Extremophiles (2008) 12:235–246
DOI 10.1007/s00792-007-0120-z

ORIGINAL PAPER

Biochemical characterization of DNA-binding proteins from *Pyrobaculum aerophilum* and *Aeropyrum pernix*

Christine D. Hardy · Patrick K. Martin

Received: 13 September 2007 / Accepted: 24 October 2007 / Published online: 7 December 2007

Abstract Several representatives of the Crenarchaeal branch of the Archaea contain highly abundant, small, positively charged proteins exemplified by the Sso7d protein from *Sulfolobus solfataricus*. These proteins bind to DNA in a non-sequence-specific manner. Using publicly available genomic sequence information, we identified a second class of small Crenarchaeal DNA-binding proteins represented by the *Pyrobaculum aerophilum* open reading frame 3192–encoded (Pae3192) protein and its paralogs. We investigated the biochemical properties of the Pae3192 protein and an orthologous protein (Ape1322b) from *Aeropyrum pernix* in side-by-side experiments with the Sso7d protein. We demonstrate that the recombinant Ape1322b, Pae3192 and Sso7d proteins bind to DNA and that the DNA-protein complexes formed are slightly different for each protein. We show that like Sso7d, Pae3192 constrains negative supercoils in DNA. In addition, we show that all three proteins raise the melting temperature of duplex DNA upon binding. Finally, we present the equilibrium affinity constants and kinetic association constants of each protein for single-stranded and double-stranded DNA.

Keywords Archaea · Crenarchaea · Chromatin · DNA-binding protein · Thermostable · Hyperthermophile

Communicated by J. N. Reeve.

C. D. Hardy (✉) · P. K. Martin
Applied Biosystems, Inc., 850 Lincoln Centre Drive,
MS 404-1, Foster City, CA 94404, USA
e-mail: christine.hardy@appliedbiosystems.com

Introduction

Many members of the Euryarchaeal branch of the Archaea contain proteins with notable similarity to eukaryotic histones (Reeve et al. 2004). These archaeal histones are generally not found in the second, more ancient branch of the Archaea, the Crenarchaea. Instead, several Crenarchaeal species contain abundant, low molecular weight DNA-binding proteins belonging to the Sso7d family. The relative abundance of Sso7d-family proteins in cellular extracts (Grote et al. 1986; Kimura et al. 1984), as well as their observed low degree of sequence preference (Lundback et al. 1998), suggests that these proteins fulfill a similar, if not identical, function to that of the archaeal histones.

The identification of the first Sso7d-family protein was reported over 20 years ago (Kimura et al. 1984). Five different proteins, Sac7a-7e, were originally purified from cellular extracts of *Sulfolobus acidocaldarius* (Grote et al. 1986), but only two corresponding genes were identified by experimental methods and in the completed genome sequence (McAfee et al. 1995; Chen et al. 2005). The additional three proteins arise as the result of distinct post-translational modifications (Choli et al. 1988b). Orthologs of the *S. acidocaldarius* proteins were also isolated from *S. solfataricus* (Choli et al. 1988a). These proteins (Sso7a-7e) are encoded by three discrete, yet almost identical, open reading frames (ORFs) in the genome of this organism (She et al. 2001). The original mechanism for the generation of these apparent gene duplications at separate loci in the chromosome is unclear. The presence of both transposase and plasmid-like gene sequences flanking the Sso7d coding regions in the chromosome suggests that they exist as a result of multiple imperfect plasmid integration/excision events (CDH and PKM, unpublished).

The Sso7d family is composed of basic proteins (pI ~10) that bind to DNA without apparent sequence bias and are rich in lysine residues, some of which have been shown to be variably mono-methylated in vivo (Choli et al. 1988a, b). Recombinant (unmethylated) forms of these proteins have been produced and studied. Lack of methylation does not appear to affect DNA binding significantly, but it does decrease the measured thermostability of the proteins in vitro (McAfee et al. 1995, 1996; Lundback et al. 1998). Definitive genetic experiments have yet to determine whether these proteins are essential for viability; it is likely, however, that the function provided by the Sso7d-family proteins is indispensable to one or more cellular processes.

A wealth of structural work involving the Sso7d and Sac7d proteins has been published, including X-ray and NMR studies (Agback et al. 1998; Edmondson et al. 1995; Gao et al. 1998; Krueger et al. 1999; Robinson et al. 1998). In contrast to the DNA binding mode of the Euryarchaeal histones (Grayling et al. 1996), these Crenarchaeal proteins bind in the minor groove of double-stranded DNA and are reportedly capable of inducing sharp kinks in the DNA of up to 61° (Robinson et al. 1998). Both Sso7d and an ortholog from *S. shibatae*, Ssh7, have been shown to constrain negative supercoils in DNA (Lopez-Garcia et al. 1998; Mai et al. 1998; Napoli et al. 2002). Interestingly, the archaeal histones, such as those from the Euryarchaeon *Methanothermus fervidus*, can wrap DNA into positive as well as negative supercoils upon binding (Musgrave et al. 1991).

Sso7d-family proteins are also able to increase the apparent melting temperature (T_m) of double-stranded DNA substrates (Baumann et al. 1994; McAfee et al. 1995), suggesting that they may perform a role in stabilizing chromosomal DNA in hyperthermophilic Crenarchaea. As observed by electron microscopy, Sso7d can coat DNA (Lurz et al. 1986), and models of multimeric forms of these proteins have been proposed (Agback et al. 1998; Krueger et al. 1999; Robinson et al. 1998).

Until recently, orthologs of the Sso7d-family proteins were identified only in different species of *Sulfolobus*. Because of their potential utility in industrial applications (Wang et al. 2004), we searched for functional homologs of Sso7d in other Crenarchaea. Here, we describe the bioinformatic identification of a similar but non-homologous class of small DNA-binding proteins present in the hyperthermophilic Crenarchaeotes, *Pyrobaculum* spp., *Aeropyrum pernix*, *Staphylothermus marinus* and *Thermoproteus tenax*.

With the goal of obtaining quantitative data on this newly described class of proteins, we purified the *Pyrobaculum aerophilum* and *A. pernix* proteins along with Sso7d and conducted a comparative biochemical characterization of the interactions of these three proteins with DNA. We examined the ability of each protein to modify the mobility of DNA in agarose gels, to constrain supercoils, and to stabilize duplex DNA. In addition, we report the equilibrium K_D values and kinetic constants describing each protein's binding to single-stranded and double-stranded DNA as measured by surface plasmon resonance (SPR) biosensor analysis.

Material and methods

Reagents and enzymes

All restriction enzymes and DNA ligases were purchased from New England Biolabs (Beverly, MA). Antibiotics and chemicals used for buffer preparation were obtained from Sigma (St. Louis, MO). All gels were stained with SYBR Gold nucleic acid gel stain (Invitrogen, Carlsbad, CA). Oligonucleotides were purchased from IDT (Coralville, IA).

Synthetic gene assembly and cloning

No authentic genomic nucleic acids were used to create the protein expression constructs. Instead, the amino acid sequences obtained from publicly available databases (Fitz-Gibbon et al. 2002; Kawarabayasi et al. 1999) were back-translated taking into account codon usage in *Escherichia coli* (Nakamura et al. 2000). The resulting DNA sequences were subdivided into fragments (~60 nt each) to facilitate synthetic gene assembly and produced as 5′-phosphorylated oligonucleotides. Reverse complementary bridging oligonucleotides (25–30 nt) containing identity to the flanking regions of both coding sequences to be joined were also produced. Oligonucleotide sequences designed for the synthetic gene assembly of *pae3192* and *ape1322b* are listed in Table 1. For the synthetic *sso7d* gene, a design described previously (Wang et al. 2004) was used as a guide for gene synthesis.

Coding and bridging oligonucleotides were mixed at a molar ratio of 1:3; an aliquot (90 μL) of this mixture was combined with annealing buffer (10×: 200 mM Tris, 400 mM NaCl, 10 mM EDTA, pH 8.0) to yield a solution at 1× final concentration. The solution was placed in an AB9700 thermal cycler equipped with a 0.5 ml sample block module (Applied Biosystems, Foster City, CA), heated to 95°C for 5 min, and cooled to 15°C using a 1% ramp rate over approximately 60 min. Multiple aliquots (5, 10, 40 μL) of this mixture were combined with Taq DNA ligase and the corresponding buffer in a 50 μL final volume. The ligation reactions were incubated at 50°C for 18 h, after which the products were purified and recovered

Table 1 Oligonucleotides used in this study

Name	Experiment	Sequence
Ape.1	Gene assembly	5′-P-ATGCCGAAGAAGGAGAAGATTAAGTTCTTCGACCTGGTCGCCAAGAAGTACTACGAG-3′
Ape.2	Gene assembly	5′-P-ACTGACAACTACGAAGTCGAGATTAAGGAGACTAAGCGCGGCAAGTTTCGCTTCGCC-3′
Ape.3	Gene assembly	5′-P-AAAGCCAAGAGCCCGTACACCGGCAAGATCTTCTATCGCGTGCTGGGCAAAGCCTAA-3′
Ape.a	Gene assembly	5′-GTAGTTGTCAGTCTCGTAGTACTT-3′
Ape.b	Gene assembly	5′-GCTCTTGGCTTTGGCGAAGCGAAA-3′
Ape.fwd	Gene assembly	5′-CGTGGGATCCCGAAGAAGGAGAAGATT-3′
Ape.rev	Gene assembly	5′-GATCTCGAGTTAGGCTTTGCCCAGCAC-3′
Pae.1	Gene assembly	5′-P-ATGTCCAAGAAGCAGAAACTGAAGTTCTACGACATTAAGGCGAAGCAGGCGTTTGAG-3′
Pae.2	Gene assembly	5′-P-ACCGACCAGTACGAGGTTATTGAGAAGCAGACCGCCCGCGGTCCGATGATGTTCGCC-3′
Pae.3	Gene assembly	5′-P-GTGGCCAAATCGCCGTACACCGGCATTAAAGTGTACCGCCTGTTAGGCAAGAAGAA ATAA-3′
Pae.a	Gene assembly	5′-GTACTGGTCGGTCTCAAACGCCTG-3′
Pae.b	Gene assembly	5′-CGATTTGGCCACGGCGAACATCAT-3′
Pae.fwd	Gene assembly	5′-CGTGGGATCCCGTCCAAGAAGCAGAAACTG-3′
Pae.rev	Gene assembly	5′-GATCTCGAGTTATTTCTTCTTGCCTAA-3′
OligoA-3′6FAM	T_m assay	5′-GTAAAACGACGGCCAGT-6FAM-3′
OligoB-5′Dabcyl	T_m assay	5′-Dabcyl-ACTGGCCGTCGTTTTAC-3′
OligoA-BioTEG	Biosensor	5′-BioTEG-GTAAAACGACGGCCAGT-3′
OligoB	Biosensor	5′-ACTGGCCGTCGTTTTAC-3′
Lip1F	Biosensor	5′-GGATTGGCTGTTCGGAGCGG-3′
Lip1R-BioTEG	Biosensor	5′-BioTEG-ATTGGATCCACCTCCACAGTGGAGACCAA-3′

P indicates phosphorylation

using the QIAquick PCR purification system (Qiagen, Valencia, CA).

Full-length ligation products were amplified by PCR using the GeneAmp High Fidelity PCR kit (Applied Biosystems). Amplification primers were designed to contain restriction endonuclease cleavage sites at the 5′ and 3′ ends of each amplicon to facilitate cloning. Each gene was cloned into a separate pET11a plasmid (Novagen/EMD Biosciences, Madison, WI) for protein expression. Gene sequences were confirmed from both directions using the BigDye Terminator v3.1 sequencing chemistry and analyzed on an AB3730*xl* DNA sequencer (Applied Biosystems).

Protein production and purification

The Ape1322b, Pae3192 and Sso7d expression constructs were transformed individually into Rosetta2/DE3 (pLysS) *E. coli* cells (Novagen/EMD). All three proteins were purified independently using the following method.

Exponentially growing cultures were induced for 2 h following the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside. Cell pellets from 500 ml culture were resuspended in 50 ml 50 mM potassium phosphate (KPi), pH 6.8, and the mixture was centrifuged to remove cell debris. The resulting extract was incubated at 70°C for 12 min to precipitate *E. coli* proteins. Following centrifugation, the supernatant was brought up to 100 mM KCl, 0.3% polyethylenimine, and incubated on ice for 30 min to precipitate the DNA. This mixture was centrifuged, and 30% ammonium sulfate was added to the recovered supernatant. Following centrifugation, the pellet, which contained contaminating proteins, was discarded. The supernatant was then brought up to 70% ammonium sulfate and incubated at room temperature for several hours. The desired protein was recovered in the pellet upon centrifugation. The pellet was dissolved in 20 ml 50 mM KPi, pH 6.8, and this mixture was run over a heparin column (GE Healthcare) and eluted with a step gradient of KCl in 50 mM KPi, pH 6.8. The Ape1322b protein eluted at 300–350 mM KCl, Pae3192 eluted at 250–300 mM KCl, and Sso7d eluted at 75–125 mM KCl. Peak fractions were pooled and concentrated using a YM-3 Centriprep device (Millipore, Billerica, MA). Protein concentrations were measured using the BCA protein assay (Pierce, Rockford, IL). Final concentrations of the purified proteins were: Ape1322b, 1.9 mg/ml; Pae3192, 2.3 mg/ml; Sso7d, 3.6 mg/ml.

The A_{280}:A_{260} ratios ($\sim$2:1) of our protein preparations indicated they were free of nucleic acids. The proteins' sequence and purity were confirmed by mass spectrometry,

which also showed that 94–99% of the N-terminal methionines were cleaved off in each protein preparation. Pae3192 migrates anomalously on polyacrylamide gels, appearing to be larger than its predicted size. This effect is not due to post-translational modifications, since none were noted by mass spectrometry, and was not seen with the Ape1322b or Sso7d proteins.

Biochemical characterization

Electrophoretic mobility shift analysis (EMSA) was performed by incubating different concentrations of protein with 10 ng linearized pET11a plasmid DNA (5.6 kb) in 60 mM KCl, 10 mM KPi, pH 6.8 (Buffer A) at 40°C for 20 min in 10 µl. For samples that were treated with sodium dodecyl sulfate (SDS) and phenol-chloroform extracted, SDS was added to 0.8% final concentration and an equal volume of 25:24:1 phenol:choloroform:isoamyl alcohol, pH 8.0 (Sigma) was added. After mixing and centrifugation, material from the aqueous layer was run on a 1% agarose gel in Tris-acetate-EDTA (TAE) buffer.

Supercoiling assays used relaxed DNA that we made by treating pT7Blue plasmid DNA (Novagen/EMD, 2.9 kb) with wheat germ topoisomerase I (Promega). Approximately 125 ng of relaxed plasmid DNA was incubated with each of the DNA-binding proteins in the presence of 10 units topoisomerase I in 10 µl Buffer A for 25 min at 37°C. Samples were then brought to 0.8% SDS and phenol-chloroform extracted as above. Half of the aqueous material was run on a 1% agarose gel in TAE buffer. The gel depicted in Fig. 3a contained 2.5 µg/ml chloroquine (Sigma) in the gel and in the running buffer. This gel was washed for 1 h in water to remove the chloroquine prior to staining.

T_m measurements were carried out in 384-well plates with a total volume of 20 µl per well. Each DNA-binding protein was mixed with 100 nM 17 bp duplex DNA (formed by the annealing of OligoA-3′6FAM and OligoB-5′Dabcyl, Table 1) in 25 mM KCl, 5 mM KPi, 20 mM Tris, pH 8.0. Fluorescence was monitored in an Applied Biosystems 7900 real-time PCR instrument using the following cycling profile: 25°C, 15 min; 25 to 95°C with a 2% ramp; 95 to 25°C with a 2% ramp. T_ms were called as the temperature at which the derivative of the denaturation curves reached a maximum.

Biosensor data was collected using a Biacore3000 instrument (GE Healthcare, Piscataway, NJ). The oligonucleotides used as substrates are listed in Table 1. The 17 bp double-stranded (ds) DNA substrate was formed by slow annealing of OligoA-BioTEG and OligoB. The 235 bp product was made by PCR using human genomic DNA as a template with primers Lip1F and Lip1R-BioTEG. Seventy response units (RUs) of 17 nt ssDNA (OligoA-BioTEG) and 70 RUs of 17 bp dsDNA were immobilized onto a streptavidin-coated chip; 57.5 RUs of the 235 bp dsDNA were immobilized. Each protein was tested at the following concentrations, each in triplicate: 6.25, 12.5, 25, 50, 100, 200, 400, 800, and 1,600 nM. All samples were in Buffer A + 0.005% Tween-20; measurements were made at 25°C. The flow rate was 100 µl/min, the injection volume was 200 µl and the dissociation observation time was 180 s. Following each run, the biosensor chip was regenerated by successive 5 µl and 8 µl injections of 2 M KCl, 10 mM KPi, 0.005% Tween-20, pH 6.8. RU_{max} values for protein binding were between 75 and 140 RUs for all samples.

K_D and RU_{max} values were calculated by plotting the average equilibrium RU level ($RU_{[P]}$) for each protein concentration ([P]) and fitting the points in KaleidaGraph (Synergy Software, Reading, PA) to a simple bimolecular reaction model using the equation $RU_{[P]}/RU_{max} = [P]/([P] + K_D)$. The stoichiometry of binding was calculated using the equation $RU_{max} = (MW_P/MW_L)\cdot RU_L\cdot s$, where MW_P is the molecular weight of the protein, MW_L is the molecular weight of the DNA, RU_L is the number of RUs of DNA immobilized on the chip, and s is the stoichiometry of protein:DNA. The apparent site sizes were calculated for the 235 bp dsDNA as 235/s. k_d values were fit to data from the protein dissociation phase of the experiments using the equation $RU_t = m_1 + m_2\cdot e^{-kt}$ where RU_t is the RU value at a given time, t; k is the k_d; and m_1 and m_2 were floated. Calculated k_d values from 15 to 18 injections for each protein on each substrate were averaged; this value is reported in Table 2. Standard errors reported in Table 2 are from the curve fits for the K_D values and are the standard errors on the mean for the k_d values. k_a values were calculated using the equation $k_a = k_d / K_D$; the errors in these values were propagated from the standard errors in the k_d and K_D measurements.

Results

Identification of Pae3192 and Ape1322b

The Pae3192 ORF and its paralogs were identified as candidate DNA-binding proteins solely by bioinformatic methods. A recent account has detailed an alternative route to the identification of these polypeptides (Luo et al. 2007). Using the characteristics of the Sso7d-family proteins as a guide, we sorted the list of predicted ORFs from the completed *P. aerophilum* genome sequence (Fitz-Gibbon et al. 2002) and searched for small, basic proteins. Among the low molecular weight polypeptides (7–10 kDa) predicted to have a high isoelectric point (pI ~10) were three virtually identical copies of an ORF annotated as

"PaRep4". These three ORFs of unknown function are named Pae3192, Pae3289, and Pae0384. (For simplicity, we will refer only to Pae3192.) An attempted pairwise alignment between Sso7d and Pae3192 revealed no significant sequence similarity.

A similar search strategy using the original annotated complete genome sequence of *A. pernix* (Kawarabayasi et al. 1999) did not reveal any candidate ORFs. However, a TBLASTN search (Altschul et al. 1990) against the nucleotide sequence database revealed a potential ORF with homology to Pae3192. This ORF has subsequently been named Ape1322b. Unlike the multiple genes coding for the Sso7 or the *P. aerophilum* proteins, the Ape1322b ORF appears only once in the genome.

The genomic dataset from the Crenarchaeon *Staphylothermus marinus* also recently became available (unpublished; US DOE Joint Genome Institute); a BLASTP search on this genome revealed an ORF with high similarity to Pae3192, Sma0380. As was the case for Ape1322b, only one copy of this sequence was identified in the *S. marinus* genome.

The sequence similarities among the proteins that form the Pae3192 family (Group I) are shown in Fig. 1. This group contains proteins from both the Desulfurococcales and Thermoproteales subdivisions of the Crenarchaea. Because of their presence in the Crenarchaea, Luo et al. proposed to name the Pae3192-family of proteins Crenarchaeal chromatin protein 1, or CC1. The Sso7d-family proteins (Group II) form a separate group, not related by sequence to the Pae3192-family, and are restricted to the Sulfolobales subdivision of the Crenarchaea.

DNA binding by Ape1322b, Pae3192 and Sso7d

While we identified the Pae3192 and Ape3122b polypeptides as having features similar to authentic DNA-binding proteins like Sso7d, no biochemical data addressing their function was available. Therefore, we created expression constructs for the Ape1322b, Pae3192 and Sso7d proteins using synthetic methods and purified the three proteins (see "Materials and methods"). These recombinant proteins were used in the comparative biochemical experiments described below.

We first tested whether the three proteins bound to dsDNA. We carried out electrophoretic mobility shift analysis (EMSA) using a 5.6 kb linear DNA as a substrate (Fig. 2). Different concentrations of each of the three proteins were incubated with the DNA substrate, and the resulting protein-DNA complexes were run on an agarose gel.

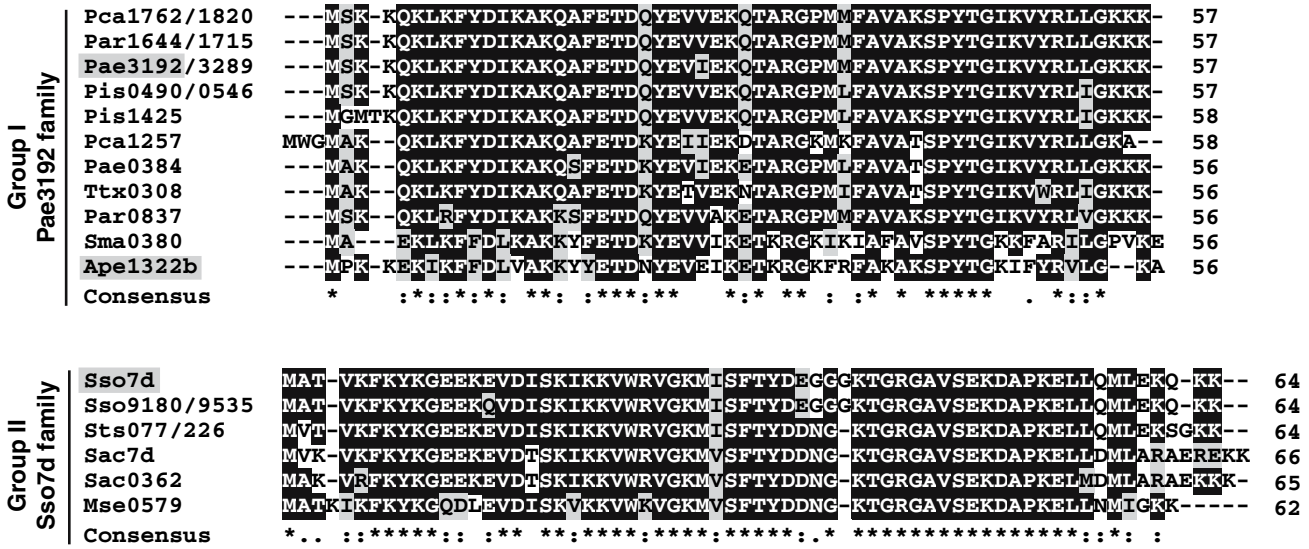

Fig. 1 Sequence alignments of Pae3192-family proteins (*Group I*) and Sso7d-family proteins (*Group II*). ORFs are indicated by their generic names based upon the organism and the gene position, or in some cases, the historic name is used. The names of the proteins used in this study are highlighted in *gray*. The number of amino acids composing each protein is indicated to the right of the sequence. An *asterisk* in the consensus sequence indicates sequence identity at that position across all sequences. Two dots signal a residue having conserved substitutions, and a single dot indicates a residue where semi-conserved substitutions have occurred. Amino acids highlighted in *black* have identity to the consensus amino acid at each position, and those highlighted in *gray* show similarity to the consensus residue. Sequence alignments were generated with CLUSTAL W (Thompson et al. 1994). Abbreviations: *Pca Pyrobaculum calidifontis*; *Par P. arsenaticum*; *Pae P. aerophilum*; *Pis P. islandicum*; *Ttx Thermoproteus tenax*; *Sma Staphylothermus marinus*; *Ape Aeropyrum pernix*; *Sso Sulfolobus solfataricus*; *Sac S. acidocaldarius*; *Sts S. tokodaii*; *Ssh S. shibatae*; *Mse Metallosphaera sedula*

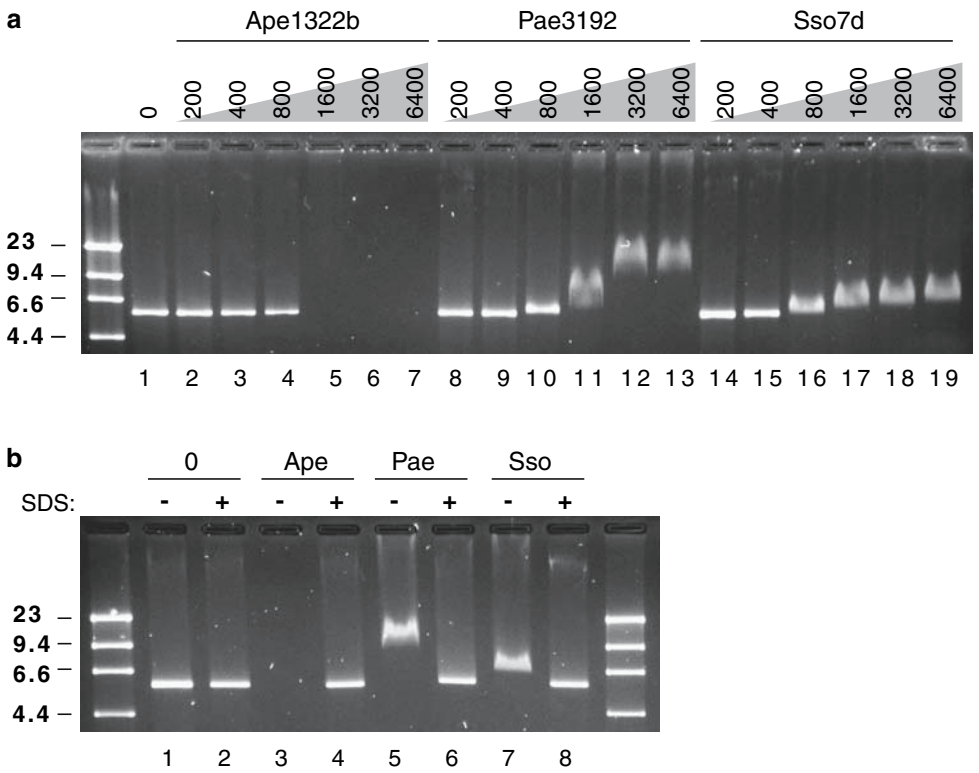

Fig. 2 Electrophoretic mobility shift analysis of DNA binding by Ape1322b, Pae3192 and Sso7d. **a** The indicated concentration (nM) of each DNA-binding protein was incubated with 10 ng of a 5.6 kb linear DNA for 20 min at 40°C in 10 µl 60 mM KCl, 10 mM potassium phosphate, pH 6.8. The binding reactions were run on a 1% agarose gel. A Lambda HindIII ladder was run in the *left-most lane*; the ladder fragments are labeled on the left side of the gel according to their sizes in kilobases. **b** Duplicate binding reactions containing 3,200 nM Ape1322b (Ape), Pae3192 (Pae) or Sso7d (Sso), or no protein (0), were set up in as in (a). SDS and phenol-chloroform were added to one of each of the duplicate reactions, and material from the aqueous layer was loaded on a 1% gel (lanes labeled "+"). The untreated binding reactions were also loaded on the gel (lanes labeled "−")

We made several observations from the EMSA experiments. First, all three proteins were able to shift the substrate DNA to slower-migrating species, indicating that the proteins interacted with the DNA. Second, the protein concentration at which the substrate began to migrate as a larger protein-DNA species was similar for all three proteins (around 800 nM).

Finally, the most surprising results from these experiments were the differences observed among the shifted complexes depending on the protein that was added. For example, at Ape1322b concentrations of 1600 nM and above, the DNA substrate appeared to vanish from the gel (Fig. 2a, lanes 5–7). We interpret this behavior as the DNA being shifted into the well while also becoming inaccessible to the fluorescent dye used to stain the DNA. When the DNA was pre-stained with ethidium bromide, the Ape1322b-bound DNA could clearly be seen in the well (not shown). We know that the disappearance of the DNA was not due to contaminating nuclease activity in our Ape1322b protein preparation because when SDS was added to the 3200 nM protein samples and the proteins removed by phenol-chloroform extraction, the DNA was recovered (Fig. 2b, lane 4). In addition, no nuclease activity was observed in our supercoiling assays (see below).

There were also differences in the way Pae3192 and Sso7d shifted the substrate in this experiment. Pae3192-bound DNA had a lower mobility than Sso7d-bound DNA (Fig. 2a, compare lanes 13 and 19). Thus, our EMSA experiments showed that Ape1322b, Pae3192 and Sso7d bind to dsDNA at low micromolar concentrations of protein and that the protein-bound species appear to be different for each protein.

Constraint of negative supercoils by Ape1322b, Pae3192 and Sso7d

Sso7d-family proteins have been shown to constrain negative supercoils in DNA (Lopez-Garcia et al. 1998; Mai et al. 1998; Napoli et al. 2002). We tested whether Ape1322b and Pae3192 are also able to constrain supercoils. For this experiment, each DNA-binding protein was incubated with a relaxed plasmid substrate (2.9 kb) in the

presence of a topoisomerase. The topoisomerase relaxes compensatory supercoils created upon protein binding. After deproteinization by addition of SDS and phenol-chloroform, the DNA contains only the supercoils that were physically constrained by the DNA-binding protein. This DNA can then be run on agarose gels with or without chloroquine to determine the handedness of the resulting supercoils.

The results for this assay are shown in Fig. 3. Increasing concentrations of each protein were incubated with relaxed plasmid DNA in the presence of wheat germ topoisomerase I. Wheat germ topoisomerase I can relax both positive and negative supercoils, so constrained supercoils of either sign will be detected in this assay. At high protein concentrations, Pae3192 and Sso7d caused the plasmid to become highly supercoiled (Fig. 3a, lanes 9 and 13), running with a mobility similar to that of negatively supercoiled plasmid DNA isolated from *E. coli* (lane 14). In contrast, the Ape1322b protein affected the topoisomer distribution only slightly, causing the formation of one or two additional negative topoisomers (lanes 4 and 5). We confirmed that all three proteins were able to bind to the relaxed DNA under these conditions by EMSA (not shown). Thus, the lack of supercoiling by Ape1322b does not result from failure of Ape1322b to bind to this substrate. Instead, we suggest that the failure of Ape1322b to generate supercoiled DNA in this assay is due to the inability of the topoisomerase to act on the Ape1322b-coated DNA (see "Discussion").

To test the handedness of the tightly supercoiled DNA generated by Pae3192 and Sso7d binding, we ran the DNA from the 3,200 nM protein samples on a gel containing 2.5 μg/ml chloroquine (Fig. 3b). Chloroquine is a DNA intercalator that causes topoisomer distributions to shift toward more positively supercoiled forms. Thus, negatively supercoiled DNA will become more relaxed in the presence of chloroquine, while relaxed DNA will become positively supercoiled. Positively supercoiled DNA then becomes even more tightly positively supercoiled under these conditions and runs ahead of a relaxed DNA marker. The Pae3192 and Sso7d-reacted plasmid (Fig. 3b, lanes 3 and 4) became more relaxed in the chloroquine gel, migrating similarly to the negatively supercoiled marker (lane 2), rather than running ahead of the relaxed marker (lane 1) as positively supercoiled DNA would have done. Thus, the Pae3192 and Sso7d proteins constrain negative supercoils in DNA.

The topoisomer distribution of the Pae3192 and Sso7d-treated plasmid DNA was approximately two to three topoisomers more relaxed than that of the negatively supercoiled plasmid DNA (Fig. 3b). This corresponds to a difference in the superhelical density (σ) of approximately

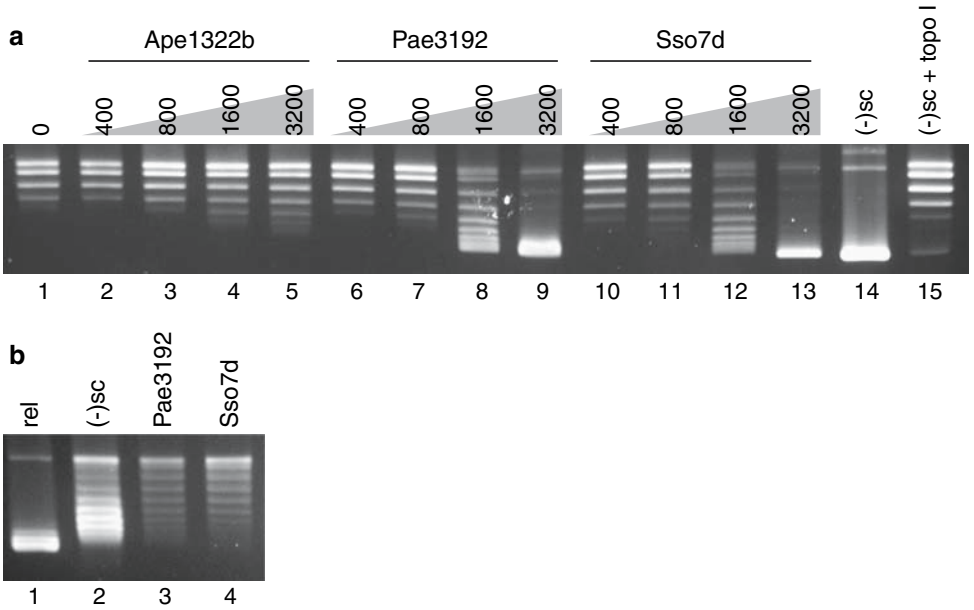

Fig. 3 Effect on DNA supercoiling by Ape1322b, Pae3192 and Sso7d. **a** The indicated concentration (nM) of Ape1322b, Pae3192 and Sso7d were incubated with approximately 125 ng of relaxed plasmid DNA (pT7Blue, 2.9 kb) in the presence of 10 units wheat germ topoisomerase (topo) I in 10 μl 60 mM KCl, 10 mM KPi, pH 6.8 for 25 min at 37°C. Negatively supercoiled pT7Blue was incubated with wheat germ topoisomerase I in the absence of DNA-binding proteins in a parallel reaction to control for topoisomerase I activity (*lane 15*). Following addition of SDS, the samples were phenol-chloroform extracted, and half of the aqueous material was loaded on a 1% agarose gel. Untreated negatively supercoiled pT7Blue DNA was run as a marker in *lane 14*. **b** The other half of the material from *lanes 9 and 13* in **a** was run on a 1% agarose gel containing 2.5 μg/ml chloroquine alongside relaxed (*lane 1*) and negatively supercoiled (*lane 2*) pT7blue DNA

0.01 for a 2.9 kb plasmid. Since negatively supercoiled plasmid DNA isolated from *E. coli* has a σ around −0.06 (Giaever et al. 1988), the Pae3192 and Sso7d-reacted DNA has a σ of approximately −0.05, which gives a rough estimate of 14 supercoils put into the plasmid upon binding of the Pae3192 or Sso7d protein.

Stabilization of duplex DNA

Sso7d has been shown to stabilize duplex DNA, raising the T_m of poly (dIdC) DNA by as much as 38°C (Baumann et al. 1994). We asked whether Ape1322b and Pae3192 would have a similar effect on dsDNA. As a substrate, we used a 17 bp duplex DNA created by the annealing of a 3′-fluorescein-labeled oligonucleotide and a complementary oligonucleotide labeled at the 5′ end with a dabcyl quencher. When this duplex melts, the fluorescent signal increases due to the relief of fluorescence quenching. The T_m of the duplex can then be determined as the inflection point in a plot of fluorescence versus temperature. When high concentrations of Ape1322b, Pae3192 or Sso7d protein were added to this DNA substrate, the T_m of the DNA increased in a protein concentration-dependent manner (Fig. 4). The pattern of the effects on T_m mirrored those seen in the EMSA and in the supercoiling experiments, with little response observed at protein concentrations under 800 nM.

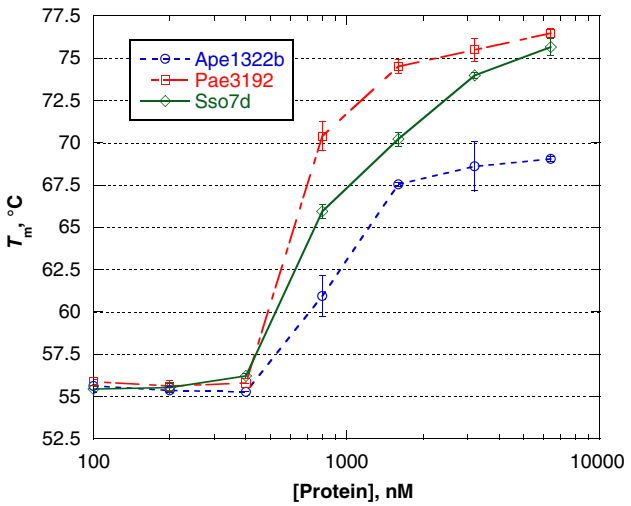

Fig. 4 Duplex DNA stabilization by Ape1322b, Pae3192 and Sso7d. Increasing amounts of each DNA-binding protein were incubated with 100 nM 17 bp duplex DNA (OligoA-3′6FAM + OligoB-5′Dab-cyl) in 20 µl 25 mM KCl, 5 mM KPi, 20 mM Tris, pH 8.0. Fluorescence was monitored in an Applied Biosystems 7900 real-time PCR instrument. The melting temperature (T_m) of the DNA in each sample was determined as the temperature at which the increase in fluorescence reached an inflection point during a slow ramp between 25 and 95°C. Error bars correspond to the standard deviation in measurements from triplicate samples

Ape1322b did not shift the T_m to the same extent as did Pae3192 and Sso7d. This may be because Ape1322b has a lower thermostability than Pae3192 and Sso7d or because it has a different effect on double-stranded DNA. Interestingly, the observed stabilization of duplex DNA was similar whether the T_m was measured during the dissociation (heating) phase of the temperature cycle or during the renaturation (cooling) phase and was repeatable during multiple cycles of heating and cooling (not shown). These results suggest that the DNA-binding proteins are able to renature into functional forms (i.e., competent to bind DNA) after heating to 95°C.

Equilibrium and kinetic parameters of DNA binding by Ape1322b, Pae3192 and Sso7d

Our data up to this point suggested that Ape1322b, Pae3192 and Sso7d may have differences in their modes of binding to DNA. To analyze these differences in a more quantitative fashion, we carried out a binding study using a Biacore3000 SPR biosensor apparatus. Biosensor experiments have the advantage of measuring both equilibrium and kinetic parameters, as well as providing information on the stoichiometry of the substrate-analyte complexes formed. We analyzed the binding of Ape1322b, Pae3192 and Sso7d to three substrates: a 17 nt ssDNA, a 17 bp dsDNA, and a 235 bp dsDNA.

We measured the K_D values of each protein for these substrates under equilibrium conditions (Table 2). An example of curve fits to the equilibrium data are shown in Fig. 5. In this example, the fraction bound (RU/RU_{max}) is plotted for the three proteins with the ssDNA substrate.

The dissociation rates (k_d) for each protein's binding to the ssDNA and dsDNA substrates are also presented in Table 2. These data were obtained during the dissociation phase of each run and had high precision. Because of the difficulties in accurately measuring association rates for small analytes using biosensor instruments (Karlsson 1999), we report association rates (k_a) calculated using the equation $k_a = K_D/k_d$. These calculated k_a values were within two-fold of the k_a values observed directly on the biosensor instrument. Also indicated in Table 2 is the number of proteins bound per substrate nucleic acid; this information was used to estimate a binding site size for the 235 bp substrate.

Four main conclusions can be drawn from the biosensor measurements. First, as demonstrated in Fig. 5, Ape1322b and Pae3192 have significantly higher affinities for ssDNA (K_D = 253 nM and 127 nM, respectively) than does Sso7d (K_D = 1,830 nM).

Second, Pae3192 and Sso7d have comparable affinities for dsDNA (K_D = 165 nM and 121 nM, respectively, on

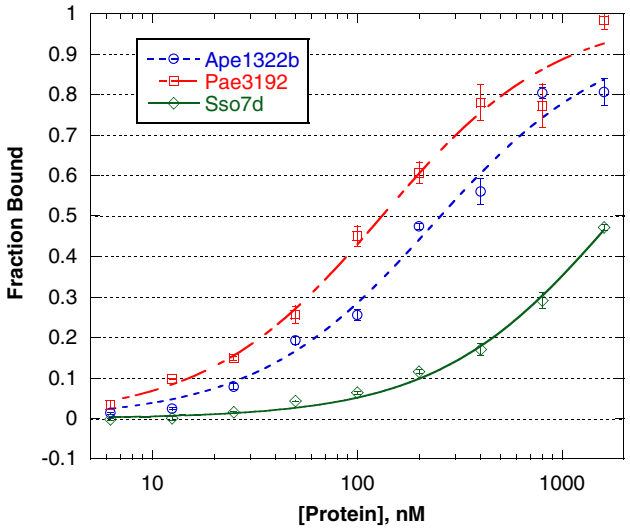

Fig. 5 Binding of Ape3192, Pae3192 and Sso7d to single-stranded DNA measured using a surface plasmon resonance biosensor instrument. The fraction bound was calculated as the RU/RU_{max} after equilibrium was reached for each protein concentration. The data points were fit to a simple bimolecular reaction model ($A + B \leftrightarrow AB$). Error bars correspond to the standard deviation in measurements from triplicate samples

the ds 17-mer), while Ape1322b appears to have lower affinity (K_D = 530 nM). The data for Ape1322b binding to dsDNA were difficult to fit because additional data points at micromolar protein concentrations would have been required to define a complete curve. Unfortunately, high concentrations of Ape1322b produced unusual signals on the biosensor chip, possibly due to non-specific interactions between the protein and the chip surface. Thus, the K_D values for Ape1322b binding to dsDNA contained sizable error. For each protein, the K_D values for binding to dsDNA were similar (within two-fold) whether the smaller 17 bp or larger 235 bp dsDNA substrate was used (Table 2).

A third observation we made from our biosensor measurements was that even though Pae3192 and Sso7d had similar K_D values for binding to dsDNA, Sso7d exhibited much faster on- and off-rates compared with those for Pae3192 (Table 2).

Finally, the binding site sizes calculated from the stoichiometry of protein binding are in the range of that published for Sac7d on calf thymus DNA (4.55 bp) (McAfee et al. 1996). The slightly larger binding site size reported here for Sso7d (6.56 bp) may reflect sequence context effects or surface effects that could limit the available binding space on the DNA substrate. Among the three proteins, Pae3192 appeared to have the smallest binding site, at 4.78 bp. This conclusion is corroborated by the observation that three Pae3192 proteins were able to bind to the 17 bp duplex while only two Ape1322b proteins or two Sso7d proteins were able to load onto this substrate (Table 2). The numbers of proteins bound per 17 nt ssDNA substrate were not integer values; this may be a result of heterogeneity in the ssDNA molecules due to secondary structure formation.

Discussion

We have presented studies that began with an *in silico* approach to gene identification. With the ever-expanding availability of genomic sequence information, it is possible to formulate and test hypotheses regarding the function of various genes and gene families without the benefit of previous experimental data. Construction of genes by completely synthetic methods allows for protein

Table 2 Equilibrium and kinetic parameters of DNA binding by Ape1322b, Pae3192 and Sso7d in 60 mM KCl, 10 mM KPi, 0.005% Tween-20, pH 6.8, at 25°C

Protein	Substrate	K_D, nM	Stoichiometry (protein:DNA)	Binding site, bp	k_a (k_d/K_D), M^{-1} s^{-1}	k_d, s^{-1}
Ape1322b	17 nt ssDNA	253 (31)	1.6 (0.05)		6.36×10^5 (0.92×10^5)	0.161 (0.012)
Ape1322b	17 bp dsDNA	530 (162)	1.9 (0.19)		4.04×10^5 (1.24×10^5)	0.220 (0.008)
Ape1322b	235 bp dsDNA	544 (156)	38.3 (3.8)	6.14 (0.61)	3.28×10^5 (1.00×10^5)	0.174 (0.012)
Pae3192	17 nt ssDNA	127 (18)	1.6 (0.06)		8.90×10^5 (1.34×10^5)	0.113 (0.007)
Pae3912	17 bp dsDNA	165 (21)	3.0 (0.12)		5.84×10^5 (0.85×10^5)	.096 (0.006)
Pae3192	235 bp dsDNA	97 (16)	49.1 (2.5)	4.78 (0.24)	7.53×10^5 (1.32×10^5)	0.073 (0.004)
Sso7d	17 nt ssDNA	1830 (325)	1.2 (0.13)		$>1.81 \times 10^6$	>3.31[a]
Sso7d	17 bp dsDNA	121 (21)	1.8 (0.09)		3.81×10^6 (0.38×10^6)	0.461 (0.029)
Sso7d	235 dsDNA	113 (9)	35.8 (0.8)	6.56 (0.14)	3.11×10^6 (0.40×10^6)	0.351 (0.015)

Standard errors are given in parenthesis

[a] Full dissociation of Sso7d from the ssDNA substrate occurred by the first data point collected following the end of protein injection. Thus, only a lower limit to the dissociation rate could be calculated based on full dissociation of the protein within 1 s

production without the need for natural cellular extracts. This is especially important in comparative metagenomic analyses, where many of the represented organisms are unculturable and the original nucleic acid sample may not be available. Using such synthetic approaches, potentially more DNA-binding proteins from uncultured marine Crenarchaea could be identified and characterized (Takai et al. 2004). Indeed, signatures with high similarity to Pae3192 can be found in shotgun sequence assembly databases like that from the microbial community of the Sargasso Sea (Venter et al. 2004).

Because structural data are not yet available for the Pae3192-family proteins, it is not known if they assume the same conformational fold as the Sso7d-family. The apparent preference of Sso7d to bind to dsDNA over ssDNA, as well as its faster on- and off-rates compared with Pae3192 (Table 2), suggest that the Pae3192-family of proteins may have a different mode of binding DNA than does Sso7d. Nevertheless, Pae3192 and Sso7d showed similar binding affinities for dsDNA, having K_D values around 100–150 nM (Table 2), and both proteins induced negative supercoiling in DNA to a similar extent (Fig. 3). In addition, Ape1322b, Pae3292 and Sso7d all raised the T_m of duplex DNA upon binding (Fig. 4). These results suggest that the two families of proteins represented by Pae3192 and Sso7d could perform similar functions in vivo.

When added in high concentrations, Pae3192 and Sso7d each added approximately 14 negative supercoils into a 2.9 kb plasmid (see Results). Using the binding site sizes shown in Table 2, we calculated that ~600 Pae3192 proteins or ~440 Sso7d proteins could bind to this plasmid. Thus, assuming that the plasmid was fully saturated for protein binding under the conditions tested, binding of ~40 molecules of Pae3192 or ~30 molecules of Sso7d resulted in the formation of one negative supercoil. These values are similar to what was previously reported for Ssh7: binding of approximately 20 Ssh7 molecules caused the formation of one negative supercoil (Mai et al. 1998). These effects on supercoiling could result from protein-induced changes in either the twist or the writhe of the DNA. In other words, the proteins may cause local unwinding of the DNA, or they may physically wrap the DNA around themselves, possibly forming nucleosome-like structures.

The K_D values for Pae3192 binding to dsDNA we observed (100–150 nM) differ by over an order of magnitude from that measured by EMSA for the Pae3192 ortholog, Ttx0308 (8 μM) (Luo et al. 2007). We hypothesize that this disparity is due to the large amount of substrate used in the Ttx0308 study and the complicated nature of EMSA as compared to the more quantitative method (SPR biosensor) we used here.

A surprising result from our studies was that even though Ape1322b and Pae3192 share 74% similarity at the amino acid level, we observed dramatic differences in the effects these proteins had on DNA. In our EMSA experiments, Ape1322b prevented the DNA/protein complexes from migrating into the agarose gel from the sample loading well and appeared to exclude dyes from staining the DNA (Fig. 2). Ape1322b binding did not result in highly supercoiled DNA in the presence of a topoisomerase (Fig. 3). We surmise that as was the case for the fluorescent DNA stains, the topoisomerase was not able to access the Ape1322b-bound DNA to relax compensatory supercoils. Alternatively, the Ape1322b-bound DNA may not be supercoiled. The Ape1322b protein also produced somewhat different results in the dsDNA stabilization assay, raising the T_m of the DNA to a lesser extent than did the Pae3192 and Sso7d proteins (Fig. 4).

The complete and sudden disappearance of substrate in our EMSA experiments at Ape1322b concentrations above 800 nM was suggestive of cooperative binding (Fig. 2). It was difficult to determine from our biosensor measurements whether Ape1322b bound to dsDNA in a cooperative manner because high concentrations of Ape1322b could not be tested. There was no evidence for cooperativity in Ape1322b binding to ssDNA (Fig. 5) or in Pae3192 and Sso7d binding to either ssDNA or dsDNA as judged by the good fit of the data to a simple bimolecular reaction model. Nevertheless, these analyses do not preclude more complicated models of binding.

Responses in our EMSA, supercoiling and T_m assays required high concentrations of protein (above 800 nM), well above the K_D values of the proteins for dsDNA (Table 2). We hypothesize that these effects result only when multiple proteins bind along individual DNA substrates. It appears that such multiply bound structures are different for Ape1322b and Pae3192. This may be a result of different protein-protein interactions and/or subtle differences in the DNA-binding interface for each protein. Ape1322b appears to form such a tight coating on DNA that even small dyes are unable to access the DNA within. Given the high abundance of proteins like Sso7d in cellular extracts, these multi-protein complexes may be the biologically relevant form of these proteins inside cells. Indeed, it was recently demonstrated that the Ttx0308 protein is expressed at high levels in *T. tenax* (Luo et al. 2007).

The Ape1322b protein and its ortholog from *S. marinus* are sequence outliers in the Pae3192-family (Fig. 1). Both *A. pernix* and *S. marinus* are members of the Desulfurococcales subset of the Crenarchaea, while the *Pyrobaculum* spp. and *T. tenax* are members of the Thermoproteales. Thus, Ape1322b and Sma0380 may form a subdivision within the Pae3192-family of proteins that has some

independent biochemical properties. It would be informative to determine whether the Sma0380 protein shares biochemical properties with Ape1322b, such as the ability to exclude dyes from DNA.

A potential complicating factor to interpreting the biological relevance of the differences in DNA binding seen with Ape1322b and Pae3192 is the possible effect of lysine methylation on protein-protein or protein-DNA interactions. For example, in vivo, Ape1322b may be hypermethylated or hypomethylated in such a way that its behavior is more similar to that of Pae3192. In addition, as is the case for Sso7d (McAfee et al. 1996), the binding of Ape1322b and Pae3192 to DNA is highly salt dependent with higher salt concentrations resulting in lower affinities (data not shown). In this study, we did not examine extensively the effect of salt on the binding behavior of the proteins. There may be certain ionic conditions under which Ape1322b and Pae3192 behave more similarly. A definitive answer on whether Ape1322b and Pae3192 form different structures with DNA may have to await molecular structures of these proteins bound to DNA.

What is the function of the Pae3192-family proteins in the cell? It has been suggested that Pae3192 may be the as of yet undiscovered single-stranded DNA-binding protein (SSB) in the *Pyrobaculum* spp. (Luo et al. 2007). Our work is not inconsistent with this idea; indeed, Pae3192 has a higher affinity for ssDNA (K_D = 127 nM) than do Ape1322b and Sso7d (Fig. 5, Table 2). However, the binding constants of Pae3192 for single- and double-stranded DNA are similar (Table 2), and we cannot rule out the possibility that there is another novel SSB in the *Pyrobaculum* spp. In addition, Pae3192-family proteins are found in at least two species that contain a known SSB, *S. marinus* and *A. pernix*, indicating that the Pae3192-family proteins are at least not the exclusive SSB in these organisms.

It is interesting to note that two completed Crenarchaeal genome sequences, those from *Hyperthermus butylicus* (a member of the Desulfurococcales) and *Thermophilum pendens* (a member of the Thermoproteales), contain neither an Sso7d-family nor a Pae3192-family protein. These species may contain yet another class of small DNA-binding proteins that fulfill an equivalent role. Or, these rather small gene sequences may be missing from the genomic sequence data sets for these organisms. It seems that the current sample size of Crenarchaeal genomes is still too small to determine from genomic analysis alone the roles of these proteins in the cell. The sequencing of more Crenarchaeal genomes and advances in the ability to perform genetic experiments in Crenarchaeal organisms may provide more insight into the roles the various DNA-binding proteins play in these unusual organisms.

Our results expand the understanding of chromatin in Archaea and highlight the naturally occurring sequence diversity among small DNA-binding proteins. The lack of sequence similarity between the Pae3192 and Sso7d families of Crenarchaeal proteins suggests that they arose from different ancestral sequences. Our characterization of these proteins' DNA interactions revealed some shared properties across families and also exposed differences within families. This study is an example of the power of combining comparative bioinformatics and comparative biochemistry to provide insights into protein structure, function and diversity.

Acknowledgments We thank Keith Bjornson for assistance with data interpretation, Chris Murray for assistance with the biosensor apparatus, and Sylvia Yuen for carrying out mass spectrometry on the purified proteins.

References

Agback P, Baumann H, Knapp S, Ladenstein R, Hard T (1998) Architecture of nonspecific protein-DNA interactions in the Sso7d-DNA complex. Nat Struct Biol 5:579–584

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Baumann H, Knapp S, Lundback T, Ladenstein R, Hard T (1994) Solution structure and DNA-binding properties of a thermostable protein from the archaeon *Sulfolobus solfataricus*. Nat Struct Biol 1:808–819

Chen L, Brugger K, Skovgaard M, Redder P, She Q, Torarinsson E, Greve B, Awayez M, Zibat A, Klenk HP, Garrett RA (2005) The genome of *Sulfolobus acidocaldarius*, a model organism of the Crenarchaeota. J Bacteriol 187:4992–4999

Choli T, Henning P, Wittmann-Liebold B, Reinhardt R (1988a) Isolation, characterization and microsequence analysis of a small basic methylated DNA-binding protein from the archaebacterium, *Sulfolobus solfataricus*. Biochim Biophys Acta 950:193–203

Choli T, Wittmann-Liebold B, Reinhardt R (1988b) Microsequence analysis of DNA-binding proteins 7a, 7b, and 7e from the archaebacterium *Sulfolobus acidocaldarius*. J Biol Chem 263:7087–7093

Edmondson SP, Qiu L, Shriver JW (1995) Solution structure of the DNA-binding protein Sac7d from the hyperthermophile *Sulfolobus acidocaldarius*. Biochemistry 34:13289–13304

Fitz-Gibbon ST, Ladner H, Kim UJ, Stetter KO, Simon MI, Miller JH (2002) Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. Proc Natl Acad Sci USA 99:984–989

Gao YG, Su SY, Robinson H, Padmanabhan S, Lim L, McCrary BS, Edmondson SP, Shriver JW, Wang AH (1998) The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA. Nat Struct Biol 5:782–786

Giaever GN, Snyder L, Wang JC (1988) DNA supercoiling in vivo. Biophys Chem 29:7–15

Grayling RA, Sandman K, Reeve JN (1996) Histones and chromatin structure in hyperthermophilic Archaea. FEMS Microbiol Rev 18:203–213

Grote M, Dijk J, Reinhardt R (1986) Ribosomal and DNA binding proteins of the thermophilic archaebacterium *Sulfolobus acidocaldarius*. Biochim Biophys Acta 873:405–413

Karlsson R (1999) Affinity analysis of non-steady-state data obtained under mass transport limited conditions using BIAcore technology. J Mol Recognit 12:285–292

Kawarabayasi Y, Hino Y, Horikawa H, Yamazaki S, Haikawa Y, Jinno K, Takahashi M, Sekine M, Baba S, Ankai A, Kosugi H, Hosoyama A, Fukui S, Nagai Y, Nishijima K, Nakazawa H, Takamiya M, Masuda S, Funahashi T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Kikuchi H (1999) Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. DNA Res 6:83–52

Kimura M, Kimura J, Davie P, Reinhardt R, Dijk J (1984) The amino-acid-sequence of a small DNA binding protein from the archaebacterium *Sulfolobus solfataricus*. Febs Letters 176:176–178

Krueger JK, McCrary BS, Wang AH, Shriver JW, Trewhella J, Edmondson SP (1999) The solution structure of the Sac7d/DNA complex: a small-angle X-ray scattering study. Biochemistry 38:10247–10255

Lopez-Garcia P, Knapp S, Ladenstein R, 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

Lundback T, Hansson H, Knapp S, Ladenstein R, Hard T (1998) Thermodynamic characterization of non-sequence-specific DNA-binding by the Sso7d protein from *Sulfolobus solfataricus*. J Mol Biol 276:775–786

Luo X, Schwarz-Linek U, Botting CH, Hensel R, Siebers B, White MF (2007) CC1, a novel crenarchaeal DNA binding protein. J Bacteriol 189:403–409

Lurz R, Grote M, Dijk J, Reinhardt R, Dobrinski B (1986) Electron microscopic study of DNA complexes with proteins from the archaebacterium *Sulfolobus acidocaldarius*. EMBO J 5:3715–3721

Mai VQ, Chen X, Hong R, Huang L (1998) Small abundant DNA binding proteins from the thermoacidophilic archaeon *Sulfolobus shibatae* constrain negative DNA supercoils. J Bacteriol 180:2560–2563

McAfee JG, Edmondson SP, Datta PK, Shriver JW, Gupta R (1995) Gene cloning, expression, and characterization of the Sac7 proteins from the hyperthermophile *Sulfolobus acidocaldarius*. Biochemistry 34:10063–10077

McAfee JG, Edmondson SP, Zegar I, Shriver JW (1996) Equilibrium DNA binding of Sac7d protein from the hyperthermophile *Sulfolobus acidocaldarius*: fluorescence and circular dichroism studies. Biochemistry 35:4034–4045

Musgrave DR, Sandman KM, Reeve JN (1991) DNA binding by the archaeal histone HMf results in positive supercoiling. Proc Natl Acad Sci USA 88:10397–10401

Nakamura Y, Gojobori T, Ikemura T (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res 28:292

Napoli A, Zivanovic Y, Bocs C, Buhler C, Rossi M, Forterre P, Ciaramella M (2002) DNA bending, compaction and negative supercoiling by the architectural protein Sso7d of *Sulfolobus solfataricus*. Nucleic Acids Res 30:2656–2662

Reeve JN, Bailey KA, Li WT, Marc F, Sandman K, Soares DJ (2004) Archaeal histones: structures, stability and DNA binding. Biochem Soc Trans 32:227–230

Robinson H, Gao YG, McCrary BS, Edmondson SP, Shriver JW, Wang AH (1998) The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. Nature 392:202–205

She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CC, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PM, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, Van der Oost J (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. Proc Natl Acad Sci USA 98:7835–7840

Takai K, Oida H, Suzuki Y, Hirayama H, Nakagawa S, Nunoura T, Inagaki F, Nealson KH, Horikoshi K (2004) Spatial distribution of marine crenarchaeota group I in the vicinity of deep-sea hydrothermal systems. Appl Environ Microbiol 70:2404–2413

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304:66–74

Wang Y, Prosen DE, Mei L, Sullivan JC, Finney M, Vander Horn PB (2004) A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. Nucleic Acids Res 32:1197–1207