

Structure, Stability, and Flexibility of Ribosomal Protein L14e from *Sulfolobus solfataricus*^{†,‡}

Stephen P. Edmondson, Jacquelyn Turri, Kelley Smith, Andrew Clark, and John W. Shriver*

Alabama High Field NMR Laboratory, Departments of Chemistry and Biological Sciences, University of Alabama in Huntsville, Huntsville, Alabama 35899

Received February 24, 2009; Revised Manuscript Received May 7, 2009

ABSTRACT: Ribosomal protein L14e is a component of the large ribosomal subunit in both archaea and eukaryotes. We report here a high-resolution NMR solution structure of recombinant L14e and show that the N-terminal 57 residues adopt a classic SH3 fold. The protein contains a tight turn between strands 1 and 2 instead of the typical SH3 RT-loop, indicating that it is unlikely to interact with neighboring ribosomal proteins using the common SH3 site for proline-rich sequences. The remainder of the protein (39 residues) forms a largely extended chain with a short helix which packs onto the surface of the SH3 domain via hydrophobic interactions. It has the potential of adopting an alternative structure to expose a hydrophobic surface for protein–protein interactions in the ribosome without disruption of the SH3 fold. ¹⁵N relaxation data demonstrate that the majority of the C-terminal chain is well-defined on the SH3 surface. The globular protein unfolds reversibly with a *T*_m of 102.8 °C at pH 7, making it one of the most stable SH3 domain proteins described to date. The structure of L14e is expected to serve as a model for other members of the L14e family, along with members of the COG2163 group, including L6e and L27e. Interestingly, the N-terminal sequence of L14e shows the greatest similarity of any *Sulfolobus* protein to the reported N-terminal sequence of Sac8b, a DNA-binding protein reported by Grote et al. ((1986) *Biochim. Biophys. Acta* 873, 405–413). The likelihood that L14e and Sac8b are the same protein is discussed.

The ribosome is the universal molecular machine which is responsible for translating mRNA into polypeptides and proteins. It is composed of about 80 proteins assembled onto rRNA to form a 70S particle in prokaryotes (eubacteria and archaea) and a larger 80S particle in eukaryotes. The intact ribosome can be separated into two subunits referred to as small and large: 30S and 50S in prokaryotes, and 40S and 60S in eukaryotes. Although the intact archaeal ribosome subunits are smaller than those in eukaryotes, the archaeal proteins are more similar to those in eukaryotes than eubacteria (1). The structure of the bacterial ribosome has been defined by crystal structures of the 30S and 50S subunits (2–4) and the intact 70S particle (5–8). Crystal structures of the archaeal 50S subunit from *Haloarcula* (9) and the intact yeast ribosome (10) have been published, along with cryomicroscopy reconstructions of the canine ribosome (11, 12).

L14e is known to be a protein component of the large ribosome subunit in eukaryotes and Archaea based on 2D gel electrophoresis and sequencing, but the structure and position in the ribosome have not been described. The L14e sequence is conserved in archaea and eukaryotes (including yeast, *Drosophila*,

rat, and human), where the protein is sometimes referred to as L14 (1, 13–16). L14e shows no homology to the bacterial L14p protein (unfortunately also sometimes referred to as L14) for which a structure has been determined (17). The archaeal protein is approximately 30% identical to the eukaryotic 60S ribosomal protein L14e (e.g., *Homo sapiens* NP_003964). Eukaryotic L14e is often significantly larger (e.g., 24 kDa for rat L14) with additional domains at the C-terminus (18) which have been identified as nuclear targeting sequences, along with a bZIP domain (19). Autoantibodies to L14e in humans have been associated with systemic lupus erythematosus (20). Although L14e is well conserved, it is not found in all archaea, and therefore it is presumably not essential, although homologous proteins may exist which can function in its place. Notably it does not occur in the halophilic archaeon *Haloarcula*, the source for the only structure of an intact archaeal ribosome available to date (21).

It is interesting to note that the N-terminal sequence of L14e from *Sulfolobus acidocaldarius* is homologous to the N-terminal sequence of Sac8b, a small, basic protein initially described by Dijk and co-workers as an 8 kDa DNA-binding protein (22, 23). *Sulfolobus* expresses a number of small, thermostable, nucleic acid binding proteins, for example, Sac7d, Sac8a, Sac8b, Sac10a, and Sac10b (22–26). The initial characterization of Sac8b was limited presumably by low expression levels. Sac8a and Sac8b differed in charge and DNA binding properties (22, 23), and gel

*This work was supported by Grant GM49686 from the National Institutes of Health to J.W.S. and S.P.E.

†Atomic coordinates and NMR data have been deposited in PDB (2kds).

‡To whom correspondence should be addressed. Phone: 256-824-2477. Fax: 256-824-6349. E-mail: shriverj@uah.edu.

filtration demonstrated that both proteins were monomeric in solution (23). No structural information was presented for 8b, other than the N-terminal 17 residue sequence (i.e., PAIEI-GYIGVETRGDEA), which demonstrated that it was unrelated to the 7 and 10 kDa proteins (23).

We report the high-resolution NMR structure of L14e from *S. solfataricus*, along with a characterization of its stability and flexibility in solution. The protein adopts an SH3 fold with an additional C-terminal tail that is largely extended, but rigidly associated with the surface of the SH3 domain. Given the high stability and classic fold, the structure is expected to serve as an excellent model of the structure of L14e in both the Archaea and eukaryotes.

MATERIALS AND METHODS

Gene Cloning, Expression, and Mutagenesis. The gene sequence for L14e was located in the *Sulfolobus solfataricus* P2 genome sequence database (27), where it is designated rpl14e (locus SSO5763) for the ribosomal protein L14e. The sequence was amplified by PCR from *Sulfolobus solfataricus* DNA, cloned into pETBlue-2 (Novagen), and expressed in RosettaBlue(DE3) (Novagen). Uniformly ^{15}N and/or ^{13}C enriched L14e for NMR studies was obtained by supplementing minimal media with $^{15}\text{NH}_4\text{Cl}$ (Isotec) and/or ^{13}C -glucose (Isotec). Protein expression was induced with IPTG (1 mM), the temperature was reduced to 27 °C for 8–10 h, cells were harvested by centrifugation, and the pellet was stored at –80 °C. For cysteine mutants used in some studies, site directed mutagenesis was used to convert cysteines 10 and 27 to alanines using the Quikchange II Site-Directed Mutagenesis kit (Stratagene, La Jolla, CA) with primers supplied by IDT (Coralville, IA). DNA sequencing by contract with Functional Biosciences (Madison, WI) was used to verify the mutations.

L14e Protein Purification. Frozen cells were thawed and suspended in 100 mL of cold (4 °C) buffer (10 mM EDTA, 10 mM Tris-HCl, pH 8.0, 0.1% Triton X-100 and 0.5 mM PMSF) and lysed by sonication. Dnase I (Sigma) was added (0.5 mg/mL), and the suspension was incubated at 37 °C for 5 min. The suspension was then incubated at 70 °C for 40 min to precipitate *E. coli* proteins, and the solution was clarified by centrifugation at 300000g. The supernatant was filtered (0.45 μm filter) and recombinant L14e was purified by cation exchange chromatography on a Hi-trap SP (Pharmacia) column equilibrated with 10 mM KH_2PO_4 (pH 7). L14e was eluted with a linear 0–1.0 M NaCl gradient at about 0.38 M NaCl. Protein purity was demonstrated by SDS gel electrophoresis, and a molecular weight of 10727 determined by mass spectrometry (by contract with Stanford University Mass Spectrometry Laboratory) indicated that the N-terminal methionine was removed by methionine aminopeptidase. The protein sequence derived from the genome sequence, NP_341932 (Figure 1), was confirmed by the NMR assignments described below. For consistency, we have maintained the residue numbering scheme used in the database sequences (e.g., NP_341932), with the NMR structure sequence beginning with proline 2. The protein concentration was determined using a calculated (28) $A_{280}^{0.1\%}$ of $0.129 \text{ mL} \cdot (\text{mg} \cdot \text{cm})^{-1}$, which includes the expected contribution from a disulfide.

Differential Scanning Calorimetry. DSC¹ measurements were performed using a MicroCal (Northampton, MA)

Extended Range VP-DSC with a cell volume of 0.5 mL. Protein samples were dialyzed against an excess of the appropriate buffer for approximately 12 h, and aliquots of protein and dialysis buffer were degassed with stirring under vacuum using the MicroCal accessory. Protein and buffer solutions were scanned at 1.5 °C/min from 5 to 125 °C under an applied pressure of 25 psi to prevent vaporization and bubble formation. Reversibility of thermal unfolding was demonstrated with repeated scans on the same sample. DSC data were analyzed using an IGOR Pro (Wavemetrics, Oregon) program (IgorDenat, available at <http://daffy.uah.edu/thermo>) as described elsewhere (29).

Circular Dichroism. CD measurements were performed on an Olis (Bogart, GA) spectrometer using 0.15 mg/mL protein solutions in 10 mM KH_2PO_4 , pH 7.0, 20 °C, with a 0.1 cm path length cell. The secondary structure of the protein was determined using CDPPro with a reference set of 37 proteins over a wavelength range of 185–240 nm (30).

NMR Spectroscopy. NMR spectra were collected on a Varian (Palo Alto, CA) 800 MHz (18.7 T field) INOVA NMR spectrometer using a triple resonance probe with triaxial pulsed field gradient capability, and a Varian 500 MHz (11.7 T field) INOVA NMR spectrometer with *z*-axis pulsed field gradient capability. Pulse sequences were either those provided in the Varian (Palo Alto, CA) BioPack, or were kindly provided by Dr. Lewis Kay (University of Toronto). All NMR spectra were collected at 30 °C unless indicated otherwise. ^1H Chemical shifts were referenced using sodium 4,4-dimethyl-4-silapentane-1-sulfonate (DSS) as an internal reference. ^{13}C and ^{15}N Chemical shifts were referenced indirectly to DSS and liquid ammonia, respectively, using the appropriate frequency ratios (31). NMR spectra were processed using NMRPipe (32), FELIX (Accelrys, San Diego, CA), or VNMR (Varian, Palo Alto, CA). NMRView (33) was used for visualization and chemical shift assignments of NMR data.

NMR samples were prepared using lyophilized ^{15}N , ^{13}C -enriched L14e which was dissolved in 700 μL of 90% $\text{H}_2\text{O}/10\%$ D_2O or in 700 μL of 99.996% D_2O . Final protein concentrations were approximately 1 mM, and the pH was adjusted to 5.0 using a Radiometer glass electrode with either HCl or NaOH for 90% $\text{H}_2\text{O}/10\%$ D_2O solutions, or DCl or KOD for 99.996% D_2O solutions. No correction was made for the deuterium isotope effect on pH.

^1H – ^{15}N residual dipolar coupling measurements were made at 11.7 T using partially aligned ^{15}N labeled L14e in a liquid crystalline media of *n*-alkyl-poly(ethylene glycol) and hexanol (34). Samples were prepared in 90% $\text{H}_2\text{O}/10\%$ D_2O with 5% C12E5 (Sigma) with a C12E5/hexanol molar ratio (*r*) of 1.0. Aligned spectra were obtained at 30 °C, and unaligned spectra were obtained at 30° after heating to 40 °C. Single bond ^1H – ^{15}N RDCs were obtained using a 3D HNCO-IPAP pulse sequence (hnco_hn_coupling_notrosy_ipap_lek) kindly provided by Lewis Kay (University of Toronto). Raw data sets contained 844 data points in the ^1H dimension and 42 increments in the ^{15}N dimension. Data were linear predicted by a factor of 2 in the ^{15}N dimension and zero filled to obtain $2048 \times 64 \times 512$ ($^1\text{H} \times ^{13}\text{C} \times ^{15}\text{N}$) data points in the final spectrum. Residual dipolar couplings were extracted using in-house NMRView scripts.

$[\text{H}]-^{15}\text{N}$ NOE and ^{15}N T_1 and $T_{1\rho}$ measurements were performed at 11.7 T using sensitivity enhanced gradient selected HSQC pulse sequences (35, 36). ^1H saturation for $[\text{H}]-^{15}\text{N}$ NOEs was obtained using a series of 120° ^1H pulses with 5 ms separation. Saturation was performed for 5 s, with a total recycle

¹Abbreviations: DSC, differential scanning calorimetry; RDC, residual dipolar coupling; rmsd, root-mean-square deviation.

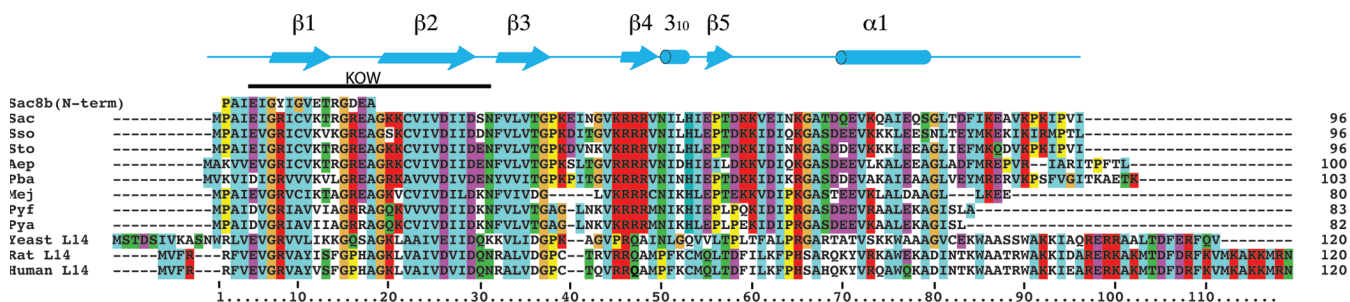


FIGURE 1: Sequence alignments of L14e proteins from *Sulfolobus acidocaldarius* (Sac), *S. solfataricus* (Sso), *S. tokodaii* (Sto), *Aeropyrum pernix* (Aep), *Pyrobaculum* (Pba), *Methanococcus janashii* (Mej), *Pyrococcus furiosus* (Pyf), *Pyrococcus abyssi* (Pya), yeast, rat, and human. The sequences of both rat and human L14e are truncated at 120 residues. Also shown is the N-terminal sequence of Sac8b as reported by Dijk et al. (23). (The N-terminal methionine was not present in the *E. coli* expressed L14e.) The regions corresponding to regular secondary structure are indicated at the top in blue, and the region of the KOW motif is indicated below.

delay of 10 s for both saturated and unsaturated experiments. Suppression of cross-correlation effects in ^{15}N $T_{1\rho}$ experiments was obtained using a train of phase alternated, random length ^1H CW pulses applied during the ^{15}N spin lock (36). T_1 delay times were 0.05, 0.10, 0.16, 0.22, 0.29, 0.36, 0.45, 0.55, and 0.66 s. $T_{1\rho}$ delay times were 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.08, and 0.10 s. R_1 and $R_{1\rho}$ data were fit with the CURVEFIT module of MODELFREE (37). R_2 relaxation rates were calculated from $R_{1\rho}$ rates as described by Korzhnev et al. (36). Duplicate $[^1\text{H}]-^{15}\text{N}$ NOE experiments and duplicates of ^{15}N T_1 and $T_{1\rho}$ experiments at three different delay times were used to estimate uncertainties in peak intensities. The errors in measured R_1 , R_2 , and NOE values averaged 0.5, 1.0, and 1.9%, respectively. The errors used for relaxation analysis were those measured or 2.5%, whichever was greater.

NMR Assignments. Backbone ^1H , ^{13}C , and ^{15}N assignments were obtained using $^1\text{H}-^{15}\text{N}$ heteronuclear single quantum coherence (HSQC), HNCA, HNCB, HNCACB, CBCA(CO)NH, HN(CO)CA and HNHA spectra collected at 18.7 T on a uniformly ^{15}N , ^{13}C -double-labeled protein sample dissolved in 90% $\text{H}_2\text{O}/10\%$ D_2O . Side chain assignments were obtained for $^{15}\text{N}-^{13}\text{C}$ -labeled L14e using HCC-TOCSY-NNH and CCC-TOCSY-NNH in 90% $\text{H}_2\text{O}/10\%$ D_2O , and HCCH-TOCSY and HCCH-COSY in 99.996% D_2O . HBCBCGCDHD (38), 2D DQF-COSY, 2D NOESY, and $^1\text{H},^{13}\text{C}$ HSQC spectra were collected on a doubly labeled protein sample in 90% $\text{H}_2\text{O}/10\%$ D_2O to obtain aromatic proton and carbon assignments. An ^{15}N -edited nuclear Overhauser effect spectrum (NOESY) (150 ms mixing time) on double labeled protein in 90% $\text{H}_2\text{O}/10\%$ D_2O , and a NOESY- ^{13}C HSQC experiment (150 ms mixing time) on protein in 99.996% D_2O were collected for NOE measurements. Backbone amide protons involved in hydrogen bonding were delineated with a $^1\text{H},^{15}\text{N}$ HSQC spectrum of ^{15}N labeled sample protein immediately after dissolving in 99.996% D_2O .

NMR Structure Calculation. Initial structures of L14e were derived using ARIA 1.2 interfaced to CNS (39) with a random chain starting structure. Initial restraints consisted of hydrogen bond and dihedral angle restraints obtained from chemical shift indices, HNHA coupling constants, NH residual dipolar couplings (RDCs), a set of unambiguous backbone NOEs obtained from a ^{15}N -edited NOESY spectrum, and sets of unassigned NOEs from ^{15}N -edited and ^{13}C -edited NOESY spectra. Initial values for the dipolar coupling (D_a) and rhombicity (R) were determined from a powder distribution of RDC's (40). In each ARIA iteration, distance restraints were calibrated and NOE assignments of the NOESY crosspeaks were made by ARIA based on an ensemble of the lowest energy structures from the

previous iteration. The top five ARIA structures were used to calculate new values for D_a and R using Module (41), the structures averaged and used as a starting structure for Xplor-NIH refinement. Restraints at this stage consisted of the ARIA-determined distance restraints, CSI-derived torsional angles, HNHA couplings, and NH RDCs. Hydrogen bonding partners of the nonexchangable NH's were identified using the top 10 out of 100 Xplor-NIH structures and used for subsequent refinements. T_1 and T_2 relaxation restraints that did not fall in suspected mobile regions, which were identified by significantly low $[^1\text{H}]-^{15}\text{N}$ NOEs, were included in subsequent refinements, and another round of Xplor-NIH (from random starting structures) was performed in which the anisotropy parameters D_a , $D_{||}/D_{\perp}$, R , and τ_c were also refined as described below. The best 10 (out of 100) structures were averaged and used as a starting structure for a final round of Xplor-NIH refinement in which D_a , $D_{||}/D_{\perp}$, R , and τ_c were fixed to the values found in the previous iteration. This final refinement also included a disulfide bond between C10 and C22 and contributions from the Xplor-NIH Ramachandran force field, but omitted the Aria ambiguous restraints. The final restraints and the statistics on the ensemble of the best 10 (out of 100) structures are summarized in Table 1. The model with the lowest backbone rmsd to the average structure was taken as representative of the ensemble.

Xplor-NIH internal dynamics were used with an initial equilibration of 3000 K for random starting structure and 1000 K for refinements starting from a defined structure, with slow cooling to 100 K with 50 K temperature decrements, followed by energy minimization. During both the equilibration and cooling phases, distance restraint force constants were increased from 5 to 50 kcal/mol, and force constants for dipolar coupling and T_1/T_2 restraints were increased during the cooling phase from 0.01 to 1.0, and 0.5 to 5.0, respectively. To find optimal values of D_a , $D_{||}/D_{\perp}$, R (rhombicity), and the overall rotational correlation time τ_c , a grid search with rigid body rotation was implemented at each temperature change during the simulated annealing. The final values characterizing the molecular alignment tensor were a D_a of -8.7 Hz and a rhombicity of 0.29. The rotational diffusion anisotropy defined by the T_1/T_2 restraints converged to a $D_{||}/D_{\perp}$ of 0.95 with a rhombicity of 0.0 and an overall τ_c of 5.2 ns. The quality of the final structures were analyzed using PROCHECK-NMR. The coordinates and resonance assignments have been deposited in the Protein Data Bank (PDB ID code 2kds).

$[^1\text{H}]-^{15}\text{N}$ NOE and ^{15}N Relaxation Data Analysis. $[^1\text{H}]-^{15}\text{N}$ NOE and ^{15}N R_1 and R_2 relaxation data were analyzed using TENSOR2 kindly provided by M. Blackledge

Table 1: Restraints and Structural Statistics for L14E^a

no. of experimental restraints ^b	
unambiguous NOE distances	1262
intramolecular	580
sequential	276
medium range	181
long range	225
ambiguous NOE distances ^c	283
H-bonds ^d	22
dihedral angle restraints	84
³ J _{HNHα} coupling constants	75
residual dipolar couplings	88
T ₁ /T ₂ relaxation ratios	48
number of restraint violations	
NOE distances with violations > 0.5 Å	2.0
dihedral angles with violations > 10°	4.0
rmsd of restraint violations	
distance restraints (Å)	0.06
dihedral angles (°)	4.17
³ J _{HNHα} (Hz)	2.04
residual dipolar couplings (Hz)	1.23
T ₁ /T ₂ relaxation ratios	0.15
rmsd from ideal geometry	
bonds (Å)	0.006
angles (°)	1.01
impropers (°)	1.32
coordinate rmsd ^e	
backbone (well-defined regions) (Å)	0.55
backbone (all residues) (Å)	1.90
Ramachandran plot statistics	
most favored (%)	76.2
allowed (%)	15.1
generously allowed (%)	6.3
disallowed (%)	2.4

^a rmsd's, violations, and Ramachandran analysis are on the ensemble of 10 best structures. ^b Number of restraints used for Xplor-NIH refinements. ^c Ambiguous restraints were omitted for final Xplor-NIH refinement. ^d Seven of the 22 H-bonds were ambiguously assigned to two or three possibilities. Each H-bond was defined by two restraints: H...O and N...O distances. ^e Coordinate rmsd's determined by superposition of 10 best Xplor-NIH structures over the well-defined regions: residues 8–12, 20–28, 32–38, 42–57, 67–83.

(<http://www.ibs.fr/ext/labos/LRMN/softs/>). The diffusion tensor elements were calculated using data for NHs only in α or β secondary structures with NOEs greater than 0.7, and R_2/R_1 ratios within 1 standard deviation of the mean. The principal elements of the diffusion tensor elements were as follows: $D_x = 2.8 \times 10^7 \text{ s}^{-1}$, $D_y = 3.5 \times 10^7 \text{ s}^{-1}$, and $D_z = 3.5 \times 10^7 \text{ s}^{-1}$. Similar results were obtained using the quadric_diffusion program kindly provided by A. Palmer (Columbia University). Order parameters characterizing backbone dynamics were obtained with TENSOR2 by fitting the ¹⁵N relaxation data according to the "model-free" approach of Lipari and Szabo and anisotropic rotation defined by the above tensor elements.

RESULTS

Initial Characterization. The *S. solfataricus* L14e gene was located in the genome database, cloned using PCR, and expressed in *E. coli* using standard procedures as described in Materials and Methods. SDS-PAGE gel electrophoresis of the recombinant protein indicated a molecular mass of approximately 12 kDa, and gel exclusion chromatography indicated a slightly larger mass of 13 kDa. Both are somewhat larger than expected from the amino acid sequence, that is, 10 861 Da. Mass spectrometry demonstrated a molecular weight of 10 727, consistent with removal of the N-terminal methionine by methionine aminopeptidase in

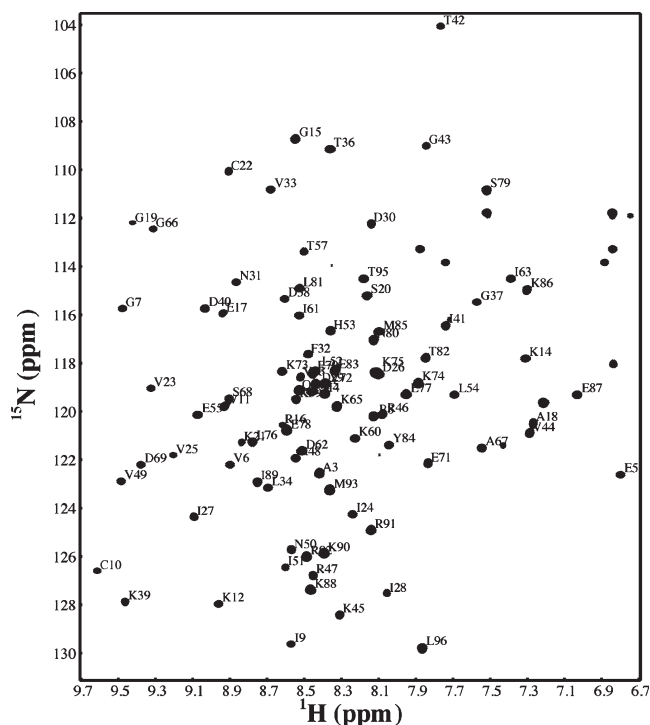


FIGURE 2: ¹H,¹⁵N HSQC spectrum (18.1 T magnetic field) of L14e at 35 °C, pH 5.0. Amide NH assignments are indicated by residue and number. Amide side chain peaks are not labeled.

E. coli. Circular dichroism spectra of purified protein indicated that only about half of the protein was composed of regular secondary structure with 6 (±2) % α -helix, 34 (±9) % β -sheet, 9 (±4) % turns, and 53 (±1) % random coil and other structures.

NMR Spectroscopy and Solution Structure. The ¹H-¹⁵N HSQC spectrum at 18.7 T was well resolved with 91 cross peaks, consistent with 95 residues and no contribution from four prolines, with one at the N-terminus (Figure 2). Backbone and side chain assignments were obtained using triple resonance experiments as described in Materials and Methods. Except for the C-terminal leucine, all backbone atoms and 96% of non-exchangeable side chain protons were assigned. Chemical shift indices indicated β -sheet secondary structure extending over residues 9–15, 20–23, 25–27, 31–36, 46–49, and 59–64, and α -helix over residues 69–79 and 82–86.

NOESY cross peak assignments were obtained using ARIA 1.2, resulting in 283 ambiguous and 1262 unambiguous NOESY cross peak assignments. Structures were obtained with Xplor-NIH using NOE distance constraints, chemical shift indices, hydrogen bonds, backbone scalar coupling constants, residual dipolar couplings, and ¹⁵N relaxation rates. Table 1 summarizes the restraints used along with statistics for the ensemble of the 10 lowest energy structures, shown in Figure 3A. The quality of the fit was excellent, with a 0.55 Å root-mean-square deviation (rmsd) for the backbone atoms in well-defined regions using the 10 best structures, and an rmsd of 1.90 Å over the entire backbone. The quality of the structure was analyzed using PROCHECK (Table 1). A ribbon drawing of a representative structure (that with the lowest root-mean-square deviation relative to the average structure) is shown in Figure 3B.

Consistent with the circular dichroism spectrum, the NMR solution structure of L14e is largely β -sheet, with a little over half of the protein (residues 1–57) composed of a β -barrel core formed by five consecutive β -strands. The turn between the first

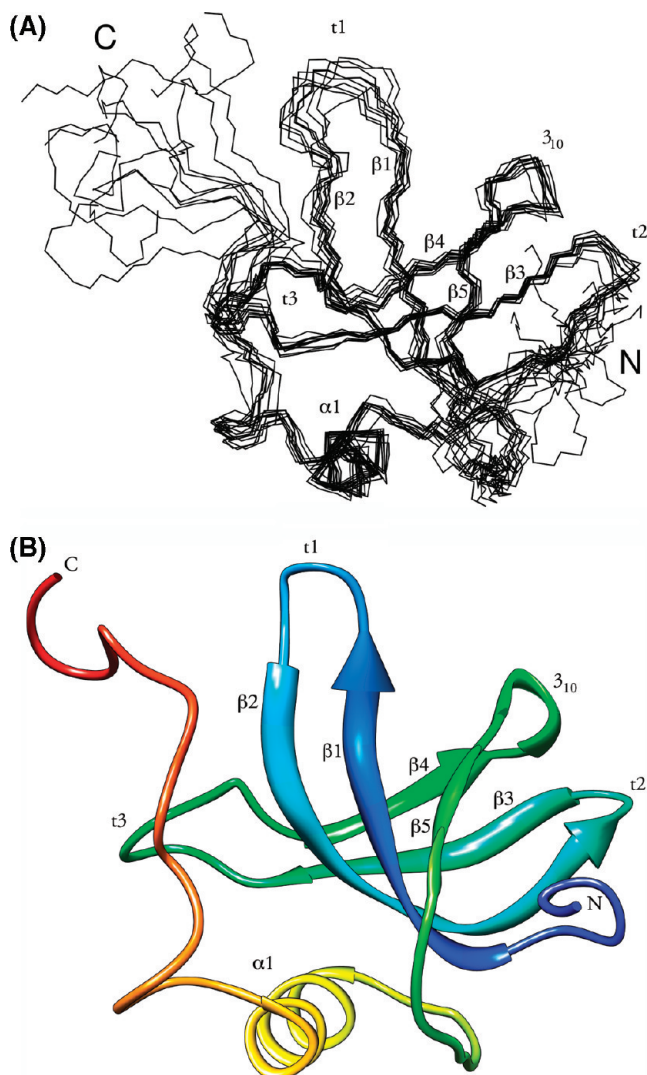


FIGURE 3: (A) Superposition of the ensemble of the 10 “best” NMR structures of L14e determined by refinement using NOEs, coupling constants, chemical shifts, ^{15}N relaxation data, and RDCs as described in the text. (B) Ribbon drawing of L14e represented by the structure in the final ensemble with the lowest root-mean-square deviation from the average structure. The orientation is similar to that in panel A.

two strands is an open loop of eight residues with little or no hydrogen bonding. The second strand is 10 residues long and contains a β -bulge which serves to accentuate left-handed twisting of the sheet. Turns 2, 3, and 4 are tight turns, and the bond angles for the three are consistent with type I turns. A 3_{10} helix bridges strand 4 and 5, allowing strand 5 to wrap around the core and complete the barrel with two hydrogen bonds with strand 1. Interestingly, proline 56 lies in the center of strand 5 and is well conserved throughout the archaeal L14e sequences (Figure 1). L14e contains two buried cysteines that lie adjacent to each other in the hydrophobic core on the first (C10) and second (C22) β -strands. The orientation of the side chains in the refined structure are consistent with a disulfide in the native fold, and both cysteines are well conserved in the archaeal L14e proteins (Figure 1).

The C-terminal portion of L14e (beginning at D58) contains a largely extended section which loops across the N-terminus of strand 1, resulting in an apparent knot. The first 10 residues stretch from the end of the last β -strand across strands 1, 2, and 3 to effectively support the β -sheet. Residues 69–79 form a helical

region that consists mainly of α -helix with a tendency for 3_{10} helix formation toward the C-terminal end of the helix. The helix and following extended chain contains a hydrophobic face composed largely of V72, L76, L81, M85, and I89 which pack against a hydrophobic patch on strands 1, 2, and 3 of the β -barrel. The importance of this interaction is reinforced by the conservation of these hydrophobic residues in the other archaeal L14e homologues (Figure 1). The packing of the C-terminal chain onto the surface of the sheet essentially doubles the size of the hydrophobic core of the protein.

The backbone rmsd's of the ensemble of NMR structures (Figure 4) indicate regions of well-defined structure (e.g., β -strands) punctuated by regions of increased rmsd values corresponding to both N- and C-terminal sequences, turns 1, 2, and 3, and the end of the last β -strand. Turn 4 consists of 3_{10} helix and is well-defined with very little structural heterogeneity in the ensemble. ^{15}N - ^1H residual dipolar coupling measurements exhibited pronounced sequence dependence (Figure 4) due to varying orientations of N–H bond vectors around the β -barrel. These were well fit by the structure refinement.

The sequence dependence of the ^{15}N order parameters (S^2) shows that increased rmsd's in some regions should not be attributed to flexibility, but are likely due to fewer constraints (Figure 4). S^2 values range from 0.85 to 0.90 over much of the protein. In addition to the terminal regions, there were two regions with enhanced flexibility compared to the main chain. These occur around the β -bulge at turn 2 (residues 27–30), and in the extended chain (residues 62–65) which stretches across β -strands 1 and 2. ^{15}N T_1/T_2 relaxation rates showed little sequence variation (data not shown) and were consistent with an axially symmetric structure, as were the results from the rotational diffusion anisotropy analysis of the relaxation data. The rotational correlation time of 5.2 ns obtained from refinement using the ^{15}N relaxation data is consistent with a monomer with a molecular mass of 10.8 kDa (i.e., τ_c (ns) \approx MW (kDa)/2).

Thermal Unfolding. As expected for a thermophile protein, L14e demonstrated significant thermal stability with a T_m of 102.8 $^\circ\text{C}$ ($\Delta H_{\text{vh}} = 145$ kcal/mol) at pH 7 and a protein concentration of 0.8 mg/mL (Figure 5). Thermal unfolding was reversible at pH 7 as judged by the reproducibility of DSC scans to the T_m . Decreasing the pH to 5 and below led to baseline artifacts and irreversible unfolding, making it impossible to obtain an estimate of the ΔC_p of unfolding using the Kirchhoff relation. Surprisingly, the T_m was slightly dependent on protein concentration, with the T_m decreasing with decreasing protein concentration (e.g., 100.8 at 0.4 mg/mL, and 100.2 $^\circ\text{C}$ at 0.2 mg/mL). Such behavior is indicative of the presence of higher order oligomers. This is consistent with the observed $\Delta H_{\text{cal}}/\Delta H_{\text{vh}}$ ratio of about 0.3, and also gel exclusion chromatography which indicated a molecular mass of 13 (± 0.3) kDa (data not shown), somewhat larger than the expected 10.8 kDa. However, as indicated above, ^{15}N relaxation data demonstrated that L14e was monomeric at 35 $^\circ\text{C}$ with a τ_c of 5.2 ns. In addition, no significant change in ^1H line width or ^1H , ^{15}N HSQC peak positions was observed with increasing temperature to 80 $^\circ\text{C}$. This would indicate that the primary species was monomeric up to at least 80 $^\circ\text{C}$. We interpret this data to indicate that L14e exists predominantly as a monomer, with a tendency to form higher order oligomers with increasing temperature, especially near 100 $^\circ\text{C}$.

To investigate a possible contribution of a C10–C27 disulfide to thermal stability, single (C27A) and double (C10A/C27A)

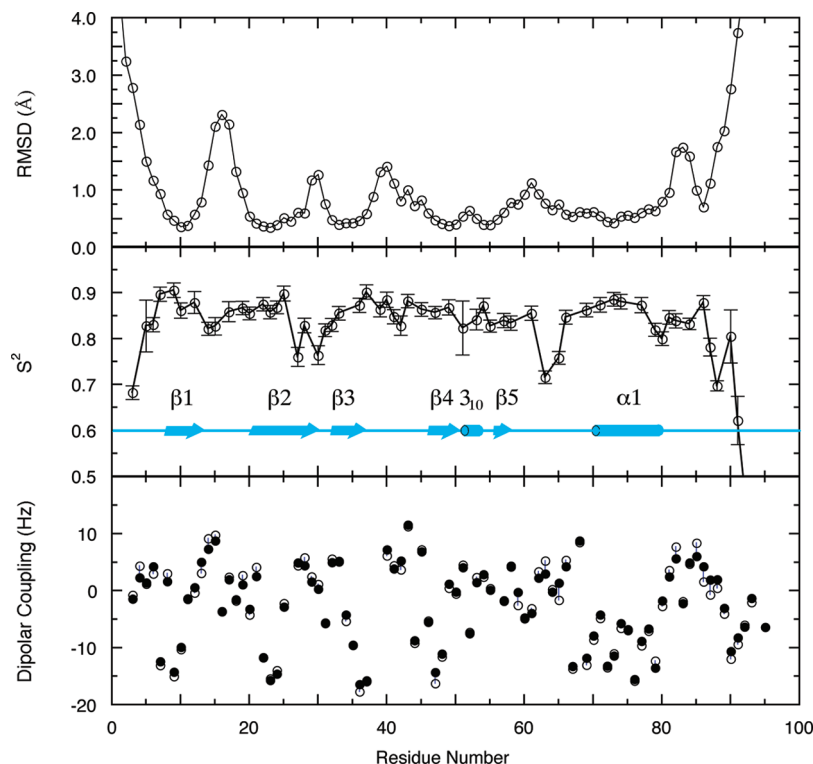


FIGURE 4: Sequence dependencies of rmsd of the ensemble of 10 “best” NMR structures of L14e (top panel); ^{15}N order parameters, S^2 (middle); and ^{15}N - ^1H residual dipolar couplings (bottom) (experimental, \circ ; calculated using refined structure, \bullet).

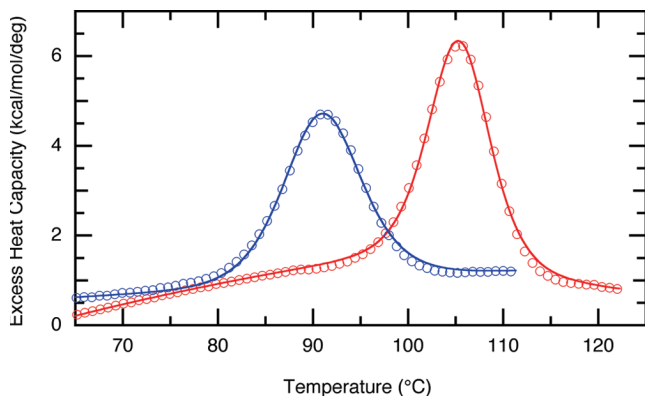


FIGURE 5: Differential scanning calorimetry of recombinant L14e (red) and the single cysteine mutant C27A (blue) in 50 mM glycine, 0.5 M KCl, pH 7. Protein concentrations: 0.63 mg/mL L14e, and 0.56 mg/mL C27A.

cysteine to alanine mutations were created. CD spectra demonstrated that the alanine substitutions resulted in negligible change in structure. In addition, ^1H , ^{15}N HSQC spectra of the mutant proteins showed minor differences except at the mutated residues, indicating that the structures of the proteins were not significantly altered by the substitutions. DSC scans of the mutant proteins were reversible at pH 7, and irreversible below pH 5, similar to the native protein. The single C22A mutant showed a decrease in T_m to 90.6 °C and a ΔH_{vh} of 97.2 kcal/mol (0.6 mg/mL). The thermodynamics of unfolding the double mutant C10A/C27A showed little additional change, with a T_m of 90.8 °C (pH 7, 0.4 mg/mL) and a ΔH_{vh} of 98.7 kcal/mol. The DSC data for both could be fit well with a ΔC_p of 1000 cal/deg/mol.

DISCUSSION

The structure of L14e is the first structure of any member of the L14e family to be reported in the literature. To our knowledge,

it is also the first structure described of a member of the COG2163 group, which includes L6e and L27e. A BLAST search of nonredundant (NCBI) genomic protein sequences using the protein sequence for *S. acidocaldarius* L14e (YP_255473) resulted in more than 66 homologous sequences (alignment scores greater than 50). Most of these have been annotated as members of the L14e family of proteins. Sequence comparisons indicate that the structures of the N-terminal domains of these proteins should be similar to L14e, that is, a classic SH3 (Src-homology region 3) domain fold with a β -barrel composed of a three-stranded antiparallel β -sheet packed orthogonally across a second three-stranded β -sheet which shares one strand (42–44). Many of these proteins contain C-terminal sequences with significant differences in length and expected domain folds. For example, the human L14e C-terminal sequence contains a bZIP domain along with a nuclear targeting sequence.

While this work was in progress, the coordinates for a lower resolution structure of L14e were deposited in the protein database (2JOY) based on a smaller NMR data set (viz. NOE distance constraints and dihedral angles). The overall folds of the two structures are similar with a backbone rmsd of 3.8 Å (Figure 6). Comparison of the SH3 domains shows the largest deviation in the loop between the first and second strands. The ^{15}N relaxation data presented here indicate that this region is no more flexible than the adjacent strands. The difference is not likely to be due to flexibility, but rather differences in the number of NMR constraints defining the region. The use of RDC's in the structure refinement presented here significantly increases the accuracy of the refinement. This is true also in the C-terminal region beyond T57, that is, beyond the SH3 fold, where the most significant difference is in the length of the helical regions.

SH3 domains have been extensively characterized because of their importance in protein–protein interactions (45, 46) with representative examples including Src (47), Csk (48), Lck (49),

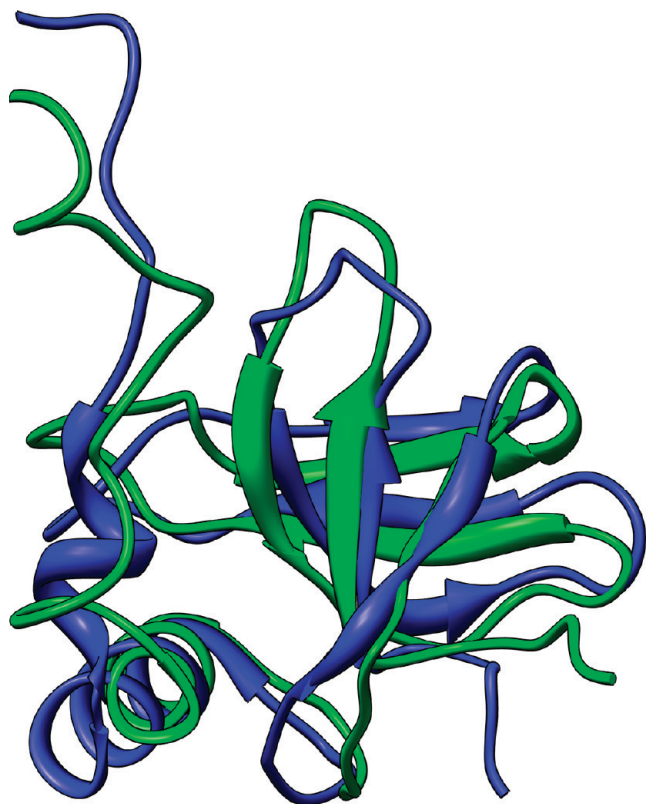


FIGURE 6: Comparison of two NMR solution structures of L14e. The structure reported here, 1KDS (blue), is compared to 2JOY (green).

Fyn (50), and Abl (50) tyrosine kinase domains, α -spectrin (43), and the adapter protein SEM-5 (51). An SH3 fold has been observed in a number of other ribosomal proteins, for example, L2 in *Bacillus stearothermophilus* (52), and L19 in *Thermus thermophilus* (6). Sequence alignments of L14e with L2 and L19 demonstrate regions of similarities that are most likely important in defining the SH3 fold. Notably, similar sequence similarities were not observed for other SH3 domain proteins such as Src, Csk, and Fyn.

SH3 domains typically recognize short (3–9 residue) proline-rich sequences with moderate affinity (e.g., a 0.001 M K_d) (46). The peptide binding site is usually a surface hydrophobic pocket defined by three loops: the RT-loop (between the first and second β -strands), the nSrc loop (between strands 2 and 3), and a 3_{10} helix (between strands 4 and 5). Interestingly, L14e contains a hydrophobic surface region defined by an nSrc loop and a 3_{10} helix. However, the RT-loop normally found in SH3 protein binding sites is absent and a tight turn is observed instead (Figure 7). The lack of an RT-loop would appear to make it difficult to define a surface with the appropriate pockets for a specific peptide ligand. However, this may not be necessary in the ribosome where multiple interactions could lead to the necessary specificity.

The stabilities of SH3 domains have been the focus of a number of studies (Fyn (53), Abl (54), α -spectrin (55), btk (56), itk (56), tec (56), drk (57), and src (58)) which have demonstrated that the fold can achieve a wide range of stability. For example, drk SH3 is marginally stable in water with significant population of the unfolded state at 25 °C, while most SH3 domains have T_m values near 70 °C. There have been few studies characterizing the stability of an SH3 fold in a thermophile protein. Sac7d and Sso7d adopt a fold that is similar to the SH3 domain, but labeling these as SH3 proteins is probably inaccurate due to the absence of

a fifth β -strand which packs against strand 1 (Figure 7). Unfortunately, a rigorous characterization of the thermodynamics of unfolding L14e is not possible due to the inability to obtain an accurate ΔC_p . Using the T_m as a measure of stability indicates that the *Sulfolobus* L14e is significantly more stable than any SH3 studied to date. The increased stability can be attributed in part to the presence of an internal disulfide. In addition, the C-terminal helix packs against a hydrophobic patch on the surface of the SH3 domain to significantly increase the size of the hydrophobic core with I63, A67, V72, L76, L81, and I89, packing against I9, V23, V25, L34, and T36. The C-terminal “meandering” sequence significantly augments the structure and more than doubles the hydrophobic core of the SH3 fold (composed of V6, C10, C 22, I24, I27, V33, V35, V44, V49, L54).

The L14e sequence from R8 to D30 shows similarity to the KOW motif, a putative RNA-binding sequence identified in the NusG bacterial transcription factor and also ribosomal protein families L24p, L26e, and L27e (59). The KOW motif is 27 residues long and appears in the NusG domain III from E196 to K222, which encompasses the first and second β -strands of an SH3 fold (60). Similarly, the R8–D30 sequence of L14e occurs in the first two β -strands of the SH3 domain. The RNA binding interactions in NusG have been attributed to electrostatic interactions with positively charged residues in the loop regions preceding, between, and following the two strands based on the L24p structure in the bacterial ribosome (60). Whether or not these can be generalized to other KOW sequences is not clear. The KOW sequence does not in general present a positively charged surface (61), although it has been pointed out that RNA interactions can occur with negatively charged protein surfaces that would normally be thought to prevent binding (60). Many conserved KOW residues (e.g., the central glycines) are more likely involved in defining the SH3 fold than in making interactions with nucleic acid. A large portion of the outer face of the two strands in the KOW sequence in L14e is not only positively charged but contains significant hydrophobic character as well (V6, I9, V11, V23, V27, I29). This face is largely occluded by association with the C-terminal sequence. Of course, the C-terminus region may adopt a different structure in complex with the ribosome to permit the KOW strands to interact with RNA. Interestingly, sequential homology between L14e and L24e indicates that L24e could also adopt an SH3 fold with additional protein sequence beyond the final β -strand of the SH3 fold.

The L14e sequence alignments shown in Figure 1 demonstrate a remarkably well-conserved stretch of four positively charged residues in β -strand 4 of the archaeal proteins ($^{45}\text{KRRR}^{48}$ in *S. solfataricus*) (Figure 8). This is a classic NLS sequence which has been demonstrated to be sufficient to drive protein localization to the nucleus in eukaryotes (62–66). Since archaea do not have nuclei, this cannot be the role of this sequence here. NLS motifs have been noted in ribosomal proteins (e.g., S12p and S17p) of bacteria, which also lack nuclei (66). In these cases, it has been argued that these “pre-existing” basic motifs may have been utilized by early eukaryotic cells for NLS sequences. The fact that the sequence is not conserved in the eukaryote examples of L14e would argue that this 4 residue basic motif did not serve as an evolutionary precursor to an NLS in L14e. Rather, it seems likely that the role is to form a highly positively charged surface on L14e that is important in defining the structure of the ribosome.

As noted in the introduction, the amino terminus of L14e is nearly identical to that of the Sac8b protein described by Dijk

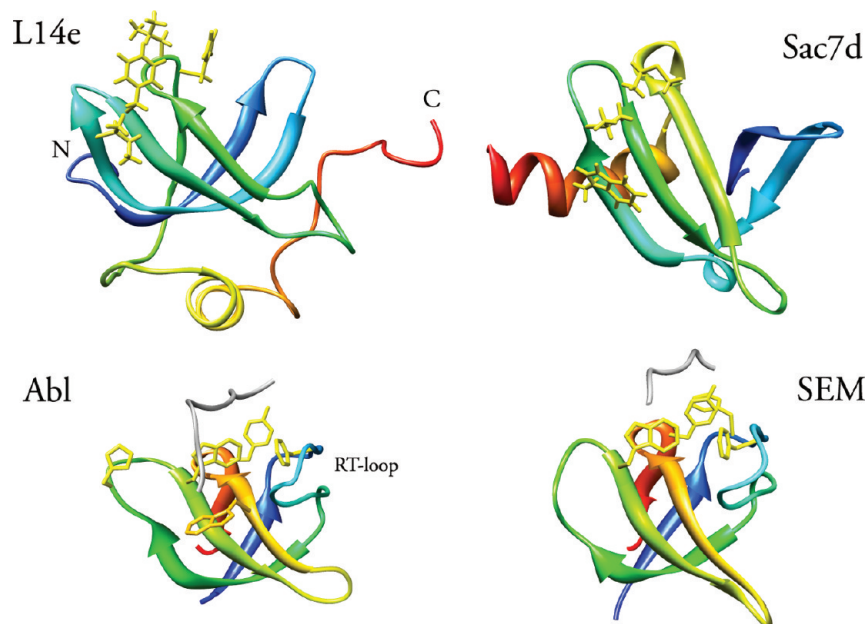


FIGURE 7: Comparison of the structure of L14e with two SH3 proteins (Abl and SEM-5) and Sac7d. The location of the RT-loop is indicated for Abl. The backbone of bound peptide fragments are shown (gray) in the peptide binding pocket of Abl and SEM-5, with protein hydrophobic side chains in yellow. (PDB files: L14e: 1KDS; Abl: 1BBZ; SEM-5: 1SEM; Sac7d: 1SAP).

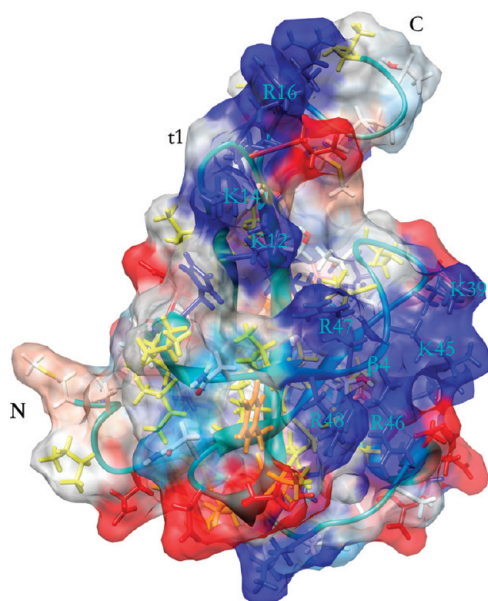


FIGURE 8: The positively charged surface of L14e defined by the KRRR sequence in β -strand β 4 (K45, R46, R47, and R48). The orientation of the protein is similar to that in Figure 3. The positive surface is extended with additional positively charged side chains of K12, K14, and R16. Notably, the following E17 is conserved in all the archaeal L14e examples (Figure 1).

et al. (23). Using the genome sequences for *S. acidocaldarius* (67), *S. tokodaii* (68), and *S. solfataricus* (27), we have searched for *Sulfolobus* sequences coding for 8 kDa proteins with N-terminal sequences similar to the sequence reported by Dijk et al. (23). Surprisingly, we find none. In fact, the most likely candidates are L14e proteins: YP_255473 in *S. acidocaldarius*, NP_376274.1 in *S. tokodaii*, and NP_341932 in *S. solfataricus*, all of which have N-terminal sequences homologous to the 17-residue sequence reported by Dijk et al. (23) except for the initiating methionine (Figure 1). L14e from *S. acidocaldarius* shows four differences that could be due to either strain differences or errors in sequencing. Sac8b was reported to be present in low amounts,

and the possibility of sequence errors due to the presence of other proteins cannot be assessed. We note that the lower than expected molecular weight reported for Sac8b is apparently real, since recombinant L14e does not migrate on SDS gels between Sac7d and Sac10a as reported for Sac8b (23). The structure indicates that proteolysis in the vicinity of the solvent exposed C-terminal helix (e.g., at K73) would result in a protein with an approximate molecular mass of 8 kDa. Whether or not this is the case will require more detailed characterization of native Sac8b from *S. acidocaldarius*.

REFERENCES

1. Lecompte, O., Ripp, R., Thierry, J. C., Moras, D., and Poch, O. (2002) Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res.* 30, 5382–5390.
2. Wimberly, B. T., Brodersen, D. E., Clemons, W. M.Jr., Morgan-Warren, R. J., Carter, A. P., Vornrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature* 407, 327–339.
3. Brodersen, D. E., Clemons, W. M.Jr., Carter, A. P., Wimberly, B. T., and Ramakrishnan, V. (2002) Crystal structure of the 30S ribosomal subunit from *Thermus thermophilus*: structure of the proteins and their interactions with 16S RNA. *J. Mol. Biol.* 316, 725–768.
4. Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., and Yonath, A. (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* 107, 679–688.
5. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science* 292, 883–896.
6. Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., and Cate, J. H. (2005) Structures of the bacterial ribosome at 3.5 Å resolution. *Science* 310, 827–834.
7. Korostelev, A., Trakhanov, S., Laurberg, M., and Noller, H. F. (2006) Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell* 126, 1065–1077.
8. Selmer, M., Dunham, C. M., Murphy, F. V. t., Weixlbaumer, A., Petry, S., Kelley, A. C., Weir, J. R., and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313, 1935–1942.
9. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289, 905–920.

10. Spahn, C. M., Beckmann, R., Eswar, N., Penczek, P. A., Sali, A., Blobel, G., and Frank, J. (2001) Structure of the 80S ribosome from *Saccharomyces cerevisiae*—tRNA-ribosome and subunit-subunit interactions. *Cell* 107, 373–386.
11. Chandramouli, P., Topf, M., Menetret, J. F., Eswar, N., Cannone, J. J., Gutell, R. R., Sali, A., and Akey, C. W. (2008) Structure of the mammalian 80S ribosome at 8.7 Å resolution. *Structure* 16, 535–548.
12. Doudna, J. A., and Rath, V. L. (2002) Structure and function of the eukaryotic ribosome: the next frontier. *Cell* 109, 153–156.
13. Tsurugi, K., Collatz, E., Todokoro, K., Ulbrich, N., Lightfoot, H. N., and Wool, I. G. (1978) Isolation of eukaryotic ribosomal proteins. Purification and characterization of the 60 S ribosomal subunit proteins La, Lb, Lf, Pl, P2, L13', L14, L18', L20, and L38. *J. Biol. Chem.* 253, 946–955.
14. Ulbrich, N., Lin, A., and Wool, I. G. (1979) Identification by affinity chromatography of the eukaryotic ribosomal proteins that bind to 5.8 S ribosomal ribonucleic acid. *J. Biol. Chem.* 254, 8641–8645.
15. Wool, I. G., Chan, Y. L., and Gluck, A. (1995) Structure and evolution of mammalian ribosomal proteins. *Biochem. Cell Biol.* 73, 933–947.
16. Alonso, J., and Santaren, J. F. (2006) Characterization of the *Drosophila melanogaster* ribosomal proteome. *J. Proteome Res.* 5, 2025–2032.
17. Davies, C., White, S. W., and Ramakrishnan, V. (1996) The crystal structure of ribosomal protein L14 reveals an important organizational component of the translational apparatus. *Structure* 4, 55–66.
18. Chan, Y. L., Olvera, J., and Wool, I. G. (1996) The primary structure of rat ribosomal protein L14. *Biochem. Biophys. Res. Commun.* 222, 427–431.
19. Tanaka, M., Tanaka, T., Harata, M., Suzuki, T., and Mitsui, Y. (1998) Triplet repeat-containing ribosomal protein L14 gene in immortalized human endothelial cell line (t-HUE4). *Biochem. Biophys. Res. Commun.* 243, 531–537.
20. Hasegawa, H., Uchiumi, T., Sato, T., Saito, A., Nakano, M., and Gejyo, F. (2002) Autoantibody against ribosomal protein L14 in patients with systemic lupus erythematosus. *Clin. Exp. Rheumatol.* 20, 139–144.
21. Klein, D. J., Moore, P. B., and Steitz, T. A. (2004) The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.* 340, 141–177.
22. Grote, M., Dijk, J., and Reinhardt, R. (1986) Ribosomal and DNA binding proteins of the thermoacidophilic archaeobacterium, *Sulfolobus acidocaldarius*. *Biochim. Biophys. Acta* 873, 405–413.
23. Dijk, J., Reinhardt, R. (1986) The structure of DNA-binding proteins from eu- and archaeobacteria, in *Bacterial Chromatin* (Gualerzi, C., Pon, C., Eds.), pp 185–218, Springer-Verlag, Berlin.
24. Lurz, R., Grote, M., Dijk, J., Reinhardt, R., and Dobrinski, B. (1986) Electron microscopic study of DNA complexes with proteins from the archaeobacterium, *Sulfolobus acidocaldarius*. *EMBO J.* 5, 3715–3721.
25. Edmondson, S. P., and Shriver, J. W. (2001) DNA binding proteins Sac7d and Sso7d from *Sulfolobus*. *Methods Enzymol.* 334, 129–145.
26. White, M. F., and Bell, S. D. (2002) Holding it together: chromatin in the archaea. *Trends Genet.* 18, 621–626.
27. She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P. M., Heikamp-de Jong, I., Jeffries, A. C., Kozera, C. J., Medina, N., Peng, X., Thi-Ngoc, H. P., Redder, P., Schenk, M. E., Theriault, C., Tolstrup, N., Charlebois, R. L., Doolittle, W. F., Duguet, M., Gaasterland, T., Garrett, R. A., Ragan, M. A., Sensen, C. W., and Van der Oost, J. (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. U.S.A.* 98, 7835–7840.
28. Gill, S., and von Hippel, P. (1989) Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data. *Anal. Biochem.* 182, 319–326.
29. Shriver, J. W., Peters, W. B., Szary, N., Clark, A. T., and Edmondson, S. P. (2001) Calorimetric analyses of hyperthermophile proteins. *Methods Enzymol.* 334, 389–422.
30. Sreerama, N., and Woody, R. W. (2000) Estimation of protein secondary structure from CD spectra: Comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set. *Anal. Biochem.* 282, 252–260.
31. Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E., and Wuthrich, K. (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. *J. Mol. Biol.* 280, 933–952.
32. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277–293.
33. Johnson, B. A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* 278, 313–352.
34. Ruckert, M., and Otting, G. (2000) Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* 122, 7793–7797.
35. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* 33, 5984–6003.
36. Korzhnev, D. M., Skrynnikov, N. R., Millet, O., Torchia, D. A., and Kay, L. E. (2002) An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. *J. Am. Chem. Soc.* 124, 10743–10753.
37. Mandel, A. M., Akke, M., and Palmer, A. G. (1995) Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.* 246, 144–163.
38. Yamazaki, T., Forman-Kay, J. D., and Kay, L. E. (1993) Two-dimensional NMR experiments for correlating Carbon-13 beta and proton delta/epsilon chemical shifts of aromatic residues in ¹³C labelled proteins via scalar couplings. *J. Am. Chem. Soc.* 115, 11054–11055.
39. Linge, J. P., Habeck, M., Rieping, W., and Nilges, M. (2003) ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* 19, 315–316.
40. Clore, G. M., Gronenborn, A. M., and Bax, A. (1998) A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J. Magn. Reson.* 133, 216–221.
41. Dosset, P., Hus, J. C., Marion, D., and Blackledge, M. (2001) A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J. Biomol. NMR* 20, 223–231.
42. Musacchio, A., Gibson, T., Lehto, V. P., and Saraste, M. (1992) SH3—an abundant protein domain in search of a function. *FEBS Lett.* 307, 55–61.
43. Musacchio, A., Noble, M., Pauptit, R., Wierenga, R., and Saraste, M. (1992) Crystal structure of a Src-homology 3 (SH3) domain. *Nature* 359, 851–855.
44. Musacchio, A., Wilmanns, M., and Saraste, M. (1994) Structure and function of the SH3 domain. *Prog. Biophys. Mol. Biol.* 61, 283–297.
45. Kay, B. K., Williamson, M. P., and Sudol, M. (2000) The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J.* 14, 231–241.
46. Mayer, B. J. (2001) SH3 domains: complexity in moderation. *J. Cell Sci.* 114, 1253–1263.
47. Xu, W., Harrison, S. C., and Eck, M. J. (1997) Three-dimensional structure of the tyrosine kinase c-Src. *Nature* 385, 595–602.
48. Ghose, R., Shekhtman, A., Goger, M. J., Ji, H., and Cowburn, D. (2001) A novel, specific interaction involving the Csk SH3 domain and its natural ligand. *Nat. Struct. Biol.* 8, 998–1004.
49. Eck, M. J., Atwell, S. K., Shoelson, S. E., and Harrison, S. C. (1994) Structure of the regulatory domains of the Src-family tyrosine kinase Lck. *Nature* 368, 764–769.
50. Musacchio, A., Saraste, M., and Wilmanns, M. (1994) High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. *Nat. Struct. Biol.* 1, 546–551.
51. Lim, W. A., Richards, F. M., and Fox, R. O. (1994) Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature* 372, 375–379.
52. Nakagawa, A., Nakashima, T., Taniguchi, M., Hosaka, H., Kimura, M., and Tanaka, I. (1999) The three-dimensional structure of the RNA-binding domain of ribosomal protein L2; a protein at the peptidyl transferase center of the ribosome. *EMBO J.* 18, 1459–1467.
53. Plaxco, K. W., Gujjarro, J. I., Morton, C. J., Pitkeathly, M., Campbell, I. D., and Dobson, C. M. (1998) The folding kinetics and thermodynamics of the Fyn-SH3 domain. *Biochemistry* 37, 2529–2537.
54. Filimonov, V. V., Azuaga, A. I., Viguera, A. R., Serrano, L., and Mateo, P. L. (1999) A thermodynamic analysis of a family of small globular proteins: SH3 domains. *Biophys. Chem.* 77, 195–208.
55. Viguera, A. R., Martinez, J. C., Filimonov, V. V., Mateo, P. L., and Serrano, L. (1994) Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition. *Biochemistry* 33, 2142–2150.
56. Knapp, S., Mattson, P. T., Christova, P., Berndt, K. D., Karshikoff, A., Vihinen, M., Smith, C. I., and Ladenstein, R. (1998) Thermal unfolding of small proteins with SH3 domain folding pattern. *Proteins* 31, 309–319.
57. Farrow, N. A., Zhang, O., Forman-Kay, J. D., and Kay, L. E. (1997) Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. *Biochemistry* 36, 2390–2402.

58. Grantcharova, V. P., and Baker, D. (1997) Folding dynamics of the src SH3 domain. *Biochemistry* 36, 15685–15692.
59. Kyripides, N. C., Woese, C. R., and Ouzounis, C. A. (1996) KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem. Sci.* 21, 425–426.
60. Steiner, T., Kaiser, J. T., Marinkovic, S., Huber, R., and Wahl, M. C. (2002) Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. *EMBO J.* 21, 4641–4653.
61. Reay, P., Yamasaki, K., Terada, T., Kuramitsu, S., Shirouzu, M., and Yokoyama, S. (2004) Structural and sequence comparisons arising from the solution structure of the transcription elongation factor NusG from *Thermus thermophilus*. *Proteins* 56, 40–51.
62. Kalderon, D., Roberts, B. L., Richardson, W. D., and Smith, A. E. (1984) A short amino acid sequence able to specify nuclear location. *Cell* 39, 499–509.
63. Chelsky, D., Ralph, R., and Jonak, G. (1989) Sequence requirements for synthetic peptide-mediated translocation to the nucleus. *Mol. Cell. Biol.* 9, 2487–2492.
64. Honkavuori, K. S., Pollard, B. D., Rodriguez, M. S., Hay, R. T., and Kemp, G. D. (2004) Dual role of the adenovirus pVI C terminus as a nuclear localization signal and activator of the viral protease. *J. Gen. Virol.* 85, 3367–3376.
65. Lange, A., Mills, R. E., Lange, C. J., Stewart, M., Devine, S. E., and Corbett, A. H. (2007) Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J. Biol. Chem.* 282, 5101–5105.
66. Peric, M., Schedewig, P., Bauche, A., Kruppa, A., and Kruppa, J. (2008) Ribosomal proteins of *Thermus thermophilus* fused to beta-galactosidase are imported into the nucleus of eukaryotic cells. *Eur. J. Cell. Biol.* 87, 47–55.
67. Chen, L., Brugger, K., Skovgaard, M., Redder, P., She, Q., Torarinsson, E., Greve, B., Awayez, M., Zibat, A., Klenk, H. P., and Garrett, R. A. (2005) The genome of *Sulfolobus acidocaldarius*, a model organism of the Crenarchaeota. *J. Bacteriol.* 187, 4992–4999.
68. Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Otsuka, R., Nakazawa, H., Takamiya, M., Kato, Y., Yoshizawa, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Masuda, S., Yanagii, M., Nishimura, M., Yamagishi, A., Oshima, T., and Kikuchi, H. (2001) Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7. *DNA Res.* 8, 123–140.