Supervisor: Dr. Alastair T. Matheson

Abstract

Six ribosomal protein genes from the sulfur dependent extreme thermophilic archaebacterium *Sulfolobus solfataricus* were cloned and sequenced. Four of these genes code for proteins that are equivalent to ribosomal proteins L11, L1, L10 and L12 in *Escherichia coli.* The other two genes code for proteins that have no equivalent in the eubacteria. The product of one of these genes was found to be equivalent to ribosomal proteins L46 from yeast (Leer *et al.* 1985a) and L39 from rat liver (Lin *et al.* 1984), while the product of the other gene shows no sequence similarity to any of the ribosomal proteins present in the data base. In *Sulfolobus*, the genes that code for ribosomal proteins L11, L1, L10 and L12 are organized in the same order as in *Escherichia coli* , that is 5' L11, L1, L10, L12 3'. The major transcript from this gene cluster was found to be a 2.5 Kb mRNA that contains the four genes. A less abundant transcript containing only the L10 and L12 gene was also detected. Upstream of the transcription initiation sites, sequences that match the consensus sequence for archaebacterial promoters (TTTAT/AA) were found. Transcription termination sites were located within or after pyrimidine rich regions. Three of the ribosomal protein genes start with unusual initiation codons, GTG in the case of the L1 and L10 genes and TTG in the case of the L11 gene. Putative Shine Dalgarno sequences, complementary to the 3' end of *Sulfolobus* 16S rRNA, were detected in the region surrounding the initiation codon. In some cases (L1 and L10 genes), the initiation codon was found to be part of this sequence. Sequence comparison of the ribosomal proteins from

Sulfolobus with those from other organisms, revealed that the *Sulfolobus* sequences are closer to those from other archaebacteria, thus supporting the existence of the archaebacterial kingdom. Comparison of the sequences of the L10 and L12 proteins from the three kingdoms revealed that the archaebacterial sequences are closer to the eukaryotes.

Dr. Alastair T. Matheson

Dr. Paul J. Romaniuk

Dr. Terry W. Pearson

Dr. Alexander McAuley

Dr. Verena J. Tunnicliffe

Dr. Robert R. Traut

Table of Contents

Abstract....ii
Table of Contents....iv
List of Tables....vii
List of Figures....viii
List of Abbreviations....xi
Acknowledgements....xv
Dedication....xvi
Introduction....1
- Main Archaebacterial Groups....3
- Phylogenetic Relationships among the Different Archaebacterial Groups....4
- Molecular Biology....8
- Transcription....12
  - RNA Polymerase....12
  - Promoters....14
  - Terminators....15
  - Messenger RNA (mRNA) and Translation Signals....16
- Ribosomes:Structure, Function, and Genetics....17
  - Structure of the Ribosome....17
  - Ribosomal RNA....25
    - 5S rRNA....25
    - 16S rRNA....26
    - 23S rRNA....27
  - Ribosomal Proteins....28
  - Protein Synthesis....31
    - Initiation....32
    - Elongation....32
    - Termination....39
  - Organization of rRNA and Ribosomal Protein Genes....39
    - rRNA Genes....39
    - Ribosomal Protein Genes....41
- The L7/L12 Domain and the Purpose of this Dissertation....44

Materials and Methods....57
- Growth of *Sulfolobus solfataricus*....57
- Other Archaebacterial Cells....58
- Bacterial Strains and Vectors....58
- Growth of the Different Bacterial Strains and Vectors....59
  - Q358, Q359, EMBL 3 and Recombinant Phages....60
  - JM103, JM105, M13mp7 and pUC 18....60
- DNA Extraction....60
- RNA Extraction....61
- Restriction Endonuclease Digestions....61
- Synthesis of Probes and Primers....62
- Labeling of Probes....63
- Dot Blots....63

Southern Blots 65
Unblots 66
Northern Blots 66
Construction of a Genomic Library of *Sulfolobus solfataricus* in EMBL 3 67
Screening of the *Sulfolobus* Library 68
Subcloning of a 1.1 Kb Pst I-Pst I Fragment into M13mp7 69
Primer Extension 74
S1 Mapping 75
Molecular Weight Markers 75
Probes for S1 Mapping and Northern Blots 75
S1 Mapping Protocol 77
Computer Programs 77
Results and Discussion 79
The Sso L12 Gene Oligonucleotide Probe 79
Conditions under which the Sso L12 Probe binds Stringently to *Sulfolobus* DNA 82
Partial Restriction Map of the Region in which the Sso L12 Gene is Located 86
Construction of a Genomic Library of *Sulfolobus solfataricus* and Isolation of the Sso L12 Gene 86
Binding of the Sso L12 Probe Mixes to DNA from other Archaebacteria 89
Subcloning and Sequencing of a 1.1 Kb Pst I - Pst I fragment that hybridizes to the 17A Probe 89
Subcloning of a 6.9 Kb Eco RI - Bam HI fragment into pUC 18 97
Mapping of the Position of the 1.1 Kb Pst I - Pst I Fragment inside the 6.9 Kb Eco RI - Bam HI Fragment 100
Sequencing of the 6.9 Kb Eco RI - Bam HI Fragment 103
The Sso L12 Gene and the Sso L12 Protein 108
Sequence Conservation between the L10 and L12 Proteins 130
Evolution of the L10 and L12 Genes and Proteins 133
The Sso L1 Gene and the Sso L1 protein 136
The Sso L11 Gene and the Sso L11 Protein 143
The Sso L46 Gene and the Sso L46 Protein 151
The Sso LX Gene and its Product 156
The Sso "Docking" Protein Gene and Its Product 162
Open Reading Frames 175
Organization and Transcription of the Sso L11, L1, L10 and L12 Ribosomal Protein Genes 181
Codon Utilization 197
Translation Signals 200
Phylogenetic Implications 202
Evolution of the Ribosome 203
Evolution of the Genetic Organization in the Three Kingdoms 206
Conclusions 208

Bibliography 212
Appendix..251
Sequence of the Primers 251
Sequence of the 6.9 Eco RI-Bam HI Fragment..252

List of Tables

Table 1 Some examples of archaebacterial genes that have been cloned and sequenced.....10
Table 2 Ribosomal components in the three kingdoms.....18
Table 3 Bacterial strains.....59
Table 4 Vectors.....59
Table 5 Restriction endonucleases.....62
Table 6 Different washing conditions for the dot blot hybridization filters.....64
Table 7 Stringent condition washes of the plaque hybridization filters.....69
Table 8 Sequencing mixes.....72
Table 9 Primers used to generate 5' labeled probes for S1 mapping and Northern blots.....76
Table 10 Codon usage in the Sso L12 gene.....110
Table 11 Predicted Amino acid composition of the Sso L12 protein.....111
Table 12 Sequence identity between the Sso L12 protein and its archaebacterial and eukaryotic counterparts.....115
Table 13 Codon utilization in the Sso L10 gene.....123
Table 14 Predicted Amino acid composition of the Sso L10 protein.....123
Table 15 Sequence identity among the L10 proteins.....124
Table 16 Codon utilization in the Sso L1 gene.....139
Table 17 Predicted Amino acid composition of the Sso L1 protein.....140
Table 18 Sequence identity between the Sso L1 protein and its archaebacterial and eubacterial counterparts.....143
Table 19 Codon utilization in the Sso L11 gene.....146
Table 20 Predicted Amino acid composition of the Sso L11 protein.....146
Table 21 Sequence identity between the Sso L11 protein and its archaebacterial and eubacterial counterparts.....148
Table 22 Codon utilization in the Sso L46 gene.....153
Table 23 Predicted Amino acid composition of the Sso L46 protein.....153
Table 24 Codon utilization in the Sso LX gene.....158
Table 25 Predicted Amino acid composition of the Sso LX protein.....158
Table 26 Codon utilization in the 5' segment (first 546 nucleotides) of the *Sulfolobus ala S* gene.....161
Table 27 Codon utilization in the Sso "docking" gene.....167
Table 28 Predicted Amino acid composition of the Sso "docking" protein.....167
Table 29 Sequence identity between the Sso "docking" protein and its eukaryotic and eubacterial counterparts.....171
Table 30 Sequence identity between the Sso "docking" protein and the 54 kDa SRP protein and 48 kDa *E.coli* protein.....172
Table 31 Codon utilization in the *Sulfolobus* ribosomal protein genes.....199
Table 32 Base composition in the wobble position in *Sulfolobus* ribosomal protein genes.....199
Table 33 Unusual initiation codons in *Sulfolobus solfataricus*.....200
Table 34 Shine-Dalgarno sequences.....201

List of Figures

Figure 1. Archaebacterial phylogenetic tree. 5
Figure 2. Models for the small and large ribosomal subunits from the three kingdoms. 19
Figure 3. Computer-Imaging models of the 30S and 50S subunits from *Escherichia coli*. 21
Figure 4. Computer-imaging models of the 50S subunit from *Sulfolobus solfataricus* and *Escherichia coli*. 23
Figure 5. Protein synthesis in the eubacteria. 33
Figure 6. Protein synthesis in the eukaryotes. 35
Figure 7. Structure of the L7/L12 ribosomal protein from *Escherichia coli*. 47
Figure 8. The L7/L12 domain. 51
Figure 9. The Sso L12 probe. 81
Figure 10. Dot blot hybridizations of the 17T/C and 17A Sso L12 gene probes to *Sulfolobus* DNA. 84
Figure 11. Hybridization of the 17A probe to different restriction endonuclease digests of *Sulfolobus* DNA. 85
Figure 12. Partial restriction map of the region containing the Sso L12 gene. 86
Figure 13. Construction of a *Sulfolobus* genomic library and isolation of a clone carrying the Sso L12 gene. 88
Figure 14. Hybridization of the 17A probe to DNA from other archaebacteria and clone E3CR-J. 90
Figure 15. Life cycle of M13 and structure of M13mp7. 92
Figure 16. Determination of the orientation of the insert in clones M13CR-3 and M13CR-8. 94
Figure 17. Sequencing strategy for the 1.1 Kb Pst I - Pst I fragment. 95
Figure 18. Example of a sequencing gel. 96
Figure 19. Structure of pUC 18 and its multiple cloning site. 98
Figure 20. Identification of clone p18CR-9. 99
Figure 21. Location of the 1.1 Kb Pst I- Pst I fragment within the 6.9 Kb Eco RI - Bam HI fragment. 101
Figure 22. Sequencing strategy and organization of the genes present in the 6.9 Kb Eco RI-Bam HI fragment. 104
Figure 23. Construction of the deletion plasmids according to the method of Lin *et al.* (1985). 105
Figure 24. Sequence of the Sso L12 gene. 109
Figure 25. Alignment of archaebacterial L12 proteins. 112
Figure 26. Sequence alignment of the Sso L12 protein and the eukaryotic P2 proteins. 113
Figure 27. Alignment of the Sso L12 protein with the eukaryotic P1 acidic proteins. 114

Figure 28. Different models for the alignment of the eubacterial and archaebacterial-eukaryotic L12 proteins. 117
Figure 29. Sequence of the Sso L10 gene. 121
Figure 30. Sequence alignment of the L10 proteins from the three kingdoms. 126
Figure 31. Sequence alignment of the modules in the L10 proteins. 129
Figure 32. Alignment of the common regions between the L10 and L12 proteins from the archaebacteria and the eukaryotes. 132
Figure 33. Identical regions present in the Sso L10 and Sso L12 genes and proteins. 133
Figure 34. Possible model for the evolution of the L10 and L12 genes in the three kingdoms. 134
Figure 35. Sequence of the Sso L1 gene. 137
Figure 36. Sequence alignment of the Sso L1 protein with its archaebacterial and eubacterial counterparts. 141
Figure 37. Sequence of the Sso L11 gene. 145
Figure 38. Sequence alignment of the Sso L11 protein and its archaebacterial and eubacterial counterparts. 147
Figure 39. Conserved region in the L11 and L1 ribosomal proteins. 151
Figure 40. Sequence of the Sso L46 gene. 152
Figure 41. Sequence alignment of the Sso L46 protein with its eukaryotic counterparts. 155
Figure 42. Common regions in the L5 family of ribosomal proteins. 155
Figure 43. Sequence of the Sso LX gene. 157
Figure 44. Sequence of the *Sulfolobus ala S* gene. 160
Figure 45. Functional domains in the *Escherichia coli* alanine tRNA synthetase. 162
Figure 46. Sequence alignment of the N-terminal region of the *Sulfolobus* alanine tRNA synthetase with the corresponding region from the *Escherichia coli* alanine tRNA synthetase. 163
Figure 47. Sequence of the gene coding for the equivalent in *Sulfolobus* of the α subunit of the docking protein. 165
Figure 48. Sequence alignment of the Sso "docking" protein and its eukaryotic and eubacterial counterparts, as well as the members of the "54 kDa family" of proteins. 169
Figure 49. Region that shows sequence similarity between the "docking protein family" and the "54 kDa protein family". 172
Figure 50. The GTP binding site in the members of the "docking protein family" and the "54 kDa protein family". 174
Figure 51. Sequence of open reading frame 1. 175
Figure 52. Sequence of open reading frame 2. 176
Figure 53. Sequence of open reading frame 3. 177
Figure 54. Sequence of open reading frame 4. 178
Figure 55. Sequence of open reading frame 5. 179
Figure 56. Sequence of open reading frame 6. 180
Figure 57. Sequence of open reading frame 7. 181

Figure 58. Overlapping stop/start codons in the ribosomal protein genes of *Sulfolobus solfataricus.* 183
Figure 59. Position of the probes used for S1 mapping, primer extension, and Northern blot hybridization. 186
Figure 60. Northern blot hybridizations. 187
Figure 61. Primer extension and nuclease protection experiments to determine the 5' end of the 2.5 Kb transcript........ 188
Figure 62. The L11 promoter region. 190
Figure 63. Nuclease protection experiments to determine the 3' end of the 2.5 Kb transcript 191
Figure 64. Transcription termination sites after the L12 gene. 192
Figure 65. Nuclease protection experiment to determine the 5' end of a 1.4 Kb transcript containing the L10 and L12 genes. 193
Figure 66. The L10 promoter. 194
Figure 67. Nuclease protection and primer extension experiments to determine the transcription initiation site of the *ala S* gene. 195
Figure 68. The *ala S* promoter. 196
Figure 69. Transcription of the L11-L1-L10-L12 genes in *Sulfolobus solfataricus.* 197

List of Abbreviations

aa-tRNA: aminoacyl -tRNA

aEF: archaebacterial elongation factor

amino acids:

A: alanine
C: cysteine
D: aspartic acid
E: glutamic acid
F: phenylalanine
G: glycine
H: histidine
I: isoleucine
K: lysine
L: leucine
M: methionine
N: asparagine
P: proline
Q: glutamine
R: arginine
S: serine
T: threonine
V: valine
W: tryptophan
Y: tyrosine

Asa: *Artemia salina*

ATP: adenosine 5' triphosphate

bp: base pairs

bases:

A: adenine
G: guanine
C: cytosine
T: thymine

Bst: *Bacillus stearothermophilus*

Cfa: *Canis familiaris*

cpm: counts per minute

DEPC: diethyl pyrocarbonate

deoxynucleotides:

dATP: deoxyadenosine 5' triphosphate

dCTP: deoxycytidine 5' triphosphate

dGTP: deoxyguanosine 5' triphosphate

dTTP: deoxythymidine 5' triphosphate

dideoxynucleotides:

ddATP: dideoxyadenosine 5' triphosphate

ddCTP: dideoxycytidine 5' triphosphate

ddGTP: dideoxyguanosine 5' triphosphate

ddTTP: dideoxythymidine 5' triphosphate

Dme: *Drosophila melanogaster*

DNA: deoxyribonucleic acid

Eco: *Escherichia coli*

EDTA: ethylenediaminetetraacetic acid

EF: elongation factor

eIF: eukaryotic initiation factor

Gdo: *Gallus domesticus*

GDP: guanosine 5' diphosphate

GTP: guanosine 5' triphosphate

Hcu: *Halobacterium cutirubrum*

Hha: *Halobacterium halobium*

Hsa: *Homo sapiens*

IF: initiation factor

IPTG: isopropylthiogalactoside

Kb: kilobase pairs

kDa: kilo Daltons

Mmu: *Mus musculus*

Mva: *Methanococcus vannielii*

O.D.: optical density

ORF: open reading frame

PIPES: piperazine-N,N'-bis [2-ethane-sulfonic acid] disodium salt

Pvu: *Proteus vulgaris*

RF: replicative form

RF-1: release factor 1

RF-2: release factor 2

Rno: *Rattus norvegicus*

rpm: revolutions per minute

RNA: ribonucleic acid

mRNA: messenger RNA

rRNA: ribosomal RNA

tRNA: transfer RNA

r.t.: room temperature

S: Svedberg sedimentation unit

Sac: *Sulfolobus acidocaldarius*

Sce: *Saccharomyces cerevisiae*

SDS: sodium dodecylsulfate

Sma: *Serratia marcescens*

Spo: *Schizosaccharomyces pombe*

SRP: signal recognition particle

SSC: sodium chloride-sodium citrate buffer (1x SSC: 0.15 M NaCl, 0.015 M sodium citrate)

Sso: *Sulfolobus solfataricus*

TAE: Tris-acetate EDTA buffer (0.04M Tris-acetate, 0.002 M Na_2EDTA, pH 8)

TBE: Tris-borate-EDTA buffer (0.089 M Tris, 0.089 M boric acid, 0.008M Na_2EDTA, pH 8)

Tris: Tris-(hydroxymethyl)-aminomethane

U.V.: ultraviolet light

Xgal: 5-dibromo 4-chloro 3-indolylgalactoside

Acknowledgements

I wish to thank my supervisor, Dr. A.T. Matheson for all his support and guidance during the course of this work. Special thanks to Lawrence Shimmin, who taught me all the basic secrets of molecular biology and to Dr. P. Dennis for allowing me to spend some time in his laboratory. I would also like to thank all the members of my committee, Dr. Romaniuk, Dr. Pearson, Dr. McAuley, Dr. Tunnicliffe and Dr. Traut for reading this dissertation and offering helpful suggestions. Many thanks to Dr. Romaniuk for synthesizing the primers used in this work and to Dr. Zillig and Dr. Pace for providing me with archaebacterial cells. The technical help of A. Louie, S. Kielland and P. Leggatt is also aknowledged. Finally, I would like to thank all the members of the Biochemistry department, but in particular Isabel, Rafael, Sonia, Margaret, Katharina, George, Peter, Kenny, Eric, Gail and DeCheng for making my stay in Victoria a memorable experience.

Para mi mamá y las dos Cristinas, con todo cariño,
por todo su apoyo y afecto durante todos estos años.

Para Toño, que me inició en el estudio
del origen y evolución de la vida.

Para Tere e Isabel

For Lynn and Kenny who introduced
me to the wonders of the bacterial world

A la memoria de Nina y Sandro

Introduction

The translation of messenger RNA into protein is a complex process that takes place on the ribosome. It represents the last step in the flow of information from DNA to RNA to protein. Since translation is the step that links the genotype with the phenotype, the evolution of the translational apparatus is closely tied to the origin of contemporary cells (Woese 1980). In order to understand the evolution of the translational apparatus, and gain further insights into the origin of modern cells, we first need to obtain a better understanding of the structure and function of the ribosome in the three primary lines of descent: the eubacteria, the archaebacteria and the eukaryotes.

Eubacterial ribosomes (*i.e.* ribosomes from *Escherichia coli*) have been extensively studied since the early 1960s, and the sequences of all their components have been determined (for reviews see Noller 1984, Giri *et al.* 1984). Our knowledge of the structure of the eukaryotic ribosome is not as complete, but many of the components of the eukaryotic ribosome have already been sequenced (for reviews, see Planta *et al.* 1986, Warner *et al.* 1986, Warner, 1989, Wool 1986). However, at the time the studies described in this dissertation were started, most of the information available regarding the structure of the components of the archaebacterial ribosome was limited to the structure of the rRNA and the N-terminal sequences of some ribosomal proteins (for reviews, see Fox 1985, Matheson 1985). For this reason, one of the objectives of the research described in this thesis was to determine the complete sequence of the ribosomal proteins that form part of the so called L12 domain in the archaebacterial ribosome. This domain is important since it is the site of interaction of the extrinsic factors (initiation, elongation and

termination factors) on the ribosome during protein synthesis, and has been extensively studied in the eubacteria and the eukaryotes (Heimark *et al.* 1976, Hamel *et al.* 1972, Girshovich *et al.* 1981, Möller *et al.* 1983, Rychlik *et al.* 1983, Sköld 1983, Traut *et al.* 1986, Moazed *et al.* 1988).

The genes that code for the different ribosomal components are also of special interest. Since they are coordinately expressed during the biogenesis of the ribosome, they are excellent models in which to study the structure, organization and expression of genes in the three kingdoms.

The organization of ribosomal protein genes is very different in the eubacteria and the eukaryotes. In the eubacteria, ribosomal protein genes are arranged in operons (for a review, see Nomura *et al.* 1984) while in the eukaryotes they are dispersed through the genome and are transcribed as single units (Mager 1988, Planta *et al.* 1986, Warner 1989). For this reason, it was of interest to determine the way in which these genes are organized in the archaebacteria. The archaebacterium *Sulfolobus solfataricus* was selected for these studies because it belongs to the sulfur-dependent thermophilic branch of the archaebacteria; a branch that is thought to be closest to the eukaryotes (Zillig 1987, Lake 1988). Furthermore, since *Sulfolobus* is an extreme thermophile (with an optimum growth temperature of 85°C), it was hoped that these studies would provide further insight into the changes in protein structure that increase their thermostability and allow these organisms to live at high temperatures.

Since our knowledge of the structure of archaebacterial ribosomes and the molecular biology of the archaebacteria has increased dramatically in the last five years, a brief review of the current knowledge of these aspects of archaebacterial biology will be presented.

Main Archaebacterial Groups

Comparative analysis of partial sequences of small subunit ribosomal RNA (rRNA) led, in the late 1970s, to the discovery that archaebacteria represent a third evolutionary line of descent different from the eubacteria (*i.e.* the true bacteria) and the eukaryotes (Woese and Fox 1977a, b). Characterization of the different members of this group has revealed that archaebacteria comprise three main phenotypes: the methanogens, the extreme halophiles and the sulfur-dependent extreme thermophiles (Woese and Wolfe 1985, Woese 1987).

The methanogens are strict anaerobes that produce energy by reducing CO_2 to CH_4. They are found in a variety of anaerobic habitats including the digestive tracts of animals and man, freshwater and marine sediments, anaerobic waste digesters and even geothermal springs and deep-sea hydrothermal vents (Whitman 1985, Jones *et al.* 1987).

The extreme halophiles are aerobic organisms that live in saline environments with salt concentrations ranging from 2.5 M to 5 M (Kushner 1985). Some extreme halophiles, like *Natrobacterium* and *Natronococcus*, live under very alkaline conditions (pH 9.5) (Tindall *et al.* 1984). Halobacteria have been isolated from salt lakes like the Great Salt Lake and the Wadi Natrun lake, the Dead Sea, salterns and spoiled salted foods.

The sulfur-dependent extreme thermophiles are either aerobic or anaerobic organisms that obtain energy from the oxidation or the reduction of sulfur, or require elemental sulfur for anabolic reactions (Stetter *et al.* 1986). They live at temperatures that range from 60°C to 110°C with optimum growth at 80°C to 90°C (Stetter and Zillig 1985, Stetter *et al.* 1986). *Pyrodictium occultum*, a member of this group, is the most thermophilic bacterium described to date. It

can grow at temperatures up to 110°C with an optimum at 105°C (Stetter *et al.* 1983, Stetter 1986). The pH requirement for growth varies among the different members of this group; some can grow at a neutral pH while others, like *Sulfolobus*, grow at a pH as low as 2 (Stetter and Zillig 1985, Stetter 1986).

Sulfur-dependent extreme thermophiles have been isolated from continental solfataric springs and mud holes; and submarine volcanic areas like hydrothermal vents and geothermally heated sea sediments. They are also found in man-made habitats such as the boiling outflows of geothermal power plants in Italy and Iceland (Stetter *et al.* 1986, Stetter 1986).

Recently, Stetter *et al.* (1987) isolated an organism from the marine hot sediments near Vulcano and Stufe di Nerone, Italy, that seems to represent a novel archaebacterial phenotype. This organism, tentatively named *Archaeoglobus fulgidus*, is an extreme thermophile (optimum growth temperature 83°C) that is able to reduce sulfate as well as to produce methane. In this respect, its metabolism seems to be intermediate between the methanogens and the sulfur-dependent thermophiles.

Phylogenetic Relationships among the Different Archaebacterial Groups

Comparison of the sequences of 16S rRNA from the different archaebacterial groups has revealed that this kingdom comprises two main branches: one corresponding to the sulfur-dependent thermophiles and the other to the methanogens and halophiles (Figure 1) (Woese and Olsen 1986, Woese 1987). However, certain thermophiles, such as *Thermoplasma*, *Thermococcus* and *Archaeoglobus* seem to be more related to the methanogen-halophile branch than to the sulfur-dependent thermophilic

branch (Woese and Olsen 1986, Woese 1987, Achenbach-Richter *et al.* 1987, Achenbach-Richter *et al.* 1988).

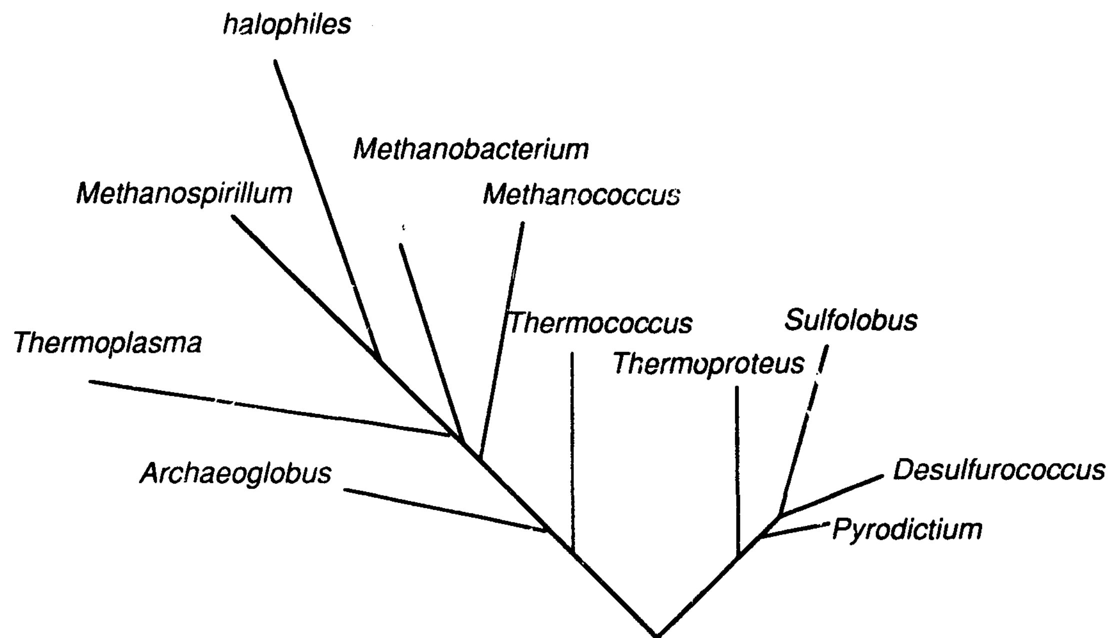


Figure 1. Archaebacterial phylogenetic tree. The tree is based on the comparison of 16S rRNA sequences. (Adapted from Woese 1987, and Achenbach-Richter *et al.* 1987)

The organisms included in the sulfur-dependent thermophilic branch share the same phenotype, while the methanogen-halophile branch is more varied (Woese and Olsen 1986). The methanogen-halophile branch includes all the archaebacterial phenotypes described until now, and contains the most rapidly evolving (*Thermoplasma acidophilum*) and the most slowly evolving (*Thermococcus celer*) lines (Woese and Olsen 1986). Within this branch, the methanogens can be divided into three groups: the *Methanobacteriales*,

Methanococcales and *Methanomicrobiales.* The extreme halophiles form a tight group that is specifically related to the *Methanomicrobiales.* These two groups are more closely related to the *Methanobacteriales* than to the *Methanococcales*, which represent the deepest division among the methanogens (Woese and Olsen 1986).

The position of *Thermoplasma acidophilum* on the tree has not been definitively established. The 16S rRNA sequence data seem to indicate that the *Thermoplasma* lineage branched between the *Methanobacteriales* and the *Methanomicrobiales*. However, the fact that the ribosomal subunits from *Thermoplasma* have a high protein content like the *Methanoccocales* and the other sulfur-dependent thermophiles (Cammarano *et al.* 1986), suggests that the *Thermoplasma* line branched earlier, perhaps between the *Methanobacteriales* and the *Methanococcales* (Woese and Olsen 1986).

The position of *Archeaeglobus*, as expected from its phenotype, is intermediate between the *Methanococcales* and *Thermococcus celer* (Achenbach-Richter *et al.* 1987).

Finally, *Thermococcus celer*, represents the deepest branch of the methanogen-halophile line. It seems to be the most slowly evolving archaebacterial line and it branches very close to the root of the archaebacterial tree (Achenbach-Richter *et al.* 1988). For these reasons, it is possible that it might even represent a third and distinct branch (Achenbach-Richter *et al.* 1988). The discovery of a second organism (*Pyrococcus woesei*) belonging to the *Thermococcus* line seems to support this idea (Zillig *et al.* 1987).

Since the thermophilic phenotype is present in both archaebacterial branches, it seems likely that it represents the ancestral archaebacterial phenotype.

Recently, Lake (1988, 1989) has presented a different tree based also on 16S rRNA sequence data, but obtained by using a new analytical method called evolutionary parsimony. This method has the advantage that it minimizes the effect of including sequences in a tree that are evolving at different rates. The tree proposed by Lake (1988, 1989), groups the methanogens and halophiles with the eubacteria, in a superkingdom designated parkaryotes; and the sulfur-dependent thermophiles (called eocytes by Lake) with the eukaryotes, in a superkingdom called karyotes; thus denying the existence of the archaebacterial kingdom. The validity of Lake's tree has recently been questioned by Achenbach-Richter *et al.* (1988) and Olsen and Woese (1989) on the basis that the statistical significance of this tree was calculated by using only the extreme halophile sequences as representatives of the methanogen branch. Achenbach-Richter *et al.* (1988) and Olsen and Woese (1989) have found that when the methanogens are used in the analysis, a similar tree to that shown in Figure 1 is also obtained with the evolutionary parsimony method.

Additional support for the archaebacterial tree has been obtained from the work of Cedegren *et al.* (1988) and Gouy and Li (1989). Cedegren *et al.* (1988) have used the 16S rRNA and 23S rRNA sequence data to construct phylogenetic trees by the maximum parsimony method, and in both cases they obtain the archaebacterial tree. Guoy and Li (1989) have analyzed the same data (16S and 23S rRNA) by the neighbour-joining as well as maximum parsimony methods, and obtained the same results. They have also applied

the evolutionary parsimony method of Lake to the 23S rRNA data and have also obtained the archaebacterial tree. Therefore, it is likely that Lake's tree is an artifact.

Molecular Biology

The genome size of the different archaebacterial groups ranges between 0.84×10^9 to 2.3×10^9 daltons and is comparable to that of the eubacteria (the size of the *E.coli* genome is 2.5×10^9 daltons) (Brown *et al.* 1989). The G-C (guanine-cytosine) content is very variable and ranges from 21% to 68% (Doolittle 1985).

Archaebacterial DNA is associated with basic DNA binding proteins similar to the histones of eukaryotes (Von Holt *et al.* 1979) and the HU proteins of the eubacteria (Briat *et al.* 1984). DNA binding proteins have been isolated from *Thermoplasma acidophilum* (DeLange *et al.* 1981), *Methanobacterium thermoautotrophicum* (Chartier *et al.* 1985), *Sulfolobus acidocaldarius* (Choli *et al.* 1988) and *Halobacterium halobium* (Ohba and Oshima 1981). All these proteins have different molecular weights and amino acid compositions.

The DNA binding protein from *Thermoplasma acidophilum* has been sequenced. It shows sequence similarity to the eukaryotic histones and to HU-1 and HU-2 from *Escherichia coli* (DeLange *et al.* 1981). *Sulfolobus acidocaldarius* contains three groups of DNA binding proteins with molecular weights of 7 (5 different proteins), 8 (2 proteins) and 10 kDa (2 proteins). Three of the five 7 kDa proteins have been sequenced and they show no homology to any eubacterial or eukaryotic DNA binding proteins (Choli *et al.* 1988). Furthermore, they are also not homologous to the *Thermoplasma* DNA binding

protein. This is somewhat surprising because the DNA binding proteins in eubacteria and eukaryotes are highly conserved (Jones *et al.* 1987).

Several rRNA, transfer RNA (tRNA) and protein coding genes have been cloned and sequenced from different archaebacterial groups. Table 1 shows some examples of the genes that have been sequenced.

Archaebacteria, like the eubacteria and the eukaryotes, use the universal genetic code. The organization of archaebacterial genes is similar to that found in the eubacteria, that is the genes are linked together in operons (Zillig *et al.* 1988, Brown *et al.* 1989). In many cases, like the tryptophan synthetase genes (*trp BA*) (Sibold and Henriquet 1988), RNA polymerase genes (Berghöfer *et al.* 1988, Zillig *et al.* 1988, Leffers *et al.* 1989) *spc* and *str* ribosomal operons (Lechner and Böck 1987, Lechner *et al.* 1988, Auer *et al.* 1989a, b), L1-L10-L12 ribosomal protein genes (Itoh 1988, Shimmin *et al.* 1989a) and the nitrogenase Fe protein genes (Souillard and Sibold 1986, Souillard *et al.* 1988), the order of the genes is similar to that found in the eubacteria, although the transcription of the genes is different. In other cases, like the *his I* and *his A* genes, the organization of the genes is completely different to that found in the eubacteria (Beckler and Reeve 1986, Cue *et al.* 1985).

Some archaebacterial genes, like their eukaryotic counterparts have introns. Introns have been identified in the tRNAs from *Sulfolobus solfataricus* (Kaine *et al.* 1983, Kaine 1987), *Thermoproteus tenax* (Wich *et al.* 1987) and *Halobacterium volcanii* (Daniels *et al.* 1985, Datta *et al.* 1989), and the 23S rRNA from *Desulfurococcus mobilis* (Kjems and Garrett 1985). There have been no reports as yet of introns in protein coding genes.

Table 1 Some Examples of Archaebacterial Genes that have been Cloned and Sequenced

Gene	Organism	Reference
16S rRNA	*Halobacterium cutirubrum*	Hui and Dennis 1985
	Halobacterium halobium	Mankin *et al.* 1985
	Halobacterium morrhuae	Leffers and Garrett 1984
	Halobacterium volcanii	Gupta *et al.* 1983
	Methanobacterium formicicum	Lechner *et al.* 1985
	Methanobacterium hungateii	Yang *et al.* 1985
	Methanobacterium thermoautotrophicum	Østergaard *et al.* 1987
	Methanococcus vannielii	Jarsch and Böck 1985a
	Thermoplasma acidophilum	Ree *et al.* 1989
	Archaeoglobus fulgidus	Achenbach-Richter *et al.* 1987
	Thermococcus celer	Achenbach-Richter *et al.* 1988
	Sulfolobus solfataricus	Olsen *et al.* 1985
	Thermoproteus tenax	Leinfelder *et al.* 1985
	Desulfurococcus mobilis	Kjems *et al.* 1987a
23S rRNA	*H. halobium*	Mankin and Kagramanova 1986
	H. morrhuae	Leffers *et al.* 1987
	M. thermoautotrophicum	Leffers *et al.* 1987
	M. vannielii	Jarsch and Böck 1985b
	D. mobilis	Leffers *et al.* 1987
	T. tenax	Kjems *et al.* 1987b
5S rRNA	*Sulfolobus acidocaldarius*	Reiter *et al.* 1987b
	M. vannielii	Wich *et al.* 1984

Table 1 ...continued

Gene	Organism	Reference
5S rRNA	*D. mobilis*	Kjems and Garrett 1987
Ribosomal Proteins		
L12	*M. vannielii*	Strobel *et al.* 1988
L10	*M. vannielii*	Köpke *et al.* 1989
L11, L1, L10, L12	*H. cutirubrum*	Shimmin and Dennis 1989
S11	*Halobacterium marismortui*	Arndt and Kimura 1988
L1, L10, L12	*H. halobium*	Itoh 1988
spc operon	*M. vannielii*	Auer *et al.* 1989a, b
RNA polymerase		
AB'B"C	*M. thermoautotrophicum*	Berghöfer *et al.* 1988
AB'B"C	*H. halobium*	Leffers *et al.* 1989
AC	*H.morrhuae*	Leffers *et al.* 1989
ABC	*S.acidocaldarius*	Pühler *et al.* 1989
Elongation factors		
aEF-1	*M. vannielii*	Lechner and Böck 1987
aEF-2	*M. vannielii*	Lechner *et al.* 1988
Genes involved in other functions		
hop	*H. halobium*	Blanck and Osterhelt 1987
bop	*H. halobium*	Dunn *et al.* 1981
brp	*H. halobium*	Betlach *et al.* 1984
his A	*Methanococcus thermolithotrophicus*	Weil *et al.* 1987
	M. vannielii	Cue *et al.* 1985
	Methanococcus voltae	Cue *et al.* 1985

Table 1 ...continued

Gene	Organism	Reference
his I	*M. vannielii*	Beckler and Reeve 1986
nif H	*M.voltae*	Souillard and Sibold 1986
nif H	*M. thermolithotrophicus*	Souillard *et al.* 1988
	Methanobacterium ivanovii	Souillard *et al.* 1988
mcr ABG	*M. thermoautotrophicum*	Bokranz *et al.* 1988
	M. vannielii	Cram *et al.* 1987
	M. voltae	Allmansberger *et al.* 1986
mcr B	*Methanosarcina barkeri*	Bokranz and Klein 1987
fdh A	*M. formicicum*	Shuber *et al.* 1986
atp A	*S. acidocaldarius*	Denda *et al.* 1988a
atp B	*S. acidocaldarius*	Denda *et al.* 1988b
atp P	*S. acidocaldarius*	Denda *et al.* 1989
arg G	*M. barkeri*	Morris and Reeve 1988
	M. vannielii	Morris and Reeve 1988
trp BA	*M. voltae*	Sibold and Heriquet 1988
gdh	*Methanobacterium bryantii*	Fabry *et al.* 1989
	Methanothermus fervidus	Fabry *et al.* 1989
	M. formicicum	Fabry *et al.* 1989

Transcription

RNA Polymerase

Archaebacteria, like eubacteria, have only one RNA polymerase, while eukaryotes have three RNA polymerases that are responsible for the transcription of different sets of genes (Zillig *et al.* 1985b). RNA polymerase I

transcribes 5.8S, 18S and 28S rRNA and the so called "small nuclear polymerase I RNA" genes (Mandal 1984, Reichel and Benecke 1984); RNA polymerase II transcribes all the protein coding genes as well as the small nuclear RNA genes (Gluzman 1985, Mangin *et al.* 1986) while RNA polymerase III transcribes the 5S rRNA, tRNA and small cytoplasmic RNA genes (Sakonju *et al.* 1980, Sharp *et al.* 1986, Reichel and Benecke 1980).

The number of subunits present in the enzymes from the three kingdoms varies. In the eubacteria there are only five subunits: $\beta\beta'\alpha_2\sigma$ (Zillig *et al.* 1976), while the eukaryotic enzymes have between 9 and 12 components (Sentenac 1985). Archaebacterial RNA polymerases have between 7 and 12 components. There is a difference in the number of components between the methanogens and halophiles, and the sulfur-dependent thermophiles. RNA polymerases from the halophiles and the methanogens have five large components A, B', B", C and D (in order of molecular weight), and about three smaller components, while the enzymes from the extreme thermophiles, including *Thermococcus* and *Thermoplasma*, have four large components, B, A, C and D (in order of decreasing molecular weight), and more than 6 smaller components (Zillig *et al.* 1985a). Note that the B component has a higher molecular weight than the A component in the extreme thermophiles.

Archaebacterial enzymes, like their eukaryotic counterparts, are insensitive to rifampicin and streptolydigin, antibiotics that inhibit eubacterial RNA polymerases. However, they are also insensitive to α-amanitin, a fungal toxin that inhibits RNA polymerase II and less efficiently RNA polymerase III (Zillig *et al.* 1985b).

Recently, the genes for the A, B, and C subunits in *Sulfolobus acidocaldarius* and the A, B', B" and C subunits in *Halobacterium halobium*

(Zillig *et al.* 1988, 1989b, Leffers *et al.* 1989) and *Methanobacterium thermoautotrophicum* (Berghöfer *et al.* 1988) have been cloned and sequenced. Sequence comparison with the equivalent genes from the eubacteria and the eukaryotes shows that the archaebacterial enzymes are closer to the eukaryotic polymerases than to the *Escherichia coli* polymerase. Sequence conservation between the archaebacterial and eukaryotic enzymes is extremely high, particularly around functionally important regions, like the substrate binding site and the zinc "finger" (Zillig *et al.* 1988, Berghöfer *et al.* 1988, Allmansberger *et al.* 1989, Leffers *et al.* 1989).

Promoters

Analysis of the conserved sequences upstream of the transcription initiation site for tRNA, rRNA and protein coding genes, has led to the proposal of a consensus sequence for archaebacterial promoters (Wich *et al.* 1986a, Reiter *et al.* 1987b, 1988, Kjems and Garrett 1987b, Zillig *et al.* 1988, Thomm and Wich 1988, Thomm *et al.* 1989). The promoter consists of two conserved boxes: box **A**, located about 25 nucleotides upstream of the transcription start site with the consensus sequence **TTA(T/A)A**, and a weakly conserved box **B**, around the transcription initiation site with the sequence **(A/T)TG(A/C)**. Initiation of transcription usually takes place at the central **G** or at a purine residue nearby (Zillig *et al.* 1988).

The structure of archaebacterial promoters, resembles that of eukaryotic RNA polymerase II promoters, which have a weakly conserved dinucleotide **CA** around the transcriptional initiation site and an **AT** rich "TATA box" motif centered about 25-30 nucleotides upstream of it (Reiter *et al.* 1988, Zillig *et al.* 1988, Thomm *et al.* 1989).

Recently, Thomm *et al.* (1988) and Brown *et al.* (1988a) have demonstrated that these conserved regions are indeed recognized by RNA polymerase. In the case of the *his A* gene, Brown *et al.* (1988a) found that the polymerase protects a 43 nucleotide fragment, which contains both boxes, from digestion by DNAse I. In the case of the gene for the C component of the methyl coenzyme M reductase, the polymerase protects a 49 nucleotide fragment, that also includes the two boxes, from digestion with exonuclease III (Thomm *et al.* 1988).

Terminators

There appears to be great variability in the sequences that determine the termination of transcription in the different archaebacterial groups. In the sulfur-dependent thermophiles, transcripts from rRNA genes (Kjems and Garrett 1987, Kjems *et al.* 1987b) and from the *Sulfolobus* phage SSV1 (Reiter *et al.* 1988) have been found to end within pyrimidine rich regions. In the methanogens, rRNA and tRNA transcripts also end in pyrimidine regions, but these are followed by a short hairpin loop (Wich *et al.* 1986a, 1986b, Østergaard *et al.* 1987). Transcripts from protein coding genes (C component of the methyl Co M reductase), on the other hand, have been found to end in an oligo T sequence <u>after</u> a hairpin loop, a structure that resembles the ρ independent terminators in the eubacteria (Müller *et al.* 1985, Allmansberger *et al.* 1986, Bokranz *et al.* 1988). In the halophiles, transcripts for rRNA genes (Chant *et al.* 1986, Chant and Dennis 1986) and protein coding genes (Chant *et al.* 1986, DasSarma *et al.* 1984, Itoh 1988, Shimmin and Dennis 1989) end in A-T rich regions preceded by a G-C rich region.

Messenger RNA (mRNA) and Translation Signals

Archaebacterial mRNA can be either monocistronic, (that is the mRNA carries the information for only one polypeptide) (DasSarma *et al.* 1984, Betlach *et al.* 1984, Shimmin and Dennis 1989) or polycistronic, (Allmansberger *et al.* 1986, Bokranz *et al.* 1988).

Unlike eukaryotic mRNAs, which have a cap structure (7mGpppXpmY) at their 5' end (Shatkin 1976), archaebacterial and eubacterial mRNAs are not capped (Brown and Reeve 1985, 1986, Ohba and Oshima 1982. Oshima *et al.* 1984). Eukaryotic mRNAs are usually polyadenylated at their 3' end (Kozak 1983). Small poly A tails have also been observed in eubacterial mRNAs but they are short and the mRNAs are unstable (Gopalakrisna *et al.* 1981). The methanogens and halophiles have mRNAs with short poly A tails, that are unstable like their eubacterial counterparts (Brown and Reeve 1985,1986); while the extreme thermophiles seem to have long poly A tracts like the eukaryotes (Ohba and Oshima 1982, Oshima *et al.* 1984).

Initiation of translation in eubacteria, is determined by the interaction of a purine rich region (the Shine-Dalgarno sequence), located between 5-13 nucleotides upstream of an initiation codon in mRNA, with a complementary pyrimidine rich region in the 3' end of the 16S rRNA (Shine and Dalgarno 1974, Gold 1988). In eukaryotes, there is no evidence of a Shine-Dalgarno type interaction between the 18S rRNA and mRNA (Kozak 1983).

Shine-Dalgarno type sequences have been detected upstream of protein coding genes from the methanogens (Cue *et al.* 1985, Souillard and Sibold 1986, Allmansberger *et al.* 1986, 1989, Lechner and Böck 1987, Souillard *et al.* 1988, Köpke and Wittmann-Liebold 1989, Brown *et al.* 1989). In the case of the halophiles, Shine-Dalgarno sequences have been identified downstream

of the initiation codon (Dunn *et al.* 1981, Betlach *et al.* 1984, Blanck and Oesterhelt 1987, Brown *et al.* 1989), while in the thermophiles, Shine-Dalgarno sequences have been observed both upstream and downstream of the initiation codon (Zillig *et al.* 1988). In some cases, two Shine-Dalgarno sequences are tandemly repeated (Zillig *et al.* 1988). However, although all these sequences could conceivably interact with the 3' end of the 16S rRNA, such interaction has not been demonstrated up to now.

Ribosomes: Structure, Function, and Genetics

Structure of the Ribosome

The ribosomes from all three kingdoms, contain two subunits: 30S and 50S in archaebacteria and eubacteria, and 40S and 60S in eukaryotes (Wittmann 1983). Each subunit contains both proteins and RNA. Table 2 gives a summary of the components of each subunit for ribosomes from the different kingdoms.

The general morphology of both subunits has been extensively studied using electron microscopy. These studies have led to the proposal of several models for the ribosomal subunits of eubacteria (particularly of *Escherichia coli*), archaebacteria and eukaryotes (see for example, Wittmann 1986, Stöffler and Stöffler-Meilicke 1986a, Oakes *et al.* 1986). Figure 2 shows the models proposed by Lake (1985) for the small and large ribosomal subunits in the three kingdoms. The main features of each subunit are also indicated in this figure.

Table 2 Ribosomal Components in the Three Kingdoms

Source	Small subunit		Large subunit	
Eubacteria[a]	number of proteins	size of rRNA	number of proteins	size of rRNA
Escherichia coli	21	16S	36[b]	5S, 23S
Eukaryotes[a]				
yeast	30-31	18S	41-45	5.8S[c], 5S, 28S
animals	≈31	18S	≈49	5.8S, 5S, 28S
plants	≈32	18S	≈47	5.8S, 5S, 28S
Archaebacteria				
Sulfolobus solfataricus [d]	28	16S	33-35	5S, 23S
Sulfolobus acidocaldarius [e]	27	16S	34	5S, 23S
Methanobacterium thermoautotrophicum [f]	22	16S	32	5S, 23S
Methanobacterium bryantii [f]	23	16S	32	5S, 23S
Methanococcus vannielii [f]	25	16S	32	5S, 23S
Halobacterium cutirubrum [g]	21	16S	32	5S, 23S

[a] Kozak 1983, [b] Wada and Sako 1987 [c] homologous to the 5' end of 23S rRNA (Jacq 1981) [d] Londei *et al.* 1983, [e] Schmid and Böck 1982, [f] Schmid *et al.* 1982, [g] Strøm and Visentin 1973

According to these models, the morphology of the small subunit seems to be very variable among the different kingdoms. Note, for example, the presence of a structure that resembles a "duck bill" in the small subunit of archaebacteria and eukaryotes, which is not present in the eubacterial small subunit (Lake 1985). The eukaryotic small subunit, also has two lobes at the bottom (Lake 1985). It is believed, although there is no direct evidence, that

these lobes correspond to extra sequences present in the 18S rRNA (Noller and Lake 1984).

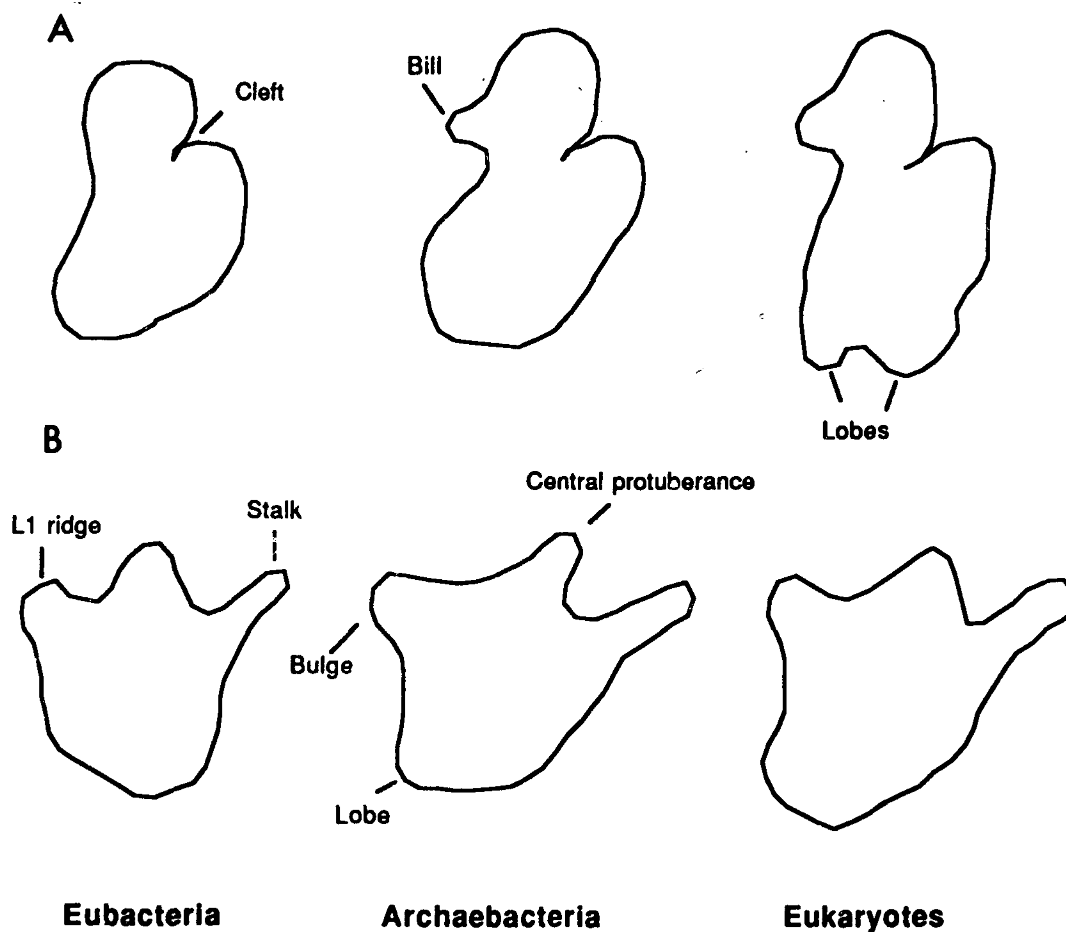


Figure 2. Models for the small and large ribosomal subunits from the three kingdoms. A. Small subunit: from left to right: *Escherichia coli*, *Thermoproteus tenax*, *Saccharomyces cerevisiae* (Drawn after Lake 1985) B. Large subunit from the same organisms (Drawn after Lake 1985)

The structure of the large subunit seems to be relatively similar in the three kingdoms (see Figure 2). However, a lobe can be seen at the bottom of the eukaryotic and archaebacterial large subunits, as well as a bulge near the L1 ridge (Lake 1985).

Lake *et al.* (1984, 1985, 1986) have used the differences in the structure of the two ribosomal subunits from different organisms to propose the existence of four kingdoms: the photocytes, which include the eubacteria and the halophiles; the archaebacteria represented by the methanogens; the eocytes which are the sulfur-dependent thermophiles; and the eukaryotes. According to this scheme, the eukaryotes are closely related to the eocytes and the archaebacteria to the photocytes.

This proposal has been severely criticized by several groups (Stöffler and Stöffler-Meilicke 1986a, Woese and Olsen 1986, Harauz *et al.* 1987), on the basis that many of the characteristics of the ribosomal subunits used to define these groups have actually been detected in other groups. For example, according to Lake *et al.* (1984), a bulge is not present in the large subunit from the photocytes. However, Harauz *et al.* (1987) have seen such a bulge in the large subunit of *Halobacterium halobium*. Furthermore, the usefulness of ribosomal morphology as a phylogenetic marker is dubious. Since many of the fine structural details used for this classification are at the border of resolution of the method, it is difficult to determine which represent real structures and which are merely artifacts produced during the sample preparation. Because of these limitations, the interpretation of the results is always subjective and thus this type of data is not useful for establishing phylogenetic relationships (Harauz *et al.* 1987).

Recently, the use of computer-imaging averaging techniques, which eliminate the subjective interpretation of electron micrographs, has allowed several authors to get a better model of the ribosomal subunits (Verschoor *et al.* 1984, 1985, Radermacher *et al.* 1987, Wagenknecht *et al.* 1988). Figure 3

shows drawings of the models for the 30S and 50S ribosomal subunits from *Escherichia coli*.

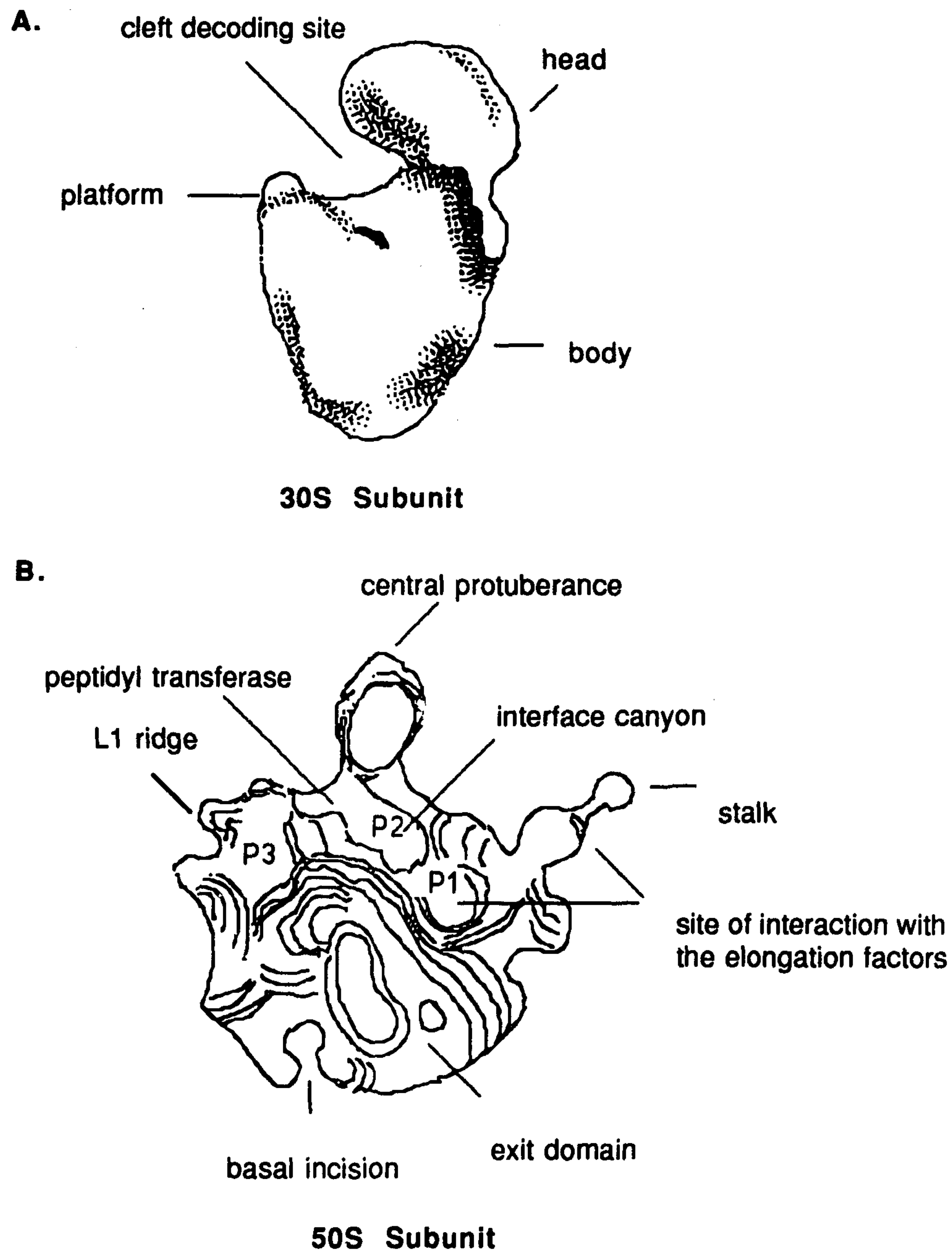


Figure 3. Computer-Imaging models of the 30S and 50S subunits from *Escherichia coli*. A. 30S subunit (Drawn from the model of Verschoor *et al.* 1984) B. Crown view of the 50S subunit. P: pockets (Drawn from the model of Radermacher *et al.* 1987).

Computer-imaging has revealed that the side protrusion of the small subunit has the shape of a platform that partially wraps around the upper part of the body of the particle (see Figure 3A) instead of a planar lobe as depicted in Figure 2. Also, the cleft formed between the lip of the platform and the head, has a cup-like shape (Verschoor *et al.* 1984). This might be important for the function of the ribosome since the cleft is the site where the mRNA binds to the ribosome (Verschoor *et al.* 1984).

In the case of the 50S subunit, computer-imaging has revealed the presence of a large groove, designated the interface canyon, which had not been detected before (see Figure 3B). This canyon probably has functional significance, because it includes the regions where the peptidyl transferase (pocket P2) and the binding sites for the elongation factors (pocket P1) have been mapped (Radermacher *et al.* 1987).

Computer-imaging techniques have also been used to study the 50S subunit from archaebacterial ribosomes (Harauz *et al.* 1987). Figure 4 shows a comparison of the models of the 50S subunit from *Sulfolobus solfataricus* and *Escherichia coli* obtained with this technique. Note that in this case we are looking at the back of the subunit, since the stalk is facing the left. The model is shown in this orientation because Harauz *et al.* (1987) were unable to obtain subunits that had the other orientation. The main differences between the 50S subunits from *S.solfataricus* and *E.coli* are the following (see Figure 4): 1] in *Sulfolobus* the central protuberance and the L1 ridge have a squarish shape while in *E.coli* they are rounded, 2] there is a notch between the central protuberance and the L1 ridge in *Sulfolobus* which is not present in *E.coli* , 3] the groove appears to be lie more to the left in *Sulfolobus* than in *E.coli* , 4] the

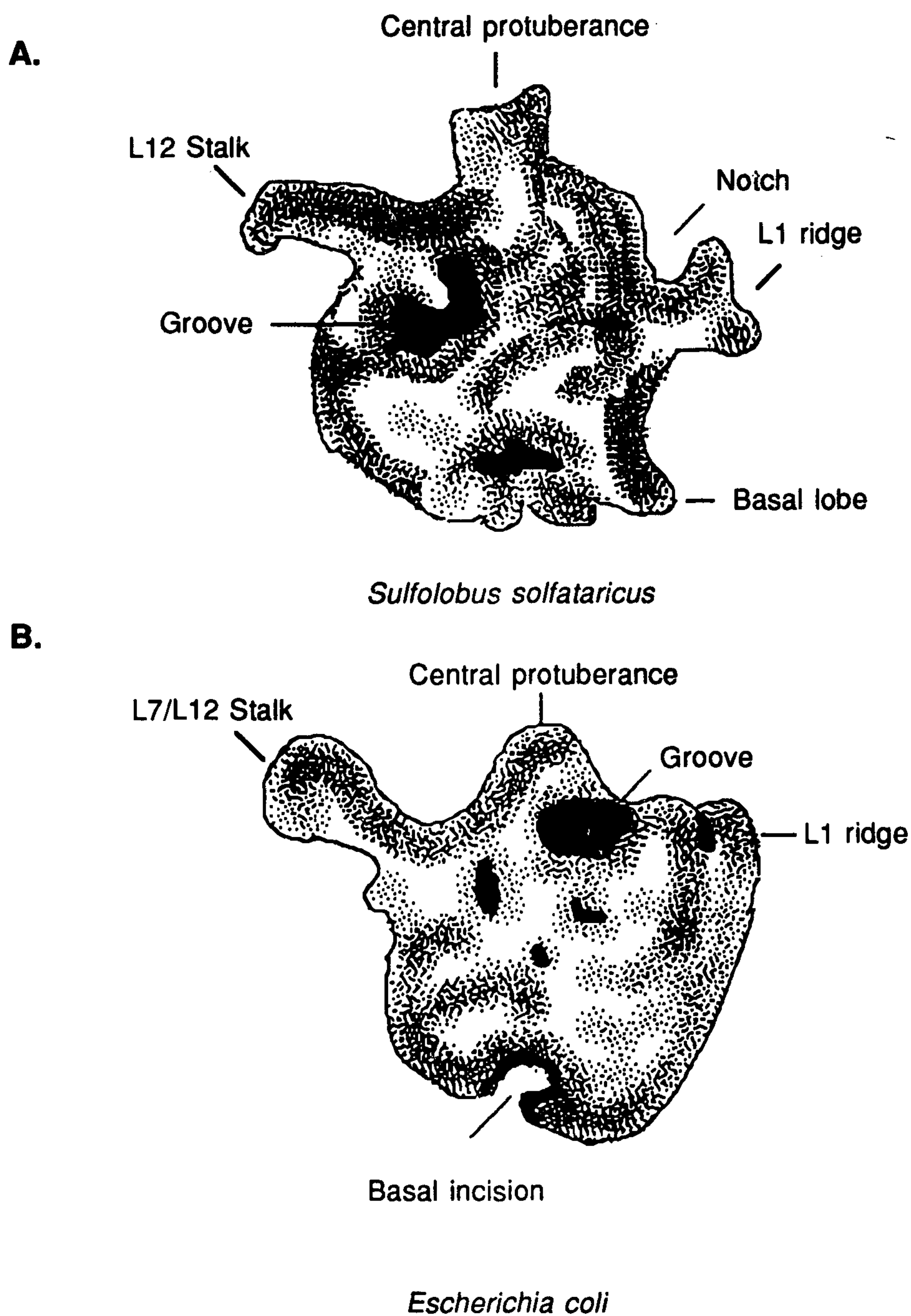


Figure 4. Computer-imaging models of the 50S subunit from *Sulfolobus solfataricus* and *Escherichia coli*. A. Rear view of the 50S subunit from *S. solfataricus* (Drawn after the model of Harauz *et al.* 1987) B. Rear view of the 50S subunit from *E. coli* (Drawn after the model of Harauz *et al.* 1987).

basal incision is not present in *Sulfolobus* , 5] the small basal lobe is not as obvious in *E.coli* as in *Sulfolobus* (Harauz *et al.* 1987). The functional significance of these differences is not known.

The structure of the ribosome has also been studied using three-dimensional image reconstruction of two-dimensional crystals. These studies have revealed the presence of a tunnel in the large subunit (Yonath *et al.* 1987). It is thought that this tunnel provides a passage for the nascent polypeptide chain to leave the ribosome (Yonath *et al.* 1987).

Recently, three-dimensional crystals of the large subunit from *Halobacterium marismortui* have been obtained that diffract to a resolution of 5.5 Å (Yonath *et al.* 1987). It is hoped that in the near future, x-ray crystallographic studies of these crystals will provide detailed information of the structure of the large subunit in the archaebacteria (Yonath *et al.* 1987).

The position of the different ribosomal proteins and rRNAs in the small and large ribosomal subunits from eubacteria, and to a lesser extent from the eukaryotes has been determined by using techniques such as cross-linking, affinity labeling, chemical and enzymatic probing, immunoelectron microscopy and neutron scattering (for reviews, see Wittmann 1986, Traut *et al.* 1986, Tolan and Traut 1981, Brimacombe *et al.* 1986, Uchiumi *et al.* 1986, Stöffler and Stöffler-Meilicke 1986b, Lake 1985, Noller 1984, Capel *et al.* 1987, Moazed *et al.* 1988, Egebjerg *et al.* 1989, Stern *et al.* 1989). Unfortunately, similar studies have not been performed with archaebacterial ribosomes.

Besides giving information about the location of the different ribosomal proteins and rRNAs, the studies mentioned above have allowed several groups to locate the functional domains of the ribosome (for a review, see Wittmann 1986). These domains are indicated in Figure 3.

The site of interaction of the mRNA, tRNAs and initiation factors (decoding site) with the ribosome is located in the small subunit; in the cleft between the head and the platform (see Figure 3A) (McKuskie-Olson and Glitz 1979, Gornicki *et al.* 1984, Stöffler and Stöffler-Meilicke 1986b).

The peptidyl transferase center of the ribosome is located in the large subunit in the valley between the L1 ridge and the central protuberance (Figure 3B) (Cooperman 1980).

The region of the ribosome involved in the binding of the elongation factors (GTPase center) is also located in the large subunit. This site includes the stalk and the region around its base (Figure 3B) (Hamel *et al.* 1972, Girshovich *et al.* 1981, Traut *et al.* 1986, Möller *et al.* 1983).

Finally, the site of exit of the nascent polypeptide chain is located on the bottom of the large subunit (Figure 3B) (Bernabeu *et al.* 1983).

In general, these functional domains have been conserved in eubacteria and eukaryotes, so it seems probable that they are also conserved in the archaebacteria (Lake 1985).

Ribosomal RNA

5S rRNA

Thirty eight archaebacterial 5S rRNA sequences have been determined (Wolters and Erdmann 1988). Comparative analysis of the sequences of 5S rRNA from the three kingdoms has led to the proposal of a general secondary structural model for 5S rRNA (for a review see, Wolters and Erdmann 19[illegible]). Although all 5S rRNAs have this general structure, each kingdom has its own particular features. For a description of the model and a detailed analysis of the differences among the three kingdoms, the reader is referred to the reviews by

Fox (1985), and Wolters and Erdmann (1988). In general, archaebacteria differ from the other two kingdoms in that there is considerable variation in the structure of the 5S rRNA from the different groups, particularly among the sulfur-dependent thermophiles (Wolters and Erdmann 1988). For this reason, a consensus model of the structure of 5S rRNA for the whole kingdom has not been proposed (Wolters and Erdmann 1988).

Although there are differences in the structure of the 5S rRNAs from the three kingdoms, reconstitution experiments, in which the 5S rRNA from *Halobacterium cutirubrum* and *Thermoplasma acidophilum* was incorporated into the large subunit of *Bacillus stearothermophilus* ribosomes (Erdmann *et al.* 1986), and *Escherichia coli* 5S rRNA was incorporated into the large subunits of *Sulfolobus solfataricus* ribosomes (Teixido *et al.* 1989), show that from a functional point of view, eubacterial and archaebacterial 5S rRNA are interchangeable. Similar experiments, to determine whether eukaryotic 5S rRNA is incorporated into archaebacterial ribosomes and archaebacterial 5S rRNA is incorporated into eukaryotic ribosomes, have not been performed as yet.

<u>16S rRNA</u>

The sequences of 14 archaebacterial 16S rRNAs have been determined (see Table 1 for references). The comparison of these sequences, as mentioned before, is the basis on which the branching order among the different archaebacterial groups has been established (see section on Phylogeny) (Woese and Olsen 1986, Woese 1987, Achenbach-Richter *et al.* 1988).

Comparison of these sequences with those of the eubacteria and eukaryotes, as in the case of 5S rRNA, has led to the proposal of a general secondary structural model for small subunit rRNA. For a description of this model and a discussion of the differences among the small subunit rRNAs from the three kingdoms, the reader is referred to the reviews by Gutell *et al.* (1985), Woese (1987), and Dams *et al.* (1988).

Nucleotides that confer resistance to antibiotics in the eubacteria have also been identified in the archaebacterial 16S rRNA sequences. A stable **G-C** pair, close to the 3' end of the molecule, at the positions equivalent to **C**1361-**G**1433 in *Escherichia coli* 16S rRNA, confers sensitivity to paromomycin, an antibiotic that affects translational fidelity (Li *et al.* 1982). All the archaebacterial sequences analyzed have this base pair, and the organisms have been shown to be sensitive (Østergaard *et al.* 1987).

The **C** at position 860 in *E.coli* 16S rRNA is involved in determining the sensitivity to streptomycin, an antibiotic that affects translational accuracy (Montandon *et al.* 1985). All the archaebacteria, so far examined, except *Desulfurococcus mobilis*, have a **U** at that position and are insensitive. *D. mobilis*, like the eubacteria, has a **C** at that position and is the only archaebacteria known until now that is sensitive to this antibiotic (Kjems *et al.* 1987a).

<u>23S rRNA</u>

The complete sequences of six archaebacterial 23S rRNAs have been determined (see Table 1 for references). Analysis of these sequences gives the same branching order as the 16S rRNA sequences, except that

Methanococcus and *Methanobacterium* branch out together (Cedergren *et al.* 1988).

Phylogenetic comparisons have also been used to generate a general secondary structural model for the 23S rRNA (28S rRNA in the eukaryotes) from the three kingdoms. For a detailed discussion of the differences among the 23S rRNAs from the three kingdoms, the reader is referred to the review by Gutell and Fox (1988).

Nucleotides involved in the sensitivity or resistance to certain antibiotics, have been identified in archaebacterial 23S rRNAs. Archaebacteria, like the eukaryotes, are insensitive to erythromycin. This is due to the presence of a guanosine or a uridine residue, instead of an adenine at the equivalent position of base 2058 of *Escherichia coli* 23S rRNA (Jarsch and Böck 1985b, Leffers *et al.* 1987). In contrast, they are sensitive to thiostrepton, because like the eubacteria, they have an adenine residue at the equivalent position of base 1067 in *E. coli* 23S rRNA (Leffers *et al.* 1987). Finally, archaebacteria are sensitive to α-sarcin, a toxin that affects eukaryotic ribosomes. This toxin cleaves the rRNA within the sequence **AGUACGAG/AGGAAC**, with the dash indicating the site of cleavage (Leffers *et al.* 1987). Eubacteria, are insensitive, because they have a **C** instead of an **A** at the penultimate position of this sequence (Leffers *et al.* 1987).

Ribosomal Proteins

The number of ribosomal proteins present in archaebacterial ribosomes seems to vary depending on the group. The halophiles, *Methanobacteriales* and *Methanomicrobiales* have small ribosomes with a size comparable to that of eubacterial ribosomes, and the number of ribosomal proteins is roughly the

same (53-54) (Strøm and Visentin 1973, Schmid *et al.* 1982, Cammarano *et al.* 1986). The sulfur-dependent thermophiles and the *Methanococcales*, on the other hand, have larger ribosomes that contain between 60-65 proteins (Schmid *et al.* 1982, Schmid and Böck 1982, Londei *et al.* 1983, Cammarano *et al.* 1986).

In contrast to other organisms, the ribosomal proteins from the extreme halophiles are very acidic (Bayley and Kushner 1964, Matheson 1985, Kimura *et al.* 1989). A correlation has been found between the number of acidic ribosomal proteins and the internal salt concentration of the organism (Matheson 1985). The number of acidic proteins decreases as the internal salt concentration decreases from the extreme halophiles to the *Methanobacteriales*, to the *Methanomicrobiales*, to the *Methanococcales* and the sulfur-dependent thermophiles (Matheson 1985). For these reason, it has been proposed that the increased acidity of the ribosomal proteins in the extreme halophiles is an adaptation that allows the proteins to maintain their structure and function in the high internal salt concentration of the cell (Bayley and Morton 1978, Eisenberg and Wachtel 1987).

The complete sequences of several ribosomal proteins from the extreme halophiles (Itoh 1988, Kimura *et al.* 1989, Spiridonova *et al.* 1989) and the methanogens (Strobel *et al.* 1988, Köpke and Wittmann-Liebold 1989, Auer *et al.* 1989b, Köpke *et al.* 1989) as well as one from the sulfur-dependent thermophiles (Matheson *et al.* 1988) have been determined. In terms of homology, archaebacterial ribosomal proteins can be divided into four groups (Kimura *et al.* 1989): 1] proteins homologous to eubacterial and eukaryotic ribosomal proteins, for example, L12 and L10 in *Methanococcus vannielii* (Strobel *et al.* 1988, Köpke *et al.* 1989), L12 in *Sulfolobus acidocaldarius*

(Matheson *et al.* 1988), L1, L10 and L11 in *Halobacterium halobium* (Itoh 1988), L11, L1, L10, L12 in *H.cutirubrum* (Shimmin *et al.* 1989a), S14, S16, S19, L25 in *H. marismortui* (Kimura *et al.* 1989) 2] proteins homologous only to eukaryotic proteins, for example S12, S15, L16 in *H. marismortui* (Kimura *et al.* 1989) 3] proteins homologous to eubacterial proteins, for example, S11, S17, S18, L23, L33 in *H. marismortui* (Kimura *et al.* 1989) 4] proteins not homologous to either eubacterial or eukaryotic proteins, for example S6, S13, L29, L31 in *H. marismortui* (Kimura *et al.* 1989). Since all the ribosomal proteins from a eukaryotic ribosome have not been sequenced, the last two categories might change when more information is available.

The nomenclature for the ribosomal proteins in eubacteria and eukaryotes is based on the separation of the proteins in a two-dimensional gel system (Kaltschmidt and Wittmann 1970). In the case of *Escherichia coli*, all the proteins are resolved in a system in which the first dimension is an 8% polyacrylamide gel, pH 8.6 and the second dimension is an 18% polyacrylamide gel, pH 4.6 (Kaltschmidt and Wittmann 1970). The proteins are usually numbered starting from the top of the gel from left to right. Proteins from the small subunit are denoted with a letter S and those of the large subunit with a letter L (Kaltschmidt and Wittmann 1970).

In the case of the archaebacterial ribosomal proteins, several different two-dimensional gel systems have been used to separate the proteins (see for example, Strøm and Visentin 1973, Schmid and Böck 1982, Schmid *et al.* 1982, Londei *et al.* 1983). For this reason, one protein may have a different number depending on the system used. In order to avoid this confusion, the ribosomal proteins of *Sulfolobus solfataricus*, studied in this work, will be designated with the number of the equivalent protein in *Escherichia coli*. The

first letter of the genus and the first two letters of the species will be used to indicate the organism. Thus, Sso L12, denotes the ribosomal protein of *Sulfolobus solfataricus* that is equivalent to the L12 protein of *E. coli.*

Protein Synthesis

Protein synthesis occurs in a series of well ordered steps which involve the interaction of different components (mRNA, tRNA, initiation factors, elongation factors, release factors) with different parts of the ribosome (see for example, Chambliss *et al.* 1980, Hardesty and Kramer 1986). The process of protein synthesis can be divided into three steps: initiation, elongation and termination.

During initiation, the initiator tRNA carrying the first amino acid (formylmethionine in eubacteria or methionine in eukaryotes and archaebacteria) binds to the initiation codon in the mRNA and occupies the P or peptidyl site of the ribosome. Elongation begins with the binding of an aminoacyl-tRNA to the A or aminoacyl site of the ribosome. A peptide bond is then formed between the amino group of the incoming aminoacyl-tRNA and the carboxyl group of the residue carried by the initiator tRNA. After the peptide bond is formed, translocation takes place and the peptidyl-tRNA moves from the A site to the P site, while the uncharged tRNA moves to the exit site (E) and leaves the ribosome. The movement of the peptidyl-tRNA from the A to the P site is accompanied by the movement of the mRNA, so that the next codon is positioned on the A site. The ribosome is then ready to start a new elongation cycle. Termination takes place when a termination codon in the mRNA is recognized by a release factor. The peptidyl-tRNA ester bond is hydrolyzed and the protein is released from the ribosome. Figures 5 and 6 show the different steps of protein synthesis in the eubacteria and eukaryotes

respectively (for a detailed review of protein synthesis in eubacteria and eukaryotes the reader is referred to Spirin 1986 and Moldave 1985). Very little is known about the individual steps of protein synthesis in the archaebacteria. However, it is probable that the general process is very similar to that found in the other two kingdoms.

Initiation

The main differences between eubacteria and eukaryotes, in terms of the different steps of protein synthesis, have been found in the initiation step (see Figures 5 and 6) (Kozak 1983). Almost nothing is known about this step in the archaebacteria, except that, as in the eukaryotes, the methionine on the initiator tRNA is not formylated (Kuchino *et al.* 1982, Gupta 1985). In view of the fact that it is in this step where the major differences between the other two kingdoms exist, it is possible that this is where we will also find differences between the archaebacteria and the other two kingdoms.

Elongation

In contrast to the initiation step, the general scheme for the elongation cycle has been conserved in eubacteria and eukaryotes, and probably in the archaebacteria as well (see Figures 5 and 6). Elongation factors (EF) equivalent to those of eubacteria (EF-Tu and EF-G) and eukaryotes (designated EF-1a and EF-2 in this case), have been isolated from the archaebacteria (designated archaebacterial elongation factors (aEF) 1 and 2) (Klink 1985).

The sensitivity of the archaebacterial factors to certain antibiotics that react specifically with either eubacterial or eukaryotic elongation factors has also

Figure 5. Protein synthesis in the eubacteria. **Initiation:** Initiation factors IF-1 and IF-3 bind to the 30S subunit and prevent its reassociation with the 50S subunit. IF-3 promotes the binding of mRNA. The AUG initiation codon is correctly positioned on the ribosome by the interaction of the Shine-Dalgarno (purine rich) sequence in the mRNA with the complementary (pyrimidine rich) region at the 3' end of the 16S rRNA. IF-2 forms a ternary complex with GTP and formylmethionine initiator tRNA. This complex binds to the ribosome at the P or peptidyl site and IF-3 is released from the ribosome. The 50S subunit binds to the 30S subunit, GTP is hydrolyzed and IF-2 leaves the ribosome in a complex with GDP. IF-1 is also released at this stage. The ribosome is then ready to start the elongation cycle. **Elongation:** At the beginning of the elongation cycle, formylmethionine-tRNA (if it is the first cycle) or a peptidyl-tRNA occupies the P site. Elongation factor EF-Tu forms a ternary complex with an amino acyl-tRNA, and GTP. The complex binds to the ribosome and positions the amino acyl-tRNA at the A or amino acyl site of the ribosome. GTP is hydrolyzed and EF-Tu leaves the ribosome in a complex with GDP. The EF-Tu•GDP complex interacts with EF-Ts and GTP, and the EF-TU•GTP complex is regenerated for a new elongation cycle. A peptide bond is formed between the amino group of the incoming amino acid and the carboxyl group of the peptidyl-tRNA at the P site. EF-G, in a complex with GTP, binds to the ribosome and promotes the translocation of the peptidyl tRNA from the A to the P site. The uncharged tRNA moves to the E or exit site and leaves the ribosome. The translocation of the peptidyl tRNA from the A to the P site is accompanied by the movement of the mRNA, so that the next codon is positioned on the A site and a new elongation cycle can begin. **Termination:** Termination takes place when a termination codon in the mRNA is recognized by a release factor. RF-1 recognizes UAA and UGA, termination codons and RF-2 recognizes UAG and UGA termination codons. RF-3 stimulates the binding of either of these two factors in the presence of GTP. The release factor promotes the hydrolysis of the ester bond between the protein and the tRNA. The protein is released from the ribosome, GTP is hydrolyzed, and the release factor, mRNA and tRNA leave the ribosome. The two subunits dissociate and are then ready to begin the translation of a new mRNA. IF: initiation factor, fmet-$tRNA_i$: formylmethionine-tRNA initiator, aa-tRNA: amino acyl tRNA, P: peptidyl site, A: amino acyl site, EF: elongation factor.

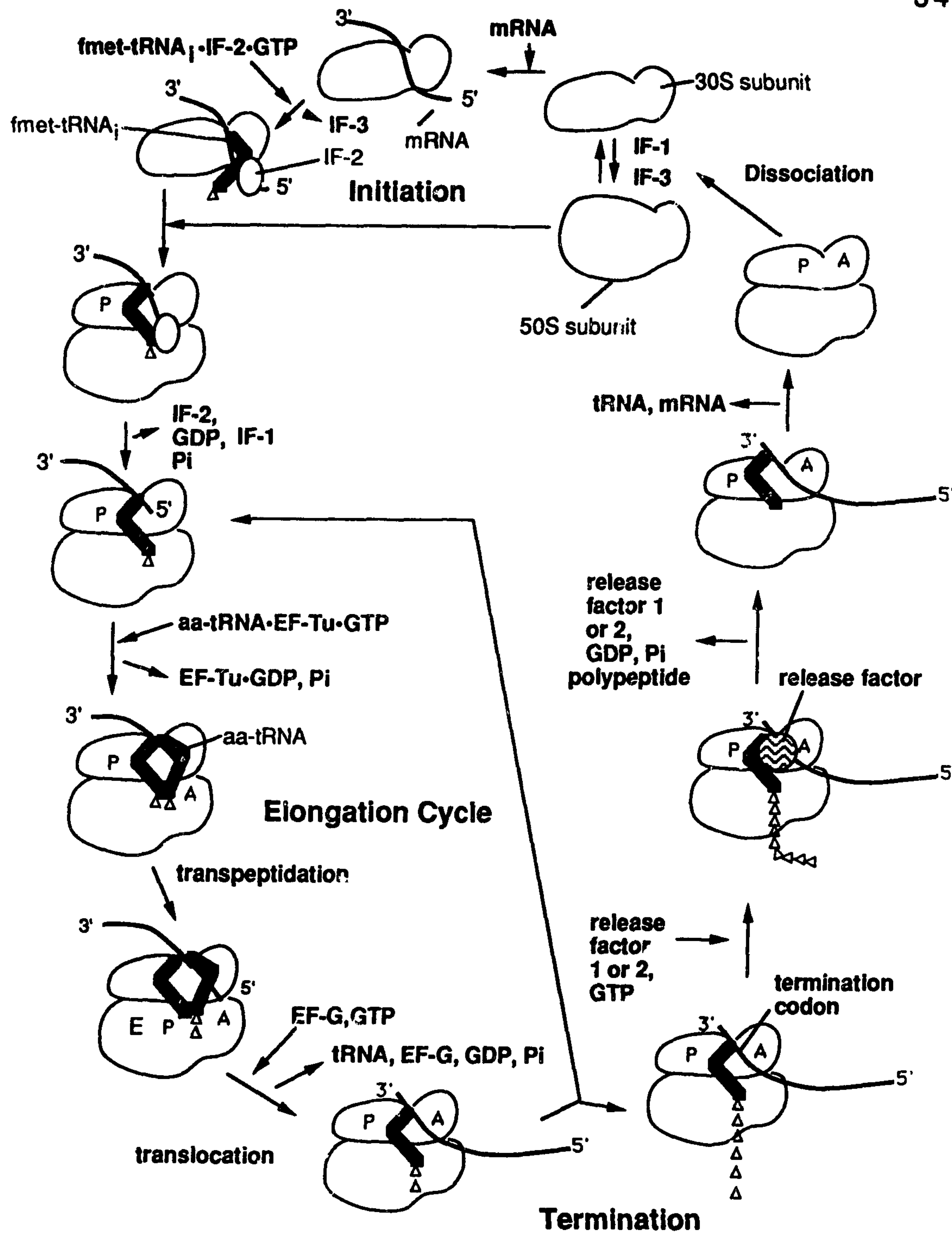


Figure 5. Protein Synthesis in Eubacteria

Figure 6. Protein synthesis in the eukaryotes. **Initiation:** Initiation factors eIF-3 and eIF-6, bind to the 40S and 60S subunits, respectively, preventing their reassociation. eIF-4C binds to the 40S subunit. Initiator methionine-tRNA forms a ternary complex with eIF-2 and GTP and binds to the 40S subunit at the P or peptidyl site. Factors eIF-1, eIF-4A, eIF-4B and eIF-4F bind to the subunit and allow it to bind to the cap (7mGpppXpmY) at the 5' end of the mRNA. ATP is required for this process. The subunit, then moves down the mRNA until it encounters an AUG initiation codon in the proper context, that is within the sequence, ACC**AUG**G (Kozak 1983). ATP is also required for this process. The binding of eIF-5 stimulates the hydrolysis of GTP, and the eIF-2•GDP complex, as well as all the other factors, leave the ribosome. The large subunit joins the small subunit, and the ribosome is ready to start the elongation cycle. The eIF-2•GDP complex reacts with the guanine exchange factor to regenerate the eIF-2•GTP complex for a new initiation cycle. **Elongation:** At the beginning of the elongation cycle, methionine-tRNA (if it is the first cycle) or a peptidyl-tRNA is located on the P site of the ribosome. The binding of an amino acyl-tRNA to the A or amino acyl site of the ribosome, requires the formation of a ternary complex between the amino acyl-tRNA, GTP and elongation factor EF-1 (the a subunit). This ternary complex binds to the ribosome, GTP is hydrolyzed and EF-1 leaves the ribosome in a complex with GDP. The b subunit of EF-1 is responsible for the regeneration of the EF-1•GTP complex to be used in the next elongation cycle. A peptide bond is formed between the amino group of the incoming amino acid and the carboxyl group of the peptidyl-tRNA at the P site. EF-2, in a complex with GTP, binds to the ribosome and promotes the translocation of the peptidyl tRNA from the A to the P site. The uncharged tRNA moves to the E or exit site and leaves the ribosome. The translocation of the peptidyl tRNA from the A to the P site is accompanied by the movement of the mRNA, so that the next codon is positioned on the A site and a new elongation cycle can begin. **Termination:** Termination takes place when a termination codon in the mRNA is recognized by the release factor. The release factor in a complex with GTP binds to the ribosome and promotes the hydrolysis of the ester bond between the protein and the tRNA. The protein is released from the ribosome, GTP is hydrolyzed, and the release factor, mRNA and tRNA leave the ribosome. The two subunits dissociate and are then ready to begin the translation of a new mRNA. eIF: eukaryotic initiation factor, met-tRNA: methionine-$tRNA_i$, aa-tRNA: amino acyl tRNA initiator, P: peptidyl site, A: amino acyl site, EF: elongation factor.

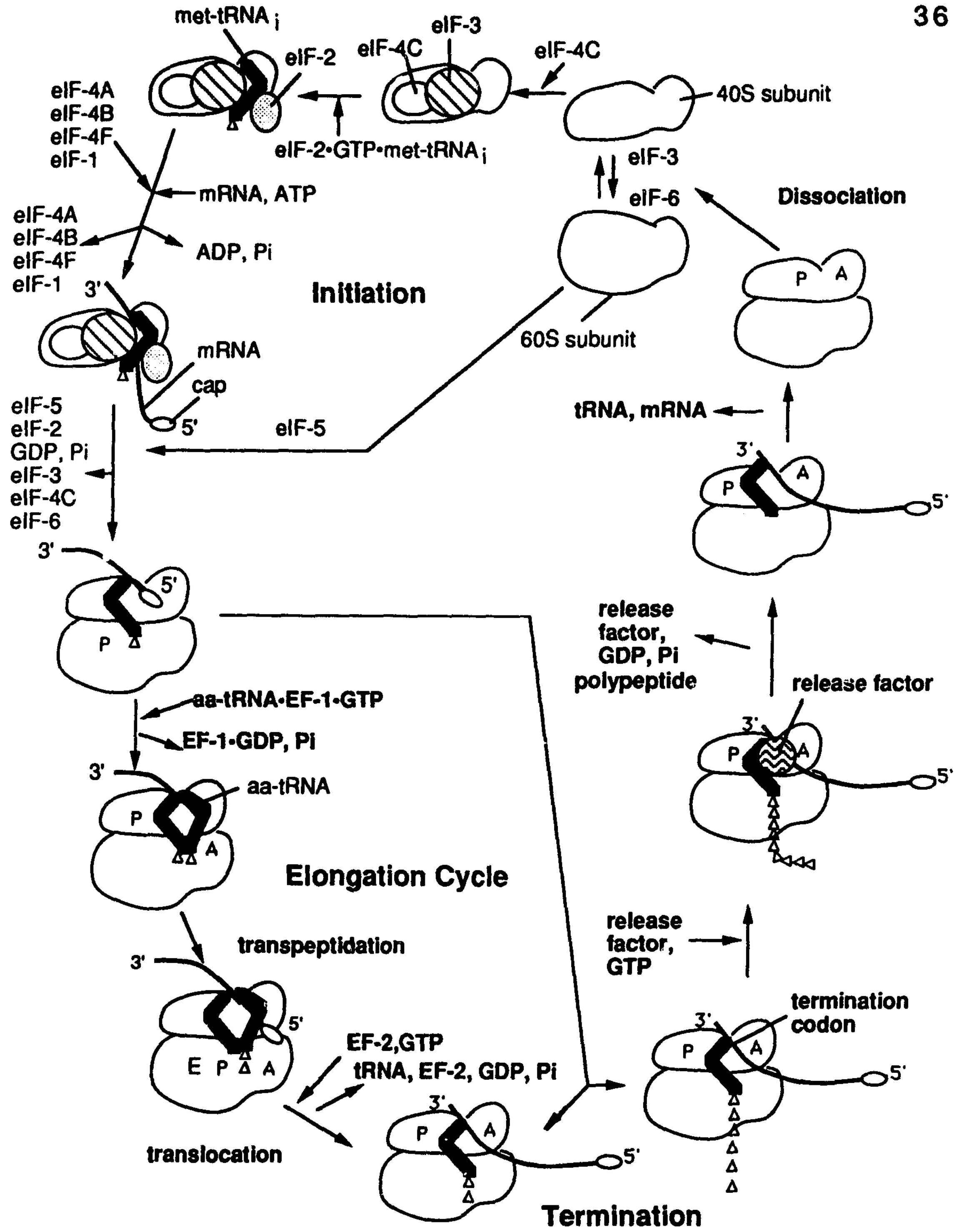


Figure 6. Protein Synthesis in Eukaryotes

been determined. Like their eukaryotic counterparts, aEF-1 factors are not affected by kirromycin, an antibiotic that prevents the EF-Tu-GDP complex from leaving the ribosome in the eubacteria (Londei *et al.* 1986). However, with respect to pulvomycin, an antibiotic that prevents the formation of the ternary complex between EF-Tu-GTP and aminoacyl-tRNA in the eubacteria, but has no effect on EF-1a, the different aEF-1 factors exhibit different sensitivities depending on the group from which they have been isolated. aEF-1 factors from the methanogens, halophiles and *Thermococcus celer* are sensitive to pulvomycin like the eubacterial factors, while those from the sulfur-dependent extreme thermophiles are not, and therefore resemble their eukaryotic counterparts (Londei *et al.* 1986). This probably reflects the existence of structural differences between the aEF-1 factors from the methanogen-halophilic branch and the sulfur-dependent thermophilic branch of the archaebacteria (Londei *et al.* 1986).

Archaebacterial aEF-2 factors have been found to be sensitive to ADP-ribosylation by diphtheria toxin (Kessel and Klink 1980, Klink 1985). In this respect, they resemble the eukaryotic factors which are also modified by diphtheria toxin (Van Ness *et al.* 1980). The site of ADP-ribosylation has been identified as a modified histidine residue called diphthamide (Van Ness *et al.* 1980). EF-G factors from eubacteria are not affected by diphtheria toxin because they lack the diphthamide residue (Van Ness *et al.* 1980).

The genes for aEF-1 and aEF-2, from the archaebacterium *Methanococcus vannielii*, have been cloned and sequenced (Lechner and Böck 1987, Lechner *et al.* 1988). Sequence comparison shows that the archaebacterial factors are more closely related to their eukaryotic counterparts than to the *Escherichia coli* elongation factors (Lechner and Böck 1987, Lechner *et al.* 1988).

The sensitivity of archaebacterial ribosomes to other antibiotics that block the elongation cycle in the other two kingdoms has been tested by using *in vitro* polypeptide synthesis systems (Böck and Kandler 1985, Hummel *et al.* 1985, Cammarano *et al.* 1985, Amils and Sanz 1986, Oliver *et al.* 1987). The antibiotics tested can be divided into three groups: 1] antibiotics that only affect eubacteria, 2] antibiotics that only affect eukaryotes, 3] antibiotics that affect eubacteria and eukaryotes (Böck and Kandler 1985, Hummel *et al.* 1985, Cammarano *et al.* 1985, Amils and Sanz 1986, Oliver *et al.* 1987). In general, it has been found that archaebacteria display a varied spectrum of responses to the three groups of antibiotics, depending on the type of organism tested (Böck and Kandler 1985, Hummel *et al.* 1985, Cammarano *et al.* 1985, Amils and Sanz 1986, Oliver *et al.* 1987). This is in contrast to the situation in the eubacteria, where a similar pattern has been observed throughout the kingdom; and the eukaryotes, where variations in the sensitivity have only been found in response to a subset of aminoglycoside antibiotics (gentamycin, kanamycin, neomycin, lividomycin, and paromycin) (Sanz *et al.* 1987).

With respect to the three groups of antibiotics tested, there appears to be a gradient of sensitivity going from the extreme sulfur-dependent thermophiles (including *Thermococcus celer* and *Thermoplasma acidophilum*), where *Sulfolobus solfataricus* is the least sensitive organism, to the halophiles and then to the methanogens, where *Methanobacterium formicicum* is the most sensitive (Cammarano *et al.* 1985, Amils and Sanz, 1986, Oliver *et al.* 1987, Sanz *et al.* 1987). This variability in the sensitivity to different antibiotics probably reflects differences in the structure of the ribosome within the archaebacteria. The correlation of these sensitivity patterns with specific changes in the structure of the ribosome will provide new insights into the

structural-functional relationships of the different components of the ribosome, as it has already done in the case of eubacteria and eukaryotes (Moazed and Noller 1989, Cundliffe 1986).

Termination

Like elongation, termination is similar in eubacteria and eukaryotes. In eubacteria, there are two release factors that recognize specific termination codons and promote the hydrolysis of the ester bond between the tRNA and the polypeptide chain: RF-1 (recognizes UAA and UGA) and RF-2 (UAG and UGA). There is also a third factor (RF-3) that stimulates the binding of the other two in the presence of GTP (Tate *et al.* 1973). Eukaryotes, on the other hand, only have one release factor that recognizes the three termination codons (Moldave 1985). GTP is also required for the binding of this factor to the ribosome. Nothing is known about the termination step of protein synthesis in the archaebacteria.

Organization of rRNA and Ribosomal Protein Genes

rRNA Genes

The organization of the rRNA genes has been highly conserved both in the eubacteria and the eukaryotes. In most eubacteria, the rRNA genes are closely linked. They are organized in an operon and transcribed in the following order: 5' 16S rRNA - 23S rRNA - 5S rRNA 3' (Nomura and Post 1980). An exception is *Thermus thermophilus*. In this organism, the 16S rRNA is not located near the 23S and 5S rRNA genes and is transcribed as a single unit (Hartmann and Erdmann 1989).

In *Escherichia coli*, there are 7 copies of the rRNA operon while in *Bacillus subtilis* there are 9 or 10 copies (Nomura *et al.* 1977, Loughney *et al.* 1982). Genes that code for tRNAs have been found in the spacer between the 16S and 23S rRNA genes in eubacteria. In *Escherichia coli*, four of the operons have a glutamic acid tRNA gene while the three others have an alanine and an isoleucine tRNA gene in the spacer (Morgan *et al.* 1977). In other eubacteria, like *B. subtilis* (Loughney *et al.* 1982) and *Anacystis nidulans* (Williamson and Doolittle 1983, Tomioka and Sugiura 1984), tRNA genes for alanine and isoleucine are also present in the spacer.

In the eukaryotes, the rRNA genes are organized as a transcriptional unit in the order 5' 18S rRNA - 5.8S rRNA - 28S rRNA 3', and are transcribed by RNA polymerase I (Planta and Meyerink 1980, Jacob 1986). These transcriptional units are tandemly repeated and their number can be very large (up to 19,300 in the amphibian *Ambystoma tigrinum*) (see Long and Dawid 1980). There are no tRNA genes located between the 18S and 28S rRNA genes in eukaryotes. The 5S rRNA gene is unlinked and is transcribed by a different RNA polymerase: RNA pol III (Long and Dawid 1980). The number of 5S rRNA genes can also be very large (up to 9000 in the frog *Xenopus borealis*) (Long and Dawid 1980).

In the archaebacteria, the organization of the rRNA genes is more variable. In the methanogens and halophiles, the genes are as closely linked as in the eubacteria and in the same order (Neumann *et al.* 1983). A gene for alanine tRNA has been detected between the 16S and 23S rRNA genes in all the methanogens and halophiles so far examined (Achenbach-Richter *et al.* 1988). In the sulfur-dependent thermophiles, *Thermococcus celer* and *Thermofilum pendens* have the same organization of rRNA genes as the methanogens and

the halophiles, except that they have an extra unlinked 5S rRNA gene (Neumann *et al.* 1983). The unlinked 5S rRNA gene in *T. celer*, has been sequenced and it is part of a small operon that includes an aspartic tRNA gene (Culham and Nazar 1988). An unlinked 5S rRNA gene is also present in *Methanococcus vannielii.* In this organism, the unlinked 5S rRNA gene is part of an operon including seven tRNA genes (Wich *et al.* 1984). In the case of *Thermoproteus tenax*, *Sulfolobus acidocaldarius* B12 and *Desulfurococcus mobilis*, the 16S and 23S rRNA genes are linked, but the 5S rRNA gene is unlinked (Neumann *et al.* 1983, Reiter *et al.* 1987b, Kjems and Garrett 1987). Finally, in *Thermoplasma acidophilum,* the three rRNA genes are unlinked (Tu and Zillig 1982, Ree *et al.* 1989). In all the sulfur-dependent thermophiles, except *Thermococcus celer*, there is no tRNA gene between the 16S and 23S rRNA genes (Achenbach-Richter *et al.* 1988).

The number of rRNA operons is also variable. There are four operons in *Methanococcus vannielii* (Jarsch *et al.* 1983), two in *Methanobacterium thermoautotrophicum* (Østergaard *et al.* 1987), *Methanothermus fervidus* (Brown *et al.* 1989), and *Halobacterium marismortui* (Mevarech *et al.* 1989), and only one in all the other archaebacteria so far examined (Neumann *et al.* 1983).

Ribosomal Protein Genes

In the eubacteria, ribosomal protein genes are organized in operons (Nomura *et al.* 1984, Ohkubo *et al.* 1987, Gross *et al.* 1989, Buttarelli *et al.* 1989). For example, in *Escherichia coli*, the ribosomal protein genes are arranged in 20 different operons. Besides ribosomal protein genes, these operons include genes that code for proteins involved in replication,

transcription, translation and other cellular functions (for a review, see Nomura *et al.* 1984, Lindahl and Zengel 1986).

The regulation of the expression of these operons has been extensively studied (for a review, see Nomura *et al.* 1984, Nomura 1986). The main control mechanism is the autogenous regulation of the synthesis of ribosomal proteins at the translational level. When ribosomal proteins are synthesized in excess, certain ribosomal proteins are able to act as repressors of translation by binding to their mRNAs. A single protein is able to block the translation of all the cistrons present in a mRNA because the translation of ribosomal proteins is coupled, that is all the proteins in an operon are translated sequentially. For this reason, when the translation of the first protein is blocked, the rest of the proteins are not translated (Nomura 1986).

Proteins L1, L10, L4, S4, S7 and S8 have been identified as repressors of their respective operons (Nomura 1986, Lindahl and Zengel 1986). Under normal conditions, these proteins bind directly to their target sites in rRNA during ribosome assembly. However, when ribosomal proteins are synthesized in excess, there is not enough rRNA to bind these proteins, so instead these proteins interact with their own mRNA and inhibit further translation (Nomura 1986). The interaction with mRNA is possible because the mRNA has a region that has a similar structure to the binding site of these proteins in rRNA (see for example, Christensen *et al.* 1984, Kearney and Nomura 1987). Thus, the competition between the binding sites on the mRNA and rRNA adjusts the production of ribosomal proteins to match the synthesis of rRNA (Nomura 1986).

In contrast to the situation in the eubacteria, the ribosomal protein genes of the eukaryotes are dispersed through the genome (D'Eustachio *et al.* 1981,

Woolford and Rosbash 1981, Planta *et al.* 1986, Warner *et al.* 1986, Mager 1988). Very few examples of genes that are closely clustered have been found, like the following pairs of ribosomal protein genes in yeast: rp29-L32 (Warner *et al.* 1985), S24-L46 (Leer *et al.* 1985a) and rp28-S16A (Molenaar *et al.* 1984) as well as the two copies of the rpS14 gene in *Drosophila melanogaster* (Brown *et al.* 1988b). However, there is no evidence that the transcription of one gene has any effect on the transcription of the other (Mager 1988).

In most cases, ribosomal protein genes in the eukaryotes are present in more than one copy. Some exceptions are the ribosomal protein genes in *Dictyostelium discoideum* (Steel and Jacobson 1986), and the rpC25, rp 1 to 6 ribosomal protein genes in *Drosophila* (Vaslet *et al.* 1980, Fabijanski and Pellegrini 1982) and L3, L25, L29, L46, S24 and rp59 in yeast (Larkin and Woolford 1983, Mager 1988).

Most of the ribosomal protein genes in yeast are duplicated, and in all the cases so far examined, both copies are functional (Woolford *et al.* 1979, Fried *et al.* 1981, Leer *et al.* 1985b, Mager 1988). In *Xenopus laevis*, two functional copies of the genes coding for ribosomal proteins L1, L14 and S19 have been detected while there are four or five copies of the S1, S8 and L32 genes (Bozzoni *et al.* 1981). In mammals, on the other hand, there are between 7 and 20 copies of each ribosomal protein gene (Monk *et al.* 1981). In general, it has been found that only one copy is functional, while the others are processed pseudogenes which are not expressed (Dudov and Perry 1984, Peled-Yalif *et al.* 1984, Wiedemann and Perry 1984, Wagner and Perry 1985, Kuzumaki *et al.* 1987, Chen and Roufa 1988).

Like most eukaryotic genes, eukaryotic ribosomal protein genes contain introns (Mager 1988). For example, the L1 gene in *Xenopus laevis* has 9

introns (Loreni *et al.* 1985), while the S17 gene in humans contains 4 introns (Chen and Roufa 1988). In yeast, the introns are usually located near the 5' end of the gene, while in other eukaryotes they can be located anywhere along the gene (Teem *et al.* 1984, Planta *et al.* 1986, Warner *et al.* 1986, Loreni *et al.* 1985, Larkin *et al.* 1987, Chen and Roufa 1988).

The regulation of the synthesis of ribosomal proteins in the eukaryotes, can occur at different levels: transcription, mRNA processing, mRNA stability, efficiency of translation and turnover of excess protein (for a detailed discussion of these regulatory mechanisms, the reader is referred to Planta *et al.* 1986, Warner *et al.* 1986, Mager 1988 and Warner 1989).

In the archeabacteria, it has recently been determined that the ribosomal protein genes are organized in operons. In many cases, the order of the genes is similar to that found in the eubacteria, but the organization of each transcriptional unit is different (see Itoh 1988, Auer *et al.* 1989b, Shimmin and Dennis 1989). The mechanism by which archaebacteria control the expression of these genes is unknown.

The L7/L12 Domain and the Purpose of this Dissertation

One of the regions of the large ribosomal subunit that has been conserved through evolution, is the so called "stalk". This structure has been observed in the ribosomes from members of the three kingdoms (see Figure 3) (Lake 1985, Harauz *et al.* 1987).

From a functional point of view, this domain is important because it is the site of interaction of the initiation, elongation and termination factors on the ribosome and is involved in triggering the GTPase activity of these factors (Hamel *et al.* 1972, Heimark *et al.* 1976, Girshovich *et al.* 1981, Möller *et al.*

1983, Rychlik *et al.* 1983, Sköld 1983, Traut *et al.* 1986, Moazed *et al.* 1988). Furthermore, this domain is also involved in the regulation of translational accuracy (Kirsebom and Isaksson 1985, Kirsebom *et al.* 1986).

In *Escherichia coli*, the stalk is formed by protein L7/L12 (Möller and Castleman 1967, Boublik *et al.* 1976, Strycharz *et al.* 1978, Möller *et al.* 1983). This is a small acidic protein, rich in alanine, that exists in two different forms. One form has the N-terminal serine residue acetylated (L7) while the other is unmodified (L12) (Terhorst *et al.* 1973). The reason for this modification is unclear, since there is no functional difference between the two forms and mutants that only produce L12, show no alterations in their phenotypes (Watson *et al.* 1975).

Physical studies of the Eco L7/L12 protein have revealed that it has a very elongated shape and a high α helical content (Österberg *et al.* 1976, Gudkov *et al.* 1977, 1978, 1981, Luer and Wong 1979). The protein also has a very flexible structure (Gudkov *et al.* 1982, Cowgill *et al.* 1984, Bushuev *et al.* 1989) and this flexibility seems to be important for the function of the ribosome since the L7/L12 protein undergoes a conformational change when EF-G binds to the ribosome (Gudkov and Gongadze 1984). The complete sequence of the Eco L7/L12 protein has been determined (see Figure 7) (Terhorst *et al.* 1973), as well as the sequence of eight L12-like proteins from other eubacteria (Itoh and Wittmann-Liebold 1978, Yaguchi *et al.* 1980, Itoh 1981, Itoh *et al.* 1982, Itoh and Higo 1983, Itoh and Otaka 1984, Garland *et al.* 1987, Matheson *et al.* 1987).

The Eco L7/L12 protein contains two well defined structural domains joined by a flexible region or hinge (see Figure 7) (Liljas 1982, Liljas and Leijonmarck 1983, Liljas *et al.* 1986a, Leijonmarck and Liljas 1987). This region is also

present in the other eubacterial proteins and is the reason why the proteins are so flexible (Bushuev *et al.* 1989).

The N-terminal domain is elongated and has a high α helix content (Luer and Wong 1979, Gudkov *et al.* 1980) while the C-terminal domain has a plum-shaped globular structure (Liljas 1982, Liljas and Leijonmarck 1983, Liljas *et al.* 1986a, Leijonmarck and Liljas 1987). The C-terminal domain has been crystallized and studied at 1.7 Å resolution (Leijonmarck and Liljas 1987). This domain is formed by three α helices and three antiparallel β strands, arranged in the order β1 α1 α2 β2 α3 β3 (see Figure 7) (Leijonmarck and Liljas 1987). The three α helices are packed on one surface, forming a layer, while the three β sheets form a second layer. In between the two layers, there is a core of hydrophobic residues (Liljas 1982, Liljas and Leijonmarck 1983, Liljas *et al.* 1986a, Leijonmarck and Liljas 1987). A similar structural motif has been observed in the Eco L30 ribosomal protein (Wilson *et al.* 1986), so it has been suggested that this might be a common structural feature of ribosomal proteins (Leijonmarck *et al.* 1988).

Recently, Rice and Steitz (1989) have shown that the C-terminal domain has a helix-turn-helix motif (residues 69-87) similar to the one found in many DNA-binding regulatory proteins. These authors speculate that this region might be involved in an interaction with RNA. However, it should be noted that the L7/L12 protein does not bind directly to rRNA. It binds to the ribosome through its interaction with ribosomal protein L10 (Dijk *et al.* 1979, Pettersson 1979, Tokimatsu *et al.* 1981).

Protein L7/L12 is unique in that four copies are present per ribosome (Hardy 1975, Subramanian 1975). It is believed that these four copies are present in the form of two dimers on the ribosome, since protein L7/L12 readily

A. monomer — hinge

B. parallel dimer

C. staggered dimer

D. antiparallel dimer

E. **Sequence of the Eco L12 protein**

N-terminal domain — hinge

SITKDQIIEAVAAMSVMDVVELISAMEEKFGVSAAA AVAVAAGPVEAAEE

C-terminal domain

KTEFDVIKAAGANKVAVIKAVRGATGLGLKEAKDLVESAPAALKEGVSKDDAEALKKALEEAGAEVEVK

F.

Figure 7. Structure of the L7/L12 ribosomal protein from *Escherichia coli*. A. Monomer showing the three domains of the protein (Drawn after Liljas 1982) B. Parallel dimer (Drawn after Liljas 1982) C. Staggered dimer (Drawn after Maassen *et al.* 1981) D. Antiparallel dimer (Drawn after Gudkov *et al.* 1980) E. Sequence of the Eco L12 protein. The three domains are indicated (Terhorst *et al.* 1973) F. Structure of the C-terminal domain (Adapted from Spirin 1986) N: N-terminal domain, C: C-terminal domain, α: α helix, β: β sheet.

forms a dimer in solution (Möller *et al.* 1983). Support for the idea that the dimer is the active form of the protein on the ribosome, comes from the observation that when the dimerization of the protein is prevented by chemical modification, the monomer is unable to interact with the ribosome (Koteliansky *et al.* 1978).

The N-terminal region of the protein has been shown to be involved in the dimerization, since oxidation of the methionine residues at positions 14, 15 and 26 or cleavage of the first 26 amino acids with cyanogen bromide prevents the formation of the dimer (Gudkov and Behlke 1978, Gudkov *et al.* 1980). It is thought that hydrophobic interactions between the two N-terminal regions of the monomers are responsible for the dimerization (Gudkov *et al.* 1980, Luer and Wong 1979).

Physical studies of the dimer have shown that it is very elongated, with a maximum length of 135 Å (Österberg *et al.* 1976). Two different models have been proposed to explain the arrangement of the two monomers in the dimer.

Gudkov *et al.* (1980) have proposed that the monomers are arranged in an antiparallel fashion. In this model, the carboxy terminal halves of each monomer are folded into a globular structure that has a hydrophobic cavity into which the first eleven amino acids of the opposite monomer are inserted during dimerization. Residues 16 to 41 form an α helix and the helices of both monomers are held together by hydrophobic interactions (see Figure 7 D).

Luer and Wong (1979), and Liljas (1982) have proposed that the monomers are arranged in a parallel fashion, with the two N-termini located at one end and the two carboxy-termini at the other end (see Figure 7 B). As in the previous model, the two N-termini are held together by hydrophobic interactions.

Although the arrangement of the monomers in the dimers has not been definitively established, the parallel model seems to be the most likely for the following reasons: 1] circular dichroism spectra of the dimer show that there is an interaction between two phenylalanine residues that cannot be explained by the antiparallel model (Luer and Wong 1979), 2] the cross-linking observed between lysine 51 of one monomer and lysine 29 of the other one, cannot be explained by the antiparallel model since these two residues would be located too far away from each other for the cross-linking to occur (Maassen *et al.* 1981), 3] the crystalline dimer of the carboxy terminal fragment of protein L7/L12 suggests a parallel alignment (Liljas 1982).

The cross-linking between lysine 51 and lysine 29, led to the proposal that the monomers are aligned in a parallel but staggered fashion since in this way both residues would be located in a better position for the cross-linking to take place than in the symmetrical arrangement (see Figure 7 C) (Maassen *et al.* 1981). However, nuclear magnetic resonance data from the dimer, indicate that the monomers have a symmetrical arrangement (Bushuev *et al.* 1984). Since the parallel symmetrical model seems the most probable, the L7/L12 dimer is drawn in this way in Figure 8.

The location of the two L7/L12 dimers on the ribosome has been the subject of a lot of controversy. Based on the observation that antibodies against the Eco L7/L12 protein only bind to the stalk of the large subunit and that in some cases up to four IgG polyclonal antibodies could be observed binding at the tip of the stalk, several authors concluded that the two dimers are aligned parallel to each other in the stalk (Strycharz *et al.* 1978, Tokimatsu *et al.* 1981). However, energy transfer measurements between fluorescent probes attached to each dimer, suggested that there is a certain contact angle

between the two dimers (Zantema *et al.* 1982, Thielen *et al.* 1984). This led to the proposal that one dimer forms the stalk while the other is bent on the body of the large subunit (Zantema *et al.* 1982, Thielen *et al.* 1984). Support for this idea came from the observation of cross-links between the L7/L12 protein and the L5 protein which is located in the central protuberance away from the stalk (Traut *et al.* 1983) and from electron microscopy studies of reconstituted ribosomal subunits, which showed that only one dimer is needed to regenerate the stalk (Möller *et al.* 1983). Recently, Olson *et al.* (1986), using monoclonal antibodies against the N- and C-terminal regions of the Eco L7/L12 protein, were able to demonstrate that only one of the dimers forms the stalk, while the other dimer is located on the body of the large subunit (see Figure 8).

As mentioned before, the two L7/L12 dimers bind to the ribosome through their interaction with ribosomal protein L10 (Dijk *et al.* 19779, Pettersson 1979, Tokimatsu *et al.* 1981). The N-terminal domain of the L7/L12 monomers is responsible for this interaction (Gudkov *et al.* 1980, Schop and Maassen 1982). The binding sites of the dimers are located in the C-terminal region of protein L10 (residues 71-165) (Gudkov *et al.* 1980).

Reconstitution experiments have shown that the two binding sites are filled at equal rates. However, one binding site is stronger than the other since conditions have been found where one dimer can be selectively removed from the ribosome (Möller *et al.* 1983, Möller and Maassen 1986). The presence of a weak and a strong binding site has also been shown with the use of fluorescein-labeled L7/L12 dimers (Zantema *et al.* 1982) and with a monoclonal antibody directed against the N-terminal region of the Eco L7/L12 protein (Tewari *et al.* 1986, Traut *et al.* 1986). This antibody causes the release of one dimer, while the other remains bound to the ribosome together with the

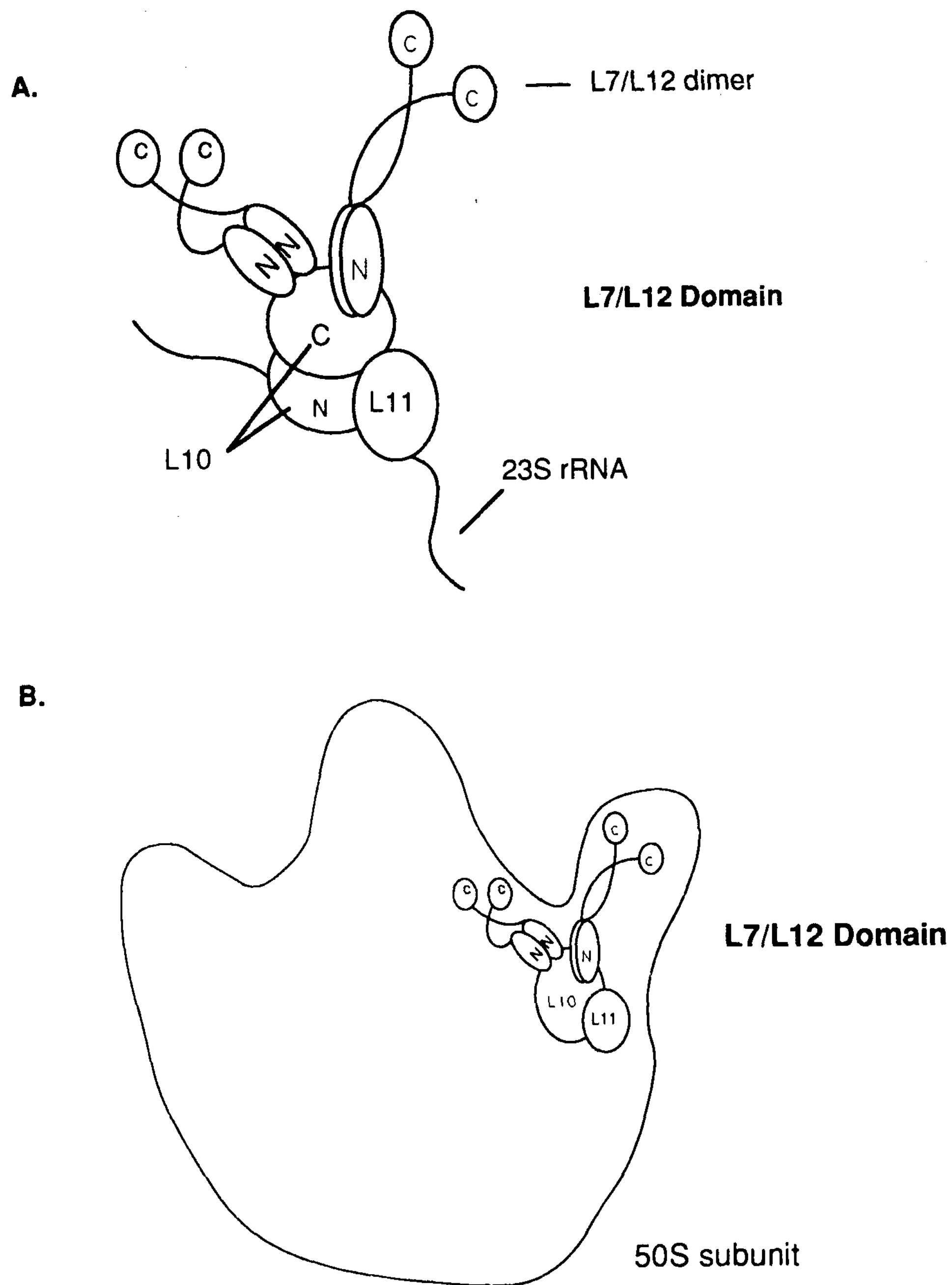


Figure 8. The L7/L12 domain. A. Structure of the L7/L12 domain. The L7/L12 dimer is shown in the parallel arrangement. B. Localization of the different components of the L7/L12 domain on the ribosome (Drawn after Olson *et al.* 1986). N: N-terminal domain, C: C-terminal domain.

antibody (Tewari *et al.* 1986, Traut *et al.* 1986).

The pentameric complex formed by the two L7/L12 dimers and one copy of protein L10 can be easily extracted from the ribosome (Pettersson *et al.* 1976, Pettersson and Liljas 1979, Pettersson 1979). In fact, at one time, the complex was erroneously identified as a ribosomal protein (L8) (Pettersson *et al.* 1976).

The L7/L12-L10 complex binds to the ribosome through the interaction of the N-terminal region of protein L10 with 23S rRNA (see Figure 8) (Gudkov *et al.* 1980). The binding site of this protein on the 23S rRNA has been determined. It includes nucleotides 1028-1124 and overlaps the binding site of another ribosomal protein: Eco L11 (residues 1052-1112) (Pettersson 1979, Dijk *et al.* 1979, Schmidt *et al.* 1981, Beauclerk *et al.* 1984). In fact, it has been found that there is mutual cooperativity in the binding of these two ribosomal proteins to the 23S rRNA (Petersson 1979, Dijk *et al.* 1979). The general structure of the whole domain is shown in Figure 8.

Acidic proteins that are equivalent to the L7/L12 *Escherichia coli* protein, have been isolated from the eukaryotes (Amons *et al.* 1979, 1982, Lin *et al.* 1982, Maassen *et al.* 1985, Wigboldus 1987, Beltrame and Bianchi 1987, Qian *et al.* 1987, Rich and Steitz 1987, Ferro and Reinach 1988, Mitsui and Tsurugi 1988 b, c, Remacha *et al.* 1988, Shimmin *et al.* 1989a). Like their eubacterial counterparts, these proteins interact with the elongation factors and are responsible for triggering their GTPase activity (MacConnell and Kaplan 1980, 1982, Sánchez-Madrid *et al.* 1979).

In the eukaryotes, these proteins can be phosphorylated, and at least in yeast, the phosphorylation seems to regulate the affinity of the proteins for the ribosome (Sánchez-Madrid *et al.* 1981b, Juan-Vidales *et al.* 1984).

The eukaryotic acidic ribosomal proteins can be divided into two families designated P1 and P2. The P1 family has a conserved tryptophan residue in the N-terminal region (position 43 in Figure 27 Results and Discussion) and only one tyrosine (Amons *et al.* 1982, Maassen *et al.* 1985, Wigboldus 1987, Ferro and Reinach 1988, Rich and Steitz 1987, Remacha *et al.* 1988, Mitsui and Tsurugi 1988 b, Shimmin *et al.* 1989a). The P2 family lacks the tryptophan residue and has two tyrosines in the N-terminal region (see Figure 26 Results and Discussion) (Amons *et al.* 1979, Lin *et al.* 1982, Maassen *et al.* 1985, Beltrame and Bianchi 1987, Qian *et al.* 1987, Rich and Steitz 1987, Mitsui and Tsurugi 1988 c, Remacha *et al.* 1988, Shimmin *et al.* 1989a). Both families have very similar highly charged C-terminal domains (see Figures 26 and 27 Results and Discussion) (Amons *et al.* 1979, 1982, Lin *et al.* 1982, Maassen *et al.* 1985, Wigboldus 1987, Beltrame and Bianchi 1987, Qian *et al.* 1987, Rich and Steitz 1987, Ferro and Reinach 1988, Mitsui and Tsurugi 1988 b, c, Remacha *et al.* 1988, Shimmin *et al.* 1989a).

All the eukaryotes examined until now, except yeast, have two identical proteins from each family (Amons *et al.* 1979, 1982, Lin *et al.* 1982, Maassen *et al.* 1985, Wigboldus 1987, Beltrame and Bianchi 1987, Qian *et al.* 1987, Rich and Steitz 1987, Ferro and Reinach 1988, Mitsui and Tsurugi 1988 b, c, Remacha *et al.* 1988, Shimmin *et al.* 1989a). Yeast, and probably other fungi (*Hansenula anomala, Geothricum lactis*) as well, have two different proteins from each family (Remacha *et al.* 1988, Shimmin *et al.* 1989a). The reason for this difference is not known.

The eukaryotic acidic proteins, like their eubacterial counterparts, are also very flexible (Gudkov *et al.* 1981, Cowgill *et al.* 1984). In fact, they have an alanine rich region located in front of the highly charged C-terminal domain,

that resembles the hinge region of the eubacterial L12 proteins (see Figures 26 and 27, Results and Discussion) (Amons *et al.* 1979, 1982, Lin *et al.* 1982, Maassen *et al.* 1985, Wigboldus 1987, Beltrame and Bianchi 1987, Qian *et al.* 1987, Rich and Steitz 1987, Ferro and Reinach 1988, Mitsui and Tsurugi 1988 b, c, Remacha *et al.* 1988, Shimmin *et al.* 1989a, Bushuev *et al.* 1989).

The eukaryotic acidic proteins can also form dimers in solution (Gudkov *et al.* 1981). Although there are no models regarding the arrangement of the eukaryotic proteins in the dimer, it is known that the P1 and P2 proteins only form homodimers, since heterodimers could not be detected by cross-linking (Uchiumi *et al.* 1987).

In the eukaryotes, the two dimers have been shown to bind to the ribosome through their interaction with a protein called P0, which appears to have the same role as protein L10 in *Escherichia coli* (Uchiumi *et al.* 1987). Thus the general structure of the stalk domain seems to have been conserved in these two kingdoms.

Acidic proteins equivalent to the eubacterial and eukaryotic L12 proteins have also been isolated from the archaebacteria (Matheson *et al.* 1980, 1988, Strobel *et al.* 1988, Itoh *et al.* 1988, Shimmin and Dennis 1989). Sequence comparison of the acidic proteins from the three kingdoms has revealed that these proteins can be divided into two groups: one formed by the eukaryotic-archaebacterial acidic proteins and another by the eubacterial proteins (for a review, see Matheson 1985, Wittmann-Liebold 1986). Alignment between the proteins from both groups has proved to be difficult, and several models have been proposed in which different regions of the proteins are rearranged to obtain an alignment (Amons *et al.* 1979, Yaguchi, *et al.* 1980, Lin *et al.* 1982, Matheson 1985, Otaka *et al.* 1985, Liljas *et al.* 1986b, Shimmin *et al.* 1989b).

These models will be discussed in detail in the section on the Sso L12 gene (Results and Discussion). However, even though it is difficult to align the proteins from both groups, there is some evidence that the eukaryotic and eubacterial proteins are functionally interchangeable (Stöffler *et al.* 1974, Wool and Stöffler 1974, Möller *et al.* 1975, Sánchez-Madrid *et al.* 1981b).

From a genetic point of view, the genes that code for these proteins are also interesting since they are organized in a very different way in eubacteria and in eukaryotes. In *Escherichia coli,* the genes that code for ribosomal proteins L12, L10 and L11 are located together with the genes for another ribosomal protein: L1, and the β and β' subunits of RNA polymerase in the *rif* region of the chromosome (89 min) (Post *et al.* 1979, Nomura *et al.* 1984). They are organized in two operons: the L11 operon formed by the L11 and L1 genes and the L10 operon containing the genes for L10 and L12 as well as the genes for the two subunits of the RNA polymerase (Post *et al.* 1979, Nomura *et al.* 1984). In contrast, in the eukaryotes, the genes that code for the acidic proteins are dispersed through the genome and each one has its own promoter (Maassen *et al.* 1985, Qian *et al.* 1987, Remacha *et al.* 1988, Shimmin *et al.* 1989a). In yeast and *Drosophila melanogaster,* there is only one copy of each gene, while in *Artemia salina* there appears to be 4 or 5 copies of each gene (Maassen *et al.* 1985, Qian *et al.* 1987, Remacha *et al.* 1988). One of the genes that codes for the acidic proteins in yeast has an intron (Remacha *et al.* 1988, Shimmin *et al.* 1989a). In *Drosophila*, the only gene that has been studied is intronless (Qian *et al.* 1987) and in *Artemia* the presence of introns has not been established (Maassen *et al.* 1985).

Since the structure of the components of the L7/L12 domain is so different in the eubacteria and the eukaryotes, one of the main objectives of this

dissertation was to determine the sequence of the different components of this domain in the archaebacteria and compare them to those from the other two kingdoms. A second objective was to determine the organization of the genes that code for the components of this domain and study their transcription, in order to get a better understanding of gene expression in the archaebacteria.

Materials and Methods

All the glassware, pipette tips and Eppendorf tubes were autoclaved. All the solutions were prepared with deionized, distilled water and autoclaved or filtered sterilized. Phenol, buffers, culture media, antibiotics and other solutions commonly used in molecular biology experiments were prepared and stored as described by Maniatis *et al.* (1982). General chemical reagents were obtained from Sigma, BDH Chemicals or Fisher Scientific. Ultrapure phenol and agarose for molecular biology were obtained from Bethesda Research Laboratories. Acrylamide, N,N'-Methylene bisacrylamide, sodium dodecylsulfate (SDS), ammonium persulfate and ultrapure urea were purchased from Serva. Tryptone, yeast extract, casamino acids and agar were obtained from Difco. N-Z amine was purchased from ICN Nutritional Biochemicals. Isopropylthiogalactoside (IPTG), vitamin B1, ribonuclease A and lysozyme were obtained from Sigma. Actinomycin D, proteinase K, tRNA and 5-dibromo 4-chloro 3-indolylgalactoside (Xgal) were purchased from Boehringer Mannheim. Calf thymus DNA, deoxyadenosine 5'-triphosphate (dATP), deoxyguanosine 5'-triphosphate (dGTP), deoxythymidine 5'-triphosphate (dTTP), deoxycytidine 5'-triphosphate (dCTP), dideoxyadenosine 5'-triphosphate (ddATP), dideoxyguanosine 5'-triphosphate (ddGTP), dideoxythymidine 5'-triphosphate (ddTTP) and dideoxycytidine 5'-triphosphate (ddCTP) were obtained from Pharmacia. M13 universal sequencing primer was purchased from New England Biolabs. The source of enzymes, radionucleotides and other special chemical reagents will be described in the following sections.

Growth of *Sulfolobus solfataricus*

Sulfolobus solfataricus (strain P1) cells were a gift of Dr. Norman Pace. They were grown (500 ml cultures) at 85°C and pH 4.0 in a medium containing per liter: 1.0 g yeast extract, 1.0 g tryptone, 3.1 g KH_2PO_4, 0.75 g KCl, 2.5 g $(NH_4)_2SO_4$, 0.2 g $MgSO_4.7H_2O$, 0.25 g $CaCl_2$ and 1 ml of a trace metal mix (Pace, personal communication). The trace metal mix contained per 100 ml: 180 mg $MgCl_2.4H_2O$, 450 mg $Na_2B_4O_7.10H_2O$, 22 mg $ZnSO_4.7H_2O$, 5 mg $CuCl_2.2H_2O$, 3 mg $NaMoO_4$, 3 mg $VOSO_4.2H_2O$ and 1 mg $CoSO_4.7H_2O$. The pH was monitored during growth and adjusted to 4.0 with H_2SO_4. When the cultures reached mid log phase (A_{540nm} 0.3) (for RNA extraction) or late log (A_{540nm} 0.6), they were rapidly cooled on ice to 4°C and harvested by spinning in a GSA rotor at 5000 rpm for 10 min at 0°C in a Sorvall RC-5B centrifuge. The cells were either used immediately or stored at -70°C

Other Archaebacterial Cells

Samples of *Methanobacterium thermoautotrophicum* and *Thermoproteus tenax* were a gift from Dr. W. Zillig. The different archaebacterial cells were kept frozen at -70°C until their DNA was extracted.

Bacterial Strains and Vectors

Table 3 shows the bacterial strains and Table 4 shows the vectors, used in this work. *Escherichia coli* Q358, Q359 and phage EMBL 3 were provided by Dr. P. Dennis. Plasmid pUC 18 was a gift of Dr. P. Romaniuk. *Escherichia coli* JM103, JM105 and phage M13mp7 were obtained from Pharmacia.

Table 3 Bacterial Strains

Strain	Genotype	Reference
Q358	*hsdR, supE, ton A*	Karn *et al.* 1980
Q359	*hsdR, supE, tonA* (P2)	Karn *et al.* 1980
JM103	*supE, thi, Δ(lac-proAB), strA, sbcB*15, *endA, hsdR*4, [F', *traD*36, *proAB, lacI* ^{q} ZΔM15]	Messing *et al.* 1981
JM105	*thi, rpsL, endA, sbcB*15, *hsdR*4, *Δ(lac-proAB)*, [F', *traD*36,*proAB, laqI* ^{q} ZΔM15]	Yanisch-Perron *et al.* 1985

Table 4 Vectors

Vector	Size (Kb)	Recombinant selection	Reference
phage EMBL3	44.0	*spi* (sensitive to P2 interference)	Frischauf *et al.* 1983
phage M13mp7	7.2	inactivation of the β-galactosidase gene, white plaques on xgal-IPTG plates	Messing *et al.* 1981
plasmid pUC 18	2.7	inactivation of the β galactosidase gene, white colonies on xgal-IPTG plates	Yanisch-Perron *et al.* 1985

Growth of the Different Bacterial Strains and Vectors

Q358, Q359, EMBL 3 and Recombinant Phages

Q358 and Q359 were grown with aeration at 37°C in NZYC media (Maniatis *et al.* 1982). Phage EMBL 3 was grown as a liquid lysate by infecting a stationary culture of Q358 at 37°C (Maniatis *et al.* 1982). Recombinant phages were grown in the same way except that Q359 was used as the host (Karn *et al.* 1980, Frischauf *et al.* 1983). After lysis of the culture, the phages were precipitated and their DNA extracted following the procedure described by Maniatis *et al.* (1982).

JM103, JM105, M13mp7 and pUC 18

JM103 was grown in M9 minimal salt medium with glucose and vitamin B1 in the absence of amino acids to prevent the loss of the F' episome which is needed for phage infection (Messing 1983).

M13mp7 and recombinant phages were grown by infecting exponentially growing JM103 cells in 2x YT media as described by Messing (1983). Replicative form phage DNA was isolated using the method for plasmid purification of Birnboim and Doly (1979).

JM105 cells were grown in LB media at 37°C as described by Messing (1983). Cells containing either pUC 18 or recombinant plasmids were grown at 37°C in LB media containing 50 µg/ml of ampicillin (Sigma) (Yanisch-Perron *et al.* 1985). Plasmids were extracted by the procedure of Birnboim and Doly (1979).

DNA Extraction

DNA from *Sulfolobus solfataricus*, *Methanobacterium thermoautotrophicum*, and *Thermoproteus tenax* was extracted following the procedure of

Marmur (1961). Briefly, the cells were suspended in 150 mM NaCl, 100 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0 and 2% SDS. The solution was boiled for 3 minutes and rapidly cooled on ice. Cell debris was spun down in a SS34 rotor at 5000 rpm for 5 minutes at 10°C in a Sorvall RC-5B centrifuge. The supernatant was removed and extracted two times with phenol and three times with chloroform. After extraction, the DNA was purified by centrifugation through a CsCl gradient as described by Maniatis *et al.* (1982). DNA was dissolved in 1 ml 10 mM Tris-HCl, 1 mM EDTA, pH 7.5 and stored at 4°C.

RNA Extraction

All the glassware and all the solutions were first treated with diethyl pyrocarbonate (DEPC) (Sigma) and then autoclaved to eliminate ribonucleases (Blumberg 1987). RNA was extracted according to the hot phenol (60°C) method of Aiba *et al.* (1981). RNA was dissolved in 1 ml of DEPC-treated sterile water and stored at -20°C.

Restriction Endonuclease Digestions

Digestion of DNA from *Sulfolobus*, other archaebacterial cells, vectors and clones with restriction endonucleases, was performed in the buffers and under the conditions suggested by the manufacturer. Table 5 shows the enzymes used in this work, their recognition sequence and their source.

Table 5 Restriction Endonucleases

Enzyme	Recognition sequence *	Source
Alu I	AG/CT	Pharmacia
Bam HI	GG/GATCC	Pharmacia
Dde I	C/TNAG	Pharmacia
Dra I	TTT/AAA	Pharmacia
Eco RI	G/AATTC	Pharmacia
Hind III	A/AGCTT	Pharmacia
Hinf I	G/ANTC	Boehringer Mannheim
Nhe I	G/CTAGC	Boehringer Mannheim
Pst I	CTGCA/G	Pharmacia
Ssp I	AAT/ATT	Biolabs
Xho I	C/TCGAG	Pharmacia
Xho II	Pu/GATCPy	Boehringer Mannheim

* N any nucleotide, Pu purines, Py pyrimidines. A dash indicates the position of the cut.

When it was necessary to perform double digests, the DNA was ethanol precipitated after the first digestion, before resuspending in the appropiate buffer for the second enzyme. Ethanol precipitations were performed by the method described by Maniatis *et al.* (1982) for 30 minutes at -70°C.

Synthesis of Probes and Primers

Primers and probes were kindly synthesized by Dr. P. Romaniuk on a Biosearch 8600 DNA Synthesizer, using the phosphite method (Sinha *et al.* 1984). After the synthesis was completed, the oligonucleotides were

deprotected by incubating them in 1 ml of concentrated ammonium hydroxide solution, first at room temperature for one hour and then at 50°C overnight (Biosearch Model 8600 Instruction Manual 1985). After the incubation, the ammonium hydroxide solution was evaporated on a RH40-11 Speed Vac Concentrator. The dry oligonucleotides were dissolved in 100 µl sterile distilled water and their concentration determined by reading their absorbance at 260 nm. Ten O.D. units of each oligonucleotide were purified by polyacrylamide gel electrophoresis and reverse phase chromatography in a C18 Sep-Pak column (Waters Associates) as described by Atkinson and Smith (1984). After purification, the oligonucleotides were stored at -20°C.

Labeling of Probes

Oligonucleotide probes were 5'-end labeled with T4 polynucleotide kinase (Pharmacia) and γ^{32}P ATP (adenosine 5'-triphosphate, 3000 Ci/mmol, New England Nuclear) following the procedure described by Maniatis *et al.* (1982). After labeling, 10 µl of the reaction were diluted to a final volume of 200 µl with water. One microliter of this dilution was dissolved in 4 ml aquasol (New England Nuclear) and counted in a Beckman LS 8100 liquid scintillation counter. Aliquots containing between 10 to 50 million cpm were used for hybridizations (see Dot Blots, Southern Blots, Unblots and Northern Blots).

Dot Blots

Dot blot hybridizations were performed in order to determine the conditions under which the oligonucleotide probe mixes for the Sso L12 gene would bind stringently to *Sulfolobus solfataricus* DNA. DNA from *Sulfolobus* and lambda phage (as a control) was spotted directly onto nitrocellulose filters (Fisher

Scientific). The filters were soaked two times in 0.5 M NaOH, 1.5 M NaCl for two minutes to denature the DNA and neutralized by soaking twice for two minutes in 0.5 M Tris-HCl, pH 7.5, 3 M NaCl. The filters were then baked for 90 minutes in an 80°C oven to fix the DNA to the membrane

Prehybridization of the filters was carried out for 3 hours at 65°C in 5 ml of 6x SSC (sodium chloride, sodium citrate) (1x SSC: 0.15 M NaCl, 0.015 M sodium citrate) and 10x Denhardt's reagent (50 x Denhardt contains per 100 ml: 1g Ficoll, 1 g polyvinylpyrrolidone, 1 g bovine serum albumin) (Maniatis *et al.* 1982). Hybridization was performed in the prehybridization solution with 50 million counts per minute of radiolabeled probe for 24 hours at 37°C. After hybridization, the filters were washed following the scheme shown in Table 6.

Table 6 Different Washing Conditions for the Dot Blot Hybridization Filters

filter	solution	time	temperature
1	6X SSC	10 min	room temperature
	6X SSC	10 min	room temperature
	6X SSC	10 min	room temperature
2	6X SSC	10 min	37°C
	1X SSC	10 min	37°C
	0.2X SSC	10 min	37°C
3	6X SSC	10 min	40°C
	1X SSC	10 min	40°C
	0.2X SSC	10 min	40°C
4	6X SSC	10 min	43°C
	1X SSC	10 min	43°C
	0.2X SSC	10 min	43°C

After washing, the filters were air dried and exposed to X-Omat-AR x ray film (Kodak) with an intensifying screen (DuPont Cronex Lightning Plus) at -70°C overnight.

Dot blots were also used to determine if the Sso L12 probe would bind to DNA from other archaebacteria. In this case, DNA from *S. solfataricus, Methanobacterium thermoautotrophicum, Thermoproteus tenax*, λ phage and phage E3CR-J (a recombinant phage carrying the Sso L12 gene) was spotted onto nitrocellulose filters. The filters were then processed and hybridized as described before.

Southern Blots

The fragments produced by the digestion of *Sulfolobus* DNA or clones, with different restriction endonucleases, were separated according to size in 0.5% to 1.0% agarose gels. Electrophoresis was performed at 50 mA for 12 hours, in Tris-acetate-EDTA buffer (TAE: 0.04 M Tris-acetate, 0.002 M Na_2 EDTA, pH 8) containing 25 µl/l of a 10 mg/ml solution of ethidium bromide in water. After electrophoresis, the gels were photographed under U.V. light with Polaroid 667 film and a yellow filter. The gels were then soaked for 30 minutes in 0.05 M NaOH, 1.5 M NaCl to denature the DNA and neutralized by soaking for another 30 minutes in 0.5 M Tris-HCl, pH 7.5, 3 M NaCl. Capillary transfer of the DNA to the nitrocellulose filter was accomplished as described by Southern (1975). After transfer, the DNA was fixed to the filter by baking it for 90 minutes in an 80°C oven. Filters were prehybridized in 10 ml of 6x SSC, 10x Denhardt for 3 hours at 65°C. Hybridization was carried out in the prehybridization solution with 50 million cpm of radiolabeled probe at 37°C overnight. After hybridization, the filters were washed under stringent conditions at 40°C

following the scheme shown in Table 6 for filter 3. The filters were air dried and exposed to X-Omat-AR x ray film with an intensifying screen at -70°C overnight.

Unblots

Gels were run and photographed as described in the previous section. The gels were denatured in 0.5 M NaOH, 150 mM NaCl for 30 minutes and neutralized in 0.5 M Tris-HCl, pH 8, 150 mM NaCl for another 30 minutes (Tsao *et al.* 1983). They were then placed on two sheets of Whatman 3MM paper, covered with plastic wrap and dried using only the vacuum in a Hoefer Scientific SE 540 Slab Gel Drier until the gel was nearly flat. The heater was then turned on and the gel dried for another hour (Wallace and Miyada 1987). Before hybridization, the gels were soaked in distilled water and the Whatman paper was removed. Hybridization was performed in 15 ml of a solution containing: 0.9 M NaCl, 50 mM sodium phosphate, 5 mM Na_2EDTA, pH 8 and 0.1% SDS (Miyada *et al.* 1985) with 50 million cpm of radiolabeled probe at 37°C overnight. After hybridization, the gels were washed under stringent conditions at 40°C following the scheme shown in Table 6 for filter 3. The gels were air dried, covered in plastic wrap and exposed to X-Omat-AR x ray film with an intensifying screen for 3 to 5 hours at -70°C.

Northern Blots

Sulfolobus RNA was separated on 1% agarose-formaldehyde gels and transferred to Z-Probe membranes (BioRad) as described by Fourney *et al.* (1988). Filters were prehybridized and hybridized as described by Miyada and Wallace (1987) when oligonucleotide probes were used and as described by

Fourney *et al.* (1988) in the case of longer probes (>50 nucleotides long) (see Probes for S1 Mapping and Northern Blots, p. 75).

In the case of oligonucleotide probes, hybridization temperature was T-12°C , where T= 4° (G+C) + 2° (A+T) (Miyada and Wallace 1987). After hybridization, the filters were washed 3 times in 6 x SSC for 15 minutes at room temperature, followed by a 1.5 minute and a 1 minute wash at the stringent temperature. Stringent temperature was T-5°C (Miyada and Wallace 1987).

Hybridization with longer probes was performed at 42°C. After hybridization, the filters were washed twice in 1 X SSC, 0.1% SDS for 20 minutes and twice with 0.1 X SSC, 0.1% SDS at 50°C for 20 minutes (Fourney *et al.* 1988). After washing, the filters were exposed to X-OMAT-AR x ray film, with an intensifying screen, at -70°C overnight.

Construction of a Genomic Library of *Sulfolobus solfataricus* in EMBL 3

DNA from phage EMBL 3 was digested with Bam HI and Eco RI. The digestion with Eco RI was performed in this case to reduce the formation of wild type phage during ligation as suggested by Frischauf *et al.* (1983). The small fragments produced by the digestion of the vector with Eco RI, were selectively removed by precipitation with 0.15 volumes 3M sodium acetate and 0.75 volumes isopropanol on ice for 5 minutes (Frischauf *et al.* 1983). *Sulfolobus* DNA was digested with Bam HI and mixed in a one to one weight ratio with digested phage DNA (Frischauf *et al.* 1983). The mix was ethanol precipitated, resuspended in ligase buffer and ligated with T4 DNA ligase (New England Biolabs) at 16°C overnight (Jessee 1984). After ligation, the DNA was

packaged *in vitro* to yield viable phages and the library amplified by plating it on a lawn of Q359 in NZYC soft agar plates as described by Maniatis *et al.* (1982). After amplification, the library was recovered and titered as described by Maniatis *et al.* (1982).

Screening of the *Sulfolobus* Library

The library was screened by plaque hybridization (Benton and Davies 1977) with the radiolabeled probe for the Sso L12 gene. Recombinant phages were plated on a lawn of Q359 in NZYC soft agar at a density of approximately 2000 plaques/petri dish and grown overnight at 37°C. The next day, the plates were stored for 3 hours at 4°C before the plaques were transferred to Colony/Plaque hybridization membranes (New England Nuclear). The filters were placed on the surface of the agar plates and left in contact with it for 3 minutes. They were then removed and the DNA from the plaques denatured and neutralized as described for the dot blots. Due to the type of membrane used, in this case it was not necessary to fix the DNA onto the filter by baking. The filters were prehybridized in 6x SSC 10x Denhardt, and 0.5% SDS for 17 hours at 65°C. Hybridization was carried out in the prehybridization solution, with 10 million cpm of radiolabeled probe for 24 hours at 37°C. After hybridization, the filters were washed under stringent conditions following the scheme shown in Table 7.

After washing, the filters were air dried and exposed to X-Omat-AR x ray film with an intensifying screen at -70°C for 4 days. Positive clones were identified and isolated. They were propagated in small-scale liquid cultures of Q359 and their DNA was extracted as described by Maniatis *et al.* (1982). Phage DNA was spotted onto nitrocellulose filters and hybridized to the Sso L12 gene

probe under the conditions described in the section on Dot Blots to check that the recombinant phages isolated were indeed true positive clones. A positive clone designated E3CR-J was identified in this way.

Table 7 Stringent Condition Washes of the Plaque Hybridization Filters

Solution	Time	Temperature
6X SSC	10 min	room temperature
6X SSC-0.5% SDS	10 min (twice)	37°C
1X SSC-0.3% SDS	10 min	37°C
0.2X SSC	10 min	37°C

Subcloning of a 1.1 Kb Pst I-Pst I Fragment into M13mp7

A 1.1 Kb Pst I-Pst I fragment hybridizing to the Sso L12 gene probe was subcloned into M13mp7 for sequencing. Phage E3CR-J was digested with Pst I and the fragments were separated on a 4% non-denaturing polyacrylamide gel (1.5 mm thick), using ΦX174 RF Hae III fragments (New England Biolabs) as molecular weight markers (Maniatis *et al.* 1982). Electrophoresis was performed in 1x Tris-Borate-EDTA buffer (1x TBE: 0.089 M Tris, 0.089 M boric acid, 0.008 M Na_2EDTA) (Maniatis *et al.* 1982), at 40 volts overnight. The gel was stained in 250 ml of 1x TBE containing 250 mg of ethidium bromide for 15 minutes. After staining, the gel was observed under U.V. light and the 1.1 Kb fragment cut out of the gel with a scalpel. The fragment was then electroeluted into a dialysis bag as described by Maniatis *et al.* (1982) except that electrophoresis was performed for 2 hours at 200 volts. In order to minimize the risk of contamination with DNAses, the buffer and the dialysis tubing were

sterilized before use. After electroelution, the buffer inside the dialysis bag was removed, extracted twice with phenol/chloroform (1:1 vol) and ethanol precipitated at -70°C overnight, to recover the DNA.

M13mp7 replicative form (RF) DNA was digested with Pst I and dephosphorylated with calf intestine alkaline phosphatase according to the procedure of Maniatis *et al.* (1982). It was then mixed in a 1:2 weight ratio with the 1.1 Kb Pst I-Pst I fragment, and the mixture was ethanol precipitated and ligated with T4 polynucleotide ligase at 16°C overnight (Jessee 1984). The ligation mix was then used to transfect JM103 competent cells following the procedure of Messing (1983). After transfection, the cells were mixed with 200 µl of fresh exponentially growing JM103 and plated on IPTG-Xgal-LB plates as described by Messing (1983). Noninfected cells are needed for plaque formation because a plaque formed by M13-infected cells is simply a zone of infected cells within a lawn of normal cells. Infected cells are distinguished from normal cells by their slower growth rate, which is twice as long as that of noninfected cells (Messing 1983).

White plaques produced by recombinant phages were isolated and used to infect small scale cultures of exponentially growing JM103 (Messing 1983). Replicative form DNA was isolated by the procedure of Birnboim and Doly (1979) and digested with Pst I. After digestion, the fragments were sized on 1.2% agarose gels (in 1x TBE) at 100 volts for 30 minutes to determine which clones had the 1.1 Kb insert. Several clones were then selected and ran on a 0.5% agarose gel, transferred to nitrocellulose filters and hybridized to the 17A probe as described in the section on Southern Blots, to determine which clones had the insert in opposite orientations. Two clones M13CR-3 and M13CR-8 were identified in this way.

Subcloning of a 6.9 Kb Eco RI-Bam HI Fragment into pUC 18

A 6.9 Kb Eco RI-Bam HI fragment hybridizing with the Sso L12 gene probe was subcloned into pUC 18 for sequencing. DNA from phage E3CR-J was digested with Bam HI and Eco RI. The fragments were separated on a 4% non-denaturing polyacrylamide gel (3 mm thick, 0.5 x TBE, 150 volts overnight), using λ Hind III fragments (Pharmacia) as molecular weight markers; and the 6.9 Kb Eco RI-Bam HI fragment was isolated and purified in the same way as the 1.1 Kb Pst I-Pst I fragment (see previous section).

After purification, the fragment was mixed in a 2 to 1 molar ratio with pUC 18 DNA that had been previously digested with Bam HI and Eco RI. The mixture was ethanol precipitated, resuspended in ligase buffer and ligated with T4 DNA ligase at 16°C overnight (Jessee 1984). This ligation mix was then used to transform competent JM105 cells following the procedure of Hanahan (1983). After transformation, 100 µl of transformed cells were plated on IPTG-Xgal-ampicillin LB plates (Yanisch-Perron *et al.* 1985) and incubated overnight at 37°C.

White colonies, containing plasmids with inserts, were picked and grown in LB-ampicillin media (see Growth of Different Bacterial Strains and Vectors). Plasmids were extracted by the method of Birnboim and Doly (1979), cut with Bam HI and Eco RI and ran in a 0.5% agarose gel (1x TBE, at 100 mA for 30 minutes) to check that the 6.9 Kb fragment was indeed present. A positive clone designated p18CR-9 was identified in this way.

Sequencing

Sequencing was performed using the dideoxynucleotide chain-termination procedure of Sanger *et al.* (1977). The composition of the sequencing mixes is

shown in Table 8 (Davies 1982).

Table 8 Sequencing Mixes*

	A	C	T	G
dCTP	62.5 µM	4 µM	90 µM	90 µM
dGTP	62.5 µM	90 µM	90 µM	4 µM
dTTP	62.5 µM	90 µM	4 µM	90 µM
ddNTP	250 µM	90 µM	250 µM	125 µM
Tris pH8	3.1 mM	4.5 mM	4.5 mM	4.5 mM
EDTA	0.06 mM	0.09 mM	0.09 mM	0.09 mM

* from Davies (1982). The final concentration in the mix is indicated.

Single stranded phage DNA (1.1Kb Pst I-Pst I fragment cloned into M13mp7) and denatured plasmid DNA (6.9 Kb Eco RI-Bam HI fragment cloned into pUC 18 and deletions plasmids) were used as templates. Phage DNA was extracted and purified as described by Messing (1983) except that the initial culture was scaled up to 100 ml to increase the yield of single stranded DNA. Plasmid DNA for sequencing was extracted and purified following the method of Krieg and Melton (1985). After purification, plasmid DNA was alkaline denatured as described by Hattori and Sakaki (1986).

Annealing of the primer to the template was performed in 10 µl of 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$ and 1 mM DTT. A 1 to 2 molar ratio (picomoles) of template to primer was used in the case of phage DNA (Messing 1983) and a 1 to 5-10 molar ratio (picomoles) of template to primer in the case of plasmid DNA (Heinrich 1986). Annealing was performed by

incubating the primer-template mix at 65°C for 15 minutes and then at room temperature for another 15 minutes.

After annealing, 40 µCi of α ^{32}P dATP (400 Ci/mmol, Amersham) were added to the primer-template mix. Three microliters of this mix were then added to four tubes containing the respective sequencing mix (C, T, A or G) and one unit of the Klenow fragment of DNA polymerase I (Pharmacia). The tubes were then incubated at 37°C. After 30 minutes, 2 µl of a chase solution (0.5 mM dATP, dGTP, dCTP and dTTP) were added to the tubes and the incubation was continued for another 30 minutes. The reactions were stopped with 6 µl of formamide dye (90% formamide, 0.25% bromophenol blue, 0.25% xylene cyanol and 10 mM EDTA, pH 7.5). The samples were boiled for 2 minutes, rapidly cooled on ice and 2.5 µl of each reaction was loaded on a 4 mm thick, 6% polyacrylamide-8 M urea sequencing gel (Davies 1982). Electrophoresis was performed at 57 watts in sequencing electrophoresis buffer (135 mM Tris, 45 mM boric acid, 2.5 mM Na_2EDTA, pH 8.9) (Anderson 1981) in an SE 1500 Poker Face Electrophoresis Apparatus (Hoefer Scientific Instruments). In order to increase the amount of sequencing data obtained from each experiment, the samples were loaded 3 times at 2 hour intervals. After electrophoresis, the gel was transferred to Whatman 3MM paper and dried in a BioRad 483 slab gel dryer for 1 hour. It was then exposed to X-Omat-K x ray film at room temperature, overnight without an intensifying screen.

In the case of the 6.9 Kb Eco RI-Bam HI fragment, a series of deletion plasmids was generated from the Bam HI site by partial digestion of the original clone (p18CR-9) with DNAse I (Boehringer Mannheim) following the procedure of Lin *et al.* (1985).

Primer Extension

Primers were 5' end labeled with T4 polynucleotide kinase and γ^{32}P ATP (3000 Ci/mmol) as described by Maniatis *et al.* (1982). After labeling, the primers were purified in 20% polyacrylamide-8 M urea gels as described by Atkinson and Smith (1984). The gels were exposed for 20 seconds to X-Omat-AR x ray film to determine the position of the labeled primers. The bands were cut out of the gel and eluted in 250 µl of 0.5 M ammonium acetate, 10 mM magnesium acetate overnight. After elution, the primers were precipitated with 3 volumes of 95% ethanol at -70°C overnight. Primers were dissolved in 30 µl water. Between 1-5 million cpm of the primer were hybridized to 10 µg of *Sulfolobus* RNA in 10 mM Tris/HCl, pH 7.5, 40 mM KCl (final volume 10 µl). This mixture was denatured at 95°C for 1 minute, transferred to a 65°C heating block and allowed to cool down to 45°C to anneal the primer to the RNA. Primer extension was performed under the conditions described by Reiter *et al.* (1987a) with 25 units of reverse transcriptase (Life Sciences) at 37°C or 50°C (Geliebter 1987) for 1 hour. The products of the reaction were ethanol precipitated and dissolved in 10 µl formamide dye. They were run on 6% polyacrylamide-8 M urea sequencing gels using the same electrophoresis conditions as described for the sequencing gels. Sequencing reactions using the same labeled primer with either DNA or RNA (Geliebter 1987) as a template, were used to determine the size of the products. Gels were dried and exposed to X-Omat-K x ray film with an intensifying screen at -70°C overnight.

S1 Mapping

Molecular Weight Markers

Phage ΦX174 RF Hae III digest (New England Biolabs) (11 fragments: 1353, 1078, 872, 603, 310, 278, 271, 234, 194, 118 and 72 base pairs) and pUC 19 digested with Dde I (9 fragments: 910, 540. 426, 409, 235 and 166 base pairs) were used as molecular weight markers for S1 mapping. Molecular weight markers were dephosphorylated with calf intestine alkaline phosphatase (Boehringer Mannheim) and 5'-end labeled with T4 polynucleotide kinase and $\gamma^{32}P$ ATP (3000 Ci/mmol) following the procedure described by Maniatis *et al.* (1982). After labeling, molecular weight markers were ethanol precipitated, counted by Cerenkov in a Beckman LS 8100 liquid scintillation counter and dissolved in water to give approximately 5000 cpm/fragment.

Probes for S1 Mapping and Northern Blots

Single stranded, 5' labeled probes for S1 mapping and Northern Blots were generated by extension of 5' labeled primers annealed to single stranded plasmid DNA (clone p18CR-9) as described by Reiter *et al.* (1987c). After extension, the double stranded DNA was cut with a restriction endonuclease to generate a defined 3' end. The fragments were then separated on 6% sequencing gels, the bands identified by autoradiography, cut out of the gel and electroeluted in 0.5 X TBE buffer at 200 volts for 1 hour. The eluate was then extracted twice with phenol/ chloroform and the DNA was ethanol precipitated. The samples were then counted by Cerenkov and dissolved in DEPC-treated water. Table 9 shows the primers and the enzymes that were used as well as the size of the fragments generated.

Table 9 Primers used to Generate 5' Labeled Probes for S1 Mapping and Northern Blots.

Primer*	**Enzyme**	**Size of Fragment** (bases)
1	Pst I	474
3	Alu I	149
7	Dde I	239
7	Hinf I	388
11	Taq I	372
15	Dra I	144

* Sequence of the primers can be found in the Appendix.

A Pst I - Bam HI fragment, labeled at its 3', end was obtained [illegible]ting clone p18CR-9 with Bam HI and Pst I, and labeling the fragments wit[illegible]inal deoxynucleotidyl transferase (Amersham) and α ^{32}P ddATP (3000 Ci/mmol, Amersham) as described in the Amersham 3' end Labeling kit. After labeling, the fragments were purified in the same way as the 5' labeled fragments.

S1 Mapping Protocol

Each labeled fragment (5000-50000 cpm) was mixed with 10 µg of *Sulfolobus* RNA and precipitated with 0.04 volumes 7.5 M potassium acetate and 2.5 volumes 95% ethanol at -70°C for 30 minutes. The RNA and the

fragment were then dissolved in 20 µl of 40 mM PIPES (piperazine-N,N'-bis[2-ethane-sulfonic acid] disodium salt), pH 6.8, 0.4 M NaCl, 1 mM EDTA, 80% deionized formamide (Kjems *et al.* 1987a); denatured at 90°C for 2 minutes and hybridized at 40°C for 3 hours. After hybridization, 300 µl of ice cold buffer (30 mM sodium acetate, pH 4.6, 280 mM NaCl, 4.5 mM $ZnSO_4$) containing 20 µg/ml of calf thymus denatured DNA and 130 units of S1 nuclease (Pharmacia), were added and the solution was incubated at 37°C for 30 minutes. The reaction was stopped by adding 75 µl of 2.5 M ammonium acetate, 50 mM EDTA and the protected fragments were precipitated with 400 µl of isopropanol and 20 µg of tRNA as a carrier at -70°C for 30 minutes. Fragments were washed once with 70% ethanol, vacuum dried and dissolved in 10 µl of formamide dye. For each fragment, a reaction in which *Sulfolobus* RNA was omitted, was run as a control. Protected sites were determined by running the S1 nuclease reactions, the controls and the intact fragments in 6% polyacrylamide-8 M urea sequencing gels. Electrophoresis was performed as described for the sequencing gels. Sequencing reactions and/or molecular weight markers were used to determine the size of the protected fragments. After electrophoresis, the gels were dried and exposed to X-Omat-K x ray film at -70°C overnight with an intensifying screen.

Computer Programs

Sequences were analyzed using the SEQ and PEP programs of BIONET. Computer searches of the data banks (NBRF, SWISS-PROT, GenBank and EMBL) were performed with the FASTP and FASTN programs of Lipmann and Pearson (1985). Sequences were initially aligned using the GENALIGN

program of BIONET (Martinez 1988) and then the alignments were adjusted by hand to minimize the number of gaps.

Results and Discussion

The Sso L12 Gene Oligonucleotide Probe

The successful isolation of a gene depends on the availability of an assay to identify the gene once it is cloned. A very widely used approach to identify genes that code for specific proteins consists of using the amino acid sequence of the protein to deduce the nucleotide sequence of its gene. This information is then used to synthesize an oligonucleotide probe that can be radiolabeled and used to screen a genomic library and identify the gene of interest (Wallace *et al.* 1981). Since information about the sequence of the amino terminal region of the Sso L12 protein was available (Matheson 1985), it was decided to use this approach in the isolation of the Sso L12 gene.

In designing a probe for the Sso L12 gene, several things were taken into consideration: redundancy of the genetic code, complexity, length and G-C content of the oligonucleotide mix. Since most amino acids are encoded by two, four or even six different codons, it is not possible to predict the exact sequence of a gene from the amino acid sequence of the protein it codes. Therefore, a probe must consist of a mixture of oligonucleotides reflecting all possible combinations (Wallace *et al.* 1981). This creates a problem. Since only one oligonucleotide in the mix will have the exact complementary sequence to the gene, as the number of different oligonucleotides present increases (*i.e.* the complexity of the mix) the abundance of the correct probe decreases, as well as the hybridization specificity of the probe (Wallace and Miyada 1987). This problem can be solved in two wavs: probes can be based on regions of the protein that contain amino acids determined by one (methionine and tryptophan) or two codons (aspartic acid, asparagine,

cysteine, histidine, phenylalanine, tyrosine, glutamic acid, glutamine and lysine), thus reducing the number of oligonucleotides that has to be synthesized (Szostak *et al.* 1979) or the mix can be divided into two or more sets of probes (Wallace and Miyada 1987).

The length of the probe determines its specificity. The longer the sequence, the less likely it is to bind non-specifically to other sequences (Lathe 1985). However, as the length of the probe increases so does the complexity of the mix; so a balance must be achieved between these two parameters. The length of the probe also influences the stability of the duplex formed when the probe binds to DNA. Longer probes bind to DNA with greater stability than shorter probes (Wahl *et al.* 1987). In general, oligonucleotides 10 to 20 bases long, have been used successfully to identify specific genes (Szostak *et al.* 1979).

The G-C (guanine and cytosine residues) content of the probe is important because it affects the stability with which the probe hybridizes to the target DNA (Wahl *et al.* 1987). A high G-C content increases the hybridization stability of the probe because guanine can form three hydrogen bonds with cytosine while adenine (A) can only form two with thymine (T). A high hybridization stability means that hybridizations can be performed under conditions in which the background is negligible and it is easy to distinguish between specific and non-specific binding (Wallace and Miyada 1987). However, when it is not possible to obtain a probe with a high G-C content, hybridizations can be performed in the presence of tetramethylammonium chloride, a reagent that binds AT rich regions in DNA and prevents the preferential melting of AT versus GC base pairs (Dilella and Woo 1987).

1. Amino terminal end sequence of the Sso L12 Protein

```
1                                      20
M E Y I Y A S L L L H A A K K E I S E E
```

(M E Y I Y A underlined) — region on which the probe is based

2. Sequence of the Oligonucleotide Probe (17 bases)

```
        A   T   A   T
ATG GA  TA  ATT TA  GC
        G   C   C   C

 M   E   Y   I   Y   A
```

(24 different oligonucleotides; G-C content 24%-47%)

3. Synthesized as:

```
              A   T       T
17A    ATG GA  TA  ATA TA GC
              G   C       C
                                  (8 oligonucleotides)
              A   T   T   T
17T/C  ATG GA  TA  AT  TA GC
              G   C   C   C

                                  (16 oligonucleotides)
```

Figure 9 The Sso L12 Gene Probe

Figure 9 shows the amino terminal end sequence of the Sso L12 protein. The amino acid sequence used for the Sso L12 gene probe is underlined. This region was selected because it contains one amino acid coded by only one codon (methionine), three amino acids coded by two codons (glutamic acid and two tyrosines), one amino acid coded by three codons (isoleucine) and only one amino acid coded by four codons (alanine). The sequence of the derived oligonucleotide probe mix is also shown in Figure 9. The probe was synthesized as a 17 base oligomer in order to use the first two bases of the alanine codon; this increased the G-C content of the probe and at the same time reduced the complexity of the mix by eliminating the ambiguity introduced by the third base in the alanine codons (the four alanine codons start with GC). The length selected is also within the range (10 to 20 bases) found to give good probe specificity and stability (Szostak *et al.* 1979). The Sso L12 probe consists of 24 different oligonucleotides with a G-C content that ranges from 24% to 47%. In order to reduce the complexity of the mix even more, the probe was synthesized as two mixes: **17A**, in which all the oligonucleotides have an **A** in the third position of the isoleucine codon and **17 T/C**, in which the oligonucleotides have a **T** or a **C** in that position (see Figure 9).

Conditions under which the Sso L12 Probe binds Stringently to *Sulfolobus* DNA

A series of dot blot experiments were performed in order to determine which of the two mixes contained the oligonucleotide with the sequence complementary to the Sso L12 gene and the conditions under which it would bind stringently to *Sulfolobus* DNA. In these experiments, the two probe mixes were radiolabeled and hybridized to *Sulfolobus* DNA bound to nitrocellulose

filters. Hybridization stringency was determined by varying the temperature and the concentration of the salt in the posthybridization washes (see Table 6, Materials and Methods) (Wahl *et al.* 1987).

Dove and Davidson (1962) have studied the effect of the cation concentration (Na^+) on the melting temperature (Tm) of DNA (Tm is defined as the temperature at which half of the double stranded DNA molecules have separated into their constituent single strands). They found that as the cation concentration decreases so does the Tm of the DNA. This is due to the fact that as the cation concentration is reduced, the degree of ionization of the phosphate groups increases creating a greater electrostatic repulsion which facilitates strand separation when fewer of the hydrogen bonds between the two chains are broken. Since the effect of reducing the cation concentration on the Tm is more pronounced for mismatched duplexes than for their perfectly matched counterparts, conditions can be found in which only perfectly matched hybrids are stable.

Wallace *et al.* (1979) have found that in the case of duplexes formed by short oligonucleotides (less than 20 bases long) and immobilized DNA, the Tm decreases by about 5°C for every mismatched base pair. This means that by increasing the temperature, one can find conditions under which only perfectly matched duplexes exist.

The effect of decreasing the cation concentration and increasing the temperature on the binding of the two probe mixes can be observed in Figure 10. Note that the **17A** probe remains bound even when the filter was washed in 0.2 x SSC at 43°C, while the **17 T/C** probe only binds under non-stringent conditions (6x SSC, room temperature). These results indicate that the **17A** probe has the oligonucleotide with the sequence that is complementary to the

Sso L12 gene. On the basis of these results and in order to reduce the background as much as possible, all the filters or gels probed with the **17A** probe were washed successively with 6x SSC, 1x SSC and 0.2x SSC at 40 °C (stringent conditions).

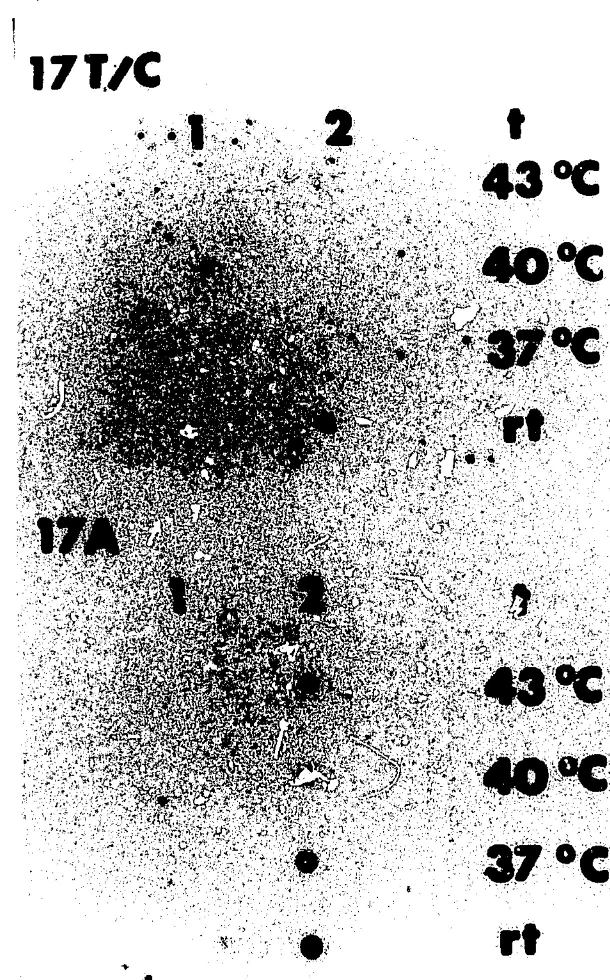


Figure 10. Dot blot hybridizations of the 17T/C and 17A Sso L12 gene probes to *Sulfolobus* DNA. *Sulfolobus* DNA bound to nitrocellulose filters was hybridized to the 17 T/C and 17A probes. After hybridization, the filters were washed under increasingly stringent conditions (higher temperature and lower ionic strength) (see Table 6, Materials and Methods) and then exposed to x-ray film. A picture of the autoradiograph obtained is shown. Lanes: 1: λ phage DNA (negative control), 2: *Sulfolobus solfataricus* DNA, t: temperature of the washes, rt: room temperature.

Figure 11. Hybridization of the 17A probe to different restriction endonuclease digests of *Sulfolobus* DNA. *Sulfolobus* DNA was digested with different restriction endonucleases. The fragments were then separated on a 0.5% agarose gel; and the dried gel was hybridized to the 17A probe. After hybridization, the gel was washed under stringent conditions and exposed to x-ray film for 5 hours. A photograph of the autoradiograph obtained is shown. The position of the molecular weight standards is shown. Lanes: a: Bam HI, b: Eco RI, c: Hind III, d: Pst I, e: Bam HI/Eco RI, f: Bam HI/Hind III, g: Bam HI/ Pst I, h: Eco RI/Hind III, i: Eco RI/Pst I, j: Hind III/Pst I.

Partial Restriction Map of the Region in which the Sso L12 Gene is Located

The **17A** probe was hybridized to *Sulfolobus* DNA cut with different restriction endonucleases in order to obtain a partial restriction map of the region in which the Sso L12 gene is located. The results of this experiment are shown in Figure 11. Using the data obtained from this experiment, the restriction map shown in Figure 12 was constructed. The Sso L12 gene or at least the region of the gene that hybridizes with the probe is located within a 1.1 Kb Pst I - Pst I fragment.

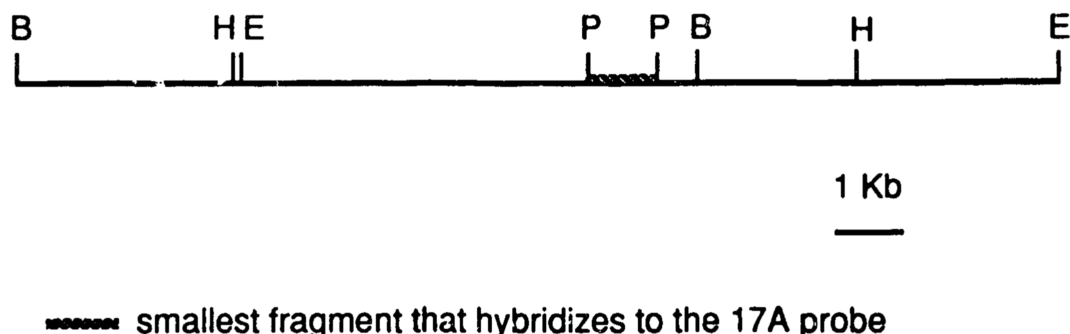


Figure 12. Partial restriction map of the region containing the Sso L12 gene. The restriction map shown was constructed using the data from the experiment shown in Figure 11. The position of the 1.1 Kb Pst I - Pst I fragment inside the 6.9 Kb Eco RI - Bam HI fragment could not be determined from the data in Figure 12. The position of this fragment was later determined by restriction mapping of a clone (p18CR-9) containing the 6.9 Kb Eco RI - Bam HI fragment (see Figure 21) and confirmed by sequencing.

Construction of a Genomic Library of *Sulfolobus solfataricus* and Isolation of the Sso L12 Gene

On the basis of the information obtained from the restriction map, it was decided to clone the 10 Kb Bam HI - Bam HI fragment that hybridizes with the 17 A probe. This fragment was selected because there was a high probability that it contained the complete Sso L12 gene as well as the genes around it. Furthermore, the Bam HI restriction site, made it very convenient for cloning since this site is present in most of the vectors available (Maniatis *et al.* 1982).

In order to clone this fragment, a genomic library was constructed by cloning a Bam HI digest of *Sulfolobus* DNA in phage EMBL 3 (see Figure 13). This vector was selected because it allows the cloning of fragments with sizes between 8 and 23 Kb and recombinant phages can be efficiently recovered by genetic and biochemical selection against the vector phage (Frischauf *et al.* 1983). The basis of the genetic selection is the presence of the *red* and *gamma* genes of λ phage in a nonessential middle fragment which is flanked by symmetrically arranged restriction sites for Sal I, Bam HI and Eco RI (see Figure 13). The expression of these two genes prevents the growth of the vector phage on *Escherichia coli* strains that are lysogenic for phage P2 (vector phage is *Spi+*, that is sensitive to P2 interference). Since the middle fragment is replaced by an insert in recombinant phages, their growth is not affected by the presence of a P2 lysogen (*Spi-*) (Frischauf *et al.* 1983, Karn *et al.* 1983). Thus, by amplifying the library in a host carrying a P2 lysogen, like Q359, wild type phages are eliminated and only recombinant phages are recovered (Frischauf *et al.* 1983, Karn *et al.* 1983). The biochemical selection consists of cleaving the middle fragment with Eco RI and removing the small fragments generated by selective precipitation of the large DNA fragments with

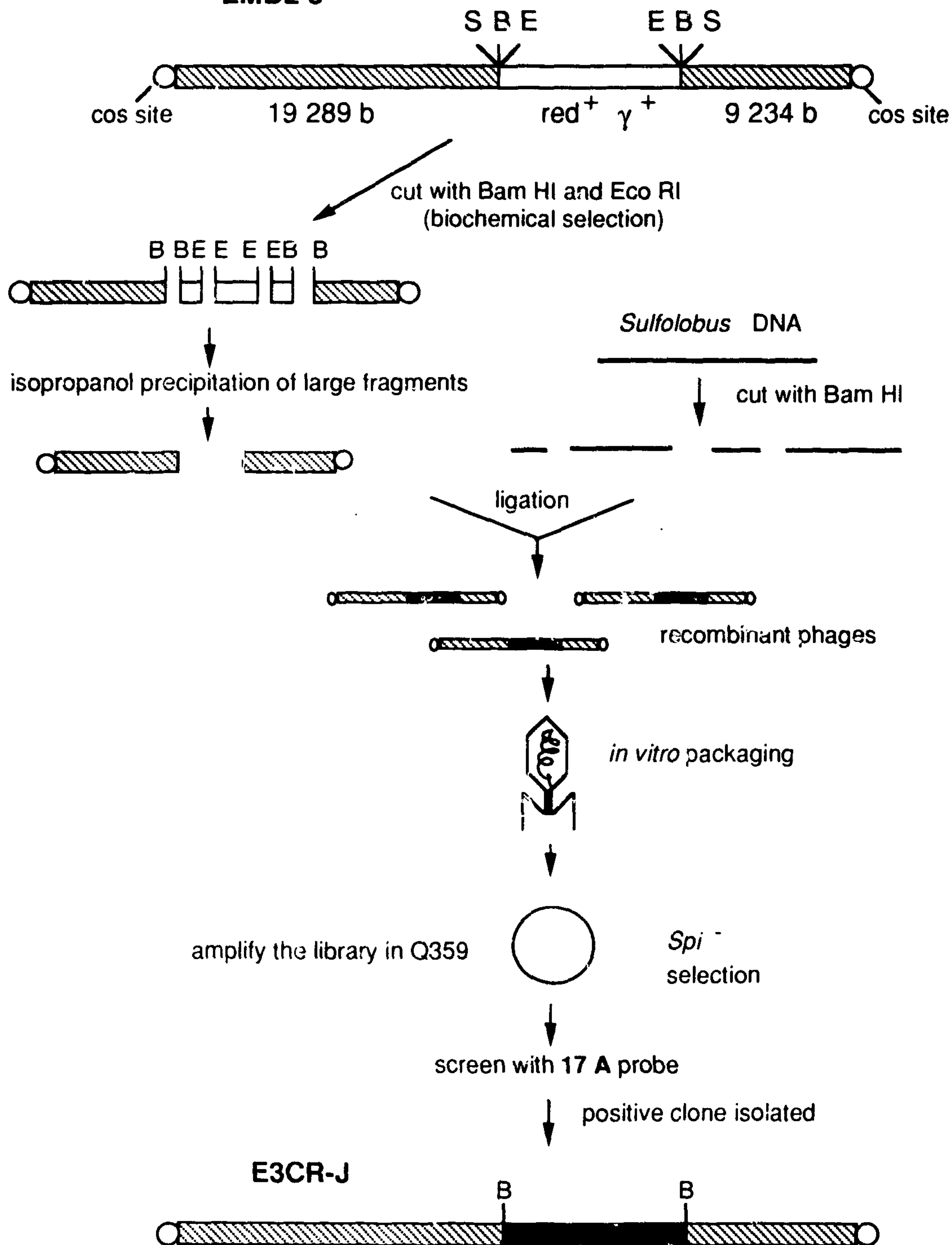


Figure 13. Construction of a *Sulfolobus* genomic library and isolation of a clone carrying the Sso L12 gene. For a discussion of the biochemical and Spi⁻ selection, see the text. B: Bam HI, E: Eco RI, S: Sal I, cos sites: single stranded cohesive ends. Structure of EMBL 3 drawn from Frischauf *et al.* 1983.

isopropanol. This prevents the religation of the original vector (Frischauf *et al.* 1983, Frischauf *et al.* 1987). Both types of selection were used in the construction of the *Sulfolobus* library (see Figure 13).

After the library was constructed, it was screened by plaque hybridization with the **17A** probe and a positive clone designated E3CR-J was isolated (see Figure 13 and Figure 14 , lane e). This clone was used as the source of DNA for the subcloning of smaller fragments for sequencing.

Binding of the Sso L12 Probe Mixes to DNA from other Archaebacteria

A series of dot blots were performed in order to check if the Sso L12 probe mixes would bind stringently to DNA from other archaebacteria. The results of this experiment are shown in Figure 14. It was found that the **17A** probe binds stringently to the DNA from the methanogen but suprisingly not to the DNA from another extreme thermophile: *Thermoproteus tenax*. The **17 T/C** probe, on the other hand, did not bind stringently to any of the samples tested (data not shown). These results indicate that the **17A** probe can also be used to clone the L12 gene from *Methanobacterium thermoautotrophicum*. The fact that neither of the Sso L12 probes binds stringently to the DNA from *T. tenax*, seems to indicate that there is sequence variability in this region of the protein among the extreme thermophiles.

Subcloning and Sequencing of a 1.1 Kb Pst I - Pst I fragment that hybridizes to the 17A Probe

Since the hybridization of the **17A** probe with different restriction digests of *Sulfolobus* DNA showed that at least part of the Sso L12 gene is located within

a 1.1 Kb Pst I-Pst I fragment (see Figure 11), it was decided to subclone this fragment and sequence it.

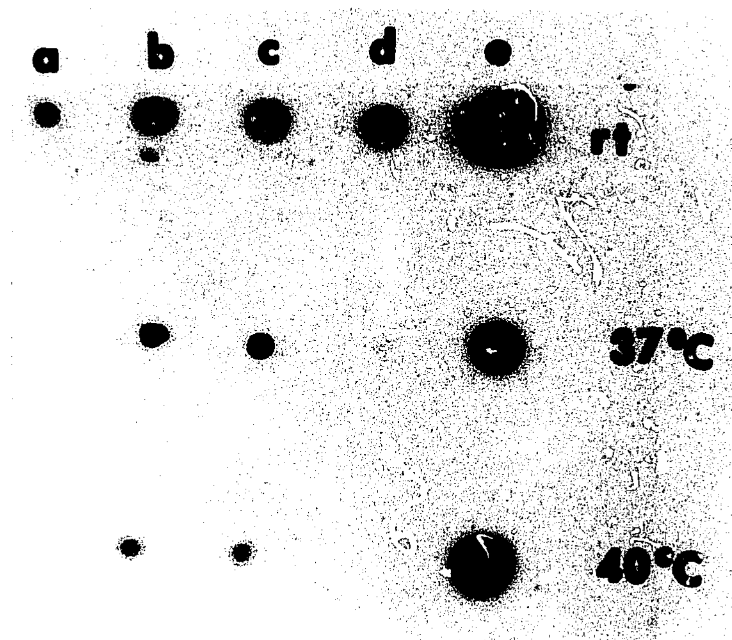
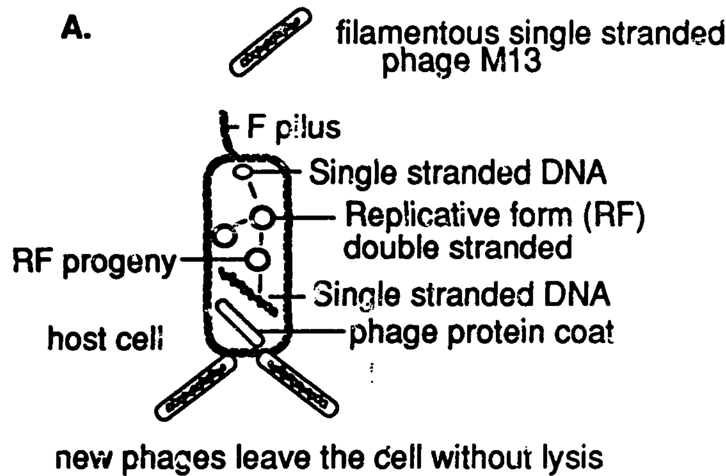
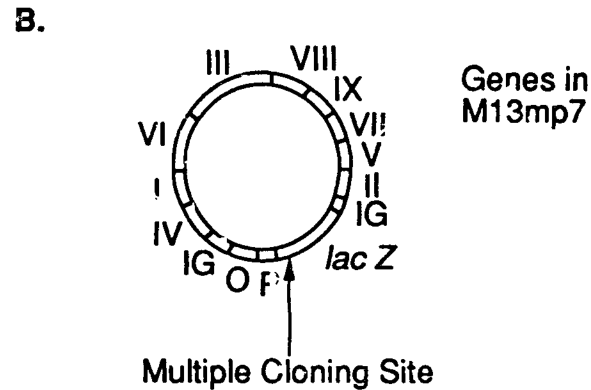


Figure 14. Hybridization of the 17A probe to DNA from other archaebacteria and clone E3CR-J. DNA from a methanogen (*Methanobacterium thermoautotrophicum*), an extreme sulfur-dependent thermophile (*Thermoproteus tenax*) and clone E3CR-J was bound to nitrocellulose filters and hybridized to the 17A probe. After hybridization, the filters were washed under increasingly stringent conditions (see Table 6, Materials and Methods). Filters were then exposed to x-ray film. A picture of the autoradiograph obtained is shown. The temperature at which the washes were performed is indicated. Lanes: a: λ phage (negative control), b: *Sulfolobus solfataricus* DNA (positive control), c: *Methanobacterium thermoautotrophicum* DNA, d: *Thermoproteus tenax* DNA, e: clone E3CR-J, rt: room temperature.

Phage M13mp7 was selected as a vector for this purpose because it provides a biological system for obtaining cloned DNA in the single stranded form required for sequencing (Messing *et al.* 1981, Messing 1983). The M13mp series of vectors are derivatives of the single stranded filamentous phage M13 (Messing *et al.* 1977, Messing 1983). During its life cycle (see Figure 15), this phage infects *Escherichia coli* cells through the F pilus. Once inside the cell, the single stranded form is converted into a double stranded circular form (replicative form or RF). This form is used to produce new single stranded DNA as well as viral proteins. After the new phages are assembled, they leave the cell without lysing the host (Messing 1983).

Figure 15 shows the structure of M13mp7 as well as the sequence of its multiple cloning site. The phage carries a fragment of the *E.coli lac* operon that contains the promoter and the first 145 amino acids of the β-galactosidase gene, and facilitates the identification of recombinant phages (see Figure 15) (Messing *et al.* 1977). This fragment is able to complement a mutation in *E.coli* cells (*lac Z* ΔM15) which deletes residues 11-41 of the α peptide of β-galactosidase (α complementation) (Messing *et al.* 1977). For this reason, when cells carrying this mutation are infected with vector phage, a functional β-galactosidase is obtained. If these cells are grown in the presence of isopropylthiogalactoside (IPTG) and 5-dibromo 4-chloro 3-indolylgalactoside (x-gal), a blue plaque is obtained. IPTG is an inducer of the *lac* operon and x-gal is a colorless compound that releases a 5- bromo 4-chloro-indigo blue color when hydrolyzed by β-galactosidase (Messing *et al.* 1977). Since the multiple cloning site is located within the α peptide sequence, the insertion of a DNA fragment disrupts the sequence and prevents the α complementation. For this reason, when cells infected with recombinant phages are grown in the





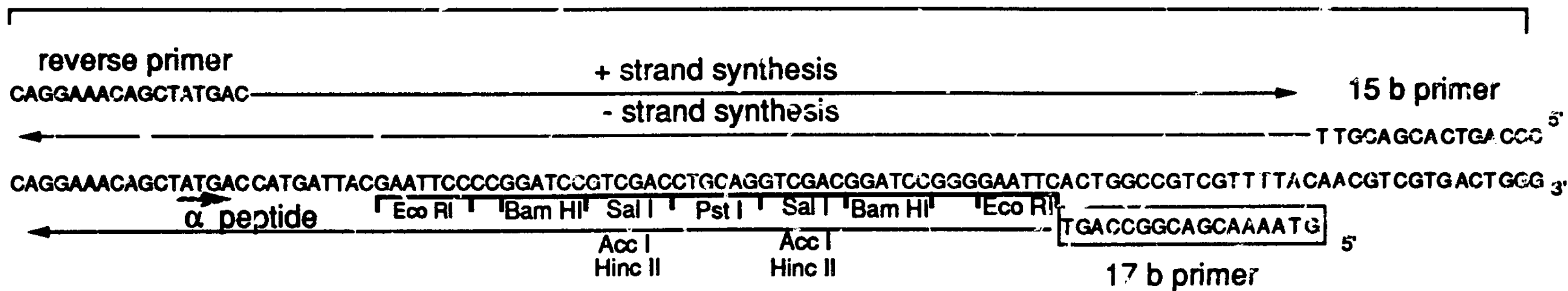


Figure 15. Life Cycle of M13 and Structure of M13mp7. A. Life Cycle of M13 (drawn from Messing 1983) B. Stucture of M13mp7. O: operator, P: promoter, *lac Z*: part of the β galactosidase gene (first 145 amino acids), viral genes I, IV, VI, VII and IX code for proteins required for viral assembly, III and VII code for the viral coat proteins, II and VIII code for proteins involved in viral DNA replication, IG: intergenic region that contains the initiation site for replication (drawn from Messing 1983) C. Sequence of the Multiple Cloning site of M13mp7.The universal primer used for sequencing is boxed (drawn from Messing *et al.* 1981). The initiation codon of the α peptide is indicated by an arrow.

presence of IPTG and x-gal, they produce white plaques and can easily be distinguished from cells carrying vector phages (Messing 1983).

Another important advantage of this vector, is the existence of commercially available universal sequencing primers, that hybridize to a region close to the multiple cloning site (see Figure 15) and permit the sequencing of any DNA fragment cloned into the M13 mp vectors (Messing 1983).

The 1.1 Kb Pst I - Pst I fragment was isolated from a Pst I digest of phage E3CR-J and cloned in M13mp7 RF (see Materials and Methods). The RF was introduced into competent JM103 cells and recombinant phages were identified by the formation of white plaques on IPTG-x-gal-LB plates. Since the 1.1 Kb fragment has the same restriction sites at both ends, it was expected that clones containing the fragment in different orientations would be obtained. In order to distinguish between these two orientations, the clones obtained were hybridized to the 17A probe (see Figure 16) In this way two clones, M13CR-3 and M13CR-8 containing the 1.1 Kb fragment in different orientations were identified. These two clones were used to sequence the 1.1 Kb fragment.

The sequencing strategy and the genes present in the 1.1 Kb fragment are shown in Figure 17. Figure 18 shows a photograph of a sequencing gel. The strategy used to sequence the 1.1 Kb fragment, consisted of obtaining sequencing data by using one of the M13 sequencing primers (the 17 base long primer (see Figure 15)), and then using the data obtained to synthesize a specific primer, complementary to the 3' end of the sequence. This new primer was then used to sequence the next 250 to 300 bases and the data obtained used to synthesize a new primer (Strauss *et al.* 1986). The sequence of all the primers used to sequence the 1.1 Kb fragment can be found in the Appendix.

Figure 16. Determination of the orientation of the insert in clones M13CR-3 and M13CR-8. DNA from several clones containing the 1.1 Kb Pst I - Pst I fragment was run on a 0.5% agarose gel and transferred to nitrocellulose filters. The filters were then hybridized to the 17A probe to determine which clones had the insert in different orientations. A picture of the autoradiograph obtained is shown. Clones M13CR-5, 7 and 8 have the insert in the same orientation because they hybridize with the probe; and clones M13CR-3, 6 and 9 have it in the opposite orientation because they do not hybridize with the probe. Lanes a: M13CR-3, b: M13CR-5, c: M13CR-6, d: M13CR-7, e: M13CR-8, f: M13CR-9.

Sequencing of the 1.1 Kb fragment revealed that only the region of the Sso L12 gene that codes for the first 70 amino acids of the Sso L12 protein is present in this fragment. Upstream of the Sso L12 gene, an open reading frame was found that was also not complete. This open reading frame was identified as the Sso L10 gene because part of the sequence of the protein that it codes for, matched the sequence of peptides obtained from the Sso L10 protein (Shimmin *et al.* 1989a). Furthermore, sequence similarity between the

product of this gene and the Eco L10 protein was also found. Since both of the genes present in the 1.1 Kb fragment were incomplete, it was necessary to subclone a bigger fragment to obtain their complete sequence.

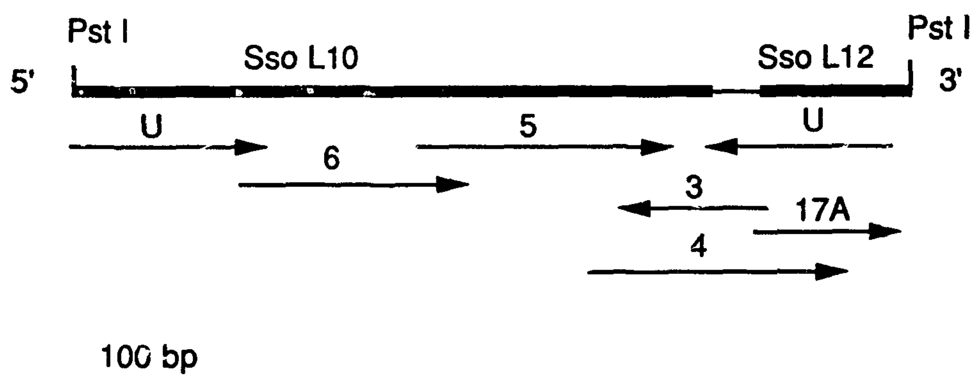


Figure 17. Sequencing strategy for the 1.1 Kb Pst I - Pst I fragment. The 3' terminal region of the Sso L10 gene and the 5' terminal region of the Sso L12 gene are located in this fragment. Arrows indicate the direction of sequencing. The numbers above the arrows refer to the primers used. The sequence of these primers can be found in the Appendix. U: universal 17 bp primer, 17A: 17A probe. Clone M13CR-8 was used to sequence the insert in the 5'- 3' direction and clone M13CR-3 in the 3' - 5' direction.

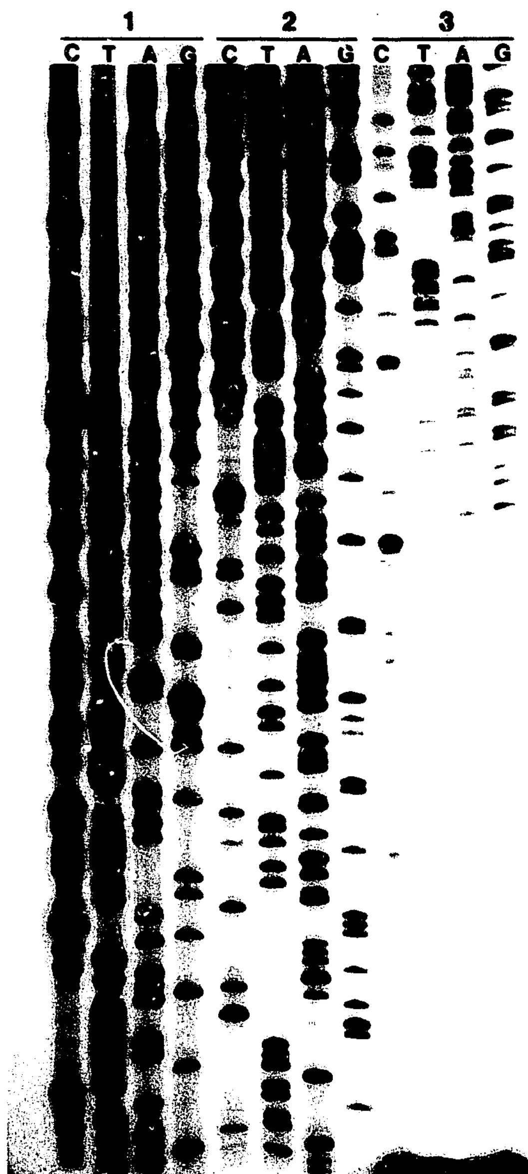


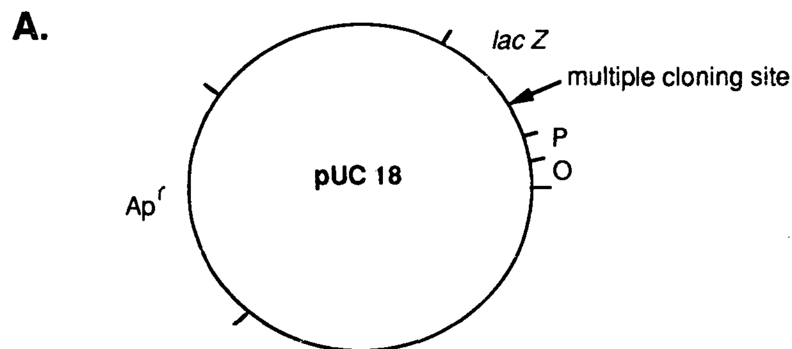
Figure 18. Example of a Sequencing gel. A picture of the autoradiograph of a sequencing gel is shown. The template in this experiment is deletion plasmid p26-4 and the primer is the universal primer. Lanes C, T, A, G : C, T, A, G sequencing reactions (see Materials and Methods). Each reaction was loaded three times at 2 hour intervals, 1: first loading, 2: second loading, 3: third loading. The sequence is read from the bottom to the top, starting with the third loading.

Subcloning of a 6.9 Kb Eco RI - Bam HI fragment into pUC 18

Since a bigger fragment was needed to obtain the complete sequence of the Sso L12 gene, the 6.9 Kb Eco RI - Bam HI fragment that hybridizes with the **17A** probe (see Figure 12), was subcloned into pUC 18. The pUC plasmids contain a Pvu II/Eco RI fragment from pBR322 which carries the β-lactamase gene (ampicillin resistance [Ap^r] and the origin of replication (Viera and Messing 1982). For this reason, cells carrying the plasmid are resistant to ampicillin and can be easily selected by growing them in the presence of this antibiotic (Viera and Messing 1982). The pUC plasmids also have multiple cloning sites that are derived from the M13mp series of vectors. Thus, when a fragment of DNA is inserted into the multiple cloning site, it prevents the α complementation, and a white, instead of a blue colony is obtained in x-gal-IPTG-ampicillin-LB plates (Viera and Messing 1982, Yanisch-Perron *et al.* 1985).

Plasmid pUC 18 was selected as a vector in this case because it allows the cloning of large fragments, which are usually very stable. Furthermore, since the multiple cloning site comes from the M13mp vectors, the same universal primer used for sequencing the 1.1 Kb fragment could be used initially to sequence the 6.9 Kb fragment. The structure of pUC 18 and the sequence of its multiple cloning site are shown in Figure 19.

The 6.9 Kb Eco RI - Bam HI fragment was isolated from a Bam HI/Eco RI digest of phage E3CR-J and ligated to pUC 18 (see Materials and Methods). The ligated plasmids were then used to transform competent JM105 cells and the transformed cells were plated on x-gal-IPTG-ampicillin-LB plates. White colonies were isolated, their plasmid DNA extracted, and sized on a 0.5% agarose gel. A plasmid with the correct size was then cut with Bam HI and Eco



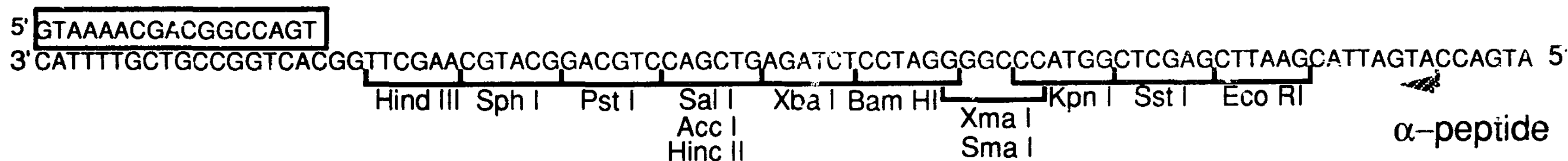


Figure 19. Structure of pUC 18 and its multiple cloning site. Ap^r: ampicillin resistance gene O: operator P: promoter and *lac Z:* part of the β galactosidase gene (the first 107 amino acids). Drawn after Yanisch-Perron *et al.* 1985. B. Multiple cloning site. The universal 17 nucleotide sequencing primer is boxed.Initiation codon of the α peptide is indicated by the arrow. Drawn after Yanisch-Perron *et al.* 1985.

RI, and the fragments were separated on a 0.5% agarose gel to check that the 6.9 Kb fragment was indeed present in this clone (see Figure 20). A clone designated p18CR-9 was isolated in this way.

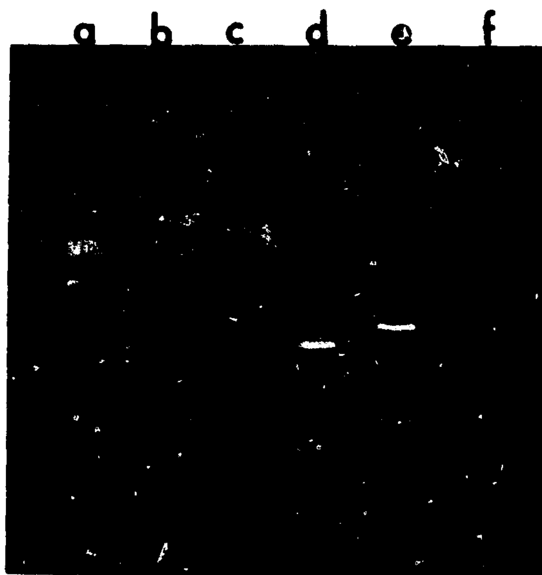


Figure 20. Identification of clone p18CR-9. Picture of a 0.5% agarose gel showing that clone p18CR-9 has the 6.9 Kb fragment. Note that the 6.9 Kb fragment is running faster in this gel than the molecular weight marker. Lanes: a: λ Hind III molecular weight fragments, from top to bottom: 23.1 Kb, 9.4 Kb, 6.6 Kb, 4.4 Kb, 2.3 Kb, 2.2 Kb, b: pUC 18, c: pUC 18 cut with Bam HI/Eco RI, d: clone p18CR-9, e: clone 18CR-9 cut with Bam HI/Eco RI, f: phage E3CR-J cut with Bam HI and Eco RI.

Mapping of the Position of the 1.1 Kb Pst I - Pst I Fragment Inside the 6.9 Kb Eco RI - Bam HI Fragment

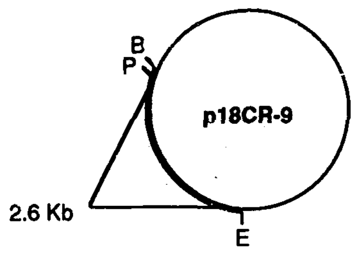
In order to design a suitable sequencing strategy to obtain the complete sequence of the Sso L12 gene, the position of the 1.1 Kb Pst I - Pst I fragment within the 6.9 Kb Eco RI - Bam HI fragment was determined. For this purpose, clone p18CR-9 was digested with the following enzymes: Pst I, Bam HI/Pst I, Eco RI/Pst I, and Eco RI/Bam HI/Pst I and the fragments were separated on a 0.5% agarose gel (see Figure 21).

The digestion with Pst I generated 3 fragments: the 1.1 Kb Pst I - Pst I fragment that contains part of the Sso L12 gene, a fragment of ≈ 7.5 Kb and a small fragment of ≈ 0.7 Kb (see Figure 21). This means that there are only three Pst I sites in clone p18CR-9. Two of these sites are within the *Sulfolobus* insert and give rise to the 1.1 Kb fragment. The other is located 12 bases upstream of the Bam HI site in the multiple cloning site (see Figures 19 and 21). Since the distance between this Pst I site and the Eco RI site that marks the border between the vector and the *Sulfolobus* insert is 2.6 Kb (the size of the vector), that means that the 7.5 Kb fragment must contain the Eco RI site, the 0.7 Kb must have the Bam HI site and therefore, the 1.1 Kb fragment must be located at about 0.7 Kb from the Bam HI site (see Figure 21). The rest of the digests confirm this result.

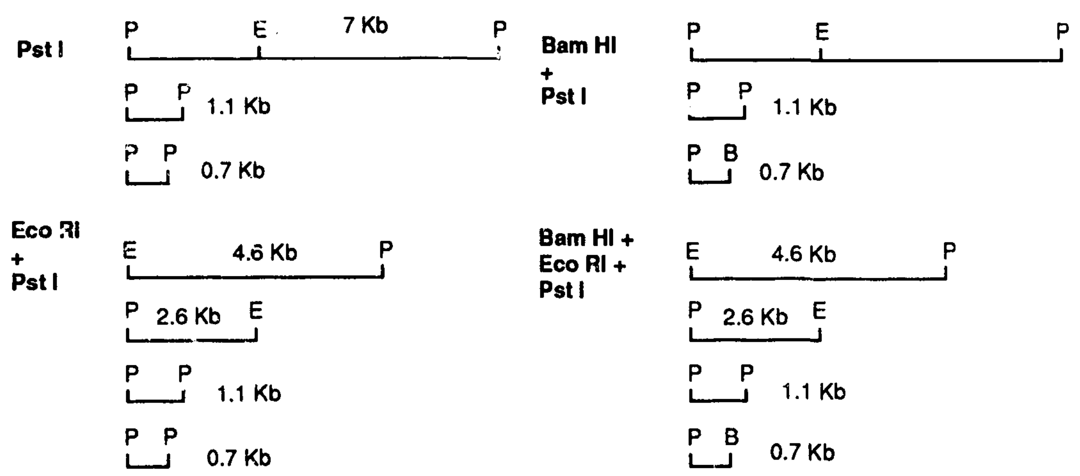
Digestion with Bam HI and Pst I generates three fragments of the same size as those produced by Pst I (see Figure 21). Since the Pst I and the Bam HI sites in the multiple cloning site are only 12 nucleotides apart, digestion with Bam HI will only cut 12 bases from the 0.7 Kb Pst I - Pst I fragment. Since the difference in size between these two fragments is to small to detect on an agarose gel, they run side by side on the gel as can be seen in Figure 21.

Figure 21. Location of the 1.1 Kb Pst I- Pst I fragment within the 6.9 Kb Eco RI - Bam HI fragment. Clone p18CR-9 was digested with several restriction endonucleases and the fragments were separated on a 0.5% agarose gel. Panel A. Picture of the agarose gel. Lanes: A: λ Hind III molecular weight standards, from top to bottom: 23.1 Kb, 9.4 Kb, 6.6 Kb, 4.4 Kb, 2.3 Kb, 2.0 Kb, B: ΦX174 RF Hae III molecular weight markers, from top to bottom: 1.3 Kb, 1.0 Kb, 0.8 Kb, 0.6 Kb, 0.3 Kb, C: Pst I digest of clone p18CR-3, D. Bam HI/Pst I digest, E: Eco RI/Pst I, F: Bam HI/Eco RI/ Pst I. Panel B. (see next page) Structure of clone p18CR-9. Vector sequences are indicated with a heavy line. Panel C. Fragments generated by the different digestions. Panel D. Location of the 1.1 Kb fragment.

B.



C. Digestion Products



D. Location of 1.1 Kb PstI - Pst I Fragment

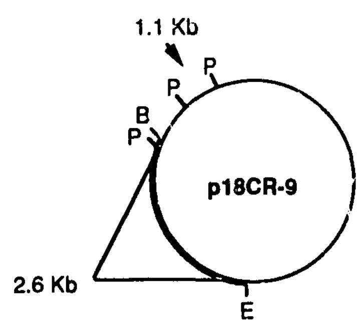


Figure 21 ...continued.

Digestion with Pst I and Eco RI, produces 4 fragments and it clearly shows that the Eco RI site is inside the 7.5 Kb fragment since this fragment is no longer present (see Figure 21). Digestion of this fragment with Eco RI generates a 2.6 Kb Eco RI - Pst I fragment which represents the vector and a 5 Kb fragment which is derived from the *Sulfolobus* insert.

Digestion with Eco RI, Bam HI and Pst I generates the same fragments as the previous digestion because the difference in size of the 0.7 Kb Pst I - Pst I and Pst I - Bam HI fragments cannot be detected in this gel (see Figure 21).

In conclusion, the results from this experiment showed that the 1.1 Kb Pst I - Pst I fragment is located at about 5 Kb from the Eco RI site and about 0.7 Kb from the Bam HI site.

Sequencing of the 6.9 Kb Eco RI - Bam HI fragment

Figure 22 shows the strategy used to sequence the complete 6.9 Kb Eco RI - Bam HI fragment. Since the 1.1 Kb Pst I - Pst I fragment is located about 700 nucleotides from the Bam HI site, sequencing of the 6.9 Kb fragment was started from this site by using the 17 bp universal sequencing primer (see Figure 19). With the data obtained, a specific primer was synthesized and used to complete the sequencing of the Sso L12 gene (see Figure 22). A second primer, derived from the 3' end of the 1.1 Kb fragment was also synthesized and used to confirm this sequence (see Figure 22).

In order to facilitate the sequencing of the rest of the 6.9 Kb fragment, a series of deletion plasmids was constructed by digesting clone p18CR-9 with DNAse I (Lin *et al.* 1985). Figure 23 shows the different steps used in the construction of these deletion plasmids.

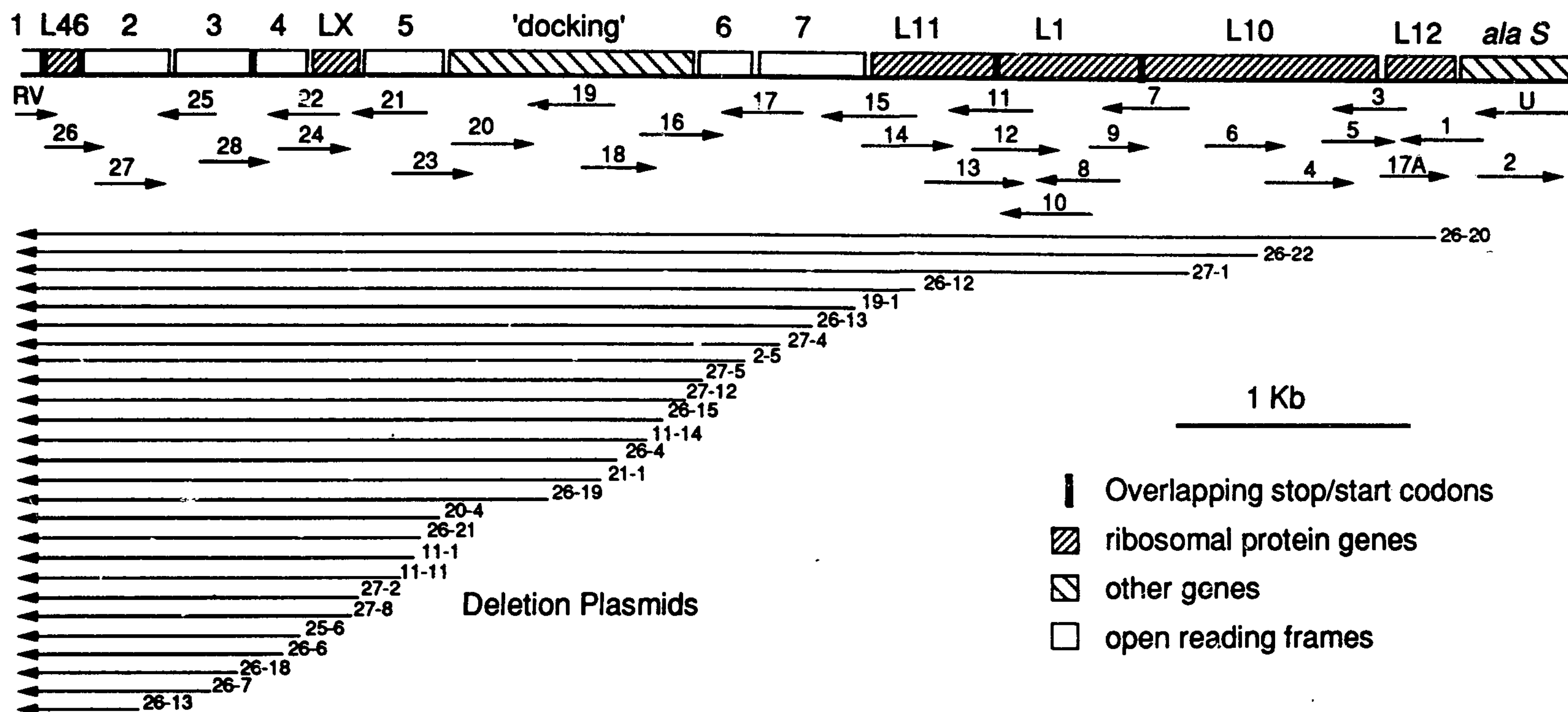


Figure 22. Sequencing strategy and organization of the genes present in the 6.9 Kb Eco RI - Bam HI fragment. Arrows indicate the direction of sequencing. Numbers above the arrows refer to the primer used. The sequence of the primers used can be found in the Appendix. Deletion plasmids are also indicated. In this case, the arrows indicate the size of the insert. Seven ribosomal protein genes were identified: L46, LX, L11, L1, L10 and L12. Besides these genes, a gene whose product shows sequence similarity to the signal recognition particle receptor (docking protein) and a gene for the alanine tRNA synthetase *(ala S)* were identified. Seven open reading frames are also present. U:universal sequencing primer, 17A: 17A probe, RV: reverse sequencing primer.

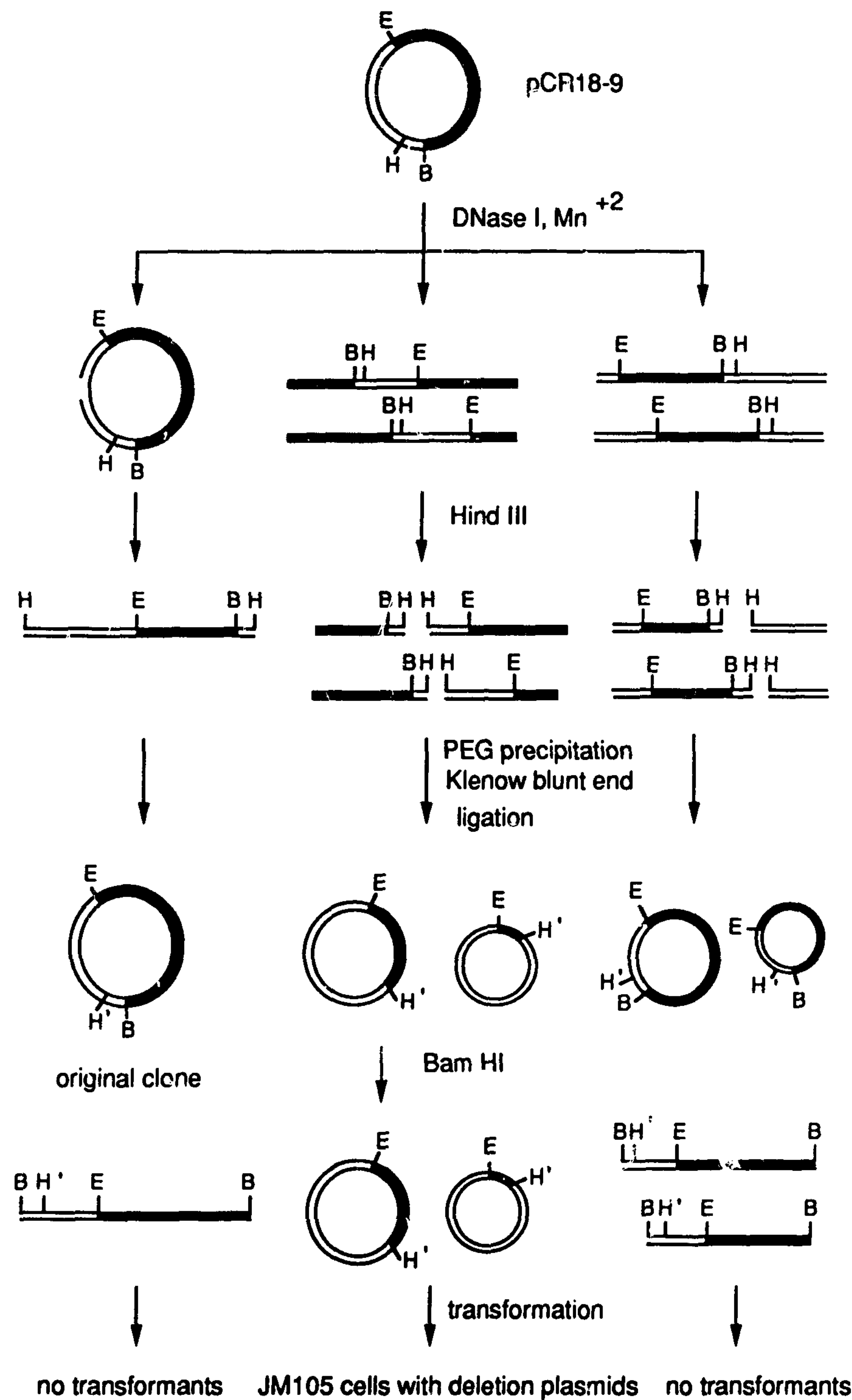


Figure 23. Construction of the deletion plasmids according to the method of Lin *et al.* (1985). For a description of the different steps, see the text. *Sulfolobus* insert shown in black. E: Eco RI, H: Hind III, H': location of original Hind III site, B: Bam HI, PEG: polyethylene glycol.

In the first step, clone p18CR-9 was digested with DNAse I. The digestion was performed in the presence of Mn^{+2}, because under these conditions, DNAse I slightly favors the cleavage of double-stranded DNA at or close to the same position on both strands (Anderson 1981). Digestion of the clone with DNAse I generated the following products (see Figure 23): 1] DNAse I cuts within the insert produced a mixture of linear double-stranded fragments with varying lengths of inserted DNA at both ends 2] DNAse I cuts within the vector produced linear double-stranded fragments with intact inserts and varying lengths of vector sequences at both ends 3] although double cuts are favored in the presence of Mn^{+2}, DNAse I will still nick many plasmids on just one strand producing open circular plasmids (Lin *et al.* 1985).

In the second step, the products of the DNAse I digestion were cleaved with Hind III. The purpose of this digestion was to orderly release part of the insert and to linearize the open circular plasmids (Lin *et al.*1985). The restriction enzyme was selected so that it cuts near the site where the primer binds so that the universal sequencing primer could be used to sequence the deletion plasmids. After digestion with Hind III, the small fragments were removed by selective precipitation of the larger fragments with polyethylene glycol (Lin *et al.* 1985).

In the third step, the fragments were made blunt ended with the Klenow fragment of DNA polymerase I and then they were ligated. Three different types of plasmids were obtained from this ligation (see Figure 23): 1] plasmids with deletions in the insert 2] plasmids with deletions in the vector 3] the original clone.

In order to eliminate the plasmids that carry an intact insert (original clone and plasmids with deletions in the vector), the plasmids were digested with a

second enzyme: Bam HI. The elimination of the plasmids containing an intact insert is based on the idea that the plasmids that have deletions in the insert should have lost this restriction site. Thus, digestion with the second enzyme linearizes the plasmids that carry an intact insert but does not affect plasmids with deletions in the insert. Since linear DNA is not suitable for transformation, only the plasmids carrying deletions in the insert will be recovered after transformation (Lin *et al.* 1985). Bam HI was selected in this case, because its single recognition site is located between the site of the first enzyme and the beginning of the insert, and thus plasmids with deletions in the insert should have lost this site (see Figure 23).

In the final step, the products of the Bam HI digestion were used to transform competent JM105 cells and the transformed cells were plated on ampicillin-LB plates. Colonies were then isolated, their plasmids were extracted (see Materials and Methods) and sized on a 0.5% agarose gel. The different plasmids were ordered according to size and used to sequence the 6.9 Kb fragment. Specific primers were used to fill the gaps when overlaps between the deletion plasmids could not be obtained and to sequence the other strand (see Figure 22). The sequence of these primers can be found in the Appendix.

The organization of the genes present in the 6.9 Kb Eco RI - Bam HI fragment is also shown in Figure 22. Besides the genes that form the L12 domain: Sso L12, Sso L10 and Sso L11, the genes of three other ribosomal proteins were identified: the Sso L1 gene that codes for a protein that is homologous to the Eco L1 protein (Post *et al.* 1979), the Sso L46 gene that codes for a small basic protein that is homologous to yeast L46 (Leer *et al.* 1985a) and rat liver L39 (Lin *et al.* 1984) ribosomal proteins and the Sso LX

gene that codes for another basic protein that shows no sequence similarity to any of the eubacterial or eukaryotic ribosomal proteins present in the data base. Besides the ribosomal proteins, an open reading frame whose product shows sequence similarity to the *Escherichia coli* alanine-tRNA synthetase (Putney *et al.* 1981a) and an open reading frame whose product shows sequence similarity to the signal recognition particle receptor (docking protein) from dogs (Lauffer *et al.* 1985) and humans (Hortsch *et al.* 1988), as well as the product of the *fts Y* gene from *Escherichia coli* (Gill *et al.* 1986), were identified. Several other open reading frames are also present, but since the sequences of their products didn't match any of the sequences present in the data base their identity is unknown.

The sequences of all these genes and their products as well as the basis for their identification will be discussed in the following sections.

The Sso L12 Gene and the Sso L12 Protein

The sequences of the Sso L12 gene and its product are shown in Figure 24. The gene has 318 nucleotides and codes for a protein of 105 amino acids. Table 10 shows the codon utilization of this gene (for a discussion, see the section on codon utilization).

The amino acid composition, pI and molecular weight of the protein are shown in Table 11. Like its eubacterial and eukaryotic counterparts (Wittmann-Liebold 1986), the Sso L12 protein is acidic (pI 4.74) and rich in alanine (17.1%).

The Sso L12 protein sequence was compared to the following archaebacterial sequences: Mva L12 (*Methanococcus vannielii*) (Strobel *et al.* 1988), Hcu L12 (*Halobacterium cutirubrum*) (Shimmin and Dennis 1989), Hha

Sso L12 Gene

1													42
ATG	GAG	TAC	ATA	TAC	GCA	AGT	TTA	CTA	TTG	CAC	GCA	GCT	AAG
M	E	Y	I	Y	A	S	L	L	L	H	A	A	K
43													84
AAA	GAA	ATA	AGT	GAG	GAA	AAC	ATA	AAG	AAT	GTA	TTA	AGT	GCA
K	E	I	S	E	E	N	I	K	N	V	L	S	A
35													126
GCA	GGA	ATA	ACA	GTG	GAC	GAA	GTT	AGA	CTA	AAA	GCT	GTA	GCA
A	G	I	T	V	D	E	V	R	L	K	A	V	A
127													168
GCT	GCT	TTA	AAA	GAA	GTA	AAT	ATA	GAC	GAA	ATA	TTG	AAG	ACA
A	A	L	K	E	V	N	I	D	E	I	L	K	T
169													210
GCT	ACT	GCA	ATG	CCT	GTA	GCA	GCA	GTA	GCA	GCC	CCT	GCA	GGT
A	T	A	M	P	V	A	A	V	A	A	P	A	G
211													252
CAA	CAG	ACA	CAG	CAG	GCA	GCT	GAG	AAG	AAA	GAA	GAA	AAG	AAA
Q	Q	T	Q	Q	A	A	E	K	K	E	E	K	K
253													294
GAA	GAA	CAA	AAG	AAA	GGA	CCA	AGC	GAA	GAA	GAG	ATA	GGC	GGA
E	E	E	K	K	G	P	S	E	E	E	I	G	G
295							318						
GGC	TTA	TCA	TCA	CTA	TTC	GGT	TAA						
G	L	S	S	L	F	G	*						

Figure 24. Sequence of the Sso L12 gene. The derived amino acid sequence of the protein is shown below the nucleotide sequence. The N-terminal sequence that corresponds to the probe is underlined.

Table 10 Codon utilization in the Sso L12 gene

UUU	Phe	0	UCU	Ser	0	UAU	Tyr	0	UGU	Cys	0
UUC	Phe	1	UCC	Ser	0	UAC	Tyr	2	UGC	Cys	0
UUA	Leu	4	UCA	Ser	2	UAA	*	1	UGA	*	0
UUG	Leu	2	UCG	Ser	0	UAG	*	0	UGG	Trp	0
CUU	Leu	0	CCU	Pro	2	CAU	His	0	CGU	Arg	0
CUC	Leu	0	CCC	Pro	0	CAC	His	1	CGC	Arg	0
CUA	Leu	3	CCA	Pro	1	CAA	Gln	1	CGA	Arg	0
CUG	Leu	0	CCG	Pro	0	CAG	Gln	3	CGG	Arg	0
AUU	Ile	0	ACU	Thr	1	AAU	Asn	2	AGU	Ser	3
AUC	Ile	0	ACC	Thr	0	AAC	Asn	1	AGC	Ser	1
AUA	Ile	7	ACA	Thr	3	AAA	Lys	6	AGA	Arg	1
AUG	Met	2	ACG	Thr	0	AAG	Lys	6	AGG	Arg	0
GUU	Val	1	GCU	Ala	6	GAU	Asp	0	GGU	Gly	2
GUC	Val	0	GCC	Ala	1	GAC	Asp	2	GGC	Gly	2
GUA	Val	5	GCA	Ala	11	GAA	Glu	12	GGA	Gly	3
GUG	Val	1	GCG	Ala	0	GAG	Glu	4	GGG	Gly	0

* chain termination

L12 (*Halobacterium halobium*) (Itoh *et al.* 1988), and Sac L12 (*Sulfolobus acidocaldarius*) (Matheson *et al.* 1988). The alignment obtained is shown in Figure 25. The Hha L12 and Sac L12 sequences are not included in this figure because the Hha L12 sequence is identical to the Hcu L12 sequence (Itoh *et al.* 1988) and the only difference between the Sso L12 and Sac L12 is that the lysine at position 46 in the Sso L12 protein is a glutamic acid in the Sac L12 sequence (Matheson *et al.* 1988). There is 48% sequence identity (only identical residues were considered) between the Sso L12 and Mva L12 proteins and 44% between the Sso L12 and Hcu L12 proteins.

Table 11 Predicted Amino Acid Composition of the Sso L12 Protein

AMINO ACID	NUMBER	MOLES %	AMINO ACID	NUMBER	MOLES %
A	18	17.1%	M	2	1.9%
C	-	-	N	3	2.9%
D	2	1.9%	P	3	2.9%
E	16	15.2%	Q	4	3.8%
F	1	1.0%	R	1	1.0%
G	7	6.7%	S	6	5.7%
H	1	1.0%	T	4	3.8%
I	7	6.7%	V	7	6.7%
K	12	11.4%	W	-	-
L	9	8.6%	Y	2	1.9%

Molecular weight: 11, 139.02 pI: 4.74

The alignment of the Sso L12 protein with the eukaryotic P2 and P1 proteins is shown in Figures 26 and 27 respectively. The Sso L12 protein can be aligned end to end with the P2 eukaryotic acidic proteins, except for two gaps (alignment positions 61-64 and 76-81) in the Sso L12 protein and small gaps in the eukaryotic sequences, located in the alanine rich region and the highly charged domain (see Figure 26).

Alignment with the P1 proteins revealed that these proteins have 6 extra amino acids at their N-termini that are not present in the Sso L12 protein. There are four gaps in the Sso L12 sequence (positions 69-72, 83-84, 101 and 109) and several gaps in the eukaryotic sequences also located in the alanine rich region and the highly charged domain (see Figure 27). Table 12 summarizes the sequence identity between the Sso L12 protein and its eukaryotic counterparts. On average, there is 33% sequence identity between the Sso L12 protein and the eukaryotic acidic proteins.

										10										20
Sso L12	M	E	Y	I	Y	A	S	L	L	L	H	A	A	K	K	E	I	S	E	E
Mva L12	M	E	Y	I	Y	A	A	L	L	L	N	S	A	N	K	E	V	T	E	E
Hcu L12	M	E	Y	V	Y	A	A	L	I	L	N	E	A	D	K	E	L	T	E	D

										30										40
Sso L12	N	I	K	N	V	L	S	A	A	G	I	T	V	D	E	V	R	L	K	A
Mva L12	A	V	K	A	V	L	V	A	G	G	I	E	A	N	D	A	R	V	K	A
Hcu L12	N	I	T	G	V	L	E	A	A	G	V	D	V	E	E	S	R	A	K	A

										50										60
Sso L12	V	A	A	A	L	K	E	V	N	I	D	E	I	L	K	T	A	T	A	M
Mva L12	L	V	A	A	L	E	G	V	D	I	A	E	A	I	A	K	A	A	I	A
Hcu L12	L	V	A	A	L	E	D	V	D	I	E	E	A	V	E	E	A	A	A	A

										70										80
Sso L12	P	V	A	A	V	A	A	P	A	-	-	-	-	-	-	-	-	-	G	Q
Mva L12	P	V	A	A	A	A	P	V	A	A	A	A	A	P	A	E	V	-	-	-
Hcu L12	P	A	A	A	P	A	A	S	G	S	D	D	E	A	A	A	D	D	G	D

										90										100
Sso L12	Q	T	Q	Q	A	A	E	K	K	E	E	K	K	E	E	E	K	K	G	P
Mva L12	-	-	-	-	-	-	-	K	K	E	E	K	K	E	D	T	T	A	A	A
Hcu L12	D	D	E	E	H	D	A	D	E	A	A	E	A	E	D	A	G	D	D	D

										110				
Sso L12	S	E	E	E	I	G	G	G	L	S	S	L	F	G
Mva L12	A	A	-	-	-	-	-	G	L	G	A	L	F	M
Hcu L12	D	E	E	P	S	G	E	G	L	G	D	L	F	G

Figure 25. Alignment of archaebacterial L12 proteins. Identical residues between the Sso L12 protein and the other two archaebacterial sequences are boxed. Sso: *Sulfolobus solfataricus*, Mva: *Methanococcus vannielii* (Strobel *et al.* 1988), Hcu: *Halobacterium cutirubrum* (Shimmin and Dennis 1989).

```
          10                  20                  30                  40
Sso L12   M E Y I Y A S L L L H A A K K - E I S E E N I K N V L S A A G I T V D E V R - L
Dme L12   M R Y V A A Y L L A V L G G K D S P A N G D L E K I L S S V G V E V D A E R - L
Rno L12   M R Y V A S Y L L A A L G G N S N P S A K D I A K I L D S V G I E A D D E R K L
Sce L12   M K Y L A A Y L L L V Q G G N A A P S A A D I K A V V E S V G A E V D E A R - I
Asa L12   M R Y V A A Y L L A A L S G N A D P S T A D I E K I L S S V G I E C N P S Q - L
Hsa L12   M R Y V A S Y L L A A L G G N S S P S A K D I K K I L D S V G I E A D D D R - L
Spo L12   M K Y L A A Y L L L T V G G K D S P S A S D I E S V L S T V G I E A E S E R - I
Sce L12'  M K Y L A A Y L L L N A A G N T P D A T K I K A - I L E S V G I E I E D E K - V

          50                  60                  70                  80
Sso L12   K A V A A A L K E V - N I D E I L K T A - - - - T A M P V A A V A A P - - - - -
Dme L12   T K V I K E L A G K - S I D D L I K E G R E K L S S M P V G G G G A V A A A D A
Rno L12   N K V I S E L N G K - N I E D V I A Q G V G K L A S V P A G G A V A V S A A P G
Sce L12   N E L L S S L E G K G S L E E I I A E G Q K K F A T V P T G G A S - - S A G P A
Asa L12   Q K V M N E L K G K - D L E A L I A E G Q T K L A S M P T G G A - - P A A A A G
Hsa L12   N K V I S E L N G K - N I E D V I A Q G I G K L A S V P A G G A V A V S A A P G
Spo L12   E T L I N E L N G K - D I D E L I A A G N E K L A T V P T G G A A S A A A P A A
Sce L12'  S S V L S A L E G K - S V D E L I T E G N E K L A A V P A A G P A S A - - - - -

          90                  100                 110
Sso L12   - A G Q Q T Q Q A A E K K E E K K E E E K K G P S E E E - I G G G L S S L F G
Dme L12   A P A A A A G G D - - K K E A K K - E E K K E E S E S E D D D M G F - A L F E
Rno L12   A A A P A A G S A P A A A E E - - - - - - K E E S E E K K D E M G F - G L F D
Sce L12   S A G A A A G G G D - A A E E K - E E E A K E E S - - - D D D M G F - G L F D
Asa L12   G A A A A P A A E A - - K E A K K - E E K K E E S E E E D E D M G F - G L F D
Hsa L12   S A A P A A G S A P A A A E E K K - D E K K E E S E E S D D D M G F - G L F D
Spo L12   A A G G A A P A A E E A A - - - K - E E A K E E - E E S D E D M G F - G L F D
Sce L12'  - G G A A A A S G D A A A E E E K E E E A A - - - E E S D D D M G F - G L F D
```

Figure 26. Sequence alignment of the Sso L12 protein and the eukaryotic P2 acidic proteins. Identical residues between the Sso L12 protein and the eukaryotic sequences are boxed. Sso: *Sulfolobus solfataricus*, Dme: *Drosophila melanogaster* (Qian *et al.* 1987), Rno: *Rattus norvegicus* (Lin *et al.* 1982),Sce: *Saccharomyces cerevisiae*, Sce L12 (corresponds to L45, Remacha *et al.* 1988), Sce L12' (corresponds to L44, Remacha *et al.* 1988), Asa: *Artemia salina* (Maassen *et al.* 1985), Hsa: *Homo sapiens* (Rich and Steitz 1987), Spo: *Schizosaccharomyces pombe* (Beltrame and Bianchi 1987).

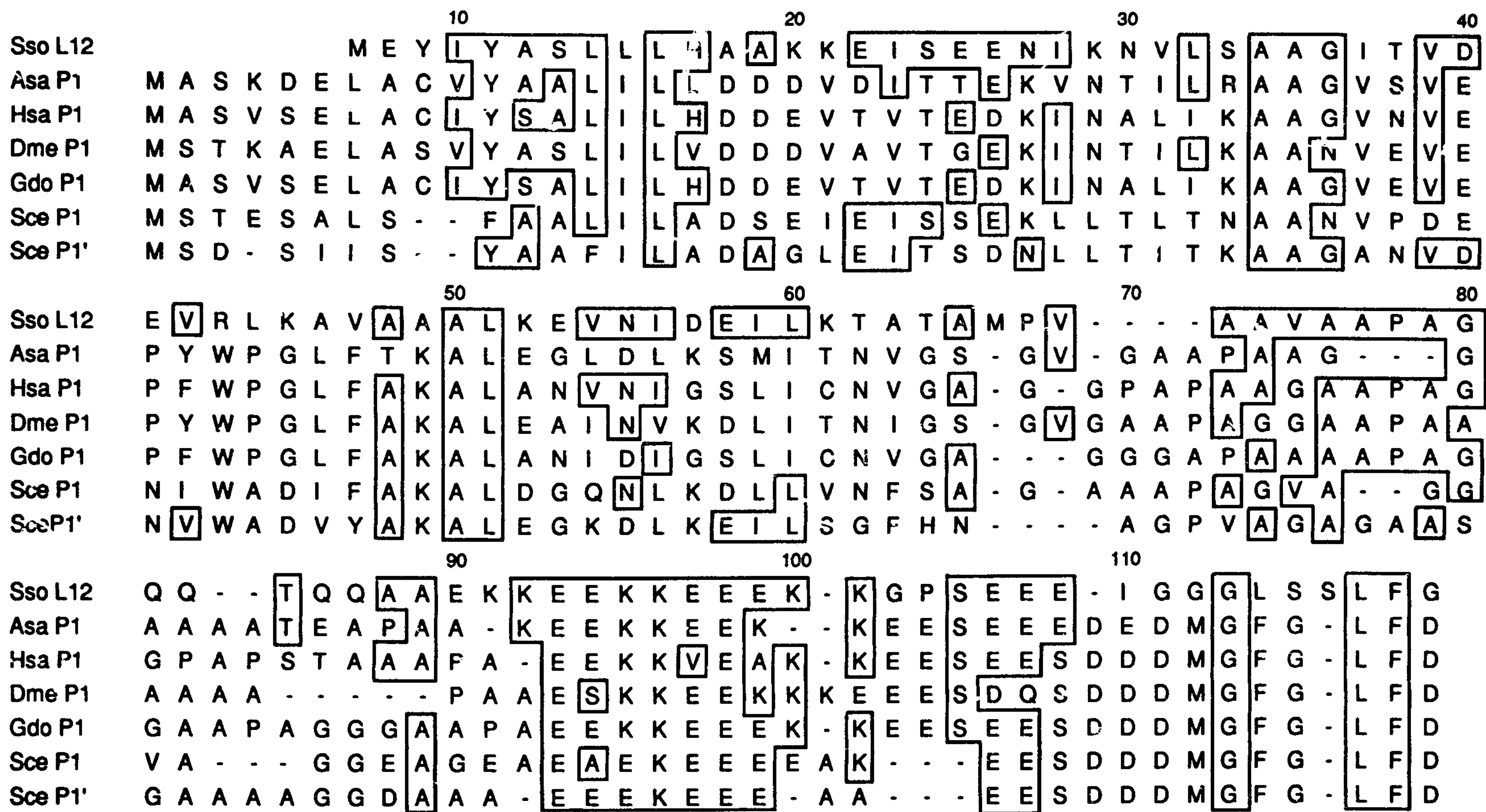


Figure 27. Alignment of the Sso L12 protein with the eukaryotic P1 acidic proteins. Identical residues between the Sso L12 protein and the eukaryotic sequences are boxed. Sso: *Sulfolobus solfataricus*, Asa: *Artemia salina* (Maassen *et al.* 1985), Hsa: *Homo sapiens* (Rich and Steitz 1987), Dme: *Drosophila melanogaster* (Wigboldus 1987), Gdo: *Gallus domesticus* (Ferro and Reinach 1988), Sce: *Saccharomyces cerevisiae*, Sce P1 (Mitsui and Tsurugi 1988b), Sce P1' (corresponds to L44') (Remacha *et al.* 1988).

Table 12 Sequence identity between the Sso L12 protein and its archaebacterial and eukaryotic counterparts.

Archaebacteria

protein[1]	length[2]	identities[3]	% identities
Mva L12	99	47	48%
Hcu L12	114	46	44%

Eukaryotic P2 Family

protein[4]	length[2]	identities[3]	% identities
Dme P2	113	37	35%
Rno P2	111	29	28%
Sce P2	110	32	31%
Asa P2	111	35	33%
Hsa P2	116	34	32%
Spo P2	111	34	32%
Sce P2'	107	32	31%

Eukaryotic P1 Family

protein[5]	length[2]	identities[3]	% identities
Asa P1	109	33	31%
Hsa P1	113	40	38%
Dme P1	111	31	30%
Gdo P1	113	38	36%
Sce P1	105	32	31%
Sce P1'	105	35	33%

[1] References can be found in Figure 25. [2] Length of the region of comparison [3] Only identical residues were considered. [4] References can be found in Figure 26. [5] References can be found in Figure 27

As mentioned in the Introduction, alignment of the archaebacterial-eukaryotic L12 proteins with their eubacterial counterparts has proved to be difficult and has led to the proposal of different models, in which different regions of the molecules are aligned to each other:

1] Amons *et al.* (1979) proposed that the archaebacterial-eukaryotic L12 proteins arose by the duplication and fusion of an eubacterial L12 protein. According to this model, there are 3 segments in the Eco L12 protein (48-64, 104-116, and 34-52) that are homologous to residues 12-30, 38-48 and 57-65 of the Hcu L12 protein respectively (see Figure 28A). This alignment is based on sequence similarity between the two proteins.

2] Yaguchi *et al.* (1980) proposed a linear alignment in which residues 35-110 of the Eco L12 protein are aligned with residues 1-75 of the Hcu L12 protein. There is 32% sequence identity between the two proteins in this alignment (see Figure 28B).

3] Lin *et al.* (1982) proposed a model in which the last 30 residues of the eubacterial L12 proteins are equivalent to the first 30 residues of the archaebacterial-eukaryotic acidic proteins (see Figure 28C). These authors postulate that the archaebacterial-eukaryotic proteins represent the ancestral sequence and that the eubacteria have moved the first 30 residues of this protein to the C-terminus. This alignment is based on sequence similarity between the two types of L12 proteins.

4] Matheson (1985) proposed a model in which the Eco L12 protein is divided into three segments: segment I, which contains the alanine rich region, segment II, which contains an acidic region and the conserved arginine and segment III, which contains the highly charged C-terminal domain (see Figure 28D). In this model, the archaebacterial-eukaryotic L12 proteins are generated by the transposition of segment I to the middle of the molecule. This alignment takes into account sequence and structural similarity.

5] Otaka *et al.* (1985, 1989) have proposed a similar model to the one proposed by Lin *et al.* (1982), but based on the results of a computer program

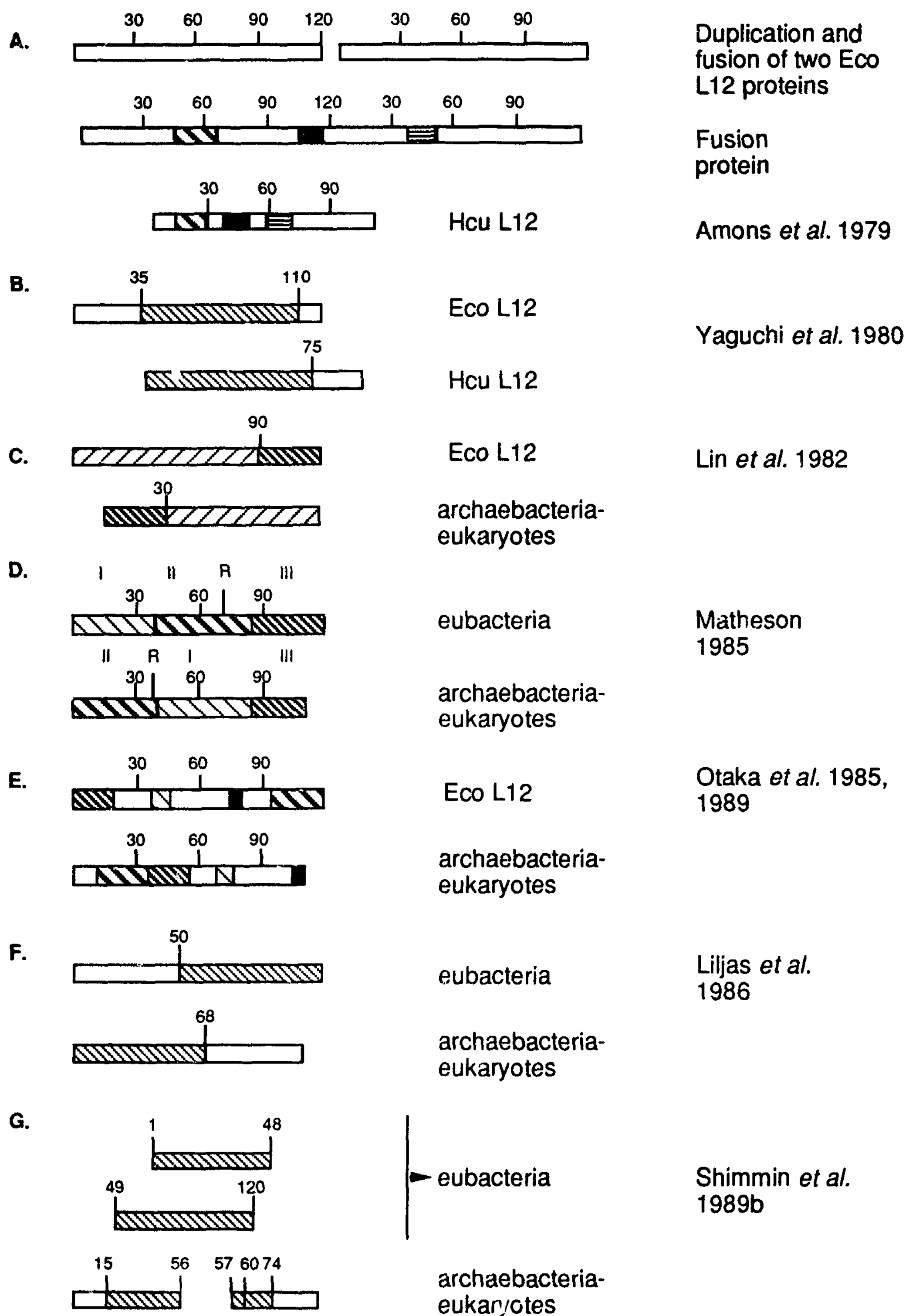


Figure 28. Different models for the alignment of the eubacterial and archaebacterial-eukaryotic L12 proteins. The filled areas represent the regions of equivalent structure between the two types of L12 proteins according to the different authors. For a description of each model, see the text.

that searches for homologies in tertiary structures. According to this model, there are 4 segments of the Eco L12 protein which have equivalents in the archaebacterial-eukaryotic acidic proteins: 96-120, 1-21, 40-48 and 75-82 (see Figure 28E). These authors believe that the eubacterial proteins represent the ancestral sequence.

6] Liljas *et al.* (1986) proposed a model based on secondary structure predictions in which the C-terminal region of the eubacterial proteins (residues 50-120) is aligned with residues 1-68 of the archaebacterial-eukaryotic proteins (see Figure 28F). According to these authors, the archaebacterial-eukaryotic proteins have an 'inverted' eubacterial structure, that is a globular N-terminal domain, followed by a hinge and an elongated C-terminus. As a consequence of this inverted structure, they propose that the archaebacterial-eukaryotic acidic proteins interact through their C-terminal domain with the ribosome, while the globular N-terminal domain is the site of interaction with the extrinsic factors. It should be noted, however, that recent studies by M. Remacha, T. Nacanda, S. Zinker, M.D. Vilella and J.P.G. Ballesta (Abst., Cold Spring Harbor: Ribosome Synthesis. 1988, p. 146) have shown that the N-terminal region rather than the C-terminal domain of the eukaryotic L12 proteins is involved in the binding to the ribosome.

7] Shimmin *et al.* (1989b) proposed a model based on sequence and structural similarity in which residues 1-48 and 49-120 of the Eco L12 protein are respectively aligned to residues 46-74 and 15-60 of the archaebacterial-eukaryotic proteins. In this model, residues 1-36 of the Eco L12 protein are aligned to residues 81-120 of the same protein and to residues 46-60 of the archaebacterial-eukaryotic acidic proteins (see Figure 28G). This model also

proposes an inverted structure for the archaebacterial-eukaryotic L12 proteins and considers them to be the ancestral type.

The fact that the eubacterial and archaebacterial-eukaryotic L12 proteins can be aligned in so many different ways, reflects the limitations of the data we have on the structure of these proteins. Sequence similarity has not been sufficient to establish a unique alignment in the case of these proteins, because their high alanine content makes it relatively easy to align different regions and still get a significant percentage of sequence similarity (models A-D). In order to solve this problem, several authors have tried to incorporate predictions on the secondary and even tertiary structure of the proteins into their models (Otaka *et al.* 1985, Liljas *et al.* 1986, Shimmin *et al.* 1989b). The limitation in this case is the reliability with which secondary and tertiary structure can be predicted from primary sequence data. For these reasons, it is difficult at the moment to decide which is the best model. Structural and functional studies of the L12 proteins from the three kingdoms are needed in order to define which are indeed the common structural-functional regions between the two types of acidic proteins.

In summary, even though there is no agreement about how to align the two types of L12 proteins, it is generally accepted that the two types are homologous and that they have suffered extensive rearrangements during the evolution of the three kingdoms.

The Sso L10 Gene and the Sso L10 Protein

The gene upstream of the Sso L12 gene was identified as the Sso L10 gene on the basis that part of the sequence of its product, matched the sequence of peptides obtained from the Sso L10 protein purified from

Sulfolobus ribosomes (Shimmin *et al.* 1989a) (see Figure 29). Furthermore, sequence similarity between the product of this gene and the Eco L10 ribosomal protein (Post *et al.* 1979) was also found.

The gene has 1008 nucleotides and codes for a protein of 335 amino acids (see Figure 29). Table 13 shows the codon usage in this gene (for a discussion, see section on codon usage) and Table 14 shows the amino acid composition of the protein.

The sequence of the Sso L10 protein was compared to the following sequences: Hcu L10 (*Halobacterium cutirubrum*) (Shimmin and Dennis 1989), Hha L10 (*Halobacterium halobium*) (Itoh *et al.* 1988), and Mva L10 (*Methanococcus vannielii*) (Köpke *et al.* 1989) from the archaebacteria; Hsa L10 (*Homo sapiens*) (Rich and Steitz 1987), Sce L10 (*Saccharomyces cerevisiae*) (Mitsui and Tsurugi 1988a) and Mmu L10 (*Mus musculus*) (Krowczynska *et al.* 1989) from the eukaryotes; and Eco L10 (*Escherichia coli*) (Post *et al.* 1979) from the eubacteria. The alignment obtained is shown in Figure 30. The Hha L10 sequence is not included in this figure because it is identical to the Hcu L10 sequence (Itoh *et al.* 1988). The Mmu L10 sequence is also not included because there are only nine differences between this sequence and the Hsa L10 protein (Krowczynska *et al.* 1989). Table 15 shows the sequence identity among the L10 proteins from the three kingdoms.

Sso L10 Gene

1													42
GTG	ATC	GGT	TTG	GCA	GTT	ACT	ACA	ACT	AAA	AAA	ATA	GCA	AAA
M	I	G	L	A	V	T	T	T	K	K	I	A	K
43													84
TGG	AAA	GTA	GAT	GAA	GTA	GCT	GAA	CTT	ACT	GAA	AAA	TTA	AAA
W	K	V	D	E	V	A	E	L	T	E	K	L	K
85													126
ACC	CAT	AAA	ACT	ATA	ATA	ATA	GCT	AAT	ATA	GAG	GGT	TTT	CCT
T	H	K	T	I	I	I	A	N	I	E	G	F	P
127													168
GCA	GAT	AAA	CTT	CAT	GAA	ATA	AGG	AAA	AAA	TTA	AGA	GGA	AAA
A	D	K	L	H	E	I	R	K	K	L	R	G	K
169													210
GCT	GAC	ATA	AAG	GTA	ACT	AAA	AAC	AAC	TTA	TTT	AAC	ATA	GCA
A	D	I	K	V	T	K	N	N	L	F	N	I	A
211													252
CTC	AAA	AAT	GCA	GGG	TAT	GAC	ACT	AAA	TTA	TTT	GAA	AGC	TAT
L	K	N	A	G	Y	D	T	K	L	F	E	S	Y
253													294
CTA	ACT	GGA	CCT	AAT	GCC	TTT	ATA	TTT	ACT	GAT	ACT	AAT	CCC
L	T	G	P	N	A	F	I	F	T	D	T	N	P
295													336
TTC	GAA	TTA	CAG	TTA	TTC	CTA	TCA	AAA	TTT	AAA	TTA	AAG	AGA
F	E	L	Q	L	F	L	S	K	F	K	L	K	R
337													378
TAT	GCC	CTG	CCA	GGA	GAT	AAG	GCA	GAT	GAG	GAG	GTT	GTT	GTA
Y	A	L	P	G	D	K	A	D	E	E	V	V	V
379													420
CCA	GCA	GGA	GAT	ACA	GGT	ATT	GCA	GCA	GGT	CCC	ATG	CTC	AGT
P	A	G	D	T	G	I	A	A	G	P	M	L	S
421													462
GTT	TTC	GGA	AAG	TTG	AAA	ATA	AAA	ACA	AAA	GTC	CAG	GAT	GGA
V	F	G	K	L	K	I	K	T	K	V	Q	D	G
463													504
AAA	ATT	CAC	ATA	TTA	CAA	GAC	ACA	ACA	GTT	GCA	AAG	CCT	GGA
K	I	H	I	L	Q	D	T	T	V	A	K	P	G

Figure 29. Sequence of the Sso L10 gene. The derived amino acid sequence of the protein is shown below the nucleotide sequence. The regions that correspond to peptides obtained from purified Sso L10 protein have been underlined (Shimmin *et al.* 1989a).

505 546
GAT GAG ATA CCA GCA GAT ATT GTA CCA ATA TTA CAA AAA CTA
D E I P A D I V P I L Q K L

547 588
GGA ATA ATG CCA GTA TAT GTA AAA TTA AAC ATT AAG ATA GCC
G I M P V Y V K L N I K I A

589 630
TAT GAT AAT GGA GTA ATA ATC CCA GGA GAT AAA TTA TCA ATT
Y D N G V I I P G D K L S I

631 672
AAC CTT GAT GAT TAT ACT AAT GAA ATC AGA AAA GCT CAT ATT
N L D D Y T N E I R K A H I

673 714
AAC GCA TTT GCA GTA GCC ACA GAA ATA GCG TAC CCA GAA CCA
N A F A V A T E I A Y P E P

715 756
AAG GTC CTT GAA TTT ACA GCA ACC AAA GCT ATG AGA AAT GCA
K V L E F T A T K A M R N A

757 798
TTA GCT CTA GCC AGT GAG ATA GGT TAT ATA ACT CAA GAG ACA
L A L A S E I G Y I T Q E T

799 840
GCA CAG GCA GTA TTT ACT AAA GCA GTA ATG AAA GCA TAC GCT
A Q A V F T K A V M K A Y A

841 882
GTT GCC TCT TCA ATA AGC GGA AAA GTA GAT TTA GGA GTA CAA
V A S S I S G K V D L G V Q

833 924
ATA CAA GCA CAA CCA CAG GTC TCT GAA CAG GCA GCT GAG AAG
I Q A Q P Q V S E Q A A E K

925 966
AAA GAA GAA AAG AAA GAA GAA GAA AAG AAA GGA CCA AGC GAA
K E E K K E E E K K G P S E

967 1008
GAA GAG ATA GGC GGA GGC TTA TCA TCA CTA TTC GGT GGA TGA
E E I G G G L S S L F G G *

Figure 29 ...Continued

Table 13 Codon utilization in the Sso L10 gene.

Codon	AA	No.	Codon	AA	No.	Codon	AA	No.	Codon	AA	No.
UUU	Phe	9	UCU	Ser	2	UAU	Tyr	7	UGU	Cys	0
UUC	Phe	4	UCC	Ser	0	UAC	Tyr	2	UGC	Cys	0
UUA	Leu	14	UCA	Ser	5	UAA	*	0	UGA	*	1
UUG	Leu	2	UCG	Ser	0	UAG	*	0	UGG	Trp	1
CUU	Leu	4	CCU	Pro	3	CAU	His	3	CGU	Arg	0
CUC	Leu	2	CCC	Pro	2	CAC	His	1	CGC	Arg	0
CUA	Leu	5	CCA	Pro	10	CAA	Gln	6	CGA	Arg	0
CUG	Leu	1	CCG	Pro	0	CAG	Gln	5	CGG	Arg	0
AUU	Ile	6	ACU	Thr	12	AAU	Asn	[illegible]	AGU	Ser	2
AUC	Ile	3	ACC	Thr	2	AAC	Asn	6	AGC	Ser	3
AUA	Ile	22	ACA	Thr	8	AAA	Lys	31	AGA	Arg	4
AUG	Met	4	ACG	Thr	0	AAG	Lys	10	AGG	Arg	1
GUU	Val	7	GCU	Ala	8	GAU	Asp	14	GGU	Gly	6
GUC	Val	3	GCC	Ala	6	GAC	Asp	3	GGC	Gly	2
GUA	Val	12	GCA	Ala	21	GAA	Glu	18	GGA	Gly	15
GUG	Val	1	GCG	Ala	1	GAG	Glu	8	GGG	Gly	1

* chain termination

Table 14 Predicted Amino Acid Composition of the Sso L10 Protein

AMINO ACID	NUMBER	MOLES %	AMINO ACID	NUMBER	MOLES %
A	36	10.7%	M	5	1.5%
C	-	-	N	13	3.9%
D	17	5.1%	P	15	4.5%
E	26	7.7%	Q	11	3.3%
F	13	3.9%	R	5	1.5%
G	24	7.1%	S	12	3.6%
H	4	1.2%	T	22	6.5%
I	31	9.2%	V	22	6.5%
K	41	12.2%	W	1	0.3%
L	28	8.3%	Y	9	2.7%

Molecular weight: 36, 548.63 pI: 9.56

The archaebacterial proteins are almost twice the size (335-352 amino acids) of the Eco L10 protein (165) and like their eukaryotic counterparts, they have a highly charged C-terminal domain.

The three archaebacterial L10 proteins can be aligned end to end. The Sso L10 and Hcu L10 proteins have 3 extra amino acids at their N-termini compared to the Mva L10 protein, while this latter protein has an extra aspartic acid residue at position 55, and five extra residues at positions 84-88 with respect to the Sso L10 and Hcu L10 proteins (see Figure 30).

Table 15 Sequence identity between the Sso L10 protein and its counterparts from the three kingdoms.

protein[1]	length[2]	identities[3]	% identities
Hcu L10	334	90	27%
Mva L10	331	109	33%
Hsa L10	316	77	24%
Mmu L10	316	77	24%
Sce L10	311	75	24%
Eco L10	165	34	21%

[1] References can be found in p 120. [2] Length of the region of comparison. [3] Only identical residues were considered.

The Sso L10 and Mva L10 proteins have suffered a large internal deletion at their C-terminal region with respect to the Hcu L10 protein (Sso L10: positions 307-324, Mva L10: positions 312-326). This deletion has eliminated an alanine rich region that precedes the highly charged C-terminal domain in the Hcu L10 protein (see Figure 30).

Sequence conservation among the three archaebacterial L10 proteins is evenly distributed through the length of the protein except for the C-terminal re-

gion (positions 300-363). In this region, the number of identities drops, mainly because in the Hcu L10 protein, the alternating lysine and glutamic acid residues (KKEEKK) present in the Sso L10 and Mva L10 proteins have been changed to aspartic acids. This increase in aspartic acid residues is thought to be an adaptation to the high intracellular salt concentration present in the extreme halophiles (Bayley and Morton 1978, Eisenberg and Wachtel 1987). Table 15 shows the percent sequence identity between the Sso L10 protein and its archaebacterial counterparts.

The archaebacterial sequences can also be aligned end to end with their slightly shorter eukaryotic counterparts (see Figure 30). The Sso L10 protein has 6 and 8 extra amino acids at its N-terminus compared to the Hsa L10 and Sce L10 proteins respectively, and an extra glycine at its C-terminus. The eukaryotic proteins, on the other hand, have 3 extra amino acids at positions 74-76 that are not present in the archaebacterial sequences. The Hsa L10 and Sce L10 proteins have an internal deletion (positions 224-249) with respect to the archaebacterial proteins. The two eukaryotic proteins, like the Hcu L10 protein have an alanine rich region preceding the highly charged C-terminal domain. As mentioned before, the Sso L10 and Mva L10 proteins lack this alanine rich region (see Figure 30). Table 15 shows the percent sequence identity between the Sso L10 protein and its eukaryotic counterparts.

The Eco L10 protein is about half the size of the archaebacterial and eukaryotic L10 proteins and lacks the highly charged C-terminal domain. The alignment of this protein with its archaebacterial and eukaryotic counterparts is shown in Figure 30. There is 21% sequence identity between the Sso L10 and the Eco L10 proteins.

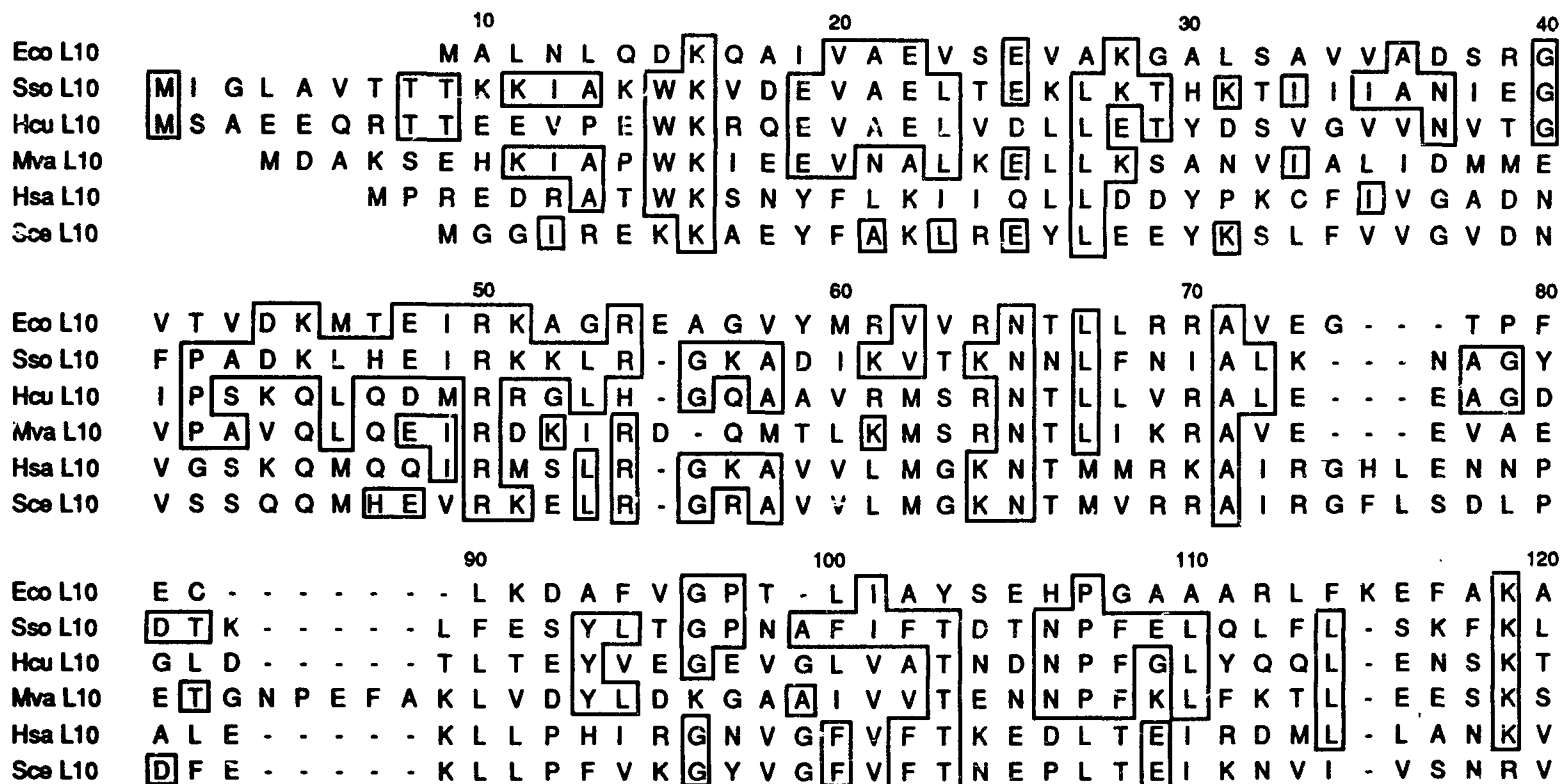


Figure 30. Sequence alignment of the L10 proteins from the three kingdoms. Identical residues between the Sso L10 protein and the other sequences are boxed. The position of three 26 amino acid modules that are repeated in the archaebacterial sequences is indicated (see text).. Eco:*Escherichia coli* (Post *et al.* 1979), Sso:*Sulfolobus solfataricus*, Hcu: *Halobacterium cutirubrum* (Shimmin and Dennis 1989), Mva: *Methanococcus vannielii* (Köpke *et al.* 1989), Hsa: *Homo sapiens* (Rich and Steitz 1987), Sce: *Saccharomyces cerevisiae* (Mitsui and Tsurugi 1988a).

```
                               130                 140                 150                 160
Eco L10   N A K F E V K A A A F E G E L I P A S Q I D R L A T L P T Y E E A - - - - - - -
Sso L10   K R Y A L P G D K A D E E V V V P A G D T G I A A G P M L S V F G K L K I K T K
Hcu L10   P A P I N A G E V A P N D I V V P E G D T G I D P G P F V G E L Q T I G A N A R
Mva L10   P A P I K G G A I A P C D I E V K S G S T G M P P G P F L S E L K A V G I P A A
Hsa L10   P A A A R A G A I A P C E V T V P A Q N T G L G P - E K T S F F Q A L G I T T K
Sce L10   P A P A R A G A V A P E D I W V R A V N T G M E P G - K T S F F Q A L G V P T K

                               170                 180                 190                 200
Eco L10   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Sso L10   V Q D G K I H I L Q D T T V A K P G D E I P A D I V P I L Q K L G I M P V Y V K
Hcu L10   I Q E G S I Q V L D D S V V T E E G E T V S D D V S N V L S E L G I E P K E V G
Mva L10   I D K G K I G I K E D K V V A K E G D V I S P K L A V V L S A L G I K P V T V G
Hsa L10   I S R G T I E I L S D V Q L I K T G D K V G A S E A T L L N M L N I S P F S F G
Sce L10   I A R G T I E I V S D V K V V D A G D K V G Q S E A S L L N L L N I S P F T F G

                               210                 220                 230                 240
Eco L10   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Sso L10   L N I K I A Y D N G V I I P G D K L S I N L D D Y T N E I R K A H I N A F A V A
Hcu L10   L V L R G V F S E G V L F T P E E L E I D V D E Y R A D I Q S A A A S A R N L S
Mva L10   L N V L G V Y E E G V I Y T S D V L R I D E E E F L G K L Q K A Y T N A F N L S
Hsa L10   L V I Q Q V F D N G S I Y N P E V L D I T E E - - - - - - - - - - - - - - - - -
Sce L10   L T V V Q V Y D N G Q V F P S S I L D I T D E - - - - - - - - - - - - - - - - -
                                                          ___________________________________
                                                                      α module
```

Figure 30 ...Continued

```
                                 250                          260                          270                          280
Eco L10   - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - I A R L M A T M K K A
Sso L10   T E I A Y P E P K V L E F T A T K A M R N A L A L A S E I G Y I T Q E T A Q A V F T
Hcu L10   V N A A Y P T E R T A P D L I A K G R G E A K S L G L Q A S V E S P D L A D D L V S
Mva L10   V N A V I P T S A T I E T I V Q K A F N D A K A V S V E S A F I T E K T A D A I L G
Hsa L10   - - - - - - - - - - - T L H S R F L E G V R N V A S V C L Q I G Y P T V A S V P H S I I
Sce L10   - - - - - - - - - - - E L V S H F V S A V S T I A S I S L A I G Y P T L P S V G H T L I

              α module                                β module

                        290                          300                          310                          320
Eco L10   S A G K L V R T L A A V R D A K E A A
Sso L10   K A V M K A Y A V A S S I S G K V D L G V Q I Q - - - - - - - - - - - - - - - - - -
Hcu L10   K A D A Q V R A L A A Q I D D E D A L P E E L Q D V D A P A A P A G G E A D T T A D
Mva L10   K A H A Q M I A V A - K L A G D E A L D D D L K E Q I S S - - - - - - - - - - - - -
Hsa L10   N G Y K R V L A L S V E T D Y T F P L A E K V K A F L A D P S A F V A A A P V A A A
Sce L10   N N Y K D L L A V A I A A S Y H Y P E I E D L V D R I E N P E K Y A A A A P A T S A

                      γ module

                      330                          340                          350                          360
Sso L10   A Q P Q V S E Q A A E K K E E K K E E E K K G P S E E E I G G - G L S S L F G G
Hcu L10   E Q S D E T Q A S E A D D A D D S D D D D D D D D G N A G A E - G L G E M F G
Mva L10   - - S A V V A T E E A P K A E T K K E E K K E E - - A A P A A - G L G L L F
Hsa L10   T T A - - - - - - - A P A A A A P A K V E A K E E S E E S D E D M G F G - L F D
Sce L10   A S G D - - - - - - - - - - A A P A E E A A A E E E E E S D D D M G F G - L F D
```

Figure 30 ...Continued

Sso α D Y T N E I R K A H I N A F A V A T E I A Y P E P K
Sso β V L E F T A T K A M R N A L A L A S E I G Y I T Q E
Sso γ T A Q A V F T K A V M K A Y A V A S S I S G K V D L

Hcu α E Y R A D I Q S A A A S A R N L S V N A A Y P T E R
Hcu β T A P D L I A K G R G E A K S L G L Q A S V E S P D
Hcu γ L A D D L V S K A D A Q V R A L A A Q I D D E D A L

Mva α E F L G K L Q K A Y T N A F N L S V N A V I P T S A
Mva β E I E T I V Q K A F N D A K A V S V E S A F I T E K
Mva γ T A F A I L G K A H A Q M I A V A - K L A G D E A L

Hsa β T L H S R F L E G V R N V A S V C L Q I G Y P T V A
Hsa γ S V P H S I I N G Y K R V L A L S V E T D Y T F P L

Sce β E L V S H F V S A V S T I A S I S L A I G Y P T L P
Sce γ S V G H T L I N N Y K D L L A V A I A A S Y H Y P E

Eco γ M A T M K K A S A G K L V R T L A A V R D A K E A A

Figure 31. Sequence alignment of the modules in the L10 proteins. The putative module in the Eco L10 protein is included for a comparison.

The three archaebacterial L10 proteins contain a sequence of 26 amino acids that appears to be repeated three times (designated modules α [positions 224-249], β [positions 250-275], and γ [positions 276-301] (Figure 30). These modules can also be identified in the eukaryotic L10 proteins, although the α module appears to be missing and sequence conservation is not as high as in the case of the archaebacterial L10 proteins (see Figure 31). In *E. coli*, the γ module, also appears to have been conserved. However, it should be noted that the identification of this region of the Eco L10 protein as the γ module, is based mainly on the alignment of the Eco L10 protein with the Hcu L10 protein. There are 9 identities between the Eco L10 and Hcu L10 proteins in this region, while there are only 4 between the Sso L10 and Eco L10 proteins (see Figure 30) (5 between Mva L10 and Eco L10, 4 between Hsa L10 and Eco L10, and 2 between Sce L10 and Eco L10). For this reason, the identification of this region as a module is still tentative. More eubacterial L10 sequences are needed to confirm if this is indeed a module. The alignment of the modules is shown in Figure 31. Note that the region of highest sequence similarity is located towards the right half of the modules.

Sequence Conservation between the L10 and L12 Proteins

Sequence similarity has been found between the highly charged C-terminal domain of the archaebacterial and eukaryotic L10 proteins and the corresponding L12 proteins (Shimmin *et al.* 1989, Ramírez *et al.* 1989b, Itoh *et al.* 1988, Köpke *et al.* 1989, Rich and Steitz 1987, Mitsui and Tsurugi 1988a) (see Figure 32). In the case of *Sulfolobus solfataricus*, the C-terminal region of the Sso L10 and Sso L12 proteins (31 amino acids) is identical, except for an extra glycine residue at the end of the L10 protein. Remarkably, the sequence

of this region is also identical at the nucleotide level (see Figure 33). The reason for this complete conservation is not known, but it probably represents a case of gene conversion (Lewin 1987). It is interesting to note in this respect, that a similar situation has been found between two viral genes from the SSV1 virus that infects *Sulfolobus* B12, where a region coding for 20 amino acids has been perfectly conserved at the nucleotide level (Reiter *et al.* 1987a).

Comparison of the L10 and L12 proteins can be extended beyond the highly charged C-terminal domain. In the case of the eukaryotic and halobacterial sequences, the conservation of the alanine rich region can also be observed (Shimmin *et al.* 1989b, Ramírez *et al.* 1989b). However, the alanine rich region is not present in the Sso and Mva L10 proteins (see Figures 30 and 32). It appears that a copy of the γ module has also been conserved, although only the archaebacterial sequences exhibit substantial sequence similarity (identities and conservative substitutions) in this region (Hcu L10-Hcu L12 [42%], Mva L10-Mva L12 [33 %], Sso L10-Sso L12 [31 %], compared to Sce L10-Sce L12 [15 %], and Hsa L10-Hsa L12 [12 %] in the eukaryotes). If the γ module has indeed been conserved, then almost 75% of the L12 protein is duplicated in the L10 archaebacterial and eukaryotic proteins (Shimmin *et al.* 1989b, Ramírez *et al.* 1989b).

Figure 32. Alignment of the common regions between the L10 and L12 proteins from the archaebacteria and the eukaryotes. Identities are indicated by a dot. The regions corresponding to the γ module, the alanine rich region and the charged region are indicated. Note that the Sso L10 and Mva L10 proteins lack an alanine rich region. Positive and negative charges are indicated.

```
Sso L10  TAQAVFTKAVMKA YA -VA-SSISGKVDLGVQIQA---QPQVSEQAAEKKEEKKEE EKKGPS EEEI GGGLSSLFGG
Sso L12  TVDEVRLKAVA AA LKEVNIDEILKTATAMPVAAVAAPAGQQTEQAAEKKEEKKEE EKKGPS EEEI GGGLSSLFG
         γ module | 'alanine rich region' | charged region

Hcu L10  LADDLVSKADAQVRALA-AQIDDEDALPEELQDVDAPAAPAGGEADTTADEQSDETQASEADDADDSDDDDDDDGNAGAEGLGEMFG
Hcu L12  GVDVEESRAKALVAALEDVDI--EEAV-EEAAAAPAAAPAASGSDDEAAADDGDDDEEADADEAAEAEDAGDDDDEEPSGEGLGDLFG
         γ module | 'alanine rich region' | charged region

Mva L10  TADAILGKAHAQMIAVAKLAGDEALDDDLKEQISSSAVVATEE---APKAETKKEEKKEE--AAPAAGLGLLF
Mva L12  GIEANDARVKALVAALEGVDIAEAIAKAAIAPVAAPVAAAAPVAAAAAPAEVKKEEKKEDTTAAAAAGLGALFM
         γ module | 'alanine rich region' | charged region

Hsa L10  SVFHSI!NGYKRVLALSVETDYTFPLAEKVKAFLADPSAFVAAAPVAAATTAAPAAAAAPA-----KVEAKEESEESDEDMGFGLFD
Hsa L12  GIEADDDRLNKVISELNGKNIEDVIAQGIGK--LASVPAGGAVAVSAAPGSAAPAAGSAPAAAEEKKDEKKEESEESDDDMGFGLFD
         γ module | 'alanine rich region' | charged region

Sce L10  SVGHTLINNYKDLLAVAIAASYHYPEIEDLVDRIENPEKYAAAAPATSAAS-----GDAAPAEEAAAEEEEESDDDMGFGLFD
Sce L12  EVDEARINELLSSLESLEEIIAEGQTKLASM-----PTGGASAGPA-SAGAAAAGGGDAA--EEKEEEAKEESDDDMGFGLFD
         γ module | 'alanine rich region' | charged region
```

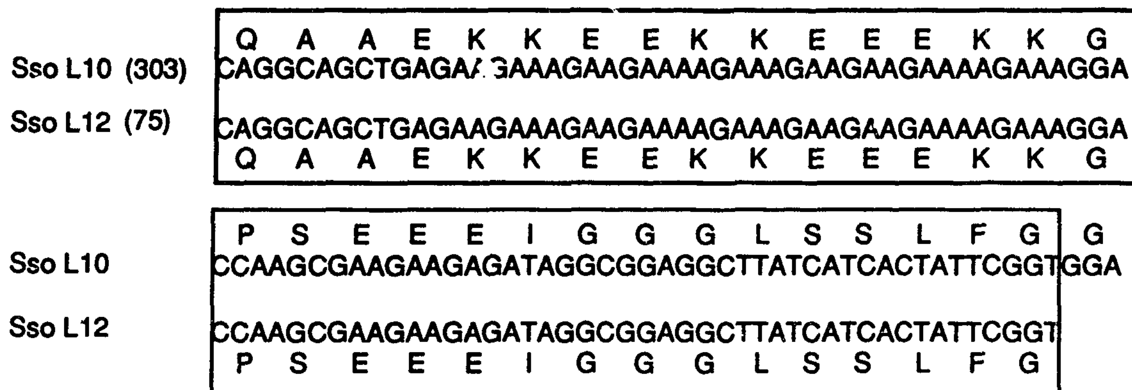


Figure 33. Identical regions present in the Sso L10 and Sso L12 genes and proteins. The amino acid sequence of the protein is shown above and below the nucleotide sequence. The identical region is boxed. The only difference in this region between the two genes is the presence of an extra glycine codon at the end of the Sso L10 gene. Numbers in parenthesis indicate the position of the residues in each protein.

Evolution of the L10 and L12 Genes and Proteins

Sequence comparison of the L10 and L12 proteins from the three kingdoms suggests that these two genes are derived from common ancestral genes and that they have undergone many alterations and rearrangements during the evolution of the three kingdoms. Based on the common structural features of these genes, and the current views regarding the origin of the three kingdoms (Woese 1987), a hypothetical model can be proposed to explain the evolution of these genes (see Figure 34).

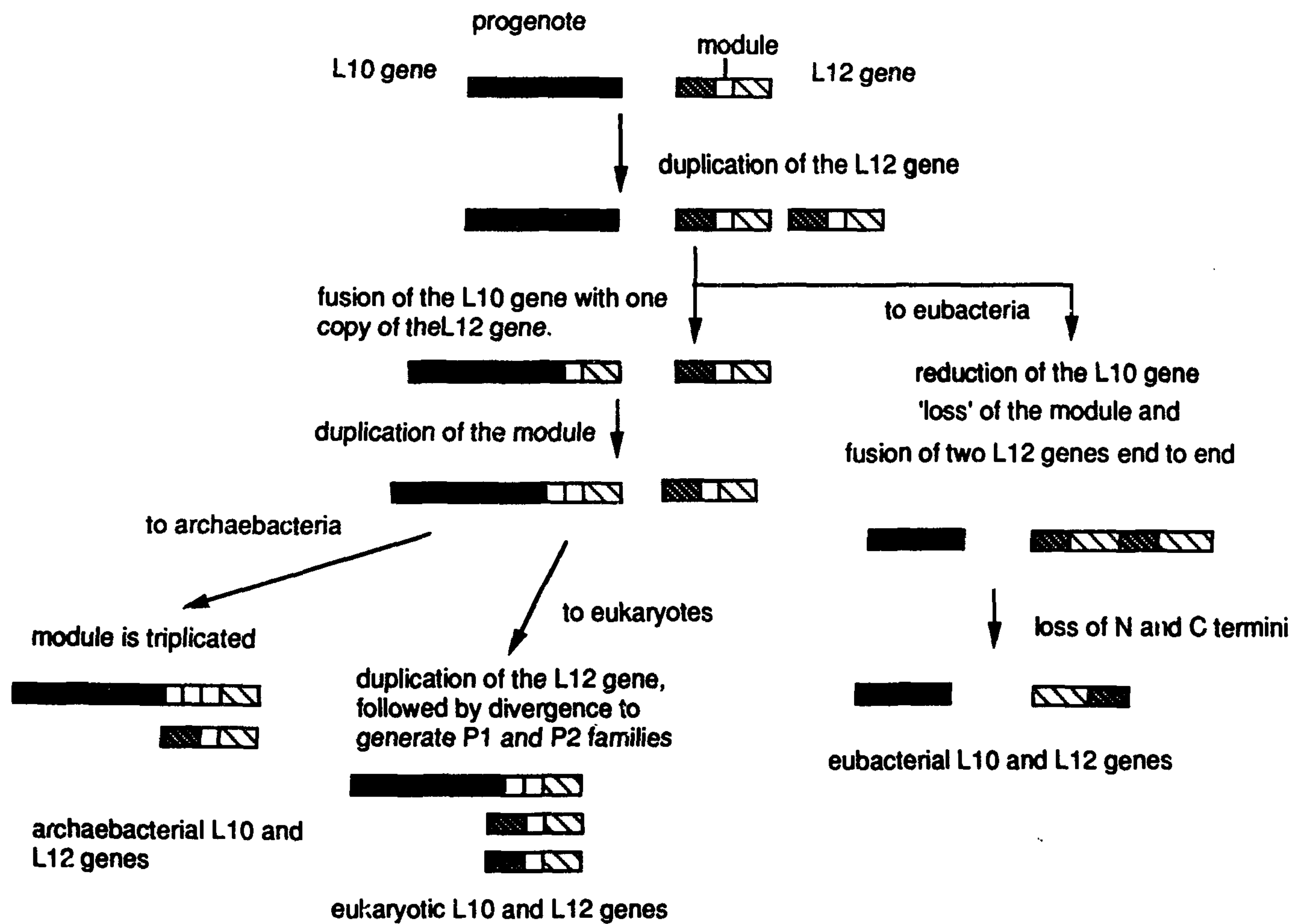


Figure 34. Possible model for the evolution of the L10 and L12 genes in the three kingdoms. For a description see the text.

In the progenote, or common ancestor, of the three kingdoms, the ancestral L10 gene probably coded for a protein that lacked the region that the L10 and L12 proteins have in common, *i.e.* the modules and the highly charged C-terminal domain (Shimmin *et al.* 1989b, Ramírez *et al.* 1989b). Since it can be assumed that traits that are present in more than one line of descent were probably present in the common ancestor, the ancestral L12 gene probably coded for a protein with a structure similar to the archaebacterial-eukaryotic L12 proteins. In the next stage, the L12 gene was duplicated. This event was later followed by a fusion of the ancestral L10 gene with one of the copies of the L12 gene, producing a hybrid gene (Shimmin *et al.* 1989b, Ramírez *et al.* 1989b). Since the modules and the highly charged C-terminal domain come from the L12 gene, the site of the fusion appears to be located at the point where the modules start. A duplication of this module generated an L10 protein with a structure similar to that found in the eukaryotes. At this stage the archaebacteria and the eukaryotes begin to diverge. In the archaebacteria, the module is duplicated once more giving rise to the present archaebacterial L10 gene with three modules. In the eukaryotes, a second duplication of the L12 gene takes place, giving rise to the P1 and P2 families.

It is difficult at the moment, to propose a unique model for the evolution of the eubacterial genes mainly because there is no agreement regarding the way in which the eubacterial and archaebacterial-eukaryotic L12 type proteins should be aligned. A very simple model that illustrates how a transposition of the type proposed by Lin *et al.* (1982) could have ocurred is illustrated in Figure 34. According to this model, the eubacteria would have diverged just before the fusion of one of the copies of the L12 gene with the L10 gene (in this model, the possible presence of such a module in the eubacterial L10 protein

is not considered since the existence of a module has not been definitively established). At this stage, loss of the C-terminal domain of the ancestral L10 gene would have generated the smaller eubacterial L10 gene seen today. The module present in the ancestral L12 protein would have also been lost at this time either by sequence divergence so that it is no longer recognizable in the modern eubacterial sequences or by actual deletion. A fusion between the two L12 genes, followed by a deletion of the N and C termini of the fusion product would have produced the present eubacterial L12 gene in which the C-terminal region of the original gene has been transposed. It should be noted that a more complex scheme has to be invoked in order to explain the origin of the eubacterial L12 proteins when other alignments, such as those proposed by Matheson (1985) or Shimmin *et al.* (1989b) are considered (see for example, Shimmin *et al.* 1989b).

The Sso L1 Gene and the Sso L1 protein

The gene upstream of the Sso L10 gene was identified as the Sso L1 gene on the basis of sequence similarity between its product and the Eco L1 protein (Post *et al.* 1979). The gene is 663 nucleotides long and codes for a protein of 221 amino acids (see Figure 35). Table 16 shows the codon usage in this gene (for a discussion, see section on codon utilization) and Table 17 shows the amino acid composition, molecular weight and pI of the protein.

In *Escherichia coli*, the L1 protein is involved in the interaction between the peptidyl-tRNA and the ribosome at the P (peptidyl) and E (exit) sites and indirectly with the GTPase center (Subramanian and Dabbs 1980, Sander 1983, Moazed and Noller 1989). Protein L1 is located on a ridge in the lateral protuberance opposite to the L12 stalk (Lake and Strycharz 1981) and binds to

1 42
GTG AAG AAA GTG TTA GCG GAT AAA GAA TCG TTG ATA GAA GCC
M K K V L A D K E S L I E A

43 84
TTA AAA CTG GCA CTT AGT ACG GAG TAT AAT GTA AAA AGG AAC
L K L A L S T E Y N V K R N

85 126
TTT ACA CAA AGT GTG GAA ATT ATA CTT ACG TTT AAA GGA ATT
F T Q S V E I I L T F K G I

127 168
GAT ATG AAG AAA GGT GAC CTT AAA CTT AGA GAA ATA GTT CCC
D M K K G D L K L R E I V P

169 210
TTA CCA AAA CAG CCA TCT AAA GCA AAG AGG GTT CTT GTT GTT
L P K Q P S K A K R V L V V

211 252
CCA TCC TCC GAA CAA CTC GAA TAT GCT AAA AAA GCA TCT CCC
P S F E Q L E Y A K K A S P

253 294
AAA GTT GTG ATA ACT AGA GAA GAA TTA CAA AAA TTA CAA GGA
N V V I T R E E L Q K L Q G

295 336
CAA AAA AGA CCA GTG AAA AAA TTA GCT ATA CAA AAT GAG TGG
Q K R P V K K L A I Q N E W

337 378
TTT TTA ATT AAC CAG GAA TCT ATG GCA CTG GCA GGA AGG ATA
F L I N Q E S M A L A G R I

379 4'
TTA GGA CCC GCC TTA GGA CCT AGA GGT AAG TTT CCT ACA CC,
L G P A L G P R G K F P T P

Figure 35. Sequence of the Sso L1 gene. The derived amino acid sequence of the protein is shown below the nucleotide sequence. The gene uses GTG as the initiation codon.

421													462
CTA	CCC	AAT	ACA	GCT	GAT	ATC	AGT	GAA	TAT	ATT	AAT	AGA	TTC
L	P	N	T	A	D	I	S	E	Y	I	N	R	F
463													504
AAG	AGA	TCA	GTT	ATA	GTA	AAG	ACC	AAG	GAT	CAG	CCA	CAA	GTT
K	R	S	V	I	V	K	T	K	D	Q	P	Q	V
505													546
CAA	GTA	TTC	ATA	GGA	ACT	GAG	GAC	ATG	AAG	CCT	GAA	GAC	TTA
Q	V	F	I	G	T	E	D	M	K	P	E	D	L
547													588
GCT	GAG	AAT	GCG	ATA	GCC	GTA	TTA	AAT	GCA	ATA	GAG	AAT	AAA
A	E	N	A	I	A	V	L	N	A	I	E	N	K
589													630
GCA	AAG	GTA	GAA	ACT	AAT	CTA	AGA	AAT	ATT	TAT	GTT	AAA	ACA
A	K	V	E	T	N	L	R	N	I	Y	V	K	T
631											663		
ACA	ATG	GGT	AAA	GCT	GTA	AAA	GTT	AAA	AGA	GCG	TGA		
T	M	G	K	A	V	K	V	K	R	A	*		

Figure 35...Continued

nucleotides 2100-2200 of the 23S rRNA (Gourse *et al.* 1981, Branlant *et al.* 1981).

The sequence of the Sso L1 protein was aligned with two other archaebacterial sequences: Hcu L1 (*Halobacterium cutirubrum*) (Shimmin and Dennis 1989) and Hha L1 (*Halobacterium halobium*) (Itoh 1988) and four eubacterial sequences: Eco L1 (*Escherichia coli*) (Post *et al.* 1979), Bst L1 (*Bacillus stearothermophilus*) (Kimura *et al.* 1985), Pvu L1 (*Proteus vulgaris*) and Sma L1 (*Serratia marcescens*) (Sor and Nomura 1987). The alignment obtained is shown in Figure 36. The Pvu L1 and Sma L1 sequences are not shown in this figure because they are virtually identical to the Eco L1 protein (Sor and Nomura 1987). The Hha L1 sequence is also not included in this figure because it only differs in one amino acid from the Hcu L1 protein (the

Hha L1 protein has an alanine instead of a valine at position 134 (see Figure 36) (Itoh 1988).

The two archaebacterial proteins can be aligned end to end with the introduction of four gaps.(see Figure 36). The Sso L1 sequence has 4 extra residues at its N-terminus and 2 extra residues at its C-terminus compared to the Hcu L1 protein. The Hcu L1 protein has 2 direct repeats: **ADDV** (residues at alignment positions 89-92 and 94-97) and **DLAD[D/E]TD** (positions: 105-111, and 115-121). There is no evidence of these repeats in the Sso L1 sequence.

Table 16 Codon utilization in the Sso L1 gene

UUU	Phe	4	UCU	Ser	3	UAU	Tyr	4	UGU	Cys	0
UUC	Phe	3	UCC	Ser	1	UAC	Tyr	0	UGC	Cys	0
UUA	Leu	11	UCA	Ser	1	UAA	*	0	UGA	*	1
UUG	Leu	1	UCG	Ser	1	UAG	*	0	UGG	Trp	1
CUU	Leu	5	CCU	Pro	4	CAU	His	0	CGU	Arg	0
CUC	Leu	1	CCC	Pro	4	CAC	His	0	CGC	Arg	0
CUA	Leu	2	CCA	Pro	5	CAA	Gln	8	CGA	Arg	0
CUG	Leu	2	CCG	Pro	0	CAG	Gln	3	CGG	Arg	0
AUU	Ile	5	ACU	Thr	3	AAU	Asn	10	AGU	Ser	3
AUC	Ile	1	ACC	Thr	1	AAC	Asn	2	AGC	Ser	0
AUA	Ile	10	ACA	Thr	5	AAA	Lys	20	AGA	Arg	8
AUG	Met	4	ACG	Thr	2	AAG	Lys	9	AGG	Arg	3
GUU	Val	9	GCU	Ala	5	GAU	Asp	4	GGU	Gly	3
GUC	Val	0	GCC	Ala	3	GAC	Asp	3	GGC	Gly	0
GUA	Val	6	GCA	Ala	7	GAA	Glu	12	GGA	Gly	6
GUG	Val	5	GCG	Ala	3	GAG	Glu	5	GGG	Gly	0

* chain termination

The two archaebacterial sequences can also be aligned end to end to their eubacterial counterparts. However, several gaps had to be introduced to obtain this alignment (see Figure 36). The eubacterial proteins are longer than the archaebacterial proteins. The Eco L1 and Bst L1 proteins have 9 extra residues at their N-terminus compared to the Sso L1 protein, and 6 and 5 extra residues respectively at their C-terminus.

The direct repeats present in the Hcu L1 protein are not evident in the eubacterial proteins. Furthermore, a gap had to be introduced in the eubacterial sequences at the region where the second repeat is located (positions 107-113) (see Figure 36).

Table 17 Predicted Amino Acid Composition of the Sso L1 Protein

AMINO ACID	NUMBER	MOLES %	AMINO ACID	NUMBER	MOLES %
A	18	8.1%	M	5	2.3%
C	-	-	N	12	5.4%
D	7	3.2%	P	13	5.9%
E	17	7.7%	Q	11	5.0%
F	7	3.2%	R	11	5.0%
G	9	4.1%	S	9	4.1%
H	-	-	T	11	5.0%
I	16	7.2%	V	19	8.6%
K	29	13.1%	W	1	0.5%
L	22	10.0%	Y	4	1.8%

Molecular weight: 24, 895.29 pI: 10.81

There are three regions of high sequence conservation among the L1 proteins from the two kingdoms: alignment positions 137-150, 193-198 and 226-230 (see Figure 36). Since these regions have been conserved in the two kingdoms, it is probable that they play an important role in the structure and/or

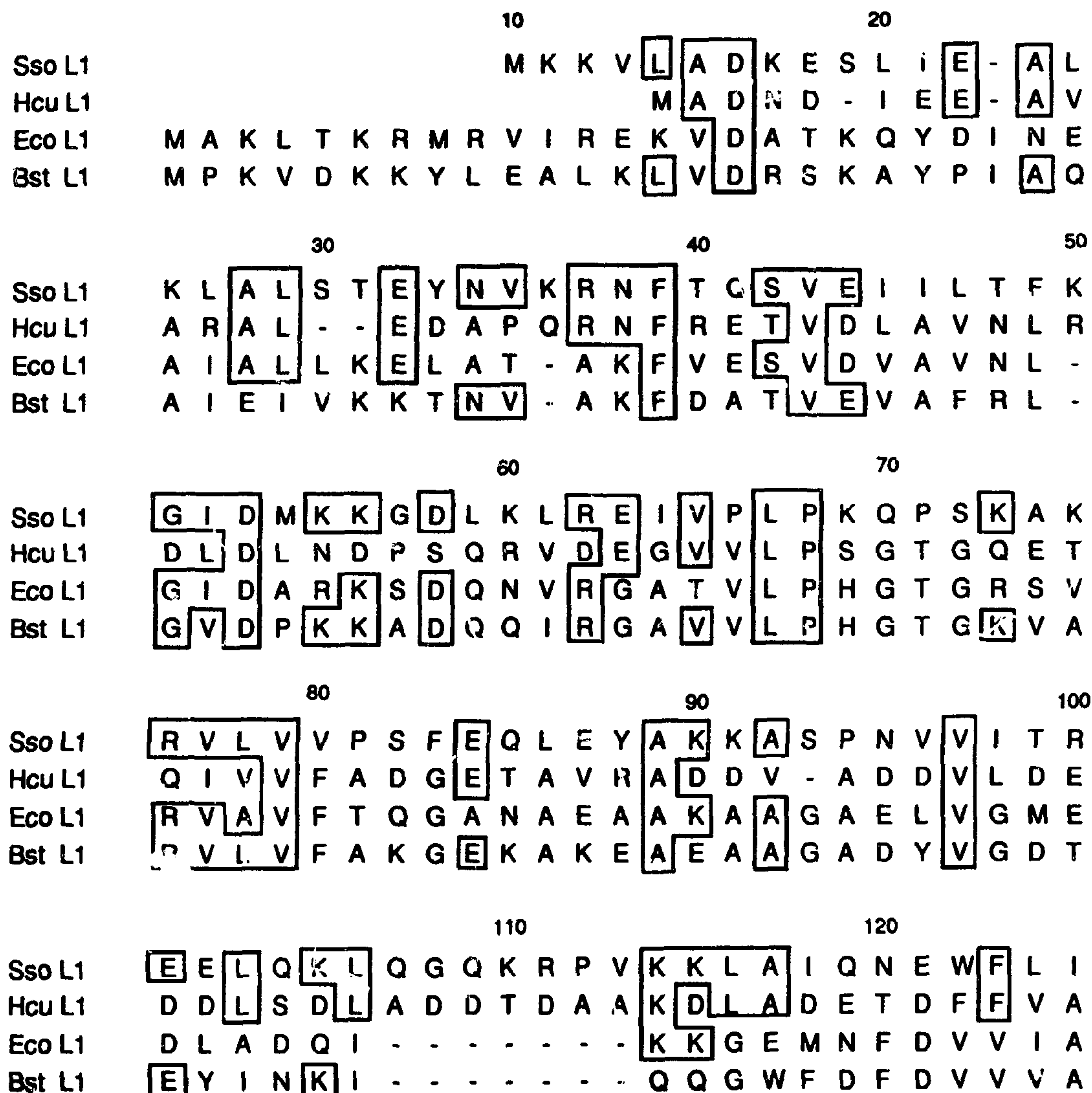


Figure 36. Sequence alignment of the Sso L1 protein with its archaebacterial and eubacterial counterparts. Identical residues between the Sso L1 protein and the other sequences are boxed. Sso: *Sulfolobus solfataricus*, Hcu: *Halobacterium cutirubrum* (Shimmin and Dennis 1989), Eco:*Escherichia coli* (Post *et al.* 1979), Bst: *Bacillus stearothermophilus* (Kimura *et al.* 1985).

```
                        130                     140                   150
Sso L1   N Q E S M A L A G R I L G P A L G P R G K F P T P
Hcu L1   E A P M M Q D I V G A L G Q V L G P R G K M P T P
Eco L1   S P D A M R V - V G Q L G Q V L G P R G L M P N P
Bst L1   T P D M M G E - V G K L G R I I G P K G L M P N P

                                  160                  170
Sso L1   L P N - - T A D I S E Y I N R F K - R S V I V K T
Hcu L1   L Q P - - D D D V V D T V N R M K - N T V Q I R S
Eco L1   K V G T V T P N V A E A V K N A K A G Q V R Y R N
Bst L1   K T G T V T F D V A K A V Q E I K A G K V E Y R V

                        180                     190                   200
Sso L1   K D Q P Q V Q V F I G T E D M K P E D L A E N A I
Hcu L1   R D R R T F H T R V G A E D M S A E D I A S N I D
Eco L1   D K N G I I H T T I G K V D F D A D K L K E N L E
Bst L1   D K A G N I H V P I G K V S F D N E K L A E N F A

                                  210                  220
Sso L1   A V L - - - - N A I E N K A K V E T N L R N I Y V
Hcu L1   V I M - - - R R L H A N L E K G P L N V D S V Y V
Eco L1   A L L V A L K K A K P T Q A K G - V Y I K K V S I
Bst L1   A V Y E A I K K A K P A A A K G - T Y V K N V T I

                        230                     240
Sso L1   K T T M G K A V K V K R A
Hcu L1   K T T M G P A V E V A
Eco L1   S T T M G N I H V P I G K V S F D N E K
Bst L1   T S T M G P G I K V D P T T V A V A Q
```

Figure 36...Continued

function of the protein. Table 18 summarizes the percent sequence identity between the Sso L1 protein and its archaebacterial and eubacterial counterparts.

As stated before, in *E.coli,* the L1 protein binds to a region in the 23S rRNA located between nucleotides 2100-2200 (Gourse *et al.* 1981, Branlant *et al.* 1981). Binding studies using the Eco L1 protein and large subunit rRNA from the archaebacteria and eukaryotes, have indicated that the protein binding site has been conserved in the three kingdoms (Zimmermann *et al.* 1980, Gourse *et al.* 1981). Although the RNA binding site on the protein has not been identified, it is possible that some of the regions of the L1 proteins that have been conserved in the two kingdoms might be involved in the interaction with rRNA (see the following section).

Table 18. Sequence identity between the Sso L1 protein and its archaebacterial and eubacterial counterparts.

protein [1]	length [2]	identities[3]	% identities
Hcu L1	214	66	31%
Eco L1	220	49	22%
Pvu L1	220	47	21%
Sma L1	220	47	21%
Bst L1	220	60	27%

[1] References can be found in p 136. [2] Length of the region of comparison [3] Only identical residues were considered

The Sso L11 Gene and the Sso L11 Protein

The gene immediately upstream of the Sso L1 gene was identified as the Sso L11 gene on the basis that the N-terminal sequence of its product matched the N-terminal amino acid sequence of the Sso L11 protein isolated from *Sulfolobus* ribosomes (Matheson 1985). Furthermore sequence similarity was found between its product and the Eco L11 ribosomal protein (Dognin and

Wittmann-Liebold 1977). The gene has 513 nucleotides and codes for a protein of 170 amino acids (see Figure 37). Table 19 shows the codon usage in this gene (for a discussion, see section on codon usage).and Table 20 the amino acid composition, molecular weight and pI of the protein.

As stated in the introduction, the L11 protein is part of the L7/L12 domain and is involved in the GTPase center of the ribosome (Schmidt *et al.* 1981, Cundliffe 1986). The sequence of the Sso L11 protein was compared to the sequences of the following L11 proteins: Eco L11 (*Escherichia coli*) (Dognin and Wittmann-Liebold 1977), Pvu L11 (*Proteus vulgaris*), and Sma L11 (*Serratia marcescens*) (Sor and Nomura 1987) from the eubacteria; Hcu L11 (*Halobacterium cutirubrum*) (Shimmin and Dennis 1989) from the archaebacteria; and the N-terminal sequence of Sce L11 (*Saccharomyces cerevisiae*) (Otaka *et al.* 1984) from the eukaryotes. The alignment obtained is shown in Figure 38. The sequences of the Pvu L11 and Sma L11 proteins are not shown in this figure because they are virtually identical to the Eco L11 protein (Sor and Nomura 1987).

Figure 38 shows that the two archaebacterial sequences can be aligned end to end with only one gap at alignment position 68 in the Hcu L11 protein. The Sso L11 protein has an extra amino acid at its N-terminus and 5 extra residues at its C-terminus. There is 40% sequence identity between the two proteins.

Although both archaebacterial proteins are longer that the Eco L11 protein, they can still be aligned end to end with the Eco L11 sequence (see Figure 38). The extra residues of the archaebacterial proteins are located at the C-terminus, with the Sso L11 protein having 31 and the Hcu L11 protein 26 extra amino acids compared to the Eco L11 protein. The Eco L11 protein, on the

Sso L11 Gene

1 42
TTG CCT ACT AAG ACA ATA AAA ATA ATG GTA GAA GGA GGT AGT
M P T K T I K I M V E G G S

43 84
GCT AAA CCA GGT CCA CCA TTG GGA CCT ACC CTA TCA CAA TTA
A K P G P P L G P T L S Q L

85 126
GGA CTT AAT GTT CAG GAA GTT GTT AAA AAA ATA AAT GAT GTT
G L N V Q E V V K K I N D V

127 168
ACA GCA CAA TTT AAA GGA ATG TCA GTA CCA GTC ACA ATT GAG
T A Q F K G M S V P V T I E

169 210
ATC GAT AGT TCA ACT AAA AAA TAT GAT ATA AAA GTT GGA GTA
I D S S T K K Y D I K V G V

211 252
CCT ACA ACA ACA TCA CTA TTA CTT AAA GCA ATA AAC GCT CAA
P T T T S L L L K A I N A Q

253 294
GAG CCC TCA GGA GAT CCA GCC CAC AAG AAG ATA GGC AAT TTA
E P S G D P A H K K I G N L

295 336
GAC TTA GAA CAA ATT GCT GAC ATA GCG ATT AAG AAA AAG CCT
D L E Q I A D I A I K K K P

337 378
CAA CTT AGC GCA AAA ACT CTA ACA GCT GCA ATA AAG AGT TTA
Q L S A K T L T A A I K S L

379 420
CTA GGA ACT GCA AGG TCA ATA GGT ATA ACT GTT GAA GGA AAA
L G T A R S I G I T V E G K

421 462
GAT CCT AAA GAT GTA ATA AAA GAG ATT GAT CAA GGA AAA TAC
D P K D V I K E I D Q G K Y

463 504
AAT GAC TTA TTA ACT AAT TAT GAG CAA AAA TGG AAT GAA GCA
N D L L T N Y E Q K W N E A

505 513
GAA GGG TGA
E G *

Figure 37. Sequence of the Sso L11 gene. The derived amino acid sequence of the protein is shown below the nucleotide sequence. The N-terminal sequence obtained from purified Sso L11 protein is underlined (Matheson 1985). The gene uses **TTG** as the initiation codon.

Table 19 Codon utilization in the Sso L11 gene

UUU	Phe	1	UCU	Ser	0	UAU	Tyr	2	UGU	Cys	0
UUC	Phe	0	UCC	Ser	0	UAC	Tyr	1	UGC	Cys	0
UUA	Leu	7	UCA	Ser	6	UAA	*	0	UGA	*	1
UUG	Leu	2	UCG	Ser	0	UAG	*	0	UGG	Trp	1
CUU	Leu	3	CCU	Pro	5	CAU	His	0	CGU	Arg	0
CUC	Leu	0	CCC	Pro	1	CAC	His	1	CGC	Arg	0
CUA	Leu	4	CCA	Pro	5	CAA	Gln	7	CGA	Arg	0
CUG	Leu	0	CCG	Pro	0	CAG	Gln	1	CGG	Arg	0
AUU	Ile	4	ACU	Thr	6	AAU	Asn	6	AGU	Ser	3
AUC	Ile	1	ACC	Thr	1	AAC	Asn	1	AGC	Ser	1
AUA	Ile	11	ACA	Thr	7	AAA	Lys	16	AGA	Arg	0
AUG	Met	2	ACG	Thr	0	AAG	Lys	6	AGG	Arg	1
GUU	Val	6	GCU	Ala	4	GAU	Asp	7	GGU	Gly	3
GUC	Val	1	GCC	Ala	1	GAC	Asp	3	GGC	Gly	1
GUA	Val	4	GCA	Ala	6	GAA	Glu	6	GGA	Gly	9
GUG	Val	0	GCG	Ala	1	GAG	Glu	4	GGG	Gly	1

* chain termination

Table 20 Predicted Amino acid composition of the Sso L11 Protein

AMINO ACID	NUMBER	MOLES %	AMINO ACID	NUMBER	MOLES %
A	12	7.0%	M	3	1.8%
C	-	-	N	7	4.1%
D	10	5.8%	P	11	6.4%
E	10	5.8%	Q	8	4.7%
F	1	0.6%	R	1	0.6%
G	14	8.2%	S	10	5.8%
H	1	0.6%	T	14	8.2%
I	16	9.4%	V	11	6.4%
K	22	12.9%	W	1	0.6%
L	15	8.8%	Y	3	1.8%

Molecular weight: 18, 220.22 pI: 9.83

```
                                10                  20
Sso L11         P T K T I K I M V E G G S A K P G P P L G P
Hcu L11           A E T I E V L V A G G Q A D P G P P L G P
Eco L11   A K K V Q A Y V K L Q V A A G M A N P S P P V G P

                  30                  40                  50
Sso L11   T L S Q L G L N V Q E V V K K I N D V T A Q F - K
Hcu L11   E L G P T P V D V Q A V V Q E I N D Q T E A F - D
Eco L11   A L G Q Q G V N I M E F C K A F N A K T D S I E K

                            60                  70
Sso L11   G M S V P V T I E I D S S T K K Y D I K V G V P T
Hcu L11   G T E V P V T I E Y E D D G S F S - I E V G V P P
Eco L11   G L P I P V V I T V Y A D R S F T - F V T K T P P

                  80                  90                  100
Sso L11   T T S L L L K A I N A Q E P S G D P A H K K I G N
Hcu L11   T A A L V K D E A G F D T G S G E P Q E N F V A D
Eco L11   A A V L L K K A A G I K S G S G K P N K D K V G K

                            110                 120
Sso L11   L D L E Q I A D I A I K K K P Q L S A K T L T A A
Hcu L11   L S I E Q L K T I A E Q K K P D L L A Y D A R N A
Eco L11   I S R A Q L Q E I A Q T K A A D M T G A D I E A M

                  130                 140                 150
Sso L11   I K S L L G T A R S I G I T V E G K D P K D V I K
Hcu L11   A K E V A G T C A S L G V T I E G E D A R T F N E
Eco L11   T R S I E G T A R S M G L V V E D

                            160                 170
Sso L11   E I D Q G K Y N D L L T N Y E Q K W N E A E G
Hcu L11   R V D D G D Y D D V L G D E L A A A
```

Figure 38. Sequence alignment of the Sso L11 protein and its archaebacterial and eubacterial counterparts. Identical residues between the Sso L11 protein and the other sequences are boxed. Sso: *Sulfolobus solfataricus*, Hcu: *Halobacterium cutirubrum* (Shimmin and Dennis 1989), Eco: *Escherichia coli* (Dognin and Wittmann-Liebold 1977).

other hand, has an N-terminal extension of 3 or 4 extra residues that is not present in the archaebacterial sequences (see Figure 38).

Two gaps had to be introduced in order to obtain the alignment of the Sso L11 protein to the Eco L11 protein shown in Figure 38. One is located at alignment position 49 in the Sso L11 sequence and the other at alignment position 68 in the Eco L11 protein. Table 21 shows the percent sequence identity between the Sso L11 protein and its archaebacterial and eubacterial counterparts.

Table 21 Sequence identity between the Sso L11 protein and its archaebacterial and eubacterial counterparts

protein [1]	length [2]	identities[3]	% identities
Hcu L11	163	65	40%
Eco L11	138	46	33%
Pvu L11	138	48	35%
Sma L11	138	47	34%

[1] References can be found in p.144 [2] Length of the region of comparison
[3] Only identical residues were considered

The archaebacterial and eubacterial L11 proteins are proline rich. Seven proline residues (alignment positions: 19, 21, 22, 24, 55, 74 and 93) have been conserved between the two kingdoms. In *E.coli*, these proline residues probably contribute to the elongated shape of the L11 protein which has an axial ratio of 6-5:1 (Giri *et al.* 1978). The conservation of these residues in the archaebacteria probably implies that the general shape of the L11 proteins from the two kingdoms is similar.

The L11 protein is the most heavily methylated ribosomal protein in *Escherichia coli* (Dognin and Wittmann-Liebold 1979) and in *Saccharomyces cerevisiae* (Sce L11) (Cannon *et al.* 1977). In *E.coli*, the residues that are modified have been identified. The α amino group of the N-terminal alanine as well as the ε amino groups of the lysines at positions 3 and 38 are trimethylated (Dognin and Wittmann-Liebold 1977). In the case of the Sce L11 protein, the identity and position of the residues that have been modified have not been determined. Although it is not known if the archaebacterial proteins are methylated, comparison of the sequences of the archaebacterial and Eco L11 proteins, shows that there has been very little if any conservation of the sites of methylation between these two kingdoms. The first two modified residues (alanine at position 1 and lysine at position 3) are located within the N-terminal extension that is unique to the Eco L11 protein. Furthermore, neither of the archaebacterial proteins has a modified N-terminal amino group (Matheson *et al.* 1984, Matheson 1985). The lysine at position 39 has been conserved between the Eco L11 and Sso L11 proteins but is not present in the Hcu L11 protein.

Comparison of the sequences of the L11 proteins from the eubacteria and the archaebacteria, revealed the presence of 3 regions of high sequence conservation between the two kingdoms: residues 12-27, 50-59 and 131-142 (see Figure 38). In these regions, glycine and proline residues have been conserved suggesting that the secondary structure of the molecule has also been conserved.

In *E.coli*, the N-terminal region of the L11 protein (residues 1-64) has been shown to be important for the interaction of the ribosome with release factor 1 (Tate *et al.* 1984). Since this domain includes the first two regions that are

highly conserved in the two kingdoms, it is possible that in the archaebacterial proteins this domain of the protein might play a similar role during the termination of translation.

The Eco L11 protein is also involved in the synthesis of ppGpp during the stringent response (Friesen *et al.* 1974, Parker *et al.* 1976). This response takes place when *E.coli* cells are subjected to amino acid starvation. Under these conditions, ppGpp is produced and the synthesis of tRNA and rRNA stops (Nomura *et al.* 1984). However, studies with the methanogens and halophiles have revealed, that at least these two archaebacterial groups lack the stringent response (Beauclerck *et al.* 1985, Chant and Dennis, personal communication). Thus, some of the functions of the L11 protein have not been conserved in the two kingdoms.

In *E.coli*, protein L11 binds to a conserved region of the 23S rRNA located between nucleotides 1052 and 1112 (Schmidt *et al.* 1981). Heterologous binding studies have revealed that the Eco L11 protein can interact with a similar region on the archaebacterial 23S rRNA and eukaryotic 28S rRNA (Beauclerck *et al.* 1985, El-Baradi *et al.* 1987). Sequencing of the rRNA fragments protected by the Eco L11 protein, has revealed that the general secondary structure of the binding site for this protein has been conserved in the three kingdoms (Beauclerck *et al.* 1985, El-Baradi *et al.* 1987). However, the rRNA binding site on the protein has yet to be determined. It is interesting to note, in this respect, that there is a region in the L11 proteins (residues 23 to 32 in Figure 38) that shows sequence similarity to a region in the L1 proteins (residues 137 to 146 in Figure 36), which are also rRNA binding proteins. The alignment of these regions is shown in Figure 39.

Sso L1 (137-146)	L	G	P	A	L	G	P	R	G	K
Hcu L1 (137-146)	L	G	Q	V	L	G	P	R	G	K
Eco L1 (137-146)	L	G	Q	V	L	G	P	K	G	L
Bst L1 (137-146)	L	G	R	I	I	G	P	K	G	L
Sso L11 (23-32)	L	G	P	T	L	S	Q	L	G	L
Hcu L11 (23-32)	L	G	P	E	L	G	P	T	P	V
Eco L11 (23-32)	V	G	P	A	L	G	Q	Q	G	V
consensus	L	G	P	X	L	G	P	B	G	hy

Figure 39. Conserved region in the L11 and L1 ribosomal proteins. The numbers in parenthesis refer to the position of these residues in each protein (see Figures 36 and 38). Identical residues are boxed. A consensus sequence is shown at the bottom. hy: hydrophobic residue (L,V, I), X: any amino acid, B: basic residue (R,K).

Glycine, proline and hydrophobic amino acids have been conserved in this region in the L11 and L1 proteins from the two kingdoms. Since the L11 and L1 proteins bind to rRNA, it is possible that this conserved region might be involved in the interaction of these proteins with rRNA.

The Sso L46 Gene and the Sso L46 Protein

The gene located at 100 nucleotides from the Eco RI site was identified as the Sso L46 gene on the basis of sequence similarity between the sequence of its product and the sequences of the Sce L46 ribosomal protein from *Saccharomyces cerevisiae* (Leer *et al.* 1985a) and Rno L39 ribosomal protein from rat liver (Lin *et al.* 1984). Furthermore, the product of this gene has been isolated from *Sulfolobus solfataricus* ribosomes and partially sequenced (Ramírez *et al.* 1989a). Since this protein shows no sequence similarity to any of the eubacterial ribosomal proteins, the number of this protein in yeast has been used to designate it.

There are two possible initiation codons, separated by only 3 nucleotides, for the Sso L46 gene (see Figure 40). Since sequencing of the Sso L46 protein revealed that the Sso L46 protein starts with a serine residue (Ramírez *et al.* 1989a), it seems likely that the second ATG codon is the real initiation codon. If the second ATG codon is the initiation codon, then the Sso L46 gene has 156 nucleotides and since the N-terminal methionine is removed post-translationally (Ramírez *et al.* 1989a), it codes for a protein of 50 amino acids (see Figure 40). Unlike the yeast Sce L46 gene, which is interrupted by a 385 nucleotide intron, located immediately after the second codon (Leer *et al.* 1985a), no intron is present in the Sso L46 gene. Table 22 shows the codon usage in this gene.

Sso L46 Gene

1													42
ATG	GAA	ATG	AGC	AAG	CAT	AAG	TCC	TTA	GGC	AAA	AAA	TTG	AGA
(M)	(E)	M	S	K	H	K	S	L	G	K	K	L	R

43													84
CTA	GGT	AAA	GCG	TTA	AAA	AGA	AAC	TCT	CCT	ATT	CCT	GCT	TGG
L	G	K	A	L	K	R	N	S	P	I	P	A	W

85													126
GTC	ATA	ATA	AAA	ACT	CAA	GCT	GAG	ATA	AGG	TTT	AAT	CCA	CTA
V	I	I	K	T	Q	A	E	I	R	F	N	P	L

127											162
AGA	AGA	AAT	TGG	AGA	AGA	AAT	AAT	TTA	AAG	GTG	TGA
R	R	N	W	R	R	N	N	L	K	V	*

Figure 40. Sequence of the Sso L46 gene. The derived amino acid sequence of the protein is shown below the nucleotide sequence. The sequence of the protein obtained from purified Sso L46 is underlined (Ramírez *et al.* 1989a).

Table 22. Codon utilization in the Sso L46 gene

UUU	Phe	1	UCU	Ser	1	UAU	Tyr	0	UGU	Cys	0
UUC	Phe	0	UCC	Ser	1	UAC	Tyr	0	UGC	Cys	0
UUA	Leu	3	UCA	Ser	0	UAA	*	0	UGA	*	1
UUG	Leu	1	UCG	Ser	0	UAG	*	0	UGG	Trp	2
CUU	Leu	0	CCU	Pro	2	CAU	His	1	CGU	Arg	0
CUC	Leu	0	CCC	Pro	0	CAC	His	0	CGC	Arg	0
CUA	Leu	2	CCA	Pro	1	CAA	Gln	1	CGA	Arg	0
CUG	Leu	0	CCG	Pro	0	CAG	Gln	0	CGG	Arg	0
AUU	Ile	1	ACU	Thr	1	AAU	Asn	4	AGU	Ser	0
AUC	Ile	0	ACC	Thr	0	AAC	Asn	1	AGC	Ser	1
AUA	Ile	3	ACA	Thr	0	AAA	Lys	5	AGA	Arg	6
AUG	Met	1	ACG	Thr	0	AAG	Lys	3	AGG	Arg	1
GUU	Val	0	GCU	Ala	2	GAU	Asp	0	GGU	Gly	1
GUC	Val	1	GCC	Ala	0	GAC	Asp	0	GGC	Gly	1
GUA	Val	0	GCA	Ala	0	GAA	Glu	0	GGA	Gly	0
GUG	Val	1	GCG	Ala	1	GAG	Glu	1	GGG	Gly	0

* chain termination

Table 23 Predicted Amino acid composition of the Sso L46 protein

AMINO ACID	NUMBER	MOLES %	AMINO ACID	NUMBER	MOLES %
A	3	5.8%	M	1	1.9%
C	-	-	N	5	9.6%
D	-	-	P	3	5.8%
E	1	1.9%	Q	1	1.9%
F	1	1.9%	R	7	13.5%
G	2	3.8%	S	3	5.8%
H	1	1.9%	T	1	1.9%
I	4	7.7%	V	2	3.8%
K	8	15.4%	W	2	3.8%
L	6	11.5%	Y	-	-

Molecular weight: 6, 064 pI: 12.9

The Sso L46 protein is very basic (pI: 12.9); containing 16 basic residues and only one acidic residue (see Table 23). The alignment of the Sso L46 protein with the Sce L46 and Rno L39 proteins is shown in Figure 41. There is 40% sequence identity between the Sso L46 and Sce L46 proteins and 46% sequence identity between the Sso L46 and Rno L39 proteins.

Two tryptophan residues (positions 25 and 43 in Figure 41) are present in this protein. This is unusual since the presence of tryptophan is very rare in ribosomal proteins (Wittmann-Liebold 1986). These residues are likely to be important for the function of the protein since they have been conserved in the two kingdoms.

As stated above, the Sso L46 protein shows no sequence similarity to any eubacterial ribosomal proteins, suggesting that this protein is absent in the eubacterial ribosome. Ribosomal proteins that show no sequence similarity to eubacterial proteins have also been isolated from other archaebacterial groups: S2, S15 and L16 from *Halobacterium marismortui* (Kimura *et al.* 1989), and the products of the *a, b, c, d* and *e* genes in the *spc* operon of *Methanococcus vannielii* as well as the products of genes *1* and *2* in the *str* operon of the same organism (Auer *et al.* 1989a, b). The significance of these proteins will be discussed in the section on the evolution of the ribosome.

Chan *et al.* (1987) have proposed that the Sce L46 and Rno L39 proteins are members of an extended family of ribosomal proteins that includes: Rno L5 from rat liver, Hsa L32 from humans and Mmu L32 from mouse. This family of proteins is characterized by the presence of short segments whose sequence has been conserved in all of its members. Although the role that these segments play in the function and/or structure of each one of these proteins is not known, their presence in such a variety of ribosomal proteins gives support

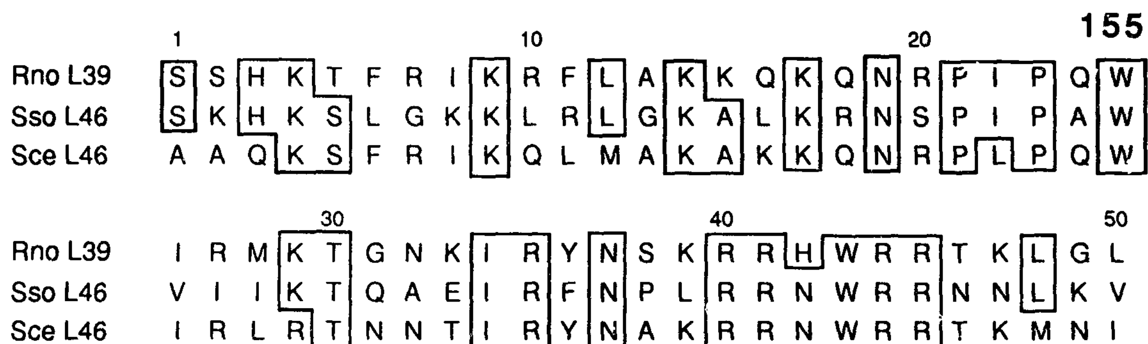


Figure 41. Sequence alignment of the Sso L46 protein with its eukaryotic counterparts. Identical residues to the Sso L46 protein are boxed. Sso: *Sulfolobus solfataricus*, Rno: *Rattus norvegicus* (Lin *et al.* 1984), Sce: *Saccharomyces cerevisiae* (Leer *et al.* 1985a)

Figure 42. Common regions in the L5 family of ribosomal proteins. The alignment of the two regions that show sequence similarity in all the members of this family of proteins is shown. Numbers in parenthesis refer to the position of the initial residue in the segment in the respective protein. Hsa: *Homo sapiens*, Mmu: *Mus musculus*, Rno: *Rattus norvegicus*, Sce: *Saccharomyces cerevisiae*, Sso: *Sulfolobus solfataricus*. Alignment drawn after Chan *et al.* 1987.

to the idea that ribosomal proteins evolved by the repeated duplication of a small number of ancestral peptides (Jue *et al.* 1980, Lin *et al.* 1987). The alignment of the regions that the members of this family have in common is shown in Figure 42. As can be seen in this figure, the Sso L46 protein is also a member of this family of ribosomal proteins.

The Sso LX Gene and its Product

The Sso LX gene was identified as a ribosomal protein gene because the sequence of its product, matches the sequence of a small basic ribosomal protein isolated from the large subunit of *Sulfolobus solfataricus* ribosomes (Louie, K.A., personal communication).

Figure 43 shows the sequence of this gene and its product. The regions that have been confirmed by amino acid sequencing of the protein are underlined. The gene has 216 nucleotides and codes for a protein of 71 amino acids. Table 24 shows the codon usage in this gene.

The protein encoded by this gene is very basic (pI: 12.02). It contains 16 basic residues and only four acidic residues (See Table 25). A search through the data base, revealed that this protein shows no sequence similarity to any of the eubacterial and eukaryotic ribosomal proteins sequenced up to now. Since all the proteins in an eukaryotic ribosome have not been sequenced, it is possible that an equivalent protein may be present in eukaryotic ribosomes.

The presence of proteins that show no sequence similarity to eubacterial or eukaryotic ribosomal proteins has also been detected in *Halobacterium marismortui* (proteins S6, S13, L29 and L31) (Kimura *et al.* 1989). Although it is probable that ribosomal proteins equivalent to these proteins will be found in

the eukaryotic ribosome, the possibility still remains that there might be ribosomal proteins that are unique to the archaebacteria.

Sso LX Gene

1													42
ATG	GCT	GAA	GTA	AAA	ATT	TTC	ATG	GTC	AGA	GGA	ACT	GCC	ATA
M	A	E	V	K	I	F	M	V	R	G	T	A	I

43													84
TTT	AGT	GCG	TCA	AGA	TTT	CCT	ACA	AGT	CAA	AAA	TAT	GTT	AGA
F	S	A	S	R	F	P	T	S	Q	K	Y	V	R

85													126
GCT	TTA	AAT	GAA	AAA	CAA	GCA	ATC	GAA	TAC	ATT	TAT	AGT	CAA
A	L	N	E	K	Q	A	I	E	Y	I	Y	S	Q

127													168
CTT	GGT	GGA	AAA	AAT	AAA	ATT	AAC	GAT	ACA	ACA	TAC	ACA	TAC
L	G	G	K	N	K	I	N	D	T	T	Y	T	Y

169													210
AAG	AGA	TCA	AAG	AAG	TTA	AGG	AAG	ATG	AAA	TCA	CAG	ACA	AGA
K	R	S	K	K	L	R	K	M	K	S	Q	T	R

211	216
CAA	TAA
Q	*

Figure 43. Sequence of the Sso LX gene. The derived amino acid sequence of the protein is shown below the nucleotide sequence. The residues that have been confirmed by amino acid sequencing of the protein are underlined (Louie, K.A., personal communication).

The Sso *ala S* Gene

The gene located downstream of the Sso L12 gene was identified as the Sso *ala S* gene on the basis of sequence similarity between its product and the alanine-tRNA synthetase from *Escherichia coli* (Putney *et al.* 1981a). Only the first 546 nucleotides of this gene are present in the 6.9 Kb Eco RI - Bam HI fragment. The sequence of this part of the gene, as well as the derived amino

Table 24 Codon utilization in the Sso LX gene

UUU	Phe	2	UCU	Ser	0	UAU	Tyr	2	UGU	Cys	0
UUC	Phe	1	UCC	Ser	0	UAC	Tyr	3	UGC	Cys	0
UUA	Leu	2	UCA	Ser	3	UAA	*	1	UGA	*	0
UUG	Leu	0	UCG	Ser	0	UAG	*	0	UGG	Trp	0
CUU	Leu	1	CCU	Pro	1	CAU	His	0	CGU	Arg	0
CUC	Leu	0	CCC	Pro	0	CAC	His	0	CGC	Arg	0
CUA	Leu	0	CCA	Pro	0	CAA	Gln	4	CGA	Arg	0
CUG	Leu	0	CCG	Pro	0	CAG	Gln	1	CGG	Arg	0
AUU	Ile	3	ACU	Thr	1	AAU	Asn	2	AGU	Ser	3
AUC	Ile	1	ACC	Thr	0	AAC	Asn	1	AGC	Ser	0
AUA	Ile	1	ACA	Thr	5	AAA	Lys	6	AGA	Arg	5
AUG	Met	3	ACG	Thr	0	AAG	Lys	4	AGG	Arg	1
GUU	Val	1	GCU	Ala	2	GAU	Asp	1	GGU	Gly	1
GUC	Val	1	GCC	Ala	1	GAC	Asp	0	GGC	Gly	0
GUA	Val	1	GCA	Ala	1	GAA	Glu	3	GGA	Gly	2
GUG	Val	0	GCG	Ala	1	GAG	Glu	0	GGG	Gly	0

* chain termination

Table 25 Predicted Amino acid composition of the Sso LX protein

AMINO ACID	NUMBER	MOLES %	AMINO ACID	NUMBER	MOLES %
A	5	6.9%	M	3	4.2%
C	-	-	N	3	4.2%
D	1	1.4%	P	1	1.4%
E	3	4.2%	Q	5	6.9%
F	3	4.2%	R	6	8.3%
G	3	4.2%	S	6	8.3%
H	-	-	T	6	8.3%
I	5	6.9%	V	3	4.2%
K	10	13.9%	W	-	-
L	3	4.2%	Y	5	6.9%

Molecular weight: 8, 329.65 pI: 12.02

acid sequence of the N-terminal region of the *Sulfolobus* alanine-tRNA synthetase are shown in Figure 44. The codon usage in this part of the gene is shown in Table 26.

Aminoacyl-tRNA synthetases play a key role in protein biosynthesis, since they catalyze the aminoacylation of each tRNA with its proper amino acid, thus ensuring the fidelity of protein translation. The reaction they catalyze consists of the esterification of the amino acid residue to the hydroxyl group at the 3' end of its cognate tRNA. This reaction is accomplished through the synthesis of an aminoacyl adenylate (condensation of an amino acid with ATP), followed by the reaction of the enzyme bound adenylate with tRNA (for a review, see Schimmel 1987).

The regions of the alanine-tRNA synthetase from *E.coli* involved in the different functions of the enzyme have been determined by introducing gene deletions (Jasin *et al.* 1983). The adenylate synthesis activity is located in the N-terminal fragment comprising residues 1-385. Extension of this fragment to residue 461 gives rise to aminoacylation activity. Finally, sequences located between amino acids 699 to 808 are important for oligomerization (the *E. coli* enzyme is a tetramer of identical subunits (α_4) (Putney *et al.* 1981b)) while the C-terminal region (residues 809 to 875) does not appear to be important for the catalytic activity of the enzyme (see Figure 45) (Jasin *et al.* 1983).

The alignment of the N-terminal region of the Sso alanine-tRNA synthetase with the corresponding region in the Eco alanine-tRNA synthetase is shown in Figure 46. This alignment reveals that the *Sulfolobus* enzyme has an N-terminal extension of 58 amino acids with respect to the *E.coli* protein. This is similar to the situation found between certain aminoacyl-tRNA synthetases of yeast and *E.coli*, where the yeast enzymes also have an N-terminal extension

1 42
ATG GTC AAA GCA AAC GAG AAC GAG TAC AGG TTA AAC TTT TTT
M V K A N E N E Y R L N F F

43 84
CTG TCA AGA GGA TAT GAA AGA AAA ATT TGT AGA TCC TGT TCT
L S R G Y E R K I C R S C S

85 126
ACC CCA TTT TGG ACT TTA GAT AAG TCA AAG GAA AAT TGT TCA
T P F W T L D K S K E N C S

127 168
GAT GTT CCT TGT ACT GAC TAT TAC TTC TTT GAT ATC AAG ATT
D V P C T D Y Y F F D I K I

169 210
AAA TCT CCC CCG TTA ACA GTG AGA GAA TCA AGA GAA AAA TTT
K S P P L T V R E S R E K F

211 252
CTA AGA TTC TTT GAA AAG AGA GGA CAC ACC ATA GTT CCA CCT
L R F F E K R G H T I V P P

253 294
AAA CCT GTA GTT GCC AAG TGG AGA GAT GAT CTT TAT CTT ACT
K P V V A K W R D D L Y L T

295 336
ATA GCT AGT ATA GTT GAC TTT CAA CCA TTT GTA ACG AGC GGA
I A S I V D F Q P F V T S G

337 378
TTA GCT AAA CCG CCT GCA AAT CCA CTG GTA GTT TCT CAG CCA
L A K P P A N P L V V S Q P

379 420
TGT ATA AGA CTA GAA GAT GTG GAT AAC GTA GGA ATA ACA TTT
C I R L E D V D N V G I T F

421 462
GGA AGG CAT TTG ACC ACA TTT GAA ATG GCT GCT CAT CAT GCA
G R H L T T F E M A A H H A

463 504
TTC AAT TAT CCT GAT AAA CAA ATT TAT TGG AAA GAC GAA ACA
F N Y P D K Q I Y W K D E T

505 546
GTA GCA TTT GCT AAG GAG TTC TTT ACA GAA GAA CTA GGG ATC
V A F A K E F F T E E L G I

Figure 44. Sequence of the *Sulfolobus ala S* gene. Only the first 546 nucleotides of this gene are present in the 6.9 Kb Eco RI-Bam HI fragment. The derived amino acid sequence of the protein is shown below the nucleotide sequence.

Table 26 Codon utilization in the 5' segment (first 546 nucleotides) of the *Sulfolobus ala S* gene

UUU	Phe	12	UCU	Ser	3	UAU	Tyr	5	UGU	Cys	5
UUC	Phe	4	UCC	Ser	1	UAC	Tyr	2	UGC	Cys	0
UUA	Leu	4	UCA	Ser	4	UAA	*	0	UGA	*	0
UUG	Leu	1	UCG	Ser	0	UAG	*	0	UGG	Trp	3
CUU	Leu	2	CCU	Pro	5	CAU	His	3	CGU	Arg	0
CUC	Leu	0	CCC	Pro	1	CAC	His	1	CGC	Arg	0
CUA	Leu	3	CCA	Pro	5	CAA	Gln	2	CGA	Arg	0
CUG	Leu	2	CCG	Pro	2	CAG	Gln	1	CGG	Arg	0
AUU	Ile	3	ACU	Thr	3	AAU	Asn	3	AGU	Ser	1
AUC	Ile	2	ACC	Thr	3	AAC	Asn	4	AGC	Ser	1
AUA	Ile	5	ACA	Thr	5	AAA	Lys	8	AGA	Arg	9
AUG	Met	2	ACG	Thr	1	AAG	Lys	6	AGG	Arg	2
GUU	Val	5	GCU	Ala	5	GAU	Asp	8	GGU	Gly	0
GUC	Val	1	GCC	Ala	1	GAC	Asp	3	GGC	Gly	0
GUA	Val	5	GCA	Ala	4	GAA	Glu	10	GGA	Gly	5
GUG	Val	2	GCG	Ala	0	GAG	Glu	3	GGG	Gly	1

* chain termination

(Mirande and Waller 1988). There are 8 aromatic residues (positions 70, 73, 74, 105, 151, 159, 167, 175 in Figure 46) that have been conserved between the two kingdoms and might have an important role in the structure and/or function of the protein. There is 34% sequence identity between the Sso and the Eco aminoacyl-tRNA synthetases over the region of comparison. Since this region is involved in adenylate synthesis in the Eco alanine-tRNA synthetase (Jasin *et al.* 1983), it is probable that this region in the Sso protein has a similar function.

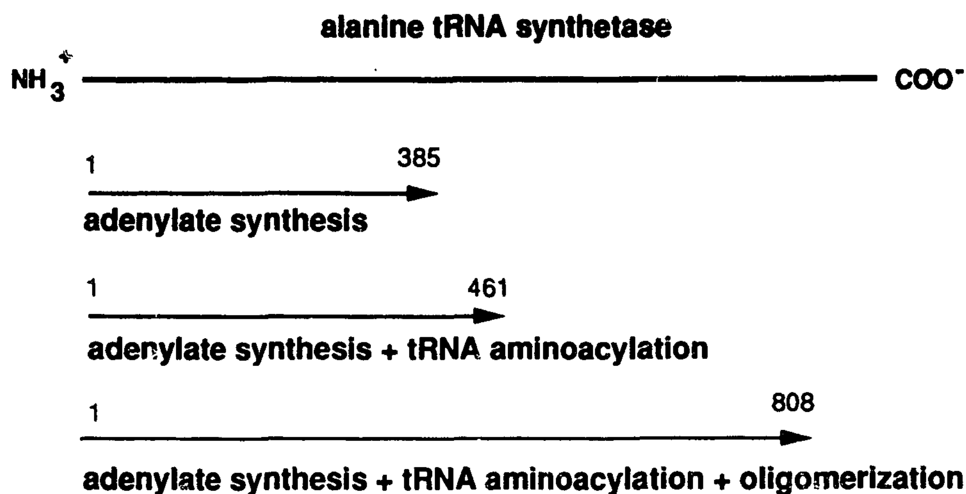


Figure 45. Functional domains in the *Escherichia coli* alanine tRNA synthetase. The protein contains 875 amino acids. Redrawn from Jasin *et al.* 1983.

Like the *E.coli* enzyme, the *Sulfolobus* enzyme lacks the sequence: **HIGH** in the N-terminal region, that is characteristic of other aminoacyl-tRNA synthetases (Webster *et al.* 1984).

As mentioned before, the alanine-tRNA synthetase in *E.coli*, has a tetrameric structure (Putney *et al.* 1981b), while in the eukaryotes, the alanine-tRNA synthetase is a monomer (Dignam and Dignam 1984). The quaternary structure of the *Sulfolobus* enzyme is unknown.

The Sso "Docking" Protein Gene and Its Product

The gene located between open reading frames 5 and 6 was identified as the probable equivalent in *Sulfolobus* of the α subunit of the signal recognition particle (SRP) receptor or docking protein, on the basis that the product of this

```
                                 10                                      20
Sso alaS   M  V  K  A  N  E  N  E  Y  R  L  N  F  F  L  S  R  G  Y  E  R  K  I  C  R
Eco alaS

                     30                                      40                     50
Sso alaS   S  C  S  T  P  F  W  T  L  D  K  S  K  E  N  C  S  D  V  P  C  T  D  Y  Y
Eco alaS

                                 60                                      70
Sso alaS   F  F  D  I  K  I  K  S  P  P  L  T  V  R  E  S  R  E  K  F  L  R  F  F  E
Eco alaS                           M  S  K  S  T  A  E  I  R  Q  A  F  L  D  F  F  H

                     80                                      90                     100
Sso alaS   K  R  G  H  T  I  V  P  P  K  P  V  V  A  K  W  R  D  D  L  Y  L  T  I  A
Eco alaS   S  K  G  H  Q  V  V  A  S  S  S  R  Y  P  I  T  D  P  T  L  L  F  T  N  A

                                 110                                     120
Sso alaS   S  I  V  D  F  Q  P  F  V  T  S  G  L  A  K  P  P  A  N  P  L  V  V  S  Q
Eco alaS   G  M  N  Q  F  K  D  -  V  F  L  G  L  D  K  R  N  Y  S  R  A  T  T  S  Q

                     130                                     140                    150
Sso alaS   P  C  I  R  -  -  -  -  L  E  D  V  D  N  V  G  I  T  F  G  R  H  L  T  T
Eco alaS   R  C  V  R  A  G  G  K  H  N  D  L  E  N  V  G  Y  T  -  A  R  H  H  T  F

                                 160                                     170
Sso alaS   F  E  M  A  A  H  H  A  F  N  Y  r  D  K  Q  I  Y  W  K  D  E  T  V  A  F
Eco alaS   F  E  M  L  G  -  -  N  F  S  F  G  D  -  -  -  Y  F  K  H  D  A  I  Q  F

                     180
Sso alaS   A  K  E  F  F  T  E  E  L  G  I  S  •  •  •
Eco alaS   A  W  E  L  L  T  S  E  K  W  F  A  •  •  •
```

Figure 46. Sequence alignment of the N-terminal region of the *Sulfolobus* alanine tRNA synthetase with the corresponding region from the *Escherichia coli* alanine tRNA synthetase. Identical residues are boxed. Sso: *Sulfolobus solfataricus*, Eco: *Escherichia coli*, (Putney *et al.* 1981a) *ala S:* alanine tRNA synthetase.

gene shows sequence similarity to the α subunit of the docking protein from dogs (Lauffer *et al.* 1985) and humans (Hortsch *et al.* 1988), as well as the product of the *fts Y* gene in *Escherichia coli* (Gill *et al.* 1986).

The gene has 1110 nucleotides and codes for a protein of 369 amino acids (see Figure 47). Table 27 shows the codon usage in this gene (for a discussion, see section on codon utilization) and Table 28 shows the amino acid composition of the protein.

In the eukaryotes, secretory, membrane and lysosomal proteins are targeted to their respective locations through a transport pathway that involves the recognition of a signal sequence encoded in the nascent protein by a cytoplasmic particle called the signal recognition particle (SRP) (for reviews, see Walter *et al.* 1984, Verner and Schatz 1988). The signal recognition particle is a ribonucleoprotein complex formed by 6 polypeptides (72, 68, 54, 19, 14 and 9 kDa) and a 300 nucleotide RNA molecule (the 7S RNA) (Siegel and Walter 1988). As soon as the signal sequence, which consists of a basic N-terminal region followed by an uninterrupted stretch of at least 7 or 8 apolar, largely hydrophobic residues (Verner and Schatz 1988), emerges from the ribosome, it is recognized by the 54 kDa subunit of the SRP (Kurzchalia *et al.* 1986). Protein synthesis is arrested and the SRP-ribosome complex moves towards the endoplasmic reticulum (Walter and Blobel 1981). The arrest of translation is due to the 9 and 14 kDa polypeptides and the Alu-like domain of the 7S RNA (100 nucleotides at the 5' end and 45 nucleotides at the 3' end of the 7S RNA that are homologous to the human Alu right monomer sequence) (Siegel and Walter 1988). Once the SRP-ribosome complex reaches the endoplasmic reticulum, it interacts with the SRP receptor or docking protein. As a consequence of this interaction, translational arrest is unblocked, the SRP

```
1                                                        42
ATA ATT TGT TTT GAT AGA TTA AAA AAA GCA TTC TCA AAT TTT
 M   I   C   F   D   R   L   K   K   A   F   S   N   F
43                                                       84
TTA GAT AAA ATT AGT GGA GAA GAG AAC AAA AAG GAA CCT GAA
 L   D   K   I   S   G   E   E   N   K   K   E   P   E
85                                                      126
ACT AGA CAA ACA GAT CAA CTA GAA AGT AAA AAA GAA GAG ACT
 T   R   Q   T   D   Q   L   E   S   K   K   E   E   T
127                                                     168
ATA CAA CAA CAG CAA AAT GTA CAA CAA CCA CAA GCG GAA AAT
 I   Q   Q   Q   Q   N   V   Q   Q   P   Q   A   E   N
169                                                     210
AAA ATT GAG CAA AAA CAG GAA AAA ATT TCA GTT CAA ACA GGT
 K   I   E   Q   K   Q   E   K   I   S   V   Q   T   G
211                                                     252
CAA GAA AAT AAG CAG GAG AAT AAA CGA TCC TTC TTT GAT TTC
 Q   E   N   K   Q   E   N   K   R   S   F   F   D   F
253                                                     294
CTG AAC TAT AAA ACG ATA AAA GAG GAC GAT CTT AAT GAC GTT
 L   K   Y   K   T   I   K   E   D   D   L   N   D   V
295                                                     336
ATA GAA GAA CTC AGA TTT CAA CTT CTT GAT TCT GAT GTA TCT
 I   E   E   L   R   F   Q   L   L   D   S   D   V   S
337                                                     378
TAT GAA GTT ACG GAA AAA ATA CTA GAA GAT CTC AAA AAT AAT
 Y   E   V   T   E   K   I   L   E   D   L   K   N   N
379                                                     420
CTG ATA GGG AAA AAG GTT AGT AGA AGA GAG GAA GTT GAG GAA
 L   I   G   K   K   V   S   R   R   E   E   V   E   E
421                                                     462
ATA GTA ATA AAT ACT CTA AAA AAA TCT ATA ACA GAG ATT TTA
 I   V   I   N   T   L   K   K   S   I   T   E   I   L
463                                                     504
ACA AAA AAT CAA AAA ACA GAT TTG ATA GAA AAA ATA AGA AGT
 T   K   N   Q   K   T   D   L   I   E   K   I   R   S
505                                                     546
AGT GGA AAG AAG CCA TTT GTA ATA ATA TTC TTT GGA GTT AAT
 S   G   K   K   P   F   V   I   I   F   F   G   V   N
```

Figure 47. Sequence of the gene coding for the equivalent in *Sulfolobus* of the α subunit of the docking protein. The derived amino acid sequence of the protein is shown below the nucleotide sequence. The gene starts with an unusual initiation codon (**ATA**) (for a discussion, see section on codon utilization).

547 ... 588
GGA GTA GGT AAA ACT ACT ACT ATA GCA AAA GTA GTA AAT ATG
G V G K T T T I A K V V N M

589 ... 630
CTG AAG AAA AAT AAC TTA TCT ACA ATT ATT GCA GCC TCA GAT
L K K N N L S T I I A A S D

631 ... 672
ACA TTT AGA GCT GCT GCA CAA GAG CAA CTA GCT TAT CAT GCA
T F R A A A Q E Q L A Y H A

673 ... 714
TCA AAA TTA GAG GTT CAA CTA ATA AGA GGA AAA TAT GGT GCT
S K L E V Q L I R G K Y G A

715 ... 756
GAT CCA GCG TCG GTA GCA TTC GAT GCC ATA AGT TTT GCT AAG
D P A S V A F D A I S F A K

757 ... 798
AGT AGA AAT ATA GAT GTA GTA CTT ATT GAT ACT GCT GGG AGA
S R N I D V V L I D T A G R

799 ... 840
ATG CAT ATA GAC TCT GAT TTA GTT GAA GAA CTT AAG AGG GTT
M H I D S D L V E E L K R V

841 ... 882
TTA AGG ATA GCA AAA CCT GAT TTT AGG ATC TTA ATA TTA GAC
L R I A K P D F R I L I L D

883 ... 924
TCG CTA GCA GGA AGT GAT GCA TTA GAA CAG GCT AGA CAT TTT
S L A G S D A L E Q A R H F

925 ... 966
GAA AAT AAT GTT GGA TAT GAT GCA GTA ATT CTT ACC AAA GTA
E N N V G Y D A V I L T K V

967 ... 1008
GAT GCA GAT GCT AAA GGT GGA ATA GCT CTT TCA TTA GCG TAT
D A D A K G G I A L S L A Y

1009 ... 1050
GAA TTG AAA AAA CCT GTA GTA TAT ATG GGA GTA GGA CAG AAT
E L K K P V V Y M G V G Q N

1051 ... 1092
TAT GAT GAT TTA ATT CCC TTC TCA CCA GAC TGG TTT GTA GAA
Y D D L I P F S P D W F V E

1093 ... 1110
AGG ATA TTC AGT AGT TAA
R I F S S '

Figure 47...Continued

Table 27 Codon utilization in the Sso "docking" gene

UUU	Phe	11	UCU	Ser	5	UAU	Tyr	8	UGU	Cys	1
UUC	Phe	7	UCC	Ser	1	UAC	Tyr	0	UGC	Cys	0
UUA	Leu	11	UCA	Ser	6	UAA	*	1	UGA	*	0
UUG	Leu	2	UCG	Ser	2	UAG	*	0	UGG	Trp	1
CUU	Leu	7	CCU	Pro	3	CAU	His	3	CGU	Arg	0
CUC	Leu	2	CCC	Pro	1	CAC	His	0	CGC	Arg	0
CUA	Leu	7	CCA	Pro	4	CAA	Gln	16	CGA	Arg	1
CUG	Leu	3	CCG	Pro	0	CAG	Gln	5	CGG	Arg	0
AUU	Ile	10	ACU	Thr	7	AAU	Asn	17	AGU	Ser	10
AUC	Ile	1	ACC	Thr	1	AAC	Asn	2	AGC	Ser	0
AUA	Ile	22	ACA	Thr	7	AAA	Lys	30	AGA	Arg	11
AUG	Met	3	ACG	Thr	2	AAG	Lys	9	AGG	Arg	4
GUU	Val	10	GCU	Ala	9	GAU	Asp	22	GGU	Gly	4
GUC	Val	0	GCC	Ala	2	GAC	Asp	5	GGC	Gly	0
GUA	Val	16	GCA	Ala	11	GAA	Glu	23	GGA	Gly	10
GUG	Val	0	GCG	Ala	3	GAG	Glu	9	GGG	Gly	2

* chain termination

Table 28 Predicted Amino acid composition of the Sso "docking" protein

AMINO ACID	NUMBER	MOLES %	AMINO ACID	NUMBER	MOLES %
A	25	6.8%	M	4	1.1%
C	1	0.3%	N	19	5.1%
D	27	7.3%	P	8	2.2%
E	32	8.7%	Q	21	5.7%
F	18	4.9%	R	16	4.3%
G	16	4.3%	S	24	6.5%
H	3	0.8%	T	17	4.6%
I	32	8.7%	V	26	7.0%
K	39	10.6%	W	1	0.3%
L	32	8.7%	Y	8	2.2%

Molecular weight: 41, 893.62 pI: 5.56

is displaced from the ribosome and co-translational transport of the polypeptide across the membrane is initiated (Gilmore *et al.* 1982 a, b, Walter *et al.* 1984). GTP is required for this process (Connolly and Gilmore 1986, Wilson *et al.* 1988).

The SRP receptor is composed of two subunits designated α and β (Tajima *et al.* 1986). The α subunits of the SRP receptors from dogs (Lauffers *et al.* 1985) and humans (Hortsch *et al.* 1988) have been sequenced. It is believed that this subunit is responsible for the interaction with SRP and the release of translational arrest (Gilmore *et al.* 1982 a). The sequence of the β subunit has not been determined and its function is not known (Tajima *et al.* 1986).

Recently, Römisch *et al.* (1989) and Bernstein *et al.* (1989) have found that the eukaryotic docking protein shows sequence similarity to the product of the *fts Y* gene in *Escherichia coli.* Although the function of this protein in *Escherichia coli* has not been determined, it is known that the gene is part of an operon that includes genes involved in cell division (Gill *et al.* 1986).

The "docking" protein from *Sulfolobus* is smaller (369 amino acids) than either the eukaryotic docking proteins (638 amino acids) (Lauffer *et al.* 1985, Hortsch *et al.* 1988) or the product of the *fts Y* gene in *Escherichia coli* (497 amino acids) (Gill *et al.* 1986). The alignment of the proteins from the three kingdoms is shown in Figure 48. The canine sequence is not included in this figure because it is very similar to the human sequence. The region of sequence similarity among the three proteins comprises residues 69-369 of the *Sulfolobus* protein, residues 303-638 of the eukaryotic proteins, and residues 183-497 of the fts Y protein. The N-terminal region of the three proteins shows no sequence similarity among the three kingdoms, although in the three proteins, this region of the molecule is highly charged (Lauffers *et al.*

```
                                    10                                      20                                      30                                      40
Sso (69)     T G Q E N - K Q E N - K R S F F D F L K - - - - - - - - - - - - - Y K T I - - K
Hsa (303)    A A Q T L T K P S A T K G T L G G M F G M L - - - - - K G L - V G S K S L - - S
fts Y (183)  I A Q E Q E K P - - T K F G F F A R L K R S L L K T K E N L G S G F I S L F R G
54 SRP (1)                               M V L A D L G R K I T S A L R S L - S N - A T I I - N
48 Eco (1)                                 M F D N L T D R L S R T L R N I - S G R G R L - - T

                                    50                                      60                                      70                                      80
Sso          E D D L N D V I E E L R F Q L L D S D V S Y E V T E K I L E D L K N N L I G K K
Hsa          R E D M E S V L D K M R D H L I A K N V A A D I A V Q L C E S V A N K L E G K V
fts Y        K K I D D D L F E E L E E Q L L I A D V G V E T T R K I I T N L T E G A S R K Q
54 SRP       E E V L N A M L K E V C T A L L E A E F N I K L V K Q L R E N V K S A I D L E E
48 Eco       E D N V K D T L R E V R M A L L E A D V A L P V V R E F I N R V K E K A V G H E

                                    90                                     100                                     110                                     120
Sso          V S R R E G V E E I V I N T L K K S I T E I L - T K N Q K T D L I E K R S S - -
Hsa          M G T F S T V T S T V K Q A L Q E S L V Q I L Q P Q R [illegible] V D M L R D I M D A Q R
fts Y        L R D A E - - - - - - - - A L Y G L L K E E M G E I L A K V D E P L N V E - - -
54 SRP       M A S G L N K R K M I Q H A V F K E L V K L V D P G V K A W T P T - - - - - - -
48 Eco       V N K S L T P G Q E F V K I V R N E L V A A M G E E N Q T L N L A - - - - - - -

                                   130                                     140                                     150                                     160
Sso          G K K P F V I I F F G V N G V G K T T T I A K V V N M L K K N N L - S T I I A A
Hsa          R Q R P Y V V T F C G V N G V G K S T N L A K I S F W L L E N G F - S V L I A A
ftsY         G K A P F V I L M V G V N G V G K T T T I G K L A R Q F E Q Q G - K S V M L A A
54 SRP       K G K Q N V I M F V G L Q G S G K T T T C S K L A Y Y Y Q R K G W K T C L I - C
48 Eco       A Q P P A V V L M A G L Q G A G K T T S V G K L G K F L R E K H K K K V L V V S
```

Figure 48. Sequence alignment of the Sso "docking" protein and its eukaryotic and eubacterial counterparts, as well as the members of the "54 kDa family" of proteins. Identical residues between the Sso "docking" protein and the other proteins are boxed. Sso: *Sulfolobus solfataricus* "docking" protein, Hsa: *Homo sapiens* docking protein (Hortsch *et al.* 1988), fts Y: product of the *fts Y* gene from *E. coli* (Gill *et al.* 1986), 54 SRP: SRP 54 kDa protein (Römisch *et al.* 1989), 48 Eco: 48 kDa protein from *E.coli* (Byström *et al.* 1983). The three consensus elements of the GTP binding site are indicated in bold letters (Römisch *et al.* 1989). Numbers in parenthesis indicate the first residue in each protein. Dots at the end of the 54 SRP and 48 Eco sequences indicate that the sequence of the protein continues.

```
                                            170                          180                          190                          200
Sso      S D T F R A A A Q E Q L - - - - - - - - - - - - - - A Y H A S K L L V Q L I R G K
Hsa      C D T F R A G A V E Q L R T H T R R L S A L H P P E K H G G R T M V Q L F E K G
fts Y    G D T F R A A A V E Q L Q V - - - - - - - - - - - - - W G Q R N N I P V I A Q H
54 SRP   A D T F R A G A F D Q L K Q N A T K A R - - - - - - - - - - - - - - I P F Y G S Y
48 Eco   A D V Y R P A A I K Q L E Q L A E Q V G - - - - - - - - - - - - - - V D F F P S D

                                            210                          220                  II      230                          240
Sso      Y G A D P A S V A F D A I S F A K S R N I D V V L I D T A G R M H I D S D L V E
Hsa      Y G K D A A G I A M E A I A F A R N Q G F D V V L V D T A G R M Q D N A P L M T
fts Y    T G A D S A S V I F D A I Q A A K S R N I D V L I A D T A G R L Q N K S H L M E
54 SRP   T E M D P V I I A S E G V E K F K N E N F E I I I V D T S G R H K Q E D S L F E
48 Eco   V G Q K P V D I A N A A L K E A K L K F Y D V L L V D T S G R L H V D E A M M D

                                            250                          260                          270                          280
Sso      E L K R V L R I A K - - - - - - - P D F R I L I L D S L A G S D A L E Q A R H F E
Hsa      A L A K L I - - - - - - - T V N T P D L V L F V G E A L V G N E A V D Q L V K F N
fts Y    E L K K I V R V M K K L D V E A P H E V M L T I D A S T G Q N A V S Q A K L F H
54 SRP   E M L Q V A - - - - - - - N A I Q P D N I V Y V M D A S I G Q A C E A Q A K A F K
48 Eco   E I K Q V H - - - - - - - A S I N P V E T L F V V D A M T G Q D A A N T A K A F N

                                            290                          300      III                 310                          320
Sso      N N V G - - - - - - - - - - Y D A V I L T K V D A - D A K G G I A L S L A Y E L
Hsa      R A L A D H S M A Q T P R L I D G I V L T K F D T I D D K V G A A I S M T Y I T
fts Y    E A V G - - - - - - - - - - L T G I T L T K L D G - T A K G G V I F S V A D Q F
54 SRP   D K V D - - - - - - - - - - V A S V I M T K L D G - H A K G G G A L S A V A A T
48 Eco   E A L P - - - - - - - - - - L T G V V L T K V D G - D A R G G A A L S I R H I T

                                            330                          340                          350
Sso      K K P V V Y M G V G Q N Y D D L I P F S P D W F V E R I F S S
Hsa      S K P I V F V G T G Q T Y C D L R S L N A K A V V A A L M K A
fts y    G L P I R Y L G V G E R I E D L R P F K A D D F I E A L F A R E D
54 SRP   K S P I I F I G T G E H I D D F E P F K T Q P F I S K L L G M G D • • •
48 Eco   G K P I K F L G V G E K T E A L E P F H P D R I A S R I L G M G D • • •
```

1985, Gill *et al.* 1986, Hortsch *et al.* 1988). Table 29 summarizes the percent sequence identity between the *Sulfolobus* "docking" protein and its eukaryotic and eubacterial counterparts (referred to from here on as the "docking protein family").

Table 29 Sequence identity between the Sso "docking" protein and its eukaryotic and eubacterial counterparts

protein[1]	length[2]	identities[3]	% identities
Hsa docking	300	104	35%
Cfa docking	300	104	35%
fts Y protein	300	122	40.6%

[1] References can be found in p 164, Hsa: *Homo sapiens*, Cfa: *Canis familiaris*
[2] Length of the region of comparison
[3] Only identical residues were considered

Römisch *et al* (1989) and Bernstein *et al.* (1989) have also found that the eukaryotic docking proteins and the product of the *fts Y* gene in *Escherichia coli* share a common domain with the 54 kDa protein of the signal recognition particle and with a 48 kDa protein of unknown function in *Escherichia coli* (referred from here on as the "54 kDa protein family"). Figure 49 shows the position of the common regions in all of these proteins. Note that in the case of the "54 kDa protein family" the region that shows sequence similarity to the "docking protein family" is restricted to the N-terminal region . The C-terminal region of the two members of this family has a very high methionine content. (Römisch *et al.* 1989). It is believed that the function of this region is to recognize the signal sequence (Bernstein *et al.* 1989). The alignment of these two proteins with the "docking protein family" is shown in Figure 48.

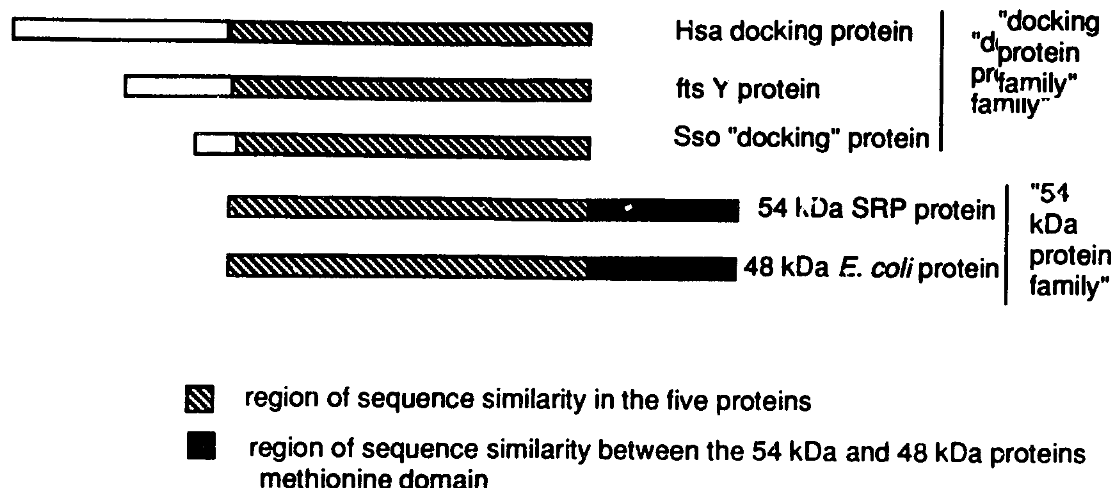


Figure 49. Region that shows sequence similarity between the "docking protein family" and the "54 kDa protein family". The N-terminal region of the members of the "docking protein family" shows little sequence similarity. The C-terminal region of the "54 kDa protein family" is characterized by a high content of methionine (Römisch *et al.* 1989). (Figure adapted from Römisch *et al.* 1989).

Table 30 Sequence identity between the Sso "docking" protein and the 54 kDa SRP protein and 48 kDa *E.coli* protein

protein[1]	length[2]	identities[3]	% identities
54 kDa SRP	289	74	26%
48 kDa Eco	289	84	29%

[1] References can be found in Figure 48 [2] Length of the region of comparison
[3] Only identical residues were considered.

Table 30 shows the the percent sequence identity between these two proteins and the Sso "docking" protein (the Sso "docking " protein is clearly not a member of the "54 kDa protein family", since it lacks the methionine domain and the sequence identity between it and the members of this family is lower than the sequence identity between it and the members of the "docking protein family").

Recently, Connolly and Gilmore (1989) have shown that the α subunit of the docking protein is a GTP binding protein and is involved in the GTP-dependent step of protein translocation across the endoplasmic reticulum membrane. The binding site is located on the C-terminal region of the α subunit and shows sequence similarity to the consensus sequence elements present in other GTP binding proteins.

The GTP binding site is located in the region that the "docking protein family" and the "54 kDa protein family" have in common, and has been conserved in all of them. Figure 50 shows a comparison of the sequence of the GTP binding site in these two protein families with those from other GTP binding proteins. As can be seen in Figure 50, the members of the "docking" and "54 kDa" protein families have sequences that match the first two elements of the consensus sequence (**GXXXXGK** and **DXXG**). These elements have been shown to be involved in interactions with the phosphate portion of the GTP molecule (Dever *et al.* 1987). [illegible]ird element (**NKXD**) is partially conserved as **TKXD** in the two families (see Figure 50). This third element has been shown to be involved in the interaction with the guanine base. The asparagine is thought to interact with the keto group of the guanine ring, the lysine forms part of a hydrophobic pocket and the aspartic acid interacts with the amino group of the guanine ring (Dever *et al.* 1987).

	I		II		III
consensus	G X X X X G K		D X X G		N K X D
Sso	G V N G V G K	75	D T A G	54	T K V D
Hsa	G V N G V G K	89	D T A G	65	T K F D
Cfa	G V N G V G K	89	D T A G	65	T K F D
fts Y	G V N G V G K	75	D T A G	61	T K F D
54 SRP	G L Q G S G K	75	D T S G	54	T K L D
48 Eco	G L O G A G K	76	D T A G	54	T K V D
EF-Tu	G H V D H G K	54	D C P G	50	N K C D
H-ras	G A G G V G K	39	D V G G	54	N K Q D
Gs	G A G E S G K	168	D V G G	64	N K Q D

Figure 50. The GTP binding site in the members of the "docking protein family" and the "54 kDa protein family". The consensus sequence that defines a GTP binding site is shown on the top (Dever *et al.* 1987). The numbers in the boxes indicate the number of residues between each element. The position of each element is indicated in Figure 48. The elements that form the GTP binding site in three other GTP binding proteins are shown for comparison. Adapted from Bernstein *et al.* 1989.

Connolly and Gilmore (1989) have suggested that since there is a deviation in the GTP binding site of the docking protein, this protein must be considered to be part of a new subfamily of GTP-binding proteins different from the four subfamilies already recognized (*i.e.* G proteins, *ras*-like proteins, elongation factors and tubulins). If this is so, the *Sulfolobus* "docking" protein would be the fifth protein identified from this new subfamily.

Although the function of the *Sulfolobus* "docking" protein is not known, the fact that a 7S RNA homologous to the eukaryotic 7S RNA has been detected in the archaebacteria (Moritz and Goebel 1985, Poritz *et al.* 1988, Struck *et al.* 1988, Kaine and Merkel 1989) raises the possibility that the "docking" protein from *Sulfolobus* might play a similar role in protein transport. Isolation of this protein and determination of its location within the *Sulfolobus* cell will be

important in determining its functional role. Evidence that the gene is expressed comes from the fact that a fragment that includes part of the C-terminal region of the gene is protected by *Sulfolobus* mRNA from digestion by nuclease S1 (data not shown).

Open Reading Frames

Seven open reading frames (ORFs) were also identified in the 6.9 Kb Eco RI - Bam HI fragment (see Figure 22). The sequences of these open reading frames and the derived amino acid sequence of their products are shown in Figures 51-57.

Open Reading Frame 1

Eco	**RI**												41
GA	**ATT**	CAA	GCA	CAA	ATA	ACT	GAT	GAT	GAA	CTA	AAA	CAA	ATA
▲	I	Q	A	Q	I	T	D	D	E	L	K	Q	I
	42												80
	TTA	GCT	CAA	CTT	AAC	TCT	CAA	ACC	AGA	AAA	GAT	TAT	AAG
	L	A	Q	L	N	S	Q	T	R	K	D	Y	K
	81										110		
	ATT	ACG	ATT	AAA	GAG	AGA	GGA	TGG	AAA	TGA			
	I	T	I	K	E	R	G	W	K	*			

Figure 51. Sequence of open reading frame 1. The derived amino acid sequence of the protein is shown below the nucleic acid sequence. Only the 3' end of this gene is present in the 6.9 Eco RI-Bam HI fragment.

Open Reading Frame 2

1													42
ATG	AGC	CAG	GAA	ACT	ACT	GCT	ACT	AAC	AAG	AAG	AGC	AAA	AAA
M	S	Q	E	T	T	A	T	N	K	K	S	K	K
43													84
CAT	CAG	AGC	TAC	AAC	AAC	AAG	AAG	AAA	GAG	GAA	CAA	AAA	CCG
H	Q	S	Y	N	N	K	K	K	E	E	Q	K	P
85													126
CAA	CAG	GCA	ACA	ACT	ACA	ACG	AAA	GAG	GAA	AAG	AAA	ACT	AAG
Q	Q	A	T	T	T	T	K	E	E	K	K	T	K
127													168
CCA	GAG	AAA	GAG	AAC	TTT	GAA	ATG	GTT	ATA	AAC	TTT	AGA	AGA
P	E	K	E	N	F	E	M	V	I	N	F	R	R
169													210
GTA	ATA	ATG	GGA	AGA	AAA	ACT	ACC	AGA	ACT	AAA	AGA	GCC	ATA
V	I	M	G	R	K	T	T	R	T	K	R	A	I
211													252
AAG	TAT	GTG	AGA	TAT	ATC	CTA	AAG	AGA	CAT	TTT	GGA	GCG	GAA
K	Y	V	R	Y	I	L	K	R	H	F	G	A	E
253													294
AAA	GTA	ATT	ATT	GAT	CCA	CTA	TTA	GCA	AAG	GCA	ATA	ACC	ATG
K	V	I	I	D	P	L	L	A	K	A	I	T	M
295													336
AAT	GGT	AGG	GAT	AAA	ATA	GTA	AGA	AGA	GTA	AGG	ATA	GCT	GTC
N	G	R	D	K	I	V	R	R	V	R	I	A	V
337											369		
AAA	AGA	ATC	GGA	GAA	AAA	AAC	ATA	TTT	AGC	TAG			
K	R	I	G	E	K	N	I	F	S	*			

Figure 52 Sequence of open reading frame 2. The derived amino acid sequence of the protein is shown below the nucleotide sequence.

Open Reading Frame 3

1 42
ATG ATC TTA GAT AAG TTA TGC TTT TTG GTA CTA ATG ATA AAT
M I L D K L C F L V L M I N

43 84
ATA CAA TTA TAC CTA AAA ATA GAC GAC AAA GAA GTA ATA GAG
I Q L Y L K I D D K E V I E

85 126
AAA ATT CAA GGA ATA CTG AAA ACA GAA ATA ATA CAG ACT ACC
K I Q G I L K T E I I Q T T

127 168
ATA TCT AAG AGC GTG CTA GTA GGT ATT CTG GTT ACA GGA AAT
I S K S V L V G I L V T G N

169 210
AAT GAT GTA ATT CTC CTA CCT AGA ACA GCT CTG GCA GAT GAA
N D V I L L P R T A L A D E

211 252
ATA AAG GTC ATA AAG GAA CAG GCT AAA GAC GTT AGA GTC GAG
I K V I K E Q A K D V R V E

253 294
GTT GTG GAT ATT AGA CCT ACT GCT TTA GGA AAT ATC ATA TTA
V V D I R P T A L G N I I L

295 336
TCC AAC ACG CAT GGT GCA CTT ATT TAC CAA GAT CTT TCT AGG
S N T H G A L I Y Q D L S R

337 345
CGG AAA TAA
R K *

Figure 53 Sequence of open reading frame 3. The derived amino acid sequence of the protein is shown below the nucleotide sequence.

As mentioned before, none of the proteins encoded by these open reading frames showed any significant sequence similarity with any of the proteins present in the data base, so their function is unknown. Although the transcription of these open reading frames was not studied in detail, evidence was obtained that indicates that at least some of them are transcribed. For example, hybridization of *Sulfolobus* mRNA with an oligonucleotide probe complementary to open reading frame 2 showed the presence of a transcript of about 1.3 Kb (data not shown) while primer extension analysis showed that open reading frames 6 and 7 are probably transcribed with the L11 gene (see the next section)

Open Reading Frame 4

1 … 42
ATG GCA GTA ATT ACA GAC AAG GCT GGT TTA GTA CAT ATT GAC
M A V I T D K A G L V H I D

43 … 84
GCA ACT GAA GAA GAG TTG AAA AAA TAT GTA CTA AAC CAG CCA
A T E E E L K K Y V L N Q P

85 … 126
GCC AGT GAA TTA TTT AAA GTA AAG TTA GAT TCT GGA ACT GTA
A S E L F K V K L D S G T V

127 … 168
AAT TTT GGG AGC GTC TTT ATA AGG AGC GGA TTA GTG GCT AAC
N F G S V F I R S G L V A N

169 … 210
AGA AAT GGA GTT CTA GTA GGT TCC TCA ACA ACG GGA CAG AGA
R N G V L V G S S T T G Q R

211 216
TTT TAA
F *

Figure 54 Sequence of open reading frame 4. The derived amino acid sequence of the protein is shown below the nucleotide sequence.

Open Reading Frame 5

1 42
ATG AGT GAA GAG GCA GAG CAA CAA CAA GCG GCA GAA TAT ATA
M S E E A E Q Q Q A A E Y I
43 84
GCA TAC CTA TAT GAT CAA GCA TCA GCA CTA AGA CAA TAT ATA
A Y L Y D Q A S A L R Q Y I
85 126
GAT ACC CTT CAG AAG AAC CTG GCA GAG GTA CTA GAA TCC TTA
D T L Q K N L A E V L E S L
127 168
GAA GCA GTA AGA GCA TCA AAA AGT GCC GTA GAC GAG ATA GGT
E A V R A S K S A V D E I G
169 210
AAG GAA AAC CAG GAA TAC TTG TTA TTT GGC GAC AGA AAG GGC
K E N Q E Y L L F G D R K G
211 252
AAT ATA GTA TTC AAG GTA AAC AGT GTA GAT AAA AAT AAG GTT
N I V F K V N S V D K N K V
253 294
CTA ATA CAC TTA GGA TTA AAT TAT TAC GCA GAA GTG ATC CAC
L I H L G L N Y Y A E V I H
295 336
AAG CAG CTA AGA AAA TTC TGG ATG ATA GAG AGC AAC AGT TGG
K Q L R K F W M I E S N S W
337 378
CTG AAG TAT CTA AAA ATA TAC AAG GTG AAT TAT CTA AGT CAA
L K Y L K I Y K V N Y L S Q
379 381
TAG
*

Figure 55 Sequence of open reading frame 5. The derived amino acid sequence of the protein is shown below the nucleotide sequence.

Open Reading Frame 6

1 42

ATG AAA GTT TTT AGT ATA ATG AAA ATT TCT GAC ATT ATC AAA

M K V F S I M K I S D I I K

43 84

AGG CTA CGA GAA GAC TGG AAG AGA ATA ATT AGT GTA GCC AAG

R L R E D W K R I I S V A K

85 126

AAA CCA GAC AAA GAT TCA TTT AAT TAC AGT ATT AGG CTT ACA

K P D K D S F N Y S I R L T

127 168

CTT TTA GTT ATG GCT GTA GTT GGT CTC ATC GCT TAC ATT GTA

L L V M A V V G L I A Y I V

169 198

CAG TTA ACT ACA TCA CTG ATA ATA AGG TGA

Q L T T S L I I R *

Figure 56 Sequence of open reading frame 6. The derived amino acid sequence of the protein is shown below the nucleotide sequence.

Open Reading Frame 7

1													42
ATG	GAG	GAC	TTC	AAG	TAC	AGA	AAT	TAT	TAC	GTG	TTA	AGA	GTT
M	E	D	F	K	Y	R	N	Y	Y	V	L	R	V
43													84
ACA	GGA	GGA	CAA	GAA	ATT	AAC	GTG	GCT	CTT	ATT	TTA	GAA	GAA
T	G	G	Q	E	I	N	V	A	L	I	L	E	E
85													126
AGA	ATA	AAA	ACA	AAT	AAT	ATA	AAT	GAA	ATC	TTT	TCT	GTA	GTA
R	I	K	T	N	N	I	N	E	I	F	S	V	V
127													168
GTC	CCA	CCT	AAT	ATA	AAG	GGT	TAT	GTG	ATA	CTT	GAA	GCT	ACC
V	P	P	N	I	K	G	Y	V	I	L	E	A	T
169													210
GGA	CCT	CAT	GTA	GTA	AAA	TTA	ATT	TCA	TCA	GGA	ATA	AGA	CAC
G	P	H	V	V	K	L	I	S	S	G	I	R	H
211													252
GTT	AAA	GGA	GTA	GCT	CAT	GGA	CTA	ATT	CAG	AAA	GAA	GAT	GTT
V	K	G	V	A	H	G	L	I	Q	K	E	D	V
253													294
ACA	AAA	TTT	GTT	TCT	AAG	TCA	GTT	GCA	TTA	CCC	GCT	GTG	AAA
T	K	F	V	S	K	S	V	A	L	P	A	V	K
295													336
GAA	GGA	GAC	TTA	GTA	GAG	GTA	ATT	AGT	GGT	CCA	TTT	AGA	GGT
E	G	D	L	V	E	V	I	S	G	P	F	R	G
337													378
ATG	CAG	GCT	CAA	GTT	GTG	AGA	GTA	GAA	TCA	ACA	AAA	AAT	GAA
M	Q	A	Q	V	V	R	V	E	S	T	K	N	E
379													420
GTA	GTT	TTA	AAT	ATT	TTA	GAA	TCC	TCA	TAT	CCT	GTT	CAG	GTT
V	V	L	N	I	L	E	S	S	Y	P	V	Q	V
421												459	
ACA	GTT	CCT	TTG	GAA	CAA	GTT	AAA	CCT	GTT	AAG	AGG	TGA	
T	V	P	L	E	Q	V	K	P	V	K	R	*	

Figure 57 Sequence of open reading frame 7. The derived amino acid sequence of the protein is shown below the nucleotide sequence.

Organization and Transcription of the Sso L11, L1, L10 and L12 Ribosomal Protein Genes

Sequencing of the 6.9 Kb Eco RI - Bam HI fragment revealed that the order of these four ribosomal protein genes in *Sulfolobus solfataricus* is identical to that found in *Escherichia coli*, that is 5' L11-L1-L10 -L12 3' (Post *et al.* 1979). These genes are also in the same order in the extreme halophile, *Halobacterium cutirubrum* (Shimmin and Dennis 1989). In *Methanococcus vannielii*, where only the genes for the L10 and L12 proteins have been cloned, the order of these two genes has also been conserved (Strobel *et al.* 1988, Köpke *et al.* 1989). However, although the order of these ribosomal protein genes has been conserved in all of these organisms, the genes surrounding them are different.

In *E. coli*, the L11 gene is preceded by the "U" gene whose product is believed to be involved in transcription termination (Downing and Dennis 1987) and the Eco L12 gene is followed by the genes coding for the β and β' subunits of RNA polymerase (Post *et al.* 1979). In *H. cutirubrum*, the L11 gene is preceded by the "NAB" gene which encodes a protein that shows sequence similarity to restriction endonucleases Eco RI and Pst I (Shimmin and Dennis 1989). The L12 gene, in this case, is followed by at least a 784 nucleotide region which does not code for any protein in either strand (Shimmin and Dennis 1989). In the case of *Sulfolobus solfataricus*, the L11 gene is preceded by two ORFs (6 and 7) which potentially code for proteins that show no sequence similarity to any of the proteins present in the data base or to the NAB product in *H. cutirubrum.* The L12 gene in *Sulfolobus* is followed by the *ala S* gene. This gene in *E. coli*, is located ≈ 700 nucleotides downstream of the *rec A* gene (Putney *et al.* 1981a). The genes that code for the equivalent

subunits of the β and β' subunits of the Eco RNA polymerase have been recently cloned from another *Sulfolobus* species: *S. acidocaldarius*, and they are located upstream of the Sac S12 gene (Zillig *et al.* 1989, Pühler *et al.* 1989).

The intergenic spaces are also different. There are 3 nucleotides between the termination codon of the Eco L11 gene and the initiation codon of the Eco L1 gene, 410 between the Eco L1 and Eco L10 genes, 66 between the Eco L10 and Eco L12 genes, and 121 between the Eco L12 and Eco *rpo B* genes in *E. coli* (Post *et al.* 1979). In *H. cutirubrum*, there are 203 nucleotides between the Hcu L11 and Hcu L1 genes, 4 between the Hcu L1 and Hcu L10 genes, and 8 between the Hcu L10 and Hcu L12 genes (Shimmin and Dennis 1989). In *Sulfolobus*, the Sso L11 and L1 genes as well as the Sso L1 and Sso L10 genes have overlapping stop/start codons (which cause a change in reading frame) (Figure 58), while there are 41 nucleotides between the Sso L10 and Sso L12 genes and 40 nucleotides between the Sso L12 and *ala S* genes.

Overlapping stop/start codons

end of L11

Sso L11-Sso L1 GCAGAAGG**GTG**AAAGAA
L1

end of L1

Sso L1-Sso L10 AAAAGAGC**GTG**ATCGGT
L10

Figure 58. Overlapping stop/start codons in the ribosomal protein genes of *Sulfolobus solfataricus.* The termination codon is overlined and the initiation codon is underlined and in bold letters..

Overlapping stop/start codons have been observed between several ribosomal protein genes in *Escherichia coli* (for example, L4-L23, L16-L29, L29-S17 (Zurawski and Zurawski 1985)) and in other eubacteria (for example, L4-L23, L16-L29, L29-S17, L15-*sec Y* in *Mycoplasma capricolum* (Ohkubo *et al.* 1987) and L4-L23 in *Yersinia pseudotuberculosis* (Gross *et al.* 1989)), as well as between genes that code for proteins with other cellular functions (for example, *gal T-gal K* (Shümperli *et al.* 1982), *trpB-trp A*, and *trpE-trpD* (Yanofsky *et al.* 1981, Askoy *et al.* 1984) in *E.coli*). They have also been detected between the chloroplast genes for the β and ε subunits of ATP synthetase (Gatenby *et al.* 1989) and, in the archaebacteria, between the genes that code for the α and β subunits of formate dehydrogenase in *Methanobacterium formicicum* (Shuber *et al.* 1986), the genes that code for 3 subunits of the RNA polymerase in *Halobacterium halobium* (B' and A, A and C, C and ORF 139) (Leffers *et al.* 1989), the A and C subunits of the RNA polymerase and the S12 and S17 genes in *Halobacterium morrhuae* (Leffers *et al.* 1989) and the B and A subunits of the RNA polymerase in *Sulfolobus acidocaldarius* (Pühler *et al.* 1989).

It has been proposed that overlapping stop/start codons represent a mechanism by which the cell couples the translation of both genes in order to obtain an equimolar synthesis of both proteins (Yanofsky *et al.* 1981, Schümperli *et al.* 1982, Askoy *et al.* 1984, Sor *et al.* 1987, Gold 1988, Gatenby *et al.* 1989). In the eubacteria, experiments in which deletions and mutations are introduced in the region preceding the overlap have given support to this idea (Schümperli *et al.* 1982, Askoy *et al.* 1984, Sor *et al.* 1987,Gatenby *et al.* 1989).

The mechanism of translational coupling is poorly understood. An overlapping stop/start codon is not strictly required for coupling, since intercistronic regions that range between 3 and 67 nucleotides have been found between genes that are translationally coupled (Sor *et al.* 1987). In general, it is thought that the initiation site for the second cistron is somehow masked or inhibited, *i.e.* by secondary structure in the mRNA, so that only translating ribosomes approaching from upstream of the mRNA, but not 'free' ribosomes, can effectively initiate translation (Sor *et al.* 1987). In the case of overlapping stop/start codons, it has been proposed that the ribosome terminates the translation of the first cistron and then is 'captured' by the interaction of the 16S rRNA with the mRNA and re-initiates translation (Gold 1988). The recent finding that there is an interaction between the 3' end of the 16S rRNA and the mRNA during the elongation step of protein synthesis seems to give support to this idea (Weiss *et al.* 1988).

Although translational coupling has not been demonstrated in the archaebacteria, the presence of overlapping stop/start codons suggests that this mechanism might also be used in this kingdom as a means of translational control.

The *in vivo* transcripts derived from the L11-L1-L10-L12 gene cluster were analyzed by primer extension, S1 mapping and Northern blot hybridization. Figure 59 shows the positions of the fragments used for S1 mapping, the probes used for Northern blot hybridizations and the primers used for the primer extension experiments.

Hybridization of *Sulfolobus* mRNA with probes (A-D in Figure 59) complementary to the four ribosomal protein genes revealed the presence of a major transcript of about 2.5 Kb (see Figure 60). Since the combined length

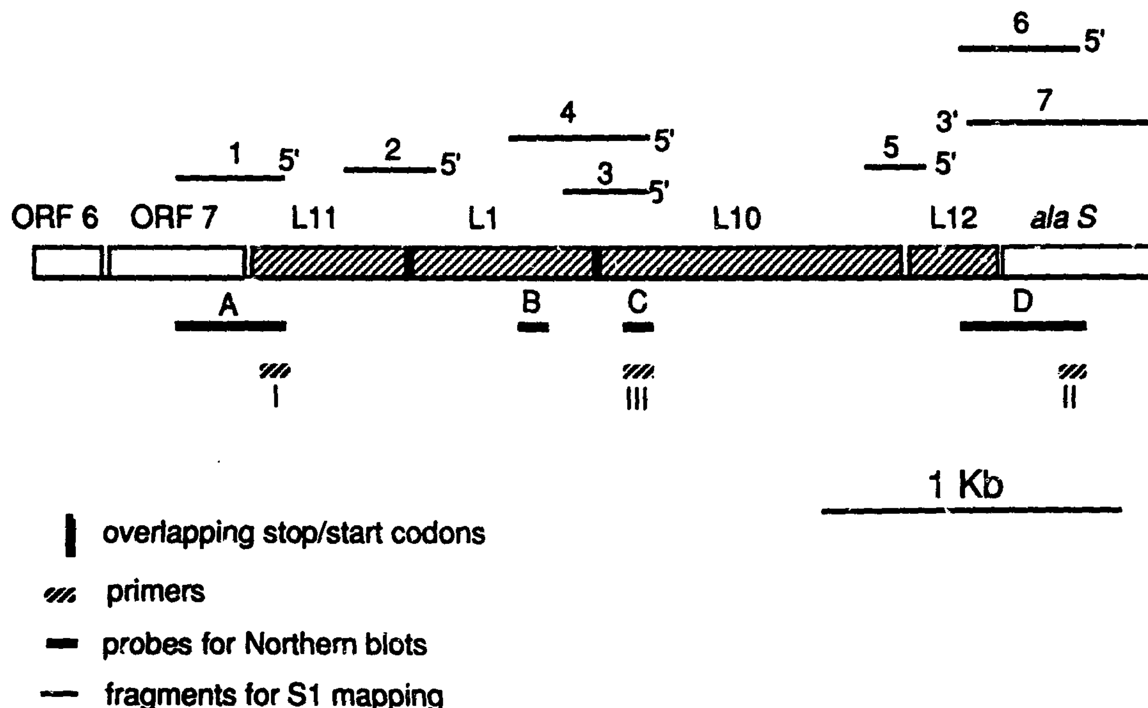


Figure 59. Position of the probes used for S1 mapping, primer extension, and Northern blot hybridization. A physical map of the chromosomal region of *Sulfolobus solfataricus* encoding the L11-L1-L10 and L12 gene cluster is shown. Straight lines indicate the position of the fragments (1-7) used for S1 mapping. The end that was labeled is indicated 1: 144 bp (15-Dra I fragment), 2: 372 bp (11-Taq I fragment), 3: 239 bp (7-Dde I fragment), 4: 388 bp(7-Hinf I fragment), 5: 149 bp (3-Alu I fragment), 6: 693 bp (Pst I- Bam HI fragment), 7: 474 bp (1-Pst I fragment). (For a description of how these fragments were generated, see Materials and Methods). Probes for Northern blots (A-D) are indicated by bars. A: 144 bp (15-Dra I), B: oligonucleotide 5' GTGCCATAGATTCCTGG 3', C: oligonucleotide 5'GTTACCTTTATGTCAGC 3', D: 474 bp (1- Pst I fragment). Hatched bars indicate the primers used for primer extension: I (5'CCCAATGGTGGACCTGG 3'), II (5' GTCAACTATACTAGCTA 3'), III (5'GTTACCTTTATGTCAGC 3').

of the four ribosomal protein genes is approximately 2.5 Kb, this result suggested that the four genes are transcribed as a single unit. The results obtained with the Northern blot experiments were later confirmed by the fact that all the fragments used to probe the intergenic regions between the four ribosomal protein genes (fragments 2, 3 and 5) were protected by *Sulfolobus* mRNA from digestion by nuclease S1 (data not shown).

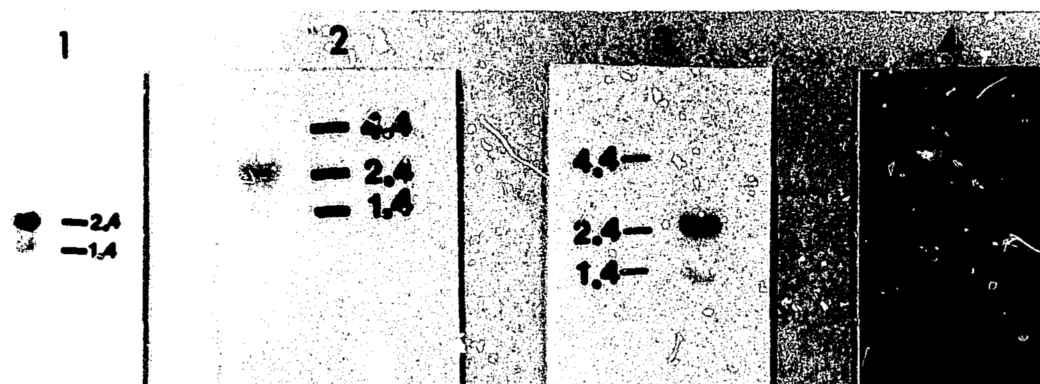


Figure 60. Northern blot hybridizations. *Sulfolobus* RNA bound to Z-probe membranes was hybridized to probes complementary to each of the four ribosomal protein genes (L11, L1, L10, L12). Hybridization conditions and the generation of the probes are described in Materials and Methods. A picture of the autoradiographs obtained is shown. Panel 1: Hybridization with probe A (144 bp fragment [15-Dra I]) complementary to the L11 gene. Panel 2: Hybridization with probe B (oligonucleotide: 5' GTGCCATAGATTCCTGG 3') complementary to the L1 gene. Panel 3: Hybridization with probe C (oligonucleotide: 5' GTTACCTTTATGTCAGC 3') complementary to the L10 gene. Panel 4: Hybridization with probe D (474 bp fragment [1-Pst I]).

The 5' end of this transcript was determined by S1 mapping, using a 144 bp fragment (fragment 1 in Figure 59) and by primer extension, using a primer that binds to a region located 63 nucleotides downstream of the initiation codon of the L11 gene (primer I) (see Figure 59). The S1 mapping results

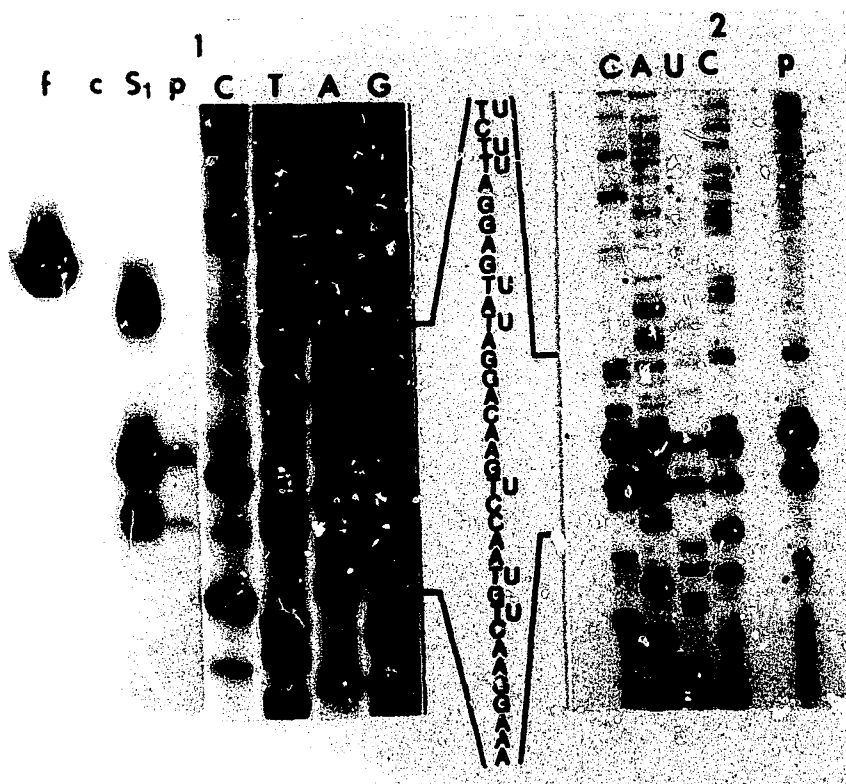


Figure 61. Primer extension and nuclease protection experiments to determine the 5' end of the 2.5 Kb transcript. A 144 bp fragment (fragment 1 in Figure 59) was used in the S1 mapping experiments and primer I (5' CCCAATGGTGGACCTGG 3') was used in the primer extension reactions. Panel 1: f: 144 bp fragment, c: fragment treated with S1 nuclease in the absence of *Sulfolobus* RNA (control) , S1: fragment treated with S1 in the presence of 10 µg of *Sulfolobus* RNA, p: primer extension reaction carried out at 37°C, lanes CTAG: DNA sequencing reactions using unlabeled primer I (because the primer lacks a phosphate group the sequencing ladder has a slightly different mobility than the fragments). Panel 2: Primer extension reactions carried out at 50°C, lanes GAUC: RNA sequencing reactions using labeled primer I (in this case the mobility of the fragment and the RNA sequencing ladder is equivalent), p: primer extension reactions. The sequence of the DNA (T) and RNA (U) is shown.

indicated the presence of three possible initiation sites (see Figure 61, panel 1 lane S1). However, primer extension analysis showed only two 5' ends which

correspond to the two lower bands observed in the S1 mapping experiment (see Figure 61, panel 1, lane p). In order to determine if the absence of the third band in the primer extension analysis is due to the presence of secondary structure in the mRNA, which would prevent the enzyme from extending the full length of the mRNA, or if this third band is an artifact, primer extension reactions were carried out at a higher temperature (50°C). The results from this experiment not only confirmed the presence of only two initiation sites but also showed that there is read through from the genes upstream into the L11 gene (see Figure 61, panel 2, lane p). Notice that there is a dark band running above the two bands that correspond to the transcription initiation sites. This band is probably due to secondary structure in the mRNA (even at this higher temperature) since its position doesn't match the position of any of the S1 bands. Thus, the upper band observed in the S1 mapping experiments is probably an artifact produced by the S1 nuclease and corresponds to full protection of the probe.

The position of the two transcription initiation sites is shown in Figure 62. The sequence **TTTAAA** that matches the consensus sequence for archaebacterial promoters (**TTTAT/AA**) proposed by Zillig *et al.* (1988) was found 27 and 35 nucleotides upstream of the two transcription initiation sites, respectively. However, the consensus sequence observed at the transcription initiation site (**A/TTGA/C**) is only partially present in the first site (**TAGA**). A second sequence **AAATATTTTAG** that partially matches the consensus sequence for promoters in the sulfur-dependent extreme thermophiles (**AAANNTTTAAA**) proposed by Brown *et al.* (1989) was also detected (see Figure 62).

The L11 Promoter

GAAGTAGTTTTAAATATTTTAGAATCCTCATATCCTGTTCAGGTTACAGTT

CCTTTGAACAAGTTAAACCTGTTAAGAGGTGACAGATAA**TTG**
Sso L11

Figure 62. The L11 promoter region. The sequence of the region upstream of the L11 gene is shown. The two transcription initiation sites are indicated by a star. The sequence that matches the consensus sequence of archaebacterial promoters (TTTAT/AA) proposed by Zillig *et al.* (1989) is underlined. The sequence that matches the consensus sequence for the promoters in the extreme thermophiles (AAANNTTTAAA) proposed by Brown *et al.* (1989) is overlined. The initiation codon of the Sso L11 gene is indicated in bold letters.

The 3' end of the transcript was mapped using a Pst I-Bam HI fragment labeled at its 3' end (fragment 7 in Figure 59). The results from this experiment are shown in Figure 63. Several transcription stops were detected, but the main one seems to be the one corresponding to the 149 bp band. The transcription termination sites are shown in Figure 64 . Note that the main transcription termination site is located 6 nucleotides upstream of the initiation codon of the Sso *ala S* gene. All of the transcription stops are located within or after pyrimidine rich regions (see Figure 64). This has been shown to be characteristic of transcription termination sites in the sulfur-dependent extreme thermophiles (Zillig *et al.* 1988).

The Northern blot experiments (Figure 60) using a probe (probe A) complementary to the L11 gene showed that besides the 2.5 Kb transcript, there is a smaller transcript of about 1.5 Kb that contains this gene.

Figure 63. Nuclease protection experiments to determine the 3' end of the 2.5 Kb transcript. A 639 bp Pst I- Bam HI fragment labeled at its 3' end was used in this experiment. f: 639 bp fragment, S1: 639 bp fragment treated with nuclease S1 in the presence of 10 µg of *Sulfolobus* RNA. The size of the bands (determined from molecular weight markers) is indicated.

Since a transcript of a similar size is not observed when *Sulfolobus* mRNA is probed with an oligonucleotide probe (probe B) complementary to the L1 gene (see Figure 60), the complete L1 gene is not present in this transcript. This means that this transcript must contain the genes upstream from the L11 gene (probably ORFs 6 and 7 because the combined length of these two genes and the L1 gene is about 1.5 Kb). The fact that the primer extension reactions

Transcription termination sites of the L11-L1-L10-L12 transcript

end of Sso L12 *ala S*

GGT**TAA**TAAGAAAGACTTTTTTACCATTTCAAATATCATTTTTACC**ATG**G

TCAAAGCAAACGAGAACGAGTACAGGTTAAACTTTTTTCTGTCAAGAGG

ATATGAAAGAAAAATTTGTAGATCCTGTTCTACCCCATTTTGGAC

Figure 64. Transcription termination sites after the L12 gene. The sequence of the region downstream of the Sso L12 gene is shown. Transcription termination sites are indicated by a star. The major termination site is indicated by an arrow. The termination codon of the Sso L12 gene and the initiation codon of the *ala S* gene are indicated in bold letters. Pyrimidine rich regions are underlined.

showed that there is read through from the genes upstream of the L11 gene supports this conclusion. Since the fragment used to probe the region between the L11 and L1 genes (fragment 2 in Figure 59) was fully protected (data not shown), the 3' end of this transcript must lie in the region between fragment 2 and probe B (see Figure 59). The 5' and 3' ends of this transcript are currently under investigation.

The Northern blots for the L10 and L12 genes also showed the presence of a second band (approximately 1.4 Kb) (see Figure 60) . Since this band is not observed when *Sulfolobus* mRNA is probed with an oligonucleotide specific for the L1 gene, and the combined length of these two genes is approximately 1.4 Kb, it is probable that this transcript only contains these two genes. S1 mapping experiments using a fragment that covers the region between the L1 and L10 genes (fragment 4 in Figure 59), revealed the presence of a very faint

band that matches a band in the primer extension analysis, using primer III (see Figure 65). The 5' end of this band was determined to be a region located 20 nucleotides upstream of the L10 gene (see Figure 66). Sequences that match the consensus sequence for archaebacterial promoters proposed by Zillig *et al.* (1988) and Brown *et al.* (1989) were detected upstream of this site (see Figure 66), so it is probably a transcription initiation site, although the possibility still exists that it is a degradation product of the 2.5 Kb transcript.

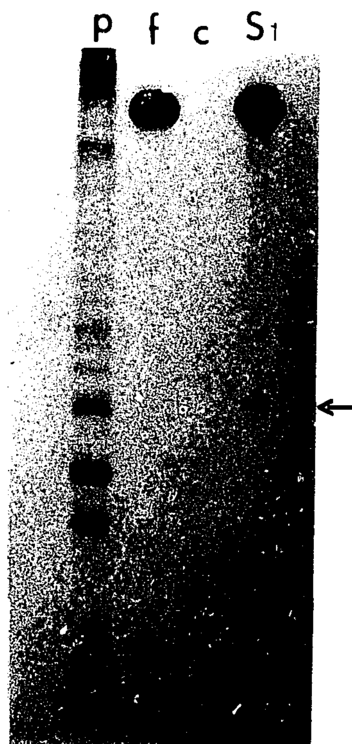


Figure 65. Nuclease protection experiment to determine the 5' end of a 1.4 Kb transcript containing the L10 and L12 genes. A 388 bp fragment (fragment 4 in Figure 59) was used in this experiment. p: primer extension reactions using primer III (5' GTTACCTTTATGTCAGC 3') at 50°C. f: 388 bp fragment, c: 388 bp fragment treated with nuclease S1 in the absence of *Sulfolobus* RNA, S1: 388 bp fragment treated with S1 nuclease in the presence of 10 µg of *Sulfolobus* RNA. The position of a very faint band in the S1 lane is indicated by an arrow.

The L10 promoter

Sso L10

AAATATTTATGTTAAAACAACAATGGGTAAAGCTGTAAAAGTTAAAAGAG**GTG**

Figure 66. The L10 promoter. The region upstream of the L10 gene is shown. The transcription initiation site is indicated by a star. A sequence that partially matches the consensus sequence (TTTAT/AA) for archaebacterial promoters proposed by Zillig *et al.* (1988) is underlined. A sequence that partially matches the consensus sequence for the promoters of the extreme sulfur-dependent thermophiles (AAANNTTTAAA) proposed by Brown *et al.* (1989) is overlined. The initiation codon of the Sso L10 gene is indicated in bold letters.

The transcription of the *ala S* gene was also studied. Primer extension analysis using primer II (see Figure 59), revealed the presence of only one transcription initiation site (see Figure 67). S1 mapping experiments using a 474 bp fragment (fragment 7 in Figure 59) confirmed this result (see Figure 67) Note that part of the probe was protected. Prolonged exposure of the primer extension experiments showed that there is read through from the L12 gene (data not shown). However, transcripts longer than 2.5 Kb were not observed in the Northern blots when a probe for the L12 gene was used (see Figure 60). It is possible that these longer transcripts were not detected because the efficiency of transfer of high molecular RNA to the hybridization membrane is usually low.

The transcription initiation site, in this case, is located 3 nucleotides upstream of the initiation codon of the *ala S* gene (see Figure 68). A sequence (**TTAATA**) that partially matches the consensus sequence for archaebacterial promoters (**TTTAT/AA**) proposed by Zillig *et al.* (1988) and a sequence

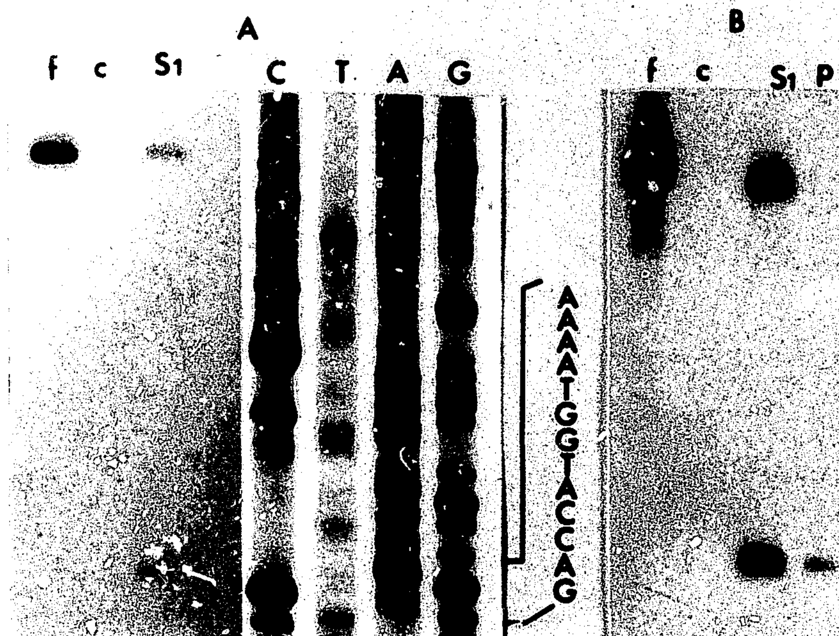


Figure 67. Nuclease protection and primer extension experiments to determine the transcription initiation site of the *ala S* gene. Panel A: S1 mapping experiments using a 474 bp fragment (fragment 7 in Figure 59). f: 474 bp fragment, c: 474 bp fragment treated with S1 nuclease in the absence of *Sulfolobus* RNA, S1: 474 bp fragment treated with S1 nuclease in the presence of 10 µg of *Sulfolobus* RNA, lanes CTAG: DNA sequencing reactions using unlabeled primer II (5' GTCAACTATACTAGCTA 3'). Panel B: f: 474 bp fragment, c: 474 bp fragment treated with S1 nuclease in the absence of *Sulfolobus* RNA, S1: 474 bp fragment treated with S1 nuclease in the presence of 10 µg of *Sulfolobus* RNA, p:.primer extension reactions using primer II, carried out at 37°C.

(**AGACTTTTTTA**) that partially matches the consensus sequence for the promoters (**AAANNTTTAAA**) of the extreme sulfur-dependent thermophiles proposed by Brown *et al.*(1989) were identified upstream of the transcription initiation site (Figure 68).

The *ala S* promoter

GTTAATAAGAAAGACTTTTTTACCATTTCAAATATCATTTTTACCATG

Figure 68. The *ala S* promoter. The sequence of the region upstream of the Sso *ala S* gene is shown. The transcription initiation site is indicated by a star. A sequence that matches the consensus sequence (TTTAT/AA) for archaebacterial promoters proposed by Zillig *et al.* (1988) is underlined. A sequence that partially matches the consensus sequence (AAANNTTTAAA) proposed by Brown *et al.* (1989) for the promoters of sulfur-dependent extreme thermophiles is overlined. The initiation codon of the *ala S* gene is indicated in bold letters.

Analysis of the *in vivo* transcripts derived from the L11-L1-L10-L12 gene cluster, revealed that the major transcript contains the four ribosomal protein genes (see Figure 69). In addition, a less abundant bicistronic transcript containing the L10 and L12 genes, and a tricistronic transcript, probably containing ORFs 6 and 7 as well as the L11 gene, were also detected. In contrast, in the extreme halophile, *Halobacterium cutirubrum*, Shimmin and Dennis (1989) found that the major transcripts are a monocistronic L11 and a tricistronic L1-L10-L12 transcripts (see Figure 69). Thus the transcription of these four genes varies within the archaebacteria.

In *Escherichia coli*, Downing and Dennis (1987) have found that the major transcript is a tetracistronic L11-L1-L10-L12 transcript, similar to the *Sulfolobus* transcript. However, bicistronic L11-L1 and L10-L12 transcripts are also found. Thus, although the order of the genes has been conserved in the archaebacteria and eubacteria the transcription products are different.

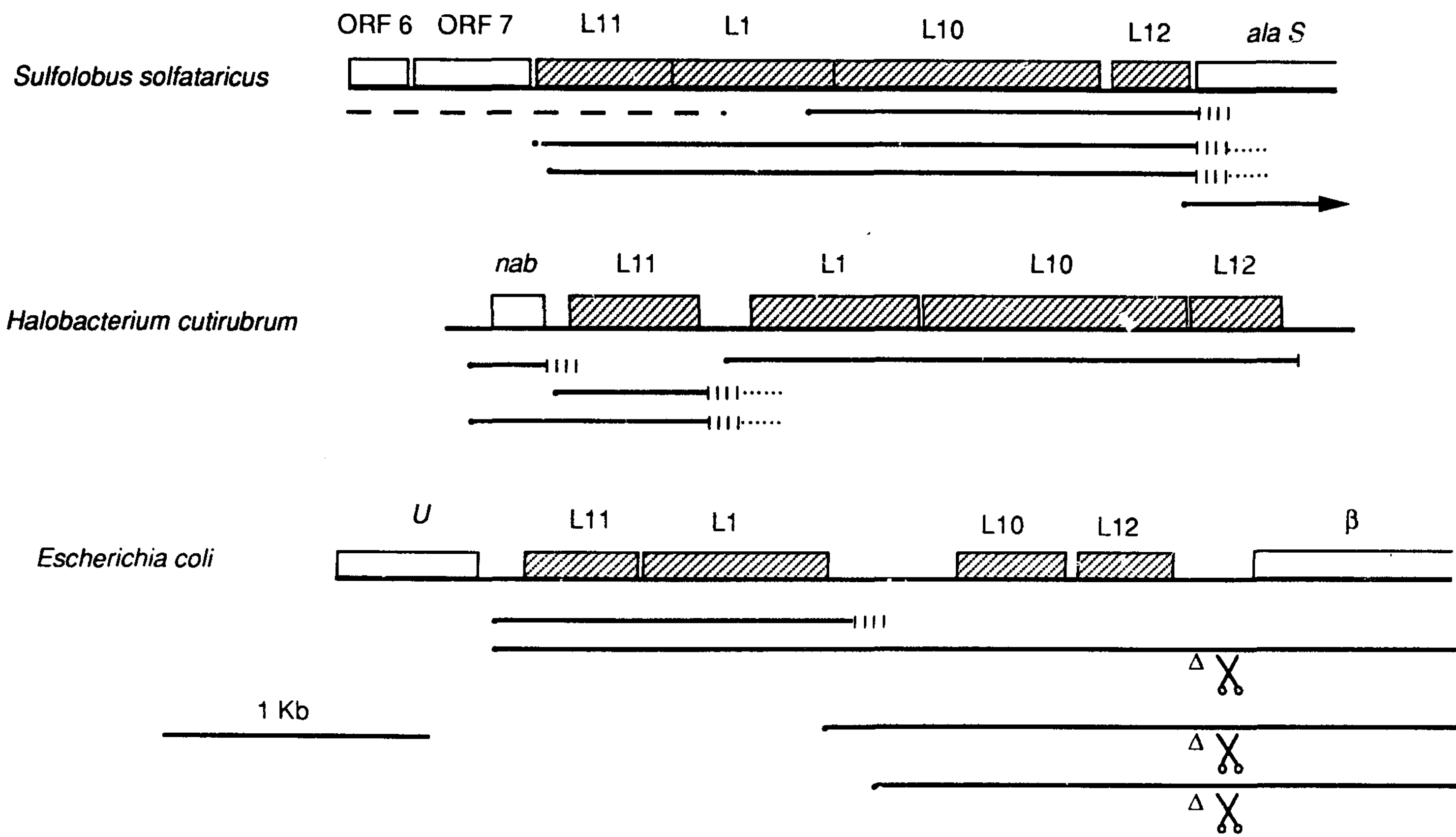


Figure 69. Transcription of the L11-L1-L10-L12 genes in *Sulfolobus solfataricus*. A comparison of the transcription pattern of the L11-L1-L10-L12 genes in *S. solfataricus, Halobacterium cutirubrum* (Shimmin and Dennis 1989), and *Escherichia coli* (Downing and Dennis 1987) is shown. Filled circles indicate the 5' ends of transcripts and vertical lines the 3' ends. Multiple vertical lines indicate several transcription stops. Dotted lines indicate read through to the genes downstream. A dashed line indicates the 1.5 Kb transcript whose 5' and 3' ends have not been determined. A triangle indicates an attenuator and scissors an RNAse III processing site within the L12-β intergenic region in *E. coli*. Only part of the *ala S* and β genes is shown. ORF: open reading frame, *ala S*: alanine synthetase gene,*nab:* nucleic acid binding protein, U: gene that codes for a protein involved in transcription termination, β: β subunit of the RNA polymerase.

Codon Utilization

Table 31 summarizes the codon utilization pattern of the six ribosomal (L11, L1, L10, L12, L46 and LX) protein genes. As expected from the low G-C content of *Sulfolobus* DNA (36%, Zillig *et al.* 1980), there is a strong bias towards having A or U in the wobble position (76%) (see Table 32). Notice that the UGU/UGC codons for cysteine and the CGN codons for arginine are rarely used in the ribosomal protein genes of *Sulfolobus*. While the case of the cysteine codons simply reflects the fact that cysteine is seldom present in ribosomal proteins (Wittmann-Liebold 1986), the case of the arginine codons does reflect a certain bias. It seems that this bias is not restricted to the ribosomal protein genes, since the CGN codons for arginine are also rarely used in the "docking" protein gene, the *ala S* gene and the open reading frames (see Tables 26 and 27, data for the open reading frames not shown). A similar situation has been observed in the ribosomal protein genes of *Methanococcus vannielii*, which also has a very low G-C content (Auer *et al.* 1989a, b). In contrast in *Halobacterium cutirubrum*, where the G-C content is very high (68%), there is a strong bias towards having a C or a G in the wobble position and the AGA/AGG codons for arginine are seldom used (Arndt and Kimura 1988, Shimmin and Dennis 1989).

However, even though *S. solfataricus* and *M. vannielii* have a low G-C content, comparison of the codon utilization pattern in the ribosomal protein genes from both organisms revealed that *Sulfolobus* shows a bias towards using G in the wobble position instead of C (see Table 32), while the reverse is true for *M. vannielii.* (Auer *et al.* 1989a, b).

Table 31 Codon Utilization in the *Sulfolobus* ribosomal protein genes

UUU	Phe	17	UCU	Ser	6	UAU	Tyr	15	UGU	Cys	0
UUC	Phe	9	UCC	Ser	2	UAC	Tyr	8	UGC	Cys	0
UUA	Leu	41	UCA	Ser	17	UAA	*	2	UGA	*	4
UUG	Leu	8	UCG	Ser	1	UAG	*	0	UGG	Trp	5
CUU	Leu	13	CCU	Pro	17	CAU	His	4	CGU	Arg	0
CUC	Leu	3	CCC	Pro	7	CAC	His	3	CGC	Arg	0
CUA	Leu	16	CCA	Pro	22	CAA	Gln	27	CGA	Arg	0
CUG	Leu	3	CCG	Pro	0	CAG	Gln	13	CGG	Arg	0
AUU	Ile	19	ACU	Thr	24	AAU	Asn	31	AGU	Ser	14
AUC	Ile	6	ACC	Thr	4	AAC	Asn	12	AGC	Ser	6
AUA	Ile	54	ACA	Thr	28	AAA	Lys	84	AGA	Arg	24
AUG	Met	16	ACG	Thr	2	AAG	Lys	38	AGG	Arg	7
GUU	Val	24	GCU	Ala	27	GAU	Asp	26	GGU	Gly	16
GUC	Val	6	GCC	Ala	12	GAC	Asp	11	GGC	Gly	6
GUA	Val	28	GCA	Ala	46	GAA	Glu	51	GGA	Gly	35
GUG	Val	8	GCG	Ala	7	GAG	Glu	22	GGG	Gly	2

* chain termination

Table 32 Base composition in the wobble position in *Sulfolobus* ribosomal protein genes

Base	number	%
A	479	49.9%
U	253	26.4%
C	95	9.9%
G	132	13.7%

Since extensive sequencing of the 5' ends of three of the ribosomal proteins genes (L11, L1 and L10) and the "docking" protein gene failed to reveal the presence of a normal ATG codon in frame, before the next termination codon, it was concluded that these genes are initiated by initiation

codons that are less frequently used (see Table 33). Evidence that these codons are indeed used as the initiation codon is only available for the Sso L11 gene where the N-terminal sequence of the protein has been determined (Matheson 1985) and for the Sso L10 gene where the amino acid composition of the N-terminal peptide has been determined (Louie, K.A., personal communication). Until the products of the other genes are isolated and their N-terminal sequence determined, the identification of the codons shown in Table 33 as the real initiation codons will remain tentative.

The presence of rarely used initiation codons has also been noted in other archaebacterial genes, such as the genes that code for the B', C and A subunits of the RNA polymerase in *Methanobacterium thermoautotrophicum* (Berghöfer *et al.* 1988) and the *nif* gene in *Methanococcus voltae* (Souillard and Sibold 1986). In this respect, the archaebacteria seem to resemble the eubacteria, where the presence of unusual initiation codons has frequently been observed (for reviews, see Kozak 1983, Gold 1988).

Table 33 Unusual initiation codons in *Sulfolobus solfataricus*

Gene	Initiation codon
Sso L11	TTG
Sso L1	GTG
Sso L10	GTG
"docking"	ATA

Translation Signals

Sequences complementary to the 3' end of the *Sulfolobus solfataricus* 16S rRNA were identified in the region around the initiation codon of some of the

genes and open reading frames (ORFs) present in the 6.9 Kb fragment (see Table 34).

Table 34 Shine-Dalgarno Sequences

3' end of the *Sulfolobus* 16S rRNA	3' **ACUCCACUAGC** 5'[1]
Shine-Dalgarno sequence	5' **TGAGGTGATCC** 3'

Gene	**Shine-Dalgarno Sequence**[2]
Sso L46	GAGAGGATGGAA**ATG**
Sso LX	AGTGATCAGT**ATG**
Sso L11	GAGGTGACAGATAA**TTG**
Sso L1	CAGAAG**GTG**
Sso L10	GAGC**GTG**ATCGGT
Sso L12	AGGTGATTATAT**ATG**
"docking"	AGGTAAAAAAGGTGA**ATA**ATT
ala S	ATTTTTACC**ATG**GTCAAAGC
ORF 2	AGGTGTGAATAA**ATG**
ORF 3	AGTGA**ATG**ATC
ORF 4	TAGGATC**ATG**GCA
ORF 5	GTGATAGAA**ATG**
ORF 6	AGTTAAAGTTAAAATTCGTAA**ATG**AAAGTTTT
ORF 7	AGGTGAAAAAGGTGCTATCA**ATG**

[1] from Olsen *et al.* 1986

[2] The initiation codon is shown in bold letters, possible Shine-Dalgarno sequences are underlined. No Shine-Dalgarno sequence was found in the region around the *ala S* gene and ORF 6.

In most cases, these sequences are located upstream of the initiation codon. However in the case of the Sso L10 and Sso L1 genes and ORF 3, the initiation codon itself seems to be part of a putative Shine-Dalgarno sequence.

Note that in the case of the "docking" protein gene, ORF 3 and 7, there are two Shine-Dalgarno sequences in tandem. This has also been observed in the case of the genes of the SSV1 virus that infects *Sulfolobus* (Reiter *et al.* 1987a). In the case of the *ala S* gene and ORF 6, sequences complementary to the 3' end of the 16S rRNA could not be found either upstream or downstream of the initiation codon. Note, however, that in the case of the *ala S* gene, the transcription initiation site is located just 3 nucleotides upstream of the initiation codon, so that the mRNA does not have a proper leader sequence. As mentioned in the Introduction, an interaction between the 3' end of the 16S rRNA and the mRNA has not yet been demonstrated in the archaebacteria.

Phylogenetic Implications

As mentioned in the Introduction, there has recently been a lot of controversy regarding the status of the archaebacteria as a separate kingdom. Lake (1988, 1989) has proposed that the extreme sulfur-dependent thermophiles, such as *Sulfolobus*, are more closely related to the eukaryotes, forming the superkingdom Karyota; while the extreme halophiles, the methanogens and the eubacteria form a separate group, the superkingdom Parkaryota. However, sequence comparison of the *Sulfolobus* ribosomal proteins with those from other organisms shows that *Sulfolobus* is more closely related to the other archaebacteria than to the members of the other two kingdoms. Thus the data presented in this dissertation, support the existence of the archaebacterial kingdom.

There is also a lot of controversy regarding the relationships among the three kingdoms. The 16S rRNA data seem to indicate that the three kingdoms

are separated from each other by an equal evolutionary distance (Woese 1987). However, comparison of the sequences of ribosomal proteins L10 and L12 from the three kingdoms shows that the archaebacterial sequences are closer to the eukaryotic sequences. A similar result has been obtained in the case of other ribosomal proteins (Auer *et al.* 1989a,b, Strobel *et al.* 1988, Köpke *et al.* 1989, Shimmin and Dennis 1989), the elongation factors (Lechner and Böck 1987, Lechner *et al.* 1988), the subunits of the RNA polymerase (Zillig *et al.* 1989b) and the H^+ ATPases (Nelson and Taiz 1989). In contrast, when the sequences of the glyceraldehyde 3-phosphate dehydrogenase are compared, the eukaryotic and eubacterial sequences seem to be closer to each other that to the archaebacterial sequences (Hensel *et al.* 1989, Fabry *et al.* 1989). At the moment, it is difficult to evaluate all of these contradictory results, particularly because in many cases only a few sequences have been analyzed. However, the fact that different molecules seem to give different relationships might mean that different genes have evolved at different rates in the three kingdoms, or it might indicate the existence of lateral gene transfer.

An additional explanation of these results has been proposed by Zillig *et al.* (1989a). According to their proposal, the reason why some eukaryotic genes are similar to eubacterial genes while others are similar to archaebacterial genes, is that the eukaryotic cell is the product of a fusion between an archaebacterium and a eubacterium. However, a problem with this model is the fact that the 16S rRNA data show the existence of three separate lines of descent. It is clear that we need more information in order to establish the relationships among the three kingdoms.

Evolution of the Ribosome

Since translation is the process that links the phenotype with the genotype, the origin and evolution of the translational apparatus is one of the key questions that must be answered in order to understand the origin of modern cells. Although we are far from having a complete picture of the evolution of the ribosome, some of the stages in this process are beginning to be defined. What I would like to do in this section is to discuss some of these stages in terms of the information provided by the archaebacterial ribosomal proteins, particularly those described in this thesis.

1]. *The 'first ribosomes' consisted only of RNA.* Woese (1980) initially proposed this idea in order to avoid the paradox of having proteins prior to the development of a mechanism to produce them. This proposal has received considerable support from the discovery of RNA catalysis (Cech and Brass 1986) and from the observation that rRNA plays a major functional role during protein synthesis (for a review, see Dalhberg 1989).

2]. *The 'first ribosomal proteins' were probably small peptides that interacted with rRNA.* It has been proposed that the original function of these peptides was to stabilize the structure of the rRNA (Maizels and Weiner 1987). Thus, ribosomal proteins that bind directly to rRNA, like the Sso L11, Sso L1 and Sso L10 proteins, probably contain domains that are directly derived from these 'ancient peptides'.

3]. *Modern ribosomal proteins were generated by gene duplication and by joining of functional domains* (Jue *et al.* 1980. Lin *et al.* 1987). Evidence of this can be seen in the modules of the archaebacterial L10 proteins, the common C-terminal domain of the archaebacterial and eukaryotic L10 and L12 proteins,

the conserved region between the L1 and L11 proteins and the common regions present in the L5 family of proteins which includes the Sso L46 protein.

4]. *Most of the evolution of the ribosome has been involved with the refinement of the mechanism of protein synthesis in order to obtain greater precision* (Woese 1980). This means that many of the ribosomal proteins present today were incorporated in order to increase the efficiency and precision of the ribosome and are thus, not essential for the translation process *per se*. Support for this comes from the isolation of mutants in *Escherichia coli*, that completely lack one ribosomal protein (for a review, see Dabbs 1986).

5]. *The 'ancestral ribosome' probably contained more ribosomal proteins than the modern eubacterial 70S ribosome* (Wool 1980, Cammarano *et al.* 1986, Auer *et al.* 1989a,b). This idea was initially proposed by Wool (1980). The discovery that there are two types of ribosomes within the archaebacteria, a small eubacterial-like ribosome with 53-54 proteins, present in the halophiles and most of the methanogens (Strøm and Visentin 1973, Schmid *et al.* 1982, Cammarano *et al.* 1986), and a larger ribosome, with 60-65 proteins, present in *Methanococcus* and the extreme thermophiles (Schmid *et al.* 1982, Schmid and Böck 1982, Cammarano *et al.* 1986), has provided new evidence to support this idea.

Since the ribosomes from *Thermococcus celer* (Cammarano *et al.* 1986), an organism that belongs to a branch that is very close to the root of the archaebacterial tree (Achenbach-Richter *et al.* 1987), are of the large type, it now seems likely that the larger ribosomes of *Methanococcus* and the extreme thermophiles, such as *Sulfolobus*, more closely resemble the 'ancestral ribosome' (Cammarano *et al.* 1986, Auer *et al.* 1989a, b).

This means that the smaller eubacterial and archaebacterial ribosomes are the product of a streamlining process, probably caused by the need for rapid growth (Wool 1980, Auer *et al.* 1989a, b). During this process, it seems likely that the functional domains of many proteins were combined, non-essential domains were eliminated and the number of proteins was reduced. Good examples of this streamlining process are the eubacterial L12 and L10 proteins, where complete domains have been modified or eliminated.

The presence of ribosomal proteins in the extreme halophiles, which have no equivalent in the eubacteria seems to indicate that the streamlining process took place independently in both lines and that this process is an example of evolutionary convergence.

In view of this scheme, the following relationships among the ribosomal proteins from the three kingdoms can be proposed:

1]. proteins conserved in the three kingdoms. Examples are the L12 and L10 proteins.

2]. proteins present only in the archaebacterial and eukaryotic ribosomes. Example Sso L46.

3]. proteins present only in the eubacterial and eukaryotic ribosomes.

4]. proteins only present in the eubacterial and archaebacterial ribosomes.

5]. unique eubacterial proteins.

6]. unique eukaryotic proteins.

7]. unique archaebacterial proteins.

Examples of the last 5 categories cannot be given at this time, since only a few archaebacterial and eukaryotic ribosomal proteins have been sequenced. Until the sequences of all the ribosomal proteins from a eukaryotic and the two

types of archaebacterial ribosomes are determined, the exact relationships among the different ribosomal proteins cannot be definitively established.

Evolution of the Genetic Organization in the Three Kingdoms

The gene organization of *Sulfolobus*, like that of the other archaebacteria, is very similar to that found in the eubacteria *i.e.* the genes are organized in transcriptional units or operons. Remarkably, in the case of the ribosomal protein genes, not only the organization but even the order of the genes is the same. However, the products of the transcription units and the promoters differ between the eubacteria and the archaebacteria. In contrast, in the eukaryotes, the ribosomal protein genes are not arranged in operons and are dispersed through the genome (Woolford and Rosbash 1981, Mager 1988). Based on the comparison of the organization of the ribosomal protein genes, a hypothetical model for the evolution of the gene organization of the three kingdoms can be proposed.

The progenote, or last common ancestor, is thought to have been an entity in which the mechanisms of transcription and translation were not fully developed (Woese 1987). For this reason, it is highly unlikely that its genes were already organized in operons. Furthermore, it is difficult to envision how once established, these transcriptional units could be broken, without serious consequences to the cell, to give rise to the eukaryotic type of organization. Therefore, it is more likely that in the progenote, each gene was transcribed independently. However, since the order of the genes is similar in eubacteria and archaebacteria, the genes must have already been organized in that order in the progenote. The reason why the genes are organized in a certain order is not clear. Since it is thought that most ribosomal proteins arose by the

duplication of a small group of "ancestral peptides" (Jue *et al.* 1980), the order might reflect a common origin or it might have some functional significance. Herold and Nierhaus (1987) have recently proposed that the order of the genes is related to the assembly process of the ribosome. They have found that in most cases (the only exceptions are L30 and L1) the proteins contained in a regulatory unit represent assembly domains. If the order of the genes is indeed related to the assembly process, this would explain why it has been conserved in the eubacteria and archaebacteria, where the synthesis of the ribosomal proteins and the assembly of the ribosome take place in the same compartment, while in the eukaryotes, where synthesis and assembly take place in different cell compartments (the cytoplasm and the nucleolus, respectively) (for a review, see Sommerville 1986) the order has not been maintained.

As the three kingdoms began to diverge, the pressures of having transcription and translation in the same cell compartment, and the need to obtain a more efficient and regulated synthesis of cell components, probably forced the eubacteria and archaebacteria to organize their genes into transcriptional units. Since this process probably took place independently in the eubacterial and archaebacterial kingdoms, the products of the transcriptional units are different. This explains why the location and the structure of the promoters is different in the eubacteria and archaebacteria. In the eukaryotes, on the other hand, there was no pressure to organize the ribosomal protein genes in transcriptional units, since transcription and translation take place in different cell compartments, and the order of the genes has not been conserved because, as stated above, the synthesis of ribosomal

proteins and the assembly of the ribosome take place in different cell compartments.

Conclusions

The results obtained from the study of the organization and transcription of the *Sulfolobus solfataricus* L11, L1, L10 and L12 ribosomal protein genes revealed that these genes are organized in an operon and that remarkably, the order of the genes is identical to that found in *Escherichia coli.* (Post *et al.* 1979). The conservation of the order of these four ribosomal protein genes in two organisms as widely separated, from an evolutionary point of view, as *Sulfolobus solfataricus* and *Escherichia coli.*, gives support to the proposal of Herold and Nierhaus (1987) that the order in which the ribosomal protein genes are organized in the chromosome is functionally important; defining different assembly domains.

From a genetic point of view, it was found that *Sulfolobus* has some features in common with the eubacteria and some with the eukaryotes. The organization of the ribosomal protein genes in operons, the use of rare initiation codons, the existence of overlapping stop/start codons and the presence of putative Shine-Dalgarno sequences are some of the eubacterial-like characteristics identified. In contrast, the nature and position of the promoters were found to resemble more closely the RNA polymerase II promoters in the eukaryotes as proposed by Zillig *et al.* (1988). From an evolutionary point of view, these results can be interpreted in terms of Woese's proposal (1987) that the last common ancestor of the three kingdoms or progenote was an entity in which the processes of transcription and translation

were still evolving (Woese 1987), giving rise to major differences in the organization and control of these genes in the three kingdoms.

Comparison of the sequences of the Sso L12 and Sso L10 ribosomal proteins with the homologous proteins from the other two kingdoms revealed that these two proteins have suffered extensive structural changes during the evolution of the three kingdoms. In general terms, two different structural groups can be identified: one comprising the archaebacterial and eukaryotic proteins and one comprising the eubacterial proteins. Since the function of these proteins has been conserved, it is now of interest to determine which regions of the proteins are carrying out the same functions in the two structural types. The fact that the genes have now been cloned will allow us to perform site-directed mutagenesis and determine which regions are functionally equivalent in the two types.

From a phylogenetic point of view, the comparison of the sequence of the ribosomal proteins from *Sulfolobus* with those from other organisms revealed that the *Sulfolobus* proteins are more closely related to the proteins from other archaebacteria. These results support the proposal that the archaebacteria are a monophyletic group (for a review, see Woese 1987). In terms of the relationships among the three kingdoms, the results obtained from the comparison of the Sso L12 and Sso L10 proteins indicate that the archaebacteria are closer to the eukaryotes. However, since information about the sequence of the eukaryotic L11 and L1 equivalents is still missing, it remains to be established if this will also be the case with these proteins.

Finally, the identification of ribosomal proteins in the archaebacteria, that have no counterparts in the eubacteria, raises several interesting questions. Where are these proteins located in the ribosome? What are their functions?

Do they have functions that have no counterpart in the eubacteria? Are other eubacterial ribosomal proteins or even domains performing the functions of these extra proteins? Why have these proteins been eliminated from the eubacteria or why have they been acquired by the archaebacteria? Future work in this area will certainly provide answers to all of these questions.

Bibliography

Achenbach-Richter, L., R. Gupta, W. Zillig, and C.R. Woese. 1988. Rooting the archaebacterial tree: the pivotal role of *Thermococcus celer* in archaebacterial evolution. System. Appl. Microbiol. **10:** 231-240.

Achenbach-Richter, L., K.O. Stetter, and C.R. Woese. 1987. A possible biochemical missing link among archaebacteria. Nature **327:** 348-349.

Aiba, H., S. Adhga, and B. de Crombrugghe. 1981. Evidence for two functional gal promoters in intact *Escherichia coli* cells. J. Biol. Chem. **256:** 11905-11910.

Allmansberger, R., M. Bokranz, L. Kröckel, J. Schallenberg, and A. Klein. 1989. Conserved gene structures and expression signals in methanogenic archaebacteria. Can. J. Microbiol. **35:** 52-57.

Allmansberger, R., C. Bollschweiler, U. Konheiser, B. Muller, E. Muth, G. Pasti, and A. Klein. 1986. Arrangement and expression of methyl CoM reductase genes in *Methanococcus voltae.* System. Appl. Microbiol. **7:** 13-17.

Amils, R., and J.L. Sanz. 1986. Inhibitors of protein synthesis as phylogenetic markers, p. 605-620. *In* B. Hardesty and G. Kramer (eds.), Structure, function and genetics of ribosomes. Springer-Verlag, New York.

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

Amons, R., W. Pluijms, J. Kriek, and W. Möller. 1982. The primary structure of protein eL12'/eL12'-P from the large subunit of *Artemia salina* ribosomes. FEBS Lett. **146:** 143-147.

Anderson, S. 1981. Shotgun DNA sequencing using cloned DNAse I-generated fragments. Nucleic Acids Res. **9:** 3015-3027.

Arndt, E. and M. Kimura. 1988. Molecular cloning and nucleotide sequence of the gene for the ribosomal protein S11 from the archaebacterium *Halobacterium marismortui.* J. Biol. Chem. **263:** 16063-16068.

Askoy, S., C.L. Squires, and C. Squires. 1984. Translational coupling of the *trp B* and *trp A* genes in *Escherichia coli* tryptophan operon. J. Bacteriol. **157:** 363-367.

Atkinson, T., and M. Smith. 1984. Solid-phase synthesis of oligodeoxyribonucleotides by the phosphite-triester method, p. 35-81. *In* M.J. Gait (ed.). Oligonucleotide Synthesis. IRL Press, Oxford.

Auer, J., K. Lechner, and A. Böck. 1989a. Gene organization and structure of two transcriptional units from *Methanococcus* coding for ribosomal proteins and elongation factors. Can. J. Microbiol. **35:** 200-204.

Auer, J., G. Spicker, and A. Böck. 1989b. Organization and structure of the *Methanococcus* transcriptional unit homologous to the *Escherichia coli* 'spectinomycin operon': implications for the evolutionary relationship of 70S and 80S ribosomes. J. Mol. Biol. **209:** 21-36.

Bayley, S.T., and D.J. Kushner. 1964. The ribosomes of the extremely halophilic bacterium *Halobacterium cutirubrum.* J. Mol. Biol. **9:** 654-669.

Bayley, S.T., and R.A. Morton. 1978. Recent developments in the molecular biology of halophiles. CRC Crit. Rev. Microbiol. **6:** 151-205.

Beauclerck, A.A.D., E. Cundliffe, and J. Dijk. 1984. The binding site for ribosomal protein complex L8 within 23S ribosomal RNA of *Escherichia coli.* J. Biol. Chem. **259:** 6559-6563.

Beauclerck, A.A.D., H. Hummel, D.J. Holmes, A. Böck, and E. Cundliffe. 1985. Studies of the GTPase domain of archaebacterial ribosomes. Eur. J. Biochem. **151:** 245-255.

Beckler, G.S., and J.N. Reeve. 1986. Conservation of primary structure in the *his I* gene of the archaebacterium *Methanococcus vannielii*, the eubacterium *Escherichia coli,* and the eukaryote *Saccharomyces cerevisiae.* Mol. Gen. Genet. **204:** 133-140.

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

Benton, W.D. , and R.W. Davies. 1977. Screening λgt recombinant clones by hybridization to single plaques *in situ.* Science **196:** 180-182.

Berghöfer, B., L. Kröckel, C. Körtner, M. Truss, J. Schallenberg, and A. Klein. 1988. Relatedness of archaebacterial RNA polymerase core subunits to their eubacterial and eukaryotic equivalents. Nucleic Acids Res. **16:** 8113-8128.

Bernabeu, C. , E. Tobin, A. Fowler, I. Zabin, and J.A. Lake. 1983. Nascent polypeptide chains exit the ribosome in the same relative position in both eukaryotes and prokaryotes. J.Cell. Biol. **96:** 1471-1474.

Bernstein , H.D., M.A. Poritz, K. Strub, P.J. Hoben, S. Brenner, and P. Walter. 1989. Model for signal sequence recognition from amino acid sequence of 54 K subunit of signal recognition particle. Nature **340:** 482-486.

Betlach, M., J. Friedmann, H.W. Boyer, and F. Pfeifer. 1984. Characterization of a halobacterial gene affecting bacterio-opsin gene expression. Nucleic Acids Res. **12:** 7949-7959.

Biosearch Model 8600 Instruction Manual. 1985. Biosearch Inc., San Rafael, California.

Birnboim, H.C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:** 1513-1523.

Blanck, A., and D. Oesterhelt. 1987. The halo-opsin gene II sequence, primary structure of halorhodopsin and comparison with bacteriorhodopsin. EMBO J. **6:** 265-273.

Blumberg, D.D. 1987. Creating a ribonuclease-free environment. Meth. in Enzymol. **152:** 20-24.

Böck, A., and O. Kandler. 1985. Antibiotic sensitivity of archaebacteria, p. 525-544. *In* C. R. Woese, and R.S. Wolfe (eds.), The Bacteria VIII: Archaebacteria. Academic Press, New York.

Bokranz, M., G. Bäumner, R. Allmansberger, D. Ankel-Fuchs, and A. Klein. 1988. Cloning and characterization of the methyl coenzyme M reductase genes from *Methanobacterium thermoautotrophicum.* J. Bacteriol. **170:** 568-577.

Bokranz, M., and A. Klein. 1987. Nucleotide sequence of the methyl coenzyme M reductase gene cluster from *Methanosarcina barkeri.* Nucleic Acids Res. **15:** 4350-4351.

Boublik, M., W. Hellman, and H.E. Roth. 1976. Localization of ribosomal protein L7/L12 in the 50S subunit of *Escherichia coli* ribosomes by electron microscopy. J. Mol. Biol. **107:** 479-490.

Bozzoni, I., E. Beccari, Z.X. Luo, and F. Amaldi. 1981. *Xenopus laevis* ribosomal protein genes: isolation of recombinant cDNA clones and study of the genomic organization. Nucleic Acids Res. **9:** 1069-1086.

Briat, J.F., S. Letoffe, R. Mache, and J. Rouviere-Yaniv. 1984. Similarity between the bacterial histone-like protein HU and a protein from spinach chloroplasts. FEBS Lett. **172:** 75-79.

Branlant, C., A. Krol, A. Machatt, and J.P. Ebel. 1981. The secondary structure of the protein L1 binding region of the ribosomal 23S RNA:

homologies with putative secondary structures of the L11 mRNA and of a region of mitochondrial 16S rRNA. Nucleic Acids Res. **9:** 293-307.

Brimacombe, R., J. Atmadja, A. Kyriatsoulis, and W. Stiege. 1986. RNA structure and RNA-protein neighborhoods in the ribosome, p. 184-202. *In* B. Hardesty and G. Kramer (eds.), Structure, function and genetics of ribosomes. Springer-Verlag, New York.

Brown, J.W., C.J. Daniels, and J.N. Reeve. 1989. Gene structure, organization and expression in archaebacteria. CRC Critical. Rev. Microbiol. **16:** 287-338.

Brown, J.W. and J.N. Reeve. 1985. Polyadenylated, non-capped RNA from the archaebacterium *Methanococcus vannielii.* J. Bacteriol. **162:** 909-917.

Brown, J. W., and J.N. Reeve. 1986. Polyadenylated RNA isolated from the archaebacterium *Halobacterium halobium*. J. Bacteriol. **166:** 686-688.

Brown, J.W., M. Thomm, G.S. Beckler, G. Frey, K.O. Stetter, and J. N. Reeve. 1988a. An archaebacterial RNA polymerase binding site and transcription initiation of the *his A* gene in *Methanococcus vannielii.* Nucleic Acids Res. 16: 135-150.

Brown, S.J., D.D. Rhoads, M.J. Stewart, B. Van Slyke, T.K. Johnson, R.E. Denell, and D.J. Roufa. 1988b. Ribosomal protein S14 is encoded by a pair of highly conserved adjacent genes on the X chromosome of *Drosophila melanogaster.* Mol. Cell. Biol. **8:** 4314-4321.

Bushuev, V.N., A.T. Gudkov, A. Liljas, and N.F. Sepetov. 1989. The flexible region of protein L12 from bacterial ribosomes studied by proton nuclear magnetic resonance. J. Biol. Chem. **264:** 4498-4505.

Bushuev, V.N., N.F. Sepetov, and A.T. Gudkov. 1984. Symmetrical structure of the L7 protein dimer FEBS Lett. **178:** 101-104.

Buttarelli, F.R., R.A. Calogeno, O. Tiboni, C.O. Gualerzi, and C.L. Pon. 1989. Characterization of the *str* operon genes from *Spirulina platensis* and their evolutionary relationship to those of other prokaryotes. Mol. Gen. Genet. **217:** 97-104.

Byström, A.S., K.J. Hjalmarsson, P.M. Wikström, and G.R. Björk. 1983. The nucleotide sequence of an *Escherichia coli* operon containing genes for the tRNA (m^1G) methyltransferase, the ribosomal proteins S16 and L19 and a 21 K polypeptide. EMBO J. **2:** 899-905.

Cammarano, P., A. Teichner, and P. Londei. 1986. Intralineage heterogeneity of archaebacterial ribosomes, evidence for two

physicochemically distinct ribosome classes within the third urkingdom. System. Appl. Microbiol. **7:** 137-146.

Cammarano, P., A. Teichner, P. Londei, M. Acca, B. Nicolaus, J.L. Sanz, and R. Amils. 1985. Insensitivity of archaebacterial ribosomes to protein synthesis inhibitors. Evolutionary implications. EMBO J. **4:** 811-816.

Cannon, M., D. Schindler, and J. Davies. 1977. Methylation of proteins in 60S ribosomal subunits from *Saccharomyces cerevisiae*. FEBS Lett. **75:** 187-191.

Capel, M.S., D.M. Engelman, B.R. Freeborn, M. Kjeldgaard, J.A. Langer, V. Ramakrishnan, D.G. Schindler, D.K. Schneider, B.P. Schoenborn, I.Y. Sillers, S. Yabuki, and P.B. Moore. 1987. A complete mapping of the proteins in the small ribosomal subunit of *Escherichia coli*. Science **238:** 1403-1406.

Cech, T.R., and B.L. Bass. 1986. Biological catalysis of RNA. Ann. Rev. Biochem. **55:** 599-626.

Cedegren, R., M.W. Gray, I. Abel, and D. Sankoff. 1988. The evolutionary relationships among known life forms. J. Mol. Evol. **28:** 98-112.

Chambliss,G., G.R. Craven, J. Davies, K. Davis, L. Kahan, and M. Nomura (eds.). 1980. Ribosomes: structure function and genetics. University Park Press, Baltimore.

Chan, Y.L., A. Lin, J. McNally, and I.G. Wool. 1987. The primary structure of rat ribosomal protein L5: a comparison of the sequence of amino acids in the proteins that interact with 5S RNA. J. Biol. Chem. **262:** 12879-12886.

Chant, J., and P.P. Dennis. 1986. Archaebacteria: transcription and processing of ribosomal RNA sequences in *Halobacterium cutirubrum.* EMBO J. **5:** 1091-1097.

Chant, J., I. Hui, D. DeJong-Wong, L.C. Shimmin, and P.P. Dennis. 1986. The protein synthesizing machinery of the archaebacterium *Halobacterium cutirubrum*: molecular characterization. System. Appl. Microbiol. **7:** 106-114.

Chartier, F., B. Laine, P. Sautiere, J.P. Touzel, and G. Albagnac. 1985. Characterization of the chromosomal protein HMb isolated from *Methanosarcina barkeri*. FEBS Lett. **183:** 119-123.

Chen, I.T., and D.J. Roufa. 1988. The transcriptionally active human ribosomal protein S17 gene. Gene **70:** 107-116.

Choli, T., B. Wittmann-Liebold, and R. Reinhardt. 1988. Microsequence analysis of DNA-binding proteins 7a, 7b, and 7e from the archaebacterium *Sulfolobus acidocaldarius*. J. Biol. Chem. **263:** 7087-7093.

Christensen, T., M. Johnsen, N.P. Fill, and J.D. Friesen. 1984. RNA secondary structure and translation inhibition: analysis of mutants in the *rpl J* leader. EMBO J. **3:** 1609-1612.

Connolly, T., and R. Gilmore. 1986. Formation of a functional ribosome-membrane junction during translocation requires the participation of a GTP-binding protein. J.Cell Biol. **103:** 2253-2261.

Connolly, T., and R. Gilmore. 1989. The signal recognition particle receptor mediates the GTP-dependent displacement of SRP from the signal sequence of the nascent polypeptide. Cell **57:** 599-610.

Cooperman, B.S. 1980. Functional sites on the *Escherichia coli* ribosome as defined by affinity labeling, p. 531-554. *In* G. Chambliss, G.R. Craven, J. Davies, K. Davis, L. Kahan, and M. Nomura (eds.), Ribosomes: structure function and genetics. University Park Press, Baltimore.

Cowgill, C.A., B.G. Nichols, J.W. Kenny, P. Butler, E.M. Bradbury, and R.R. Traut. 1984. Mobile domains in ribosomes revealed by proton nuclear magnetic resonance. J. Biol. Chem. **259:** 15257-15263.

Cram, D.S., B.A. Sherf, R.T. Libby, R.J. Mattaliano, K.L. Ramachandran, and J.N. Reeve. 1987. Structure and expression of the genes *mcr BDCGA*, which encode subunits of component C of methyl coenzyme M reductase in *Methanococcus vannielii*. Proc. Natl. Acad. Sci. U.S.A. **84:** 3992-3996.

Cue, D., G.S. Beckler, J.N. Reeve, and J. Konisky. 1985. Structure and sequence divergence of two archaebacterial genes. Proc. Natl. Acad. Sci. U.S.A. **82:** 4207-4211.

Culham, D.E., and R.E. Nazar. 1988. An unlinked 5S ribosomal RNA gene in the archaebacterium *Thermococcus celer*. Mol. Gen. Genet. **212:** 382-385.

Cundliffe, E. 1986. Involvement of specific portions of ribosomal RNA in defined ribosomal functions: a study utilizing antibiotics, p. 586-604. *In* B. Hardesty and G. Kramer (eds.), Structure, function and genetics of ribosomes. Springer-Verlag, New York.

Dahlberg, A.E. 1989. The functional role of ribosomal RNA in protein synthesis. Cell **55:** 525-529.

Dams, E., L. Hendricks, Y. Van de Peer, J.M. Neefs, G. Smits. I. Vandenbempt, and R. De Wachter. 1988. Compilation of small ribosomal RNA sequences. Nucleic Acids Res. **16:** r87-r137.

Daniels, C.J., R. Gupta, and W.F. Doolittle. 1985. Transcription and excision of a large intron in the tRNA trp gene of an archaebacterium, *Halobacterium volcanii.* J. Biol. Chem. **260:** 3132-3134.

DasSarma, S., U. RajBhandary, and H.S. Khorana. 1984. Bacterio-opsin mRNA in wild type and bacterio-opsin deficient *Halobacterium halobium* strains. Proc. Natl. Acad. Sci. U.S.A. 81: 125-129.

Datta, P.K., L.K. Hawkins, and R. Gupta. 1989. Presence of an intron in elongator methionine-tRNA of *Halobacterium volcanii.* Can. J. Microbiol. **35:** 189-194.

Davies. R.W. 1982. DNA sequencing, p. 117-172. *In* D. Rickwood, and B.D. Hames (eds.), Gel electrophoresis of nucleic acids. IRL Press, Oxford.

De Lange, R.J., L.C. Williams, and D.G. Searcy. 1981. A histone-like protein (HTa) from *Thermoplasma acidophilum* II. Complete amino acid sequence. J. Biol. Chem. **256:** 905-911.

Denda, K., J. Konishi, T. Oshima, T. Date, and M. Yoshida. 1989. A gene encoding the proteolipid subunit of *Sulfolobus acidocaldarius* ATPase complex. J. Biol. Chem. **264:** 7119-7121.

Denda, K., J. Konishi, T. Oshima, T. Date, and M. Yoshida. 1988a. The membrane-associated ATPase from *Sulfolobus acidocaldarius* is distantly related to F_1-ATPase as assessed from the primary structure of its α subunit. J. Biol. Chem. **263:** 6012-6015.

Denda, K., J. Konishi, T. Oshima, T. Date, and M. Yoshida. 1988b. Molecular cloning of the β-subunit of a possible non-F_0F_1 type ATP synthetase from the acidothermophilic archaebacterium, *Sulfolobus acidocaldarius.* J. Biol. Chem. **263:** 17251-17254.

D'Eustachio, P., O. Meyuhas, F. Ruddle, and R.P. Perry. 1981. Chromosomal distribution of ribosomal protein genes in the mouse. Cell **24:** 307-312.

Dever, T.E., M.J. Glynias, and W.C. Merrick. 1987. GTP-binding proteins: three consensus sequence elements with distinct spacing. Proc. Natl. Acad. Sci. U.S.A. **84:** 1814-1818.

Dignam, S.S. and J.D. Dignam. 1984. Glycyl- and alanyl-tRNA synthetases from *Bombyx mori*: purification and properties. J. Biol. Chem. **259:** 4043-4048.

Dijk, J., R.A. Garrett, and R. Müller. 1979. Studies on the binding of the ribosomal protein complex L7/L12 - L10 and protein L11 to the 5' one third of the 23S RNA: a functional center of the 50S subunit. Nucleic Acids Res. **6:** 2717-2729.

Dilella, A.G., and S.L.C. Woo. 1987. Hybridization of genomic DNA to oligonucleotide probes in the presence of tetramethylammonium chloride. Meth. in Enzymol. **152:** 447-451.

Dognin, M.J. and B. Wittmann-Liebold. 1977. The primary structure of L11, the most heavily methylated protein from *Escherichia coli* ribosomes. FEBS Lett. **[illegible]:** 342-346.

Doolittle, W.F. 1985. Genome Structure in Archaebacteria, p. 545-560. *In* C. R. Woese, and R.S. Wolfe (eds.), The Bacteria VIII: Archaebacteria. Academic Press, New York

Dove, W.F., and N. Davidson. 1962. Cation effects on the denaturation of DNA. J. Mol. Biol. **5:** 467-478.

Downing, W., and P. Dennis. 1987. Transcription products from the *rpl* KAJL-*rpo* BC gene cluster. J. Mol. Biol. **194:** 609-620.

Dudov, K.P., and R.P. Perry. 1984. The gene encoding the mouse ribosomal protein L32 contains a uniquely expressed intron-containing gene and an unmutated processed gene. Cell **37:** 457-468.

Dunn, R., J. McCoy, M. Simsek, A. Majumdar, S.H. Chang, and U.L. RajBhandary. 1981. The bacteriorhodopsin gene. Proc. Natl. Acad. Sci. U.S.A. **78:** 6744-6748.

Egebjerg, J., S. Douthwaite, and R.A. Garrett. 1989. Antibiotic interactions at the GTPase-associated center within *Escherichia coli* 23S rRNA. EMBO J. **8:** 607-611.

Eisenberg, H., and E.J. Wachtel. 1987. Structural studies of halophilic proteins, ribosomes and organelles of bacteria adapted to extreme salt concentrations. Ann. Rev. Biophys. Biophys. Chem. **16:** 69-92.

El-Baradi, T.T.A.L., V.C.H.F. de Regt, S.W.C. Einerhand, J. Teixido, R.J. Planta, J.P.G. Ballesta, and H.A. Raue. 1987. Ribosomal proteins EL11 from *Escherichia coli* and L15 from *Saccharomyces cerevisiae* bind to the same site on both yeast 26S and mouse 28S rRNA. J. Mol. Biol. **195:** 909-917.

Erdmann, V.A., T. Pieler, J. Wolters, M. Digweed, D. Vogel, and R. Hartmann. 1986. Comparative structural and functional studies on small ribosomal RNAs, p. 164-183. *In* B. Hardesty and G. Kramer (eds.), Structure, function and genetics of ribosomes. Springer-Verlag, New York.

Fabijanski, S. and M. Pellegrini. 1982. Isolation of a cloned DNA segment containing a ribosomal protein gene of *Drosophila melanogaster.* Gene **18:** 267-276.

F[illegible]ry, S., J. Lang, T. Niermann, M. Vingron, and R. Hensel. 1989. Nucleotide sequence of the glyceraldehyde-3 phosphate dehydrogenase gene from the mesophilic methanogenic archaebacteria *Methanobacterium bryantii* and *Methanobacterium formi[illegible]icum.* Eur. J. Biochem. **1[illegible]9:** 405-413.

Ferro, J. A., and F.C. Reinach. 1988. The complete sequence of a chicken- muscle cDNA encoding the acidic ribosomal protein P1. Eur. J. Biochem. **177:** 513-516.

Fourney, R.M., J. Miyakoshi, R.S. Day III, and M. C. Paterson. 198[illegible]. Northern Blotting: efficient RNA staining and transfer. Focus **10:** 5-7.

Fox, G.E. 198[illegible]. The structure and evolution of archaebacterial ribosomal RNA, p.257-[illegible]. *In* C. R. Woese, and R.S. Wolfe (eds.), The Bacteria VIII: Archaebacteria. Academic Press, New York.

Fried, H.M., N.J. Pearson, C.H. Kim, and J.R. Warner. 1981. The genes for fifteen ribosomal proteins of *Saccharomyces cerevisiae*. J. Biol. Chem. **256:** 10176-10183.

Frischauf, A. 1987. Construction and characterization of a genomic library in λ. Meth. in Enzymol. **152:** 190-199.

Frischauf, A., H. Lehrach, A. Poustka, and N. Murray. 1983. Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. **170:** 827-842.

Friesen, J.D., N.P. Fiil, J.M. Parker, and W.A. Haseltine. 1974. A new relaxed mutant of *Escherichia coli* with an altered 50S ribosomal subunit. Proc. Natl. Acad. Sci. U.S.A. **71:** 3465-3469.

Garland, W.G., K.A. Louie, A.T. Matheson, and A. Liljas. 1987. The complete amino acid sequence of the ribosomal 'A' protein (L12) from *Bacillus stearothermophilus*. FEBS Lett. **220:** 43-46.

Gatenby, A.A., S.J. Rothstein, and M. Nomura. 1989. Translational coupling of the maize chloroplast *atp B* and *atp E* genes. Proc. Natl. Acad. Sci. U.S.A. **86:** 4066-4070.

Geliebter, J. 1987. Dideoxynucleotide sequencing of RNA and uncloned cDNA. Focus **9:** 5-8.

Gill, D.R., G.F. Hatfull and G.P.C. Salmond. 1986. A new cell division operon in *Escherichia coli.* Mol. Gen. Genet. **205:** 134-145.

Gilmore, R., G. Blobel, and P. Walter. 1982a. Protein translocation across the endoplasmic reticulum I. Detection in the microsomal membrane of a receptor for the signal recognition particle. J. Cell Biol. **95:** 463-499.

Gilmore, R., G. Blobel, and P. Walter. 1982b. Protein translocation across the endoplasmic reticulum. II. Isolation and characterization of the signal recognition particle receptor. J. Cell Biol. **95:** 470-472.

Giri, L., J. Dijk, H. Labischinski, and H. Bradaczek. 1978. Shape of protein L11 from the 50S ribosomal subunit of *Escherichia coli.* Biochemistry **17:** 745-749.

Giri, L., W.E. Hill, and H.G. Wittmann. 1984. Ribosomal proteins: their structure and spacial arrangement in prokaryotic ribosomes. Adv. Prot. Chem. **36:** 1-78.

Girshovich, A.S., T.V. Kurtskhalia, Y.A. Guchinnikov, and V.D. Vasiliev. 1981. Localization of the elongation factor G on *Escherichia coli* ribosome. FEBS Lett. **130:** 54-59.

Gluzman, Y. (ed.) 1985. Eukaryotic transcription. Cold Spring Harbor Laboratory, Cold Spring Harbor.

Gold, L. 1988. Post-transcriptional regulatory mechanisms in *Escherichia coli.* Ann. Rev. Biochem. **57:** 199-233.

Gopalakrisna, Y., D. Langley, and N. Sarkar. 1981. Detection of high levels of polyadenylate-containing RNA in bacteria by the use of a single-step RNA isolation procedure. Nucleic Acids Res. **9:** 3345-3354.

Gornicki, P., K. Nurse, W. Hellmann, M. Boublik, and J. Ofengand. 1984. High resolution localization of the tRNA anticodon interaction on *Escherichia coli* 30S ribosomal subunit. J. Biol. Chem. **259:** 10493-10498.

Gourse, R.L., D.L. Thurlow, S.A. Gerbi, and R.A. Zimmermann. 1981. Specific binding of a prokaryotic ribosomal protein to a eukaryotic ribosomal RNA: implications for evolution and autoregulation. Proc. Natl. Acad. Sci. U.S.A. **78:** 2722-2726.

Gouy, M. and W.H. Li. 1989. Phylogenetic analysis based on rRNA sequences supports the archaebacterial rather than the eocyte tree. Nature **339:** 145-147.

Gross, U., J.H. Chen, D.H. Kono, J.G. Lobo, and D.Ty. Yu. 1989. High degree of conservation between ribosomal proteins of *Yersenia pseudotuberculosis* and *Escherichia coli.* Nucleic Acids Res. **17:** 3601-3602.

Gudkov, A.T., and J. Behlke. 1978. The N-terminal sequence of protein L7/L12 is responsible for its dimerization. Eur. J. Biochem. **90:** 309-312.

Gudkov, A.T., J. Behlke, N.N. Utiurin, and V.I. Lim. 1977. Tertiary and quaternary structure for ribosomal protein L7 in solution. FEBS Lett. **82:** 125-129.

Gudkov, A.T., and G.M. Gongadze. 1984. The L7/L12 proteins change their conformation upon interaction of EF-G with ribosomes. FEBS Lett. **176:** 32-36.

Gudkov, A.T., G.M. Gongadze, V.N. Bushuev, and M.S. Okon. 1982. Proton nuclear magnetic resonance study of the ribosomal protein L7/L12 *in situ.* FEBS Lett. **138:** 229-232.

Gudkov, A.T., N.N. Khechinashvili, V.N. Bushuev. 1978. Studies on the structure of protein L7/L12 from *Escherichia coli* ribosomes. Eur. J. Biochem. **90:** 313-318.

Gudkov, A.T., L.G. Tumanova, G.M. Gongadze, and V.N. Bushuev. 1980. Role of different regions of ribosomal proteins L7 and L10 in their complex formation and in the interaction with the ribosomal 50S subunit. FEBS Lett. **109:** 34-38.

Gudkov, A.T., S.Y. Venyaminov, R. Amons, W. Möller, and T. Itoh. 1981. A physical study of acidic ribosomal proteins from *Artemia salina, Bacillus subtilis* and *Micrococcus lysodeikticus.* FEBS Lett. **136:** 235-238.

Gupta, R. 1985. Transfer ribonucleic acids of archaebacteria, p. 311-343. *In* C. R. Woese, and R.S. Wolfe (eds.), The Bacteria VIII: Archaebacteria. Academic Press, New York.

Gupta, R., J.M. Lanter, and C.R. Woese. 1983. Sequence of the 16S ribosomal RNA from *Halobacterium volcanii*, an archaebacterium. Science **221:** 656-659.

Gutell, R.R., and G.E. Fox. 1988. A compilation of large subunit RNA sequences presented in structural format. Nucleic Acids Res. **16:** r175-r269.

Gutell, R.R., B. Weiser, C.R. Woese, and H.F. Noller. 1985. Comparative anatomy of 16S-like ribosomal RNA. Prog. in Nucleic Acid Res. and Mol. Biol. **32:** 155-216.

Hamel, E., M. Koka, and T. Nakamoto. 1972. Requirement of an *Escherichia coli* 50S ribosomal protein for effective interaction of the ribosome with T and G factors and guanosine triphosphate. J. Biol. Chem. **247:** 805-814.

Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166:** 557-580.

Harauz, G., M. Stöffler-Meilicke, and M. Van Heel. 1987. Characteristic views of prokaryotic 50S ribosomal subunits. J. Mol. Evol. **26:** 347-357.

Hardesty, B. and G. Kramer (eds.). 1986. Structure, function, and genetics of ribosomes. Springer-Verlag, Berlin.

Hardy, S.J.S. 1975. The stoichiometry of the ribosomal proteins of *Escherichia coli.* Mol. Gen. Genet. **140:** 253-274.

Hartmann, R.K., and V.A. Erdmann. 1989. *Thermus thermophilus* 16S rRNA is transcribed from an isolated transcription unit. J. Bacteriol. **171:** 2933-2941.

Hattori, M., and Y. Sakaki. 1986. Dideoxy-sequencing method using denatured plasmid templates. Anal. Biochem. **152:** 232-238.

Heimark, R.L., J.W.B. Hershey, and R.R. Traut. 1976. Cross-linking of initiation factor IF2 to proteins L7/L12 in 70S ribosomes of *Escherichia coli.* J. Biol. Chem. **251:** 7779-7784.

Heinrich, P. 1986. Guidelines for quick and simple plasmid sequencing. Boehringer Mannheim GmbH Biochemica, Mannheim.

Hensel, R., P. Zwickl, S. Fabry, J. Lang and P. Palm. 1989. Sequence comparison of glyceraldehyde 3-phosphate dehydrogenases from the three urkingdoms: evolutionary implications. Can. J. Microbiol. **35:** 81-85.

Hortsch, M., S. Labeit, and D.I. Meyer. 1988. Complete cDNA sequence coding for human docking protein. Nucleic Acids Res. **16:** 361-362.

Hui, I., and P. Dennis. 1985. Characterization of rRNA gene clusters in *Halobacterium cutirubrum.* J. Biol. Chem. **260:** 899-906.

Hummel, H., U. Bar, G. Heller, and A. Böck. 1985. Antibiotic sensitivity pattern of *in vitro* polypeptide synthesis from *Methanosarcina barkeri* and *Methanospirillum hungatei.* System. Appl. Microbiol. **6:** 125-131.

Itoh, T. 1988. Complete nucleotide sequence of the ribosomal 'A' protein operon from the archaebacterium *Halobacterium halobium.* Eur. J. Biochem. **176:** 297-303.

Itoh, T. 1981. Primary structure of an acidic ribosomal protein from *Micrococcus lysodeikticus*. FEBS Lett. **127:** 67-70.

Itoh, T., and K.I. Higo. 1983. Complete amino acid sequence of an L7/L12 - type ribosomal protein from *Rhodopseudomonas spheroides*. Biochim. Biophys. Acta **744:** 105-109.

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

Itoh, T., and E. Otaka. 1984. Complete amino acid sequence of an L7/L12-type ribosomal protein from *Desulfovibrio vulgaris* Miyazaki. Biochim. Biophys. Acta **789:** 229-233.

Itoh, T., M. Sugiyama, and K.I. Higo. 1982. The primary structure of an acidic ribosomal protein from *Streptomyces griseus*. Biochim. Biophys. Acta **701:** 164-172.

Itoh, T., and B. Wittmann-Liebold. 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.

Jacob, S.T. 1986. Transcription of eukaryotic ribosomal RNA gene. Mol. and Cell. Biochem. **70:** 11-20.

Jacq, B. 1981. Sequence homologies between eukaryotic 5.8S rRNA and the 5' end of prokaryotic 23S rRNA: evidence for a common evolutionary origin. Nucleic Acids Res. **9:** 2913-2932.

Jarsch, M., J. Altenbuchner, and A. Böck. 1983. Physical organization of the genes for ribosomal RNA in *Methanococcus vannielii*. Mol. Gen. Genet. **189:** 41-47.

Jarsch, M., and A. Böck. 1985a. Sequence of the 16S ribosomal RNA gene from *Methanococcus vannielii:* evolutionary implications. System. Appl. Microbiol. **6:** 54-59.

Jarsch, M., and A. Böck. 1985b. Sequence of the 23S rRNA gene from the archaebacterium *Methanococcus vannielii*: evolutionary and functional implications. Mol. Gen. Genet. **200:** 305-312.

Jasin, M., L. Regan, and P. Schimmel. 1983. Modular arrangement of functional domains along the sequence of an aminoacyl tRNA synthetase. Nature **306:** 441-447.

Jessee, J. 1984. Use of competent cells for DNA transformations. Focus **6:** 5-6.

Jones, W.J., D.P. Nagle, and W.B. Whitman. 1987. Methanogens and the diversity of archaebacteria. Microbiol. Rev. **51:** 135-177.

Juan-Vidales, F., M.T. Saenz-Robles, and J.P.G. Ballesta. 1984. Acidic proteins of the large ribosomal subunit in *Saccharomyces cerevisiae.* Effect of phosphorylation. Biochemistry **23:** 390-396.

Jue, R.A., N.W. Woodbury, and R.F. Doolittle. 1980. Sequence homologies among *Escherichia coli* ribosomal proteins: evidence for evolutionarily related groupings and internal duplications. J. Mol. Evol. **15:** 129-148.

Kaine, B.P. 1987. Intron-containing tRNA genes of *Sulfolobus solfataricus.* J. Mol. Evol. **25:** 248-254.

Kaine, B.P., R. Gupta, and C.R. Woese. 1983. Putative introns in tRNA genes of prokaryotes. Proc. Natl. Acad. Sci. U.S.A. **80:** 3309-3312.

Kaine, B.P., and U.L. Merkel. 1989. Isolation and characterization of the 7S RNA gene from *Methanococcus voltae.* J. Bacteriol. **171:** 4261-4266.

Kaltschmidt, E., and H.G. Wittmann. 1970. Ribosomal proteins VII. Two-dimensional polyacrylamide gel electrophoresis for finger printing of ribosomal proteins. Anal. Biochem. **36:** 401-412.

Karn, J., S. Brenner, and L. Barnett. 1983. New bacteriophage lambda vectors with positive selection for cloned inserts. Meth. in Enzymol. **101:** 3-19.

Karn, J., S. Brenner, L. Barnett, and G. Cesareni. 1980. Novel bacteriophage λ cloning vector. Proc. Natl. Acad. Sci. U.S.A. **77:** 5172-5176.

Kearney, K.R., and M. Nomura. 1987. Secondary structure of the autoregulatory mRNA binding site of ribosomal protein L1. Mol. Gen. Genet. **210:** 60-68.

Kessel, M., and F. Klink. 1980. Archaebacterial elongation factor is ADP-ribosylated by diphtheria toxin. Nature 287: 250-251.

Kimura, M., E. Arndt, T. Hatakeyama, T. Hatakeyama, and J. Kimura. 1989. Ribosomal proteins in halobacteria. Can. J. Microbiol. **35:** 195-199.

Kimura, M., J. Kimura, and K. Ashman. 1985. The complete primary structure of ribosomal proteins L1, L14, L15, L23, L24 and L29 from *Bacillus stearothermophilus.* Eur. J. Biochem. **150:** 491-497.

Kirsebom, L.A., R. Amons, and L.A. Isaksson. 1986. Primary structures of mutationally altered ribosomal protein L7/L12 and their effects on cellular growth and translational accuracy. Eur. J. Biochem. **156:** 669-675.

Kirsebom, L.A., and L.A. Isaksson. 1985. Involvement of ribosomal protein L7/L12 in control of translational accuracy. Proc. Natl. Acad. Sci. U.S.A. **82:** 717-721.

Kjems, J., and R.A. Garrett. 1985. An intron in the 23S ribosomal RNA gene of the archaebacterium *Desulfurococcus mobilis*. Nature **318:** 675-677.

Kjems, J., and R.A. Garrett. 1987. Novel expression of the ribosomal RNA genes in the extreme thermophile and archaebacterium *Desulfurococcus mobilis*. EMBO J. **6:** 3521-3530.

Kjems, J., R.A. Garrett, and W. Ansorge. 1987a. The sequence of the 16S RNA gene and its flanking region from the archaebacterium *Desulfurococcus mobilis*. System. Appl. Microbiol. **9:** 22-28.

Kjems, J., H. Leffers, R.A. Garrett, G. Wich, W. Leifelder, and A. Böck. 1987b. Gene organization, transcription signals and processing of the single ribosomal RNA operon of the archaebacterium *Thermoproteus tenax*. Nucleic Acids Res. **15:** 4821-4835.

Klink, F. 1985. Elongation factors, p. 379-340. *In* C. R. Woese, and R.S. Wolfe (eds.), The Bacteria VIII: Archaebacteria. Academic Press, New York.

Köpke, A.K.E., G. Baier, and B. Wittmann-Liebold. 1989. An archaebacterial gene from *Methanococcus vannielii* encoding a protein homologous to the ribosomal protein L10 family. FEBS Lett. **247:** 167-172.

Köpke, A.K.E., and B. Wittmann-Liebold. 1989. Comparative studies of ribosomal proteins and their genes from *Methanococcus vannielii* and other organisms. Can. J. Microbiol. **35:** 11-20.

Koteliasky, V., S.P. Domogatsky, and A.T. Gudkov. 1978. Dimer state of protein L7/L12 and EF-G dependent reactions on ribosomes. Eur. J. Biochem. **90:** 319-323.

Kozak, M. 1983. Comparison of initiation of protein synthesis in prokaryotes, eukaryotes and organelles. Microbiol Rev. **47:** 1-45.

Krieg, P., and D. Melton. 1985. Plasmid preparation, p. DB3. *In* Promega Catalogue. Promega Biotec Molecular Biologicals, Wisconsin.

Krowczynska, A.M., M. Coutts, S. Makrides, and G. Brawerman. 1989. The mouse homologue of the human acidic ribosomal phosphoprotein

P0: a highly conserved polypeptide that is under translational control. Nucleic Acids Res. **15:** 6408.

Kuchino, Y., M. Ihara, Y. Yabusaki, and S. Nishimura. 1982. Initiator tRNAs from archaebacteria show common unique sequence characteristics. Nature **298:** 684-685.

Kurzchalia, T.V., M. Wiedmann, A.S. Girshovich, E.S. Bochkareva, H. Bielka, and T.A. Rapoport. 1986. The signal sequence of nascent preprolactin interacts with the 54 KD polypeptide of the signal recognition particle. Nature **320:** 634-636.

Kushner, D.J. 1985. The halobacteriaceae, p. 171-214. *In* C. R. Woese, and R.S. Wolfe (eds.), The Bacteria VIII: Archaebacteria. Academic Press, New York.

Kuzumaki, T., T. Tanaka, K. Ishikawa, and K. Ogata. 1987. Rat ribosomal protein L35a multigene family: molecular structure and characterization of three L35a-related pseudogenes. Biochim. Biophys. Acta **909:** 99-106.

Lake, J.A. 1985. Evolving ribosome structure: domains in archaebacteria, eubacteria, eocytes and eukaryotes. Ann. Rev. Biochem. **54:** 507-530.

Lake, J.A. 1988. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. Nature **331:** 184-188.

Lake, J.A. 1989. Origin of the eukaryotic nucleus: eukaryotes and eocytes are genotypically related. Can. J. Microbiol. **35:** 109-118.

Lake, J.A., M.W. Clark, E. Henderson, S. Fay, M. Oakes, A. Scheiman, P. Thornber, and R.A. Mah. 1985. Eubacteria, halobacteria, and the origin of photosynthesis: the photocytes. Proc. Natl. Acad. Sci. U.S.A. **82:** 3716-3720.

Lake, J.A., E. Henderson, M.W. Clark , A. Scheinman, and M. Oakes. 1986. Mapping evolution with three dimensional ribosome structure. System. Appl. Microbiol. **7:** 131-136.

Lake, J.A., E. Henderson, M. Oakes, and M.W. Clark. 1984. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. Proc. Natl. Acad. Sci. U.S.A. **81:** 3786-3790.

Lake, J.A., and W.A. Strycharz. 1981. Ribosomal proteins L1, L17, L27, from *Escherichia coli* localized at single sites on the large subunit by immune electron microscopy. J. Mol. Biol. **153:** 979-992.

Larkin, J.C., J.R. Thompson and J.L. Woolford, Jr. 1987. Structure and expression of the *Saccharomyces cerevisiae CRY1* gene: a highly conserved ribosomal protein gene. Mol. Cell. Biol. **7:** 1764-1775.

Larkin, J.C., and J.L. Woolford, Jr. 1983. Molecular cloning and analysis of the *CRY1* gene : a yeast ribosomal protein gene. Nucleic Acids Res. **11:** 403-420.

Lathe, R. 1985. Synthetic oligonucleotide probes deduced from amino acid sequence data: theoretical and practical considerations. J. Mol. Biol. **183:** 1-12.

Lauffer, L., P.D. Garcia, R.N. Harkins, L. Coussens, A. Ullrich, and P. Walter. 1985. Topology of signal recognition particle receptor in endoplasmic reticulum membrane. Nature **318:** 334-338.

Lechner, K., and A. Böck. 1987. Cloning and nucleotide sequence of the gene for an archaebacterial protein synthesis elongation factor Tu. Mol. Gen. Genet. **208:** 523-528.

Lechner, K., G. Heller, and A. Böck. 1988. Gene for the diphtheria toxin-susceptible elongation factor 2 from *Methanococcus vannielii*. Nucleic Acids Res. **16:** 7817-7826.

Lechner, K., G. Wich, and A. Böck. 1985. The nucleotide sequence of the 16S rRNA gene and flanking regions from *Methanobacterium formicicum*: the phylogenetic relationship between methanogenic and halophilic archaebacteria. System. Appl. Microbiol. **6:** 157-163.

Leer, R.J., M.C. van Raamsdonk-Duin, P. Kraakman, W.H. Mager, and R.J. Planta. 1985a. The genes for yeast ribosomal proteins S24 and L46 are adjacent and divergently transcribed. Nucleic Acids Res. **13:** 701-709.

Leer, R.J., M.C. van Raamsdonk-Duin, C.M.T. Molenaar, H.M.N. Witsenboer, W.H. Mager, and R.J. Planta. 1985b. Yeast contains two functional genes coding for ribosomal protein S10. Nucleic Acids Res. **13:** 5027-5039.

Leffers, H., and R.A. Garrett. 1984. The nucleotide sequence of the 16S ribosomal RNA gene of the archaebacterium *Halococcus morrhuae*. EMBO J. **3:** 1613-1619.

Leffers, H., F. Gropp, F. Lottspeich, W. Zillig, and R.A. Garrett. 1989. Sequence, organization, transcription and evolution of RNA polymerase subunit genes from the archaebacterial extreme halophiles *Halobacterium halobium* and *Halococcus morrhuae*. J. Mol. Biol. **206:** 1-17.

Leffers, H., J. Kjems, L. Østergaard, N. Larsen, and R.A. Garrett. 1987. Evolutionary relationships amongst archaebacteria: a comparative study of 23S ribosomal RNAs of a sulfur-dependent extreme thermophile, an extreme halophile and a thermophilic methanogen. J. Mol. Biol. **195:** 43-61.

Leijonmarck, M., K. Appelt, J. Badger, A. Liljas, K.S. Wilson, and S.W. White. 1988. Structural comparison of the prokaryotic ribosomal proteins L7/L12 and L30. Proteins: structure, function and genetics **3:** 243-251.

Leijonmarck, M., and A. Liljas. 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.

Leinfelder, W., M. Jarsch, and A. Böck. 1985. The phylogenetic position of the sulfur-dependent archaebacterium *Thermoproteus tenax:* sequence of the 16S rRNA gene. System. Appl. Microbiol. **6:** 164-170.

Lewin, B. 1987. Genes III. John Wiley & Sons, New York.

Li, M., M. Tzagoloff, K. Underbrink-Lyon, and N.C. Martin. 1982. Identification of the paromomycin-resistance mutation in the 5S rRNA gene of yeast mitochondria. J. Biol. Chem. **257:** 5921-5928.

Liljas, A. 1982. Structural studies of ribosomes. Prog. Biophys. Molec. Biol. **40:** 161-228.

Liljas, A. L.A. Kirsebom, and M. Leijonmarck. 1986a. Structural studies of the factor binding domain, p. 379-390. *In* B. Hardesty and G. Kramer (eds.), Structure, function and genetics of ribosomes. Springer-Verlag, New York.

Liljas, A., and M. Leijonmarck. 1983. Crystallographic studies of the protein biosynthesis system, p. 245-273. *In* G. Pifat and J.N. Herak (eds.) Supramolecular structure and function. Plenum Publ. Co., New York.

Liljas, A., S. Thirup, and A.T. Matheson. 1986b. Evolutionary aspects of ribosome-factor interactions. Chemica Scripta **26B:**109-119.

Lin, A., Y.L. Chan, R. Jones, and I.G. Wool. 1987. The primary structure of rat ribosomal protein S12: the relationship of rat S12 to other ribosomal proteins and a correlation of the amino acid sequence of rat and yeast ribosomal proteins. J. Biol. Chem. **262:** 14343-14351.

Lin, A., J. McNally, and I.G. Wool. 1984. The primary structure of rat liver ribosomal protein L39. J. Biol. Chem. **259:** 487-490.

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

Lin, H., S. Lei, and G. Wilcox. 1985. An improved DNA sequencing strategy. Anal. Biochem. **147:** 114-119.

Lindahl, L., and J.M. Zengel. 1986. Ribosomal genes in *Escherichia coli.* Ann. Rev. Genet. **20:** 297-326.

Lipman, D.J., and W.R. Pearson. 1985. Rapid and sensitive protein similarity searches. Science 227: 1435-1441.

Londei, P., J.L. Sanz, S. Altamura, H. Hummel, P. Cammarano, R. Amils, A. Böck, and H. Wolf. 1986. Unique antibiotic sensitivity of archaebacterial polypeptide elongation factors. J. Bacteriol. **167:** 265-271.

Londei, P., A. Teichner, P. Cammarano, M. De Rosa, and A. Gambacorta. 1983. Particle weights and protein composition of the ribosomal subunits of the extremely thermoacidophilic archaebacterium *Caldariella acidophila.* Biochem. J. **209:** 461-470.

Long, E.O., and I.B. Dawid. 1980. Repeated genes in eukaryotes. Ann. Rev. Biochem. **49:** 727-764.

Loreni, F., I. Ruberti, I. Bozzoni, P. Pierandrei-Amaldi, and F. Amaldi. 1985. Nucleotide sequence of the L1 ribosomal protein gene of *Xenopus laevis*: remarkable sequence homology among introns. EMBO J. **4:** 3483-3488.

Loughney, K., E. Lund, and J.E. Dahlberg. 1982. tRNA genes are found between the 16S and 23S rRNA genes in *Bacillus subtilis.* Nucleic Acids Res. **10:** 1607-1624.

Luer, C.A., and K.P. Wong. 1979. Conformation of *Escherichia coli* ribosomal protein L7/L12 in solution: hydrodynamic, spectroscopic, and conformation prediction studies. Biochemistry **18:** 2019-2027.

Maassen, J.A., E.N. Schop, J.H.G.M. Brands, F.J. Van Hemert, J.A. Lenstra, and W. Möller. 1985. Molecular cloning and analysis of cDNA sequence for two ribosomal proteins from *Artemia.* Eur. J. Biochem. **149:** 609-616.

Maassen, J.A., E.N. Schop, and W. Möller. 1981. Structural analysis of ribosomal protein L7/L12 by the heterobifunctional cross-linker 4-(6-formyl-3 azidophenoxy) butyrimidate. Biochemistry **20:** 1020-1025.

MacConnell, W.P., and N.O. Kaplan. 1982. The activity of the acidic phosphoproteins from the 80S rat liver ribosome. J. Biol. Chem. **257:** 5359-5366.

MacConnell, W.P., and N.O. Kaplan. 1980. The role of ethanol extractable proteins from the 80S rat liver ribosome. Biochem. Biophys. Res. Commun. **92:** 46-52.

Mager, W.H. 1988. Control of ribosomal protein gene expression. Biochim. Biophys. Acta **949:** 1-15.

Maizels, N., and A.M. Weiner. 1987. Peptide-specific ribosomes, genomic tags, and the origin of the genetic code. Cold Spring Harbor Symp. Quant. Biol. **52:** 743-749.

Mandal, R.K. 1984. The organization and transcription of eukaryotic ribosomal RNA genes. Prog. Nucleic Acid Res. and Mol. Biol. **31:** 115-160.

Mangin, M., M. Ares, and A.M. Weiner. 1986. Human U2 small nuclear RNA genes contain an upstream enhancer. EMBO J. **5:** 987-995.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: a laboratory manual. Cold Spring Harbor Laboratory: Cold Spring Harbor.

Mankin, A.S., and V.K. Kagramanova. 1986. Complete nucleotide sequence of the single ribosomal RNA operon of *Halobacterium halobium:* secondary structure of the archaebacterial 23S rRNA. Mol. Gen. Genet. **202:** 152-161.

Mankin, A.S., V.K. Kagramanova, N.L. Teterina, P.M. Rubtsov, E.N. Belova, A.M. Kopylov, L.A. Baratova, and A.A. Bogdanov. 1985. The nucleotide sequence of the gene coding for the 16S rRNA from the archaebacterium *Halobacterium halobium.* Gene **37:** 181-189.

Marmur, J. 1961. A procedure for the isolation of deoxyribonucleic acid from micro-organisms. J. Mol. Biol. **3:** 208-218.

Martinez, H.M. 1988. A flexible multiple sequence alignment program. Nucleic Acids Res. **16:** 1683-1691.

Matheson. A.T. 1985. Ribosomes of Archaebacteria, p. 345-377. *In* C. R. Woese, and R.S. Wolfe (eds.), The Bacteria VIII: Archaebacteria. Academic Press, New York.

Matheson, A.T., K.A. Louie, and A. Böck. 1988. The complete amino acid sequence of the ribosomal A protein (L12) from the archaebacterium *Sulfolobus acidocaldarius.* FEBS Lett. **231:** 331-335.

Matheson, A.T., K.A. Louie, B.D. Tak, and M. Zuker. 1987. The primary structure of the ribosomal A-protein (L12) from the halophilic eubacterium *Haloanaerobium praevalens.* Biochimie **69:** 1013-1020.

Matheson, A.T., M. Yaguchi, W.E. Balch, and R.S. Wolfe. 1980. 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, A.T., M. Yaguchi, P. Christensen, C.F. Rollin, and S. Hasnain. 1984. Purification, properties, and N-terminal amino acid sequence of certain 50S ribosomal subunit proteins from the archaebacterium *Halobacterium cutirubrum*. Can. J. Bioc. and Cell Biol. **62:** 426-433.

McKuskie-Olson, H., and D.G. Glitz. 1979. Ribosome structure: localization of 3' end of RNA in small subunit by immuno electron microscopy. Proc. Natl. Acad. Sci. U.S.A. **76:** 3769-3773.

Messing, J. 1983. New M13 vectors for cloning. 1983. Meth. in Enzymol. **101:** 20-78.

Messing, J., R. Crea, and P.H. Seeburg. 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9:** 309-321.

Messing, J., B. Groenborn, B. Müller-Hill, and P.H. Hofschneider. 1977. Filamentous coliphage M13 as a cloning vehicle: Insertion of a Hind II fragment of the *lac* regulatory region in M13 replicative form *in vitro*. Proc. Natl. Acad. Sci. U.S.A. **74:** 3642-3643.

Mevarech, M., S. Hirsh-Twizer, S. Goldman, E. Yakobson, H. Eisenberg and P.P. Dennis. 1989. Isolation and characterization of the rRNA gene clusters of *Halobacterium marismortui*. J. Bacteriol. **171:** 3479-3485.

Mirande, M., and J.P. Waller. 1988. The yeast lysyl-tRNA synthetase gene. J. Biol. Chem. **263:** 18443-18451.

Mitsui, K., and K. Tsurugi. 1988a. cDNA and deduced amino acid sequence of 38 KDa-type acidic ribosomal protein A0 from *Saccharomyces cerevisiae*. Nucleic Acids Res. **16:** 3573.

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

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

Miyada, C.G., C. Klofelt, A.A. Reyes, E. McLaughlin-Taylor, and R.B. Wallace. 1985. Evidence that polymorphism in the murine major

histocompatibility complex may be generated by the assortment of subgene sequences. Proc. Natl. Acad. Sci. U.S.A. **82:** 2890-2894.

Miyada, C.G., and R.B. Wallace. 1987. Oligonucleotide hybridization techniques. Meth. in Enzymol. **154:** 94-107.

Moazed, D., and H.F. Noller. 1989. Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. Cell **57:** 585-597.

Moazed, D., J.M. Robertson, and H.F. Noller. 1988. Interaction of elongation factors EF-G and EF-Tu with a conserved loop in 23S RNA. Nature **334:** 262-364.

Moldave, K. 1985. Eukaryotic protein synthesis. Ann. Rev. Biochem. **54:** 1109-1149.

Molenaar, C.M.T., L.P. Woudt, A.E.M. Jansen, W.H. Mager, and R.J. Planta. 1984. Structure and organization of two linked ribosomal protein genes in yeast. Nucleic Acids Res. **12:** 7345-7358.

Möller, W., and H. Castleman. 1967. Characterization of an acidic protein in 50S ribosomes of *Escherichia coli.* Nature **215:** 1293-1295.

Möller, W., and J.A. Maassen. 1986. On the structure, function and dynamics of L7/L12 from *Escherichia coli* ribosomes, p. 309-325. *In* B. Hardesty and G. Kramer (eds.), Structure, function and genetics of ribosomes. Springer-Verlag, New York.

Möller, W., P.I. Schrier, J.A. Maassen, A. Zantema, E. Schop, H. Reinalda, A.F.M. Cremers, and J.E. Mellema. 1983. Ribosomal proteins L7/L12 of *Escherichia coli*: localization and possible mechanism in translation. J. Mol. Biol. **163:** 553-573.

Möller, W., L.I. Slobin, R. Amons, and D. Richter. 1975. Isolation and characterization of two acidic proteins of 60S ribosomes from *Artemia salina* cysts. Proc. Natl. Acad. Sci. U.S.A. **72:** 4744-4748.

Monk, R.J., O. Meyuhas, and R.P. Perry. 1981. Mammals have multiple genes for individual ribosomal proteins. Cell **24:** 301-306.

Montandon, P., P. Nicolas, P. Schürmann, and E. Stutz. 1985. Streptomycin-resistance of *Euglena gracilis* chloroplasts: identification of a point mutation in the 16S rRNA gene in an invariant position. Nucleic Acids Res. **13:** 4299-4310.

Morgan, E.A., T. Ikemura, and M. Nomura. 1977. Identification of spacer tRNA genes in individual ribosomal RNA transcription units of *Escherichia coli.* Proc. Natl. Acad. Sci. U.S.A. **74:** 2710-2714.

Moritz, A., and W. Goebel. 1985. Characterization of the 7S RNA and its gene from halobacteria. Nucleic Acids Res. **13:** 6969-6980.

Morris, C.J., and J.N. Reeve. 1988. Conservation of structure in the human gene encoding argininosuccinate synthetase and the *arg G* genes of the archaebacteria *Methanosarcina barkeri* MS and *Methanococcus vannielii.* J. Bacteriol. **170:** 3125-3130.

Müller, B., R. Allmansberger, and A. Klein. 1985. Termination of a transcription unit comprising highly expressed genes in the archaebacterium *Methanococcus voltae.* Nucleic Acids Res. **13:** 6439-6445.

Nelson, N., and L. Taiz. 1989. Evolution of H^+ ATPases. Trends in Bioc. Sci. **14:** 113-116.

Neumann, H., A. Gierl, J. Tu, J. Leibrock, D. Staiger, and W. Zillig. 1983. Organization of the genes for ribosomal RNA in archaebacteria. Mol. Gen. Genet. **192:** 66-72.

Noller, H.F. 1984. Structure of ribosomal RNA. Ann. Rev. Biochem. **53:** 119-162.

Noller, H.F., and J.A. Lake. 1984. Ribosomes, structure and function: localization of rRNA, p. 217-297. *In* E. Bittar (ed.). Membrane Structure and Function Vol. 6. John Wiley and Sons, New York.

Nomura, M. 1986. Regulation of the synthesis of ribosomes and ribosomal components in *Escherichia coli* : translational regulation and feedback loops, p. 199-220. *In* I. Booth, and C. Higgins (eds.) Symposium of the Society for General Microbiology: regulation of gene expression. Cambridge University Press, Cambridge.

Nomura, M., R. Gourse, and G. Baughman. 198[illegible] Regulation of the synthesis of ribosomes and ribosomal components. Ann. Rev. Biochem. **53:** 75-117.

Nomura, M., E. Morgan, and S.R. Jaskunas. 1977. Genetics of Bacterial Ribosomes. Ann. Rev. Genet. **11:** 297-347.

Nomura, M., and L.E. Post. 1980. Organization of ribosomal genes and regulation of their expression in *Escherichia coli,* p. 761-791. *In* G. Chambliss, G.R. Craven, J. Davies, K. Davis, L. Kahan, and M. Nomura (eds.), Ribosomes: structure function and genetics. University Park Press, Baltimore.

Oakes, M., E. Henderson, A. Scheinman, M. Clark, and J.A. Lake. 1986. Ribosome structure, function and evolution: mapping ribosomal RNA, proteins, and functional sites in three dimensions, p. 47-67. *In* B. Hardesty and

G. Kramer (eds.), Structure, function and genetics of ribosomes. Springer-Verlag, New York.

Ohba, M., and T. Oshima. 1981. DNA-protein complex from an extreme halophile: a histone-like protein in archaebacteria, p. 543-549. *In* Y. Wolman (ed.). Origin of life, Reidel Publ. Co., Dordrecht.

Ohkubo, S., A. Muto, Y. Kawauchi, F. Yamao, and S. Osawa. 1987. The ribosomal protein cluster of *Mycoplasma capricolum.* Mol. Gen. Genet. **210:** 314-322.

Oliver, J.L., J.L. Sanz, R. Amils, and A. Marin. 1987. Inferring the phylogeny of archaebacteria: the use of ribosomal sensitivity to protein-synthesis inhibitors. J. Mol. Evol. **24:** 281-288.

Olsen, G.J., N.P. Pace, M. Nuell, B.P. Kaine, R. Gupta, and C.R. Woese. 1985. Sequence of the 16S rRNA gene from the thermoacidophilic archaebacterium *Sulfolobus solfataricus* and its evolutionary implications. J. Mol. Evol. **22:** 301-307.

Olsen, G.J., and C.R. Woese. 1989. A brief note concerning archaebacterial phylogeny. Can. J. Microbiol. **35:** 119-123.

Olson, H. M. , A. Sommer, D.S. Tewari, R.R. Traut, and D.G. Glitz. 1986. Localization of two epitopes of protein L7/L12 to both the body and stalk of the large ribosomal subunit: immune electron microscopy using monoclonal antibodies. J. Biol. Chem. **261:** 6924-6932.

Oshima, T., M. Obha, and T. Wakagi. 1984. Some biochemical properties of an acido-thermophilic archaebacterium *Sulfolobus acidocaldarius.* Origins of Life **14:** 665-669.

Österberg, R., B. Sjöberg, A. Liljas, and I. Pettersson. 1976. Small-angle x-ray scattering and cross-linking study of the proteins L7/L12 from *Escherichia coli* ribosomes. FEBS Lett. **66:** 48-51.

Østergaard, L. N. Larsen, H. Leffers, J. Kjems, and R. Garrett. 1987. A ribosomal RNA operon and its flanking region from the archaebacterium *Methanobacterium thermoautotrophicum*, Marburg strain: transcription signals, RNA structure and evolutionary implications. System. Appl. Microbiol. **9:** 199-209.

Otaka, E., K. Higo, and T. Itoh. 1984. Yeast ribosomal proteins. Mol. Gen. Genet. **195:** 544-546.

Otaka, E., T. Ooi, T. Kumazaki, and T. Itoh. 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.

Otaka, E., T. Ooi, and K. Suzuki, 1989. Examination of protein sequence homologies VI. The evolution of *Escherichia coli* L7/L12 equivalent ribosomal proteins ('A' proteins), and the tertiary structure. Protein Seq. Data Anal. **2:** 395-402.

Parker, J., R.J. Watson, J.D. Friesen, and N.P. Fiil. 1976. A relaxed mutant with an altered ribosomal protein L11. Mol. Gen. Genet. **144:** 111-114.

Peled-Yalif, E., I. Cohen-Binder, and O. Meyuhas. 1984. Isolation and characterization of four mouse ribosomal protein L18 genes that appear to be processed pseudogenes. Gene **29:** 157-166.

Pettersson, I. 1979. Studies on the RNA and protein binding sites of the *Escherichia coli* ribosomal protein L10. Nucleic Acids Res. **6:** 2637-2646.

Pettersson, I., S.J.S. Hardy, and A. Liljas. 1976. The ribosomal protein L8 is a complex of L7/L12 and L10. FEBS Lett. **64:** 135-138.

Pettersson, I., and A. Liljas. 1979. The stoichiometry and reconstitution of a stable protein complex from *Escherichia coli* ribosomes. FEBS Lett. **98:** 139-144.

Planta, R.J., W.H. Mager, R.J. Leer, L.P. Woudt, H.A. Raue, and El-Baradi, T.T.A.L. 1986. Structure and expression of ribosomal protein genes in yeast, p. 699-718. *In* B. Hardesty and G. Kramer (eds.), Structure, function and genetics of ribosomes. Springer-Verlag, New York.

Planta, R.J., and J.H. Meyerink. 1980. Organization of the ribosomal RNA genes in eukaryotes, p.871-887. *In* G. Chambliss, G.R. Craven, J. Davies, K. Davis, L. Kahan, and M. Nomura (eds.), Ribosomes: structure function and genetics. University Park Press, Baltimore.

Post, L.E., G.D. Strycharz, M. Nomura, H. Lewis, and P.P. Dennis. 1979. Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit in *Escherichia coli.* Proc. Natl. Acad. Sci. U.S.A. **76:** 1697-1701.

Poritz, M.A., K. Strub, and P. Walter. 1988. Human SRP RNA and *Escherichia coli* 4.5S RNA contain a highly homologous structural domain. Cell **55:** 4-6.

Pühler, G., F. Lottspeich, and W. Zillig. 1989. Organization and nucleotide sequence of the genes encoding the large subunits A, B, and C of the DNA-dependent RNA polymerase of the archaebacterium *Sulfolobus acidocaldarius.* Nucleic Acids Res. **17:** 4517-4534.

Putney, S.D., N.J. Royal, H. Neuman de Vegvar, W.C. Herlihy, K. Biemann and P. Schimmel. 1981a. Primary structure of a large aminoacyl-tRNA synthetase. Science **213:** 1497-1500.

Putney, S.D., R.T. Sauer and P. Schimmel. 1981b. Purification and properties of alanine tRNA synthetase from *Escherichia coli:* a tetramer of identical subunits. J. Biol. Chem. **256:** 198-204.

Putney, S.D., and P. Schimmel. 1981. An aminoacyl tRNA synthetase binds to a specific DNA sequence and regulates its gene transcription. Nature **291:** 632-635.

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

Radermacher, M., T. Wagenknecht, A. Verschoor, and J. Frank. 1987. Three-dimensional structure of the large ribosomal subunit from *Escherichia coli.* EMBO J. **6:** 1107-1114.

Ramírez, C., K.A. Louie, and A.T. Matheson. 1989a. A small basic ribosomal protein in *Sulfolobus solfataricus* equivalent to L46 in yeast: structure of the protein and its gene. FEBS Lett. **250:** 416-418.

Ramírez, C., L.C. Shimmin, C.H. Newton, A.T. Matheson, and P.P. Dennis. 1989b. Structure and evolution of the L11, L1, L10 and L12 equivalent ribosomal proteins in eubacteria, archaebacteria and eukaryotes. Can. J. Microbiol. **35:** 234-244.

Ree, H.K., K. Cao, D.L. Thurlow, and R.A. Zimmermann. 1989. The structure and organization of the 16S ribosomal RNA gene from the archaebacterium *Thermoplasma acidophilum.* Can. J. Microbiol. **35:** 124-133.

Reichel, R., and B.J. Benecke. 1980. Reinitiation of synthesis of small cytoplasmic RNA species K and L in isolated HeLa cell nuclei *in vitro.* Nucleic Acids Res. **8:** 225-234.

Reichel, R. and B.J. Benecke. 1984. Localization of small nuclear polymerase I RNA sequences at the 5' end of the human rDNA transcription unit. EMBO J. **3:** 473-479.

Reiter, W.D., P. Palm, A. Henschen, F. Lottspeich, W. Zillig, and B. Grampp. 1987a. Identification and characterization of the genes encoding three structural proteins of the *Sulfolobus* virus-like, SSV1 particle. Mol. Gen. Genet. **206:** 144-153.

Reiter, W.D., P. Palm, W. Voos, J. Kaniecki, B. Grampp, W. Schulz, and W. Zillig. 1987b. Putative promoter elements for the ribosomal RNA genes of the thermoacidophilic archaebacterium *Sulfolobus* sp. strain B12. Nucleic Acids Res. **15:** 5581-5595.

Reiter, W.D., P. Palm, S. Yeats, and W. Zillig. 1987c. Gene expression in archaebacteria: physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. Mol. Gen. Genet. **209:** 270-275.

Reiter, W.D., P. Palm, and W. Zillig. 1988. Analysis of transcription in the archaebacterium *Sulfolobus* indicates that archaebacterial promoters are homologous to eukaryotic pol II promoters. Nucleic Acids Res. **16:** 1-19.

Remacha, M., M.T. Sáenz-Robles, M. Vilella, and J.P.G. Ballesta. 1988. Independent genes coding for three acidic proteins of the large ribosomal subunit from *Saccharomyces cerevisiae*. J. Biol. Chem. **263:** 9094-9101.

Rice, P.A., and T.A. Steitz. 1989. Ribosomal protein L7/L12 has a helix-turn-helix motif similar to that found in DNA-binding regulatory proteins. Nucleic Acids Res. **17:** 3757-3762.

Rich, B.E., and J.A. Steitz. 1987. Human acidic ribosomal phosphoproteins P0, P1, and P2: analysis of cDNA clones, *in vitro* synthesis and assembly. Mol. Cell. Biol. **7:** 4065-4074.

Römisch, K., J. Webb, J. Herz, S. Prehn, R. Frank, M. Vingron, and B. Dobberstein. 1989. Homology of 54 K protein of signal recognition particle, docking protein and two *Escherichia coli* proteins with putative GTP-binding domains. Nature **340:** 478-482.

Rychlik, W., O.W. Odom, and B. Hardesty. 1983. Localization of the elongation factor Tu binding site on *Escherichia coli* ribosomes. Biochemistry **22:** 85-93.

Sakonju, S., D.F. Bogenhagen, and D.D. Brown. 1980. A control region in the center of the 5S RNA gene directs specific initiation of transcription: I. the 5' border of the region. Cell **19:** 13-25.

Sánchez-Madrid, F., R. Reyes, P. Conde, and J.P.G. Ballesta. 1979. Acidic ribosomal proteins from eukaryotic cells: effect on ribosomal functions. Eur. J. Biochem. **98:** 409-416.

Sánchez-Madrid, F., F. Juan-Vidales, and J.P.G. Ballesta. 1981a. Effect of phosphorylation on the affinity of acidic proteins from *Saccharomyces cerevisiae* for the ribosomes. Eur. J. Biochem. **114:** 609-613.

Sánchez-Madrid, F., F. Juan-Vidales, and J.P.G. Ballesta. 1981b. Functional role of acidic ribosomal proteins. Interchangeability of proteins from bacterial and eukaryotic cells. Biochemistry **20:** 3263-3266.

Sander, G. 1983. Ribosomal protein L1 from *Escherichia coli:* its role in the binding of tRNA to the ribosome and in elongation factor G-dependent GTP hydrolysis. J. Biol. Chem. **258:** 10098-10102.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74:** 5463-5467.

Sanz, J.L., S. Altamura, I. Mazziotti, R. Amils, P. Cammarano, and P. Londei. 1987. Unique antibiotic sensitivity of an *in vitro* polypeptide synthesis system from the archaebacterium *Thermoplasma acidophilum.* Phylogenetic implications. Mol. Gen. Genet. **207:** 385-394.

Schallenberg, J., M.Moes, M. Truss, W. Reiser, M. Thomm, K.O. Stetter, and A. Klein. 1988. Cloning and physical mapping of RNA polymerase genes from *Methanobacterium thermoautotrophicum* and comparison of homologies and gene orders with those of RNA polymerase genes from other methanogenic archaebacteria. J. Bacteriol. **170:** 2247-2253.

Schimmel, P. 1987. Aminoacyl tRNA synthetases: general structure-function relationships in the polypeptides and recognition of transfer RNAs. Ann. Rev. Biochem. **56:** 125-158.

Schmid , G. and A. Böck. 1982. The ribosomal protein composition of the archaebacterium *Sulfolobus*. Mol. Gen. Genet. **185:** 498-501.

Schmid, G., T. Pecher, and A. Böck. 1982. Properties of the translational apparatus of archaebacteria. Zbl.Bakt.Hyg., I.Abt.Orig.C **3:** 209-217.

Schmidt, F.J., J. Thompson, K. Lee, J. Dijk, and E. Cundliffe. 1981. The binding site for ribosomal protein L11 within 23S ribosomal RNA of *Escherichia coli.* J. Biol. Chem. **256:** 12301-12305.

Schop, E.N., and J.A. Maassen. 1982. Characterization of the region on protein L7/L12 involved in binding to ribosomal particles. Eur. J. Biochem. **128:** 371-375.

Schümperli, D., K. Mc.Kenney, D.A. Sobieski, and M. Rosenberg. 1982. Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon. Cell **30:** 865-871.

Sentenac, A. 1985. Eukaryotic RNA polymerases. CRC Critical Rev. Biochem. **19:** 31-90.

Sharp, S.J., J. Schaack, L. Cooley, D.J. Burke, and D. Söll. 1986. Structure and transcription of eukaryotic tRNA genes. CRC Critical Rev. Biochem. **19:** 107-144.

Shatkin, A.J. 1976. Capping of eukaryotic mRNAs. Cell **9:** 645-653.

Shimmin, L.C., and P.P. Dennis. 1989a. 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, L.C., C.H. Newton, C. Ramírez, J. Yee, W.L. Downing, A. Louie, A.T. Matheson, and P.P. Dennis. 1989. Organization of genes encoding the L11, L1, L10 and L12 equivalent ribosomal proteins in eubacteria, archaebacteria and eukaryotes. Can. J. Microbiol. **35:** 164-170.

Shimmin, L.C., C. Ramírez, A.T. Matheson, and P.P. Dennis. 1989b. Sequence alignment and evolutionary comparison of the L10 equivalent and L12 equivalent ribosomal proteins from archaebacteria, eubacteria, and eukaryotes. J. Mol. Evol. *in press.*

Shine, J., and L. Dalgarno. 1974. The 3' terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. U.S.A. **71:** 1342-1346.

Shuber, A.P., E.C. Orr, M.A. Recny, P.F. Schendel, H.D. May, N.L. Shauer, and J.G. Ferry. 1986. Cloning, expression and nucleotide sequence of the formate dehydrogenase genes from *Methanobacterium formicicum.* J. Biol. Chem. **261:** 12942-12947.

Sibold, L., and M. Henriquet. 1988. Cloning of the *trp* genes from the archaebacterium *Methanococcus voltae*: nucleotide sequence of the *trp BA* genes. Mol. Gen. Genet. **214:** 439-450.

Siegel, V., and P. Walter. 1988. Each of the activities of signal recognition particle (SRP) is contained within a distinct domain: analysis of biochemical mutants of SRP. Cell **52:** 39-49.

Sinha, N.D., J. Biernat, J. McManus, and H. Koster. 1984. Polymer support oligonucleotide synthesis. XVIII. Use of β cyanoethyl-N,N-dialkylamino-N-morpholine phosphoramidite of deoxynucleosides for the synthesis of DNA fragments simplifying deprotection and isolation of the final product. Nucleic Acids Res. **12:** 4539-4557.

Sköld, S. 1983. Chemical crosslinking of elongation factor G to the 23S RNA in 70S ribosomes from *Escherichia coli.* Nucleic Acids Res. **11:** 4923-4932.

Sommerville, J. 1986. Nucleolar structure and ribosome biogenesis. Trends in Bioc. Sci. **11:** 438-442.

Sor, F., M. Bolotin-Fukuhara, and M. Nomura. 1987. Mutational alterations of translational coupling in the L11 ribosomal protein operon of *Escherichia coli*. J. Bacteriol. **169:** 3495-3507.

Sor, F., and M. Nomura. 1987. Cloning and DNA sequence determination of the L11 ribosomal protein operon of *Serratia marcescens* and *Proteus vulgaris*: translational feedback regulation of *Escherichia coli* L11 operon by heterologous L1 proteins. Mol. Gen. Genet. **210:** 52-59.

Souillard, N., and L. Sibold. 1986. Primary structure and expression of a gene homologous to *nif H* (nitrogenase Fe protein) from the archaebacterium *Methanococcus voltae*. Mol. Gen. Genet. **203:** 21-28.

Souillard, N., M. Magot, O. Possot, and L. Sibold. 1988. Nucleotide sequence of regions homologous to *nif H* (nitrogenase Fe protein) from the nitrogen-fixing archaebacteria *Methanococcus thermolithotrophicum* and *Methanobacterium ivanovii:* evolutionary implications. J. Mol. Evol. **27:** 65-76.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:** 503-517.

Spiridonova, V.A., A.S. Akhmanova, V.K. Kagramanova, A.K.E. Köpke, and A.S. Mankin. 1989. Ribosomal protein gene cluster of *Halobacterium halobium:* nucleotide sequence of the genes coding for S3 and L29 equivalent ribosomal proteins. Can. J. Microbiol. **35:** 153-159.

Spirin, A.S. 1986. Ribosome Structure and Protein Biosynthesis. Benjamin Cummings Publ. Co. , Menlo Park.

Steel, L.F., and A. Jacobson. 1986. Ribosomal proteins are encoded by single copy genes in *Dictyostelium discoideum*. Gene **41:** 165-172.

Stern, S., T. Powers, L-M. Changchien, and H.F. Noller. 1989. RNA-protein interactions in 30S ribosomal subunits: folding and function of 16S rRNA. Science **244:** 783-790.

Stetter, K.O. 1986. Diversity of extremely thermophilic archaebacteria, p. 39-74. *In* T.D. Brock (ed.). Thermophiles: general, molecular and applied microbiology. John Wiley and Sons, New York.

Stetter, K.O., H. Konig, and E. Stackebrandt. 1983. *Pyrodictium* gen. nov., a new genus of submarine disc-shaped sulfur-reducing archaebacteria growing optimally at 105 °C. System. Appl. Microbiol. **4:** 535-551.

Stetter, K.O., G. Lauerer, M. Thomm, and A. Neuner. 1987. Isolation of extremely thermophilic sulfate reducers: evidence for a novel branch of archaebacteria. Science **236:** 822-824.

Stetter, K.O., A. Segerer, W. Zillig, G. Huber, G. Fiala, R. Huber, and H. Konig. 1986. Extremely thermophilic sulfur-metabolizing archaebacteria. System. Appl. Microbiol. **7:** 393-397.

Stetter, K.O., and W. Zillig. 1985. *Thermoplasma* and the thermophilic sulfur-dependent archaebacteria, p. 85-170. *In* C. R. Woese, and R.S. Wolfe (eds.), The Bacteria VIII: Archaebacteria. Academic Press, New York.

Stöffler, G., and M. Stöffler-Meilicke. 1986a. Electron microscopy of archaebacterial ribosomes. System. Appl. Microbiol. **7:** 123-130.

Stöffler, G., and M. Stöffler-Meilicke. 1986b. Immuno electron microscopy on *Escherichia coli* ribosomes, p. 28-46. *In* B. Hardesty and G. Kramer (eds.), Structure, function and genetics of ribosomes. Springer-Verlag, New York.

Stöffler, G., I.G. Wool, A. Lin, and K.H. Rak. 1974. The identification of the eukaryotic ribosomal proteins homologous with *Escherichia coli* proteins L7 and L12. Proc. Natl. Acad. Sci. U.S.A. **71:** 4723-4726.

Strauss, E.C., J.A. Kobori, G. Sui, and L.E. Hood. 1986. Specific-primer-directed DNA sequencing. Anal. Biochem. **154:** 353-360.

Strobel, O., A.K.E. Köpke, R.M. Kamp, A. Böck, and B. Wittmann-Liebold. 1988. Primary structure of the archaebacterial *Methanococcus vannielii* ribosomal protein L12. J. Biol. Chem. **263:** 6538-6546.

Strøm, A.R., and L.P. Visentin. 1973. Acidic ribosomal proteins from the extreme halophile, *Halobacterium cutirubrum:* the simultaneous separation, identification, and molecular weight determination. FEBS Lett. **37:** 274-280.

Struck, J.C.R., H.Y. Toschka, T. Specht, and V.A. Erdmann. 1988. Common structural features between eukaryotic 7SL RNAs, eubacterial 4.5S RNA and scRNA and archaebacterial 7S RNA. Nucleic Acids Res. **16:** 7740.

Strycharz, W.A., M. Nomura, and J.A. Lake. 1978. Ribosomal proteins L7/L12 localized at a single region of the large subunit by immune electron microscopy. J. Mol. Biol. **126:** 123-140.

Subramanian, A.R. 1975. Copies of L7 and L12 and heterogeneity of the large subunit of *Escherichia coli* ribosome. J. Mol. Biol. **95:** 1-8.

Subramanian, A.R., and E.R. Dabbs. 1980. Functional studies on ribosomes lacking protein L1 from mutant *Escherichia coli.* Eur. J. Biochem. **112:** 425-430.

Szostak, J.W., J.I. Stiles, B.K. Tye, P. Chiu, F. Sherman, and R. Wu. 1979. Hybridization with synthetic oligonucleotides. Meth. in Enzymol. **68:** 419-428.

Tajima. S., L. Lauffer, V. L. Rath, and P. Walter. 1986. The signal recognition particle receptor is a complex that contains two distinct polypeptide chains. J. Cell Biol. **103:** 1167-1178.

Tate, W.P., A.L. Beaudet, and C.T. Caskey. 1973. Influence of guanine nucleotides and elongation factors on interaction of release factors with the ribosome. Proc. Natl. Acad. Sci. U.S.A. **70:** 2350-2352.

Tate, W.P., M.J. Dognin, M. Noah, M. Stöffler-Meilicke, and G. Stöffler. 1984. The N-terminal domain of *Escherichia coli* ribosomal protein L11: its three dimensional location and its role in the binding of release factor 1. J. Biol. Chem. **259:**7317-7324.

Teem, J.L., N. Abovich, N.F. Kaufer, W.F. Schwindinger, J.R. Warner, A. Levy, J.L. Woolford, Jr., R.J. Leer, M.M.C. van Raamsdonk-Duin, W.H. Mager, R.J. Planta, L. Schultz, J.D. Friesen, H. Fried, and M. Rosbash. 1984. A comparison of yeast ribosomal protein gene DNA sequences. Nucleic Acids Res.**12:** 8295-8312.

Teixido, J., S. Altamura, P. Londei, and R. Amils. 1989. Structural and functional exchangeability of 5S RNA species from the eubacterium *Escherichia coli* and the thermoacidophilic archaebacterium *Sulfolobus solfataricus.* Nucleic Acids Res. **17:** 845-851.

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

Tewari, D.S., A. Sommers, and R.R. Traut. 1986. The selective release of one of the two L7/L12 dimers from the *Escherichia coli* ribosome induced by a monoclonal antibody to the NH_2-terminal region. J. Biol. Chem. **261:** 6919-6923.

Thielen, T.P. G.M., J. A. Maassen, J. Krieb, and W. Möller. 1984. Dimer structure of the ribosomal protein L7/L12 probed by energy transfer. Biochemistry **23:** 6668-6674.

Thomm, M., B.A. Sherf, and J.N. Reeve. 1988. RNA polymerase-binding and transcription initiation sites upstream of the methyl reductase operon of *Methanococcus vannielii.* J. Bacteriol. **170:** 1958-1961.

Thomm, M., and G. Wich. 1988. An archaebacterial promoter element for stable RNA genes with homology to the TATA box of higher eukaryotes. Nucleic Acids Res. **16:** 151-163.

Thomm, M., G. Wich, J.W. Brown, G. Frey, B.A. Sherf, and G.S. Beckler. 1989. An archaebacterial promoter sequence assigned by RNA polymerase binding experiments. Can. J. Microbiol. **35:** 30-35.

Thurlow, D.L., and R.A. Zimmermann. 1982. Evolution of protein binding regions of archaebacterial, eubacterial, and eukaryotic ribosomal RNAs, p. 347. *In* O. Kandler (ed.) Archaebacteria. Gustav Fischer Verlag, Stuttgart.

Tindall, B.J., H.N.M. Ross, and W.D. Grant. 1984. *Natrobacterium* gen. nov. and *Natronococcus* gen. nov. two new genera of haloalkaliphilic archaebacteria. System. Appl. Microbiol. **5:** 41-57.

Tokimatsu, H., W.A. Strycharz, and A.E. Dahlberg. 1981. Gel electrophoretic studies on ribosomal proteins L7/L12 and the *Escherichia coli* 50S subunit. J. Mol. Biol. **152:** 397-412.

Tolan, D.R., and R.R. Traut. 1981. Protein topography of the 40S ribosomal subunit from rabbit reticulocytes shown by cross-linking with 2-iminothiolane. J. Biol. Chem. **256:** 10129-10136.

Tomioka, N. and M. Sugiura. 1984. Nucleotide sequence of the 16S-23S spacer region in the *rrnA* operon from a blue-green alga, *Anacystis nidulans*. Mol. Gen. Genet. **193:** 427-430.

Traut, R.R., J.M. Lambert, and J.W. Kenny. 1983. Ribosomal protein L7/L12 cross-links to proteins in separate regions of the 50S ribosomal subunit of *Escherichia coli.* J. Biol. Chem. **258:** 14592-14598.

Traut, R.R., D.S. Tewari, A. Sommer, G.R. Gavino, H.M. Olson, and D.G. Glitz. 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, p. 286-308. *In* B. Hardesty and G. Kramer (eds.), Structure, function and genetics of ribosomes. Springer-Verlag, New York.

Tsao, S.G.S., C.F. Brunk, and R.E. Perlmann. 1983. Hybridization of nucleic acids directly in agarose gels. Anal. Biochem. **131:** 365-

Tu, J., and W. Zillig. 1982. Organization of rRNA structural genes in the archaebacterium *Thermoplasma acidophilum*. Nucleic Acids Res. **10:** 7231-7245.

Uchiumi, T., M. Kikuchi, K. Terao, and K. Ogata. 1986. Cross-linking study on protein topography of rat liver 60S ribosomal subunits with 2-iminothiolane. J. Biol. Chem. **260:** 5675-5682.

Uchiumi, T., A.J. Wahba, and R.R. Traut. 1987. Topography and stoichiometry of acidic proteins in large ribosomal subunits from *Artemia salina* as determined by crosslinking. Proc. Natl. Acad. Sci. U.S.A. **84:** 5580-5584.

Van Ness, B.G., J.B. Howard, and J.W. Bodley. 1980. ADP ribosylation of elongation factor 2 by diphtheria toxin. J. Biol. Chem. **255:** 10710-10716.

Vaslet, C.A., P. O'Connell, M. Izquierdo, and M. Robash. 1980. Isolation and mapping of a cloned ribosomal protein gene of *Drosophila melanogaster*. Nature **285:** 674-676.

Verner, K., and G. Schatz. 1988. Protein translocation across membranes. Science **241:** 1307-1313.

Verschoor, A., J. Frank, and M. Boublik. 1985. Investigation of the 50S ribosomal subunit by electron microscopy and image analysis. J. of Ultrastructure Res. **92:** 180-189.

Verschoor, A., J. Frank, M. Radermacher, and T. Wagenknecht. 1984. Three-dimensional reconstruction of the 30S ribosomal subunit from randomly oriented particles. J. Mol. Biol. **178:** 677-698.

Vieira, J., and J. Messing. 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19:** 259-268.

Von Holt, C., W. N. Strickland, W.F. Bandt, and M.S. Strickland. 1979. More histone structures. FEBS Lett. **100:** 201-218.

Wada, A., and T. Sako. 1987. Primary structures of and genes for new ribosomal proteins A and B in *Escherichia coli*. J. Biochem. **101:** 817-820.

Wagenknecht, T., R. Grassucci, and J. Frank. 1988. Electron microscopy and computer image averaging of ice-embedded large ribosomal subunits from *Escherichia coli* . J. Mol. Biol. **199:** 137-147.

Wagner, M., and R. Perry. 1985. Characterization of the multigene family encoding the mouse S16 ribosomal protein : strategy for distinguishing an expressed gene from its processed pseudogene counterparts by an analysis of total genomic DNA. Mol. Cell. Biol. **5:** 3560-3576.

Wahl, G.M., S.L. Berger, and A.R. Kimmel. 1987. Molecular hybridization of immobilized nucleic acids: theoretical concepts and practical considerations. Meth. in Enzymol. **152:** 399-407.

Wallace, R.B., M.J. Johnson, T. Hirose, T. Miyake, E.H. Kawashima, and K. Itakura. 1981. The use of synthetic oligonucleotides as hybridization probes II. Hybridization of oligonucleotides of mixed sequence to rabbit β-globin DNA. Nucleic Acids Res. **9:** 879-894.

Wallace, R.B., and C.G. Miyada. 1987. Oligonucleotide probes for the screening of recombinant DNA libraries. Meth. in Enzymol. **152:** 432-442.

Wallace, R.B., J. Shaffer, R.F. Murphy, J. Bonner, T. Hirose, and K. Itakura. 1979. Hybridization of synthetic oligodeoxyribonucleotides to ΦX174 DNA: the effect of single base pair mismatch. Nucleic Acids Res. **6:** 3543-3557.

Walter, P., and G. Blobel. 1982. Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. Nature **299:** 691-698.

Walter, P., and G. Blobel. 1981. Translocation of proteins across the endoplasmic reticulum III. Signal recognition protein (SRP) causes signal sequence and site specific arrest of chain elongation that is released by microsomal membranes. J. Cell Biol. **91:** 557-561.

Walter, P., R. Gilmore, and G. Blobel. 1984. Protein translocation across the endoplasmic reticulum. Cell **38:** 5-8.

Warner, J.R. 1989. Synthesis of Ribosomes in *Saccharomyces cerevisiae.* Microbiol. Rev. **53:** 256-271.

Warner, J.R., E.A. Elion, M.D. Dabeva, and W.F. Schwindinger. 1986. The ribosomal genes of yeast and their regulation, p. 719-732. *In* B. Hardesty and G. Kramer (eds.), Structure, function and genetics of ribosomes. Springer-Verlag, New York.

Warner, J.R., G. Mitra, W.F. Schwindinger, M. Studeny, and H.M. Fried. 1985. *Saccharomyces cerevisiae* coordinates the accumulation of yeast ribosomal proteins by modulating mRNA splicing, translational initiation and protein turnover. Mol. Cell. Biol. **5:** 1512-1521.

Watson, R.J., J. Parker, N.P. Fiil, J.G. Flaks, and J.D. Friesen. 1975. New chromosomal location for structural genes of ribosomal proteins. Proc. Natl. Acad. Sci. U.S.A. **72:** 2765-2769.

Webster, T., H. Tsai, G.A. Mikula, and P. Schimmel. 1984. Specific sequence homology and three-dimensional structure of an aminoacyl transfer RNA synthetase. Science **226:** 1315-1317.

Weil, C.F., G.S. Beckler, and J.N. Reeve. 1987. Structure and organization of the *his A* gene of the thermophilic archaebacterium *Methanococcus thermolithotrophicus.* J. Bacteriol. **169:** 4857-4860.

Weiss, R.B., D.M. Dunn, A.E. Dahlberg, J.F. Atkins, and R.F. Gesteland. 1988. Reading frame switch caused by base-pair formation between the 3' end of 16S rRNA and the mRNA during elongation of protein synthesis in *Escherichia coli.* EMBO J. **7:** 1503-1507.

Whitman, W.B. 1985. Methanogenic bacteria, p. 3-84. *In* C. R. Woese, and R.S. Wolfe (eds.), The Bacteria VIII: Archaebacteria. Academic Press, New York.

Wich, G., H. Hummel, M. Jarsch, U. Bär, and A. Böck. 1986a. Transcriptional signals for stable RNA genes in *Methanococcus.* Nucleic Acids Res. **14:** 2459-2479.

Wich, G., M. Jarsch, and A. Böck. 1984. Apparent operon for a 5S ribosomal RNA gene and for tRNA genes in the archaebacterium *Methanococcus vannielii.* Mol. Gen. Genet. **196:** 146-151.

Wich, G., W. Leinfelder, and A. Böck. 1987. Genes for stable RNA in the extreme thermophile *Thermoproteus tenax:* introns and transcription signals. EMBO J. **6:** 523-528.

Wich, G., L. Sibold, and A. Böck. 1986b. Genes for tRNA and their putative expression signals in *Methanococcus.* System. Appl. Microbiol. **7:** 18-25.

Wiedemann, L., and R. Perry. 1984. Characterization of the expressed gene and several processed pseudogenes for the mouse ribosomal protein L30 gene family. Mol. Cell. Biol **4:** 2518-2528.

Wigboldus, J.D. 1987. cDNA and deduced amino acid sequence of *Drosophila* rp21c, another 'A'-type ribosomal protein. Nucleic Acids Res. **15:** 10064.

Williams, J.G., and P.J. Mason. 1985. Hybridization in the analysis of RNA, p. 139-160. *In* B.D. Hames, and S.J. Higgins (eds.). Nucleic acid hybridization: a practical approach. IRL Press, Oxford.

Williamson, S.E., and W.F. Doolittle. 1983. Genes for $tRNA^{ile}$ and $tRNA^{ala}$ in the spacer between the 16S and 23S rRNA genes of a blue-green alga: strong homology to chloroplast tRNA genes and tRNA genes of the *Escherichia coli rrn D* gene cluster. Nucleic Acids Res. **11:** 223-235.

Wilson, K.S., K. Appelt, J. Badger, I. Tanaka, and S.W. White. 1986. Crystal structure of a prokaryotic ribosomal protein. Proc. Natl. Acad. Sci. U.S.A. **83:** 7251-7255.

Wilson, C., T. Connolly, T. Morrison, and R. Gilmore. 1988. Integration of membrane proteins into the endoplasmic reticulum requires GTP. J. Cell Biol. **107:** 69-77.

Wittmann, H.G. 1983. Architecture of prokaryotic ribosomes. Ann. Rev. Biochem. **52:** 35-65.

Wittmann, H.G. 1986. Structure of ribosomes, p. 1-27. *In* B. Hardesty and G. Kramer (eds.), Structure, function and genetics of ribosomes. Springer-Verlag, New York.

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

Woese, C.R. 1987. Bacterial evolution. Microbiol. Rev. **51:** 221-271.

Woese, C.R. 1980. Just so stories and Rube Goldberg machines: speculations on the origin of the protein synthetic machinery, p. 357-373. *In* G. Chambliss, G.R. Craven, J. Davies, K. Davis, L. Kahan, and M. Nomura (eds.), Ribosomes: structure function and genetics. University Park Press, Baltimore.

Woese, C.R., and G.E. Fox. 1977a. The concept of cellular evolution. J. Mol. Evol. **10:** 1-6.

Woese, C.R., and G.E. Fox. 1977b. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. U S.A. **74:** 5088-5090.

Woese, C.R., and G.J. Olsen. 1986. Archaebacterial phylogeny: perspectives on the urkingdoms. System. Appl. Microbiol. **7:** 161-177.

Woese, C.R., and R.S. Wolfe (eds.). 1985. The Bacteria VIII: Archaebacteria. Academic Press, New York.

Wolters, J., and V.A. Erdmann. 1988. Compilation of 5S rRNA and 5S rRNA gene sequences. Nucleic Acids Res. **16:** r1-r10.

Wool, I.G. 1986. Studies of the structure of eukaryotic (mammalian) ribosomes, p. 391-411. *In* B. Hardesty and G. Kramer (eds.), Structure, function and genetics of ribosomes. Springer-Verlag, New York.

Wool, I.G. 1980. The structure and function of eukaryotic ribosomes, p. 797-824. *In* G. Chambliss, G.R. Craven, J. Davies, K. Davis, L. Kahan, and M.

Nomura (eds.), Ribosomes: structure function and genetics. University Park Press, Baltimore.

Wool, I.G., and G. Stöffler. 1974. Structure and function of eukaryotic ribosomes, p.417-460. *In* M. Nomura, A. Tissieres, and P. Lengyel (eds.). The ribosome. Cold Spring Harbor Laboratory, Cold Spring Harbor.

Woolford, J.L.,Jr., L.M. Hereford, and M. Rosbash. 1979. Isolation of cloned DNA sequences containing ribosomal protein genes from *Saccharomyces cerevisiae.* Cell **18:** 1247-1259.

Woolford, J.L.,Jr., and M. Rosbash. 1981. Ribosomal protein genes rp39 (10-78), rp39 (11-40), rp51, and rp52 are not contiguous to other ribosomal protein genes in the *Saccharomyces cerevisiae* genome. Nucleic Acids Res.**9:** 5021-5036.

Yaguchi, M., A.T. Matheson, L.P. Visentin, and M. Zuker. 1980. Molecular evolution of the alanine-rich, acidic ribosomal A protein, p.585-599. *In* S. Osawa, H. Ozeki, H. Uchida, and T. Yura (eds.). Genetics and evolution of RNA polymerase, tRNA and ribosomes. Univ. of Tokio Press, Tokio.

Yang, D., B.P. Kaine, and C.R. Woese. 1985. The phylogeny of archaebacteria. System. Appl. Microbiol. **6:** 251-256.

Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC 19 vectors. Gene **33:** 103-119.

Yanofsky, C., T. Platt, I. Crawford, B. Nichols, G. Christie, H. Horowitz, M. Van Cleemput, and A. Wu. 1981. The complete nucleotide sequence of the tryptophan operon of *Escherichia coli.* Nucleic Acids Res. **9:** 6647-6668.

Yonath, A., K.R. Leonard, S. Weinstein, and H.G. Wittmann. 1987. Approaches to the determination of the three-dimensional architecture of ribosomal particles. Cold Spring Harbor Symp. Quant. Biol. **52:** 729-741.

Zantema, A., J.A. Maassen, J. Kriek, and W. Möller. 1982. Fluorescence studies on the location of L7/L12 relative to L10 in the 50S ribosome of *Escherichia coli.* Biochemistry **21:** 3077-3082.

Zillig, W. 1987. Eukaryotic traits in archaebacteria: could the eukaryotic cytoplasm have arisen from archaebacterial origin? Ann.N.Y. Acad. of Sciences **503:** 78-82.

Zillig, W., I. Holz, H. Klenk, J. Trent, S. Wunderl, D. Janekovic, E. Imsel, and B. Hass. 1987. *Pyrococcus woesei*, sp. nov., an ultra-

thermophilic marine archaebacterium, representing a novel order, *Thermococcales*. System. Appl. Microbiol. **9:** 62-70.

Zillig, W., H.P. Klenk, P. Palm, H. Leffers, G. Pühler, F. Gropp, and R.A. Garrett. 1989a. Did eukaryotes originate by a fusion event? Endocyt. Cell Res. in press.

Zillig, W., H.P. Klenk, P. Palm, G. Pühler, F. Gropp, R.A. Garrett, and H. Leffers. 1989b. The phylogenetic relations of DNA-dependent RNA polymerases of archaebacteria, eukaryotes and eubacteria. Can. J. Microbiol. **35:** 73-80.

Zillig, W., P. Palm, and A. Heil. 1976. Function and reassembly of subunits of DNA-dependent RNA polymerase, p. 101-126. *In* R. Losick, and M. Chamberlin (eds.). RNA polymerase. Cold Spring Harbor Laboratory, Cold Spring Harbor.

Zillig, P. Palm, W. Reiter, F. Gropp, G. Pühler, and H. Klenk. 1988. Comparative evaluation of gene expression in archaebacteria. Eur. J. Biochem. **173:** 473-482.

Zillig, W., R. Schnabel, F. Gropp, and W. Reiter. 1985a. The evolution of the transcription apparatus, p. 45-72. *In* K.H. Schleifer, and E. Stackebrandt (eds.). Evolution of Prokaryotes. Academic Press, New York.

Zillig, W., K.O. Stetter, R. Schnabel, and M. Thomm. 1985b. DNA-dependent RNA polymerases of the archaebacteria, p. 499-524. *In* C. R. Woese, and R.S. Wolfe (eds.), The Bacteria VIII: Archaebacteria. Academic Press, New York.

Zillig, W., K.O. Stetter, S. Wunderl, W. Schulz, H. Priess, and I. Scholz. 1980. The *Sulfolobus-"Caldariella"* group: taxonomy on the basis of the structure of DNA-dependent RNA polymerases. Arch. Microbiol. **125:** 259-269.

Zimmermann, R.A., D. L. Thurlow, R.S. Finn, T.L. Marshand, and L.K. Ferrett. 1980. Conservation of specific protein-RNA interactions in ribosome evolution. *In* S. Osawa, H. Ozeki, H. Uchida, and T. Yura (eds.). Genetics and evolution of RNA polymerase, tRNA and ribosomes. Univ. of Tokio Press, Tokio.

Zurawski, G., and S.M. Zurawski. 1985. Structure of the *Escherichia coli* S10 ribosomal protein operon. Nucleic Acids Res. **13:** 4521-4526.

APPENDIX

Sequence of the primers used to sequence the 6.9 Kb Eco RI - Bam HI Fragment

1. 5' GTCAACTATACTAGCTA 3'
2. 5' CAGTGGACGAAGTTAG 3'
3. 5' GCGTATATGTACTCCAT 3'
4. 5' CGCTGTTGCCTCTTCAA 3'
5. 5' GCCYATGATAATGGAGT 3'
6. 5' GCCAGGAGATAAGGCAG 3'
7. 5' GTTACCTTTATGTCAGC 3'
8. 5' ACCTTTGCTTTATTCTC 3'
9. 5' ACCAAGGATCAGCCACAA 3'
10. 5' GTGCCATAGATTCCTGG 3'
11. 5' AACGATTCTTTATCCGC 3'
12. 5' ACTGCAAGGTCAATAGG 3'
13. 5' CGCTCAAGAGCCCTCAG 3'
14. 5' CCAGGTCCACCATTGGG 3'
15. 5' CCCAATGGTGGACCTGG 3'
16. 5' GCAGGCTCAAGTTGTG 3'
17. 5' CCTCCATTGATAGCACC 3'
18. 5' GGTGGAATAGCTCTTTC 3'
19. 5' GTCTATATGCATTCTCC 3'
20. 5' GCTGCTGCACAAGAGCA 3'
21. 5' TGCTCTTGTGCAGCAGC 3'
22. 5' CCTCAACTTCCTCTCTTC 3'
23. 5' CAGTTGGCTGAAGTATC 3'
24. 5' AGTGATCCACAAGCAGC 3'
25. 5' GCTGCTTGTGGATCACT 3'
26. 5' GCCAGGAAACTACTGCT 3'
27. 5' CCGCAACAGGCAACAAC 3'
28. 5' GACGCAACTGAAGAAGA 3'

Sequence of the 6.6 Kb Eco RI - Bam HI Fragment. Initiation codons are underlined, termination codons are overlined.

```
ORF 1
GAATTCAAGCACAAATAACTGATGATGAACTAAAACAAATATTAGCTCAA     50

CTTAACTCTCAAACCAGAAAAGATTATAAGATTACGATTAAAGAGAGAGG     100
      Sso L46
ATGGAAATGAGCAAGCATAAGTCCTTAGGCAAAAAATTGAGACTAGGTAA     150

AGCGTTAAAAAGAAACTCTCCTATTCCTGCTTGGGTCATAATAAAAACTC     200

AAGCTGAGATAAGGTTTAATCCACTAAGAAGAAATTGGAGAAGAAATAAT     250
                 ORF 2
TTAAAGGTATGAATAAATGAGCCAGGAAACTACTGCTACTAACAAGAAGA     300

GCAAAAAACATCAGAGCTACAACAACAAGAAGAAAGAGGAACAAAAACCG     350

CAACAGGCAACAACTACAACGAAAGAGGAAAAGAAAACTAAGCCAGAGAA     400

AGAGAACTTTGAAATGGTTATAAACTTTAGAAGAGTAATAATGGGAAGAA     450

AAACTACCAGAACTAAAAGAGCCATAAAGTATGTGAGATATATCCTAAAG     500

AGACATTTTGGAGCGGAAAAAGTAATTATTGATCCACTATTAGCAAACGC     550

AATAACCATGAATGGTAGGGATAAAATAGTAAGAAGAGTAAGGATAGCTG     600

TCAAAAGAATCGGAGAAAAAAACATATTTAGCTAGACTTGCAATTAAGAG     650
   ORF 3
TGAATGATCTTAGATAAGTTATGCTTTTTGGTACTAATGATAAATATACA     700

ATTATACCTAAAAATAGACGACAAAGAAGTAATAGAGAAAATTCAAGGAA     750

TACTGAAAACAGAAATAATACAGACTACCATATCTAAGAGCGTGCTAGTA     800

GGTATTCTGGTTACAGGAAATAATGATGTAATTCTCCTACCTAGAACAGC     850

TCTGGCAGATGAAATAAAGGTCATAAAGGAACAGGCTAAAGACGTTAGAG     900

TCGAGGTTGTGGATATTAGACCTACTGCTTTAGGAAATATCATATTATCC     950

AACACGCATGGTGCACTTATTTACCAAGATCTTTCTAGGCGGAAATAAAT     1000

AAGGTAAAGAAAGCATGCAGATGATACTGCAATTAAGGGTACAATAGCAA     1050
                    ORF4
ATATAATTACAGTAGGATCATGGCAGTAATTACAGACAAGGCTGGTTTAG     1100

TACATATTGACGCAACTGAAGAAGAGTTGAAAAAATATGTACTAAACCAG     1150

CCAGCCAGTGAATTATTTAAAGTAAAGTTAGATTCTGGAACTGTAAATTT     1200
```

```
TGGGAGCGTCTTTATAAGGAGCGGATTAGTGGCTAACAGAAATGGAGTT      1250
CTAGTAGGTTCCTCAACAACGGGACAGAGATTTTAAGAATCCAAAGAGCA     1300
                                           Sso LX
TTTAGTGATCAGTATGGCTGAAGTAAAAATTTTCATGGTCAGAGGAACTG     1350
CCATATTTAGTGCGTCAAGATTTCCTACAAGTCAAAAATATGTTAGAGCT     1400
TTAAATGAAAAACAAGCAATCGAATACATTTATAGTCAACTTGGTGGAAA     1450
AAATAAAATTAACGATACAACATACACATACAAGAGATCAAAGAAGTTAA     1500
GGAAGATGAAATCACAGACAAGACAATAAGAGATTTAGCAAAGCTAGATA     1550
                       ORF 5
AAATTATAATGTGATAGAAATGAGTGAAGAGGCAGAGCAACAACAAGCGG     1600
CAGAATATATAGCATACCTATATGATCAAGCATCAGCACTAAGACAATAT     1650
ATAGATACCCTTCAGAAGAACCTGGCAGAGGTACTAGAATCCTTAGAAGC     1700
AGTAAGAGCATCAAAAAGTGCCGTAGACGAGATAGGTAAGGAAAACCAGG     1750
AATACTTGTTATTTGGCGACAGAAAGGGCAATATAGTATTCAAGGTAAAC     1800
AGTGTAGATAAAAATAAGGTTCTAATACACTTAGGATTAAATTATTACGC     1850
AGAAGTGATCCACAAGCAGCTAAGAAAATTCTGGATGATAGAGAGCAACA     1900
GTTGGCTGAAGTATCTAAAAATATACAAGGTGAATTATCTAAGTCAATA      1950
GACGCTTATAATCAAATTGCAGAAATTCTTTCCCAAGTGCAAGGTAAAAA     2000
    Sso "docking protein" gene
AGGTGAATAATTTGTTTTGATAGATTAAAAAAAGCATTCTCAAATTTTTT     2050
AGATAAAATTAGTGGAGAAGAGAACAAAAAGGAACCTGAAACTAGACAAA     2100
CAGATCAACTAGAAAGTAAAAAAGAAGAGACTATACAACAACAGCAAAAT     2150
GTACAACAACCACAAGCGGAAAATAAAATTGAGCAAAAACAGGAAAAAAT     2200
TTCAGTTCAAACAGGTCAAGAAAATAAGCAGGAGAATAAACGATCCTTCT     2250
TTGATTTCCTGAAGTATAAAACGATAAAAGAGGACGATCTTAATGACGTT     2300
ATAGAAGAACTCAGATTTCAACTTCTTGATTCTGATGTATCTTATGAAGT     2350
TACGGAAAAAATACTAGAAGATCTCAAAAATAATCTGATAGGGAAAAAGG     2400
TTAGTAGAAGAGAGGAAGTTGAGGAAATAGTAATAAATACTCTAAAAAAA     2450
TCTATAACAGAGATTTTAACAAAAAATCAAAAAACAGATTTGATAGAAAA     2500
AATAAGAAGTAGTGGAAAGAAGCCATTTGTAATAATATTCTTTGGAGTTA     2550
```

ATGGAGTAGGTAAAACTACTACTATAGCAAAAGTAGTAAATATGCTGAAG 2600

AAAAATAACTTATCTACAATTATTGCAGCCTCAGATACATTTAGAGCTGC 2650

TGCACAAGAGCAACTAGCTTATCATGCATCAAAATTAGAGGTTCAACTAA 2700

TAAGAGGAAAATATGGTGCTGATCCAGCGTCGGTAGCATTCGATGCCATA 2750

AG TTTTGCTAAGAGTAGAAATATAGATGTAGTACTTATTGATACTGCTGG 2800

GAGAATGCATATAGACTCTGATTTAGTTGAAGAACTTAAGAGGGTTTTAA 2850

GGATAGCAAAACCTGATTTTAGGATCTTAATATTAGACTCGCTAGCAGGA 2900

AGTGATGCATTAGAACAGGCTAGACATTTTGAAAATAATGTTGGATATGA 2950

TGCAGTAATTCTTACCAAAGTAGATGCAGATGCTAAAGGTGGAATAGCTC 3000

TTTCATTAGCGTATGAATTGAAAAAACCTGTAGTATATATGGGAGTAGGA 3050

CAGAATTATGATGATTTAATTCCCTTCTCACCAGACTGGTTTGTAGAAAG 3100

ORF 6

GATATTCAGTAGTTAAAGTTAAAATTCGTAAATGAAAGTTTTTAGTATAA 3150

TGAAAATTTCTGACATTATCAAAAGGCTACGAGAAGACTGGAAGAGAATA 3200

ATTAGTGTAGCCAAGAAACCAGACAAAGATTCATTTAATTACAGTATTAG 3250

GCTTACACTTTTAGTTATGGCTGTAGTTGGTCTCATCGCTTACATTGTAC 3300

ORF 7

AGTTAACTACATCACTGATAATAAGGTGAAAAAGGTGCTATCAATGGAGG 3350

ACTTCAAGTACAGAAATTATTACGTGTTAAGAGTTACAGGAGGACAAGAA 3400

ATTAACGTGGCTCTTATTTTAGAAGAAAGAATAAAAACAAATAATATAAA 3450

TGAAATCTTTTCTGTAGTAGTCCCACCTAATATAAAGGGTTATGTGATAC 3500

TTGAAGCTACCGGACCTCATGTAGTAAAATTAATTTCATCAGGAATAAGA 3550

CACGTTAAAGGAGTAGCTCATGGACTAATTCAGAAAGAAGATGTTACAAA 3600

ATTTGTTTCTAAGTCAGTTGCATTACCCGCTGTGAAAGAAGGAGACTTAG 3650

TAGAGGTAATTAGTGGTCCATTTAGAGGTATGCAGGCTCAAGTTGTGAGA 3700

GTAGAATCAACAAAAAATGAAGTAGTTTTAAATATTTTAGAATCCTCATA 3750

TCCTGTTCAGGTTACAGTTCCTTTGGAACAAGTTAAACCTGTTAAGAGGT 3800

Sso L11

GACAGATAATTGCCTACTAAGACAATAAAAATAATGGTAGAAGGAGGTAG 3850

TGCTAAACCAGGTCCACCATTGGGACCTACCCTATCACAATTAGGACTTA 3900

```
ATGTTCAGGAAGTTGTTAAAAAAATAAATGATGTTACAGCACAATTTAAA     3950
GGAATGTCACTACCAGTCACAATTGAGATCGATAGTTCAACTAAAAAATA    4000
TGATATAAAAGTTGGAGTACCTACAACAACATCACTATTACTTAAAGCAA    4050
TAAACGCTCAAGAGCCCTCAGGAGATCCAGCCCACAAGAAGATAGGCAAT    4100
TTAGACTTAGAACAAATTGCTGACATAGCCATTAAGAAAAAGCCTCAACT    4150
TAGCGCAAAAACTCTAACAGCTGCAATAAAGAGTTTACTAGGAACTGCAA    4200
GGTCAATAGGTATAACTGTTGAGAAGGAAAAGATCCTAAAGATGTAATAA    4250
AAGAGATTGATCAAGGAAAATACAATGACTTATTAACTAATTATGAGCAA    4300
                           Sso L1
AAATGGAATGAAGCAGAAGGGTGAAGAAAGTGTTAGCGGATAAAGAATCG    4350
TTGATAGAAGCCTTAAAACTGGCACTTAGTACGGAGTATAATGTAAAAAG    4400
GAACTTTACACAAAGTGTGGAAATTATACTTACGTTTAAAGGAATTGATA    4450
TGAAGAAAGGTGACCTTAAACTTAGAGAAATAGTTCCCTTACCAAAACAG    4500
CCATCTAAAGCAAAGAGGGTTCTTGTTGTTCCATCCTTCGAACAACTCGA    4550
ATATGCTAAAAAAGCATCTCCCAATGTTGTGATAACTAGAGAAGAATTAC    4600
AAAAATTACAAGGACAAAAAAGACCAGTGAAAAAATTAGCTATACAAAAT    4650
GAGTGGTTTTTAATTAACCAGGAATCTATGGCACTGGCAGGAAGGATATT    4700
AGGACCCGCCTTAGGACCTAGAGGTAAGTTTCCTACACCTCTACCCAATA    4750
CAGCTGATATCACTGAATATATTAATAGATTCAAGAGATCAGTTATAGTA    4800
AAGACCAAGGATCAGCCACAAGTTCAAGTATTCATAGGAACTGAGGACAT    4850
GAAGCCTGAAGACTTAGCTGAGAATGCGATAGCCGTATTAAATGCAATAG    4900
AGAATAAAGCAAAGGTAGAAACTAATCTAAGAAATATTTATGTTAAAACA    4950
                                     Sso L10
ACAATGGGTAAAGCTGTAAAAGTTAAAAGAGCGTGATCGGTTTGGCAGTT    5000
ACTACAACTAAAAAAATAGCAAAATGGAAAGTAGATGAAGTAGCTGAACT    5050
TACTGAAAAATTAAAAACCCATAAAACTATAATAATAGCTAATATAGAGG    5100
GTTTTCCTGCAGATAAACTTCATGAAATAAGGAAAAAATTAAGAGGAAAA    5150
GCTGACATAAAGGTAACTAAAAACAACTTATTTAACATAGCACTCAAAAA    5200
TGCAGGGTATGACACTAAATTATTTGAAAGCTATCTAACTGGACCTAATG    5250
```

CCTTTATATTTACTGATACTAATCCCTTCGAATTACAGTTATTCCTATCA 5300

AAATTTAAATTAAAGAGATATGCCCTGCCAGGAGATAAGGCAGATGAGGA 5350

GGTTGTTGTACCAGCAGGAGATACAGGTATTGCAGCAGGTCCCATGCTCA 5400

GTGTTTTCGGAAAGTTGAAAATAAAAACAAAAGTCCAGGATGGAAAAATT 5450

CACATATTACAAGACACAACAGTTGCAAAGCCTGGAGATGAGATACCAGC 5500

AGATATTGTACCAATATTACAAAAACTAGGAATAATGCCAGTATATGTAA 5550

AATTAAACATTAAGATAGCCTATGATAATGGAGTAATAATCCCAGGAGAT 5600

AAATTATCAATTAACCTTGATGATTATACTAATGAAATCAGAAAAGCTCA 5650

TATTAACGCATTTGCAGTAGCCACAGAAATAGCGTACCCAGAACCAAAGG 5700

TCCTTGAATTTACAGCAACCAAAGCTATGAGAAATGCATTAGCTCTAGCC 5750

AGTGAGATAGGTTATATAACTCAAGAGACAGCACAGGCAGTATTTACTAA 5800

AGCAGTAATGAAAGCATACGCTGTTGCCTCTTCAATAAGCGGAAAAGTAG 5850

ATTTAGGAGTTCAAATACAAGCACAACCACAGGTCTCTGAACAGGCAGCT 5900

GAGAAGAAAGAAGAAAAGAAAGAAGAAGAAAAGAAAGGACCAAGCGAAGA 5950

AGAGATAGGCGGAGGCTTATCATCACTATTCGGTGGATGAGTTTAAATAT 6000

Sso L12

CACCTCTTTATGTATTTTAAGGTGATTATATATGGAGTACATATACGCAA 6050

GTTTACTATTGCACGCAGCTAAGAAAGAAATAAGTGAGGAAAACATAAAG 6100

AATGTATTAAGTGCAGCAGGAATAACAGTGGACGAAGTTAGACTAAAAGC 6150

TGTAGCAGCTGCTTTAAAAGAAGTAAATATAGACGAAATATTGAAGACAG 6200

CTACTGCAATGCCTGTAGCAGCAGTAGCAGCCCCTGCAGGTCAACAGACA 6250

CAGCAGGCAGCTGAGAAGAAAGAAGAAAAGAAAGAAGAAGAAAAGAAAGG 6300

ACCAAGCGAAGAAGAGATAGGCGGAGGCTTATCATCACTATTCGGTTAAT 6350

Sso *ala S*

AAGAAAGACTTTTTTACCATTTCAAATATCATTTTTACCATGGTCAAAGC 6400

AAACGAGAACGAGTACAGGTTAAACTTTTTTCTGTCAAGAGGATATGAAA 6450

GAAAAATTTGTAGATCCTGTTCTACCCCATTTTGGACTTTAGATAAGTCA 6500

AAGGAAAATTGTTCAGATGTTCCTTGTACTGACTATTACTTCTTTGATAT 6550

CAAGATTAAATCTCCCCCGTTAACAGTGAGAGAATCAAGAGAAAAATTTC 6600

TAAGATTCTTTGAAAAGAGAGGACACACCATAGTTCCACCTAAACCTGTA 6650

GTTGCCAAGTGGAGAGATGATCTTTATCTTACTATAGCTAGTATAGTTGA 6700

CTTTCAACCATTTGTAACGAGCGGATTAGCTAAACCGCCTGCAAATCCAC 6750

TGGTAGTTTCTCAGCCATGTATAAGACTAGAAGATGTGGATAACGTAGGA 6800

ATAACATTTGGAAGGCATTTGACCACATTTGAAATGGCTGCTCATCATGC 6850

ATTCAATTATCCTGATAAACAAATTTATTGGAAAGACGAAACAGTAGCA 6900

TTGCTAAGGAGTTCTTTACAGAAGAACTAGGGATCC 6936