



**HAL**  
open science

# La protéine MC1 d'archaeabactérie : reconnaissance de séquences particulières

Guillaume de Vuyst

► **To cite this version:**

Guillaume de Vuyst. La protéine MC1 d'archaeabactérie : reconnaissance de séquences particulières. Biochimie [q-bio.BM]. Université d'Orléans, 2004. Français. NNT : . tel-00008337v2

**HAL Id: tel-00008337**

**<https://theses.hal.science/tel-00008337v2>**

Submitted on 28 Jul 2005

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

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



**THESE**  
**PRESENTEE**  
**A L'UNIVERSITE D'ORLEANS**  
**POUR OBTENIR LE GRADE DE**  
**DOCTEUR DE L'UNIVERSITE D'ORLEANS**

Discipline :

*Aspects Moléculaires et Cellulaires de la Biologie*

PAR

**Guillaume DE VUYST**

**LA PROTEINE MC1**  
**D'ARCHAEABACTERIE : RECONNAISSANCE**  
**DE SEQUENCES PARTICULIERES.**

Soutenue le : 1<sup>er</sup> avril 2004

**MEMBRES DU JURY :**

- |  |                            |
|--|----------------------------|
| - Pr. Francis DELMOTTE, LBLGC Université d'Orléans | <b>Président</b>           |
| - Dr. Brigitte HARTMANN, CNRS/IBPC Paris V         | <b>Rapporteur</b>          |
| - Dr. Eric LE CAM, CNRS/IGR Villejuif              | <b>Rapporteur</b>          |
| - Dr. Françoise CULARD, CNRS/CBM d'Orléans         | <b>Directrice de thèse</b> |



**THESE**  
**PRESENTEE**  
**A L'UNIVERSITE D'ORLEANS**  
**POUR OBTENIR LE GRADE DE**  
**DOCTEUR DE L'UNIVERSITE D'ORLEANS**

Discipline :

*Aspects Moléculaires et Cellulaires de la Biologie*

PAR

**Guillaume DE VUYST**

**LA PROTEINE MC1**  
**D'ARCHAEABACTERIE : RECONNAISSANCE**  
**DE SEQUENCES PARTICULIERES.**

Soutenue le : 1<sup>er</sup> avril 2004

**MEMBRES DU JURY :**

- |  |                            |
|--|----------------------------|
| - Pr. Francis DELMOTTE, LBLGC Université d'Orléans | <b>Président</b>           |
| - Dr. Brigitte HARTMANN, CNRS/IBPC Paris V         | <b>Rapporteur</b>          |
| - Dr. Eric LE CAM, CNRS/IGR Villejuif              | <b>Rapporteur</b>          |
| - Dr. Françoise CULARD, CNRS/CBM d'Orléans         | <b>Directrice de thèse</b> |

Je remercie messieurs Paul VIGNY et Jean Claude BELŒIL pour m'avoir accueilli au Centre de Biophysique Moléculaire.

Je remercie également Mme Françoise CULARD pour son accueil au sein de son équipe de recherche et pour m'avoir guidé pendant ces années de thèse.

Je tiens à exprimer ma reconnaissance aux Docteurs Brigitte HARTMANN et Eric LE CAM pour avoir accepté d'examiner et de juger ce travail. Je remercie également le professeur Francis DELMOTTE pour avoir accepté d'être le président de mon jury de thèse.

Je tiens aussi à remercier les Docteurs Michel CHARLIER et M Daniel GENEST avec qui j'ai effectué des collaborations très enrichissantes.

Je remercie chaleureusement M Alain GERVAIS pour sa précieuse aide technique ainsi que toutes les personnes avec qui j'ai interagi pendant ces années de thèse. J'ai particulièrement apprécié leur disponibilité, leur gentillesse et leur bonne humeur.

Je n'oublie pas ma famille et mes amis qui ont fait preuve de tolérance à mon égard et qui m'ont encouragés dans les moments difficiles.

Ce travail à été soutenu par une bourse du Ministère de la Recherche et de l'Enseignement.

# TABLE DES MATIERES

ABREVIATIONS .....	1
INTRODUCTION.....	4
LES ARCHAEABACTERIES .....	6
1) Généralités : .....	7
1.1) Le génome : .....	8
1.2) La réplication : .....	9
1.3) La transcription : .....	12
2) Les protéines architecturales des archaeobactéries : .....	14
2.1) Protéines architecturales à motif histone : .....	15
2.2) Protéines architecturales sans motif histone : .....	19
2.2.1) Chez les Sulfolobales : .....	19
2.2.2) Protéines de la famille Alba : .....	22
2.2.3) HTa : .....	24
2.2.4) MC1 : .....	25
PROTEINES ARCHITECTURALES DES EUBACTERIES .....	31
1) Introduction : le nucléoïde : .....	32
2) La protéine FIS : .....	34
3) La protéine HU : .....	36
4) La protéine IHF : .....	39
5) La protéine H-NS : .....	42

OBJECTIF DU TRAVAIL DE THESE .....	46
MATERIELS ET METHODES .....	48
1) Préparation de la protéine MC1 : .....	49
2) Isolement des séquences par la méthode SELEX : .....	49
2.1) Purification des oligonucléotides : .....	49
2.2) Synthèse de la librairie aléatoire d'ADN double brin : .....	50
2.3) Sélection et purification de fragments de haute affinité pour la protéine MC1 : .....	50
2.4) Amplification par PCR des ADN de forte affinité : .....	51
2.5) Marquage radioactif : .....	52
3) Caractérisation des séquences issues du SELEX : .....	52
3.1) Clonage des séquences issues du SELEX : .....	52
3.2) Mesure de la vitesse de dissociation des complexes : .....	53
3.3) Expériences de compétition directe : .....	54
3.4) Mesure de la constante de dissociation des complexes MC1/séquences préférentielles : .....	54
4) Etude du mode de fixation de la protéine : .....	54
4.1) Préparation de l'ADN double brin de 59 pb radiomarké sur un seul brin : .....	54
4.2) Séquençage des purines : .....	55
4.3) Empreintes au DMS : .....	55
4.4) Interférences au DMS : .....	56
4.5) Empreintes aux radicaux OH <sup>•</sup> : .....	56
4.6) Empreintes au KMnO <sub>4</sub> : .....	57
4.7) Traitement des gels d'empreinte après migration : .....	57
4.8) Stoechiométrie de l'interaction MC1/séquence consensus : .....	57

5) Etude de l'effet des rayonnements ionisants sur les fixations préférentielles de MC1 :.....	57
5.1) Préparation des fragments ADN : .....	57
5.2) Irradiation de la protéine MC1 :.....	58
5.3) Expériences de compétition pour la fixation sur des ADN linéaires : .....	58
5.4) Expériences de compétition pour la fixation sur des ADN circulaires :.....	59
6) Récapitulatif des différents fragments utilisés : .....	60
<b>RESULTATS .....</b>	<b>61</b>
<b>ISOLEMENT ET CARACTERISATION DES SEQUENCES.....</b>	<b>62</b>
1) La méthode SELEX : .....	63
1.1) Généralités : .....	63
1.2) Facteurs influençant la méthode SELEX : .....	66
2) Sélection de sites préférentiels pour la protéine MC1 : .....	68
2.1) Choix des conditions expérimentales : .....	68
2.2) Déroulement des cycles de sélection : .....	69
2.3) Mise en évidence de la sélection de sites préférentiels :.....	70
3) Analyse des séquences sélectionnées :.....	75
3.1) Détermination des séquences : .....	75
3.2) Deuxième expérience de sélection : .....	76
3.3) Mise en évidence d'une séquence consensus :.....	78
3.4) Comparaison de l'affinité de MC1 pour chacune des séquences issues du SELEX : .....	79
3.5) Détermination de l'affinité de l'interaction MC1/séquence consensus : .....	82
4) Discussion : .....	85
<b>ETUDE DU MODE DE FIXATION DE LA PROTEINE .....</b>	<b>94</b>

1) Principe de la méthode : .....	95
2) Empreintes et interférences au DMS : .....	95
2.1) Généralités : .....	95
2.2) Les empreintes : .....	96
2.3) Les interférences : .....	99
3) Empreintes aux radicaux hydroxyle : .....	100
3.1) Généralités : .....	100
3.2) Résultats : .....	101
3.3) Stœchiométrie de l'interaction MC1/séquence consensus : .....	105
4) Discussion : .....	106

## ETUDE DE L'EFFET DES RAYONNEMENTS IONISANTS SUR LES FIXATIONS PREFERENTIELLES DE MC1 ..... 109

1) Introduction : .....	110
2) Résultats : .....	112
2.1) Effet des radiations $\gamma$ sur la reconnaissance MC1/séquence préférentielle : .....	113
2.2) Effet des radiations $\gamma$ sur la reconnaissance préférentielle MC1/cercles : ..	116
2.3) Effets des rayonnements ionisants sur la structure du complexe MC1/ADN : .....	117
3) Discussion : .....	118

## CONCLUSIONS ET PERSPECTIVES ..... 120

## BIBLIOGRAPHIE ..... 124

## ANNEXE ..... 146



## TABLE DES ILLUSTRATIONS

Figure 1 : répartition des organismes dans le monde du vivant ainsi que dans les quatre domaines des archaeobactéries.....	7
Tableau 1 : répartition des différentes ADN polymérase chez les archaeobactéries, les eubactéries et les eucaryotes. ....	10
Figure 2 : modèles de l'initiation de la réplication dans les trois domaines du vivant. ....	11
Figure 3 : schéma montrant deux exemples de répression de la transcription.....	14
Figure 4 : répartition des différentes protéines architecturales dans le règne des archaeobactéries. ....	15
Figure 5 : structure RMN d'un monomère de HMf. ....	16
Figure 6 : modèle montrant la fixation d'un dimère (HMfA) <sub>2</sub> sur l'ADN.....	17
Figure 7 : structure tridimensionnelle de deux complexes Sac 7d/ADN.....	20
Figure 8 : structure de la protéine Alba.....	23
Figure 9 : alignement des séquences des protéines MC1 de différents organismes. ....	26
Figure 10 : structure RMN de la protéine MC1. ....	27
Figure 11 : visualisation des minicercles d'ADN nus ou complexés avec MC1.....	29
Figure 12 : visualisation des complexes MC1/site I. ....	29
Tableau 2 : niveau d'expression de certaines protéines de type histone au cours des différentes phases de culture de bactéries <i>E. coli</i> .....	33
Figure 13 : structure d'un dimère de FIS. ....	35
Figure 14 : modèle de la fixation de HU sur des structures particulières. ....	37
Figure 15 : structure RMN d'un dimère de HU ( <i>Anabaena</i> ) fixé à l'ADN. ....	38
Figure 16 : structure cristallographique d'un dimère de IHF fixé à l'ADN.....	41
Figure 17 : structure de la protéine H-NS. ....	45

Figure 18 : les différentes parties des fragments SELEX. ....	64
Figure 19 : récapitulatif des différentes étapes d'une expérience SELEX.....	65
Tableau 3 : récapitulatif du taux de sélection obtenu pour chaque cycle SELEX. ....	70
Figure 20 : dissociation des complexes MC1/fragments du cycle 0 ou du cycle 10.....	71
Figure 21 : dissociation des complexes formés avec les fragments du cycle 0 et du cycle 10.	72
Tableau 4 : temps de demi-vie et vitesse de dissociation des complexes obtenus avec les ADN des cycles 0 et 10.....	73
Figure 22 : évolution du temps de demi-vie des complexes MC1/fragments SELEX au cours des cycles SELEX. ....	73
Figure 23 : mesure des vitesses de dissociation des complexes MC1/aptamères et MC1/site I. .....	74
Figure 24 : résultat du séquençage des clones issus du cycle 10. ....	75
Figure 25 : compétition directe entre les fragments S2-6 et S2-0 pour la fixation de la protéine MC1.....	77
Figure 26 : fréquence des bases dans les 13 fragments séquencés.....	78
Tableau 5 : cinétiques de dissociation des complexes formés entre la protéine MC1 et chacune des séquences issues du SELEX. ....	80
Figure 27 : compétition directe entre la séquence consensus et une séquence aléatoire pour la fixation de la protéine MC1. ....	82
Figure 28 : analyse par gel retard de la fixation de la protéine MC1 sur un fragment contenant la séquence consensus (26 pb préférentiel). ....	83
Figure 29 : mesure des vitesses de dissociation des complexes MC1/site consensus et MC1/séquence aléatoire. ....	84
Tableau 6 : temps de demi-vie et vitesses de dissociation des complexes obtenus avec les ADN du cycle 0 ou la séquence consensus. ....	85
Figure 30 : alignement de la séquence du site I et de la séquence C20.....	86
Figure 31 : recherche des séquences préférentielles de fixation dans les génomes d'archaeobactéries. ....	87

Tableau 7 : localisation des séquences trouvées par « blastn » dans les génomes de <i>M. acetivorans</i> C2A et <i>M. mazei</i> Goe1. ....	88
Figure 32 : duplex utilisé pour la simulation de dynamique moléculaire. ....	91
Figure 33 : identification des angles dièdres du squelette phosphodiester. ....	92
Figure 34 : action du DMS sur l'ADN. ....	96
Figure 35 : empreinte au DMS. ....	97
Figure 36 : analyse quantitative de l'expérience d'empreinte au DMS. ....	98
Figure 37 : interférence au DMS. ....	100
Figure 38 : empreintes aux radicaux OH <sup>•</sup> . ....	102
Figure 39 : densitogrammes des gels présentés Figure 38. ....	103
Figure 40 : analyse quantitative des données de la Figure 39. ....	104
Figure 41 : titration d'un fragment de 26 pb contenant la séquence consensus par des concentrations croissantes de protéine MC1. ....	106
Figure 42 : représentation plane résumant les résultats obtenus par les méthodes d'empreintes. ....	107
Figure 43 : radiolyse de l'eau par les rayonnements ionisants (rayons X, γ...). ....	110
Tableau 8 : principaux produits de réaction entre un radical OH <sup>•</sup> et certains acides aminés. ....	111
Figure 44 : inactivation de la protéine MC1 par les rayons γ. ....	113
Figure 45 : effet des rayons γ sur la reconnaissance des fragments issus du SELEX par MC1. ....	114
Figure 46 : effet des rayons γ sur la reconnaissance du site naturel (site I) par MC1. ....	115
Figure 47 : effet des rayons γ sur la reconnaissance des ADN circulaire par MC1. ....	116
Figure 48 : dessin représentant des complexes entre la protéine MC1 et des ADN de différentes longueurs possédant différents angles de courbures. ....	117

## **ABBREVIATIONS**

- Alba : Acetylation Lower Binding Affinity
- BRE : B Recognition Element
- Csp : Cold Shock Protein
- DMS : diméthylsulfate
- DTT : dithiotréitol
- FEN-1 : Flap Endonuclease 1
- FIS : Factor for Inversion Stimulation
- H2A, H2B, H3, H4 : histones eucaryotiques
- HMfA, HMfB : Histone de *Methanococcus fervidus* A ou B
- HMG : High Mobility Group protéin
- HMk ( ou MkaH) : histone de *Methanopyrus kandleri*
- HMt : histone de *Methanobacterium thermoautotrophicum*
- H-NS :Histone like Nucleoid Structurating protein
- Hta : histone de *Thermoplasma acidophilum*
- HU : Heat Unstable protein ; protéine de type histone chez les eubactéries
- IHF : Integration Host Factor
- IPTG : isopropylthio  $\beta$ -D galactoside
- KMnO<sub>4</sub> : permanganate de potassium
- LB : milieu de culture Luria-Bertani
- LEF-1 : Lymphoid Enhancer binding Factor 1
- Lrp : Leucine Responsive regulatory Protein
- Lrs 14 : homologue de Lrp
- MC1 : Metanogen Chromosomal protein 1
- MCM : Mini Chromosome Maintenance
- MDR1 : Metal Dependent Repressor 1
- OBP : Origin Binding Protein

Orc : Origin recognition complex

pb : paire de bases

PCNA : Proliferating Cell Nuclear Antigen

PCR : Polymerase Chain Reaction

PEG : polyéthylène glycol

RFC : Replication Factor C

R : purines

RMN : résonance magnétique nucléaire

RT-PCR : Retro Transcription Polymerase Chain Reaction

SAB : Sérum Albumine Bovine

SAP : Shrimp Alkaline Phosphatase

Sac 7d : protéine de la famille Sul 7d chez *Sulfolobus acidocaldarius*

SELEX : Systematic Evolution of Ligands by EXponential enrichment

SRY : sex determining factor

Sso 7d : protéine de la famille Sul 7d chez *Sulfolobus solfataricus*

StpA : Suppressor of *td* mutant phenotype A (protéine paralogue de H-NS)

Sul 7d : protéine de 7 kDa des sulfolobales

TBE : Tris-borate, 45 mM ; EDTA, 1 mM

TBP : TATA Binding Factor

TE : Tris-HCl, 100 mM ; EDTA, 10 mM

TF... : Transcription Factor...

TFB : Transcription Factor B (paralogue de TFIIB chez les archaeobactéries)

Y : pyrimidines

## **INTRODUCTION**

Les archaeobactéries forment avec les eubactéries et les eucaryotes l'un des trois domaines du vivant. Les archaeobactéries sont des organismes ubiquitaires très abondants qui occupent des niches écologiques variées : glaces, sols, océans.

La place de ces organismes procaryotes dans l'évolution reste encore controversée : pour certains auteurs les archaeobactéries sont des organismes intermédiaires entre les eubactéries et les eucaryotes, alors que pour d'autres les trois domaines du vivant se sont différenciés à partir d'un même ancêtre commun.

Les archaeobactéries ont su s'adapter à des conditions de vie extrêmes (température, pression, pH) et possèdent des protéines qui sont potentiellement intéressantes pour les biotechnologies (l'ADN polymérase Pfu par exemple).

Bien que certains aspects de la régulation des processus biologiques des archaeobactéries soient semblables à ceux trouvés chez les eubactéries ou les eucaryotes, ces organismes possèdent aussi des propriétés propres. Cette observation est particulièrement vraie dans le cas du métabolisme de leur ADN.

Un certain nombre de protéines très abondantes servant à la compaction de l'ADN ont été isolées chez les archaeobactéries. Différentes études ont été réalisées *in vitro* afin de caractériser leurs interactions avec l'ADN (recherche de séquences préférentielles de fixation ; changement de conformation de l'ADN). La structure de ces protéines architecturales ainsi que celle des complexes protéine/ADN a également été déterminée. Certaines de ces protéines possèdent des homologues chez les eubactéries (protéines de type HU) ou chez les eucaryotes (protéines type histones), d'autres au contraire sont spécifiques des archaeobactéries (MC1, Sul 7d). L'analyse des différents génomes d'archaeobactéries actuellement séquencés a permis de montrer qu'un même organisme peut comporter des protéines architecturales spécifiques des archaeobactéries et des protéines architecturales trouvées dans les autres domaines du vivant : il n'existe donc pas un mécanisme unique et conservé de compaction du génome de ces organismes.



# **LES ARCHAEBACTERIES**

## 1) Généralités :

Depuis les travaux de Woese et collaborateurs basés sur l'étude des séquences des ARN ribosomiques, le monde du vivant est divisé en trois domaines (Woese et coll., 1990) : les eucaryotes, les eubactéries et les archaeobactéries. Le domaine des archaeobactéries regroupe des organismes très divers aux niches écologiques variées et souvent très particulières (température extrême, forte concentration en sel, pH acide, absence d'oxygène...). D'un point de vue phylogénique, le domaine des archaeobactéries est actuellement divisé en quatre règnes (Woese et coll., 1990 ; Barns et coll., 1996 ; Huber et coll., 2002 ; Figure 1) : 1) les **crenarchaeas** qui regroupent les organismes thermophiles soufre dépendant (Sulfolobales, Pyrodictiales, Thermoprotéales) ; 2) les **euryarchaeas** comprenant les Thermococcales, les Methanococcales, les Méthanobactéries et les halophiles extrêmes ; 3) les **korarchaeas** vivant dans les sédiments des sources chaudes ; 4) les **nanoarchaeas** qui regroupent des organismes de petite taille vivant en symbiose avec certaines crenarchaeas des fumées noires sous-marines.

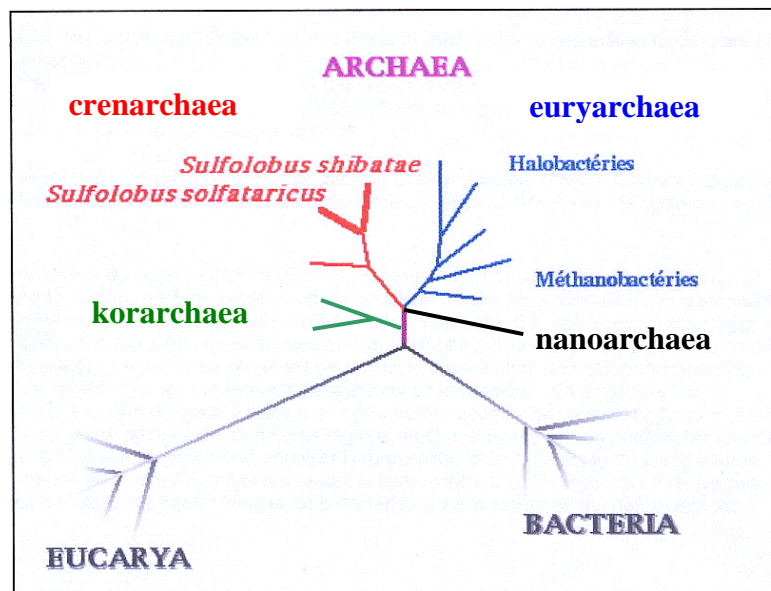


Figure 1 : répartition des organismes dans le monde du vivant ainsi que dans les quatre domaines des archaeobactéries.

Les archaeobactéries ont une morphologie proche de celle des eubactéries (bactéries vraies) : absence de noyau, d'organelle et de cytosquelette, taille du génome. Pourtant, l'étude du fonctionnement des archaeobactéries a rapidement montré que ces organismes possèdent des similarités avec les cellules eucaryotes (couplage entre la réplication et la division cellulaire ; transcription ; réplication).

### 1.1) Le génome :

Hormis pour *Nanoarchaeum equitans* qui a un génome de  $0.5 \times 10^6$  pb, la taille du génome des archaeobactéries varie entre  $1.6 \times 10^6$  pb (*Thermoplasma*) et  $4.6 \times 10^6$  pb (*Methanosarcina*). La proportion de bases GC est globalement de 42 % (22 % au niveau des régions intergéniques). Actuellement, 17 génomes d'archaeobactéries sont totalement séquencés et 2 le sont partiellement.

A partir de la séquence complète du génome de *Methanococcus jannaschii*, Bult et collaborateurs ont montré que la plupart des gènes impliqués dans le métabolisme ressemblent à ceux des eubactéries alors que les gènes impliqués dans les différents processus biologiques tels que la transcription, la réplication et la traduction ressemblent à ceux des eucaryotes (Bult et coll., 1996). D'autre part, si 44 % des gènes ressemblent à ceux des eubactéries et des eucaryotes, 56 % sont totalement différents de tout autre gène déjà décrit.

L'analyse des génomes a permis de montrer que dans tous les organismes séquencés (sauf *Halobacterium*), des séquences d'environ 30 pb sont répétées sous forme de « clusters » (Karlín et coll., 2002). Suivant les organismes, le nombre de « clusters » varie entre 2 et 9 et le nombre de répétitions par cluster entre 18 et 124. D'autre part, la présence de séquences palindromiques de 4 à 6 pb a été mise en évidence. Il est donc possible que, comme les eubactéries, les archaeobactéries possèdent un système de restriction.

Les travaux de Takayanagi et collaborateurs ont révélé par microscopie électronique l'existence de particules ressemblant aux nucléosomes eucaryotes au niveau de l'ADN génomique ou plasmidique de certaines euryarchaeas (Takayanagi et coll., 1992). Ces particules sont cependant plus petites que celles trouvées chez les eucaryotes (9.4 nm au lieu de 11 nm).

Actuellement, les études réalisées sur les crenarchaeas n'ont pas permis d'isoler de protéine provoquant un enroulement de l'ADN semblable à celui observé avec les histones des eucaryotes. Les travaux de microscopie en contraste de phase menés en 1997 par Poplawski et Bernander ont montré que le degré d'organisation du chromosome de *Sulfolobus solfataricus* et *S. acidocaldarius* est plus grand au cours de la phase de croissance exponentielle qu'au cours de la phase stationnaire.

## 1.2) La réplication :

Afin de comprendre le mode de réplication des archaeobactéries, il est nécessaire de déterminer d'une part la ou les origine(s) de réplication et d'autre part la machinerie de réplication. D'après les différents travaux réalisés dans ce domaine, on peut dire que les archaeobactéries utilisent un mode de réplication bactérien avec une machinerie de type eucaryote.

En 2000, l'équipe de Forterre a montré par des analyses statistiques des séquences que, chez *Pyrococcus abyssi*, il existe une origine de réplication unique. Les études ont montré que cette origine permet une réplication bidirectionnelle (Myllykallio et coll., 2000). Par ailleurs, à proximité de cette origine de réplication, on trouve les gènes codants pour les protéines fixant l'origine de réplication (OBP) et pour l'ADN polymérase. Les auteurs ont donc conclu que les archaeobactéries possèdent une origine de réplication similaire à celle des eubactéries. Cette origine de réplication des archaeobactéries a été dénommée OriC.

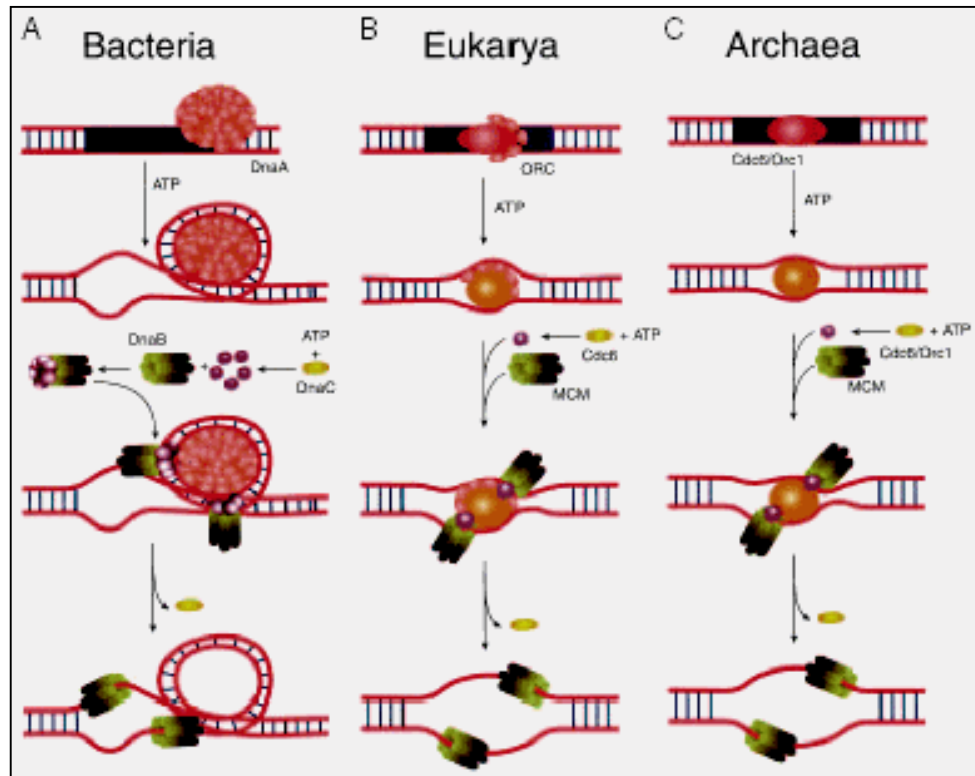
De nombreux travaux ont montré que la machinerie de réplication des archaeobactéries est très proche de celle des cellules eucaryotes (pour revue Giraldo, 2003 ; MacNeill, 2001). Lors de l'initiation, l'origine de réplication est reconnue par le complexe protéique Orc1/Cdc6. La protéine Orc1 (origin recognition complex 1) est l'orthologue des protéines Orc1-6 de la levure. La protéine Cdc6 possède elle aussi des orthologues dans les cellules eucaryotes. La partie Cdc6 du complexe Orc1/Cdc6 permet le recrutement de l'hélicase MCM (Mini Chromosome Maintenance) qui est composée de une à quatre sous-unité(s) (6 unités chez les eucaryotes). L'association de ces trois protéines est appelée complexe pré-répliatif. La Figure 2 (p 11) représente de façon schématique l'initiation de la réplication dans les trois domaines du vivant : bactéries (partie A) ; eucaryotes (partie B) ; archaeobactéries (partie C).

Le complexe pré-répliatif est ensuite rejoint par les protéines permettant l'élongation : primase, ADN polymérase, facteurs accessoires (RFC, PCNA...). La primase va permettre la synthèse d'une amorce sur laquelle va se fixer l'ADN polymérase. La nature exacte de ces ADN polymérases est mal connue car elle dépend en partie de l'organisme étudié. En effet, chez les euryarchaeas, on trouve des polymérases monomériques de la famille B (polBI) et des polymérases dimériques de la famille D (polD). En revanche, les crenarchaeas ne codent pas pour les polymérases de la famille D mais pour plusieurs types de polymérases de la famille B. Une partie de ces polymérases B est semblable à celles trouvées chez les euryarchaeas (polBI), l'autre partie est spécifique des crenarchaeas (polBII). Au sein de la famille de polymérases B, on trouve les polymérases eucaryotes  $\alpha$ ,  $\delta$ ,  $\varepsilon$ ,  $\zeta$ ,  $V$  et la polymérase eubactérienne polIII (Tableau 1). La famille D de polymérases ne contient que la polymérase D des euryarchaeas. Pour permettre le bon déroulement de la réplication, la polymérase a besoin de différents facteurs accessoires : PCNA (Proliferating Cell Nuclear Antigen), RFC (Replication Factor C), FEN-1 (Flap ENdonuclease 1), RNase H. Les facteurs PCNA et RFC rendent la polymérase processive, FEN-1 et la RNase H interviennent dans la suppression de l'amorce synthétisée par la primase. Tous ces facteurs accessoires sont des orthologues de ceux trouvés dans les cellules eucaryotes.

**Tableau 1 : répartition des différentes ADN polymérases chez les archaeobactéries, les eubactéries et les eucaryotes.**

Domain	Family					
	A	B	C	D	X	?
Archaea						
Euryarchaeota		Pol BI		Pol D		
Crenarchaeota		Pol BI, Pol BII				
Bacteria	Pol I	Pol II	Pol III			
Eukarya	$\gamma$	$\alpha$ , $\delta$ , $\varepsilon$ , $\zeta$ , $V$			$\beta$	$\eta^*$

\* This is a DNA polymerase yet to be assigned a family.



**Figure 2 : modèles de l'initiation de la réplication dans les trois domaines du vivant.**

A : L'initiation de la réplication chez les bactéries commence par la reconnaissance et la fixation de l'origine de réplication par la protéine DnaA. Cet événement provoque une dénaturation locale de l'ADN. Un hexamère  $[DnaC : ATP]_6$  va ensuite interagir avec la protéine DnaB. Le complexe DnaB/ $[DnaC : ATP]_6$  va se fixer sur la protéine DnaA. L'hydrolyse de l'ATP par DnaC entraîne le relargage de l'hélicase DnaA. B : l'initiation de la réplication chez les eucaryotes débute par la fixation sur l'origine de réplication de Orc. La dénaturation locale de l'ADN se produit grâce à la fixation de Orc seul ou associé à d'autres facteurs. La protéine Cdc6 liée à de l'ATP et à l'hélicase MCM s'associe à Orc au niveau de l'origine de réplication. L'hydrolyse de l'ATP par Cdc6 permet un changement de conformation qui provoque la libération de l'hélicase. Il est probable que d'autres facteurs interviennent dans ce processus. C : initiation de la réplication chez les archaeobactéries. Le complexe Cdc6/Orc1 reconnaît et fixe l'origine de réplication. Cette fixation provoque la dénaturation locale de l'ADN. L'hélicase MCM et un complexe Cdc6 lié à de l'ATP s'associent au complexe Cdc6/Orc1 fixé à l'origine. L'hydrolyse de l'ATP par le complexe Cdc6/Orc1 permet la libération de l'hélicase. (d'après Kelman et Kelman, 2003).

Il est important de noter que, même si de façon générale la machinerie de réplication des archaeobactéries est proche de celle des eucaryotes, tout n'est pas encore compris. En effet, aucun gène codant pour les protéines Orc1 et Cdc6 n'a pu être isolé chez *Methanococcus jannaschii*. Comme il est peut probable que la divergence de cet organisme soit suffisamment grande pour empêcher l'identification de ces protéines par des méthodes classiques de recherche dans les bases de données, il est envisageable que *M. jannaschii* ait

développé une autre stratégie de reconnaissance de l'origine de réplication. Par ailleurs, le groupe de Forterre n'a pas pu définir l'origine de réplication dans un certain nombre des archaeobactéries étudiées (notamment *M. jannaschii* ; Myllykallio et coll., 2000). Les auteurs proposent deux explications : 1) la méthode utilisée n'est pas suffisamment précise ; 2) ces organismes possèdent plusieurs origines de réplication (comme les cellules eucaryotes).

### 1.3) La transcription :

La transcription des archaeobactéries est une mosaïque entre les caractéristiques eucaryotes et bactériennes : mécanisme d'initiation voisin de celui des cellules eucaryotes, facteurs de régulations semblables à ceux trouvés dans les eubactéries.

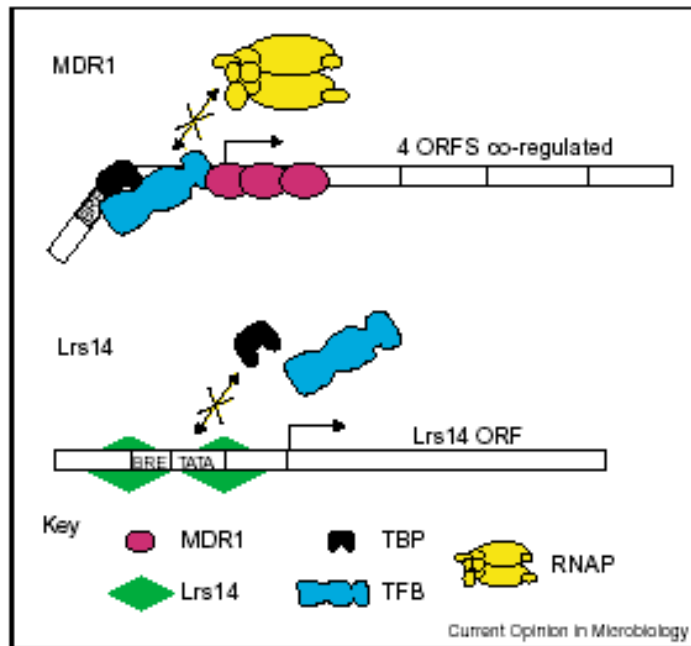
Il a été démontré que l'ARN polymérase des archaeobactéries est unique (comme celle des eubactéries) et composée d'un grand nombre de sous-unités (Langer et coll., 1995). Par ailleurs, des études immunologiques ont mis en évidence une réaction croisée entre des sérums dirigés contre plusieurs sous-unités des ARN polymérases eucaryotes (en particulier l'ARN polymérase II) et l'ARN polymérase des archaeobactéries (Schnabel et coll., 1983). Ainsi, on peut dire que l'ARN polymérase des archaeobactéries est plus proche de celle des eucaryotes que de celle des eubactéries. Cependant, contrairement à son homologue eucaryote, le domaine C terminal de l'ARN polymérase des archaeobactéries n'est pas impliqué dans la régulation de la transcription.

L'étude de différents génomes archaeobactériens a montré que ces organismes comportent des séquences de type boîte TATA (environ 30 pb en amont du site d'initiation) et BRE (élément de reconnaissance du facteur B). Des recherches dans les banques de données ont donc été menées afin de déterminer si les archaeobactéries comportent des protéines homologues à la TBP (TATA Binding Protein) et au facteur de transcription TFIIB eucaryote. Ces recherches ont révélé qu'une protéine homologue à la TBP est bien présente chez les archaeobactéries. La TBP des archaeobactéries, tout comme son homologue eucaryote, possède deux répétitions directes incomplètes à son extrémité C terminale. De même, TFB (homologue de TFIIB) possède un domaine C terminal organisé de la même façon que celui de TFIIB : doigt de zinc et motif hélice-tour-hélice permettant la fixation à l'ADN.

Pour former complètement leur complexe de préinitialisation, les cellules eucaryotes ont besoin de l'intervention des facteurs de transcription TFIIA, F, E, H. Des facteurs de transcription homologues ont été recherchés sans succès dans les banques de données contenant les génomes d'archaeobactéries (Bult et coll., 1996 ; Klenk et coll., 1997 ; Smith et coll., 1997 ; Kawarabayasi et coll., 1998). Ce résultat est cohérent avec plusieurs observations. Tout d'abord, la transcription des archaeobactéries peut être initialisée *in vitro* par la formation d'un complexe ternaire TBP-TFB-ARN polymérase (Hausner et coll., 1996). De plus, dans les cellules eucaryotes, les facteurs de transcription TFIIH et TFIIF servent à réguler l'activité de l'ARN polymérase en phosphorylant (TFIIH) ou en déphosphorylant (TFIIF) le domaine C terminal de la polymérase. Le domaine C terminal de l'ARN polymérase des archaeobactéries n'étant pas impliqué dans les processus de régulation de la transcription, les facteurs TFIIH et TFIIF ne sont pas nécessaires.

En 1999, Kyrpides et Ouzounis ont cherché à identifier les facteurs de régulation de la transcription chez les archaeobactéries. Pour cela, ils ont recherché dans le génome de quatre archaeobactéries (*M. jannaschii*, *A. fulgidus*, *M. thermoautotrophicum* et *P. horikoshii*) des protéines homologues aux facteurs de régulations trouvés chez les eubactéries et les eucaryotes. Les résultats ont montré que les génomes étudiés possèdent 168 facteurs de régulation homologues à ceux trouvés chez les eubactéries, 51 facteurs homologues à ceux des eucaryotes et 61 facteurs présents dans les deux règnes (Kyrpides et Ouzounis, 1999). Ces résultats montrent que la régulation de la transcription des archaeobactéries est plus proche du système bactérien que du système eucaryote. Parmi les facteurs de régulation bactériens qui sont présents dans les archaeobactéries, on peut citer Nus A et Nus G. Ces deux facteurs ont un rôle anti-terminateur dans les bactéries. De nombreux travaux montrent que les répresseurs de la transcription peuvent intervenir de différentes manières (pour revue, Reeve, 2003 ; Figure 3) : 1) compétition avec la TBP pour la fixation sur la boîte TATA ; 2) compétition avec TFB pour la fixation sur son site (BRE) ; 3) compétition avec l'ARN polymérase pour la fixation sur le site d'initiation de la transcription. Un même répresseur de transcription peut combiner plusieurs des stratégies citées précédemment.





**Figure 3 : schéma montrant deux exemples de répression de la transcription.**

En haut : le facteur MDR1 (en magenta) se fixe sur trois sites opératoires en aval du promoteur empêchant ainsi le recrutement de l'ARN polymérase (en jaune) par les facteurs TBP et TFB (respectivement en noir et en bleu). En bas : le facteur Lrs 14 (en vert) se fixe sur des opérateurs qui recouvrent la boîte TATA et le BRE empêchant ainsi la fixation de la TBP et de la TFB. (d'après Bell et Jackson 2001).

## 2) Les protéines architecturales des archaeobactéries :

Les archaeobactéries ont-elles une chromatine ? Actuellement, il n'y a pas de réponse définitive à cette question mais, il est clair qu'il n'existe pas un mécanisme unique et conservé pour compacter le génome des archaeobactéries. La Figure 4 montre la répartition des protéines architecturales à l'intérieur du règne des archaeobactéries. Certaines de ces protéines ont un motif semblable à celui des histones eucaryotes, d'autres non.

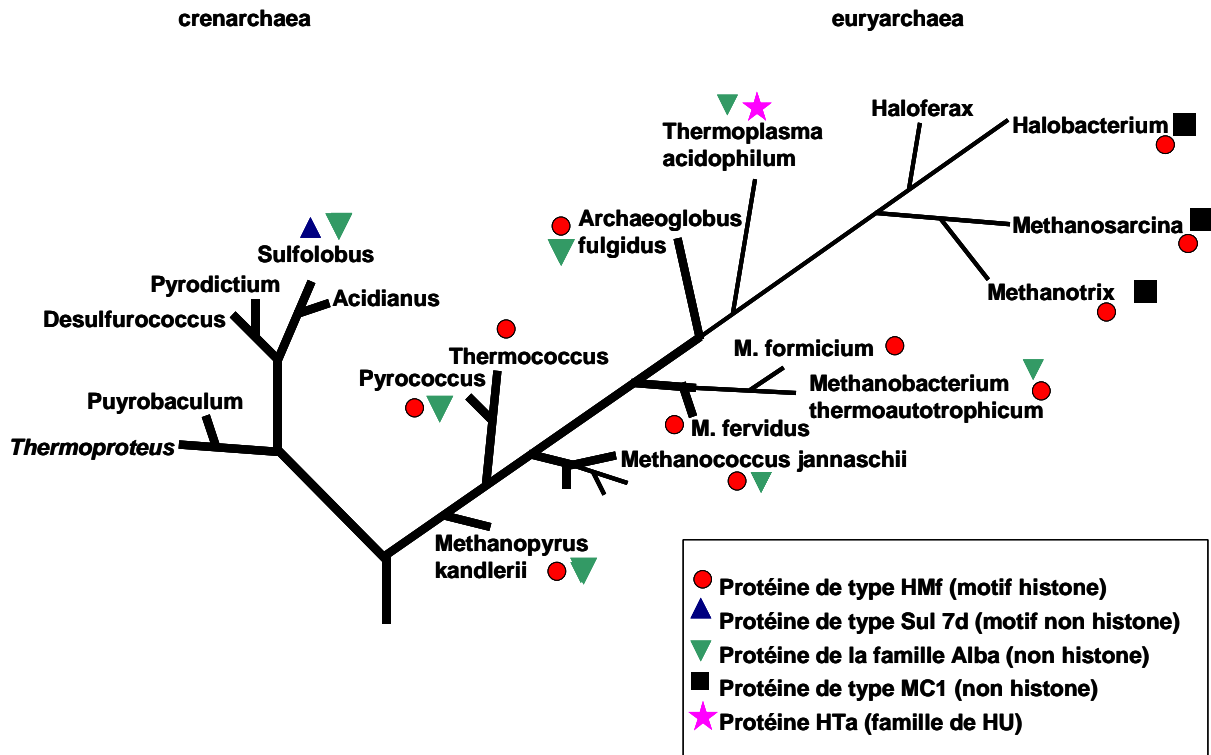


Figure 4 : répartition des différentes protéines architecturales dans le règne des archaeobactéries.

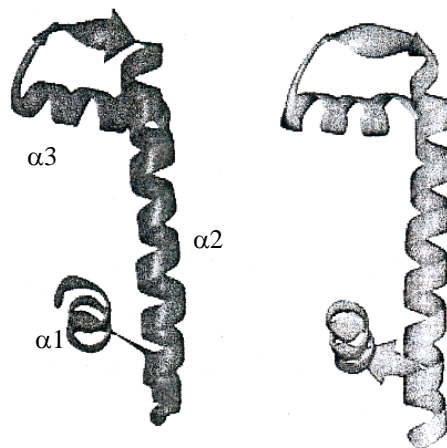
### 2.1) Protéines architecturales à motif histone :

Comme cela est montré dans la Figure 4, les seules archaeobactéries à posséder des protéines architecturales à motif histone sont les euryarchaeas. Actuellement une trentaine de séquences d’histones archaeobactériennes sont disponibles (Sandman et coll., 1998 ; Higashibata et coll., 1999). Ces séquences possèdent 65 à 85 % d’homologie entre elles.

La protéine la plus étudiée est HMf (histone de *Methanococcus fervidus*). Chez *M. fervidus* (température optimale de croissance de 83°C) on trouve un gène codant pour la protéine HMfA (68 résidus) et un gène codant pour la protéine HMfB (69 résidus). Au cours des différentes phases de croissance de *M. fervidus*, le rapport HMfA/HMfB varie (Sandman et coll., 1994). Des expériences de gel retard ont montré que les dimères (HMfA)<sub>2</sub>, (HMfB)<sub>2</sub> et (HMfA+HMfB) possèdent des caractéristiques différentes

(Decannier et coll., 2000). Ainsi, il a été démontré que les dimères (HMfA)<sub>2</sub> et (HMfB)<sub>2</sub> ont une thermostabilité différente *in vitro* et que celle-ci est dépendante de la concentration en protéines (Li et coll., 1998).

La structure de HMfB sous forme monomérique et dimérique a été déterminée par RMN (Starich et coll., 1996 ; Figure 5). Cette structure est semblable à celle des histones eucaryotes : chaque monomère est composé d'une grande hélice  $\alpha$  (notée  $\alpha_2$ ) qui est reliée à deux petites hélices  $\alpha$  (notées  $\alpha_1$  et  $\alpha_3$ ) par deux boucles  $\beta$  qui forment des boucles. En revanche, contrairement aux histones eucaryotes, la protéine HMfB ne possède pas d'extensions N et C terminales (sites de régulations post-traductionnelles importantes pour la biologie de la cellule ; Strahl et Allis, 2000).



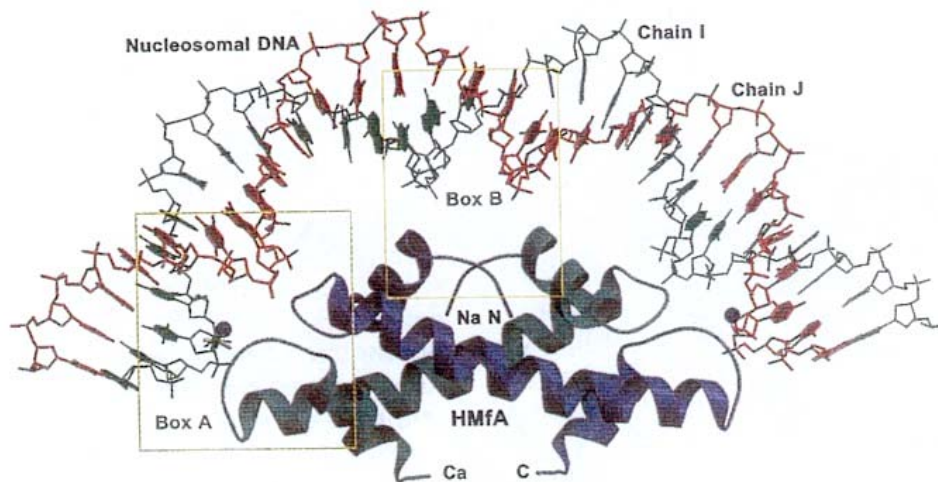
**Figure 5 : structure RMN d'un monomère de HMf.**

La structure tridimensionnelle du motif structural du monomère HMfB (à gauche) est comparée avec celle de l'histone H4 de poulet (à droite). (d'après Starich et coll., 1996).

En solution, HMf est trouvée majoritairement sous forme dimérique. En revanche, lorsque de l'ADN est présent dans le milieu, deux dimères s'assemblent en un tétramère qui va se fixer à l'ADN. L'association en dimère se fait grâce à des interactions antiparallèles entre l'hélice  $\alpha_2$  de chaque monomère et par des interactions parallèles entre les boucles L1/L2a et L1a/L2 (le « a » représente le deuxième monomère ; Decanniere et

coll., 2000 ; Figure 6). L'existence d'une structure octamérique se liant à l'ADN n'a pas été démontrée.

L'utilisation de différentes protéines mutantes a permis de montrer que la partie N terminale ainsi que l'hélice  $\alpha 2$  et les boucles L1 et L2 sont impliquées dans l'interaction avec l'ADN (Soares et coll., 2000). L'observation des séquences des différentes histones a montré que l'extrémité N terminale comporte toujours deux prolines (Decanniere et coll., 2000). Lors de la formation du dimère, ces prolines vont former une tétrade qui va interagir avec l'ADN. Cette tétrade de prolines a été recherchée dans les séquences des histones des cellules eucaryotes et n'a pas été trouvée.



**Figure 6 : modèle montrant la fixation d'un dimère (HMfA)<sub>2</sub> sur l'ADN.**

Le modèle est généré par remplacement dans le nucléosome eucaryote du dimère (H3+H4) par (HMfA)<sub>2</sub>. La paire L1-L2a et la région N terminale en interaction avec l'ADN sont encadrées (respectivement box A et B). (d'après Decanniere et coll., 2000).

Les tétramères d'HMf reconnaissent préférentiellement les séquences riches en AT (Widlund et coll., 1997) ou courbées par des suites d'adénines en phase avec l'ADN (Howard et coll., 1992). Par ailleurs des études ont montré que des séquences comportant 6 ou 8 répétitions du triplet CTG (Sandman et Reeve, 1999) et la séquence codant pour l'ARN 7S (Pereira et Reeve, 1999) sont elles aussi reconnues de façon préférentielle. L'ensemble des séquences préférentiellement reconnues par HMf a été recherché de

manière systématique par la méthode SELEX (Bailey et coll., 2000). Ces expériences n'ont pas permis de déterminer de séquence consensus de fixation de HMf. En revanche, les séquences issues du SELEX possèdent les caractéristiques des séquences préférentiellement reconnues par l'octamère d'histones : alternance de régions riches en G/C et de régions riches en A/T en phase avec le pas de l'hélice de l'ADN.

Le tétramère d'histones HMf va permettre d'enrouler 90 à 150 pb d'ADN en une structure de type nucléosome observable au microscope électronique (Sandman et coll., 1990). Ces particules sont proche de celles obtenues avec le tétramère d'histones (H3+H4)<sub>2</sub> qui permet d'enrouler 120 pb d'ADN et laisse une empreinte de 70 pb sur l'ADN (Dong et van Holde., 1991).

Une protéine homologue à la protéine HMf, la protéine HMt, a été isolé chez *Methanobacterium thermoautotrophicum*. Les deux sous-unités HMtA et HMtB présentent avec HMfA et HMfB plus de 60 % d'homologie. Comme HMf, la protéine HMt présente la capacité de compacter l'ADN (Tabassum et coll., 1992). La visualisation des complexes entre HMt et l'ADN par microscopie électronique révèle également la présence de structures ressemblant aux nucléosomes des eucaryotes (Pereira et coll., 1997).

Dans l'archaeabactérie *Methanopyrus kandleri*, dont la température optimale de croissance est de 112°C, on trouve une protéine histone inhabituelle. La différence majeure entre cette protéine, nommée HMk ou MkaH, et les autres histones, est qu'elle contient une seule chaîne peptidique qui comprend deux motifs histones qui s'organisent en une structure pseudo-dimérique. Les études par cristallographie ont montré que la structure du pseudo-dimère est superposable à celle des dimères des histones eucaryotiques ou archaeabactériennes (déviations des structures : 1.5 à 2.7 Å ; Fahrner et coll., 2001). La protéine HMk possède des propriétés qui sont spécifiques aux histones des archaeabactéries (par exemple l'absence d'extensions N et C terminales) et des propriétés spécifiques aux histones eucaryotiques (par exemple l'absence de la tétrade de proline).

En absence d'ADN, HMk est sous forme dimérique et comprend donc, comme le tétramère (H3+H4)<sub>2</sub>, 4 motifs histones associés. Malgré la forte homologie structurale entre un dimère de HMk et un tétramère d'histones, les observations par microscopie électronique n'ont pas fait apparaître une structure en collier de perles mais des boucles

avec un trou au centre (Pavlov et coll., 2002). Les auteurs attribuent cette apparence atypique au mode de coloration utilisée (acétate d'uranyle) et mentionnent que des résultats similaires ont déjà été observés avec les autres histones.

## 2.2) Protéines architecturales sans motif histone :

### 2.2.1) *Chez les Sulfolobales :*

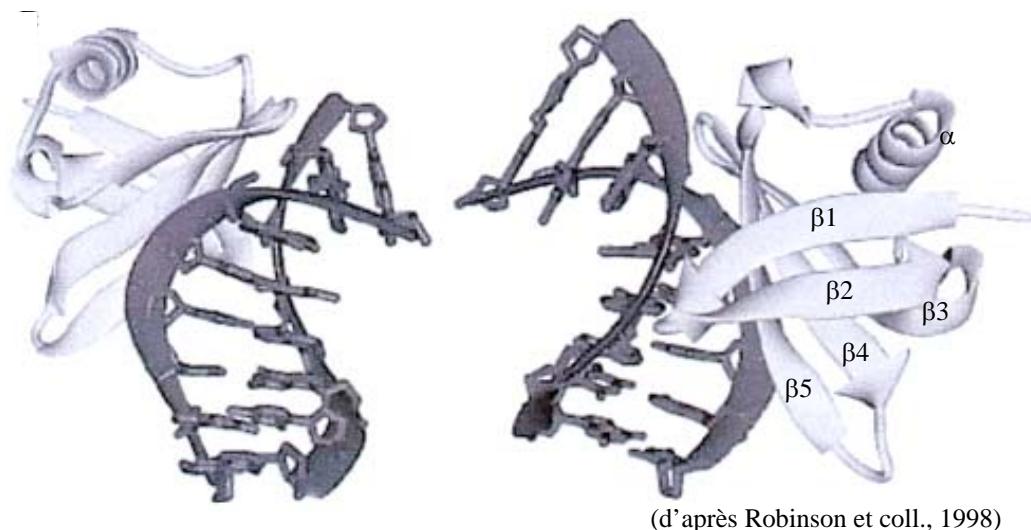
Les sulfolobales sont des archaeobactéries de la famille des crenarchaeas. Ces organismes vivent au niveau des points chauds volcaniques à des températures optimales pouvant aller de 60 à 85°C (Stetter et coll., 1990). Ces archaeobactéries produisent des petites protéines basiques de 7 kDa) fortement conservées dans les différentes espèces. L'ensemble de ces protéines constitue la famille Sul 7d (protéine de 7 kDa des sulfolobales). Les représentants les plus connus de cette famille sont Sso 7d chez *Sulfolobus solfataricus* et Sac 7d chez *S. acidocaldarius*. Ces deux protéines possèdent environ 90 % d'homologie de séquence. Les protéines de la famille Sul 7d représentent environ 5 % des protéines solubles isolées des sulfolobales.

La structure des protéines Sso 7d et Sac 7d a été résolue par RMN (respectivement Baumann et coll., 1994 ; Edmondson et coll., 1995). Les résultats de ces deux études montrent que les deux protéines ont une structure tridimensionnelle identique. Elles s'organisent en une structure globulaire compacte comprenant un cœur protéique hydrophobe relativement petit (11 acides aminés). La structure tertiaire des protéines Sso 7d et Sac 7d est constituée de trois brins de feuillets  $\beta$  antiparallèles (brins  $\beta_3$ ,  $\beta_4$ ,  $\beta_5$ ) superposés de façon orthogonale à deux autres brins de feuillets  $\beta$  (brins  $\beta_1$ ,  $\beta_2$ ) puis d'une hélice  $\alpha$  C terminale (Figure 7). Des études comparatives ont montré que ces deux protéines possèdent une homologie structurale avec les domaines SH3 et les chromo domaines (respectivement Baumann et coll., 1994 ; Ball et coll. 1997). Dans les cellules eucaryotes, les protéines à domaine SH3 sont impliquées dans la transduction du signal et les protéines possédant un chromo domaine ont un rôle dans la répression de l'expression des gènes. Les conséquences biologiques de cette homologie de structure entre les protéines de la famille Sul 7d et les protéines à domaine SH3 ou à chromo domaine ne sont pas encore comprises. Par ailleurs, les études de Gao et collaborateurs ont montré que la

structure des protéines de la famille Sul 7d peut s'aligner avec celle des protéines de réponse au choc thermique froid d'*E. coli* (CspA/CspB ; Gao et coll., 1998).

L'étude des complexes a montré que la fixation des protéines de la famille Sul 7d se fait sans spécificité de séquence, de façon monomérique et non coopérative via le petit sillon de l'ADN (Mc Afee et coll., 1995 ; Baumann et coll., 1994). Les mêmes travaux ont montré que la taille du site de fixation est de 4 pb pour Sac 7d et 6 pb pour Sso 7d. Pour se fixer à l'ADN, les protéines de la famille Sul 7d intercalent les chaînes latérales de plusieurs résidus hydrophobes (entre autres la tyrosine 7, la valine 25 et la méthionine 28 respectivement brins  $\beta 1$ ,  $\beta 3$  et  $\beta 4$ ). Les interactions sont stabilisées par des liaisons hydrogène entre l'ADN, le tryptophane 23 et l'alanine 44 (respectivement brin  $\beta 3$  et  $\beta 5$ ).

En 2001, des travaux ont montré que contrairement aux histones des eucaryotes, les méthylations hétérogènes qui touchent 5 des 14 lysines des protéines de la famille de Sul 7d n'affectent pas l'affinité des protéines pour l'ADN (Edmondson et Shriver, 2001). Le rôle de ces méthylations n'est pas encore compris mais les observations montrent qu'après un choc thermique leur nombre augmente (Baumann et coll., 1994).



**Figure 7 : structure tridimensionnelle de deux complexes Sac 7d/ADN.**

La fixation des protéines de la famille Sul 7d provoque un changement important de la structure de l'ADN. Différentes études ont montré un élargissement du petit sillon de l'ADN ainsi que la formation d'un coude important d'environ 60° (Robinson et coll., 1998 ; Agback et coll., 1998). Le groupe de Forterre a montré que la fixation de la protéine Sso 7d sur l'ADN provoque un déroulement local de la double hélice (Lopez-Garcia et coll., 1998). Les auteurs émettent l'hypothèse que la protéine Sso 7d est impliquée dans le maintien du degré de compaction de l'ADN des sulfolobales. En effet, ces organismes vivant à haute température, leur ADN a tendance à se surenrouler positivement (Lopez-Garcia et coll., 1997). La protéine Sso 7d, en générant un déroulement de l'ADN serait donc garante de l'intégrité de l'ADN. Le mécanisme de fixation des protéines de la famille Sul 7d ainsi que certaines conséquences de leur fixation sont semblables aux résultats obtenus pour la protéine fixant la boîte TATA (TBP) ou les boîtes HMG SRY et LEF-1 (Werner et coll., 1996).

Par ailleurs, la fixation des protéines de la famille de Sul 7d provoque une augmentation d'environ 40°C de la température de fusion des ADN double brin (Baumann et coll., 1994) et facilite l'appariement de brins complémentaires d'ADN à haute température (Guagliardi et coll., 1997).

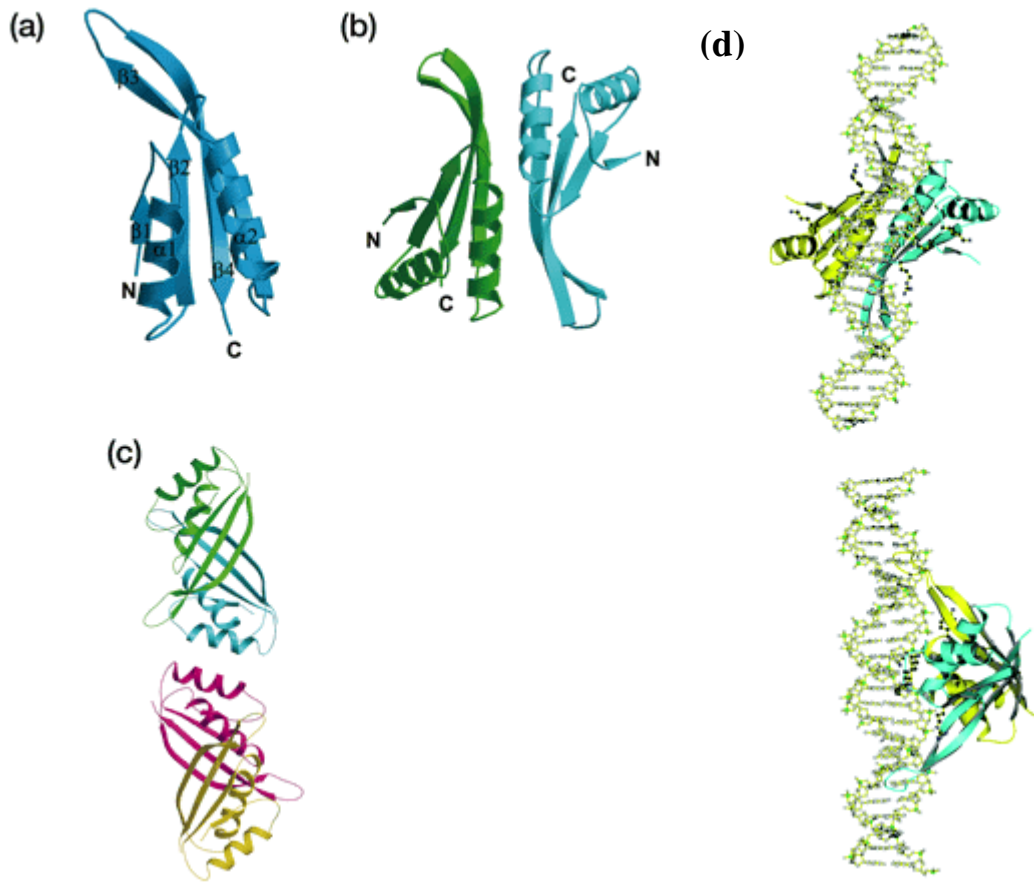
Certaines études récentes ont suggéré que la protéine Sso 7d interviendrait dans les phénomènes de recombinaison en facilitant la résolution des jonctions de Holliday en interagissant avec la protéine Hjc (enzyme de résolution des jonctions de Holliday par catalyse ; Kvaratskhelia et coll., 2002). En 2000, le groupe de Rossi a montré que la protéine Sso 7d possède une activité ATPase (dépendante des cations bivalents) et que sa structure en présence d'ATP est différente de celle en présence d'ADP (Guagliardi et coll., 2000). Dans la même étude, les auteurs montrent que lorsque la protéine est dans sa configuration « ATP », elle est capable de reconnaître et de dissoudre les agrégats protéiques. La dissolution des agrégats nécessite l'hydrolyse de l'ATP. La fonction ATPase est indépendante de la fixation à l'ADN ou de la reconnaissance des agrégats protéiques. Il est important de noter qu'actuellement, Sso 7d est la seule protéine d'archaeobactérie à posséder cette fonction de dissolution des agrégats protéiques. Dans les autres organismes (bactéries, cellules eucaryotes), cette fonction est assurée par les protéines HSP (Heat Shock Protein) 60, 70 et 100 qui n'ont pas d'homologue connu chez les archaeobactéries.



### 2.2.2) Protéines de la famille Alba :

Chez les sulfobactéries, on trouve une deuxième famille de protéines abondantes (4 à 5 % des protéines solubles). Contrairement aux protéines de la famille Sul 7d qui ne sont présentes que dans les sulfobactéries (crenarchaeas), des protéines de la famille Alba (Acetylation Lower Binding Affinity) ont été isolées des euryarchaeas et des crenarchaeas (Figure 4, p15) et des homologues ont même été trouvés chez les eucaryotes (plantes, *leishmania*, trypanosome ; Bell et coll., 2002). Les protéines de la famille Alba ont une taille moyenne de 10 kDa. De façon surprenante, quand un organisme possède une protéine de la famille Alba, il code aussi pour une autre protéine chromatinienne : histone, protéine de la famille Sul 7d, ou homologue de la famille HU. Par ailleurs, les organismes qui ne codent pas pour une protéine de la famille Alba possèdent en général une autre protéine architecturale (par exemple MC1 chez les Methanocarsines).

La surexpression chez *E. coli* de la protéine Alba de *S. solfataricus* a permis de déterminer sa structure cristallographique. Cette structure montre qu'en solution la protéine Alba est sous forme dimérique. Chaque monomère comprend deux grands feuillets ( $\beta 3$  et  $\beta 4$ ) formant une épingle à cheveux flexible qui s'étend à partir du cœur protéique constitué de deux alternances feuillet  $\beta$ /hélice  $\alpha$  (Wardleworth et coll., 2002 ; Figure 8 (a)). Dans la même étude, les auteurs ont démontré que cette structure est semblable à celle retrouvée dans la partie N terminale de la DNase I ou dans la partie C terminale du facteur bactérien d'initiation de la transcription IF3. En 2003, le groupe de Marmorstein a montré que la structure de la protéine Alba de *A. fulgidus* est identique à celle de *S. solfataricus* (Zhao et coll., 2003) et que les interactions entre deux monomères ont lieu grâce à des contacts entre des résidus de l'hélice  $\alpha 2$  et l'épingle à cheveux constituée des feuillets  $\beta 3$  et  $\beta 4$  (Figure 8 (b)).



**Figure 8 : structure de la protéine Alba.**

(a), (b), (c) structure des protéines de la famille Alba (*A. fulgidus*) respectivement sous forme monomérique, dimérique et tétramérique (d'après Zhao et coll., 2003). (d) : structure d'un complexe formé entre un dimère de protéine de la famille Alba (*S. solfataricus*) et l'ADN. L'image du bas est la même que celle du haut après rotation de 90° par rapport à l'axe de l'ADN (d'après Wardleworth et coll., 2002).

Les expériences de Zhao et collaborateurs (2003) ont montré que la fixation des protéines de la famille Alba a lieu par association de deux dimères. Ce tétramère provoque une courbure de 60° de l'ADN (Zhao et coll., 2003). La tétramérisation de la protéine se fait grâce à des interactions hydrophobes entre l'hélice  $\alpha 1$  et la partie C terminale de l'hélice  $\alpha 2$ . Dans cette structure, les deux dimères sont orientés de manière antiparallèle (Figure 8 (c)). L'association entre l'ADN double brin et les protéines de la famille Alba se fait sans spécificité de séquence et de manière coopérative en un site dont la taille est estimée à 15 pb (Xue et coll., 2000 ; Bell et coll., 2002). Au sein d'un tétramère, chaque dimère va insérer les bras constitués des feuilletts  $\beta 3$  et  $\beta 4$  à l'intérieur du petit sillon de

l'ADN (comme le font les protéines de la famille HU/IHF chez les eubactéries). Cette insertion permet l'interaction des lysines 16 et 17 du cœur protéique avec le grand sillon de l'ADN (Wardleworth et coll., 2002 ; Figure 8 (d)).

La fixation des protéines de la famille Alba sur l'ADN induit des supertours négatifs (Xue et coll., 2000). Les observations au microscope électronique réalisées par l'équipe de Dobrinski ont montré que la fixation de la protéine ne provoque pas la compaction de l'ADN (Lurz et coll., 1986).

En 2002, les travaux de Bell et collaborateurs ont montré que les protéines Alba et la désacétylase Sir2 interagissent (Bell et coll., 2002). Les auteurs ont montré que la protéine Alba de *S. solfataricus* possède deux sites d'acétylation (lysines 16 et 17) et que les formes acétylées et non acétylées existent *in vivo*. Des études comparatives d'affinité menées par gel retard ont montré que la forme non acétylée de la protéine Alba fixe environ 30 fois plus l'ADN que la forme acétylée. Les lysines 16 et 17 étant impliquées dans l'interaction avec l'ADN, les auteurs ont supposé que la fixation des protéines de la famille Alba est augmentée par la désacétylation des lysines 16 et 17. Ces hypothèses ont été confirmées par les travaux de Zhao et collaborateurs (2003). Dans ces travaux, les auteurs montrent que les lysines désacétylées par la protéine Sir2 sont impliquées à la fois dans l'interaction avec l'ADN et dans le processus de tétramérisation de la protéine Alba de *A. fulgidus*. La protéine Sir2 est donc capable d'augmenter l'affinité de la protéine Alba en permettant la tétramérisation de la protéine. Il est important de noter que les protéines de la famille Alba sont actuellement les seules protéines d'archaeobactéries dont la fixation à l'ADN est régulée par un processus d'acétylation/désacétylation. Cette régulation de la fixation de protéines chromatiniennes par acétylation/désacétylation est connue chez les protéines histones eucaryotes.

### 2.2.3) HTa :

La protéine HTa est une petite protéine basique de 9.9 kDa très affine pour l'ADN, issue d'une archaeobactérie thermoacidophile *Thermoplasma acidophilum*. Sa séquence complète a été déterminée par Delange et collaborateurs en 1981 et révèle une certaine homologie avec celle de la protéine HU d'*E. coli* (Delange et coll., 1981 ; Oberto et coll.,

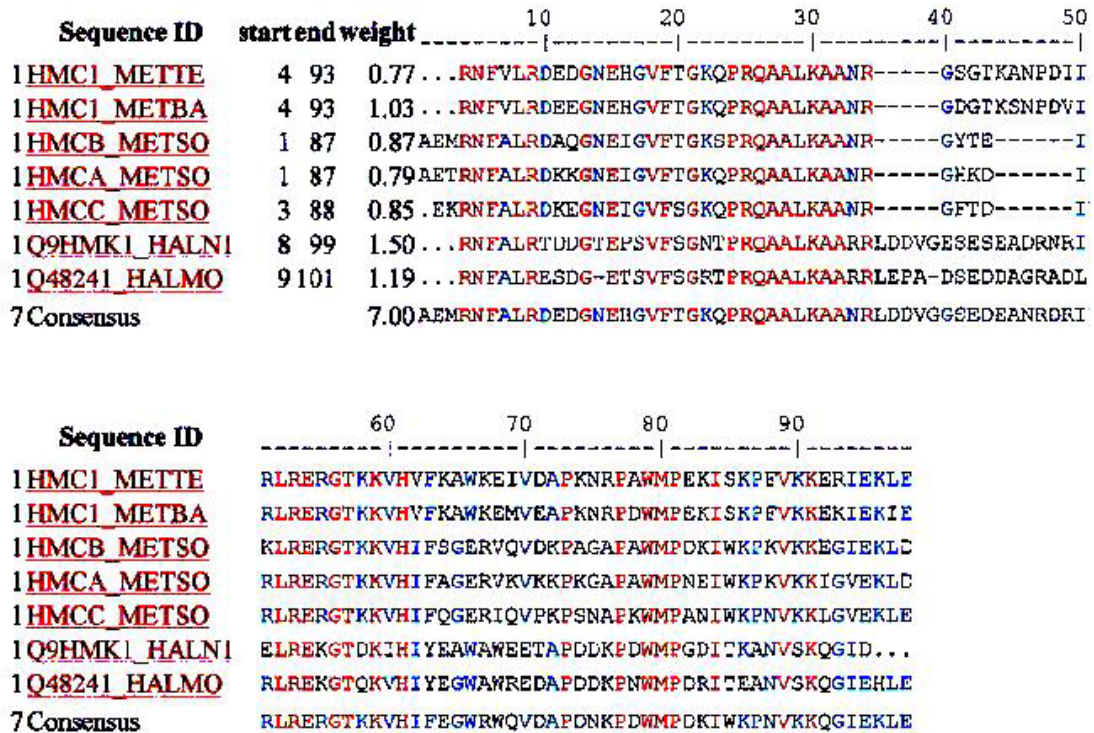
1994). Les acides aminés conservés chez toutes les protéines HU et sensés intervenir dans la fixation de la protéine sur l'ADN se retrouvent dans la protéine HTa.

#### 2.2.4) MC1 :

Les protéines MC (Metanogen Chromosomal protein) forment une famille de protéines présentes chez les méthanosarcines. La première protéine (nommée MC1) a été isolée de la souche *M. barkeri* (Chartier et coll., 1985) puis, 4 autres protéines de cette famille ont été séquencées et caractérisées (Chartier et coll., 1988, 1989a). L'analyse des différents génomes d'archaeobactéries montre que plusieurs gènes peuvent être présents dans un même organisme et qu'une protéine très similaire à MC1 existe chez les halophiles (Ng et coll., 2000).

La comparaison des séquences des différentes protéines MC1 (Figure 9) montre que la protéine possède différentes régions très conservées : 24-35 ; 50-61 (très basique) et 79-81 (strictement conservée). Chez les méthanosarcines, on peut aussi noter la présence de cinq prolines strictement conservées aux positions 24, 73, 77, 81 et 87 (Chartier et coll., 1989 a, b).

La protéine MC1 que nous étudions au laboratoire est isolée de l'archaeobactérie *Methanosarcina thermophila* (aussi nommée CHTI 55). La protéine MC1 est une petite protéine basique (charge nette de + 12) de 93 acides aminés (11 kDa). Du fait de la grande quantité de résidus basiques et acides (respectivement 24 % et 14 %) cette protéine est nettement hydrophile (Laine et coll., 1986) et ne comporte pas de domaine hydrophobe. Avec une concentration cellulaire estimée à 1 protéine pour 170 pb (Imbert et coll., 1988), la protéine MC1 est très abondante *in vivo*. Les trois variants trouvés chez *Methanotrix sohengonii* (notés MC1a, MC1b, MC1c) présentent 60 % d'homologie avec la protéine MC1 CHTI 55 (Chartier et coll., 1989 a).

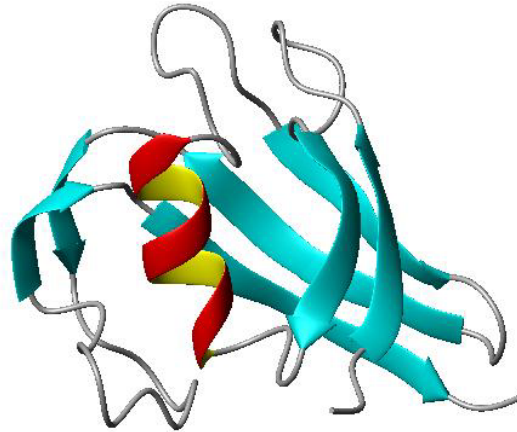


**Figure 9 : alignement des séquences des protéines MC1 de différents organismes.**

METTE : *Methanosarcina thermophila*; METBA : *M. barkeri*; METSO : *Methanotrix. sohengenii*; HALN1 : *halobacterium sp*; HALMO : *halococcus morrhuae*. Les acides aminés notés en rouge sont strictement conservés, les acides aminés notés en bleu sont très conservés, ceux notés en noir ne le sont pas.

Seule la structure et les propriétés de la protéine sont actuellement connues. Ainsi, différentes études effectuées *in vitro* ont montré que la protéine compacte l'ADN et le protège contre les rayons ionisants et la dénaturation thermique.

La structure tridimensionnelle de la protéine MC1 a été étudiée par RMN au laboratoire dans l'équipe de G. Lancelot par F. Paquet. Les résultats font apparaître une structure composée de 2 brins  $\beta$  disposés orthogonalement à 3 autres brins  $\beta$ , d'une hélice  $\alpha$  et d'une grande boucle mal structurée (Figure 10). Les recherches dans les banques de données ont montré que le repliement de la protéine MC1 constitue un nouveau motif structural. En 1991, les expériences de photo-pontage réalisées par Katouzian-Safadi et collaborateurs ont montré que la protéine MC1 interagit avec l'ADN via sa partie C terminale (résidus 74 à 86 localisés dans une boucle peu structurée).



**Figure 10 : structure RMN de la protéine MC1.**

PDB id : 1kom. (d'après Paquet et coll.).

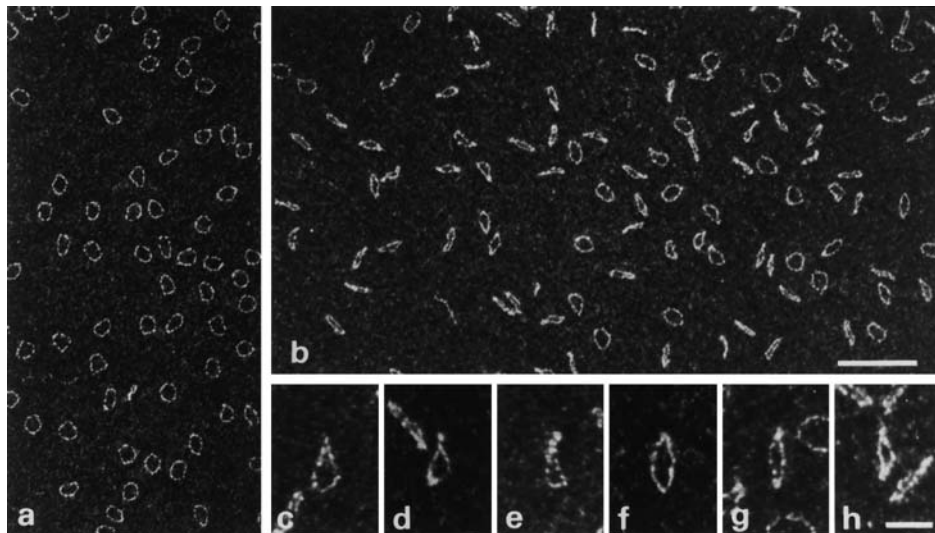
L'utilisation de la technique de gel retard a permis de déterminer les caractéristiques de la fixation de MC1 sur l'ADN (Culard et coll., 1993). Les résultats ont montré que MC1 se fixe sur les ADN double brin sous forme monomérique et non coopérative. Dans ces expériences, les auteurs ont montré que la taille du site exclu est de 11 pb.

Les expériences réalisées par Paradinas et collaborateurs ont montré que la protéine MC1 reconnaît spécifiquement les jonctions à 4 voies mais pas les jonctions à 3 voies ou les ADN présentant des « bulges ». L'affinité de MC1 pour les jonctions à 4 voies est 60 fois plus forte que pour les ADN linéaires et n'est pas dépendante de la concentration en cations bivalents (Paradinas et coll., 1998). Ces résultats semblent indiquer que c'est la structure fine des jonctions à 4 voies (par exemple la largeur du petit sillon) qui est reconnue par la protéine et non l'angle formé entre les bras de la jonction. Dans la même étude, les auteurs démontrent par des expériences d'empreintes moléculaires aux radicaux hydroxyle que la protéine MC1 se fixe au centre des jonctions à 4 voies. Les résultats des empreintes au permanganate de potassium ( $\text{KMnO}_4$ ) ont montré que la fixation de la protéine MC1 entraîne un changement de structure des jonctions à 4 voies.

En 1995, Toulmé et collaborateurs ont réalisé l'étude des complexes formés entre la protéine MC1 et des minicercles d'ADN relâchés (Toulmé et coll., 1995). L'analyse sur

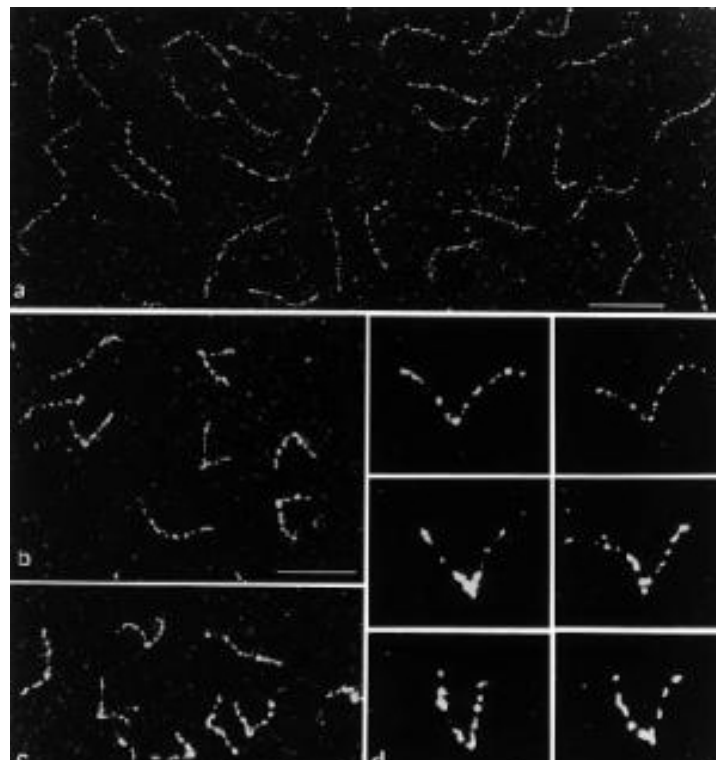
gel retard des différents complexes obtenus a permis de montrer la formation préférentielle du complexe composé de 2 protéines MC1 par minicercle d'ADN (complexe C2). Ce complexe est très stable et sa demi-vie est de 130 min dans un tampon contenant 80 mM de NaCl (moins de 1 min pour les complexes obtenus avec de l'ADN linéaire). Le profil de migration des différents complexes montre que la protéine MC1 compacte l'ADN. En effet, les complexes composés de 1 à 8 protéines migrent plus vite que les minicercles relâchés libres (le complexe C2 a la vitesse de migration la plus grande). L'analyse du profil de dissociation des complexes C2 a montré que lorsqu'on augmente la concentration en sel dans le milieu, les deux protéines MC1 se dissocient en même temps (le complexe composé d'une protéine fixée sur un minicercle n'est pas observé). L'analyse par microscopie électronique des complexes C2 a montré que la fixation des protéines se fait de manière diamétralement opposée. Ainsi, les minicercles complexés prennent une forme de fuseau (Figure 11). Les images de microscopie électronique ont aussi permis de déterminer les points structuraux suivants : 1) la longueur du contour de l'ADN est la même en présence et en absence de protéine donc l'ADN ne s'enroule pas autour de la protéine ; 2) la fixation de protéine sur les minicercles relâchés provoque un changement de conformation des minicercles ; 3) aucune boucle d'ADN résultant d'interaction MC1-MC1 n'est observée. En 1996, les travaux de Teyssier et collaborateurs ont montré que la fixation de MC1 sur des minicercles surenroulés négativement est plus forte que sur des minicercles relâchés. Ainsi, la protéine MC1 est capable de se fixer préférentiellement sur les topoisomères  $-2$  (demi-vie de 280 min dans un tampon contenant 200 mM NaCl et 10 mM  $MgCl_2$  ; Teyssier et coll., 1996).

En 1994, les expériences de Teyssier et collaborateurs ont montré que malgré une fixation non spécifique forte, la protéine MC1 est capable de reconnaître spécifiquement une séquence d'ADN localisée en amont du gène codant pour la methyl CoM reductase (Teyssier et coll., 1994). Dans cette étude, les auteurs ont montré que la fixation de la protéine MC1 sur ce site (dénommé site I) induit une forte courbure de l'ADN et la formation de sites hypersensibles à la DNase I. Les études par microscopie électronique ont permis de déterminer que l'angle de courbure résultant de la fixation de MC1 est de  $116^\circ$  (Le Cam et coll., 1999, Figure 12). Certains aspects structuraux restent cependant à déterminer : direction de la courbure, mode d'interaction avec l'ADN.



**Figure 11 : visualisation des minicercles d'ADN nus ou complexés avec MC1.**

Les minicercles de 207 pb sont incubés en absence (a) ou en présence (b-h) de MC1 à un rapport protéine/minicercle de 7. Les barres représentent 100 nm (a-b) et 25 nm (c-h). (d'après Toulmé et coll., 1995).



**Figure 12 : visualisation des complexes MC1/site I.**

Fragment de 176 pb contenant le site I libre (a) ou complexé à MC1 (b, c, d). (d'après Le Cam et coll., 1999).



De nombreuses questions se posent quant à la fonction de la protéine MC1. Ainsi, on peut se demander quelles sont les conditions qui permettent l'expression de la protéine MC1. En effet, l'analyse des génomes des méthanosarcines et des halobactéries a montré que ces organismes possèdent également une séquence codante pour une protéine de la famille de HMf. Pourtant, aucune protéine de type HMf n'a été isolée de ces organismes. Des travaux, réalisés chez *Thermococcus zilligii* (Dinger et coll. 2000), ont montré que les gènes codants pour les protéines histones HTz1 et HTz2 ne sont plus exprimés quand la culture est en phase stationnaire. En revanche, une protéine dont la séquence est proche de celle de la protéine Alba est trouvée dans les cultures en phase stationnaire. Les auteurs interprètent ces résultats comme étant un moyen mis en place par cet organisme pour moduler la structure de son ADN (donc l'expression génique) suivant les conditions de culture. Il est possible que MC1 joue le même rôle chez les méthanosarcines et les halophiles.

# **PROTEINES ARCHITECTURALES DES EUBACTERIES**

## 1) Introduction : le nucléoïde :

Sur l'ADN bactérien, un grand nombre de protéines sont fixées. Cet ensemble ADN/protéine est nommé nucléoïde. La composition exacte du nucléoïde est difficile à déterminer car elle dépend de plusieurs facteurs (Robinow et Kellenberg, 1994) : méthode d'isolement, milieu de culture, phase de croissance de la culture. De façon générale, on trouve dans le nucléoïde des protéines de type histone, des ARN polymérase, des chaînes naissantes d'ARN et des fragments de membrane bactérienne.

Les résultats d'observations directes au microscope électronique de nucléoïdes isolés montrent que la majorité du nucléoïde s'étale en rosette. Dans cette structure, de larges boucles émergent d'une partie centrale plus dense. Chaque boucle est appelée sous-domaine, la partie centrale est dénommée « nœud ». Chez *E. coli*, on trouve entre 20 et 50 sous-domaines par nucléoïdes (Hinnebusch et coll., 1997). La structure des sous-domaines est organisée en supertours. Dans les bactéries *Borrelia burgdorferi* et *B. hermsii*, les auteurs ont observé des nucléoïdes composés d'un réseau d'ADN sans « nœud » dense.

Au sein du nucléoïde, les protéines de type histone ont plusieurs fonctions : 1) structuration de l'ADN ; 2) régulation globale des processus biologiques tels que la réplication, la transcription et la recombinaison. Actuellement, 4 protéines de type histone sont reconnues pour être les constituants majeurs du nucléoïde (FIS, HU, IHF, H-NS). Une étude visant à déterminer la concentration de ces 4 protéines (ainsi que celle de 8 autres) au cours des différentes phases de culture d'*E. coli* a été menée par western blot (Ali Azam et coll., 1999). Les résultats (Tableau 2) montrent que le niveau d'expression de ces protéines au cours des différentes phases de culture est cohérent avec leurs fonctions connues. Ainsi, la protéine FIS qui a pour rôle principal l'augmentation de la transcription des gènes nécessaires à la croissance cellulaire (gènes codants pour les ARN ribosomiques et les ARN de transfert) n'est exprimée que pendant la phase de croissance exponentielle. La protéine Dps dont le rôle est de protéger l'ADN contre les stress oxydatifs ou de nutrition a un taux maximal d'expression pendant la phase stationnaire. Les fonctions des protéines HU, IHF et H-NS seront décrites de façon détaillée dans la suite de ce chapitre.

**Tableau 2 : niveau d'expression de certaines protéines de type histone au cours des différentes phases de culture de bactéries *E. coli*.**

Les valeurs correspondent au nombre de copies par cellule. FIS : factor for inversion stimulation ; HU : heat unstable nucleoid protein ; Dps : DNA binding protein from starved cell ; IHF : integration host factor ; H-NS : histone like nucleoid structuring protein.

	FIS	HU	Dps	IHF	H-NS
phase exponentielle	60000	50000	6000	12000	20000
début phase stationnaire	0	30000	120000	55000	15000
fin phase stationnaire	0	15000	180000	30000	10000

Les protéines de type histone participent à l'organisation du nucléoïde avec les enzymes de topologie (topoisomérase I, gyrase). En effet, lors de la réplication et de la transcription, le déplacement de la fourche va générer des supertours positifs en aval et des supertours négatifs en amont. La topoisomérase I a pour rôle la suppression des supertours positifs. La gyrase permet de relâcher les supertours négatifs. Ces deux enzymes sont donc impliquées dans le maintien du niveau de compaction de l'ADN. L'observation des nucléoïdes isolés a montré que des traitements à la RNase, à la topoisomérase IV ou au lysozyme provoquent une décondensation du nucléoïde (Bendich, 2001). Les traitements à la RNase semblent montrer que les ARN interviennent dans la compaction du nucléoïde *in vitro*. Pourtant, ceci est controversé. En effet, des travaux ont montré par l'utilisation d'anticorps dirigés contre l'ARN polymérase que la synthèse d'ARN n'a pas lieu au niveau du « nœud » du nucléoïde mais dans les régions moins condensées (Durrenberger et coll., 1988 ; Kellenberger, 1990). Par ailleurs, les expériences à la topoisomérase argumentent en faveur d'une stabilisation du centre du nucléoïde par des « nœuds » ou toute autre association de chaîne pouvant être détruite par le passage d'un brin d'ADN autour de l'autre.

Certains facteurs, au contraire, provoquent des contraintes qui vont avoir tendance à étirer le chromosome. L'analyse des nucléoïdes isolés a montré que l'ARN polymérase est

retrouvée de manière très abondante. Cette polymérase, pour permettre la synthèse d'ARN, nécessite l'ouverture de la double hélice donc une décondensation de l'ADN. Dans la bactérie, la traduction des ARN en protéines commence avant que la transcription ne soit terminée (mécanismes couplés). De plus, les protéines membranaires en cours de synthèse vont s'ancrer rapidement à la membrane. La conséquence de ces deux observations est que le nucléoïde est accroché à la membrane de la cellule. Comme cette membrane est continuellement en croissance, elle va avoir tendance à étirer le nucléoïde. Le fait que le nucléoïde soit ancré à la membrane explique la présence de fragments de membrane dans les préparations de nucléoïde.

## 2) **La protéine FIS** :

FIS est une protéine homodimérique trouvée chez *E. coli* ainsi que chez un grand nombre d'eubactéries. Chaque monomère est composé de 98 acides aminés. Comme cela a été mentionné dans la partie précédente, cette protéine est un des constituants majeurs du nucléoïde (Pan et coll., 1996) et elle n'est exprimée qu'au cours de la phase exponentielle de croissance (Ali Azam et coll., 1999).

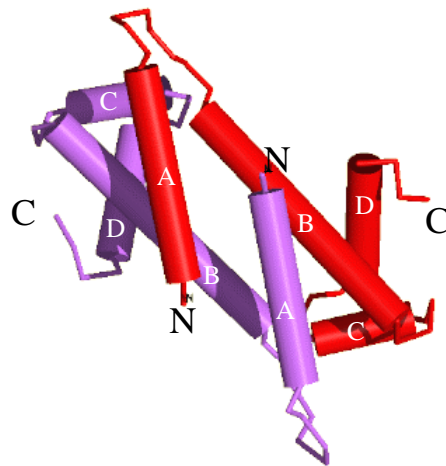
La protéine FIS a tout d'abord été isolée pour son rôle dans la régulation des réactions de recombinaison faisant intervenir l'invertase Hin (Johnson et coll., 1986 ; Koch et Kahmann, 1986). Par la suite, d'autres propriétés ont été identifiées : 1) activation ou répression de la transcription de certains gènes nécessaires à la croissance cellulaire ; 2) intervention comme co-facteur dans la réplication, la division cellulaire et la séparation des chromosomes ; 3) participation à la recombinaison site spécifique du phage lambda (Ball et Johnson, 1991a, b).

La protéine FIS reconnaît de façon spécifique des séquences dégénérées de 15 pb possédant des répétitions inversées (GNN(C/T)(A/T)NN(A/T)NN(T/C)(G/A)NNC ; Pan et coll., 1996 ; Bokal et coll., 1995 ; Bruist et coll., 1987). L'affinité de la protéine pour ces séquences est environ 50 fois plus grande que pour des séquences quelconques.

La structure cristallographique montre que chaque monomère est constitué de 4 hélices  $\alpha$  (notée A, B, C, D) séparées par des boucles. Les hélices A et B s'organisent en un cœur protéique, les hélices C et D en un motif hélice-tour-hélice. (Kostrewa et coll.,

1992). Au sein du dimère, chaque monomère est assemblé de manière antiparallèle grâce à des interactions hydrophobes (Figure 13). Ces interactions se produisent entre l'hélice A d'un monomère et l'hélice B du deuxième monomère. L'interaction de chaque motif hélice-tour-hélice avec deux grands sillons adjacents de l'ADN permet la fixation du dimère. Dans le complexe, l'ADN a une courbure de 90° dirigée vers le petit sillon. Dans la même étude, les auteurs démontrent à l'aide de protéines mutantes que la partie N terminale est impliquée dans l'interaction avec l'invertase.

Une étude récente montre par quels mécanismes la protéine FIS régule la transcription au niveau du promoteur *tyrT* (Auner et coll., 2003). Les données obtenues montrent que la régulation est la conséquence de deux phénomènes : 1) augmentation de la transcription par fixation de FIS sur UAS (Upstream Activating Sequence) et par interaction avec l'ADN polymérase ; 2) réduction de la transcription par réduction du degré de surenroulement global de la double hélice.



**Figure 13 : structure d'un dimère de FIS.**

### 3) La protéine HU :

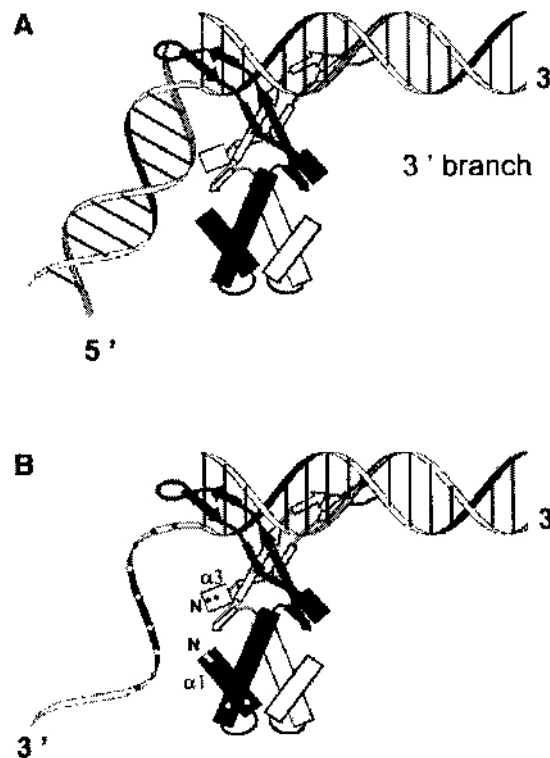
HU est une protéine de petite taille (18 kDa), fortement basique et très abondante (15000 à 50000 copies par cellule suivant le stade de culture). La protéine HU est composée de 2 sous-unités et est homodimérique dans tous les organismes sauf chez les entérobactéries où elle peut être hétérodimérique (ex. *E. coli*). Chez *E. coli*, les 2 sous-unités sont notées  $\alpha$  et  $\beta$  et elles sont respectivement codées par les gènes *hupA* et *hupB*. L'homologie de séquence de ces 2 sous-unités est de 70 %. Dans les bactéries *E. coli* en phase exponentielle de croissance, on trouve majoritairement l'homodimère  $\alpha_2$ . En revanche, lorsque les bactéries sont en phase stationnaire, l'hétérodimère  $\alpha\beta$  représente 90 à 95 % des protéines HU isolées.

Initialement isolées comme un facteur de transcription permettant l'activation de certains gènes du bactériophage lambda (Rouvière-Yaniv et Gros, 1975), elle s'avère être impliquée dans un grand nombre de phénomènes biologiques : 1) transposition du phage Mu (Craigie et coll., 1985) ou du transposon Tn10 (Morisato et Kleckner, 1987) ; 2) inversion de gènes (Wada et coll., 1989) ; 3) initiation de la réplication de l'ADN (Bramhill et Kornberg, 1988 ; Hwang et Kornberg, 1992) ; 4) stimulation de l'ARN polymérase T7 (Morales et coll., 2002) ; 5) modulation de la fixation spécifique de protéines régulatrices comme les protéines CRP, IHF, LexA (Flashner et Gralla, 1988 ; Bonnefoy et Rouvière-Yaniv, 1992 ; Preobrajenskaya, 1994).

Par ailleurs, la protéine HU a été décrite comme permettant la formation de structures ressemblant à celles des nucléosomes eucaryotes (Rouvière-Yaniv et coll., 1979). Cependant, cette fonction de compaction de l'ADN est soumise à de nombreuses controverses et ceci pour plusieurs raisons. Tout d'abord, les structures de type nucléosomique observées en 1979 n'ont pas pu être reproduites. Ensuite, des études récentes montrent un effet inverse (Dame et Gossen, 2002).

La fixation de la protéine HU est indépendante de la séquence d'ADN. En revanche, HU reconnaît avec une haute affinité les ADN simple brin ou les ADN double brin contenant une coupure simple brin, une délétion d'un ou deux nucléotides (Castaing et coll., 1995) ou une jonction à quatre voies (Bonnefoy et coll., 1994). Des expériences ont montré que les souches d'*E. coli* déficientes en gènes *hupA* et *hupB* sont hypersensibles aux radiations  $\gamma$  (Boubrik et Rouvière-Yaniv, 1995) et aux UV (Li et Waters, 1998). Ces résultats peuvent être interprétés de plusieurs façons. Il est en effet possible de déduire de

ces expériences que la protéine HU permet de protéger l'ADN des radiations en agissant comme un écran. Il est aussi possible d'envisager que la protéine HU reconnaît les lésions provoquées par les radiations et qu'elle permet ainsi le recrutement du système de réparation de ces lésions (Castaing et coll., 1995). En 2000, Kamashev et Rouvière-Yaniv ont proposé un modèle de la reconnaissance des ADN double brin comportant une coupure simple brin ou une extrémité 3' sortant par la protéine HU (Kamashev et Rouvière-Yaniv, 2000). Les résultats de cette étude montrent que la protéine HU reconnaît une partie ADN double brin à l'aide de la pince formée par des brins  $\beta$  et une partie double brin ou simple brin à l'aide du « corps » d'hélices  $\alpha$  (Figure 14).



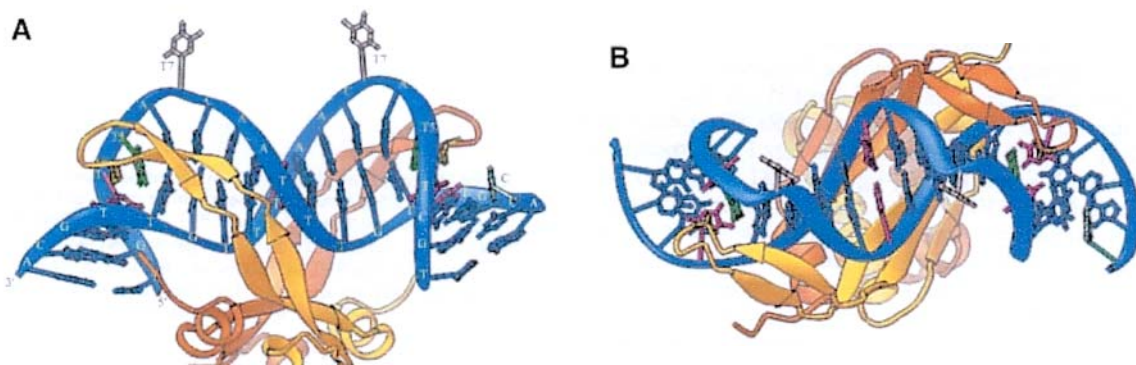
**Figure 14 : modèle de la fixation de HU sur des structures particulières.**

ADN possédant une coupure simple brin (A) ou une extrémité 3' sortant (B). (d'après Kamashev et Rouvière-Yaniv, 2000).

Le mode de dénaturation de la protéine HU a été étudié par la méthode de « differential scan display » (Ramstein et coll., 2003). Les résultats de cette étude montrent



que la protéine HU d'*E. coli* se dénature en deux étapes ( $N_2 \leftrightarrow I_2 \leftrightarrow 2D$  ;  $N_2$  : dimère natif ;  $I_2$  : intermédiaire dimérique ;  $2D$  : deux monomères dénaturés) alors que les protéines HU des autres organismes se dénaturent par un mécanisme à une seule étape ( $N_2 \leftrightarrow 2D$ ). Le passage de l'état  $N_2$  à  $I_2$  est dépendant de la température et de la force ionique du milieu. L'analyse par dichroïsme circulaire montre que l'intermédiaire  $I_2$  possède 50 % d'hélicité  $\alpha$  de moins que le dimère natif  $N_2$ . De plus, les auteurs ont montré que le profil de dénaturation des différents dimères de HU ( $\alpha_2$ ,  $\alpha\beta$ ,  $\beta_2$ ) n'est pas identique. Ainsi, dans les conditions physiologiques, on trouve environ 60 % de dimères  $\alpha_2$  et  $\alpha\beta$  natifs alors que le dimère  $\beta_2$  est principalement sous la forme intermédiaire (85 %). Les expériences réalisées avec l'intermédiaire  $I_2$  ont montré qu'il n'est plus capable de reconnaître avec une haute affinité les ADN comportant une lésion simple brin. Les auteurs avancent deux hypothèses quant au(x) rôle(s) physiologique(s) de  $I_2$  : 1) la forme  $I_2$  étant plus flexible que la forme  $N_2$ , elle pourrait faciliter l'échange de chaînes entre les deux homodimères  $\alpha_2$  et  $\beta_2$ , permettant ainsi la synthèse *de novo* de l'hétérodimère  $\alpha\beta$  ; 2) dans la conformation  $I_2$ , la flexibilité existant à l'interface des deux monomères constituant un homodimère  $\beta_2$  pourrait expliquer son implication dans la réponse au choc froid.



**Figure 15 : structure RMN d'un dimère de HU (*Anabaena*) fixé à l'ADN.**

Chaque dimère est représenté d'une couleur différente. A : vue de profil. B : vue de dessus (d'après Swinger et coll., 2003).

Bien que la structure cristallographique de la protéine HU de *B. stearothermophilus* soit connue depuis longtemps (Tanaka et coll., 1984), ce n'est que récemment que la structure du complexe a été déterminée (Swinger et coll., 2003 ; Figure 15). Pour déterminer cette structure, les auteurs ont utilisé la protéine HU issue de la cyanobactérie *Anabaena*. Les résultats montrent que dans le complexe HU/ADN il y a une intercalation de deux prolines entre les plateaux de paires de bases (chacune localisée à l'extrémité des « bras » composés de deux feuilletts  $\beta$  antiparallèles). Cette intercalation a lieu via le petit sillon de l'ADN. A ce niveau, la taille de la zone de contact HU/ADN est de 9 pb. A chaque site d'intercalation des prolines l'ADN forme un coude (zone où l'empilement des plateaux de paires de bases est détruit). Au niveau de ces coudes, l'ADN est légèrement sous enroulé par rapport à un ADN de forme B canonique. En revanche la région d'ADN compris entre ces deux coudes est plutôt surenroulée. L'interaction HU/ADN est ensuite stabilisée par des contacts au niveau du « corps » de la protéine (composé de trois hélices  $\alpha$ ). La taille totale du site de fixation de HU sur l'ADN est d'environ 37 pb (interaction au niveau du « bras » plus interaction au niveau du « corps »). Comme la fixation de la protéine HU provoque une courbure de l'ADN variant entre 105 et 140°, on peut dire que l'ADN est flexible au sein du complexe. Par ailleurs, les données cristallographiques ont montré que la courbure de l'ADN n'a pas lieu dans un plan.

#### **4) La protéine IHF :**

La protéine IHF (Integration Host Factor) a d'abord été isolée comme étant un facteur permettant l'insertion site spécifique du phage lambda dans le génome de son hôte. IHF se compose de deux sous-unités notées  $\alpha$  et  $\beta$  qui possèdent une homologie de 30 % de l'une par rapport à l'autre. La sous-unité  $\alpha$  (11 kDa) est codée par le gène *himA* et la sous-unité  $\beta$  (9 kDa) par le gène *hip*. *In vivo*, la protéine IHF est impliquée dans des phénomènes comme l'inversion de gène, la réplication, le transfert de plasmide ou le contrôle de la transposition (Friedman et coll., 1988).

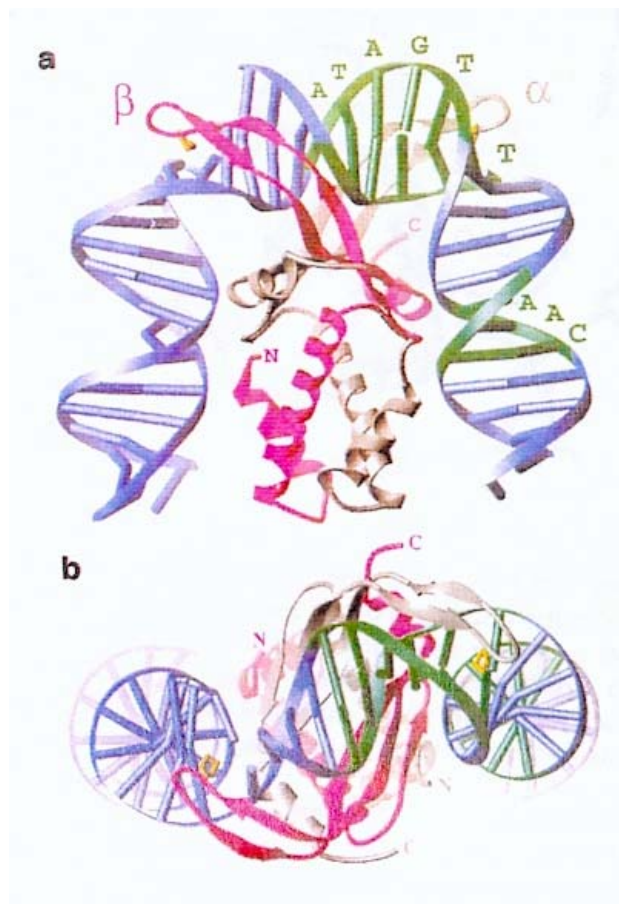
En 1984, Craig et Nash ont débuté l'étude de l'interaction IHF/ADN par des méthodes d'empreintes moléculaires à la DNase I et au DMS (Craig et Nash, 1984). Les résultats de ces travaux ont permis de montrer que la protéine IHF a des sites préférentiels

de fixation. Certains de ces sites, d'une longueur de 30 à 35 pb, sont localisés dans la région d'attachement du phage lambda à son hôte (site attP). Des études ont montré que la constante d'équilibre de dissociation ( $K_d$ ) des complexes spécifiques est 1000 fois plus grande que celle des complexes non spécifiques (Mengeritsky et coll., 1993 ; Yang et Nash, 1994, Wang et coll., 1995).

La fixation de la protéine IHF sur l'ADN est dépendante à la fois de la présence de la séquence consensus (WATCAANNNTTR ; W=A ou T, R=A ou G, N=A, T, G, C) et des séquences voisines. L'observation des séquences des sites naturels de fixation de la protéine IHF a montré que du côté 5' de certains sites de fixation on trouve une région riche en AT. Ainsi, le fragment attP comporte un site possédant une région riche en AT (noté H') et deux sites sans région riche en AT (noté H1 et H2). Des études ont été réalisées pour déterminer le rôle des répétitions AT en 5' du site de fixation (Hales et coll., 1996). Dans ces études, les auteurs ont observé que l'ajout d'une séquence riche en AT en 5' du site H1 provoque une augmentation de l'affinité de IHF et une courbure induite plus importante. En revanche, la capacité à recombinaison de l'ADN est diminuée. Toujours dans le but d'étudier l'importance de la région localisée en 5' du site de fixation spécifique, des études par SELEX ont été menées (Goodman et coll., 1999). Pour cela, la région riche en AT du site H' est remplacée par 13 pb aléatoires. La détermination des constantes d'équilibre de dissociation des différents complexes réalisés à partir des produits issus du SELEX fait apparaître des différences d'affinité variant d'un facteur 100. Contrairement à toute attente, les séquences isolées par SELEX ne possèdent pas de régions riches en AT. Par ailleurs, aucun lien n'a pu être fait entre l'affinité de la protéine pour une séquence et la capacité de cette séquence à favoriser la recombinaison de l'ADN par insertion ou par excision.

Dans un premier temps, la structure des complexes IHF/ADN a été étudiée par des méthodes d'empreintes moléculaires. Ainsi, les résultats des empreintes aux radicaux OH• et au DMS ont montré que la protéine se fixe via le petit sillon de l'ADN (Yang et Nash, 1989). En utilisant la pluramicyne comme sonde chimique, une autre équipe a montré que la fixation de la protéine IHF provoque un déroulement de l'ADN de part et d'autre du site de fixation (Sun et coll., 1996). Dans les mêmes travaux, les auteurs montrent à l'aide d'empreintes réalisées au permanganate de potassium ( $KMnO_4$ ) que, malgré la forte courbure de l'ADN, la fixation de la protéine IHF ne provoque pas d'ouverture de la double hélice d'ADN. La même année, le groupe de Nash a résolu la structure des

complexes IHF/ADN par cristallographie aux rayons X (Rice et coll., 1996 ; Figure 16). Cette structure confirme que la protéine IHF se fixe à l'ADN exclusivement via son petit sillon en provoquant une courbure importante de l'ADN (Robertson et coll., 1988). La valeur de l'angle de courbure de l'ADN dans le cristal ( $160^\circ$ ) est proche de celle obtenue par microscopie électronique et par des méthodes de gel retard ( $140^\circ$  ; Kosturko et coll., 1989 ; Thompson et coll., 1988). La structure des complexes IHF/ADN est donc très proche de celle des complexes HU/ADN. On peut cependant noter que la courbure induite par IHF est supérieure à celle induite par HU ( $105\text{-}140^\circ$  ; cf. partie 3). Cette forte courbure donne une forme caractéristique en U aux complexes IHF/ADN.



**Figure 16 : structure cristallographique d'un dimère de IHF fixé à l'ADN.**

Les prolines impliquées dans l'interaction sont repérées en jaune ; la séquence consensus de fixation est notée en vert. a : vue de profil. b : vue de dessus. (d'après Rice et coll., 1996).

## 5) La protéine H-NS :

La protéine H-NS (Histone like Nucleoid Structural protein) a tout d'abord été isolée comme facteur de transcription bactérien. Cette petite protéine de 137 acides aminés (15.4 kDa) est fortement conservée chez les entérobactéries. Elle est très abondante dans la cellule (entre 10000 et 20000 copies suivant le stade de croissance). Dans les entérobactéries, on trouve une protéine possédant 58 % d'identité avec H-NS. Cette protéine, nommée StpA, comporte un grand nombre de similitudes avec la protéine H-NS tant au niveau structural qu'au niveau fonctionnel. Pourtant, les expériences ont montré que l'intégralité des phénotypes mutants due à la délétion de H-NS ne peut pas être compensée par StpA (Free et Dorman, 1997). Les protéines H-NS et StpA possèdent donc un certain nombre de fonctions distinctes. On peut citer comme exemple le rôle spécifique de StpA dans l'appariement de chaînes d'ARN complémentaires (Sondén et Uhlin, 1996 ; Cusick et Belfort, 1998).

La protéine H-NS joue un rôle central dans l'expression des gènes d'*E. coli*. En effet, des études génétiques ont montré que la délétion de la protéine H-NS affecte l'expression de 5 % des gènes (Hommiais et coll., 2001). Dans la majorité des cas, les gènes régulés par H-NS sont impliqués dans les phénomènes de virulences (Prosseda et coll., 1998 ; Nye et coll., 2000) ou d'adaptation aux changements environnementaux (Brandi et coll., 1999). On peut aussi noter que la protéine H-NS régule sa propre expression (Spurio et coll., 1997).

Pour réguler l'expression des gènes, la protéine agit au niveau de la transcription selon deux mécanismes : 1) action directe au niveau des promoteurs ; 2) « silencing ». Le « silencing » est un mécanisme d'action à distance dans lequel la fixation préférentielle en un site relativement éloigné du promoteur à réguler (200 à 500 pb) provoque un changement global de la structure de l'ADN. Ce changement global permet l'inhibition de l'initiation de la transcription au niveau du promoteur régulé. Des études ont montré que la distance maximale d'action du mécanisme de « silencing » est d'environ 800 pb (Fletcher et Csonka, 1995).

La détermination de l'état d'oligomérisation de la protéine H-NS n'est pas un problème simple. En effet, les données thermodynamiques ont montré que, suivant les conditions expérimentales, la protéine H-NS s'assemble soit en tétramère soit en dimère (Ceschini et coll., 2000). Ainsi, lorsque la concentration en protéine est supérieure à  $10^{-7}$  M,

la forme tétramérique est majoritaire. La présence de cations monovalents, quant à elle, favorise la formation des dimères de H-NS. Une autre étude réalisée par le groupe de Kochoyan a montré qu'en solution la forme monomérique de H-NS est fortement défavorisée du fait de l'exposition au solvant de 12 résidus hydrophobes (Bloch et coll., 2003)

La protéine H-NS reconnaît préférentiellement les séquences d'ADN qui sont intrinsèquement courbes ou riches en dA+dT (Dame et coll., 2001 ; Rimsky et coll., 2001). Ces régions sont souvent localisées à proximité des promoteurs des gènes qui sont régulés par H-NS (Tanaka et coll., 1991). Des empreintes moléculaires au DMS ont été réalisées (Tippner et coll., 1994 ; Afflerbach et coll., 1999). Les résultats de ces empreintes ont montré que la protéine H-NS se fixe sur une face de l'ADN via son grand sillon. Ces mêmes études ont aussi permis de mettre en évidence une expansion et une distorsion du petit sillon opposé.

Des expériences de permutation circulaire ont montré que la fixation de la protéine H-NS provoque une courbure de l'ADN (Spurio et coll., 1997). Les travaux de Tupper et collaborateurs ont montré que lorsque la concentration en protéine H-NS est faible, la fixation de la protéine sur la cible provoque des supertours négatifs de l'ADN (Tupper et coll., 1994).

Dans le but de comprendre le mode de fixation de la protéine H-NS sur sa cible, des études visant à déterminer la structure de la protéine H-NS ont été réalisées. Il faut cependant noter qu'actuellement la structure de la protéine H-NS n'a pas encore été déterminée ni par cristallographie ni par RMN. Les données bibliographiques portent donc soit sur la partie N terminale soit sur la partie C terminale de la protéine.

En 1996, Ueguchi et Mizuno ont obtenu des mutants comportant différentes délétions sur la protéine H-NS (Ueguchi et Mizuno, 1996). Ces mutants ont permis de mettre en évidence l'implication du domaine C terminal de H-NS dans la reconnaissance protéine/ADN. Ces travaux ont aussi permis de montrer que la partie N terminale (résidus 1 à 15) est impliquée dans les phénomènes de répression de la transcription et que la partie centrale (résidus 16 à 90) joue un rôle dans l'oligomérisation de la protéine.

En 1995, Shindo et collaborateurs ont déterminé la structure du domaine C terminal (résidus 91 à 137) par RMN (Figure 17 (B) ; Shindo et coll., 1995). Les résultats montrent que cette région de la protéine H-NS est constituée de deux brins  $\beta$  antiparallèles, d'une

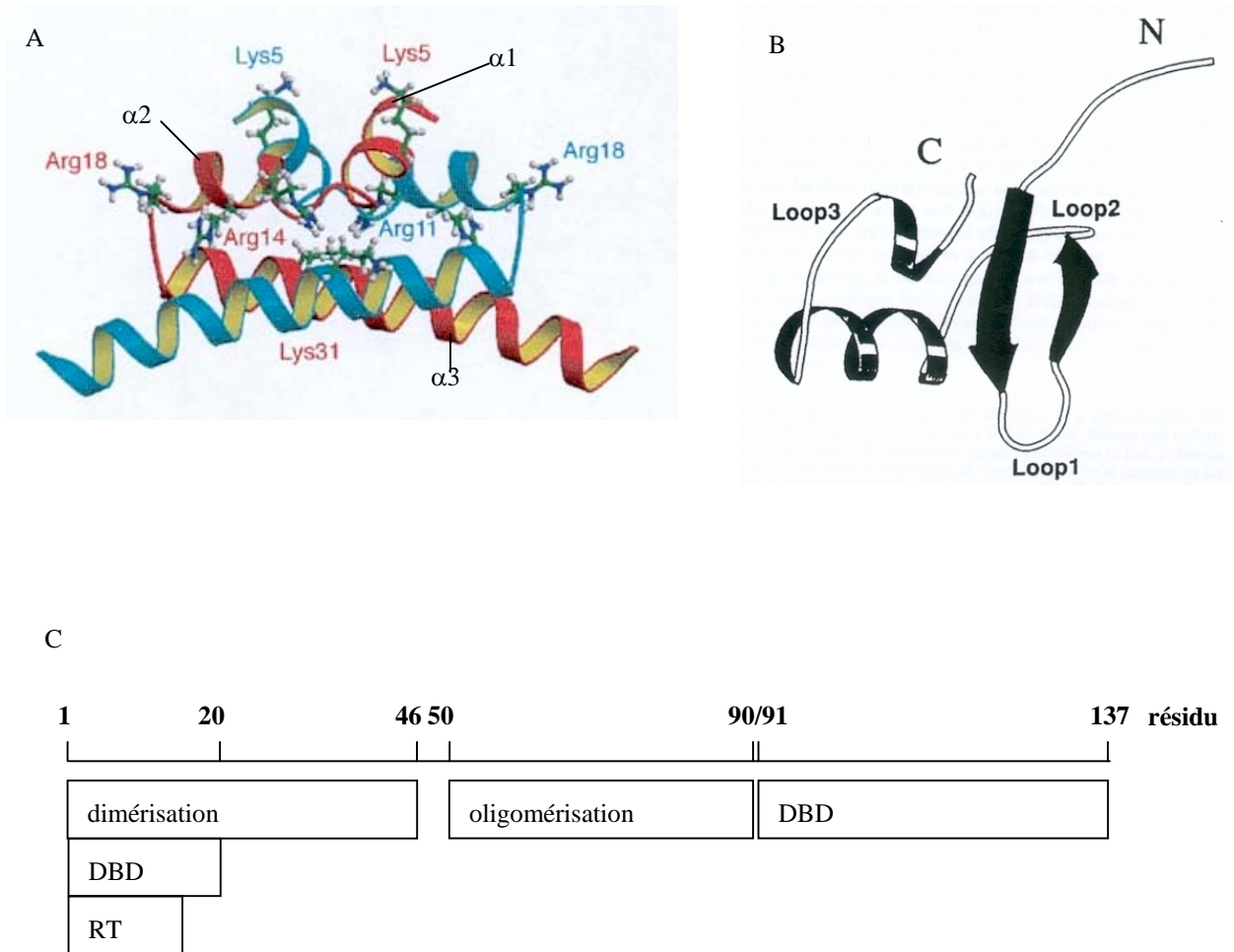
hélice  $\alpha$  et d'une hélice  $3_{10}$ . L'arrangement tridimensionnel de ces structures fait apparaître un nouveau motif de fixation à l'ADN. Dans la même étude, les auteurs montrent par des expériences de gel retard que le peptide constitué des résidus 91 à 137 est capable de se fixer à l'ADN mais avec une affinité moins forte que celle de la protéine complète.

En 1997, l'équipe de Mizuno complète les travaux publiés en 1996 et montre par des expériences similaires que les résidus 21 à 63 sont impliqués dans le processus de dimérisation de H-NS (Ueguchi et coll., 1997).

En 1996, Williams et collaborateurs montrent grâce à l'utilisation de mutants que la partie N terminale de la protéine H-NS (résidus 1 à 64) est impliquée dans le processus de multimérisation (Williams et coll., 1996). La structure de la partie N terminale (résidus 1 à 46) de la protéine H-NS est obtenue par RMN (Bloch et coll., 2003). Ces résultats font apparaître un nouveau motif de dimérisation. Ce motif est composé de 3 hélices  $\alpha$ . Les hélices  $\alpha_2$  et  $\alpha_3$  forment un U, l'hélice  $\alpha_1$  étant perpendiculaire aux deux autres et dans un même plan (Figure 17 (A)).

Un modèle de la fixation de H-NS sur l'ADN a été proposé par le groupe de Rimsky (Badaut et coll., 2002). Ce modèle se décompose en deux étapes : 1) fixation d'un dimère de H-NS sur une séquence préférentiellement reconnue (étape de nucléation) ; 2) recrutement d'autres protéines H-NS qui sont fixées ailleurs dans le génome (notamment au niveau des boîtes Pribnow ; étape de polymérisation).

Suite à leurs travaux, Bloch et collaborateurs précisent le modèle précédemment décrit (Bloch et coll., 2003). D'après cette étude, deux monomères s'associent par des interactions hydrophobes au niveau de leur partie N terminale pour former un dimère. Le dimère va ensuite se lier à l'ADN par sa partie C terminale et par les résidus 1 à 20. A la suite de la fixation à l'ADN, la région centrale flexible (résidus 46 à 90) va se structurer. Cette nouvelle organisation de la protéine va permettre aux dimères présents sur l'ADN d'interagir via leur partie centrale pour que la protéine se tétramérise. Cette interaction entre les protéines fixées à l'ADN pourrait jouer un rôle dans le processus de courbure de l'ADN.



**Figure 17 : structure de la protéine H-NS.**

A : structure RMN du domaine N terminal (résidus 1 à 46) d'après Bloch et coll., 2003. B : structure RMN du domaine C terminal (résidus 91 à 137) d'après Shindo et coll., 1995. C : représentation schématique du rôle des différents domaines de la protéine H-NS. DBD : domaine de fixation à l'ADN ; RT : répression de la transcription.

La Figure 17 C représente de façon schématique l'état actuel de nos connaissances sur la localisation et le rôle des différents domaines de H-NS. Des études complémentaires sont nécessaires pour déterminer de façon précise les facteurs qui déterminent le mode de fixation de la protéine sur l'ADN ainsi que l'état d'oligomérisation de la protéine en solution.



## **OBJECTIF DU TRAVAIL DE THESE**

Des travaux précédemment réalisés au laboratoire avaient permis d'isoler, de manière fortuite, une séquence préférentielle de fixation de la protéine MC1 (site I). Dans la continuité de ces travaux, l'objectif de ma thèse a été double :

- isoler, par une méthode indépendante des résultats préalablement obtenus, l'ensemble des séquences préférentiellement reconnues par MC1.
- utiliser ces séquences de forte affinité pour caractériser le mode de fixation de la protéine sur l'ADN.

Ces expériences pouvaient nous permettre de répondre à différentes questions : quelle est l'affinité de la protéine pour ses sites préférentiels ? ; la protéine « lit-elle » de manière directe ou indirecte la séquence d'ADN ? ; quelle est la nature des acides aminés intervenant dans l'interaction ?

Dans le but d'isoler les séquences de haute affinité pour la protéine MC1 de *Methanosarcina thermophila*, la méthode SELEX (Systematic Evolution of Ligands by EXponential enrichment) a été mise en œuvre. Cette technique qui comprend une étape de sélection de l'ADN par MC1 et une étape d'amplification des séquences sélectionnées est décrite dans la première partie de la thèse.

La détermination des constantes physico-chimiques de l'interaction MC1/séquences issues de la sélection a été réalisée par des méthodes utilisant les gels retard : mesure de vitesse de dissociation, mesure de constante d'équilibre (première partie).

Pour caractériser l'interaction d'un point de vue moléculaire, différentes méthodes ont été utilisées :

- le mode de fixation de la protéine a été étudié par des méthodes d'empreintes moléculaires au DMS et au radicaux  $\text{OH}^\bullet$  (deuxième partie).
- à défaut de pouvoir réaliser des expériences de mutagenèse dirigée (la protéine n'est pas clonée), nous avons utilisé les rayonnements ionisants pour modifier sélectivement certains acides aminés et en déduire leur implication dans l'interaction (dernière partie).

## **MATERIELS ET METHODES**

## 1) Préparation de la protéine MC1 :

La souche *Methanosarcina thermophila* (encore appelée *Methanosarcuna sp CHTI 55*) d'où est issue la protéine MC1, est fournie par J.P. Touzel (Station de Technologie Alimentaire de l'INRA, Villeneuve d'Ascq). Le milieu utilisé pour faire pousser les bactéries (milieu BCYT) est riche en acétate (Touzel et coll., 1983). La croissance des bactéries s'effectue à 55°C, pH 7 sous mélange N<sub>2</sub>+CO<sub>2</sub> (Touzel et coll., 1985).

La protéine MC1 est préparée au laboratoire par A. Gervais essentiellement par chromatographie d'affinité sur colonne d'ADN/cellulose.

Après rupture des parois bactériennes à la presse de French, les complexes ADN/protéines sont dissociés par élévation de la force ionique à 2 M NaCl. Les archaeobactéries sont centrifugées pendant 1 h à 10000 g. Après élimination des débris cellulaires, les acides nucléiques sont précipités par le polyéthylène glycol puis centrifugés 1 h à 10000 g. La force ionique du surnageant contenant les protéines est abaissée à 50 mM par dialyse. Le surnageant est ensuite chargé sur une colonne ADN/cellulose et les protéines sont éluées par un gradient de NaCl. La fraction enrichie en protéine MC1 est purifiée et concentrée sur une colonne échangeuse d'ions de sulfopropyl à flux rapide.

La concentration de la protéine MC1 est mesurée par spectrométrie d'absorption à 280 nm, avec un coefficient d'extinction molaire de 11000 M<sup>-1</sup>.cm<sup>-1</sup>. Elle est conservée dans du tampon phosphate 10 mM pH 6.8 à une concentration de 5x10<sup>-4</sup> M et stockée à -20°C. Les dilutions de protéine sont réalisées soit dans du TE soit dans du TE/NaCl 150 ou 60 mM. Les dilutions de concentration inférieure à 10<sup>-5</sup> M ne sont pas conservées.

## 2) Isolement des séquences par la méthode SELEX :

### 2.1) Purification des oligonucléotides :

Pour réaliser les expériences de SELEX, nous avons utilisé des oligonucléotides synthétisés par la société Eurogentec. Dans un premier temps les oligonucléotides sont purifiés sur gel dénaturant. Pour cela, l'échantillon d'ADN est repris dans du tampon dénaturant (formamide/TE/bleu de bromophénol ; 50/50/traces), chauffé pendant 3 min à 95°C puis chargé sur gel d'électrophorèse dénaturant (polyacrylamide (R=19/1) 12 ou 15 % suivant la taille des oligonucléotides, 7 M urée, dans du tampon TBE). La migration

a lieu pendant 7 h à la température ambiante sous 24 V/cm. Après électrophorèse, la bande d'ADN majoritaire est découpée sous UV par la méthode d'ombrage (UV shadowing). L'ADN est ensuite élué du gel par incubation dans du tampon Tris HCl (pH 7.4), 20 mM ; EDTA (pH 8), 1 mM ; NaCl, 20 ou 100 mM (suivant la taille du fragment) pendant la nuit à 55°C. La fin de la purification a lieu par passage de la solution d'ADN sur une colonne échangeuse d'ions (colonne Elutip ; Schleicher & Schuell) selon les recommandations du fabricant. Les oligonucléotides sont alors repris dans du TE.

## 2.2) Synthèse de la librairie aléatoire d'ADN double brin :

Pour synthétiser les fragments d'ADN double brin, l'oligonucléotide 65N<sub>15</sub> (5' CAGGTCAGTTCAGCGGATCCTGTCG-(N)<sub>15</sub>.GAGGCGAATTCAGTGCAACTGCAGC 3' ; (N)<sub>15</sub> séquence aléatoire de 15 nucléotides) est hybridé avec l'oligonucléotide Rev (5' GCTGCAGTTGCACTGAATTCGCCTC 3'). Pour cela, une solution contenant 5.4 µM de chaque oligonucléotide est chauffée à 95°C pendant 3 min (tampon : Tris-HCl (pH 7.2), 50 mM ; MgSO<sub>4</sub>, 10 mM ; DTT, 0.1 mM) puis laissée à refroidir doucement jusqu'à la température ambiante. Par la suite, la polymérase de Klenow (0.3 U.µL<sup>-1</sup>) et 10 mM de chaque dNTP sont ajoutés à la solution d'ADN préparée précédemment. La polymérisation du fragment double brin a lieu pendant 2 h à 37°C. Afin de purifier les ADN double brin de 65 pb, la solution de polymérisation est chargée sur un gel de polyacrylamide (R=19/1) 12 % après ajout de 1/10 de volume de tampon de charge (glycérol/TE/bleu de bromophénol ; 50/50/traces). La migration a lieu pendant 5h sous 14V/cm dans du tampon TBE. La bande d'ADN majoritaire (ADN 65pbN<sub>15</sub>) est ensuite découpée sous UV, éluée du gel et l'ADN est purifié sur colonne Elutip comme cela a déjà été décrit dans la partie 2.1).

## 2.3) Sélection et purification de fragments de haute affinité pour la protéine MC1 :

Pour isoler les séquences d'ADN de forte affinité pour la protéine MC1, la librairie d'ADN double brin comportant une région aléatoire de 15 pb (65pbN<sub>15</sub>) est criblée par la protéine MC1.

Pour cela, les complexes MC1/ADN sont réalisés par incubation (30 min, température ambiante) dans du tampon TE/NaCl 150 mM de 50 nM d'ADN 65pbN<sub>15</sub> radiomarqué en 5' avec la protéine MC1. La quantité de protéine MC1 est ajustée de façon

à obtenir 5 à 10 % de sélection à chaque cycle. Le taux de sélection est égal à la fraction d'ADN présente dans les complexes. Les complexes sont ensuite purifiés de l'ADN libre par la méthode de gel retard. Pour cela, la solution d'ADN contenant les complexes est chargées sur un gel de polyacrylamide (R=19/1) 8 % et la migration a lieu à température ambiante pendant 5 h sous 14V/cm dans du tampon TBE. Le gel est ensuite autoradiographié et la bande retardée (contenant les complexes MC1/ADN de forte affinité) est découpée. L'ADN est élué du gel et purifié à l'aide d'une colonne Elutip comme cela a déjà été décrit dans la partie 2.1).

#### 2.4) Amplification par PCR des ADN de forte affinité :

Après sélection et purification, les ADN possédant une forte affinité pour la protéine MC1 sont amplifiés. Cette étape d'amplification permet d'obtenir le matériel suffisant pour réaliser un nouveau cycle de sélection.

L'amplification des fragments 65pbN<sub>15</sub> (5' CAGGTCAGTTCAGCGGATC CTGTTCG-(N)<sub>15</sub>.GAGGCGAATTCAGTGCAACTGCAGC 3' ; (N)<sub>15</sub> séquence aléatoire de 15 nucléotides) est réalisée par PCR avec les amorces suivantes : Fwd 5' CAGGTCAGTTCAGCGGATCCTGTTCG 3' (le site de coupure par Bam HI est souligné) ; Rev 5' GCTGCAGTTGCACTGAATTCGCCTC 3' (le site de coupure par Eco RI est souligné). La réaction de PCR a lieu dans les conditions suivantes : Tris-HCl (pH 8.8), 20 mM ; KCl, 10 mM ; (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 10 mM ; MgSO<sub>4</sub>, 2 mM ; Sérum Albumine Bovine (SAB), 0.1 mg/ml ; triton X-100, 0.1 % ; dNTP's, 0.2 mM de chaque ; Pfu DNA polymérase, 0.03 U.µl<sup>-1</sup> ; matrice, 2 nM ; amorces, 0.2 µM de chaque. Après une dénaturation initiale de 5 min à 95°C, on réalise 20 fois le cycle PCR suivant : deshybridation 1 min à 94°C, polymérisation 1 min à 65°C, extension 2 min à 72°C.

Après la PCR, la Pfu DNA polymerase est éliminée de l'ADN par une extraction au phénol/chloroforme. Les traces de phénol sont éliminées par plusieurs lavages à l'éther. L'ADN est ensuite précipité à l'éthanol en présence d'acétate de sodium (0.3 M). Après centrifugation et lavage à l'éthanol 70 %, le culot d'ADN est repris dans 10 µl de TE. L'ADN est purifié de la même façon que celle décrite dans le paragraphe 2.1) mais on utilise des conditions de gel non dénaturantes : tampon de charge glycérol/TE/bleu de bromophénol (50/50/traces) ; gel de polyacrylamide (R=19/1), 8 %. La migration a lieu pendant 4 h sous 14 V/cm à la température ambiante dans du tampon TBE).

### 2.5) Marquage radioactif :

Les ADN purifiés issus de la PCR ne sont pas phosphorylés donc ils peuvent être radiomarqués directement avec du [ $\gamma^{32}\text{P}$ ]-ATP (activité spécifique : 5000 Ci.mmol<sup>-1</sup>). Pour réaliser la phosphorylation, la T<sub>4</sub> polynucléotides kinase (0.12 U.μL<sup>-1</sup>) et 0.5 μM de [ $\gamma^{32}\text{P}$ ]-ATP sont mis en présence de l'ADN (1 μM) pendant 1 h à 37°C (tampon : Tris-HCl (pH 7.6), 5 mM ; MgCl<sub>2</sub>, 1 mM ; DTT, 0.5 mM ; spermidine-HCl, 0.01 mM ; EDTA (pH 8), 0.01 mM). Pour augmenter le nombre de fragments phosphorylés, un excès d'ATP froid est ajouté. La solution est incubée une seconde fois à 37°C pendant 15 min. L'ATP non incorporé est éliminé par filtration grâce au système microcon YM-30 (amicon) puis l'ADN est repris soit dans du TE soit dans du TE/NaCl 150 mM.

## 3) Caractérisation des séquences issues du SELEX :

### 3.1) Clonage des séquences issues du SELEX :

En fin d'expérimentation SELEX, nous avons obtenu un mélange d'ADN possédant une forte affinité pour la protéine MC1. Afin d'isoler chacune des séquences sélectionnées par la protéine, il est nécessaire de réaliser un clonage.

Les ADN issus du 10<sup>ème</sup> cycle SELEX sont digérés 3 h à 37°C par les enzymes de restriction Eco RI et Bam HI dans un tampon contenant Tris-acétate (pH 7.9), 33 mM ; acétate de magnésium, 10 mM ; acétate de potassium, 66 mM ; SAB, 0.1 mg.mL<sup>-1</sup>. Parallèlement, du plasmide pGEM3Zf(-) est digéré dans les mêmes conditions par les mêmes enzymes de restriction et déphosphorylé par la Shrimp Alkaline Phosphatase (SAP) selon les recommandations du fournisseur (Promega). La ligation entre le fragment de 31 pb (444 fmol) et le plasmide pGEM3Zf(-) déphosphorylé (148 fmol) est réalisée par incubation pendant la nuit à 4°C dans 20 μL de tampon Tris-HCl (pH 7.6), 50 mM ; MgCl<sub>2</sub>, 10 mM ; ATP, 1 mM ; DTT, 1 mM ; PEG 8000, 5 %.

Les bactéries DH5α sont rendues compétentes par différents lavages dans du tampon CaCl<sub>2</sub>, 60 mM ; PIPES 10 mM ; glycérol 15 %. Elles sont transformées par choc thermique (10 min à 4°C puis 2 min à 42°C et retour à 4°C) par le fragment de 31 pb ligué au plasmide pGEM3Zf(-) (5 ng d'ADN pour 50 μL de bactéries compétentes). Les bactéries sont ensuite étalées sur des boîtes de culture contenant du milieu LB/agarose additionné d'ampicilline, d'IPTG et de bleu X-Gal (respectivement 100 μg.mL<sup>-1</sup> ; 0.5 mM ;

40  $\mu\text{g}.\text{ml}^{-1}$ ). Après une nuit de culture à 37°C, les colonies blanches sont repiquées en milieu liquide LB additionné d'ampicilline (100  $\mu\text{g}.\text{mL}^{-1}$ ) et incubées une nuit à 37°C sous agitation. L'ADN est extrait des bactéries à l'aide du kit NucleoSpin Plasmid de Macherey Nagel selon les recommandations du fournisseur. L'ADN est séché sous vide et séquencé par la société MWG-biotech.

### 3.2) Mesure de la vitesse de dissociation des complexes :

Les fragments d'ADN utilisés dans ces expériences, sont ceux qui sont obtenus par amplification par PCR des fragments clonés dans le plasmide pGEM3Zf(-). Les amorces utilisées pour cette amplification sont : 5' CGACTCACTATAGGGCGAAT 3' (65Fwd) et 5' GTCGACTCTAGAGGATCCTG 3' (65Rev). La séquence obtenue après la PCR est : CGACTCACTATAGGGCGAATTCGCCTC-CLONE $\alpha$ -CGACAGGATCCTCTAGAGTCGAC.

Dans le but de mesurer la vitesse de dissociation des complexes formés entre la protéine MC1 et les séquences amplifiées ci-dessus, 5 nM d'ADN radiomarqués en 5' sont mis en présence d'une quantité suffisante de protéine MC1 pour obtenir uniquement des complexes monoprotéiques. La réaction a lieu dans un tampon Tris-HCl (pH 7.4), 10 mM ; EDTA (pH 8), 1 mM ; NaCl, 60 ou 100 mM pendant 30 min à 4°C. Un excès de compétiteur froid (pUC) est ensuite ajouté à la réaction ci-dessus. La concentration finale de compétiteur froid est 350 fois plus grande que celle de l'ADN radiomarqué (en équivalent taille<sup>1</sup>). La dissociation a lieu à 4°C pendant différents temps. La réaction est stoppée par chargement sur gel de polyacrylamide (R=19/1) 8 %. La migration a lieu à 4 °C pendant 5 h sous 14 V/cm.

La constante de vitesse de dissociation est donnée par :

$$k_d = \ln 0.5/\Gamma$$

$k_d$  : vitesse de dissociation ;  $\Gamma$  : temps de demi-vie (s)

---

<sup>1</sup> Equivalent taille : le plasmide ayant une taille 49.8 fois plus grande que celle du fragment de 65 pb, 1 M de plasmide correspond à 49.8 M de fragment de 65 pb.



### 3.3) Expériences de compétition directe :

Dans le but de déterminer la constante d'affinité relative de la protéine MC1 pour la séquence consensus issue des expériences de SELEX et pour un ADN aléatoire, des expériences de compétition directe sont réalisées. Pour cela, les deux oligonucléotides suivants sont synthétisés : pour la fixation non préférentielle, GAGA(N<sub>13</sub>)GAG (20pbN<sub>13</sub> ; N<sub>13</sub> : séquence aléatoire de 13 pb) ; pour la fixation préférentielle, GCCTCAAAAACACACACCCCCGACAG (26pbC2). Les duplex radiomarqués en 5' (cf. partie 2.5)) sont mélangés en quantité équivalente ( $8 \times 10^{-12}$  M) et incubés pendant 20 min à 4°C avec des concentrations en protéine MC1 variant entre 0 et 235 nM. Cette réaction a lieu dans du tampon contenant NaCl, 0.1 M ; SAB, 15 µg/ml ; TE. Les complexes sont séparés de l'ADN libre par migration sur un gel de polyacrylamide (R=30/1), 8 % pendant 3 h à 4°C sous 14 V/cm.

La constante relative d'équilibre ( $K_{rel}$ ) est déterminée par la relation A:

$$K_{rel} = ([ADN]_{npréf} * [C]_{préf}) / ([ADN]_{préf} * [C]_{npréf}) \quad (A)$$

où [ADN] est la concentration de fragment libre préférentiel (préf) ou non (npréf) et [C] la concentration de complexe préférentiel (préf) ou non (npréf).

### 3.4) Mesure de la constante de dissociation des complexes MC1/séquences préférentielles :

Les expériences sont réalisées dans les mêmes conditions que dans la partie précédente mais les complexes sont formés soit avec l'ADN préférentiel (26pbC2) soit avec l'ADN non préférentiel (20pbN<sub>13</sub>).

## 4) Etude du mode de fixation de la protéine :

### 4.1) Préparation de l'ADN double brin de 59 pb radiomarqué sur un seul brin :

Pour limiter la radiolyse induite par le  $[\gamma^{32}P]$ -ATP, les conditions de marquage radioactif présentées en 2.5) ont été modifiées. L'ADN de 59 bases (55 pmol) contenant le brin supérieur de la séquence issue du clone 2 (5' GCAGTTGCACTGAATTCGCCTC-AAAAACACACACCCC-CGACAGGATCCGCTGAACTGAC 3') est mis en présence de 1 pmol de  $[\gamma^{32}P]$ -ATP (activité spécifique : 5000 Ci.mmol<sup>-1</sup>). La réaction de

phosphorylation en 5' a lieu pendant 10 min à 37°C dans du tampon Tris-HCl (pH 7.6), 70 mM ; MgCl<sub>2</sub>, 10 mM ; DTT, 5 mM et dans un volume final de 20 µL. L'ATP non incorporé est éliminé par tamisage moléculaire (colonne SigmaSpin Post-Reaction Purification Columns) selon les recommandations du fabricant (Sigma) et l'ADN est repris dans du TE.

L'ADN de 59 pb radiomarké en 5' sur un seul brin est réalisé par hybridation dans du TE/NaCl 150 mM du brin radiomarké en 5' (10<sup>-6</sup>M) avec du brin complémentaire non radiomarké (10<sup>-5</sup>M). Pour que l'hybridation ait lieu, la solution est chauffée à 95°C pendant 2 min puis laissée à refroidir lentement jusqu'à la température ambiante.

#### 4.2) Séquençage des purines :

Pour se repérer le long de la séquence d'ADN, un séquençage selon la méthode de Maxam et Gilbert est réalisé. Pour cela, l'ADN de 59 pb préparé en 4.1) (40 nM) est incubé avec 1 µL de glycogène et 50 µL d'acide formique (solution commerciale) pendant 5 min à 20°C. La réaction est stoppée par ajout de 180 µL de solution « hydrazine stop » (acétate de sodium (pH 7.0), 0.3 M ; EDTA (pH 8.0), 0.1 M). L'ADN est ensuite précipité à l'éthanol en présence d'acétate de sodium 0.3 M, centrifugé et lavé à l'éthanol 80 %. Les dommages aux bases induits par l'acide formique sont révélés par incubation pendant 30 min à 95°C avec 100 µL de pipéridine dilués au 1/10. La pipéridine est éliminée par lyophilisation et l'ADN repris dans 20 µL d'eau. Cette étape de lavage est exécutée trois fois puis l'ADN est séché et repris dans du tampon de charge dénaturant (formamide/eau/xylène cyanol, 50/50/traces).

#### 4.3) Empreintes au DMS :

Les complexes MC1/ADN sont obtenus en incubant pendant 20 min à 20°C 25 nM d'ADN de 59 pb (cf. partie 4.1)) avec différentes quantités de protéine (0 à 25 nM) dans du tampon TE/NaCl 60 mM ; cacodylate de sodium, 50 mM. L'ADN est méthylé par incubation avec du DMS (concentration finale : 0.5 %) pendant 30 s à 20°C. La réaction de méthylation est arrêtée par ajout de 1/3 de volume de solution stop (acétate de sodium, 1.5 M ; EDTA (pH 8.0), 1 mM ; β mércaptoéthanol, 1 M). A ce stade, l'ADN est purifié par extraction au phénol/chloroforme (cf. partie 2.4)), précipité à l'éthanol en présence d'acétate de sodium (0.3 M) et les bases méthylées sont révélées par traitement à la

pipéridine (voir le paragraphe 4.2) pour les détails). Les échantillons sont ensuite dénaturés pendant 3 min à 95°C puis chargés sur gel d'électrophorèse dénaturant (polyacrylamide (R=19/1) 8 % , 7 M urée, dans du tampon TBE). La migration s'effectue à 35 W pendant 1 h 30 à la température ambiante. Le traitement du gel après migration est décrit dans le paragraphe 4.7).

#### 4.4) Interférences au DMS :

L'ADN de 59 pb marqué sur le brin inférieur (cf. partie 4.1) ; 25 nM) est méthylé par incubation dans du DMS (0.25 % final) pendant 30 sec à 20°C. La réaction de méthylation est arrêtée par ajout de la solution stop (cf. partie 4.3)). L'ADN est purifié par extraction au phénol/chloroforme comme cela a déjà été décrit (cf. partie 2.4)). Les complexes sont réalisés par incubation pendant 20 min à 20°C de l'ADN méthylé précédemment (25 nM) avec différentes concentrations de protéine MC1 (24 et 40 nM). Les complexes sont ensuite séparés de l'ADN libre par la technique de gel retard comme cela est décrit dans le paragraphe 2.3). Les sites méthylés des ADN venant des fractions correspondant aux complexes ou à l'ADN libre sont révélés par traitement à la pipéridine (cf. partie 4.2) et chargés sur un gel de séquence comme cela est décrit dans la partie 4.3).

#### 4.5) Empreintes aux radicaux OH<sup>•</sup> :

Les complexes MC1/ADN sont obtenus en incubant 25 nM d'ADN de 59 pb (cf. partie 4.1)) avec différentes quantités de protéine (0 à 50 nM) dans du TE/NaCl 60 mM (volume final : 15 µL). Après 20 min d'incubation à 20°C, on ajoute simultanément 1.5 µL des solutions suivantes fraîchement préparées : 0.4 mM Fe/0.8 mM EDTA ; peroxyde d'hydrogène, 0.3 % ; ascorbate de sodium, 20 mM. Le mélange est incubé 2 min à 20°C puis la réaction est stoppée par ajout de 2.25 µL de thiourée 0.1 M et 0.45 µL d'EDTA 80 mM. La protéine est éliminée de l'ADN par extraction au phénol/chloroforme. Les traces de phénol sont enlevées par plusieurs extractions à l'éther. L'ADN est ensuite précipité à l'éthanol en présence d'acétate de sodium (0.3 M) et de 1 µL de glycogène. Après centrifugation, l'ADN est lavé avec de l'éthanol 80 % et les coupures sont révélées et visualisées de la même façon que celle décrite dans le paragraphe 4.3).

#### 4.6) Empreintes au KMnO<sub>4</sub> :

Les complexes MC1/ADN sont obtenus en incubant 25 nM d'ADN de 59 pb radiomarqués en 5' sur un brin (cf. partie 4.1)) avec différentes quantités de protéine (0 à 25 nM) dans du tampon TE/NaCl 60 mM ; cacodylate de sodium, 50 mM. Après incubation 20 min à 20°C, l'ADN est oxydé par ajout de KMnO<sub>4</sub> (2 mM). La réaction d'oxydation est arrêtée après incubation pendant 2 min à 20°C par l'ajout de β mercaptéthanol (1.1 M final). Les étapes de purification de l'ADN avant le gel de séquence ainsi que le gel de séquence sont les mêmes que celles décrites dans la partie 4.3).

#### 4.7) Traitement des gels d'empreinte après migration :

Après migration, les gels sont fixés par un bain d'éthanol/acide acétique/eau (20/10/70) pendant 15 min puis rincés dans un bain d'eau pendant 10 min. Les gels sont ensuite séchés sous vide puis densitométrés à l'aide d'un storm 860 (Molecular Dynamics) et quantifiés à l'aide du logiciel Image Quant 5.1 (Molecular Dynamics).

#### 4.8) Stœchiométrie de l'interaction MC1/séquence consensus :

Pour déterminer la stœchiométrie de l'interaction, une forte concentration (2 μM) de fragment de 26 pb contenant la séquence issue du clone 2 est incubée avec des concentrations de protéine MC1 variant de 0 à 5 μM. La réaction est incubée à 4°C pendant 20 min dans du tampon Tris-HCl (pH 7.4), 10 mM ; EDTA (pH 8), 1 mM ; NaCl, 60 mM ; glycérol 8 % ; traces de bleu de bromophénol. Les complexes sont ensuite chargés sur un gel de polyacrylamide (R=30/1), 8 % et la migration a lieu 4°C pendant 3 h sous 14 V/cm.

### **5) Etude de l'effet des rayonnements ionisants sur les fixations préférentielles de MC1 :**

#### 5.1) Préparation des fragments ADN :

Les ADN issus du 10<sup>e</sup> cycle SELEX sont digérés par les enzymes de restrictions Eco RI et Bam HI (Promega) pendant 4 h à 37°C dans du tampon Tris-acétate (pH 7.9), 33 mM ; acétate de magnésium, 10 mM ; acétate de potassium, 66 mM ; SAB, 0.1 mg/mL.

Les fragments de 31 pb résultants de cette digestion sont purifiés sur un gel de polyacrylamide (R=19/1) 8 % (migration pendant 2 h 30 dans du tampon TBE sous 14V/cm et à la température ambiante). L'ADN est extrait du gel et purifié sur une colonne Elutip comme cela a été décrit dans le paragraphe 2.1). L'ADN est ensuite déphosphorylé par la phosphatase SAP (Promega) selon les recommandations du fournisseur. Pour finir, l'ADN est radiomarqué au [ $\gamma^{32}\text{P}$ ]-ATP comme décrit dans la partie 2.5).

### 5.2) Irradiation de la protéine MC1 :

Une solution aqueuse de protéine MC1 est irradiée par des rayons  $\gamma$  de 0.6 MeV dans un irradiateur IBL437 contenant une source de  $^{137}\text{Cs}$  (CisBio International). Le débit de dose au point d'intérêt (maintenu à 0°C dans un bain d'eau glacée) est de 9 Gy.min<sup>-1</sup>. Ce débit de dose est mesuré par la dosimétrie de Frick. Sauf indication contraire, la protéine est irradiée à une concentration de 2  $\mu\text{M}$  (tampon : phosphate de sodium (pH 7.5), 1 mM ; NaCl, 60 mM) et la dose reçue est de 125 Gy.

### 5.3) Expériences de compétition pour la fixation sur des ADN linéaires :

Dans le but d'étudier l'effet des rayonnements  $\gamma$  sur l'interaction MC1/ADN linéaires, différentes quantités de protéine MC1 irradiée à 125 Gy (0 à 80 nM) ou native (5 à 20 nM) sont incubées pendant 15 min à 4°C dans un tampon contenant 5 nM du fragment de 31 pb préparé ci-dessus (cf. paragraphe 5.1)), 5 nM d'ADN radiomarqué de 65 pb avant les cycles SELEX (65pbS<sub>0</sub>), 150 mM de NaCl et 1 mM de phosphate de potassium pH 7.25. Après incubation, on ajoute 1/10 de volume de tampon de charge (glycérol/phosphate de potassium pH(7.25)/bleu de bromophénol/xylène cyanol, 50/50/traces/traces) et les échantillons sont chargés sur un gel de polyacrylamide (R=19/1) 6 % dans du tampon TBE. Après migration de 4 h à 4°C sous 14 V/cm, le gel est fixé pendant 10 min dans un bain éthanol/acide acétique/eau (10/10/80), séché, exposé puis scanné à l'aide du storm 860.

De la protéine MC1 irradiée à 300 Gy (concentration lors de l'irradiation : 13.6  $\mu\text{M}$ ) est utilisée pour réaliser le même type d'expérience que celle décrite ci-dessus. Dans cette expérience, la protéine est incubée avec un fragment préférentiel de 184 pb contenant le site I et un fragment non préférentiel de 225 pb. Le tampon d'incubation

contient NaCl, 100 mM ; ADN, 3.5 nM de chaque. Les concentrations de protéine utilisées sont précisées pour chacun des gels.

5.4) Expériences de compétition pour la fixation sur des ADN circulaires :

Les minicercles d'ADN sont préparés par incubation de l'ADN de 207 pb (0.1 µg/ml), possédant des extrémités cohésives AATT, avec la ligase T<sub>4</sub> du phage λ (invitrogen). L'ADN circulaire fermé est séparé de l'ADN linéaire et de l'ADN relâché par un gel préparatif de polyacrylamide 4 % (R=19/1). Le tampon de migration est composé de TBE auquel on a ajouté 0.6 µg/ml de bromure d'éthidium. Pour plus de précisions, voir Toulmé et collaborateurs, 1995.

La protéine irradiée à 125 Gy (cf. 5.2)) est incubée avec un mélange contenant une quantité équimolaire (1.5 nM) d'ADN circulaire de 207 pb et d'ADN linéaire de 207 pb. Ces ADN ne contiennent pas de séquence préférentielle de fixation pour MC1. L'incubation a lieu pendant 15 min à 4°C dans un tampon contenant 100 mM de NaCl et 1 mM de phosphate de potassium pH 7.25. L'électrophorèse sur gel de polyacrylamide est réalisé dans les mêmes conditions que celles décrites dans la partie 5.3).

**6) Récapitulatif des différents fragments utilisés :**

type	nom	séquence (5'-3')
fragments SELEX	65pbS0	CAGGTCAGTTCAGCGGATCCTGTGC-(N) <sub>15</sub> GAGGCGAATTCAGTGCAACTGCAGC
	S2-0	CAGGTCAGTTCAGCGGATCCTGTGC- NNNAAAAACACAC(C/A)CCC(C/A)NNN- AGGCGAATTCAGTGCAACTGCAGC
fragments préférentiels	26pbC2	GCCTCAAAAACACACACCCCCGACAG
	31 pb	AATTCGCCTC-CLONE2-CGACAG
	65 clone alpha	CGACTCACTATAGGGCGAATTCGCCTC-CLONEalpha- CGACAGGATCCTCTAGAGTCGAC
	64 pb	ACTGTAAAACCTTACTTCTGTGACTCCATGGTCTATATAAAAA AGAGCATCATCAGGCTTGTGA
184 pb	TCCGGATCTTTTCATGTATTCAGCTCTTAAAACCGGATCAGT CGCAAAGACTGTAAAACCTTACTTCTGTGACTCCATGGTCTA TATAAAAAAGAGCATCATCAGGCTTGTGAAACCTTTCCG CTTGTGTACCCAGGTGTATGGGGATCTTGAACCTGAGATA AAAAAGAGACAAGATCTTTCAG	
fragments non préférentiels	20pbN <sub>13</sub>	GAGA(N) <sub>13</sub> GAG
	52pbS2-0	CAGGTCAGTTCAGCGGATCCTGTGC- NNNAAAAACACAC(C/A)CCC(C/A)NNN-AGGCG
	225 pb	GTGGGATTATGCCCGGAATAATAGGGGAGTTATTCAGGAT AAGTCCATTTTCCGTAAAGTTTGCAAATCTCATTAGAATAGC TCCTTTTACACCCATAGTTTCAAGATCATTGAGGGTTTTATC AAGGACCTCTCTGTCAATTTACTCCCGGAATGATAACCATCG CTCCGTATACATCACAATGAGCACAGAAGTCCCGTAGAAC CTGTATAGACGCTTCAGGT
	207 pb	CCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACT GGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTA GCTCACTCATTAGGCACCCAGGCTTTACACTTTATGCTTC CGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTT CACACAGGAAACAGCTATGACCATGATTACGGATTCAGT G
fragments empreintes	59 b brin sup	GCAGTTGCACTGAATTCGCCTC-AAAAACACACACCCC- CGACAGGATCCGCTGAACTGAC
	59 b brin inf	CGTCAACGTGACTTAAGCGGAG-TTTTTGTGTGTGGGG- GCTGTCCTAGGCGACTTGACTG
amorces PCR	Fwd	CAGGTCAGTTCAGCGGATCCTGTGC
	Rev	GCTGCAGTTGCACTGAATTCGCCTC
	65Fwd	CGACTCACTATAGGGCGAAT
	65Rev	GTCGACTCTAGAGGATCCTG

## **RESULTATS**



# **ISOLEMENT ET CARACTERISATION DES SEQUENCES**

## 1) La méthode SELEX :

### 1.1) Généralités :

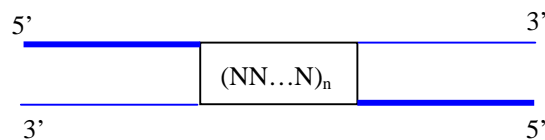
Les méthodes de sélection de sites *in vitro* permettent d'isoler, de façon systématique, des ADN ou des ARN possédant une capacité à interagir fortement et préférentiellement avec une cible donnée. Plusieurs méthodes similaires de sélection de sites préférentiels de fixation de protéines ont été décrites simultanément. Ainsi, le groupe de L. Gold (université du Colorado) décrit une méthode qu'il dénomme SELEX (Systematic Evolution of Ligands by EXponential enrichment) dans un travail visant à déterminer le mode de fixation de l'ADN polymérase du phage T4 sur son ARN messager (Tuerk et coll., 1990). Le groupe de W. Wright (Dallas) qui cherchait à caractériser la fixation de la myogénine sur l'ADN propose une méthode qu'il nomme CASTing (Cyclic Amplification and Selection of Targets) et détermine ainsi une séquence consensus de fixation pour ce facteur de transcription spécifique du tissu musculaire (Wright et coll., 1991). Un travail concernant les sites de fixation du facteur de transcription SP1 est également décrit en 1990 (Thiesen et Bach, 1990), la méthode est alors dénommée TDA (Target Detection Assay). Ces méthodes font toutes appel aux deux mêmes étapes de sélection et d'amplification successives.

Depuis ces travaux, la méthode SELEX (cette dénomination est actuellement la plus usitée) a été utilisée dans de nombreuses études et avec des molécules très différentes (petites molécules chimiques, peptides, protéines, acides nucléiques voire des cellules entières et des tissus). Cet outil très puissant est maintenant utilisé par de nombreuses firmes biotechnologiques pour des projets à visée thérapeutique ou de diagnostique.

En 1990, Ellington et Szostak créent le terme d'aptamère. Les aptamères sont des ARN ou des ADN isolés par un processus de sélection itératif et possédant une forte affinité pour une cible (ou capables de catalyser une réaction chimique). De manière inattendue, des aptamères possédant des affinités très importantes ( $K_d$  de l'ordre de  $10^{-10}$ - $10^{-12}$  M) ont été obtenus avec des protéines dont la fonction n'est pas d'interagir avec les acides nucléiques (Gold et coll., 1997). De façon tout aussi surprenante, des interactions

spécifiques ont été déterminées entre des ARN et des protéines connues pour interagir avec l'ADN (Cassiday et Maher, 2002).

Chaque fragment d'ADN utilisé pour la sélection comprend une partie centrale dont la séquence est engendrée de manière aléatoire. Cette séquence aléatoire est entourée de séquences identiques pour tous les fragments (Figure 18). Ces séquences servent pour les étapes de PCR et de clonage des fragments.



**Figure 18 : les différentes parties des fragments SELEX.**

Les parties en bleu sont identiques pour tous les fragments ; les traits bleus épais sont complémentaires des amorces de PCR. La partie en noir schématise la séquence aléatoire.

Lors de l'étape de sélection (Figure 19) la protéine (ou la cible d'intérêt) est utilisée pour isoler les fragments de plus forte affinité. Pour cela, une quantité suffisante de protéine pour fixer 5 à 10 % de la banque d'ADN est incubée avec les fragments. Les complexes sont ensuite séparés des ADN libres (étape de partition) et amplifiés par PCR (étape d'amplification).

Pour une expérience de sélection d'ARN les molécules sélectionnées par les protéines sont tout d'abord amplifiées par RT-PCR, puis la banque d'ADN ainsi obtenue est transformée en une nouvelle banque d'ARN par une étape de transcription *in vitro*.

L'étape de partition peut être réalisée de différentes façons : immunoprécipitation (Pollock et coll., 1990), fixation sur membrane de nitrocellulose (Tuerk et coll., 1990), colonne d'affinité (Ellington et Szostak, 1990 ; Chittenden et coll., 1991), gradient de sucrose (Lowary et Widom, 1998), gel retard (Blackwell et coll., 1990 ; Cui et coll., 1995) ou avec l'aide de billes modifiées (David-cordonnier et coll., 1998). Plus récemment des méthodes à haut débit ont été proposées. Ainsi Hao et collaborateurs (2003), proposent une

méthode utilisant la résonance plasmonique de surface où un cycle de sélection est effectué en moins d'une heure.

Deux catégories de banques de fragments peuvent être préparées. Le plus fréquemment, la séquence aléatoire est obtenue par synthèse chimique en utilisant le mélange des quatre bases pour chaque position. Une deuxième stratégie, plus complexe à mettre en œuvre, consiste à utiliser l'ensemble des séquences d'un génome (SELEX génomique). Le succès d'une expérience de SELEX génomique dépend alors fortement de la qualité de la banque de départ. Plusieurs méthodes permettant d'éviter toute sous ou sur représentation de certaines séquences ont été proposées (Singer et coll., 1997). De par leur origine, les séquences isolées par SELEX génomique ont une signification *in vivo*, cela n'est pas forcément le cas avec les séquences isolées par SELEX « non génomique ».

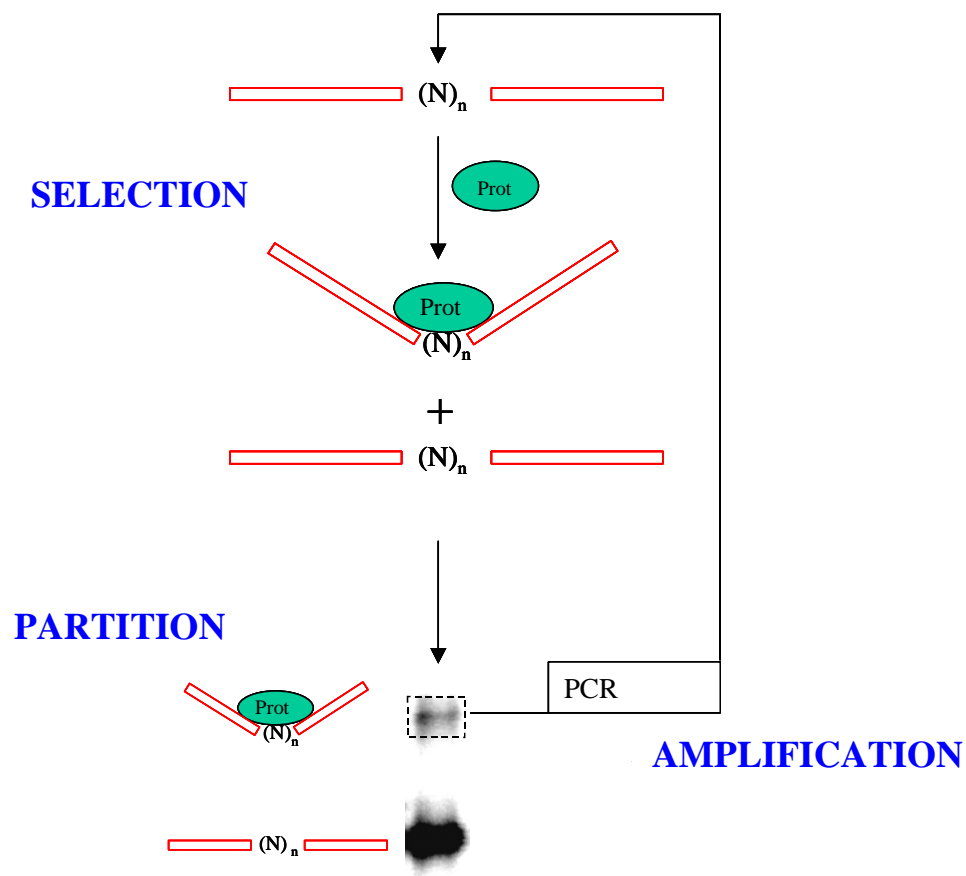


Figure 19 : récapitulatif des différentes étapes d'une expérience SELEX.

## 1.2) Facteurs influençant la méthode SELEX :

Du fait de sa performance, la méthode SELEX a suscité plusieurs études théoriques (Irvine et coll., 1991 ; Vant-Hull et coll., 1998). Ces études visent d'une part à déterminer des conditions expérimentales pour s'affranchir du risque de perte de séquences de haute affinité pour la cible et d'autre part à optimiser le nombre de cycles nécessaires.

Un des facteurs importants est **la longueur de la séquence aléatoire** car elle détermine la complexité (nombre de séquences différentes) de la librairie de fragments. On comprend aisément que plus le nombre de séquences différentes sera grand, plus le nombre de cycles à réaliser sera élevé. La longueur de la séquence aléatoire se situe le plus fréquemment entre 8 et 20 pb, ce qui représente, en théorie, entre  $4^8$  et  $4^{20}$  molécules différentes possibles au départ de la sélection. Si le nombre de séquences de haute affinité est faible, la sélection sera plus difficile. Ainsi lorsque L. Gold obtient deux séquences de forte affinité (60 fois plus) pour l'ADN polymérase du phage T4 à partir d'une librairie d'ARN composée de  $6 \times 10^4$  séquences différentes, il réalise la sélection de deux séquences ne représentant, dans la banque de départ, qu'une molécule sur 30000 (Tuerck et coll., 1990).

Dans la littérature, il est souvent mentionné que, pour que l'ensemble des séquences soit représenté dans la librairie de départ, la partie aléatoire ne doit pas dépasser 25 pb. En fait, cette longueur maximale de 25 pb n'est pas une valeur absolue. Le nombre de séquences différentes effectivement présentes dépend évidemment de la quantité de fragments utilisés. Ainsi, pour qu'une banque contenant une séquence aléatoire de 25 pb ( $4^{25}$  molécules différentes) soit complète, il faut une quantité initiale de fragment au moins égale à 2 nmol ( $4^{25} / 6 \times 10^{23}$ ). Pour un fragment de 75 pb, ces conditions correspondent à environ 10  $\mu$ g d'ADN (conditions difficilement compatibles avec la PCR). Dans des conditions standard de PCR (1 pmol de fragment soit  $10^{12}$  molécules), il n'est pas possible de représenter totalement une banque de fragments contenant plus de 20 pb aléatoires.

Cependant, même si une banque de fragments aléatoires ne peut pas être entièrement représentée, la méthode SELEX est applicable. Dans ce cas, les séquences obtenues ne seront pas forcément optimales pour la fixation de la protéine. Des études dans de telles conditions (travail sur une banque incomplète) ont été réalisées avec les histones. Dans une de ces études, les auteurs construisent une banque contenant trois régions de

séquences aléatoires (de 72 à 76 pb ) soit  $10^{132}$  séquences différentes possibles. Pour débiter leurs expériences, les auteurs ont utilisé 1.6  $\mu\text{g}$  d'ADN. Leur banque de départ ne pouvait donc comprendre que  $5 \times 10^{12}$  séquences différentes soit une infime fraction de l'ensemble des séquences possibles (seulement une séquence sur  $10^{120}$ ; Lowary et Widom, 1998). Une solution dénommée « dopage » permet d'optimiser l'affinité des séquences obtenues par SELEX à partir d'une librairie incomplète. Pour cela, les informations connues sur les sites de fixation sont utilisées et seulement certaines positions sont rendues aléatoires (Knight et Yarus 2003).

Un facteur primordial pour le processus de sélection est **le degré de la spécificité de reconnaissance**. Cette spécificité correspond au rapport des constantes d'équilibre entre la fixation pour les séquences de haute affinité et celle pour les séquences aléatoires de départ. Plus ce rapport est grand, moins le nombre de cycles à accomplir est important. Ainsi dans le cas des protéines chaperonnes d'E. coli CspB, C, D (cold shock protein ; protéines avec une faible spécificité de reconnaissance) il a fallu 10 cycles de sélection pour isoler 11 séquences préférentielles de fixation parmi une librairie de fragments contenant  $4^{15}$  ( $10^9$ ) séquences différentes (Phadtare et Inouye, 1999). En revanche dans le cas du facteur de transcription AP-2 $\alpha$  (intervenant dans la régulation de la prolifération cellulaire et de l'embryogénèse de vertébrés) 3 cycles ont été suffisants pour déterminer une séquence consensus de fixation à partir d'une librairie de fragments contenant  $4^{16}$  ( $4 \times 10^9$ ) séquences différentes (Mohibullah et coll., 1999). D'après Irvine et collaborateurs (1991), pour une librairie de  $10^n$  séquences aléatoires, n cycles seraient toujours suffisants pour obtenir une sélection.

Un autre point important est **le rapport signal/bruit** de la méthode utilisée pour séparer les complexes des ADN libres. Quelle que soit la méthode utilisée le « signal » correspond à l'ensemble des ADN correctement séparés, le « bruit » étant les séquences qui n'ont pas été correctement séparées. Le « bruit » peut avoir deux origines distinctes : 1) des ADN libres peuvent être entraînés dans la fraction contenant les complexes ; 2) les complexes peuvent être déstabilisés lors de l'étape de sélection. Quelle qu'en soit la cause un mauvais rapport signal/bruit entraînera toujours une augmentation du nombre de cycles à réaliser.

Par ailleurs, il est important de contrôler le **taux de sélection** (pourcentage de fragments complexés). En effet dans des conditions où trop de complexes sont formés, la protéine ne choisit plus les sites de forte affinité. En revanche, si la quantité de protéine est trop faible, une séquence de forte affinité, faiblement représentée, risquerait d'être perdue. Il est clair que les différents paramètres sont liés. Ainsi, le risque de perte de fragments dépend également de **la quantité totale de fragments utilisés pour la sélection**. Comme pour l'étape d'amplification par PCR, la quantité totale de fragments à utiliser pour la sélection est dépendante de la longueur de la séquence aléatoire. D'après Irvine et collaborateurs (1991), dans le cas où l'on n'a aucune information sur l'affinité de la protéine pour les séquences préférentiellement reconnues il est important de ne sélectionner que 6 % des fragments pour ne pas perdre de séquences et avoir un nombre de cycles raisonnable.

## **2) Sélection de sites préférentiels pour la protéine MC1 :**

### **2.1) Choix des conditions expérimentales :**

Le choix de la longueur de la séquence aléatoire s'est fait, pour une part, en fonction de données concernant l'interaction « non spécifique » de la protéine. Ainsi, différentes études montrent que le site minimum de fixation (site exclu) est de 11 pb (Culard et coll., 1993 ; Isabelle et coll., 1993). La longueur de la partie aléatoire devait donc être d'au moins 11 pb et ne pouvait dépasser 21 pb, sinon deux protéines auraient pu s'y fixer. Nous savions, d'autre part, que des duplex de 15 pb de séquence proche de celle du site I avaient une longueur suffisante pour permettre une interaction forte avec la protéine (Thèse de C. Paradinas, 1998). D'après ces données et pour ne pas partir d'un nombre trop grand de séquences à sélectionner, nous avons choisi d'utiliser une banque de fragments contenant une partie aléatoire de 15 pb.

Outre la région aléatoire, les fragments se composent de deux séquences de 25 pb servant pour la fixation des amorces de PCR (voir Matériels et Méthodes p60). Nous avons choisi des amorces utilisées dans deux études précédentes pour des expériences SELEX (Harley et coll., 1994, David-cordonnier et coll., 1998). Au total les fragments ont une longueur de 65 pb (de manière « non spécifique » 5 protéines peuvent s'y fixer).

En théorie, la séquence aléatoire choisie peut former  $4^{15}$  (soit  $10^9$ ) molécules différentes, cependant, l'orientation des séquences n'ayant *a priori* pas d'importance, ce nombre peut être divisé par deux. Au moment de la première amplification par PCR, nous avions 0.8 pmol de fragments soit  $4.8 \times 10^{11}$  molécules. Nous pouvions dès lors considérer que nous avions en moyenne 960 copies de chaque séquence ( $4.8 \times 10^{11} / 5 \times 10^8$ ). Ce large excès nous assurait que chaque séquence possible était présente dans la banque de départ.

L'étape de sélection a été réalisée à une concentration de fragment de 50 nM (et une quantité de 1.1 µg d'ADN). Pour les premiers cycles, la concentration en protéine (50 nM) a été choisie de sorte que 10 % des fragments soient complexés. Ce choix, supérieur à celui préconisé dans les études théoriques (Irvine et coll., 1991 ; Vant-Hull et coll., 1998), est celui adopté dans les travaux réalisés avec l'octamère d'histones (Lowary et Widom, 1998). Il permet d'éviter une perte des séquences.

Les complexes MCI/ADN étant stables pendant une électrophorèse sur gel de polyacrylamide, nous avons utilisé ce moyen pour séparer les complexes des fragments libres. Cette méthode, bien qu'assez longue à réaliser, présente plusieurs avantages : 1) le rapport signal/bruit est relativement élevé, ce qui limite le nombre de cycles à effectuer ; 2) des informations sur le processus de sélection peuvent être directement obtenues. Ainsi, une augmentation de la quantité de complexes monoprotéiques (pour une concentration en protéine donnée) et une variation dans la mobilité des complexes indiquent que le processus d'enrichissement a commencé.

## 2.2) Déroulement des cycles de sélection :

Lors des sélections successives, la quantité de complexes formés a été contrôlée par quantification des gels retards. Comme le montrent les résultats présentés dans le Tableau 3 nous avons, à partir du cycle 6, décidé d'utiliser des conditions de plus en plus stringentes. En effet nous pouvions considérer que la sélection avait effectivement commencé et que le risque de perte de séquences préférentielles avait fortement diminué. La quantité de protéine a alors été diminuée pour que seulement 6 puis 5 % de l'ADN soit complexé (les derniers cycles ont été réalisés avec une concentration en protéine de 5 nM).



**Tableau 3 : récapitulatif du taux de sélection obtenu pour chaque cycle SELEX.**

Le taux de sélection est égal au pourcentage d'ADN complexé. Ce pourcentage est déterminé par quantification des gels de polyacrylamide à l'aide du logiciel Image Quant (Molecular Dynamics).

Cycle	1	2	3	4	5	6	7	8	9	10
Taux de sélection	10	11	10	11.8	11.5	8.4	6	7.7	4.5	4.5

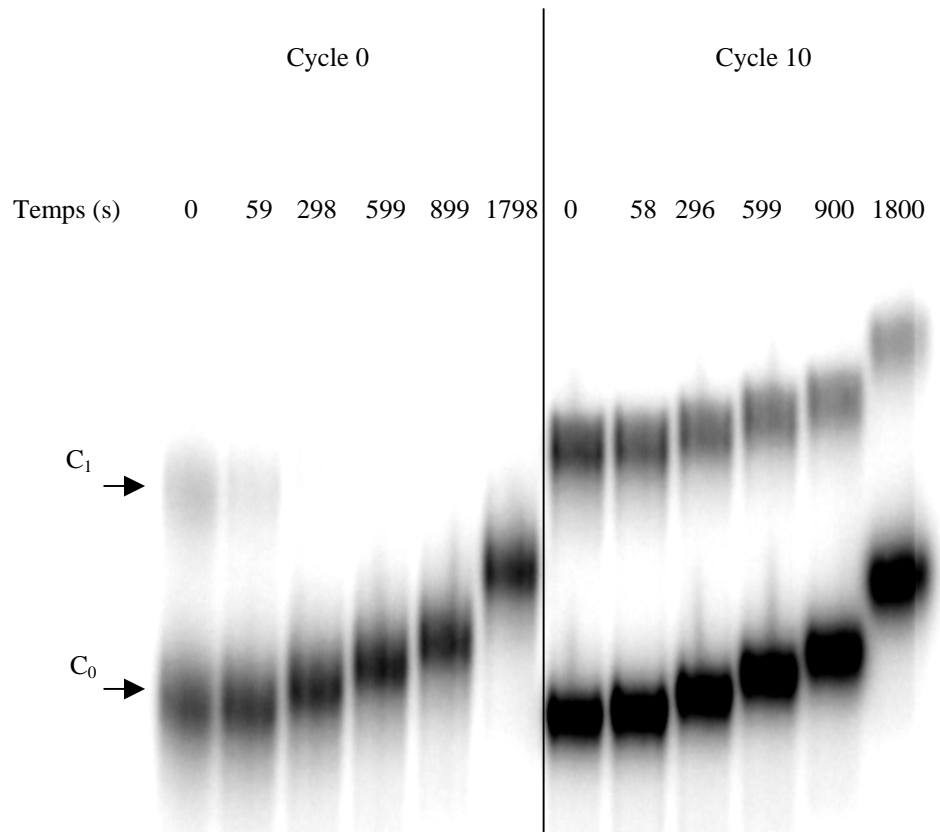
2.3) Mise en évidence de la sélection de sites préférentiels :

Pour mesurer l'enrichissement en séquences préférentielles de fixation nous avons choisi de mesurer les vitesses de dissociation des complexes. Cette méthode nous a semblé la plus précise pour quantifier de faibles enrichissements (nous pensions que l'enrichissement était faible car la concentration en protéine n'avait pu être réduite que d'un facteur 10).

Dans ces expériences, les complexes MC1/ADN sont, dans un premier temps, formés puis, lorsque l'équilibre est atteint, un large excès d'ADN non radiomarqué est ajouté (temps 0). Des aliquotes de ce mélange sont ensuite chargés sur un gel retard à différents temps. Chaque molécule de protéine qui se dissocie du complexe se trouve piégée par l'ADN compétiteur, de sorte que nous observons une diminution progressive de la quantité de complexes en fonction du temps.

Le gel de la Figure 20 permet de comparer :

- 1) la quantité de complexes formés avec les deux populations de fragments (temps 0).
- 2) leur cinétique de dissociation.

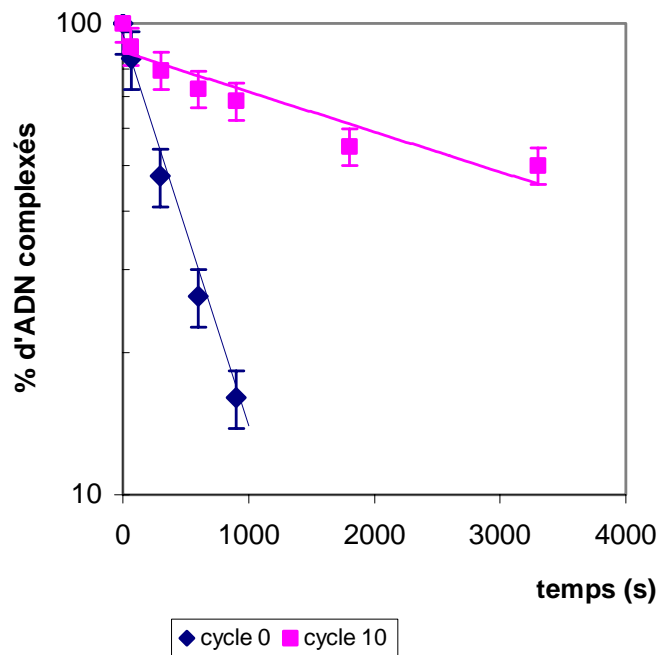


**Figure 20 : dissociation des complexes MC1/fragments du cycle 0 ou du cycle 10.**

Gel retard. Les complexes sont formés par incubation de 5 nM de fragment SELEX avec 2.5 nM de protéine MC1 dans un tampon TE contenant 60 mM de NaCl. A l'équilibre, du plasmide non radiomarqué est ajouté à la concentration finale de 1.8  $\mu$ M (équivalent taille). Les temps d'incubation avec le plasmide figurent en haut de chaque piste. C<sub>0</sub> : ADN libre ; C<sub>1</sub> : ADN dans le complexe. La migration a lieu pendant 5 h sous 14 V/cm à 4°C sur un gel de polyacrylamide 8 % (R=19/1).

Alors que la même concentration en protéine est utilisée pour les deux types de fragments, on observe que la proportion d'ADN complexé est supérieure avec l'ADN du cycle 10 (43 % par rapport à 19 %). Cette différence montre qu'il y a eu sélection de séquences de plus haute affinité. Il apparaît par ailleurs, que les complexes formés avec l'ADN du cycle 10 migrent légèrement moins vite que ceux formés avec les fragments de départ. Il est vraisemblable que l'ADN du cycle 10 adopte dans le complexe une conformation différente de celle adoptée par l'ADN de départ (courbure plus grande). Il est aussi possible que ce retard soit dû à un temps de vie des complexes plus long (temps pendant lequel l'ADN est courbé plus long).

Les vitesses de dissociation des complexes sont également nettement différentes. Les complexes formés avec les ADN de départ se dissocient plus rapidement. Ainsi, après 30 min, l'ADN du cycle 0 est complètement libre alors que 70 % des complexes sont encore présents avec l'ADN du cycle 10 (Figure 21).



**Figure 21 : dissociation des complexes formés avec les fragments du cycle 0 et du cycle 10.**

L'analyse quantitative de plusieurs gels retard identiques à celui de la Figure 20 permet de calculer le pourcentage d'ADN complexés en fonction du temps. Les pourcentages d'ADN complexés ont été normalisés par les valeurs obtenues au temps 0.

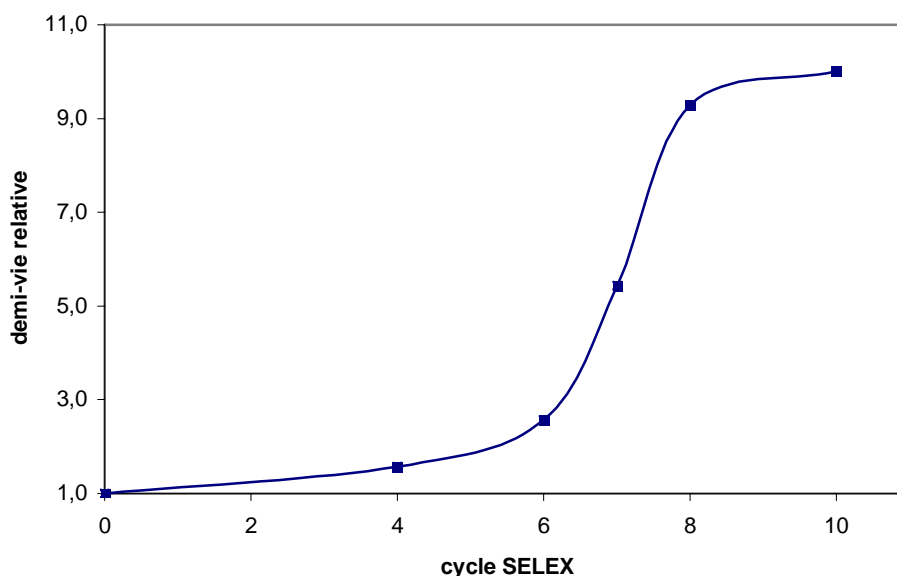
A partir des courbes (Figure 21), il est possible de déterminer les temps de demi-vie et les vitesses de dissociation des complexes. Les résultats présentés dans le Tableau 4 montrent qu'après le 10<sup>ème</sup> cycle, la vitesse de dissociation des complexes est environ 10 fois plus faible.

**Tableau 4 : temps de demi-vie et vitesse de dissociation des complexes obtenus avec les ADN des cycles 0 et 10.**

$\Gamma$  : temps de demi-vie ;  $k_d$  : vitesse de dissociation

	$\Gamma$ (s)	$k_d$ ( $s^{-1}$ )
Cycle 0	$300 \pm 15 \%$	$2.3 \cdot 10^{-3} \pm 15 \%$
Cycle 10	$3300 \pm 10 \%$	$2.1 \cdot 10^{-4} \pm 10 \%$

Pour décider de la poursuite ou de l'arrêt des cycles de sélection nous avons mesuré l'évolution du temps de demi-vie des complexes au cours des différents cycles (Figure 22).



**Figure 22 : évolution du temps de demi-vie des complexes MC1/fragments SELEX au cours des cycles SELEX.**

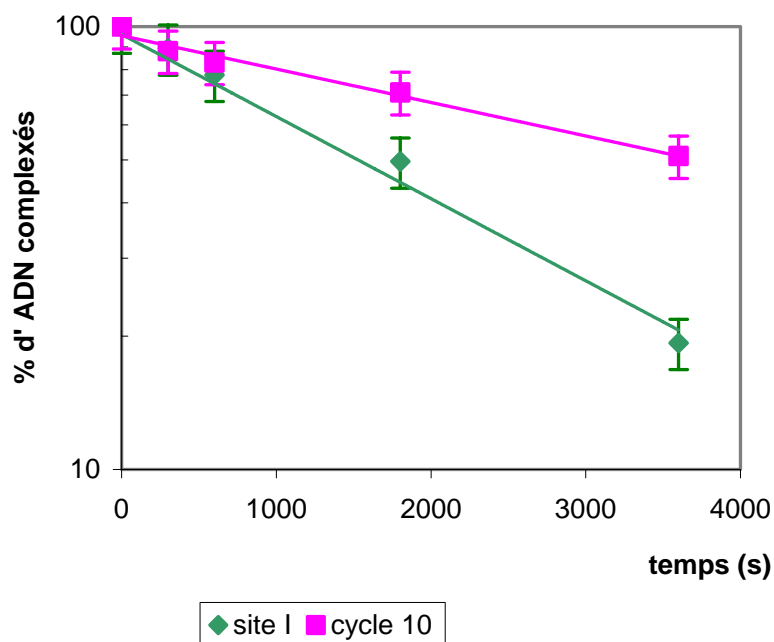
La demi-vie relative est égale au temps de demi-vie des complexes formés avec l'ADN du cycle considéré divisé par le temps de demi-vie des complexes formés avec l'ADN initial. Les données présentées sur ce graphique sont les valeurs moyennes obtenues par plusieurs expériences indépendantes.

L'augmentation de la stabilité des complexes est nette à partir du cycle 6. Pour les cycles 6 à 8, le temps de demi-vie des complexes augmente de façon importante, puis la

courbe atteint un plateau. La proportion de fragments de forte affinité pour la protéine MC1 n'augmentant plus de façon importante depuis le cycle 8 nous avons arrêté le processus de sélection.

A ce stade de notre travail, nous avons voulu comparer l'affinité de la protéine pour les ADN du cycle 10 à celle pour le site I préalablement déterminé (Teyssier et coll., 1994). Pour cela, nous avons réalisé deux types d'expériences : 1) compétitions directes entre des fragments de taille voisine (64 et 65 pb) ; 2) mesures de vitesse de dissociation des différents complexes. Les résultats obtenus par ces deux méthodes sont équivalents et seule l'expérience concernant la stabilité des complexes est présentée.

Le graphique présenté Figure 23 montre le résultat de la quantification de plusieurs expériences de mesure de vitesse de dissociation : la dissociation des complexes MC1/aptamères est plus lente que celle des complexes MC1/site I.



**Figure 23 : mesure des vitesses de dissociation des complexes MC1/aptamères et MC1/site I.**

Les conditions expérimentales sont identiques à celles de la Figure 20. Les pourcentages d'ADN complexés ont été normalisés par les valeurs obtenues au temps 0.

Le temps de demi-vie des complexes formés avec le site I est de 1800 s ( $\pm 10\%$ ) alors que le temps de demi-vie des complexes formés avec les ADN du cycle 10 est de 3300 s ( $\pm 10\%$ ).

Cette augmentation de stabilité (facteur 2) nous montre que nous avons sélectionné des fragments de plus forte affinité pour la protéine MC1 que ceux précédemment isolés.

### 3) Analyse des séquences sélectionnées :

#### 3.1) Détermination des séquences :

Les ADN issus du cycle 10 ont été digérés par les enzymes de restriction Eco RI et Bam HI, insérés dans du plasmide pGEM 3Zf(-) et clonés dans des bactéries DH5 $\alpha$ . L'ADN extrait de 19 clones indépendants a été fourni à la société MWG-biotech pour être séquencé.

Sur les 19 clones seulement 13 ont pu être séquencés sans ambiguïté.

<b>Clone 1</b>	5 ' AATTCGCCTC <u>AAAAACA</u> CCACCCCC CGACAGGATCCTC 3 '
<b>Clone 2</b>	AATTCGCCTC <u>AAAAACA</u> CACACCCC CGACAGGATCCTC
<b>Clone 3</b>	AATTCGCCTC <u>AAAAACA</u> CACACCTC CGACAGGATCCTC
<b>Clone 4</b>	AATTCGCCTC <u>AAAAACA</u> CCACCCCC CGACAGGATCCTC
<b>Clone 8</b>	AATTCGCCTC <u>AAAAACA</u> CACCCCC CGACAGGATCCTC
<b>Clone 10</b>	AATTCGCCTC <u>AAAAACA</u> AAACACACA CGACAGGATCCTC
<b>Clone 11</b>	AATTCGCCTC <u>AAAAACA</u> CACACCCA CGACAGGATCCTC
<b>Clone 12</b>	AATTCGCCTC <u>AAAAACA</u> CACACCAC CGACAGGATCCTC
<b>Clone 13</b>	AATTCGCCTC <u>AAAAACA</u> CACCCCC CGACAGGATCCTC
<b>Clone 14</b>	AATTCGCCTC <u>AAAAACA</u> CACACCAC CGACAGGATCCTC
<b>Clone 16</b>	AATTCGCCTC <u>AAAAACA</u> CACACCCA CGACAGGATCCTC
<b>Clone 18</b>	AATTCGCCTC <u>AAAAACA</u> CACACCCA CGACAGGATCCTC
<b>Clone 20</b>	AATTCGCCTC <u>AAAAACA</u> GCACACCC CGACAGGATCCTC

**Figure 24 : résultat du séquençage des clones issus du cycle 10.**

La colonne centrale correspond aux 15 pb aléatoires du cycle 0. Les colonnes de gauche et droite correspondent aux séquences constantes des fragments. Les bases soulignées et bleues (colonne centrale) sont identiques pour toutes les séquences.

L'alignement des séquences montre que les 13 séquences sont très conservées. Plusieurs clones ont la même séquence (clones 1 et 4, clones 11, 16 et 18, clones 12 et 14), et au total 9 séquences différentes ont été obtenues. Ceci semble indiquer que le nombre de clones séquencés est représentatif de l'ensemble de la banque. Sur les 15 paires de bases aléatoires de départ, 6 sont strictement conservées après la sélection et toutes les séquences comportent en 5' la séquence AAAACA. Les 8 bases en 3' du site sont presque exclusivement des adénines ou des cytosines. Seule la séquence C3 (issue du clone 3) comprend une thymine et la séquence C20 comprend une guanine.

Les résultats font apparaître également deux points qui ont retenu notre attention : d'une part la sélection s'est effectuée sur la totalité de la séquence aléatoire, d'autre part l'orientation de la séquence (par rapport aux séquences adjacentes constantes) est la même pour les 13 séquences. On pouvait dès lors se demander si les séquences adjacentes n'avaient pas joué un rôle lors de l'étape de sélection.

Dans le but de tester cette hypothèse et afin de savoir si les séquences isolées correspondaient ou non au site complet, nous avons effectué une deuxième sélection (SELEX 2).

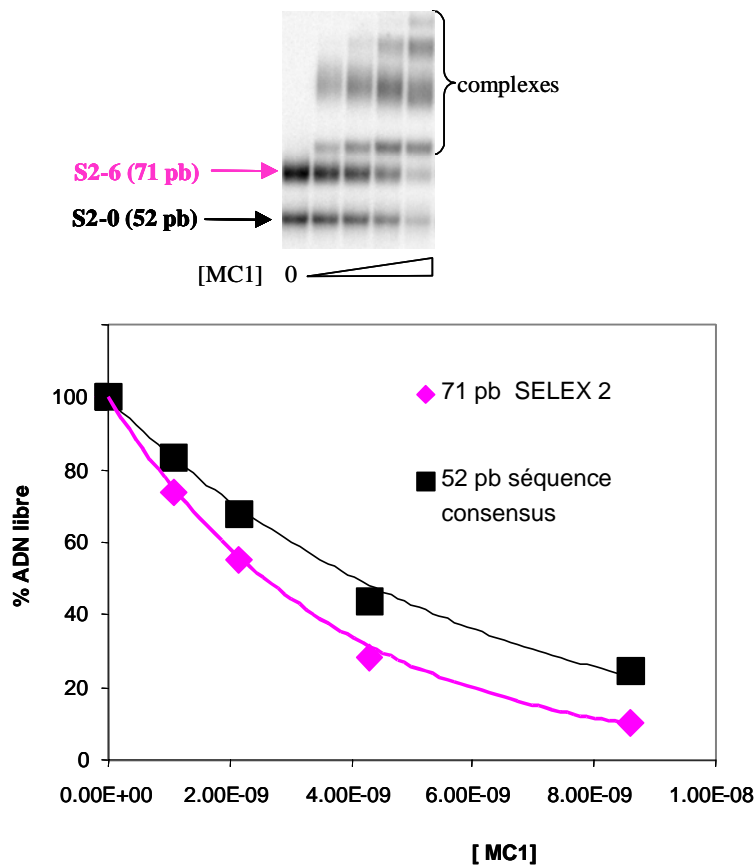
### 3.2) Deuxième expérience de sélection :

Les fragments utilisés dérivent de la séquence du clone 2 (la justification complète de notre choix de séquence apparaît lors de la comparaison détaillée des séquences sélectionnées ; p78). Un mélange de cytosine et d'adénine a été conservé en positions 11 et 15 et, 3 pb complètement aléatoires ont été ajoutées en 5' et 3' des fragments (nommés S2-0 ; NNNAACACAC(A/C)CCC(C/A)NNN). A chaque extrémité de ces fragments, on retrouve les séquences constantes utilisées lors de la 1<sup>ère</sup> sélection (25 pb chacune). La taille totale des fragments est donc de 71 pb.

Les cycles du SELEX 2 ont été réalisés dans les mêmes conditions expérimentales que pour la première série d'expériences. Six cycles ont ainsi été réalisés avec un taux de sélection de 6 %.

Pour comparer l'affinité de la protéine MC1 pour les ADN issus du 6<sup>ème</sup> cycle (S2-6) à celle de la protéine pour les ADN de départ (S2-0) nous avons réalisé des compétitions directes entre les deux populations de fragments.

Dans l'expérience montrée Figure 25, ces fragments sont soit entiers, 71 pb (S2-6), soit digérés par Eco RI, 52 pb (S2-0), mais l'expérience inverse a également été réalisée (71 pb pour les fragments S2-0 et 52 pb pour S2-6).



**Figure 25 : compétition directe entre les fragments S2-6 et S2-0 pour la fixation de la protéine MC1.**

Les deux ADN radiomarqués sont mélangés à une concentration finale de  $8 \times 10^{-12}$  M (pour chaque duplex) et incubés avec des concentrations en protéine respectivement de gauche à droite de 0 ; 1 ; 2 ; 4 et 8 nM. Le tampon de fixation utilisé contient 0.1 M de NaCl et 15  $\mu\text{g}/\text{mL}$  de SAB. La migration a lieu pendant 3 h sous 14 V/cm à 4°C sur un gel de 8 % polyacrylamide (R=30/1).

L'analyse quantitative du gel (Figure 25) montre que, en présence de concentrations croissantes de protéine, les fragments libres S2-6 disparaissent plus vite que



les fragments S2-0. Cette différence est reproductible et ne dépend pas de la longueur des fragments utilisés (car le même résultat est obtenu avec l'expérience complémentaire).

Nous avons considéré que cette différence d'affinité était faible (inférieure à 2) et que le fait d'avoir utilisé un mélange pour les positions 11 et 15 (adénines et cytosines) avait également pu contribuer à cette augmentation d'affinité (cf. Tableau 5, p 80, séquence C2, C11, C8).

Nous n'avons pas poursuivi plus loin cette étude, et nous n'avons pas cherché à déterminer les séquences.

### 3.3) Mise en évidence d'une séquence consensus :

L'histogramme présenté Figure 26 permet de comparer les nucléotides entourant la séquence AAAACA (positions 2-7) strictement conservée dans tous les clones.

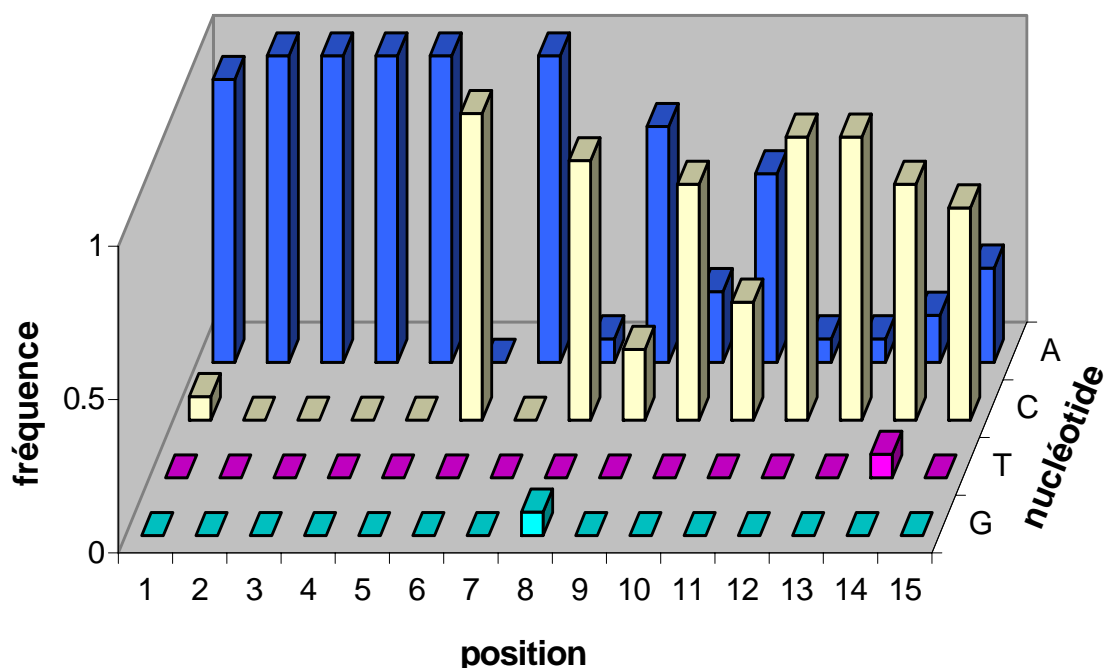


Figure 26 : fréquence des bases dans les 13 fragments séquencés.

Plusieurs positions sont bien conservées :

- position 1 : une adénine pour 12 des 13 séquences.
- positions 12 et 13 : 12 des 13 séquences comprennent des cytosines.
- position 14 : une cytosine est présente dans 10 séquences sur 13.
- positions 8 à 10 : séquence CAC retrouvée dans 10 séquences sur 13.

Deux positions apparaissent moins bien conservées :

- position 15 : 9 séquences ont des cytosines et 4 des thymines.
- position 11 : c'est la position la plus variable, elle est composée majoritairement par une adénine (8 adénines pour 5 cytosines).

La séquence consensus suivante a pu être établie :

**5' AAAAACACAC(A/C)CCC(C/A) 3'**

**3' TTTTGTGTG(T/G)GGG(G/T) 5'**

Cinq séquences sélectionnées sur 13 (clones 2, 11, 16, 18, 8) correspondent à la séquence consensus. Ce résultat confirme que nous sommes allés jusqu'au bout du processus de sélection.

La séquence consensus de la protéine MC1 ne ressemble à aucun site de fixation connu. En revanche, sa partie strictement conservée (AAAACA) est proche de la séquence de sites consensus de fixations connus (David-Cordonnier et coll., 1998) :

tyrosine kinase c-Ab11 (A(A/C)AACAA(A/C) ; 5 pb sur 8)

protéines à boîte HMG : - LEF-1 (AACAAAG ; 4 pb sur 7)

- SRY ((A/T)AACAA(A/T) ; 5 pb sur 7)

### 3.4) Comparaison de l'affinité de MC1 pour chacune des séquences issues du SELEX :

Nous avons, dans un premier temps, comparé les stabilités des complexes formés avec les 9 séquences différentes sélectionnées. Ces expériences ont été réalisées avec des

fragments de 65 pb obtenus par PCR à partir des différents plasmides (cf. Matériels et méthodes p53).

Les complexes formés avec ces séquences étant nettement plus stables que ceux obtenus avec des séquences aléatoires, nous avons pu réaliser les études dans un tampon contenant 100 mM de NaCl (dans ces conditions les complexes avec les fragments de départ ont des temps de demi-vie très courts et la cinétique de dissociation ne peut être étudiée par gel retard).

Les résultats de ces expériences sont résumés dans le Tableau 5. Les valeurs correspondent aux valeurs moyennes obtenues par trois expériences indépendantes pour chaque séquence.

Les vitesses de dissociation des différents complexes varient au maximum d'un facteur 6. L'ensemble des résultats est cohérent avec la séquence consensus établie précédemment. Les trois séquences formant les complexes les plus stables (C2, C11, C8) correspondent à la séquence consensus.

*Pour alléger le texte qui suit, les complexes formés entre la protéine MC1 et une séquence C $\alpha$  seront nommés complexes C $\alpha$ .*

**Tableau 5 : cinétiques de dissociation des complexes formés entre la protéine MC1 et chacune des séquences issues du SELEX.**

Les séquences sont nommées en fonction du clone dont elles sont extraites. Le tampon de fixation utilisé est le même que celui utilisé Figure 20 mais la force ionique est augmentée à 100 mM NaCl. Les bases en rouge gras sont celles qui diffèrent de la séquence consensus (notée en vert) ;  $k_d$  : vitesse de dissociation.

clone	séquence	demi-vie (s)	$k_d$ ( $\times 10^{-4}$ ) ( $s^{-1}$ )	précision (%)
C2	<b>AAAACACACACCCC</b>	1750	3.96	7
C11	AAAACACACACCC <b>A</b>	1425	4.86	3
C8	AAAACACAC <b>CCCC</b>	1350	5.13	1
C12	AAAACACACAC <b>AC</b>	1300	5.33	1
C20	AAAACAG <b>CACACCC</b>	950	7.30	11
C13	<b>C</b> AAAACACAC <b>CCCC</b>	700	9.90	15
C3	AAAACACACAC <b>CTC</b>	550	12.60	12
C4	AAAACAC <b>CACCCCC</b>	400	17.33	14
C10	AAAACA <b>AACACACA</b>	300	23.10	16

De façon générale, on peut dire que plus le nombre de substitutions par rapport à la séquence consensus est grand, moins les complexes sont stables. Cependant certaines séquences n'obéissent pas à cette remarque. Ainsi le complexe C3 (une seule substitution) est moins stable que le complexe C20 (5 substitutions).

Certaines comparaisons donnent des résultats surprenants qui ne permettent pas une interprétation directe des différences observées :

- le complexe C12 est 2.5 fois plus stable que le complexe C3, alors que ces deux séquences ont une seule substitution (en position 14). Le remplacement d'une pyrimidine (cytosine) par une autre pyrimidine (thymine) perturbe plus l'interaction qu'un remplacement par une purine (adénine).
- de même, le complexe C20 est plus de 2.4 fois plus stable que le complexe C4 alors que ces deux séquences ont 13 pb identiques sur 15 et que la séquence C4 est la plus proche de la séquence consensus. Il est possible que la séquence C20 possède, grâce à la répétition du di-nucléotide CA, une flexibilité plus grande que la séquence C4.

Certaines comparaisons nous donnent cependant des informations directement accessibles :

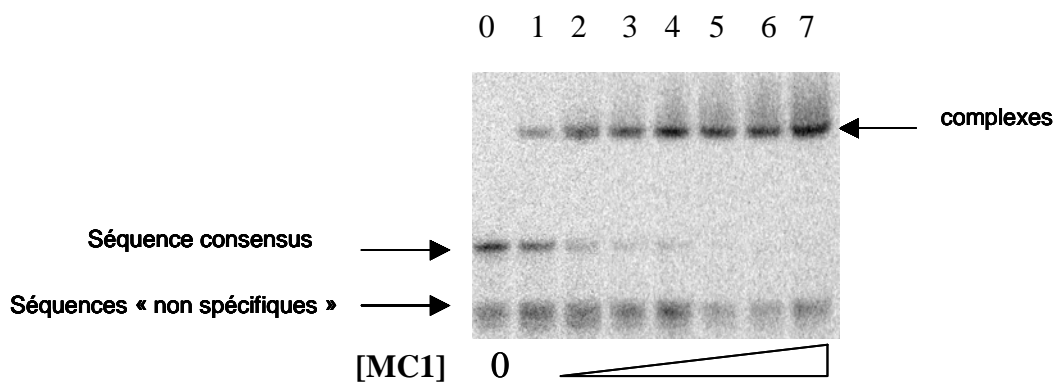
- la comparaison des stabilités des complexes C8 et C13 montre l'importance de la répétition d'adénines. En effet, le remplacement de l'adénine par une cytosine en position 1 (séquence C13) provoque une augmentation de la vitesse de dissociation d'un facteur 2.
- les complexes C10 sont fortement défavorisés. Il est vraisemblable que la séquence C10, (seulement 3 substitutions) ne possède plus certaines caractéristiques importantes pour l'interaction. Ainsi, la suite de trois adénines (positions 7 à 9) rend probablement rigide une région qui doit être flexible pour favoriser l'interaction.

### 3.5) Détermination de l'affinité de l'interaction MC1/séquence consensus :

Pour comparer l'affinité de la protéine MC1 pour la séquence consensus à celle « non-séquence » spécifique pour une séquence aléatoire, plusieurs expériences ont été réalisées : mesures des constantes d'équilibre et des constantes de vitesse de dissociation, compétition directe entre fragments.

Dans toutes ces expériences, la séquence appelée séquence consensus est celle ayant la plus forte affinité pour la protéine (AAAAACACACACCCC).

Le gel de la Figure 27 montre une compétition entre un fragment de 20 pb contenant une séquence aléatoire de 13 pb (20pbN<sub>13</sub>) et un fragment de 26 pb contenant la séquence consensus pour la fixation de la protéine MC1.



**Figure 27 : compétition directe entre la séquence consensus et une séquence aléatoire pour la fixation de la protéine MC1.**

La concentration en fragment est de  $8 \times 10^{-12}$  M (pour chaque duplex). La concentration en protéine est respectivement de gauche à droite 0 ; 1.6 ; 3.2 ; 6.5 ; 13 ; 39 ; 117 ; 235 nM. Le tampon de fixation contient 0.1 M de NaCl et 15  $\mu$ g/mL de SAB. La migration a lieu pendant 3 h sous 14 V/cm à 4 °C sur un gel de 8 % polyacrylamide (R=30/1).

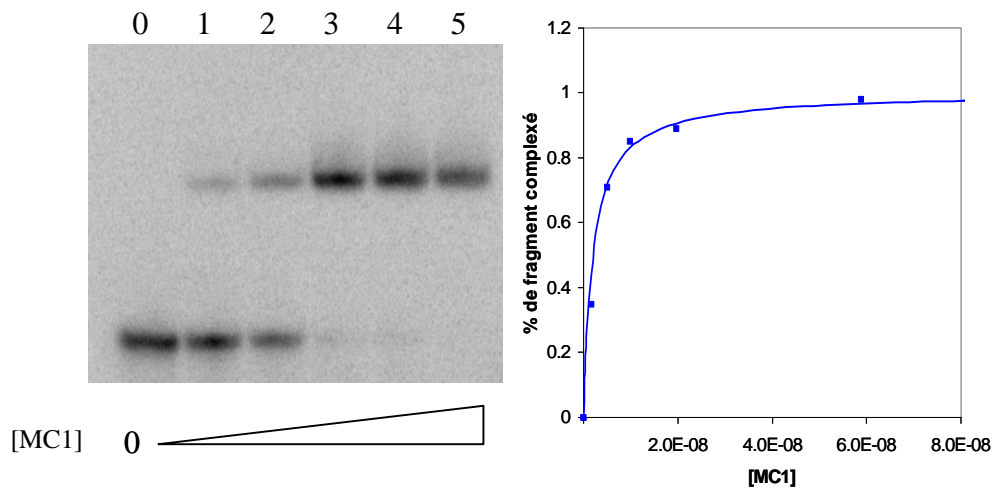
Les deux types de complexes ne sont pas séparés, mais la comparaison des disparitions progressives des deux ADN libres avec les concentrations croissantes de protéine, montre clairement que la protéine choisit de se fixer préférentiellement sur la séquence consensus : ainsi, alors que 75 % des fragments contenant la séquence consensus sont complexés, 95 % des fragments de séquence aléatoire sont encore à l'état libre (piste 2).

L'analyse suivant la relation A (matériels et méthodes p54) nous donne une valeur moyenne pour trois expériences de :

$$K_{d \text{ npréf}}/K_{d \text{ préf}} = 45$$

où  $K_{d \text{ npréf}}$  et  $K_{d \text{ préf}}$  sont les constantes d'équilibre de dissociation respectivement avec le fragment de séquence aléatoire et le fragment contenant la séquence consensus.

La Figure 28 montre un gel retard correspondant à une titration de la séquence consensus par des concentrations croissantes en protéine.

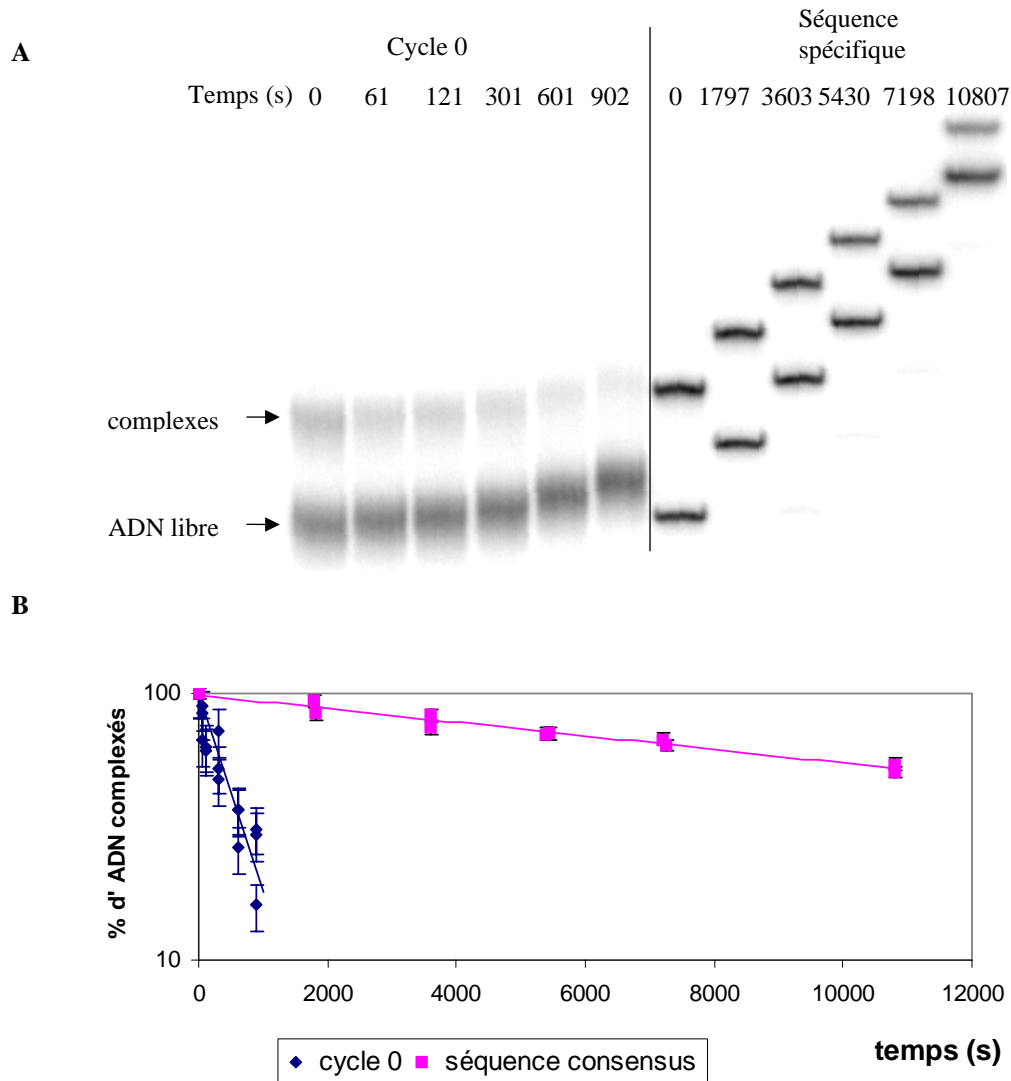


**Figure 28 : analyse par gel retard de la fixation de la protéine MC1 sur un fragment contenant la séquence consensus (26 pb préférentiel).**

Les concentrations en protéine sont respectivement pour les pistes 0 à 5 de 0 ; 1.6 ; 5 ; 9.8 ; 19.6 ; 58.8 nM. La concentration en ADN est de  $8 \times 10^{-12}$  M. La fixation est réalisée en présence de 100 mM NaCl, 15  $\mu$ g /ml de SAB, dans un tampon TE. La migration a lieu pendant 3 h sous 14 V/cm à 4 °C sur un gel de 8 % polyacrylamide (R=30/1).

La quantification du gel permet de calculer la fraction d'ADN complexé en fonction de la concentration en protéine. Les valeurs obtenues s'ajustent avec celles calculées pour une constante apparente de dissociation de  $K_d = 2$  nM (équation utilisée : cf. Le Cam et coll., 1999). Lorsque la même expérience est réalisée avec une séquence aléatoire (fragment 20pbN<sub>13</sub>) un  $K_d$  de 100 nM est obtenu (résultats non montrés).

Les cinétiques de dissociation (Figure 29) ont été réalisées avec des fragments de 65 pb dans un tampon de fixation contenant 60 mM NaCl (cf. Matériels et Méthode p53).



**Figure 29 : mesure des vitesses de dissociation des complexes MC1/site consensus et MC1/séquence aléatoire.**

A : gel retard. B : résultat de la quantification de plusieurs gels retard. Les conditions expérimentales sont identiques à celles de la Figure 20.

A partir des courbes (Figure 29), il est possible de déterminer les temps de demi-vie et les vitesses de dissociation des complexes. Les résultats présentés dans le Tableau 6

montrent que les complexes formés avec la séquence consensus sont environ 38 fois plus stables que ceux formés avec des séquences aléatoires.

**Tableau 6 : temps de demi-vie et vitesses de dissociation des complexes obtenus avec les ADN du cycle 0 ou la séquence consensus.**

$\Gamma$  : temps de demi-vie ;  $k_d$  : vitesses de dissociation

	$\Gamma$ (s)	$k_d$ ( $s^{-1}$ )
Cycle 0	$300 \pm 15 \%$	$2.3 \cdot 10^{-3} \pm 15 \%$
Séquence consensus	$11400 \pm 7 \%$	$6.08 \times 10^{-5} \pm 7 \%$

Les méthodes à l'équilibre nous donnent des résultats équivalents : l'affinité pour la séquence consensus est 45 à 50 fois supérieure à celle obtenue pour les séquences aléatoires. Les mesures de cinétique de dissociation nous donnent des résultats un peu différents (facteur de 38). Le plus probable est que cette différence soit due à la méthode utilisée (peu précise pour des vitesses de dissociation rapides) et non à un effet sur la vitesse d'association (des études par résonance plasmonique de surface auraient été plus précises).

#### 4) Discussion :

Après tout processus de sélection *in vitro*, on peut se demander si les conditions expérimentales ont pu influencer les résultats. Tout d'abord, nous pouvons dire que les amorces utilisées lors de la PCR n'ont pas (ou peu) influencé le processus de sélection et ceci pour plusieurs raisons : 1) leur séquence est très différente de celles reconnues par la protéine ; 2) en terme d'affinité, les résultats des deux expériences SELEX sont très proches ; 3) les contacts protéine/ADN ont lieu dans la région sélectionnée (cf. partie suivante). En revanche, comme nous sommes allés loin dans la sélection (les séquences



obtenues sont très proches les unes des autres), il est possible que des séquences de forte affinité mais faiblement représentées aient été perdues. Il serait sans doute intéressant de rechercher les séquences des fragments obtenus en début de sélection (par exemple au cycle 7) afin de savoir si des séquences totalement différentes sont également reconnues.

Les séquences sélectionnées sont-elles présentes dans le génome des méthanosarcines ?

On peut aligner la séquence du site I, trouvée sur un fragment d'ADN de *Methanosarcina barkeri* (Teyssier et coll., 1994), avec les différentes séquences issues de la sélection. Le site I fait clairement partie de la même famille de séquences. Le meilleur alignement possible est obtenu avec la séquence du clone 20 (9 bases sur 15 sont identiques ; Figure 30).

site I	A	A	A	A	A	G	A	G	C	A	T	C	A	T	C
séquence C20	A	A	A	A	A	C	A	G	C	A	C	A	C	C	C

**Figure 30 : alignement de la séquence du site I et de la séquence C20.**

Les bases notées en fuchsia gras sont identiques pour les deux séquences, les bases notées en bleu sont différentes.

L'affinité de la protéine pour le site I est vraisemblablement du même ordre de grandeur que celle pour la séquence C20, soit entre 10 et 20 fois supérieure à celle pour une séquence aléatoire (nos différents résultats sont cohérents avec cette hypothèse). La fixation préférentielle de la protéine MC1 avait précédemment été estimée 2000 fois plus forte pour le site I que pour une séquence quelconque. La différence entre les deux résultats peut s'expliquer par le fait que le fragment de 209 pb utilisé par Teyssier et collaborateurs (1994) contient certainement plusieurs sites préférentiels de fixation. En effet, ce fragment comporte en plus du site I, 4 répétitions de 4 à 6 adénines. Or, nous savons maintenant que les répétitions d'adénines sont des éléments importants pour la reconnaissance.

Pour savoir si la séquence consensus (AAAAACACACACCCC) est présente dans les génomes des méthanosarcines, nous avons interrogé la banque de données NCBI à l'aide du programme « blastn » (réglage par défaut des paramètres).

Les résultats obtenus sont montrés dans la Figure 31. Pour l'ensemble des génomes d'archaeobactéries testés, les séquences trouvées ont une taille comprise entre 11 et 14 bases. De manière générale, on voit que le nombre de séquences proches de la séquence consensus est plus important chez les méthanosarcines que chez les autres espèces (seule exception, *Ferroplasma acidarmanus*).

Afin de déterminer si cette plus grande occurrence de séquence consensus chez les méthanosarcines est significative, nous avons calculé, de façon simplifiée, le nombre statistique de séquences de 11 à 14 bases que l'on peut trouver dans les génomes d'archaeobactéries. Pour cela, nous avons considéré que les quatre bases ont la même probabilité d'être trouvées et que la taille moyenne du génome des méthanosarcines est de  $4 \times 10^6$  pb. Dans ces conditions, le nombre statistique de séquences de 11 à 14 pb est proche de 3. Les résultats obtenus pour les méthanosarcines sont 6 à 7 fois supérieurs à la statistique alors que ceux obtenus pour les autres organismes sont proches de la statistique (seule exception, *F. acidarmanus*).

Archaea	Nombre de séquences
<b>Crenarchaeota</b>	
<b>Desulfurococcales</b>	
P <input type="checkbox"/> <i>Aeropyrum pernix</i>	4
<b>Sulfolobales</b>	
P <input type="checkbox"/> <i>Sulfolobus solfataricus</i>	2
P <input type="checkbox"/> <i>Sulfolobus tokodaii</i>	3
<b>Thermoproteales</b>	
P <input type="checkbox"/> <i>Pyrobaculum aerophilum</i> str. IM2	1
<b>Euryarchaeota</b>	
<b>Archaeoglobales</b>	
P <input type="checkbox"/> <i>Archaeoglobus fulgidus</i> DSM 4304	0
<b>Halobacteriales</b>	
P <input type="checkbox"/> <i>Halobacterium</i> sp. NRC-1	4
<b>Methanobacteriales</b>	
P <input type="checkbox"/> <i>Methanothermobacter thermautotrophicus</i> str. Delta H	4
<b>Methanococcales</b>	
P <input type="checkbox"/> <i>Methanocaldococcus jannaschii</i>	4
<b>Methanopyrales</b>	
P <input type="checkbox"/> <i>Methanopyrus kandleri</i> AV19	0
<b>Methanosarcinales</b>	
P <input type="checkbox"/> <i>Methanosarcina acetivorans</i> C2A	21
P <input type="checkbox"/> <i>Methanosarcina barkeri</i>	17
P <input type="checkbox"/> <i>Methanosarcina mazei</i> Goe1	19
<b>Thermococcales</b>	
P <input type="checkbox"/> <i>Pyrococcus abyssi</i>	0
P <input type="checkbox"/> <i>Pyrococcus furiosus</i> DSM 3638	1
P <input type="checkbox"/> <i>Pyrococcus horikoshii</i>	0
<b>Thermoplasmatales</b>	
P <input type="checkbox"/> <i>Ferroplasma acidarmanus</i>	12
P <input type="checkbox"/> <i>Thermoplasma acidophilum</i>	1
P <input type="checkbox"/> <i>Thermoplasma volcanium</i>	0

**Figure 31 : recherche des séquences préférentielles de fixation dans les génomes d'archaeobactéries.**

La recherche est effectuée à l'aide du programme « blastn » proposé par la base de données NCBI. Le génome des espèces surlignées en jaune est complètement séquencé, celui des espèces surlignées en vert ne l'est pas complètement. Les chiffres représentent le nombre de séquences trouvées dans les différents organismes. Toutes les séquences trouvées ont entre 11 et 14 bases en commun avec la séquence consensus testée.

Les séquences trouvées par « blastn » ont été localisées dans les génomes de *M. acetivorans C2A* et *M. mazei Goe1* (Tableau 7). Malheureusement, aucune localisation particulière n'a pu être déterminée.

**Tableau 7 : localisation des séquences trouvées par « blastn » dans les génomes de *M. acetivorans C2A* et *M. mazei Goe1*.**

organisme	nombre de séquences localisées dans un gène	nombre de séquences localisées entre deux gènes
<i>M acetivorans C2A</i>	4/21	16/21
<i>M. mazei Goe1</i>	10/19	9/19

Notre recherche des séquences dans le génome des archaeobactéries est, de toute évidence, incomplète et préliminaire. Le plus probable est que la protéine soit capable de reconnaître beaucoup de séquences différentes avec une très faible spécificité. Si de telles séquences étaient proches dans un génome il pourrait y avoir des régions plus « riches » en protéines que d'autres (cf. résultats de Caroline Teyssier).

#### Que peut reconnaître la protéine dans les séquences sélectionnées ?

Deux remarques peuvent être faites :

- compte tenu de la faible spécificité de reconnaissance de la protéine pour la séquence consensus, le plus probable est qu'elle n'établisse pas d'interaction spécifique avec les bases.
- la plupart des séquences sélectionnées et la séquence consensus ont la même organisation : une répétition de 4 à 5 adénines côté 5', une partie centrale riche en dinucléotide CA et une répétition de 3 ou 4 cytosines en 3'. Les caractéristiques structurales particulières de chacune de ces parties sont vraisemblablement utilisées par la protéine pour effectuer une lecture indirecte des séquences.

*Répétitions d'adénines :*

Il est connu depuis plus d'une vingtaine d'années que des ADN comprenant des répétitions de 4 à 6 adénines présentent une courbure intrinsèque. De nombreuses études montrent que cette courbure peut jouer un rôle important dans différents processus biologiques comme l'initiation de la transcription ou la compaction de l'ADN. La détermination de la structure de ces ADN a donc fait l'objet de nombreuses études par RMN, cristallographie et modélisation moléculaire (pour revue, Crothers et Shakked, 1999). Outre une courbure, ces séquences présentent différentes caractéristiques : 1) petit sillon de plus en plus étroit de 5' vers 3' ; 2) angle de propeller twist important (ce qui rend l'ADN rigide ; El Hassan et Calladine, 1996) ; 3) présence d'eau et d'ions à des positions préférentielles au niveau du petit sillon. L'amplitude de la courbure varie entre 9 et 28° suivant les études. Lorsqu'un ADN possède plusieurs répétitions d'adénines en phase avec le pas de l'hélice la courbure devient observable par différentes méthodes comme la circularisation de fragments par une ligase et l'électrophorèse. Ainsi des études menées par électrophorèse sur gel ont montré que les séquences bordant les répétitions d'adénines n'influencent que de  $\pm 10\%$  l'amplitude de la courbure (Haran et coll., 1994). Des études à l'aide de différents types de microscopies ont également été réalisées (Cognet et coll., 1999 ; Scipioni et coll., 2002), ces méthodes ont l'avantage de pouvoir appréhender la part due à la flexibilité de l'ADN dans les phénomènes observés.

Deux modèles ont été proposés pour expliquer la courbure intrinsèque des répétitions d'adénines : le modèle dénommé « wedge » proposé par le groupe de Trifonov et le modèle dénommé « jonctions » proposé par le groupe de Crothers. Dans le modèle « wedge » (Trifonov et Sussman, 1980), chaque pas de la répétition d'adénine présente une courbure. Cette courbure est principalement due à un angle de roll négatif (l'angle de tilt peut être soit négatif soit positif). La modification de ces angles provoque une courbure à chaque pas dans une direction qui comprime le petit sillon de l'ADN. Dans le modèle de « jonction » (Wu et Crothers, 1984), la courbure a lieu en 5' et en 3' de la répétition d'adénines qui est considérée comme étant droite. Les causes de cette courbure ne sont pas identiques de part et d'autre de la répétition d'adénines : angles de roll et de tilt positifs du côté 5' (compression du grand sillon) ; angle de tilt négatif du côté 3' (compression du squelette phosphodiester du brin contenant les adénines). Dans une telle situation, la courbure globale de l'ADN est équivalente à une compression du petit sillon (si l'on se place au milieu de la répétition d'adénines). Il est important de noter que ni le modèle de

« wedge » ni le modèle de « jonction » ne permettent d'expliquer l'ensemble des structures obtenues par RMN ou par cristallographie. Ainsi, Crothers et Shakked proposent un modèle intermédiaire dans lequel les séquences comportant une répétition d'adénines possèdent un roll négatif à l'intérieur de la répétition et un roll positif au niveau des jonctions avec les séquences bordantes (Crothers et Shakked, 1999).

Les protéines peuvent différencier les deux types de courbures dues à la présence de répétitions d'adénines. Ainsi le facteur de transcription LrpC ne reconnaît que les répétitions de type  $C(A)_n$  en phase avec l'ADN et dont la courbure correspond au modèle de type « jonction » (Beloin et coll., 2003).

#### *Région riche en C*

Les études réalisées par électrophorèse sur gel montrent que les ADN comprenant des répétitions de 4 à 5 guanines en phase (donc de cytosines sur le brin complémentaire) ne sont pas courbes (Koo et coll., 1986). Mais les données obtenues sur la chromatine montrent que ces séquences participent cependant à la compaction de l'ADN en facilitant la compression du grand sillon (McCall et coll., 1985 ; Drew et Travers, 1985). Par ailleurs des études par dichroïsme circulaire et par spectrométrie infrarouge à transformée de Fourier montrent que les régions contenant des répétitions de cytosines ont une facilité à réaliser des transitions ADN B/ADN A (Lindqvist et Gräslund, 2001 ; Dornberger et coll. 2001).

#### *di-nucléotides CA*

Des approches théoriques et expérimentales ont montré que les pas CpA sont particulièrement flexibles. Ils possèdent une énergie d'empilement des bases faible, un angle de roll positif et un angle de twist faible (Dickerson 1998 ; Parker et coll., 2000). Des expériences de circularisation de fragments montrent que plus le nombre de pas CpA est grand, plus la flexibilité anisotrope de l'ADN est importante (Lyubchenko et coll., 1993). Cette flexibilité des pas CpA est utilisée par différentes protéines (CAP, histones, répresseur lambda) pour induire des coudes dans la double hélice.

Une des spécificités des pas CpA est de favoriser la transition des phosphates de la conformation la plus courante dénommée BI ( $-160^\circ < \epsilon - \xi < 20^\circ$ ) à la conformation BII ( $20^\circ < \epsilon - \xi < 200^\circ$  ; les angles dièdres  $\epsilon$  et  $\xi$  sont montrés Figure 33). Cette conformation des phosphates perturbe les paramètres hélicoïdaux du di-nucléotide qui lui est associé : le roll devient négatif, l'angle de twist augmente et les bases sont déplacées vers le grand sillon.

De plus, les deux oxygènes non estérifiés des phosphates changent de conformation, ce qui modifie leur accessibilité aux ligands. Dans cette conformation, on observe une compression du petit sillon, et une diminution de la profondeur du grand sillon (Djuranovic et Hartmann, 2003 ; Djuranovic et Hartmann, 2004). Ces études ont également montré que la probabilité de trouver deux phosphates en conformation BII se faisant face (c'est à dire sur un même plateau de paire de base) est plus grande que celle de rencontrer deux phosphates adjacents en conformation BII . Des expériences de RMN (Tisné et coll., 1998) et de simulation de dynamique moléculaire (Djuranovic et Hartmann, 2004) ont montré que la fixation de la protéine NFκB est influencée par la conformation des phosphates. Ainsi, la présence de phosphates en conformation BII/BII au niveau des paires de bases TpG•CpA et CpA•TpG (respectivement en 5' et 3' du site de fixation) confère à l'ADN une grande malléabilité qui lui permet de prendre une conformation adaptée à la fixation de la protéine.

#### Etude par simulation de dynamique moléculaire.

Afin d'obtenir des informations directes sur la structure et la flexibilité de la séquence consensus, une étude par simulation de dynamique moléculaire a été réalisée par Samia ACI (collaboration avec l'équipe de D. Genest). Pour ces expériences *in silico*, un duplex d'ADN contenant la séquence consensus (Figure 32) est entouré par des contre-ions et une boîte d'eau. Après équilibration du système à 300 K, la simulation est effectuée pendant 5 ns à température et pression constantes. Les valeurs moyennes des coordonnées obtenues pour la dernière nanoseconde sont analysées à l'aide du logiciel CURVE (Lavery et Sklenar, 1989).

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
brin sup	C	p	A	p	A	p	A	p	A	p	C	p	A	p	C	p	C
brin inf	G	p	T	p	T	p	T	p	T	p	G	p	T	p	G	p	G

**Figure 32 : duplex utilisé pour la simulation de dynamique moléculaire.**

Les bases notées sur fond vert sont localisées dans la région la plus courbe de l'ADN. Les guanines notées en rouge possèdent leur N7 accessible au solvant. Les bases sont numérotées par rapport à l'extrémité 5' du brin supérieur de la séquence consensus. Les bases de la position 0 sont ajoutées pour permettre la stabilisation du fragment lors de la simulation. Les phosphates en noir alternent entre la conformation BI et la conformation BII.



sillon autour de G12 (2-5 Å au lieu de 7-11 Å). A ce niveau du site le petit sillon est donc profond et étroit, ce qui favorise certainement l'interaction avec la protéine (voir. chapitre suivant).

La simulation de dynamique moléculaire montre également que certains dinucléotides ont leur phosphate qui alterne entre les conformations BI et BII. Six phosphates sont concernés par cette transition au niveau des pas : A2pA3 et C8pA9 sur le brin supérieur et G13pG12, G12pT11, T11pG10, G10pT9 sur le brin inférieur. Aucune configuration de type BII/BII n'est observée. Ceci est en accord avec les données de la littérature montrant que dans un environnement 5' RCpAY 3 (ce qui est le cas pour les 3 pas CpA du site), les conformations BII/BII se faisant face sont rares. On observe par contre que 4 phosphates adjacents sont en conformation BII.

Il est vraisemblable que la protéine MC1 reconnaisse certaines des caractéristiques structurales de ces séquences comme la dimension particulière du petit sillon et la conformation des phosphates.

Il serait maintenant intéressant de déterminer l'implication des phosphates dans l'interaction (substitution par des méthylphosphonates, éthylation) en particulier pour savoir si les phosphates en équilibre entre les conformations BI et BII interagissent avec la protéine. Une telle implication des transitions BI/BII a été proposée pour expliquer le processus de lecture indirecte des séquences par le répresseur de l'opéron tryptophane (Wellenzohn et coll., 2002). Les auteurs montrent que les transitions arrivent de façon synchrone avec l'élaboration et la destruction des liaisons hydrogène.



## **ETUDE DU MODE DE FIXATION DE LA PROTEINE**

## 1) Principe de la méthode :

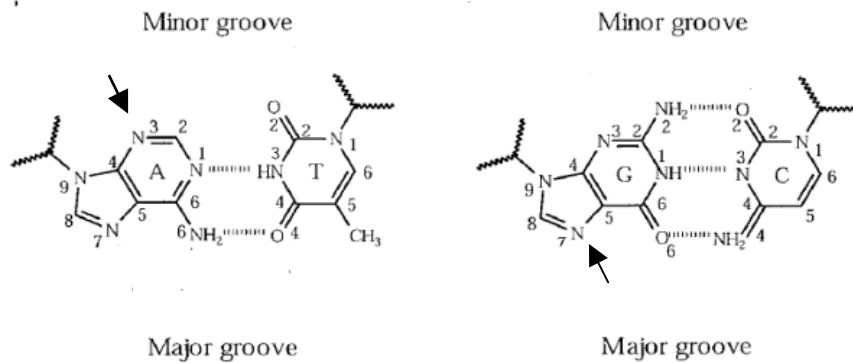
Afin de déterminer les caractéristiques moléculaires de l'interaction MC1/séquence consensus, nous avons utilisé différentes méthodes d'empreinte moléculaire. Dans de telles expériences, l'ADN subit un traitement modéré (à l'aide d'une sonde chimique ou enzymatique) qui, directement ou indirectement, conduira à la coupure du squelette phosphodiester. Il existe un grand nombre de sondes qui ont toutes des propriétés différentes. Certaines modifient l'ADN par la fixation d'un groupement chimique (traitement au diméthylsulfate par exemple), d'autres comme les radicaux hydroxyle vont attaquer directement le squelette de la double hélice. Le principe de base de ces méthodes est que l'ADN complexé et l'ADN libre vont réagir différemment. Certains nucléotides deviendront moins réactifs parce que la protéine gênera l'action de la sonde alors que d'autres deviendront plus réactifs car la fixation de la protéine provoquera des modifications structurales rendant certaines positions plus accessibles. L'analyse se fait par comparaison, sur gel de polyacrylamide dénaturant, de la réactivité obtenue pour chaque nucléotide en présence et en absence de protéine.

Une méthode complémentaire de l'empreinte moléculaire est l'interférence. Dans ce cas, le traitement avec la sonde chimique a lieu avant la formation des complexes protéine/ADN. Alors que les expériences d'empreinte identifient essentiellement des proximités, les interférences peuvent permettre de définir plus précisément les bases (ou les phosphates) intervenant dans le processus d'interaction.

## 2) Empreintes et interférences au DMS :

### 2.1) Généralités :

Le diméthylsulfate (DMS) est un agent chimique qui méthyle spécifiquement les purines (Figure 34). Le DMS réagit avec le N3 des adénines par le petit sillon et avec le N7 des guanines par le grand sillon. La fixation du groupement chimique sur les purines rend ces bases labiles en milieu basique. Le traitement au DMS est donc suivi d'un traitement à la soude ou à la pipéridine afin de permettre la  $\beta$  élimination de la base modifiée. Il en résulte une coupure de l'ADN en 3' de la base modifiée.



**Figure 34 : action du DMS sur l'ADN**

Les positions méthylées par le DMS sont repérées par une flèche.

## 2.2) Les empreintes :

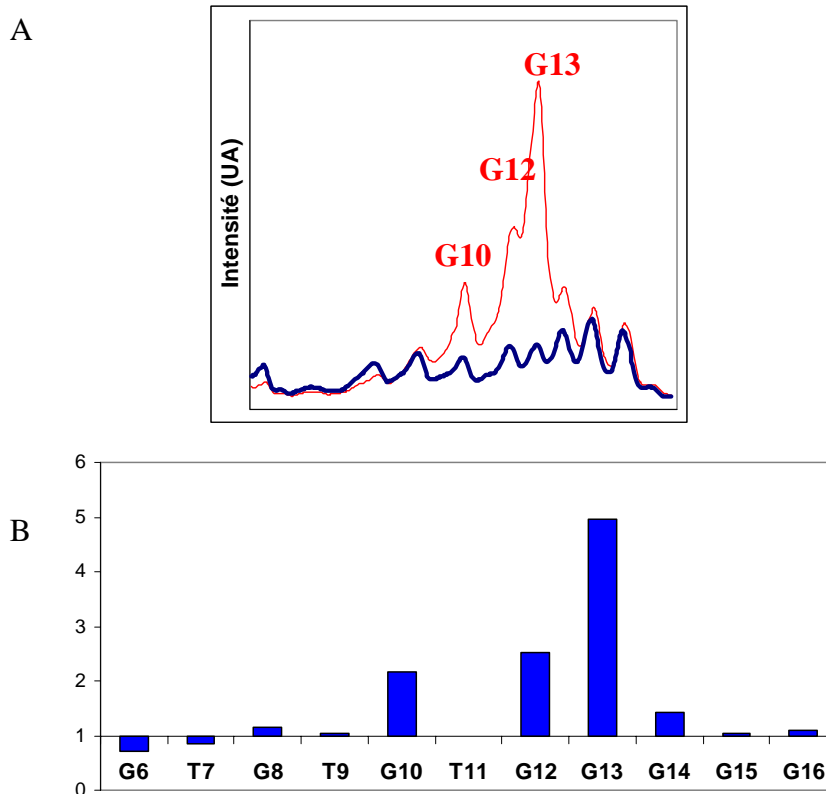
Les expériences ont été réalisées avec un fragment de 59 pb contenant la séquence consensus (séquence p60). Les conditions expérimentales choisies permettent de méthyler entre 10 et 20 % de l'ADN. A ce taux de méthylation, on peut estimer que la majorité des fragments ne contiennent qu'une seule base modifiée et qu'il n'y aura, en moyenne, qu'une seule coupure par fragment.

Nos essais avec le brin supérieur contenant les adénines se sont avérés infructueux, et les conditions permettant d'obtenir un niveau satisfaisant de méthylation des adénines n'ont pas été obtenues. Ceci est en partie dû au fait que les adénines sont beaucoup moins réactives au DMS que les guanines (en moyenne 5 fois moins). Les résultats ne sont donc présentés que pour le brin inférieur qui comprend 7 guanines dans la séquence consensus.

Le gel présenté Figure 35 montre le résultat de l'empreinte laissée par la protéine MC1. La comparaison du profil de coupure en présence et en absence de protéine montre clairement que certaines guanines de la séquence consensus deviennent hypersensibles au DMS en présence de protéine.



L'absence de protection indique que les azotes en position 7 des guanines (pointés côté grand sillon), restent accessibles au DMS. La protéine MC1 n'est vraisemblablement pas fixée dans le grand sillon de l'ADN. A l'inverse, les amplifications montrent que le changement de conformation de l'ADN rend le côté grand sillon de trois guanines plus accessible au DMS (par une ouverture du grand sillon par exemple).



**Figure 36 : analyse quantitative de l'expérience d'empreinte au DMS.**

Les pistes (0 et 2) du gel de la Figure 35 correspondant à la méthylation de l'ADN en absence et en présence de protéine (25 nM) ont été quantifiées. A : densitogrammes. En rouge (trait fin) : profil en absence de protéine. En bleu (trait épais) : profil en présence de protéine. B : histogramme correspondant au rapport des probabilités de coupures en présence et en absence de protéine.

Le  $\text{KMnO}_4$  a la propriété d'oxyder la double liaison 5,6 des thymines des ADN dont la structure est fortement modifiée par rapport à une forme B classique de l'ADN. Après action du  $\text{KMnO}_4$ , un traitement alcalin permet de couper l'ADN en 3' des thymines qui ont été oxydées.

Nous avons utilisé cette méthode en particulier pour avoir accès à la conformation de la thymine en position 11 localisée entre deux guanines hypersensibles au DMS. Le résultat de cette expérience (non montré) ne fait apparaître aucune différence de réactivité entre l'ADN libre et l'ADN en présence de protéine. Ce résultat surprenant suggère que la conformation de la thymine 11 n'est pas (ou peu) affectée par l'interaction avec la protéine.

### 2.3) Les interférences :

Pour compléter cette étude nous avons effectué des expériences d'interférence au DMS. Comme pour les empreintes, les expériences ont été réalisées avec le brin inférieur qui comprend les guanines. Dans ces expériences, le taux de méthylation a été de 20 %. Les complexes ont ensuite été réalisés avec un large excès d'ADN permettant à la protéine de choisir entre les fragments méthylés et ceux non méthylés. Après purification et traitement alcalin les ADN (mélange initial et ADN complexés) ont été séparés sur un gel de séquence dénaturant.

La Figure 37 montre le résultat obtenu dans des conditions où seulement 16 % de l'ADN est complexé. La partie B de la figure montre la quantification des résultats et compare les coupures obtenues sur les fragments de départ et sur les fragments choisis par la protéine. Pour chaque position le rapport des probabilités de coupure est calculé. Un rapport inférieur à 1 indique que la protéine choisit d'interagir avec les fragments non méthylés, et qu'un groupement méthyle sur le N7 des guanines défavorise l'interaction.

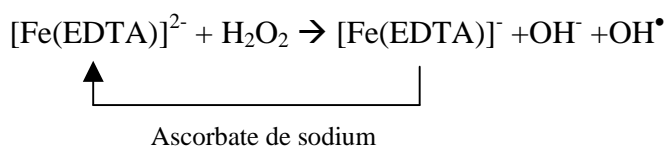
La méthylation des guanines 6, 8, 10 et 15 n'affecte pas l'interaction alors que la méthylation des guanines 12 à 14 défavorise l'interaction. Ce résultat peut paraître contradictoire avec celui obtenu par la méthode d'empreinte. Cependant l'interprétation la plus probable de ces résultats est que la méthylation des 3 guanines gêne l'interaction, non par un effet direct mais, parce qu'elle réduit la flexibilité de la double hélice, empêchant ainsi un changement de conformation de cette zone du site (changement entraînant une hypersensibilité des guanines au DMS). Au total il ressort de ces expériences que la protéine ne se fixe probablement pas dans le grand sillon de l'ADN.



Cette méthode d’empreinte présente plusieurs avantages :

- les coupures sont, pour l’essentiel, indépendantes de la séquence (les ADN courbes avec le petit sillon moins accessible ont cependant des taux de coupure plus faibles).
- grâce à leur petite taille les radicaux  $\text{OH}^\bullet$  produisent une empreinte qui reflète l’accessibilité du solvant et dont la résolution est très forte (1 pb).
- les réactions peuvent être réalisées dans des conditions expérimentales variées (force ionique, pH, température). Seuls les « scavengers » de radicaux  $\text{OH}^\bullet$  (comme le glycérol ou les thiols) peuvent considérablement réduire les taux de coupure.

Pour réaliser notre étude, nous avons utilisé les radicaux  $\text{OH}^\bullet$  produits par la réaction de Fenton :



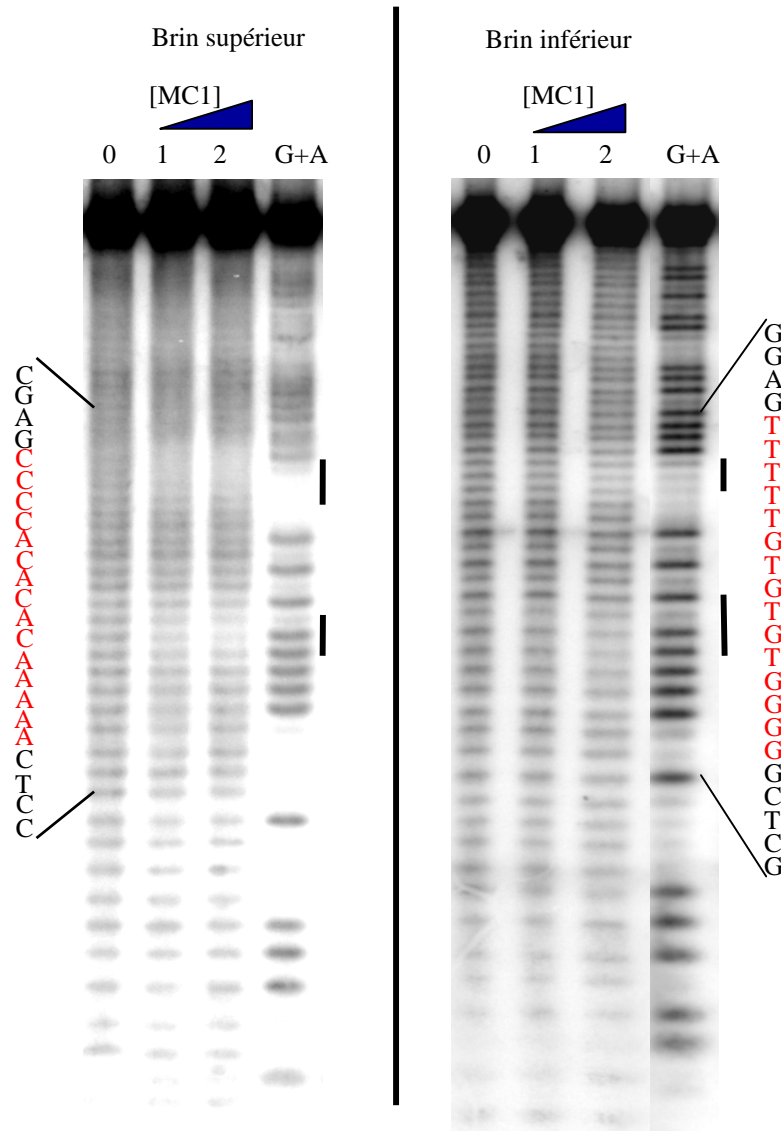
Dans cette réaction, le complexe Fe/EDTA est utilisé pour réduire la liaison O-O du peroxyde d’hydrogène ( $\text{H}_2\text{O}_2$ ). Au cours de cette réduction, le complexe Fe/EDTA va passer d’un état FeII à FeIII, et le peroxyde d’hydrogène va fournir des radicaux hydroxyle ( $\text{OH}^\bullet$ ) et des ions hydroxyde ( $\text{OH}^-$ ). L’addition d’ascorbate de sodium permet, par la réduction du FeIII en FeII, la création d’un cycle catalytique de telle sorte que de faibles concentrations en réactif sont nécessaires pour obtenir un profil de coupure suffisant.

### 3.2) Résultats :

La Figure 38 montre le résultat d’une expérience d’empreinte par les radicaux  $\text{OH}^\bullet$ . Les coupures ont été réalisées sur un fragment contenant la séquence consensus en absence et en présence de protéine.



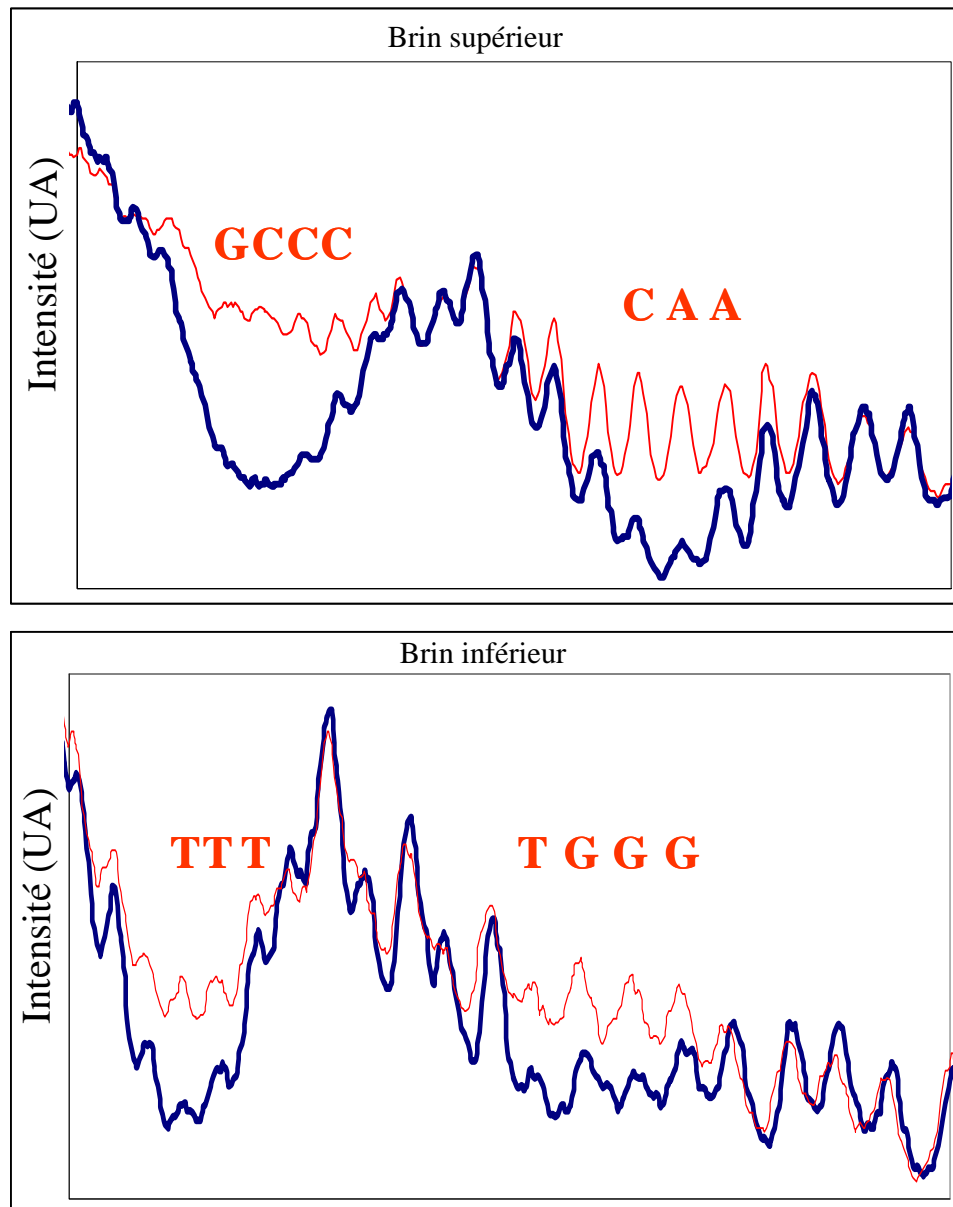
Sur chacun des deux brins deux empreintes sont visibles. Ces deux zones de protection apparaissent toujours de manière concomitante et dans des conditions où un seul complexe est visible sur les gels retard. Elles sont donc causées par la fixation d'une seule protéine.



**Figure 38 : empreintes aux radicaux OH•.**

Autoradiographie d'un gel de polyacrylamide montrant les empreintes observées sur les deux brins de la séquence consensus. La piste 0 est obtenue par l'attaque du fragment d'ADN en absence de protéine et les pistes 1 et 2 sont obtenues en présence respectivement, de 25 et 50 nM de protéine. Les zones protégées de l'attaque des radicaux OH• sont repérées par un trait gras. La séquence des ADN est reportée à côté de chaque gel. La séquence consensus est notée en rouge. Les complexes sont réalisés dans un tampon contenant 60 mM de NaCl. La piste G+A correspond au séquençage de purines selon la méthode de Maxam et Gilbert.

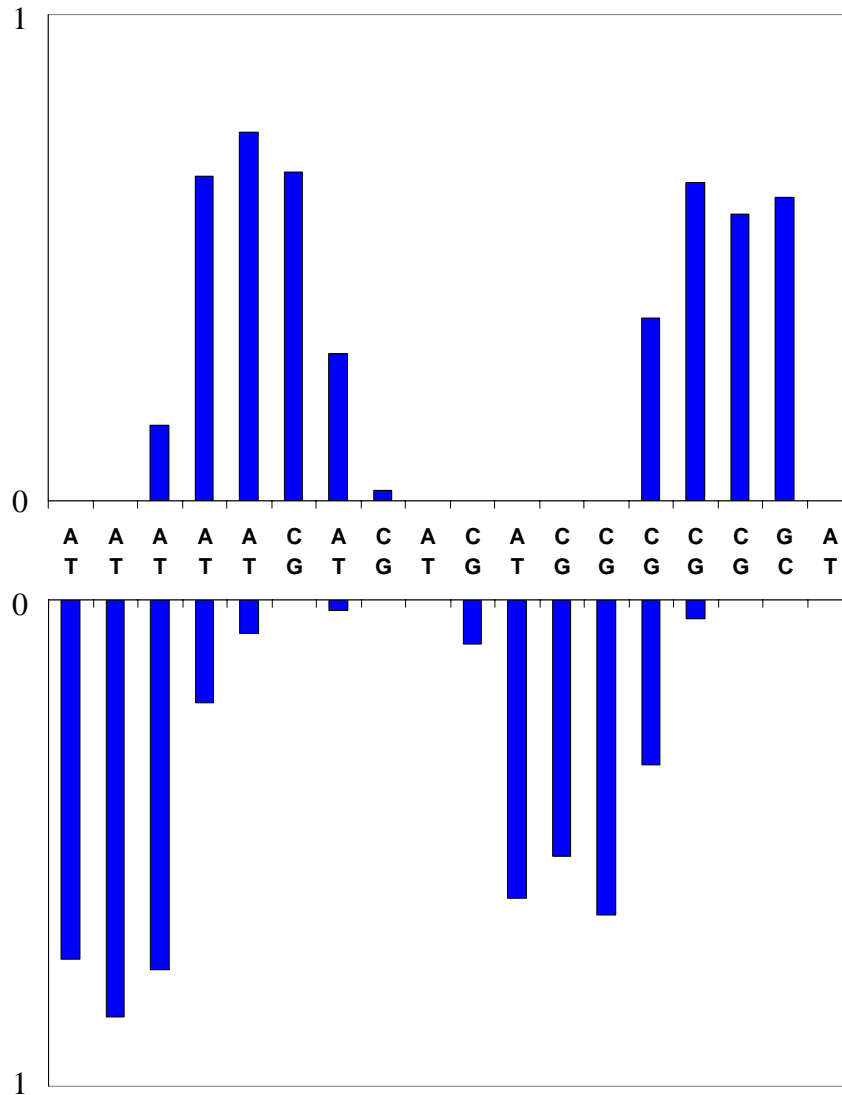
La Figure 39 montre les densitogrammes des gels et permet de comparer les coupures obtenues en absence et en présence de protéine.



**Figure 39 : densitogrammes des gels présentés Figure 38.**

En rouge (trait fin) : profil en absence de protéine. En bleu (trait épais) : profil en présence de protéine (50 nM).

Sur chacun des deux brins, 2 importantes protections respectivement de 3 et 4 nucléotides (5' vers 3') sont visibles.



**Figure 40 : analyse quantitative des données de la Figure 39.**

Les barres représentent le rapport  $(D-C)/D$  où D est l'aire de chaque pic d'ADN libre, et C l'aire du même pic en présence de protéine.

La Figure 40 montre qu'il y a un décalage entre les protections des deux brins. Pour les deux zones protégées, ce décalage est de 3 ou 4 bases vers le côté 3'. Ce profil de protection est caractéristique d'une empreinte au niveau du petit sillon.

Nous ne pouvons pas exclure que ces deux empreintes soient causées par une diminution de l'accessibilité des sucres due à un rétrécissement du petit sillon (courbure

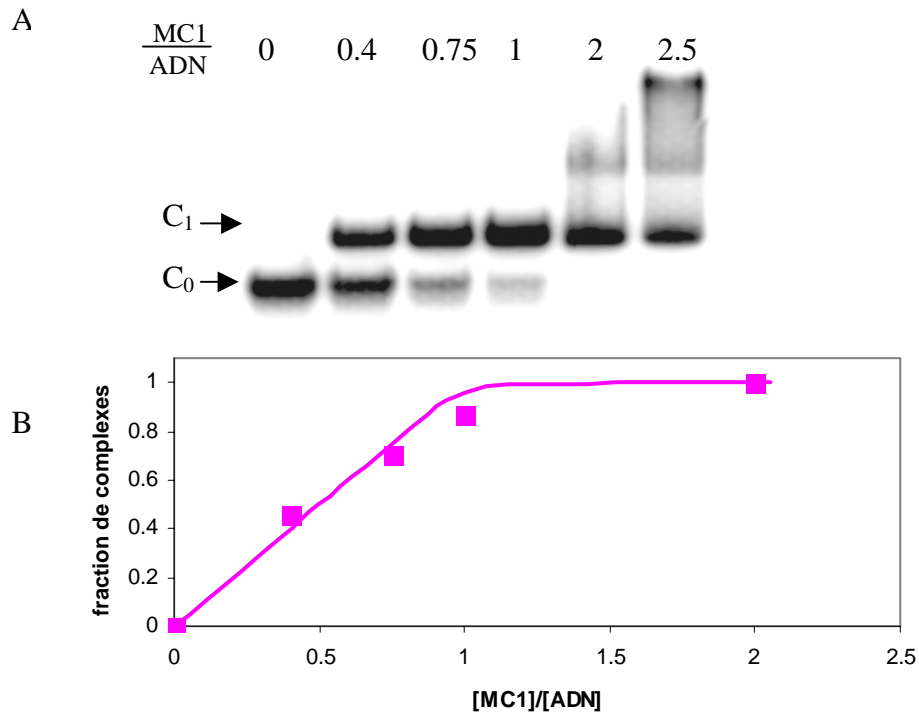
vers le petit sillon) et non directement à l'interaction de la protéine. Nous avons cependant considéré qu'une telle hypothèse était peu probable (à notre connaissance, aucune conclusion de ce type n'est donnée dans la littérature). De plus, des expériences précédemment réalisées au laboratoire ont montré que la fixation de la distamycine dans le petit sillon rentre en compétition avec la fixation de la protéine MC1.

### 3.3) Stœchiométrie de l'interaction MC1/séquence consensus :

A ce stade de notre travail, nous nous sommes demandés si nos résultats étaient compatibles avec une fixation monomérique de la protéine MC1. En effet, l'empreinte aux radicaux OH<sup>•</sup> (et la séquence consensus) est étonnamment étendue pour une protéine de la taille de MC1.

Nous avons alors pensé que ces résultats pouvaient être dus à la fixation d'un dimère et non à celle d'un monomère.

Pour vérifier ce point nous avons déterminé la stœchiométrie de l'interaction MC1/séquence consensus. Dans l'expérience montrée Figure 41, un duplex de 26 pb comportant la séquence consensus est mis en présence de concentrations croissantes de protéine. Les conditions (forte concentration en ADN) permettent à chaque protéine de se fixer à l'ADN. L'analyse quantitative du gel montre que la saturation du fragment se produit pour un rapport de un monomère pour un fragment : l'interaction MC1/séquence spécifique a lieu avec une protéine monomérique (comme montré précédemment par Culard et collaborateurs (1993), dans une étude concernant l'interaction MC1/séquences aléatoires).



**Figure 41 : titration d'un fragment de 26 pb contenant la séquence consensus par des concentrations croissantes de protéine MC1.**

A : gel retard montrant l'apparition progressive des complexes. L'interaction a été réalisée en présence de 60 mM de NaCl (tampon TE). La concentration en fragment est de 2  $\mu$ M. B : analyse quantitative du gel montrant la fraction d'ADN complexé en fonction du rapport protéine/fragment. La concentration en protéine est calculée en considérant que la protéine est monomérique.

#### 4) Discussion :

Le schéma de la Figure 42 reprend les résultats obtenus avec le DMS et les radicaux OH $\cdot$ . Pour chaque brin, les protections sont localisées en deux zones et au niveau du petit sillon. Les deux zones protégées étant séparées de 10 à 11 pb (de milieu à milieu), on peut penser que la fixation de la protéine se fait par une face de l'ADN. Du côté 3' de la séquence consensus, la fixation de la protéine dans le petit sillon rend le grand sillon opposé plus accessible. La région centrale ne semble pas interagir avec la protéine.



à une compression du petit sillon et un élargissement du grand sillon opposé. Dans le premier cas la protéine est localisée à l'extérieur de la courbure et c'est la configuration inverse qui existe pour le deuxième cas. Dikerson (1998) définit plusieurs modes de courbure et différencie une courbure continue sur l'ensemble du site de fixation (TBP, MAT $\alpha$ 1/MAT $\alpha$ 2) d'une courbure localisée et brutale du fait de la formation d'un ou deux coude(s) dans la double hélice (IHF, CAP). Il montre aussi comment chacune des familles de courbures induites peut donner lieu à des courbures planes (CAP, MAT $\alpha$ 1/MAT $\alpha$ 2) ou non (TBP, IHF).

D'après nos résultats, il semble clair que la fixation de MC1 induit la formation d'au moins un coude dans la double hélice (dans la partie 3' du site) qui comprime le petit sillon et élargit le grand sillon (3 guanines sont hypersensibles au DMS en présence de protéine). Le plus probable est donc que la protéine se trouve sur la face interne de la courbure.

Il serait intéressant de réaliser d'autres types d'empreintes (DNase I) afin de détecter la présence d'un deuxième coude qui pourrait expliquer comment la protéine arrive à courber si fortement l'ADN (Le Cam et coll., 1999). La présence de deux coudes permettrait également d'expliquer les résultats de Larquet et collaborateurs (1996) montrant que la fixation de la protéine entraîne un vrillage des minicercles. La présence de deux coudes augmenterait les ressemblances avec la protéine IHF qui se fixe à l'ADN grâce à deux bras « étreignant » la double hélice (à l'extrémité de chaque bras se trouve une proline qui s'intercale dans l'ADN en provoquant deux coudes séparés de 9 pb qui compressent localement le petit sillon). Par la fixation dans le petit sillon, l'interaction par des bras en feuillets  $\beta$ , la faible spécificité de reconnaissance, la localisation à l'intérieur de la courbure, la protéine MC1 ressemble à IHF. Par contre, du point de vue de sa structure, la protéine MC1 ressemble plus aux protéines Sul 7d (elles sont de plus toutes deux monomériques) qui courbent l'ADN grâce à l'intercalation de résidus hydrophobes mais dans ce cas la protéine est à l'extérieur de la courbure (avec une compression du grand sillon). Nos résultats d'empreinte moléculaire ne semblent, par contre, pas compatibles avec ce mode de fixation.

**ETUDE DE L'EFFET DES RAYONNEMENTS IONISANTS  
SUR LES FIXATIONS PREFERENTIELLES DE MC1**

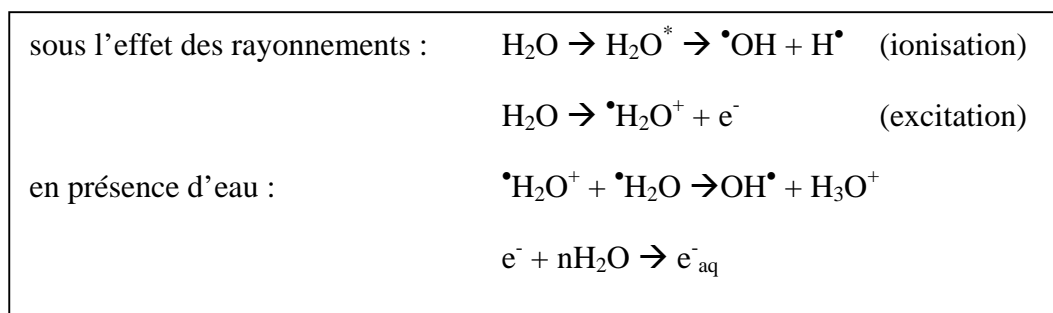


## 1) Introduction :

Différents travaux montrent que les rayonnements ionisants provoquent des dommages sur les protéines. Ainsi, les rayons  $\gamma$  détruisent le complexe formé entre le répresseur et l'opérateur de l'opéron lactose (Eon et coll., 2001). Dans cette étude, les auteurs ont montré qu'au sein du complexe, c'est la protéine qui est la plus sensible aux rayons  $\gamma$ . En effet, après irradiation, l'ajout de répresseur non irradié permet la reconstitution des complexes (ceci n'est pas observé si on ajoute de l'opérateur non irradié). D'autres résultats obtenus sur ce système montrent que l'inactivation de la protéine n'est pas due à des coupures de la chaîne peptidique par les rayons ionisants mais très probablement à des oxydations provoquées sur les chaînes latérales (Charlier et coll., 2002).

Un modèle prenant en compte la structure quaternaire du répresseur (association de deux dimères, chaque dimère possédant deux sites de fixation à l'ADN) a été construit pour expliquer cette inactivation (Charlier et coll., 2002). Ce modèle montre qu'au moins deux types de dommages par protomère sont nécessaires pour inactiver le répresseur.

L'eau contenue dans les solutions biologiques est très sensible aux rayonnements ionisants. Sa radiolyse donne de nombreuses espèces dont les plus réactives sont le radical hydroxyle ( $\text{OH}^\bullet$ ), l'électron solvaté ( $e^-_{\text{aq}}$ ) et le radical hydrogène ( $\text{H}^\bullet$ ) (Figure 43). Ces espèces peuvent réagir entre elles et avec les différentes molécules présentes dans les solutions. En présence d'oxygène, les agents réducteurs ( $\text{H}^\bullet$  et  $e^-_{\text{aq}}$ ) sont rapidement éliminés de sorte que le radical  $\text{OH}^\bullet$  est de loin l'espèce la plus réactive.



**Figure 43 : radiolyse de l'eau par les rayonnements ionisants (rayons X,  $\gamma$ ...).**

Les radicaux OH<sup>•</sup> ont la propriété d'oxyder les protéines en provoquant des coupures de chaîne peptidique, des pontages entre résidus et des oxydations au niveau des chaînes latérales des acides aminés. Au sein d'une protéine, le premier événement observé est l'oxydation de la chaîne latérale des acides aminés (Maleknia et coll., 1999). Des exemples de produits d'oxydation sont donnés dans le Tableau 8. Les phénomènes de pontage des acides aminés sont le résultat de la stabilisation des produits d'oxydation des résidus aromatiques (phénylalanine, tyrosine) ou soufrés (cystéine, méthionine). En utilisant des peptides de 5 à 14 résidus possédant des acides aminés modifiés sur leur chaîne latérale, l'ordre de réactivité des acides aminés vis à vis des radicaux OH<sup>•</sup> a été déterminé : Cys, Met >>>Phe, Tyr ≥ Trp > Pro > His, Leu (Maleknia et coll., 1999).

**Tableau 8 : principaux produits de réaction entre un radical OH<sup>•</sup> et certains acides aminés.**

ACIDES AMINES	PRODUITS DE REACTION
tyrosine	3, 4 dihydro-phenylalanine (DOPA)
	bityrosine
	tyrosine hydroperoxyde
phénylalanine	o-, m-tyrosine, tyrosine
	dimère
tryptophane	N-formyl cynurénine
	cynurénine
	5-hydroxytryptophane
	7-hydroxytryptophane
méthionine	méthionine sulfoxide
cystéine	cystine
proline	hydroxyproline
leucine	hydroxyleucine
valine	hydroxyvaline

(d'après Davies et Dean, 1997)

Un certain nombre de travaux ont montré que les rayonnements ionisants peuvent être utilisés pour obtenir des informations sur la structure des protéines (Maleknia et coll.,

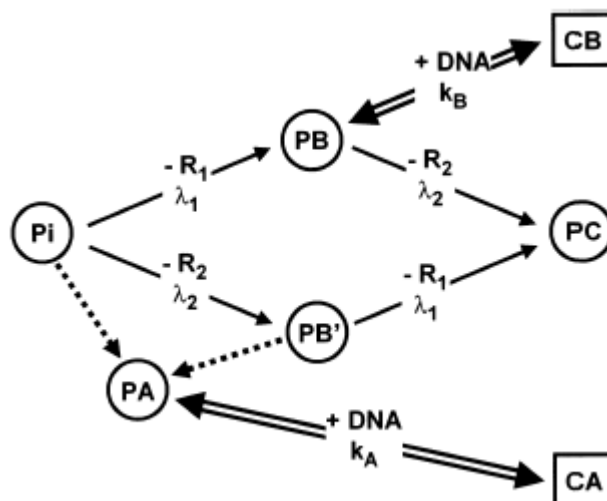
2001). Une technique, nommée « empreinte » permet de déterminer l'accessibilité des acides aminés au solvant au sein de la protéine. En effet, les observations ont montré que les réactions induites par les radicaux  $\text{OH}^\bullet$  sont modulées par l'accessibilité au solvant des différentes régions de la protéine. Les produits formés suite à la radiolyse étant le plus souvent stables, leur cinétique d'apparition pourra être suivie par une analyse en spectrométrie de masse. L'utilisation d'un « stop flow » couplé à un irradiateur permettant de réaliser de la radiolyse pulsée est intéressante. Ainsi, la cinétique du repliement d'une protéine a pu être étudiée, notamment lors de son association avec l'ADN ou après une modification de la force ionique du milieu (Maleknia et coll., 2001).

Des études par « empreinte » ont été réalisées sur la TATA binding protein entière (TBP ; Rashidzadeh et coll., 2003). Les résultats permettent de préciser les données obtenues par cristallographie et montrent que le domaine N terminal, que l'on pensait non structuré, adopte une conformation qui est dépendante de la présence d'ADN.

## 2) Résultats :

Au sein d'une collaboration avec l'équipe de Michel Charlier, les effets des rayonnements ionisants sur la protéine MC1 sont étudiés au laboratoire. Ces travaux ont donné lieu à un article figurant à la fin de ce mémoire. L'ensemble des résultats obtenus montre que la fixation de la protéine MC1 sur l'ADN est inactivée par les rayonnements ionisants. Les doses nécessaires à cette inactivation sont plus grandes si la protéine est irradiée au sein d'un complexe que si la protéine est irradiée seule. Cette protection importante de l'ADN se fait sans formation de pontage. Ces travaux ont permis d'établir un modèle en deux étapes expliquant l'inactivation de la protéine (Figure 44) :

- 1- Pour de faibles doses, la protéine reste active (PB) et un nouveau type de complexe est formé (CB). A ce stade, la constante d'équilibre d'association avec la protéine irradiée est environ 1.5 fois plus faible qu'avec la protéine native (respectivement  $6.8 \times 10^7 \text{ M}^{-1}$  et  $1 \times 10^8 \text{ M}^{-1}$ ).
- 2- Pour des doses plus fortes, la protéine perd totalement sa capacité d'interaction avec l'ADN (PC).



**Figure 44 : inactivation de la protéine MC1 par les rayons  $\gamma$ .**

Pi : protéine MC1 native ; PA : protéine donnant un complexe « natif » ; PB, PB' : protéine MC1 avec un dommage (respectivement résidu R<sub>1</sub> et R<sub>2</sub>) ; PC : protéine MC1 avec un dommage sur les résidus R<sub>1</sub> et R<sub>2</sub> ; CA : complexe « natif » ; CB : nouveau complexe ;  $\lambda_1, \lambda_2$  : vitesses de formation des produits (respectivement  $28 \times 10^{-4}$  et  $17 \times 10^{-4}$  Gy) ;  $K_{A, B}$  : constantes d'équilibres.

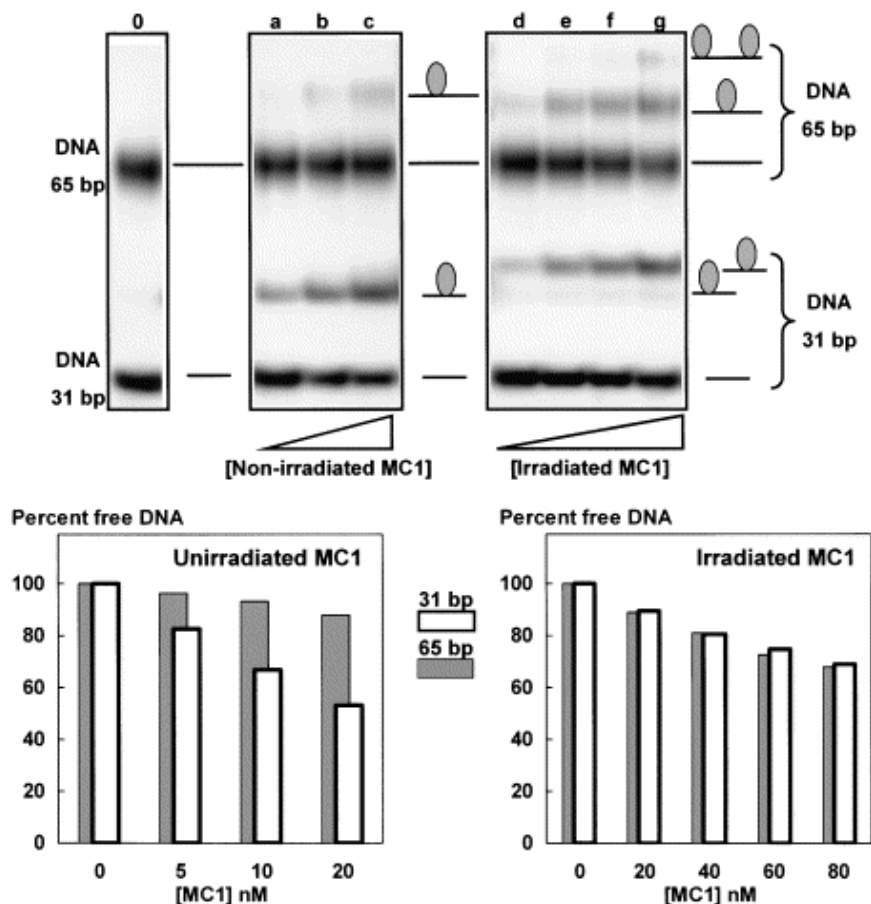
Le but de mon travail au sein de cette collaboration a été de déterminer l'effet des rayons  $\gamma$  sur les processus de reconnaissance de la protéine MC1. Comme ces processus sont dépendants de la séquence et de la structure de l'ADN, les fragments issus des expériences de SELEX, un fragment contenant le site I et des minicercles d'ADN ont été utilisés.

### 2.1) Effet des radiations $\gamma$ sur la reconnaissance MC1/séquence préférentielle :

Des compétitions directes entre deux fragments, l'un contenant un site préférentiel de fixation, l'autre de séquence aléatoire ont été réalisées. L'ADN ayant une séquence préférentielle de fixation (31 pb) est obtenu par digestion des fragments issus du cycle 10 des expériences de SELEX. L'ADN de séquence aléatoire (65 pb) correspond au mélange de fragments utilisé pour réaliser le premier cycle SELEX.

Comme montré précédemment, la protéine MC1 native se fixe préférentiellement sur l'ADN de 31 pb et peu de complexes sont obtenus avec l'ADN de séquence aléatoire de 65 pb (Figure 45). Avec la protéine irradiée, des concentrations plus fortes en MC1 sont nécessaires pour observer la formation de complexe. De façon surprenante, on observe que

les deux ADN disparaissent dans les mêmes proportions avec la protéine irradiée (partie droite de la figure), ce qui n'est pas le cas avec la protéine native (partie gauche de la figure). Ces résultats montrent que la protéine MC1 irradiée n'est plus capable de reconnaître préférentiellement les séquences issues de l'expérience de sélection *in vitro*.

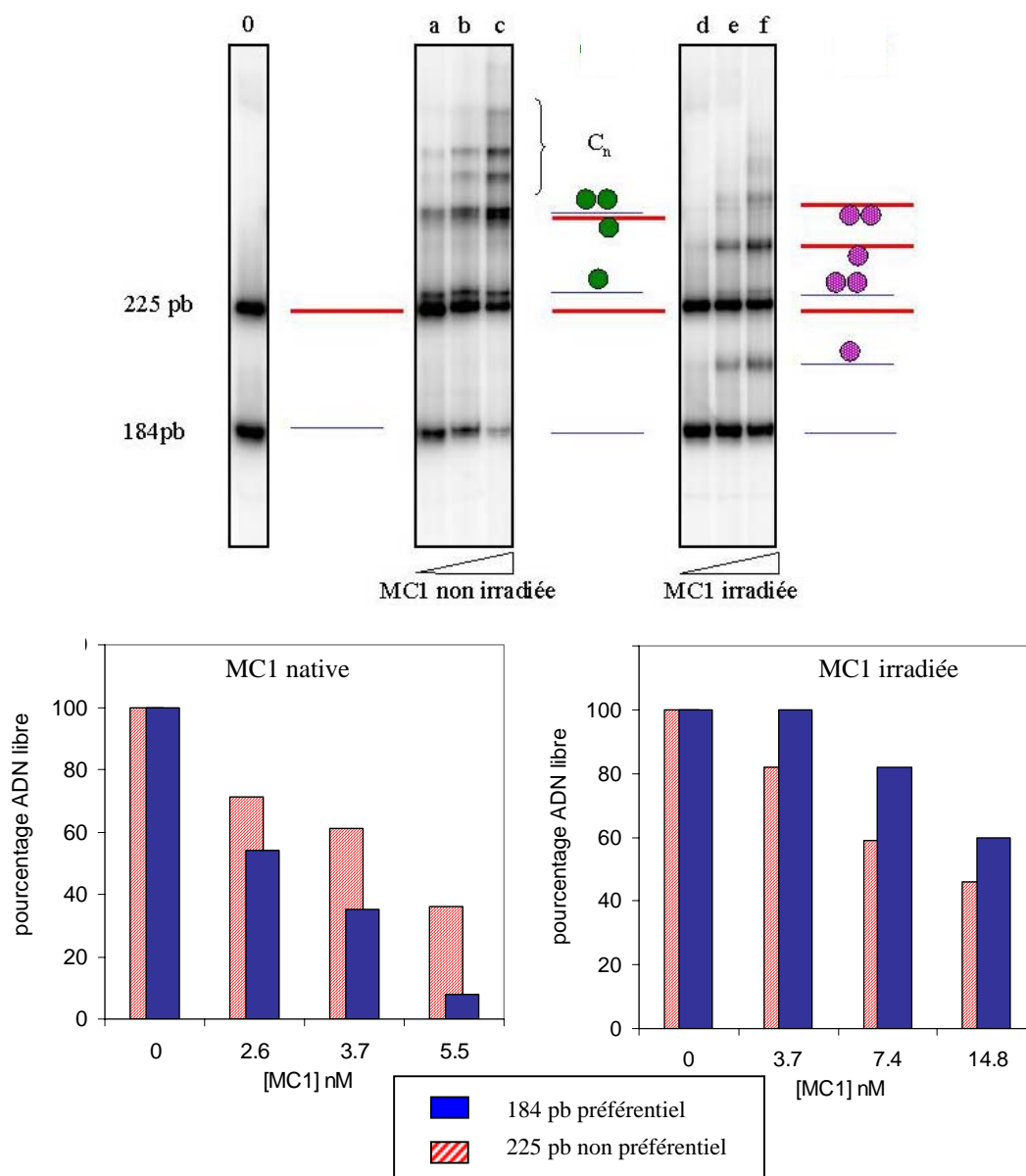


**Figure 45 : effet des rayons  $\gamma$  sur la reconnaissance des fragments issus du SELEX par MC1.**

En haut : gel retard montrant les résultats obtenus lorsque l'on incube la protéine MC1 irradiée ou non avec un mélange équimolaire (5 nM) de fragment de séquence préférentielle de 31 pb ou aléatoire de 65 pb. 0 : pas de protéine ; a, b et c : protéine native, concentrations respectives 5 ; 10 et 20 nM ; d, e et f protéine irradiée (125 Gy, 2  $\mu$ M), concentrations respectives 20 ; 40 ; 60 et 80 nM. En bas : histogramme montrant le pourcentage d'ADN libre (31 et 65 pb) en fonction de la concentration en protéine.

Le même type d'expérience a été réalisé avec un fragment de 184 pb contenant le site I et un fragment sans séquence préférentielle de 225 pb. L'analyse quantitative de ces expériences (bas de la Figure 46) montre que la protéine MC1 native se fixe préférentiellement sur l'ADN de 184 pb contenant le site I. En revanche, lorsque la protéine est irradiée, le phénomène inverse est observé. Cette expérience confirme que la

protéine irradiée n'est plus capable de reconnaître son site préférentiel. Le fait que la protéine irradiée se fixe légèrement plus sur l'ADN sans séquence préférentielle peut s'expliquer par la présence d'un plus grand nombre de sites potentiels sur cet ADN (environ 20 sites contre 16 sites pour l'ADN de 184 pb).



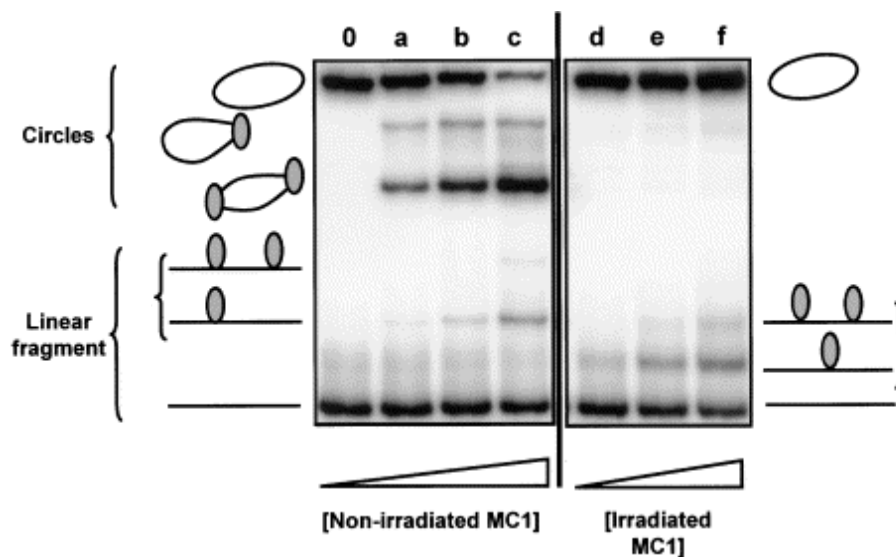
**Figure 46 : effet des rayons  $\gamma$  sur la reconnaissance du site naturel (site I) par MC1.**

En haut : gel retard montrant les résultats obtenus lorsque l'on incube la protéine MC1 irradiée ou non avec un mélange équimolaire (3.5 nM) de fragment avec ou sans site I (respectivement 184 pb et 225 pb). 0 : pas de protéine ; a, b et c : protéine native, concentrations respectives 2.6 ; 3.7 et 5.5 nM ; d, e et f protéine irradiée (300 Gy, 13.6  $\mu$ M), concentrations respectives 3.7 ; 7.4 et 14.8 nM. En bas : histogramme montrant le pourcentage d'ADN libre (31 et 65 pb) en fonction de la concentration en protéine.

2.2) Effet des radiations  $\gamma$  sur la reconnaissance préférentielle MC1/cercles :

Des expériences de compétition directe entre un ADN circulaire et un ADN linéaire pour la fixation de la protéine MC1 ont été réalisées. Les ADN utilisés ont une taille de 207 pb et ne contiennent pas de séquence préférentielle de fixation pour la protéine MC1.

Les résultats d'une expérience sont présentés Figure 47. Cette figure montre que la protéine native (partie gauche du gel) ne se fixe que très peu sur l'ADN linéaire. Comme cela a déjà été démontré dans des travaux précédents (Toulmé et coll., 1995 ; Teyssier et coll., 1996), la protéine MC1 a une grande affinité pour l'ADN circulaire et une majorité de complexes contenant deux protéines est formée. Lorsque la protéine MC1 est irradiée (partie droite du gel), aucun complexe n'est obtenu avec les ADN circulaires alors que des complexes comprenant une ou deux protéines fixées sur l'ADN linéaire sont visibles. Ces expériences montrent que non seulement la protéine irradiée n'est plus capable de reconnaître préférentiellement les minicercles d'ADN, mais aussi que ces structures courbes sont devenues de très mauvais ligands.



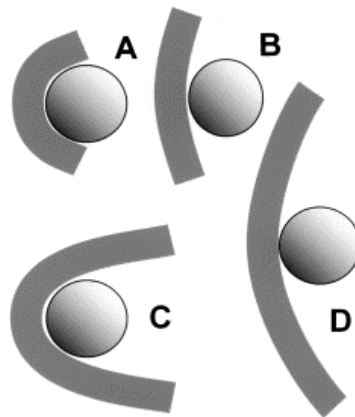
**Figure 47 : effet des rayons  $\gamma$  sur la reconnaissance des ADN circulaire par MC1.**

La protéine MC1 irradiée ou non est incubée avec un mélange contenant une concentration équimolaire (1.5 nM) d'ADN circulaire de 207 pb et d'ADN linéaire de 207 pb. 0 : ADN seul ; a, b et c : protéine native, concentrations respectives de 4.5 ; 9 et 18 nM ; d, e et f : protéine irradiée (125 Gy, 2  $\mu$ M), concentrations respectives 50 ; 100 et 200 nM.

2.3) Effets des rayonnements ionisants sur la structure du complexe MC1/ADN :

La fixation préférentielle d'une protéine sur les ADN circulaires est couramment admise comme étant une preuve que la fixation de la protéine induit une courbure de l'ADN (Sitlani et Crothers, 1998). La protéine MC1 irradiée ne se fixant plus sur les minicercles d'ADN, on peut penser qu'elle n'est plus capable de courber l'ADN. Afin de vérifier cette hypothèse, des expériences de circularisation d'un ADN de 86 pb par la ligase T4 en présence de protéine MC1, irradiée ou non, ont été réalisées au laboratoire (cf. article en annexe). Ces expériences, montrent que la proportion de cercles obtenus en présence de protéine MC1 irradiée est très nettement inférieure à la proportion de cercles obtenus en présence de protéine native (même pour des concentrations fortes en protéine). La capacité de la protéine MC1 irradiée à courber l'ADN est donc beaucoup plus faible que celle de la protéine native.

Dans ces conditions, comment expliquer les résultats obtenus avec les fragments de 31 pb issus du SELEX (complexe obtenu avec la protéine native migrant plus vite que le complexe obtenu avec la protéine irradiée ; Figure 45) ?



**Figure 48 : dessin représentant des complexes entre la protéine MC1 et des ADN de différentes longueurs possédant différents angles de courbures.**

(A) et (B) : fragment d'ADN court. (A) possède une courbure plus importante que (B) mais il migre plus vite car il a une forme globulaire. (C) et (D) : fragment d'ADN long. (D) possède une courbure moins importante que (C) et il migre plus vite car son coefficient de diffusion dans le gel est plus petit.

Une explication possible aux résultats obtenus avec les fragments de 31 pb est présentée Figure 48. Lorsqu'une protéine se fixe sur un ADN court en induisant une forte



courbure, un complexe globulaire est obtenu (A). La vitesse de migration dans le gel d'un tel complexe est plus grande que celle d'un complexe possédant une courbure plus faible (B). En revanche, comme montré par le groupe de Crothers (Wu et Crothers, 1984), lorsqu'une protéine se fixe sur un ADN long, la migration est d'autant plus lente que la courbure induite est grande (D migre plus vite que C). Des expériences complémentaires ont montré que pour un ADN de 50 pb, la migration du complexe obtenu avec la protéine irradiée est identique à celle du complexe formé avec la protéine native. Les parties A et B de la Figure 48 peuvent correspondre à l'association du fragment de 31 pb respectivement avec la protéine native et la protéine irradiée. Les parties C et D de cette figure peuvent correspondre à l'association des ADN de 65, 184, 207 et 225 pb respectivement avec la protéine native et la protéine irradiée.

### **3) Discussion :**

L'ensemble des résultats obtenus permet de compléter le modèle présenté Figure 44. A faible dose, on peut constater que la protéine garde sa capacité à fixer l'ADN mais la reconnaissance préférentielle des séquences et des minicercles d'ADN est perdue. Dans ces conditions, on observe un nouveau type de complexe dans lequel la courbure de l'ADN est plus faible que dans les complexes obtenus avec la protéine native. A forte dose, la protéine MC1 perd totalement sa capacité à former des complexes avec l'ADN.

L'oxydation d'un (ou plusieurs) acide(s) aminé(s) est responsable de cette inactivation partielle de la protéine (la spectrométrie de masse et les gels SDS/PAGE montrent qu'il n'y a pas coupure de la protéine). Il est donc vraisemblable que cet (ou ces) acide(s) aminé(s) soi(en)t directement ou indirectement (via la structure de la protéine) nécessaire(s) à la reconnaissance des séquences et à la courbure de l'ADN.

Quels peuvent être les acides aminés responsables de cette inactivation ? MC1 ne comporte ni cystéine ni tyrosine qui sont très réactives aux radicaux  $\text{OH}^\bullet$ , mais comprend 1 méthionine, 2 tryptophanes, 4 phénylalanines et 2 histidines également facilement oxydables. Des études par fluorescence (cf. article en annexe) montrent que les deux tryptophanes sont détruits au cours de l'irradiation. Le tryptophane 74, localisé dans la boucle non structurée de la protéine est beaucoup plus accessible au solvant que le

tryptophane 61 localisé à l'extrémité du 4<sup>ème</sup> brin  $\beta$  (Paquet et coll., en préparation ; Figure 10, p27). Par ailleurs, différentes études montrent que le tryptophane 74 est très probablement proche ou au niveau du site de fixation de la protéine. Ainsi, ce tryptophane peut être photochimiquement ponté à l'ADN (Katouzian-Safadi et coll., 1993) et des études préliminaires par RMN montrent que ses raies de résonances sont modifiées en présence d'ADN (communication personnelle F. Paquet). L'oxydation du tryptophane 74 par les rayons  $\gamma$  est donc une hypothèse possible pour expliquer l'inactivation de la protéine MC1. Cependant, la méthionine 75, très réactive, est dans le même environnement et peut aussi conduire à l'inactivation. Des études sont en cours pour déterminer les dommages produits pendant l'irradiation (P. Lopez-Pereira, (DEA) : étude de l'inactivation photochimique de MC1 ; C. Buré : analyse par spectrométrie de masse).

## **CONCLUSIONS ET PERSPECTIVES**

Dans la première partie de ce travail, la méthode SELEX a été utilisée pour isoler des séquences préférentiellement reconnues par la protéine MC1. Pour cela, nous avons utilisé une librairie contenant une partie aléatoire de 15 pb (soit  $10^9$  séquences différentes). Dans nos conditions de sélection, 10 cycles ont été suffisants pour obtenir un enrichissement maximal en séquences de forte affinité. Les 9 séquences obtenues sont très proches les unes des autres et une séquence consensus a pu être déterminée (AAAAACACAC(A/C)CCC(C/A)). Ce résultat est relativement surprenant car, lorsque la méthode SELEX a été utilisée avec d'autres protéines architecturales telles que l'octamère d'histone, les protéines HMf ou certaines protéines HMGB un grand nombre de séquences différentes a été isolés et aucune séquence consensus n'a été obtenue. (Lowary et Widom, 1998 ; Wildund et coll., 1997 ; Sandman et coll., 2001 ; Klass et coll., 2003). Avec les histones, des analyses par transformée de Fourier ont été nécessaires pour déterminer certaines caractéristiques des séquences isolées (alternance de séquences G/C et A/T en phase avec le pas de l'hélice d'ADN). Avec une protéine de type HMG-D un simple enrichissement en dinucléotides CG a été obtenu.

La protéine MC1 possède une affinité 40 à 50 fois plus forte pour les séquences issues du SELEX que pour des séquences aléatoires. Ce rapport, assez faible, est toutefois du même ordre que ceux obtenus avec d'autres protéines architecturales dont l'interaction avec l'ADN est considérée comme essentiellement non spécifique. Ainsi, si un rapport d'affinité d'environ 1000 a été obtenu avec l'octamère d'histones (Lowary et Widom, 1998), des facteurs plus faibles ont été observés avec les protéines HMf et les protéines de la famille de CspA (respectivement 9 à 18 et 2 ; Bailey et coll., 2000 ; Phadtare et Inouye, 1999).

Cette faible spécificité de reconnaissance a-t-elle une signification *in vivo* ? L'occurrence 6 à 7 fois supérieure à la statistique de ces séquences dans le génome des Méthanosarcines est-elle significative ? Plusieurs variants étant présents dans certains organismes, sommes-nous, comme dans le cas des protéines CspA, en présence d'une famille de protéines où chaque membre a des caractéristiques distinctes de reconnaissance de la séquence de l'ADN ? Seule la dernière question pourrait trouver une réponse par la poursuite de ce travail et la réalisation d'autres expériences SELEX. Il est toutefois important de rappeler que plus une protéine est abondante plus une faible reconnaissance peut avoir un importance *in vivo*. Ainsi la protéine FIS ou les protéines CspA peuvent elles

être effectivement fixées sur des sites définis alors que leur spécificité de reconnaissance est très faible.

Que reconnaît la protéine MC1 dans la séquence consensus ? L'ensemble de nos résultats, en particulier la faible spécificité, est en faveur d'une lecture indirecte de la séquence. Certaines comparaisons de stabilité des complexes vont également dans ce sens. Ainsi les séquences C12 et C3, qui ne varient que pour la position 14 (A→T), ont une différence de stabilité d'un facteur 2.5. Le petit sillon ne fournissant pas assez d'atomes donneurs ou accepteurs de liaisons hydrogène pour différencier les paires de bases AT et TA, c'est donc bien via la structure et la flexibilité de l'ADN que se fait la reconnaissance (Seeman et coll., 1976). Les études par simulation de dynamique moléculaire montrent que la séquence consensus a effectivement des caractéristiques très particulières. Il est vraisemblable que la protéine utilise à la fois la légère courbure présente sur la séquence et son profil de flexibilité pour courber l'ADN de manière importante. Les études concernant l'effet des rayonnements ionisants vont également dans ce sens en montrant le lien étroit entre la reconnaissance de la séquence et la courbure de l'ADN.

Comment la protéine interagit-elle avec l'ADN et par quels mécanismes le courbe-t-elle ? Seules des études structurales sur un complexe pourraient réellement répondre à cette question. Nous espérons que la séquence consensus que nous avons déterminée contribuera à la réalisation de tels complexes. Nos résultats donnent cependant des informations intéressantes. Ils montrent en particulier que la fixation de la protéine MC1 a lieu sur une face de l'ADN via le petit sillon et que les contacts sont localisés en deux régions séparées par 10 pb. En cela MC1 ressemble aux histones et l'ADN semble entourer la protéine. Les données de la littérature montrent qu'il existe différents mécanismes possibles pour courber l'ADN. Ainsi, Strauss et Maher (1994), ont montré comment la neutralisation des charges portées par les phosphates peut être à l'origine d'une courbure de l'ADN. Ce mécanisme est en particulier utilisé par les histones. Etant donnée la forte charge positive portée par la protéine MC1 (+12), cette hypothèse est vraisemblable. Par ailleurs, les résultats obtenus avec la protéine irradiée laissent supposer que le tryptophane 74 et/ou la méthionine 75 (localisés dans la boucle peu structurée de MC1) sont impliqués dans la reconnaissance et dans la courbure de l'ADN. Dans ce cas, on peut envisager que ces acides aminés s'intercalent dans le petit sillon, permettant ainsi la formation de liaisons hydrophobes et d'interaction de Van der Waals avec l'ADN, comme cela est observé avec différentes protéines HMGB ou avec les protéines Sul 7d. Dans les exemples cités ci-

dessus, les acides aminés intercalés agissent comme des coins qui vont provoquer une courbure de l'ADN dans la direction opposée à la protéine. D'après nos empreintes moléculaires ceci ne semble pas être le cas avec la protéine MC1 (courbure dirigée vers la protéine). Evidemment, les deux modes de courbure (neutralisation de charges et intercalation d'acides aminés) ne sont pas exclusifs et les protéines HU et IHF utilisent d'ailleurs les deux mécanismes pour courber fortement l'ADN.

Ni les données structurales ni les études physico-chimiques ne permettront de déterminer les processus biologiques dans lesquels est impliquée la protéine. L'intérêt croissant pour développer des outils génétiques chez les archaebactéries devrait par contre amener de précieuses informations sur ce point. Actuellement de telles méthodologies ont été mises au point pour différents organismes mais pas encore pour les méthanosarcines (Bitan-Banin et coll., 2003 ; Toffin et coll., 2002).

## **BIBLIOGRAPHIE**

- Afflerbach, H., Schroder, O. and Wagner, R. (1999) Conformational changes of the upstream DNA mediated by H-NS and FIS regulate E. coli RrnB P1 promoter activity. *J Mol Biol*, **286**, 339-353.
- Agback, P., Baumann, H., Knapp, S., Ladenstein, R. and Hard, T. (1998) Architecture of nonspecific protein-DNA interactions in the Sso7d-DNA complex. *Nat Struct Biol*, **5**, 579-584.
- Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S. and Ishihama, A. (1999) Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *J Bacteriol*, **181**, 6361-6370.
- Auner, H., Buckle, M., Deufel, A., Kutateladze, T., Lazarus, L., Mavathur, R., Muskhelishvili, G., Pemberton, I., Schneider, R. and Travers, A. (2003) Mechanism of transcriptional activation by FIS: role of core promoter structure and DNA topology. *J Mol Biol*, **331**, 331-344.
- Badaut, C., Williams, R., Arluison, V., Bouffartigues, E., Robert, B., Buc, H. and Rimsky, S. (2002) The degree of oligomerization of the H-NS nucleoid structuring protein is related to specific binding to DNA. *J Biol Chem*, **277**, 41657-41666.
- Bailey, K.A., Pereira, S.L., Widom, J. and Reeve, J.N. (2000) Archaeal histone selection of nucleosome positioning sequences and the procaryotic origin of histone-dependent genome evolution. *J Mol Biol*, **303**, 25-34.
- Ball, C.A. and Johnson, R.C. (1991) Multiple effects of Fis on integration and the control of lysogeny in phage lambda. *J Bacteriol*, **173**, 4032-4038.
- Ball, C.A. and Johnson, R.C. (1991) Efficient excision of phage lambda from the Escherichia coli chromosome requires the Fis protein. *J Bacteriol*, **173**, 4027-4031.
- Ball, L.J., Murzina, N.V., Broadhurst, R.W., Raine, A.R., Archer, S.J., Stott, F.J., Murzin, A.G., Singh, P.B., Dmaille, P.J. and Laue, E.D. (1997) Structure of the chromatin binding (chromo) domain from mouse modifier protein 1. *Embo J*, **16**, 2473-2481.
- Barns, S.M., Delwiche, C.F., Palmer, J.D. and Pace, N.R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA



- sequences. *Proc Natl Acad Sci U S A*, **93**, 9188-9193.
- Baumann, H., Knapp, S., Lundback, T., Ladenstein, R. and Hard, T. (1994) Solution structure and DNA-binding properties of a thermostable protein from the archaeon *Sulfolobus solfataricus*. *Nat Struct Biol*, **1**, 808-819.
- Bell, S.D., Botting, C.H., Wardleworth, B.N., Jackson, S.P. and White, M.F. (2002) The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. *Science*, **296**, 148-151.
- Bell, S.D. and Jackson, S.P. (2001) Mechanism and regulation of transcription in archaea. *Curr Opin Microbiol*, **4**, 208-213.
- Beloin, C., Jeusset, J., Revet, B., Mirambeau, G., Le Hegarat, F. and Le Cam, E. (2003) Contribution of DNA conformation and topology in right-handed DNA wrapping by the *Bacillus subtilis* LrpC protein. *J Biol Chem*, **278**, 5333-5342.
- Bendich, A.J. (2001) The form of chromosomal DNA molecules in bacterial cells. *Biochimie*, **83**, 177-186.
- Bitan-Banin, G., Ortenberg, R. and Mevarech, M. (2003) Development of a gene knockout system for the halophilic archaeon *Haloferax volcanii* by use of the *pyrE* gene. *J Bacteriol*, **185**, 772-778.
- Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N. and Weintraub, H. (1990) Sequence-specific DNA binding by the c-Myc protein. *Science*, **250**, 1149-1151.
- Bloch, V., Yang, Y., Margeat, E., Chavanieu, A., Auge, M.T., Robert, B., Arold, S., Rimsky, S. and Kochoyan, M. (2003) The H-NS dimerization domain defines a new fold contributing to DNA recognition. *Nat Struct Biol*, **10**, 212-218.
- Bokal, A.J.t., Ross, W. and Gourse, R.L. (1995) The transcriptional activator protein FIS: DNA interactions and cooperative interactions with RNA polymerase at the *Escherichia coli* *rrnB* P1 promoter. *J Mol Biol*, **245**, 197-207.
- Bonnefoy, E. and Rouviere-Yaniv, J. (1992) HU, the major histone-like protein of *E. coli*, modulates the binding of IHF to *oriC*. *Embo J*, **11**, 4489-4496.
- Bonnefoy, E., Takahashi, M. and Yaniv, J.R. (1994) DNA-binding parameters of the HU protein of *Escherichia coli* to cruciform DNA. *J Mol Biol*, **242**, 116-129.

- Boubrik, F. and Rouviere-Yaniv, J. (1995) Increased sensitivity to gamma irradiation in bacteria lacking protein HU. *Proc Natl Acad Sci U S A*, **92**, 3958-3962.
- Bramhill, D. and Kornberg, A. (1988) A model for initiation at origins of DNA replication. *Cell*, **54**, 915-918.
- Brandi, A., Spurio, R., Gualerzi, C.O. and Pon, C.L. (1999) Massive presence of the Escherichia coli 'major cold-shock protein' CspA under non-stress conditions. *Embo J*, **18**, 1653-1659.
- Bruist, M.F., Glasgow, A.C., Johnson, R.C. and Simon, M.I. (1987) Fis binding to the recombinational enhancer of the Hin DNA inversion system. *Genes Dev*, **1**, 762-772.
- Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S. and Venter, J.C. (1996) Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science*, **273**, 1058-1073.
- Cassiday, L.A. and Maher, L.J., 3rd. (2002) Having it both ways: transcription factors that bind DNA and RNA. *Nucleic Acids Res*, **30**, 4118-4126.
- Castaing, B., Zelwer, C., Laval, J. and Boiteux, S. (1995) HU protein of Escherichia coli binds specifically to DNA that contains single-strand breaks or gaps. *J Biol Chem*, **270**, 10291-10296.
- Ceschini, S., Lupidi, G., Coletta, M., Pon, C.L., Fioretti, E. and Angeletti, M. (2000) Multimeric self-assembly equilibria involving the histone-like protein H-NS. A thermodynamic study. *J Biol Chem*, **275**, 729-734.
- Charlier, M., Eon, S., Seche, E., Bouffard, S., Culard, F. and Spothem-Maurizot, M. (2002) Radiolysis of lac repressor by gamma-rays and heavy ions: a two-hit model for protein inactivation. *Biophys J*, **82**, 2373-2382.
- Chartier, F., Laine, B., Belaiche, D. and Sautiere, P. (1989 a) Primary structure of the chromosomal proteins MC1a, MC1b, and MC1c from the archaeobacterium Methanotheroxobacterium soehngenii. *J Biol Chem*, **264**, 17006-17015.

- Chartier, F., Laine, B., Belaiche, D., Touzel, J.P. and Sautiere, P. (1989 b) Primary structure of the chromosomal protein MC1 from the archaeobacterium *Methanosarcina* sp. CHTI 55. *Biochim Biophys Acta*, **1008**, 309-314.
- Chartier, F., Laine, B. and Sautiere, P. (1988) Characterization of the chromosomal protein MC1 from the thermophilic archaeobacterium *Methanosarcina* sp. CHTI 55 and its effect on the thermal stability of DNA. *Biochim Biophys Acta*, **951**, 149-156.
- Chartier, F., Laine, B., Sautiere, P., Touzel, J.P. and Albagnac, G. (1985) Characterisation of the chromosomal protein HMB isolated from *Methanosarcina barkeri*. *FEBS Lett*, **183**, 119-123.
- Chittenden, T., Livingston, D.M. and Kaelin, W.G., Jr. (1991) The T/E1A-binding domain of the retinoblastoma product can interact selectively with a sequence-specific DNA-binding protein. *Cell*, **65**, 1073-1082.
- Cognet, J.A., Pakleza, C., Cherny, D., Delain, E. and Cam, E.L. (1999) Static curvature and flexibility measurements of DNA with microscopy. A simple renormalization method, its assessment by experiment and simulation. *J Mol Biol*, **285**, 997-1009.
- Craig, N.L. and Nash, H.A. (1984) E. coli integration host factor binds to specific sites in DNA. *Cell*, **39**, 707-716.
- Craigie, R., Arndt-Jovin, D.J. and Mizuuchi, K. (1985) A defined system for the DNA strand-transfer reaction at the initiation of bacteriophage Mu transposition: protein and DNA substrate requirements. *Proc Natl Acad Sci U S A*, **82**, 7570-7574.
- Crothers, D.M. and Shakkled, Z. (1999) DNA bending by adenine-thymine tracts. *in Oxford Handbook of Nucleic Acid Structure*, ed. Neidle, S. (Oxford Univ Press, Oxford), 455-470.
- Cui, Y., Wang, Q., Stormo, G.D. and Calvo, J.M. (1995) A consensus sequence for binding of Lrp to DNA. *J Bacteriol*, **177**, 4872-4880.
- Culard, F., Laine, B., Sautiere, P. and Maurizot, J.C. (1993) Stoichiometry of the binding of chromosomal protein MC1 from the archaeobacterium, *Methanosarcina* spp. CHTI55, to DNA. *FEBS Lett*, **315**, 335-339.

- Cusick, M.E. and Belfort, M. (1998) Domain structure and RNA annealing activity of the Escherichia coli regulatory protein StpA. *Mol Microbiol*, **28**, 847-857.
- Dame, R.T. and Goosen, N. (2002) HU: promoting or counteracting DNA compaction? *FEBS Lett*, **529**, 151-156.
- Dame, R.T., Wyman, C. and Goosen, N. (2001) Structural basis for preferential binding of H-NS to curved DNA. *Biochimie*, **83**, 231-234.
- David-Cordonnier, M.H., Hamdane, M., Bailly, C. and D'Halluin, J.C. (1998) The DNA binding domain of the human c-Abl tyrosine kinase preferentially binds to DNA sequences containing an AAC motif and to distorted DNA structures. *Biochemistry*, **37**, 6065-6076.
- Davies, M.J. and Dean, R.T. (1997) *In Radical-mediated protein oxydation from chemistry to medicine*, Oxford University Press, New-York.
- Decanniere, K., Babu, A.M., Sandman, K., Reeve, J.N. and Heinemann, U. (2000) Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon Methanothermus fervidus. *J Mol Biol*, **303**, 35-47.
- DeLange, R.J., Williams, L.C. and Searcy, D.G. (1981) A histone-like protein (HTa) from Thermoplasma acidophilum. II. Complete amino acid sequence. *J Biol Chem*, **256**, 905-911.
- Dickerson, R.E. (1998) DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res*, **26**, 1906-1926.
- Dinger, M.E., Baillie, G.J. and Musgrave, D.R. (2000) Growth phase-dependent expression and degradation of histones in the thermophilic archaeon Thermococcus zilligii. *Mol Microbiol*, **36**, 876-885.
- Djuranovic, D. and Hartmann, B. (2003) Conformational characteristics and correlations in crystal structures of nucleic acid oligonucleotides: evidence for sub-states. *J Biomol Struct Dyn*, **20**, 771-788.
- Djuranovic, D. and Hartmann, B. (2004) DNA fine structure and dynamics in crystals and in solution: the impact of BI/BII backbone conformations. *Biopolymers*, **73**, 356-368.
- Dong, F. and van Holde, K.E. (1991) Nucleosome positioning is determined by the

- (H3-H4)<sub>2</sub> tetramer. *Proc Natl Acad Sci U S A*, **88**, 10596-10600.
- Dornberger, U., Spackovj, N., Walter, A., Gollmick, F.A., Sponer, J. and Fritzsche, H. (2001) Solution structure of the dodecamer d-(CATGGGCC-CATG)<sub>2</sub> is B-DNA. Experimental and molecular dynamics study. *J Biomol Struct Dyn*, **19**, 159-174.
- Drew, H.R. and Travers, A.A. (1985) DNA bending and its relation to nucleosome positioning. *J Mol Biol*, **186**, 773-790.
- Durrenberger, M., Bjornsti, M.A., Uetz, T., Hobot, J.A. and Kellenberger, E. (1988) Intracellular location of the histonelike protein HU in Escherichia coli. *J Bacteriol*, **170**, 4757-4768.
- Edmondson, S.P., Qiu, L. and Shriver, J.W. (1995) Solution structure of the DNA-binding protein Sac7d from the hyperthermophile Sulfolobus acidocaldarius. *Biochemistry*, **34**, 13289-13304.
- Edmondson, S.P. and Shriver, J.W. (2001) DNA binding proteins Sac 7d and Sso 7d from sulfolobus. *Methods Enzymol*, **334**, 129-145.
- el Hassan, M.A. and Calladine, C.R. (1996) Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J Mol Biol*, **259**, 95-103.
- Ellington, A.D. and Szostak, J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818-822.
- Eon, S., Culard, F., Sy, D., Charlier, M. and Spothem-Maurizot, M. (2001) Radiation disrupts protein-DNA complexes through damage to the protein. The lac repressor-operator system. *Radiat Res*, **156**, 110-117.
- Fahrner, R.L., Cascio, D., Lake, J.A. and Slesarev, A. (2001) An ancestral nuclear protein assembly: crystal structure of the Methanopyrus kandleri histone. *Protein Sci*, **10**, 2002-2007.
- Flashner, Y. and Gralla, J.D. (1988) DNA dynamic flexibility and protein recognition: differential stimulation by bacterial histone-like protein HU. *Cell*, **54**, 713-721.
- Fletcher, S.A. and Csonka, L.N. (1995) Fine-structure deletion analysis of the transcriptional silencer of the proU operon of Salmonella typhimurium. *J Bacteriol*, **177**, 4508-4513.

- Free, A. and Dorman, C.J. (1997) The *Escherichia coli* *stpA* gene is transiently expressed during growth in rich medium and is induced in minimal medium and by stress conditions. *J Bacteriol*, **179**, 909-918.
- Friedman, D.I. (1988) Integration host factor: a protein for all reasons. *Cell*, **55**, 545-554.
- Gaillard, C. and Strauss, F. (1994) Association of poly(CA).poly(TG) DNA fragments into four-stranded complexes bound by HMG1 and 2. *Science*, **264**, 433-436.
- Gao, Y.G., Su, S.Y., Robinson, H., Padmanabhan, S., Lim, L., McCrary, B.S., Edmondson, S.P., Shriver, J.W. and Wang, A.H. (1998) The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA. *Nat Struct Biol*, **5**, 782-786.
- Giraldo, R. (2003) Common domains in the initiators of DNA replication in Bacteria, Archaea and Eukarya: combined structural, functional and phylogenetic perspectives. *FEMS Microbiol Rev*, **26**, 533-554.
- Gold, L., Brown, D., He, Y., Shtatland, T., Singer, B.S. and Wu, Y. (1997) From oligonucleotide shapes to genomic SELEX: novel biological regulatory loops. *Proc Natl Acad Sci U S A*, **94**, 59-64.
- Goodman, S.D., Velten, N.J., Gao, Q., Robinson, S. and Segall, A.M. (1999) In vitro selection of integration host factor binding sites. *J Bacteriol*, **181**, 3246-3255.
- Guagliardi, A., Cerchia, L., Moracci, M. and Rossi, M. (2000) The chromosomal protein *sso7d* of the crenarchaeon *Sulfolobus solfataricus* rescues aggregated proteins in an ATP hydrolysis-dependent manner. *J Biol Chem*, **275**, 31813-31818.
- Guagliardi, A., Napoli, A., Rossi, M. and Ciaramella, M. (1997) Annealing of complementary DNA strands above the melting point of the duplex promoted by an archaeal protein. *J Mol Biol*, **267**, 841-848.
- Hales, L.M., Gumport, R.I. and Gardner, J.F. (1996) Examining the contribution of a dA+dT element to the conformation of *Escherichia coli* integration host factor-DNA complexes. *Nucleic Acids Res*, **24**, 1780-1786.
- Hamiche, A., Carot, V., Alilat, M., De Lucia, F., O'Donohue, M.F., Revet, B. and Prunell, A. (1996) Interaction of the histone (H3-H4)<sub>2</sub> tetramer of the

- nucleosome with positively supercoiled DNA minicircles: Potential flipping of the protein from a left- to a right-handed superhelical form. *Proc Natl Acad Sci U S A*, **93**, 7588-7593.
- Hao, D., Ohme-Takagi, M. and Yamasaki, K. (2003) A modified sensor chip for surface plasmon resonance enables a rapid determination of sequence specificity of DNA-binding proteins. *FEBS Lett*, **536**, 151-156.
- Haran, T.E., Kahn, J.D. and Crothers, D.M. (1994) Sequence elements responsible for DNA curvature. *J Mol Biol*, **244**, 135-143.
- Harley, V.R., Lovell-Badge, R. and Goodfellow, P.N. (1994) Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res*, **22**, 1500-1501.
- Hausner, W., Wettach, J., Hethke, C. and Thomm, M. (1996) Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase. *J Biol Chem*, **271**, 30144-30148.
- Higashibata, H., Fujiwara, S., Takagi, M. and Imanaka, T. (1999) Analysis of DNA compaction profile and intracellular contents of archaeal histones from *Pyrococcus kodakaraensis* KOD1. *Biochem Biophys Res Commun*, **258**, 416-424.
- Hinnebusch, B.J. and Bendich, A.J. (1997) The bacterial nucleoid visualized by fluorescence microscopy of cells lysed within agarose: comparison of *Escherichia coli* and spirochetes of the genus *Borrelia*. *J Bacteriol*, **179**, 2228-2237.
- Ho, P.S. (1994) The non-B-DNA structure of d(CA/TG)<sub>n</sub> does not differ from that of Z-DNA. *Proc Natl Acad Sci U S A*, **91**, 9549-9553.
- Hommais, F., Krin, E., Laurent-Winter, C., Soutourina, O., Malpertuy, A., Le Caer, J.P., Danchin, A. and Bertin, P. (2001) Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. *Mol Microbiol*, **40**, 20-36.
- Howard, M.T., Sandman, K., Reeve, J.N. and Griffith, J.D. (1992) HMf, a histone-related protein from the hyperthermophilic archaeon *Methanothermus fervidus*, binds preferentially to DNA containing phased tracts of adenines. *J*

- Bacteriol*, **174**, 7864-7867.
- Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C. and Stetter, K.O. (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature*, **417**, 63-67.
- Hwang, D.S. and Kornberg, A. (1992) Opening of the replication origin of *Escherichia coli* by DnaA protein with protein HU or IHF. *J Biol Chem*, **267**, 23083-23086.
- Imbert, M. and Laine, B. (1988) The deoxyribonucleoprotein complex from the archaeobacterium *Methanosarcina barkeri* : characterisation and ultrastructural localisation of the chromosomal protein MC1. *Can. J. Microbiol.*, **34**, 931-937.
- Irvine, D., Tuerk, C. and Gold, L. (1991) SELEXION. Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. *J Mol Biol*, **222**, 739-761.
- Isabelle, V., Franchet-Beuzit, J., Sabattier, R., Laine, B., Spothem-Maurizot, M. and Charlier, M. (1993) Radioprotection of DNA by a DNA-binding protein: MC1 chromosomal protein from the archaeobacterium *Methanosarcina* sp. CHT155. *Int J Radiat Biol*, **63**, 749-758.
- Johnson, R.C., Bruist, M.F. and Simon, M.I. (1986) Host protein requirements for in vitro site-specific DNA inversion. *Cell*, **46**, 531-539.
- Kamashev, D. and Rouviere-Yaniv, J. (2000) The histone-like protein HU binds specifically to DNA recombination and repair intermediates. *Embo J*, **19**, 6527-6535.
- Karlin, S., Brocchieri, L., Trent, J., Blaisdell, B.E. and Mrazek, J. (2002) Heterogeneity of genome and proteome content in bacteria, archaea, and eukaryotes. *Theor Popul Biol*, **61**, 367-390.
- Katouzian-Safadi, M., Blazy, B., Cremet, J.Y., Le Caer, J.P., Rossier, J. and Charlier, M. (1993) Photo-cross-linking of CRP to nonspecific DNA in the absence of cAMP. DNA interacts with both the N- and C-terminal parts of the protein. *Biochemistry*, **32**, 1770-1773.
- Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T.,



- Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K. and Kikuchi, H. (1998) Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. *DNA Res*, **5**, 55-76.
- Kellenberger, E. (1990) Intracellular organisation of the bacterial genome. *The Bacterial Chromosome*, 173-186.
- Kelman, L.M. and Kelman, Z. (2003) Archaea: an archetype for replication initiation studies? *Mol Microbiol*, **48**, 605-615.
- Kim, J.L., Nikolov, D.B. and Burley, S.K. (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, **365**, 520-527.
- Klass, J., Murphy, F.V.t., Fouts, S., Serenil, M., Changela, A., Siple, J. and Churchill, M.E. (2003) The role of intercalating residues in chromosomal high-mobility-group protein DNA binding, bending and specificity. *Nucleic Acids Res*, **31**, 2852-2864.
- Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyripides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B., Venter, J.C. and et al. (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature*, **390**, 364-370.
- Knight, R. and Yarus, M. (2003) Analyzing partially randomized nucleic acid pools: straight dope on doping. *Nucleic Acids Res*, **31**, e30.
- Koch, C. and Kahmann, R. (1986) Purification and properties of the *Escherichia coli* host factor required for inversion of the G segment in bacteriophage Mu. *J Biol Chem*, **261**, 15673-15678.
- Koo, H.S., Wu, H.M. and Crothers, D.M. (1986) DNA bending at adenine . thymine tracts. *Nature*, **320**, 501-506.
- Kostrewa, D., Granzin, J., Stock, D., Choe, H.W., Labahn, J. and Saenger, W. (1992) Crystal structure of the factor for inversion stimulation FIS at 2.0 Å resolution. *J Mol Biol*, **226**, 209-226.

- Kosturko, L.D., Daub, E. and Murialdo, H. (1989) The interaction of E. coli integration host factor and lambda cos DNA: multiple complex formation and protein-induced bending. *Nucleic Acids Res*, **17**, 317-334.
- Krylov, A.S., Zasedateleva, O.A., Prokopenko, D.V., Rouviere-Yaniv, J. and Mirzabekov, A.D. (2001) Massive parallel analysis of the binding specificity of histone-like protein HU to single- and double-stranded DNA with generic oligodeoxyribonucleotide microchips. *Nucleic Acids Res*, **29**, 2654-2660.
- Kvaratskhelia, M., Wardleworth, B.N., Bond, C.S., Fogg, J.M., Lilley, D.M. and White, M.F. (2002) Holliday junction resolution is modulated by archaeal chromatin components in vitro. *J Biol Chem*, **277**, 2992-2996.
- Kyrpides, N.C. and Ouzounis, C.A. (1999) Transcription in archaea. *Proc Natl Acad Sci U S A*, **96**, 8545-8550.
- Laine, B., Chartier, F., Imbert, M., Lewis, R. and Sautiere, P. (1986) Primary structure of the chromosomal protein HMB from the archaeobacteria *Methanosarcina barkeri*. *Eur J Biochem*, **161**, 681-687.
- Langer, D., Hain, J., Thuriaux, P. and Zillig, W. (1995) Transcription in archaea: similarity to that in eucarya. *Proc Natl Acad Sci U S A*, **92**, 5768-5772.
- Larquet, E., Le Cam, E., Fourcade, A., Culard, F., Furrer, P. and Delain, E. (1996) [Complementarity of microscopies in the structural analysis of DNA minicircles associated to protein MC1]. *C R Acad Sci III*, **319**, 461-471.
- Lavery, R. and Sklenar, H. (1989) Defining the structure of irregular nucleic acids: conventions and principles. *J Biomol Struct Dyn*, **6**, 655-667.
- Le Cam, E., Culard, F., Larquet, E., Delain, E. and Cognet, J.A. (1999) DNA bending induced by the archaeobacterial histone-like protein MC1. *J Mol Biol*, **285**, 1011-1021.
- Li, S. and Waters, R. (1998) *Escherichia coli* strains lacking protein HU are UV sensitive due to a role for HU in homologous recombination. *J Bacteriol*, **180**, 3750-3756.
- Li, W.T., Grayling, R.A., Sandman, K., Edmondson, S., Shriver, J.W. and Reeve, J.N. (1998) Thermodynamic stability of archaeal histones. *Biochemistry*, **37**, 10563-10572.

- Lindqvist, M. and Graslund, A. (2001) An FTIR and CD study of the structural effects of G-tract length and sequence context on DNA conformation in solution. *J Mol Biol*, **314**, 423-432.
- Lopez-Garcia, P. and Forterre, P. (1997) DNA topology in hyperthermophilic archaea: reference states and their variation with growth phase, growth temperature, and temperature stresses. *Mol Microbiol*, **23**, 1267-1279.
- Lopez-Garcia, P., Knapp, S., Ladenstein, R. and Forterre, P. (1998) In vitro DNA binding of the archaeal protein Sso7d induces negative supercoiling at temperatures typical for thermophilic growth. *Nucleic Acids Res*, **26**, 2322-2328.
- Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R. and Wright, P.E. (1995) Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature*, **376**, 791-795.
- Lowary, P.T. and Widom, J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol*, **276**, 19-42.
- Lucas, S., Toffin, L., Zivanovic, Y., Charlier, D., Moussard, H., Forterre, P., Prieur, D. and Erauso, G. (2002) Construction of a shuttle vector for, and spheroplast transformation of, the hyperthermophilic archaeon *Pyrococcus abyssi*. *Appl Environ Microbiol*, **68**, 5528-5536.
- Lurz, R., Grote, M., Dijk, J., Reinhardt, R. and Dobrinski, B. (1986) Electron microscopic study of DNA complexes with protein from archaeobacterium *Sulfolobus acidocaldarius*. *Embo J*, **5**, 3715-3721.
- Lyubchenko, Y.L., Shlyakhtenko, L.S., Appella, E. and Harrington, R.E. (1993) CA runs increase DNA flexibility in the complex of lambda Cro protein with the OR3 site. *Biochemistry*, **32**, 4121-4127.
- MacNeill, S.A. (2001) Understanding the enzymology of archaeal DNA replication: progress in form and function. *Mol Microbiol*, **40**, 520-529.
- Maleknia, S.D., Brenowitz, M. and Chance, M.R. (1999) Millisecond radiolytic modification of peptides by synchrotron X-rays identified by mass spectrometry. *Anal Chem*, **71**, 3965-3973.

- Maleknia, S.D., Ralston, C.Y., Brenowitz, M.D., Downard, K.M. and Chance, M.R. (2001) Determination of macromolecular folding and structure by synchrotron x-ray radiolysis techniques. *Anal Biochem*, **289**, 103-115.
- Marc, F., Sandman, K., Lurz, R. and Reeve, J.N. (2002) Archaeal histone tetramerization determines DNA affinity and the direction of DNA supercoiling. *J Biol Chem*, **277**, 30879-30886.
- McAfee, J.G., Edmondson, S.P., Datta, P.K., Shriver, J.W. and Gupta, R. (1995) Gene cloning, expression, and characterization of the Sac7 proteins from the hyperthermophile *Sulfolobus acidocaldarius*. *Biochemistry*, **34**, 10063-10077.
- McCall, M., Brown, T. and Kennard, O. (1985) The crystal structure of d(G-G-G-G-C-C-C-C). A model for poly(dG).poly(dC). *J Mol Biol*, **183**, 385-396.
- Mengeritsky, G., Goldenberg, D., Mendelson, I., Giladi, H. and Oppenheim, A.B. (1993) Genetic and biochemical analysis of the integration host factor of *Escherichia coli*. *J Mol Biol*, **231**, 646-657.
- Mohibullah, N., Donner, A., Ippolito, J.A. and Williams, T. (1999) SELEX and missing phosphate contact analyses reveal flexibility within the AP-2[alpha] protein: DNA binding complex. *Nucleic Acids Res*, **27**, 2760-2769.
- Morales, P., Rouviere-Yaniv, J. and Dreyfus, M. (2002) The histone-like protein HU does not obstruct movement of T7 RNA polymerase in *Escherichia coli* cells but stimulates its activity. *J Bacteriol*, **184**, 1565-1570.
- Morisato, D. and Kleckner, N. (1987) Tn10 transposition and circle formation in vitro. *Cell*, **51**, 101-111.
- Murphy, K.E. and Stringer, J.R. (1986) RecA independent recombination of poly[d(GT)-d(CA)] in pBR322. *Nucleic Acids Res*, **14**, 7325-7340.
- Myllykallio, H., Lopez, P., Lopez-Garcia, P., Heilig, R., Saurin, W., Zivanovic, Y., Philippe, H. and Forterre, P. (2000) Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science*, **288**, 2212-2215.
- Ng, W.V., Kennedy, S.P., Mahairas, G.G., Berquist, B., Pan, M., Shukla, H.D., Lasky, S.R., Baliga, N.S., Thorsson, V., Sbrogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Welti, R., Goo, Y.A., Leithauser, B., Keller, K., Cruz, R., Danson,

- M.J., Hough, D.W., Maddocks, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Dale, H., Isenbarger, T.A., Peck, R.F., Pohlschroder, M., Spudich, J.L., Jung, K.W., Alam, M., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P., Omer, A.D., Ebhardt, H., Lowe, T.M., Liang, P., Riley, M., Hood, L. and DasSarma, S. (2000) Genome sequence of Halobacterium species NRC-1. *Proc Natl Acad Sci U S A*, **97**, 12176-12181.
- Nye, M.B., Pfau, J.D., Skorupski, K. and Taylor, R.K. (2000) Vibrio cholerae H-NS silences virulence gene expression at multiple steps in the ToxR regulatory cascade. *J Bacteriol*, **182**, 4295-4303.
- Oberto, J., Drlica, K. and Rouviere-Yaniv, J. (1994) Histones, HMG, HU, IHF: Meme combat. *Biochimie*, **76**, 901-908.
- Pan, C.Q., Finkel, S.E., Cramton, S.E., Feng, J.A., Sigman, D.S. and Johnson, R.C. (1996) Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. *J Mol Biol*, **264**, 675-695.
- Paradinas, C., Gervais, A., Maurizot, J.C. and Culard, F. (1998) Structure-specific binding recognition of a methanogen chromosomal protein. *Eur J Biochem*, **257**, 372-379.
- Pavlov, N.A., Cherny, D.I., Jovin, T.M. and Slesarev, A.I. (2002) Nucleosome-like complex of the histone from the hyperthermophile Methanopyrus kandleri (MkaH) with linear DNA. *J Biomol Struct Dyn*, **20**, 207-214.
- Pereira, S.L., Grayling, R.A., Lurz, R. and Reeve, J.N. (1997) Archaeal nucleosomes. *Proc Natl Acad Sci U S A*, **94**, 12633-12637.
- Pereira, S.L. and Reeve, J.N. (1999) Archaeal nucleosome positioning sequence from Methanothermus fervidus. *J Mol Biol*, **289**, 675-681.
- Phadtare, S. and Inouye, M. (1999) Sequence-selective interactions with RNA by CspB, CspC and CspE, members of the CspA family of Escherichia coli. *Mol Microbiol*, **33**, 1004-1014.
- Pollock, R. and Treisman, R. (1990) A sensitive method for the determination of protein-DNA binding specificities. *Nucleic Acids Res*, **18**, 6197-6204.
- Poplawski, A. and Bernander, R. (1997) Nucleoid structure and distribution in thermophilic Archaea. *J Bacteriol*, **179**, 7625-7630.

- Preobrajenskaya, O., Boullard, A., Boubrik, F., Schnarr, M. and Rouviere-Yaniv, J. (1994) The protein HU can displace the LexA repressor from its DNA-binding sites. *Mol Microbiol*, **13**, 459-467.
- Prosseda, G., Fradiani, P.A., Di Lorenzo, M., Falconi, M., Micheli, G., Casalino, M., Nicoletti, M. and Colonna, B. (1998) A role for H-NS in the regulation of the virF gene of Shigella and enteroinvasive Escherichia coli. *Res Microbiol*, **149**, 15-25.
- Ramstein, J., Hervouet, N., Coste, F., Zelwer, C., Oberto, J. and Castaing, B. (2003) Evidence of a thermal unfolding dimeric intermediate for the Escherichia coli histone-like HU proteins: thermodynamics and structure. *J Mol Biol*, **331**, 101-121.
- Rashidzadeh, H., Khrapunov, S., Chance, M.R. and Brenowitz, M. (2003) Solution structure and interdomain interactions of the Saccharomyces cerevisiae "TATA binding protein" (TBP) probed by radiolytic protein footprinting. *Biochemistry*, **42**, 3655-3665.
- Reeve, J.N. (2003) Archaeal chromatin and transcription. *Mol Microbiol*, **48**, 587-598.
- Rice, P.A., Yang, S., Mizuuchi, K. and Nash, H.A. (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell*, **87**, 1295-1306.
- Rimsky, S., Zuber, F., Buckle, M. and Buc, H. (2001) A molecular mechanism for the repression of transcription by the H-NS protein. *Mol Microbiol*, **42**, 1311-1323.
- Robertson, C.A. and Nash, H.A. (1988) Bending of the bacteriophage lambda attachment site by Escherichia coli integration host factor. *J Biol Chem*, **263**, 3554-3557.
- Robinow, C. and Kellenberger, E. (1994) The bacterial nucleoid revisited. *Microbiol Rev*, **58**, 211-232.
- Robinson, H., Gao, Y.G., McCrary, B.S., Edmondson, S.P., Shriver, J.W. and Wang, A.H. (1998) The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature*, **392**, 202-205.
- Rouviere-Yaniv, J. and Gros, F. (1975) Characterization of a novel, low-molecular-weight DNA-binding protein from Escherichia coli. *Proc Natl Acad Sci U S A*, **72**, 3428-3432.

- Rouviere-Yaniv, J., Yaniv, M. and Germond, J.E. (1979) E. coli DNA binding protein HU forms nucleosomelike structure with circular double-stranded DNA. *Cell*, **17**, 265-274.
- Sandman, K., Bailey, K.A., Pereira, S.L., Soares, D., Li, W.T. and Reeve, J.N. (2001) Archaeal histones and nucleosomes. *Methods Enzymol*, **334**, 116-129.
- Sandman, K., Grayling, R.A., Dobrinski, B., Lurz, R. and Reeve, J.N. (1994) Growth-phase-dependent synthesis of histones in the archaeon *Methanothermus fervidus*. *Proc Natl Acad Sci U S A*, **91**, 12624-12628.
- Sandman, K., Krzycki, J.A., Dobrinski, B., Lurz, R. and Reeve, J.N. (1990) HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. *Proc Natl Acad Sci U S A*, **87**, 5788-5791.
- Sandman, K., Pereira, S.L. and Reeve, J.N. (1998) Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. *Cell Mol Life Sci*, **54**, 1350-1364.
- Sandman, K. and Reeve, J.N. (1999) Archaeal nucleosome positioning by CTG repeats. *J Bacteriol*, **181**, 1035-1038.
- Schnabel, R., Thomm, M., Schahn-Gerardy, R., Zillig, W., Stetter, K., O. and Huet, J. (1983) Structural homology between different archaeobacterial DNA-dependent RNA polymerases analysed by immunological comparison of their components. *Embo J*, **2**, 751-755.
- Scipioni, A., Anselmi, C., Zuccheri, G., Samori, B. and De Santis, P. (2002) Sequence-dependent DNA curvature and flexibility from scanning force microscopy images. *Biophys J*, **83**, 2408-2418.
- Seeman, N.C., Rosenberg, J.M. and Rich, A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc Natl Acad Sci U S A*, **73**, 804-808.
- Shimajiri, S., Arima, N., Tanimoto, A., Murata, Y., Hamada, T., Wang, K.Y. and Sasaguri, Y. (1999) Shortened microsatellite d(CA)<sub>21</sub> sequence down-regulates promoter activity of matrix metalloproteinase 9 gene. *FEBS Lett*, **455**, 70-74.

- Shindo, H., Iwaki, T., Ieda, R., Kurumizaka, H., Ueguchi, C., Mizuno, T., Morikawa, S., Nakamura, H. and Kuboniwa, H. (1995) Solution structure of the DNA binding domain of a nucleoid-associated protein, H-NS, from *Escherichia coli*. *FEBS Lett*, **360**, 125-131.
- Singer, B.S., Shtatland, T., Brown, D. and Gold, L. (1997) Libraries for genomic SELEX. *Nucleic Acids Res*, **25**, 781-786.
- Sitlani, A. and Crothers, D.M. (1998) DNA-binding domains of Fos and Jun do not induce DNA curvature: an investigation with solution and gel methods. *Proc Natl Acad Sci U S A*, **95**, 1404-1409.
- Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Reeve, J.N. and et al. (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J Bacteriol*, **179**, 7135-7155.
- Soares, D.J., Sandman, K. and Reeve, J.N. (2000) Mutational analysis of archaeal histone-DNA interactions. *J Mol Biol*, **297**, 39-47.
- Sodèn, B. and Uhin, B.E. (1996) Coordinated and differential expression of the histone-like protein in *Escherichia coli* : regulation and function of the H-NS analogue StpA. *Embo J*, **15**, 4970-4980.
- Spurio, R., Falconi, M., Brandi, A., Pon, C.L. and Gualerzi, C.O. (1997) The oligomeric structure of nucleoid protein H-NS is necessary for recognition of intrinsically curved DNA and for DNA bending. *Embo J*, **16**, 1795-1805.
- Starich, M.R., Sandman, K., Reeve, J.N. and Summers, M.F. (1996) NMR structure of HMfB from the hyperthermophile, *Methanothermus fervidus*, confirms that this archaeal protein is a histone. *J Mol Biol*, **255**, 187-203.
- Stetter, K., O., Fiala, G., Huber, R. and Seegerer, A. (1990) Hyperthermophilic microorganisms. *FEMS Microbiology Letters*, **75**, 117-124.
- Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications. *Nature*, **403**, 41-45.



- Strauss, J.K. and Maher, L.J., 3rd. (1994) DNA bending by asymmetric phosphate neutralization. *Science*, **266**, 1829-1834.
- Stringer, J.R. (1985) Recombination between poly[d(GT).d(CA)] sequences in simian virus 40-infected cultured cells. *Mol Cell Biol*, **5**, 1247-1259.
- Suck, D., Lahm, A. and Oefner, C. (1988) Structure refined to 2Å of a nicked DNA octanucleotide complex with DNase I. *Nature*, **332**, 464-468.
- Sun, D., Hurley, L.H. and Harshey, R.M. (1996) Structural distortions induced by integration host factor (IHF) at the H' site of phage lambda probed by (+)-CC-1065, pluramycin, and KMnO<sub>4</sub> and by DNA cyclization studies. *Biochemistry*, **35**, 10815-10827.
- Swinger, K.K., Lemberg, K.M., Zhang, Y. and Rice, P.A. (2003) Flexible DNA bending in HU-DNA cocrystal structures. *Embo J*, **22**, 3749-3760.
- Tabassum, R., Sandman, K.M. and Reeve, J.N. (1992) HMT, a histone-related protein from *Methanobacterium thermoautotrophicum* delta H. *J Bacteriol*, **174**, 7890-7895.
- Tae, H.J., Luo, X. and Kim, K.H. (1994) Roles of CCAAT/enhancer-binding protein and its binding site on repression and derepression of acetyl-CoA carboxylase gene. *J Biol Chem*, **269**, 10475-10484.
- Takayanagi, S., Morimura, S., Kusaoke, H., Yokoyama, Y., Kano, K. and Shioda, M. (1992) Chromosomal structure of the halophilic archaeobacterium *Halobacterium salinarium*. *J Bacteriol*, **174**, 7207-7216.
- Tanaka, I., Appelt, K., Dijk, J., White, S.W. and Wilson, K.S. (1984) 3-Å resolution structure of a protein with histone-like properties in prokaryotes. *Nature*, **310**, 376-381.
- Tanaka, K., Muramatsu, S., Yamada, H. and Mizuno, T. (1991) Systematic characterization of curved DNA segments randomly cloned from *Escherichia coli* and their functional significance. *Mol Gen Genet*, **226**, 367-376.
- Teyssier, C., Laine, B., Gervais, A., Maurizot, J.C. and Culard, F. (1994) Archaeobacterial histone-like protein MC1 can exhibit a sequence-specific binding to DNA. *Biochem J*, **303 ( Pt 2)**, 567-573.

- Teyssier, C., Toulme, F., Touzel, J.P., Gervais, A., Maurizot, J.C. and Culard, F. (1996) Preferential binding of the archaeobacterial histone-like MC1 protein to negatively supercoiled DNA minicircles. *Biochemistry*, **35**, 7954-7958.
- Thiesen, H.J. and Bach, C. (1990) Target Detection Assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. *Nucleic Acids Res*, Vol. 18, pp. 3203-3209.
- Thompson, J.F., Snyder, U.K. and Landy, A. (1988) Helical-repeat dependence of integrative recombination of bacteriophage lambda: role of the P1 and H1 protein binding sites. *Proc Natl Acad Sci U S A*, **85**, 6323-6327.
- Tippner, D., Afflerbach, H., Bradaczek, C. and Wagner, R. (1994) Evidence for a regulatory function of the histone-like Escherichia coli protein H-NS in ribosomal RNA synthesis. *Mol Microbiol*, **11**, 589-604.
- Toulme, F., Le Cam, E., Teyssier, C., Delain, E., Sautiere, P., Maurizot, J.C. and Culard, F. (1995) Conformational changes of DNA minicircles upon the binding of the archaeobacterial histone-like protein MC1. *J Biol Chem*, **270**, 6286-6291.
- Touzel, J.P., Petroff, D. and Albagnac, G. (1985) Isolation and characterisation of a new thermophilic Methanosarcina, the strain CHTI 55. *System. Appl. Microbiol.*, **6**, 66-71.
- Trifonov, E.N. and Sussman, J.L. (1980) The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proc Natl Acad Sci U S A*, **77**, 3816-3820.
- Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505-510.
- Tupper, A.E., Owen-Hughes, T.A., Ussery, D.W., Santos, D.S., Ferguson, D.J., Sidebotham, J.M., Hinton, J.C. and Higgins, C.F. (1994) The chromatin-associated protein H-NS alters DNA topology in vitro. *Embo J*, **13**, 258-268.
- Ueguchi, C. and Mizuno, T. (1996) Purification of H-NS protein and its regulatory effect on transcription in vitro. *Methods Enzymol*, **274**, 271-276.
- Ueguchi, C., Seto, C., Suzuki, T. and Mizuno, T. (1997) Clarification of the dimerization domain and its functional significance for the Escherichia coli

- nucleoid protein H-NS. *J Mol Biol*, **274**, 145-151.
- Vant-Hull, B., Payano-Baez, A., Davis, R.H. and Gold, L. (1998) The mathematics of SELEX against complex targets. *J Mol Biol*, **278**, 579-597.
- Varshavsky, A. (1996) The N-end rule: functions, mysteries, uses. *Proc Natl Acad Sci U S A*, **93**, 12142-12149.
- Wada, M., Kutsukake, K., Komano, T., Imamoto, F. and Kano, Y. (1989) Participation of the hup gene product in site-specific DNA inversion in *Escherichia coli*. *Gene*, **76**, 345-352.
- Wang, S., Cosstick, R., Gardner, J.F. and Gumpert, R.I. (1995) The specific binding of *Escherichia coli* integration host factor involves both major and minor grooves of DNA. *Biochemistry*, **34**, 13082-13090.
- Wardleworth, B.N., Russell, R.J., Bell, S.D., Taylor, G.L. and White, M.F. (2002) Structure of Alba: an archaeal chromatin protein modulated by acetylation. *Embo J*, **21**, 4654-4662.
- Wellenzohn, B., Flader, W., Winger, R.H., Hallbrucker, A., Mayer, E. and Liedl, K.R. (2002) Indirect readout of the trp-repressor-operator complex by B-DNA's backbone conformation transitions. *Biochemistry*, **41**, 4088-4095.
- Werner, M.H., Gronenborn, A.M. and Clore, G.M. (1996) Intercalation, DNA kinking, and the control of transcription. *Science*, **271**, 778-784.
- Werner, M.H., Huth, J.R., Gronenborn, A.M. and Clore, G.M. (1995) Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell*, **81**, 705-714.
- Widlund, H.R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P.E., Kahn, J.D., Crothers, D.M. and Kubista, M. (1997) Identification and characterization of genomic nucleosome-positioning sequences. *J Mol Biol*, **267**, 807-817.
- Williams, R.M., Rimsky, S. and Buc, H. (1996) Probing the structure, function, and interactions of the *Escherichia coli* H-NS and StpA proteins by using dominant negative derivatives. *J Bacteriol*, **178**, 4335-4343.
- Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) Towards a natural system of

- organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A*, **87**, 4576-4579.
- Wright, W.E., Binder, M. and Funk, W. (1991) Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. *Mol Cell Biol*, **11**, 4104-4110.
- Wu, H.M. and Crothers, D.M. (1984) The locus of sequence-directed and protein-induced DNA bending. *Nature*, **308**, 509-513.
- Xue, H., Guo, R., Wen, Y., Liu, D. and Huang, L. (2000) An abundant DNA binding protein from the hyperthermophilic archaeon *Sulfolobus shibatae* affects DNA supercoiling in a temperature-dependent fashion. *J Bacteriol*, **182**, 3929-3933.
- Yang, C.C. and Nash, H.A. (1989) The interaction of *E. coli* IHF protein with its specific binding sites. *Cell*, **57**, 869-880.
- Yang, S.W. and Nash, H.A. (1994) Specific photocrosslinking of DNA-protein complexes: identification of contacts between integration host factor and its target DNA. *Proc Natl Acad Sci U S A*, **91**, 12183-12187.
- Zhao, K., Chai, X. and Marmorstein, R. (2003) Structure of a Sir2 substrate, Alba, reveals a mechanism for deacetylation-induced enhancement of DNA binding. *J Biol Chem*, **278**, 26071-26077.

## **ANNEXE**

## Response of a DNA-binding Protein to Radiation-induced Oxidative Stress

This paper is a homage to the memory of Professor Claude Hélène, who was our mentor, both in science and in human relationships

**Françoise Culard, Alain Gervais, Guillaume de Vuyst  
Mélanie Spothem-Maurizot and Michel Charlier\***

*Centre de biophysique  
moléculaire, CNRS, rue  
Charles-Sadron, F-45071  
Orléans Cedex 2, France*

The DNA-binding protein MC1 is a chromosomal protein extracted from the archaebacterium *Methanosarcina* sp. CHTI55. It binds any DNA, and exhibits an enhanced affinity for some short sequences and structures (circles, cruciform DNA). Moreover, the protein bends DNA strongly at the binding site. MC1 was submitted to oxidative stress through  $\gamma$ -ray irradiation. In our experimental conditions, damage is essentially due to hydroxyl radicals issued from water radiolysis.

Upon irradiation, the regular complex between MC1 and DNA disappears, while a new complex appears. In the new complex, the protein loses its ability to recognise preferential sequences and DNA circles, and bends DNA less strongly than in the regular one. The new complex disappears and the protein becomes totally inactivated by high doses.

A model has been proposed to explain these experimental results. Two targets,  $R_1$  and  $R_2$ , are concomitantly destroyed in the protein, with different kinetics.  $R_2$  oxidation has no effect on the regular binding, whereas  $R_1$  oxidation modifies the functioning of MC1: loss of preferential site and structure recognition, weaker bending. The destruction of both  $R_1$  and  $R_2$  targets leads to a total inactivation of the protein. This model accounts for the data obtained by titrations of DNA with irradiated proteins.

When the protein is irradiated in the complex with DNA, bound DNA protects its binding site on the protein very efficiently.

The highly oxidisable tryptophan and methionine could be the amino acid residues implicated in the inactivation process.

© 2003 Elsevier Science Ltd. All rights reserved

*Keywords:* oxidative stress; radiolysis; chromosomal protein; protein–nucleic acid interaction; protein oxidation

\*Corresponding author

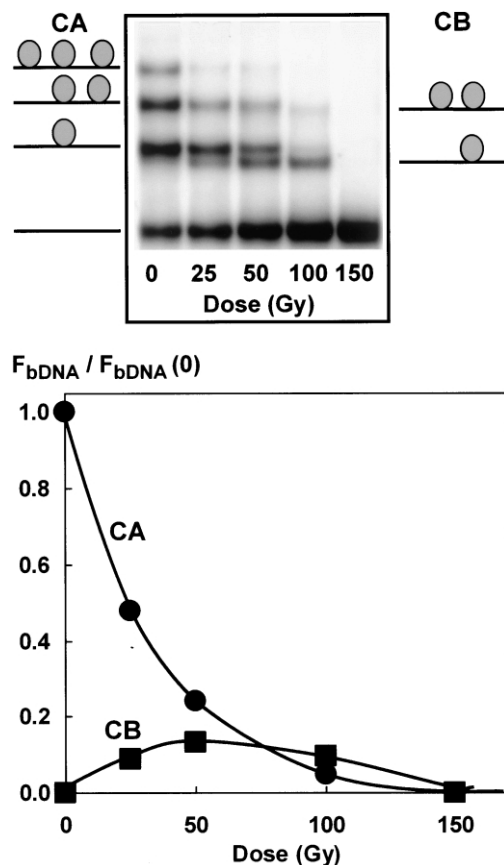
### Introduction

Proteins are involved in the regulation of gene expression, DNA structuring and packaging, cellular signalling, enzymatic processes, etc. Oxidation-induced damage to the proteins may alter these functions, and thus, lead to deleterious effects. Environmental, as well as physiological, factors, such as ageing, diseases and oxidative stress, participate in this general process.<sup>1,2</sup> The chemical agents responsible for these disorders are commonly called reactive oxygen species.<sup>3</sup>

Among these reactive oxygen species, the hydroxyl radical  $\text{OH}^\cdot$ , the superoxide radical ion  $\text{O}_2^-$  and hydrogen peroxide  $\text{H}_2\text{O}_2$  can be generated by water radiolysis.<sup>3</sup> Therefore, radiation chemistry of dilute aerated aqueous solutions of protein appears as a good tool to study the effects of oxidative stress. Several damages to the peptide chain or to the side-chains of amino acid residues have been identified, and some of their mechanisms of formation described.<sup>4–6</sup> The most sensitive amino acids, cysteine, tryptophan, tyrosine and methionine, bear aromatic or sulphur-containing side-chains.<sup>7–9</sup>

In our studies of the effects of oxidative stress on proteins, we are especially interested in DNA-binding proteins. In a previous work, we showed

E-mail addresses of the corresponding authors:  
[micharli@cnrs-orleans.fr](mailto:micharli@cnrs-orleans.fr); [culard@cnrs-orleans.fr](mailto:culard@cnrs-orleans.fr)



**Figure 1.** Top, Retardation gel electrophoresis of a mixture of 64 bp specific DNA fragments (100 nM) with irradiated protein (100 nM). The protein was irradiated at a concentration of 2  $\mu$ M. Bottom, Fraction of DNA in CA and CB complexes (data from upper part).  $F_{bDNA}$  is the fraction of DNA bound to irradiated protein in either CA or CB complexes.  $F_{bDNA}(0)$  is the fraction of DNA bound to non-irradiated protein (dose 0) in CA complexes.

that upon irradiation, a DNA-binding protein, the regulation protein *lac* repressor from *Escherichia coli*, lost its ability to bind DNA. Moreover, we observed that bound DNA protected the protein against oxidative attack.<sup>10</sup> Although it was necessary to damage at least two targets to inactivate a binding site,<sup>11</sup> the inactivation appeared as a one-step process, i.e. only two types of proteins (regarding their binding function) were present: active and inactive. Here, using another DNA-binding protein, a more sophisticated two-step inactivation process is described.

The protein studied, the MC1 protein, is a chromosomal protein extracted from the archaebacterium *Methanosarcina* sp. CHTI55. It is present naturally in cells at the ratio of one protein molecule for 100 bp, and its physiological role is probably to participate in the structuring of the archaean chromosome.<sup>12</sup> This DNA-binding protein contains 93 amino acid residues, with a positive global charge at neutral pH (24 basic residues,

Arg, Lys and His, and 11 acidic residues, Glu and Asp).

MC1 binds any DNA, but shows a preferential binding for specific sequences.<sup>13</sup> Moreover, and like HMGB proteins,<sup>14,15</sup> it exhibits an enhanced affinity for circular DNA<sup>16</sup> and for cruciform structures.<sup>17</sup> Binding to DNA of a specific sequence generates a strong curvature of the DNA, with a bending angle of 110°.<sup>18</sup> This curvature induces a significant reduction of the migration velocity of the complexes, as observed by gel-retardation electrophoresis.<sup>13</sup> In this work, the binding ability of the protein and the induced curvature of DNA were used as markers for the protein activity.

## Results

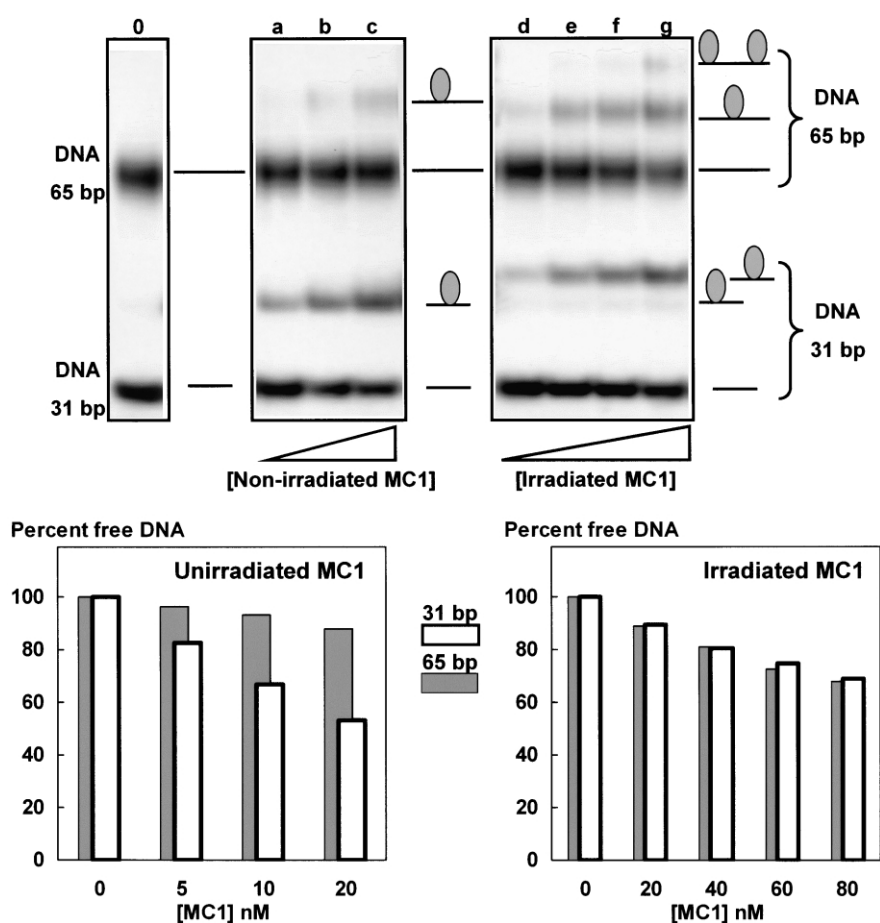
Several DNA fragments were used to test the DNA-binding ability of the irradiated protein. In some of these DNA fragments (64 bp and 31 bp), a sequence of 15 bp exhibits an enhanced affinity for the protein, as evidenced by footprinting in the case of the 64 bp fragment,<sup>13</sup> or selected using the SELEX method in the case of the 31 bp fragment (see Materials and Methods). These DNA fragments are called specific fragments, or specific DNA. The other fragments (65, 50, 86 and circular or linear 207 bp), called here non-specific fragments, or non-specific DNA, do not contain any particular sequence whose affinity for the protein would be greater than that of an average DNA.

The length of the binding site of the protein is between 11 (excluded site for non-specific binding) and 15 bp.<sup>13,19</sup> With fragments longer than several site lengths, and for high concentrations of protein, complexes of one DNA fragment with more than one molecule of MC1 can be formed. These complexes will be called 1:1, 1:2, or 1:3 complexes, the second digit representing the number of bound protein molecules per DNA fragment. In the case of a specific DNA fragment, one protein is bound to the specific site, the others being distributed randomly along the rest of the fragment. The 26 bp fragment is too short to bind more than one MC1 protein molecule.

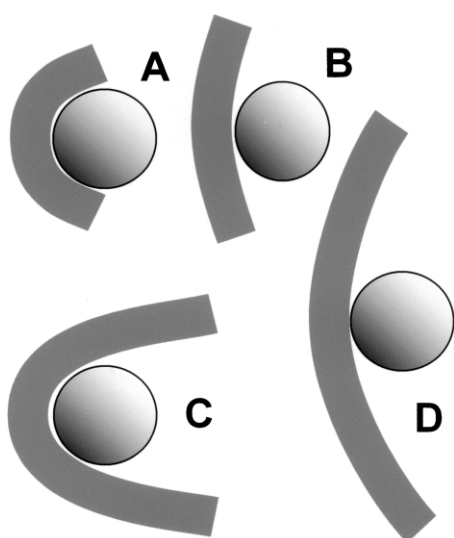
### Irradiation alters DNA-binding properties. A new complex is formed

MC1 protein is mixed with a 64 bp specific DNA. Figure 1 shows that with the non-irradiated protein, we obtain a ladder of 1:1, 1:2, 1:3 complexes (called CA complexes) containing one DNA fragment to which one, two or three proteins are bound.

With the irradiated protein, we observe that the higher the dose, the smaller the amount of CA complex. Concomitantly, we observe the emergence of a new CB complex, as a very well defined band, which then disappears at higher doses. Thus, for doses higher than 100 Gy, only the CB complex is detected. This means that ionising radiation induces a phenomenon more complex than a



**Figure 2.** Top, Retardation gel electrophoresis of mixtures of either non-irradiated or irradiated protein with an equimolar mixture (5 nM each) of specific 31 bp and non-specific 65 bp DNA fragments. Lane 0, no protein. lanes a, b and c, non-irradiated protein at 5 nM, 10 nM and 20 nM, respectively; lanes d, e, f and g, 125 Gy irradiated protein at 20 nM, 40 nM, 60 nM and 80 nM, respectively. The protein was irradiated at a concentration of 2  $\mu$ M. Bottom, Percentage of free 31 bp specific and free 65 bp non-specific DNA *versus* protein concentration.



**Figure 3.** A cartoon representing complexes between MC1 protein and DNA fragments of different lengths and bending angles. (A) and (B), Small fragment of DNA, (A) (strong bending) migrates faster than (B). (C) and (D), Large fragment of DNA, (D) (small bending) migrates faster than (C).

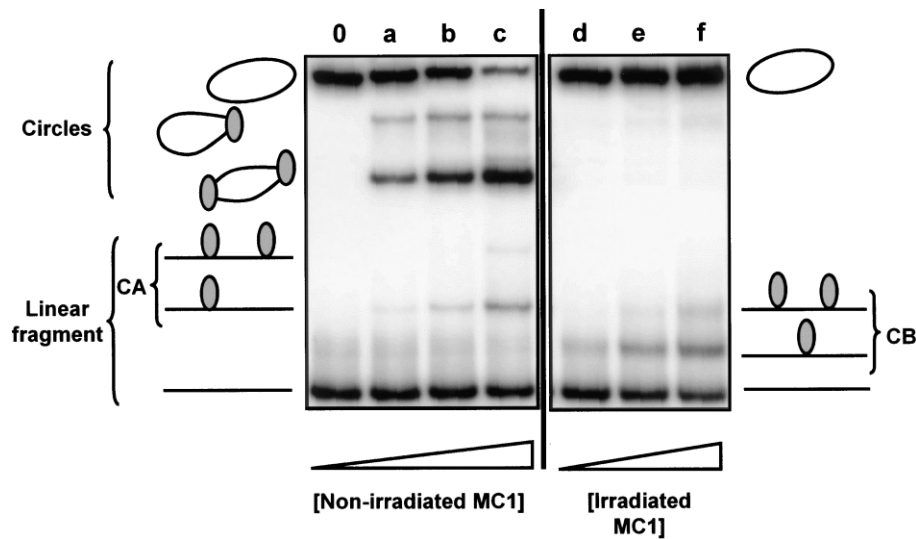
simple inactivation of the protein: some damaged proteins still bind DNA, but in a different manner.

The 64 bp DNA fragment bears in its central part a sequence that is recognised preferentially by the protein. According to Le Cam and co-workers,<sup>18</sup> MC1 binding induces a bending angle of about 110° at this site. This very strong bending leads to a reduction of the CA complex migration velocity. The enhanced mobility of CB complex may be due to (i) an averaged distribution of the bound protein everywhere along the DNA fragment due to the lack of specific site recognition, and/or (ii) a smaller bending angle at the binding site.<sup>20</sup>

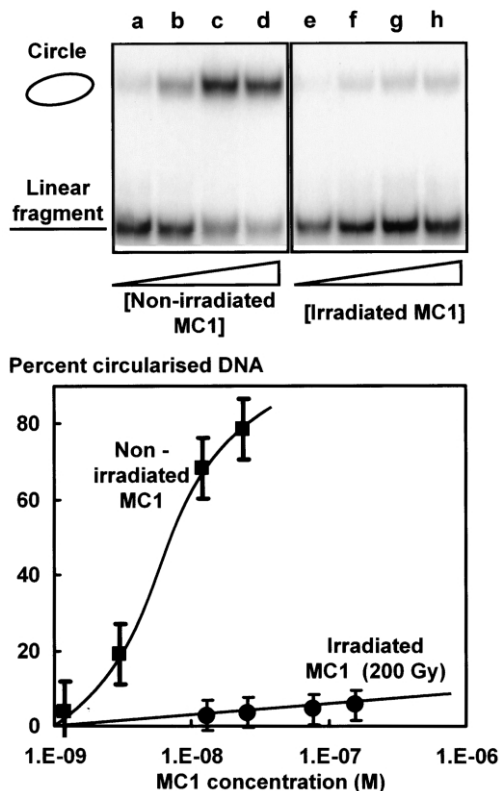
#### Discrimination between specific and non-specific DNA

Non-irradiated and irradiated protein were added to an equimolar mixture of 31 bp specific and 65 bp non-specific DNA fragments. [Figure 2](#) shows that the non-irradiated protein binds the 31 bp specific fragments preferentially, giving the CA complex, whereas it practically does not bind the 65 bp non-specific fragments.





**Figure 4.** Retardation gel electrophoresis of a mixture of either non-irradiated or irradiated (125 Gy) protein and an equimolar (1.5 nM each) mixture of open linear and circular closed 207 bp DNA fragments. The protein was irradiated at a concentration of 2  $\mu$ M. 0, No protein; a, b and c, non-irradiated protein at 4.5 nM, 9 nM and 18 nM, respectively; d, e and f, irradiated protein at 50 nM, 100 nM and 200 nM, respectively.



**Figure 5.** Circularisation assay. Left part, Polyacrylamide gel electrophoresis of 86 bp DNA, circularised with T4 ligase in the presence of various concentrations of either non-irradiated or 200 Gy irradiated MC1 protein. DNA concentration, 100 pM. Protein concentrations: (a)  $1.2 \times 10^{-9}$  M, (b)  $3 \times 10^{-9}$  M, (c)  $1.2 \times 10^{-8}$  M, (d)  $2.4 \times 10^{-8}$  M, (e)  $1.3 \times 10^{-8}$  M, (f)  $2.6 \times 10^{-8}$  M, (g)  $7.8 \times 10^{-8}$  M and (h)  $1.56 \times 10^{-7}$  M. Right part, Plot of the percentage circularised 86 bp DNA fragment as a function of the protein concentration (logarithmic scale).

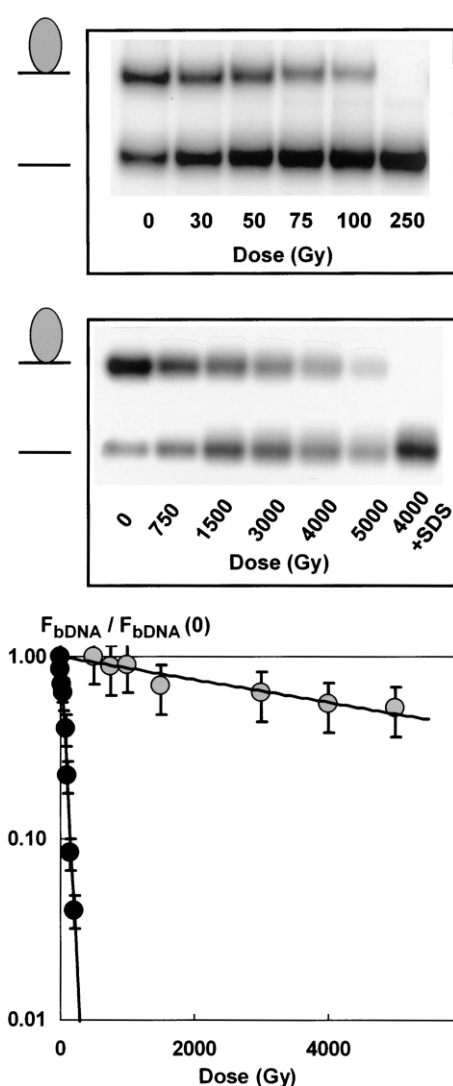
When the protein is irradiated, the CA complex with the 31 bp specific DNA disappears, and the CB complex is formed. With the 65 bp non-specific DNA, 1:1 and 1:2 CB complexes are formed.

Quantification of free DNA remaining unbound in their respective bands shows that the irradiated protein does not discriminate anymore between specific and non-specific DNA.

We remark that, compared to CA, CB complex migrates slightly more rapidly with 65 bp non-specific DNA fragments, as expected for a smaller bending angle,<sup>20</sup> and more slowly for 31 bp specific fragments. In the latter case (31 bp specific fragment), in the CA complex, the bending angle is large, and consequently the DNA should be wrapped on the protein, in close contact along most of its length. The diffusion coefficient of such a globular compact object would be relatively small. In the case of the CB complex, if the bending angle is smaller, the end parts of the DNA fragment are not in contact with the protein so tightly, and the diffusion coefficient may be larger (Figure 3). This could explain the slowing of the CB complex in the case of the 31 bp specific DNA fragment. This fact has been observed in the case of a 26 bp specific DNA fragment used for the model design (see below). As a consequence, a fragment length between 31 bp and 65 bp should lead to CA and CB complexes displaying the same velocity. It is probably the case for the 50 bp fragments (see below).

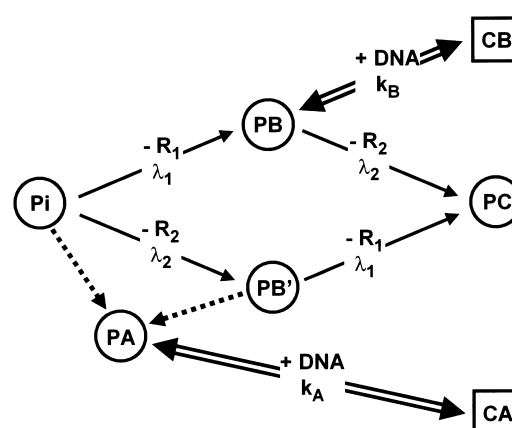
#### Reduction of bending: inactivation of circles recognition and loss of circularisation ability

An equimolar solution of open-linear and circularised 207 bp non-specific DNA fragment was mixed with increasing amounts of either



**Figure 6.** Top, Retardation gel electrophoresis of mixtures of 50 bp non-specific DNA fragments (7 nM) with irradiated protein (9 nM). The protein was irradiated at a concentration of 2  $\mu$ M. Middle, Retardation gel electrophoresis of irradiated mixtures of 50 bp non-specific DNA fragments (7 nM) and MC1 protein (12 nM). The concentrations in the irradiated mixtures were 2  $\mu$ M for protein and 1.16  $\mu$ M for DNA. Bottom, Fraction of DNA bound in the complexes as a function of irradiation dose (data from top and middle parts).  $F_{bDNA}$  is the fraction of DNA bound to the irradiated protein in the complexes.  $F_{bDNA}(0)$  is the fraction of DNA bound to the non-irradiated protein (dose 0).

non-irradiated or irradiated protein (Figure 4). In this case, the non-irradiated protein binds linear DNA slightly, and only a few 1:1 CA complexes are formed. As expected from previous works,<sup>16,21</sup> the non-irradiated protein has a great affinity for the circular DNA, since 1:1 and principally 1:2 complexes are formed. Preferential binding to circular DNA is generally interpreted as proof that the binding of the protein induces the curvature of DNA.<sup>20</sup> The irradiated protein does not bind the



**Figure 7.** A scheme representing the different pathways described in the hypotheses sustaining the model of degradation of the protein and the complexation with DNA.

circles. It binds only linear fragments to give 1:1 and 1:2 CB complexes.

Circularisation of DNA upon protein binding can be used to reveal the ability of the protein to bend DNA, as shown with HMGB protein.<sup>22</sup> The 86 bp non-specific DNA fragments were circularised with phage T4 ligase in the presence of increasing amounts of either non-irradiated or irradiated (200 Gy) MC1 protein. Very low concentrations of non-irradiated protein allow the efficient circularisation of the fragment, whereas large concentrations of irradiated protein are needed even for very small amounts of circles (Figure 5).

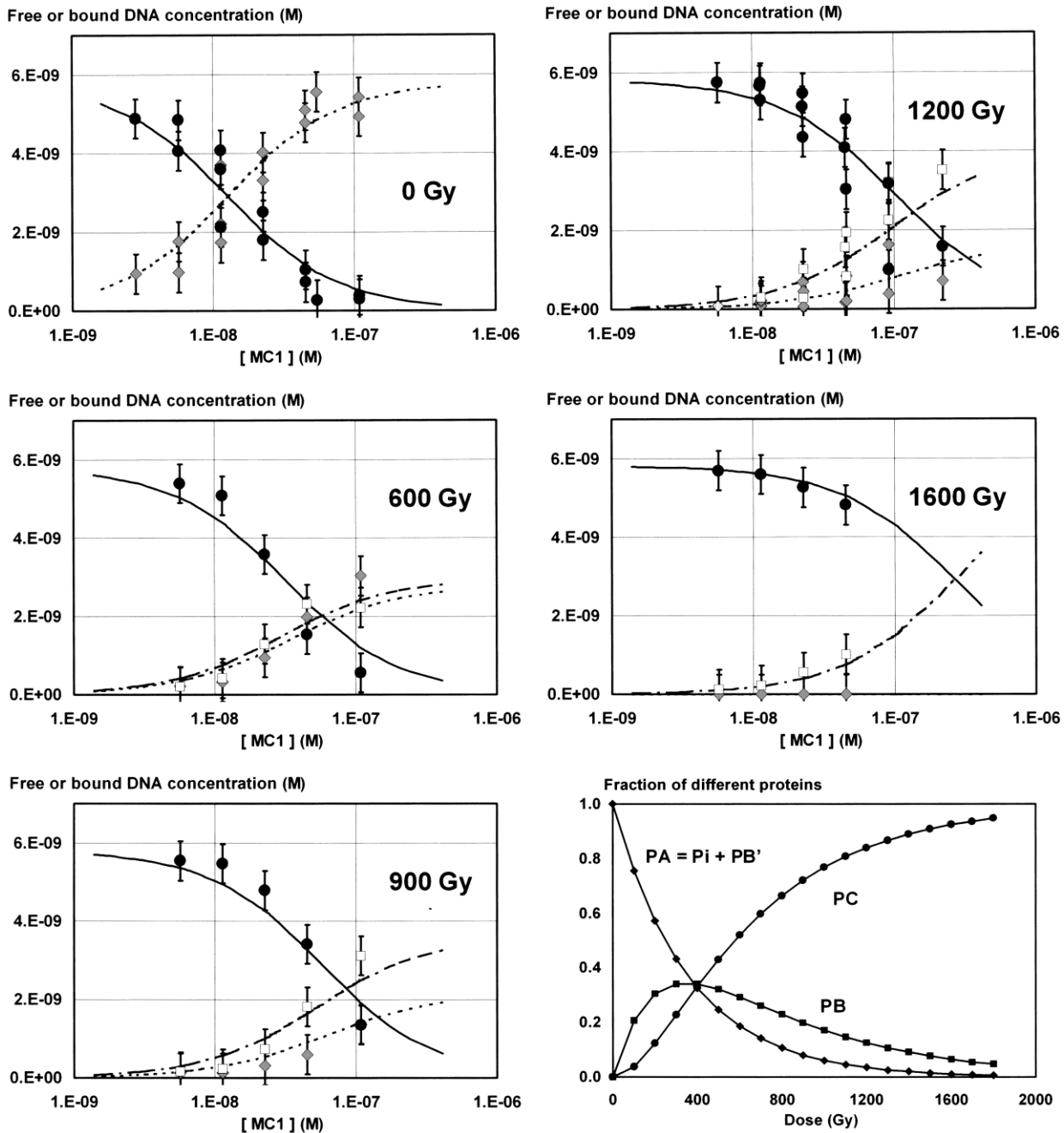
Both these experiments demonstrate the radiation-induced loss of bending power of the MC1 protein.

### Radioprotection of the protein by DNA

Irradiated protein was mixed with labelled 50 bp non-specific DNA, and the samples were submitted to retardation gel electrophoresis (Figure 6). As a result, the fraction of DNA present in the upper band corresponding to the complex decreases exponentially with increasing dose. Thus, the binding ability of the protein decreases when the dose increases.

In this case of a 50 bp non-specific fragment, only one band of complex can be observed. CA and CB complexes cannot be separated by retardation gel electrophoresis, since they may have equal migration velocities (see above).

To investigate the possible protection of MC1 by bound DNA, we irradiated the DNA-protein complex. When the 50 bp non specific DNA-protein mixture was irradiated after mixing, we observed an exponential decrease in the DNA-binding ability, but with a considerably smaller dose-response kinetic constant (Figure 6). Addition of SDS to a sample irradiated at 4000 Gy allowed a complete



**Figure 8.** Five panels (0, 600, 900, 1200 and 1600 Gy) show the results of the titration experiments obtained by retardation gel electrophoresis. The concentration of total DNA was 5.8 nM. The concentration of free DNA (black circles) and of DNA bound as the CA (grey diamonds) and CB (open squares) complexes are plotted as a function of the added protein concentration. In these experiments, the protein was irradiated at a concentration of 20  $\mu$ M. The curves (continuous for DNA, dotted for CA and dash/dot for CB) are calculated using the  $k_{A'}$ ,  $k_B$ ,  $\lambda_1$  and  $\lambda_2$  parameters giving the best fit (root-mean-square minimisation). The sixth panel (right bottom) shows the fraction of different types of proteins calculated using the above-determined  $\lambda_1$  and  $\lambda_2$  parameters.

recovery of free DNA. This shows that the upper bands attributed to a mixture of CA and CB complexes actually contain complexes, and not some irradiation-induced cross-linked species.

Thus the bound DNA protects the protein in the complex against radiolytic damage, and the protection factor in our experimental conditions is of the order of 100.

#### Model. Titrations of DNA by irradiated protein

MC1 protein is a monomer, and we demonstrated in a previous work that it binds DNA in the monomeric form.<sup>23</sup> Consequently, the distribution of both CA and CB complexes *versus* irradiation dose (Figure 1) cannot be explained by a simple destruction of the binding site on the

protein. This observation suggests a more sophisticated model involving two targets,  $R_1$  and  $R_2$ , strongly implicated in the DNA-binding process, each of them susceptible to destruction upon irradiation. Our model was drawn (Figure 7) from the following hypotheses.

(1) Both targets  $R_1$  and  $R_2$ , are sensitive to radiolysis (dose  $D$ ) with dose-response kinetic constants  $\lambda_1$  and  $\lambda_2$ , respectively:

$$\frac{d[R_1]}{dD} = -\lambda_1[R_1]$$

$$\frac{d[R_2]}{dD} = -\lambda_2[R_2]$$

The non-damaged MC1 protein (concentration  $[P_0]$  before irradiation) is called  $P_i$ . The proteins with either  $R_1$  or  $R_2$  destroyed are called  $P_B$  and  $P_B'$ , respectively. The protein with both  $R_1$  and  $R_2$  destroyed is called  $P_C$ .

(2) Damage to other less sensitive sites of the protein are not taken into account. We can therefore write:

$$[P_i] = [P_0]e^{-(\lambda_1+\lambda_2)D} \quad (1)$$

$$[P_B] = [P_0](1 - e^{-\lambda_1 D})e^{-\lambda_2 D} \quad (2)$$

$$[P_B'] = [P_0](1 - e^{-\lambda_2 D})e^{-\lambda_1 D} \quad (3)$$

$$[P_C] = [P_0](1 - e^{-\lambda_1 D})(1 - e^{-\lambda_2 D}) \quad (4)$$

(3) As long as  $R_1$  remains intact, and even if  $R_2$  is destroyed, the protein keeps its full activity (it behaves as the native protein, and leads to the CA complex). Such a protein is called PA:

$$[PA] = [P_i] + [P_B'] = [P_0]e^{-\lambda_1 D} \quad (5)$$

The destruction of  $R_1$  and the conservation of  $R_2$  lead to a protein that binds DNA differently. This protein, called  $P_B$ , gives the CB complex.

Destruction of both  $R_1$  and  $R_2$  residues leads to the protein  $P_C$ , which does not bind DNA.

(4) When DNA is added to the irradiated protein solution, the complexation with PA and  $P_B$  proteins follows the equilibrium equations:



The equilibrium constants  $k_A$  and  $k_B$  are equal to:

$$k_A = \frac{[CA]}{[PA_f][DNA_f]} \quad (6)$$

$$k_B = \frac{[CB]}{[P_B_f][DNA_f]} \quad (7)$$

where subscript f corresponds to free species.

The total concentration of protein is:

$$[P_0] = [PA] + [PB] + [PC] \quad (8)$$

with  $[PA] = [PA_f] + [CA]$  and  $[PB] = [P_B_f] + [CB]$

The total concentration of DNA is:

$$[DNA] = [DNA_f] + [CA] + [CB] \quad (9)$$

The experiments supporting this model were performed as follows. Mixtures of  $^{32}P$ -labeled DNA fragments at 5.8 nM and increasing concentrations of irradiated protein were submitted to retardation gel electrophoresis. In each lane, the amounts of DNA either free or complexed as CA or CB were quantified. A short fragment of 26 bp was chosen to avoid the presence of 1:2 complexes.

Titration with the non-irradiated protein allowed us to calculate  $k_A$ . Introducing  $k_A$  into the calculation and solving the system of equations (6)–(9),  $k_B$  can be found if the proportions of PA, PB and PC for each irradiation dose, i.e.  $\lambda_1$  and  $\lambda_2$ , are known. The determination of  $\lambda_1$  and  $\lambda_2$ , and consequently of  $k_B$ , was done by minimising the root-mean-square differences between experimental data and calculated values over all experimental results. We have obtained:

$$k_A = 1.0(\pm 0.3) \times 10^8 \text{ M}^{-1}$$

$$k_B = 6.8(\pm 1.5) \times 10^7 \text{ M}^{-1}$$

$$\lambda_1 = 28(\pm 2) \times 10^{-4} \text{ Gy}^{-1}$$

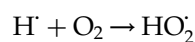
$$\lambda_2 = 17(\pm 4) \times 10^{-4} \text{ Gy}^{-1}$$

Figure 8 shows the experimental data and the fitted titration curves calculated from the above values. The calculated fractions of the different types of proteins (PA, PB and PC) are plotted in the same Figure. The shape of PA and PB curves is consistent with the pattern of complexes shown in Figure 1.

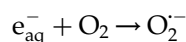
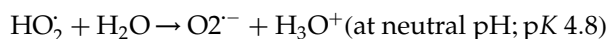
## Discussion

In our experimental conditions, dilute aerated aqueous solutions of protein were irradiated by photons of energy close to 0.6 MeV. Consequently, the direct effects following the ionisation of the protein by Compton or photoelectric effect can be neglected. In fact, the protein is attacked by some reactive species issued from the radiolysis of water, which are primarily  $\text{OH}^\cdot$  (oxidant) and  $\text{H}^\cdot$  radicals and hydrated electrons  $e_{\text{aq}}^-$  (reducers).

In the presence of  $3 \times 10^{-4} \text{ M}$  oxygen (air-saturated solutions), the reducing agents  $\text{H}^\cdot$  and  $e_{\text{aq}}^-$  are eliminated rapidly, since their rate constants of reaction with the molecular oxygen are  $2.1 \times 10^{10} \text{ M}^{-1} \text{ s}^{-1}$  and  $1.9 \times 10^{10} \text{ M}^{-1} \text{ s}^{-1}$ , respectively:<sup>7</sup>



and:



Hydroxyl radical  $\text{OH}^\cdot$  is susceptible to dimerise to produce hydrogen peroxide:



with a rate constant of  $5.5 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$ .<sup>7</sup> In fact, this reaction is not complete, due to the low concentration of  $\text{OH}^\cdot$ , and the steady-state concentration of  $\text{OH}^\cdot$  in the bulk remains large enough to attack the protein.

As a result, the main oxidative species present in the solution are  $\text{OH}^\cdot$ ,  $\text{O}_2^{\cdot-}$  and  $\text{H}_2\text{O}_2$ , which all contribute to oxidative stress. The hydroxyl radical  $\text{OH}^\cdot$  is by far the most reactive towards biological material.

For some experiments, the protein was irradiated at low concentration, i.e.  $2 \mu\text{M}$ . As shown in Figure 1, the initial radiolytic yield for CA complex destruction (slope at the origin of the inactivation curve) is around  $4 \times 10^{-8} \text{ mol J}^{-1}$ . For protein irradiated at high concentration, i.e.  $20 \mu\text{M}$  (Figure 8), the initial radiolytic yield is of the same order of magnitude. In fact, in both cases, we operated near or at the plateau of the dilution curve.<sup>24</sup> In pure water, the radiolytic yield of  $\text{OH}^\cdot$  radical formation is  $2.8 \times 10^{-7} \text{ mol J}^{-1}$ , seven times the radiolytic yield of CA complex destruction. This does not mean that seven  $\text{OH}^\cdot$  radicals are needed to inactivate the CA complex formation. In fact,  $\text{OH}^\cdot$  radicals have reacted with themselves, and with scavengers. In our experiments, the irradiation medium was not pure water, but a buffer solution (1 mM potassium phosphate, pH 7.25) containing unavoidable impurities and traces of oxidisable compounds used for bacteria growing, protein purification and conservation. All of these products may be good or even excellent  $\text{OH}^\cdot$  radical scavengers. Thus, not all the  $\text{OH}^\cdot$  radicals produced by water radiolysis attack the protein.

To understand the behaviour of the irradiated protein, some features concerning the DNA-MC1 binding have to be recalled here. MC1 protein is a basic protein (24% Arg + Lys) probably involved in the structuring of the chromosome. In methanosarcina, it is naturally bound to DNA, in an average amount of one protein molecule per 100 bp.<sup>12</sup> *In vitro*, purified MC1 is able to bind to any DNA by electrostatic interaction. Moreover, some preferential structures or sequences have been identified. The protein has an enhanced affinity for circular<sup>23</sup> and cruciform DNA,<sup>17</sup> and for short sequences present in the genome of the archæum.<sup>13</sup> This sequence specificity, evidenced by DNA footprinting and retardation gel electrophoresis, corresponds to an increase of about two orders of magnitude in the binding constant (data not shown).

The most important result is that upon

irradiation, the MC1 protein loses its genuine DNA interaction properties. For low doses the DNA-binding ability of the protein is preserved, but one observes: (1) the loss of circle recognition and circularisation ability that demonstrate the loss of DNA bending power; (2) the loss of recognition of specific-binding sites. At this stage, damage is sufficient to abolish the specific site recognition and to induce different structural modifications to the bound DNA.

For larger doses, the binding ability is abolished completely. This behaviour supports our model very well. What does this model mean in terms of complex formation during the irradiation process?

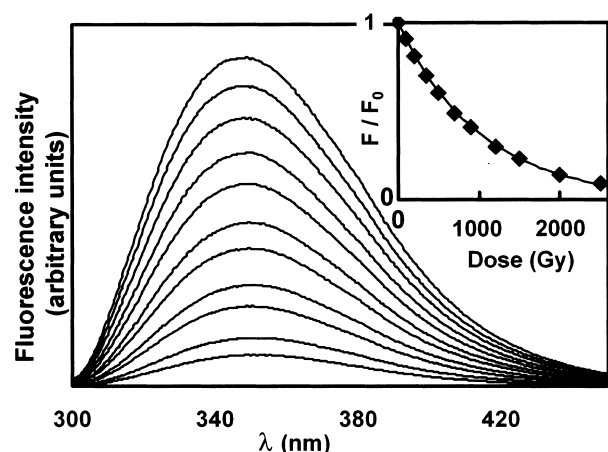
On one hand, one (or may be several) amino acid side-chain,  $R_1$ , is oxidised rapidly. This amino acid residue is involved in the DNA-binding process, implicated strongly in the recognition of the specific site, and participates in DNA bending. Its destruction leads to the protein PB, which gives the CB complex. It is likely that several factors are involved in the bending process of DNA.  $R_1$  is probably only one of these factors, and thus its destruction could not abolish the bending completely. The protein is positively charged, and the neutralisation of DNA phosphate groups by the non-destroyed Lys and Arg may be sufficient to bend the fragments, but to a lower extent than in the CA complex.<sup>25</sup>

On the other hand, another (or may be several other) amino acid side-chain,  $R_2$ , is oxidised concomitantly but more slowly. Its involvement in DNA binding and bending is not determining, and it is not implicated directly in the recognition of the specific site process. The destruction of  $R_2$  alone does not prevent an almost normal DNA binding and bending. Protein with only  $R_2$  oxidised corresponds to PB', and gives, like native MC1, a CA complex.

The occurrence of both damages on the same protein leads to the complete loss of DNA-binding activity, as far as we can conclude from the absence of complexes in the retardation gels.

In the model presented above, we implicitly consider that  $R_1$  and  $R_2$  belong to a unique-binding site. Another model could be considered with two contemporary, geometrically distinct-binding sites. The site that includes  $R_1$  would be preferential. The other-binding site, which includes  $R_2$ , would be weaker, and would be revealed only after the destruction of the first one. The first site would be the only one detectable on the non-irradiated protein if the binding constants for the two sites are highly different. But the two binding constants  $k_A$  and  $k_B$  are not very different, and therefore the second complex should have been detected when increasing the concentration of the non-irradiated protein. Since we never observed this second complex with the non-irradiated protein, even at very high concentration, this latter model can be withdrawn.

The protection due to DNA binding (Figure 5) is very efficient. Such a protection of protein by DNA



**Figure 9.** Fluorescence spectra of MC1 protein irradiated with increasing doses. Top to bottom: 0, 100, 200, 350, 500, 700, 900, 1200, 1500, 2000, 2500 Gy. The protein concentration was  $1 \times 10^{-4}$  M; the pathlength was 5 mm. Insert: relative decrease of the fluorescence spectra areas (from 320 nm to 380 nm) as a function of the dose.

has been observed in the *E. coli lac* repressor–*lac* operator complex. This protection is mainly due to the masking of the binding site of the protein by DNA.<sup>10</sup> Moreover in the cases of *lac* repressor–operator and MC1–DNA interacting systems, protection of DNA by the bound protein has been observed<sup>19,26</sup> and was interpreted in terms of masking of the binding sites on DNA.<sup>27</sup> There is cross-protection of the two partners of the complex: one partner protects its binding site on the other partner.

The protection factor due to DNA binding is more important in the case of MC1 (100) than in the case of *lac* repressor (4.5).<sup>11</sup> This can be due to a difference in their binding process. The MC1 is a small monomeric protein, with only one binding site for DNA, whereas the repressor is a large tetrameric allosteric protein, with two DNA-binding sites. Its activity is under the control of an effector, whose binding site is located far away from those of DNA. There are probably several pathways, in addition to the destruction of the directly interacting residues, to inactivate, at least partially, the repressor. The effector-binding site, for instance, is not protected by the bound DNA. Such an observation was made by inducing photochemical modification of Trp residues.<sup>28</sup>

Which amino acids have been damaged? MC1 protein contains neither cysteine nor tyrosine, which are among the most oxidisable amino acids. But it contains two tryptophan, one methionine, four phenylalanine and two histidine residues, whose rate constants for OH<sup>•</sup> attack are high,<sup>7</sup> in the range  $1.3 \times 10^{10}$ – $5 \times 10^9$  M<sup>-1</sup> s<sup>-1</sup>. Some of these residues are obviously the best candidates for being responsible for the protein inactivation. Fluorescence measurements of the irradiated protein show that indeed tryptophan residues are destroyed until abolishment of the intrinsic fluor-

escence of the protein (Figure 9). Using photo-cross-linking with 5-bromouracil-substituted DNA, we showed previously that Trp74 is implicated in the DNA-binding process.<sup>29</sup> Furthermore, NMR spectroscopy data recently demonstrated that this residue is exposed to the solvent, in an external loop of the protein, and that its resonance parameters are significantly modified upon DNA binding (F. Paquet, personal communication).

Met75 could be a target for the oxidative process. It is located in the same structural environment as Trp74, and the oxidation of methionine residues has been evoked in the protein response to the oxidative stress.<sup>30</sup>

Besides Trp74 and Met75, other amino acid residues could be responsible, at various levels, for the oxidative inactivation of the protein. Work is in progress in our laboratory to identify the damage induced by the oxidative stress.

## Materials and Methods

The MC1 protein was purified in our laboratory from *Methanosarcina* sp. CHT155 as described.<sup>12,31</sup> The protein concentration was determined by absorption spectrophotometry, using a molecular absorbance of 11,000 M<sup>-1</sup> cm<sup>-1</sup> at 280 nm.

Several labeled DNA fragments were prepared for the experiments. In all cases, labelling at the 5'-end was done using phage T4 polynucleotide kinase (Eurogentec), and [ $\gamma$ -<sup>32</sup>P]ATP (5000 Ci mmol<sup>-1</sup>; Amersham).

The 64 bp and 50 bp fragments were obtained by *Hin*I digestion of a 114 bp DNA fragment obtained by PCR using a pUC 9-derived plasmid containing 1.2 kb sequence upstream from the gene coding for the methyl CoM in the *Methanosarcina* genome.<sup>13</sup> The sequence of the top-strand of the 64 bp duplex is: 5'-ACTCCATGGTCTATATAAAAAAGAGCATCATCAGGCTTGTTGAAACCTTTCCGCTTGTGTACC. The sequence recognised preferentially by the protein is in bold. The 50 bp fragment (sequence not shown) does not exhibit any preferential-binding region.

The 207 bp DNA minicircles were obtained as described.<sup>16</sup> Briefly, a 207 bp DNA fragment with AATT cohesive ends at a concentration of 0.1  $\mu$ g/ml was incubated in the presence of phage T4 DNA ligase (Invitrogen). The circular closed DNA was separated from linear species and nicked minicircles by preparative polyacrylamide gel electrophoresis (4% (w/v) acrylamide, 0.2% (w/v) bisacrylamide) in the presence of 0.6  $\mu$ g/ml of ethidium bromide, in gel and buffer.

The 26 bp DNA fragment was prepared from the commercial oligonucleotide (Eurogentec) 5'-GAAAAA GAGCATCATCAGGCTTGTT. After labelling at the 5' end, the oligonucleotide was annealed to the complementary strand by incubation at 80 °C for one minute, and then cooled slowly to room temperature. The sequence recognised preferentially by the protein is in bold.

The 65 bp labeled pool of random sequence bearing DNA fragments was obtained by annealing the 5'-labeled oligonucleotide: 5'-CAGGTCAGTTCAGC GGATCCTGTCG-(N)<sub>15</sub>-GAGGCGAATTCAGTGCAACT GCAGC, where N represents any A, T, G or C base, with the complementary strand.

After 14 cycles of SELEX procedure using decreasing amounts of MC1 protein, 65 bp fragments with increased affinity for the protein were selected. The 31 bp DNA duplexes result from cutting of these selected fragments with *EcoRI* and *BamHI* enzymes (underlined sites).

The 86 bp DNA fragment was obtained by PCR on the plasmid pSL1180, using the two primers 5'-CGAGA GAATTCTGTCATGATGATCATGCAATT, and 5'-TTA CGAATTCGAATGGCCATG, followed by digestion with *EcoRI*, and purified by electrophoresis on a 10% polyacrylamide gel.

### Gel-binding assays

Complexes were formed by addition of the protein to DNA duplexes at various concentrations of protein and DNA in 1 mM potassium phosphate (pH 7.25), in the presence of various concentrations of NaCl: 60 mM for experiments shown in Figures 1, 5, and 6, 150 mM for Figure 2, and 100 mM for Figure 4. After equilibration for 15 minutes at 4 °C, samples were mixed with 0.2 volume of loading buffer (0.01% (w/v) bromophenol blue, 0.01% (w/v) xylene cyanol, 50% (v/v) glycerol, 1 mM potassium phosphate (pH 7.25), 60 mM NaCl). They were loaded onto a polyacrylamide gel (6% acrylamide, 0.2% bisacrylamide) in TBE buffer (89 mM Tris, 89 mM boric acid, 1 mM EDTA, pH 8). Electrophoreses were run at 14 V m<sup>-1</sup> for four hours at 4 °C. After drying, gels were quantified using a  $\beta$ -scanner (Storm, Molecular Dynamics).

### Circularisation assay

The circularisation assays were performed in 50  $\mu$ l samples of ligase buffer (50 mM Tris-HCl (pH7.6), 10 mM MgCl<sub>2</sub>, 1 mM ATP, 1 mM DTT, 5% (w/v) polyethylene glycol 8000). The <sup>32</sup>P-labelled 86 bp fragments with *EcoRI* ends was 100 pM. Various concentrations of either non-irradiated or irradiated MC1 protein were added (Figure 5).

Before the addition of 0.5 unit of T4 ligase (Invitrogen), 10  $\mu$ l of the DNA-MC1 mixture was submitted to retardation polyacrylamide (6%) gel electrophoresis, to control complex formation.

The incubation with ligase was performed at 37 °C for two hours. The products of ligation were extracted with phenol/chloroform (1:1, v/v), precipitated with ethanol and loaded onto a polyacrylamide gel (8% acrylamide, 0.4% bisacrylamide). Circular DNA was identified by its resistance to exonuclease III (Promega) digestion.

### Irradiations

Aqueous solutions of either protein alone, or a mixture of protein and DNA, were irradiated by  $\gamma$ -rays of 0.6 MeV in a IBL437 irradiator containing <sup>137</sup>Cs sources (CisBio International). The dose rate at the point of interest in a water/ice bath maintained at 0 °C was 9 Gy min<sup>-1</sup> measured by Frick's dosimetry.

The protein was generally irradiated at a low concentration, 2  $\mu$ M, in 1 mM potassium phosphate buffer (pH 7.5), 60 mM NaCl. Under these conditions, the irradiation doses ranged from 0 Gy to 250 Gy to observe a total inactivation of the protein. For the titration experiments used to establish the model, higher concentrations (20  $\mu$ M) of protein were needed. It was then necessary to increase the doses, in the range 0 Gy to

2500 Gy, in order to observe the same inactivation (see Discussion).

### Fluorescence spectra

Fluorescence spectra in the tryptophan band were recorded from 300 to 450 nm with a Jobin-Yvon Fluoromax2 spectrofluorimeter. The excitation wavelength was 295 nm, and the bandwidths were 5 nm for both excitation and emission. Spectra were corrected for the buffer background and Raman effect of water.

### Acknowledgements

We thank Dr Françoise Paquet, who gave us the structure of the MC1 protein established by NMR before publication. Part of this work was supported by La Ligue contre le Cancer (Comité du Loiret).

### References

- Berlett, B. S. & Stadtman, E. R. (1997). Protein oxidation in aging, disease and oxidative stress. *J. Biol. Chem.* **272**, 20313–20316.
- Sohal, R. S. (2002). Role of oxidative stress and protein oxidation in the aging process. *Free Radic. Biol. Med.* **33**, 37–44.
- Buxton, G. V. (1987). Radiation chemistry of the liquid state: (1) water and homogeneous aqueous solutions. In *Radiation Chemistry. Principles and Applications* (Farhataziz, & Rodgers, A. J., eds), pp. 321–349, VCH Publishers, New York, Weinheim.
- Garrison, W. M. (1987). Reaction mechanisms in the radiolysis of peptides, polypeptides, and proteins. *Chem. Rev.* **87**, 381–398.
- Stadtman, E. R. (1993). Oxidation of free amino acids and amino acid residues in protein by radiolysis and by metal-catalysed reactions. *Annu. Rev. Biochem.* **62**, 797–821.
- Dean, R. T., Fu, S., Stocker, R. & Davies, M. J. (1997). Biochemistry and pathology of radical-mediated protein oxidation. *Biochem. J.* **324**, 1–18.
- Buxton, G. V., Greenstock, C. L., Helman, W. P. & Ross, A. B. (1988). Critical review of rate constants for reactions of hydrated electrons, hydrogen atoms and hydroxyl radicals ( $\text{OH}^\bullet/\text{O}^\bullet$ ) in aqueous solution. *J. Phys. Chem. Ref. Data*, **17**, 513–886.
- Finley, E. L., Dillon, J., Crouch, R. K. & Schey, K. L. (1998). Radiolysis-induced oxidation of bovine  $\alpha$ -crystallin. *Photochem. Photobiol.* **68**, 9–15.
- Audette, M., Blouquit, Y. & Houée-Lévin, C. (2000). Oxidative dimerization of protein: role of tyrosine accessibility. *Arch. Biochem. Biophys.* **376**, 217–220.
- Eon, S., Culard, F., Sy, D., Charlier, M. & Spothem-Maurizot, M. (2001). Radiation disrupt protein-DNA complexes through damage to the protein. The lac repressor-operator system. *Radiat. Res.* **156**, 110–117.
- Charlier, M., Eon, S., Sèche, É., Bouffard, S., Culard, F. & Spothem-Maurizot, M. (2002). Radiolysis of lac repressor by  $\gamma$ -rays and heavy ions: a two-hit model for protein inactivation. *Biophys. J.* **82**, 2373–2382.
- Chartier, F., Laine, B., Sautière, P., Touzel, J. P. &

- Albagnac, G. (1985). Characterization of the chromosomal protein Hmb isolated from *Methanosarcina barkeri*. *FEBS Letters*, **183**, 119–123.
13. Teyssier, C., Laine, B., Gervais, A., Maurizot, J.-C. & Culard, F. (1994). The archaeobacterial histone-like protein MC1 can exhibit a sequence specific binding to DNA. *J. Biochem.* **303**, 567–573.
  14. Webb, M. & Thomas, J. O. (1999). Structure-specific binding of the two tandem HMG boxes of HMG1 to four-way junction DNA is mediated by the A domain. *J. Mol. Biol.* **294**, 373–387.
  15. Webb, M., Payet, D., Keng-Boon, L., Travers, A. & Thomas, J. O. (2001). Structural requirements for cooperative binding of HMG1 to DNA minicircles. *J. Mol. Biol.* **309**, 79–88.
  16. Toulmé, F., Le Cam, E., Teyssier, C., Delain, E., Sautière, P., Maurizot, J.-C. & Culard, F. (1995). Conformational changes of DNA minicircles upon the binding of the archaeobacterial histone-like protein MC1. *J. Biol. Chem.* **270**, 6286–6291.
  17. Paradinas, C., Gervais, A., Maurizot, J.-C. & Culard, F. (1998). Structure-specific binding recognition of a methanogen chromosomal protein. *Eur. J. Biochem.* **257**, 372–379.
  18. Le Cam, E., Culard, F., Larquet, E., Delain, E. & Cognet, J. A. (1999). DNA bending induced by the archaeobacterial histone-like protein MC1. *J. Mol. Biol.* **285**, 1011–1021.
  19. Isabelle, V., Franchet-Beuzit, J., Sabattier, R., Laine, B., Spothheim-Maurizot, M. & Charlier, M. (1993). Radioprotection of DNA by a DNA-binding protein: MC1 chromosomal protein from the archaeobacterium *Methanosarcina* sp. CHT155. *Int. J. Radiat. Biol.* **63**, 749–758.
  20. Sitlani, A. & Crothers, D. M. (1998). DNA-binding domains of Fos and Jun do not induce DNA curvature: an investigation with solution and gel methods. *Proc. Natl Acad. Sci. USA*, **95**, 1404–1409.
  21. Teyssier, D. C., Toulmé, F., Touzel, J.-P., Gervais, A., Maurizot, J.-C. & Culard, F. (1996). Preferential binding of the archaeobacterial histone like MC1 protein to negatively supercoiled DNA minicircles. *Biochemistry*, **35**, 7954–7958.
  22. Grasser, K. D., Teo, S.-H., Lee, K.-B., Broadhurst, R. W., Rees, C., Hardman, C. H. & Thomas, J. O. (1998). DNA-binding properties of the tandem HMG boxes of high-mobility-group protein 1 (HMG1). *Eur. J. Biochem.* **253**, 787–795.
  23. Culard, F., Laine, B., Sautière, P. & Maurizot, J.-C. (1993). Stoichiometry of the binding of chromosomal protein MC1 from the archaeobacterium, *Methanosarcina* sp. CHT155, to DNA. *FEBS Letters*, **315**, 335–339.
  24. Ferradini, C. & Jay-Gerin, J.-P. (1999). La radiolyse de l'eau et des solutions aqueuses: historique et réalité. *Can. J. Chem.* **77**, 1542–1575.
  25. Strauss, J. K. & Maher, L. J., III (1994). DNA bending by asymmetric phosphate neutralisation. *Science*, **266**, 1829–1834.
  26. Franchet-Beuzit, J., Spothheim-Maurizot, M., Sabattier, R., Blazy-Baudras, B. & Charlier, M. (1993). Radiolytic footprinting.  $\beta$  rays,  $\gamma$  photons and fast neutrons probe DNA–protein interactions. *Biochemistry*, **32**, 2104–2110.
  27. Begusova, M., Eon, S., Sy, D., Culard, F., Charlier, M. & Spothheim-Maurizot, M. (2001). Radiosensitivity of DNA in a specific protein–DNA complex: the *lac* repressor–*lac* operator complex. *Int. J. Radiat. Biol.* **77**, 645–654.
  28. Spothheim-Maurizot, M., Culard, F. & Charlier, M. (1987). Photochemical modifications of *lac* repressor. II. Tryptophan photochemistry as a probe in studying the allosteric behaviour of the protein. *Photochem. Photobiol.* **46**, 15–21.
  29. Katouzian-Safadi, M., Laine, B., Chartier, F., Cremet, J.-Y., Belaiche, D., Sautiere, P. & Charlier, M. (1991). Determination of the DNA-interacting region of the archaeobacterial chromosomal protein MC1. Photocrosslinks with 5-bromouracil-substituted DNA. *Nucl. Acids Res.* **19**, 4937–4941.
  30. Dow, L. K., Changela, A., Hefner, H. E. & Churchill, M. E. A. (1997). Oxidation of a critical methionine modulates DNA binding of the *Drosophila Melanogaster* high mobility group protein, HMG-D. *FEBS Letters*, **414**, 514–520.
  31. Laine, B., Culard, F., Maurizot, J.-C. & Sautière, P. (1991). The chromosomal protein MC1 from archaeobacterium *Methanosarcina* sp. CHT155 induces DNA bending and supercoiling. *Nucl. Acids Res.* **19**, 3041–3045.

Edited by J. O. Thomas

(Received 2 December 2002; received in revised form 20 February 2003; accepted 7 March 2003)



## **Guillaume DE VUYST**

Mots clefs : interaction protéine/ADN, MC1, empreintes moléculaires, SELEX, courbure ADN

### **La protéine MC1 d'archaeobactérie : reconnaissance de séquences particulières.**

La protéine MC1 est une petite protéine structurale très abondante chez *Methanosarcina thermophila* (archaeobactérie). Nous avons utilisé la méthode SELEX (Systematic Evolution of Ligands by EXponential enrichment) pour rechercher ses séquences préférentielles de fixation. Après 10 cycles de sélection, une séquence consensus avec une affinité 50 fois plus forte que celle pour une séquence aléatoire a été déterminée. Cette séquence a été étudiée par simulation de dynamique moléculaire. Le mode de fixation de MC1 sur l'ADN a ensuite été déterminé par des empreintes moléculaires (DMS, OH<sup>\*</sup>). Les résultats montrent une fixation par une face et dans le petit sillon de l'ADN. Une reconnaissance par lecture indirecte des séquences est probable. L'oxydation de certains acides aminés (Trp, Met) par les rayons gamma entraîne une modification des propriétés de la protéine : perte de la reconnaissance de structures et séquences particulières et diminution de sa capacité à courber l'ADN.

---

Key words : Protein-DNA interaction; MC1; molecular footprinting; SELEX, DNA bending

### **The archaeal MC1 protein : recognition of special sequences**

The structural MC1 protein is a small and very abundant protein in *Methanosarcina thermophila*. To identify DNA sequences that are recognised by this protein we have performed SELEX experiments (Systematic Evolution of Ligands by EXponential enrichment) with a pool of DNA duplexes containing a central randomised N15 region. The alignment of DNAs after 10 rounds of selection reveals a 15 bp consensus sequence. The protein affinity for the consensus sequence is about 50 times higher than for the initial pool. The binding mode of the protein on DNA has been studied by molecular footprintings (DMS, hydroxyl radicals). The protein interacts via the minor groove at two distinct sites located on the same face of the DNA. The more likely is that the protein recognises the DNA sequences via an indirect readout process. The gamma radiation-induced oxidation of certain amino acids (Trp, Met) lead to the loss of recognition of particular sequences and of bent DNA is lost.