

50376
1989
121

UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE
FLANDRES ARTOIS

50376
1989
121

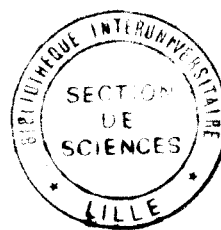
N° d'ordre: 343

THESE

Présentée pour l'obtention du grade de
DOCTEUR DE L'UNIVERSITE
DES SCIENCES ET TECHNIQUES DE LILLE
FLANDRES ARTOIS
Option Biochimie

par

François CHARTIER



**LA PROTEINE CHROMOSOMALE NC1 ISOLEE D'ARCHAEBACTERIES
DE LA FAMILLE DES METHANOSARCINACEAE
STRUCTURE PRIMAIRE ET INTERACTION AVEC L'ADN**

Soutenue le 2 mai 1989 devant la commission d'examen:

President:	Professeur	J. MONTREUIL
Rapporteurs:	Professeur	J. KREMBEL
	Docteur	M. CHARLIER
Examineurs:	Docteur	P. SAUTIERE
	Docteur	B. LAINE
	D.R. INRA	G. ALBAGNAC

A mes professeurs, aux chercheurs, aux techniciens qui m'ont formé, m'ont conseillé et épaulé,

Aux membres de ma famille et à mes amis qui ont toujours su me soutenir et m'encourager,

A tous, j'adresse un grand MERCI.

Vous avez tous contribué à l'élaboration de cette thèse.

Ce travail a été réalisé à l'Unité de Recherche Associée 409 du CNRS "Structure et Interactions des Protéines Associées au DNA des Eucaryotes et des Procaryotes" (Directeur : Docteur Pierre SAUTIERE) au sein de l'Institut de Recherches sur le Cancer (Directeur : Professeur Gérard BISERTE) dans le groupe "Protéines associées au DNA des Procaryotes" dirigé par le Docteur Bernard LAINE.

Ce travail a été effectué en collaboration avec la Station de Technologie Alimentaire de l'INRA de Villeneuve d'Ascq (Directeur : Guy ALBAGNAC, Directeur de Recherches INRA). Dans le cadre du Pôle Régional des Anaérobies de la Région Nord/Pas-de-Calais, il a bénéficié des contrats EPR 50 9180, FRT 86 E2 et FRT 87 C2. Il a également bénéficié de l'aide du CNRS et de l'Université de Lille II.

En outre, ce travail n'aurait pu être mené à bien sans l'attribution d'une allocation de recherche par le Ministère de la Recherche et de l'Enseignement Supérieur.

Ce travail a fait l'objet des publications et des communications par affiches suivantes :

Publications

- Characterization of the chromosomal protein HMb isolated from *Methanosarcina barkeri*.
CHARTIER, F., LAINE, B., SAUTIERE, P., TOUZEL, J.P. and ALBAGNAC, G. (1985)
FEBS Lett. 183, 119-123.

- Primary structure of the chromosomal protein HMb from the archaeobacterium *Methanosarcina barkeri*.
LAINE, B., CHARTIER, F., IMBERT, M., LEWIS, R. and SAUTIERE, P. (1986)
Eur. J. Biochem. 161, 681-687.

- Features of the primary structure of the *Methanosarcina barkeri* chromosomal protein HMb compared to those of eukaryotic and prokaryotic chromosomal proteins.
IMBERT, M., CHARTIER, F., LAINE, B. and SAUTIERE, P. (1986)
in Biology of Anaerobic Bacteria (H.C. Dubourguier et al., eds)
pp 236-241, Elsevier Science Publisher.

- Characterization of the chromosomal protein MC1 from the thermophilic archaeobacterium *Methanosarcina* sp CHTI 55 and its effect on the thermal stability of DNA.
CHARTIER, F., LAINE, B. and SAUTIERE, P. (1988)
Biochim. Biophys. Acta 951, 149-156.

- High-performance liquid chromatographic separation of variants of chromosomal proteins from prokaryotes.
CHARTIER, F., CREVEL, G., LAINE, B. and SAUTIERE, P. (1989)
J. Chromatogr. 466, 331-337.

- Primary structure of the chromosomal protein MC1 from the archaeobacterium *Methanosarcina* sp CHTI 55.
CHARTIER, F., LAINE, B., BELAÏCHE, D., TOUZEL, J.P. and SAUTIERE, P. (1989)
Biochim. Biophys. Acta sous presse.

- Primary structure of the chromosomal proteins MC1a, MC1b and MC1c from the archaeobacterium *Methanotherix soehngeni*.
CHARTIER, F., LAINE, B., BELAÏCHE, D. and SAUTIERE, P. (1989)
J. Biol. Chem. soumis à publication.

Communications par affiches

- Characterization of the chromosomal protein HMB isolated from the archaeobacterium *Methanosarcina barkeri*.
CHARTIER, F., IMBERT, M., LAINE, B. and SAUTIERE, P.
Proceeding of the EMBO workshop on Molecular Genetics of Archaeobacteria and the international workshop on Biology and Biochemistry of Archaeobacteria, Munich, Juin 1985 (KANDLER, O. and ZILLIG, W., eds) pp 403, Gustav Fisher Verlag. Stuttgart, New-York.
- Structural characteristics of the *Methanosarcina barkeri* protein HMB.
IMBERT, M., CHARTIER, F., LAINE, B. and SAUTIERE, P.
13th International Congress of Biochemistry 25-30 Août 1986. Amsterdam, The Netherlands.
- Characteristics of a chromosomal protein isolated from different species of Methanosarcinaceae : implication in phylogeny.
CHARTIER, F., IMBERT, M., LAINE, B., SAUTIERE, P. and TOUZEL, J.P.
Proceeding of the first international seminar on Biology of Anaerobic Bacteria, Villeneuve d'Ascq, 17-18 Juin 1986. (H.C. DUBOURGUIER *et al.*, eds) pp 242-244, Elsevier Science Publisher.
- La protéine chromosomale de l'archaebactérie *Methanosarcina* sp CHTI 55. Propriétés structurales et effet sur la stabilité thermique du DNA.
CHARTIER, F., LAINE, B. et SAUTIERE, P.
XIV Forum des Jeunes Chercheurs, Lyon, 1-15 Septembre 1987.
- Separation of the variants of chromosomal proteins from prokaryotes using ion-exchange and reversed-phase HPLC.
CHARTIER, F., CREVEL, G., LAINE, B. and SAUTIERE, P.
Proceeding of the 2nd Forum on peptides held in Nancy 2-6 May 1988 (A.AUBRY *et al.*, eds) pp 73-76. Les Editions INSERM.

TABLE DES MATIÈRES

	Pages
INTRODUCTION	1
<u>CHAPITRE I</u>	
GÉNÉRALITÉS	3
I - 1 INTRODUCTION	4
I - 2 LES ARCHAEABACTERIES METHANOGENES	6
I - 2-1 Position phylogénétique des bactéries méthanogènes	6
I - 2-2 Caractéristiques générales des bactéries méthanogènes	9
I - 2-3 Taxonomie des bactéries méthanogènes	9
I - 3 CONDENSATION DE L'ADN	11
I - 3-1 Condensation de l'ADN dans la cellule eucaryote	11
I - 3-1-1 Le nucléosome ou la sous-unité désoxyribonucléoprotéique de base	11
I - 3-1-2 Les histones	12
I - 3-2-2 Les protéines chromosomales non histones	13
I - 3-2 Condensation de l'ADN chez les eubactéries	15
I - 3-2-1 Propriétés de la protéine H-NS	17
I - 3-2-2 Propriétés de la "DNA-binding protein II"	18
I - 3-3 Condensation de l'ADN chez les archaebac- téries	21
I - 3-3-1 La protéine chromosomale HTa de <i>Thermoplasma acidophilum</i>	22
I - 3-3-2 Les protéines chromosomales de <i>Sulfolobus acidocaldarius</i>	23
I - 3-3-3 Les protéines chromosomales des bactéries méthanogènes	24
BIBLIOGRAPHIE	27

CHAPITRE II

ISOLEMENT ET CARACTÉRISATION DE LA PROTÉINE CHROMOSOMALE MC1 DE DIFFÉRENTES SOUCHES DE BACTÉRIES MÉTHANOGENÈS	35
II - 1 PRESENTATION DES SOUCHES ETUDIÉES	36
II - 2 ISOLEMENT DE LA PROTEINE MC1	38
II - 2-1 Mise en évidence de l'association de la protéine MC1 avec l'ADN	38
II - 2-1-1 Préparation du complexe désoxy- ribonucléoprotéique sur gradient de métrizamide	40
II - 2-1-2 Préparation du complexe désoxy- ribonucléoprotéique par combinai- son de l'ultracentrifugation et de la chromatographie de tamisage moléculaire	40
II - 2-2 Préparation de la protéine MC1 par chromato- graphie d'affinité	41
<i>PUBLICATION N°1</i>	44
<i>PUBLICATION N°2</i>	49
<i>PUBLICATION N°3</i>	57
II - 3 CARACTERISTIQUES DE LA PROTEINE MC1 ISOLEE DE DIFFE- RENTES SOUCHES MESOPHILES ET THERMOPHILES DE METHANO- SARCINACEAE	64
II - 3-1 Comportement de la protéine MC1 sur colonne d'ADN-cellulose	64
II - 3-2 Mobilités électrophorétiques	67
II - 3-3 Compositions en acides aminés	67
II - 3-4 Cartes peptidiques de la protéine MC1 de différentes souches du genre <i>Methanosarcina</i>	71
II - 4 CONCLUSION DES ETUDES DE CARACTERISATION DE LA PRO- TEINE MC1	73

CHAPITRE III

STRUCTURE PRIMAIRE DES PROTÉINES MC1	74
III - 1 STRATEGIES	75
III - 2 RESULTATS ET DISCUSSION	77
III - 2-1 Contribution personnelle	77
<i>PUBLICATION N°4</i>	78
<i>PUBLICATION N°5</i>	86
<i>PUBLICATION N°6</i>	112
III - 2-2 Caractéristiques de la structure pri- maire de la protéine MC1	143
III - 2-3 Etude de la variabilité structurale de la protéine MC1	143
III - 2-4 Etude comparative de la protéine MC1 et d'autres protéines se liant à l'ADN	149
<i>PUBLICATION N°7</i>	150
III - 3 CONCLUSION	156

CHAPITRE IV

ETUDE DES INTERACTIONS DE LA PROTÉINE MC1 AVEC L'ADN	157
IV - 1 EFFET DE LA PROTEINE MC1 SUR LA STABILITE THERMIQUE DE L'ADN	158
IV - 2 EFFET DE LA PROTEINE MC1 SUR LA TRANSCRIPTION DE L'ADN	160
IV - 3 DETERMINATION DU DOMAINE DE LA PROTEINE MC1 QUI INTERAGIT AVEC L'ADN	162
IV - 3-1 Principe de la méthode	162
IV - 3-2 Résultats des expériences de pontage	164
IV - 3-3 Etude du complexe ADN-protéine MC1 pontée	167
CONCLUSIONS ET PERSPECTIVES	173
BIBLIOGRAPHIE	176

I N T R O D U C T I O N

La découverte de la division des procaryotes en deux lignées distinctes, les eubactéries et les archaebactéries a suscité un développement important des recherches sur la structure des macromolécules des archaebactéries. L'ensemble de ces travaux a permis de montrer que les archaebactéries occupent une position phylogénétique intermédiaire entre les eubactéries et les eucaryotes. Lors de mon arrivée au laboratoire en 1983, seule la protéine chromosomale de l'archaebactérie *Thermoplasma acidophilum* avait été caractérisée (SEARCY, 1975) et l'état des connaissances sur les protéines chromosomales d'autres archaebactéries se limitait à la description de leur migration électrophorétique sur gel de polyacrylamide (THOMM et coll, 1982). Au sein de l'Unité de Recherche Associée 409 CNRS, le groupe spécialisé sur l'isolement et la caractérisation des protéines chromosomales des eubactéries a élargi son thème de recherche aux protéines des archaebactéries méthanogènes en collaboration avec la Station de Technologie Alimentaire de l'INRA à Villeneuve d'Ascq. Cette étude a pour objectif de définir les caractéristiques structurales qui permettent à ces protéines d'organiser l'ADN et de les comparer à celles des protéines chromosomales des eubactéries et des eucaryotes.

Nos recherches ont porté sur les points suivants :

- 1) Isolement et caractérisation de la protéine chromosomale de *Methanosarcina barkeri* souche MS.
- 2) Participation à la détermination de la structure primaire de cette protéine.
- 3) Détermination de la structure primaire des protéines MC1 de *Methanosarcina* sp CHTI 55 et de *Methanotherix soehngenii* (trois variants a, b et c).
- 4) Etude de l'effet de la protéine MC1 de *Methanosarcina* sp CHTI 55 sur la stabilité thermique et la transcription de l'ADN.
- 5) Participation à la détermination de la région de la protéine MC1 qui interagit avec l'ADN.

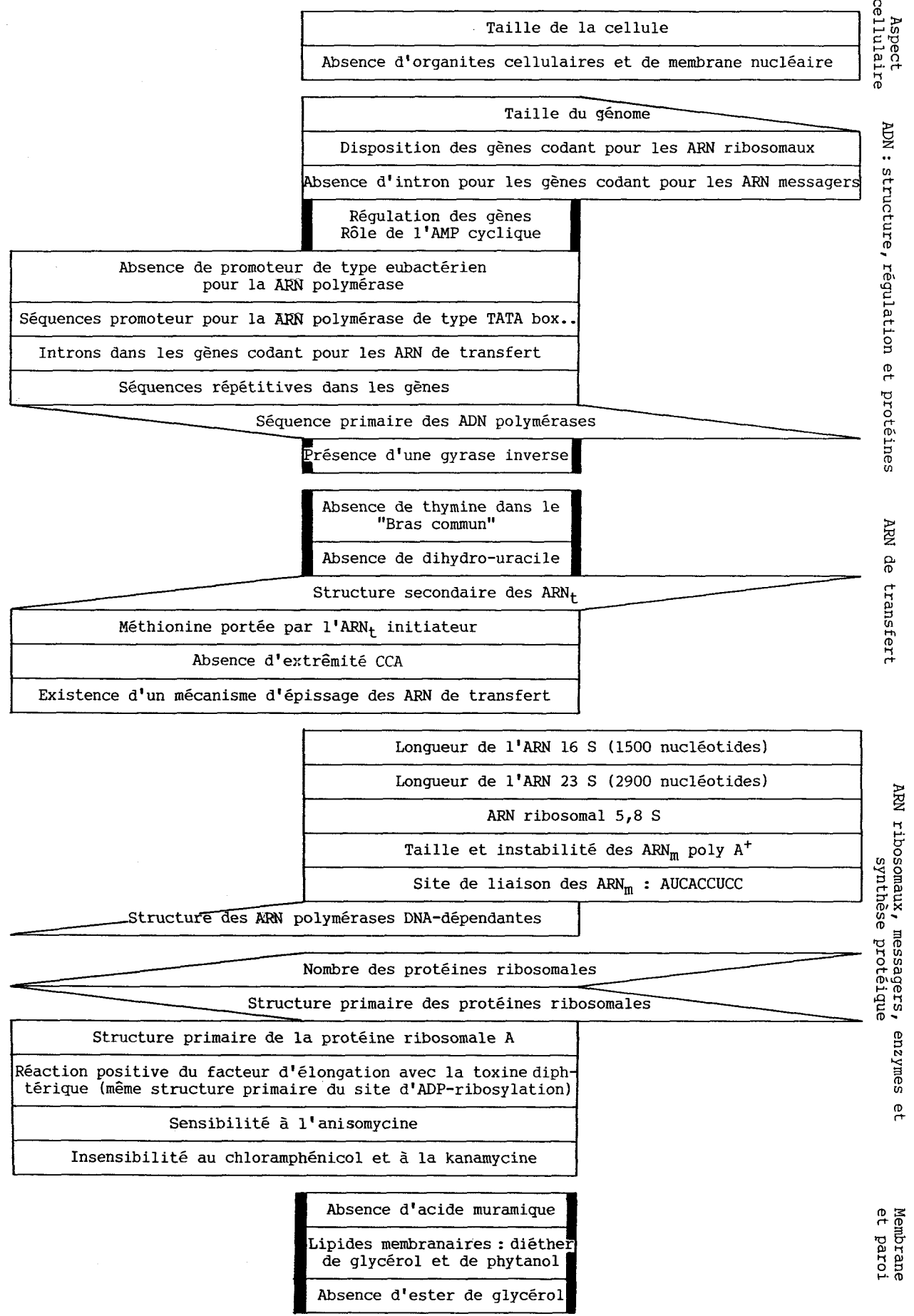
CHAPITRE I

G É N É R A L I T É S

EUCARYOTES

ARCHAEBACTÉRIES

EUBACTÉRIES



Après une brève présentation des bactéries méthanogènes nous préciserons les caractéristiques des protéines chromosomales dans les trois lignées : encaryotes, eubactéries et archaeobactéries.

I - 2 LES ARCHAEBACTÉRIES MÉTHANOGÈNES

En 1776, Alessandro VOLTA a découvert que de "l'air combustible" se dégagait des marécages et des cours d'eau riches en végétaux en décomposition. Il s'écoula presque deux siècles avant que l'on ne découvre en 1956 que cet "air combustible" était du méthane synthétisé par les bactéries méthanogènes. Ces bactéries interviennent à la fin du processus de dégradation des matières organiques. Les premières étapes de ce processus sont assurées par une population bactérienne complexe qui conduit à la formation d'ammoniaque, d'acétates et d'un mélange gazeux H_2/CO_2 . Les bactéries méthanogènes éliminent du milieu ces composés dont l'accumulation est toxique pour les autres bactéries. Des traces fossiles des bactéries méthanogènes, notamment des lipides membranaires, peuvent être mis en évidence dans tous les pétroles et charbons (OURISSON et coll., 1984). Cela montre bien que les bactéries méthanogènes qui ont un rôle écologique essentiel sont présentes depuis longtemps sur l'ensemble du globe.

I - 2-1 POSITION PHYLOGENETIQUE DES BACTERIES METHANOGENES

La phylogénie des archaeobactéries a largement été établie à partir de l'étude des ARN ribosomaux (WOESE et FOX, 1977). L'arbre phylogénétique proposé par WOESE et FOX (1977) a été vivement contesté par HORI et coll. (1982) ainsi que par

LAKE (1985, 1988) qui n'utilisent pas les mêmes critères pour déterminer la phylogénie des espèces. Il semble évident que l'utilisation d'un seul critère est insuffisante pour établir un arbre phylogénétique définitif et que seule l'acquisition de données multiples sur un nombre important de macromolécules permettra de préciser la phylogénie des archaebactéries (LEINFELDER et coll., 1985 ; HUNT et coll., 1985). Les arbres phylogénétiques établis à partir du séquençage complet des ARN 16 S (WOESE et OLSEN, 1986) ou par des techniques d'hybridation ARN ribosomiaux/ADN (KLENK et coll., 1986) sont concordants. Les arbres phylogénétiques les plus récents tiennent compte de données supplémentaires telles que les séquences des ARN polymérases ADN-dépendantes. Par souci de clarté, nous présentons dans la figure 2 l'arbre phylogénétique le plus récent proposé par ZILLIG et coll. (1988).

Le règne des archaebactéries comporte deux embranchements majeurs. Le premier regroupe trois ordres : les bactéries méthanogènes, les bactéries halophiles extrêmes et le genre isolé : *Thermoplasma*. Le second regroupe les bactéries thermophiles extrêmes c'est-à-dire les Thermoprotéales et les Sulfolobales. Les Thermococcales qui regroupent les bactéries thermophiles extrêmes vivant en milieu halophile modéré se placent en position intermédiaire entre les deux embranchements majeurs.

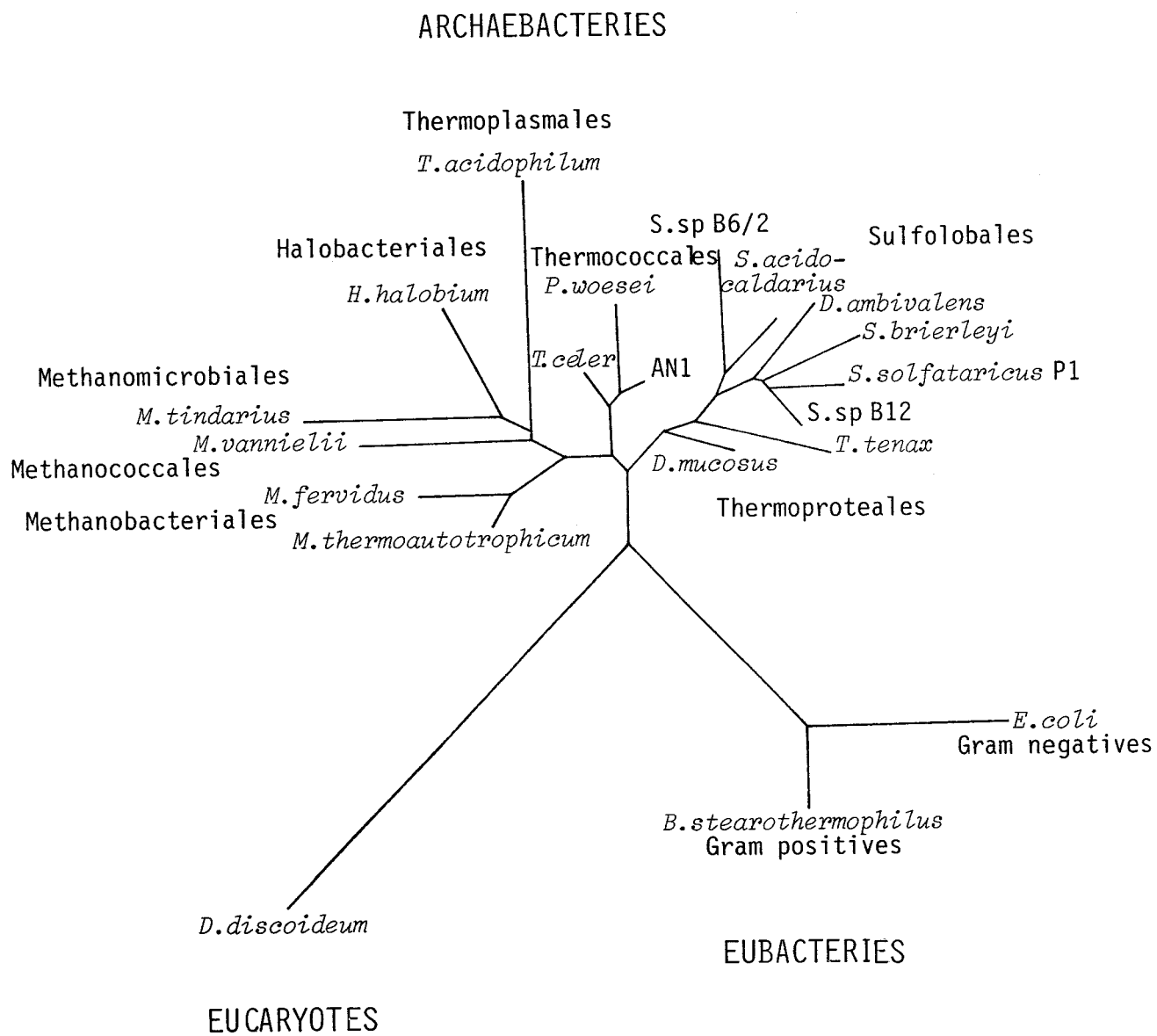


FIGURE 2 - Arbre phylogénétique selon ZILLIG et coll. (1988)

I - 2-2 CARACTERISTIQUES GENERALES DES BACTERIES METHANOGENES

Les bactéries méthanogènes présentent toutes les morphologies que l'on trouve chez les eubactéries (cocci, filaments, bâtonnets, sarcines...) mais contrairement aux eubactéries, elles ne se développent que sur un nombre limité de substrats : le mélange hydrogène/gaz carbonique, le formate, le méthanol, les méthylamines et l'acétate. Les bactéries méthanogènes sont anaérobies strictes : on les trouve donc uniquement dans les milieux à faible potentiel d'oxydoréduction. Elles possèdent des coenzymes spécifiques tels le coenzyme M ou les facteurs F₄₂₀ et F₄₃₀ qui jouent un rôle clef dans le mécanisme de biosynthèse du méthane. Notons que chez les bactéries méthanogènes l'assimilation du CO₂ ne se fait pas par le cycle de CALVIN.

I - 2-3 TAXONOMIE DES BACTERIES METHANOGENES

La taxonomie des bactéries méthanogènes reste basée sur les travaux de BALCH et coll. (1979) qui, à partir d'études partielles de séquences des ARN 16 S, distinguent trois ordres principaux : les Méthanobactériales, les Méthanococcales et les Méthanomicrobiales. L'ordre des Méthanomicrobiales comprend deux familles : les Methanomicrobiaceae et les Methanosarcinaceae (Tableau I). La famille des Methanosarcinaceae regroupe le genre *Methanosarcina* et le genre *Methanotherix*. Toutes les souches du genre Methanosarcinaceae se caractérisent par une croissance sur acétate, méthanol ou méthylamines, seule l'espèce *Methanosarcina barkeri* est capable d'utiliser le mélange H₂/CO₂.

TABLEAU I

Classification des bactéries méthanogènes selon BALCH et coll.1979,
revue par STACKEBRANDT et coll. 1982.

Ordre	Famille	Genres	Espèces		
Méthanobactériales	Methanobacteriaceae	<i>Methanobacterium</i>	<i>M. formicicum</i>		
			<i>M. bryantii</i>		
			<i>M. thermoautotrophicum</i>		
		<i>Methanobrevibacter</i>	<i>M. ruminantium</i>		
			<i>M. arboriphilus</i>		
			<i>M. smithii</i>		
Méthanococcales	Methanococcaceae	<i>Methanococcus</i>	<i>M. vannielli</i>		
			<i>M. voltae</i>		
Méthanomicrobiales	Methanomicrobiaceae	<i>Methanomicrobium</i>	<i>M. mobile</i>		
			<i>M. caryaci</i>		
			<i>M. maris nigri</i>		
		<i>Methanospirillum</i>	<i>M. hungatei</i>		
			Methanosarcinaceae	<i>Methanosarcina</i>	<i>M. barkeri</i>
				<i>Methanotherix</i>	<i>M. soehngenii</i>

Le genre *Methanosarcina* se caractérise par la formation d'agrégats irréguliers. Il n'y a que chez *Methanosarcina mazei* que l'on trouve un stade en forme de cocci au cours du cycle de croissance (ROBINSON, 1986). Le contenu en guanine et cytosine (G + C) varie au sein de ce genre entre 38,8 et 51 %. Le genre *Methanotherix* appartient également à la famille des Methanosarcinaceae (STACKEBRANT et coll., 1982) (Tableau I), bien qu'il soit morphologiquement très différent du genre *Methanosarcina*.

I - 3 CONDENSATION DE L'ADN

I - 3-1 CONDENSATION DE L'ADN DANS LA CELLULE EUCARYOTE

Chez les eucaryotes, la plus grande partie de l'ADN est contenue dans le noyau et se trouve impliquée dans une structure complexe appelée chromatine. Celle-ci est constituée d'ADN, d'ARN et de protéines. L'ADN est associé à une grande variété de protéines parmi lesquelles on distingue les histones et les protéines chromosomales non-histones. Les histones jouent un rôle prépondérant dans l'organisation du motif de base de la compaction de l'ADN : le nucléosome, dont la structure globulaire donne un aspect en collier de perles à la chromatine observée en microscopie électronique.

I - 3-1-1 Le nucléosome ou la sous-unité désoxyribonucléoprotéique de base

La digestion douce de la chromatine par une endonucléase extraite de staphylocoque a permis de mettre en évidence une structure répétitive au sein de la chromatine. Les fragments

d'ADN de cette structure contiennent 200 paires de bases (pb) ou des multiples de cette longueur (HEWISH et BURGOYNE, 1973).

Le nucléosome obtenu à partir de la chromatine de la plupart des cellules eucaryotes est un complexe désoxyribonucléoprotéique contenant 195 ± 5 pb d'ADN enroulées autour d'un octamère d'histones formé de 2 molécules de chacune des histones H2A, H2B, H3 et H4. Une molécule d'histone H1 en position externe par rapport à cette structure, verrouille environ 166 pb d'ADN autour de l'octamère (KORNBERG, 1974, 1977). NOLL et KORNBERG (1977) parviennent à définir deux structures communes à toutes les chromatines quelle que soit leur origine :

- un nucléosome minimum qui renferme 168 pb, l'octamère d'histones et une molécule d'histone H1, structure que SIMPSON (1978) appelle "chromatosome".
- la "core particle" qui consiste en l'octamère d'histones autour duquel s'enroulent 146 pb d'ADN (1,75 tour) (KORNBERG, 1974 ; PRUNELL et coll., 1979)

I - 3-1-2 Les histones

Le rapport pondéral histone/ADN est voisin de 1. Les histones sont des protéines de faible taille moléculaire (M_r 11.000 à 22.000), riches en acides aminés basiques (lysine et arginine). Elles sont pauvres en acides aminés aromatiques ou soufrés et ne contiennent pas de tryptophane. D'une façon générale, la structure primaire des histones est caractérisée par une asymétrie de répartition des charges positives dans la chaîne polypeptidique. La majorité (70 %) des résidus basiques des histones de la "core particle" (H2A, H2B, H3, H4)

est située dans le quart amino-terminal de ces protéines. Ces régions de structure flexible sont stabilisées lorsqu'elles interagissent avec les groupes phosphate de l'ADN. Les régions centrales de ces molécules, particulièrement conservées au cours de l'évolution, sont organisées en structures globulaires et sont le siège des interactions protéine-protéine. Les histones H1 et H5 sont composées d'un domaine globulaire de nature hydrophobe en position centrale et de deux domaines de structure inorganisée de caractère très basique en positions amino- et carboxy-terminales.

Les histones sont sujettes à des modifications post-synthétiques : acétylation, phosphorylation, méthylation, ADP-ribosylation et ubiquitination (MARTINAGE, 1984 ; MATTHEWS et WATERBORG, 1985) en relation avec le cycle cellulaire. Les caractéristiques structurales des histones sont résumées dans la figure 3.

I - 3-1-3 Les protéines chromosomales non-histones

Les protéines non-histones constituent un ensemble très hétérogène de protéines possédant des fonctions différentes. Sur la base d'un critère de mobilité électrophorétique on distingue deux groupes parmi ces protéines : les protéines LMG (Low Mobility Group) et les protéines HMG (High Mobility Group). Les protéines HMG sont associées aux régions transcriptionnellement actives de la chromatine et leur affinité pour l'ADN est inférieure à celle des histones. Elles sont caractérisées par un taux élevé de résidus basiques (≥ 25 %) et acides (20 à 30 %), une concentration forte en proline (≥ 7 %) et un faible taux d'acides aminés aromatiques. Les protéines HMG sont réparties en deux sous-groupes : l'un

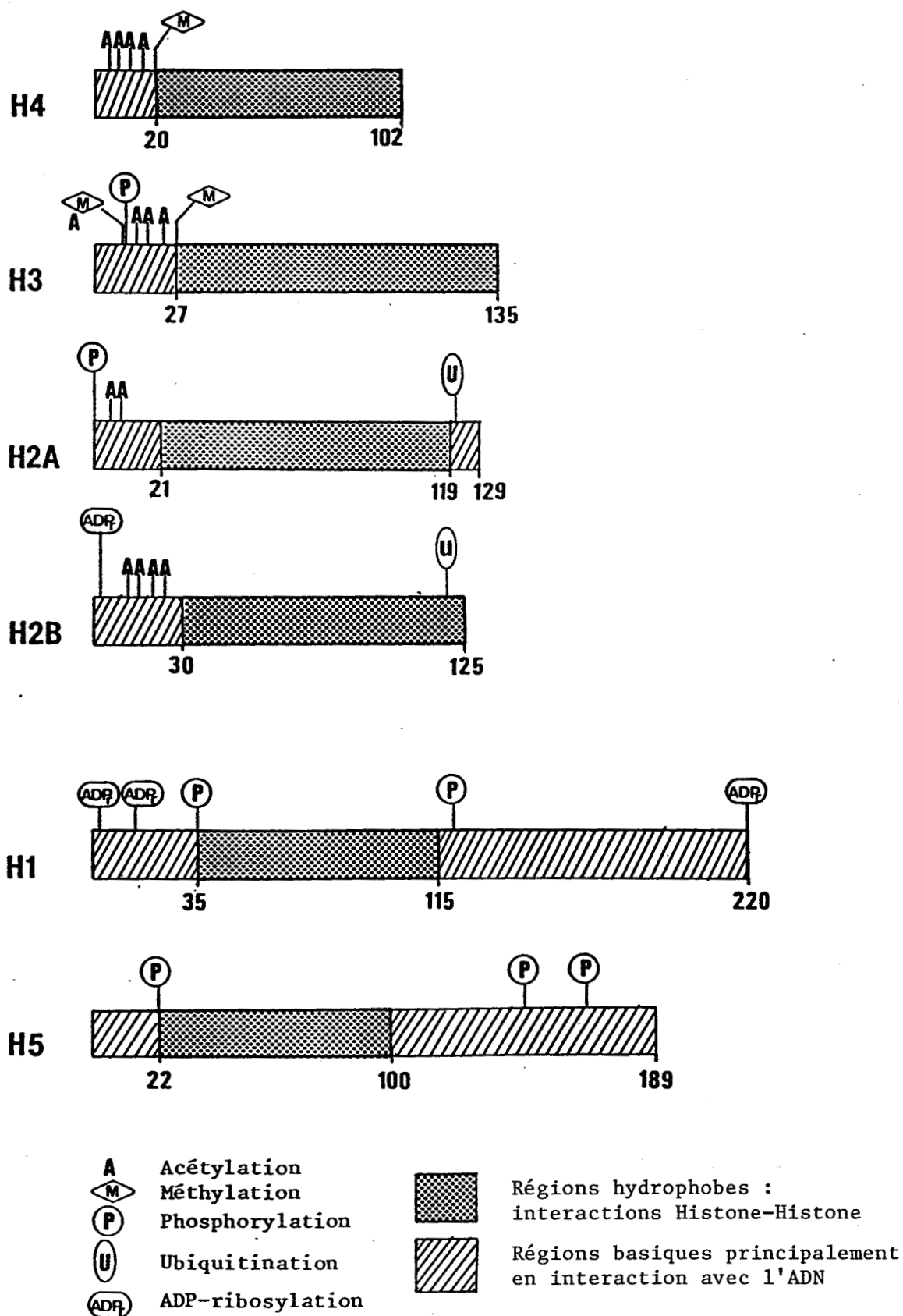


FIGURE 3 - Caractéristiques structurales des histones selon D. KMIECIK (1985).

renferme les protéines de masse moléculaire élevée (HMG 1 et 2, $M_r \approx 28.000$) et l'autre des protéines de faible masse moléculaire (HMG 14 et 17, $M_r \approx 11.000$; HMG 20 ou ubiquitine, $M_r \approx 9.000$). Les protéines HMG 1 et 2 sont structurées sur une large gamme de pH et de force ionique tandis que les protéines HMG 14 et 17 restent déstructurées. Les protéines HMG 1 et 2 se caractérisent par une région C-terminale presque exclusivement constituée par des résidus d'acides aminés acides. Par exemple, sur les 70 résidus C-terminaux de la protéine HMG 1 de thymus de veau, il y a 45 résidus d'acide aspartique et d'acide glutamique (WALKER et coll., 1980). Comme les histones, les protéines HMG sont soumises à des modifications post-synthétiques : acétylation, ADP-ribosylation, méthylation et phosphorylation.

Au terme de ce bref rappel concernant les protéines chromosomales des eucaryotes une place particulière doit être faite aux protéines basiques intermédiaires et aux protamines. Ces protéines apparaissent au cours de la spermiogenèse et viennent en remplacement des histones. Dans le spermatozoïde la compaction de la chromatine atteint un degré très élevé et l'activité génomique est nulle. Les protamines qui sont responsables de cette compaction intense se caractérisent par leur petite taille moléculaire et leur petit nombre d'acides aminés constitutifs. Ces protéines sont extrêmement riches en acides aminés basiques (jusqu'à 80 % d'arginine) et pauvres en acides aminés de caractère aromatique ou hydrophobe.

I - 3-2 CONDENSATION DE L'ADN CHEZ LES EUBACTERIES

La cellule procaryotique ne contient pas de noyau délimité par une membrane nucléaire. Cependant, l'observation en microscopie électronique a montré que l'ADN chromosomique

n'est pas dispersé dans toute la cellule et se trouve localisé dans une région restreinte appelée "corps nucléaire" ou "nucléoïde" (KELLENBERGER et coll., 1958). Dans la cellule bactérienne, l'ADN est associé à des cations bivalents abondants comme le magnésium, à des polyamines comme la spermine et à des protéines (BJORNSTI et coll., 1986). Ces protéines sont des partenaires nécessaires à sa compaction et à sa régulation (PETTIJOHN, 1982).

L'ADN est également associé à la membrane bactérienne, soit directement, soit par l'intermédiaire d'un mésosome. L'association de l'ADN à la membrane, et la grande instabilité des complexes ADN-protéine rendent très délicate la préparation des nucléoïdes (PETTIJOHN, 1982). La digestion douce de la chromatine par les nucléases endogènes a permis à VARSHAVSKY et coll. (1977) d'obtenir de longs fragments de complexes désoxyribonucléoprotéiques débarrassés des protéines membranaires. Les composants protéiques majeurs de ces complexes sont l'ARN polymérase ainsi que deux protéines B1 (M_r 17.000) et B2 (M_r 10.000). Les protéines B1 et B2 correspondent respectivement aux protéines H1 (SPASSKY et BUC, 1977) et HU (ROUVIERE-YANIV et GROS, 1975) qui ont été isolées directement à partir de broyats bactériens. La protéine HU est encore appelée "DNA-binding protein II" et la protéine H1, protéine H-NS ce qui évite la confusion avec l'histone H1. Avec le développement des études sur les protéines chromosomales des eubactéries, la nomenclature HU est apparue inadaptée, nous utiliserons donc dans ce mémoire le terme "DNA binding protein II". A côté de ces protéines, YAMAZAKI et coll. (1984) ont mis en évidence cinq autres protéines acido-solubles associées à l'ADN. Seules les protéines H-NS et les protéines II qui sont abondantes, ont fait l'objet d'études approfondies.

I - 3-2-1 Propriétés de la protéine H-NS

La protéine H-NS est constituée de trois variants : a, b et c qui se distinguent par la valeur de leur pH iso-électrique (pH_i). Le nombre de copies des variants b et c ne varie pas au cours du cycle de croissance bactérienne tandis que le nombre de copies du variant a augmente fortement lors de la phase stationnaire (SPASSKY et coll., 1984 ; RIMSKY et SPASSKY, 1986).

** Propriétés structurales de la protéine H-NS*

FALCONI et coll. (1988) ont séquencé l'un des variants de la protéine H-NS sans préciser lequel des 3 variants cités ci-dessus. Ce variant est un polypeptide de 136 acides aminés qui se caractérise par sa richesse en acide glutamique, en alanine et en leucine. Cette protéine contient trois résidus de tyrosine et un résidu de tryptophane. La charge nette de la protéine est nulle mais la détermination de sa séquence primaire a permis de mettre en évidence une zone à caractère très hydrophile et riche en résidus basiques. Selon ces auteurs, cette séquence (résidus 82 à 114) qui contient aussi l'unique résidu de tryptophane présent dans la molécule pourrait correspondre à la zone d'interaction de la protéine avec l'ADN. La protéine H-NS ne présente pas d'homologie significative de séquence avec d'autres protéines.

** Interactions de la protéine H-NS avec l'ADN*

La protéine H-NS dont la forme fonctionnelle dans le complexe désoxyribonucléoprotéique serait un dimère, présente une affinité importante pour l'ADN bicaténaire (FRIEDRICH et coll., 1988). La protéine H-NS stabilise l'ADN contre la dénaturation thermique et inhibe fortement l'initiation de

la transcription de l'ADN *in vitro* (GUALERZI et coll., 1986). La protéine H-NS est aussi capable de replier et de compacter fortement l'ADN bicaténaire (SPASSKY et coll., 1984).

I - 3-2-2 Propriétés de la "DNA-binding protein II"

La "DNA-binding protein II" a été identifiée dans un grand nombre d'espèces d'eubactéries, toutefois sa fonction n'est pas encore clairement définie.

* *Propriétés structurales de la "DNA-binding protein II"*

La "DNA-binding protein II" est un polypeptide légèrement basique de 90 résidus qui ne contient ni cystéine, ni tryptophane. De manière générale, la "DNA-binding protein II" est également dépourvue de tyrosine. Dans plusieurs espèces eubactériennes, cette protéine est un complexe hétérologue constitué de deux chaînes polypeptidiques de structure primaire étroitement apparentée (LAINE et coll., 1980, 1983 ; KHANAKA et coll., 1985).

La structure tridimensionnelle de la "DNA-binding protein II" de *Bacillus stearothermophilus* a été déterminée avec une résolution de 3 Å (Fig.4) par diffraction des rayons X. Le dimère de la "DNA-binding protein II" correspond certainement à son état fonctionnel lors de sa liaison avec l'ADN. Dans le dimère les segments de structure organisée (hélice α et feuillet β) forment une structure globulaire d'où émergent les deux bras externes, qui délimitent un réceptacle où peut se loger la double hélice d'ADN. La comparaison des structures primaires de "DNA-binding protein II" d'origines diverses a montré que les segments 44-50 et 60-65 permettant la formation de coudes β entre les feuillets β_1 et β_2 et à

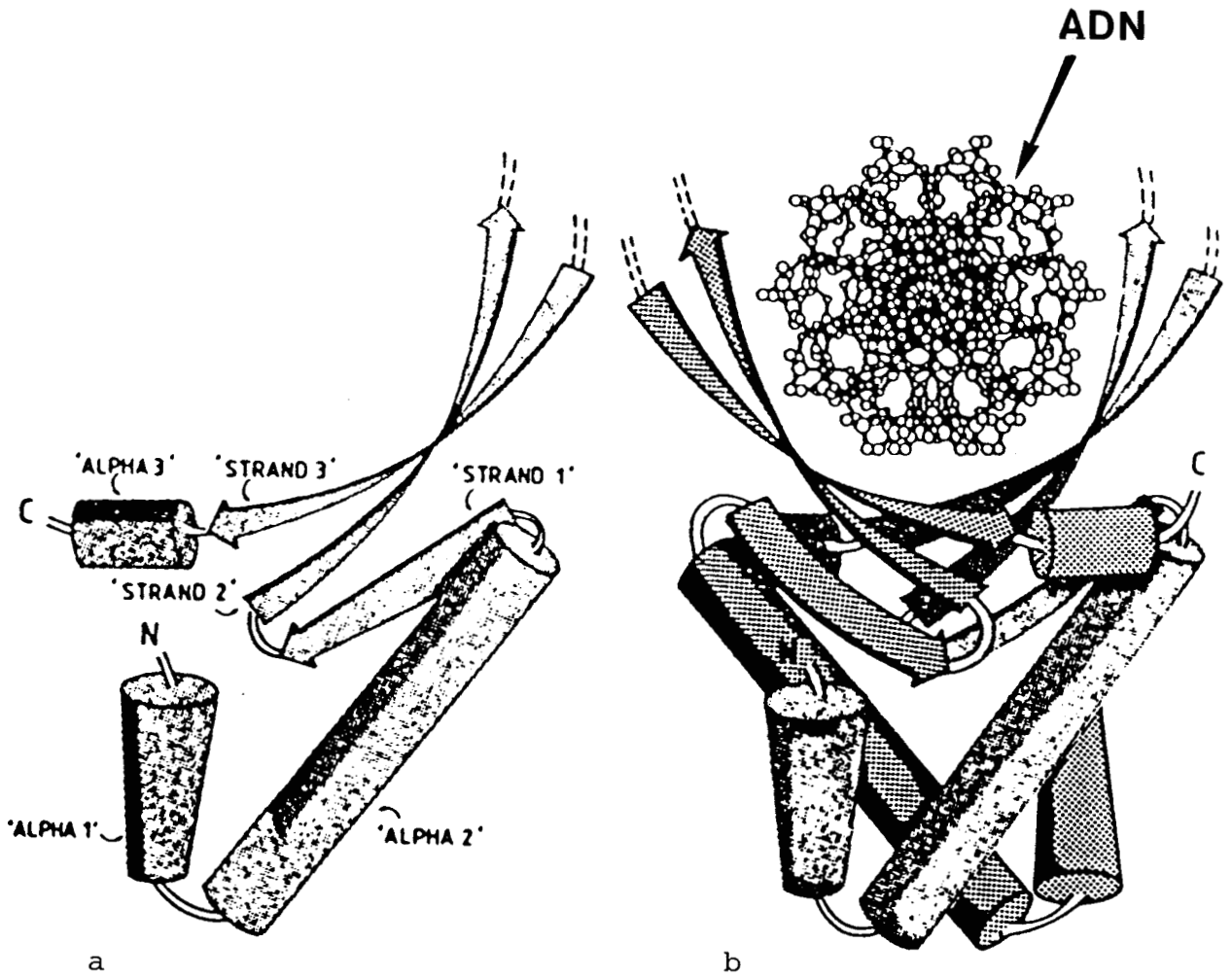


FIGURE 4 a - Structure du monomère de la "DNA-binding protein II"
 b - Schéma montrant l'interaction du dimère de la "DNA-binding protein II" avec l'ADN.
 (TANAKA et coll., 1984).

l'extrémité du bras externe sont particulièrement bien conservés. En outre trois résidus d'arginine (positions 55, 58 et 61) placés sur le bras externe sont bien conservés et forment le site d'interaction de la protéine avec l'ADN par des liaisons électrostatiques (LAMMI et coll., 1984).

* *Interaction de la "DNA-binding protein II" avec l'ADN*

La "DNA-binding protein II" stabilise l'ADN contre la dénaturation thermique (ROUVIERE-YANIV et coll., 1979 ; MIANO et coll., 1982). L'augmentation de la température de fusion de l'ADN est proportionnelle à la quantité de "DNA-binding protein II". IMBER et coll. (1987) ont montré que la longueur minimale d'ADN susceptible d'être associée au dimère de "DNA binding protein II" correspond à 8 paires de bases. Lorsque l'ADN est associé à la "DNA-binding protein II" le pas de l'hélice est réduit de 10,5 à 8,5 paires de bases (BROYLES et PETTIJOHN, 1986) : l'ADN se trouve donc fortement compacté en des structures superhéliçoïdales. En présence de topoisomérase I, la "DNA-binding protein II" d'*E.coli* introduit 18 supertours négatifs dans l'ADN circulaire relâché du phage SV40. La "DNA-binding protein II" forme avec cet ADN des structures globulaires de 18 nm de diamètre et dans ce complexe le degré de condensation de l'ADN est similaire à celui observé dans des complexes reconstitués avec de l'ADN de phage SV40 et les histones (ROUVIERE-YANIV et coll., 1979). *In vivo* l'ADN d'*E.coli* est divisé en une centaine de "domaines" à l'intérieur desquels il est maintenu dans un état de torsion négative (SINDEN et PETTIJOHN, 1981). Les "domaines" sont indépendants c'est-à-dire que le relâchement de la superhélicité provoqué par la coupure d'un brin d'ADN à l'intérieur d'un "domaine", ne se propage pas dans les "domaines" adjacents.

Selon certains auteurs, la "DNA-binding protein II" jouerait un rôle dans la régulation de la transcription ou la répllication de l'ADN. Ainsi DIXON et KORNBERG (1984) ont montré

que la "DNA-binding protein II" stimule la réplication de l'ADN double brin de phages recombinants ou de plasmides portant l'origine de réplication d'*E.coli*. Par ailleurs, LOSSO et coll. (1982) ont observé une inhibition de la transcription de l'ADN de source eucaryote ou procaryote en présence de "DNA-binding protein II".

Enfin à l'aide d'anticorps anti"DNA-binding protein II" et des techniques de microscopie électronique, BJORNSTI et coll. (1986) ont localisé la "DNA-binding protein II" dans les régions du nucléoïde contenant des ribosomes. Ces observations confirment celles de SURYANARAYANA et SUBRAMANIAN (1978) qui ont trouvé la "DNA-binding protein II" associée à la sous-unité 30 S du ribosome d'*E.coli*. L'ensemble de ces résultats laisse supposer que la "DNA-binding protein II" jouerait un rôle dans la transcription. En fait, le rôle de cette protéine sur la transcription et la réplication de l'ADN résulte probablement de son effet sur la conformation de l'ADN (DIXON et KORNBERG, 1984).

I - 3-3 CONDENSATION DE L'ADN CHEZ LES ARCHAEABACTERIES

Lors d'une étude électrophorétique et immunochimique des protéines chromosomales chez différentes espèces d'eubactéries et d'archaeobactéries, THOMM et coll. (1982) ont souligné l'hétérogénéité des protéines chromosomales chez les archaeobactéries. Des travaux plus complets ont été réalisés sur les protéines chromosomales de *Thermoplasma acidophilum*, de *Sulfolobus acidocaldarius* et au sein de notre laboratoire sur les protéines chromosomales des Méthanosarcinaceae. A notre connaissance, aucune publication concernant les protéines chromosomales des archaeobactéries halophiles, n'est parue à ce jour.

I - 3-3-1 La protéine chromosomale HTa de *Thermoplasma acidophilum*.

La protéine HTa (M_r 9.934) représente le composant protéique acido-soluble majeur du complexe désoxyribonucléoprotéique de *Thermoplasma acidophilum* (SEARCY, 1975). Cette protéine est capable de protéger l'ADN contre la dénaturation thermique et de favoriser sa renaturation (STEIN et SEARCY, 1978). La digestion par la nucléase staphylococcale de complexes protéine HTa/ADN natifs ou reconstitués conduit à l'obtention de fragments de 40 paires de bases (SEARCY et STEIN, 1980) : environ 20 % de l'ADN serait protégé. Ces fragments d'ADN sont associés à 4 molécules de protéine HTa. Ces mêmes auteurs ont pu observer en microscopie électronique des structures globulaires de 5 à 6 nm de diamètre qui consisteraient chacune en une boucle d'ADN enroulée autour d'un tétramère de protéine HTa. La stabilité du tétramère protéique serait assurée par des liaisons hydrophobes dans lesquelles les résidus de phénylalanine (5 par molécule de protéine) seraient impliqués (SEARCY, 1986). Cette organisation en sous-unités répétitives rappelle celles qui ont été mises en évidence dans la chromatine des eucaryotes. Une étude de la fluorescence de l'unique résidu de tyrosine présent dans la protéine HTa associée à l'étude des spectres de dichroïsme circulaire et d'absorption UV a conduit SEARCY et ses collaborateurs (1988) à suggérer que le résidu de tyrosine était situé dans le site d'interaction avec l'ADN. Par ailleurs, se fondant sur la richesse en résidus basiques de la protéine HTa (22 % des résidus) et la présence d'un triplet et de quatre doublets basiques le long de la chaîne polypeptidique, SEARCY et coll. n'hésitent pas à assimiler la protéine HTa à une protéine "histone-like". Cependant la détermination de la structure primaire de la protéine HTa (DELANGÉ et coll., 1981a,b) a clairement montré qu'elle est

étroitement apparentée à la "DNA-binding protein II" des eubactéries et non aux histones. En effet, les séquences hautement conservées de la "DNA-binding protein II" (séquences 44-50 et 60-65) qui jouent un rôle important dans la fonction de cette protéine, sont présentes dans la protéine HTa (TANAKA et coll., 1984).

I - 3-3-2 Les protéines chromosomales de *Sulfolobus acidocaldarius*.

Chez l'archaebactérie thermophile-dépendante du soufre *Sulfolobus acidocaldarius*, GREEN et coll. (1983) ont caractérisé deux protéines chromosomales majeures : une protéine acido-soluble H_{Sa} (M_r 14.500) et une protéine non acido-soluble N_HSa (M_r 36.000). La protéine H_{Sa} se caractérise par sa richesse en leucine (14,7 %) et en acides aminés basiques (15,7 %), aucune information n'est donnée sur la composition en acides aminés de la protéine N_HSa. Ultérieurement, trois autres classes de protéines chromosomales de faibles tailles moléculaires (M_r 7.000, 8.000 et 10.000) ont été isolées de *Sulfolobus acidocaldarius* (DIJK et REINHARDT, 1986). Chacune de ces classes correspondant en fait, à un mélange d'au moins deux protéines. Ainsi la classe des protéines de 7 kDa est constituée de quatre variants dont les séquences en acides aminés sont étroitement apparentées (KIMURA et coll., 1984 ; CHOLI et coll., 1988). Ces protéines se caractérisent par la présence de lysine monométhylée et par leur richesse en acides aminés basiques (de 20 à 22 % de lysine et de 3 à 6 % d'arginine), chacune de ces protéines contient deux résidus de tyrosine et un résidu de tryptophane. Ces protéines ne présentent aucune homologie de séquence avec d'autres protéines chromosomales ou non.

Seulement 5 % de l'ADN des complexes désoxyribonucléo-protéiques natifs de *Sulfolobus acidocaldarius* sont protégés contre la digestion par la nucléase staphylococcale. Cette digestion aboutit à l'obtention de fragments de tailles hétérogènes dont la longueur moyenne correspond à un oligonucléotide de 25 paires de bases (GREEN et coll., 1983). L'étude en microscopie électronique de complexes formés avec les différentes classes de protéines et de l'ADN simple ou double brin ne montre pas de structure de type globulaire, ni de diminution significative de la longueur de l'ADN (LURTZ et coll., 1986). Ces observations suggèrent que la fonction de ces protéines ne serait pas de compacter l'ADN comme le font les histones. REDDY et SURYANARAYANA (1988) ont isolé du nucléoïde de *Sulfolobus acidocaldarius* quatre protéines : A, B, C et C', de faibles tailles moléculaires (M_r 12.000, 11.500, 10.500 et 9.000 respectivement). Ces protéines ont un taux de lysine élevé (19 à 24 %) et trois d'entre elles sont capables de stabiliser l'ADN contre la dénaturation thermique. Malheureusement, ces auteurs n'ont donné aucune indication sur les relations éventuelles entre les protéines qu'ils ont isolées et celles décrites précédemment.

I - 3-3-3 Les protéines chromosomales des bactéries méthanogènes .

THOMM et coll. (1982) ont identifié une protéine non acido-soluble de masse moléculaire apparente 14.000 associée à l'ADN de *Methanobacterium thermoautotrophicum* . En 1984, nous avons isolé une protéine acido-soluble de M_r 14.400 qui est le composant protéique majeur du complexe désoxyribonucléo-protéique de *Methanosarcina barkeri* (CHARTIER et coll., 1985). En l'absence d'une nomenclature précise concernant les protéines chromosomales des procaryotes, nous avons alors

adopté la nomenclature H en usage pour désigner les protéines chromosomales chez les eubactéries ; la lettre H signifie "Histone-like" et les deux lettres suivantes correspondent aux initiales du genre et de l'espèce bactérienne. Dans cette nomenclature la protéine chromosomale de *Methanosarcina barkeri* a été appelée protéine HMB. Avec le développement de nos recherches sur les protéines chromosomales de bactéries méthanogènes, cette nomenclature est apparue rapidement inadaptée. En outre, le terme "histone-like" s'est avéré totalement impropre. Nous avons donc proposé le terme de protéine MCl qui signifie "Methanogen chromosomal protein 1" pour désigner le type de protéine chromosomale que nous avons identifiée dans *Methanosarcina barkeri*.

Cette protéine a été caractérisée au laboratoire par sa composition en acides aminés, par sa structure primaire et sa structure secondaire et par l'étude du complexe désoxyribonucléoprotéique de *Methanosarcina barkeri*. Seuls ces deux derniers points qui constituent l'essentiel de la thèse de Marlène IMBERT (1987) seront résumés dans ce chapitre.

La structure secondaire de la protéine MCl de *Methanosarcina barkeri* a été étudiée au moyen de méthodes statistiques de prévision de structure ainsi que par dichroïsme circulaire et spectroscopie infrarouge. La protéine MCl est caractérisée par la quasi-absence de structure α hélicoïdale mais elle renferme de courts fragments en feuillets β antiparallèles. Cette dernière structure disparaît lorsque la force ionique augmente. De plus des études de spectroscopie de fluorescence ont montré que le tryptophane en position 61 est au coeur d'une région structurée enfouie dans la structure tertiaire de la protéine. Il faut noter que la quasi-absence de structure α hélicoïdale et le dépliement de la chaîne polypeptidique provoqué par une augmentation de la force ionique sont deux caractéristiques qui différencient fortement la protéine MCl de *M.barkeri*, de la "DNA-binding protein II" d'*E.coli* et des histones.

La protéine MC1 a été localisée *in situ* par observation en microscopie électronique, de cryosections de *Methanosarcina barkeri* en utilisant un double marquage immunochimique de l'ADN et de la protéine MC1 (IMBERT et coll., 1988). Ainsi ces auteurs ont montré que la protéine MC1 est localisée dans les régions riches en ADN. Ces études confirment que la protéine MC1 est bien une protéine chromosomale. La protéine MC1 est présente sur l'ADN avec un rapport pondéral protéine/ADN égal à 0,1 (soit le même rapport que celui décrit pour la "DNA-binding protein II" mais dix fois plus petit que le rapport histone/ADN chez les eucaryotes). Dans le complexe désoxyribonucléoprotéique isolé, la protéine MC1 ne forme pas avec l'ADN de structures globulaires répétitives et contrairement aux histones et à la protéine HTa de *Thermoplasma acidophilum*, la protéine MC1 semble incapable de protéger l'ADN contre la digestion par la nucléase staphylococcale. IMBERT et coll. (résultats non publiés) ont aussi déterminé la structure primaire de la protéine MC1 isolée de *Methanosarcina* sp MST-A1. La séquence de cette protéine est très proche de celle de *Methanosarcina barkeri* MS puisqu'il n'y a que dix changements dont sept sont conservatifs.

BIBLIOGRAPHIE

- BALCH, W.E., FOX, G.E., MAGRUM, L.J., WOESE, C.R. and WOLFE, R.S. (1979)
Methanogens : reevaluation of a unique biological group.
Microbiol. Rev. 43, 260-296.
- BJORNSTI, M.A., HOBOT, J.A., KELUS, A.S., VILLIGER, W. and KELLENBERGER, E. (1986)
New electron microscopic data on the structure of the nucleoid and their functional consequences.
in Bacterial chromatin (GUALERZI, C.O. and PON, C.L. eds) pp 64-81, Springer-Verlag.
- BROYLES, S.S. and PETTIJOHN, D.E. (1986)
Interaction of the *Escherichia coli* HU protein with DNA. Evidence for formation of nucleosome-like structures with altered DNA helical pitch.
J. Mol. Biol. 187, 47-60.
- CHARTIER, F., LAINE, B., SAUTIERE, P., TOUZEL, J.P. and ALBAGNAC, G. (1985)
Characterization of the chromosomal protein Hmb isolated from *Methanosarcina barkeri*.
FEBS Lett. 183, 119-123.
- CHOLI, T., WITTMANN-LIEBOLD, B. and REINHARDT, R. (1988)
Microsequence analysis of DNA-binding proteins 7a, 7b and 7c from the archaebacterium *Sulfolobus acidocaldarius*.
J. Biol. Chem. 263, 7087-7093.
- DANIELS, C.J., GUPTA, R. and DOOLITTLE, W.F. (1985)
Transcription and excision of a large intron in the tRNA (trp) gene of an archaebacterium *Halobacterium volcanii*.
J. Biol. Chem. 260, 3132-3134.
- DELANGE, R.J., GREEN, G.R. and SEARCY, D.G. (1981)
A histone-like protein (HTa) from *Thermoplasma acidophilum*. I Purification and properties.
J. Biol. Chem. 256, 900-904.
- 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.
- DIJK, J. and REINHARDT, R. (1986)
The structure of DNA-binding proteins from Eu- and Archaeobacteria.
in Bacterial chromatin (GUALERZI, C.O. and PON, C.L. eds) pp 186-218, Springer-Verlag.

- DIXON, N.E. and KORNBERG, A. (1984)
Protein HU in the enzymatic replication of the chromosomal origin of *Escherichia coli*.
Proc. Natl. Acad. Sci. USA 81, 424-428.
- FALCONI, M., GUALTIERI, M.T., LA TEANA, A., LOSSO, M.A. and PON, C.L. (1988)
Proteins from the prokaryotic nucleoid : primary and quaternary structure of the 15-kD *Escherichia coli* DNA binding protein H-NS.
Molecular Microbiol. 2, 323-329.
- FOX, G.E., STACKEBRANDT, E., HESPELL, R.B., GIBSON, J., MANILOFF, J., DYER, T.A., WOLFE, R.S., BALCH, W.E., TANNER, R.S., MAGRUM, L.J., ZABLEN, L.B., BLAKEMORE, R., GUPTA, R., BONEM, L., LEWIS, B.J., STAHL, D.A., LUEHRSEN, K.R., CHEN, K.N. and WOESE, C.R. (1980)
The phylogeny of Procaryotes.
Science 209, 457-463.
- FRIEDRICH, K., GUALERZI, C.O., LAMMI, M., LOSSO, M.A. and PON, C.L. (1988)
Proteins from the prokaryotic nucleoid. Interaction of nucleic acids with the 15 kDa *Escherichia coli* histone like protein H-NS.
FEBS Lett. 229, 197-202.
- GREEN, G.R., SEARCY, D.G. and DELANGE, R.G. (1983)
Histone-like protein in the Archaeobacterium *Sulfolobus acidocaldarius*.
Biochim. Biophys. Acta 741, 251-257.
- GROPP, F., REITER, W.D., SENTENAC, A., ZILLIG, W., SCHNABEL, R., THOMM, M. and STETTER, K.O. (1986)
Homologies of components of DNA-dependant RNA polymerases of archaeobacteria, eukaryotes and eubacteria.
System. Appl. Microbiol. 7, 95-101.
- GUALERZI, C.O., LOSSO, M.A., LAMMI, M., FRIEDRICH, K., PAWLIK, R.T., CANONACO, M.A., GIANFRANCESCHI, G., PINGOUD, A. and PON, C.L. (1986)
Proteins from the prokaryotic nucleoid. Structural and functional characterization of the *Escherichia coli* DNA-binding proteins NS (HU) and H-NS.
In Bacterial chromatin. (GUALERZI, C.O. and PON, C.L. eds) Heidelberg, Springer-Verlag, pp 101-134.
- GUPTA, R. (1984)
Halobacterium volcanii tRNAs ; identification of 41 tRNAs covering all amino acids, and the sequences of 33 class I tRNAs.
J. Biol. Chem. 259, 9461-9471.

- HEWISH, D.R. and BURGOYNE, L.A. (1973)
Chromatin sub-structure. The digestion of chromatin DNA at regular spaced sites by a nuclear deoxyribonuclease. Biochem. Biophys. Res. Commun. 52, 504-510.
- HORI, H., ITOH, T. and OSAWA, S. (1982)
The phylogenic structure of the metabacteria. Zbl. Bakt. Hyg. I, Abt. Orig. C3, 18-30
- HUNT, L.T., GEORGE, D.G. and BARKER, W.C. (1985)
The prokaryote-eukaryote interface. Biosystems 18, 223-240.
- IMBER, R., KIMURA, M., GROCH, N. and HEINEMANN, U. (1987)
DNA-binding properties and primary structure of HB protein from *Bacillus globigii*. Eur. J. Biochem. 165, 547-552.
- IMBERT, M. (1987)
La protéine chromosomale MCl isolée d'archaebactéries du genre *Methanosarcina*. Etudes structurales, localisation ultra-structurale et étude du complexe désoxyribonucléoprotéique. Thèse de Doctorat d'Université, Lille.
- IMBERT, M., LAINE, B., PRENSIER, G., TOUZEL, J.P. and SAUTIERE, P. (1988)
The deoxyribonucleoprotein complex from the archaebacterium *Methanosarcina barkeri* : characterization and ultrastructural localization of the chromosomal protein MCl. Canadian J. Microbiol. 34, 931-937.
- KAINE, P., GUPTA, R. and WOESE, C.R. (1983)
Putative introns in tRNA genes of prokaryotes. Proc. Natl. Acad. Sci. USA 80, 3309-3312.
- KANDLER, O. and KONIG, H. (1985)
Cell envelopes of Archaeobacteria. In The bacteria, a treatise on structure and function (WOESE, C.R. and WOLFE, R.S. eds) Academic Press, New York, Vol. VIII, 413-452.
- KELLENBERGER, E., RYTER, A. and SECHAUD, J. (1958)
Electron microscopic study of DNA-containing plasms. II - Vegetative and phage DNA as compared with normal bacterial nucleoids in different physiological states. J. Biophys. Biochem. Cytol. 4, 671-678.
- KHANAKA, H., LAINE, B., SAUTIERE, P. and GUILLAUME, J. (1985)
Characterization and primary structures of DNA-binding HU-type proteins from *Rhizobiaceae*. Eur. J. Biochem. 147, 343-349.

- KIMURA, M., KIMURA, J., DAVIE, P., REINHARDT, R. and DIJK, J. (1984)
The amino acid sequence of a small DNA binding protein from the archaebacterium *Sulfolobus solfataricus*.
FEBS Lett. 176, 176-178.
- KLENK, H-P., HAAS, B., SCHWASS, V. and ZILLIG, W. (1986)
Hybridization homology : a new parameter for the analysis of phylogenetic relations, demonstrated with the urkingdom of the archaebacteria.
J. Mol. Evol. 24, 167-173.
- KLIMCZAK, L.J., GRUMMT, F., BURGER, K.J. (1986)
Purification and characterization of DNA polymerase from the archaebacterium *Sulfolobus acidocaldarius*.
Nucleic Acids Research 13, 5269-5282.
- KMIECIK, D. (1985)
Etudes structurales d'histones d'invertébrés marins. Histone H2A de l'érythrocyte de Siponcle. Histone H1 spécifique du sperme de *Platynereis dumerilii*.
Thèse de Doctorat d'Etat, Lille.
- KORNBERG, R.D. (1974)
Chromatin structure : a repeating unit of histone and DNA chromatin structure is based on a repeating unit of eight histone molecule and about 200 DNA base pairs.
Science 184, 868-871.
- KORNBERG, R.D. (1977)
Structure of chromatin.
Ann. Rev. Biochem. 46, 931-954.
- LAINE, B., KMIECIK, D., SAUTIERE, P., BISERTE, G. and COHEN-SOLAL, M. (1980)
Complete amino acid sequence of DNA binding proteins HU-1 and HU-2 from *Escherichia coli*.
Eur. J. Biochem. 103, 447-461.
- LAINE, B., BELAÏCHE, D., KHANAKA, H. and SAUTIERE, P. (1983)
Primary structure of the DNA-binding protein HRm from *Rhizobium meliloti*.
Eur. J. Biochem. 131, 325-331.
- LAKE, J.A. (1985)
Evolving ribosome structure : domains in archaebacteria, eubacteria, eocytes and eukaryotes.
Ann. Rev. Biochem. 54, 507-530.
- LAKE, J.A. (1988)
Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences.
Nature 331, 184-186.

- LAMMI, M., PACI, M. and GUALERZI, C.O. (1984)
The interaction between protein NS and DNA involves the oligomeric form of the protein and at least one Arg residue.
FEBS Lett. 170, 99-104.
- LANGWORTHY, T.A. (1985)
Lipids of archaebacteria.
In The Bacteria, a treatise on structure and function (WOESE, C.R. and WOLFE, R.S., eds) Academic Press, New York, Vol. VIII, 453-491.
- LEINFELDER, W., JARSCH, M. and BÖCK, A. (1985)
The phylogenetic position of the sulfur-dependent archaebacterium *Thermoproteus tenax*: sequence of the 16S rRNA gene.
System. Appl. Microbiol. 6, 164-170.
- LOSSO, M.A., MIANO, A., GIANFRANCESCHI, G.L. and GUALERZI, C.O. (1982)
Proteins from the prokaryotic nucleoid. II. Inhibition of DNA transcription by NS1 and NS2 (HU proteins).
Biochem. Int. 5, 423-427.
- LURTZ, R., GROTE, M., DIJK, J., REINHARDT, R. and DOBRINSKI, B. (1986)
Electron microscopic study of DNA complexes with proteins from the archaebacterium *Sulfolobus acidocaldarius*.
EMBO J. 5, 3715-3721.
- MARTINAGE, A. (1984)
Histones H2A et H2B de gonades d'étoile de mer. Structure primaire. Phosphorylation des histones libres ou associées au DNA.
Thèse de Doctorat d'Etat, Lille.
- MATHESON, A.T., LOUIE, K.A. and HENDERSON, G.N. (1986)
The evolution of the ribosomal 'A' protein (L12) in archaebacteria.
System. Appl. Microbiol. 7, 147-150.
- MATTHEWS, H.R. and WATERBORG, J.H. (1985)
Reversible modifications of nuclear proteins and their significance.
In The enzymology of post-transcriptional modification of proteins (FREEDMAN, R.B. and HAWKINS, H.C., eds) pp 125-185. Academic Press Inc, London.
- MIANO, A., LOSSO, M.A., GIANFRANCESCHI, G.L. and GUALERZI, C.O. (1982)
Proteins from the prokaryotic nucleoid. I - Effect of NS1 and NS2 (HU) proteins on the thermal stability of DNA.
Biochem. Int. 5, 415-422.

- NOLL, M. and KORNBERG, R.D. (1977)
Action of micrococcal nuclease on chromatin and the location of histone H1.
J. Mol. Biol. 109, 393-404.
- OURISSON, G., ALBRECHT, P. and ROHMER, M. (1984)
L'origine microbienne des combustibles fossiles.
Pour la Science 84, 56-66.
- PETTIJOHN, D.E. (1982)
Structure and properties of the bacterial nucleoid.
Cell 30, 667-669.
- PRUNELL, A., KORNBERG, R.D., LUTTER, L.C., KLUG, A., LEVITT, M. and CRICK, P.H.C. (1979)
Periodicity of deoxyribonuclease I digestion of chromatin.
Science 204, 855-858.
- REDDY, T.R. and SURYANARAYANA, T. (1988)
Novel histone-like DNA-binding proteins in the nucleoid from the acidothermophilic archaeobacterium *Sulfolobus acidocaldarius*.
Biochim. Biophys. Acta 949, 87-96.
- RIMSKY, S. and SPASSKY, A. (1986)
Escherichia coli protein H1_a strongly compacts DNA *in vitro*.
in Bacterial chromatin (GUALERZI, C.O. and PON, C.L., eds)
Heidelberg : Springer-Verlag, pp.167-174.
- ROBINSON, R.W. (1986)
Life cycles in the methanogenic archaeobacterium *Methanosarcina mazei*.
Appl. Environ. Microbiol. 52, 17-27.
- ROUVIERE-YANIV, J. and GROS, F. (1975)
Characterization of a novel, low-molecular-weight DNA-binding protein from *Escherichia coli*.
Proc. Natl. Acad. Sci. USA 72, 3428-3432.
- ROUVIERE-YANIV, J., YANIV, M. and GERMOND, J.E. (1979)
E. coli DNA binding protein HU forms nucleosome-like structure with circular double stranded DNA.
Cell 17, 265-274.
- SEARCY, D.G. (1975)
Histone-like protein in the prokaryote *Thermoplasma acidophilum*.
Biochim. Biophys. Acta 395, 535-547.
- SEARCY, D.G. (1986)
The archaeobacterial histone "HTa".
in Bacterial chromatin (GUALERZI, C.O. and PON, C.L., eds)
Heidelberg , Springer-Verlag, pp 64-81.

- SEARCY, D.G., MONTENAY-GARESTIER, T., LASTON, D.J. and HELENE, C. (1988)
Tyrosine environment and phosphate binding in the archaeobacterial histone-like protein HTa.
Biochim. Biophys. Acta 953, 321-333.
- SEARCY, D.G. and STEIN, D.B. (1980)
Nucleoprotein subunit structure in an unusual prokaryotic organism : *Thermoplasma acidophilum*.
Biochim. Biophys. Acta 609, 180-195.
- SIMPSON, R.T. (1978)
Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones.
Biochemistry 17, 5524-5531.
- SINDEN, R.R. and PETTIJOHN, D.E. (1981)
Chromosomes in living *Escherichia coli* cells are segregated into domains of supercoiling.
Proc. Natl. Acad. Sci. USA 78, 224-228.
- SPASSKY, A. and BUC, H.C. (1977)
Physico-chemical properties of a DNA binding protein : *Escherichia coli* factor H1.
Eur. J. Biochem. 81, 79-80.
- SPASSKY, A., RIMSKY, S., GARREAU, H. and BUC, H. (1984)
Hla, an *E. coli* DNA-binding protein which accumulated in stationary phase, strongly compacts DNA *in vitro*.
Nucleic Acids Research 12, 5321-5340.
- STACKEBRANDT, E., SEEWALD, E., LUDWIG, W., SCHEIFER, K.H. and HUSER, B.A. (1982)
The phylogenic position of *Methanotherix soehngenii*. Elucidated by a modified technique of sequencing oligonucleotides from 16S rRNA.
Zbl. Bakt. Hyg. I, Abt. Orig. C3, 90-100.
- STEIN, D.B. and SEARCY, D.G. (1978)
Physiologically important stabilization of DNA by a prokaryotic histone-like protein.
Science 202, 219-221.
- SURYANARAYANA, T. and SUBRAMANIAN, A.R. (1978)
Specific association of two homologous DNA-binding proteins to the native 30 S ribosomal subunits of *Escherichia coli*.
Biochim. Biophys. Acta 520, 342-347.
- 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.

- THOMM, M., STETTER, K.O. and ZILLIG, W. (1982)
Histone-like proteins in Eu- and Archaeobacteria.
Zbl. Bakt. Hyg. I, Abt. Orig. C3, 128-139.
- VARSHAVSKY, A.J., NEDOSPASOV, S.A., BAKAYEV, V.V., BAKAYEVA, T.G.
and GEORGIEV, G.P. (1977)
Histone-like proteins in the purified *Escherichia coli*
deoxyribonucleoprotein.
Nucleic Acids Research 4, 2725-2745.
- VASHAKIDZE, R.P. and PRANGISHVILI, D.A. (1987)
Simple repetitive sequences in the genomes of archaeobac-
teria.
FEBS Lett. 216, 217-220.
- WALKER, J.M., GOODERHAM, K., HASTINGS, J.R., MAYES, E. and
JOHNS, E.W. (1980)
The primary structures of non-histone chromosomal proteins
HMG 1 and 2.
FEBS Lett. 122, 264-270.
- WILLEKENS, P., HUYSMANS, E., VANDENBERGHE, A. and DE WACHTER, R.
(1986)
Archaeobacterial 5 S ribosomal RNA : nucleotide sequence
in two methanogen species, secondary structure models and
molecular evolution.
System. Appl. Microbiol. 7, 151-159.
- WOESE, C.R. and FOX, G.E. (1977)
Phylogenetic structure of the prokaryotic domain : the
primary kingdoms.
Proc. Natl. Acad. Sci. USA 74, 5088-5090.
- WOESE, C.R. (1981)
Les archaebactéries.
Pour la Science 46, 56-69.
- WOESE, C.R. and OLSEN, G. (1986)
Archaeobacterial phylogeny : perspectives on the urkingdoms.
System. Appl. Microbiol. 7, 161-177.
- YAMAZAKI, K., NAGATA, A., KANO, Y. and IMAMOTO, F. (1984)
Isolation and characterization of nucleoid proteins from
Escherichia coli.
Mol. Gen. Genet. 196, 217-224.
- ZILLIG, W., PALM, P., REITER, W.D., GROPP, F., PÜHLER, G. and
KLENK, H.P. (1988)
Comparative evaluation of gene expression in archaeobacteria.
Eur. J. Biochem. 173, 473-482.

CHAPITRE II

ISOLEMENT ET CARACTÉRISATION DE LA PROTÉINE CHROMOSOMALE

MCI DE DIFFÉRENTES SOUCHES DE BACTÉRIES MÉTHANOGENÈS

II - 1 PRÉSENTATION DES SOUCHES ÉTUDIÉES

Depuis 1984, les recherches se sont développées dans notre laboratoire pour isoler et caractériser la ou les protéines chromosomales des bactéries méthanogènes. Notre groupe a particulièrement étudié la protéine MC1 isolée de cinq souches mésophiles ou thermophiles de bactéries appartenant toutes à la famille des Methanosarcinaceae. Mes travaux ont été consacrés à l'étude de la protéine MC1 isolée de trois souches du genre *Methanosarcina* : *Methanosarcina barkeri* MS, *Methanosarcina* sp CHTI 55 et *Methanosarcina thermophila* TM1 et d'une souche du genre *Methanothrix* : *Methanothrix soehngeni*. La cinquième souche choisie est *Methanosarcina* sp MST-A1 dont la protéine MC1 a été isolée et caractérisée par Marlène IMBERT (1987).

Les principaux caractères morphologiques et physiologiques des bactéries que nous avons étudiées sont rassemblés dans le tableau II. Ces bactéries se différencient principalement par leur température optimale de croissance et par la nature des substrats qu'elles peuvent métaboliser : ainsi, seule *Methanosarcina barkeri* est capable d'utiliser le mélange H₂/CO₂. Les souches de *Methanosarcina* présentent également des aspects en culture différents par la taille et la forme des agrégats qu'elles produisent. Les bactéries du genre *Methanothrix*, bien que reliées à la famille des Methanosarcinaceae, ont une forme de bâtonnets associés en filaments qui se regroupent pour former des amas spécifiques.

TABLEAU II

Caractéristiques générales des souches de bactéries méthanogènes étudiées.

	Température de croissance	Substrats de croissance	(G + C) %	Morphologie	Aspect en culture
<i>Methanosarcina barkeri</i>	Mésophile 35°C	H ₂ /CO ₂ , acétate méthylamines méthanol	39	Sarcines	Agrégats irréguliers fins
<i>Methanosarcina</i> MST-A1	Thermophile 55°C	acétate méthylamines méthanol	39,3	Sarcines	Agrégats plus réguliers
<i>Methanosarcina</i> sp CHTI 55	Thermophile 55°C	acétate méthylamines méthanol	39,3	Sarcines	Gros agrégats irréguliers
<i>Methanosarcina thermophila</i> TM1	Thermophile 50°C	acétate méthylamines méthanol	42	Sarcines	Gros agrégats irréguliers
<i>Methanotherix soehngenii</i>	Mésophile 35°C	acétate	52,6	Filaments	Aspect cotonneux

II - 2 ISOLEMENT DE LA PROTÉINE MC1

II - 2-1 MISE EN EVIDENCE DE L'ASSOCIATION DE LA PROTEINE MC1 AVEC L'ADN

La préparation des protéines associées à l'ADN des procaryotes fait généralement appel à une première étape de chromatographie d'affinité sur colonne d'ADN-cellulose. Cette étape permet d'identifier et de préparer en quantité importante les protéines qui présentent une affinité pour l'ADN, mais l'affinité pour l'ADN n'est pas spécifique aux protéines chromosomales. Il est donc nécessaire d'isoler d'abord un complexe ADN-protéine(s) natif pour identifier sans ambiguïté la ou les protéines chromosomales présentes dans ce complexe. Les protéines de ce complexe sont analysées par électrophorèse sur gel de polyacrylamide dans différents systèmes tampons. Ces protéines sont éventuellement purifiées par chromatographie liquide haute performance (CLHP) en phase inverse et caractérisées par leur composition en acides aminés. Les caractéristiques de ces protéines sont ensuite comparées à celles des protéines préparées par chromatographie d'affinité sur colonne d'ADN-cellulose. L'identité des caractéristiques des protéines obtenues par les deux méthodes permet de nous assurer que les protéines préparées sur la colonne d'ADN-cellulose sont bien les protéines chromosomales recherchées.

Les préparations des complexes désoxyribonucléoprotéiques ont été effectuées selon deux méthodes différentes : dans le cas de *Methanosarcina barkeri*, nous avons utilisé la centrifugation isopycnique (méthode A) et dans le cas de *Methanosarcina* sp CHTI 55, nous avons utilisé, en combinaison, l'ultracentrifugation et la chromatographie de tamisage moléculaire (méthode B). Les principes de ces méthodes sont décrits dans la figure 5.

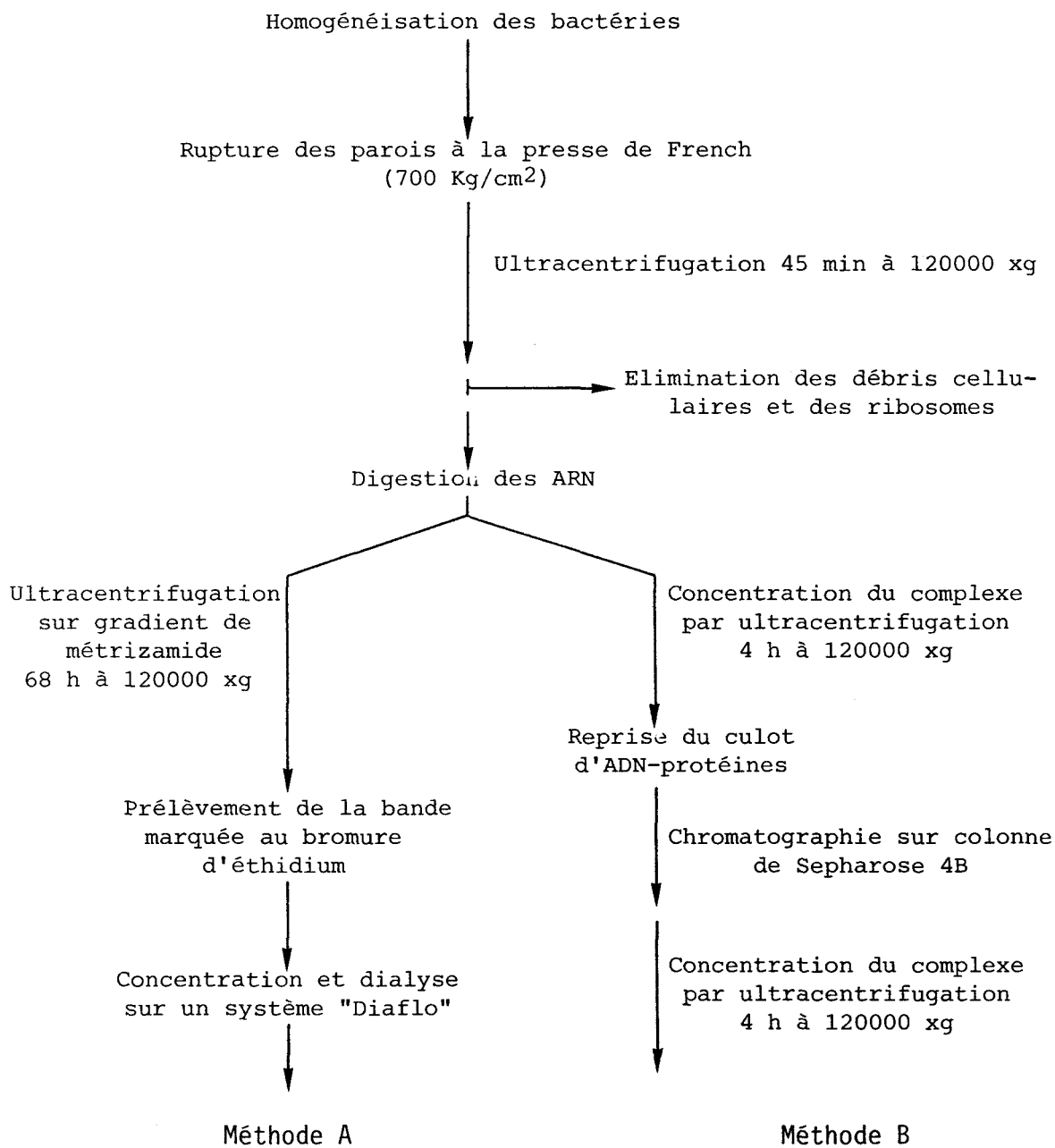


FIGURE 5 - Méthodes de préparation des complexes désoxyribonucléoprotéiques.

Méthode A : par ultracentrifugation isopycnique sur métrizamide.

Méthode B : par combinaison de l'ultracentrifugation et de la chromatographie de tamisage moléculaire.

II - 2-1-1 Préparation du complexe désoxyribonucléoprotéique sur gradient de métrizamide

L'ultracentrifugation sur gradient de métrizamide a été utilisée pour étudier le complexe désoxyribonucléoprotéique de *Methanosarcina barkeri*. Nous avons pu mettre en évidence pour la première fois qu'une protéine basique de $M_r \sim 10.000$, que nous avons appelée protéine MC1, représente le composant protéique majeur de ce complexe. Les résultats obtenus sont présentés dans la publication n°1 :

Characterization of the chromosomal protein HMB isolated from *Methanosarcina barkeri*.
 CHARTIER, F., LAINE, B., SAUTIERE, P., TOUZEL, J.P. and ALBAGNAC, G.
 (1985)
 FEBS Lett. 183, 119-123.

II - 2-1-2 Préparation du complexe désoxyribonucléoprotéique par combinaison de l'ultracentrifugation et de la chromatographie de tamisage moléculaire

La méthode utilisée pour la préparation du complexe désoxyribonucléoprotéique de *Methanosarcina barkeri* s'est avérée inutilisable dans le cas de *Methanosarcina* sp CHTI 55 : de nombreuses macromolécules de densité voisine de celle du complexe désoxyribonucléoprotéique ne sont pas séparées du complexe sur le gradient de métrizamide. Ces molécules ont pu être facilement éliminées par passage du complexe sur colonne de Sepharose 4B, après concentration par ultracentrifugation. Chez *Methanosarcina* sp CHTI 55, l'ADN est associé à deux protéines : la protéine MC1 de $M_r \sim 10.000$ et la protéine MC2 de $M_r \sim 17.000$, qui représentent respectivement 80 et 14% des composants protéiques du complexe désoxyribonucléoprotéique.

Les résultats obtenus sont présentés dans la publication n°2 :

Characterization of the chromosomal protein MCl from the thermophilic archaeobacterium *Methanosarcina* sp CHTI 55 and its effect on the thermal stability of DNA.
CHARTIER,F., LAINE,B. and SAUTIERE,P. (1988)
Biochim. Biophys.Acta 951, 149-156.

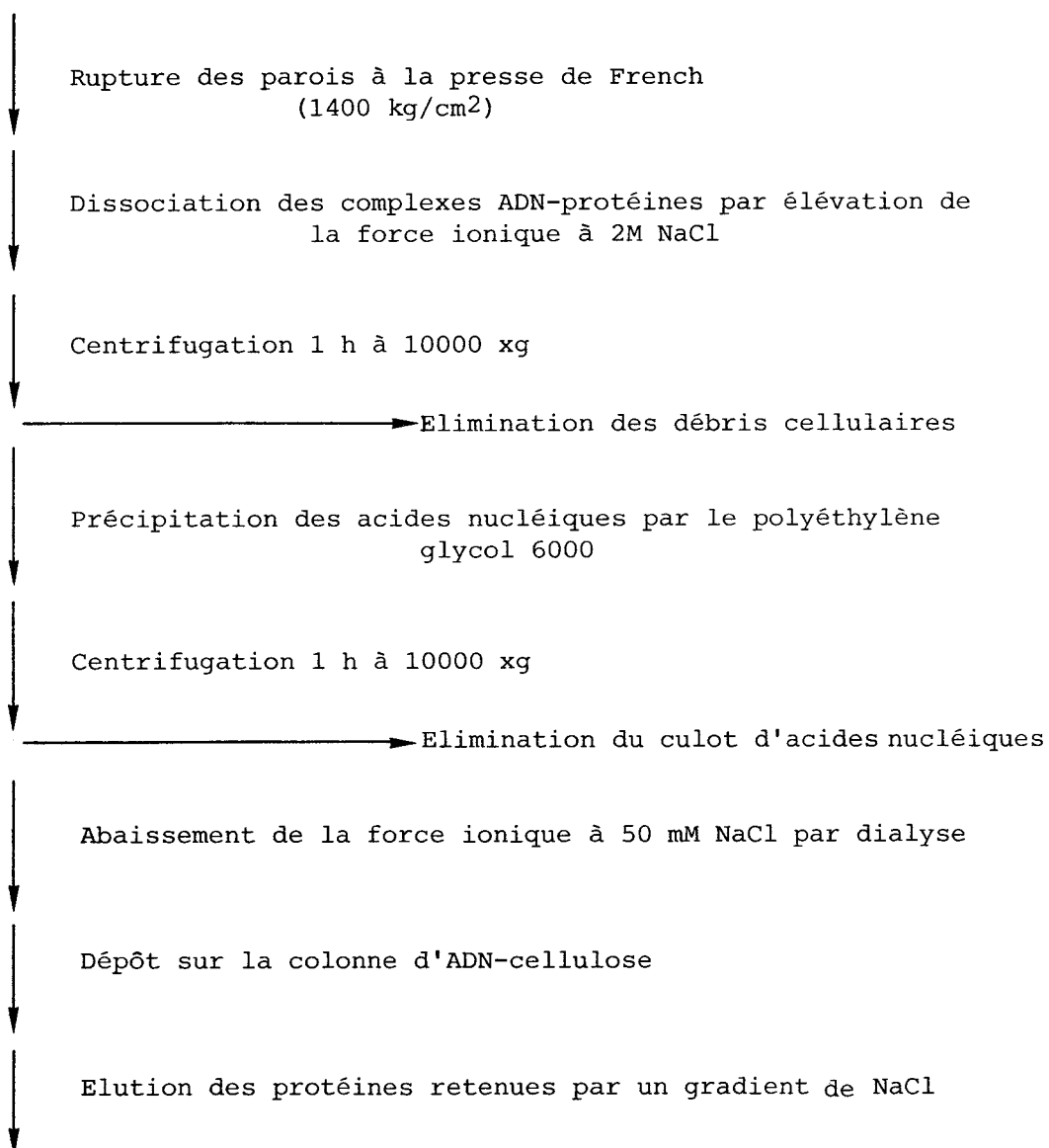
Cette dernière méthode présente l'avantage de ne pas être dénaturante, d'éviter l'étape difficile de l'élimination de la métrizamide et de fournir un complexe qui peut être directement étudié.

II - 2-2 PREPARATION DE LA PROTEINE MCl PAR CHROMATOGRAPHIE D'AFFINITE

La protéine MCl est préparée par chromatographie d'affinité sur colonne d'ADN ponté à de la cellulose selon la méthode décrite par ALBERTS et HERRICK (1971). Les différentes étapes de la préparation sont présentées dans la figure 6. Après enrichissement, les protéines MCl sont purifiées, soit par chromatographie de tamisage moléculaire, soit par chromatographie liquide haute performance en phase inverse.

L'isolement des complexes désoxyribonucléoprotéiques a permis d'identifier la protéine MCl de *Methanosarcina barkeri* et de *Methanosarcina* sp CHTI 55. La préparation de ces complexes est délicate et les rendements en sont faibles. Nous avons donc préparé la protéine MCl par chromatographie d'affinité sur colonne d'ADN-cellulose. Cette dernière méthode a également été utilisée pour préparer la protéine MCl d'autres souches de bactéries méthanogènes afin d'en étudier la variabilité structurale. Cette étude avait aussi pour but de rechercher des différences au niveau moléculaire parmi des bactéries thermophiles proches d'un point de vue phénotypique et génotypique mais présentant des aspects macroscopiques différents (voir tableau II). Les bactéries thermophiles

Homogénéisation des bactéries



Fraction enrichie en protéine MCl

FIGURE 6 - Principe de la préparation de la protéine MCl par chromatographie d'affinité.

qui ont fait l'objet de cette dernière étude sont : *Methanosarcina* sp CHTI 55, *Methanosarcina* MST-A1 et *Methanosarcina thermophila* TM1. La protéine MCl de *Methanosarcina* MST-A1 a été étudiée par Marlène IMBERT (1987). La protéine MCl isolée de *Methanosarcina thermophila* TM1 est en tous points identique à la protéine MCl de *Methanosarcina* sp CHTI 55 (voir sous chapitre II-3, p. 64), son étude structurale n'a donc pas été développée.

Chez *Methanotherix soehngenii*, nous avons pu caractériser trois variants structuraux de la protéine MCl, que nous avons appelés MCl_a, MCl_b et MCl_c. En dépit de leurs tailles moléculaires voisines et de la faible différence de charge des variants b et c, observée en électrophorèse sur gel de polyacrylamide en milieu acide, ces variants ont pu être séparés en une seule étape par CLHP en phase inverse après enrichissement sur colonne d'ADN-cellulose. Leur séparation et caractérisation sont décrites dans la publication n°3 :

High-performance liquid chromatographic separation of variants of chromosomal proteins from prokaryotes.
CHARTIER, F., CREVEL, G., LAINE, B. and SAUTIERE, P. (1989)
J.Chromatogr. 466, 331-337.

Characterization of the chromosomal protein HMb isolated from *Methanosarcina barkeri*

François Chartier, Bernard Laine*, Pierre Sautiere, Jean-Pierre Touzel⁺ and Guy Albagnac⁺

Unité Associée CNRS 409, Place de Verdun, 59045 Lille Cédex and ⁺INRA, Station de Technologie Alimentaire, BP39, 59651 Villeneuve d'Ascq Cédex, France

Received 12 December 1984; revised version received 23 January 1985

The acid soluble protein HMb is the major component of the archaeobacteria *Methanosarcina barkeri* nucleoprotein complex. This protein of M_r 14 500 is rich in basic amino acids. Protein HMb differs from eubacterial HU-type chromosomal proteins by its molecular size and physico-chemical characteristics.

Archaeobacteria *Methanogen* *Chromosomal protein*

1. INTRODUCTION

Prokaryotes comprise two phylogenetically distinct groups: the eubacteria and the archaeobacteria which include thermoacidophiles, halophiles and methanogens [1]. Structure of macromolecules such as 5 S rRNA and elongation factor 2 proteins and properties of DNA-dependent RNA polymerase strongly suggest a closer relationship of archaeobacteria with eukaryotes than with eubacteria [2-4]. In eukaryotic cells, packaging of the DNA inside the nucleus is mainly ensured by basic proteins called histones. In eubacteria, HU-type proteins (M_r ~ 9500) are involved in DNA condensation [5-7]. Another protein called H1 (M_r ~ 15 500) which may also participate in DNA condensation has been identified in *E. coli* [8]. The change in chromosomal protein composition between eukaryotes and eubacteria explains the fundamental difference in their chromatin organization [9].

The organization of the DNA in archaeobacteria has yet to be elucidated. However, chromosomal proteins have been characterized in two strains of

thermoacidophiles: the protein HTa from *Thermoplasma acidophilum* [10,11] and two proteins of M_r 14 500 and 36 000, respectively, from *Sulfolobus acidocaldarius* [12]. The amino acid sequence of the former is clearly homologous to that of HU-type proteins [6,7,13]; the latter are not related to the HU-type proteins [12].

An electrophoretic study has shown the presence of a DNA-associated protein of M_r 14 300 in *Methanobacterium thermoautotrophicum* [14] but, up to now, no chromosomal protein has been isolated from methanogens and halophiles. This paper deals with the characterization of the major chromosomal protein isolated from the methanogen *Methanosarcina barkeri*.

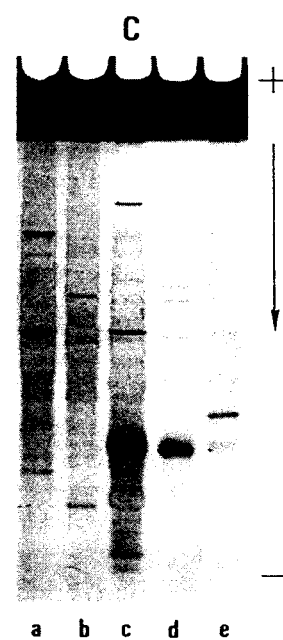
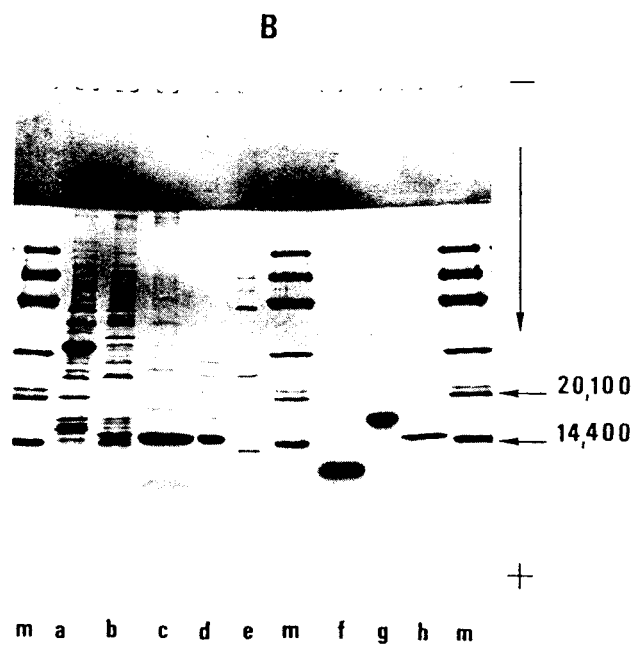
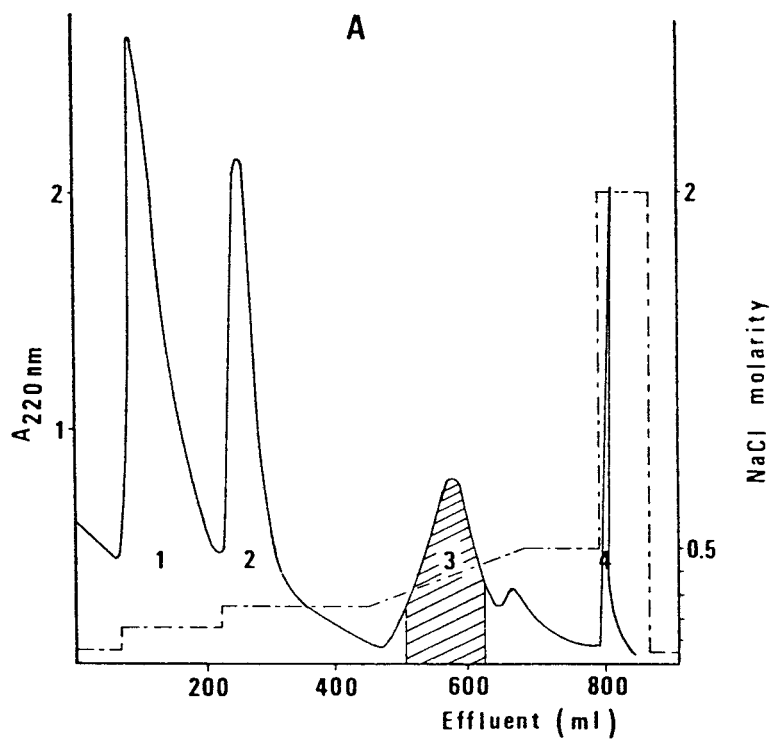
2. MATERIALS AND METHODS

M. barkeri, strain MS, was provided by Dr J.G. Zeikus (University of Wisconsin, Madison, USA) and was grown at 35°C and pH 7 under 80% H₂/20% CO₂ in the basal medium containing carbonate, yeast extract and trypticase (BCYT) described in [15]. Cells were harvested in the post-exponential phase of growth.

The nucleoprotein complex was isolated as in [14] with the following modifications: bacteria (15 g) were suspended in 30 ml of 10 mM Tris-HCl,

Abbreviations: DFP, diisopropylfluorophosphate; PTH, phenylthiohydantoin

* To whom correspondence should be addressed



200 mM NaCl, 1 mM Na₂EDTA at pH 7.4 in the presence of 0.1 mM DFP, and isopycnic centrifugation was performed for 68 h at 120 000 × *g*. The acid-soluble proteins were extracted from the nucleoprotein complex overnight at 4°C with 0.25 M HCl under gentle agitation. Proteins were precipitated by 6 volumes of acetone at -20°C and redissolved in 0.01 M HCl.

Protein HMB was prepared by affinity chromatography on a DNA-cellulose column as described in [16] with the following modifications: bacteria were suspended in 10 mM sodium phosphate, 1 mM Na₂EDTA buffer adjusted to pH 6.8 (buffer A) containing 2 M NaCl, 0.1 mM DFP and 1 mM 2-mercaptoethanol, and the DNA-cellulose made with *Rhizobium meliloti* DNA was eluted with increasing concentrations of NaCl in buffer A as indicated in fig. 1A. Protein HMB was purified by gel filtration chromatography on a Sephadex G100 column (90 × 1.6 cm) equilibrated and eluted with 0.01 M HCl saturated with chloroform.

Analytical gel electrophoreses and amino acid analyses were performed as indicated in [16,17], respectively.

3. RESULTS

A large number of protein species are retained on DNA bound to cellulose, but all of them cannot be considered as chromosomal proteins [12]. Therefore the preparation of the nucleoprotein complex is essential for identifying the proteins associated with the DNA in vivo. The *M. barkeri* nucleoprotein complex was prepared in the

presence of 0.2 M NaCl to avoid non-specific binding of cytoplasmic or membranous proteins [14]. A protein of apparent *M_r* 14 500 is the major component extracted from the nucleoprotein by 0.25 M HCl (fig. 1B, lane c). By analogy with the *T. acidophilum* protein HTa and *S. acidocaldarius* protein HSA, this protein was named HMB.

Since protein HMB was recovered with a low yield from the nucleoprotein, it was prepared simultaneously by affinity chromatography on a DNA-cellulose column. The *M. barkeri* protein fractions eluted from the DNA-cellulose column (fig. 1A) were analysed by polyacrylamide gel electrophoresis in the presence of SDS (fig. 1B) and in 0.9 M acetic acid, 2.5 M urea (fig. 1C). The identity of the major protein eluted in fraction 3 at a molarity of NaCl between 0.32 and 0.40 M with protein HMB was ascertained by the following criteria: first, it is the only protein recovered by affinity chromatography showing an electrophoretic mobility identical to that of protein HMB in two different electrophoretic systems; secondly, both proteins have the same retention times on reverse-phase high-pressure liquid chromatography (not shown) and identical amino acid compositions.

Protein HMB was purified to homogeneity by gel filtration chromatography on Sephadex G100 as assessed by polyacrylamide gel electrophoresis (Fig. 1B, lane h). Moreover, the purity of the protein was confirmed by the identification of only one amino acid derivative (PTH-Ser) at its N-terminus.

The amino acid composition of protein HMB is presented in table 1. Protein HMB is devoid of cys-



Fig. 1. (A) Elution pattern of *M. barkeri* proteins fractionated on a DNA-cellulose column. The column (15 × 2.5 cm) was eluted at a flow rate of 40 ml/h with a gradient of NaCl (---) in buffer A (see section 2). The effluent was monitored at 220 nm (—) and fractions of 6.7 ml were collected. Protein HMB is eluted in the hatched peak (peak 3). (B) SDS-polyacrylamide gel electrophoresis of *M. barkeri* and *E. coli* DNA binding proteins. lanes a, b, d, e correspond respectively to fractions 1, 2, 3, 4 eluted from the DNA-cellulose in (A); c, acid extract of the nucleoprotein complex; f, *E. coli* protein HU; g, *E. coli* protein HI; h, protein HMB; m, *M_r* markers: phosphorylase b (*M_r* 94 000), bovine serum albumin (*M_r* 67 000), ovalbumin (*M_r* 43 000), carbonic anhydrase (*M_r* 30 000) soybean trypsin inhibitor (*M_r* 20 100), α-lactalbumin (*M_r* 14 400). Samples (7 μg) were treated with 5% SDS, 2% 2-mercaptoethanol for 3 min at 100°C and run at 40 mA for 150 min on a 5–30% polyacrylamide gradient gel [22]. Electrode buffer (Tris-glycine, pH 8.3) and gel buffer (Tris-HCl, pH 8.9) were made 0.1% in SDS. The gels were stained with 0.5% Coomassie blue R-250 in acetic acid/ethanol/water (1:2:7, v/v/v) for 15 h. The gels were destained by diffusion in acetic acid/ethanol/water (4:5:31, v/v/v). (C) Polyacrylamide gel electrophoresis of *M. barkeri* DNA binding proteins in 0.9 M acetic acid, 2.5 M urea. Lanes a–e are as in (B). Samples (3 μg) were dissolved in 0.01 M HCl, 8 M urea, 0.5 M 2-mercaptoethanol and were run at 22 mA for 4 h on a gel containing 17% acrylamide. Gel was stained and destained as in (B).

Table 1
Amino acid compositions of protein HMB and of other chromosomal proteins

Amino acids	Archaeobacteria			Eubacteria	Eukaryote	
	HMB		HSa	HU	HCc	
	<i>M. barkeri</i>		<i>S. acidocaldarius</i>	<i>E. coli</i>	<i>Cryptocodinium cohnii</i>	
	(mol%)	(Residues)	[12]	[11]	[16]	[19]
			(mol%)	(mol%)	(mol%)	(mol%)
Aspartic acid	10.3	12.7 (13) ^c	9.1	5.6	8.9	5.5
Threonine	4.3 ^a	5.3 (5)	3.7	5.6	6.8	7.2
Serine	3.4 ^a	4.2 (4)	8.2	7.8	4.5	4.5
Glutamic acid	13.0	16.0 (16)	10.8	13.3	9.5	9.0
Proline	6.4	7.9 (8)	3	3.3	2.6	4.3
Glycine	6.8	8.3 (8)	9.5	6.7	7.2	7.6
Alanine	6.4	7.9 (8)	7.1	7.8	18.5	15.8
Cysteine	0 ^b	0	1.8	0	0	trace
Valine	7.0 ^c	8.6 (9)	4.5	8.9	7.4	7.3
Methionine	2.3	2.9 (3)	1.1	1.1	1.2	0.9
Isoleucine	4.1 ^c	5.1 (5)	4.1	7.8	6.6	3.1
Leucine	3.5 ^c	4.3 (4)	14.7	3.3	6.9	7.5
Tyrosine	0	0	2.8	1.1	0	3.6
Phenylalanine	4.1	5.1 (5)	1.8	5.6	3.3	2.6
Histidine	2.2	2.7 (3)	0.2	0	0.9	2.1
Lysine	15.7	19.3 (19)	10.2	15.6	11.1	17.8
Arginine	8.9	11.0 (11)	5.3	6.7	4.7	3.6
Tryptophan	1.5 ^d	1.9 (2)	nd	0	0	nd
Total	99.9	123				
M_r		14 360	14 500	9934	9500	12 000
% basic amino acids	26.8		15.7	22.3	16.7	23.5

^aValues for threonine and serine were obtained by linear extrapolation to zero hydrolysis time

^bCysteine was determined as cysteic acid after performic acid oxidation [20]

^c72 h hydrolysis values

^dDetermined after hydrolysis in 3 N mercaptoethanesulfonic acid [21]

^eNumbers in parentheses are the nearest integers

nd = not determined

teine and tyrosine, it contains two residues of tryptophan. This protein is characterized by a high content of basic amino acids (~27%) and dicarboxylic amino acids (~23%). With 123 amino acid residues, the calculated M_r of protein HMB is 14 360.

4. DISCUSSION

The acid-soluble protein HMB of M_r 14 360 constitutes the major protein of the nucleoprotein

complex of *M. barkeri* (strain MS). Protein HMB is not homologous to any of the eubacterial or archaeobacterial DNA binding proteins. It differs in size and in amino acid composition from HU-type proteins [11,16] and from the *Sulfolobus solfatarius* DNA binding protein [18]. Chromosomal proteins from *M. thermoautotrophicum* [14] and from *S. acidocaldarius* (protein HSa) [12] have a molecular size similar to that of protein HMB but exhibit different physico-chemical characteristics. Thus, the protein HMB is acid soluble whereas the

M. thermoautotrophicum protein is not. On the other hand, protein H_{Sa} is distinct from protein H_{Mb} particularly with respect to the amounts of lysine, arginine and leucine.

Elucidation of protein H_{Mb} amino acid sequence will allow us to determine if this protein shares some structural homology with HU-type proteins or with histones. Since phylogenetic data indicate that archaebacteria are more closely related to eukaryotes than to eubacteria, it is of interest to investigate whether protein H_{Mb} may represent an ancestral histone. It is also of interest to know if this protein can be related to the basic protein H_{Cc} isolated from a lower eukaryote, the dinoflagellate *Cryptothecodinium cohnii* which lacks histones [19].

Knowledge of the properties of protein H_{Mb} and of archaebacterial nucleoid structure will serve as a model for better understanding the radical difference in chromatin organization between eubacteria and eukaryotes.

ACKNOWLEDGEMENTS

The authors are grateful to M.J. Dupire, T. Ernout and A. Hemez for their skillful technical assistance. The work was supported by a grant from the Centre National de la Recherche Scientifique.

REFERENCES

- [1] Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., Zablen, L.B., Blakemore, R., Gupta, R., Bonem, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N. and Woese, C.R. (1980) *Science* 209, 457-463.
- [2] Hori, H., Itoh, T. and Osawa, S. (1982) *Zbl. Bakt. Hyg. I, Abt. Orig. C3*, 18-30.
- [3] Kessel, M. and Klink, F. (1980) *Nature* 287, 250-251.
- [4] Huet, J., Schnabel, R., Sentenac, A. and Zillig, W. (1983) *EMBO J.* 2, 1291-1294.
- [5] Rouvière-Yaniv, J., Yaniv, M. and Germond, J.E. (1979) *Cell* 17, 265-274.
- [6] Laine, B., Bélaïche, D., Khanaka, H. and Sautière, P. (1983) *Eur. J. Biochem.* 131, 325-331.
- [7] Tanaka, I., Appelt, K., Dijk, J., White, S.W. and Wilson, K.S. (1984) *Nature* 310, 376-381.
- [8] Spassky, A., Rimsky, S., Garreau, H. and Buc, H. (1984) *Nucleic Acids Res.* 12, 5321-5340.
- [9] Pettijohn, D.E. (1982) *Cell* 30, 667-669.
- [10] Searcy, D.G. (1975) *Biochim. Biophys. Acta* 395, 535-547.
- [11] De Lange, R.J., Green, G.R. and Searcy, D.G. (1981) *J. Biol. Chem.* 256, 900-904.
- [12] Green, G.R., Searcy, D.G. and De Lange, R.J. (1983) *Biochim. Biophys. Acta* 741, 251-257.
- [13] De Lange, R.J., Williams, L.C. and Searcy, D.G. (1981) *J. Biol. Chem.* 256, 905-911.
- [14] Thomm, M., Stetter, K.O. and Zillig, W. (1982) *Zbl. Bakt. Hyg. I, Abt. Orig. C3*, 128-139.
- [15] Touzel, J.P. and Albagnac, G. (1983) *FEMS Lett.* 16, 241-245.
- [16] Laine, B., Bélaïche, D., Sautière, P. and Biserte, G. (1982) *Biochem. Biophys. Res. Commun.* 106, 101-107.
- [17] Laine, B., Sautière, P., Spassky, A. and Rimsky, S. (1984) *Biochem. Biophys. Res. Commun.* 119, 1147-1153.
- [18] Kimura, M., Kimura, J., Davie, P., Reinhardt, R. and Dijk, J. (1984) *FEBS Lett.* 176, 176-178.
- [19] Rizzo, P.J. and Morris, R.L. (1984) *Bio Systems* 16, 211-216.
- [20] Hirs, C.H.W. (1967) *Methods Enzymol.* 11, 197-199.
- [21] Penke, B., Ferenczi, R. and Kovacs, K. (1974) *Anal. Biochem.* 60, 45-50.
- [22] Laemmli, U.K. (1970) *Nature* 227, 680-685.

Biochimica et Biophysica Acta, 951 (1988) 149–156
Elsevier

149

BBA 91860

Characterization of the chromosomal protein MC1 from the thermophilic archaeobacterium *Methanosarcina* sp. CHTI 55 and its effect on the thermal stability of DNA

François Chartier, Bernard Laine and Pierre Sautiere

Unité Associée CNRS 409, Institut de Recherches sur le Cancer, Lille (France)

(Received 17 March 1988)

Key words: Chromosomal protein; MC1; DNA stability; Deoxyribonucleoprotein complex; Methanogen; (*M. barkeri*)

In the deoxyribonucleoprotein complex of *Methanosarcina* sp. CHTI 55, DNA is associated with two proteins, named MC1 (methanogen chromosomal protein 1) (M_r 10 760) and MC2 (M_r 17 000). Protein MC1, the most abundant of these proteins, is closely related to the *Methanosarcina barkeri* MS protein MC1. The effect of *Methanosarcina* sp. CHTI 55 protein MC1 on the thermal stability of DNA has been studied in native deoxyribonucleoprotein complex, as well as in reconstituted complexes, and it has been compared to the effect of *E. coli* DNA-binding protein II. Both proteins are able to protect DNA against thermal denaturation, but the differences observed in the melting profiles suggest that they interact by different mechanisms. Moreover, our studies indicate that one molecule of protein MC1 protects eight base pairs of DNA.

Introduction

Protein MC1, the major chromosomal protein of the archaeobacterium *M. barkeri* strain MS [1], is a polypeptide of 93 amino acid residues (M_r 10 757), mainly characterized by a high content of charged residues (15% acidic residues and 27% basic residues), which are distributed all along the polypeptide chain [2]. Its primary structure does not show any homology with that of chromosomal proteins, which play a role in DNA condensation, such as histones in eukaryotes, DNA-binding protein II in eubacteria [3] or archaeobacterial proteins

HTa and 7 kDa isolated from *Thermoplasma acidophilum* [4] and *Sulfolobus acidocaldarius* [5], respectively. The association of the protein MC1 with DNA in vivo has been evidenced by its presence in the isolated deoxyribonucleoprotein complex [1] and by the localization of protein MC1 in the DNA-rich areas on cryosections of bacteria cells [6]. However, the role of protein MC1 remains unknown.

In view of a possible role of protein MC1 in the organisation of DNA in methanogenic bacteria, we have investigated its interaction with DNA by thermal denaturation experiments. This method has been proved to be a useful tool to study the structural properties of chromatin [7] and the interaction of eukaryotic and bacterial chromosomal proteins with DNA [8–16]. Our study was performed with the protein MC1 isolated from *Methanosarcina* sp. CHTI 55, which gives higher growth yields than the mesophilic *M. barkeri* strain

Abbreviations: proteins MC1 and MC2, methanogen chromosomal proteins 1 and 2.

Correspondence: B. Laine, Unité Associée CNRS 409, Institut de Recherches sur le Cancer, Place de Verdun, 59045 Lille, France.

150

MS. In contrast to the latter, *Methanosarcina* sp. CHTI 55 is thermophilic and unable to grow in the presence of formate and $H_2 + CO_2$ [17].

Native deoxyribonucleoprotein complex was isolated from *Methanosarcina* sp. CHTI 55 and its protein components were characterized. The thermal denaturation studies were performed on the native complex and on DNA-protein MC1 reconstituted complexes.

Materials and Methods

Methanosarcina sp. CHTI 55 (DSM 2902) was provided by J.P. Touzel (Station de Technologie Alimentaire de l'I.N.R.A., Villeneuve d'Ascq, France) and was grown as described in Ref. 17. Pancreatic ribonuclease (from bovine pancreas, EC 3.1.27.5, purchased from Sigma) was incubated for 10 min at $80^\circ C$ before using to inactivate any trace of deoxyribonuclease.

Preparation of the deoxyribonucleoprotein complex and of DNA

The deoxyribonucleoprotein complex was prepared by a method adapted from Ref. 18. All the steps were performed at $4^\circ C$, except the RNAase treatment. 10 g of frozen cells were suspended in 50 ml of 10 mM sodium phosphate/200 mM NaCl (pH 6.8) in the presence of 10 mM Na_2 -EDTA and 0.1 mM diisopropylfluorophosphate. Cells were disrupted at 700 kg/cm^2 in a French press. Cell debris and ribosomes were removed by centrifugation at $120\,000 \times g$ for 45 min, and then the supernatant was treated with RNAase (50 $\mu\text{g/ml}$) for 15 min at room temperature and centrifuged at $120\,000 \times g$ for 4 h. The pelleted deoxyribonucleoprotein complex was resuspended in 10 mM sodium phosphate/200 mM NaCl (pH 6.8) in the presence of 0.5% NaN_3 and applied to a Sepharose 4B column ($120 \times 2.5\text{ cm}$) equilibrated and eluted with the same buffer at a flow rate of 9 ml/h. The deoxyribonucleoprotein complex eluted in the first peak: the fractions with absorbance ratios $A_{260}/A_{230} > 1.6$ and $A_{260}/A_{280} > 1.7$ were pooled, clarified by centrifugation at $10\,000 \times g$ for 1 h and dialysed for 24 h against 0.15 mM sodium citrate/1.5 mM NaCl (pH 7.0).

To prepare the DNA from *Methanosarcina* sp. CHTI 55, the bacterial deoxyribonucleoprotein

complex pelleted by centrifugation at $120\,000 \times g$ for 4 h was deproteinised by the method described in Ref. 19. The DNA was precipitated by 2 vol. ethanol at $-20^\circ C$ overnight and pelleted at $5000 \times g$ for 15 min. The DNA was redissolved in the column buffer and applied to the Sepharose 4B column. The DNA fractions having an elution volume identical to that of the deoxyribonucleoprotein complex were pooled and dialysed for 24 h against 100 vol. 0.15 mM sodium citrate/1.5 mM NaCl (pH 7.0). Chicken erythrocyte DNA was prepared as described in Ref. 20.

Preparation of the protein

The protein MC1 was prepared as described in Ref. 1 and purified in non-denaturing conditions by selective precipitation with ammonium sulphate [6]. The protein was desalted by gel filtration chromatography on a Sephadex G-25 column equilibrated and eluted with 0.15 mM sodium citrate/1.5 mM NaCl (pH 7.0). The protein concentration was determined by amino acid analysis.

Complex reconstitution

The complexes were prepared directly by mixing DNA (20 $\mu\text{g/ml}$) and increasing amounts of protein as indicated in Fig. 3. The complexes were kept for 24 h at room temperature before thermal denaturation experiments.

Thermal denaturation of DNA

Thermal denaturation of DNA in the presence or absence of protein was performed in 0.15 mM sodium citrate/1.5 mM NaCl (pH 7.0). Absorbance was monitored at 260 nm with a Kontron Uvikon 810 spectrophotometer equipped with a Haacke thermostat and a Kontron Uvikon recorder 21. The frequency of absorbance determination was 1/min. The rates of heating and cooling were 0.5 and $1\text{ }^\circ\text{C}/\text{min}$, respectively. The calculation method was as developed in Ref. 21.

Two-dimensional analytical electrophoresis

The deoxyribonucleoprotein complex was analysed by two-dimensional electrophoresis using the apparatus described by Kerckaert [22]. However, the apparatus was modified as follows for the preparation of the slab gel and for the various steps to carry out the two-dimensional electrophoresis.

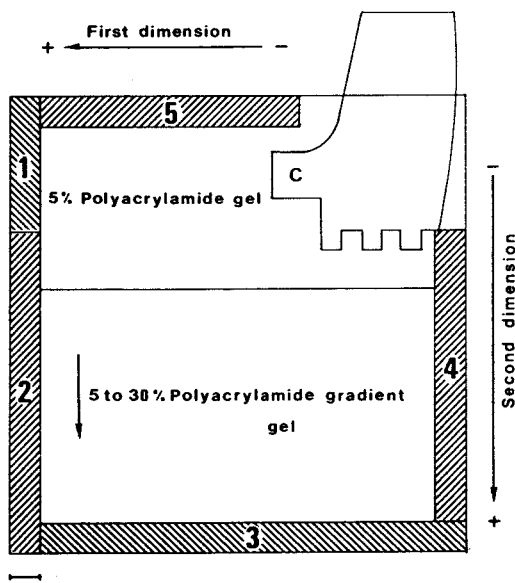


Fig. 1. Assembly of the chamber used in two-dimensional gel electrophoresis. Bar (lower left) represents 1 cm and the spacers were 1.5 mm thick.

Preparation of the slab gel. Glass plates and spacers were assembled as indicated in Fig. 1 to form a gel chamber, which is tightly sealed with rubber. An acrylamide solution in 0.75 M Tris-HCl (pH 8.9) used to form a 5–30% polyacrylamide gradient gel was poured to about 5 cm from the top of the plates. When polymerized, the first gel was overlaid with a 5% polyacrylamide gel in buffer A: 89 mM Tris-boric acid (pH 8.5)/2 mM $\text{Na}_2 \cdot \text{EDTA}$ /13 mM NaCl. The geometry of the comb used in our experiment has been calculated in such a way that one sample loading chamber of 300 μl volume (noted C in Fig. 1) is available for the first dimension, whereas three sample loading chambers of 100 μl volume are available for the second dimension.

Electrophoretic conditions. To carry out electrophoresis in the first dimension, the comb and the spacer 1 (see Fig. 1) were removed, the gel was turned 90° anti-clockwise and the deoxyribonucleoprotein complex loaded in chamber C was submitted to electrophoresis as described in Ref. 22 at 100 V constant voltage in the 5% polyacrylamide gel with the buffer A as electrode buffer. Electrophoresis was stopped before the bromophenol blue reached the bottom of the gel.

After completion of the first dimension, spacers 3 and 5 (see Fig. 1) were removed, the slab gel was turned 90° clockwise and electrophoresis was run in the 5–30% polyacrylamide gel at 50 mA constant current for 2 h in the electrode buffer B (16.5 mM Tris-HCl/130 mM glycine (pH 8.3)) containing 0.1% SDS. To reveal DNA, gel was stained with ethidium bromide in ethanol/acetic acid/water (125:100:775, v/v). Afterwards, proteins were stained as described in Ref. 23. The gel was scanned with a Vernon PHI 4 densitometer.

Results

To prepare the deoxyribonucleoprotein complex of *Methanosarcina* sp. CHTI 55, we had to modify the usual procedure of bacteria disruption [19], since lysozyme is inefficient on the cell wall of methanogens which lacks murein. This modification enabled us to answer to two contradictory aims: (i) to have a good yield of disruption; (ii) to avoid destruction of DNA-protein complexes and excessive splits of DNA. Under the conditions used (see Materials and Methods), about 90% cell disruption was obtained. The combination of ultracentrifugation and gel filtration chromatography yielded deoxyribonucleoprotein complexes with a length of DNA ranging from 3 to 10 kbp (as determined by agarose gel electrophoresis (data not shown)). The deproteinized DNA from *Methanosarcina* sp. CHTI 55 has been chromatographed under the same conditions in order to have DNA length suitable for thermal denaturation assays and to have a valuable comparison with the results obtained with the native deoxyribonucleoprotein complex. The proteins of the deoxyribonucleoprotein complex were analysed by two-dimensional electrophoresis on a polyacrylamide slab gel: the first dimension was run under non-denaturing conditions, whereas the second one was performed in the presence of SDS (Fig. 2). When revealed with ethidium bromide, DNA appears to have been trapped in the first millimeters of the 5% acrylamide gel in the first dimension. The proteins detected after two-dimensional electrophoresis were those which remained associated with the DNA under non-denaturing conditions; their electrophoretic pattern was identical to that of the deoxyribonucleoprotein com-

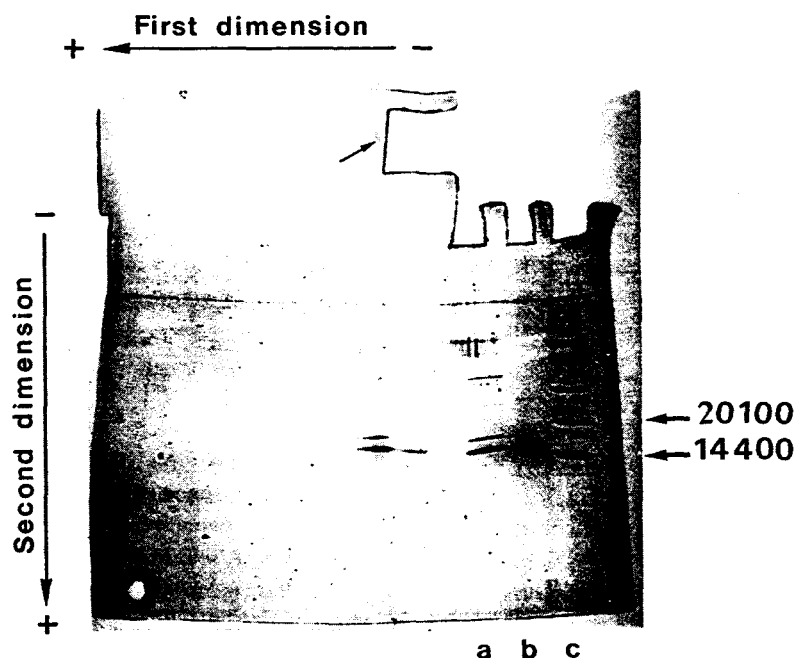


Fig. 2. Two-dimensional gel electrophoresis of the deoxyribonucleoprotein complex from *Methanosarcina* sp. CHTI 55. First dimension: the sample of the bacterial DNA-protein complex (150 μ g) was diluted with 1 vol. buffer A/glycerol (75:25, v/v) (see Materials and Methods) containing bromophenol blue and loaded onto the gel. Second dimension: lane a, bacterial DNA-protein complex (150 μ g); lane b, purified protein MC1 (10 μ g); lane c, low-molecular-weight markers (Pharmacia). Samples were dissolved in electrode buffer B/glycerol (75:25, v/v) (see Materials and Methods) containing bromophenol blue. Arrow indicates the position of DNA-protein complex after the first dimension.

plex analysed directly in the presence of SDS (lane a in Fig. 2). Lack of high-molecular-weight components migrating faster than the deoxyribonucleoprotein complex in the first dimension ascertained the purity of this complex.

The DNA of *Methanosarcina* sp. CHTI 55 is mainly complexed with two proteins of apparent relative molecular mass 14000 and 17000, which represent 80 and 14% of the total proteins associated with the DNA (as determined by densitometric scanning). The smaller protein corresponds to the protein MC1, which has been simultaneously prepared from whole cells by affinity chromatography on a DNA-cellulose column and purified by gel filtration chromatography (Fig. 2, lane b). The proteins MC1 from *Methanosarcina* sp. CHTI 55 and from *M. barkeri* appear closely related: both proteins contain 93 amino acid residues and exhibit very similar amino acid compositions (Table I). Their contents of basic residues are identical

and only slight differences were observed in the amount of hydrophobic residues.

The dissociation of the deoxyribonucleoprotein complex by 0.25 M HCl did not allow us to obtain the 17 kDa protein (protein MC2) with a good yield and in reasonably pure form. Therefore, this protein was not studied further.

The effect of the *Methanosarcina* sp. CHTI 55 protein MC1 on the thermal stability of DNA was studied on reconstituted complexes. In first experiments, the complexes were reconstituted with DNA from *Methanosarcina* sp. CHTI 55. This DNA presented a melting band at 58°C, whereas derivative melting profiles obtained in the presence of protein MC1 were biphasic (Fig. 3A). Each melting assay was performed at least in duplicate and reproducible derivative melting profiles were obtained. Increasing amounts of protein MC1 caused a reduction of the melting band of free DNA (T_m 58°C) and a concomitant increase

TABLE I
AMINO ACID COMPOSITION OF THE *METHANOSARCINA* SP. CHTI 55 CHROMOSOMAL PROTEIN MC1

Results are expressed in numbers of residues per mol of protein. Numbers in parentheses are the nearest integer.

	<i>Methanosarcina</i> sp. CHTI 55	<i>M. barkeri</i> MS ^d
Asp	10.2 (10)	10
Thr ^a	3.8 (4)	4
Ser ^a	2.9 (3)	3
Glu	9.7 (10)	12
Pro	5.9 (6)	6
Gly	5.7 (6)	6
Ala	8.2 (8)	6
Cys	0.0 (0)	0
Val ^b	6.4 (6)	7
Met	0.9 (1)	2
Ile ^b	5.2 (5)	4
Leu	4.2 (4)	3
Tyr	0.0 (0)	0
Phe	4.1 (4)	4
His	1.9 (2)	2
Lys	12.9 (13)	14
Arg	9.3 (9)	8
Trp ^c	1.9 (2)	2
Total	93	93

^a Values for threonine and serine were obtained by linear extrapolation to zero hydrolysis time.

^b 72 h hydrolysis values.

^c Determined after hydrolysis in 3 M mercaptoethanesulfonic acid [27].

^d The amino acid composition of *M. barkeri* protein MC1 [2] is given for comparison.

of a melting band at 70°C (T'_m), which corresponds to DNA regions protected by the protein MC1. It is noticeable that with protein MC1-to-DNA ratios (by weight) higher than 0.5, the melting temperature of the unprotected DNA was slightly raised, whereas the second melting point remained constant. Since DNA of these reconstituted complexes and deoxyribonucleoprotein complexes isolated from the bacteria were similarly prepared, their derivative melting profiles can be reliably compared.

The deoxyribonucleoprotein complex showed the same melting point as free DNA, but the descending slope of the melting profile of the former shifted towards higher temperatures and presented a shoulder near 69°C (Fig. 4). This result can be explained by taking into account the

low protein-to-DNA ratio of the isolated deoxyribonucleoprotein complex, close to 0.1, as determined in the chromatin of the closely related strain *M. barkeri* [6]. Indeed, when the bacterial DNA is complexed with the protein MC1 at a protein-to-DNA ratio equal to 0.25, the second melting band is partially masked by a trail of the first melting band, which was not observed with naked DNA.

Occurrence of biphasic melting profiles led us to apply the quantitative method developed by Li [21], which allows to calculate the length of DNA protected per molecule of protein. This method is based on the equation: $r = B(AT'_m/AT)$ where

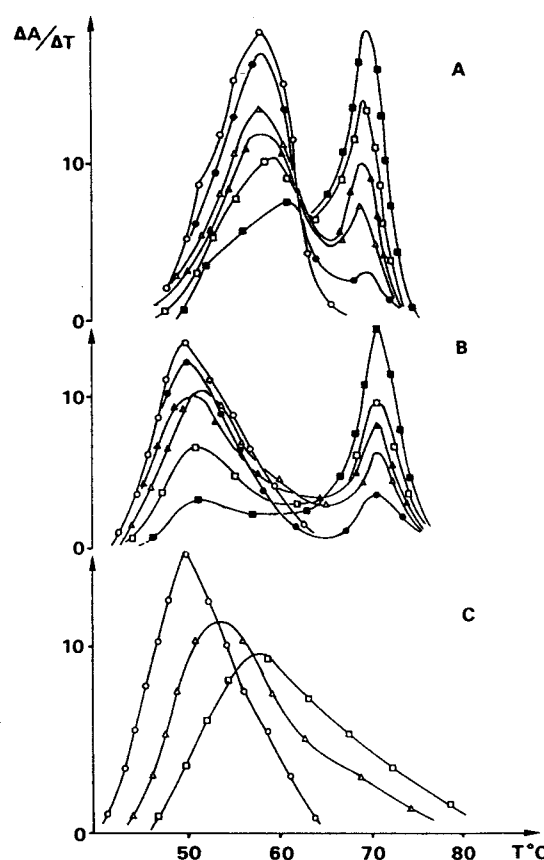


Fig. 3. Derivative melting profiles obtained with reconstituted complexes. (A) protein MC1 and DNA from *Methanosarcina* sp. CHTI 55; (B) protein MC1 and chicken erythrocyte DNA, (C) *E. coli* DNA-binding protein II and chicken erythrocyte DNA. The ratios (w/w) of input protein-to-DNA were: 0 (○—○); 0.22 (●—●); 0.55 (△—△); 0.77 (▲—▲); 1.12 (□—□); 1.60 (■—■).

154

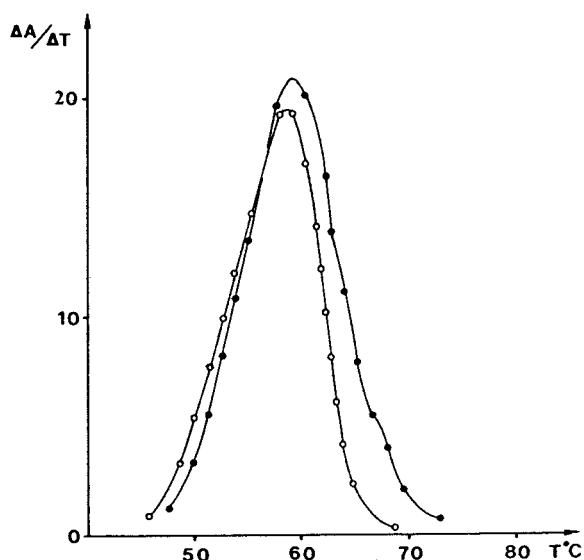


Fig. 4. Derivative melting profile obtained with the deoxyribonucleoprotein complex from *Methanosarcina* sp. CHTI 55 (○—○) compared to that of deproteinized bacterial DNA (●—●).

AT'_m and AT are the areas under the melting band at T'_m and the total area under the whole melting curve, respectively, r is the protein-to-DNA ratio (in mol of protein/nucleotide) and B the number of protein molecules per protected basepair. In order to avoid the overlap of the two melting bands and therefore to reach the best estimation of their areas, a DNA with a lower melting point

was used. Chicken erythrocyte DNA presented a transition (T_m) at 51°C and a second transition appeared at 70°C upon addition of protein MC1 (Fig. 3C). The evolution of the derivative melting profiles was similar to that observed with the bacterial DNA. From the equation cited above, the number of protected basepairs was calculated for each reconstituted complex and for both DNAs. The plotting of the number of protected basepairs versus the amount of protein gave a linear curve, from which it was determined that 8 basepairs were protected by a molecule of protein.

The comparison of the effect of protein MC1 and *E. coli* DNA-binding protein II was performed with chicken erythrocyte DNA to avoid specific interaction of protein MC1 with the host DNA. Addition of *E. coli* DNA-binding protein II to chicken erythrocyte DNA caused an increase of the transition temperature (T_m), which is proportional to the amount of protein (3 and 6 C° for protein-to-DNA ratios equal to 0.5 and 1 by weight, respectively) (Fig. 3C).

Upon cooling, the thermal renaturation profiles of DNA from *Methanosarcina* sp. CHTI 55 and chicken erythrocyte were similar. The extent of thermal renaturation appeared to be independent of the presence of protein MC1 down to 56°C, and then increased substantially with higher amounts of protein MC1. DNA-binding protein II also caused a dose-dependent extent of chicken

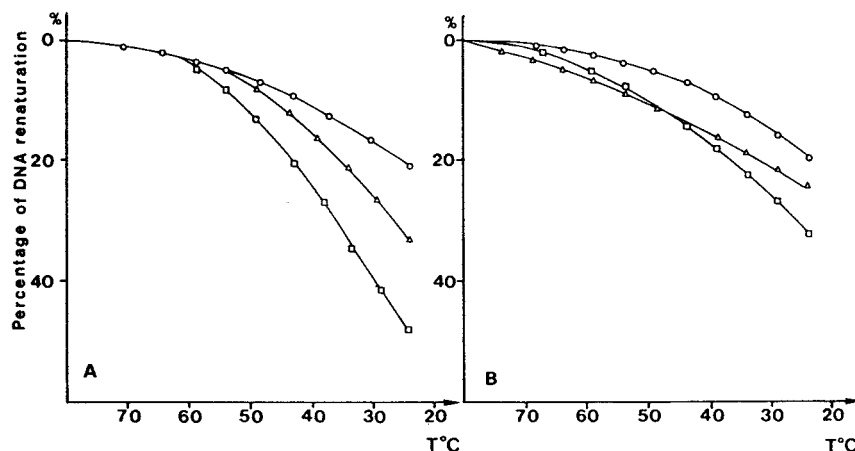


Fig. 5. Renaturation profiles of chicken erythrocyte DNA. (A) in the presence of *Methanosarcina* sp. CHTI 55 protein MC1; (B) in the presence of *E. coli* DNA-binding protein II. The ratios (w/w) of input protein-to-DNA were: 0 (○—○); 0.5 (Δ—Δ); 1.0 (□—□).

erythrocyte DNA renaturation, but the amounts of renaturated DNA were lower than those obtained with protein MC1 at identical protein-to-DNA ratios (Fig. 5).

Discussion

The association of ultracentrifugation with gel exclusion chromatography is efficient to obtain a deoxyribonucleoprotein complex in pure form. The complex was prepared in the presence of 0.2 M NaCl to avoid nonspecific adsorption of protein to the chromosomal DNA. Using a low disrupting pressure in the French press cell, this method is mild enough to yield a complex with DNA fragments suitable for thermal denaturation experiments. This method gives the following advantages over isopycnic centrifugation on a metrizamide gradient: (i) it is not denaturing; (ii) it avoids the difficulty of the complete elimination of metrizamide; (iii) it yields a more homogeneous DNA-protein complex. When using isopycnic centrifugation to isolate the DNA-protein complex of *Methanosarcina* sp. CHTI 55, several contaminants of identical *S* values were observed.

On the basis of amino acid composition (Table I) and sequence data (unpublished results), the protein MC1 from *Methanosarcina* sp. CHTI 55 appears closely related to the protein MC1 from *M. barkeri* MS. This latter protein has been shown to be associated with DNA in vivo by immunocytochemical localization of DNA and protein MC1 in *M. barkeri* cryosections [6]. The deoxyribonucleoprotein complex from *M. barkeri* has been prepared by isopycnic centrifugation and by the method described in this paper. Whatever the method used, protein MC1 is by far the major counterpart of DNA [1,6]. In the deoxyribonucleoprotein complex of *Methanosarcina* sp. CHTI 55, protein MC1 is also the major protein, and a second protein, named MC2 (M_r 17000), accounts for 14% of the protein complement of this complex. The protein MC2 is not encountered in *M. barkeri* MS, where other proteins of larger molecular size are present in very low amounts [6].

Protein MC1 from *Methanosarcina* sp. CHTI 55 exhibits an effective DNA protection effect against thermal denaturation, as shown with com-

plexes reconstituted either with bacterial or eukaryotic DNA. Protein MC1 also causes a substantial increase in the extent of DNA renaturation. The slight raise of the melting point of free DNA, observed with protein-to-DNA ratios larger than 0.5, can be attributed to a tendency of short DNA stretches located between stabilized regions to melt at a higher temperature than long stretches of entirely unprotected DNA. Such an effect has been described for histones [8,12] and for the chromosomal protein HTa from the thermoacidophilic archaeobacterium *T. acidophilum* [11]. Biphasic thermal denaturation profiles are obtained with core particles [12] and with reconstituted complexes formed between DNA and HMG1 or HMG2 proteins [10] or protamines [9]. But protamines are a particular case, since they completely neutralize the phosphate groups in DNA-bound regions. With this phenomenon, also observed with polylysine, the second transition occurs at 98°C. It can be inferred from our thermal denaturation studies that 8-bp-long stretches of DNA are protected by one molecule of protein MC1; this corresponds to 5.8 amino acids per nucleotide. This value is close to that obtained with HMG1 protein (six amino acids per nucleotide) or HMG2 (five amino acids per nucleotide). In contrast with protein MC1 and eukaryotic chromosomal proteins (histones, HMG proteins and protamines), eubacterial DNA-binding protein II causes a shift of the DNA melting point. A shift of the transition temperature of various DNAs induced by the *E. coli* DNA-binding protein II was also observed by Miano et al. [13]. This finding suggests that DNA-binding protein II and protein MC1 interact differently with DNA. The existence of a mechanism of interaction with DNA different from that of core histones has already been reported for histone H1, which also causes a shift of the DNA melting point [12]. The lack of a biphasic melting profile prevents us from determining the number of protected nucleotides per molecule of DNA-binding protein II. Using a different procedure consisting of the binding of chromosomal protein to deoxyoligonucleotides of variable length, Imber et al. [24] have shown that DNA-binding protein II from *Bacillus globigii* covers 8 bp. The *E. coli* DNA-binding protein II affects transcription and stimulates the initiation

of DNA replication [25]. It has been postulated that these properties are correlated with the ability of this protein to wrap DNA and to promote DNA loops [25].

On the other hand, the intracellular ionic strength of Methanosarcinaceae [26], which is much higher than that of 0.15 mM sodium citrate/1.5 mM NaCl (pH 7.0), allows stabilization of the host DNA up to the temperature of optimal growth rate (55°C). Therefore, the protective effect of protein MC1 against thermal denaturation of DNA probably does not come under a physiological role. Nevertheless, this protective effect is a characteristic shared with histones and the DNA-binding protein II, which are involved in packaging of DNA. This characteristic suggests that protein MC1 may play a structural role in the chromatin of methanogenic bacteria. However, this structural role most probably arises through a way different from that of eukaryotic histones: observations in electron microscopy clearly show the lack of nucleosomal structure in *M. barkeri* MS chromatin [6]. Moreover, the protein-to-DNA ratio in the bacteria is too low to compact the whole chromatin. On the other hand, the protein may locally introduce a DNA structure which is stabilized against thermal denaturation and is favorable to modulate the expression of the genome of *Methanosarcina* sp. CHTI 55.

Acknowledgements

The authors are indebted to Dr. J.P. Touzel (Station de Technologie Alimentaire, INRA, Villeneuve d'Ascq) for helpful advice and for providing the bacterial cells. They also thank Dr M. Daune for stimulating discussions. They are grateful to A. Lemaire and M.J. Dupire for skilful technical assistance and to T. Ernout for kindly typing this manuscript. This work was supported by grants from the Centre National de la Recherche Scientifique, from the Pôle des Anaérobies de la région Nord-Pas de Calais and from the Université de Lille II.

References

- 1 Chartier, F., Laine, B., Sautière, P., Touzel, J.P. and Albagnac, G. (1985) FEBS Lett. 183, 119–123.
- 2 Laine, B., Chartier, F., Imbert, M., Lewis, R. and Sautière, P. (1986) Eur. J. Biochem. 161, 681–687.
- 3 Laine, B., Bélaïche, D., Khanaka, H. and Sautière, P. (1983) Eur. J. Biochem. 131, 325–331.
- 4 DeLange, R.J., Williams, L.C. and Searcy, D.G. (1981) J. Biol. Chem. 256, 905–911.
- 5 Kimura, M., Kimura, J., Davie, P., Reinhart, R. and Dijk, J. (1984) FEBS Lett. 176, 176–178.
- 6 Imbert, M., Laine, B., Prensier, G., Touzel, J.P. and Sautière, P. (1988) Can. J. Microbiol., in press.
- 7 Subirana, J.A. (1973) J. Mol. Biol. 74, 363–386.
- 8 Wilhelm, F.X., De Murcia, G.M., Champagne, M.H. and Daune, M.P. (1974) Eur. J. Biochem. 45, 431–443.
- 9 Willmitzer, L., Bode, J. and Wagner, K.G. (1977) Nucleic Acids Res. 4, 149–175.
- 10 Yu, S.S., Li, H.J., Goodwin, G.H. and Johns, E.W. (1977) Eur. J. Biochem. 78, 497–502.
- 11 Stein, D.B. and Searcy, D.G. (1978) Science 202, 219–221.
- 12 Weischet, W.O., Tatchell, K., Van Holde, K.E. and Klump, H. (1978) Nucleic Acids Res. 5, 139–160.
- 13 Miano, A., Losso, M.A., Gianfranceschi, G.L. and Gualerzi, C.O. (1982) Biochem. Int. 5, 415–422.
- 14 Green, G.R., Searcy, D.G. and DeLange, R.G. (1983) Biochim. Biophys. Acta 741, 251–257.
- 15 Zierer, R., Grote, M., Dijk, J. and Wilson, K. (1986) FEBS Lett. 194, 235–241.
- 16 Hermann, A.L., Cadilla, C.L., Cacheiro, L.H., Carne, A.F. and Olins, D.E. (1987) Eur. J. Cell Biol. 43, 155–162.
- 17 Touzel, J.P., Petroff, D. and Albagnac, G. (1985) System. Appl. Microbiol. 6, 66–71.
- 18 Searcy, D.G. and Stein, D.B. (1980) Biochim. Biophys. Acta 609, 180–195.
- 19 Johnson, J.L. (1981) in Manual of Methods for General Bacteriology (Gerhardt et al., eds.), pp. 453–455, American Society for Microbiology, Washington, DC.
- 20 Kay, E.R.M., Simmons, M.S. and Dounce, A.L. (1952) J. Am. Chem. Soc. 74, 1724–1726.
- 21 Li, H.J. (1978) Methods in Cell Biol. 18, 385–415.
- 22 Kerckaert, J.P. (1978) Anal. Biochem. 84, 354–360.
- 23 Stephano, J.L., Gould, M. and Rojas-Galicia, L. (1986) Anal. Biochem. 152, 308–313.
- 24 Imber, R., Kimura, M., Groch, N. and Heinemann, U. (1987) Eur. J. Biochem. 165, 547–552.
- 25 Drlica, K. and Rouvière-Yaniv, J. (1987) Microbiol. Rev. 51, 301–319.
- 26 Sprott, G.D., Shaw, K.M. and Jarrell, K.F. (1984) J. Biol. Chem. 259, 12602–12608.
- 27 Penke, B., Ferenczy, R. and Kovacs, K. (1974) Anal. Biochem. 60, 45–50.

Journal of Chromatography, 466 (1989) 331–337
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 21 186

HIGH-PERFORMANCE LIQUID CHROMATOGRAPHIC SEPARATION OF VARIANTS OF CHROMOSOMAL PROTEINS FROM PROKARYOTES

FRANÇOIS CHARTIER*, GILLES CREVEL, BERNARD LAINE and PIERRE SAUTIERE
URA 409 CNRS, Institut de Recherches sur le Cancer de Lille, Place de Verdun, 59045 Lille (France)
(First received September 14th, 1988; revised manuscript received December 13th, 1988)

SUMMARY

The separation of variants of chromosomal proteins exhibiting closely related amino acid compositions has been achieved using weak cation-exchange or reversed-phase high-performance liquid chromatography. The purity of the isolated proteins has been ascertained by polyacrylamide gel electrophoresis and in several cases by micro-sequencing.

INTRODUCTION

In the nucleoid of prokaryotes, acid-soluble proteins of molecular weights (M_r) ranging from 7000 to 16000 daltons were shown to be involved in DNA condensation¹. In several cases, these proteins are heterodimers constituted of polypeptides of closely related amino acid sequences which are referred to as variants. As an example, the number of sequence differences between variants of the DNA-binding protein II ($M_r \approx 10000$ daltons) ranges from only one residue in *Thermoplasma acidophilum*² to 27 residues in *Escherichia coli*³. Moreover, most of the observed changes are conservative. Highly resolving chromatographic methods are therefore required to separate variants of chromosomal proteins with a view to their characterization by amino acid composition and sequence analyses. Ion-exchange and reversed-phase high-performance liquid chromatography (HPLC) have been shown to be suitable for resolving complex mixtures of basic proteins differing in molecular mass, charge and hydrophobicity such as ribosomal proteins^{4–8}. This paper deals with the application of HPLC to the separation of variants of eubacterial and archaebacterial chromosomal proteins.

MATERIALS AND METHODS

HPLC-grade acetonitrile was obtained from Carlo Erba and sequanal grade trifluoroacetic acid (TFA) from Pierce. Water for HPLC was provided by an Elgastat UHQ water apparatus. All other chemicals were proanalysis grade. Saline buffers were filtered through a 0.22- μm Millipore filter.

HPLC separations were performed with Beckman equipment consisting of

a pair of Model 126 Altex pumps and a Model 167 variable wavelength detector controlled by a Tandon computer.

Preparation of chromosomal proteins

Proteins were prepared by affinity chromatography on a DNA-cellulose column as described in ref. 9 for eubacterial DNA-binding protein II and as in ref. 10 for the protein MCI from the archaeobacterium *Methanotherix soehngeni*. The eubacterial proteins eluted from the DNA-cellulose column were desalted on a Sephadex G-25 column equilibrated and eluted in 10 mM hydrochloric acid, and freeze-dried. The protein MCI was dialysed against water and concentrated in a SpeedVac apparatus.

For cation-exchange chromatography, the proteins were dissolved in 200 μ l of 10 mM sodium acetate buffer, pH 5.6 containing 8 M urea. The variants were separated on a TSK IEX CM-2SW column (250 mm \times 4.6 mm) (Beckman) equilibrated in 10 mM sodium acetate buffer (pH 5.6) containing 6 M urea and eluted with a linear gradient of NaCl in the same buffer. Proteins were then desalted on a reversed-phase column as described in ref. 11.

Reversed-phase chromatography was performed on a Ultrapore C₈ column from Beckman (particle size 5 μ m, pore size 300 Å, column size 250 mm \times 4.6 mm). The proteins dissolved in 10 mM hydrochloric acid were loaded on the column equilibrated in 0.05% TFA in water and eluted with a gradient of acetonitrile in 0.05% TFA.

Analytical gel electrophoreses and amino acid analyses were performed as described in ref. 9.

RESULTS AND DISCUSSION

Cation-exchange HPLC

The separation by ion-exchange HPLC of the two variants 1 and 2 of the DNA-binding protein II from *E. coli* is shown in Fig. 1. In preliminary experiments performed at pH 7.0 with 10 mM sodium phosphate, the two variants were coeluted in a single fraction, whereas at pH 6.0 with 10 mM sodium phosphate or sodium acetate they were partially resolved (data not shown). The use of 10 mM sodium acetate at pH 5.6 gave the best resolution. The separation of these variants appears mainly dependent on the pH and, to a lesser extent, on the nature of the salt used as the buffer. Sequence analyses³ indicate that the only difference in net charge between the two variants of this protein can be brought by the presence in variant 2 of one histidine residue. Since the pK' of the imidazole group is equal to 6.0, the protonation of the histidine residue probably plays an important role in the retention time of this protein. On the other hand, the presence of 6 M urea in the buffer is necessary to dissociate the two polypeptide chains which tightly bind together to form a stable dimer corresponding to the functional state of the protein^{1,2}.

The chromatogram obtained with the *Azotobacter vinelandii* DNA-binding protein II using the sodium acetate buffer at pH 5.6 is presented in Fig. 2. Variants 1 and 2 were eluted in fractions 1 and 2 respectively. By contrast with *E. coli* variants which are in equimolar amounts, the variant 1 is about twice as abundant as the variant 2 in *A. vinelandii*. The amino acid compositions of variants 1 and 2 present only slight differences except in the amounts of threonine and glutamic acid (Table I). The variants display the same total number of lysine plus arginine and differ by the presence of two histidine residues in variant 2.

HPLC OF CHROMOSOMAL PROTEINS FROM PROKARYOTES

333

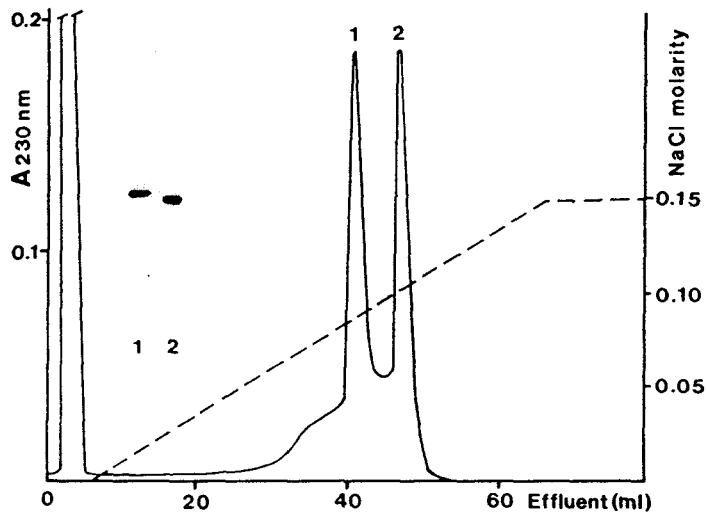


Fig. 1. Separation of the variants of the *E. coli* DNA-binding protein II (2 mg injected) on a TSK IEX CM-2SW column (250 mm \times 4.6 mm). --- = Linear gradient of sodium chloride in 10 mM sodium acetate pH 6 containing 6 M urea; flow-rate 1 ml/min. Fractions of 1 ml were collected. Insert: polyacrylamide gel electrophoresis (PAGE) of variants 1 (lane 1) and 2 (lane 2) in 0.9 M acetic acid containing 6.25 M urea and 0.38% Triton X-100. Samples (3 μ g) dissolved in 10 mM hydrochloric acid, 8 M urea, 0.5 M 2-mercaptoethanol were run at 22 mA for 3 h at room temperature in the gel containing 17% acrylamide. The gel was stained and destained according to ref. 14.

Reversed-phase HPLC

The DNA-binding protein II from *Synechococcus* PCC 7002 was desorbed from the DNA-cellulose column together with a contaminant protein of $M_r \approx 16000$ daltons. Using a C_8 Ultrapore column eluted with a linear gradient of acetonitrile in 0.05% TFA, the contaminant protein was eluted in fraction 1 whereas the DNA-

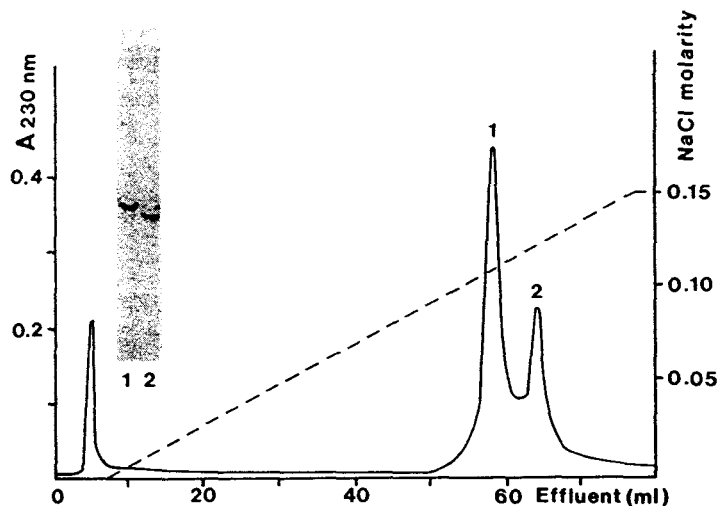


Fig. 2. Separation of the variants of the *A. vinelandii* DNA-binding protein II (2.5 mg injected) on a TSK IEX CM-2SW column (250 mm \times 4.6 mm). The protein was chromatographed as indicated in Fig. 1. Insert: PAGE of variants 1 (lane 1) and 2 (lane 2) in 0.9 M acetic acid containing 2.5 M urea. Preparation of samples and electrophoretic conditions were as in Fig. 1.

TABLE I

AMINO ACID COMPOSITIONS OF VARIANTS OF DNA-BINDING PROTEIN II FROM EUBACTERIA

Results are expressed as the number of residues per mol of protein. Numbers in parentheses are the nearest integers. Values for *Synechococcus* 7002 variants are from ref. 15.

Amino acid	<i>A. vinelandii</i>		<i>Synechococcus</i> 7002	
	Variant 1	Variant 2	Variant 1	Variant 2
Asp	9.1 (9)	7.9 (8)	8	7
Thr	4.2 (4)	8.3 (8)	6	6
Ser	3.9 (4)	4.1 (4)	5	6
Glu	5.3 (5)	8.7 (9)	12	11
Pro	3.8 (4)	2.1 (2)	5	5
Gly	9.1 (9)	9.2 (9)	7	7
Ala	18.0 (18)	14.0 (14)	12	11
Val	8.0 (8)	6.3 (6)	9	8
Met	0.7 (1)	0.7 (1)	4	4
Ile	6.7 (7)	5.1 (5)	4	4
Leu	5.2 (5)	7.0 (7)	3	3
Phe	3.0 (3)	3.0 (3)	4	4
His	0 (0)	1.8 (2)	1	2
Lys	8.7 (9)	9.7 (10)	11	10
Arg	4.0 (4)	3.1 (3)	4	6
Total	90	91	94	95

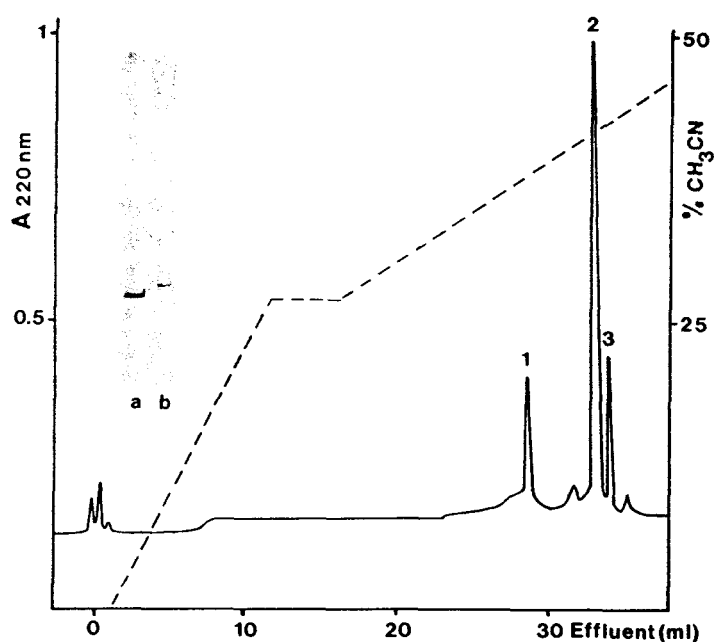


Fig. 3. Separation of variants of *Synechococcus* DNA-binding protein II (0.3 mg injected) on an Ultrapore C₈ column (250 mm × 4.6 mm). --- = Gradient of acetonitrile in 0.05% TFA, flow-rate 1 ml/min. Fractions of 0.5 ml were collected. Insert: PAGE of variants 1 (lane a) and 2 (lane b) in 0.9 M acetic acid containing 6.25 M urea. Preparation of samples and electrophoretic conditions were as in Fig. 1.

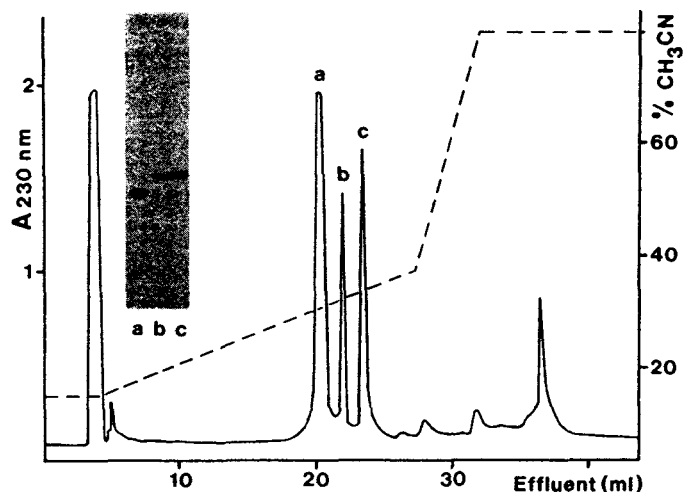


Fig. 4. Separation of variants of the protein MC1 from the archaebacterium *Methanotherix soehngeni* on an Ultrapore C₈ column (250 mm × 4.6 mm). The crude fraction eluted from the DNA-cellulose was concentrated in a Speedvac apparatus and injected. --- = Gradient of acetonitrile in 0.05% TFA; flow-rate 1 ml/min. Fractions of 0.5 ml were collected. Insert: PAGE of variants a (lane a), b (lane b) and c (lane c) in 0.9 M acetic acid containing 2.5 M urea. Preparation of samples and electrophoretic conditions were as in Fig. 1.

TABLE II

AMINO ACID COMPOSITIONS OF THE VARIANTS OF *Methanotherix soehngeni* PROTEIN MC1

Results are expressed as mol per 100 mol. n.d. = Not determined.

Amino acid	MC1a	MC1b	MC1c
Asp	7.2	8.7	9.1
Thr ^a	3.5	3.6	2.6
Ser ^a	0.0	2.3	2.5
Glu	10.6	12.3	13.0
Pro	6.0	6.0	6.6
Gly	9.2	9.4	8.9
Ala	10.4	11.8	8.5
Cys	0.0	0.0	0.0
Val	7.1	6.0 ^b	5.8
Met	1.1	2.1	2.2
Ile	6.8	5.9 ^b	7.4
Leu	4.9	5.0	5.9
Tyr	0.0	1.1	0.0
Phe	3.5	3.5	4.4
His	1.9	1.2	1.1
Lys	18.5	12.6	13.3
Arg	9.3	8.5	8.7
Trp	n.d.	n.d.	n.d.
Total	100.0	100.0	100.0

^a Values obtained by linear extrapolation to zero hydrolysis time.

^b 72-h Hydrolysis values.

binding protein II was resolved in two variants obtained in fractions 2 and 3 (Fig. 3). The high resolution of the column is demonstrated by the separation of these variants which exhibit similar numbers of amino acid residues with aromatic or bulky aliphatic side chains (Table I)¹³.

The chromosomal protein MC1 isolated from the archaebacterium *Methanotherix soehngeni* is a complex constituted of three variants named a, b and c which were eluted from the Ultrapore C₈ column in fractions 1, 2 and 3 respectively (Fig. 4). The three variants have about the same molecular size ($M_r \approx 11\,000$ daltons). From their electrophoretic migration in acid-urea polyacrylamide gel (see insert, Fig. 4), variants b and c have a similar charge whereas variant a is more basic. The three variants were separated in a single step according to their hydrophobicities (Table II). Micro-sequence analyses of the three polypeptides clearly show that they are obtained in pure form and are structurally related to each other, which indicates that they represent variants of the same protein. Taking into account the low amount of protein available and the complexity of the DNA-cellulose fraction from which the protein MC1 variants were prepared, purification of the three variants in a single step shows the high efficiency of the method.

CONCLUSIONS

Variants of chromosomal proteins isolated from different prokaryotic organisms have been separated by ion-exchange or reversed-phase HPLC. The results demonstrate the efficiency of these methodologies for the separation of variants exhibiting only slight differences in amino acid compositions within each bacterial strain.

Previous chromatography of *E. coli* DNA-binding protein II on carboxymethyl-cellulose gave similar results³ but required a much longer time (1 week *versus* 2 h) and six times as much mobile phase as does HPLC. Moreover, the sensitivity of HPLC is suitable to identify variants when low amounts of proteins are available and/or when the stoichiometry of one out of the variants is very low.

ACKNOWLEDGEMENTS

The authors are indebted to Professor J. Guillaume (Université des Sciences et Techniques de Lille Flandres Artois) and B. Gruson (Station de Technologie Alimentaire, Institut National de la Recherche Agronomique, Villeneuve d'Ascq) for providing the bacterial cells. They are grateful to A. Lemaire, M. J. Dupire and A. Hemez for skilful technical assistance and to T. Ernout for kindly typing this manuscript. They thank also Dr. Anselme (Beckman France) for helpful advice. This work was supported by grants from the Centre National de la Recherche Scientifique, from the Pôle des Anaerobies de la Région Nord-Pas de Calais and from the Université de Lille II.

REFERENCES

- 1 K. Drlica and J. Rouvière-Yaniv, *Microbiol. Rev.*, 51 (1987) 301.
- 2 R. J. DeLange, L. C. Williams and D. G. Searcy, *J. Biol. Chem.*, 256 (1981) 905.
- 3 B. Laine, D. Kmiecik, P. Sautière, G. Biserte and M. Cohen-Solal, *Eur. J. Biochem.*, 103 (1980) 447.

HPLC OF CHROMOSOMAL PROTEINS FROM PROKARYOTES

337

- 4 A. R. Kerlavage, L. Kahan and B. S. Cooperman, *Anal. Biochem.*, 123 (1982) 342.
- 5 R. M. Kamp and B. Wittmann-Liebold, *FEBS Lett.*, 167 (1984) 59.
- 6 P. N. Dalrymple, S. Gupta, F. Regnier and L. L. Houston, *Biochim. Biophys. Acta*, 755 (1983) 157.
- 7 R. M. Kamp, A. Bosserhoff, D. Kamp and B. Wittmann-Liebold, *J. Chromatogr.*, 317 (1984) 181.
- 8 M. Capel, D. Datta, C. R. Nierras and G. R. Craven, *Anal. Biochem.*, 158 (1986) 179.
- 9 B. Laine, D. Bélaiche, P. Sautière and G. Biserte, *Biochem. Biophys. Res. Commun.*, 106 (1982) 101.
- 10 F. Chartier, B. Laine, P. Sautière, J. P. Touzel and G. Albagnac, *FEBS Lett.*, 183 (1985) 119.
- 11 T. Pohl and R. M. Kamp, *Anal. Biochem.*, 160 (1987) 388.
- 12 I. Tanaka, K. Appelt, J. Dijk, S. W. White and K. S. Wilson, *Nature (London)*, 310 (1984) 376.
- 13 T. Sasagawa, T. Okuyama and D. C. Teller, *J. Chromatogr.*, 240 (1982) 329.
- 14 J. L. Stephano, M. Gould and L. Rojas-Galicia, *Anal. Biochem.*, 152 (1986) 308.
- 15 G. Crevel, B. Laine, P. Sautière and C. Galleron, *Biochim. Biophys. Acta*, 1007 (1989) 36.

II - 3 CARACTÉRISTIQUES DE LA PROTÉINE MC1 ISOLÉE DE DIFFÉRENTES SOUCHES MÉSOPHILES ET THERMOPHILES DE METHANOSARCINACEAE

II - 3-1 COMPORTEMENT DE LA PROTEINE MC1 SUR COLONNE D'ADN- CELLULOSE

Les résultats obtenus lors de la préparation de la protéine MC1 par chromatographie d'affinité sur colonne d'ADN d'*Escherichia coli* appellent les commentaires suivants :

- dans le cas des souches du genre *Methanosarcina* : *Methanosarcina barkeri*, *Methanosarcina* sp CHTI 55 et *Methanosarcina thermophila* TM1, les diagrammes d'élution de la colonne d'ADN-cellulose sont très voisins. Un diagramme type est présenté dans la figure 7.
- dans le cas de *Methanotherix soehngeni*, la protéine MC1 est élue dans deux pics distincts mais mal résolus (Figure 8 α). L'analyse par CLHP (Figures 8 β , γ et δ) de ces deux pics montre que le variant b est essentiellement élué dans le premier pic (fraction 1, figure 8 α) avec de nombreux contaminants. Le variant c est essentiellement élué dans la fraction 2 tandis que le variant a représente le composant majeur de la fraction 3 (figure 8 α). Ce résultat s'explique par la composition en acides aminés de ces trois variants; il apparaît en effet clairement que l'ordre d'élution des variants est relié à la richesse en acides aminés basiques : variant b (22,3%), variant c (23,1%) et variant a (29,7%).

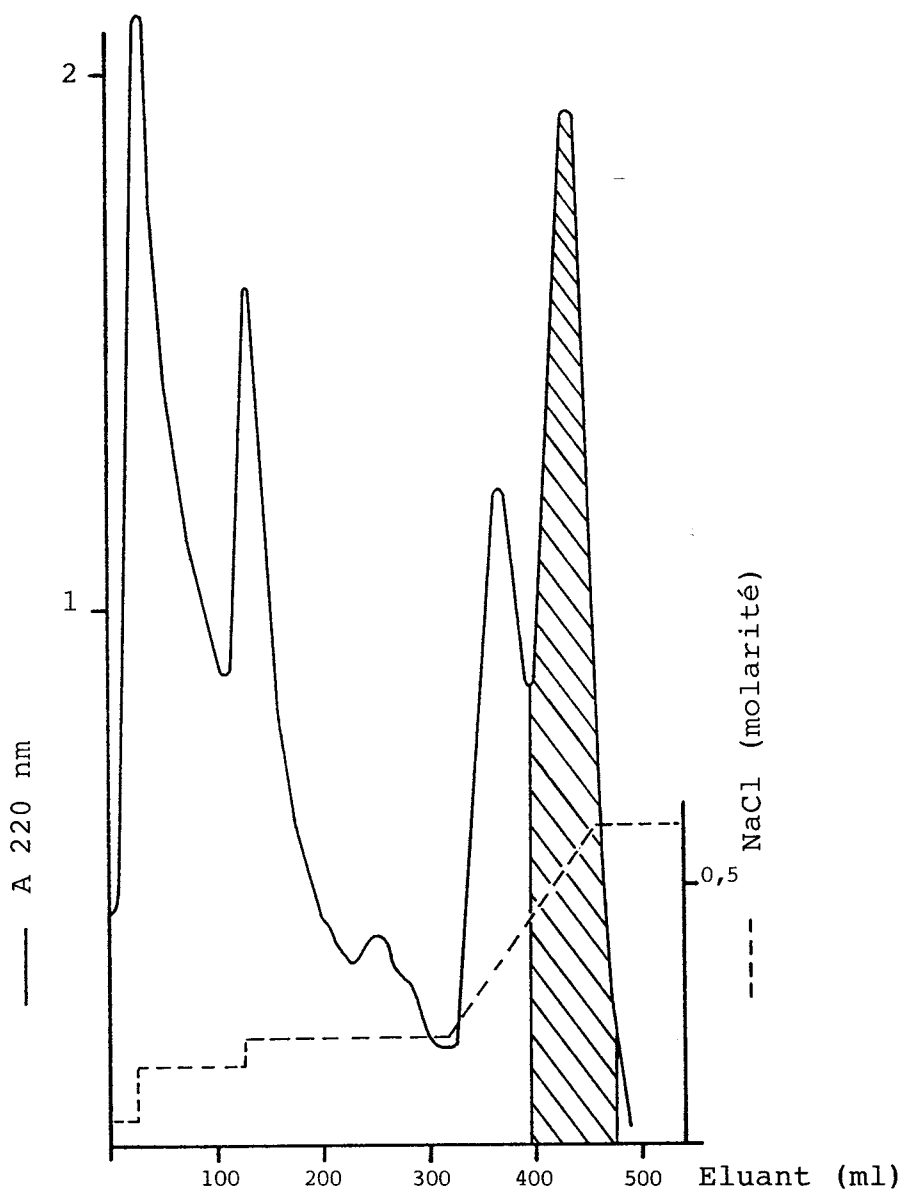


FIGURE 7 - Chromatographie d'affinité sur colonne d'ADN-cellulose d'un homogénat d'une bactérie du genre *Methanosarcina*.

La colonne d'ADN-cellulose (20 x 2,5 cm) est équilibrée dans un tampon phosphate de sodium 10 mM, NaCl 50 mM, avec un débit de 40 ml/h. Les fractions sont éluées par un gradient de NaCl (-----), des fractions de 6,6 ml sont collectées. Le pic hachuré correspond à la fraction enrichie en protéine MCl.

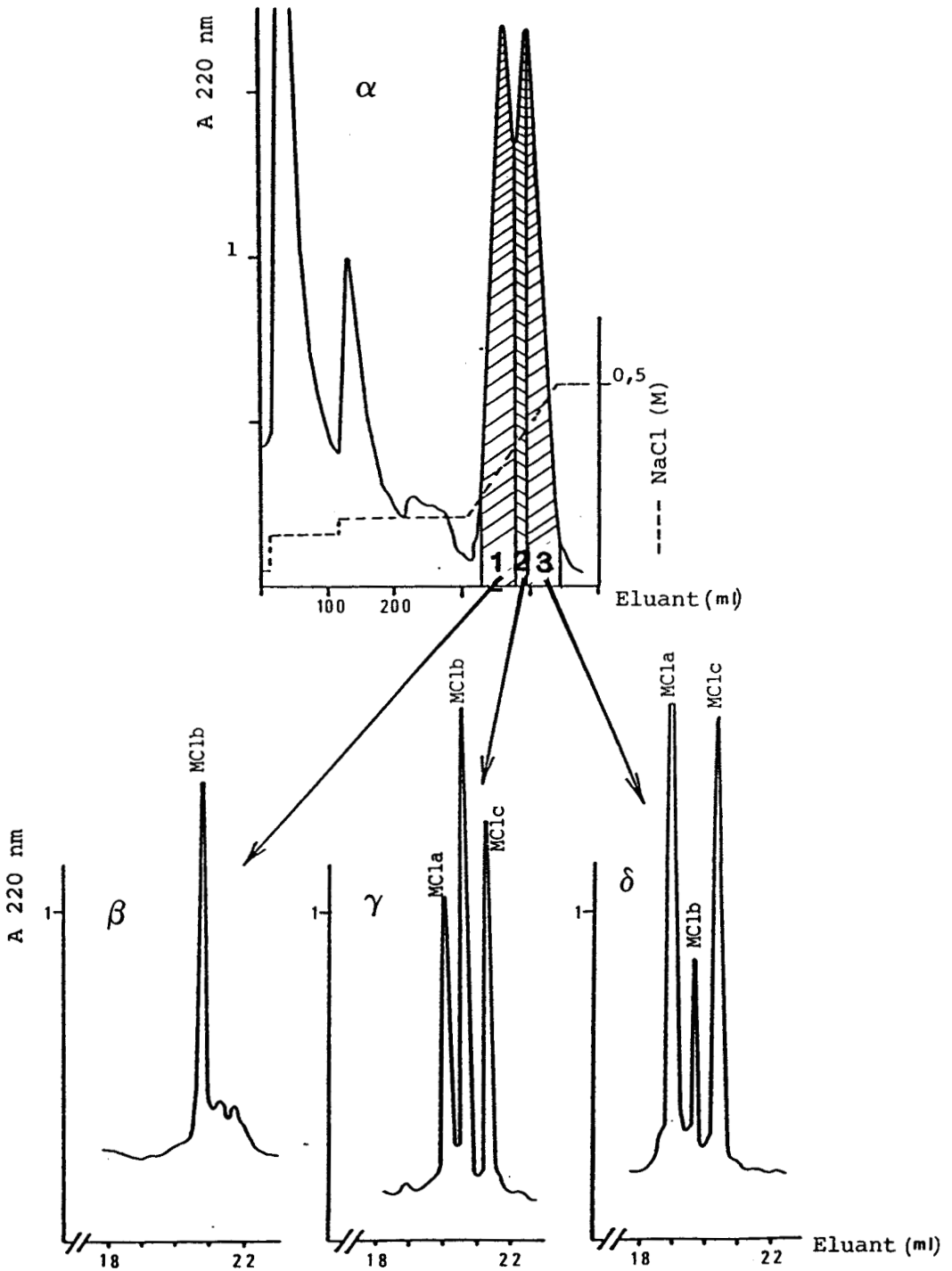


FIGURE 8 - Préparation des variants de protéine MCl de *Methanotrix soehngeni*.

α : chromatographie d'affinité sur la colonne d'ADN-cellulose. Les conditions sont décrites dans la figure 7. Les pics hachurés correspondent aux fractions enrichies en protéine MCl.

Les variants de protéine MCl contenus dans chaque fraction sont ensuite purifiés par CLHP en phase inverse sur colonne de C₈ (β, γ, δ). Les modalités de chromatographie sur la colonne de C₈ sont décrites dans la publication n°3.

II - 3-2 MOBILITES ELECTROPHORETIQUES

Toutes les protéines de type MCl que nous avons purifiées présentent une masse moléculaire apparente voisine de 14.400 sur gel de polyacrylamide en présence de SDS (Figure 9). En fait la masse moléculaire calculée à partir des données structurales de ces protéines varie entre 9.760 et 10.760. Ce comportement électrophorétique anormal en gel dénaturant est étroitement relié à la basicité des protéines. Un tel phénomène a déjà été décrit pour d'autres protéines basiques telles que l'histone H4.

La migration électrophorétique sur gel de polyacrylamide en milieu acide acétique 0,9 M - urée 2,5 M des protéines MCl que nous avons étudiées est présentée dans la figure 10. Les protéines MCl ont des migrations électrophorétiques voisines sauf la protéine MCl de *Methanosarcina* MST-A1 qui a une migration inférieure et la protéine MCl_a de *Methanotherrix soehngeni* qui a une migration nettement plus cathodique. Le comportement de cette dernière protéine s'explique par son taux élevé de résidus basiques (voir paragraphe suivant).

II - 3-3 COMPOSITIONS EN ACIDES AMINES

Les compositions en acides aminés de la protéine MCl des souches étudiées sont présentées dans le tableau III.

La protéine MCl se caractérise par sa richesse en acides aminés basiques (de 23 à 29 %) et en acides aminés dicarboxyliques (de 11 à 15 %) et par la présence de deux résidus de tryptophane. La protéine MCl est dépourvue de cystéine et de tyrosine, à l'exception du variant MCl_b de *Methanotherrix soehngeni* qui renferme un résidu de tyrosine.

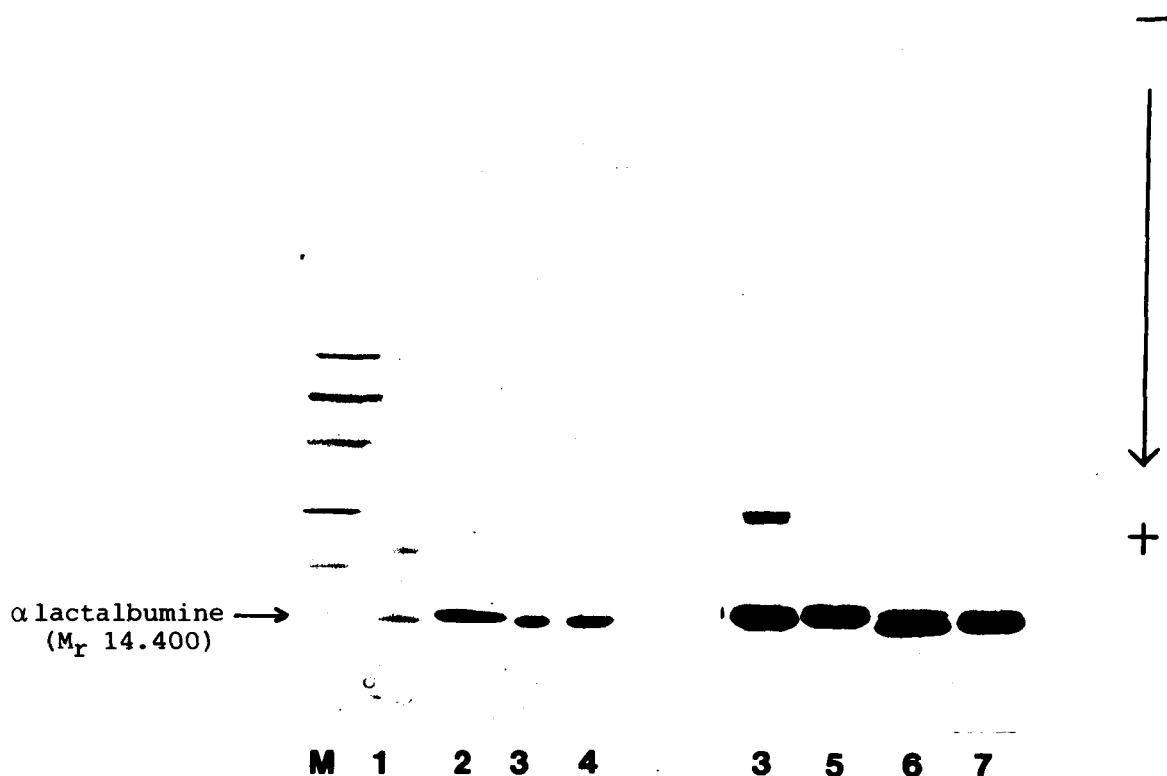


FIGURE 9 - Mobilité électrophorétique de la protéine MCl de souches de Methanosarcinaceae en présence de SDS.

M : témoins de protéines de faibles masses moléculaires (Kit Pharmacia).

La protéine MCl a été purifiée à partir de :

1 : *Methanosarcina barkeri*

2 : *Methanosarcina* MST-A1

3 : *Methanosarcina* sp CHTI 55

4 : *Methanosarcina thermophila* TM1

5, 6 et 7 : variants a, b et c de la protéine MCl de *Methanotherix soehngeni* respectivement

Les conditions électrophorétiques sont décrites dans la publication n°1.

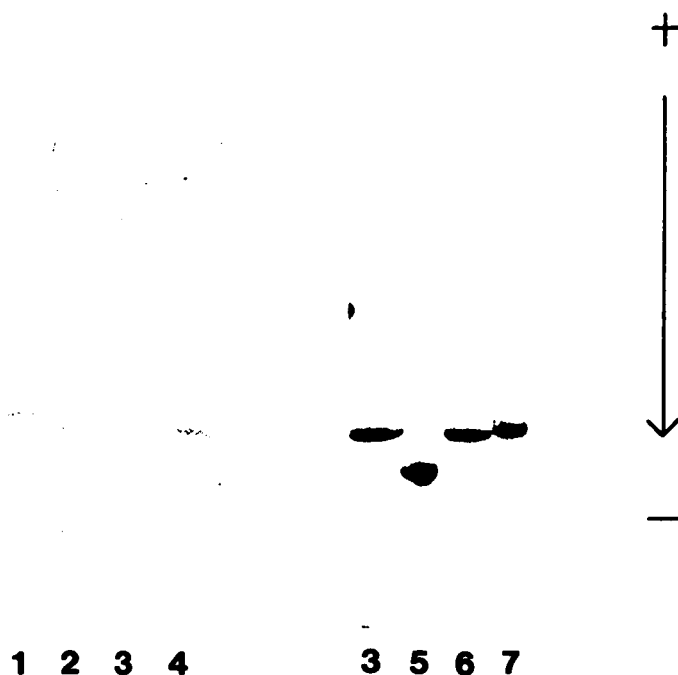


FIGURE 10 - Mobilité électrophorétique en milieu acide acétique (0,9M), urée (2,5M) de la protéine MCl isolée de différentes souches de Methanosarcinaceae.

Protéine MCl isolée de :

- 1 : *Methanosarcina* MST-A1
- 2 : *Methanosarcina barkeri*
- 3 : *Methanosarcina* sp CHTI 55
- 4 : *Methanosarcina thermophila* TM1
- 5, 6 et 7 : variants a, b et c de la protéine MCl de *Methanotrix soehngeni* respectivement.

Les conditions électrophorétiques sont décrites dans la publication n°1.

TABLEAU III

Compositions en acides aminés de la protéine MC1 isolée de Methanosarcinaceae
Les résultats sont exprimés en nombre de résidus d'acides aminés par mole de protéine.

Origine Acides aminés	<i>Methanosarcina</i>			<i>Thermophila</i> TM1	<i>Methanotherix soehngeni</i> variants		
	<i>barkeri</i> MS	MST-A1*	sp CHTI 55		a	b	c
Acide aspartique	10	9	10	10	7	7	9
Thréonine	4	4	4	4	3	3	2
Sérine	3	3	3	3	0	2	2
Acide glutamique	12	11	10	10	9	10	11
Proline	6	6	6	6	5	5	6
Glycine	6	6	6	6	8	8	7
Alanine	6	8	8	8	9	10	7
Cystéine	-	-	-	-	-	-	-
Valine	7	6	6	6	6	5	5
Méthionine	2	1	1	1	1	2	2
Isoleucine	4	5	5	5	6	5	7
Leucine	3	4	4	4	4	4	5
Tyrosine	0	0	0	0	0	1	0
Phénylalanine	4	4	4	4	3	3	4
Histidine	2	2	2	2	2	1	1
Lysine	14	13	13	13	16	12	12
Arginine	8	8	9	9	8	7	8
Tryptophane**	2	2	2	2	2	2	2
Total	93	93	93	93	89	87	90
Nombre d'acides aminés dicarboxyliques**	14	12	12	?	10	11	10
Nombre d'acides aminés basiques	24	23	24	24	26	20	21
Charge nette	+ 10	+ 11	+ 12	?	+ 16	+ 9	+ 11

* selon M.IMBERT (1987)

** déterminé d'après la séquence

Les compositions en acides aminés de la protéine MCl obtenue à partir de bactéries du genre *Methanosarcina* sont très voisines. Les compositions en acides aminés de la protéine MCl obtenue à partir de *Methanosarcina* sp CHTI 55 et de *Methanosarcina thermophila* TM1 sont même rigoureusement identiques. Les variants de la protéine MCl du genre *Methanotherix* se différencient nettement de la protéine MCl des souches du genre *Methanosarcina*. De plus, ces variants présentent de nombreuses différences entre eux ; ainsi le variant a se distingue par l'absence de sérine et un taux élevé de lysine, le variant b est caractérisé par la présence d'un résidu de tyrosine et le variant c présente un taux élevé d'acides aminés dicarboxyliques. Des différences significatives de structures primaires doivent être attendues.

II - 3-4 CARTES PEPTIDIQUES DE LA PROTEINE MCl DE DIFFÉRENTES SOUCHES DU GENRE *METHANOSARCINA*

L'étude de la variabilité structurale des protéines MCl de bactéries thermophiles étroitement apparentées du point de vue phénotypique et génotypique a été effectuée par comparaison des cartes peptidiques obtenues par CLHP en phase inverse. Cette technique sensible, reproductible et non destructive permet un repérage rapide et une identification sans ambiguïté des séquences variables.

Les cartes peptidiques obtenues après coupure des liaisons aspartyles des protéines MCl des quatre souches du genre *Methanosarcina* sont présentées dans la figure 11. Il apparaît clairement que les cartes peptidiques des protéines MCl de *Methanosarcina* sp CHTI 55 et de *Methanosarcina thermophila* TM1 sont rigoureusement superposables. De plus, les compositions

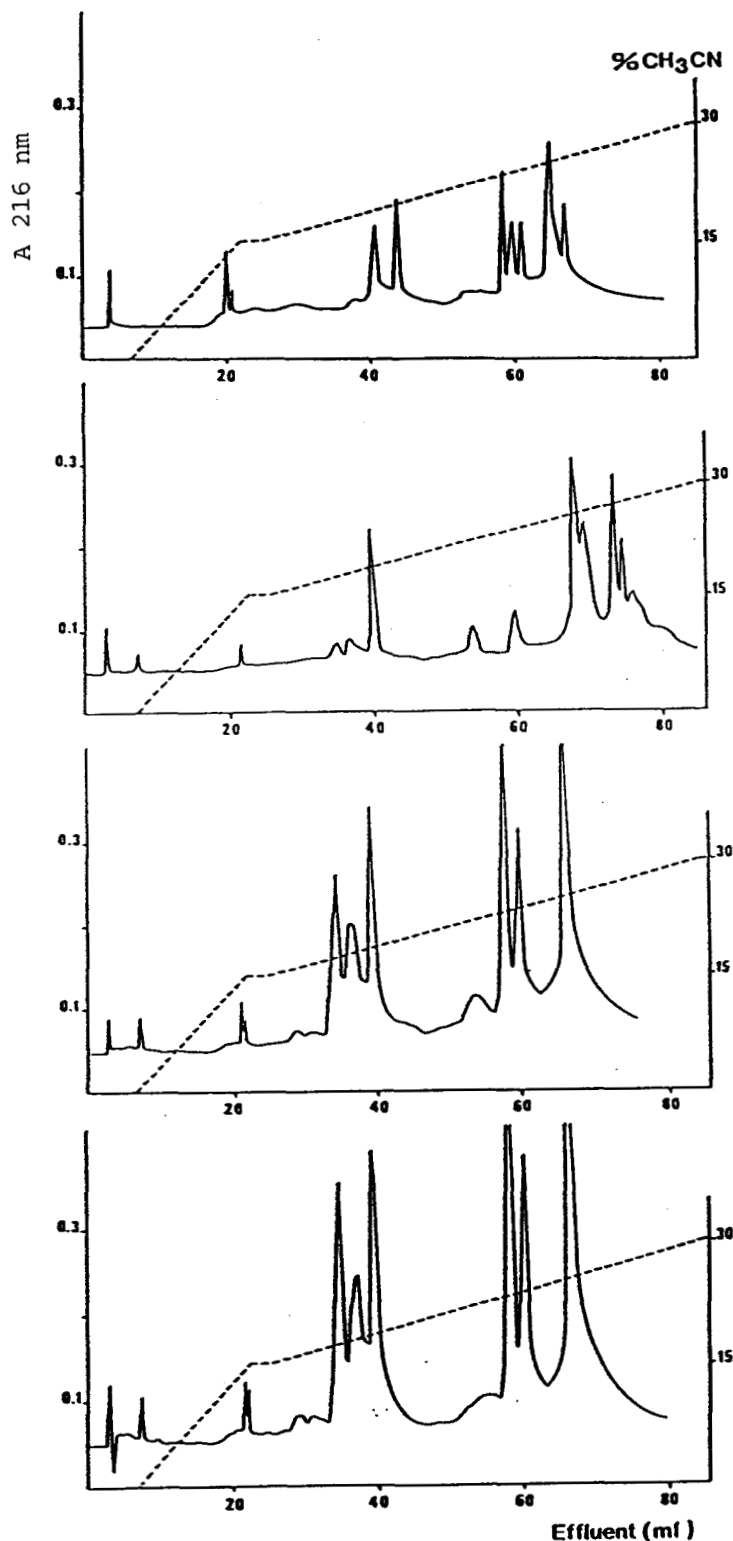


FIGURE 11 - Cartes peptidiques obtenues par coupure des liaisons aspartyles de la protéine MCl de souches du genre *Methanosarcina*.

Les peptides sont séparés sur une colonne de phase inverse C₁₈ éluée par un gradient d'acétonitrile dans le TFA 0,05 %. Les peptides ont été obtenus à partir de la protéine MCl de :

- A : *Methanosarcina barkeri*
- B : *Methanosarcina* MST-A1
- C : *Methanosarcina* sp CHTI 55
- D : *Methanosarcina thermophila* TM1

Les conditions de coupure chimique et de chromatographie sont décrites dans la publication n°4.

en acides aminés de ces peptides sont totalement identiques. Nous pouvons conclure que les résidus d'acide aspartique occupent les mêmes positions dans les séquences de ces deux protéines et que ces séquences sont vraisemblablement identiques. Par conséquent, seule la structure primaire de la protéine MC1 de *Methanosarcina* sp CHTI 55 a été complètement déterminée. Les cartes peptidiques obtenues avec les protéines MC1 de *Methanosarcina* MST-A1 et de *Methanosarcina barkeri* sont légèrement différentes de celles citées précédemment. Ces différences rendent compte des variations structurales observées lors de la détermination de la séquence en acides aminés de ces protéines comme nous le verrons dans le chapitre III.

II - 4 CONCLUSION DES ÉTUDES DE CARACTÉRISATION DE LA PROTÉINE MC1

Nous avons isolé et caractérisé dans chacune des trois souches du genre *Methanosarcina* une protéine MC1. Les protéines MC1 de *Methanosarcina* sp CHTI 55 et de *Methanosarcina thermophila* TM1 sont très étroitement apparentées et probablement identiques, ce qui est en accord avec les données phénotypiques et génotypiques observées par J.P.TOUZEL. Ces deux dernières protéines sont légèrement différentes de la protéine MC1 de *Methanosarcina barkeri*. Chez *Methanotherix soehngenii*, nous avons mis en évidence la présence de trois variants a, b et c de la protéine MC1. Le variant a se distingue des autres variants par sa richesse en acides aminés basiques et le variant b par la présence d'un résidu de tyrosine.

CHAPITRE III

STRUCTURE PRIMAIRE DES PROTÉINES MCI

III - 1 STRATÉGIES

Dans un premier temps, chaque protéine a été soumise à la dégradation de Edman automatisée. En général, nous avons pu ainsi établir environ la moitié de la séquence de la protéine (de 44 résidus pour la protéine MCl de *Methanosarcina barkeri* jusqu'à 63 résidus pour la protéine MClb de *Methanotherrix soehngeni*). Nous avons ensuite cherché à obtenir un nombre restreint de longs fragments peptidiques faciles à isoler avec de bons rendements. La protéine a été coupée spécifiquement par des méthodes chimiques et enzymatiques.

Pour la protéine MCl de *Methanosarcina barkeri*, qui a été la première à être séquencée, le choix des coupures a été fait en fonction de sa composition en acides aminés. Ensuite la stratégie a évolué en fonction des résultats progressivement acquis. L'évolution de ces stratégies est présentée dans la figure 12.

La coupure des liaisons aspartyles par l'acide acétique 0,25 M à 105°C qui s'était révélée très intéressante pour la détermination de la structure primaire de la protéine MCl de *Methanosarcina barkeri* et de *Methanosarcina* sp CHTI 55, a été abandonnée pour l'étude structurale des variants de protéine MCl de *Methanotherrix soehngeni* car ces protéines possèdent des résidus d'acide aspartique très proches de l'extrémité C-terminale. De même, l'utilisation de l'endoprotéinase Arg-C a été abandonnée en raison du manque de spécificité de cette enzyme. Nous avons en effet observé des coupures au niveau des résidus de lysine, et de nombreuses coupures partielles au niveau des résidus d'arginine dans la région centrale de la protéine, ce qui rend l'hydrolysate complexe, et réduit le rendement des peptides. Par contre, la coupure au niveau des résidus de méthionine par le bromure de cyanogène s'est

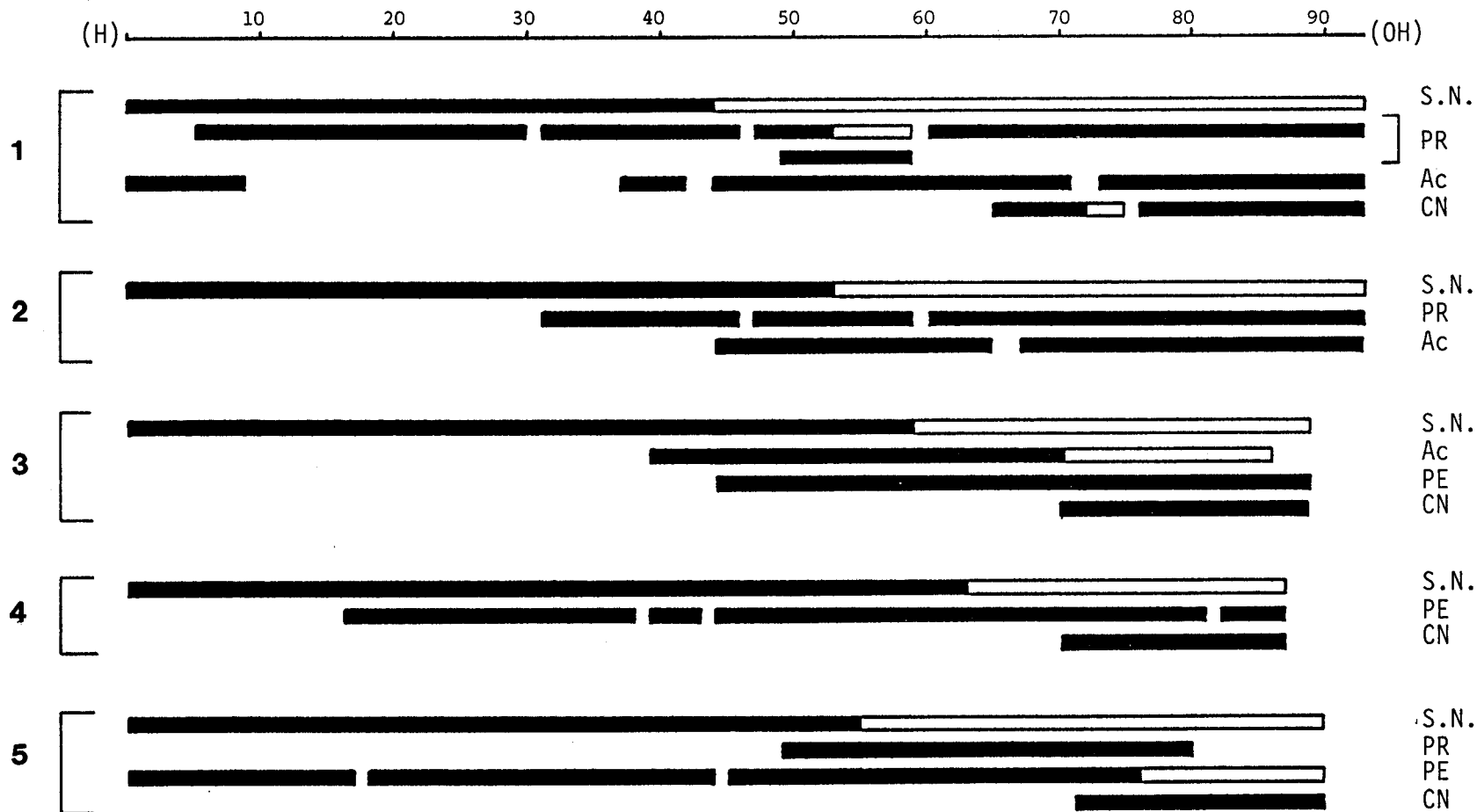


FIGURE 12 - Stratégies adoptées pour la détermination de la séquence en acides aminés de la protéine MCl.

S.N. : séquence N-terminale de la protéine MCl établie par microséquençage automatique (dégradation de Edman).

Peptides obtenus par hydrolyse de la protéine MCl avec l'acide acétique 0.25M (Ac), avec le bromure de cyanogène (CN), avec l'endoprotéinase Arg-C (PR), avec l'endoprotéinase Glu-C (PE). Les rectangles noirs représentent les séquences déterminées dans les fragments peptidiques. La protéine MCl a été isolée de : 1, *Methanosarcina barkeri*; 2, *Methanosarcina* sp CHTI 55 ; 3, 4 et 5 variants MCl a, b et c de *Methanotherix soehngeni* respectivement.

avérée fort utile pour l'obtention du peptide C-terminal des variants de la protéine MC1 de *Methanothrix soehngenii*. La structure de ce peptide d'environ 20 résidus a pu être complètement déterminée par microséquençage. Enfin, l'utilisation de l'endoprotéinase Glu-C de *Staphylococcus aureus* a donné des résultats particulièrement intéressants malgré la présence de nombreux résidus d'acide glutamique dans la protéine MC1. En fait, plusieurs liaisons glutamyles sont résistantes à l'action de cette enzyme tandis qu'au niveau des autres liaisons glutamyles la coupure est totale. Ces observations sont discutées dans la publication n°6 (voir sous chapitre III-2).

III - 2 RÉSULTATS ET DISCUSSION

III - 2-1 CONTRIBUTION PERSONNELLE

Notre contribution personnelle à l'élucidation de la structure primaire de la protéine MC1 de *Methanosarcina barkeri* se situe dans la détermination de la séquence aminoterminal de la protéine (résidus 1-44) et dans l'étude des peptides obtenus par coupure chimique des liaisons aspartyles et des liaisons méthionyles.

Ces travaux ont fait l'objet de la publication n°4 :

Primary structure of the chromosomal protein HMB from the archaeobacterium *Methanosarcina barkeri*.
LAINE, B., CHARTIER, F., IMBERT, M., LEWIS, R. and SAUTIERE, P. (1986)
Eur.J.Biochem. 161, 681-687.

Eur. J. Biochem. 161, 681–687 (1986)
© FEBS 1986

Primary structure of the chromosomal protein HMb from the archaeobacteria *Methanosarcina barkeri*

Bernard LAINE¹, François CHARTIER¹, Marlène IMBERT¹, Randy LEWIS² and Pierre SAUTIERE¹

¹ Unité Associée 409 du Centre National de la Recherche Scientifique, Lille

² Department of Biochemistry, University of Wyoming, Laramie

(Received July 2/September 5, 1986) — EJB 86 0703

The amino acid sequence of the protein HMb, a protein of 93 residues (M_r 10757) which represents the major acid-soluble component of the *Methanosarcina barkeri* nucleoprotein complex, has been established from automated sequence analysis of the protein and from structural data provided by peptides derived from cleavage of the protein at aspartic acid, arginine and methionine residues.

The protein HMb is mainly characterized by a high amount of charged residues (15% of acidic residues and 26.8% of basic residues) which are distributed all along the polypeptide chain. The amino acid sequence of the protein HMb is not homologous to any eubacterial, archaeobacterial or eukaryotic chromosomal proteins known up to now.

Methanogens constitute with thermoacidophiles and halophiles a group of prokaryotic organisms called archaeobacteria [1]. Although some macromolecules in archaeobacteria exhibit physico-chemical properties similar to those found in eubacteria or in eukaryotes, others such as membrane lipids, cell wall peptidoglycans or enzymes have unique structural features which place archaeobacteria in a distinct line of descent [1, 2]. The fundamental differences observed in the packaging of the DNA between eubacteria and eukaryotes result from the nature of their chromosomal proteins. Packaging of the DNA inside the nucleus of eukaryotic cells is mainly ensured by histones whereas in eubacteria, HU-type proteins (M_r 9500) are involved in DNA condensation [3–5]. It is therefore of great interest to investigate the features of chromosomal proteins from archaeobacteria. Chromosomal proteins have been isolated from various archaeobacteria: three strains of thermoacidophiles, *Thermoplasma acidophilum* [6], *Sulfolobus acidocaldarius* [7] and *Sulfolobus solfataricus* [8] and one strain of methanogens *Methanosarcina barkeri* [9]. Only two of these proteins have been sequenced: on the criteria of their primary structure, the protein HTa from *T. acidophilum* [10] is homologous to the HU-type proteins [11] whereas the protein (M_r 7000) from *S. solfataricus* [8] is related neither to histones nor to HU-type proteins.

The protein HMb is a low-molecular-mass polypeptide (M_r 10757) which has been shown to be the major acid-soluble component of the *Methanosarcina barkeri* nucleoprotein complex [9]. We report here the complete amino acid sequence of this protein. The primary structure of protein HMb is not homologous to any another chromosomal protein from prokaryotes or eukaryotes.

Correspondence to B. Laine, Unité Associée CNRS 409, Place de Verdun, F-59045 Lille Cedex, France

Abbreviations. SDS, sodium dodecyl sulfate; HPLC, high-pressure liquid chromatography.

Enzymes. Carboxypeptidase A (EC 3.4.17.1); carboxypeptidase B (EC 3.4.17.2); carboxypeptidase Y (EC 3.4.16.1); endoproteinase Arg-C (EC 3.4. —. —).

MATERIALS AND METHODS

Materials

Methanosarcina barkeri (strain DSM 800) was grown as indicated in [9]. The protein HMb was isolated as in [9]. Endoproteinase Arg-C from mouse submaxillary gland and carboxypeptidase Y were purchased from Boehringer. Carboxypeptidase A treated with phenylmethylsulfonyl fluoride and carboxypeptidase B treated with diisopropyl-fluorophosphate were from Sigma. Cyanogen bromide was from Eastman Kodak (Rochester, USA). Acetonitrile for reverse-phase high-pressure liquid chromatography was from Carlo Erba. Reagents and solvents for the liquid-phase sequencer were as indicated in [12]. Reagents and solvents for the gas-phase sequencer were from Applied Biosystems.

Cleavage at aspartyl residues

Protein HMb (200 nmol) was hydrolysed with 1 ml of 0.25 M acetic acid in an evacuated sealed tube for 15 h at 105°C. The hydrolysate was freeze-dried prior to fractionation.

Cleavage with the endoproteinase Arg-C

Protein HMb (300 nmol) was dissolved in 2.2 ml of 0.1 M sodium bicarbonate pH 7.9 and digested with endoproteinase Arg-C at 37°C for 4 h. Enzyme was added at the beginning of the reaction and after 2 h; the final enzyme/substrate ratio was 1/50 (w/w). The digestion was stopped by lowering the pH with formic acid. The hydrolysate was freeze-dried prior to fractionation.

Cleavage at methionyl residues

Protein HMb (1 μ mol) was dissolved in 4 ml of 70% (v/v) formic acid. Cyanogen bromide was added in a 100-fold molar excess over one methionine residue. The reaction mixture was

flushed with N₂ and incubated for 24 h under gentle stirring at 37°C in the dark. The reaction mixture was then evaporated in a vacuum concentrator (Savant Instruments) to 0.7 ml; the concentration was repeated twice after washing with 1 ml water. The peptide solution was made 6 M in guanidine hydrochloride and submitted to fractionation.

Fractionation of peptides

Peptides generated by cleavage with 0.25 M acetic acid or with the endoproteinase Arg-C were fractionated by reverse-phase high-pressure liquid chromatography as in [11], on a C₁₈ µBondapak column (Waters Associates) and using gradient systems indicated in Fig. 1.

Peptides generated by cleavage with cyanogen bromide were fractionated by chromatography on a Sephadex G-50 superfine column (95 × 1.6 cm) equilibrated and eluted with 0.01 M HCl saturated with chloroform.

Amino acid analyses

Peptide samples were hydrolysed in evacuated sealed tubes at 110°C for 24 h in 6 M HCl.

The hydrolysates of peptides generated by cleavage with cyanogen bromide were treated with a pyridine acetate buffer (pyridine/glacial acetic acid/water, 25:1:224; pH 6.5) at 105°C for 1 h in order to convert homoserine lactone to homoserine.

Amino acid analyses were performed on a Beckman 119 CL amino acid analyser, using a single-column system. The column was packed with a Durrum DC 6A resin.

Carboxy-terminal analyses

The carboxy-terminal sequence of protein HMB and of the carboxy-terminal peptide obtained by cleavage at methionine residues (peptide CN-3) were determined simultaneously. Each substrate (28 nmol) was successively digested with carboxypeptidases Y, A and B. After every step, an 8-nmol aliquot of each digest was acidified with formic acid and immediately freeze-dried. The released amino acids were analysed on the amino acid analyser. Digestion by carboxypeptidase Y (56 µg) was performed for 7 h at 37°C in 0.1 M acetate pyridine buffer (0.35 ml) adjusted at pH 5.5. The carboxypeptidase Y was then denatured in a boiling water bath for 2 min and the buffer was removed by evaporation. The dry residue was then dissolved in 0.2 M *N*-methylmorpholine acetate pH 8.0 (0.25 ml) and the digestion by carboxypeptidase A (11 µg; 0.55 unit) was allowed to proceed at 37°C for 0.5 h. Carboxypeptidase B (1 µg; 0.14 unit) was then added and the digestion was carried on for 1 h at 37°C.

Peptides derived by cleavage of the protein with endoproteinase Arg-C were digested successively by carboxypeptidase B (0.6 µg; 0.08 unit) for 15 min and by carboxypeptidase A (3.6 µg; 0.18 unit) for 15 min. Peptides (12 nmol) were dissolved in 0.2 M *N*-methylmorpholine acetate pH 8.0 (0.24 ml) and the digestion was performed at 37°C. Aliquots (5 nmol) were treated as above for amino acid analyses.

Sequence determination

Automated Edman degradation of the protein HMB and of peptides was carried out on a gas-phase sequencer Applied Biosystems using the 02 N VAC Program.

Table 1 Edman degradation of protein HMB

The Edman degradation of protein HMB was performed on the gas-phase sequencer using the 02 N VAC program. A single run was carried out with 12.9 nmol of protein. Phenylthiohydantoin derivatives of amino acids were identified and quantified by HPLC. The repetitive yield calculated from valine residues at positions 7 and 18 was 93 %.

Cycle	Identified residue	Yield (nmol)	Cycle	Identified residue	Yield (nmol)
1	Ser	3.50	25	Arg	0.17
2	Asn	7.95	26	Gln	0.21
3	Thr	4.31	27	Ala	0.28
4	Arg	1.81	28	Ala	0.41
5	Asn	5.69	29	Leu	0.35
6	Phe	10.99	30	Lys	0.31
7	Val	6.69	31	Ala	0.37
8	Leu	8.97	32	Ala	0.41
9	Arg	2.75	33	Asn	0.20
10	Asp	4.10	34	Arg	0.10
11	Glu	4.41	35	Gly	0.16
12	Glu	4.28	36	Asp	0.09
13	Gly	3.19	37	Gly	0.12
14	Asn	5.08	38	Thr	0.03
15	Glu	4.13	39	Lys	-
16	His	1.41	40	Ser	-
17	Gly	2.28	41	Asn	0.05
18	Val	2.94	42	Pro	0.14
19	Phe	4.07	43	Asp	0.04
20	Thr	1.13	44	Val	0.07
21	Gly	1.66	45	nd	-
22	Lys	0.58	46	nd	-
23	Gln	0.30	47	Leu	0.25
24	Pro	0.31			

nd : not determined.

Alternatively automated Edman degradation of large peptides was performed on a liquid-phase Beckman 890 C sequencer using the 0.33 M quadrol program (11 1978) in the presence of polybrene [13]. Manual Edman degradation of peptides was carried out as in [14]. The phenylthiohydantoin derivatives of amino acids were identified by high-pressure liquid chromatography as indicated in [12].

Nomenclature of peptides

Peptides derived from cleavage at aspartyl and methionyl residues are designated by the symbols Ac- and CN-, respectively, whereas peptides generated by digestion with the endoproteinase Arg-C were designated by PR-. Peptides were numbered according to their position in the amino acid sequence of the protein.

RESULTS

The sequence of the amino-terminal half of the protein HMB (residues 1–44) was determined by automated Edman degradation of the protein (Table 1). The remainder of the sequence was elucidated from data provided by large adjacent or overlapping peptides generated by highly specific cleavage of the protein.

Cleavage at aspartic acid residues

Peptides derived from hydrolysis of protein HMB with 0.25 M acetic acid were fractionated by reverse-phase high pressure liquid chromatography (Fig. 1 A). Five peptides were obtained directly in pure form. Their amino acid composition are given in Table 2. Due to partial deamidation of glutamyl residues, peptide Ac-2 was eluted in fractions 2 and 3.

Only the large peptides Ac-4 and Ac-5 which overlap the carboxy-terminal half of the protein (residues 44–93) were completely sequenced by automated Edman degradation performed with the gas-phase sequencer (Table 3).

Cleavage by the endoproteinase Arg-C

Peptides generated by cleavage of the protein HMB with endoproteinase Arg-C were fractionated by reverse-phase high-pressure liquid chromatography (Fig. 1 B). Each of the nine fractions obtained corresponds to a pure peptide. Only

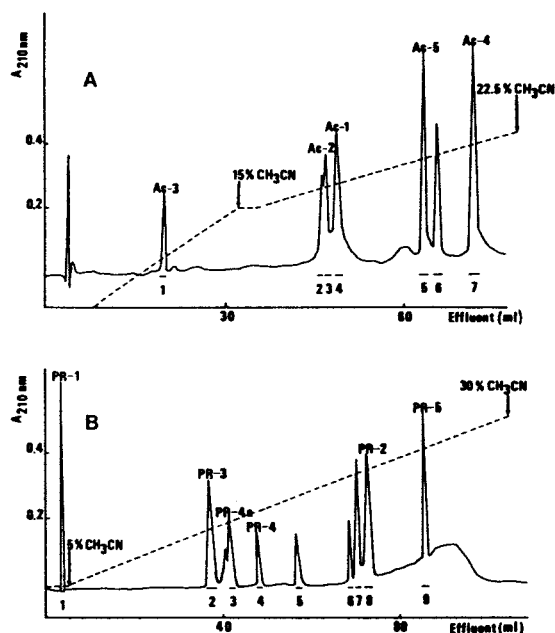


Fig. 1. Fractionations of peptides by reverse-phase HPLC. Chromatography was performed on a C_{18} μ Bondapak column (30×0.39 cm); (---) gradient of acetonitrile in 0.05% trifluoroacetic acid; (—) absorbance at 210 nm. Fractions were numbered by order of elution. Peptides used for sequence determination were numbered according to their position in the sequence of the protein. (A) Peptides derived by cleavage of protein HMB at aspartyl residues. 40 nmol of hydrolysate were loaded; flow rate was 1.5 ml/min; fractions of 0.75 ml were collected. (B) Peptides generated by hydrolysis of protein HMB with the endoproteinase Arg-C. 65 nmol of digest were loaded; flow rate was 2.0 ml/min; fractions of 0.8 ml were collected.

the amino acid compositions of peptides PR-1, PR-2, PR-3, PR-4, PR-4a and PR-5, which account for the entire sequence of the protein HMB, are given in Table 2.

Data obtained by automated Edman degradation performed on the gas-phase sequencer (for peptide PR-5) and on the liquid-phase sequencer (for peptides PR-2 and PR-4a) are given in Tables 3 and 4 respectively.

Besides the usual lack of cleavage of the Arg-Pro bond at position 71-72, the endoproteinase Arg-C did not cleave either the bond Arg-Asp at position 9-10 or the two Arg-Gly bonds at positions 34-35 and 50-51. The cluster Asp¹⁰-Glu-Glu¹² makes the environment of the susceptible bond Arg-Asp too acidic and the activity of the endoproteinase Arg-C may be hampered by the conformation of the polypeptide chain which has a high potential for β -turn structure at these glycine and proline residues. On the other hand, unexpected cleavage was observed at two Lys-Ala bonds at positions 30-31 and 59-60; the hydrolysis of both bonds was complete.

Half of the protein molecules were cleaved at the Arg⁴-Asn⁵ bond but only 16% of the molecules were cleaved at the Arg²⁵-Gln²⁶ bond; therefore, in addition to peptides PR-1 (residues 1-4) and PR-2 (residues 5-30), two peptides covering the sequences 1-30 and 1-25 were obtained in fractions 7 and 6 respectively. The bonds Arg-Leu at position 46-47 and Arg-Glu at position 48-49 were partially cleaved. Consequently, in addition to peptides PR-3 (residues 31-46) and PR-4 (residues 47-59), two peptides covering the sequences 31-48 and 49-59 (peptide PR-4a) were obtained with a yield of 8% and 10% respectively; the former was eluted in fraction 5 and the latter was eluted in fraction 3.

Cleavage at methionyl residues

Fragments derived from cleavage of the protein HMB at methionyl residues with cyanogen bromide, were separated by gel-filtration chromatography on a Sephadex G-50 superfine

Table 2 Amino acid compositions of peptides derived from protein HMB

	Ac-1 Residues 1-9	Ac-2 Residues 11-35	Ac-3 Residues 37-42	Ac-4 Residues 44-72	Ac-5 Residues 74-93	PR-1 Residues 1-4	PR-2 Residues 5-30	PR-3 Residues 31-46	PR-4 Residues 47-59	PR-4a Residues 49-59	PR-5 Residues 60-93	CN-1 Residues 1-64	CN-2 Residues 65-75	CN-3 Residues 76-93	Protein HMB
Aspartic acid	2.0 (2)	2.1 (2)	1.0 (1)	1.5 (1)		1.0 (1)	3.2 (3)	3.9 (4)			1.8 (2)	7.6 (8)	1.8 (2)		9.6 (10)
Threonine	1.0 (1)	1.1 (1)	1.0 (1)	0.8 (1)		1.1 (1)	1.1 (1)	1.0 (1)	1.0 (1)	0.8 (1)		3.6 (4)			3.8 (4)
Serine	0.8 (1)		1.1 (1)		1.3 (1)	0.9 (1)		1.0 (1)			1.0 (1)	1.8 (2)		0.9 (1)	2.9 (3)
Glutamic acid		4.7 (5)		3.3 (3)	4.2 (4)		5.0 (5)		1.1 (1)	1.0 (1)	6.0 (6)	7.3 (7)	1.3 (1)	4.2 (4)	12.1 (12)
Proline		1.1 (1)	1.0 (1)	1.8 (2)	2.1 (2)		1.0 (1)	1.0 (1)			4.0 (4)	2.4 (2)	1.6 (2)	1.8 (2)	6.0 (6)
Glycine		3.4 (3)	1.1 (1)	1.8 (1)			3.0 (3)	2.2 (2)	1.1 (1)	1.2 (1)		5.7 (6)			6.3 (6)
Alanine		3.6 (4)		2.1 (2)			2.0 (2)	2.1 (2)			2.0 (2)	5.0 (5)	1.0 (1)		6.0 (6)
Valine	1.1 (1)	1.2 (1)		4.1 ^a (4)	1.4 (1)		1.9 (2)	1.0 (1)	1.7 (2)	1.9 (2)	2.0 (2)	4.7 (5)	1.0 (1)	1.3 (1)	6.5 (7)
Methionine				0.8 (1)	1.2 (1)						2.0 (2)	1.1 ^c (1)	1.0 ^c (1)	0.0 ^c (0)	1.9 (2)
Isoleucine				1.2 ^a (1)	2.6 ^a (3)			0.8 (1)			2.9 (3)	0.9 (1)		2.8 (3)	3.9 (4)
Leucine	1.0 (1)	1.1 (1)		1.3 (1)		1.9 (2)			1.0 (1)			3.2 (3)			3.2 (3)
Phenylalanine	1.1 (1)	1.0 (1)		1.0 (1)	1.2 (1)		1.8 (2)		0.9 (1)	1.0 (1)	1.1 (1)	2.9 (3)		1.1 (1)	3.9 (4)
Histidine		0.9 (1)		1.0 (1)			0.9 (1)		0.8 (1)	1.0 (1)		1.6 (2)			2.0 (2)
Lysine		1.9 (2)	0.9 (1)	5.0 (5)	5.7 (6)		1.8 (2)	0.7 (1)	3.0 (3)	3.2 (3)	8.2 (8)	7.2 (7)	1.3 (1)	5.8 (6)	14.4 (14)
Arginine	2.1 (2)	1.8 (2)		3.6 (4)		0.9 (1)	2.4 (2)	2.0 (2)	2.0 (2)	0.9 (1)	1.1 (1)	7.1 (7)	1.0 (1)		8.4 (8)
Tryptophan				1 ^b (1)	1 ^b (1)						2 ^b (2)	1 ^b (1)	1 ^b (1)		1.9 (2)
TOTAL	9	24	6	29	20	4	26	16	13	11	34	64	11	18	93
Fraction Number	4	2 + 3	1	7	5	1	8	2	4	3	9	1	3	2	
% yield of recovery	49	23	36,7	19,8	38,5	47	32	51,5	16	10	22,5	36	52	61	

Values are expressed as number of residues per mole of peptide. The numbers of residues in parentheses are from the established sequence. No correction was made for degradative losses of threonine and serine during hydrolysis of peptides. a, 72 h hydrolysis values. b, determined spectrophotometrically. c, determined as homoserine.

Table 3 Edman degradation of protein HMb peptides performed on the gas-phase sequencer.

The runs were performed once for each peptide using the D2 N VAC program.

Phenylthiohydantoin derivatives of amino acids were identified and quantified by HPLC.

Peptides	Ac-4	Ac-5	PR-5	CN-3				
Quantity of material subjected to degradation	6 nmol	3 nmol	6 nmol	6 nmol				
Repetitive yield (%)	89.7	89.1	91.7	89.9				
Cycle	Identified residue	Yield (nmol)	Identified residue	Yield (nmol)	Identified residue	Yield (nmol)	Identified residue	Yield (nmol)
1	Val	5.00	Trp	0.81	Ala	5.00	Pro	2.99
2	Ile	4.90	Met	2.00	Trp	3.17	Glu	4.50
3	Arg	3.91	Pro	1.40	Lys	5.00	Lys	3.09
4	Leu	7.21	Glu	1.17	Glu	6.00	Ile	2.93
5	Arg	3.87	Lys	2.00	Met	5.00	Ser	0.45
6	Glu	3.80	Ile	0.76	Val	5.00	Lys	2.90
7	Arg	3.60	Ser	0.53	Glu	4.00	Pro	2.56
8	Gly	2.61	Lys	1.00	Ala	4.00	Phe	2.18
9	Thr	3.27	Pro	0.69	Pro	4.00	Val	1.36
10	Lys	3.07	Phe	0.74	Lys	4.00	Lys	1.43
11	Lys	4.19	Val	1.00	Asn	4.00	Lys	1.88
12	Val	2.55	Lys	1.00	Arg	2.44	Glu	1.74
13	His	1.45	Lys	1.00	Pro	3.00	Lys	1.88
14	Val	1.67	Glu	0.37	Asp	3.00	Ile	1.01
15	Phe	2.19	Lys	0.11	Trp	1.02	Glu	2.73
16	Lys	2.45	Ile	0.24	Met	3.00	Lys	1.15
17	Ala	2.55	Glu	0.36	Pro	3.00	Ile	0.78
18	Trp	0.17	Lys	0.16	Glu	2.06		
19	Lys	2.01	Ile	0.21	Lys	3.00		
20	Glu	1.07			Ile	1.64		
21	Met	1.00			Ser	0.23		
22	Val	0.86			Lys	1.00		
23	Glu	0.93			Pro	2.00		
24	Ala	0.88			Phe	1.01		
25	Pro	0.73			Val	0.72		
26	Lys	0.44			Lys	1.00		
27	Asn	0.32			Lys	1.00		
28	Arg	0.21			Glu	0.99		
29					Lys	1.00		
30					Ile	0.69		
31					Glu	0.93		
32					Lys	1.00		
33					Ile	0.66		

column (Fig. 2). Their amino acid compositions are presented in Table 2.

The lack of homoserine in peptide CN-3 and the data showing that peptide CN-3 and protein HMb have the same carboxy-terminal sequence determined as Ile-Glu-OH (Table 5) led us to assign the peptide CN-3 to the carboxy-terminal position in the protein. The complete sequence of peptide CN-3 was determined by automated Edman degradation performed on the gas-phase sequencer (Table 3).

Complete amino acid sequence of the protein HMb

The complete amino acid sequence of the protein HMb is presented in Fig. 3. The sequence of the amino-terminal half of the protein HMb (residues 1–44) was determined by automated Edman degradation of the protein. This sequence was confirmed by the data provided by peptides Ac-1, Ac-3,

PR-1, PR-2 and PR-3. The sequence of the carboxy-terminal half of the protein HMb (residues 44–93) was unambiguously established from sequence analyses of peptides Ac-4, Ac-PR-4, PR-5, CN-2 and CN-3. Alignment of fragments Ac (residues 37–42) and Ac-4 (residues 44–72) was provided by the overlapping peptide PR-3 (residues 31–46). Alignment of fragments Ac-4 and Ac-5 (residues 74–93) was deduced from sequence analysis of peptide PR-5 (residues 60–93). The assignment of peptides Ac-5 and PR-5 to the carboxy-terminal position in protein HMb was supported by the sequence analyses of fragment CN-3.

The sequence data led us to discover that the apparent molecular size of protein HMb determined by SDS/polyacrylamide gel electrophoresis was overestimated: the protein HMb migrates as a protein of M_r 14 500 which corresponds to a polypeptide chain of 123 amino acid residues [9]. But all the sets of peptides generated by chemical or enzymatic

cleavages of the protein HMB (including a cleavage at tryptophan residues, data not shown) we never found peptides which might overlap a sequence beyond the residue at position 93. This was confirmed by determining the carboxy-terminal sequence of the protein HMB. The relative molecular mass of the protein HMB calculated from the sequence data is 10757. The abnormal electrophoretic migration of protein HMB in SDS/polyacrylamide gel is due to its high content of basic amino acids. Such an abnormal electrophoretic behaviour was also observed for calf thymus histone H4: protein HMB and histone H4 (calculated M_r 11280) exhibit the same migration which is slightly slower than that of α -lactalbumin (M_r 14400) (Fig. 4).

DISCUSSION

The protein HMB is mainly characterized by a high amount of charged residues (15% of acidic residues and 26.8% of basic residues) which are distributed all along the polypeptide chain (Fig. 5). However, three sequences enriched in charged residues are obvious in the molecule: an acidic sequence (residues 10–15), a basic sequence (residues 46–

56) and the carboxy-terminal sequence (residues 85–93) where the four positive charges are balanced by the negative charges given by three glutamic acid residues and the free α -carboxyl group.

Most of the hydrophobic residues are associated in doublets or triplets. They are probably involved in strong hydrophobic interactions which might account for the high tendency of the protein HMB to aggregate during its purification process. The four residues of proline at positions 68, 72, 76 and 82 give to this region of the protein a highly folded

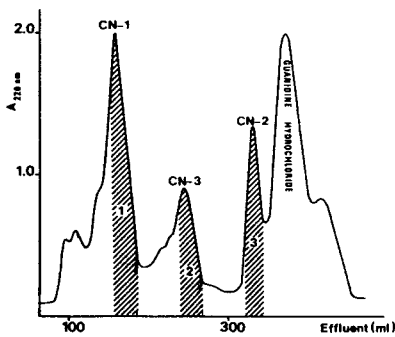


Fig. 2. Fractionation by gel-filtration chromatography of peptides derived from cleavage at methionyl residues. Peptides (0.5 μ mol in 0.7 ml of 6 M guanidine hydrochloride, 0.01 M HCl) were loaded on a Sephadex G-50 superfine column (90 \times 1.6 cm). The column was equilibrated and eluted with 0.01 M HCl saturated with chloroform. 1.3-ml fractions were collected at a flow rate of 4.0 ml/h. (—) Absorbance at 220 nm

Table 4. Edman degradation of protein HMB peptides performed on the liquid-phase sequencer.

The runs were performed once for each peptide using the 0.33 M quadrol program (11.1978) in the presence of polybrene. Phenylthiohydantoin derivatives of amino acids were identified and quantified by HPLC and when necessary by gas chromatography indicated by GC.

Peptide	PR-2	PR-4a		
Quantity of material subjected to degradation	65 nmol	50 nmol		
Repetitive yield (%)	77.7	82.8		
Cycle	Residue	Yield (nmol)	Residue	Yield (nmol)
1	Asn	17.9	Glu	50
2	Phe	37.1 GC	Arg	15.9
3	Val	29.4 GC	Gly	17.2
4	Leu	23.3	Thr	2.7
5	Arg	14.5	Lys	8.3
6	Asp	19.5	Lys	8.3
7	Glu	1.5	Val	7.3 GC
8	Glu	1.7	His	3.8
9	Gly	6.0	Val	5.0 GC
10	Asn	0.8	Phe	3.4 GC
11	Glu	2.0	Lys	2.4
12	His	-	-	-
13	Gly	2.8	-	-
14	Val	0.6 GC	-	-
15	Phe	1.4 GC	-	-
16	Thr	-	-	-
17	Gly	0.4	-	-
18	Lys	-	-	-
19	Gln	-	-	-
20	Pro	-	-	-
21	Arg	-	-	-
22	nd	-	-	-
23	Ala	-	-	-
24	Ala	-	-	-

nd : not determined.

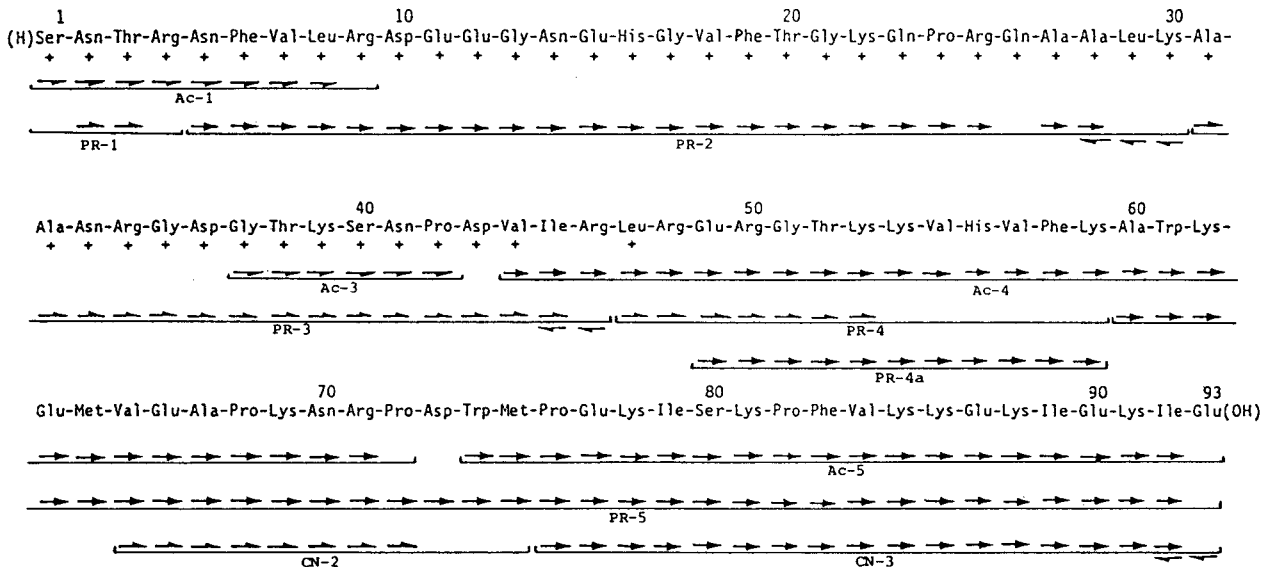


Fig. 3. Complete amino acid sequence of the *Methanosarcina barkeri* protein HMB. Methods used for determination of the sequence are indicated as follows: +++ automated Edman degradation of the protein; -> automated Edman degradation of peptides; <- manual Edman degradation of peptides; <- carboxypeptidase hydrolysis

conformation where β turns are placed side by side as predicted by the statistical method of Fasman et al. [15].

The protein HMB does not show any amino acid sequence homology either to eubacterial or to archaeobacterial chromosomal proteins. With an amount of basic residues equal to 27% and a net charge of +10, the protein HMB is much more basic than eubacterial HU-type proteins: the *Escherichia coli* HU protein contains 16.7% basic residues and has a net charge of +4. Moreover the distribution of basic residues in the two types of proteins is quite different; in HU-type proteins which lack highly basic sequences, the basic residues are scattered along the polypeptide chain (Fig. 5). The protein HTa, isolated from the archaeobacteria *T. acidophilum* [10], has an amino acid sequence clearly homologous to that of eubacterial HU-type proteins [11]. However a higher basicity

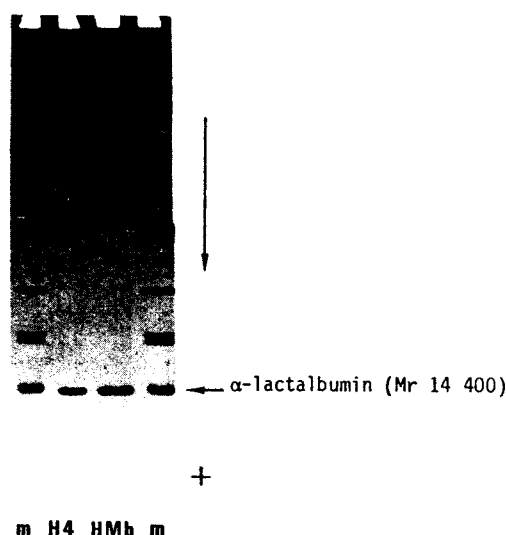


Fig. 4. SDS/polyacrylamide gel electrophoresis of *Methanosarcina barkeri* protein HMB and calf thymus histone H4. The M_r markers in lane m are: phosphorylase b (94000), bovine serum albumin (67000), ovalbumin (43000), carbonic anhydrase (30000), soybean trypsin inhibitor (20100), α -lactalbumin (14400). Samples (7 μ g) were treated with 5% SDS, 2% 2-mercaptoethanol for 2 min at 100°C and run at 40 mA for 150 min on a 5–30% polyacrylamide gradient gel [17]. Electrode buffer (Tris/glycine, pH 8.3) and gel buffer (Tris/HCl, pH 8.9) were made 0.1% in SDS. The gels were stained with 0.5% Coomassie blue R-250 in acetic acid/ethanol/water (1:2:7, v/v/v) for 15 h. The gel was destained by diffusion in acetic acid/ethanol/water (4:5:31, v/v/v)

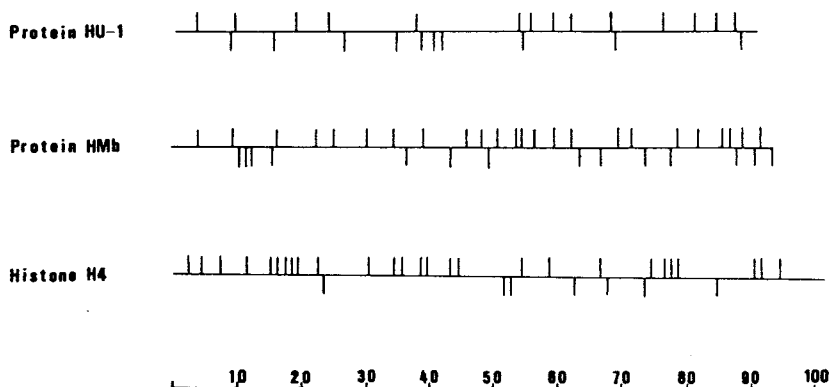


Fig. 5. Distribution of basic (\perp) and acidic (∇) residues in protein HU-1 (*Escherichia coli*) [18]; protein HMB (*Methanosarcina barkeri*) and histone H4 (*calf thymus*) [19]

(net charge equal to +13) and the presence of clusters of two or three basic residues, differentiate the protein HTa from eubacterial HU-type proteins. The DNA-binding protein isolated from *Sulfolobus solfataricus* [8] differs from protein HMB in size (M_r 7000) and in primary structure. Proteins of $M_r \approx 14500$ and exhibiting physico-chemical properties different from those of protein HMB have been characterized in *Methanosarcina thermoautotrophicum* [16] and in *Sulfolobus acidocaldarius* (protein HSa) [7]. Unlike protein HMB, the former is not acid-soluble and the protein HSa is distinct from the protein HMB particularly with respect to the amounts of lysine, arginine and leucine.

On the other hand, the protein HMB presents some analogies with the smallest core histone, the histone H4 in: (a) its molecular size; (b) its tendency to form aggregates and (c) the large number of basic residues, several of which are associated in doublets or in sequences where they are separated from each other by one or two non-basic residues. However the following differences between these two proteins have to be emphasized: (a) due to its high content in dicarboxylic amino acids, the protein HMB is less basic than histone H4 which has a net charge equal to +19; (b) the protein HMB contains two residues of tryptophan, an amino acid residue which is not encountered in any histones; (c) in histone H4, the highly basic character is located in the amino-terminal half of the polypeptide chain whereas an opposite distribution is encountered in protein HMB: two thirds of its basic residues are located in the carboxy-terminal half of the polypeptide (Fig. 5); (d) no significant amino acid sequence homology has been detected between the protein HMB and calf histone H4 by a computer-based survey.

Table 5 Determination of the carboxy-terminal sequences of the protein HMB and of peptide CN-3.

Results are expressed in number of residues released in the digests (see Methods).

Carboxypeptidase Y for 7 h	protein HMB peptide CN-3	Glu 1.0, Ile 0.6 Glu 0.9, Ile 0.0
Carboxypeptidase A for 0.5 h	protein HMB peptide CN-3	Glu 1.0, Ile 0.9 Glu 0.9, Ile 0.8

Circular dichroic spectroscopy indicates that the protein HMB is almost devoid of α -helix structure. We are now looking for the conformational state of the protein and the way it can organize the DNA in the archaeobacterial nucleoid. We also have to determine whether the protein HMB is able by itself to play this function or if it must be associated with another chromosomal protein, such as the 7-kDa protein found in the *Methanosarcina barkeri* deoxyribonucleoprotein complex [9].

The authors are grateful to Dr J. P. Touzel (Station de Technologie Alimentaire, INRA, Villeneuve d'Ascq) who provided the cultures of *Methanosarcina barkeri* and for valuable discussions. They are indebted to A. Lemaire, M. J. Dupire, and W. Kruggel for skilful technical assistance and to M. Lestiennes for editorial assistance. They also thank the *service commun de séquence Unités 409 CNRS/16 INSERM* for sequencer runs. This work was supported by grants from the *Centre National de la Recherche Scientifique*, from the *Pôle Régional Anaérobies* and from the *Fondation à la Recherche Médicale*.

REFERENCES

1. Fox, G. E., Stackebrandt, E., Hespell, R. B., Gibson, J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Balch, W. E., Tanner, R. S., Magrum, L. J., Zablen, L. B., Blakemore, R., Gupta, R., Bonem, L., Lewis, B. J., Stahl, D. A., Luehrsén, K. R., Chen, K. N. & Woese, C. R. (1980) *Science (Wash. DC)* **209**, 457–463.
2. Fox, G. E. (1985) in *The bacteria, a treatise on structure and function* (Woese, C. R. & Wolfe, R. S., eds) vol. VIII, pp. 257–301, Academic Press, New York.
3. Pettijohn, D. E. (1982) *Cell* **30**, 667–669.
4. Rouvière-Yaniv, J., Yaniv, M. & Germond, J. E. (1979) *Cell* **17**, 265–274.
5. Broyles, S. S. & Pettijohn, D. E. (1986) *J. Mol. Biol.* **187**, 47–60.
6. Searcy, D. G. (1975) *Biochim. Biophys. Acta* **395**, 535–547.
7. Green, G. R., Searcy, D. G. & De Lange, R. J. (1983) *Biochim. Biophys. Acta* **741**, 251–257.
8. Kimura, M., Kimura, J., Davie, P., Reinhardt, R. & Dijk, J. (1984) *FEBS Lett.* **176**, 176–178.
9. Chartier, F., Laine, B., Sautière, P., Touzel, J. P. & Albagnac, G. (1985) *FEBS Lett.* **183**, 119–123.
10. De Lange, R. J., Williams, L. C. & Searcy, D. G. (1981) *J. Biol. Chem.* **256**, 905–911.
11. Laine, B., Bélaïche, D., Khanaka, H. & Sautière, P. (1983) *Eur. J. Biochem.* **131**, 325–331.
12. Laine, B., Bélaïche, D., Sautière, P. & Biserte, G. (1982) *Biochem. Biophys. Res. Commun.* **106**, 101–107.
13. Tarr, G. E., Beecher, J. F., Bell, M. & McKeaw, D. (1978) *Anal. Biochem.* **84**, 622–627.
14. Edman, P. & Henschen, A. (1975) in *Protein sequence determination* (Needleman, S. B., ed.) 2nd edn, pp. 232–279, Springer-Verlag, Berlin, Heidelberg, New York.
15. Fasman, G. D., Chou, P. Y. & Adler, A. J. (1976) *Biophys. J.* **16**, 1201–1238.
16. Thomm, M., Stetter, K. O. & Zillig, W. (1982) *Zentralbl. Bakteriol. Hyg. I. Abt. Orig. C3*, 128–139.
17. Laemmli, U. K. (1970) *Nature (Lond.)* **227**, 680–685.
18. Laine, B., Kmičik, D., Sautière, P., Biserte, G. & Cohen-Solal, M. (1980) *Eur. J. Biochem.* **103**, 447–461.
19. De Lange, R. J., Fambrough, D. M., Smith, E. L. & Bonner, J. (1969) *J. Biol. Chem.* **244**, 319–334.

La détermination de la structure primaire de la protéine MC1 de *Methanosarcina* sp CHTI 55 et des trois variants de la protéine MC1 de *Methanotherix soehngeni* a fait l'objet des publications n°5 et 6.

PUBLICATION N°5

Primary structure of the chromosomal protein MC1 from the archaeobacterium *Methanosarcina* sp CHTI 55.
CHARTIER, F., LAINE, B., BELAÏCHE, D., TOUZEL, J.P. and SAUTIERE, P.
(1989)
Biochim. Biophys. Acta, sous presse.

PUBLICATION N°6

Primary structure of the chromosomal proteins MC1a, MC1b and MC1c from the archaeobacterium *Methanotherix soehngeni*.
CHARTIER, F., LAINE, B., BELAÏCHE, D. and SAUTIERE, P. (1989)
J. Biol. Chem. soumis pour publication.

PRIMARY STRUCTURE OF THE CHROMOSOMAL PROTEIN MC1 FROM THE ARCHAE-
BACTERIUM *METHANOSARCINA* sp CHTI 55.

François CHARTIER¹, Bernard LAINE¹, Denise BELAÏCHE¹,
Jean-Pierre TOUZEL² and Pierre SAUTIERE¹.

¹Université de Lille II. Unité de Recherche Associée au Centre
National de la Recherche Scientifique n°409, Institut de
Recherches sur le Cancer, Place de Verdun 59045 LILLE FRANCE

²Station de Technologie Alimentaire de l'Institut National de
la Recherche Agronomique, B.P.39, 59651 VILLENEUVE D'ASCQ FRANCE

Correspondence and proofs should be sent to : François CHARTIER

Unité de Recherche Associée CNRS 409, Place de Verdun

59045 LILLE, FRANCE

F.CHARTIER, B.LAINE, D.BELAÏCHE and P.SAUTIERE, Université de Lille II, Unité de Recherche Associée au Centre National de la Recherche Scientifique n°409, Institut de Recherches sur le Cancer, Place de Verdun 59045 LILLE FRANCE.

J.P.TOUZEL, Station de Technologie Alimentaire de l'Institut National de la Recherche Agronomique, B.P.39 59651 VILLENEUVE D'ASCQ FRANCE.

Abbreviations :

Protein MC1, methanogen chromosomal protein 1.

Key words :

Archaeobacteria, Chromosomal protein, Primary structure, Transcription.

Enzymes :

Endoproteinase Arg-C (E.C.3.4.21.40.)

E.coli RNA polymerase (E.C.2.7.7.6.)

This paper is dedicated to the memory of Professor Gérard Biserte who pioneered research on chromosomal basic proteins in Lille.

SUMMARY

The DNA of the thermophilic archaebacterium *Methanosarcina* sp CHTI 55 has been shown to be associated with two proteins called MC1 and MC2 of molecular weight 11 kDa and 17 kDa [Chartier *et al*, Biochim. Biophys. Acta (1988)951,149-156]. The most abundant of these proteins, protein MC1, can protect DNA against thermal denaturation. In the present paper we report the covalent structure of protein MC1 and its effect on transcription of DNA *in vitro*.

The covalent structure was determined from automated sequence analysis of the protein and from structural data provided by peptides derived from cleavage of the protein at aspartic acid and arginine residues. The amino acid sequence of protein MC1 from *Methanosarcina* sp CHTI 55 is closely related to that of the protein MC1 (previously called HMB) isolated from *Methanosarcina barkeri* strain MS : among the nine substitutions observed between the two proteins seven are conservative. Transcription of DNA *in vitro* is stimulated by protein MC1 at low protein-to-DNA ratio but is inhibited at ratio higher than 0.1 (w/w) which is the one determined in the bacterial deoxyribonucleoprotein complex.

INTRODUCTION

In the archaeobacterium *Methanosarcina barkeri* MS, a mesophilic strain of methanogenic bacterium, the main component associated to DNA is a small basic polypeptide (M_r 10,760) called protein MC1 [1]. Protein MC1 has been localized in DNA-rich areas of *M.barkeri* cryosections, and does not form stable structural units reminiscent of eukaryotic nucleosomes [2]. Furthermore, its primary structure differs significantly from that of histones and eubacterial chromosomal proteins [3]. On the other hand, on the basis of molecular size and amino acid composition, protein MC1 does not resemble DNA-associated proteins isolated from other archaeobacteria such as *Thermoplasma acidophilum* and *Sulfolobus acidocaldarius* [4-6]. In addition, no sequence similarity has been found between *M.barkeri* protein MC1 [3] and the *T.acidophilum* protein HTa [7] or the three DNA-binding proteins of the 7 kDa group isolated from *S.acidocaldarius* [8].

Protein MC1 has also been encountered in the thermophilic strain *Methanosarcina* sp CHTI 55 together with a less abundant protein called MC2 (M_r 17,000) [9]. The protein MC1 from *Methanosarcina* sp CHTI 55 has an amino acid composition closely related to that of

M.barkeri MS protein MC1 and has been shown to protect DNA against thermal denaturation [9] which suggests that it may be involved in packaging of DNA. We report in the present paper the complete amino acid sequence of this protein and its effect on the transcription of DNA *in vitro*.

MATERIALS AND METHODS

Materials

Methanosarcina sp CHTI 55 (strain 2902) was grown as indicated in [9]. The chromosomal protein MC1 was isolated as indicated in [1] but the protein used for the transcription assays was purified in non-denaturing conditions by selective precipitation with ammonium sulphate [2]. Endoproteinase Arg-C and RNA polymerase were purchased from Boehringer. Acetonitrile for reversed-phase high performance liquid chromatography was from Carlo Erba. Reagents and solvents for the gas-phase sequencer were from Applied Biosystems.

Methods

DNA transcription *in vitro* assay

Each incubation mixture contained in 100 μ l : 50 mM Tris-HCl, pH 7.4 ; 10 mM MgCl₂ ; 10 mM MnCl₂ ; 0.5 mM dithiotreitol ; 100 mM KCl ; 0.25 unit of *E.coli* RNA polymerase ; 10 μ g of chicken erythrocyte DNA ; 0.5 mM each of CTP, GTP, UTP and 6 μ M of ³H-ATP (16.6 ci/mol, Amersham). The trichloroacetic acid insoluble radioactivity was measured after filtration of the mixture on glass fiber.

Sequence analysis

** Cleavage at aspartyl residues*

Protein MC1 (100 nmol) was hydrolysed with 1 ml of 0.25 M acetic acid in an evacuated sealed tube for 16 h at 105°C. The hydrolysate was freeze-dried prior to fractionation.

** Cleavage with the endoproteinase Arg-C*

Protein MC1 (240 nmol) was dissolved in 2.6 ml of 0.1 M ammonium bicarbonate buffer pH 8.0 and digested with endoproteinase Arg-C at 37°C for 4 h using an enzyme-to-substrate ratio of 1/50 (w/w). The digestion was stopped by lowering the pH with formic acid, and the hydrolysate was freeze-dried prior to fractionation.

** Fractionation of peptides*

Peptides were fractionated by reversed-phase HPLC as in [3] on a C₁₈ μ Bondapak column (Waters Associates) using gradient systems indicated in the legend to figure 3.

** Amino acid analyses*

Amino acid analyses of peptides were performed as in [3].

** Automated Edman degradation*

Automated Edman degradation of the protein MC1 and of peptides was carried out on a gas-phase sequencer Applied Biosystems 470A using the 03 RPTH Program. The phenylthiohydantoin derivatives of amino acids (PTH amino acids) were identified by reversed-phase HPLC on a column (200 x 2.1 mm) of Hypersil ODS 5 (Hewlett-Packard) using a Hewlett-Packard 1090 chromatograph. Separation of PTH amino acids was performed as in [10].

** Nomenclature of peptides*

Peptides derived from cleavage at aspartyl residues were designated by Ac-, whereas peptides generated by digestion with endoproteinase Arg-C were designated by PR-. Peptides were numbered according to their position in the amino acid sequence of the protein.

RESULTS

Effect of protein MC1 on DNA transcription *in vitro*

As a prerequisite to this assay it was checked that the protein MC1 did not exhibit affinity for nucleotides nor an enzymatic activity similar to that of RNA polymerase. For the assay, the activity of *E.coli* RNA polymerase was measured as a function of increasing amounts of protein MC1 associated to DNA (Fig.1). At low protein-to-DNA ratios (0.05 and 0.1 by weight) the RNA polymerase activity was strongly stimulated and reached 160 per cent of the activity observed with naked DNA. When higher amounts of protein were added to the DNA, the stimulation was no longer observed : on the contrary, the RNA polymerase activity appeared to be strongly inhibited. The time course study of RNA polymerase activity performed at a protein-to-DNA ratio equal to 0.1 by weight (Fig.2) indicates that the transcription linearly increases up to 60 min whereas with naked DNA a loss of activity can be observed as early as 15 min.

Sequence analysis of protein MC1

The amino terminal sequence of the protein MC1 was determined up to residue 53 by automated Edman degradation of the protein (Fig.4). Large adjacent or overlapping peptides generated by specific cleavage at aspartic acid residues and with the endoproteinase Arg-C allowed us to elucidate the remainder of the protein sequence.

** Cleavage at aspartyl residues*

Peptides derived from hydrolysis of protein MC1 with 0.25 M acetic acid were fractionated by reversed-phase HPLC (Fig.3-A). Four peptides were obtained directly in pure form. Their amino acid compositions are given in Table I. Due to partial deamidation of glutamine residues, peptide Ac-2 (residues 13-42) was eluted in fractions 1 and 2. Only the large peptides Ac-3 and Ac-4 which overlap the carboxy-terminal half of the protein (residues 44-93) were sequenced by automated Edman degradation (Fig.4).

** Cleavage with endoproteinase Arg-C*

Peptides generated by cleavage of the protein with endoproteinase Arg-C were fractionated by reversed-phase HPLC (Fig.3-B). The amino

acid compositions of peptides PR-1, PR-2, PR-3 and PR-4 which together account for the entire sequence of protein MCl are given in Table I. In fact peptides PR-2 and PR-3, which cover the adjacent sequences 31-46 and 47-59 respectively, were obtained together in equimolar amounts in fraction 1. When this fraction was submitted to gas phase sequencing, the sequence of peptide PR-3 could be easily determined since sequence of peptide PR-2 was already known from data provided by sequencing of the intact protein. In addition the sequence from residue 47 to residue 59 had already been determined by data provided by peptide Ac-3.

** Complete amino acid sequence of protein MCl*

The complete amino acid sequence of the protein MCl is presented in Fig.4. The sequence of the amino terminal half of the protein (residues 1 to 53) was determined by two runs of automated Edman degradation on the intact protein. Peptide Ac-3 allowed us to determine the sequence up to residue 65. The carboxy-terminal sequence of the protein MCl was established from data provided by peptide Ac-4 (residues 67-93) and the overlapping peptide PR-4 (residues 60-93).

DISCUSSION

Our results clearly show that the activity of RNA polymerase *in vitro* greatly depends on the amounts of protein MC1 associated to DNA. This activity is strongly stimulated with low amounts of protein MC1 and is inhibited when the protein-to-DNA ratio comes up to 0.2. It is worth pointing out that the maximum stimulation is obtained at a protein-to-DNA ratio close to 0.1 which is the ratio found in the native deoxyribonucleoprotein complex of *Methanosarcina barkeri*. It is obvious that this amount of protein associated to DNA appears sufficient to keep the reaction curve linear up to 60 min, while with naked DNA, a loss of linearity can be observed as early as 15 min.

Previous studies performed on complexes reconstituted with protein-to-DNA ratios higher than 0.25 showed that protein MC1 is able to protect DNA against thermal denaturation. Unfortunately due to the low sensitivity of the method, these studies could not be performed with lower ratios. In fact the protective effect probably

does not correspond to the physiological role of the protein [9] but is merely a property shared with other chromosomal proteins that are involved in packaging of DNA. This indicates that protein MC1 probably has an effect on the organization of DNA.

Furthermore it has been shown that proteins such as gyrase and the *E.coli* DNA-binding protein II which modify the supercoiling of DNA, modulate metabolic processes involving DNA as substrate or template and particularly DNA transcription [11-12]. The DNA wrapping activity of *E.coli* DNA binding protein II may promote DNA loop formation or changes in DNA dynamic flexibility and accessibility to proteins involved in DNA expression [13-14]. One emerging hypothesis is that at physiological ratio, protein MC1 allows the formation of a DNA conformation favorable to the activity of RNA polymerase whereas at higher ratios, the protein hinders binding of RNA polymerase through stabilization of the double-stranded DNA as observed in thermal denaturation studies [9]. Data obtained by these studies together with those provided by transcription assays will lead us to further analyse changes in DNA topology induced by protein MC1.

Analysis of the digest of protein MC1 with endoproteinase Arg-C calls for some comments about the specificity of this enzyme. Unexpectedly the bonds Lys-Ala at position 30-31 and 59-60 were completely cleaved whereas the bond Lys-Ala at position 39-40 was resistant to proteolysis. In *Methanosarcina barkeri* protein MC1, the bonds Lys-Ala at positions 30-31 and 59-60 were also completely cleaved [3]. It can be quoted that the two cleaved bonds are preceded by a hydrophobic cluster, while the bond at position 39-40 is preceded by a hydrophilic cluster. The activity of the endoproteinase Arg-C may be influenced by the hydrophobicity of the neighbouring sequence or may be hampered by the vicinity of the proline at position 42 which probably introduces a β turn.

The sequence of the protein MC1 from *Methanosarcina* sp CHTI 55 is close to that of the protein MC1 from *Methanosarcina barkeri* strain MS. Their relative molecular masses calculated from sequence data are 10,665 and 10,757 respectively. Nine substitutions distributed all along the polypeptide chain can be observed between these



two proteins (Fig.5). Seven of these substitutions are conservative. The two others involve the replacement of an aspartate residue by a serine residue or an alanine residue at positions 36 and 73 respectively. Only the substitution Asp/Ser at position 36 requires the change of two bases in the codon. Both proteins have the same number of basic residues but due to the loss of two acidic residues, the protein MC1 from *Methanosarcina* sp CHTI 55 has a net charge equal to + 12, versus + 10 for the homologous protein from *Methanosarcina barkeri*. One of the main differences between *Methanosarcina barkeri* and *Methanosarcina* sp CHTI 55, which belong to the same genus of Methanosarcinaceae, is their optimal temperature of growth : the former is mesophilic and grows up at 35°C whereas the latter is thermophilic and grows up at 55°C. This difference of growth temperature could induce structural constraints. Since the protein MC1 appears to be highly conserved within Methanosarcinaceae, it seems to be of great interest to study chromosomal proteins from the other genus of Methanosarcinaceae, the genus *Methanothermobacter*, and from other families of

Methanogens. The study of protein MC1 from *Methanotherix soehngeni* allowed us to show that this protein is indeed a mixture of three variants exhibiting numerous differences in their amino acid composition [15]. Preliminary sequence data suggest that these variants are structurally related to each other but differ significantly from protein MC1 isolated from the species *Methanosarcina barkeri* strain MS taken as reference.

ACKNOWLEDGEMENTS

The authors are grateful to Professor A.Verbert and to Dr. R.Cacan for helpful advice and stimulating discussions. They are indebted to A.Lemaire, M.J.Dupire, J.Herno and A.Hémez for technical assistance and to T.Ernout for kindly typing the manuscript. This work was supported by grants from the Centre National de la Recherche Scientifique, from the Pôle des Anaérobies de la Région Nord-Pas de Calais and from the Université de Lille II.

REFERENCES

- 1 Chartier,F., Laine,B., Sautière,P., Touzel,J.P. & Albagnac,G.
(1985) *FEBS Lett.* 183, 119-123.
- 2 Imbert,M., Laine,B., Prensier,G., Touzel,J.P. & Sautière,P.(1988)
Can.J.Microbiol. 34, 931-937.
- 3 Laine,B., Chartier,F., Imbert,M., Lewis,R. & Sautière,P. (1986)
Eur.J.Biochem. 161, 681-687.
- 4 Searcy,D.G. (1975) *Biochim.Biophys.Acta* 395, 535-547.
- 5 Green,G.R., Searcy,D.G. & Delange,R.J. (1983) *Biochim.Biophys.*
Acta 741, 251-257.
- 6 Dijk,J. & Reinhardt,R. (1986) in *Bacterial chromatin* (Gualerzi
C.O., Pon,C.L.,eds) pp.185-218, Springer-Verlag, Berlin,Heidelberg.
- 7 Delange,R.J., Williams,L.C. & Searcy,D.G. (1981) *J.Biol.Chem.*
256, 905-911.
- 8 Choli,T., Wittmann-Liebold,B. & Reinhardt,R. (1988) *J.Biol.Chem.*
263, 7087-7093.
- 9 Chartier,F., Laine,B. & Sautière,P. (1988) *Biochim.Biophys.Acta*
951, 149-156.

- 10 Chauvière, M., Martinage, A., Briand, G., Sautière, P. & Chevaillier, P. (1989) *Eur. J. Biochem.* in press.
- 11 Cozzarelli, M.R. (1980) *Science* 207, 953-960.
- 12 Drlica, K. & Rouvière-Yaniv, J. (1987) *Microbiol. Rev.* 51, 301-319.
- 13 Johnson, R.C., Bruist, M.F. & Simon, M.I. (1986) *cell* 46, 531-539.
- 14 Flashner, Y. & Gralla, J.D. (1988) *cell* 54, 713-721.
- 15 Chartier, F., Crevel, G., Laine, B. & Sautière, P. (1989)
J. Chromatogr. 466, 331-337.

Fig.1 - *E.coli* RNA polymerase activity as a function of the amount of protein MC1 added to DNA. Incubation times were 45 min. Results are expressed as percentages of the RNA polymerase activity obtained with naked DNA used as reference.

Fig.2 - Comparison of the enzymatic time courses of *E.coli* RNA polymerase in absence of protein MC1 (○—○) and in presence of protein MC1 (●—●) protein-to-DNA ratio of 0.1 (w/w).

Fig.3 - Fractionation of peptides by reversed-phase HPLC. Chromatography was performed on a C₁₈ μBondapak column (30 x 0.39 cm) : (- - -) gradient of acetonitrile in 0.05 % trifluoroacetic acid at a flow rate of 1.5 ml/min ; (—) absorbance at 206 nm. Fractions were numbered by order of elution. Peptides used for sequence were numbered according to their position in the sequence of the protein. (A) Peptides derived from cleavage of protein MC1 at aspartyl residues. 50 nmol of hydrolysate were loaded. (B) Peptides generated by hydrolysis of protein MC1 with endoproteinase Arg-C. 60 nmol of digest were loaded.

Fig.4 - Complete amino-acid sequence of the *Methanosarcina* sp CHTI 55 protein MC1. Methods used for determination of the sequence are indicated as follows : +++ automated Edman degradation of the protein ; \longrightarrow automated Edman degradation of peptides. PR- : peptides generated by digestion of the protein with endoproteinase Arg-C. Ac- : peptides obtained by cleavage of the protein at aspartyl residues.

Fig.5 - Structural homologies between protein MC1 from *Methanosarcina* sp CHTI 55 and protein MC1 from *Methanosarcina barkeri*. The sequence shown is that of protein MC1 from *Methanosarcina* sp CHTI 55. Only the changes in sequence of protein MC1 from *Methanosarcina barkeri* are indicated below. The boxes indicate non conservative changes.

Table I . Amino acid composition of peptides derived from protein MCl. Values are expressed as number of residues per mole of peptide . The numbers of residues in parentheses are from the established sequence. No correction was made for degradative losses of threonine and serine during hydrolysis of peptides. Values for protein MCl from *Methanosarcina* sp CHTI 55 are from [9].

Amino acids	Ac-1 1-9	Ac-2 13-42	Ac-3 44-65	Ac-4 67-93	PR-1 1-30	PR-2 & 3 31-59	PR-4 60-93	Protein MCl
Aspartic acid	1.9 (2)	3.0 (3)		1.2 (1)	4.7 (5)	3.3 (3)	2.4 (2)	10
Threonine	0.9 (1)	1.8 (2)	1.2 (1)		1.7 (2)	2.1 (2)		4
Serine	0.7 (1)	1.2 (1)		0.6 (1)	1.4 (1)	2.8 (1)	1.1 (1)	3
Glutamic acid		3.4 (3)	2.4 (2)	4.1 (4)	4.0 (4)	1.3 (1)	4.9 (5)	10
Proline		1.7 (1)		4.4 (4)	0.9 (1)	1.1 (1)	3.8 (4)	6
Glycine		5.2 (5)	1.5 (1)		3.4 (3)	3.2 (3)		6
Alanine		4.7 (5)	1.6 (1)	2.1 (2)	2.2 (2)	1.8 (3)	3.1 (3)	8
Valine	1.1 (1)	1.0 (1)	2.6 (3)	1.3 (1)	1.7 (2)	2.2 (2)	2.4 (2)	6
Methionine				tr. (1)			0.6 (1)	1
Isoleucine			1.8 (3)	1.8 (2)		1.4 (2)	2.6 (3)	5
Leucine	1.0 (1)	1.1 (1)	1.1 (1)	1.0 (1)	2.1 (2)	0.9 (1)	1.1 (1)	4
Phenylalanine	1.1 (1)	1.0 (1)	1.0 (1)	1.0 (1)	2.0 (2)	0.9 (1)	1.3 (1)	4
Histidine		1.1 (1)	0.7 (1)		1.1 (1)	0.9 (1)		2
Lysine		2.7 (3)	3.9 (4)	5.4 (6)	2.0 (2)	3.7 (4)	6.2 (6)	13
Arginine	1.9 (2)	2.0 (2)	2.8 (3)	2.0 (2)	2.7 (3)	3.1 (4)	2.2 (2)	9
Tryptophan			n.d.(1)	n.d.(1)			n.d.(2)	2
Total	9	30	22	30	30	29	34	93
Fraction number	3	1 + 2	5	4	2	1	3	
% yield of recovery	60	49	63	43	25	9	20	

tr. , traces ; n.d., not determined.

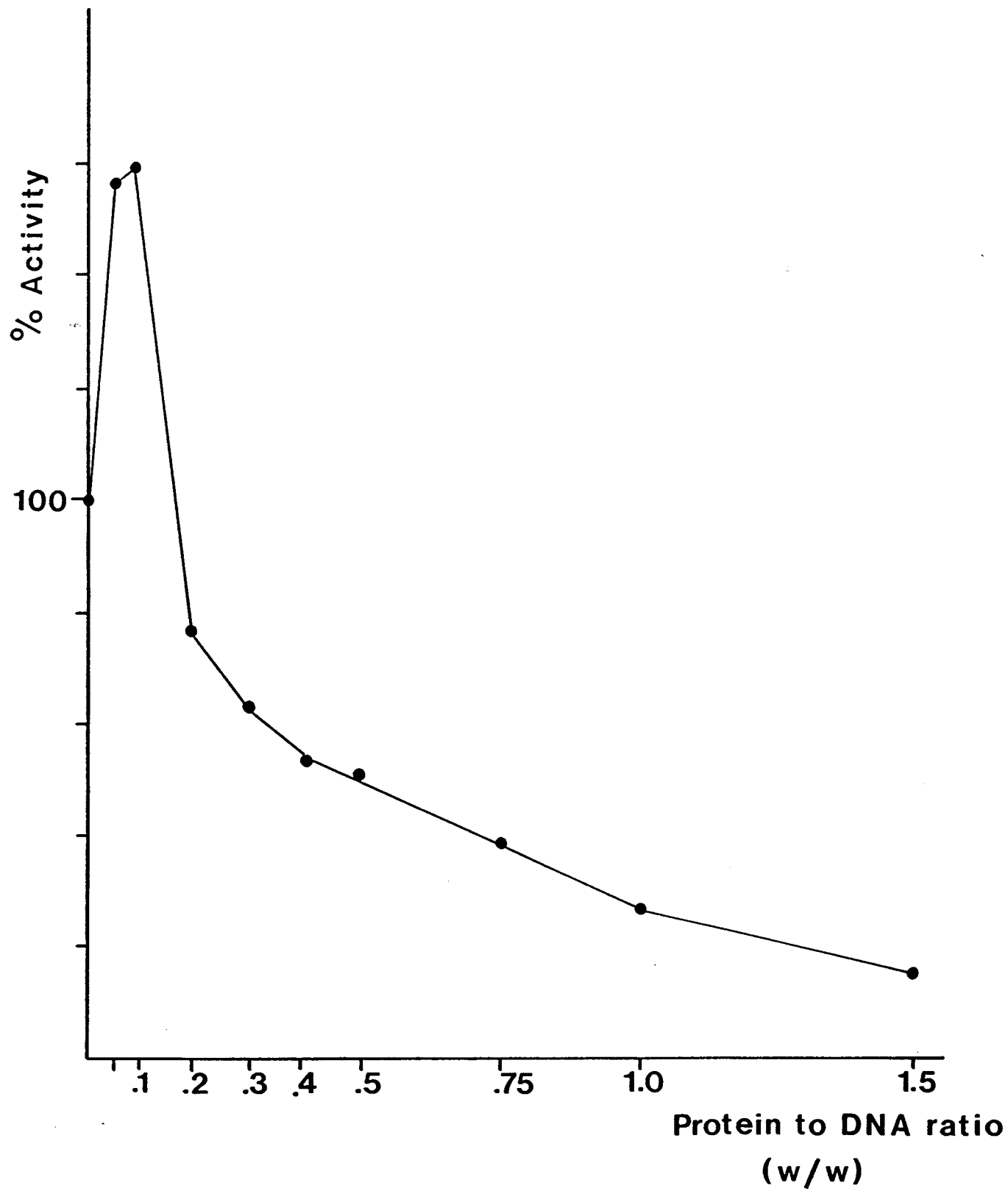


Figure 1

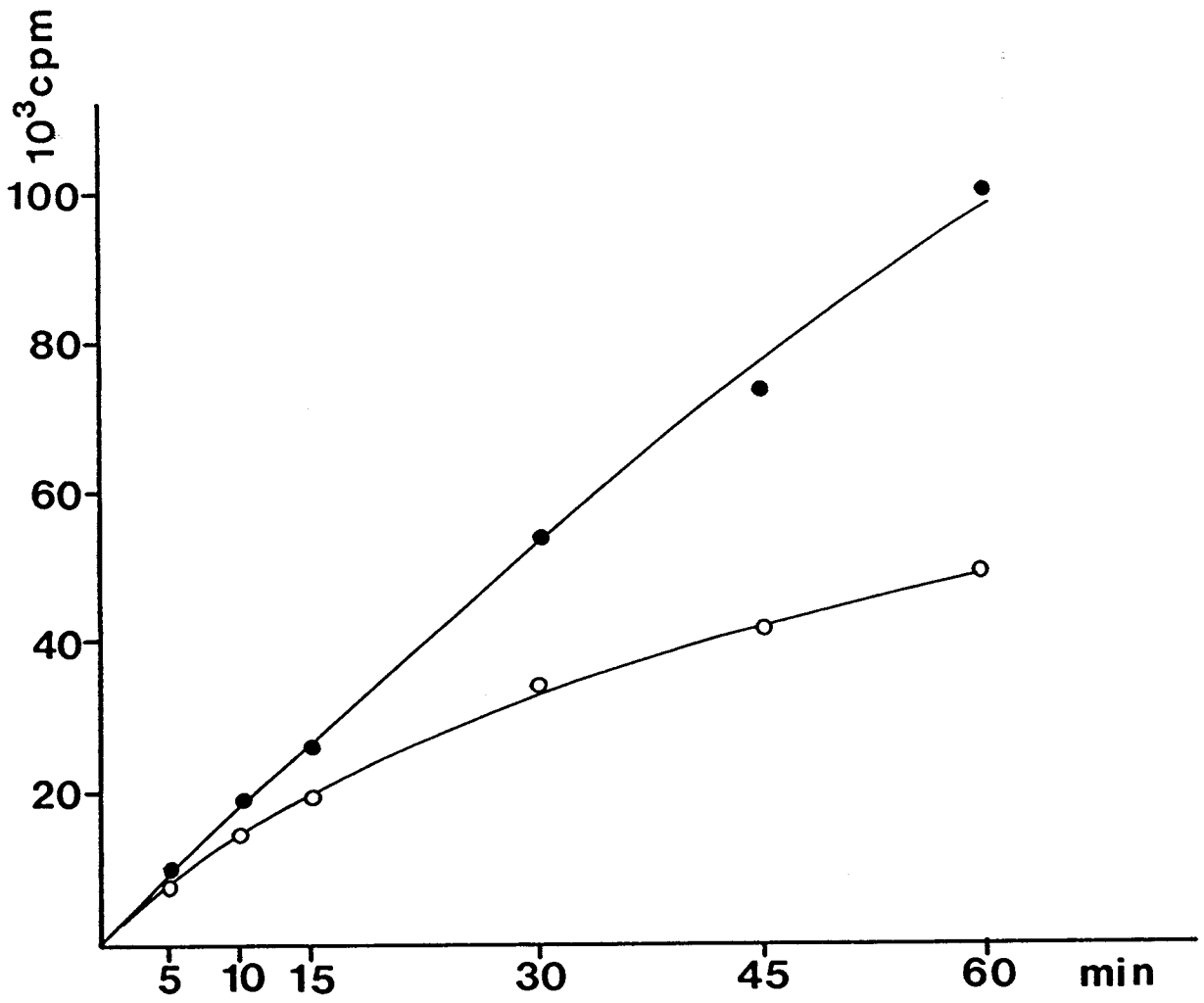


Figure 2

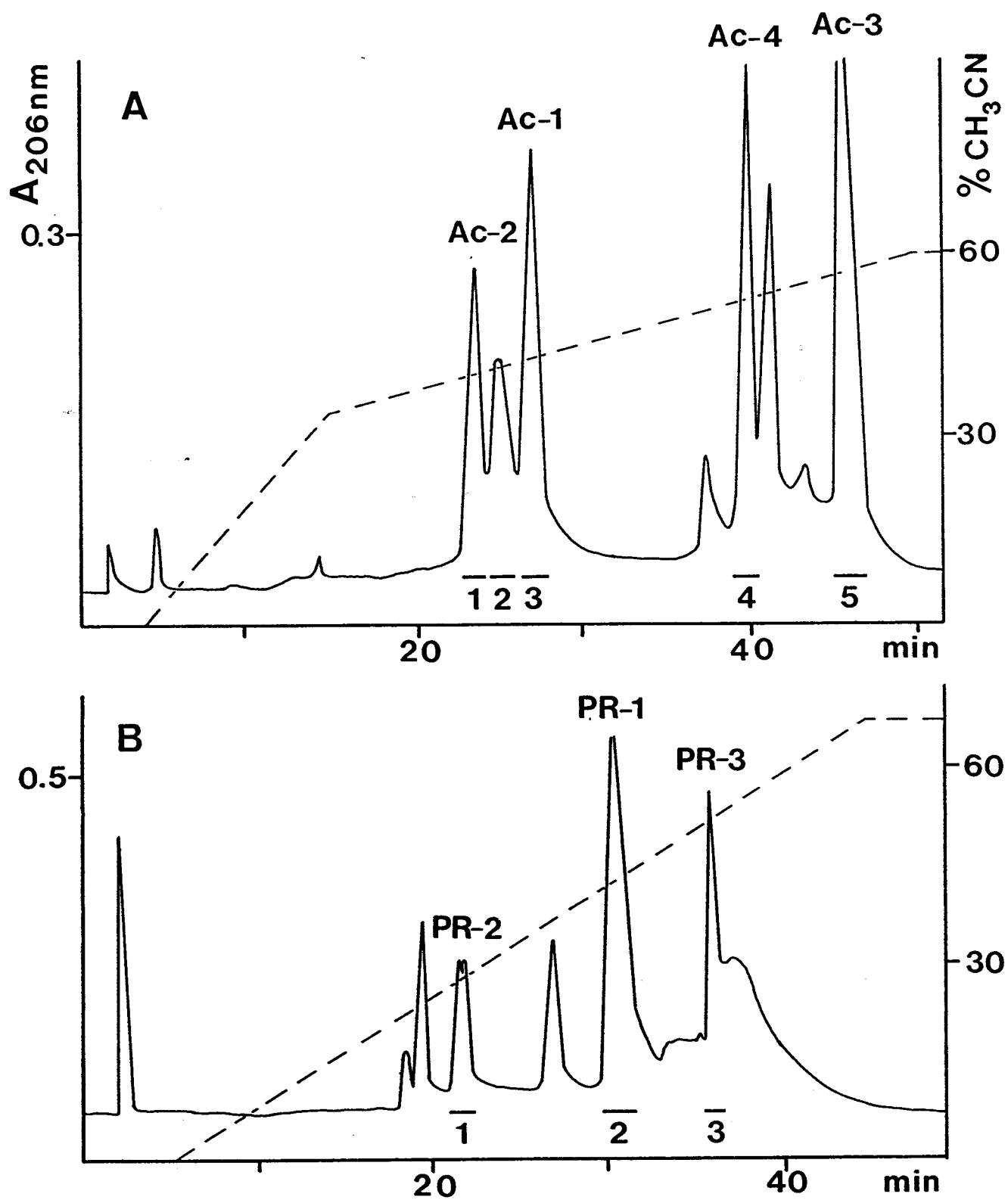


Figure 3

PRIMARY STRUCTURE OF THE CHROMOSOMAL PROTEINS MC1a, MC1b AND MC1c FROM THE
ARCHAEBACTERIUM *METHANOTHRIX SOEHNGENII*.

François Chartier*, Bernard Laine, Denise Bélaïche and Pierre Sautière

Université de Lille II - Unité de Recherche Associée 409 au Centre National
de la Recherche Scientifique, Institut de Recherches sur le Cancer,
Place de Verdun 59045 Lille, France.

* To whom correspondence should be sent.

Running title

Methanotherix soehngeni chromosomal proteins MC1.

SUMMARY

The chromosomal protein MC1 of *Methanotherix soehngeni* is a family of three variants a, b and c. These are small basic polypeptides of 89, 87 and 90 residues respectively. Their primary structures have been determined from automated sequence analyses of the intact proteins and from structural data provided by peptides derived from the variants by cleavage at aspartic acid, glutamic acid, arginine and methionine residues. By comparison with variant b taken as reference, variants a and c present 18 and 24 differences respectively. The extent of sequence homologies between protein MC1 from *Methanotherix soehngeni* and proteins MC1 from two other species of Methanosarcinaceae is only 60 %. The sequences 17-35 and 45-58 of the protein MC1 appear well conserved. Deletions are observed in region 36-44. Many changes, most of them non-conservative, occur in the carboxy-terminal third of the protein. However, proline residues at positions 68, 72, 76 and 82 remain strictly conserved. Predictive methods for secondary structures indicate a low content of α helix and β sheet structures in proteins MC1.

INTRODUCTION

The protein MC1, a basic polypeptide of $M_r \sim 10\ 000$, has been shown to be associated to DNA *in vivo* (Imbert *et al.*, 1988) and to be the major chromosomal protein in two different species of Methanosarcinaceae : *Methanosarcina barkeri* and *Methanosarcina* sp. CHTI 55 (Chartier *et al.*, 1985 ; Chartier *et al.*, 1988). Protein MC1 of *Methanosarcina* sp. CHTI 55 is able to stabilize DNA against thermal denaturation (Chartier *et al.*, 1988) and to stimulate DNA transcription *in vitro* (Chartier *et al.*, 1989). These effects are likely to be correlated with a structural role of the protein in the bacterial chromatin. The primary structure of protein MC1 differs significantly from those of chromosomal proteins isolated from eubacteria, eukaryotes or other archaeobacteria (Laine *et al.*, 1986 ; DeLange *et al.*, 1981 ; Choli *et al.*, 1988). The close relationship observed between the primary structure of protein MC1 isolated from *Methanosarcina barkeri* and *Methanosarcina* sp. CHTI 55, which belong to the same genus of the family Methanosarcinaceae, led us to study the chromosomal proteins in the other genus of Methanosarcinaceae, the genus *Methanotherix* (Stackebrandt *et al.*, 1982). *Methanotherix soehngeni* strain FE is mesophilic and, contrary to the species cited above, is a filamentous bacterium and grows only on acetate (Touzel *et al.*, 1988). *Methanotherix soehngeni* protein MC1 is indeed a complex of three variants which have been separated by reversed-phase HPLC (Chartier *et al.*, 1989). In this paper, we report the complete amino acid sequences of the three variants of protein MC1 isolated from *Methanotherix soehngeni*. Comparison of primary structures of these variants with protein MC1 from *Methanosarcina* strains shows that sequences 17-35 and 45-58 are well conserved whereas the carboxy-terminal third of the molecule is highly variable.

RESULTS

The strategy and sequence data which allowed us to determine the complete sequences of the variants a, b and c of *M.soehngeni* protein MC1 are presented in figures 1, 2 and 3 respectively. The fractionation diagrams of the peptides derived from these proteins are given in the Miniprint Section. Since the elution profiles of peptides derived by cleavage of the three variants with cyanogen bromide are identical, only the one obtained with the variant MC1a is presented.

Protein MC1a

The amino-terminal sequence of protein MC1a was determined up to residue 59 by microsequencing of the intact protein (Fig.1). Cleavage of the protein at aspartic acid residues yielded peptide Ac-3 covering the sequence 39-87. Microsequencing of this peptide provided sequence data for the region 39-70 of the protein. The remainder of the sequence was unambiguously elucidated by sequencing peptide PE-3 (residues 44-89) derived from cleavage of the protein with endoproteinase Glu-C and peptide CN-2 (sequence 70-89) obtained by cleavage of the protein with cyanogen bromide. The amino acid compositions of these peptides are presented together with that of the intact protein in the Miniprint Section (Table I).

Protein MC1b

The amino-terminal sequence of protein MC1b was determined up to residue 63 by microsequencing of the intact protein (Fig.2). Cleavage of the protein with endoproteinase Glu-C yielded peptides PE-4 and PE-5 (residues 44-81 and 82-87 respectively) which correspond to the carboxy-terminal part of the protein. Although peptide PE-5 coeluted with peptide PE-3 (residues 39-43)(in

fraction 1 of Fig.S3) its sequence could be easily determined since the structure of peptide PE-3 was already known from data provided by sequencing the intact protein. The alignment of peptides PE-4 and PE-5 was established by the overlapping peptide CN-2 obtained after cleavage of the protein at the methionyl residue at position 69. The amino acid composition of these peptides are given in Table II in the Miniprint Section together with that of the intact protein MC1b.

Protein MC1c

The amino-terminal sequence of protein MC1c was determined up to residue 55 by microsequencing of the intact protein (Fig.3). The sequence of the median region of the protein was elucidated by data provided by peptides PE-3 (residues 45-90) and PR-1 (residues 49-80) obtained by cleavage of the protein with endoproteinase Glu-C and endoproteinase Arg-C respectively. The remainder of the sequence was determined by sequencing the peptide CN-2 obtained by cleavage of the protein at the methionyl residue at position 70. The amino acid compositions of these peptides are given in Table III in the Miniprint Section together with that of the intact protein MC1c.

In the course of the sequence analyses of protein MC1 variants, we took advantage of an unspecific cleavage with endoproteinase Arg-C of two Lys-Lys bonds at positions 48-49 and 80-81 in protein MC1c which yielded peptide PR-1 (residues 49-80)(Fig.3), and of a lack of cleavage of some glutamyl bonds with endoproteinase Glu-C in the three variants. Indeed, endoproteinase Glu-C failed to cleave the glutamyl bonds located in the vicinity of the amino- and carboxy-termini of the peptide chains as observed for the glutamyl bonds at positions 2 and 84 in proteins MC1a and MC1b, and at positions 3, 85 and 88 in protein MC1c. The bonds Glu-Arg at position 55-56 in variants a and b and 56-57 in

variant c, Glu-Ile at position 72-73 in protein MC1a and Glu-Gly at position 13-14 in protein MC1c are resistant to hydrolysis with endoproteinase Glu-C. This is not due to the nature of the amino acid adjacent to the glutamic acid residue since the enzyme cleaved identical bonds with glutamic acid at positions 43 in protein MC1a, 15, 38, 43 and 81 in protein MC1b and 16 and 44 in protein MC1c (see Figures 1, 2 and 3). The fact that all these resistant bonds are located in sequences having a high propensity to be in a β -turn conformation, is a possible explanation for their resistance to cleavage with endoproteinase Glu-C.

DISCUSSION

Considering the yields in which they were obtained, variants a, b and c represent approximately 44, 30 and 26 per cent of the protein MC1 respectively. The three proteins have slight differences in size : MC1a, MC1b and MC1c contain 89, 87 and 90 residues respectively and have relative molecular masses of 10 112, 9 762 and 10 323 respectively, as calculated from their amino acid sequences. These proteins appear structurally related to each other, thus demonstrating that they represent variants of the same protein. Among these three variants , variant b was taken as reference because it is the most closely related to the protein MC1 from *Methanosarcina barkeri* and *Methanosarcina* sp. CHTI 55. Variant a differs from variant b by five conservative substitutions, eleven non conservative substitutions and two extra residues at the carboxy-terminus of the molecule (Fig.4). Variant c differs from variant b by ten conservative substitutions, eleven non conservative substitutions and by the presence of one additional residue at the amino-terminus and two additional residues at the carboxy-terminus. These substitutions do not introduce signifi-

cant differences in the overall distribution of charged residues except in the variant MC1a which is much more basic than the variants b and c : the net charge is +16 for variant a, versus +9 and +11 for variants b and c respectively. The extra positive charges in protein MC1a introduce two doublets of basic residues, Lys-Lys and His-Lys at positions 11-12 and 36-37 respectively, in addition to a basic sequence located between residues 56 and 63 (Fig.5). By comparison with proteins MC1a and MC1b, protein MC1c contains an additional proline residue at position 60.

The presence of variants of protein MC1 has not been observed previously, but variants of chromosomal proteins have been described in the archaebacteria *Thermoplasma acidophilum* (DeLange *et al.*,1981) and *Sulfolobus acidocaldarius* (Choli *et al.*,1988). In the protein HTa from *Thermoplasma acidophilum* the second polypeptide differs from the first by one additional residue at its amino-terminus. In the protein 7 of *Sulfolobus acidocaldarius*, the variant b differs from the variant a by two deletions at the carboxy-terminus, and the variant e differs from variant a by three substitutions and three additional residues at the carboxy-terminus.

The comparison of the amino acid sequences of the variants of protein MC1 from *Methanotherxix soehngeni* with those of homologous proteins isolated from other species is presented in Fig.6. Two regions, where only a few substitutions occur, appear well conserved ; these are region 17-35, which is rich in glycine and alanine, and region 45-58 which is the most basic sequence in the protein MC1. These two conserved regions together represent 40 % of the whole molecule. They are separated by a highly variable amino acid residue at position 36 and a region where deletions accounting for a total of six amino acid residues are observed. However these changes do not introduce a significant modification in the secondary structure of this region which is predicted to be in a coil

structure in the protein MC1 isolated from *Methanosarcina barkeri* and *Methanosarcina* sp.CHTI 55 (Fig.7). An other marked change in the amino-terminal sequence of the protein MC1 is the loss of the acidic triplet at position 10-12. The carboxy-terminal third of the protein MC1, particularly the sequence 59-73, is the most variable. Indeed numerous changes are encountered in this region. Among these changes, the substitutions Ala-Gly at position 60 is very important because it introduces a β turn as predicted by the predictive methods used in our studies (Fig.7). This β turn plays a role of breaker of the α helix formed with residues 58 through 66 in *M.barkeri* MC1 (Fig.7). Furthermore, at position 63, the replacement of the glutamic acid residue, which is a helix promotor, by a hydrophobic residue results in the formation of a β sheet strand between residues 63 and 66 in variants a and b of *Methanotherix soehngeni* protein MC1. In variant c, the additional proline residue at position 66 hinders the formation of the β sheet strand. It must be emphasized that in the highly variable sequence the bulky hydrophobic residues at positions 65, 74, 75 and 79 and the four proline residues at positions 68, 72, 76 and 82 are strictly conserved. Moreover the three proline residues at positions 68, 72 and 76 give rise to three adjacent β turns : such a particular conformation could play a crucial role in the function of the protein MC1.

Previous studies show that the primary structures of protein MC1 from two species belonging to the genus *Methanosarcina* exhibit 90 % of sequence homology, and differ only by two non conservative substitutions and seven conservative substitutions. The present work shows that protein MC1 from a strain of the other genus of Methanosarcinaceae, the genus *Methanotherix*, differs much more. The extent of sequence homologies between MC1 proteins isolated from the two genus is about 60 %. The comparison of methanogen chromosomal proteins with chromosomal proteins from other Archaeobacteria such as protein HTa from

Thermoplasma acidophilum (DeLange *et al.*,1981) and DNA-binding proteins 7a, 7b, 7e from *Sulfolobus acidocaldarius* (Choli *et al.*,1988) show that these proteins are not structurally related.

Work is now in progress to determine the amino acid residues involved in interaction with DNA. Preliminary data indicates that the carboxy-terminal third of the protein may contain the binding site to DNA. Studies of 16 S RNA and DNA-dependant RNA polymerases (Thomm *et al.*,1986 ; Klenk *et al.*,1986) show that methanogens form a very heterogeneous group. At the present time, very little is known about the chromosomal proteins in other families of methanogenic bacteria. It will be of interest to screen these bacteria for chromosomal proteins.

Acknowledgements

The authors are grateful to A.Lemaire, M.J.Dupire and J.Herno for technical assistance, and T.Ernout for kindly typing this manuscript. They are indebted to B.Gruson who provided the *Methanotherix soehngeni* cells. This work was supported by grants from the Centre National de la Recherche Scientifique, from the Pôle des Anaérobies de la Région Nord-Pas-de-Calais and from the Université de Lille II.

REFERENCES

- Chartier, F., Laine, B., Sautière, P., Touzel, J.P. and Albagnac, G. (1985) FEBS Lett. 183, 119-123.
- Chartier, F., Laine, B. and Sautière, P. (1988) Biochim. Biophys. Acta 951, 149-156.
- Chartier, F., Crevel, G., Laine, B. and Sautière, P. (1989) J. Chromatogr. 466, 331-337.
- Chartier, F., Laine, B., Bélaïche, D., Touzel, J.P. and Sautière, P. (1989) Biochim. Biophys. Acta, submitted to publication.
- Chauvière, M., Martinage, A., Briand, G., Sautière, P. and Chevaillier, P. (1989) Eur. J. Biochem. 180,
- Choli, T., Wittman-Liebold, B. and Reinhardt, R. (1988) J. Biol. Chem. 263, 7087-7093.
- Chou, P.Y. and Fasman, G.D. (1978) Ann. Rev. Biochem. 47, 251-276.
- DeLange, R.J., Williams, L.C. and Searcy, D.G. (1981) J. Biol. Chem. 256, 905-911.
- Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J.P. (1987) FEBS Lett. 224, 149-155.
- Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97-120.
- Imbert, M., Laine, B., Prensier, G., Touzel, J.P. and Sautière, P. (1988) Can. J. Microbiol. 34, 931-937.
- Klenk, H.P., Haas, B., Schwass, V. and Zillig, W. (1986) J. Mol. Evol. 24, 167-173.
- Laine, B., Chartier, F., Imbert, M., Lewis, R. and Sautière, P. (1986) Eur. J. Biochem. 161, 681-687.

Levin, J.M., Robson, B. and Garnier, J. (1986) FEBS Lett. 205, 303-308.

Stackebrandt, E., Seewaldt, E., Ludwig, M., Schleifer, K.H. and Huser, B.A. (1982)
Zentralbl. Bakteriol. Mikrobiol. Hyg. Abt. 1 Orig C3 : 90-100.

Thomm, M., Madon, J. and Stetter, K.O. (1986) Biol. Chem. Hoppe-Seyler, 367, 473-481.

Touzel, J.P., Prensier, G., Roustan, J.L., Thomas, I., Dubourguier, H.C. and Albagnac, G.
(1988) Int. J. Syst. Bacteriol. 38, 30-36.

Abbreviations

Protein MC1, methanogen chromosomal protein 1

Enzymes

Endoproteinase Glu-C (E.C.3.4.21.19.)

Endoproteinase Arg-C (E.C.3.4.21.40.)

This paper is dedicated to the memory of Professor Gérard Biserte who pioneered research on basic chromosomal proteins in Lille.

FIGURE LEGENDS

Figure 1.

Complete amino acid sequence of *Methanotherix soehngeni* protein MC1a. Methods used for the determination of the sequence are indicated as follows : +++ , automated Edman degradation of the protein ; —> , automated Edman degradation of peptides. PE, Ac, and CN indicate peptides derived from cleavage with endoproteinase Glu-C, dilute acetic acid and cyanogen bromide respectively.

Figure 2.

Complete amino acid sequence of *Methanotherix soehngeni* protein MC1b. Methods and nomenclature of peptides are as indicated in legend to Fig.1.

Figure 3.

Complete amino acid sequence of *Methanotherix soehngeni* protein MC1c. Methods and nomenclature of peptides are as indicated in legend to Fig.1. PR- : peptide obtained by digestion of the protein with endoproteinase Arg-C.

Figure 4.

Alignment of the amino acid sequences of the chromosomal proteins MC1a, b and c from *Methanotherix soehngeni*. Protein MC1b (2) taken as reference ; (1) protein MC1a ; (3) protein MC1c. For the two latter proteins, only the changes are indicated.

Figure 5.





Distribution of charged residues in proteins MC1a, b and c from *Methanotherix soehngeni*. , basic residues ; , acidic residues.

Figure 6.

Comparison of amino acid sequences of proteins MC1 from different methanogenic strains. *Methanosarcina barkeri* (1) ; *Methanosarcina* sp.CHTI 55 (2) ; *Methanotherix soehngeni* : variant a (3), variant b (4), variant c (5). Variant c was aligned for maximum homology. The amino acid deletions are represented by ----. Strictly conserved proline residues are indicated by arrows.

Figure 7.

Prediction of secondary structures of protein MC1 : *Methanosarcina barkeri* (1) ; *Methanotherix soehngeni* : variant a (2), variant b (3) and variant c (4). Results for *Methanosarcina barkeri* protein MC1 are from (Imbert,M., Laine,B., Helbecque,N., Mornon,J.P. and Sautière,P., manuscript in preparation). Four predictive methods are used (see Material and Methods). For clarity of the figure only structures predicted by at least two methods were considered in the joint prediction presented. Dotted lines correspond to the deletions.

The symbols represent residues in α helix (, β sheet (, β turn (, random coil is not presented.

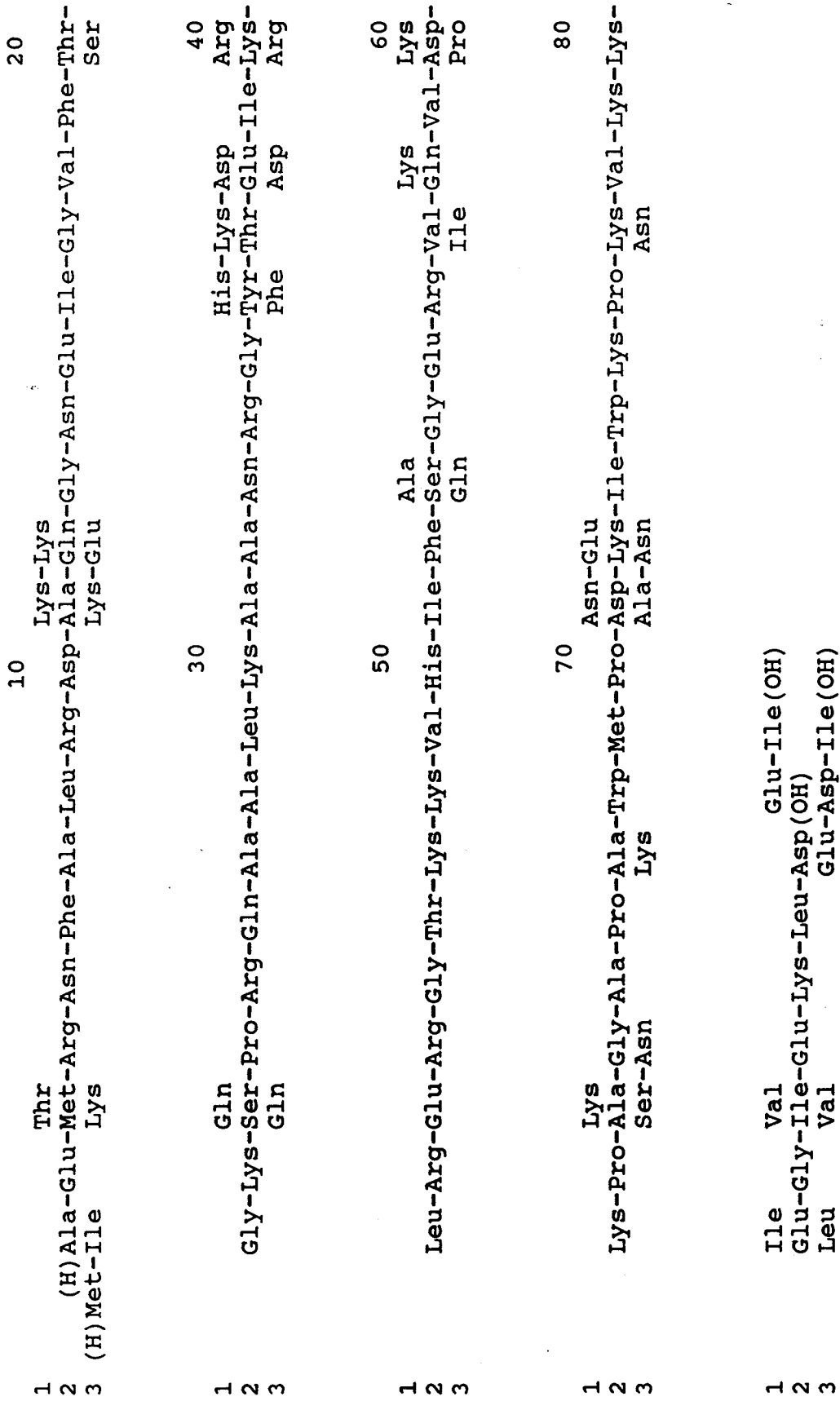


Figure 4

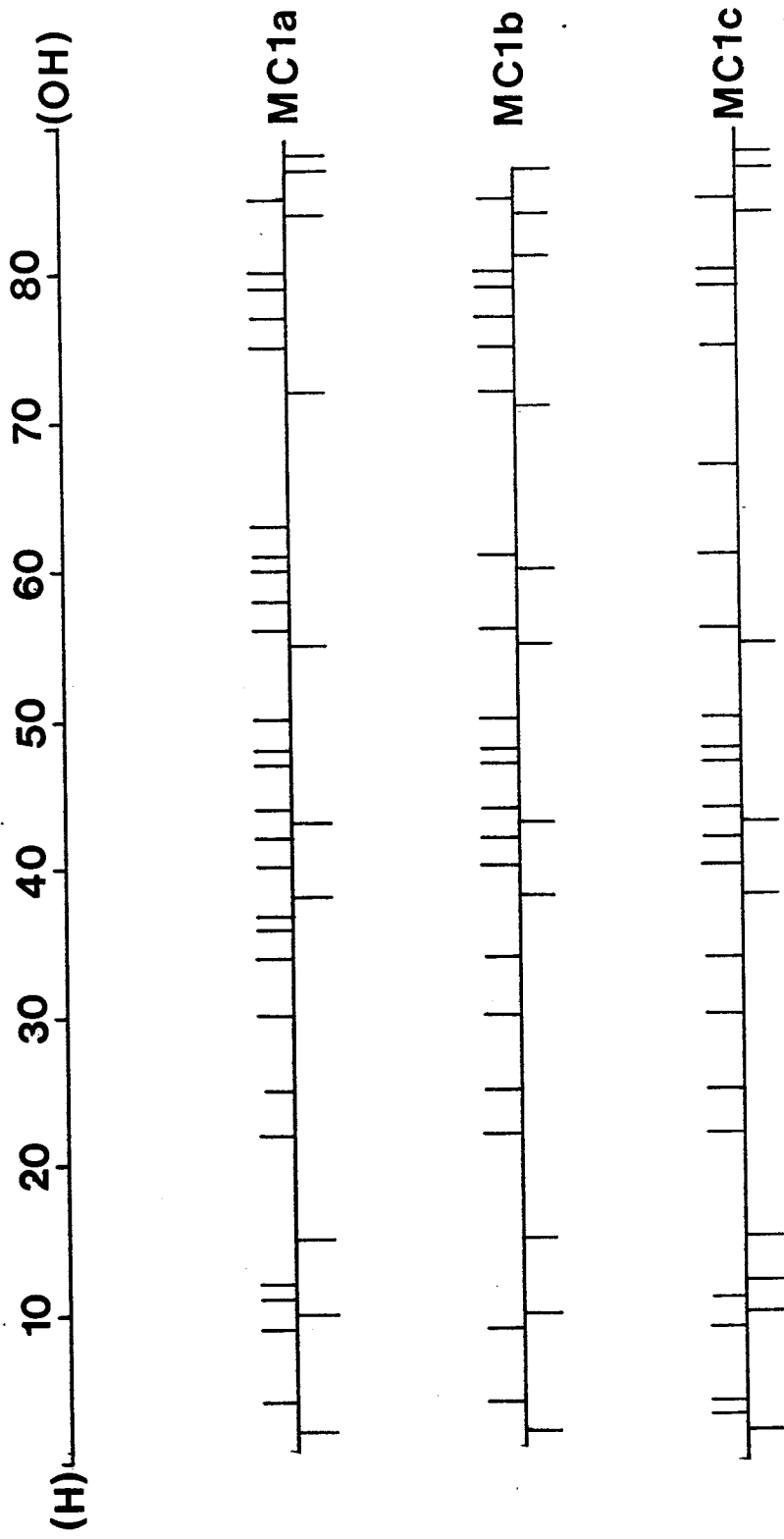


Figure 5

		10		20
1	(H) Ser-Asn-Thr-Arg-Asn-Phe-Val-Leu-Arg-Asp-Glu-Glu-Gly-Asn-Glu-His-Gly-Val-Phe-Thr-			
2			Asp	
3	(H) Ala-Glu	Ala	Lys-Lys	Ile
4	(H) Ala-Glu-Met	Ala	Ala-Gln	Ile
5	(H) Met-Ile-Glu-Lys	Ala	Lys	Ile Ser

		30		40
1	Gly-Lys-Gln-Pro-Arg-Gln-Ala-Ala-Leu-Lys-Ala-Ala-Asn-Arg-Gly-Asp-Gly-Thr-Lys-Ser-			
2				Ser Ala
3				His-----
4	Ser			Tyr-----
5				Phe-----

		50		60
1	Asn-Pro-Asp-Val-Ile-Arg-Leu-Arg-Glu-Arg-Gly-Thr-Lys-Lys-Val-His-Val-Phe-Lys-Ala-			
2		Ile		
3	-----	-----		Ile Ala-Gly
4	-----Glu-----	Lys		Ile Ser-Gly
5	-----	-----		Ile Gln-Gly

		70		80
1	Trp-Lys-Glu-Met-Val-Glu-Ala-Pro-Lys-Asn-Arg-Pro-Asp-Trp-Met-Pro-Glu-Lys-Ile-Ser-			
2		Ile Asp		Ala
3	Glu-Arg-Val-Lys	Lys-Lys	Gly-Ala	Ala Asn-Glu Trp
4	Glu-Arg-Val-Gln	Asp-Lys	Ala-Gly-Ala	Ala Asp Trp
5	Glu-Arg-Ile-Gln	Pro-Lys	Ser Ala Lys	Ala-Asn Trp

		90		
1	Lys-Pro-Phe-Val-Lys-Lys-Glu-Lys-Ile-Glu-Lys-Ile-Glu(OH)			
2		Arg		Leu
3	Lys	Ile-Gly-Val		Leu-Asp-Glu-Ile(OH)
4	Lys	Gly		Leu-Asp(OH)
5	Asn	Leu-Gly-Val		Leu Asp-Ile(OH)

Figure 6

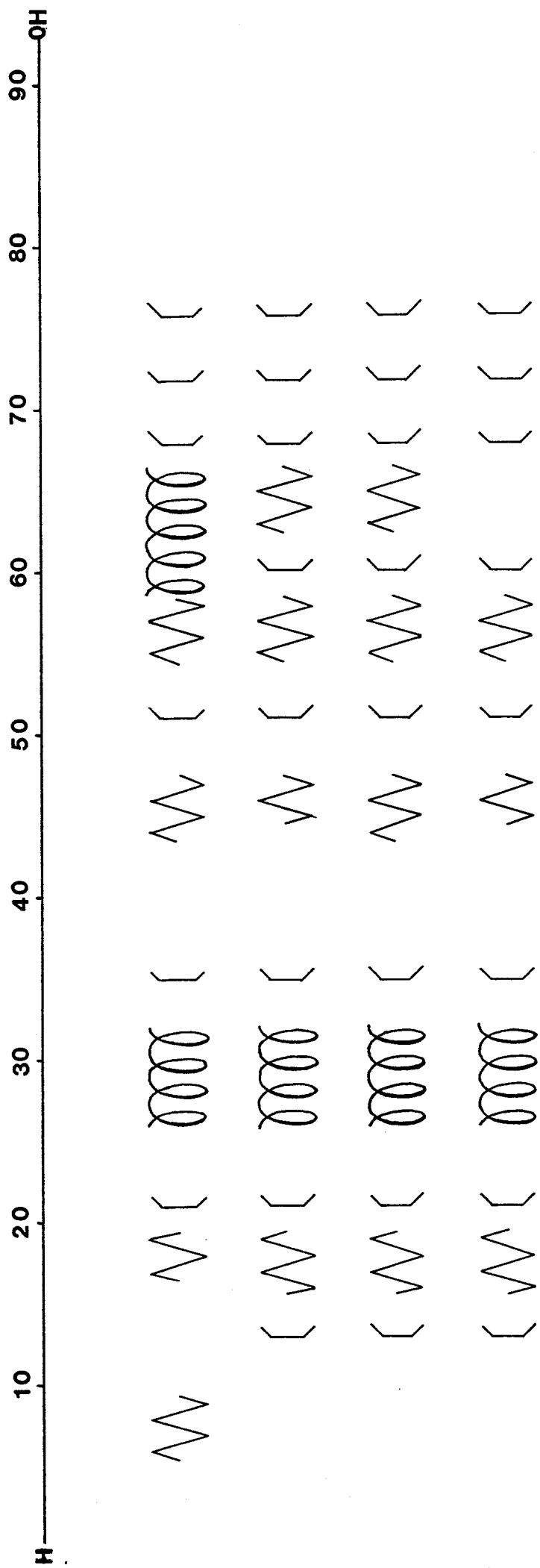


Figure 7

Supplementary Material to :

Primary structure of the chromosomal proteins MC1a, MC1b and MC1c from the archaebacterium *Methanothermobacter thermoautotrophicus*

by

François Chartier, Bernard Laine, Denise Bélaïche and Pierre Sautière

EXPERIMENTAL PROCEDURES

Materials. *Methanothermobacter thermoautotrophicus* strain FE (DSM 3013) was grown as indicated in (Touzel *et al.*, 1988). Endoproteinase Arg-C was from Pierce and endoproteinase Glu-C was from Miles Laboratories. Cyanogen bromide was from Eastman Kodak (Rochester, USA). Acetonitrile for reverse-phase high-pressure liquid chromatography was from Carlo Erba. Reagents and solvents for the gas-phase sequencer were from Applied Biosystems.

Isolation of protein MC1 and separation of its three variants were performed as described in (Chartier *et al.*, 1989).

Cleavage at aspartyl residues. The protein MC1a (100 nmol) was hydrolyzed with 1 ml of 0.25 M acetic acid in an evacuated sealed tube for 16 h at 105°C.

Cleavage at methionyl residues. To 100 nmol of each variant, dissolved in 1 ml of 70% (v/v) formic acid, cyanogen bromide was added in a 100 fold molar excess over one methionyl residue. The reaction mixture was flushed with N₂ and incubated for 24 h under gentle stirring at 37°C in the dark. The reaction mixture was then treated as described in (Laine *et al.*, 1986).

Cleavage with endoproteinase Arg-C. The protein MC1c (80 nmol) was dissolved in 1 ml of 10 mM ammonium bicarbonate buffer pH 8.1, and digested with endoproteinase Arg-C at 37°C for 4 h, using a final enzyme-to-substrate ratio of 1/50 (w/w). The digestion was stopped by lowering the pH with formic acid and the hydrolysate was freeze-dried prior to fractionation.

Cleavage with endoproteinase Glu-C. 100 nmol of variants a and b and 40 nmol of variant c were dissolved in 1 ml of 0.1 M ammonium acetate buffer pH 4.0 and digested with endoproteinase Glu-C at an enzyme-to-substrate ratio of 1/40 (w/w) for 18 h at 37°C. The digestion was stopped by heating at 100°C for 2 min.

Fractionation of peptides. Peptides generated by cleavage with cyanogen bromide were fractionated by chromatography on a Sephadex G50 superfine column (0.7x70 cm) equilibrated and eluted with 0.01 M HCl at a flow-rate of 6 ml/h.

Enzymatic digests and peptides derived from cleavage with dilute acetic acid were fractionated by reversed-phase high-pressure liquid chromatography on a C₁₈ μBondapak column (Waters Associates) using the gradient systems indicated in the legends of figures.

Amino acid analyses. Peptide samples were hydrolyzed in evacuated sealed tubes at 110°C for 24 h in 6 M HCl. Amino acid analyses were performed on a Beckman 6300 amino acid analyser.

Sequence determination. Microsequencing was carried out on a gas-phase sequencer Applied Biosystems using the O3 RPTH Program. The phenylthiohydantoin derivatives of amino acids were identified by high-pressure liquid chromatography as indicated in (Chauvière *et al.*, 1989).

Nomenclature of peptides. Peptides derived from cleavage of the variants by acetic acid and cyanogen bromide are indicated by the symbols Ac- and CN- respectively, whereas peptides generated by digestion with the endoproteinase Arg-C and with the endoproteinase Glu-C were designated by PR- and PE- respectively. Peptides were numbered according to their position in the amino acid sequence of the proteins.

Secondary structure predictions. The secondary structures were predicted using the methods of Chou and Fasman (1978), Garnier *et al.* (1978), Levin *et al.* (1978) and Gaboriaud *et al.* (1987). Only structures predicted by at least two methods were considered in the proposed structures presented in figure 7.

TABLE I
Amino acid compositions of peptides derived from protein MC1a

Peptide : Residues :	PE-3 44-89	Ac-3 39-87	CN-2 70-89	Protein
Asp	2.2 (2)	1.8 (2)	2.2 (2)	6.2 (7)
Thr	0.8 (1)	1.2 (1)		3.0 (3)
Ser				0.0 (0)
Glu	4.3 (4)	4.2 (4)	3.0 (3)	9.2 (9)
Pro	4.2 (4)	3.3 (4)	1.7 (2)	5.2 (5)
Gly	4.3 (4)	4.8 (4)	1.2 (1)	8.0 (8)
Ala	3.1 (3)	3.8 (3)		9.0 (9)
Val	5.4 (5)	5.2 (5)	2.1 (2)	6.2 (6)
Met	0.8 (1)	0.7 (1)		0.9 (1)
Ile	3.9 (4)	3.8 (4)	2.9 (3)	5.9 (6)
Leu	1.1 (1)	2.1 (2)	1.2 (1)	4.2 (4)
Phe	1.0 (1)	1.1 (1)		3.0 (3)
His	1.0 (1)	1.0 (1)		1.7 (2)
Lys	11.3 (11)	10.1 (11)	4.6 (5)	16.1 (16)
Arg	2.0 (2)	4.0 (4)		8.1 (8)
Trp	+ (2)	+ (2)	+ (1)	2 ^s (2)
Total	46	50	20	89
Yield of recovery	50%	19%	62%	

The numbers in parentheses are from the established sequence. No correction was made for degradative losses of threonine during hydrolysis of peptides. Values for the protein were calculated from previous data (Chartier *et al.*, 1989).

+ : presence evidenced spectrophotometrically.

s : determined spectrophotometrically.

Amino acid compositions of peptides derived from protein MC1b

Peptide :	PE-2	PE-3 & 5	PE-4	CN-2	Protein
Residues :	16-38	39-43 + 82-87	44-81	70-87	
Asp	1.4 (1)	1.0 (1)	2.3 (2)	2.2 (2)	7.0 (7)
Thr	1.7 (2)		0.7 (1)		2.7 (3)
Ser	0.8 (1)		1.0 (1)		2.3 (2)
Glu	2.2 (2)	2.2 (2)	3.1 (3)	2.2 (2)	9.8 (10)
Pro	0.9 (1)		4.0 (4)	1.8 (2)	4.8 (5)
Gly	2.8 (3)	0.9 (1)	2.8 (3)	1.4 (1)	8.5 (8)
Ala	4.1 (4)		3.2 (3)		10.2 (10)
Val	0.9 (1)		3.8 (4)	1.2 (1)	5.2 (5)
Met			0.8 (1)		1.6 (2)
Ile	0.8 (1)	2.2 (2)	1.7 (2)	2.0 (2)	4.7 (5)
Leu	1.0 (1)	2.2 (2)		1.2 (1)	4.0 (4)
Tyr	0.8 (1)				1.0 (1)
Phe	0.9 (1)		0.8 (1)		2.8 (3)
His			0.8 (1)		1.2 (1)
Lys	1.8 (2)	2.0 (2)	7.8 (8)	5.8 (6)	11.8 (12)
Arg	2.0 (2)	1.0 (1)	1.8 (2)		7.0 (7)
Trp			+ (2)	+ (1)	2 ^s (2)
Total	23	10	38	18	87
Yield of recovery	80%	75%	38%	49%	

The numbers in parentheses are from the established sequence. Values are uncorrected for partial destruction or incomplete hydrolysis. Values for the protein were calculated from our previous data (Chartier *et al.*, 1989).

+ : presence evidenced spectrophotometrically.

s : determined spectrophotometrically.

TABLE III

Amino acid compositions of peptides derived from protein MClc

Peptide :	PE-1	PE-2	PE-3	PR-1	CN-2	Protein
Residues :	1-16	17-44	45-90	49-80	70-90	
Asp	3.0 (3)	2.1 (2)	4.0 (4)	3.6 (3)	2.9 (3)	8.7 (9)
Thr		1.1 (1)	1.0 (1)			1.9 (2)
Ser		1.1 (1)	1.2 (1)	0.9 (1)		2.3 (2)
Glu	3.0 (3)	3.1 (3)	5.2 (5)	2.7 (3)	2.0 (2)	10.9 (11)
Pro		0.9 (1)	5.1 (5)	5.7 (5)	1.9 (2)	5.7 (6)
Gly	1.2 (1)	3.1 (3)	3.3 (3)	0.9 (1)	1.2 (1)	7.5 (7)
Ala	1.2 (1)	4.0 (4)	2.4 (2)	2.4 (2)	1.2 (1)	7.5 (7)
Val		0.9 (1)	3.5 (4)	3.0 (3)	2.0 (2)	5.2 (5)
Met	0.9 (1)		0.7 (1)	0.9 (1)		1.7 (2)
Ile	1.1 (1)	1.9 (2)	3.7 (4)	3.0 (3)	1.9 (2)	6.6 (7)
Leu	1.1 (1)	2.1 (2)	2.3 (2)		2.0 (2)	5.0 (5)
Tyr						
Phe	1.0 (1)	2.0 (2)	1.5 (1)	0.7 (1)		3.8 (4)
His			1.0 (1)	0.6 (1)		1.2 (1)
Lys	2.0 (2)	1.7 (2)	8.0 (8)	5.6 (5)	4.0 (4)	11.9 (12)
Arg	2.0 (2)	3.9 (4)	2.5 (2)	0.7 (1)		7.8 (8)
Trp			+ (2)	+ (2)	+ (1)	2 ^s (2)
Total	16	28	46	32	19	90
Yield of recovery	45%	16,7%	17,5%	30%	65%	

Calculations were as indicated in the legend to Table I.

+ : presence evidenced spectrophotometrically.

s : determined spectrophotometrically.

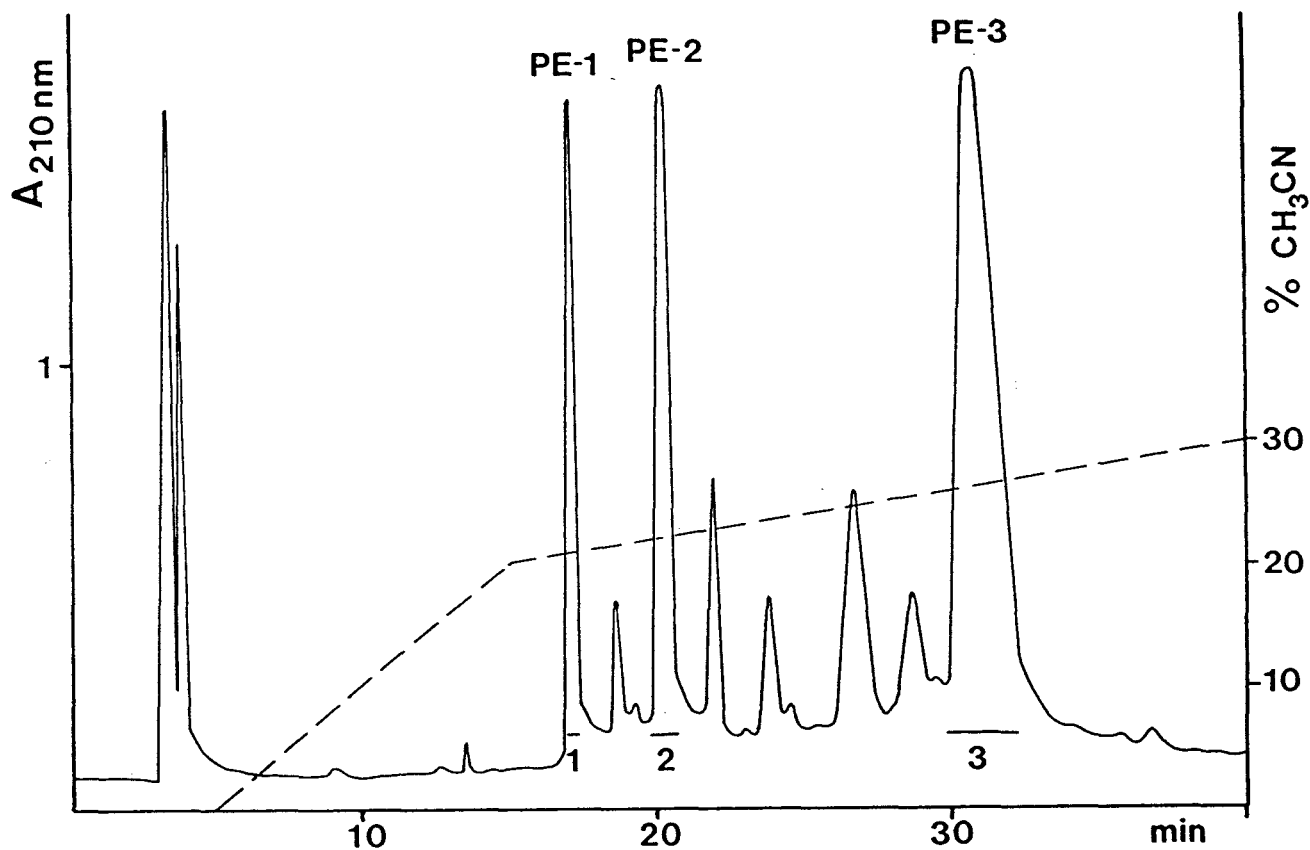


Figure S1 : Fractionation by RP HPLC of peptides obtained by digestion of protein MC1a with endoproteinase Glu-C. Chromatography was performed on a C₁₈ μ Bondapak column (30 x 0.39 cm). 30 nmol of digest were injected. (----) gradient of acetonitrile in 0.05 % trifluoroacetic acid at a flow rate of 1 ml/min ; (—) absorbance at 210 nm. Fractions of interest were pooled as indicated by solid bars and numbered by order of elution. Peptides were numbered according to their position in the sequence.

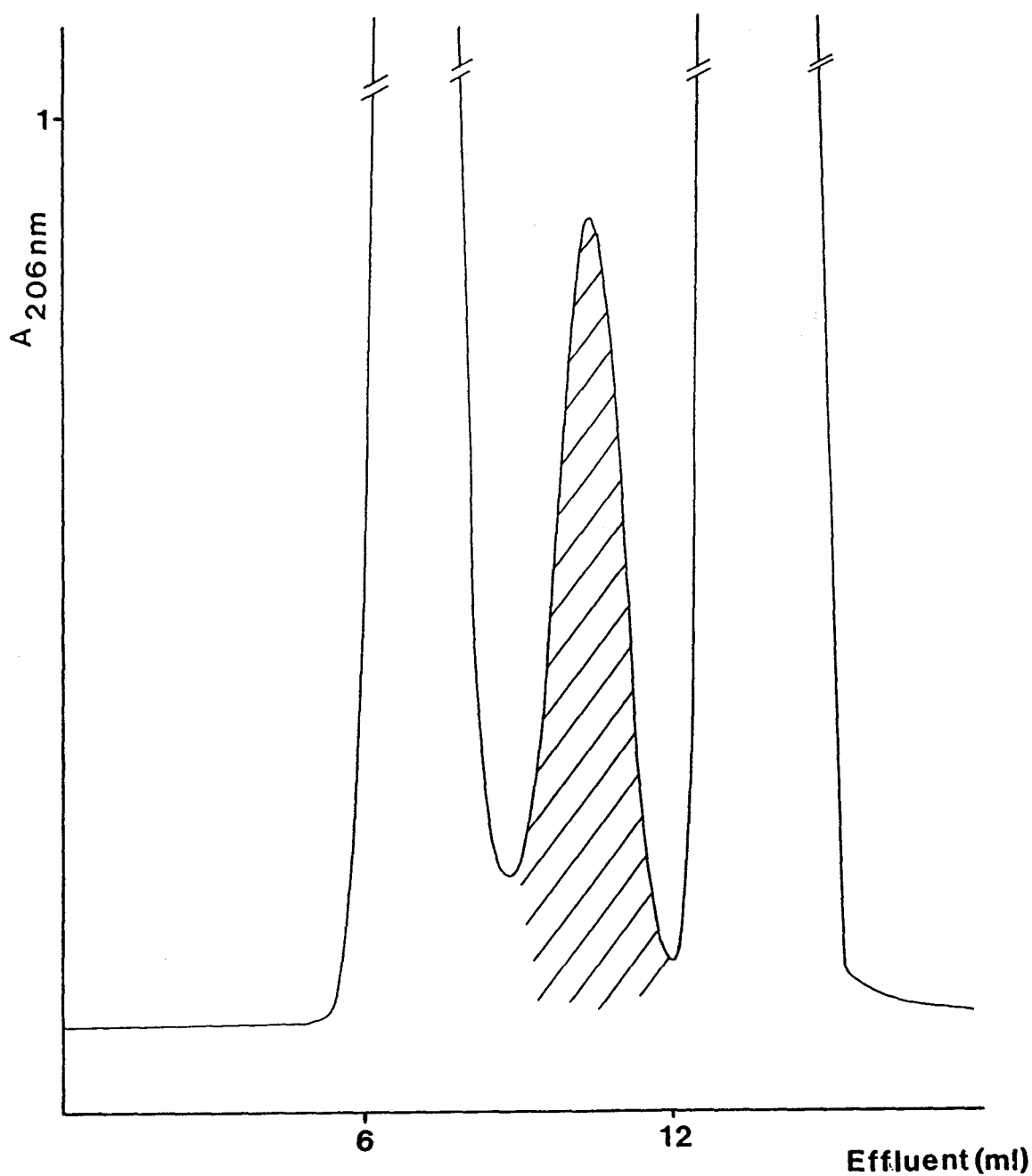


Figure S2 : Fractionation by gel filtration chromatography of peptides derived from protein MC1a by cleavage at methionyl residues. Peptides (100 nmols) were loaded on to a Sephadex G50 superfine column (0.7 x 60 cm). The column was equilibrated and eluted with 0.01 M HCl a a flow rate of 6 ml/h. (—) absorbance at 206 nm. The hatched peak corresponds to the C-terminal fragment CN-2.

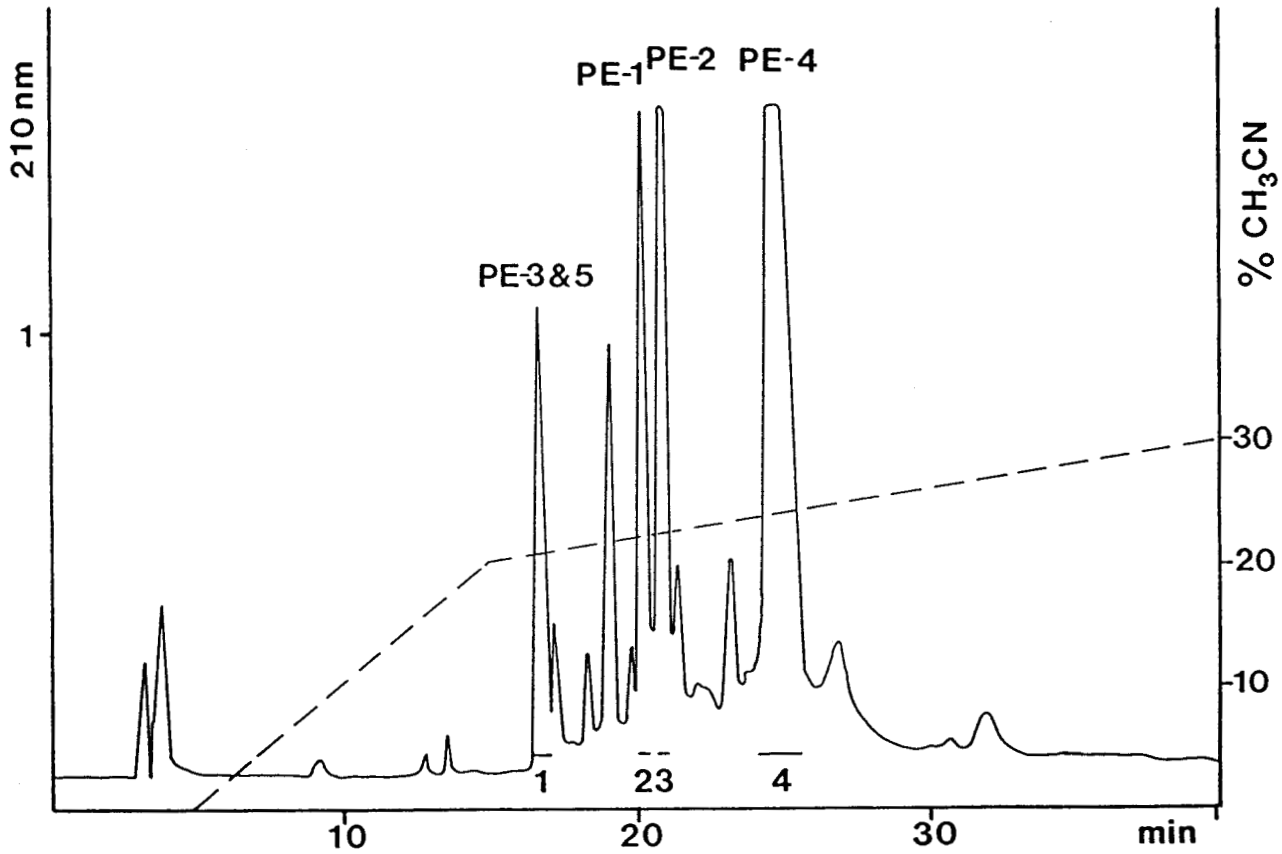


Figure S3 : Fractionation by RP-HPLC of peptides obtained by digestion of protein MClb with endoproteinase Glu-C. 40 nmol of digest were injected. Conditions and numbering of peptides as indicated in legend to figure S1.

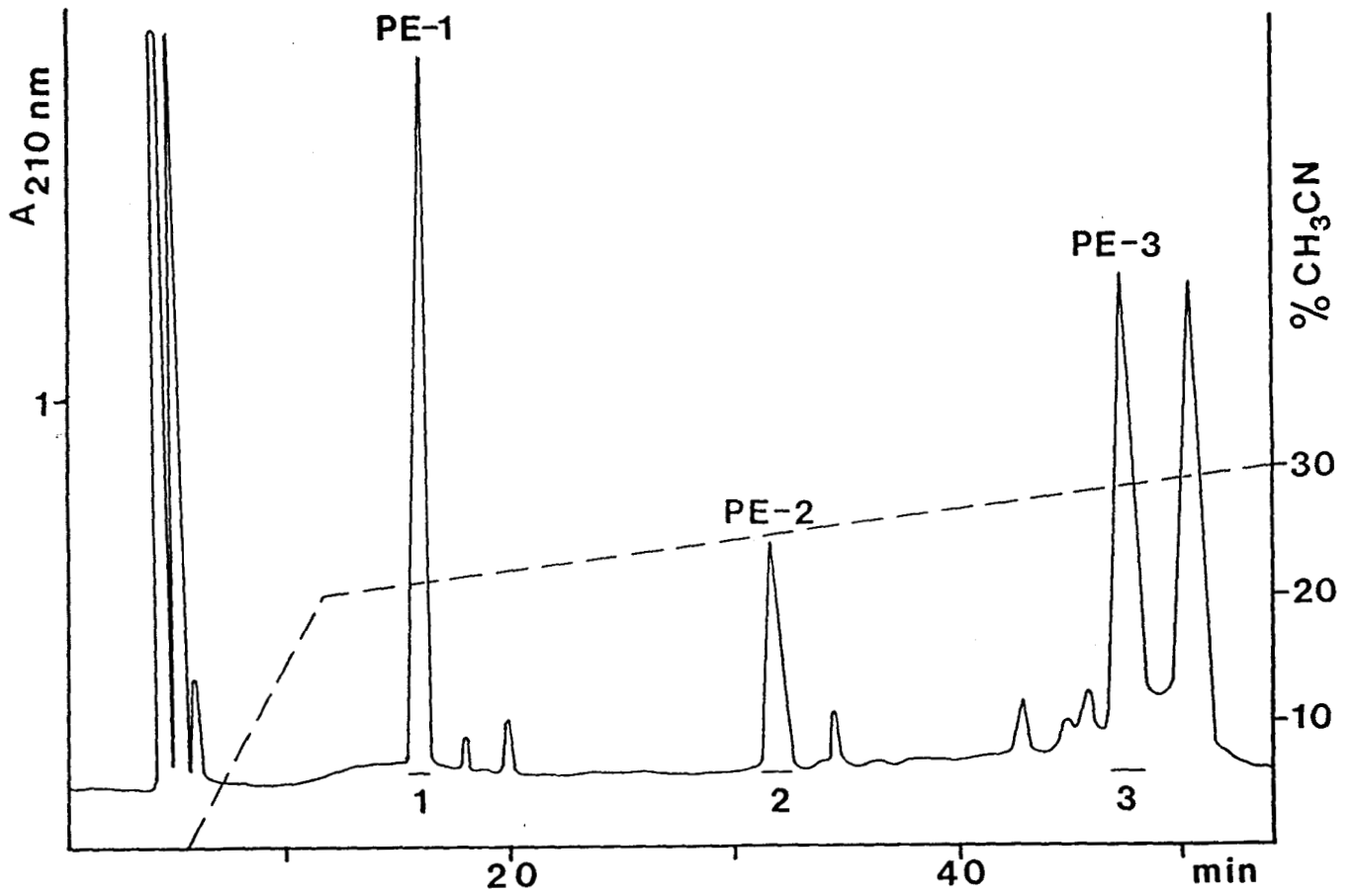


Figure S4 : Fractionation by RP-HPLC of peptides obtained by digestion of protein MC1c with endoproteinase Glu-C. 40 nmol of digest were injected. Conditions and numbering of peptides are as indicated in legend to figure S1.

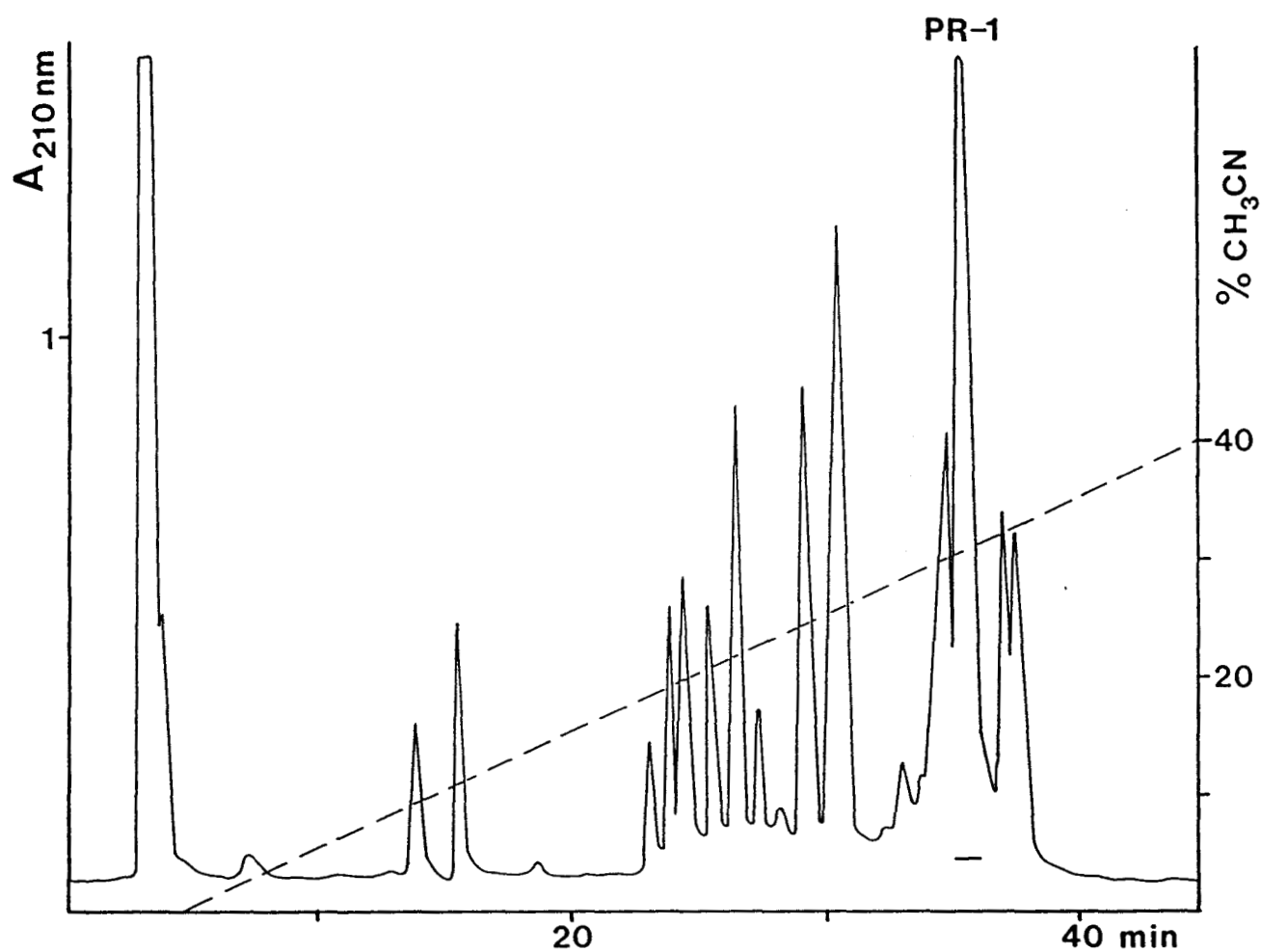


Figure S5 : Fractionation by RP-HPLC of peptides obtained by digestion of protein MClc with endoproteinase Arg-C. 30 nmol of digest were injected. Conditions are as indicated in legend to figure S1. Only peptide PR-1 was used for sequence determination.

III - 2-2 CARACTERISTIQUES DE LA STRUCTURE PRIMAIRE DE LA PROTEINE MC1

La protéine MC1 est une protéine de faible taille moléculaire ($M_r \sim 9.760$ à 10.760) constituée de 87 à 93 résidus d'acides aminés selon son origine. La protéine MC1 se caractérise par un taux élevé de résidus acides (11 à 15 %) et basiques (23 à 29 %) ce qui rend la molécule très hydrophile. Les acides aminés basiques sont répartis tout au long de la séquence de la protéine, certains de ces résidus sont associés en doublets ou dans des séquences du type basique-X-basique où X est un acide aminé quelconque. La séquence centrale (résidus 46-56) est particulièrement basique (charge nette + 5 dans toutes les protéines étudiées) tandis que dans la région carboxy-terminale, les charges positives sont compensées par les charges négatives des résidus d'acide aspartique et d'acide glutamique. La protéine MC1 se caractérise aussi par une accumulation de quatre résidus de proline dans la région 68 à 82. Ces résidus de proline doivent donner à cette région une structure particulière dont nous reparlerons par la suite.

III - 2-3 ETUDE DE LA VARIABILITE STRUCTURALE DE LA PROTEINE MC1

Les études comparatives des séquences (Figure 13) montrent qu'à l'intérieur du groupe des Méthanosarcines thermophiles (*Methanosarcina* sp CHTI 55, MST-A1 et TM1), la protéine MC1 présente une forte invariabilité structurale. La comparaison des cartes peptidiques de la protéine MC1 des souches sp CHTI 55 et TM1 suggère que ces deux protéines sont identiques. Les protéines MC1 des souches sp CHTI 55 et MST-A1 diffèrent par

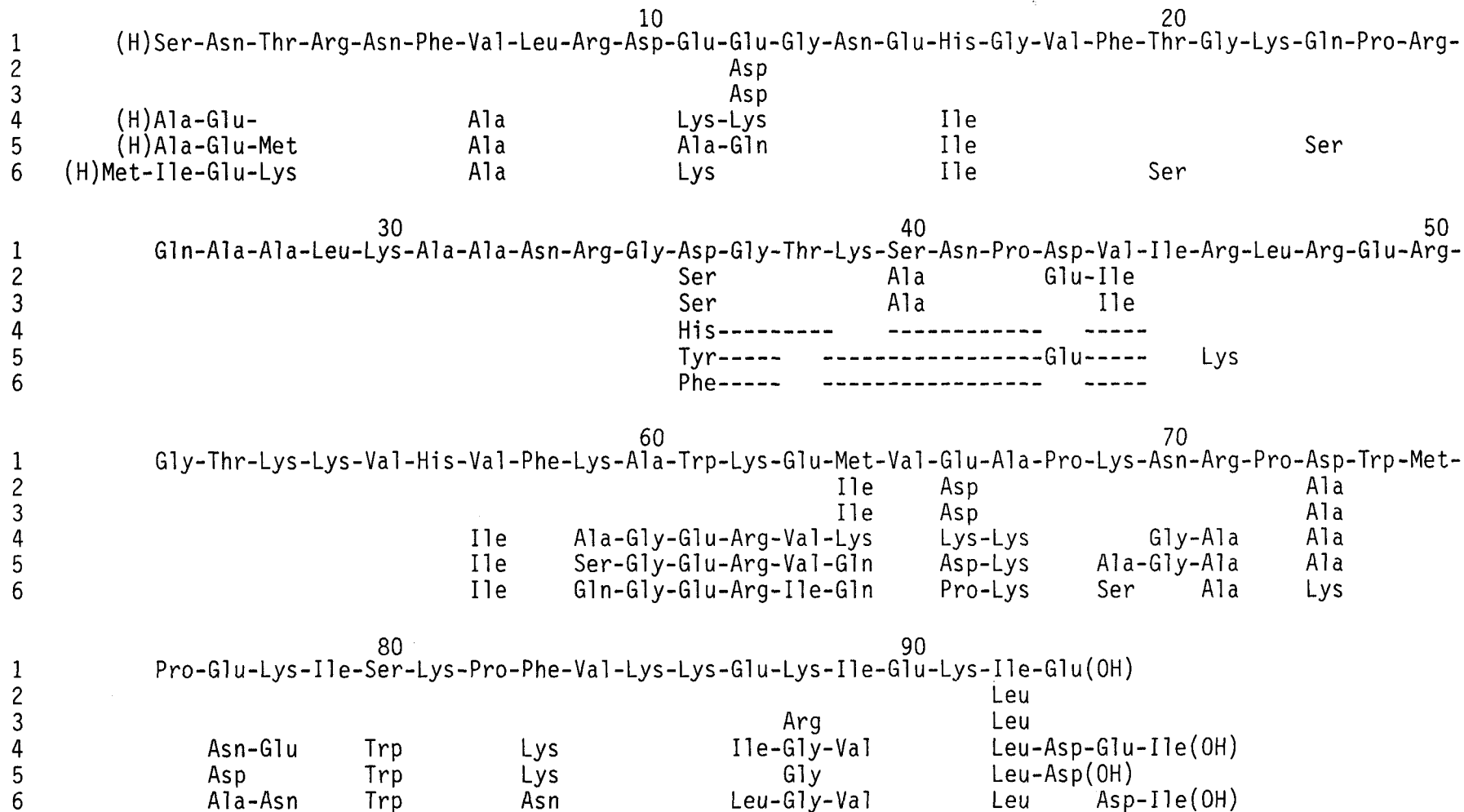


FIGURE 13 - Comparaison de la structure primaire de la protéine MC1 isolée de différentes souches de Methanosarcinaceae.

La protéine MC1 de *Methanosarcina barkeri* (1) est prise comme référence pour les autres protéines. Seules les substitutions sont indiquées. (----) représente des délétions en acides aminés.
 2 : protéine MC1 de *Methanosarcina* MST-A1 selon M. IMBERT (1987); 3 : protéine MC1 de *Methanosarcina* sp CHTI 55 ; 4,5 et 6 : variants a,b et c de *Methanotherix soehngenii*.

une seule substitution conservative Glu → Asp en position 43. Quant aux acides aminés aux positions 36 et 88 de la protéine MC1 de *Methanosarcina* MST-A1, ils n'ont pu être identifiés de manière directe. Toutefois l'analyse des acides aminés des peptides recouvrant ces positions indique la présence probable d'un résidu de sérine en position 36 et d'un résidu de lysine en position 88. Par rapport à la protéine MC1 de ces trois souches thermophiles, la protéine MC1 de la bactérie mésophile *Methanosarcina barkeri* présente davantage de substitutions. Par exemple, lorsque nous la comparons à la protéine MC1 de *Methanosarcina* sp CHTI 55 nous dénombrons 9 substitutions dont 7 sont conservatives. Parmi ces substitutions une seule nécessite le changement de deux des bases du codon.

Quand nous étendons la comparaison au genre *Methanotherix*, les changements deviennent plus nombreux et plus importants. On observe une variabilité importante entre les trois variants de la protéine MC1 de *Methanotherix soehngeni*. Par rapport au variant b, le variant a diffère par cinq substitutions conservatives, onze substitutions non conservatives et l'addition de deux acides aminés à son extrémité C-terminale tandis que le variant c présente dix substitutions conservatives, onze substitutions non conservatives et l'addition d'un résidu de méthionine en position amino-terminale et de deux résidus d'acides aminés en position C-terminale (voir figure 13).

Le bilan des différences entre les protéines MC1 qui ont été complètement séquencées, en prenant la protéine MC1 de *Methanosarcina barkeri* comme référence, est présenté dans le tableau IV. Par rapport à cette structure de référence, le nombre de différences est de 36 pour les variants a et b et de 37 pour le variant c de la protéine MC1 de *Methanotherix soehngeni*.

TABLEAU IV

Bilan des changements dans la séquence des protéines MC1 comparées à la protéine MC1 de *Methanosarcina barkeri*

	Changements		Délétion	Addition	Total
	conservatifs	non conservatifs			
<i>Methanosarcina</i> sp CHTI 55					
MC1	6	3			9
<i>Methanotherix soehngeni</i>					
MC1a	7	21	6	2 en C-t	36
MC1b	11	19	6		36
MC1c	6	22	6	1 en N-t 2 en C-t	37

L'alignement des séquences fait ressortir deux régions hautement conservées, séparées par une région de délétions et une région très variable correspondant au tiers C-terminal de la protéine. La région de délétions que l'on trouve dans les variants de la protéine MC1 de *Methanotherrix soehngeni* concerne en fait trois segments dont la longueur varie entre 1 et 4 acides aminés qui représentent ensemble 6 acides aminés. Les deux séquences hautement conservées (résidus 13 à 35 et 45 à 58) jouent certainement un rôle important dans la protéine. Elles correspondent au domaine structuré de la protéine comme l'indique la synthèse des quatre méthodes de prédictions de structures secondaires utilisées* (Figure 14). Ce domaine comprend trois feuillets β et une hélice α séparés chacun par un coude β . Dans le tiers C-terminal variable de la protéine, tous les acides aminés de la séquence 59 à 73 changent excepté le résidu de valine en position 65 et les deux résidus de proline en positions 68 et 72. Il faut noter également que les deux résidus de proline en positions 76 et 82 sont aussi strictement conservés. Les méthodes de prédictions de structures secondaires indiquent de fortes probabilités de coudes β au niveau des prolines 68, 72 et 76. La stricte conservation de cette structure en coudes β juxtaposés semble indiquer qu'elle joue un rôle essentiel dans la fonction de la protéine MC1. L'apparition d'un résidu de glycine en position 60 dans les variants de la protéine MC1 de *Methanotherrix soehngeni* augmente la probabilité d'un coude β et semble responsable de la transformation du fragment d'hélice α_{58-66} en un segment en feuillet β_{63-66} . La présence d'un résidu de proline supplémentaire en position 66 dans la protéine MC1c diminue fortement les probabilités de structure dans la zone 63-66 de cette protéine.

* Les méthodes de prédictions de structures utilisées sont les méthodes de : CHOU et FASMAN (1978), LEVIN *et al* (1986), GARNIER *et al* (1978), GABORIAUD *et al* (1987).

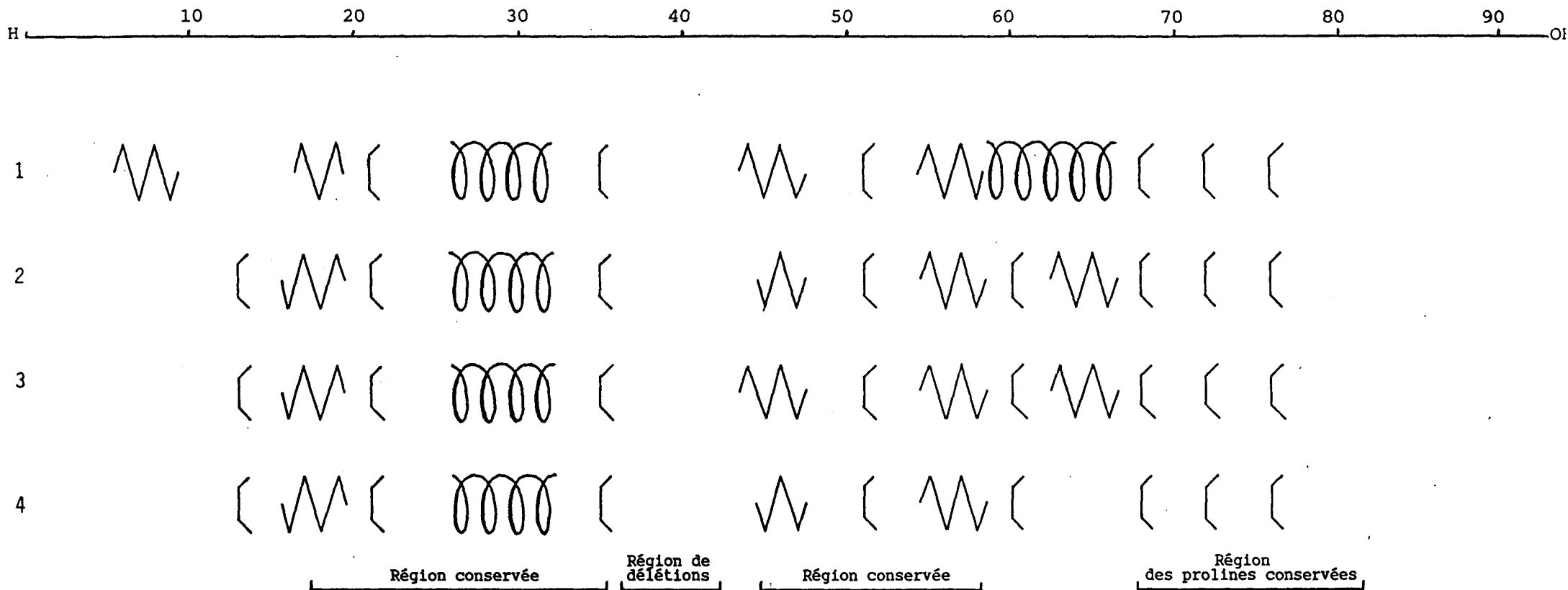





FIGURE 14 - Structures secondaires de la protéine MC1 isolée des différentes souches de Methanosarcinaceae déterminées par les méthodes prédictives.

1 : protéine MC1 de *Methanosarcina barkeri* et de *Methanosarcina* sp CHTI 55 ; 2,3,4 : variants a,b et c de la protéine MC1 de *Methanotherxix soehngeni*.

Les structures en hélice α , en feuillet β et en coude β sont représentées respectivement par les symboles :

 ,
  ,
 
 Les résultats présentés pour la protéine MC1 de *Methanosarcina barkeri* sont de IMBERT (1987)

III - 2-4 ETUDE COMPARATIVE DE LA PROTEINE MCl ET D'AUTRES
PROTEINES SE LIANT A L'ADN

La comparaison de la séquence en acides aminés de la protéine MCl avec celle d'autres protéines chromosomales eucaryotes et procaryotes a fait l'objet de la publication n°7 :

Features of the primary structure of the *Methanosarcina barkeri* chromosomal protein HMb compared to those of eukaryotic and prokaryotic chromosomal proteins.

IMBERT, M., CHARTIER, F., LAINE, B. and SAUTIERE, P. (1986)
in *Biology of Anaerobic Bacteria* (H.C. Dubourguier *et al*) pp 236-241, Elsevier Science Publishers B.V., Amsterdam The Netherlands.

La protéine MCl se différencie des protéines chromosomales d'eubactéries et d'eucaryotes par l'absence de domaines hydrophobes. De plus, elle se différencie des "DNA-binding proteins II" d'eubactéries par un taux important de résidus chargés et par sa charge positive nette relativement élevée. Bien que son caractère basique la rapproche des histones et de la protéine HTa de *Thermoplasma acidophilum*, elle ne montre aucune homologie de séquence avec ces protéines. Les "DNA-binding proteins" 7a, 7b et 7e isolées de *Sulfolobus acidocaldarius* (CHOLI et coll., 1988) présentent des pourcentages d'acides aminés basiques proches de ceux des protéines MCl mais, mise à part une séquence Ala-Pro-Lys que l'on retrouve dans la protéine MCl, ces protéines ont des structures primaires très différentes de celles de la protéine MCl.

236

Biology of Anaerobic Bacteria, edited by H.C. Dubourguier et al.
Elsevier Science Publishers B.V., Amsterdam, 1986 — Printed in The Netherlands

FEATURES OF THE PRIMARY STRUCTURE OF THE *METHANOSARCINA BARKERI* CHROMOSOMAL
PROTEIN HMB COMPARED TO THOSE OF EUKARYOTIC AND PROKARYOTIC CHROMOSOMAL PROTEINS

M. IMBERT, F. CHARTIER, B. LAINE and P. SAUTIERE

Unité Associée 409 CNRS, Institut de Recherches sur le Cancer, Place de Verdun
59045 Lille Cédex, France.

SUMMARY

The basic protein HMB is the major component of the *Methanosarcina barkeri* nucleoprotein complex. The main features of protein HMB primary structure are described and compared to those of eukaryotic, eubacterial and archaeobacterial chromosomal proteins. Protein HMB is not homologous to any of these chromosomal proteins.

INTRODUCTION

The composition and the organization of the chromosome in eukaryotes and eubacteria are quite different (ref.1). Inside the nucleus of eukaryotic cells packaging of DNA is mainly ensured by basic proteins called histones, whereas in eubacteria, HU-type proteins are involved in DNA condensation. The discovery of archaeobacteria, a line of descent distinct from eubacteria and eukaryotes (ref. 2), led several authors to investigate the protein composition of their nucleoid. Chromosomal proteins have been isolated from three strains of thermoacidophiles : *Thermoplasma acidophilum* (ref.3), *Sulfolobus acidocaldarius* (ref.4), and *Sulfolobus solfataricus* (ref.5). Up to now, only one chromosomal protein (protein HMB) has been characterized from a methanogenic bacteria : *Methanosarcina barkeri* strain MS (ref.6). The main features of protein HMB primary structure are described and compared to those of eukaryotic, eubacterial and archaeobacterial chromosomal proteins.

RESULTS AND DISCUSSION

Protein HMB is the major component of the acid extract of the *Methanosarcina barkeri* deoxyribonucleoprotein complex. It is a polypeptide of 93 amino acid residues (Mr 10,757) which tightly binds to the double stranded DNA : a sodium chloride molarity of 0.47 M is necessary to desorb the protein HMB from a column of *E.coli* DNA bound to cellulose. The protein HMB has a high content of basic residues (Lys 14 ; Arg 8 and His 2) and of acidic residues (Asp 4 and Glu 10). Half of the basic residues are clustered in sequences such as B-B-X-B (residues 53-56 and 85-88) or B-X-B (residues 46-48, 48-50, 70-72) where B and X respec-

238

TABLE 1
Chemical characteristics of chromosomal proteins

	EUKARYOTES				EUBACTERIA	ARCHAEBACTERIA			
	H2B	H2A	H3	H4	HU	HTa	HMb	HSa	7kD
Molecular size	13770	14000	15300	11280	9500	9930	10760	14500	7000
Basic residues %	25	23	24	26	17	22	27	16	27
Net charge	+18	+16	+20	+16	+4	+13	+10	nd	+5

The proteins HU, HTa, HMb, HSa and 7kD are respectively isolated from *E.coli*, *T.acidophilum*, *M.barkeri*, *S.acidocaldarius* and *S.solfataricus*. The values given for histones H2B, H2A, H3 and H4 are those of calf thymus histones.

stitutions by comparison with *E.coli* protein HU-1 but the amino acid residues which play a crucial role in the function of the protein are strictly conserved (ref.10). With 102 residues (Mr 11,280) histone H4 is the smallest of the five species of histones, it is the most conserved during evolution and constitutes with histone H3 a tetramer (H3-H4)₂ which is essential in the formation and the stability of the nucleosome protein core.

The comparison of the hydrophobicity diagrams indicates that the protein HMb is radically different from the other proteins (Fig.2). The protein HMb exhibits a marked hydrophilic character due to its high content of charged residues (42% of the total amino acid residues) distributed all along the polypeptide chain. On the contrary large hydrophobic domains are encountered in the three other proteins ; they play an important role in the conformation of these proteins and in the formation of oligomers. In HU-type proteins, including the archaeobacterial protein HTa, these hydrophobic domains are well conserved and are involved in hydrophobic interactions between the structural segments of the protein (α helices or β stranded sheets). These hydrophobic interactions give a globular shape to the dimer HU which corresponds to the functional state of this chromosomal protein (ref.10). On the other hand, the protein HMb behaves as a monomer and preliminary data indicate that, when dissociated from the DNA, it does not exhibit a conformation as well organized than that of HU-type proteins.

The protein HMb also differs from the HU-type proteins in its amount of basic residues and in its net charge (Table 1). Moreover in protein HMb, half of the basic residues are clustered and constitute highly basic sequences (Fig.3). Such a distribution is not encountered in eubacterial HU-type proteins where the basic residues are scattered all along the polypeptide chain. Only the *T.acidophilum* protein HTa displays a strong basicity with a net charge equal to +13 and with the presence of clusters of two or three basic residues.

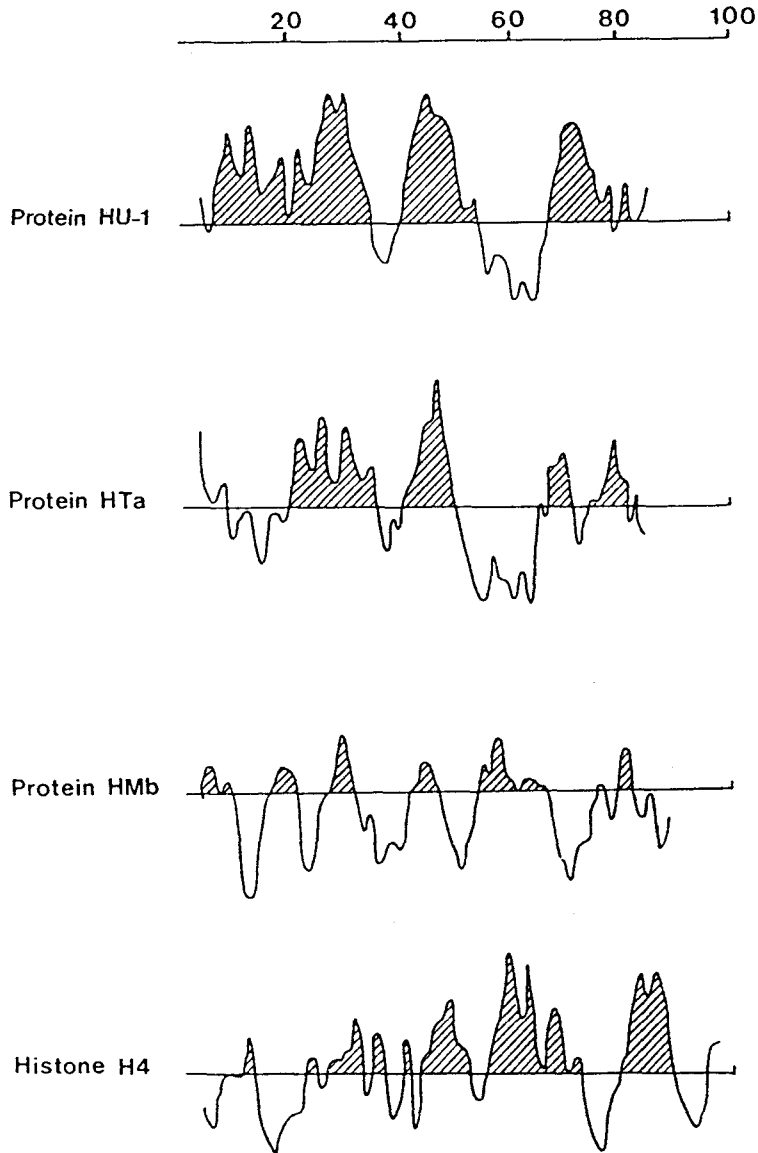


Fig.2. Hydrophobicity diagrams of proteins HU-1 (*E.coli*), HTa (*T.acidophilum*), HMB (*M.barkeri*) and calf thymus histone H4. The diagrams were drawn according to (ref.11) and using a computer program. Hydrophobic sequences appear in shaded areas and hydrophilic sequences in non-shaded areas.

The amount of basic residues in protein HMB is similar to that found in histones. However a higher content of dicarboxylic amino acid residues makes the net charge of protein HMB lower than that of histones. Moreover, unlike in protein HMB, acidic residues are not found in the sequences of histones where basic residues are clustered. Consequently the basic domains of histones have a very high net charge : the amino-terminal half of histone H4 (residues 1 through 45)

240

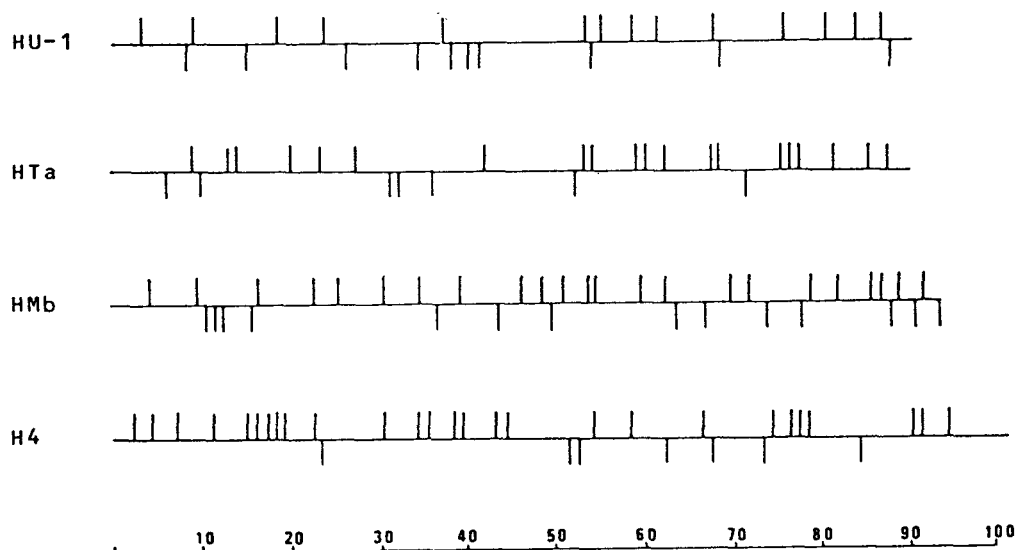


Fig.3. Distribution of basic (\uparrow) and acidic (\downarrow) residues in proteins HU-1 (*E. coli*)(ref.7), HTa (*T.acidophilum*)(ref.8), HMb (*M.barkeri*) and histone H4 (calf thymus)(ref.9).

has a net charge of +13. A last character differentiates protein HMb from histones : it is the presence in protein HMb of two residues of tryptophan, an amino acid residue which is not encountered in any histone.

In conclusion, by its high basicity and the presence of basic clusters, the protein HMb appears to be closer to eukaryotic histones and to *T.acidophilum* protein HTa than to eubacterial HU-type proteins. However any sequence homology can be observed between protein HMb and histones or protein HTa.

Work is in progress to localize the protein HMb in the nucleoid of *Methanosarcina barkeri* using an immunochemical method. We also are looking for the conformational state of the protein and the way it can condense the DNA of this bacteria.

The authors are grateful to Dr J.P.Touzel from the Station de Technologie Alimentaire, INRA, Villeneuve d'Ascq, for valuable discussions and for growing the bacteria. They are indebted to A.Lemaire-Poitau, T.Ernout and A.Hemez for skiful technical assistance. This work was supported by grants from the Centre National de la Recherche Scientifique and the Pôle des Anaérobies de la Région Nord-Pas de Calais.

REFERENCES

- 1 D.E. Pettijohn, Structure and properties of the bacterial nucleoid, Cell, 30 (1982) 667-669.
- 2 C.R. Woese and G.E. Fox, Phylogenetic structure of the prokaryotic domain : the primary kingdoms, Proc.Natl.Acad.Sci.USA,74 (11)(1977) 5088-5090.

- 3 D.G. Searcy, Histone-like protein in the prokaryote *Thermoplasma acidophilum*, *Biochim.Biophys.Acta*, 395 (1975) 535-547.
- 4 G.R. Green, D.G. Searcy and R.J. De Lange, Histone-like protein in the archaeobacterium *Sulfolobus acidocaldarius*, *Biochim.Biophys.Acta*, 741 (1983) 251-257.
- 5 M. Kimura, J. Kimura, P. Davie, R. Reinhardt and J. Dijk, The amino acid sequence of a small DNA-binding protein from the archaeobacterium *Sulfolobus solfataricus*, *FEBS Letters*, 176(1)(1984) 176-178.
- 6 F. Chartier, B. Laine, P. Sautière, J.P. Touzel and G. Albagnac, Characterization of the chromosomal protein HMB isolated from *Methanosarcina barkeri*, *FEBS Letters*, 183 (1)(1985) 119-123.
- 7 B. Laine, D. Kničcik, P. Sautière, G. Biserte and M. Cohen-Solal, Complete amino acid sequences of DNA-binding proteins HU-1 and HU-2 from *Escherichia coli*, *Eur.J.Biochem.*, 103 (1980) 447-461.
- 8 R.J. De Lange, L.C. Williams and D.G. Searcy, A histone-like protein (HTa) from *Thermoplasma acidophilum* II. Complete amino acid sequence, *J.Biol.Chem.* 256 (2)(1981) 905-911.
- 9 R.J. De Lange, D.M. Fambrough, E.L. Smith and J. Bonner, Calf and Pea histone IV. II. The complete amino acid sequence of calf thymus histone IV ; presence of ϵ -N-acetyllysine, *J.Biol.Chem.*, 244 (1969) 319-334.
- 10 I. Tanaka, K. Applet, J. Dijk, S.W. White and K.S. Wilson, 3-Å resolution structure of a protein with histone-like properties in prokaryotes, *Nature*, 310 (1984) 376-381.
- 11 J. Kyte and R. Doolittle, A simple method for displaying the hydropathic character of a protein, *J.Mol.Biol.*, 157 (1982) 105-132.

III - 3 CONCLUSION

Nous avons déterminé la structure primaire des protéines MCl de *Methanosarcina barkeri*, de *Methanosarcina* sp CHTI 55 et des trois variants de la protéine MCl de *Methanotherix soehngeni*.

La protéine MCl ne présente d'homologie de séquence avec aucune autre protéine chromosomale isolée d'autres organismes. Elle se caractérise par l'absence de zone hydrophobe due à la distribution de ses résidus d'acides aminés acides et basiques tout le long de la molécule, et par une succession de prolines dans sa région C-terminale.

Nous avons pu montrer que les régions très conservées de la protéine MCl correspondent aux régions structurées de la protéine et que le tiers C-terminal dont la séquence est très variable, conserve une structure en coudes β juxtaposés. Avec un faible taux d'hélice α et quelques courts fragments en feuillet β , la protéine MCl, considérée isolément, apparaît ainsi peu structurée. Il est toutefois possible qu'en interaction avec l'ADN, la conformation de la protéine soit différente.

CHAPITRE IV

ÉTUDE DES INTERACTIONS DE LA PROTÉINE MC1 AVEC L'ADN

Les études d'interactions protéine MCl-ADN ont été réalisées avec la protéine MCl de *Methanosarcina* sp CHTI 55 qui est plus facile à purifier que les autres protéines MCl. La protéine MCl utilisée pour ces études a été préparée en conditions non dénaturantes (IMBERT et coll., 1987).

IV - 1 EFFET DE LA PROTÉINE MCl SUR LA STABILITÉ THERMIQUE DE L'ADN

L'effet de la protéine MCl sur la stabilité thermique de l'ADN a été décrit dans la publication n°2 (page 49). L'interaction de la protéine MCl avec l'ADN provoque l'apparition d'un second point de fusion à 70°C et une diminution du signal correspondant à la fusion de l'ADN nu. L'amplitude du second point de fusion est directement proportionnelle à la quantité de protéine présente (Figure 15 A & B). L'application de la méthode de calcul de LI (1978) nous a permis d'établir que la protéine MCl est capable de protéger une longueur d'ADN de huit paires de bases. Des profils de fusion biphasiques de l'ADN sont également observés en présence de l'octamère d'histones (WEISCHET et coll., 1978), de protamines (WILLMITZER et coll., 1977) ou de protéines HMG (YU et coll., 1977). La "DNA-binding protein II" (ou protéine HU) d'*Escherichia coli* provoque un glissement du point de fusion et non pas un profil biphasique (Figure 15-1C). Ce résultat suggère qu'elle fait intervenir un type d'interactions différent de celui de la protéine MCl. La faible quantité de protéine MCl associée à l'ADN (rapport protéine/ADN 1 : 10 en masse) (IMBERT et coll., 1987), suggère que le rôle de la protéine MCl dans la bactérie ne serait pas d'assurer la stabilité thermique de l'ADN. D'autre part, la force ionique intracellulaire suffit à empêcher la fusion de l'ADN à la température de croissance de la bactérie.

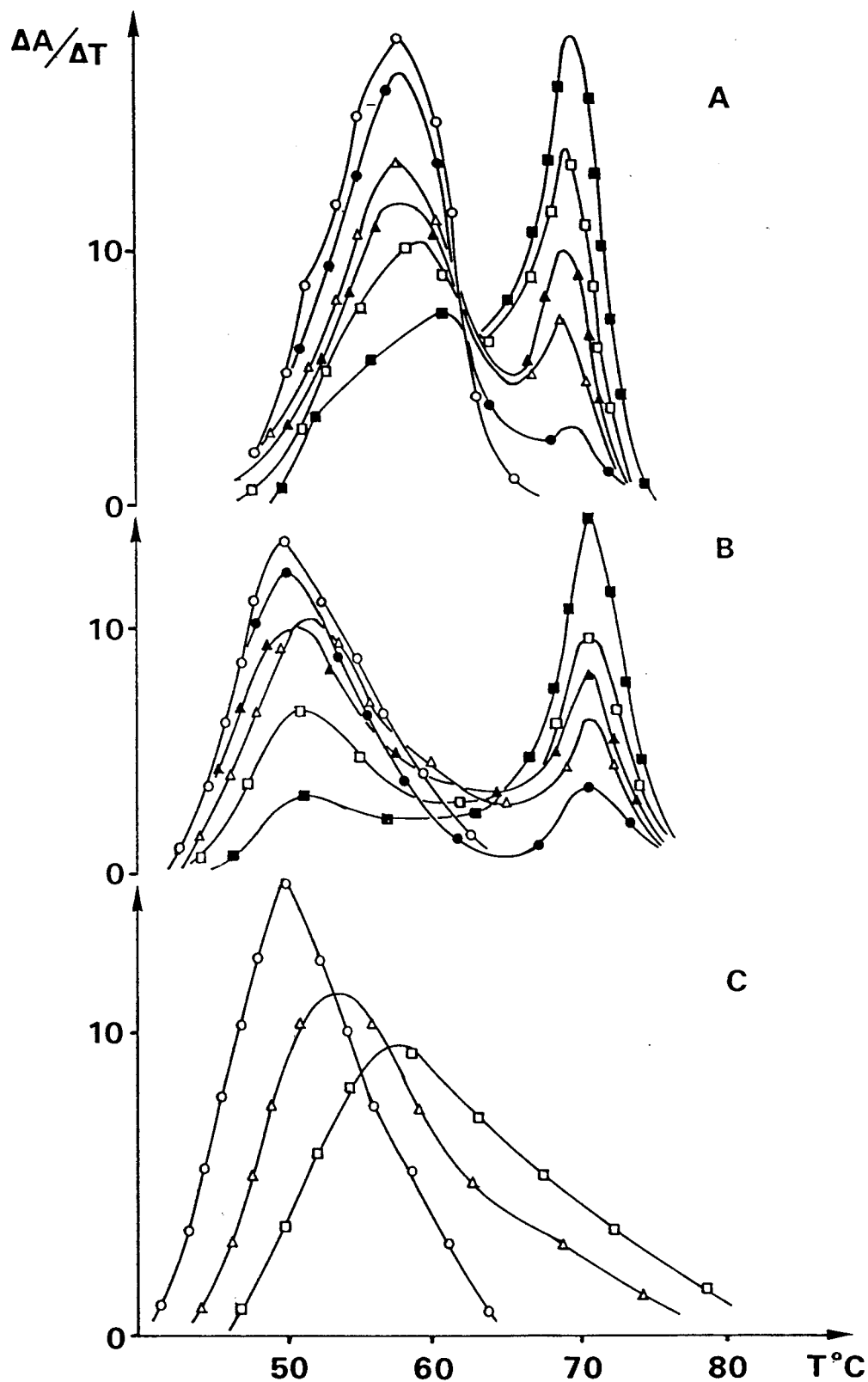


FIGURE 15 - Courbes des dérivées des profils de fusion de l'ADN en présence de protéine.

A : Protéine MCl et ADN préparés à partir de *Methanosarcina* sp CHTI 55 ; B : Protéine MCl de *Methanosarcina* sp CHTI 55 et ADN d'érythrocyte de poulet ; C : "DNA-binding protein II" d'*Escherichia coli* et ADN d'érythrocyte de poulet. Les rapports pondéraux protéine/ADN sont : 0 (○—○) ; 0,22 (●—●) ; 0,55 (△—△) ; 0,77 (▲—▲) ; 1,12 (□—□) ; 1,60 (■—■). Les conditions expérimentales sont décrites dans la publication n°2.

Nous pensons donc que l'effet observé n'est probablement qu'une conséquence d'un rôle structural de la protéine sur l'ADN. La conformation de l'ADN induite par la protéine serait plus stable que celle de l'ADN nu. De plus, nous avons montré que la protéine MCl favorise la renaturation de l'ADN.

IV - 2 EFFET DE LA PROTÉINE MCl SUR LA TRANSCRIPTION DE L'ADN

L'effet de la protéine MCl sur la transcription de l'ADN a été décrit dans la publication n°5. La protéine MCl ne possède pas d'activité ARN polymérase, ni d'affinité pour les nucléotides libres. A des rapports pondéraux protéine MCl/ADN proches du rapport déterminé dans le complexe désoxyribonucléoprotéique de *Methanosarcina barkeri*, la protéine MCl favorise fortement la synthèse d'ARN *in vitro* par l'ARN polymérase d'*Escherichia coli* (Figure 16). Par contre, on observe rapidement une chute de la synthèse d'ARN dès que le rapport protéine MCl/ADN augmente.

L'effet de la protéine MCl sur la transcription résulte probablement de son rôle structural sur l'ADN suggéré par les études de stabilité thermique. La structure de l'ADN en interaction avec la protéine MCl serait plus stable que celle de l'ADN nu et favorable à l'action de l'ARN polymérase. L'inhibition de transcription observée avec des rapports protéine/ADN plus élevés correspondrait à une trop grande structuration de l'ADN.

Il serait intéressant de poursuivre ces travaux par des études complémentaires de l'effet de la protéine MCl sur la transcription et la répllication de l'ADN et par des études physico-chimiques visant à observer les changements de conformation de l'ADN en présence de protéine MCl.

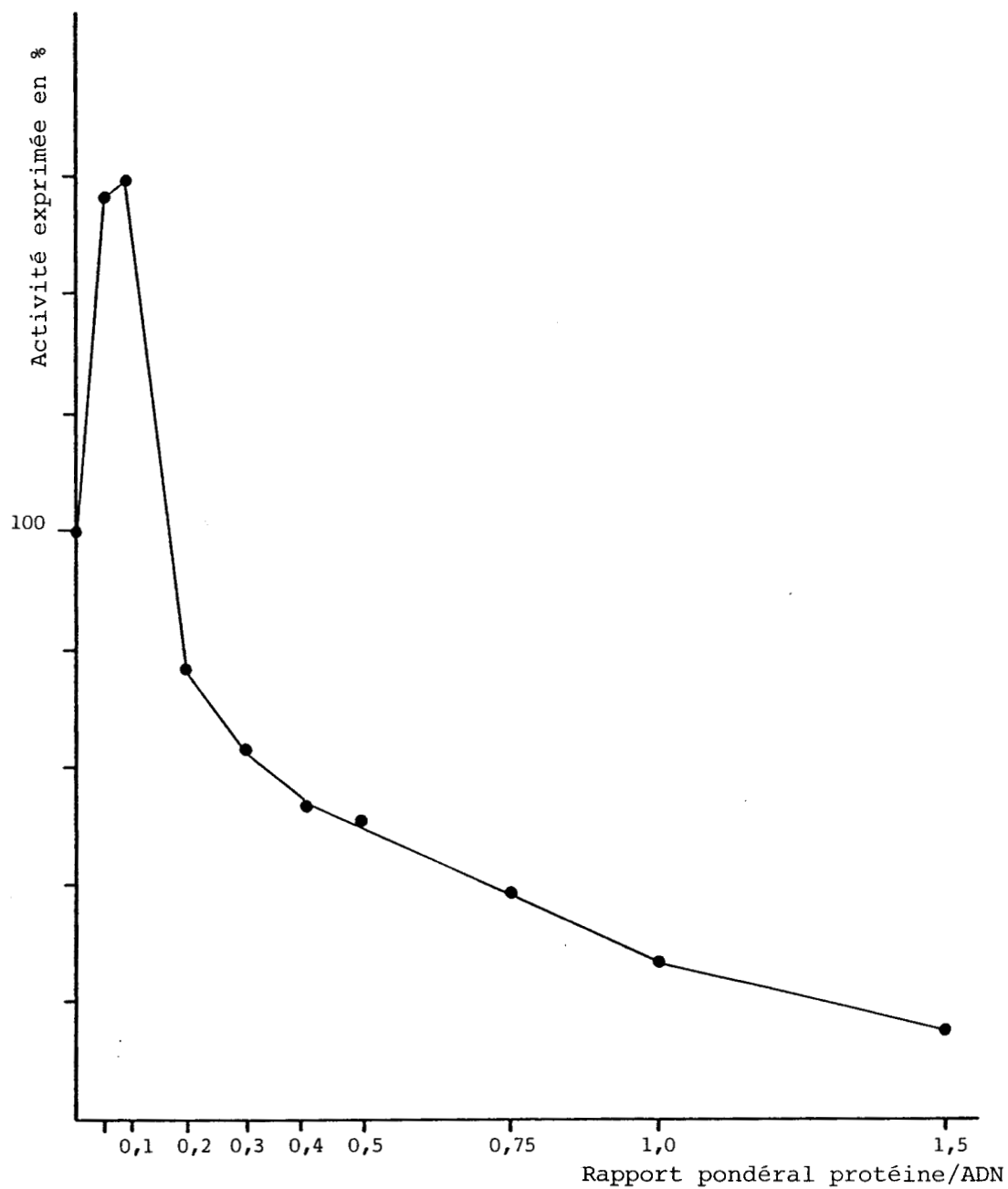


FIGURE 16 - Activité de l'ARN polymérase d'*Escherichia coli* en fonction de la quantité de protéine MCl ajoutée à l'ADN.

Les conditions expérimentales sont décrites dans la publication n°5.

IV - 3 DÉTERMINATION DU DOMAINE DE LA PROTÉINE MCl QUI INTERAGIT AVEC L'ADN

Ce travail a été entrepris avec Bernard LAINE en collaboration avec le Docteur Michel CHARLIER au Centre de Biophysique Moléculaire du CNRS (Orléans).

IV - 3-1 PRINCIPE DE LA METHODE

La méthode que nous avons choisie consiste à ponter la protéine et l'ADN par une liaison covalente. Le complexe ainsi obtenu est ensuite traité par une endoprotéase qui libère les peptides non impliqués dans le pontage. L'analyse de ces peptides permet par déduction d'identifier la séquence peptidique pontée à l'ADN. Le schéma de cette étude est présenté dans la figure 17. Nous avons utilisé un pontage photochimique qui a l'avantage de ne ponter que les acides aminés à proximité immédiate de l'ADN dans le complexe ADN-protéine. Cette technique évite l'introduction de réactifs chimiques étrangers au complexe ADN-protéine.

Pour réaliser les pontages, il faut utiliser un ADN dont les thymidines ont été substituées par de la 5 bromo-désoxyuridine (5BrU). Cet ADN est obtenu par la culture de bactéries *Escherichia coli* souche B/r T⁻ dans un milieu enrichi en 5BrU (4 mg/ml). L'isolement de l'ADN se fait par la méthode de MARMUR (1961) et le pourcentage de substitutions est déterminé par la mesure de la densité de flottation sur CsCl (HACKETT et HANAWALT, 1966). Les taux de substitutions obtenus sont compris entre 70 et 80 %. Les complexes sont irradiés avec une lampe au mercure équipée de filtres ne laissant

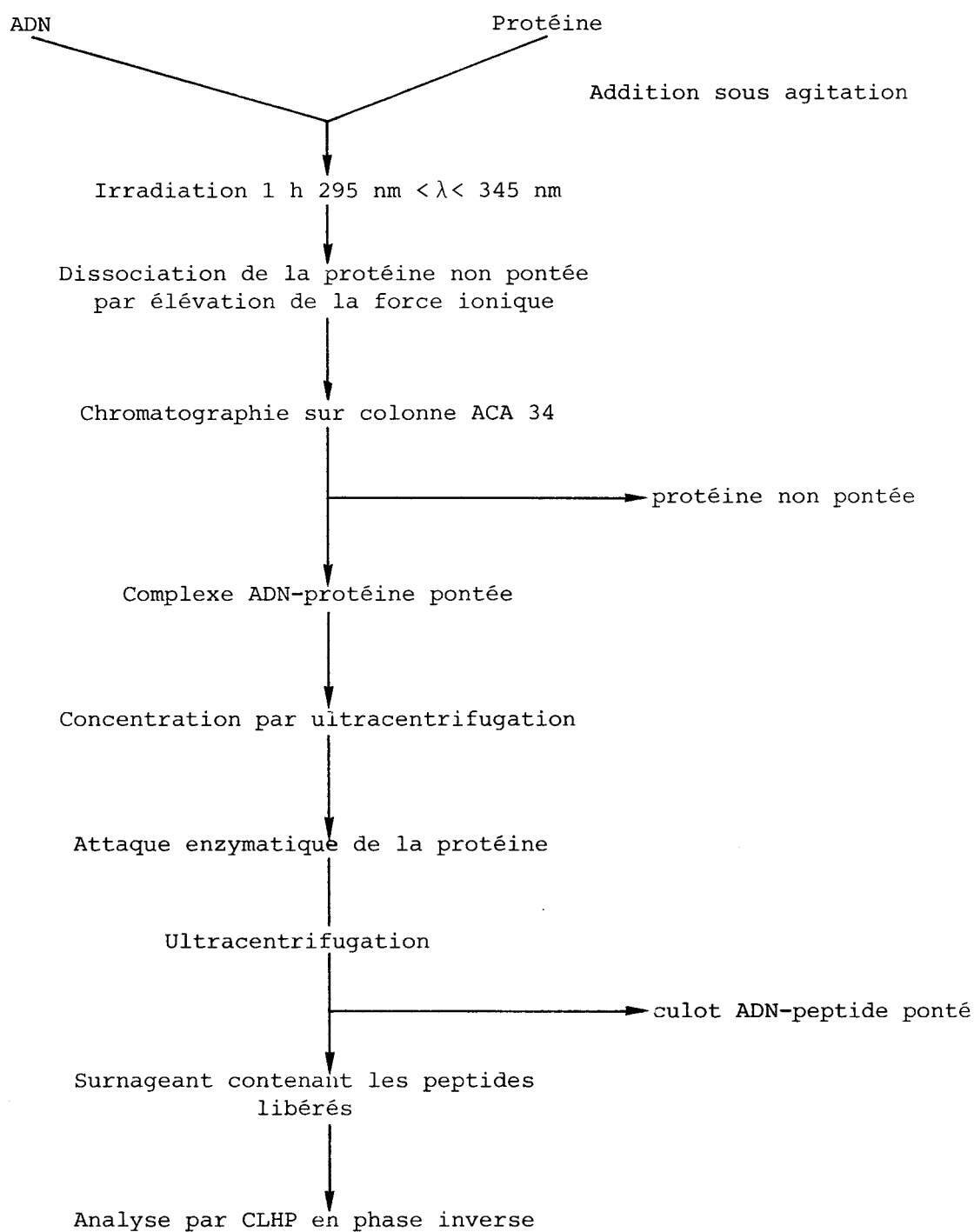


FIGURE 17 - Schéma de la méthode d'étude du domaine d'interaction de la protéine MCl avec l'ADN.

passer que les rayonnements de longueur d'onde comprise entre 295 et 345 nm. Ainsi, ni les bases normales de l'ADN, ni les acides aminés aromatiques de la protéine ne sont photolysés de façon significative. Seule la 5BrU absorbe à ces longueurs d'ondes, et donne lieu, par libération d'un radical Br^\cdot à la formation d'un radical très réactif en position C_5 de l'uracile. Ce radical est susceptible de former une liaison covalente avec un groupement réactif de la protéine se trouvant à proximité. Il faut cependant remarquer qu'avec cette méthode, une absence de pontage ne signifie pas nécessairement l'absence de contact de la protéine avec l'ADN.

Les complexes ADN-protéine MCl sont reconstitués par mélange à faible force ionique (phosphate de potassium 10 mM, pH 7,2), dans un rapport de 1 molécule de protéine pour 15 paires de bases. Ces complexes sont irradiés 1 h puis amenés en milieu de force ionique élevée (phosphate de potassium 0,2 M, KCl 0,5 M, pH 7,2) pour dissocier les protéines non pontées. Celles-ci sont séparées du complexe par chromatographie de tamisage moléculaire sur colonne d'ACA 34 équilibrée et éluée par le tampon phosphate de potassium 0,2 M, KCl 0,5 M, pH 7,2. Une lecture du chromatogramme par mesure de fluorescence dans la bande du tryptophane permet de doser la quantité de protéine pontée.

IV - 3-2 RESULTATS DES EXPERIENCES DE PONTAGE

Un exemple type de séparation du complexe et de la protéine non pontée est présenté dans la figure 18. La mesure de la fluorescence permet d'estimer le rendement de fixation covalente de la protéine sur l'ADN. Pour effectuer ce calcul il est indispensable de tenir compte de l'effet d'écran de

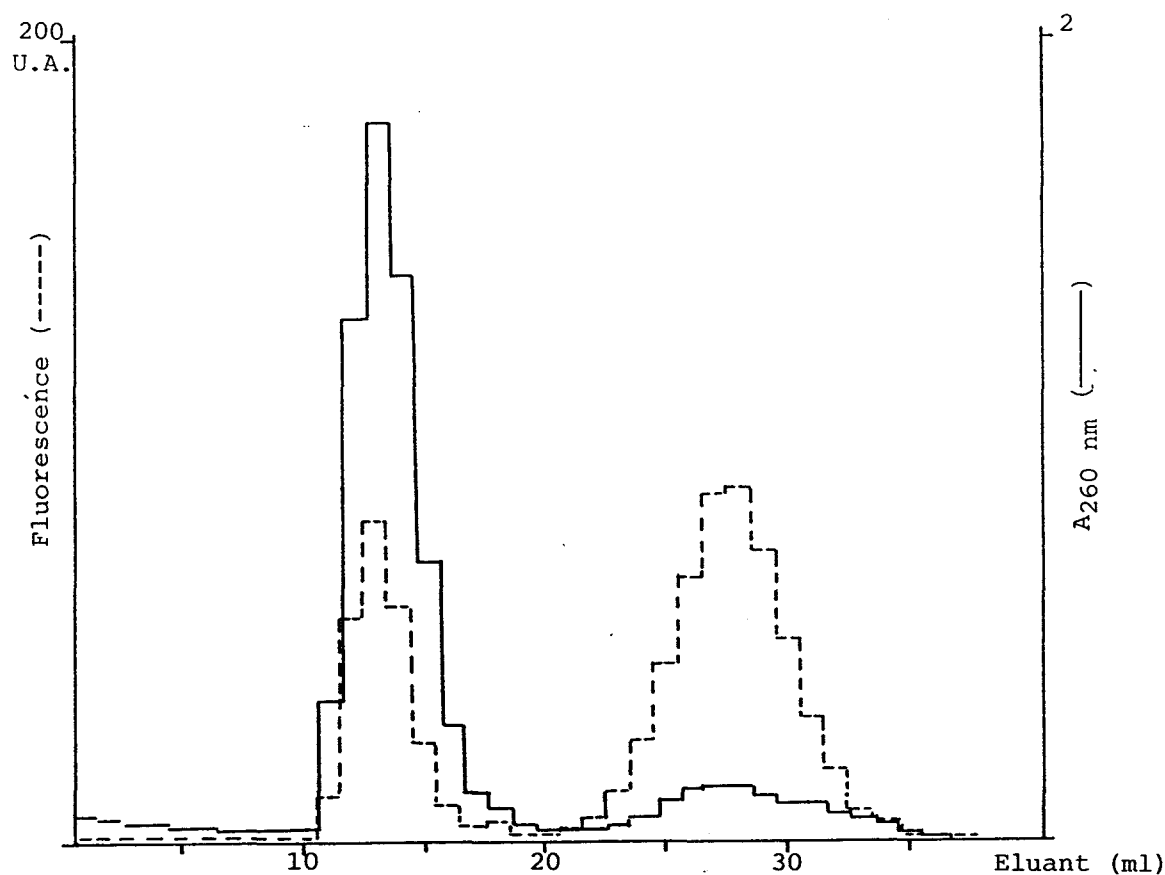


FIGURE 18 - Séparation des complexes protéine-ADN de la protéine non pontée sur colonne d'ACA 34.

La colonne d'ACA 34 (2 x 20 cm) est équilibrée et éluée avec un tampon phosphate de potassium 0,2 M, KCl 0,5 M, pH 7,2, avec un débit de 30 ml/h. Des fractions de 1 ml sont collectées. La chromatographie est suivie par lecture de la fluorescence dans la bande du tryptophane (excitation à 290 nm, émission à 345 nm) et par mesure de l'absorbance à 260 nm. La fluorescence est exprimée en unité arbitraire (U.A.)

l'ADN qui absorbe une partie de la lumière d'excitation des tryptophanes à 290 nm. Nous appliquons donc la formule établie par HELENE et coll. (1971).

$$IF = IF_0 \frac{d}{d_0} \frac{1 - 10^{-d_0}}{1 - 10^{-d}}$$

dans laquelle :

d = absorption totale de la solution à la longueur d'onde d'excitation

d_0 = absorption de la protéine à la longueur d'onde d'excitation

IF_0 = Intensité de fluorescence mesurée

IF = Intensité de fluorescence corrigée

Les rendements que nous avons obtenus sont proches de 20 %.

Lorsque nous irradiions des complexes reconstitués avec de l'ADN non substitué en milieu de force ionique faible, nous n'observons pas de fluorescence au niveau du pic d'ADN après passage sur colonne d'ACA 34. Ceci montre que les pontages font intervenir la bromo-uracile.

Lorsqu'un complexe protéine MCl-ADN substitué est reconstitué puis déposé, sans avoir été irradié, sur une colonne d'ACA 34 équilibrée en milieu de faible force ionique, nous retrouvons l'intégralité de la fluorescence au niveau du pic d'ADN. Ce qui montre que même aux rapports protéine/ADN élevés que nous avons utilisés, la totalité de la protéine est complexée à l'ADN.

Enfin, lorsque la protéine est ajoutée à l'ADN en milieu de force ionique élevée (milieu dissociant correspondant aux conditions de séparation sur la colonne d'ACA 34) et que ce mélange est irradié, la totalité de la fluorescence est retrouvée au niveau de la protéine libre. Ceci signifie que les pontages que nous observons ne sont pas liés à un simple phénomène de rencontre aléatoire entre la protéine et l'ADN au cours de l'irradiation.

IV - 3-3 ETUDE DU COMPLEXE ADN-PROTEINE MC1 PONTEE

Les résultats présentés dans cette partie ne sont en fait que les résultats d'une étude préliminaire destinée à vérifier la validité de notre protocole d'étude. Nous ne présentons ici que l'analyse des peptides libres (voir figure 17). L'analyse du peptide ponté à l'ADN n'a pas été possible en raison des faibles rendements que nous avons obtenus.

Les complexes pontés ADN-protéine MC1 ont été concentrés par ultracentrifugation (5 h à 223.000 xg) et dialysés contre un tampon bicarbonate d'ammonium 0,1 M - NaCl 0,7 M, pH 7,8. La protéine a été soumise à l'action de la protéase Glu-C de *Staphylococcus aureus* qui agit spécifiquement du côté carboxylique des acides glutamiques. Cette enzyme a été choisie en raison de la répartition de ces acides aminés dans la protéine MC1 (voir figure 20-A). De façon à éviter la rupture du pontage photochimique en milieu acide (PANDAY et MODAK, 1988), nous avons préféré utiliser l'hydrolyse enzymatique à pH 7,8 plutôt qu'à pH 4 comme cela a été fait pour l'étude des variants de protéine MC1 de *Methanotherix soehngeni*. Afin d'éviter que les sites potentiels de coupure ne soient masqués par l'ADN, nous avons utilisé une force ionique suffisante (NaCl 0,7 M) pour dissocier la protéine de l'ADN. Un échantillon de protéine MC1 non irradiée a été hydrolysé dans les mêmes conditions de façon à servir de témoin. L'hydrolyse enzymatique a été effectuée à 37°C pendant 18 h avec un rapport pondéral enzyme/substrat de 1/50. Les peptides ont été fractionnés par CLHP en phase inverse sur colonne de C₁₈.

La carte peptidique obtenue après coupure de la protéine témoin est présentée dans la figure 19. Les cinq peptides obtenus ont été identifiés par leur composition en acides aminés et éventuellement par leur séquence amino-terminale. Ils

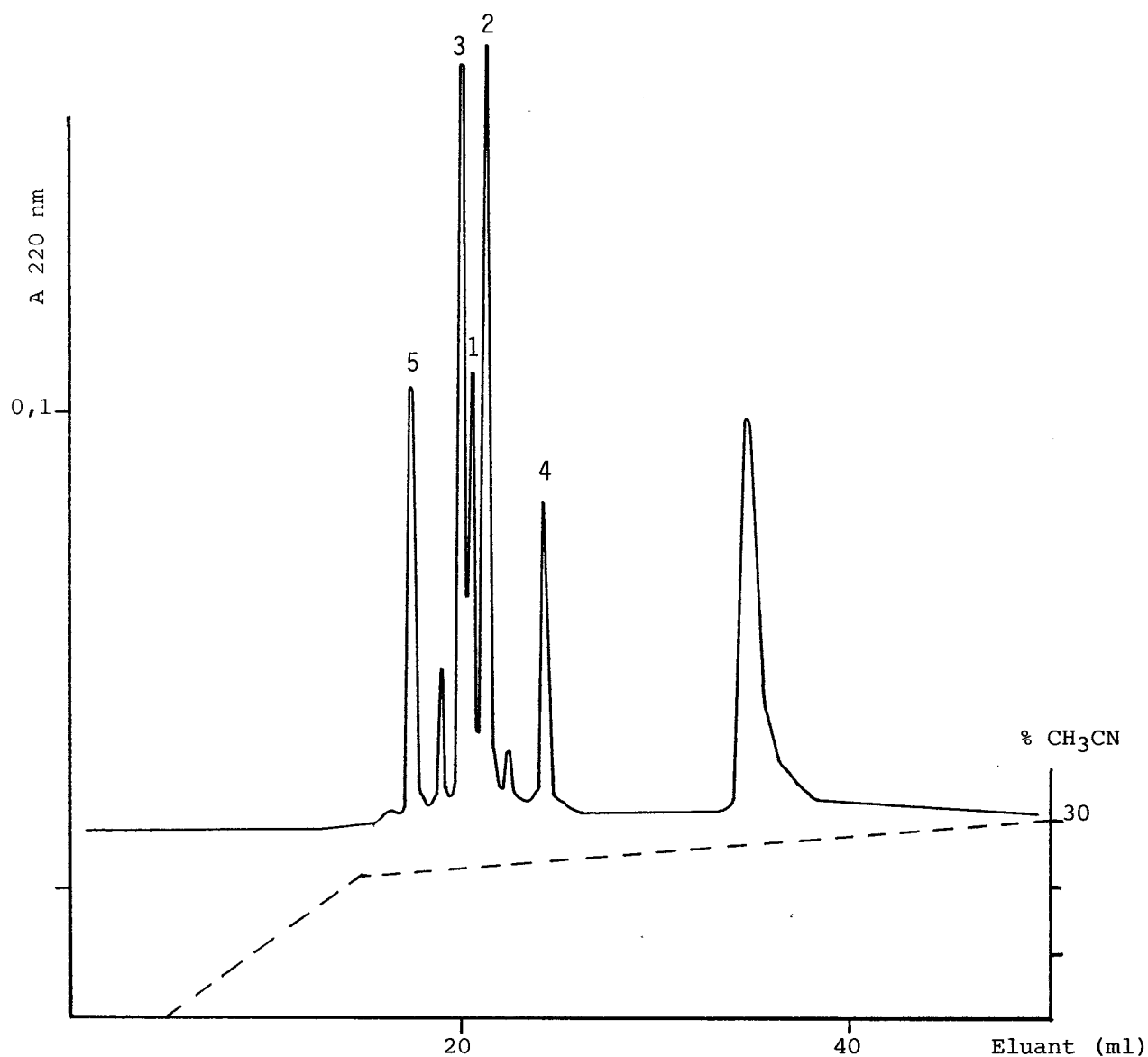


FIGURE 19 - Carte peptidique de la protéine MCl après coupure par l'endoprotéinase V8 en tampon bicarbonate d'ammonium 0,1 M, NaCl 0,7 M, pH 7,8.

La chromatographie est effectuée sur une colonne C₁₈ μ Bondapak (300 x 3,9 mm) ; (-----) gradient d'acétonitrile dans le TFA à 0,05% ; (——) absorbance à 220 nm. Les fractions sont numérotées selon la position du peptide dans la séquence de la protéine.

recouvrent la totalité de la protéine MCl (Figure 20). Les résultats obtenus sur cette protéine nous ont permis de contrôler l'activité de l'enzyme dans les conditions utilisées. La carte peptidique obtenue à partir du complexe ADN-protéine est présentée dans la figure 21. Les trois peptides obtenus couvrent les régions 1-49, 1-63 et 88-93 de la protéine (Figure 20). Nous n'avons pas retrouvé de peptide riche en proline caractéristique de la région 64-87 ; il semble donc que ce peptide corresponde au domaine de la protéine ponté à l'ADN. Nous avons montré au cours du chapitre III que ce peptide correspond à une région très variable de la protéine MCl. Toutefois, nous avons remarqué que les quatre résidus de proline de cette séquence (positions 68, 72, 76, 82) sont strictement conservés. Il en est de même des résidus hydrophobes en positions 65, 74, 75, 79 et 84 et des trois résidus de lysine en positions 81, 85 et 86. Cette séquence est très vraisemblablement organisée en coudes β juxtaposés.

Lors de cette étude préliminaire, nous avons observé que l'emploi de l'ultracentrifugation pour la concentration du complexe ADN-protéine et la séparation du complexe ADN-peptide des peptides libres donne des rendements très faibles. Nous avons donc repris cette étude en utilisant un nouveau protocole présenté dans la figure 22. Dans ce protocole, le complexe ADN-protéine obtenu par pontage photochimique est concentré par précipitation à l'éthanol. L'ADN est ensuite digéré par la nucléase staphylococcale et la DNase I avant hydrolyse enzymatique de la protéine par l'endoprotéinase Glu-C de façon à éviter de faire agir cette protéine en milieu de force ionique élevée.

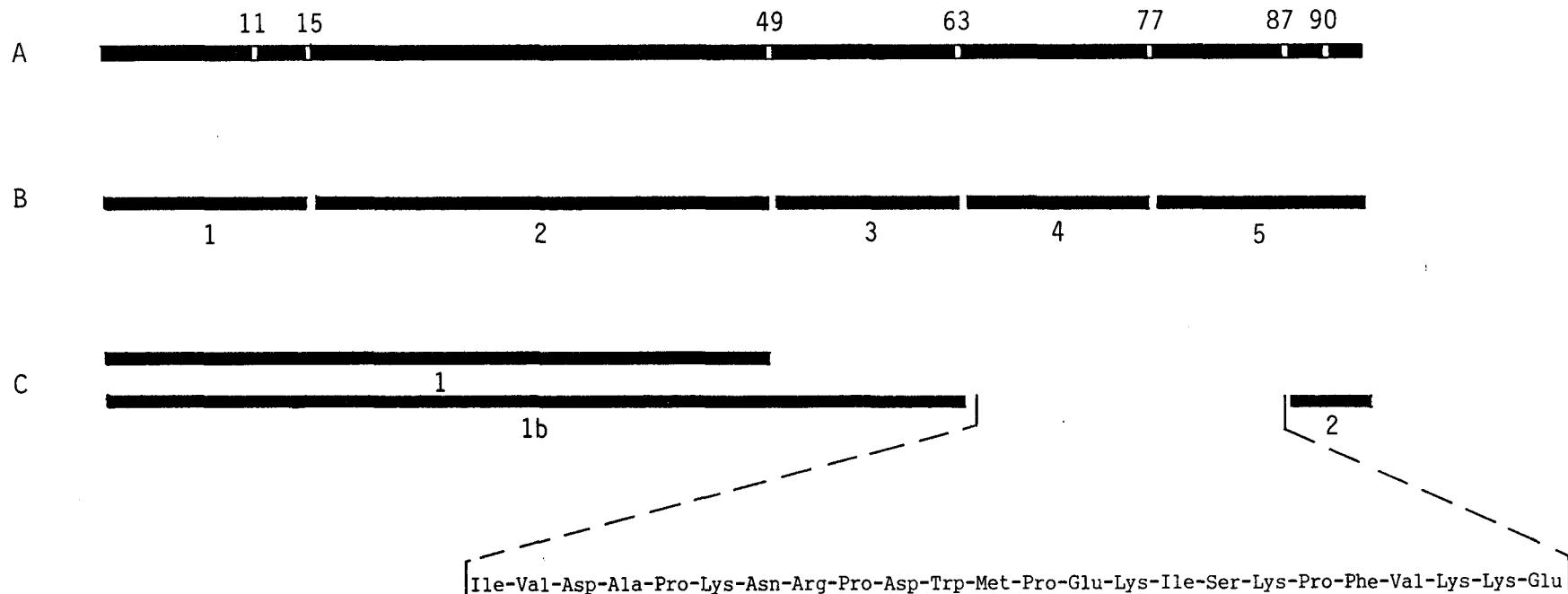


FIGURE 20 - Bilan des peptides obtenus après coupure par l'endoprotéinase V8 de la protéine MC1 libre et pontée à l'ADN.

- A : Sites potentiels de coupures de la protéine MC1.
- B : Peptides obtenus par coupure de la protéine MC1 témoin.
- C : Peptides récupérés après coupure de la protéine pontée à l'ADN. La séquence indiquée en toutes lettres est celle du peptide manquant qui correspond vraisemblablement à la région pontée à l'ADN.

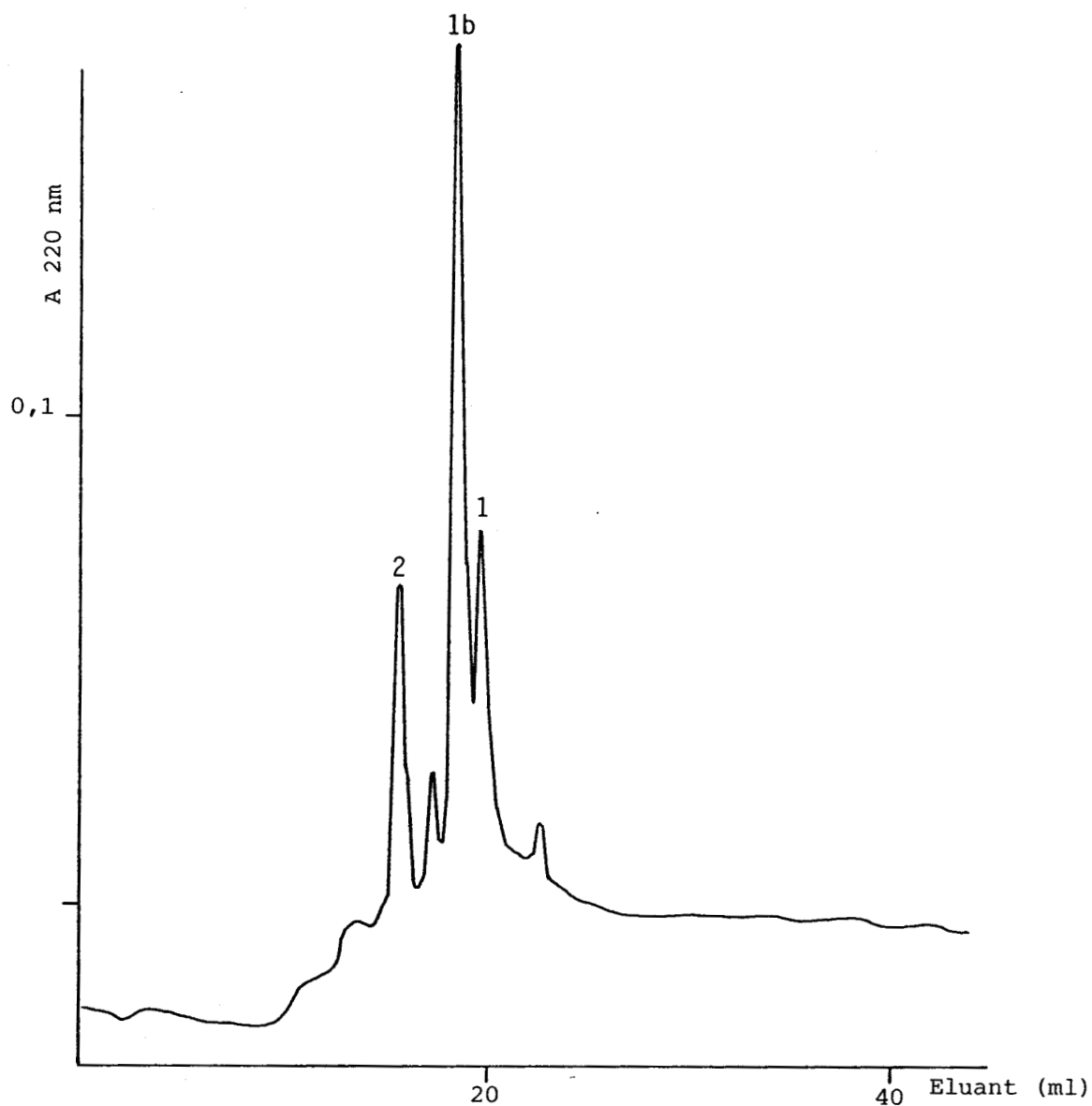


FIGURE 21 - Carte peptidique de la protéine MCl pontée à l'ADN après coupure par l'endoprotéinase V8.

Les conditions de chromatographie sont décrites dans la légende de la figure 20.

Les fractions ont été numérotées en fonction de la position du peptide dans la séquence de la protéine.

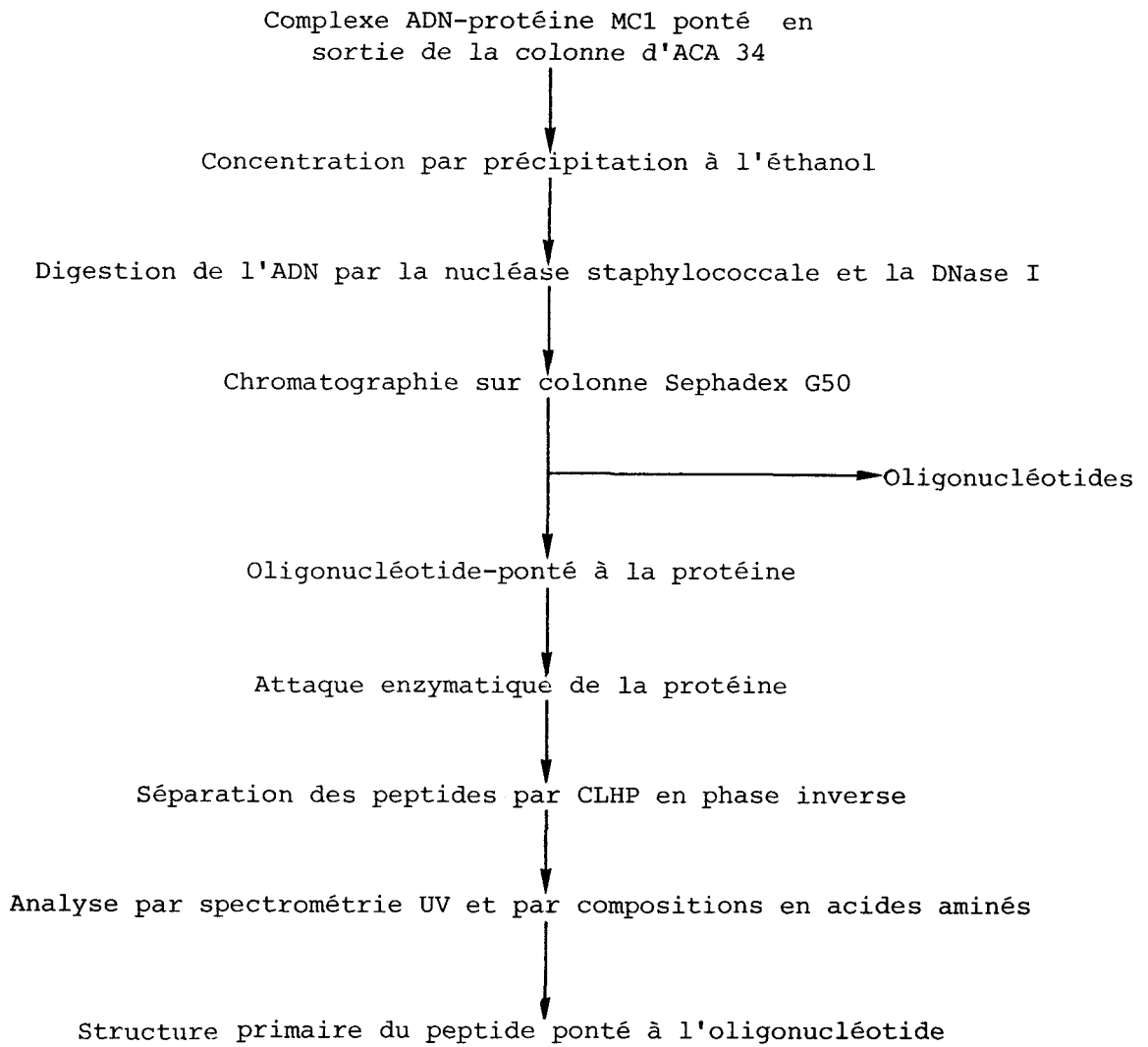


FIGURE 22 - Principe du nouveau protocole d'étude du site de pontage de la protéine MCl sur l'ADN.

CONCLUSIONS ET PERSPECTIVES

Nous avons isolé et caractérisé pour la première fois la protéine chromosomale d'archaebactéries de la famille des Methanosarcinaceae. Nous avons appelé cette protéine MC1 pour "Methanogen Chromosomal Protein 1".

Nous avons contribué à la détermination de la séquence en acides aminés de la protéine MC1 de *Methanosarcina barkeri*. Cette protéine basique de 93 résidus d'acides aminés ne présente aucune homologie de structure primaire avec les protéines chromosomales d'eubactéries ("DNA-binding protein II" et protéine HNS) et des eucaryotes (histones, protéines non-histones) et d'autres archaebactéries comme *Thermoplasma acidophilum* et *Sulfolobus acidocaldarius*.

Dans le cadre d'études comparatives, nous avons déterminé la séquence en acides aminés de la protéine MC1 de *Methanosarcina* sp CHTI 55 et des trois variants de la protéine MC1 de *Methanotherix soehngeni*.

Nous avons établi que :

- Deux séquences, les séquences 17 à 35 et 45 à 58 sont très conservées et correspondent aux régions structurées de la protéine. Ceci suggère que ces régions seraient impliquées dans l'organisation spatiale de la protéine et/ou dans des interactions protéine-protéine.
- Dans le tiers C-terminal variable de la protéine, il y a stricte conservation des résidus de proline donnant une structure en coudes β successifs.
- D'après les études réalisées sur les complexes ADN-protéine MC1 pontée, la région 64 à 87 de la protéine contient le site d'interaction de la protéine avec l'ADN.

D'autre part, nous avons montré que la protéine MC1 est capable de stabiliser l'ADN contre la dénaturation thermique et de stimuler de façon significative la transcription de l'ADN par l'ARN polymérase d'*Escherichia coli*, lorsqu'elle est complexée à l'ADN à des rapports voisins du rapport physiologique (protéine/ADN = 0,1 en poids). Ces deux effets peuvent être les conséquences de l'organisation de l'ADN par la protéine MC1.

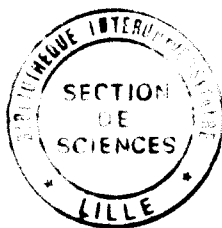
La protéine MC1 constitue un nouveau modèle de protéine chromosomale. Il serait très intéressant d'étudier de manière plus approfondie son mode d'interaction avec l'ADN ainsi que la conformation que la protéine impose à l'ADN.

L'étude des conséquences de l'interaction de la protéine MC1 avec l'ADN sur les activités enzymatiques liées à la régulation de l'expression de l'ADN et la recherche de facteurs pouvant influencer la fixation de la protéine MC1 sur l'ADN seraient également dignes d'intérêt.

BIBLIOGRAPHIE

- ALBERTS, B. and HERRICK, G. (1971)
DNA-cellulose chromatography.
in Methods Enzymol. (COLOWICK, S.P. and KAPLAN, N.O., eds)
21, 198-217, Academic Press, New-York.
- CHOLI, T., WITTMANN-LIEBOLD, B. and REINHARDT, R. (1988)
Microsequence analysis of DNA-binding proteins 7a, 7b and
7e from the archaebacterium *Sulfolobus acidocaldarius*.
J. Biol. Chem. 263, 7087-7093.
- CHOU, P.Y. and FASMAN, G.D. (1978)
Empirical predictions of protein conformation.
Ann. Rev. Biochem. 47, 251-276.
- GABORIAUD, C., BISSERY, V., BENCHETRIT, T. and MORNON, J.P. (1987)
Hydrophobic cluster analysis, an efficient new way to com-
pare and analyse amino acid sequence.
FEBS Lett. 224, 149-155.
- GARNIER, J., OSGUTHORPE, D.J. and ROBSON, B. (1978)
Analysis of the accuracy and implication of simple methods
for predicting the secondary structure of globular proteins.
J. Mol. Biol. 120, 97-120.
- HACKETT, P. Jr and HANAWALT, P. (1966)
Selectivity for thymine over 5 bromo-uracyl by a thymine
requiring bacterium.
Biochim. Biophys. Acta 123, 356-363.
- HELENE, C., BRUN, F. and YANIV, M. (1971)
Fluorescence studies of interactions between *Escherichia coli*
valyl tRNA synthetase and its substrate.
J. Mol. Biol. 58, 349-365.
- IMBERT, M. (1987)
La protéine chromosomale MC1 isolée d'archaebactéries du
genre *Methanosarcina*. Etudes structurales, localisation ul-
trastructurale et étude du complexe désoxyribonucléopro-
téique.
Thèse de Doctorat d'Université, Lille
- LEVIN, J.M., ROBSON, B. and GARNIER, J. (1986)
An algorithm for secondary structure determination in pro-
teins based on sequence similarity.
FEBS Lett. 205, 303-308.

- LI, H.J. (1978)
Thermal denaturation analysis of chromatin and DNA-nuclear protein complexes.
Methods in Cell Biol. 18, 385-396.
- MARMUR, J. (1961)
A procedure for the isolation of deoxyribonucleic acid from micro-organisms.
J. Mol. Biol. 3, 208-218.
- PANDEY, V.N. and MODAK, M.J. (1988)
Purification of nucleotide-linked peptide.
J. Chromatogr. 444, 133-139.
- WILLMITZER, L., BODE, J. and WAGNER, K.G. (1977)
Phosphorylated protamines. I. Binding stoichiometry and thermal stability of complexes with DNA.
Nucleic Acids Res. 4, 149-175.
- WEICHET, W.O., TATCHELL, K., VAN HOLDE, K.E. and KLUMA, H. (1978)
Thermal denaturation of nucleosomal core particles.
Nucleic Acids Res. 5, 139-160.
- YU, S.S., LI, H.J., GOODWIN, G.H. and JOHNS, E.W. (1977)
Interaction of non-histone chromosomal proteins HMG 1 and HMG 2 with DNA.
Eur. J. Biochem. 78, 497-502.



Résumé

Les archaebactéries constituent une troisième branche divergente de l'ancêtre commun aux eubactéries et aux eucaryotes. L'étude des protéines chromosomales de ces bactéries constitue un apport original permettant d'éclairer les relations entre les procaryotes et les eucaryotes au niveau de la composition et de l'organisation de la chromatine. Les travaux ont porté sur la protéine chromosomale MC1 des archaebactéries méthanogènes de la famille des Methanosarcinaceae.

La structure primaire de la protéine MC1 isolée de *Methanosarcina barkeri*, de *Methanosarcina* sp CHTI 55 et des trois variants de la protéine MC1 de *Methanotherix soehngenii* a été déterminée. La protéine MC1 est un polypeptide basique de 93 résidus d'acides aminés, caractérisé par un nombre élevé de résidus chargés distribués tout au long de la chaîne polypeptidique. Les études de variabilité structurale de cette protéine montrent que les séquences 17 à 35 et 45 à 58 sont très conservées et que le tiers carboxy-terminal est le plus variable. Toutefois dans cette région, quatre résidus de proline sont strictement conservés et donnent à cette séquence une structure particulière formée de trois coudes β juxtaposés. Cette région semble correspondre au domaine d'interaction avec l'ADN.

La protéine MC1 ne présente aucune homologie de séquence avec les protéines chromosomales des eucaryotes ou des eubactéries.

La protéine MC1 est capable de stabiliser l'ADN contre la dénaturation thermique, ce qui suggère qu'elle a un rôle structural sur l'ADN. Quand la protéine MC1 est complexée à l'ADN dans un rapport protéine/ADN de 0.1 en poids (rapport physiologique), on observe une forte stimulation de la transcription de l'ADN *in vitro*.

Mots clefs : Archaebactéries - Bactéries méthanogènes - Methanosarcinaceae - Protéine chromosomale - Structure primaire - Variabilité structurale - Interactions ADN-protéine - Dénaturation thermique de l'ADN - Transcription de l'ADN.