JOURNAL OF BACTERIOLOGY, Jan. 2007, p. 403–409 Vol. 189, No. 2
0021-9193/07/$08.00+0 doi:10.1128/JB.01246-06

CC1, a Novel Crenarchaeal DNA Binding Protein▽

Xiao Luo,[1] Uli Schwarz-Linek,[1] Catherine H. Botting,[1] Reinhard Hensel,[2] Bettina Siebers,[2] and Malcolm F. White[1]*

Centre for Biomolecular Sciences, University of St. Andrews, St. Andrews, Fife KY16 9ST, United Kingdom,[1] and Microbiology I, University Duisburg-Essen, Universitätsstr. 5, 45117 Essen, Germany[2]

Received 8 August 2006/Accepted 21 October 2006

The genomes of the related crenarchaea *Pyrobaculum aerophilum* and *Thermoproteus tenax* lack any obvious gene encoding a single-stranded DNA binding protein (SSB). SSBs are essential for DNA replication, recombination, and repair and are found in all other genomes across the three domains of life. These two archaeal genomes also have only one identifiable gene encoding a chromatin protein (the Alba protein), while most other archaea have at least two different abundant chromatin proteins. We performed a biochemical screen for novel nucleic acid binding proteins present in cell extracts of *T. tenax*. An assay for proteins capable of binding to a single-stranded DNA oligonucleotide resulted in identification of three proteins. The first protein, Alba, has been shown previously to bind single-stranded DNA as well as duplex DNA. The two other proteins, which we designated CC1 (for crenarchaeal chromatin protein 1), are very closely related to one another, and homologs are restricted to the *P. aerophilum* and *Aeropyrum pernix* genomes. CC1 is a 6-kDa, monomeric, basic protein that is expressed at a high level in *T. tenax*. This protein binds single- and double-stranded DNAs with similar affinities. These properties are consistent with a role for CC1 as a crenarchaeal chromatin protein.

All organisms use chromatin proteins, which allow reversible compaction of their genomes in a manner that allows DNA replication, gene expression, recombination, and repair. In eukarya the nucleosome, comprising a histone octamer, is the universal unit for DNA packaging. In prokaryotes, however, there is a much wider diversity of chromatin proteins. The bacterial chromatin proteins include HU, IHF, Fis, and H-NS, and there is variable distribution of these proteins across the bacterial family tree (25). The archaea have a different set of chromatin proteins, including a true histone protein, which organizes DNA into tetrameric nucleosomes (20), and the dimeric Alba protein, which can form extended filamentous fibers with DNA (12, 14). The distribution of these proteins is complex, with the histones primarily restricted to the euryarchaea (with the notable exception of *Cenarchaeum symbiosum*, a marine crenarchaeote [7]) and Alba present in most archaea with the exception of the *Methanosarcinales* and halophiles. A number of other chromatin proteins, including MC1, 7kMk, and HU, are present in a much more restricted subset of the archaea (19, 23) (Table 1).

It has been noted previously that almost all archaeal species encode more than one type of chromatin protein (23); for example, histones and Alba are both present in *Pyrococcus* species, while *Sulfolobus* species possess Alba and the Sul7 protein and members of the *Methanosarcinales* have histones and the MC1 protein. Purified chromatin proteins tend to form extended filaments when they are bound to DNA (14), and it has been suggested that expression of multiple chromatin proteins provides a mechanism for the control of this phenomenon in vivo (19). Notable exceptions to this rule are the genomes of *Pyrobaculum aerophilum* and *Aeropyrum pernix*, which have one and two copies of the Alba gene, respectively, but no other genes for known chromatin proteins (9).

In addition to double-stranded DNA (dsDNA) binding proteins, all organisms contain single-stranded DNA (ssDNA) binding proteins (SSBs), which bind, sequester, and protect ssDNA generated during DNA replication, recombination, and repair. SSBs in all three domains of life are related and based on the oligonucleotide binding (OB) fold domain (16). The crenarchaeal SSB of *Sulfolobus solfataricus* is monomeric, with a domain organization similar to that of bacterial SSBs (21) but a structure closely related to that of the eukaryal SSB replication protein A (13). The euryarchaeal SSBs have various domain organizations that can resemble the heterotrimeric structure of eukaryal replication protein A quite closely (6). The sequence signature of the SSB OB fold is conserved and can therefore be detected quite readily using bioinformatics. Intriguingly, the genome of *P. aerophilum* lacks any obvious candidate *ssb* gene; this genome is the only published genome in this category from any domain of life (22). The genome sequence of the related crenarchaeon *Thermoproteus tenax* is currently being annotated (unpublished data). Like *P. aerophilum*, *T. tenax* also lacks a second chromatin protein (it has one gene for Alba) and any obvious SSB (B. Siebers, unpublished data). These observations suggest that these crenarchaea harbor a novel or highly divergent protein for ssDNA binding.

Given the likelihood that *T. tenax* and *P. aerophilum* encode unusual double- and/or single-stranded DNA binding proteins, we used a biochemical approach to purify and identify such proteins from *T. tenax*, based on an assay for DNA binding. We detected two major protein fractions that retarded a DNA oligonucleotide significantly, and the proteins were purified to homogeneity and identified by mass spectrometry. The first protein turned out to be a *Thermoproteus* Alba ortholog, TtxAlba. The second protein was a

* Corresponding author. Mailing address: Centre for Biomolecular Sciences, University of St. Andrews, St. Andrews, Fife KY16 9ST, United Kingdom. Phone: 44-1334-463432. Fax: 44-1334-462595. E-mail: mfw2@st-and.ac.uk.

▽ Published ahead of print on 3 November 2006.

TABLE 1. Summary of archaeal chromatin proteins

Order	Phylum[a]	Genus	No. of genes encoding:		Other protein (no. of genes)
			Histones	Alba	
Sulfolobales	CA	*Sulfolobus*	0	2	Sul7 (2 or 3)
Desulfurococcales	CA	*Aeropyrum*	0	2	CC1[b]
Thermoproteales	CA	*Pyrobaculum*	0	1	CC1 (3)[b]
	CA	*Thermoproteus*	0	1	CC1 (3)[b]
Thermoplasmales	EA	*Thermoplasma*	0	1	HU
		Picrophilus	0	1	HU
Thermococcales	EA	*Pyrococcus*	2	1	
Methanopyrales	EA	*Methanopyrus*	4	2	7kMk
		Nanoarchaeum	2	1	
Methanococcales	EA	*Methanocaldococcus*	5	1	
		Methanococcus	2	1	
Archaeglobales	EA	*Archaeoglobus*	2	2	
Methanobacteriales	EA	*Methanothermobacter*	3	1	
Methanosarcinales	EA	*Methanosarcina*	1	0	MC1 (1 or 2)
Halophiles	EA	*Halobacterium*, *Haloarcula*	1–4	0	MC1

[a] CA, crenarchaea; EA, euryarchaea.
[b] Identified in this study.

novel family consisting of two highly related proteins, which we designated CC1 (for crenarchaeal chromatin protein 1). Homologs of CC1 are restricted to *T. tenax*, *P. aerophilum*, and *A. pernix*. CC1 of *T. tenax* binds both ssDNA and dsDNA in a highly cooperative fashion, and it is present at a high level in the cell. These observations are consistent with an important role for CC1 as a chromatin protein in the *Thermoproteus-Pyrobaculum* (*Thermoproteales*) and *Aeropyrum* (*Desulfurococcales*) cell lineages.

MATERIALS AND METHODS

Growth, cell lysis, purification, and identification of chromatin proteins from *T. tenax*. Mass cultures of *T. tenax* Kra 1 (= DSM 2078) (27) were grown at 86°C in an enamelled 100-liter fermentor (Braun Biotec International, Melsungen, Germany) in a medium described by Brock et al. (3), as reported previously (4). Cell lysis, centrifugation, and chromatography were carried out at 4°C. Cells (13.6 g) were thawed in 90 ml lysis buffer (30 mM morpholineethanesulfonic acid [MES], 50 mM NaCl, 1 mM dithiothreitol [DTT], 1 mM EDTA; pH 6.5) and immediately sonicated three times for 2 min with cooling. The lysate was centrifuged at 20,000 × *g* three times for 30 min. The supernatant was passed through a 0.45-μm HT Tuffryn membrane and then was applied to a HiTrap Heparin HP 5-ml column (GE Healthcare) equilibrated with buffer A (20 mM MES, 1 mM DTT, 1 mM EDTA; pH 6.5). A linear gradient of 0 to 1,000 mM NaCl in buffer A was used to elute proteins. Fractions were collected and analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and they were assayed to determine their abilities to retard an ssDNA oligonucleotide in an electrophoretic mobility shift assay (EMSA). The relevant fractions were pooled, diluted threefold with buffer A, and loaded onto a Resource 30S cation-exchange column (GE Healthcare). A linear gradient of 0 to 1,000 mM NaCl in buffer A was used to elute proteins. Fractions were analyzed as described above, and relevant fractions were concentrated and applied to a Superose 12 HR 10/30 gel filtration column (GE Healthcare) that was developed with buffer A containing 200 mM NaCl. Active fractions identified by the EMSA coincided with a single, essentially homogeneous polypeptide observed by SDS-PAGE. The proteins were identified by tryptic digestion and nano liquid chromatography-electrospray ionization-mass spectrometry/mass spectrometry.

Oligonucleotides. For all DNA binding studies the following 45-mer oligonucleotide labeled with an internal fluorescein moiety to allow fluorescent detection was utilized: 5′-CCGAATAGCGGAATTCACGAGTACCTGCGGCCTC GAGGGA(fluoro-dT)CCGT. This oligonucleotide had an internal fluorescein because it was designed originally as part of a construct to investigate nucleotide excision repair in archaea. Oligonucleotides with fluorescein modifications at the 5′ end yield results similar to those reported here. For experiments requiring a duplex DNA species, the fluorescent oligonucleotide was annealed with an unmodified complementary oligonucleotide, 5′-ACGGATCCCTCGAGGCCGCA GGTACTCGTGAATTCCGCTATTCGG.

Gel electrophoretic retardation assay. During protein purification, fractions eluting from chromatography columns were tested to determine their abilities to retard a fluorescent ssDNA oligonucleotide in binding buffer (20 mM MES [pH 6.5], 1 mM DTT, 1 mM EDTA, 0.2 mg/ml bovine serum albumin). Aliquots (4 μl) of each fraction were incubated with the fluorescent oligonucleotide (final concentration, 2 μM) in a 5-μl (total volume) mixture. After incubation for 15 min at 20°C, 1/6 volume loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 20% Ficoll type 400) was added, and samples were loaded onto 10% acrylamide gels and electrophoresed in 1× Tris-borate-EDTA buffer at 60 V for 120 min to separate bound and free DNA species. The DNA was visualized using a Fuji FLA5000 imager with a blue laser.

Determination of DNA binding affinity. Different concentrations of purified CC1 or TtxAlba were incubated with either the single-stranded oligonucleotide (2 μM) or the double-stranded form (1 μM) in binding buffer (20 mM MES [pH 6.5], 1 mM DTT, 1 mM EDTA, 0.2 mg/ml bovine serum albumin) in 5-μl (total volume) mixtures. After incubation for 15 min at 20°C, 1/6 volume loading buffer was added, and samples were loaded onto 12% acrylamide gels and electrophoresed in 1× Tris-borate-EDTA buffer at 30 V for 240 min to separate bound and free DNA species. Binding was quantified using a Fuji FLA5000 imager with a blue laser.

Secondary structure analysis by circular dichroism. Far-UV (260 nm to 180 nm) circular dichroism (CD) spectra were recorded at room temperature with a Jasco J-810 spectropolarimeter using quartz cells with a path length of 0.05 cm with Ttx1853 in D_2O (0.13 mg/ml). The D_2O was present so that proton nuclear magnetic resonance (NMR) experiments could be conducted, and it was considered equivalent to H_2O for the CD spectroscopy analysis. The CD spectra from six wavelength scans were averaged and corrected using a D_2O blank. The secondary structure content was predicted using the DICHROWEB web server (24) with the CDSSTR algorithm and reference data set 3.

NMR. NMR experiments were performed at 298 K with a Bruker DRX500 spectrometer equipped with a z-shielded gradient triple-resonance probe and the XWIN-NMR 3.5 control and processing software. Pure Ttx1853 protein was dialyzed overnight against 20 mM phosphate buffer (pH 6.5) using a dialysis cassette (molecular mass cutoff, 3.5 kDa; Pierce) and was concentrated to 570 μl with a centrifugal ultrafiltration unit (molecular mass cutoff, 3 kDa; Pall). Thirty microliters of D_2O (5%) and sodium azide (0.02%) was added to the sample. The final protein concentration was approximately 50 μM. A one-dimensional ^1H spectrum was acquired using a standard pulse sequence with Watergate water suppression. A total of 8,192 scans were accumulated, the sweep width was set to 12,500 Hz, and the digital resolution was 1.52 Hz/point.

To reduce the signal from water, the NMR sample was freeze-dried, dissolved in 0.5 ml D_2O, freeze-dried again, and dissolved in 0.6 ml D_2O. The final protein concentration in the sample was approximately 30 μM. One-dimensional ^1H spectra without water suppression (8,192 scans) were acquired shortly after

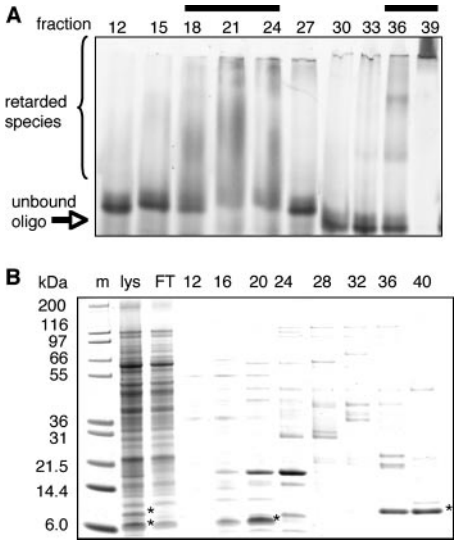

FIG. 1. Detection of an ssDNA binding activity in extracts of *T. tenax*. (A) Proteins were fractionated with a HiTrap Heparin HP column, and aliquots were tested to determine their abilities to retard a fluorescent single-stranded oligonucleotide in an EMSA. Fraction numbers are indicated above the gel. Significant retarded species were centered on fractions 21 and 39. (B) SDS-PAGE analysis of fractionated proteins. The crude cell lysate (lys) and unbound fractions (FT), as well as selected bound fractions, were separated by SDS-PAGE and detected by Coomassie blue staining. Two small protein species visible in the crude lysate (indicated by asterisks) coincided with the two areas of gel retardation observed by EMSA. Lane m contained markers.

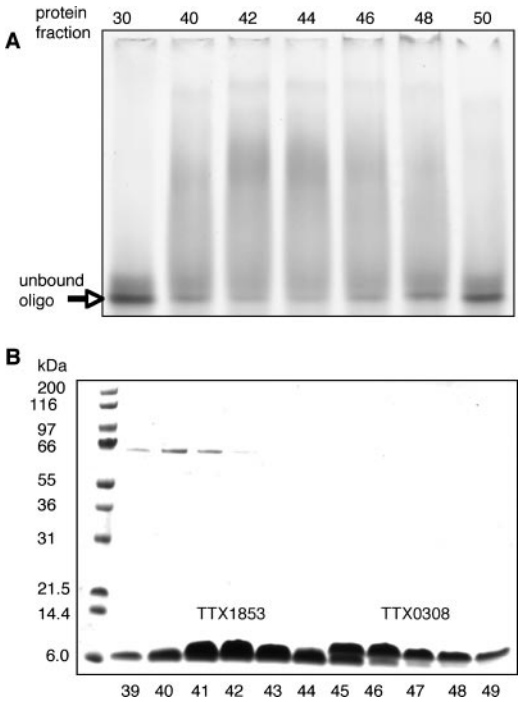

FIG. 2. Purification of novel DNA binding proteins from *T. tenax*. (A) Fractions 18 to 25 from the heparin column were pooled and fractionated using a Resource 30S cation-exchange column, and fractions were tested to determine their abilities to retard the fluorescent oligonucleotide. Retarded DNA species were observed in fractions 40 to 48. (B) SDS-PAGE analysis of proteins fractionated by cation-exchange chromatography. Two overlapping protein species were apparent in the fractions that gave rise to retarded DNA species, and these species were identified by mass spectrometry as Ttx1853/1420 (Fig. 3) and Ttx0308. The fractions eluting from the column are indicated below the gel.

sample preparation, after 24 h, and after 48 h. The sample was kept at room temperature between the experiments.

Mass spectrometry: intact mass measurement. A protein sample (15 μl; 5 pM/μl) was desalted online using a MassPrep online desalting cartridge (2.1 by 10 mm) eluted with increasing acetonitrile concentrations (2% acetonitrile–98% aqueous 1% formic acid to 98% acetonitrile–2% aqueous 1% formic acid) and was delivered to an electrospray ionization mass spectrometer (LCT; Micromass, Manchester, United Kingdom) which had previously been calibrated using myoglobin. An envelope of multiply charged signals was obtained and deconvoluted using the MaxEnt1 software to obtain the molecular mass of the protein.

Mass spectrometry: protein identification. A band from an SDS-PAGE gel was excised to obtain four approximately 1-mm cubes. These cubes were then subjected to in-gel digestion, using a ProGest Investigator in-gel digestion robot. Briefly, the gel cubes were destained by washing them with acetonitrile and were subjected to reduction and alkylation before digestion with trypsin at 37°C. The peptides were extracted with 10% formic acid and concentrated to 20 μl (SpeedVac; ThemoSavant). They were then separated using an UltiMate nanoLC (LC Packings, Amsterdam, The Netherlands) equipped with a PepMap C_{18} trap and column. The eluent was sprayed into a Q-Star Pulsar XL tandem mass spectrometer (Applied Biosystems, Foster City, CA) and analyzed in information-dependent acquisition mode. Mass spectrometry/mass spectrometry data for doubly and triply charged precursor ions were analyzed using the ProID software (Applied Biosytems), searching against a database containing the protein translations of the *T. tenax* open reading frames. The data were searched with a tolerance of 0.2 Da for the precursor and fragment ions, trypsin as the cleavage enzyme, one missed cleavage, and carbamidomethyl modification of cysteines and methionine oxidation selected as possible modifications. The sample was identified as the protein hit with the most significant score, and this score was significantly above the score for any other match and had the largest number of matching peptides.

RESULTS

To screen for abundant DNA binding proteins from *T. tenax*, 14 g cells was lysed and subjected to heparin chromatography. Proteins were eluted using a salt gradient, and fractions were analyzed by SDS-PAGE (Fig. 1A). Protein fractions were tested to determine their abilities to bind to and hence retard a 45-nucleotide DNA oligonucleotide tagged with a fluorescein reporter. Significant retardation of the DNA was detected in two different regions of the eluant, centered on fractions 18 to 24 and 36 to 40 (Fig. 1A). These regions coincided with the elution points of two small proteins present at high levels in the crude extract (Fig. 1B). The protein eluting around fractions 36 to 40 was identified by mass spectrometry as the *T. tenax* Alba (Sac10b) ortholog. Alba is known to interact with single-stranded DNA, double-stranded DNA, and RNA (10, 14, 15). The TtxAlba protein was purified to homogeneity by gel filtration chromatography prior to all further characterization.

Fractions 18 to 24 corresponding to binding activity were pooled and subjected to cation-exchange chromatography on a Resource 30S column. The protein fractions that retarded the fluorescent oligonucleotide were centered on fractions 40 to 48 (Fig. 2A). This region corresponded to two small protein species migrating in SDS-PAGE with a mass of around 7 kDa (Fig. 2B). Mass spectrometry (nano liquid chromatography-electrospray ionization-mass spectrometry/mass spectrometry) coupled with tryptic digestion and interrogation of a *T. tenax*

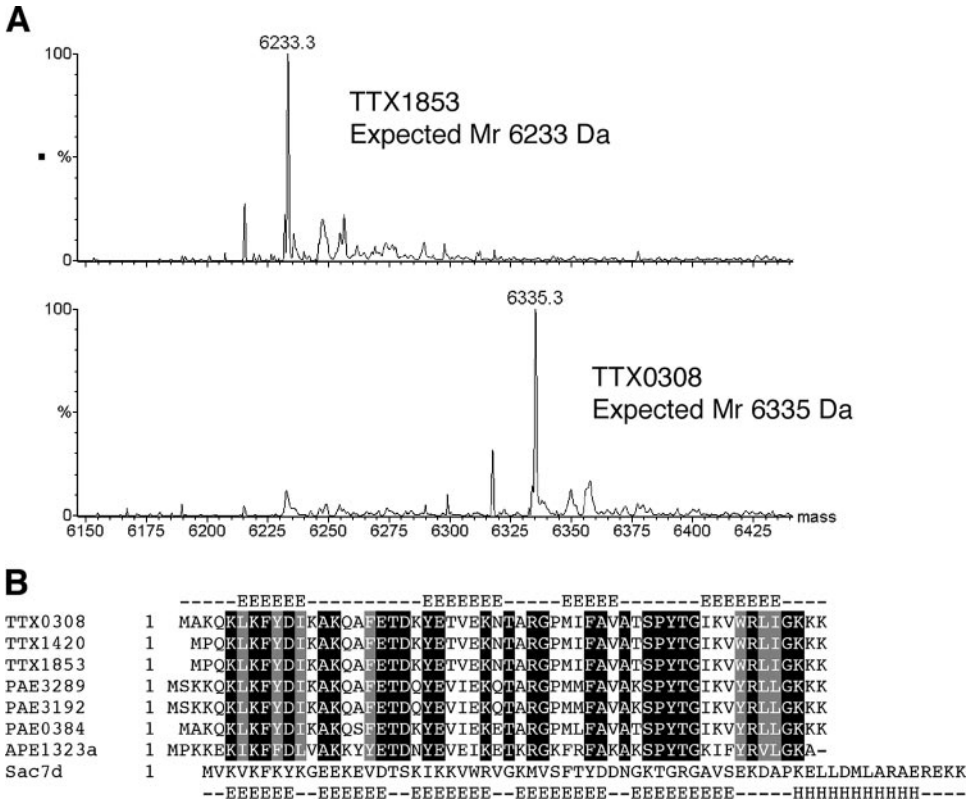

FIG. 3. Novel family of DNA binding proteins in *T. tenax*, *P. aerophilum*, and *A. pernix*. (A) Characterization of purified DNA binding proteins by mass spectrometry. Following final purification by gel filtration chromatography, the Ttx1853/1420 and Ttx0308 proteins were analyzed by mass spectrometry. The expected masses of Ttx1853/1420 and Ttx0308 following posttranslational cleavage of the N-terminal methionine residue are 6,235 Da and 6,337 Da, respectively, in good agreement with experimental determinations. (B) Sequence alignment of the novel DNA binding protein family, CC1, including homologs from *T. tenax*, *P. aerophilum*, and *A. pernix.* The secondary structure prediction from the JPRED program is shown above the sequences (E, predicted β-sheet). Below the CC1 alignment, the sequence of the Sul7 protein Sac7d from *Sulfolobus acidocaldarius*, along with its known secondary structure (E, sheet; H, helix), is shown for comparison.

protein sequence database resulted in positive identification of the protein eluting first as Ttx1853/1420 (encoded by two *T. tenax* genes that have identical gene products) and of the protein eluting in peak 2 as Ttx0308. The two fractions were purified independently by gel filtration chromatography on a Superose-12 column, yielding homogeneous proteins that were characterized by mass spectrometry, which yielded molecular masses in good agreement with the expected masses for these two proteins after posttranslational cleavage of the N-terminal methionine residues (Fig. 3A). Using a size exclusion column calibrated with molecular weight standards, we investigated the likely quaternary structure of Ttx0308. The protein eluted from the column with a retention time equivalent to an estimated molecular mass of 7.6 kDa, which is consistent with a monomeric composition for the protein (data not shown).

Ttx1853/1420 and Ttx0308 have very similar primary structures, exhibiting >95% sequence identity. The only homologous proteins in the database are a family consisting of two proteins from *P. aerophilum* and one protein from *A. pernix* (Fig. 3B). Interestingly, three encoding genes were also identified in *P. aerophilum* (PAE0384, PAE3289, PAE3192), and two of them also have the same gene product (PAE3289, PAE3192), whereas *A. pernix* harbors only one gene homolog (APE1323a). The *T. tenax* protein Ttx1853/Ttx1420 exhibits 86% (PAE0384), 84% (PAE3289, PAE3192), and 50% (APE1323a) identity with these proteins. We propose the designation CC1 for this family, since these proteins are found in the crenarchaeal orders *Thermoproteales* and *Desulfurococcales* and since this nomenclature is consistent with the nomenclature used for the MC1 (methanogen chromatin protein 1) family (5, 17). The small size, limited phylogenetic distribution, very basic isoelectric point (calculated pI, around 9.9), and presence of multiple copies of the CC1 gene are all highly reminiscent of the Sul7 family of proteins found in *Sulfolobus* species (23). However, there was no detectable similarity between CC1 and any other protein at the amino acid level (Fig. 3B). Secondary structure prediction using the JPRED server (8) suggested the presence of four β-sheets and no α-helices. This compares to five β-sheets followed by one α-helix in the larger Sac7d protein, which adopts a fold reminiscent of the OB fold found in the SSBs, with a triple-stranded β-sheet forming the DNA binding interface (18). There is no clear conservation in CC1 of the Sul7 residues involved in the interface with DNA (Fig. 3B). Thus, we identified a novel, abundant DNA binding protein from the *Pyrobaculum-Thermoproteus-Aeropyrum* (*Thermoproteales* and *Desulfurococcales*) group of the crenarchaea.

To assess the role of the proteins in more detail, we deter-

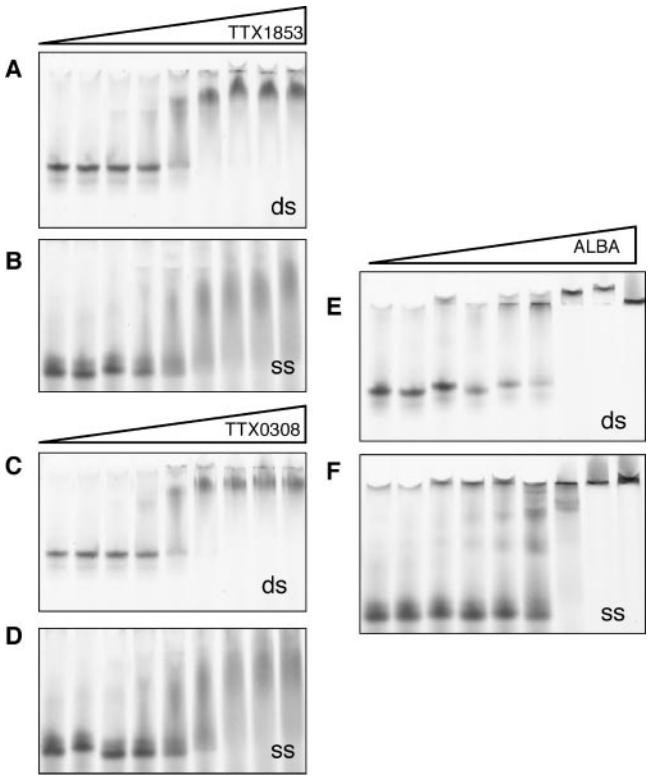

FIG. 4. Comparison of DNA binding affinities for single- and double-stranded DNA species for the CC1 and TtxAlba proteins. Fluorescent double-stranded (ds) or single-stranded (ss) DNA species were incubated with different concentrations of Ttx1853/1420 (A and B), Ttx0308 (C and D), and Alba (E and F) in binding buffer for 15 min at 20°C prior to gel electrophoresis. For Ttx1853/1420 and Ttx0308 the protein concentrations in the lanes from left to right were 0, 0.98, 1.95, 3.9, 7.8, 15.6, 31.2, 62.5, and 125 μM, respectively. For TtxAlba the protein concentrations in the lanes from left to right were 0, 0.5, 1.0, 1.9, 3.8, 7.5, 15, 30, and 60 μM, respectively.

mined the DNA binding affinities of Ttx1853/1420 and Ttx0308 for the 45-nucleotide fluorescent ssDNA oligonucleotide used in the initial assay and in parallel for a 45-bp duplex DNA having the same sequence. For comparison, we carried out the same analysis for TtxAlba (Fig. 4). The apparent dissociation constant for both *T. tenax* proteins was estimated to be around 8 μM for both the ssDNA and dsDNA species (Fig. 4A to D, lane 5). For comparison, the TtxAlba protein bound slightly more tightly to dsDNA than to ssDNA, consistent with results obtained with Alba from *S. solfataricus* (C. Jelinska and M. White, unpublished data). CC1 binding transitions were sharp for both ssDNA and dsDNA, suggesting that there was highly cooperative binding, as has been noted previously for the Alba protein (12).

The secondary structure of the CC1 Ttx1853/1420 was evaluated by circular dichroism spectroscopy in the far-UV region. The spectrum obtained was characteristic of a folded protein with a high β-sheet content (Fig. 5). The secondary structure of CC1 was estimated using the DICHROWEB online server, which yielded values of 52% β-sheet, 26% β-turn, and 21% random coil and a low normalized root mean square deviation of 0.014 (24), in good agreement with the secondary structure prediction analysis based on the sequence. NMR spectroscopy

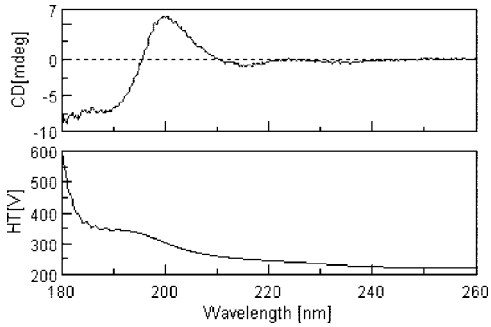

FIG. 5. Far-UV CD spectroscopy of CC1 suggests a secondary structure rich in β-sheets. The data are the results of a far-UV circular dichroism analysis of a 0.13-mg/ml solution of Ttx1853/1420 in D_2O. The spectrum suggests a content consisting of approximately 50% β-sheets and minimal α-helices.

was used to further characterize the folded state of the CC1 Ttx1853/1420. Due to limited protein availability only one-dimensional ^1H spectra in water and D_2O were acquired. The well-dispersed spectra had several features characteristic of folded proteins (26). The presence of high-field shifted methyl proton signals (−0.6 to 0.5 ppm) (Fig. 6) is an indicator of ring current shielding effects as they occur in the hydrophobic core of a protein fold. Due to the strong water signal no information regarding proton signals near the water frequency (4.68 ppm) could be obtained from the water sample. After the protein was placed into D_2O, however, α proton resonances downfield of the water frequency (4.8 to 5.6 ppm) were observed (Fig. 6). Protein NMR signals in this spectral region are characteristic of amino acid residues involved in β-sheets. Finally, despite twofold exchange of the protein into deuterated solvent, several residual signals for amide protons were observed, even after 24 h and 48 h of incubation. These slowly exchanging protons were strongly indicative of a stable folded state for CC1.

DISCUSSION

The two major archaeal chromatin proteins, Alba and the archaeal histone, have a wide phylogenetic distribution. All archaeal species have either one of these proteins, and many encode both proteins (Table 1). In addition, there are a number of abundant, unrelated chromatin proteins with a much narrower distribution, including Sul7, MC1, and 7kMk, and here we add the CC1 protein to this list. Identification of the CC1 protein fills the gaps that were previously present in Table 1, and all sequenced archaeal genomes are now known to encode more than one chromatin protein. CC1 is in many respects a typical archaeal chromatin protein, and its physical properties are highly reminiscent of those of the *Sulfolobus* Sul7 protein. Nevertheless, it has no clear relationship at the amino acid sequence level to any previously described protein. This emphasizes the sheer diversity of chromatin proteins in the prokaryotic world. Structural studies, now under way, may shed more light on the origins of the CC1 protein, as protein folds often provide evidence for deep evolutionary relationships when the primary sequence has diverged to the point where homology can no longer be detected. The prediction

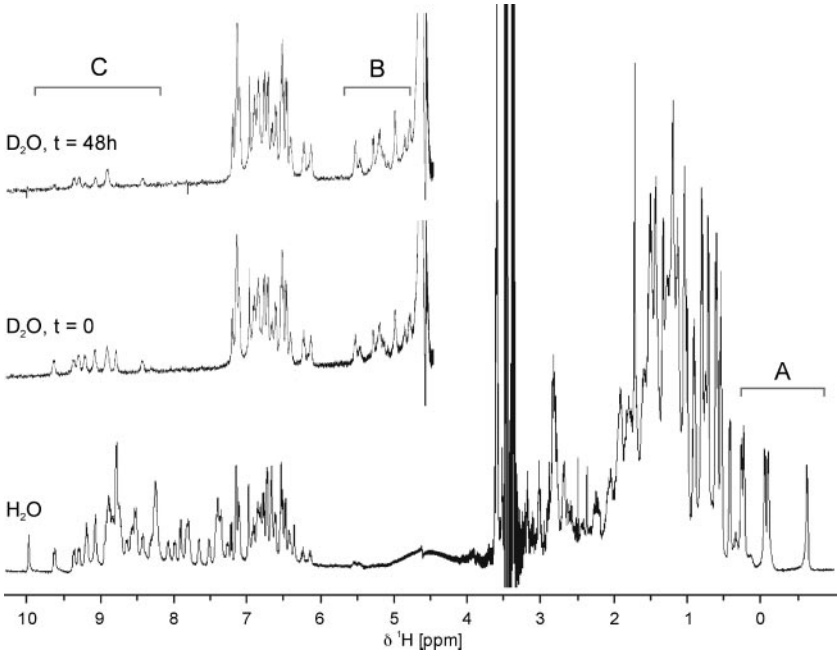

FIG. 6. One-dimensional ^{1}H NMR spectrum of Ttx1853/1420. The insets show the downfield regions of spectra in D_2O, acquired shortly after the sample was dissolved (zero time) and after 48 h. High-field shifted proton resonances (A), α proton signals indicative of β-sheets (B), and slowly exchanging amide proton resonances (C) are shown. The strong truncated signals around 3.5 ppm belong to a residual buffer component (MES) that was not removed completely in the sample preparation process.

based on sequence analysis, CD, and NMR spectroscopy of a high percentage of β-sheet structure for CC1 is something that CC1 has in common with both MC1 (17) and Sul7 (1) and indeed with the OB fold (16).

In contrast, single-stranded DNA binding proteins are much more highly conserved, and SSBs based on the basic unit of the OB fold are ubiquitous across the three domains of life (11, 13). Notable exceptions to this are the genomes of *P. aerophilum* and *T. tenax*, suggesting that there is something unusual about the SSBs of this group of organisms. Our functional screen for novel SSBs in *T. tenax* revealed only two major ssDNA binding activities, which have been identified as Alba and CC1. Both proteins bind double-stranded DNA as well as single-stranded DNA, and the former is also known to bind RNA with a reasonably high affinity (2, 10). We are therefore left with the question of whether Alba and proteins such as CC1 are bona fide ssDNA binding proteins in vivo. What is certain is that hyperthermophilic organisms, including *P. aerophilum* and *T. tenax*, are likely to require abundant, functional SSBs to protect ssDNA, which is much more susceptible to DNA damage than double-stranded DNA is. The functional screen described in this paper suggests that small chromatin proteins, such as Alba and CC1, are currently the best candidates for such a role in the *Thermoproteus-Pyrobaculum* lineage and may also contribute to the protection of ssDNA in other archaea in which a canonical SSB is present.

ACKNOWLEDGMENTS

We thank Jane Potter for help with CD spectroscopy, Thomas Knura for *T. tenax* cell growth, and Alex Houston and Sally Shirran for mass spectrometric assistance.

We thank the MRC for funding C.H.B. Mass spectrometry at the University of St. Andrews is funded by the Wellcome Trust.

REFERENCES

1. **Agback, P., H. Baumann, S. Knapp, R. Ladenstein, and T. Hard.** 1998. Architecture of nonspecific protein-DNA interactions in the Sso7d-DNA complex. Nat. Struct. Biol. **5:**579–584.
2. **Aravind, L., L. M. Iyer, and V. Anantharaman.** 2003. The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism. Genome Biol. **4:**R64.
3. **Brock, T. D., K. M. Brock, R. T. Belly, and R. L. Weiss.** 1972. Sulfolobus: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch. Mikrobiol. **84:**54–68.
4. **Brunner, N. A., and R. Hensel.** 2001. Nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase from Thermoproteus tenax. Methods Enzymol. **331:**117–131.
5. **Chartier, F., B. Laine, and P. Sautiere.** 1988. Characterization of the chromosomal protein MC1 from the thermophilic archaebacterium Methanosarcina sp. CHTI 55 and its effect on the thermal stability of DNA. Biochim. Biophys. Acta **951:**149–156.
6. **Chedin, F., E. M. Seitz, and S. C. Kowalczykowski.** 1998. Novel homologs of replication protein A in archaea: implications for the evolution of ssDNA-binding proteins. Trends Biochem. Sci. **23:**273–277.
7. **Cubonová, L., K. Sandman, S. J. Hallam, E. F. Delong, and J. N. Reeve.** 2005. Histones in crenarchaea. J. Bacteriol. **187:**5482–5485.
8. **Cuff, J. A., M. E. Clamp, A. S. Siddiqui, M. Finlay, and G. J. Barton.** 1998. JPred: a consensus secondary structure prediction server. Bioinformatics **14:**892–893.
9. **Fitz-Gibbon, S. T., H. Ladner, U. J. Kim, K. O. Stetter, M. I. Simon, and J. H. Miller.** 2002. Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum aerophilum. Proc. Natl. Acad. Sci. USA **99:**984–989.
10. **Guo, R., H. Xue, and L. Huang.** 2003. Ssh10b, a conserved thermophilic archaeal protein, binds RNA in vivo. Mol. Microbiol. **50:**1605–1615.
11. **Iftode, C., Y. Daniely, and J. A. Borowiec.** 1999. Replication protein A (RPA): the eukaryotic SSB. Crit. Rev. Biochem. Mol. Biol. **34:**141–180.
12. **Jelinska, C., M. J. Conroy, C. J. Craven, A. M. Hounslow, P. A. Bullough, J. P. Waltho, G. L. Taylor, and M. F. White.** 2005. Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA packaging. Structure **13:**963–971.
13. **Kerr, I. D., R. I. M. Wadsworth, L. Cubbedu, W. Blankenfeldt, J. H. Naismith, and M. F. White.** 2003. Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of Sulfolobus SSB. EMBO J. **22:**2561–2570.
14. **Lurz, R., M. Grote, J. Dijk, R. Reinhardt, and B. Dobrinski.** 1986. Electron microscopic study of DNA complexes with proteins from the Archaebacterium Sulfolobus acidocaldarius. EMBO J. **5:**3715–3721.
15. **Marsh, V. L., S. Y. Peak-Chew, and S. D. Bell.** 2005. Sir2 and the acetyl-

transferase, Pat, regulate the archael chromatin protein, Alba. J. Biol. Chem. **280:**21122–21128.
16. **Murzin, A. G.** 1993. OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. EMBO J. **12:**861–867.
17. **Paquet, F., F. Culard, F. Barbault, J. C. Maurizot, and G. Lancelot.** 2004. NMR solution structure of the archaebacterial chromosomal protein MC1 reveals a new protein fold. Biochemistry **43:**14971–14978.
18. **Robinson, H., Y. G. Gao, B. S. McCrary, S. P. Edmondson, J. W. Shriver, and A. H. Wang.** 1998. The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. Nature **392:**202–205.
19. **Sandman, K., and J. N. Reeve.** 2005. Archaeal chromatin proteins: different structures but common function? Curr. Opin. Microbiol. **8:**656–661.
20. **Sandman, K., D. Soares, and J. N. Reeve.** 2001. Molecular components of the archaeal nucleosome. Biochimie **83:**277–281.
21. **Wadsworth, R. I., and M. F. White.** 2001. Identification and properties of the crenarchaeal single-stranded DNA binding protein from *Sulfolobus solfataricus*. Nucleic Acids Res. **29:**914–920.
22. **White, M. F.** 2003. Archaeal DNA repair: paradigms and puzzles. Biochem. Soc. Trans. **31:**690–693.
23. **White, M. F., and S. D. Bell.** 2002. Holding it together: chromatin in the archaea. Trends Genet. **18:**621–626.
24. **Whitmore, L., and B. A. Wallace.** 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res. **32:**W668–W673.
25. **Wu, L. J.** 2004. Structure and segregation of the bacterial nucleoid. Curr. Opin. Genet. Dev. **14:**126–132.
26. **Wüthrich, K.** 1986. NMR of proteins and nucleic acids. Wiley, New York, NY.
27. **Zillig, W., K. O. Stetter, W. Schäffer, D. Janekovic, S. Wunderl, I. Holz, and P. Palm.** 1981. Thermoproteales: a novel type of extremely thermoacidophilic anaerobic archaebacteria isolated from Icelandic solfatares. Zentbl. Bakteriol. Hyg. Abt. 1 Orig. C **2:**205–227.