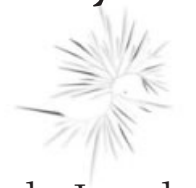
Archaea: Evolution, Physiology, and Molecular Biology
Edited by Roger A. Garrett, Hans-Peter Klenk

19

Translational mechanisms and protein synthesis



Paola Londei

Introduction

Translation is the last step of the gene expression pathway, whereby the genetic message carried over from the DNA in the form of a messenger RNA (mRNA) is finally converted into the final product, a polypeptide chain. Given its fundamental importance for all cells, translation is well conserved in its basic aspects across the three domains of life. Nevertheless, it has been known for a long time that the protein synthetic machineries of the Bacteria and the Eukarya present clear-cut differences in composition and complexity. Until recently, conventional wisdom had it that there were two different versions of the translational apparatus: the simple, streamlined one possessed by the Bacteria and the more complex one found in eukaryotic cells. This dichotomy seemed to be the logical result of the different organization and lifestyles of eukaryotic and prokaryotic cells. The "simpler" Bacteria, whose basic evolutionary strategy consisted of maximizing the velocity of growth and multiplication, had gene expression machineries made of fewer and, often, smaller components. The "complex" eukaryotic cells, many of which live in a highly integrated environment, had to add much sophistication to the basic design of their gene expression apparatus.

The discovery of the third domain of life, the Archaea, has challenged in many ways the classical textbook dichotomy between prokaryotes and eukaryotes. As far as cellular organization is concerned, all known Archaea are unicellular prokaryotes. However, a host of phylogenetic, molecular, and genomic studies have now shown that the Archaea are clearly separated from the other prokaryotic kingdom, the Bacteria. Moreover, the rooting of the universal tree of life has revealed that the Archaea have shared a tract of evolutionary path with the eukaryotes before branching out as a separate domain (Iwabe et al., 1989). This is reflected in the presence, at the molecular level, of many intriguing similarities between the Eukarya and the Archaea, regarding especially the structure and composition of the cellular machineries for gene expression. In recent years, for instance, an elegant series of experimental studies have shown that the archaeal transcriptional apparatus can be regarded as a "basic" or simplified version of the eukaryal one. The structure of the archaeal translational apparatus also exhibits an unexpected complexity, including several components, especially translational factors, that are found in Eukarya but have no counterparts in Bacteria (Bell & Jackson, 1998). Therefore, the study of the translational machinery in Archaea is not only interesting in its own right, but promises to yield new and exciting insights into the evolutionary history of the protein-synthetic mechanism. This chapter presents a survey of the most recent advances in the understanding of archaeal translation, focusing especially on the data that highlight the novel and Eukarya-like features of the archaeal translational apparatus.

Archaeal ribosomes

It has been known for a long time that archaeal ribosomes are composed of subunits that sediment at 30S and 50S and contain 16S, 23S, and 5S rRNAs, thus resembling their bacterial counterparts. However, the existence of certain similarities between archaeal and eukaryal ribosomes was suggested over two decades ago. Compositional analyses of the ribosomes of various archaeal species revealed that several of them were protein-richer than the bacterial particles, especially as far as the small subunit was concerned. Moreover, the ribosomes of certain sulfur-dependent thermophiles were shown to have morphological characteristics reminiscent of the eukaryal particle (Amils et al., 1993, and references therein).

Nowadays, the composition of archaeal ribosomes can be analyzed extensively taking advantage of the over 20 complete genomic sequences representative of archaeal diversity. Genomic studies have essentially confirmed the early surmise that archaeal ribosomes, although being closer in size to bacterial ones, have specific affinities with their eukaryal counterparts. A survey of the composition of archaeal ribosomes is presented in Table 19.1.

Archaeal ribosomal RNAs generally resemble bacterial ones in both size and structure. Most small subunit (SSU) rRNAs are 1400–1500 nucleotides in size; the smallest one, with 1344 nucleotides, is found in the parasitic species *Nanoarchaeum equitans* (Table 19.1). The large subunit (LSU) rRNAs have sizes comprising between 2850 and 3100 nucleotides, while 5S rRNAs have 119–132 nucleotides. Interestingly, the genome of the crenarcheon *Aeropyrum pernix* is unique in containing a 167 nucleotide long homolog of 5.8S rRNA, an rRNA specific to eukaryal LSU. However, there are no experimental observations available to confirm the presence of a separate 5.8S rRNA in *A. pernix* large ribosomal subunits. Both LSU and SSU archaeal rRNAs may contain introns (Lykke-Andersen et al., 1997).

Unlike the majority of Bacteria, most Archaea are endowed with single copies of the rRNA-coding genes. Exceptions are found among the methanogens and halophiles, which may have up to three copies of 16S and 23S RNA genes and up to four copies of 5S RNA genes (Table 19.1). The 16S and 23S rRNA genes are adjacent and co-transcribed in most Archaea; the exceptions are the Thermoplasmales and *N. equitans*, where these genes are located far apart in the genome and are transcribed independently. Also, the Archaea are unlike the Bacteria (and like the Eukarya) in having 5S RNA genes normally unlinked from the genes encoding the 16S and 23S rRNAs. Only the halophiles and some methanogens have Bacteria-like rRNA operons, including the 5S RNA genes and often also tRNA genes inserted as spacers between the rRNA genes. However, extra copies of the 5S RNA genes may be found unlinked from the rRNA operons in several species (e.g. *Methanococcus maripaludis*).

Where the similarity of archaeal ribosomes to the eukaryal ones becomes most apparent is in their complement of proteins. Archaeal genomes include a total of 68 r-protein families, 28 belonging to the SSU and 40 to the LSU (Lecompte et al., 2002). Thirty-four of these (15 in the SSU and 19 in the LSU) are universal proteins, having identifiable homologs in the ribosomes of all three domains of life. Another 33 (13 SSU and 20 LSU) are specifically shared by archaeal and eukaryal ribosomes and are not found in Bacteria, while no r-proteins are shared by the Bacteria and the Archaea to the exclusion of Eukarya. Only one r-protein (LXa in the LSU) is unique to the Archaea (Lecompte et al., 2002). These data show that, as far as their protein composition is concerned, archaeal ribosomes can be envisaged as a somewhat smaller version of the eukaryotic particles, which contain 78 r-protein families, including all of the 68 found in Archaea plus another ten Eukarya-specific ones. By contrast, bacterial ribosomes have altogether 57 r-protein families, 23 of which are exclusive of the Bacteria (Lecompte et al., 2002).

Interestingly, the protein composition of archaeal ribosomes is not constant over the domain, but presents a certain heterogeneity that correlates with the branching order of the various species (Lecompte et al., 2002). Thus, the deep-branching Crenarcheota tend to have protein-richer ribosomes than the Euryarchaeota (Table 19.1). The protein-richest ribosome is that of the crenarcheon *A. pernix*, which is endowed with the full complement of 68 r-proteins. On the other hand, the ribosomes of the late-branching halobacteria and Thermoplasmales have only 58 proteins, thus coming closer to the bacterial size. This evolutionary trend toward a "lighter" ribosome is not observed in the Bacteria or the Eukarya, where the protein composition of ribosomes remains essentially

Table 19.1 Ribosome composition in Archaea.

	23S	16S	5S	SSU prot.	LSU prot.
Crenarcheota					
*A. pernix**	4413†	1423	19/132	28	39
P. aerophylum	3024	2210‡	130	28	38
S. solfataricus	3049	1496	121	28	37
S. tokodaii	3012	1445	125	28	37
Euryarchaeota					
Thermococcales					
P. furiosus	3048	1446	121/125	25	37
P. horikoshii	3857	1494	121 (2)	25	37
P. abyssi	3017	1502	121 (2)	25	37
T. kodakaraensis	3028	1497	125 (2)	25	37
Methanopyrales					
M. kandleri	3097	1511	132	25	37
Methanobacteriales					
M. thermoautotrophicus	3028/3034	1478 (2)	126/128	25	36
Methanococcales					
M. jannaschii	2889/2948	1474/1477	119 (2)	25	36
M. maripaludis	2956 (3)	1391 (3)	114 (4)	25	36
Methanosarcinales					
M. mazei	2892 (3)	1473 (3)	134 (2) 132 (1)	25	34
M.acetivorans	2831 2848 2948	1429 (3)	134 (2) 132 (1)	25	34
Archeoglobales					
A. fulgidus	2931	1491	123	25	34
Halobacteriales					
H. marismortui	2923 (3)	1471 (3)	121 (3)	25	32
Halobacterium sp. NRC-1	2905	1472	122	25	32
Thermoplasmata					
T. acidophilum	3044	1470	122	25	32
T. volcanium	2906	1469	122	25	32
Nanoarcheota					
N. equitans	2861	1344	122		

The figures indicate the number of nucleotides in each rRNA gene. The number of genes, if more than one, is indicated in parentheses, except when the genes are of different length, in which case their size is shown in full.

*Has a 5.8S of 167 nt; †gene containing two introns; ‡gene containing one intron.

constant, excepting only some parasitic species where a few proteins may be missing (Lecompte et al., 2002). The functional significance of ribosomal protein loss in late-branching Archaea is unclear at present. However, it should be observed that none of the dispensable proteins belongs to the set of universally conserved ones, which are probably essential for an efficient ribosomal function in all cells.

The organization of ribosomal protein (r-protein) genes in archaeal genomes presents very interesting aspects. In Bacteria, about half of the r-protein genes are clustered in the two large spectinomycin (spc) and S10 operons, whose structure is largely conserved in most species. Likewise, in the Archaea over one-third of the r-protein genes are included in a few large clusters closely resembling the bacterial spc and

S10 operons in the type and order of genes. Most of the proteins belonging in these clusters are universal ones, suggesting that this genetic organization was already present in the last universal common ancestor (LUCA) of extant cells and predated the radiation of the three primary domains. The alternative hypothesis is that similar gene clustering is due to convergent evolution, namely positive selection because of some functional advantage independently operating in both prokaryotic domains. However, any such advantage is not immediately apparent, at least in present-day organisms, since the clusters can be broken, and frequently are, in both Bacteria and Archaea. Moreover, no information is available about the transcription patterns of the archaeal spc- and S10-like clusters, making it difficult to tell to what extent they are organized into functional operons that may resemble the bacterial ones.

However, other clusters of ribosomal protein genes are known to be organized and also regulated in a way similar to that observed in Bacteria. A well studied case is the *Methanococcus vannielii* L1 ribosomal protein operon, encoding the r-proteins L1, L10, and L12, which is transcribed as a single polycistronic mRNA (Kraft et al., 1999). This operon is subjected to autogenous translational regulation, namely its translation can be repressed by the protein encoded by the first cistron (L1). The regulatory protein L1 is a 23S rRNA-binding protein that under normal conditions interacts preferentially with its binding site on the ribosomal RNA. However, when in excess because of blocked or reduced ribosome synthesis, L1 also binds to a specific regulatory target site of its mRNA, thereby inhibiting translation of all three cistrons in the operon. The regulatory mRNA site, a structural mimic of the rRNA binding site for L1, is located within the L1 gene about 30 nucleotides downstream of the ATG initiation codon (Kraft et al., 1999). A similar regulatory mechanism also exists in *Methanococcus thermolitotrophicus* and *Methanocaldococcus jannaschii*; however, its presence in other Archaea is more doubtful.

The three-dimensional architecture of archaeal ribosomes has also been explored in some detail. Early electron microscopy observations showed that the ribosomes of sulfur-dependent thermophiles displayed morphological characteristics similar to those of the eukaryotic particles (Lake, 1985). This was especially true for the small ribosomal subunits, which possessed a "bill" on the head and "lobes" on the body similar to those observed in their eukaryal, but not bacterial, counterparts. These features were, however, absent in the ribosomes of halophiles and some methanogens, a fact that correlates well with the larger protein content of the ribosomes of sulfur-dependent themophiles (Table 19.1).

More recently, the three-dimensional structure of the large ribosomal subunit of the halophilic archaeon *Haloarcula marismortui* has been solved at high resolution by crystallographic analysis (Ban et al., 2000). Due to the lack of comparative data, however, it is difficult to discriminate between features that may be specific to archaeal (or halophilic) ribosomes and features that are common to all large ribosomal subunits. As an example, the exceptionally compact and "monolithic" quaternary packing observed in *H. marismortui* 50S particles (Ban et al., 2000) might be due, at least in part, to adaptation to a hypersaline cellular environment. Moreover, *H. marismortui* ribosomes have an RNA and protein content comparable to that of the bacterial ribosomes, making it difficult to detect archaeal-specific features as the position of the extra proteins in the three-dimensional structure. For instance, the one archaeal-specific protein (LXa) is missing from the genome of *H. marismortui* and other halophiles. Nevertheless, some specific observations can be made, such as that concerning the protein L7ae, shared by the Eukarya and the Archaea but not present in Bacteria. L7ae was initially identified as a ribosomal protein; however, it was later found to behave also as a component of the machinery for rRNA post-transcriptional modification. In fact, L7ae has homology with the eukaryal protein snu13p, which is the RNA-binding element of the snoRNPs involved in post-transcriptional modification of the rRNA transcripts (Kuhn et al., 2002). L7ae is clearly seen in the three-dimensional structure of *H. marismortui* 50S subunit, showing that it is a bona fide ribosomal protein; however, in agreement with its multifunctional character, it is located at the periphery of the subunit and is one of the few r-proteins that make contact with only one rRNA domain. Its function in the ribosome is uncertain (Ban et al., 2000). It is to be expected that the unraveling of more structures of whole ribosomal subunits will allow us in the future to learn more about any architectural features that may be unique to archaeal ribosomes.

Archaeal mRNAs

The structure and organization of mRNAs is another aspect of translation that presents clear-cut differences in Bacteria and Eukarya. Bacterial mRNAs are mostly polycistronic, always uncapped and devoid of long poly(A) tails. In addition, bacterial mRNAs are endowed with *cis*-acting sequences, the Shine–Dalgarno (SD) motifs that enhance the efficiency of ribosome binding to the translation initiation regions of the various cistrons in a polycistronic message. In contrast, eukaryal mRNAs are monocistronic, have 5′ cap structures and long poly(A) tails, and lack SD-motifs for mRNA/ribosome recognition. As the Archaea are prokaryotes with compact genomes, they were expected to have mRNAs similar in organization to the bacterial ones. In fact, the first studies of archaeal transcripts had uncovered the presence of polycistronic mRNAs endowed with SD sequences and coordinately translated into several polypeptides (Shimmin et al., 1989). However, recent *in silico* genome-wide studies analyzing the position of transcription start signals, initiation codons, and potential ribosome-binding motifs such as SD sequences have revealed interesting unique aspects of mRNA structure in Archaea.

Several years ago, genomic analyses of the thermophilic Archaea *Sulfolobus solfataricus* and *Pyrobaculum aerophilum* revealed that in these species a large proportion of mRNAs were predicted to be leaderless, i.e. to lack entirely, or almost so, a 5′ untranslated region ahead of the translation start codons (Sensen et al., 1996; Slupska et al., 2001). This constituted an unexpected unique feature of archaeal mRNAs, as leaderless messengers are rarely encountered in both the Bacteria and the Eukarya.

More recent surveys including a larger number of species have extended and refined those initial observations. It has been found that the archaeal genomes so far sequenced form two distinct groups as far as the structure of transcripts is concerned (Torarinsson et al., 2005). Group A genomes, including several (but not all) Crenarcheota, Euryarchaeota such as the Thermoplasmales, halobacteria and *N. equitans*, putatively produce a high proportion (about 50% on average) of leaderless transcripts. In some of these genomes the genes located internally in (presumably) polycistronic transcripts are preceded by clearly identifiable SD motifs. However, in other group A genomes, such as those of *N. equitans* and *P. aerophilum*, the internal cistrons do not appear to possess evident SD-like sequences.

Group B genomes, on the other hand, produce few leaderless transcripts and, accordingly, usually possess SD motifs ahead of the initiation codons of both the first and the internal genes in operons (or of genes in monocistronic transcripts). Group B genomes include a diverse array of species, mostly methanogens but also Crenarcheota as *A. pernix* and *Hyperthermus butylicus* and the Pyrococcales.

The presence of different types of mRNA organization in Archaea – "leaderless" messages as well as messages of more conventional bacterial type – has prompted several speculations about their respective evolutionary status. Some investigators have proposed that leaderless mRNAs are an evolutionary relic, i.e. represent the ancestral kind of mRNA, possibly the one prevalent at the LUCA stage. The most compelling evidence in favor of this hypothesis is that leaderless mRNAs are universally translatable (at least *in vitro*) by archaeal, bacterial, and eukaryotic ribosomes (Grill et al., 2000). Since "normal" eukaryotic mRNAs are poorly, if at all, translated in bacterial systems (and vice versa) this is a remarkable fact that argues for a common conserved mechanism for leaderless translation, predating the branching of the primary domains. Another recent observation supporting the ancestral nature of leaderless mRNAs is that in the protozoan *Giardia lamblia* most mRNAs are leaderless (Li & Wang, 2004). However, the "primitive" status of *G. lamblia* is uncertain: the species occupies a deep branch of the eukaryal evolutionary tree, but this may be an artefact of evolutionary analysis, due to an abnormally fast mutation rate, a frequent occurrence in parasitic organisms like *G. lamblia*.

Further evidence in favor of the ancestral nature of leaderless mRNAs is that their translation, at least in Bacteria, seems to have no stringent requirement for initiation factors, especially if carried out by preformed 70S ribosomes (Udagawa et al., 2004). Since only a very restricted set of IFs is common to all three primary domains (see below), translational initiation in the absence or semi-absence of accessory factors can be envisaged as a primitive condition.

However, there are also data arguing against the "primitivity" of leaderless mRNAs. First, there is the study mentioned above showing that most leaderless-mRNA-rich group A genomes are found in late-branching archaeal species, while early-branching

species tend to have a prevalence of leadered mRNAs with SD motifs (Torarinsson et al., 2005). The latter would therefore represent the likely "ancestral" mRNA structure. If so, the prevalence of leaderless mRNAs in later-evolved, and especially in extremely thermophilic, archaeal species may have a physiological reason that currently escapes our understanding.

Second, there is reason to think that the polycistronic arrangement of genes and the SD motifs predate the branching of the primary domains from the common root of the tree of life (Londei, 2005). In fact, as observed previously for the case of certain ribosomal protein genes, groups of genes clustered (and sometimes transcribed) in the same or a similar order are frequently observed in Bacteria and Archaea. It is very unlikely, albeit not impossible, that this situation is the result of convergent evolution. Interestingly, short-branch Archaea in which most cistrons are endowed with SD motifs, such as *A. pernix* and *H. butylicus*, use AUG, GUG, and UUG as initiation codons in roughly the same proportion, while in most other species AUG is by far the prevalent initiation signal (Torarinsson et al., 2005). This suggests that a "primitive" function of the SD motifs may have been that of ensuring a correct ribosome positioning on the translation initiation region independently of the presence of an optimal initiation codon.

A better understanding of the evolutionary status of leaderless mRNAs may be reached when the mechanism for their translation will be unraveled. This mechanism is likely to be quite distinct from that operating on the leadered mRNAs (Tolstrup et al., 2000). As illustrated in the next section, the notion that the Archaea normally employ two different mechanisms for translating leaderless and leadered mRNAs has recently received some experimental confirmation. These studies are uncovering a very interesting aspect of archaeal translation that may have profound implications for the understanding of the evolution of the mechanism for mRNA/ribosome recognition.

Translational mechanism in Archaea

Initiation

mRNA–ribosome interaction Initiation is the step of translation that has incurred the greatest evolutionary divergence. Eukaryotic ribosomes normally identify the translational start site by a "scanning" mechanism, whereby the 40S subunits, aided by many protein factors and carrying pre-bound met-tRNAi, slide along the message until the initiation codon is found and codon–anticodon interaction takes place. In contrast, bacterial 30S ribosomes can bind directly to the translation initiation regions of individual cistrons by the pairing between the SD sequence on the mRNA and the anti-SD sequence on the SSU rRNA. Only three initiation factors (compared to over a dozen in eukaryotes) participate in bacterial initiation.

The Archaea, as prokaryotes endowed with polycistronic mRNAs, were expected to have a translation initiation mechanism of bacterial type. However, as explained above, the discovery of the abundance of leaderless mRNAs in the third domain of life has suggested the existence of two different mechanisms for archaeal translational initiation.

The first experimental data supporting the notion of two distinct modes of mRNA/ribosome recognition in Archaea were obtained from *in vitro* studies of translation in the crenarcheon *Sulfolobus solfataricus*. The functional relevance of the SD motifs for the decoding of *S. solfataricus* leadered mRNAs was demonstrated by the fact that the disruption of such motifs by site-directed mutagenesis completely inhibited the translation of the following cistrons (Condo et al., 1999). Moreover, it was shown that in a polycistronic mRNA each ORF can be translated independently of the others if preceded at the correct distance by an SD motif.

However, the most interesting fact revealed by the work in *S. solfataricus* was that the *in vitro* translation of the mRNAs whose SD motifs had been disrupted could be rescued by deleting entirely the 5′ untranslated region, thereby rendering the mRNA leaderless (Condo et al., 1999). These results strongly support the notion that a second specific mechanism exists for initiating translation on leaderless messages, operating independently of the SD-motif-based one.

The mechanistic details of ribosome interaction with leadered and leaderless mRNAs are as yet poorly understood. Recent *in vitro* studies carried out with purified translational components of *S. solfataricus* (Benelli et al., 2003) revealed that the 30S ribosomal subunits can interact directly and strongly with leadered mRNAs possessing SD motifs even in the absence of any other translational component,

including initiator tRNA. The leaderless mRNAs, by contrast, were unable to form binary complexes with 30S subunits unless met-tRNAi was added to the samples (Benelli et al., 2003). These results suggest that ribosomal recognition of an initiation codon at the 5′ end of a leaderless mRNA requires codon–anticodon pairing, as previously observed for leaderless mRNA translation in *E. coli* (Grill et al., 2000). It should be pointed out that eukaryotic 40S ribosomes also need to carry met-tRNAi in order to recognize the initiation codon at the end of the scanning process.

The reason why the Archaea should keep two distinct mechanisms for mRNA/ribosome recognition is unclear at present, also because there are not enough data on the molecular details of either of them. An especially important task for future research will be to determine the function of the protein factors involved in archaeal translational initiation.

Translation initiation factors Translation initiation factors (IFs) are very interesting from an evolutionary point of view, since they differ to a large extent in the Bacteria and the Eukarya. Only three IFs exist in Bacteria. The principal one, called IF2, is an RNA-binding G-protein of about 90 kDa that performs the essential task of promoting the correct binding of the initiator tRNA (f-met-tRNA) to the ribosomal P site. The other two factors, IF1 and IF3, assist initiation by hindering premature subunit association (both) and by discouraging recognition of non-optimal initiation codons (IF3) (Gualerzi & Pon, 1990).

The Eukarya, in contrast, have an elaborate set of IFs. The cap-binding factor (termed eIF4F), absent in Bacteria, recognizes the 5′ cap structure and unwinds secondary structures in the mRNA, thus allowing ribosome binding. The preinitiation complex "scanning" the mRNA in quest of the initiator AUG codon is composed of the 40S subunits, met-tRNAi, and the proteins eIF1, eIF1A, and eIF3 (Pestova & Kolupaeva, 2002). eIF1 and eIF1A are both required for the correct identification of the start codon, while eIF3, among other things, connects the ribosome with the cap-binding factor eIF4F. Met-tRNAi binding to the 40S subunits is promoted by the G-protein eIF2, a hetero-trimeric complex not homologous to the bacterial factor IF2 (Kyrpides & Woese, 1998a). Adaptation of met-tRNAi in the P site is accompanied by the hydrolysis of the eIF2-bound GTP, whereupon the factor dissociates from the ribosome. However, eIF2 has no spontaneous GTPase activity and needs a GTPase activator factor, called eIF5, to trigger GTP hydrolysis. Moreover, the reactivation of eIF2-GDP obligatorily requires a GTP/GDP exchange factor, the pentameric protein eIF5B (Kimball, 1999). After the establishment of the codon–anticodon interaction, the factor eIF5B, also a G-protein and a homolog of bacterial IF2, stimulates subunit joining and thereby the formation of the monomeric ribosome 80S (Pestova et al., 2000).

The elaborate mechanism for translation initiation in eukaryotes is usually explained by invoking the greater complexity of eukaryotic over prokaryotic cells. Therefore, it was very surprising to discover that archaeal genomes contain genes encoding homologs of most eukaryotic factors, only excepting those involved in cap recognition (Bell & Jackson, 1998).

A summary of the putative translation IFs identified in Archaea on the basis of sequence homologies with known proteins in the other primary domains is shown in Table 19.2. Of the six proteins listed in the table, four are universal, i.e. have homologs in all domains of life, while another two are shared by the Archaea and the Eukarya to the exclusion of the Bacteria. Remarkably, no factor is shared by the Archaea and the Bacteria to the exclusion of Eukarya, once more stressing the evolutionary closeness of the gene expression machineries in the Archaea and the Eukarya. No archaeal-specific IF has been found so far, but it is possible that some will be identified following a more accurate biochemical and genetic analyses of archaeal initiation.

At present, very little is known about the function of the putative initiation factors in Archaea. The one archaeal factor to which a definite function can be assigned on the score of experimental data is the trimeric protein homologous to the eukaryal factor eIF2, here termed a/eIF2.

eIF2 has a central importance in eukaryal translation, as it interacts specifically with the initiator tRNA (met-tRNAi) and carries it to the 40S ribosomal subunits (Kimball, 1999). In Bacteria, the same essential function is carried out by the monomeric protein also called IF2, which, however, has no homology with any of the eIF2 subunits (Kyrpides & Woese, 1998a). The Eukarya do have a homolog of bacterial IF2 (termed eIF5B), which, however, does not interact with met-tRNAi but promotes the joining of the large ribosomal subunit to the

Table 19.2 Translation initiation factors in Archaea.

Factor name	E homolog	B homolog	Structure	Function in A	Function in other domains
aIF1A	eIF1A	IF1	Li & Hoffman, 2001	not determined	B: stimulates IF2 E: assists scanning
aIF2	eIF5B	IF2	Roll-Mecak et al., 2000	not determined	B: binds fmet-tRNA E: subunit joining
aSUI1	eIF1/SUI1	YCiH (some phyla)	Cort et al., 1999	not determined	B: unknown E: fidelity factor
a/eIF2 (αβγ)	eIF2 (αβγ)		γ, Roll-Mecak et al., 2004 β, Cho & Hoffman, 2002 α, Yatime et al., 2005	binds met-tRNAi	binds met-tRNAi
aIF6	eIF6		Groft et al., 2000	not determined	inhibits subunit association
aIF5A	eIF5A	EFP	Kim et al., 1998	not determined	B: formation 1st peptide bond E: undetermined

A, Archaea; B, Bacteria; E, Eukarya.

preinitiation complex (Pestova et al., 2000). The divergence of the tRNAi binding factors in Bacteria and Eukarya has been customarily attributed to the greater sophistication of translational regulation in Eukarya. In fact, eIF2 is central player in the regulation of eukaryal protein synthesis; the phosphorylation of its α-subunit, triggered by various stress signals, inhibits GTP/GDP exchange, thereby blocking the recycling of the factor and shutting off translation (Colthurst et al., 1987). However, the fact that the Archaea resemble the eukaryotes in having both eIF2-like and IF2-like factors shows that cellular complexity probably has nothing to do with the usage of these translation initiation factors.

Like eIF2, a/eIF2 is composed of three subunits that associate to form a hetero-trimeric complex (Yatime et al., 2004, 2005; Pedulla et al., 2005). The γ-subunit (about 45 kDa) and the α-subunit (about 30 kDa) have sizes comparable to those of their eukaryal homologs, while the archaeal β polypeptide (about 15 kDa) is much smaller than the eukaryal one, which is about 50 kDa in size and often the largest component of the trimeric complex. In fact, archaeal IF2-β is reduced to a conserved domain containing a zinc-finger motif while lacking the eukaryal-specific domains responsible for the interaction with the guanine nucleotide exchange factor eIF2B and with the GTPase activator eIF5. This agrees with the observation that all Archaea lack a homolog of eIF5 as well as a complete eIF2B. Archaeal genomes do include homologs of the α, β, and δ subunits of eIF2B, but lack counterparts of the γ and ε subunits that catalyze guanine nucleotide exchange on eIF2. Therefore, it is probable that the archaeal homologs of the eIF2B α, β, and δ proteins have a function unrelated to guanine nucleotide exchange (Kyrpides & Woese, 1998a).

Crystal or NMR structures are available for all three separate subunits of a/eIF2. As shown in Plate 19.1, the γ-subunit has a striking resemblance to the elongation factor 1A (formerly EF-Tu in Bacteria) (Schmitt et al., 2002; Roll-Mecak et al., 2004), in agreement with the fact that it contains the guanine-nucleotide binding domain and is principally involved in the interaction with met-tRNAi. a/eIF2-γ also contains a zinc-finger motif of uncertain function (Plate 19.1). The structures of the archaeal and eukaryal α-subunits are compared in Plate 19.2. Both proteins have a similar folding including three domains. The N-terminal domain (domain 1) has a β-barrel structure frequently observed in many RNA-binding proteins. The C-terminal domain (domain 3) contains an αββαβ module that is found in a large number of proteins and has been proposed to be an ancestral RNA binding motif (Yatime et al., 2005). An interesting feature of a/eIF2-α is the exposed loop in domain 1 (Plate 19.2), which is conserved in structure and contains the serine residue phosphorylated

in the eukaryal factor. The archaeal α-polypeptides also contain a Ser residue in the same loop, although it occupies a slightly different position with respect to its eukaryal counterpart.

The function of a/eIF2 from *Pyrococcus abyssi* (Yatime et al., 2004) and *S. solfataricus* (Pedulla et al., 2005) has been explored by *in vitro* biochemical assays using the factor reconstituted from the cloned recombinant subunits. These studies have revealed that a/eIF2, like its eukaryal counterpart, binds specifically met-tRNAi and carries it to the ribosomes. However, a number of features differentiate functionally the archaeal and the eukaryal proteins. One regards the nature of the tRNA binding site: an αγ dimer of a/eIF2 is necessary and sufficient to achieve a stable interaction with met-tRNAi, while in the case of eIF2 met-tRNAi binding seems to involve principally the γ and β subunits (Das et al., 1982). The α-polypeptide of the eukaryal factor appears to have mainly a regulatory function.

Another very relevant difference is that a/eIF2 has a similar affinity for GDP and GTP and therefore does not require a guanine nucleotide exchange factor to be reactivated (Pedulla et al., 2005). This finding is consistent with the lack of a complete homolog of eIF2B in archaeal genomes (Kyrpides & Woese, 1998a). According to this observation, a/eIF2 should not be subjected to a eukaryal-type functional regulation based on the inhibition of guanine nucleotide exchange upon phosphorylation of the α-subunit. However, it has been reported recently that the α-subunit of *Pyrococcus horikoshii* a/eIF2 is phosphorylated by a specific protein kinase (Tahara et al., 2004). The function of this modification is unknown, but it cannot be related to the regulation of GTP/GDP exchange. Perhaps it controls the function of the factor by regulating the formation of the trimer or its interaction with the ribosome. The solution of the problem may help us to understand why the archaeal/eukaryal branch, unlike the bacterial one, originally evolved a trimeric tRNAi binding factor.

Finally, unlike its eukaryal counterpart, a/eIF2 probably does not require a companion GTPase activator factor. GTP hydrolysis on eukaryal eIF2 is triggered by the helper factor eIF5, which has no recognizable homolog in Archaea. It is therefore likely that a/eIF2 has an intrinsic, ribosome-triggered GTPase activity, although this has not yet been demonstrated experimentally. Alternatively, a/eIF2 may be helped by a new and still unidentified GTPase activator.

The function of all of the other putative archaeal IFs remains undetermined, although crystal or NMR structures are available for most of them. A particularly interesting protein is aIF2, homologous to bacterial IF2 and eukaryal eIF5B and therefore one the few universally conserved IFs (Kyrpides & Woese, 1998b). Despite its conservation in all primary domains, this factor seems to have diverged in function, since in Bacteria it binds f-met-tRNAi and carries it to the ribosome, while in Eukarya it appears to promote the joining of the ribosomal subunits in a late stage of initiation (Pestova et al., 2000).

To date, there are few published experimental data about the function of the archaeal IF2-like factor. The only study performed *in vivo* has shown that *M. jannaschii* aIF2 can partially rescue yeast mutants lacking eIF5B (Lee et al., 1999), thus demonstrating that aIF2 is to some extent functionally homologous to eIF5B. On the other hand, preliminary data have been obtained *in vitro* suggesting that *S. solfataricus* aIF2 promotes the binding of met-tRNAi to the ribosome (Londei, 2005). Thus, aIF2 would seem to have properties somewhat intermediate between those of the bacterial and the eukaryal protein, but more data are needed to understand its function.

Structurally, archaeal IF2 proteins are smaller than their eukaryal and bacterial homologs, since they lack the long and poorly conserved N-terminal tracts of uncertain function present in both IF2 and eIF5B. Crystallographic studies on the *Methanothermobacter thermoautotrophicus* aIF2 (Roll-Mecak et al., 2000) have revealed that it is characteristically shaped as a chalice (Plate 19.3). The globular "cup" of the chalice (N-terminal region) includes the guanine-nucleotide-binding domain and a β-barrel domain probably involved in the interaction with the ribosomes. The "stem" of the chalice is a long α-helix, while the globular "base" (domain IV) corresponds to the C-terminal domain known to bind f-met-tRNA in bacterial IF2 (Guenneugues et al., 2000). The aIF2 (and eIF5B) domain IV has, however, lost the capacity for tRNA binding because of some critical amino acid substitutions in the relevant region.

Despite the evident divergence in their tRNA-binding capacity, the universal conservation of the

IF2-like proteins suggests that they still have some common function in all cells. This function may consist of promoting the interaction of the ribosomal subunits, but this has yet to be proven for the archaeal protein. An interesting common feature of all IF2-like factors is their ability to interact with another universal initiation factor, the protein termed eIF1A/aIF1A in Eukarya and Archaea and IF1 in Bacteria. Experimental evidence for a direct interaction of the eukaryal proteins eIF5B and eIF1A has been obtained (Marintchev et al., 2003). In contrast, bacterial IF2 and IF1 do not form a complex in solution, but may interact on the surface of the ribosome, as suggested by earlier cross-linking data (Boileau et al., 1983) and by a more recent cryo-electron microscopy study (Allen et al., 2005). Experimental data on archaeal IF1A indicate that it resembles its eukaryal counterpart in being able to interact stably with aIF2 in solution (Londei, 2005). The complex between the universal factors IF1/IF1A and IF2/IF5B is likely to be an ancestral feature of translation initiation, whose significance will be fully understood once more data are available on the archaeal proteins.

The small protein termed aSUI (or aIF1) has homologous counterparts in all Eukarya (where it is known as SUI1 in yeast and as eIF1 in vertebrates) and in a limited number of bacterial species, including *E. coli*, where it is called YciH (Cort et al., 1999). A phylogenetic analysis has shown that SUI1 is very likely an ancestral factor that has been lost secondarily by most bacteria, possibly because its function has been replaced by another protein (Londei, 2005). In Archaea, aSUI1 interacts with the 30S subunits but its precise function in translation initiation has yet to be determined. In Eukarya, SUI1/eIF1 is an essential protein that controls the fidelity of initiation codon recognition and probably also of elongation (Cui et al., 1998).

A very interesting factor shared specifically by the Archaea and the Eukarya is the 25 kDa protein called aIF6 (eIF6). The function of this factor in the Eukarya has been studied in some detail, but remains somewhat enigmatic. In yeast, eIF6 is an essential protein that is found both in the nucleolus and in the cytoplasm, where it associates with the 60S ribosomal subunits (Basu et al., 2003). The main phenotype observed in conditional mutants lacking the factor is a defect in the synthesis of 60S ribosomes, specifically a block in the processing of the rRNA 26S precursor (Basu et al., 2001). However, the cytoplasmic, 60S-bound eIF6 behaves as a ribosome anti-associating factor, preventing the formation of 80S particles and thereby inhibiting protein synthesis. According to a recent report, the dissociation of eIF6 from mammalian ribosomes requires the phosphorylation of the factor, which takes place when certain environmental cues activate a specific kinase (Ceci et al., 2003). Thus, eIF6 would resemble eIF2 in being a general regulator of protein synthesis. However, it remains unclear whether the two functions described for eIF6 indeed coexist, and which is the relationship between them, if any. Clearly, the functional study of the archaeal factor will be of great help in advancing our understanding of the cellular role of this interesting protein.

A last protein generally included in the translation initiation factors is the universally conserved polypeptide known as aIF5A/eIF5A in Archaea and Eukarya and as EFP in Bacteria. As the bacterial name implies, this protein can be regarded as a specialized elongation factor since it appears to catalyze the formation of the first peptide bond at the end of the initiation process in Bacteria (Glick et al., 1979). It seems probable that the function of this factor is conserved in all cells; however, the structure of archaeal IF5A differs to some extent from that of bacterial EFP. The latter is composed of three β-barrel domains and has an L-shaped structure reminiscent of a tRNA; it seems to bind both ribosomal subunits and to stimulate the peptidyl transferase center on the 50S particle (Hanawa-Suetsugu et al., 2004). The archaeal factor (structures are available for *M. jannaschii*, *P. aerophylum*, and *Pyrococcus horikoshii* aIF5A) is somewhat shorter than its bacterial homolog. It includes only two β-barrel domains and has a rod-like shape rather than a L-shape, and therefore it may interact preferentially with the 50S subunit (Kim et al., 1998). On the basis of structural comparisons, it has been proposed that archaeal/eukaryal IF5A evolved from an EFP-like common ancestor following the deletion of one of its three domains, and that perhaps another still unidentified protein has replaced functionally the missing domain in Archaea and Eukarya (Hanawa-Suetsugu et al., 2004). A remarkable feature of aIF5A, shared with its eukaryal homolog, is the presence of a uniquely modified lysine known as hypusine (N-ε-(4-aminobutyl-2-hydroxy) lysine), whose functional role is poorly understood.

Elongation and termination

The process of elongation is the basic biochemical core of protein synthesis and as such is extremely well conserved in evolution (see Table 19.3). All cells make use of two elongation factors, EF1 and EF2 (also known as EF-Tu and EFG in Bacteria). Like the majority of the components of the translational apparatus, the archaeal elongation factors have the closest homology with the eukaryal ones. Indeed, elongation factors-based evolutionary trees have first allowed us to place the root of the universal tree between the Archaea and the Bacteria, identifying the Archaea and the Eukarya as sister domains (Iwabe et al., 1989).

Termination, like initiation, has incurred a certain divergence in the primary domains of cell descent (see Table 19.3). In Bacteria and Eukarya, the specific task of recognizing the stop codons is performed by the class-1 termination factors, which release the completed polypeptide by promoting the hydrolysis of the ester bond anchoring it to the tRNA in the P-site. Bacteria possess two class-1 termination factors: RF1, recognizing UAA and UAG, and RF2, recognizing UAA and UGA. By contrast, the Eukarya appear to employ a single factor (eRF1) to recognize all three stop codons (Kisselev & Buckingham, 2000). All archaeal genomes include genes encoding a polypeptide homologous to eRF1 (termed aRF1), while no counterparts of bacterial RF2 have been detected. Therefore, the Archaea appear to resemble the Eukarya in using a single factor for stop codon recognition. That this is in fact the case has been demonstrated by the observation that *M. jannaschii* aRF1 can promote termination on eukaryotic ribosomes (Dontsova et al., 2000). Despite exhaustive *in silico* analyses, no meaningful similarity has ever been detected between the bacterial and the archaeal/eukaryal class-1 RF, which therefore seem to belong to two distinct protein families (Kisselev & Buckingham, 2000). Given the functional similarity between aRF1 and eRF1, the archaeal proteins were expected to have a structure comparable to that of their eukaryal counterpart. However, aRF1 appears to lack entirely a C-terminal domain present in both bacterial and eukaryal class-1 RF (Kisselev & Buckingham, 2000). Very likely, this reflects the fact that the archaeal genomes do not include any homolog of the class-2 RF, present in both the Bacteria and Eukarya, where they are termed respectively RF3 and eRF3. Class-2 RF are G proteins that do not participate in the peptide release reaction itself. The function of bacterial RF3 has been analyzed in some detail: briefly, its main task seems to be to accelerate the recycling of class-1 RFs after translational termination (Zavialov et al., 2002). Class-1 RFs interact with class-2 RFs by means of the C-terminal domains that are lacking in archaeal class-1 RFs. The Archaea also lack any apparent homolog of a bacterial-specific termination factor called RRF.

Thus, the data so far available suggest that the Archaea are endowed with a simplified version of the eukaryal translation termination mechanism, based on a single class-1 RF and dispensing with both the RF3 and the RRF proteins. Obviously, detailed experimental studies are needed to tell whether the Archaea possess unique termination factors that may take up the role played by the RF3s and/or RRF in the other two domains.

Conclusion and prospects

The study of the translational apparatus and of the protein synthesis mechanism in Archaea is still in its

Table 19.3 Translation elongation and termination factors in the primary domains.

	Archaea	Eukarya	Bacteria	Function
Elongation factors	aEF1A	eEF1A	EFT	Adapts aa-tRNA in ribosomal A site
	aEF2	eEF2	EFG	Promotes translocation
Termination factors	aRF1	eRF1	RF1/RF2	Stop codon recognition
		eRF3	RF3	Recycling of RF1
			RRF	Ribosome recycling

infancy, but the relatively few data available are revealing an interesting scenario of "hybrid" features whose detailed understanding will give new and exciting insights to the evolutionary history of the protein synthetic machinery. Foremost questions to be addressed in the near future regard the understanding of the mechanism for translation of leaderless mRNAs and the unraveling of the function of the putative translation initiation factors. A particularly interesting task will be to determine whether the Archaea make use of translational regulation mechanisms based on the phosphorylation of the translational factors that the Archaea share with the Eukarya, i.e. a/eIF2 and aIF6. The unraveling of the functional and regulatory role of these proteins, besides being interesting in itself, will be important to clarify some still obscure aspects of their function in eukaryotes.

A further subject almost completely unexplored is the mechanism of archaeal translational termination, which seems to be based on a single, eukaryal-like termination factor. Once more, the study of archaeal termination is likely to shed light on the unclear aspects of the corresponding eukaryal process, especially on the function of eRF3, a factor that in Archaea is apparently missing. Thus, translational studies in Archaea will not only improve our knowledge of the basic workings of the gene expression machinery, but also be precious to orient future research in the eukaryotic field.

Acknowledgments

Thanks are due to the people in the author's laboratory who worked on archaeal translation in the past few years: Dario Benelli, Enzo Maone, Maria Grazia Pantano, Nadia Pedullà, Rocco Palermo, Laura Peri, and Alessandra Barbazza. Thanks are also due to Udo Blaesi and David Hasenhörl of Vienna University Biozentrum for fruitful collaboration and helpful discussions. The work in the author's laboratory was supported by MURST projects PRIN 2000 "Translational initiation of leadered and leaderless mRNAs in Bacteria and Archaea: comparative analysis of the functions of IF2, an essential and universally conserved translation initiation factor" and PRIN 2002 "Evolution of the gene expression machinery: characterization of a set of universally conserved factors modulating translational initiation."