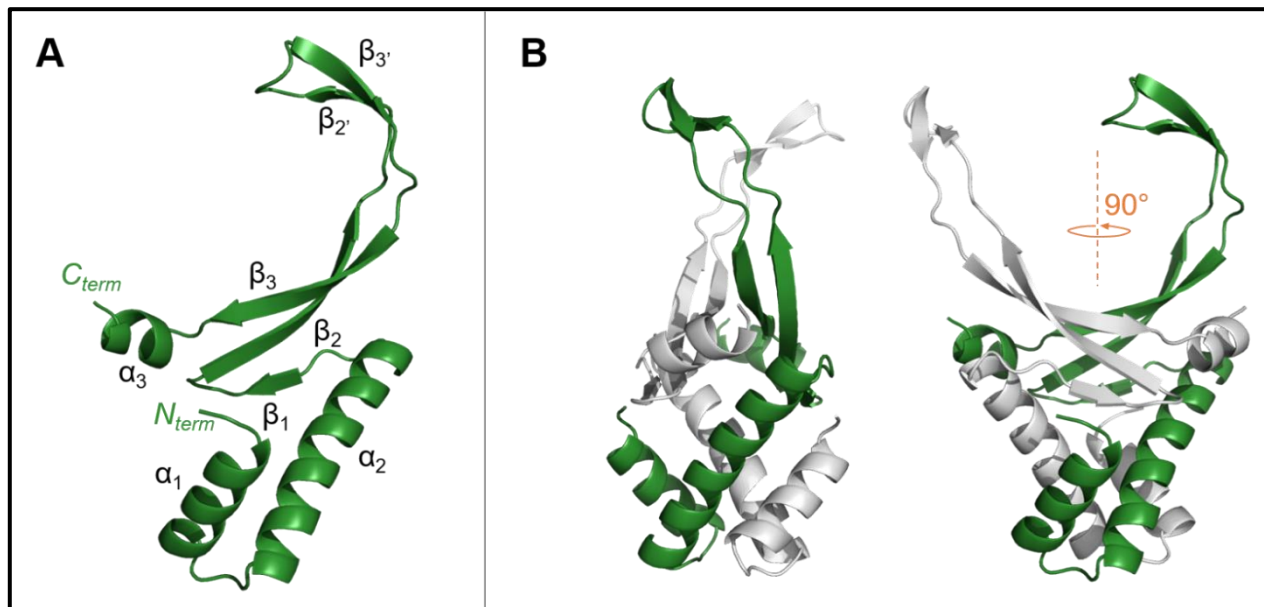


Figure 5 : **Structure tridimensionnelle des protéines de la superfamille HU/IHF.** A – structure d'une unité protéique. B – structure d'un dimère selon deux orientations.

Toutes les protéines de la superfamille HU/IHF interagissent avec l'ADN. Elles sont des régulateurs de la transcription et peuvent courber l'ADN. Cette interaction est globalement non spécifique pour les protéines HU et spécifique pour les protéines des deux autres groupes ^[41]. Les protéines HU reconnaissent préférentiellement certaines organisations (ou structurations) de l'ADN alors que les protéines IHF reconnaissent un motif séquentiel ^[42]. Une différence notable entre HU et IHF est que le premier peut interagir avec d'autres HU afin de courber l'ADN de façon coopérative et induire un surenroulement négatif de l'ADN, alors que la protéine IHF interagit toujours indépendamment des autres IHF ^[6]. Une autre grande différence est l'angle de la courbure : IHF courbe plus sévèrement l'ADN que HU (respectivement $\sim 160^\circ$ et $\sim 70^\circ$) ^[38,41]. L'interaction se fait en revanche, de façon homologue pour toutes les protéines de la superfamille qui ont une préférence pour les ADNs préalablement courbés. Elle se fait au niveau :

- des **bras flexibles** qui entourent l'ADN et les prolines présentes à leurs extrémités (Pro65 et Pro63 respectivement pour IHF et HU d'*E. coli*) qui s'intercalent entre les paires de bases dans le **petit sillon**.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

- de la « **selle** » formée par les deux **feuilletés β** à trois brins de la protéine. Ce domaine de liaison à l'ADN est nommé domaine DBD dans certains articles scientifiques. Le complexe est stabilisé par des interactions électrostatiques dues à la présence de nombreux résidus basiques (Arg et Lys).
- du corps de la protéine, formé des **domaines HTH**, qui stabilise le complexe par des interactions hydrophobes.

Les complexes d'*HU* avec l'ADN

HU interagit de façon non-spécifique avec l'ADN, mais montre des affinités différentes pour divers motifs de celui-ci. Elle interagit préférentiellement avec des **ADNs double-brins présentant une cassure** ($K_D = 8$ nM) ^[34], **des délétions** (ou « bulges ») sur un des deux brins ($K_D = 4$ nM) ^[43], et avec des ADNs en **jonction à 3 ou 4 brins** ($K_D = 4$ nM) ^[44] plutôt qu'avec des ADNs linéaires ($K_D > 500$ nM) ^[45] ou courbés ($K_D = 100$ nM) ^[43]. Outre les différences d'affinité de la protéine pour certains motifs d'ADN, HU est capable de compacter ce dernier selon deux modes : en le **courbant** et/ou en l'**enroulant** (Figure 6).

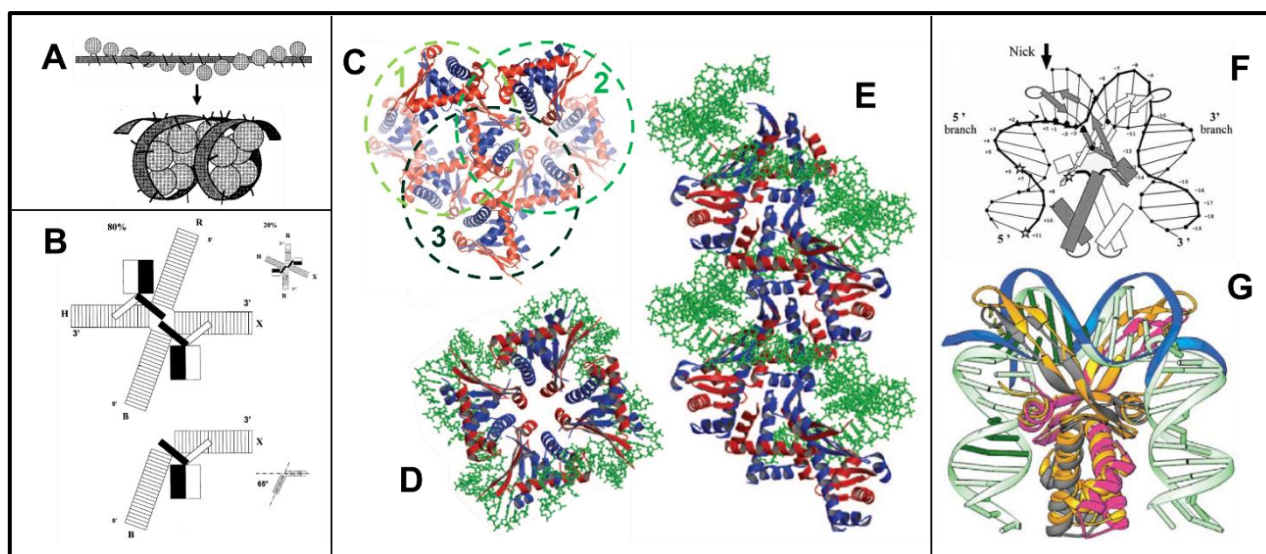


Figure 6 : Modèles de structure de HU complexé à un ADN. A – Modèle du mécanisme de surenroulement (adapté de Tanaka et al., 1995) ^[46]. B – Représentation schématique du complexe HU/ADN (adapté de Kamashev et al., 1999) ^[47] avec un ADN en jonction 4 voies (haut) et présentant une cassure simple brin (bas). C à E – Modèle de surenroulement basé sur un réseau d'empilements cristallins de la protéine (adapté de Guo and Adhya, 2007) ^[48]. C représente la vue du dessus du réseau d'empilements cristallins (1 : arrangement en unité octamérique d'une spirale multimérique droite ; 2 : arrangement en unité octamérique d'une spirale multimérique gauche ; 3 : arrangement en octamère). D représente la vue de dessus et E la vue de côté. F – Modèle du complexe avec un ADN présentant une cassure simple brin (adapté de Kamashev et al., 1999) ^[47]. G – Superposition des structures RX du complexe ADN/IHF et ADN/HU (adapté de Swinger et Rice, 2004) ^[41].

HU est principalement un **courbeur d'ADN**. En 1999, Dmitri Kamashev et ses collaborateurs parviennent à localiser la protéine HU sur l'ADN par une technique d'empreinte aux radicaux libres

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

(« phenanthroline DNA footprinting ») ^[47]. Cette technique repose sur le fait que des radicaux (OH•) sont générés et vont couper le squelette désoxyribose-phosphate de l'ADN, marqué au ³²P. Cette coupure se fait par attaque oxydative au niveau du désoxyribose (liaison C_{3'}-C_{4'}) et au niveau des bases ^[49]. Les zones protégées par l'interaction avec une protéine sont protégées de la dégradation. Cette méthode a révélé que la protéine HU interagit, par ses bras flexibles, **avec le petit sillon de l'ADN**. Les chercheurs ont également montré que, dans le cas d'un ADN ayant une cassure simple brin, les bras flexibles de HU se fixent au niveau de la coupure et le corps au niveau de l'extrémité 3' du brin clivé (Figure 6B,F). La protéine induit sur ce type d'ADN une **courbure de 65°** à l'endroit de la cassure. En 2003, Kerren K. Swinger et ses collaborateurs, complètent la caractérisation du complexe en montrant que HU est capable d'induire et/ou de stabiliser différents ADNs présentant des angles de courbure allant de **105° à 140°** ^[41,42]. Cette variabilité serait liée aux nombreux contacts entre les **9 paires de bases** de l'ADN impliquées et les bras flexibles de la protéine, ainsi qu'aux multiples contacts supplémentaires entre l'ADN et le corps de la protéine (Figure 6G).

En plus d'être un courbeur, des données cristallographiques du complexe HU $\alpha\beta$ /ADN suggèrent que HU peut contribuer au mécanisme de **surenroulement négatif de l'ADN *in vivo*** en s'oligomérisant de façon hélicoïdale le long de l'ADN ^[50]. Le packing cristallin observé dans ces structures cristallographiques suggère que les interactions au sein du complexe, entre les dimères de HU et l'ADN et entre les dimères, sont des interactions électrostatiques qui impliquent très probablement les phosphates des nucléotides et les arginines des bras de la protéine. En 1986, Steven S. Broyles et David E. Pettijohn s'intéressent au temps de demi-vie ($t_{1/2}$) du complexe et montrent que la **force ionique influe grandement sur la dissociation** de ce dernier ^[51]. Ils émettent eux-aussi l'hypothèse que HU pourrait se structurer en un cylindre autour duquel l'ADN s'enroulerait, engendrant des surenroulements, de façon toroïdale. Ils s'appuient, pour cette hypothèse, sur des données de dichroïsme circulaire. En 1995, Hiromitsu Tanaka et ses collaborateurs réalisent une étude montrant que la protéine HU peut induire des surenroulements négatifs sur un ADN circulaire indépendamment de la présence de topoisomérases, en particulier d'ADN gyrase ^[46]. HU y interagit sous la forme d'un oligomère. Il faut environ **10 dimères de la protéine pour faire un supertour d'ADN**. Plus précisément, les dimères de HU se fixent de façon non-spécifique sur un ADN quelconque toutes les **8-9 paires de bases** environ et créent des contacts avec les dimères adjacents, réduisant ainsi la taille d'un tour de double-hélice d'environ 0,5 bp par dimère de protéine. Ce mécanisme aboutit à la formation de solénoïdes gauches d'un tour tous les 10 dimères de HU (Figure 6A). Ce modèle de solénoïde est conforté par le modèle de surenroulement proposé par Fusheng Guo et Sankar Adhya en 2007 ^[48]. Ce modèle illustré sur la Figure 6E modélise l'enroulement d'un double-brin d'ADN sur la structure cristallographique d'octamères de la protéine HU (Figure 6C-E).

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

1.1.2. HU chez *E.coli*

Codage et régulation des gènes

HU est codée par **un seul gène pour la majorité des espèces bactériennes et archéennes**. Chez *E. coli* en revanche, comme chez toutes les entérobactéries ^[52] et les bactéries du genre *Vibrio* ^[53], la protéine HU est codée par deux gènes : *hupA* ^[54] et *hupB* ^[55]. Ces deux gènes codent respectivement pour les chaînes protéiques ***EchU α*** et ***EchU β*** ^[40]. HU d'*E. coli* existe ainsi sous trois formes dimériques : deux homodimères (***EchU α ₂*** et ***EchU β ₂***) et un hétérodimère (***EchU $\alpha\beta$***) ^[56].

Sur le chromosome bactérien, les gènes *hupA* et *hupB* sont à des positions relativement éloignées et ne font donc pas partie du même opéron (**90,5 min** ^[54] et **9,7 min** ^[55] respectivement pour *hupA* et *HupB* - Figure 7A). De ce positionnement, il est déduit que seuls les homodimères peuvent être immédiatement formés suite à l'**expression non-concomitante des gènes *hup*** et que l'hétérodimère serait formé dans un second temps à partir des homodimères. En 2003, il a été montré que l'hétérodimère se formait **spontanément** à partir des deux homodimères ^[57]. L'échange de chaînes peut se réaliser *in vitro* à faible (T = 4 °C) comme à haute température (T = 37 °C) en mélangeant simplement les homodimères purifiés. Cette formation spontanée serait liée à une affinité plus grande de la chaîne *EchU β* pour son homologue *EchU α* que pour elle-même ($K_D[\beta \rightarrow \alpha] < K_D[\beta \rightarrow \beta]$) ^[58].

Au niveau de la régulation de la transcription des gènes, *hupA* est précédé d'un unique promoteur (**P**), alors que *hupB* est précédé de 4 promoteurs (**P1, P2, P3, P4**) dont deux peu actifs (P1 et P4). La transcription de ces gènes varie au cours du cycle de croissance de la bactérie et en réponse à certains stress comme une irradiation ou un choc froid ^[59–61]. La transcription des gènes est **autorégulée de façon négative**, avec une répression plus importante du gène *hupA* que du gène *hupB* ^[59]. Elle est également régulée par les protéines **CRP** (pour « Cyclic AMP Receptor Protein ») et **Fis** (Figure 7B). La protéine CRP, activateur de classe I, **active** la transcription de *hupA* et celle de *hupB* au niveau du promoteur P3. La protéine **Fis**, de son côté, **active** le promoteur de *hupA* mais **réprime** la transcription de *hupB* au niveau des promoteurs P2 et P3. Le promoteur P4 de *hupB* est peu actif. Il peut toutefois être activé sous certaines conditions et en particulier dans le cas d'un choc froid ^[62]. Le promoteur P1 est encore moins actif dans le sens où les transcrits issus de ce promoteur ne sont pas détectables.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine EchU

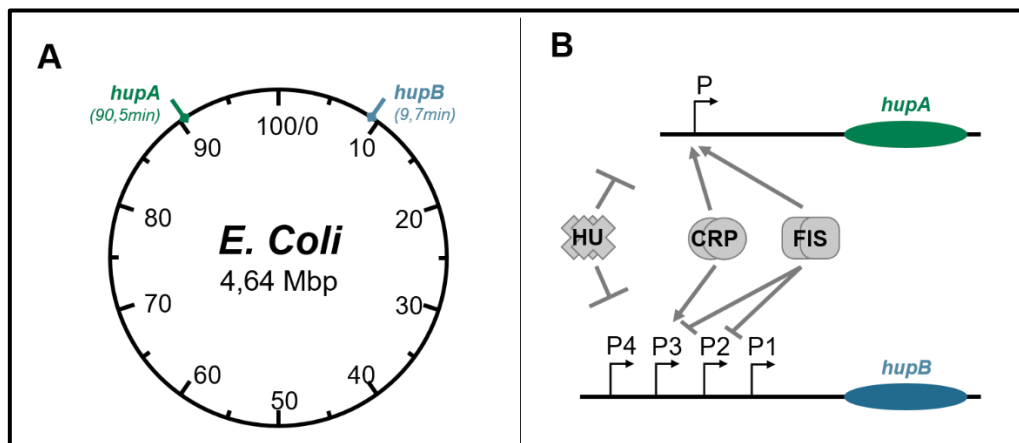


Figure 7 : **Schéma simplifié présentant les gènes *hupA* et *hupB*.** A – Leur disposition sur la carte du génome de *E. coli*. B – Leur régulation par les protéines HU, CRP et Fis. Les flèches grises indiquent une activation par la protéine concernée et les traits gris indiquent une répression.

La prédominance des dimères de EchU au cours du cycle bactérien

La courbe de croissance du cycle bactérien se découpe en **4 phases majeures**. La première phase est une **phase de latence**. Elle correspond au temps nécessaire aux bactéries pour s'adapter à leur nouveau substrat. La seconde phase est la **phase de croissance exponentielle**. Elle correspond à une augmentation du nombre de bactéries dans le milieu. La troisième phase est la **phase stationnaire**. Elle correspond à un appauvrissement du milieu de culture où la multiplication des bactéries compense leur perte par autolyse. La dernière phase est la **phase de déclin**. Cette phase commence lorsque toutes les ressources nutritives du milieu sont épuisées. Il persiste une croissance mais qui ne compense plus la mort des bactéries.

Pour ce qui est de la protéine HU d'*E. coli*, les proportions des sous-unités *EchU α* et *EchU β* varient au cours du cycle bactérien (Figure 8A) ^[63]. **Durant la phase de croissance exponentielle, quatre fois plus de sous-unités *EchU α* que *EchU β* sont produites, alors qu'en phase stationnaire les sous-unités sont produites en quantité égale.** Cette différence de proportion est probablement liée à la régulation de la transcription des gènes par les protéines CRP et Fis. En effet, en phase exponentielle, la protéine Fis est très abondante. Comme Fis active la transcription de *hupA* et réprime celle de *hupB*, on comprend aisément que le dimère majoritaire pendant cette phase est l'homodimère *EchU α ₂*. En phase stationnaire, en revanche, Fis devient très peu abondant, et les deux gènes sont activement transcrits, activés par CRP. La proportion équivalente des deux sous-unités abondamment synthétisées et la formation spontanée de l'hétérodimère rendent le dimère *EchU $\alpha\beta$* majoritaire dans la cellule. Cette nouvelle espèce améliore la survie de la bactérie en phase stationnaire.

Dans des conditions standards de culture (milieu LB, pour « Luria Bertani Broth » à 37 °C), **l'homodimère *EchU β ₂* n'est pas indispensable à la croissance ni à la survie des bactéries** ^[63]. Ce dimère devient cependant essentiel à leur survie en cas de choc froid ^[62]. En effet, si au cours

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

de son cycle, la culture de bactéries est brutalement soumise à une faible température (10 °C), l'homodimère *EchU* β_2 devient l'espèce majoritaire dans la cellule (Figure 8B). Dans le cas d'un choc froid, la quantité de protéine Fis diminue ^[64]. La transcription du gène *hupA* est dans ces conditions totalement réprimée. De plus les transcrits de *hupA* sont peu stables à basse température. La transcription reste inchangée au niveau des promoteurs P2 et P3 de *hupB* et le promoteur P4 est lui, plus actif. **La répression de *hupA* et l'activation de *hupB* en cas de choc froid, font que l'homodimère *EchU* β_2 devient majoritaire.** *EchU* β_2 est donc essentiel à l'adaptation de *E. coli* suite à un choc froid. Une fois la culture replacée dans des conditions standards, la régulation des gènes est de nouveau réalisée par les protéines CRP et Fis, et l'homodimère *EchU* β_2 redevient non essentiel pour les cellules.

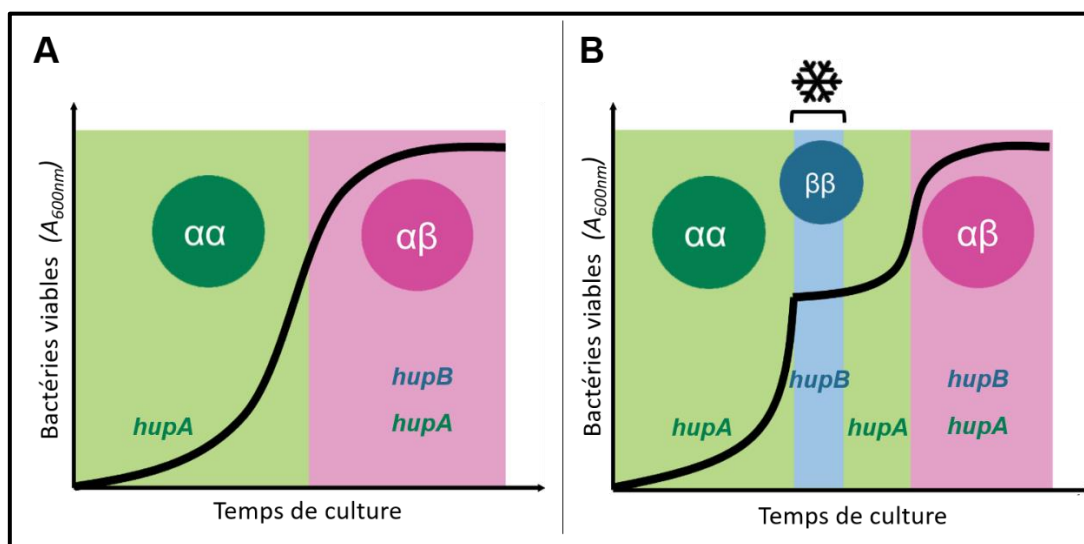


Figure 8 : **Prédominance des dimères de *EchU* au cours du cycle bactérien.** A – Dans des conditions standards (37 °C). B – Phase adaptative suite à un choc froid (10 °C, ❄️).

La prédominance des différents dimères de HU au cours du cycle bactérien est liée à des **fonctions biologiques propres** à chacune des formes dimériques. Ainsi la forme homodimérique *EchU* α_2 est importante pour la phase exponentielle de croissance des bactéries, la forme hétérodimérique est importante pour leur survie en phase stationnaire de croissance, et l'homodimère *EchU* β_2 est essentiel pour leur survie en cas de choc froid. Ces fonctions propres à chaque dimère doivent être liées à une différence au niveau de leur **interaction avec l'ADN**, de leur **structure** et/ou de leurs **propriétés physico-chimiques** qu'il faut caractériser pour mieux comprendre les **mécanismes cellulaires médiés par cette protéine**.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

1.1.3. Structure de *EchU*

Les **structures tridimensionnelles (3D)** des trois dimères de *EchU* ont été résolues par cristallographie aux rayons X : *EchU* α_2 en 2002 (pdb :1MUL) ^[57], *EchU* β_2 en 2014 (pdb :4P3V) et *EchU* $\alpha\beta$ en 2006 (pdb :2O97) ^[48].

Les chaînes protéiques *EchU* α et *EchU* β sont constituées toutes les deux de **90 acides aminés**, et possèdent de nombreux résidus lysine et alanine (respectivement 11 et 15 pour *EchU* α , et 9 et 19 pour *EchU* β) ^[40]. Elles ont des masses molaires et des points isoélectriques proches (9,53 kDa / 9,57 et 9,23 kDa / 9,69 respectivement pour *EchU* α et *EchU* β), partagent **70 % d'identité** et **90 % d'homologie** de séquence (Figure 9). Les structures des trois dimères de *EchU* sont, comme attendu, **homologues** et correspondent à celles de toutes les protéines de la superfamille HU/IHF (Figure 5).

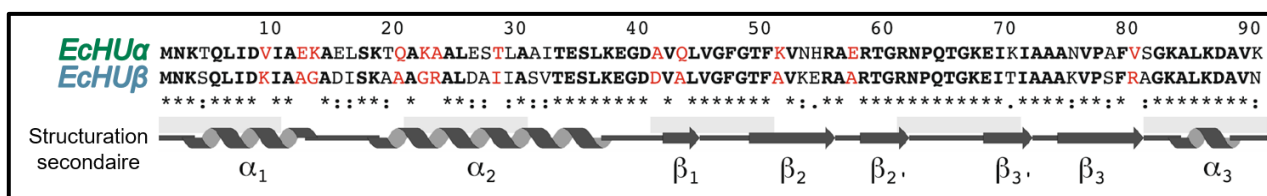


Figure 9 : Séquence en acides aminés et éléments de structure secondaire des chaînes protéiques *EchU* α et *EchU* β . Les résidus identiques sont écrits en gras et représentés par « * », les résidus homologues sont représentés par « : » et les résidus différents sont écrits en rouge (adapté de Le Meur, et al., 2015) ^[65].

1.1.4. Interaction de *EchU* avec différents types d'ADN

De nombreuses études ont été menées sur *EchU* et sur ses interactions avec différentes structures d'ADN en se focalisant sur l'hétérodimère *EchU* $\alpha\beta$. *EchU* $\alpha\beta$ se fixe de manière modérée sur les brins d'oligonucléotides, avec une légère préférence pour les ADNs simple brin et les ARNs plutôt que pour les ADNs double-brins ^[66]. *EchU* α_2 présente une affinité légèrement inférieure à celle de *EchU* $\alpha\beta$ mais beaucoup plus grande que *EchU* β_2 pour les ADNs double-brins ($K_D[\alpha\beta] \leq K_D[\alpha\alpha] \ll K_D[\beta\beta]$). Il est à noter que **les homodimères ont bien plus d'affinité pour les ADNs simple brin que n'en a l'hétérodimère** ^[67]. La protéine HU se lie préférentiellement à des ADNs courbés et à des ADNs cruciformes par exemple, plutôt qu'à des ADNs linéaires ^[43,58]. Cette étude a été complétée par des données sur les trois dimères (Figure 10). L'évaluation précise des constantes de dissociation (K_D) des différents dimères de HU d'*E. coli* indique que l'hétérodimère *EchU* $\alpha\beta$ et l'homodimère *EchU* α_2 ont une affinité du **même ordre de grandeur** pour les divers motifs d'ADN testés. Toutefois, la constante de dissociation de l'hétérodimère est le plus souvent plus faible. L'homodimère *EchU* β_2 possède, quant à lui, une **affinité très inférieure aux deux autres dimères** pour les divers motifs d'ADN. De manière général, on observe une différence de K_D d'au moins un ordre de grandeur dans chacun des cas testés. La seule exception est dans le **cas**

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

de l'ADN cruciforme avec une faible concentration en sels (15 mM NaCl). Dans ces conditions, la constante de dissociation de *EchU* $\alpha\beta$ est du même ordre que celle de *EchU* α_2 pour le motif d'ADN. Les homodimères sont tous les deux bien moins affins que l'hétérodimère dans ces mêmes conditions.

Table 1. DNA binding constants (K_D), of HU $\alpha\beta$, HU α_2 and HU β_2 protein for linear DNA, cruciform DNA, nick DNA, gap-1 DNA and gap-2 DNA under stringent (200 mM NaCl) and non-stringent (15 mM NaCl) conditions

A. High salt conditions (200 mM NaCl)					
	Linear DNA	Cruciform DNA	Nick DNA	Gap-1 DNA	Gap-2 DNA
$\alpha\beta$	2500	20.0	40.0	20.0	8.0
α_2	9000	25.0	40.0	11.0	3.0
β_2	66,000	100.0	1200.0	1200.0	400.0
B. Low salt conditions (15 mM NaCl)					
	Linear DNA	Cruciform DNA	Nick DNA	Gap-1 DNA	Gap-2 DNA
$\alpha\beta$	400	7.5	4.0	2.0	1.0
α_2	1000	25.0	7.5	2.0	1.5
β_2	29,000	50.0	330.0	47.0	47.0

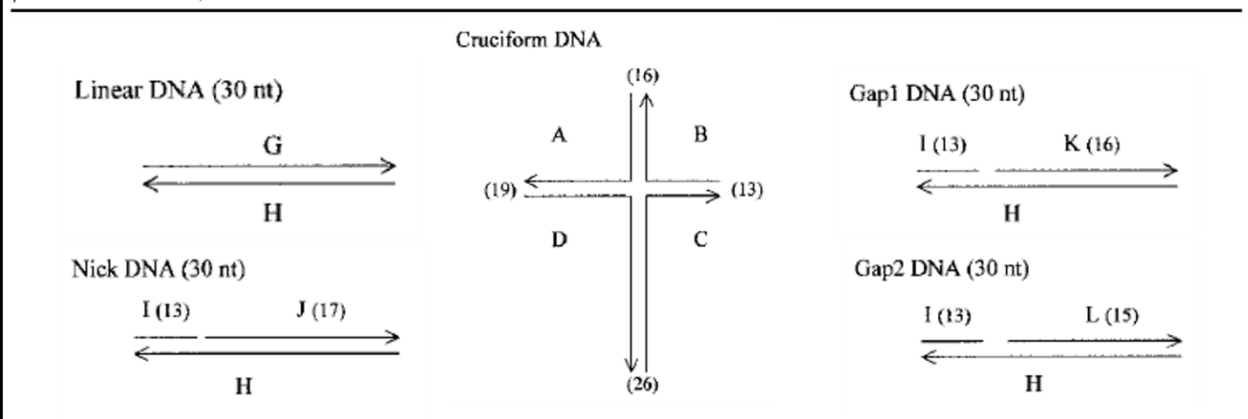


Figure 10 : Affinité des protéines *EchU* $\alpha\beta$, *EchU* α_2 et *EchU* β_2 à différentes structures d'ADNs synthétiques (K_D) : ADN linéaire (linear DNA), ADN cruciforme (cruciform DNA), ADN présentant une coupure simple brin (nick DNA), ADN présentant une (gap-1 DNA) ou deux (gap-2 DNA) délétions simple brin en conditions sévères (200 mM NaCl) et non-sévères (15 mM NaCl). Les flèches indiquent les terminaisons 3' des oligonucléotides (issus de Pinson et al., 1999) ^[58].
Haut - Table des K_D . B - représentation des structures d'ADN.

1.1.5. Propriétés physico-chimiques de *EchU*

Les dimères de *EchU* présentent des propriétés physico-chimiques différentes qui leur permettent d'interagir différemment avec l'ADN, notamment des **propriétés thermodynamiques**. La température de transition (T_m , correspondant à la **température à laquelle 50 % des protéines sont dénaturées**, a été mesurée pour chaque dimère de *E. coli* par calorimétrie (DSC) et indique non pas une mais **deux températures de transition pour chacun des dimères : T_{m1} et T_{m2}** (Tableau 1) ^[57]. Cette étude met en évidence une **dénaturation** de la protéine **en deux étapes**. Les dimères de HU passent d'abord d'un état natif (N_2) à un état intermédiaire thermodynamiquement stable, puis de cet état intermédiaire à un état totalement dissocié et dénaturé (2D). Dans ce processus, l'état intermédiaire peut être dimérique (I_2) ou monomérique (2I). Afin de le déterminer, les auteurs ont fait varier la concentration en protéine lors de leurs expériences. En effet, dans le cas d'une

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

transition du type bimoléculaire $X_2 \leftrightarrow 2X$, la température de transition varie en fonction de la concentration en protéine, alors qu'elle reste constante dans le cas d'une transition de type monomoléculaire $X_2 \leftrightarrow X'_2$. En passant de 1 mg.mL⁻¹ à 4,5 mg.mL⁻¹ de protéine, il a été montré, pour chacun des dimères, une faible variation de Tm1 ($\Delta T_{m1} \sim 1$ °C) et une plus forte variation de Tm2 ($\Delta T_{m2} \sim 6$ °C) (Tableau 1). Le processus de dénaturation de la protéine passe donc par une **forme intermédiaire dimérique et stable**, selon le schéma : $N_2 \leftrightarrow I_2 \leftrightarrow 2D$ où N_2 correspond à l'état natif, I_2 à un état intermédiaire dimérique et 2D aux monomères déstructurés (D pour dénaturé). Les valeurs des Tm obtenues pour *EchU* $\alpha\beta$ et *EchU* α_2 sont très proches (~ 3 °C de différence) et plus élevées que celles de *EchU* β_2 (au moins 10 °C de différence).

NaCl		<i>EchU</i> α_2		<i>EchU</i> β_2		<i>EchU</i> $\alpha\beta$	
		1 mg/ml	4.5 mg/ml	1 mg/ml	4.5 mg/ml	1 mg/ml	4.5 mg/ml
0.2 M	t_{m1}	41	40	27	26	38	41
	Δt_{m1}		-1		-1		+3
	ΔH_1	52.5	51.0	44.6	45.7	40.3	46.8
	ΔS_1	167	162	148	152	124	149
	ΔC_{p1}	0.9	1.3	1.3	1.2	1.4	1.5
	t_{m2}	50	57	56	59	53	62
	Δt_{m2}		+7		+3		+9
	ΔH_2	18.8	14.0	44.3	20	19.1	26.3
	ΔS_2	39	25	115	42	34	60
	ΔC_{p2}	1.5	1.3	0.6	1.5	1.4	1.5
1 M	t_{m1}	59		50		60	
	ΔH_1	49.7		39.3		52.4	
	ΔS_1	148		120		156	
	ΔC_{p1}	0.55		0.9		0.59	
	t_{m2}	69		65		72	
	ΔH_2	42.8		33.7		57.5	
	ΔS_2	104		78		148	
	ΔC_{p2}	0.55		0.9		0.2	
	Δt_{m1}^*	+19		+23		+22	
	Δt_{m2}^*	+20		+9		+19	

$\Delta t_{ma} = t_{ma}$ difference between 1 mg/ml and 4.5 mg/ml of HU dimer and $\Delta t_{ma}^* = t_{ma}^*$ difference between 0.2 M and 1 M NaCl. t_m , Δt_m and Δt_m^* are in °C, ΔH (kcal mol dimer⁻¹), ΔS (cal K⁻¹ mol dimer⁻¹), ΔC_p (kcal K⁻¹ mol dimer⁻¹); ΔH , ΔS , ΔC_p values are for 37 °C.

Tableau 1 : Paramètres thermodynamiques des dimères de *EchU* déterminés par calorimétrie (DSC). Les valeurs des températures de transition des deux étapes du dépliement de la protéine (Tm1 et Tm2) en présence de 200 mM de chlorure de sodium sont mises en évidence dans l'encadré en bleu (adapté de Ramstein et al, 2003) ^[57].

1.2. Mise en évidence d'une conformation intermédiaire d'*EchU*

1.2.1. Premières descriptions structurales de l'intermédiaire

En combinant les données précédentes de DSC à une étude de dichroïsme circulaire (CD), un **premier modèle structural de l'état intermédiaire de HU d'*E. coli*** a été proposé ^[57]. Pour ce faire, des spectres CD à basse température (entre 4 °C et 10 °C) et à une température correspondant à 10 °C au-dessus de Tm1 pour chaque dimère ont été collectés. Le pourcentage d'éléments de structure secondaire contenus dans chaque dimère a été évalué à partir des spectres de CD et par la méthode de prédiction de Johnson (Tableau 2).

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EcHU*

Les structurations secondaires des trois dimères, obtenues à basse température, sont homologues à celles retrouvées dans la structure cristallographique d'*EcHU α_2* (pdb : 1MUL). On retrouve dans chacun de ces quatre cas environ 45 % de structuration en hélice α et environ 19 % de structuration en feuillet β . Les données obtenues par dichroïsme circulaire indiquent plus de structuration en tour et en autres structurations pour les dimères en solution que pour le cristal d'*EcHU α_2* , probablement parce que le signal CD des hélices écrase le signal des brins β . A haute température, on note une **perte d'hélicité de l'ordre de 50 % à 60 %** (60,7 % pour *EcHU α_2* , 54,2 % pour *EcHU β_2* et 50,8 % pour *EcHU $\alpha\beta$*) par rapport à celles des dimères à basse température. On peut également noter une légère augmentation de la structuration en feuillets β qui se rapproche de celle observée dans la structure cristalline (probablement que la diminution du pourcentage d'hélice α permet de mieux estimer la valeur du pourcentage de feuillet β).

		Secondary structure content (%)			
Protein		α -Helix	β -Sheet	Turn	Remainder
<i>CD calculated</i> ^a					
<i>EcHUα_2</i>	Native ^b	43.2	16.7	17.1	22.6
	50 °C ^d	17.0	20.7	17.6	44.2
<i>EcHUβ_2</i>	Native ^c	54.8	20.5	4.7	20.1
	43 °C ^d	25.1	22.6	20.3	31.6
<i>EcHU$\alpha\beta$</i>	Native ^b	42.7	15.8	18.2	22.9
	50 °C ^d	21.0	21.5	18.8	38.2
Structural model ^e		43.8	21.1	15.7	21.4

^a From the method of Johnson.
^b Between 4 °C and 10 °C.
^c At 4 °C.
^d Temperature corresponding to (t_{m1} + above 10 °C).
^e Determined from the crystal structure of the *EcHU α_2* homodimer. For this calculation, we assumed that residues in the protein arms that are not visible in the density map contain essentially random coil residues.

Tableau 2 : Eléments de structuration secondaire des dimères calculées par dichroïsme circulaire (adapté de Ramstein et al., 2003) ^[57]. La perte d'hélicité est mise en évidence par l'encadré bleu.

A partir de toutes ces données, les auteurs ont proposé un **premier modèle structural du mécanisme de dénaturation** de la protéine (Figure 11). Dans ce mécanisme en deux étapes, la conformation native correspond à celle obtenue par cristallographie. Pour la conformation intermédiaire, les auteurs suggèrent que la forte perte d'hélicité vient de la perte de la grande hélice (α_2) engendrant une ouverture partielle de la cavité hydrophobe de la protéine. La dissociation aboutit à deux monomères totalement déstructurés.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

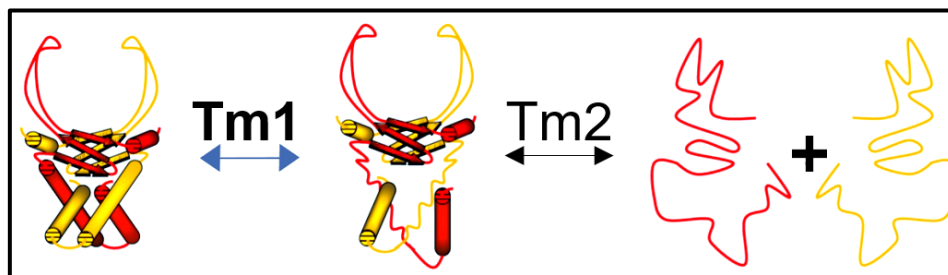


Figure 11 : Schéma du dépliement de *EchU* basé sur les données de CD et de DSC (issu de Ramstein et al., 2003) ^[57].

A la suite de ce premier modèle, une description plus fine de la transition de l'état natif N_2 à l'état intermédiaire I_2 de *EchU* β_2 a été proposée sur la base de simulations de dynamiques moléculaires (MD) ^[68]. La technique employée pour cette étude est la modélisation moléculaire « tout atome » à très haute température. Cette technique permet d'étudier le mouvement de chaque atome sur une échelle de temps de l'ordre de la dizaine de nanosecondes. Le modèle de *EchU* β_2 , représentant 58 016 atomes, a été étudié à des températures de 127 °C et de 227 °C. Il était nécessaire pour l'étude d'augmenter fortement la température afin de palier à l'échelle de temps atteignable par les calculs de dynamique moléculaire. La Figure 12 présente les résultats obtenus pour l'une des simulations à 127 °C (modèle $C_4(MD_{1^{400}})$) et nous permet de suivre l'évolution du dimère sur 24 nanosecondes. A t_0 , les sous-unités protéiques de *EchU* β_2 sont symétriques mais évoluent indépendamment au cours de la simulation. Le modèle placé à très haute température montre que les feuillets β de la « selle » de la protéine sont globalement peu déstructurés au cours de celle-ci. Cependant, on peut noter la disparition du brin β_1 pour l'une des deux sous-unités (soit *EchU* β). Cette simulation met en évidence une **perte totale des brins β_2 et β_3** appartenant aux bras flexibles de la protéine, une **déstructuration partielle** au cours du temps des **hélices α_1 et α_2** de chaque sous-unité et une **perte totale de l'hélice terminale α_3** .

Dans ce modèle, **l'hélice terminale serait la clef structurale permettant le déclenchement de l'ouverture de la cavité hydrophobe de HU.**

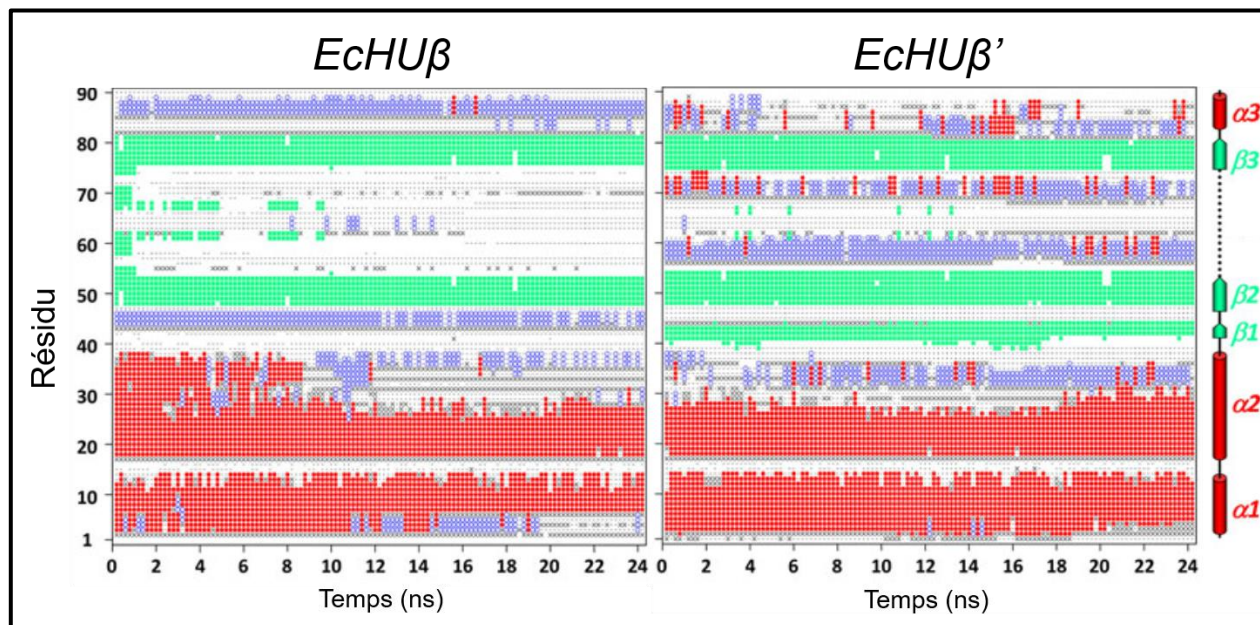


Figure 12 : **Suivi de l'évolution au cours du temps des éléments de structure secondaire de chaque résidu dans *EchUβ₂*** (adapté de Garnier et al., 2011) ^[68]. *EchUβ* et *EchUβ'* correspondent aux sous-unités protéiques de *EchUβ₂*. Les structurations secondaires principales sont représentées par des points rouges pour l'hélice α , des points bleus pour les hélices 3_{10} et π , et par des points verts pour les feuillets β .

1.2.2. Première caractérisation de l'échange entre les conformations, par RMN

Les déplacements chimiques (CS) du squelette peptidique, propres aux différents dimères de *EchU*, ont été déterminés par Résonance Magnétique Nucléaire (RMN) dans leur état natif ^[65]. L'évolution des valeurs des déplacements chimiques $^1\text{H}_\text{N}$ et $^{15}\text{N}_\text{H}$ des deux homodimères en fonction de la température a été suivie. Il a été observé l'apparition de nouveaux signaux sur les spectres ^{15}N -HSQC dans le cas de chaque dimère. Ces nouveaux signaux, augmentent en intensité avec l'élévation de la température alors que d'autres, correspondant aux résidus dans leur conformation native, diminuent d'intensité. Une étude préliminaire de dynamique (échange ZZ) a permis d'identifier ces nouveaux signaux comme étant ceux de la conformation intermédiaire I_2 ^[69]. Cette étude préliminaire a permis de commencer l'attribution séquentielle de la protéine dans sa conformation intermédiaire et d'apporter une information sur la vitesse d'échange entre les deux conformations. **La vitesse de l'échange N_2/I_2 est de l'ordre de la seconde⁻¹ (s^{-1}) pour l'homodimère *EchUα₂* à 37 °C. L'échange N_2/I_2 dans l'homodimère *EchUβ₂* n'a pas pu être quantifié pour *EchUβ₂* à 27 °C car il est plus rapide que pour l'homodimère *EchUα₂*.**

L'attribution séquentielle à haute température des homodimères a pu être complétée. **37 % des résidus** présentent un déplacement chimique différent au niveau du squelette peptidique entre les deux conformations en particulier au niveau de l'hélice α_3 . La prédiction des angles dièdres φ et ψ a été réalisée par DANGLE ^[70] sur la conformation intermédiaire. Elle prédit la **perte totale de l'hélice terminale**, ce qui concorde avec le modèle précédemment proposé par Norbert Garnier et ses

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

collaborateurs ^[68]. En revanche, il n'y a aucune indication de la perte des hélices α_1 et α_2 . Concernant les feuillets β , les données de prédiction montrent une perte totale du brin β_3 de la « selle » dans le cas de *EchU* α_2 et une formation partielle dans le cas de *EchU* β_2 . Les brins β_1 et β_2 semblent, pour leur part, maintenus.

1.2.3. Proposition d'un modèle de la conformation intermédiaire combinant les données RMN et MD

Une étude par simulation « gros grains » de dynamiques moléculaires et les données de RMN précédentes ont permis de proposer un modèle de la transition N_2/I_2 plus abouti. La dynamique d'échange entre les deux conformations étant de l'ordre de la milliseconde, les simulations tout atome auraient requis des températures bien trop élevées pour modéliser la transition. Pour les simulations « gros grains », ce sont des ensembles d'atomes qui sont modélisés par des sphères, appelées « grains » et chaque ensemble possède les propriétés équivalentes (la polarité notamment) au groupe d'atomes qu'il modélise. Avec ce type de simulation, la résolution atomique est perdue mais l'échelle de temps accessible est bien plus grande : de l'ordre de la microseconde à la milliseconde. En contrepartie, la méthode ne permet pas de simuler des pertes de structuration secondaire. La dynamique structurale de chacun des homodimères a été simulée sur une durée de dix microsecondes pour des températures de 77 °C et 97 °C. Lors de ces simulations, **les hélices du domaine HTH sont suivies par mesure de distances intra chaînes** entre les résidus des hélices α_1 et α_2 (résidus 3-25 et 7-22) **et inter chaînes** entre les résidus de l'hélice α_1 et α_2' (résidus 6-31' et 14-23') et entre les résidus de l'hélice α_1 et du brin β_1' (résidus 1-40' et 3-44'). Elles sont également suivies par la mesure de l'angle entre les hélices α_1 et α_2 des deux chaînes (angles α_1 - α_2 , α_1 - α_2' et α_2 - α_2'). Les résultats de simulation montrent un écartement des deux domaines HTH (**de 5 Å à 10 Å**) et un changement dans l'orientation des hélices de chaque domaine d'**environ 20°** (Figure 13A). Du fait du maintien des structures secondaires pendant la simulation, l'hélice α_3 est maintenue. Les auteurs notent cependant une **augmentation significative de la distance entre l'hélice α_3 et le haut de l'hélice α_2** pour une des deux chaînes à 97 °C, dans le cas de *EchU* β_2 . Ce mouvement précède celui du domaine HTH de cette même chaîne. Cette étude permet de proposer un nouveau modèle de la conformation intermédiaire montrant une **ouverture importante de la cavité hydrophobe en accord avec les données RMN** (Figure 13B).

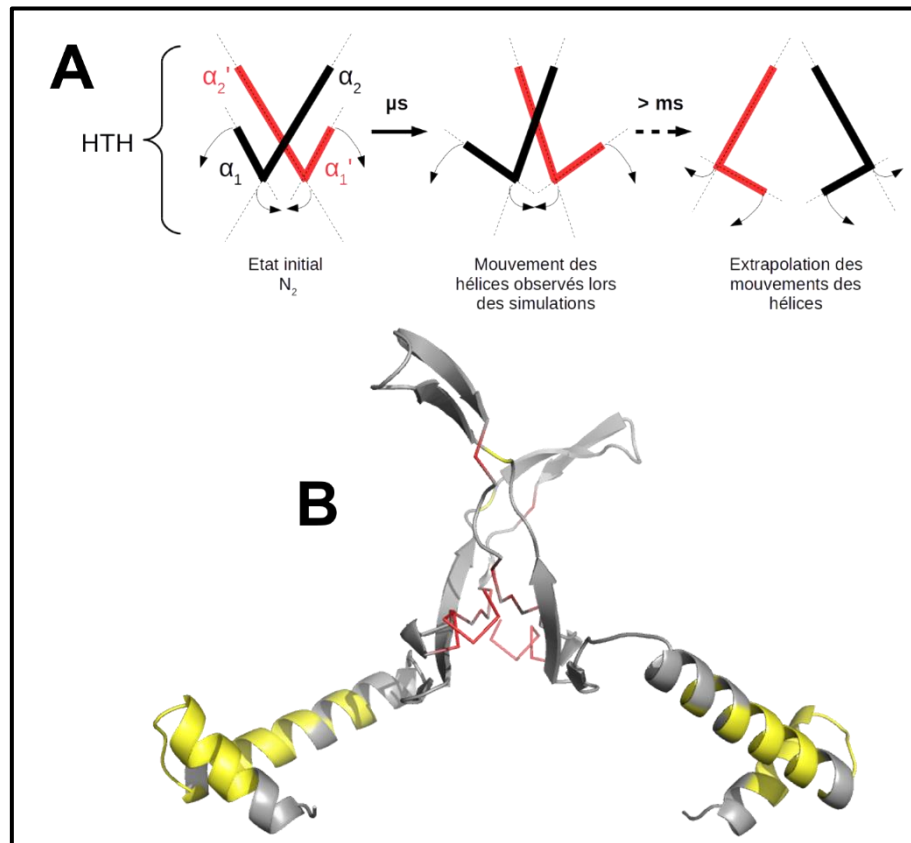


Figure 13 : **Modèle structural de l'intermédiaire déduit des expériences RMN et de simulations « gros grains »** (adapté de la thèse de R. Le Meur, 2015) ^[69]. A – Mouvements des hélices observés en simulation (μs) et extrapolation (ms). B – Modèle de la conformation intermédiaire de $ECHU\alpha_2$.

1.3. L'hypothèse d'une forme tétramérique transitoire dans la formation de l'hétérodimère

1.3.1. Hypothèse sur le mécanisme de formation de l'hétérodimère

In vivo comme *in vitro*, HU est un dimère qui peut s'oligomériser lorsqu'il est impliqué dans un complexe avec l'ADN. Étant donné l'éloignement entre les deux gènes codant pour les chaînes de la protéine, celle-ci ne peut logiquement être **synthétisée que sous la forme d'homodimères**. De ce fait, on peut aisément proposer que l'obtention de l'hétérodimère dans la cellule proviendrait d'une combinaison/dissociation des homodimères. Cette hypothèse est d'autant plus probable qu'*in vitro*, lorsque les homodimères sont mis en présence l'un de l'autre, **l'hétérodimère est formé spontanément** ^[57]. Pour expliquer cette formation spontanée, deux voies sont envisageables : soit les homodimères se dissocient puis leurs chaînes se recombinent en hétérodimères, soit les homodimères se combinent en oligomères puis se dissocient en hétérodimères. D'après les données thermodynamiques, la seconde voie est la plus probable, énergétiquement parlant. En effet, il a été montré précédemment, que pour passer d'une forme dimérique à une forme

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EcHU*

monomérique la protéine a besoin d'une quantité de chaleur importante ^[57]. La nature préférant des mécanismes peu gourmands en énergie, la voie la plus probable est celle du **passage par un état oligomérique**, à caractériser, **des homodimères avant de se dissocier en hétérodimères**.

1.3.2. L'oligomère transitoire

L'oligomère le plus simple pour le mécanisme proposé serait un **tétramère formé de deux chaînes *EcHU α* et de deux chaînes *EcHU β*** . Cette forme tétramérique n'a pour le moment, pas été caractérisée ni observée d'après la littérature scientifique.

Dans le cadre de travaux préliminaires de mon équipe d'accueil, l'équilibre conformationnel de HU et en particulier la forme intermédiaire I_2 de chaque homodimère pourrait être la clef pour la formation du tétramère transitoire (Figure 14). Dans ce modèle, l'hétérotétramère serait **stabilisé par des interactions entre les feuillets β** des dimères sous leur forme intermédiaire I_2 , impliquant notamment **les phénylalanines** présentes sur chaque chaîne (Phe47, Phe50 et Phe79), et par de **nouvelles interactions hydrophobes au niveau des domaines HTH** d'un homodimère avec ceux de l'autre homodimère pour reconstituer le corps de l'hétérodimère.

Le mécanisme de formation se ferait ainsi en trois étapes : le passage de chaque homodimère de leur conformation native à leur conformation intermédiaire, puis l'association des intermédiaires en un hétérotétramère transitoire, et finalement la dissociation de ce tétramère en deux hétérodimères.

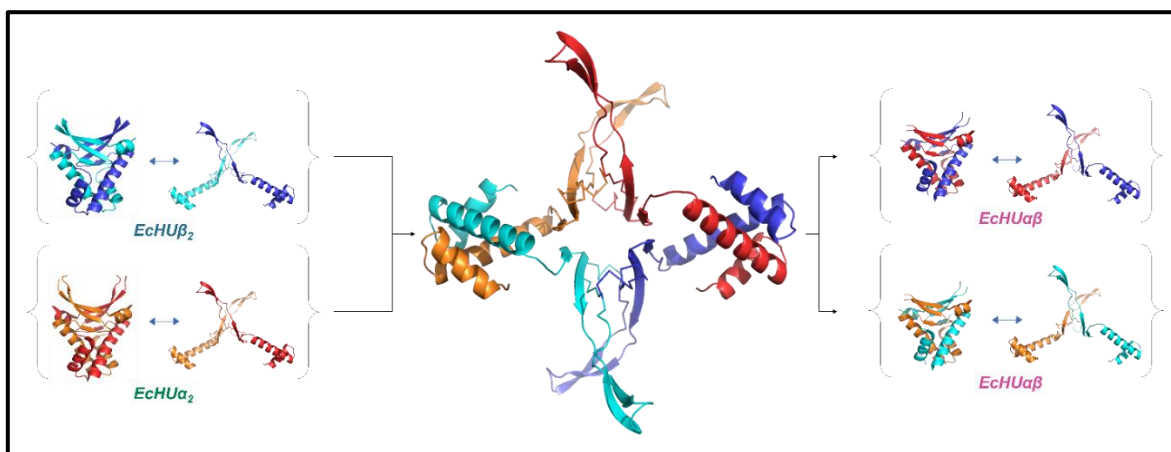


Figure 14 : Modèle du tétramère transitoire et schéma du mécanisme de formation de l'hétérodimère à partir des structures cristallographiques et des modèles « gros-grain » (adapté de thèse RLM, 2015) ^[69].

1.3.3. Mise en évidence de formes tétramériques

Dans le cadre d'une étude préliminaire, Rémy Le Meur, au cours de ses travaux de thèse et en collaboration avec le groupe de spectrométrie de masse (MS) de l'équipe, a mis en place des conditions minimales pour suivre la **cinétique de formation de l'hétérodimère par spectrométrie**

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

de masse en condition native. Ce type d'expérience, détectant des ions de différents rapports masse-sur-charge, peut permettre de caractériser le changement d'état oligomérique du système. Lors de ces suivis cinétiques, **l'apparition de l'hétérodimère** a pu être quantifiée en se basant sur l'ion +9 de l'hétérodimère. Il est aussi possible d'observer l'apparition de **plusieurs formes tétramériques** pendant le suivi (<1 %) au niveau des ions +13 et +15 (Figure 15).

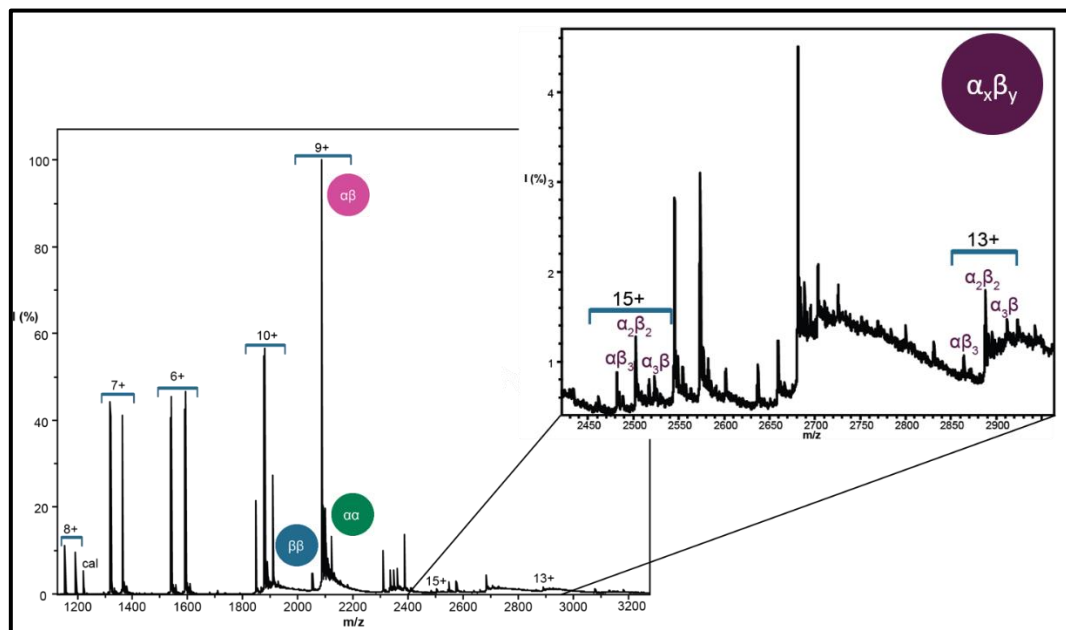


Figure 15 : Spectre de masse en conditions natives de suivi de la formation de l'hétérodimère (adapté de Thèse RLM, 2015) ^[69].

Lors d'une seconde étude par spectrométrie de masse en conditions natives, la possibilité d'un échange de chaîne au sein d'un même échantillon d'homodimère a été évaluée. Sur le spectre de masse de chacun des homodimères, on peut observer des signaux correspondant à une forme tétramérique ^[69]. Afin de mieux caractériser ce signal, des suivis cinétiques ont été réalisés à partir d'échantillons d'homodimère marqué ¹⁵N et d'échantillons du même homodimère non marqué. L'utilisation d'un dimère marqué ¹⁵N et d'un dimère non marqué permet de suivre un **changement de masse** dans l'échantillon si l'échange se produit. L'expérience a permis d'observer l'apparition d'un homodimère hémimarké et de le quantifier. Des signaux correspondant à des formes tétramériques ont aussi été observés et quantifiés. En fin de cinétique, les échantillons étaient composés **de 40 % à 50 % d'homodimères hémimarkés**. Ces expériences ont mis en évidence que **les homodimères peuvent échanger leurs chaînes** et que ce système d'échange entre les homodimères est **combinatoire**, contrairement au système global qui est lui séquentiel. Le rôle physiologique que pourrait jouer un échange de chaînes entre deux molécules d'un même homodimère n'est pas trivial. Il pourrait soit simplement découler de l'existence du mécanisme d'échange de chaînes nécessaire à la formation de l'hétérodimère, soit être un mécanisme intrinsèque du système HU de type auto-chaperonnage (=self-chaperoning).

1.4. Conclusion : *EchU*, un système d'étude de choix, mais un système d'étude complexe

HU est la protéine majoritaire associée au nucléoïde des bactéries. Elle est présente chez toutes les bactéries. C'est un dimère qui peut échanger ses chaînes et qui présente une grande homologie de structure et de séquence entre les espèces. Chez *E. coli* en particulier, elle est abondante, et existe sous trois formes dimériques. De plus, la souche *E. coli* est un **modèle** de bactérie très étudié en laboratoire. Caractériser l'échange de chaînes chez cette espèce permet **comprendre les bases du système d'échange de HU d'un point de vue structural et cinétique**. Cette caractérisation totale permettra de discriminer les éléments clefs du mécanisme. Une fois caractérisés, ces éléments pourront être bloqués afin d'empêcher la protéine de remplir ses fonctions biologiques. Cette compréhension du mécanisme est nécessaire à des fins **thérapeutiques** et/ou **agronomiques**. Elle permettrait par exemple le développement de nouveaux antibiotiques (en travaillant sur les protéines HU d'espèces comme *Pseudomonas aeruginosa* ou *Mycobacterium tuberculosis*), ou de conservateurs (en travaillant sur HU β_2 d'*E. coli*).

Un mécanisme complexe à décrire

A partir de l'ensemble des travaux de recherches réalisés sur la protéine HU d'*E. coli*, on constate que l'échange de chaîne de cette protéine est un mécanisme complexe aux multiples états :

- Chacun des dimères est en équilibre entre deux conformations (N_2/I_2) ;
- Les homodimères sont, à eux seuls, des systèmes combinatoires (α_4 et β_4) ;
- La formation de l'hétérodimère se fait à partir des homodimères ($\alpha\alpha + \beta\beta \rightarrow 2 \alpha\beta$) ;
- Cette formation passe par un état tétramérique transitoire ($\alpha_2\beta_2$) ;
- Il a été observé plusieurs formes hétérotétramériques transitoires ($\alpha\beta_3$, $\alpha_2\beta_2$ et $\alpha\beta_3$) ;
- Et le tétramère est issu de l'association des conformations intermédiaires ($I_2 + I_2 \rightarrow T_4$).

Afin de décrire le système, il a fallu dans un premier temps, trouver une **représentation schématique simple de ce mécanisme d'échange complexe** en tenant compte de toutes les étapes d'échange connues et/ou observées. De cette réflexion, j'ai pu construire le schéma présenté sur la Figure 16. Il est la représentation, sous la forme d'un grand triangle, du mécanisme global à un temps t . Chaque côté de ce triangle correspond à un dimère en équilibre entre sa conformation native et intermédiaire. Chacun des dimères peut former des tétramères en fonction du dimère avec lequel il s'associe. **Ce schéma montre plusieurs points de blocage de l'échange qui sont les éléments centraux du triangle à savoir les formes tétramériques**. En effet, si on parvient à bloquer le tétramère, on bloquera totalement l'échange de chaîne et il n'y aura plus de dimère de HU pouvant assurer les fonctions biologiques de la protéine.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

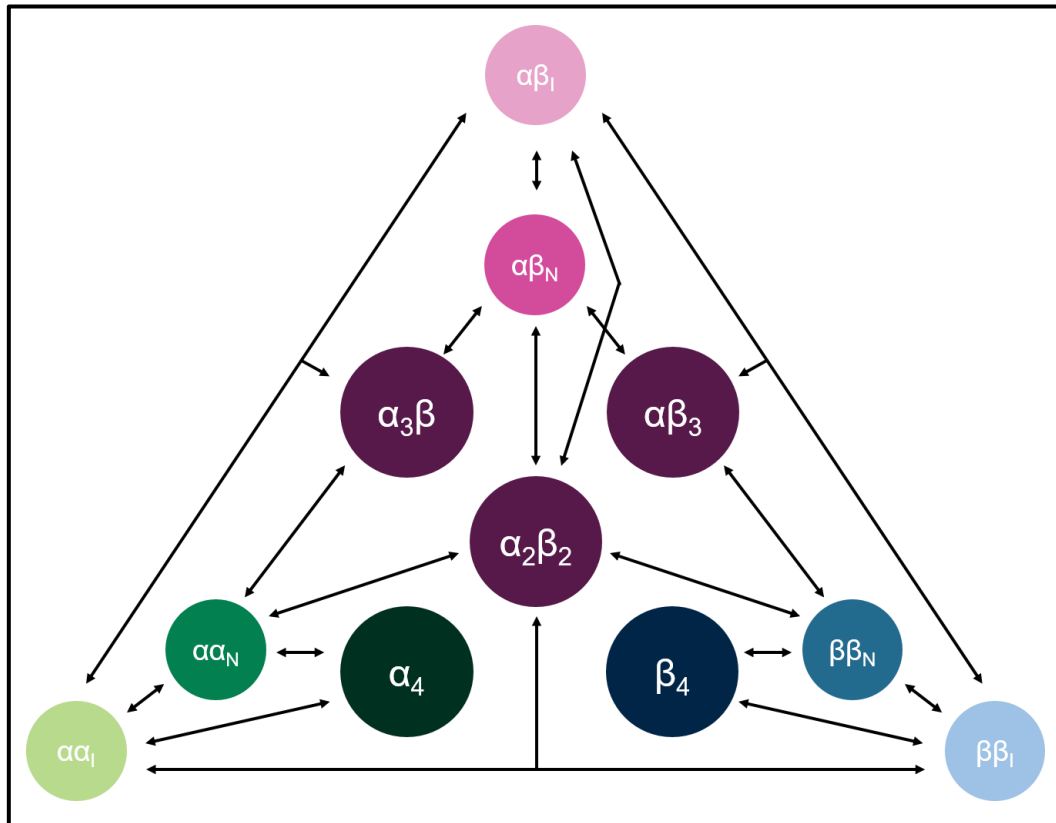


Figure 16 : Schéma du mécanisme complet d'échange de chaînes de la protéine HU d'*E. coli*.
 α = chaîne EchU α ; β = chaîne EchU β ; N = conformation native ; I = conformation intermédiaire.

Chapitre 2. Matériel et méthodes pour la description des *EchU*

Dans ce chapitre, nous présentons les différentes méthodes ainsi que les paramètres utilisés pour l'étude du système de la protéine HU d'*E. coli* par RMN.

L'ensemble des expériences par RMN, pour l'étude du mécanisme d'échange de chaînes de *EchU*, a été enregistré sur un spectromètre Bruker 700 MHz Avance III HD équipé d'une crysonde triple résonance inverse TCI 5 mm, avec bobine de gradient Z. Les différents spectres ont été traités avec le logiciel TopSpin™ puis analysés en utilisant le logiciel CcpNmr^[71]. Les calculs de structure ont ensuite été réalisés sous contrainte RMN par le logiciel ARIA^[72].

2.1. Du plasmide à l'échantillon de protéine pure

Les expériences de RMN sur des protéines commencent par l'obtention d'une quantité suffisante de cette protéine pure en solution afin de pouvoir par la suite l'étudier.

Le protocole de production et de purification des protéines HU d'*E. coli* a déjà été présenté et illustré par Rémy Le Meur dans son manuscrit de thèse^[69] et vous est résumé ci-après (Figure 17). Une description plus détaillée de la stratégie autour des différentes étapes de production et de purification de protéines est réalisée dans la Partie 2.2.1.1 du manuscrit qui concerne la mise en place du protocole de production et de purification de la protéine MC1.

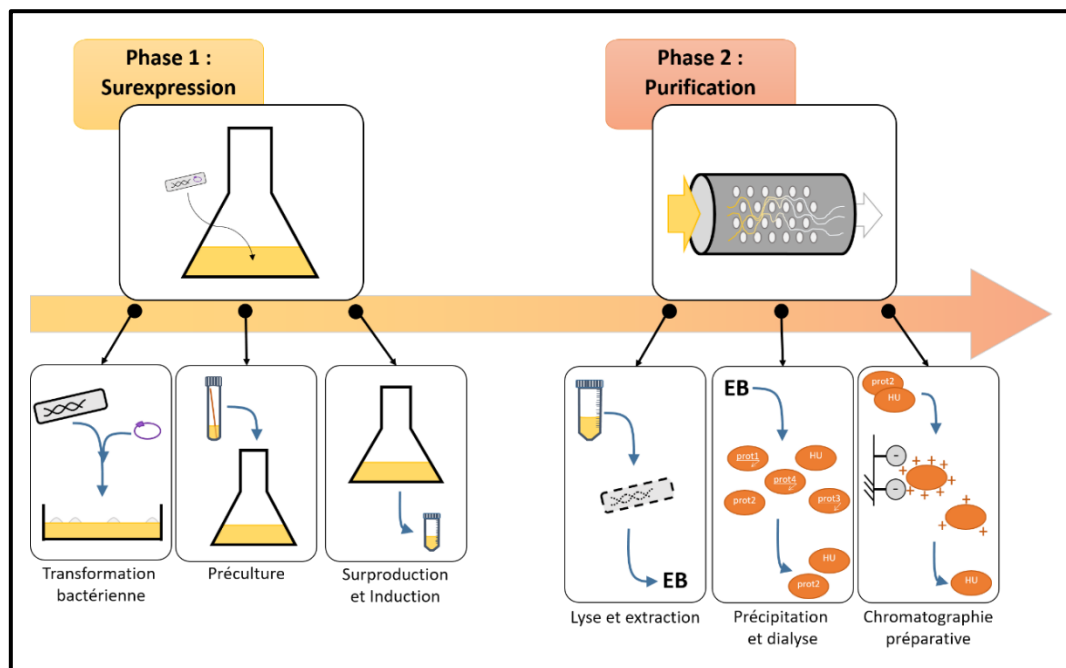


Figure 17 : Schéma de préparation des échantillons de protéines HU.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

2.1.1. Surexpression de *EchU* non marquée ou marquée (^{15}N , $^{15}\text{N}^{13}\text{C}$)

Les protocoles de surexpression de protéines marquées et non marquées sont différents. Lorsque la protéine est marquée (^{15}N ou $^{15}\text{N}^{13}\text{C}$), la surproduction doit se faire en milieu minimum afin de s'assurer que la protéine d'intérêt synthétisée sera marquée à 100 % avec l'isotope souhaité. Pour la surexpression de protéines non marquées (NM) en revanche, la surproduction se fait dans un milieu de culture standard. L'ensemble des solutions et la composition des différents tampons utilisés sont présentés dans le Tableau 3.

Volume				
nm	¹⁵ N, ¹⁵ N ¹³ C	Solution		
3mL	3mL	Ampicilline 100mg.mL ⁻¹ (Ampi)		
1,5mL	1mL	IPTG 1M		
3L	3L	LB 20g.L ⁻¹		
-	600mL	Milieu M9 base sans NH ₄ Cl		
-	500mL	Milieu M9 complet ¹⁵ N ¹³ C (ou ¹⁵ N)		
100mL	100mL	Tampon de lavage des cellules		
Milieu M9 base sans NH ₄ Cl (1L, pH7.2)		Milieu M9 base ¹⁵ N (1L, pH7.2)	*Milieu M9 complet ¹⁵ N ¹³ C (500 mL)	Tampon de lavage des cellules
8,83g Na ₂ HPO ₄ , 2H ₂ O 3,10g KH ₂ PO ₄ 517mg NaCl		8,83g Na ₂ HPO ₄ , 2H ₂ O 3,10g KH ₂ PO ₄ 517mg NaCl 1,03g ¹⁵ NH ₄ Cl	483,5mL Milieu M9 base ¹⁵ N 10mL Glucose ¹³ C 20% 500μL MgSO ₄ 1M 150μL CaCl ₂ 1M 500μL Thiamine 1mg.mL ⁻¹ (Vitamine B1) 5,35mL H ₂ O	10mM Tris/HCl pH8 100mM EDTA pH8 100mM NaCl

Tableau 3 : **Composition des solutions et des tampons pour la surproduction de protéines HU dans *E. coli*.**

Transformation bactérienne

50 μL de bactéries BL21-DE3 (calcium compétentes) sont incubées 30 min dans un bain de glace avec 100 ng de plasmide (pJES *hupA/B*). Un choc thermique est ensuite appliqué en incubant les bactéries 45 secondes à 42 °C. Les bactéries sont ensuite rapidement remises dans la glace afin de refortifier les parois. Au bout de 5 minutes, l'échantillon est enrichi avec 1 mL de milieu de culture (LB) et placé 1-2 h sous agitation (160 rpm à 37 °C). 100 μL de culture non diluée ou diluée au 10^{ème} sont ensuite étalés sur des boîtes de Pétri contenant 20 mL LB/Agar supplémentés de 20 μL d'Ampicilline (Ampi). Les boîtes sont finalement incubées à 37 °C pendant la nuit. Les colonies de HUα₂ ou de HUβ₂ obtenues sont utilisées pour ensemercer des cultures plus importantes.

Production de protéines HUα₂ ou HUβ₂ non marquées

100 mL de LB supplémentés de 100 μL d'Ampi sont ensemencés avec une colonie de bactéries BL21-DE3 ayant intégré le plasmide pJES *hupA/B* puis incubés pendant la nuit (160 rpm à 37 °C, OVN). Cette préculture est ensuite culotée (5000 rpm à 4 °C pendant 20 min) et utilisée pour

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

ensemencer les cultures de travail (soit 2,4 L LB + 2,4 mL Ampi) telles que la turbidité de départ de ces cultures de travail soit comprise entre 0,1 et 0,15 UA_{600nm}. Celles-ci sont de nouveau incubées sous agitation (160 rpm à 37 °C). Entre 0,7 et 1 UA_{600nm}, l'expression est induite par ajout de 0,5 mM final d'IPTG (soit 1,2 mL). Après 2 h d'induction sous agitation (160 rpm à 37 °C pendant 2 h), les bactéries sont culotées (6000 rpm à 4 °C pendant 20 min), avant d'être reprises dans un tampon de lavage (90 mL) et reculotées (5000 rpm à 4 °C pendant 30 min) en Falcon de 50 mL. Les culots humides obtenus au final sont conservés à -20 °C.

Les rendements obtenus avec ce protocole sont de l'ordre de **2 à 3 g de culot par litre de milieu culture de travail**.

Production de protéines HUα₂ ou HUβ₂ marquées ¹⁵N ou ¹⁵N¹³C

De la même façon que précédemment, 100 mL de LB supplémentés de 100 µL d'Ampi sont ensemencés avec une colonie de bactéries BL21-DE3 ayant intégré le plasmide pJES *hupA/B* puis incubés pendant la nuit (160 rpm à 37 °C). Cette préculture est ensuite culotée (5000 rpm à 4 °C pendant 20 min) et utilisée pour ensemencer les cultures de travail (soit 2,4 L LB + 2,4 mL Ampi) telles que la turbidité de départ de ces cultures de travail soit comprise entre 0,1 et 0,15 UA_{600nm}. Celles-ci sont de nouveau incubées sous agitation (160 rpm à 37 °C). Entre 0,7 et 1 UA_{600nm}, elles sont culotées (6000 rpm à 4 °C pendant 20 min) puis reprises avec la solution M9 base sans NH₄Cl stérile et froide (soit 600 mL) avant d'être de nouveau culotées (6000 rpm à 4 °C à 20 min). Les culots sont repris avec 100 mL du milieu M9 complet marqué (marqué ¹⁵N¹³C ou marqué ¹⁵N) et transférés dans une nouvelle culture de travail contenant les 400 mL de milieu M9 complet restants supplémentés de 500 µL d'Ampi. La seule différence entre la production de protéine marquée ¹⁵N¹³C et marquée ¹⁵N se fait à cette étape au niveau de la composition du tampon : le glucose marqué ¹³C du milieu M9 complet est remplacé par du glucose non marqué (* dans le [Tableau 3](#)). La culture est incubée une heure sous agitation (160 rpm à 37 °C). L'expression est ensuite induite par ajout de 500 µL d'IPTG. Après 3 h d'incubation (160 rpm à 37 °C), les bactéries sont culotées (6000 rpm à 4 °C pendant 20 min) puis reprises avec le tampon de lavage (soit 90 mL) avant d'être transférées dans des tubes Falcon de 50 mL. Les bactéries sont une nouvelle fois culotées (5000 rpm à 4 °C pendant 30 min) puis les culots humides obtenus sont stockés à -20 °C.

Les rendements obtenus avec ce protocole sont de l'ordre de **3 à 4 g de culot pour 2,4 L de milieu culture de travail et 500 mL de milieu minimum**.

2.1.2. Purification des protéines HUα₂ et HUβ₂

Le protocole présenté ci-dessous est identique quel que soit l'homodimère, marqué ou non, de *ECHU* à purifier. Les rendements obtenus avec ce protocole sont de **6 à 8 mg de protéines purifiées par gramme de culot bactérien**.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

Le Tableau 4 présente la composition des différents tampons utilisés dans le protocole.

TpA [4X]	Tp Lyse	Tp lavage/précipitation SA	Tp dialyse	
100mM Tris/HCl pH7.5 4mM EDTA pH8 8% Glycérol 0,4mM PMSF	[1X] TpA 400mM NaCl 10mM β-mercaptoéthanol 0,1mM PMSF	1X TpA 400mM NaCl 2,5mM βmercapto 290g.L ⁻¹ (NH ₄) ₂ SO ₄ 0,1mM PMSF	[1X] TpA 50mM NaCl 2,5mM βmercapto 0,1mM PMSF	
Tp1	Tp2	Tp3	Tp4	Tp final (pH7.6)
[1X] TpA 2,5mM βmercapto 0,1mM PMSF	[1X] TpA 1M NaCl 2,5mM βmercapto	25mM Tris/HCl pH7.5 2mM EDTA pH8 2,5mM βmercapto	25mM Tris/HCl pH7.5 1M NaCl 2mM EDTA pH8 2,5mM βmercapto	20mM NaP 200mM NaCl 0,05mM DTT 10% D ₂ O

Tableau 4 : Composition des tampons pour la purification de protéines HU de *E. coli*.

Lyse et extraction

Les culots humides sont repris avec le Tp lyse (10 mL.g⁻¹ de culot). On ajoute ensuite à la suspension du triton X-100 (~0,1 % final) et du lysozyme (~0,5 g.L⁻¹ final), puis on laisse incuber 10 min à 30 °C. On réalise alors 3 cycles de congélation (azote liquide) / décongélation (bain marie 30 °C). Le lysat est ensuite ultra centrifugé (35000 rpm à 4 °C pendant 1 h) et le surnageant (« **EB** » pour **extrait brut**) est récupéré pour la suite.

Précipitation au sulfate d'ammonium et dialyse

On ajoute progressivement à l'extrait brut du sulfate d'ammonium sous agitation à 4 °C (jusqu'à 0,29 g.mL⁻¹). Après 3 h d'agitation, l'extrait saturé en sulfate d'ammonium est récupéré avec le Tp lavage/précipitation SA et ultra centrifugé (35000 rpm à 4 °C pendant 1 h). Le surnageant est récupéré et dialysé contre le Tp1 pendant 1 h sous agitation (150-200 rpm) à 4 °C. Il est ensuite dialysé contre le Tp dialyse pendant la nuit (150 rpm à 4 °C).

Chromatographie préparative

Le dialysat est filtré à 0,45 μ m puis chargé sur une colonne de chromatographie échangeuse de cations SP-FF (préalablement équilibrée dans 95:5 Tp1/Tp2). Après un lavage en 10 volumes de colonne (CV) avec le mélange 90:10 Tp1/Tp2, la **fraction « FF »** d'intérêt est éluée en mode isocratique avec un mélange 40:60 Tp1/Tp2.

La fraction « FF » est ensuite diluée avec le Tp1 jusqu'à une concentration en NaCl finale d'environ 85 mM et chargée sur une colonne de chromatographie échangeuse d'ions SP-HP (équilibrée dans 95:5 Tp1/Tp2). Après un lavage en 10 CV avec le mélange 90:10 Tp1/Tp2, la **fraction « HP »** rassemblée et contenant la protéine, est éluée par un gradient linéaire de 10 à 60% Tp2 sur 20 CV.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

Une fois concentrée jusqu'à 3-4 mL, la fraction « HP » est chargée sur une colonne de tamisage moléculaire AcA54, pour « Acrylamide 5 %, Agarose 4 % », (équilibrée dans Tp2) puis éluée avec Tp2 sur 1,5 CV. Les fractions collectées contenant la protéine sont rassemblées et constituent la **fraction « AcA »**.

La fraction « AcA » est alors concentrée et filtrée à 0,2 µm. Elle est ensuite diluée quinze fois avec le Tp3 et chargée sur une colonne de chromatographie faiblement échangeuse de cations 20CM (équilibrée dans 90:10 Tp3/Tp4). Après un lavage en 20 CV avec le mélange 90:10 Tp3/Tp4, la protéine est éluée par un gradient linéaire de 10 % à 50 % de Tp4 sur 120 CV. Les fractions collectées contenant la protéine sont rassemblées et constituent la **fraction « 20CM »**.

Cette fraction est concentrée et son tampon est échangé pour le Tp final (20 mM NaP, 200 mM NaCl, 0,05 mM DTT, 10 % D₂O). L'échantillon de protéine pure est conservé à -80 °C.

2.1.3. Préparation des échantillons pour les différentes études par RMN

L'ensemble des échantillons de dimère de *EchU* ont été préparés dans un tampon (20 mM NaP, 200 mM NaCl, 0,05 mM DTT, 10 % D₂O), et analysés dans un tube 3 mm standard.

Les échantillons d'hétérodimère ont été systématiquement préparés à partir d'échantillons de chaque homodimère avec différents ratios (x:y *EchU*α₂/*EchU*β₂) où x et y correspondent aux proportions respectives de *EchU*α₂ et de *EchU*β₂, marqués ou non.

Pour la détermination de la structure spatiale des *EchU*

Un échantillon de chaque homodimère a été préparé à 0,5 mM pour la détermination des structures natives à basse température.

Trois échantillons d'hétérodimère ont été préparés à 1 mM pour la détermination de la structure du dimère natif et intermédiaire :

- *EchU*αβ non marqué (1:1 *EchU*α₂/*EchU*β₂) ;
- *EchU*αβ marqué ¹⁵N¹³C sur la chaîne β (3:2 *EchU*α₂/*EchU*β₂ marqué ¹⁵N¹³C) ;
- *EchU*αβ marqué ¹⁵N¹³C sur la chaîne α (2:3 *EchU*α₂ marqué ¹⁵N¹³C/*EchU*β₂).

Pour le suivi en température de l'évolution des populations N₂/I₂

Les échantillons utilisés pour les études de relaxation, ainsi qu'un échantillon à 0,5 mM de *EchU*αβ doublement marqué ¹⁵N¹³C/¹⁵N (1:1 *EchU*α₂ marqué ¹⁵N¹³C/*EchU*β₂ marqué ¹⁵N), ont servi à l'enregistrement des spectres ¹⁵N-HSQC en température des trois dimères.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

Pour les études de relaxation

Des échantillons des différents dimères de *EchU* ont été préparés à 0,5 mM pour les études de relaxation par RMN :

- *EchU* α_2 marqué ^{15}N ;
- *EchU* β_2 marqué ^{15}N ;
- *EchU* $\alpha\beta$ marqué ^{15}N sur la chaîne β (3:2 *EchU* α_2 /*EchU* β_2 marqué ^{15}N) ;
- *EchU* $\alpha\beta$ marqué ^{15}N sur la chaîne α (2:3 *EchU* α_2 marqué ^{15}N /*EchU* β_2).

Pour l'étude cinétique du système

Différents échantillons d'homodimères marqués et non marqués ont été utilisés pour l'étude de la cinétique globale du système d'échange de chaînes :

- *EchU* α_2 marqué ^{15}N à 0,62 mM et *EchU* β_2 marqué $^{15}\text{N}^{13}\text{C}$ à 0,45 mM (cinétique 1 et titrage 1) ;
- *EchU* α_2 non marqué à 2,66 mM et *EchU* β_2 marqué $^{15}\text{N}^{13}\text{C}$ à 3,29 mM (cinétiques 2 et 3) ;
- *EchU* α_2 marqué $^{15}\text{N}^{13}\text{C}$ à 1,74 mM et *EchU* β_2 non marqué à 8,01 mM (cinétiques 4 et 5) ;
- *EchU* α_2 non marqué à 0,57 mM et *EchU* β_2 marqué ^{15}N à 0,11 mM (titrage 2 et 3).

2.2. Détermination de la structure spatiale d'une protéine

La détermination des structures des protéines par RMN se déroule en **3 étapes principales** (Figure 18).

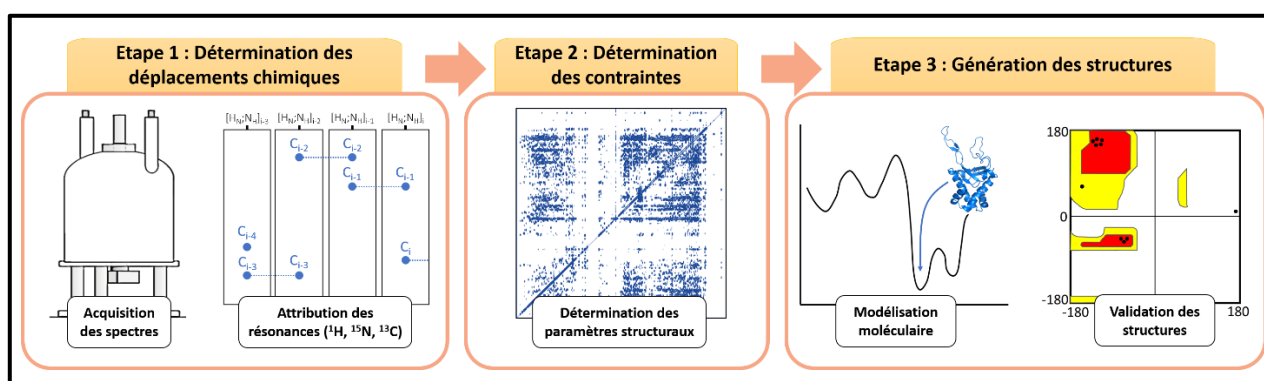


Figure 18 : Etapes de la détermination de structure spatiale de protéine par RMN.

La **première étape** consiste à déterminer les **déplacements chimiques des noyaux ^1H , ^{15}N , ^{13}C** lorsque la protéine est doublement marquée isotopiquement. La RMN ne permet cependant pas de déterminer la structure spatiale d'une protéine de façon directe à partir de ces déplacements chimiques bien que des programmes de prédiction existent comme CS Rosetta [73].

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

La **seconde étape** est de déterminer des **contraintes RMN** qui peuvent être utilisées dans un calcul de modélisation moléculaire. Ces contraintes peuvent être de 3 types :

- des **contraintes de distances**, dérivées des intensités (ou des volumes) des pics hors diagonale observés sur les spectres de type NOESY (pour « Nuclear Overhauser Spectroscopy »). Ces intensités, I , correspondant à des corrélations dipolaires, sont reliées à la distance qui sépare deux hydrogènes proches dans l'espace, par la relation $I = \frac{k}{d^6}$, où k est une constante qui dépend du temps de corrélation de la protéine [74]. Ces mesures fournissent des informations très locales sur l'environnement des hydrogènes ($\leq 6 \text{ \AA}$) et permettent d'obtenir de nombreuses contraintes de distance (distances intra chaînes et/ou inter chaînes) et permet de reconstruire la structure de la molécule. De plus, on observe des enchaînements de corrélations spécifiques : dans le cas d'hélices α , $d_{\alpha N}(i, i+3)$ et $d_{\alpha N}(i, i+4)$ et dans le cas de feuillets β des corrélations intenses $d_{\alpha N}(i, j)$ et $d_{\alpha\alpha}(i, j)$. Elles correspondent à des distances courtes ($\sim 2 \text{ \AA}$) intra ou inter chaînes qui permettent de définir les liaisons hydrogène existantes dans la structure de la protéine d'intérêt. Dans ce sens, il est important de recueillir un maximum de données pour affiner la structure de la molécule.
- des **contraintes d'angle dièdre** (=angles de torsion) obtenues soient par une mesure expérimentale des constantes de couplage, soit par prédiction à partir des déplacements chimiques. Dans le premier cas, les constantes de **couplage scalaire** $^3J_{\text{HNH}\alpha}$ (couplage entre deux noyaux H_N et H_α) peuvent être déterminées sur des spectres de type COSY DQF [74]. Elles permettent, en utilisant la relation de Karplus, d'obtenir une valeur expérimentale de l'angle dièdre ϕ [74]. Cette méthode est encore utilisée sur des petits peptides ou des miniprotéines non marquées. Dans le deuxième cas, plusieurs programmes peuvent être utilisés pour prédire ces valeurs d'angles comme DANGLE [70], PREDITOR [75], ou TALOS [76,77]. Tous utilisent une base de données contenant des structures de protéines connues ainsi que les déplacements chimiques expérimentaux associés à laquelle ils confrontent la séquence et les déplacements chimiques expérimentaux de la protéine étudiée.
- des **contraintes d'orientation des liaisons** dérivées de la mesure des couplages dipolaires résiduels (RDCs) [78,79]. Ils sont principalement utilisés pour affiner une structure. Les RDCs n'ont pas été utilisés dans cette étude.

La **troisième étape** consiste à réaliser le calcul de la structure grâce à un logiciel de modélisation moléculaire sous contraintes RMN (CNS [79,80], AMBER [81], CHARMM [82,83], etc.). Ces logiciels intègrent la séquence primaire de la protéine, les contraintes de distances déterminées sur les spectres NOESY, les contraintes de distances liées à la présence de liaisons hydrogène ou de ponts disulfure (le cas échéant), les contraintes d'angles dièdres (expérimentaux ou prédits) et les

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

contraintes issues de la mesure des RDCs (le cas échéant). Ces logiciels bien que très performants, nécessitent, avant de pouvoir les utiliser, d'avoir attribué tous les pics des spectres NOESY manuellement et d'en avoir extrait l'information sur les distances. Il existe maintenant des logiciels qui permettent de combiner l'attribution automatique des spectres NOESY et le calcul de structure comme ARIA [72].

2.2.1. Première étape : Attribution

Dans le cas de grosses protéines, l'attribution de l'ensemble des noyaux observables (^1H , ^{15}N , ^{13}C), par marquage isotopique de la protéine et par l'acquisition de spectres 3D/4D hétéronucléaires, est nécessaire pour l'attribution séquentielle et la distinction des corrélations spatiales de la macromolécule.

Attribution du squelette peptidique

L'attribution du squelette peptidique (H_N , N_H , C' , C_α , H_α) étendu aux C_β , H_β de *EchU $\alpha\beta$* est réalisée grâce à l'utilisation d'une série d'expériences hétéronucléaires complémentaires de routine : ^{15}N -HSQC [84], HNCO [85], HN(CA)CO [85], HNCA [85], HN(CO)CA [86], CBCANH [87] et CBCA(CO)NH [87]. Dans notre cas, elles ont été enregistrées à 20 °C et à 37 °C sur les deux échantillons d'hétérodimère hémimarqués $^{15}\text{N}^{13}\text{C}$.

L'expérience ^{15}N -HSQC (pour « Heteronuclear Single Quantum Coherence ») nous donne les déplacements chimiques des corrélations $^1\text{H}_\text{N}$ - $^{15}\text{N}_\text{H}$ de la protéine. Les autres expériences correspondent à des expériences hétéronucléaires dont le nom correspond aux noyaux observés et au chemin de cohérence suivi (Figure 19). Ces expériences vont, en général, par paires (HNCO/HN(CA)CO, HNCA/HN(CO)CA, etc.) afin de faciliter l'identification des signaux inter résiduels des signaux intra résiduels et donc permettre l'attribution séquentielle de la protéine [74].

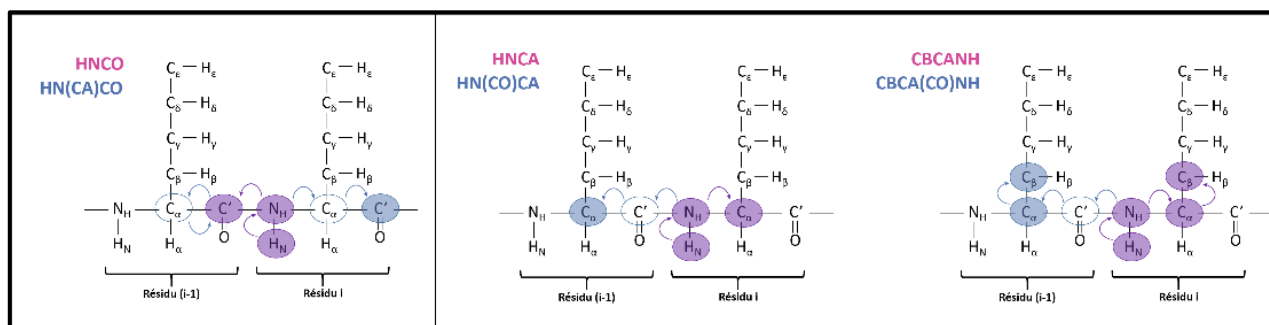


Figure 19 : Illustration de la stratégie d'attribution du squelette peptidique étendu aux corrélations β . Schéma des chemins de cohérence (flèches) des différentes expériences tridimensionnelles hétéronucléaires, des noyaux observés (cercle plein) et relais (cercle vide). Pour une paire d'expériences, les éléments relatifs à la première sont en bleu et les éléments communs aux deux expériences sont en violet.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

Par exemple, l'expérience HNCA permet l'observation des noyaux H_N , N_H et C_α , chacun correspondant à une des trois dimensions du spectre. Le chemin de cohérence est réalisé grâce à un transfert d'aimantation qui se fait à travers les couplages scalaires. Dans le cas de l'expérience HNCA, les valeurs des constantes de couplage scalaire $N_H(i)-C_\alpha(i)$ et $N_H(i)-C_\alpha(i-1)$ sont très proches (respectivement 11 Hz et 7 Hz). De ce fait, le transfert est réalisé de façon intra résiduel, mais aussi inter résiduel. Il est donc possible d'observer la corrélation du couple $H_N(i)-N_H(i)$ avec le carbone $C_\alpha(i)$ intra résiduel et avec le carbone $C_\alpha(i-1)$ inter résiduel. L'expérience HN(CO)CA est ensuite utilisée. Cette dernière permet d'observer uniquement les corrélations inter résiduelles. En effet, sur cette expérience le chemin de cohérence $H_N(i)-N_H(i)-C'(i-1)-C_\alpha(i-1)$ est imposé du fait de la grande différence entre les valeurs de constantes de couplage $C'(i)-C_\alpha(i)$ et $C'(i)-C_\alpha(i-1)$ (respectivement 55 Hz et inférieure à 10 Hz) et seule la corrélation avec le carbone $C_\alpha(i-1)$ est observée.

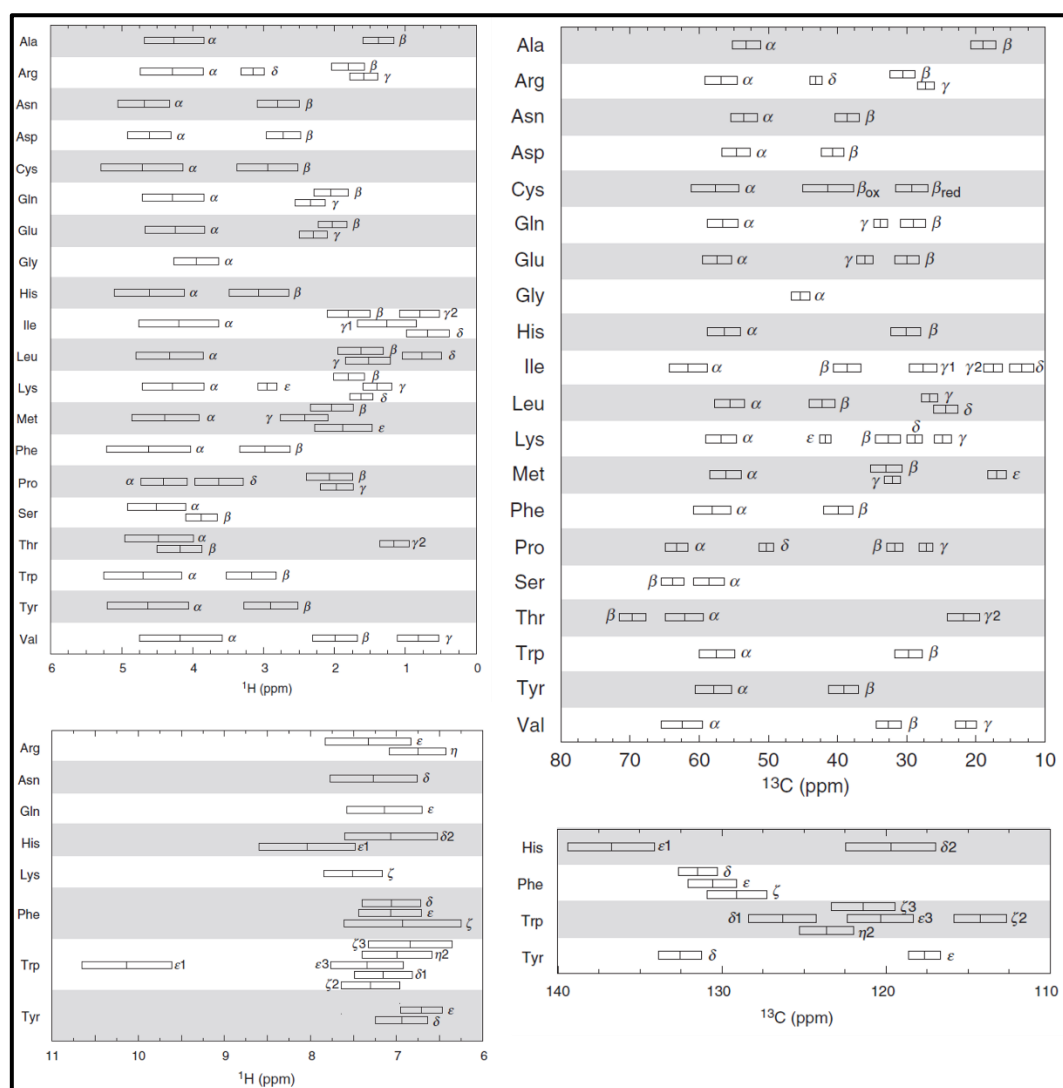


Figure 20 : Distribution moyenne des déplacements chimiques 1H et ^{13}C des acides aminés (adapté de Cavanagh et al., 2007) ^[74]. Déplacements chimiques des chaînes aliphatiques 1H (gauche) et ^{13}C (milieu), et non aliphatiques (droite) 1H (haut) et ^{13}C (bas).

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

Il est à noter que l'identification et l'attribution de la majorité des corrélations scalaires est facilitée par l'utilisation de tables de déplacements chimiques ([Figure 20](#)). En théorie, une seule paire d'expériences est nécessaire pour identifier tous les résidus et leur enchaînement. Expérimentalement, nous sommes limités par la sensibilité des expériences ainsi que par leur résolution. De plus, certains noyaux, en particulier les noyaux C' et C α , ont une faible gamme de valeurs de déplacement chimique, ce qui empêche de réaliser l'attribution séquentielle uniquement avec les couples d'expériences HNCO/HN(CA)CO et HNCA/HN(CO)CA. Il faut donc les combiner avec les expériences CBCANH/CBCA(CO)NH. La valeur de déplacement chimique des C β lève la majorité des ambiguïtés.

Attribution des chaînes latérales

L'attribution des chaînes latérales de *EchU* $\alpha\beta$ est ensuite réalisée à partir de quatre expériences hétéronucléaires de routine : CC(CO)NH ^[88], H(CC)(CO)NH ^[88], HBHANH, HBHA(CO)NH ^[87] et HCCH-TOCSY ^[89]. **Pour cette étude, elles ont été réalisées à 20 °C et à 37 °C sur les deux échantillons d'hétérodimère hémimarqués ¹⁵N¹³C.**

On peut commencer à compléter l'attribution grâce à l'utilisation des expériences HBHANH/HBHA(CO)NH qui apportent des informations sur les hydrogènes (H α , H β), données essentielles à l'attribution des chaînes latérales. L'expérience CC(CO)NH permet, quant à elle, d'attribuer les carbones de la chaîne latérale du résidu précédent au-delà du C β (H_N(i)-N_H(i)- -C_{lat}(i-1)). L'expérience H(CC)(CO)NH permet de commencer à compléter l'attribution des hydrogènes de la chaîne latérale du résidu précédent au-delà du H β (H_N(i)-N_H(i)- -H_{lat}(i-1)). Enfin, l'expérience HCCH-TOCSY permet de compléter l'attribution en reliant chaque carbone d'un résidu i à l'ensemble des hydrogènes H_{lat}(i) de la chaîne latérale du résidu.

Etude en température et perturbations de déplacements chimiques

Afin de faciliter l'attribution des signaux ¹H_N-¹⁵N_H des deux conformations de *EchU* $\alpha\beta$ à haute température, nous avons, en parallèle de l'attribution des signaux amides, réalisé une **étude en température des trois dimères**.

Les expériences ¹⁵N-HSQC enregistrées en fonction de la température ont été réalisées :

- pour l'homodimère *EchU* β_2 , entre 5 °C et 33 °C (5, 10, 15, 18, 19, 20, 21, 24, 27, 30 et 33 °C) ;
- pour l'homodimère *EchU* α_2 , entre 5 °C et 52 °C (5, 10, 15, 20, 24, 27, 30, 33, 35, 38, 41, 43, 45, 47, 49 et 52 °C) ;
- pour l'hétérodimère *EchU* $\alpha\beta$ marqué sur la chaîne α , entre 5 °C et 37 °C (5, 10, 13, 15, 18, 20, 22, 25, 27 et 37 °C) ;

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

- pour l'hétérodimère *EchU* $\alpha\beta$ marqué sur la chaîne β (5, 8, 10, 13, 15, 18, 20, 23, 25, 27, 30, 33 et 37 °C) ;
- pour l'hétérodimère *EchU* $\alpha\beta$ marqué sur les deux chaînes, entre 5 °C et 52 °C (5, 10, 16, 22, 27, 30, 33, 35, 37, 39, 41, 43, 45, 47, 49 et 52 °C).

Les données d'intensités normalisées ont ensuite été traitées dans Grace© et dans Origin© pour les différents calculs.

Le ratio de chaque état (N_2/I_2) de *EchU* est quantifié à partir de l'équation (1), adaptée de l'équation de Boltzmann.

$$p(I, T^\circ) = p(T_{\max}) + \frac{p(T_{\min}) - p(T_{\max})}{1 + e^{\frac{T^\circ - T_m}{A}}} \quad (1)$$

où $p(I, T^\circ)$ correspond à la population (N_2 ou I_2) à une température T° , $p(T_{\max})$ correspond à la population à la température la plus élevée, $p(T_{\min})$ correspond à la population à la température la plus faible et T_m est la température de transition du dimère, correspondant à la température pour laquelle $p(T_m) = \frac{p(T_{\max}) + p(T_{\min})}{2}$.

Les perturbations de déplacements chimiques (CSPs) entre les conformations N_2/I_2 sont déterminées à partir de l'équation (2).

$$CSP = \sqrt{\left(\omega_{^1H, N_2} - \omega_{^1H, I_2}\right)^2 + \left(\frac{\omega_{^{15}N, N_2} - \omega_{^{15}N, I_2}}{6}\right)^2} \quad (2)$$

Où $\omega_{^1H}$ et $\omega_{^{15}N}$ correspondent aux déplacements chimiques 1H_N et $^{15}N_H$. Les indices, N_2 et I_2 , correspondent aux états natif et intermédiaire, respectivement. Le facteur 6 permet de corriger la différence de largeur spectrale entre les atomes 1H et ^{15}N (de l'ordre de 5 ppm et 30 ppm).

2.2.2. Deuxième étape : Détermination des contraintes RMN

Après l'attribution de l'ensemble des noyaux observables, des paramètres expérimentaux sont convertis en paramètres structuraux et utilisés pour le calcul de structure sous forme de contraintes.

Comme introduit plus haut, des contraintes de distances sont nécessaires pour réaliser le calcul de structure. Elles ont été obtenues sur les expériences hétéronucléaires ^{15}N -HSQC NOESY^[84], ^{13}C -HSQC NOESY^[90] et l'expérience homonucléaire 1H NOESY^[91] que nous avons toutes enregistrées avec un **temps de mélange de 100 ms, à 20 °C et à 37 °C**, sur l'hétérodimère de HU (protéine hémimarquée $^{15}N^{13}C$ dans le cas des expériences hétéronucléaires). Elles avaient été précédemment enregistrées sur les deux homodimères avec le même temps de mélange à 20 °C pour *EchU* α_2 et à 15 °C pour *EchU* β_2 .

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EcHU*

L'expérience ^1H NOESY est la plus sensible et fournit l'ensemble des corrélations spatiales de la protéine qu'elles soient intra ou inter chaînes. L'expérience ^{15}N -HSQC NOESY fournit l'ensemble des corrélations spatiales des hydrogènes amino ($\text{H(i)}-\text{N(i)}$) et l'expérience ^{13}C -HSQC NOESY fournit l'ensemble des corrélations spatiales des hydrogènes aliphatiques et aromatiques ($\text{H(i)}-\text{C(i)}$).

Théoriquement, le logiciel ARIA ^[72] permet l'attribution automatique des pics des différents spectres NOESY. Dans notre cas, il a fallu réaliser des attributions manuelles afin d'aider ARIA à reconnaître les signaux provenant des deux chaînes de l'hétérodimère (dont les déplacements chimiques étaient très proches), mais aussi des deux conformations de l'hétérodimères, N_2 et I_2 , lorsque celles-ci étaient présentes (37 °C).

Des contraintes de distances dues à la présence de liaisons hydrogène ont aussi été introduites au fur et à mesure des calculs.

Le programme **DANGLE** (pour « Dihedral ANGles from Global Likelihood Estimates ») ^[70], inclus dans le logiciel CcpNmr, a été utilisé pour prédire les angles Phi (φ) et Psi (ψ).

Dans le cas des dimères de *EcHU*, le dernier « run » a été réalisé sur 500 structures pour les différentes itérations et 200 structures ont été affinées dans l'eau, avec les nombres de contraintes présentées dans le [Tableau 5](#).

Protéine	Contraintes		
	distance	Liaison hydrogène	Angle dièdre
<i>EcHU</i> α_2 (N_2 – 20°C)	3344	72	340
<i>EcHU</i> β_2 (N_2 – 15°C)	3622	95	309
<i>EcHU</i> $\alpha\beta$ (N_2 – 20°C)	3699	71	296
<i>EcHU</i> $\alpha\beta$ (N_2 – 37°C)	3034	73	315
<i>EcHU</i> $\alpha\beta$ (I_2 – 37°C)	2840	33	223

Tableau 5 : Nombre et type de contraintes utilisées dans les calculs de structure des dimères de *EcHU*.

2.2.3. Troisième étape : Modélisation moléculaire et calcul de structure

Le principe de l'étape de modélisation est de générer des structures compatibles avec les données expérimentales (contraintes RMN) et réalistes d'un point de vue géométrique (zone favorable du diagramme de Ramachandran).

Modélisation moléculaire

Pour ce faire, la technique utilisée dans nos calculs de structures a été le recuit simulé. Il s'agit d'une méthode de dynamique moléculaire à haute température. Elle permet de s'affranchir de la plupart des barrières énergétiques et ainsi d'explorer l'ensemble de l'espace conformationnel pour la

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

génération de structures. Le protocole de recuit simulé comporte plusieurs étapes durant lesquelles la contribution des contraintes expérimentales augmente au fur et à mesure du calcul afin de permettre à la structure de s'adapter à ces dernières. Ces contraintes sont également pondérées de façons différentes en fonction de leur type (distance inter hydrogène, liaison hydrogène, angle dièdre). Enfin, les potentiels électrostatiques ne sont pris en compte que durant l'étape finale réalisée dans une boîte d'eau. Ce type de modélisation moléculaire sous contraintes a été, pour notre étude, réalisé par le **logiciel ARIA** [72].

ARIA (pour « Ambiguous Restraints for Iterative Assignment ») est un logiciel d'attribution semi-automatique des spectres NOESY et de calcul de structures qui fonctionne en interface avec le logiciel CNS. L'utilisation simultanée des logiciels ARIA et CNS (pour « Crystallography & NMR System ») permet de réaliser l'attribution des pics NOEs et le calcul des structures 3D de manière itérative (Figure 21). Le logiciel ARIA connaît par ailleurs la notion de contrainte de distance ambiguë. Il réduit les attributions au fil des itérations afin de ne garder que celles qui contribuent à l'intensité des signaux. L'utilisation des contraintes de distance ambiguë permet de traiter les hydrogènes équivalents (méthyles, etc.) comme des NOEs ambigus.

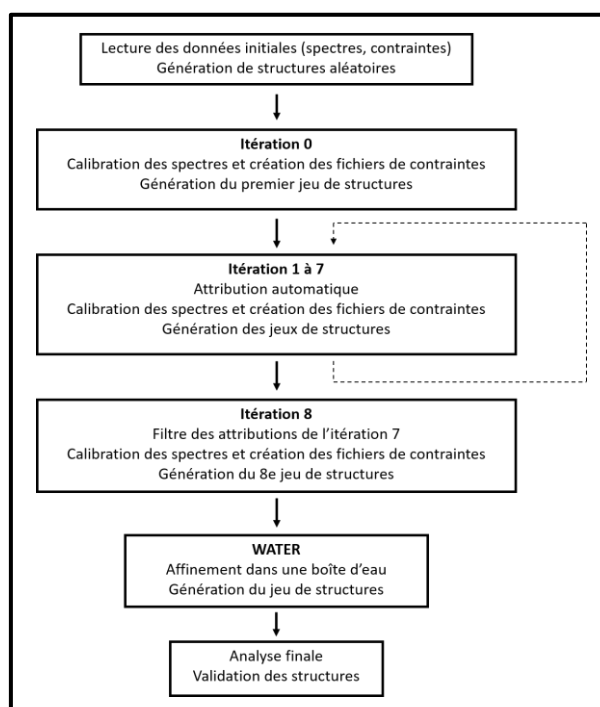


Figure 21 : Schéma simplifié du processus d'attribution et de calcul de structures itératif de ARIA.

Validation des structures

Une fois le meilleur jeu de structures généré par Aria, nous avons eu à évaluer ces structures. Seules les 20 meilleures structures sont utilisées pour les statistiques finales. Afin de choisir ces dernières, nous nous sommes basés sur différents critères :

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

- le nombre de violations des contraintes de distances ($n_{\text{Viol}>0,5\text{\AA}}$ ou $n_{\text{Viol}>0,3\text{\AA}}$ le plus proche de 0 possible) ;
- les valeurs énergétiques des structures ($E_{\text{NOE}} \leq 200 \text{ kcal.mol}^{-1}$; $E_{\text{totale}} \leq -5000 \text{ kcal.mol}^{-1}$) ;
- le nombre de violations des liaisons hydrogène ($n_{\text{Viol}>0,5\text{\AA}}$ ou $n_{\text{Viol}>0,3\text{\AA}}$ le plus proche de 0 possible) ;
- le nombre de violations des contraintes d'angle dièdre ($(n_{\text{Viol}>10^\circ}$ ou $n_{\text{Viol}>10^\circ}$ le plus proche de 0 possible).

Les structures finalement sélectionnées pour représenter la protéine d'intérêt en solution sont analysées en utilisant Procheck-NMR [92,93], CASTp3 [94], PDBePISA [95] et PROMOTIF [96].

2.3. Les principaux processus d'échange accessibles en RMN

La RMN est une technique puissante qui permet de **quantifier, avec une résolution atomique, des phénomènes dynamiques** extrêmement rapides (à l'échelle de la picoseconde) comme très lents (au-delà du millier de secondes). On peut quantifier des vitesses d'échange en se focalisant sur les déplacements chimiques et leur variation pendant une expérience RMN. Quand un échange est lent à l'échelle de la RMN, à chaque état correspond un déplacement chimique. Quand, à l'inverse, l'échange est rapide, le déplacement chimique du signal correspond à la moyenne des deux états (Figure 22) [97].

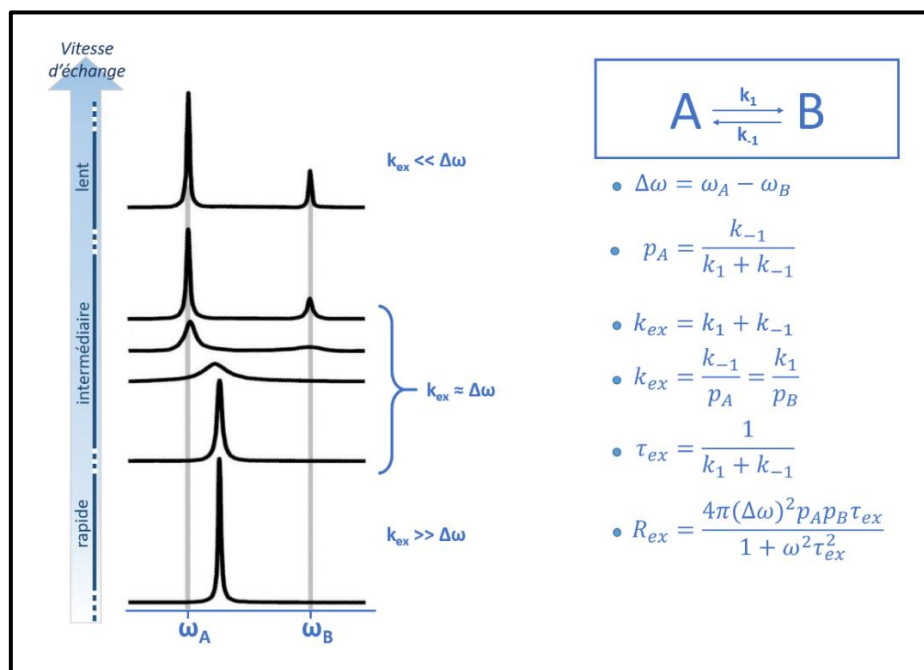


Figure 22 : L'effet de l'échange chimique entre deux états, A et B, sur des spectres RMN. (adapté de Kleckner et Foster, 2011) [98].

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

Nous avons réalisé des expériences de relaxation sur les trois dimères de HU et à différentes températures :

- pour $^{15}\text{N-EcHU}\beta_2$ à 5, 10, 15, 20 et 27 °C ;
- pour $^{15}\text{N-EcHU}\alpha_2$ à 15, 20, 27, 38 et 41 °C ;
- pour $^{15}\text{N-EcHU}\alpha^*\beta$ (* indique la chaîne marquée) à 20, 27 et 38 °C ;
- pour $^{15}\text{N-EcHU}\alpha\beta^*$ à 15, 20, 27 et 38 °C.

Ces expériences se complètent et permettent d'obtenir des informations au niveau atomique sur la dynamique (**vitesse d'échange : k_{ex} en s^{-1}**) de la protéine HU sur une large gamme de temps (Figure 23). Elles sont décrites dans les paragraphes suivants.

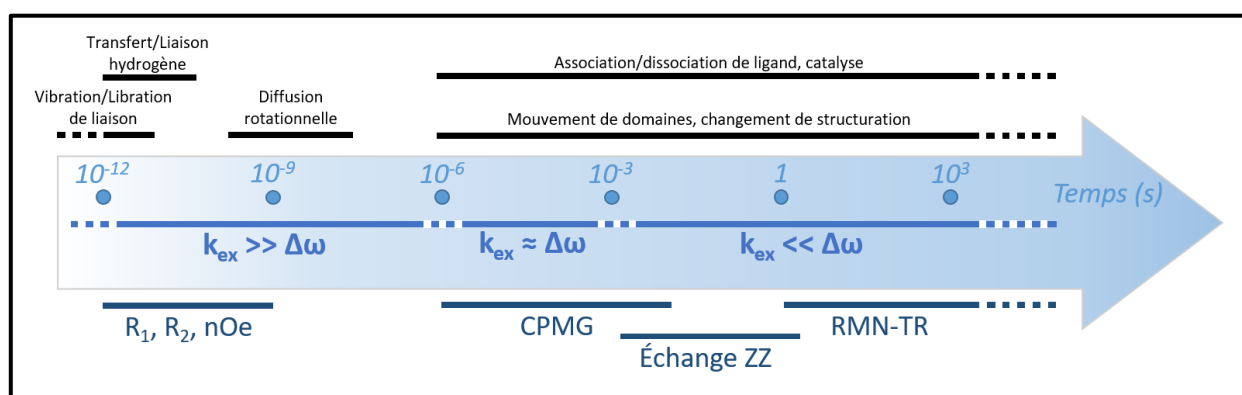


Figure 23 : Echelle de temps observable en RMN des mouvements biomoléculaires. L'échelle de temps est en secondes, les types de mouvements sont au-dessus et les expériences RMN pour les quantifier sont en-dessous.

2.3.1. Processus d'échange rapide (ps-ns)

Les processus d'échange rapide sont liés à la dynamique interne des protéines comme les phénomènes de vibration (oscillation) et/ou de libration (balancement) des liaisons. Les expériences de relaxation ^1H - ^{15}N dans cette échelle de temps nous fournissent des informations sur la dynamique interne de la protéine du squelette peptidique. Les constantes d'échange (k_{ex}) correspondant à ce type de processus vont de la gigaseconde $^{-1}$ à teraseconde $^{-1}$.

La relaxation longitudinale et transversale

La vitesse de relaxation longitudinale $\mathbf{R}_1$ est exprimée en s^{-1} et correspond à l'inverse du temps de relaxation longitudinale, T_1 . Elle correspond, comme illustré sur la Figure 24, à la vitesse du retour à l'équilibre de l'aimantation du système dans le plan du champ magnétique principal B_0 (axe parallèle à B_0). Les expériences de relaxation T_1 ont été enregistrées avec un délai de recyclage de 2 s en faisant varier le délai de relaxation : 10, 100, 200, 300, 500, 800, 1000, 1500 ms.

La vitesse de relaxation transversale $\mathbf{R}_2$ est exprimée en s^{-1} et correspond à l'inverse du temps de relaxation transversale T_2 . Elle correspond à la vitesse du retour à l'équilibre de l'aimantation du

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

système dans le plan perpendiculaire au champ magnétique principal B_0 (Figure 24). Les expériences de relaxation T_2 ont été enregistrées avec un délai de recyclage de 2 s en faisant varier le délai de relaxation : 10, 30, 50, 70, 90, 110, 130, 150, 210, 310 ms.

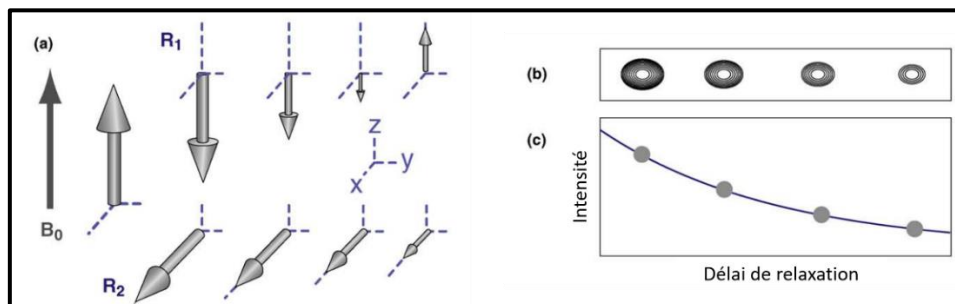


Figure 24 : **Relaxation de spin** (adapté de Mittermaier et Kay., 2009) ^[97]. a) Retour à l'équilibre d'aimantation suite à une perturbation par des impulsions radiofréquences. b,c) Diminution des intensités des pics obtenus pour des délais de plus en plus longs dans les expériences de relaxation par RMN.

L'effet Overhauser nucléaire

Lorsque deux spins sont suffisamment proches dans l'espace, la perturbation de l'un influe sur l'évolution de l'autre. Cette perturbation correspond à l'effet Overhauser nucléaire, η_{NH} . Dans le cas des protéines, la mesure de cet effet correspond à un rapport d'intensité entre l'intensité des pics provenant des expériences avec et en absence de présaturation. L'erreur sur les valeurs de η_{NH} est calculée à partir des rapports signaux sur bruit.

Les spectres ^{15}N NOE ont été enregistrés en utilisant une période de présaturation de 2 s et un délai de relaxation de 2 s également. Les spectres de référence sont enregistrés avec un délai équivalent de 4 s sans présaturation.

2.3.2. Processus d'échange intermédiaire (μs - ms)

L'expérience de **Relaxation Dispersion Carr-Purcell-Meiboom-Gill (RD-CPMG)** permet de quantifier des échanges intermédiaires à intermédiaires-rapides, même lorsque l'état minoritaire ne semble pas visible (souvent lié à un élargissement très petit des pics correspondant à des résidus présentant deux états avec des déplacements chimiques très proches). Pour ce type d'expérience, les propriétés de relaxation sont évaluées à partir d'une séquence CPMG $[\tau_{CPMG}-\pi-\tau_{CPMG}]_N$ constituée de 2 N délais τ_{CPMG} et de N impulsions π . La vitesse d'échange effective, $R_{2,\text{eff}}$, y est évaluée en fonction de la fréquence ν_{CPMG} à laquelle les impulsions π sont appliquées pendant le délai de relaxation T_{relax} (Figure 25).

$$T_{\text{relax}} = 2 \times N \times \tau_{CPMG} \quad (3)$$

$$R_{2,\text{eff}} = \frac{1}{4 \times \tau_{CPMG}} \quad (4)$$

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

$$R_{2,eff} = \frac{-\ln(\frac{I}{I_0})}{T_{relax}} \quad (5)$$

$$R_{2,eff} = R_2^0 + R_{ex} \quad (6)$$

Les équations (3) à (6), où N correspond au nombre d'impulsions π appliquées pendant le délai de relaxation T_{relax} ; I correspond à l'intensité mesurée et I_0 à l'intensité mesurée sur un spectre de référence (sans délai T_{relax}) ; et R_2^0 correspond à la vitesse de relaxation transversale, permettent de remonter à la valeur de la **vitesse d'échange**, R_{ex} [98,99].

Les expériences ont été enregistrées en utilisant la séquence d'impulsion directement disponible dans le catalogue des programmes Bruker [99] en fixant un délai T_{relax} de 80 ms pour les dimères de *EchU* et avec une série de fréquences ν_{CPMG} (0, 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 500, 600, 700, 800 et 900 Hz). Un délai de recyclage de 2 s a été utilisé. Les intensités extraites des spectres ont ensuite été traitées sur le logiciel NESSY [100] pour obtenir les valeurs de R_{ex} , de k_{ex} et les populations, p_A et p_B , de chaque état impliqué dans l'échange.

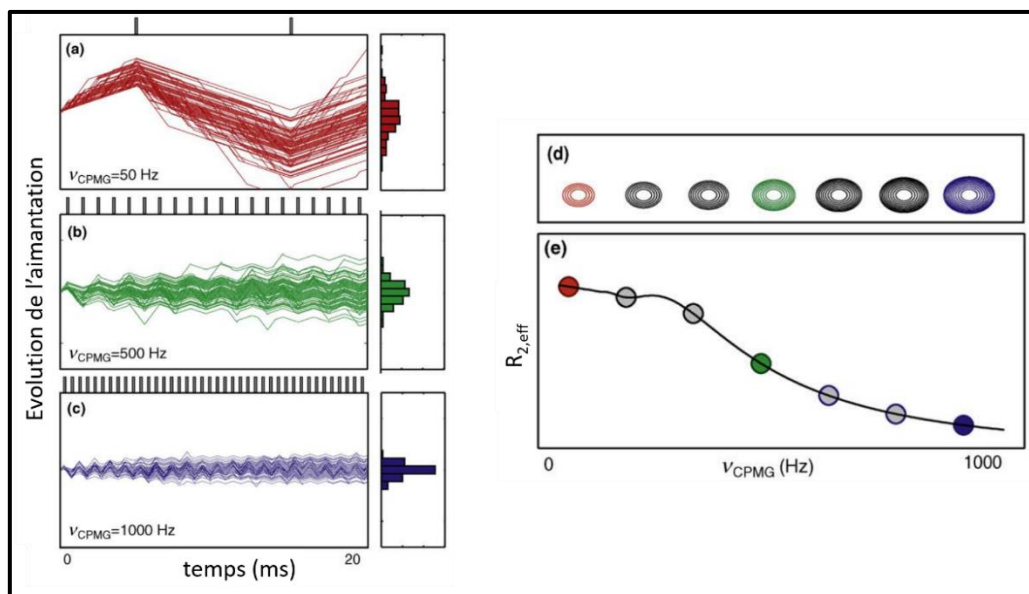


Figure 25 : Expérience de Relaxation Dispersion CPMG (adapté de Mittermaier et Kay, 2009) [97]. a-c) Simulation de trajectoires de signaux pour un ensemble de noyaux échangeant de façon stochastique entre deux environnements avec des déplacements chimiques différents pendant un délai T_{relax} de 20 ms. Les barres verticales, en haut de chaque simulation, représentent les impulsions de refocalisation appliquées pendant T_{relax} . d,e) Simulation des intensités des pics obtenus pour des fréquences ν_{CPMG} de plus en plus grandes.

2.3.3. Processus d'échange lent (>ms)

L'échange ZZ

L'expérience d'échange ZZ (ou EXSY pour « Exchange SpectroscopY ») est utilisée pour caractériser des processus d'échange lent, allant de la milliseconde à la seconde (k_{ex} entre $0,1 \text{ s}^{-1}$

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

et 10 s^{-1}). Cette expérience permet une quantification en suivant l'échange d'aimantation entre les deux états du signal (NN et II), en fonction du temps de mélange T (Figure 26). L'évolution du système dans le temps est décrite, comme publié par Martin Tollinger et ses collaborateurs ^[99], par les équations (7) à (10). Cette quantification n'est toutefois possible que dans le cas où l'état minoritaire résonne en donnant un signal suffisamment intense pour être observé. De plus, la constante d'échange (k_{ex}) doit être inférieure à la valeur des vitesses de relaxation longitudinale de chacun des états ($R_{1,N}$ et $R_{1,I}$ des états NN et II respectivement). Dans le cas contraire, les signaux diminuent en raison de la relaxation plus rapide que le transfert de population ^[101].

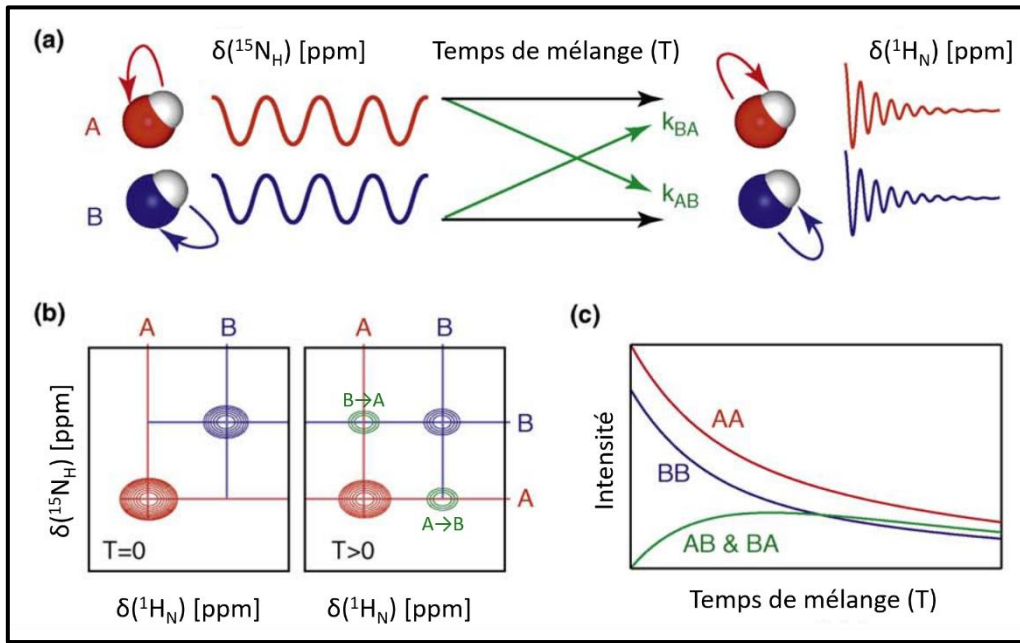


Figure 26 : Spectroscopie d'échange, EXSY ou échange ZZ (adapté de Mittermaier et Kay, 2009) ^[97]. Représentation schématique du schéma d'impulsion (a) et spectres simulés (b) pour une paire de spins ^1H - ^{15}N en échange lent entre deux états, A et B. c) Dépendance des intensités des pics par rapport au temps de mélange T . Les éléments en rouge, bleu et vert correspondent respectivement aux états A, B et à l'échange conformationnel entre A et B pendant T .

$$I_{NN}(T) = \frac{I_{NN}(0) \times A_N}{\lambda_1 - \lambda_2} \times ((\lambda_1 - a_{11}) \times e^{-\lambda_2 T} - (\lambda_2 - a_{11}) \times e^{-\lambda_1 T}) \quad (7)$$

$$I_{II}(T) = \frac{I_{II}(0) \times A_I}{\lambda_1 - \lambda_2} \times ((\lambda_1 - a_{22}) \times e^{-\lambda_2 T} - (\lambda_2 - a_{22}) \times e^{-\lambda_1 T}) \quad (8)$$

$$I_{NI}(T) = \frac{I_{NN}(0) \times A_I}{\lambda_1 - \lambda_2} \times a_{21} \times (e^{-\lambda_1 T} - e^{-\lambda_2 T}) \quad (9)$$

$$I_{IN}(T) = \frac{I_{II}(0) \times A_N}{\lambda_1 - \lambda_2} \times a_{12} \times (e^{-\lambda_1 T} - e^{-\lambda_2 T}) \quad (10)$$

où $\lambda_{1,2} = \frac{(a_{11} + a_{22}) \pm \sqrt{(a_{11} - a_{22})^2 + 4k_{N \rightarrow I}k_{I \rightarrow N}}}{2}$; $a_{11} = R_{1,N} + k_{N \rightarrow I}$; $a_{22} = R_{1,I} + k_{I \rightarrow N}$; $a_{12} = -k_{I \rightarrow N}$; $a_{21} = -k_{N \rightarrow I}$; $I_{NN}(0)$ et $I_{II}(0)$ correspondent respectivement à l'aimantation des états NN et II au départ du temps de mélange ($T=0$) ; $I_{NN}(T)$ et $I_{II}(T)$ correspondent aux pics autos ; $I_{NI}(T)$ et $I_{IN}(T)$ correspondent

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

aux pics croisés ; $k_{N \rightarrow I}$ et $k_{I \rightarrow N}$ correspondent aux constantes de vitesse d'échange de l'état NN à II et inversement, respectivement ; les facteurs A_N et A_I correspondent à l'efficacité de transfert de cohérence, $^{15}\text{N}_H \rightarrow ^1\text{H}_N$, pendant l'étape d'INEPT inverse de la séquence pour l'aimantation associée aux états NN et II, respectivement.

Nous avons utilisé les expériences d'échange ZZ afin de quantifier la vitesse d'échange entre les conformations natives et intermédiaires de nos trois dimères à différentes températures. Elles ont été réalisées avec une série de délais T de 50, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900 et 1000 ms et un délai de recyclage de 2 s. Les volumes des pics du résidu 90 ont été extraits puis traités avec les équations (7) à (10) par le programme Grace© pour le calcul des constantes de vitesse d'échange ($k_{N \rightarrow I}$ et $k_{I \rightarrow N}$) en prenant compte des valeurs expérimentales de R_1 .

La RMN en temps réel

La RMN en temps réel (RMN-TR) est une approche simple et très puissante. Elle permet d'observer des procédés de dynamique d'échange supérieurs ou égaux à la seconde en quantifiant directement les intensités des signaux RMN. Ce type d'expérience peut permettre de suivre l'intensité d'un signal RMN en fonction du temps, au cours d'une réaction chimique par exemple. La limitation pour ce type d'expérience est le temps d'acquisition d'un spectre : le signal y est moyenné sur le temps total d'acquisition. Bien que la rapidité de l'acquisition en RMN-TR soit limitante, le développement d'expériences SOFAST a permis d'élargir l'échelle de temps accessible par cette méthode.

Nous avons utilisé la RMN-TR dans notre étude afin d'effectuer des suivis cinétiques de la formation de l'hétérodimère *EchUaβ* dans différentes conditions. Les détails de l'étude de la cinétique globale sont présentés dans la Partie 1.2.4 qui suit. Selon le cas, des expériences ^{15}N -HSQC [84] ou SOFAST ^{15}N -HMQC [102] ont été enregistrées.

Les intensités des pics des résidus d'intérêt, en particuliers des signaux de l'hétérodimère formé, ont été extraits puis traités avec les équations (11) à (18) par le programme Grace© pour les différentes modélisations de la cinétique globale.

2.4. Détermination de la cinétique globale d'un système biochimique par RMN

Un système peut se trouver en équilibre thermodynamique ou hors équilibre. Dans le premier cas, le système n'évolue pas d'un point de vue macroscopique. Dans le second cas, le système subit une **transformation**. Il évolue au cours du temps d'un état initial à un état final et il est caractérisé par l'**avancement chimique** ξ ($\xi = \frac{dn_i}{\varepsilon_i}$ avec n_i la quantité de matière du constituant i et ε_i le coefficient stœchiométrique associé à i) et la **vitesse de réaction** v_r ($v_r = \frac{d\xi}{dt}$). La transformation peut être incomplète, limitée ou totale.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

Lorsque la réaction est incomplète, elle peut correspondre à un **équilibre chimique**. Il est le résultat de deux réactions antagonistes simultanées, et est associé une **constante d'équilibre K**. Cette constante correspond au rapport des constantes de vitesse des deux réactions antagonistes ($K = k_1/k_{-1}$). Elle est liée à la température (T, en kelvin) par la relation : $\Delta_r G^0(T) = -RT \ln(K)$ où $\Delta_r G^0(T)$ est l'enthalpie libre standard de réaction et R est la constante des gaz parfaits.

Une réaction incomplète peut aussi être une réaction totale ou limitée qui n'a pas encore atteint son état final. L'état de la réaction est alors déterminé par le calcul du **taux d'avancement** (Figure 27) :

$$\tau = \frac{\xi}{\xi_{max}}.$$

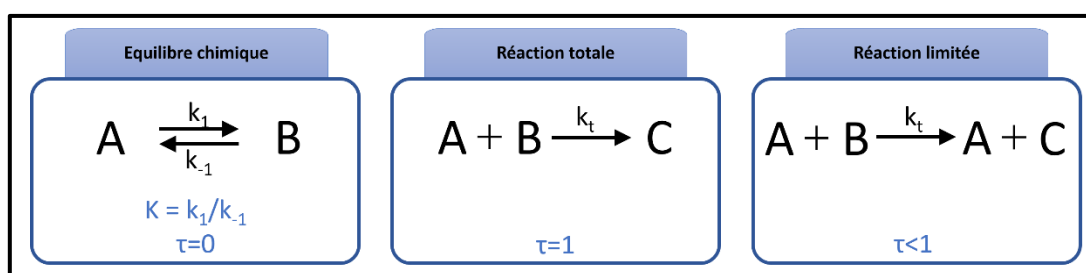


Figure 27 : **Types de réaction chimique.**

Dans le cas d'une réaction totale ou limitée, au moins un des réactifs est totalement consommé par la transformation. Le réactif consommé le plus rapidement est le **réactif limitant**. Dans le cas d'un mécanisme réactionnel, c'est-à-dire une réaction en plusieurs étapes, la réaction élémentaire la plus lente impose sa vitesse (= **réaction limitante**) et on peut déterminer l'ordre global de la réaction (2.4.2). La vitesse de la réaction peut également être modulée par différents facteurs cinétiques tels que la concentration des réactifs, la température ou la présence de catalyseurs. La constante de vitesse de la réaction est, par exemple, liée à la température par la loi d'Arrhenius ($k_t(T) = A \times e^{-\frac{E_a}{RT}}$, où E_a est l'énergie d'activation et A est une constante à déterminer).

2.4.1. Etude de la réaction globale

Afin de caractériser la réaction de formation de l'hétérodimère, une première étude par titrage a été menée. Celle-ci avait pour but de déterminer si l'échange de chaînes pouvait être considéré comme une réaction totale, limitée ou en équilibre chimique.

Protocole de titrage

Nous avons, pour quantifier la formation de l'hétérodimère, réalisé trois titrages de l'homodimère *ECHU* β_2 par l'homodimère *ECHU* α_2 .

Le **premier titrage** (titrage1), réalisé à 20 °C, correspond à des ajouts de 10 % de *ECHU* α_2 ^{15}N à 0,62 mM dans un échantillon de *ECHU* β_2 $^{15}\text{N}^{13}\text{C}$ à 0,45 mM. Ce premier essai a permis de suivre

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

l'apparition de l'hétérodimère au niveau des deux chaînes et la disparition des signaux de l'homodimère *ECHU* β_2 . L'échantillon titrant a été remplacé à 4 °C entre chaque ajout. L'ensemble du titrage a été réalisé à partir de l'acquisition d'une succession de spectres ^{15}N -HSQC de 9 minutes et 35 secondes ($d_1 = 1$ s, $ns = 4$). Le protocole de ce titrage a été déterminé à partir du tableau d'avancement théorique (Tableau 6).

Avancement temps (n° ajout)	$\beta\beta$ (nmol)	+	$\alpha\alpha$ (nmol)	$\rightarrow$	$2\alpha\beta$ (nmol)	V_{tot} (μL)	V_{ajout} (μL)	ratio
$t_i(0)$	90		0		0	200	0	1 : 0
$t_i(1)$	90		9		0	214,63	14,63	1 : 0,1
$t_f(1)$	81		0		18			
$t_i(2)$	81		9		18	229,26	14,63	1 : 0,2
$t_f(2)$	72		0		36			
$t_i(3)$	72		9		36	243,89	14,63	1 : 0,3
$t_f(3)$	63		0		54			
$t_i(4)$	63		9		54	258,52	14,63	1 : 0,4
$t_f(4)$	54		0		72			
$t_i(5)$	54		9		72	273,15	14,63	1 : 0,5
$t_f(5)$	45		0		90			
$t_i(6)$	45		9		90	287,78	14,63	1 : 0,6
$t_f(6)$	36		0		108			
$t_i(7)$	36		9		108	302,41	14,63	1 : 0,7
$t_f(7)$	27		0		126			
$t_i(8)$	27		9		126	317,04	14,63	1 : 0,8
$t_f(8)$	18		0		144			
$t_i(9)$	18		4,5		144	324,36	7,32	1 : 0,85
$t_f(9)$	13,5		0		153			
$t_i(10)$	13,5		4,5		153	331,68	7,32	1 : 0,9
$t_f(10)$	9		0		162			
$t_i(11)$	9		3		162	336,56	4,88	1 : 0,93
$t_f(11)$	6		0		168			
$t_i(12)$	6		3		168	341,44	4,88	1 : 0,97
$t_f(12)$	3		0		174			
$t_i(13)$	3		3		174	346,32	4,88	1 : 1,0
$t_f(13)$	0		0		180			

Tableau 6 : **Tableau d'avancement théorique et volumes des différents ajouts du premier titrage de l'homodimère *ECHU* β_2 par l'homodimère *ECHU* α_2 .**

Le **second** et le **troisième titrage** (titrage 2 à 27 °C et titrage 3 à 37 °C). Les deux protocoles sont donnés en Annexe 3 et Annexe 4. Pour ces nouveaux suivis, nous avons enregistré une succession de spectres ^{15}N -HMQC SOFAST de 6,5 minutes ($d_1 = 0,1$ s, $ns = 32$).

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

Pour les trois titrages, les ajouts ont été réalisés à température ambiante avec le montage présenté en [Figure 28](#). Le paramétrage des expériences RMN était préalablement réalisé sur l'échantillon de *EchU* β_2 à titrer puis réajusté après chaque ajout (impulsion 90°, fréquence de la porteuse).

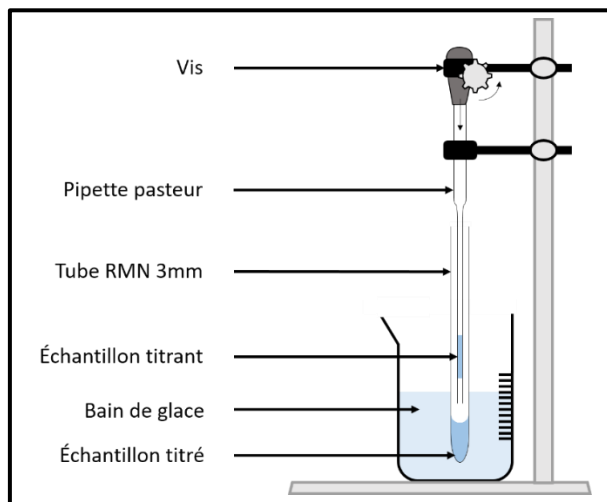


Figure 28 : Schéma du montage pour les essais de titrage de EchU β_2 par EchU α_2 .

Modélisation du titrage

Les différents titrages ont été modélisés par l'équation (11) ci-après :

$$R(\alpha\beta_{app}) = C \times R(\alpha\alpha_{aj}) \quad (11)$$

Où $R(\alpha\beta_{app}) = \frac{I(\beta_{app})}{I(\beta_{app}) + I(\beta_{disp})}$ correspond au ratio de l'apparition de l'hétérodimère, C est une constante à déterminer et $R(\alpha\alpha_{aj})$ correspond à la proportion de réactif ajoutée.

2.4.2. Caractérisation de la vitesse de réaction globale

Nous avons ensuite cherché à caractériser la constante de vitesse du système global à différentes températures.

Modélisation des cinétiques

L'ordre général d'une réaction est $\nu = k \times [A]^x \times [B]^y$ pour une réaction $n A + m B \rightarrow \text{produits}$. x et y sont des nombres souvent entiers, appelés ordres partiels de la réaction, respectivement par rapport aux réactifs A et B. Ils peuvent être différents des coefficients stœchiométriques n et m de l'équation chimique en exemple. La somme $x + y$ correspond à l'ordre global de la réaction.

CINETIQUE D'ORDRE 0

Dans une cinétique d'ordre 0, un seul réactif donne un ou plusieurs produits, comme dans le cas d'une décomposition par exemple. La loi de vitesse dans ce cas est indépendante de la concentration du réactif, [A]. Elle est exprimée en $\text{mol.L}^{-1}.\text{s}^{-1}$.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

$$\text{Loi de vitesse} \quad v = k \times [A]^0 \quad (12)$$

$$\text{Loi de vitesse intégrée} \quad [A] = [A]^0 - kt \quad (13)$$

CINETIQUE D'ORDRE 1

Dans une cinétique d'ordre 1, un seul réactif donne un ou plusieurs produits, comme dans le cas de l'ordre 0. La loi de vitesse dans ce cas-ci est dépendante de la concentration du réactif, $[A]$. Elle est exprimée en s^{-1} .

$$\text{Loi de vitesse} \quad v = k \times [A] \quad (14)$$

$$\text{Loi de vitesse intégrée} \quad [A] = [A]^0 \times e^{-kt} \quad (15)$$

CINETIQUE D'ORDRE 2

Dans une cinétique d'ordre 2, un ou deux réactifs donnent un ou plusieurs produits. La loi de vitesse dans ce cas-ci est dépendante de la concentration d'un réactif, $[A]^2$ ou de deux réactifs, $[A][B]$. Elle est exprimée en $L.mol^{-1}.s^{-1}$.

$$\text{Loi de vitesse} \quad v = k \times [A]^2 (= k \times [A] \times [B]) \quad (16)$$

$$\text{Loi de vitesse intégrée} \quad [A] = \frac{[A]^0}{1 + kt \times [A]^0} \quad (17)$$

(cas d'un réactif)

$$\text{Loi de vitesse intégrée} \quad \ln \frac{[A]^0 \times [B]}{[B]^0 \times [A]} = ([B]^0 - [A]^0) \times kt \quad (18)$$

(cas de deux réactifs)

Protocole de suivi cinétique

Le protocole des suivis cinétiques est donné dans le Tableau 7.

Suivi cinétique	Température	Temps ajout	Temps du suivi	Séquence	Ratio α/β	Nombre d'expériences
1	20 °C	9'40	19,0 h	¹⁵ N-HSQC	0,1:1	118
2	10 °C	6'38	10,9 h	¹⁵ N-HMQC SOFAST	2:1	100
3	20 °C	7'03	12,0 h	¹⁵ N-HMQC SOFAST	1,7:1	210
4	10 °C	6'16	28,3 h	¹⁵ N-HMQC SOFAST	1:1,5	260
5	10 °C	6'26	22,9 h	¹⁵ N-HMQC SOFAST	1:1,5	210

Tableau 7 : Description des suivis cinétiques de la formation de l'hétérodimère.

Chapitre 3. Etape clé pour l'échange de chaînes : le dimère intermédiaire

Dans ce chapitre, nous nous intéressons à la première étape du mécanisme d'échange de la protéine *ECHU* d'un point de vue structural. Il s'agit principalement de **déterminer la structure des dimères dans leurs deux conformations (native et intermédiaire) et d'en étudier la dynamique interne par RMN**. Pour ce faire, une stratégie en quatre étapes a été menée. Dans un premier temps, nous avons suivi l'évolution de l'équilibre thermodynamique N_2/I_2 des dimères sur une large gamme de températures. Dans un second temps, nous avons cherché à résoudre les structures de la conformation native de chaque dimère à basse température ($\sim T_{m1} - 10^\circ\text{C}$). Dans un troisième temps, nous avons cherché à attribuer les résonnances et résoudre les structures de la conformation native de l'hétérodimère à une température correspondant à la température de transition entre les deux états ($\sim T_{m1}$, déterminé par DSC). Pour rappel à cette température, nous avons autant de protéine *ECHU* dans sa conformation native qu'intermédiaire. Enfin, nous avons cherché à résoudre la structure de la conformation intermédiaire de *ECHU* $\alpha\beta$ à cette même température ($\sim T_{m1}$).

3.1. N_2/I_2 : un équilibre thermodynamique très dépendant de la température

A basse température, la forme native de *ECHU* devrait être majoritaire voire unique. Quand on augmente la température, on augmente la proportion de la forme intermédiaire. L'équilibre thermodynamique N_2/I_2 est caractérisé par une température de transition, **T_{m1}** , correspondant à la température à laquelle 50 % de la protéine est dans la conformation N_2 et 50 % dans la conformation I_2 . **Cette température est propre à chaque dimère de *ECHU*.**

Une première étude de l'évolution des signaux des dimères en température avait été réalisée sur *ECHU* β_2 [68] et sur *ECHU* α_2 [69]. L'étude des homodimères était basée sur la caractérisation des perturbations de déplacement chimique en température et des intensités de signaux. Les perturbations de déplacement chimique ont permis la détermination de **coefficients de température** liés à la probabilité de formation de liaisons hydrogène stables. La caractérisation des intensités de signaux avait mis en évidence **trois profils d'intensité** :

- Les résidus du premier profil ont une intensité qui reste stable avec l'augmentation de la température (résidus du début de l'hélice α_2 principalement) ;
- Les résidus du second profil ont une intensité qui reste stable jusqu'à une certaine température (T_{m1}) puis qui décroît pour des températures supérieures (résidus dont les signaux se dédoublent et présentent donc un échange lent entre leur deux conformations native et intermédiaire) ;

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

- Les résidus du troisième profil ont une intensité qui décroît continuellement avec l'augmentation de la température (résidus au niveau des bras et des extrémités des structures secondaires).

En augmentant la température de l'échantillon des trois dimères, nous avons pu suivre par RMN, le **dédoublage de signaux** sur le spectre ^{15}N -HSQC (Figure 29). En effet, comme attendu, certains résidus de chacun de ces dimères présentent deux signaux distincts (30 à 40 % des résidus selon le dimère), caractéristiques d'un processus d'échange lent à l'échelle de la RMN. Le premier signal correspond à la forme native et diminue en intensité avec l'augmentation de la température, alors que le second gagne en intensité. Ce second signal correspond à la **conformation intermédiaire**.

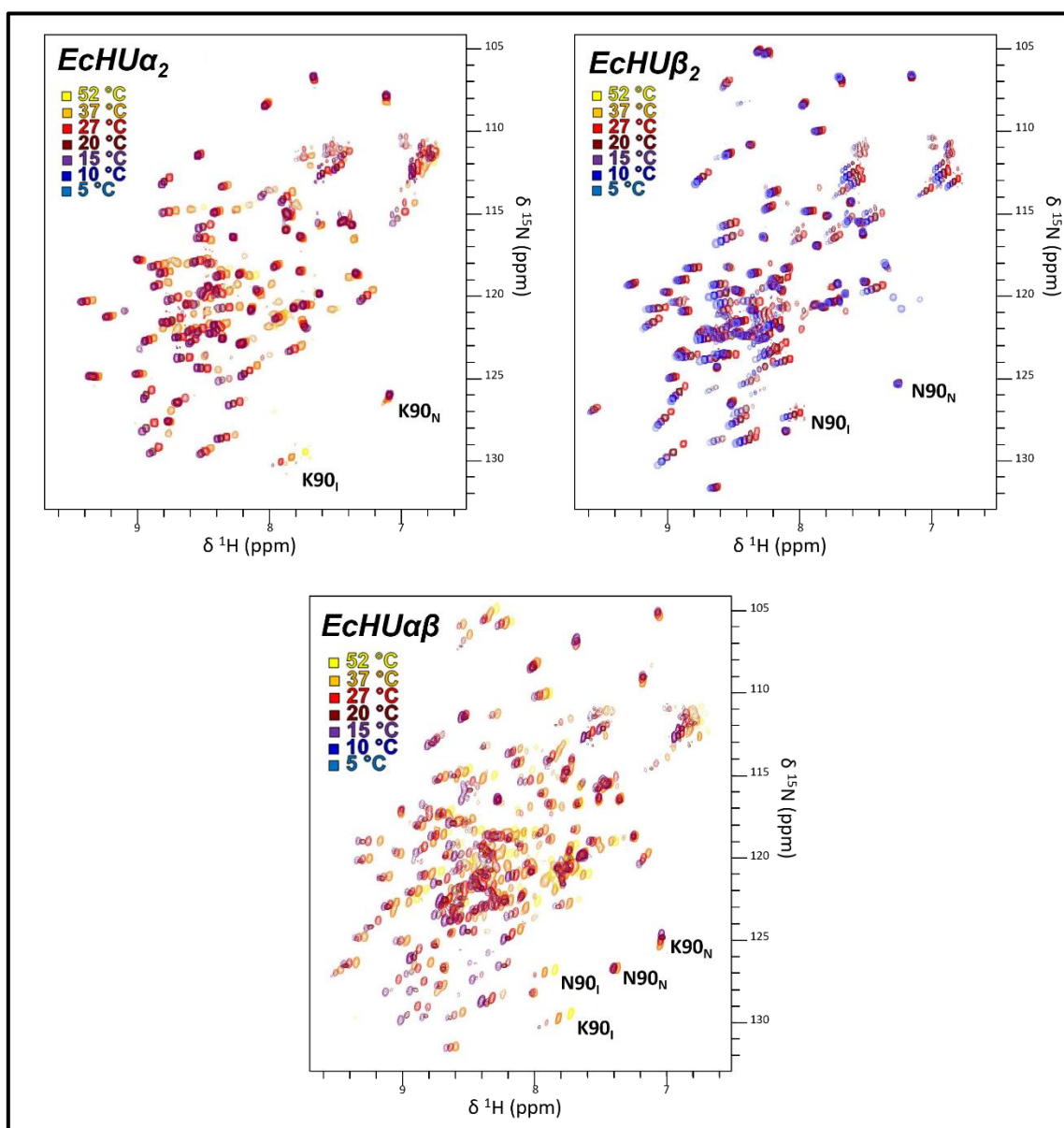


Figure 29 : Suivi de l'évolution des populations N_2/I_2 avec la température sur chacun des dimères *EchU* par des ^{15}N -HSQC. Le résidu C-terminal, Asn90 ou Lys90, est indiqué sur les spectres. Les pics correspondants à la forme native et intermédiaire sont désignés respectivement par l'indice N et I.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

D'un point de vue quantitatif, les signaux du **résidu 90** (Lys90 et Asn90 respectivement pour les chaînes *EchU α* et *EchU β*), ont été particulièrement intéressants pour le suivi en température. Nous avons, grâce à eux, pu quantifier les populations N_2/I_2 sur une large gamme de température (de 5 °C à 33 °C pour *EchU β_2* et de 5 °C à 52 °C pour *EchU α_2* et *EchU $\alpha\beta$*) et déterminer les valeurs de température de transition (T_{m1}) avec une grande précision.

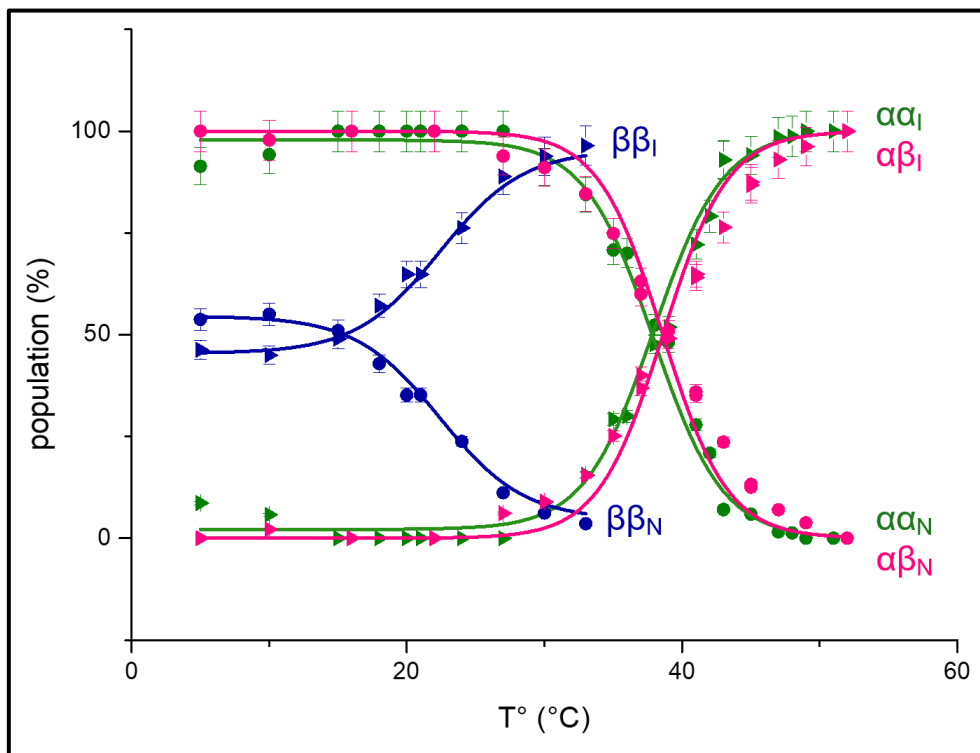


Figure 30 : Suivi des intensités normalisées du résidu 90 des dimères de *EchU* à différentes températures. En bleu, *EchU β_2* , en vert, *EchU α_2* et en rose, *EchU $\alpha\beta$* . N = natif. I = intermédiaire.

Les intensités du résidu terminal ont d'abord été normalisées sur l'ensemble des spectres en température et pour les différentes chaînes protéiques (Figure 30). Nous avons ensuite appliqué l'équation (1), dérivée de l'équation de Boltzmann, pour quantifier les populations N_2/I_2 à n'importe quelle température et déterminer la température du point d'inflexion (T_m). Les températures de transition, T_{m1} , correspondant, par définition, à la température à laquelle $p(I,T) = 50$, ont été déterminées. Les résultats obtenus sont présentés dans le Tableau 8.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

	état	R ²	p(T _{min})	Δ(p(T _{min}))	p(T _{max})	Δp((T _{max}))	T _m [°C]	Δ(T _m) [°C]	A [°C]	Δ(A) [°C]	T _{m1} [°C]	Δ(T _{m1}) [°C]
RMN	αα _N	0,98997	97,8	1,2	0	1,8	37,9	0,2	2,5	0,2	37,8	0,6
	αα _I		1,8		100							
	ββ _N	0,96506	54,6	1,4	4,5	2,6	22,4	0,5	3,0	0,5	15,5	2,9
	ββ _I		45,4		95,5							
	αβ _N	0,99463	94,1	0,6	0	1,4	38,3	0,2	3,8	0,2	37,8	0,5
	αβ _I		5,9		100							
DSC	αα _N	0,99997	99,62	0,02	0	0,02	40,28	0,01	3,66	0,01	40,30	0,02
	αα _I		0,38		100							
	ββ _N	0,9998	88,79	0,11	0	0,03	26,56	0,02	4,76	0,01	25,40	0,04
	ββ _I		11,21		100							
	αβ _N	0,99989	98,88	0,0491	0	0,03	40,44	0,01	3,82	0,01	40,40	0,02
	αβ _I		1,12		100							

Tableau 8 : Caractérisation des populations N₂/I₂ d'EchU en température par RMN et par DSC. T_{m1} est calculé pour p(I, T) = 50 à partir de l'équation contenant les coefficients déterminés par Origin©. L'incertitude sur T_{m1} ont été déterminées par la méthode différentielle de propagation des incertitudes.

Les valeurs de température de transition obtenues pour l'homodimère *EchUα₂* et l'hétérodimère *EchUαβ* sont identiques à l'incertitude près (**T_{m1αα} = T_{m1αβ} = 37,8 °C**). Cette valeur est légèrement inférieure à la température mesurée précédemment par DSC (**T_{m1αα} = 40,3 °C ± 0,02 °C** et **T_{m1αβ} = 40,4 °C ± 0,02 °C**), mais est bien du même ordre de grandeur. Ces valeurs de température de transition sont supérieures à la valeur de température optimale de croissance des bactéries *E.coli* (37 °C), la forme N₂ est donc majoritaire à cette température optimale.

Qualitativement, nous avons observé que l'évolution des signaux sur les spectres ¹⁵N-HSQC relatifs à *EchUα₂* et à *EchUαβ* était similaire. A 5 °C, l'ensemble des signaux des résidus correspondent aux déplacements chimiques de la forme native. **A 20 °C**, le résidu 90 est le premier à apparaître à une valeur de déplacement chimique correspondant à la forme intermédiaire. En continuant d'augmenter la température, le résidu 89 commence à apparaître au déplacement chimique correspondant à la forme I₂. Puis assez rapidement, il en est de même pour les résidus 88, 87 et 86, et jusqu'à l'apparition de la totalité des signaux intermédiaires à 52 °C. Cette observation est en accord avec l'hypothèse selon laquelle l'hélice α₃ serait le premier élément de structure secondaire à se déstructurer et que cette déstructuration permettrait de déclencher la transition entre l'état natif et intermédiaire.

La valeur de la température de transition obtenue pour l'homodimère *EchUβ₂* est bien plus basse (**T_{m1ββ} = 15,5 °C ± 2,9 °C**) que pour les deux autres dimères et est très différente de celle mesurée par DSC (**T_{m1ββ} = 25,4 °C ± 0,04 °C**). Le cas de *EchUβ₂*, est, en effet, un peu plus complexe à analyser. Contrairement à ce qui avait été observé ou relevé jusqu'à présent, nous avons constaté qu'à la température la plus basse testée (5 °C), des résidus comme Asn90 se trouvaient déjà en équilibre entre les formes N₂ et I₂. De plus les résonances correspondantes étaient toutes deux fortement intenses. Cette observation inattendue montre que *EchUβ₂* n'est quasiment jamais à plus

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

de 55 % sous sa forme native quelle que soit la température en solution. Cela ne semblait pas être le cas lors des expériences de DSC.

Ce dimère est très peu compétent pour interagir avec l'ADN à température physiologique et est impliqué dans la réponse au choc froid. Ceci nous indique que la conformation N_2 serait la conformation nécessaire pour interagir avec l'ADN, alors que la conformation I_2 pourrait, non seulement permettre l'échange de chaînes et donc la création de l'hétérodimère, mais aussi constituer une réserve de protéines non compétentes vis-à-vis de l'interaction avec l'ADN, mais pouvant le devenir suite à un changement de conditions de vie des bactéries.

3.2. Caractérisation structurale de la conformation native à basse température

L'attribution du squelette peptidique de chacun des dimères est disponible sur la « Biological Magnetic Resonance Bank » sous les codes 25328 (*EchU α_2*), 25329 (*EchU β_2*) et 25330 (*EchU $\alpha\beta$*)^[65]. L'attribution de l'ensemble des noyaux observables (^1H , ^{15}N , ^{13}C) des trois dimères ainsi que la détermination de leurs structures natives à basse température ont été réalisées. A noter que dans le cas de *EchU α_2* à 20 °C et de *EchU β_2* à 15 °C, il s'agissait de réaliser l'affinement de structures préliminaires obtenues avant mon arrivée au sein du laboratoire^[69]. Les trois structures seront déposées à la PDB.

3.2.1. Attribution et prédiction des éléments de structure secondaire

Avant le lancement des calculs de structure nous avons terminé l'attribution de l'ensemble des résonances des chaînes *EchU* dans les trois dimères. A partir de ces résonances, nous avons pu prédire la structure secondaire des différentes chaînes protéiques et en étudier la dynamique. Nous avons également pu prédire les angles dièdres utilisés dans les calculs.

Attribution des résonances

Nous avons obtenu un pourcentage d'attribution des résonances ^1H , ^{15}N et ^{13}C du **squelette peptidique** des chaînes de HU dans chacun des trois dimères, supérieur à **95 %** (96 % pour *EchU α_2* , 97 % pour *EchU β_2* , 96 % et 97 % respectivement pour *EchU α* et *EchU β* dans *EchU $\alpha\beta$*). Nous avons également pu attribuer plus de **80 % de l'ensemble des hydrogènes des chaînes latérales** de chaque chaîne de HU (91 % pour *EchU α_2* , 93 % pour *EchU β_2* , 89 % et 83 % respectivement pour la chaîne α et la chaîne β dans *EchU $\alpha\beta$*). Le récapitulatif du pourcentage d'attribution des dimères à basse température (à 20 °C pour *EchU α_2* et *EchU $\alpha\beta$* , à 15 °C pour *EchU β_2*) est présenté en illustration dans le Tableau 9. Les déplacements chimiques des trois dimères seront également déposés à la BRMB.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

Attribution		<i>EcHUα₂</i>	<i>EcHUβ₂</i>	<i>EcHUαβ</i>	
				chaîne α	chaîne β
squelette peptidique	total	96 %	97 %	95,9 %	97,5 %
chaînes latérales	H	91 %	93 %	88,5 %	83,3 %
	non-H	69 %	69 %	73,0 %	68,2 %

Tableau 9 : Attribution des dimères de *EchU* dans leur conformation native, à 20 °C pour *EcHUα₂* et *EcHUαβ*, à 15 °C pour *EcHUβ₂*.

Prédiction de structure secondaire

Les prédictions d'angle dièdre, à partir des attributions du squelette peptidique des homodimères, avaient déjà été réalisées et étaient en accord avec les structures cristallographiques ^[69]. Nous avons donc complété cette étude en réalisant la prédiction des angles dièdres pour chaque chaîne *EchU* dans l'hétérodimère grâce au programme DANGLE ^[70]. La **Figure 31** illustre les résultats obtenus. Quelle que soit la chaîne α ou β, on y observe que :

- les résidus 3 à 13 et 18 à 37 se structureraient en hélice α et formeraient le domaine HTH ;
- les résidus 83 à 88 devraient former l'hélice α₃ terminale ;
- les résidus 41 à 44, 48 à 54 et 73 à 81 se structureraient en brins β et formeraient le feuillet β antiparallèle correspondant à la « selle » de la protéine ;
- les bras flexibles des chaînes protéiques formeraient un feuillet β antiparallèle avec les brins β_{2'} et β_{3'} prédits au niveau des résidus 58 à 61 et 68 à 71.

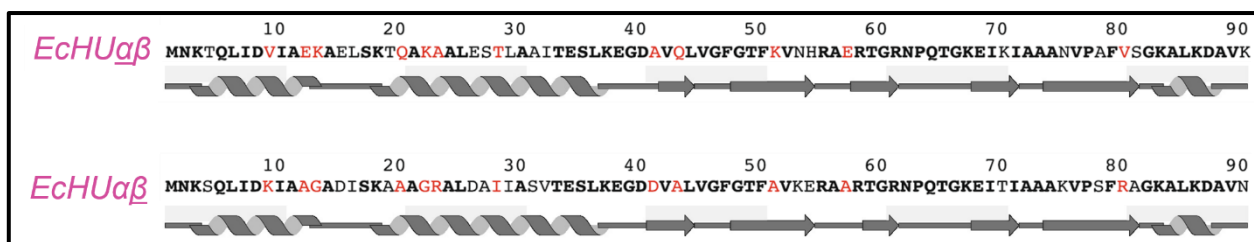


Figure 31 : Prédiction des éléments de structure secondaire des chaînes *EcHUα* et *EcHUβ* de l'hétérodimère natif, à 20 °C.

En combinant ces nouveaux résultats à ceux obtenus sur les homodimères, nous pouvons conclure que les prédictions obtenues sur les trois dimères de *EchU* sont **totale**ment en accord avec les **structures cristallographiques** déposées à la Protein Data Bank (PDB : 1MUL, 4P3V et 2O97 respectivement pour *EcHUα₂*, *EcHUβ₂* et *EcHUαβ*). Sans surprise, ils se structurent donc de façon identique aux chaînes protéiques de la superfamille HU/IHF.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

3.2.2. Structure de la conformation native en solution

Au cours de sa thèse, Rémy Le Meur avait obtenu des structures préliminaires (non affinées) des homodimères en prenant en compte une **symétrie C2**. Cette méthode de calcul permet à ARIA ^[72] de réaliser l'attribution automatique des spectres NOESY de façon plus efficace. De plus, il y a peu d'attributions ambiguës, ce qui permet aux structures de converger plus rapidement. Toutefois, un biais non négligeable est introduit : les bras, qui sont flexibles et n'ont aucune raison d'adopter des conformations symétriques à un temps *t*, apparaissent dans tous les modèles comme symétriques.

Pour cette étude, le mode de calcul choisi pour déterminer les structures des trois dimères est de ne prendre en compte **aucune symétrie** afin que les bras flexibles de chaque chaîne puissent adopter des conformations différentes dans chaque modèle.

Le Tableau 10 regroupe le nombre et le type de contraintes (distances dérivées des NOE, liaison hydrogène, angles dièdres) utilisées lors du dernier « run » pour chacun des dimères. De plus, il présente le résultat des analyses statistiques de validation de structure (Procheck-NMR ^[92,93]).

	<i>EchUα₂</i>	<i>EchUβ₂</i>	<i>EchUαβ</i>
Contraintes RMN			
Contraintes de distance par chaîne			
NOE totaux	1672	1811	3699
Liaisons hydrogènes	72	95	71
Contraintes d'angle dièdres			
	340	309	296
Statistiques structurales calculées sur les 20 meilleurs modèles sélectionnés			
Violations moyennes par structure			
NOEs ≥ 0,3 Å	2,85	3,15	1,25
Liaisons hydrogènes ≥ 0,5 Å	0	0,55	0
Angles dièdres ≥ 15°	0,05	0,1	0
RMSD moyen - backbone (Å)			
résidus 1 à 55 et 75 à 90	0,33	0,57	0,31 (0,24 - chaîne α ; 0,37 - chaîne β)
RMSD moyen - total (Å)			
résidus 1 à 55 et 75 à 90	0,69	1,00	0,70 (0,62 - chaîne α ; 0,78 - chaîne β)
Diagramme de Ramachandran			
Régions les plus favorables	84,4%	81,6%	81,7%
Régions favorables	15,0%	15,4%	15,4%
Régions autorisées	0,3%	2,2%	1,7%
Régions interdites	0,2%	0,7%	1,2%

Tableau 10 : **Contraintes RMN et statistiques structurales de la conformation native de *EchUα₂* à 20 °C, de *EchUβ₂* à 15 °C et de l'hétérodimère à 20 °C.**

Le dernier « run », permettant d'obtenir les structures pour chacun des trois dimères, a été réalisé en prenant 500 structures par itération. Les 200 structures de plus faibles énergies de la dernière itération ont été affinées dans l'eau et nous ont permis d'obtenir 200 structures finales. Parmi celles-ci, les 20 structures respectant au mieux les données expérimentales, la topologie des acides aminés et ayant les énergies les plus basses (totales, électrostatiques, de Van der Waals) ont été

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

sélectionnées afin de **représenter l'ensemble de structure représentatif de chaque dimère de HU**.

L'homodimère *EchUα₂* a été modélisé avec 1672 contraintes de distance, 72 contraintes de liaison hydrogène et 340 contraintes d'angle dièdre. Si l'on compare les nouvelles statistiques ([Tableau 10](#)) aux statistiques obtenues lors du calcul prenant en compte la symétrie de type C2 ([Annexe 2](#)), il n'y a pas de réelle amélioration de la structure précédemment obtenue, mais chaque bras explore de façon indépendante un espace conformationnel important.

L'homodimère *EchUβ₂* a été modélisé avec 1811 contraintes de distance, 95 contraintes de liaison hydrogène et 309 contraintes d'angle dièdre. Cette fois-ci, toutes les statistiques obtenues lors de notre calcul sont améliorées par rapport à celles qui prenaient en compte la symétrie de type C2. En particulier, le nombre de violations de distances dérivées des NOEs supérieures à 0,3 Å et d'angles dièdres supérieurs à 15° passent respectivement de 10 à 3 et de 9 à 0,1. De la même façon que pour *EchUα₂*, les bras flexibles adoptent des conformations indépendantes l'une de l'autre.

L'hétérodimère *EchUαβ* a été modélisé avec 3699 contraintes de distance (~1800 pour *EchUα* et ~1800 pour *EchUβ*), 71 contraintes de liaison hydrogène et 296 contraintes d'angle dièdre.

La [Figure 32](#) présente les structures tridimensionnelles de *EchU* obtenues par RMN (superposition des 20 modèles sélectionnés) ainsi que les structures RX déposées à la PDB.

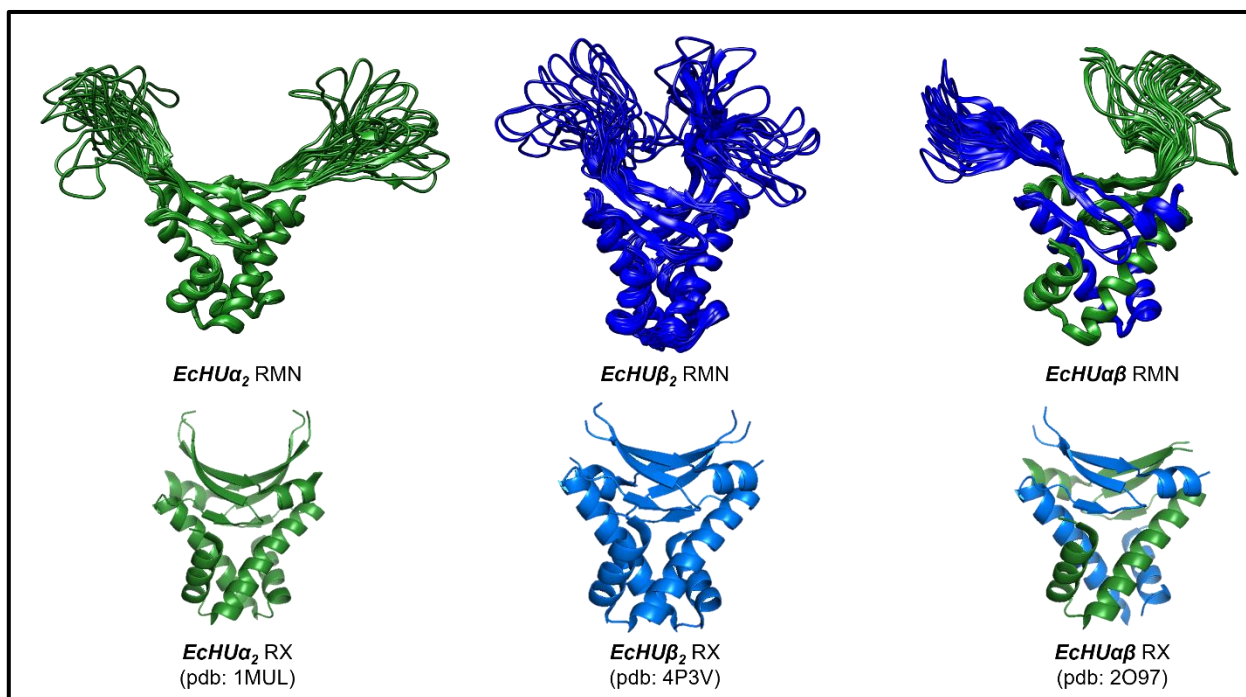


Figure 32 : Superposition des 20 meilleurs modèles de structure de la conformation native de *EchU*. A gauche, *EchUα₂* à 20 °C par RMN et structure RX (pdb : 1MUL), au centre *EchUβ₂* à 15 °C et structure RX (pdb : 4P3V), et à droite, *EchUαβ* à 20 °C et structure RX (pdb : 2O97). Les chaînes α et β sont représentées respectivement en vert et en bleu.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

On observe une excellente superposition des structures avec un **RMSD global** de **0,69 Å** pour l'homodimère $\alpha\alpha$, **1,00 Å** pour l'homodimère $\beta\beta$ et **0,70 Å** pour l'hétérodimère $\alpha\beta$ (0,62 Å pour la chaîne α et 0,78 Å pour la chaîne β). Si nous comparons ces structures obtenues par RMN à celles déposées à la PDB, nous observons un RMSD de 1,25 Å pour *EchU* α_2 , 1,81 Å pour *EchU* β_2 et 1,25 Å pour *EchU* $\alpha\beta$.

Comme pour les homodimères, les structures calculées pour l'hétérodimère dans sa conformation native correspondent aux structures de la superfamille des HU/IHF. Les structures affinées des différents dimères de HU d'*E. coli* sont très proches des structures cristallographiques connues. L'alignement structural (structures RMN *versus* RX) de chacun des trois dimères révèle des différences assez subtiles au niveau du squelette peptidique. En particulier, les extrémités de β_2 et β_3 qui sont prolongées par les bras flexibles ne se superposent pas. Ceci est tout à fait logique puisque la RMN permet d'obtenir dans chaque modèle des conformations différentes des bras qui ont forcément un impact non négligeable sur les extrémités auxquelles ces bras sont liés covalamment.

Il est connu que, bien que les structures des différentes protéines HU soient très similaires, ces protéines présentent des différences de thermostabilité. Cette thermostabilité peut être modérée [57,103–106] ou élevée [29,103,105,107–111]. De nombreuses études ont montré que la température de dénaturation (T_m) augmente lorsque la protéine est plus concentrée ou lorsque la force ionique de la solution est plus grande. Cela suggère que la stabilité des dimères est surtout due à la présence d'interactions hydrophobes [57,104,108]. Par exemple, toutes les phénylalanines des protéines HU sont fortement impliquées dans la formation du cœur hydrophobe présent à l'interface du dimère. Plus la séquence de HU contient de phénylalanines, plus la protéine HU est thermostable [105,108]. Cependant, d'autres facteurs structuraux ne sont pas négligeables comme la présence d'un plus ou moins grand nombre de liaisons hydrogène ou de ponts salins à l'interface du dimère [29,57,103,104,107–111].

Afin d'analyser plus finement les structures présentées sur la [Figure 32](#), nous avons déterminé les interactions polaires et hydrophobes présentes à l'interface des deux chaînes protéiques pouvant être importantes pour leur thermostabilité.

Concernant l'étude des interactions polaires, nous avons utilisé l'outil interactif PDBePISA (pour « Protein Interfaces Surfaces and Assemblies ») [95] permettant la caractérisation des interfaces macromoléculaires. En particulier, il propose une analyse approfondie et automatique des interactions protéine-protéine telles que le nombre de **liaisons hydrogène et de ponts salins inter chaînes**.

Les interfaces des structures en solution obtenues pour l'homodimère *EchU* α_2 présentent cinq liaisons hydrogène (Met1-Ala41, Lys3-Gln43, Thr28-Leu6, Lys37-Leu85 et Gln43-Met1) et un pont

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

salin (Met1-Asp40). Dans la structure cristallographique, nous caractérisons six liaisons hydrogène dont trois sont communes aux structures RMN (**Met1-Ala41**, **Lys3-Gln43**, Lys3-Val45, **Gln43-Met1**, Asn75-Val89 et Asn75-Lys90) et aucun pont salin.

Les interfaces des structures en solution obtenues pour l'homodimère *ECHU* β_2 présentent sept liaisons hydrogène (Met1-Asp40, Met1-Ala43, Gly13-Arg23, Asp15-Arg23, Arg23-Ala14, Lys75-Asn90 et Lys86-Lys75) et deux ponts salin (Met1-Asp40, Asp15-Arg23). Dans la structure cristallographique, nous caractérisons quatre liaisons hydrogène dont trois sont communes aux structures RMN (**Arg23-Ala14**, Lys37-Ala88, **Ala43-Met1**, et **Lys75-Asn90**) et un seul pont salin (**Met1-Asp40**).

Les interfaces des structures en solution obtenues pour l'hétérodimère *ECHU* $\alpha\beta$ présentent neuf liaisons hydrogène (Asn2 $_{\alpha}$ -Asp41 $_{\beta}$, Lys13 $_{\alpha}$ -Ala27 $_{\beta}$, Glu15 $_{\alpha}$ -Arg23 $_{\beta}$, Thr33 $_{\alpha}$ -Leu85 $_{\beta}$, Gln43 $_{\alpha}$ -Met1 $_{\beta}$, Gln43 $_{\alpha}$ -Asn2 $_{\beta}$, Arg55 $_{\alpha}$ -Lys86 $_{\beta}$, Asn75 $_{\alpha}$ -Val89 $_{\beta}$ et Val89 $_{\alpha}$ -Lys75 $_{\beta}$) et trois ponts salin (Met1 $_{\alpha}$ -Asp40 $_{\beta}$, Asp40 $_{\alpha}$ -Met1 $_{\beta}$, Glu15 $_{\alpha}$ -Arg23 $_{\beta}$). Dans la structure cristallographique, nous caractérisons onze liaisons hydrogène dont deux sont communes aux structures RMN (Met1 $_{\beta}$ -Asp40 $_{\alpha}$, Met1 $_{\beta}$ -Ala41 $_{\alpha}$, Met1 $_{\beta}$ -Gln43 $_{\alpha}$, Lys3 $_{\beta}$ -Gln43 $_{\alpha}$, **Gly13 $_{\beta}$ -Ser27 $_{\alpha}$** , Arg23 $_{\beta}$ -Ala14 $_{\alpha}$, Glu34 $_{\beta}$ -Lys13 $_{\alpha}$, Asp40 $_{\beta}$ -Met1 $_{\alpha}$, Asp41 $_{\beta}$ -Met1 $_{\alpha}$, **Ala43 $_{\beta}$ -Met1 $_{\alpha}$** et Ala43 $_{\beta}$ -Lys3 $_{\alpha}$) et trois ponts salins dont deux sont communs (**Met1 $_{\beta}$ -Asp40 $_{\alpha}$** , **Asp40 $_{\beta}$ -Met1 $_{\alpha}$** , Glu34 $_{\beta}$ -Lys13 $_{\alpha}$).

Cette analyse montre qu'il existe **trois zones principales** où des interactions polaires inter chaînes sont présentes dans les trois dimères : le N-terminal de l'hélice α_1 est stabilisé sur le brin β_1 , le haut de l'hélice α_2 est en contact avec l'hélice α_3 , elle-même stabilisée le brin β_3 (Figure 33).

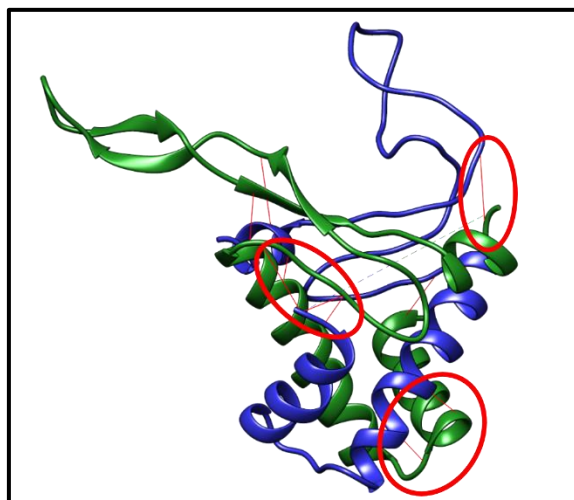


Figure 33 : Mise en évidence des trois zones d'interactions polaires inter chaînes.

Si nous regardons le nombre d'interactions polaires inter chaînes, qui existent dans les trois dimères, obtenu sur les structures RMN et cristallographiques (9 à 10 liaisons hydrogène et 1 à 2 ponts salins), il semble que ces dernières n'ont qu'un impact mineur sur la capacité relative des trois

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

dimères à échanger leur conformation native en leur conformation intermédiaire. Cette observation est tout à fait en accord avec les études précédentes déjà analysées sur les protéines d'HU.

Il a, par contre, été mis en évidence, qu'il existe une corrélation entre la taille de la poche hydrophobe présente à l'interface des deux chaînes protéiques de nos dimères (entre le HTH et la « selle ») et la température T_{m1} . Plus cette poche hydrophobe serait grande, plus T_{m1} serait bas. Cette mise en évidence a été réalisée uniquement sur les structures cristallographiques des trois dimères de *EchU*. Nous avons donc utilisé l'outil disponible en ligne CASTp3^[94], avec une sonde de 1,2 Å de rayon, afin d'infirmer ou de confirmer cette mise en évidence. La [Figure 34](#) présente les résultats obtenus.

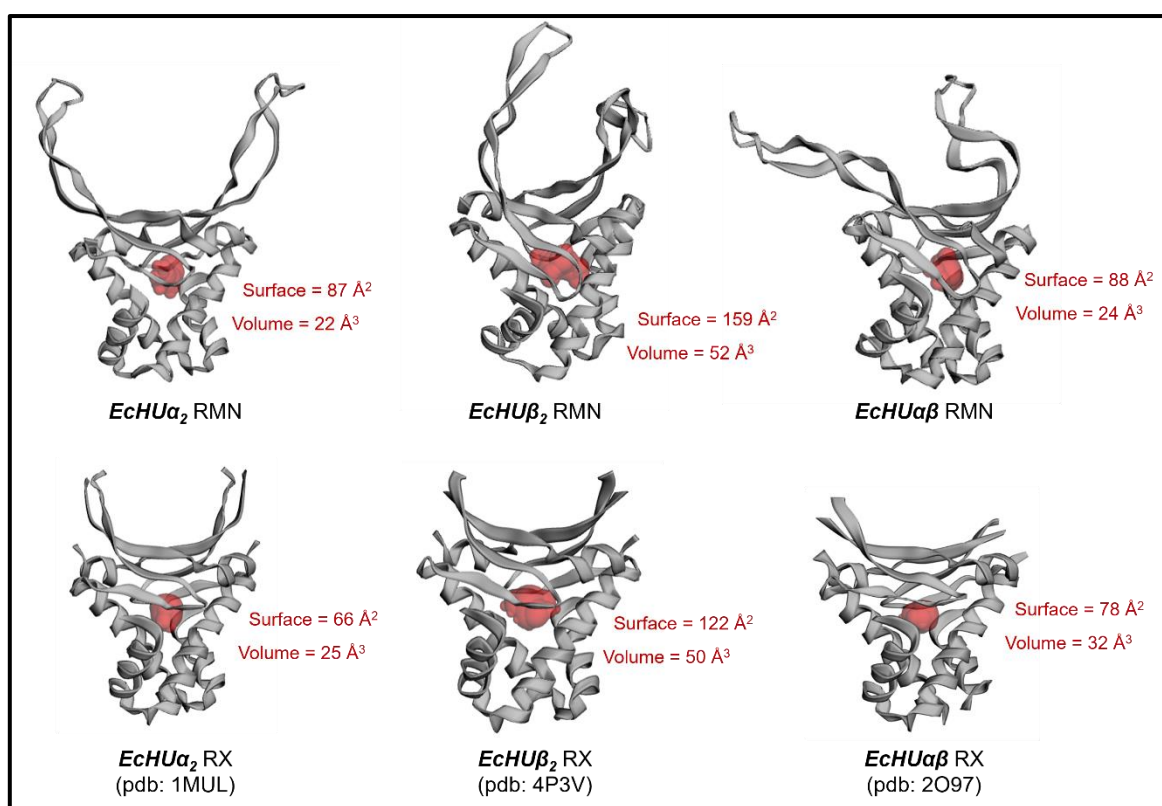


Figure 34 : Poches hydrophobes des dimères de *EchU*. En haut, les valeurs calculées pour les structures en solution par RMN et en bas, celles des structures cristallographiques. Respectivement à gauche, au milieu et à droite, les valeurs correspondant aux dimères *EchUα₂* à 20 °C, *EchUβ₂* à 15 °C et *EchUαβ* à 20 °C par le programme CASTp3^[94].

Si nous regardons à présent la poche hydrophobe qui existe dans les trois dimères (obtenus sur les structures RMN et RX) ainsi que les résidus qui y sont impliqués, nous trouvons dans chaque cas une poche entre le sous-domaine HTH et la « selle ».

Les volumes des poches obtenus pour les dimères *EchUα₂*, $V_{\text{RMN}} = 22 \text{ Å}^3$ ou $V_{\text{RX}} = 25 \text{ Å}^3$, et *EchUαβ*, $V_{\text{RMN}} = 24 \text{ Å}^3$ ou $V_{\text{RX}} = 32 \text{ Å}^3$, sont du même ordre de grandeur. Celui de la poche de l'homodimère *EchUβ₂* est quant à lui plus grand, $V_{\text{RMN}} = 52 \text{ Å}^3$ ou $V_{\text{RX}} = 50 \text{ Å}^3$. Cette observation peut être reliée au nombre de résidus définissant cette cavité hydrophobe. Pour les trois dimères,

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

elle est bordée par les résidus **Leu6**, **Leu25**, **T_α/I_β28**, **L_α/I_β29**, **I_α/V_β32**, **Leu44**, **Phe47** et **Phe50**, auxquels s'ajoutent les résidus Val42, Gly48 et Thr49 de la chaîne α dans *EchUα₂*, ou les résidus Lys3 de la chaîne α et Val42 et Val45 de la chaîne β dans le cas de l'hétérodimère.

Il semble donc clair que la taille de la poche hydrophobe soit bien un des points clés permettant à nos dimères de présenter des propriétés thermodynamiques différentes. Toutes ces analyses vont nous permettre d'avoir une référence pour l'analyse des structures (N₂ et I₂) des dimères à haute température.

3.2.3. Caractérisation de la dynamique interne

La dynamique interne des résidus, à l'échelle de la picoseconde à la nanoseconde, peut être déterminée grâce à des expériences de relaxation (R₁, R₂, NOE) sur les signaux amides ¹H_N-¹⁵N_H. Ces expériences ont été réalisées sur les trois dimères (à 20 °C pour *EchUα₂* et *EchUαβ*, à 15 °C pour *EchUβ₂*). Dans le cas des homodimères, il s'agit de répéter des expériences qui avaient déjà été réalisées [69] et dans le cas de *EchUβ₂* déjà publiées [68]. Les expériences sur l'hétérodimère vont donc compléter les connaissances de la dynamique interne des trois dimères (Figure 35).

Les trois dimères présentent des profils de relaxation tout à fait comparables. Ils sont relativement plats :

- $\langle R_1 \rangle = (0,8 \pm 0,1) \text{ s}^{-1}$ pour *EchUα₂*, *EchUβ₂* et *EchUα* dans l'hétérodimère et $\langle R_1 \rangle = (0,8 \pm 0,3) \text{ s}^{-1}$ pour *EchUβ* dans l'hétérodimère ;
- $\langle R_2 \rangle = (21,8 \pm 2,9) \text{ s}^{-1}$ pour l'homodimère *EchUα₂*, $\langle R_2 \rangle = (21,7 \pm 5,0) \text{ s}^{-1}$ pour l'homodimère *EchUβ₂* et $\langle R_2 \rangle = (21,9 \pm 7,8) \text{ s}^{-1}$ et $\langle R_2 \rangle = (22,3 \pm 5,7) \text{ s}^{-1}$ pour respectivement *EchUα* et *EchUβ* dans l'hétérodimère ;
- $\langle \eta_{NH} \rangle = 0,8 \pm 0,1$ pour *EchUα₂*, *EchUβ₂* et $\langle \eta_{NH} \rangle = 0,8 \pm 0,2$ pour *EchUαβ* dans l'hétérodimère.

Toutefois, les résidus du bras, de Arg55 à Ala74, présentent des valeurs de R₂ et de η_{NH} respectivement inférieures à 20 s⁻¹ et à 0,8, et par une valeur de R₁ supérieure à 1 s⁻¹. La dynamique importante des bras est tout à fait attendue et est compatible avec une exploration conformationnelle importante de ceux-ci.

Les nouvelles mesures de relaxation sur l'homodimère *EchUβ₂* ne confirment pas que les résidus Val32 et Thr33 ont une dynamique accrue à l'échelle de la ps-ns. Cette dynamique avait été observée en dynamique moléculaire où, durant la simulation, un « kink » de l'hélice α₂ apparaissait. Il est à noter, qu'au vu des connaissances actuelles sur l'échange N₂/I₂ (échange lent), il est très improbable que ce « kink » puisse se produire à l'échelle de la ps-ns (échange rapide).

L'augmentation de la dynamique de la Phe79 est, quant à elle, confirmée. Elle a été expliquée grâce à l'analyse de trajectoires de simulation de dynamique moléculaire. Au cours de celle-ci, la chaîne

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

latérale de la Phe79 change d'orientation en alternant entre une orientation enfouie dans le cœur hydrophobe et une orientation exposée au solvant [68].

Pour *EchU* $\alpha\beta$, les expériences doivent être répétées avec des spectres ayant une meilleure résolution afin de permettre une meilleure quantification des intensités et s'assurer de la reproductibilité des mesures.

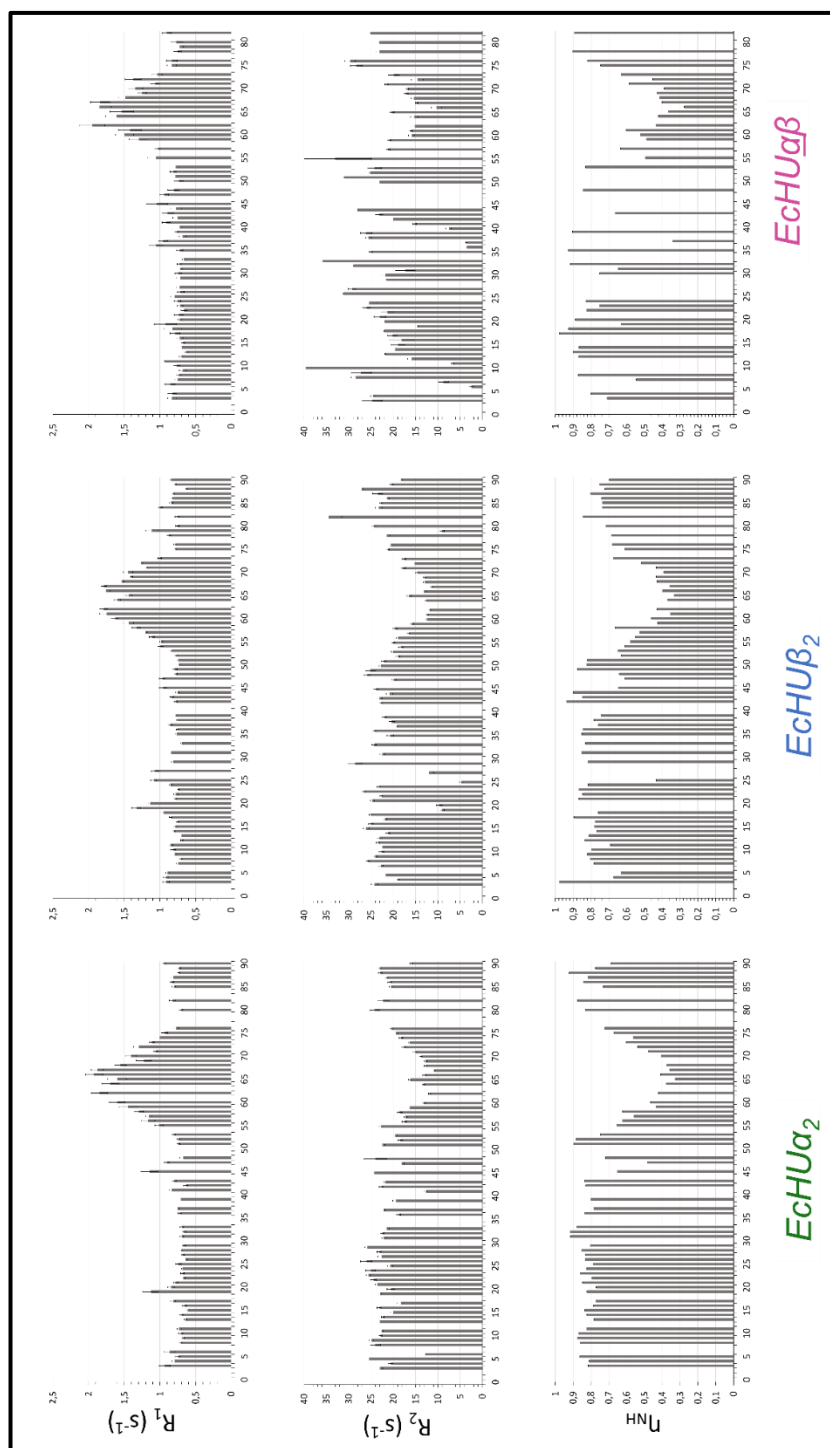


Figure 35 : Relaxation ^{15}N R_1 , R_2 et NOE du squelette peptidique des dimères de *EchU*. Valeurs obtenues à 20 °C pour *EchU* α_2 et *EchU* $\alpha\beta$ et à 15 °C pour *EchU* β_2 .

3.3. Caractérisation structurale de la conformation native à haute température de l'hétérodimère

Une fois les structures natives des trois dimères caractérisées à basse température ($\sim T_{m1} - 10$ °C, déterminé par DSC), nous nous sommes concentrés sur la résolution de la structure de la forme intermédiaire de l'hétérodimère à T_{m1} . Pour rappel, à cette température, le dimère adopte deux conformations également peuplées, N_2 et I_2 , en échange lent. Il nous a donc fallu réussir à distinguer les signaux provenant de chacune des formes, que ce soit sur les spectres standards d'attribution de la chaîne peptidique et des chaînes latérales, ou sur les spectres de type NOESY.

3.3.1. Caractérisation des résonances

Attribution des résonances

L'utilisation des spectres ^{15}N -HSQC du suivi en température des résonances $^1\text{H}_\text{N}$ - $^{15}\text{N}_\text{H}$ des résidus natifs a permis, en parallèle des spectres 3D/4D d'attribution des résonances, de réduire les difficultés liées à la distinction des signaux des deux conformations.

Nous avons obtenu un pourcentage d'attribution des résonances ^1H , ^{15}N et ^{13}C du **squelette peptidique** supérieur à **80 %** (83,3 % et 87,4 % respectivement pour *EchU α* et *EchU β* dans *EchU $\alpha\beta$*). Nous avons également pu attribuer plus de **80 % de l'ensemble des hydrogènes des chaînes latérales** de chaque chaîne de HU (82,7 % et 88,2 % respectivement pour *EchU α* et *EchU β* dans *EchU $\alpha\beta$*). Le récapitulatif du pourcentage d'attribution de l'hétérodimère à 37 °C, est présenté en illustration dans le Tableau 11.

Attribution		<i>EchU$\alpha\beta$</i> (37°C - N_2)	
		chaîne α	chaîne β
total	H	83,3%	87,4%
squelette peptidique	H	89,6%	87,7%
	non-H	90,7%	88,9%
chaînes latérales	H	82,7%	88,2%
	non-H	68,8%	73,9%

Tableau 11 : Attribution de l'hétérodimère de *EchU* dans sa conformation native, à 37 °C.

Prédiction de structure secondaire

A partir des résonances déterminées précédemment, nous avons pu réaliser la prédiction des angles dièdres en utilisant DANGLE ^[70] pour chaque chaîne de l'hétérodimère (Figure 36).

Les deux chaînes sont prédites pour se structurer de la même façon à basse et à haute température. Les résidus 3 à 13, 18 à 37 devraient se structurer en hélice α et former le domaine HTH et les résidus 83 à 88 devraient former l'hélice α_3 terminale. Les résidus 41 à 44, 48 à 54 et 73 à 81

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

devraient se structurer en brins β et devraient former les deux feuillets β antiparallèles correspondant à la « selle » de la protéine. L'augmentation de température semble avoir une légère influence sur les éléments de structure secondaire présents dans les bras (β_2 et β_3). Ces deux brins seraient beaucoup plus courts voire inexistant à haute température.

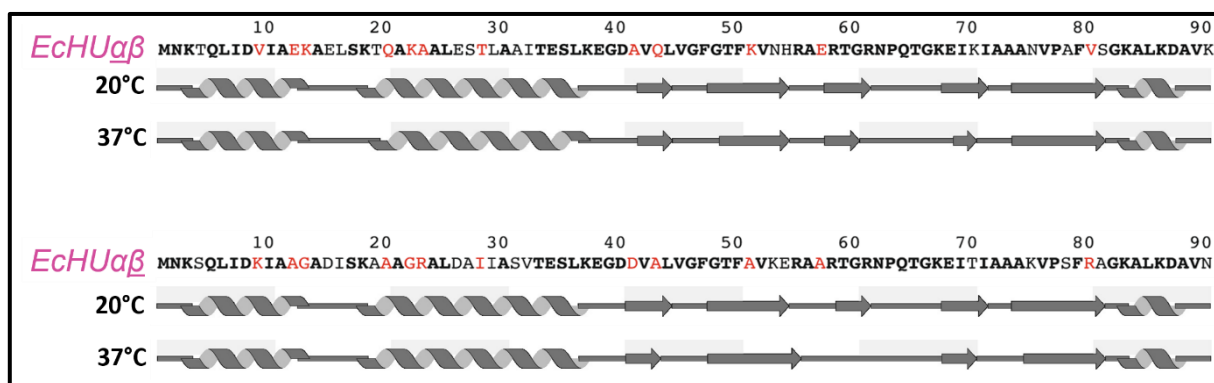


Figure 36 : Prédiction des éléments de structure secondaire des chaînes *EchUα* et *EchUβ* pour la conformation native de l'hétérodimère à 20 °C et à 37 °C.

3.3.2. Structure de la conformation native à haute température

Le Tableau 12 regroupe le nombre et le type de contraintes (distances dérivées des NOE, liaison hydrogène, angles dièdres) utilisées lors du dernier « run » pour l'hétérodimère à 37 °C. De plus, il présente le résultat des analyses statistiques de validation de structure (Procheck-NMR ^[92,93]).

<i>EchUαβ</i> (37°C - N ₂)	
Contraintes RMN	
Contraintes de distance par chaîne	
NOE totaux	3034
Liaisons hydrogènes	74
	(38 - chaîne α ; 36 - chaîne β)
Contraintes d'angle dièdres	
	316
	(158 - chaîne α ; 158 - chaîne β)
Statistiques structurales	
20 meilleurs modèles	
Violations moyennes par structure	
NOEs ≥ 0,3 Å	3,75
Liaisons hydrogènes ≥ 0,5 Å	0
Angles dièdres ≥ 10°	0,1
RMSD moyen - backbone (Å)	
résidus 1 à 55 et 75 à 90	0,355 (0,297 - chaîne α ; 0,412 - chaîne β)
RMSD moyen - total (Å)	
résidus 1 à 55 et 75 à 90	0,766 (0,707 - chaîne α ; 0,826 - chaîne β)
Diagramme de Ramachandran	
Régions les plus favorables	80,2%
Régions favorables	17,8%
Régions autorisées	1,6%
Régions interdites	0,4%

Tableau 12 : Contraintes RMN et statistiques structurales de la conformation native de l'hétérodimère à 37 °C.

Ce dernier « run » a été réalisé en prenant 500 structures par itération. Les 200 structures de plus faibles énergies de la dernière itération ont été affinées dans l'eau et nous ont permis d'obtenir 200

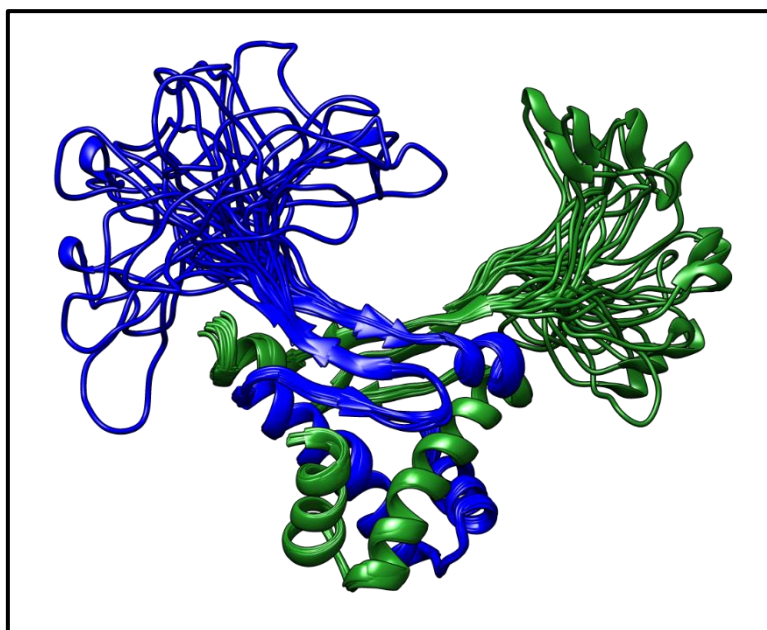
Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

structures finales. Parmi celles-ci, les 20 structures respectant au mieux les données expérimentales, la topologie des acides aminés et ayant les énergies les plus basses (totales, électrostatiques, de Van der Waals) ont été sélectionnées afin de représenter l'ensemble de structure le plus représentatif de la conformation native de *EchUαβ* à 37 °C.

L'hétérodimère *EchUαβ* a été modélisé avec 3034 contraintes de distance, 74 contraintes de liaison hydrogène et 316 contraintes d'angle dièdre. A noter, l'augmentation de la température et l'émergence des signaux de I_2 sur les spectres doivent être la cause de la perte d'environ 600 contraintes de distance (3699 contraintes pour N_2 à 20 °C).

La [Figure 37](#) présente la superposition des 20 modèles sélectionnés pour représenter la structure de la conformation native de l'hétérodimère à Tm1.

On observe une excellente superposition des structures avec un **RMSD global** de **0,77 Å** pour l'hétérodimère αβ (0,71 Å pour la chaîne α et 0,83 Å pour la chaîne β). Pour rappel, la superposition des structures de l'hétérodimère natif à basse température présentait un RMSD global de **0,70 Å** (0,62 Å pour la chaîne α et 0,78 Å pour la chaîne β).



*Figure 37 : Superposition des 20 meilleurs modèles de la conformation native de *EchUαβ* à 37 °C. La chaîne α et la chaîne β sont représentées respectivement en vert et en bleu.*

Evolution de la structure native à haute température

Lorsque l'on regarde le repliement global du dimère, il semble avoir peu évolué avec la température. Seuls les bras semblent particulièrement touchés. Ils présentent une diminution, voire une perte, de structuration en brin β, β₂' et β₃'. Ces deux brins sont beaucoup plus courts voire inexistantes (selon la chaîne et selon le modèle) à haute température. De façon général, les brins, bien qu'identifiés et

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

caractérisés par les logiciels d'analyse d'éléments de structure secondaire tels que PROMOTIF ^[96], ne sont pas toujours dessinés par les logiciels de visualisation de structure.

Lorsque l'on regarde plus finement les structures natives à basse et à haute température en les superposant, on trouve un **RMSD de 1,92 Å** et ce, malgré l'exclusion des bras du calcul (Figure 38). En effet, bien que le repliement de la superfamille des HU soit conservé, des différences locales sur l'ensemble des deux chaînes sont observables : les feuillets formant la « selle » sont maintenus bien qu'en étant plus courts. Concernant les sous-domaines HTH, les deux hélices α_1 et α_2 ont légèrement changé d'orientation : rotation de chacune des hélices de l'ordre de 5° à 10° par rapport à la structure N₂ à basse température. De plus on peut noter que l'hélice α_2 est légèrement incurvée (point de courbure vers le milieu de l'hélice). Concernant la « selle », la longueur des brins β_2 et β_3 des deux chaînes est très dépendante du positionnement dans l'espace du bras flexible et donc diffère sur tous les modèles. Le positionnement de l'hélice α_3 , dépendant non seulement de l'orientation du feuillet à trois brins antiparallèles, mais aussi de l'orientation de l'hélice α_2 (Figure 33), reflète les différences observées au niveau des deux sous-domaines (HTH et « selle »). L'ensemble de ces différences montre une **structure conservée**, mais moins compacte qu'à 20 °C.

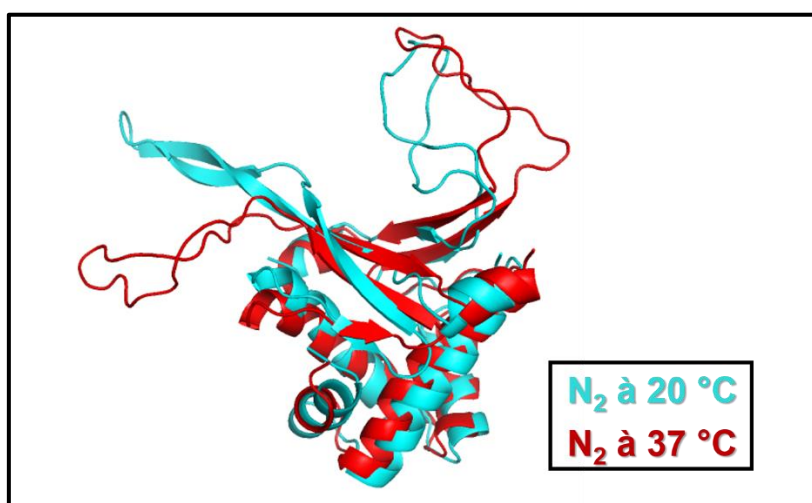


Figure 38 : **Superposition des deux meilleurs modèles structuraux de l'hétérodimère natif à basse et haute température.** Les structures de EchU $\alpha\beta$ à 20 °C et 37 °C sont représentées respectivement en bleu et en rouge.

Afin d'analyser plus finement les structures obtenues, les interactions polaires et hydrophobes, présentes à l'interface des deux chaînes protéiques, ont été analysées avec PDBePISA ^[95] et CASTp3 ^[94].

L'hétérodimère EchU $\alpha\beta$ à 37 °C présente, au niveau de son interface, **dix liaisons hydrogène** (Met1 $_{\alpha}$ -Ser35 $_{\beta}$, Met1 $_{\alpha}$ -Asp40 $_{\beta}$, Lys13 $_{\alpha}$ -Glu34 $_{\beta}$, Ala14 $_{\alpha}$ -Arg23 $_{\beta}$, Glu15 $_{\alpha}$ -Arg23 $_{\beta}$, Lys37 $_{\alpha}$ -Val89 $_{\beta}$, Glu38 $_{\alpha}$ -Asn2 $_{\beta}$, Glu38 $_{\alpha}$ -Lys9 $_{\beta}$, Asn75 $_{\alpha}$ -Asn90 $_{\beta}$ et Lys90 $_{\alpha}$ -Lys75 $_{\beta}$) et **deux ponts salins** (Met1 $_{\alpha}$ -Asp40 $_{\beta}$, Glu15 $_{\alpha}$ -Arg23 $_{\beta}$). La taille de la cavité hydrophobe est de **209,4 Å³** (Figure 39), confirmant

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

que la structure à 37 °C est moins compacte qu'à 20 °C. La cavité est délimitée par les résidus **Leu6**, **Leu25**, **L_α/I_β29**, **I_α/V_β32**, **Val42**, **Q_α/A_β43**, **Leu44**, **Val45**, **Phe47** et **Phe50** (résidus communs aux deux chaînes), auxquels s'ajoutent les résidus Ile10 et Thr28 de la chaîne α, et les résidus Met1, Asn2, Lys3, Ile7, Gly22, Asp26, Ile28, Ser31 et Gly48 de la chaîne β de l'hétérodimère. Il apparaît que l'augmentation de la taille de la poche soit liée au changement d'orientation des hélices α₁ et α₂ du sous-domaine HTH. En particulier, au niveau de la chaîne β de l'hétérodimère.

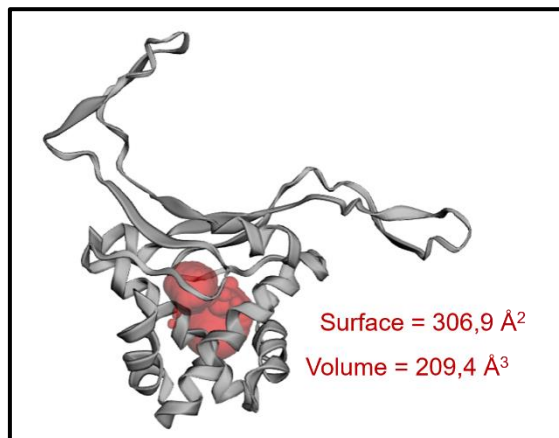


Figure 39 : Cavité hydrophobe de EchUαβ dans sa conformation native à 37 °C, calculée par le programme CASTp3 ^[94].

L'analyse de la structure native à haute température confirme que les interactions polaires, bien que nécessaires à la stabilisation du dimère natif, ne sont pas la source du déclenchement de la transition structurale N₂/I₂. En revanche, l'augmentation significative de la taille de la poche hydrophobe rend la protéine de moins en moins compacte. Elle provoque des changements d'orientation des deux sous-domaines de la protéine et une courbure des hélices de plus en plus grande, jusqu'à provoquer une déstabilisation suffisante des interactions polaires. Nous savons que les premiers signaux observables de la forme intermédiaire proviennent de la déstructuration de l'hélice α₃, **il semble donc que les premières interactions polaires qui doivent disparaître seraient les liaisons hydrogène : Lys37_α-Val89_β, Asn75_α-Asn90_β et Lys90_α-Lys75_β.**

3.4. Caractérisation structurale de la conformation intermédiaire

Maintenant que nous avons identifié l'ensemble des signaux de l'hétérodimère dans sa conformation native à haute température, il nous reste à attribuer l'ensemble des noyaux observables (¹H, ¹⁵N, ¹³C) de la forme intermédiaire à 37 °C et d'en déterminer la structure.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EcHU*

3.4.1. Attribution de l'hétérodimère dans sa conformation intermédiaire

L'attribution des déplacements chimiques de la conformation intermédiaire a été faite à partir des spectres attribués de la conformation native à haute température, du suivi en température des résonances $^1\text{H}_\text{N}$ - $^{15}\text{N}_\text{H}$ et de l'exploitation des expériences d'échange CPMG et ZZ.

Pour rappel, environ 36 % des résidus de l'hétérodimère présentent un échange lent à 37 °C (dédoublément de pics observables sur les spectres ^{15}N -HSQC). Nous nous sommes demandés s'il était possible que la conformation N_2 et I_2 puissent présenter un échange intermédiaire (échelle de la micro à la milliseconde) pour les autres résidus. Afin d'infirmer ou de confirmer cette possibilité, des expériences RD ^{15}N -CPMG ont été réalisées et nous ont permis de confirmer une absence totale d'échange intermédiaire. **Cette information nous a permis d'affirmer que les signaux non dédoublés pouvaient être attribués aux deux conformations.**

L'attribution de la forme intermédiaire de l'hétérodimère a ainsi pu être complétée. Les pourcentages d'attribution sont répertoriés [Tableau 13](#) et la [Figure 40](#) présente les spectres ^{15}N -HSQC de l'hétérodimère attribués à 20 °C et à 37 °C.

Attribution		<i>EcHU</i> $\alpha\beta$ (37°C - I_2)	
		chaîne α	chaîne β
total	H	83,7%	88,8%
squelette peptidique	H	82,6%	81,2%
	non-H	81,1%	79,6%
chaînes latérales	H	82,9%	89,4%
	non-H	68,1%	67,8%

[Tableau 13](#) : Attribution de l'hétérodimère de *EcHU* dans sa conformation intermédiaire à 37 °C.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

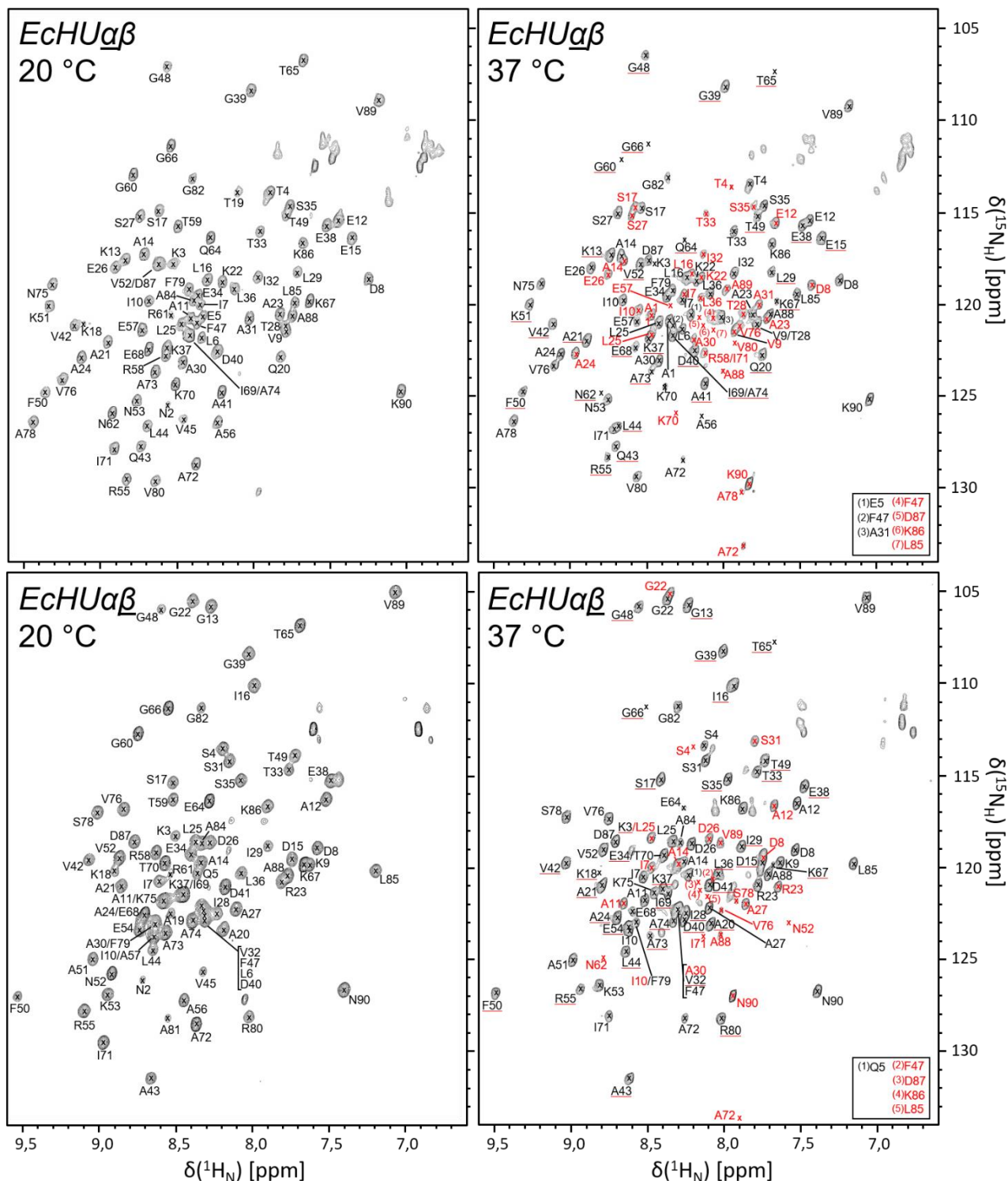


Figure 40 : Spectres ^{15}N -HSQC des chaînes α et β de *EchU $\alpha\beta$* à 20 °C et à 37 °C.
 Représentation à gauche et à droite des spectres attribués respectivement à 20 °C et à 37 °C. Les attributions de la forme native sont en noir, en rouge pour la forme intermédiaire et soulignées lorsque les pics des deux formes sont superposés.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

Perturbation des déplacements chimiques

La [Figure 41](#) présente les perturbations de déplacement chimique $^1\text{H}_\text{N}$ - $^{15}\text{N}_\text{H}$ entre la conformation native haute température et la forme intermédiaire. Les résidus du bras ne présentent pas de CSP, ce qui est attendu. En effet, les bras ne peuvent pas être considérés comme pouvant changer de conformation du fait de leur grande flexibilité. Les résidus des brins β_1 et β_2 de la « selle » ne présentent pas de CSP. La majorité des résidus du brin β_3 sont impactés, indiquant que ce brin ou que des interactions longue distance pourraient avoir disparu. Les résidus impliqués dans les hélices α_1 et α_2 sont faiblement influencés par le changement de conformation ($0,4 > \text{CSP} \geq 0,1$ ppm). Ceci indique que les hélices sont globalement maintenues, mais qu'il existerait un changement d'environnement chimique. Ceci semble confirmer notre hypothèse d'une dissociation partielle ou total des deux domaines HTH. Enfin, l'hélice C-terminale de la protéine montre de grandes valeurs de CSP ($> 0,5$ ppm) pouvant correspondre à un changement de conformation (perte d'hélicité).

Ces résultats sont tout à fait compatibles avec ce qui avait été observé pour les homodimères ([Figure 42](#)). Les structures des formes intermédiaires des trois dimères pourraient être très similaires.

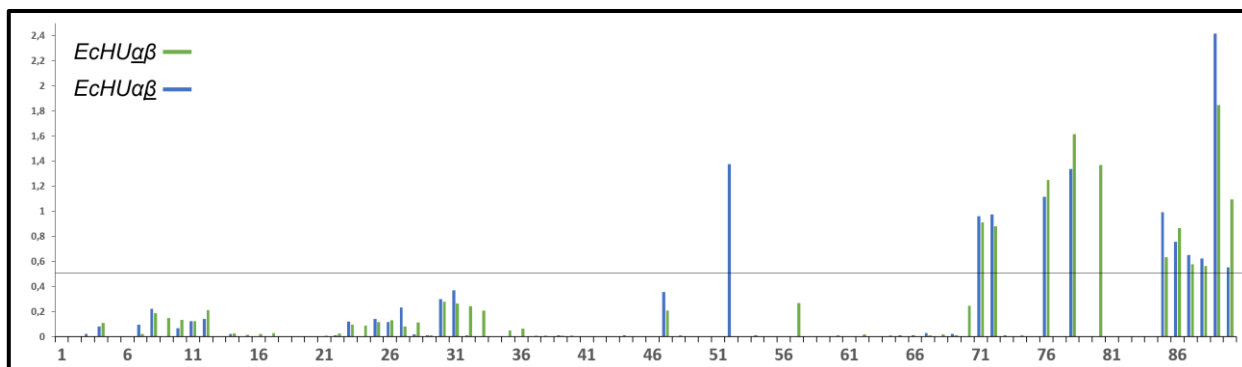


Figure 41 : Perturbation des déplacements chimiques entre la conformation native et intermédiaire de *ECHUαβ* à 37 °C. Les valeurs correspondant à la chaîne α et à la chaîne β sont représentées respectivement en vert et en bleu.

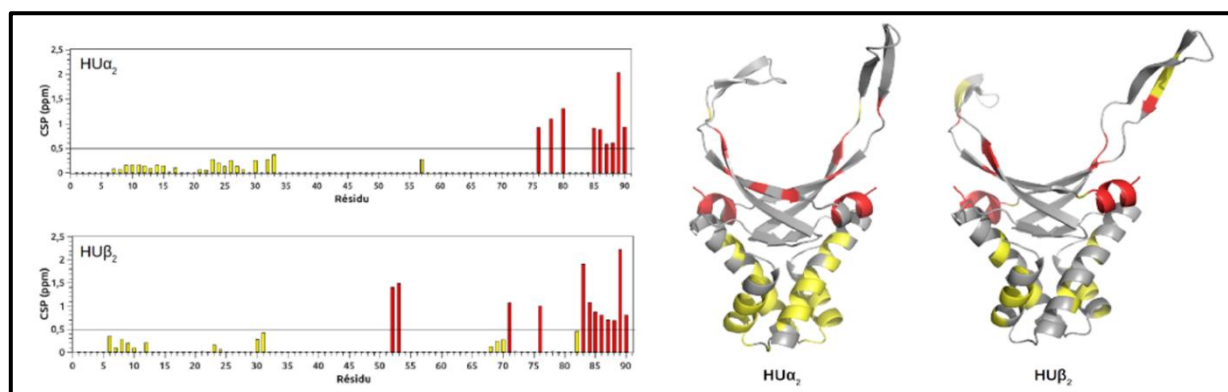
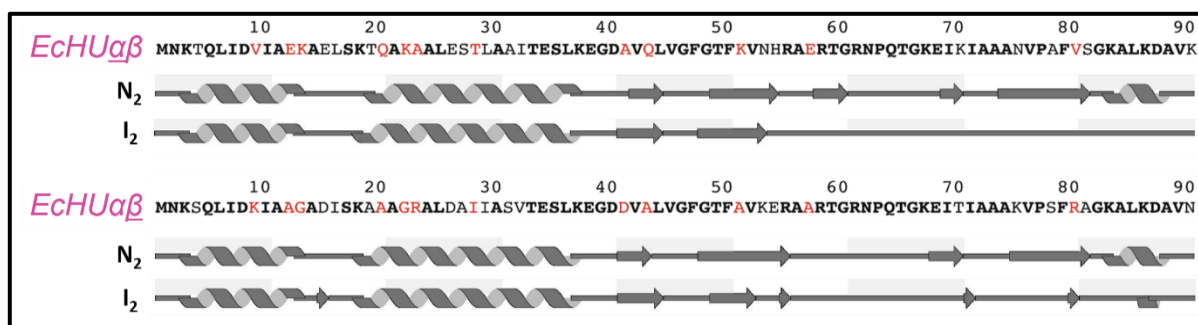


Figure 42 : Perturbation des déplacements chimiques entre les conformations natives et intermédiaires des homodimères (adapté de la thèse de Rémy Le Meur, 2015) ^[69].

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

Prédiction de structure secondaire :

A partir des résonances déterminées précédemment, nous avons réalisé la prédiction des angles dièdres en utilisant DANGLE ^[70] pour chaque chaîne de l'hétérodimère ([Figure 43](#)). Ces prédictions confirment que les hélices α_1 et α_2 seraient maintenues, que le feuillet β pourrait n'être formé que de deux brins (β_1 et β_2) et finalement que l'hélice C-terminale ne serait plus présente. Elles sont, de plus, en accord avec les résultats obtenus sur les homodimères ^[69], confirmant que **les structures des intermédiaires seraient très semblables**.



[Figure 43](#) : **Prédictions des éléments de structure secondaire entre la conformation native (N_2) et intermédiaire (I_2) de l'hétérodimère.**

3.4.2. Structure de l'intermédiaire à haute température

L'hétérodimère *EchUαβ* a été modélisé avec 2840 contraintes de distance, 74 contraintes de liaison hydrogène et 298 contraintes d'angle dièdre. Le dernier « run » a été réalisé en prenant 500 structures par itération. Les 200 structures de plus faibles énergies de la dernière itération ont été affinées dans l'eau et nous ont permis d'obtenir 200 structures finales. Parmi celles-ci, les 20 structures respectant au mieux les données expérimentales, la topologie des acides aminés et ayant les énergies les plus basses (totales, électrostatiques, de Van der Waals) ont été sélectionnées afin de représenter **l'ensemble de structure représentatif de la conformation intermédiaire de *EchUαβ* à 37 °C**. Le [Tableau 14](#) regroupe le nombre et le type de contraintes (distances dérivées des NOE, liaison hydrogène, angles dièdres) utilisées lors du dernier « run » de calcul. De plus, il présente le résultat des analyses statistiques de validation de structure (Procheck-NMR ^[92,93]).

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

<i>EchUαβ</i> (37°C - I ₂)	
Contraintes RMN	
Contraintes de distance par chaîne	
NOE totaux	2840
Liaisons hydrogènes	74 (39 - chaîne α ; 35 - chaîne β)
Contraintes d'angle dièdres	280 (138 - chaîne α ; 142 - chaîne β)
Statistiques structurales	20 meilleurs modèles
Violations moyennes par structure	
NOEs ≥ 0,3 Å	5,90
Liaisons hydrogènes ≥ 0,5 Å	0
Angles dièdres ≥ 10°	0,2
RMSD moyen - backbone (Å)	
résidus 1 à 37 (HTH)	0,603 (0,610 - chaîne α ; 0,595 - chaîne β)
résidus 38 à 54 (feuillet β)	0,578 (0,334 - chaîne α ; 0,821 - chaîne β)
résidus 55 à 80	0,864 (0,669 - chaîne α ; 1,060 - chaîne β)
résidus 81 à 90 (C-term)	3,067 (2,430 - chaîne α ; 3,704 - chaîne β)
RMSD moyen - total (Å)	
résidus 1 à 37 (HTH)	0,972 (0,982 - chaîne α ; 0,963 - chaîne β)
résidus 38 à 54 (feuillet β)	0,985 (0,714 - chaîne α ; 1,255 - chaîne β)
résidus 55 à 80	1,389 (1,149 - chaîne α ; 1,629 - chaîne β)
résidus 81 à 90 (C-term)	3,863 (3,300 - chaîne α ; 4,426 - chaîne β)
Diagramme de Ramachandran	
Régions les plus favorables	54,9%
Régions favorables	35,7%
Régions autorisées	7,3%
Régions interdites	2,1%

Tableau 14 : Contraintes RMN et statistiques structurales de la conformation intermédiaire de l'hétérodimère à 37 °C.

La [Figure 44A](#) présente la superposition des 20 modèles sélectionnés pour représenter la structure de la conformation intermédiaire de l'hétérodimère à Tm1. La structure affinée de la conformation I₂ est **d'apparence globulaire et a perdu son axe de symétrie**. Cet aspect globulaire est lié au repliement du bras de chaque chaîne sur l'hélice α₂ de la chaîne à laquelle il appartient. Il est à noter qu'aucune contrainte de distance n'a été imposée entre les résidus des bras et ces hélices. Les bras restent complètement flexibles dans la conformation intermédiaire. La sélection d'une conformation particulière est certainement le résultat d'un artefact de la méthode de sélection des structures au cours du processus itératif : elle permet d'avoir des structures de plus faibles énergies (Totale, électrostatiques et Van der Waals), tout en respectant les contraintes expérimentales. Nous nous concentrerons donc sur la description de la structure sans prendre en compte les bras.

L'utilisation du logiciel PROMOTIF ^[96] permet de mettre en évidence que les deux domaines HTH ([Figure 44B et C](#)) sont toujours présents, mais sont maintenant formés d'une succession d'hélices et de tours. Pour la chaîne *EchUα*, l'hélice jusqu'à présent nommée **α₁ dans les structures natives** est maintenant composée d'une **hélice α**, résidus Val9 à Ala14, précédée d'un tour hélicoïdal de type I, résidus Lys3 à Asp8. L'hélice jusqu'à présent nommée **α₂ dans les structures natives** est à présent composée de **deux hélices 3₁₀** (Ala31 à Thr33 et Ser35 à Lys37) précédées d'un tour hélicoïdal de type I (Leu25 à Leu29). Pour la chaîne *EchUβ*, la première « hélice » du **motif HTH** est de **type I**, longue de 9 résidus (Lys3 à Ala11), alors que la seconde « hélice » est composée de **deux hélices α** (Arg23 à Asp26 et Ile29 à Glu38).

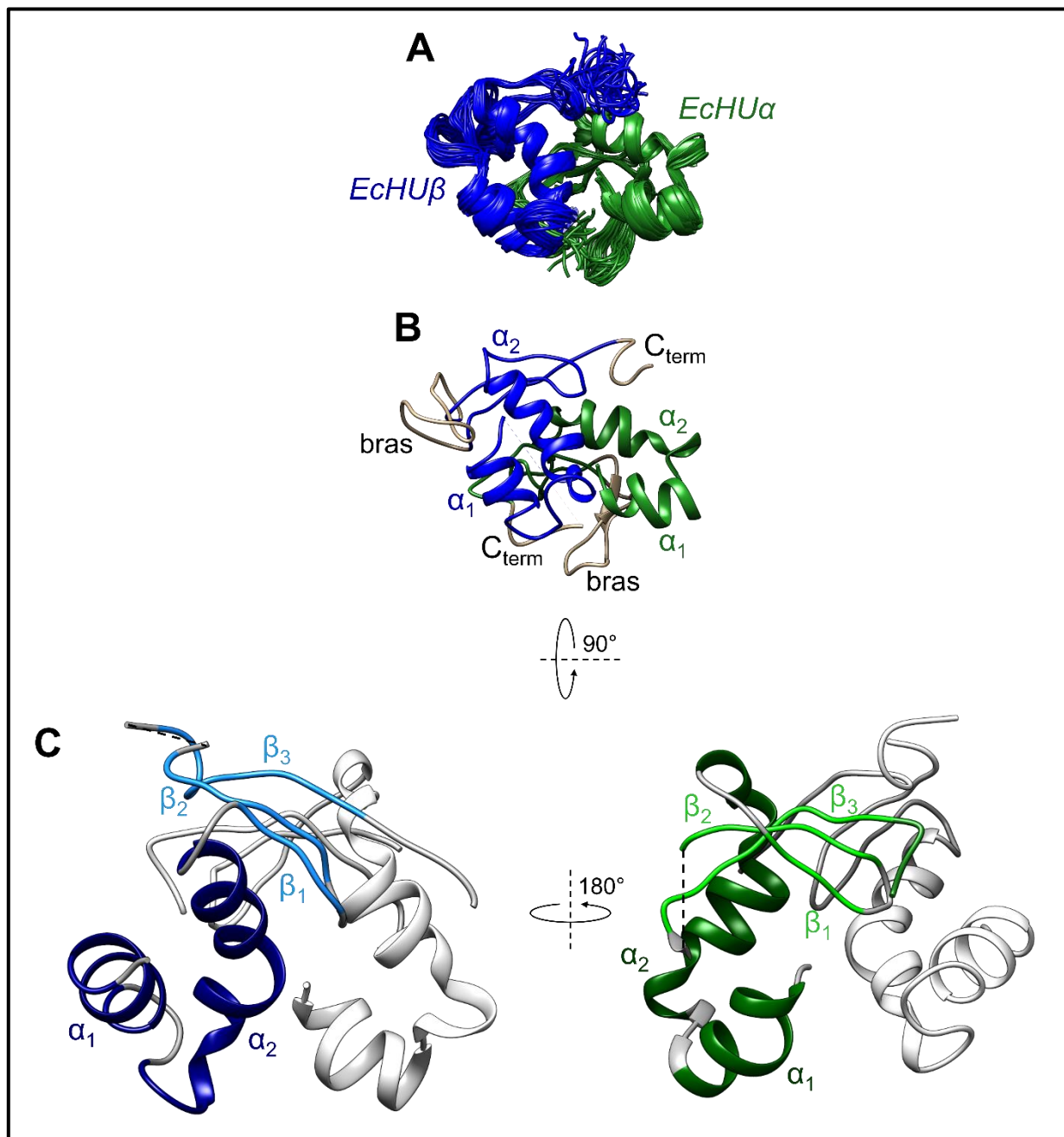


Figure 44 : Structure tridimensionnelle de l'hétérodimère *EcHU* $\alpha\beta$ dans sa conformation intermédiaire, à 37 °C. A – Superposition des 20 meilleurs modèles. B – Représentation du modèle de plus basse énergie. C – Deux autres orientations du modèle de plus basse énergie où les bras et l'extrémité C-terminale ont été masqués. La chaîne α et la chaîne β sont représentées respectivement en vert et en bleu. Les appellations α_i et β_i correspondent ici aux éléments de structure secondaire présents dans la conformation native.

La conformation intermédiaire présente un pourcentage d'hélicité de **24,4 %** (α : 13,3 % ; β : 3,3 % et type I : 7,8 %). Ce pourcentage d'hélicité est du même ordre de grandeur que celui observé par les expériences de dichroïsme circulaire (21%) [57].

Les deux domaines HTH se sont dissociés. Les hélices formant chaque domaine HTH s'éloignent l'une de l'autre de **3 Å** et interagissent avec un angle de **160°** dans chacune des deux chaînes, alors

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

qu'elles sont à une distance de 7,8 Å et interagissent avec un angle de 120° dans la forme native. Il y a donc une **réorientation non négligeable des deux « hélices » lors de la transition N₂/I₂**. Ceci avait déjà été observé lors de l'établissement du premier modèle de l'intermédiaire de *EchU*β₂ obtenu par modélisation moléculaire « tout atome »^[68], mais aussi lors de la modélisation des modèles intermédiaires « gros grains » des deux homodimères^[69].

Il est à noter que, sur la totalité du dimère, seuls deux brins β sont détectés au niveau de la chaîne *EchU*α, β₁ (Val42 à Leu44) et β₂ (Gly48 à Phe50). **La « selle » ne peut plus être définie officiellement comme l'assemblage de deux feuillets β, mais reste identifiable** (Figure 44C). Elle est maintenue par des liaisons hydrogène (observées sur la structure) ainsi que par le « π-stacking » des phénylalanines 47 et 50 sur au moins une des deux chaînes.

On observe au niveau :

- du sous domaine HTH (résidus 1 à 37), un **RMSD global de 0,97 Å** (0,98 Å pour la chaîne α et 0,96 Å pour la chaîne β) ;
- des brins β₁ et β₂ (résidus 38 à 54), un **RMSD global de 0,99 Å** (0,71 Å pour la chaîne α et 1,26 Å pour la chaîne β) ;
- de l'extrémité C-terminale (résidus 81 à 90), un **RMSD global de 3,86 Å** (3,30 Å pour la chaîne α et 4,43 Å pour la chaîne β).

La forme intermédiaire de l'hétérodimère *EchU*αβ à 37 °C ne présente, au niveau de son interface, plus qu'une **seule liaison hydrogène** et un **unique pont salin** situés **entre les résidus Lys3_α et Asp15_β** (utilisation de PDBePISA^[95]) et une cavité hydrophobe de **164,4 Å³** (Figure 45) se présentant sous la forme d'un **sillon** dont une extrémité est ouverte vers le solvant.

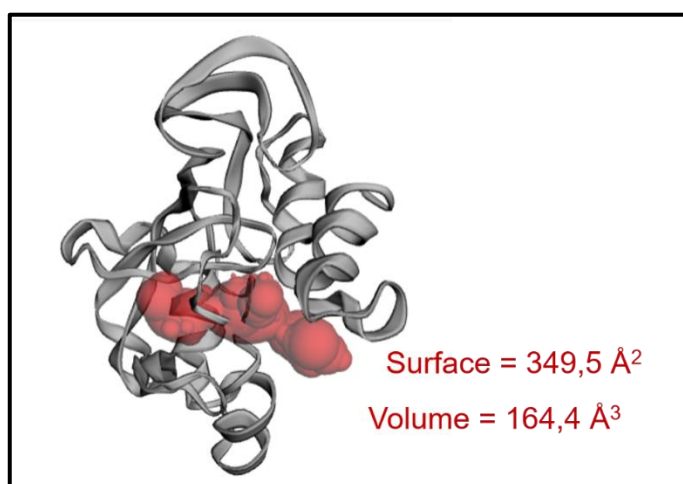


Figure 45 : Cavité hydrophobe de *EchU*αβ dans sa conformation intermédiaire à 37 °C, calculée par le programme CASTp3^[94].

Les résidus définissant cette cavité sont : Met1, Asn2, Lys3, Ile7, Ala41, Val42, Gln43, Leu44, Val45, Thr59, Asn62, Lys67, Glu68, Ile69, Lys70, Leu85, Ala88, Val89 et Lys90 de la chaîne α et Ala14,

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

Asp15, Ile16, Ala21, Arg23, Ala24, Ile28 et Ile29 de la chaîne β . Il ne s'agit donc pas de la poche hydrophobe existant dans le dimère natif. Les résidus de la poche initiale ($K_{\alpha}3$, Leu6, Leu25, $T_{\alpha}/I_{\beta}28$, $L_{\alpha}/I_{\beta}29$, $I_{\alpha}/V_{\beta}32$, $V_{\beta}42$, Leu44, $V_{\beta}45$, Phe47 et Phe50) sont importants pour la stabilisation de la forme native, en permettant entre autres l'association des deux domaines HTH. Il est intéressant de regarder où ils se positionnent dans cette structure intermédiaire dite « ouverte ». Parmi ces résidus, Lys3, Leu6, Leu25 et Leu 29 de *EchU α* , ainsi que Ile28, Leu44, Val45 et Phe47 de *EchU β* sont maintenant **accessibles au solvant** (Figure 46). Ils forment un **patch hydrophobe à la surface de I_2** et **pourraient permettre l'association de deux dimères dans leur conformation intermédiaire et donc favoriser la formation du tétramère transitoire**.

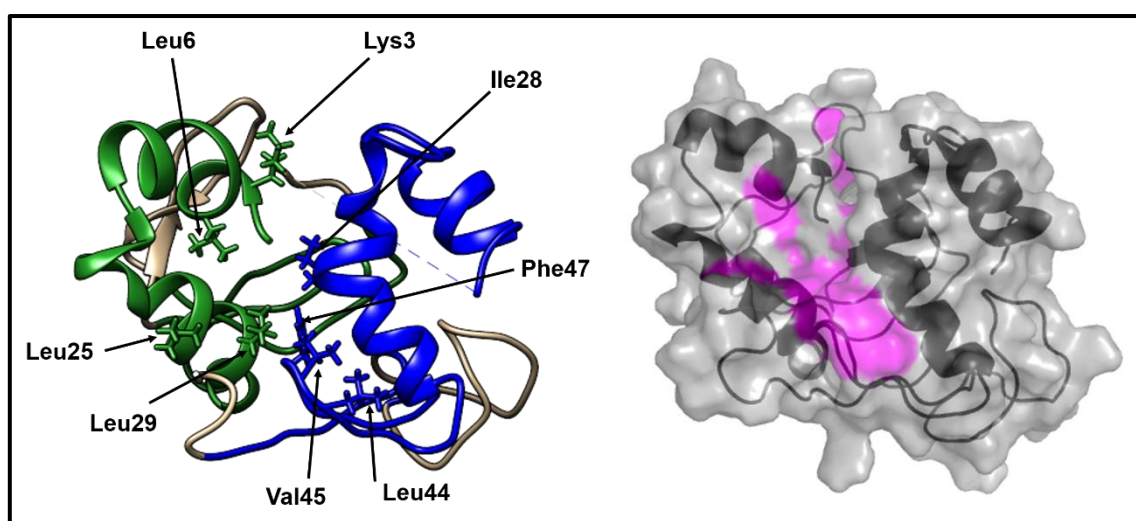


Figure 46: Position des résidus considérés comme des déterminants structuraux de la stabilité de la forme native. Gauche – représentation en bâton sur le squelette de la conformation intermédiaire. La chaîne α et la chaîne β sont représentées respectivement en vert et en bleu. Droite – mise en évidence d'un patch hydrophobe (rose) à la surface de I_2 .

3.5. Conclusion

Dans ce chapitre, nous nous sommes intéressés à la première étape du mécanisme d'échange de chaînes de la protéine *EchU* d'un point de vue structural. Nous avons présenté **cinq structures déterminées par RMN**. Les trois premières sont celles des conformations natives des trois dimères de *EchU* à faible température. Ces structures sont typiques des structures des protéines de la superfamille HU/IHF. Les dimères natifs de *EchU* présentent trois zones d'interactions polaires inter chaînes permettant : d'une part, le positionnement du N-terminal de l'hélice α_1 sur le brin β_1 et d'autre part, la stabilisation de l'hélice α_3 en établissant des contacts entre les résidus de celle-ci avec le haut de l'hélice α_2 et le brin β_3 . Une poche hydrophobe est présente entre la « selle » et les domaines HTH. Il a été établi que les résidus définissant cette poche ainsi que la taille de celle-ci sont des points clés permettant d'expliquer que les trois dimères de *EchU* présentent des propriétés thermodynamiques propres.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EcHU*

Ces structures nous ont par la même occasion permis de mettre en place un protocole de calcul n'incluant aucune contrainte de symétrie. Nous avons alors résolu la structure de la conformation native de l'hétérodimère à une température correspondant à la température de transition entre les états N_2 et I_2 ($\sim T_m1$, déterminée par RMN et confirmant celle obtenue par DSC). A cette température, la structure du dimère natif ne présente pas de différence majeure avec celle obtenue à basse température. Nous avons toutefois noté une diminution, voire une perte, de structuration au niveau des brins β_2' et β_3' des bras flexibles et une forte augmentation du volume de la cavité hydrophobe (209 Å contre 24 Å à basse température).

Enfin, nous avons résolu la structure de la conformation intermédiaire de *EcHU* $\alpha\beta$ à 37 °C. Cette nouvelle structure présente un dimère partiellement dissocié (éloignement des domaines HTH) et déstructuré (perte d'hélicité et diminution de la longueur des brins β). Les hélices de chaque domaine HTH se réorientent afin de quasiment s'aligner (angle de 160°). Les deux feuillets β formant la « selle » de la protéine ne sont plus définis comme tel sur la structure intermédiaire bien que celle-ci soit encore reconnaissable. Le dimère présente à sa surface un patch hydrophobe présentant des résidus qui étaient enfouis dans le cœur hydrophobe du dimère natif. Ce patch pourrait être le motif de reconnaissance des conformations intermédiaire entre elles entraînant la formation du tétramère transitoire.

Chapitre 4. Etude de la cinétique globale de la réaction

Ce chapitre est consacré à l'étude du mécanisme d'échange de chaînes de la protéine HU d'*E. coli* d'un point de vue cinétique. Ce mécanisme d'échange peut être décrit comme une réaction se déroulant en trois étapes successives : la transition de la conformation native de chaque homodimère à la conformation intermédiaire homodimérique, puis l'association de deux formes intermédiaires en un tétramère transitoire qui se dissocie finalement en deux hétérodimères. A partir de l'ensemble des travaux de recherches réalisés sur la protéine HU d'*E. coli*, il a été montré que la cinétique du système global pouvait être résumé à l'équation-bilan présentée sur la Figure 47.

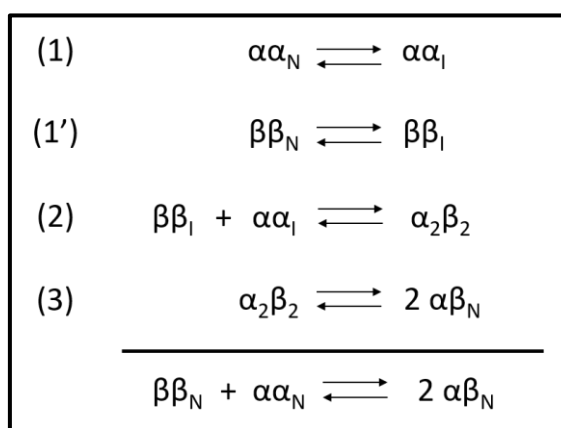


Figure 47 : **Equation-bilan du mécanisme global de formation de l'hétérodimère.**

Toutefois, le mécanisme complet d'échange de chaîne implique de nombreuses formes oligomériques de la protéine telles que des homotétramères (α_4 et β_4). Pour rappel, le mécanisme complet a pu être schématisé par grand triangle (Figure 48). Chaque côté de ce triangle correspond à un dimère en équilibre entre sa conformation native et intermédiaire. Chacun des dimères peut former des tétramères en fonction du dimère avec lequel il s'associe. **Le but de cette étude est de déterminer les constantes de vitesse des trois étapes principales.**

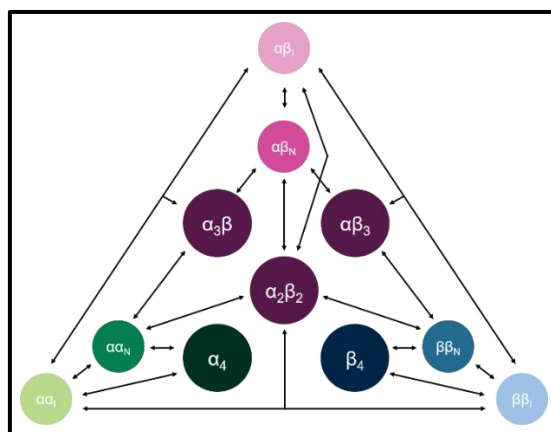


Figure 48 : **Schéma du mécanisme complet d'échange de chaînes de la protéine HU d'*E. coli*.**
 α = chaîne EchU α ; β = chaîne EchU β ; N = conformation native ; I = conformation intermédiaire.

4.1. Titrage de l'échange de chaînes

Cette première phase de caractérisation de la formation de *EchU* $\alpha\beta$ est réalisée avec des échantillons à l'équilibre, c'est-à-dire pour lesquels aucun échange n'a lieu à l'échelle macroscopique (Figure 49).

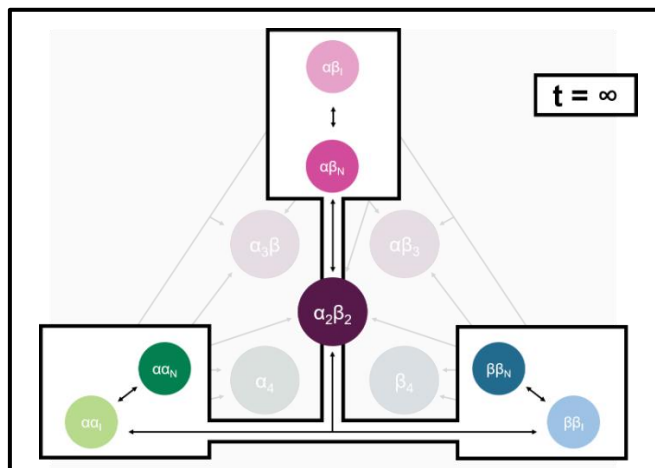


Figure 49 : Etude cinétique du système de *EchU* par titrage à l'équilibre.

Cet équilibre est observé à un temps considéré comme infini après l'ajout des homodimères. En effet, il est connu que si l'on met en présence les deux homodimères, l'hétérodimère se forme spontanément avec une cinétique lente. Mais cette réaction peut-elle être considérée comme totale ?

Afin de répondre à cette question, nous avons réalisé un premier titrage, à 20 °C, de la formation de l'hétérodimère à partir d'un échantillon d'homodimère *EchU* β_2 (ou $\beta\beta$) doublement marqué (titrage 1). Cet échantillon a été titré par des ajouts successifs d'une quantité connue d'homodimère *EchU* α_2 ($\alpha\alpha$) simplement marqué jusqu'à un ratio de 1 pour 1 des deux homodimères (Tableau 6).

Nous avons réalisé cette première étude avec des échantillons marqués car cela nous permettait de :

- suivre la formation de l'hétérodimère en suivant les déplacements chimiques des résidus de chaque chaîne sur un spectre ^{15}N -HSQC ;
- caractériser l'apparition éventuelle de signaux correspondant aux homodimères et en particulier de l'homodimère titrant.

Nous avons toutefois dû prendre des précautions pour cette analyse. Bien que l'utilisation d'un échantillon marqué soit un avantage pour notre étude, elle entraîne des complications dans l'analyse des spectres. En effet, ils contiennent les signaux cumulés de l'homodimère titrant, de l'homodimère titré, de l'hétérodimère formé (les deux chaînes) dans leur forme native et potentiellement

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

intermédiaire, en fonction de la température d'étude. Ceci provoque un grand nombre de superpositions et ne permet donc pas d'étudier de façon quantitative tous les résidus.

Pour ce titrage, nous avons suivi des résidus qui présentaient, dans l'homodimère titrant, l'homodimère titré et l'hétérodimère formé, des déplacements chimiques suffisamment différents et des signaux isolés (Figure 50). Les résidus sélectionnés pour l'analyse appartiennent à plusieurs zones des chaînes protéiques :

- les résidus Asp8 des chaînes *EchU α* et *EchU β* , situés au niveau de l'hélice α ;
- le résidu Glu15 de la chaîne *EchU α* et le résidu Ile16 de la chaîne *EchU β* ; situés au niveau du tour du motif « HTH » ;
- le résidu Ser27 de la chaîne *EchU α* et les résidus Gly22 et Ser35 de la chaîne *EchU β* , situés dans l'hélice α_2 ;
- les résidus Val42 (brin β_1) et Lys51 (brin β_2), Ala78 et Val80 (brin β_3) de la chaîne *EchU α* et les résidus Val76, Ser78 et Arg80 (brin β_3) sur la chaîne *EchU β* , situés au niveau de la « selle » ;
- les résidus Val89 et Lys90 de la chaîne *EchU α* et les résidus Leu85, Val89 et Asn90 sur de la chaîne *EchU β* , situés dans l'hélice C-terminale α_3 .

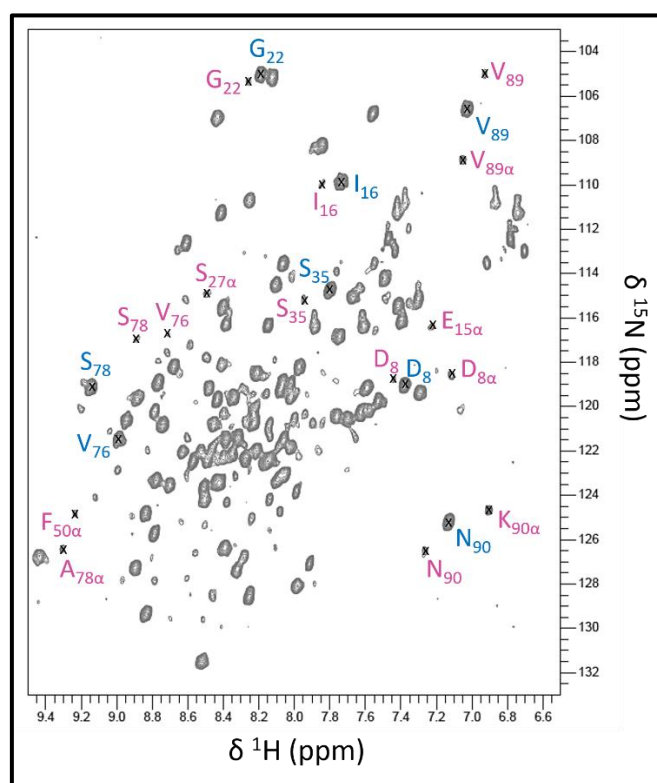


Figure 50 : ^{15}N -HSQC de *EchU β_2 :EchU α_2* au ratio 1:0,1 du titrage. Les résidus D8 α , E15 α , S27 α , F50 α , A78 α , V89 α et K90 α représentés en rose montre l'apparition des signaux correspondant à la chaîne *EchU α* dans l'hétérodimère. Les résidus D8, I16, G22, S35, V76, S78, V89 et N90 représentés en bleu et en rose représentent respectivement la disparition des signaux correspondant à la chaîne *EchU β* dans l'homodimère $\beta\beta$ et l'apparition des signaux à la chaîne *EchU β* dans l'hétérodimère.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

Nous avons suivi l'apparition de l'hétérodimère ainsi que la disparition de l'homodimère titré en utilisant l'équation (11) donnée dans le Chapitre 2. La [Figure 51A](#) montre le suivi de la formation de l'hétérodimère mesurée avec l'augmentation des intensités des pics de sa chaîne α au cours du titrage. Elles ont été doublement normalisées afin de tenir compte du facteur dilution lié aux ajouts et d'obtenir les valeurs de fraction molaire de l'hétérodimère ($f_{hetero} = \frac{I_i}{I_{finale}}$). Les données expérimentales ne peuvent pas être analysées en utilisant une droite. Il semble donc que l'utilisation des signaux de la chaîne α n'est pas la bonne méthode d'analyse du titrage. Toutefois, il apparaît clairement que les fractions molaires dérivées des intensités mesurées lors du troisième ajout semblent avoir des valeurs aberrantes. La source de cette aberration n'est pas due à une erreur de manipulation de l'échantillon lors du titrage, car les ajouts suivants ne donnent pas de résultats aberrants, mais doit certainement provenir d'un problème de paramétrage des expériences ^{15}N -HSQC.

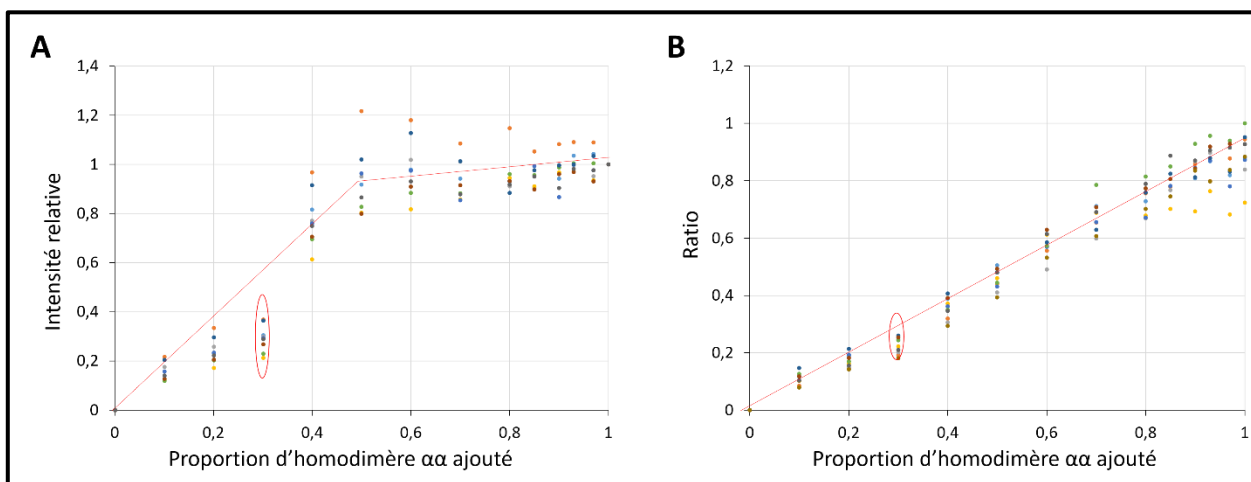


Figure 51 : Suivi de la formation de l'hétérodimère lors du titrage de $\text{EchU}\beta_2$ par $\text{EchU}\alpha_2$ selon deux méthodes. A – Tracé de l'apparition des résidus D8, E15, S27, V42, K51, A78, V80, V89 et K90 de la chaîne $\text{EchU}\alpha$ de l'hétérodimère avec $f_{hetero} = \frac{I_i}{I_{finale}}$. B – Tracé de l'apparition des résidus D8, I16, G22, S35, V76, S78, R80, L85, V89 et N90 de la chaîne $\text{EchU}\beta$ de l'hétérodimère avec $R(\alpha\beta) = \frac{I(\alpha\beta)}{I(\alpha\beta)+I(\beta\beta)}$. Les points aberrants sont entourés en rouge.

La [Figure 51B](#) montre le suivi de la formation de l'hétérodimère mesuré en suivant l'augmentation des intensités des pics de sa chaîne β au cours du titrage. La fraction molaire d'hétérodimère est ici déterminée par le rapport $R(\alpha\beta) = \frac{I(\alpha\beta)}{I(\alpha\beta)+I(\beta\beta)}$. Avec cette méthode, les données expérimentales de l'ensemble du titrage, excepté celles obtenues lors du troisième ajout, peuvent être modélisées par une droite passant par l'ordonnée à l'origine, d'équation $R(\alpha\beta) = 0,92 R(\alpha\alpha)$ avec un R^2 de 0,9937. La deuxième méthode permettant une meilleure quantification de l'apparition d' $\text{EchU}\alpha\beta$ (ou $\alpha\beta$) a été choisie pour poursuivre notre étude.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

La Figure 52 présente une analyse plus rigoureuse du suivi de la formation de l'hétérodimère ($R(\alpha\beta) = \frac{I(\alpha\beta)}{I(\alpha\beta)+I(\beta\beta)}$) et de la disparition du dimère titré ($R(\beta\beta) = \frac{I(\beta\beta)}{I(\alpha\beta)+I(\beta\beta)}$) lors du titrage de *ECHU* β_2 par *ECHU* α_2 . L'incertitude de chaque point de données expérimentales a été déterminée en propageant les incertitudes sur les concentrations, sur les volumes ajoutés et sur les intensités des signaux des différents résidus par la méthode du calcul de la différentielle. Ces incertitudes en abscisses et en ordonnées nous ont permis de déterminer les droites de pentes minimum et maximum afin d'obtenir l'incertitude sur la valeur de la pente représentant nos données. Nous trouvons que nos données peuvent être représentées par deux droites de pente $0,92 \pm 0,06$ (formation de l'hétérodimère) et $-0,92 \pm 0,06$ (disparition du dimère titré). Ces deux droites se coupent au point $(0,54 \pm 0,04 ; 0,50 \pm 0,04)$. Nous pouvons en conclure que, dans nos conditions, **la réaction de formation de l'hétérodimère est une réaction pouvant être considérée comme totale.**

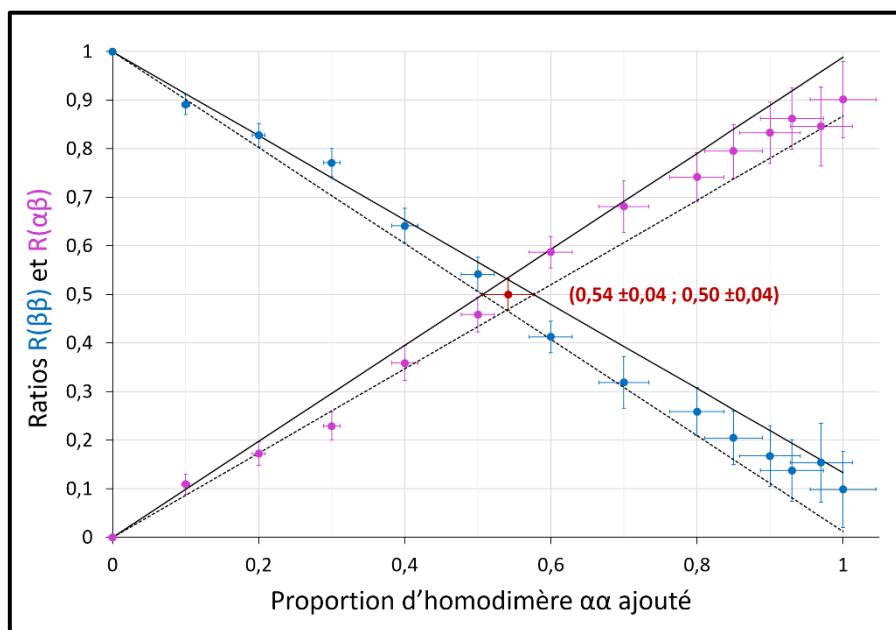


Figure 52 : **Suivi de la formation de l'hétérodimère et de la disparition de l'homodimère titré lors du titrage de *ECHU* β_2 par *ECHU* α_2 .**

Suite à notre première expérience de titrage, deux autres titrages ont été réalisés afin de confirmer les résultats précédents et ce à plusieurs températures. Les titrages 2 et 3 ont été réalisés, avec le même lot d'échantillons mais, à de plus hautes températures (27 °C et 37 °C). Travailler à ces températures nous a, non seulement, permis d'atteindre plus rapidement l'équilibre pour chaque point de titrage, mais aussi de regarder l'influence de la température sur la réaction (le cas échéant). Cela est particulièrement intéressant dans le sens où nous cherchons à caractériser l'ensemble du mécanisme d'échange et que nous savons qu'il est **fortement dépendant de la température.**

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

Sachant que les signaux permettant de suivre la réaction de façon quantitative ne sont que ceux de l'hétérodimère et ceux du dimère titrant, nous avons choisi de titrer un échantillon d'homodimère *EchU* β_2 marqué par un échantillon d'homodimère *EchU* α_2 non marqué. Tout comme pour le titrage précédent, les résidus suivis doivent présenter des déplacements chimiques suffisamment différents et des signaux isolés. Il faut, de plus, que l'intensité de ces résidus ne soit pas trop influencée par les changements de température afin de pouvoir réaliser les mesures avec la même précision à 27 °C et 37 °C, qu'à 10 °C.

Nous nous sommes focalisés sur le résidu **Ile16**, car il est positionné au milieu du tour du motif « HTH », ne présente pas de structuration secondaire et est totalement exposé au solvant. A l'issue des **deux nouveaux titrages**, nous avons obtenu des tracés typiques du titrage d'une réaction pouvant être qualifiée de **quasi-totale ou totale** : une droite de pente croissante passant par l'origine suivie d'un plateau. La droite représente l'apparition de l'hétérodimère alors que le plateau marque que la totalité de l'hétérodimère a été formée. Au niveau de ce plateau, la totalité du réactif titré est consommée.

Les droites obtenues pour les titrages réalisés à 27 °C et à 37 °C sont d'équation respectives $R(\alpha\beta) = (0,97 \pm 0,04) R(\alpha\alpha)$ et $R(\alpha\beta) = (0,84 \pm 0,08) R(\alpha\alpha)$.

Ces expériences permettent de confirmer que, dans nos conditions expérimentales, **la réaction de formation de l'hétérodimère est une réaction pouvant être considérée comme totale**.

4.2. Etude cinétique de l'échange de chaînes

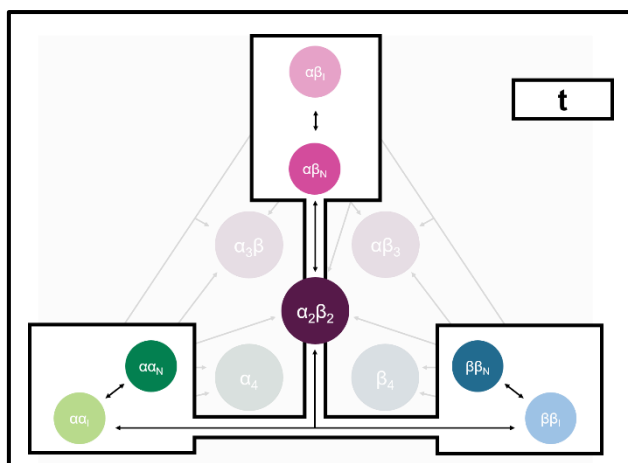


Figure 53 : Etude cinétique du système de EchU par suivi cinétique de la réaction globale.

Une fois la réaction caractérisée, nous avons cherché, à déterminer la vitesse de formation de l'hétérodimère (Figure 53). Pour la déterminer, nous avons utilisé de nombreux suivis cinétiques :

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

- trois ont été réalisés par spectrométrie de masse en conditions natives et en **conditions stœchiométriques**, chacun à une température différente (20 °C, 27 °C et 37 °C) ^[69] ;
- cinq ont été réalisés par RMN avec un **réactif en excès** et à deux températures différentes (10 °C et 20 °C).

Pour réaliser les cinétiques par RMN, l'échantillon contenant le réactif en excès est placé dans un bain de glace. Au moment de l'ajout du réactif limitant, un chronomètre est lancé et l'échantillon est rapidement placé dans le spectromètre RMN pré-paramétré sur l'échantillon de réactif en excès. Le chronomètre n'est interrompu qu'une fois la première expérience ¹⁵N-HSQC ou ¹⁵N-HMQC SOFAST en cours d'acquisition. Pendant ce temps de latence, la température de l'échantillon n'est ni constante ni mesurable.

L'utilisation d'échantillons en conditions stœchiométriques permet généralement de déterminer l'ordre global de la réaction, alors que l'utilisation d'échantillons avec un réactif en excès permet de déterminer l'ordre partiel du réactif minoritaire. Pour déterminer l'ordre de notre réaction, il nous a fallu établir les équations cinétiques pouvant régir celle-ci en considérant que la réaction est **bimoléculaire et totale**. Dans le cas, d'une réaction bimoléculaire totale, la loi de vitesse s'écrit : $v(t) = k \times [A]_t^p \times [B]_t^q$ dans laquelle p et q correspondent aux ordres partiels des réactifs.

Dans un premier temps, nous avons écrit l'équation cinétique de la formation de *ECHUαβ* en considérant que les **ordres partiels des réactifs étaient de 1** et que les **conditions stœchiométriques** étaient respectées (Tableau 15).

	αα	+	ββ	→	2 αβ
t₀	a ₀		b ₀ = a ₀		0
t	a ₀ - x(t)		a ₀ - x(t)		2 x(t)

Tableau 15 : Tableau d'avancement de la formation d'hétérodimère en conditions stœchiométriques.

Nous pouvons alors simplifier la loi de vitesse : $v(t) = k \times [A]_t^2$.

Si l'on cherche à résoudre l'équation de la vitesse avec cette loi de vitesse, nous obtenons :

$$v(t) = \frac{-d[A]_t}{dt} = k \times [A]_t^2$$

$$\frac{d[A]_t}{[A]_t^2} = -k dt$$

$$\int_0^{[A]_t} \frac{d[A]_t}{[A]_t^2} = -k \int_0^t dt$$

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

$$\frac{1}{[A]_t} = \frac{1}{[A]_0} + kt$$

Si les données expérimentales peuvent être modélisées en utilisant cette équation, alors on peut conclure que la cinétique de la réaction est d'ordre 2.

Dans un second temps, nous avons cherché à établir l'équation cinétique de la formation de *EchUαβ* en considérant que les **ordres partiels des réactifs étaient de 1** et que l'un des réactifs est en **excès** (Tableau 16). La loi des vitesses s'écrit alors :

$$v(t) = \frac{-d[A]_t}{dt} = k \times [A]_t [B]_t$$

	αα	+	ββ	→	2 αβ
t_0	a_0		$b_0 (> a_0)$		c_0
T	$a_0 - x(t)$		$b_0 - x(t)$		$c_0 + 2 x(t)$

Tableau 16 : **Tableau d'avancement de la formation d'hétérodimère en conditions non stœchiométriques.**

Elle peut être intégrée pour trouver $\ln \frac{[A]^0 \times [B]_t}{[B]^0 \times [A]_t} = ([B]^0 - [A]^0) \times kt$.

Si l'excès est très important ($[B]_0 \gg [A]_0$, au moins un ordre de grandeur), alors la concentration $[B]$ varie peu et la vitesse devient quasi proportionnelle à la concentration de $[A]$ seul. L'évolution du réactif A se rapproche alors d'une cinétique monomoléculaire. La loi des vitesses s'écrit alors :

$$v(t) = \frac{-d[A]_t}{dt} = k_{obs} \times [A]_t \text{ avec } k_{obs} = k[B]_t$$

$$\frac{d[A]_t}{[A]_t} = -kdt$$

$$[A]_t = [A]_0 e^{-kt}$$

Cette situation est souvent exploitée expérimentalement dans le but de simplifier la cinétique. Il s'agit de la méthode de dégénérescence de l'ordre et le système y correspond à une cinétique de **pseudo-ordre 1**.

La Figure 54A présente le suivi cinétique de la formation de l'hétérodimère de *EchU* à 20 °C obtenu par spectrométrie de masse en conditions natives. Les données peuvent être modélisées par une équation correspondant à une **cinétique d'ordre 2**. La constante de vitesse mesurée est de **$k = 1,4 \cdot 10^{-2} \text{ mM}^{-1} \cdot \text{s}^{-1}$** . Nous avons alors réalisé le suivi de la formation de l'hétérodimère en présence d'un excès de *EchUβ₂* (450 μM) par rapport à *EchUα₂* (45 μM) à la même température afin de nous placer dans une situation de pseudo-ordre 1 (Figure 54B). Malheureusement, les données ne

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EcHU*

peuvent pas être modélisées par l'équation permettant d'analyser une cinétique d'ordre 1 ou de pseudo-ordre 1, mais par celle d'une cinétique d'ordre 2. La constante de vitesse mesurée est de $1,9 \cdot 10^{-3} \text{ mM}^{-1} \cdot \text{s}^{-1}$. Ces deux vitesses sont différentes de plus d'un ordre de grandeur, ce qui n'est absolument pas négligeable. De plus, l'ajustement informatique obtenu sur les données de spectrométrie de masse est de meilleure qualité, ce qui nous laisse penser que c'est la valeur de la constante de vitesse déterminée par cette méthode qui doit être prise en compte pour continuer l'étude cinétique du mécanisme d'échange de chaîne.

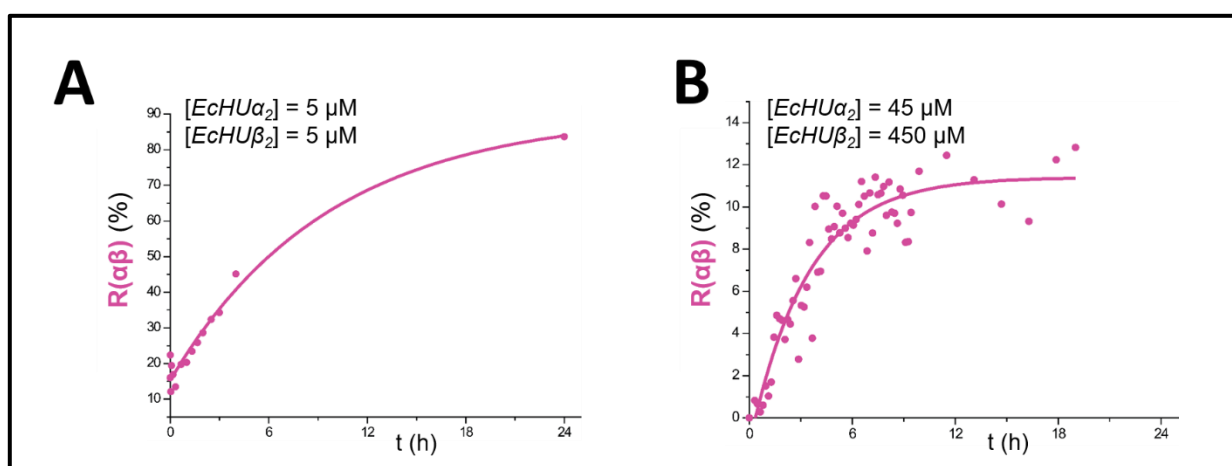


Figure 54 : **Suivi cinétique de la formation de l'hétérodimère de *EcHU* à 20 °C.** A – en conditions stœchiométriques par spectrométrie de masse en conditions natives. B – dans le cas d'excès de l'un des réactifs par RMN en solution.

Nous avons toutefois complété les données expérimentales en analysant deux autres cinétiques dont le suivi avait été réalisé par spectrométrie de masse à 27°C et à 37 °C, mais aussi avec des suivis de cinétiques réalisés par RMN. Deux de ces suivis ont été réalisés à 10 °C avec des concentrations en réactifs différentes (variation de la concentration de l'homodimère *EcHUβ₂*) : pour le premier suivi (cinétique 4), nous étions partis d'un mélange de 200 μL contenant 87,52 nmol de *EcHUα₂* $^{15}\text{N}^{13}\text{C}$ et 160,2 nmol de *EcHUβ₂* et pour le second (cinétique 5), d'un mélange de 200 μL contenant de nouveau 87,52 nmol de *EcHUα₂* $^{15}\text{N}^{13}\text{C}$ et 200,3 nmol de *EcHUβ₂*. Deux autres suivis cinétiques ont été réalisés afin de quantifier la variation de vitesse de réaction en fonction de la température (cinétique 2 et 3).

Le Tableau 17 récapitule les différentes constantes de vitesse obtenues par RMN et par spectrométrie de masse en conditions natives. Les résultats obtenus par spectrométrie de masse montrent clairement une augmentation, tout à fait logique, de la constante de vitesse d'échange avec la température (un ordre de grandeur pour environ 10 °C).

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

Suivi cinétique par MS	Température	$[\alpha\alpha]_0$ (μM)	$[\beta\beta]_0$ (μM)	k_t ($\text{mM}^{-1}.\text{s}^{-1}$)	Δk_t ($\text{mM}^{-1}.\text{s}^{-1}$)
1	20 °C	0,045	0,45	$1,4 \cdot 10^{-2}$	$1,4 \cdot 10^{-3}$
2	27 °C	0,60	1,20	$1,4 \cdot 10^{-1}$	$1,4 \cdot 10^{-2}$
3	37 °C	0,60	1,07	3,4	$3,4 \cdot 10^{-1}$

Suivi cinétique par RMN	Température	$[\alpha\alpha]_0$ (mM)	$[\beta\beta]_0$ (mM)	k_t ($\text{mM}^{-1}.\text{s}^{-1}$)	Δk_t ($\text{mM}^{-1}.\text{s}^{-1}$)
1	20 °C	0,045	0,45	$1,9 \cdot 10^{-3}$	$1,9 \cdot 10^{-4}$
2	10 °C	0,60	1,20	$1,6 \cdot 10^{-4}$	$1,6 \cdot 10^{-5}$
3	20 °C	0,60	1,07	$5,1 \cdot 10^{-4}$	$5,1 \cdot 10^{-5}$
4	10 °C	0,44	0,80	$1,3 \cdot 10^{-4}$	$1,3 \cdot 10^{-5}$
5	10 °C	0,44	1,00	$5,7 \cdot 10^{-5}$	$5,7 \cdot 10^{-6}$

Tableau 17 : Récapitulatif des différents suivis cinétiques et des valeurs de constante de vitesse d'échange obtenues.

Les résultats obtenus par RMN semblent plus contrastés. En effet, les valeurs de constantes d'échange obtenues à 20 °C sont très différentes l'une de l'autre (presque un ordre de grandeur). A 10 °C, deux vitesses sur les trois mesurées ont des valeurs compatibles entre elles.

Nous avons alors voulu vérifier que la loi d'Arrhenius, $k_{obs}(T) = A \times e^{-\frac{E_a}{RT}}$, pouvait être appliquée à nos mesures expérimentales. La [Figure 55](#) montre que la majorité de nos valeurs expérimentales vérifient cette loi d'Arrhenius en formant une droite d'équation : $\ln(k) = \frac{-34\,161}{T} + 111\,510$ où T correspond à la température en Kelvin.

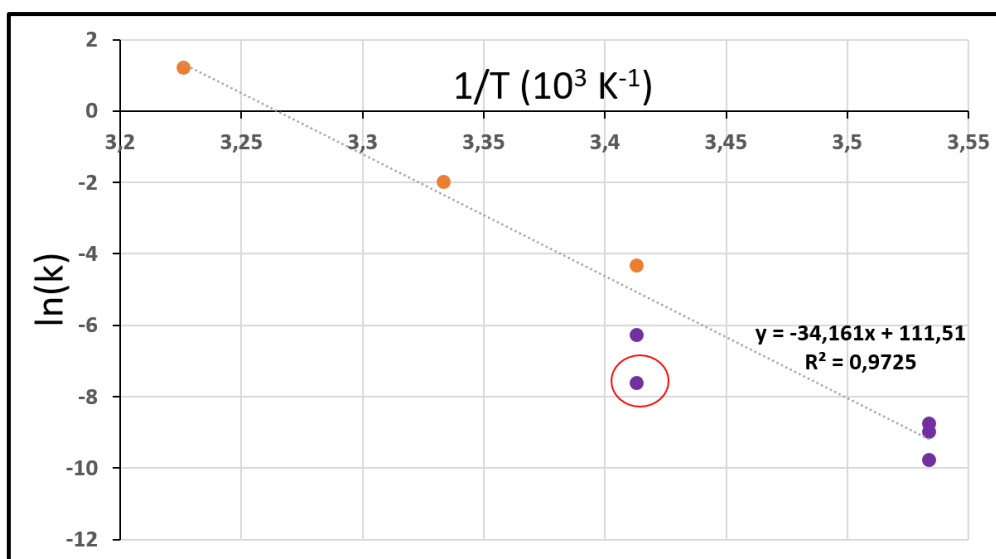


Figure 55 : Vérification de la loi d'Arrhenius à partir des données cinétiques obtenues par RMN (en violet) et spectrométrie de masse en conditions natives (en orange).

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

Cette droite nous permet de conclure que les constantes de vitesse mesurées par RMN et par spectrométrie de masse sont tout à fait compatibles entre elles, excepté pour celles obtenues à 20 °C. Il faudrait répéter les cinétiques à cette température. De plus, on pourrait enregistrer des cinétiques à d'autres températures afin de confirmer que nous avons bien caractérisé la constante de vitesse de la réaction globale à toutes les températures au travers de la droite d'équation : $\ln(k) = \frac{-34161}{T} + 111510$, caractérisée par cette étude.

4.3. Caractérisation cinétique des étapes intermédiaires

La troisième phase de la caractérisation de notre système cinétique est, après la caractérisation de la cinétique globale, la caractérisation des cinétiques des étapes intermédiaires (Figure 56). Cette caractérisation est nécessaire pour décrire totalement le système et en déduire l'étape limitante.

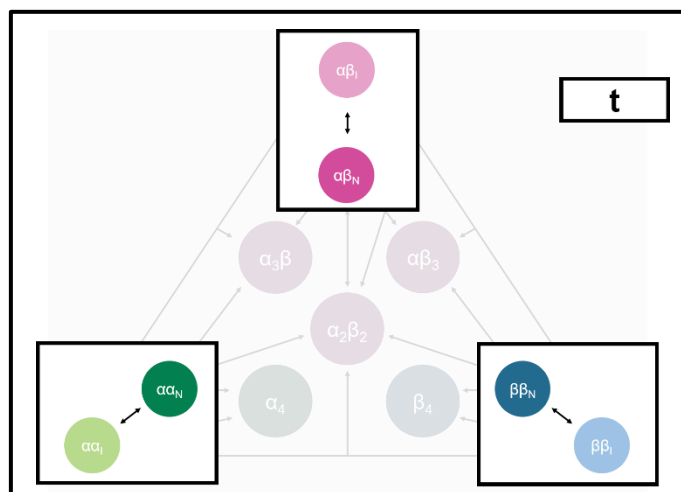
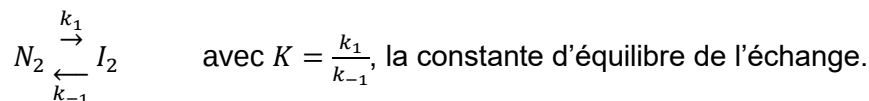


Figure 56 : Etude cinétique du système de *ECHU* par la caractérisation des réactions intermédiaires, en particulier l'échange N_2/I_2 .

La première étape intermédiaire à caractériser est l'échange entre les conformations N_2 et I_2 des différents dimères. Cet échange est caractérisé d'un point de vue cinétique par :



Si on considère que $[N_2]_0 = a$ et $[I_2]_0 = 0$ et qu'à l'équilibre, $[N_2]_{EQ} = a - x_{EQ}$ et $[I_2]_{EQ} = x_{EQ}$, la loi des vitesses $v_r = -\frac{d[N_2]_t}{dt} = k_1[N_2]_t - k_{-1}[I_2]_t$ permet d'établir que :

$$[I_2]_t = -\frac{k_1 a}{k_{-1} + k_1} (e^{-(k_1 + k_{-1})t} - 1)$$

$$\text{Et que, } x_{EQ} = \frac{k_1 a}{k_{-1} + k_1}$$

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

L'échange entre les deux formes dimériques est lent à l'échelle de la RMN et a été précédemment caractérisé à partir d'une série d'expériences d'échange ZZ sur l'homodimère *EchU α_2* à 37 °C et sur l'homodimère *EchU β_2* à 27 °C [69]. Lors de cette première étude, des valeurs de constante de vitesse ont pu être déterminées pour l'homodimère *EchU α_2* à 37 °C, $k_1 = 2,4 \text{ s}^{-1}$ et $k_{-1} = 7,7 \text{ s}^{-1}$. Il en a été déduit que la constante d'échange K, pour ce dimère à 37 °C, était de $k_1/k_{-1} = 0,31$. En revanche, seul un ordre de grandeur, très supérieur à la seconde, avait pu être déterminé pour l'homodimère *EchU β_2* à 27 °C du fait de l'intensité trop faible des signaux croisés pour permettre une évaluation quantitative.

Pour mon étude et afin de caractériser le système d'échange complet de la protéine, je me suis appuyée sur ces données et les ai complétées avec des données supplémentaires sur les trois dimères à différentes températures. L'ensemble des données ont été traités avec les équations présentées dans le chapitre « Matériel et méthodes » de cette partie (Partie 1.2.3.3).

Pour compléter l'étude sur l'homodimère *EchU α_2* , nous avons enregistré les expériences d'échange à 20, 27, 38 et 41 °C. Pour ce dimère nous avons pu quantifier l'échange N_2/I_2 pour la majorité des températures. Les résultats obtenus sont résumés dans le Tableau 18.

T°	27 °C	38 °C	41 °C
k_1 (ou $k_{N \rightarrow I}$) en s^{-1}	0,48 s^{-1}	2,23 s^{-1}	3,66 s^{-1}
k_{-1} (ou $k_{I \rightarrow N}$) en s^{-1}	1,28 s^{-1}	2,19 s^{-1}	0,50 s^{-1}
$k_{\text{ex}} = k_1 + k_{-1}$ en s^{-1}	1,76 s^{-1}	4,42 s^{-1}	4,16 s^{-1}
K	0,38	0,98	7,32

Tableau 18 : **Constantes de vitesse d'échange N_2/I_2 de l'homodimère *EchU α_2* pour différentes températures.**

Dans l'ensemble, les résultats sont cohérents en présentant une augmentation de la constante de vitesse k_1 et de la constante d'échange K avec la température. Il faudra répéter les expériences afin de s'assurer d'un meilleur rapport signal sur bruit (en augmentant la durée des acquisitions et, peut-être, la concentration des échantillons) mais aussi de la répétabilité des mesures de constantes k_1 et k_{-1} . Il est à noter que nos résidus présentent une **dynamique interne (R_1) très différente dans leurs deux conformations N_2/I_2** . Cette différence a bien été prise en compte dans le calcul, mais la mesure des R_1 sur les pics de la conformation intermédiaire ne peut pas, si les expériences de relaxation R_1 ne sont pas répétées, être considérée comme précise. Une mauvaise mesure des R_1 de la conformation intermédiaire peut entraîner une erreur de calcul sur la constante de vitesse k_{-1} .

Nous pouvons tout de même conclure que l'ordre de la vitesse d'échange N_2/I_2 mesurée pour l'homodimère *EchU α_2* est vingt à trente fois plus rapide que la vitesse de réaction du mécanisme d'échange de chaînes. On peut donc en conclure que, la transition N_2/I_2 de *EchU α_2* n'en est pas l'étape limitante.

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *EchU*

Nous avons réalisé les expériences d'échange ZZ sur l'homodimère ***EchUβ₂*** à 15, 20 et à 27 °C. Nous voulions ralentir l'échange afin de réussir à le quantifier à plus basse température. A 15 °C et à 20 °C, bien que l'échange ait lieu ([Partie 1.3.1](#)), les signaux croisés ne sont pas assez intenses pour être observés à 15 °C et sont trop faibles pour être quantifiés à 20 °C. A 27 °C, nous avons réussi à déterminer la constante de vitesse k_1 . Connaissant les populations N_2/I_2 à 27 °C, nous avons déterminé que la concentration d'équilibre de la forme intermédiaire est de 0,43 mM et en avons déduit la valeur de k_{-1} : **k_1 51,8 s⁻¹ et $k_{-1} \sim 8,4$ s⁻¹**, ce qui correspond à une constante d'équilibre **$K \approx 6$** .

On peut donc en conclure que, quelle que soit la température, la vitesse d'échange **N_2/I_2 de *EchUβ₂*** est plus rapide que celle de *EchUα₂*. Cet échange **ne peut donc en aucun cas être l'étape limitante du mécanisme d'échange de chaînes de *EchU***.

La plupart des expériences d'échange ZZ ont été enregistrées après les expériences RD ¹⁵N-CPMG. Ces dernières étaient utilisées pour la discrimination des signaux lors de l'attribution de la protéine et pour la recherche de forme intermédiaire de la protéine qui pourraient être présentes dans l'échantillon. Il s'est avéré que le train CPMG de la séquence détériorait nos échantillons et explique que nous n'ayons pas pu enregistrer plus d'expériences d'échange ZZ. Il nous faudra absolument compléter l'étude par des mesures sur l'hétérodimère. Toutefois, les propriétés thermodynamiques de celui-ci sont très proches de celles de l'homodimères *EchUα₂*. **Nous pouvons donc supposer que la transition N_2/I_2 de l'hétérodimère est caractérisée par une constante de vitesse d'échange très proche de celle de *EchUα₂***.

L'étape limitante doit donc être une étape impliquant le tétramère transitoire que ce soit son association ou sa dissociation. S'il s'agissait de sa dissociation, il y aurait une accumulation du tétramère dans la solution et celui-ci serait observable. **L'étape limitante ne peut donc être que l'association de deux I_2 pour former le tétramère transitoire et la constante de vitesse de cette étape doit être du même ordre de grandeur que la vitesse de réaction de la réaction globale.**

4.4. Conclusion : la formation du tétramère transitoire comme étape limitante de la réaction

L'ensemble de cette étude, bien que nécessitant encore des expériences complémentaires pour confirmer et compléter nos résultats, nous a permis de caractériser la cinétique globale de formation de l'hétérodimère. Ce mécanisme d'échange, décrit comme une réaction se déroulant en trois étapes successives, correspond à une réaction totale d'ordre 2 dont l'étape limitante est la formation du tétramère transitoire.

Actuellement, seule la spectrométrie de masse en conditions natives a permis d'observer des formes tétramériques. Bien que la caractérisation cinétique de cette forme transitoire ne soit pas possible

Partie 1. Vers une description complète du mécanisme d'échange de chaînes de la protéine *ECHU*

pour le moment, nous pouvons d'ores et déjà proposer quelques pistes pour les déterminer expérimentalement. Il faudrait réussir :

- à bloquer la cinétique à l'étape de formation du tétramère. Pour ce faire, il faudrait stabiliser ou bloquer la protéine dans sa conformation tétramérique par mutagenèse par exemple. Cela permettrait de confirmer que la vitesse globale de la réaction correspond bien à la vitesse de formation du tétramère. Cette confirmation se ferait par quantification de l'accumulation de la forme tétramérique.

- à libérer le tétramère pour quantifier la disparition de cette forme.

- dans un troisième temps, à réaliser ce jeu de blocage/libération pour toutes les formes tétramériques impliquées dans le mécanisme global.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Chapitre 1. Caractérisation d'une courbure en V de l'ADN induite par MC1

1.1. MC1 : une protéine uniquement présente chez les Archées

1.1.1. Les Archées en bref

On les connaît psychrophiles, thermophiles ou hyperthermophiles (vivant à des températures inférieures à 0 °C, comprises entre 50 °C et 70 °C ou supérieures à 80 °C) ^[112], anaérobies (vivant dans des environnements privés d'oxygène) ou aérobies (vivant dans des milieux où l'oxygène est présent), acidophiles (pH<4), alcalinophiles (pH>9), halophiles (vivant dans des environnements riches en sels) ^[113], méthanogènes (produisant du méthane), etc. Elles ont été très longtemps considérées comme des « bactéries ancestrales » et/ou des « bactéries de l'extrême », mais il a été montré que les **Archées** constituent **l'un des trois domaines du Vivant** ^[114,115] et sont, en réalité, **présentes partout** ^[116]. On les retrouve aussi bien dans des biotopes usuels (le sol, la mer, les intestins d'animaux, etc.) que dans des biotopes extrêmes (les geysers, les lacs salés, les sources chaudes volcaniques, etc.) comme l'illustre la Figure 57. Vivant dans des environnements très variés, elles se présentent sous de multiples formes avec des modes de vie et d'adaptation divers. De ce fait, elles sont des modèles originaux pour aider et inspirer les scientifiques dans leur recherche du **décryptage des mécanismes de la nature et le développement de nouvelles technologies** ^[112].

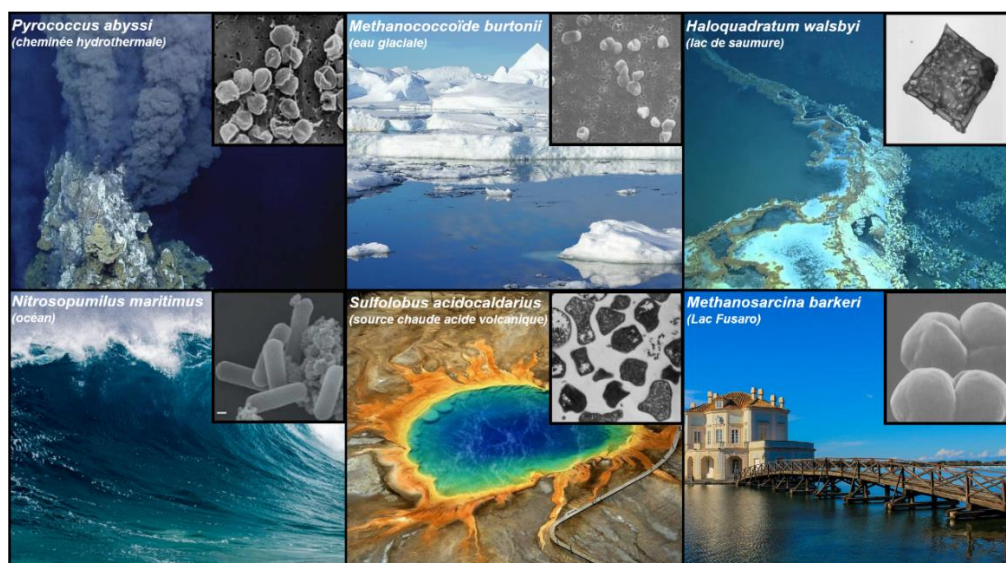


Figure 57 : Diversité des Archées et de leurs biotopes. Les espèces et leur environnement correspondant sont représentés de gauche à droite et de haut en bas : *Pyrococcus abyssi* (anaérobie) ; *Methanococcoïde burtonii* (psychrophile) ; *Haloquadratum walsbyi* (halophile) ; *Nitrosopumilus maritimus* (chimioolithoautotrophe) ; *Sulfolobus acidocaldarius* (acidophile) et *Methanosarcina barkeri* (méthanogène).

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Caractérisation des Archées

Les Archées ont été caractérisées pour la première fois en 1977 par Carl R. Woese et George E. Fox à partir d'une **analyse phylogénétique basée sur le séquençage** des Acides RiboNucléiques (ARN) ribosomiques, ARNr, de multiples organismes ^[117]. Il est présent chez tous les organismes vivants et a été choisi pour comparer la phylogénie des espèces (ARNr 16S pour les Bactéries et ARNr 18S pour les Eucaryotes) car il présente des régions fortement conservées permettant d'établir des relations évolutives entre des espèces très éloignées. Les séquences des ARNr de treize espèces différentes ont été analysées deux à deux et un coefficient d'association entre ces dernières, S_{AB} , a pu être calculé. Ce coefficient correspond à deux fois le nombre d'oligonucléotides communs entre les deux séquences testées divisé par le nombre total d'oligonucléotides. La valeur de ce coefficient est d'autant plus grande que les espèces sont proches. Pour résumer les résultats, une **matrice de comparaison des espèces**, présentée sur le Tableau 19 a été construite. Sans surprise, cette matrice montre que toutes les espèces eucaryotes testées appartiennent au même groupe avec un S_{AB} supérieur à 0,19 (espèces 1 à 3 sur le Tableau 19). En revanche, ce qui étaient inattendus à l'époque, c'est qu'elle montre clairement que les espèces procaryotes sont divisées en deux groupes : les bactéries (espèces 4 à 9 sur le Tableau 19) et les « bactéries » méthanogènes (espèces 10 à 13 sur le Tableau 19). Il a alors été proposé une nouvelle vision du Vivant, en trois domaines, où les « bactéries » méthanogènes n'appartenaient plus au domaine Bactéries mais à celui des **Archées**.

Table 1. Association coefficients (S_{AB}) between representative members of the three primary kingdoms													
	1	2	3	4	5	6	7	8	9	10	11	12	13
1. <i>Saccharomyces cerevisiae</i> , 18S	—	0.29	0.33	0.05	0.06	0.08	0.09	0.11	0.08	0.11	0.11	0.08	0.08
2. <i>Lemna minor</i> , 18S	0.29	—	0.36	0.10	0.05	0.06	0.10	0.09	0.11	0.10	0.10	0.13	0.07
3. L cell, 18S	0.33	0.36	—	0.06	0.06	0.07	0.07	0.09	0.06	0.10	0.10	0.09	0.07
4. <i>Escherichia coli</i>	0.05	0.10	0.06	—	0.24	0.25	0.28	0.26	0.21	0.11	0.12	0.07	0.12
5. <i>Chlorobium vibrioforme</i>	0.06	0.05	0.06	0.24	—	0.22	0.22	0.20	0.19	0.06	0.07	0.06	0.09
6. <i>Bacillus firmus</i>	0.08	0.06	0.07	0.25	0.22	—	0.34	0.26	0.20	0.11	0.13	0.06	0.12
7. <i>Corynebacterium diphtheriae</i>	0.09	0.10	0.07	0.28	0.22	0.34	—	0.23	0.21	0.12	0.12	0.09	0.10
8. <i>Aphanocapsa</i> 6714	0.11	0.09	0.09	0.26	0.20	0.26	0.23	—	0.31	0.11	0.11	0.10	0.10
9. Chloroplast (<i>Lemna</i>)	0.08	0.11	0.06	0.21	0.19	0.20	0.21	0.31	—	0.14	0.12	0.10	0.12
10. <i>Methanobacterium thermoautotrophicum</i>	0.11	0.10	0.10	0.11	0.06	0.11	0.12	0.11	0.14	—	0.51	0.25	0.30
11. <i>M. ruminantium</i> strain M-1	0.11	0.10	0.10	0.12	0.07	0.13	0.12	0.11	0.12	0.51	—	0.25	0.24
12. <i>Methanobacterium</i> sp., Cariaco-isolate JR-1	0.08	0.13	0.09	0.07	0.06	0.06	0.09	0.10	0.10	0.25	0.25	—	0.32
13. <i>Methanosarcina barkeri</i>	0.08	0.07	0.07	0.12	0.09	0.12	0.10	0.10	0.12	0.30	0.24	0.32	—

Tableau 19 : **Matrice des coefficients d'association (S_{AB}) entre différents organismes des trois domaines du Vivant** (adapté de Woese et Fox, 1977) ^[117]. Les organismes 1 à 3 sont des Eucaryotes, 4 à 9 des Bactéries et de 10 à 13 des Archées. Les S_{AB} entre organismes du même groupe sont encadrés en bleu et sont supérieurs à 0,19.

La mise en évidence tardive du domaine des Archées est principalement liée à la **forte ressemblance morphologique et physiologique entre les Archées et les Bactéries**. En effet, avant les travaux de Carl R. Woese et George E. Fox, seuls des critères phénotypiques tels que la morphologie ou la capacité à résister à une certaine gamme de température étaient exploités pour

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

distinguer les espèces. Les Archées, tout comme les Bactéries, présentent des formes variées, une taille de l'ordre du micromètre (0,1 μm à 10 μm) et sont dépourvues de noyau. En revanche, les Archées présentent des caractéristiques moléculaires, biochimiques et métaboliques originales, comme la méthanogenèse chez les Archées méthanogènes. Le progrès des techniques d'analyse a notamment permis de caractériser la membrane des Archées ^[118,119]. Elle est constituée d'un type de phospholipides qui diffèrent totalement de celui des Eucaryotes et des Bactéries. Ces phospholipides sont des éthers d'isoprénol et non des esters d'acides gras comme pour les espèces des deux autres domaines ^[120–122]. De plus, le protéome des Archées est bien plus proche de celui des Eucaryotes que de celui des bactéries. En effet, il n'est pas rare de constater que les Archées et les Eucaryotes utilisent la même famille de protéines pour réaliser les mêmes fonctions au sein de la cellule, alors que ces fonctions sont assurées par des protéines d'une autre famille chez les Bactéries. Par exemple, de nombreuses **protéines impliquées dans le traitement de l'information génétique**, telles que les **histones** ^[19], sont présentes chez les Eucaryotes et les Archées alors qu'elles sont absentes chez les Bactéries.

Phylogénie des Archées

La classification des espèces dans le domaine des Archées est régulièrement réajustée en fonction des découvertes et de la caractérisation de nouvelles souches d'archées.

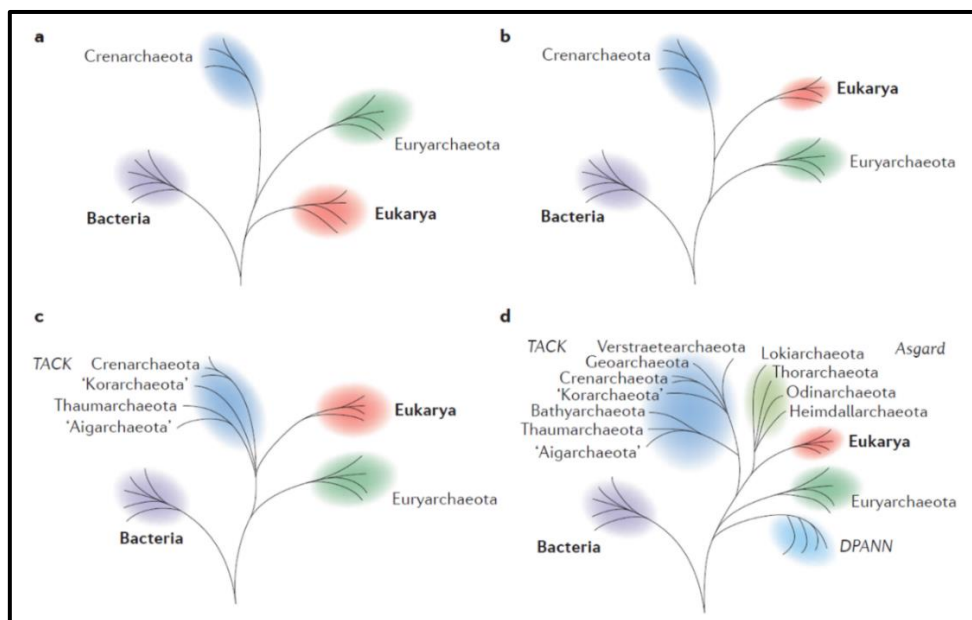


Figure 58 : Evolution de la classification des Archées dans l'Arbre phylogénétique du Vivant (issu de Eme et al., 2017) ^[121]. (a) dans les années 90, arbre à trois domaines avec deux phylums décrivant les archées ; (b) dans les années 90, arbre à deux domaines avec une proximité phylogénétique entre les Eucaryotes et les Crénarchées ; (c) au début des années 2000, proposition du superphylum TACK ; (d) début des années 2010, arbre à deux domaines avec 3 superphylums et 17 embranchements décrivant les Archées.

Description du domaine des Archées jusqu'au début des années 2010

Entre le début des années 1990 et 2010, le domaine des Archées est passé de deux embranchements majeurs (les Euryarchées et des Crénarchées, (Figure 58a) à **six embranchements** (les Euryarchées, les Nanoarchées, les Korarchées, les Thaumarchées, les Aigarchées et les Crénarchées (Figure 58c)).

Au début des années 1990, Carl R. Woese et ses collaborateurs ont caractérisé les **deux embranchements majeurs** qui correspondent à la plupart des archées cultivables en laboratoire : les **Euryarchées** et les **Crénarchées** (Figure 58a) ^[1]. L'embranchement des Euryarchées regroupe toutes les espèces méthanogènes, les halobactéries et certaines espèces hyperthermophiles comme *Thermococcus kodakarensis* ^[123]. Chez les Crénarchées on retrouve beaucoup d'espèces anaérobies obligatoires comme *Desulfurococcus kamchatkensis* ^[124] et d'espèces thermophiles.

En 1996, un troisième embranchement est proposé par Susan M. Barns et ses collaborateurs, celui des **Korarchées** grâce à la caractérisation de la séquence d'ARNr de deux nouvelles espèces : pJP27 et pJP78 ^[125]. Il s'agit d'Archées thermophiles qui, comme le proposent des études génétiques, se seraient différenciées assez tôt des Euryarchées et des Crénarchées du fait de la dispersion géographique et des différences de température et de salinité des milieux colonisés ^[126,127].

En 2002, Harald Huber et ses collaborateurs, ont découvert un « symbiote nanométrique hyperthermophile » : *Nanoarchaeum equitans* ^[128]. Avec un diamètre de 350 nm à 500 nm, l'espèce compte parmi les plus petites formes de vie connues. Ils ont alors proposé l'embranchement des **Nanoarchées** qui ne contient, à ce jour, qu'un membre unique. L'existence de cet embranchement a été sujet à débat, car il aurait pu s'agir d'une espèce très spécialisée issue de l'ordre des *Thermococcales*, appartenant à l'embranchement des Euryarchées ^[115,128].

En 2008, est apparu le phylum des **Thaumarchées** proposé par Céline Brochier-Armanet et ses collaborateurs ^[129]. Il contient des organismes chimiolithoautotrophes par oxydation de l'ammoniac (des organismes pouvant oxyder l'azote et capables de réduire le CO₂ atmosphérique), comme *Nitrososphaera viennensis* ^[130], qui pourraient jouer un rôle important dans les cycles biogéochimiques tels que le cycle de l'azote ou le cycle du carbone ^[131,132].

Enfin, en 2011 a été proposé l'embranchement des **Aigarchées** par Takuro Nunoura et ses collaborateurs suite à la découverte de *Candidatus Caldiarchaeum subterraneum*, un organisme supposément chimiolithoautotrophe par oxydation de sulfure ^[133]. Cet embranchement, proche de celui des Thaumarchées, regroupe des organismes largement retrouvés dans des environnements géothermiques ^[134].

Description du domaine des Archées après 2011

Depuis une dizaine d'années, la classification des espèces a énormément évolué avec, notamment, l'arrivée du séquençage haut-débit et le développement de la métagénomique. Cette évolution a permis d'arriver à une **nouvelle représentation de l'arbre phylogénique du Vivant** ([Figure 58d](#)) ^[121,135]. Au niveau des Archées, on compte actuellement **trois superembranchements** (TACK, DPANN et Asgard) regroupant au moins vingt embranchements, et les Euryarchées (seul embranchement à ne pas être inclus dans un superembranchement).

Les embranchements Thaumarchées, Aigarchées, Crénarchées, et Korarchées donnent leurs noms au superembranchement **TACK** ^[136]. Celui-ci contient actuellement trois embranchements supplémentaires (les Bathyarchées, les Géoarchées et les Verstraetearchées).

Les Nanoarchées appartiennent avec les embranchements Diapherotrites, Parvarchées, Aenigmarchées, et Nanohaloarchées au superembranchement **DPANN** ^[137]. Celui-ci contient actuellement quatre embranchements supplémentaires (les Micrarchées, les Woesearchées, les Pacearchées et les Altiarchées). Ce superembranchement est principalement caractérisé par des espèces aux petites cellules et avec des petits génomes.

Le superembranchement **Asgard** ^[138,139] a été proposé en 2015 par Anja Spang et ses collaborateurs ^[138] et possède actuellement quatre embranchements (les Lokiarchées, les Thorarchées, les Odinararchées et les Heimdallarchées). Il est décrit comme partageant un ancêtre commun avec le domaine des Eucaryotes ^[135,138,140,141] nommé **LAECA** (pour « Last Archaeo-Eukaryotic Common Ancestor ») ^[140], même si cette proximité phylogénétique a été critiquée par certains chercheurs. En effet, Violette Da Cunha et ses collaborateurs le décrivent comme plus proche des Euryarchées (et non des Eucaryotes) ^[142,143].

Cette nouvelle proposition de classification des espèces est toujours en évolution ^[144,145] et montre une grande proximité phylogénétique entre les Eucaryotes et les Archées. Cette nouvelle phylogénie pourrait, si on reprend la formulation d'Anja Spang et de ses collaborateurs, « combler le fossé entre les procaryotes et les eucaryotes » ^[138], mais ne suffit pas pour résoudre le débat sur la description du Vivant en deux ou en trois domaines ([Figure 59](#)) ^[146].

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

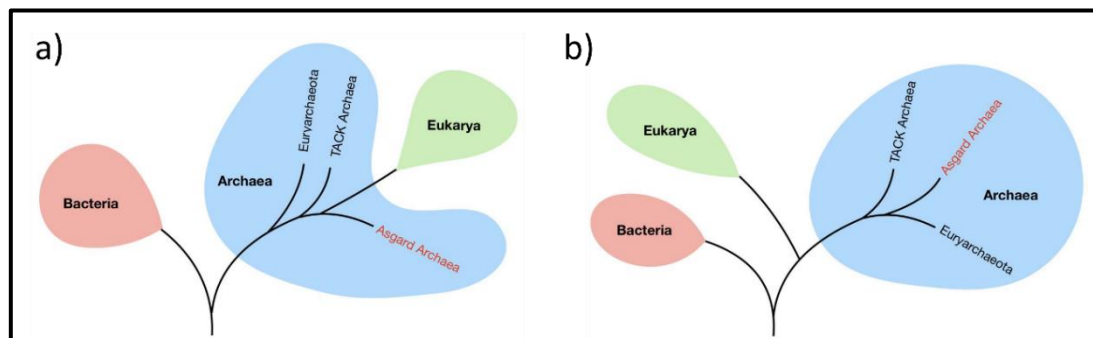


Figure 59 : **Différents arbres phylogénétiques résultant de l'inclusion du superphylum des Asgards** (adapté de MacLeod et al., 2019) ^[139]. a) L'arbre du Vivant résultant d'analyses phylogénétiques de marqueurs conservés, de gènes d'ARN ribosomique et de protéines ribosomiques. b) L'arbre du Vivant résultant d'analyses phylogénétiques des gènes de l'ARN polymérase.

1.1.2. Les principales protéines architecturales d'Archées

Chaque génome d'Archée possède **au moins deux types de protéines chromatiennes** présentant des propriétés architecturales différentes ^[147]. Ces protéines architecturales peuvent être des histones, **similaires à celles retrouvées chez les Eucaryotes**, ou de petites protéines basiques, fonctionnant de façon **analogue à celles retrouvées chez les bactéries**, les NAPs ^[148]. Les mieux caractérisées d'entre elles sont illustrées sur la [Figure 60](#) et sont décrites dans ce paragraphe à l'exception de MC1 décrite dans le paragraphe 1.1.3. Une description plus complète de la distribution des protéines de structure majeurs de l'ADN dans le Vivant, en accord avec la phylogénie actuelle et focalisée également sur les Archées, est proposée en [Annexe 5](#).

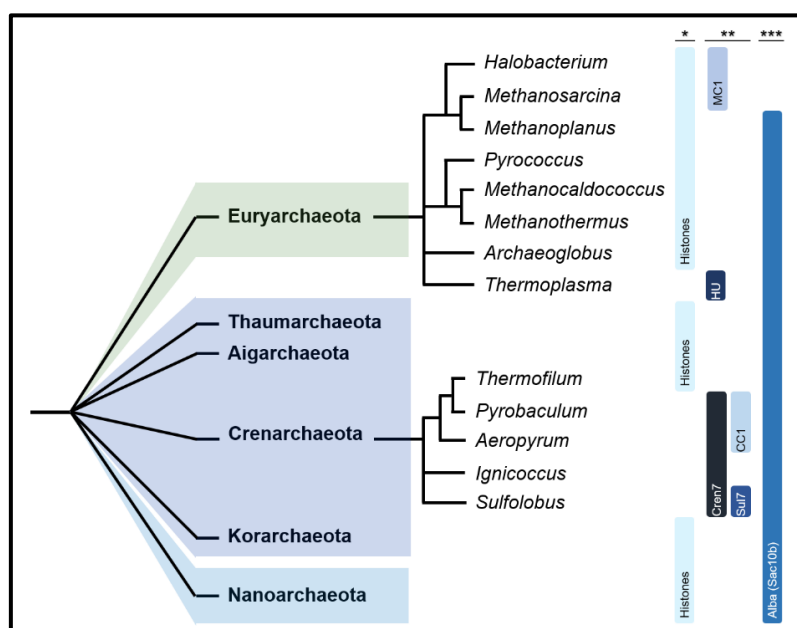


Figure 60 : **Distribution des protéines de structure majeures du nucléoïde archéen** (adapté de Zhang et al. 2012) ^[19]. Les embranchements sont représentés en gras et les genres en italique. Les astérisques correspondent aux différents modes principaux de compaction de l'ADN : *enrouleurs, **courbeurs et ***ponteurs.

Les histones

Comme chez les Eucaryotes, de nombreuses Archées possèdent des **histones** ^[149]. On les retrouve chez la grande majorité des Archées comme les Euryarchées à l'exception du genre *Thermoplasma*, les Crénarchées du genre *Thermofilum*, l'ensemble des Thaumarchées, etc. Ces protéines sont structurées en **trois hélices α séparées par deux boucles courtes** : le motif typique des histones ^[150]. De manière générale chez les Archées, les histones sont dépourvues des extensions N- et/ou C-terminales présentes chez les Eucaryotes. Cette protéine se présente sous la forme de dimères en solution (homo- et/ou hétérodimères), et sous la forme de tétramères ou l'oligomères en présence d'ADN ^[150,151]. C'est notamment le cas des histones **HMfA** et **HMfB** (pour « Histone [A et B] de *Methanothermobacter fervidus* ») de *Methanothermobacter fervidus* qui se structurent sous leur forme libre *in vivo* en homodimère et en hétérodimère ^[152]. Ce dimère est l'assemblage de deux petites sous-unités structurées par trois hélices alpha (α_1 , α_2 et α_3 , où α_2 est la plus grande hélice) et par deux boucles courtes (LP₁ et LP₂) en deux **motifs HTH**. Les deux sous-unités HMf interagissent de façon non-covalente en formant un « X » dans l'espace au niveau des hélices α_2 . Les hélices α_1 et α_3 d'une sous-unité interagissent, quant à elles, respectivement avec les parties C-terminale et N-terminale de l'hélice α_2 de l'autre sous-unité (Figure 61a). Sous leur forme complexée avec l'ADN, les dimères de HMf forment un tétramère, ou restent sous forme dimérique, et enroulent l'ADN double brin en induisant sur ce dernier des **surenroulements positifs** (Figure 61b) ^[153]. Au sein de ce complexe, les parties N-terminales des chaînes du dimère d'histone interagissent avec l'ADN en se plaçant dans la partie **concave** de la courbure induite. Là où les histones d'Eucaryotes s'assemblent en octamères sur l'ADN, les histones archéennes forment elles, des filaments de taille variable avec l'ADN appelés hypernucléosomes. Bien que les protéines des deux domaines diffèrent par leur structure quaternaire, il est à noter que les **résidus importants**, responsables de la liaison à l'ADN ainsi que ceux qui sont responsables des interactions interdimériques, sont le plus **souvent conservés** parmi les Archées ^[149].

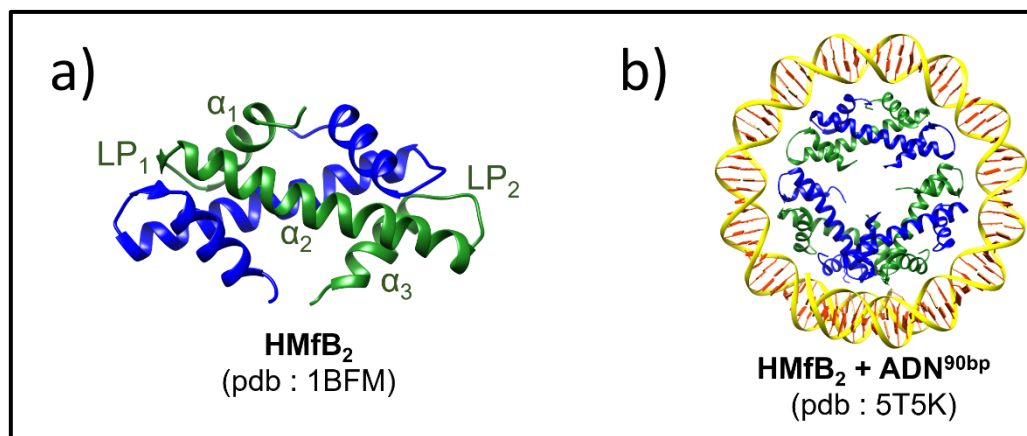


Figure 61 : Structure tridimensionnelle de HMfB₂. a) en solution par RMN (pdb : 1BFM). b) complexé à un ADN^{90bp} par cristallographie aux rayons X (pdb : 5T5K).

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

La protéine HU

Chez les Archées dépourvues d'histone, les Euryarchées du genre *Thermoplasma*, on retrouve la protéine commune à toutes les bactéries et largement décrite dans la première partie de ce manuscrit : la protéine **HU**. Pour rappel, cette protéine dimérique est structurée par un **cœur compact** d'hélices α et d'une « selle » de brins β ainsi que par deux **bras flexibles** en feuillet β (Figure 5). Elle interagit de façon non spécifique avec l'ADN montrant cependant une meilleure affinité pour les ADN courbés. Les bras flexibles du dimère interagissent avec l'ADN et le cœur compact se retrouve dans la partie **concave** de la courbure induite par la protéine sur l'ADN (Figure 6).

Il a été montré par des données de bioinformatique ^[29] qu'au moins trois organismes archéens possédaient au moins un gène codant pour la protéine HU : *Thermoplasma acidophilum* ^[154–156], *Thermoplasma volcanium* ^[29,157] et *Picrophilus torridus* ^[158]. La protéine a d'ailleurs été identifiée et caractérisée chez deux de ces organismes. **HTa** de *Thermoplasma acidophilum* présente, comme illustré sur la Figure 62, une grande homologie de séquence avec HU d'*E. coli* (48,9 % avec *EcHU α* et 47,8 % avec *EcHU β*) et semble jouer un rôle dans la protection de l'ADN contre des dénaturations thermiques et enzymatiques (digestion par la nucléase micrococcale). **HUTvo** de *Thermoplasma volcanium* présente, quant à elle, 53 % d'homologie de séquence et 33 % d'identité avec HU de *T. maritima* (Figure 62), et se structure de façon homologue à **HUTmar**.

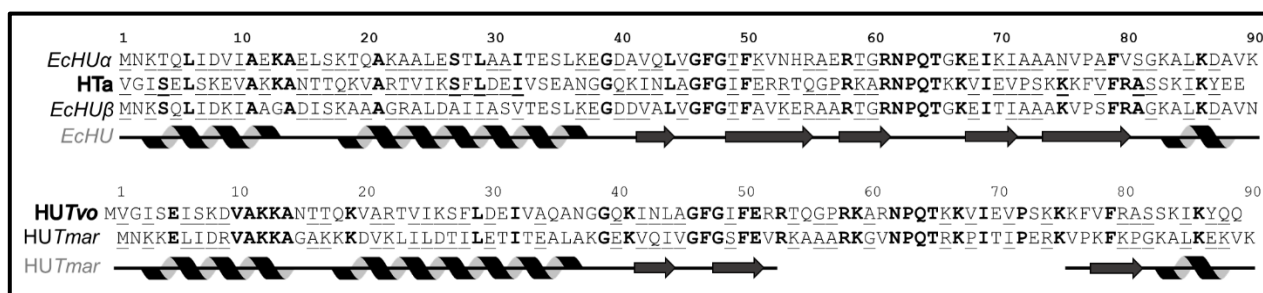


Figure 62 : **Alignement de séquences de HU d'Archées.** (haut) alignement de HTa avec HU d'*E. coli* (adapté de DeLange et al., 1980) ^[154] ; (bas) alignement de HUTvo avec HUTmar (adapté de Orfaniotou et al., 2008) ^[29].

La protéine Alba

La protéine **Alba** (= Sso10b ou Sac10b) est la **NAP la plus répandue** dans le domaine des archées. Elle est présente chez toutes les Archées à l'exception de deux genres d'Euryarchées : *Halobacterium* et *Methanosarcina*. Cette protéine très basique existe en solution sous forme dimérique ^[159]. Il s'agit généralement d'un homodimère d'environ 20 kDa, sauf chez *Sulfolobus solfataricus*, où la protéine est codée par deux gènes homologues et existe sous 3 formes : deux **homodimères** et un **hétérodimère** ^[160,161]. Chaque sous-unité se structure avec l'enchaînement : $\beta_1 - \alpha_1 - \beta_2 - \alpha_2 - \beta_3 - \beta_4$ (Figure 63). Dans l'espace, les deux hélices sont parallèles et les brins β

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

sont positionnés de façon orthogonales avec : β_1 - β_2 parallèle et β_2 - β_4 - β_3 antiparallèles. Ce sont les interactions entre les feuillets des deux sous-unités qui stabilisent le dimère. Alba est un **ponteur** d'ADN qui interagit de façon non spécifique avec les ADNs double brin. Elle interagit au niveau de la boucle entre β_1 et α_1 et au niveau de la boucle du motif en épingle à cheveux formé par les brins β_3 et β_4 terminaux. Un pont se forme lorsque deux dimères sont impliqués respectivement dans un complexe avec un ADN double brin. Pour chaque pont, une seule sous-unité des deux dimères protéiques est impliquée en interagissant par leur hélice α_1 . Alba se localise alors dans la partie concave de la boucle d'ADN formée [162].

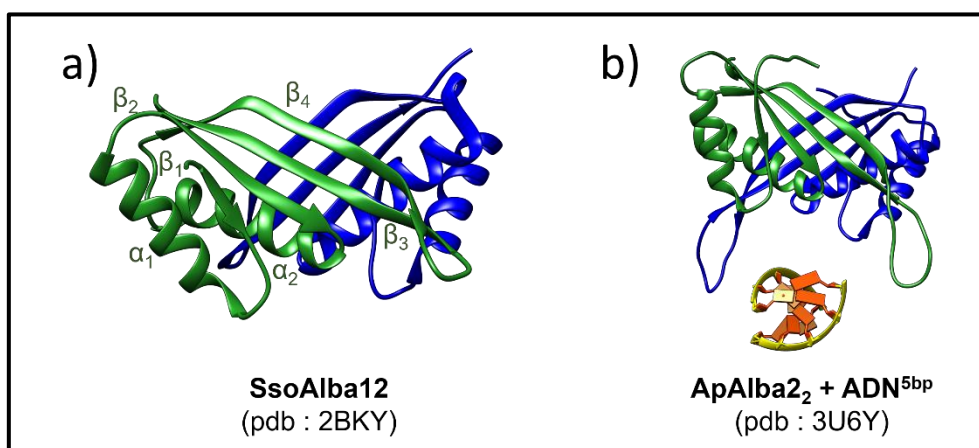


Figure 63 : **Structure tridimensionnelle d'Alba par cristallographie aux rayons X.** a) hétérodimérique libre (pdb : 2BKY). b) homodimérique et complexé à un ADN^{5bp} (pdb : 3U6Y).

La protéine Sul7

La protéine **Sul7** (=Sso7d ou Sac7d) est présente chez les Crénarchées du genre *Sulfolobus* [163,164]. Cette petite protéine **monomérique** et basique d'environ 7 kDa est un **courbeur** d'ADN. Elle se structure en un feuillet- β de deux brins antiparallèles (β_1 et β_2), d'un second feuillet de trois brins antiparallèles (β_3 , β_4 et β_5) et d'une hélice α C-terminale. Elle interagit au niveau du petit sillon de l'ADN sur **quatre paires de bases** et se place sur la partie **convexe** induite par la courbure de l'ADN complexé (Figure 64a). Pour la protéine Sac7d présente chez *Sulfolobus acidocaldarius*, une archée hyperthermophile, cette courbure de l'ADN est indépendante de la séquence en oligonucléotides et a été caractérisée par cristallographie aux rayons X par un angle de **61°** [164,165].

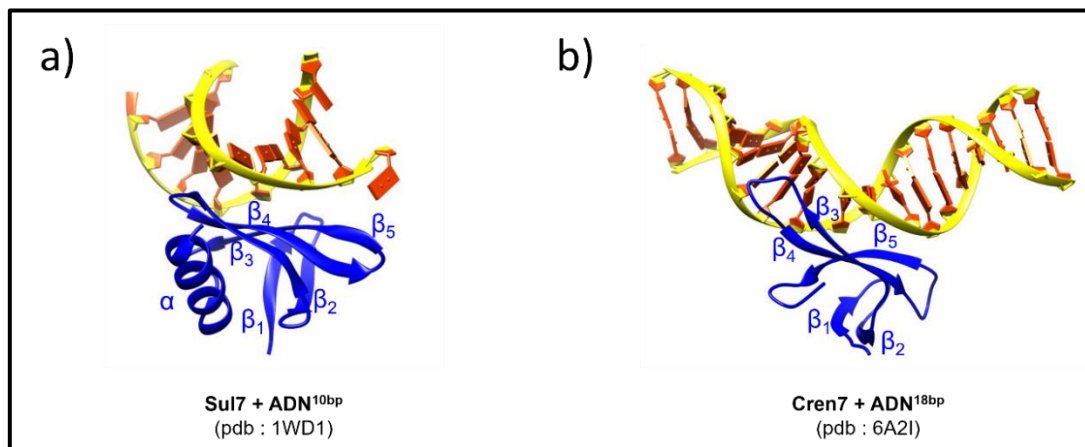


Figure 64 : Structure tridimensionnelle des protéines *Sul7* et de *Cren7* complexées à un ADN par cristallographie aux rayons X. a) complexé à un ADN^{10bp} (pdb : 1WD1). b) complexé à un ADN^{18bp} (pdb : 6A2I).

La protéine *Cren7*

La protéine **Cren7** est présente chez toutes les Crénarchées à l'exception du genre *Thermofilum*. Cette petite protéine **monomérique** et basique d'environ 7 kDa est très conservée parmi les différentes espèces d'archée [166]. Cren7 présente une **homologie de structure avec Sul7** bien que les deux NAPs diffèrent au niveau de leur composition en acides aminés (Figure 65). Cren7 se structure également en un feuillet- β de deux brins antiparallèles (β_1 et β_2) et d'un second feuillet de trois brins antiparallèles (β_3 , β_4 et β_5) [167]. Elle interagit au niveau du petit sillon de l'ADN sur environ **huit paires de bases** et se place sur la partie **convexe** de la courbure qu'elle induit sur l'ADN (Figure 64b) [165].

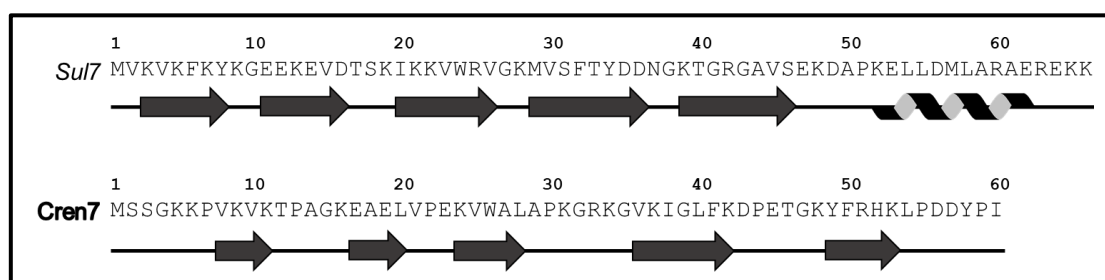


Figure 65 : Séquences primaires et secondaires de *Sul7* et *Cren7*. Séquences de *Cren7* chez *Saccharolobus Solfataricus* (pdb : 3LWH) et de *Sul7* chez *Sulfolobus acidocaldarius* (pdb : 1AZP).

La protéine *CC1*

La protéine **CC1** (pour « Crenarchaeal Chromatin protein 1 ») est présente chez les Crénarchées du genre *pyrobaculum* et *aeropyrum*. Cette petite protéine **monomérique** d'environ 6 kDa est la moins bien caractérisée d'un point de vue structural. Actuellement, seules des données de dichroïsme circulaire sont disponibles pour décrire la protéine [168]. L'analyse par dichroïsme circulaire montre que la protéine ne se structure pas en hélice, mais qu'elle présente 52 % de feuillet β et

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

26 % de tour β . De part une disponibilité limitée en termes de quantité de protéine, seule une étude 1D ^1H a pu être réalisée par RMN. Cette étude préliminaire a permis d'évaluer la structuration de la protéine grâce à l'observation de déplacements chimiques indicatifs de la présence de brins β (Ha résonant entre 4,8 ppm et 5,6 ppm), mais aussi de la présence d'un cœur hydrophobe (blindage de méthyls par effet de courant de cycle – 0,6 à 0,5 ppm) et d'une structuration stable (protons peu échangeables en milieu deutéré). La protéine semble donc être principalement structurée en **feuillet β** . Au niveau de ses interactions avec l'ADN, la protéine ne semble pas reconnaître de séquence spécifique et se fixe avec une affinité similaire aux ADNs simple et double brin.

1.1.3. La protéine MC1

MC1 de *Methanosarcina thermophila* CHTI55

Découverte en 1985 chez *Methanosarcina barkeri* ^[169,170], la protéine **MC1** (=HMb) pour « Methanogen Chromatin protein 1 » n'est retrouvée que chez les classes Methanomicrobia et Halobacteria d'Euryarchée. Chez *Methanosarcina thermophila* CHTI55, MC1 est la protéine architecturale la plus abondante dans les conditions de croissance en laboratoire. Elle est étudiée depuis une trentaine d'années au Centre de Biophysique Moléculaire par différents groupes de recherche afin de comprendre et de caractériser les mécanismes (compaction, réparation, transcription et recombinaison de l'ADN) dans lesquelles elle est ou pourrait être impliquée au sein des Methanomicrobia et Halobacteria.

Cette petite protéine monomérique (~10 kDa) est composée de **93 résidus** dont 24 sont basiques (9 Arg, 13 Lys, 2 His) et 12 sont acides (8 Glu, 4 Asp). Il est à noter que ces séquences ne présentent **aucune homologie avec des protéines connues**, ce qui laissait présager que sa structure tridimensionnelle serait originale et que ses mécanismes d'action pourraient être d'un nouveau type.

Cette protéine, **structuralement à part des autres NAPs**, participe à l'organisation du génome. Bien que les mécanismes biologiques impliquant cette protéine soient encore inconnus, il a été montré que celle-ci courbe et compacte l'ADN ^[21,171,172] et qu'elle est impliquée dans la transcription de l'ADN ^[173] et la division cellulaire ^[174,175]. Elle montre une **forte affinité** pour l'ADN **sans spécificité de séquence** ($K_D \sim 10^{-7}$ M), mais, tout comme HU, MC1 montre néanmoins une préférence pour les ADNs préalablement courbés : des mini-ADNs circulaires négativement surenroulés ($K_D < 10^{-8}$ M) ou des ADNs structurés en jonction quatre voies ($K_D \sim 2,5 \cdot 10^{-8}$ M). MC1 interagit avec l'ADN sous forme monomérique toutes les **cent paires de bases** dans la cellule. Elle interagit sur au moins **quinze paires de bases** en induisant sur l'ADN une **courbure de 116°** (observé par microscopie électronique, [Figure 66](#)) ^[102]. C'est la première fois qu'une petite protéine monomérique montre une telle capacité.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

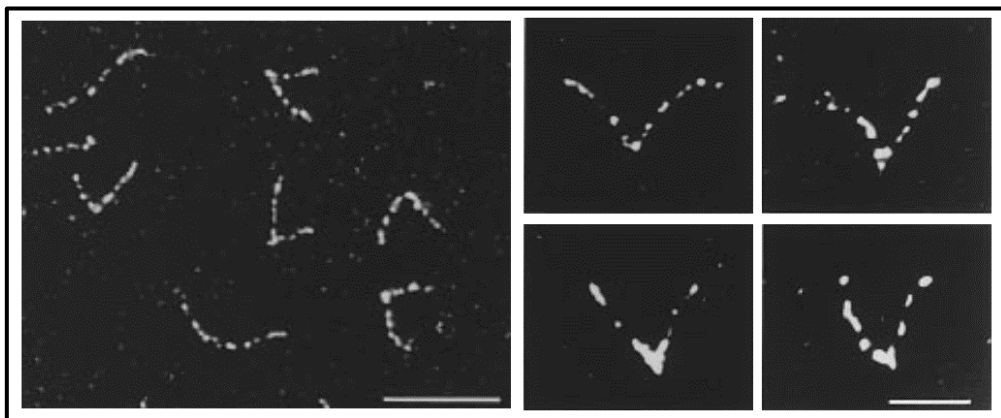


Figure 66 : Images de microscopie électronique de complexes ADN^{176bp}/MC1 (adapté de Le Cam et al, 1999) ^[172]. Les ADN^{176bp} présentent un site préférentiel pour la protéine en leur centre. Les barres d'échelle correspondent à 50 nm (gauche) et 20 nm (droite).

La **structure tridimensionnelle** de MC1 a été résolue au laboratoire par RMN dans un premier temps sur une protéine extraite et donc non marquée (pdb : 1T23) ^[176], puis sur une protéine recombinante doublement marquée (pdb : 2KHL) ^[22]. L'obtention de cette dernière, ainsi que l'évolution des logiciels permettant le calcul par modélisation moléculaire sous contraintes RMN, ont permis d'affiner la structure tridimensionnelle (**Figure 67**). De plus, des mesures de relaxation R_1 , R_2 et NOE ont pu être réalisées.

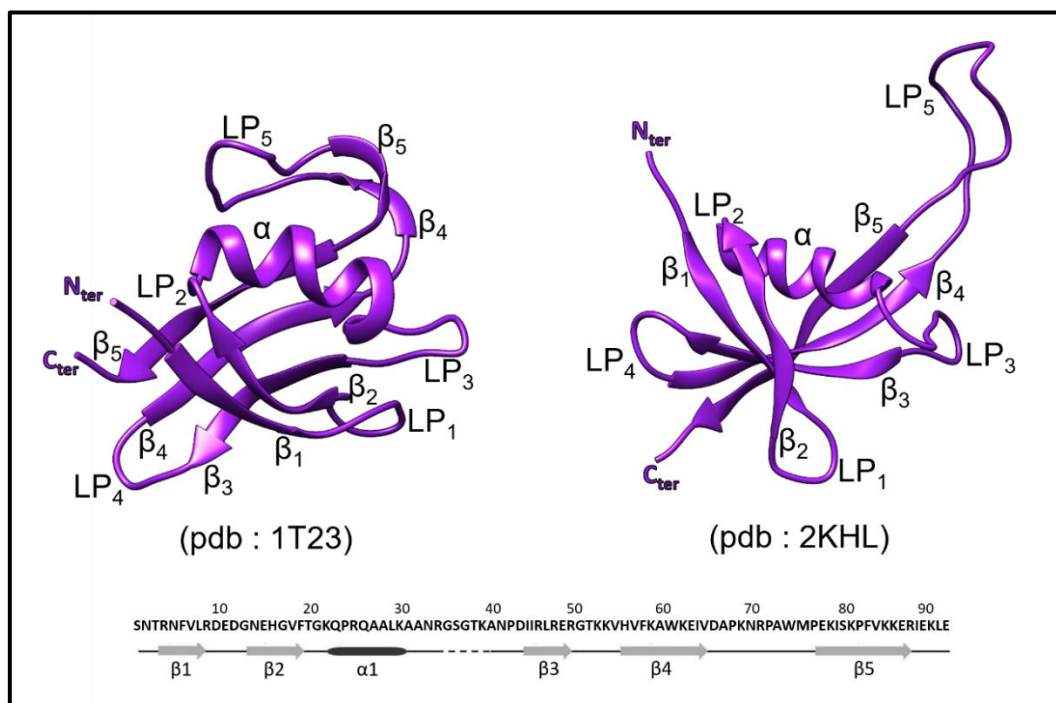


Figure 67 : Structure de la protéine MC1 de *M. thermophila* en solution. Structure tridimensionnelle de MC1 extraite par RMN ¹H à gauche (pdb : 1T23) ^[176] et recombinante par RMN hétéronucléaire à droite (pdb : 2KHL) ^[22]. Séquence de MC1 et prédiction de structuration secondaire en bas.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

La forme globale de la protéine consiste en un **pseudo tonneau** présentant une **extension au niveau des brins β_4 - β_5** formant une boucle appelée « bras » (LP5 – Ala67 à Glu77). Cette boucle, décrite sur la première structure comme en interaction avec le cœur de la protéine, est en réalité flexible et exposée au solvant. MC1 est structurée en une hélice α (Arg25 à Arg34) et deux feuillets β par l'enchaînement **β_1 - β_2 - α - β_3 - β_4 -LP5- β_5** . Le premier feuillet β est formé des deux brins antiparallèles β_1 (Thr3 à Leu8) et β_2 (Gly17 à Lys22). Le second feuillet β est formé des trois brins antiparallèles β_3 (Asp43 à Glu49), β_4 (Lys54 à Asp66) et β_5 au (Lys78 à Lys91). Ces deux feuillets présentent des déformations ou « bulges » au niveau de Gly17 (β_1) et de Ile89 (β_5).

L'alignement des séquences des protéines MC1 d'autres souches d'archées (Figure 68) montre une grande conservation de la séquence dans les différentes espèces (43 % d'identité en moyenne). Lorsque l'on met en perspective cet alignement et la structure de MC1, on remarque que les résidus de l'hélice α ainsi que de nombreux résidus hydrophobes présents au niveau du bras de MC1 (Pro68, Pro72, Trp74, Met75 et Pro76) sont **très conservés** et sont donc certainement importants pour l'activité de la protéine. Une étude de mutagenèse dirigée a par ailleurs montré que les résidus du bras LP5, **Trp74** et **Met75**, sont impliqués dans l'interaction avec l'ADN^[177]. D'un point de vue structural, il existerait une différence majeure entre les MC1 des *Methanobacteria* et des *Halobacteria* qui possèderaient une boucle LP3 plus longue.

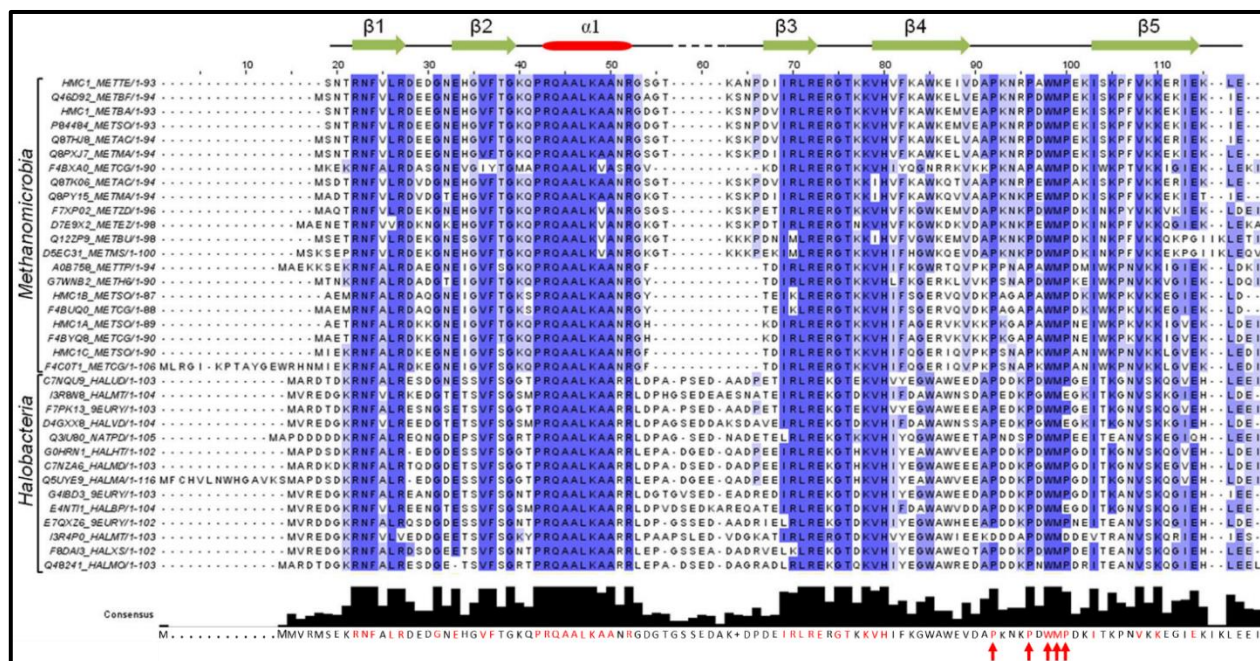


Figure 68 : Alignement de la séquence de MC1 de *Methanosarcina thermophila* (HMC1_METTE) avec des protéines MC1 des classes *Methanomicrobia* et *Halobacteria* d'Archée (adapté de Paquet et al., 2014)^[23]. Les éléments de structure secondaire HMC1_METTE sont donnés au-dessus de l'alignement. Les résidus conservés sont notés en rouge sur la séquence consensus.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Cette structure peut également être comparée à celles d'autres NAPs (Figure 69).

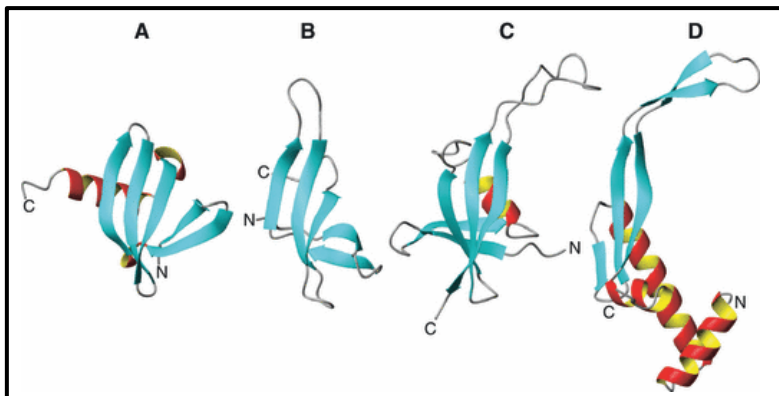


Figure 69 : Structure de protéines chromatiniennes d'Archées, différentes des histones et des homologues d'Alba (issu de Paquet et al., 2010) ^[22]. A – Sul7 (PDB : 1AZP) et B – Cren7 (PDB : 3LWI) de *Sulfolobus* (Crénarchées). C – MC1 (PDB : 2KHL) de *Methanosarcina* (Euryarchées). D – unité monomérique de HU (PDB : 1P71) de *Thermoplasma* (Euryarchées).

MC1 présente des **similarités avec Sul7 et Cren7**. Les trois protéines possèdent un repliement de type SH3, c'est-à-dire présentant un feuillet β à trois brins (β_3 - β_4 - β_5) ^[22]. De plus il a été montré que MC1 possédait comme Sul7, un **tryptophane très exposé au solvant** (Trp74) ^[178,179]. Ce résidu possède pour Sul7, un rôle dans l'interaction de la protéine avec l'ADN ^[176]. D'autres homologies entre les séquences en acides aminés de MC1 et de Sul7 ont été observées. En effet, plusieurs combinaisons de résidus présents dans Sul7 (Gly26-Lys39, Gly37-X-Lys39), et lui permettant d'interagir avec l'ADN, sont retrouvés chez MC1 (Gly21-Lys22, Gly37-Thr38-Lys39, Gly51-Thr52-Lys53) ^[176]. MC1 diffère cependant de Sul7 et de Cren7 par son potentiel électrostatique (Figure 70). Le feuillet β_3 - β_4 - β_5 de Cren7 et de Sul7 interagit avec l'ADN sur la même face, alors qu'au niveau du cœur de la protéine MC1, les résidus chargés positivement sont placés de sorte que **l'interaction à l'ADN se ferait sur la face opposée du feuillet β_3 - β_4 - β_5** . MC1 présente également des **similarités avec la protéine HU**. La protéine MC1 possède, tout comme l'unité monomérique de HU, un cœur stabilisé par des interactions hydrophobes et un bras flexible impliqué dans l'interaction à l'ADN.

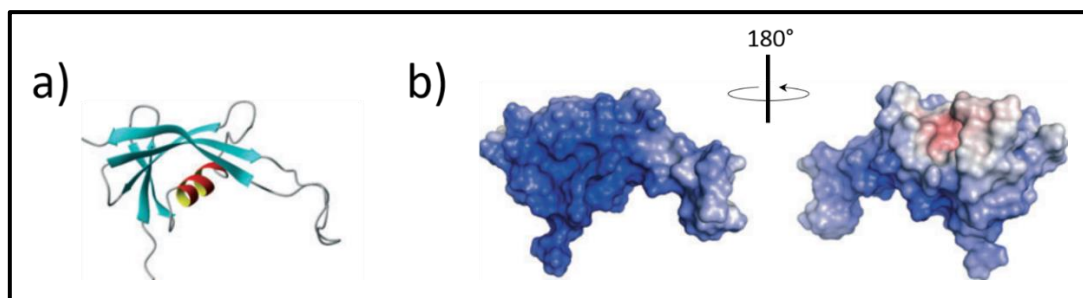


Figure 70 : Potentiel électrostatique de la protéine MC1 (adapté de Paquet et al., 2010) ^[22]. a) Diagramme en ruban de la structure de la protéine en solution. b) Surface de MC1 accessible aux solvants, codée par couleur en fonction de sa charge (le bleu et le rouge correspondent respectivement aux régions basiques et acides).

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Le complexe ADN^{15bp}/MC1

Des études sur MC1 de la souche *Methanosarcina thermophila* CHTI55 ont permis d'élucider d'un point de vue atomique par RMN, l'interaction de la protéine avec un ADN de 15 paires de bases dont la séquence a été déterminée par la méthode SELEX (pour « Systematic Evolution of Ligands by EXponential enrichment ») [180,181]. La séquence ainsi déterminée est la **séquence consensus [AAAAACACAC(A/C)CCC(C/A)]** pour laquelle MC1 a une affinité particulièrement forte ($K_D < 2$ nM). Afin de faciliter l'attribution des spectres RMN de l'ADN, la séquence [AAAAACACACACCA] a été choisie.

La comparaison des déplacements chimiques du squelette peptidique et des valeurs R_1 , R_2 et NOE entre la forme libre de la protéine et sa forme complexée a permis de conclure que la surface d'interaction de la protéine était composée des résidus Arg25 et Lys30 de l'hélice α , des résidus Pro72 à Pro76 de la boucle LP₅ et des résidus Lys85, Lys86, Arg88 et Ile89 du brin β_5 (Figure 71).

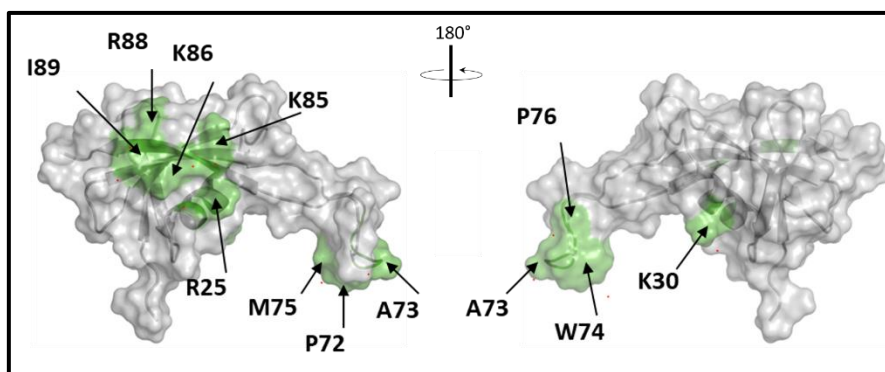


Figure 71 : **Surface d'interaction de MC1 avec l'ADN^{15bp}** (issue de Paquet et al., 2014) [23].

Au niveau de l'ADN, la séquence qui a été sélectionnée montre des propriétés structurales, dynamiques et d'interactions connues [23]. Elle se décompose en trois zones (Figure 72) :

- un **A-tract** qui, en général, possède un petit sillon plus petit que la normal présentant les charges négatives des phosphates. Celles-ci deviennent, de fait, très accessibles à des résidus protéiques chargés positivement tels que les arginines et les lysines ;
- un **triplet de CA (CpA)₃** qui est très flexible et permet à l'ADN de se courber ;
- une **succession de GC** qui présente un petit sillon élargi favorisant les interactions hydrophobes, telles que les intercalations.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

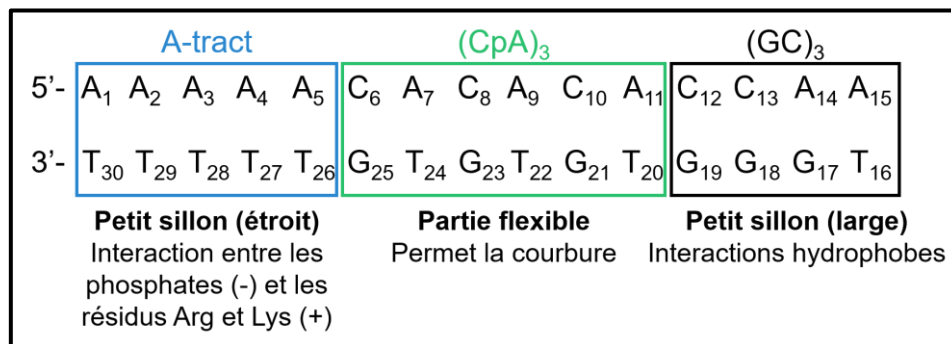


Figure 72 : Séquence d'ADN^{15bp} et ses propriétés d'interaction.

Les déplacements chimiques des hydrogènes imino de Gua25 et Thy26 ainsi que ceux des H1' de Cys6, Thy14 et Thy16 sont très impactés par l'interaction de l'ADN avec MC1 indiquant qu'il y a au moins deux points de contact entre l'ADN et la protéine avec confirmation que cette interaction provoque une courbure non négligeable de l'ADN.

Afin de confirmer l'orientation de la protéine par rapport à l'ADN, une sonde paramagnétique (dThy-EDTA-Mn²⁺) a été introduite sur l'ADN à l'extrémité 5', soit du brin supérieur (nécessite l'ajout d'une paire de bases), soit du brin inférieur. Des mesures d'atténuation de signaux ¹H de la protéine par effet paramagnétique ont ensuite été réalisées. Elles ont montré que les résidus Lys69 à Lys81 appartenant au bras de MC1 (LP₅) sont proches de la Thy16, alors que les résidus Lys54 et Val55 sont proches de la Thy(-1) et donc de Ade1.

A partir de l'ensemble de ces données, des modèles du complexe 15 paires de bases ont été modélisés grâce à la combinaison de deux méthodes :

- BioSpring, une méthode de modélisation de type « gros grain » où les deux partenaires sont considérés comme rigides et pour lesquels seuls les potentiels électrostatiques permettent de réaliser le docking ;
- HADDOCK, une méthode permettant de simuler des objets semi-flexibles ou flexibles en utilisant des contraintes ambiguës, déterminées par diverses techniques expérimentales, pour conduire le docking.

Les modèles obtenus, (Figure 73) ont permis une première description structurale du complexe ADN^{15bp}/MC1. La protéine interagit du **côté concave de l'ADN**. Elle interagit au niveau du **petit sillon**, sur **deux zones** de l'ADN. La première zone correspond au petit sillon composé de l'enchaînement de nucléotides **A1-A5**. Cet enchaînement est reconnu par les résidus basiques **Lys85**, **Lys86** et **Arg88** avec un positionnement de la **chaîne latérale du résidu Ile89** dans le petit sillon. La seconde zone correspond au petit sillon composé de l'enchaînement de nucléotides **CCCA** terminal. Cette zone interagit avec les chaînes latérales hydrophobes des résidus **Pro72**, **Trp74** et **Met75** de la boucle LP₅ de MC1. Enfin, le résidu **Arg25** est apparu comme étant essentiel pour stabiliser, par sa chaîne latérale positive, l'**enchaînement de CpA**

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

formant le grand sillon, où est induite une forte **courbure** de l'ADN. Ces modèles ont été validés par mutagenèse dirigée.

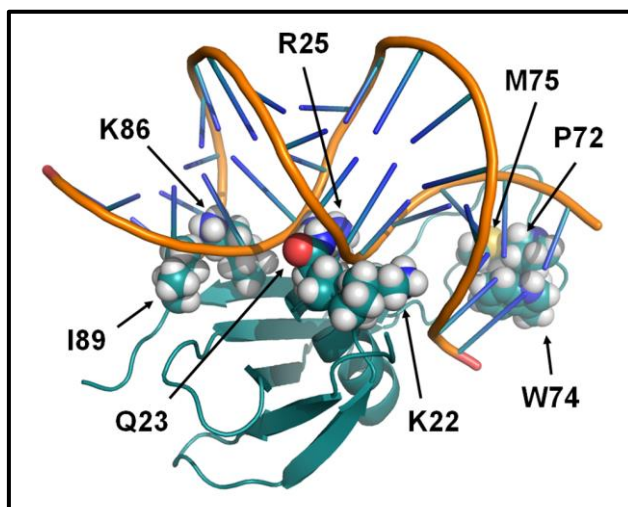


Figure 73 : Modèle de l'interaction de MC1 avec un ADN^{15bp} obtenue par HADDOCK (issue de Paquet et al., 2014) ^[23]. Les chaînes latérales des résidus Lys22, Gln23, Arg25, Pro72, Trp74, Met75, Lys86 et Ile89 sont représentées en sphères.

Cette première description du modèle ainsi que le marquage ¹⁵N¹³C du brin supérieur de l'ADN, a permis de réaliser l'attribution de l'ensemble des résonances du complexe ADN^{15bp}/MC1 ^[24] ainsi que l'attribution des spectres NOESY (intra-partenaires, mais aussi inter-partenaires). Afin d'avoir encore plus de contraintes, des mesures de RDC ont été réalisées sur la protéine et l'ADN dans leur forme complexée. Il a alors été possible de calculer la structure du complexe ^[25].

Conformément à ce qui avait été observé sur les modèles obtenus par docking, la protéine interagit du côté concave de l'ADN, au niveau des petits sillons de l'ADN^{15bp} (Figure 74A). Cependant, la structure obtenue n'est pas identique que ce soit au niveau de la structure de chaque partenaire, mais aussi de leur positionnement l'un par rapport à l'autre. En effet, la protéine a légèrement modifié sa conformation, l'ADN présente non seulement une courbure importante, mais aussi une torsion impressionnante et de nouveaux résidus en interaction ont été identifiés (Figure 35B).

Au niveau de la protéine, on observe que l'hélice est composée d'une hélice α (Pro24-Ala31) suivie d'une hélice 3_{10} (Ala32-Arg34). Elle a pivoté de 10° autour de Ala32 permettant ainsi aux chaînes latérales des résidus Gln23, Arg25 et Gln26 d'être idéalement positionnées pour interagir avec l'ADN. Pour ce qui est des autres éléments de structure secondaire, les brins β forment à présent un unique feuillet β à 5 brins antiparallèles accompagné d'une rigidification de la boucle LP₅, liée à son interaction avec l'ADN. De plus, le « bulge » présent au niveau du brin β_5 (Ile89) dans la forme libre de la protéine a disparu.

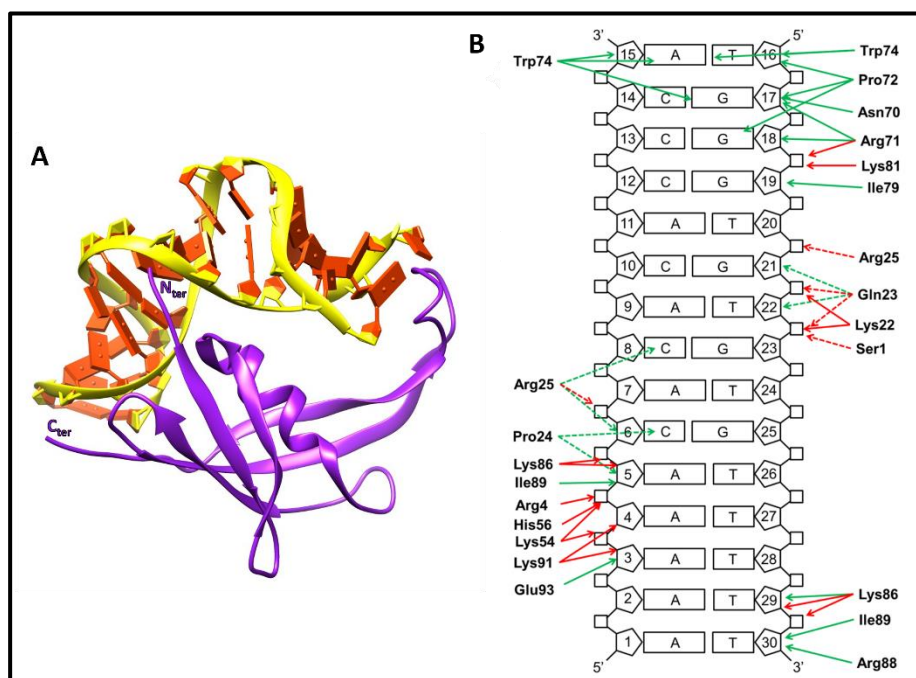


Figure 74 : Détermination structurale du complexe ADN^{15bp}/MC1 par RMN (adapté de Loth et al., 2019) [25]. A – Structure tridimensionnelle du complexe (pdb : 2NBJ). B – Interactions entre l'ADN^{15bp} et la protéine MC1. Les flèches vertes et rouges correspondent respectivement aux corrélations NOEs observées et aux liaisons hydrogène formées. Les flèches en trait continu et discontinu correspondent respectivement aux interactions avec le petit et avec le grand sillon de l'ADN.

Au niveau de l'ADN, la courbure induite par l'interaction avec la protéine a pu être caractérisée par l'existence de deux points de courbure. Le premier correspond à un angle de ($55^\circ \pm 2^\circ$) au niveau de A₄pA₅/T₂₆pT₂₇ et le second correspond à un angle plus important ($70^\circ \pm 3^\circ$) au niveau de C₁₀pA₁₁/T₂₀pG₂₁. L'ADN a une **courbure en forme de V**. Cependant cette courbure n'est pas réalisée dans un plan et présente un angle dièdre, ou angle de torsion, de ($106^\circ \pm 3^\circ$). Cette torsion impressionnante nous a laissé supposer que MC1 pourrait avoir un **mode d'action secondaire** en étant un **enrouleur d'ADN** et être impliquée dans le surenroulement négatif du nucléoïde. C'est dans ce cadre que mon travail sur la protéine MC1 a commencé.

1.2. Complexe ADN^{23bp}/MC1

Avec la torsion que MC1 provoque, on réussit très vite à imaginer qu'il pourrait exister un **troisième point de contact** sur la protéine lors d'une interaction avec des ADN plus longs. Cette idée est confortée par la présence d'un troisième point de courbure au sein du complexe à 15 paires de bases ($\theta_{C14pA15/T16pG17} \sim 40^\circ$), liée à l'intercalation du résidu Trp74 à l'extrémité du duplex et à la présence du bulge du brin β_1 (Gly17) qui est idéalement positionné dans le complexe MC1/ADN^{15bp} pour permettre une interaction avec un ADN plus long.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Dans la littérature, un nouveau complexe protéine-ADN a attiré notre attention, le complexe TFAM-ADN mitochondrial ^[182]. De façon surprenante, la séquence ADN utilisée pour réaliser ce complexe ressemble énormément à la séquence 15bp utilisée pour MC1. On y retrouve : la partie riche en nucléotides AT contenant un A-tract (encadré bleu), suivie de la partie riche en enchaînements CpA (encadré vert) et de la partie riche en nucléotides CG (encadré noir). Nous avons donc choisi de s'inspirer de la fin de cette séquence [**AACTAACA**] (encadré rouge) pour rallonger notre ADN (Figure 75).

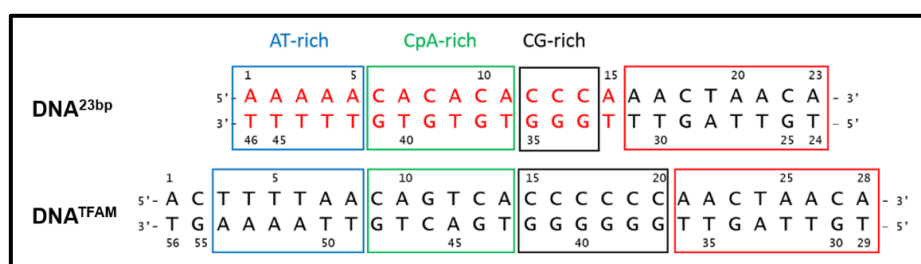


Figure 75 : séquences d'ADN d'interaction préférentielle des protéines TFAM et MC1 (issu de Loth et al., 2019) ^[25]. Les deux séquences sont en partie identiques (encadré rouge) et homologues sur trois domaines (correspondants à la séquence du 15bp) : partie riche en CG (encadré noir), partie riche en CpA (encadré vert) et partie riche en AT (encadré bleu).

Nous avons ensuite, à partir de cette séquence d'ADN^{23bp} et de la structure du complexe ADN^{15bp}/MC1, pu construire un modèle du complexe ADN^{23bp}/MC1 (Figure 76). Celui-ci prédit un nouveau point de contact sur MC1 au niveau des résidus Glu15 à Gly17 (cette zone correspond à une boucle dans MC1 libre, LP₂). Nous avons donc réalisé des expériences de biochimie pour estimer l'affinité du nouveau complexe. Celle-ci s'est avérée meilleure que pour le complexe précédent ($K_D = 1,2 \pm 0,5$ nM).

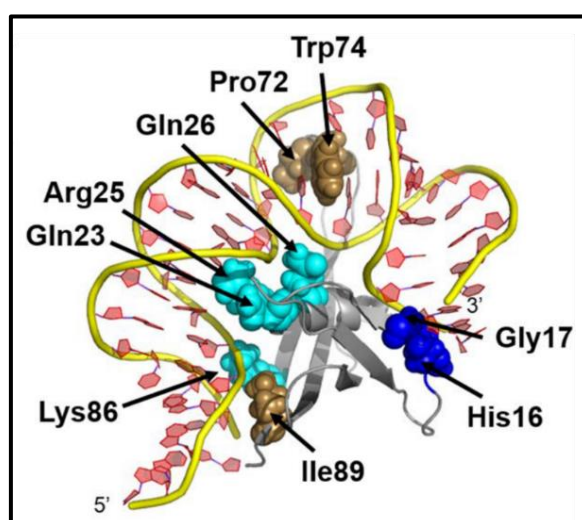


Figure 76 : Modèle de l'interaction de MC1 avec un ADN^{23bp} (issu de Paquet et al., 2019) ^[25]. Les chaînes latérales des résidus His16, Gly17, Gln23, Arg25, Gln26, Pro72, Trp74, Lys86 et Ile89 sont représentées en sphères.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Afin de caractériser le complexe ADN^{23bp}/MC1, nous avons réalisé une étude par RMN focalisée sur les signaux amides de la protéine.

1.2.1. Matériel et méthodes

Préparation des échantillons

Les surproductions et purifications de la protéine MC1 de type sauvage marquées ¹⁵N (¹⁵N-MC1) de *M. Thermophila* ont été précédemment décrites [22].

La phase de surexpression est réalisée à partir du plasmide pET24a-mc1 transformé dans des bactéries compétentes BL21(DE3). Ces bactéries sont ensuite surproduites dans un milieu LB et induites 2 h en présence d'IPTG.

La phase de purification est ensuite réalisée en 5 étapes :

- Lyse et extraction cellulaire par congélation/décongélation ;
- Chromatographie sur colonne échangeuse de cations (SP-FF) ;
- Chromatographie sur une colonne de tamisage moléculaire (AcA54) ;
- Chromatographie sur colonne échangeuse d'ions (MonoS) ;
- Enfin une étape de dialyse (1 kDa – 1 mL.cm⁻¹).

L'échantillon RMN de ¹⁵N-MC1 libre a été préparé à une concentration 0,73 mM dans un tampon phosphate (10 mM NaP pH6.0, 100 mM NaCl, 1 mM EDTA, 10 % D₂O) et analysé dans un tube 3 mm standard.

L'échantillon RMN de complexe ADN^{15bp}/¹⁵N-MC1 a été préparé à 1 mM dans un tampon phosphate (10 mM NaP pH6.0, 100 mM NaCl, 1 mM EDTA, 10 % D₂O) et analysé dans un tube 3 mm standard [23]. Les brins d'oligonucléotides synthétiques, (5'- A₁AAAACACACACCCA₁₅-3') et son complémentaire (5'- T₁₆GGGTGTGTGTTTT₃₀-3'), achetés chez Eurogentec avec une pureté estimée à 80 % (OliGold - oligonucleotide quality), ont été utilisés en abondance isotopique naturelle.

L'échantillon RMN de complexe ADN^{23bp}/¹⁵N-MC1 a été préparé à une concentration de 0,67 mM et analysé dans un tube 3 mm standard. Les deux brins complémentaires d'oligonucléotides synthétiques également achetés chez Eurogentec avec une pureté estimée à 80 % (OliGold - oligonucleotide quality), (5'- A₁AAAACACACACCCAACTAACA₂₃-3' et 5'- T₂₄GTTAGTTTGGGTGTGTGTTTT₄₆-3') ont été utilisés en abondance isotopique naturelle.

Résonance Magnétique Nucléaire : expériences de relaxation-dispersion

L'ensemble des expériences RMN a été enregistré à 25 °C sur un spectromètre Bruker 700 MHz Avance III HD équipé d'une cryosonde triple résonance inverse TCI 5 mm, avec bobine de

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

gradient Z et/ou sur un spectromètre Bruker 600 MHz Avance III équipé d'une sonde inverse triple résonance TXI 5 mm.

Les expériences de relaxation-dispersion ^{15}N de Carr-Purcell-Meiboom-Gill (RD ^{15}N -CPMG), présentées dans la [Partie 1.2.3.2](#) (Matériel et méthodes pour l'étude de la protéine *EchU*), ont été enregistrées en utilisant la séquence d'impulsion directement disponible dans le catalogue des programmes Bruker ^[99]. Ces expériences ont été enregistrées en fixant un délai T_{relax} de 80 ms pour la protéine seule et le complexe de 15 paires de bases, et un délai de 40 ms pour le complexe de 23 paires de bases. Les expériences ont été réalisées avec une série de fréquences ν_{CPMG} (0, 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 500, 600, 700, 800 et 900 Hz). Les spectres ont été analysés avec le logiciel CcpNmr ^[71]. Les intensités extraites ont ensuite été traitées avec le logiciel NESSY ^[100] pour obtenir les valeurs de R_{ex} , de k_{ex} et les populations, p_A et p_B , de chaque état impliqué dans l'échange.

1.2.2. Perturbations de déplacement chimique

Le spectre ^{15}N -HSQC du complexe ADN^{23bp}/MC1 a été enregistré, attribué ([Figure 78](#)) et comparé à ceux de la protéine seule ([Figure 77](#)) ^[22] et de la protéine complexée à l'ADN^{15bp} ^[24].

Les spectres ^{15}N -HSQC des deux complexes sont, comme attendu, assez similaires. La protéine reste globalement inchangée par l'interaction à l'ADN quelle que soit la longueur de ce dernier. Certaines différences sont toutefois à noter.

De nouveaux CSPs apparaissent au niveau des résidus **Val7** et **Arg9** du brin β_1 , ainsi que des résidus **Glu15**, **His16** et **Gly17** de la boucle LP₁ de MC1, dans le complexe à 23 paires de bases. L'allongement du double brin d'ADN a également induit un changement dans l'environnement de la protéine, observable sur le spectre par la présence de signaux plus larges pour certains résidus comme **Asn70**, **Arg71**, ou **Glu90** dans la dimension ^1H , et **Gly17** dans la dimension ^{15}N . Pour certains autres résidus, comme en particulier **Gln26**, **Lys30** et **Ala32** de l'hélice α , **Lys54** du brin β_4 , **Trp74**, **Met75** et **Glu77** de la boucle LP₅, **Ser80**, **Lys81**, **Phe83** et **Lys86** du brin β_5 , l'élargissement du signal est important à tel point que ces derniers ne sont plus détectables sur le spectre ^{15}N -HSQC. Ces élargissements de signal nous ont menés à l'hypothèse que la protéine pourrait être en échange intermédiaire entre deux conformations.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

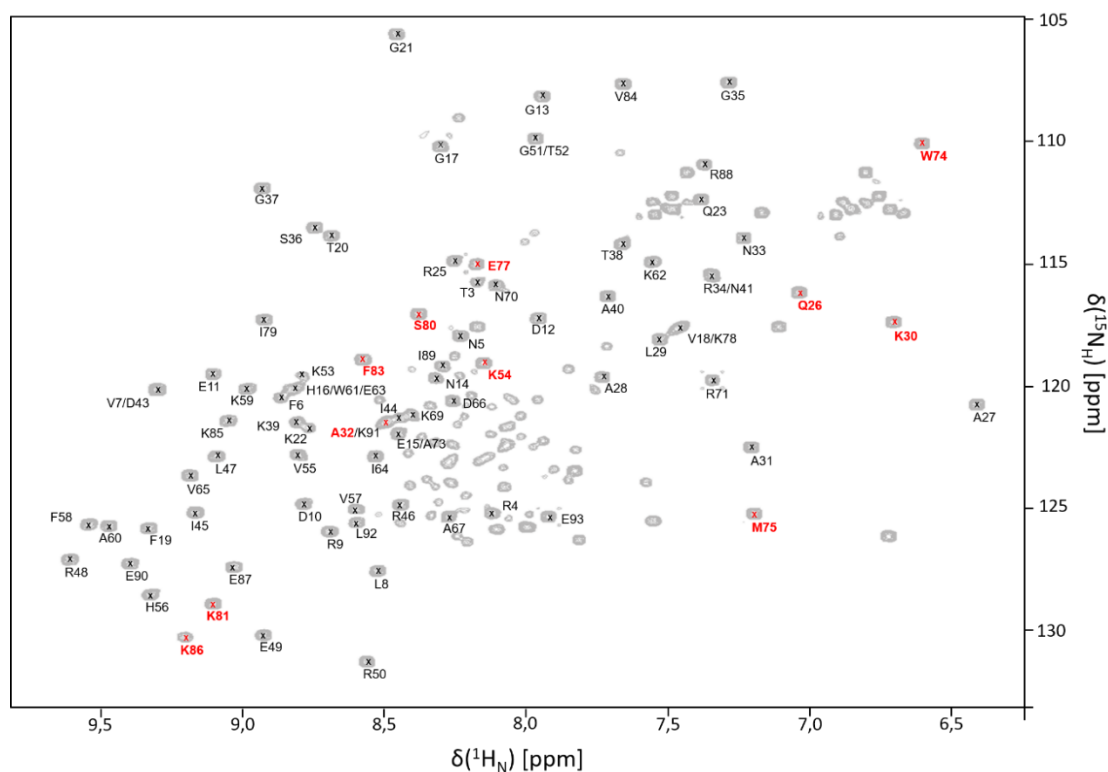


Figure 77 : Spectre 700 MHz ^{15}N -HSQC de la protéine MC1 marquée ^{15}N [0,73 mM], (100 mM NaCl, 1 mM EDTA, 10 mM tampon phosphate pH6.0), à 25 °C. Les signaux correspondants aux chaînes latérales ne sont pas identifiés sur le spectre. Les résidus indiqués en rouge ne sont plus observables sur le spectre du complexe à 23bp.

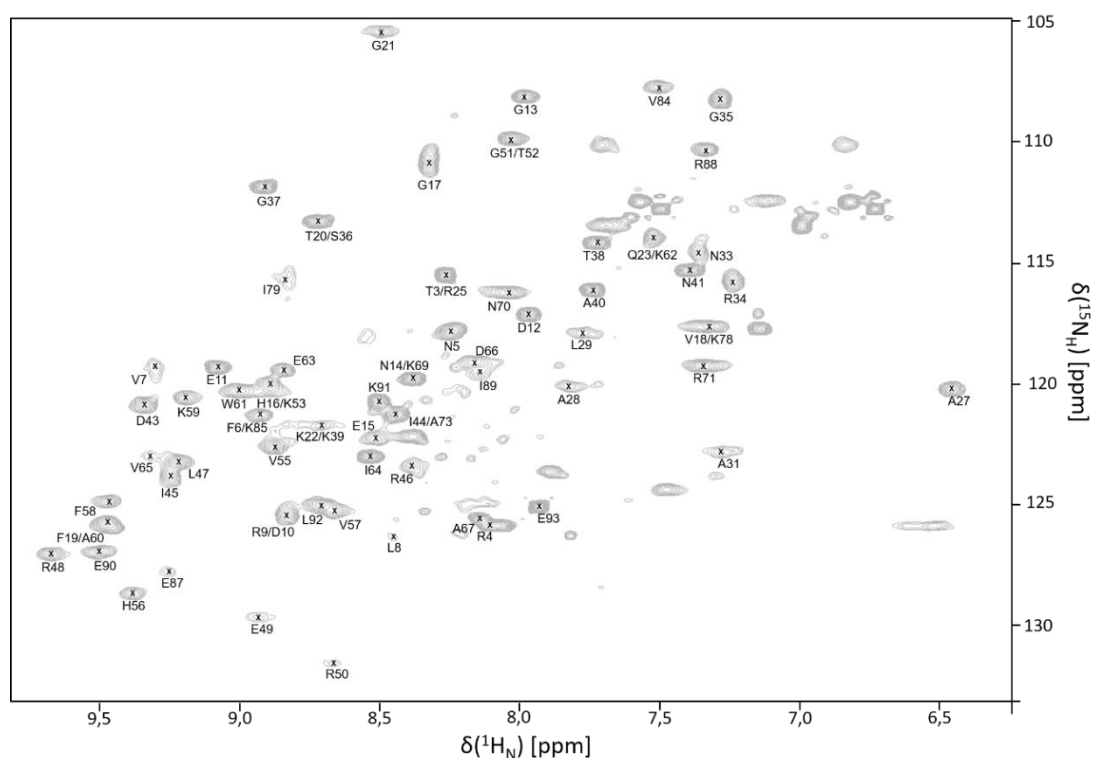


Figure 78 : Spectre 700 MHz ^{15}N -HSQC de la protéine MC1 marquée ^{15}N et complexée à un ADN de 23bp [0,67 mM], (100 mM NaCl, 1 mM EDTA, 10 mM tampon phosphate pH6.0), à 25 °C. Les signaux correspondants aux chaînes latérales ne sont pas identifiés sur le spectre.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

1.2.1. Relaxation Dispersion Carr-Purcell-Meiboom-Gill

Afin de caractériser cette potentielle dynamique d'échange, une étude complémentaire par RMN a été réalisée à l'aide des **expériences de relaxation dispersion ^{15}N -Carr-Purcell-Meiboom-Gill (RD ^{15}N -CPMG)**. Cette étude a été menée sur la protéine libre, le complexe ADN ^{15}bp /MC1 et le complexe ADN ^{23}bp /MC1.

Les expériences réalisées sur la protéine ^{15}N -MC1 libre ou complexée, permettent une étude atomique de l'échange. La limite de ces expériences est que seuls les signaux non superposés sont exploitables et quantifiables. Plusieurs résidus d'intérêt dont Val7, His16 et Ala32 n'ont pas pu être analysés.

Etude CPMG de la protéine seule

Les expériences réalisées sur la protéine ^{15}N -MC1 libres ont été menées sur un spectromètre à 600 MHz et à 700 MHz. Pour cette première étude, nous attendions une absence d'échange du fait que la protéine est libre en solution et qu'il n'est pas connue que celle-ci échange avec elle-même.

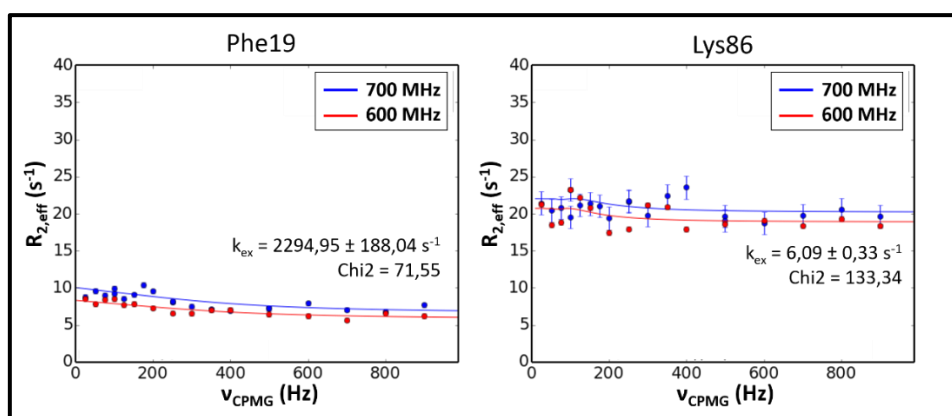


Figure 79 : Courbes de l'évolution du signal CPMG des résidus Phe19 et Lys86 de la protéine MC1 en solution.

Globalement, les expériences réalisées sur la protéine montrent une absence totale d'échange pour l'ensemble des résidus quantifiables sur les spectres ^1H - ^{15}N de relaxation dispersion ([Annexe 6](#) et [Annexe 7](#)). Seul l'affinement, dans le logiciel NESSY, des signaux des résidus **Phe19** du brin β_2 ($k_{\text{ex}} = 2294,95 \pm 188,04 \text{ s}^{-1}$) et **Lys86** du brin β_5 ($k_{\text{ex}} = 6,09 \pm 0,33 \text{ s}^{-1}$) ont permis d'obtenir une valeur de constante d'échange ([Figure 79](#)). Ces résidus sont proches des « bulges » (Gly17 et Arg88-Ile89) et cette observation pourrait donc être une conséquence de leur présence. Cependant, les résidus Gly17, Arg88 et Ile89, qui sont observables et dont les signaux peuvent être analysés, montrent une absence totale d'échange ([Annexe 6](#) pour Gly17, et [Annexe 7](#) pour Arg88 et Ile89). De plus, les expériences de dynamique moléculaire réalisées sur la protéine MC1 ont montré que la dynamique au niveau de ces bulges étaient liées à un flip-flop de liaisons hydrogène qui se produit à l'échelle de la nanoseconde ^[22]. En conclusion, bien que le programme ait été capable de nous proposer une

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

valeur de k_{ex} avec un Chi2 relativement faible pour deux résidus, Phe19 et Lys86, la forme des tracés ainsi que les observations énoncées ci-dessus, nous permet de réfuter la présence d'un échange au niveau de ces résidus.

Etude CPMG du complexe ADN^{15bp}/MC1

Les courbes d'évolution CPMG des signaux des résidus, issues de l'analyse du complexe ADN^{15bp}/MC1, ont été réalisées à 700 MHz. Elles sont globalement similaires à celles de la protéine libre.

Les résidus **Arg71** ($k_{ex} = 1407,48 \pm 202,75 \text{ s}^{-1}$) et **Glu77** ($k_{ex} = 594,23 \pm 115,00 \text{ s}^{-1}$) de la boucle LP₅ ainsi que **Ile79** ($k_{ex} = 1222,38 \pm 302,22 \text{ s}^{-1}$) du début du brin β_5 sont en échange ([Figure 80](#)). Ces trois résidus sont soit dans le bras qui interagit avec l'ADN (Arg71 forme une liaison hydrogène avec le sucre de Gua18), soit très proche de la zone d'interaction (Ile79 présente des NOEs avec Gua19). L'échange pourrait être le résultat d'un léger glissement (ou « sliding ») du bras à l'intérieur du petit sillon de l'ADN. En général, le « sliding » est observable grâce à un élargissement des signaux dans la dimension ^1H [183]. Cet élargissement dans la dimension proton est observé pour Arg71 et Glu77 ce qui appuie l'hypothèse du léger « sliding ».

Une nouvelle fois, les expériences réalisées sur la protéine impliquée dans le complexe ADN^{15bp}/MC1 montrent une absence d'échange dans l'échelle de temps accessible à ce type d'expériences pour la majorité des résidus quantifiables sur les spectres ^1H - ^{15}N de relaxation dispersion ([Annexe 8](#) et [Annexe 9](#)).

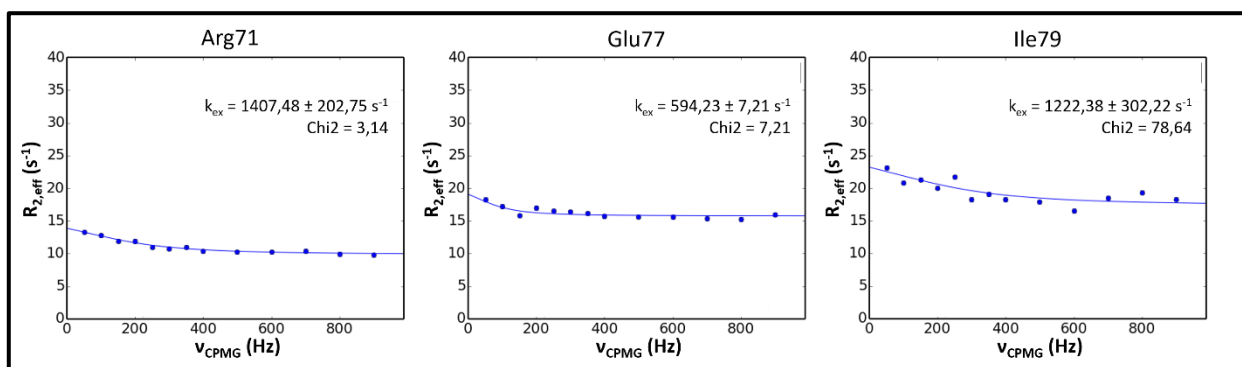


Figure 80 : *Courbes de l'évolution du signal CPMG des résidus Arg71, Glu77 et Ile79 de la protéine MC1 dans le complexe à 15bp.*

Etude CPMG du complexe ADN^{23bp}/MC1

Les courbes issues de l'analyse des données RMN à 700 MHz ont permis d'observer que de nombreux résidus étaient en échange dans le complexe à 23 paires de bases (Graphes des courbes de relaxation en [Annexe 10](#), [Tableau 20](#) et [Figure 81](#)).

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Acide Aminé	k_{ex} (s ⁻¹)	Incertitude (s ⁻¹)
Glu15	774,89	69,66
Gly17	814,06	19,14
Arg34	1323,01	159,55
Gly35	1231,79	57,59
Asp43	586,36	56,47
Ile44	1022,54	79,88
Ile45	882,54	33,06
Leu47	1456,20	129,14
Lys59	25,07	13,13
Asp66	3195,02	264,17
Arg71	2395,64	210,40
Ala73	2638,04	312,94
Lys91	945,33	66,59

Tableau 20 : Constantes d'échange issues des expériences de Relaxation Dispersion ¹⁵N-CPMG sur le complexe ADN^{23bp}/MC1.

D'après le modèle (Figure 76) les acides aminés **Glu15** et **Gly17** sont supposés être impliqués dans un nouveau point de contact avec un ADN plus long. Les résultats de CPMG dans le complexe montrent qu'ils sont effectivement en échange avec des constantes d'échange respectives de 775 ± 70 s⁻¹ et 814 ± 19 s⁻¹ (Figure 81) entre deux conformations également peuplées. Notre hypothèse est que **le complexe** ADN^{23bp}/MC1 est en **échange entre deux conformations** : celle correspondant à une interaction avec un ADN par deux points de contact (mis en évidence avec l'ADN^{15bp}), et celle correspondant à une interaction avec l'ADN par trois points de contact, expliquant ainsi la dynamique de l'ensemble des résidus situés au pied de la courbure de l'ADN. Ce qui n'est pas surprenant puisque que l'ADN^{23bp} n'est pas « optimisé » pour former un complexe stable avec les trois points de contact.

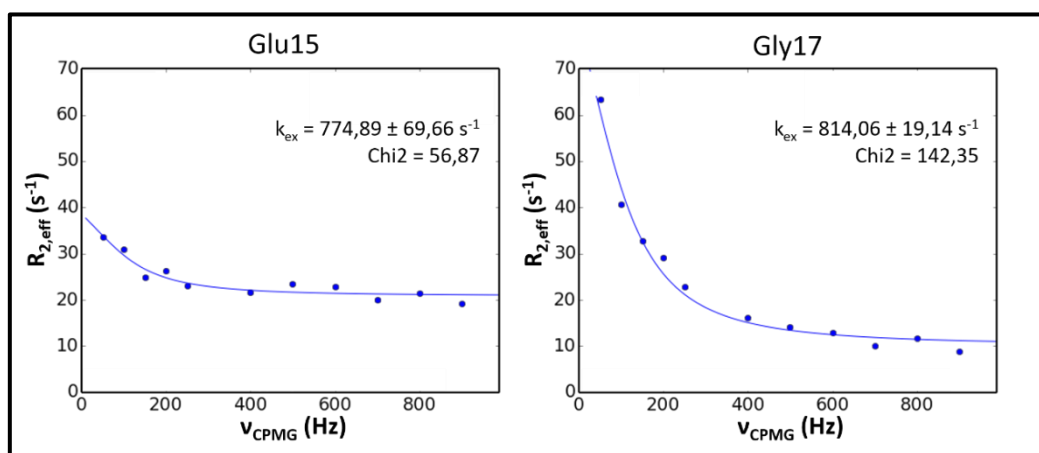


Figure 81 : Courbe de l'évolution du signal CPMG des résidus Glu15 et Gly17 de la protéine MC1 dans le complexe ADN^{23bp}/MC1.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Il semble toutefois surprenant de voir autant d'autres acides aminés en échange avec l'ADN^{23pb}. La [Figure 82](#) représente la protéine MC1 dans le complexe à 23bp où sont mis en évidence :

- Les résidus situés **au niveau des deux points de contacts établis** et montrant un échange (en marron). Parmi ces résidus, on retrouve ceux présentant les plus grandes constantes d'échange ($>2.10^3 \text{ s}^{-1}$) : Asp66, Arg71 et Ala73 de la boucle LP₅, en interaction avec le petit sillon élargi de l'ADN (partie riche en GC). Le résidu Lys91, présent au niveau du petit sillon plus étroit (partie A-tract), est en échange plus lent.
- Les résidus Arg34, Gly35, Asp43, Ile44 et Ile45 situés **à proximité du troisième point de contact** et montrant un échange (en violet). Ces résidus ont des constantes d'échange comprises entre 500 s^{-1} et 1500 s^{-1} .
- Le résidu Lys59 (en rose), montrant un échange intermédiaire-lent ($k_{\text{ex}} = 25,07 \text{ s}^{-1} \pm 13 \text{ s}^{-1}$), et situé à proximité des résidus influencés par le troisième point de contact.
- Les résidus situés **au niveau du troisième point de contact**, Glu15 et Gly17 (en bleu).

Le modèle présenté ici n'est certainement pas correct et il faudrait des données structurales supplémentaires pour l'améliorer. Cependant, au vu de cette figure, il est possible d'imaginer que l'ADN pourrait s'enrouler autour de MC1 en interagissant avec tous les acides aminés cités plus haut. Cela nécessiterait un repositionnement des deux partenaires et une légère adaptation conformationnelle de MC1.

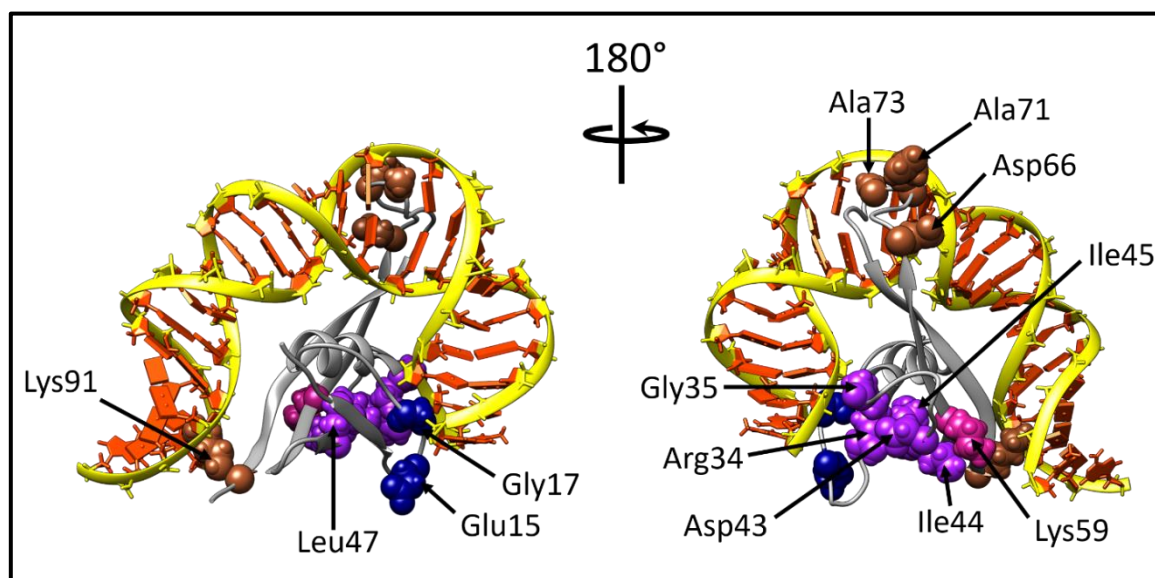


Figure 82 : Interaction dynamique de MC1 avec un ADN^{23bp}. Les résidus Glu15, Gly17, Arg34, Gly35, Asp43, Ile44, Ile45, Leu47, Lys59, Asp66, Arg71, Ala73 et Lys90, montrant un échange par RD-CPMG, sont représentés en sphères.

1.3. Conclusion

La structure 3D du complexe MC1-DNA^{15bp} montre que MC1 est capable d'induire une courbure importante de l'ADN. Cette courbure est associée à un angle de torsion de 106° et à une conformation en V de l'ADN. L'utilisation de tous les résultats obtenus au cours des trente dernières années au sein du laboratoire sur MC1 nous permet de proposer un mécanisme de complexation en trois étapes. Premièrement, le fait que l'ADN libre soit très dynamique et flexible permet à MC1 de s'associer au petit sillon du A-tract. Puis, dans un second temps, la conformation en V-turn de l'ADN est stabilisée par l'insertion ou l'intercalation des chaînes latérales de la Pro72 et du Trp74 dans le petit sillon de l'ADN. La courbure de l'ADN ne peut être maintenue que grâce à la présence de Gln23, Arg25 et Gln26 dont les chaînes latérales tamponnent les phosphates du grand sillon qui est devenu très étroit. Enfin, dans un troisième temps, MC1 peut interagir avec des ADN plus longs en établissant un troisième point de contact. Ceci nous a permis de faire l'hypothèse que non seulement MC1 est un courbeur de l'ADN, mais qu'elle doit aussi avoir un mode de compaction secondaire en tant qu'enrouleur de l'ADN.

De plus, cette structure permet d'imaginer que les complexes ADN/MC1 pourraient être des partenaires importants pour permettre la réalisation de différents mécanismes en lien avec l'ADN. En effet, la courbure de l'ADN permet de rapprocher dans l'espace des sites à l'origine distants sur un ADN, ce qui est primordial pour la recombinaison. De plus, l'élargissement du grand sillon pourrait permettre à une autre protéine nécessaire à la réparation, à la réplication ou à la transcription, de venir se fixer sur l'ADN. *In vitro*, il a été montré que l'activité des ARN polymérases d'*E. coli* peut être stimulée par l'addition de MC1 à faible concentration par rapport à la quantité d'ADN, mais que, si la protéine est ajoutée à forte concentration, l'activité est alors inhibée [173]. De ce constat, nous faisons l'hypothèse que MC1 pourrait réguler la transcription avec un mécanisme de compétition similaire à celui des histones.

Pour conclure, la détermination structurale du complexe ADN/MC1 nous ouvre de nouvelles possibilités pour étudier et comprendre les différents rôles de cette petite protéine monomérique en lien avec sa capacité à moduler l'accessibilité de l'ADN lors des mécanismes de réparation, de transcription et de réplication.

L'ensemble des travaux réalisés sur la dynamique du complexe ADN/MC1 a été valorisé par la publication d'un article scientifique ci-joint.

OPEN

New protein-DNA complexes in archaea: a small monomeric protein induces a sharp V-turn DNA structure

Karine Loth^{1,2}, Justine Largillière¹, Franck Coste¹, Françoise Culard¹, Céline Landon¹, Bertrand Castaing¹, Agnès F. Delmas¹ & Françoise Paquet¹

MC1, a monomeric nucleoid-associated protein (NAP), is structurally unrelated to other DNA-binding proteins. The protein participates in the genome organization of several *Euryarchaea* species through an atypical compaction mechanism. It is also involved in DNA transcription and cellular division through unknown mechanisms. We determined the 3D solution structure of a new DNA-protein complex formed by MC1 and a strongly distorted 15 base pairs DNA. While the protein just needs to adapt its conformation slightly, the DNA undergoes a dramatic curvature (the first two bend angles of 55° and 70°, respectively) and an impressive torsional stress (dihedral angle of 106°) due to several kinks upon binding of MC1 to its concave side. Thus, it adopts a V-turn structure. For longer DNAs, MC1 stabilizes multiple V-turn conformations in a flexible and dynamic manner. The existence of such V-turn conformations of the MC1-DNA complexes leads us to propose two binding modes of the protein, as a bender (primary binding mode) and as a wrapper (secondary binding mode). Moreover, it opens up new opportunities for studying and understanding the repair, replication and transcription molecular machineries of Archaea.

In the three domains of life, DNA-binding proteins are involved in genome organization, packaging it into eukaryotic organelles or into the cell (in bacteria or archaea), while efficiently accommodating DNA-based processes such as transcription, replication and repair. Each organism has a characteristic set of DNA-binding proteins that has unique regulatory features. Advances in this field have undoubtedly benefited from high-resolution structures of DNA-protein complexes, as well as analyses of the global and local effects of these proteins on chromatin structure.

DNA-binding proteins are classified in three categories: benders, bridgers and wrappers¹. Focusing on archaea, Alba and histone proteins are the most widely represented^{2–4}. They belong to the groups of bridgers and wrappers, respectively. Alba is present in all archaea, except in two classes of *Euryarchaea* (*Methanomicrobia* and *Haloarchaea*). In archaea lacking Alba, the Methanogen Chromosomal protein 1 (MC1) is present. Alba shows a bimodal DNA-binding behavior by bridging or stiffening the DNA⁵. The Alba family of proteins exhibits functional diversities: genome packaging and organization, transcriptional and translational regulation (through acetylation/deacetylation mechanisms), RNA metabolism, development and differentiation processes⁶. Histone proteins are present in all *Euryarchaea* (except *Thermoplasmatales*) and in some *Crenarchaea*. The homo- and heterodimers of HMfA and HMfB proteins, from the hyperthermophile *Methanothermobacter fervidus*, are the benchmark of the wrapper histone proteins⁷. They oligomerize in the presence of DNA, further forming tetramers and even hexamers. Besides, in order to compact DNA, histone proteins modulate transcriptional regulation by interplaying with transcription factors⁸.

Several other archaeal nucleoid-associated proteins (NAPs) belonging to the benders are less widely distributed among the archaeal phyla. In *Crenarchaea* lacking histone proteins, Cren7 is always present whereas Sul7 (also known as Sso7d or Sac7d) is restricted to *Sulfolobus*^{8,9}. They non-specifically bind to DNA by inserting

¹Centre de Biophysique Moléculaire, Centre National de la Recherche Scientifique UPR 4301, rue Charles Sadron, F-45071, Orléans, Cedex 2, France. ²UFR Collegium Sciences et Techniques, Université d'Orléans, rue de Chartres, 45100, Orléans, France. Correspondence and requests for materials should be addressed to K.L. (email: karine.loth@cncrs-orleans.fr) or F.P. (email: francoise.paquet@cncrs-orleans.fr)

hydrophobic residues into the minor groove, bending the DNA by up to 60° ^{10–12}. HTa and HUTvo belonging to the HU structural superfamily are found in *Thermoplasma* species, which are the only *Euryarchaea* lacking histone proteins^{13,14}. The dimeric HU-family proteins present a compact core of α -helices with two emerging flexible β -ribbon arms¹⁴. They form low-affinity HU-DNA complexes without sequence specificity and high-affinity complexes with flexible or bent DNA, such as single-strand nicks and gaps, and bulged and branched DNA structures^{15–17}.

The MC1 protein studied here belongs to *Methanosarcina thermophila*. In laboratory growth conditions, MC1 is the most abundant architectural protein in *Methanosarcina thermophila* CHT155¹⁸ and was primarily classified as histone-like protein^{19–21}. Depending on the growth conditions, MC1 could compact and organize DNA in a number of different ways. It was also proposed to be involved in DNA transcription (*in vitro* essays)²² and cellular division (proteomics by LCMS)^{23,24}, through as yet unknown mechanisms. This small basic monomeric protein of 93 residues (UniProt P12770) is structurally unrelated to other DNA-binding proteins^{25,26}. Its affinity, determined by gel retardation essays under direct competitive conditions, for any double-stranded DNA is high ($K_D \sim 10^{-7}$ M) without strong sequence specificity but with a conformational preference for bent DNA, such as negatively supercoiled DNA minicircles ($K_D < 10^{-8}$ M)²⁰ and four-way junctions ($K_D \sim 2.5 \times 10^{-8}$ M)²⁷. Binding of two MC1 molecules per DNA minicircle induces two diametrically opposed kinks visible by electron microscopy²¹ and binding of MC1 to linear 176 base pairs (bp) DNA duplexes leads to a bend angle estimated to be 116° ¹⁹. When DNA is bound to MC1, it is protected against thermal denaturation¹⁸ and radiolysis²⁸.

Recently, combining NMR and biochemical data through data-driven docking, we proposed an atypical compaction mechanism of DNA by MC1²⁹. We describe here the solution structure of an MC1-DNA^{15bp} complex, uncovering the first structure of a monomeric protein bound to the concave side of a strongly bent and distorted DNA (PDB code 2NBJ). This new data allowed us to establish that MC1, as a monomer, is able to induce a V-turn on DNAs. The existence of such a conformation opens new insights into MC1-related and MC1-based biological processes, including transcription, replication and repair.

Results

In this work, a new type of a protein-DNA complex structure, formed by MC1 and a DNA^{15bp} was determined using NMR spectroscopy data. To avoid overlaps in NMR resonance peaks, we chose the [AAAAACACACACCCA] DNA^{15bp} sequence from the consensus [AAAAACACAC(A/C)CCC(C/A)] sequences revealed by a SELEX (systematic evolution of ligands by exponential experiment) procedure³⁰.

We observed six NOEs between the protons of the side chain of Pro72 with the sugar protons of T₁₆, G₁₇ and G₁₈ located in the minor groove. Twelve other NOEs were unambiguously identified to the side chain protons of Trp74 (HB1, HB2, HD1, HE3 and HZ2) in contact with H2 and sugar protons of A₁₅ (H1') and T₁₆ (H2', H3', H5') and amino protons H21/22 of G₁₇, all located in the minor groove. Moreover, we observed eight NOEs between the side chain protons of Ile89 (HB, HD1#) and the sugar protons of A₅ (H5') and T₃₀ (H1', H4', H5' and H5'') located in the minor groove. The amino side chain of Gln23 gave observable NOEs with sugar protons of T₂₂ and amino protons of C6 located in the major groove. The use of NOE restraints is generally not sufficient to determine global structural features such as bending of nucleic acids. RDC constraints, known to improve the precision and accuracy of both the local and global structures of the double helix^{31,32}, were measured on both the DNA (the labelled strand) and protein. Their addition in the calculation was an absolute necessity to resolve the 3D structure of the complex.

The type and number of restraints, and the structural statistics for the twelve structures deposited to the PDB are reported in Table 1. Superimposition of these structures clearly shows that they are well defined (Fig. 1).

MC1 slightly adapts its conformation upon DNA binding. Small conformational changes of the 3D structure of MC1 occur upon DNA binding (Fig. 2a). In its free form, secondary structure elements of MC1 consist of an α -helix (Arg25-Ala32) and two β -sheets²⁶. The first sheet is composed of two antiparallel β -strands, β 1 (Arg4-Arg9) and β 2 (Glu15-Gly21), while the second contains three antiparallel β -strands β 3 (Asp43-Arg48), β 4 (Val55-Val65), and β 5 (Ile79-Glu90). Each β -sheet contains one antiparallel β -bulge composed of Leu8, His16, Gly17 and Val57, Glu87, Arg88, respectively (Fig. 2b). Upon DNA binding, the helix of MC1 appears longer and is, indeed, composed of an α -helix (Pro24-Ala31) followed by a small 3_{10} helix (Ala32-Arg34). The helix rotates by 10° around Ala32, the β 1-strand by 10° around Val7 and the β 2-strand by 30° around Val18, leading the side chain of Gln23, Arg25 and Gln26 to be optimally positioned for an interaction with DNA (Fig. 2c). Due to the displacement of both β 1 and β 2 strands, a unique five β -stranded antiparallel β -sheet is formed, and the first β -bulge (Leu8, His16, and Gly17) disappears. The second β -bulge (Val57, Glu87 and Arg88) also disappears, due to interactions of Lys86, Arg88 and Ile89 (β 5-strand) in the DNA minor groove. Furthermore, MC1 adopts an extended and less mobile conformation of the arm (Ala67-Glu77) due to its binding to DNA²⁶.

MC1-DNA^{15bp} interactions. MC1 binds the DNA through two main areas of contact into the minor groove of the DNA (Fig. 3a). The first one, in the A-tract, is dominated by the formation of hydrogen bonds between the side chains of Lys54, His56, Lys86 and Lys91, and DNA phosphate groups (Fig. 3b). Note that the side chain of Arg4 interacts in the major groove and Van der Waals' interactions occurring between the non-polar side chain of Ile89 and sugar groups of A₅ and T₃₀ strengthen these interactions. The second is dominated by hydrophobic interactions through the insertion of Pro72 and the intercalation of Trp74 between the C₁₄pA₁₅ step. To reinforce these interactions, Arg71 and Lys81 make electrostatic contacts with the sugar-phosphate backbone of G₁₈pG₁₉. Between these two areas of contact, the DNA is composed of a more flexible sequence (CpA)₃ in which the major groove is compressed. The presence of the positive side chain of Arg25 is essential to neutralize the repulsive negative charges of the phosphates belonging to the closely spaced C₆pA₇ and T₂₀pG₂₁ steps in the major groove (Fig. 3b). The side chains of Lys22, Gln23 and Gln26 participate in the stabilization of the complex by forming

Number of experimental restraints			
<i>Distance restraints from NOEs</i>			
MC1		1209	
DNA ^{15bp}		735	
Intermolecular		424	
<i>H-bonds</i>		50	
MC1 (secondary structures)		142	
DNA (base-pairing)		70	
Intermolecular		72	
<i>Residual dipolar coupling (RDC)</i>		0	
MC1		113	
DNA ^{15bp}		82	
Average violations per structure*			
<i>Average distance restraint violations > 0.5 Å</i>			
MC1		10.2 (1.4%)	
DNA ^{15bp}		0.3 (0.07%)	
Intermolecular		0.6 (1.2%)	
<i>H-bonds > 0.5 Å</i>			
MC1 (secondary structures)		0	
DNA ^{15bp} (base-pairing)		0	
RDC constraints rmsd (Hz)			
MC1		0.61	
DNA ^{15bp}		0.86	
Ramachandran analysis of MC1			
Residues in favored regions (%)		68.4	
Residues in additional allowed regions (%)		24.4	
Residues in generously allowed regions (%)		4.1	
Residues in disallowed regions (%)		3.1	
Average pairwise rmsd (Å)	Backbone		All heavy atoms
MC1	0.45 ± 0.10		0.67 ± 0.10
DNA ^{15bp}	0.76 ± 0.15		0.72 ± 0.13

Table 1. Experimental restraints and structure statistics for MC1-DNA^{15bp} complex. *No distance restraint violation observed at 0.7 Å and only an average of eleven non-recurring violations at 0.5 Å (ten for the protein and one intermolecular).

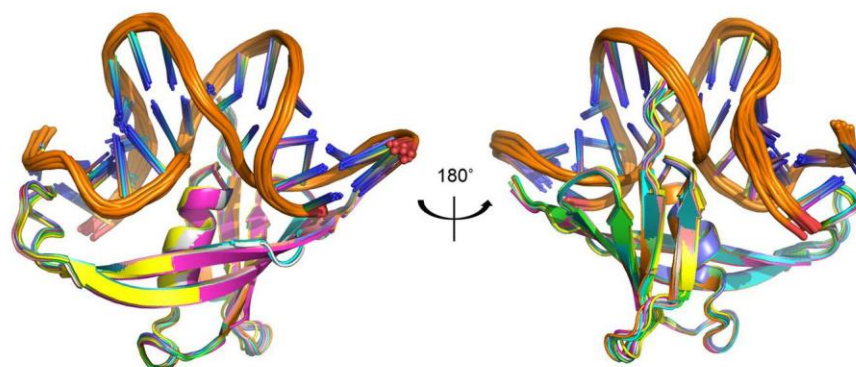


Figure 1. Overlay of the twelve-deposited structures of the MC1-DNA^{15bp} in cartoon representation. The rmsd for all heavy atoms of the complex is equal to 0.7 Å.

hydrogen bonds with the sugar-phosphate backbone of G₂₁P_{T22}P_{G23} steps, Gln26 in the minor groove, Gln23 in the major groove and Lys22 either in the minor or the major groove depending on the structures. Residues essential to DNA-binding and -bending were highlighted and confirmed by site-directed mutagenesis²⁹. MC1, as many NAPs, interacts with DNA using a combination of the two main and often interrelated readout mechanisms: recognition of the DNA shape (e.g., narrow A-tract minor groove, wide CpCpC major groove, kinking, bending) and recognition of bases (e.g., hydrogen bonds and hydrophobic interactions into the major and minor grooves)³³.

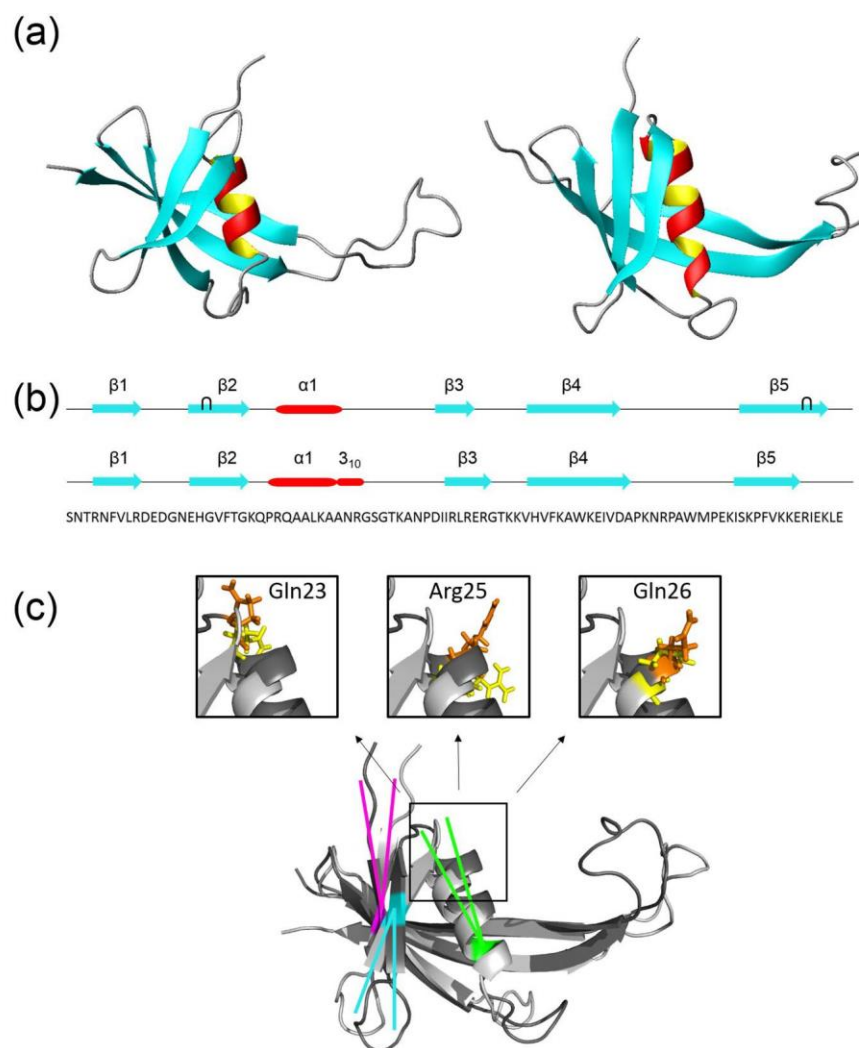


Figure 2. MC1 slightly adapts its conformation upon binding to the DNA^{15bp}. (a) 3D solution structures of MC1 in its free (left) and bound (right) forms in cartoon representation. The β -strands and the α -helix are colored in blue and red, respectively. (b) Primary and secondary structures of MC1 free (top) and bound (bottom): Blue arrows and red cylinder are for β - and α -structures, respectively, and black wickets for bulges. (c) Overlay of the cartoon representations of the free (dark grey) and bound (light grey) conformations of the MC1 protein. In order to highlight the conformational changes of MC1 upon binding, three couples of axes are represented: rotation of the α -helix by 10° around Ala32 in green, of the β 1-strand by 10° around Val7 in magenta and of the β 2-strand by 30° around Val18 in cyan. Close-up pictures of the side chains of Gln23, Arg25 and Gln26, in yellow (free MC1) and orange (bound MC1), show that their orientation have changed upon binding in order to be optimally positioned for an interaction with DNA. Alignment of the two MC1 structures, in their free and bound forms, give a rmsd of 4 Å (all residues) and a rmsd of 2.3 Å (without the arm).

MC1 imposes a severe bend and a torsion on DNA^{15bp} leading to a V-turn DNA structure. The MC1-DNA^{15bp} complex reveals large conformational perturbations in the DNA structure, which exhibits a dramatic bending, unstacked base pairs and anomalous groove widths. The DNA conformation analysis with Curves^{34,35} and 3DNA³⁶ programs cannot correctly define the base-pair parameters and the geometry of the grooves because of the very unusual conformation of the DNA, which cannot be easily classified as a B- or an A-DNA conformation. However, the roll angle profile shows two sharp peaks, which reflect distortions of base stacking owing to acute DNA bending by two kinks: the first of $(55 \pm 2)^\circ$ between A₄pA₅/T₂₆pT₂₇ ($\langle \text{tilt} \rangle = -13^\circ$, $\langle \text{roll} \rangle = -45^\circ$, $\langle \text{twist} \rangle = 50^\circ$), and the second of $(70 \pm 3)^\circ$ between C₁₀pA₁₁/T₂₀pG₂₁ ($\langle \text{tilt} \rangle = 31^\circ$, $\langle \text{roll} \rangle = -71^\circ$, $\langle \text{twist} \rangle = 20^\circ$). Negative roll results in bending of the oligonucleotide toward the minor groove, which becomes very narrow at the kinks (Fig. 4a). These two bend angles, resulting principally from large tilt angles at the site of kinking, are not coplanar and create a dihedral angle of $(106 \pm 3)^\circ$ (Fig. 4b). Between the two kinks, another striking feature is the C₈pA₉/T₂₂pG₂₃ junction ($\langle \text{tilt} \rangle = -28^\circ$, $\langle \text{roll} \rangle = 17^\circ$, $\langle \text{twist} \rangle = 50^\circ$, A₉ in north pucker) which leads a smooth bend towards the major groove. From A₇ to A₉, the minor groove is particularly wide and the major groove abnormally narrow.

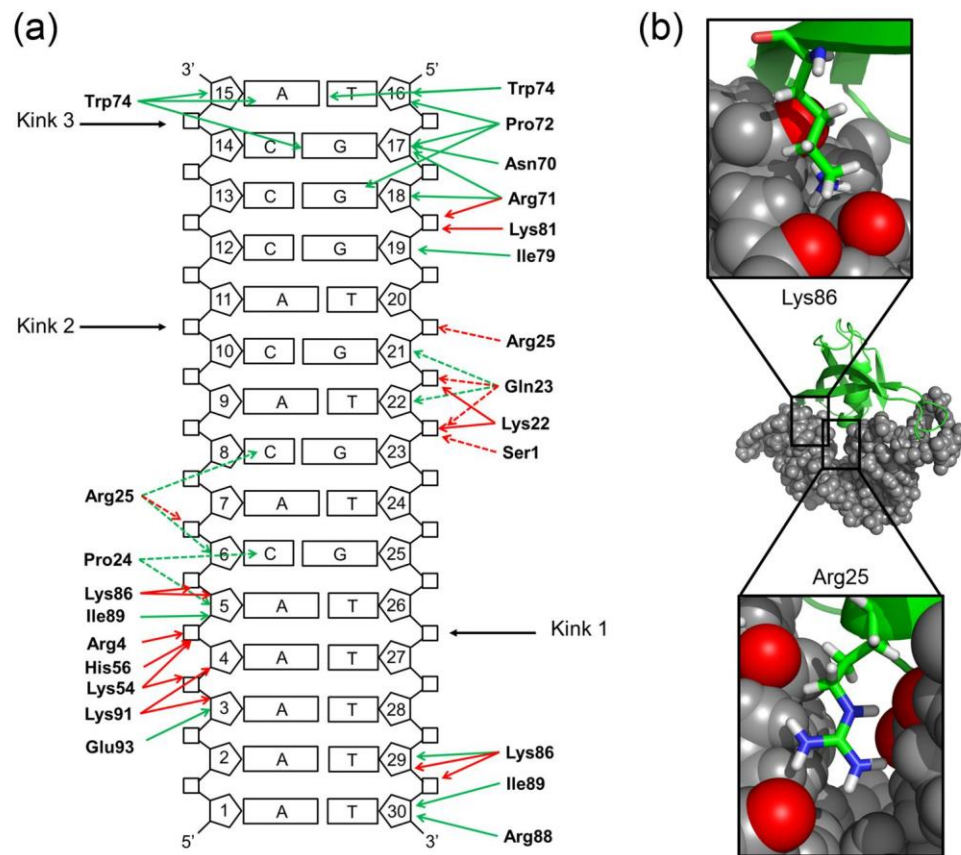


Figure 3. (a) Interactions between MC1 and DNA^{15bp}. Green and red arrows represent observed NOEs and H-bonds formed in the complex, respectively. Dashed arrows are for interactions in the major groove and standard arrows for interactions in the minor groove. Kinks are represented by black arrows. (b) Neutralization of the phosphates by side chains of Lys86 and Arg25 is essential to the compression of the minor and major grooves, respectively.

In agreement with the ability of MC1 to restrain negative supercoils, our structure shows that MC1 introduces both unwinding and negative writhe³⁷. Positive base pair tilt values and/or north pucker of the sugars are characteristic of an A-DNA-like structure. A₁pA₂/T₂₉pT₃₀, A₅pC₆/G₂₅pT₂₆, C₁₀pA₁₁/T₂₀pG₂₁ and C₁₃pC₁₄/G₁₈pG₁₉ exhibit such tilt values and the sugars of A₅, C₁₀, A₁₁, and T₂₇ adopt a north pucker (C3'-endo, C2'-exo and C4'-exo conformations). It is known that the B to A transformation leads to a rotation of -3.3° for every transformed base-pair, resulting in an overall untwisting of DNA³⁸. Moreover, the average twist in the MC1-DNA^{15bp} is 35°/bp compared with $\sim 36^\circ$ /bp or $\sim 32.7^\circ$ /bp for a relaxed B-form DNA and a relaxed A-form DNA, respectively. Although the average twist is close to that of a relaxed B-form, unwinding is locally observed around A₇pC₈/G₂₃pT₂₄, C₁₀pA₁₁/T₂₂pG₂₃ and A₁₁pC₁₂/G₁₉pT₂₀ steps with twist angle values lower than 30°. As the two kink angles are not coplanar (*vide supra*), the orientation of the resulting torsional angle is consistent with a negative supercoiling.

A third kink evaluated around 40° occurs between the two terminal base pairs C₁₄pA₁₅/T₁₆pG₁₇ upon the insertion of Pro72 and the intercalation of Trp74 (Fig. 5a). As A₁₅ and T₁₆ are unpaired, the intercalation of Trp74 implies a great perturbation in the stacking with the previous base pair preventing the correct determination of the associated base pairs parameters (tilt, roll and twist). To further extend our exploration of the DNA conformations upon MC1 binding, we decided to study a complex with a longer DNA.

A third contact observed with a longer DNA: dynamic nature of the V-turn structures. We took inspiration in the U-turn formed by the DNA^{TFAM} promoter upon interaction with TFAM, the major mitochondrial NAP comprising a tandem of HMG-box domains³⁹. Indeed, the sequence of DNA^{TFAM} has three domains (AT-rich, CpA-rich and CG-rich domains) similar to DNA^{15bp} and a supplementary domain of 8 bp which was chosen to lengthen our DNA^{15bp} to DNA^{23bp} (Fig. 5b). A structural model of the MC1-DNA^{23bp} complex was built starting from the MC1-DNA^{15bp} complex structure and new interactions with His16, and Gly17 were observed in the minor groove (Fig. 5c). This supplementary contact point, due to the bend and the torsion of the DNA, stabilizes another V-turn conformation of the DNA.

We validated the use of the chimeric DNA^{23bp} using EMSA (electrophoretic mobility shift assay) experiments (Fig. 6a). A 2.3-fold increase in affinity to MC1 was observed when the DNA was 8 bp longer ($K_{Dapp} = (1.2 \pm 0.5)$ nM for MC1-DNA^{23bp} and $K_{Dapp} = (3.1 \pm 0.6)$ nM for MC1-DNA^{15bp}). This strongly suggested that additional protein residues are involved in the interaction with DNA^{23bp}.

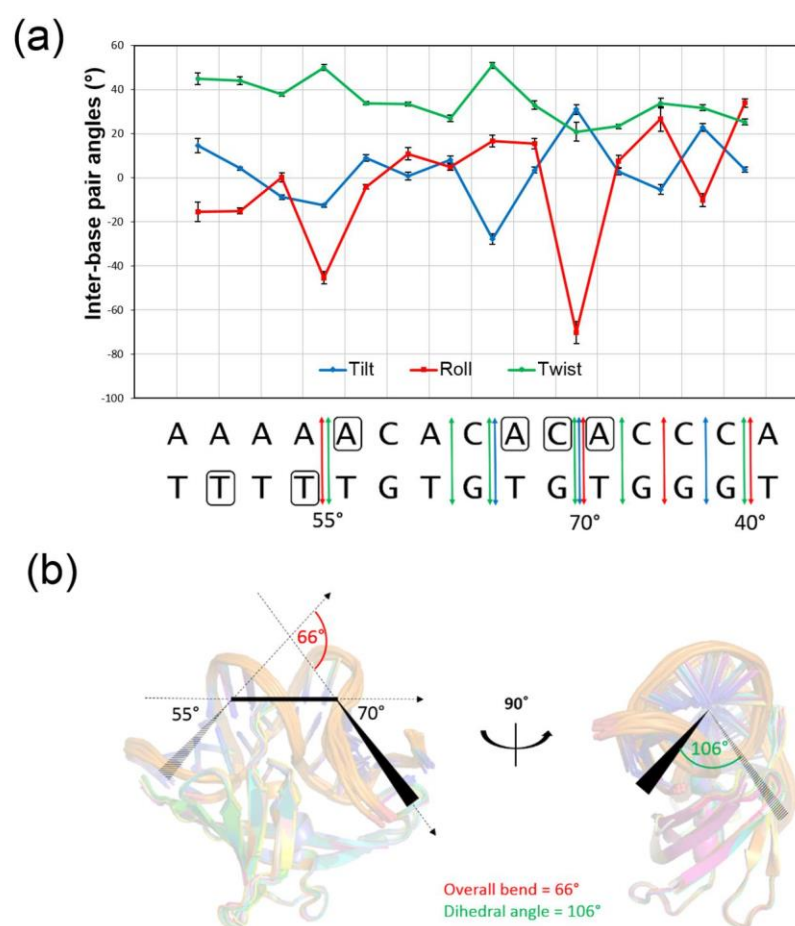


Figure 4. (a) Top: Tilt (blue), roll (red) and twist (green) angles profiles in the MC1-DNA^{15bp} structure. Bottom: DNA^{15bp} sequence, colored arrows summarized the base pairs parameters differing from standard values, boxed letters indicate sugars in north pucker conformations and values of the bend angles are given. (b) Superimposition of the 12 deposited structures of the MC1-DNA^{15bp} in cartoon representation. The two kinks are represented as well as the torsional angle.

A ¹⁵N-HSQC spectrum of the MC1-DNA^{23bp} complex revealed that this longer DNA affects both the local environment and the dynamics of the protein. Chemical shift perturbations (CSPs) between the free protein and the protein bound to DNA^{23bp} were calculated for all of the observable residues. We observed CSPs for all of the residues already highlighted in the MC1-DNA^{15bp} complex²⁶ but most importantly, new CSPs were observed for Val7, Arg9, Glu15, His16 and Gly17. Moreover, some peaks were broader than expected, leading us to hypothesize that the protein is in an intermediate exchange between at least two conformations at the NMR timescale⁴⁰. In particular, the Gln26, Lys30, Lys54, Trp74, Met75, Lys81, Phe83 and Lys86 residues exhibited such a broadened linewidth that they were undetectable. To probe the intermediate chemical exchange regime for the bound MC1, (¹H-¹⁵N) CPMG relaxation dispersion experiments were performed. Unfortunately, the resonances of some key residues, especially His16, were overlapped, preventing us from obtaining reliable measurements. However, a slow intermediate chemical exchange was observed for Glu15 and Gly17 with k_{ex} of $(775 \pm 73) \text{ s}^{-1}$ and $(814 \pm 17) \text{ s}^{-1}$, respectively (Fig. 6b). Both residues were in exchange between two equally populated conformations, one close to the MC1-DNA^{15bp} complex structure (longer DNA but only two contact areas) and one corresponding to a different V-turn in which MC1 is bound to DNA^{23bp} with a supplementary contact point.

Discussion

The 3D structure of the MC1-DNA^{15bp} complex uncovers a very strong bend in the DNA^{15bp} featuring an unprecedented torsional angle of 106°, and a V-turn DNA conformation. It is the first time that atomic data enable to corroborate data from electron micrographs showing the sharp bend angle of 116° in MC1-DNA^{176bp} complexes¹⁹.

The combination of results from our previous²⁶ and present studies agree with a three-step mechanism of complexation. Firstly, the dynamic nature of the free DNA^{15bp} structure enables MC1 to transiently select narrow minor groove segments (A-tract), which can be further stabilized through Coulombic interactions with the electropositive binding surface of the protein (Arg4, Lys54, His56, Lys86, Lys91). The widths of the DNA minor groove varies with sequence and can be a major determinant of DNA shape recognition by proteins⁴¹. For example, the NAP Fis protein selects targets primarily through indirect recognition mechanisms involving the shape of the minor groove (intrinsically narrow minor groove) and sequence-dependent induced fits over adjacent

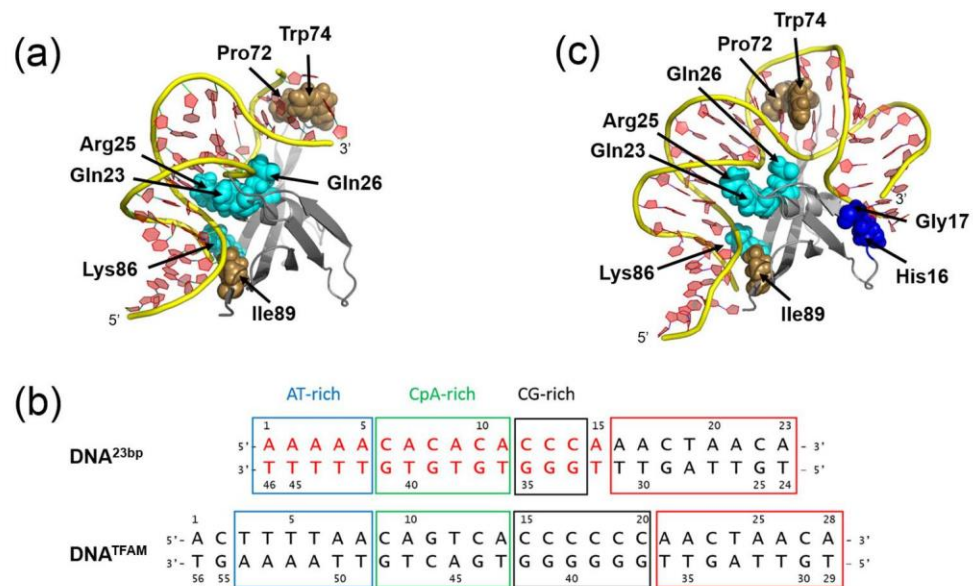


Figure 5. (a) Cartoon representation of the MC1-DNA^{15bp} complex. Side chains of key residues are represented as spheres (blue for hydrophilic and brown for hydrophobic residues, respectively). The side chains of Pro72, Trp74, Lys86 and Ile89 contact the minor groove of the DNA, whereas the side chains of Gln23, Arg25 and Gln26 contact the major groove of the DNA. (b) Sequences of the chimeric DNA^{23bp} and the DNA^{TFAM}. The first 15 base pairs of the chimeric DNA^{23bp} are identical to those of the DNA^{15bp} (written in red). Both sequences have three domains, an AT-rich (blue boxed), a CpA-rich (green boxed) and a CG-rich (black boxed) domains. The eight base pairs chosen from DNA^{TFAM} to lengthen our DNA^{15bp} into DNA^{23bp} are red boxed. (c) Cartoon representation of the model of the MC1-DNA^{23bp} complex. New contacts in the DNA minor groove are predicted for His16 and Gly17 side-chains (blue spheres).

major groove interfaces. Fis-DNA X-ray structures have revealed that narrow minor grooves containing A/T-rich sequences are compressed to a width that is about half of that observed for canonical B-form DNA^{42,43} and that the neutralization of the proximal phosphates is due to the side chain of Lys90. In MC1, the minor groove compression is particularly important between C6 and T29 (just after the A-tract) and is facilitated by the neutralization of their phosphates by the side chain of Lys86 (Fig. 3b). Secondly, the kinked DNA^{15bp} V-turn conformation is stabilized by direct hydrophobic contacts via the intercalation or insertion of hydrophobic side chains (Pro72 and Trp74) into the minor groove. The resulted dramatic bend is made possible through the neutralization of the repulsive negative charges of the phosphates belonging to the narrow major groove by the Arg25 side chain with the help of Gln23 and Gln26 side chains (Fig. 3b). Thirdly, MC1 is able to stabilize a V-turn on a longer DNA because of a supplementary contact in the minor groove. The affinity of MC1 for this supplementary area is probably weak (non-specific interactions) resulting in an equilibrium between two different V-turn conformations of the MC1-DNA^{23bp} complex.

A few other DNA binding proteins are able to impose such a drastic DNA conformational change: two eukaryotic mitochondrial proteins (human TFAM^{39,44} and its yeast counterpart Abf2p⁴⁵) and the prokaryotic nucleoid IHF^{46,47} and HU⁴⁸ family proteins (Table 2, Fig. 7). At the atomic level, MC1, IHF, HU and TFAM use similar mechanisms to strongly bend DNA: binding in the minor groove, neutralization of the negative charges of the DNA backbone and intercalation of hydrophobic residues. However, the HMG boxes of TFAM interact into the convex face of the DNA curvature, whereas IHF/HU/MC1 interact with the concave side. IHF and TFAM create two almost coplanar kinks on DNA upon binding, leading to the DNA being bent by 163° and 180°, respectively, thus reversing the direction of the DNA helix forming a U-turn (Table 2). On the contrary, HU and MC1 induce two non-coplanar kinks, resulting in a medium (40–73°) or strong (106°) dihedral angle on the DNA. Torsion has a major impact on the overall bend angle. In the case of MC1 and HU, the bend angle values are then consistent with a V-turn DNA conformation.

Our 3D structure helps us to understand how a small monomeric protein, such as MC1 with a completely asymmetric conformation, can sharply bend the DNA and impose a strong torsional angle. As previously known, MC1 is a bender but might also have a secondary binding mode as a wrapper protein. Indeed, some NAPs, as HU (bender and wrapper) or as Fis (bender and bridger) exhibit dual architectural properties that are likely dependent on protein concentration or DNA binding sequence¹. In the IHF/HU complexes, under-twisting of the DNA near the kinks is partially compensated by over-twisting of the DNA between the kinks, whereas, in the MC1-DNA^{15bp}, there is an overtwisting at the first kink and near the second kink, which is compensated by undertwisting between the kinks and after the second kink. This new mode of torsion is a key point to allow a small monomeric protein to impose a V-turn conformation on the DNA. The flexible and dynamic nature of the V-turn conformation can be modulated through the intrinsic dynamics of the protein (adaptability of its core through small local rearrangement and of its arm through a more or less extended conformation) and/

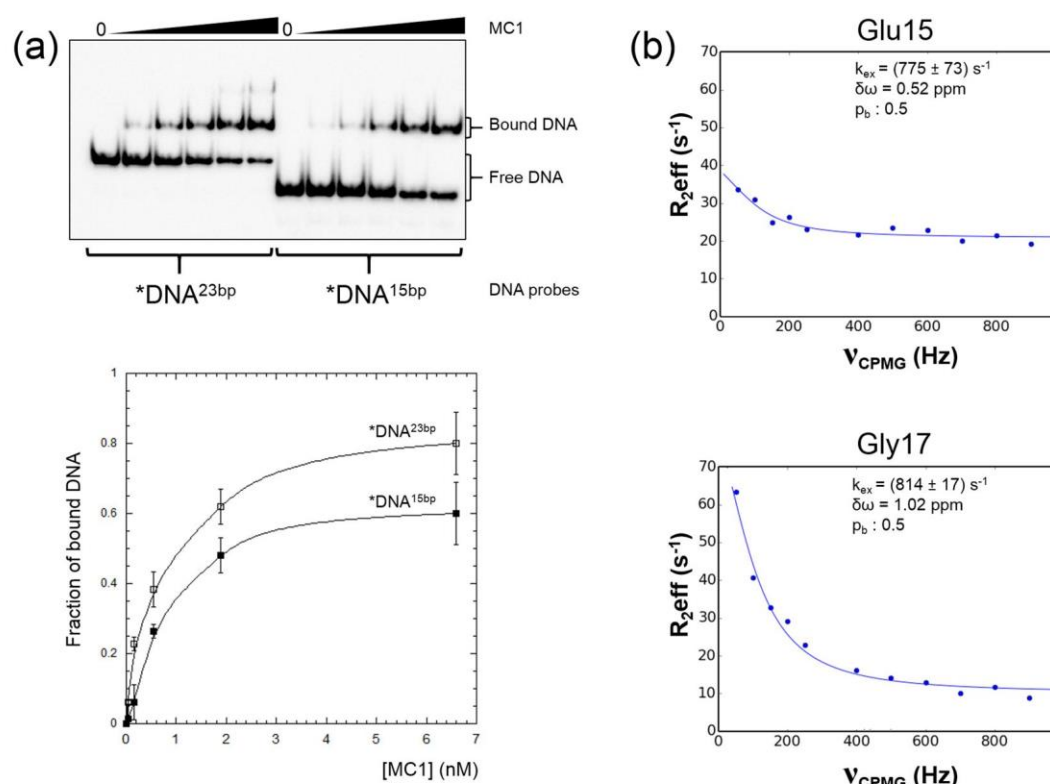


Figure 6. Binding of MC1 protein to two different DNA probes of 17 bp (*DNA^{15bp}) and 25 bp (*DNA^{23bp}). (a) Top: Electrophoresis gel of EMSA reaction mixtures (10 μ l) prepared at 4 °C by mixing DNA (0.05 nM) and increasing MC1 protein concentrations from 0.04 nM to 6.6 nM. Bottom: Binding curves obtained from quantification using ImageQuant software of the EMSA experiments. The fraction of bound DNA is plotted as a function of protein concentration [MC1]. (b) Relaxation dispersion curves, fitted using the NESSY software, for Glu15 and Gly17 of the MC1 protein in complex with the DNA^{23bp}. Data were fitted to a two-site, fast exchange model where p_a and p_b are the populations of the two state models ($p_a + p_b = 1$), k_{ex} is the exchange constant and $\delta\omega$ is the chemical shift difference between states.

Protein-DNA complex	Kink 1 (°)	Kink 2 (°)	Overall bend (°)	Dihedral angle (°)	PDB code [ref]
MC1-15bp	55	70	66	106	2NBJ
AHU2-19bp	77	82	105	73	1P78 ⁴⁸
TR3-19bp	74	84	106	72	1P71 ⁴⁸
AHU6a-17bp	91	85	130	50	1P51 ⁴⁸
AHU6b-17bp	89	82	139	40	1P51 ⁴⁸
IHF-35bp	93	87	163	15	1IHF ⁴⁶
Hbb-35bp	~90	~90	~180	<5	2NP2 ⁴⁷
TFAM-22bp	~90	~90	~180	0	3TQ6 ⁴⁴
TFAM-28bp	~90	~90	~180	0	3TMM ³⁹

Table 2. Comparison of the kinks angles, overall bend angles and dihedral angles for DNAs in complex with the MC1 (this work), HU⁴⁸, IHF^{46,47} and TFAM^{39,44} proteins.

or additional interactions with DNAs, such as those observed with the MC1-DNA^{23bp} complex. Indeed, MC1, contrary to IHF and like HU, binds DNA without sequence specificity and selects structural features of DNA for binding^{17,49}.

More than just a histone-like protein involved in the maintenance of negative supercoiling and chromosomal compaction, the flexibility and dynamics of the MC1-DNA complexes observed in solution could also reveal anchor points (i.e. molecular targets) for other molecular partners involved in various DNA transactions. The formation of such higher-order protein-DNA architectures requires the conformation of the DNA template to be bent or distorted. Bending brings distant sites of the DNA into proximity, which is necessary for the site-directed recombination process, whereas negative supercoiling favors unwinding and is likely to facilitate processes in which proteins need access to the DNA bases, such as repair, replication and transcription molecular machineries.

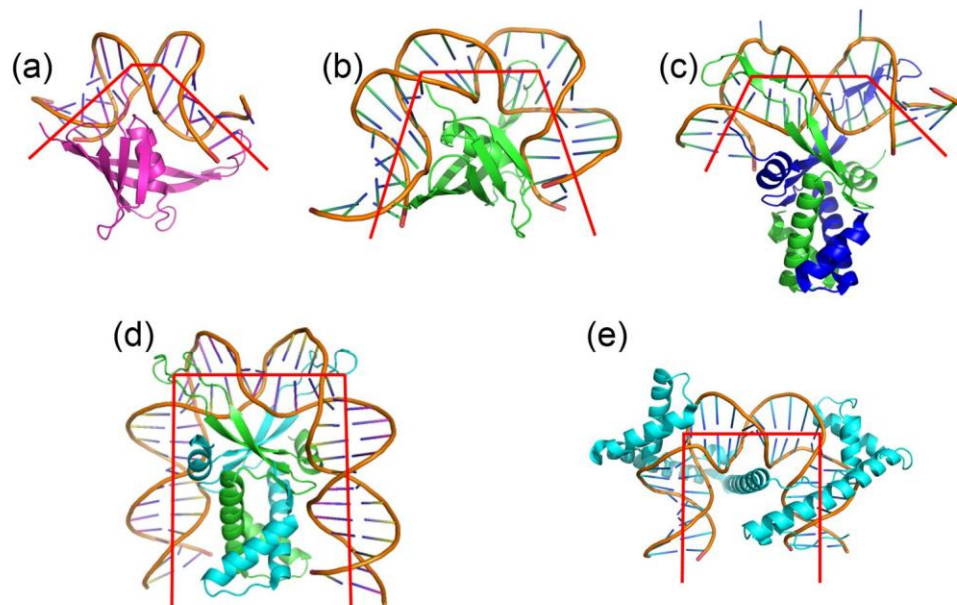


Figure 7. Cartoon representation of (a) MC1-DNA^{15bp} (PDB code 2NBJ), (b) model of MC1-DNA^{23bp}, (c) HU-DNA (PDB code 1P71), (d) IHF-DNA (PDB code 1HF) and, (e) TFAM-DNA (PDB code 3TMM) complexes. Red lines show the curvature of the DNA axis.

The binding mode of MC1 results in the formation of a particularly widened minor groove in the A₇pC₈pA₉/T₂₂pG₂₃pT₂₄ region, permitting ready access to the DNA by other DNA-binding proteins. For example, architectural DNA-binding proteins can modulate the DNA glycosylase activity. DNA glycosylases, which initiate the base excision repair (BER) process, recognize and extrude the damaged base, stabilizing it into an extrahelical conformation inside the active site pocket. To achieve this base-flipping mechanism, DNA at the lesion site is bent by inserting an intercalating residue triad inside the minor groove. The damaged nucleotide is then expelled from the DNA double helix by the major groove. We previously showed that the *E. coli* HU could stimulate the turnover of the Formamidopyrimidin-DNA glycosylase (Fpg)⁵⁰. Like HU, MC1 could play a role in DNA repair *in vivo* in Archaea.

The knowledge of which proteins are involved in modulating chromatin structure in archaeal organisms is incomplete, and views on the interplay between chromatin proteins, such as MC1, and transcription are still unknown. The archaeal transcription machinery is closely related to the RNA polymerase II system in terms of subunit composition, structure, use of general factors and molecular mechanisms⁸. There are two putative transcriptional regulatory mechanisms involving chromatin proteins in Archaea. The first involves histone proteins, which compete for DNA-binding to the promoter. Their steric and torsional effects limit the binding of basal transcription factors to the DNA^{8,51}. *In vitro*, the *E. coli* RNA polymerase activity is stimulated by the addition of MC1 at low protein to DNA ratios, but is inhibited at higher ratios²². Such a bell-shape effect was also observed in the experiments showing the stimulation of Fpg by HU⁵⁰. We propose that MC1 uses the same transcriptional regulatory mechanism of competition as histones. The second mechanism involves the post-translational modification of NAPs, especially acetylation of the Alba protein⁶, which modifies their DNA-binding properties. MC1- α from *Methanosarcina mazei* strain Gö1 was described to be specifically methylated (Lys37) by the methyltransferase Gö1-SET, *in vitro*⁵². In MC1, this Lys37 is replaced by a serine, located into the loop between the helix and the β 3-strand, a region which has not been implicated in the interaction surface of the MC1-DNA^{15bp} complex. Among the MC1 family proteins, this loop presents a great variability in the number and nature of amino acids (see alignment in ref.²⁹). It often contains at least one lysine, which could be methylated. Since methylation does not alter the DNA-binding interface of MC1- α , the role of this modification needs to be clarified. Cren7 and Sul7 are also both known to be methylated at several lysine residues without any known impact *in vitro*⁵³. Such post-translational modifications of MC1- α , Sul7, Cren7, and thus MC1, can alter their affinity for other molecular partners.

Finally, the V-shape structure of the MC1-DNA complexes provides a new perspective for the high affinity of MC1 to four-way junctions²⁷. Indeed, cruciform structures are fundamentally important for a wide range of biological processes, including replication and recombination. MC1 has been shown to be involved in cellular division in promoting replication and high rates of growth in the psychrophilic methanogen *Methanococcoides burtonii* after heat-shock at 23 °C^{23,24}. In *Halobacterium salinarum*, a growth-phase-dependent effect on the relative expression of MC1 and histone proteins was shown⁵⁴. The Holliday junction-resolvase Hjc, conserved in archaea, specifically recognizes four-way DNA junctions, cleaving them without sequence preference to generate recombinant DNA duplex. Hjc, which is retro-inhibited by its Holliday-junction cleavage product (no turnover *in vitro*) is released by the addition of the architectural double-stranded DNA-binding protein Sso7d⁵⁵. MC1 could regulate the Hjc enzymatic activity in methanosarcinal, as Sso7d does in *Sulfolobus*⁵⁵.

In summary, we have determined the 3D NMR structure of a MC1-DNA^{15bp} complex showing the remarkable ability of a small and monomeric protein to induce DNA bending and DNA torsional constraints. Our structural

work highlights a new and very efficient way, used by MC1, to produce a DNA V-turn. Indubitably, MC1 largely contributes to the organization of the genome of methanosarcinal and haloarchaeal classes of *Euryarchaea*. The structural determination of MC1-DNA complexes opens up new opportunities for studying and understanding the different emerging roles of MC1 to modulate DNA accessibility to transcription and replication machineries.

Materials and Methods

NMR sample preparation. The MC1-DNA^{15bp} complex samples (20 kDa) were prepared in 10 mM phosphate buffer pH 6, 100 mM NaCl, 1 mM EDTA, 10% D₂O and were previously described: ¹⁵N MC1-DNA^{15bp} complex²⁹ (1 mM) and ¹⁵N MC1-¹³C¹⁵N DNA^{15bp} complex (1.18 mM)⁵⁶. The first strand (5'-A₁AAAACACACACCA₁₅-3') is uniformly ¹⁵N and ¹³C-labelled and the second strand (5'-T₁₆GGGTGTGTGTTTTT₃₀-3') is in natural abundance. To measure residual dipolar couplings, the filamentous bacteriophage Pf1 (50 mg/ml, Asla Biotech. Ltd., Latvia) was added in the ¹⁵N MC1-¹³C¹⁵N DNA^{15bp} complex sample as the orienting medium. The corresponding final concentration was ¹⁵N MC1-¹³C¹⁵N DNA^{15bp} complex (0.85 mM) with Pf1 bacteriophage (10 mg/ml).

The ¹⁵N MC1-DNA^{23bp} complex sample (25 kDa) was prepared in the same way to obtain a final concentration of 0.67 mM.

NMR experiments and structure calculations of the MC1-DNA^{15bp} complex. All NMR experiments were performed at 298 K on a 700 MHz Bruker Avance III HD spectrometer equipped with a cryoprobe. All data were processed using Topspin Bruker and assignments were performed with the CcpNmr suite⁵⁷. Interproton distances were derived from NOESY data sets run at 120 ms mixing time. ¹D_{NH} RDCs were measured by using 2D IPAP ¹H-¹⁵N and ¹H-¹³C HSQC experiments acquired on a 600 MHz Varian UNITYINOVA spectrometer for both the isotropic and anisotropic conditions³².

Structures were calculated with NOE distance, hydrogen-bond, and RDC constraints using the HADDOCK web server based on restrained MD/SA (XPLORE/CNS)⁵⁸. The H-bond restraints used for the calculation of the 3D structures (Table 1) were deduced from the cross peaks observed in the NOESY spectrum. For MC1, they were set in accordance with the observation of typical long or medium NOE cross peaks network for β-sheets and α-helices respectively – H^N/H^N, H^N/H^α, H^α/H^α. For DNA, they were set in accordance with the observation of NOE cross peaks between the H1 imino protons of the guanines and the H4 amino protons of the cytosines and, between the H3 imino protons of the thymines and the H2 aromatic protons of the adenines. Our precedent model, obtained by data-driven docking²⁹, allowed us to calculate the expected protein-DNA contacts which helped us to unambiguously identify 50 intermolecular distance restraints derived from NOEs. RDC restraints were incorporated at the last iteration with the proper parameters for MC1 (Da = 18.94 and R = 0.58) and DNA (Da = 13.8 and R = 0.36). RDC restraints were fitted and analyzed with the MODULE program⁵⁹.

Docking was started from the whole ensemble of the 15 lowest-energy MC1 free structures (2KHL.pdb) and a B-standard DNA structure obtained with 3D-DART⁶⁰. The first step of the protocol⁵⁸ described for a protein-DNA complex which consisted of a rigid-body docking (1000 models) followed by a refinement stage (200 models). Both protein and DNA were successively defined to be fully flexible on all their length (first run), then fully flexible for residues 63–80 (MC1) and 12–15/16–18 (DNA^{15bp}) and finally automatically semi-flexible throughout the different runs, 40 in total. The final step of the structure refinement was performed in explicit water (200 models). The twelve structures with the lowest interaction energies and lowest constraints violations were selected for further analysis. Protein analysis was performed with Procheck. The figures were prepared with PyMOL⁶¹ and Molmol⁶².

Relaxation experiments. ¹H-¹⁵N CPMG experiments were recorded with $\nu_{\text{CPMG}} = 0, 50, 100, 150, 200, 250, 300, 350, 400, 500, 600, 700, 800$ and 900 Hz, and a CPMG constant time delay, T_{cp} of 80 ms and 40 ms for MC1 and MC1-DNA^{23bp}, respectively⁶³. Peak intensities were plotted as effective relaxation rates and relaxation dispersion curves were fitted using the Nussy software⁶⁴.

Model construction of the MC1-DNA^{23bp} complex. DNA^{23bp} was constructed with 3D-Dart starting from bound DNA^{15bp} structure and a canonical B-DNA of 8 bp was added to the terminal A₁₅T₁₆ base pair. By substituting DNA^{23bp} in the MC1-DNA^{15bp} complex structure and minimization, we obtained a model of the MC1-DNA^{23bp} complex.

Electrophoretic mobility-shift assays. Two ³²P-labeled DNA duplexes *DNA^{23bp} and *DNA^{15bp}, containing the NMR used sequences with an additional base pair at each extremity, were respectively prepared by annealing a 25-mer (5'-CAAAAACACACACCCAACTAACAG) and a 17-mer (5'-CAAAAACACACACCCAG) oligonucleotide to their complementary strand, as previously described²⁰. EMSA reaction mixtures (10 μl) were prepared at 4 °C by mixing DNA duplexes (0.05 nM) and increasing MC1 protein concentrations from 0.04 nM to 6.6 nM, in binding buffer (10 mM Tris-HCl, 400 mM NaCl, 200 μg.ml⁻¹ BSA, and 10% (v/v) glycerol, pH 7.5), followed by incubation for 30 min at 4 °C. The different mixtures were then loaded onto a 12% polyacrylamide gel (29:1 acrylamide:bisacrylamide) in 0.5 TBE buffer (44.5 mM Tris-HCl, pH 8.3, 44.5 mM boric acid, 0.5 mM EDTA). Electrophoresis was run at 14 V/cm, for 2 hours at 10 °C. After drying, gels were scanned with a β-scanner (Typhoon Trio) and quantified using ImagQuant software. The binding curves were fitted to a single binding site model using the equation $Y = [\text{MC1}] / ([\text{MC1}] + K_{\text{Dapp}})$ where Y is the fraction of bound DNA, [MC1] is the concentration of free protein, and K_{Dapp} is the apparent dissociation constant. Under our experimental conditions, the free protein concentration is reasonably approximated by the total protein concentration. The obtained values correspond to the mean (±SD) of three independent experiments.

Assession numbers. NMR restraints and atomic coordinates for the MC1-DNA^{15bp} complex have been deposited with the RCSB PDB under the accession code PDB ID: 2NBJ.

References

- Luijsterburg, M. S., White, M. F., van Driel, R. & Dame, R. T. The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Crit. Rev. Biochem. Mol. Biol.* **43**, 393–418, <https://doi.org/10.1080/10409230802528488> (2008).
- Sandman, K., Krzycki, J. A., Dobrinski, B., Lurz, R. & Reeve, J. N. HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermobacter fervidus*, is most closely related to histones. *Proc. Natl. Acad. Sci. USA* **87**, 5788–5791 (1990).
- Starich, M. R., Sandman, K., Reeve, J. N. & Summers, M. F. NMR structure of HMfB from the hyperthermophile, *Methanothermobacter fervidus*, confirms that this archaeal protein is a histone. *Journal of molecular biology* **255**, 187–203, <https://doi.org/10.1006/jmbi.1996.0016> (1996).
- Sandman, K. & Reeve, J. N. Archaeal chromatin proteins: different structures but common function? *Curr. Opin. Microbiol.* **8**, 656–661, <https://doi.org/10.1016/j.mib.2005.10.007> (2005).
- Laurens, N. *et al.* Alba shapes the archaeal genome using a delicate balance of bridging and stiffening the DNA. *Nat. Commun.* **3**, 1328, <https://doi.org/10.1038/ncomms2330> (2012).
- Goyal, M., Banerjee, C., Nag, S. & Bandyopadhyay, U. The Alba protein family: structure and function. *Biochim. Biophys. Acta* **1864**, 570–583, <https://doi.org/10.1016/j.bbapap.2016.02.015> (2016).
- Marc, F., Sandman, K., Lurz, R. & Reeve, J. N. Archaeal histone tetramerization determines DNA affinity and the direction of DNA supercoiling. *J. Biol. Chem.* **277**, 30879–30886, <https://doi.org/10.1074/jbc.M203674200> (2002).
- Peeters, E., Driessen, R. P., Werner, F. & Dame, R. T. The interplay between nucleoid organization and transcription in archaeal genomes. *Nat. Rev. Microbiol.* **13**, 333–341, <https://doi.org/10.1038/nrmicro3467> (2015).
- White, M. F. & Bell, S. D. Holding it together: chromatin in the Archaea. *Trends Genet.* **18**, 621–626 (2002).
- Robinson, H. *et al.* The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature* **392**, 202–205, <https://doi.org/10.1038/32455> (1998).
- Chen, C. Y. *et al.* Probing the DNA kink structure induced by the hyperthermophilic chromosomal protein Sac7d. *Nucleic Acids Res.* **33**, 430–438, <https://doi.org/10.1093/nar/gki191> (2005).
- Guo, L. *et al.* Biochemical and structural characterization of Cren7, a novel chromatin protein conserved among Crenarchaea. *Nucleic Acids Res.* **36**, 1129–1137, <https://doi.org/10.1093/nar/gkm1128> (2008).
- DeLange, R. J., Williams, L. C. & Searcy, D. G. A histone-like protein (HTa) from *Thermoplasma acidophilum*. II. Complete amino acid sequence. *J. Biol. Chem.* **256**, 905–911 (1981).
- Orfanidou, F. *et al.* The stability of the archaeal HU histone-like DNA-binding protein from *Thermoplasma volcanium*. *Extremophiles: life under extreme conditions* **13**, 1–10, <https://doi.org/10.1007/s00792-008-0190-6> (2009).
- Pontiggia, A., Negri, A., Beltrame, M. & Bianchi, M. E. Protein HU binds specifically to kinked DNA. *Mol. Microbiol.* **7**, 343–350 (1993).
- Castaing, B., Zelwer, C., Laval, J. & Boiteux, S. HU protein of *Escherichia coli* binds specifically to DNA that contains single-strand breaks or gaps. *J. Biol. Chem.* **270**, 10291–10296 (1995).
- Kamashev, D., Balandina, A. & Rouviere-Yaniv, J. The binding motif recognized by HU on both nicked and cruciform DNA. *EMBO J.* **18**, 5434–5444, <https://doi.org/10.1093/emboj/18.19.5434> (1999).
- Chartier, F., Laine, B. & Sautiere, P. Characterization of the chromosomal protein MC1 from the thermophilic archaeobacterium *Methanosarcina* sp. CHT1 55 and its effect on the thermal stability of DNA. *Biochim. Biophys. Acta* **951**, 149–156; 0167-4781(88)90035-8 (1988).
- Le Cam, E., Culard, F., Larquet, E., Delain, E. & Cognet, J. A. DNA bending induced by the archaeobacterial histone-like protein MC1. *Journal of molecular biology* **285**, 1011–1021, <https://doi.org/10.1006/jmbi.1998.2321> (1999).
- Teyssier, C. *et al.* Preferential binding of the archaeobacterial histone-like MC1 protein to negatively supercoiled DNA minicircles. *Biochemistry* **35**, 7954–7958, <https://doi.org/10.1021/bi952414o> (1996).
- Toulme, F. *et al.* Conformational changes of DNA minicircles upon the binding of the archaeobacterial histone-like protein MC1. *J. Biol. Chem.* **270**, 6286–6291 (1995).
- Chartier, F., Laine, B., Belaiche, D., Touzel, J. P. & Sautiere, P. Primary structure of the chromosomal protein MC1 from the archaeobacterium *Methanosarcina* sp. CHT1 55. *Biochim. Biophys. Acta* **1008**, 309–314 (1989).
- Williams, T. J. *et al.* Global proteomic analysis of the insoluble, soluble, and supernatant fractions of the psychrophilic archaeon *Methanococcus burtonii*. Part I: the effect of growth temperature. *J. Proteome Res.* **9**, 640–652, <https://doi.org/10.1021/pr900509n> (2010).
- Goodchild, A., Raftery, M., Saunders, N. F., Guilhaus, M. & Cavicchioli, R. Biology of the cold adapted archaeon, *Methanococcus burtonii* determined by proteomics using liquid chromatography-tandem mass spectrometry. *J. Proteome Res.* **3**, 1164–1176, <https://doi.org/10.1021/pr0498988> (2004).
- Paquet, F., Culard, F., Barbault, F., Maurizot, J. C. & Lancelot, G. NMR solution structure of the archaeobacterial chromosomal protein MC1 reveals a new protein fold. *Biochemistry* **43**, 14971–14978, <https://doi.org/10.1021/bi048382z> (2004).
- Paquet, F. *et al.* Refined solution structure and backbone dynamics of the archaeal MC1 protein. *FEBS J.* **277**, 5133–5145, <https://doi.org/10.1111/j.1742-4658.2010.07927.x> (2010).
- Paradinas, C., Gervais, A., Maurizot, J. C. & Culard, F. Structure-specific binding recognition of a methanogen chromosomal protein. *Eur. J. Biochem.* **257**, 372–379 (1998).
- Isabelle, V. *et al.* Radioprotection of DNA by a DNA-binding protein: MC1 chromosomal protein from the archaeobacterium *Methanosarcina* sp. CHT155. *Int. J. Radiat. Biol.* **63**, 749–758 (1993).
- Paquet, F. *et al.* Model of a DNA-protein complex of the architectural monomeric protein MC1 from Euryarchaea. *PLoS one* **9**, e88809, <https://doi.org/10.1371/journal.pone.0088809> (2014).
- De Vuyst, G., Aci, S., Genest, D. & Culard, F. Atypical recognition of particular DNA sequences by the archaeal chromosomal MC1 protein. *Biochemistry* **44**, 10369–10377, <https://doi.org/10.1021/bi0474416> (2005).
- Vermeulen, A., Zhou, H. & Pardi, A. Determining DNA global structure and DNA bending by application of NMR residual dipolar couplings. *J. Am. Chem. Soc.* **122**, 9638–9647 (2000).
- Hansen, M. R., Hanson, P. & Pardi, A. Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. *Methods Enzymol.* **317**, 220–240 (2000).
- Rohs, R. *et al.* Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **79**, 233–269, <https://doi.org/10.1146/annurev-biochem-060408-091030> (2010).
- Blanchet, C., Pasi, M., Zakrzewska, K. & Lavery, R. CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res.* **39**, W68–73, <https://doi.org/10.1093/nar/gkr316> (2011).
- Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **37**, 5917–5929, <https://doi.org/10.1093/nar/gkp608> (2009).
- Lu, X. J. & Olson, W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31**, 5108–5121 (2003).
- Laine, B., Culard, F., Maurizot, J. C. & Sautiere, P. The chromosomal protein MC1 from the archaeobacterium *Methanosarcina* sp. CHT1 55 induces DNA bending and supercoiling. *Nucleic Acids Res.* **19**, 3041–3045 (1991).
- Saenger, W. *Principles of nucleic acid structure*. (Springer-Verlag, 1984).
- Ngo, H. B., Kaiser, J. T. & Chan, D. C. The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA. *Nat. Struct. Mol. Biol.* **18**, 1290–1296, <https://doi.org/10.1038/nsmb.2159> (2011).

40. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta* **1814**, 942–968, <https://doi.org/10.1016/j.bbapap.2010.10.012> (2011).
41. Rohs, R. *et al.* The role of DNA shape in protein–DNA recognition. *Nature* **461**, 1248–1253, <https://doi.org/10.1038/nature08473> (2009).
42. Hancock, S. P. *et al.* Control of DNA minor groove width and Fis protein binding by the purine 2-amino group. *Nucleic Acids Res.* **41**, 6750–6760, <https://doi.org/10.1093/nar/gkt357> (2013).
43. Stella, S., Cascio, D. & Johnson, R. C. The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. *Genes Dev.* **24**, 814–826, <https://doi.org/10.1101/gad.1900610> (2010).
44. Rubio-Cosials, A. *et al.* Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter. *Nat. Struct. Mol. Biol.* **18**, 1281–1289, <https://doi.org/10.1038/nmsb.2160> (2011).
45. Chakraborty, A. *et al.* DNA structure directs positioning of the mitochondrial genome packaging protein Abf2p. *Nucleic Acids Res.* **45**, 951–967, <https://doi.org/10.1093/nar/gkw1147> (2017).
46. Rice, P. A., Yang, S., Mizuuchi, K. & Nash, H. A. Crystal structure of an IHF–DNA complex: a protein-induced DNA U-turn. *Cell* **87**, 1295–1306 (1996).
47. Mouw, K. W. & Rice, P. A. Shaping the *Borrelia burgdorferi* genome: crystal structure and binding properties of the DNA-bending protein Hbb. *Mol. Microbiol.* **63**, 1319–1330, <https://doi.org/10.1111/j.1365-2958.2007.05586.x> (2007).
48. Swinger, K. K., Lemberg, K. M., Zhang, Y. & Rice, P. A. Flexible DNA bending in HU–DNA cocystal structures. *EMBO J.* **22**, 3749–3760, <https://doi.org/10.1093/emboj/cdg351> (2003).
49. Kamashev, D. & Rouviere-Yaniv, J. The histone-like protein HU binds specifically to DNA recombination and repair intermediates. *EMBO J.* **19**, 6527–6535, <https://doi.org/10.1093/emboj/19.23.6527> (2000).
50. Le Meur, R. *et al.* The nucleoid-associated protein HU enhances 8-oxoguanine base excision by the formamidopyrimidine–DNA glycosylase. *Biochem. J.* **471**, 13–23, <https://doi.org/10.1042/BJ20150387> (2015).
51. Gehring, A. M., Walker, J. E. & Santangelo, T. J. Transcription regulation in Archaea. *J. Bacteriol.* **198**, 1906–1917, <https://doi.org/10.1128/JB.00255-16> (2016).
52. Manzur, K. L. & Zhou, M. M. An archaeal SET domain protein exhibits distinct lysine methyltransferase activity towards DNA-associated protein MC1- α . *FEBS Lett.* **579**, 3859–3865, <https://doi.org/10.1016/j.febslet.2005.05.026> (2005).
53. Driessen, R. P. & Dame, R. T. Structure and dynamics of the crenarchaeal nucleoid. *Biochem. Soc. Trans.* **41**, 321–325, <https://doi.org/10.1042/BST20120336> (2013).
54. Dulmage, K. A., Todor, H. & Schmid, A. K. Growth-phase-specific modulation of cell morphology and gene expression by an archaeal histone protein. *mBio* **6**, e00649–00615, <https://doi.org/10.1128/mBio.00649-15> (2015).
55. Kvaratskhelia, M. *et al.* Holliday junction resolution is modulated by archaeal chromatin components *in vitro*. *J. Biol. Chem.* **277**, 2992–2996, <https://doi.org/10.1074/jbc.M109496200> (2002).
56. Loth, K., Landon, C. & Paquet, F. Chemical shifts assignments of the archaeal MC1 protein and a strongly bent 15 base pairs DNA duplex in complex. *Biomol. NMR Assign.* **9**, 215–217, <https://doi.org/10.1007/s12104-014-9577-8> (2015).
57. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696, <https://doi.org/10.1002/prot.20449> (2005).
58. de Vries, S. J., van Dijk, M. & Bonvin, A. M. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **5**, 883–897, <https://doi.org/10.1038/nprot.2010.32> (2010).
59. Dosset, P., Hus, J. C., Marion, D. & Blackledge, M. A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J. Biomol. NMR* **20**, 223–231 (2001).
60. van Dijk, M. & Bonvin, A. M. 3D-DART: a DNA structure modelling server. *Nucleic Acids Res.* **37**, W235–239, <https://doi.org/10.1093/nar/gkp287> (2009).
61. De Lano, W. L. *Pymol*. (De Lano Scientific, 2002).
62. Koradi, R., Billeter, M. & Wuthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14**, 51–55, 29–32;0263785596000094 [pii] (1996).
63. Tollinger, M., Skrynnikov, N. R., Mulder, F. A., Forman-Kay, J. D. & Kay, L. E. Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.* **123**, 11341–11352 (2001).
64. Bieri, M. & Gooley, P. R. Automated NMR relaxation dispersion data analysis using NESSY. *BMC Bioinformatics* **12**, 421, <https://doi.org/10.1186/1471-2105-12-421> (2011).

Acknowledgements

We wish to express our gratitude to Dr. Gérard Lancelot, now retired, who initiated the NMR RDC experiments. We thank Dr. Stéphane Bourge for helping us to minimize the structural model of the MC1–DNA^{23bp} complex.

Author Contributions

K.L. and F.P. designed the experiments, analyzed the data and wrote the paper. J.L. performed the CPMG experiments. F.Cu. conducted the EMSA study. J.L. and F.Co. purified MC1 and the oligonucleotides for preparing DNA^{23bp}. Together with the rest of the authors, B.C., A.F.D. and C.L. participated in discussion and contributed to manuscript writing.

Additional Information

Competing Interests: The authors declare 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) 2019

Chapitre 2. Vers une caractérisation aux rayons X du complexe ADN^{23bp}/MC1

Ce chapitre concerne l'étude du complexe ADN^{23bp}/MC1 par d'autres méthodes de biophysique : la diffusion des rayons X aux petits angles (SAXS) et la cristallographie aux rayons X (RX). Cette étude a nécessité l'optimisation des conditions expérimentales permettant de **stabiliser une conformation unique du complexe**.

2.1. Matériel et méthodes

2.1.1. Préparation des échantillons de complexe

Tout comme pour une étude structurale par RMN, la détermination de structure par cristallographie nécessite d'avoir une quantité importante de protéine avec un degré de pureté le plus proche possible des 100 %. L'obtention des échantillons de départ se fait, dans notre cas, en **trois phases** : la surexpression de la protéine d'intérêt, sa purification et sa complexation avec un duplex d'ADN (Figure 83).

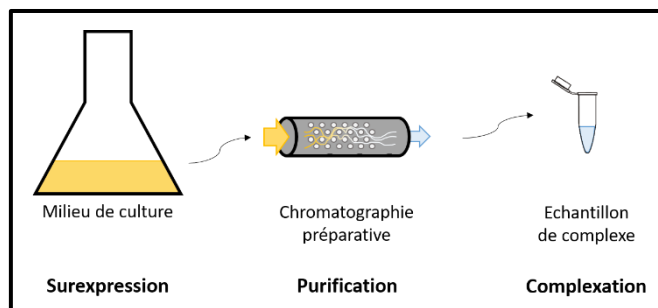


Figure 83 : Schéma des principales phases pour la préparation d'échantillons de protéine.

Préparation des complexes de ADN^{23bp}/¹⁵N-MC1

L'échantillon de protéines MC1 de type sauvage marquées ¹⁵N (¹⁵N-MC1) avait été préparé tel que présenté dans le matériel et méthodes du chapitre précédent (Partie 2.1.2.1).

Pour les essais de cristallographie, l'échantillon de protéine a été échangé pour un tampon F (20 mM Hepes pH7.6, 100 mM NaCl), concentré à 2,6 mM et stocké à 4 °C.

La phase de complexation de la protéine a été ensuite réalisée, comme pour les complexes de MC1t, pendant la thèse selon la méthode détaillée dans le paragraphe « Préparation des échantillons de MC1t : Complexation de la protéine » (p148).

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Préparation des échantillons de MC1t : Surexpression de la protéine

Pour l'expression et la purification de la protéine MC1t, nous avons utilisé le plasmide **pET28b-mc1** construit à partir du plasmide pET-28b(+) (pET pour « plasmid Expression vector under T7 control »). Il contient les éléments de base suivants : une origine de réplication, un gène de résistance à la Kanamycine, un gène régulateur et notre séquence d'intérêt placée entre deux sites de restriction (Figure 84).

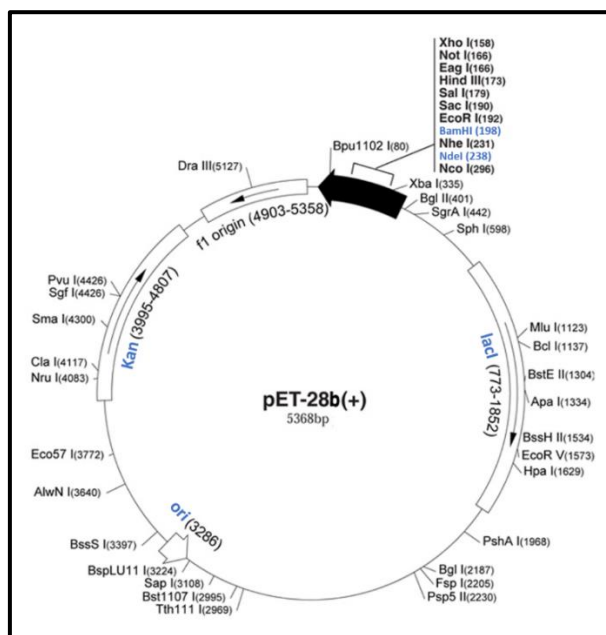


Figure 84 : Carte du plasmide pET-28b(+). Les éléments d'intérêt sont écrits en bleu.

L'**origine de réplication (ori)** permet d'initier la réplication du plasmide. Dans un premier temps, un complexe protéique spécifique se forme au niveau de cette séquence. Dans un second temps, ce complexe va permettre d'ouvrir l'ADN double-brin afin de commencer la réplication du plasmide.

Le **gène de résistance** à un antibiotique, ici à la Kanamycine (**Kan**), permet aux bactéries ayant intégré le plasmide, et donc cette séquence, de survivre en présence de l'antibiotique. Ces gènes de résistance codent le plus souvent des enzymes qui peuvent dégrader ou modifier l'antibiotique pour le rendre inactif (oxydation, etc.).

Le **gène régulateur (LacI)** précède l'opéron lactose. Il code pour une protéine régulatrice qui inhibe l'expression des gènes de l'opéron lactose en se fixant sur l'ADN au niveau de l'opérateur lac. Ce répresseur est exprimé quelles que soient les conditions de croissance de la bactérie. En revanche, en présence d'allolactose (un isomère du lactose), le répresseur se complexe au sucre et perd son affinité pour l'opérateur (Figure 85). L'opéron lactose peut ainsi être exprimé (induction). Cette activation peut également être réalisée en présence d'un réactif de biologie moléculaire spécifique : l'isopropyl β -D-1-thiogalactopyranoside (IPTG), un analogue de l'allolactose. Contrairement à

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

l'allolactose, l'IPTG n'est pas métabolisable : sa concentration ne varie pas dans la culture cellulaire ce qui rend l'induction par l'IPTG constante et stable.

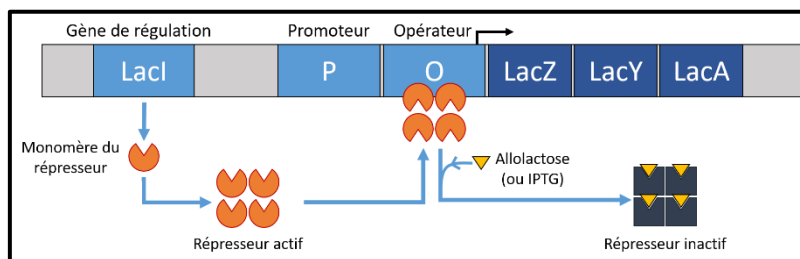


Figure 85 : Régulation de la transcription de l'opéron lactose.

La **séquence d'intérêt** est placée entre les sites de restriction *NdeI* et *BamHI*. Elle permet d'exprimer la protéine MC1t fusionnée avec une étiquette « poly-Histidines » et un site de clivage. Il est connu que le greffage d'une étiquette de type « poly-Histidines », ici de six histidines, en N ou en C terminus d'une protéine permet de simplifier les protocoles de purification grâce à l'utilisation d'une résine d'affinité. L'étiquette peut ensuite être retirée par l'action d'une enzyme, ici la thrombine, si un site de clivage, ici « **LVPRGS** », a été prévu dans la construction du plasmide entre l'étiquette et la protéine d'intérêt (Figure 86).

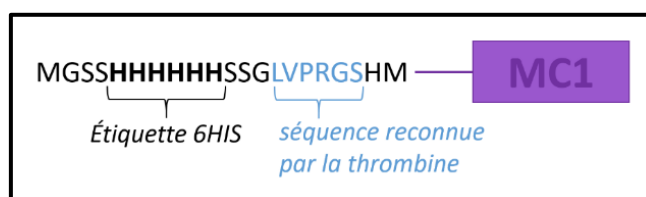


Figure 86 : Construction de la protéine MC1t fusionnée avec une étiquette « 6HIS » et exprimée par le gène d'intérêt.

La surexpression des protéines se fait en **trois étapes** illustrées sur la Figure 87 : une transformation bactérienne, une surproduction des bactéries et une induction du milieu par ajout d'IPTG.

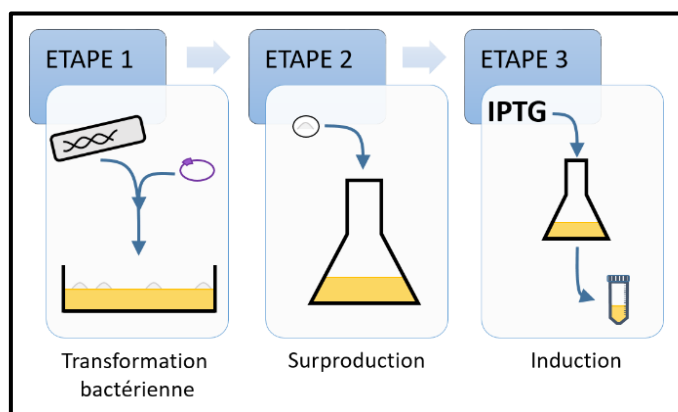


Figure 87 : Principales étapes pour la surexpression de protéines.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

ETAPE 1 : Transformation bactérienne

Pour la transformation bactérienne, 50 μL de bactéries calcium-compétentes (BL21-DE3) sont incubées avec 100 ng du plasmide pET28b-mc1 pendant 30 minutes dans la glace. Les BL21-DE3 sont des bactéries rendues compétentes à l'incorporation d'ADN par un traitement au chlorure de calcium. Un choc thermique est ensuite appliqué en incubant les bactéries 45 secondes à 42 °C. Cette étape permet, en synergie de l'action du calcium présent (toujours sous forme CaCl_2), une grande fragilisation des parois bactériennes. Ces deux actions facilitent l'entrée des plasmides dans les bactéries. Les bactéries sont ensuite rapidement remises dans la glace afin de refortifier les parois. Au bout de 5 minutes, l'échantillon est enrichi avec 1 mL de milieu de culture et placé sous agitation (160 rpm à 37 °C pendant 1 h). Le milieu de culture utilisé est le LB, un milieu nutritif standard (10 g.L^{-1} Tryptone, 5 g.L^{-1} extrait de levures, 5 g.L^{-1} NaCl) pour la croissance de bactéries *E. coli*.

Après l'heure d'incubation, 20 μL de culture non diluée ou diluée au dixième sont étalés sur des boîtes de Pétri contenant du LB/Agar supplémenté de Kanamycine (50 $\mu\text{g.mL}^{-1}$). Les boîtes sont ensuite incubées à 37 °C pendant la nuit.

Les colonies de bactéries ainsi obtenues correspondent à des amas de bactéries pouvant appartenir à deux catégories et qui peuvent coexister sur les boîtes :

- Les bactéries ayant intégré le plasmide et contenant donc un gène de résistance à l'antibiotique du milieu. Les colonies formées se présentent sous la forme de disque blanc d'environ 1 mm avec un contour régulier ;
- Les bactéries résistantes à la Kanamycine mais dépourvues du plasmide d'intérêt. Les colonies formées sont très proches les unes des autres et présentent des tailles et des formes irrégulières. La résistance des bactéries à l'antibiotique est liée au relargage des bactéries de colonies voisines qui, elles, possèdent le plasmide et partagent leurs défenses dans leur environnement proche.

Pour ensemer des cultures plus importantes et réaliser l'étape de surproduction, une seule colonie ayant intégré le plasmide est nécessaire.

ETAPE 2 : Surproduction

Les surproductions des protéines fusionnées avec une étiquette « 6HIS » (6HIS-MC1t) dans *E. coli* se font de manière analogue à celles utilisées pour les protéines sauvages. On réalise une première pré-culture à partir de la colonie sélectionnée à l'étape précédente, en la plaçant dans 1 mL de milieu LB supplémenté de Kanamycine (50 $\mu\text{g.mL}^{-1}$). Après 8 h d'incubation à 37 °C sous agitation (160 rpm), le milieu de culture contenant les bactéries est trouble et ces dernières sont collectées par centrifugation (10000 rpm à 4 °C pendant 5 min). Les culots sont resuspendus dans un

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

erlenmeyer contenant 150 mL de LB supplémenté de Kanamycine ($50 \mu\text{g.mL}^{-1}$). La seconde pré-culture est laissée à 37°C sous agitation pendant la nuit (160 rpm à 37°C , OVN).

La seconde pré-culture est utilisée pour ensemençer les cultures de travail, soit 4 x 500 mL de milieu LB supplémenté de Kanamycine ($50 \mu\text{g.mL}^{-1}$), à une turbidité comprise entre 0,1 $\text{UA}_{600\text{nm}}$ et 0,15 $\text{UA}_{600\text{nm}}$. Pour ce faire, l'absorbance à 600 nm de la pré-culture, devenue trouble, est mesurée diluée au dixième. Le volume nécessaire à l'obtention de la turbidité de départ dans nos cultures de travail est ensuite centrifugé (4000 rpm à 4°C pendant 20 min). Les culots sont resuspendus dans les cultures de travail et les absorbances de départ sont mesurées avant de placer les cultures sous agitation (160 rpm à 37°C). L'absorbance est mesurée ponctuellement afin de suivre la croissance des bactéries et de repérer leur phase de croissance exponentielle. C'est au cours de cette phase que les bactéries se multiplient le plus rapidement et que le nombre potentiel de cellules pouvant synthétiser la protéine est le plus grand.

ETAPE 3 : Induction

Entre 0,8 et 1,0 $\text{UA}_{600\text{nm}}$, soit après 3 à 4 h de croissance, la population de bactéries est en fin de croissance exponentielle. L'expression est à ce moment-là induite par ajout de 1 mM final d'IPTG. Après 2 h d'induction, les bactéries sont collectées par centrifugation (5000 rpm - rotor SLA-3000 - à 4°C pendant 45 min). Les culots sont ensuite resuspendus dans du tampon de lavage frais (à 4°C), transférés dans des tubes Falcon de 50 ml et centrifugés de nouveau (4000 rpm à 4°C pendant 30 min). Ce tampon permet de retirer les protéases et les toxines excrétées par les cellules tout en maintenant une salinité suffisante pour ne pas déstabiliser les parois bactériennes. Différents tampons de lavages peuvent être utilisés. Ce choix dépend du protocole de purification des protéines. Dans notre cas, la purification nécessite l'utilisation d'une résine d'affinité. Afin de ne pas dégrader cette dernière, le tampon de lavage utilisé doit être dépourvu d'inhibiteurs de protéases ou de tout autre additif contenant des chélateurs tel que l'éthylènediaminetétraacétique (EDTA), ou de forts agents réducteurs tels que le dithiothréitol (DTT) ou le β -mercaptoéthanol (β mercapto). Nous avons ici choisi d'utiliser du LB frais pour laver les cellules. Une nouvelle fois le surnageant est éliminé. Les culots humides sont pesés et conservés à -20°C .

Préparation des échantillons de MC1t : Purification de la protéine

Comme brièvement introduit plus tôt, l'étiquette d'affinité présente un atout majeur pour la purification de protéines. Elle permet des purifications bien plus rapides (possiblement en une seule étape), efficace et avec de très bons rendements. Ici, la purification est réalisée en **six étapes** car l'étiquette de la protéine doit lui être retirée (Figure 88).

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

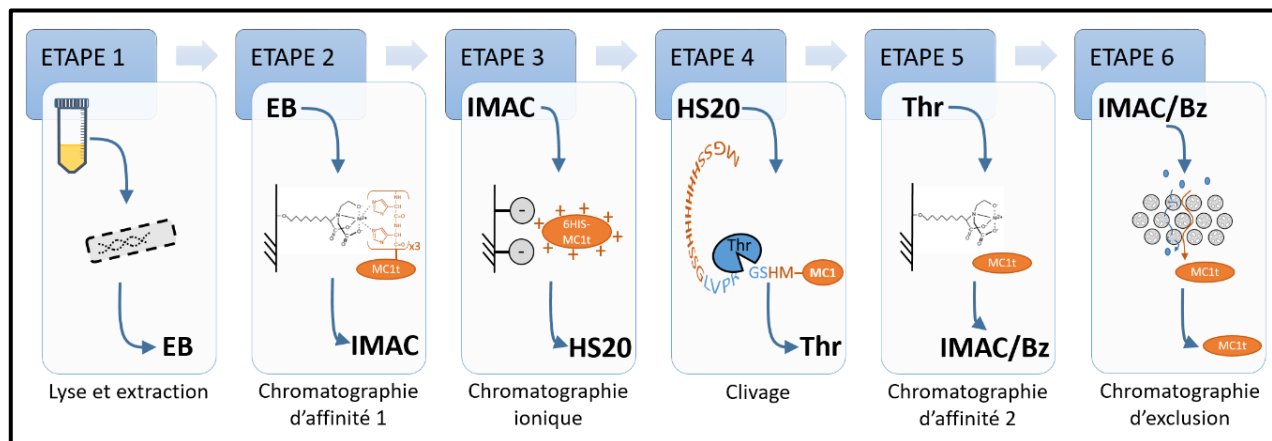


Figure 88 : Schéma des différentes étapes de purification de la protéine MC1t.

ETAPE 1 : Lyse et Extraction

La **première étape** est une **étape de lyse et d'extraction**. Les culots bactériens sont repris dans un tampon de lyse (20 mM Hepes pH7.6, 1 M NaCl, 5 mM imidazole, 1 mM PMSF) à raison de 10 mL de tampon par gramme de culot bactérien. La suspension cellulaire est ensuite supplémentée par du lysozyme (1 mg.mL⁻¹ final), permettant de détruire les parois bactériennes. Après 30 min d'incubation au bain marie (à 30 °C), on réalise 3 à 5 cycles de congélation (azote liquide) / décongélation (bain-marie à 30 °C) suivi de 5 à 8 cycles de sonication (1 min [5" pulse:5" repos], 40 % puissance sur sonicateur 750 W équipé d'une sonde 13 mm). La suspension passe d'abord par une phase visqueuse, correspondant à la libération de l'ADN, puis redevient fluide lorsque l'ADN est suffisamment fractionné. Le lysat obtenu est centrifugé (12500 rpm - rotor SS34 - à 4 °C pendant 1 h) et le surnageant est ensuite récupéré pour l'étape suivante. Il constitue l'**extrait brut** (« EB »).

ETAPE 2 : Chromatographie d'affinité n°1

Pour la **seconde étape**, avons utilisé une colonne de **chromatographie d'affinité sur ions métalliques divalents immobilisés (IMAC)**. Les ions métalliques divalents sont fixés sur la colonne (phase stationnaire) et interagissent avec les noyaux imidazoles de l'étiquette « 6HIS » de la protéine éluée, permettant à cette dernière d'être spécifiquement retenue sur la colonne. Le passage d'une solution concentrée en imidazole (phase mobile) permet dans un second temps d'éluer 6HIS-MC1t. Cet échange est possible par compétition entre les noyaux imidazoles libres et ceux de l'étiquette « 6HIS » liés ioniquement à la phase stationnaire.

Une colonne de **chromatographie d'affinité sur ions nickels immobilisés** (Ni-NTA Agarose, Qiagen) de 5 mL a été utilisée. Ce type de colonne est moins spécifique qu'une colonne Cobalt, mais plus facile à régénérer et présente une grande capacité de fixation. La protéine 6HIS-MC1t se fixant aussi bien sur l'une ou l'autre colonne, la colonne Nickel a, de par sa capacité de fixation plus élevée, été retenue dans le protocole optimisé.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Après avoir équilibré la colonne dans un tampon équivalent au tampon de lyse (=Tp1), l'extrait brut y est chargé puis lavé avec environ 10 volumes de colonne (CV) de Tp1. Un second lavage est effectué avec 10 CV d'un mélange 90:10 Tp1/Tp2 (Tp2 = 20 mM Hepes pH7.6, 1 M NaCl, 200 mM imidazole, 1 mM PMSF). Le second lavage permet d'éluer les composés qui se seraient fixés de façon non spécifique sur la colonne. La protéine d'intérêt est ensuite éluée en mode isocratique avec 100 % de Tp2. Ces étapes sont suivies par spectroscopie UV à une absorbance de 280 nm. L'imidazole absorbe également à 280 nm mais il n'est pas nécessaire de soustraire son signal pour évaluer quels autres composés sont élués.

ETAPE 3 : Chromatographie échangeuse d'ions

Pour la **troisième étape**, une colonne de **chromatographie fortement échangeuse de cations** (POROS HS/20 10 mm D x 100 mm L PEEK) de 7,9 mL est utilisée. Cette étape permet de s'assurer que l'échantillon est pur et homogène avant le clivage de l'étiquette « 6HIS ». La fraction « IMAC » précédente a été diluée trois fois dans un tampon dépourvu de sels et d'imidazole, Tp3 (20 mM Hepes pH7.6), afin de diminuer la force ionique globale (~160 mM NaCl final) de l'échantillon. Elle est ensuite chargée sur la colonne préalablement équilibrée dans un mélange 60:40 Tp3/Tp4 (Tp4 = Tp3 + 1 M NaCl). La protéine est éluée par un gradient linéaire sur 50 CV jusqu'à un mélange 10:90 Tp3/Tp4. Les fractions de protéine éluée ont été rassemblées pour l'étape suivante et constituent la fraction globale « **HS20** ».

ETAPE 4 : Clivage de l'étiquette 6HIS

La **quatrième étape** correspond au **clivage de l'étiquette « 6HIS » par la thrombine (Thr)**. Pour cette étape, on utilise un kit de clivage à la thrombine commercial (Novagen).

Pour cette étape, la fraction « HS20 », issue de l'**ETAPE 3**, est concentrée jusqu'à ~5 mg.mL⁻¹ et son tampon est échangé par centricon Amicon (3 kDa), pour un tampon moins concentré en sel, le Tp5 (20 mM Hepes pH7.6, 500 mM NaCl). La thrombine est, en parallèle, diluée au 200^{ième} avec un tampon de dilution (50 mM citrate de sodium pH6.5, 200 mM NaCl, 0,1 % PEG-8000, 50 % glycérol) fourni dans le kit (*Thrombin Dilution/Storage Buffer*).

L'échantillon de clivage est préparé pour un volume final ($V_{\text{tot}} = Q \times 30 \mu\text{L}$) lié à la quantité de protéine à cliver ($Q/100 \mu\text{g}$) avec : $Q \times 100 \mu\text{g}$ de 6HIS-MC1t (~3 mg.mL⁻¹ final) ; $Q \times 3 \mu\text{L}$ de tampon de clivage 10X (200 mM Tris/HCl pH8.4, 1,5 M NaCl, 25 mM CaCl₂) ; et $Q \times 0,6 \mu\text{L}$ de thrombine préalablement diluée au 200^{ième}. Le volume final, V_{tot} , est ajusté avec de l'eau ultrapure. L'échantillon est ensuite placé à 4 °C sous agitation pendant la nuit (35 rpm à 4 °C).

Le produit de digestion « **Thr** » est obtenue à la fin du temps de clivage.

ETAPE 5 : Inhibition de la thrombine et Chromatographie d'affinité n°2

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

La **cinquième étape** permet d'**inhiber l'action de l'enzyme** et de **purifier la protéine clivée**. Pour ce faire, le produit de digestion « Thr » est d'abord passé sur une pré-colonne de benzamidine de 1 mL. La benzamidine est un inhibiteur de la trypsine et des protéases à sérine, telle que la thrombine. C'est elle qui va stopper l'étape de clivage. Cette pré-colonne est placée en amont de la colonne d'affinité sur ions métalliques de l'**ETAPE 2**. Les deux colonnes sont préalablement équilibrées avec le Tp5. Après la charge, le système est lavé avec 10 CV du Tp5. En théorie et comme expliqué à l'**ETAPE 2**, la phase de lavage doit suffire à faire éluer la protéine dépourvue de son étiquette et ne retenir que la portion non clivée de protéine. En effet, sans l'étiquette « 6HIS » la protéine perd son affinité pour la colonne. Expérimentalement, il a été nécessaire d'ajouter une grande quantité d'imidazole (100 mM) pour éluer MC1t. Même dépourvue de son étiquette, il semble que la protéine reste très affine pour la colonne. La protéine clivée est ainsi élue pendant le lavage, avec 100 % de Tp2 (soit 200 mM imidazole) en une fraction « **IMAC/Bz** ».

ETAPE 6 : Chromatographie par filtration sur gel

La **sixième et dernière étape** est une étape de **chromatographie par filtration sur gel (chromatographie d'exclusion stérique)**. Elle permet d'obtenir un échantillon contenant la protéine pure et sans agrégats, en séparant les molécules par leur volume hydrodynamique. La fraction « IMAC/Bz », élue précédemment, est concentrée et le tampon est échangé pour le Tp5 jusqu'à un volume de 1 mL par centricon Amicon (3 kDa). Elle est ensuite chargée sur une colonne de chromatographie par filtration sur gel (Superdex S75) de 120 mL préalablement équilibrée avec le Tp4. L'élution se fait en mode isocratique avec 100 % Tp4. Les fractions contenant MC1t sont ensuite rassemblées, concentrées et le tampon est échangé pour le tampon F (20 mM Hepes pH7.6, 100 mM NaCl) par centricon Amicon (3 kDa).

Préparation des échantillons de MC1t : Complexation avec l'ADN^{23bp}

La cristallogénèse est fonction de nombreux paramètres. Pour obtenir des structures très haute résolution par diffraction des rayons X, l'un des objectifs serait d'avoir la meilleure affinité possible entre la protéine et l'ADN et de favoriser un empilement plus stable des complexes pour la formation des cristaux.

Les ADN sélectionnés pour réaliser les études structurales et dynamiques par RMN ont été choisis afin d'avoir un minimum de recouvrement entre les signaux ¹H de l'ADN et permettre ainsi leur attribution. Ce problème est inexistant lorsque les études structurales sont réalisées par diffraction des rayons X. Nous sommes ainsi repartis de la séquence ADN^{23bp} choisie pour la construction du modèle par RMN et avons changé la paire de bases AT en position 15 pour la paire de bases CG.

Afin de stabiliser l'empilement des doubles brins d'ADN, deux autres duplex à « bords cohésifs » ont également été retenus. Ils correspondent au duplex précédent mais présentent un décalage au niveau de la séquence de leur brin inférieur (Figure 89). L'étude a ainsi été menée sur le duplex **D₁**

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

dit « bout franc » composé du brin SUP^{23bp} et du brin INF₁^{23bp}, sur le duplex **D₂** ayant un décalage d'une base vers le côté 5' du brin inférieur (SUP^{23bp}/ INF₂^{23bp}) et le duplex **D₃** ayant un décalage d'une base vers le côté 3' sur le brin inférieur (SUP^{23bp}/ INF₃^{23bp}).

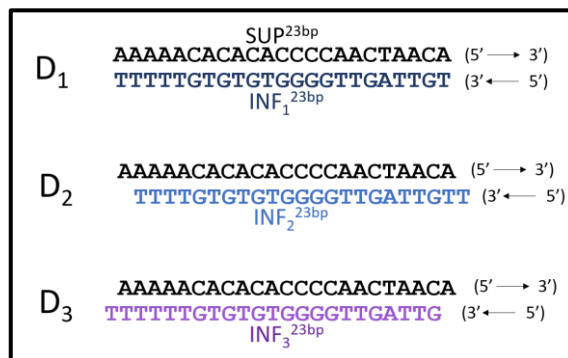


Figure 89 : Illustration des différents duplex D_i d'ADN utilisés pour la caractérisation aux rayons X.

La préparation des complexes de MC1t avec les différents duplex se fait en **3 étapes** (Figure 90). D'abord les oligonucléotides de synthèse sont purifiés, puis hybridés en duplex. Enfin, les duplex sont complexés avec la protéine pour l'étude de cristallographie.

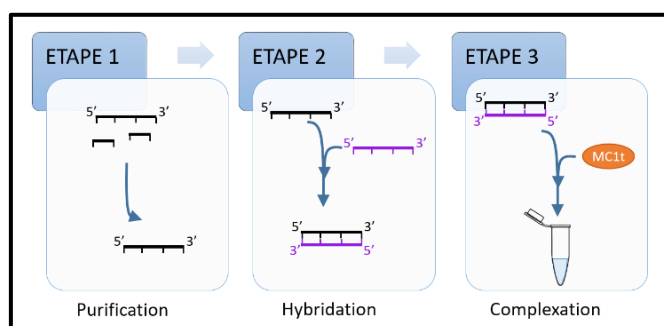


Figure 90 : Schéma des différentes étapes de préparation des complexes d'ADN^{23bp} avec la protéine MC1t.

ETAPE 1 : Purification des oligonucléotides

Les oligodéoxyribonucléotides synthétiques de 23 paires de bases illustrés sur la Figure 89, ont été achetés chez Eurogentec avec une pureté estimée à 70 % (SePOP - Selective Precipitation Optimised Process - oligonucleotide quality). Cette pureté n'étant pas suffisante pour des essais de cristallogénèse, une étape de purification supplémentaire est réalisée par chromatographie échangeuse d'anions.

Après avoir été solubilisés dans de l'eau ultrapure, Les échantillons d'oligonucléotide sont d'abord élués sur une colonne de **chromatographie échangeuse d'anions** (MonoQ 10/100GL) de 8 mL préalablement équilibrée dans un tampon A + 20 % d'un tampon B (TpA : 10 mM NaOH, 50 mM NaCl et TpB : 10 mM NaOH, 1 M NaCl) soit 240 mM NaCl final. La méthode utilisée pour le brin

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

supérieur consiste en un plateau initial avec 30 % TpB puis un gradient de 30 % à 60 % de TpB. Pour les brins inférieurs, le plateau initial est de 60 % de tampon B et le gradient de TpB varie de 60 % à 90 %.

Les fractions éluées ont ensuite été dessalées par gravité en **chromatographie de phase inverse** (Sep-pack C₁₈). Pour ce faire, les fractions sont diluées par deux dans un tampon D (TpD : 100 mM Tris/HCl pH7.5, 10 mM AcOH). Chaque fraction diluée a été chargée sur la colonne préalablement équilibrée dans l'eau ultrapure et lavée par 3 CV d'eau ultrapure. Elle a ensuite été éluée par 1/10 CV de 80:20 ACN/H₂O suivi de 1/5 CV 100 % d'acétonitrile (ACN).

Les fractions collectées ont été séchées avant d'être ressolubilisées dans 500 µL de tampon final F (TpF : 20 mM Hepes pH7.6, 100 mM NaCl). Les concentrations des échantillons finaux d'oligonucléotides ont été déterminées par la mesure de l'absorbance à 260 nm et ces derniers sont stockés à +4 °C.

ETAPE 2 : Hybridation en duplex d'ADN

Les oligonucléotides sont ensuite **hybridés** en double brin d'ADN pour réaliser des échantillons de duplex D_i à 4 mM dans un tampon H (TpH : 5 mM CaCO₃/Na⁺ pH6.5, 0,2 M KCl, 0,5 mM EDTA). Ces échantillons sont tous préparés avec des ratios de 1,0 mol du brin SUP_{RX}^{23bp} pour 1,1 mol du brin INF_i^{23bp}. Pour ce faire, on prélève un volume correspondant à la quantité d'oligonucléotide SUP_{RX}^{23bp} nécessaire pour le duplex (soit 76,4 µL) et on l'ajoute à un volume de brin INF_i^{23bp}, correspondant à 1,1 fois la quantité de brin supérieur (soit 184,9 µL de INF₁^{23bp}, 142,9 µL de INF₂^{23bp}, ou 171,9 µL de INF₃^{23bp}). Le mélange est ensuite séché et repris dans 100 µL de tampon H. Cet échantillon est chauffé 2 minutes au bain-marie à 80 °C (Tm^{oligo} = 73,6 °C) pour permettre à l'hybridation entre les brins supérieurs et inférieurs de se faire. Puis le bain-marie est placé à 4 °C de façon à assurer une descente en température de l'échantillon relativement lente et permettre une hybridation totale de l'échantillon de duplex. Les différents échantillons de double brin d'ADN (« D_i », où i correspond à INF_i^{23bp}) sont ensuite conservés à 4 °C.

ETAPE 3 : Complexation

Les **complexes ADN/protéine** notés C_i, où i correspond au numéro du brin inférieur du duplex D_i utilisé, ont été préparés à des concentrations comprises entre 1,0 mM et 1,5 mM dans le Tampon F (20 mM Hepes pH7.6, 100 mM NaCl) en respectant un ratio 1,1:1,0 où la proportion en duplex d'ADN est majoritaire par rapport à celle de la protéine. Ce ratio permet de s'assurer qu'un maximum de protéine MC1t est impliqué dans un complexe ADN/protéine sans que cela ne consomme trop de duplex. Les échantillons de complexe sont placés 1 h à 4 °C puis 30 min à température ambiante avant d'être utilisés pour les essais suivants ou stockés à 4 °C.

2.1.2. Cristallographie aux rayons X de protéines

La cristallographie aux rayons X est une technique basée sur la diffraction des rayons X sur un cristal de biomolécule et permettant d'observer l'organisation tridimensionnelle de cette biomolécule.

Principe de la cristallogenèse

La cristallogenèse correspond à la cristallisation de macromolécules biologiques. Cette cristallisation est une transition de phase. La macromolécule passe d'une phase soluble à une phase solide appelée cristal. Cette évolution de la macromolécule, liée aux variations de son environnement, peut être décrite sous la forme d'un diagramme de phases (Figure 91). Sur ce diagramme, la courbe de solubilité représente la limite où cristal et solution sont en équilibre thermodynamique. Pour que la molécule cristallise, il lui faut dépasser cette courbe de solubilité et passer à un **état sursaturé**. Il s'en suit trois scénarios possibles. Dans le premier, la sursaturation est trop forte et la macromolécule précipite (**zone de précipitation**). Dans le second, la sursaturation est trop faible et les premiers germes de cristaux sont incapables de se former (**zone métastable**). Dans le troisième et meilleur scénario, la sursaturation est suffisante. La macromolécule se trouve alors dans un état de sursaturation qui lui permet d'initier sa cristallisation (**zone de nucléation**) et de former des germes de cristaux (= **nucléation**). La formation de ces germes consomme de la macromolécule et le système passera dans la **zone métastable**. Dans cette zone métastable, seule la croissance cristalline peut avoir lieu. Les cristaux y croissent jusqu'à l'obtention d'un équilibre entre la concentration en agent de cristallisation dans la goutte et dans le puits.

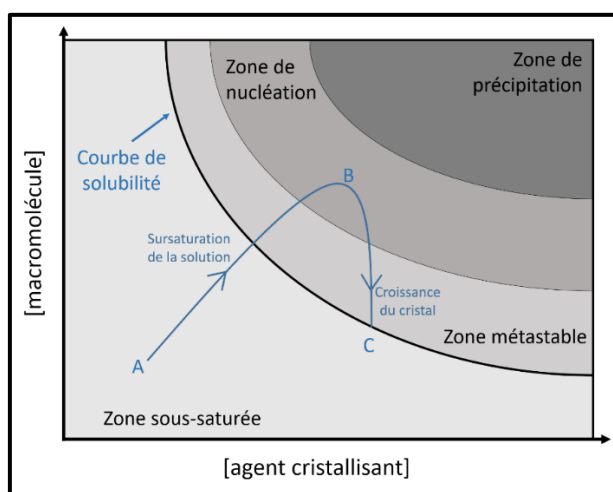
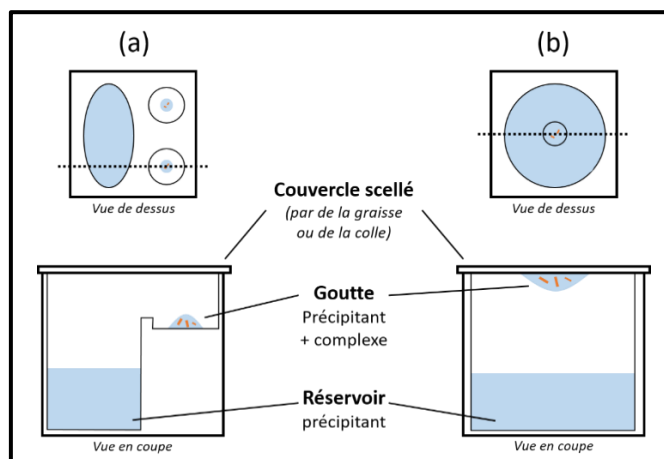


Figure 91 : Diagramme de phase à deux dimensions d'une macromolécule en fonction de sa propre concentration et de la concentration d'un agent cristallisant (adapté de Sauter et Giegé, 2001) ^[184]. A = macromolécule en solution. B = nucléation du cristal. C = cristal.

Différentes méthodes ont été développées pour la cristallisation des macromolécules biologiques. Parmi les plus employées, on retrouve la technique du « batch » qui consiste à placer la macromolécule directement en condition de sursaturation pour permettre une nucléation plus rapide.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Il existe aussi les méthodes de dialyse, de diffusion libre à l'interface ou de diffusion de vapeur. Ces techniques, moins brutales que le « batch », sont basées sur une augmentation progressive de la concentration d'agent cristallisant. La méthode de **diffusion de vapeur**, simple à mettre en place et adaptable à des volumes réduits (notamment par l'utilisation de robot de cristallisation comme le Mosquito), est illustrée sur la [Figure 92](#). Elle peut être réalisée de deux façons différentes : par un système de goutte assise ou par un système de goutte suspendue.



[Figure 92](#) : *Illustration de cristallogenèse par la méthode d'échange de vapeur (a) goutte assise et (b) goutte suspendue.*

La technique de **diffusion de vapeur** nécessite l'utilisation d'un **système clos**. Celui-ci est constitué d'une lamelle de verre scellée par un joint de graisse ou de colle sur un puits. Dans chacun des cas, une **goutte de solution** (de volume V) contenant la macromolécule, le tampon et l'agent cristallisant (=précipitant) est mise à équilibrer contre un **réservoir** (de volume $\sim 10 \times V$) contenant l'agent cristallisant à une concentration plus élevée et le tampon. L'équilibration se fait ensuite par **évaporation des composés volatils** (comme l'eau) de la goutte vers le réservoir, jusqu'à ce que les tensions de vapeur à la surface des deux compartiments soient identiques. Durant cette équilibration, la composition du réservoir est considérée comme constante : le volume du réservoir étant bien plus grand que celui de la goutte. Au niveau de la goutte en revanche, le **processus d'équilibration concentre l'agent cristallisant et la macromolécule**. Cela diminue la solubilité de la macromolécule et, dans le cas de conditions favorables (sursaturation suffisante), l'amène à cristalliser (nucléation et croissance).

Diffraction des RX

Une fois les cristaux de macromolécules obtenus, la diffraction des rayons X est utilisée pour « cartographier » la maille cristalline d'un **monocristal**. En effet, un cristal correspond à un arrangement périodique des macromolécules dans les trois dimensions de l'espace, et cet arrangement, placé dans un faisceau de rayons X, est à l'origine du phénomène de diffraction observé sur les **clichés de diffraction** ([Figure 93](#)). Ce phénomène est possible car la distance entre

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

deux atomes est du même ordre de grandeur que la longueur d'onde des rayons X : l'échelle de l'Angström.

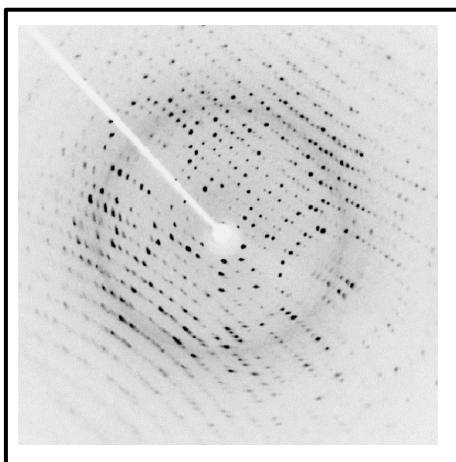


Figure 93 : Cliché de diffraction aux rayons X d'un cristal de lysozyme.

Pour les essais de diffractions, on ajoute dans un premier temps un cryoprotectant au cristal sélectionné, afin d'éviter qu'il ne soit détérioré lors de sa congélation. Le cristal est, dans un second temps, monté sur un lasso lui-même fixé sur une tête goniométrique. La tête goniométrique va permettre de tourner le cristal au niveau du goniomètre et ainsi de faire varier l'angle du faisceau incident θ (angle de Bragg). Après avoir gelé le cristal dans un bain d'azote liquide, celui-ci est monté sur le goniomètre et soumis à un flux d'azote à 100 K grâce à une canne de réfrigération. Ce système, illustré sur la [Figure 94](#), permet principalement de limiter les dommages dues radiations sur le cristal.

Une fois le cristal centré, il est soumis au faisceau de rayons X pour différents angles θ . Pour chaque angle, un cliché de diffraction est enregistré et sur chacun de ces clichés on observe des **taches de diffractions**. Les taches de diffractions dépendent de la nature des atomes et de leurs positions dans la maille et permettent de caractériser entièrement le motif cristallin.

Expérimentalement, la résolution des données de diffraction enregistrées, découle d'un traitement numérique des positions et des intensités des différentes taches de diffraction. Des programmes informatiques calculent une **carte de densité électronique**, liée à la **phase** et au facteur de structure, qui est peu à peu affinée, jusqu'à obtenir la structure tridimensionnelle complète de la macromolécule.

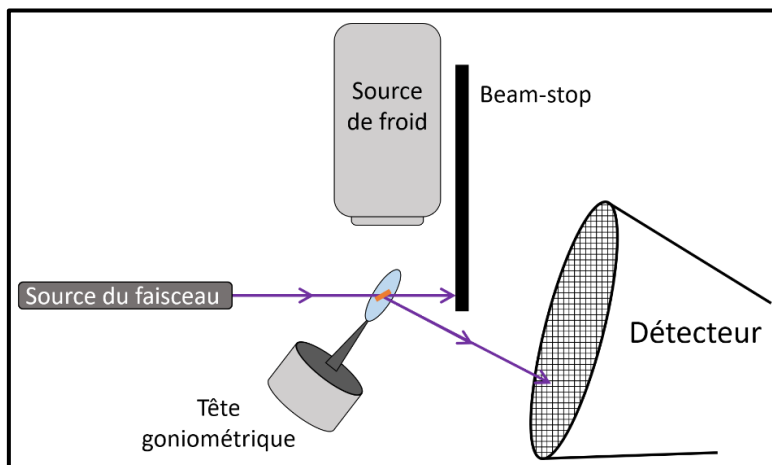


Figure 94 : *Schéma simplifié du dispositif de diffraction des rayons X.*

Préparation des échantillons et réalisation des expériences

Afin de trouver rapidement des conditions dans lesquelles le complexe ADN^{23bp}/MC1 cristallise, plusieurs **tests de criblages** ont été effectués. Pour ces essais, différentes solutions de kits de cristallogénèse ont été utilisés. La composition de l'ensemble des solutions de ces kits est présentée en Annexe 11 à Annexe 16.

Les tests de criblage ont été réalisés à 20 °C avec un système de goutte assise de 200 nL (contre 200 µL dans le réservoir) :

- sur les trois constructions de complexe **D_i/15N-MC1 à 0,5 mM final** avec les kits JCSG+ de Molecular Dimensions, NUCLEIX et CLASSICS de chez Qiagen ;
- puis sur les trois constructions de complexe **D_i/MC1t (C_i) à 0,6 mM final** avec les kits WIZARD de Emerald Biostructures, CRYO de Sigma, NUCLEIX et COMPAS de Qiagen.

Les **essais d'optimisation** ont ensuite été réalisés avec un système de goutte suspendue de 1,6 µL ou de 2,4 µL (contre 750 µL dans le réservoir).

Pour les **essais de diffraction**, les monocristaux de complexe ont d'abord été plongés dans une solution cryoprotectante (20 % glycérol final). Les expériences préliminaires de diffraction des rayons X sur les monocristaux, ont été menées à l'European Synchrotron Radiation Facility (ESRF, Grenoble, France) sur la ligne de lumière ID29.

Un premier jeu de données, correspondant aux différentes images de diffraction sur un cristal du complexe D₃/MC1t (C₃), a été collecté à 100 K à une résolution de 6,28 Å avec le logiciel MXCuBE ^[185]. Les réflexions de chaque image du jeu de données ont été traitées par le logiciel XDS ^[186] puis analysés par le logiciel CCP4 ^[187].

2.1.3. Diffusion de rayons X aux petits angles (SAXS)

Principe de la méthode

La diffusion des rayons X aux petits angles permet d'étudier des structures de molécules ou d'assemblages moléculaires dans leurs conditions quasi-natives ^[188]. Il s'agit d'une technique de basse résolution (autour de 20 Å) permettant de caractériser l'enveloppe d'une macromolécule par l'interaction d'un faisceau de rayons X avec les couches électroniques internes des atomes de la macromolécule en produisant une déviation de ce faisceau. Elle apporte principalement des informations sur la forme globale du système, sur son état d'oligomérisation, sur son rayon de giration et sur la présence éventuelle de différentes espèces en solution. Cette technique fournit ainsi des informations structurales complémentaires à la RMN ou à la cristallographie.

Pour ce faire, l'échantillon de macromolécule pure est élué et le flux d'échantillon est soumis au faisceau de rayons X au niveau d'un capillaire pendant que des clichés de diffusion sont enregistrés en continu. Ces clichés (2D) sont ensuite transformés informatiquement en courbes de diffusion (1D). Elle représente le logarithme de l'intensité diffusée, $\text{Log}(I(q))$, par rapport au vecteur de diffusion, q (équation (19)), et correspond à une représentation 1D de l'enveloppe de la macromolécule (Figure 95). L'analyse de cette courbe (analyse de Guinier, analyse de distribution des distances par paires (= analyse de Porod)) permet ensuite de déterminer la taille de la macromolécule, sa forme, ainsi que son degré d'oligomérisation grâce à différents paramètres (masse molaire MW, rayon de giration R_g , intensité à l'origine $I(0)$, dimension maximale D_{max} , polydispersité) ^[189]. Elle permet également d'identifier le désordre de la macromolécule en évaluant qualitativement son état de repliement et sa flexibilité (représentation de Katky).

Ces paramètres sont ensuite utilisés dans différents programmes informatiques de calcul issus la suite de programmes d'ATSAS ^[190] (GNOM ^[191], DAMMIN ^[192], MONSA ^[192], CRY SOL ^[193]).

$$\text{Vecteur de diffusion } q \text{ (}\text{\AA}^{-1}\text{)} \quad q = \frac{4\pi}{\lambda} \sin \theta \quad (19)$$

$$\text{Loi de Guinier} \quad \ln(I(q)) = \ln(I(0)) - \frac{R_g^2 q^2}{3} \quad (20)$$

$$\text{Loi de Porod} \quad V = I(0) \frac{2\pi^2}{q} \quad (21)$$

Où λ correspond à la longueur d'onde (Å) du faisceau de rayons X et θ au double de l'angle de diffusion.

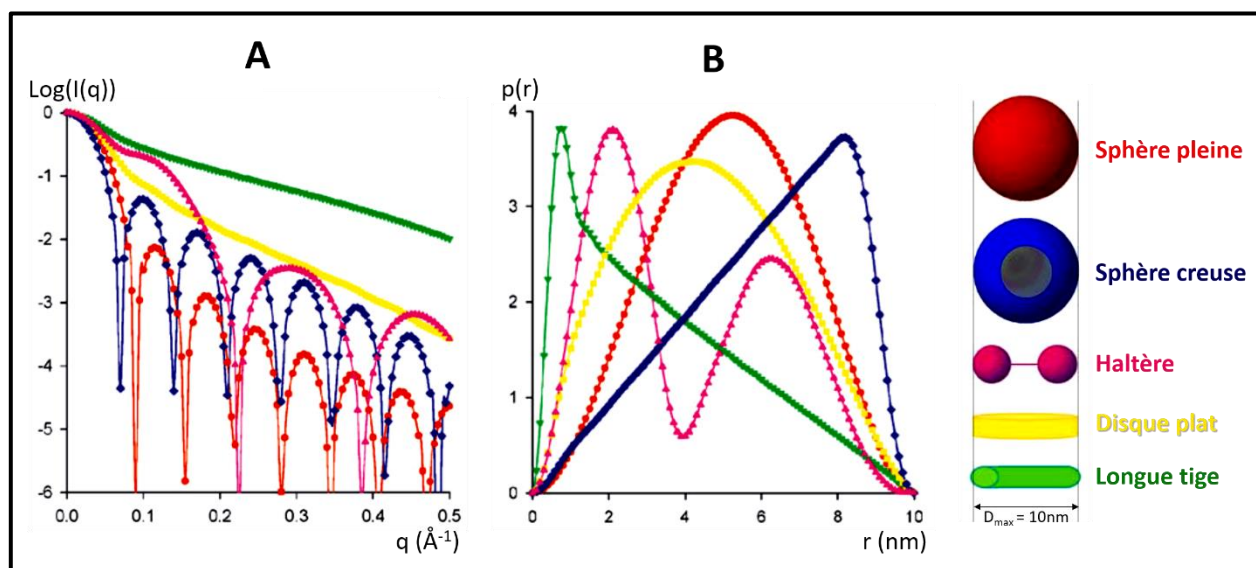


Figure 95 : Intensités de diffusion et fonctions de distribution de la distance de formes géométriques simples (adapté de Svergun et Koch, 2003) ^[194]. Courbes de diffusion (A) et fonctions de distribution de la distance par paire radiale (B) pour différentes formes géométriques ayant le même D_{max} .

Préparation des échantillons et réalisation de l'expérience

Les expériences de SAXS ont été menées au synchrotron SOLEIL (Source Optimisée de Lumière d'Énergie Intermédiaire du LURE, Saint-Aubin, France) sur la ligne de lumière SWING (**S**mall and **W**ide angle X-ray scattering**ING**).

Des échantillons de complexe, de duplex d'ADN et de la protéine ont été préparés et analysés dans un tampon TpSAXS (20 mM Hepes pH7.4, 200 mM NaCl, 2 % glycérol, 1 % TCEP) pour obtenir :

- le complexe D_1 /MC1t (C_1) à 1,4 mM et à 1,0 mM ;
- le duplex d'ADN D_1 à 1,0 mM ;
- la protéine MC1t à 2,37 mM et à 1,0 mM.

Les échantillons ont été chargés en amont d'une colonne de chromatographie d'exclusion stérique (Bio SEC-3-300, Agilent) préalablement équilibrée dans 100 % TpSAXS. L'acquisition des données est faite dès la sortie de la colonne et une seule injection par protéine a été nécessaire. La séparation des différentes espèces et agrégats en solution par la colonne de chromatographie a permis d'isoler le pic d'intérêt que l'on souhaite étudier. Les données de SAXS, collectées au niveau du pic d'élution avec une distance entre l'échantillon et le détecteur fixée à 2 m, ont ensuite été traitées par le logiciel FOXTROT ^[195]. Les différentes enveloppes ont été affinées avec la suite de programmes du logiciel ATSAS ^[190].

2.2. Optimisation du complexe pour l'étude

Afin de déterminer la structure du **complexe de MC1 avec un ADN^{23bp}**, nous avons besoin de conditions expérimentales permettant de **stabiliser une conformation unique du complexe**, en espérant que celle-ci soit la conformation comprenant les trois points de contacts. Il m'est apparu que la meilleure méthode pour déterminer cette structure était la **cristallographie aux rayons X (RX)**, du fait de la taille de ce complexe, des difficultés rencontrées lors de la détermination de la structure du complexe ADN^{15bp}/MC1, du coût d'un marquage des deux brins de l'ADN (aurait été absolument nécessaire à une étude RMN) et sûrement d'un marquage différent de la protéine (deutération ou marquage ¹³C des méthyles).

2.2.1. Surexpression et purification de la protéine MC1

Les premiers essais de cristallographie ont d'abord été réalisés sur la protéine native du fait de la disponibilité d'un échantillon pur marqué ¹⁵N dans l'équipe.

Les rendements obtenus avec le protocole optimisé au sein du laboratoire étaient : pour la phase de production, d'environ **1,5 g de culot bactérien par litre de culture de travail** et pour la phase de purification, d'environ **7 mg de protéines par gramme de culot bactérien**.

La préparation de l'échantillon de ¹⁵N-MC1 pur étant antérieur au début du projet de thèse (11/2008), je ne décrirai en détail que la surexpression et la purification de la protéine MC1t, non marquée, pour lesquelles j'ai été impliquée.

2.2.2. Surexpression et purification de la protéine MC1t

Pour l'étude par cristallographie, nous avons cherché à produire les cristaux de complexe stable de la **protéine MC1t (= GSHM-MC1)** avec un ADN^{23bp} et à résoudre sa structure 3D. Cette protéine correspond à MC1 native précédée de quatre acides aminés en N-terminal (Gly(-3), Ser(-2), His(-1), et Met0). L'extrémité N-terminale n'étant pas directement impliquée dans l'interaction à l'ADN, en particulier au niveau des trois points de contact, l'ajout des quatre résidus ne devrait pas empêcher ou modifier la formation du complexe. Au contraire, cette construction devrait nous permettre d'une part, de simplifier la purification de la protéine avec de meilleurs rendements, et d'autre part, d'utiliser un complexe possiblement plus stable pour la cristallographie.

Nous avons mis au point un nouveau protocole de production et de purification adapté à la protéine MC1t à partir d'une construction de plasmide réalisée au sein du laboratoire : **pET28b-mc1**.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Surexpression de la protéine MC1t fusionnée avec une étiquette « 6HIS »

ETAPE 1 : Transformation bactérienne

L'intégration du plasmide pET28b-mc1, par transformation bactérienne, dans des cellules (BL21-DE3) d'*E. coli* a permis la croissance de plusieurs colonies de bactéries (Figure 96). Ces colonies sont issues d'une préculture, non diluée ou diluée au dixième, étalée sur des boîtes de Pétri supplémentées de Kanamycine ($50 \mu\text{g.mL}^{-1}$).

Pour la boîte préparée en conditions non diluées (Figure 96a), les colonies sont nombreuses et globalement isolées les unes des autres. La seconde boîte, en conditions diluées (Figure 96b), montre également des colonies isolées et plus espacées que pour la première boîte. L'ensemble des colonies isolées sont utilisables pour ensemercer des cultures plus importantes. Nous avons choisi, par praticité, d'utiliser les colonies prélevées sur la seconde boîte pour les pré-cultures suivantes.

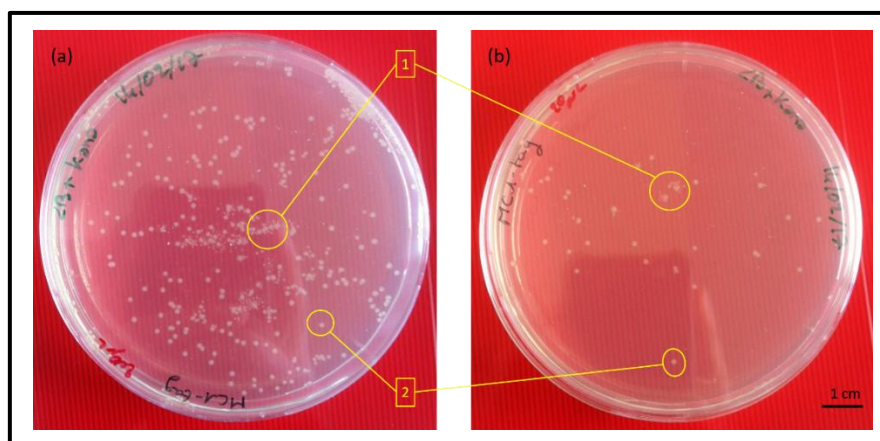


Figure 96 : Boîtes de culture cellulaire pour la surproduction de 6HIS-MC1t (a) à partir de 20 μl de culture non diluée et (b) à partir de 20 μl de culture diluée au dixième. 1 : colonies présentant une antibiorésistance de proximité. 2 : colonies isolées ayant intégré le plasmide.

ETAPES 2 et 3 : Surproduction et induction

Afin de réaliser l'ensemble de l'étude, le protocole de surexpression a été réalisé de nombreuses fois et les résultats de chaque surexpression étaient comparables. Nous avons récupéré pour chaque surexpression des culots d'environ **4 g pour 2 L de culture de travail**.

Les surproductions ainsi que l'induction par ajout d'IPTG ont été suivies par mesure régulière de l'absorbance à 600 nm (Figure 97a). Après 3 à 4 h de croissance, 3,5 h sur l'exemple présenté, l'expression a été induite par ajout de 1 mM final d'IPTG et contrôlée par électrophorèse sur gel SDS-PAGE 16 % (Figure 97b).

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

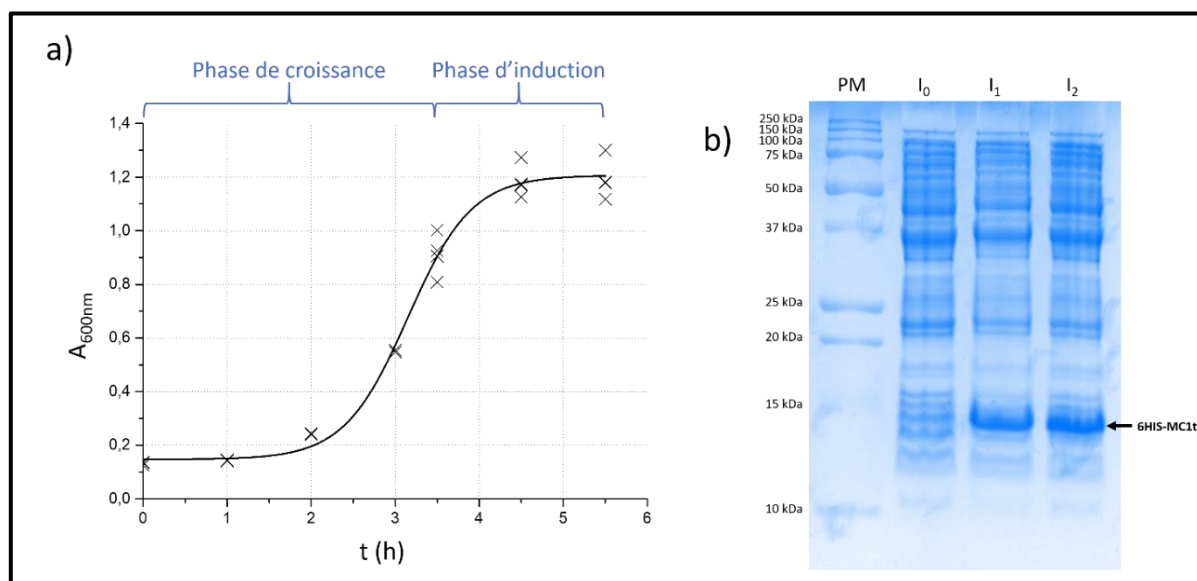


Figure 97 : Evolution des cultures de bactéries au cours de la surproduction de 6HIS-MC1t.
a) mesure des absorbances à 600 nm de quatre cultures de travail réalisées en parallèle. b) extrait du gel SDS-PAGE 16 % de contrôle de l'étape d'induction. PM = marqueur de poids moléculaire ; I_0 à I_2 = culture respectivement au moment de l'ajout de l'IPTG, 1 h après l'ajout et 2 h après l'ajout.

Purification de la protéine MC1t

La purification de la protéine MC1t se fait en six étapes comme expliqué dans le paragraphe « matériel et méthodes ». Je ne détaillerai ci-dessous, que les étapes ayant nécessité à un ajustement ou une optimisation à savoir les étapes 2, 3, 4 et 6.

ETAPE 2 : Chromatographie d'affinité n°1

Après extraction en présence de lysozyme, l'extrait brut de protéines est chargé sur une colonne de **chromatographie d'affinité sur ions métalliques divalents immobilisés (IMAC)**.

Un **premier test** avait été effectué sur une colonne de chromatographie d'affinité sur ions cobalt immobilisés (HisPur Cobalt Resin, Thermo Scientific) de 5 mL. Après avoir équilibré la colonne dans un tampon équivalent au tampon de lyse, Tp1 (20 mM Hepes pH7.6, 1 M NaCl, 5 mM imidazole, 1 mM PMSF), 35 mL d'extrait brut issus de 1,75 g de culots bactériens, ont été chargés puis lavés avec 10 CV de Tp1. Un second lavage a été effectué avec 10 CV d'un mélange 90:10 Tp1/Tp2 (Tp2 = 20 mM Hepes pH7.6, 1 M NaCl, 200 mM imidazole, 1 mM PMSF). La protéine d'intérêt a ensuite été éluée par un gradient linéaire de 20 CV jusqu'à 100 % de Tp2.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

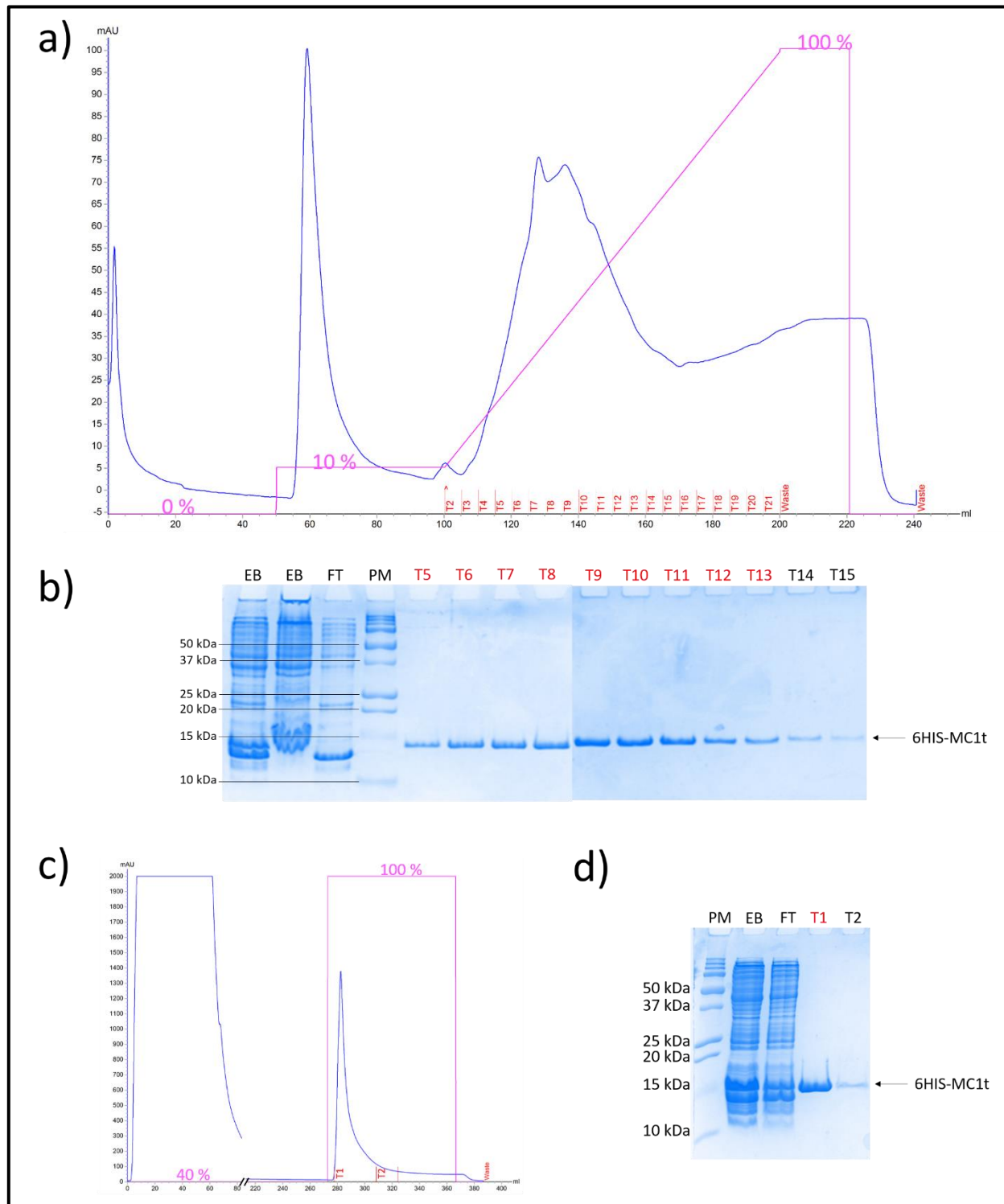


Figure 98 : Analyse de l'élution sur colonne IMAC. Chromatogramme de l'élution en mode gradient (a) et isocratique (c) ; analyse sur gel SDS-PAGE 16 % de l'élution en mode gradient (b) et isocratique (d). La proportion de Tp2 et l'absorbance à 280 nm sont représentées respectivement en rose et en bleu sur les chromatogrammes. PM = marqueur de poids moléculaire ; EB = extrait brut ; FT = partie éluee pendant la charge de l'échantillon ; T_i = fraction éluee i .

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Le chromatogramme de suivi de l'absorbance à 280 nm de ce premier test montre qu'au moins un composé est élué en présence de 20 mM d'imidazole et qu'au moins un autre composé est élué entre 35 mM et 140 mM d'imidazole ([Figure 98a et b](#)).

Lors de cette première purification, les fractions **T5** à **T13** contenant la protéine 6HIS-MC1t ont été rassemblées et formaient la fraction « **IMAC** » de 58 mL utilisée pour l'étape suivante.

Une analyse complémentaire par spectrométrie de masse a montré que le composé élué pendant le second lavage correspondait également à la protéine 6HIS-MC1t. Il s'est avéré qu'une faible quantité d'EDTA était présente lors du premier test et avait limité la fixation de 6HIS-MC1t sur la colonne.

Pour les purifications suivantes, nous avons privilégié une colonne de chromatographie d'affinité sur ions nickels immobilisés de même volume de colonne que la précédente. Lors de ces essais, l'ensemble de la protéine restait spécifiquement fixé sur la colonne pendant les deux lavages. Elle a été spécifiquement éluée en mode isocratique avec 200 mM d'imidazole afin de réduire le facteur dilution lié à l'élution. La protéine a été dans ce cas éluée en une seule fraction « **IMAC** » de 30 mL dans l'exemple illustré ([Figure 98c et d](#)).

ETAPE 3 : Chromatographie échangeuse d'ions

La fraction « IMAC » de 30 mL, et diluée trois fois avec le Tp3 pauvre en sels, est ensuite chargée sur une colonne de **chromatographie fortement échangeuse de cations** (POROS HS/20 10 mm D x 100 mm L PEEK de 7,9 mL). La protéine est alors éluée entre 600 mM et 790 mM de NaCl.

Trois maxima sont observés à chaque élution et semblent tous, d'après les résultats sur gel, correspondre à de la protéine d'intérêt. Afin de s'assurer une homogénéité de l'échantillon pour les essais finaux, seules les fractions du pic majoritaire sont rassemblées et concentrées. Dans l'exemple illustré sur la [Figure 99](#), les fractions **T50** à **T63** ont été rassemblées pour l'étape suivante et constituent la fraction globale « **HS20** » (63 mL).

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

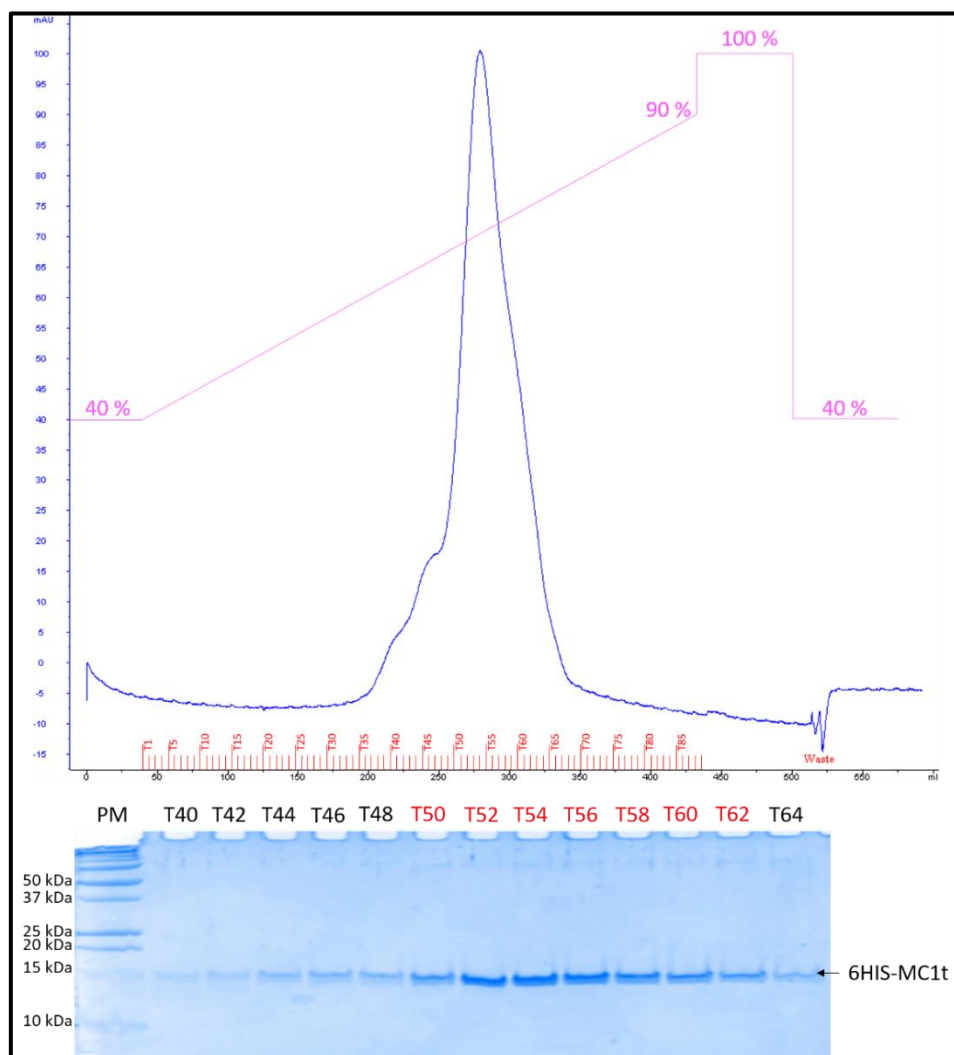


Figure 99 : Analyse de l'élution de la colonne HS20. Chromatogramme de l'élution et analyse sur gel SDS PAGE 16 %. Respectivement en rose et en bleu sur le chromatogramme : la proportion de Tp2 et l'absorbance à 280 nm. PM = marqueur de poids moléculaire ; T_i = fraction éluee i .

ETAPE 4 : Clivage de l'étiquette 6HIS

La fraction globale « HS20 » est ensuite utilisée pour l'étape de **clivage de l'étiquette « 6HIS » par la thrombine (Thr)**. Pour cette étape, nous avons dû, dans un premier temps, adapter un kit de clivage à la thrombine commercial (*Thrombin Cleavage Capture Kit* de Novagen) à notre protéine. Pour ce faire, la quantité d'enzyme en présence de l'échantillon, et le temps de mélange, t_i , pour le clivage ont dû être déterminés. De plus, il est nécessaire de diluer la thrombine (dans le tampon de dilution fourni dans le kit - *Thrombin Dilution/Storage Buffer*) avant l'ajout dans l'échantillon de protéine. Il a donc fallu déterminer le ratio, $R_i(\text{Thr})$, de cette première dilution.

La fraction « HS20 », issue de l'ETAPE 3, est concentrée jusqu'à $0,60 \text{ mg.mL}^{-1}$ pour un volume de 1,6 mL, et son tampon est échangé pour le Tp5 (20 mM Hepes pH7.6, 500 mM NaCl). Des échantillons tests de clivage (E_i) sont ensuite préparés à partir du kit de Novagen et de son protocole.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Le test d'optimisation a été réalisé avec plusieurs temps de mélange (2 h, 4 h et 16 h) et différents ratios de thrombine (1:25, 1:50, 1:100, 1:200 et 1:400) afin de déterminer les meilleures conditions de clivage de 6HIS-MC1t. Un contrôle négatif (E_0) est effectué en parallèle avec un temps de mélange de 24 h et en absence de thrombine.

Les échantillons de clivage ont été préparés pour un volume final, V_{tot} , de 30 μL avec : 10 μg de 6HIS-MC1t (soit 20 μL à 0,47 mM) ; 3 μL de tampon de clivage 10X ; 0,6 μL de thrombine préalablement diluée ; et 6,4 μL d'eau ultrapure. Les échantillons ont été placés à 4 °C sous agitation pendant 16 h avec prélèvement aux différents temps de mélange (35 rpm à 4 °C à $t+X$ h). Les différents produits de digestion ont ensuite été analysés par gel SDS-PAGE 16 % ([Figure 100](#)).

Pour notre protocole, nous souhaitons que la consommation de thrombine ne soit pas trop importante et que le clivage soit réalisé pendant la nuit (soit entre 10 et 16 h). D'après les résultats des gels, le clivage est total après 4 h de mélange quand la thrombine est utilisée avec un ratio supérieur ou égal à (1:100 Thrombin/Buffer). Il est total après 16 h de mélange quand la thrombine est utilisée avec un ratio supérieur ou égal à (1:400 Thrombin/Buffer).

Nous avons retenu, pour le protocole optimisé, les conditions de clivage de l'essai **E₁₂** pour lequel le ratio est de **1:200** qui permet donc l'utilisation d'une faible quantité de thrombine tout en s'assurant une certaine flexibilité sur le temps de mélange (entre 10 h et 16 h de mélange) pour obtenir un clivage total.

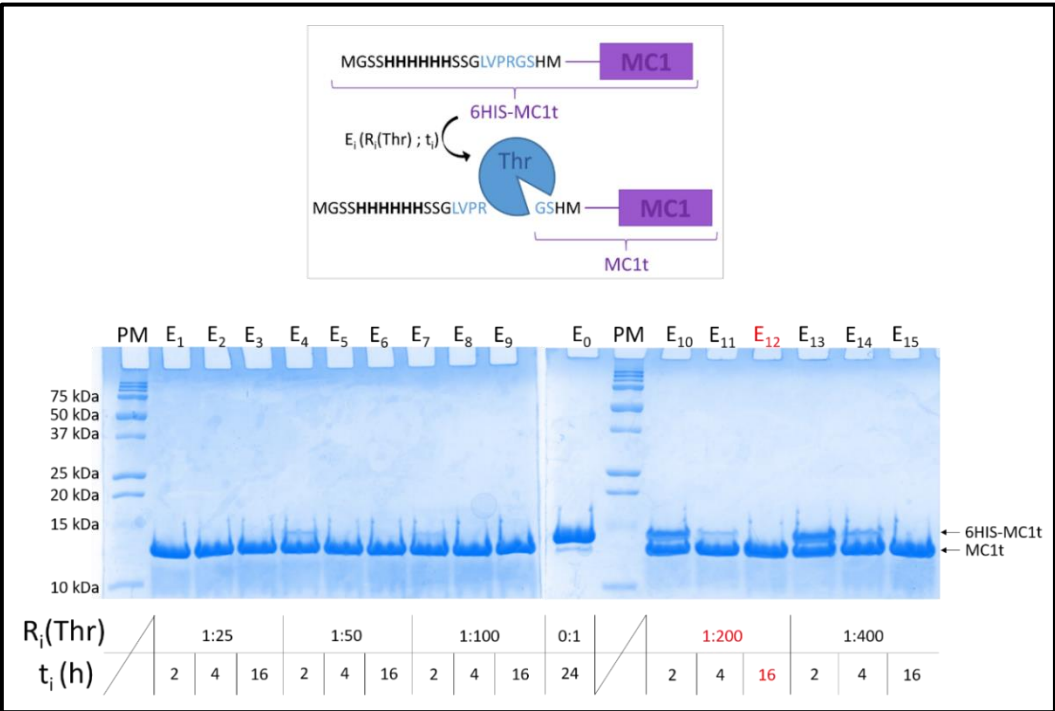


Figure 100 : Clivage de l'étiquette « 6HIS » à la thrombine. Schéma (haut) et analyse sur gel SDS-PAGE 16 % (bas). PM = marqueur de poids moléculaire ; E_i = essai i ; $R_i(\text{Thr})$ = ratio de thrombine ; t_i = temps de mélange.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

ETAPE 6 : Chromatographie par filtration sur gel

La fraction « IMAC/Bz » de 12 mL est obtenue à l'issue de l'**ETAPE 5** (inhibition de la thrombine et deuxième chromatographie d'affinité) par chromatographie IMAC en mode isocratique. Elle est d'abord concentrée et le tampon est échangé pour le Tp5 jusqu'à un volume de 1 mL par centricon Amicon (3 kDa). Elle est ensuite chargée sur une colonne de chromatographie par filtration sur gel (S75) de 120 mL préalablement équilibrée avec le Tp4. La protéine est alors éluée entre 75 mL et 90 mL de Tp4 (**Figure 101**).

Les fractions **T40** à **T51** contenant MC1t pure sont rassemblées, concentrées et le tampon est échangé pour le tampon F (20 mM Hepes pH7.6, 100 mM NaCl) par centricon Amicon (3 kDa). Nous avons obtenu à l'issu de cette étape 150 μ L de protéine pure à 2,06 mM (issus d'un culot de 2,66g).

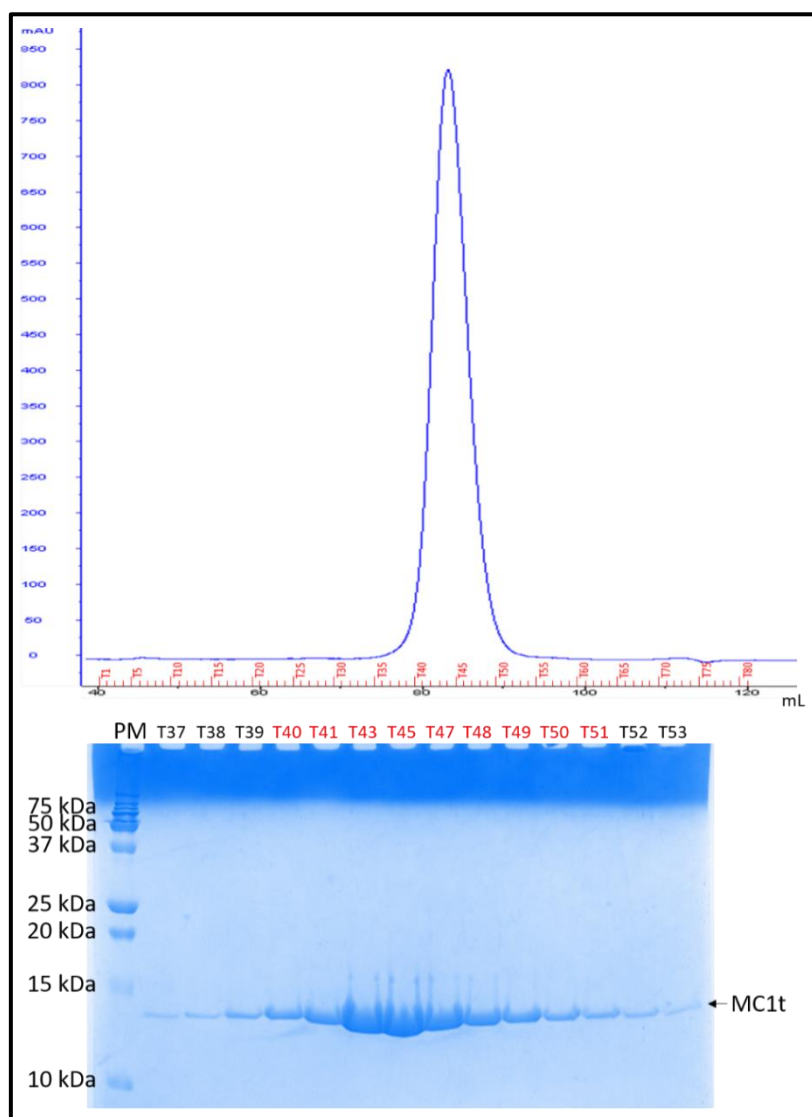


Figure 101 : Analyse de l'élution de la colonne S75. Chromatogramme de l'élution et analyse sur gel SDS PAGE 16 %. Suivi de l'absorbance à 280 nm représenté en bleu. PM = marqueur de poids moléculaire ; T_i = fraction éluée i .

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Rendement de la purification de la protéine MC1t

Un rendement d'en moyenne **2 mg** de protéine purifiée par gramme de culot de bactéries est obtenu en fin de chaque purification et la pureté de l'échantillon a systématiquement été confirmée par une analyse en spectrométrie de masse (Figure 102, $MW_{th} = 11076$ Da et $MW_{exp} = 11077$ Da).

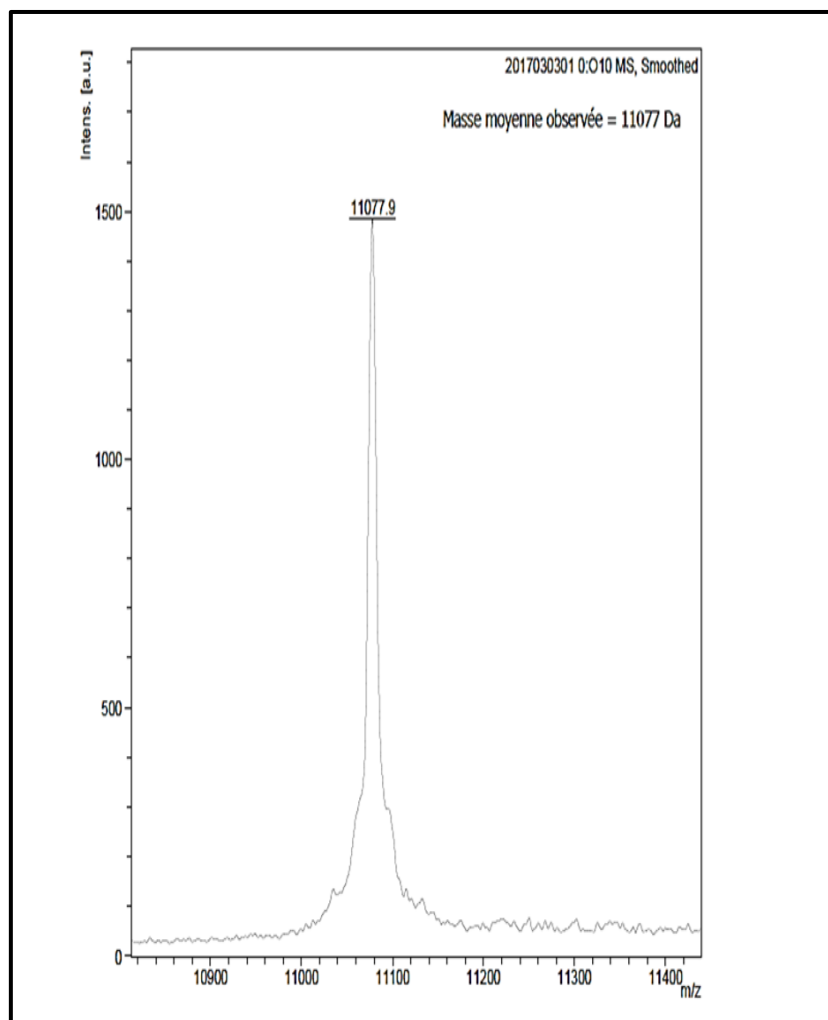


Figure 102 : Spectre de masse déconvolué de la protéine MC1t pure.

2.2.1. Complexation avec l'ADN^{23bp}

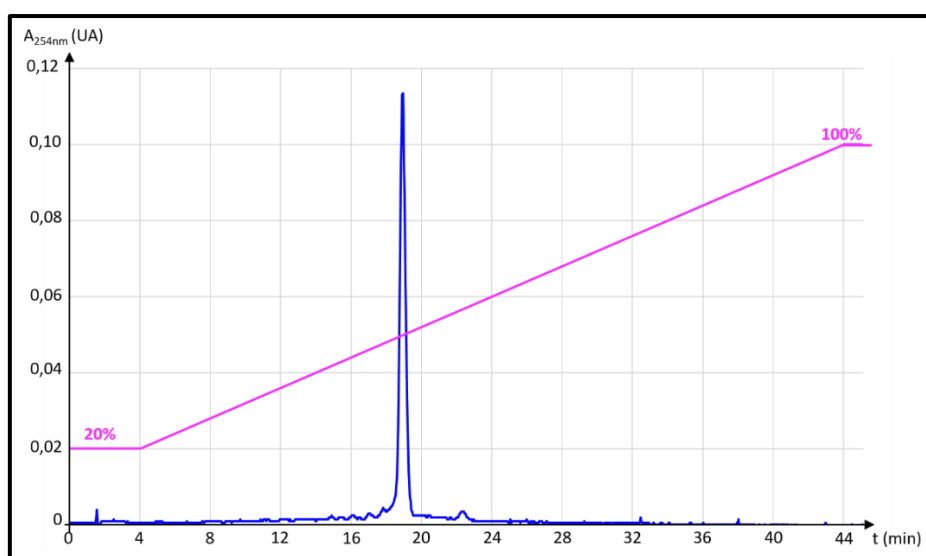
Purification des oligonucléotides de synthèse

Les oligodéoxyribonucléotides synthétiques de 23 paires de bases utilisés pour l'étude, ont été achetés avec une pureté estimée à 70 % (21,19 mg de SUP_{RX}^{23bp}, 13,22 mg de INF₁^{23bp}, 13,07 mg de INF₂^{23bp} et 12,88 mg de INF₃^{23bp}). Cette pureté n'étant pas suffisante pour des essais de cristallogénèse, une étape de purification supplémentaire est réalisée par chromatographie échangeuse d'anions.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Des tests analytiques sur une colonne de **chromatographie échangeuse d'anions** (MonoQ 10/100GL) de 8 mL ont été effectués afin d'optimiser spécifiquement le gradient d'élution de chaque brin avec cette colonne. Ces tests analytiques ont été réalisés avec des concentrations d'échantillon de 5,0 μM pour le brin $\text{SUP}_{\text{RX}}^{23\text{bp}}$ et 7,4 μM pour les brins $\text{INF}_i^{23\text{bp}}$.

La première méthode testée avec le brin $\text{SUP}_{\text{RX}}^{23\text{bp}}$ consistait en un plateau avec un mélange 80:20 TpA/TpB sur 2 CV puis un gradient jusqu'à 100 % TpB sur 20 CV. Un maximum d'absorbance à 254 nm a été obtenu pour cette élution à 49,84 % de TpB soit ~700 mM NaCl ([Figure 103](#)). Pour l'optimisation, le pourcentage de tampon B du plateau initial et celui du gradient ont été ajustés afin d'éluer le plus sélectivement les oligonucléotides testés.



[Figure 103](#) : **Chromatogramme de l'élution de $\text{SUP}_{\text{RX}}^{23\text{bp}}$** . Respectivement en rose et en bleu, le suivi de la proportion de TpB et de l'absorbance à 254 nm.

Au final, la méthode optimisée pour le brin supérieur a permis d'éluer l'oligonucléotide $\text{SUP}_{\text{RX}}^{23\text{bp}}$ à **46,0 % de TpB**. La méthode optimisée pour les brins inférieurs a permis d'éluer $\text{INF}_1^{23\text{bp}}$, $\text{INF}_2^{23\text{bp}}$ et $\text{INF}_3^{23\text{bp}}$ respectivement à 80,6 %, 81,1 % et 82,6 % de TpB.

Suite au dessalage par **chromatographie de phase inverse**, nous avons finalement obtenu des échantillons d'oligonucléotides de 500 μL dans le tampon final F (TpF : 20 mM Hepes pH 7,6, 100 mM NaCl). Les concentrations de ces échantillons finaux ont pu être déterminées par la mesure de l'absorbance à 260 nm (5,23 mM pour $\text{SUP}_{\text{RX}}^{23\text{bp}}$, 2,38 mM pour $\text{INF}_1^{23\text{bp}}$, 3,08 mM pour $\text{INF}_2^{23\text{bp}}$ et 2,56 mM pour $\text{INF}_3^{23\text{bp}}$).

Cette étape nous permet d'obtenir des brins d'oligonucléotides de meilleure pureté avec 76,41 % de rendement en moyenne (85,37 % pour $\text{SUP}_{\text{RX}}^{23\text{bp}}$, 64,52 % pour $\text{INF}_1^{23\text{bp}}$, 84,47 % pour $\text{INF}_2^{23\text{bp}}$ et 71,27 % pour $\text{INF}_3^{23\text{bp}}$).

Hybridation et complexation

Les oligonucléotides purifiés ont ensuite pu être **hybridés** en double brin d'ADN pour réaliser des échantillons de duplex D_i où i correspond au numéro du brin inférieur (INF_i^{23bp}).

Les duplex sont ensuite complexés à la protéine MC1t pour obtenir les échantillons C_i , où i correspond au numéro du brin inférieur (INF_i^{23bp} et D_i), finaux de complexe. Ces échantillons ont été préparés à des concentrations de 1,2 mM et 1,5 mM dans le tampon F. Tous les échantillons ont été stockés à 4 °C.

2.3. Vers une description atomique du complexe

Une fois les échantillons de complexe préparés, il nous a fallu obtenir des cristaux exploitables pour une analyse par rayons X.

2.3.1. Optimisation de la cristallogénèse

Tests de criblage sur les complexes $D_i/^{15}N$ -MC1

Les complexes $D_i/^{15}N$ -MC1, ont tout d'abord été testés à 0,5 mM final avec les kits JCSG+ de Molecular Dimensions, NUCLEIX et CLASSICS de chez Qiagen, soit **288 conditions au total**. Cette méthode de criblage nous a permis d'évaluer rapidement le comportement du complexe pour une grande variété de conditions de cristallisation.

Nous avons observé que pour 81 % des conditions testées sur nos trois complexes, aucun cristal n'était obtenu (Figure 104, JCSG+ 2.11 - $D_1/^{15}N$ -MC1). En effet, dans 82,6 % des conditions testées pour $D_1/^{15}N$ -MC1, 82,3 % pour $D_2/^{15}N$ -MC1 et 79,5 % pour $D_3/^{15}N$ -MC1, la goutte contenant le complexe est restée limpide. Le complexe se trouve dans ce cas en sous-saturation. Ce résultat peut être lié à une concentration trop faible du complexe pour accéder à la zone de nucléation et donc à la formation de cristal.

Nous avons observé l'apparition de précipités pour 14 % des conditions testées (11,8 % pour $D_1/^{15}N$ -MC1, 13,5 % pour $D_2/^{15}N$ -MC1 et 16,0 % pour $D_3/^{15}N$ -MC1). Ces observations indiquent une trop forte saturation du complexe au sein de la goutte (Figure 104, JCSG+ 2.11 – $D_2/^{15}N$ -MC1).

Enfin, nous avons pu observer des éléments qui dépolarisent, sous la forme de sphérulites, d'aiguilles (Figure 104, JCSG+ 2.11 – $D_3/^{15}N$ -MC1) ou de cristaux (Figure 104, JCSG+ 1.44 – $D_2/^{15}N$ -MC1), pour 5 % des conditions testées (5,6 % pour $D_1/^{15}N$ -MC1, 4,2 % pour $D_2/^{15}N$ -MC1 et 4,5 % pour $D_3/^{15}N$ -MC1). Ce sont ces conditions que nous avons cherché à optimiser.

Au final trois conditions de cristallisation ont été retenues à l'issue de ce premier criblage :

- complexe $D_3/^{15}N$ -MC1 en présence de 0,2 M nitrate d'ammonium, 20 % PEG 3350 ;

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

- complexe $D_3/^{15}\text{N-MC1}$ en présence de 100 mM Tris pH8.0, 40 % MPD ;
- complexe $D_2/^{15}\text{N-MC1}$ en présence de 160 mM acétate de calcium, 80 mM cacodylate de sodium pH6.5, 14,4 % PEG 8k, 20 % glycérol.

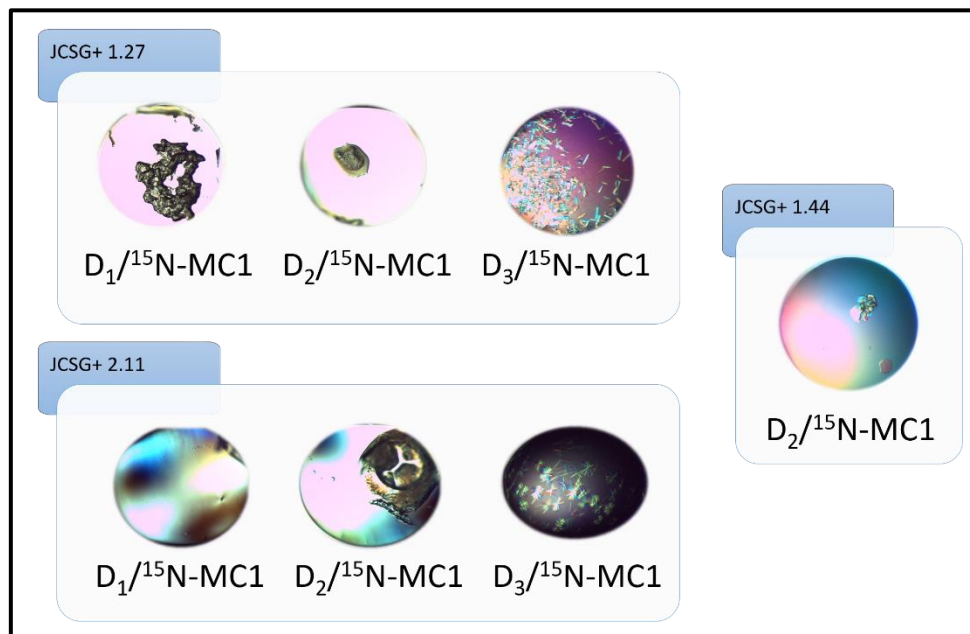


Figure 104 : Gouttes issues de tests de criblage sur les complexes $D_i/^{15}\text{N-MC1}$ dans les conditions JCSG+ 1.27 (0,2 M nitrate d'ammonium, 20 % PEG 3350), 2.11 (160 mM acétate de calcium, 80 mM cacodylate de sodium pH6.5, 14,4 % PEG 8k, 20 % glycérol) et 1.44 (100 mM Tris pH8.0, 40 % MPD).

Optimisation des conditions de cristallisation sur les complexes $D_i/^{15}\text{N-MC1}$

Nous avons alors cherché à reproduire sur des volumes plus gros (gouttes de 2 μL) les conditions retenues et à les optimiser afin d'améliorer la qualité des premiers cristaux (taille, forme) obtenus.

La reproduction des gouttes préparées avec les conditions 1.27 du kit JCSG+ (0,2 M nitrate d'ammonium, 20 % PEG 3350) et 1.44 du kit JCSG+ (100 mM Tris pH8.0, 40 % MPD) sur les trois complexes de $D_i/^{15}\text{N-MC1}$ n'a pas donné de résultat positif. Les gouttes sont restées limpides dans ces conditions. Elles ont donc été écartées de notre étude.

Les gouttes préparées avec la condition 2.11 du kit JCSG+ (160 mM acétate de calcium, 80 mM cacodylate de sodium pH6.5, 14,4 % PEG 8k, 20 % glycérol) ont, en revanche, donné des cristaux sous forme d'aiguilles avec le complexe $D_3/^{15}\text{N-MC1}$ une journée seulement après leur préparation.

De ce fait, afin d'améliorer la qualité et la taille des cristaux, différents paramètres ont été modifiés :

- la concentration de complexe $D_3/^{15}\text{N-MC1}$ (0,33 mM, 0,5 mM et 0,67 mM final) ;
- la concentration du sel (0, 50 mM, 100 mM et 150 mM acétate de calcium) ;
- le pH du tampon (100 mM cacodylate de sodium pH5, pH6 et pH7) ;
- le pourcentage de l'agent précipitant (8 %, 11 %, 14 %, 15 %, 16 %, 17 % et 20 % PEG 8k) ;

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

- le pourcentage d'agent cryoprotectant (5 %, 10 % et 20 % glycérol).

Au total 66 nouvelles conditions ont été testées à 20 °C et nous ont permis d'obtenir des cristaux dans 25 cas. Ces nombreux tests nous ont permis de faire les observations suivantes :

- la formation des cristaux est favorisée avec des concentrations en sels plus élevées (150 mM) ;
- la cristallisation est favorisée en présence de 5 % de glycérol ;
- la concentration en complexe ne semble pas influencer la quantité de cristaux formés ;
- les autres paramètres ne semblent pas avoir d'influence sur la formation des cristaux.

Les **conditions retenues** pour la formation des cristaux sont : 0,33 mM final de complexe $D_3/^{15}\text{N}$ -MC1, 150 mM d'acétate de calcium, 100 mM de cacodylate de sodium pH6, 17 % de PEG 8k et 5 % de glycérol (Figure 105).

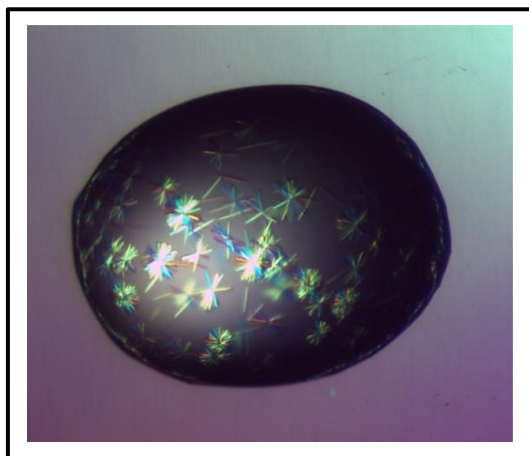


Figure 105 : Cristaux du complexe $D_3/^{15}\text{N}$ -MC1 en conditions 150 mM d'acétate de calcium, 100 mM de cacodylate de sodium pH 6, 17 % de PEG 8k et 5 % de glycérol.

Tests de criblage sur les complexes $D_i/\text{MC1t}$

Une seconde série de tests de criblage a été réalisée sur les trois complexes optimisés ($D_i/\text{MC1t}$ ou C_i) à 0,6 mM final dans 288 conditions supplémentaires. Pour cette deuxième série, les kits WIZARD de Emerald Biostructures, CRYO de Sigma, NUCLEIX et COMPAS de Qiagen ont été utilisés.

Dans 79 % des conditions, aucun cristal n'a été observé (82,3 % pour C_1 , 81,3 % pour C_2 et 72,4 % pour C_3). Comme pour la première série de tests de criblage, nous sommes majoritairement dans des conditions de sous-saturation.

Des précipités sont apparus pour 15 % des conditions (12,8 % pour C_1 , 12,8 % pour C_2 et 19,3 % pour C_3) et des éléments qui dépolarisent pour 6 % des conditions (4,9 % pour C_1 , 5,9 % pour C_2 et 8,3 % pour C_3) ont été observés.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Nous nous sommes ensuite focalisés uniquement sur l'amélioration des cristaux les plus gros et dont les faces étaient les mieux définies pour l'étape d'optimisation ([Figure 106](#)). Ces cristaux ont été obtenus avec le complexe C₃ en présence de 0,5 M KCl, 12 % PEG 8k et 10 % glycérol.

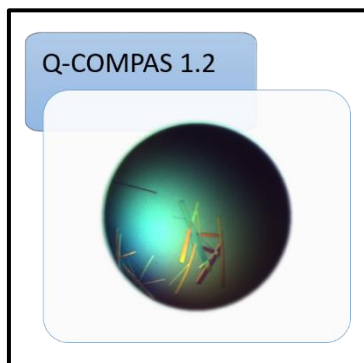


Figure 106 : Cristaux du complexe D₃/MC1t (ou C₃) dans les conditions COMPAS 1.2 (0,5 M KCl, 12 % PEG 8k, 10 % glycérol).

Optimisation des conditions de cristallisation sur le complexe D₃/MC1t

Afin de réaliser l'optimisation, nous sommes passés sur des gouttes de plus gros volumes (2 µL) contenant chacune un de nos trois complexes (D_i/MC1t) avec d'une part les solutions commerciales des kits et de d'autre part des solutions réalisées au laboratoire.

Les gouttes préparées avec la condition 1.2 du kit COMPAS (0,5 M KCl, 12 % PEG 8k, 10 % glycérol) issue du second criblage ont donné des cristaux sous forme d'un amas d'aiguilles uniquement pour le complexe C₃.

Nous avons ensuite fait varier la température de cristallisation (20 °C et 4 °C), la concentration en complexe (0,4 mM, 0,6 mM et 0,8 mM final), la concentration en sel (0,15 mM, 0,2 mM, 0,25 mM, 0,3 mM, 0,35 mM, 0,4 mM, 0,5 mM, 1 M KCl), le pourcentage de l'agent précipitant (12 %, 14 %, 16 %, 17 % et 18 % PEG 8k) ainsi que le type d'agent précipitant (14 % PEG 3350, PEG 5k, PEG 6k, PEG 8k et PEG 10k). Tous les essais ont été réalisés avec un pourcentage constant de glycérol (10 %, agent cryoprotectant).

Au total 132 conditions ont été testées et nous avons obtenu des cristaux dans 34 de ces conditions ([Figure 107](#)). Les cristaux, considérés comme étant exploitables pour une étude structurale, ont été obtenus en deux jours, à partir de gouttes de complexe « frais » (préparées dans les 48 h). Il est à noter que, pour l'ensemble des gouttes contenant des cristaux, nous avons observé la formation d'une coque de précipité à la surface de la goutte. Cette coque a dû être rompue afin de pouvoir isoler un cristal du complexe.

Les **conditions retenues** pour la formation des cristaux sont : 0,8 mM final de complexe C₃, 250 mM de KCl, 14 % de PEG 8k et 10 % de glycérol.

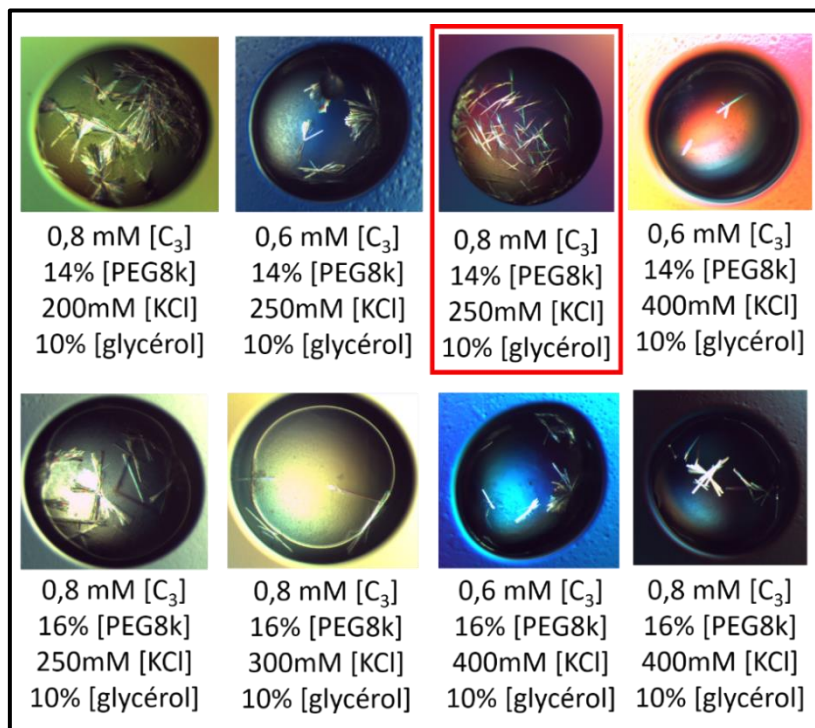


Figure 107 : Optimisation de la cristallogénèse sur le complexe C₃ à 20 °C. Les meilleures conditions de cristallisations sont encadrées en rouge.

2.3.2. Tests de diffraction des rayons X au synchrotron

Les cristaux retenus ont été testés au synchrotron ESRF de Grenoble sur la ligne de lumière ID29.

Le cliché de diffraction obtenu sur les **cristaux de D₃/15N-MC1** (0,33 mM final de complexe, 150 mM d'acétate de calcium, 100 mM de cacodylate de sodium pH6, 17 % de PEG 8k et 5 % de glycérol) présentait trop peu de signaux et ne nous a pas permis d'obtenir d'information structurale.

La Figure 108 présente un premier cliché de diffraction qui a pu être enregistré à partir de l'un des cristaux de C₃ (8 mM final de complexe, 250 mM de KCl, 14 % de PEG 8k et 10 % de glycérol). On peut y observer des signaux espacés périodiquement. Ces signaux sont peu nombreux et de faible intensité. De par leur proximité et leur périodicité, ces signaux nous permettent d'attester qu'il s'agit bien du cliché de diffraction d'une macromolécule biologique et non d'un sel.

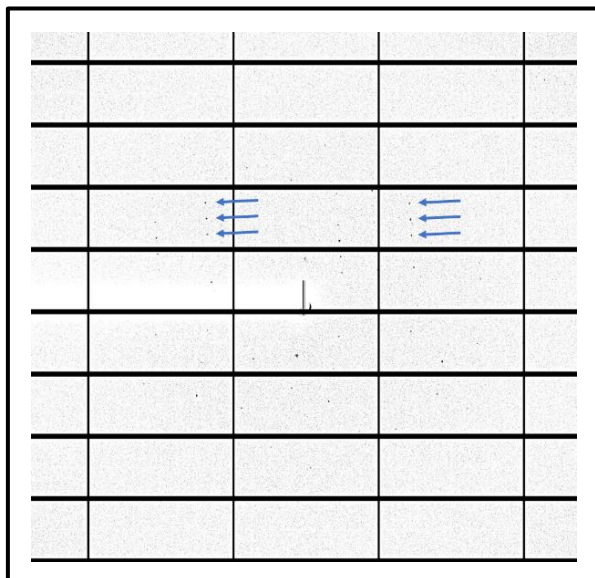


Figure 108 : Cliché de diffraction par rayons X d'un cristal de complexe C₃. L'enchainement des flèches bleus reflète la périodicité de la maille cristalline dans une dimension.

Un ensemble de clichés de diffraction pour lesquels l'angle d'incidence du faisceau de rayons X sur le cristal varie ($\Delta\theta=0,15^\circ$) a été enregistré. De par la faible résolution des clichés obtenus, le traitement informatique de ces derniers ne nous a permis que d'obtenir une idée du groupe spatial du cristal (P4₁2₁2 ou P4₃2₁2), ainsi que ses paramètres de maille ($a, b = 88,83 \text{ \AA}$; $c = 174,08 \text{ \AA}$; $\alpha, \beta, \gamma = 90^\circ$).

Les résultats obtenus avec les cristaux de C₃ sont prometteurs mais ne nous ont pas permis de résoudre la structure du complexe. Une nouvelle analyse par diffraction des rayons X sur de jeunes cristaux de C₃ et l'éventuel changement du cryoprotectant pourrait peut-être aboutir à l'enregistrement de clichés de diffraction plus résolus et donc permettre l'obtention de la structure tridimensionnelle du complexe.

2.4. Vers une description de l'enveloppe spatiale du complexe

Afin de compléter les données obtenues précédemment par RMN et par cristallographie aux rayons X, j'ai mené des études de diffusion de rayons X aux petits angles (SAXS). Nous avons ainsi pu caractériser les enveloppes de différents objets biologiques : la protéine MC1t seule, un duplex d'ADN^{23bp} « bouts francs » seul (D₁) et le complexe D₁/MC1t (C₁).

2.4.1. Essais de diffusion SAXS

Les données, collectées sur la ligne de lumière SWING (synchrotron SOLEIL), ont toutes pu être traitées complètement à l'exception de celles du complexe C₁ à 1 mM pour lequel le D_{max} n'a pas pu être déterminé.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

L'analyse de Guinier des courbes de diffusion a été possible pour chaque objet et a permis en premier lieu de nous conforter sur l'état de non agrégation de nos échantillons ([Figure 109](#)). Si les échantillons s'étaient agrégés, nous aurions observé, pour des valeurs du vecteur de diffusion q inférieure à $0,1 \text{ \AA}^{-1}$, une importante décroissance du signal. En effet, plus un objet est volumineux, plus son $\Delta(\text{Log}(I(q)))$ sera grand pour de faibles valeurs de q . C'est ce que nous pouvons observer avec nos trois objets ($\text{MC1t} = 11,08 \text{ kDa}$; $D_1 = 14,09 \text{ kDa}$ et $C_1 = 25,16 \text{ kDa}$).

La courbe de diffusion de D_1 peut être assimilable à une droite ce qui correspond à une forme allongée attendue pour le duplex. La courbe de diffusion du complexe peut, quant à elle, être assimilée à une courbe en cloche. Cette allure de courbe correspond à un objet plutôt globulaire. La courbe de la protéine libre correspond à une forme majoritairement globulaire.

Les paramètres d'intensité de diffusion à l'origine $I(0)$ et de rayon de giration R_g , estimés à partir de cette analyse (Guinier), ont été reportés dans le [Tableau 21](#) tout comme la distance maximale, D_{max} , et une estimation du volume de l'enveloppe (Porod).

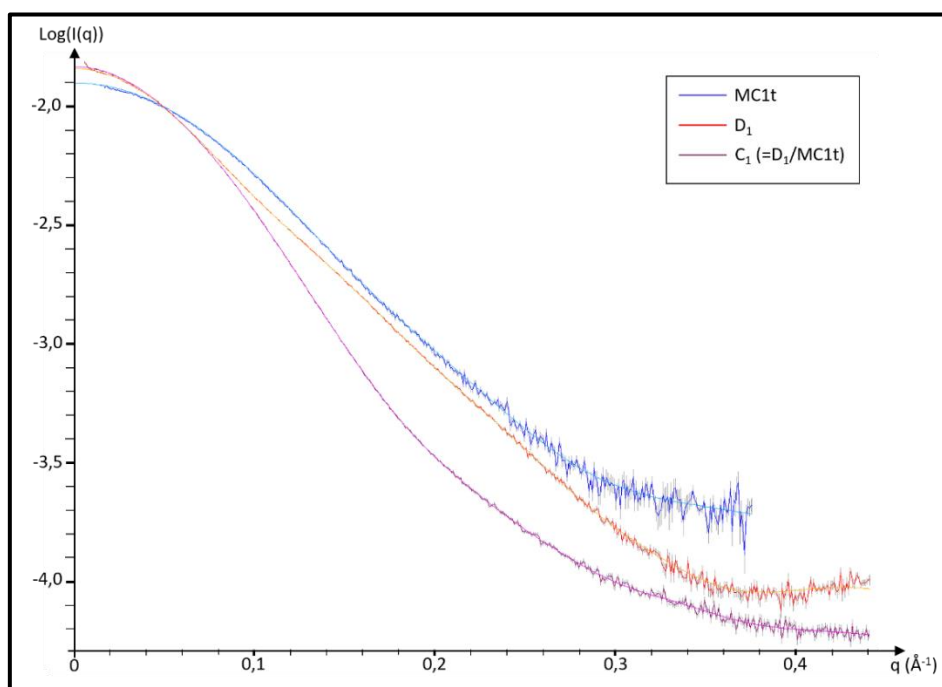


Figure 109 : *Superposition des courbes de diffusion SAXS de MC1t, de D_1 et de C_1 respectivement en bleu, en rouge et en violet.*

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

	MC1t		D1/MC1t		D1
Paramètres de collecte des données					
Instrument	SEC-SAXS - ligne de lumière SWING (SOLEIL)				
colonne d'exclusion stérique	Bio SEC-3 300Å (Agilent)				
Longueur d'onde (Å)	1,03319713	1,03317	1,03319713	1,03317	1,03317
gamme q (Å ⁻¹)	0,01-0,61	0,01-0,53	0,01-0,61	0,01-0,53	0,01-0,53
Temps d'exposition (s)	0,94	0,88	0,94	0,88	0,88
Distance (mm)	1789	2087	1789	2087	2087
Température (°C)	16	16	16	16	16
concentration (mM)	2,37	1,0	1,42	1,0	1,0
concentration (g.L ⁻¹)	26,26	11,08	35,73	25,16	14,09
Paramètres structuraux					
I(0) (Å ⁻¹) (Guinier)	2,9.10 ⁻² ± 3,2.10 ⁻⁵	1,3.10 ⁻² ± 1,3.10 ⁻⁵	1,3.10 ⁻¹ ± 5,0.10 ⁻⁵	1,2.10 ⁻¹ ± 4,5.10 ⁻⁵	5,7.10 ⁻² ± 2,3.10 ⁻⁵
Rg (Å) (Guinier)	17,20 ± 0,38	16,42 ± 1,21	20,55 ± 0,13	21,39 ± 1,30	21,51 ± 1,16
D _{max} (Å) (P(r))	52,83	55,27	71,67	70,58	75,79
Volume estimé (Å ³) (P(r))	15900	15400	28600	34300	17800
Programmes utilisés					
Réduction des données	Foxtrot	Foxtrot	Foxtrot	Foxtrot	Foxtrot
Analyse des données	PRIMUS	PRIMUS	PRIMUS	PRIMUS	PRIMUS
Validation des modèles	DAMMIN CRY SOL	-	DAMMIN MONSA CRY SOL	-	-

Tableau 21 : Paramètres des données SAXS des échantillons analysés. Les valeurs ont été estimées pour chaque échantillon à partir des courbes de diffusion.

La Figure 56 présente les tracés de Kratky. Ils permettent l'analyse de la flexibilité intrinsèque des objets étudiés (Figure 110).

Le tracé de la protéine MC1 met en évidence une partie flexible assez importante et une partie rigide. Ce résultat est en accord avec la structure tridimensionnelle de la protéine libre, les données de relaxation et de dynamique moléculaire qui ont été obtenues précédemment et qui montrent que la protéine possède un corps globulaire pouvant être considéré comme rigide et un bras flexible (la boucle LP5) [22]. Le duplex D₁ apparaît, quant à lui, comme étant très flexible, ce qui était aussi un résultat attendu. Le complexe apparaît comme étant plus rigide que les deux partenaires libres, ce qui est en accord avec les données structurales et de relaxation obtenues sur le complexe ADN^{15bp}/MC1 qui montre que le bras de MC1 devient rigide après complexation et que le complexe ADN/protéine est stable dans le temps (aucune dissociation ou dégradation n'est constatée sur les spectres ¹⁵N-HSQC après plusieurs mois de conservation à 4 °C).

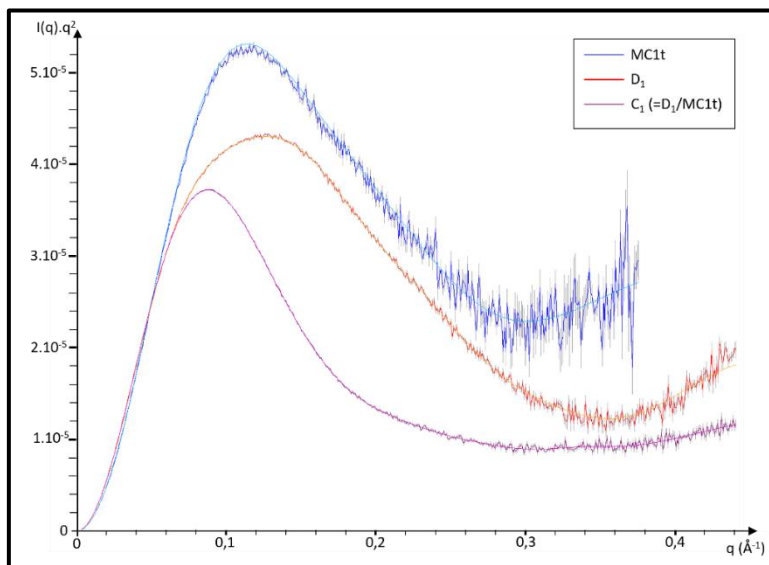


Figure 110 : *Superposition des tracés de Kratky de MC1t, de D₁ et de C₁ respectivement en bleu, en rouge et en violet.*

2.4.2. Calculs d'enveloppe

Suite à l'enregistrement et à la caractérisation des courbes de diffusion, nous avons pu, en utilisant la suite de programmes d'ATSAS^[190], réaliser différents calculs d'enveloppe et de courbe de diffusion pour la protéine MC1t libre et du complexe.

Enveloppes générées par DAMMIN

Nous avons, en première approche, utilisé le logiciel DAMMIN^[192] qui permet de générer des enveloppes lorsque rien n'est connu d'un objet hormis sa courbe de diffusion. Même si les modèles obtenus sont imprécis et hypothétiques, ils nous ont donné une idée de la forme globale des deux objets (Figure 111).

Les valeurs de Chi2 (χ^2) issues de l'affinement des courbes expérimentales et théoriques déterminées à partir des enveloppes générées sont proches de zéro (inférieures à 5). L'enveloppe étant générée uniquement à partir de la courbe de diffusion expérimentale, cette faible valeur était attendue et sert plutôt de contrôle positif au processus de calcul.

L'enveloppe générée pour MC1t ne peut pas être décrite comme étant globulaire. Toutefois, cette enveloppe contient une zone centrale relativement globulaire et deux extensions (Figure 111A). Ces extensions sont opposées l'une par rapport à l'autre et sont de tailles très différentes. En se basant sur la structure tridimensionnelle déterminée par RMN de la protéine libre (pdb : 2KHL), il semble évident que la zone centrale correspond au corps de la protéine et que l'extension dont la taille est la plus grande peut correspondre au bras (LP5) de la protéine dans sa conformation étendue. La seconde extension, pourrait correspondre aux quatre résidus en N-terminal (Gly(-3), Ser(-2), His(-1), et Met0) de la protéine MC1t qui sont absents de la séquence et donc de la structure de MC1.

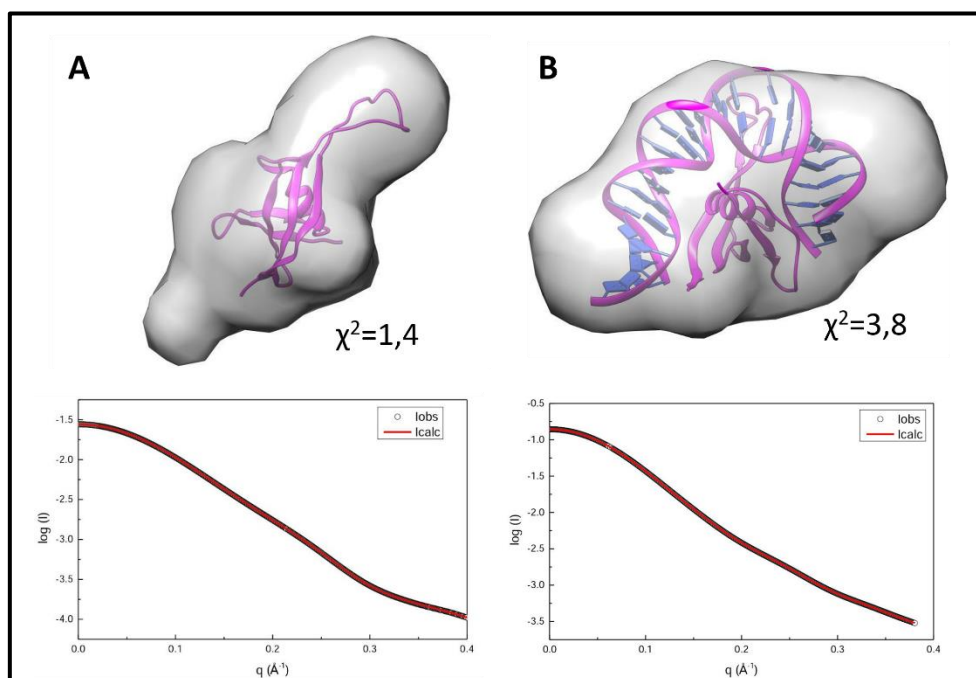


Figure 111 : *Enveloppe et courbe de diffusion générés par DAMMIN. A – Sur la protéine MC1t. B – Sur le complexe C₁.*

L'enveloppe générée pour le complexe est, comme attendu, plus globulaire que celle de la protéine libre. Cependant, il est très difficile de positionner le modèle du complexe dans cette enveloppe. Le positionnement, considéré comme étant le plus optimal semble indiquer, qu'en moyenne, l'ADN n'a pas la même courbure et/ou la même torsion que dans le modèle proposé. Cette enveloppe n'est cependant pas aberrante. D'après les résultats du Chapitre 1, deux hypothèses sont à évoquer et à discuter. La première hypothèse consiste à dire que, dans le complexe ADN^{23bp}/MC1, l'ADN se trouve dans deux conformations présentant, pour l'une, deux points de contact dans son petit sillon avec MC1 et, pour l'autre, trois points de contact. Si la durée d'existence de la première est supérieure à celle de la seconde, alors cette enveloppe est tout à fait compatible avec les données obtenues dans le chapitre précédent. Toutefois, il serait étonnant que la conformation ne contenant que deux points de contact puisse être considérée comme rigide. Ce qui semble en contradiction avec l'analyse du tracé de Kratky réalisé un peu plus haut. La seconde hypothèse consiste à dire que notre modèle est à revoir afin de prendre en compte cette enveloppe comme une nouvelle donnée. Cette hypothèse a déjà été soulevée dans le chapitre précédent.

Courbes de diffusion générées par CRY SOL

Afin d'aller plus loin dans la compréhension des données de SAXS, nous avons utilisé le logiciel CRY SOL ^[193]. Celui-ci permet de générer des courbes de diffusion à partir de structures tridimensionnelles obtenues par RMN ou de par diffraction de rayons X et de les comparer à des courbes de diffusion expérimentales (Figure 112).

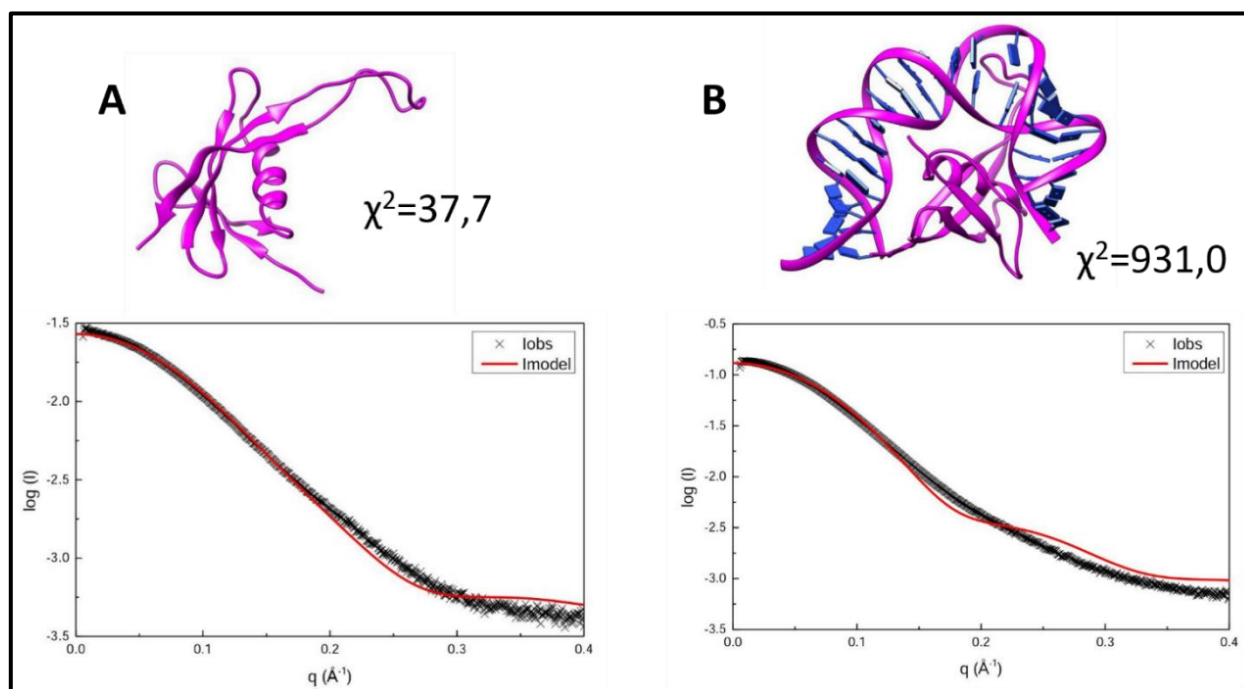


Figure 112 : Courbes de diffusion générées par CRYSOLOG. A – Sur la protéine MC1t à partir de la pdb 2KHL. B – Sur le complexe C₁ à partir du modèle de structure déterminé par modélisation moléculaire à partir des données RMN.

La courbe de diffusion de la protéine MC1 libre en solution a été calculée à partir du jeu des 20 structures déposées à la banque de données sur les protéines (pdb : 2KHL). La comparaison de cette courbe de diffusion « théorique » avec notre courbe SAXS expérimentale nous donne une valeur de Chi2 de **37,7** avec une parfaite superposition pour des valeurs de q inférieures à $0,2 \text{ \AA}^{-1}$ (Figure 112A). Pour des valeurs de q supérieures à $0,2 \text{ \AA}^{-1}$, on observe une différence non négligeable entre les deux courbes. Elle est certainement liée à deux paramètres qui ne peuvent pas être pris en compte par le logiciel. Premièrement, la courbe théorique ne prend en compte que 20 structures de la protéine MC1. La combinaison de ces 20 modèles représente bien, d'un point de vue structural, les différentes positions du bras qui découlent de sa flexibilité, mais le logiciel ne peut en aucun cas prendre en compte toutes les positions intermédiaires pouvant exister entre ces 20 modèles. Deuxièmement, les modèles ne contiennent pas les quatre résidus supplémentaires en N-terminal, Gly(-3), Ser(-2), His(-1), et Met0), présents dans MC1t. Sachant cela, la superposition des deux courbes est tout à fait satisfaisante et les données mesurées par SAXS sont en accord avec les données structurales. Ce résultat nous permet d'avoir un contrôle positif sur l'utilisation de ce logiciel et donc d'analyser les résultats obtenus sur le complexe.

Le modèle du complexe ADN^{23bp}/MC1, a été utilisé pour le calcul de la courbe de diffusion expérimentale donnée sur la Figure 112B. Cette courbe, superposée à la courbe SAXS de C₁, nous donne une valeur de Chi2 très élevée (**Chi2 = 931,0**). Nous observons que les deux courbes diffèrent dès que les valeurs de q sont supérieures à $0,1 \text{ \AA}^{-1}$. Ce résultat semble confirmer que le modèle proposé ne correspond pas à la structure du complexe en solution.

Partie 2. Étude structurale et dynamique d'une protéine associée au nucléoïde d'archée en interaction avec l'ADN : MC1

Enveloppes générées par MONSA

Pour finir, nous avons utilisé en troisième approche le logiciel MONSA ^[192]. Ce logiciel est particulièrement adapté au calcul d'enveloppes de complexes ADN/protéine. Il génère, tout comme DAMMIN, des enveloppes *ab initio* mais peut, en plus, utiliser deux phases pour son calcul, chacune générant l'enveloppe et la courbe de diffusion de l'un des objets du complexe.

L'enveloppe générée pour le complexe C₁ est présentée sur la [Figure 113](#). Comme pour les enveloppes générées par DAMMIN, les valeurs de Chi2 (χ^2) calculées sont proches de zéro (inférieures à 5) et nous servent de contrôle au bon déroulement du processus de calcul.

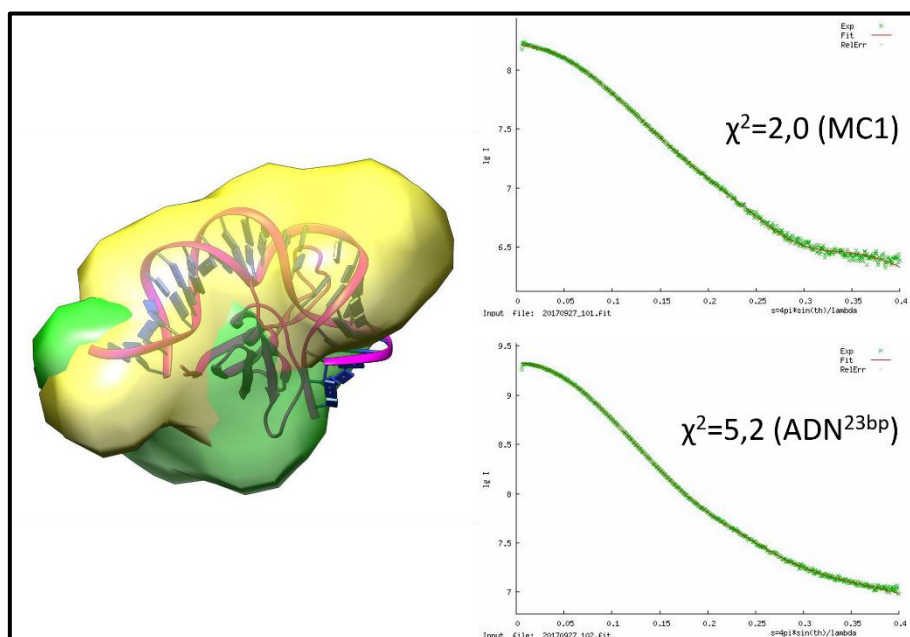


Figure 113 : Enveloppe et courbe de diffusion générés par MONSA sur le complexe C₁. En jaune et en vert sont représentés respectivement l'enveloppe générée pour la protéine MC1t (phase 1) et pour l'ADN D₁ (phase 2).

Cette enveloppe obtenue pour le complexe C₁ est quasi ellipsoïdale. La partie supérieure, en jaune sur la [Figure 113](#) correspond au duplex d'ADN D₁. La partie inférieure, en vert sur la [Figure 113](#) correspondant à la protéine MC1t.

En intégrant le modèle du complexe 23 bp au cœur de l'enveloppe, on constate que ni la protéine, ni l'ADN ne peuvent être positionnés de façon satisfaisante. De plus, le logiciel MONSA n'étant pas particulièrement adapté dans le cas de changements de conformation en solution, la modélisation d'enveloppes par ce logiciel est, pour la caractérisation de notre complexe, limitée. Nous devons donc finalement conclure qu'il faut construire un nouveau modèle afin de prendre en compte les enveloppes déterminées à partir des mesures expérimentales de SAXS.

2.5. Conclusion : vers une détermination électronique de la structure

L'ensemble des données et des résultats discutés dans ce chapitre nous a permis d'obtenir des pistes pour la résolution par diffraction des rayons X de la structure du complexe.

Le protocole optimisé pour la préparation du complexe ADN^{23bp}/MC1t nous a permis d'obtenir rapidement, et avec une très haute pureté, le complexe. Les essais de cristallogenèse nous ont permis de déterminer des conditions favorables à la formation des cristaux de complexe et pourront être encore améliorées (croissance de plus gros cristaux et meilleure conservation). Malgré tout, d'après les premiers essais de diffraction des rayons-X, ces cristaux sont encourageants pour déterminer la structure tridimensionnelle du complexe. Toutefois, il serait peut-être nécessaire de réaliser une optimisation de la séquence de l'ADN selon la méthode SELEX sur une longueur de 23bp comme cela avait été réalisé pour les 15 premières paires de bases. Cette méthode peut s'avérer longue et onéreuse.

Les données de SAXS nous ont permis de conclure que l'enveloppe du complexe ADN/MC1 pouvait être déterminée, mais qu'elle ne correspond pas au modèle structural proposé. Un nouveau modèle pourrait donc être généré afin de prendre en compte ces nouvelles données.

Conclusion générale

Conclusion générale

Les protéines architecturales régulent de nombreuses fonctions biologiques en lien avec l'ADN et sont des acteurs importants, existant dans les trois domaines du Vivant. Comprendre le fonctionnement de la machinerie génétique permet de mieux comprendre l'ensemble des mécanismes cellulaires. C'est dans la recherche de cette connaissance détaillée que s'inscrit mon travail de thèse. Il a été mené sur deux systèmes de protéines associées au nucléoïde : HU (protéine bactérienne, *E.coli*) et MC1 (protéine archéenne, *M. thermophila*).

HU est une protéine dimérique retrouvée chez toutes les bactéries. Elle est la protéine majoritaire associée au nucléoïde de ces dernières. Chez la bactérie *E. coli*, elle est codée par deux gènes, *hupA* et *hupB*, et existe sous trois formes dimériques qui **échan**gent leurs chaînes par un **mécanisme remarquable, mais complexe**. Par ce travail de thèse, nous avons pu compléter les connaissances du système d'échange de *EcHU* tant d'un point de vue structural que cinétique.

Nous nous sommes d'abord intéressés à la première étape du mécanisme d'échange de chaînes de la protéine *EcHU*, à savoir la **transition N₂/I₂ d'un point de vue structural**. Pour ce faire, cinq structures de la protéine ont été déterminées par RMN : la structure de chaque dimère dans sa conformation native à basse température, la structure de l'hétérodimère à sa température de transition (**Tm1**) dans sa conformation native et dans sa conformation intermédiaire. Au cours de cette étude, nous avons pu montrer que les résidus bordant la **poche hydrophobe**, présente entre la « selle » et les domaines HTH, dans la conformation native de la protéine se retrouvaient pour une partie d'entre eux, exposés au solvant. Au-delà du fait que les résidus de cette poche seraient des déterminants structuraux permettant d'expliquer que les trois dimères de *EcHU* présentent des propriétés thermodynamiques différentes, ceux-ci pourraient former le motif de reconnaissance des conformations intermédiaires et ainsi entraîner la formation du tétramère transitoire.

Nous avons ensuite cherché à caractériser les **étapes principales du mécanisme d'échange de chaînes** menant à la formation de l'hétérodimère d'un point de vue cinétique ($\alpha\alpha_N + \beta\beta_N \rightarrow \alpha\alpha_I + \beta\beta_I \rightarrow \alpha_2\beta_2 \rightarrow 2 \alpha\beta_N$). Cette étude nous a permis de caractériser la cinétique globale du système comme étant, quelle que soit la température, une **réaction totale d'ordre 2** dont l'étape limitante est la formation du tétramère transitoire. Même à basse température (10 °C), la formation de l'hétérodimère semble être limitée par la formation du tétramère et non par l'échange N₂/I₂.

Ces nouveaux éléments nous confortent dans l'idée que la forme tétramérique transitoire est la **clef** structurale et cinétique du mécanisme global d'échange de chaîne intrinsèque aux *EcHU* et probablement à l'ensemble des HU bactériennes. Ces dernières seraient capables de **s'auto-chaperonner** pour moduler leur interaction à l'ADN et augmenter leur temps de demi-vie *in vivo*.

Conclusion générale

La protéine **MC1** n'est retrouvée que chez les classes *Methanomicrobia* et *Halobacteria* d'Euryarchée. Elle est étudiée depuis une trentaine d'années au Centre de Biophysique Moléculaire par différents groupes de recherche afin de comprendre et de caractériser les mécanismes génétiques, tels que la transcription et la recombinaison de l'ADN, dans lesquelles elle est ou pourrait être impliquée. L'ensemble de l'étude, menée ces trente dernières années, associée aux données dynamiques et structurales réalisées sur la protéine complexée à un ADN de 15 paires de bases et 23 paires de bases, nous ont permis de proposer un **mécanisme de complexation en trois étapes**. Dans un premier temps, MC1 se rapproche de l'ADN et s'associe au petit sillon du A-tract grâce à ses résidus basiques. Dans un second temps, son bras flexible lui permet de stabiliser des conformations courbées de l'ADN en intercalant les chaînes latérales de la Pro72 et du Trp74 dans le petit sillon. MC1 peut également interagir avec des ADNs plus longs en établissant un **troisième point de contact**. Les données préliminaires, obtenues par cristallographie aux rayons X et par diffusion des rayons X de la structure du complexe avec un ADN plus long, n'ont pas permis de déterminer la structure tridimensionnelle du complexe. Elles nous ont toutefois permis de décrire l'**enveloppe du complexe** en solution et déterminer des **conditions de cristallogenèse** suffisantes pour aboutir à la caractérisation de cristaux du complexe à haute résolution.

L'ensemble de ces résultats, nous a permis de faire l'hypothèse que non seulement MC1 est un courbeur avéré de l'ADN, mais que cette petite protéine monomérique doit aussi avoir un mode de compaction secondaire en tant qu'enrouleur de l'ADN impliquant le troisième point de contact, tout comme c'est le cas pour la protéine HU chez les bactéries.

Références bibliographiques

1. Woese, C.R. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. 1990, 4.
2. Gaba, T. vector version: E. A phylogenetic tree of living things, based on RNA data and proposed by Carl Woese, showing the separation of bacteria, archaea, and eukaryotes (public domain illustration).; 2006.
3. Luijsterburg, M.S.; White, M.F.; van Driel, R.; Dame, R.Th. The Major Architects of Chromatin: Architectural Proteins in Bacteria, Archaea and Eukaryotes. *Critical Reviews in Biochemistry and Molecular Biology* 2008, 43(6), 393–418 10.1080/10409230802528488.
4. Champoux, J.J. DNA Topoisomerases: Structure, Function, and Mechanism. *Annual Review of Biochemistry* 2001, 70(1), 369–413 10.1146/annurev.biochem.70.1.369.
5. Ketron, A.C.; Osheroff, N. DNA Topology and Topoisomerases. In *Molecular Life Sciences*; E. Bell, Ed.; Springer New York: New York, NY, 2014, 1–19 10.1007/978-1-4614-6436-5_127-3.
6. Dorman, C.J. Function of Nucleoid-Associated Proteins in Chromosome Structuring and Transcriptional Regulation. *Journal of Molecular Microbiology and Biotechnology* 2014, 24(5–6), 316–331 10.1159/000368850.
7. Browning, D.F.; Grainger, D.C.; Busby, S.J. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Current Opinion in Microbiology* 2010, 13(6), 773–780 10.1016/j.mib.2010.09.013.
8. Winardhi, R.S.; Yan, J.; Kenney, L.J. H-NS Regulates Gene Expression and Compacts the Nucleoid: Insights from Single-Molecule Experiments. *Biophysical Journal* 2015, 109(7), 1321–1329 10.1016/j.bpj.2015.08.016.
9. Luger, K.; Mäder, A.W.; Richmond, R.K.; Sargent, D.F.; Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 1997, 389(6648), 251 10.1038/38444.
10. Kornberg, R.D. Structure of chromatin. *Annual Review of Biochemistry* 1977, 46, 931–954 10.1146/annurev.bi.46.070177.004435.
11. Crane-Robinson, C. Linker histones: History and current perspectives. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 2016, 1859(3), 431–435 10.1016/j.bbagr.2015.10.008.
12. Reeves, R. Nuclear functions of the HMG proteins. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 2010, 1799(1), 3–14 10.1016/j.bbagr.2009.09.001.
13. Postnikov, Y.V.; Bustin, M. Functional interplay between histone H1 and HMG proteins in chromatin. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 2016, 1859(3), 462–467 10.1016/j.bbagr.2015.10.006.
14. Reeves, R. High mobility group (HMG) proteins: Modulators of chromatin structure and DNA repair in mammalian cells. *DNA Repair* 2015, 36, 122–136 10.1016/j.dnarep.2015.09.015.
15. Verma, S.C.; Qian, Z.; Adhya, S.L. Architecture of the Escherichia coli nucleoid. *PLoS Genetics* 2019, 15(12) 10.1371/journal.pgen.1008456.
16. Dillon, S.C.; Dorman, C.J. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology* 2010, 8(3), 185–195 10.1038/nrmicro2261.
17. Tolstorukov, M.Y.; Virnik, K.; Zhurkin, V.B.; Adhya, S. Organization of DNA in a bacterial nucleoid. *BMC Microbiology* 2016, 16(1), 22 10.1186/s12866-016-0637-3.
18. Bell, S.D.; White, M.F. Archaeal Chromatin Organization. In *Bacterial Chromatin*; R.T. Dame, C.J. Dorman, Eds.; Springer Netherlands: Dordrecht, 2010, 205–217 10.1007/978-90-481-3473-1_10.
19. Zhang, Z.; Guo, L.; Huang, L. Archaeal chromatin proteins. *Science China Life Sciences* 2012, 55(5), 377–385 10.1007/s11427-012-4322-y.
20. Brunk, C.F.; Martin, W.F. Archaeal Histone Contributions to the Origin of Eukaryotes. *Trends in Microbiology* 2019, 27(8), 703–714 10.1016/j.tim.2019.04.002.
21. Toulmé, F.; Cam, E.L.; Teyssier, C.; Delain, E.; Sautière, P.; Maurizot, J.-C.; et al. Conformational Changes of DNA Minicircles upon the Binding of the Archaeobacterial Histone-like Protein MC1. *Journal of Biological Chemistry* 1995, 270(11), 6286–6291 10.1074/jbc.270.11.6286.

22. Paquet, F.; Loth, K.; Meudal, H.; Culard, F.; Genest, D.; Lancelot, G. Refined solution structure and backbone dynamics of the archaeal MC1 protein. *The FEBS Journal* 2010, 277(24), 5133–5145 10.1111/j.1742-4658.2010.07927.x.
23. Paquet, F.; Delalande, O.; Goffinont, S.; Culard, F.; Loth, K.; Asseline, U.; et al. Model of a DNA-Protein Complex of the Architectural Monomeric Protein MC1 from Euryarchaea. *PLOS ONE* 2014, 9(2), e88809 10.1371/journal.pone.0088809.
24. Loth, K.; Landon, C.; Paquet, F. Chemical shifts assignments of the archaeal MC1 protein and a strongly bent 15 base pairs DNA duplex in complex. *Biomolecular NMR Assignments* 2015, 9(1), 215–217 10.1007/s12104-014-9577-8.
25. Loth, K.; Largillière, J.; Coste, F.; Culard, F.; Landon, C.; Castaing, B.; et al. New protein-DNA complexes in archaea: a small monomeric protein induces a sharp V-turn DNA structure. *Scientific Reports* 2019, 9(1), 1–12 10.1038/s41598-019-50211-2.
26. Rouvière-Yaniv, J.; Gros, F. Characterization of a novel, low-molecular-weight DNA-binding protein from *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 1975, 72(9), 3428–3432.
27. Alberts, B.; Herrick, G. [11] DNA-cellulose chromatography. In *Methods in Enzymology*; Academic Press: 1971, 198–217 10.1016/S0076-6879(71)21013-2.
28. Itzhaki, R.F.; Gill, D.M. A micro-biuret method for estimating proteins. *Analytical Biochemistry* 1964, 9(4), 401–410 10.1016/0003-2697(64)90200-3.
29. Orfaniotou, F.; Tzamalīs, P.; Thanassoulas, A.; Stefanidi, E.; Zees, A.; Boutou, E.; et al. The stability of the archaeal HU histone-like DNA-binding protein from *Thermoplasma volcanium*. *Extremophiles* 2008, 13(1), 1 10.1007/s00792-008-0190-6.
30. Datta, C.; Jha, R.K.; Ahmed, W.; Ganguly, S.; Ghosh, S.; Nagaraja, V. Physical and functional interaction between nucleoid-associated proteins HU and Lsr2 of *Mycobacterium tuberculosis*: altered DNA binding and gene regulation. *Molecular Microbiology* 2019, 111(4), 981–994 10.1111/mmi.14202.
31. Oberto, J.; Nabti, S.; Jooste, V.; Mignot, H.; Rouvière-Yaniv, J. The HU Regulon Is Composed of Genes Responding to Anaerobiosis, Acid Stress, High Osmolarity and SOS Induction. *PLoS ONE* A. Imhof, Ed.; 2009, 4(2), e4367 10.1371/journal.pone.0004367.
32. Wada, M.; Kano, Y.; Ogawa, T.; Okazaki, T.; Imamoto, F. Construction and characterization of the deletion mutant of hupA and hupB genes in *Escherichia coli*. *Journal of Molecular Biology* 1988, 204(3), 581–591 10.1016/0022-2836(88)90357-9.
33. Dri, A.M.; Rouvière-Yaniv, J.; Moreau, P.L. Inhibition of cell division in hupA hupB mutant bacteria lacking HU protein. *Journal of Bacteriology* 1991, 173(9), 2852–2863.
34. Castaing, B.; Zelwer, C.; Laval, J.; Boiteux, S. HU Protein of *Escherichia coli* Binds Specifically to DNA That Contains Single-strand Breaks or Gaps. *Journal of Biological Chemistry* 1995, 270(17), 10291–10296 10.1074/jbc.270.17.10291.
35. Burroughs, A.M.; Kaur, G.; Zhang, D.; Aravind, L. Novel clades of the HU/IHF superfamily point to unexpected roles in the eukaryotic centrosome, chromosome partitioning, and biologic conflicts. *Cell Cycle* 2017, 16(11), 1093–1103 10.1080/15384101.2017.1315494.
36. Dey, D.; Nagaraja, V.; Ramakumar, S. Structural and evolutionary analyses reveal determinants of DNA binding specificities of nucleoid-associated proteins HU and IHF. *Molecular Phylogenetics and Evolution* 2017, 107, 356–366 10.1016/j.ympev.2016.11.014.
37. Kamashev, D.; Agapova, Y.; Rastorguev, S.; Talyzina, A.A.; Boyko, K.M.; Korzhenevskiy, D.A.; et al. Comparison of histone-like HU protein DNA-binding properties and HU/IHF protein sequence alignment. *PLOS ONE* D. Chatterji, Ed.; 2017, 12(11), e0188037 10.1371/journal.pone.0188037.
38. Grove, A. Functional evolution of bacterial histone-like HU proteins. *Current Issues in Molecular Biology* 2011, 13(1), 1–12.
39. Mende, L.; Timm, B.; Subramanian, A.R. Primary structures of two homologous ribosome-associated DNA-binding proteins of *Escherichia coli*. *FEBS Letters* 1978, 96(2), 395–398 10.1016/0014-5793(78)80446-3.
40. Laine, B.; Kmiecik, D.; Sautiere, P.; Biserte, G.; Cohen-Solal, M. Complete Amino-Acid Sequences of DNA-Binding Proteins HU-1 and HU-2 from *Escherichia coli*. *European Journal of Biochemistry* 1980, 103(3), 447–461 10.1111/j.1432-1033.1980.tb05968.x.

41. Swinger, K.K.; Rice, P.A. IHF and HU: flexible architects of bent DNA. *Current Opinion in Structural Biology* 2004, 14(1), 28–35 10.1016/j.sbi.2003.12.003.
42. Swinger, K.K.; Lemberg, K.M.; Zhang, Y.; Rice, P.A. Flexible DNA bending in HU–DNA cocrystal structures. *The EMBO Journal* 2003, 22(14), 3749–3760 10.1093/emboj/cdg351.
43. Pontiggia, A.; Negri, A.; Beltrame, M.; Bianchi, M.E. Protein HU binds specifically to kinked DNA. *Molecular Microbiology* 1993, 7(3), 343–350 10.1111/j.1365-2958.1993.tb01126.x.
44. Kamashev, D.; Rouviere-Yaniv, J. The histone-like protein HU binds specifically to DNA recombination and repair intermediates. *The EMBO Journal* 2000, 19(23), 6527–6535 10.1093/emboj/19.23.6527.
45. Bonnefoy, E.; Takahashi, M.; Yaniv, J.R. DNA-binding Parameters of the HU Protein of *Escherichia coli* to Cruciform DNA. *Journal of Molecular Biology* 1994, 242(2), 116–129 10.1006/jmbi.1994.1563.
46. Tanaka, H.; Yasuzawa, K.; Kohno, K.; Goshima, N.; Kano, Y.; Saiki, T.; et al. Role of HU proteins in forming and constraining supercoils of chromosomal DNA in *Escherichia coli*. *Molecular & general genetics: MGG* 1995, 248(5), 518–526 10.1007/bf02423446.
47. Kamashev, D.; Balandina, A.; Rouviere-Yaniv, J. The binding motif recognized by HU on both nicked and cruciform DNA. *The EMBO Journal* 1999, 18(19), 5434–5444 10.1093/emboj/18.19.5434.
48. Guo, F.; Adhya, S. Spiral structure of *Escherichia coli* HU $\alpha\beta$ provides foundation for DNA supercoiling. *Proceedings of the National Academy of Sciences of the United States of America* 2007, 104(11), 4309–4314 10.1073/pnas.0611686104.
49. Papavassiliou, A.G.; Heumann, H. 1,10-Phenanthroline-Copper Ion Nuclease Footprinting of DNA-Protein Complexes in Situ Following Mobility-Shift Electrophoresis Assays. In *DNA-Protein Interactions: Principles and Protocols*; G. Geoff Kneale, Ed.; Humana Press: Totowa, NJ, 1994, 43–78 10.1385/0-89603-256-6:43.
50. Tanaka, I.; Appelt, K.; Dijk, J.; White, S.W.; Wilson, K.S. 3-Å resolution structure of a protein with histone-like properties in prokaryotes. *Nature* 1984, 310(5976), 376–381 10.1038/310376a0.
51. Broyles, S.S.; Pettijohn, D.E. Interaction of the *Escherichia coli* HU protein with DNA: Evidence for formation of nucleosome-like structures with altered DNA helical pitch. *Journal of Molecular Biology* 1986, 187(1), 47–60 10.1016/0022-2836(86)90405-5.
52. Oberto, J.; Rouviere-Yaniv, J. *Serratia marcescens* contains a heterodimeric HU protein like *Escherichia coli* and *Salmonella typhimurium*. *Journal of Bacteriology* 1996, 178(1), 293–297.
53. Phan, N.Q.; Uebanso, T.; Shimohata, T.; Nakahashi, M.; Mawatari, K.; Takahashi, A. DNA-Binding Protein HU Coordinates Pathogenicity in *Vibrio parahaemolyticus*. *Journal of Bacteriology* 2015, 197(18), 2958–2964 10.1128/JB.00306-15.
54. Kano, Y.; Osato, K.; Wada, M.; Imamoto, F. Cloning and sequencing of the HU-2 gene of *Escherichia coli*. *Molecular and General Genetics MGG* 1987, 209(2), 408–410 10.1007/BF00329674.
55. Kano, Y.; Yoshino, S.; Wada, M.; Yokoyama, K.; Nobuhara, M.; Imamoto, F. Molecular cloning and nucleotide sequence of the HU-1 gene of *Escherichia coli*. *Molecular & general genetics: MGG* 1985, 201(2), 360–362 10.1007/bf00425687.
56. Rouvière-Yaniv, J.; Kjeldgaard, N.O. Native *Escherichia coli* HU protein is a heterotypic dimer. *FEBS Letters* 1979, 106(2), 297–300 10.1016/0014-5793(79)80518-9.
57. Ramstein, J.; Hervouet, N.; Coste, F.; Zelwer, C.; Oberto, J.; Castaing, B. Evidence of a Thermal Unfolding Dimeric Intermediate for the *Escherichia coli* Histone-like HU Proteins: Thermodynamics and Structure. *Journal of Molecular Biology* 2003, 331(1), 101–121 10.1016/S0022-2836(03)00725-3.
58. Pinson, V.; Takahashi, M.; Rouviere-Yaniv, J. Differential binding of the *Escherichia coli* HU, homodimeric forms and heterodimeric form to linear, gapped and cruciform DNA 11 Edited by J. Karn. *Journal of Molecular Biology* 1999, 287(3), 485–497 10.1006/jmbi.1999.2631.
59. Kohno, K.; Wada, M.; Kano, Y.; Imamoto, F. Promoters and autogenous control of the *Escherichia coli* hupA and hupB genes. *Journal of Molecular Biology* 1990, 213(1), 27–36 10.1016/S0022-2836(05)80119-6.
60. Claret, L.; Rouviere-Yaniv, J. Regulation of HU α and HU β by CRP and FIS in *Escherichia coli*. *Journal of Molecular Biology* 1996, 263(2), 126–139 10.1006/jmbi.1996.0564.

61. Boubrik, F.; Rouviere-Yaniv, J. Increased sensitivity to gamma irradiation in bacteria lacking protein HU. *Proceedings of the National Academy of Sciences of the United States of America* 1995, 92(9), 3958–3962.
62. Giangrossi, M.; Giuliodori, A.M.; Gualerzi, C.O.; Pon, C.L. Selective expression of the beta-subunit of nucleoid-associated protein HU during cold shock in *Escherichia coli*. *Molecular Microbiology* 2002, 44(1), 205–216.
63. Claret, L.; Rouviere-Yaniv, J. Variation in HU composition during growth of *Escherichia coli*: the heterodimer is required for long term survival¹¹Edited by M. Gottesman. *Journal of Molecular Biology* 1997, 273(1), 93–104 10.1006/jmbi.1997.1310.
64. Brandi, A.; Spurio, R.; Gualerzi, C.O.; Pon, C.L. Massive presence of the *Escherichia coli* “major cold-shock protein” CspA under non-stress conditions. *The EMBO journal* 1999, 18(6), 1653–1659 10.1093/emboj/18.6.1653.
65. Le Meur, R.; Loth, K.; Culard, F.; Castaing, B.; Landon, C. Backbone assignment of the three dimers of HU from *Escherichia coli* at 293 K: EchU α 2, EchU β 2 and EchU $\alpha\beta$. *Biomolecular NMR Assignments* 2015, 9(2), 359–363 10.1007/s12104-015-9610-6.
66. Drlica, K.; Rouviere-Yaniv, J. Histone-like proteins of bacteria. *Microbiological Reviews* 1987, 51(3), 301–319.
67. Bonnefoy, E.; Rouvière-Yaniv, J. HU and IHF, two homologous histone-like proteins of *Escherichia coli*, form different protein-DNA complexes with short DNA fragments. *The EMBO Journal* 1991, 10(3), 687–696.
68. Garnier, N.; Loth, K.; Coste, F.; Augustyniak, R.; Nadan, V.; Damblon, C.; et al. An alternative flexible conformation of the *E. coli* HU β ₂ protein: structural, dynamics, and functional aspects. *European Biophysics Journal* 2011, 40(2), 117–129 10.1007/s00249-010-0630-y.
69. Le Meur, R. Etude structurale du mécanisme d'échange de chaînes des dimères de la protéine HU d'*E. coli*. [thesis] Orléans: 2015.
70. Cheung, M.-S.; Maguire, M.L.; Stevens, T.J.; Broadhurst, R.W. DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *Journal of Magnetic Resonance* 2010, 202(2), 223–233 10.1016/j.jmr.2009.11.008.
71. Vranken, W.F.; Boucher, W.; Stevens, T.J.; Fogh, R.H.; Pajon, A.; Llinas, M.; et al. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function, and Bioinformatics* 2005, 59(4), 687–696 10.1002/prot.20449.
72. Linge, J.P.; Habeck, M.; Rieping, W.; Nilges, M. ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* 2003, 19(2), 315–316 10.1093/bioinformatics/19.2.315.
73. Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J.M.; Liu, G.; et al. Consistent blind protein structure generation from NMR chemical shift data. *Proceedings of the National Academy of Sciences of the United States of America* 2008, 105(12), 4685–4690 10.1073/pnas.0800256105.
74. Cavanagh, J.; Fairbrother, W.J.; Palmer, A.G.; Rance, M.; Skelton, N.J. *Protein NMR Spectroscopy*; Elsevier: 2007, 10.1016/B978-0-12-164491-8.X5000-3.
75. Berjanskii, M.V.; Neal, S.; Wishart, D.S. PREDITOR: a web server for predicting protein torsion angle restraints. *Nucleic Acids Research* 2006, 34(suppl_2), W63–W69 10.1093/nar/gkl341.
76. Cornilescu, G.; Delaglio, F.; Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of biomolecular NMR* 1999, 13(3), 289–302 10.1023/a:1008392405740.
77. Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR* 2009, 44(4), 213–223 10.1007/s10858-009-9333-z.
78. Chen, K.; Tjandra, N. The Use of Residual Dipolar Coupling in Studying Proteins by NMR. In *Topics in current chemistry*; 2012, 47–67.
79. Brünger, A.T.; Adams, P.D.; Clore, G.M.; DeLano, W.L.; Gros, P.; Grosse-Kunstleve, R.W.; et al. Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallographica Section D Biological Crystallography* 1998, 54(5), 905–921 10.1107/S0907444498003254.
80. Brünger, A.T. Version 1.2 of the Crystallography and NMR system. *Nature Protocols* 2007, 2(11), 2728–2733 10.1038/nprot.2007.406.

81. Salomon-Ferrer, R.; Case, D.A.; Walker, R.C. An overview of the Amber biomolecular simulation package. *WIREs Computational Molecular Science* 2013, 3(2), 198–210 10.1002/wcms.1121.
82. Brooks, B.R.; Brucoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* 1983, 4(2), 187–217 10.1002/jcc.540040211.
83. Brooks, B.R.; Brooks, C.L.; MacKerell, A.D.; Nilsson, L.; Petrella, R.J.; Roux, B.; et al. CHARMM: The Biomolecular Simulation Program. *Journal of computational chemistry* 2009, 30(10), 1545–1614 10.1002/jcc.21287.
84. Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S.J.; et al. A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *Journal of Biomolecular NMR* 1994, 4(2), 301–306 10.1007/BF00175254.
85. Kay, L.E.; Xu, G.Y.; Yamazaki, T. Enhanced-Sensitivity Triple-Resonance Spectroscopy with Minimal H₂O Saturation. *Journal of Magnetic Resonance, Series A* 1994, 109(1), 129–133 10.1006/jmra.1994.1145.
86. Lescop, E.; Schanda, P.; Brutscher, B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *Journal of Magnetic Resonance* 2007, 187(1), 163–169 10.1016/j.jmr.2007.04.002.
87. Muhandiram, D.R.; Kay, L.E. Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *Journal of Magnetic Resonance, Series B* 1994, 103(3), 203–216 10.1006/jmrb.1994.1032.
88. Carlomagno, T.; Maurer, M.; Sattler, M.; Schwendinger, M.G.; Glaser, S.J.; Griesinger, C. PLUSH TACSy: Homonuclear planar TACSy with two-band selective shaped pulses applied to C α ,C' transfer and C β ,Caromatic correlations. *Journal of Biomolecular NMR* 1996, 8(2), 161–170 10.1007/BF00211162.
89. Kay, L.E.; Xu, G.Y.; Singer, A.U.; Muhandiram, D.R.; Formankay, J.D. A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-Chain ¹H and ¹³C Correlations in H₂O Samples of Proteins. *Journal of Magnetic Resonance, Series B* 1993, 101(3), 333–337 10.1006/jmrb.1993.1053.
90. Davis, A.L.; Keeler, J.; Laue, E.D.; Moskau, D. Experiments for recording pure-absorption heteronuclear correlation spectra using pulsed field gradients. *Journal of Magnetic Resonance* (1969) 1992, 98(1), 207–216 10.1016/0022-2364(92)90126-R.
91. Hwang, T.L.; Shaka, A.J. Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *Journal of Magnetic Resonance, Series A* 1995, 112(2), 275–279 10.1006/jmra.1995.1047.
92. Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 1993, 26(2), 283–291 10.1107/S0021889892009944.
93. Laskowski, R.A.; Rullmann, J.A.C.; MacArthur, M.W.; Kaptein, R.; Thornton, J.M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR* 1996, 8(4), 477–486 10.1007/BF00228148.
94. Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Research* 2018, 46(W1), W363–W367 10.1093/nar/gky473.
95. Krissinel, E.; Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology* 2007, 372(3), 774–797 10.1016/j.jmb.2007.05.022.
96. Hutchinson, E.G.; Thornton, J.M. PROMOTIF--a program to identify and analyze structural motifs in proteins. *Protein Science : A Publication of the Protein Society* 1996, 5(2), 212–220.
97. Mittermaier, A.K.; Kay, L.E. Observing biological dynamics at atomic resolution using NMR. *Trends in Biochemical Sciences* 2009, 34(12), 601–611 10.1016/j.tibs.2009.07.004.
98. Kleckner, I.R.; Foster, M.P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et biophysica acta* 2011, 1814(8), 942–968 10.1016/j.bbapap.2010.10.012.
99. Tollinger, M.; Skrynnikov, N.R.; Mulder, F.A.A.; Forman-Kay, J.D.; Kay, L.E. Slow Dynamics in Folded and Unfolded States of an SH3 Domain. *Journal of the American Chemical Society* 2001, 123(46), 11341–11352 10.1021/ja011300z.

100. Bieri, M.; Gooley, P.R. Automated NMR relaxation dispersion data analysis using NESSY. *BMC Bioinformatics* 2011, 12, 421 10.1186/1471-2105-12-421.
101. Palmer, A.G.; Kroenke, C.D.; Patrick Loria, J. [10] - Nuclear Magnetic Resonance Methods for Quantifying Microsecond-to-Millisecond Motions in Biological Macromolecules. In *Methods in Enzymology*; T.L. James, V. Dötsch, U. Schmitz, Eds.; Academic Press: 2001, 204–238 10.1016/S0076-6879(01)39315-1.
102. Schanda, P.; Kupče, Ě.; Brutscher, B. SOFAST-HMQC Experiments for Recording Two-dimensional Deteronuclear Correlation Spectra of Proteins within a Few Seconds. *Journal of Biomolecular NMR* 2005, 33(4), 199–211 10.1007/s10858-005-4425-x.
103. Christodoulou, E.; Vorgias, C.E. The thermostability of DNA-binding protein HU from mesophilic, thermophilic, and extreme thermophilic bacteria. *Extremophiles* 2002, 6(1), 21–31 10.1007/s007920100235.
104. Welfle, H.; Misselwitz, R.; Welfle, K.; Groch, N.; Heinemann, U. Salt-dependent and protein-concentration-dependent changes in the solution structure of the DNA-binding histone-like protein, HBSu, from *Bacillus subtilis*. *European Journal of Biochemistry* 1992, 204(3), 1049–1055 10.1111/j.1432-1033.1992.tb16727.x.
105. Wilson, K.S.; Vorgias, C.E.; Tanaka, I.; White, S.W.; Kimura, M. The thermostability of DNA-binding protein HU from bacilli. *Protein Engineering, Design and Selection* 1990, 4(1), 11–22 10.1093/protein/4.1.11.
106. Altukhov, D.A.; Talyzina, A.A.; Agapova, Y.K.; Vlaskina, A.V.; Korzhenevskiy, D.A.; Bocharov, E.V.; et al. Enhanced conformational flexibility of the histone-like (HU) protein from *Mycoplasma gallisepticum*. *Journal of Biomolecular Structure & Dynamics* 2016, 36(1), 45–53 10.1080/07391102.2016.1264893.
107. Christodoulou, E.; Vorgias, C.E. Cloning, overproduction, purification and crystallization of the DNA binding protein HU from the hyperthermophilic eubacterium *Thermotoga maritima*. *Acta Crystallographica Section D* 1998, 54(5), 1043–1045 10.1107/S0907444998000341.
108. Boyko, K.M.; Rakitina, T.V.; Korzhenevskiy, D.A.; Vlaskina, A.V.; Agapova, Y.K.; Kamashev, D.E.; et al. Structural basis of the high thermal stability of the histone-like HU protein from the mollicute *Spiroplasma melliferum* KC3. *Scientific Reports* 2016, 6 10.1038/srep36366.
109. Kawamura, S.; Kakuta, Y.; Tanaka, I.; Hikichi, K.; Kuhara, S.; Yamasaki, N.; et al. Glycine-15 in the Bend between Two α -Helices Can Explain the Thermostability of DNA Binding Protein HU from *Bacillus stearothermophilus*. *Biochemistry* 1996, 35(4), 1195–1200 10.1021/bi951581l.
110. Kawamura, S.; Abe, Y.; Ueda, T.; Masumoto, K.; Imoto, T.; Yamasaki, N.; et al. Investigation of the Structural Basis for Thermostability of DNA-binding Protein HU from *Bacillus stearothermophilus*. *Journal of Biological Chemistry* 1998, 273(32), 19982–19987 10.1074/jbc.273.32.19982.
111. Ruiz-Sanz, J.; Filimonov, V.V.; Christodoulou, E.; Vorgias, C.E.; Mateo, P.L. Thermodynamic analysis of the unfolding and stability of the dimeric DNA-binding protein HU from the hyperthermophilic eubacterium *Thermotoga maritima* and its E34D mutant. *European Journal of Biochemistry* 2004, 271(8), 1497–1507 10.1111/j.1432-1033.2004.04057.x.
112. Straub, C.T.; Counts, J.A.; Nguyen, D.M.N.; Wu, C.-H.; Zeldes, B.M.; Crosby, J.R.; et al. Biotechnology of extremely thermophilic archaea. *FEMS Microbiology Reviews* 2018, 42(5), 543–578 10.1093/femsre/fuy012.
113. Gunde-Cimerman, N.; Plemenitaš, A.; Oren, A. Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. *FEMS Microbiology Reviews* 2018, 42(3), 353–375 10.1093/femsre/fuy009.
114. Adam, P.S.; Borrel, G.; Brochier-Armanet, C.; Gribaldo, S. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *The ISME Journal* 2017, 11(11), 2407–2425 10.1038/ismej.2017.122.
115. Rosenberg, E.; DeLong, E.F.; Lory, S.; Stackebrandt, E.; Thompson, F., eds. *The Prokaryotes*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2014, 10.1007/978-3-642-38954-2.
116. Prieur, D. *Les bactéries de l'extrême*; De Boeck Supérieur: 2017.
117. Woese, C.R.; Fox, G.E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences* 1977, 74(11), 5088–5090 10.1073/pnas.74.11.5088.

118. Kates, M. The phytanyl ether-linked polar lipids and isoprenoid neutral lipids of extremely halophilic bacteria. *Progress in the Chemistry of Fats and other Lipids* 1977, 15(4), 301–342 10.1016/0079-6832(77)90011-8.
119. Koga, Y.; Morii, H. Biosynthesis of Ether-Type Polar Lipids in Archaea and Evolutionary Considerations. *Microbiology and Molecular Biology Reviews* 2007, 71(1), 97–120 10.1128/MMBR.00033-06.
120. Pohlschroder, M.; Pfeiffer, F.; Schulze, S.; Halim, M.F.A. Archaeal cell surface biogenesis. *FEMS Microbiology Reviews* 2018, 42(5), 694–717 10.1093/femsre/fuy027.
121. Eme, L.; Spang, A.; Lombard, J.; Stairs, C.W.; Ettema, T.J.G. Archaea and the origin of eukaryotes. *Nature Reviews Microbiology* 2017, 15(12), 711–723 10.1038/nrmicro.2017.133.
122. Jordan, S.F.; Nee, E.; Lane, N. Isoprenoids enhance the stability of fatty acid membranes at the emergence of life potentially leading to an early lipid divide. *Interface Focus* 2019, 9(6) 10.1098/rsfs.2019.0067.
123. Fukui, T.; Atomi, H.; Kanai, T.; Matsumi, R.; Fujiwara, S.; Imanaka, T. Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Research* 2005, 15(3), 352–363 10.1101/gr.3003105.
124. Kublanov, I.V.; Bidjjeva, S.Kh.; Mardanov, A.V.; Bonch-Osmolovskaya, E.A. *Desulfurococcus kamchatkensis* sp. nov., a novel hyperthermophilic protein-degrading archaeon isolated from a Kamchatka hot spring. *International Journal of Systematic and Evolutionary Microbiology*, 2009, 59(7), 1743–1747 10.1099/ijs.0.006726-0.
125. Barns, S.M.; Delwiche, C.F.; Palmer, J.D.; Pace, N.R. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proceedings of the National Academy of Sciences of the United States of America* 1996, 93(17), 9188–9193.
126. Elkins, J.G.; Podar, M.; Graham, D.E.; Makarova, K.S.; Wolf, Y.; Randau, L.; et al. A korarchaeal genome reveals insights into the evolution of the Archaea. *Proceedings of the National Academy of Sciences* 2008, 105(23), 8102–8107 10.1073/pnas.0801980105.
127. Auchtung, T.A.; Shyndriayeva, G.; Cavanaugh, C.M. 16S rRNA phylogenetic analysis and quantification of Korarchaeota indigenous to the hot springs of Kamchatka, Russia. *Extremophiles* 2011, 15(1), 105–116 10.1007/s00792-010-0340-5.
128. Huber, H.; Hohn, M.J.; Rachel, R.; Fuchs, T.; Wimmer, V.C.; Stetter, K.O. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 2002, 417(6884), 63–67 10.1038/417063a.
129. Brochier-Armanet, C.; Boussau, B.; Gribaldo, S.; Forterre, P. Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nature Reviews Microbiology* 2008, 6(3), 245–252 10.1038/nrmicro1852.
130. Tournai, M.; Stieglmeier, M.; Spang, A.; Könneke, M.; Schintlmeister, A.; Urich, T.; et al. *Nitrososphaera viennensis*, an ammonia oxidizing archaeon from soil. *Proceedings of the National Academy of Sciences* 2011, 108(20), 8420–8425 10.1073/pnas.1013488108.
131. Spang, A.; Hatzenpichler, R.; Brochier-Armanet, C.; Rattei, T.; Tischler, P.; Spieck, E.; et al. Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. *Trends in Microbiology* 2010, 18(8), 331–340 10.1016/j.tim.2010.06.003.
132. Brochier-Armanet, C.; Gribaldo, S.; Forterre, P. A DNA topoisomerase IB in Thaumarchaeota testifies for the presence of this enzyme in the last common ancestor of Archaea and Eucarya. *Biology Direct* 2008, 3, 54 10.1186/1745-6150-3-54.
133. Nunoura, T.; Takaki, Y.; Kakuta, J.; Nishi, S.; Sugahara, J.; Kazama, H.; et al. Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Research* 2011, 39(8), 3204–3223 10.1093/nar/gkq1228.
134. Hua, Z.-S.; Qu, Y.-N.; Zhu, Q.; Zhou, E.-M.; Qi, Y.-L.; Yin, Y.-R.; et al. Genomic inference of the metabolism and evolution of the archaeal phylum Aigarchaeota. *Nature Communications* 2018, 9(1), 1–11 10.1038/s41467-018-05284-4.
135. Spang, A.; Caceres, E.F.; Ettema, T.J.G. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* 2017, 357(6351) 10.1126/science.aaf3883.
136. Guy, L.; Ettema, T.J.G. The archaeal “TACK” superphylum and the origin of eukaryotes. *Trends in Microbiology* 2011, 19(12), 580–587 10.1016/j.tim.2011.09.002.

137. Dombrowski, N.; Lee, J.-H.; Williams, T.A.; Offre, P.; Spang, A. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiology Letters* 2019, 366(2) 10.1093/femsle/fnz008.
138. Spang, A.; Saw, J.H.; Jørgensen, S.L.; Zaremba-Niedzwiedzka, K.; Martijn, J.; Lind, A.E.; et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 2015, 521(7551), 173–179 10.1038/nature14447.
139. MacLeod, F.; Kindler, G.S.; Wong, H.L.; Chen, R.; Burns, B.P. Asgard archaea: Diversity, function, and evolutionary implications in a range of microbiomes. *AIMS Microbiology* 2019, 5(1), 48–61 10.3934/microbiol.2019.1.48.
140. Zaremba-Niedzwiedzka, K.; Caceres, E.F.; Saw, J.H.; Bäckström, D.; Juzokaite, L.; Vancaester, E.; et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 2017, 541(7637), 353–358 10.1038/nature21031.
141. Spang, A.; Eme, L.; Saw, J.H.; Caceres, E.F.; Zaremba-Niedzwiedzka, K.; Lombard, J.; et al. Asgard archaea are the closest prokaryotic relatives of eukaryotes. *PLOS Genetics* A. Rokas, Ed.; 2018, 14(3), e1007080 10.1371/journal.pgen.1007080.
142. Da Cunha, V.; Gaia, M.; Gabelle, D.; Nasir, A.; Forterre, P. Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLOS Genetics* 2017, 13(6), e1006810 10.1371/journal.pgen.1006810.
143. Da Cunha, V.; Gaia, M.; Nasir, A.; Forterre, P. Asgard archaea do not close the debate about the universal tree of life topology. *PLOS Genetics* A. Rokas, Ed.; 2018, 14(3), e1007215 10.1371/journal.pgen.1007215.
144. Zhou, L.; Zhou, Z.; Lu, Y.-W.; Ma, L.; Bai, Y.; Li, X.-X.; et al. The newly proposed TACK and DPANN archaea detected in the production waters from a high-temperature petroleum reservoir. *International Biodeterioration & Biodegradation* 2019, 143, 104729 10.1016/j.ibiod.2019.104729.
145. Probst, A.J.; Ladd, B.; Jarett, J.K.; Geller-McGrath, D.E.; Sieber, C.M.K.; Emerson, J.B.; et al. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nature Microbiology* 2018, 3(3), 328–336 10.1038/s41564-017-0098-y.
146. Fournier, G.P.; Poole, A.M. A Briefly Argued Case That Asgard Archaea Are Part of the Eukaryote Tree. *Frontiers in Microbiology* 2018, 9 10.3389/fmicb.2018.01896.
147. Peeters, E.; Driessen, R.P.C.; Werner, F.; Dame, R.T. The interplay between nucleoid organization and transcription in archaeal genomes. *Nature Reviews Microbiology* 2015, 13(6), 333–341 10.1038/nrmicro3467.
148. Sanders, T.J.; Marshall, C.J.; Santangelo, T.J. The Role of Archaeal Chromatin in Transcription. *Journal of Molecular Biology* 2019, 431(20), 4103–4115 10.1016/j.jmb.2019.05.006.
149. Henneman, B.; van Emmerik, C.; van Ingen, H.; Dame, R.T. Structure and function of archaeal histones. *PLoS Genetics* 2018, 14(9) 10.1371/journal.pgen.1007582.
150. Mattioli, F.; Bhattacharyya, S.; Dyer, P.N.; White, A.E.; Sandman, K.; Burkhart, B.W.; et al. Structure of Histone-based Chromatin in Archaea. *Science (New York, N.Y.)* 2017, 357(6351), 609–612 10.1126/science.aaj1849.
151. Bailey, K.A.; Chow, C.S.; Reeve, J.N. Histone stoichiometry and DNA circularization in archaeal nucleosomes. *Nucleic Acids Research* 1999, 27(2), 532–536 10.1093/nar/27.2.532.
152. Decanniere, K.; Sandman, K.; Reeve, J.N.; Heinemann, U. Crystallization and preliminary X-ray characterization of the *Methanothermobacter thermautotrophicus* histones HMfA and HMfB. *Proteins: Structure, Function, and Bioinformatics* 1996, 24(2), 269–271 10.1002/(SICI)1097-0134(199602)24:2<269::AID-PROT16>3.0.CO;2-L.
153. Musgrave, D.R.; Sandman, K.M.; Reeve, J.N. DNA binding by the archaeal histone HMf results in positive supercoiling. *Proceedings of the National Academy of Sciences of the United States of America* 1991, 88(23), 10397–10401.
154. DeLange, R.J.; Williams, L.C.; Searcy, D.G. A histone-like protein (HTa) from *Thermoplasma acidophilum*. II. Complete amino acid sequence. *Journal of Biological Chemistry* 1981, 256(2), 905–911.

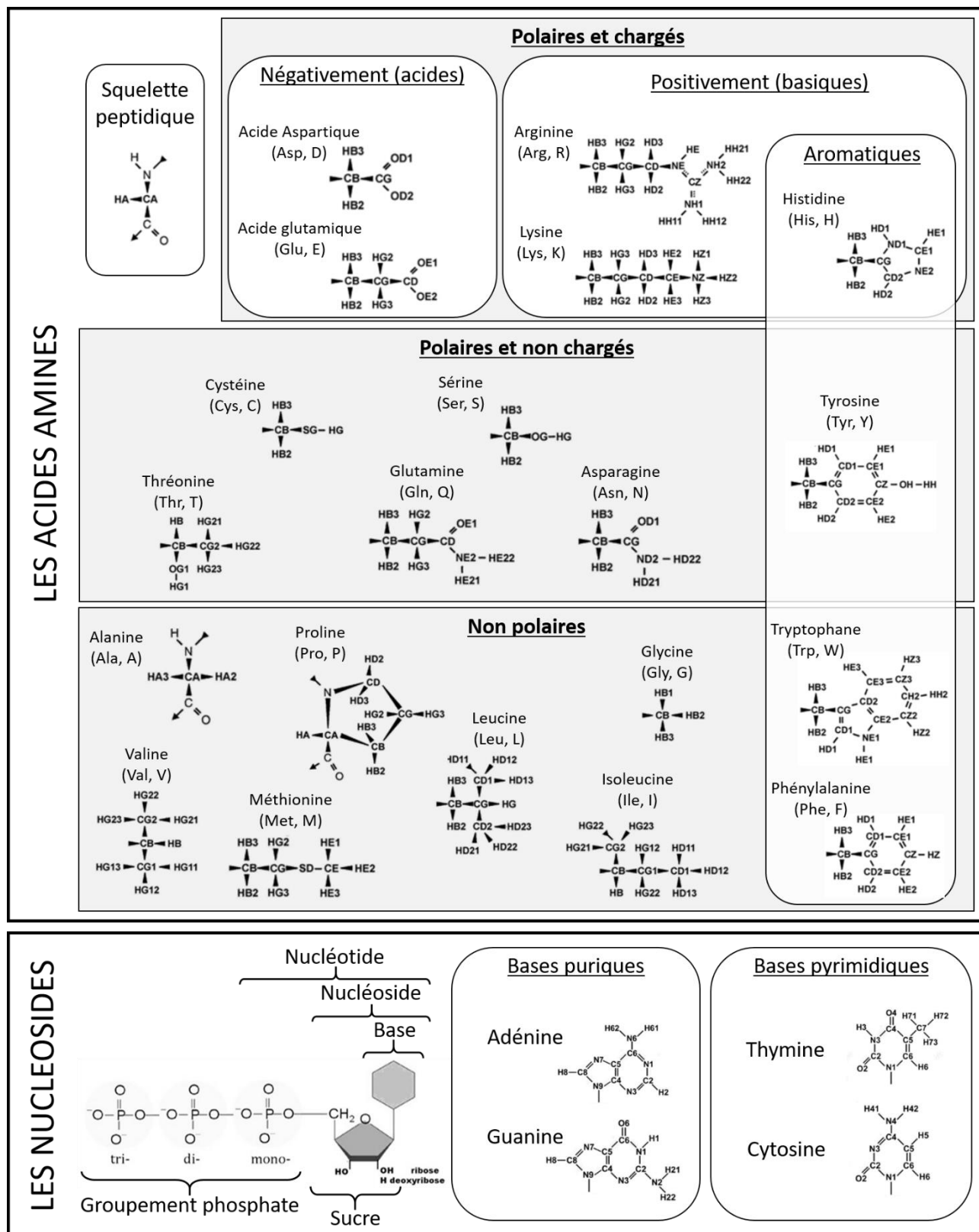
155. Hocher, A.; Rojec, M.; Swadling, J.B.; Esin, A.; Warnecke, T. The DNA-binding protein HTa from *Thermoplasma acidophilum* is an archaeal histone analog. *bioRxiv* 2019, 564930 10.1101/564930.
156. Stein, D.B.; Searcy, D.G. Physiologically important stabilization of DNA by a prokaryotic histone-like protein. *Science* 1978, 202(4364), 219–221 10.1126/science.694528.
157. Kawashima, T.; Amano, N.; Koike, H.; Makino, S.; Higuchi, S.; Kawashima-Ohya, Y.; et al. Archaeal adaptation to higher temperatures revealed by genomic sequence of *Thermoplasma volcanium*. *Proceedings of the National Academy of Sciences of the United States of America* 2000, 97(26), 14257–14262.
158. Fütterer, O.; Angelov, A.; Liesegang, H.; Gottschalk, G.; Schleper, C.; Schepers, B.; et al. Genome sequence of *Picrophilus torridus* and its implications for life around pH 0. *Proceedings of the National Academy of Sciences of the United States of America* 2004, 101(24), 9091–9096 10.1073/pnas.0401356101.
159. Wardleworth, B.N.; Russell, R.J.M.; Bell, S.D.; Taylor, G.L.; White, M.F. Structure of Alba: an archaeal chromatin protein modulated by acetylation. *The EMBO Journal* 2002, 21(17), 4654–4662 10.1093/emboj/cdf465.
160. Jelinska, C.; Conroy, M.J.; Craven, C.J.; Hounslow, A.M.; Bullough, P.A.; Waltho, J.P.; et al. Obligate Heterodimerization of the Archaeal Alba2 Protein with Alba1 Provides a Mechanism for Control of DNA Packaging. *Structure* 2005, 13(7), 963–971 10.1016/j.str.2005.04.016.
161. Goyal, M.; Banerjee, C.; Nag, S.; Bandyopadhyay, U. The Alba protein family: Structure and function. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 2016, 1864(5), 570–583 10.1016/j.bbapap.2016.02.015.
162. Laurens, N.; Driessen, R.P.C.; Heller, I.; Vorselen, D.; Noom, M.C.; Hol, F.J.H.; et al. Alba shapes the archaeal genome using a delicate balance of bridging and stiffening the DNA. *Nature Communications* 2012, 3, 1328 10.1038/ncomms2330.
163. Napoli, A.; Zivanovic, Y.; Bocs, C.; Buhler, C.; Rossi, M.; Forterre, P.; et al. DNA bending, compaction and negative supercoiling by the architectural protein Sso7d of *Sulfolobus solfataricus*. *Nucleic Acids Research* 2002, 30(12), 2656–2662.
164. Robinson, H.; Gao, Y.-G.; McCrary, B.S.; Edmondson, S.P.; Shriver, J.W.; Wang, A.H.-J. The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature* 1998, 392(6672), 202–205 10.1038/32455.
165. Driessen, R.P.C.; Dame, R.T. Structure and dynamics of the crenarchaeal nucleoid. *Biochemical Society Transactions* 2013, 41(1), 321–325 10.1042/BST20120336.
166. Guo, L.; Feng, Y.; Zhang, Z.; Yao, H.; Luo, Y.; Wang, J.; et al. Biochemical and structural characterization of Cren7, a novel chromatin protein conserved among Crenarchaea. *Nucleic Acids Research* 2008, 36(4), 1129–1137 10.1093/nar/gkm1128.
167. Kaur, G.; Iyer, L.M.; Subramanian, S.; Aravind, L. Evolutionary convergence and divergence in archaeal chromosomal proteins and Chromo-like domains from bacteria and eukaryotes. *Scientific Reports* 2018, 8(1), 1–10 10.1038/s41598-018-24467-z.
168. Luo, X.; Schwarz-Linek, U.; Botting, C.H.; Hensel, R.; Siebers, B.; White, M.F. CC1, a Novel Crenarchaeal DNA Binding Protein. *Journal of Bacteriology* 2007, 189(2), 403–409 10.1128/JB.01246-06.
169. Chartier, F.; Laine, B.; Sautiere, P.; Touze, J.-P.; Albagnac, G. Characterization of the chromosomal protein HMB isolated from *Methanosarcina barkeri*. *FEBS Letters* 1985, 183(1), 119–123 10.1016/0014-5793(85)80967-4.
170. Laine, B.; Chartier, F.; Imbert, M.; Lewis, R.; Sautiere, P. Primary structure of the chromosomal protein HMB from the archaeobacteria *Methanosarcina barkeri*. *European Journal of Biochemistry* 1986, 161(3), 681–687 10.1111/j.1432-1033.1986.tb10493.x.
171. Laine, B.; Culard, F.; Maurizot, J.C.; Sautière, P. The chromosomal protein MC1 from the archaeobacterium *Methanosarcina* sp. CHTI 55 induces DNA bending and supercoiling. *Nucleic Acids Research* 1991, 19(11), 3041–3045.
172. Cam, E.L.; Culard, F.; Larquet, E.; Delain, E.; Cognet, J.A.H. DNA Bending Induced by the Archaeobacterial Histone-like Protein MC1. *Journal of Molecular Biology* 1999, 285(3), 1011–1021 10.1006/jmbi.1998.2321.
173. Chartier, F.; Laine, B.; Belaïche, D.; Touzel, J.-P.; Sautière, P. Primary structure of the chromosomal protein MC1 from the archaeobacterium *Methanosarcina* sp. CHTI 55. *Biochimica*

- et Biophysica Acta (BBA) - Gene Structure and Expression 1989, 1008(3), 309–314 10.1016/0167-4781(89)90021-3.
174. Williams, T.J.; Burg, D.W.; Raftery, M.J.; Poljak, A.; Guilhaus, M.; Pilak, O.; et al. Global Proteomic Analysis of the Insoluble, Soluble, and Supernatant Fractions of the Psychrophilic Archaeon *Methanococcoides burtonii* Part I: The Effect of Growth Temperature. *Journal of Proteome Research* 2010, 9(2), 640–652 10.1021/pr900509n.
175. Goodchild, A.; Raftery, M.; Saunders, N.F.W.; Guilhaus, M.; Cavicchioli, R. Biology of the Cold Adapted Archaeon, *Methanococcoides burtonii* Determined by Proteomics Using Liquid Chromatography-Tandem Mass Spectrometry. *Journal of Proteome Research* 2004, 3(6), 1164–1176 10.1021/pr0498988.
176. Paquet, F.; Culard, F.; Barbault, F.; Maurizot, J.-C.; Lancelot, G. NMR Solution Structure of the Archaeobacterial Chromosomal Protein MC1 Reveals a New Protein Fold. *Biochemistry* 2004, 43(47), 14971–14978 10.1021/bi048382z.
177. Buré, C.; Goffinont, S.; Delmas, A.F.; Cadene, M.; Culard, F. Oxidation-sensitive Residues Mediate the DNA Bending Abilities of the Architectural MC1 Protein. *Journal of Molecular Biology* 2008, 376(1), 120–130 10.1016/j.jmb.2007.11.057.
178. Gao, Y.-G.; Su, S.-Y.; Robinson, H.; Padmanabhan, S.; Lim, L.; McCrary, B.S.; et al. The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA. *Nature Structural Biology* 1998, 5(9), 782–786 10.1038/1822.
179. Agback, P.; Baumann, H.; Knapp, S.; Ladenstein, R.; Härd, T. Architecture of nonspecific protein–DNA interactions in the Sso7d–DNA complex. *Nature Structural Biology* 1998, 5(7), 579–584 10.1038/836.
180. De Vuyst, G. LA PROTEINE MC1 D'ARCHAEABACTERIE : RECONNAISSANCE DE SEQUENCES PARTICULIERES 2004.
181. De Vuyst, G.; Aci, S.; Genest, D.; Culard, F. Atypical Recognition of Particular DNA Sequences by the Archaeal Chromosomal MC1 Protein †. *Biochemistry* 2005, 44(30), 10369–10377 10.1021/bi0474416.
182. Ngo, H.B.; Kaiser, J.T.; Chan, D.C. The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA. *Nature Structural & Molecular Biology* 2011, 18(11), 1290–1296 10.1038/nsmb.2159.
183. Loth, K.; Gnida, M.; Romanuka, J.; Kaptein, R.; Boelens, R. Sliding and target location of DNA-binding proteins: an NMR view of the lac repressor system. *Journal of Biomolecular NMR* 2013, 56(1), 41–49 10.1007/s10858-013-9723-0.
184. sauter, claude; giegé, richard La cristallogénèse des macromolécules biologiques. *Regard sur la biochimie* 2001.
185. Gabadinho, J.; Beteva, A.; Guijarro, M.; Rey-Bakaikoa, V.; Spruce, D.; Bowler, M.W.; et al. MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments. *Journal of Synchrotron Radiation* 2010, 17(Pt 5), 700–707 10.1107/S0909049510020005.
186. Kabsch, W. XDS. *Acta Crystallographica Section D: Biological Crystallography* 2010, 66(Pt 2), 125–132 10.1107/S0907444909047337.
187. Winn, M.D.; Ballard, C.C.; Cowtan, K.D.; Dodson, E.J.; Emsley, P.; Evans, P.R.; et al. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* 2011, 67(Pt 4), 235–242 10.1107/S0907444910045749.
188. Kikhney, A.G.; Svergun, D.I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Letters* 2015, 589(19, Part A), 2570–2577 10.1016/j.febslet.2015.08.027.
189. Nallet, F. Introduction à la diffusion aux petits angles. *École thématique de la Société Française de la Neutronique* 2010, 11, 17–42 10.1051/sfn/201011003.
190. Franke, D.; Petoukhov, M.V.; Konarev, P.V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H.D.T.; et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography* 2017, 50(4), 1212–1225 10.1107/S1600576717007786.
191. Svergun, D.I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* 1992, 25(4), 495–503 10.1107/S0021889892001663.

192. Svergun, D.I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical Journal* 1999, 76(6), 2879–2886.
193. Svergun, D.; Barberato, C.; Koch, M.H.J. CRY SOL– a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography* 1995, 28(6), 768–773 10.1107/S0021889895007047.
194. Svergun, D.I.; Koch, M.H.J. Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics* 2003, 66(10), 1735–1782 10.1088/0034-4885/66/10/R05.
195. David, G.; Pérez, J. Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. *Journal of Applied Crystallography* 2009, 42(5), 892–900 10.1107/S0021889809029288.

Annexes

Annexes



Annexe 1 : Structure primaire et nomenclature IUPAC des acides aminés et des nucléosides.

	HU α_2	HU β_2
Contraintes RMN		
Contraintes de distance par chaîne		
NOE totaux	1638	1846
Liaisons hydrogène	0	101
Contraintes d'angle dièdres	168	163
Statistiques structurales		
	Modèles 1-10	Modèles 1-10
Violations moyennes par structure		
NOEs ≥ 0.3 Å	1,2%	10%
Liaisons hydrogène ≥ 0.5 Å	0	0
Angles dièdres $\geq 10^\circ$	3,6%	9,2%
RMSD moyen (Å) (1-55 75-90)	0,3	0,24
Diagramme de Ramachandran		
Régions les plus favorables	87,6%	77,9%
Régions favorables	11,9%	21,9%
Régions autorisées	0,5%	0,2%
Régions interdites	0%	0%

Annexe 2 : Contraintes RMN et statistiques structurales sur un jeu de 10 modèles pour chaque homodimère de EchU, en cours d'affinement (issu de la thèse de Rémy Le Meur, 2015) ^[69].

Paramètres d'acquisition

Exp	¹⁵ N-HMQC SOFAST
d1	0,1s
ns	32
durée exp	6'30
T°	27°C

Paramètres d'échantillon

nom	Dimère titré		Dimère tirant
	EchUβ ₂ ¹⁵ N (ββ)	EchUα ₂ nm (αα)	
C° (mM)	0,07	0,25	
V (μL)	198		
n (nmol)	14,40		

Titrage par ajouts cumulés

Avancement temps (n° ajout)	ββ (nmol)	+	αα (nmol)	→	2 αβ (nmol)	V _{tot} (μL)	V _{ajout} (μL)	ratio
t(0)	14,40		0,00		0,00	198	0	0
t(1)	14,40		0,72		0,00			
t(1)	13,68		0,00		1,45	200,9	2,9	1 : 0,050
t(2)	13,68		0,72		1,45			
t(2)	12,95		0,00		2,90	203,8	2,9	1 : 0,101
t(3)	12,95		0,72		2,90			
t(3)	12,23		0,00		4,35	206,7	2,9	1 : 0,151
t(4)	12,23		0,72		4,35			
t(4)	11,51		0,00		5,80	209,6	2,9	1 : 0,201
t(5)	11,51		0,72		5,80			
t(5)	10,78		0,00		7,25	212,5	2,9	1 : 0,252
t(6)	10,78		0,72		7,25			
t(6)	10,06		0,00		8,70	215,4	2,9	1 : 0,302
t(7)	10,06		0,72		8,70			
t(7)	9,33		0,00		10,15	218,3	2,9	1 : 0,352
t(8)	9,33		0,72		10,15			
t(8)	8,61		0,00		11,59	221,2	2,9	1 : 0,403
t(9)	8,61		0,72		11,59			
t(9)	7,88		0,00		13,04	224,1	2,9	1 : 0,453
t(10)	7,88		0,72		13,04			
t(10)	7,16		0,00		14,49	227	2,9	1 : 0,503
t(11)	7,16		0,72		14,49			
t(11)	6,43		0,00		15,94	229,9	2,9	1 : 0,553
t(12)	6,43		0,72		15,94			
t(12)	5,71		0,00		17,39	232,8	2,9	1 : 0,604

Avancement temps (n° ajout)	ββ (nmol)	+	αα (nmol)	→	2 αβ (nmol)	V _{tot} (μL)	V _{ajout} (μL)	ratio
t(13)	5,71		0,72		17,39	235,7	2,9	1 : 0,654
t(13)	4,98		0,00		18,84			
t(14)	4,98		0,72		18,84			
t(14)	4,26		0,00		20,29	238,6	2,9	1 : 0,704
t(15)	4,26		0,72		20,29			
t(15)	3,53		0,00		21,74	241,5	2,9	1 : 0,755
t(16)	3,53		0,72		21,74			
t(16)	2,81		0,00		23,19	244,4	2,9	1 : 0,805
t(17)	2,81		0,72		23,19			
t(17)	2,08		0,00		24,64	247,3	2,9	1 : 0,855
t(18)	2,08		0,72		24,64			
t(18)	1,36		0,00		26,09	250,2	2,9	1 : 0,906
t(19)	1,36		0,72		26,09			
t(19)	0,64		0,00		27,54	253,1	2,9	1 : 0,956
t(20)	0,64		0,72		27,54			
t(20)	0,00		0,09		28,81	256	2,9	1 : 1,006
t(21)	0,00		0,81		28,81			
t(21)	0,00		0,81		28,81	258,9	2,9	1 : 1,057
t(22)	0,00		1,54		28,81			
t(22)	0,00		1,54		28,81	261,8	2,9	1 : 1,107
t(23)	0,00		2,99		28,81			
t(23)	0,00		2,99		28,81	267,6	5,8	1 : 1,207
t(24)	0,00		4,44		28,81			
t(24)	0,00		4,44		28,81	273,4	5,8	1 : 1,308
t(25)	0,00		5,89		28,81			
t(25)	0,00		5,89		28,81	279,2	5,8	1 : 1,409
t(26)	0,00		7,34		28,81			
t(26)	0,00		7,34		28,81	285	5,8	1 : 1,509

Paramètres d'acquisition

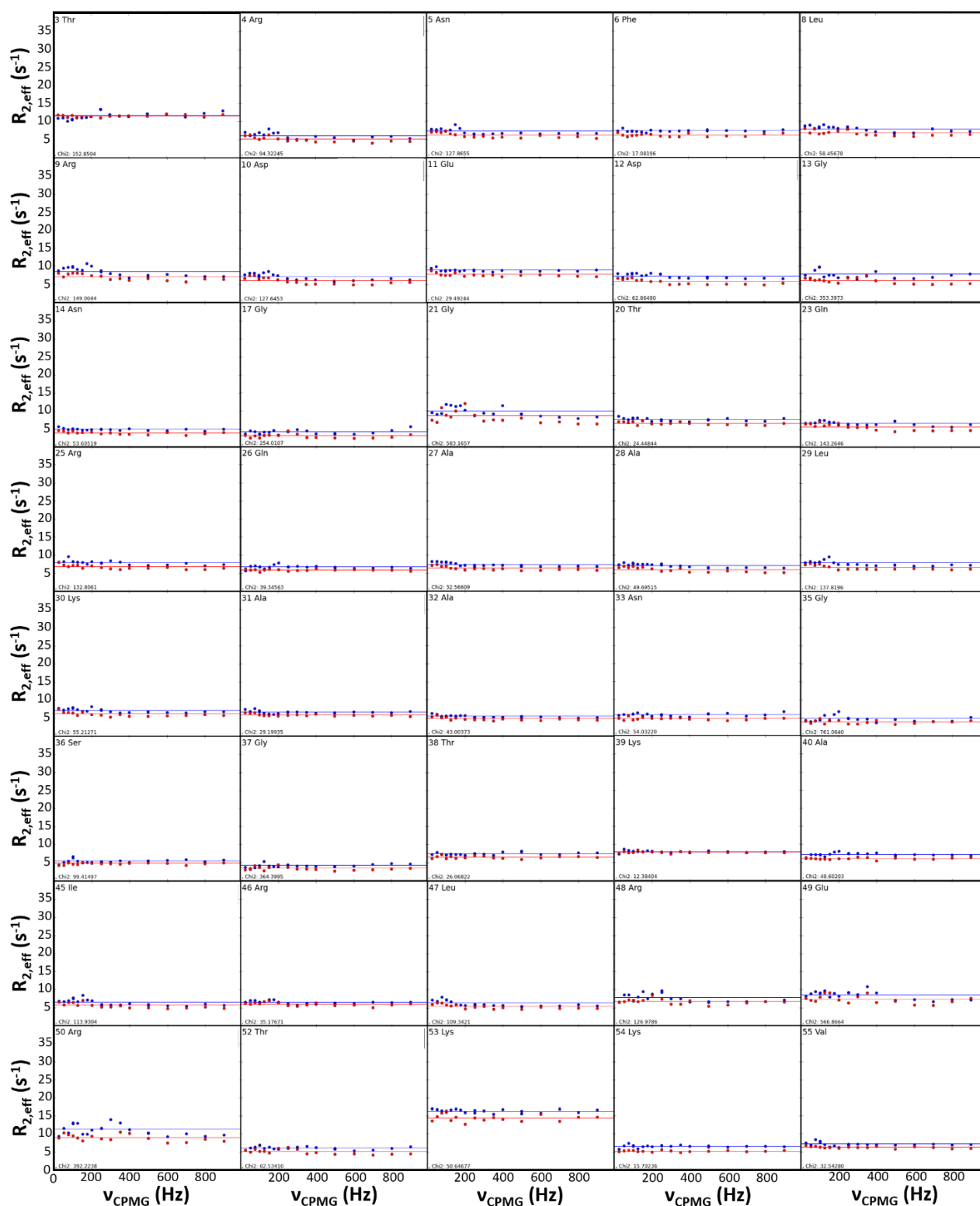
Exp	¹⁵ N-HMOC SOFAST
d1	0.1s
ns	32
durée exp	6'30
T°	37°C

Paramètres d'échantillon

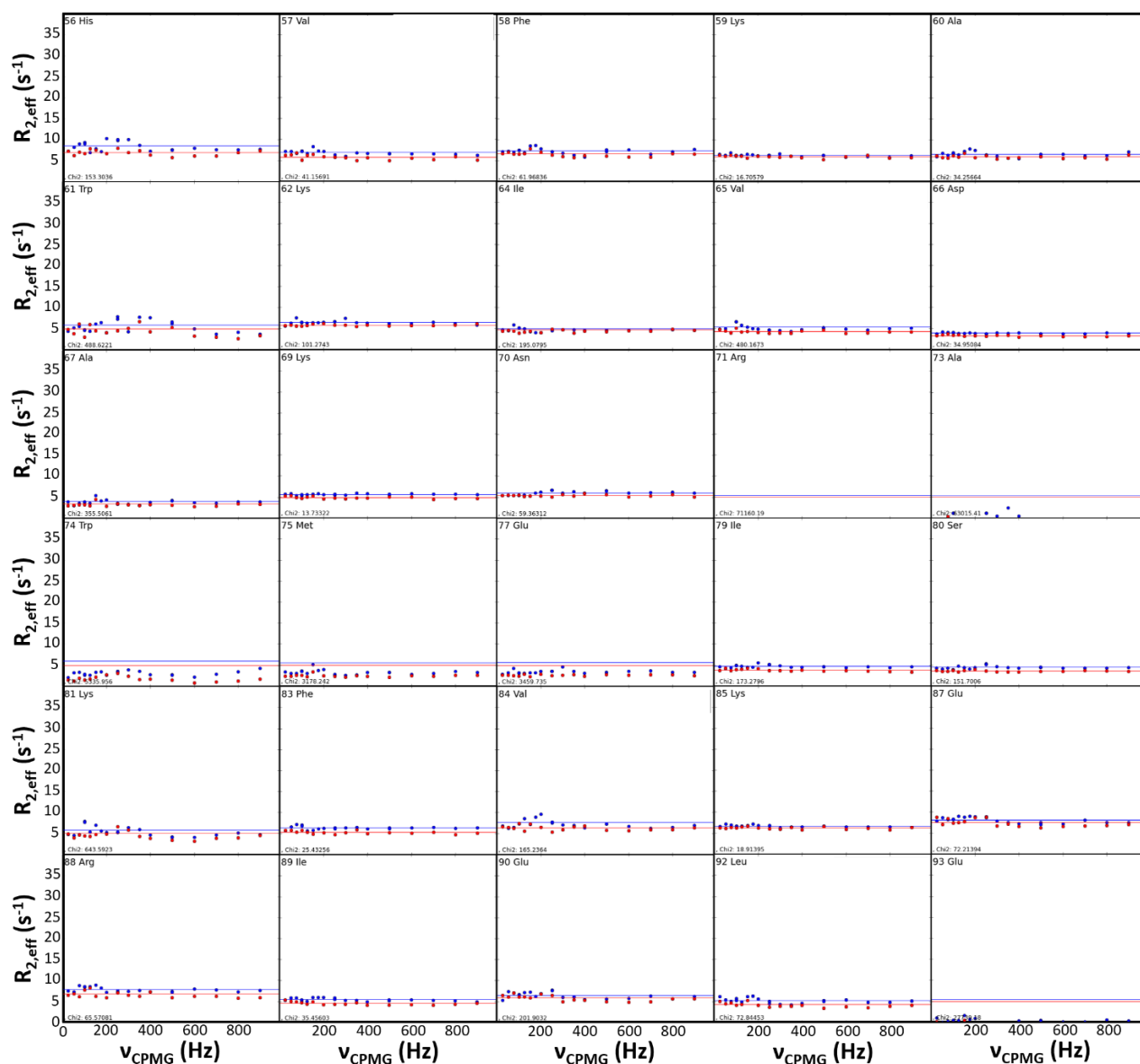
nom	Dimère titré	Dimère titrant
C° (mM)	<i>EchU</i> β ₂ - ¹⁵ N (ββ)	<i>EchU</i> α ₂ nm (αα)
V (μL)	0.07	0.25
n (nmol)	200	
	14.55	

Titration par ajouts cumulés

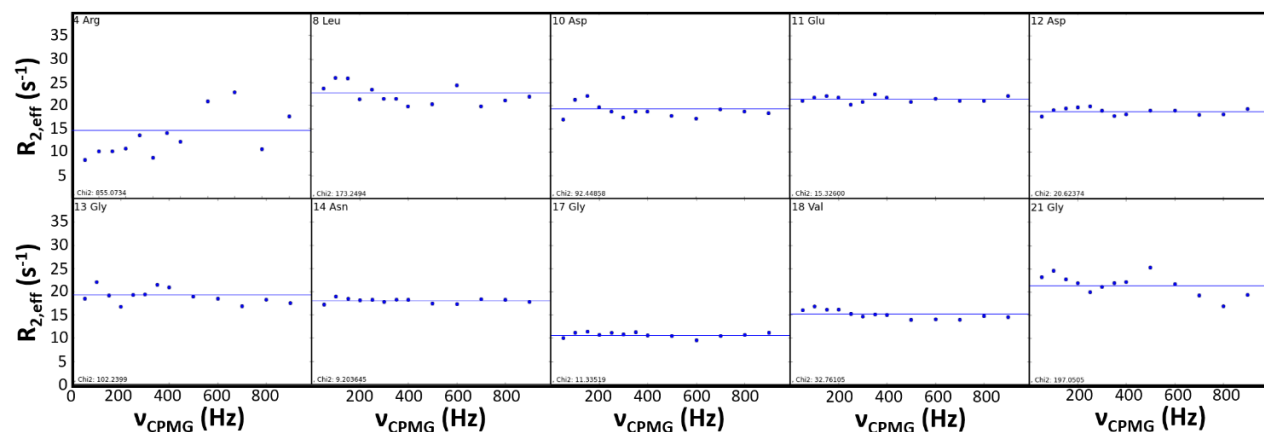
Avancement temps (n° ajout)	ββ (nmol)	+	αα (nmol)	→	2 αβ (nmol)	Avancement temps (n° ajout)	ββ (nmol)	+	αα (nmol)	→	2 αβ (nmol)	V _{tot} (μL)	V _{ajout} (μL)	ratio
t _i (0)	14.55		0,00		0,00	t _i (13)	7,05		0,50		14,99	232	2	1 : 0.550
t _i (1)	14.55		0,50		0,00	t _i (14)	6,55		0,00		15,99			
t _i (1)	14.05		0,00		1,00	t _i (15)	6,05		0,50		16,99	234	2	1 : 0.584
t _i (2)	14.05		0,50		1,00	t _i (16)	5,55		0,00		17,99			
t _i (2)	13.55		0,00		2,00	t _i (17)	5,05		0,50		18,99	236	2	1 : 0.618
t _i (3)	13.55		0,50		2,00	t _i (18)	4,55		0,00		19,99			
t _i (3)	13.05		0,00		3,00	t _i (19)	4,05		0,50		20,99	238	2	1 : 0.653
t _i (4)	13.05		0,50		3,00	t _i (20)	3,55		0,00		21,99			
t _i (4)	12.55		0,00		4,00	t _i (21)	3,05		0,50		22,99	240	2	1 : 0.687
t _i (5)	12.55		0,50		4,00	t _i (22)	2,56		0,00		23,99			
t _i (5)	12.05		0,00		5,00	t _i (23)	2,06		0,50		24,99	244	4	1 : 0.756
t _i (6)	12.05		0,50		5,00	t _i (24)	1,56		0,00		25,99			
t _i (6)	11.55		0,00		6,00	t _i (25)	1,06		0,50		26,99	248	4	1 : 0.824
t _i (7)	11.55		0,50		6,00	t _i (26)	0,56		0,00		27,99			
t _i (7)	11.05		0,00		7,00	t _i (27)	0,06		0,50		28,99	252	4	1 : 0.893
t _i (8)	11.05		1,00		7,00	t _i (28)	0,00		0,44		29,10			
t _i (8)	10.05		0,00		9,00	t _i (29)	0,00		0,44		29,10	256	4	1 : 0.962
t _i (9)	10.05		0,50		9,00	t _i (30)	0,00		1,44		29,10			
t _i (9)	9.55		0,00		10,00	t _i (31)	0,00		1,44		29,10	260	4	1 : 1.030
t _i (10)	9.55		1,00		10,00	t _i (32)	0,00		2,44		29,10			
t _i (10)	8.55		0,00		11,99	t _i (33)	0,00		2,44		29,10	264	4	1 : 1.099
t _i (11)	8.55		1,00		11,99	t _i (34)	0,00		3,44		29,10			
t _i (11)	7.55		0,00		13,99	t _i (35)	0,00		3,44		29,10	268	4	1 : 1.198
t _i (12)	7.55		0,50		13,99	t _i (36)	0,00		3,44		29,10			
t _i (12)	7.05		0,00		14,99							272	4	1 : 1.397



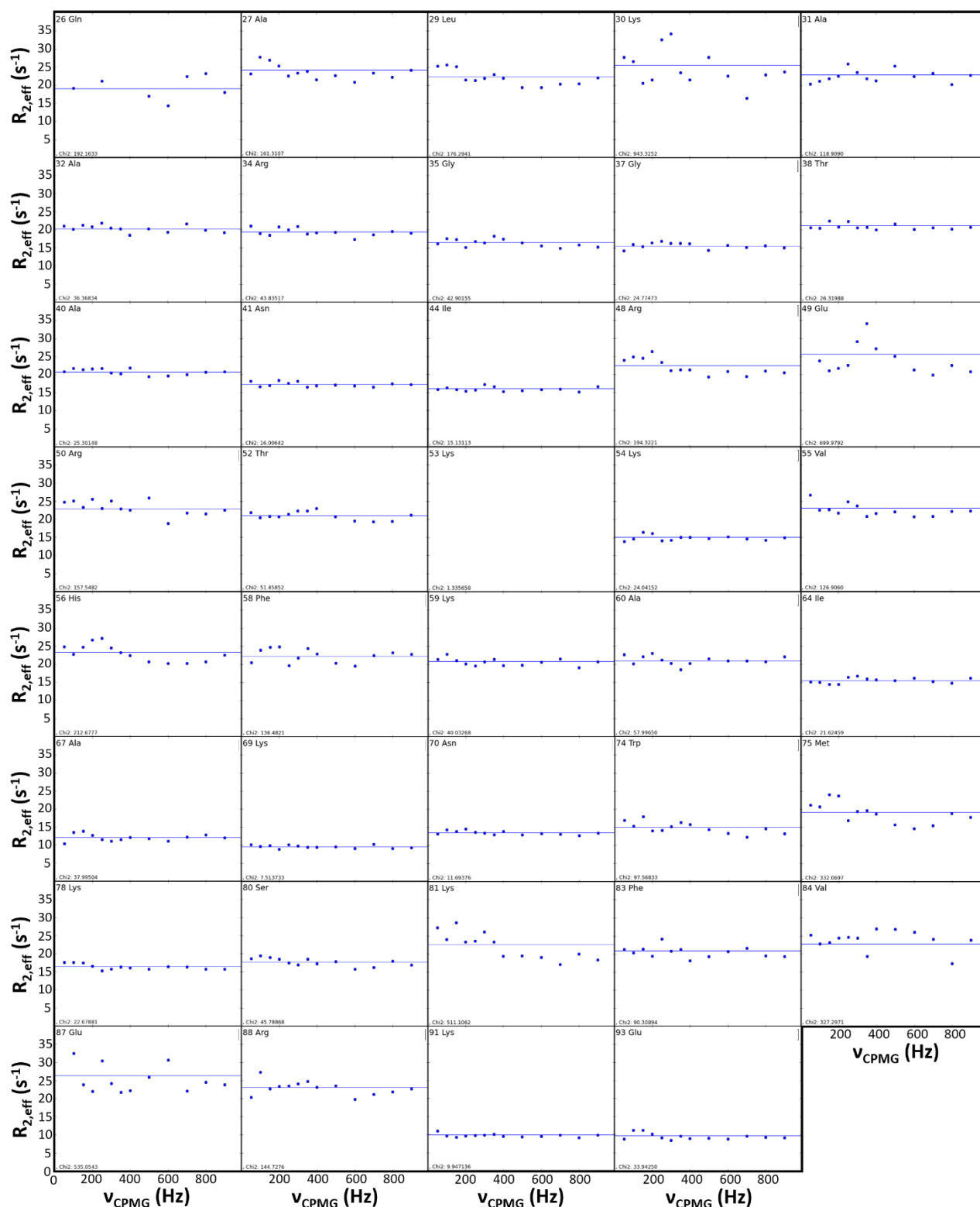
Annexe 6 : *Courbes de l'évolution du signal CPMG des résidus de la protéine MC1 en solution (1/2). Les données en bleu et en rouge ont été enregistrées respectivement sur un spectromètre de 700 MHz et de 600 MHz.*



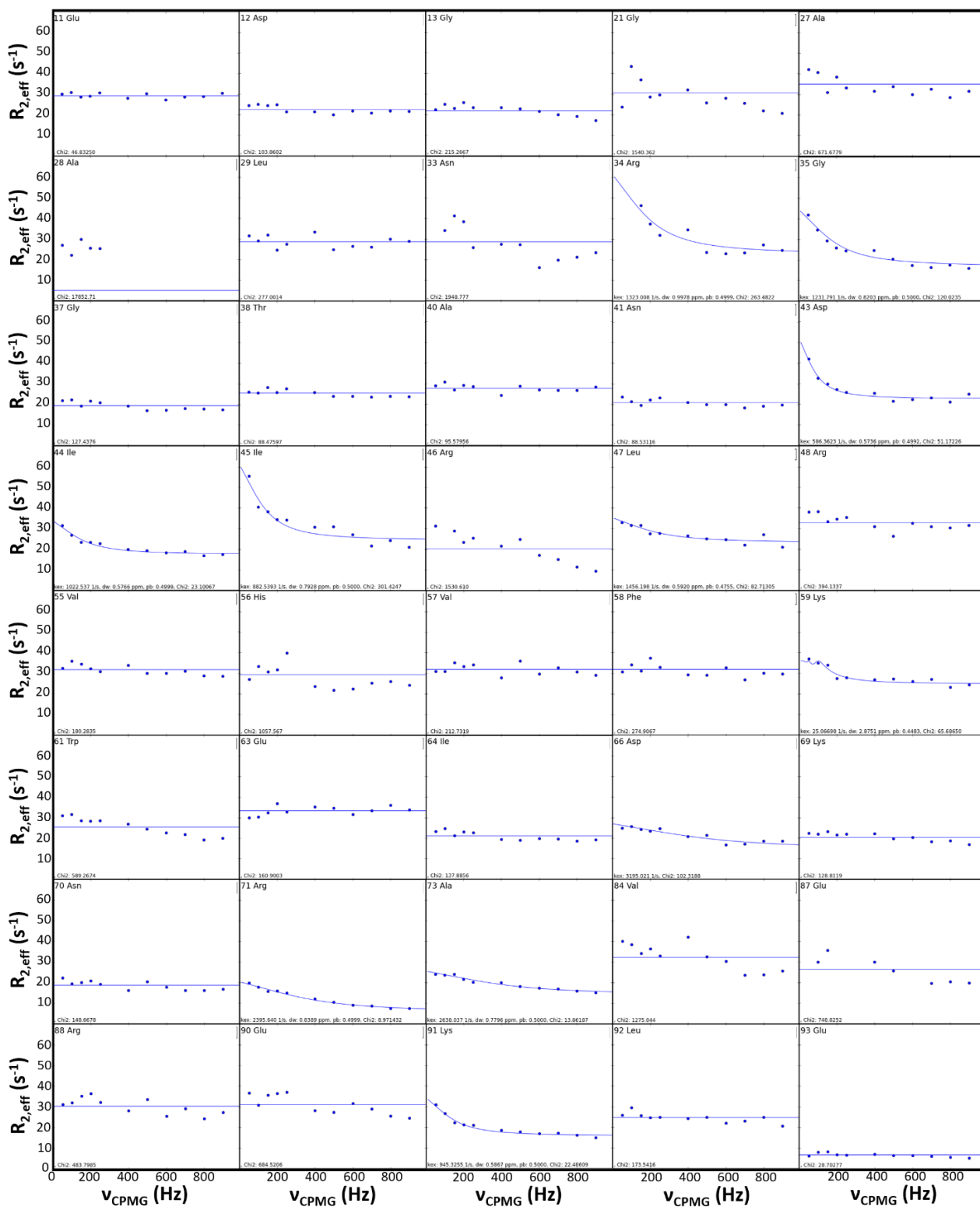
Annexe 7 : Courbes de l'évolution du signal CPMG des résidus de la protéine MC1 en solution (2/2). Les données en bleu et en rouge ont été enregistrées respectivement sur un spectromètre de 700 MHz et de 600 MHz.



Annexe 8 : Courbes de l'évolution du signal CPMG des résidus de la protéine MC1 dans le complexe à 15bp (1/2).



Annexe 9 : *Courbes de l'évolution du signal CPMG des résidus de la protéine MC1 dans le complexe à 15bp (2/2).*



Annexe 10 : *Courbes de l'évolution du signal CPMG des résidus de la protéine MC1 dans le complexe à 23bp.*

Well	Kit #	Salt	Buffer	pH	Precipitant
A1	1.1	0.2M Lithium sulfate	0.1M Sodium acetate	4.5	50% w/v PEG 400
B1	1.2	none	0.1M Sodium citrate	5.5	20% w/v PEG 3000
C1	1.3	0.2M Ammonium citrate dibasic	none		20% w/v PEG 3350
D1	1.4	0.02M Calcium chloride dihydrate	0.1M Sodium acetate	4.6	30% v/v MPD
E1	1.5	0.2M Magnesium formate dihydrate	none		20% w/v PEG 3350
F1	1.6	0.2M Lithium sulfate	0.1M Phosphate/citrate	4.2	20% w/v PEG 1000
G1	1.7	none	0.1M CHES	9.5	20% w/v PEG 8000
H1	1.8	0.2M Ammonium formate	none		20% w/v PEG 3350
A2	1.9	0.2M Ammonium chloride	none		20% w/v PEG 3350
B2	1.10	0.2M Potassium formate	none		20% w/v PEG 3350
C2	1.11	0.2M Ammonium phosphate monobasic	0.1M Tris	8.5	50% v/v MPD
D2	1.12	0.2M Potassium nitrate	none		20% w/v PEG 3350
E2	1.13	0.8M Ammonium sulfate	0.1M Citrate	4.0	none
F2	1.14	0.2M Sodium thiocyanate	none		20% w/v PEG 3350
G2	1.15	none	0.1M BICINE	9.0	20% w/v PEG 6000
H2	1.16	none	0.1M HEPES	7.5	10% w/v PEG 8000 / 8% v/v Ethylene glycol
A3	1.17	none	0.1M Sodium cacodylate	6.5	40% v/v MPD / 5% w/v PEG 8000
B3	1.18	none	0.1M Phosphate/citrate	4.2	40% v/v Ethanol / 5% w/v PEG 1000
C3	1.19	none	0.1M Sodium acetate	4.6	8% w/v PEG 4000
D3	1.20	0.2M Magnesium chloride hexahydrate	0.1M Tris	7.0	10% w/v PEG 8000
E3	1.21	none	0.1M Citrate	5.0	20% w/v PEG 6000
F3	1.22	0.2M Magnesium chloride hexahydrate	0.1M Sodium cacodylate	6.5	50% v/v PEG 200
G3	1.23	1.6M Sodium citrate tribasic dihydrate	none	6.5	none
H3	1.24	0.2M Potassium citrate tribasic monohydrate	none		20% w/v PEG 3350
A4	1.25	0.2M Sodium chloride	0.1M Phosphate/citrate	4.2	20% w/v PEG 8000
B4	1.26	1M Lithium chloride	0.1M Citrate	4.0	20% w/v PEG 6000
C4	1.27	0.2M Ammonium nitrate	none		20% w/v PEG 3350
D4	1.28	none	0.1M HEPES	7.0	10% w/v PEG 6000
E4	1.29	0.8M Sodium phosphate monobasic monohydrate 0.8M Potassium phosphate monobasic	0.1M Sodium HEPES	7.5	none
F4	1.30	none	0.1M Phosphate/citrate	4.2	40% v/v PEG 300
G4	1.31	0.2M Zinc acetate dihydrate	0.1M Sodium acetate	4.5	10% w/v PEG 3000
H4	1.32	none	0.1M Tris	8.5	20% v/v Ethanol
A5	1.33	none	0.1M Sodium/potassium phosphate	6.2	25% v/v 1,2-Propanediol / 10%v/v Glycerol
B5	1.34	none	0.1M BICINE	9.0	10% w/v PEG 20.000 / 2% v/v 1,4-Dioxane
C5	1.35	2M Ammonium sulfate	0.1M Sodium acetate	4.6	none
D5	1.36	none	none		10% w/v PEG 1000 / 10% w/v PEG 8000
E5	1.37	none	none		24% w/v PEG 1500 / 20% v/v Glycerol
F5	1.38	0.2M Magnesium chloride hexahydrate	0.1M Sodium HEPES	7.5	30% v/v PEG 400
G5	1.39	0.2M Sodium chloride	0.1M Sodium/potassium phosphate	6.2	50% v/v PEG 200
H5	1.40	0.2M Lithium sulfate	0.1M Sodium acetate	4.5	30% w/v PEG 8000
A6	1.41	none	0.1M HEPES	7.5	70% v/v MPD
B6	1.42	0.2M Magnesium chloride hexahydrate	0.1M Tris	8.5	20% w/v PEG 8000
C6	1.43	0.2M Lithium sulfate	0.1M Tris	8.5	40% v/v PEG 400
D6	1.44	none	0.1M Tris	8.0	40% v/v MPD
E6	1.45	0.17M Ammonium sulfate	none		25.5% w/v PEG 4000 / 15% v/v Glycerol
F6	1.46	0.2M Calcium acetate hydrate	0.1M Sodium cacodylate	6.5	40% v/v PEG 300
G6	1.47	0.14M Calcium chloride dihydrate	0.07M Sodium acetate	4.6	14% v/v 2-Propanol / 30% v/v Glycerol
H6	1.48	0.04M Potassium phosphate monobasic	none		16% w/v PEG 8000 / 20% v/v Glycerol
A7	2.1	1M Sodium citrate tribasic dihydrate	0.1M Sodium cacodylate	6.5	none
B7	2.2	2M Ammonium sulfate 0.2M Sodium chloride	0.1M Sodium cacodylate	6.5	none
C7	2.3	0.2M Sodium chloride	0.1M HEPES	7.5	10% v/v 2-Propanol
D7	2.4	1.26M Ammonium sulfate 0.2M Lithium sulfate	0.1M Tris	8.5	none
E7	2.5	none	0.1M CAPS	10.5	40% v/v MPD
F7	2.6	0.2M Zinc acetate dihydrate	0.1M Imidazole	8.0	20% w/v PEG 3000
G7	2.7	0.2M Zinc acetate dihydrate	0.1M Sodium cacodylate	6.5	10% v/v 2-Propanol
H7	2.8	1M Ammonium phosphate dibasic	0.1M Sodium acetate	4.5	none
A8	2.9	1.6M Magnesium sulfate heptahydrate	0.1M MES	6.5	none
B8	2.10	none	0.1M BICINE	9.0	10% w/v PEG 6000
C8	2.11	0.16M Calcium acetate hydrate	0.08M Sodium cacodylate	6.5	14.4% w/v PEG 8000 / 20% v/v Glycerol
D8	2.12	none	0.1M Imidazole	8.0	10% w/v PEG 8000
E8	2.13	0.05M Cesium chloride	0.1M MES	6.5	30% v/v Jeffamine® M-600
F8	2.14	3.2M Ammonium sulfate	0.1M Citrate	5.0	none
G8	2.15	none	0.1M Tris	8.0	20% v/v MPD
H8	2.16	none	0.1M HEPES	7.5	20% v/v Jeffamine® M-600
A9	2.17	0.2M Magnesium chloride hexahydrate	0.1M Tris	8.5	50% v/v Ethylene glycol
B9	2.18	none	0.1M BICINE	9.0	10% v/v MPD
C9	2.19	0.8M Succinic acid	none	7.0	none
D9	2.20	2.1M DL-Malic acid	none	7.0	none
E9	2.21	2.4M Sodium malonate dibasic monohydrate	none	7.0	none
F9	2.22	1.1M Sodium malonate dibasic monohydrate	0.1M HEPES	7.0	0.5% v/v Jeffamine® ED-2003
G9	2.23	1M Succinic acid	0.1M HEPES	7.0	1% w/v PEG 2000 MME
H9	2.24	none	0.1M HEPES	7.0	30% v/v Jeffamine® M-600
A10	2.25	none	0.1M HEPES	7.0	30% v/v Jeffamine® ED-2003
B10	2.26	0.02M Magnesium chloride hexahydrate	0.1M HEPES	7.5	22% w/v Poly(acrylic acid sodium salt) 5100
C10	2.27	0.01M Cobalt(II) chloride hexahydrate	0.1M Tris	8.5	20% w/v Polyvinylpyrrolidone
D10	2.28	0.2M TMAO	0.1M Tris	8.5	20% w/v PEG 2000 MME
E10	2.29	0.005M Cobalt(II) chloride hexahydrate 0.005M Cadmium chloride hemi(pentahydrate) 0.005M Magnesium chloride hexahydrate 0.005M Nickel(II) chloride hexahydrate	0.1M HEPES	7.5	12% w/v PEG 3350
F10	2.30	0.2M Sodium malonate dibasic monohydrate	none		20% w/v PEG 3350
G10	2.31	0.1M Succinic acid	none		15% w/v PEG 3350
H10	2.32	0.15M DL-Malic acid	none		20% w/v PEG 3350
A11	2.33	0.1M Potassium thiocyanate	none		30% w/v PEG 2000 MME
B11	2.34	0.15M Potassium bromide	none		30% w/v PEG 2000 MME
C11	2.35	2M Ammonium sulfate	0.1M Bis-Tris	5.5	none
D11	2.36	3M Sodium chloride	0.1M Bis-Tris	5.5	none
E11	2.37	0.3M Magnesium formate dihydrate	0.1M Bis-Tris	5.5	none
F11	2.38	1M Ammonium sulfate	0.1M Bis-Tris	5.5	1% w/v PEG 3350
G11	2.39	none	0.1M Bis-Tris	5.5	25% w/v PEG 3350
H11	2.40	0.2M Calcium chloride dihydrate	0.1M Bis-Tris	5.5	45% v/v MPD
A12	2.41	0.2M Ammonium acetate	0.1M Bis-Tris	5.5	45% v/v MPD
B12	2.42	0.1M Ammonium acetate	0.1M Bis-Tris	5.5	17% w/v PEG 10.000
C12	2.43	0.2M Ammonium sulfate	0.1M Bis-Tris	5.5	25% w/v PEG 3350
D12	2.44	0.2M Sodium chloride	0.1M Bis-Tris	5.5	25% w/v PEG 3350
E12	2.45	0.2M Lithium sulfate	0.1M Bis-Tris	5.5	25% w/v PEG 3350
F12	2.46	0.2M Ammonium acetate	0.1M Bis-Tris	5.5	25% w/v PEG 3350
G12	2.47	0.2M Magnesium chloride hexahydrate	0.1M Bis-Tris	5.5	25% w/v PEG 3350
H12	2.48	0.2M Ammonium acetate	0.1M HEPES	7.5	45% v/v MPD

Annexe 11 : Composition des 96 solutions des kits 1 et 2 JCSG-plus™ (JCSG+) de Molecular Dimensions.

Well	Kit #	Salt	Buffer	pH	Precipitant
A1	1.1	0.01 M Cobalt chloride	0.1 M Sodium acetate	4.6	1.0 M 1,6-Hexanediol
B1	1.2	none	0.1 M tri-Sodium citrate	5.6	2.5 M 1,6-Hexanediol
C1	1.3	0.2 M Magnesium chloride	0.1 M TRIS	8.5	3.4 M 1,6-Hexanediol
D1	1.4	2.0 M Ammonium sulfate	none		5 % (v/v) Isopropanol
E1	1.5	none	0.1 M HEPES sodium salt	7.5	10 % (v/v) Isopropanol / 20 % (w/v) PEG 4000
F1	1.6	0.2 M Calcium chloride	0.1 M Sodium acetate	4.6	20 % (v/v) Isopropanol
G1	1.7	none	0.1 M tri-Sodium citrate	5.6	20 % (v/v) Isopropanol / 20 % (w/v) PEG 4000
H1	1.8	0.2 M tri-Sodium citrate	0.1 M HEPES sodium salt	7.5	20 % (v/v) Isopropanol
A2	1.9	0.2 M tri-Sodium citrate	0.1 M Sodium cacodylate	6.5	30 % (v/v) Isopropanol
B2	1.10	0.2 M Magnesium chloride	0.1 M HEPES sodium salt	7.5	30 % (v/v) Isopropanol
C2	1.11	0.2 M Ammonium acetate	0.1 M TRIS.HCl	8.5	30 % (v/v) Isopropanol
D2	1.12	1.5 M Sodium chloride	none		10 % (v/v) Ethanol
E2	1.13	none	0.1 M TRIS	8.5	20 % (v/v) Ethanol
F2	1.14	none	none		25 % (v/v) Ethylene glycol
G2	1.15	0.02 M Calcium chloride	0.1 M Sodium acetate	4.6	30 % (v/v) MPD
H2	1.16	0.2 M Sodium chloride	0.1 M Sodium acetate	4.6	30 % (v/v) MPD
A3	1.17	0.2 M Ammonium acetate	0.1 M tri-Sodium citrate	5.6	30 % (v/v) MPD
B3	1.18	0.2 M Magnesium acetate	0.1 M Sodium cacodylate	6.5	30 % (v/v) MPD
C3	1.19	0.2 M tri-Sodium citrate	0.1 M HEPES sodium salt	7.5	30 % (v/v) MPD
D3	1.20	0.5 M Ammonium sulfate	0.1 M HEPES	7.5	30 % (v/v) MPD
E3	1.21	0.2 M Ammonium phosphate	0.1 M TRIS	8.5	50 % (v/v) MPD
F3	1.22	none	0.1 M HEPES	7.5	70 % (v/v) MPD
G3	1.23	none	0.1 M TRIS	8.5	25 % (v/v) tert-Butanol
H3	1.24	none	0.1 M tri-Sodium citrate	5.6	35 % (v/v) tert-Butanol
A4	1.25	none	none		0.4 M Ammonium phosphate
B4	1.26	none	0.1 M tri-Sodium citrate	5.6	1.0 M Ammonium phosphate
C4	1.27	none	0.1 M TRIS.HCl	8.5	2.0 M Ammonium phosphate
D4	1.28	none	0.1 M HEPES	7.5	2.0 M Ammonium formate
E4	1.29	none	0.1 M Sodium acetate	4.6	2.0 M Ammonium sulfate
F4	1.30	none	0.1 M TRIS.HCl	8.5	2.0 M Ammonium sulfate
G4	1.31	none	none		2.0 M Ammonium sulfate
H4	1.32	0.1 M Sodium chloride	0.1 M HEPES	7.5	1.6 M Ammonium sulfate
A5	1.33	0.01 M Cobalt chloride	0.1 M MES	6.5	1.8 M Ammonium sulfate
B5	1.34	0.2 M K/Na tartrate	0.1 M tri-Sodium citrate	5.6	2.0 M Ammonium sulfate
C5	1.35	none	none		1.0 M Imidazole pH 7.0
D5	1.36	none	none		0.4 M K/Na tartrate
E5	1.37	none	0.1 M HEPES sodium salt	7.5	0.8 M K/Na tartrate
F5	1.38	none	0.1 M Imidazole	6.5	1.0 M Sodium acetate
G5	1.39	0.05 M Cadmium sulfate	0.1 M HEPES	7.5	1.0 M Sodium acetate
H5	1.40	none	0.1 M Sodium cacodylate	6.5	1.4 M Sodium acetate
A6	1.41	none	0.1 M Sodium acetate	4.6	2.0 M Sodium chloride
B6	1.42	0.1 M Sodium phosphate 0.1 M Potassium phosphate	0.1 M MES	6.5	2.0 M Sodium chloride
C6	1.43	none	0.1 M HEPES	7.5	4.3 M Sodium chloride
D6	1.44	none	0.1 M HEPES sodium salt	7.5	1.4 M tri-Sodium citrate
E6	1.45	none	none		1.6 M tri-Sodium citrate pH 6.5
F6	1.46	none	0.1 M HEPES sodium salt	7.5	0.8 M Sodium phosphate / 0.8 M Potassium phosphate
G6	1.47	none	0.1 M Sodium acetate	4.6	2.0 M Sodium formate
H6	1.48	none	none		4.0 M Sodium formate
A7	2.1	none	0.1 M BICINE	9.0	2 % (v/v) Dioxane / 10 % (w/v) PEG 20000
B7	2.2	1.6 M Ammonium sulfate	0.1 M MES	6.5	10 % (v/v) Dioxane
C7	2.3	none	none		35 % (v/v) Dioxane
D7	2.4	0.5 M Sodium chloride	0.1 M tri-Sodium citrate	5.6	2 % (v/v) Ethylene imine polymer
E7	2.5	1.5 M Ammonium sulfate	0.1 M TRIS	8.5	12 % (v/v) Glycerol
F7	2.6	0.5 M Sodium chloride 0.01 M Magnesium chloride	none		0.01 M CTAB
G7	2.7	0.01 M Ferric chloride	0.1 M tri-Sodium citrate	5.6	10 % (v/v) Jeffamine M-600
H7	2.8	none	0.1 M HEPES	7.5	20 % (v/v) Jeffamine M-600
A8	2.9	0.5 M Ammonium sulfate	0.1 M tri-Sodium citrate	5.6	1.0 M Lithium sulfate
B8	2.10	0.01 M Nickel chloride	0.1 M TRIS	8.5	1.0 M Lithium sulfate
C8	2.11	none	0.1 M HEPES sodium salt	7.5	1.5 M Lithium sulfate
D8	2.12	none	0.1 M BICINE	9.0	2.0 M Magnesium chloride
E8	2.13	none	none		0.2 M Magnesium formate
F8	2.14	none	0.1 M MES	6.5	1.6 M Magnesium sulfate
G8	2.15	none	0.1 M TRIS.HCl	8.5	8 % (w/v) PEG 8000
H8	2.16	none	0.1 M HEPES	7.5	10 % (w/v) PEG 8000
A9	2.17	0.5 M Lithium sulfate	none		15 % (w/v) PEG 8000
B9	2.18	0.2 M Zinc acetate	0.1 M Sodium cacodylate	6.5	18 % (w/v) PEG 8000
C9	2.19	0.2 M Calcium acetate	0.1 M Sodium cacodylate	6.5	18 % (w/v) PEG 8000
D9	2.20	0.2 M Magnesium acetate	0.1 M Sodium cacodylate	6.5	20 % (w/v) PEG 8000
E9	2.21	0.05 M Potassium phosphate	none		20 % (w/v) PEG 8000
F9	2.22	0.2 M Ammonium sulfate	0.1 M Sodium cacodylate	6.5	30 % (w/v) PEG 8000
G9	2.23	0.2 M Sodium acetate	0.1 M Sodium cacodylate	6.5	30 % (w/v) PEG 8000
H9	2.24	0.2 M Ammonium sulfate	none		30 % (w/v) PEG 8000
A10	2.25	2.0 M Ammonium sulfate	0.1 M HEPES sodium salt	7.5	2 % (v/v) PEG 400
B10	2.26	0.2 M Calcium chloride	0.1 M HEPES sodium salt	7.5	28 % (v/v) PEG 400
C10	2.27	0.1 M Cadmium chloride	0.1 M Sodium acetate	4.6	30 % (v/v) PEG 400
D10	2.28	0.2 M Magnesium chloride	0.1 M HEPES sodium salt	7.5	30 % (v/v) PEG 400
E10	2.29	0.2 M tri-Sodium citrate	0.1 M TRIS.HCl	8.5	30 % (v/v) PEG 400
F10	2.30	0.1 M Sodium chloride	0.1 M BICINE	9.0	20 % (w/v) PEG 550 MME
G10	2.31	0.01 M Zinc sulfate	0.1 M MES	6.5	25 % (w/v) PEG 550 MME
H10	2.32	none	none		10 % (w/v) PEG 1000 / 10 % (w/v) PEG 8000
A11	2.33	none	none		30 % (w/v) PEG 1500
B11	2.34	0.01 M Nickel chloride	0.1 M TRIS	8.5	20 % (w/v) PEG 2000 MME
C11	2.35	0.2 M Ammonium sulfate	0.1 M Sodium acetate	4.6	30 % (w/v) PEG 2000 MME
D11	2.36	none	0.1 M Sodium acetate	4.6	8 % (w/v) PEG 4000
E11	2.37	0.2 M Ammonium sulfate	0.1 M Sodium acetate	4.6	25 % (w/v) PEG 4000
F11	2.38	0.2 M Ammonium acetate	0.1 M Sodium acetate	4.6	30 % (w/v) PEG 4000
G11	2.39	0.2 M Ammonium acetate	0.1 M tri-Sodium citrate	5.6	30 % (w/v) PEG 4000
H11	2.40	0.2 M Magnesium chloride	0.1 M TRIS.HCl	8.5	30 % (w/v) PEG 4000
A12	2.41	0.2 M Lithium sulfate	0.1 M TRIS.HCl	8.5	30 % (w/v) PEG 4000
B12	2.42	0.2 M Sodium acetate	0.1 M TRIS.HCl	8.5	30 % (w/v) PEG 4000
C12	2.43	0.2 M Ammonium sulfate	none		30 % (w/v) PEG 4000
D12	2.44	0.2 M Ammonium sulfate	0.1 M MES	6.5	30 % (w/v) PEG 5000 MME
E12	2.45	none	0.1 M HEPES	7.5	10 % (w/v) PEG 6000 / 5 % (v/v) MPD
F12	2.46	2.0 M Sodium chloride	none		10 % (w/v) PEG 6000
G12	2.47	none	0.1 M HEPES	7.5	20 % (w/v) PEG 10000 / 8 % (v/v) Ethylene glycol
H12	2.48	none	0.1 M MES	6.5	12 % (w/v) PEG 20000

Annexe 12 : Composition des 96 solutions des kits 1 et 2 CLASSICS Suite de Qiagen.

Well	Kit #	Salt	Buffer	pH	Precipitant
A1	1.1	0.1 M Potassium chloride	none		12 % (w/v) PEG 8000 / 5 % (w/v) Glycerol
B1	1.2	0.5 M Potassium chloride	none		12 % (w/v) PEG 8000 / 10 % (w/v) Glycerol
C1	1.3	0.2 M Ammonium sulfate	none		15 % (w/v) PEG 8000
D1	1.4	0.5 M Lithium sulfate	none		15 % (w/v) PEG 8000
E1	1.5	0.2 M Sodium acetate	0.1 M MES	6.5	15 % (w/v) PEG 8000
F1	1.6	0.05 M Ammonium sulfate	0.1 M Sodium citrate		15 % (w/v) PEG 8000
G1	1.7	0.2 M Calcium acetate	0.1 M HEPES	7.5	18 % (w/v) PEG 8000
H1	1.8	0.1 M Sodium acetate	0.1 M HEPES	7.5	18 % (w/v) PEG 8000 / 2 % (w/v) Isopropanol
A2	1.9	0.2 M Lithium sulfate	0.1 M TRIS	8.5	18 % (w/v) PEG 8000
B2	1.10	none	0.1 M HEPES	7.5	20 % (w/v) PEG 8000
C2	1.11	0.2 M Magnesium acetate	0.1 M MES	6.5	20 % (w/v) PEG 8000
D2	1.12	none	0.1 M CHES	9.5	20 % (w/v) PEG 8000
E2	1.13	0.2 M Ammonium sulfate	0.1 M MES	6.5	22 % (w/v) PEG 8000
F2	1.14	0.2 M Lithium chloride	none		25 % (w/v) PEG 8000
G2	1.15	0.2 M Ammonium sulfate	none		30 % (w/v) PEG 8000
H2	1.16	none	0.1 M Sodium acetate	4.6	8 % (w/v) PEG 10000
A3	1.17	none	0.1 M Imidazole	8.0	14 % (w/v) PEG 10000
B3	1.18	none	0.1 M TRIS	8.5	16 % (w/v) PEG 10000
C3	1.19	0.1 M Sodium chloride	0.1 M TRIS	8.5	18 % (w/v) PEG 10000 / 20 % (w/v) Glycerol
D3	1.20	none	0.1 M HEPES	7.5	20 % (w/v) PEG 10000
E3	1.21	none	0.1 M TRIS	8.5	30 % (w/v) PEG 10000
F3	1.22	none	0.1 M MES	6.5	10 % (w/v) PEG 20000
G3	1.23	0.1 M Magnesium chloride	0.1 M TRIS	8.5	17 % (w/v) PEG 20000
H3	1.24	none	none		20 % (w/v) PEG 20000
A4	1.25	0.01 M Sodium acetate	none		50 % (w/v) MPD / 15 % (w/v) Ethanol
B4	1.26	0.05M Sodium chloride	0.05 M Sodium acetate		50 % (w/v) MPD / 20 % (w/v) Isopropanol
C4	1.27	0.1 M Ammonium phosphate	0.1 M TRIS	8.5	50 % (w/v) MPD
D4	1.28	none	none		55 % (w/v) MPD
E4	1.29	0.01 M Calcium chloride	0.1 M Sodium acetate	4.6	60 % (w/v) MPD
F4	1.30	0.02 M Sodium acetate	none		60 % (w/v) MPD
G4	1.31	none	0.1 M MES	6.5	70 % (w/v) MPD
H4	1.32	none	0.1 M TRIS	8.5	70 % (w/v) MPD
A5	1.33	0.01 M Calcium chloride	0.1 M TRIS	8.5	20 % (w/v) Methanol
B5	1.34	none	0.1 M TRIS	8.5	2 % (w/v) Ethanol
C5	1.35	none	0.1 M HEPES	7.5	5 % (w/v) Ethanol / 5 % (w/v) MPD
D5	1.36	0.2 M Sodium chloride	0.1 M TRIS	8.5	5 % (w/v) Ethanol / 5 % (w/v) MPD
E5	1.37	none	0.1 M TRIS	8.5	10 % (w/v) Ethanol
F5	1.38	none	0.1 M Sodium acetate	4.6	12 % (w/v) Ethanol / 4 % (w/v) PEG 400
G5	1.39	none	0.1 M TRIS	8.5	14 % (w/v) Ethanol / 5 % (w/v) Glycerol
H5	1.40	none	0.1 M TRIS	8.5	18 % (w/v) Ethanol
A6	1.41	none	none		20 % (w/v) Ethanol
B6	1.42	none	none		20 % (w/v) Ethanol / 10 % (w/v) Glycerol
C6	1.43	0.1 M Sodium acetate	none		30 % (w/v) Ethanol / 10 % (w/v) PEG 6000
D6	1.44	none	none		45 % (w/v) Ethanol
E6	1.45	0.01 M Sodium acetate	none		50 % (w/v) Ethanol
F6	1.46	0.05 M Sodium acetate	none		60 % (w/v) Ethanol / 1.5 % (w/v) PEG 6000
G6	1.47	0.1 M Sodium chloride	none		60 % (w/v) Ethanol
H6	1.48	0.01 M Magnesium sulfate	0.1 M TRIS	8.5	2 % (w/v) Isopropanol
A7	2.1	none	0.1 M HEPES	7.5	5 % (w/v) Isopropanol
B7	2.2	0.2 M Calcium chloride	0.1 M Sodium acetate	4.6	10 % (w/v) Isopropanol
C7	2.3	0.2 M Sodium citrate	0.1 M HEPES	7.5	10 % (w/v) Isopropanol
D7	2.4	0.01 M Magnesium chloride	0.1 M TRIS	8.5	10 % (w/v) Isopropanol
E7	2.5	0.05 M Sodium chloride	0.1 M TRIS	8.5	12 % (w/v) Isopropanol
F7	2.6	0.2 M Sodium citrate	0.1 M MES	6.5	15 % (w/v) Isopropanol
G7	2.7	0.2 M Sodium citrate	0.1 M HEPES	7.5	15 % (w/v) Isopropanol
H7	2.8	0.2 M Magnesium chloride	0.1 M HEPES	7.5	15 % (w/v) Isopropanol
A8	2.9	0.2 M Ammonium acetate	0.1 M TRIS	8.5	15 % (w/v) Isopropanol
B8	2.10	0.2 M Calcium chloride	0.1 M Sodium acetate	4.6	20 % (w/v) Isopropanol
C8	2.11	0.2 M Sodium citrate	0.1 M HEPES	7.5	20 % (w/v) Isopropanol
D8	2.12	0.1 M Magnesium chloride	0.1 M HEPES	7.5	25 % (w/v) Isopropanol
E8	2.13	0.2 M Sodium citrate	0.1 M MES	6.5	30 % (w/v) Isopropanol
F8	2.14	0.2 M Magnesium chloride	0.1 M HEPES	7.5	30 % (w/v) Isopropanol
G8	2.15	0.2 M Ammonium acetate	0.1 M TRIS	8.5	30 % (w/v) Isopropanol
H8	2.16	0.1 M Calcium chloride	0.1 M TRIS	8.5	25 % (w/v) tert-Butanol
A9	2.17	none	0.1 M Sodium citrate	5.6	35 % (w/v) tert-Butanol
B9	2.18	none	none		0.2 M Ammonium dihydrogen phosphate
C9	2.19	none	none		0.2 M Potassium/Sodium tartrate
D9	2.20	none	none		0.2 M Magnesium acetate
E9	2.21	none	none		0.4 M Ammonium dihydrogen phosphate
F9	2.22	none	none		0.4 M Potassium/Sodium tartrate
G9	2.23	none	0.1 M TRIS	8.5	0.4 M Potassium/Sodium tartrate
H9	2.24	0.2 M Sodium citrate	none		0.5 M Ammonium dihydrogen phosphate
A10	2.25	none	0.1 M Imidazole	8.0	0.5 M Sodium acetate
B10	2.26	none	0.1 M HEPES	7.5	0.7 M Sodium citrate
C10	2.27	none	0.1 M TRIS	8.5	0.7 M Lithium sulfate
D10	2.28	none	0.1 M HEPES	7.5	0.8 M Potassium/Sodium tartrate
E10	2.29	none	0.1 M Sodium citrate	5.6	1.0 M Ammonium dihydrogen phosphate
F10	2.30	none	0.1 M TRIS	8.5	1.0 M Ammonium dihydrogen phosphate
G10	2.31	0.01 M Nickel chloride	0.1 M TRIS	8.5	1.0 M Lithium sulfate
H10	2.32	none	0.1 M Imidazole	8.0	1.0 M Sodium acetate
A11	2.33	none	0.1 M Sodium acetate	4.6	1.0 M Sodium formate
B11	2.34	none	0.1 M MES	6.5	1.4 M Sodium acetate
C11	2.35	none	0.1 M HEPES	7.5	1.4 M Sodium citrate
D11	2.36	none	0.1 M TRIS	8.5	1.5 M Lithium sulfate
E11	2.37	none	1 M Sodium citrate	6.5	none
F11	2.38	none	0.1 M MES	6.5	1.6 M Magnesium sulfate
G11	2.39	none	0.1 M MES	6.5	1.6 M Potassium/Sodium tartrate
H11	2.40	none	0.1 M MES	6.5	2.0 M Ammonium formate
A12	2.41	none	0.1 M TRIS	8.5	2.0 M Ammonium dihydrogen phosphate
B12	2.42	none	none		2.0 M Sodium formate
C12	2.43	none	0.1 M TRIS	8.5	2.0 M Magnesium chloride
D12	2.44	0.2 M Sodium acetate	0.1 M MES	6.5	2.0 M Sodium chloride
E12	2.45	none	0.1 M Sodium acetate	4.6	2.0 M Sodium formate
F12	2.46	none	0.1 M TRIS	8.5	1.0 M Ammonium dihydrogen phosphate / 30 % (w/v) Glycerol
G12	2.47	none	0.1 M HEPES	7.5	4.0 M Sodium chloride
H12	2.48	none	none		3.0 M Sodium formate

Annexe 13 : Composition des 96 solutions des kits 1 et 2 COMPAS de Qiagen.

Well	Kit #	Salt	Buffer	pH	Precipitant
A1	1.1	0.01 M Magnesium chloride	0.05 M MES	5.6	2 M Lithium sulfate
B1	1.2	0.01 M Magnesium acetate	0.05 M MES	5.6	2.5 M Ammonium sulfate
C1	1.3	0.1 M Magnesium acetate	0.05 M MES	5.6	20 % (v/v) MPD
D1	1.4	0.2 M Potassium chloride	0.05 M MES	5.6	10 % (v/v) PEG 400
E1	1.5	0.2 M Potassium chloride	0.05 M MES	5.6	5 % (w/v) PEG 8000
F1	1.6	0.1 M Ammonium sulfate	0.05 M MES	5.6	20 % (w/v) PEG 8000
G1	1.7	0.02 M Magnesium chloride	0.05 M MES	6.0	15 % (v/v) Isopropanol
H1	1.8	0.1 M Ammonium acetate	0.05 M MES	6.0	0.6 M Sodium chloride
A2	1.9	0.1 M Potassium chloride	0.01 M Magnesium chloride	6.0	10 % (v/v) PEG 400
B2	1.10	0.005 M Magnesium sulfate	0.05 M MES	6.0	5 % (w/v) PEG 4000
C2	1.11	0.01 M Magnesium chloride	0.05 M Sodium cacodylate	6.0	1 M Lithium sulfate
D2	1.12	0.01 M Magnesium sulfate	0.05 M Sodium cacodylate	6.0	1.8 M Lithium sulfate
E2	1.13	0.015 M Magnesium acetate	0.05 M Sodium cacodylate	6.0	1.7 M Ammonium sulfate
F2	1.14	0.1 M Potassium chloride	0.05 M Sodium cacodylate	6.0	15 % (v/v) Isopropanol
G2	1.15	0.04 M Magnesium chloride	0.05 M Sodium cacodylate	6.0	5 % (v/v) MPD
H2	1.16	0.04 M Magnesium acetate	0.05 M Sodium cacodylate	6.0	30 % (v/v) MPD
A3	1.17	0.2 M Potassium chloride	0.01 M Calcium chloride	6.0	10 % (w/v) PEG 4000
B3	1.18	0.01 M Magnesium acetate	0.05 M Sodium cacodylate	6.5	1.3 M Lithium sulfate
C3	1.19	0.01 M Magnesium sulfate	0.05 M Sodium cacodylate	6.5	2 M Ammonium sulfate
D3	1.20	0.1 M Ammonium acetate	0.015 M Magnesium acetate	6.5	10 % (v/v) Isopropanol
E3	1.21	0.2 M Potassium chloride	0.005 M Magnesium chloride	6.5	10 % (w/v) 1,6-Hexanediol
F3	1.22	0.08 M Magnesium acetate	0.05 M Sodium cacodylate	6.5	15 % (v/v) PEG 400
G3	1.23	0.2 M Potassium chloride	0.01 M Magnesium chloride	6.5	10 % (w/v) PEG 4000
H3	1.24	0.2 M Ammonium acetate	0.01 M Calcium chloride	6.5	10 % (w/v) PEG 4000
A4	1.25	0.08 M Magnesium acetate	0.05 M Sodium cacodylate	6.5	30 % (w/v) PEG 4000
B4	1.26	0.2 M Potassium chloride	0.1 M Magnesium acetate	6.5	10 % (w/v) PEG 8000
C4	1.27	0.2 M Ammonium acetate	0.01 M Magnesium acetate	6.5	30 % (w/v) PEG 8000
D4	1.28	0.05 M Magnesium sulfate	0.05 M HEPES Sodium salt	7.0	1.6 M Lithium sulfate
E4	1.29	0.01 M Magnesium chloride	0.05 M HEPES Sodium salt	7.0	4 M Lithium chloride
F4	1.30	0.01 M Magnesium chloride	0.05 M HEPES Sodium salt	7.0	1.6 M Ammonium sulfate
G4	1.31	0.005 M Magnesium chloride	0.05 M HEPES Sodium salt	7.0	25 % (v/v) PEG MME 550
H4	1.32	0.2 M Potassium chloride	0.01 M Magnesium chloride	7.0	20 % (w/v) 1,6-Hexanediol
A5	1.33	0.2 M Ammonium chloride	0.01 M Magnesium chloride	7.0	30 % (w/v) 1,6-Hexanediol
B5	1.34	0.1 M Potassium chloride	0.005 M Magnesium sulfate	7.0	15 % (v/v) MPD
C5	1.35	0.1 M Potassium chloride	0.01 M Magnesium chloride	7.0	5 % (v/v) PEG 400
D5	1.36	0.1 M Potassium chloride	0.01 M Calcium chloride	7.0	10 % (v/v) PEG 400
E5	1.37	0.2 M Potassium chloride	0.025 M Magnesium sulfate	7.0	20 % (v/v) PEG 200
F5	1.38	0.2 M Ammonium acetate	0.15 M Magnesium acetate	7.0	5 % (w/v) PEG 4000
G5	1.39	0.1 M Ammonium acetate	0.02 M Magnesium chloride	7.0	5 % (w/v) PEG 8000
H5	1.40	0.01 M Magnesium chloride	0.05 M TRIS.HCl	7.5	1.6 M Ammonium sulfate
A6	1.41	0.1 M Potassium chloride	0.015 M Magnesium chloride	7.5	10 % (v/v) PEG MME 550
B6	1.42	0.01 M Magnesium acetate	0.05 M TRIS.HCl	7.5	5 % (v/v) Isopropanol
C6	1.43	0.05 M Ammonium acetate	0.01 M Magnesium chloride	7.5	10 % (v/v) MPD
D6	1.44	0.2 M Potassium chloride	0.05 M Magnesium chloride	7.5	10 % (w/v) PEG 4000
E6	1.45	0.025 M Magnesium sulfate	0.05 M TRIS.HCl	8.5	1.8 M Ammonium sulfate
F6	1.46	0.005 M Magnesium sulfate	0.05 M TRIS.HCl	8.5	35 % (w/v) 1,6-Hexanediol
G6	1.47	0.1 M Potassium chloride	0.01 M Magnesium chloride	8.5	30 % (v/v) PEG 400
H6	1.48	0.2 M Ammonium chloride	0.01 M Calcium chloride	8.5	30 % (w/v) PEG 4000
A7	2.1	2.5 mM Spermine	80 mM Magnesium chloride	0.05 M HEPES	7.5 none
B7	2.2	2.25 mM Spermine 1 mM Copper sulfate	18 mM Magnesium chloride	0.05 M Sodium cacodylate	6.0 9 % (v/v) Isopropanol
C7	2.3	0.9 mM Spermine 1.8 mM Cobalt(III)hexamine chloride	18 mM Magnesium chloride	0.05 M Sodium cacodylate	6.5 9 % (v/v) Isopropanol
D7	2.4	2.25 mM Spermine	18 mM Magnesium chloride	0.05 M Sodium cacodylate	6.5 9 % (v/v) Isopropanol
E7	2.5	2.25 mM Spermine 0.9 mM Cobalt(III)hexamine chloride	18 mM Magnesium chloride	0.05 M Sodium cacodylate	7.0 4.5 % (v/v) MPD
F7	2.6	2.25 mM Spermine	36 mM Magnesium chloride	0.05 M Sodium cacodylate	6.5 5 % (v/v) PEG 400
G7	2.7	2.0 mM Cobalt(III)hexamine chloride	10 mM Magnesium chloride	0.05 M Sodium succinate	5.5 10 % (v/v) Isopropanol
H7	2.8	1.0 mM Spermine	20 mM Magnesium chloride	0.05 M Sodium cacodylate	6.0 15 % (v/v) Ethanol
A8	2.9	1.0 mM Spermine 1.0 mM Cobalt(III)hexamine chloride	20 mM Magnesium chloride	0.05 M Sodium cacodylate	7.0 15 % (v/v) Ethanol
B8	2.10	1.0 mM Spermidine	5 mM Magnesium chloride	0.05 M Sodium cacodylate	7.0 10 % (v/v) tert-Butanol
C8	2.11	2.5 mM Spermine	30 mM Magnesium chloride	0.05 M Sodium cacodylate	7.0 5 % (v/v) PEG 400
D8	2.12	2.0 mM Cobalt(III)hexamine chloride	100 mM Magnesium chloride	0.05 M Sodium cacodylate	6.5 5 % (v/v) Isopropanol
E8	2.13	1.0 mM Cobalt(III)hexamine chloride	10 mM Magnesium chloride	0.05 M TRIS	8.0 20 % (v/v) Ethanol
F8	2.14	1.0 mM Spermine	20 mM Magnesium chloride	0.05 M HEPES	7.5 5 % (w/v) PEG 8000
G8	2.15	2.5 mM Spermine	20 mM Magnesium chloride	0.05 M Sodium cacodylate	6.0 5 % (w/v) PEG 4000
H8	2.16	2.5 mM Spermine 5 mM Calcium chloride	10 mM Magnesium chloride	0.05 M Sodium cacodylate	6.0 10 % (v/v) Isopropanol
A9	2.17	2.25 mM Spermine 0.9 mM Spermidine	9 mM Magnesium chloride 1.8 mM Cobalt(III)hexamine chloride	0.05 M Sodium cacodylate	7.0 5 % (v/v) PEG 400
B9	2.18	2.5 mM Spermine 1 mM Copper sulfate	10 mM Magnesium chloride	0.05 M Sodium cacodylate	6.5 10 % (v/v) Isopropanol
C9	2.19	1.0 mM Spermine 2 mM Calcium chloride	20 mM Magnesium chloride	0.05 M Sodium cacodylate	6.0 10 % (v/v) 1,5-Hexanediol
D9	2.20	1.0 mM Spermidine	15 mM Magnesium chloride	0.05 M HEPES	7.5 10 % (v/v) Dioxane
E9	2.21	3.0 mM Spermine	15 mM Magnesium chloride	0.05 M Sodium cacodylate	6.0 10 % (w/v) PEG 400
F9	2.22	18 mM Calcium chloride	2.5 mM Spermine	0.05 M Sodium cacodylate	6.5 9 % (v/v) Isopropanol
G9	2.23	1.0 mM Cobalt(III)hexamine chloride 80 mM Calcium chloride	2.0 mM Spermine	0.05 M Sodium cacodylate	6.5 none
H9	2.24	2.5 mM Cobalt(III)hexamine chloride	5 mM Magnesium chloride	0.05 M Sodium cacodylate	6.5 none
A10	2.25	1.0 mM Spermine	30 mM Magnesium chloride	0.05 M Sodium cacodylate	6.5 1.3 M Lithium sulfate
B10	2.26	200 mM Calcium acetate	0.05 M Sodium cacodylate	6.0	5 % (v/v) Isopropanol
C10	2.27	1.0 mM Cobalt(III)hexamine chloride	100 mM Magnesium chloride	0.05 M Sodium cacodylate	6.5 10 % (v/v) Ethanol
D10	2.28	2.5 mM Spermidine	10 mM Magnesium chloride	0.05 M Sodium cacodylate	6.0 2.5 M Sodium chloride
E10	2.29	200 mM tri-Na citrate	10 mM Magnesium chloride	0.05 M Sodium cacodylate	6.5 5 % (v/v) Isopropanol
F10	2.30	10.0 mM Spermine	15 mM Magnesium chloride	0.05 M Sodium cacodylate	6.5 2.0 M Lithium sulfate
G10	2.31	1.0 mM Spermine	20 mM Magnesium chloride	0.05 M Sodium cacodylate	6.5 2.0 M Ammonium sulfate
H10	2.32	1.5 mM Spermine	10 mM Magnesium chloride	0.05 M Sodium cacodylate	6.5 3.0 M Ammonium sulfate
A11	2.33	1.0 mM Spermine	15 mM Magnesium chloride	0.05 M HEPES	7.5 1.0 M Ammonium sulfate
B11	2.34	200 mM Calcium acetate	0.05 M Sodium cacodylate	6.0	2.5 M Sodium chloride
C11	2.35	1.0 mM Cobalt(III)hexamine chloride	200 mM Calcium acetate	0.05 M Sodium cacodylate	6.0 2.0 M Lithium chloride
D11	2.36	5.0 mM Spermidine 1.0 mM Cobalt(III)hexamine chloride	15 mM Magnesium chloride	0.05 M Sodium cacodylate	6.5 2.0 M Sodium chloride
E11	2.37	100 mM Sodium chloride	200 mM Magnesium chloride	0.05 M Sodium cacodylate	6.5 20 % (w/v) PEG 1000
F11	2.38	50 mM Magnesium chloride	0.05 M TRIS	7.5	1.0 M Sodium tartrate
G11	2.39	200 mM Magnesium chloride	0.05 M TRIS	7.5	2.5 M Sodium chloride
H11	2.40	200 mM Magnesium chloride	0.05 M Sodium cacodylate	6.0	2.5 M Potassium chloride
A12	2.41	200 mM Magnesium chloride	0.05 M TRIS	8.0	15 % (w/v) Ethanol
B12	2.42	5.0 mM Spermidine	15 mM Magnesium chloride	0.05 M Sodium cacodylate	6.0 2.0 M Lithium sulfate
C12	2.43	0.5 mM Spermine 100 mM Sodium chloride	20 mM Magnesium acetate	0.05 M Sodium cacodylate	6.0 25 % (v/v) MPD
D12	2.44	0.5 mM Spermine	20 mM Magnesium chloride	0.05 M Sodium succinate	5.5 3.0 M Ammonium sulfate
E12	2.45	5 mM Cobalt(III)hexamine chloride	0.05 M Sodium cacodylate	6.5	2.5 M Potassium chloride
F12	2.46	2.0 mM Cobalt(III)hexamine chloride	50 mM Magnesium chloride	0.05 M Sodium cacodylate	6.5 1.5 M Lithium sulfate
G12	2.47	2.0 mM Cobalt(III)hexamine chloride 30 mM Calcium chloride	1.0 mM Spermine	0.05 M Sodium cacodylate	6.5 2.0 M Lithium chloride
H12	2.48	50 mM Spermine	0.05 M Sodium cacodylate	6.5	10 mM Magnesium chloride

Annexe 14 : Composition des 96 solutions des kits 1 et 2 NUCLEIX de Qiagen.

Well	Kit #	Salt	Buffer	pH	Precipitant
A1	1.1	none	0.1M CHES	9.5	20% (w/v) PEG-8000
B1	1.2	0.1M NaCl	0.1M HEPES	7.5	10% (v/v) 2-propanol
C1	1.3	none	0.1M CHES	9.5	15% (v/v) ethanol
D1	1.4	0.1M MgCl ₂	0.1M imidazole	8.0	35% (v/v) 2-methyl-2,4-pentanediol
E1	1.5	none	0.1M CAPS	10.5	30% (v/v) PEG-400
F1	1.6	none	0.1M citrate	5.5	20% (w/v) PEG-3000
G1	1.7	0.1M Zn(OAc) ₂	0.1M MES	6.0	10% (w/v) PEG-8000
H1	1.8	none	0.1M citrate	5.5	2.0 M (NH ₄) ₂ SO ₄
A2	1.9	none	0.1M acetate	4.5	1.0 M (NH ₄) ₂ HPO ₄
B2	1.10	none	0.1M Tris	7.0	20% (w/v) PEG-2000 MME
C2	1.11	0.1M Li ₂ SO ₄	0.1M MES	6.0	20% (v/v) 1,4-butanediol
D2	1.12	0.1M Ca(OAc) ₂	0.1M imidazole	8.0	20% (w/v) PEG-1000
E2	1.13	none	0.1M cacodylate	6.5	1.26 M (NH ₄) ₂ SO ₄
F2	1.14	none	0.1M cacodylate	6.5	1.0 M sodium citrate
G2	1.15	0.1M Li ₂ SO ₄	0.1M imidazole	8.0	10% (w/v) PEG-3000
H2	1.16	none	0.1M Na/K phosphate	6.2	2.5 M NaCl
A3	1.17	0.1M Li ₂ SO ₄	0.1M acetate	4.5	30% (w/v) PEG-8000
B3	1.18	0.1M NaCl	0.1M imidazole	8.0	1.0 M K/Na tartrate
C3	1.19	none	0.1M Tris	7.0	20% (w/v) PEG-1000
D3	1.20	0.1M NaCl	0.1M imidazole	8.0	0.4 M NaH ₂ PO ₄ /1.6 M K ₂ HPO ₄
E3	1.21	none	0.1M HEPES	7.5	20% (w/v) PEG-8000
F3	1.22	none	0.1M Tris	8.5	10% (v/v) 2-propanol
G3	1.23	0.1M MgCl ₂	0.1M imidazole	8.0	15% (v/v) ethanol
H3	1.24	0.1M NaCl	0.1M Tris	7.0	35% (v/v) 2-methyl-2,4-pentanediol
A4	1.25	0.1M MgCl ₂	0.1M Tris	8.5	30% (v/v) PEG-400
B4	1.26	none	0.1M CHES	9.5	10% (w/v) PEG-3000
C4	1.27	0.1M Li ₂ SO ₄	0.1M CAPS	10.5	1.2 M NaH ₂ PO ₄ /0.8 M K ₂ HPO ₄
D4	1.28	0.1M NaCl	0.1M HEPES	7.5	20% (w/v) PEG-3000
E4	1.29	0.1M NaCl	0.1M CHES	9.5	10% (w/v) PEG-8000
F4	1.30	0.1M NaCl	0.1M acetate	4.5	1.26 M (NH ₄) ₂ SO ₄
G4	1.31	0.1M NaCl	0.1M phosphate-citrate	4.2	20% (w/v) PEG-8000
H4	1.32	none	0.1M Na/K phosphate	6.2	10% (w/v) PEG-3000
A5	1.33	0.1M Li ₂ SO ₄	0.1M CAPS	10.5	2.0 M (NH ₄) ₂ SO ₄
B5	1.34	none	0.1M imidazole	8.0	1.0 M (NH ₄) ₂ HPO ₄
C5	1.35	none	0.1M acetate	4.5	20% (v/v) 1,4-butanediol
D5	1.36	none	0.1M imidazole	8.0	1.0 M sodium citrate
E5	1.37	none	0.1M imidazole	8.0	2.5 M NaCl
F5	1.38	0.1M Li ₂ SO ₄	0.1M CHES	9.5	1.0 M K/Na tartrate
G5	1.39	0.1M Li ₂ SO ₄	0.1M phosphate-citrate	4.2	20% (w/v) PEG-1000
H5	1.40	0.1M Ca(OAc) ₂	0.1M MES	6.0	10% (v/v) 2-propanol
A6	1.41	none	0.1M CHES	9.5	30% (w/v) PEG-3000
B6	1.42	none	0.1M Tris	7.0	15% (v/v) ethanol
C6	1.43	none	0.1M Na/K phosphate	6.2	35% (v/v) 2-methyl-2,4-pentanediol
D6	1.44	0.1M Ca(OAc) ₂	0.1M acetate	4.5	30% (v/v) PEG-400
E6	1.45	none	0.1M acetate	4.5	20% (w/v) PEG-3000
F6	1.46	0.1M Ca(OAc) ₂	0.1M imidazole	8.0	10% (w/v) PEG-8000
G6	1.47	0.1M Li ₂ SO ₄	0.1M Tris	8.5	1.26 M (NH ₄) ₂ SO ₄
H6	1.48	0.1M Zn(OAc) ₂	0.1M acetate	4.5	20% (w/v) PEG-1000

Annexe 15 : Composition des 48 solutions du kit 1 WIZARD de Emerald Biostructures.

Well	Kit #	Salt	Buffer	pH	Preprecipitant	Cryoprotectant
A1	1.1	0.02M calcium chloride dihydrate	0.1M sodium acetate trihydrate	4.6	30% (v/v) MPD	none
B1	1.2	none	none		0.26M potassium sodium tartrate tetrahydrate	35% (v/v) glycerol
C1	1.3	none	none		0.26M ammonium phosphate monobasic	35% (v/v) glycerol
D1	1.4	none	0.075M TRIS/HCl	8.5	1.5M ammonium sulfate	25% (v/v) glycerol
E1	1.5	0.2M sodium citrate tribasic dihydrate	0.1M HEPES sodium	7.5	30% (v/v) MPD	none
F1	1.6	0.16M magnesium chloride hexahydrate	0.08M TRIS/HCl	8.5	24% (w/v) PEG-4000	20% (v/v) glycerol
G1	1.7	none	0.07M sodium cacodylate trihydrate	6.5	0.98M sodium acetate trihydrate	30% (v/v) glycerol
H1	1.8	0.14M sodium citrate tribasic dihydrate	0.07M sodium cacodylate trihydrate	6.5	21% (v/v) 2-propanol	30% (v/v) glycerol
A2	1.9	0.17M ammonium acetate	0.085M sodium citrate tribasic dihydrate	5.6	25.5% (w/v) PEG-4000	15% (v/v) glycerol
B2	1.10	0.17M ammonium acetate	0.085M sodium acetate trihydrate	4.6	25.5% (w/v) PEG-4000	15% (v/v) glycerol
C2	1.11	none	0.07M sodium citrate tribasic dihydrate	5.6	0.7M ammonium phosphate monobasic	30% (v/v) glycerol
D2	1.12	0.18M magnesium chloride hexahydrate	0.09M HEPES sodium	7.5	27% (v/v) 2-propanol	10% (v/v) glycerol
E2	1.13	0.2M sodium citrate tribasic dihydrate	0.1M TRIS/HCl	8.5	30% (v/v) PEG-400	none
F2	1.14	0.19M calcium chloride dihydrate	0.095M HEPES sodium	7.5	26.6% (w/v) PEG-400	5% (v/v) glycerol
G2	1.15	0.17M ammonium sulfate	0.085M sodium cacodylate trihydrate	6.5	25.5% (w/v) PEG-8000	15% (v/v) glycerol
H2	1.16	none	0.075M HEPES sodium	7.5	1.125M lithium sulfate monohydrate	25% (v/v) glycerol
A3	1.17	0.17M lithium sulfate monohydrate	0.085M TRIS/HCl	8.5	25.5% (w/v) PEG-4000	15% (v/v) glycerol
B3	1.18	0.16M magnesium acetate tetrahydrate	0.08M sodium cacodylate trihydrate	6.5	16% (w/v) PEG-8000	20% (v/v) glycerol
C3	1.19	0.16M ammonium acetate	0.08M TRIS/HCl	8.5	24% (v/v) 2-propanol	20% (v/v) glycerol
D3	1.20	0.16M ammonium sulfate	0.08M sodium acetate trihydrate	4.6	20% (w/v) PEG-4000	20% (v/v) glycerol
E3	1.21	0.2M magnesium acetate tetrahydrate	0.1M sodium cacodylate trihydrate	6.5	30% (v/v) MPD	none
F3	1.22	0.17M sodium acetate trihydrate	0.085M TRIS/HCl	8.5	25.5% (w/v) PEG-4000	15% (v/v) glycerol
G3	1.23	0.2M magnesium chloride hexahydrate	0.1M HEPES sodium	7.5	30% (v/v) PEG-400	none
H3	1.24	0.14M calcium chloride dihydrate	0.07M sodium acetate trihydrate	4.6	14% (v/v) 2-propanol	30% (v/v) glycerol
A4	1.25	none	0.07M imidazole	5.6	0.7M sodium acetate trihydrate	30% (v/v) glycerol
B4	1.26	0.2M ammonium acetate	0.1M sodium citrate tribasic dihydrate	5.6	30% (v/v) MPD	none
C4	1.27	0.14M sodium citrate tribasic dihydrate	0.07M HEPES sodium	7.5	14% (v/v) 2-propanol	30% (v/v) glycerol
D4	1.28	0.17M sodium acetate trihydrate	0.085M sodium cacodylate trihydrate	6.5	25.5% (w/v) PEG-8000	15% (v/v) glycerol
E4	1.29	none	0.065M HEPES sodium	7.5	0.52M potassium sodium tartrate tetrahydrate	35% (v/v) glycerol
F4	1.30	0.17M ammonium sulfate	none		25.5% (w/v) PEG-8000	15% (v/v) glycerol
G4	1.31	0.17M ammonium sulfate	none		25.5% (w/v) PEG-4000	15% (v/v) glycerol
H4	1.32	none	none		1.5M ammonium sulfate	25% (v/v) glycerol
A5	1.33	none	none		3.6M sodium formate	10% (v/v) glycerol
B5	1.34	none	0.07M sodium acetate trihydrate	4.6	1.4M sodium formate	30% (v/v) glycerol
C5	1.35	none	0.075M HEPES sodium	7.5	0.6M sodium phosphate monobasic monohydrate	25% (v/v) glycerol
D5	1.36	none	0.065M TRIS/HCl	8.5	0.6M potassium phosphate monobasic	
E5	1.37	none	0.07M sodium acetate trihydrate	4.6	5.2% (w/v) PEG-8000	35% (v/v) glycerol
F5	1.38	none	0.09M HEPES sodium	7.5	5.6% (w/v) PEG-4000	30% (v/v) glycerol
G5	1.39	none	0.085M HEPES sodium	7.5	0.26M sodium citrate tribasic dihydrate	10% (v/v) glycerol
H5	1.40	none	0.095M sodium citrate tribasic dihydrate	5.6	1.7% (v/v) PEG-400	15% (v/v) glycerol
A6	1.41	none	0.085M HEPES sodium	7.5	1.7M ammonium sulfate	
B6	1.42	0.04M potassium phosphate monobasic	none		19% (v/v) 2-propanol	5% (v/v) glycerol
C6	1.43	none	none		19% (w/v) PEG-4000	
D6	1.44	none	none		8.5% (v/v) 2-propanol	15% (v/v) glycerol
E6	1.45	0.16M zinc acetate dihydrate	0.08M sodium cacodylate trihydrate	6.5	17% (w/v) PEG-4000	
F6	1.46	0.16M calcium acetate hydrate	0.08M sodium cacodylate trihydrate	6.5	16% (w/v) PEG-8000	20% (v/v) glycerol
G6	1.47	none	0.08M sodium acetate trihydrate	4.6	14.4% (w/v) PEG-8000	20% (v/v) glycerol
H6	1.48	none	0.08M TRIS/HCl	8.5	14.4% (w/v) PEG-8000	20% (v/v) glycerol
A7	1.49	0.8M lithium sulfate monohydrate	none		1.6M ammonium sulfate	20% (v/v) glycerol
B7	1.50	0.4M lithium sulfate monohydrate	none		1.6M ammonium phosphate monobasic	20% (v/v) glycerol
					1.6% (w/v) PEG-8000	20% (v/v) glycerol
					12% (w/v) PEG-8000	20% (v/v) glycerol

Annexe 16 : Composition des 50 solutions du kit CRYO de Sigma.

Article dans une revue :

. Karine Loth, [Justine Largillière](#), Franck Coste, Françoise Culard, Céline Landon, *et al.* **New protein-DNA complexes in archaea: a small monomeric protein induces a sharp V-turn DNA structure.** *Scientific Reports*, Nature Publishing Group, 2019, 9 (1), ([10.1038/s41598-019-50211-2](#)). ([hal-02306224](#))

Communication orale :

. Justine Largillière, Rémy Le Meur, David Guinguène, Norbert Garnier, Céline Landon, *et al.* **Vers une description cinétique et structurale complète du mécanisme d'échange de chaînes de la protéine HU d'*Escherichia coli*.** *GERM* 2019, May 2019, Oléron, France. ([hal-02361645](#))

Vulgarisation scientifique :

. [Justine Largillière](#). **L'échange de chaînes de la protéine HU comme nouvelle cible contre les entérobactéries.** *Ma thèse en 180 secondes*, édition 2019, Mar 2019, Orléans, France. ([hal-02362865](#))

Poster :

. Justine Largillière, Rémy Le Meur, David Guinguène, Norbert Garnier, Céline Landon, *et al.* **To a complete kinetics and structural description of the EchU chain exchange system.** *Journée Fédération de recherche "Physique et Chimie du Vivant"*, Apr 2019, Orléans, France. ([hal-02362578](#))

. Justine Largillière, Rémy Le Meur, David Guinguène, Norbert Garnier, Céline Landon, *et al.* **To a complete kinetics and structural description of the EchU chain exchange system.** *NMR : a tool for biology XITH*, Jan 2019, Paris, France. ([hal-02362588](#))

. Justine Largillière, Rémy Le Meur, David Guinguène, Norbert Garnier, Céline Landon, *et al.* **To a complete kinetics and structural description of the EchU chain exchange system.** *31e colloque Biotechnocentre*, Oct 2018, Seillac, France. ([hal-02362818](#))

. Justine Largillière, Rémy Le Meur, Norbert Garnier, Céline Landon, A.F. Delmas, *et al.* **Native and Intermediate conformations of the three HU dimers of *E. coli*.** *XXVIII ICMRBS 2018*, Aug 2018, Dublin, Ireland. ([hal-02362585](#))

. Justine Largillière, Rémy Le Meur, David Guinguène, Céline Landon, A.F. Delmas, *et al.* **First kinetics description of the EchU chain exchange system.** *XXVIII ICMRBS 2018*, Aug 2018, Dublin, Ireland. ([hal-02362586](#))

. Karine Loth, [Justine Largillière](#), Franck Coste, Françoise Culard, Bertrand Castaing, *et al.* **New protein-DNA complexes in archaea: a small monomeric protein induces a V-turn structure.** *XXVIII ICMRBS 2018*, Aug 2018, Dublin, Ireland. ([hal-02362584](#))

. Karine Loth, [Justine Largillière](#), Franck Coste, Françoise Culard, Céline Landon, *et al.* **New protein-DNA complexes in archaea: a small monomeric protein induces a sharp V-turn DNA structure.** *EuroMar 2018*, Jul 2018, Nantes, France. ([hal-02362581](#))

. Karine Loth, Françoise Culard, [Justine Largillière](#), Céline Landon, Bertrand Castaing, *et al.* **New protein-DNA complexes in archaea: a monomeric protein induces a U-turn structure.** *5e séminaire scientifique du LIA MiR-TANGO*, May 2018, Orléans, France. ([hal-02362804](#))

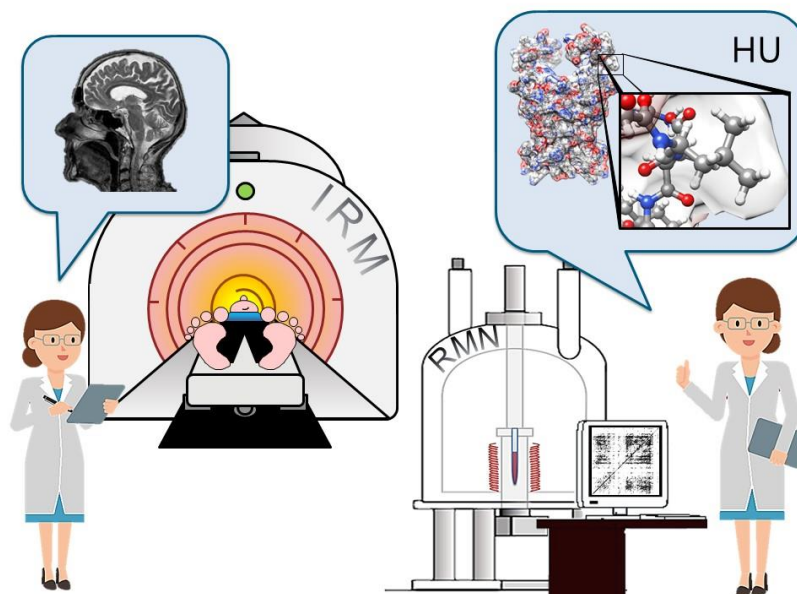
. Justine Largillière, Rémy Le Meur, Martine Cadène, Bertrand Castaing, Karine Loth, *et al.* **Premières études cinétiques du mécanisme d'échange de chaînes de la protéine HU d'*Escherischia coli*.** *16e colloque scientifique de l'ADOC : Converger pour Renforcer*, Apr 2018, Orléans, France. ([hal-02362579](#))

. Justine Largillière, Rémy Le Meur, Martine Cadène, Bertrand Castaing, Karine Loth, *et al.* **Premières études cinétiques du mécanisme d'échange de chaînes de la protéine HU d'*Escherischia coli*.** *Journée Fédération de recherche "Physique et Chimie du Vivant"*, Jan 2018, Orléans, France. ([hal-02362577](#))

. Karine Loth, [Justine Largillière](#), Franck Coste, Françoise Culard, A.F. Delmas, *et al.* **New protein-DNA complexes in archaea: a monomeric protein induces a U-turn structure.** *Structural biology and biophysics of RNA-protein complexes*, Nov 2017, Orléans, France. ([hal-02362576](#))

. Justine Largillière, Rémy Le Meur, Martine Cadène, Bertrand Castaing, Karine Loth, *et al.* **Première description cinétique du mécanisme d'échange de chaînes de la protéine HU d'*Escherichia coli*.** *30ème colloque Biotechnocentre*, Oct 2017, Seillac, France. ([hal-02361632](#))

. Karine Loth, Françoise Culard, [Justine Largillière](#), Céline Landon, Bertrand Castaing, *et al.* **The MC1-DNA complex solution structure reveals that a monomeric protein can induce a U-turn in archaea.** *NMR: a tool for biology XTh*, Jan 2017, Paris, France. ([hal-02361613](#))



Avez-vous déjà passé une **I-R-M** ?

Eh bien, je suis comme ces médecins qui vous accueillent là-bas...

Regardez : Blouse, Lunettes !

Alors, qu'est-ce qui peut bien nous différencier ?

Je vais vous le dire : c'est la **taille** de nos patients.

Il vous faut imaginer un être très petit, bien plus petit qu'un insecte. Mon patient fait **un-mi-cro-mètre** ! Il est un million de fois plus petit qu'un Homme. Il s'agit d'une **bactérie** que vous connaissez probablement. On la retrouve en grande partie dans notre tube digestif.

C'est **Es-che-ri chia-co-li**.

Ma patiente est comme nous, elle possède des « **organes** » qui lui sont **vitaux**. Parmi ceux-ci il y a l'**A-D-N**, **GARDIEN** de l'information génétique. Mais l'**ADN** n'est pas seul. Afin de protéger ce précieux « **organe** », il existe de nombreux éléments biologiques.

Pour mon projet de thèse, je m'intéresse à l'un de ces éléments en **particulier** : une protéine organisatrice de l'ADN des bactéries.

On l'appelle **H-U**.

Mon objectif est, d'une part de décrire la **structure 3D** de H-U et d'autre part d'étudier sa **dynamique**.

Comment faire cela... ?

Pour commencer voyez ma protéine comme un assemblage de **deux chaînes**, où chaque chaîne est un enchaînement de petites **billes aimantées** qui s'organise dans l'espace.

Ces billes peuvent **se repousser ou s'attirer**.

C'est l'**addition des forces de tous ces petits aimants** qui structure ma protéine. Et ce sont aussi ces forces qui lui donnent sa **dynamique**... soit entre les deux chaînes de ma protéine, soit avec d'autres protéines.

C'est bon ? Vous voyez mieux H-U à présent ?

Maintenant rappelez-vous,

Je travaille avec un super IRM à protéine que l'on appelle **spectromètre à Résonance Magnétique Nucléaire : R-M-N**. J'utilise de **grands** champs magnétiques pour en détecter de plus petits, dépendants de leur environnement.

J'obtiens alors un **spectre R-M-N**. Avec toutes ses taches, on peut le voir comme une carte codée ou encore une empreinte digitale.

Il est **propre** à ma protéine. Et pour le décoder, c'est avant tout une question de **pratique**.

Chaque tache correspond à une **distance** entre deux billes aimantées. On peut voir cette distance comme... un petit élastique.

Et c'est l'**assemblage dans l'espace** de tous ces petits élastiques, qui me permet de décrire la structure 3D de ma protéine.

En parallèle, j'étudie aussi la **dynamique** de H-U.

Si l'on revient à mon spectre RMN avec ses tâches. Je vous ai fait **imaginer** des élastiques pour représenter les distances...

Eh bien... Ce n'est pas par hasard !

Un élastique, contrairement à une baguette ou à un cordon, peut **s'étirer ou se rétracter**. Et cette dynamique peut être observé par RMN grâce à l'étude de la **forme**, ou de la **taille** des tâches sur le spectre par exemple. Ces tâches qui, je vous le rappelle, sont liées à l'environnement de mes billes aimantées.

A la fin de mon étude, je ferai un compte-rendu de toutes mes analyses sur H-U. Cela afin d'établir, avec l'aide de confrères, un **diagnostic général** sur cette protéine, et sur son utilité pour la bonne santé de notre patiente commune : *Escherichia coli*.

Annexe 18 : Texte et illustration de la présentation au concours Ma thèse en 180 secondes, édition 2019.

Justine LARGILLIERE

Architecture moléculaire et dynamique de protéines histone-like de bactérie et d'archée

HU est une protéine bactérienne qui est impliquée dans de nombreuses fonctions liées à l'ADN. Elle est présente sous forme de trois dimères chez *E. coli* (deux homodimères et un hétérodimère). Lorsque les deux homodimères sont mélangés *in vitro*, ils échangent leurs chaînes pour former l'hétérodimère. Mon travail a consisté à caractériser, structuralement et cinétiquement, ce mécanisme d'échange qui peut être décrit comme une réaction d'ordre 2 se déroulant en 3 étapes : d'une conformation native (N_2) de chaque homodimère à une conformation intermédiaire (I_2 , partiellement dissociée et déstructurée), puis la formation d'un tétramère transitoire (étape limitante) qui se dissocie finalement en deux hétérodimères. Les résidus considérés comme étant les déterminants structuraux permettant la transition entre N_2 et I_2 ont pu être déterminés. Ces résidus, enfouis dans N_2 , forment un patch hydrophobe sur la surface de I_2 . Ce patch peut être impliqué dans la reconnaissance des chaînes de HU et permettrait la formation du tétramère.

MC1 participe à l'organisation du génome de plusieurs archées, à la transcription de l'ADN et à la division cellulaire par des mécanismes inconnus. Nous présentons la structure d'un complexe formé par MC1 avec un ADN de 15 paires de bases. Alors que la protéine a besoin d'adapter sa conformation légèrement, l'ADN subit une courbure dramatique et une torsion impressionnante. Une telle conformation en V du complexe et un modèle structural de MC1 avec un ADN plus long nous ont amené à proposer un nouveau mode de liaison de la protéine en tant qu'« enrouleur ». Des expériences de diffraction RX et de SAXS ont été réalisées sur ce complexe. Malheureusement, la structure n'a pas pu être résolue en raison du manque de données de diffraction et les données SAXS ont invalidé le modèle. Ces résultats confirment que MC1 est une protéine atypique, qui stabilise de multiples conformations en V de l'ADN de manière flexible et dynamique.

Mots clés : structure, dynamique, échange, protéine interagissant avec l'ADN, RMN, ADN

Molecular and dynamic architecture of Bacteria and Archaea histone-like protein

HU is an essential bacterial protein that is involved in many functions related to DNA. It is present as three dimers in *E.coli* (two homodimers and one heterodimer). If the two homodimers are mixed *in vitro*, they exchange their chains to spontaneously form the heterodimer. My work was to characterize, structurally and kinetically, this exchange mechanism that can be described as a second order reaction of three successive steps: from a native conformation of each homodimer into an intermediate homodimer conformation (partially unfolded and dissociated), followed by the formation of a transient tetramer (limiting step) which finally dissociates into two heterodimers. The key residues allowing the protein to switch from the native to the intermediate state has been determined. Whereas, there are buried in the native conformation, they are forming a hydrophobic patch at the surface of the intermediate one. This patch could mediate the association of the intermediate conformation in order to form the tetramer.

MC1 participates in the genome organization of several archaea and in DNA transcription and cellular division through unknown mechanisms. We discuss the solution structure of a complex formed by MC1 with a strongly distorted 15 base pairs DNA. While the protein just needs to adapt slightly its conformation, the DNA undergoes a dramatic curvature and an impressive torsion. Such a V-turn conformation of the complex lead us to propose a new binding mode for the protein as a wrapper and a structural model of MC1 with a longer DNA. XR diffraction and SAXS experiments were then carried out on this new complex. Unfortunately, the structure could not be solved due to the lack of diffraction data and the SAXS data invalidated the model. These results confirm that MC1 is an atypical protein, which stabilizes multiple V-turn conformations of the DNA in a flexible and dynamic manner.

Keywords: structure, dynamic, exchange, DNA binding protein, NMR, DNA



**Centre de Biophysique Moléculaire
CNRS UPR 4301
Avenue de la recherche Scientifique - ORLEANS**

