Crystal Structure of Ribosomal Protein L30e from the Extreme Thermophile *Thermococcus celer*: Thermal Stability and RNA Binding[†,‡]

Yu Wai Chen,[§] Mark Bycroft,[§] and Kam-Bo Wong*[,||]

Cambridge University Chemical Laboratory and Centre for Protein Engineering, Medical Research Council Centre, Hills Road, Cambridge CB2 2QH, U.K., and Department of Biochemistry, The Chinese University of Hong Kong, Shatin, Hong Kong

Received November 7, 2002; Revised Manuscript Received January 14, 2003

ABSTRACT: We report here the high-resolution crystal structure of the ribosomal protein L30e from the hyperthermophilic archaeon *Thermococcus celer* determined at cryo-temperature. When it is compared with its mesophilic homologue, L30e from yeast, a number of structural features that can enhance thermostability are revealed. Disordered residues corresponding to a large RNA-binding loop in yeast L30e are well structured in the *T. celer* protein. The overall charge of *T. celer* L30e is near neutral, whereas that of the yeast homologue is highly positive. This is the result of an increase in the number of acidic residues at the expense of polar residues, Asn, Ser, and Thr. Extensive ion pair networks are found on the molecular surface. Exposed nonpolar surface areas are reduced in the *T. celer* protein. Its side chain atoms preferably form hydrogen bonds with main chain atoms. Taken together, these factors contribute to high protein stability. The roles of well-conserved L30e residues are studied and found to be important in defining a very compact overall structure and in maintaining the structure of the RNA binding site. By comparing it with the yeast homologue, we also identified the residues that are responsible for RNA binding and built a model to illustrate how L30e binds to an RNA kink turn motif.

The survival of extreme thermophilic organisms at elevated temperatures requires, among many factors, that their body contents be heat resistant. The proteins of these organisms must be able to withstand temperatures that are close to or even higher than the boiling point of water and remain functional. There are tremendous interests in delineating the structural determinants underlying thermostability. Apart from offering a deeper understanding of protein evolution, this knowledge holds great promise in designing and engineering enzymes to be more cost-effective for industrial use.

The archaebacterium *Thermococcus celer* grows optimally at 85 °C. We have cloned the gene encoding a vital protein of the ribosome machinery from *T. celer* and purified the protein to homogeneity (*1*). L30e (formerly known as the L32; *2*, *3*) is a universal component of the large subunit of the eukaryotic and archaeal ribosomes. In yeast, L30e is essential for growth (*2*) and its synthesis is doubly autoregulated (*4*). On overproduction, yeast L30e binds to its own pre-mRNA and disrupts splicing. It can also bind to the start codon of its mature mRNA and inhibits translation. Autoregulation of L30e is not conserved in some other species (*5*, *6*). L30e is considered to play an important role in ribosome structure and function (*7*). One report suggested that L30e is involved in bridging the large and small ribosomal subunits (*8*).

L30e from *T. celer* is a highly stable protein having a melting temperature of 95 °C (K. B. Wong, unpublished data). As a first step toward understanding the structural basis of thermostability, we undertook the determination of its structure by X-ray crystallography. The solution structure of a mesophilic homologue, the yeast L30e protein, in both free and mRNA-bound forms, has been reported (*7*, *9*). By comparing the thermophilic and mesophilic structures, we can identify important structural principles contributing to thermostability. We also studied the RNA-binding motif of this protein and proposed a model for its interaction with the *T. celer* ribosomal RNA.

EXPERIMENTAL PROCEDURES

Protein Overproduction, Purification, Crystallization, and Data Collection. These procedures have been published previously (*1*).

Structure Solution and Refinement. The structure was determined by molecular replacement, and calculations were performed with AMoRe (*10*, *11*). In the Protein Data Bank (PDB),[1] there are two NMR structures of L30e from yeast (free and complexed with mRNA, PDB entries 1ck9 and 1cn9, respectively) that can serve as search models. Unfor-

[†] Supported by Grant CUHK4243/00M from the Research Grants Council of Hong Kong and a grant (project code 2030253) from the Research Committee of The Chinese University of Hong Kong. K.-B.W. acknowledges summer research grants from The Chinese University of Hong Kong. Y.W.C. is supported by Wellcome Trust Grant 061836.

[‡] The atomic coordinates and structure factors have been deposited in the Protein Data Bank as entry 1h7m.

* To whom correspondence should be addressed: Department of Biochemistry, The Chinese University of Hong Kong, Shatin, Hong Kong. Telephone: +852-2609-8024. Fax: +852-2603-5123. E-mail: kbwong@cuhk.edu.hk.

[§] Cambridge University Chemical Laboratory and Centre for Protein Engineering.

[||] The Chinese University of Hong Kong.

[1] Abbreviations: PDB, Protein Data Bank; NMR, nuclear magnetic resonance; rms, root-mean-square; KT, kink turn.

10.1021/bi027131s CCC: $25.00
Published on Web 02/19/2003

tunately, we were unable to find a solution with these models. In retrospect, this may be due to the relatively large overall differences (see later) in the atomic coordinates of these models when compared to those of the target structure. Another structural homologue is the L7Ae protein component of the 2.4 Å crystal structure of the large ribosomal subunit of *Haloarcula marismortui* (*12*). At the time this work was undertaken, the L7Ae model deposited in the PDB contained only Cα atoms (PDB entry 1ffk, chain E). An all-atom model was built with the aid of the EMBL-EBI MaxSprout server (http://www.ebi.ac.uk/dali/maxsprout/; *13*) and was used as a search model. Not surprisingly, this crude model did not reveal a molecular replacement solution.

The structure reported here was determined when a parallel NMR study of this protein yielded a good-quality intermediate model (K. B. Wong, unpublished data) consisting of 25 conformers. A search procedure similar to that recommended by Chen et al. (*14*) was adopted. Ensemble search models were employed in rotation searches, and all peaks were recycled into the subsequent translation search in the two enantiomorphic space groups $P6_1$ and $P6_5$. In both rotation and translation searches, data in the range of 15−3.5 Å were used. Translation function peaks were pooled and sorted according to the correlation coefficient, and then subjected to rigid-body refinement, using data in the range of 10−3.5 Å. The search model was an ensemble with intact hydrophobic core residues, while long surface side chains are truncated to Cβ. With this model, a solution was obtained in space group $P6_1$ with the highest correlation coefficient of 50.7 that can be clearly distinguished from the highest noise peak having a correlation coefficient of 43.7 (*15*). This top peak also corresponds to the lowest *R* factor of 0.463, compared to an *R* factor of 0.495 for the closest noise peak. The best peak in $P6_5$ has a correlation coefficient of 43.7 and an *R* factor of 0.509 and is hardly distinguishable from other peaks. The molecular replacement solution has no spatial overlap with its symmetry-related neighbors when examined in a molecular graphics program.

The NMR conformer with the least rms deviation from the mean structure was used for refinement. This model was subjected to alternating cycles of computational refinement and manual rebuilding. Throughout the course of refinement, 7.5% of the data was left out for cross validation and a bulk solvent model was used. In the early refinement cycles, we employed torsional molecular dynamics (*16*) implemented in the program *CNS* (*17*) using the maximum likelihood target function. Only overall and grouped *B* factors (one for main chain and one for side chain atoms in each residue) were refined at this stage. Manual rebuilding was performed with the program *O* (*18*) guided by σ_A-weighted maps with $2mF_o - DF_c$ and $mF_o - DF_c$ coefficients (*19*), as well as maps generated by a density modification protocol (*20*). All maps were calculated with *CNS*. After seven rounds, the refinement essentially converged, and *Refmac* (*21*) was employed for further computational refinement. Alternate conformations were introduced to explain the side chain electron density of three residues; solvent molecules were revised, and individual atomic *B* factors were refined. The quality of the final model was analyzed with *PROCHECK* (*22*) and found to be excellent. The final model has an *R* factor of 0.17 and an R_{free} of 0.23, for all data to 1.96 Å, and the refinement statistics are listed in Table 1.

Table 1: Crystallographic Data and Refinement Statistics for *T. celer* L30e

Crystal Information	
space group	$P6_1$
unit cell dimensions (Å)	
$a = b$	48.32
c	86.42
Refinement Statistics	
resolution range (Å)	24.2−1.96
R_{cryst}[a]	0.170
R_{free}[b]	0.225
model	
no. of protein atoms	734
no. of water atoms	120
average *B* factors (Å^2)	
all atoms	21.1
protein main chain	18.3
protein side chain	21.7
solvent molecules	32.1
rms deviation from ideality	
bonds (Å)	0.028
bond angles (deg)	2.22
coordinate error (Å)	
estimated from R_{free}	0.15
estimated from maximum likelihood	0.11
geometry, Ramachandran plot analysis[c]	
most favored region (%)	91.7
additional allowed region (%)	8.3
generously allowed and disallowed regions (%)	0.0

[a] $R_{cryst} = \sum||F_o| - |F_c||/\sum|F_o|$, where F_o and F_c are the observed and calculated amplitudes, respectively. [b] The free set contains 618 (7.5% of total) reflections. [c] Results of *PROCHECK* for 84 non-glycine, non-proline residues.

Structure Analysis. We used the criteria that are based on the work of Barker and Hubbard (*23*) to identify hydrogen bonds. In addition, we have included Cα-H groups as hydrogen bond donors and aromatic rings as acceptors. Hydrogen atoms of *T. celer* L30e were generated by the program *REDUCE* (*24*). Hydrogen bond analysis was performed with the program *HBPLUS* (*25*). The minimized average model of the yeast L30e NMR structure (PDB entry 1ck2) was used for comparison. For this purpose, disordered loop residues and poorly defined long surface side chains were excluded from the study to ensure that only the most reliable and best defined hydrogen bonds were counted.

Solvent accessibilities were calculated with *NACCESS* (*26*), with a probe radius of 1.4 Å. Disordered residues from both termini were excluded from analysis, for both the *T. celer* structure and the yeast average model. Graphical representation of electrostatic potential on the molecular surface was calculated with *GRASP* (*27*). Ion pairs were inspected manually. Internal cavities were analyzed with *VOIDOO* (*28*) with a probe radius of 1.2 Å and a minimum volume of 10 Å^3 as the threshold.

Modeling of the L30e−rRNA Complex. The *T. celer* rRNA sequence was obtained from the European Large Subunit Ribosomal RNA Database (*29*). The RNA kink turn model was built with the graphics program *O*. The putative *T. celer* rRNA sequence that is predicted to bind L30e (see later) is highly similar with the sequence of kink turn 7 of the *H. marismortui* large ribosomal subunit (PDB entry 1jj2; *30*). A model was built with the phosphate backbone preserved and correct nucleotide bases introduced. This rRNA model was docked onto the *T. celer* L30e protein by referencing the interaction between the spliceosomal 15.5 kDa protein

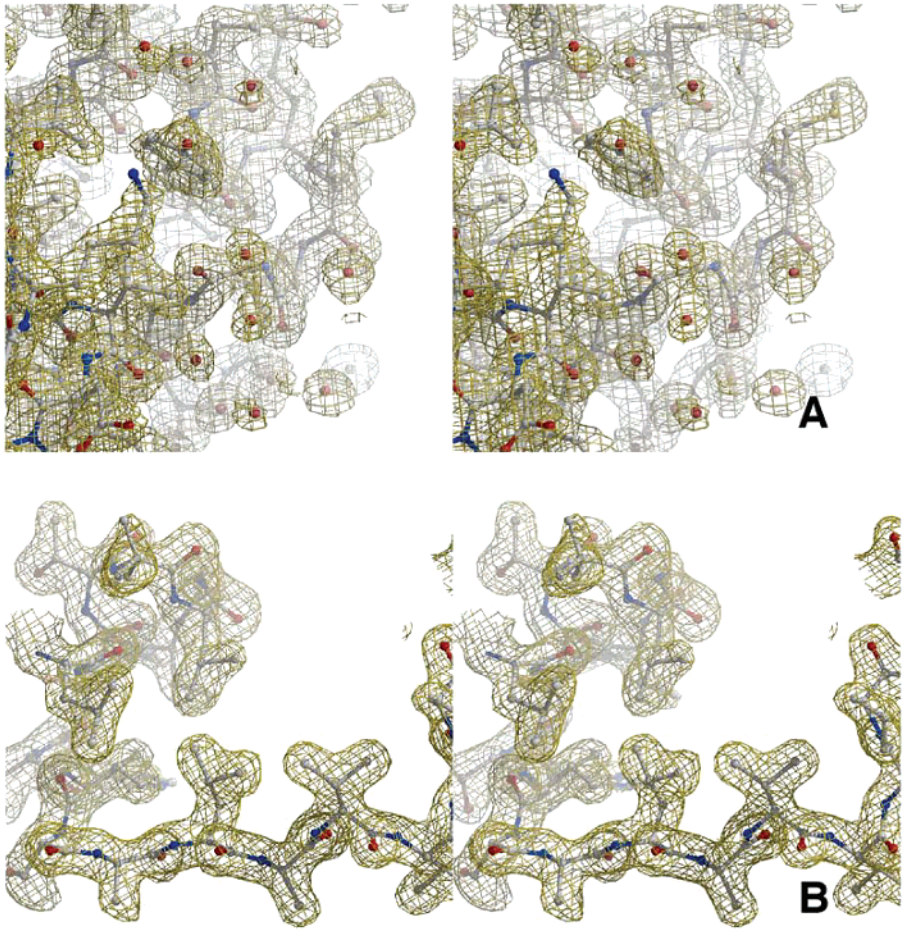

FIGURE 1: Typical electron density calculated with the final model. (A) Atomic model of the loop connecting α2 and *β*2. (B) Residues 72−88 of *T. celer* L30e, a region corresponding to a disordered loop in yeast L30e. Both panels are shown with a $2mF_o - DF_c$ electron density map contoured at 1.2*σ*. This figure was generated with *bobscript* (*47*) and rendered with *Raster3d* (*48*).

and the U4 snRNA (PDB entry 1e7k). The binary model was then subjected to three runs (200 cycles each) of energy minimization performed with *CNS*. The first run employed large harmonic restraints (harmonic restraint constant, $k_{harm} = 10$) imposed on protein atoms that are not involved in RNA binding. During the second run, atoms in loops close to the RNA-binding site were also allowed to move unrestrained while medium harmonic restraints ($k_{harm} = 4$) were applied to the remaining non-RNA-binding atoms. The last run used weak harmonic restraints ($k_{harm} = 2$) on all atoms in the model. The final complex model has an rms deviation in bond lengths of 0.0023 Å and an rms deviation in bond angles of 0.58°. This model is available upon request from the authors.

RESULTS

Quality of the Structure Model

The protein sample used for crystallization contains at the N-terminus two plasmid-encoded residues ("Gly-Ser") which are numbered −2 and −1, respectively, according to the standard PDB scheme. The final model of the *T. celer* L30e crystal structure consists of 97 residues (−1 to 96). The first residue, Gly−2, and the last four C-terminal residues (97−100) have no interpretable electron densities and are excluded from the model. There is a break in main chain electron density between residues 76 and 77, this region (described later) being the worst defined of the whole model. Three arginine residues (21, 39, and 76), Pro77, and His78 have disordered side chain atoms. The side chain electron densities of Met18, Ser67, and Ser80 are best described by multiple conformations. Apart from that described above, the structure is well-defined with a mean *B* factor of 21 Å^2. Typical electron density calculated with the final model is shown in Figure 1A.

Table 2: Secondary Structure Definition of L30e

	start residue	end residue
helix 1 (α1)	Asp2	Gly14
strand 1 (*β*1)	Lys15	Met18
helix 2 (α2)	Gly19	Met29
strand 2 (*β*2)	Leu34	Ala38
helix 3 (α3)	Arg42	Ser56
strand 3 (*β*3)	Pro59	Phe63
helix 4 (α4)	Thr66	Leu74
strand 4 (*β*4)	Ser81	Val86
helix 5 (H5)	Arg92	Leu96

Overall Structure

The topology of *T. celer* L30e has an α/*β*/α sandwich arrangement (Figure 2A). The secondary structure consists of five helices [four α-helices (α1−α4) and a 3_{10}-helix (H5)] interspersed with four *β*-strands following a regular alternating pattern (Table 2). The four strands constitute a central mixed parallel/antiparallel *β*-sheet. Helices α2 and α3 pack

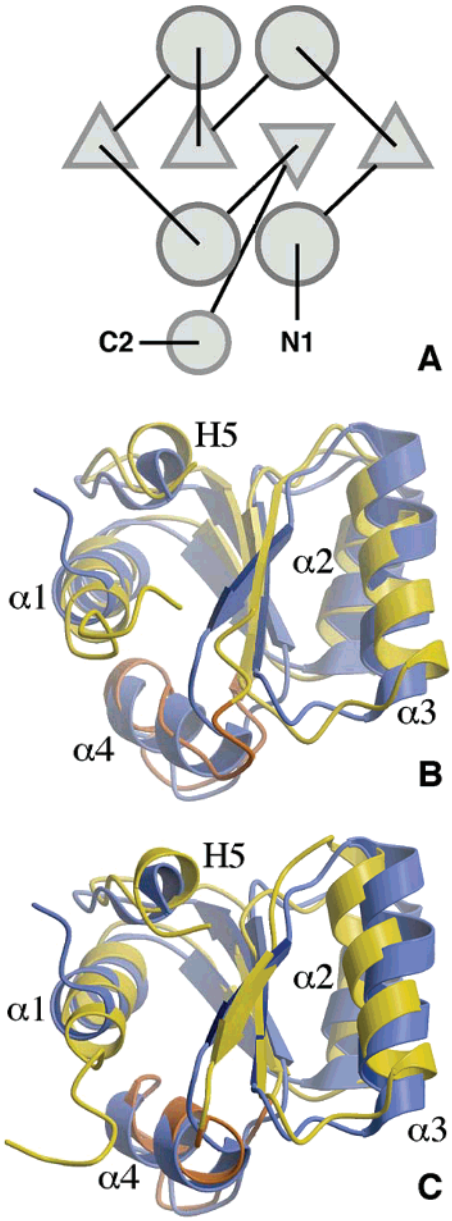

FIGURE 2: Overall structure of *T. celer* L30e. (A) Topology of *T. celer* L30e prepared with the protein topology server (http://www.sander.embl-ebi.ac.uk/tops/). Comparison of the overall structure of *T. celer* L30e (blue) with (B) free yeast L30e (yellow) and with (C) yeast L30e with bound pre-mRNA. Residues in a disordered loop in free yeast L30e are highlighted in orange. Panels B and C were generated with *bobscript* (*47*) and rendered with *Raster3d* (*48*).

on one face of the sheet and run parallel to the strands. Helices α1, α4, and H5 pack on another face and run at a right angle to the strands. There are six type I β-turns in the structure: at the C-terminus of α2 (residues 27−30), at the loop connecting β2 to α3 (residues 38−41), at both termini of α3 (residues 42−45 and 54−57), and a double β-turn at the C-terminus of α4 (residues 71−74 and 72−75). A type II′ β-turn is identified at residues 88−91, in the loop connecting β4 to H5. The overall fold is characteristic of the L30e family (Figure 2B,C).

Comparison with the T. celer L30e NMR Structure

First, it should be noted that the crystallography work was performed at cryo-temperature (100 K), whereas all NMR data were collected at near-room temperatures. When main chain atoms (Cα, C, N, and O) of residues 1−96 of the two structures are superimposed, the rms deviation in atomic positions is 0.82 Å, indicating that the two structures are essentially identical.

Comparison with Yeast L30e NMR Structures

Overall Structure. The overall fold and all the secondary structure elements of the *T. celer* and yeast L30e proteins are well-conserved (Figure 2B,C). The two structures have an overall rms deviation of 2.0 Å in Cα atomic positions when the well-defined residues are compared (residues 1−96 of the *T. celer* structure aligned with residues 8−103 of the yeast structure). The RNA-bound yeast L30e structure has a similar overall rms deviation of 2.1 Å when compared with the *T. celer* structure. However, the secondary structure definition of the *T. celer* structure clearly more resembles the yeast RNA-bound structure than the RNA-free structure. A noticeable feature of the RNA-free yeast L30e is the presence of a large disordered region (residues 72−88) which becomes structured on binding to its pre-mRNA (*7*; Figure 2B,C). In striking contrast, this segment is well-ordered in the *T. celer* L30e crystal (Figure 1B) and NMR structures, both of them do not contain RNA. *T. celer* L30e residues 65−80, corresponding to the disordered loop residues in the yeast RNA-free structure, form an α-helix (α4) as well as a trailing loop connecting α4 to strand β4 (Figures 1B and 2B,C). The peptide backbone conformation in this region is very similar to that of the yeast RNA-bound structure, with a local rms deviation in Cα atomic positions of 1.3 Å (Figure 2C). Only three residues in this region have atomic *B* factors exceeding 40 Å^2. These residues, Arg76, Pro77, and His78, are the worst modeled in this structure. Presumably, these residues are more dynamic in nature and are likely to play important roles in RNA binding, as implicated in the yeast L30e structures (see later). A second region where major structural differences occur includes residues 38−45 (residues 45−52 in yeast L30e), also close to the RNA-binding site. This is another region where large structural differences between the RNA-free and RNA-bound yeast L30e structures are observed. In the *T. celer* structure, this corresponds to a loop connecting β2 to α3 and is well-ordered. In the yeast RNA-free structure, this loop has a totally different conformation and is slightly disordered. Again, the backbone conformation of this region in the *T. celer* structure resembles that of the RNA-bound yeast L30e structure (Figure 2C), with the exception of two residues: Pro43 and Asp44 (which correspond to yeast Val50 and Leu51, respectively).

Amino Acid Composition. The comparison of amino acid contents is given in Table 3. It is noteworthy that *T. celer* has significantly fewer Asn, Leu, Lys, Ser, and Thr residues, while it has more Ala, Arg, Asp, and Glu residues. The *T. celer* protein has more charged residues at the expense of polar uncharged residues. As a result, the overall charge balance of the two proteins is very different. At physiological pH, yeast L30e has a net charge of +8 (seven negatively charged and 15 positively charged residues) whereas *T. celer* L30e has a net charge of +1 (13 negatively charged and 14 positively charged residues). The increase in the number of acidic residues renders the *T. celer* protein a near-neutral molecule.

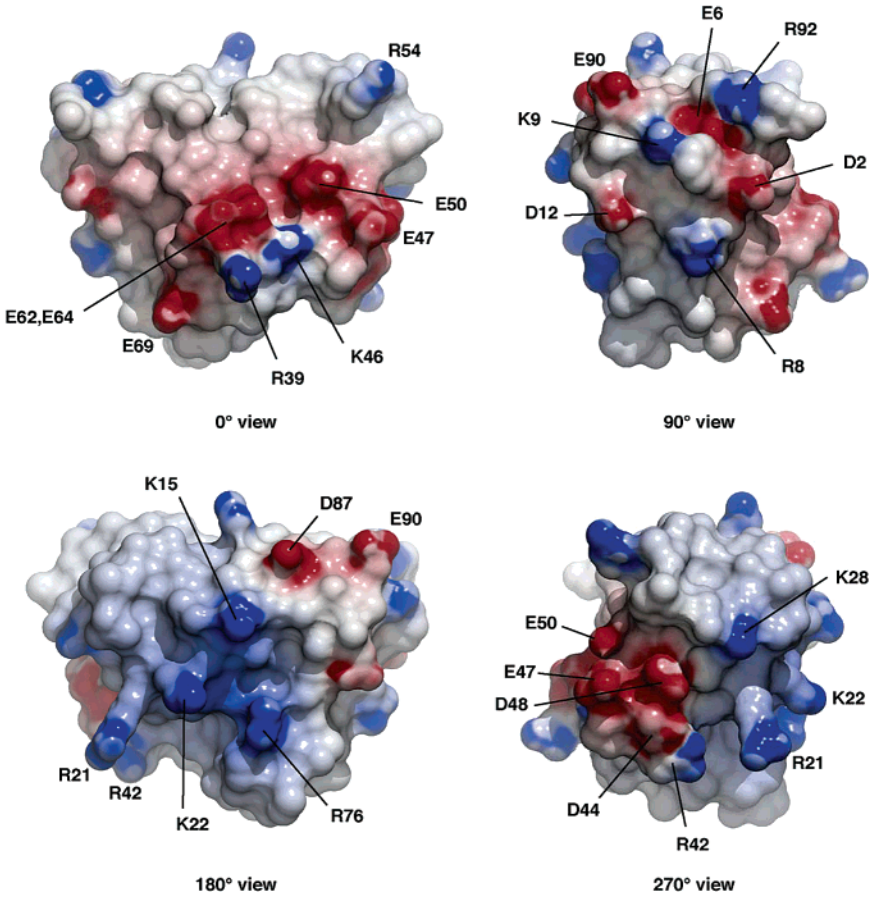

FIGURE 3: Surface electrostatic potential of *T. celer* L30e. Shown here are four views obtained by rotating the respective molecules along a vertical axis. The view at 0° is at the same orientation as that in Figure 2. The molecular surface is calculated with *GRASP* (*27*) and rendered with *POV4GRASP* (http://pov4grasp.free.fr/). The surface colors red, white, and blue represent electrostatic potential values of −10, 0, and 10 kT/e, respectively.

Molecular Surface. The differences in amino acid compositions are expressed at the surfaces of the molecules. The total solvent accessibility of both yeast and *T. celer* L30e are similar and are approximately 5400 Å^2. However, the ratios of nonpolar to polar surface areas are very different. The *T. celer* protein has 56% nonpolar surface and 44% polar surface, whereas the yeast homologue has 68 and 32%, respectively. The surface distribution of electrostatic potentials is very different in the two proteins. The yeast protein has a predominantly positively charged surface contributed by evenly located side chains. The surface charge distribution of *T. celer* L30e is heterogeneous (Figure 3), with an asymmetric arrangement of positive and negative charges. The positively charged group is located at the RNA binding surface (see later) and consists of Lys15, Arg21, Lys22, Lys28, Arg42, and Arg76 (Figure 3, 180° view and 270° view). Adjacent to these positively charged residues, there is a negatively charged group of acidic residues on helix α3 consisting of Asp44, Glu47, Asp48, and Glu50 (Figure 3, 270° view). The rest of the surface is covered by a network of ion pairs (Figure 3, 0° and 90° views).

Ion Pair Network. Two extensive clusters of charged residues are found on the exterior of *T. celer* L30e contributing favorable short- to medium-range (up to 10 Å apart) electrostatic interactions. One cluster involves mostly side chains of residues in helix α1: Asp2, Glu6, Arg8, Lys9, Asp12, Glu90, and Arg92 (Figure 3, 90° view). The second cluster involves Arg39, Lys46, Glu47, Glu50, Glu62, Glu64, Glu69, and Arg54 (Figure 3, 0° view).

Table 3: Comparison of Amino Acid Compositions of *T. celer* and Yeast L30e Proteins[a]

	T. celer	yeast	*T. celer* − yeast
hydrophobic			
Ala	11	8	3
Leu	9	13	−4
polar			
Asn	1	4	−3
Ser	5	10	−5
Thr	4	7	−3
charged			
Arg	7	3	4
Asp	5	2	3
Glu	8	5	3
Lys	7	12	−5
total hydrophobic	45	43	2
total polar	17	31	−14
total charged	27	22	5
charged − polar	10	−9	19

[a] Only amino acids that differ in three or more counts are listed.

Hydrophobic Core. Both *T. celer* and yeast L30e have similar hydrophobic contents (Table 3), and their cores are similar in size and contributed by well-conserved hydrophobic residues. In both cases, core residues pack efficiently and have no internal cavity.

Hydrogen Bonds. *T. celer* L30e has 68 ± 1 main chain to main chain (MC−MC) hydrogen bonds, including four with

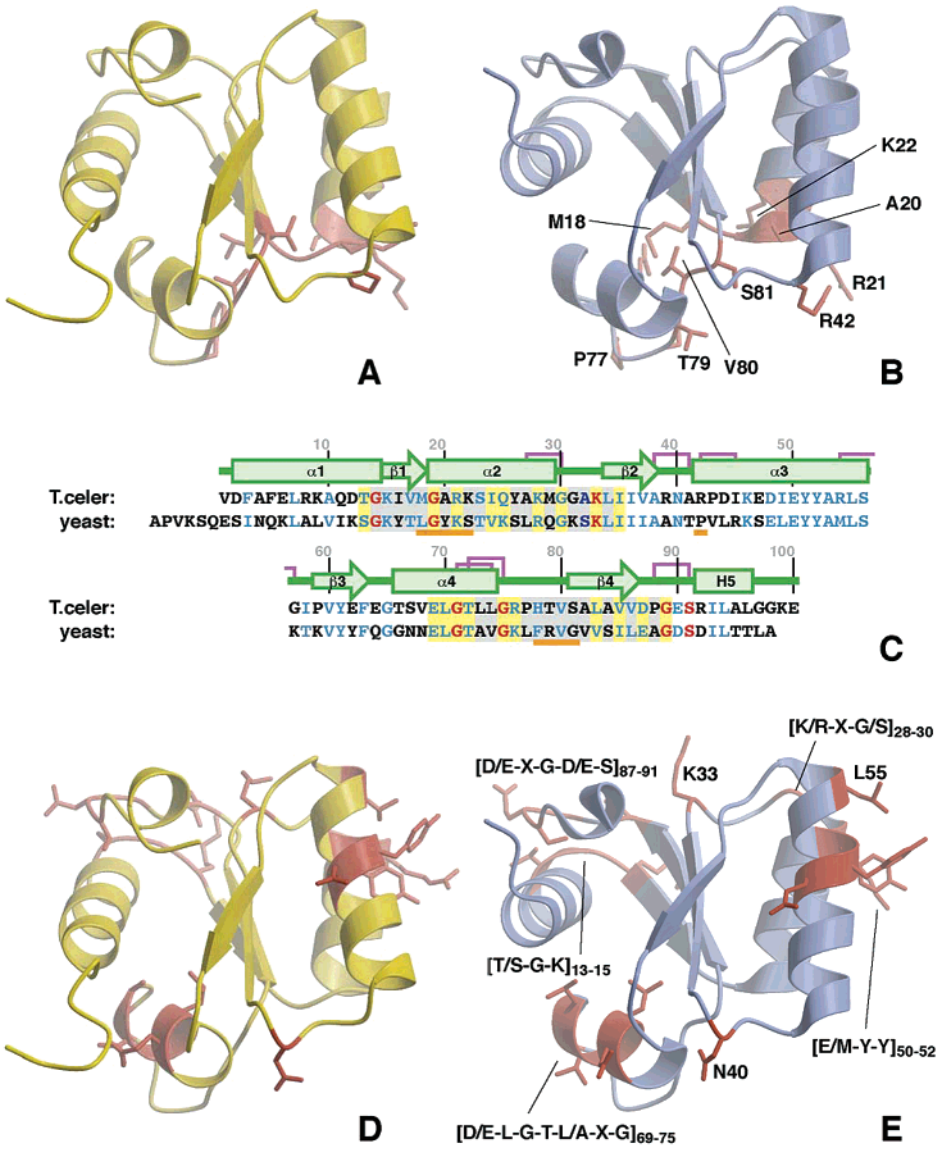

FIGURE 4: Important residues of *T. celer* L30e. The residues that are responsible for RNA binding are shown in orange for (A) yeast L30e and (B) *T. celer* L30e. (C) Sequence alignment of *T. celer* and yeast L30e sequences with the secondary structure of the *T. celer* protein (box for strand and arrow for helix) shown at the top. β-Turns are denoted with purple linking bars. Residues involved in RNA binding are underlined. The two L30e PROCITE signatures (see the text) are enclosed in gray boxes with homologous residues in the signatures highlighted with yellow boxes. Residues that are absolutely conserved among the L30e family are shown in red, and homologous (including size, shape, charge, polarity, and hydrophobicity considerations) residues are shown in blue. Highly conserved residues are shown in orange for (D) yeast L30e and (E) *T. celer* L30e.

Cα-H groups as donors. There are 17 $\pm$ 2 main chain to side chain (MC$-$SC) hydrogen bonds and one side chain to side chain (SC$-$SC) hydrogen bond in the *T. celer* structure. The side chains of Arg42 and Asp44 contributed two salt bridges at the N-terminus of helix α3. When atoms in the disordered loop are excluded, the yeast L30e structure has 65 MC$-$MC, 8 MC$-$SC, and 13 SC$-$SC hydrogen bonds. There are also two salt bridges contributed by side chains of Arg52 and Glu55, and Lys53 and Glu57. In a comparison of the two structures, the main chain hydrogen bond pattern that defines the secondary structure elements is well-conserved and the hydrogen bond counts are also similar. Both structures have two salt bridges. The major differences are in the hydrogen bonds contributed by side chain atoms. Most of the side chain hydrogen bonds of *T. celer* are formed with main chain atoms. In yeast, many more SC$-$SC hydrogen bonds are found while the number of MC$-$SC hydrogen bonds is only half of the number found in the *T. celer* structure.

Modeling of RNA Binding in the Ribosome

Residues responsible for RNA binding can be inferred from the yeast NMR structure and from sequence alignment (Figure 4). In the *T. celer* L30e protein, these residues are Met18, Gly19, Ala20, Arg21, Lys22, Arg42, Pro77, His78, Thr79, Val80, and Ser81 (Figure 4B). Two of these residues, Gly19 and Val80, are conserved between the two species and are responsible for binding the RNA backbone at the kink. In addition, the side chain of Arg76 may also be involved in interacting with the backbone of the RNA.

Three similar protein$-$RNA complex structures of the L30e/L7Ae family have now been reported (*9*, *30*, *31*). They all exhibit similar modes of binding to an RNA motif called the kink turn (KT) or the GA motif (*30*, *32*, *33*). We

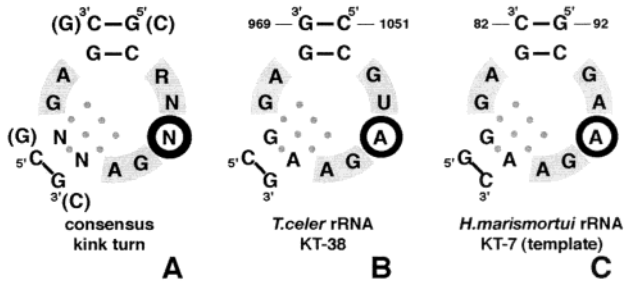

FIGURE 5: Ribosomal RNA motif that binds L30e. (A) Consensus kink turn motif (*30*). (B) Putative *T. celer* rRNA that binds L30e. A sequence that corresponds to KT-38 of *H. marismortui* rRNA. (C) KT-7 of *H. marismortui* rRNA used as a template for modeling. N denotes any nucleotide; R denotes a purine. Watson−Crick base pairing is shown as a solid line, and wobble base pairing is shown as dots. The nucleotide that twists out of the RNA backbone is circled.

attempted to build a model of *T. celer* L30e interacting with its native rRNA. A previous study has located a highly conserved region of the 23S rRNA to which yeast L30e binds (*8*). However, the homologous rRNA sequence in *T. celer* cannot be satisfactorily modeled as a kink turn motif (not shown). There is another region of the *T. celer* rRNA sequence, homologous to KT-38 of the *H. marismortui* rRNA, that is a putative binding site of L30e (Figure 5B). This *T. celer* rRNA sequence (nucleotides 961−970 and 1050−1062) satisfies all the characteristics of a kink turn (Figure 5A): a Watson−Crick base-paired G-C canonical stem, an asymmetric internal loop with three unpaired bases, and a sheared G·A noncanonical stem. The sequence of the *T. celer* rRNA is very similar to the sequence of KT-7 found in the *H. marismortui* 50S ribosomal subunit crystal structure (Figure 5C; *30*). It is the most straightforward model of the *T. celer* rRNA using KT-7 as a template that we can build (Figure 6A). Subsequently, the binary complex between the *T. celer* rRNA and L30e can be assembled (Figure 6B). This model offers a good understanding of how *T. celer* L30e interacts with RNA and confirms the important residues responsible for binding.

DISCUSSION

Thermal Stability. Extreme thermostability of proteins is a fundamentally interesting biological phenomenon and has been extensively studied. Over the years, a number of proteins from hyperthermophiles have been dissected and compared with their mesophilic homologues (see, for example, refs *34*−*36*). A whole spectrum of structural features have been identified that can enhance thermostability. It is consensus now that most proteins use a subset of this collection and that the overall thermostability is a combinatorial effect (for reviews, see refs *37*−*39*). Apparently, the only common feature that can be found in *most* of these proteins is an increase in the number of ion pairs (*36*, *37*). In this study, we observed that the *T. celer* L30e protein also employed a combination of structural features to enhance thermostability, and all of these are consistent with previous findings. The amino acid composition is delicately optimized for stability. In *T. celer* L30e, we saw an exchange of arginine for lysine because of better stability (*39*). A reduction in the numbers of asparagine, serine, and threonine residues can help in minimizing deamidation damage (*40*).

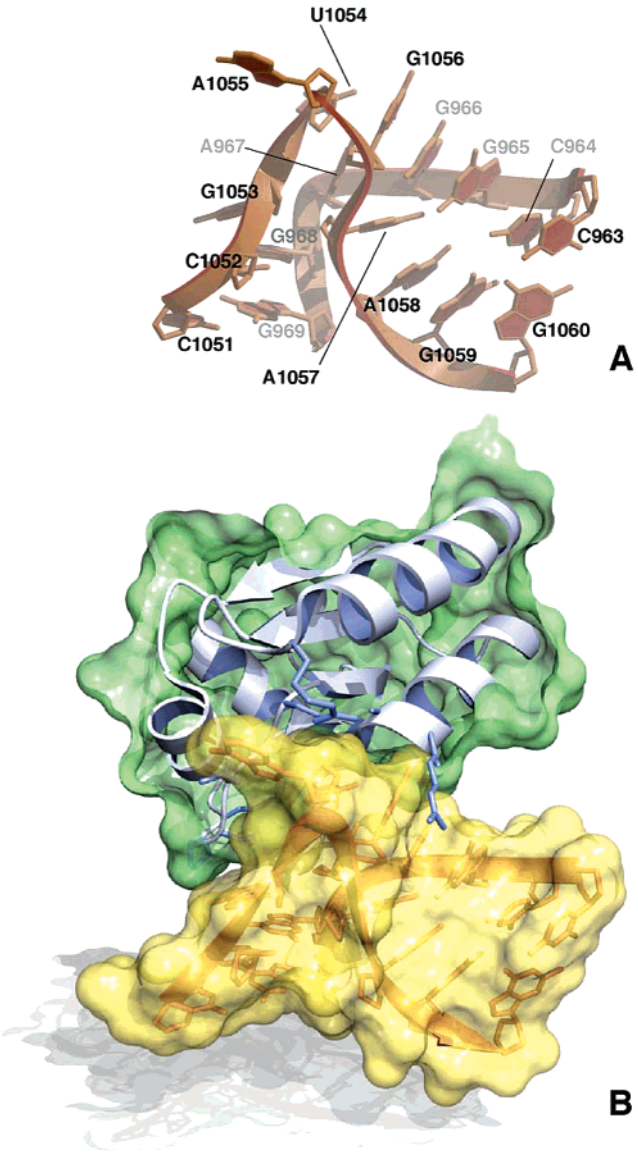

FIGURE 6: Model of L30e binding to ribosomal RNA. (A) Model of a *T. celer* rRNA kink turn to which L30e binds. (B) Model of the binary complex of *T. celer* L30e (green) and the rRNA kink turn (yellow) on the *T. celer* ribosome. The L30e surface is cut away to show the protein secondary structures (light blue). Residues that are responsible for RNA binding are shown with their side chains. The surfaces are generated with *GRASP* and atomic models with *POVSCRIPT+* (http://people.brandeis.edu/~fenn/povscript/) and rendered with *POV4GRASP*.

The difference in polar and nonpolar surfaces between the thermophilic protein and the mesophilic protein is more than 10%. The solvent exposure of nonpolar residues is minimized by adjusting the proportion of polar to nonpolar residues. The increase in surface charges has another stabilizing effect: it confers upon the protein a near-neutral overall charge, and repulsion of charges of the same polarity, which promotes unfolding, is suppressed (*41*). It is interesting to note that many *Escherichia coli* ribosomal proteins are indeed "natively unfolded" when studied in isolation from the ribosome because they carry high positive charges (*42*). The heterogeneous distribution of surface charges of *T. celer* L30e forms an ionic network and establishes favorable electrostatic interactions. The increase in the number of charged residues

in *T. celer* L30e is at the expense of uncharged polar residues, and this is consistent with other works (*38*, *40*, *43*).

The *T. celer* L30e crystal structure is more rigid than its mesophilic homologue. For example, helix α4 of *T. celer* L30e is well-structured, while that of the yeast homologue is disordered in the absence of an RNA substrate. There have been suggestions from previous works that overall rigidity can play a part in enhancing thermostability (*39*). Others are sceptical of whether rigidity is the cause of thermostability or the effect (*37*). Our analysis is further complicated by the fact that we are comparing structures that are determined at very different temperatures: the *T. celer* L30e crystal data was collected at 100 K, while the yeast L30e NMR data were obtained at 283 and 303 K. The structuring of the disordered loop in *T. celer* L30e may simply be the consequence of restrained atomic motions at low temperatures. At optimal temperatures, this RNA-binding motif may become more flexible. This possibility can be clarified by performing further NMR experiments at temperatures closer to the optimal growth temperature of *T. celer*. The fact that the *T. celer* protein has most of its side chain hydrogen bonds formed with main chain atoms does argue for a role of overall rigidity because the movement of side chain atoms is restricted by the protein backbone, whereas side chain to side chain hydrogen bonds can allow concerted side chain movements.

RNA Binding, Autoregulation, and Protein Evolution. Archaea are among the most ancient organisms that survived the harshest conditions in the early history of the earth. Yeast came in much later along the evolution path when the environment became milder and can sustain most life forms. The comparison of archaeal proteins with their eukaryotic homologues thus sheds light on how they evolve. It is conceivable that archaeal proteins are optimized for thermostability whereas yeast proteins are optimized for function, when the environmental selection pressure for thermostability is relaxed.

Being a near-neutral molecule, *T. celer* L30e interacts with RNA only when they are in proximity. Here the heterogeneous surface charge distribution comes into effect. A group of negatively charged side chains, on helix α3, is positioned adjacent to the positively charged RNA-binding site (Figure 3, 270° view), driving precise docking of the RNA. This is probably required to compensate for the lack of long-range electrostatic attraction. On the contrary, the yeast protein is an overall positively charged molecule and can interact with an RNA molecule from a distance. The RNA binding in the yeast protein is by induced fit. When the requirement for strict structural rigidity is relaxed, the yeast protein can exploit conformational adaptability to introduce a better binding mechanism. In addition, the flexibility of the RNA binding motif in yeast L30e allows tolerance of different substrates (pre-mRNA and mRNA), thus offering the tremendous advantage of feedback regulation. This broad substrate specificity of L30e from higher species is demonstrated by cross-species inhibition of splicing (*4*, *8*).

The degree of sequence homology among the L30e family proteins is very high (Figure 4C). Many of these highly conserved residues have a structural role. These include most hydrophobic residues that form the protein core, most noticeably the residues that constitute the two central β-strands, β2 (residues 34−38) and β4 (residues 82−86). Among all L30e sequences, there exists a region of high-level homology consisting of a "Thr/Ser-Gly-Lys" motif (residues 13−15; denoted as [T/S-G-K]:13−15 from here on) at the junction of α1 and β1 and a [D/E-X-G-D/E-S]:87−91 motif at the loop connecting β4 and H5 in its proximity (Figure 4E). The conserved residues in this "α1/β1···β4/H5 region" have important structural implications. The absolutely conserved Gly14, as the last residue of α1, allows an abrupt secondary structure transition from helix to strand without any loop residues. The absolutely conserved Gly89 allows a tight β-turn to be formed connecting β4 and helix 5. Thr13 (yeast L30e Ser20) forms a hydrogen bond bridging α1 and the nearby loop connecting β4 and helix 5. Ser91 serves as the N-cap and stabilizes helix 5. Taken together, the α1/β1···β4/H5 region is important in defining the proper conformation of the nearby RNA-binding residues (residues 18−22, Figure 4A,D, Figure 4B,E). Another highly conserved motif, [D/E-L-G-T-L/A-X-G]:69−75, is present at the C-terminal half of α4 that is close to another group of RNA-binding residues (residues 78−81, Figure 4E). The two absolutely conserved glycines (Gly71 and Gly75) make it possible to form a double β-turn at the C-terminus of α4. Glycine is rarely found in the middle of an α-helix (*44*), and its presence usually implies some local destabilization. Presumably, this highly conserved motif confers structural flexibility on the adjacent RNA-binding residues. Helices α2 and α3 contain a highly conserved [E/M-Y-Y]:50−52 motif at the third turn of α3 and a [K/R-X-G/S]:28−30 motif near the C-terminus of α2 in its proximity (Figure 4D,E). Of particular interest is the presence of the highly conserved Leu55 that is largely accessible to the solvent near this region. Together, these residues form a characteristic conserved surface that is distant from the RNA-binding site (Figure 4D,E). Long and bulky side chains on this surface are mostly exposed to the solvent and are likely to be involved in other protein or RNA interactions that are essential for L30e function. There are also a number of isolated polar or charged residues that are well-conserved, including the absolutely conserved Lys33 (Figure 4D,E), [N/T]:40, [K/R]:46, [D/E]:48, and [D/E/N]:87. Presumably, these are also important for protein or RNA interactions.

Role of L30e and Its Interaction with Ribosomal RNA. L30e is a component in most archaeal and eukaryotic ribosomes. However, some archaea, for instance, *Halobacterium halobium* (*45*) and *Haloarcula marismortui* (*12*), do not have L30e. In *H. marismortui*, structural homologue L7Ae (formerly identified as HMS6) shares a similar RNA-binding motif with L30e (*46*). L7Ae is unlikely to be a L30e homologue in this species for two reasons. First, according to PROSITE analysis (http://www.expasy.ch/prosite/), it is clear that L30e and L7Ae have distinct signatures [PROSITE references: L30e, PS00709 and PS00993 (Figure 4C); L7Ae, PS01082] and belong to different protein families. Second, in most eukaryotic organisms, both L30e and L7Ae are present.

We attempted to locate the L30e binding site on the *T. celer* ribosome. The rRNA sequence that was implicated in the work of Vilardell et al. (*8*) is not likely to be the correct L30e binding site because it cannot adopt a kink turn conformation. In the *H. marismortui* 23S rRNA, six kink turns are found. When mapped to the *T. celer* rRNA sequence, five of these (except KT-58, which may form a

double helix using wobble base pairing) can satisfy the consensus requirements for a kink turn. However, only KT-38 does not interact with any ribosomal proteins in the *H. marismortui* ribosome. We suspect that the *T. celer* rRNA that is homologous to KT-38 of *H. marismortui* is the putative L30e-binding site. By using molecular modeling, a good protein−RNA interaction complex can be built without major structural rearrangements. A preliminary electrophoretic mobility shift assay showed that *T. celer* L30e does bind *T. celer* rRNA KT-38 and the interaction is stronger than its binding with the rRNA corresponding to the region identified by Vilardell et al. (K. B. Wong, unpublished data)

When this complex was modeled on the whole *H. marismortui* 50S ribosome, we found that KT-38 was not accessible for binding. L30e atoms clash with those of ribosomal protein L18e. Further, neighboring rRNA in helix 28 and 29, a region that is conserved between *H. marismortui* and *T. celer*, would interfere with L30e binding. Although we can build a convincing model out of L30e interacting with a kink turn RNA, KT-38, it is inconclusive whether we have located the true ribosome-binding site. It is possible that L30e binds to the *T. celer* ribosome via a kink turn located elsewhere that is not conserved between *T. celer* and *H. marismortui*, since the latter does not bind L30e.

L30e is a good model system for structural studies. The structure of *T. celer* L30e and its comparison with the yeast homologue revealed many features that can enhance thermostability. With this knowledge in hand, we are now in a good position to examine individual structural features in greater detail by protein engineering. The structure reported here is of the highest resolution among the L30e/L7Ae family proteins determined to date and shall be informative in future ribosome studies.

ACKNOWLEDGMENT

We thank Professor Alan Fersht for helpful discussions and continuous support during sabbatical visits of K.-B.W. to the Centre for Protein Engineering.

REFERENCES

1. Wong, K. B., Wang, W. K., Proctor, M. R., Bycroft, M., and Chen, Y. W. (2001) *Acta Crystallogr. D57*, 865−866.
2. Dabeva, M. D., and Warner, J. R. (1987) *J. Biol. Chem. 262*, 16055−16059.
3. Mager, W. H., Planta, R. J., Ballesta, J.-P. G., Lee, J. C., Mizuta, K., Suzuki, K., Warner, J. R., and Woolford, J. (1997) *Nucleic Acids Res. 25*, 4872−4875.
4. Li, B., Vilardell, J., and Warner, J. R. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 1596−1600.
5. Wilson, K., Uyetake, L., and Boothroyd, J. C. (2000) *Mol. Biochem. Parasitol. 111*, 199−205.
6. Nakasone, K., Kenmochi, N., Toku, S., and Tanaka, T. (1993) *Biochim. Biophys. Acta 1174*, 75−78.
7. Mao, H., and Williamson, J. R. (1999) *J. Mol. Biol. 292*, 345−359.
8. Vilardell, J., Yu, S. J., and Warner, J. R. (2000) *Mol. Cell 5*, 761−766.
9. Mao, H., White, S. A., and Williamson, J. R. (1999) *Nat. Struct. Biol. 6*, 1139−1147.
10. Navaza, J. (1994) *Acta Crystallogr. A50*, 157−163.
11. Navaza, J., and Saludjian, P. (1997) *Methods Enzymol. 276*, 581−594.
12. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) *Science 289*, 905−920.
13. Holm, L., and Sander, C. (1991) *J. Mol. Biol. 218*, 183−194.
14. Chen, Y. W., Dodson, E. J., and Kleywegt, G. J. (2000) *Structure 8*, R213−R220.
15. Chen, Y. W. (2001) *Acta Crystallogr. D57*, 1457−1461.
16. Rice, L. M., and Brünger, A. T. (1994) *Proteins 19*, 277−290.
17. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., GrosseKunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905−921.
18. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
19. Read, R. J. (1986) *Acta Crystallogr. A42*, 140−149.
20. Abrahams, J. P., and Leslie, A. G. W. (1996) *Acta Crystallogr. D52*, 30−42.
21. Collaborative Computational Project No. 4 (1994) *Acta Crystallogr. D50*, 760−763.
22. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
23. Baker, E. N., and Hubbard, R. E. (1984) *Prog. Biophys. Mol. Biol. 44*, 97−179.
24. Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999) *J. Mol. Biol. 285*, 1735−1747.
25. McDonald, I. K., and Thornton, J. M. (1994) *J. Mol. Biol. 238*, 777−793.
26. Hubbard, S. J., and Thornton, J. M. (1993) *NACCESS*, http://wolf.bms.umist.ac.uk/naccess/.
27. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins 11*, 281−296.
28. Kleywegt, G. J., and Jones, T. A. (1994) *Acta Crystallogr. D50*, 178−185.
29. Wuyts, J., De Rijk, P., Van de Peer, Y., Winkelmans, T., and De Wachter, R. (2001) *Nucleic Acids Res. 29*, 175−177.
30. Klein, D. J., Schmeing, T. M., Moore, P. B., and Steitz, T. A. (2001) *EMBO J. 20*, 4214−4221.
31. Vidovic, I., Nottrott, S., Hartmuth, K., Lührmann, R., and Ficner, R. (2000) *Mol. Cell 6*, 1331−1342.
32. Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R., and Luhrmann, R. (1999) *EMBO J. 18*, 6119−6133.
33. Winkler, W. C., Grundy, F. J., Murphy, B. A., and Henkin, T. M. (2001) *RNA 7*, 1165−1172.
34. Vogt, G., and Argos, P. (1997) *Folding Des. 2*, S40−S46.
35. Sterner, R., and Liebl, W. (2001) *Crit. Rev. Biochem. Mol. Biol. 36*, 39−106.
36. Szilágyi, A., and Závodszky, P. (2000) *Struct. Folding Des. 8*, 493−504.
37. Petsko, G. A. (2001) *Methods Enzymol. 334*, 469−478.
38. Jaenicke, R., and Böhm, G. (1998) *Curr. Opin. Struct. Biol. 8*, 738−748.
39. Vieille, C., and Zeikus, G. J. (2001) *Microbiol. Mol. Biol. Rev. 65*, 1−43.
40. Haney, P. J., Badger, J. H., Buldak, G. L., Reich, C. I., Woese, C. R., and Olsen, G. J. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 3578−3583.
41. Karshikoff, A., and Ladenstein, R. (2001) *Trends Biochem. Sci. 26*, 550−556.
42. Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) *Proteins 41*, 415−427.
43. Cambillau, C., and Claverie, J.-M. (2000) *J. Biol. Chem. 275*, 32383−32386.
44. Williams, R. W., Chang, A., Juretic, D., and Loughran, S. (1987) *Biochim. Biophys. Acta 916*, 200−204.
45. Klenk, H.-P., Schwass, V., and Zillig, W. (1991) *Nucleic Acids Res. 19*, 6047.
46. Koonin, E. V., Bork, P., and Sander, C. (1994) *Nucleic Acids Res. 22*, 2166−2167.
47. Esnouf, R. M. (1997) *J. Mol. Graphics Modell. 15*, 132−134.
48. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol. 277*, 505−524.

BI027131S