NMR Solution Structure of the Archaebacterial Chromosomal Protein MC1 Reveals a New Protein Fold‡

Françoise Paquet,*,§ Françoise Culard,§ Florent Barbault,|| Jean-Claude Maurizot,§ and Gérard Lancelot§

Centre de Biophysique Moléculaire, CNRS, Rue Charles Sadron, 45071 Orléans Cedex 2, France, and Laboratoire ITODYS, Université Paris 7, 1 rue Guy de la Brosse, 75005 Paris, France

Received July 29, 2004; Revised Manuscript Received September 17, 2004

ABSTRACT: The three-dimensional structure of methanogen chromosomal protein 1 (MC1), a chromosomal protein extracted from the archaebacterium *Methanosarcina* sp. CHTI55, has been solved using ^{1}H NMR spectroscopy. The small basic protein MC1 contains 93 amino acids (24 basic residues against 12 acidic residues). The main elements of secondary structures are an α helix and five β strands, arranged as two antiparallel β sheets (a double one and a triple one) packed in an orthogonal manner forming a barrel. The protein displays a largely hydrophilic surface and a very compact hydrophobic core made up by side chains at the interface of the two β sheets and the helix side facing the interior of the protein. The MC1 solution structure shows a globular protein with overall dimensions in the range of 34−40 Å, which potentially corresponds to a DNA-binding site of 10−12 base pairs. The presumed DNA-binding site is located on the sequence comprising residues K62−P82, which is formed by a part of strands II2 and II3 belonging to the triple-stranded antiparallel β sheet and a loop flanked by prolines P68 and P76. The tryptophan W74 that is expected to play a key role in the DNA-binding according to photocross-linking experiments was found completely exposed to the solvent, in a good position to interact with DNA. The overall fold of MC1, characterized by its linking $\beta-\beta-\alpha-\beta-\beta-$loop$-\beta$, is different from other known DNA-binding proteins. Its structure suggests a different DNA-binding mode than those of the histone-like proteins HU or HMGB. Thus, MC1 may be classified as a member of a new family.

Methanogen chromosomal protein 1 (MC1)[1] is the major chromosomal protein of various species of Methanosarcinaceae, a family of methanogenic archaea. This protein is naturally present in cells at the ratio of 1 protein per 100 bp. It has been shown to be associated to DNA *in vivo* (*1*), and its physiological role is probably to participate in the structuring of the archaeal chromosome. The protein MC1 studied here has been isolated from *Methanosarcina* sp. CHTI55 and sequenced (*2*−*4*). It is a 93 amino acid residue basic protein with a net charge of +12 at neutral pH (24 basic residues: 9 Arg, 13 Lys, and 2 His against 12 acidic residues: 8 Glu and 4 Asp) and is devoid of cysteines and tyrosines. Its capacity to bend DNA was characterized quantitatively (*5*). Similarly to the histone-like protein HU, MC1 is able to bind, compact, and strongly bend DNA (*6*−*8*). In contrast to HU, MC1 has a high affinity for linear DNA. Moreover and like HMGB proteins, it exhibits and enhances affinities for circular and cruciform structures (*9*). Electron microscopy and atomic force spectroscopy observations showed that this histone-like protein induced sharp bending of DNA (*7*, *8*) and supercoiling (*6*). Although many DNA-interaction studies were devoted to MC1 (*5*−*8*, *10*−*12*), little is known about its three-dimensional structure. Circular dichroism and infrared studies indicated a low amount of α-helix content and the presence of antiparallel β strands (*1*). We report, here, the three-dimensional solution structure of the histone-like MC1 in high-salt conditions.

MATERIALS AND METHODS

Nuclear Magnetic Resonance (NMR) Sample Preparation. Natural MC1 was purified in our laboratory from *Methanosarcina* sp. CHTI55 as previously described (*6*). NMR samples were prepared in 90% H_2O/10% D_2O. MC1 did not bear lyophilisation, and the NMR sample in 100% D_2O is directly obtained by elution with the deutered buffer during its HPLC purification.

NMR Data Collection, Methods, and Analysis. At low salt concentration, the ^{1}H NMR of MC1 showed a low-resolved spectrum with broadened resonance lines, suggesting aggregation processes of the protein. After we experimented with numerous conditions of salt, pH, and temperature, it was concluded that a protein concentration of 1.74 mM in 70 mM sodium acetate buffer at pH 5.3 and 800 mM sodium chloride and a temperature in the range of 300 K resulted in optimum quality NMR spectra.

As the protein was extracted from archaea, only unlabeled material was available. Therefore, exclusively homonuclear

‡ Coordinates for MC1 have been deposited in the Protein Data Bank (PDB ID code 1T23).

* To whom correspondence should be addressed. Telephone: 332 38 25 76 92. Fax: 332 38 25 63 15 17. E-mail: paquet@cnrs-orleans.fr.

§ Centre de Biophysique Moléculaire, UPR 4301 affiliated to the University of Orléans and to INSERM.

|| Laboratoire ITODYS.

[1] Abbreviations: NMR, nuclear magnetic resonance; MC1, methanogen chromosomal protein 1; DQF−COSY, double-quantum-filtered−correlation spectroscopy; TOCSY, total correlation spectroscopy; ECOSY, exclusive COSY; NOESY, nuclear Overhauser enhancement spectroscopy; NOE, nuclear Overhauser effect; RMSD, root-mean-square distance.

10.1021/bi048382z CCC: $27.50
Published on Web 11/04/2004

Table 1: Restraints and Structural Statistics for MC1 Structures

distance restraints		
sequential, $\|i - j\| = 1$	214	
medium range, $1 < \|i - j\| \leq 4$	82	
long range, $\|i - j\| > 4$	177	
hydrogen bonds	37	
total distance restraints	1024	
average by residue	11	
refinement statistics		
for selected 8 conformers		
NOE violations >0.4 Å	0	
pairwise RMSD (Å)	backbone atoms	heavy atoms
global	1.43 ± 0.39	2.32 ± 0.42
secondary structure elements (3−8, 17−22, 25−34, 43−49, 54−66, 78−91)	0.79 ± 0.17	1.75 ± 0.21
β I1 and I2 (3−8, 17−22)	0.45 ± 0.13	1.43 ± 0.46
β II1, II2, and II3 (43−49, 54−66, 78−91)	0.78 ± 0.24	1.77 ± 0.23
β II1 (43−49)	0.63 ± 0.29	1.87 ± 0.43
β II2 (54−66)	0.39 ± 0.10	1.08 ± 0.20
β II3 (78−91)	0.78 ± 0.28	1.92 ± 0.31
α helix (25−34)	0.27 ± 0.10	1.30 ± 0.22
arm (67−77)	0.62 ± 0.31	1.48 ± 0.43
backbone dihedral angles (Ramachandran plot)		
most favored region (%)	58.7	
allowed region	38.6	
generously allowed	1.3	
disallowed	1.3	

spectra were recorded, including double-quantum-filtered−correlation spectroscopy (DQF−COSY), exclusive correlation spectroscopy (ECOSY) for stereospecific assignments, and total correlation spectroscopy (TOCSY) with MLEV17, and an isotropic period of 40 and 70 ms and NOESY with mixing times of 100, 150, and 200 ms were collected on 500 MHz BRUKER AMX and 600 MHz VARIAN Unity INOVA spectrometers equipped with pulse-field z gradient. Standard BRUKER and VARIAN library pulses programs were acquired in the phase-sensitive mode, with quadrature detection in both directions using the States method and WATERGATE water suppression (*13*).

Chemical shifts were referenced to DSS. Spectra were analyzed within the program XEASY (*14*), and assignments of nuclear Overhauser effect (NOE) cross peaks were performed using the automatic assignment program NOAH, which is a part of the DYANA 1.5 package (*15*). Cross peaks in nuclear Overhauser enhancement spectroscopy (NOESY) spectra were recorded in 90% H_2O/10% D_2O with a mixing time of 100 ms. NOE peak volumes were measured in XEASY using the standard integration protocol. Peak distances were calibrated by using the program CALIBA included in DYANA (*15*).

Structure Calculation Protocol. A total of 100 starting structures were generated using random dihedral angles (φ, ψ) and a low target function to produce structures with correct local geometry. These basic structures were subjected to a torsion-angle-simulated annealing protocol within DYANA 1.5 (*15*). A total of 50 structures consistent with the NMR restraints (Table 1) and with target functions <2 Å^2 were subjected to further Cartesian dynamics with the CHARMM force field (*16*) implemented in XPLOR 3.851 (*17*). The upper-limit distance restraints in DYANA were then converted to XPLOR upper-limit restraints. These 50 structures were calculated using a simulated annealing protocol without electrostatic term and a single quartic repulsive van der Waals potential. This protocol started with 15 ps dynamics at high temperature (500 K) and followed by a cooling step of 20 ps dynamics, where the temperature was lowered to 300 K. Then, each structure was minimized by using the Powell algorithm until convergence (generally less than 2000 steps), with a full electrostatic term and attractive van der Waals potential. Finally, 8 structures with low potential energies and showing no experimental distance violation larger than 0.4 Å were considered as representative of the solution structure of MC1 and selected for structural analysis using the software programs MOLMOL 2.6 (*18*), PROMOTIF (*19*), PROCHECK 3.5 (*20*), and PROCHECK−NMR (*21*).

RESULTS AND DISCUSSION

Sequential Assignment and Secondary Structure. The assignment procedure was complicated by the fact that only 50% of the side chains exhibited a good fingerprint on TOCSY spectra. However, COSY spectra in H_2O showed the greatest part of the NH−C_αH connectivities, and the NOESY spectra were of sufficiently good quality for structural analysis (Figure 1 in the Supporting Information). The two adjacent groups of alanine (A27, A28 and A31, A32) were good entry points, and 50% of the residues were assigned by using the standard assignment procedure on COSY, TOCSY, and NOESY spectra (*22*). The other residues were assigned simultaneously by sequential assignment via NOESY spectra (*23*). Finally, the NH and C_αH resonances were assigned for all of the residues. The greatest part of the proton resonances of the side chains were assigned except for Arg4, Arg9, Glu87, and Arg88.

The secondary structure elements of MC1 were identified by the detection of specific pattern of short-, medium-, and long-range NOEs (Figure 1). Thus, residues R25−R34 form an α helix that is well-defined by medium $d_{\alpha,N}(i, i + 3)$ and $d_{\alpha,\beta}(i, i + 3)$ and small $d_{\alpha,N}(i, i + 4)$ NOEs. Two β-sheet structures were well-defined by the network of NOEs (Figure 2). The β strands were organized in a double-stranded antiparallel β sheet formed by residues T3−L8 and G17−K22 (strand I1 and I2) onto which a triple-stranded antiparallel β sheet was packed in an orthogonal manner, involving residues D43−E49, K54−D66, and K78−K91 (strands II1, II2, and II3, respectively). All of these secondary structure elements, except strands II2 and II3, are linked by short segments clustered in loop regions involving residues R9−H16, Q23−P24, G35−P42, and R50−K53. Strands II2 and II3 are connected by a segment (A67−E77) involving 11 residues having NOE connectivities between them and other residues of the protein. MC1 contains 6 prolines (residues P24, P42, P68, P72, P76, and P82) all of which display strong $d(\alpha_i \rightarrow \delta_{i+1})$ NOEs in the D_2O NOESY spectrum ($i + 1$ corresponds to Pro). This confirms that the preceding amide bonds are all in the trans conformation. The specific secondary structure elements are generally in agreement (Figure 1) with those suggested by the α-proton chemical-shift indices (*24*). Thr3 that is involved in the beginning of the β strand I1 presented a negative deviation (−0.23 ppm). We suggest that it is a consequence of a dynamic effect induced by its location near the N-terminal part of the protein, as already observed for the first residue of the motif (*25*).

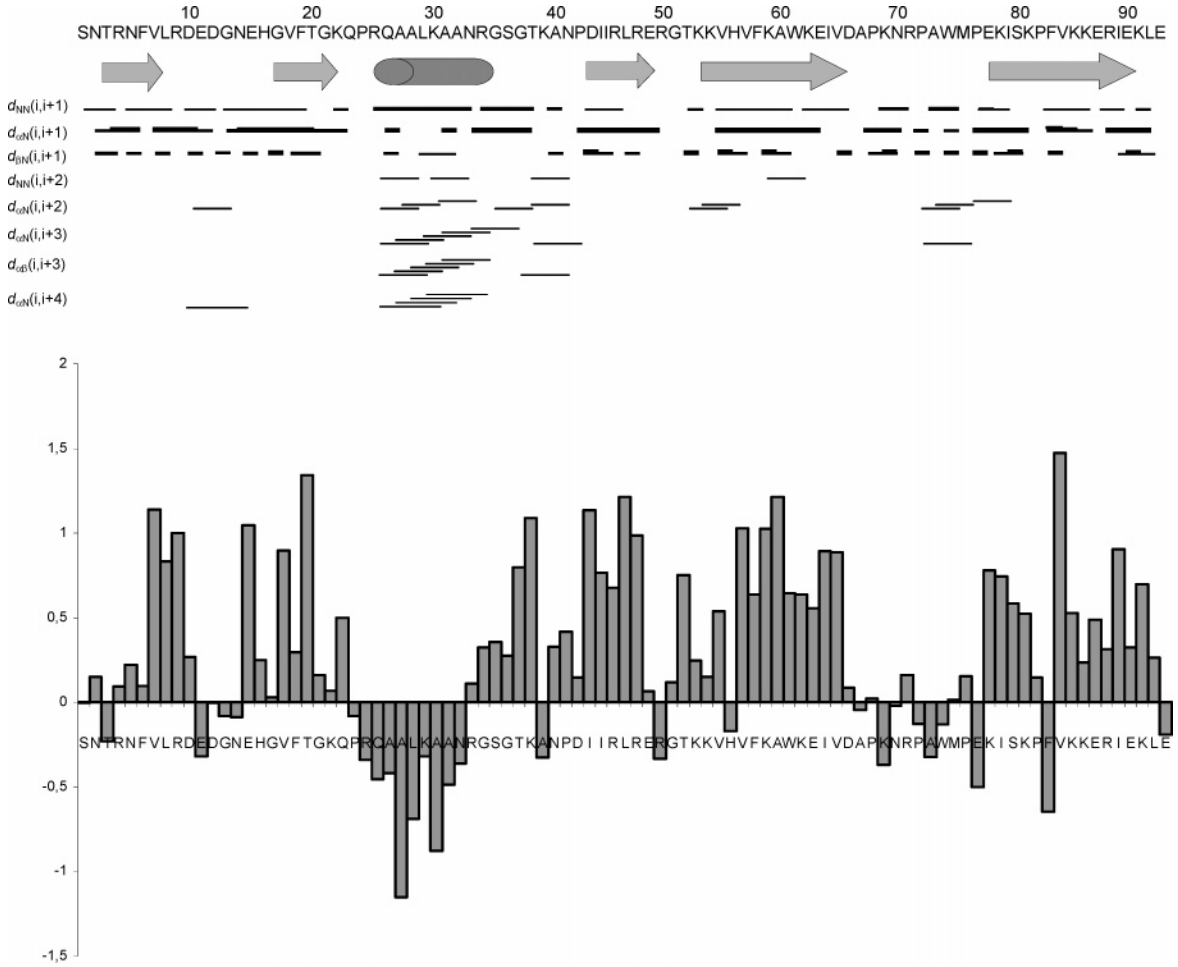

FIGURE 1: (A) Schematic representation of short- and medium-range NOE connectivities involving NH, Hα, and Hβ protons observed for the protein MC1. The thickness of the lines indicates the intensities of the NOEs. (B) ^{1}Hα chemical-shift indices of MC1 computed as the difference between the observed ^{1}Hα chemical shift and the published random-coil value (*24*). The average value of the two glycine Hα is indicated on the scheme.

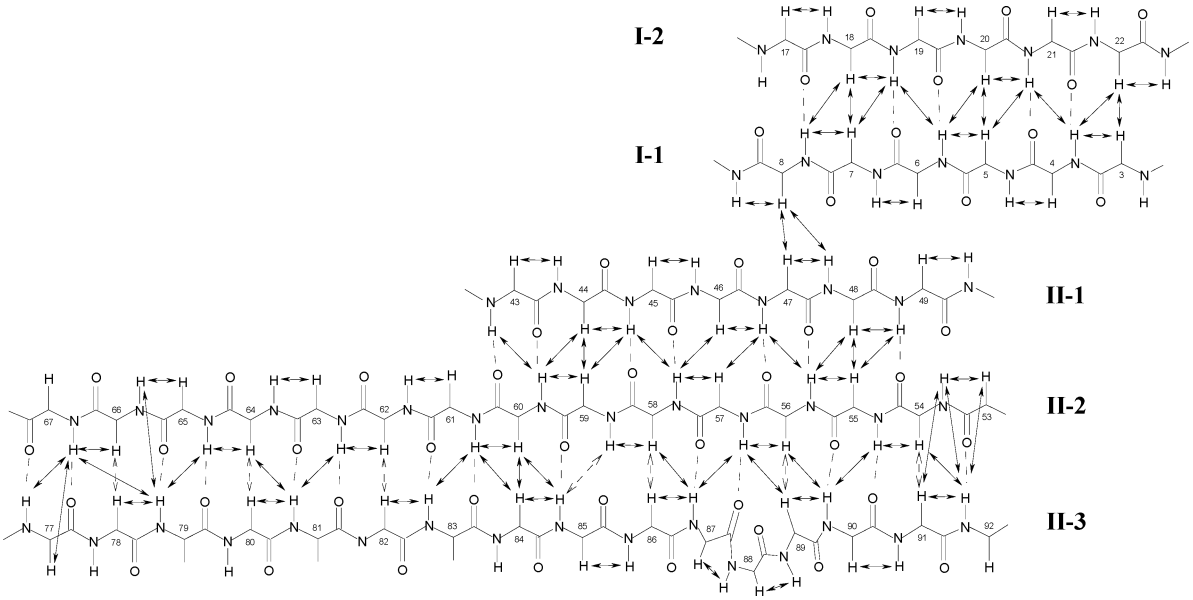

FIGURE 2: Schematic diagram of the five-stranded β barrel of MC1. The antiparallel double-stranded I1 and I2, the antiparallel triple-stranded II1, II2, and II3, and the antiparallel wide bulge (His56, Arg88, and Ile89) are shown. Arrows represent observed NOEs between protons of the sheet. Dashed arrows represent NOEs that could be present but are not included in the structure calculations because of the overlap. Dashed lines correspond to hydrogen bonds subsequently introduced as restraints in the structure calculations.

Tertiary Structure of MC1. The three-dimensional conformation of MC1 was calculated using a combination of NMR data (Table 1), as described in the Materials and Methods. The backbone atoms of all residues from 8 refined

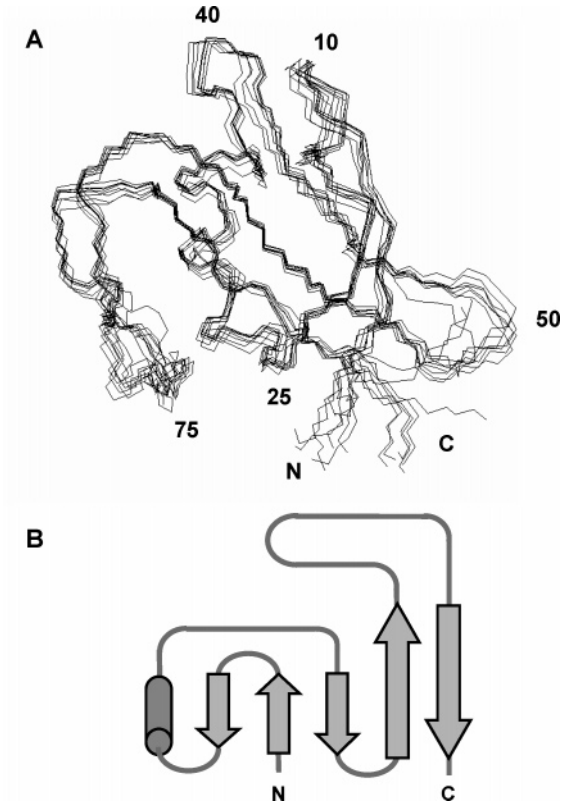

FIGURE 3: (A) Solution structure of the core of MC1. An ensemble of 8 refined structures created by MOLMOL (*18*) are represented by superimposition of the secondary structure elements (4−8, 17−21, 25−34, 43−49, 53−61, and 83−91). N and C termini and selected residues are labeled. (B) Topology diagram of MC1 showing the arrangement of the secondary structure elements, $\beta-\beta-\alpha-\beta-\beta-$loop$-\beta$.

structures were superimposed using MOLMOL (*18*) and are shown in Figure 3. The best fit was determined to be 0.79 ± 0.17 Å when the backbone atoms of the secondary structure elements are superimposed (3−8, 17−22, 25−34, 43−49, 54−66, and 78−91). The pairwise root-mean-square distances (RMSDs) are given in Table 1 for different segments of MC1. It is worth noting that each secondary structure element gives a separately good superimposition, especially β sheet I and α helix. The poor definition of the whole protein with a RMSD of 1.43 ± 0.39 Å is the result of the relative position of these secondary structure elements linked with 6 loops. Analysis of the φ, ψ angles for 8 refined structures of MC1 indicates that most of the backbone torsion angles (97.3%) lie within the allowed regions of the Ramachandran plot. Residues in the disallowed and generously allowed regions (2.6%) are positioned in γ turns (Table 1).

MC1 is folded into a compact globular unit consisting of a double-stranded antiparallel β sheet and a right-handed α helix, onto which an orthogonal triple-stranded antiparallel β sheet is packed. The superimposition of the 8 final structures (Figure 3) showed that the other regions are made up of well-defined turns and of an arm containing 11 residues linking the two last strands (II2 and II3), which were folded into a less-defined loop. This arm is oriented toward the α helix (Figure 5) by the last residues of the β strands, II2 (K54−D66) and II3 (K78−K91). The medium definition of the arm region can result from local flexibility, which is

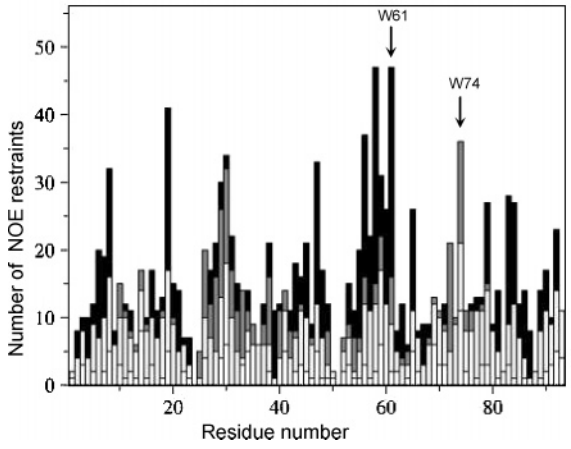

FIGURE 4: Distance restraints data for MC1 generated by DYANA 1.5 (*15*) [long-range distances $|i - j| > 4$ (black), medium-range distances $1 < |i - j| \leq 4$ (dark gray), sequential distances $|i - j| = 1$ (light gray), and intraresidual distances (white)]. Contrary to Trp 74, Trp 61 presented a lot of long-range NOEs, suggesting a quite different exposure to the solvent.

reflected by the small number of NMR restraints for some residues (Figure 4).

An analysis of the eight structures with PROMOTIF (*19*) shows a high percent of typical turns: a β turn of type I′ for residues Asp10−Gly13 (90%), several β turns of type IV for residues Glu23−Glu26, Thr38−Asn41, Lys39−Pro42, Glu49−Thr52, Ala67−Asp70, Pro68−Arg71, Pro72−Met75, and Ala73−Pro76, a β turn of type VIII for residues Gly21−Pro24 (80%), two classical γ turns for residues Thr38−Ala40 and Pro68−Asn70, and one inversed γ turn for residues Glu77−Ile79. In addition, an antiparallel wide bulge made up of residues His56 (β), Arg88 (β), and Ile89 (β) is present in all of the structures. It is worth noting that His56 presented a negative chemical-shift index, whereas it is implied in a β-sheet structure (Figure 1), suggesting a deviation from the orientation of the CO bonds in this peculiar conformation in relation to a classical antiparallel β sheet.

A network of hydrogen bonds between side chains and other chemical groups stabilized (i) the β-sheet structures: 18 hydrogen bonds between the β strands I1 and I2, I1 and II1, II1 and II2, and II2 and II3; (ii) the interaction between the turns: NH(Glu11) and NH(Asn12) with O(Asn41); and (iii) the turn structures: HN(His16) with Oδ(Asp10) and HN and Hγ1(Thr52) with Oϵ(Glu49). We can note that the Thr52 hydroxyl proton presented a lot of connectivities in the NOESY spectra. The Glu49−Arg50−Gly51−Thr52 region forms a well-defined turn that is stabilized by three hydrogen bonds. The carboxyl of Glu49 acts as an acceptor to both the Thr52 amide proton and the Thr52 hydroxyl proton. Further, the hydroxyl group of Thr52 acts as an acceptor to the Lys54 amino proton. This network of hydrogen bonds effectively shields the hydroxyl proton from the solvent and accounts for its visibility in the spectra.

The protein displayed a largely hydrophilic surface with many basic residues and a very compact hydrophobic core made up by side chains at the interface of the two β sheets, in particular, the aromatics Phe6, Phe19, His56, and Phe58, the aliphatic chains of Leu8, Leu47, Ala60, Val84, and Ile89, and the alanines Ala27, Ala28, Ala31, and Ala32 of the α helix (Figures 5 and 6). The helix is polar on one side facing the solution (Arg25, Gln26, Lys30, Asn33, and Arg34) and

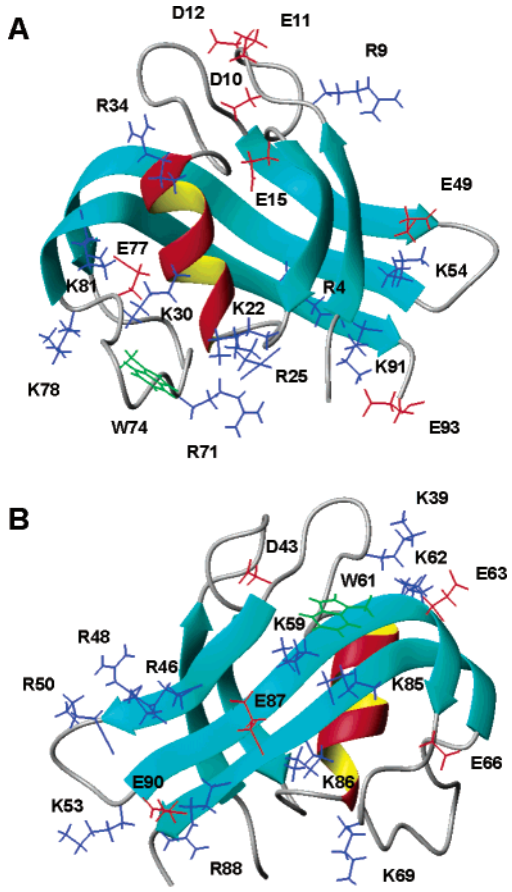

FIGURE 5: Ribbon diagrams of the best solution structure of the protein MC1 created using MOLMOL (*18*). (A) Front view of the protein. (B) Back view of the protein. The positively charged (Lys and Arg, in blue), negatively charged (Glu and Asp, in red) and tryptophan (in green) side chains are represented for each side of the protein. Most of the charged side chains are directed toward the solvent.

hydrophobic on the other side facing the interior of the protein (Ala28, Ala29, Ala31, and Ala32). The tryptophan Trp61 was found in the interaction with the phenylalanine Phe83 as suggested by NOEs between the protons of the two residues and the upfield-shifted resonances of the aromatic protons of Phe83. Both of these residues are partially masked from the solvent (solvent accessibility of 14 and 13% for Trp61 and Phe83, respectively) by their neighboring residues belonging to the loop 36−44 and the β strands 62−65 and 79−82. In contrast, Trp74, which is implied in DNA-binding (*26*), was located in the middle of the arm and was exposed to the solvent, as suggested by its lack of long-range NOEs (Figure 4) and its solvent accessibility (52%).

The loop 10−15 containing Asp10, Glu11, Asp12, and Glu15 and the C-terminal part containing Glu87, Glu90, and Glu93 formed two regions with high negative potential located on the opposite faces of the protein (Figures 5 and 6). These regions can play a key role in the self-association of the protein at low ionic strength. Some basic residues formed a salt bridge with the carboxylic group of the side chains of acid residues, which stabilized the β sheets in most of the structures, Lys54 with Glu49 ($d[O\epsilon - H\zeta] = 1.39 \pm 0.01$ Å) and Lys78 with Asp66 ($d[O\epsilon - H\zeta] = 1.38 \pm 0.01$ Å), or in a few structures, Lys59 with Glu87 ($d[O\epsilon_2 - H\zeta] = 1.45 \pm 0.05$ Å in 25% of the structures) and Arg71 with Gln23 ($d[O\epsilon - H\epsilon] = 1.51 \pm 0.02$ Å in 25% of the structures). The other lysine and arginine side chains were found exposed to the solvent and distributed around the protein. Some of them (Lys22, Arg25, Lys30, Lys69, Arg71, Lys78, Lys81, Lys85, Lys86, Arg88, and Lys91) located in regions surrounding the arm, were well-positioned to bind DNA phosphate groups and are good candidates to participate in the interaction of the arm with DNA.

Comparison with Histone or Histone-like Protein Structures. Sequence homology was found with DNA-binding proteins from different methanobacteria such as MC1 from *Methanosarcina barkeri*, MC1A, MC1B, and MC1C from *Methanothrix soehngenii*, from archaea such as Q48241 from *Halococcus morrhuae*, and Q9HMK1 from *Halobacterium* sp. (Figure 7), but today, the structure of all of these proteins is known. For these 7 proteins, the residues corresponding to the β strands I1, I2, and II1 as well as the α helix are fully conserved and their conformation can be predicted like

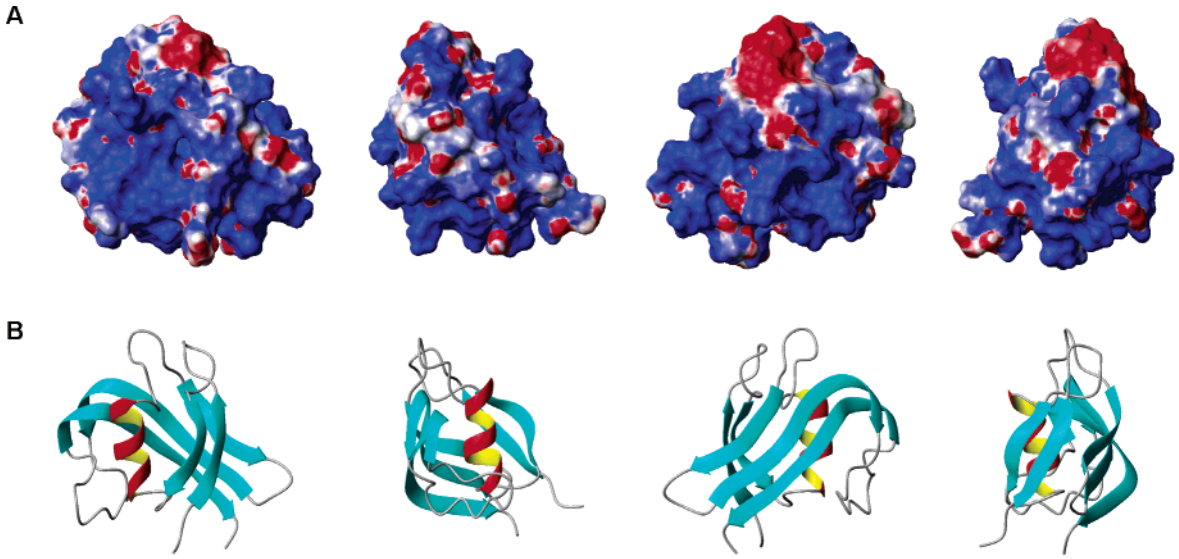

FIGURE 6: Four orthogonal views of (A) the MC1 electrostatic potential surface map (positively and negatively charged residues are colored blue and red, respectively) and (B) the corresponding ribbon diagrams created in MOLMOL (*18*). MC1 shows a highly positively charged surface favorable for electrostatic interactions with phosphates of DNA.

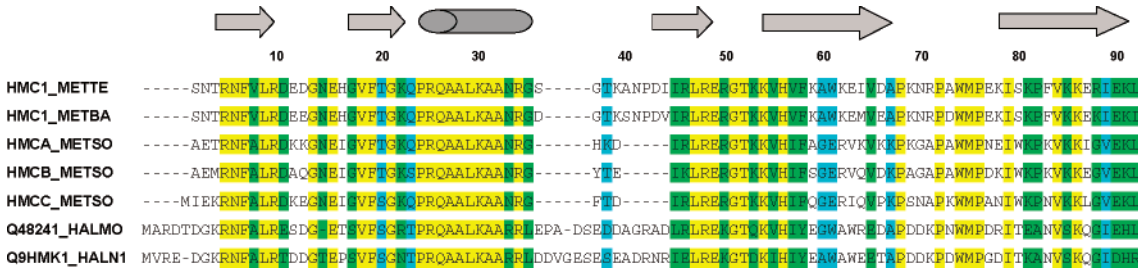

FIGURE 7: Sequence alignment of MC1s for different archaeal organisms, including MC1 from *Methanosarcina thermophila* (HMC1-METTE) and *Methanosarcina barkeri* (HMC1-METBA), MC1A, MC1B, and MC1C from *Methanothrix soehngenii* (HMCA_METSO, HMCB_METSO, and HMCC_METSO, respectively), Q48241 from *Halococcus morrhuae* and Q9HMK1 from *Halobacterium* sp. Yellow, green, and light blue indicate the level of homology (100, 70, and 50%, respectively). The secondary structure of MC1 is indicated below the alignment. The sequence alignment was created using the Swiss-Prot data bank from the primary accession number P12770 for HMC1_METTE.

those of the corresponding elements in MC1. The β strand II2 is also well-conserved in MC1 from *Methanosarcina barkeri*, but those from other proteins of Figure 7 show several substitutions between residues 59 and 65. Some mutations are also observed in the last β strand, but their small number suggests that the triple-stranded β-sheet motif is conserved in these proteins. Figure 7 shows that the less conserved part of the sequence was the loop located between the α helix and the β strand II1 (residues 36−44 in HMC1-METTE), whose length varied from 3 to 14 residues. The solution structure (Figure 5) clearly shows that the part 36−39 is buried in the β barrel and that the part 40−44 interacts with the loop 10−15 and Trp61, suggesting that a shorter loop will be sufficient to hand the respective position of the α helix and the β strand II1. It can be suggested that 10 residues of this loop do not play an important role in the common functions of the homologous MC1.

Among the known DNA-binding protein structures, one looks like MC1; it is the extreme-heat and pressure-resistant Sso7d protein (*27*). However, this small globular basic protein from the thermoacidophilic archeon *Sulfolobus solfataricus* has no sequence homology with MC1. Its binding distorts the DNA conformation and introduces significant unwinding of the double helix (*28*). Their folding topologies possess some similarities. Both MC1 and Sso7d are characterized by a double-stranded β sheet that packed against a triple-stranded β sheet by making a barrel. Sso7d also possesses an α helix but is located at its C-terminal part and not between the double and triple β-sheet motif like in MC1. Each β strand is linked by short turns, contrary to MC1, where a large loop connects strands II2 and II3 (*29*). Moreover, the long arm, which plays a crucial role in the binding of MC1 (*26*), was not present in Sso7d, and this protein binds DNA by placing its β triple strand across the DNA minor groove. A striking feature of MC1 and some other proteins (Figure 7) is that the five highly conserved prolines (among a total of six) participate in the folding of the β strands and the arm.

The CAD domain of Caspase-activated DNase (CAD), which degrades chromosomal DNA after being released from its inhibitor ICAD, is classified as a ubiquitin superfold (*30*, *31*). Like MC1, a double-and triple-stranded β sheet are linked to an α helix, but its strand 5 is antiparallel to the strand 3 and not to the strand 4. Another deep difference lies in the fact that all of the interacting β strands are antiparallel in MC1, while the strands 1 and 5 of the CAD domain are parallel.

Indeed, the topology of MC1 is characterized by the linking $\beta_{I1}-\beta_{I2}-\alpha-\beta_{II1}-\beta_{II2}-$loop$-\beta_{II3}$, where all of the interacting β strands are antiparallel and its overall fold is different from other known DNA-binding proteins. MC1 may thus be classified in a family with a new protein fold.

Possible DNA-Binding Mode of MC1. Photochemical cross-linking of MC1 on 5-bromouracil-substituted DNA showed that the cross-linked polypeptide chain is located within residues Lys69−Lys87 (*26*). This presumed DNA-binding site comprised residues K62−P82, which is formed by a part of the strands II2 and II3 belonging to the triple-stranded antiparallel β sheet and a loop flanked by prolines P68 and P76. This sequence contains 11 highly conserved residues found in every MC1 shown in Figure 7. Interestingly, the well-conserved lysines Lys81, Lys85, and Lys86 were found close in the three-dimensional structure and directed toward the solvent in ideal position to connect the phosphate groups of DNA. The nonconserved Lys78 was hydrogen-bonded to the aspartic acid Asp66. Moreover, the tryptophan Trp74 in the loop was found completely exposed to the solvent (Figure 5) and ideally placed to interact with DNA as suggested by photocross-links (*26*). It is worth noting that the other tryptophan Trp61 was located on the other side of the protein, slightly less exposed to the solvent than Trp74 because Trp61 participates in interactions of the strand II2 with the β turn 40−44. The 4 highly conserved prolines, which are present in the sequence Lys69−Lys87, participate in the conformation of the arm and suggest a key role in the DNA-binding mode. A lot of hydrogen bonds or electrostatic bonds and a proline participated to curve the region Lys59−Glu63 and Lys81−Lys85, giving a peculiar orientation to the arm, possibly the most favorable position to bind DNA. The side chain of the totally conserved arginine Arg25 belonging to the α helix was located near the charged parts of the side chains of Lys22, Lys69, and Arg71, making a zone of highly positive potential near the end of the arm (Figure 6).

For discussion, it is interesting to examine the DNA-binding mode of the basic Ssod7 protein previously mentioned. The solution structure of the Ssod7 protein shows that the face of the triple-stranded β sheet displays a continuous region of positive electrostatic potential in Sso7d (*32*) and is oriented to span the minor groove of ADN in the NMR solution structure of the complex (*28*). Interestingly, turns that interact with DNA in Sso7d (Gly26−Lys39 and Gly37−X−Lys39) were also found in MC1: Gly21−Lys22, Gly37−Thr38−Lys39, and Gly51−Thr52−Lys53.

Such a turn (Gly21-Lys22) is located near the arm, in a favorable position to interact with DNA in MC1. We note that like MC1, Sso7d possesses a tryptophan completely exposed to the solvent, which plays a major role in DNA binding (*28*, *33*). These interacting turns connect strands II1 to II2 and II2 to II3 in Sso7d, and the tryptophan is near this first turn. In MC1, these turns are far from the arm, which contains the Trp74, and suggest that the DNA-binding mode of MC1 should be different from those of Sso7d.

Another interaction mode including a β-sheet motif has also been described for the *Bacillus stearothermophilus* protein HU. The DNA-binding protein HU is quite different from MC1 because it consists of a dimer formed by two identical monomers of 90 residues each, is folded in 6 β sheets and 2 α helices (3 β sheets and an α helix by the monomer) (*34*, *35*), and its β arms interact in the minor groove of DNA and invoke dramatic bending (*36*, *37*). NMR studies have shown that the tips of the arms are highly flexible and suggested a DNA-binding model in which the tips of the arms wrapped around the DNA. Interestingly, the region of residues 55−74 is basically an antiparallel β hairpin, which contains a double flip over at prolines 63 and 72 in both strands of the β arm. The curved shape of the β-hairpin arms resulting from this backbone geometry is believed to improve the interaction with DNA. Such a conformation was also found in the arm of MC1 suggesting a preadaptation of the shape of the arm to bind DNA.

In summary, the solution structure reported here will greatly help to understand the DNA-binding mode of this new folded-motif protein. A complete NMR study of this complex requires a ^{15}N-,^{13}C-labeled protein, and in our effort to further characterize MC1, the cloning of a synthetic gene encoding MC1 and expressed in *Escherichia coli* is now in progress in our laboratory.

ACKNOWLEDGMENT

We thank Alain Gervais for extracting and purifying the sample of MC1.

SUPPORTING INFORMATION AVAILABLE

Figure 1A shows the dαN NOE connectivities in the NOESY spectrum observed with a mixing time of 100 ms at 28 °C. Figure 1B indicates the sequential dαN(i, $i + 1$) way (red) for the residues Q26−G35 and the dαN(i, $i + 3$) and dαN(i, $i + 4$) NOEs connectivities labeled in green. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Imbert, M., Laine, B., Helbecque, N., Mornon, J. P., Henichart, J. P., and Sautiere, P. (1990) Conformational study of the chromosomal protein MC1 from the archaebacterium *Methanosarcina barkeri*, *Biochim. Biophys. Acta 1038*, 346−354.
2. Laine, B., Chartier, F., Imbert, M., Lewis, R., and Sautiere, P. (1986) Primary structure of the chromosomal protein HMb from the archaebacteria *Methanosarcina barkeri*, *Eur. J. Biochem. 161*, 681−687.
3. Chartier, F., Laine, B., and Sautiere, P. (1988) Characterization of the chromosomal protein MC1 from the thermophilic archaebacterium *Methanosarcina* sp. CHTI55 and its effect on the thermal stability of DNA, *Biochim. Biophys. Acta 951*, 149−156.
4. 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. CHTI55, *Biochim. Biophys. Acta 1008*, 309−314.
5. Le 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.
6. Laine, B., Culard, F., Maurizot, J. C., and Sautiere, P. (1991) The chromosomal protein MC1 from the archaebacterium *Methanosarcina* sp. CHTI55 induces DNA bending and supercoiling, *Nucleic Acids Res. 19*, 3041−3045.
7. 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.
8. 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 archaebacterial histone-like protein MC1, *J. Biol. Chem. 270*, 6286−6291.
9. 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.
10. Isabelle, V., Franchet-Beuzit, J., Sabattier, R., Laine, B., Spotheim-Maurizot, M., and Charlier, M. (1993) Radioprotection of DNA by a DNA-binding protein: MC1 chromosomal protein from the archaebacterium *Methanosarcina* sp. CHTI55, *Int. J. Radiat. Biol. 63*, 749−758.
11. 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.
12. 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.
13. Piotto, M., Saudek, V., and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions, *J. Biomol. NMR 2*, 661−665.
14. Bartels, C., Xia, T., Guntert, P., Billeter, M., and Wüthrich, K. (1995) The program Xeasy for computer-supported NMR spectral analysis, *J. Biomol. NMR 5*, 1−10.
15. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J. Mol. Biol. 273*, 283−298.
16. Brooks, B., Brucoli, R., Olafson, B. O., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem. 4*, 187−217.
17. Brünger, A. T. (1992) *XPLOR: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.
18. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: A program for display and analysis of macromolecular structures, *J. Mol. Graphics 14*, 51−55.
19. Hutchinson, E. G., and Thornton, J. M. (1996) PROMOTIF−A program to identify and analyze structural motifs in proteins, *Protein Sci. 5*, 212−220.
20. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
21. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK−NMR: Programs for checking the quality of protein structures solved by NMR, *J. Biomol. NMR 8*, 477−486.
22. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience Publication, New York.
23. Englander, S. W., and Wand, A. J. (1987) Main-chain-directed strategy for the assignment of ^{1}H NMR spectra of proteins, *Biochemistry 26*, 5953−5958.
24. Zhang, H., Neal, S., and Wishart, D. S. (2003) RefDB: A database of uniformly referenced protein chemical shifts, *J. Biomol. NMR 25*, 173−195.
25. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The chemical shift index: A fast and simple method for the assignment of protein secondary structure through NMR spectroscopy, *Biochemistry 31*, 1647−1651.
26. Katouzian-Safadi, M., Laine, B., Chartier, F., Cremet, J. Y., Belaiche, D., Sautiere, P., and Charlier, M. (1991) Determination of the DNA-interacting region of the archaebacterial chromosomal

protein MC1. Photocrosslinks with 5-bromouracil-substituted DNA, *Nucleic Acids Res. 19*, 4937−4941.
27. Consonni, R., Santomo, L., Fusi, P., Tortora, P., and Zetta, L. (1999) A single-point mutation in the extreme heat- and pressure-resistant sso7d protein from *Sulfolobus solfataricus* leads to a major rearrangement of the hydrophobic core, *Biochemistry 38*, 12709−12717.
28. Agback, P., Baumann, H., Knapp, S., Ladenstein, R., and Hard, T. (1998) Architecture of nonspecific protein−DNA interactions in the Sso7d−DNA complex, *Nat. Struct. Biol. 5*, 579−584.
29. Baumann, H., Knapp, S., Lundback, T., Ladenstein, R., and Hard, T. (1994) Solution structure and DNA-binding properties of a thermostable protein from the archaeon *Sulfolobus solfataricus*, *Nat. Struct. Biol. 1*, 808−819.
30. Uegaki, K., Otomo, T., Sakahira, H., Shimizu, M., Yumoto, N., Kyogoku, Y., Nagata, S., and Yamazaki, T. (2000) Structure of the CAD domain of caspase-activated DNase and interaction with the CAD domain of its inhibitor, *J. Mol. Biol. 297*, 1121−1128.
31. Otomo, T., Sakahira, H., Uegaki, K., Nagata, S., and Yamazaki, T. (2000) Structure of the heterodimeric complex between CAD domains of CAD and ICAD, *Nat. Struct. Biol. 7*, 658−662.
32. Baumann, H., Knapp, S., Karshikoff, A., Ladenstein, R., and Hard, T. (1995) DNA-binding surface of the Sso7d protein from *Sulfolobus solfataricus*, *J. Mol. Biol. 247*, 840−846.
33. Gao, Y. G., Su, S. Y., Robinson, H., Padmanabhan, S., Lim, L., McCrary, B. S., Edmondson, S. P., Shriver, J. W., and Wang, A. H. (1998) The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA, *Nat. Struct. Biol. 5*, 782−786.
34. Tanaka, I., Appelt, K., Dijk, J., White, S. W., and Wilson, K. S. (1984) 3 Å resolution structure of a protein with histone-like properties in prokaryotes, *Nature 310*, 376−381.
35. White, S. W., Appelt, K., Wilson, K. S., and Tanaka, I. (1989) A protein structural motif that bends DNA, *Proteins 5*, 281−288.
36. Vis, H., Mariani, M., Vorgias, C. E., Wilson, K. S., Kaptein, R., and Boelens, R. (1995) Solution structure of the HU protein from *Bacillus stearothermophilus*, *J. Mol. Biol. 254*, 692−703.
37. Boelens, R., Vis, H., Vorgias, C. E., Wilson, K. S., and Kaptein, R. (1996) Structure and dynamics of the DNA binding protein HU from *Bacillus stearothermophilus* by NMR spectroscopy, *Biopolymers 40*, 553−559.

BI048382Z