

Model of a DNA-Protein Complex of the Architectural Monomeric Protein MC1 from *Euryarchaea*

Françoise Paquet^{1*}, Olivier Delalande², Stephane Goffinont¹, Françoise Culard¹, Karine Loth¹, Ulysse Asseline¹, Bertrand Castaing¹, Celine Landon¹

1 Centre de Biophysique Moléculaire, Centre National de la Recherche Scientifique UPR 4301, Université d'Orléans, Orleans, France, **2** Faculté des Sciences Pharmaceutiques et Biologiques, Institut de Génétique et Développement de Rennes, Centre National de la Recherche Scientifique UMR 6290, Université de Rennes1, Rennes, France

Abstract

In *Archaea* the two major modes of DNA packaging are wrapping by histone proteins or bending by architectural non-histone proteins. To supplement our knowledge about the binding mode of the different DNA-bending proteins observed across the three domains of life, we present here the first model of a complex in which the monomeric Methanogen Chromosomal protein 1 (MC1) from *Euryarchaea* binds to the concave side of a strongly bent DNA. In laboratory growth conditions MC1 is the most abundant architectural protein present in *Methanosarcina thermophila* CHTI55. Like most proteins that strongly bend DNA, MC1 is known to bind in the minor groove. Interaction areas for MC1 and DNA were mapped by Nuclear Magnetic Resonance (NMR) data. The polarity of protein binding was determined using paramagnetic probes attached to the DNA. The first structural model of the DNA-MC1 complex we propose here was obtained by two complementary docking approaches and is in good agreement with the experimental data previously provided by electron microscopy and biochemistry. Residues essential to DNA-binding and -bending were highlighted and confirmed by site-directed mutagenesis. It was found that the Arg25 side-chain was essential to neutralize the negative charge of two phosphates that come very close in response to a dramatic curvature of the DNA.

Citation: Paquet F, Delalande O, Goffinont S, Culard F, Loth K, et al. (2014) Model of a DNA-Protein Complex of the Architectural Monomeric Protein MC1 from *Euryarchaea*. PLoS ONE 9(2): e88809. doi:10.1371/journal.pone.0088809

Editor: Jun-Tao Guo, The University of North Carolina at Charlotte, United States of America

Received: October 3, 2013; **Accepted:** January 11, 2014; **Published:** February 18, 2014

Copyright: © 2014 Paquet et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Funds result from the CNRS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: francoise.paquet@cnrs-orleans.fr

Introduction

The genomic DNA of all organisms across the three domains of life needs to be compacted and functionally organized. Wrapping-proteins, bending-proteins and bridging-proteins are thus involved and it appears that the underlying mechanisms are similar among *Bacteria*, *Archaea* and *Eukaryota* [1]. The two major modes of DNA packaging are 1) wrapping by histone proteins in *Eukaryota* (H2, H3 and H4 core histones) and *Archaea* (HMf histones), and 2) bending by architectural non-histone proteins in *Bacteria* (HU/IHF/Fis) and *Archaea* (Cren7/Sul7/MC1). DNA-bridging proteins have also been found in *Eukaryota* (H1 linker histone), *Archaea* (Alba) and *Bacteria* (H-NS). Alba is the second most widely distributed archaeal chromatin protein after the archaeal histones [2,3]. All *Crenarchaea* and *Euryarchaea* encode Alba with the exception of *Methanosarcina*, *Halobacteria* and *Thermoplasma* species [4]. Possibly not coincidentally, MC1 is present when Alba is absent. *Thermoplasma acidophilum*, another *Euryarchaea*, which is known to lack archaeal histones, encodes the HTa protein, a homolog of bacterial chromatin HU [5,6].

MC1 is a member of the archaeal bending proteins and is the most abundant architectural protein present in *Methanosarcina thermophila* CHTI55 in laboratory growth conditions [7]. This small basic monomeric protein of 93 residues is structurally unrelated to other DNA-binding proteins [8,9]. Its global fold consists of a pseudo barrel with an extension of the β -sheet (β 4– β 5)

forming an arm (LP5). The secondary structure elements, namely an α -helix (Arg25–Arg34) and five β -strands β 1 (Arg4–Arg9), β 2 (Glu15–Gly21), β 3 (Asp43–Glu49), β 4 (Lys54–Asp66), and β 5 (Lys78–Glu90), are all antiparallel and packed to each other. An antiparallel β -bulge composed of Val57, Glu87 and Arg88 is always present. The secondary structure elements are connected by five loops LP1 (Asp10–Asn14), LP2 (Lys22–Pro24), LP3 (Gly35–Pro42), LP4 (Arg50–Lys53), and LP5 (Ala67–Glu77).

MC1 protects DNA against thermal denaturation and shapes DNA by binding to it [7]. Its affinity for any double-stranded DNA is high ($K_D \approx 100$ nM) and it recognizes and preferentially binds to bent DNA, such as four-way junctions (≈ 4 -fold affinity) [10] and negatively supercoiled DNA minicircles (> 10 -fold affinity) [11]. More recently a SELEX (systematic evolution of ligands by exponential experiment) procedure revealed that it preferentially binds to a linear 15 base pair (bp) motif [AAAAACACAC(A/C)CCC(C/A)] with a particularly strong affinity ($K_D \approx 2$ nM) [12], making it possible to investigate the DNA-binding mode of the protein by NMR spectroscopy. Visualization of DNA duplexes (176 bp) by electron microscopy revealed that the binding of MC1 induces sharp kinks so that the overall bend angle is estimated at 116° [13]. MC1 is too small to be directly visualized by electron microscopy but the relatively large segment of DNA (20–30 bp) that is protected from DNase I suggests that DNA is wrapped around MC1 [14]. Hydroxyl radical footprinting together with a distamycin competition experiment demonstrated that the protein

binds to DNA through its minor groove and that the binding site is composed of two areas of contact separated by approximately 10 bp [12]. Many other examples of architectural proteins that interact exclusively with the minor groove of DNA are found in the literature [15]: the TATA-binding protein (TBP), the male sex determining factor SRY, the lymphoid enhancer-binding factor 1 (LEF-1), the integration host factor (IHF) and the high mobility group I (HMGI). All of these minor groove-binding proteins bind with high affinity and varying degrees of sequence-specificity. Moreover, they exhibit very different global folds and use different strategies, or combinations of strategies, for recognition and binding.

We present here a new type of DNA-protein complex in which the monomeric protein MC1 binds on the concave side of the strongly bent DNA. Models were obtained by two successive and complementary docking approaches. The first one is based on a coarse-grained and interactive approach: a flexible and initially mildly-kinked double-strand DNA is driven into the electrostatic potential of the free static MC1 protein, thus producing a more kinked structure of the double-strand obtained upon protein interaction. The second step is flexible docking. It takes into account i) the 15 NMR models of the free MC1 protein to model the flexibility of the LP5 loop with regard to the core of the protein and ii) the 15 bp oligonucleotide generated in the first step and guided by the defined DNA-protein interface as ambiguous interaction restraints. Models of the DNA/MC1 complex are in good agreement with the previous experimental data provided by electron microscopy and biochemistry. Residues essential to the DNA-binding and -bending are highlighted by the structural complex and confirmed by site-directed mutagenesis.

Materials and Methods

DNA sequences

Only top strands are shown:

15 bp used for NMR and All Atom docking (**AAAAACACA-CACCCA**)

21 bp used for coarse grained docking (**CTCAAAAAACACA-CACCCACGA**)

26 bp used for EMSA (**GCCTCAAAAAACACACCCCCG-GACAG**).

NMR sample preparation

The free protein NMR sample was prepared by concentrating ^{15}N -MC1 to 1.6 mM (10 mM phosphate buffer pH 6, 100 mM NaCl, 1 mM EDTA, 10% D_2O) as described previously [9].

Synthetic oligodeoxyribonucleotides (OligoGold oligonucleotide quality) were purchased from Eurogentec (Liège, Belgium). The single-stranded oligodeoxyribonucleotides were characterized by NMR and annealed at a 1:1 ratio. The free duplex concentration in the NMR sample was 2.5 mM (10% D_2O).

The ^{15}N -MC1/DNA complex was prepared by slowly adding the 7.5 mM DNA duplex solution (10 mM phosphate buffer pH 6, 100 mM NaCl, 1 mM EDTA, 10% D_2O) to the 1.6 mM protein solution to give a final concentration of 1 mM.

The EDTA-C2-dT phosphoramidite was purchased from Eurogentec (Liège, Belgium). The EDTA labeled strands DNA1* and DNA2* were solid-phase synthesized, purified by RP-HPLC and characterized by ^1H NMR. After annealing of the complementary strands, coordination of manganese was achieved by addition of 1.1 molar equivalent of MnCl_2 to DNA1* or DNA2* solutions at pH 6. Chelex-100 beads (BioRad) were added to the DNA* Mn^{2+} solutions to remove metal excess. NMR samples for $^1\text{H}_\text{N}$ - Γ_2 paramagnetic relaxation enhancement (PRE)

measurements were prepared by mixing the purified protein ^{15}N -MC1 and the DNA* Mn^{2+} at a molar ratio of 1:1.2. The final complex concentration was 0.35 mM in 10 mM phosphate buffer pH 6, 100 mM NaCl and 10% D_2O .

NMR spectroscopy

All of the NMR experiments were performed on a 600 MHz Varian $^{\text{UNITY}}$ INOVA spectrometer at 299 K. DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) was used as reference. Spectra were processed using NMRPIPE [16], and analyzed with NMRVIEW [17].

The ^{15}N and $^1\text{H}_\text{N}$ chemical shifts for free MC1 have already been reported [9] and those for bound MC1 were assigned from a combination of ^1H - ^{15}N HSQC (heteronuclear single quantum coherence) and 2D ^{15}N NOESY-HSQC (mixing time of 150 ms) spectra. Chemical shift perturbation (CSP) data of the protein were calculated and analyzed with SAMPLEX [18].

The ^1H NMR resonances of the free 15-mer dsDNA were assigned from a combination of NOESY (nuclear Overhauser effect spectroscopy) (mixing times of 150 and 300 ms) and TOCSY (total correlation spectroscopy) (mixing time of 60 ms) spectra. The bound form was studied using NOESY (mixing times of 150 and 300 ms) and TOCSY (mixing time of 80 ms) spectra.

$^1\text{H}_\text{N}$ - Γ_2 values were determined as a difference in transverse relaxation rates (R_2) for the paramagnetic and diamagnetic states of the bound protein. R_2 for the diamagnetic state was given previously [9] and R_2 for the paramagnetic state was measured with R_2 relaxation delays of 2, 3, 4, 5, 6, 7, 8, 10, 12, 14, 16, 18, 20 and 24 ms. Volumes for the amide ^{15}N - ^1H cross peaks were measured using NMRVIEW [17] and fitted with a single exponential decay function.

Structure analysis

The Adaptive Poisson-Boltzmann Solver (APBS) program [19] was used within PyMOL to display the results of the calculations as an electrostatic potential molecular surface. The Platinum software [20] was used to calculate and visualize molecular hydrophobic/hydrophilic properties using the concept of "Molecular Hydrophobicity Potential" (MHP). Sequence homology and alignment were performed using online Protein BLAST software (blast.ncbi.nlm.nih.gov). The structures of nucleic acids were analysed using the CURVES+ software [21]. The figures were drawn with VMD [22], PyMOL [23] and MOLMOL [24].

Docking of DNA onto the MC1 protein

Several steps were used to obtain DNA/MC1 complexes and are schematically reported in Supplementary Figure S1.

Coarse-grained docking with BIOSPRING. An interactive docking experiment was performed using the BioSpring program [25]. Flexible double-strand DNA was driven into the electrostatic potential of a static coarse-grained molecular shape of the free MC1 protein corresponding to the first model of the 2KHL PDB structures. DNA should be strongly curved in the final complex, as observed from electron microscopy experiments [13]. Taking this into account and in order to obtain a better convergence of the simulation, two kinked-DNA conformers were extracted from DNA-protein complexes available in the PDB (1A74, recognition DNA sequence of I-PpoI [26] and 1YTB, yeast TATA-box) and were used as starting structures. Given the 10 bp distance between the two contact regions that was observed experimentally, a 15 bp oligonucleotide would be too short to achieve an accurate docking procedure; a 21 bp oligonucleotide was therefore modeled. This coarse-grained ligand was simulated as a flexible double-stranded DNA following the protocol proposed in our previous study and

using a 9 Å distance cutoff to form the elastic network [25]. The supplementary information that was used to interactively build the system and to place DNA in contact with the protein was as follows: i) the protein makes a double minor groove contact spaced by approximately ten base pairs in the double helix [12]; ii) Trp74 and Met75 are key residues to establish one of these contacts [27]; and iii) surface contact occurs near the T16 extremity (paramagnetic probe). We performed five replicas for each docking by considering each of the two different DNA structural patterns (1A74 and 1YTB).

High resolution model reconstruction of the DNA. All-atom 21 bp oligonucleotide models containing the specified sequence (AAAAACACACACCA) and a high curvature were generated using the 3D-Dart server [28]. Ten most relevant all-atom 3D-Dart models were selected after superimposition (best RMSD of 3 Å) onto the low-resolution oligonucleotides from the DNA-MC1 complex generated by the BioSpring approach. The DNA-MC1 atomic models were obtained after shortening the 21 bp oligonucleotide to the 15 bp length used experimentally.

High resolution docking using HADDOCK. The default protein-DNA docking protocol described by van Dijk and Bonvin [29] which consists of i) rigid-body docking (1000 models), ii) a semi-flexible refinement stage (200 models), and iii) final refinement in explicit solvent (200 models), was used for all the docking runs using the HADDOCK web server [30]. The 15 models of the free protein were used to model the flexibility of the LP5 loop (2KHL.pdb). DNA structures were extracted from the 10 MC1-DNA models obtained by the interactive docking. Each DNA model was docked to the 15 free conformations of MC1 in solution. Planarity and base pairing restraints were used during all stages of the docking to preserve the helical conformation of DNA. Both protein and DNA were defined to be semi-flexible on all their length after the rigid-body docking stage. Ambiguous interaction restraints based on experimental information (CSPs, dynamics, paramagnetic probes, mutations...) were used: the side chain protons of Lys86, Arg88 and Ile89 were constrained to approach at least one proton belonging to A2A3A4A5C6:G25T26T27T28T29 with a distance of 5 ± 1 Å. The side chain protons of Pro72, Trp74 and Met75 were constrained to approach at least one proton belonging to C13C14A15:T16G17G18 with a distance of 5 ± 1 Å (Supplementary Figure S2).

MC1 mutagenesis

Mutagenesis of MC1 was performed using standard protocols for site-directed mutagenesis with the plasmid pET24b-MC1 encoding for the MC1 protein. All of the mutants were verified by DNA sequencing of the entire gene. The proteins were expressed in BL-21(DE3) cells transformed with the appropriate plasmid and purified by three column chromatography steps (SP-Sepharose, Ultrogel AcA54, and Mono-S). The protein concentration was determined by absorption spectrophotometry, using the molecular absorbance coefficient of $11,000 \text{ M}^{-1} \text{ cm}^{-1}$ at 280 nm for wild-type and all mutants, except for the mutant Trp74Phe for which the molecular absorbance coefficient used was $5600 \text{ M}^{-1} \text{ cm}^{-1}$. Protein purity evaluated by SDS-PAGE was found to be better than 95 percent (Supplementary Figure S3). Correct molecular weights were observed by maldi-TOF mass spectrometry.

Electrophoretic mobility-shift assays

The duplex DNA (26 bp) used for EMSA (electrophoretic mobility-shift assays) experiments comprises the consensus sequence that was previously determined by SELEX [12]. The single-stranded oligonucleotide (MWG-Eurofins) was first

^{32}P -labeled at its 5' end then annealed to its complementary sequence in 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 150 mM NaCl by heating at 90°C for 3 min and slow cooling. EMSA reaction mixtures (10 μl) were prepared at 4°C by mixing DNA duplex and MC1 protein at concentrations indicated in the legend to each figure (Supplementary Figure S4), in binding buffer (10 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 15 $\mu\text{g}\cdot\text{ml}^{-1}$ BSA, and 8% (v/v) glycerol, pH 7.5), followed by incubation for 15 min at 20°C. The different mixtures were loaded onto a polyacrylamide gel in TBE buffer (89 mM Tris-HCl, pH 8.3, 89 mM boric acid, 1 mM EDTA). Electrophoresis was run at 14 V/cm, for 1 hour (K_D measurement) or 3 hours (bending effect) at 20°C. After drying, gels (Supplementary Figure S4A) were quantified using a β -scanner (Typhoon Trio, Molecular Dynamics). The binding curves (Supplementary Figure S4B, C) were fitted to a single binding site model using the equation $\{Y = [\text{MC1}] / ([\text{MC1}] + K_{Dapp})\}$ where Y is the fraction of bound DNA, [MC1] is the concentration of free protein, and K_{Dapp} is the apparent dissociation constant. Under our experimental conditions, the free protein concentration is reasonably approximated by the total protein concentration.

Results

Determination of the DNA-binding surface of MC1

As attested by the superimposition of ^1H - ^{15}N HSQC spectra between the free and the DNA-bound form of MC1 (Supplementary Figure S5), the overall structure of the protein is largely unchanged upon binding. Measurement of $^1\text{H}_N$ and ^{15}N chemical shifts on both the free and bound forms of MC1 yielded the chemical shift perturbations (CSP) along the sequence of MC1. These CSP data were analyzed using the SAMPLEX software [18] and are reported in Supplementary Figure S5. Significantly perturbed residues were located in five sites: the α -Helix from Pro24 to Arg34, three residues in the center of the β 3-strand (Ile45-Leu47), six residues belonging to the β 4-strand (Phe58-Glu63), part of the loop LP5 (Pro72-Pro76) and practically all of the residues constituting the β 5-strand (Val84-Glu90). It is worth noting that chemical shifts can report on both direct interaction and indirect effects such as remote conformational changes [31–33]. To refine our selection these CSP data were compared with the electrostatic potential of MC1 and with the flexible regions of MC1 defined by NMR relaxation data [9]. First, the CSP of residues located in the core of MC1 correlate with the basic patch constituted by the positive charges of Arg4, Lys22, Arg25, Lys30, Lys53, Lys54, His56, Arg71, Lys81, Lys85, Lys86, Arg88 and Lys91. Second, we have previously observed that Ile89 and the loop LP5 (67–77), mainly composed of hydrophobic residues (Ala67, Pro68, Pro72, Ala73, Trp74, Met75 and Pro76), possess considerable internal motions on the nanosecond time scale in the free protein and become much less mobile after binding with DNA [9], these residues have significant CSP too. In summary, we assume that residues of MC1, which present CSP after DNA-binding and belong to the basic surface or to the flexible regions, define the DNA-binding surface of MC1 (Figure 1). These residues belong to the α -helix (Arg25, Lys30), the LP5 arm (Pro72, Ala73, Trp74, Met75 and Pro76) and the β 5-strand (Lys85, Lys86, Arg88 and Ile89).

NMR features of the DNA after MC1-binding

Exchangeable imino protons. The hydrogen-bonded imino protons from the guanine and thymine nucleotides observed in the very low field region of the ^1H NMR spectrum are highly sensitive indicators of the Watson-Crick base pairing and stability of DNA

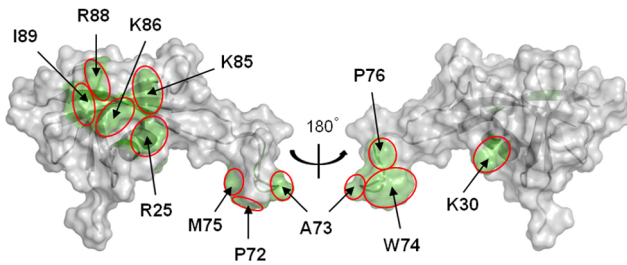


Figure 1. DNA-binding surface of MC1. The DNA-binding surface of MC1 is represented in green on the solvent-accessible surface. Residues which combine significant CSP and positive charge, or CSP and internal dynamics, are labeled.
doi:10.1371/journal.pone.0088809.g001

duplexes in solution. In the free 15 bp DNA, all of the thymine and guanine imino protons, with the exception of the terminal residues T16 and T30, were observable at 10°C indicating normal base stacking of a B-form for the duplex. In the bound DNA these imino protons plus T30 were observable at 25°C. MC1 stabilizes the duplex pairing of DNA as already observed for the DNA melting profiles with MC1 [7].

The largest CSP for the imino protons after binding were measured for G25 and T26 (>0.2 ppm), the imino proton of G25 is shielded and that of T26 is deshielded (Supplementary Figure S6A). This could be due to a local environment change such as a kink or an intercalation of amino acid side-chains between G25 and T26.

Nonexchangeable protons. The spectral region corresponding to the H1' (6.5–5 ppm), and H2, H6/H8 (8.4–7 ppm) DNA protons is quite free of protons belonging to the protein in the NOESY spectra of the DNA-MC1 complex (Supplementary Figure S7). Thus, it was possible to assign these protons except those belonging to G18, G19 and T20 due to overlapping. Their chemical shifts were compared to those of the free DNA and we observed that H1' of C6, C14 and T16 were particularly affected after binding (Supplementary Figure S6B). The H1' protons of the deoxyribose cycle point towards the minor groove; thus, the CSPs of H1' express perturbation in the minor groove. Moreover, the chemical environment of all the protons of the C6 nucleotide appears to be changed, supporting the hypothesis of a kink or an intercalation at the C6:G25 base pair level.

Paramagnetic probes attached to nucleic acid

Paramagnetic relaxation enhancement (PRE) of ^1H -nuclei is a well-known source of qualitative and quantitative long-range distance information [34]. To determine qualitatively and rapidly the bound orientation of the protein relative to the DNA, we introduced a paramagnetic label (dT-EDTA- Mn^{2+}) at the 5'-extremity of each DNA strand. The phosphoramidite for dT-EDTA is commercially available and can be incorporated at any chosen position using a DNA synthesizer. We synthesized two oligonucleotides containing a single specific binding site: the first labeled on the TG-rich strand at the 5'-terminal position (DNA1*) and the second labeled on the AC-rich strand by adding one thymine at the 5'-terminal position (DNA2*). ^1H - ^{15}N HSQC spectra of MC1 in the two complexes DNA1*/MC1 and DNA2*/MC1 showed no chemical shift differences but broadened signals, attesting that spin labeling does not change the DNA-protein interaction or the protein structure in the complex. $^1\text{H}_\text{N}$ - Γ_2 PRE is defined as the difference in the transverse relaxation rates of the paramagnetic (after adding MnCl_2) and diamagnetic states. $^1\text{H}_\text{N}$ - Γ_2 PREs were measured for the backbone amide groups ($^1\text{H}_\text{N}$).

Large magnitude $^1\text{H}_\text{N}$ - Γ_2 PREs were observed only for the regions close to the dT-EDTA- Mn^{2+} probes. A significant relaxation enhancement was observed for the residues Lys69-Lys81 belonging to the same face of the arm LP5 and located in the regions closest to the dT-EDTA- Mn^{2+} site of the DNA1*/ Mn^{2+} /MC1 complex, whereas the residues Lys54 and Val55 close to the dT-EDTA- Mn^{2+} site of the DNA2*/ Mn^{2+} /MC1 complex exhibited $^1\text{H}_\text{N}$ - Γ_2 values that were slightly higher (Figure 2 and supplementary Figure S8). These data immediately provided qualitative information about the orientation of the protein on the DNA duplex. Residues Lys69-Lys81 belonging to the arm LP5 are undeniably close to T16 and residues Lys54-Val55 are close to the 5'-supplementary thymine added before A1.

Molecular models of the DNA-MC1 complex

We performed a preliminary interactive docking experiment at low-resolution by simulating the association of a static molecular shape for MC1 with a flexible augmented elastic network model (aENM) for the oligonucleotide. All ten runs - i.e. driving of DNA double-strand through the electrostatic grid - strongly converged to highly reproducible DNA positions with an RMSD of 2 Å found on the phosphate groups (Figure 3A). This RMSD value is reasonable for low-resolution simulations. From these coarse-grained structures, all-atom reconstruction was performed to propose models of the DNA-MC1 complex (Figure 3B).

The average DNA curvature obtained from the 10 models is about $109 \pm 6^\circ$, which is a much greater angle than that of the starting DNA conformers (1A74: 49.3° and 1YTB: 71.4°) and in agreement with electron microscopy results [13]. The analysis of these models highlights two areas of contact in the minor groove located at the extremities of the 15 bp DNA. The first area is located in the β_5 -strand around Lys85, Lys86, Arg88 and Ile89 contacting the A-tract. The second area is part of the loop LP5 around Asn70, Arg71, Pro72, Trp74 and Met75 which contacts T16, G17 and G18. Strong additional electrostatic contacts involve A5C6 and the T20G21 phosphate backbone with Arg25 side-chains. These predicted models were in agreement with the experimental DNA-binding surface of MC1 defined above (Figure 1). However, the docking was performed with only one rigid protein model for MC1 and was not representative of the flexibility of the LP5 loop with regard to the core of the protein. We therefore performed a second docking with the commonly used HADDOCK program using the 15 free conformations of MC1 in solution and 10 15 bp DNA models extracted from the previous docking. Both the protein and DNA were considered semi-flexible molecules. Eight models of the DNA-MC1 complex were selected from the more populated clusters ($\sim 140/200$ pdb). Their bound DNA is bent with an average curvature of $104 \pm 21^\circ$ (Supplementary Figure S9). The angle of curvature depends on the position of the loop LP5 with regard to the core of MC1. As in the first models, the protein hangs on to the extremities of the DNA by inserting the expected residues Pro72, Trp74, Met75, Lys85, Lys86, Arg88 and Ile89 in the minor groove. The side chain of Lys22 is most probably involved in the minor groove interaction as well (Figure 4). As a consequence of this dramatic curvature, the A5pC6 and T20pG21 phosphates move closer to each other and the side chains of Arg25 and Gln23, which were not constrained by our ambiguous interaction restraints (AIRs) during the docking, are positioned in the major groove (Figure 4). In the center of the 15 bp oligonucleotide (CACACA region) the minor groove is widened and shallow whereas the major groove is narrow and deep (Supplementary Figure S9).

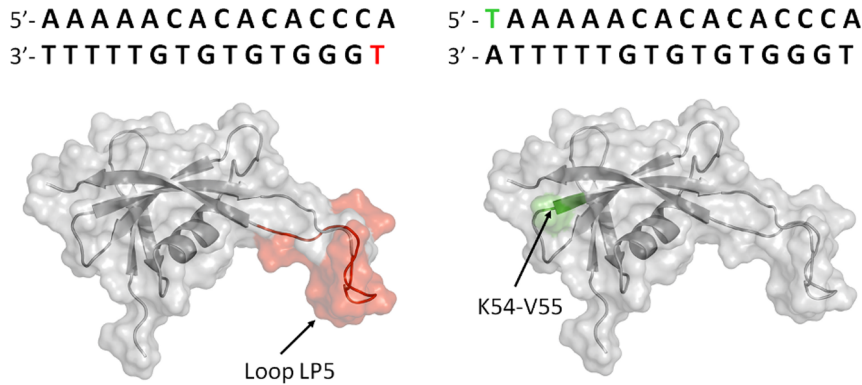


Figure 2. Relative orientation of the DNA extremity with respect to MC1. Sequence of the two oligonucleotides DNA1* (top) and DNA2* (bottom) including a paramagnetic probe; the color-coded T (red or green) represents dT-EDTA-Mn²⁺. The effect of each paramagnetic probe is shown on the three-dimensional surface of MC1 in red and green respectively.
doi:10.1371/journal.pone.0088809.g002

Functional validation of the DNA-MC1 model

Directed mutagenesis was conducted to determine whether amino acids predicted in our models to be in areas of contact with DNA are required for DNA binding. In a previous study, the DNA-binding affinity of Trp61, Trp74 and Met75 mutants was compared with that of the wild-type protein [27]. We showed that the two residues located in the LP5 arm are likely involved in the interaction since substituting Trp74 for Phe and Met75 for Leu led

to a decrease in the DNA-binding affinity. On the other hand, and in accordance with our current model, Trp61 located in the β 4-strand is probably not directly involved in DNA binding since its substitution for Phe had no effect. In the same study, we also showed that Trp74 and Met75 are involved in DNA bending since their substitution by Ala induced a loss of the capacity of the protein to recognize bent DNA as well as a strong reduction in the protein's ability for DNA bending. Finally, we observed that the substitution Pro76Ala does not affect DNA binding (unpublished results). To complete this analysis, we constructed a new series of MC1 mutants. Altogether, eight proteins were compared for their ability to bind a 26 bp DNA containing the consensus sequence used in the NMR experiments. EMSA were used to separate bound from free DNA (Figure 5 and Supplementary Figure S4) and to determine K_D values for each mutant (Table 1). The DNA affinities of the mutants are clearly in accordance with our model of DNA-MC1 binding, implying three area sites. First, we attest that LP5 is involved in DNA binding; substituting Arg71 or Pro72 for Ala induced a slight decrease in the affinity (\approx 3-fold) whereas the conservative substitution Trp74Phe induced a larger effect (7-fold). A second area involves Ile89 located in the β 5-strand, whose substitution for Ala greatly affected the interaction (12.8-fold

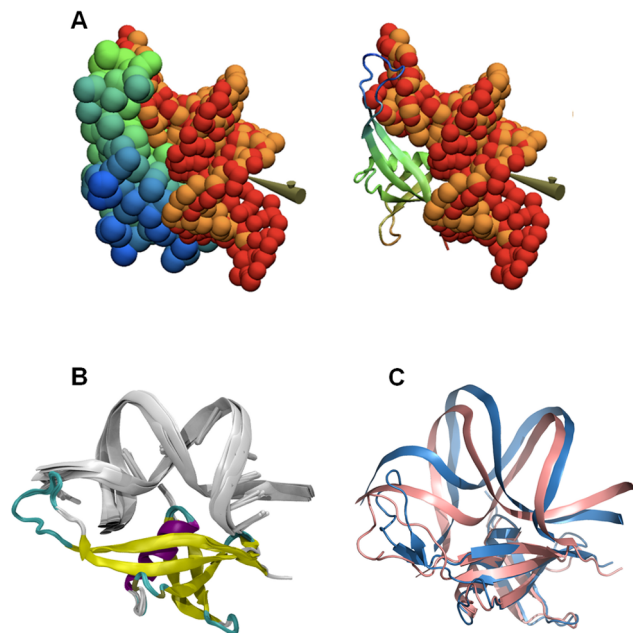


Figure 3. Proposed 3D model for MC1 bound to DNA. (A) Superimposition of two positions for double-strand DNA (orange and red spheres) after docking on the MC1 static molecular shape (green-blue, cartoon or spheres) by interactive driving of the flexible ligand through electrostatic potential of the protein, in accordance with the experimental data. (B) Ten final models for the DNA-MC1 complex (silver cartoon and colored cartoon respectively) proposed after all-atom reconstruction. Average DNA axis curvature measured by the Curves program on these models is $109 \pm 6^\circ$. (C) Superimposition of 2 of the 8 models of DNA-MC1 complex resulting from HADDOCK. The DNA angle curvature seems to be correlated to the position of the flexible protein LP5 loop: 122° for model 1 (blue) and 81° for model 8 (pink).
doi:10.1371/journal.pone.0088809.g003

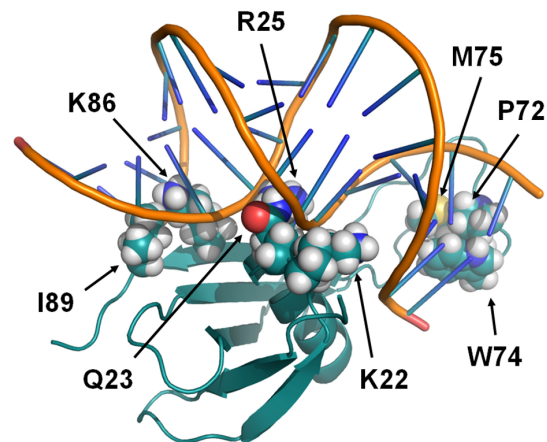


Figure 4. Model of MC1 interaction with the bent 15 bp DNA. The side chains of K22, P72, W74, M75, K86 and I89 contact the minor groove of DNA and that of R25, with the possible help of Q23, neutralizes the negative phosphates in the shrunk major groove.
doi:10.1371/journal.pone.0088809.g004

affinity). Finally, the third point of DNA-MC1 contact was validated by substitutions of Gln23 and Arg25 located in the α -helix. The replacement of Gln23 by a Glu distinctly decreased the binding (7.5-fold affinity), and the effect was larger still with Arg25, whose replacement by Ala or Gln greatly affected the binding (>100 -fold affinity). According to the data in the literature, the loss of an electrostatic interaction has a greater effect on DNA binding than the loss of a hydrophobic bond [35]. Mutation of Arg25 strongly affected the binding, which further argues that Arg25 is involved in an electrostatic contact with the DNA backbone.

EMSA also provided evidence for DNA bending. We previously observed an expected behavior with short DNA (<30 bp): a complex in which DNA is not or is only slightly bent migrates less rapidly than a complex in which DNA is tightly bent [27,36]. Clearly, complexes with the mutants Arg25Ala/Gln, Pro72Ala, Trp74Phe and Ile89Ala migrate less rapidly than complexes with the WT protein (Figure 5); these four residues are therefore likely to be involved in DNA bending. On the contrary, Arg71Ala forms a complex that migrates slightly faster than WT MC1 and the Gln23Ala protein has no visible effect on the mobility of the complex. The trend in the mutant capacity to bend DNA is: Arg71Ala $>$ WT \approx Gln23Ala $>$ Ile89Ala $>$ Pro72Ala $>$ Trp74Phe \approx Arg25Ala/Gln.

Discussion

Our experimental results converge to a protein/DNA model, in which the monomeric protein MC1 interacts on the concave side of a strongly bent DNA: 1) MC1 bears structural similarities to the small basic architectural proteins Sul7d and Cren7, belonging to the *Sulfolobus* strains of the *Crenarchaea* subdomain [37,38], and interacts with the DNA minor groove. However, the later proteins bind on the convex side of the DNA curvature (Figure 6); and 2) Protein interactions with the concave side of DNA curvature have only been observed so far for dimeric proteins, such as histone-like

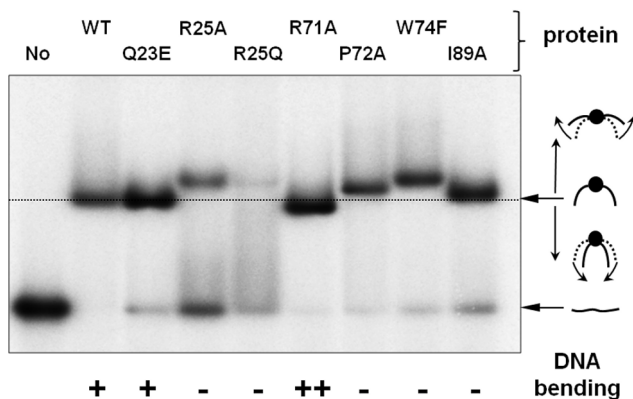


Figure 5. DNA bending ability of the WT and mutant MC1 proteins evaluated by EMSA. EMSA experiment was performed as described for K_{Dapp} measurement: 0.1 nM of 5'-[32 P]-labeled 26 bp DNA were incubated with the WT or mutant MC1 protein (at 20 nM final concentration, excepting R25A and R25Q mutant versions for which 100 nM was used). At equilibrium, assays were analyzed by EMSA as described in the *Materials & Methods* section. After 3 hours of electrophoresis, the gel was dried and visualized by autoradiography. The relative electrophoretic mobility of the protein/DNA complex provides an evaluation of the bending ability of the MC1 version considered. With this short DNA duplex, an apparent greater mobility of the nucleoprotein complex is expected for a protein with a greater DNA bending.

doi:10.1371/journal.pone.0088809.g005

HU or IHF, bound to U-shape DNA [39] (Figure 6). DNA-MC1 recognition is probably based on the shape readout such as minor groove narrowing, kink and bending [40]. Slight differences in minor groove shape, leading to slight differences in electrostatic potential may allow a fine grained recognition [41].

Specific minor groove interactions

Minor groove narrowing in the A-tract. A-tracts are known to rigidify the double-strand helix of DNA [42] and have a bending propensity toward their narrow minor grooves [43]. The binding of arginine residues to the narrow minor groove is a widely used mode for DNA-protein recognition [44,45]. The narrow minor groove of MC1 induced by the A1-A5 tract is recognized by the basic residues Lys85, Lys86 and Arg88. In their proximity, the great flexibility of the bulge (Val57, Glu87, Arg88) [9] is probably essential for the specific positioning of the Ile89 side-chain in the A-tract minor groove. Site-directed mutagenesis of MC1 coupled to binding experiments (Figure 5 and Table 1) showed that the side-chain of Ile89 is important for binding and contributes to the curvature of DNA. Some architectural proteins are known to use additional intercalating residues in the AT rich minor groove [15] such as the Phe-Met dipeptide for SOX17 protein [46] and Pro for IHF (HU) dimer [47].

Flexible CpA steps. Among the 10 possible dinucleotides, CpA, TpA and TpG are the most flexible steps, as they are weakly stabilized through base stacking interactions [40]. The MC1 sequence contains three consecutive CpA steps, giving the 15 bp sequence great flexibility. Moreover, the cytosine C6 is adjacent to the A-tract and, with this specific sequence, it was reported that bending and kinking were able to enhance each other [48]. This great flexibility is certainly responsible for the dramatic bending induced by MC1. Hence, the presence of the positive side-chain of Arg25 is essential to neutralize the repulsive negative charges of the phosphates belonging to the closely spaced A5pC6 and T20pG21 steps.

Minor groove widening in the CCCA sequence. The hydrophobic side-chains of Pro72, Trp74 and Met75 residues located in the mobile arm of MC1 are well positioned to bind the widened minor groove composed of the CCC sequence with possible intercalation in the terminal CpA step. Differences in the hydrophobicity of the mutated proteins could explain the observed differences in DNA bending (Supplementary Figure S10). Clearly, the mutation of Arg71 into Ala enhances the hydrophobicity of the LP5 arm, which may explain the strong DNA curvature observed for the Arg71Ala mutant. Indeed, hydrophobic contacts with bases are used by many architectural proteins such as TBP, SRY and LEF-1 that only contact the minor groove. Recently [49], a structural study of the CCAAT-binding complex revealed a very shallow and widened minor groove around the CCA steps with a DNA kink stabilized by the intercalation of a phenylalanine side-chain at the CpA step.

DNA-recognition by the Euryarchaeal MC1 family

Three regions of MC1 are essential to the DNA-recognition: the first is composed of the basic amino acids Lys85, Lys86, Arg88 and the isoleucine Ile89 located on the triple-stranded β -sheet; the second region involves the hydrophobic residues Pro72, Trp74 and Met75 in the arm; and to connect these two regions the Arg25 residue located on the first turn of the α -helix. These residues, except for Lys85, Arg88 and Ile89, are conserved among the other MC1 proteins from *Methanomicrobia* and *Halobacteria*, which are two classes of *Euryarchaea* (Figure 7). Arg88 is often replaced by Gly or Lys, which are other residues capable of contacting the narrow minor groove [50,51], and Ile89 is sometimes substituted with Val,

Table 1. Apparent dissociation constant (K_{Dapp}) for the WT and mutant MC1 proteins obtained from EMSA experiments.

MC1	WT	Q23E	R25A	R25Q	R71A	P72A	W74F	I89A
K_{Dapp} (nM)	1.2±0.1	9.05±2	161±14	552±34	3.6±0.2	4.2±0.5	8.4±1	15.4±2
DNA Bending	+	+	-	-	++	-	-	-

Binding curves (Supplementary Figure S3) obtained from EMSA experiments were fitted with the equation: $Y = [MC1] / ([MC1] + K_{Dapp})$, where Y is the fraction of the bound 26 bp DNA, and [MC1] is the free protein concentration approximated by the total protein concentration in our experimental conditions. Each K_{Dapp} value is the mean ± standard deviation of three independent experiments.

doi:10.1371/journal.pone.0088809.t001

positioning a Γ -methyl group in the same way. An important point to note is the complete conservation of the residues Pro24 to Lys30, belonging to the α -helix, which seems to ensure the correct positioning of the Arg25 side-chain. In our opinion, the proposed model built for *Methanosarcina thermophila* MC1 is representative of MC1 from *Methanomicrobia* and *Halobacteria* species and more generally from the *Euryarchaeal* subdomain (Figure 7).

Bending proteins, used by all organisms across the three domains of life to compact genomic DNA, belong to distinct families of small, basic and abundant chromatin proteins (Figure 6). Many 3D structures of DNA complexes with bending proteins are deposited in the Protein Data Bank and most of them were obtained by crystallography: In *Bacteria*, 5 structures are known with HU [52-54], 11 with IHF [52-57], and 12 with Fis [58]. The dimeric proteins contact the concave side of DNA through three recognition sites in the minor groove. In *Eukaryota*, many tridimensional structures of HMG box-DNA complexes have been determined [46,59]. The monomeric HMG proteins wrap around the DNA by minor groove recognition. In *Archaea*, only tridimensional structures of DNA-protein complexes belonging to the *Crenarchaea* subdomain are known: 6 structures with Sul7d (Sso7d/Sac7d) [60-62], and 3 structures with Cren7 [38,63]. These monomeric archaeal proteins, like HMG proteins, contact

the convex side of the DNA curvature. To our knowledge, the known solution structures are obtained only for proteins wrapping around DNA, such as HMG [64-67] and Sul7d [60-62], with many intermolecular NOEs between the protein and DNA. For the DNA-MC1 complex, the recognition type is completely different, in that the monomeric protein contacts the concave side of the DNA curvature.

In summary, the DNA-MC1 binding site is composed of two areas of contact in the minor groove. First, the narrow minor groove induced by the A1-A5 tract is recognized by the basic residues Lys85, Lys86 and Arg88 of MC1. In their proximity, the great flexibility of the bulge (Val57, Glu87, Arg88) seems to be important for the specific positioning of the Ile89 side-chain in the A-tract minor groove. Second, the hydrophobic side-chains of Pro72, Trp74 and Met75 residues located in the mobile arm of MC1 are well positioned to bind the minor groove composed of the CCCA sequence, with possible intercalation in the terminal CpA step. Then, three central and flexible CpA steps adjacent to the A-tract are responsible for the dramatic bending of DNA. The presence of the positive side-chain of Arg25 is essential to neutralize the repulsive negative charges of the phosphates in the major groove. To our knowledge, this is the first description of the interaction of a monomeric protein that takes place on the

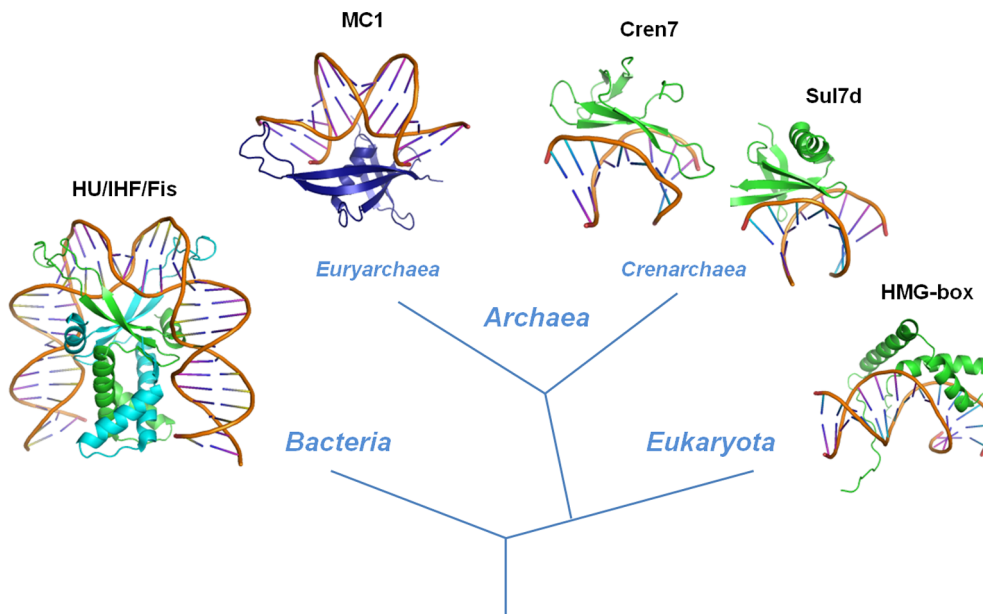


Figure 6. Representative 3D-structures of complexes with DNA-bending proteins among the three domains of life. The dimeric bacterial protein IHF (PDB 1IHF) and the monomeric *Euryarchaeal* protein MC1 contact the concave side of the DNA curvature. In *Eukaryota* and *Crenarchaea* subdomain the proteins SRY (HMG-box protein) (PDB 1J46), Sul7d (PDB 1AZP) and Cren7 (PDB 3KXT) contact the convex side of the DNA curvature.

doi:10.1371/journal.pone.0088809.g006

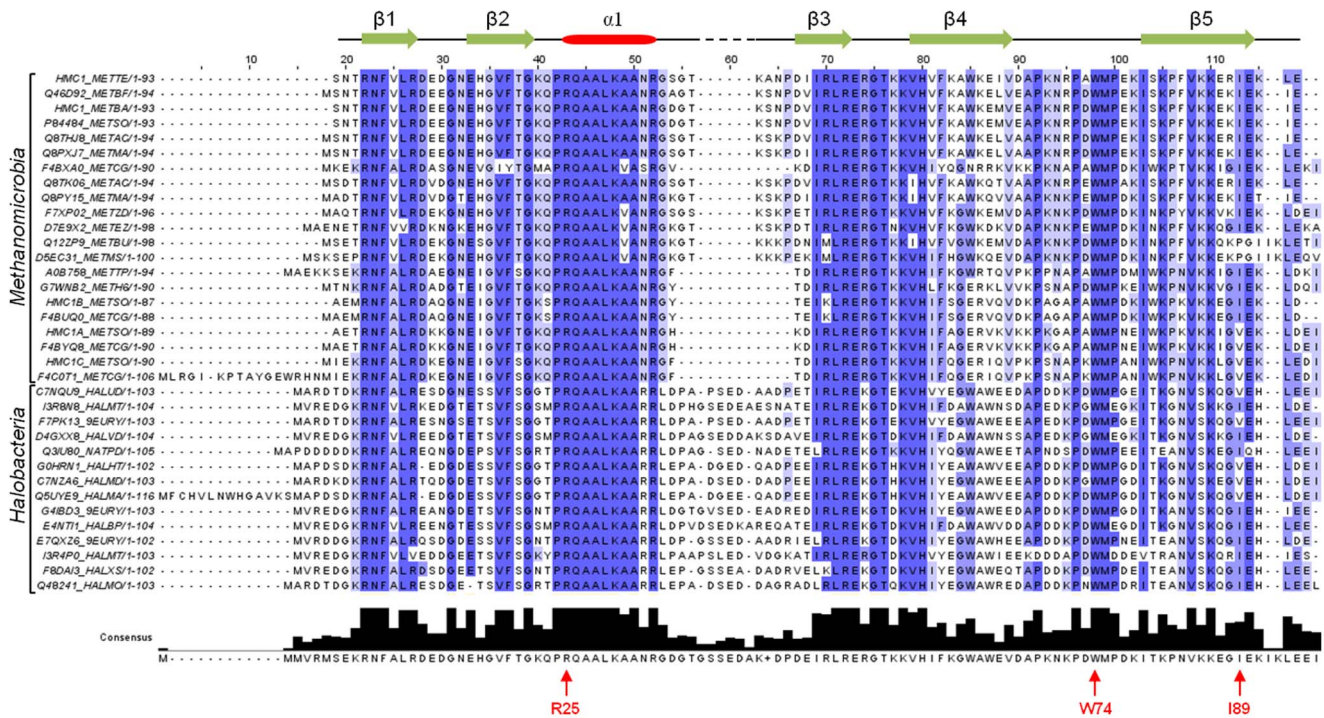


Figure 7. Sequence alignment of MC1 from *Methanosarcina thermophila* (HMC1_METTE) with other MC1 from the *Methanomicrobia* and *Halobacteria* classes. Secondary-structure elements of MC1(HMC1_METTE) are given above the sequences. Key residues indicated under the consensus sequence respect the numbering of the MC1 (HMC1_METTE) sequence.
doi:10.1371/journal.pone.0088809.g007

minor groove and on the concave side of the DNA curvature. This atypical DNA-MC1 structural model adds to our knowledge on DNA condensation observed across the three domains of life.

Supporting Information

Figure S1 Protocol used for the modeling of DNA-MC1 complexes. The three successive steps used are from left to right: i) docking using BIOSPRING which gives coarse-grained models of the complex, ii) high-resolution model reconstruction of DNA with the specified sequence, and iii) high-resolution docking using HADDOCK which gives all-atom models of the complexes. (TIF)

Figure S2 Ambiguous Interaction Restraints used in HADDOCK runs. The side-chain protons of Lys86, Arg88 and Ile89 are constrained to approach at least one proton belonging to A₂A₃A₄A₅C₆:G₂₅T₂₆T₂₇T₂₈T₂₉ with a distance of 5 ± 1 Å (green rectangle). The side chain protons of Pro72, Trp74 and Met75 are constrained to approach at least one proton belonging to C₁₃C₁₄A₁₅:T₁₆G₁₇G₁₈ with a distance of 5 ± 1 Å (red rectangle). (TIF)

Figure S3 SDS-PAGE analysis of purified proteins. 2 µg of each protein (MC1 WT and mutants as indicated) were loaded on an EZ-run™ gel (Fischer Scientific). Staining with Instant-Blue™ Coomassie followed by quantification using ImageQuant software showed purity higher than 95%. (TIF)

Figure S4 Comparative DNA binding properties of the WT and mutant MC1 proteins. 5'-[³²P]-26 bp DNA duplex (0.1 nM) containing the MC1 15 bp high affinity sequence was incubated with increased concentrations of WT and mutants

versions of MC1 (0.05–10 nM protein range). At equilibrium, assays were analyzed by EMSA under conditions defined in the *Materials & Methods* section. (A) Example of a gel autoradiography obtained with the WT MC1 protein. (B) and (C) After autoradiography, the bands corresponding to the free and bound DNA probe were quantified and each point of titration experiments represents the mean value obtained for three independent experiments. Apparent dissociation constants (K_D) were extracted from these curves by fitting to a single binding site using the equation: $Y = [MC1] / ([MC1] + K_D)$ and reported in table 1. (TIF)

Figure S5 Analysis of MC1 CSP upon DNA-binding. (A) Superimposition of the ¹⁵N HSQC spectra of the free protein (blue) and the bound protein (red); H^N and N chemical shifts as a function of free MC1 chemical shifts. A line was added to highlight the linear trend. (B) CSP values as a function of residue number. (C) Analysis of CSP with SAMPLEX: the MC1 protein backbone is shown as a cartoon; colored regions correspond to residues showing no (grey), medium (orange) or strong (red) chemical shift variations upon complexation of oligonucleotide to protein as recorded in NMR experiments; the five loops and key residues are labeled and the MC1 sequence with CSP is reported. Residues are located in five sites: Pro24-Arg34 (α-Helix), Ile45-Leu47 (β3-strand), Phe58-Glu63 (β4-strand), Pro72-Pro76 (loop LP5) and Val 84-Glu90 (β5-strand). (TIF)

Figure S6 CSP analysis of DNA resonances upon MC1-binding. (A) CSP of the exchangeable protons upon MC1-binding: the bound amino protons (NHb), the free amino protons (NHf) and the imino protons (NH) are indicators of the Watson-Crick base pairing. (B) CSP of the nonexchangeable protons after

MC1-binding. The anomeric H1' protons of deoxyribose and the aromatic H2 protons of adenine are located in the DNA minor groove. The H6 protons of cytosine/thymine, the H8 protons of adenine/guanine, the H5 protons of cytosine and the methyl group of thymine are located in the DNA major groove. (TIF)

Figure S7 Assignment of the 15 bp DNA protons after binding to MC1. The NOESY spectrum was recorded on the complex (1 mM) in 10 mM phosphate buffer pH 6, 100 mM NaCl, 1 mM EDTA, 10% D₂O at 26°C with a mixing time of 120 ms on a 600 MHz spectrometer. (TIF)

Figure S8 Intermolecular PREs for the MC1/DNA complex. Intermolecular PRE ¹H_N-Γ₂ profiles obtained for each EDTA-Mn²⁺ DNA. (TIF)

Figure S9 DNA groove geometry in the complex. The 8 models of 15 bp DNA bound to MC1 were analyzed using Curves+: (A) Schematic view of the bound DNA in the model 1 of the complex; the global helical axis is represented in blue and the backbone spines in red. (B) The DNA axis total bend, calculated for each model by Curves+, corresponds to the cumulative axis

References

- Luijsterburg MS, White MF, van Driel R, Dame RT (2008) The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Crit Rev Biochem Mol Biol* 43: 393–418.
- Reeve JN (2003) Archaeal chromatin and transcription. *Mol Microbiol* 48: 587–598.
- White MF, Bell SD (2002) Holding it together: chromatin in the Archaea. *Trends Genet* 18: 621–626.
- Sandman K, Reeve JN (2005) Archaeal chromatin proteins: different structures but common function? *Curr Opin Microbiol* 8: 656–661.
- DeLange RJ, Green GR, Searcy DG (1981) A histone-like protein (HTa) from *Thermoplasma acidophilum*. I. Purification and properties. *J Biol Chem* 256: 900–904.
- DeLange RJ, Williams LC, Searcy DG (1981) A histone-like protein (HTa) from *Thermoplasma acidophilum*. II. Complete amino acid sequence. *J Biol Chem* 256: 905–911.
- Chartier F, Laine B, Sautiere P (1988) Characterization of the chromosomal protein MC1 from the thermophilic archaeobacterium *Methanosarcina* sp. CHTI 55 and its effect on the thermal stability of DNA. *Biochim Biophys Acta* 951: 149–156.
- Paquet F, Culard F, Barbault F, Maurizot JC, Lancelot G (2004) NMR solution structure of the archaeobacterial chromosomal protein MC1 reveals a new protein fold. *Biochemistry* 43: 14971–14978.
- Paquet F, Loth K, Meudal H, Culard F, Genest D, et al. (2010) Refined solution structure and backbone dynamics of the archaeal MC1 protein. *FEBS J* 277: 5133–5145.
- Paradinas C, Gervais A, Maurizot JC, Culard F (1998) Structure-specific binding recognition of a methanogen chromosomal protein. *Eur J Biochem* 257: 372–379.
- Teyssier C, Toulme F, Touzel JP, Gervais A, Maurizot JC, et al. (1996) Preferential binding of the archaeobacterial histone-like MC1 protein to negatively supercoiled DNA minicircles. *Biochemistry* 35: 7954–7958.
- De Vuyst G, Aci S, Genest D, Culard F (2005) Atypical recognition of particular DNA sequences by the archaeal chromosomal MC1 protein. *Biochemistry* 44: 10369–10377.
- Le Cam E, Culard F, Larquet E, Delain E, Cognet JA (1999) DNA bending induced by the archaeobacterial histone-like protein MC1. *J Mol Biol* 285: 1011–1021.
- Teyssier C, Laine B, Gervais A, Maurizot JC, Culard F (1994) Archaeobacterial histone-like protein MC1 can exhibit a sequence-specific binding to DNA. *Biochem J* 303 (Pt 2): 567–573.
- Bewley CA, Gronenborn AM, Clore GM (1998) Minor groove-binding architectural proteins: structure, function, and DNA recognition. *Annu Rev Biophys Biomol Struct* 27: 105–131.
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6: 277–293.
- Johnson BA, Blevins RA (1994) NMR View: a computer program for the visualization and analysis of NMR data. *J Biomol NMR* 4: 603–614.
- Krzeminski M, Loth K, Boelens R, Bonvin AM (2010) SAMPLEX: automatic mapping of perturbed and unperturbed regions of proteins and complexes. *BMC Bioinformatics* 11: 51.
- Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98: 10037–10041.
- Pyrkov TV, Chugunov AO, Krylov NA, Nolde DE, Efremov RG (2009) PLATINUM: a web tool for analysis of hydrophobic/hydrophilic organization of biomolecular complexes. *Bioinformatics* 25: 1201–1202.
- Blanchet C, Pasi M, Zakrzewska K, Lavery R (2011) CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res* 39: W68–73.
- Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14: 33–38, 27–38.
- De Lano WL (2002) Pymol. South San Francisco, CA: De Lano Scientific.
- Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* 14: 51–55, 29–32.
- Saladin A, Amourda C, Poulain P, Ferey N, Baaden M, et al. (2010) Modeling the early stage of DNA sequence recognition within RecA nucleoprotein filaments. *Nucleic Acids Res* 38: 6313–6323.
- Flick KE, Jurica MS, Monnat RJ, Jr., Stoddard BL (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature* 394: 96–101.
- Bure C, Goffinont S, Delmas AF, Cadene M, Culard F (2008) Oxidation-sensitive residues mediate the DNA bending abilities of the architectural MC1 protein. *J Mol Biol* 376: 120–130.
- van Dijk M, Bonvin AM (2009) 3D-DART: a DNA structure modelling server. *Nucleic Acids Res* 37: W235–239.
- van Dijk M, Bonvin AM (2010) Pushing the limits of what is achievable in protein-DNA docking: benchmarking HADDOCK's performance. *Nucleic Acids Res* 38: 5634–5647.
- de Vries SJ, van Dijk M, Bonvin AM (2010) The HADDOCK web server for data-driven biomolecular docking. *Nat Protoc* 5: 883–897.
- Schumann FH, Riepl H, Maurer T, Gronwald W, Neidig KP, et al. (2007) Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. *J Biomol NMR* 39: 275–289.
- Zuiderweg ER (2002) Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* 41: 1–7.
- Stratmann D, Boelens R, Bonvin AM (2011) Quantitative use of chemical shifts for the modeling of protein complexes. *Proteins* 79: 2662–2670.
- Clore GM, Iwahara J (2009) Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev* 109: 4108–4139.
- Klass J, Murphy FVt, Fouts S, Serenil M, Changela A, et al. (2003) The role of intercalating residues in chromosomal high-mobility-group protein DNA binding, bending and specificity. *Nucleic Acids Res* 31: 2852–2864.

36. Culard F, Gervais A, de Vuyst G, Spothem-Maurizot M, Charlier M (2003) Response of a DNA-binding protein to radiation-induced oxidative stress. *J Mol Biol* 328: 1185–1195.
37. Chen CY, Ko TP, Lin TW, Chou CC, Chen CJ, et al. (2005) Probing the DNA kink structure induced by the hyperthermophilic chromosomal protein Sac7d. *Nucleic Acids Res* 33: 430–438.
38. Zhang Z, Gong Y, Guo L, Jiang T, Huang L (2010) Structural insights into the interaction of the crenarchaeal chromatin protein Cren7 with DNA. *Mol Microbiol* 76: 749–759.
39. Swinger KK, Rice PA (2004) IHF and HU: flexible architects of bent DNA. *Curr Opin Struct Biol* 14: 28–35.
40. Rohs R, Jin X, West SM, Joshi R, Honig B, et al. (2010) Origins of specificity in protein-DNA recognition. *Annu Rev Biochem* 79: 233–269.
41. Joshi R, Passner JM, Rohs R, Jain R, Sosinsky A, et al. (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* 131: 530–543.
42. El Hassan MA, Calladine CR (1997) Conformational characteristics of DNA: empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps. *Phil Trans R Soc Lond A* 355: 43–100.
43. Ababneh AM, Ababneh ZQ, Large CC (2008) DNA A-tracts bending: polarization effects on electrostatic interactions across their minor groove. *J Theor Biol* 252: 742–749.
44. Cordeiro TN, Schmidt H, Madrid C, Juarez A, Bernado P, et al. (2011) Indirect DNA readout by an H-NS related protein: structure of the DNA complex of the C-terminal domain of Ler. *PLoS Pathog* 7: e1002380.
45. Rohs R, West SM, Sosinsky A, Liu P, Mann RS, et al. (2009) The role of DNA shape in protein-DNA recognition. *Nature* 461: 1248–1253.
46. Palasingam P, Jauch R, Ng CK, Kolatkar PR (2009) The structure of Sox17 bound to DNA reveals a conserved bending topology but selective protein interaction platforms. *J Mol Biol* 388: 619–630.
47. Ellenberger T, Landy A (1997) A good turn for DNA: the structure of integration host factor bound to DNA. *Structure* 5: 153–157.
48. Tolstorukov MY, Colasanti AV, McCandlish DM, Olson WK, Zhurkin VB (2007) A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning. *J Mol Biol* 371: 725–738.
49. Huber EM, Scharf DH, Hortschansky P, Groll M, Brakhage AA (2012) DNA Minor Groove Sensing and Widening by the CCAAT-Binding Complex. *Structure* 20: 1757–1768.
50. Escalante CR, Nistal-Villan E, Shen L, Garcia-Sastre A, Aggarwal AK (2007) Structure of IRF-3 bound to the PRDIII-I regulatory element of the human interferon-beta enhancer. *Mol Cell* 26: 703–716.
51. Panne D, Maniatis T, Harrison SC (2007) An atomic model of the interferon-beta enhancosome. *Cell* 129: 1111–1123.
52. Swinger KK, Lemberg KM, Zhang Y, Rice PA (2003) Flexible DNA bending in HU-DNA cocrystal structures. *EMBO J* 22: 3749–3760.
53. Swinger KK, Rice PA (2007) Structure-based analysis of HU-DNA binding. *J Mol Biol* 365: 1005–1016.
54. Mouw KW, Rice PA (2007) Shaping the *Borrelia burgdorferi* genome: crystal structure and binding properties of the DNA-bending protein Hbb. *Mol Microbiol* 63: 1319–1330.
55. Rice PA, Yang S, Mizuuchi K, Nash HA (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* 87: 1295–1306.
56. Bao Q, Chen H, Liu Y, Yan J, Droge P, et al. (2007) A divalent metal-mediated switch controlling protein-induced DNA bending. *J Mol Biol* 367: 731–740.
57. Lynch TW, Read EK, Mattis AN, Gardner JF, Rice PA (2003) Integration host factor: putting a twist on protein-DNA recognition. *J Mol Biol* 330: 493–502.
58. Stella S, Cascio D, Johnson RC (2010) The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. *Genes Dev* 24: 814–826.
59. Rubio-Cosials A, Sidow JF, Jimenez-Menendez N, Fernandez-Millan P, Montoya J, et al. (2011) Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter. *Nat Struct Mol Biol* 18: 1281–1289.
60. Su S, Gao YG, Robinson H, Liaw YC, Edmondson SP, et al. (2000) Crystal structures of the chromosomal proteins Sso7d/Sac7d bound to DNA containing T-G mismatched base-pairs. *J Mol Biol* 303: 395–403.
61. Gao YG, Su SY, Robinson H, Padmanabhan S, Lim L, et al. (1998) The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA. *Nat Struct Mol Biol* 5: 782–786.
62. Agback P, Baumann H, Knapp S, Ladenstein R, Hard T (1998) Architecture of nonspecific protein-DNA interactions in the Sso7d-DNA complex. *Nat Struct Mol Biol* 5: 579–584.
63. Feng Y, Yao H, Wang J (2010) Crystal structure of the crenarchaeal conserved chromatin protein Cren7 and double-stranded DNA complex. *Protein Sci* 19: 1253–1257.
64. Williams DC, Jr., Cai M, Clore GM (2004) Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-kDa Oct1-Sox2-Hoxb1-DNA ternary transcription factor complex. *J Biol Chem* 279: 1449–1457.
65. Masse JE, Wong B, Yen YM, Allain FH, Johnson RC, et al. (2002) The *S. cerevisiae* architectural HMGB protein NHP6A complexed with DNA: DNA and protein conformational changes upon binding. *J Mol Biol* 323: 263–284.
66. Murphy EC, Zhurkin VB, Louis JM, Cornilescu G, Clore GM (2001) Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occurring point mutation. *J Mol Biol* 312: 481–499.
67. Stott K, Tang GS, Lee KB, Thomas JO (2006) Structure of a complex of tandem HMG boxes and DNA. *J Mol Biol* 360: 90–104.