Evolution of translational initiation: new insights from the archaea

Paola Londei *

Dipartimento di Biochimica Medica e Biologia Medica (DIBIM), Università di Bari – Piazzale Giulio Cesare, 70124 Bari, Italy

Received 31 July 2004; received in revised form 22 September 2004; accepted 28 October 2004

First published online 20 November 2004

Abstract

Recent in silico and experimental data have shed new light on the mechanism and components of translational initiation in archaea. The available data about the structure of archaeal mRNAs, mRNA/ribosome interaction and archaeal translation initiation factors are reviewed and analyzed in the conceptual framework of the evolution of translational initiation. A model of the initiation step of translation in the Last Universal Common Ancestor of extant cells is presented and discussed.

Keywords: Archaea; Leaderless mRNA; Translation initiation; Translation initiation factors; Last Universal Common Ancestor

Contents

1. Introduction ............................................................................ 185
2. mRNAs in the archaea: “Leadered” and “Leaderless” messages. ........................................ 186
3. Translation initiation factors ................................................................. 188
   3.1. Translation initiation factors in the primary domains . ......................................................... 188
   3.2. The IF2-like factors: universally conserved proteins with divergent functions? . . . . . . 189
   3.3. a/eIF2: the met-tRNAi binding factor ................................................................................... 191
   3.4. aIF1A and aSU1 ..................................................................................................................... 192
   3.5. aIF5A and aIF6 ....................................................................................................................... 194
4. Concluding remarks: towards a reconstruction of the ancestral translation initiation mechanism ................. 195
   4.1. Structure of the mRNAs and initiation mechanism ................................................................. 195
   4.2. Initiation factors in ancestral translation .................................................................................. 195
   4.3. Models of ancestral translational initiation ............................................................................... 197
Acknowledgements ..................................................................... 197
References .............................................................................. 197

© 2004 Federation of European Microbiological Societies. Published by Elsevier B.V. All rights reserved.

* Address for correspondence: Dpt. Biotecnologie Cellulare ed Ematologia, Universitàdi Roma La Sapienza – Policlinico Umberto I, Viale Regina Elena 324 – 00161 Roma, Italy. Tel.: +39 06 4940463; fax: +39 06 4462891.
E-mail address: londei@bce.uniroma1.it.

1. Introduction

Translation, the last phase of the gene expression pathway, is highly conserved in evolution as regards its basic mechanism and components. However, initiation,
the first (and rate-limiting) step of translation, is known
to have incurred an extensive divergence in the bacteria
and the eukaryotes. Bacterial initiation entails direct
recognition of the translation initiation region on the
mRNA by the small ribosomal subunit (30S). This inter-
action is guided by the pairing of complementary RNA
sequences, the purine-rich Shine–Dalgarno (SD) motif
on the mRNA and the pyrimidine-rich anti-Shine–
Dalgarno (anti-SD) sequence on the 16S rRNA [1,2].
Ribosomal recognition of the translation start site is
modulated by three accessory factors, the monomeric
proteins known as IF1, IF2 and IF3 [3,4].

By contrast, translation initiation in eukaryotes fol-

ows a mechanism of considerable intricacy and com-
plexity. mRNA/ribosome recognition does not rely on
a direct mRNA/rRNA interaction and requires the help
of many different factors. The current consensus model
holds that the 40S ribosomal subunit, in a complex with
numerous accessory proteins, binds in the vicinity of the
capped 5' end of the mRNA and then slides in a 3' direc-
tion until the initiation codon – normally the first AUG
in any frame – is encountered [2,5,6]. Eukaryotic ribo-
somes do not seem to be able to land directly on the
mRNA unless the latter is equipped with special “inter-
nal ribosome entry sequences” (IRES) [7]. The factors
assisting eukaryotic initiation are about 10, many of
which complex multimeric proteins [8–10]. Several of
them are required for the unwinding of secondary struc-
tures in the mRNA, a necessary preliminary step for
ribosomal binding to the message. The picture is further
complicated by the peculiar structure of eukaryotic
mRNAs, which appear to adopt a circular shape owing
to the interaction between the poly (A) binding proteins
at the 3' end with the complex of factors recognizing the
cap at the 5' end [11,12].

The canonical textbook explanation for such diver-
gent translation initiation mechanisms is that they reflect
the different organization of the gene expression appar-
atus in prokaryotes and eukaryotes. In the bacteria, tran-
scription and translation are contemporaneous and
mRNAs are usually polycistronic. Therefore, direct
mRNA/ribosome recognition via the SD/anti-SD inter-
action is necessary for speeding up the decoding of the
more distal open reading frames (ORFs) in a polycis-
tronic message. On the other hand, a large complement
of initiation factors is superfluous since simple unicel-
ular organisms can dispense with an elaborate transla-
tional regulation as that enjoyed by the more
sophisticated and differentiated eukaryotic cells.

This neat scenario was challenged when the archaea
came into the picture. The expectation was that the ar-
chaea, being prokaryotes, had a translation initiation
apparatus similar to the bacterial one. There was no
structural or physiological reason to expect anything
different. Indeed, early data coming from translational
studies in the methanogens and the halophiles indicated
that the archaea did possess polycistronic mRNAs, of-
ten – but not always – endowed with clearly identifiable
SD motifs ahead of the initiation codons [13–15 and ref-
ences therein].

The discovery came therefore as a big surprise, fol-
lowing the sequencing of the first complete archaeal gen-
ome, that the archaea are endowed with over 10 genes
encoding proteins homologous to eukaryotic initiation
factors [16,17]. To date, the list of putative translation
initiation factors found in all completely sequenced
archaeal genomes (21 at the time of this writing) in-
cludes homologues of most of the known eukaryotic fac-
tors, excepting only those involved in cap recognition
(Table 1).

Such data elicited, of course, many questions and
comments. Were these proteins actually involved in
translational initiation and, if so, why such a wealth of
translation initiation factors in a prokaryotic domain?
Most comments favoured the idea that initiation in ar-
chaea had a somewhat hybrid character, resembling
bacteria in the structure of mRNAs (polycistronic orga-
nization and SD motifs) and eukarya in the complement
of initiation factors [16,17]. This, in a sense, was a mat-
ter of fact: but what does it mean in terms of the actual
mechanism of initiation in archaea? Addressing these
questions required experimental data; some of which,
although slowly, have been forthcoming in the last few
years. In the following paragraphs, I shall try to review
what is currently known about initiation in archaea and
what the emerging picture can teach us about the broa-
der topic of the evolution of the mechanism and appara-
tus for translational initiation.

2. mRNAs in the archaea: “Leadered” and “Leaderless”
messages

The complete sequencing of numerous archaeal gen-
ome in the last few years has yielded a wealth of infor-
mation on the nature and the organization of genetic
information in the third kingdom of life.

Like bacterial ones, archaeal genomes are at most a
few megabases in size, and contain little non-coding
DNA. Several introns have been detected in RNA-cod-
ing genes [18], while to date only one example of an
interrupted protein-coding gene has been found [19].
Some proteins, however, contain inteins [20].

As in bacteria, genes are tightly clustered and may
form polycistronic transcription units, some of which
resemble closely in structure the corresponding bacterial
unit. Gene clusters tend to be larger and more bacteria-
like in methanogens and halophiles, while the same
genes may be unlinked in the thermoacidophiles [14
and references therein].

The presence of polycistronic transcripts translated
into several polypeptides has been experimentally veri-
fied in a number of cases [13 and references therein]. Sometimes, mechanisms of translational regulation similar to the bacterial ones have been detected, as for instance autogenous regulation by ribosomal proteins on their encoding mRNAs [21].

As regards the mechanism of initiation, the presence of SD sequences in archaeal mRNAs (and that of the corresponding anti-SD sequence in the 16S ribosomal RNA) has been known for many years [15 and references therein], although the experimental verification of their function has been obtained only recently [22]. However, already two decades ago some unusual mRNAs had been found in halophilic archaea, which lacked entirely, or almost so, a 5′ untranslated region [23]. In these “leaderless” mRNAs, the initiation codon is located right at the 5′ end or just a few nucleotides downstream; SD sequences are obviously missing. To explain how the leaderless mRNAs of halophiles could be translated it was originally proposed that they were endowed with “cryptic” SD sequences, located a few bases downstream of the initiation codon; these sequences were recognized by the small ribosomal subunit with the customary RNA/RNA interaction mechanism [23]. A similar mechanism was proposed to account for the translation of the few leaderless mRNAs existing in bacteria and bacteriophages, such as the transcript of the λ repressor protein gene. In this latter case, it was suggested that a special sequence downstream of the initiation codon, termed the “downstream box”, mediated direct mRNA/ribosome interaction [24]. However, the “downstream box” sequence would not base pair with the anti-SD motif, but with another tract of the 16S ribosomal RNA spanning nucleotides 1469–1483 of 16S rRNA (E. coli numbering) [24,25]. The “downstream box” concept could be extended to the translation of any prokaryotic leaderless mRNA lacking “cryptic” SD motifs.

In summary, until recently mRNA/ribosome interaction in archaea was believed to be based on an mRNA/rRNA interaction as in bacteria, the only oddity being the presence of a sizeable amount of leaderless mRNAs whose translation, however, also relied on a mechanism essentially similar to the canonical SD/anti-SD pairing. This view has been first challenged by in silico investigations of the putative structure of the transcripts in the thermophilic archaea Sulfolobus solfataricus and Pyrobaculum aerophilum [26–28]. From the position of the presumptive promoters, the authors deduced that a large proportion of mRNAs were leaderless in these species. However, plausible “downstream boxes” or “cryptic” SD sequences were seldom identifiable. In the case of polycistronic mRNAs, the first cistron was very often leaderless, while the “internal” ones usually did possess well-defined SD motifs ahead of the initiation codons. On the basis of these data, it was proposed that two distinct mechanisms for initiating translation existed in archaea: one of bacterial type, based on SD motifs and operating on the internal cistrons of polycistronic mRNAs, and another of unknown nature employed for initiating translation of leaderless cistrons [27,29].
The first set of experimental data supporting the notion of two distinct mRNA types and translation mechanism in archaea was obtained using *S. solfataricus* as the model organism. With the aid of a cell-free translation system, it was demonstrated that SD motifs, when present, are essential for translational initiation [22]. In fact, the disruption of SD motifs by site-directed mutagenesis was found to completely inhibit in vitro translation, a situation more extreme than that found in bacteria where disruption of SD sequences usually reduces but not abolishes protein production. However, similar studies recently conducted in vivo on *Halobacterium salinarium* seem to indicate that in halophiles the disruption of SD motifs has less extreme consequences than in *Sulfolobus*, leading to a reduction of translational efficiency but not to a total inhibition of protein synthesis [30].

The most remarkable finding, however, 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 (UTR), namely by rendering the mRNA leaderless [22]. This result argued strongly against the presence of a cryptic SD motif or a “downstream box”, as it appeared very unlikely that a normal leadered mRNA should possess a masked ribosome binding sequence just in case it would somehow lose its 5′UTR. It suggested, instead, that when the SD-based mechanism was no longer operative, even an artificial leaderless mRNA could be translated using the alternative route, existing to operate on natural leaderless messages.

The first hints as to the features of the alternative initiation mechanism came from studies performed on bacterial leaderless mRNAs. The “downstream box” notion was beginning to be challenged in the bacterial domain by studies indicating that no mRNA/16S rRNA interactions were involved in the decoding of leaderless mRNAs in *E. coli* [31,32]. A study performed in vivo using mRNAs endowed with 5′ terminal as well as internal initiation codons [33] demonstrated that the presence of an excess of initiation factor 2 (IF2) favoured “leaderless” translation over leadered, SD-based, one. As IF2 in bacteria is the factor that promotes f-met-tRNA entry in the ribosomal P site, these results suggested that 30S ribosomal subunits pre-loaded with initiator tRNA were a prerequisite for leaderless translation. This meant that the recognition of a 5′ terminal AUG codon obligatorily required the interaction with the anticodon of the ribosome-bound f-met-tRNA, without any previous mRNA/ribosome interaction. Such a mechanism is reminiscent of eukaryotic initiation. However, perhaps the most interesting aspect of leaderless mRNAs is that, regardless of their source, they appear to be universally translatable by bacterial, eukaryal and archaeal ribosomes [33]. Therefore, “leaderless” initiation can be considered neither “bacterial” nor “eukaryal”, but might be a very old mechanism still operative in the three domains of life (see discussion).

Further information and support for the existence of two different mechanisms for translational initiation in archaea came from studies addressing the mechanism of ribosome/mRNA interaction in *S. solfataricus* [34].

Briefly, it was found that, in the presence of SD motifs, *Sulfobolus* small ribosomal subunits interacted directly and strongly in vitro with leadered mRNAs in the absence of any other translational component. Unlike their bacterial counterparts, the binary 30S/mRNA complexes thus formed were very stable. Leaderless mRNAs, by contrast, were unable to form binary complexes with 30S subunits. A 30S subunit/leaderless mRNA interaction could be detected only in the presence of met-tRNAi [34] supporting the idea that codon–anticodon pairing was required for initiation site recognition, as previously observed for leaderless mRNA translation in *E. coli* [33].

Of course, the presence of two different mechanisms for ribosome/mRNA recognition in the archaea raises a host of questions. Why two mechanisms? Are there any physiological reasons why “leaderless” translation is seemingly so widespread in archaea, especially in the hyperthermophiles, despite being rather exceptional in bacteria and eukarya? Or is leaderless translation a molecular fossil, carried over from the Last Universal Common Ancestor (LUCA) and retained preferentially in the archaea because of their slow evolutionary rate? Far more is necessary to know in order to begin to answer some of these questions. A very important task is to understand the function of the initiation factors in the two types of initiation.

3. Translation initiation factors

3.1. Translation initiation factors in the primary domains

As it is well known from a wealth of genetic and biochemical studies, the bacteria and eukarya have widely different sets of translation initiation factors. Bacteria, which are always very parsimonious in their endowment of cellular components, possess only three translation initiation factors (IF). The pivotal function of adapting f-met-tRNA in the ribosomal P site is performed by IF2, a 90 kDa guanine-nucleotide-binding (G) protein. IF1, a small polypeptide of 71 amino acids, stimulates the activity of IF2 and IF3 in the formation of the 30S ternary complex and modulates the interaction of IF2 with the ribosome. IF3 acts a “fidelity factor”, discouraging initiation at non-optimal codons such as AUU [4].

Eukaryotes by contrast, are endowed with a much more elaborate set of IFs [8,10]. Several factors, absent in bacteria, are required to unfold the mRNA (which
comes from the nucleus tightly packed with certain proteins) so as to allow its interaction with the ribosome. The cap-binding factor (termed eIF4F) is a complex of three polypeptides [8]. The actual cap-binding function is performed by eIF4E, a relatively small protein of 25 kDa. The other two members of the eIF4F complex are eIF4A, an RNA helicase, and eIF4G, a large modular protein that interacts with several other components, including the mRNA and the multimeric factor eIF3 [35].

No SD-anti SD interaction exists in eukarya, and ribosome/mRNA recognition is assisted by a complex set of factors. To “scan” the mRNA in search of the initiation codon the eukaryotic 40S subunits must be loaded with initiator tRNA (met-tRNAi). Such loading is accomplished by the trimeric factor eIF2, a G protein, which interacts with met-tRNAi only in the active, GTP-bound form. Following the successful interaction of met-tRNAi with the 40S subunit, GTP is hydrolyzed to GDP and eIF2 falls off the ribosome [36]. Despite the similar name and function, eukaryal IF2 has no homology with bacterial IF2; also, unlike the latter, it requires a GDP/GTP exchange factor to be reactivated after GTP hydrolysis. The exchange factor is a pentameric GDP/GTP exchange factor to be reactivated after GTP hydrolysis. Following the successful interaction of met-tRNAi with the 40S subunit, GTP is hydrolyzed to GDP and eIF2 falls off the ribosome [36]. Despite the similar name and function, eukaryal IF2 has no homology with bacterial IF2; also, unlike the latter, it requires a GDP/GTP exchange factor to be reactivated after GTP hydrolysis. The exchange factor is a pentameric factor called eIF2B. Moreover, unlike IF2, eIF2 has no intrinsic GTPase activity; the hydrolysis of the eIF2-bound GTP is triggered by an accessory GTPase, called eIF5 [36].

Other factors required for a successful “scanning” are eIF1, eIF1A, and eIF3; the latter interacts with eIF4F, the cap-binding complex [10]. Once the initiator AUG codon has been identified a further factor, eIF5B, is required to stimulate the joining of the 60S ribosomal subunit to the 40S initiation complex, thereby forming the complete 80S ribosome [37]. Like eIF2, eIF5B is a G protein: upon 60S subunit joining, GTP must be hydrolyzed to GDP to dissociate the factor from the newly assembled 80S ribosome. However, the formation of the 80S ribosomes per se does not require GTP hydrolysis [38]. The most interesting feature of eIF5B is that it has homology with bacterial IF2, as will be discussed later in better detail.

Still another factor, eIF5A, is required immediately after the assembly of the 80S ribosome to stimulate the formation of the first peptide bond [39]. To complete this brief survey, the recently characterized factor eIF6 must also be mentioned. This protein seems to be a very important regulator of eukaryotic translation: it interacts with the 60S ribosomal subunit preventing its joining with the 40S particle. Therefore, 80S ribosomes cannot form efficiently unless eIF6 is released from the 60S subunit. Such release requires the phosphorylation of the factor, which is controlled by means of several regulatory kinase cascades [40].

How do archaea compare with the other two domains in their complement of translation initiation factors? As noted previously, the recent sequencing of many archaeal genomes has overturned the expectations of most investigators, i.e. that the archaea, being prokaryotes, should possess a restricted set of initiation factors similar to the bacterial one. To the contrary, archaeal genomes appear to have a decidedly “eukaryotic” complement of (putative) translation initiation factors. As shown in Table 1, all archaeal genomes sequenced so far include genes encoding for homologues of most of the eukaryotic initiation factors. The only factors apparently restricted to the eukaryotic domain are the members of the eIF4F cap-binding complex, eIF3 and the GTPase eIF5.

Of course, homology in primary sequence does not necessarily mean homology in function. Once the candidate factors are identified by in silico studies, that they indeed have a role in translation must be demonstrated experimentally. Despite their scarcity, the available experimental studies of archaeal putative initiation factors have begun to delineate a very interesting scenario that will be presented in the next paragraphs.

3.2. The IF2-like factors: universally conserved proteins with divergent functions?

An essential function performed by translation initiation factors is initiator tRNA selection and adaptation in the ribosomal P site. As briefly described above, different and unrelated proteins, IF2 and eIF2, perform this task in bacteria and eukarya respectively. Thus, until recently it was generally believed that the initiator tRNA binding factor had evolved independently in the bacterial and eukaryal domains.

This belief was, however, repeatedly challenged in the last few years. First, genome sequencing in archaea revealed the presence of genes encoding proteins homologous to both bacterial IF2 and the three subunits of eukaryal IF2. Almost contemporaneously, a homologue of bacterial IF2 was recognized in the completely sequenced yeast genome. The first experimental study addressing the function of the yeast IF2 homologue revealed that was indeed a translation factor, unable to bind directly met-tRNAi but apparently capable of stimulating its binding to the ribosomes [41]. However, the factor was not essential, since yeast mutants lacking it were viable albeit very slow-growing. Later, the function of the rodent IF2-like protein was unravelled in a series of elegant experiments that made use of in vitro systems with purified proteins and ribosomes. These studies established that the rodent IF2-like factor is required primarily to promote the joining of the 60S subunits to the 48S initiation complex, thereby facilitating the formation of 80S ribosomes [37]. The factor was then dubbed eIF5B, a designation which, although sensible in the context of eukaryal initiation, somewhat obscures the very special evolutionary status of this protein. In
fact, being represented in bacteria, archaea and eukarya, eIF5B/IF2 is one of the few universal translation initiation factors [42]. This means that the IF2-like proteins were already present in the LUCA; their conservation in all modern members of the three primary domains strongly suggests that they still share an important common function. Yet, as explained above, the available experimental information on bacterial IF2 and eukaryal eIF5B indicates that these factors have incurred a marked functional divergence.

To understand something more about the hypothetical shared function of the IF2-like factors, and to get some insight into the evolution of the tRNAl binding factors, it becomes very important to unravel the function of the archaeal homologues of IF2/eIF5B and of eIF2. Why should archaea possess a eukaryotic set of these proteins? Have the archaeal proteins the same functions as those of their eukaryal counterparts?

To date, the only published experimental data about the function of the archaeal IF2-like proteins show that M. jannaschii IF2 homologue can partially rescue yeast mutants lacking eIF5B [43]. These results suggest at least some degree of functional homology between the eukaryal and archaeal factors. Yet, as will be explained below, more-detailed studies suggest some caution before concluding that the archaeal and eukaryal IF2-like proteins have entirely overlapping functions. For this reason I shall henceforth refer to the archaeal factor as aIF2 rather than as aIF5B as some investigators do.

In contrast with the paucity of functional data, there is much detailed information about the structure of aIF2. Crystallographic studies on the M. thermoautotrophicum aIF2 have revealed that it is characteristically shaped as a chalice [44]. The globular “cup” of the chalice (N-terminal region) includes three domains, the first of which is the guanine-nucleotide-binding domain. The “stem” of the chalice is a long alpha-helix, while the globular “base” (domain IV) corresponds to the C-terminal domain known to bind f-met-tRNA in bacterial IF2 [45]. It must be pointed out that the aIF2 proteins are generally smaller than their eukaryal and bacterial homologues, since they lack the long and poorly conserved N-terminal tracts of uncertain function present in both IF2 and eIF5B.

Our laboratory is currently studying the function in archaeal translation of S. solfataricus aIF2. The factor has been obtained in recombinant form by standard cloning and over-expression procedures, and it has been tested in a number of in vitro assays to unravel its role in translational initiation. Moreover, natural aIF2 has been produced by translation of its mRNA in a cell-free system for protein synthesis prepared from S. solfataricus cell lysates [22].

The assays performed with the recombinant protein [77] have revealed that aIF2 is a ribosome-dependent, thermophilic GTPase that interacts preferentially with 50S subunits and 70S ribosomes. Like its eukaryal homologue, aIF2 does not bind met-tRNAi or any other tRNA, including bacterial f-met-tRNA. Nevertheless, when the recombinant protein is added to a purified system containing only ribosomes and met-tRNAi it stimulates strongly the binding of the latter to the ribosomes, a function reminiscent of that of its bacterial counterpart (Fig. 1). Moreover, when aIF2 is over-expressed in vitro by translating its mRNA in a cell-free system for protein synthesis, it displays a most remarkable stimulatory effect on the translation of other mRNAs, both leadered and leaderless, suggesting that aIF2 function is common to both initiation mechanisms (Fig. 2; Benelli, D., Maone, E., Londei, P., unpublished results). Stimulation of met-tRNAi/ribosome interaction is a good candidate for such function.

Stabilization of met-tRNAi binding to the ribosomes was also the function originally proposed for the yeast IF2 homologue, following the observation that slow-growing yeast strains lacking eIF5B could be partially rescued by over-expressing initiator tRNA [41]. These data have been integrated with those obtained in higher eukarya [37] in the proposal that eukaryal eIF5B stabilizes met-tRNAi binding to the ribosome while stimulating the formation of the 80S ribosome [46].

However, despite all our efforts, so far we have been unable to produce any evidence that archaeal IF2 has a role in facilitating ribosomal subunit joining. This, of course, may only be due to technical problems, but it must be observed that also in the bacterial domain...
protein synthesizing systems were prepared as described by [22].

In eukarya, it is well established that the fundamental function of adapting met-tRNAi into the ribosomal P site is carried out by the trimeric factor eIF2, whose component subunits are termed α, β and γ [8]. Ensuring an efficient met-tRNAi/ribosome interaction is particularly crucial in the eukaryal domain as the 40S subunit must carry met-tRNAi to successfully identify the initiation codon during the mRNA scanning process. Therefore, eIF2 is a preferential target of mechanisms regulating eukaryotic translation initiation. Inhibition of eIF2 activity is obtained through the phosphorylation of its α subunit; this is triggered by a number of stress signals which activate certain kinase cascades [47]. When the α subunit is phosphorylated, the factor becomes unable to interact with the guanine nucleotide exchange factor eIF2B. As a result, eIF2 remains locked in the inactive GDP-bound form and translation is severely inhibited or entirely shut off [36].

The other two subunits of eIF2, β and γ, perform different tasks. The γ subunit contains the guanine nucleotide binding site and part, at least, of the tRNA binding site [48], while the β subunit, besides being probably involved in tRNA binding, includes the domains for the interaction with the accessory factors, i.e. the GTP/GDP exchange protein eIF2B and the GTPase eIF5 [49,50]. In summary, eIF2 is a very complex factor, which lies at the centre of an elaborate network of interactions regulating its activity.

Until recently, the prevalent rationale for the divergence of the tRNAi binding factors in bacteria and eukarya was that the latter had to evolve a brand-new protein to achieve a more finely tuned and sophisticated regulation of translational initiation than that required by the simple prokaryotic cells. Needless to say, this view has been challenged by the discovery that all archaeal genomes so far sequenced include homologues of the three component subunits of eukaryal eIF2. That these proteins associate to form a trimeric complex has been demonstrated in a couple of recent works on P. abyssi a/eIF2 [51,52].

Trimeric a/eIF2 is decidedly smaller in size than its eukaryal counterpart, owing principally to the much-reduced length of its β subunit. In all archaea, the γ subunit is the largest protein (about 45 kDa) followed by the α subunit (about 30 kDa). The β subunit of only 15 kDa comes a far third, while in the eukarya it is generally about the same size as the γ subunit (about 50 kDa). The archaeal β polypeptide is so much smaller than its eukaryal counterpart because it lacks the domains, which, in the eukaryal subunit, interact with eIF2B (the guanine nucleotide exchange factor) and eIF5 (the GTPase). This agrees with the observation that all archaean lack a homologue of eIF5 as well as a complete eIF2B. Archael genomes do include homologues of the α, β and δ subunits of eIF2B, which in the eukaryotic factor have a regulatory role. However, no counterparts of the catalytic subunits γ and ε have been detected. Therefore, it is probable that the archaeal homologues of eIF2B α, β and δ subunits have a function unrelated to guanine nucleotide exchange [53].

The crystal structure of the IF2B α subunit of P. horikoshii has been solved and it has been proposed that it exists in a complex with the β and δ homologues [54].

The small archaeal β subunit is almost entirely occupied by a conserved domain containing a zinc-finger
motif whose function in the eukaryal polypeptide is still uncertain, although it has been proposed to bind mRNA [55].

Has a/eIF2 the same function as its eukaryal homologue? Recently, it has been demonstrated that the factor from Pyrococcus abyssi, reconstituted in vitro from the cloned recombinant subunits, is able to interact preferentially with met-tRNAi. Apparently, the dimer formed by the α and γ subunit is sufficient for a stable binding of met-tRNAi; however, the γ subunit by itself is also able to interact with the tRNA, although in a much less stable fashion [52]. The recent determination of the crystal structure of the isolated γ subunit of P. abyssi a/eIF2 [51] has also shown that it contains a complete tRNA binding domain similar to that found in the translational elongation factor Tu, another G protein showing homology with the γ subunit of eIF2 and a/eIF2.

It seems therefore established that a/eIF2 is the initiator tRNA binding factor in the archaea, although the details of its function remain to be worked out. Of the many open questions, one regards the nature of the tRNA binding site: the work of Yatine et al. [52], as well as data from our laboratory suggests that a αγ dimer is necessary and sufficient for a stable interaction with met-tRNAi (Fig. 3). This differs from the situation in eukarya, where the γ subunit also has the largest share in met-tRNAi binding [48], but its principal helper seems to be the β subunit [56], while the α polypeptide has been implicated mainly in the regulation of the factor activity. However, the eukaryal β subunit has evidently incurred an extensive evolutionary rearrangement, increasing its size by the addition of new domains. Therefore, it is quite possible that it has also achieved a role in tRNA binding, absent in its archaeal homologue.

Another question regards the regulation of guanine nucleotide binding to a/eIF2. As explained previously, the complex of the eukaryal factor with GDP is several orders of magnitude more stable that that with GTP. As a consequence, following GTP hydrolysis, the eIF5B factor is required to catalyze GDP/GTP exchange in order to reactivate eIF2. As a complete IF5B is missing in the archaea, it is reasonable to suppose that a/eIF2 can dispense of a GTP/GDP exchange factor. If so, no eukaryal-type functional regulation of a/eIF2 based on the inhibition of guanine nucleotide exchange should exist in archaea. However, it has been reported recently that the α subunit of P. horikoshii a/eIF2 is phosphorylated by a specific protein kinase [57]. If these results are confirmed, determining the function of α subunit phosphorylation of archaenal a/eIF2 will be a quite interesting task, which may help to clarify why the archaean, unlike bacteria but like eukarya, have adopted a trimeric met-tRNAi binding factor.

The role of the a/eIF2 β subunits remains at the moment elusive. In our laboratory, we have so far failed to detect any involvement of this protein in any of the factor’s functions, although it must be remembered that these are in vitro experiments. Further work is necessary to identify the conserved role of the a/eIF2 β subunit along the archaenal/eukaryal evolutionary line.

Another problem that remains as yet unsolved is whether a/eIF2 possesses an intrinsic GTPase activity or whether it requires an accessory GTPase as its eukaryal counterpart. As discussed earlier, GTP hydrolysis on eukaryal eIF2 is triggered by the helper factor eIF5, which has no recognizable homologue in archaea. Therefore, a/eIF2 either has an autonomous GTPase activity or is helped by a new and still unidentified GTPase. Some years ago Kyrpides and Woese [53] suggested that the companion GTPase of a/eIF2 might be aIF2. As said previously, aIF2 does possess an intrinsic ribosome-dependent GTPase activity, and it is just possible that, upon interacting with the ribosome, it would promote the hydrolysis of its own bound GTP and of that bound to a/eIF2. This hypothesis is currently being investigated in our laboratory.

3.4. aIF1A and aSUI1

As shown in Table 1, the list of putative archaenal initiation factors homologous to eukaryal proteins includes two polypeptides known as eIF1A and eIF1 (also termed SUI1 in yeast). The most interesting feature of these proteins is that they have homologues also in the

---

**Fig. 3.** The αγ dimer of a/eIF2 is necessary and sufficient for met-tRNAi binding. About 50 pmol each of purified recombinant α, β and γ subunits of S. solfataricus a/eIF2 were incubated, individually and in all possible combinations, with an excess of S. solfataricus tRNA charged with 35S met. The samples were incubated for 15 min at 65 °C and the amount of met-tRNA bound to the proteins was determined by filtration through nitrocellulose disks (0.22 μm).
bacterial domain and therefore, like IF2, are universally conserved initiation factors [42].

Bacterial IF1 and archaeal/eukaryal IF1A constitute the IF1-like factor family. Bacterial IF1 is a small polypeptide of 71 amino acids essential for cell survival [58]. It has been shown to interact with the A site on the 30S ribosomal subunits [59]. Its reported functions include: (1) the enhancement of the rate of 70S ribosome dissociation and subunit association, (2) the stimulation of the activity of IF2 and IF3 in the formation of the 30S ternary complex, (3) the modulation of the interaction of IF2 with the ribosome, increasing its affinity for the 30S ribosomal subunit when IF1 is bound and indirectly favouring its release when IF1 is ejected [4].

The eukaryal IF1-like protein, aIF1A, is a 17-kDa polypeptide that is essential for cell survival as its counterpart in bacteria. The factor has been shown to stimulate the binding of met-tRNAi in the P site [60] and, in collaboration with eIF1 (SUI1), to promote ribosome scanning and correct initiation codon selection [61].

Interestingly, it has been shown that eIF1A interacts with eIF5B (the IF2 homologue), forming a stable complex even outside the ribosome [62,63]. As some earlier data suggested that bacterial IF1 and IF2 may also interact [64], the proposal has been put forward that a complex between the IF1-like and the IF2-like proteins is a universal feature of translation initiation in all domains [46]. The function of such a complex would be that of occupying the ribosomal A site during the early stages of initiation, thereby stimulating the correct access of tRNAi in the P site and promoting the joining of the large subunit to the initiation complex.

However, this model is based mainly on the properties of the eukaryal proteins and requires more data on their homologues in the other two domains to be validated. For instance, a convincing demonstration that bacterial IF1 and IF2 interact is still lacking; in any event, a putative complex of these proteins would not be stable in a free state and would form only on the surface of the ribosome [63].

Our laboratory is currently investigating the role of aIF1A in archaeal translation initiation. The protein of *S. solfataricus* has been obtained by PCR cloning and over-expression in *E. coli*. In vitro assays with the recombinant *S. solfataricus* aIF1A have shown that it enhances the activity of aIF2 in promoting the entry of met-tRNAi in the ribosomal P site, suggesting an interaction between the two proteins. Indeed, when mixed in solution in equimolar amounts under appropriate conditions, in the absence of ribosomes or other translational components, the two factors can be co-immunoprecipitated, indicating that they form a sufficiently stable complex in vitro (Fig. 4).

These results, albeit still preliminary, are in agreement with the hypothesis that a complex of the IF1-like and the IF2-like factors is a common feature of the three primary domains. More experimental data are required, however, to work out the structural and functional features of the archaeal aIF1A/aIF2 complex, and especially to determine whether it has any role in promoting subunit joining, supposedly an important aspect of its universally conserved function [46].

Another interesting protein, essential for eukaryotic translation and present in all archaea so far examined is the factor known as eIF1 (in higher eukaryotes) or SUI1 (in yeast). In eukarya, it seems to have an important role in ensuring the accuracy of initiation codon selection and in controlling translational fidelity during elongation [65]. In bacteria, however, it is well established that the “fidelity factor” for initiation codon selection is IF3, a strictly bacterial protein absent in both archaea and eukarya (Table 1) and unrelated to eIF1/SUI1. Certain bacterial species, including *E. coli*, do possess a homologue of eIF1/SUI1, whose function is still unknown [66]. Therefore, eIF1/SUI1 is a third universally conserved factor, which is, however, present only in a minority of bacterial phyla while being found in all archaea and eukarya [42].

The peculiar distribution in the bacteria of the SUI1-like factor (known as YciH in *E. coli*) is amenable to two different interpretations. The protein may have arisen in the archaeal/eukaryal branch and may have been acquired later by some bacterial phyla by horizontal gene transfer. Alternatively, it may be a truly ancestral factor, i.e. already present in the LUCA and later lost in most bacterial phyla during the separate evolution of the
bacterial domain. Phylogenetic trees drawn by comparing the available sequences of SUI1-like proteins in the three domains favour the latter interpretation. In fact, bacterial, archaeal and eukaryal SUI1 homologues group into clearly distinct clades (Fig. 5). If the bacteria had acquired the protein by horizontal gene transfer from the archaea or the eukarya, bacterial SUI1 homologues would have clustered either with their archaeal or with their bacterial counterparts.

The *S. solfataricus* SUI1 homologue (aSUI1) has been cloned and purified in our laboratory. The most relevant result so far obtained is that the protein is tightly bound to the small ribosomal subunits, where it remains during both translational initiation and elongation (Fig. 6). Currently, we are investigating whether the over-expression of aSUI1 in a cell-free protein synthesizing system exerts any effect on the translation of leadered or leaderless mRNAs endowed with AUG or other initiation codons.

3.5. aIF5A and aIF6

Scarce or no information is available about these archaeal factors. IF5A is actually an interesting protein: it belongs to the universal set of factors (Table 1) and appears to have a conserved function in bacteria and eukarya, namely that of promoting the formation of the first peptide bond [39]. This makes it more similar to a specialized elongation factor than to an initiation factor, and in fact, the protein is termed EF (i.e. elongation factor) P in the bacteria (see Table 1).

The function of archaeal IF5A has not been investigated experimentally, although it is likely to be similar to that of its bacterial and eukaryal counterparts. Curiously, the presence of this protein in the archaeal lineage has been known for a long time since, like its eukaryal counterpart, it contains a unique hyper-modified form of lysine known as hypusine [67]. The structure of aIF5A from *M. jannaschii* has been solved by crystallographic techniques [68].

Studying the function of the archaeal IF6 homologue should be high on the agenda of investigators working in the translation initiation field. As explained previously, eukaryal IF6 inhibits the association of the ribosomal subunits when bound to the 60S particle in its non-phosphorylated form. Phosphorylation of eIF6 triggers its release from the 60S subunits thereby allowing the assembly of 80S ribosomes and stimulating translation [40]. It will be very interesting to see whether a similar kind of regulation also operates in the archaea.
4. Concluding remarks: towards a reconstruction of the ancestral translation initiation mechanism

As will be clear from the above overview of the state of the art, the study of translation initiation in archaea is still in its infancy and either mechanism or components of the process are far from being fully understood. Nevertheless, the information available from the archaea has greatly changed our perception of the evolution of translation initiation. Although it is too early to reconstruct in a meaningful way the evolutionary history of translation initiation, some speculations can be made that will also help to shape future research in the field.

4.1. Structure of the mRNAs and initiation mechanism

As described in a previous paragraph, there is enough evidence to state with some certainty that the archaea employ two different initiation mechanisms to translate leadered and leaderless mRNAs [34]. Since there is no apparent physiological reason that may explain this fact, a possibility is that “leaderless” and “leadered” initiation have different evolutionary histories. According to this view, “leaderless” initiation would be the ancestral mechanism, carried over from the LUCA and retained preferentially in the archaea because of their slower evolutionary rate [33,34].

Several observations argue in favour of the hypothesis that “leaderless” initiation is the ancestral mechanism. The most compelling evidence is that leaderless mRNAs are universally translatable (at least in vitro) by archaeal, bacterial and eukaryotic ribosomes [33]. 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 underlying leaderless translation. The recent observation that in the protozoan Giardia lamblia most mRNAs are leaderless [69] may also be taken as supporting the ancestral nature of such mRNAs, depending on whether or not G. lamblia is classified as a primitive eukaryote. It is known that Giardia occupies a deep branch of the eukaryal evolutionary tree, but it is still doubtful whether this indicates true primitivity or is an artefact of evolutionary analysis, due to an abnormally fast “ticking” of the molecular clock, a frequent occurrence in parasitic organisms like G. lamblia.

If leaderless mRNAs are indeed the ancestral type of message, what about the SD-motif based, direct mRNA/ribosome interaction also found in present-day archaea? The logical surmise is that this mechanism has arisen to allow an efficient decoding of polycistronic messages, especially of the cistrons located internally in a long mRNA. Since extant bacteria also possess polycistronic mRNAs and SD motifs, a possibility is that the evolution of the SD-based mechanism of ribosome/mRNA interaction has preceded the branching of the bacterial domain off the root of tree of life, i.e. has occurred while still at the LUCA stage of cellular evolution. If so, one may envisage that the LUCA underwent several steps in the evolution of translation initiation, starting, in a more primitive stage, with monocistronic leaderless mRNAs to achieve, in a later stage, also polycistronic mRNAs endowed with SD motifs in the internal cistrons but still leaderless as far as the initial cistron was concerned. This is the situation still prevalent in the Archaean domain, especially in the Crenarcheota. The alternative scenario would be the independent evolution of polycistronic mRNAs and SD/anti-SD interaction in bacteria and archaea, which, although possible, seems less likely, considering that the order and number of genes in several operons are conserved between bacteria and archaea [14].

If polycistronic mRNAs and the SD/anti-SD based mRNA/ribosome interaction were already present in the LUCA, it follows that they have been lost during the evolution of Eukaryotes. The evolutionary pressure underlying such loss has been, in all probability, the separation between nucleus and cytoplasm leading in turn to the separation between transcription and translation [70]. On the other hand, the “ribosome scanning” mechanism of present-day Eukaryotes may be envisaged as an elaborate version of leaderless initiation: it is based on ribosome binding at the 5' end of the message, with the added functions of RNA unfolding and directional ribosome movement to hit the correct initiation AUG codon, which is recognized mainly by means of base-pairing with the anticodon of the met-tRNAi. Notwithstanding the added sophistications, the basic mechanism has been conserved to the extent that modern eukaryotic ribosomes are still able to translate leaderless mRNAs from bacterial sources [33].

To gain a better understanding of the evolutionary status of “leaderless” initiation we also need to determine which initiation factors it requires. In bacteria, translation of leaderless mRNAs has been shown to be stimulated by an excess of IF2, perhaps because this factor promotes the loading of initiator tRNA on the 30S subunits [33]. Recent studies have suggested that bacterial leaderless mRNAs may be read by 70S ribosomes, even without the aid of any factors [71,72]. On the other hand, the protozoan Giardia, whose mRNAs are all leaderless, seems to be endowed with a normal eukaryotic complement of initiation factors [69]. It is hoped that the study of archaea may shed light on these seemingly contradictory evidences.

4.2. Initiation factors in ancestral translation

The previous survey of the available information about archaeal initiation factors will have made clear that we need to know much more about the function of these proteins. A glance to Table 1 will show that...
the majority of the archaeal IFs (4 over 6) are in fact universally conserved proteins. The other two factors (a/eIF2 and aIF6) are shared only with the eukaryotes and therefore arose in the common ancestor of archaea and eukarya, after the branching of the bacterial domain. As yet, no archaeal-specific translation initiation factor has been identified, but this is not very surprising since the search for initiation factors has been carried out looking for homologies with the bacterial or eukaryal proteins. Obviously, any archaeal-specific IF would have escaped detection: therefore, it is entirely possible that archaeal translation initiation makes use of unique factors still to be discovered.

Needless to say, it would be of the utmost interest to unravel the function of the universally conserved factors in ancestral initiation. At present, however, only a few speculations are possible. As shown in Table 1, the set of ancestral factors includes four proteins: the IF1-like factors, the IF2-like factors, the SUI1- like factors and the EF-P/IF5A proteins. Three of these (IF1, IF2, IF5A/EF-P) are ubiquitous in the three primary domains. The SUI1-like factors are not truly universal, since they are missing in the majority of bacterial phyla, indicating that they are probably not essential in modern bacterial translation. However, the reconstruction of the evolutionary tree of the SUI1-like factors (see Fig. 5) suggests that they, too, were present in the LUCA and have later been lost in most bacterial species. It is not known whether bacterial SUI1 is functional in the species that still possess it; likewise, we are still ignorant of the function in translation of archaeal SUI1. At present, it is therefore impossible to speculate meaningfully on the hypothetical ancestral function of this factor.

As to the truly universal factors, the EF-P/IF5A proteins seem to have the same function in bacteria and eukarya, namely stimulating the formation of the first peptide bond [39]. Even if the function of archaeal IF5A has not been studied yet, it is safe enough to assume that it will turn out to be similar to that of its eukaryal and bacterial counterparts. Therefore, the EF-P/IF5A factors can be looked at as a sort of specialized elongation factors that catalyzed the formation of the first peptide bond in ancestral and modern cells alike.

As regards the IF2-like proteins, the available data on their present function in the primary domains suggest that their ancestral role consisted in promoting the interaction of tRNAi with the ribosomal P site. The other function detected in eukaryotic cells, stimulating ribosomal subunit joining, requires further data in the bacteria and the archaea to be validated as universal.

It is debatable whether the ancestral IF2-like factor was capable of direct tRNA binding. In present-day cells, only bacterial IF2 has this function, while its eukaryal and archaeal counterparts do not bind met-tRNAi. Therefore, the tRNA-binding capacity may have been acquired during the evolution of bacterial IF2, or may have been lost by the archaeal/eukaryal IF2 homologues. At present, it is impossible to decide between these possibilities; however, it must be pointed out that bacterial IF2 is specialized for interacting with f-met-tRNAi, and it is known that the formyl group plays an important role in this interaction [45]. Since formylation of tRNA-bound methionine is a unique bacterial feature, it is perhaps easier to imagine that the ancestral IF2 had no special tRNA-binding capacity, but could indirectly favour the entry and stabilization of tRNA in the P site by interacting with the ribosome in correspondence of the A site.

In ancestral translation (as well as in modern one), the IF2-like factor worked in close collaboration with the IF1-like protein, also a universally conserved factor. As we have seen, the available experimental evidence suggests that the IF1-like and the IF2-like factors interact in all cells to form a complex designed to occupy the ribosomal A site during translational initiation [46,63]. Further work on the bacterial and archaeal IF2-like and IF1-like proteins is needed to confirm this surmise and to work out the structural and functional features of the hypothetical universal complex.

To complete the picture, it is worthwhile to spend a few words on the possible evolutionary history of the eukaryal/archaeal met-tRNAi binding factor, the trimeric protein eIF2. The γ subunit of this factor belongs to a family of tRNA-binding G proteins that includes the elongation factor Tu1-A and its specialized version SELB [73]. The latter is the factor that selects the selenocysteine tRNA during the synthesis of selenoproteins in all primary domains [74–76]. It is interesting to note that the γ subunit of eukaryal and archaeal IF2 has the best homology with SELB [73]. This indicates that a/eIF2 γ arose by gene duplication from the elongation factor Tu at about the same time as SELB; each of the two proteins became specialized for interacting with a single aminoacyl-tRNA species, while the elongation factor Tu remained capable of binding any aminoacyl-tRNA. Probably, a/eIF2 acquired a selective affinity for met-tRNAi during the evolution of the common ancestor of archaea and eukarya.

It is probable that in the eukaryal/archaeal line the γ subunit of IF2 initially bound met-tRNAi as a monomer. This is indicated by the fact that the γ polypeptide hosts the tRNA binding domain and can still interact weakly with met-tRNAi in isolation [52]. The later addition of the α and β subunits improved the met-tRNAi binding capacity of the factor and added regulatory functions, at least in the eukaryal lineage. However, a full understanding of the evolution of this factor must await a better knowledge of its structure and function in the archaea.
4.3. Models of ancestral translational initiation

In Fig. 7, the above considerations are combined in a hypothetical model of ancestral translation initiation. Our assumptions are that ancestral mRNAs were leaderless, and, at least in early LUCA, monocistronic. The first stages of initiation were assisted only by the IF1-like and the IF2-like factors. An IF1/IF2 complex, probably assembled on the surface of the ribosome, occupied the ribosomal A site thereby guiding the tRNAi in the P site. Once tRNAi was stably bound in the P site, the ribosome could interact with the leaderless mRNA by codon–anticodon pairing. In early LUCA, ribosome/mRNA interaction possibly required complete 70S ribosomes rather than isolated small subunits as it happens in modern initiation. Initiation complexes formed by isolated small subunits may have evolved later, together with the “invention” of the SD–anti-SD interaction.

Two slightly different scenarios are conceivable depending on the assumptions made about the interaction of ancestral IF2 with tRNAi. If the factor was capable of direct tRNA binding, f-met-tRNA was probably already used as the initiator tRNA species. Therefore, the bacterial branch retained the streamlined ancestral situation, while the archaeal/eukaryal line evolved another factor for the binding of met-tRNAi and lost the tRNA binding domain on IF2/IF5B.

Assuming that the ancestral IF2 had no tRNA binding-capacity, it is probable that the tRNAi chose the ribosomal P site primarily by negative default, the A site being occupied by the IF1/IF2 complex. Perhaps at this stage any tRNA could have been used as the initiator species, depending on the codon present at the 5’ end of the leaderless mRNA. Later, bacterial IF2 acquired the capacity of selective f-met tRNA binding, while the archaeal/eukaryal branch evolved a different met-tRNAi binding factor. Much of the uncertainty in drawing these models is due to the fact that we still do not really understand why different initiator tRNAs should be used in the bacteria and in the archaea/eukarya, or why (and when) AUG was chosen as the preferential initiation codon.

Some aspects of the above models are, however, amenable to experimental testing. No doubt an exciting future awaits investigators interested in the evolution of translation.

Acknowledgements

Thanks are due to the people in my lab who expertly and enthusiastically worked on archaeal initiation in the last few years: Alessandra Berardi, Dario Benelli, Enzo Maone, Nadia Pedulla, Rocco Palermo, Laura Peri and Alessandra Barbazza. I especially thank Prof. Piero Cammarano for many helpful discussions and for working out the SUI1 phylogenetic tree.

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”.

References


mammalian elongation factor for selenoprotein translation. EMBO J. 19, 4796–4805.
