Published online 20 December 2007

Nucleic Acids Research, 2008, Vol. 36, No. 4 ***1129–1137***
doi:10.1093/nar/gkm1128

Biochemical and structural characterization of Cren7, a novel chromatin protein conserved among Crenarchaea

Li Guo[1], Yingang Feng[2], Zhenfeng Zhang[1], Hongwei Yao[2], Yuanming Luo[1], Jinfeng Wang[2] and Li Huang[1,*]

[1]State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, 3A Datun Road, Beijing 100101, P. R. China and [2]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, P.R. China

Received October 30, 2007; Revised December 3, 2007; Accepted December 4, 2007

ABSTRACT

Archaea contain a variety of chromatin proteins consistent with the evolution of different genome packaging mechanisms. Among the two main kingdoms in the Archaea, Euryarchaeota synthesize histone homologs, whereas Crenarchaeota have not been shown to possess a chromatin protein conserved at the kingdom level. We report the identification of Cren7, a novel family of chromatin proteins highly conserved in the Crenarchaeota. A small, basic, methylated and abundant protein, Cren7 displays a higher affinity for double-stranded DNA than for single-stranded DNA, constrains negative DNA supercoils and is associated with genomic DNA *in vivo*. The solution structure and DNA-binding surface of Cren7 from the hyperthermophilic crenarchaeon *Sulfolobus solfataricus* were determined by NMR. The protein adopts an SH3-like fold. It interacts with duplex DNA through a β-sheet and a long flexible loop, presumably resulting in DNA distortions through intercalation of conserved hydrophobic residues into the DNA structure. These data suggest that the crenarchaeal kingdom in the Archaea shares a common strategy in chromatin organization.

INTRODUCTION

All cells must package their genomic DNA into a small space while providing controlled access to DNA for replication, recombination, repair and gene expression. Life appears to have evolved different strategies for genome packaging. In Eukarya, DNA is wrapped around the histone core of nucleosome, the basic structural unit for DNA packaging in the chromatin (1). In bacteria, chromosomal DNA is folded into a compact structure called nucleoid. Although the details of DNA packaging in bacterial cells are still lacking, the general organization of bacterial chromatin seems quite different from that of eukaryotic chromatin. An array of nucleoid-associated proteins (e.g. HU, IHF, H-NS, Fis and Lrp) have been identified, but their contributions to the overall structure of the bacterial nucleoid remain unclear (2,3).

The situation is intriguingly complex in Archaea. The Euryarchaeota contain archaeal histones that share a common ancestry with the histone fold regions of the eukaryotic nucleosome core histones (4,5). Extensive biochemical studies have revealed that archaeal histones form structures analogous to the eukaryotic H3/H4 tetrasome (5). Presumably, archaeal histones contribute to chromatin compaction and accessibility *in vivo*. Interestingly, histones are absent from the Crenarchaeota, with the exception of some species, e.g. psychrophilic Crenarchaea from marine environments (6).

Efforts to understand DNA packaging in Crenarchaea have led to the isolation of a range of suspected chromatin proteins (e.g. Alba, CC1 and Sul7d) (7–9). However, all of these proteins, with the exception of Alba, appear to have a restricted phylogenetic distribution (4). Alba is highly conserved among both archaeal kingdoms (5). But, recent studies suggest that Alba binds to both DNA and RNA *in vivo* (10,11). So the physiological role of this protein family remains to be understood. CC1 denotes a family of DNA-binding proteins recently identified in the crenarchaeal orders *Thermoproteales* and *Desulfurococcales* (9). Since CC1 binds equally well to both dsDNA and ssDNA, it is probably not involved in DNA packaging and may

*To whom correspondence should be addressed. Tel: +86 10 64807430; Fax: +86 10 64807429; Email: huangl@sun.im.ac.cn.
Correspondence may also be addressed to Jinfeng Wang. Tel: +86 10 64888490; Fax: +86 10 64872026; Email: jfw@sun5.ibp.ac.cn.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

instead play a role in the protection of ssDNA in these organisms, which lack a canonical ssDNA-binding protein (SSB) (9). Sul7d is restricted to the order *Sulfolobales*. It includes a family of small, basic and abundant proteins that bind much more strongly to dsDNA than to ssDNA *in vitro* and are associated with the genomic DNA *in vivo* (11–13). These properties, together with the ability of the protein to constrain DNA supercoils and compact DNA, are consistent with a role for Sul7d in genome packaging. So far, no chromatin proteins have been found that are conserved among all phylogenetic branches of Crenarchaea. Whether Crenarchaea employ a conserved mechanism in chromatin organization is unknown. An answer to this question will not only shed light on the evolution of chromatin organization but also increase our knowledge about regulation of gene expression in Crenarchaea.

In the present study, we report the identification of a family of small and basic DNA-binding proteins, denoted Cren7, that are highly conserved in Crenarchaea. Biochemical and structural analyses show that Cren7 is a chromatin protein. Our results suggest that the majority of Crenarchaea share a common strategy in chromatin organization.

MATERIALS AND METHODS

Preparation of native and recombinant Cren7 proteins

Native Cren7 was purified from *Sulfolobus shibatae* using the purification protocol described previously for Ssh10b (14) with modifications. Recombinant Cren7 was prepared by amplifying the gene encoding the protein (SSO6901) from *Sulfolobus solfataricus* by PCR, cloning it into pET30a. The recombinant protein was subsequently overproduced in *Escherichia coli* strain Rosetta 2 (DE3) plysS and purified. Detailed purification procedures for native and recombinant Cren7 proteins are in Supplementary Data.

Protein identification

Identification of Cren7 was carried out by subjecting the purified protein to SDS–PAGE. The protein was digested in-gel with trypsin, and tryptic fragments were analyzed by LC-MS, as described (15).

Quantitative immunoblotting

Samples were taken from a growing *S. solfataricus* culture at various cell densities and centrifuged. Intracellular levels of Cren7 were measured by immunoblotting using anti-Cren7 antibodies, as described (11). Anti-Cren7 antibodies were raised in rabbit using purified recombinant Cren7.

Electrophoretic mobility shift assays (EMSA)

Radiolabeling of DNA fragments and EMSAs were performed as described previously (11). A 60 bp dsDNA fragment (5′-GATCCCCCAATGCTTCGTTTCGTATCA CACACCCCAAAGCCTTCTGCTTTGAATGCTGCC) and a 60 nt ssDNA fragment [poly(GACT)$_{15}$] were used in this study.

Thermal denaturation of dsDNA

Recombinant Cren7 was mixed with dsDNA (poly[dAdT]–poly[dAdT]) in 10 mM potassium phosphate (pH 7.0). Thermal denaturation of the DNA was measured at A_{260} in a temperature range from 25°C to 90°C on a DU800 UV/Visible spectrophotometer (Beckman). The melting temperature (T_m) was obtained using DU800 software.

Association of Cren7 with nucleic acids *in vivo*

The nature of nucleic acids bound by Cren7 in the cell was determined by chromatin fractionation and by *in vivo* UV-crosslinking followed by co-immunoprecipitation, as described previously (10,11).

Nick closure assays

Plasmid pBR322 containing a single nick per circular molecule was prepared by digestion with the nicking enzyme Nb.Bpu10I (Fermentas). The nicked plasmid (0.5 μg) was mixed with recombinant Cren7 in 20 mM Tris–Cl (pH7.5), 1 mM DTT and 100 μg/ml BSA and ligated with T4 DNA ligase (Fermentas) for 5 min at 25°C in a final volume of 15 μl. Samples were deproteinized and analyzed by agarose gel electrophoresis in the presence or absence of chroloquine, as described previously (16).

NMR spectroscopy

Cren7 were uniformly labeled with $^{15}N/^{13}C$ by growing the recombinant strain in M9-minimal medium containing $^{15}NH_4Cl$ and ^{13}C-glucose as the sole nitrogen and carbon sources, respectively. The labeled proteins were purified as described above. Samples for NMR measurements contained 1.0–2.0 mM Cren7, 90% H_2O/10% D_2O and 0.01% sodium 2,2-dimethylsilapentane-5-sulfonate (DSS) in 50-mM potassium phosphate buffer (pH 6.0).

All NMR experiments were carried out at 310 K on a Bruker DMX 600 MHz spectrometer with a triple resonance cryo-probe. NMR experiments for ^{1}H, ^{15}N, and ^{13}C resonance assignments included: 2D ^{1}H-^{1}H TOCSY, ^{1}H-^{1}H NOESY, ^{1}H-^{1}H DQF-COSY, ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC, 3D HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, HBHA(CBCA)CONH, ^{1}H-^{15}N TOCSY-HSQC, (H)CCH-TOCSY and HC(C)H-COSY. Four NOE experiments: 2D ^{1}H-^{1}H NOESY, 3D ^{1}H-^{15}N NOESY-HSQC and ^{1}H-^{13}C NOESY-HSQC for aliphatic and aromatic regions with mixing times of 200 ms were carried out to obtain the distance constraints. Dihedral constraints were generated using TALOS (17). All NMR spectra were processed and analyzed using Felix software (Accelrys Inc.). Proton chemical shifts and ^{15}N and ^{13}C chemical shifts were referenced to internal DSS and indirectly to DSS, respectively (18).

Initial structures of Cren7 were generated using CANDID module of CYANA software (19). The NOE assignments given by CANDID were checked manually and the structures were refined in explicit water using CNS software (20) and RECOORDScript (21). 100 structures

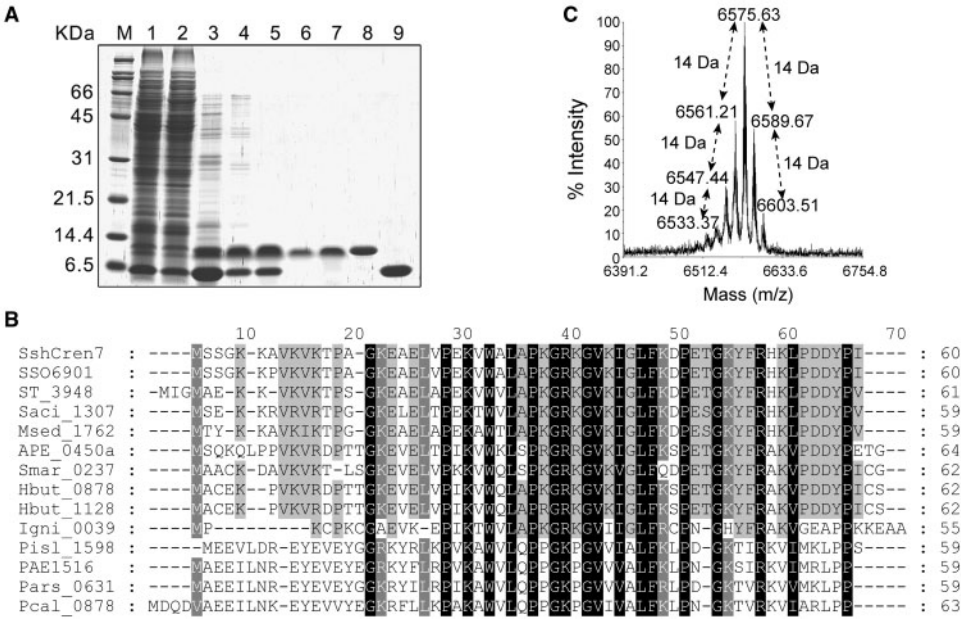

Figure 1. Identification of the Cren7 protein family in Crenarchaea. (**A**) Purification of the *S. shibatae* Cren7 protein. Lane 1, cell lysate; lane 2, ammonium sulfate fraction; lane 3, SP Sepharose peak fractions; lane 4, Resource S peak fractions; lane 5, heparin peak fractions; lane 6, flow-through fractions from anti-Ssh7 antibody affinity column; lane 7, recombinant SsoCren7; lane 8, native Ssh10b; lane 9, native Ssh7; M, molecular weight standards with molecular weights indicated. (**B**) Sequence alignment of Cren7 homologs in Crenarchaea. Sequences are from *Sulfolobus shibatae* (SshCren7), *Sulfolobus solfataricus* P2 (SSO6901), *Sulfolobus tokodaii* str.7 (ST_3948), *Sulfolobus acidocaldarius* DSM 639 (Saci_1307), *Metallosphaera sedula* DSM5348 (Msed_1762), *Aeropyrum pernix* K1(APE_0450a), *Staphylothermus marinus* F1 (Smar_0237), *Hyperthermus butylicus* DSM5456 (Hbut_0878 and Hbut_1128), *Ignicoccus hospitalis* KIN4/I(Igni_0039), *Pyrobaculum islandicum* DSM4184 (Pisl_1589), *Pyrobaculum aerophilum* str. IM2 (PAE1516), *Pyrobaculum arsenaticum* DSM13514 (Pars_0631) and *Pyrobaculum calidifontis* JCM 11548 (Pcal_0878). (**C**) MALDI-TOF MS spectrum of native Cren7 from *S. shibatae*. A cluster of peaks with a mass difference of 14 kDa between adjacent peaks are shown.

were calculated, and 20 structures with lowest energies were selected for final analysis. The quality of the structures was analyzed with MOLMOL (22) and PROCHECK-NMR (23). Structure figures were prepared using MOLMOL. Structural similarity was searched using Dali server (24) and SSM server (25).

Binding of Cren7 to a DNA duplex (5′-GCGAATT CGC) in 50-mM potassium phosphate buffer (pH 6.0) was monitored by ^{1}H-^{15}N HSQC using the ^{15}N-labeled protein (0.2 mM). The backbone assignments of the Cren7–dsDNA complex were obtained in CBCA(CO)NH and HNCA experiments, in which $^{15}N/^{13}C$-labeled Cren7 (0.2 mM) was mixed with a slight excess amount of dsDNA.

Data deposition

Chemical shift assignments of *S. solfataricus* Cren7 have been deposited in the BioMagResBank under an accession number 15415. The atomic coordinates of *S. solfataricus* Cren7 and all restraints have been deposited in the Protein Data Bank with accession code 2JTM.

RESULTS

Identification of Cren7

In the process of purifying Ssh7, a member of the Sul7d family from *S. shibatae*, we co-purified a small, abundant protein (Figure 1A). LC-MS analysis identified this protein as a homolog of the product of an unknown gene (SSO6901) in the *S. solfataricus* P2 genome. Sequence comparison shows that this novel protein is highly conserved among the crenarchaeotal lineage of the Archaea (Figure 1B). All genome-sequenced Crenarchaea except for *Thermofilum pendens* Hrk5 and *Cenarchaeum symbiosum* encode a homolog of this protein. It is present in the orders *Sulfolobales*, *Thermoproteales* and *Desulfurococcales*. Since it was later shown to be a chromatin protein, we designate this novel protein family as Cren7 following the nomenclature used for Sul7d.

Sulfolobus shibatae Cren7 contains 60 amino acid residues (theoretical M_r: 6637) and is highly basic (calculated pI: 9.87). The protein has 22 charged residues including 12 lysines and 2 arginines. N-terminal microsequencing revealed the lack of the initiator methionine in the protein. The native protein appeared to be methylated at multiple sites since a cluster of peaks differing by 14 kDa between adjacent peaks were detected by MALDI-TOF-TOF in the protein sample (Figure 1C). Mass fingerprinting showed that the protein was methylated at these sites to various extents (Supplementary Figures S1 and S2 and Supplementary Table S1). Cren7 exists as a monomer in solution and is stable at high temperature (Supplementary Data, Supplementary Figure S3).

Cren7 is a chromatin protein

We first investigated the interaction of Cren7 with DNA by EMSA. Cren7–DNA complexes initially generated

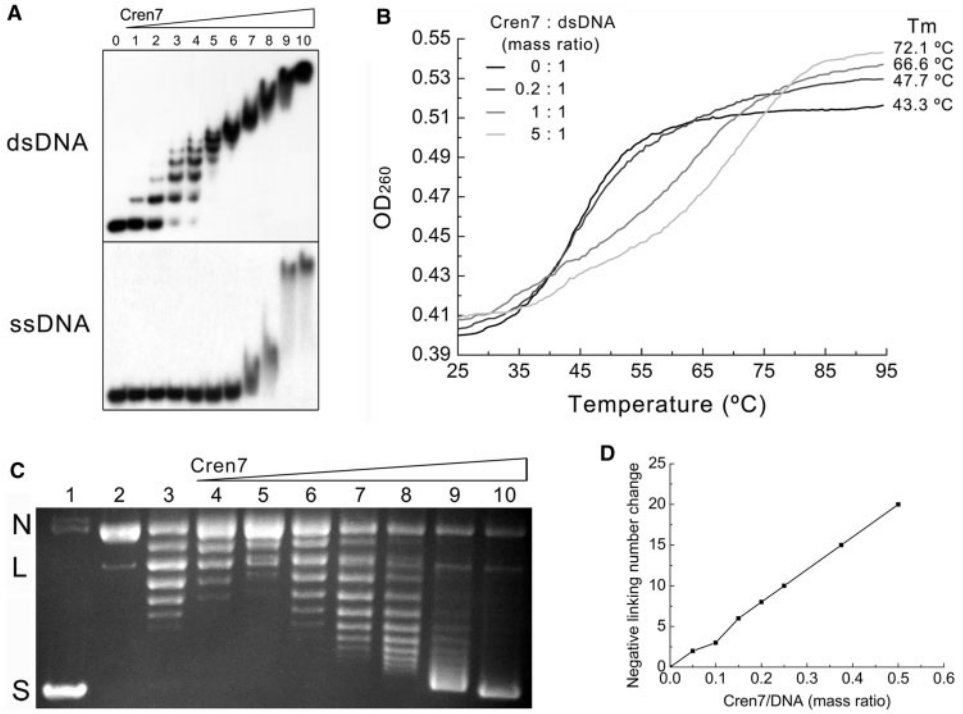

Figure 2. Interaction of Cren7 with DNA *in vitro*. (**A**) Comparison of binding of Cren7 to dsDNA and ssDNA. Recombinant Cren7 was incubated with a radiolabeled 60 bp dsDNA or 60 nt ssDNA. Protein–DNA complexes were subjected to electrophoresis in polyacrylamide. Protein concentrations were 0, 0.02, 0.04, 0.08, 0.16, 0.31, 0.63, 1.3, 2.5, 5 and 10 μM, respectively. (**B**) Effect of binding by Cren7 on thermal stability of dsDNA. Thermal denaturation of poly[dAdT]–poly[dAdT] in the presence and absence of Cren7 was determined by monitoring changes in UV absorbance at 260 nm. (**C**) Nick closure analysis of the ability of Cren7 to constrain DNA supercoils. Single-nicked plasmid pBR322 was incubated with Cren7 at various protein/DNA mass ratios and ligated with T4 DNA ligase at 25°C. Samples were deproteinized and subjected to agarose gel electrophoresis. Lane 1, negatively supercoiled pBR322; lane 2, single-nicked pBR322; lanes 3–10, topoisomers of pBR322 ligated in the presence of Cren7 at the following protein/DNA mass ratios: 0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.375 and 0.5, respectively. N, nicked circular plasmid; L, linear plasmid; S, supercoiled plasmid. (**D**) Plot of the linking number change of nick-closed plasmid against the Cren7/DNA mass ratio. The linking change of pBR322 ligated in the presence or absence of Cren7 was measured by resolving topoisomers on agarose gels in the presence or absence of chloroquine and band counting.

resolvable shifts and, after a binding ratio of ~10 bp per protein molecule was reached, formed increasingly slow-migrating smears as protein concentration increased (Figure 2A). Therefore, the gel retardation pattern of Cren7 was biphasic, with a low-binding density phase and a high-binding density one. Cren7 probably bound to DNA with a binding site size of ~10 bp in the low-binding density phase, but switched to a new binding mode in the high-binding density phase. This new binding mode presumably involved a change in protein–DNA and/or protein–protein contacts. The inability to resolve large Cren7–DNA complexes formed at high protein concentrations by EMSA prevented an accurate determination, based on the relationship between the maximum number of shifts and the size of the dsDNA fragment, of the binding site size of the protein in the high-density-binding phase. However, judging from the lowest protein/DNA ratio at which plasmid pBR322 was maximally retarded by Cren7 in an agarose gel, we obtained a binding density of ~3 bp per protein molecule for the high-density-binding mode (data not shown). In comparison, no biphasic gel retardation pattern was observed for Sul7d, which has a binding site size of 3.3–4.6 bp (13,26). Both native *S. shibatae* Cren7 and recombinant *S. solfataricus* Cren7 showed similar affinities for the dsDNA fragment with apparent dissociation constants (K_D) between 0.08 and 0.16 μM under low salt conditions (data not shown). Therefore, the binding affinity of Cren7 for DNA was not affected by the post-translational modifications of the protein. In addition, the Cren7 proteins from the two *Sulfolobus* species showed no detectable differences in DNA binding, as expected from the fact that the two proteins are nearly identical, differing only at one amino acid position (A7 in the former and P7 in the latter) in a nonconserved region (Figure 1B). We also found that the protein showed little sequence specificity in DNA binding since similar gel retardation patterns were generated when DNA fragments of different sequences were used (data not shown).

Like Sul7d, Cren7 bound only weakly to ssDNA (apparent K_D = ~2.5 μM) in low salt, and generated no resolvable shifts on a labeled ssDNA (Figure 2A) (12,13), indicating that Cren7 binds preferentially to dsDNA over ssDNA. This observation suggests that Cren7 might affect the T_m of dsDNA. Indeed, we found that the T_m of poly[dAdT]–poly[dAdT] was raised by 23.3°C and 28.8°C, respectively, in the presence of Cren7 at protein/DNA mass ratios of 1 and 5 (Figure 2B). So, binding by Cren7 significantly increased the stability of dsDNA against thermal denaturation.

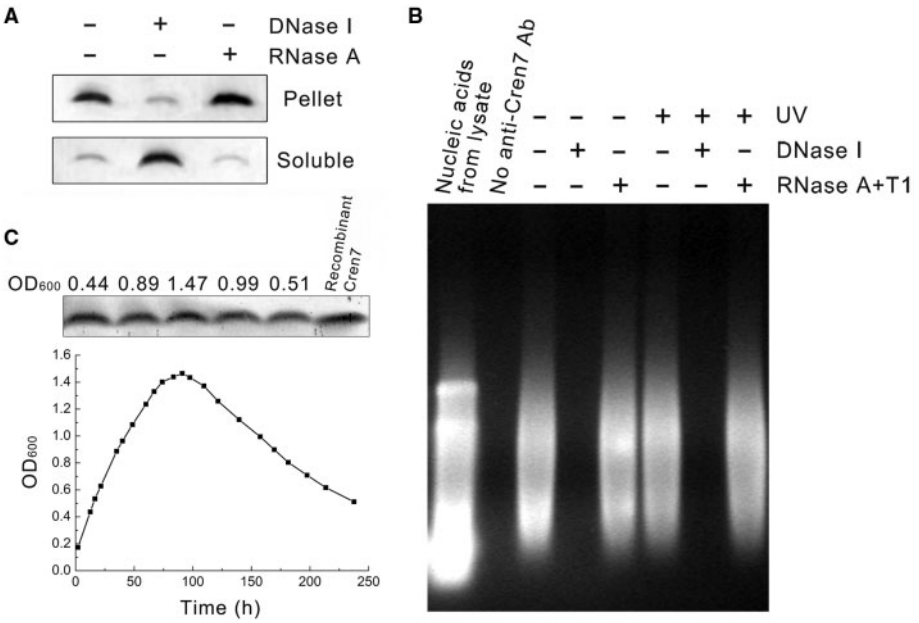

Figure 3. Identification of the binding target of Cren7 in the cell. (**A**) Chromosomal fractionation. *S. solfataricus* cells were resuspended in Bugbuster™ protein extraction reagent. Samples were treated with RNase A or DNase I. After centrifugation, the pellets were resuspended and subjected to polyacrylamide gel electrophoresis. Cren7 was detected by immunoblotting using anti-Cren7 antibodies. (**B**) Co-immunoprecipitation. *S. solfataricus* cells were irradiated with UV light or left untreated. After cell lysis, extracts were subjected to co-immunoprecipitation with anti-Cren7 antibodies. Nucleic acids co-precipitated with Cren7 were treated with either DNase I or RNase A + T1, and electrophoresed in agarose. (**C**) Abundance of Cren7 in *S. solfataricus*. *S. solfataricus* was grown at 78°C. Samples were taken at different stages during the growth, centrifuged and subjected to SDS–PAGE. Cren7 was detected by immunoblotting using anti-Cren7 antibodies.

We then determined if binding by Cren7 affected the geometry of DNA using a nick closure assay (16). A single-nicked plasmid became supercoiled when ligated at 25°C in the presence of the protein (Figure 2C). Comparison of the mobilities of plasmid topoisomers in agarose gels in the presence and absence of chloroquine indicated that the protein constrained the DNA in negative supercoils (data not shown). The number of constrained supercoils was proportional to the protein/DNA mass ratio in the tested range (Figure 2D). Binding of ~10 protein molecules resulted in one negative supercoil being constrained. The ability of the protein to constrain negative DNA supercoils at 45°C, 60°C and 80°C did not differ measurably from that at 25°C (data not shown). Taken together, Cren7 is about twice as efficient as, but otherwise resembles, Sul7d in constraining DNA supercoils (16,27).

As a small, basic and methylated protein capable of efficient binding to dsDNA and constraining negatively DNA supercoils, Cren7 may be a chromatin protein *in vivo*. To verify this possibility, we examined the association of Cren7 with nucleic acids in the *S. solfataricus* cell by using two approaches. In the first approach, we performed chromatin fractionation as described previously (10). As shown in Figure 3A, the bulk of Cren7 was in the insoluble chromatin-containing fraction. Treatment of the pellet with DNase I, but not RNase A, released a significant proportion of Cren7, indicating that the protein was associated with chromosomal DNA *in vivo*. In the second approach, we irradiated *S. solfataricus* cells with UV light to induce the formation of protein–nucleic acid cross-links, lyzed the cells and subjected the lysates to immunoprecipitation with antibodies against Cren7. Nucleic acids co-immunoprecipitated with Cren7 were treated with either DNase I or RNase A + T1. In both cross-linked and untreated samples, DNA, but not RNA, was co-immunoprecipitated with Cren7 (Figure 3B). These data indicate that Cren7 is a chromatin DNA-binding protein.

A protein serving a structural role in chromatin usually exists in abundance. So we measured the cellular level of *S. solfataricus* Cren7 by quantitative immunoblotting. Samples taken from various stages during the growth of *S. solfataricus* appeared to contain similar amounts of Cren7 (Figure 3C). The protein constituted ~1% of the total cellular protein. Based on this estimate, the mass ratio of Cren7 to genomic DNA was ~0.8 *in vivo* (16). This number translates into one Cren7 molecule per 12 bp of DNA. In other words, Cren7 binds ~1/4 of the genomic DNA in the cell. By comparison, Ssh7 amounts to 4.8% of the cellular protein in *S. shibatae* (16), and HmfA/HmfB are ~1% of the cellular protein in *Methanothermus fervidus* (28).

Solution structure and DNA-binding surface of Cren7

The solution structure of Cren7 was determined on the basis of 1623 NOE-derived distance restraints, 72 dihedral constraints and 30 hydrogen bond restraints derived from multidimensional NMR spectroscopy (Supplementary Table S2). It has a 'β-barrel' structure including two antiparallel β-sheets. β-sheet 1 is composed of two β-strands (β1: 8–11 and β2: 17–20) while β-sheet 2 contains three

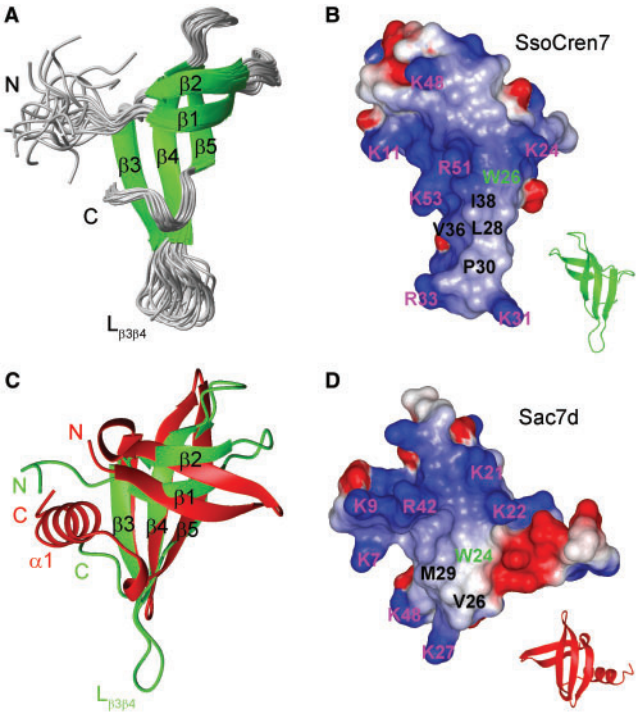

Figure 4. Overall solution structure of *S. solfataricus* Cren7. (**A**) Ribbon representation of 20 Cren7 solution structures. The backbone of five β-strands are shown in green and labeled. (**B**) Molecular surface of Cren7 colored by electrostatic potential. Blue, positively charged; red, negatively charged; white, neutral. Some charged or hydrophobic residues on the molecular surface are labeled. A ribbon representation is also shown for reference. (**C**) Structure alignment of Cren7 (in green) and Sac7d (in red, PDB code: 1AZP) by SSM. (**D**) Molecular surface of Sac7d colored by electrostatic potential. For comparison, the Sac7d protein is aligned in the same orientation as Cren7 in Figure 4B.

β-strands (β3: 24–28, β4: 36–42, and β5: 49–53) (Figure 4A). The side-chains of hydrophobic residues V8, V10, A18, L20, V25, F41 and F50 of the five β-strands make a tight packing within the 'β-barrel'. The C-terminal region (L54-I60) also participates in the hydrophobic packing through residues L54 and Y58 and, therefore, is structurally ordered. On the other hand, the N-terminal region (M1-K6) of the protein is unstructured. Among the four loops linking the five β-strands, loop $L_{β3β4}$ (A29-G35) is most flexible. These structural features of the protein are supported by dynamic data (Supplementary Data and Supplementary Figure S4). Side chains of L28 in β3, V36 and I38 in β4 and P30 in loop $L_{β3β4}$ form a notable solvent-exposed hydrophobic region (Figure 4B). Side chains of K11, K24, K48, R51 and K53 from the β-strands, and K31 and R33 from loop $L_{β3β4}$ face outside of the 'β-barrel', forming a positively charged region on the electrostatic surface of the protein (Figure 4B). Most of these residues are highly conserved, indicating that proteins of the Cren7 family have similar structures as well as surface charge distributions (Figure 1B).

DNA-binding surface of Cren7 was determined by monitoring signal changes in two-dimensional ^{1}H-^{15}N HSQC spectra following titration of the protein with a 10-bp duplex DNA (Supplementary Figure S5). As the DNA was titrated into the mixture, changes were observed for the cross peaks of residues involved in DNA binding. No further changes in the HSQC spectra occurred at DNA/protein molar ratios >1, suggesting a stoichiometry of one for the Cren7–dsDNA complex (Supplementary Data). Approximately one-third of the residues in Cren7 showed significant chemical shift changes in the HSQC spectra during dsDNA titration, indicating the presence of a large DNA-binding surface on the protein. In addition to the chemical shift changes, two cross peaks for side chains of R33 (on loop $L_{β3β4}$) and R51 (in β5), which were not observed in the HSQC spectrum of free protein because of amide-water proton exchange, appeared in the spectrum at the saturating level of dsDNA. Therefore, the two Arg residues are located at the binding interface of the protein–DNA interaction, and the amide-water proton exchange was prevented by bound dsDNA. The binding surface, as obtained by mapping these residues on the structure of Cren7, spans residues on strands β3, β4, β5, the C-terminus of β1 and loop $L_{β3β4}$ (Figure 5A–C). It consists mainly of the positively charged surface region and the solvent-exposed hydrophobic region. It was noted that the conformational flexibility of loop $L_{β3β4}$ was significantly reduced upon binding of the protein to DNA, as evidenced by a significant increase in ^{1}H-^{15}N NOE value (Supplementary Figure S4).

Dali and SSM searches show that the structure of Cren7 is a SH3-like fold. Although this fold was originally identified in proteins that bind poly-proline peptides, it has since been found in proteins with a number of other functions (29,30). A typical SH3 fold contains five β-strands and a short 3_{10}-helix between β4 and β5. Cren7 lacks both the C-terminal β5 and the 3_{10}-helix, but has a structured loop at the C-terminus. The other β-strands of Cren7 are similar to SH3, except that the β2 found in a typical SH3 fold is split into two sequential β-strands in Cren7. Interestingly, Sac7d, a Sul7d protein from *Sulfolobus acidocaldarius*, is the most relevant hit among those identified SH3-like proteins in the searches although the two proteins share no significant similarity at the amino-acid sequence level. The RMSD between Cren7 and Sac7d aligned by Dali is 2.6 Å (with a *Z*-score of 2.8), indicating a high structural similarity between the two proteins. Among 60 residues in Cren7 and 66 residues in Sac7d, 44 are structurally aligned by Dali. The 'β-barrel' parts of the two proteins are aligned best (Figures 4C and 5A). The two proteins also have similar solvent-exposed hydrophobic regions and positively charged regions on β-sheet 2, which form the DNA-binding surfaces (Figure 4B and C). A number of residues on the DNA-binding surfaces of the two proteins are conserved or similar. These include positively charged residues (K11, K24 and R51), hydrophobic residues (L28, V36 and L40) and a tryptophan residue (W26) in Cren7 (Figure 5D and E). However, several notable structural differences exist between Cren7 and Sac7d. At the DNA-binding surface, Cren7 has a long and highly flexible loop ($L_{β3β4}$), whereas only a small hinge exists at the corresponding position in Sac7d. In addition, the solvent-exposed hydrophobic region of Cren7 (L28, P30, V36, I38) is larger than that of Sul7d (V26, M29). Cren7 has a flexible N-terminal

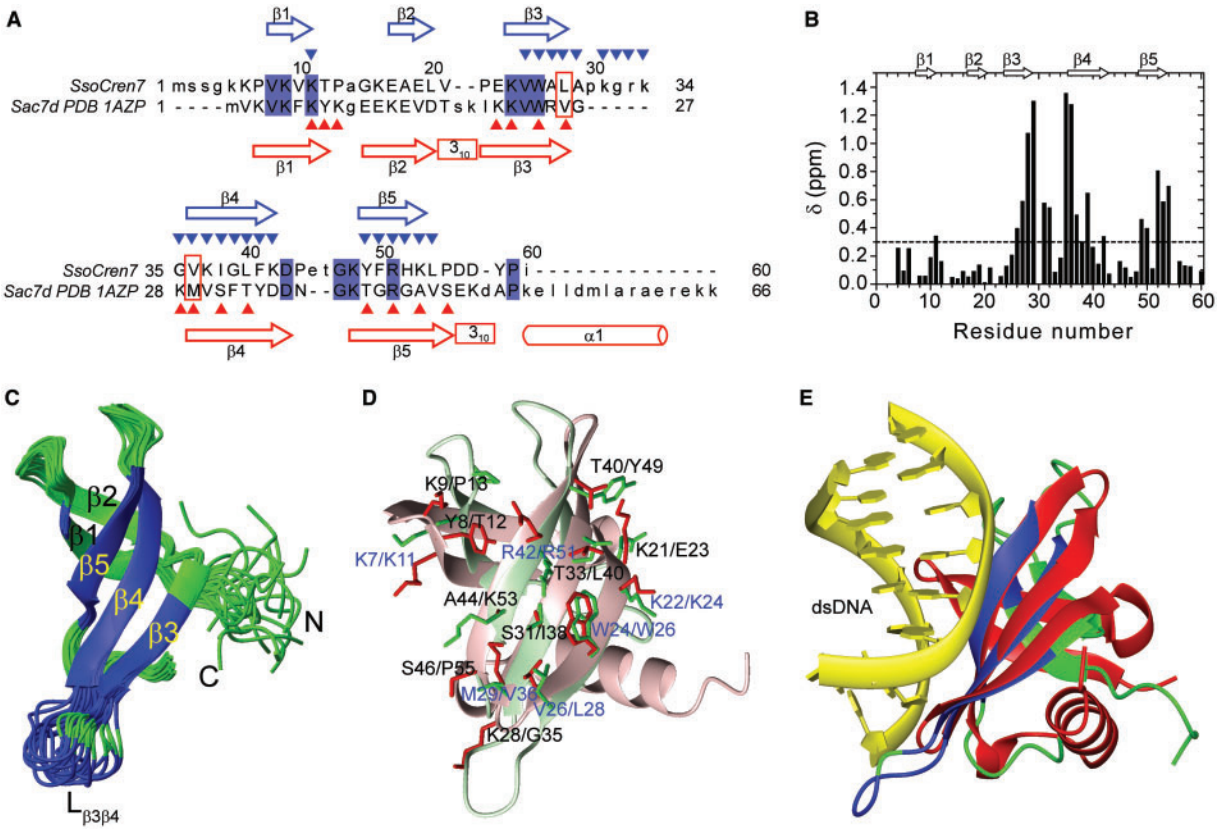

Figure 5. DNA-binding surface of Cren7. (**A**) Structure-based sequence alignment of Cren7 and Sac7d. Red triangles represent the DNA-binding residues of Sac7d; blue triangles indicate the possible binding residues inferred from the chemical shift changes caused by the binding of dsDNA. Residues conserved between the two proteins are shown in blue. DNA-intercalating residues in Sac7d and the corresponding residues in Cren7 are in red box. The secondary structural elements of the two proteins are also shown. Cren7, blue; Sac7d, red. (**B**) ^{1}H-^{15}N combined chemical shift differences (δ) between dsDNA-bound Cren7 and free Cren7. $\delta = (\delta_{HN}^2 + \delta_N^2/6)^{1/2}$, where δ_{HN} and δ_N are the chemical shift differences of amide hydrogen and nitrogen atoms, respectively. The secondary structural elements are indicated above the histogram. (**C**) Mapping of the dsDNA-binding surface of Cren7. Residues with significant chemical shift differences or peak intensity changes during the titration are shown in blue. (**D**) Comparison of DNA-binding sites of Cren7 and Sac7d. Side-chains of the DNA-interacting residues in Sac7d are displayed in red and those of the corresponding residues of Cren7 are in green. **(E)** Cren7 is aligned with Sac7d in the Sac7d–dsDNA complex. Cren7, green and blue, as in Figure 5C; Sac7d, red.

loop which is longer than that of Sac7d. On the other hand, the C-terminal segment of Sac7d is 15 residues longer than that of Cren7 and forms a helix which is absent in Cren7.

DISCUSSION

In contrast to the wide distribution of archaeal histones in the Euryarchaeota, no chromatin proteins conserved at the kingdom level in and unique to the Crenarchaeota have been found. The discrepancy between the two main archaeal lineages in this regard is puzzling. Identification of Cren7, a highly conserved protein family in Crenarchaea, has offered important clues to the issue. Biochemical and structural studies reported in the present work suggest that Cren7 is a chromatin protein. Crenarchaea whose genome has been sequenced so far, encode Cren7 homologs with only two known exceptions. The conservation of Cren7 among Crenarchaea points strongly to the possibility that this archaeal lineage shares a common feature in genome packaging and/or regulation of gene expression. The Crenarchaea known to lack this protein are *T. pendens* Hrk5 and *C. symbiosum*. Intriguingly and, probably not incidentally, both of these organisms encode an archaeal histone (5,31). Therefore, it is tempting to speculate that Cren7 plays a role in Crenarchaea similar to that of archaeal histones in Euryarchaea. Clearly, this hypothesis needs to be tested with more crenarchaeal genomes. It is also of interest to find out if there are Archaea that contain neither Cren7 nor histones.

An interesting finding of the present study is that Cren7 bears significant resemblance in structure to Sul7d, whose structure has been extensively investigated (12,26,32,33), although the two proteins are unrelated at the amino acid sequence level. Comparative analyses of the similarities and the differences between the two proteins in amino acid sequence, structure and DNA-binding properties may offer valuable insight into the mechanisms of DNA binding and distortion by Cren7. It has been shown that binding by Sac7d produces, through intercalation of the side chains of V26 and M29 (26,34), a sharp single-step kink (~60°) associated with helix unwinding in dsDNA, which induces negative DNA supercoiling (32,35).

The two intercalating amino acid residues are substituted for by two conserved hydrophobic residues (L28 and V36) in the Cren7 family (Figure 1B), which appear to be properly positioned in the protein structure (Figure 5D) to serve a role similar to that of V26 and M29 in Sac7d. So it is likely that Cren7 also causes a sharp kink in bound dsDNA. Notably, however, Cren7 differed from Sac7d in the ability to induce DNA distortions. In our topological assays, Cren7 was significantly more efficient than Sul7d in constraining negative DNA supercoils, suggesting that the former is probably able to induce greater helix unwinding and larger kinks in DNA than the latter. This difference may be due to the presence of a longer flexible loop ($L_{\beta3\beta4}$) and a larger solvent-exposed hydrophobic patch at the DNA-binding surface in Cren7 than in Sac7d.

Additional major structural differences between Cren7 and Sul7d are present in the terminal regions of the two proteins. Cren7 lacks the C-terminal α-helix present in Sac7d but possesses a flexible N-terminal loop absent in Sac7d. Both of these terminal regions are located opposite to the DNA-binding face of the protein and, therefore, are likely involved in protein–protein interactions. This notion is consistent with the finding that Cren7 was capable of forming large complexes with duplex DNA fragments at higher protein concentrations.

The presence of long and flexible loop $L_{\beta3\beta4}$ and N-terminal tail in Cren7, but not in Sac7d, also suggests that the interaction of the two proteins with DNA may be regulated differently. Residue K31 on loop $L\beta_3\beta_4$ in Cren7 is partially methylated, and the N-terminal tail of Cren7 contains two serine and two lysine residues, which are often sites of phosphorylation and methylation or acetylation, respectively, in eukaryotic histones (36–38). Taken together, the ability of Cren7 to constrain negative supercoils, form large complexes with dsDNA and undergo potential regulatory modifications is consistent with a role for the protein in chromosomal organization.

Archaea of the genus *Sulfolobus* synthesize copious amounts of both Cren7 and Sul7d. It seems unlikely that either protein is an evolutionary fossil. So a question arises as to how the two proteins are evolutionarily and functionally related. Presumably, both proteins serve important roles in chromosomal organization and their functions are not entirely redundant. The differences between the two proteins in function may have led to the loss in sequence similarity between them. Given its wide distribution, Cren7 is probably the best candidate for a major chromatin protein in Crenarchaea. The remarkable ability of the protein to distort DNA conformation is consistent with a role in DNA supercoiling and compaction. So our results reinforce the suggestion that Archaea employ two major strategies in chromatin compaction (26,33): one involves archaeal histones and the other uses the minor-groove binding proteins such as Cren7. While the histone-based strategy is primarily employed by the Euryarchaeota, the Cren7-based strategy may predominate in the Crenarchaeota. In *Sulfolobus* species, Cren7 and Sul7d (and, possibly, other unidentified chromatin proteins) may participate in chromosomal organization by forming functional complexes and/or interacting in a coordinated manner with DNA. The role of Sul7d is probably served by functional equivalents in other Crenarchaea.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Danxia Yu for help in recombinant protein preparation. This work was supported by National Basic Research Program of China (2004CB719603 to L.H.); National Natural Science Foundation of China (30470025 and 30621005 to L.H, 30570375 to J.W.); Chinese Academy of Sciences (KSCX2-YW-G-023 to L.H., KSCX1-SW-17 to J.W.). Funding to pay the Open Access publication charges for this article was provided by National Natural Science Foundation of China (30470025).

Conflict of interest statement. None declared.

REFERENCES

1. Luger,K., Mader,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. *Nature*, **389**, 251–260.
2. Wu,L.J. (2004) Structure and segregation of the bacterial nucleoid. *Curr. Opin. Genet. Dev.*, **14**, 126–132.
3. Dame,R.T. (2005) The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol. Microbiol.*, **56**, 858–870.
4. Reeve,J.N. (2003) Archaeal chromatin and transcription. *Mol. Microbiol.*, **48**, 587–598.
5. Sandman,K. and Reeve,J.N. (2005) Archaeal chromation proteins: different structures but common function? *Curr. Opin. Microbiol.*, **8**, 656–661.
6. Cubonova,L., Sandman,K., Hallam,S.J., DeLong,E.F. and Reeve,J.N. (2005) Histones in Cenarchaea. *J. Bacteriol.*, **187**, 5482–5485.
7. Grote,M., Dijk,J. and Reinhardt,R. (1986) Ribosomal and DNA binding proteins of the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius*. *Biochem. Biophys. Acta*, **873**, 405–413.
8. Bell,S.D., Botting,C.H., Wardleworth,B.N., Jackson,S.P. and White,M.F. (2002) The interaction of Alba, a conserved archaeal, chromatin protein, with Sir2 and its regulation by acetylation. *Science*, **296**, 148–151.
9. Luo,X., Schwarz-Linek,U., Botting,C.H., Hensel,R., Siebers,B. and White,M.F. (2007) CC1, a novel crenarchaeal DNA binding protein. *J. Bacteriol.*, **189**, 403–409.
10. Marsh,V.L., Peak-Chew,S.Y. and Bell,S.D. (2005) Sir2 and the acetyltransferase, Pat, regulate the archaeal chromatin protein, Alba. *J. Biol. Chem.*, **280**, 21122–21128.
11. Guo,R., Xue,H. and Huang,L. (2003) Ssh10b, a conserved thermophilic archaeal protein, binds RNA in vivo. *Mol. Microbiol.*, **50**, 1605–1615.
12. 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.
13. Chen,X.L., Guo,R., Huang,L. and Hong,R. (2002) Evolutionary conservation and DNA binding properties of the Ssh7 proteins from *Sulfolobus shibatae*. *Sci. China Ser. C-Life Sci.*, **45**, 583–592.
14. Xue,H., Guo,R., Wen,Y.F., Liu,D.X. and Huang,L. (2000) An abundant DNA binding protein from the hyperthermophilic archaeon *Sulfolobus shibatae* affects DNA supercoiling in a temperature-dependent fashion. *J. Bacteriol.*, **182**, 3929–3933.
15. Xiang,X.Y., Chen,L.M., Huang,X.X., Luo,Y.M., She,Q.X. and Huang,L. (2005) *Sulfolobus tengchongensis* spindle-shaped virus

STSV1: Virus-host interactions and genomic features. *J. Virol.*, **79**, 8677–8686.
16. Mai,V.Q., Chen,X.L., Hong,R. and Huang,L. (1998) Small abundant DNA binding proteins from the thermoacidophilic archaeon *Sulfolobus shibatae* constrain negative DNA supercoils. *J. Bacteriol.*, **180**, 2560–2563.
17. Cornilescu,G., Delaglio,F. and Bax,A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
18. Markley,J.L., Bax,A., Arata,Y., Hilbers,C.W., Kaptein,R., Sykes,B.D., Wright,P.E. and Wuthrich,K. (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids - (IUPAC Recommendations 1998). *Pure Appl. Chem.*, **70**, 117–142.
19. Herrmann,T., Guntert,P. and Wuthrich,K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.*, **319**, 209–227.
20. Brunger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D-Biol. Crystallogr.*, **54**, 905–921.
21. Nederveen,A.J., Doreleijers,J.F., Vranken,W., Miller,Z., Spronk,C.A.E.M., Nabuurs,S.B., Guntert,P., Livny,M., Markley,J.L. *et al.* (2005) RECOORD: A recalculated coordinate database of 500 + proteins from the PDB using restraints from the BioMagResBank. *Proteins*, **59**, 662–672.
22. Koradi,R., Billeter,M. and Wuthrich,K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 51–55.
23. Laskowski,R.A., Rullmann,J.A.C., 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.
24. Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.
25. Krissinel,E. and Henrick,K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D-Biol. Crystallogr.*, **60**, 2256–2268.
26. Robinson,H., Gao,Y.G., McCrary,B.S., Edmondson,S.P., Shriver,J.W. and Wang,A.H.J. (1998) The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature*, **392**, 202–205.
27. Lopez-Garcia,P., Knapp,S., Ladenstein,R. and Forterre,P. (1998) In vitro DNA binding of the archaeal protein Sso7d induces negative supercoiling at temperatures typical for thermophilic growth. *Nucleic Acids Res.*, **26**, 2322–2328.
28. Sandman,K., Krzycki,J.A., Dobrinski,B., Lurz,R. and Reeve,J.N. (1990) HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. *Proc. Natl Acad. Sci. USA*, **87**, 5788–5791.
29. Dalgarno,D.C., Botfield,M.C. and Rickles,R.J. (1997) SH3 domains and drug design: ligands, structure, and biological function. *Biopolymers*, **43**, 383–400.
30. Kishan,K.V.R. and Agrawal,V. (2005) SH3-like fold proteins are structurally conserved and functionally divergent. *Curr. Protein Pept. Sci.*, **6**, 143–150.
31. Sandman,K. and Reeve,J.N. (2006) Archaeal histones and the origin of the histone fold. *Curr. Opin. Microbiol.*, **9**, 520–525.
32. 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.
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.J. (1998) The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA. *Nat. Struct. Biol.*, **5**, 782–786.
34. Chen,C.Y., Ko,T.P., Lin,T.W., Chou,C.C., Chen,C.J. and Wang,A.H.J. (2005) Probing the DNA kink structure induced by the hyperthermophilic chromosomal protein Sac7d. *Nucleic Acids Res.*, **33**, 430–438.
35. Napoli,A., Zivanovic,Y., Bocs,C., Buhler,C., Rossi,M., Forterre,P. and Ciaramella,M. (2002) DNA bending, compaction and negative supercoiling by the architectural protein Sso7d of *Sulfolobus solfataricus*. *Nucleic Acids Res.*, **30**, 2656–2662.
36. Nowak,S.J. and Corces,V.G. (2004) Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. *Trends Genet.*, **20**, 214–220.
37. Hansen,J.C., Lu,X., Ross,E.D. and Woody,R.W. (2006) Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J. Biol. Chem.*, **281**, 1853–1856.
38. Grant,P.A. (2001) A tale of histone modifications. *Genome Biol.*, **2**, REVIEWS0003.