Atypical Recognition of Particular DNA Sequences by the Archaeal Chromosomal MC1 Protein†

Guillaume De Vuyst, Samia Aci, Daniel Genest, and Françoise Culard*

Centre de Biophysique Moléculaire, CNRS, rue Charles-Sadron, 45071 Orléans Cedex 2, France

Received December 7, 2004; Revised Manuscript Received April 12, 2005

ABSTRACT: The MC1 protein is a chromosomal protein likely involved in the DNA compaction of some methanogenic archaea. This small and monomeric protein, structurally unrelated to other DNA binding proteins, bends DNA sharply. By studying the protein binding to various kinds of kinked DNA, we have previously shown that MC1 is able to discriminate between different deformations of the DNA helix. Here we investigate its capacity to recognize particular DNA sequences by using a SELEX procedure. We find that MC1 is able to preferentially bind to a 15 base pair motif [AAAAACACAC(A/C)CCCC]. The structural parameters of this sequence are characterized by molecular dynamics simulation experiments, and the binding mode of the protein to the DNA is studied by footprinting experiments. Our results strongly suggest that the protein realizes an indirect readout of the DNA sequence by binding to the DNA minor groove.

There is no single mechanism of archaeal genome packaging, and a variety of structural proteins, isolated from these prokaryotic organisms, have been characterized (*1*). These abundant proteins use different strategies to bind and to compact DNA, and they likely participate in different cellular processes involving DNA sequence accessibility. The proteins of the HMf family, containing an α-helical histone fold, have been found in different species of the euryarchaeota subdomain (*2*−*4*). These archaeal histones have no N- and C-terminal extensions, and unlike their eukaryal counterparts, they are probably not posttranslationally modified but they can form nucleosome-like structures similar to the H3/H4 tetramer core (*5*, *6*). The *Sulfolobus* strains, belonging to the Crenarchaeota subdomain, lack histones but contain small proteins called Sso7d, Sac7d, or Ssh7d according to their origin that are involved in DNA bending and compaction (*7*−*10*). Some archaea may use several ways to compact their DNA. Indeed, the sequenced genomes of the *Methanosarcina* species (euryarchaeota subdomain) contain one gene coding for a true archaeal histone as well as genes coding for structural proteins of the MC1 family. Under our laboratory conditions no archaeal histone seems to be expressed, and it is plausible that these different chromosomal proteins exhibit growth phase-dependent expression. In this way, a dramatic decrease in the level of the archaeal histones HTz1 and HTz2 with the culture age of the thermophilic archaeon *Thermococcus zilligii* has been reported (*11*).

Proteins of the MC1 family are present in different methanosarcinales (*12*, *13*) and are predicted to be present in *Halococcus*, *Halobacterium*, and *Haloferax*. Recent data from several sequenced genomes (*14*, *15*) as well as earlier protein sequences data show that two or three conserved variants (*16*) can be present in some species. These small proteins (87−94 residues) are characterized by a large number of charged residues (respectively 20−24 basic and 10−14 acidic amino acids) distributed all along the protein. The three-dimensional structure of the *Methanosarcina thermophila* MC1 protein, determined by NMR spectroscopy (*17*), reveals a novel fold comprising one α helix and five β strands arranged as two antiparallel β sheets packed in an orthogonal manner, forming a barrel (GenBank Accession Number 1T23).

MC1 proteins bind double-stranded DNA as a monomer, in a noncooperative way and with an excluded site size of 10−11 bp (*18*). Electron microscopy experiments have shown that MC1 binding induces sharp kink(s) on DNA (*19*, *20*) so that the overall DNA bend is estimated at 116° (*21*). Related to its capacity to induce strong DNA conformational changes, the MC1 protein is able to discriminate between different deformations of the DNA helix. Thus the protein recognizes and strongly binds to four-way junctions and to minicircle DNA whereas it binds very badly to bulged DNA (*20*, *22*). The binding to DNA minicircles of 207 bp in length is particularly strong and highly depends on the DNA topology (*23*).

The chromosomal proteins are considered as unspecific DNA binding proteins that strongly bind to DNA whatever its sequence. However, several studies have shown that some of them have a higher affinity for particular DNA sequences. Thus, the archaeal and eucaryal histones recognize different kinds of intrinsically curved and easily flexible sequences (*24*−*26*). Such preferential bindings can modify the availability of the genetic material and can participate in the regulation of gene expression. In a previous work, a preferential binding site for MC1 was found by chance in the *Methanosarcina barkeri* genome. The binding was sufficiently strong to produce a clear DNase I footprint and to perform circular permutation assays (*23*). In this work

† This work was supported with a Ph.D. grant from La Région Centre (to S.A.).

* To whom correspondence should be addressed. Tel: 33 2 38 25 56 29. Fax: 33 2 38 63 15 17. E-mail: culard@cnrs-orleans.fr.

10.1021/bi0474416 CCC: $30.25
Published on Web 07/09/2005

we systematically examine the MC1 DNA binding recognition by using a SELEX[1] approach (*27*). Besides the search for preferential binding, the aim of this study is to better understand the DNA binding mode of the protein. To complete our investigation, a 5 ns molecular dynamics (MD) simulation of a naked consensus sequence was performed in order to identify its intrinsic structural properties.

EXPERIMENTAL PROCEDURES

Materials. The MC1 protein (MW 11 kDa) from the strain *Methanosarcina* sp. CHTI 55 was purified as previously described (*12*, *21*). The protein concentration was determined by absorption spectrophotometry using a molecular absorbance of 11000 M^{-1} cm^{-1} at 280 nm.

All oligonucleotides were purified by denaturing polyacrylamide gel electrophoresis. DNA labeling was performed using T4 polynucleotide kinase and [γ-^{32}P]ATP (Amersham, 3000 Ci·$mmol^{-1}$).

SELEX Experiments. For the first SELEX experiment 65-mer oligonucleotides including 15 random nucleotides [5′ CAGGTCAGTTCAGCGGATCCTGTCG-$(N)_{15}$-GAGGCGAATTCAGTGCAACTGCAGC 3′] and two PCR primer sequences corresponding to the first 25 bases (forward: 5′ CAGGTCAGTTCAGCGGATCCTGTCG 3′) and complementary to the last 25 bases (reverse: 5′ GCTGCAGTTGCACTGAATTCGCCTC 3′) were chemically synthesized. The random sequence library was generated by a primer extension reaction carried out with the purified 65-mers as template and the reverse primer in a 50 μL Klenow reaction mixture. It was gel purified and ^{32}P-labeled. To perform the selections, approximately 1 μg of the library DNA (final concentration 45 nM) in 500 μL of binding buffer was incubated for 30 min at room temperature with the protein MC1 (360 nM for the initial five selection cycles and 36 nM in the later cycles) (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 150 mM NaCl). MC1 protein−DNA complexes were isolated by electrophoresis on two 8% native polyacrylamide gels (19:1 acrylamide:bisacrylamide) in Tris−borate−EDTA buffer. The gel regions containing the complexes were localized by autoradiography, excised, and soaked overnight at 37 °C in elution buffer (20 mM Tris-HCl, pH 7.4, 1 mM EDTA, 100 mM NaCl). The eluted DNA was purified on an Elutip column (Schleicher and Schuell) and ethanol precipitated. A fraction of the eluate (about 10%) was used for PCR amplification with 0.2 μM each primer and 0.2 mM each deoxynucleoside triphosphate. Amplifications were done first by 5 min at 94 °C followed for a total of 20 cycles by 1 min at 94 °C, 1 min at 65 °C, and 1 min at 72 °C. The PCR products were purified on 12% polyacrylamide gels (acrylamide:bisacrylamide, 19:1). DNA amplified from the tenth cycle of selection was cut by *Eco*RI and *Bam*HI at the constant regions and ligated into pGEM-3Z f(−) plasmids (Promega). The recombinant vectors were transformed into *Escherichia coli* strains DH5α and plated on selective medium. White colonies were picked up, and inserts were determined by automated sequencing (MWG-Biotech).

The second SELEX experiment was performed by using basically the same protocol with a 71 base oligonucleotide (5′ CTGCAGTTGCACTGAATTCGCCTCNNNAAAAACACACMCCCMNNNCGACAGGATCCGCTGAACTGACCTG 3′, where M is C or A) as initial DNA. The 25 bp at each end are identical to those of the fragments used for the first SELEX experiment, and the same primers were used for the PCR amplifications. Six rounds of selection were performed with a protein amount (60 ng, final concentration 36 nM) that shifted about 5% of the DNA fragments.

Searches in the two completely *Methanosarcina* sequenced genomes (*14*, *15*) were conducted using the blastn search program with the threshold set at 1000 at http://www.ncbi.nlm.nih.gov.

Electrophoretic Mobility Shift Assays (EMSA). To determine the apparent dissociation constant (K_D) of the MC1 protein for the consensus sequence and for a random sequence, we used DNA fragments of respectively 26 and 27 bp. These synthetic DNA fragments were composed of a ^{32}P-labeled oligonucleotide [26 bp, 5′ GCCTCAAAAACACACACCCCCGACAG 3′; 27 bp, 5′ CGCCTC-$(N)_{15}$-CGACAG 3′] annealed to the complementary strand by incubation at 80 °C for 1 min and slowly cooled to room temperature. DNA duplexes (8 pM) were incubated with 0.5−1000 nM protein in the presence of 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 100 mM NaCl, and 15 μg/mL BSA to a total volume of 15 μL. Samples were incubated at 4 °C for 30 min, mixed with $^1/_5$ volume of loading buffer (50% glycerol in binding buffer), and loaded on an 8% native polyacrylamide gel (30:1 acrylamide:bisacrylamide) in Tris−borate−EDTA buffer. Electrophoreses were run at 14 V/cm for 4 h at 4 °C. The fraction of bound duplexes was quantified using a Storm 860 (Molecular Dynamics). The binding curves were fitted to a single binding site model using the equation $Y = ([MC1]/K_D)/(1 + [MC1]/K_D)$, where Y is the fraction of bound DNA and [MC1] is the concentration of free protein.

The ratio of preferential (p) and random (r) association binding constants (K_p/K_r) is related to the free energy differences $\Delta\Delta G$ according to $\Delta\Delta G = -RT \ln(K_p/K_c)$.

The complex stability was determined as previously described (*20*, *23*). Briefly, protein−DNA complexes were first formed under the indicated conditions; then unlabeled pUC8 plasmid (117 μM/base pair) was added as competitor DNA. Aliquots of the mixture were loaded onto running gels (previously described) at different times. To calculate the half-lives of the complexes, the percentage of bound DNA was plotted versus incubation time (logarithmic scale) and fitted to a straight line (*28*). The experiments described in Table 1 were obtained with 65 bp DNA fragments prepared by PCR with the various derived pGEM-3Z f(−) plasmids, each containing a selected fragment as template and two primers: forward, 5′ GTCGACTCTAGAGGATCCTG 3; reverse, 5′ CGACTCACTATAGGGCGAAT 3′.

To study the effect of distamycin A on the MC1−DNA complexes, a ^{32}P-labeled 26 bp DNA fragment (1 nM) containing the consensus sequence (same sequence as above) was first incubated with MC1 (6 nM) for 15 min in the presence of 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, and 60 mM NaCl, to a total volume of 10 μL; then 1 μL of various solutions of distamycin (Sigma) at concentrations from 0.5 to 15 μM was added. The different mixtures were incubated for 40 min and loaded on a gel as previously described.

Footprinting and Interference Experiments. These experiments were performed with a 59 bp duplex DNA (5′ GC-

[1] Abbreviations: MC, methanogen chromosomal; SELEX, *s*ystematic *e*volution of *l*igands by *ex*ponential enrichment; MD, molecular dynamics.

AGTTGCACTGAATTCGCCTCAAAAACACACACCCCC-GACAGGATCCGCTGAACTGAC 3′) first 5′-^{32}P-end-labeled on one strand prior to duplex formation.

Hydroxyl radical footprinting assays were performed as described previously (*22*) using 25 nM DNA and 50 nM MC1 protein in 10 mM Tris-HCl, pH 7.4, and 60 mM NaCl, incubated at 20 °C for 20 min. A fresh mixture of 1.5 μL of sodium ascorbate (20 mM), $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$ (0.4 mM), EDTA (0.8 mM), and H_2O_2 (0.3%) was added to a 15 μL aliquot to initiate the reaction. After 2 min at 20 °C, 2.25 μL of thiourea (0.1 M) and 0.45 μL of 80 mM EDTA were added to stop the reaction. DNA was extracted with phenol/chloroform and then precipitated with ethanol. The DNA samples and G + A (Maxam−Gilbert) reactions were electrophoresed on 8% denaturing polyacrylamide/7 M urea gels in TBE buffer. Gels were dried and quantified on a Molecular Dynamics phosphorimager (Storm 860 system).

For DMS protection experiments 25 nM DNA fragments labeled on the bottom strand were incubated with 0−25 nM MC1 protein in 10 mM Tris-HCl, pH 7.4, 60 mM NaCl, and 50 mM sodium cacodylate, at 20 °C for 20 min. The protein−DNA complexes were treated with DMS (0.5% final) for 30 s at 20 °C, and the reaction was terminated by the addition of 1 M β-mercaptoethanol (1/3 v/v). The samples were phenol treated, ethanol precipitated, and dissolved in 100 μL of 10% piperidine. The samples were then incubated at 95 °C for 30 min, vacuum-dried, suspended in 8 μL of 50% formamide, and analyzed on urea−acrylamide gel as described above.

For the interference assay, the labeled DNA was methylated (0.25% final DMS) for 30 s at 20 °C, and the reaction was stopped by the addition of β-mercaptoethanol as described above and ethanol precipitated. The modified DNA was then incubated with 250 nM MC1 protein for 20 min on ice. The bound and free DNAs were separated on an 8% native polyacrylamide gel, eluted, and analyzed after piperidine cleavage as described above.

MD Simulation. A 16 bp sequence containing the consensus sequence d($^{-1}$CAAAAACACACACCCC^{15})·d(^{15}GGGGTGTGTGTTTTTG^{-1}) was simulated using the AMBER7 package (*29*) with the parm 98 force field (*30*). The canonical B form DNA was taken as the starting structure. Thirty Na^+ counterions were initially randomly distributed around the molecule with the following conditions: (i) each Na^+ is associated to a specific phosphorus atom with a distance smaller than 7 Å, (ii) the distance between any two ions is larger than 6 Å, and (iii) the distance between any ion and any DNA atom is larger than 5 Å. The system was hydrated with a truncated octahedron box of explicit water, leading to a total of 31167 atoms. The whole system was first energy minimized with 100 steps of steepest descent followed by 1900 steps of conjugate gradient methods. MD simulation started with a progressive increase in the temperature from 0 to 300 K during 25 ps, restraining counterions and DNA atom positions. During 32 ps these restraints were progressively removed at 300 K first for the ions and then for DNA. This equilibration period was ended with 500 ps of free MD simulations at 300 K. Finally, the production period consisted of 4.5 ns of free MD simulation in the NPT ensemble with storage of atomic coordinates every 0.5 ps. The SHAKE algorithm (*31*) was applied to maintain constant the covalent bonds including a hydrogen atom, and the particle mesh Ewald (PME) method was used to calculate electrostatic interactions (*32*). The integration time step for solving the equations of motion was 2×10^{-15} s.

The different structural parameters and the curvature of the double helix were calculated with the program CURVES (*33*). The similarity between two structures of the molecule was quantified by RMSD $= \{(1/N)\sum(r_{i1} - r_{i2})^2\}^{1/2}$, where r_{i1} and r_{i2} are the position vectors of atom i in structures 1 and 2, respectively, N is the number of atoms of the molecule, and the summation was performed over the N atoms. The solvent (water) accessible surface area (SASA) of each atom was calculated with POPS (*34*) on the server located at http://ibivu.cs.vu.nl/programs/popswww/.

RESULTS

In Vitro Selection of MC1 Preferential Binding Sites. The DNA pool of our SELEX experiment consisted of 65 bp DNA molecules in which the 25 bp at each extremity were constant and the central 15 bp were random. We chose the randomized sequence length on the basis of previous experiments with short DNA duplexes (*18*) and in order that a single molecule could bind to this part of the fragments. For the first five cycles of selection an amount of MC1 protein that bound roughly 10% of the DNA was used, and the amount was then progressively decreased until only 5% of the DNA was selected. At each cycle, the complexes were separated from the free DNA by native gel electrophoresis as described in Experimental Procedures. The progressive enrichment of MC1 binding sites was followed by measuring the stability of the complexes. In these experiments, the free proteins are trapped by a large excess of cold competitor DNA, and the proteins remaining on the labeled DNA are assessed by loading aliquots of the dissociation mixture on running gel at various times. Loss in the fraction of labeled bound DNA versus time can then be used to follow the dissociation kinetics of the complexes.

Figure 1 compares the formation and the stability of MC1−DNA complexes with SELEX DNA fragments before (pre-SELEX fragments) and after 10 cycles of selection. Both the amount of complexes and their kinetics of dissociation show that selection has occurred. Indeed, fewer complexes are visible at time 0 (i.e., before the addition of the competitor DNA) with the pre-SELEX DNA fragments than with the selected fragments (14% versus 31% of the total amount of DNA, respectively, in lanes A and G). Once formed, the complexes with the pre-SELEX fragments dissociate very rapidly (hardly any complex is visible after about 15 min) while with the selected DNA fragments about 50% of the complexes present at time 0 are still present after 30 min (Figure 1B). As we found no improvement by adding an 11th round, the selection was stopped after round 10.

Analysis of the Selected Binding Sites. The selected DNAs were then digested, cloned, and sequenced (Experimental Procedures). Figure 2 shows the nucleotide sequence of 15 independent clones. Although each clone was unique, four sequences appeared twice in the sequence population (sequences 1 and 4, 2 and 31, 11 and 16, and 12 and 28). Only 11 different sequences were obtained, and all of them contained in their 5′ portion an AAAACA motif. Clearly,

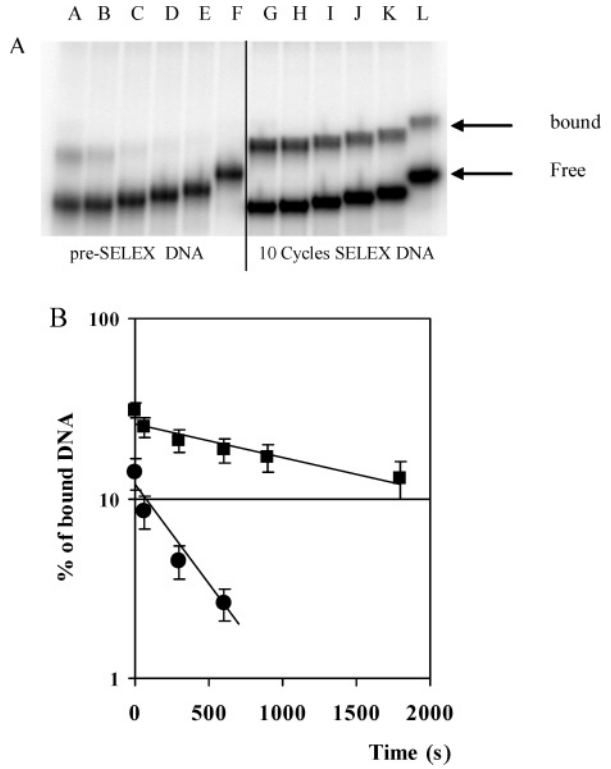

FIGURE 1: Dissociation kinetics of the MC1 protein from DNA fragments. (A) EMSA. The left part of the gel (lanes A−F) corresponds to the MC1 binding to a population of DNA of random sequence (pre-SELEX fragments), and the right part (lanes G−L) corresponds to the binding to a population of fragments after 10 cycles of selection. Proteins (2.5 nM) and DNA (5 nM) were allowed to assemble for 30 min at 4 °C before addition of plasmid competitor DNA (117 μM/base pair). Samples were loaded on the running gel at various times after the addition of the competitor DNA: 0 (lanes A and G), 59 s (lanes B and H), 298 s (lanes C and I), 598 s (lanes D and J), 900 s (lanes E and K), and 1800 s (lanes F and L). The binding buffer was 60 mM NaCl in TE (10 mM Tris-HCl, pH 7.5, 1 mM EDTA). (B) Kinetics of dissociation of preformed complexes with the pre-SELEX DNA fragments (circles) and with the population of DNA fragments after 10 rounds of selection (squares). The percentage of bound DNA was obtained by quantification of the gel using a β scanner (Storm; Molecular Dynamics) and plotted with a logarithmic scale versus the incubation time(s) after the addition of competitor DNA.

we have carried out the selection process until only one family of DNA binding sites has been obtained and the selection has occurred over the whole of the random region. A consensus binding site of 15 bases [AAAAACACAC(A/C)CCCC] can be deduced from this sequence alignment with a degree of conservation at each of the 15 positions ranging from 73% to 100%.

The sequence alignment also shows that all of the selected sequences have the same orientation in relation to the constant adjacent sequences, raising the possibility that these adjacent sequences may have played a role in the selection process so that the preferred binding sequences could be longer than 15 bp. To test this possibility, we decided to perform a second SELEX experiment with a pool of 71 bp DNA fragments in which the central 21 bp corresponded to the 15 bp consensus sequence flanked at each side by 3 random base pairs (the complete sequence of the fragments is in Experimental Procedures). Six rounds of selection were performed in conditions similar to those used previously, and the binding properties of this pool of DNA fragments were studied. Since the amount of complexes and their stability were the same as those with DNA fragments of round 0 of this second SELEX experiment (not shown), we concluded that there was no enrichment in sequences of higher affinity and no sequencing was performed. Clearly, the sequences adjacent to the consensus sequence do not significantly contribute to the MC1 binding.

Table 1: Summary of Half-Lives of Complexes Formed with Some Selected DNA Fragments[a]

sequence	position															half-life (min)
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
2	A	A	A	A	A	C	A	C	A	C	A	C	C	C	C	29 ± 2
11	...														A	24 ± 1
8	...										C				...	23 ±1
12	...													A	...	22 ± 1
13	C	...									C				...	12 ± 2
3	...													T	...	9 ±1
10	...							A	...				A	...	A	5 ±1

[a] The half-life of each complex was determined from three or more kinetic progress curves as in Figure 1. The binding buffer was 100 mM NaCl in TE.

Comparison of the Stability of the Various Complexes. The stability of the complexes formed with seven of the selected sequences was compared by measuring their dissociation kinetics as in Figure 1 but with a higher salt concentration (100 versus 60 mM NaCl). This experiment was performed with PCR DNA fragments of 65 bp. As expected, the complexes formed with these fragments are significantly more stable than complexes formed with random DNA sequences. Indeed, in the presence of 100 mM NaCl, complexes with the pool of pre-SELEX fragments dissociate immediately and cannot be studied by EMSAs. The strongest complex is formed with sequence 2 containing $A_5(CA)_3C_4$, which corresponds to the consensus sequence (Table 1). Three other fragments that contain one substitution in regard to sequence 2 also have quite a strong stability: fragment 8 that contains a C/G in place of an A/T in position 11 and that also corresponds to the consensus sequence; fragments 11 and 12 that contain an A/T versus C/G respectively in positions 14 and 15. On the other hand, a T/A in position 14 destabilizes the complexes (fragment 3). The presence of an A/T in position 1 seems important for the binding since substitution by C/G decreases the stability of the complexes (complexes with fragment 13 are nearly twice less stable than those formed with fragment 8). The weakest complex is formed with fragment 10, which has three A/T in place of G/C (these three substitutions destabilize the complexes by a factor of 5).

EMSAs were carried out to analyze the binding of MC1 to small DNA fragments with or without the consensus sequence under conditions where only one complex is formed. Figure 3 shows the fractional DNA saturation versus the protein concentration with a 26 bp DNA containing sequence 2 [$A_5(CA)_3C_4$] and with a 27 bp DNA containing 15 random base pairs. These data show the high affinity of MC1 for the selected DNA (K_D = 2 nM) and the lower affinity of the protein for the random sequences (K_D = 100 nM). A nearly 50-fold preference for the consensus sequence was also obtained by direct competition between different DNA fragments and by kinetics measurements (not shown).

Biochemical Analyses of the Complex MC1 Consensus Sequence. First, to identify sugars that are in close proximity to the MC1 protein in the complexes, we performed hydroxyl

		position															
name		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	t c	A	**A**	**A**	**A**	**A**	**C**	**A**	C	C	A	C	C	C	C	C	c g
2	t c	A	**A**	**A**	**A**	**A**	**C**	**A**	C	A	C	A	C	C	C	C	c g
3	t c	A	**A**	**A**	**A**	**A**	**C**	**A**	C	A	C	A	C	C	T	C	c g
4	t c	A	**A**	**A**	**A**	**A**	**C**	**A**	C	C	A	C	C	C	C	C	c g
8	t c	A	**A**	**A**	**A**	**A**	**C**	**A**	C	A	C	C	C	C	C	C	c g
10	t c	A	**A**	**A**	**A**	**A**	**C**	**A**	A	A	C	A	C	A	C	A	c g
11	t c	A	**A**	**A**	**A**	**A**	**C**	**A**	C	A	C	A	C	C	C	A	c g
12	t c	A	**A**	**A**	**A**	**A**	**C**	**A**	C	A	C	A	C	C	A	C	c g
13	t c	C	**A**	**A**	**A**	**A**	**C**	**A**	C	A	C	C	C	C	C	C	c g
16	t c	A	**A**	**A**	**A**	**A**	**C**	**A**	C	A	C	A	C	C	C	A	c g
20	t c	A	**A**	**A**	**A**	**A**	**C**	**A**	G	C	A	C	A	C	C	C	c g
26	t c	A	**A**	**A**	**A**	**A**	**C**	**A**	C	A	C	A	C	C	C	G	c g
28	t c	A	**A**	**A**	**A**	**A**	**C**	**A**	C	A	C	A	C	C	A	C	c g
30	t c	A	**A**	**A**	**A**	**A**	**C**	**A**	C	C	A	C	G	C	C	C	c g
31	t c	A	**A**	**A**	**A**	**A**	**C**	**A**	C	A	C	A	C	C	C	C	c g
G		0	0	0	0	0	0	0	1	0	4	0	1	0	0	1	
A		14	15	15	15	15	0	15	1	11	0	9	1	1	2	3	
T		0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
C		1	0	0	0	0	15	0	13	4	11	6	13	14	12	11	
%		93	100	100	100	100	100	100	87	73	73	100	87	93	80	73	
		A	**A**	**A**	**A**	**A**	**C**	**A**	**C**	**A**	**C**	**A/C**	**C**	**C**	**C**	**C**	

FIGURE 2: Consensus sequence identified by the SELEX experiment. The upper part shows the alignment of 15 sequenced clones obtained after 10 cycles of selection. The sequence of the 15 positions that were random before the selection is in capital letters, and the 2 bp at each side of this sequence are in lowercase letters. The sequences that are identical in all of the fragments are in bold letters and are highlighted in gray for the randomized part. The lower part shows the number and the percentage of the sequences with the indicated base at that position. The consensus sequence deduced from this alignment is also shown.

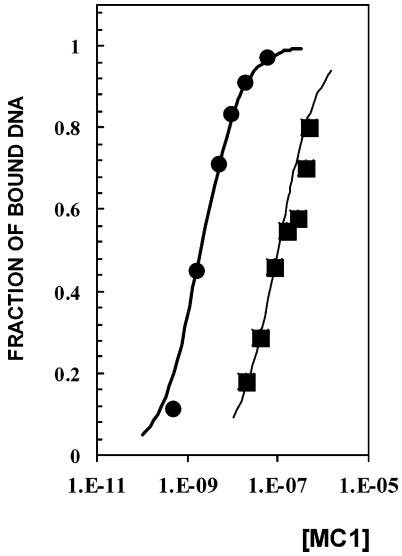

FIGURE 3: Representative MC1 protein−DNA binding curves with the consensus sequence and with random sequences. The fraction of bound DNA was obtained by EMSA experiments as a function of molar protein concentration [MC1]. The lanes are the best obtained with the equilibrium constant $K_D = 100$ nM for the random sequences (squares) and $K_D = 2$ nM for the consensus sequence (circles). The binding buffer is 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 100 mM NaCl, and 15 μg/mL BSA.

radical footprinting experiments on a DNA fragment containing the sequence $A_5(CA)_3C_4$ at its center. Densitometric scanning of the resulting gels shows protections on two regions of each DNA strand (Figure 4A, top strand, and Figure 4B, bottom strand). The intensities of the bands were quantitatively analyzed, and a schematic representation of the results is shown on Figure 4C. The areas of protection are 3−5 nt long, and the centers of these protections are nearly one full helical turn apart and offset by a few nucleotides in the 3′ direction on complementary strands. Such a pattern is characteristic of a protein binding to the minor groove mainly to one face of the DNA helix. To differently investigate the DNA binding mode of the protein, we have also performed experiments with distamycin A, a minor groove binder, that preferentially recognizes short AT tracts and that is a probe of protein−DNA interactions (*35*, *36*). Figure 4D shows that distamycin strongly destabilizes MC1 binding to a SELEX DNA fragment. The same result was obtained when the consensus sequence was preincubated with distamycin and MC1 was then added, clearly indicating that there is competition between MC1 and distamycin for binding to the consensus sequence. On the whole, our results strongly suggest that the protein binds to DNA through the minor groove.

To assess for DNA−protein contacts at the major groove, we used the alkylating agent DMS, which methylates guanine bases at their N7 position. These experiments were done only with the bottom strand that contains seven guanines. The comparison of DNA methylation in the presence and absence of the protein shows no guanine protection but three enhancements of the guanine's reactivity (Figure 5). These results suggest that there is a strong distortion of the DNA duplex on this stretch of the binding site (G10−G13) but no major groove contact.

We also performed DMS interference experiments. After a limited methylation of guanines with DMS, the DNA was incubated with MC1 and electrophoresed through a native polyacrylamide gel. After the scission of eluted DNA at the methylated positions, the DNA fragments were separated on a denaturing sequencing gel (Figure 6A). Comparison of the pattern of band intensities from protein-free DNA (lane F) and the DNA−protein complex (lane B) shows that the methylation of at least one guanine (G12) interferes with the protein binding (Figure 6B). This result may appear to conflict with the previous one since no contact in the major groove is observed at that position. The most likely explanation is that this result is due to an indirect effect and that the

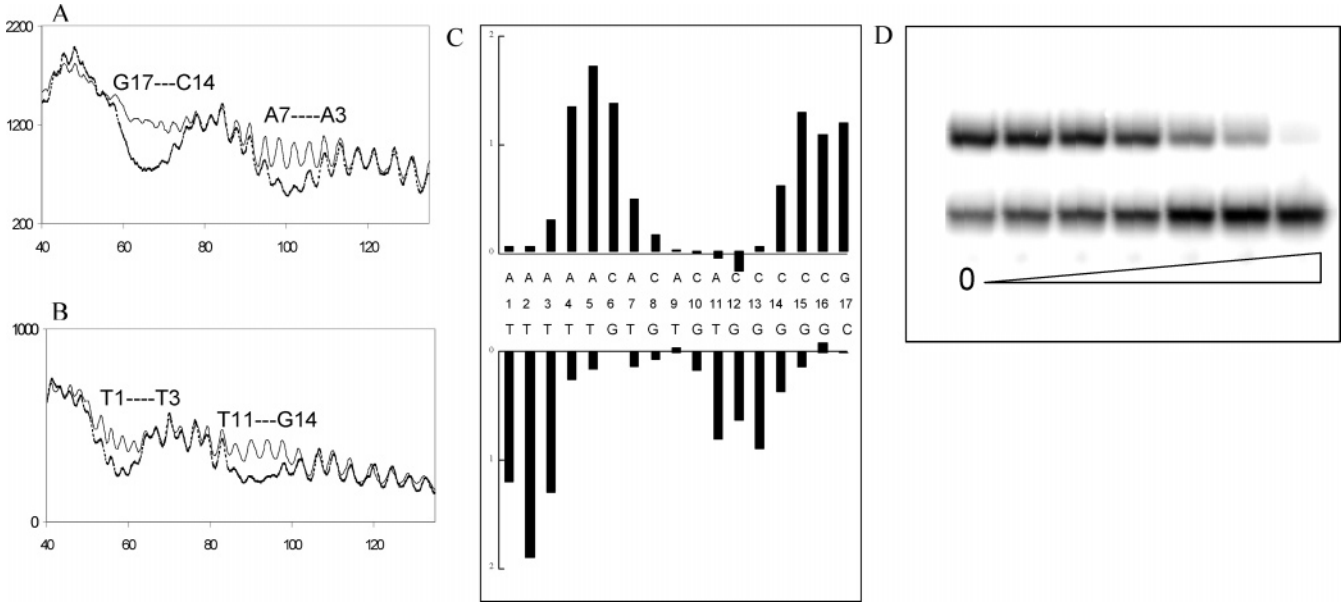

FIGURE 4: (A−C) Hydroxyl radical footprinting of the MC1 protein bound to the consensus sequence. The data represent scans quantitating the cleavage of the top strand (A) and of the bottom strand (B) by hydroxyl radicals. The thin lines correspond to the cleavage pattern for free DNA, and the thick lines correspond to the cleavage pattern for DNA in the presence of the protein. Regions of protection are indicated. The histogram shows the relative cleavage frequency of the bound DNA compared to the unbound control (C). (D) Effect of distamycin on MC1−DNA complexes. A ^{32}P-labeled 26 bp DNA (1 nM) were first incubated with the MC1 protein (6 nM) for 15 min before the addition of distamycin for 40 min. The mixture was then loaded on a 8% native polyacrylamide gel. The distamycin A concentration was 0, 0.04, 0.09, 0.18, 0.35, 0.71, and 1.4 μM.

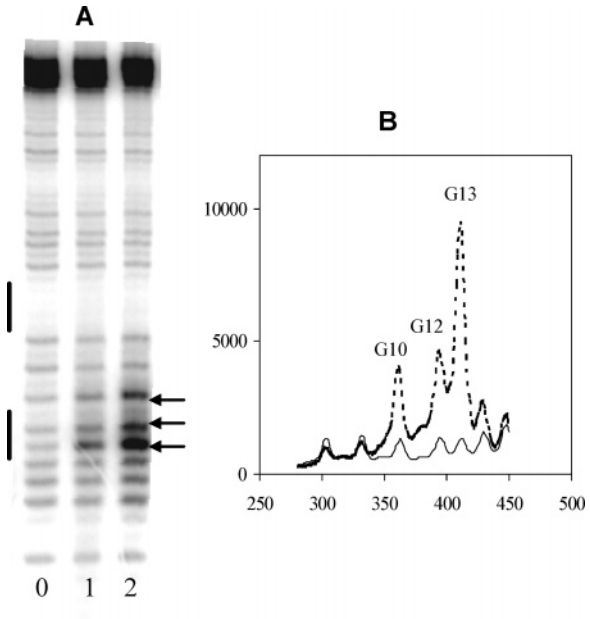

FIGURE 5: DMS protection of MC1 binding to the bottom strand of the consensus sequence. (A) Storage phosphor autoradiogram of a DMS footprint experiment. The reactions are performed in the absence of protein (lane 0) and in the presence of 12.5 nM (lane 1) or 25 nM MC1 protein (lane 2). The position of guanines whose DMS reactivity is enhanced by the MC1 binding is indicated (arrows), and the two regions protected from hydroxyl radical attacks are indicated (vertical lines). (B) Scans of the gel presented in the part A. The continuous line corresponds to lane 0 (free DNA), and the dotted line corresponds to lane 2 (DNA in the presence of the protein MC1).

guanine methylation prevents or limits the conformational change undergone by the DNA within the protein−DNA complex.

MD Experiments. To obtain information on the conformation and on the flexibility of the sequences preferentially recognized by MC1, we performed a MD simulation on a 16 bp DNA duplex containing the consensus sequence plus a 5′ end C/G base pair to stabilize this extremity (this base pair is numbered −1 to keep the same numeration of the sequence shown on Figure 6). The RMSD from canonical B-DNA has the same value (3.9 Å) when averaged over the first 3.5 ns of the simulation or over the last nanosecond. The different structural parameters exhibit also the same average values when they are calculated over the first or the last part of the simulation. Therefore, it is not expected that statistically significant differences will be observed by prolonging the simulation by a factor of 2 or 3.

The sugar ring pucker and glycosidic angles of all nucleotides are characteristic of a B conformation except for A9, the pucker of which exhibits a north conformation. Dihedral angles describing the backbone have also classical average values with the exception of $\alpha = +60°$ for G13 and $\gamma = -165°$ for G12 instead of $(-70 \pm 12)°$ and $(+55 \pm 10)°$, respectively, for all of the other nucleotides. Only G13 is in the BII conformation on average, all other nucleotides being in BI. The major structural feature is an axial curvature of 30° due to three additive successive kinks located at C8/G8, A9/T9, and C10/G10, which curve the oligonucleotide toward the major groove at the level of A9/T9 (Figure 7). These kinks are observed throughout the simulation. Two other less pronounced kinks are observed. The first one located between the A4/T4 and A5/T5 curves of the DNA by approximately 10°. The second one (approximately 20° on average) is observed during the simulation between C12/G12 and C14/G14 but without a preferential direction, reflecting the flexibility of the sequence at this level. Globally, the overall end-to-end curvature of the helix axis is in the range 10−50° during the simulation with an average value of 26°, which is mainly due to the bending around A9/T9.

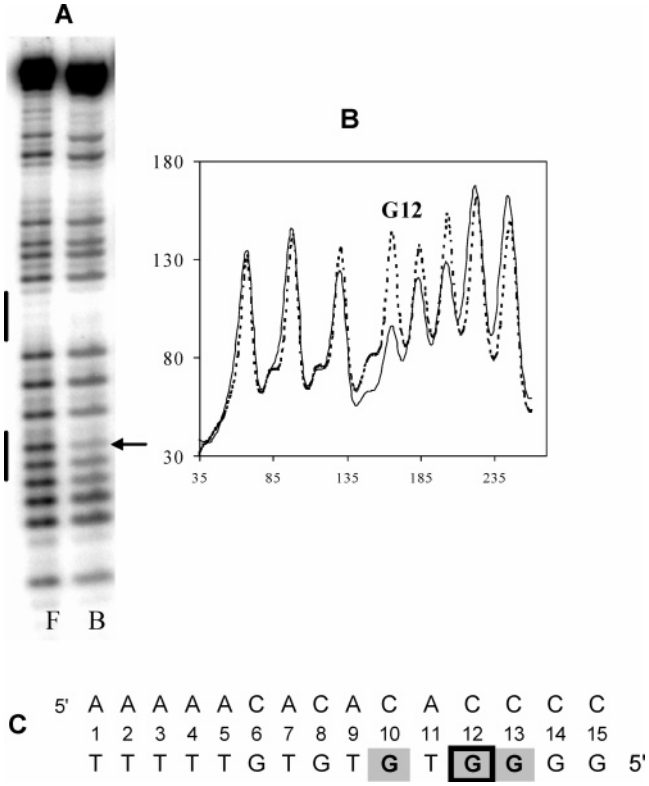

FIGURE 6: DMS interference analysis. A DNA fragment containing the sequence consensus was randomly methylated by DMS and incubated with the MC1 protein. Bound (B) and free (F) DNA fractions were separately isolated from a nondenaturing gel, then cleaved at the methylated bases, and finally loaded on a sequencing gel. (A) Storage phosphor autoradiogram of the results. The position of guanine G12 is indicated by an arrow, and the two regions protected from hydroxyl radical attacks are indicated by vertical lines. (B) Densitometric scans of lane B (continuous line) and of lane F (dotted line) are shown. (C) The position of the guanine whose premethylation strongly interferes with the MC1 binding is boxed, and the three guanines whose DMS reactivity is enhanced by the MC1 binding are highlighted in gray.

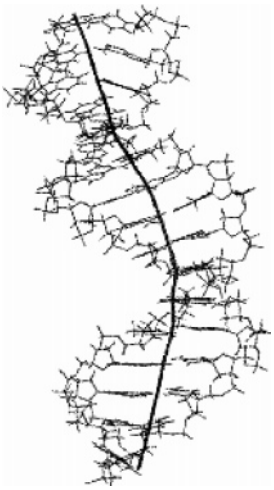

FIGURE 7: Duplex axis display of the minimized average structure obtained from MD simulation. The A track is on the top.

Because some guanines have particular conformations or high flexibility (G12, G13, G14), the solvent average accessible area of their N7 atoms was investigated. The values which are obtained (8.3−13.6 Å^2) are clearly larger than those of a classical B-DNA (7.5 ± 0.7 Å^2).

The top strand is found to be more flexible than the other one, with seven phosphate residues between A2 and A11 adopting temporary short-time BII conformations. The second strand appears quite rigid in the 3′ end region between nucleotides T1 and G6 and more flexible in the 5′ end region with five residues between T7 and G12 performing a few short time incursions toward the BII conformation.

In summary, the sequence is curved, and the CG-rich half part which contains nucleotides with some deviation from the usual BI form of DNA is more flexible than the AT-rich one.

DISCUSSION

Starting from a random DNA library of 4^{15} (10^9) different molecules, we have shown that the structural MC1 protein recognizes some particular DNA sequences. Higher affinity sequences have been detected from the fifth cycle, and the SELEX experiment has been stopped at the tenth cycle (at the end of the selection process). The selected sequences are highly conserved, and a 15 bp consensus sequence with only one partially degenerated base pair [AAAAACACAC-(A/C)CCCC] is obtained. This sequence motif is surprisingly long for such a small monomeric protein and longer than the excluded binding site (10−11 bp) that was previously determined using DNA fragments of random sequence (*18*). On the other hand, the sequence specificity of the protein is quite low (approximately a 50-fold greater affinity for the consensus sequence relative to a random substrate).

Obtaining a strong affinity sequence ($K_D = 2$ nM) allowed us to investigate the DNA binding mode of the protein. Hydroxyl radical footprintings together with a distamycin competition experiment suggest that the protein binds to DNA through the minor groove. The binding site is composed of two areas of contact separated by nearly 10 bp. The major groove between these two regions does not seem to be in interaction with the protein as suggested by the DMS footprinting results.

Many, if not all, architectural proteins interact with DNA exclusively through minor groove contacts (*37*). Small monomeric proteins such as the Sso7d/Sac7d protein family (*9*, *10*) or the HMG box of the HMGB1/2 protein family (*38*−*40*) strongly bend DNA by intercalating hydrophobic amino acids between base steps in a widened minor groove. MC1 might use a similar mode of DNA bending. Trp74 or Met75 whose oxidations seem to be responsible for a concomitant loss of the protein DNA bending and specificity (*41*) could be good candidates to accomplish such an intercalation process. In the complexes, the main body of the MC1 protein would then be located on the convex side of the bent DNA. Taking into account the overall dimensions of the protein (in the range of 30−40 Å), this assumption does not seem to be in agreement with the two protein−DNA contacts detected by the footprinting experiments. It is possible that MC1 is located inward of the bent DNA. In that way, the numerous cationic amino acids (9 arginines and 13 lysines) exposed at the surface of the protein could contribute to the DNA bending by an asymmetric phosphate neutralization process (*42*, *43*). Of course, only structural studies will be able to decide between these conflicting hypotheses. Using the consensus sequence such studies are now in progress in our laboratory.

Given its very low binding specificity ($\Delta\Delta G = -2.2$ kcal/mol) for a relatively long consensus sequence, it is unlikely that MC1 sets up direct specific interactions with some functional groups of DNA bases. The protein probably binds to DNA by an indirect mechanism, i.e., by recognizing structural characteristics of the double helix. So, the A_4CA motif, which is present in all of the selected sequences, is certainly important in the recognition process. One may speculate that the protein differentiates some particularities of the minor groove of the adenine tract (*44*, *45*). In addition, the energy cost necessary to deform the DNA is probably lower with the selected sequences either because the conformation of these naked sequences looks like the one in the complex and/or because these sequences are more easily deformable. Our MD simulation data support both hypotheses. So the protein could recognize one or several DNA kinks that induce a global curvature of the consensus sequence. This recognition is doubtless very subtle since the protein does not preferentially bind to curved DNA containing phased A tracts (unpublished results) and binds very poorly to DNA containing bulged adenine bases (*22*). The 3′ end of the consensus sequence might be recognized because this G/C-rich region (i) deviates from a classical B DNA and (ii) is flexible. It is therefore interesting to point out that the accessibility of N7 of G12 and G13 is high both in the free state of the DNA (as shown by water accessibility calculations) and in the complexes (as shown by the enhanced DMS reactivity). A direct comparison between the experimental and calculated accessibilities is, however, not obvious as in the first case the probe (DMS) is much larger than the probe (water) in the latter case. The higher accessibility of water to N7 of particular guanines certainly reflects the larger flexibility of the molecule at this level. In addition, the global flexibility of the consensus sequence as well as the incursions of some phosphate residues toward the BII conformation might also be recognized by MC1, but our data on these parameters cannot be interpreted without knowledge of the structure of the MC1−DNA complex.

We have searched how frequently the selected sequences are present in the two completely sequenced genomes of methanosarcinales (*Methanosarcina acetivorans* and *Methanosarcina mazei*). No exact match of the sequence consensus was found, but the motif $A_5CACACA$ (included in sequences comprising from 11 to 14 bp of the consensus sequence) appears 20 times in each genome. This occurrence is slightly higher (about 7 times) than the one predicted by random chance, but as the location of these sequences in the genomes seems unpredictable, the biological significance of these sequences cannot be assessed by our in vitro work. In several respects our results resemble those obtained with eukaryotic histones for which (i) the binding specificities are quite low [$\Delta\Delta G < -4$ kcal/mol (*46*)] and (ii) the strongest natural positioning sequences have affinities lower than those obtained by nonnatural sequences selected by SELEX. Naturally, we do not know whether, as in the case of histones, some preferential positioning of the MC1 protein exists. If so, various factors (such as the presence of specific DNA binding proteins and the topological state of the DNA) should be involved, but our results suggest that the DNA sequence could also take part in such a process. In any case, the understanding of this particular protein−DNA sequence recognition which remains to be elucidated is interesting since this process is certainly linked to the physicochemical properties of the double helix.

ACKNOWLEDGMENT

We thank Alain Gervais for the preparation of the protein and Jean Claude Maurizot for helpful comments.

REFERENCES

1. Reeve, J. N. (2003) Archaeal chromatin and transcription, *Mol. Microbiol. 48*, 587−598.
2. Bailey, K. A., Marc, F., Sandman, K., and Reeve, J. N. (2002) Both DNA and histone fold sequences contribute to archaeal nucleosome stability, *J. Biol. Chem. 277*, 9293−9301.
3. Musgrave, D., Zhang, X., and Dinger, M. (2002) Archaeal genome organization and stress responses: implications for the origin and evolution of cellular life, *Astrobiology 2*, 241−253.
4. Musgrave, D. R., Sandman, K. M., and Reeve, J. N. (1991) DNA binding by the archaeal histone HMf results in positive supercoiling, *Proc. Natl. Acad. Sci. U.S.A. 88*, 10397−10401.
5. Reeve, J. N., Bailey, K. A., Li, W. T., Marc, F., Sandman, K., and Soares, D. J. (2004) Archaeal histones: structures, stability and DNA binding, *Biochem. Soc. Trans. 32*, 227−230.
6. Tomschik, M., Karymov, M. A., Zlatanova, J., and Leuba, S. H. (2001) The archaeal histone fold protein HMf organizes DNA into bona fide chromatin fibers, *Structure (Cambridge) 9*, 1201−1211.
7. Napoli, A., Zivanovic, Y., Bocs, C., Buhler, C., Rossi, M., Forterre, P., and Ciaramella, M. (2002) DNA bending, compaction and negative supercoiling by the architectural protein Sso7d of *Sulfolobus solfataricus*, *Nucleic Acids Res. 30*, 2656−2662.
8. Choli, T., Henning, P., Wittmann-Liebold, B., and Reinhardt, R. (1988) Isolation, characterization and microsequence analysis of a small basic methylated DNA-binding protein from the Archaebacterium, *Sulfolobus solfataricus*, *Biochim. Biophys. Acta 950*, 193−203.
9. Krueger, J. K., McCrary, B. S., Wang, A. H., Shriver, J. W., Trewhella, J., and Edmondson, S. P. (1999) The solution structure of the Sac7d/DNA complex: a small-angle X-ray scattering study, *Biochemistry 38*, 10247−10255.
10. Su, S., Gao, Y. G., Robinson, H., Liaw, Y. C., Edmondson, S. P., Shriver, J. W., and Wang, A. H. (2000) Crystal structures of the chromosomal proteins Sso7d/Sac7d bound to DNA containing T-G mismatched base-pairs, *J. Mol. Biol. 303*, 395−403.
11. Dinger, M. E., Baillie, G. J., and Musgrave, D. R. (2000) Growth phase-dependent expression and degradation of histones in the thermophilic archaeon *Thermococcus zilligii*, *Mol. Microbiol. 36*, 876−885.
12. Chartier, F., Laine, B., and Sautiere, P. (1988) Characterization of the chromosomal protein MC1 from the thermophilic archaebacterium *Methanosarcina* sp. CHTI 55 and its effect on the thermal stability of DNA, *Biochim. Biophys. Acta 951*, 149−156.
13. Chartier, F., Laine, B., Belaiche, D., Touzel, J. P., and Sautiere, P. (1989) Primary structure of the chromosomal protein MC1 from the archaebacterium *Methanosarcina* sp. CHTI 55, *Biochim. Biophys. Acta 1008*, 309−314.
14. Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R. A., Martinez-Arias, R., Henne, A., Wiezer, A., Baumer, S., Jacobi, C., Bruggemann, H., Lienard, T., Christmann, A., Bomeke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H. P., Gunsalus, R. P., Fritz, H. J., and Gottschalk, G. (2002) The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea, *J. Mol. Microbiol. Biotechnol. 4*, 453−461.
15. Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., Macdonald, P., FitzHugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D., Brown, A., Allen, N., Naylor, J., Stange-Thomann, N., DeArellano, K., Johnson, R., Linton, L., McEwan, P., McKernan, K., Talamas, J., Tirrell, A., Ye, W., Zimmer, A., Barber, R. D., Cann, I., Graham, D. E., Grahame, D. A., Guss, A. M., Hedderich, R., Ingram-Smith, C., Kuettner, H. C., Krzycki, J. A., Leigh, J. A., Li, W., Liu, J., Mukhopadhyay, B., Reeve, J. N., Smith, K., Springer, T. A., Umayam, L. A., White, O., White, R. H., Conway de Macario, E., Ferry, J. G., Jarrell, K. F., Jing, H., Macario, A. J., Paulsen, I., Pritchett, M., Sowers, K. R., Swanson, R. V., Zinder, S. H., Lander, E., Metcalf, W. W., and Birren, B. (2002)

The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity, *Genome Res. 12*, 532−542.
16. Chartier, F., Laine, B., Belaiche, D., and Sautiere, P. (1989) Primary structure of the chromosomal proteins MC1a, MC1b, and MC1c from the archaebacterium *Methanothrix soehngenii*, *J. Biol. Chem. 264*, 17006−17015.
17. Paquet, F., Culard, F., Barbault, F., Maurizot, J. C., and Lancelot, G. (2004) NMR solution structure of the archaebacterial chromosomal protein MC1 reveals a new protein fold, *Biochemistry 43*, 14971−14978.
18. Culard, F., Laine, B., Sautiere, P., and Maurizot, J. C. (1993) Stoichiometry of the binding of chromosomal protein MC1 from the archaebacterium, *Methanosarcina* spp. CHTI55, to DNA, *FEBS Lett. 315*, 335−339.
19. Larquet, E., Le Cam, E., Fourcade, A., Culard, F., Furrer, P., and Delain, E. (1996) Complementarity of microscopies in the structural analysis of DNA minicircles associated to protein MC1, *C. R. Acad. Sci., Ser. III 319*, 461−471.
20. Toulme, F., Le Cam, E., Teyssier, C., Delain, E., Sautiere, P., Maurizot, J. C., and Culard, F. (1995) Conformational changes of DNA minicircles upon the binding of the arconogi, T. M., Alley, S. C., Harwood, E. A., Hopkins, P. B., and Robinson, B. H. (2002) Phosphate backbone neutralization increases duplex DNA flexibility: a model for protein binding, *Proc. Natl. Acad. Sci. U.S.A. 99*, 4156−4160.
21. Cam, E., Culard, F., Larquet, E., Delain, E., and Cognet, J. A. (1999) DNA bending induced by the archaebacterial histone-like protein MC1, *J. Mol. Biol. 285*, 1011−1021.
22. Paradinas, C., Gervais, A., Maurizot, J. C., and Culard, F. (1998) Structure-specific binding recognition of a methanogen chromosomal protein, *Eur. J. Biochem. 257*, 372−379.
23. Teyssier, C., Toulme, F., Touzel, J. P., Gervais, A., Maurizot, J. C., and Culard, F. (1996) Preferential binding of the archaebacterial histone-like MC1 protein to negatively supercoiled DNA minicircles, *Biochemistry 35*, 7954−7958.
24. Thastrom, A., Lowary, P. T., Widlund, H. R., Cao, H., Kubista, M., and Widom, J. (1999) Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences, *J. Mol. Biol. 288*, 213−229.
25. Lowary, P. T., and Widom, J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning, *J. Mol. Biol. 276*, 19−42.
26. Bailey, K. A., Pereira, S. L., Widom, J., and Reeve, J. N. (2000) Archaeal histone selection of nucleosome positioning sequences and the procaryotic origin of histone-dependent genome evolution, *J. Mol. Biol. 303*, 25−34.
27. He, Y. Y., Stockley, P. G., and Gold, L. (1996) In vitro evolution of the DNA binding sites of *Escherichia coli* methionine repressor, MetJ, *J. Mol. Biol. 255*, 55−66.
28. Riggs, A. D., Bourgeois, S., and Cohn, M. (1970) The lac repressor-operator interaction. 3. Kinetic studies, *J. Mol. Biol. 53*, 401−417.
29. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., III, Wang, J., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (2002) AMBER 7, University of California, San Francisco.
30. Cheatham, T. E., III, Cieplak, P., and Kollman, P. A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat, *J. Biomol. Struct. Dyn. 16*, 845−862.
31. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes, *J. Comput. Phys. 23*, 327−341.
32. Darden, T., York, D., and Pederson, L. J. (1993) Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems, *J. Chem. Phys. 103*, 8577−8593.
33. Lavery, R., and Sklenar, H. (1997) *Handbook for CURVES 5.2, Helical Analysis of Irregular Nucleic Acids*, Laboratory for Theoretical Biochemistry, CNRS, Paris, France.
34. Fraternali, F., and Cavallo, L. (2002) Parameter optimized surfaces (POPS): analysis of key interactions and conformational changes in the ribisome, *Nucleic Acids Res. 30*, 2950−2960.
35. Roque, A., Orrego, M., Ponte, I., and Suau, P. (2004) The preferential binding of histone H1 to DNA scaffold-associated regions is determined by its C-terminal domain, *Nucleic Acids Res. 32*, 6111−6119.
36. Chai, S., and Alsonso, J. C. (1996) Distamycin-induced inhibition of formation of a nucleoprotein complex between the terminase small subunit G1P and the non-encapsidated end (pacL site) of *Bacillus subtilis* bacteriophage SPP1, *Nucleic Acids Res. 24*, 282−288.
37. Bewley, C. A., Gronenborn, A. M., and Clore, G. M. (1998) Minor groove-binding architectural proteins: structure, function, and DNA recognition, *Annu. Rev. Biophys. Biomol. Struct. 27*, 105−131.
38. Murphy, F. V. T., Sweet, R. M., and Churchill, M. E. (1999) The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition, *EMBO J. 18*, 6610−6618.
39. Masse, J. E., Wong, B., Yen, Y. M., Allain, F. H., Johnson, R. C., and Feigon, J. (2002) The *S. cerevisiae* architectural HMGB protein NHP6A complexed with DNA: DNA and protein conformational changes upon binding, *J. Mol. Biol. 323*, 263−284.
40. Thomas, J. O., and Travers, A. A. (2001) HMG1 and 2, and related "architectural" DNA-binding proteins, *Trends Biochem. Sci. 26*, 167−174.
41. Culard, F., Gervais, A., De Vuyst, G., Spotheim-Maurizot, M., and Charlier, M. (2003) Response of a DNA-binding protein to radiation-induced oxidative stress, *J. Mol. Biol. 328*, 1185−1195.
42. Okonogi, T. M., Alley, S. C., Harwood, E. A., Hopkins, P. B., and Robinson, B. H. (2002) Phosphate backbone neutralization increases duplex DNA flexibility: a model for protein binding, *Proc. Natl. Acad. Sci. U.S.A. 99*, 4156−4160.
43. Strauss, J. K., and Maher, L. J., III (1994) DNA bending by asymmetric phosphate neutralization, *Science 266*, 1829−1834.
44. Dickerson, R. E. (1998) DNA bending: the prevalence of kinkiness and the virtues of normality, *Nucleic Acids Res. 26*, 1906−1926.
45. Katahira, M., Sugeta, H., Kyogoku, Y., Fujii, S., Fujisawa, R., and Tomita, K. (1988) One- and two-dimensional NMR studies on the conformation of DNA containing the oligo(dA)oligo(dT) tract, *Nucleic Acids Res. 16*, 8619−8632.
46. Thastrom, A., Bingham, L. M., and Widom, J. (2004) Nucleosomal locations of dominant DNA sequence motifs for histone-DNA interactions and nucleosome positioning, *J. Mol. Biol. 338*, 695−709.

BI0474416