Gene Cloning, Expression, and Characterization of the Sac7 Proteins from the Hyperthermophile *Sulfolobus acidocaldarius*†

James G. McAfee,‡ Stephen P. Edmondson, Prasun K. Datta,§ John W. Shriver,* and Ramesh Gupta*

Department of Medical Biochemistry, School of Medicine, Southern Illinois University, Carbondale, Illinois 62901-4413

Received March 28, 1995; Revised Manuscript Received May 19, 1995®

ABSTRACT: The genes for two Sac7 DNA-binding proteins, Sac7d and Sac7e, from the extremely thermophilic archaeon *Sulfolobus acidocaldarius* have been cloned into *Escherichia coli* and sequenced. The *sac7d* and *sac7e* open reading frames encode 66 amino acid (7608 Da) and 65 amino acid (7469 Da) proteins, respectively. Southern blots indicate that these are the only two Sac7 protein genes in *S. acidocaldarius*, each present as a single copy. Sac7a, b, and c proteins appear to be carboxy-terminal modified Sac7d species. The transcription initiation and termination regions of the *sac7d* and *sac7e* genes have been identified along with the promoter elements. Potential ribosome binding sites have been identified downstream of the initiator codons. The *sac7d* gene has been expressed in *E. coli*, and various physical properties of the recombinant protein have been compared with those of native Sac7. The UV absorbance spectra and extinction coefficients, the fluorescence excitation and emission spectra, the circular dichroism, and the two-dimensional double-quantum filtered ^{1}H NMR spectra of the native and recombinant species are essentially identical, indicating essentially identical local and global folds. The recombinant and native proteins bind and stabilize double-stranded DNA with a site size of 3.5 base pairs and an intrinsic binding constant of 2×10^7 M^{-1} for poly[dGdC]·poly[dGdC] in 0.01 M KH_2PO_4 at pH 7.0. The availability of the recombinant protein permits a direct comparison of the thermal stabilities of the methylated and unmethylated forms of the protein. Differential scanning calorimetry demonstrates that the native protein is extremely thermostable and unfolds reversibly at pH 6.0 with a T_m of approximately 100 °C, while the recombinant protein unfolds at 92.7 °C.

Small basic DNA-binding proteins have been isolated from various archaea, some of which have been shown to be associated with the nucleoid or chromatin and presumably perform a histone-like or helix-stabilizing function in these organisms (Searcy, 1975; Stein & Searcy, 1978; Searcy & Delange, 1980; Thomm et al., 1982; Grote et al., 1986; Lurz et al., 1986; Choli et al., 1988a,b; Reddy & Suryanarayana, 1989; Sandman et al., 1990), although the actual function of many of these proteins has not been demonstrated. HTa protein from the thermophilic archaeon *Thermoplasma acidophilum* shows considerable homology to eukaryotic histones and *Escherichia coli* HU protein (Searcy, 1975; Searcy & Delange, 1980). Hmf1 and Hmf2, two DNA binding proteins from *Methanothermus fervidus*, are also homologous to some of the eukaryotic histones (Sandman et al., 1990).

Sulfolobus, a thermoacidophilic archaeon, expresses a number of small basic DNA-binding proteins ranging in molecular weight from 7000 to 10 000 (Kimura et al., 1984; Grote et al., 1986; Choli et al., 1988a). These have no apparent homology to any of the histones. Much of the early work on these proteins resulted from a search for chromatin proteins that might stabilize the genomic DNA at the high growth temperature. *Sulfolobus acidocaldarius* grows optimally in the range of 70−80 °C, while *Sulfolobus solfataricus* grows optimally at approximately 75−85 °C. The G+C base composition of *Sulfolobus* DNA is about 40%, and its cellular salt concentration is relatively low, making a helix-stabilizing protein presumably necessary (Reddy & Suryanarayana, 1988). The 7 kDa class of proteins has been presented as a likely candidate given that they are present in relatively large amounts in the cell (Grote et al., 1986; Choli et al., 1988a,b).

Five proteins have been isolated in the 7 kDa class from *S. acidocaldarius* (Kimura et al., 1984; Choli et al., 1988b), and have been labeled Sac7a[1] through Sac7e, in order of increasing basicity. Four of these, Sac7a, b, d, and e, have been sequenced (Figure 1) (Kimura et al., 1984; Choli et al., 1988b), and only minor differences among them have been noted. The sequence of Sac7c has not been reported. The number of genes encoding the 7 kDa proteins of *S. acidocaldarius* has not been determined. Comparison of the

† This work was supported by the Biotechnology Research and Development Corporation (J.W.S. and S.P.E.) and the National Institutes of Health (GM49686) (J.W.S. and S.P.E.). A preliminary report of this work was given at the Swedish Biophysical Society Meeting, June 4−6, 1990, Lovanger, Sweden, and at the Biophysical Society Meeting, Feb 9−13, 1992, Houston, TX.

* Authors to whom correspondence should be sent. Phone: 618-453-6479 or 618-453-6466. Fax: 618-453-6440. E-mail: rgupta@som.siu.edu or jshriver@som.siu.edu.

‡ Present address: Vanderbilt University, Department of Molecular Biology, 1161 21st Ave. South, Nashville TN 37235.

§ Present address: Department of Pharmacology and Molecular Biology, Chicago Medical School, N. Chicago, IL 60064-3095.

® Abstract published in *Advance ACS Abstracts,* July 15, 1995.

[1] Abbreviations: DSM, Deutsche Sammlung für Mikroorganismen; IPTG, isopropyl β-D-thiogalactopyranoside; NMR, nuclear magnetic resonance; COSY, correlation spectroscopy; DQF-COSY, double-quantum filtered correlation spectroscopy; DSC, differential scanning calorimetry; CD, circular dichroism; Sac7, a group of 7 kDa DNA-binding proteins from *Sulfolobus acidocaldarius,* individually referred to as Sac7a, Sac7b, Sac7c, Sac7d, and Sac7e, in order of increasing basicity; Sso7, a group of 7 kDa DNA-binding proteins from *Sulfolobus solfataricus.*

0006-2960/95/0434-10063$09.00/0

amino acid sequences indicates that there must be at least two separate genes coding the 7d and 7e species. The high degree of similarity observed in the primary sequence of the 7d and 7e proteins suggests that two genes arose through gene duplication. Sac7a and Sac7b are truncated versions of the Sac7d protein, most likely resulting from truncated genes, posttranslational processing, or degradation during isolation.

Specific ϵ-aminomonomethylation of lysines 4 and 6 is characteristic of the Sac7a, b, and d proteins, while Sac7e is monomethylated at lysines 6, 62, and 63 (residue 4 is an arginine in Sac7e) (Kimura et al., 1984; Choli et al., 1988b). No lysine methylation has been detected in the C-terminus of Sac7a, b, or d, presumably since there are no lysines at positions 62 and 63 in these proteins, although Sac7d contains lysines at positions 64 and 65. The Sso7d protein from *S. solfataricus* is monomethylated at lysines 4 and 6 and also at lysines 62, 64, and 65 (Choli et al., 1988a). The role of lysine monomethylation has not been determined but is most likely nontrivial given the specificity (there are 12–14 lysines in these proteins) and the occurrence in both *S. acidocaldarius* and *S. solfataricus* proteins. Baumann et al. (1994) have recently shown that an increase in Sso7d methylation occurs upon heat shock and indicate that methylation may be directly related to protein stability. However, methylation may be an incidental response to an increase in methylase activity directed at other processes. Methylation may also increase the reversibility of the unfolding process rather than changing the stability. A direct calorimetric measurement of the unfolding and stability of these proteins has not been reported.

The Sac7 proteins would appear to be ideal models for studies of protein folding and stability given their small size, the absence of cysteine, and expected high thermostability. Biophysical analyses of these proteins is hampered, however, by the inability to selectively isolate a homogenous isoform in large quantities. The differential methylation of individual 7 kDa proteins could further complicate quantitative studies of structure and stability as well as DNA binding. Therefore, we have cloned and expressed the gene encoding the Sac7d species in *E. coli* to facilitate elucidation of the solution structure of the protein by NMR with high resolution, probing of the thermostability and DNA-binding properties of the protein by site-directed mutagenesis, and determination of the role of methylation. The availability of recombinant protein allows for a direct comparison of the stability of the methylated and unmethylated proteins. In the process of cloning the *sac7d* gene, the gene for Sac7e has also been cloned and sequenced; and we have delineated the transcription initiation and termination regions of the *sac7d* and *sac7e* genes along with the promoter elements.

An initial structure of the native Sso7d protein has been recently published by Baumann et al. (1994), and a high-resolution structure of the homogeneous, recombinant Sac7d protein has been completed (Edmondson, Qiu, and Shriver, manuscript submitted). There are significant differences between these structures, and it remains to be determined if these can be attributed to sequence differences, lysine methylation, or quality of data due to heterogeneity in the native preparation. The spectroscopic, DNA binding, and calorimetric comparisons of the native and recombinant Sac7 proteins reported here indicate little difference in structure, but significant difference in thermostability.

MATERIALS AND METHODS

Strains of Microorganisms. *E. coli* strain DH5αF'IQ [F' *lacI*^{q}ZΔM15/Δ (*lacZYA-argF*) *recA*1 *hsdR*17(r_k^- m_k^+)] was purchased from Gibco BRL. *E. coli* strains HMS174 (F$^-$ *recA* r^-_{k12} m^+_{k12} Rif^r), BL21 (F$^-$ *ompT* r^-_B m^-_B), and their derivatives were generous gifts from F. William Studier (Studier et al., 1990). *E. coli* strain CJ236 (*dut*$^-$ *ung*$^-$) was obtained from Jack Parker (Southern Illinois University, Carbondale, IL). *S. solfataricus* P2 and *S. acidocaldarius* DG6 were gifts from Dennis Grogan (Grogan, 1989, 1991). *S. acidocaldarius* (DSM 639) and *S. solfataricus* P1 (DSM 5354) were purchased from Deutsche Sammlung für Mikroorganismen (DSM).

The *Sulfolobus* strain used here was received from W. Zillig (originally called *S. solfataricus* P1). We have isolated a single colony of our organism on solid medium (Grogan, 1989) and have compared the *Hin*dIII, *Eco*RI, and *Sal*I restriction fragment patterns of its genomic DNA with two strains of *S. acidocaldarius* (DG6 and DSM639) and two strains of *S. solfataricus* (DSM5354 and P2) according to Grogan (1989). In each case the restriction pattern of our organism is identical to the *S. acidocaldarius* strains and is distinctly different from the *S. solfataricus* strains. This has been further substantiated by Southern analysis of genomic DNA using Sac7 protein gene specific oligonucleotides (see Results). We have designated our laboratory strain as *S. acidocaldarius* RGJM. There has been confusion in the literature regarding the identity of the strains of two *Sulfolobus* species used in various laboratories at different times. Zillig (1993) has recently addressed this issue and tried to clarify the confusion.

Growth of Microorganisms. *E. coli* strains were grown in Luria Bertani media (1% bactotryptone/1% NaCl/0.5% yeast extract) by standard methods (Sambrook et al., 1989). Small scale cultures of *Sulfolobus* (10–200 mL) were grown in Brock's medium (Brock et al., 1972) at 75 °C, supplemented with 0.2% sucrose. Large scale *Sulfolobus* cultures were grown either in 10 L polypropylene carboy at 78 to 80 °C or in a 16 L VirTis glass fermentor at 70–72 °C with vigorous aeration using DeRosa's medium (DeRosa & Gambacorta, 1975) supplemented with 0.1% glucose and 0.1% glutamic acid.

Enzymes and Chemicals. Restriction enzymes, alkaline phosphatase, T4 DNA ligase, T4 DNA polymerase and T4 polynucleotide kinase were purchased from New England Biolabs, Brisco Ltd., BRL, or United States Biochemical Co. [^{32}P]H_3PO_4 and 5'-[α-^{35}S]adenosine thiotriphosphate triethylammonium salt were purchased from ICN Biochemicals Inc. and Amersham Co., respectively. Sequenase version 2.0 DNA sequencing kit was obtained from United States Biochemical Co. Specific deoxyoligonucleotides were purchased from Research Genetics. The list of the oligonucleotides used in this work is presented in Table 1. Difco bacterial media were purchased from Fisher Scientific. CM52 was obtained from Whatman and Sephacryl S-100-HR from Sigma Chemical Co. All other chemicals were reagent grade and obtained primarily from Fisher Scientific, J. T. Baker Co., and Sigma Chemical Co. Laboratory water was routinely purified to 18.3 MΩ resistance with a recycling Barnstead Nanopure system.

Genomic DNA Isolation. Cells from 10–20 ml cultures were pelleted and resuspended in 0.2–0.3 mL of 10 mM

Tris-HCl, pH 8.0/1 mM EDTA/1% SDS. This solution was extracted once each with equal volumes of phenol, phenol/chloroform/isoamylalcohol (25:24:1), and chloroform/isoamyl alcohol (24:1). Sodium acetate (3 M, pH 5.2) was added to the final aqueous phase to a concentration of 0.3 M, followed by DNA precipitation with three volumes of ice-cold ethanol. The DNA was spooled onto a thin glass rod, washed in 70% ethanol, and air dried. The DNA was dissolved in 10 mM Tris-HCl, pH 8.0/1 mM EDTA.

Cloning, Hybridization, and Sequencing. The preparation of a *Pst*I genomic library of *S. acidocaldarius* RGJM in *E. coli* strain DH5αF′IQ and screening of the library by colony hybridization was according to published procedures (Berger & Kimmel, 1987; Sambrook et al., 1989). Southern and dot blot hybridizations were carried out using nitrocellulose membranes according to the manufacturer's protocols (Schleicher & Schuell) which are based on the method of Southern (1975). The preparation of [γ -^{32}P]ATP and 5′ ^{32}P-end-labeling of oligonucleotides was by standard methods (Johnson & Walseth, 1979; Gupta, 1984; Sambrook et al., 1989). DNA was sequenced by the dideoxy chain termination method (Sanger et al., 1977) using a Sequenase version 2.0 kit. The final sequences were determined from both strands. The standard universal primers for Stratagene's pBluescript vectors (Short et al., 1988) and specifically synthesized oligonucleotides were used in sequencing reactions. DNA sequences were analyzed using the computer program DNA Inspector IIe (Textco Co.).

Primer Extension. Total RNA from *S. acidocaldarius* RGJM was isolated by previously published procedures (Emory & Belasco, 1990). The primer extension assay was conducted as described in the Promega "Protocol and Applications" manual.

Oligonucleotide-Directed Mutagenesis. Procedures for the oligonucleotide directed mutagenesis were those outlined in the Bio-Rad Muta-Gene manual and are based on Kunkel's method (Kunkel et al., 1987) using *E. coli dut*$^-$*ung*$^-$ strains. We were unable to propagate the substrate for oligonucleotide directed mutagenesis, pBluescript KS+/*sac7d* (see Results for the description and nomenclature of the plasmids), in *E. coli* strain CJ236 (*dut*$^-$ *ung*$^-$). Therefore, we used DH5αF′IQ as the host cell for the production of single-stranded template and as the recipient for transformation with mutagenized plasmid and modified the procedure for the selection of mutant plasmid. Colonies arising from transformation with the plasmids from the mutagenesis reaction to create the *Nde*I site were pooled and grown as a mixed culture. Plasmids isolated from these cells were digested with *Nde*I and separated on a 0.8% agarose gel. Linear plasmids were isolated from the gel, recircularized, and again used to transform DH5αF′IQ. Plasmids were then extracted from individual colonies and screened for the presence of an *Nde*I restriction site by digestion with the enzyme. Final confirmation of the desired mutation in the plasmids was obtained by sequencing.

Gene Expression. For gene expression, pET-3b/*sac7d* was transformed into *E. coli* strain BL21 (DE3) pLysS (Studier et al., 1990). For protein isolation, a 10 mL culture of this transformant was grown overnight in LB broth containing ampicillin (200 μg/mL) and chloramphenicol (27 μg/mL). From this, 0.6−1 mL was used to inoculate 50 mL of fresh medium. At an A_{600} of 0.3−0.6, 25 mL of the culture was diluted into 1 L of new medium. The culture was induced upon reaching an A_{600} of 0.8−0.95 by adding IPTG to a final concentration of 0.4 mM. A small aliquot of each culture was taken prior to induction to assay for expression and plasmid stability as described by Studier et al. (1990). Cultures were harvested at 1 h postinduction and stored at −70 °C.

Protein Isolation and Purification. E. coli cells containing recombinant protein were thawed slowly and resuspended in 100 mL of 10 mM Tris-HCl, pH 7.5/0.5 mM phenylmethanesulfonyl fluoride, and the cells were lysed by repeated freezing and thawing along with brief sonication on ice. To isolate native protein, *Sulfolobus* cells were suspended in 0.05 M KH_2PO_4 buffer (pH 6.8) and lysed by sonication on ice. DNase I (20 mg/100 mL) was added to lysed cells, and the suspension was incubated at 37 °C for 5 min followed by centrifugation at 280000*g* for 60 min. The supernatant was cooled on ice and dialyzed in SpectraPor CE 1000 MWCO tubing against 0.2 M H_2SO_4 overnight at 4 °C. The resulting precipitate was removed by centrifugation at 180000*g* for 30 min, and the supernatant was dialyzed four times against 20 mM Tris-HCl, pH 7.4/1 mM EDTA. A small amount of precipitate was removed by centrifugation, and the supernatant was applied to a CM-52 ion exchange column equilibrated with 20 mM Tris-HCl (pH 7.4). The protein was eluted with a linear NaCl gradient (0.0−0.3 M) with both the native and recombinant Sac7 proteins giving a primary peak at approximately 0.2 M NaCl. Further purification was accomplished by gel exclusion chromatography on Sephacryl S-100-HR in 0.02 M Tris-HCl (pH 7.4).

The identity and purity of the 7 kDa proteins were monitored by nonreducing SDS gel electrophoresis (Schägger & von Jagow, 1987). The recombinant protein showed a single band that comigrated with the mixture of Sac7 native proteins isolated from *S. acidocaldarius* (Figure 2) and was absent in preparations from control *E. coli* cells lacking the recombinant plasmid (data not shown). The Sso7 proteins run slightly ahead of Sac7 proteins, consistent with a molecular weight of 7020 (calculated from the sequence). The Schägger−von Jagow gel used here did not resolve the individual Sac7 and Sso7 native species. The identity of the recombinant Sac7d protein was confirmed by comparison of the double-quantum filtered COSY spectra of native Sac7 and recombinant Sac7d proteins (see below) and by the consistency of the sequence specific ^{1}H NMR assignments with the expected sequence (Edmondson, Qiu, and Shriver, submitted).

In earlier studies the recombinant protein was isolated by a different procedure (McAfee, 1993). *E. coli* cells were lysed and DNase treated as above but without sonication. The pH of the supernatant was adjusted to 1.5 with 5 M H_2SO_4. After 45 min on ice and centrifugation, the supernatant was neutralized with 10 N NaOH. The mixture was incubated in a water bath at 70 °C for 2 h, followed by centrifugation. The supernatant was dialyzed three times with 1 mM NaH_2PO_4 buffer (pH 7.0) followed by CM-52 chromatography as above.

Molecular Weight Determination. Approximate molecular weights of the native and recombinant Sac7 proteins were determined by gel exclusion chromatography on Sephacryl S-100-HR. Cytochrome *c*, myoglobin, carbonic anhydrase, and bovine serum albumin were used as molecular weight standards, and blue dextran and DNP-alanine were used to measure the column void and total volumes, respectively.

The molecular weights were determined as described by Mayes (1984).

Phosphorylation and Glycosylation Assays. Phosphate analysis was performed by the method of Fiske and Subbarow (Fiske & Subbarow, 1925; Leloir & Cardini, 1957). Small aliquots of Sac7 (0.95 mL of a 0.5 mg/mL solution in 0.02 M Tris-HCl, pH 7.0) were incubated at 37 °C for 1 h with 0.05 mL of bovine intestinal alkaline phosphatase (2.5 mg/mL in 0.01 M Tris-HCl, pH 9.8). The protein was precipitated with 0.10 mL of concentrated perchloric acid, incubated on ice for 10 min, and centrifuged for 5 min at 13 000 rpm. To 0.90 mL of supernatant was added 2.0 mL of distilled water, 1.0 mL of 5 N H_2SO_4, 1.0 mL of 2.5% ammonium molybdate, and 0.10 mL of reducing agent [prepared fresh by dissolving 0.25 g of reducing mixture (sodium bisulfite, sodium sulfite, and 1-amino-2-naphthol-4-sulfonic acid in a 46:46:8 ratio) in 10 mL of water]. The solutions were allowed to stand for 20 min, and the absorbance was measured at 660 nm. A standard curve was prepared using known amounts of a 0.01 M KH_2PO_4 solution. *O*-Phosphoserine, treated with alkaline phosphatase as described for Sac7 gave quantitative recovery of phosphate.

The phenol−sulfuric acid reaction was used to assay carbohydrate content of Sac7 protein (Debois et al., 1956; Hirs, 1967). To 1.0 mL aliquots of Sac7 protein solution (0.3 mg/mL) was added 0.25 mL of 80% phenol and 2.5 mL of concentrated sulfuric acid. After mixing, the solutions were left at room temperature for 10 min and then placed in a 25 °C water bath for 20 min. The absorbance was measured at 489 nm. Known amounts of α-D-glucose were used to construct a standard curve.

Protein Extinction Coefficient. Ultraviolet and visible spectra were recorded on a Cary 210 spectrophotometer at 25 °C. The wavelength accuracy was checked using benzene vapor and found to be accurate to within ±0.3 nm, and the absorbsance accuracy was checked using potassium chromate in 0.05 M KOH (Gordon & Ford, 1972) and found to be accurate to within 1%.

The extinction coefficients of both the native Sac7 and recombinant Sac7d proteins were determined by measuring the amino acid concentration using the ninhydrin reaction (Moore & Stein, 1954) for a sample of known absorbance. A standard curve was prepared using amino acid standard H (Pierce Biochemicals) and converted into leucine molar equivalents. The concentration of amino acid standards was checked using tyrosine with an extinction coefficient of $\epsilon_{274.5}$ = 1340 in 0.1 M HCl. The molar concentration of amino acid residues in the samples was calculated by dividing leucine equivalents by the average color yield based on the amino acid composition (Moore & Stein, 1954). The average color yields for Sac7d, lysozyme, and RNase A were 1.0, 1.05, and 1.06, respectively. The extinction coefficients of lysozyme and RNase A standards were checked by this procedure and found to be within 1% of published values. The procedure gave an extinction coefficient of 1.03 ± 0.05 mL/(mg·cm) for both native and recombinant proteins.

The extinction coefficients were also determined by the method of van Iersel et al. (1985) immediately following chromatography of the proteins on Sephadex G-50 in 0.01 M NaH_2PO_4 buffer (pH 6.5). A flat (±0.0005 absorbance units) spectrophotometer baseline was programmed using the same buffer which had been used to equilibrate the column. Protein spectra were collected on samples directly from the gel exclusion column, generally using only those samples with an absorbance less than 2.0 at 205 nm to minimize the effects of stray light. The reproducibility of the A_{280}/A_{205} ratio using different aliquots collected through the protein peak as it eluted from the column was found to be on the order of 99%. The linear relationship between the extinction coefficient at 280 nm and the ratio of the absorbance at 280 and 205 nm was confirmed in our hands using bovine α-chymotrypsin (Worthington), hen egg white lysozyme (Sigma), bovine pancreatic ribonuclease A (Sigma), avidin (Sigma), β-lactoglobulin (Sigma), and bovine serum albumin (Sigma). A linear fit of the standards yielded a standard curve such that

$$\epsilon_{280}^{0.1\%} = 35.76\frac{A_{280}}{A_{205}} - 0.04$$

with a correlation coefficient of 0.999 and a standard deviation for the slope of 0.62 and 0.03 for the *y* intercept. The extinction coefficients for the native and recombinant protein were found to be identical with this technique at 1.18 mL/(mg·cm) with a standard deviation of 0.008 mL/(mg·cm).

The extinction coefficients were also calculated to be 1.08 mL/(mg·cm) in 6 M guanidine hydrochloride, based on the amino acid content of the protein using the procedure of Edelhoch (Edelhoch, 1967; Gill & von Hippel, 1989) assuming ϵ_{Tyr} = 1280 M^{-1} cm^{-1}, ϵ_{Trp} = 5690 M^{-1} cm^{-1} in 6 M guanidine hydrochloride. An increase in absorbance of 3.5% was noted upon denaturation of the protein with 6 M GdnHCl, so the calculated extinction coefficient of the folded protein was corrected to 1.05 mL/(mg·cm). The estimated error was taken to be ±0.04 with a maximal error of ±0.15 (Gill & von Hippel, 1989).

Circular Dichroism. Circular dichroism spectra of purified native Sac7 and recombinant Sac7d proteins were measured at room temperature in a 0.01 cm path length cylindrical cell on an AVIV 62DS spectropolarimeter. CD data were collected at 1 nm intervals using averaging times of 15−30 s/nm, depending on the signal-to-noise ratio. Relatively high signal-to-noise ratios made signal averaging of multiple scans unnecessary. The spectral bandwidth was 1.5 nm. Baselines were measured using water and subtracted from the sample CD. Sample concentrations ranged from 0.2 to 0.7 mg/mL. Protein concentrations were determined from UV absorption spectra measured in 1 cm cuvettes. The molar CD per peptide bond was determined using standard procedures (Johnson, 1984) along with the UV extinction coefficient determined above. CD spectra were smoothed as described by Savitsky and Golay (1964). The CD was calibrated at 290.5 nm with *d*-camphor-10-sulfonic acid using $\Delta\epsilon^{0.1\%}$ = 2.36, and the ratio $\Delta\epsilon_{192.5}/\Delta\epsilon_{290.5}$ was −2.10 (Chen & Yang, 1977).

The fractions of protein secondary structures were determined by fitting the CD spectra from 260 to 184 nm in 2 nm intervals using the variable selection method of Johnson (Manavalan & Johnson, 1987). The results reported are the averages plus or minus one standard deviation of all possible combinations of 22 reference proteins taken 19 at a time that (1) have secondary structure components greater than −0.05, (2) have sums of secondary structures between 0.9 and 1.1, and (3) have an rms error between measured and calculated CD spectra less than 0.21 $\Delta\epsilon$ units. The number

of fits meeting this selection criteria were greater than 250 for native and recombinant protein.

Nuclear Magnetic Resonance. NMR spectra were collected on a Varian 500 MHz NMR spectrometer with the magnet installed on a TMC Micro-g triangular antivibration table. All data were collected at 35 °C in 90% H_2O/10% D_2O, pH 4.1, with a protein concentration of approximately 10 mM. The pH was adjusted with DCl and NaOD using a Radiometer glass electrode and was not corrected for the deuterium isotope effect (Bundi & Wüthrich, 1979). The chemical shifts are referenced to the water resonance at 4.73 ppm at 35 °C [measured relative to sodium 4,4-dimethyl-4-silapentane sulfonate (DSS) in a separate experiment without protein].

Phase-sensitive double-quantum filtered COSY (DQF-COSY) spectra were collected using standard procedures (Rance et al., 1983). Typically, 1024 data points were collected in the t_2 domain with 512 increments in the t_1 domain, each the sum of 32 scans with a 3 s relaxation delay. The spectral widths in both dimensions was 6000 Hz. The water peak was diminished in all experiments by presaturation during the relaxation delay. Both carrier and decoupler frequencies were set equal to the water resonance frequency in all experiments (Zuiderweg et al., 1986).

The NMR data were transferred to a Silicon Graphics workstation for Fourier transformation and further data manipulation using FELIX 2.1 (BioSym). The data were zero-filled to 2048 data points in both dimensions and treated with a Lorentzian to Gaussian apodization function prior to Fourier transformation.

Differential Scanning Calorimetry. Differential scanning calorimetry was performed with a Microcal MC2 calorimeter. Temperature calibration was monitored using sealed samples supplied by Microcal. Heat flow accuracy was periodically monitored by applying pulses of known magnitude using the internal heater. In addition, ribonuclease A (Sigma, R5250) was used as a benchmark test protein and shown to unfold at pH 2.2 [0.1 M KCl, 0.02 M glycine, ϵ_{280} = 0.69 mL/(mg·cm), MW 13 700] with a T_m of 36.0 °C, a ΔH_{cal} of 74.1 kcal/mol, and a ΔH_{vh} of 74.8 kcal/mol ($\Delta H_{cal}/\Delta H_{vh}$ ratio of 1.00 ± 0.01), in good agreement with the published values of Tiktopulo and Privalov (1974).

Protein solutions were exhaustively dialyzed against the indicated buffer overnight. The sample cell was loaded with 1.229 mL of protein solution, and the reference cell was filled with the last dialysis buffer. Approximately 30 psi of nitrogen was applied to the cells during each scan to minimize degassing during heating. Samples were not degassed, but, instead, the sample was heated repetitively three times in the DSC instrument by scanning to 35 °C (i.e., below any denaturation endotherm), followed by rapid cooling. This procedure resulted in the flattest and most reproducible instrumental baselines.

All DSC experiments were under computer control using an IBM PC computer interfaced to the Microcal MC2 instrument. A scan rate of 1 deg/min was used in all experiments. The computer interface and data collection software were supplied by Microcal. Multiple, repetitive scans were performed on the same sample to check for reversibility, with identical cooling and equilibration times between scans.

The DSC raw data, in the form of heat flow (mcal/min) as a function of temperature, was transferred to a Macintosh Quadra computer for analysis. The raw data were converted to excess heat capacity (kcal/deg·mol) by dividing each data point by the scan rate and the concentration of protein in the sample cell. All baselines were corrected by subtraction of DSC scans of the buffer against which the protein had been dialyzed. The heat capacity data was fit by using in-house nonlinear least-squares fitting routines to obtain the midpoint temperature of the transition and both the calorimetric and van't Hoff enthalpies. The basis of the programs has been described elsewhere (Shriver & Kamath, 1990).

Fluorescence. Fluorescence titration measurements were performed on an SLM 8000C spectrofluorimeter with 4 nm excitation and 8 nm emission slit widths. Binding titrations were performed with excitation at 295 nm and emission monitored at 350 nm. Reverse titrations were performed by adding aliquots of concentrated nucleotide solutions to a known concentration of protein in a 4 mL fluorescence quartz cell with stirring using a magnetic "flea" within the cell. Nucleic acid concentrations were determined spectrophotometrically using an extinction coefficient of 8400 L/(cm·mol) for poly[dGdC]·poly[dGdC] (Wells, 1970) and 6600 L/(cm·mol) for poly[dAdT]·poly[dAdT] (Inman, 1962). All experiments were performed at 25 °C. The fluorescence intensity was constant at high DNA concentrations, and thus no correction was made for the inner filter effect. Apparently, any decrease in fluorescence due to the inner filter effect was balanced by other effects, such as scattering by the DNA-protein complexes. Photobleaching was not observed during the titrations. Binding parameters were obtained by using a simple, noncooperative McGhee–von Hippel model (McGhee & von Hippel, 1974).

DNA Stabilization. Thermal denaturation studies of DNA and DNA–protein complexes were performed on a Cary 210 spectrophotometer equipped with water-jacketed cuvette holders and a circulating water bath calibrated to within ±0.3 °C. Melting curves are scaled to an A_{262} of 1.0 at 20 °C for the DNA component of DNA–protein mixtures.

Sequence Analysis. BLAST (Altshul et al., 1990) searching and alignment were performed using the NCBI server (blast@ncbi.nlm.nih.gov) against the "nr" (nonredundant) sequence database (including Brookhaven Protein Data Bank, January 1994 release; SWISS-PROT Release 29.0, June 1994; PIR Release 41.0, June 30, 1994; CDS Translations from GenBank Release 83.0, June 15, 1994, Kabat Sequences of Proteins of Immunological Interest Release 5.0, August, 1992; TFD Transcription Factor Database Release 7.0, June 1993). BLITZ and FASTA searches of the latest SWISS-PROT database were performed using the EMBL servers (blitz@embl-heidelberg.de and fasta@embl-heidelberg.de). Database retrieval was performed using the GDB/Accessor (Johns Hopkins University) available from ftp.gdb.org. MacPattern (Fuchs, 1991) (fuchs@embl-heidelberg.de) was utilized for BLOCKS (Henikoff & Henikoff, 1991) and PROSITE (Bairoch, 1992) analysis on a Quadra 700 (BLOCKS database Version 7.01 was utilized with 2679 entries and PROSITE database version 12.0, June 1994, was used with 1021 entries, both obtained from the NCBI ftp site ncbi.nlm.nih.gov.) The MacVector software package (IBI) was utilized for protein secondary structure analysis.

Table 1: List of Oligonucleotides

oligonucleotide[a]	sequence[b]	position[c]
A	NACYTCYTTYTCYTCNCC	230−247
B	GGGAGCTTYAARTAYAARGGNGARGA[d]	218−237
C	GGGGTACCRTTRTCRTCRTANGTRAA[d]	296−317
D	TCTTAACAAATTATTTTATTT	398−418
E	GCCCTTTATACCTTCCCCTTA	398−418
F	CCTGTCTTACCATTGTCGTC	305−324
G	CCTTCACCATATGAGGTCAAGTTATC[e]	187−212
H	GACTTAACTTAATACCG	143−159

[a] Oligonucleotides A, B, and C were derived from amino acids 9−14, 5−11, and 31−38, respectively, of the Sac7 proteins (Figure 1). These amino acid sequences are identical in the four Sac7 proteins. [b] N = A, G, C, or T; Y = C or T; R = A or G. [c] Nucleotide positions correspond to those in Figure 3. Sequences of oligonucleotides A, C, D, E, F, and G are complementary to the sequences shown in Figure 3. Oligonucleotides D and E correspond to the same positions (Figure 3) for *sac7d* and *sac7e*, respectively. [d] Oligonucleotides B and C have six and four additional nucleotides, respectively, at the 5′ termini which are not derived from the amino acid sequence of the protein. [e] Sequence of the primer used for oligonucleotide directed mutagenesis. The underlined G replaces a T in the *sac7d* gene sequence creating an *Nde*I restriction site.

RESULTS

Gene Cloning and Sequence. *Pst*I digested genomic DNA of *S. acidocaldarius* RGJM was shotgun cloned in the vector pUC19 and transformed into *E. coli*, DH5αF′IQ. Approximately 10 000 transformants were screened by colony hybridization to a mixed oligonucleotide probe (oligonucleotide A, Table 1) derived from residues 9−14 of the published amino acid sequence of the *S. acidocaldarius* 7 kDa proteins (Kimura et al., 1984; Choli et al., 1988a). [The published amino acid sequences for Sac7a, b, d, and e are identical over this range (Figure 1) as well as over the ranges for oligonucleotides B and C.] Tentative positive clones were restreaked onto selective media and screened a second time with the same probe. Plasmids isolated from a number of these positive clones were then independently hybridized to three different mixed probes (oligonucleotides A, B, and C, Table 1) by dot blot hybridization. Two clones were isolated which hybridized to all three probes. Plasmids isolated from these cells were partially sequenced using oligonucleotide B as a primer. One of the genes corresponded with the published protein sequence for the carboxy-terminal half of the Sac7d protein of *S. acidocaldarius* (Kimura et al., 1984; Choli et al., 1988a) with the exception of one additional lysine at the carboxy terminus, and the other corresponded to the Sac7e sequence. The genes which matched the Sac7d and 7e proteins have been designated *sac7d* and *sac7e*, respectively.

Agarose gel analysis of the plasmids carrying the *sac7d* (pUC19/*sac7d*) and *sac7e* (pUC19/*sac7e*) genes indicated that the cloned *Pst*I fragments were greater than 15 kb in size. Southern blot hybridizations of oligonucleotide C to the restriction digests of pUC19/*sac7d* indicated that *sac7d* gene was present on a slightly less than 800 bp *Eco*RI fragment. Preliminary sequencing of pUC19/*sac7d* using oligonucleotide B as a primer indicated the presence of an *Eco*RI site 61 bases downstream of the termination codon of the protein. Since the published sequence of Sac7d protein consists of 64 amino acids (Kimura et al., 1984; Choli et al., 1988a), the second *Eco*RI site was expected to be upstream of the start codon. Thus, the *Eco*RI fragment hybridizing to probe C was expected to contain the entire coding region of the gene. This *Eco*RI fragment was subcloned in the vector pBluescript KS+ to produce pBluescript KS+/*sac7d*, and the sequence of *sac7d* gene was determined (Figure 3). The sequence of the *sac7e* gene (Figure 3) was obtained directly from the pUC19/*sac7e* using primers complementary to the coding region of the gene.

The GenBank accession numbers for the *sac7d* and *sac7e* gene sequences reported here are M87569 and L08891, respectively.

Sequence Analysis and Gene Copy Number. The start of transcription for both *sac7d* and *sac7e* genes was determined using primer extension analysis (Figure 4). Specific primers (oligonucleotides D and E, Table 1) that were complementary to residues 398−418 (Figure 3) of the two genes were used. A single start site was observed for each of the two genes which occurs on a guanosine residue eight nucleotides upstream from the initiation codon. These guanosine residues are present within perfect archaeal "B box" consensus sequences (consensus $\frac{A}{T}TG\frac{A}{C}$ (Zillig et al., 1988). A sequence resembling the archaeal "A-box" motif (consensus $TTTA\frac{A}{T}A$) is seen 24 and 23 nucleotides upstream from the transcription start site for the *sac7d* and *sac7e* genes, respectively (Figure 3). The "A-box" of *sac7d* has a five base match with the consensus sequence, while that for the *sac7e* has only four matches.

Oligonucleotide F (Table 1) was used to probe genomic blots of three *S. acidocaldarius* (RGJM, DG6, and DSM639) and two *S. solfataricus* (DSM5354 and P2) strains (Figure 5A). Oligonucleotide F is complementary to a region coding for residues 34−40 (Figure 1) which are identical for all the *S. acidocaldarius* 7 kDa proteins (DDNGKTG) and significantly different from that of *S. solfataricus* (DEGGGKTG, two substitutions and an insertion). Two *Hin*dIII restriction fragments (~3.0 and ~4.6 kb) were recognized by the probe in all three *S. acidocaldarius* strains, while no hybridization to the *S. solfataricus* strains was observed. This observation reinforces the assignment of the RGJM strain (our laboratory strain) as an *S. acidocaldarius* strain. The results indicate that the putative genes encoding all of the Sac7 proteins are present on the two *Hin*dIII restriction fragments of ~3.0 and ~4.6 kb in size. Genomic blots of *Eco*RI, *Hin*dIII, and *Pst*I digested *S. acidocaldarius* RGJM DNA were also probed with the common oligonucleotide F (Figure 5B), and in each case hybridization to two bands was observed. One band in each hybridized to oligonucleotide H, specific for an untranscribed region upstream of the *sac7d* gene (Figure 5C). Results of the hybridizations of various restriction digests of the original pUC/*sac7d* and pUC/*sac7e* clones to appropriate oligonucleotides (data not shown) corroborated the results in Figure 5 and also indicated that the original clones had a single copy of a *sac7* gene. The 3.0 and 4.6 kb *Hin*dIII fragments can be correlated with the *sac7d* and *sac7e* genes, respectively. The data indicate that there are only two *sac7* genes in *S. acidocaldarius* genome, each being present as a single copy. This reinforces the conclusion that Sac7a and Sac7b are proteolytically truncated versions of the Sac7d protein.

Protein Sequence Analysis. The *sac7d* open reading frame can encode a 66 amino acid protein with a calculated molecular weight of 7608, and the *sac7e* encodes a 65 amino acid protein with a calculated molecular weight of 7469

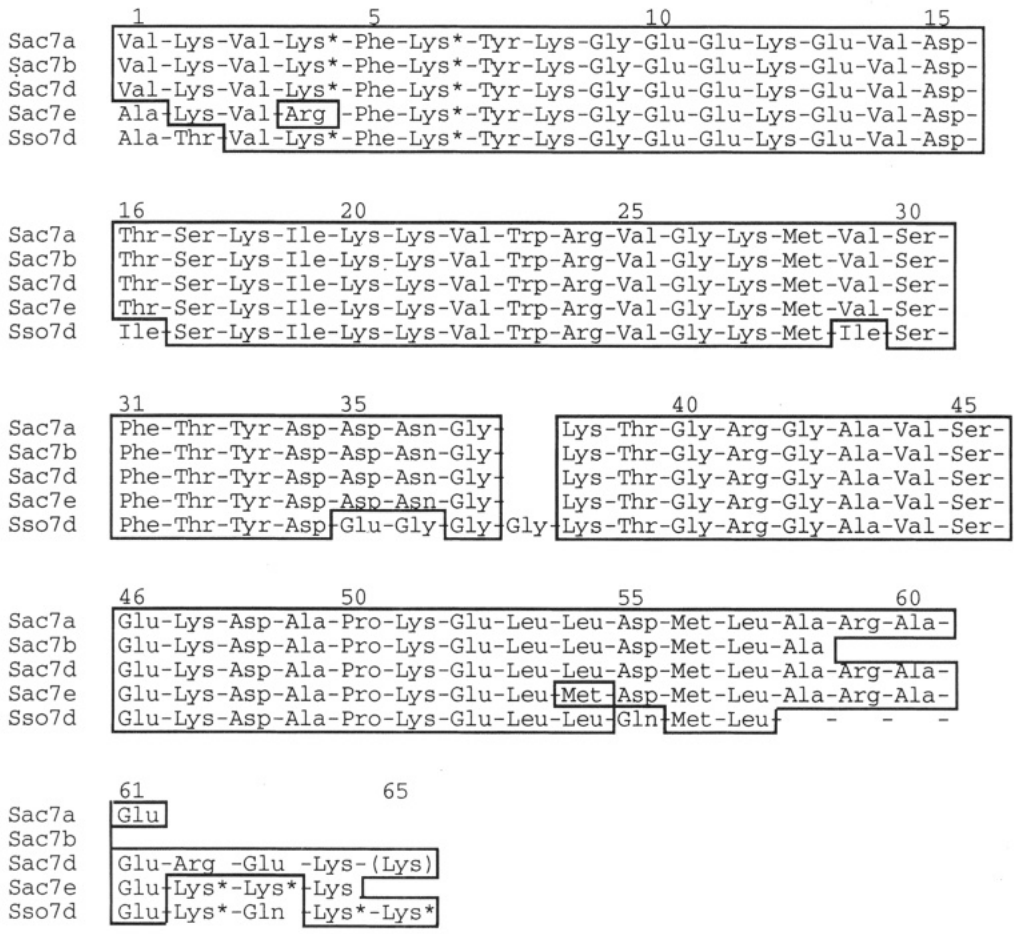

FIGURE 1: Amino acid sequences of the Sac7a, b, d, and e proteins [after Kimura et al. (1984) and Choli et al. (1988b)] and the Sso7d protein [after Choli et al. (1988a)]. [Note that the sequence reported by Kimura et al. (1984) was claimed to be for Sso7d but was later shown to be for Sac7d (Choli et al., 1988a).] Numbering is according to the Sac7d sequence without the initiator methionine. Regions homologous to the Sac7d protein are outlined. Sac7a, b, and d differ only in length. Lysines which are monomethylated to some extent in the native protein are indicated with asterisks. The additional C-terminal lysine coded by the *sac7d* gene described here which was not indicated in the published protein sequence is enclosed in parentheses.

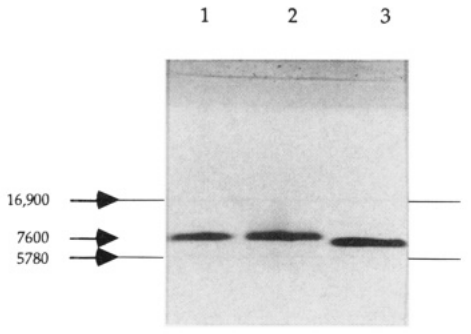

FIGURE 2: Schägger and von Jagow (1987) polyacrylamide nonreducing SDS gel of purified native Sac7 proteins (lane 1), recombinant Sac7d (lane 2), and native Sso7 (lane 3) proteins stained with Coomasie Brilliant Blue G-250 (Bio-Rad). The molecular weight of the Sso7 protein is 7019 based on the published protein sequence (Choli et al., 1988a), while that of the Sac7d is 7608 based on the DNA sequence presented here. The band positions of myoglobin (MW 16 900) and insulin (MW 5780) are indicated for comparison.

(including initiator methionines). Secondary structure analysis of the sequences of the Sac7d and Sac7e proteins was performed with both the Chou−Fasman (Chou & Fasman, 1974, 1978) and the Robson−Garnier algorithms (Robson & Suzuki, 1976; Garnier et al., 1978). Both methods predict the occurrence of significant α-helix (52%) in both proteins extending from approximately Lys9 to Lys28 and from Gly43 to Ala59. Only the Chou−Fasman algorithm predicts a small amount of β-sheet (12%) extending from Lys22 to Lys29 and from Ser31 to Asp36. Reverse turns are predicted near Asp36 and Gly43. These predictions are not consistent with the solution structure of the Sac7d protein which has been determined by 2D NMR (Edmondson, Qiu, and Shriver, manuscript submitted).

Recombinant Gene Expression. The *sac7d* gene (in pBluescript KS+/*sac7d*) was modified by converting the hexanucleotide sequence containing the initiation codon (AATATG) to an *Nde*I site (CATATG) by oligonucleotide G (Table 1) directed mutagenesis to produce pBluescript KS+/*sac7d*(Nd). The *Nde*I−*Bam*HI fragment of pBluescript KS+/*sac7d*(Nd) carrying the coding region of *sac7d* gene was then subcloned into the *Nde*I−*Bam*HI site of pET-3b (Studier et al., 1990) to give pET-3b/*sac7d*, and transformed into HMS174 (DE3), HMS174 (DE3) pLysS, BL21 (DE3), and BL21 (DE3) pLysS (Studier et al., 1990). The plasmid could be established in all of these strains except BL21 (DE3). Furthermore, in transformed BL21 (DE3) pLysS, the growth of the organism is impaired and cultures lyse within 60−70 min after induction with IPTG. On the other hand, the growth of HMS174 strains were not significantly effected by the presence of the plasmid, and lysis was not observed in cultures after 3 h postinduction. The absence of impaired growth in the presence of the plasmid in these

```
Sac 7d                                                  GAATTCTTAT
Sac 7e                                                  ..........

1       GTTCTATAGCGTAATTATGAACAGTTGTATAACTCCTTTAGAGAATAAAT
        CTTAGACGACAAACCTGTAAATAGTATAGTAAATAATGCTATAAATGAAT

51      TATATTTCAATATTACTAATTATTGTACTGGATTCCCCATAAAATTGTAT
        ATGGTGGTACTCCTCAGATAAATTTCACAAAAGTTAGGGCTATTTTGAAA

101     ACATTATATAGGAAAAATAATTTGAGGTAGTCTCATAAGTATGACTTAAC
        TAAATTGTAATGTGATACTAATGATATTTGGATATTAATGTAATACTGGT
                  (A-box)                       *
151     TTAATACCGTAAGGTTTATTTATGACAATATCGTAAGATAACTTGACCTA
        ATATTAATGATAATATTAATTAATGGCGAATTTAAGATATACATGAACAA
                                              *
           M  V  K  V  K  F  K  Y  K  G  E  E  K  E  V  D
201     ATATGGTGAAGGTAAAGTTCAAGTATAAGGGTGAAGAGAAAGAAGTAGAC
        ATATGGCAAAAGTCAGGTTTAAGTATAAGGGTGAAGAGAAAGAAGTAGAC
           M  A  K  V  R  F  K  Y  K  G  E  E  K  E  V  D

         T  S  K  I  K  K  V  W  R  V  G  K  M  V  S  F  T
251     ACTTCAAAGATAAAGAAGGTTTGGAGAGTAGGCAAAATGGTGTCCTTTAC
        ACTTCAAAGATAAAGAAGGTCTGGAGAGTTGGCAAAATGGTGTCCTTTAC
         T  S  K  I  K  K  V  W  R  V  G  K  M  V  S  F  T

           Y  D  D  N  G  K  T  G  R  G  A  V  S  E  K  D
301     CTATGACGACAATGGTAAGACAGGTAGAGGAGCTGTAAGCGAGAAAGATG
        CTATGACGACAATGGTAAGACAGGTAGAGGAGCTGTAAGCGAAAAAGACG
           Y  D  D  N  G  K  T  G  R  G  A  V  S  E  K  D

        A  P  K  E  L  L  D  M  L  A  R  A  E  R  E  K  K
351     CTCCAAAAGAATTATTAGACATGTTAGCAAGAGCAGAAAGAGAGAAGAAA
        CTCCAAAAGAACTAATGGACATGTTAGCAAGAGCAGAAAAGAGAAGAGTAA
        A  P  K  E  L  M  D  M  L  A  R  A  E  K  K  K stop

        stop
401     TAAAATAATTTGTTAAGAAAATCTTCATATAAATTCTTTTTATTTTCTTG
        GGGGAAGGTATAAAGGGCTTTTTAAATGTCAAAAGTTTTTATTTTTTGTT

451     TTTTAATTTATTAGAATTC.............................
        GCATTTCAACTTTAGAAGATCTTTTATAATAGCCTAAATTTCTTTCATGT

501     ..................................................
        GGAGTTTTTCCGCTATTCTTAGGCTTCGATAATAATAATGGTAATGTTGA

551     ......
        AGTATT
```

FIGURE 3: Nucleotide sequences of the *sac7d* and *sac7e* genes. The top and bottom sequences are the nucleotide sequence for the *sac7d* and *sac7e* genes, respectively (aligned using the coding region of each gene). Numbering starts with the *sac7e* sequence. The amino acid sequence coded for by each gene is shown above (*sac7d*) or below (*sac7e*) each nucleotide sequence. Putative promoter (A- and B-boxes) and termination elements are underlined in the 5′ and 3′ noncoding regions of each sequence. Amino acid and nucleotide differences in the coding region of each gene are also indicated by underlines. The G at the start of transcription (in the B-box) for each gene is indicated with an asterisk.

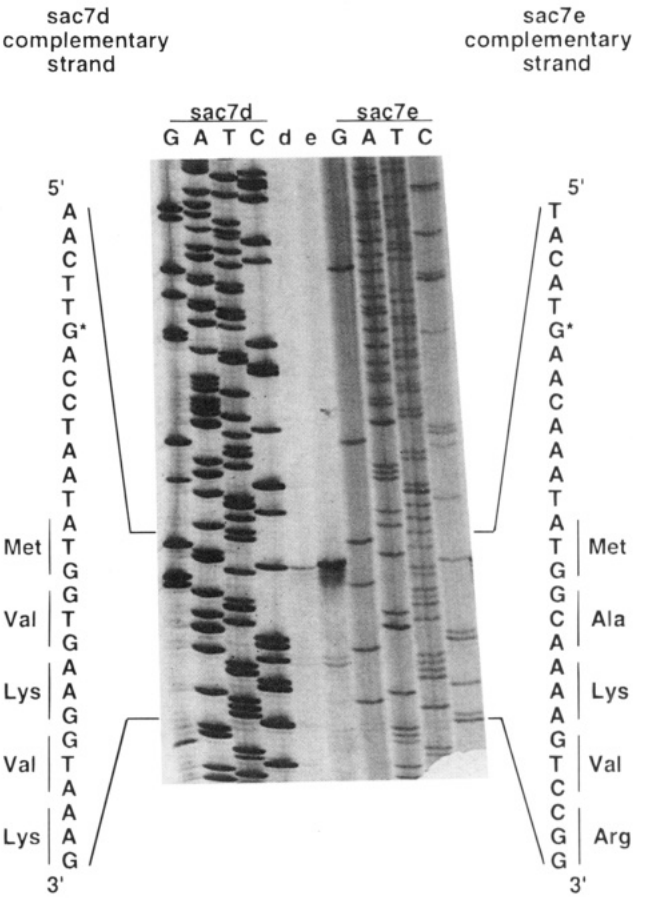

FIGURE 4: Determination of the *in vivo* start of transcription for the *sac7d* and *sac7e* genes by primer extension analysis. *sac7d* (lane d) and *sac7e* (lane e) specific oligonucleotides D and E, respectively [which are complementary to residues 398−418 (Figure 3)], were used to prime the synthesis of a complementary strand of DNA from total *S. acidocaldarius* RNA. These same oligonucleotides were also primers in the dideoxy sequencing reactions used as markers for the *sac7d* (pBSKS+/*sac7d*) and *sac7e* genes (pUC19/*sac7e*) indicated. The sequences written on the left and right are complementary to the ones observed in the autoradiogram in the marked region. The start of transcription is indicated in each sequence by an asterisk. The first five coded amino acids of each protein are also indicated along side each complementary strand sequence.

strains was correlated with a lack of Sac7d protein accumulation. In contrast to HMS174 strains, BL21 and its derivatives lack the *omp*T outer membrane protease and are deficient in the *lon*A protease (Studier et al., 1990). The *omp*T protease has been shown to be responsible for T7 RNA polymerase degradation during protein purification from *E. coli* (Grodberg & Dunn, 1988). Thus, it appears that in the absence of stringent regulation of T7 RNA polymerase synthesis prior to induction with IPTG, or proteolytic degradation of the Sac7d protein, the protein accumulates to lethal levels. However, because significant amounts of the Sac7d protein do not accumulate in HMS174 strains, we have utilized BL21 (DE3) pLysS for subsequent expression and purification of the protein.

Spectroscopic and Chemical Characterization. The UV spectra of native and recombinant Sac7 proteins were essentially identical, as expected, given the presence of a single tryptophan and two tyrosines and two phenylalanines in all proteins. The calculated extinction coefficient based on amino acid composition is 1.05 mL/(mg·cm) at 280 nm, in good agreement with the value of 1.03 mL/(mg·cm) determined by ninhydrin analysis. The extinction coefficients were also determined by using the ratio of absorbance at 280 and 205 nm (see Materials and Methods). The empirical nature of this method might lead to some question of its accuracy, but the high correlation of the results from the six standards is extraordinary ($r = 0.999$), and the reproducibility of the A_{280}/A_{205} ratio measurement is high leading to an expected error of 0.6%. The ratio method demonstrates that the extinction coefficients of the native and recombinant protein are identical, viz., the mean of all extinction coefficient measurements (native and recombinant combined) using this method was 1.18 mL/(mg·cm) with a standard deviation of 0.008 mL/(mg·cm). The final extinction coefficient for both the recombinant and native proteins is taken to be 1.09 mL/(mg·cm), the mean of the three independent measurements, with a standard error of ±0.04 (calculated by propagating the errors of the three measurements). The extinction coefficient was shown to be pH independent from 2 to 10.

The fluorescence excitation and emission spectra of the native Sac7 and recombinant Sac7d proteins were also essentially identical (data not shown). In addition, the fluorescence emission spectrum was essentially that expected

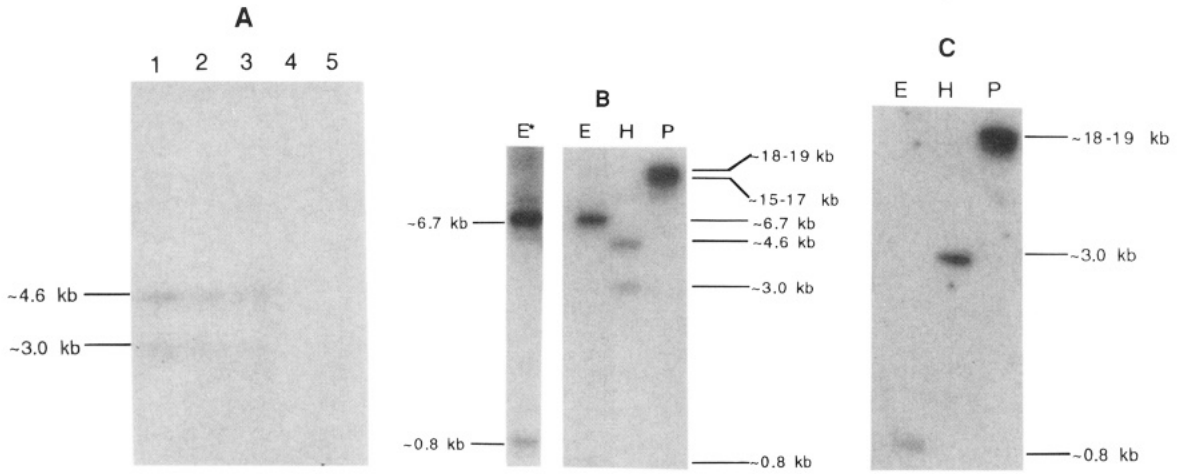

FIGURE 5: Southern analysis of *Sulfolobus* genomic DNA. (A) Autoradiogram of a Southern blot of *Hin*dIII digests of genomic DNA from *S. acidocaldarius* (RGJM) (lane 1), *S. acidocaldarius* (DG6) (lane 2), *S. acidocaldarius* (DSM639) (lane 3), *S. solfataricus* (DSM5354) (lane 4), and *S. solfataricus* (P2) (lane 5) probed with oligonucleotide F. The approximate sizes of the restriction fragments hybridizing to oligonucleotide F are indicated. (B) Autoradiogram of a Southern blot of *Eco*RI (lane E), *Hin*dIII (lane H), and *Pst*I (lane P) digested *S. acidocaldarius* RGJM genomic DNA hybridized with oligonucleotide F. Two closely spaced bands in lane P are clearly evident in the original autoradiogram. Lane E* is a second independent *Eco*RI experiment to clearly demonstrate the 0.8 kb fragment. (C) Similar to panel B except that the DNA was probed with oligonucleotide H.

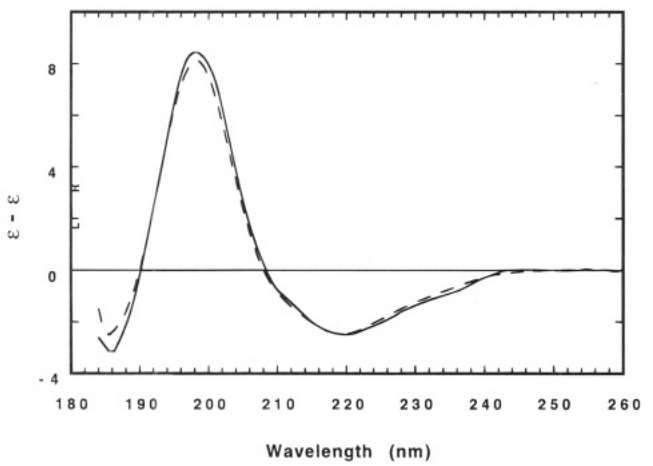

FIGURE 6: Circular dichroism spectra of native Sac7 (solid line, 0.26 mg/mL) and recombinant Sac7d (dashed line, 0.66 mg/mL) proteins in 0.01 M KH_2PO_4, pH 7.0.

for a free tryptophan, indicating that the single tryptophan is highly solvent exposed in both proteins. Notably, the fluorescence emission spectra show a small shift upon DNA binding (data not shown), indicating that the exposure of the tryptophan changes slightly upon DNA binding. The CD spectra of native Sac7 and recombinant Sac7d proteins were also essentially identical (Figure 6). The variable selection method of Johnson (Manavalan & Johnson, 1987) indicates that both the native and recombinant Sac7 proteins are composed of 31% helix (both α- and 3_{10}-helix), 22−25% β-sheet, 0−2% turn, and 42−45% nonrepetitive structure.

The DQF-COSY spectra of the native and recombinant Sac7 proteins are remarkably similar (Figure 7). The native spectrum shows some additional correlation peaks, most likely due to the presence of 7a, b, c, d, and e isoforms in the native preparation and posttranslational modifications (e.g., monomethylation of lysines) in *Sulfolobus*. The essential identity of the chemical shifts for the native and recombinant proteins indicates again that the recombinant and native proteins are folded similarly. The extensive number of alpha protons shifted downfield of the water line at 4.7 ppm indicates the presence of significant β-sheet structure (Wishart et al., 1992). The wide chemical shift dispersion has permitted an essentially complete assignment of the proton resonances and determination of the solution structure (Edmondson, Qiu, and Shriver, manuscript submitted).

No phosphorylation or glycosylation of either the native or recombinant proteins could be detected. The recombinant protein differs from the native by containing the initiator methionine. The recombinant protein also contains an additional C-terminal lysine which was not reported in the amino acid sequence (Kimura et al., 1984), although it remains to be determined if this is an error in the protein sequence or if the lysine is actually removed posttranslationally.

DNA Binding. The binding of Sac7 proteins to DNA is associated with a significant quenching of the intrinsic fluorescence of the single tryptophan (Trp23) in both the native and recombinant Sac7 proteins (Figure 8). Binding of poly[dGdC]•poly[dGdC] in 0.01 M KH_2PO_4 at pH 7.0 leads to a maximal fluorescence quenching of the native protein by 88% and the recombinant Sac7d protein by 87%. Poly[dAdT]•poly[dAdT] shows a maximal quenching of 84% for both proteins (data not shown). The binding data can be fit using the McGhee and von Hippel model (McGhee and von Hippel, 1974) without cooperative interactions assuming a linear relationship between fractional quenching and protein binding. The poly[dGdC]•poly[dGdC] data can be fit with an intrinsic association constant of 2×10^7 M^{-1} for both native and recombinant Sac7d protein and site sizes of 7 bases (3.5 base pairs) and 6.8 bases for native and recombinant protein, respectively. Poly[dAdT]•poly[dAdT] appears to bind slightly weaker with an association constant of 1×10^7 M^{-1} for both proteins and site sizes of 7.5 bases for native protein and 6.8 bases for recombinant protein.

The binding of Sac7 to poly[dAdT]•poly[dAdT] significantly stabilizes the DNA double helix against thermal denaturation. The UV melting curve of poly[dAdT]•poly[dAdT] in 0.01 M KH_2PO_4 is very sharp and has a T_m of 43.5 °C (Figure 9). In the presence of native Sac7d protein,

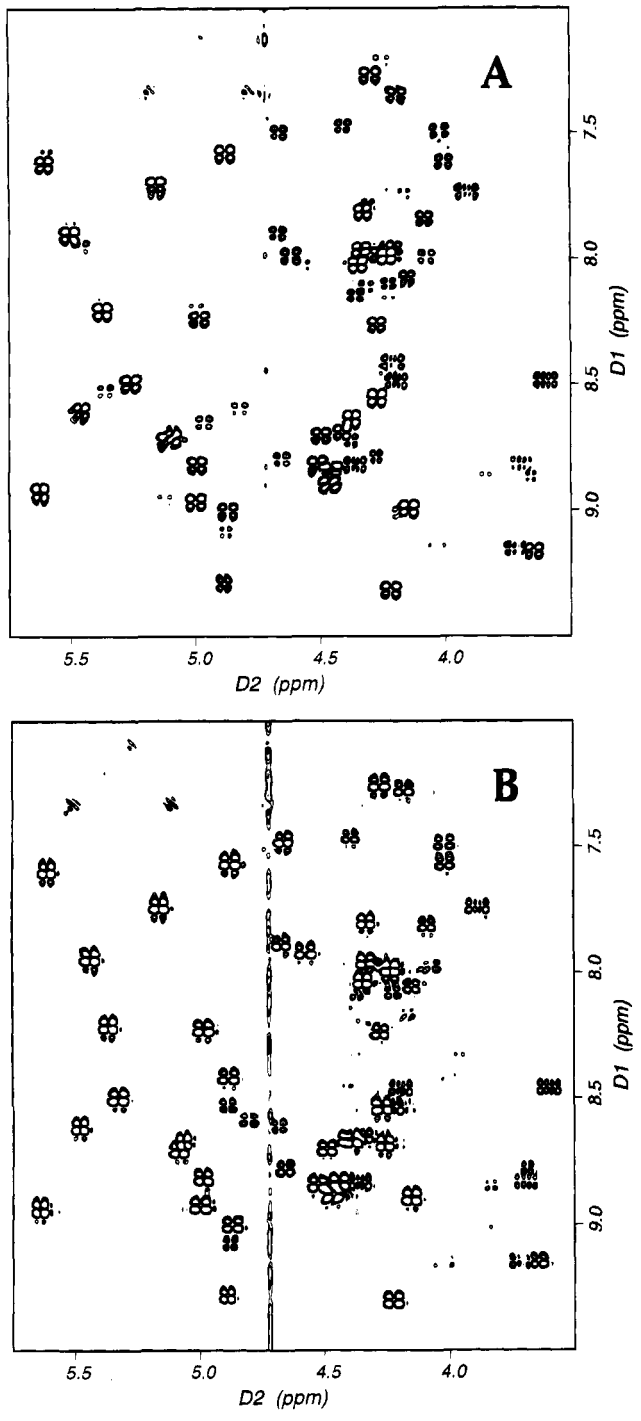

FIGURE 7: Double-quantum filtered (DQF-COSY) α to amide proton correlation spectra of the native Sac7 (A) and recombinant Sac7d (B) proteins at 35 °C in 90% H_2O/10% D_2O, pH 4.1. The protein concentrations in both spectra were approximately 10 mM.

the melting profile of poly[dAdT]·poly[dAdT] broadens and the T_m increases. At the highest protein concentration used in this series of experiments, the DNA melting temperature was increased about 33 °C above that of poly[dAdT]·poly-[dAdT] alone. The recombinant protein increases the T_m of poly[dAdT]·poly[dAdT] by a similar amount. However, the recombinant protein differs in that it aggregates as the double-stranded poly[d(AT)] melts. CD measurements of the suspension, and the supernatant after allowing the aggregate to settle, indicate no major conformational changes during aggregation of the protein–DNA mixture.

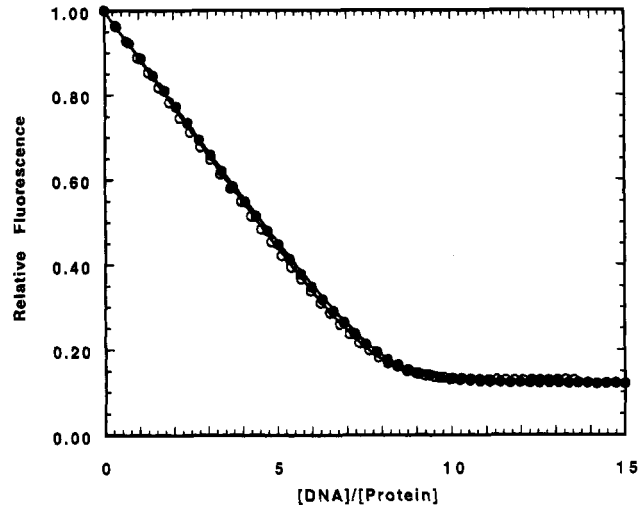

FIGURE 8: Reverse titrations of the native Sac7 (solid circles) and recombinant Sac7d (open circles) proteins with poly[dGdC]·poly-[dGdC] at pH 7.0 (0.01 M KH_2PO_4), 25 °C with 6.6 μM Sac7 proteins and 7.3 μM Sac7d. The smooth curves through the data are overlays of simulations using a noncooperative McGhee–von Hippel model (McGhee & von Hippel, 1974). For the native Sac7 proteins this corresponds to a site size of 7 bases (3.5 base pairs), maximal quenching of 88%, and an intrinsic association constant of 2×10^7 M^{-1}. For the recombinant Sac7d protein this corresponds to a site size of 6.8 bases (3.4 base pairs), maximal quenching of 87%, and an association constant of 2×10^7 M^{-1}.

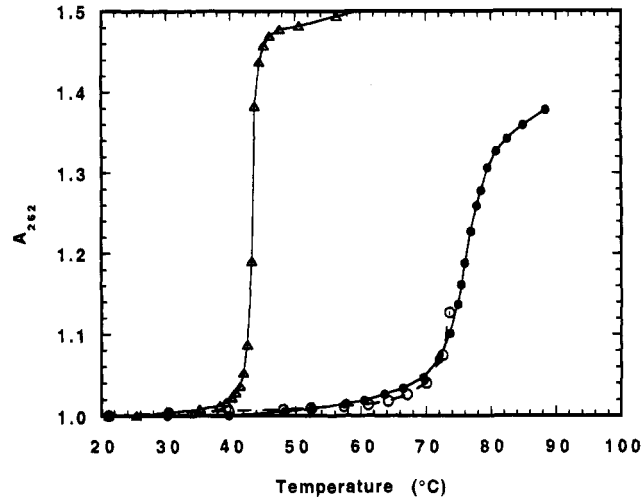

FIGURE 9: Thermal denaturation of poly[dAdT]·poly[dAdT] monitored by changes in UV absorbance at 262 nm in 0.01 M KH_2PO_4, pH 7.0. The melting of poly[dAdT]·poly[dAdT] is shown alone (open triangles), with native Sac7 proteins (solid circles), and with recombinant Sac7d (open circles). The concentration of poly-[dAdT]·poly[dAdT was 70 μM (nucleotides), and the concentration of protein was 350 μM.

Thermal Stability. Sac7 proteins are highly thermostable, as expected from their origin. Native Sac7 and recombinant Sac7d samples heated to 100 °C showed no precipitation or cloudiness, although some increase in scattering was noticeable in the UV spectrum. The proteins unfold reversibly as indicated by the observation of similar endotherms with repetitive DSC scans up to 100 °C.

The native Sac7 proteins show a DSC endotherm at pH 6.0 (0.01 M KH_2PO_4, 0.1 M KCl, 0.001 M EDTA) with a T_m of 99.0–100.2 °C (data not shown). By comparison, the native Sso7 protein has a T_m of 99.4 °C under similar conditions (data not shown). A precise midpoint for the unfolding transition is difficult to define since data above

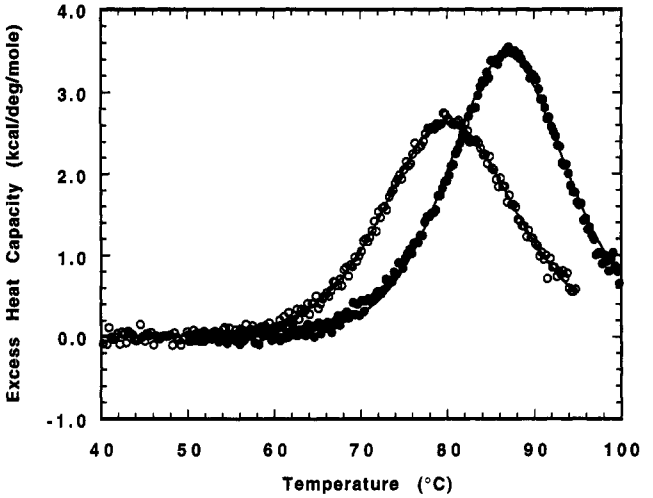

FIGURE 10: Differential scanning calorimetry (DSC) of native Sac7 (solid circles) and recombinant Sac7d (open circles) proteins at pH 4.0 (0.3 M KCl, 0.05 M potassium acetate). Protein concentrations were 1.5 mg/mL of native Sac7 proteins and 1.38 mg/mL of recombinant Sac7d. Smooth curves through the data are nonlinear least-squares fits with T_m = 80.3 °C, ΔH_{cal} = 53.0 kcal/mol, ΔH_{vh} = 49.6 kcal/mol, for the recombinant protein; and T_m = 86.8 °C, ΔH_{cal} = 56.4 kcal/mol, ΔH_{vh} = 60.3 kcal/mol for the native protein.

100 °C cannot be collected in water in the MC2 calorimeter. Notably, the unfolding of the native Sac7 proteins is remarkably reversible, as indicated by essentially 100% reproducibility of successive scans on the same sample following cooling. The recombinant Sac7d protein unfolds at pH 6.0 (0.01 M KH_2PO_4, 0.1 M KCl, 0.001 M EDTA) with a T_m of 92.7 °C, or approximately 7 °C less than the native.

A reliable analysis of the DSC endotherms requires a more complete delineation of the endotherm which can be obtained by lowering the pH and increasing the salt concentration to shift the endotherms to lower temperature. At pH 4.0 (0.05 M potassium acetate, 0.3 M KCl) the native protein unfolds with a T_m of 86.8 °C (Figure 10). The endotherm can be fit with a van't Hoff enthalpy of 60.3 kcal/mol and a calorimetric enthalpy of 56.4 kcal/mol, i.e., a $\Delta H_{cal}/\Delta H_{vh}$ of 0.94, indicating that the native protein exists as a monomer under these conditions and unfolds in an all-or-none fashion with no significant, populated intermediates.

The recombinant Sac7d protein similarly unfolds reversibly at pH 4.0 (0.05 M potassium acetate, 0.3 M KCl) but with a midpoint temperature of 80.3 °C (Figure 10), or 6.5 °C less than the native protein. It unfolds with a van't Hoff enthalpy of 49.6 kcal/mol, and a calorimetric enthalpy of 53.0 kcal/mol, i.e., a $\Delta H_{cal}/\Delta H_{vh}$ of 1.07. The identity, within experimental error, of the calorimetric and van't Hoff enthalpies indicates that the recombinant protein also exists as a monomer under these conditions and unfolds via a two-state reaction.

DISCUSSION

We report here the cloning and sequencing of two genes from *S. acidocaldarius* coding for Sac7 proteins which correspond to Sac7d and Sac7e. The *sac7d* and *sac7e* genes differ at only 16 positions within the coding region (underlined in Figure 3); three of these differences are transversions, while the rest are transitions. The *sac7d* and *sac7e* genes code for 66 and 65 amino acid proteins, respectively. The deduced amino acid sequences are in complete agreement with the published sequences for both proteins (Kimura et al., 1984; Choli et al., 1988a) with the exception of initiator methionines at the amino termini and an additional lysine (Lys66) at the carboxy terminus of the Sac7d protein in the deduced sequence. The additional lysine can be explained either by a failure to discern the final lysine in the amino acid sequencing of the Sac7d or by posttranslational carboxy-terminal processing to produce the mature protein. It should be noted that Sac7d, Sac7e, and Sso7d all terminate with at least two lysine residues (Figure 1).

The data presented here indicate that there are only two Sac7 protein genes in *S. acidocaldarius*. Genes coding for Sac7 proteins other than Sac7d and e could not be detected. The failure to detect genes for the Sac7a and b proteins and the fact that the proteins appear to be simply truncated at the carboxy termini to various extents suggest that Sac7a and b result from either posttranslational modification at the carboxy terminus or by proteolysis during protein isolation and purification.

Promoter elements consistent with the archaeal "A-box" and "B-box" consensus sequences have been located upstream of the *sac7d* and *sac7e* protein coding sequences. The agreement of the "A-box" sequence of *sac7d* with the consensus "A-box" sequence is greater than that for the *sac7e*. This difference between the "A-box" of the promoter elements in the two genes may explain the higher levels of Sac7d relative to Sac7e *in vivo* (Grote et al., 1986).

There is significant sequence similarity in the regions of *sac7d* and *sac7e* extending from the 5′ end of the "A box" to the initiation codon when the corresponding "A-" and "B-" boxes are aligned. The two sequences also have similarly placed pyrimidine rich regions downstream of their termination codons. These regions show similarity to the transcription termination signals described for the *Sulfolobus* virus-like particle, SSV1, where transcription termination has been shown to occur within pyrimidine-rich regions directly 3′ of the consensus TTTTTYT [reviewed in Brown et al. (1989)]. Northern analysis of *S. acidocaldarius* RGJM RNA probed with an oligonucleotide (oligonucleotide F, Table 1) complementary to the common sequence at residues 305–324 of the two *sac7* genes (Figure 3) showed hybridization to a single size of transcripts (Shao and Gupta, unpublished results), indicating that both transcripts terminate in similarly placed regions. Thus, it is likely that the conserved oligopyrimidine sequences of the two genes contain the transcription termination signals.

Although the regions associated with transcription termination are highly homologous, the sequences between these regions and the termination codons are significantly different in the *sac7d* and *sac7e* genes. Similarly, though the regions encompassing the putative core promoter elements in the two genes ("A-" and "B-" boxes) share extensive homology, the sequences 5′ of the "A-box" show less similarity. It would appear that sufficient time has elapsed since the supposed original gene duplication for the two sequences to diverge. The conservation of cis-regulatory elements along with coding regions in the two genes indicates that there is a selective pressure to maintain not only the expression of both gene products but also a large part of their sequence. It is not clear if there is more than one form of the Sso7 proteins.

A typical ribosome binding site sequence upstream of initiator ATG is not observed in either of the two *sac7* genes

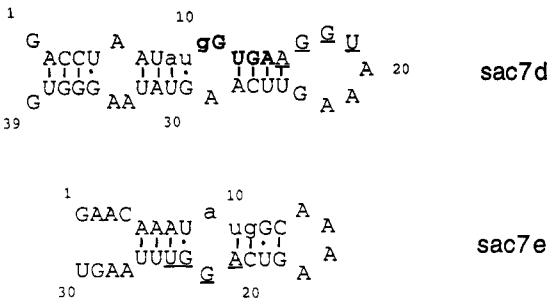

FIGURE 11: Potential secondary structures for the 5′-terminal regions of the *sac7* RNA transcripts determined using Mulfold (Jaeger et al., 1989a,b; Zuker, 1989). Initiator codons are shown in lower case. Putative ribosome binding sequences GGUGA and AGGU are indicated in bold and underlined formats, respectively. Note that the AGGU sequences within the two transcripts are located at different positions.

(Figure 3). This is not unusual, since many other *Sulfolobus* genes also lack these sites (Amils et al., 1993; Dalgaard & Garrett, 1993). However, potential ribosome binding sites are observed downstream of the initiator codons of the two *sac7* genes which have precedents in other archaea. The ribosome binding sites in certain halobacterial genes, which have very short or no 5′ untranslated regions, occur within loops of potential hairpin structures in the 5′ regions of the transcripts (Brown et al., 1989; Amils et al., 1993). The hairpin arrangement probably exposes these sites for interaction with 16S rRNA. We note that the 5′ regions of the two *sac7* transcripts can be folded into secondary structures as shown in Figure 11. The sequence UCACCU near the 3′ end of 16S rRNA of *Sulfolobus* (Woese et al., 1984; Olsen et al., 1985) potentially can either form five base pairs with GGUGA within codons 1−3 or form four base pairs with AGGU within codons 3−4 of the *sac7d* transcript. Corresponding sequences in the sac7e transcript are GGCAA and AAGU, respectively, which cannot form similar pairs with the 16S rRNA. However, further downstream in the *sac7e* transcript, there is AGGU within codons 5−6, which can form four base pairs with the same UCACCU sequence of the 16S rRNA; the corresponding site in sac7d is less efficient AAGU. Parts of these potential ribosome binding sites do occur within single-stranded regions (Figure 11), as are the cases for the above mentioned halobacterial genes. The differences between the sequences and locations of the potential ribosome binding sites of the two *sac7* transcripts, along with the previously mentioned differences in the "A-box" sequences, may also explain the higher synthesis of Sac7d protein.

Kimura et al. (1984) have previously noted that the clustering of lysines in the amino terminus of these proteins is reminiscent of that observed in eukaryotic HMG proteins. Choli et al. (1988b) have also pointed out a slight sequence similarity with E2A DNA-binding protein from adenovirus. An extensive search of the currently available sequence databases showed no significant homologies between the Sac7d protein and any known chromatin or DNA-binding protein. A BLAST search using the Sac7d sequence picked up a 100% homology with the amino-terminal sequence (only 12 amino-terminal residues are known) of a small protein (accession number S21168) from *S. solfataricus* which apparently catalyzes disulfide bond formation (Guagliardi et al., 1992). This report should be viewed with caution due to the loss of activity upon cation exchange chromatography of the protein. BLAST also picked up a high homology to a reported p2 ribonuclease (Fusi et al., 1993) from *S. solfataricus* with a sequence identical to the Sso7d protein (Choli et al., 1988a). RNase activity for the 7 kDa proteins is surprising and remains to be confirmed. Preliminary experiments indicate that the recombinant Sac7d protein does not have RNase activity (Edmondson and Shriver, unpublished results). The BLAST search also picked up some weak homology with the 30S ribosomal protein S5 from *E. coli* (P02356) and heat shock protein X16 from the African clawed frog (A22175). A FASTA search using the Sac7d sequence revealed some homology with elongation factor 1-δ (P29692), 30S ribosomal protein S8 (P24353), and DNA-directed RNA polymerase subunit A′ (P31813). A PROSITE search using the Sac7d sequence revealed phosphocreatine kinase phosphorylation sites at residues 17−19 (TSK), 40−42 (TGR), and 46−48 (SEK), and creatine kinase II phosphorylation sites at 33−36 (TYDD), and 46−49 (SEKD). A BLOCKS analysis provided a single meaningful match with ribosomal S5 protein.

We have expressed the *sac7d* gene in the tightly controlled BL21(DE3)pLysS *E. coli* expression system developed by Studier et al. (1990) using the pET series of plasmids. Accumulation of the *sac7d* gene product appears to be lethal in *E. coli*. This is indicated perhaps most clearly by our inability to establish the pET-3b/*sac7d* construct in BL21(DE3). The additional regulation provided by the T7 lysozyme inhibition of T7 polymerase appears to be required. The purified, recombinant protein can be isolated with reasonable yield, e.g., typically, about 1 mg of protein per g of wet weight *E. coli* cells is obtained, or approximately twice that obtained for the native protein from *S. acidocaldarius*. We have been unsuccessful in expressing the *sac7e* gene, possibly due to its usage of codons rare in *E. coli*.

The recombinant Sac7d protein appears to be essentially identical to the native Sac7 proteins in all respects except for stability. The UV spectral extinction coefficients are identical, as are the fluorescence excitation and emission spectra. This is perhaps not surprising given that both are largely due to a single tryptophan on the surface of the protein (Edmondson, Qiu, and Shriver, manuscript submitted) [see also Baumann et al. (1994) for the structure of Sso7d], although the two tyrosines should be sensitive to differences in structure. CD spectra are more sensitive to differences in secondary structure content, and the spectra of the two proteins are essentially identical, again indicating similar structures for native and recombinant protein.

Analyses of the CD spectra using the variable selection method of Johnson (Manavalan & Johnson, 1987) indicate that Sac7d consists of 31% helix and 22−25% β-sheet. This differs from the 52% α-helix, 12% β-sheet predicted by sequence analysis algorithms in this work and the 15% α-helix, 15% β-sheet predicted by Choli et al. (1988a) using the average of four different prediction methods. All of these methods significantly underestimate the amount of β-sheet in Sac7d (42%) as determined from the NMR solution structure (Edmondson, Qiu, and Shriver, manuscript submitted) [see also Baumann et al. (1994)]. However, the helical content determined by CD (31%) is close to that of the NMR solution structure (22% α-helix, 11% 3_{10}-helix). An analysis of the CD spectrum of Sac7e (Dijk & Reinhardt, 1986) using the PG method (Provencher & Glockner, 1981) gave a much better estimate of β-sheet content (44%) but underestimated

the helical content (15%). The CD spectrum reported for Sac7e (Dijk & Reinhardt, 1986) differs quantitatively from that of native Sac7 and recombinant Sac7d presented here. Further, the inability of the CD analyses to accurately estimate the secondary structure content suggests that at least part of the secondary structure contributions to the CD spectra of the Sac7 proteins are not well represented in these sets of reference proteins.

A more detailed, atomic level comparison of the structures of the recombinant and native proteins can be obtained from NMR. The "fingerprint" region of double-quantum filtered COSY spectra of proteins shows the chemical shift correlations of alpha and NH protons and is exquisitely sensitive to the structure of the protein [see, for example, Wishart et al. (1992)]. This permits a qualitative comparison of the structure of the backbone of the two proteins which is more detailed than that provided by optical spectra comparisons. The fingerprint regions of native and recombinant Sac7d protein are remarkably similar, indicating that the two proteins have very similar backbone folding patterns.

The binding of the Sac7 proteins to double stranded DNA leads to a dramatic decrease in intrinsic tryptophan fluorescence. The large signal allows for essentially noise-free titrations and accurate comparisons of the native and recombinant protein binding function. The data presented here indicate an affinity of 2×10^7 M^{-1} and site size of 3.5 base pairs for poly[dGdC]·poly[dGdC]. The agreement of quantitative binding parameters obtained for the native and recombinant proteins is additional evidence for essentially identical global folds for the two proteins. These binding studies are the first quantitative analysis of the binding of the Sac7 proteins to DNA.

Various prior studies of the 7 kDa DNA-binding proteins from *Sulfolobus* have characterized the binding to nucleic acids in a qualitative manner. Electron micrographs of the 7 kDa proteins from *S. acidocaldarius* complexed with DNA indicated that the helix becomes increasingly compacted with increasing ratios of protein to DNA (Dijk & Reinhardt, 1986; Lurz et al., 1986). Filter binding studies confirmed that the 7 kDa proteins had an affinity for pBR322 DNA even at relatively high salt concentrations (e.g., 0.265 M NaCl) which was comparable to that observed for *E. coli* HU protein (Grote et al., 1986; Choli et al., 1988a). Characterization of the affinity for DNA in this work was in terms of percent bound at a specific ratio of protein to DNA. DNA-melting studies have also been performed on a small DNA-binding protein from *S. acidocaldarius*, HSNP-C′, with an amino acid composition similar to the Sac7e protein, although the sequence has not been presented The protein increases the T_m of double-stranded DNA (Reddy & Suryanarayana, 1989). In addition, this protein demonstrated a significant quenching of its intrinsic tryptophan fluorescence upon DNA binding, although no quantitative analysis of the titrations was performed.

Baumann et al. (1994) have recently presented some fluorescence binding data for the homologous Sso7 proteins from *S. solfataricus*. A quantitative analysis of the titrations was not performed, but a visual inspection of the data indicates a binding site size for double-stranded DNA of six base pairs in low salt (0.02 M Tris, pH 7.4), nearly twice that presented here. Assuming a site size of 3−6 base pairs, the binding affinity in low salt is approximately 0.5 to 1 × 10^6 M^{-1}. The thermal stability of poly[dIdC]·poly[dIdC] was increased by approximately 40 °C in 5 mM Tris (pH 7.0).

The unfolding of both the native and recombinant proteins is reversible, allowing for detailed, accurate characterization of the thermodynamics of folding. In contrast to all other physical parameters studied here, the energetics of folding of the recombinant Sac7d protein differs significantly from that of the native Sac7 proteins. The native protein unfolds at pH 6.0 at 100 °C, remarkable given the absence of any metal cofactors or disulfides. Surprisingly, the recombinant protein unfolds with a T_m 6.5 °C less than the native. The lower enthalpy of unfolding of the recombinant protein is not surprising and most likely results from a positive heat capacity change associated with unfolding. Any shift to lower temperature of an endotherm associated with a positive ΔC_p will lead to a decrease in enthalpy since

$$\Delta C_p = \left(\frac{\partial \Delta H}{\partial T}\right)_P$$

It is generally thought that a positive ΔC_p of unfolding is due to the exposure of internal hydrophobic residues (Sturtevant, 1977; Privalov & Gill, 1988). The magnitude of the change observed here is consistent with that observed for other globular proteins (Privalov & Gill, 1988).

Maras et al. (1992) have previously noted that specific lysine monomethylation of glutamate dehydrogenase from *S. solfataricus* might be responsible for enhanced thermal stability of this enzyme relative to homologous mesophile forms. Baumann et al. (1994) have presented mass spectroscopic evidence correlating methylation of the Sso7 protein with growth temperature, and they have suggested that such a modification might be related to the stability of the protein. The most straightforward way to determine if methylation increases the thermostability of the protein would be to compare the stabilities of the protein in its methylated and unmethylated forms. Demethylation of the native protein is not a trivial control experiment given the lack of commercially available demethylases and most importantly the specificity of reported demethylases (Paik & Kim, 1980). In the absence of a demethylase, the preparation of an unmethylated form is best accomplished using recombinant protein. We have demonstrated here a significant difference in the thermostability of native and recombinant Sac7 protein. The only known difference between these proteins is the ϵ-aminomonomethylation of lysines 5 and 7 in the native protein and the initiating methionine in the recombinant protein. The lack of Lys66 in the reported amino acid sequence of the native protein is presumably a sequencing error, and this will be investigated in the NMR analysis of the native protein. No other posttranslational modification, such as phosphorylation or glycosylation, of the native or recombinant Sac7 proteins was detectable. The current evidence, therefore, strongly indicates that *Sulfolobus* can increase the thermostability of some of its proteins by specific lysine monomethylation.

We note that the level of specific methylation of Sac7 is variable and incomplete, i.e., the native preparation is heterogeneous (Kimura et al., 1984; Choli et al., 1988a,b). Choli et al., (1988b) report that the degree of monomethylation of lysine 4 is 70%, 25%, and 20% in native Sac7a, Sac7b, and Sac7d, respectively; while that for lysine 6 is 50%, 40%, and 50%, respectively. Heterogeneity would be

expected to lead to broadening of the endotherm, rather than narrowing (see Figure 10). It would appear, therefore, that stabilization might not require complete methylation of the specific lysines.

Interestingly, we have been unable to increase the stability of the recombinant Sac7d protein by nonspecific, reductive methylation (McCrary and Shriver, unpublished results), a process which leads to predominantly dimethylation (Means & Feeney, 1971). Monomethylation changes the pK_a of the ϵ-amino group from 9.25 to 10.63, while dimethylation has little further effect giving a pK_a of 10.78 (Paik & Kim, 1980). Trimethylation returns the pK_a to 9.8. Given the small change in pK_a and the fact the difference is observed even at pH 4.0, it is doubtful that an effect of monomethylation on stability might be electrostatic in origin. A structural explanation of the difference in stability must await a more detailed comparison of the structures of the native and recombinant proteins. The spectroscopic data presented here would indicate that the structural differences are slight.

ACKNOWLEDGMENT

We thank Drs. Paul Hargrave and Hugh McDowel (University of Florida) for amino-terminal sequencing of the 7 kDa proteins of *Sulfolobus acidocaldarius*, F. William Studier (Brookhaven) for providing the expression vectors and *E. coli* strains HMS174 and BL21 and their derivatives, Jack Parker (Southern Illinois University) for supplying *E. coli* CJ236, Dennis Grogan (University of Cinncinati) for providing *S. acidocaldarius* DG6 and *S. solfataricus* P2 and advice for characterization of different *Sulfolobus* strains, and Lingshi Qiu (Southern Illinois University) for purifying the Sac7 and Sso7 proteins from *Sulfolobus* strains. We also thank Dr. Ignatius Gomes for valuable discussions and Neelima Reddy for her assistance in the laboratory. Finally, we thank an anonymous reviewer for pointing out the GGTGA sequence as a potential ribosome binding site.

REFERENCES

Altshul, S., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990) *J. Mol. Biol. 215*, 403−410.

Amils, R., Cammarano, P., & Londei, P. (1993) in *The Biochemistry of Archaea (Archaebacteria)* (Kates, M., Kushner, D. J., & Matheson, A. T., Eds.) pp 393−438, Elsevier, New York.

Bairoch, A. (1992) *Nucleic Acids Res. 20*, 2013−2018.

Baumann, H., Knapp, S., Lundback, T., Ladenstein, R., & Hard, T. (1994) *Struct. Biol. 1*, 808−819.

Berger, L. S., & Kimmel, A. R. (1987) *Guide to Molecular Cloning Techniques*, Academic Press, New York.

Brock, T., Brock, K., Belly, R., & Weiss, R. (1972) *Arch. Microbiol. 84*, 54−68.

Brown, J., Daniels, C., & Reeve, J. (1989) *CRC Crit. Rev. Microbiol. 16*, 287−338.

Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 285−298.

Chen, G. C., & Yang, J. T. (1977) *Anal. Lett. 10*, 1195.

Choli, T., Henning, P., Wittmann-Liebold, B., & Reinhardt, R. (1988a) *Biochim. Biophys. Acta 950*, 193−203.

Choli, T., Wittmann-Liebold, B., & Reinhardt, R. (1988b) *J. Biol. Chem. 263*, 7087−7093.

Chou, P. Y., & Fasman, G. (1974) *Biochemistry 13*, 222−245.

Chou, P. Y., & Fasman, G. (1978) *Annu. Rev. Biochem. 47*, 251−276.

Dalgaard, J. Z., & Garrett, R. A. (1993) in *The Biochemistry of Archaea (Archaebacteria)* (Kates, M., Kushner, D. J., & Matheson, A. T., Eds.) pp 535−563, Elsevier, New York.

Debois, B., Gilles, K., Hamilton, J., Rebers, P., & Smith., F. (1956) *Anal. Chem. 28*, 350.

DeRosa, M., & Gambacorta, A. (1975) *J. Gen. Microbiol. 86*, 156−164.

Dijk, J., & Reinhardt, R. (1986) in *Bacterial Chromatin* (Gualerzi, C. O., & Pon, C. O., Eds.) Springer-Verlag, Berlin.

Edelhoch, H. (1967) *Biochemistry 6*, 1948−1954.

Emory, S. A., & Belasco, J. G. (1990) *J. Bacteriol. 172*, 4472−4481.

Fiske, C., & Subbarow, Y. (1925) *J. Biol. Chem. 66*, 375−400.

Fuchs, R. (1991) *Comput. Appl. Biosci. 7*, 105−106.

Fusi, P., Tedeschi, G., Aliverti, A., Ronchi, S., Tortora, P., & Guerritore, A. (1993) *Eur. J. Biochem. 211*, 305−310.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97−120.

Gill, S., & von Hippel, P. (1989) *Anal. Biochem. 182*, 319−326.

Gordon, A., & Ford, R. (1972) *The Chemist's Companion: A Handbook of Practical Data, Techniques, and References*, John Wiley, New York.

Grodberg, J., & Dunn, J. J. (1988) *J. Bacteriol. 170*, 1245−1253.

Grogan, D. (1989) *J. Bacteriol. 171*, 6710−6719.

Grogan, D. W. (1991) *J. Bacteriol. 173*, 7725−7727.

Grote, M., Dijk, J., & Reinhardt, R. (1986) *Biochim. Biophys. Acta 873*, 405−413.

Guagliardi, A., Cerchia, L., De Rosa, M., Rossi, M., & Bartolucci, S. (1992) *FEBS Lett. 303*, 27−30.

Gupta, R. (1984) *J. Biol. Chem. 259*, 9461−9471.

Henikoff, S., & Henikoff, J. G. (1991) *Nucleic Acids Res. 19*, 6565−6572.

Hirs, C. W. (1967) *Methods Enzymol. 11*, 411−413.

Inman (1962) *J. Mol. Biol. 5*, 172.

Jaeger, J. A., Turner, D. H., & Zuker, M. (1989a) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7706−7710.

Jaeger, J. A., Turner, D. H., & Zuker, M. (1989b) *Methods Enzymol. 183*, 281−306.

Johnson, R., & Walseth, T. F. (1979) *Adv. Cyclic Nucleotide Res. 10*, 135−167.

Johnson, W. C., Jr. (1984) *Food Analysis Principles and Techniques, Volume 2, Physicochemical Techniques*, (Gruenwedel, D. W., & Whitaker, J. R., Eds.) Marcel Dekker, Inc., New York.

Kimura, M., Kumura, J., Phillips, D., Reinhardt, R., & Dijk, J. (1984) *FEBS Lett. 176*, 176−178.

Kunkel, T., Roberta, J. D., & Zakour, R. A. (1987) *Methods Enzymol. 154*, 367−382.

Leloir, L., & Cardini, C. (1957) *Methods Enzymol. 3*, 840−850.

Lurz, R., Grote, M., Dijk, J., Reinhardt, R., & Dobrinski, B. (1986) *EMBO J. 5*, 3715−3721.

Manavalan, P., & Johnson, W. C. (1987) *Anal. Biochem. 167*, 76−85.

Maras, B., Conslavi, V., Chiaraluce, R., Politi, L., De Rosa, M., Bossa, F., Scandurra, R., & Barra, D. (1992) *Eur. J. Biochem. 203*, 81−87.

Mayes, E. L. V. (1984) in *Methods in Molecular Biology, Volume 1: Proteins* (Walker, J. M., Ed.) Humana Press, Clifton, NJ.

McAfee, J. (1993) Ph.D. Dissertation, Genetic and DNA Binding Studies on a 7 kDa Protein from *Sulfolobus acidocaldarius*, Southern Illinois University, Carbondale, IL.

McGhee, J., & von Hippel, P. (1974) *J. Mol. Biol. 86*, 469−489.

Means, G., & Feeney, R. (1971) *Chemical Modification of Proteins*, Holden-Day, Inc., San Francisco, CA.

Moore, S., & Stein, W. H. (1954) *J. Biol. Chem. 211*, 907−913.

Olsen, G. J., Pace, N. R., Nuell, M., Kaine, B. P., Gupta, R., & Woese, C. R. (1985) *J. Mol. Evol. 22*, 301−307.

Paik, W., & Kim, S. (1980) *Protein Methylation*, John Wiley and Sons, New York.

Privalov, P., & Gill, S. (1988) *Adv. Protein Chem. 39*, 191−234.

Provencher, S. W., & Glöckner, J. (1981) *Biochemistry 20*, 33−37.

Rance, M., Bodenhausen, G., Wagner, G., Ernst, R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479−485.

Reddy, T. R., & Suryanarayana, T. (1988) *Biochim. Biophys. Acta 949*, 87−96.

Reddy, T. R., & Suryanarayana, T. (1989) *J. Biol. Chem. 264*, 17298−17308.

Robson, B., & Suzuki, E. (1976) *J. Mol. Biol. 107*, 327−356.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Sandman, K., Krzycki, J. A., Dobrinski, B., Lurz, R., & Reeve, J. N. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5788−5791.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.
Savitsky, A., & Golay, M. J. E. (1964) *Anal. Chem. 36*, 1627.
Schägger, H., & von Jagow, G. (1987) *Anal. Biochem. 166*, 368−379.
Searcy, D. G. (1975) *Biochim. Biophys. Acta 395*, 535−547.
Searcy, D. G., & Delange, R. J. (1980) *Biochim. Biophys. Acta 609*, 197−200.
Short, J., Fernandez, J. M., Sorge, J. A., & Huse, W. D. (1988) *Nucleic Acids Res. 16*, 7583−7600.
Shriver, J. W., & Kamath, U. (1990) *Biochemistry 29*, 2556−2564.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503−517.
Stein, D. B., & Searcy, D. G. (1978) *Science 202*, 219−221.
Studier, W. F., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 66−89.
Sturtevant, J. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2236−2240.
Thomm, M., Stetter, K. O., & Zillig, W. (1982) *Zbl. Bakt. Hyg., I. ABT. Orig. C3*, 128−139.
Tiktopulo, E. I., & Privalov, P. (1974) *Biophys. Chem. 1*, 349−357.
van Iersel, J., Jzn, J. F., & Duine, J. (1985) *Anal. Biochem. 151*, 196−204.
Wells (1970) *J. Mol. Biol. 54*, 465.
Wishart, D., Sykes, B., & Richards, F. (1992) *Biochemistry 31*, 1647−1651.
Woese, C. R., Gupta, R., Hahn, C. M., Zillig, W., & Tu, J. (1984) *System. Appl. Microbiol. 5*, 97−105.
Zillig, W. (1993) *Nucleic Acids Res. 21*, 5273.
Zillig, W., Palm, P., Reiter, W.-D., Gropp, F., Puhler, G., & Klenk, H.-P. (1988) *Eur. J. Biochem. 173*, 473.
Zuker, M. (1989) *Science 244*, 48−52.
Zuiderweg, E. R. P., Hallenga, K., & Olejniczak, E. T. (1986) *J. Magn. Reson. 70*, 336−343.

BI950704Z