Mohammad Wadud Bhuiya,^{a*}
Jimmy Suryadi,^b Zholi Zhou^c and
Bernard Andrew Brown II^{d*}

^aConagen Inc., Saint Louis, MO 63132, USA,

^bDepartment of Chemistry, Wake Forest
University, Winston-Salem, NC 27109, USA,

^cBristol-Myers Squibb, Syracuse, NY 13221,
USA, and ^dWomble Carlyle Sandridge and Rice
LLP, One West Fourth Street, Winston-Salem,
NC 27101, USA

Correspondence e-mail:

wadud@conagen-inc.com, bebrown@wcsr.com

Received 13 March 2013

Accepted 5 August 2013

PDB Reference: L7Ae, 2fc3

Structure of the *Aeropyrum pernix* L7Ae multifunctional protein and insight into its extreme thermostability

Archaeal ribosomal protein L7Ae is a multifunctional RNA-binding protein that directs post-transcriptional modification of archaeal RNAs. The L7Ae protein from *Aeropyrum pernix* (Ap L7Ae), a member of the Crenarchaea, was found to have an extremely high melting temperature (>383 K). The crystal structure of Ap L7Ae has been determined to a resolution of 1.56 Å. The structure of Ap L7Ae was compared with the structures of two homologs: hyperthermophilic *Methanocaldococcus jannaschii* L7Ae and the mesophilic counterpart mammalian 15.5 kD protein. The primary stabilizing feature in the Ap L7Ae protein appears to be the large number of ion pairs and extensive ion-pair network that connects secondary-structural elements. To our knowledge, Ap L7Ae is among the most thermostable single-domain monomeric proteins presently observed.

1. Introduction

Aeropyrum pernix (Ap) L7Ae is a multifunctional RNA-binding protein that recognizes the K-turn motif in ribosomal RNA, box C/D sRNAs and box H/ACA sRNAs and initiates small nucleolar ribonucleoprotein complex (sRNP) assembly (Kuhn *et al.*, 2002; Omer *et al.*, 2002; Tran *et al.*, 2003; Suryadi *et al.*, 2005; Biswas *et al.*, 2011). Box C/D and box H/ACA sRNPs direct the post-transcriptional nucleotide modification of ribosomal and other RNAs (Tollervey, 1996; Kiss, 2001; Bachellerie *et al.*, 2002; Terns & Terns, 2002; Decatur & Fournier, 2003). The occurrence of box C/D and box H/ACA guide RNAs in both eukarya and archaea suggests that these RNP enzymes arose before the divergence of these two kingdoms more than two billion years ago (Gaspin *et al.*, 2000; Omer *et al.*, 2000; Speckmann *et al.*, 2002; Tang *et al.*, 2002; Tran *et al.*, 2004).

Most archaeal organisms are hyperthermophiles that have been isolated from marine and continental volcanic environments (Bhuiya *et al.*, 2005). As most of the proteins from these organisms are stable at or above 373 K (Adams, 1993), they have drawn attention and interest for both practical and fundamental reasons. From small-scale experiments to large-scale industrial processes, it is often necessary or advantageous to use proteins that are structurally and functionally stable at high temperatures, and thermophilic proteins have been used in various fields directly or as models for protein engineering (Bruins *et al.*, 2001; Rothschild & Mancinelli, 2001). How thermophilic proteins achieve their extraordinary stability is not fully understood. To elucidate the molecular mechanism of the increased stability of thermophilic proteins, numerous studies have compared them with their mesophilic homologs. Several factors have been suggested to enhanced thermostability, such as the presence of extensive ion-pair networks on the protein surface (Karshikoff & Ladenstein, 2001; Lebbink *et al.*, 2002; Yip *et al.*, 1998; Sakuraba *et al.*, 2008; Yoneda *et al.*, 2012), hydrophobic interactions (Bhuiya *et al.*, 2005; Kim *et al.*, 2012), tight hydrophobic packing of the core (Eriksson *et al.*, 1992; Takano *et al.*, 1995), greater numbers of hydrogen bonds (Hennig *et al.*, 1997; Xiang *et al.*, 2004), protein loop shortening (Esposito *et al.*, 2002; Guy *et al.*, 2003; Van Boxtael *et al.*,

2003), increased hydrophilicity of the surface (Fersht & Serrano, 1993), decreased hydrophobic accessible surface area (Knapp *et al.*, 1999; Zhang *et al.*, 2001; Akbulut *et al.*, 2013), improved helix capping (Serrano & Fersht, 1989), additional aromatic–aromatic interactions (Burley & Petsko, 1985) and a decrease in the main-chain flexibility (Matthews *et al.*, 1987; Herning *et al.*, 1992). There is no unifying mechanism found for the increased stability. Rather, it is apparent that different thermophilic proteins have particular strategies for temperature adaptation and use combinations of subtle intramolecular interactions to gain enhanced thermal stability (Szilágyi & Závodszy, 2000; Vieille & Zeikus, 2001; Bezsudnova *et al.*, 2012). The highest denaturation temperature directly measured for a protein is near 423 K for the trimeric CutA1 protein from *Pyrococcus horikoshii* (Tanaka *et al.*, 2006). Wong *et al.* (2001) reported that L30e from *Thermococcus celer*, which has a melting temperature (T_m) of 367 K, is the one of the most stable monomeric proteins (Wong *et al.*, 2001).

Here, we report a detailed structural analysis of L7Ae from the crenarchaeal species *A. pernix* (Ap L7Ae) and a detailed biophysical and structural comparison to homologous proteins from euryarchaea (*Methanocaldococcus jannaschii* L7Ae; Mj L7Ae) and mesophilic eukaryotes (human/mouse 15.5kD).

2. Experimental procedures

2.1. Protein expression and purification

The L7Ae gene was cloned from *A. pernix* genomic DNA (ATCC) using PCR and ligated into the *Nde*I and *Hind*III restriction sites of the pET28a plasmid (Novagen). This construct produced an N-terminally His₆-tagged protein that could be cleaved with thrombin to remove the affinity tag; three vector-encoded amino acids (GSH) remained at the N-terminus after cleavage. The Ap L7Ae protein was overexpressed in *Escherichia coli* Rosetta(DE3) pLysS strain (Novagen) at 310 K by induction with isopropyl β -D-1-thiogalactopyranoside (1 mM) in LB broth. After 3 h of further growth, the cells were harvested by centrifugation, washed once with phosphate-buffered saline, re-centrifuged and frozen at 193 K. The protein was isolated from thawed cell pellets by lysis with a French press, centrifugation at 25 000g, incubation of the supernatant solution with Ni-NTA resin (Qiagen), loading onto a low-pressure column and running an imidazole step gradient as suggested by the manufacturer. Fractions containing the L7Ae protein were pooled, treated with thrombin to remove the His₆ tag and dialyzed into buffer for ion-exchange chromatography (20 mM Tris, 20 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 1 mM PMSF pH 7.5; Q anion-exchange buffer). The dialysate was further purified by FPLC anion-exchange chromatography using a RESOURCE Q column (Amersham Biosciences) and eluted using a linear concentration gradient from 20 mM to 1 M NaCl in Q anion-exchange buffer. The resulting fractions containing L7Ae were dialyzed into 10 mM HEPES–KOH pH 7.5, 20 mM NaCl, concentrated using centrifugal filters and stored at a concentration of $\sim$ 1 mM at 253 K. Yields were typically 35 mg purified protein per litre of bacterial culture. The Mj L7Ae and human/mouse 15.5kD proteins were expressed and purified using analogous methods to those used for Ap L7Ae.

2.2. Circular-dichroism spectroscopic experiments

Circular-dichroism experiments were performed by monitoring the ellipticity in 1 nm steps from 200 to 350 nm using an Aviv 215 spectrometer. L7Ae samples (5 μ M in 1 ml volumes) were prepared in 20 mM Na₂HPO₄ pH 7.5 and analyzed in 0.2 or 0.5 cm quartz

cuvettes at 298 K. Sodium perchlorate and guanidine hydrochloride (Gdn–HCl) denaturation experiments were accomplished by preparing individual samples consisting of 5 μ M L7Ae in 20 mM Na₂HPO₄ buffer pH 7.5 containing the desired concentration of denaturant or salt; concentrations ranged from 0 to 9 M for sodium perchlorate and from 0 to 6 M for Gdn–HCl. Samples were equilibrated at 298 K for at least 15 min before acquiring the spectra. A plot of the averaged ellipticity at 222 nm *versus* concentration was constructed and the free energy of denaturation (ΔG_D) was obtained by fitting the curve with a two-state denaturation model using *Origin* 7.0 (MicroCal) (Pace, 1986).

Thermal denaturation experiments were performed on native and denaturant-containing samples by first collecting a wavelength spectrum at 298 K. The sample was then cooled to 275 K and slowly heated from 275 to 383 K in 2 K steps while the ellipticity was monitored at 222 nm. After the temperature had been held constant at 383 K for 2 min, the sample was re-cooled to 275 K while monitoring the ellipticity at 222 nm. The sample was returned to 298 K and a final wavelength spectrum was collected. The last two steps tested the reversibility of the denaturation. The heating and cooling rates were 20 K min^{−1}, with 0.5 min equilibrations at each temperature; ellipticity data were collected for 10 s periods. Melting experiments were performed in duplicate and the ellipticity values at each temperature were averaged. The melting temperature (T_m) and the van't Hoff enthalpy of denaturation (ΔH_{vH}) were determined from a plot of the averaged ellipticity at 222 nm *versus* temperature (in Kelvin). The resulting curve was fitted to a two-state denaturation model implemented in the program *Origin* 7.0 (Marky & Breslauer, 1987; Puglisi & Tinoco, 1989).

2.3. Crystallization

L7Ae samples were screened for crystallization using the hanging-drop vapor-diffusion technique, and conditions for crystallization were obtained using several sparse-matrix screens (Hampton Research). Initial lead conditions appeared to require zinc salts for crystallization, and final conditions were derived by adding ZnCl₂ to the crystallization conditions established for the *M. jannaschii* L7Ae protein (Mj L7Ae; Suryadi *et al.*, 2005). Optimal crystals were obtained by mixing 3 μ l Ap L7Ae ($\sim$ 1 mM $\simeq$ 13 mg ml^{−1}) with 2 μ l 20% polyethylene glycol 4000, 10% 2-propanol, 100 mM HEPES–KOH pH 7.5, 1 mM ZnCl₂ and suspending the drops above the same solution in the well. Long needle-like crystals were visible after 24 h at 298 K and grew to maximum size after 4–5 d.

2.4. Structure determination and refinement

A single crystal was harvested from the mother liquor with a nylon loop, mounted on a goniometer and directly cooled in a liquid-nitrogen vapor stream at 100 K. Data were collected using a Rigaku MicroMax-007 microfocus rotating-anode generator coupled with a Rigaku Saturn 92 CCD detector at a distance of 45 mm from the crystal. Images (360) were collected as 5 s exposures in 0.5° oscillations (a 180° hemisphere) and were indexed and scaled using Rigaku *CrystalClear* v.1.3.6. This resulted in 34 416 reflections in the 65–1.54 Å resolution range, with 34 157 reflections above 3 σ and 13 926 unique reflections. Unit-cell parameters and data-reduction statistics are summarized in Table 1.

The Ap L7Ae structure was phased by the molecular-replacement method using *EPMR* (Kissinger *et al.*, 1999) and a model of Ap L7Ae based on the Mj L7Ae structure (Suryadi *et al.*, 2005). Based on the Matthews coefficient, one protein monomer was expected in the asymmetric unit. The solution from *EPMR* had a correlation coeffi-

Table 1Crystallographic data statistics for *A. pernix* L7Ae (PDB entry 2fc3).

Values in parentheses are for the highest resolution bin.

Space group	$P2_12_12_1$
Unit-cell parameters	
a (Å)	35.05
b (Å)	49.19
c (Å)	65.05
$\alpha = \beta = \gamma$ (°)	90
Matthews coefficient (Å ³ Da ⁻¹)	1.86
Solvent content (%)	33.40
Resolution range (Å)	65.05–1.54
No. of reflections	
Total	34157
Unique	13926 (217)
Completeness (%)	81.5 (15.8)
Multiplicity	2.45 (1.08)
$R_{\text{merge}}^{\dagger}$ (%)	2.5 (28.9)
$\langle I/\sigma(I) \rangle$	23.7 (2.3)
Reduced χ^2	0.95 (1.36)
Wilson B factor (Å ²)	17.53

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I(hkl)$ is the observed intensity of the i th measurement of reflection hkl and $\langle I(hkl) \rangle$ is the mean intensity of reflection hkl .

cient of 0.535 and an R_{cryst} of 46.3% (50–4.0 Å resolution). Refinement was initially performed using *CNS* with 5% of the data excluded from refinement for the calculation of R_{free} (Brünger *et al.*, 1998). Rigid-body refinement of the initial solution produced a model with an R_{cryst} of 47.7% and an R_{free} of 48.3% in the 50–2.0 Å resolution range. One round of simulated annealing, B -factor refinement and energy minimization significantly improved the model and resulted in a structure with an R_{cryst} of 32.2% and an R_{free} of 35.4% (50–1.56 Å). At this stage, manual fitting was initiated using *XtalView* (McRee, 1999). Cross-validated σ_A -weighted composite OMIT maps or σ_A -weighted $2F_o - F_c$, $5F_o - 3F_c$ and $F_o - F_c$ maps calculated in *CNS* were used for fitting and model building. Density modification, implemented in *CNS*, was also used to reduce model bias and to improve maps. About ten iterations of manual fitting and refinement with *CNS* in the resolution range 27–1.6 Å were performed, yielding a model with an R_{cryst} of 21.8% and an R_{free} of 23.6%. High-resolution refinement was carried out with *REFMAC5* (Winn *et al.*, 2011; Murshudov *et al.*, 2011). Eight cycles of restrained refinement with isotropic B factors were carried out, riding H atoms were added to the protein model and ten additional cycles of restrained refinement were performed. Finally, individual anisotropic B factors were refined, including contributions from riding H atoms (eight cycles), which produced a model with an R_{cryst} of 17.7% and an R_{free} of 22.5% in the resolution range 32.4–1.56 Å. The final model was analyzed using the *CCP4* programs *SFCHECK* and *PROCHECK* (Winn *et al.*, 2011; Vaguine *et al.*, 1999; Laskowski *et al.*, 1993). Refinement statistics are listed in Table 2. Coordinates and structure factors have been deposited in the Protein Data Bank as entry 2fc3.

2.5. Structure and sequence analyses

The Ap L7Ae secondary structure was assigned using *DSSP* (Kabsch & Sander, 1983), surface atoms were analyzed by *GETAREA* (Fraczkiewicz & Braun, 1998) and intermolecular interactions, solvent accessibility and molecular volumes were analyzed using the *CCP4* programs *CONTACT*, *NCONT*, *ACT*, *SURFACE* and *VOLUME* (Winn *et al.*, 2011). Hydrogen bonding was analyzed with *HBPLUS* (McDonald & Thornton, 1994). Ion pairs were identified as potential ionic interactions between two oppositely charged residues; they were characterized in three distance ranges: 4, 6 and 8 Å. The protein-sequence data used in generating Fig. 5 are based on

Table 2Refinement statistics for *A. pernix* L7Ae (PDB entry 2fc3).

Values in parentheses are for the highest resolution bin.

Resolution (Å)	32.44–1.56
$R_{\text{cryst}}^{\dagger}$ (%)	17.71 (24.3)
$R_{\text{free}}^{\ddagger}$ (%)	22.55 (87.5)
Luzzati coordinate error (Å)	0.21
Luzzati coordinate error, R_{free} (Å)	0.24
Luzzati σ_A (Å)	0.12
Luzzati σ_A , R_{free} (Å)	0.17
$F_o - F_c$ correlation coefficient	0.959
$F_o - F_{c,\text{free}}$ correlation coefficient	0.933
R.m.s.d. from ideal geometry	
Bond lengths (Å)	0.014
Bond angles (°)	2.012
Ramachandran plot analysis (%)	
Most favored regions	98.1 [105 of 107 amino acids]
Additional allowed regions	2.8 [3 of 107 amino acids]
Glycine or proline residues	12.1 [15 of 124 amino acids]
Terminal residues	1.6 [2 of 124 amino acids]
B factors (Å ²)	
Average	21.35
Protein backbone	18.69
Protein side chains	21.01
Solvent molecules	33.38
Total No. of atoms	1076
No. of protein atoms	957
No. of waters	119

$\dagger R_{\text{cryst}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$. $\ddagger R_{\text{free}}$ was monitored with 695 reflections out of 13 042 (~5%) that were excluded from refinement.

NCBI accession Nos. NP_148187 (*A. pernix* L7Ae), NP_126357 (*Pyrococcus abyssi* L7Ae), NP_248198 (*M. jannaschii* L7Ae), NP_069598 (*Archaeoglobus fulgidus* L7Ae), 1JJ2F (*Haloarcula marismortui* L7Ae), NP_010888 (*Saccharomyces cerevisiae* Snu13p), NP_004999 (*H. sapiens* 15.5kD), CAD98427 (*Cryptosporidium parvum* Nhp2p), CAC08452 (*H. sapiens* Nhp2), P29160 (*Thermococcus celer* L30) and NP_011485 (*S. cerevisiae* L30e).

3. Results and discussion

3.1. The thermal and chemical stability of Ap L7Ae is extraordinary

Thermal denaturation studies were performed on the Ap and Mj L7Ae proteins and the mouse 15.5kD protein. The human and mouse 15.5kD homologs are virtually identical; the mouse 15.5kD that we have used in these circular-dichroism studies has an arginine at sequence position 9 (NCBI accession No. BAB23329), whereas the human 15.5kD homolog which was used in previous structural studies has a lysine at this position (NCBI accession No. NP_004999) (Vidovic *et al.*, 2000). Circular-dichroism thermal denaturation results are shown in Fig. 1. Mouse 15.5kD had the lowest melting temperature (325.5 ± 0.34 K), followed by *M. jannaschii* L7Ae (365.4 ± 0.50 K) and Ap L7Ae (>383 K under the conditions used for these experiments). Please note that in Fig. 1 the L7Ae protein is not fully denatured at 383 K but is shown for reference to the mouse 15.5kD and Mj L7Ae proteins. Based on CD melting experiments performed in sodium perchlorate or guanidine hydrochloride (Gdn-HCl), we conclude that the apparent melting temperature of Ap L7Ae is between 393 and 413 K (see below). To our knowledge, this is one of the highest apparent melting temperatures recorded for a single-domain monomeric protein. It is quite remarkable that such similar proteins have such markedly different melting temperatures. Mouse 15.5kD has 44 identical residues (33.3%) and 22 similar residues (16.7%) compared with Ap L7Ae, and their structures are practically identical (~1.1 Å r.m.s.d.). Likewise, Mj L7Ae has 71 identical residues (56% identity) and 16 similar residues (12.5%) when compared with Ap L7Ae and the structures have a 0.92 Å r.m.s.d. Yet, the Ap

L7Ae melting temperature differs by ~ 88 and ~ 48 K from those of 15.5kD and Mj L7Ae, respectively. Clearly, factors other than primary sequence and tertiary structure contribute to the marked thermal stability of Ap L7Ae.

In an effort to further examine the stability of Ap L7Ae, we performed more aggressive thermal denaturation experiments by varying the pH and the ionic strength and by adding chaotropic denaturants. We thought that the large number of charged residues in Ap L7Ae ostensibly contributed to its thermal stability through ion pairs and presumed that changing the pH of the solution would affect the melting temperature. Samples of Ap L7Ae were prepared in 20 mM sodium phosphate at the desired pH (from pH 1 to 12) and were incubated for 15 min at room temperature; a CD wavelength spectrum was acquired (at 298 K) and they were then thermally denatured while observing the ellipticity at 222 nm. Fig. 2(a) shows a typical wavelength spectrum of Ap L7Ae under the standard conditions (20 mM Na_2HPO_4 pH 7.5 at 298 K). The spectrum has characteristic bands indicative of α -helices and β -strands, and is similar to those observed for Mj L7Ae and mouse 15.5kD. Extremes of low and high pH were found to markedly destabilize Ap L7Ae (Fig. 2b); the protein readily denatures as a function of temperature at highly acidic pH values (e.g. 1, 2 and 3) and highly basic pH values (e.g. 11 and 12) (Table 3). However, the melting temperatures were over 383 K at pH values of 4–10. These results suggest that when the acidic amino acids are protonated or basic amino acids are deprotonated the structural stability of the protein is compromised. Aspartate and glutamate side chains have pK_a values of around 4.0 (3.9 and 4.1, respectively) and lysine and arginine have pK_a values of above 10 (10.5 and 12.5, respectively). At pH values below the pK_a values for the acidic residues or above the pK_a values for the basic residues the ion pairs would be disrupted and the protein destabilized. The role of ion pairs in stabilizing the Ap L7Ae structure will be further discussed below.

We also desired to evaluate the effects of electrostatic screening agents, such as high concentrations of salt and chaotropic denaturants, on the melting temperature of the protein. Initial melting experiments were performed with varying concentrations of NaCl; however, it was not possible to thermally denature Ap L7Ae protein in NaCl even at concentrations up to 5 M. We then sought a salt that could be used at higher concentrations and performed experiments with NaClO_4 . These experiments were successful and Ap L7Ae could readily be melted in 5 M NaClO_4 . We performed NaClO_4 concentration-dependent wavelength CD scans (Fig. 2c) and Ap L7Ae began to denature at room temperature at NaClO_4 concentrations of above 7 M. Sodium perchlorate concentration-dependent melting experiments were used to extrapolate the melting temperature under native conditions (Fig. 2e and Table 3). Based on these experiments, the native melting temperature for Ap L7Ae is 420 ± 2 K.

Gdn-HCl concentration-dependent CD wavelength scans were used to generate an isothermal (298 K) equilibrium unfolding curve (Fig. 2d). The midpoint of the Gdn-HCl-induced transition is 6 ± 3 M and the $\Delta G_{\text{d,H}_2\text{O}}$ for denaturation is $12 \pm 3 \text{ kcal mol}^{-1}$ ($49 \pm 13 \text{ kJ mol}^{-1}$) with an m value of $2 \pm 0.5 \text{ kcal mol}^{-1} \text{ M}^{-1}$ ($7 \pm 2 \text{ kJ mol}^{-1} \text{ M}^{-1}$). Further experiments examined thermal denaturation at varying concentrations of Gdn-HCl (Fig. 2e).

The Ap L7Ae protein could also be thermally denatured at Gdn-HCl concentrations above 2 M. The extrapolated apparent melting temperature at zero concentration of denaturant was 409 ± 2 K for Gdn-HCl. This is 11 K lower than the native melting temperature determined with NaClO_4 , but according to the Hofmeister series Gdn-HCl is a much more potent chaotrope than sodium perchlorate. We estimate that the native melting temperature of Ap L7Ae is

approximately 415 K based on both the NaClO_4 and Gdn-HCl studies. It is notable to mention that even under conditions containing 6 M Gdn-HCl the melting temperature of Ap L7Ae was 327 K, which is still higher than that of mouse 15.5kD (325 K) under our standard buffer conditions (20 mM Na_2HPO_4 pH 7.5). Collectively, these results suggest that Ap L7Ae is stabilized by electrostatic interactions, as the melting temperature decreases concomitant with increasing concentrations of ionic chaotropes. However, other factors such as amino-acid composition, secondary-structural elements, hydrophobic core and packing density could be involved, because the melting temperature is higher than the mesophilic mouse 15.5kD homolog even in 6 M Gdn-HCl, which should abrogate ionic interactions and disrupt the solvent-imposed hydrophobic effect.

3.2. Structure determination and quality of the model

Ap L7Ae was crystallized using conditions similar to those for Mj L7Ae (Suryadi *et al.*, 2005); however, zinc salts were required for optimal crystal growth. Diffraction data extending to 1.54 Å resolution were collected on our home source from a single crystal (Table 1) and the structure was solved by the molecular-replacement method using a model of the Ap L7Ae protein based on the Mj L7Ae structure (Suryadi *et al.*, 2005). The structure was refined using *CNS* (Brünger *et al.*, 1998) and *REFMAC5* (Murshudov *et al.*, 2011) to a final resolution of 1.56 Å with excellent crystallographic statistics and stereochemical parameters (Table 2).

The final refined model contains 124 of the 127 Ap L7Ae amino-acid residues. The N-terminal vector-encoded GSH residues remaining after thrombin cleavage and wild-type MSK residues did not have adequate density to permit placement, indicating that they are mobile or disordered in the crystal. The remainder of the structure was easily modeled into the density and encompassed the majority of the residues completely. The final model consists of 957 protein atoms and 119 waters. Analysis of the Ramachandran plot showed that 105 of 107 amino acids (98.1%; excluding glycine, proline

Figure 1
Thermal stability comparison of eukaryotic and archaeal L7Ae-family homologs. Circular-dichroism thermal denaturation of mouse 15.5kD (squares), Mj L7Ae (circles) and Ap L7Ae (triangles). Note the marked increase in the stability of Mj L7Ae compared with that of mouse 15.5kD ($\Delta T_m \approx 313$ K). Ap L7Ae is not fully (e.g. 100%) denatured under the conditions used for this experiment and by extrapolation we estimate that the T_m is at least 88 K greater than that of mouse 15.5kD and ~ 48 K greater than Mj L7Ae.

or terminal residues) were in the most favored regions of the plot and that three residues (2.8%) were in additionally allowed regions. The

average isotropic displacement parameter (B factor) was $21.35 \text{ \AA}^2$, with values of $18.69 \text{ \AA}^2$ for the peptide backbone, $21.01 \text{ \AA}^2$ for side

Figure 2

Chemical and thermal stability of the Ap L7Ae protein evaluated by circular-dichroism spectroscopy. (a) Wavelength spectrum (200–300 nm) of the Ap L7Ae protein at 298 K. (b) Thermal denaturation as a function of pH. The Ap L7Ae protein was thermally stable in the pH 4–10 range but was readily denatured at pH 1, 2, 3, 11 and 12. (c, d) Ellipticity at 222 nm as a function of sodium perchlorate concentration (c) or guanidine hydrochloride concentration (d) at 298 K. (e) Melting temperatures as a function of denaturant concentrations. Ap L7Ae was thermally denatured in varying concentrations of sodium perchlorate (squares) or Gdn-HCl (circles). The extrapolated melting temperatures for zero concentrations of denaturant were $420.5 \pm 2.3 \text{ K}$ for sodium perchlorate and $409.5 \pm 1.8 \text{ K}$ for Gdn-HCl.

chains and $33.38 \text{ \AA}^2$ for solvent atoms. The mean anisotropic displacement of the protein atoms was $0.803 \pm 0.088 \text{ \AA}$ and that for the solvent atoms was $0.628 \pm 0.126 \text{ \AA}$.

3.3. Structure description

The Ap L7Ae tertiary structure has the familiar α - β - α sandwich fold characteristic of this family of proteins, with alternating α -helices and β -sheets that form the globular domain (Fig. 3). The α -helices are predominately on the exterior of the domain and surround the central β -sheet formed by three parallel ($\beta 1$ - $\beta 3$) and one antiparallel ($\beta 4$) β -strands. The hydrophobic core is comprised of $\beta 2$ - $\beta 4$, while the surrounding α -helices are amphipathic. As observed in other structures, the N-terminal loop 1 organizes a three-residue aromatic trefoil with Phe9, Tyr77 and Tyr79 and creates a hydrophobic barrier on this side of the protein. The other side of the protein is sealed by $\beta 1$, which is formed predominately by lysine and arginine residues (Fig. 3). The loop formed by residues Ala90-Ala97 (loop 9) extends

Figure 3 Tertiary structure of Ap L7Ae. (a) Ribbon diagram of the structure. (b) Topological diagram of the Ap L7Ae structure showing the α - β - α motif and amino acids comprising the secondary structure. Colors correspond to identical regions in both diagrams.

Table 3 Thermal denaturation parameters for *A. pernix* L7Ae as a function of pH, NaClO₄ or Gdn-HCl.

1 cal = 4.186 J.

Agent	<i>T_m</i> (K)	ΔH (kcal mol ⁻¹)	ΔS (cal mol ⁻¹)	ΔG (kcal mol ⁻¹)
pH				
1	340 ± 0.2	64 ± 3	187 ± 8	13 ± 3
2	353 ± 0.2	72 ± 3	205 ± 8	17 ± 3
3	387 ± 31	82 ± 4	213 ± 21	24 ± 5
4–10	>383	—	—	—
11	375 ± 1	78 ± 4	208 ± 10	21 ± 4
NaClO ₄ concentration (M)				
0†	430 ± 2	105 ± 2	250 ± 4	37 ± 2
4	385 ± 4	97 ± 9	253 ± 4	28 ± 9
5	380 ± 4	84 ± 4	221 ± 11	24 ± 4
6	371 ± 1	95 ± 6	256 ± 16	25 ± 6
7	365 ± 1	91 ± 5	249 ± 14	23 ± 5
8	355 ± 1	90 ± 5	253 ± 13	21 ± 5
9	354 ± 1	89 ± 5	251 ± 14	20 ± 5
Gdn-HCl concentration (M)				
0†	409 ± 2	121 ± 8	295 ± 18	40 ± 8
2	383 ± 8	80 ± 3	208 ± 10	23 ± 3
3	368 ± 0.4	82 ± 3	224 ± 9	21 ± 3
4	354 ± 0.3	63 ± 2	179 ± 8	14 ± 3
5	343 ± 0.3	50 ± 2	146 ± 7	10 ± 2
6	327 ± 0.5	41 ± 3	126 ± 10	7 ± 3

† Extrapolated to zero concentration of denaturant.

outwards over $\alpha 4$ and forms a prominent region for interacting with K-turn RNA motifs (Fig. 3a). The archaeal L7Ae homologs appear to have the broadest RNA-target binding specificity, and residues in the loop 9 region are believed to be responsible for conferring RNA-target specificity to the L7Ae family of proteins (Hamma & Ferré-D'Amaré, 2004; Suryadi *et al.*, 2005). Several groups have shown that archaeal L7Ae homologs can interact with K-turn regions of box C/D and C'/D' motifs, box H/ACA RNAs and ribosomal RNAs. In contrast, the eukaryotic homolog 15.5kD/Snu13p has a more limited RNA-binding specificity; 15.5kD recognizes the box C/D region but cannot interact with box C'/D'. Loop 9 was found to be remarkably similar among the structures of archaeal L7Ae homologs (from *A. fulgidus*, *M. jannaschii*, *P. abyssi* and *H. marismortui*) and to differ in the eukaryotic 15.5kD, Snu13p and Nhp2p proteins. The eukaryotic 15.5kD/Snu13p homologs have a loop 9 consensus sequence of C-G-V-S-R-P-V-I, whereas the archaeal L7Aes have an A-G-L/I-X-V-X'-A-A sequence.

3.4. Structural comparison to homologous proteins

Among the L7Ae homologs, the sequence and structure of Ap L7Ae are very similar to those of Mj L7Ae and human 15.5kD, but the thermal stabilities of these three proteins are very different (Fig. 1). The primary sequences of several L7Ae homologs for which structural data are available are aligned in Fig. 4. The topologies of the L7Ae-family members contain no large deletions or insertions and overall the secondary-structure elements are similar. The major difference between the Ap and Mj L7Ae proteins is a small gap of three amino acids in the primary structure (Ala19-Glu21 in Ap L7Ae).

Although the tertiary structures of the Ap L7Ae, Mj L7Ae and 15.5kD proteins have remarkable similarities, they also show several differences. The Ap L7Ae structure can be superimposed with the Mj L7Ae and human 15.5kD structures with root-mean-square deviations (r.m.s.d.s) of 0.92 and $\sim 1.1 \text{ \AA}^2$, respectively. The major differences are the lengths of the N- and C-terminal α -helices. In Ap L7Ae, helices $\alpha 1$ and $\alpha 5$ are longer by one full turn (~ 5 residues) and 1.5 turns (seven residues), respectively, whereas helix $\alpha 2$ is one half-

Figure 4

Sequence and secondary-structural alignment of selected L7Ae-family members. The sequences for the determined structures of L7Ae, 15.5kD/Snu13p, L30 and Nhp2p homologs have been aligned and the secondary structures are indicated. No structure exists for human Nhp2p, but the sequence is shown for comparison with the other homologs. Residues in α -helices are represented by straight lines, 3_{10} -helices are indicated by dashed lines and β -sheets are specified by saw-toothed lines. Loops, turns and coils are indicated by blank spaces. Sequence-alignment conservation is indicated as follows: asterisks, identical residues; colons, strong conservation; periods, semi-conservation.

turn (two residues) shorter than that in Mj L7Ae. In Ap L7Ae helix $\alpha 5$ begins with a three-residue 3_{10} -helical region and then extends a further 18 residues, creating a 21-residue α -helical terminus to the protein. Other members of the L7Ae family for which structures have been determined typically have between 12 and 16 residues in the terminal α -helix. A similar 3_{10} -initiated and elongated C-terminal α -helix was also observed in the structure of *P. abyssi* L7Ae (Charron *et al.*, 2004).

3.5. Structural plasticity

Isotropic temperature factors (*B* factors) are a measure of the isotropic displacement of atoms by thermal motion, conformational disorder and static lattice disorder; they can also be indicators of flexible regions of a molecule (Petsko & Ringe, 1984). The absolute values of *B* factors are dependent on the resolution, the crystal temperature during data collection and the programs used to refine the structure, but relative differences can be compared if these factors are similar among the structures being evaluated. The data sets for Ap L7Ae and Mj L7Ae were both collected at 100 K and were refined using CNS (Brünger *et al.*, 1998) and the resolution ranges are similar (1.5 and 1.45 Å, respectively). The data set for human 15.5kD (bound to RNA) was collected at 100 K and refined using X-PLOR

(Brünger, 1993) and the resolution is somewhat lower (2.9 Å). Thus, we can make some qualitative comparisons of the *B* factors among the three structures. The average *B* factors for all protein atoms are 20 Å² for Ap L7Ae and 15 Å² for Mj L7Ae and human 15.5kD, and the values for the main-chain atoms are 19, 12 and 13 Å², respectively. A plot of the *B* factors versus main-chain atoms (*C*^α) of the three structures is given in Supplementary Fig. S1¹. The average main-chain *B* factors were analyzed in order to identify flexible regions. The *B* factors for loop L10 in Ap L7Ae and Mj L7Ae are 23 and 13 Å², respectively, which are greater than the overall average *B* factors (19 and 12 Å², respectively) and are likely to indicate a flexible region. In human 15.5kD, the loops L2, L7 and L10 appear to be flexible because their *B* factors (17 and 22 Å², respectively) are above the average *B* factor (13 Å²). The shortening of loops L2 and L10 (described above), the high flexibility of loop L10 and their location on the same side of the protein structure (Fig. 3a) suggest that they may bind RNA and/or protein partners during sRNP assembly. Large relative fluctuations can be observed in the 3_{10} -helix region (residues 59–62), where the *B* factors for *A. pernix* L7Ae (26 Å²) are greater than the average value (19 Å²). This is consistent with the prediction

¹ Supplementary material has been deposited in the IUCr electronic archive (Reference: BE5231).

Table 4Amino-acid compositions of *A. pernix*, *M. jannaschii* and *H. sapiens* L7Ae/15.5kD proteins.

Properties	<i>A. pernix</i>	<i>M. jannaschii</i>	<i>H. sapiens</i>
Growth temperature (K)	366	359	310
T_m (K)	>383	365	325
pI	6.6	5.5	8.7
Charged amino acids	45 (23+/22−)	37 (18+/19−)	31 (18+/13−)
Arg	8	1	6
Lys	14	16	10
His	1	1	2
Asp	5	4	5
Glu	17	15	8
Hydrophobic amino acids†	54	55	55
Hydrophilic amino acids‡	46	44	44
Neutral amino acids§	27	18	27
Nonpolar amino acids¶	72	68	65
Uncharged polar amino acids††	10	12	32
Charged amino acids – polar amino acids	35	25	−1
Thermally labile amino acids‡‡	4	9	21

† Hydrophobic amino acids: Ala, Cys, Phe, Ile, Leu, Met and Val. ‡ Hydrophilic amino acids: Asp, Glu, His, Lys, Asn, Gln and Arg. § Neutral amino acids: Gly, Pro, Ser, Thr, Trp and Tyr. ¶ Nonpolar amino acids: Ala, Phe, Gly, Ile, Leu, Met, Pro, Val, Trp and Tyr. †† Uncharged polar amino acids: Cys, Asn, Gln, Ser and Thr. ‡‡ Thermally labile amino acids: Cys, Met, Asn and Gln.

that the 3₁₀-helix region interacts with RNA and that residues Pro59, Glu61 and Ile62 of the 3₁₀-helix in Ap are highly conserved in the Mj L7Ae, Af L7Ae and human 15.5kD proteins (Fig. 4; Suryadi *et al.*, 2005; Vidovic *et al.*, 2000; Hamma & Ferré-D'Amaré, 2004; Moore *et al.*, 2004).

3.6. Ion-pair networks

Recent structural studies of hyperthermophilic proteins have revealed that increases in the number of ion pairs and the formation of ion-pair networks significantly contribute to the thermal stability (Karshikoff & Ladenstein, 2001; Lebbink *et al.*, 1999, 2002; Yip *et al.*, 1998; DeDecker *et al.*, 1996; Guy *et al.*, 2003; Hennig *et al.*, 1995; Lim *et al.*, 1997; Russell *et al.*, 1997; Pappenberger *et al.*, 1997; Vetriani *et al.*, 1998; Merz *et al.*, 1999; Bae & Phillips, 2005; Bezsudnova *et al.*, 2012; Sakuraba *et al.*, 2008; Yoneda *et al.*, 2012). The total numbers of charged residues in Ap L7Ae, Mj L7Ae and the human 15.5kD protein are 45 (35%), 37 (32%) and 31 (24%), respectively (Table 4). Not surprisingly, most of the charged residues are located on the protein surface. In Ap L7Ae, we found that a large number of lysine residues had been exchanged for arginines (Arg8, Arg31, Arg112 and Arg122) compared with Mj L7Ae (Fig. 4).

Using a cutoff distance of 6.0 Å, we identified 24 ion pairs in Ap L7Ae, 17 ion pairs in Mj L7Ae and only five ion pairs in human 15.5kD (Table 5, Fig. 5 and Supplementary Table S1). Mutagenic studies have suggested that ion-pair networks are energetically more favorable than the equivalent number of single ion-pair interactions (Horovitz *et al.*, 1990). An extensive number of interactions, 19 out of 24 ion pairs (79%), are involved in networks in Ap L7Ae. In contrast, only nine out of 17 ion pairs (53%) are involved in networks in Mj L7Ae and no ion-pair networks are observed in the human 15.5kD protein (Fig. 5c). In Ap L7Ae an extensive number of residues are involved in three-, four- and six-residue ion-pair networks (Table 5). Two three-residue networks (Glu44–Lys73–Glu72 and Glu17–Arg85–Glu88) and three four-residue networks (Arg8–Asp71–Arg122–Val127, Lys84–Asp56–Lys83–Asp58 and Lys33–Glu38–Arg41–Lys45) are observed in Ap L7Ae (Fig. 5a), whereas only one three-residue network (Lys6–Glu66–Lys116) and one four-residue network (Lys28–Glu33–Lys36–Arg40) are found in Mj L7Ae (Fig. 5b). In Ap L7Ae, a single four-residue network forms 15 ionic interactions (Arg8–

Table 5Ion pairs, hydrogen bonds and buried surface areas in the *A. pernix*, *M. jannaschii* and *H. sapiens* L7Ae/15.5kD proteins.

Properties	<i>A. pernix</i>	<i>M. jannaschii</i>	<i>H. sapiens</i>
No. of ion pairs			
Total	33	25	9
≤4 Å	9	8	3
4–6 Å	15	9	2
6–8 Å	9	8	4
Ion-pair interactions			
Total	101	64	21
≤4 Å	22	13	3
4–6 Å	38	19	8
6–8 Å	41	32	10
Ion-pair networks			
Total (≤6 Å)	6	3	0
Three-residue networks	2	1	0
Four-residue networks	3	1	0
Five-residue networks	0	1	0
Six-residue networks	1	0	0
Interactions in largest ion network (≤6 Å)	15	7	1
Hydrogen bonds†			
Total	113	106	113
Backbone–backbone	82	74	86
Backbone–side chain	18	17	19
Side chain–side chain	13	15	8
Surface atoms‡	508	488	529
Surface residues	38	33	46
Buried atoms	449	358	439
Buried residues	50	53	49
Total surface area (Å ²)	6500	5900	7000
Polar surface area (Å ²)	2500	2100	3000
Apolar surface area (Å ²)	4000	3800	4100
ΔSA total§ (Å ²)	15000	14400	14700
ΔSA polar (Å ²)	9700	8700	9700
ΔSA apolar (Å ²)	6100	6200	6100

† Calculated with HBPLUS (McDonald & Thornton, 1994). ‡ Calculated with GETAREA (Frackiewicz & Braun, 1998). § Calculated by subtracting the solvent-accessible surface area calculated for the folded protein from that of the tripeptide Ala-X-Ala, which is a model for the unfolded polypeptide (Hubbard *et al.*, 1991).

Asp71–Arg122–Val127), whereas the equivalent residues in Mj L7Ae form only three ion-pair interactions (Lys6–Glu66–Lys116). The largest ion-pair network in Ap L7Ae consists of six residues that form 15 ionic interactions (Glu21, Lys24, Arg25, Arg27, Glu28 and Asp106). The analogous network in Mj L7Ae has only seven interactions among five residues (Asp101–Lys23–Glu104–Lys111–Glu16). These observations suggest that a markedly increased number of ion pairs as well as the optimization of ion-pair clusters contribute most significantly to the thermostability of Ap L7Ae.

3.7. Hydrogen bonds

The Ap L7Ae structure contains approximately 169 hydrogen bonds between main-chain and side-chain atoms from 168 hydrogen-bond donor moieties and 184 acceptor groups as assigned using DSSP (Kabsch & Sander, 1983). The numbers of backbone-to-backbone hydrogen bonds that define the secondary-structure elements are similar in both the Ap L7Ae (82) and human 15.5kD (86) proteins, but there are fewer in Mj L7Ae (74) (Table 5). The major differences are in the number of hydrogen bonds contributed by side-chain atoms. The thermostable proteins Ap L7Ae and Mj L7Ae have more side chain-to-side chain hydrogen bonds (13 and 15, respectively) than does human 15.5kD (eight). In addition, most of the side-chain hydrogen bonds form ion pairs in Ap L7Ae and Mj L7Ae. Treatment of Ap L7Ae with Gdn–HCl disrupted its hydrogen bonds and the thermostability of the protein decreased to 327 K (Table 1). This result clearly indicates that hydrogen bonds contribute to the thermostability of Ap L7Ae. The same feature has been reported to enhance the compactness of the structure and to play a role in the

thermostability of *Bacillus pumilus* lipase (Akbulut *et al.*, 2013) and *Thermococcus sibiricus* alcohol dehydrogenase (Bezsudnova *et al.*, 2012). Comparative analyses of hydrogen bonds suggest that they may also play a minor role in the thermostability of the Mj L7Ae and human 15.5kD proteins. In Ap L7Ae, approximately 116 hydrogen-bond interactions are made with solvent molecules and the solvent

shell surrounding the protein is involved in at least 124 inter-molecular interactions. Collectively, there are about 3620 protein–protein contacts within 3.6 Å and about 292 protein–water contacts as analyzed using the CCP4 program CONTACT (Winn *et al.*, 2011).

4. Conclusions

A large number of proteins from hyperthermophilic organisms have been compared with their mesophilic homologs and the resulting observation is that most proteins use a combination of structural features to enhance thermostability. The primary feature that is found in most thermostable proteins is an increased number of ion-pair interactions (Karshikoff & Ladenstein, 2001; Lebink *et al.*, 1999, 2002; Yip *et al.*, 1998; DeDecker *et al.*, 1996; Guy *et al.*, 2003; Hennig *et al.*, 1995; Lim *et al.*, 1997; Russell *et al.*, 1997; Pappenberger *et al.*, 1997; Vetriani *et al.*, 1998; Merz *et al.*, 1999; Bae & Phillips, 2005; Bezsudnova *et al.*, 2012; Sakuraba *et al.*, 2008; Yoneda *et al.*, 2012). In this study, we have determined that Ap L7Ae employs a combination of features to enhance its thermostability. The amino-acid composition is delicately optimized for stability. There is a significant decrease in the number of thermolabile amino acids, an increase in the number of proline residues in optimal helical positions, a large number of charged surface residues and a marked exchange of lysine residues for arginines in Ap L7Ae. Superposition of the Ap L7Ae, Mj L7Ae and 15.5kD structures shows that the hydrophobic core is highly conserved but appears to be optimized in Ap L7Ae for maximum van der Waals interactions and packing. A striking number of charged surface residues in Ap L7Ae permit favorable electrostatic interactions and the formation of extensive ion-pair networks. The amino-acid composition of Ap L7Ae also contributes to protein folding and stability. Another stabilization feature is the shortening of two loop regions, which increases the rigidity among the connected secondary-structural elements. Other stabilization features are the number of hydrogen bonds, the burial of more hydrophobic atoms, a decrease in the solvent-accessible surface area and a slight increase in the packing density. The relative contributions of these stabilization features need to be assessed by the analyses of site-directed mutants using future biophysical and structural studies. This study has shown that a combination of sequence and structural features collectively contribute to the remarkable thermal stability of this highly conserved multifunctional protein.

We thank Professor E. Stuart Maxwell (North Carolina State University) and his laboratory for reading the manuscript and useful discussions. We also thank Professor Rebecca Alexander (Wake Forest University) for comments on the manuscript. We thank Theodore B. Gupton Jr, Sunnie Leigh Flowers and C. Leigh Allen for experimental assistance. This research was supported by grants from the Wake Forest University Science Research Fund, the American Cancer Society (IRG 093-035-09) and the National Institutes of Health (GM069699) to BAB and from the Wake Forest University Summer Undergraduate Research Fund to TBG, SLF and CLA.

References

- Adams, M. W. W. (1993). *Annu. Rev. Microbiol.* **47**, 627–658.
- Akbulut, N., Tuzlakoglu Öztürk, M., Pijning, T., İşsever Öztürk, S. & Gümüşel, F. (2013). *J. Biotechnol.* **164**, 123–129.
- Bachelier, J.-P., Cavaillé, J. & Hüttenhofer, A. (2002). *Biochimie*, **84**, 775–790.
- Bae, E. & Phillips, G. N. (2005). *J. Biol. Chem.* **280**, 30943–30948.
- Bezsudnova, E. Y., Boyko, K. M., Polyakov, K. M., Dorovatovskiy, P. V., Stekhanova, T. N., Gumerov, V. M., Ravin, N. V., Skryabin, K. G., Kovalchuk, M. V. & Popov, V. O. (2012). *Biochimie*, **94**, 2628–2638.

Figure 5 Schematic view of ion pairs in the Ap L7Ae (a), Mj L7Ae (b) and human 15.5kD (c) proteins. The wide lines, arrows and thin lines represent α -helices, β -strands and loops, respectively. The red and blue circles enclosing a numeral represent positively and negatively charged residues, respectively, at the indicated position. The red, blue and black dotted lines represent ion-pair networks within distances of 4, 6 and 8 Å, respectively.

- Bhuiya, M. W., Sakuraba, H., Ohshima, T., Imagawa, T., Katunuma, N. & Tsuge, H. (2005). *J. Mol. Biol.* **345**, 325–337.
- Biswas, S., Buhrman, G., Gagnon, K., Mattos, C., Brown, B. A. II & Maxwell, E. S. (2011). *Biochemistry*, **50**, 2907–2918.
- Bruins, M. E., Janssen, A. E. & Boom, R. M. (2001). *Appl. Biochem. Biotechnol.* **90**, 155–186.
- Brünger, A. T. (1993). *X-PLOR Version 3.1*. New Haven: Yale University Press.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst. D* **54**, 905–921.
- Burley, S. K. & Petsko, G. A. (1985). *Science*, **229**, 23–28.
- Charron, C., Manival, X., Cléry, A., Senty-Ségault, V., Charpentier, B., Marmier-Gourrier, N., Branlant, C. & Aubry, A. (2004). *J. Mol. Biol.* **342**, 757–773.
- Decatur, W. A. & Fournier, M. J. (2003). *J. Biol. Chem.* **278**, 695–698.
- DeDecker, B. S., O'Brien, R., Fleming, P. J., Geiger, J. H., Jackson, S. P. & Sigler, P. B. (1996). *J. Mol. Biol.* **264**, 1072–1084.
- Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P. & Matthews, B. W. (1992). *Science*, **255**, 178–183.
- Espósito, L., Sica, F., Raia, C. A., Giordano, A., Rossi, M., Mazzarella, L. & Zagari, A. (2002). *J. Mol. Biol.* **318**, 463–477.
- Fersht, A. R. & Serrano, L. (1993). *Curr. Opin. Struct. Biol.* **3**, 75–83.
- Fraczkiewicz, R. & Braun, W. (1998). *J. Comput. Chem.* **19**, 319–333.
- Gaspin, C., Cavallé, J., Erauso, G. & Bachellerie, J.-P. (2000). *J. Mol. Biol.* **297**, 895–906.
- Guy, J. E., Isupov, M. N. & Littlechild, J. A. (2003). *J. Mol. Biol.* **331**, 1041–1051.
- Hamma, T. & Ferré-D'Amaré, A. R. (2004). *Structure*, **12**, 893–903.
- Hennig, M., Darimont, B., Sterner, R., Kirschner, K. & Jansonius, J. N. (1995). *Structure*, **3**, 1295–1306.
- Hennig, M., Sterner, R., Kirschner, K. & Jansonius, J. N. (1997). *Biochemistry*, **36**, 6009–6016.
- Herning, T., Yutani, K., Inaka, K., Kuroki, R., Matsushima, M. & Kikuchi, M. (1992). *Biochemistry*, **31**, 7077–7085.
- Horowitz, A., Serrano, L., Avron, B., Bycroft, M. & Fersht, A. R. (1990). *J. Mol. Biol.* **216**, 1031–1044.
- Hubbard, S. J., Campbell, S. F. & Thornton, J. M. (1991). *J. Mol. Biol.* **220**, 507–530.
- Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
- Karshikoff, A. & Ladenstein, R. (2001). *Trends Biochem. Sci.* **26**, 550–556.
- Kim, T., Joo, J. C. & Yoo, Y. J. (2012). *J. Biotechnol.* **161**, 49–59.
- Kiss, T. (2001). *EMBO J.* **20**, 3617–3622.
- Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst. D* **55**, 484–491.
- Knapp, S., Kardinahl, S., Hellgren, N., Tibbelin, G., Schäfer, G. & Ladenstein, R. (1999). *J. Mol. Biol.* **285**, 689–702.
- Kuhn, J. F., Tran, E. J. & Maxwell, E. S. (2002). *Nucleic Acids Res.* **30**, 931–941.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- Lebbink, J. H., Consalvi, V., Chiaraluce, R., Berndt, K. D. & Ladenstein, R. (2002). *Biochemistry*, **41**, 15524–15535.
- Lebbink, J. H., Knapp, S., van der Oost, J., Rice, D., Ladenstein, R. & de Vos, W. M. (1999). *J. Mol. Biol.* **289**, 357–369.
- Lim, J.-H., Yu, Y. G., Han, Y. S., Cho, S., Ahn, B.-Y., Kim, S.-H. & Cho, Y. (1997). *J. Mol. Biol.* **270**, 259–274.
- Marky, L. A. & Breslauer, K. J. (1987). *Biopolymers*, **26**, 1601–1620.
- Matthews, B. W., Nicholson, H. & Becktel, W. J. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 6663–6667.
- McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.
- McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
- Merz, A., Knöchel, T., Jansonius, J. N. & Kirschner, K. (1999). *J. Mol. Biol.* **288**, 753–763.
- Moore, T., Zhang, Y., Fenley, M. O. & Li, H. (2004). *Structure*, **12**, 807–818.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst. D* **67**, 355–367.
- Omer, A. D., Lowe, T. M., Russell, A. G., Ebhardt, H., Eddy, S. R. & Dennis, P. P. (2000). *Science*, **288**, 517–522.
- Omer, A. D., Ziesche, S., Ebhardt, H. & Dennis, P. P. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 5289–5294.
- Pace, C. N. (1986). *Methods Enzymol.* **131**, 266–280.
- Pappenberger, G., Schurig, H. & Jaenicke, R. (1997). *J. Mol. Biol.* **274**, 676–683.
- Petsko, G. A. & Ringe, D. (1984). *Annu. Rev. Biophys. Bioeng.* **13**, 331–371.
- Puglisi, J. D. & Tinoco, I. (1989). *Methods Enzymol.* **180**, 304–325.
- Rothschild, L. J. & Mancinelli, R. L. (2001). *Nature (London)*, **409**, 1092–1101.
- Russell, R. J., Ferguson, J. M., Hough, D. W., Danson, M. J. & Taylor, G. L. (1997). *Biochemistry*, **36**, 9983–9994.
- Sakuraba, H., Yoneda, K., Asai, I., Tsuge, H., Katunuma, N. & Ohshima, T. (2008). *Biochim. Biophys. Acta*, **1784**, 563–571.
- Serrano, L. & Fersht, A. R. (1989). *Nature (London)*, **342**, 296–299.
- Speckmann, W. A., Li, Z.-H., Lowe, T. M., Eddy, S. R., Terns, R. M. & Terns, M. P. (2002). *Curr. Biol.* **12**, 199–203.
- Suryadi, J., Tran, E. J., Maxwell, E. S. & Brown, B. A. II (2005). *Biochemistry*, **44**, 9657–9672.
- Szilágyi, A. & Závodszky, P. (2000). *Structure Fold. Des.* **8**, 493–504.
- Takano, K., Ogasahara, K., Kaneda, H., Yamagata, Y., Fujii, S., Kanaya, E., Kikuchi, M., Oobatake, M. & Yutani, K. (1995). *J. Mol. Biol.* **254**, 62–76.
- Tanaka, T., Sawano, M., Ogasahara, K., Sakaguchi, Y., Bagautdinov, B., Katoh, E., Kuroishi, C., Shinkai, A., Yokoyama, S. & Yutani, K. (2006). *FEBS Lett.* **580**, 4224–4230.
- Tang, T.-H., Bachellerie, J.-P., Rozhdestvensky, T., Bortolin, M.-L., Huber, H., Drungowski, M., Elge, T., Brosius, J. & Hüttenhofer, A. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 7536–7541.
- Terns, M. P. & Terns, R. M. (2002). *Gene Expr.* **10**, 17–39.
- Tollervay, D. (1996). *Science*, **273**, 1056–1057.
- Tran, E., Brown, J. & Maxwell, E. S. (2004). *Trends Biochem. Sci.* **29**, 343–350.
- Tran, E. J., Zhang, X. & Maxwell, E. S. (2003). *EMBO J.* **22**, 3930–3940.
- Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst. D* **55**, 191–205.
- Van Boxstael, S., Cunin, R., Khan, S. & Maes, D. (2003). *J. Mol. Biol.* **326**, 203–216.
- Vetriani, C., Maeder, D. L., Tolliday, N., Yip, K. S.-P., Stillman, T. J., Britton, K. L., Rice, D. W., Klump, H. H. & Robb, F. T. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 12300–12305.
- Vidovic, I., Nottrott, S., Hartmuth, K., Lührmann, R. & Ficner, R. (2000). *Mol. Cell*, **6**, 1331–1342.
- Vieille, C. & Zeikus, G. J. (2001). *Microbiol. Mol. Biol. Rev.* **65**, 1–43.
- Winn, M. D. *et al.* (2011). *Acta Cryst. D* **67**, 235–242.
- Wong, K.-B., Wang, W. K., Proctor, M. R., Bycroft, M. & Chen, Y. W. (2001). *Acta Cryst. D* **57**, 865–866.
- Xiang, T., Liu, Q., Deacon, A. M., Koshy, M., Kriksunov, I. A., Lei, X. G., Hao, Q. & Thiel, D. J. (2004). *J. Mol. Biol.* **339**, 437–445.
- Yip, K. S.-P., Britton, K. L., Stillman, T. J., Lebbink, J., de Vos, W. M., Robb, F. T., Vetriani, C., Maeder, D. & Rice, D. W. (1998). *Eur. J. Biochem.* **255**, 336–346.
- Yoneda, K., Sakuraba, H., Araki, T. & Ohshima, T. (2012). *J. Biol. Chem.* **287**, 12966–12974.
- Zhang, X., Meining, W., Fischer, M., Bacher, A. & Ladenstein, R. (2001). *J. Mol. Biol.* **306**, 1099–1114.