J Mol Evol (1989) 29:448–462

Journal of Molecular Evolution

Sequence Alignment and Evolutionary Comparison of the L10 Equivalent and L12 Equivalent Ribosomal Proteins from Archaebacteria, Eubacteria, and Eucaryotes

Lawrence C. Shimmin,[1] Celia Ramirez,[2] Alistair T. Matheson,[2] and Patrick P. Dennis[1]

[1] Department of Biochemistry, The University of British Columbia, Vancouver, British Columbia V6T 1W5, Canada
[2] Department of Biochemistry and Microbiology, The University of Victoria, Victoria, British Columbia V8W 2Y2, Canada

Summary. The genes corresponding to the L10 and L12 equivalent ribosomal proteins (L10e and L12e) of *Escherichia coli* have been cloned and sequenced from two widely divergent species of archaebacteria, *Halobacterium cutirubrum* and *Sulfolobus solfataricus*. The deduced amino acid sequences of the L10e and L12e proteins have been compared to each other and to available eubacterial and eucaryotic sequences. We have identified the human P0 protein as the eucaryotic L10e. The L10e proteins from the three kingdoms were found to be colinear. The eubacterial L10e protein is much shorter than the archaebacterial–eucaryotic proteins because of two large deletions, one internal and one at the carboxy terminus. The archaebacterial and eucaryotic L12e proteins were also colinear; the eubacterial protein is homologous to the archaebacterial and eucaryotic L12e proteins, but has suffered rearrangement through what appear to be gene fusion events. Intraspecies comparisons between L10e and L12e sequences indicate the archaebacterial and eucaryotic L10e proteins contain a partial copy of the L12e protein fused to their carboxy terminus. In the eubacteria most of this fusion has been removed by the carboxy terminal deletion. Within the L12e-derived region, a 26-amino acid-long internal modular sequence reiterated thrice in the archaebacterial L10e, twice in the eucaryotic L10e, and once in the eubacterial L10e was discovered. This modular sequence also appears to be present as a single copy in all L12e proteins and may play a role in L12e dimerization, L10e–L12e complex formation, and the function of L10e–L12e complex in translation. From these sequence comparisons a model depicting the evolutionary progression of the L10e and L12e genes and proteins from the primordial state to the contemporary archaebacterial, eucaryotic, and eubacterial states is presented.

Key words: *Halobacterium* — *Sulfolobus* — Translation — Ribosome — Evolution

Offprint requests to: P.P. Dennis

Introduction

In the progenote state the gradual development of a template-directed protein synthesis apparatus is believed to have accompanied the progression from a catalytic RNA world to a ribonucleoprotein world (Darnell and Doolittle 1986). Although the molecular details have been refined independently for efficiency and accuracy, the basic features of the translation apparatus have been preserved in each of the three lines of descent from the primordial ancestor (i.e., in the eubacteria, the eucaryotes, and the archaebacteria; Woese and Fox 1977). The central component of the translation apparatus in all contemporary organisms is a ribonucleoprotein particle, the ribosome. The ribosome utilizes an mRNA template to align and polymerize amino acids (carried on adaptor tRNAs) into proteins. During the polymerization cycle, conformational rearrangements occur within the structure of the ribosome; many of these are mediated by interactions with extrinsic protein factors and the concomitant hydrolysis of GTP (Burma et al. 1986).

The eubacterial ribosomal "A" protein complex

forms the stalk structure on the large ribosomal subunit (Strycharz et al. 1978; Kastner et al. 1981; Marquis et al. 1981; Moller et al. 1983; Traut et al. 1986), and is comprised of four copies (a pair of dimers) of the L12e protein bound to a single copy of the L10e protein (Osterberg et al. 1977; Gudkov et al. 1978; Pettersson and Liljas 1979). It is likely that the L10e protein binds to the ribosomal RNA cooperatively with the L11e protein (Pettersson 1979); the L11e binding site on the large-subunit RNA has been characterized well (Schmidt et al. 1981; El-Baradi et al. 1987).

In *Escherichia coli,* the genes encoding the L11, L1, L10, and L12 proteins are contained within a 3-kbp region of genomic DNA. The region has been sequenced (Post et al. 1979), and the regulation and expression of the genes within it have been characterized extensively (Lindahl and Zengel 1986; Downing and Dennis 1987). We have cloned the corresponding genes from two widely divergent archaebacterial species, *Halobacterium cutirubrum* NRCC 34001 and *Sulfolobus solfataricus* P1, and the L10e and four different L12e genes from the eucaryote *Saccharomyces cerevisiae* (our unpublished data; Shimmin and Dennis 1989; Shimmin et al. 1989). Recently, sequences for *S. cerevisiae* genes similar to ours have appeared (Mitsui and Tsurugi 1988a,b,c; Remacha et al. 1988). Amazingly, for both archaebacteria the same four genes are linked, and the order as present in *E. coli* is preserved perfectly. In this paper we have utilized the L10e and L12e archaebacterial amino acid sequences (derived from the nucleic acid sequences and confirmed by partial and complete protein sequencing) to construct an interkingdom alignment for the two proteins and to characterize at the molecular level the evolutionary divergence that has occurred within the ribosomal "A" protein complex. The complete nucleotide sequence and genetic characterization of the *H. cutirubrum* and *S. solfataricus* clones and a discussion of the evolution of the L11e and L1e genes (and proteins) will be published elsewhere (Ramirez et al. 1989; Shimmin and Dennis 1989).

There are presently available 5 complete amino acid sequences of L10e proteins [1 eubacterial (Post et al. 1979), 2 archaebacterial, and 2 eucaryotic (Rich and Steitz 1987; Mitsui and Tsurugi 1988a; our unpublished data) and 27 complete amino acid sequences of L12e proteins [9 eubacterial (Terhorst et al. 1973; Itoh and Wittmann-Liebold 1978; Itoh 1981a; Bartsch et al. 1982; Itoh et al. 1982; Itoh and Higo 1983; Falkenberg et al. 1986; Garland et al. 1987; Matheson et al. 1987), 5 archaebacterial (Itoh et al. 1988; Matheson et al. 1988; Strobel et al. 1988), and 13 eucaryotic (Amons et al. 1979; Itoh 1981b; Amons et al. 1982; Lin et al. 1982; Beltrame and Bianchi 1987; Qian et al. 1987; Rich and Steitz 1987; Wigboldus 1987; Mitsui and Tsurugi 1988b,c; Remacha et al. 1988; our unpublished data)]. Although little is known of the structure or function of the L10e proteins, the L12e protein, especially Eco[1] L12, has been studied extensively. Eco L12 is a highly elongated molecule (Osterberg et al. 1976) composed of an N- and a C-terminal domain connected by an alanine–proline-rich region that is believed to function as a hinge (Leijonmarck et al. 1981). The N-terminal domain dimerizes spontaneously (Gudkov and Behlke 1978) and contains the site for binding L12 to L10 (Koteliansky et al. 1978). The C-terminal domain has been crystallized as a dimer, and shows a very compact structure of alternating alpha helices and beta sheets (Leijonmarck and Liljas 1987). The structure contains an anion (potential GTP) binding site, a putative dimerization site, and a conserved face for interaction with extrinsic translation factors that bind to the ribosome sequentially during protein synthesis. The functions of five translation factors (IF-2, EF-Tu, EF-G, RF-1, and RF-2) are known to depend upon Eco L12. The first three of these factors associate with the ribosome in complex with a GTP molecule to promote a structural rearrangement before GTP is hydrolyzed and the factor is released from the ribosome (see Liljas 1982 for a review).

Alignment comparisons of L12e proteins from all three kingdoms have been made previously. The archaebacterial and eucaryotic proteins show an end-to-end linear correspondence (Matheson et al. 1979). The eubacterial protein cannot, however, be aligned easily with its archaebacterial–eucaryotic counterpart. This has resulted in an enigma: a perplexing series of alignments derived from a variety of sequence and structural criteria, some of which are equally meritorious but apparently mutually exclusive. Alignments based on duplications (Amons et al. 1979; Jue et al. 1980), linear correspondence (Yaguchi et al. 1980; Wittmann-Liebold 1986), transpositions (Lin et al. 1982; Matheson 1985; Otaka et al. 1985), and conservation of structural features (Liljas et al. 1986) have been proposed. All of these alignments consider the evolution of the L12e gene (and protein) in isolation.

In this paper we present an alignment for the L10e genes from eubacteria, archaebacteria, and eucaryotes and demonstrate that the human P0 complementary DNA (cDNA) sequenced by Rich and Steitz

[1] Abbreviations used as organism identifiers in protein names are as follows: Asa, *Artemia salina*; Bst, *Bacillus stearothermophilus*; Dme, *Drosophila melanogaster*; Eco, *Escherichia coli*; Hcu, *Halobacterium cutirubrum*; Hsa, *Homo sapiens*; Mly, *Micrococcus lysodeikticus*; Sso, *Sulfolobus solfataricus*; and Sce, *Saccharomyces cerevisiae*

(1987) encodes a protein homologous to Eco L10. We have identified and characterized extended regions of homology between and within the L10e and L12e proteins. From this new information we provide an interkingdom alignment for the L12e proteins that incorporates many of the features of previous alignments and explains some of their enigmatic features. Finally, we propose a model that integrates the genetic and structural evolution of the L10e and L12e genes and proteins.

Materials and Methods

The alignments were based on the sequence similarity alignment given by the FASTP protein alignment program and optimized by manually maximizing the amino acid identities (Lipman and Pearson 1985). Precise placement of gaps was decided by maximizing conservative substitutions at the amino acid level and identities at the nucleotide level. The L10e alignments are based on five complete and one partial amino acid sequence (two each of eubacterial, archaebacterial, and eucaryotic). The alignment of the L10e proteins from positions 1 to 218 (Fig. 1) was based solely on sequence similarity. In addition to sequence similarity, in some cases known or hypothesized structure–function relationships were utilized for alignment of the L10e proteins for positions 219–374 (Fig. 1) and for the L12e proteins over their entire length. The L12e alignments are based on nine eubacterial, five archaebacterial, and eight eucaryotic sequences, although only two species from each kingdom are shown in Fig. 2. It is impossible to state explicitly the relative importance of sequence similarity versus structure–function for these alignments. For example, the ala–pro-rich region in Eco L12 is believed to function as a flexible hinge between the N-terminus (which binds L12 to L10) and the C-domain (which binds translation factors); the ala–pro-rich regions in L10e therefore have been aligned based on the hypothesis that they serve a similar function. The merit of the alignment therefore must be considered within both a structure–function and a sequence similarity context.

The 26 amino acid modules are too short and have diverged too greatly for any single module to have a statistically significant match to any other module. However, the modules as a group have a statistically significant match. To establish this, two hypothetical proteins of 194 amino acids were constructed from the tandem L10e modules such that a linear comparison of the two proteins yielded all the intraspecies L10e matches as indicated in all the A, B, and C lines of Fig. 3, i.e.,

Protein 1 Hcu α β γ Sso α β γ Hsa β Eco β′
Protein 2 Hcu β γ α Sso β γ α Hsa γ Eco γ′

The actual match score was calculated manually from the PAM 250 matrix of Dayhoff (1978). Simulated random match scores for each artificial protein versus jumbled versions of the second artificial protein were generated with the RDF program (Lipman and Pearson 1985). The significance (z) of the overall module match was calculated by subtracting the random match value from the actual match value and dividing by the standard deviation of the randomized match values. A value of z of 10 or greater is considered to be a significant homology.

Results

Alignment of conserved amino acid residues in homologous proteins from distantly related organisms often requires the introduction of gaps into one or more of the sequences. At the nucleic acid level such a gap generally corresponds either to a simple deletion of codons from one ancestral gene or to an insertion of codons into the other ancestral gene. It is often not possible to distinguish which event has occurred. The terms "deletion" and "insertion" are used imprecisely in this paper to refer, respectively, to the gapped and extended sequences with no bias as to which event actually occurred.

Sequence Alignments of the L10 Equivalent Proteins

The archaebacterial L10e genes from *H. cutirubrum* and *S. solfataricus* encode proteins of 352 and 337 amino acids, respectively. The end-to-end alignment of the two proteins is illustrated in Fig. 1. The proteins exhibit greater than 30% sequence identity, with no deletions or insertions through the first 302 positions (common scale). Identities beyond position 302 are negligible except for the extreme carboxy terminus (positions 367–373). Although no identities occur between positions 344 and 365, this region in both proteins is rich in charged amino acids: the *S. solfataricus* region contains 9 glutamic acid and 6 lysine residues and the *H. cutirubrum* region contains 12 aspartic acid and 2 glutamic acid residues. The high concentration of acidic residues in proteins from halophilic archaebacteria is believed to be an adaptation to the high intracellular ionic strength (Bayley and Morton 1978). The *H. cutirubrum* protein also contains an ala–pro-rich region (positions 320–329) that precedes the charged region and that has been deleted in the *S. solfataricus* protein.

The archaebacterial L10e proteins can be aligned with the eucaryotic Hsa L10e and the shorter eubacterial Eco L10 proteins (Fig. 1). The significance of the archaebacterial to eubacterial protein sequence matches (z = 10 for Hcu L10e vs Eco L10 and z = 10 for Sso L10e vs Eco L10) is within the range regarded as indicative of certain homology. The ancestral gene encoding the 165-amino acid-long eubacterial *E. coli* protein appears to have suffered a large internal deletion (positions 141–258, common scale; Fig. 1), a 3′ terminal truncation (position 297 and beyond), and five shorter deletion or insertion events, one of which (positions 15 and 16) removed the unique and conserved tryptophan residue. A partial amino acid sequence of the Bst L10e protein indicates that it shares the features of the Eco L10 protein (Garland, Louie, Matheson, unpublished results).

The very high statistical significance of the archaebacterial versus eucaryotic protein sequence matches (z = 42 for Hcu L10e vs Hsa L10e and z

Fig. 1. Ribosomal protein L10e amino acid sequence alignment. The predicted L10e amino acid sequences derived from the corresponding gene sequences from the eubacterium *Escherichia coli* (Post et al. 1979), the archaebacteria *Halobacterium cutirubrum* and *Sulfolobus solfataricus*, and the eucaryote *Homo sapiens* (Rich and Steitz 1987) are aligned. Amino acid identities are recorded in lines **A** to **D** as follows: **A** solid boxes are identities between Eco L10 and Hsa L10e; **B** solid boxes are identities between Eco L10, Hcu L10e, and Sso L10e, stippled boxes are identities between Eco L10 and Hcu L10e, and open boxes are identities between Eco L10 and Sso L10e; **C** solid diamonds are intrakingdom identities between Hcu L10e and Sso L10e; **D** solid boxes are identities between Hsa L10e, Hcu L10e, and Sso L10e, stippled boxes are identities between Hsa L10e and Hcu L10e, and open boxes are identities between Hsa L10e and Sso L10e. Gaps required for alignment are indicated as dashes (–). The amino acid position scale is below the Hsa L10e sequence and the proteins are aligned over 374 amino acid positions.

= 25 for Sso L10e vs Hsa L10e) unequivocally demonstrates that these proteins are homologous. During the course of evolution following the divergence of archaebacteria and eucaryotes, the archaebacteria appear to have suffered a deletion (positions 305–319) preceding the ala–pro-rich region. The gene encoding the ancestral Hsa L10e protein appears to have suffered two internal deletions (positions 219–244 and 332–351) and five short deletion–insertion events. The deletion at position 332–351 follows the ala–pro-rich sequence and extends into the region of high amino acid charge density, which is also present in the Hcu L10e and Sso L10e proteins but missing from the Eco L10 protein.

Using the alignments presented in Fig. 1, pairwise sequence comparisons have been made among the four L10e proteins. The intrakingdom Hcu L10e and Sso L10e proteins are 27% identical in amino acid sequence, with a deletion or insertion (DI) index of one. Interkingdom comparisons yielded identity values of 15–25% and DI indexes of five to seven (Table 1). The eubacterial Eco L10 and the eucaryotic Hsa L10e proteins were the most dissimilar; in fact, the match is not significant ($z = 2$ for Hsa L10e vs Eco L10). The two archaebacterial proteins appeared to be midway between and equally related to both the eubacterial and eucaryotic proteins.

Sequence Alignments of the L12 Equivalent Proteins

The amino acid sequences for a number of eubacterial, eucaryotic, and archaebacterial L12e proteins have previously been determined. Alignment of two

typical but distantly related eubacterial (*E. coli* and *M. lysodeikticus*), two archaebacterial (*H. cutirubrum* and *S. solfataricus*), and two eucaryotic (*H. sapiens* and *S. cerevisiae*) sequences are presented (Fig. 2). Intrakingdom alignments and comparisons (Table 2) are made without difficulty and indicate that (1) the eubacterial sequences were most similar (63% identities, with a DI index of one), (2) the eucaryotic proteins were of intermediate similarity (57% identities, with a DI index of four), and (3) the archaebacterial proteins were least similar [42% identities, with a DI index of one; changes of lys and glu residues to asp during the adaptation to high salt for the Hcu L12e protein almost certainly contribute to this dissimilarity (Bayley and Morton 1978)].

The archaebacterial and eucaryotic L12e proteins can be aligned with each other end to end with the initiation methionine at position 59 (common scale). The eubacterial L12e protein could not be made to fit this pattern. A number of alignments between the eubacterial protein and the archaebacterial–eucaryotic proteins have been proposed. The alignment illustrated (Fig. 2) maximizes identities at both the amino acid and the nucleotide levels and preserves predicted structural features in the L12e protein alignments. At least two regions within the eubacterial protein appear to have homologous domains in the archaebacterial and eucaryotic proteins. The first domain common to the proteins from all three groups is located near the amino terminus of the archaebacterial and eucaryotic proteins and in the middle of the eubacterial protein (positions 74–123, common scale). Between the two archaebacteria, *H. cutirubrum* and *S. solfataricus,* this domain exhibited 48% (20/42) amino acid identity. Interkingdom similarities for this domain range from a minimum of 16% similarity between Hsa L12e and Hcu L12e to a maximum of 35% between Eco L12 and Hcu L12e (Table 3). In the Eco L12 protein this region contains the putative surface for factor interaction and sites utilized for L12 dimerization (Leijonmarck and Liljas 1987).

Table 1. The genes encoding the L10e proteins: amino acid and nucleotide sequence identities

	Average length[a]	Amino acid identities[b]	DI index[c]	Nucleotide identities[b]
Hcu/Sso	334	90 (27%)	1	383 (38%)
Eco/Hcu	169	40 (24%)	6	187 (37%)
Eco/Sso	169	35 (21%)	6	184 (36%)
Hsa/Hcu	330	82 (25%)	7	415 (42%)
Hsa/Sso	322	77 (24%)	6	319 (33%)
Eco/Hsa	163	25 (15%)	5	151 (31%)

[a] The average length in amino acids over the region of comparison of the two sequences; determined as the length of the comparison region minus half of the total gaps. Comparisons with Eco L10 are shortened by the 118-residue deletion (positions 141 to 258, Fig. 1)

[b] Amino acid and nucleotide identities are the number of perfect matches over the region of comparison. The protein alignment is illustrated in Fig. 1

[c] DI index is the number of deletions (or insertions) required to achieve alignment of the two sequences within the region of comparison

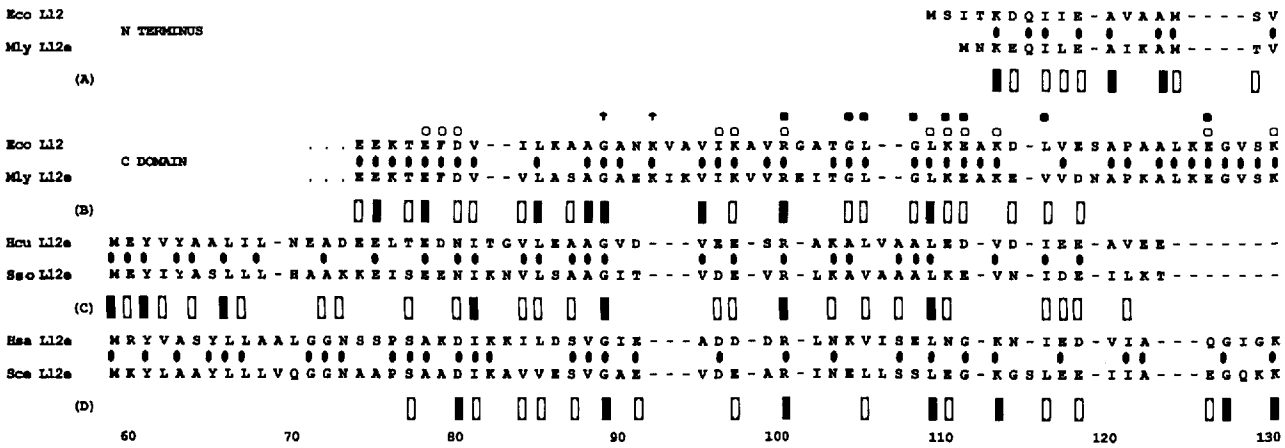

Fig. 2. Ribosomal protein L12e amino acid sequence alignment. The amino acid sequences from two eubacteria (*Escherichia coli* and *Micrococcus lysodeikticus*), two archaebacteria (*Halobacterium cutirubrum* and *Sulfolobus solfataricus*), and two eucaryotes (*Homo sapiens* and *Saccharomyces cerevisiae*) are illustrated. The N-terminal 58 amino acid positions of the eubacterial protein have no direct counterpart in the archaebacterial or eucaryotic proteins; rather, this segment exhibits a degree of sequence similarity with its own C-domain (positions 1–44 align with positions 109–152). In the eubacterial protein, position 58 is fused to position 74 and divides the protein into the N-terminus and the C-domain. The intervening positions, beginning at position 59, form the unique N-terminus of the archaebacterial and eucaryotic proteins. Amino acid identities are indicated as follows: Solid diamonds (♦) indicate intrakingdom identities between Eco L12 and Mly L12e, Hcu L12e and Sso L12e, and Hsa L12e and Sce L12e. Line **A** illustrates similarities between the N-terminal and C-domain regions of the eubacterial proteins. Lines **B**, **C**, and **D** illustrate interkingdom similarities between eubacteria (C-domain) and archaebacteria, between archaebacteria and eucaryotes, and between eubacteria (C-domain) and eucaryotes, respectively.

Table 2. The L12e proteins: intrakingdom amino acid identities

	Average length[a]	Amino acid identities[b]	DI index[c]
Eubacteria (Eco/Mly)	120	76 (63%)	1
Archaebacteria (Hcu/Sso)	110	46 (42%)	1
Eucaryotes (Hsa/Sce)	113	64 (57%)	4

Alignments are as illustrated in Fig. 2. The average length, amino acid identities, and DI index are as defined in the legend to Table 1

Table 3. The L12e common globular domain: amino acid and nucleotide sequence identities

	Average length[a]	Amino acid identities[b]	DI index[c]	Nucleotide identities[b]
Hcu/Sso	42	20 (48%)	0	60 (48%)
Eco/Hcu	43	15 (35%)	6	61 (47%)
Eco/Sso	43	12 (28%)	6	55 (43%)
Hsa/Hcu	43	7 (16%)	0	56 (43%)
Hsa/Sso	43	13 (30%)	0	54 (42%)
Eco/Hsa	42	8 (19%)	6	44 (35%)

The common globular domain within the L12e proteins occurs between positions 74 and 118 as illustrated in Fig. 2. The average length, amino acid and nucleotide identities, and DI index are as defined in the legend to Table 1

The second region that appeared to be common to the L12e protein from all three kingdoms was the ala–pro-rich sequence located at positions 149–166 in the archaebacterial and eucaryotic proteins and between positions 41 and 58 in the eubacterial protein. Even within kingdoms the length and sequence of these ala–pro-rich sequences are highly variable, with substitutions occurring between alanine, proline, serine, threonine, and glycine. They are believed to be unstructured and to function as hinges between domains of the L12e proteins (Leijonmarck et al. 1981), accounting for the observed high mobility of the C-domain of Eco L12 (Tritton 1978; Cowgill et al. 1984).

In the archaebacterial and eucaryotic L12e proteins the ala–pro hinge precedes a region in the carboxy terminal domain that contains a high concentration of acidic and basic amino acid residues. A similar motif, an ala–pro-rich hinge followed by a high charge density domain, was also observed near the carboxy terminus of the Hcu L10e and Hsa L10e proteins. The Sso L10e protein lacks most or all of the ala–pro-rich hinge but retains the region of high charge density.

When attempts were made to align the amino terminus of the eubacterial L12e protein (positions 3–38, Fig. 2), it was found to be more similar to its own carboxy terminus (positions 111–146) than to any sequence within the archaebacterial or eucaryotic proteins. For this intramolecular alignment there were 12 of 31 (39%) amino acid identities and 49 of 93 (53%) nucleotide identities, with a DI index of three. The second half of this intramolecular complementarity in the eubacterial protein appears to align with regions of the archaebacterial and eu-

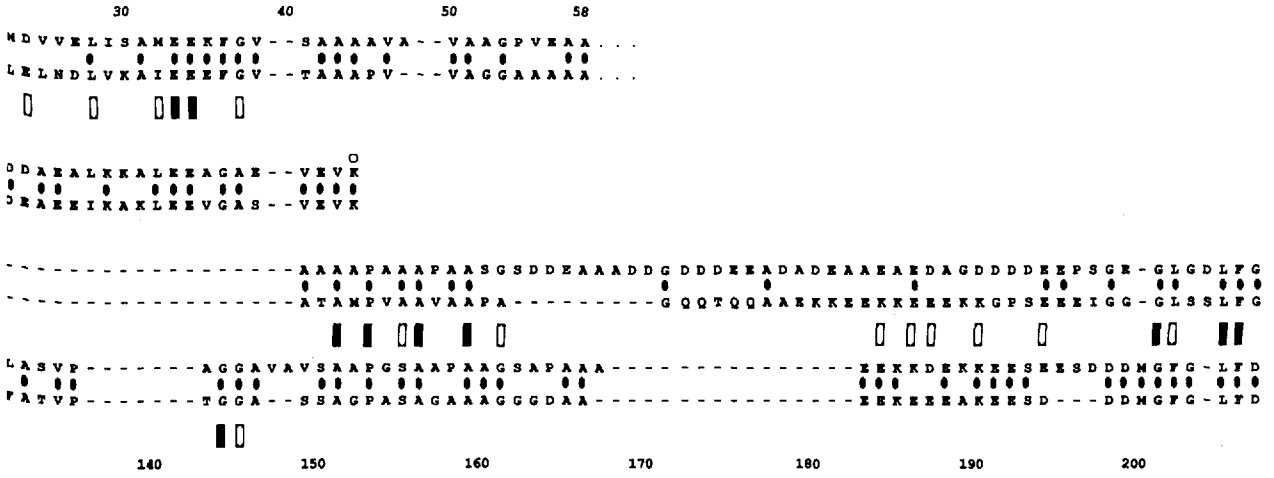

Fig. 2. Continued. Solid boxes (■) represent total identity, whereas open boxes (□) represent one or more conservative amino acid replacements. Conservative replacements are defined as substitutions within the following groups: (1) F, L, I, M, V; (2) R, K, E, D, N, Q, H; (3) A, S; (4) A, G; (5) S, T; and (6) F, Y. Gaps required for alignment are indicated as dashes (–). The region of positions 41–58 within the N-terminus of the eubacterial protein and approximate positions 149–166 of the archaebacterial and eucaryotic proteins are homologous alanine–proline-rich hinge regions. Residues participating in known structural features of the Eco L12 protein are illustrated immediately above its C-domain sequence. Open squares (□) represent residues involved in the conserved face, filled squares (■) represent residues involved in the dimerization site, and hatch marks (+) represent residues involved in the anion (potential GTP) binding site. The amino acid sequence scale is indicated at the top and bottom and the proteins are aligned over 207 amino acid positions.

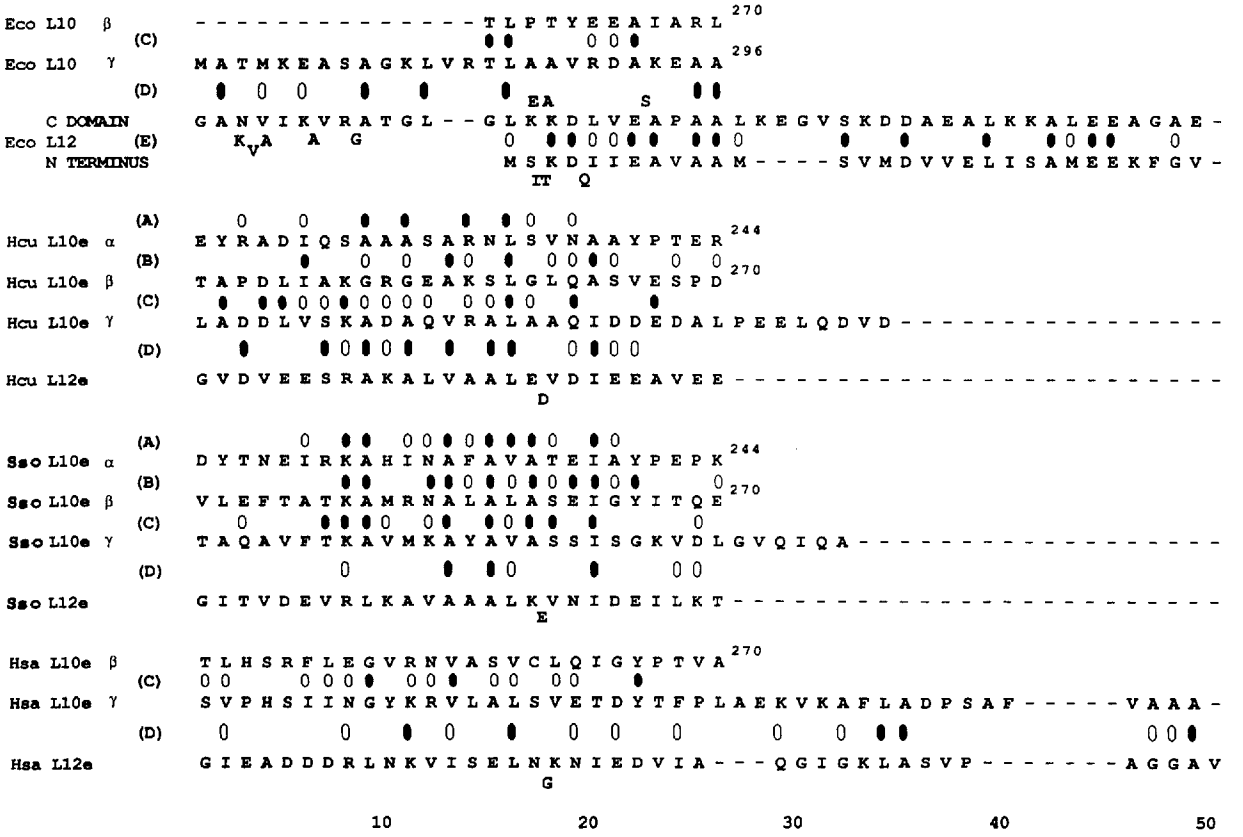

Fig. 3. Amino acid identities and similarities in the C-terminus and module regions of the L10e and L12e proteins. Four intraspecies comparisons of L10e and L12e amino acid sequences are presented from top to bottom: *Escherichia coli, Halobacterium cutirubrum, Sulfolobus solfataricus,* and *Homo sapiens.* The L10e α, β, and γ sequences are the three 26-residue-long module repeats. For Eco L12, where the protein has undergone major rearrangements and alterations during eubacterial evolution, a complete and a partial copy of the module appear to be present in the C-domain and N-terminus, respectively. In HSA L10e one copy of the module is not present.

caryotic proteins that are interrupted by deletion (approximate positions 124–146 in Fig. 2).

Homologies within and between the L10e and L12e Proteins

It has already been noted that both the L10e and L12e proteins from archaebacteria contain a region of high amino acid charge density near their carboxy terminal ends. When compared, the L10e and L12e proteins were found to exhibit a high degree of sequence similarity at their carboxy terminal ends (Fig. 3). For the two *S. solfataricus* proteins, the 31 carboxy terminal residues that contain the region of high charge density were found to be identical except for an extra glycine at the end of the L10e sequence. Remarkably, conservation at the nucleotide level was also perfect. For the *H. cutirubrum* proteins the degree of amino acid identity was highly significant, although less pronounced.

When these L10e–L12e archaebacterial alignments were extended into the central regions of the proteins, a modular sequence 26 amino acids in length was discovered. The module was found once in the L12e proteins and tandemly reiterated three times in the L10e proteins. The three L10e module copies were designated α, β, and γ. These modular sequence domains are separated from the high charge density domain by the ala–pro-rich hinges in Hcu L10e, Hcu L12e, and Sso L12e. In the Sso L10e protein, the ala–pro-rich hinge is mostly or completely absent.

Sequence similarity at the carboxy terminal ends of the eucaryotic Hsa L10e and Hsa L12e proteins has been noted (Rich and Steitz 1987). The region of similarity includes the ala–pro-rich hinge and the high charge density domain. When the alignment was extended, a single copy of the module domain was discovered in the Hsa L12e protein sequence and two tandemly reiterated copies were discovered in the Hsa L10e protein sequence. The L10e α module apparently either was never generated in the ancestral eucaryotic gene or was removed by a deletion covering positions 219–244 (Fig. 1).

The eubacterial Eco L10 and Eco L12 proteins do not exhibit sequence similarity at their carboxy terminal ends; neither protein contains the carboxy terminal region of high amino acid charge density

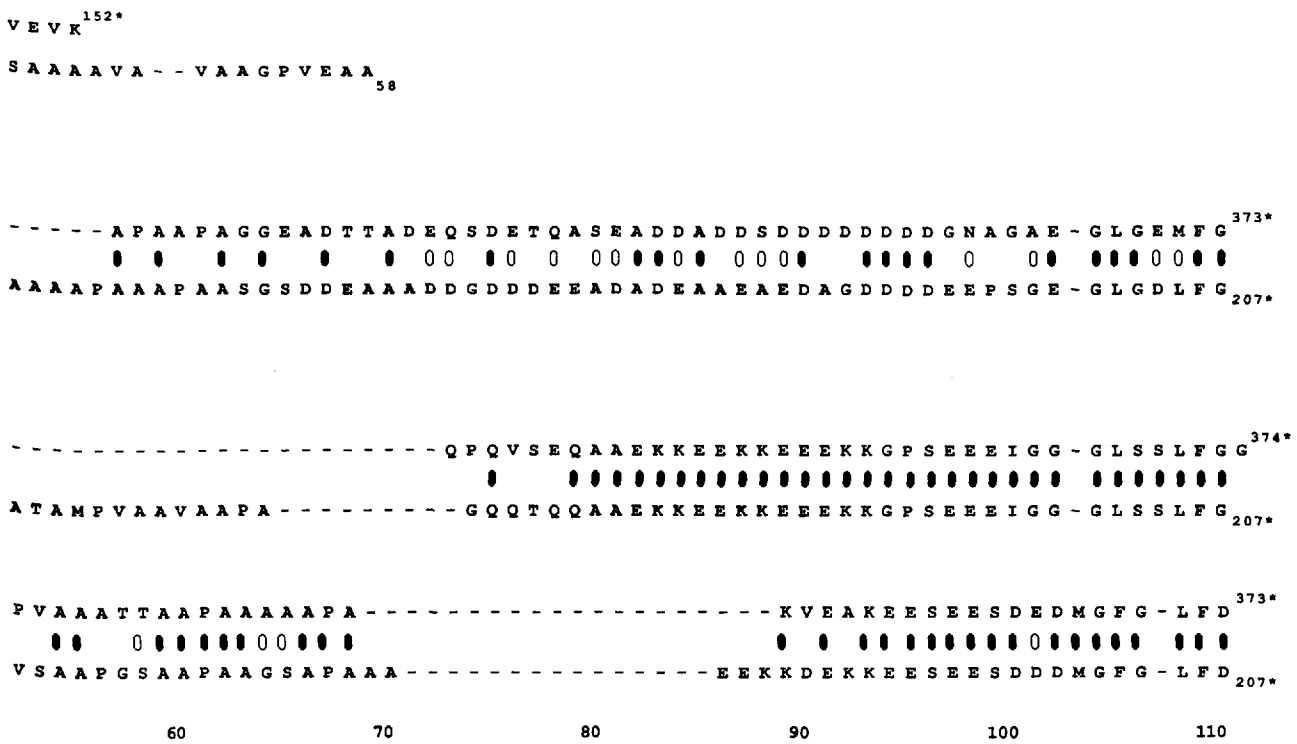

Fig. 3. Continued. Identities (◆) and conservative replacements (◇) are as follows: line **A** compares L10e α with γ; line **B** compares L10e α with β; line **C** compares L10e β with γ; line **D** compares L10e γ module with the L12e module, and for *H. cutirubrum*, *S. solfataricus*, and *H. sapiens* compares the C-termini of L10e and L12e. The numbers at the ends of the sequences designate the position number of the terminal amino acid of modules and proteins (from Figs. 1 and 2). Residues representing the C-termini of the respective proteins are identified (*). Conservative replacements are as defined in the legend to Fig. 2.

found in the archaebacterial and eucaryotic equivalent proteins. When the Eco L10 and Eco L12 sequences were examined in relationship to their alignment with the corresponding archaebacterial and eucaryotic proteins (as illustrated in Figs. 1 and 2), potential single intact and partial copies of the residual modular sequence were discovered (Fig. 3).

Although matches between individual modules have low statistical significances, the modules when treated as a group were highly significant ($z = 15$ by the RDF program; see Materials and Methods). In addition, the presence of a gap of precisely 26 residues eliminating the α module within the eucaryotic L10e proteins (both Hsa L10e and Sce L10e) strengthens the evidence for the existence of the modules (our unpublished data).

Discussion

The L10 Equivalent Proteins

We have aligned four L10e proteins representing the three kingdoms: one eubacterial (*E. coli*), two archaebacterial (*H. cutirubrum* and *S. solfataricus*), and one eucaryotic (*H. sapiens*). The archaebacterial Hcu L10e and Sso L10e proteins are structurally well conserved, with only a single deletion eliminating the hinge region in Sso L10e. The low amino acid sequence similarity (27%) is at least partially the result of modification required for adaptation to high salt in halophilic archaebacteria.

The eubacterial Eco L10 protein is homologous to the Hcu L10e and Sso L10e proteins by virtue of both sequence similarity and genetic linkage. In all three organisms the L10e genes are contained at the same position within the conserved L11e, L1e, L10e, L12e tetragenic cluster. The probability of nonhomologous genes that perform analogous functions (i.e., factor binding GTPase center) fortuitously clustering in two separate lineages is exceedingly remote. The sequence similarity and structural features (modules, L12e-like C-terminus) of the alignment of eucaryotic Hsa L10e with Hcu L10e and Sso L10e unequivocally indicate that these proteins are homologous and thus, despite very low similarity (15% identity at the amino acid level) and statistical significance ($z = 2$), the eucaryotic Hsa L10e must be the homologue of the eubacterial Eco L10.

The Hsa L10e protein is known to form a complex with the L12e proteins analogous to the L10–L12 complex of *E. coli* (Rich and Steitz 1987). Rich and Steitz (1987) independently recognized that the extreme carboxy 30 amino acids of the human P0 protein (here identified as Hsa L10e) were highly similar to the carboxy terminal sequences conserved among eucaryotic L12e proteins; because of insufficient information they were unable to identify P0 as the L10e protein. Thus archaebacterial proteins may serve as a link in identifying proteins that have diverged too greatly between the eubacterial and eucaryotic kingdoms to be demonstrably homologous by sequence similarity methods.

The highly charged C-terminal sequences of the archaebacterial–eucaryotic L10e proteins can be aligned with the C-termini of the corresponding L12e proteins (Fig. 3). From the C-terminus, similarity extends through the ala–pro-rich hinge region to position 271 (Fig. 1) in L10e and position 89 (Fig. 2) in L12e. It is interesting to note that deletions–insertions appearing in the archaebacterial–eucaryotic L12e protein appear in similar, sometimes identical, positions in the L10e proteins (Fig. 3). This suggests that these sequences are likely still functional homologues and that selection or possibly recombination events preserve the similarity of the sequences. The only anomaly in this pattern is the lack of the ala–pro-rich hinge feature in the Sso L10e protein. The absolute identity of the Sso L10e and Sso L12e proteins from positions 79 to 110 (Fig. 3) at both the amino acid and nucleic acid level suggests a very recent recombinational restoration event. If this event removed the hinge region leaving only one proline at position 74 (Fig. 3), then other thermoacidophilic archaebacteria may retain the extended hinge region. Within the eucaryotes (*H. sapiens, A. salina, D. melanogaster,* and *S. cerevisiae*) the C-terminus of the multiple copies of the L12e (and, where known, L10e) proteins is highly conserved within a species but variable in sequence and length between species (our unpublished data; Ramirez et al. 1989). This is indicative of periodic recombination or a gene conversion-like mechanism maintaining the virtual identity of these charged sequences within species. The presence in *S. solfataricus* of identical carboxy termini in the L10e and L12e proteins indicates that this mechanism may be extremely ancient and widely distributed.

We have demonstrated the existence of a statistically significant ($z = 15$ by RDF; see Materials and Methods) module of 26 amino acids in the L10e proteins, thrice repeated in archaebacteria, twice in eucaryotes, and a complete plus a possible partial copy in eubacteria (Fig. 3). Similarity with the L12e proteins indicates that the module also exists in that protein. Several features of the module region are noteworthy. Similarity between the modules is greatest in the central section (positions 8–21, Fig. 3) of the modules within L10e (31% amino acid identities and 29% amino acid conservative substitutions). The flanking regions (positions 1–7 and 22–26, Fig. 3) have similarity that is close to random for sequences of this amino acid composition (10% identical and 15% amino acid conservative substitutions). The unique arginine of the L12e proteins (position 8, Fig. 3) generally aligns with a positively charged (five lysines) or hydrophilic (one glutamic acid, one asparagine, two serines) residue. The most highly conserved residue appears to be the hydrophobic residue in position 16 (11 leucines, 3 valines, 1 methionine). Alanine is also highly conserved in a specific pattern (positions 9, 13, 15, and 17, Fig. 3); this is most apparent in the Sso L10e protein, where of the 16 alanines present within the three modules, 12 align perfectly at these positions.

The eucaryotic Hsa L10e protein is missing one module (Figs. 1 and 3); which of the three modules is actually missing cannot be ascertained, although the alignment given in Fig. 1 (that is, missing the α module) yields the highest level of sequence similarity. The Sce L10e gene also has only two copies of the module (Newton et al. 1989). Similarly, the Eco L10 complete module matches well with the Hcu L10e γ module (nine identical amino acid residues) and because of this is aligned at the γ position. The 12 amino acids preceding the Eco L10 module may represent a second partial module; its alignment preserves the highly conserved hydrophobic leucine residue at position 16 (Fig. 3).

Short sequences flanking the triple module in Hcu L10e are strikingly similar; amino acid positions 210–218 (PEELEIDVD, Fig. 1) compared with 297–304 (PEELQDVD) have, excluding a one amino acid gap, seven out of eight amino acid residues and 21 out of 24 nucleotides identical (Shimmin and Dennis 1989). Whether this nearly perfect direct repeat in the DNA was involved in the module duplication process is unknown.

The L12 Equivalent Proteins

Biophysical studies on the Eco L12 protein indicate that the N-terminal domain spontaneously dimerizes and binds to the Eco L10 protein (Koteliansky et al. 1978). The C-terminal domain forms a compact structure that crystallizes as a dimer, contains an anion binding site, and may interact through a conserved face with extrinsic translation factors during the protein synthesis cycle (Leijonmarck and Liljas 1987). The two domains are separated by an ala–pro-rich region believed to function as a flexible hinge (Leijonmarck et al. 1981).

The alignment of the L12e proteins illustrated in

Fig. 2 implies that the N-terminal end of the archaebacterial–eucaryotic proteins contains the dimerization, anion, and factor binding domains. The alignment suggests that the ancestral globular domain comprised 75–80 amino acids. The eubacteria have lost approximately 15 amino acids (including the unique 2 conserved tyrosine residues) on the amino side of the region of the conserved face, whereas archaebacteria and eucaryotes have lost approximately 25 and 10 amino acids, respectively, from the carboxy side of the region of the conserved face. However, each has retained the dimerization, anion, and factor binding domains. The ala–pro-rich hinge region is easily identified on the amino and carboxy sides of the globular domain of the eubacterial and archaebacterial–eucaryotic proteins, respectively. The archaebacterial–eucaryotic L12e C-terminal ends are very similar both to each other and to the C-terminal end of the corresponding L10e proteins. It is possible that these conserved C-terminal sequences are utilized for auto association and assembly of the "A" protein complex and thus may be analogous (not homologous) in function to the N-terminal end of the eubacterial L12e protein (Liljas et al. 1986).

Some eucaryotes (*A. salina, D. melanogaster, H. sapiens,* and *S. cerevisiae*) are known to have two different types of L12e genes: L12e type I and L12e type II (Amons et al. 1982; Rich and Steitz 1987; Wigboldus 1987; our unpublished data). The type I proteins are most similar to the archaebacterial L12e proteins. The L12e type II proteins have an extended N-terminus, lack arginine at position 100, and contain tryptophan at position 102 (Fig. 2). In *S. cerevisiae* there are four different L12e proteins; a recent gene duplication has resulted in two copies of each type of L12e protein (our unpublished data). The two types are generally highly similar from the hinge to the carboxy terminus (likely due to gene conversion events) but differ substantially in the N-terminal region. The functional significance of two different L12e-like proteins in eucaryotes and apparently only one in archaebacteria and eubacteria remains to be elucidated.

The putative copy of the module present in Eco L12 (positions 89–123, Fig. 2) contains the L12 dimerization site of the globular domain (primarily positions 100–116). Alignment of the L12e proteins with the L10e proteins (Fig. 3) revealed that the dimerization site in the C-domain of Eco L12 (positions 8–20) was aligned with the region of highest conservation in the L10e modules (approximate positions 8–21). This would suggest that these modules may be reiterative L12 dimerization sites. Furthermore, the N-terminal end of the eubacterial L12e protein appears to be a duplication of part of this same dimerization site (Fig. 2). This may explain the tendency of the Eco L12 N-terminus to dimerize spontaneously.

The presence of these putative dimerization modules in all L10e proteins suggests a mechanism for interaction with the L12e proteins. It is known that the Eco L12 globular domain undergoes a conformational change upon interaction with extrinsic translation factors (Gudkov and Gongadze 1984). If this conformational change exposes the L12 dimerization site then the L12 protein could conceivably fold about the hinge and thus bring the L12 dimerization site into interaction with the dimerization site of the L10 module. This would bring the extrinsic translation factor to the ribosome surface in a specific orientation. It is possible that the multiple modules in the L10e proteins also serve as multiple interaction sites for the L12e protein. If other ribosomal proteins contain the module sequence, then the L12e protein (with bound translation factor) could be targeted to various sites on the ribosome surface. This mechanism of action would be possible for all types of "A" protein complexes.

A summary of the structural and functional features of the L10e and L12e proteins is illustrated in Fig. 4. Previous models of the interkingdom relationships of the structural and functional features of the L12e proteins have considered the evolution of the L12e genes (and proteins) in isolation (Amons et al. 1979; Jue et al. 1980; Yaguchi et al. 1980; Lin et al. 1982; Matheson 1985; Otaka et al. 1985; Wittmann-Liebold 1985; Liljas et al. 1986). We have considered the interkingdom alignments, structural and functional features, and evolution (see below) of the L12e genes (and proteins) in concert with those of the related L10e genes (and proteins). The most likely evolutionary events are those that preserve the structure and function of the L10e–L12e complex. We believe the alignments presented for the L10e, L12e, and interprotein relationship between the L10e and L12e proteins permit preservation of the structure and function of the complex and resolve some of the enigmatic features of previous models presented for evolution of the L12e protein.

The Evolution of the L10e and L12e Genes and Proteins

Functional, structural, and sequence information indicates that the genes encoding the contemporary L10e and L12e proteins were derived from single ancestral genes present in the common primordial ancestor. During evolution these genes have undergone numerous alterations and rearrangements within the progenote and in the independent lines of descent to produce a variety of products (Fig. 4).

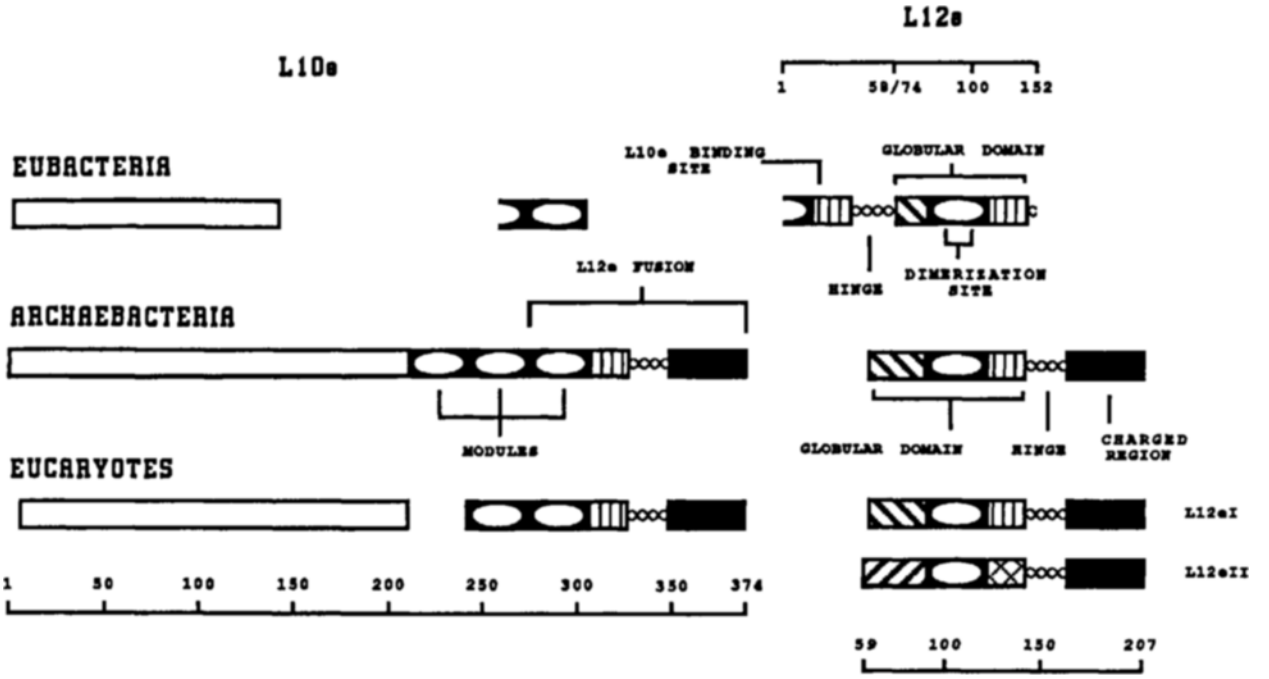

Fig. 4. Structural summary of the L10e and L12e proteins of eubacteria, archaebacteria, and eucaryotes. The structural features of the L10e and L12e proteins from the eubacteria (top), archaebacteria (center), and the eucaryotes (bottom) are illustrated with amino acid scales corresponding to the sequence scales utilized in the L10e and L12e alignments of Figs. 1 and 2. The archaebacterial L12e protein is composed of a globular domain containing a copy of the 26-residue-long module, a hinge, and the charged carboxy terminus. The archaebacterial L10e contains a fusion of three-quarters of a copy of the L12e protein and a triplication of the modular sequence present in the L12e part of the fusion. The eucaryotic proteins are very similar to their archaebacterial counterparts; there exist two types of L12e protein (i.e., L12e type I, which is similar to the archaebacterial L12e, and L12e type II, which differs in the globular domain) and the L10e protein has only two modules. The eubacterial proteins have undergone substantial alterations. The L10e protein has a large internal deletion and only one complete and one partial module, and the carboxy terminal sequences containing the hinge and highly charged regions are truncated. The L12e protein retains the globular domain with the internal module and dimerization site, but the hinge has been relocated to the amino terminal side of the globular domain. The N-terminal end, responsible for dimerization of the L12e proteins and binding to the L10e protein, is partially derived from a module. The upper amino acid scale is unique to the eubacterial L12e protein and has a fusion of position 58 with position 74. The archaebacterial and eucaryotic L12e proteins correspond with the lower L12e scale.

By recognizing common and conserved features in the proteins encoded by these genes, it is possible not only to suggest a structure for the ancestral genes but also to construct a model that integrates the hypothetical genetic and structural evolution of the L10e and L12e genes and proteins in eucaryotes, archaebacteria, and eubacteria (Fig. 5).

We suggest that initially in the progenote state there existed a single ancestral L10e protein at least 210 amino acids long that lacked the modular sequence and the highly charged C-terminus and a single ancestral L12e gene composed of an N-terminal globular domain and highly charged C-terminus separated by a hinge (i.e., structurally similar to the archaebacterial and eucaryotic L12e proteins). The highly charged C-terminal end of the L12e type proteins may have been responsible for or facilitated the binding to the L10e protein and/or the dimerization of the L12e protein. Duplication of the L12e gene [presumably to ensure the elevated stoichiometry of the L12e dimer(s) in the ribosome] and subsequent divergence produced the type I and type II genes found in contemporary eucaryotes; the type II gene was possibly lost during the formation of the L11e–L1e–L10e–L12e gene cluster in the archaebacterial and eubacterial lineages. A further duplication of the L12e type I gene provided an extra copy for the gene fusion event that created a splice between L10e and one of the copies of L12e; the fusion junction was possibly immediately preceding the conserved basic residue at position 8 of Fig. 3, where an intron occurs in one of the Sce L12e type II genes (Remacha et al. 1988). If the module contains a dimerization site as has been suggested, this would allow specific targeting of the globular domain of the L12e protein to the fusion protein. Duplication of the module resulted in the present eucaryotic state and a second module duplication produced the archaebacterial state (Figs. 4 and 5).

At this point the eubacteria diverged from the archaebacteria and eucaryotes. If the 12-amino acid partial Eco L10 β module is real, then eubacteria probably evolved from a three-module ancestor. If the Eco L10 β module is a fortuitous match, then

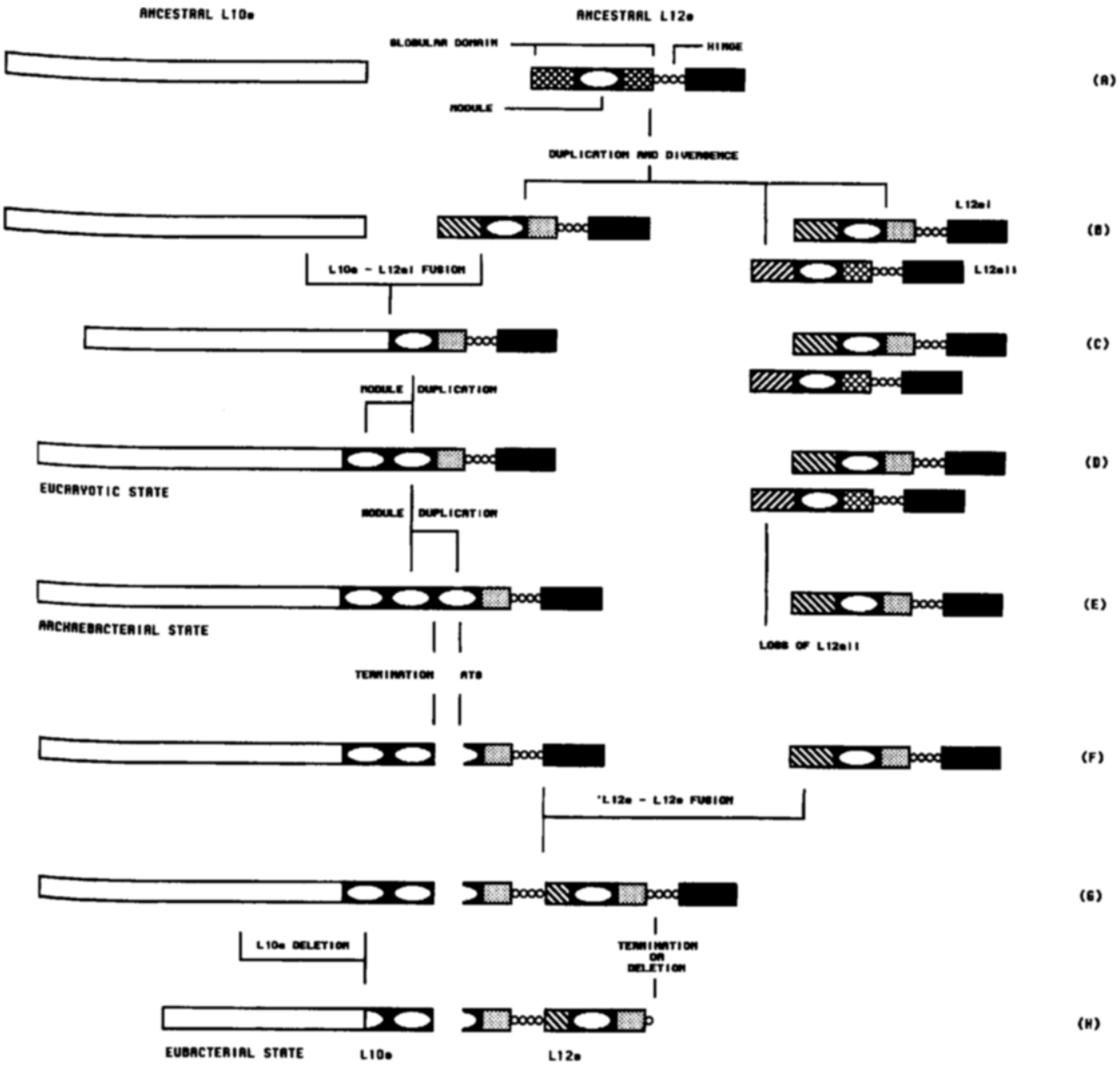

Fig. 5. A model depicting the evolution of the primordial L10e and L12e genes. Illustrated is a model to demonstrate how simple rearrangements might explain the evolutionary divergence and contemporary relationships among eubacterial, archaebacterial, and eucaryotic L10e and L12e genes. The stages and intermediates are illustrated from **A** to **H**: **A** L10e and L12e gene structures in the progenote. **B** Duplication of the L12e and divergence resulting in the L12e type I and II genes. A further duplication results in two copies of L12e type I. **C** A deletion fuses the 3′ portion of an L12e type I gene copy to the L10e gene. **D** Duplication within the L10e–L12e fusion gene of a 26-codon-long module originally from the L12e gene sequence results in the contemporary eucaryotic state. **E** A second duplication of the module within L10e and loss of the L12e type II gene results in the contemporary archaebacterial state. **F** The eubacterial state may have arisen from either the eucaryotic or archaebacterial states; for simplicity only descent from the archaebacterial state is illustrated. Within the eubacterial line, translation stop and start codons are generated within the fusion gene to produce two separate and nonoverlapping open reading frames. Part of the distal module is lost. **G** 'L12e is fused by deletion with the second copy of the ancestral L12e gene to produce a 'L12e–L12e hybrid. **H** Truncation of the 'L12e–L12e hybrid generates the contemporary eubacterial L12e. An internal deletion shortens the L10e to the modern eubacterial state.

evolution may have occurred from a two-module ancestor. The eucaryotic state of two modules may have arisen from deletion of a module from the three-module archaebacterial state. Thus it is impossible to determine the branching order of eubacteria, archaebacteria, and eucaryotes from the present data. For simplicity we describe the derivation of the eubacterial state as if arising from a three-module L10e ancestor.

Four additional steps are required to achieve the contemporary eubacterial state. First, within the L10e fusion gene a translation start site is generated

in the γ module and a translation stop is generated upstream at the 3′ end of the β module, resulting in production of a short 'L12e peptide. The partial module of this short peptide permits binding to the L10e protein through the module(s) of the L10e protein. Second, the 'L12e gene fuses to the L12e gene, removing the unique carboxy terminal end of the 'L12e protein and preserving the functional factor binding globular domain in the L12e protein. Third, there is deletion or termination of the unique carboxy terminal sequence of the 'L12e–L12e protein, leaving the present eubacterial L12e state (Figs. 4 and 5). The L12e protein binds to L10e through the partial module of its N-terminus. The modern L12e of eubacteria have ragged N-terminal ends, starting between positions −4 and +3 on Fig. 2. This may represent fine tuning of the N-terminal binding domain during the evolution of the primary eubacterial lineages. Finally, there is an internal deletion in L10e, resulting in the shortened contemporary eubacterial L10e gene (Figs. 4 and 5).

The eubacteria and the eucaryotes have clearly evolved independently of the archaebacteria. There are presently two competing phylogenies for the evolution of the archaebacteria. Woese suggested that the archaebacteria form a single phylogenetic clade: one of three urkingdoms, the archaebacteria, the eubacteria, and the eucaryotes (Woese and Fox 1977; Woese and Olsen 1986). Lake suggested a polyphyletic origin of the archaebacteria, i.e., the halobacteria and methanogens grouped with the eubacteria, and the thermoacidophiles with the eucaryotes: all five groups with urkingdom status (Lake 1988). The L10e and L12e genes are extremely ancient, their evolution occurring in a series of discrete steps over the interval from the preprogenote state, through the primary speciation event giving rise to the urkingdoms, and extending well into the main eubacterial lineages. The evolution of the L10e and L12e proteins exhibits a series of discrete alterations over the interval of contention (that is, during the primary speciation event giving rise to the urkingdoms). A discrete phylogenetic tree of the evolution of the translation apparatus may eventually be constructed over the contentious time span if during this time a sufficient number of ribosomal proteins either first appeared or share the complex alterations exhibited by the L10e and L12e proteins.

Acknowledgments. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (to ATM), the Medical Research Council of Canada (to PPD), and the Office of Naval Research of the USA (to PPD). PPD is a Fellow of the Canadian Institute for Advanced Research. LCS was partially supported by a University of British Columbia Graduate Fellowship. We thank Dr. Russell Doolittle for his comments.

References

Amons R, Pluijms W, Moller W (1979) The primary structure of ribosomal protein eL12/eL12-P from *Artemia salina* 80S ribosomes. FEBS Lett 104:85–89

Amons R, Pluijms W, Kriek T, Moller W (1982) The primary structure of protein eL12′/eL12′-P from the large subunit of *Artemia salina* ribosomes. FEBS Lett 146:143–147

Bartsch M, Kimura M, Subramanian AR (1982) Purification, primary structure, and homology relationships of a chloroplast ribosomal protein. Proc Natl Acad Sci USA 79:6871–6875

Bayley ST, Morton RA (1978) Recent developments in the molecular biology of extremely halophilic bacteria. CRC Crit Rev Microbiol 6:151–205

Beltrame M, Bianchi ME (1987) Sequence of the cDNA for one acidic ribosomal protein of *Schizosaccharomyces pombe.* Nucleic Acids Res 15:9089

Burma DP, Srivastava S, Srivastava AK, Mahanti S, Dash D (1986) Conformational change of 50S ribosomes during protein synthesis. In: Hardesty B, Kramer G (eds) Structure, function, and genetics of ribosomes. Springer-Verlag, New York, pp 438–453

Cowgill, CA, Nichols BG, Kenny JW, Butler P, Bradbury EM, Traut RR (1984) Mobile domains in ribosomes revealed by proton nuclear magnetic resonance. J. Biol Chem 259:15257–15263

Darnell J, Doolittle WF (1986) Speculations on the early course of evolution. Proc Natl Acad Sci USA 83:1271–1275

Dayhoff MO (1978) Atlas of protein sequence and structure, ed 5. MacGregor and Warner, Silver Spring MD, pp 345–352

Downing WL, Dennis PP (1987) Transcription products from the *rpl*KAJL–*rpo*BC gene cluster. J. Mol Biol 194:609–620

El-Baradi TTAL, de Regt CHF, Einerhand SWC, Teixido J, Planta RJ, Ballesta JPG, Raue HA (1987) Ribosomal proteins EL11 from *Escherichia coli* and L15 from *Saccharomyces cerevisiae* bind to the same site in both yeast 26S and mouse 28S rRNA. J Mol Biol 195:909–917

Falkenberg P, Yaguchi M, Roy C, Zuker M, Matheson AT (1986) The primary structure of the ribosomal A-protein (L12) from the moderate halophile NRCC 41227. Biochem Cell Biol 64: 675–680

Garland WG, Louie KA, Matheson AT, Liljas A (1987) The complete amino acid sequence of the ribosomal 'A' protein (L12) from *Bacillus stearothermophilus.* FEBS Lett 220: 43–46

Gudkov AT, Behlke J (1978) The N-terminal sequence of L7/L12 is responsible for its dimerization. Eur J Biochem 90: 309–312

Gudkov AT, Gongadze GM (1984) The L7/L12 proteins change their conformation upon interaction of EF-G with ribosomes. FEBS Lett 176:32–36

Gudkov AT, Tumanova LG, Venyaminov SY, Khechinashvili NN (1978) Stoichiometry and properties of the complex between ribosomal proteins L7 and L10 in solution. FEBS Lett 93:215–218

Itoh T (1981a) Primary structure of an acidic ribosomal protein from *Micrococcus lysodeikticus.* FEBS Lett 127:67–70

Itoh T (1981b) Primary structure of an acidic ribosomal protein YPA1 from *Saccharomyces cerevisiae.* Isolation and characterization of peptides and the complete amino acid sequence. Biochim Biophys Acta 671:16–24

Itoh T, Higo KI (1983) Complete amino aid sequence of an L7/L12-type ribosomal protein from *Rhodopseudomonas spheroides.* Biochim Biophys Acta 744:105–109

Itoh T, Wittmann-Liebold B (1978) The primary structure of

Bacillus subtilis acidic ribosomal protein B-L9 and its comparison with *Escherichia coli* proteins L7/L12. FEBS Lett 96: 392–394

Itoh T, Sugiyama M, Higo KI (1982) The primary structure of an acidic ribosomal protein from *Streptomyces griseus*. Biochim Biophys Acta 701:164–172

Itoh T, Kamazaki T, Sugiyama M, Otaka E (1988) Molecular cloning and sequence analysis of the ribosomal 'A' protein gene from the archaebacterium, *Halobacterium halobium*. Biochim Biophys Acta 949:110–118

Jue RA, Woodbury NW, Doolittle RF (1980) Sequence homologies among *E. coli* ribosomal proteins: evidence for evolutionarily related groupings and internal duplications. J Mol Evol 15:129–148

Kastner B, Stoffler-Meilicke M, Stoffler G (1981) Arrangement of the subunits in the ribosome of *Escherichia coli*: demonstration by immunoelectron microscopy. Proc Natl Acad Sci USA 78:6652–6656

Koteliansky VE, Domogatsky SP, Gudkov AT (1978) Dimer state of protein L7/L12 and EF-G dependent reactions on ribosomes. Eur J Biochem 90:319–323

Lake JA (1988) Origin of the eucaryotic nucleus determined by rate-invariant analysis of rRNA sequences. Nature 331:184–186

Leijonmarck M, Liljas A (1987) Structure of the C-terminal domain of the ribosomal protein L7/L12 from *Escherichia coli* at 1.7 Å. J Mol Biol 195:555–580

Leijonmarck M, Pettersson I, Liljas A (1981) Structural studies on the protein L7/L12 from *E. coli* ribosomes. In: Balaban M, Sussmann JL, Traub W, Yonath A (eds) Structural aspects of recognition and assembly in biological macromolecules. ISS, Rehovot, pp 761–777

Liljas A (1982) Structural studies of ribosomes. Prog Biophys Mol Biol 40:161–228

Liljas A, Thirup S, Matheson AT (1986) Evolutionary aspects of ribosome-factor interactions. Chem Scr 26B:109–119

Lin A, Wittmann-Liebold B, McNally J, Wool IG (1982) The primary structure of the acidic phosphoprotein P2 from rat liver 60S ribosomal subunits. J Biol Chem 257:9189–9197

Lindahl L, Zengel J (1986) Ribosomal genes in *Escherichia coli*. Annu Rev Genet 20:297–326

Lipman DJ, Pearson WR (1985) Rapid and sensitive protein similarity searches. Science 227:1435–1441

Marquis DM, Fahnestock SR, Henderson E, Woo D, Schwinge S, Clarke MW, Lake JA (1981) The L7/L12 stalk, a conserved feature of the prokaryotic ribosome, is attached to the large subunit through its N terminus. J Mol Biol 150:121–132

Matheson AT (1985) Ribosomes of archaebacteria. In: Woese CR, Wolfe RS (eds) The bacteria, vol 8. The Archaebacteria. Academic Press, New York, pp 761–777

Matheson AT, Yaguchi M, Balch WE, Wolfe RS (1979) Sequence homologies in the N-terminal region of the ribosomal 'A' proteins from *Methanobacterium thermoautotrophicum* and *Halobacterium cutirubrum*. Biochim Biophys Acta 626: 162–169

Matheson AT, Louie KA, Tak BD, Zuker M (1987) The primary structure of the ribosomal A-protein (L12) from the halophilic eubacterium *Haloanaerobium praevalens*. Biochemie 69:1013–1020

Matheson AT, Louie KA, Bock A (1988) The complete amino acid sequence of the ribosomal A protein (L12) from the archaebacterium *Sulfolobus acidocaldarius*. FEBS Lett 231: 331–335

Mitsui K, Tsurugi K (1988a) cDNA and deduced amino acid sequence of acidic ribosomal protein A0 from *Saccharomyces cerevisiae*. Nucleic Acids Res 16:3573

Mitsui K, Tsurugi K (1988b) cDNA and deduced amino acid sequence of acidic ribosomal protein A1 from *Saccharomyces cerevisiae*. Nucleic Acids Res 16:3574

Mitsui K, Tsurugi K (1988c) cDNA and deduced amino acid sequence of acidic ribosomal protein A2 from *Saccharomyces cerevisiae*. Nucleic Acids Res 16:3575

Moller W, Schrier PI, Maasen JA, Zantema A, Schop E, Reinalda H, Cremers AFM, Mellema JE (1983) Ribosomal proteins L7/L12 of *Escherichia coli*. Localization and possible molecular mechanism in translation. J Mol Biol 163:553–573

Osterberg R, Sjoberg B, Liljas A, Pettersson I, Kurland CJ (1976) Small angle X-ray scattering and crosslinking study of the proteins L7/L12 from *Escherichia coli* ribosomes. FEBS Lett 66:48–51

Osterberg R, Sjoberg B, Liljas A, Pettersson I, Kurland CJ (1977) Small angle X-ray scattering study of the protein complex of L7/L12 and L10 from *Escherichia coli* ribosomes. FEBS Lett 73:22–24

Otaka E, Ooi T, Kumazaki T, Itoh T (1985) Examination of protein sequence homologies: II. Six *Escherichia coli* L7/L12 type ribosomal 'A' protein sequences from eukaryotes and metabacteria, contrasted with those from prokaryotes. J Mol Evol 22:342–350

Pettersson I (1979) Studies on the RNA and protein binding sites of the *E. coli* ribosomal protein L10. Nucleic Acids Res 6:2637–2646

Pettersson I, Liljas A (1979) The stoichiometry and reconstitution of a stable protein complex from *Escherichia coli* ribosomes. FEBS Lett 98:39–144

Post LE, Strycharz GD, Nomura M, Lewis H, Dennis PP (1979) Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit in *Escherichia coli*. Proc Natl Acad Sci USA 76:1697–1701

Qian S, Zhang J, Kay MA, Jacobs-Lorena M (1987) Structural analysis of the *Drosophila* rpA1 gene, a member of the eucaryotic 'A' type ribosomal protein family. Nucleic Acids Res 15:987–1003

Ramirez C, Shimmin LC, Newton CH, Matheson AT, Dennis PP (1989) Structure and evolution of the L11, L1, L10 and L12 equivalent ribosomal proteins in eubacteria, archaebacteria and eucaryotes. Can J Microbiol 35:234–244

Remacha M, Saenz-Robles MT, Vilella MD, Ballesta JPG (1988) Independent genes coding for three acidic proteins of the large ribosomal subunit from *Saccharomyces cerevisiae*. J Biol Chem 263:9094–9101

Rich BE, Steitz JA (1987) Human acidic ribosomal phosphoproteins P0, P1, and P2: analysis of cDNA clones, *in vitro* synthesis, and assembly. Mol Cell Biol 7:4065–4074

Schmidt FJ, Thompson J, Lee K, Diik J, Cundliffe E (1981) The binding site for ribosomal protein L11 within 23S ribosomal RNA of *Escherichia coli*. J Biol Chem 256:12301–12305

Shimmin LC, Dennis PP (1989) Characterization of the L11, L1, L10 and L12 equivalent ribosomal protein gene cluster of the halophilic archaebacterium *Halobacterium cutirubrum*. EMBO J 8:1225–1235

Shimmin LC, Newton CH, Ramirez C, Yee J, Downing WL, Louie A, Matheson AT, Dennis PP (1989) Organization of genes encoding the L11, L1, L10 and L12 equivalent ribosomal proteins in eubacteria, archaebacteria and eucaryotes. Can J Microbiol 35:164–170

Strobel O, Kopke AKE, Kamp RM, Bock A, Wittmann-Liebold B (1988) Primary structure of the archaebacterial *Methanococcus vannielii* ribosomal protein L12. J Biol Chem 263: 6538–6546

Strycharz WA, Nomura A, Lake JA (1978) Ribosomal proteins

L7/L12 localized at a single region of the large subunit by immune electron microscopy. J Mol Biol 126:123–140

Terhorst C, Moller W, Laursen R, Wittmann-Liebold B (1973) The primary structure of an acidic protein from 50-S ribosomes of *Escherichia coli* which is involved in GTP hydrolysis dependent on elongation factors G and T. Eur J Biochem 34: 138–152

Traut RR, Tewari DS, Sommer A, Gavino GR, Olson HM, Glitz DG (1986) Protein topography of ribosomal functional domains: effects of monoclonal antibodies to different epitopes in *Escherichia coli* protein L7/L12 on ribosome function and structure. In: Hardesty B, Kramer G (eds) Structure, function, and genetics of ribosomes. Springer-Verlag, New York, pp 286–308

Tritton TR (1978) Spin-labelled ribosomes. Biochemistry 17: 3959–3964

Wigboldus JD (1987) cDNA and deduced amino acid sequence of *Drosophila* rp21C, another 'A'-type ribosomal protein. Nucleic Acids Res 15:10064

Wittmann-Liebold B (1986) Ribosomal proteins: their structure and evolution. In: Hardesty B, Kramer G (eds) Structure, function, and genetics of ribosomes. Springer-Verlag, New York, pp 326–361

Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA 74:5088–5090

Woese CR, Olsen GJ (1986) Archaebacterial phylogeny: perspectives on the urkingdoms. System Appl Microbiol 7:161–177

Yaguchi M, Matheson AT, Visentin LP, Zuker M (1980) Molecular evolution of the alanine-rich, acidic ribosomal A protein. In: Osawa S, Ozeki H, Uchida H, Yura Y (eds) Genetics and evolution of RNA polymerase, tRNA and ribosomes. University of Tokyo Press, Tokyo, pp 585–599

Received January 7, 1989/Revised and accepted March 6, 1989