Bridging Domains

A Comparison between Information Processing in Archaea and Eukarya

Bart de Koning

Thesis committee

Promotor
Prof. Dr J. van der Oost
Personal chair at the Laboratory of Microbiology
Wageningen University

Co-promotor
Dr S.J.J. Brouns
Assistant professor at the Laboratory of Microbiology
Wageningen University

Other members
Prof. Dr B. Siebers, University Duisburg-Essen, Germany
Prof. Dr A.J.M. Driessen, University of Groningen
Dr R.T. Dame, Leiden University
Prof. Dr D. Weijers, Wageningen University

Bridging Domains

A Comparison between Information Processing in Archaea and Eukarya

Bart de Koning

Table of Contents

1 Introduction

Luckily most people are aware that they are surrounded by a living world that goes well beyond humanity. Most people do not realise that what they consider to be the living world, is in fact just the tip of the living iceberg. The majority of the world's biomass is not fixed in huge rainforest trees, or in vast areas of wheat. It is not even fixed in the numerous insects that bug us wherever we go. Instead most of the world's biomass is fixed, exchanged, and recycled in a living world that is several orders of magnitude smaller, a world that is invisible for the human eye. We might not like it, but earth is actually the dominion of prokaryotes.

These relatively primitive creatures are often associated with awful diseases, as some constantly try to attack our human bodies, and with hygienic nightmares, as some constantly threat to devastate our bathrooms and kitchens. However, unlike our main associations suggest, the vast majority of the prokaryotic world is harmless to humans, and some are even beneficial. We, ourselves, carry along on our skin and within our gut, a microbiome that has ten times more cells than we have human cells in our body, constituting a genetic diversity that is much larger than the mere 30.000 genes that make up our own genome. Being responsible for vitamin K production, for protection against pathogens, and for degradation of potential harmful metabolites, they are in fact our primary guards that protect us from far worse.

The evolutionary tree of life

These prokaryotes, whether they induce illnesses, or protect us from it, whether they live in our bathrooms, kitchens or guts, are often bacterial species. However, next to the bacteria there is another distinct form of prokaryotes. In the second half of the last century, during the pioneering, laborious work of characterizing the prokaryotic world by molecular profiling of rRNAs, Carl Woese and George Fox stumbled upon prokaryotic species that appeared to differ significantly from bacteria. In fact their ribosomal RNA (rRNA) was much more like the rRNA of eukaryotic species like plants and animals [1]. This revolutionary model was supported by work of Wolfram Zillig, performed during the same decade, in which he found a similar dichotomy within the prokaryotic RNA polymerases [2]. These hallmark studies revealed that the living world was not divided in merely two branches, prokaryotic and eukaryotic, but in fact was divided into three main lineages, called Domains: the familiar Eucarya (or Eukaryotes), Bacteria, and the newly discovered unknown Archaea [3, 4]. This proclamation caused turmoil in the field of taxonomy. A field that was mainly dealing with ordering of the (visible) eukaryotic world, and back then, molecular profiling to aid in this ordering was new, and therefore highly controversial. This deviation into three domains has often been, and still is, the topic of fierce debates [5, 6]. Since Lynn Margulis posted her symbiotic theories for the origin of the eukaryotic cell [7, 8], the question remained whether it is possible to draw the evolutionary lines of all living species as a simple dichotomous tree. We know, since then, that some of these branches merge again, and later split again into other branches [9]. The analyses necessary to disentangle the evolutionary processes are even more complicated by the fact that horizontal gene transfer, which is the transfer of genetic information between (often unrelated) organisms other than via inheritance, is a very common phenomenon in the prokaryotic world. An alternative view on this evolutionary picture comes from James Lake, who postulated the rings of life theory – with one ring to rule them all –, providing evidence that the eukaryotic genome is in fact the result of a very complex fusion of several distinct evolutionary lines, or rings in fact, of descent of archaeal and a bacterial origin [10, 11]. In 2008 Cox et al. even complicated the view, providing additional support for the Eocyte hypothesis for the origin of

the eukaryotic lineage. This hypothesis claims that the eukaryotic branch node lies not outside of the archaeal branch, as Woese et al. proposed, but within, as a sister group to the crenarchaeotes – a group they called the Eocytes. This does not only imply that the Eukaryotes and the crenarchaeotes share a direct common ancestor, that is unrelated to the euryarchaeotes, but also implies that the archaeal domain is actually paraphyletic, and should be divided between the Crenarchaeota and the Euryarchaeota [6]. Disentangling these lines of descent has suffered always from a great lack of archaeal sequences to provide evolutionary resolution within the rooting of the eukaryotic branch. However the recent drop in sequencing costs has led to an increasing number of metagenomic, and single-cell genomic studies, aiming to explore the worlds microbial diversity. Each of these efforts constantly added a great number of "new" archaeal species, belonging to "new" archaeal branches besides the already known Crenarchaeota and Euryarchaeota [3]: the Diapherotrites, Parvarchaeota, Aenigmarchaeota (forming toghether with the Nanoarchaeota [12] and the Nanohaloarchaeota [13] the DPANN superphylum) and, Aigarchaeota [14] (forming together with the Thaumarchaeota [15], the Crenarchaeota and the Korarchaeota [16] the TACK superphylum) [5, 17–19]. The constant additions of new genomes from branches unknown before, increases the resolution for deep branching phylogenetic studies. With the current genomes available today it might be still too early to give a definitive call for the origin of the eukaryotic cell, but support for a progenitor originating from within the TACK superphylum is increasing [17].

Eukaryotic like Information Processing in a Archaeal Context

Controversial thoughts or not, nobody can deny the evidence that the eukaryotic cell is a mixture of both bacterial as archaeal features [6, 9, 20, 21]. The close phylogenetic relation of archaea and eukaryotes is even more obvious when the comparison is restricted to systems involved in information processing: the flow of information through DNA replication, gene expression and protein synthesis [21]. The archaeal complexes involved in DNA replication, transcription, translation, and even RNA and protein turnover, have more similarity with their eukaryotic counterparts than with their bacterial ones. Research on the archaeal systems, performed during the last decade, provided numerous new insights into the functionality of these systems, which, in turn, provided new insights into their more complex eukaryotic counterparts as well. In the second chapter of this thesis, we will review current knowledge about information processing in archaea, and focus on the relationship of these systems with their bacterial and eukaryotic counterparts. It shows that information processing throughout the living world is related, but that the relationship between the archaea and eukaryotes is stronger (Chapter 2).

Archaea as Models

Like bacteria, archaea are simple, unicellular, organisms, with relatively small circular genomes, and without cell compartments, although in a few prokaryotic lineages (e.g., in Cyanobacteria, Planctomycetes, and Crenarchaeotes) some kind of compartmentalization was found (see [22] for an evolutionary note). In the beginning archaea were especially associated with extreme environments. They were found on places where it was expected that life would be impossible: hyperthermophiles are found even around marine hydrothermal vents where temperatures can exceed 100°C, acidophiles live in terrestrial vulcanic springs at pH values below 2, piezophiles thrive at spots with pressures of more than 100 MPa, and even die at normal atmospheric pressures. Archaea clearly expanded our rather limited view on the boundaries of – earthly – life. However, nowadays, we realise more and more that the archaeal branch does not only encompass creatures that are extremely good in surviving uninhabitable areas, but are present in large numbers in "normal" habitats as well. Being responsible for up to ~40% of the total marine prokaryotes in shallow zones, and for the majority at deeper levels, they have a important role in the marine carbon and nitrogen cycles [23–25].

The clear relationship between archaeal and eukaryotic information processing, poses the interesting question whether archaeal cells could serve as model organisms to study the more complex eukaryotic systems. Due to their less complex cell structure and lifestyle, studying information processing systems in archaea could reveal key characteristics about these systems that could be easily overlooked in the eukaryotic systems. In addition, some archaea are easily cultivable due to their preference for culturing conditions at which other organisms will not survive.

During the last decade, a number of archaea had their genomes being sequenced, and genetic tools have been developed for several model archaea [26]. Although these genetic tools are still far from being established in ways known to their eukaryotic and bacterial counterparts. Therefore we investigated whether we could use these tools to elucidate yet unknown features of some eukaryotic information processing proteins during the efforts described in this thesis.

Protein Bridges between Domains

In 1999 Aravind and Koonin published the results of a bioinformatic study for transcriptional regulators, selected for the presence of a Helix-Turn-Helix (HTH) motif, in all domains of life [27]. Because the RNA Polymerase complex (RNAP) of eukaryotes and archaea is highly similar (Chapter 2), they investigated whether archaea – unlike bacteria – would contain regulatory elements that resembled counterparts of eukaryotic transcription regulation. One such element draw their special attention: Multiprotein Bridging Factor 1 (MBF1), the only HTH domain protein that was highly conserved in the archaeo-eukaryotic lineage, but completely absent in bacteria. MBF1 was first isolated from silk moths, and characterized in yeast to be a transcriptional co-activator. A transcriptional co-activator serves as a connector between DNA binding regulatory elements and the core transcriptional machinery. In archaea, however, its function was unknown: all known interaction partners in eukaryotes, besides the Tata-box Binding Protein (TBP) – a core transcription factor –, were known to be absent in

archaea. During the decade that followed, investigations revealed more and more about the functionality of MBF1 in eukaryotic cells, especially about its role in stress response within plants. About its role in archaea on the other hand little progress was made in the meantime. A complete overview of the knowledge about MBF1 at the time of the start of the project, from literature, sequence alignments, and orthology comparisons is provided in Chapter 3.

An approach often undertaken to explore the functionality of a gene, is to investigate the effects of a knockout mutation on the fitness of an organism under several circumstances to understand what happens if it is absent. Therefore the aim of the project described in chapter 4 of this thesis, was to construct a knockout strain within a well characterized crenarchaeal species (*Sulfolobus solfataricus*), and expose this mutant to several stress situations and monitor the growth characteristics. Key element within this study was to do large scale comparisons to reach the statistical power necessary to provide evidence for gene-knockout related defects. The standard growth procedure for this organism did not allow such comparisons, and therefore a less known method was adopted, which is described in Chapter 4. Unfortunately, gene disruption mutagenesis does not always reveal the functionality of a protein. Measuring the growth characteristics can only monitor the gross outcome of a very complex interplay between the biological reactions that occur within living cells. Although that information is invaluable for identifying the role for a certain component within the complex behaviour of a living cell, it will not give exact information about the molecular reactions it is involved in. To investigate the molecular basis of the growth behaviour we observed, we did multiple interaction studies with MBF1 of *Sulfolobus solfataricus*, in which we tried several state of the art high throughput methods to find an answer to the question whether or not MBF1 was acting as a transcriptional activator in archaeal cells as well, with a quite unexpected answer (Chapter 4 & 5).

Archaeosine

MBF1 is present and highly conserved in the archaeo-eukaryotic lineage, but completely absent in the bacterial domain of life. In contrast, the TGT family of proteins is present in all three lineages, however with a clear dichotomy in function. In the bacterial and eukaryotic clades of the TGT family, it is responsible for the exchange of a guanine into a queuosine, important for distinction between highly alike nucleotides at the wobble position in the anti-codon loop of several tRNAs. However, within the archaea, TGT is responsible for one of the archaeal hallmarks: the archaeosine incorporation at position 15 of tRNAs. The dichotomy between both roles is absolute: creatures with both modes of action do not exist, or are at least not known up to today. In the early days of archaeal research, archaea where mainly associated with extremophilic growth behaviour. Therefore, the early discovery of this archaeosine modification, soon led to common view that it was an archaeal trait obtained to sustain its harsh environments. To verify these speculations, we constructed an *Sulfolobus solfataricus* – which is a thermophilic organism – knockout mutant for this gene as well, and subjected it to different growth temperatures (Chapter 6).

Closing the circle

Our pet model organism during these studies was *Sulfolobus solfataricus.* This organism already has a quite extensive toolbox [26], but lacked the possibility to create mutants that provided easy purification of target proteins, without the need for insertion of foreign genetic elements. As a proof of principle we investigated the possibility to create an in-frame genomic insertion tag, for a protein complex that since the discovery 40 years ago provided already numerous insights for its eukaryotic counterpart: RNAP. Recent models showed the existence of a protruding stalk in both archaea and eukaryotes. This stalk, which is important for initiation, and for proliferation as well, is composed of two conserved subunits: Rpo4 and Rpo7. As this stalk like structure is protruding from the complex, we designed a purification tag on the outermost location available, as to prevent possible side effects of the tag, and to make substrate binding to an affinity matrix as likely as possible. These efforts where described in the last experimental chapter of this thesis (Chapter 7).

To conclude and have a more general elaboration of the findings presented within this thesis, a reflection on this work has been provided in the last chapter.

References

1. Woese CR, Fox GE: **Phylogenetic structure of the prokaryotic domain: the primary kingdoms.** *Proc. Natl. Acad. Sci. U. S. A.* 1977, **74**:5088–90.
2. Zillig W, Stetter KO, Janeković D: **DNA-dependent RNA polymerase from the archaebacterium Sulfolobus acidocaldarius.** *Eur. J. Biochem.* 1979, **96**:597–604.
3. Woese CR, Kandler O, Wheelis ML: **Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.** *Proc. Natl. Acad. Sci. U. S. A.* 1990, **87**:4576–9.
4. Albers S-V, Forterre P, Prangishvili D, Schleper C: **The legacy of Carl Woese and Wolfram Zillig: from phylogeny to landmark discoveries.** *Nat. Rev. Microbiol.* 2013, **11**:713–9.
5. Guy L, Ettema TJG: **The archaeal "TACK" superphylum and the origin of eukaryotes.** *Trends Microbiol.* 2011, **19**:580–7.
6. Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM: **The archaebacterial origin of eukaryotes.** *Proc. Natl. Acad. Sci. U. S. A.* 2008, **105**:20356–61.
7. Sagan L: **On the origin of mitosing cells**. *J. Theor. Biol.* 1967, **14**:255–74.
8. Knoll AH: **Lynn Margulis, 1938-2011.** *Proc. Natl. Acad. Sci. U. S. A.* 2012, **109**:1022.
9. Zillig W, Klenk H-P, Palm P, Leffers H, Pühler G, Gropp F, Garrett RA: **Did Eukaryotes Originate by a Fusion Event?** *Endocytobiosis Cell Res.* 1989, **6**:1–25.
10. Rivera MC, Lake JA: **The ring of life provides evidence for a genome fusion origin of eukaryotes.** *Nature* 2004, **431**:152–5.
11. Lake JA, Sinsheimer JS: **The deep roots of the rings of life.** *Genome Biol. Evol.* 2013, **5**:2440–8.
12. Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, Stetter KO: **A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont.** *Nature* 2002, **417**:63–7.
13. Narasingarao P, Podell S, Ugalde JA, Brochier-Armanet C, Emerson JB, Brocks JJ, Heidelberg KB, Banfield JF, Allen EE: **De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities.** *ISME J.* 2012, **6**:81–93.
14. Nunoura T, Takaki Y, Kakuta J, Nishi S, Sugahara J, Kazama H, Chee G-J, Hattori M, Kanai A, Atomi H, Takai K, Takami H: **Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group.** *Nucleic Acids Res.* 2011, **39**:3204–23.

15. Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P: **Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota.** *Nat. Rev. Microbiol.* 2008, **6**:245–52.

16. Elkins JG, Podar M, Graham DE, Makarova KS, Wolf Y, Randau L, Hedlund BP, Brochier-Armanet C, Kunin V, Anderson I, Lapidus A, Goltsman E, Barry K, Koonin E V, Hugenholtz P, Kyrpides N, Wanner G, Richardson P, Keller M, Stetter KO: **A korarchaeal genome reveals insights into the evolution of the Archaea.** *Proc. Natl. Acad. Sci. U. S. A.* 2008, **105**:8102–7.

17. Guy L, Saw JH, Ettema TJG: **The Archaeal Legacy of Eukaryotes: A Phylogenomic Perspective.** *Cold Spring Harb. Perspect. Biol.* 2014.

18. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T: **Insights into the phylogeny and coding potential of microbial dark matter.** *Nature* 2013, **499**:431–7.

19. Gribaldo S, Brochier-Armanet C: **Time for order in microbial systematics.** *Trends Microbiol.* 2012, **20**:209–10.

20. Rivera MC, Jain R, Moore JE, Lake JA: **Genomic evidence for two functionally distinct gene classes.** *Proc. Natl. Acad. Sci. U. S. A.* 1998, **95**:6239–44.

21. Yutin N, Makarova KS, Mekhedov SL, Wolf YI, Koonin E V: **The deep archaeal roots of eukaryotes.** *Mol. Biol. Evol.* 2008, **25**:1619–30.

22. McInerney JO, Martin WF, Koonin E V, Allen JF, Galperin MY, Lane N, Archibald JM, Embley TM: **Planctomycetes and eukaryotes: a case of analogy not homology.** *Bioessays* 2011, **33**:810–7.

23. Herndl GJ, Reinthaler T, Teira E, van Aken H, Veth C, Pernthaler A, Pernthaler J: **Contribution of Archaea to total prokaryotic production in the deep Atlantic Ocean.** *Appl. Environ. Microbiol.* 2005, **71**:2303–9.

24. Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA: **Isolation of an autotrophic ammonia-oxidizing marine archaeon.** *Nature* 2005, **437**:543–6.

25. Karner MB, DeLong EF, Karl DM: **Archaeal dominance in the mesopelagic zone of the Pacific Ocean.** *Nature* 2001, **409**:507–10.

26. Leigh JA, Albers S-V, Atomi H, Allers T: **Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales**. *FEMS Microbiol. Rev.* 2011, **35**:577–608.

27. Aravind L, Koonin E V: **DNA-binding proteins and evolution of transcription regulation in the archaea.** *Nucleic Acids Res.* 1999, **27**:4658–70.
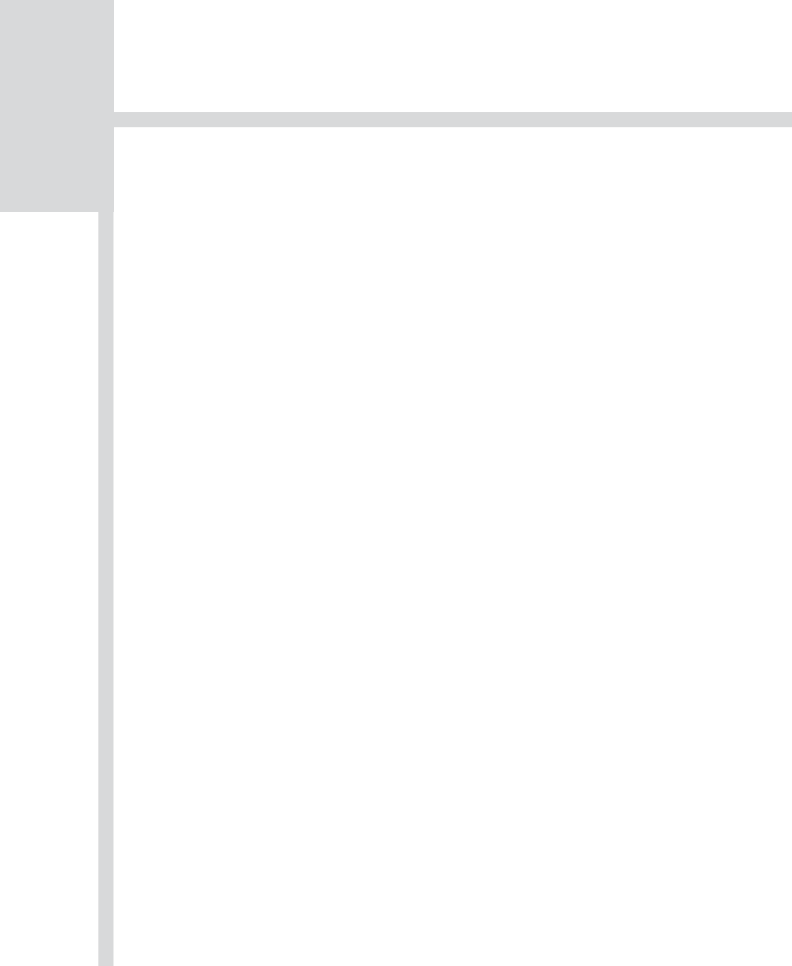
2 Fidelity in Archaeal Information Processing

Bart de Koning, Fabian Blombach, Stan J.J. Brouns, John van der Oost

Abstract

A key element during the flow of genetic information in living systems is fidelity. The accuracy of DNA replication influences the genome size as well as the rate of genome evolution. The large amount of energy invested in gene expression implies that fidelity plays a major role in fitness. On the other hand, an increase in fidelity generally coincides with a decrease in velocity. Hence, an important determinant of the evolution of life has been the establishment of a delicate balance between fidelity and variability. This paper reviews the current knowledge on quality control in archaeal information processing. While the majority of these processes are homologous in Archaea, Bacteria, and Eukaryotes, examples are provided of non-orthologous factors and processes operating in the archaeal domain. In some instances, evidence for the existence of certain fidelity mechanisms has been provided, but the factors involved still remain to be identified.

Introduction

Francis Crick first announced his central dogma of molecular biology in 1958: the flow of sequential information that occurs in living cells, including replication of stored information (DNA), as well as expression of this information via messengers (mRNA) to functional proteins [1]. This dogma turned out to be a solid basis for molecular biology, although additional roles of (small) regulatory and metabolic RNA have been recognized more recently [2]. A key element during this transfer of genetic information is fidelity: the final accuracy depends on the combined error rates of the processes that constitute the whole chain.

From the ancient RNA world on, replication fidelity has been a major limiting factor of the amount of information stored. It has been proposed that on average less than one error per replicated genome is tolerated, as higher error rates lead to a so-called "error catastrophe" with a fatal amount of progeny not being viable [3–5]. The same rule applies also for extant cellular life in which double-stranded DNA is used for storage of genetic information. The increase in genome size was allowed by the increased stability of DNA [6] and by considerably lower error rates in DNA replication [7]. One might expect a continuous selection towards the highest possible fidelity. However, a very high level of fidelity in replication will negatively affect both the genome's adaptation potential, and the replication velocity and costs, posing the risk of being out-competed by more efficient rival organisms [8, 9]. Overall, the delicate balance between fidelity and mutation rate is in itself a trait of organisms and can differ between individuals and species [10]. For some species it is even known to change upon environmental signals and may vary between different locations within the same genome [11]. Fidelity of information processing is thus a major factor driving the evolution of cellular life.

Transcription and Translation show significantly higher error rates than replication. Although the risk on affecting progeny is lower, erroneous gene expression might influence the error rate of replication indirectly e.g. when the replication machinery is affected [12]. On the one hand, inaccurate gene expression may lead to the production of non-functional proteins, and as such to a decreased fitness, i.e. generating selective pressure for increasing fidelity. On the other hand, increasing the fidelity of Transcription and Translation also correlates with decreasing velocity, what also has an impact on fitness. Hence, natural evolution leaves a narrow range for varying the level of fidelity [13, 14].

In this review, we will whenever possible focus on the systems of Archaea that contribute to accurate replication and expression of their genetic information. While the majority of the archaeal processes are well conserved in Bacteria, and/or Eukaryotes, a number of examples will be described of factors and processes that appear to be restricted to the archaeal domain. Despite the fact that research on Archaea is generally lagging behind that of the other two domains, the successful development of several Archaea as model organisms has recently lead to some first insight in their mechanisms to control fidelity of information processing.

Replication

Fidelity in replication is the result of three separate processes: (i) base selection, (ii) proofreading, and (iii) post-synthetic correction [15, 16]. These three processes contribute to very accurate DNA replication: incorporating a mistake only once every $10^6 - 10^{10}$ nucleotides for DNA-based microorganisms. Interestingly, the genomic mutation rate (the number of mutations per replicated genome) is quite constant for all DNA-based microorganisms, including bacteriophages, bacteria and fungi: roughly 0.003-0.004 (Drake's rule [7]), what is largely below the above mentioned predicted upper limit of 1 error per replicated genome [4]. Surprisingly, it has recently been found that a thermophilic bacterium (*Thermus thermophilus*) and a thermophilic archaeum (*Sulfolobus acidocaldarius*) have error rates that are 5-fold lower, supporting the concept that there is an evolved balance between the need for fidelity and the cost of reducing the mutation rate [17].

After a brief description of polymerases in living systems, the three separate processes will be discussed in more detail. The last paragraph will discuss systems that organisms have evolved to overcome misincorporations.

DNA polymerases

DNA is polymerized by DNA-dependent DNA polymerases (DNAPs) that can be classified into various families based upon their sequence similarity. Most replication-related DNAPs and primases belong to DNAP family B. Like Bacteria and Eukaryotes, Archaea contain multiple DNAPs. *Sulfolobus solfataricus* for example contains three family B DNAPs (B1 to B3) and one family Y DNAP (Dpo4) [18]. Crenarchaeota are restricted to family B for their replicative polymerases, while Euryarchaeota, Korarchaeota, Nanoarchaeota, and Thaumarchaeota use both a family B and a family D DNAP [19]. There is biochemical evidence that in these species the family B DNAP replicates the leading strand, while the family D DNAP replicates the lagging strand [20]. Deviation between leading and lagging strand replication has been found in other domains of life as well [21, 22]. Lagging strand replication involves Okazaki fragments that are produced by a lagging strand replicative DNAP, initially extending an RNA primer generated by a primase, a family B RNA polymerase. Archaea possess homologs of eukaryotic primase proteins (PriS ans PriL) that can synthesize both RNA as DNA oligonucleotides *in vitro*, but seem to prefer RNA polymerization *in vivo* [23, 24]. Interestingly the B family replicative DNAPs of Archaea contain an uracil-specific pocket that scans the template for the presence of uracil ahead of the polymerase. This feature is apparently lost in eukaryotic and bacterial DNAPs, although they still possess the reminiscent pocket structure. If uracil is encountered the archaeal polymerase stalls, presumably until the uracil is removed by Base Excision Repair (BER) or until a TransLesion Synthesis (TLS) DNAP takes over [25, 26]. TLS is a process in which the regular replicative DNAP is substituted by a translesion DNAP. Translesion polymerases, often family Y DNAPs, allow replication to occur past otherwise impassable DNA lesions. This adaptation however has led to a considerably lower fidelity than in case of replicative DNAPs. Dpo4 from *Sulfolobus solfataricus* is a family Y TLS DNAP. Dpo4 has a spacious solvent-exposed active site in comparison to replicative DNAPs that permits accurate bypass of the 8-oxoguanine oxidation product of guanine. 8-oxoguanine preferentially base-pairs to adenine, however in the active site a stabilizing hydrogen bond network fixes 8-oxoguanine in such position that the correct preference for cytosine is restored [15, 27].

Base selection

The highest contribution to fidelity during DNA replication is brought about by base selection. Soon after the initial suggestion by Watson and Crick that selection was the result of hydrogen bonding of complementary bases [28], it became clear that the free-energy differences between correct and incorrect base pairs could only account for error rates of approximately 0.01 [16]. Although the removal of water from the active site of the DNAP leads to elevated ΔG values, improving the selectivity between correct and incorrect base pairings [29], studies with base analogs that lost the capacity to create hydrogen bonds revealed the importance of base pair geometry. In addition, structural studies showed that a Watson-Crick pair, of which all four are nearly identical in shape and size, fits nicely into the base pair binding pocket of DNAP, while non-Watson-Crick base pairs presumably cause steric clashes (reviewed in [15, 30]).

Proofreading

Like the replicative DNAPs of the Bacteria and Eukaryotes, both family B as family D DNAPs from Archaea possess intrinsic proofreading capabilities [26, 31, 32]. Because these enzymes are thermostable, and have intrinsic proofreading, they are of commercially interest as exemplified by the high fidelity Pfu DNAP from *Pyrococcus furiosus* in polymerase chain reactions. Comparisons between wild-type polymerases with intrinsic proofreading capabilities and exonuclease-deficient mutants show that on average proofreading improves fidelity between 3-100 fold. For *Sulfolobus solfataricus* DNAP B1, the commercially available DNAP (Vent™ pol) from *Thermococcus litoralis* and their respective exonuclease-deficient mutants, it was measured to improve approximately 3 fold, a similar increase as observed for *E. coli* DNA pol III [15, 32]. DNAPs have prolonged interaction with the newly generated duplex DNA. Mismatches are recognized because of abnormal base pair geometry, and generally result in considerably decreased elongation rate. In DNAPs that have intrinsic or associated 3' → 5' exonuclease activity, elongation rate drops below the exonuclease rate upon mismatch recognition, leading to removal of mismatched nucleotides. Polymerases without intrinsic exonuclease activity can either recruit another protein that has exonuclease activity, or can dissociate and allow another polymerase with intrinsic exonuclease activity to take over.

Other errors generated during elongation, at approximately the same rate as mismatches, are single-base deletions and slightly less frequently single-base insertions. These "indels" can occur by (i) DNA strand slippage, (ii) mis-insertion that is followed by primer relocation, or (iii) misalignment at the polymerase active site, and can occur especially at repetitive sequences. Whereas proofreading corrects mismatches at a high rate, this mechanism is relatively inefficient in correcting indels, especially if the repetitive elements are longer. Strand slippage for example occurs often upstream of the polymerase, is therefore not sensed and does not decrease the elongation rate, preventing the exonuclease activity from taking over (reviewed in [15, 30]).

Post-synthetic Correction

Mismatches or indels that slipped through the proofreading process, or that are introduced by mutagenic factors, are to be repaired by post-synthetic correction. Organisms generally have a set of distinct systems, designed to repair a specific class of damage, each with a different fidelity rate. A repair system directly connected to replication is Mismatch Repair (MMR). This system removes base substitutions and indels on the newly synthesized strand directly after replication. MMR increases fidelity of replication almost 100-fold [15]. In Bacteria and Eukaryotes essential proteins required for MMR belong to the MutS and MutL family. These

two families are largely absent in the archaeal domain. Archaeal homologs have only been found in some euryarchaeal species, probably the result of a horizontal gene transfer from bacterial origin [33]. Deletion mutants of a variety of MutS and MutL homologs in *Halobacterium salinarum*, including a MutS double mutant, had only little effect on mutation rates, indicating that these genes are not essential for MMR in this species [34]. A MutS2 ortholog is also present in the euryarchaeote *Pyrococcus furiosus* and it was shown to have ATPase and DNA binding activity, but no specific MMR activity [35]. Despite the general absence of MutS and MutL in Archaea, it is found that spontaneous base pair substitution rates in *S. acidocaldarius* are an order of a magnitude lower than MMR-proficient *E. coli* suggesting the existence of a powerful, yet unknown MMR system in Archaea [17]. During MMR, a key step is to identify which of the two strands is the (correct) parental strand and which one the (mutated) daughter. In some bacterial systems, the methylated strand is considered to be the parental strand, a signal for MutH to cleave opposite of a methylated GATC sequence near the mismatch [36]. Other Bacteria, Eukaryotes, and Archaea, use other mechanisms to distinguish between the strands that are not yet fully understood. It is believed that in Eukaryotes the newly synthesized daughter strand contains discontinuities, caused by the separate Okazaki fragments during lagging strand replication and by reinitiation or low-level incorporation of dUMP during leading strand replication. Archaea may also use the incorporation of uracils as a marker for the daughter strand as well, in line with the fact that DNA replication in Archaea cannot pass uracils on the template strand [37].

Excision Repair

Two additional repair systems that repair single strand damage by using the complementary strand as a template, include (i) Base Excision Repair (BER) used to remove regularly occurring small, non-helix-distorting base lesions (e.g. modification by de-purinations and de-aminations) and involves DNA glycosylases, and (ii) Nucleotide Excision Repair (NER) used to remove bulky distortions in the helix (e.g. thymine dimers formed by oxidative stress or UV). The BER system appears to be functional in Archaea, as archaeal BER-related thermostable N-glycosylases have been characterized [38–42]. In contrast, the archaeal NER system appears to lack important damage-recognition proteins, but has structure-specific nucleases, homologous to eukaryotic NER nucleases [37]. UV stress experiments with *Sulfolobus acidocaldarius* show evidence for the existence of an archaeal NER system, as its repair capacity is at least half the capacity of NER-proficient *E. coli* [43, 44]. Especially life at elevated temperatures asks for efficient repair systems, as spontaneous decomposition reactions are accelerated under these conditions [6]. The high temperatures characterising the habitat of *Sulfolobus* species causes high rates of depurinations and deaminations. Although most of these types of damage are removed by BER, the apparent absence of key factors for both NER and MMR has been referred to as "the great irony" [37].
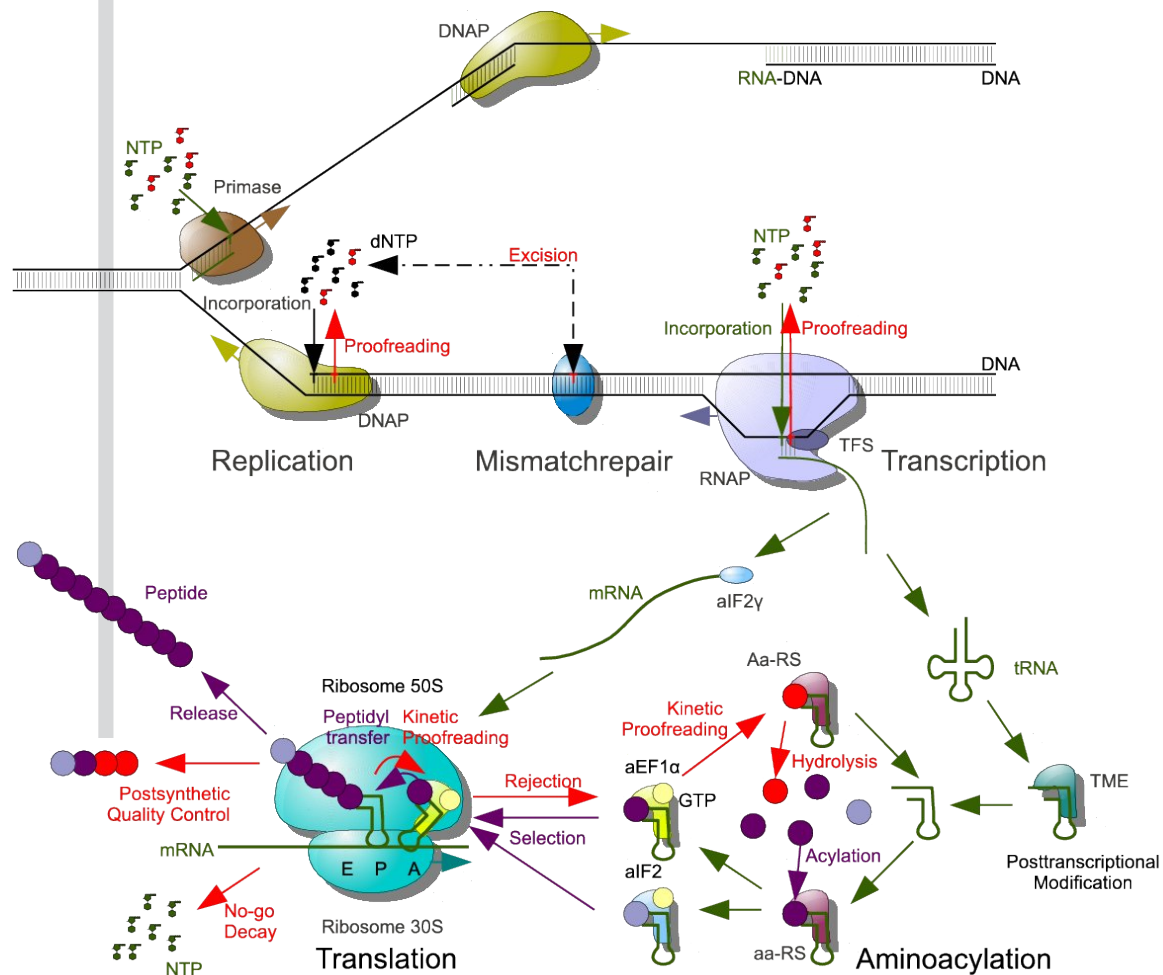
Figure 1. Overview of the processes involved in genetic Information Processing in Archaea. (TFS Transcription Factor S; TME, tRNA modifying enzymes; aa-RS, aminoacyl-tRNA synthetase; aIF2(γ), archaeal Initiation Factor 2(γ); aEF1α, archaeal Elongation Factor 1α).

Transcription

During transcription mRNAs are generated by a DNA-dependent RNA polymerase (RNAP). The polymerization reactions of RNA and DNA shows several similarities, e.g. the course of nucleic acids through the active centre, and the mechanism of substrate binding, as reflected by the similar location of the two metal binding sites in the active sites of both polymerases [45]. Despite these similarities there are also several differences: (i) RNA polymerization incorporates NTPs instead of dNTPs, (ii) most RNAPs, with the exception of bacteriophage and mitochondrial RNAPs, are complexes that consist of 5-15 polypeptide subunits, in contrast to most DNAPs and primases that contain a single or only a few subunits, and (iii) while in DNAPs the newly formed DNA duplex persists, the newly formed RNA is removed from the DNA-RNA hybrid in RNAPs after which the original DNA duplex is restored [45]. Two processes are relevant in terms of transcription fidelity: base selection, and proofreading; post-synthetic correction of RNA does not exist, although some systems exist to monitor the quality of the transcripts that are used as templates during translation. These surveillance systems occur mainly during translation and will be discussed in that section (below). Although the fidelity of the transcription process is considerably lower than that of the replication process, it has been reported to be less than one error every 10^5 nucleotides that are being transcribed in organisms ranging from *E. coli* to wheat [46–48].

RNA polymerases

The RNAP of Bacteria a relatively simple complex consisting of 5 subunits. In addition, a set of up to 20 sigma factors allows for promoter selection in response to changing conditions. Eukaryotes use up to five variant RNAP complexes (I – V) that are responsible for transcription of distinct genes: ribosomal RNAs (RNAP I), protein-coding messenger RNAs (RNAP II), transfer RNAs and other small non-coding RNAs (RNAP III). RNAP IV and RNAP V are restricted to plants and transcribe small RNAs involved in silencing [49]. RNAP I and III are similar to RNAP II, but have some additional subunits that vary between the two. Archaea, in contrast, have only a single RNAP complex that contains 12 orthologous subunits of the eukaryotic RNAP II. There appear to be minor variations among the complexes of the archaeal phyla [50, 51]. For instance the RNAP from *Sulfolobus shibatae* has an additional subunit in comparison to the eukaryotic RNAP II (Rpo13) that has been proposed to play a role in the formation of the transcription bubble [52]. The subunits of these RNA polymerases can be assigned to three different functional groups: (i) the "catalytic core" (the large subunits A'A", and B'B"; in some Archaea these subunits are fused as in Bacteria and Eukaryotes) that harbours the active site, (ii) the "assembly platform" (D, N, L, and P), and (iii) the "auxiliary subunits" (H, K, F, E, and Rpo13). The latter auxiliary set is the part of the complex that differs between the archaeal and the different eukaryotic RNAPs. These subunits that are not required for *in vitro* transcription, but important to stabilize interactions with RNA (F/E stalk), DNA (H and Rpo13) and transcription factors (F/E stalk). Additionally, the F/E stalk is found to be important for processivity during elongation, and correct recognition of weak terminators during termination [51, 53]. Recently it was shown that subunit H is required during promoter opening and initial transcription, and that it, in contrast to its eukaryotic counterpart Rpb5, undergoes a structural rearrangement in the transition from initiation complex to elongation complex that might be specific for archaeal RNAPs [54]. It was also shown recently that *in vitro* reconstitution of the archaeal RNAP is similar in the presence or absence of subunit P. Apparently it does not play a key role in establishing the assembly platform *in vitro*. In addition,

subunit P seems to be involved in open complex formation [55]. Interestingly, a putative ortholog of Rpc34, which is a part of the eukaryotic RNAP III, has recently been found to be present in all crenarchaeal and thaumarchaeal genomes, as well as in several euryarchaeal genomes. This finding suggests that in Archaea the single RNAP might use a variable set-up of auxiliary proteins to transcribe different sets of transcripts [56]. Archaeal RNAPs can be reconstituted from single heterologously expressed subunits in contrast to eukaryotic RNAPs [57, 58]. Recent success with a hybrid archaeal enzyme that contain subunits Rpb5 and Rpb12 from Eukaryotes confirms the high structural similarity of the archaeal and the eukaryotic RNAPs [55, 59].

NTP selection and induced fit

RNAPs discriminate NTPs over dNTPs by recognizing the 2'-hydroxyl group of incoming NTPs. Selection of NTPs by RNAPs is performed by measuring the base pair geometry, in a similar manner as in DNAPs, in a two step process. In the pre-insertion state of the open active center the NTP can come in. If the NTP is complementary to the template nucleotide, the catalytic subunit undergoes a conformational change to the closed state, after which NTP is delivered to the insertion site. This rearranges the active site in such a way that it promotes polymerization by induced fit. If a non-complementary nucleotide is incorporated, the complex enters an off-line state, in which elongation is slowed down considerably [60].

Proofreading

Incorporation of a non-complementary nucleotide induces an inactivated state, in which the nucleotide is frayed. The fraying sites of the RNAP overlap with the NTP-binding site, and as such the frayed nucleotide does not allow elongation to proceed. This paused RNAP complex favours backtracking, a process in which the RNAPs moves one nucleotide backwards. During this process the misincorporated nucleotide is moved from the fraying site to the proofreading site. Multi-subunit RNA polymerases contain an intrinsic nucleolytic RNA cleavage activity that hydrolyses a phosphodiester bond to remove the last two nucleotides as a dinucleotide, resulting in a new RNA 3'-OH group and an empty NTP-binding site. This restores an active on-line state ready for elongation again [60]. This process of backtracking and subsequent cleavage is transcriptional proofreading, and was also described in Archaea. In contrast to Bacteria and Eukaryotes it was found that elongation in Archaea could not continue after misincorporation, but stalled completely instead. TFS, like its eukaryotic homolog TFIIS and its bacterial non-orthologous counterparts GreA/GreB, is known to induce the cleavage activity by direct interaction with the active centre of the polymerase through the nucleotide entrance pore, and could therefore rescue stalled elongation complexes. Stalling of the elongation complex in Archaea appears to be an important trigger for TFS induced cleavage *in vitro* [61]. *Methanopyrus kandleri* has lost TFS during its evolution. Interestingly, this organisms shows a higher mutation rate in comparison with closely related organisms, making it difficult to reconstruct its phylogeny. Especially genes encoding proteins related to transcription are affected, and could include compensatory mutations for the loss of TFS [12].

Protein synthesis

The overall mis-sense substitution rate of *in vivo* bacterial protein synthesis by ribosomes is in the range of 6×10^{-4} to 5×10^{-3} per amino acid [62, 63]. In line with those findings are measurements of the rate of misreading in *Sulfolobus in vitro* translation systems: 3×10^{-3} incorrect leucine incorporations per amino acid on a poly(U) template [64]. Rates of misincorporations during replication, transcription, and aminoacyl-tRNA synthesis are all lower, showing that the final step, the translation process itself, is decisive with respect to fidelity of protein synthesis. The importance of fidelity during protein synthesis is reflected in the organization and evolution of the genetic code. The presumable primordial genetic code that codes for an original set of 10 amino acids [65], as well as the 20 amino acid genetic code, operating in extant cellular life forms, are relatively robust, as most misincorporations will result in substitutions by physico-chemically related amino acids that only in rare occasions will lead to a non-functional protein [65]. Fidelity was thus a key determinant in the evolution of the genetic code. Two separate processes are distinguished during protein synthesis: the coupling of amino acids to their respective tRNAs by a set of specific aminoacyl-tRNA synthetases, and the actual translation itself by ribosomes. In the next paragraphs both processes will be discussed, after which it will be concluded with an overview of the mRNA surveillance systems that are used to avoid the re-use of erroneous templates.

tRNA modification

tRNA molecules are among the most strongly modified RNAs. This mainly concerns nucleotides that are located within the 3D-core and in the anticodon arm, especially at the wobble position N34 and position N37 (conventional numbering). At present over 120 different post-transcriptional modifications of nucleotides have been described, ranging from quite simple methylations to very complex multistep transformations [66]. These nucleotide modifications are important for cellular functionality of tRNAs: they lower conformational flexibility, improve (thermal) stability, and improve aminoacylation rate and specificity. Interestingly, it is known that lack of modification in *in vitro* translation systems can be compensated by excess of magnesium ions, indicating the importance to lower flexibility of tRNAs for translation (reviewed in [67]). Modifications of the wobble position N34 are common in all three domains of life and contribute to accuracy and efficacy of decoding during translation. These modifications are specific and vary between tRNAs. In contrast to unsplit codon boxes in the genetic code, tRNAs coding the split codon boxes are always modified at N34, suggesting that modifications play an important role in increasing the discriminative characteristics between near-identical codons. Remarkably, many modifications of N34 are restricted to specific phylogenetic Domains, or even to lower taxonomic groups, and come with an enormous diversity. This suggests that the corresponding modification enzymes evolved after the divergence of the three domains, and that the extension of the primordial code, and the accompanying increasing need for higher discrimination capacity, has led to a multitude of solutions (reviewed in [68]).

One of these wobble modifications in Bacteria and Eukaryotes is the conversion from G34 to queuosine. The replacement of guanine in this process is catalysed by the enzyme tRNA-guanine Transglycosylase. In Archaea a related enzyme catalyses also the replacement step in the conversion from a guanine to a the positively charged archaeosine at position 15 [69]. G15 is part of the Levitt base-pair, which is the base-pair between N15 of the D-loop and N48 in the variable loop at the start of the T-loop, [70] and also interacts with N59 in the

T-loop (Figure 2) [71]. These interactions between the D- and the T-loops establish the L-shape of the tRNA, indicating that formation of archaeosine is involved in stabilization of the RNA molecule. In Eukaryotes and Bacteria, where position 15 is not restricted to a G, and other variants of the Levitt base-pair exists, stabilization of the Levitt base-pair is brought about by Mg^{2+} binding. Interestingly, binding of a metal, which is less stable at high temperatures than chemical modification, is not compatible with archaeosine formation, suggesting distinct evolutionary mechanisms to stabilize the L-shaped structure of tRNAs between the Domains [72]. For modification of the deeply buried position 15, but probably also for other modifications, the tRNA has to adopt a different configuration, the λ-form. The energetics involved in such rearrangement suggest that modification enzymes might act together in a tRNA maturation complex [71]. Modifications in tRNAs are important for fidelity, processivity, and velocity of translation as they can directly affect decoding for example by modifications in the anti-codon loop or in sites that are recognized by aminoacyl-tRNA synthetases, or indirectly by decreasing the flexibility and increasing the stability of the molecule.

Aminoacylation

The specific coupling of amino acids to their tRNAs yields aminoacyl-tRNAs (aa-tRNAs) and is catalyzed by specific aminoacyl-tRNA synthetases (aaRSs). Two classes (I and II) of aaRSs are distinguished on the basis of their structural topology of the active site [73]. Class I aaRSs, are generally monomeric, attach to the minor groove of the tRNA acceptor stem, and aminoacylate the terminal adenosine of the tRNA at the 2'-OH position, while Class II are generally multimeric, attach to the major groove, and aminoacylate the 3'-OH position [74]. Aminoacylation is a two-step process. First the amino acid is activated using ATP, forming the intermediate aminoacyl-adenylate. Once activated, the amino acid is transferred to the 3' adenosine of the corresponding tRNA [74]. In Archaea and Eukaryotes aaRSs are often organized in higher order complexes that contain multiple aaRSs and other cellular factors, e.g. the large multi-aminoacyl-tRNA synthetase complex in *Haloarcula marismortui* that might harbour all aaRSs [75], or the LysRS-LeuRS-ProRS complex in *Methanothermobacter thermoautotrophicus* that increases the kinetics of LysRS and ProRS [76]. In Eukaryotes complex formation is sometimes also associated with other non-canonical functions like translational silencing, transcriptional control, or anti-apoptosis (reviewed in [74]).

A cell might contain over 25 different types of aa-tRNAs [77]. For translation purposes there are 20 canonical elongator tRNAs, usually acylated by the corresponding synthetases, and an initiator aa-tRNA, acylated by methionyl tRNA synthetase. In Bacteria and eukaryotic organelles, the initiator Met-tRNA^{Met} is subsequently formylated by a specific formyltransferase, in contrast to the situation in Eukaryotes and Archaea. In addition, a small number of non-canonical elongator tRNAs have been discovered (selenocysteinyl-tRNA, and pyrrolysyl-tRNA; see below). After coupling, aa-tRNAs are screened for their correctness by the translation elongation factor EF-Tu (eEF1A/aEF1α in Eukaryotes and Archaea) and delivered to the ribosome, with the exceptions of the initiator aa-tRNA that is verified and delivered by translation initiation factors, and selenocysteinyl-tRNA that is verified and delivered by SelB.

The second major group of aa-tRNAs is composed by mis-acylated translation substrates. A part is due to mistakes by the synthetases. Because elongation factor EF-Tu verifies aa-tRNAs before delivery to the ribosomes, and due to rapid editing by synthetases these errors are low: in most cases once in 10^6 events or less [78, 79]. aaRSs have special editing domains, which are located at a distant position from the synthetic domain, to decouple amino acids from
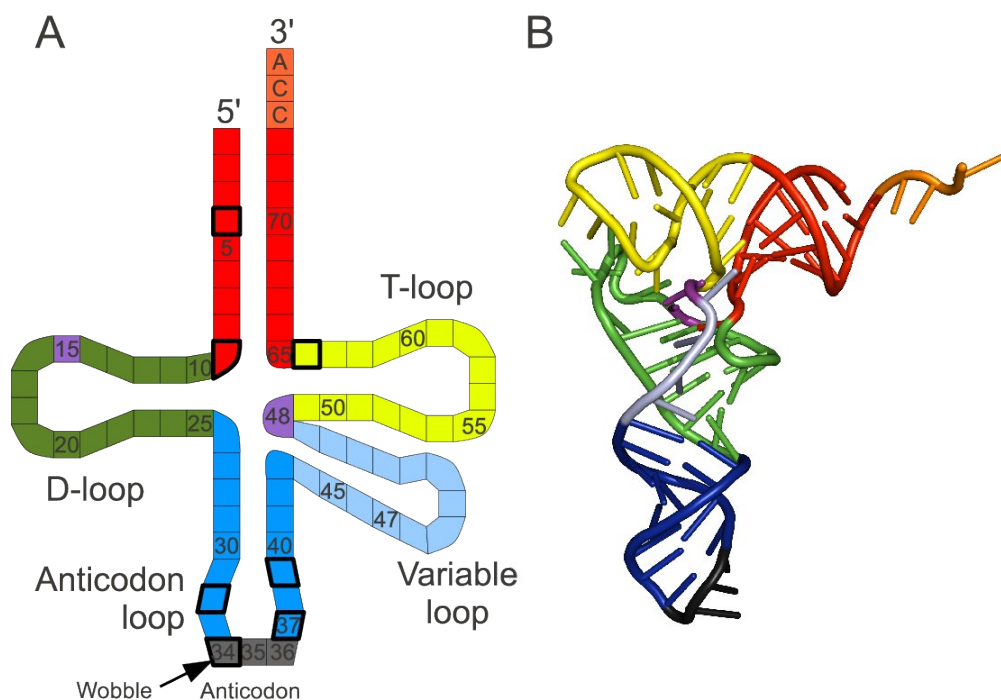
Figure 2. tRNAs. (A) Schematic representation, showing the D-loop (green), anticodon loop (blue) that harbours the anticodon (grey), variable loop that is variable in length (light blue), the T-loop (yellow), the acceptor stem (red), and the CCA aminoacyl binding site (orange). The Levitt base pair is coloured purple. Nucleotides with thick boxes are often modified with variable modifications. (B) Tertiary structure of a yeast tRNAPhe, coloured similar to (A). Figure is rendered with PyMOL from data deposited in the Protein Data Bank (1ehz).

mis-acylated tRNAs. It has been suggested that amino acid selection of the aaRSs depends on a double sieve mechanism, in which the substrate selection at the editing site is the inverse of the substrate selection at the synthetic site. E.g. during coupling at the synthetic site, the amino acids larger than the cognate will be rejected. Then subsequent translocation to the editing site takes place where amino acids smaller than the cognate will be removed [80]. Unfortunately this model is not complete as the editing site of some aaRSs can still edit on the basis of substrate selection present at the synthetic site. In a more recent model for class I aaRSs, it is proposed that the resting state of an aaRSs has the CCA of a bound tRNA at the editing site. When the intermediate aminoacyl-adenylate is formed in the synthetic site, the CCA of the tRNA is translocated to the synthetic site, allowing aminoacyl transfer from the adenylate to the CCA. After that the aminoacylated-CCA is translocated back to the editing site, allowing inspection, and subsequent hydrolysis or release of the aa-tRNA. This model uses two translocation actions providing the opportunity for kinetic proofreading (discussed below) [78]. Besides the editing domains available in aaRSs themselves, free-standing editing proteins, homologs to aaRSs that lack the acylation domain, also exists in all three domains [81–83].

In addition to accidentally mis-acylated tRNAs, there is also a group of aa-tRNAs that is deliberately mis-acylated by aminoacyl-tRNA synthetases, and are subjected to pre-translational amino acid modification. In a large number of Archaea, for example *Methanothermobacter thermautotrophicus* [84], and Bacteria glutamate and aspartate are coupled to tRNA^{Gln} and tRNA^{Asn} respectively by a nondiscriminating aaRS, and then converted by a tRNA amidotransferase into Gln-tRNA^{Gln} and Asn-tRNA^{Asn}. Other deliberate mis-acylation pathways include cysteinyl-tRNA^{Cys} (via *O*-phosphoseryl-tRNA^{Cys}) in methanogenic Archaea [85], and selenocysteinyl-tRNA^{Sec} (via seryl-tRNA^{Sec}) [86] (reviewed in [77]).

Translation

Polymerization of amino acids is catalyzed by the ribosome, a large ribonucleoprotein complex that consists of 3-4 ribosomal RNAs and a large number of ribosomal proteins [87]. Archaeal translation is initiated by recognition of the small ribosomal subunit (30S) of an initiation codon, and the formation of the initiation complex, which includes the initiation factors, the initiator tRNA (Met-tRNA^{Met}) and mRNA. When the initiation complex is formed the large subunit (50S) joins and the monomeric 70S ribosome is formed. Several mechanisms are known for initiation site recognition. Best known for prokaryotes is the mechanism that is associated with a Shine-Dalgarno (SD) motif that is recognized by the anti-SD motif on the 16S rRNA of the 30S. Although it is best known, it is not primarily used by all Bacteria or Archaea. *Sulfolobus* and *Pyrobaculum* for example use the SD mechanism only on distal cistrons of polycistronic transcripts, and not for the first cistron [88, 89]; in addition, *Haloarchaea*, hardly make use of this mechanism at all [90]. In Eukaryotes, that are devoid of the SD mechanism, the 40S cannot interact directly with mRNA, but needs mediation by the 5'-cap binding complex eIF4F. After binding of mRNA, it scans the RNA for an initiation codon by moving in the 3' direction. Once located, the 60S joins the complex, the initiation factors leave, and elongation can start [91]. Less frequently, Eukaryotes use an IRES dependent recognition mechanism, in which the complex IRES structures, that are located in the 5'-UTR, are recognized by IRES-binding transacting factors that are involved in recruitment of the small subunit [92]. All three domains of life also contain leaderless mRNAs, transcripts that start with 5'-terminal initiation codons, and that can be efficiently translated by all ribosomes regardless of the source [93, 94]. While leaderless transcripts are rare in Bacteria and Eukaryotes, they are abundant in many Archaeal species, being the primary mechanisms for monocistronic mRNAs and opening cistrons [88–90, 95]. It is thought that these leaderless transcripts are relics of primitive translation systems [93]. Recently a novel mechanism has been identified in *Haloarchaea*, and although the exact molecular details are unknown, it has been demonstrated to act on transcripts that do not contain SD nor IRES motifs, however the efficiency of their translation depends on the 5'-UTR sequence involved [94].

On the basis of structural and chemical similarities between the homologous systems, translation elongation in Archaea is most likely very similar to that in Bacteria and Eukaryotes. Bacterial translation elongation occurs as follows: first a ternary complex, which consist of an aminoacyl-tRNA, elongation factor EF-Tu (eEF1A/aEF1α in Eukaryotes/Archaea), and GTP, is delivered to the Aminoacyl (A)-site. This complex reacts with the peptidyl-tRNA harboring the Peptidyl (P)-site. During this reaction, that is discussed below in more detail, the peptidyl is transferred to the aminoacyl-tRNA, elongating the nascent chain by one amino acid. Third the peptidyl-tRNA in the A-site and the deacylated tRNA in the P-site move one position to the P

and the Exit (E)-site respectively, leaving the A-site empty and ready for a new round. Energy for this translocation, in which also the accompanying mRNA moves accordingly, is delivered by GTP hydrolysis by EF-G (eEF2/aEF2 in Eukaryotes and Archaea). Accuracy of the ribosome depends on (i) kinetic proofreading, (ii) induced fit, and (iii) post-peptidyl transfer quality control that will be discussed in more detail in the next paragraphs (reviewed in [14]).

tRNA selection by kinetic proofreading and induced fit

Kinetic proofreading is a mechanism that allows discrimination between small energetic differences with low error rates by repeated usage of those differences in distinct separate steps and by coupling them to high energy intermediates. The error rate drops exponentially proportional to the number of repetitions [96, 97]. During translation elongation, the energetic difference between the codon and anti-codon is measured first during the encounter between ribosome and the ternary complex (initial selection), and then again after hydrolysis of GTP, which is irreversible, when the ribosome associates with either the ternary complex (with GDP instead of GTP), or the free aminoacyl-tRNA when EF-Tu is dissociated (proofreading) [14].

Recent models based on bacteria show that decoding is composed out of seven steps [98]: (1) *Initial binding.* The exceptionally fast codon-independent interaction of the ternary complex to the ribosome is determined by EF-Tu and the ribosome, probably with a key role for the L7/L12 stalk. (2) *Codon recognition.* The formation of a complementary codon-anticodon at the decoding centre, what is reflected by a correct (presumably Watson-Crick) geometry, induces conformational changes in the 16S rRNA, while near-cognate geometry induces a different structural change that leads to an almost 1000-fold higher dissociation rate, although recognition rates remain almost similar. (3) *GTPase activation.* The GTP hydrolysis rate is increased by binding of cognate tRNAs compared to near-cognate binding. The local 16S rRNA conformational changes upon cognate binding (step 2) lead to a closed conformation of the 30S ribosomal subunit. This conformational signal is communicated to the 50S ribosomal subunit and affects EF-Tu GTP hydrolysis. Near-cognate binding induces a different structural change in the decoding centre what most probably does not lead to the closed conformation of the 30S subunit, and thereby does not affect EF-Tu GTP hydrolysis. Slowing down hydrolysis increases discrimination capacity, however at the cost of velocity. (4) *GTP hydrolysis*. The rate of GTP hydrolysis by EF-Tu depends on the activation state of EF-Tu. (5) *Conformational change of EF-Tu.* EF-Tu changes from the GTP-form to the GDP form. This conformational change is limited by the rate of inorganic phosphate release. EF-Tu releases the aa-tRNA probably during the transition. (6) *Accommodation or rejection.* After release, the 3' end of the aa-tRNA has to move almost 70 Å from its binding site on EF-Tu to the Peptidyl Transferase Centre (PTC), while the codon-anticodon interaction should remain intact. Accommodation in the PTC of cognate aa-tRNA is rapid and efficient, in contrast to near-cognate aa-tRNA that is mostly rejected because of the low stability of binding and lower rate for accommodation. (7) *Peptidyl-transfer.* The Peptidyl chain is transferred from the aa-tRNA on the P-site to the aa-tRNA on the A-site, elongating the nascent peptide with one amino acid. Initial selection occurs in step (1) to (3), while proofreading occurs in step (6).

The ribosome not only uses kinetic proofreading to improve its selectivity it also uses an additional principle to further improve it: induced-fit. Induced-fit is a principle in which the correct substrate induces a conformational change leading to an acceleration of the desired process, while an incorrect substrate has the opposite effect. During decoding a correct codon-

anticodon interaction accelerates both GTPase activation and accommodation steps, while a non-correct near-cognate interaction inhibits both steps, leading to rejection [98]. It was reported that near-cognate tRNAs showed an increase in GTP consumption relative to the amount of amino acids incorporated, while other non-cognate tRNAs did not. This suggests that non-cognate tRNAs are already rejected during the second step, whereas near-cognate are expelled during the fifth step after GTP hydrolysis, showing the importance of kinetic proofreading and induced fit for reliable descrimination between cognate and near-cognate tRNAs [99].

Post-Peptidyl Transfer Quality Control

The molecular characteristics of post-peptidyl transfer quality control (PPQC) have only recently been discovered in detail using bacterial *in vitro* systems [100], but it is likely to occur also in Eukaryotes and Archaea. Like proofreading in nucleotide polymerization, the ribosome senses mismatching after the polymerization reaction (peptidyl transfer in this case). However, where in replication and transcription exonucleases could erase the mistake, the ribosome should take more drastic actions to undo translation errors: abortive termination of the nascent peptide chain. Trigger for PPQC is mismatching between tRNA and template at the P-site of the ribosome. A mismatch at the P-site increases selection of non-cognate tRNA at the A-site dramatically. After peptidyl transfer and subsequent translocation, the nascent chain contains two wrong subsequent amino acids, and both E-site as P-site harbour a mismatching tRNA. Mismatching at both E- and P-site leads to strongly stimulated release of the nascent peptide chain, increasing the rate constants for release in a range comparable to tRNA selection due to increased binding of Release Factors [100].

Termination

Translation terminates when a stop codon reaches the A-site. Unlike other codons a stop-codon is recognized by proteins that mimic tRNAs: class-1 release factors (RF1s). These factors induce hydrolysis of peptidyl-tRNA, disconnecting the nascent chain from the tRNA. While Bacteria use two release factors (RF1 and RF2) that recognizes different stop-codon pairs (UAA/UAG and UAA/UGA respectively), most Archaea and Eukaryotes have a single one (aRF1 and eRF1 respectively) that recognize all three stop-codons. Results from experiments with genuine archaeal release factors and archaeal/eukaryotic chimeras in eukaryotic *in vitro* translation systems suggest similar mechanisms for both [101]. An interesting variant on this theme is found in pyrrolysine-utilizing Archaea. Pyrrolysine (Pyl), the 22nd amino acid, is only found in some archaeal species belonging to the Methanosarcinaceae and two Bacteria (Desulfitobacterium hafniense and an uncultured δ-proteobacteria) [102]. It is encoded by the amber stop codon (UAG). Pyl-tRNA[Pyl] is normally recognized by EF-Tu, implicating normal incorporation during elongation [103]. *Methanosarcina barkeri* contains two RFs of which only one appears to be active in termination: it was found that aRF1-1 (at least when combined as a archaeal/eukaryotic chimera) had a lower release efficiency for the UAG codon than for UAA or UGA. Comparative genomics also showed that pyrrolysine-utilizing Archaea avoid UAG as a stop codon. This suggests that in these Archaea the genetic code is changed to incorporate Pyrrolysine instead of termination [101]. Reassigning stop codons is not restricted to Archaea: the Eukaryotic ciliates *Tetrahymena thermophila* and *Euplotes aediculates* reassigned stop codons, UAG and UAA to glutamine, and UGA to cysteine respectively, and changed specificity of their eRF1s accordingly [104, 105]. More prominent and present in all domains of

life is selenocysteine (Sec), the 21st amino acid. In Archaea selenoproteins are found in *Methanococcus*, *Methanocaldococcus* and *Methanopyrus* species [106]. Sec is encoded by the opal stop codon (UGA). However in contrast to Pyr incorporation, Sec incorporation needs a special elongation factor (SelB) that via an extended domain recognizes a mRNA hairpin loop downstream of a UGA codon (the selenocysteine insertion element or SECIS) [107]. Binding of Sec-tRNA^{Sec}-SelB-GTP to this structure leads to insertion of Sec at in-frame UGA codons. In contrast to Bacteria, the SECIS element is located outside of the coding region in Archaea and Eukaryotes, while the archaeal and eukaryotic SelB contain considerably shorter extensions. To overcome the distance, Eukaryotes evolved an additional adapter protein (SBP2). Additionally, it was found that the ribosomal protein L30 binds SECIS elements and influences Sec insertion. Although a similar mechanism is proposed for Archaea the adapter protein is not found [86].

mRNA surveillance

In Eukaryotes mRNA quality control processes exist that act during translation to ensure the quality of the transcripts. These processes, called mRNA surveillance, are dependent on the eukaryotic release factors eRF1 and eRF3 and their paralogs Dom34 (synonym Pelota), and Hbs1 and Ski7 respectively. Three mRNA surveillance pathways are known in Eukaryotes: (i) Nonsense Mediated Decay (NMD): when premature stop codons are encountered, (ii) No-go Decay (NGD): to release stalled ribosomes, and (iii) Non-stop Decay (NSD): to rescue ribosomes that have read through a stop codon. NMD and NSD are restricted to Eukaryotes: as eRF3 and other components of the NMD system are missing in Archaea. Ski7, necessary for NSD, is even only present in the *Saccharomycetales*, although recent findings suggest that Hbs1 could take over what could mean that NSD is more widespread in Eukaryotes. In contrast, Dom34, necessary for No-go Decay is also found in Archaea. This suggests that NGD might be functionally present, although Hbs1p and eRF3 are missing in Archaea [108]. In Eukaryotes a ternary complex Dom34-Hbs1-GTP is formed, similar to the formation of the eRF1-eRF3-GTP complex used in eukaryotic translation termination. This Dom34-Hbs1-GTP ternary complex is able to recognize stalled ribosomes, leading to endonucleolytic cleavage of mRNA by Dom34 [109, 110]. In Archaea aRF1 is able to terminate translation without the help of a RF3 ortholog, what could imply that the paralogous Dom34 might be able to perform NGD without the help of an Hbs1p ortholog [108].

To rescue stalled ribosomes Bacteria have a system that uses an intermediate between tRNA and mRNA: tmRNA, a RNA molecule with a tertiary structure similar to tRNAs, but with an extended anticodon loop that contains a mRNA-like ORF. If the ribosome stalls, because a transcript is finished without a proper termination, tmRNA in concert with SmpB and EF-Tu binds to the empty A-site of the stalled ribosome. After translocation to the P-site the mRNA-like ORF located in the anticodon loop of the tmRNA takes over the role of messenger, and encodes for a degradation tag and ends with a proper stop-codon. After release the nascent peptide is thus tagged for degradation, and the ribosomal subunits are released again. This system seems to be restricted to Bacteria as tmRNA genes have not been identified in Eukaryotes, with the small exception of a few eubacterial-like organelles, or Archaea [111]. Interestingly, in investigations of archaeal protein degradation in *Methanococcus jannaschii,* green fluorescent proteins tagged with a ssrA-extension were used. The ssrA extension is the 11 amino acid degradation tag encoded on the tmRNA, which gene was designated *ssrA*. Tagged proteins showed a rapid unfolding and degradation while untagged proteins did not [112].

Turnover of RNA and proteins

RNA decay

Beside above mentioned mRNA surveillance during translation, more general systems are involved in RNA turnover. Main component in these mechanisms in Archaea is the exosome, a protein complex that includes Rrp41 and Rrp42, a homolog of RNasePH, a bacterial phosphorolytic nuclease, and Rrp4 and Csl4, containing KH/S1 RNA-binding domains. The archaeal exosome is responsible for $3' \rightarrow 5'$ degradation of RNA, as well as for 3' polyadenylation. This complex is similar to the bacterial PNPase and the eukaryotic exosome. All three have a double-doughnut-like structure with a central hole with a core ring of six RNasePH-type subunits. The narrow neck of the archaeal structure only allows single-stranded RNA devoid of secondary structures, suggesting a regulatory role for cofactors, as observed in Eukaryotes [113–115].

Polyadenylation occurs mainly on fragmented molecules as part of a RNA decay pathway in Bacteria, Archaea and eukaryotic cell organelles, and recently has been described for nuclear genes from Eukaryotes as well. Although, in contrast, in Eukaryotes poly(A) tails are also added to mature 3' ends of most nuclear encoded, full-length, mRNAs for proper translation initiation, and mRNA stability. The general scheme of RNA $3' \rightarrow 5'$ degradation in prokaryotes is as follows: (1) removal of the 5' pyrophosphate (2) endonucleolytic cleavage of the transcript, (3) poly-adenylation of cleavage products, and (4) rapid exonucleolytic degradation of poly-adenylated products. In *Sulfolobus* the exosome is able to generate a heteromeric poly(A)-rich tail and use NDPs as a substrate. It has been suggested that polyadenylation is used to overcome secondary RNA structures that otherwise cannot pass the exosome neck. Interestingly, halophilic Archaea, together with several methanogenic Archaea, like *Haloferax*, and *Methanococcus*, are the only known organisms that lack polyadenylation, and do not contain an exosome or PNPase. In these organisms poly(A)-independent RNA degradation is performed by RNase R [116–119].

Eukaryotes also use another pathway for mRNA degradation that involves $5' \rightarrow 3'$ exonucleases, like Xrn1p. Eukaryotic transcripts are protected against this rapid form of decay by a 5' cap. To prevent transcripts from being decapped unintentionally, they are protected by the eukaryotic translation initiation factor eIF4E [120]. In Archaea, mRNAs are similarly protected from $5' \rightarrow 3'$ decay by binding of the γ-subunit of the archaeal translation initiation factor aIF2 to the 5' end. The similarities between both systems suggests that $5' \rightarrow 3'$ decay is common to all domains of life [121]. Additionally, this protection offers a mechanism to discriminate between new versus already translated transcripts. After translation aIF2 is removed from the mRNA, making the mRNA vulnerable to $5' \rightarrow 3'$ decay as soon as translation is terminated. Interestingly, a tight coupling beyond the use of an intiation factor to protect mRNA, exists between transcription and translation in Archaea, as it has been found that multiple rounds of translation already start before transcription is finished [122]. This tight interplay might have to be extended to mRNA degradation as well, what would provide Archaea with a very efficient and short information processing pipeline.

The protein waste bin

The 20S proteasome, present in Eukaryotes, Archaea, and actinobacteria, is a barrel shaped complex that consists of four heptameric rings of α- and β-type subunits in a α7β7β7α7 configuration. Other Bacteria use the simpler HslV protease, that is structurally related to the β-type subunits of the 20S proteasomes. The function of the proteasome is to breakdown

proteins into short peptides that in turn can be further degraded to amino acids by peptidases to be recycled in protein synthesis or in metabolism. The proteasome is therefore an essential component for protein turnover and to maintain protein quality control by degrading misfolded and denatured proteins [123, 124].

The protease domains of β-type subunits are located on the inside of the barrel. This creates a tightly regulated environment, to circumvent uncontrolled protein breakdown. In Eukaryotes the 20S proteasome can be capped by 19S regulatory particles (a combination of a Rpt and Rpn proteins forming a base and a lit), on one side (26S proteasome), or on both sides (30S proteasome). These caps play a role in recognition and degradation of polyubiquitin tagged substrates. Archaea encode orthologs of Rpt called Proteasome-activating Nucleotidase (PAN). PAN is able to unfold proteins in a ATP-dependent manner, can open the axial gate of the 20S proteasome, and subsequently translocates the substrate into the 20S core. Interestingly, as mentioned earlier, the archaeal PAN is able to distinguish between a ssrA-tagged or untagged green fluorescent protein, which suggest a role for ssrA-tagging in peptide degradation in Archaea, although a tmRNA system responsible for ssrA-tagging has not been identified [112, 123, 124].

In Eukaryotes, ubiquitin and ubiquitin-like proteins, small stable proteins that contain a β-grasp fold and that can be attached to a wide variety of other proteins, play an important role in targeted degradation by the proteasome. Ubiquitylation is also used in a number of other non-proteolytic mechanisms like endocytosis, intracellular trafficking, chromotin-mediated regulation of transcription, DNA repair. Discrimination between those target processes is thought to be dependent on the differences in ubiquitin chains [125]. Ubiquitin-targeted degradation is used for quality control in Eukaryotes. Misfolded proteins in the cytosol are recognized by chaperones, because of their toxic hydrophobic surfaces. These chaperones recruit ubiquitylation enzymes (e.g. CHIP), that attach a polyubiquitin chain to the misfolded protein after which it is degraded or refolded [126]. In the lumen of the endoplasmatic reticulum, N-linked oligosaccharides indicate the folding stage, but also appear to keep track of the time a polypeptide resides within the lumen. If misfolding occurs and the polypeptide is trapped within the lumen, they are directed to a ubiquitin ligase and targeted for destruction [127].

Although ubiquitin-like tagging was long thought to be restricted to Eukaryotes, an ubiquitin-like tagging system was recently revealed in *Haloferax volcanii* that is able to tag proteins with small archaeal modifier proteins (SAMPs). SAMPs are small proteins that contain a β-grasp fold and a C-terminal diglycine motif similar to ubiquitin, and are widespread among the Archaea. It was shown that SAMPs are coupled to a wide range of proteins. SAMP1 appears to target proteins for destruction by the proteasome [128]. Alternative signalling objectives might also be present, as SAMP2 was also found to be coupled to a wide range of proteins like SAMP1, but showed decreasing levels in proteasomal mutant strains [128]. It seems to be likely that systems similar to eukaryotic ubiquitin-targeted systems, like targeted destruction, is also present in the archaeal domain. Opening up a potential role for the proteasome in regulation of protein levels, quality control against misfolded proteins, and recycling of non-functional polypeptides in archaeal cells.

Concluding Remarks

It is obvious that a certain level of fidelity of genetic information processing is of major importance to the cell, in order to maintain the delicate balance between accuracy on the one hand and velocity on the other. At the moment, a rather complete picture is emerging for the three main polymerization reactions related to genetic information processing in living cells in general. More and more is known about the mechanisms and the role of factors that contribute to fidelity in these systems. To some extent these crucial cellular processes have successfully been studied in selected Archaea. Despite this progress, however, it is obvious that insight in fidelity-related mechanisms in Archaea is still relatively scarce. For that reason, extrapolations on the basis of analogous systems of Bacteria and Eukaryotes have been used in this overview to bridge the gaps in our understanding of the archaeal counterparts. Although we think that most of the described processes work similarly in Archaea, we cannot rule out that such generalisations may in some instances turn out to be an oversimplification of the actual situation. As many Archaea thrive in extreme environments, it will be very interesting to learn how fidelity mechanisms of these extremophilic organisms are adapted to overcome these harsh conditions. It is therefore anticipated that Archaea will continue to play an important role in future research to elucidate details on the intriguing systems that control the fidelity of information processing.

References

1. Crick FHC: **Central dogma of molecular biology.** *Nature* 1970, **227**:561–3.
2. Mattick JS: **Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms.** *Bioessays* 2003, **25**:930–9.
3. Eigen M: **Selforganization of matter and the evolution of biological macromolecules.** *Naturwissenschaften* 1971, **58**:465–523.
4. Eigen M, Schuster P: **The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle.** *Naturwissenschaften* 1977, **64**:541–65.
5. Wolf YI, Koonin E V: **On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization.** *Biol Direct* 2007, **2**:14.
6. Lindahl T: **Instability and decay of the primary structure of DNA.** *Nature* 1993, **362**:709–15.
7. Drake JW, Charlesworth B, Charlesworth D, Crow JF: **Rates of spontaneous mutation.** *Genetics* 1998, **148**:1667–86.
8. De Visser JAGM, Rozen DE: **Limits to adaptation in asexual populations.** *J Evol Biol* 2005, **18**:779–88.
9. Furió V, Moya A, Sanjuán R: **The cost of replication fidelity in human immunodeficiency virus type 1.** *Proc Biol Sci* 2007, **274**:225–30.
10. Loh E, Salk JJ, Loeb LA: **Optimization of DNA polymerase mutation rates during bacterial evolution.** *Proc Natl Acad Sci U S A* 2010, **107**:1154–9.
11. Metzgar D, Wills C: **Evolutionary changes in mutation rates and spectra and their influence on the adaptation of pathogens.** *Microbes Infect* 2000, **2**:1513–22.
12. Brochier C, Forterre P, Gribaldo S: **Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the Methanopyrus kandleri paradox.** *Genome Biol* 2004, **5**:R17.

13. Thompson RC, Karim AM: **The accuracy of protein biosynthesis is limited by its speed: high fidelity selection by ribosomes of aminoacyl-tRNA ternary complexes containing GTP[gamma S].** *Proc Natl Acad Sci U S A* 1982, **79**:4922–6.

14. Zaher HS, Green R: **Fidelity at the molecular level: lessons from protein synthesis.** *Cell* 2009, **136**:746–62.

15. Kunkel TA: **Evolving views of DNA replication (in)fidelity.** *Cold Spring Harb Symp Quant Biol* 2009, **74**:91–101.

16. Loeb LA, Kunkel TA: **Fidelity of DNA synthesis.** *Annu Rev Biochem* 1982, **51**:429–57.

17. Drake JW: **Avoiding dangerous missense: thermophiles display especially low mutation rates.** *PLoS Genet* 2009, **5**:e1000520.

18. She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CC, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PM, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, et al.: **The complete genome of the crenarchaeon Sulfolobus solfataricus P2.** *Proc Natl Acad Sci U S A* 2001, **98**:7835–40.

19. Tahirov TH, Makarova KS, Rogozin IB, Pavlov YI, Koonin E V: **Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol epsilon and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors.** *Biol Direct* 2009, **4**:11.

20. Henneke G, Flament D, Hübscher U, Querellou J, Raffin J-P: **The hyperthermophilic euryarchaeota Pyrococcus abyssi likely requires the two DNA polymerases D and B for DNA replication.** *J Mol Biol* 2005, **350**:53–64.

21. Fukui T, Yamauchi K, Muroya T, Akiyama M, Maki H, Sugino A, Waga S: **Distinct roles of DNA polymerases delta and epsilon at the replication fork in Xenopus egg extracts.** *Genes Cells* 2004, **9**:179–91.

22. Sanders GM, Dallmann HG, McHenry CS: **Reconstitution of the B. subtilis replisome with 13 proteins including two distinct replicases.** *Mol Cell* 2010, **37**:273–81.

23. Lao-Sirieix S, Bell SD: **The heterodimeric primase of the hyperthermophilic archaeon Sulfolobus solfataricus possesses DNA and RNA primase, polymerase and 3'-terminal nucleotidyl transferase activities.** *J Mol Biol* 2004, **344**:1251–63.

24. Liu L, Komori K, Ishino S, Bocquier AA, Cann IK, Kohda D, Ishino Y: **The archaeal DNA primase: biochemical characterization of the p41-p46 complex from Pyrococcus furiosus.** *J Biol Chem* 2001, **276**:45484–90.

25. Barry ER, Bell SD: **DNA replication in the archaea.** *Microbiol Mol Biol Rev* 2006, **70**:876–87.

26. Russell HJ, Richardson TT, Emptage K, Connolly BA: **The 3'-5' proofreading exonuclease of archaeal family-B DNA polymerase hinders the copying of template strand deaminated bases.** *Nucleic Acids Res* 2009, **37**:7603–11.

27. Broyde S, Wang L, Rechkoblit O, Geacintov NE, Patel DJ: **Lesion processing: high-fidelity versus lesion-bypass DNA polymerases.** *Trends Biochem Sci* 2008, **33**:209–19.

28. Watson JD, Crick FHC: **Genetical implications of the structure of deoxyribonucleic acid.** *Nature* 1953, **171**:964–7.

29. Petruska J, Sowers LC, Goodman MF: **Comparison of nucleotide interactions in water, proteins, and vacuum: model for DNA polymerase fidelity.** *Proc Natl Acad Sci U S A* 1986, **83**:1559–62.

30. Kunkel TA, Bebenek K: **DNA replication fidelity.** *Annu Rev Biochem* 2000, **69**:497–529.

31. Jokela M, Eskelinen A, Pospiech H, Rouvinen J, Syväoja JE: **Characterization of the 3' exonuclease subunit DP1 of Methanococcus jannaschii replicative DNA polymerase D.** *Nucleic Acids Res* 2004, **32**:2430–40.

32. Zhang L, Lou H, Guo L, Zhan Z, Duan Z, Guo X, Huang L: **Accurate DNA synthesis by Sulfolobus solfataricus DNA polymerase B1 at high temperature.** *Extremophiles* 2010, **14**:107–17.

33. Lin Z, Nei M, Ma H: **The origins and early evolution of DNA mismatch repair genes--multiple horizontal gene transfers and co-evolution.** *Nucleic Acids Res* 2007, **35**:7591–603.

34. Busch CR, DiRuggiero J: **MutS and MutL are dispensable for maintenance of the genomic mutation rate in the halophilic archaeon Halobacterium salinarum NRC-1.** *PLoS One* 2010, **5**:e9045.

35. Vijayvargia R, Biswas I: **MutS2 family protein from Pyrococcus furiosus.** *Curr Microbiol* 2002, **44**:224–8.

36. Au KG, Welsh K, Modrich P: **Initiation of methyl-directed mismatch repair.** *J Biol Chem* 1992, **267**:12142–8.

37. Grogan DW: **Stability and repair of DNA in hyperthermophilic Archaea.** *Curr Issues Mol Biol* 2004, **6**:137–44.

38. Birkeland N-K, Anensen H, Knaevelsrud I, Kristoffersen W, Bjørås M, Robb FT, Klungland A, Bjelland S: **Methylpurine DNA glycosylase of the hyperthermophilic archaeon Archaeoglobus fulgidus.** *Biochemistry* 2002, **41**:12697–705.

39. Kiyonari S, Tahara S, Uchimura M, Shirai T, Ishino S, Ishino Y: **Studies on the base excision repair (BER) complex in Pyrococcus furiosus.** *Biochem Soc Trans* 2009, **37**(Pt 1):79–82.

40. Kiyonari S, Tahara S, Shirai T, Iwai S, Ishino S, Ishino Y: **Biochemical properties and base excision repair complex formation of apurinic/apyrimidinic endonuclease from Pyrococcus furiosus.** *Nucleic Acids Res* 2009, **37**:6439–53.

41. Knaevelsrud I, Ruoff P, Anensen H, Klungland A, Bjelland S, Birkeland NK: **Excision of uracil from DNA by the hyperthermophilic Afung protein is dependent on the opposite base and stimulated by heat-induced transition to a more open structure.** *Mutat Res* 2001, **487**:173–90.

42. Sartori AA, Schär P, Fitz-Gibbon S, Miller JH, Jiricny J: **Biochemical characterization of uracil processing activities in the hyperthermophilic archaeon Pyrobaculum aerophilum.** *J Biol Chem* 2001, **276**:29979–86.

43. Schmidt KJ, Beck KE, Grogan DW: **UV stimulation of chromosomal marker exchange in Sulfolobus acidocaldarius: implications for DNA repair, conjugation and homologous recombination at extremely high temperatures.** *Genetics* 1999, **152**:1407–15.

44. Wood ER, Ghané F, Grogan DW: **Genetic responses of the thermophilic archaeon Sulfolobus acidocaldarius to short-wavelength UV light.** *J Bacteriol* 1997, **179**:5693–8.

45. Cramer P: **Common structural features of nucleic acid polymerases.** *Bioessays* 2002, **24**:724–9.

46. Blank A, Gallant JA, Burgess RR, Loeb LA: **An RNA polymerase mutant with reduced accuracy of chain elongation.** *Biochemistry* 1986, **25**:5920–8.

47. Rosenberger RF, Foskett G: **An estimate of the frequency of in vivo transcriptional errors at a nonsense codon in Escherichia coli.** *Mol Gen Genet* 1981, **183**:561–3.

48. De Mercoyrol L, Corda Y, Job C, Job D: **Accuracy of wheat-germ RNA polymerase II. General enzymatic properties and effect of template conformational transition from right-handed B-DNA to left-handed Z-DNA.** *Eur J Biochem* 1992, **206**:49–58.

49. Landick R: **Functional divergence in the growing family of RNA polymerases.** *Structure* 2009, **17**:323–5.

50. Koonin E V, Makarova KS, Elkins JG: **Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerases are conserved in hyperthermophilic Crenarchaeota and "Korarchaeota".** *Biol Direct* 2007, **2**:38.

51. Werner F: **Structure and function of archaeal RNA polymerases.** *Mol Microbiol* 2007, **65**:1395–404.

52. Korkhin Y, Unligil UM, Littlefield O, Nelson PJ, Stuart DI, Sigler PB, Bell SD, Abrescia NGA: **Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure.** *PLoS Biol* 2009, **7**:e1000102.

53. Hirtreiter A, Grohmann D, Werner F: **Molecular mechanisms of RNA polymerase--the F/E (RPB4/7) complex is required for high processivity in vitro.** *Nucleic Acids Res* 2010, **38**:585–96.

54. Grünberg S, Reich C, Zeller ME, Bartlett MS, Thomm M: **Rearrangement of the RNA polymerase subunit H and the lower jaw in archaeal elongation complexes.** *Nucleic Acids Res* 2010, **38**:1950–63.

55. Reich C, Zeller M, Milkereit P, Hausner W, Cramer P, Tschochner H, Thomm M: **The archaeal RNA polymerase subunit P and the eukaryotic polymerase subunit Rpb12 are interchangeable in vivo and in vitro.** *Mol Microbiol* 2009, **71**:989–1002.

56. Blombach F, Makarova KS, Marrero J, Siebers B, Koonin E V, van der Oost J: **Identification of an ortholog of the eukaryotic RNA polymerase III subunit RPC34 in Crenarchaeota and Thaumarchaeota suggests specialization of RNA polymerases for coding and non-coding RNAs in Archaea.** *Biol Direct* 2009, **4**:39.

57. Naji S, Grünberg S, Thomm M: **The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation.** *J Biol Chem* 2007, **282**:11047–57.

58. Werner F, Weinzierl RO.: **A Recombinant RNA Polymerase II-like Enzyme Capable of Promoter-Specific Transcription**. *Mol Cell* 2002, **10**:635–646.

59. Thomm M, Reich C, Grünberg S, Naji S: **Mutational studies of archaeal RNA polymerase and analysis of hybrid RNA polymerases.** *Biochem Soc Trans* 2009, **37**(Pt 1):18–22.

60. Sydow JF, Cramer P: **RNA polymerase fidelity and transcriptional proofreading.** *Curr Opin Struct Biol* 2009, **19**:732–9.

61. Lange U, Hausner W: **Transcriptional fidelity and proofreading in Archaea and implications for the mechanism of TFS-induced RNA cleavage.** *Mol Microbiol* 2004, **52**:1133–43.

62. Bouadloun F, Donner D, Kurland CG: **Codon-specific missense errors in vivo.** *EMBO J* 1983, **2**:1351–6.

63. Edelmann P, Gallant J: **Mistranslation in E. coli.** *Cell* 1977, **10**:131–7.

64. Londei P, Altamura S, Sanz JL, Amils R: **Aminoglycoside-induced mistranslation in thermophilic archaebacteria.** *Mol Gen Genet* 1988, **214**:48–54.

65. Novozhilov AS, Koonin E V: **Exceptional error minimization in putative primordial genetic codes.** *Biol Direct* 2009, **4**:44.

66. Czerwoniec A, Dunin-Horkawicz S, Purta E, Kaminska KH, Kasprzak JM, Bujnicki JM, Grosjean H, Rother K: **MODOMICS: a database of RNA modification pathways. 2008 update.** *Nucleic Acids Res* 2009, **37**(Database issue):D118–21.

67. Helm M: **Post-transcriptional nucleotide modification and alternative folding of RNA.** *Nucleic Acids Res* 2006, **34**:721–33.

68. Grosjean H, de Crécy-Lagard V, Marck C: **Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes.** *FEBS Lett* 2010, **584**:252–64.

69. Sabina J, Söll D: **The RNA-binding PUA domain of archaeal tRNA-guanine transglycosylase is not required for archaeosine formation**. *J Biol Chem* 2006, **281**:6993–7001.

70. Levitt M: **Detailed molecular model for transfer ribonucleic acid.** *Nature* 1969, **224**:759–63.

71. Ishitani R, Nureki O, Nameki N, Okada N, Nishimura S, Yokoyama S: **Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme**. *Cell* 2003, **113**:383–394.

72. Oliva R, Tramontano A, Cavallo L: **Mg2+ binding and archaeosine modification stabilize the G15 C48 Levitt base pair in tRNAs.** *RNA* 2007, **13**:1427–36.

73. Eriani G, Delarue M, Poch O, Gangloff J, Moras D: **Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs.** *Nature* 1990, **347**:203–6.

74. Hausmann CD, Ibba M: **Aminoacyl-tRNA synthetase complexes: molecular multitasking revealed.** *FEMS Microbiol Rev* 2008, **32**:705–21.

75. Goldgur Y, Safro M: **Aminoacyl-tRNA synthetases from Haloarcula marismortui: an evidence for a multienzyme complex in a procaryotic system.** *Biochem Mol Biol Int* 1994, **32**:1075–83.

76. Praetorius-Ibba M, Hausmann CD, Paras M, Rogers TE, Ibba M: **Functional association between three archaeal aminoacyl-tRNA synthetases.** *J Biol Chem* 2007, **282**:3680–7.

77. Ibba M, Söll D: **Aminoacyl-tRNAs: setting the limits of the genetic code**. *Genes Dev* 2004, **18**:731–738.

78. Francklyn CS: **DNA polymerases and aminoacyl-tRNA synthetases: shared mechanisms for ensuring the fidelity of gene expression.** *Biochemistry* 2008, **47**:11695–703.

79. Jakubowski H, Goldman E: **Editing of errors in selection of amino acids for protein synthesis.** *Microbiol Rev* 1992, **56**:412–29.

80. Fersht AR: **Enzymic editing mechanisms and the genetic code.** *Proc R Soc Lond B Biol Sci* 1981, **212**:351–79.

81. Korencic D, Ahel I, Schelert J, Sacher M, Ruan B, Stathopoulos C, Blum P, Ibba M, Söll D: **A freestanding proofreading domain is required for protein synthesis quality control in Archaea.** *Proc Natl Acad Sci U S A* 2004, **101**:10260–5.

82. Wong F-C, Beuning PJ, Silvers C, Musier-Forsyth K: **An isolated class II aminoacyl-tRNA synthetase insertion domain is functional in amino acid editing.** *J Biol Chem* 2003, **278**:52857–64.

83. Ahel I, Korencic D, Ibba M, Söll D: **Trans-editing of mischarged tRNAs.** *Proc Natl Acad Sci U S A* 2003, **100**:15422–7.

84. Oshikane H, Sheppard K, Fukai S, Nakamura Y, Ishitani R, Numata T, Sherrer RL, Feng L, Schmitt E, Panvert M, Blanquet S, Mechulam Y, Söll D, Nureki O: **Structural basis of RNA-dependent recruitment of glutamine to the genetic code.** *Science* 2006, **312**:1950–4.

85. Sauerwald A, Zhu W, Major TA, Roy H, Palioura S, Jahn D, Whitman WB, Yates JR, Ibba M, Söll D: **RNA-dependent cysteine biosynthesis in archaea.** *Science* 2005, **307**:1969–72.

86. Stock T, Rother M: **Selenoproteins in Archaea and Gram-positive bacteria.** *Biochim Biophys Acta* 2009, **1790**:1520–32.

87. Lecompte O, Ripp R, Thierry J-C, Moras D, Poch O: **Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale.** *Nucleic Acids Res* 2002, **30**:5382–90.

88. Benelli D, Maone E, Londei P: **Two different mechanisms for ribosome/mRNA interaction in archaeal translation initiation.** *Mol Microbiol* 2003, **50**:635–43.

89. Slupska MM, King AG, Fitz-Gibbon S, Besemer J, Borodovsky M, Miller JH: **Leaderless transcripts of the crenarchaeal hyperthermophile Pyrobaculum aerophilum.** *J Mol Biol* 2001, **309**:347–60.

90. Brenneis M, Hering O, Lange C, Soppa J: **Experimental characterization of Cis-acting elements important for translation and transcription in halophilic archaea.** *PLoS Genet* 2007, **3**:e229.

91. Kozak M: **Pushing the limits of the scanning mechanism for initiation of translation.** *Gene* 2002, **299**:1–34.

92. López-Lastra M, Rivas A, Barría MI: **Protein synthesis in eukaryotes: the growing biological relevance of cap-independent translation initiation.** *Biol Res* 2005, **38**:121–46.

93. Benelli D, Londei P: **Begin at the beginning: evolution of translational initiation.** *Res Microbiol* 2009, **160**:493–501.

94. Hering O, Brenneis M, Beer J, Suess B, Soppa J: **A novel mechanism for translation initiation operates in haloarchaea.** *Mol Microbiol* 2009, **71**:1451–63.

95. Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R: **A single-base resolution map of an archaeal transcriptome.** *Genome Res* 2010, **20**:133–41.

96. Hopfield JJ: **Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity.** *Proc Natl Acad Sci U S A* 1974, **71**:4135–9.

97. Ninio J: **Kinetic amplification of enzyme discrimination.** *Biochimie* 1975, **57**:587–95.

98. Rodnina M V, Gromadski KB, Kothe U, Wieden H-J: **Recognition and selection of tRNA in translation.** *FEBS Lett* 2005, **579**:938–42.

99. Thompson RC, Stone PJ: **Proofreading of the codon-anticodon interaction on ribosomes.** *Proc Natl Acad Sci U S A* 1977, **74**:198–202.

100. Zaher HS, Green R: **Quality control by the ribosome following peptide bond formation.** *Nature* 2009, **457**:161–6.

101. Alkalaeva E, Eliseev B, Ambrogelly A, Vlasov P, Kondrashov FA, Gundllapalli S, Frolova L, Söll D, Kisselev L: **Translation termination in pyrrolysine-utilizing archaea.** *FEBS Lett* 2009, **583**:3455–60.

102. Lee MM, Jiang R, Jain R, Larue RC, Krzycki J, Chan MK: **Structure of Desulfitobacterium hafniense PylSc, a pyrrolysyl-tRNA synthetase.** *Biochem Biophys Res Commun* 2008, **374**:470–4.

103. Théobald-Dietrich A, Frugier M, Giegé R, Rudinger-Thirion J: **Atypical archaeal tRNA pyrrolysine transcript behaves towards EF-Tu as a typical elongator tRNA.** *Nucleic Acids Res* 2004, **32**:1091–6.

104. Ito K, Frolova L, Seit-Nebi A, Karamyshev A, Kisselev L, Nakamura Y: **Omnipotent decoding potential resides in eukaryotic translation termination factor eRF1 of variant-code organisms and is modulated by the interactions of amino acid sequences within domain 1.** *Proc Natl Acad Sci U S A* 2002, **99**:8494–9.

105. Kervestin S, Frolova L, Kisselev L, Jean-Jean O: **Stop codon recognition in ciliates: Euplotes release factor does not respond to reassigned UGA codon.** *EMBO Rep* 2001, **2**:680–4.

106. Stock T, Selzer M, Rother M: **In vivo requirement of selenophosphate for selenoprotein synthesis in archaea.** *Mol Microbiol* 2010, **75**:149–60.

107. Leibundgut M, Frick C, Thanbichler M, Böck A, Ban N: **Selenocysteine tRNA-specific elongation factor SelB is a structural chimaera of elongation and initiation factors.** *EMBO J* 2005, **24**:11–22.

108. Atkinson GC, Baldauf SL, Hauryliuk V: **Evolution of nonstop, no-go and nonsense-mediated mRNA decay and their termination factor-derived components.** *BMC Evol Biol* 2008, **8**:290.

109. Doma MK, Parker R: **Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation.** *Nature* 2006, **440**:561–4.

110. Graille M, Chaillet M, van Tilbeurgh H: **Structure of yeast Dom34: a protein related to translation termination factor Erf1 and involved in No-Go decay.** *J Biol Chem* 2008, **283**:7145–54.

111. Moore SD, Sauer RT: **The tmRNA system for translational surveillance and ribosome rescue.** *Annu Rev Biochem* 2007, **76**:101–24.

112. Benaroudj N, Zwickl P, Seemüller E, Baumeister W, Goldberg AL: **ATP hydrolysis by the proteasome regulatory complex PAN serves multiple functions in protein degradation.** *Mol Cell* 2003, **11**:69–78.

113. Hartung S, Hopfner K-P: **Lessons from structural and biochemical studies on the archaeal exosome.** *Biochem Soc Trans* 2009, **37**(Pt 1):83–7.

114. Lorentzen E, Basquin J, Conti E: **Structural organization of the RNA-degrading exosome.** *Curr Opin Struct Biol* 2008, **18**:709–13.

115. Slomovic S, Portnoy V, Yehudai-Resheff S, Bronshtein E, Schuster G: **Polynucleotide phosphorylase and the archaeal exosome as poly(A)-polymerases.** *Biochim Biophys Acta* 2008, **1779**:247–55.

116. Celesnik H, Deana A, Belasco JG: **Initiation of RNA decay in Escherichia coli by 5' pyrophosphate removal.** *Mol Cell* 2007, **27**:79–90.

117. Portnoy V, Evguenieva-Hackenberg E, Klein F, Walter P, Lorentzen E, Klug G, Schuster G: **RNA polyadenylation in Archaea: not observed in Haloferax while the exosome polynucleotidylates RNA in Sulfolobus.** *EMBO Rep* 2005, **6**:1188–93.

118. Portnoy V, Schuster G: **RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R.** *Nucleic Acids Res* 2006, **34**:5923–31.

119. Slomovic S, Laufer D, Geiger D, Schuster G: **Polyadenylation of ribosomal RNA in human cells.** *Nucleic Acids Res* 2006, **34**:2966–75.

120. Newbury SF: **Control of mRNA stability in eukaryotes.** *Biochem Soc Trans* 2006, **34**(Pt 1):30–4.

121. Hasenöhrl D, Lombo T, Kaberdin V, Londei P, Bläsi U: **Translation initiation factor a/eIF2(-gamma) counteracts 5' to 3' mRNA decay in the archaeon Sulfolobus solfataricus.** *Proc Natl Acad Sci U S A* 2008, **105**:2146–50.

122. French SL, Santangelo TJ, Beyer AL, Reeve JN: **Transcription and translation are coupled in Archaea.** *Mol Biol Evol* 2007, **24**:893–5.

123. Humbard MA, Zhou G, Maupin-Furlow JA: **The N-terminal penultimate residue of 20S proteasome alpha1 influences its N(alpha) acetylation and protein levels as well as growth rate and stress responses of Haloferax volcanii.** *J Bacteriol* 2009, **191**:3794–803.

124. Maupin-Furlow JA, Humbard MA, Kirkland PA, Li W, Reuter CJ, Wright AJ, Zhou G: **Proteasomes from structure to function: perspectives from Archaea.** *Curr Top Dev Biol* 2006, **75**:125–69.

125. Hochstrasser M: **Origin and function of ubiquitin-like proteins.** *Nature* 2009, **458**:422–9.

126. Kubota H: **Quality control against misfolded proteins in the cytosol: a network for cell survival.** *J Biochem* 2009, **146**:609–16.

127. Hirsch C, Gauss R, Horn SC, Neuber O, Sommer T: **The ubiquitylation machinery of the endoplasmic reticulum.** *Nature* 2009, **458**:453–60.

128. Humbard MA, Miranda H V, Lim J-M, Krause DJ, Pritz JR, Zhou G, Chen S, Wells L, Maupin-Furlow JA: **Ubiquitin-like small archaeal modifier proteins (SAMPs) in Haloferax volcanii.** *Nature* 2010, **463**:54–60.

3 Role of multiprotein bridging factor 1 in archaea: bridging the domains?

Bart de Koning, Fabian Blombach, Hao Wu, Stan J.J. Brouns and John van der Oost

Abstract

MBF1 (multiprotein bridging factor 1) is a highly conserved protein in archaea and eukaryotes. It was originally identified as a mediator of the eukaryotic transcription regulator BmFTZ-F1 (*Bombyx mori* regulator of *fushi tarazu*). MBF1 was demonstrated to enhance transcription by forming a bridge between distinct regulatory DNA-binding proteins and the TATA-box-binding protein. MBF1 consists of two parts: a C-terminal part that contains a highly conserved helix-turn-helix, and an N-terminal part that shows a clear divergence: in eukaryotes, it is a weakly conserved flexible domain, whereas, in archaea, it is a conserved zinc-ribbon domain. Although its function in archaea remains elusive, its function as a transcriptional co-activator has been deduced from thorough studies of several eukaryotic proteins, often indicating a role in stress response. In addition, MBF1 was found to influence translation fidelity in yeast. Genome context analysis of *mbf1* in archaea revealed conserved clustering in the crenarchaeal branch together with genes generally involved in gene expression. It points to a role of MBF1 in transcription and/or translation. Experimental data are required to allow comparison of the archaeal MBF1 with its eukaryotic counterpart.

Introduction

Because most information-processing pathways in eukaryotes are highly similar to that of archaea, it is tempting to use the generally less complex archaeal systems as models. The archaeal transcription machinery, for example, is highly similar to the eukaryotic RNAP (RNA polymerase) II machinery. The core subunits of the RNA polymerase are present in all archaea, together with orthologs of some general transcription factors: TBP (TATA-box-binding protein), TF (transcription factor) IIB (TFB in archaea), and the α-subunit of TFIIE (TFE in archaea) [1]. In contrast, the majority of the known eukaryotic transcription factors are absent. A protein that is reported to be present in almost all archaea and known as a transcriptional co-activator in eukaryotes is MBF1 (multiprotein bridging factor 1) [2]. The function of this protein in archaea remains poorly understood. However, its function as a transcriptional co-activator in eukaryotes has been described in numerous studies since it was discovered in *Bombyx mori* [3]. Interestingly, MBF1 has also been reported to influence fidelity of translation in yeast [4]. In the present paper, we review the current knowledge of MBF1 and its possible role(s) in archaea: is MBF1 involved in transcription or in translation, or does it have a dual function?

Phylogenetic distribution

MBF1 was first reported in 1994 in *B. mori* as a mediator of BmFTZ-F1, the equivalent of the FTZ-F1 regulator of the *Drosophila fushi tarazu* (*ftz*) gene that is associated with body segmentation. It was demonstrated that a heterodimer of MBF1 and MBF2 (which is not related to MBF1 and is found only in moths), could enhance transcription of a FTZ-F1-dependent transcript by bridging between BmFTZ-F1 and TBP [3, 5–8]. Early genomic comparisons revealed that a gene coding for MBF1 is present in all eukaryotic genomes [2, 6, 9–11]. In humans (h), MBF1 is present in two isoforms: hMBF1α and hMBF1β, which originates from alternative splicing. The α mRNA was found in all the tissues tested, whereas the β form was only detected in pancreas and HeLa cells. The β form contains a very acidic tail, possibly an adaptation to the pancreatic environment [12]. The plant *Arabidopsis thaliana* (At) has three distinct copies in its genome, encoding AtMBF1a, AtMBF1b, and AtMBF1c. All three paralogues appear to have distinct biological roles [13]. These distinct MBF1 proteins and MBF1 proteins of other plants can be divided into two subgroups on the basis of their primary structure and different exon/intron patterns [14].

As to the distribution of MBF1 among prokaryotes, none of the sequenced bacterial genomes contains an orthologue of *mbf1*, whereas almost all archaea are equipped with one. This seems a logical distribution for a transcriptional co-activator: eukaryotes and the archaea have a high similarity in their transcription initiation complex, whereas the bacterial system is very different. Most archaeal genomes contain a single copy of the *mbf1* gene; however, all of the sequenced genomes of the Halobacteriales contain a second copy that lacks the C-terminal amino acids (Figure 1).


Abbreviations used: MBF1, Multiprotein Bridging Factor 1; EDF-1, Endothelial cell Differentiation Factor 1; TBP, TATA-box Binding Protein; TFB, Transcription Factor B; TFE, Transcription Factor E; RNAPII, RNA Polymerase II machinery; FTZ-F1, regulator of *fushi tarazu*; HTH, Helix-Turn-Helix; bZIP, basic-leucine zippers.

Interestingly, the archaeal genomes that do not possess an *mbf1* gene are the only two species that are completely sequenced within the recently proposed new phylum of the Thaumarchaeota [15]: *Cenarchaeum symbiosum* and *Nitrosopumilus maritimus*. Apparently, the gene is lost in this archaeal branch (Figure 2).

Sequence and structure

In a comparative genomics analysis of HTH (helix-turn-helix) domains, which are generally responsible for protein-DNA interactions, Aravind and Koonin [2] found that most of the predicted HTH-transcriptional regulators found in archaea appeared to be of the bacterial type. This came as a surprise because of the aforementioned relationship between the archaeal transcription initiation system and the eukaryotic RNAPII. It was reported that the only conserved classical HTH domain that is vertically inherited in almost all archaea and all eukaryotes was MBF1. NMR studies of hMBF1 and BmMBF1 indeed revealed a well structured C-terminal HTH core. It contains four α-helices and a conserved tail, proposed to be responsible for maintaining domain stability [16, 17]. Binding of MBF1 to DNA-binding regulators occurs via the central region, residues 35-113 in *B. mori*, that also contains the HTH-core [6]. Binding of MBF1 to TBP in yeast (y) requires an aspartate residue at position 112 (Asp^{112}) of yMBF1, located in the third helix of the C-terminal part, and a glutamine at position 68 (Gln^{68}) of yTBP, located on top of the saddle-shaped molecule. In eukaryotic and archaeal species, several combinations of amino acids are present on the corresponding locations. Altered yMBF1 and yTBP proteins with combinations that exist in Nature show an *in vitro* interaction, whereas unnatural combinations (e.g. Gln^{68}/Lys^{112} and Glu^{68}/Asp^{112}) do not, suggesting co-evolution of MBF1 and TBP [18]. This holds true for the archaeal combinations, although experimental evidence for an interaction between archaeal TBP and MBF1 is currently lacking.

In rats, MBF1 interacts with calmodulin [9]. This is in agreement with the presence in mammalian MBF1 of an IQ motif that interacts with calmodulin in the cytoplasm [19, 20]. Binding is increased by calcium and appears to be a function that is distinct from its
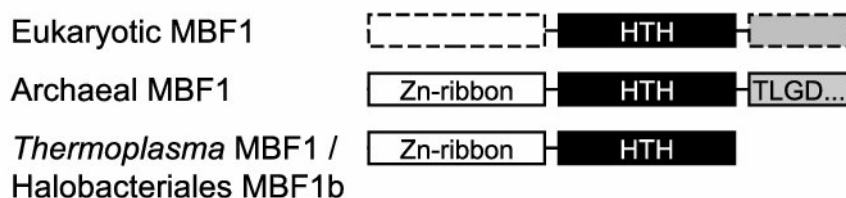
Figure 1. Schematic overview of conserved domains in eukaryotic and archaeal MBF1. White box with broken line: N-terminal domain of eukaryotic MBF1 (IPR013729); white box: zinc ribbon of archaeal MBF1; black box: four-helical HTH domain conserved in all MBF1s (IPR001387); grey box: conserved C-terminus of archaeal MBF1 harbouring a conserved motif (T/SL/MGD/E), including MBF1a of the Halobacteriales, but lacking in the paralogue MBF1b found in Halobacteriales (NP2072A, VNG1483C, rrnAC0872 and HQ2874A) and in MBF1 from *Thermoplama* spp. (Ta0948 and TV1117) ; grey box with broken line: C-terminus of eukaryotic MBF1, which habours conserved sequence motifs as well, and is partly lacking in plant MBF1. Locus tags and InterPro assignments (http://www.ebi.ac.uk/interpro/) are noted within parentheses.

involvement in transcription activation, as impairing one does not impair the other [21]. Both calmodulin and the IQ motif are absent from archaea. A phosphorylation site is located in close proximity of the IQ motif. Phosphorylation of MBF1 disrupts calmodulin binding and leads to translocation to the nucleus [19, 20]. In mammals, phosphorylation also appears to be necessary for MBF1-potentiated expression *in vivo* during cardiac hypertrophy, which is a stress response to increased workload [21]. Phosphorylation has also been reported for potato MBF1. It could be increased by elicitors derived from *Phytophthora infestans* and is inhibited by calcium chelators and calmodulin antagonists, suggesting that phosphorylation of MBF1 is carried out by a calcium-dependent serine/threonine kinase [22]. Unlike its mammalian counterpart, plant MBF1 appears to be localized predominantly in the nucleus [23].

In contrast with the high homology found in the C-terminal part of MBF1, the N-terminal part shows a clear divergence between eukaryotes and archaea. In eukaryotes, it has low conservation, and NMR studies indicate a relatively high flexibility [16, 17]. In archaea, it contains a conserved zinc-ribbon motif, predicted on the basis of its two pairs of cysteine residues. The presence of this zinc ribbon in archaea and its absence in eukaryotes could indicate distinct functions. Whereas, in eukaryotes, MBF1 binding to DNA by itself was not observed, the presence of a zinc ribbon might indicate that archaeal MBF1 functions as a single activator in archaea [18]. However, at present, no experimental support for this proposal is available.

Transcription activation

After its first description, MBF1 has been studied in several eukaryotic organisms. It does not affect histone modification, but rather acts as a bridging factor between a DNA-binding protein and the entire TFIID complex [TBP and TAFs (TBP-associated factors)] [3, 5, 6, 12, 17, 19, 20, 24–27]. DNA-binding proteins that have been reported to interact with MBF1 belong to two protein families: the family of steroid/nuclear hormone receptors, and the family of the bZIPs (basic leucine zippers). Established interactions between MBF1 and members of the former family are FTZ-F1, Ad4BP (adrenal 4-binding protein; a human orthologue of FTZ-F1), and liver receptor homologue 1, liver X receptor α and PPARγ (peroxisome-proliferator-activated receptor γ), non-steroid nuclear receptors, involved in lipid metabolism in humans [3, 5–7, 25]. Known interaction partners belonging to the bZIP family include GCN4 (general control non-derepressible 4), CREB (cAMP-response-element-binding protein) and CREBP1, ATF (activating transcription factor) 1, Jun and Fos, TDF (tracheae defective factor) and HaHB4 (*Helianthus annuus* homeobox-leucine zipper protein 4) [12, 21, 24, 25, 28, 29]. Jun, Fos and ATF family members dimerize into homo- or hetero-dimers, all called AP-1 (activator protein 1). Different AP-1 dimers have different physiological effects. It has been reported, for example, that, for activation of the atrial natriuretic peptide, MBF1 interacts with c-Jun, but not with c-Fos or ATF, suggesting that MBF1 may be able to discriminate between highly similar proteins. Interestingly, MBF1 is not always essential, as other promoters can be activated by c-Jun in the absence of MBF1 [21, 30].

In yeast, MBF1 is essential for GCN4-dependent activation of *HIS3,* encoding an imidazoleglycerol-phosphate dehydratase that is responsible for a step in histidine biosynthesis. Without activation, a basic level of constitutive expression of this gene still allows growth. However disruption of *mbf1* generates a phenotype that is sensitive to 3-aminotriazole, an

inhibitor of HIS3 enzyme [24]. In human endothelial cell cultures, MBF1 represses endothelial cell differentiation. Therefore hMBF1 is also called EDF-1 (endothelial cell differentiation factor1) [10, 31]. In *Drosophila*, *mbf1*-null mutants are viable, but show severe tracheal and central nervous system defects. However, no segmentation defects similar to *ftz-f1* mutants and no decrease in expression of a FTZ-F1-dependent reporter gene were observed, indicating that MBF1 is not a crucial co-activator for FTZ-F1-dependent transcription *in vivo* in fruitflies [28]. These reports indicate involvement of MBF1 in different physiological processes in phylogenetically distinct eukaryotic organisms. Although it is completely conserved in eukaryotes, *mbf1* knockouts appear viable under tested conditions.

Translation fidelity

Unexpectedly, *mbf1* was found to be synonymous with the frameshift-suppression gene *suf13* in yeast [4]. The suf13-1 mutant strains as well as a directed *suf13* (*mbf1*)-gene-knockout strain exhibit an increased rate of ribosomal +1 frameshifting for several different reporter gene constructs harbouring frameshift mutations [4, 32, 33]. It remains elusive how a transcription co-activator influences translation fidelity. The authors speculate about a role of MBF1 in RNA polymerase III transcription of tRNA genes [4]. Decreased levels of tRNA would increase translational pausing and thereby the probability of frameshift events. TBP is also part of the RNAPIII general transcription factor TFIIIB. Similarly to its role in RNAPII transcription activation, MBF1 could therefore interact with the RNAPIII transcription machinery via TBP, but, up to now, no experimental evidence links MBF1 to RNAPIII transcription initiation.

An alternative scenario would be an involvement of MBF1 in the biogenesis of ribosomes and tRNA, possibly indirectly by transcription regulation of factors contributing to these pathways. In addition, it cannot be ruled out that MBF1 interacts directly with the ribosome during translation as it is also present in the cytoplasm [23, 28, 34, 35]. Both translation and tRNA gene transcription are central parts of information processing. A function of MBF1 in these processes would explain the remarkable evolutionary conservation of MBF1 in eukaryotes and archaea.

Stress response

In vivo analyses in plants indicate a link between MBF1 and stress response, including the ethylene response. MBF1s from *Arabidopsis,* potato, tomato, *Retama raetam* (white weeping broom) and tobacco have been reported to be up-regulated upon biotic and abiotic stresses, such as pathogen infection, drought, salinity, application of ethylene and, especially, heat and oxidative stress. AtMBf1c also reacts on application of abscisic acid and salicylic acid. It controls the heat-stress-response network. AtMBF1a is linked to salt stress, and is also, together with AtMBF1b, differentially regulated during plant development [11, 14, 22, 23, 27, 35–40]. Overexpression in plants leads to a higher tolerance to a number of stress factors, without suppressing plant growth. It also leads to higher levels of trehalose, which is generally believed to be important in stress signaling in plants [35, 40, 41]. In *Drosophila*, it reduces sensitivity to oxidative stress by protecting the transcriptional activator d-Jun against oxidation [34]. The latter effect is found in mammals too, where MBF1 appears to be a crucial factor in the response against oxidative stress [30]. During hyperthrophy, enhanced *mbf1* gene expression leads to

elevated protein levels in mammalian cardiomyocyte cultures, and, in turn, results in activation of hyperthrophic gene expression [21]. Interestingly, MBF1 appears also to be a target for viral abuse: MBF1 of tobacco (NtMBF1a) and *Arabidopsis* (AtMBF1a and AtMBF1b) have been reported to interact with two viral Movement Proteins, responsible for cell-to-cell movement [42].

In eukaryotes MBF1 plays an important role in stress response. This might also be true for the archaeal MBF1, although *mbf1* is not up-regulated during oxidative stress (B. Wiedenheft, M. Young and J. van der Oost, unpublished work).

Genomic context

In prokaryotes, genomic context analysis is a powerful tool in the functional prediction of genes [43]. Especially in crenarchaeotes, the gene context of *mbf1* shows a high degree of conservation encompassing several genes (with predicted function): *pan* (proteasome-activating nucleotidase [44]), *hflX* (G–protein of the HflX family [45]), *tgt* (tRNA-guanine transglycosylase [46]), *tfe* and *tfb* (coding for archaeal homologues of transcription initiation factors E and B respectively) and *rpoG* (hypothetical RNA-polymerase subunit G [47]) (Fig. 2). The proteasome-activating nucleotidase is homologous with the ATPases of the 19S particle of the eukaryotic proteasome and regulates the entry of folded proteins into the proteolytic core particle of the proteasome. The HflX family of G-proteins is largely uncharacterized, but a function linked to the translation machinery has been predicted [48, 49]. In euryarchaeotes, the genomic context of *mbf1* is less conserved; however, proximity to *pan* occurs regularly. Apart from this, the genomic context is only conserved at higher taxonomic levels. Interestingly, although *mbf1* is absent from Thaumarchaeota, a part of the chromosome shows similarities to the genomic context of *mbf1* in *Thermophilum pendens*, except that *mbf1* has been replaced by a gene encoding a putative phosphate-uptake regulator.

It is not possible to give a clear functional prediction based on the observed context in crenarchaeotes. However, the conserved clustering (and probably the co-regulation) with genes involved in gene expression at different levels point to a basal function of MBF1 in gene expression as well, most likely at the level of translation and/or transcription.

Figure 2. Genome context analysis of mbf1 in archaea. In *Sulfolobus solfataricus*, *mbf1* resides in a large gene cluster that is conserved in distantly related Crenarchaea. COG classification is provided. (1) COG0459 (thermosome β subunit); (2) COG1405 (transcription initiation factor B, TFB); (3) COG1676 (tRNA-splicing endonuclease); (4) no COG assigned (RNA polymerase subunit G, RpoG); (5) COG1958 [small nuclear RNP (RNA-binding protein)]; (6) COG0343 (tRNA-ribosyltransferase, TGT); (7) COG1938 [DUF75; ATP-GRASP (Golgi reassembly stacking protein) superfamily; ligase (C/N, C/S-CoA)]; (8) COG1370 [PUA domain, RNA binding; in many Euryarchaea fused to TGT (COG0343)]; (9) COG1222 [AAA + (ATPase associated with various cellular activities)-type ATPase, *pan*); (10) COG1813 (MBF1); (11) COG2262 (HflX-family GTPase); (12) COG1303 (DUF127); (13) COG1675 (transcription initiation factor E, TFE); (14) COG0270 (site-specific DNA methylase); (15) COG4080 (RecB-family nuclease); (16) COG0456 (acetyltransferase); (17) COG1628 (DUF99); (18) COG1308 [nascent polypeptide-associated complex α subunit; BTF3 (basal transcription factor 3)]; and (19) COG1844 (DUF356). Dotted arrows indicate the presence of a gene with another classification. Slashes indicate a separation of more than two other genes. The classification and order of the species is based upon trees by Brochier-Armanet et al. [15] and Elkins et al. [50].

Functional prediction for archaeal MBF1

MBF1 is a conserved protein present in all eukaryotes and archaea, except the two characterized members of the Thaumarchaeota. The protein consists of an N-terminal part that differs between eukaryotes and archaea, and a very well conserved C-terminal part. The observation that the N-terminal domain of human MBF1α is not required for binding to TBP [12] suggests that this interaction can also be found in archaea. Such a role would fit with the absence of MBF1 in the bacterial domain where the transcription machinery is very different, and with the aforementioned indications of MBF1/TBP co-evolution [18]. However the divergence in the N-terminal part may point to a different mode of action, possibly a direct interaction with DNA. This is supported by the absence of bZIP proteins in the archaeal domain. An interesting other possibility is the finding that yMBF1 (directly or indirectly) influences translation fidelity.

The genomic context is certainly interesting, but unfortunately not conclusive either. The strongest co-occurence of archaeal *mbf1* is with *pan*, the gene encoding a protein associated with protein degradation. Other conserved genes, mainly found in crenarchaeotes, encode proteins in one way or another associated with transcription or translation. Potential experimental approaches to gain insight into the physiological function of the archaeal MBF1 include a broad spectrum of complementary biochemical, genetic and genomic studies (*in vivo* and *in vitro*): comparison of wild-type, *mbf1* knockout, and MBF1 overproduction strains at the level of transcription and translation. In the absence of such experimental data, however, comparison of the archaeal MBF1 with its eukaryotic counterpart still seems to be a bridge too far.

Acknowledgements

References

1. Bell SD: **Archaeal transcriptional regulation-variation on a bacterial theme?** *Trends Microbiol* 2005, **13**:262–5.

2. Aravind L, Koonin E V: **DNA-binding proteins and evolution of transcription regulation in the archaea.** *Nucleic Acids Res* 1999, **27**:4658–70.

3. Li FQ, Ueda H, Hirose S: **Mediators of activation of fushi tarazu gene transcription by BmFTZ-F1.** *Mol Cell Biol* 1994, **14**:3013–21.

4. Hendrick JL, Wilson PG, Edelman II, Sandbaken MG, Ursic D, Culbertson MR: **Yeast frameshift suppressor mutations in the genes coding for transcription factor Mbf1p and ribosomal protein S3: evidence for autoregulation of S3 synthesis.** *Genetics* 2001, **157**:1141–58.

5. Li FQ, Takemaru K, Goto M, Ueda H, Handa H, Hirose S: **Transcriptional activation through interaction of MBF2 with TFIIA.** *Genes Cells* 1997, **2**:143–53.

6. Takemaru K i, Li FQ, Ueda H, Hirose S: **Multiprotein bridging factor 1 (MBF1) is an evolutionarily conserved transcriptional coactivator that connects a regulatory factor and TATA element-binding protein.** *Proc Natl Acad Sci U S A* 1997, **94**:7251–6.

7. Liu QX, Ueda H, Hirose S: **Comparison of sequences of a transcriptional coactivator MBF2 from three Lepidopteran species Bombyx mori, Bombyx mandarina and Samia cynthia.** *Gene* 1998, **220**:55–9.

8. Liu QX, Ueda H, Hirose S: **MBF2 is a tissue- and stage-specific coactivator that Is regulated at the step of nuclear transport in the silkworm Bombyx mori.** *Dev Biol* 2000, **225**:437–46.

9. Smith ML, Johanson RA, Rogers KE, Coleman PD, Slemmon JR: **Identification of a neuronal calmodulin-binding peptide, CAP-19, containing an IQ motif.** *Brain Res Mol Brain Res* 1998, **62**:12–24.

10. Dragoni I, Mariotti M, Consalez GG, Soria MR, Maier JA: **EDF-1, a novel gene product down-regulated in human endothelial cell differentiation.** *J Biol Chem* 1998, **273**:31119–24.

11. Zegzouti H, Jones B, Frasse P, Marty C, Maitre B, Latche A, Pech J-C, Bouzayen M: **Ethylene-regulated gene expression in tomato fruit: characterization of novel ethylene-responsive and ripening-related genes isolated by differential display**. *Plant J* 1999, **18**:589–600.

12. Kabe Y, Goto M, Shima D, Imai T, Wada T, Morohashi K i, Shirakawa M, Hirose S, Handa H: **The role of human MBF1 as a transcriptional coactivator.** *J Biol Chem* 1999, **274**:34196–202.

13. Tsuda K, Tsuji T, Hirose S, Yamazaki K: **Three Arabidopsis MBF1 homologs with distinct expression profiles play roles as transcriptional co-activators.** *Plant Cell Physiol* 2004, **45**:225–31.

14. Tsuda K, Yamazaki K-I: **Structure and expression analysis of three subtypes of Arabidopsis MBF1 genes.** *Biochim Biophys Acta* 2004, **1680**:1–10.

15. Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P: **Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota.** *Nat Rev Microbiol* 2008, **6**:245–52.

16. Mishima M, Ozaki J, Ikegami T, Kabe Y, Goto M, Ueda H, Hirose S, Handa H, Shirakawa M: **Resonance assignments, secondary structure and 15N relaxation data of the human transcriptional coactivator hMBF1 (57-148).** *J Biomol NMR* 1999, **14**:373–6.

17. Ozaki J, Takemaru KI, Ikegami T, Mishima M, Ueda H, Hirose S, Kabe Y, Handa H, Shirakawa M: **Identification of the core domain and the secondary structure of the transcriptional coactivator MBF1.** *Genes Cells* 1999, **4**:415–24.

18. Liu Q-X, Nakashima-Kamimura N, Ikeo K, Hirose S, Gojobori T: **Compensatory change of interacting amino acids in the coevolution of transcriptional coactivator MBF1 and TATA-box-binding protein.** *Mol Biol Evol* 2007, **24**:1458–63.

19. Mariotti M, De Benedictis L, Avon E, Maier JA: **Interaction between endothelial differentiation-related factor-1 and calmodulin in vitro and in vivo.** *J Biol Chem* 2000, **275**:24047–51.

20. Ballabio E, Mariotti M, De Benedictis L, Maier JAM: **The dual role of endothelial differentiation-related factor-1 in the cytosol and nucleus: modulation by protein kinase A.** *Cell Mol Life Sci* 2004, **61**:1069–74.

21. Busk PK, Wulf-Andersen L, Strøm CC, Enevoldsen M, Thirstrup K, Haunsø S, Sheikh SP: **Multiprotein bridging factor 1 cooperates with c-Jun and is necessary for cardiac hypertrophy in vitro.** *Exp Cell Res* 2003, **286**:102–14.

22. Zanetti ME, Blanco FA, Daleo GR, Casalongué CA: **Phosphorylation of a member of the MBF1 transcriptional co-activator family, StMBF1, is stimulated in potato cell suspensions upon fungal elicitor challenge.** *J Exp Bot* 2003, **54**:623–32.

23. Sugikawa Y, Ebihara S, Tsuda K, Niwa Y, Yamazaki K-I: **Transcriptional coactivator MBF1s from Arabidopsis predominantly localize in nucleolus.** *J Plant Res* 2005, **118**:431–7.

24. Takemaru K, Harashima S, Ueda H, Hirose S: **Yeast coactivator MBF1 mediates GCN4-dependent transcriptional activation.** *Mol Cell Biol* 1998, **18**:4971–6.

25. Brendel C, Gelman L, Auwerx J: **Multiprotein bridging factor-1 (MBF-1) is a cofactor for nuclear receptors that regulate lipid metabolism.** *Mol Endocrinol* 2002, **16**:1367–77.

26. Millership JJ, Waghela P, Cai X, Cockerham A, Zhu G: **Differential expression and interaction of transcription co-activator MBF1 with TATA-binding protein (TBP) in the apicomplexan Cryptosporidium parvum.** *Microbiology* 2004, **150**(Pt 5):1207–13.

27. Arce DP, Tonón C, Zanetti ME, Godoy AV, Hirose S, Casalongué CA: **The potato transcriptional co-activator StMBF1 is up-regulated in response to oxidative stress and interacts with the TATA-box binding protein.** *J Biochem Mol Biol* 2006, **39**:355–60.

28. Liu Q-X, Jindra M, Ueda H, Hiromi Y, Hirose S: **Drosophila MBF1 is a co-activator for Tracheae Defective and contributes to the formation of tracheal and nervous systems.** *Development* 2003, **130**:719–28.

29. Zanetti ME, Chan RL, Godoy A V, González DH, Casalongué CA: **Homeodomain-leucine zipper proteins interact with a plant homologue of the transcriptional co-activator multiprotein bridging factor 1.** *J Biochem Mol Biol* 2004, **37**:320–4.

30. Miotto B, Struhl K: **Differential gene regulation by selective association of transcriptional coactivators and bZIP DNA-binding domains.** *Mol Cell Biol* 2006, **26**:5969–82.

31. De Benedictis L, Mariotti M, Dragoni I, Maier JA: **Cloning and characterization of murine EDF-1.** *Gene* 2001, **275**:299–304.

32. Costanzo MC, Mueller PP, Strick CA, Fox TD: **Primary structure of wild-type and mutant alleles of the PET494 gene of Saccharomyces cerevisiae.** *Mol Gen Genet* 1986, **202**:294–301.

33. Culbertson MR, Gaber RF, Cummins CM: **Frameshift suppression in Saccharomyces cerevisiae. V. Isolation and genetic properties of nongroup-specific suppressors.** *Genetics* 1982, **102**:361–78.

34. Jindra M, Gaziova I, Uhlirova M, Okabe M, Hiromi Y, Hirose S: **Coactivator MBF1 preserves the redox-dependent AP-1 activity during oxidative stress in Drosophila.** *EMBO J* 2004, **23**:3538–47.

35. Suzuki N, Bajad S, Shuman J, Shulaev V, Mittler R: **The transcriptional co-activator MBF1c is a key regulator of thermotolerance in Arabidopsis thaliana.** *J Biol Chem* 2008, **283**:9269–75.

36. Godoy A V., Zanetti ME, San Segundo B, Casalongué CA: **Identification of a putative Solanum tuberosum transcriptional coactivator up-regulated in potato tubers by Fusarium solani f. sp. eumartii infection and wounding.** *Physiol Plant* 2001, **112**:217–222.

37. Pnueli L, Hallak-Herr E, Rozenberg M, Cohen M, Goloubinoff P, Kaplan A, Mittler R: **Molecular and biochemical mechanisms associated with dormancy and drought tolerance in the desert legume Retama raetam.** *Plant J* 2002, **31**:319–30.

38. Rizhsky L, Liang H, Mittler R: **The combined effect of drought stress and heat shock on gene expression in tobacco.** *Plant Physiol* 2002, **130**:1143–51.

39. Rizhsky L, Liang H, Shuman J, Shulaev V, Davletova S, Mittler R: **When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress.** *Plant Physiol* 2004, **134**:1683–96.

40. Kim M-J, Lim G-H, Kim E-S, Ko C-B, Yang K-Y, Jeong J-A, Lee M-C, Kim CS: **Abiotic and biotic stress tolerance in Arabidopsis overexpressing the multiprotein bridging factor 1a (MBF1a) transcriptional coactivator gene.** *Biochem Biophys Res Commun* 2007, **354**:440–6.

41. Suzuki N, Rizhsky L, Liang H, Shuman J, Shulaev V, Mittler R: **Enhanced tolerance to environmental stress in transgenic plants expressing the transcriptional coactivator multiprotein bridging factor 1c.** *Plant Physiol* 2005, **139**:1313–22.

42. Matsushita Y, Miyakawa O, Deguchi M, Nishiguchi M, Nyunoya H: **Cloning of a tobacco cDNA coding for a putative transcriptional coactivator MBF1 that interacts with the tomato mosaic virus movement protein.** *J Exp Bot* 2002, **53**:1531–2.

43. Ettema TJG, de Vos WM, van der Oost J: **Discovering novel biology by in silico archaeology.** *Nat Rev Microbiol* 2005, **3**:859–69.

44. Medalia N, Sharon M, Martinez-Arias R, Mihalache O, Robinson C V, Medalia O, Zwickl P: **Functional and structural characterization of the Methanosarcina mazei proteasome and PAN complexes.** *J Struct Biol* 2006, **156**:84–92.

45. Wu H, Sun L, Brouns SJJ, Fu S, Akerboom J, Li X, van der Oost J: **Purification, crystallization and preliminary crystallographic analysis of a GTP-binding protein from the hyperthermophilic archaeon Sulfolobus solfataricus.** *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2007, **63**(Pt 3):239–41.

46. Sabina J, Söll D: **The RNA-binding PUA domain of archaeal tRNA-guanine transglycosylase is not required for archaeosine formation**. *J Biol Chem* 2006, **281**:6993–7001.

47. Koonin E V, Makarova KS, Elkins JG: **Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerases are conserved in hyperthermophilic Crenarchaeota and "Korarchaeota".** *Biol Direct* 2007, **2**:38.

48. Leipe DD, Wolf YI, Koonin E V, Aravind L: **Classification and evolution of P-loop GTPases and related ATPases.** *J Mol Biol* 2002, **317**:41–72.

49. Brown ED: **Conserved P-loop GTPases of unknown function in bacteria: an emerging and vital ensemble in bacterial physiology.** *Biochem Cell Biol* 2005, **83**:738–46.

50. Elkins JG, Podar M, Graham DE, Makarova KS, Wolf Y, Randau L, Hedlund BP, Brochier-Armanet C, Kunin V, Anderson I, Lapidus A, Goltsman E, Barry K, Koonin E V, Hugenholtz P, Kyrpides N, Wanner G, Richardson P, Keller M, Stetter KO: **A korarchaeal genome reveals insights into the evolution of the Archaea.** *Proc Natl Acad Sci U S A* 2008, **105**:8102–7.

Phenotypic analysis of MBF1 from the archaeaon *Sulfolobus solfataricus* reveals a regulatory function beyond the level of transcription

4

Bart de Koning, Fabian Blombach, Rie Matsumi, Sonja V. Albers, Pawel Sierocinski, Olga Pougovkina, Thijs J.G. Ettema, Harryuki Atomi, Martine Roovers, Stan J.J. Brouns, John van der Oost

Abstract

Impressive progress has recently been achieved at the level of structural and functional analyses of the eukaryotic-like protein complexes responsible for processing of genetic information in archaea (replication, transcription and translation). Understanding the regulation of these processes is less advanced, however it is anticipated that these key processes in archaea are controlled by highly conserved proteins. A candidate global regulator is Multi-protein Bridging Factor1 (MBF1). This protein is highly conserved in the archaeo-eukaryotic lineage. Eukaryotic MBF1 has been reported to have a role in transcription regulation, by establishing a physical bridge between certain transcriptional regulators and the RNA polymerase II transcription machinery. In addition, a genetic link of yeast MBF1 with translation fidelity has been reported. To investigate the possible role of the archaeal MBF1 orthologs in transcription and translation, we generated a gene disruption mutant in the crenarchaeon *Sulfolobus solfataricus*. While growth of the *S. solfataricus* Δ*mbf1* strain was indistinguishable from that of the parent strain under standard conditions, it was found that the Δ*mbf1* strain has a prolonged lag phase upon transfer from rich to minimal medium, and appears to be more sensitive to stationary phase stress, to sulphate limitation stress, as well as to the ribosome-targeting antibiotic paromomycin. These results suggest involvement in adaptation to changing growth conditions. Transcriptome profiles at standard growth conditions revealed only minor differences between Δ*mbf1* and its parental strain, which contradicts with its presumed role as a transcription regulator. This study provides evidence that archaeal MBF1 is likely to have a role in adaptation to various stresses like its eukaryotic counterpart, but in contrast it does not appear to function as a transcription regulator.

Introduction

The evolution of life has resulted in numerous regulatory systems to differentially gear cellular processes, in order to reach the highest possible metabolic efficiency. In general, such control mechanisms are important for regulation and integration of the entire metabolic network, but, in particular, this is true for the processing of genetic information: (i) replication to copy parental genomic DNA for offspring cells, (ii) transcription to express genes and generate corresponding messenger RNAs, and (iii) translation to convert the RNA message into proteins. Major progress has been accomplished in revealing the mechanistic details of these systems. Moreover, it became clear that the degree of conservation of macromolecules (proteins, rRNA) that are associated with information processing in the three domains of life - Bacteria, Archaea and Eukaryotes - is higher between the archaeal and eukaryotic domain than between the others [1–14].

During the past decade, archaeal information processing systems have been used as models to study the more complex eukaryotic counterparts [6, 8, 11, 14]. In contrast to the increasing knowledge gathered about the basic archaeal information processing machineries, the global regulation of these machineries is relatively poorly understood. Apart from bioinformatical evidence on global networks in archaea [15–17], little experimental data is available on how the different information processing systems are coordinated and integrated. In order to understand global regulation in archaea and to identify the factors involved therein, it is of particular interest to study those candidate regulators that show a high level of conservation across the archaeal phyla. Given the extended homology between archaeal and eukaryotic information processing machineries, it seems possible that regulators mediating between the different information processing machineries might have remained conserved in both archaea and eukaryotes.

The helix-turn-helix (HTH) motif is a ubiquitous motif found in many proteins functioning as transcription regulators or basal transcription factors [18]. In 1999 Aravind and Koonin described that the only classical HTH protein that is conserved in all archaea and eukaryotes, based on the complete genome sequences available at that time, was a small protein called Multi-protein Bridging Factor 1 (MBF1) [19]. More recent reanalyses of the phylogenetic distribution of MBF1 confirmed its strict conservation in the archaeo-eukaryotic branch, with the single exception of species belonging to the marine Thaumarchaeota (*Nitrosopumilus sp., Cenarchaeum sp., and Nitrosoarchaeum sp.*), who are devoid of the *mbf1* gene [20, 21]. The terrestrial Thaumarchaeota – at least *Nitrososphaera gargensis* and *Caldiarchaeum subterraneum* – as well as members belonging to the recently identified DPANN superfamily of archaea [22] seem to contain at least one copy of the *mbf1* gene. However because most genomes belonging to this latter superfamily are still incomplete, getting absolute answers about its distribution within the archaeal superfamilies is still difficult.

Several studies on eukaryotic MBF1 revealed a role in transcription regulation in a variety of organisms, ranging from yeast and fly to plant and man [23–43]. No direct interaction was found between MBF1 proteins and their target promoters, but it was found to act by bridging between a transcriptional regulator - which often belongs either to the basic region leucine

---

Abbreviations: MBF1, Multi-protein Bridging Factor 1; HTH, helix-turn-helix; PBL2025, *Sulfolobus solfataricus* PBL2025; Δ*mbf1*, *Sulfolobus solfataricus* Δ*mbf1*

zipper family or to the nuclear hormone receptor family - and the TATA-box binding protein of the transcription initiation complex [35, 36]. In addition, it has also been reported that MBF1 in yeast appears to have a function beyond transcription regulation: besides stimulation of the *HIS3* gene [36] it also influences translation fidelity at the level of reading frame maintenance [44, 45].

In contrast to the above mentioned studies on eukaryotic MBF1, limited experimental analysis of the archaeal counterpart protein has been published to date. Obviously a detailed molecular characterization is a prerequisite to understand why MBF1 remained conserved in nearly all archaeal and eukaryotic species despite the separation of billions of years of evolution. Homology between the archaeal and eukaryotic MBF1 orthologs is restricted to the core part of the C-terminal half of the protein, which covers the helix-turn-helix domain. The archaeal MBF1 contains a zinc-ribbon domain at the N-terminal side where eukaryotes appear to have a unique domain [20] (Chapter 3). In a recent study, archaeal MBF1 from both *Thermoproteus therax* and *Methanosarcina mazei* was unable to rescue a deletion mutant in yeast for growth in the presence of aminotriazole, even in chimeric constructs that contained both the eukaryotic N-terminal domain and the archaeal C-terminal domain [21].

To investigate the possible role of the archaeal MBF1 orthologs in transcription and translation, we generated a gene disruption mutant in the crenarchaeon *Sulfolobus solfataricus*. The mutant and its parental strains were compared with respect to growth characteristics, transcriptome profiles, and sensitivity to ribosome-targeting antibiotics. Both these results and the outcomes of a thorough molecular interaction analysis published elsewhere [46] (Chapter 5) suggest that archaeal MBF1 orthologs might exert their regulatory role beyond the level of transcription.

Material and Methods

Genetic manipulation of Sulfolobus solfataricus

Strains used in this study were *Sulfolobus solfataricus* PBL2025 (PBL2025), a derivative of strain 98/2 with a deletion that spans the ORFs SSO3004 to SSO3050, including SSO3019 that encodes a beta-galactosidase (LacS) necessary for growth on lactose as carbon source [47], and a PBL2025 derivative *S. solfataricus* Δ*mbf1* (Δ*mbf1*).

Δ*mbf1* was produced as follows. Flanking regions of *mbf1* (SSO0270) and *lacS*, including promoter and terminator, were amplified by PCR from genomic DNA using the following primers: for *lacS* amplification BG2009 and BG2010, for the upstream flank BG2019 and BG2020, and for the downstream flank BG2017 and BG2018 (Table 1). The suicide recombination plasmid pWUR443 was made by introducing the *lacS* gene between the upstream and downstream flanks into the multiple cloning site of pUC29.

Electroporation of *S. solfataricus* was performed as described previously [48]. Electrocompetent *S. solfataricus* PBL2025 cells were prepared from cultures grown on Brock's medium with 0.1% (w/v) tryptone and 0.4% (w/v) sucrose by washing the cells twice with a 20 mM sucrose solution. Plasmids were used to transform 50 µl of the competent cells by electroporation (2 mm cuvette, 1.5 kV, 400 Ω, and 25 µF). After electroporation, cells were immediately transferred to deionized water and placed for 1 minute on ice, followed by an

incubation step of 10 minutes at 75°C. Cells were subsequently transferred to prewarmed medium containing 0.4% (w/v) lactose. After initial growth, cells were transferred to fresh 0.4% (w/v) lactose containing Brock's medium, grown again to OD_{600} of approximately 1.0, and plated on plates containing Brock's medium supplemented with 0.1% (w/v) tryptone and 1% (w/v) gelrite. The presence of LacS was tested by spraying the plates with X-gal solution (5 mg/ml in 20% DMF solution). Blue colonies were picked, grown in tryptone containing medium, and analysed for the presence of *lacS* and *mbf1* genes by PCR. Genomic DNA was isolated using the QuickPick SML gDNA Kit (Bionobile). Plating and culturing were repeated until homogeneous cultures were obtained.

S. solfataricus growth

S. solfataricus strains were grown in modified Brock's medium at 75°C using an incubator (New Brunswick) shaking at 120 rpm, or an oil bath (New Brunswick) at 180 rpm [49]. Sucrose (0.4% w/v) and tryptone (0.1% w/v) were used as carbon source. To avoid evaporation Erlenmeyer flasks with elongated necks were used. In addition *S. solfataricus* was grown in tubes (30 mm diameter) containing 20 ml medium and closed with Silicosen T-32 plugs (Hirschmann Laborgeräte) to reduce water evaporation. Growth was monitored by measuring the optical density at 600 nm (OD_{600}) in a Hitachi spectrophotometer U-1500.

For sulphate limitation experiments, the medium composition was altered as follows. $MgSO_4$ was replaced by $MgCl_2$ (0.21 gl^{-1}) and the pH of the medium was adjusted with hydrochloric acid instead of sulphuric acid. In this way, the sulphate concentration was lowered to 1 μM compared to 11 mM in the original medium. Cultures were first grown in duplicate in 50 ml normal medium containing 0.4% sucrose, washed by low centrifugation (5,000×g, 3 min, room temperature) and resuspending in sulphate limited medium, and then transferred in duplicate to tubes containing either sulphate limited medium with 0.4% sucrose in duplicate or standard medium with 0.4% sucrose serving as control.

Table 1. *Primers used in this study.* d: downstream, u: upstream, fw: forward, rev: reverse.

Primer	Name	Sequence 5'→3' (restriction site underlined)	Restr. site
Primers used for genetic manipulation of S. solfataricus			
BG2009	*lacS* fw	CCGGCC<u>GGATCC</u>CTATATCAATCTCTTTTTGAAAGTGC	BamHI
BG2010	*lacS* rev	GCGCGC<u>GGTACC</u>GTAACAGTATTAAATCTAAATGAC	KpnI
BG2017	d.flank *mbf1* fw	GCGCGC<u>CTGCAG</u>GGGGAGTATAATTCTTTGGACAG	PstI
BG2018	d.flank *mbf1* rev	GCGCGC<u>TCTAGA</u>TCTCAAGCTAGGCAATTAGAG	XbaI
BG2019	u.flank *mbf1* fw	GCGCGC<u>GCGGCCGC</u>TGCCTTTACTATCACTGCATG	NotI
BG2020	u.flank *mbf1* rev	GCGCGC<u>ATGCAT</u>GGCCAAGAGCTGCATTAGCG	NsiI
Primers used in PCR analysis and Southern blot hybridization			
BG2961	*mbf1* probe fw	TTGTAGGTTTCAATTTACCACTTTC	-
BG2962	*mbf1* probe rev	CCAAAATGGAAAATGCTGAA	-
BG2637	*lacS* probe fw	GGGGGCCATGGACTCATTTCCAAATAGCTTTAGG	-
BG2638	*lacS* probe rev	CCATAGAGGTAATGGCCAATGATACATG	-

Transcriptome analysis

Transcriptome analysis was performed as described previously in detail [49]. In brief, cultures of PBL2025 and Δ*mbf1* were grown to an OD_{600} of approximately 0.6, cooled down in liquid nitrogen and ice water, and harvested by centrifugation (4,000×g, 10 minutes, 4°C). Total RNA was extracted using the MirVana miRNA isolation kit (Ambion). cDNA was synthesized using Superscript III reverse transcriptase (Invitrogen) and a mixture of nucleotides including 1:4 TTP and aminoallyl-dUTP (Ambion), labelled using Alexa dyes 647 and 555 (Invitrogen), and hybridized to a 70-mer oligonucleotide DNA microarray containing 8,860 spots, covering approximately 3,500 *S. solfataricus* genes, in duplicate (Ocimum Biosolutions). For *mbf1* two different oligonucleotides were included. Arrays were scanned using a GenePix Pro 4000B scanner (Axon), and data was analysed using GenePixPro 6.0 (Axon), Midas software (TIGR), and Excel (Microsoft).

Southern blot hybridization

Southern blots were performed as has been described earlier [50]. Genomic DNA was extracted from *Sulfolobus* strains by phenol extraction as described before [49]. Genomic DNA was treated with RNase A (10 µg/ml) overnight at 4°C. Extracted DNA was digested with AflII (New England Biolabs) and with HindIII (New England Biolabs), electrophoresed on 0.2% agarose gel, and transferred to a nytran membrane (Perkin Elmer) via capillary transfer. DNA was immobilized by incubating the membranes for 2 hours at 90°C. PCR-generated probes against *mbf1* and *lacS* (Table 1) were labelled with Digoxigenin using DIG High Prime according to the manufacturer's protocol (Roche). Prehybridization of the membranes was performed by incubating the membranes 4 hours in hybridization buffer (50% freshly deionized formamide, 5×SSC, 2% blocking reagent (Roche), 0.1% Na-lauroylsarcosine, 0.02% SDS) at 60°C. Hybridization with the probes was done overnight at 60°C in fresh hybridization buffer. Membranes were washed twice in 2×SSC, 0.1% SDS, twice in 0.2×SSC, 0.1% SDS at 60°C, once in maleic acid buffer (0.1 M maleic acid, pH 7.5, 0.15 M NaCl) with 0.3% Tween20, incubated for 30 minutes in maleic acid buffer with 2% blocking reagent followed by an incubation for 30 minutes in the same solution with Anti-Digoxigenin-AP (Roche). Then they were washed again twice in maleic acid buffer with 0.3% Tween20. After two final washes and in Assay buffer (100 mM diethanolamine, pH 10, 1 mM $MgCl_2$, 100 mM NaCl), membranes were incubated with 1:100 diluted CDP-star in the supplied buffer (New England Biolabs). Signals were captured on BioMax light films (Kodak) and films were scanned with a GS800 densitometer (BioRad).

Immunodetection of S. solfataricus MBF1

Rabbit Antiserum against the C-terminal helix-turn-helix domain of *S. solfataricus* MBF1 (residues 57-165) was generated at Eurogentec. To purify the antibodies, the final bleed was filtered through a 0.2 µM filter, purified over Protein A-agarose (Sigma), and eluted at low pH according to standard procedures. Antibodies were concentrated by ultrafiltration and buffer exchanged to PBS. Antibodies were reacted with Digoxigenin-3-O-methylcarbonyl-ε-aminocaproic acid-N-hydroxysuccinimide ester (Roche) in a molar ratio of 1:10 according to manufacturer's protocol (Roche).

PBL2025 and Δ*mbf1* cell extracts were separated by bis-Tris SDS-PAGE and transferred overnight to Nitrocellulose membranes (0.2 µm pore size, BioRad) in 10 mM CAPS (pH 11.0), 10% methanol at 10 V using a tank transfer system. Efficient transfer was verified by PonceauS staining. Filters were washed twice in TBS (20mM Tris-HCl, pH 7.9; 150 mM NaCl) and

incubated for 1hr in blocking solution (0.2 % i-block (Applied Biosystems), 0.1 % Tween-20 in TBS). Filters were then incubated with anti-MBF1 antibody diluted 1:500 in blocking solution for 1 hr. After three 15 minute washes in 50-100 ml TTBS (TBS, 0.2 % Tween-20), filters were incubated with Anti-Digoxigenin-AP, Fab fragments (Roche) diluted 1:1500 in blocking solution for 30 min. After three more TTBS washes, and an additional 5 min TBS wash, detection of MBF1 was performed by chemiluminescence using 1:100 diluted CDP-Star reagent (New England Biolabs) and Kodak Biomax films for signal capture.

tRNA methylation assays

500 ml cultures of PBL2025 and Δ*mbf1* were grown to an OD_{600} of approximately 1.0 in duplicate. Cells were pooled and spun down (4,000×g for 10 minutes at 4°C). PBL2025 cell extracts were prepared by sonication from 0.5 g cells resuspended in 3 ml 50 mM Tris-HCl buffer (pH 8). tRNAs were extracted from PBL2025 and Δ*mbf1* as described [51]. 80 μg of each tRNA preparation was incubated with 30 µl PBL2025 cell extract in the presence of 25 nCi [methyl-^{14}C] S-adenosyl-L-methionine (50 mCi/mmol; GE Healthcare) for 1 hour at 70°C. *In vitro* methylation of tRNA was subsequently monitored by purifying the tRNAs and measuring incorporation of ^{14}C by scintillation counting [52, 53].

In addition, the isolated tRNAs were completely digested by nuclease P1, and separated by 2-dimensional thin layer chromatography on 20 × 20 cm cellulose plates (Merck) using isobutyric acid/concentrated NH_4OH/water (66:1:33 v/w/v) for the first dimension and 0.1 M $NaPO_4$ (pH 6.8)/$(NH_4)_2SO_4$/*n*-propanol (100:60:2 v/w/v) for the second dimension. The radioactively labelled nucleotides were detected by autoradiography [52, 53].

Results

Disruption mutagenesis

The *mbf1* gene has been reported to be the first gene of a larger operon in *S. solfataricus* [20, 54] that also contains genes coding for an HflX-type GTPase [55], the general transcription factor TFE [56], a DNA methyltransferase, and a hypothetical protein. The expression of these genes could potentially have been affected by the *mbf1* disruption as well. To avoid accidental overexpression of these genes, *lacS* and its promotor and terminator were inserted in reverse orientation (Fig. 1A). Disruption mutagenesis of the *mbf1* gene *S. solfataricus* was successfully accomplished. After mutagenesis, the selection marker *lacS* was present, while the *mbf1* gene could not be detected by PCR (Fig. 1B), and Southern blot hybridization (data not shown). Immunodetection also revealed loss of MBF1 protein in the disruption mutant (Fig. 1C).

Growth characteristics

The conservation of *mbf*1 in all eukaryotes and almost all archaea suggests a crucial function for this gene in the regulation of cellular metabolism. However, growth of Δ*mbf1* was not significantly different from PBL2025 under standard laboratory growth conditions (Fig. 2ABC). Comparison of PBL2025 with Δ*mbf1* revealed similar growth characteristics on media supplemented with different sugars or peptide mixtures as growth substrates in presence or absence of additional vitamins (Fig. 2D). Furthermore, Δ*mbf1* was able to grow on media containing lactose unlike PBL2025 due the introduction of *lacS* as a selection marker during the disruption process.
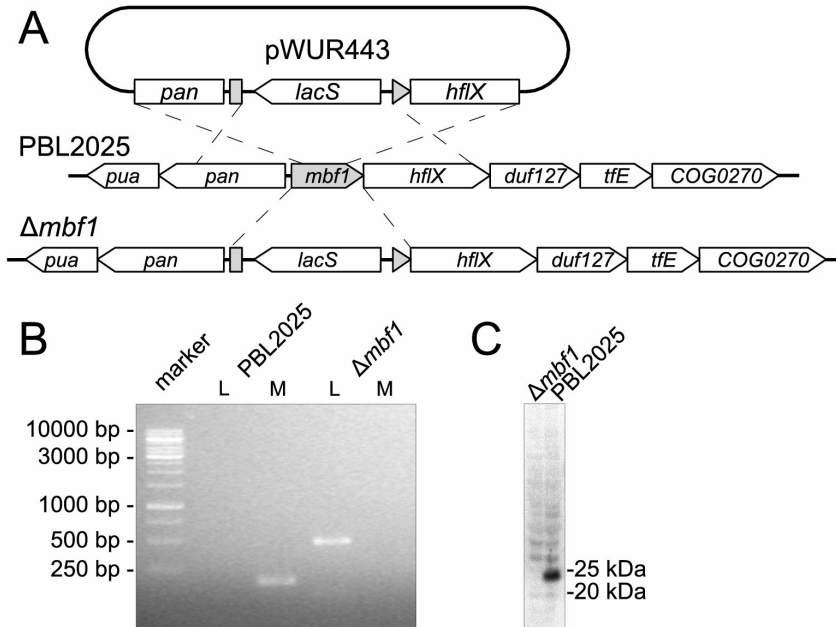
Figure 1. Disruption of mbf1 in S. solfataricus. (A) Overview of the genetic organization in *S. solfataricus* after recombination with pWUR443 plasmid. (B) PCR using primer sets to detect the presence of *lacS* (L) or *mbf1* (M) for PBL2025 and Δ*mbf1*. (C) Immunodetection of MBF1 in Δ*mbf1* and PBL2025.
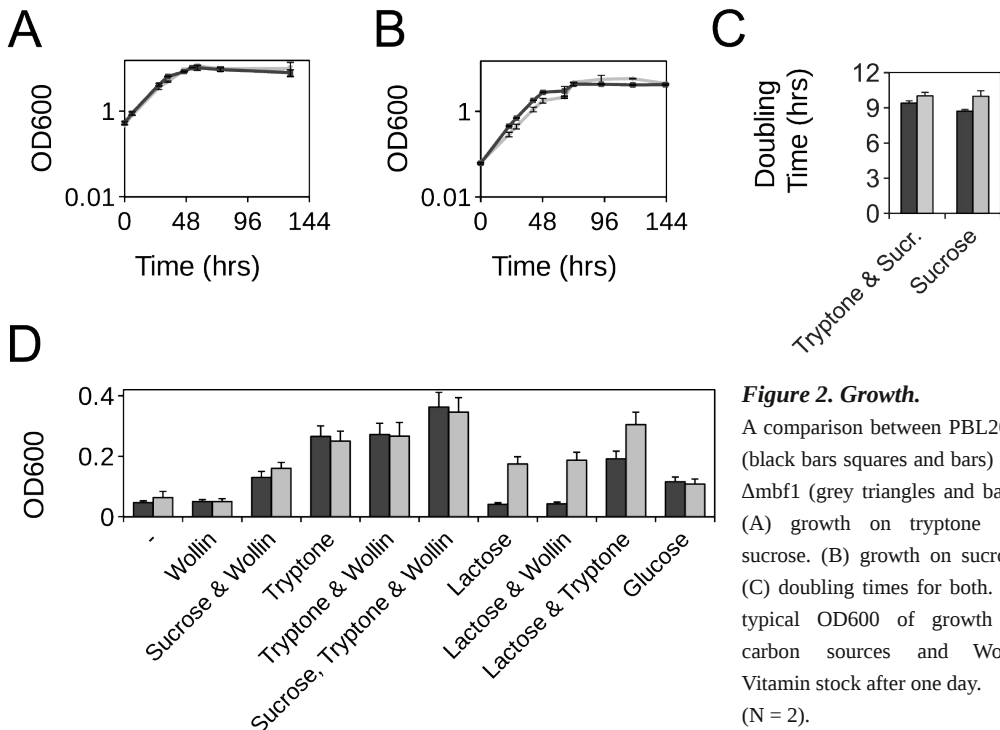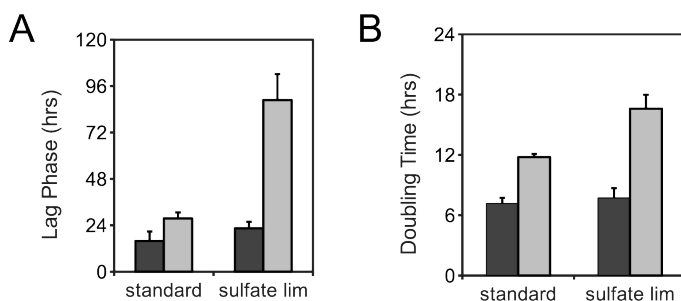

Figure 2. Growth.

A comparison between PBL2025 (black bars squares and bars) and Δmbf1 (grey triangles and bars). (A) growth on tryptone and sucrose. (B) growth on sucrose. (C) doubling times for both. (D) typical OD600 of growth on carbon sources and Wollin Vitamin stock after one day. (N = 2).
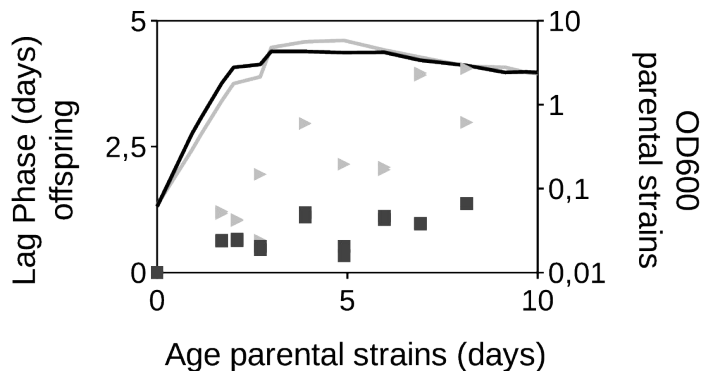
Figure 3. Relationship between lag phase upon transfer to fresh medium and the age of the parental culture. Lines depicts the growth of the parental cultures. PBL2025 (black squares and line) and Δ*mbf1* (grey triangles and line).
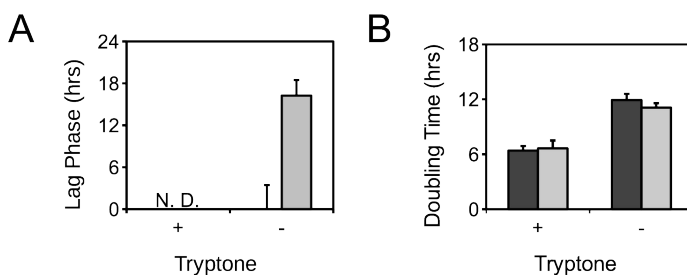
Figure 4. Aminoacid starvation experiment. Cultures are grown in medium with 0.1% tryptone and 0.4% sucrose. On day 1, cultures were transferred in duplo to medium with and without 0.1% tryptone, both contained 0.4% sucrose. The experiment is performed twice. Growth characteristics are shown for PBL2025 (black) and Δ*mbf1* (grey).

Figure 5. Sulphate limitation experiment. Cultures are grown on normal medium ($[SO_4]$ 11 mM) and on sulphate limited medium ($[SO_4]$ 1.0 µM), both contained 0.4% sucrose. Characteristics of growth are shown for PBL2025 (black) and Δ*mbf1* (grey), a lag phase could not be detected (N. D.) for growth of both strains on tryptone and sucrose. Experiment is performed two times in duplicate.

To test the survival of Δ*mbf1* during the stationary growth phase, aliquots of a culture growing on 0.4% sucrose were transferred at different time points to new medium (of similar composition) in duplicate and growth was monitored. When the parental cultures entered the stationary growth phase, this led to a significant increase in the lag phase of the newly transferred offspring cultures, while this effect was less pronounced for PBL2025 (Fig. 3).

A distinct phenotype for Δ*mbf1* was found in response to nutrient stress. Transfer of Δ*mbf1* and PBL2025 from medium containing tryptone and sucrose to medium containing only sucrose as carbon source caused a prolonged lag phase of Δ*mbf1* (Fig. 4A). No apparent difference in growth rate between PBL2025 and Δ*mbf1* was observed, but for both strains the doubling time increased upon amino acid starvation (Fig. 4B). In addition to tryptone omission, severe sulphate limitation (1.0 μM versus 11 mM) as nutrient stress was tested as well. Severe sulphate limitation caused an increased lag phase as well as a significantly reduced growth rate of Δ*mbf1* (Fig. 5AB). Surprisingly, the growth rate of Δ*mbf1* appeared to be also affected by an unidentified, additional stress factor, because doubling time of Δ*mbf1* in the standard medium raised already to levels associated with stress response, as observed for example for PBL2025 and Δ*mbf1* during late lag phase transfers (data not shown) and in the starvation experiment (Fig. 4B). This was probably caused by washing of the cells in sulphate deprived medium that was carried out in order to prepare the inoculum to avoid unintended transfer of sulphate. A heat shock, applied by a short term elevation of the temperature of the medium (up to 93°C for 4 hrs), caused no measurable alterations in growth characteristics (data not shown).

Transcriptome profiles

In eukaryotes, a well defined role of MBF1 in transcription regulation has been reported [23–38] and it has been suggested that the archaeal MBF1 ortholog might have a similar function [19, 31]. Recently Ying et al. published the first outcomes of an RNAseq experiment on a eukaryotic *mbf1* disruption in *Beauveria bassiana* (a fungal insect pathogen), but outcomes

Table 2. Significantly upregulated and downregulated genes in Δmbf1 in comparison to PBL2025*.

ID	Annotation	Log$_2$ ratio	StDev
Upregulated			
SSO3019	β-glycosidase (lacS)	6.33	1.17
SSO2621	hypothetical protein	1.40	0.51
Downregulated			
SSO2146	conserved hypothetical protein	-1.25	0.14
SSO2693	acylaminoacyl-peptidase, putative (apeH-3)	-1.25	0.25
SSO0270	multiprotein bridging factor 1	-3.20	1.04[**]
SSO0284	hypothetical protein conserved within sulfolobales	-3.43	0.50
SSO2527	prolidase (Xaa-Pro dipeptidase) (pepQ-like3)	-4.18	0.41
SSO2514	3-hydroxyacyl-CoA dehydrogenase/enoyl CoA hydratase	-4.60	0.57
SSO0872	hypothetical two-domain protein conserved in archaea	-4.74	0.76

*Only spots that are considered significant ($p < 0.01$) and that are measured in more than 66% of the arrays (minimum 8 out of 12) were taken into account.

**Two different probes for mbf1 (Sso0270) were present on the array.

did not reveal a clear regulon, on the contrary, a wide range of cell processes was found to be affected, with a large number of transporters amongst them [41]. In order to test a role of archaeal MBF1 in transcriptional control, we conducted a transcriptome analysis that could potentially reveal specific transcription regulation patterns for MBF1. Transcriptome profiles were compared between PBL2025 and Δ*mbf1* in mid-exponential growth phase under normal growth conditions. Surprisingly, only few genes were differentially expressed apart from the disrupted *mbf1* gene and the introduced *lacS* selection marker gene (Table 2). Down regulated genes include: two genes encoding hypothetical proteins, one restricted to the *Sulfolobales*, the other more commonly found in archaea, and a gene that encodes a 3-hydroxyacyl-CoA dehydrogenase/eonyl CoA hydratase. The genes that cluster with *mbf1* in an operon, like *hflX*, and *tfe* [54], were not found to be differentially expressed.

Sensitivity to ribosome-targeting antibiotics

In yeast, *mbf1* deletion leads to an altered sensitivity to some aminoglycoside antibiotics targeting the ribosome [45]. In addition, it was also reported that yeast MBF1 also influences reading frame maintenance by the ribosome [44, 45, 57]. It has remained unclear how this phenotype is linked to the role of MBF1 as transcription co-activator. In contrast, these results suggest a role of MBF1 in translation fidelity, albeit by a so far unknown molecular mechanism. In order to investigate whether the archaeal MBF1 affects translation fidelity in a similar way, we tested the Δ*mbf1* and PBL2025 strains for their sensitivity to several ribosome-targeting antibiotics. Although the archaeal translation machinery appeared to be rather insensitive to many antibiotics, a limited number of antibiotics have been shown to inhibit translation, or to interfere with tRNA selection *in vitro* [58, 59]. Neomycin, sisomycin, puromycin, and tetracycline cause modest inhibition of archaeal translation [58], whereas paromomycin causes misreading, but no inhibition of translation [59]. No effect on translation was found using kanamycin. The ethanol concentrations, used to dissolve tetracycline, did not result in growth inhibition either.

Growth rates of Δ*mbf1* and PBL2025 were compared in the presence of varying concentrations of several different antibiotics (Fig. 6). Neomycin, and sisomycin did not appear to inhibit growth of PBL2025 even at high concentrations (data not shown). Tetracycline and
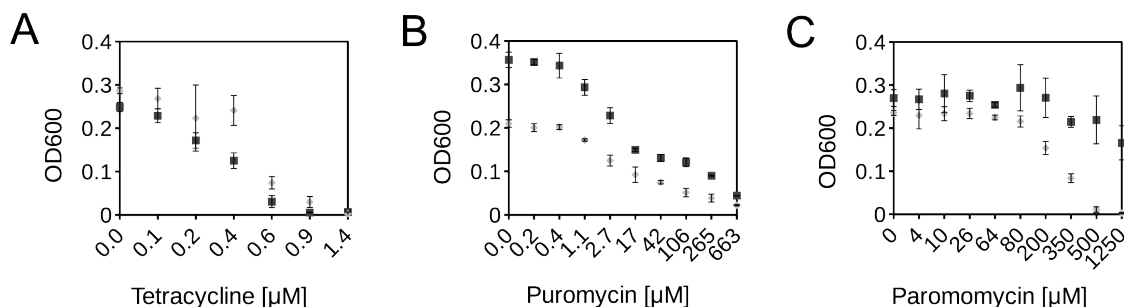

Figure 6. Effect of antibiotics on growth. Growth on 0.4% sucrose medium in the presence of varying amounts of tetracycline (A), puromycin (B), and paromomycin (C) after one day for PBL2025 (black squares) and Δmbf1 (grey diamonds). (tetracycline N = 3, paromomycin N = 3, puromycin N = 2).

puromycin were found to inhibit the growth of PBL2025 and Δ*mbf1* equally (Fig. 6AB). In contrast, paromomycin affected growth of Δ*mbf1* stronger than that of PBL2025 (Fig. 6C).

tRNA methylation

The increased sensitivity to ribosome targeting antibiotic paromomycin indicated that MBF1 might have a translation-related function. It was hypothesized that the translation related phenotype observed for the *mbf1* deletion strain in yeast might be due to a role of MBF1 in tRNA synthesis [45]. tRNAs are heavily modified nucleic acids [60], and these modifications are known to influence fidelity of translation [61, 62]. As a first test whether tRNA modification is altered in Δ*mbf1*, the extent of tRNA methylation was studied. An *in vitro* methylation assay was used that measures the incorporation of ^{14}C-methyl groups into the tRNAs in order to compare the level of tRNA methylation between both strains. Furthermore, it was tested whether a specific subset of nucleotides was affected in its methylation by disruption of *mbf1*. Nucleotides were prepared by complete nuclease digestion of tRNA and separated using thin layer chromatography. However, no differences in methylation were observed between PBL2025 and Δ*mbf1* (Fig. 7).
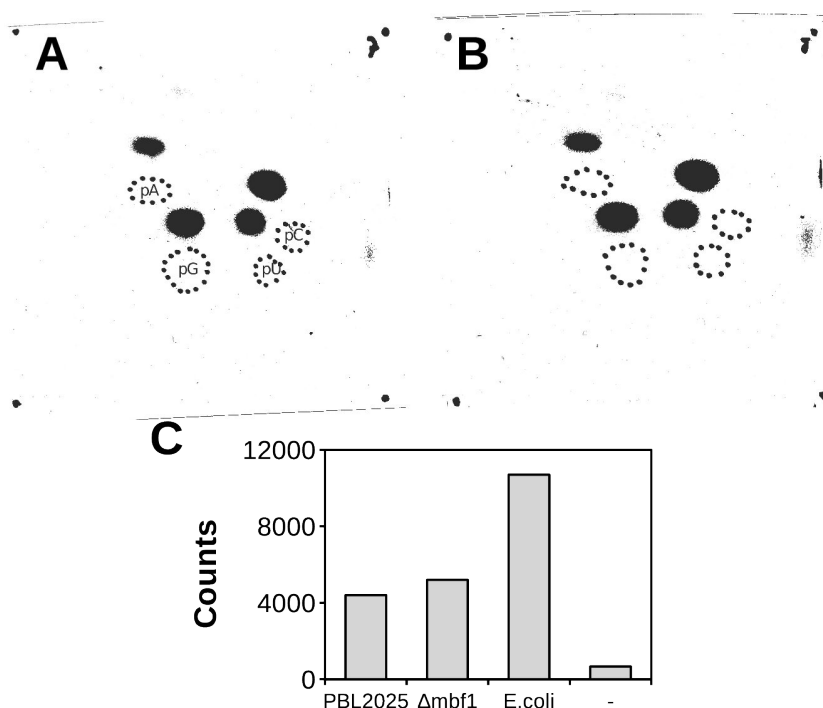
Figure 7. tRNA methylation analyses. Thin layer chromatography separations of digested tRNAs obtained from PBL2025 (A) and Δ*mbf1* (B). Comparison between ^{14}C-methyl group incorpation into tRNAs by scintillation counting; - indicates a blanco measurement.

Discussion

Disruption of the *mbf1* gene in the crenarchaeon *S. solfataricus* is not lethal. We found *mbf1* deletion in the euryarchaeon *Thermococcus kodakaraensis* to be not lethal as well (Matsumi, Atomi, de Koning, and Van der Oost, unpublished results). These observations are in agreement with the fact that *mbf1* mutants are also viable in a diverse range of eukaryotic species (yeast, fungi, fruit-flies, and plants) [27, 36, 38, 41, 45, 63, 64]. Hence, *mbf1* appears to be non-essential (definition according to Jordan et al. [65]) in both eukaryotes as well as in archaea.

The disruption of *mbf1* in *S. solfataricus* leads to a prolonged lag phase after exposing the cells to a diverse range of stresses. Generally, this could be due to an increased rate of cell death during the exposure to stress conditions, or to a decreased rate of adaptation of cell metabolism to a change in growth conditions. The latter assumption is supported by the finding that in one case also the growth rate of Δ*mbf1* on standard medium was impaired, i.e. after washing of the cells in sulphate deprived medium. Apparently the Δ*mbf1* cells that survived the treatment were still in the process of adaptation, while the PBL2025 cells grew at a normal pace. Thus, although archaeal MBF1 is not directly involved in carbon metabolism, it is involved in the adaptation of cells to a change in growth conditions. Notably, this matches observations made for several eukaryotic *mbf1* disruption strains that grow normally under standard growth conditions but are more sensitive to different stress conditions. In yeast, a Δ*mbf1* strain is susceptible to histidine starvation and shows increased sensitivity to antibiotics that target the ribosome. Under standard conditions, however, normal growth is observed [36, 45]. In *Drosophila melanogaster*, a Δ*mbf1* strain showed increased sensitivity to oxidative stress [27], while in *Arabidopsis thaliana* deletion of the three *mbf1* paralogs results in reduced thermotolerance and high stress susceptibility during germination and early growth [38, 63, 66]. In addition, in plants, like *Arabidopsis*, tomato, potato, sorghum, and rice, and in other animals, like crayfish, expression experiments of either wildtype strains or *mbf1* overexpression strains shows a clear relation of *mbf1*, especially orthologs of *mbf1c*, to stress tolerance pathways [23, 38, 41, 67–77].

Transcriptome analysis revealed that only a very limited number of genes is significantly up or down regulated in *S. solfataricus* Δ*mbf1* (Table 2). Moreover, the regulated genes do not give a coherent picture concerning their physiological function and they do not seem to be highly conserved as some occur only in the Sulfolobales. Thus, they are unlikely to represent a regulon of *S. solfataricus* MBF1. These results are therefore in marked contrast to the prediction for MBF1 being a transcription regulator in archaea, although it cannot be ruled out that the role of the archaeal MBF1 in transcription regulation is limited only to stress conditions. Moreover, transcriptome profiles obtained from a *Thermococcus kodakaraensis mbf1* disruption mutant under normal laboratory growth conditions did not reveal any significant differences to the parental strain, with the sole exception of *mbf1* and the selection marker gene encoding a Anthranilate synthase (Matsumi, Atomi, de Koning, and Van der Oost, unpublished results), suggesting that archaeal MBF1 orthologs generally do not have a role in transcription regulation.

Our analysis revealed an increased sensitivity of Δ*mbf1* to paromomycin as deduced from a significantly different inhibition profile between PBL2025 and Δ*mbf1*. Sensitivity to tetracycline and puromycin remained similar to that of PBL2025. It is unknown whether these antibiotics act on ribosomes of *S. solfataricus* similarly *in vivo* and *in vitro* [58, 59]. The stability of these antibiotics is likely to be affected by the low pH and high temperature growth conditions of *S. solfataricus*. However, the pronounced effect of paromomycin on translational

misreading *in vitro* suggests that paromomycin is likely to interfere with translation *in vivo* as well. Although *in vitro* misreading assays performed thus far did not reveal significant differences in misreading rates between cell lysates prepared from Δ*mbf1* and PBL2025 in the presence of paromomycin [46] (Chapter 6), but other effects such as stop codon read-through are likely to have more drastic effects *in vivo* especially in the long run [78].

The molecular basis for the contribution of MBF1 to translation fidelity observed in yeast is unknown. However, the fact that the altered aminoglycoside sensitivity was found both for yeast and *S. solfataricus* Δ*mbf1* strains suggests that the role of MBF1 in translation fidelity might be conserved in eukaryotes and archaea. In an attempt to get a first insight into the mechanism how *mbf1* gene deletion affects translation fidelity, tRNA methylation has been investigated, as tRNA modification is known to affect translation fidelity [79, 80]. However, tRNA modification did not appear to be different between PBL2025 and Δ*mbf1*, but MBF1 might alternatively play a role in the assembly or functioning of the ribosome. This alternative role is further supported by evidence for a physical interaction of MBF1 with the small ribosomal subunit in *S. solfataricus* that we published recently [46] (Chapter 6).

Recently it was shown that, despite the high conservation in the C-terminal HTH domain, archaeal *mbf1*s could not rescue growth of a yeast *mbf1* deletion mutant in the presence of aminotriazole [21]. It was reasoned that this might be caused by the incapability of archaeal MBF1 to bind GCN4, and that the interaction with TBP might still be present. This could provide an alternative mode of action for transcription regulation than co-activation, for example by direct binding of DNA. This, however, does not comply with our findings that there are hardly expression alterations found in *mbf1* deletion mutants, the increased sensitivity to paromomycin, and the binding of MBF1 to the small ribosomal subunit. Even more, despite numerous efforts (amongst others pull down assays, Y2H screenings, and immunological co-purification), the proposed interaction with the archaeal TBP could not be detected (unpublished results). Archaeal TBP remained even undetected in a high throughput screening for archaeal MBF1 interaction partners, which revealed mainly ribosomal proteins [46]. Alternatively, the transcriptional mode of action might be acquired relatively late during eukaryotic evolution, and that the original and probably the main mode of function for archaeal and eukaryotic MBF1 might be a role during translation. This is additionally supported by another recent findings in eukaryotes. In yeast, MBF1 was found in a screening for RNA binding proteins [81]. Although the RNA binding domain appeared to be the N-terminal domain, which is not conserved between archaea and eukaryotes [20, 21]. This yeast finding appears to be evolutionary conserved among eukaryotes as both in human HEK293 [82] and mouse embryonic stem cells [83] MBF1 (called EDF1 in vertebrates) were also present at significant levels in recent large screenings for mRNA binding proteins.

In both archaea and eukaryotes *mbf1* appears to be a non-essential gene that functions in the adaptation to changing growth conditions and, in higher organisms, development. We did not obtain any evidence that this function in archaea might be a regulatory function on the level of transcription. However, archaeal MBF1 might play some role in translation fidelity, as has been also demonstrated in yeast. Recent findings point at a direct interaction between MBF1 and the translation apparatus, which could be the molecular basis of these observations. Further biochemical experiments to elucidate the role of MBF1 during translation in both eukaryotic and archaeal organisms will be required to verify the hypothesis that MBF1 acts as a translational factor in both archaea and eukaryotes.

Acknowledgements

References

1. Bell SD, Jackson SP: **Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features.** *Trends Microbiol* 1998, **6**:222–8.
2. Blombach F, Makarova KS, Marrero J, Siebers B, Koonin E V, van der Oost J: **Identification of an ortholog of the eukaryotic RNA polymerase III subunit RPC34 in Crenarchaeota and Thaumarchaeota suggests specialization of RNA polymerases for coding and non-coding RNAs in Archaea.** *Biol Direct* 2009, **4**:39.
3. Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM: **The archaebacterial origin of eukaryotes.** *Proc Natl Acad Sci U S A* 2008, **105**:20356–61.
4. Ettema TJG, de Vos WM, van der Oost J: **Discovering novel biology by in silico archaeology.** *Nat Rev Microbiol* 2005, **3**:859–69.
5. Huet J, Schnabel R, Sentenac A, Zillig W: **Archaebacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type.** *EMBO J* 1983, **2**:1291–4.
6. Korkhin Y, Unligil UM, Littlefield O, Nelson PJ, Stuart DI, Sigler PB, Bell SD, Abrescia NGA: **Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure.** *PLoS Biol* 2009, **7**:e1000102.
7. Kyrpides NC, Woese CR: **Archaeal translation initiation revisited: the initiation factor 2 and eukaryotic initiation factor 2B alpha-beta-delta subunit families.** *Proc Natl Acad Sci U S A* 1998, **95**:3726–30.
8. Lundgren M, Andersson A, Chen L, Nilsson P, Bernander R: **Three replication origins in Sulfolobus species: synchronous initiation of chromosome replication and asynchronous termination.** *Proc Natl Acad Sci U S A* 2004, **101**:7046–51.
9. McGeoch AT, Bell SD: **Extra-chromosomal elements and the evolution of cellular DNA replication machineries.** *Nat Rev Mol Cell Biol* 2008, **9**:569–74.
10. Myllykallio H, Lopez P, López-García P, Heilig R, Saurin W, Zivanovic Y, Philippe H, Forterre P: **Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon.** *Science* 2000, **288**:2212–5.
11. Olsen GJ, Woese CR: **Archaeal genomics: an overview.** *Cell* 1997, **89**:991–4.
12. Przytycka TM, Jothi R, Aravind L, Lipman DJ: **Differences in evolutionary pressure acting within highly conserved ortholog groups.** *BMC Evol Biol* 2008, **8**:208.
13. Yutin N, Makarova KS, Mekhedov SL, Wolf YI, Koonin E V: **The deep archaeal roots of eukaryotes.** *Mol Biol Evol* 2008, **25**:1619–30.
14. De Koning B, Blombach F, Brouns SJJ, Van Der Oost J: **Fidelity in Archaeal Information Processing**. *Archaea Vancouver BC* 2010, **2010**:1–34.
15. Berthon J, Cortez D, Forterre P: **Genomic context analysis in Archaea suggests previously unrecognized links between DNA replication and translation.** *Genome Biol* 2008, **9**:R71.
16. Berthon J, Fujikane R, Forterre P: **When DNA replication and protein synthesis come together.** *Trends Biochem Sci* 2009, **34**:429–34.

17. Koonin E V, Wolf YI, Aravind L: **Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach.** *Genome Res* 2001, **11**:240–52.

18. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM: **The many faces of the helix-turn-helix domain: transcription regulation and beyond.** *FEMS Microbiol Rev* 2005, **29**:231–62.

19. Aravind L, Koonin E V: **DNA-binding proteins and evolution of transcription regulation in the archaea.** *Nucleic Acids Res* 1999, **27**:4658–70.

20. De Koning B, Blombach F, Wu H, Brouns SJJ, Van Der Oost J: **Role of multiprotein bridging factor 1 in archaea: bridging the domains?** *Biochem Soc Trans* 2009, **37**(Pt 1):52–57.

21. Marrero Coto J, Ehrenhofer-Murray AE, Pons T, Siebers B: **Functional analysis of archaeal MBF1 by complementation studies in yeast.** *Biol Direct* 2011, **6**:18.

22. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T: **Insights into the phylogeny and coding potential of microbial dark matter.** *Nature* 2013, **499**:431–7.

23. Arce DP, Tonón C, Zanetti ME, Godoy AV, Hirose S, Casalongué CA: **The potato transcriptional co-activator StMBF1 is up-regulated in response to oxidative stress and interacts with the TATA-box binding protein.** *J Biochem Mol Biol* 2006, **39**:355–60.

24. Bolognese F, Pitarque-Martì M, Lo Cicero V, Mantovani R, Maier JAM: **Characterization of the human EDF-1 minimal promoter: involvement of NFY and Sp1 in the regulation of basal transcription.** *Gene* 2006, **374**:87–95.

25. Brendel C, Gelman L, Auwerx J: **Multiprotein bridging factor-1 (MBF-1) is a cofactor for nuclear receptors that regulate lipid metabolism.** *Mol Endocrinol* 2002, **16**:1367–77.

26. Busk PK, Wulf-Andersen L, Strøm CC, Enevoldsen M, Thirstrup K, Haunsø S, Sheikh SP: **Multiprotein bridging factor 1 cooperates with c-Jun and is necessary for cardiac hypertrophy in vitro.** *Exp Cell Res* 2003, **286**:102–14.

27. Jindra M, Gaziova I, Uhlirova M, Okabe M, Hiromi Y, Hirose S: **Coactivator MBF1 preserves the redox-dependent AP-1 activity during oxidative stress in Drosophila.** *EMBO J* 2004, **23**:3538–47.

28. Kabe Y, Goto M, Shima D, Imai T, Wada T, Morohashi K i, Shirakawa M, Hirose S, Handa H: **The role of human MBF1 as a transcriptional coactivator.** *J Biol Chem* 1999, **274**:34196–202.

29. Li FQ, Ueda H, Hirose S: **Mediators of activation of fushi tarazu gene transcription by BmFTZ-F1.** *Mol Cell Biol* 1994, **14**:3013–21.

30. Li FQ, Takemaru K, Goto M, Ueda H, Handa H, Hirose S: **Transcriptional activation through interaction of MBF2 with TFIIA.** *Genes Cells* 1997, **2**:143–53.

31. Liu Q-X, Nakashima-Kamimura N, Ikeo K, Hirose S, Gojobori T: **Compensatory change of interacting amino acids in the coevolution of transcriptional coactivator MBF1 and TATA-box-binding protein.** *Mol Biol Evol* 2007, **24**:1458–63.

32. Mariotti M, De Benedictis L, Avon E, Maier JA: **Interaction between endothelial differentiation-related factor-1 and calmodulin in vitro and in vivo.** *J Biol Chem* 2000, **275**:24047–51.

33. Babini E, Hu X, Parigi G, Vignali M: **Human multiprotein bridging factor 1 and Calmodulin do not interact in vitro as confirmed by NMR spectroscopy and CaM-agarose affinity chromatography.** *Protein Expr Purif* 2011, **80**:1–7.

34. Millership JJ, Waghela P, Cai X, Cockerham A, Zhu G: **Differential expression and interaction of transcription co-activator MBF1 with TATA-binding protein (TBP) in the apicomplexan Cryptosporidium parvum.** *Microbiology* 2004, **150**(Pt 5):1207–13.

35. Takemaru K i, Li FQ, Ueda H, Hirose S: **Multiprotein bridging factor 1 (MBF1) is an evolutionarily conserved transcriptional coactivator that connects a regulatory factor and TATA element-binding protein.** *Proc Natl Acad Sci U S A* 1997, **94**:7251–6.

36. Takemaru K, Harashima S, Ueda H, Hirose S: **Yeast coactivator MBF1 mediates GCN4-dependent transcriptional activation.** *Mol Cell Biol* 1998, **18**:4971–6.

37. Tsuda K, Yamazaki K-I: **Structure and expression analysis of three subtypes of Arabidopsis MBF1 genes.** *Biochim Biophys Acta* 2004, **1680**:1–10.

38. Suzuki N, Bajad S, Shuman J, Shulaev V, Mittler R: **The transcriptional co-activator MBF1c is a key regulator of thermotolerance in Arabidopsis thaliana.** *J Biol Chem* 2008, **283**:9269–75.

39. López-Victorio CJ, Velez-delValle C, Beltrán-Langarica A, Kuri-Harcuch W: **EDF-1 downregulates the CaM/Cn/NFAT signaling pathway during adipogenesis.** *Biochem Biophys Res Commun* 2013, **432**:146–51.

40. Liu D, Wang Y, Ye Y, Yin G, Chen L: **Distinct molecular basis for endothelial differentiation: gene expression profiles of human mesenchymal stem cells versus umbilical vein endothelial cells.** *Cell Immunol* 2014, **289**:7–14.

41. Ying S-H, Ji X-P, Wang X-X, Feng M-G, Keyhani NO: **The transcriptional co-activator multiprotein bridging factor 1 from the fungal insect pathogen, Beauveria bassiana, mediates regulation of hyphal morphogenesis, stress tolerance and virulence.** *Environ Microbiol* 2014, **16**:1879–97.

42. Mendes-Pereira AM, Sims D, Dexter T, Fenwick K, Assiotis I, Kozarewa I, Mitsopoulos C, Hakas J, Zvelebil M, Lord CJ, Ashworth A: **Genome-wide functional screen identifies a compendium of genes affecting sensitivity to tamoxifen.** *Proc Natl Acad Sci U S A* 2012, **109**:2730–5.

43. Mihály Z, Kormos M, Lánczky A, Dank M, Budczies J, Szász MA, Győrffy B: **A meta-analysis of gene expression-based biomarkers predicting outcome after tamoxifen treatment in breast cancer.** *Breast Cancer Res Treat* 2013, **140**:219–32.

44. Culbertson MR, Gaber RF, Cummins CM: **Frameshift suppression in Saccharomyces cerevisiae. V. Isolation and genetic properties of nongroup-specific suppressors.** *Genetics* 1982, **102**:361–78.

45. Hendrick JL, Wilson PG, Edelman II, Sandbaken MG, Ursic D, Culbertson MR: **Yeast frameshift suppressor mutations in the genes coding for transcription factor Mbf1p and ribosomal protein S3: evidence for autoregulation of S3 synthesis.** *Genetics* 2001, **157**:1141–58.

46. Blombach F, Launay H, Snijders AP, Zorraquino V, Wu H, de Koning B, Brouns SJJ, Ettema T, Camilloni C, Cavalli A, Vendruscolo M, Dickman MJ, Cabrita LD, La Teana A, Benelli D, Londei P, Christodoulou J, van der Oost J: **Archaeal MBF1 binds to 30S and 70S ribosomes via its helix-turn-helix domain.** *Biochem J* 2014, **462**:373–84.

47. Schelert J, Dixit V, Hoang V, Simbahan J, Drozda M, Blum P: **Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon Sulfolobus solfataricus by use of gene disruption.** *J Bacteriol* 2004, **186**:427–37.

48. Albers S-V, Driessen AJM: **Conditions for gene disruption by homologous recombination of exogenous DNA into the Sulfolobus solfataricus genome.** *Archaea* 2008, **2**:145–9.

49. Zaparty M, Esser D, Gertig S, Haferkamp P, Kouril T, Manica A, Pham TK, Reimann J, Schreiber K, Sierocinski P, Teichmann D, van Wolferen M, von Jan M, Wieloch P, Albers S V, Driessen AJM, Klenk H-P, Schleper C, Schomburg D, van der Oost J, Wright PC, Siebers B: "**Hot standards" for the thermoacidophilic archaeon Sulfolobus solfataricus.** *Extremophiles* 2010, **14**:119–42.

50. Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning: A Laboratory Manual.* Second. New York: Cold Spring Harbor Laboratory Press; 1989.

51. Buck M, Connick M, Ames BN: **Complete analysis of tRNA-modified nucleosides by high-performance liquid chromatography: the 29 modified nucleosides of Salmonella typhimurium and Escherichia coli tRNA.** *Anal Biochem* 1983, **129**:1–13.

52. Droogmans L, Roovers M, Bujnicki JM, Tricot C, Hartsch T, Stalon V, Grosjean H: **Cloning and characterization of tRNA (m1A58) methyltransferase (TrmI) from Thermus thermophilus HB27, a protein required for cell growth at extreme temperatures.** *Nucleic Acids Res* 2003, **31**:2148–56.

53. Roovers M, Kaminska KH, Tkaczuk KL, Gigot D, Droogmans L, Bujnicki JM: **The YqfN protein of Bacillus subtilis is the tRNA: m1A22 methyltransferase (TrmK).** *Nucleic Acids Res* 2008, **36**:3252–62.

54. Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R: **A single-base resolution map of an archaeal transcriptome.** *Genome Res* 2010, **20**:133–41.

55. Wu H, Sun L, Blombach F, Brouns SJJ, Snijders APL, Lorenzen K, van den Heuvel RHH, Heck AJR, Fu S, Li X, Zhang XC, Rao Z, van der Oost J: **Structure of the ribosome associating GTPase HflX.** *Proteins* 2010, **78**:705–13.

56. Bell SD, Brinkman AB, van der Oost J, Jackson SP: **The archaeal TFIIEalpha homologue facilitates transcription initiation by enhancing TATA-box recognition.** *EMBO Rep* 2001, **2**:133–8.

57. Costanzo MC, Mueller PP, Strick CA, Fox TD: **Primary structure of wild-type and mutant alleles of the PET494 gene of Saccharomyces cerevisiae.** *Mol Gen Genet* 1986, **202**:294–301.

58. Cammarano P, Teichner A, Londei P, Acca M, Nicolaus B, Sanz JL, Amils R: **Insensitivity of archaebacterial ribosomes to protein synthesis inhibitors. Evolutionary implications.** *EMBO J* 1985, **4**:811–6.

59. Londei P, Altamura S, Sanz JL, Amils R: **Aminoglycoside-induced mistranslation in thermophilic archaebacteria.** *Mol Gen Genet* 1988, **214**:48–54.

60. Grosjean H: **Nucleic acids are not boring long polymers of only four types of nucleotides: a guided tour**. In *DNA RNA Modif Enzym Struct Mech Funct Evol*. Orsay: Landes Bioscience; 2009:1–18.

61. Ogle JM, Ramakrishnan V: **Structural insights into translational fidelity.** *Annu Rev Biochem* 2005, **74**:129–77.

62. Urbonavicius J, Qian Q, Durand JM, Hagervall TG, Björk GR: **Improvement of reading frame maintenance is a common function for several tRNA modifications.** *EMBO J* 2001, **20**:4863–73.

63. Arce DP, Godoy AV, Tsuda K, Yamazaki K-I, Valle EM, Iglesias MJ, Di Mauro MF, Casalongué CA: **The analysis of an Arabidopsis triple knock-down mutant reveals functions for MBF1 genes under oxidative stress conditions.** *J Plant Physiol* 2010, **167**:194–200.

64. Schönig B, Vogel S, Tudzynski B: **Cpc1 mediates cross-pathway control independently of Mbf1 in Fusarium fujikuroi.** *Fungal Genet Biol* 2009, **46**:898–908.

65. Jordan IK, Rogozin IB, Wolf YI, Koonin E V: **Essential genes are more evolutionarily conserved than are nonessential genes in bacteria.** *Genome Res* 2002, **12**:962–8.

66. Di Mauro MF, Iglesias MJ, Arce DP, Valle EM, Arnold RB, Tsuda K, Yamazaki K, Casalongué CA, Godoy AV: **MBF1s regulate ABA-dependent germination of Arabidopsis seeds.** *Plant Signal Behav* 2012, **7**:188–92.

67. Zegzouti H, Jones B, Frasse P, Marty C, Maitre B, Latche A, Pech J-C, Bouzayen M: **Ethylene-regulated gene expression in tomato fruit: characterization of novel ethylene-responsive and ripening-related genes isolated by differential display**. *Plant J* 1999, **18**:589–600.

68. Suzuki N, Rizhsky L, Liang H, Shuman J, Shulaev V, Mittler R: **Enhanced tolerance to environmental stress in transgenic plants expressing the transcriptional coactivator multiprotein bridging factor 1c.** *Plant Physiol* 2005, **139**:1313–22.

69. Kim M-J, Lim G-H, Kim E-S, Ko C-B, Yang K-Y, Jeong J-A, Lee M-C, Kim CS: **Abiotic and biotic stress tolerance in Arabidopsis overexpressing the multiprotein bridging factor 1a (MBF1a) transcriptional coactivator gene.** *Biochem Biophys Res Commun* 2007, **354**:440–6.

70. Aglawe SB, Fakrudin B, Patole CB, Bhairappanavar SB, Koti R V, Krishnaraj PU: **Quantitative RT-PCR analysis of 20 transcription factor genes of MADS, ARF, HAP2, MBF and HB families in moisture stressed shoot and root tissues of sorghum.** *Physiol Mol Biol Plants* 2012, **18**:287–300.

71. Bechtold U, Albihlal WS, Lawson T, Fryer MJ, Sparrow PAC, Richard F, Persad R, Bowden L, Hickman R, Martin C, Beynon JL, Buchanan-Wollaston V, Baker NR, Morison JIL, Schöffl F, Ott S, Mullineaux PM: **Arabidopsis HEAT SHOCK TRANSCRIPTION FACTORA1b overexpression enhances water productivity, resistance to drought, and infection.** *J Exp Bot* 2013, **64**:3467–81.

72. Hommel M, Khalil-Ahmad Q, Jaimes-Miranda F, Mila I, Pouzet C, Latché A, Pech JC, Bouzayen M, Regad F: **Over-expression of a chimeric gene of the transcriptional co-activator MBF1 fused to the EAR repressor motif causes developmental alteration in Arabidopsis and tomato**. *Plant Sci* 2008, **175**:168–177.

73. Li S, Zhou X, Chen L, Huang W, Yu D: **Functional characterization of Arabidopsis thaliana WRKY39 in heat stress.** *Mol Cells* 2010, **29**:475–83.

74. Liu H, Chen R, Zhang Q, Peng H, Wang K: **Differential gene expression profile from haematopoietic tissue stem cells of red claw crayfish, Cherax quadricarinatus, in response to WSSV infection.** *Dev Comp Immunol* 2011, **35**:716–24.

75. Qin D, Wang F, Geng X, Zhang L, Yao Y, Ni Z, Peng H, Sun Q: **Overexpression of heat stress-responsive TaMBF1c, a wheat (Triticum aestivum L.) Multiprotein Bridging Factor, confers heat tolerance in both yeast and rice.** *Plant Mol Biol* 2014.

76. Guo W-L, Chen R-G, Du X-H, Zhang Z, Yin Y-X, Gong Z-H, Wang G-Y: **Reduced tolerance to abiotic stress in transgenic Arabidopsis overexpressing a Capsicum annuum multiprotein bridging factor 1.** *BMC Plant Biol* 2014, **14**:138.

77. Baloglu MC, Inal B, Kavas M, Unver T: **Diverse expression pattern of wheat transcription factors against abiotic stresses in wheat species.** *Gene* 2014, **550**:117–22.

78. Rospert S, Rakwalska M, Dubaquié Y: **Polypeptide chain termination and stop codon readthrough on eukaryotic ribosomes.** *Rev Physiol Biochem Pharmacol* 2005, **155**:1–30.

79. Agris PF: **Decoding the genome: a modified view.** *Nucleic Acids Res* 2004, **32**:223–38.

80. Grosjean H, de Crécy-Lagard V, Marck C: **Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes.** *FEBS Lett* 2010, **584**:252–64.

81. Klass DM, Scheibe M, Butter F, Hogan GJ, Mann M, Brown PO: **Quantitative proteomic analysis reveals concurrent RNA-protein interactions and identifies new RNA-binding proteins in Saccharomyces cerevisiae.** *Genome Res* 2013, **23**:1028–38.

82. Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, Bonneau R, Selbach M, Dieterich C, Landthaler M: **The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts.** *Mol Cell* 2012, **46**:674–90.

83. Kwon SC, Yi H, Eichelbaum K, Föhr S, Fischer B, You KT, Castello A, Krijgsveld J, Hentze MW, Kim VN: **The RNA-binding protein repertoire of embryonic stem cells.** *Nat Struct Mol Biol* 2013, **20**:1122–30.

5

Archaeal MBF1 binds to 30S and 70S ribosomes via its helix-turn-helix domain

Fabian Blombach, Helene Launay, Ambrosius P.L. Snijders, Violeta Zorraquinoa, Hao Wu, Bart de Koning, Stan J.J. Brouns, Thijs J. G. Ettema, Carlo Camilloni, Andrea Cavalli, Michele Vendruscolo, Mark J. Dickman, Lisa D. Cabrita, Anna La Teana, Dario Benelli, Paola Londei, John Christodoulou and John van der Oost

Abstract

MBF1 (multi-protein bridging factor 1) is a protein containing a conserved HTH (helix-turn-helix) domain in both eukaryotes and archaea. Eukaryotic MBF1 has been reported to function as a transcriptional co-activator that physically bridges transcription regulators with the core transcription initiation machinery of RNA polymerase II. In addition, MBF1 has been found to be associated with polyadenylated mRNA in yeast as well as mammalian cells. aMBF1 (archaeal MBF1) is very well conserved among most archaeal lineages; however, its function has so far remained elusive. To address this, we have conducted a molecular characterization of this aMBF1. Affinity purification of interacting proteins indicates that aMBF1 binds to ribosomal subunits. On sucrose density gradients, aMBF1 co-fractionates with free 30S ribosomal subunits as well as with 70S ribosomes engaged in translation. Binding of aMBF1 to ribosomes does not inhibit translation. Using NMR spectroscopy, we show that aMBF1 contains a long intrinsically disordered linker connecting the predicted N-terminal zinc-ribbon domain with the C-terminal HTH domain. The HTH domain, which is conserved in all archaeal and eukaryotic MBF1 homologues, is directly involved in the association of aMBF1 with ribosomes. The disordered linker of the ribosome-bound aMBF1 provides the N-terminal domain with high flexibility in the aMBF1-ribosome complex. Overall, our findings suggest a role for aMBF1 in the archaeal translation process.

Introduction

Archaea and eukaryotes share a common set of proteins involved in genetic information processing (transcription, translation and replication), including several proteins containing HTH (helix-turn-helix) domains [1–3]. Most of these proteins carry out functions within the core transcription machinery in both archaea and eukaryotes. This includes the eukaryotic protein MBF1 (multi-protein bridging factor 1) that has been shown to act as transcriptional co-activator, transmitting the signal from eukaryote-specific transcription factors to the core transcription machinery by physically bridging these factors with the TBP (TATA-box-binding protein) via the HTH domain of MBF1 [4–7].

Besides its characterized function as transcription co-activator, previous studies suggest that eukaryotic MBF1 might be a moonlighting protein. In yeast, a frameshift mutation in the *mbf1* sequence as well as deletion of the entire *mbf1* gene have been shown to alter the rate of ribosomal frameshifting as well as the sensitivity of the strains to aminoglycoside antibiotics including paromomycin [8–10]. In addition, yeast MBF1 has been recently shown to co-purify with Pab1 [poly(A)-binding protein 1]. The interaction is sensitive to RNase treatment, suggesting that MBF1 is associated with polyadenylated mRNA [11]. Furthermore, yeast MBF1 binds directly to RNA via the less-conserved N-terminal domain [11]. Similarly, human MBF1 was also identified as an mRNA-binding protein in embryonic stem cells and HEK (human embryonic kidney)-293 cells [12, 13].

When the first archaeal genomes became available, aMBF1 (archaeal MBF1) orthologues were identified on the basis of sequence homology encompassing the HTH domain [2]; however, aMBF1 has remained functionally uncharacterized ever since. The evolutionary conservation of TBP across all eukaryotes and archaea might suggest that TBP and aMBF1 also interact in archaea [5]. However, the fact that experimental investigations using chimaeric constructs bearing HTH domains originating from different archaeal species were unable to functionally replace the endogenous HTH domain of MBF1 in yeast hints at a functional difference between archaeal and eukaryotic MBF1 [14]. This is corroborated by the observation that the well-conserved C-terminal HTH domain of archaeal and eukaryotic MBF1 is linked to an N-terminal MBF1-specific domain in eukaryotic MBF1, and to a distinct N-terminal domain (predicted zinc-ribbon fold) in aMBF1 orthologues [2, 15].

In the present study, we make an important step forward towards elucidating the function of aMBF1 by presenting a biochemical and functional characterization of an aMBF1 orthologue. The results show that aMBF1 from *Sulfolobus solfataricus* binds to the small ribosomal subunit during translation via its conserved HTH motif. These results suggest an unexpected physiological function for aMBF1 linked to translation.

Accession numbers

Co-ordinates and structure factors of the aMBF1 (archaeal multi-protein bridging factor 1) helix-turn-helix domain have been deposited in the PDB under code 2MEZ. The NMR assignment data have been deposited in the BMRB under accession number 19028.

Experimentals

Molecular cloning of mfb1 and in vitro translation templates

The mbf1 gene from *S. solfataricus* P2 (GeneID: 1455418) [16] was PCR-amplified from genomic DNA using forward primer 5'-GCGCGCATATGCAAGCTAATAGTGAAGAATAC-3' and reverse primer 5'-GCGCGCTCGAGCTTCTTTCCCTCTTTAATATTTACC-3' and cloned into vector pET26b via NdeI and XhoI restriction sites (underlined), resulting in plasmid pWUR298. For molecular cloning of the isolated N-terminal domain of aMBF1 (aMBF1-N) encompassing amino acids 1-58, the reverse primer was exchanged for 5'-GCGCGG-CCCCTCGAGCTTACGTGTTTCGCTTTTCTTAC-3' (resulting in plasmid pWUR557). For molecular cloning of the isolated C-terminal domain of aMBF1 (aMBF1-C) encompassing amino acids 57-165, the forward primer was exchanged for 5'-GCGCGGC-CCATATGCGTAAGAAAGCCACTCTTAAACCACC-3' (resulting in plasmid pWUR300). As no complementary stop codons were included in the reverse primers, all three proteins were designed to have a C-terminal His_6-tag.

Heterologous expression and purification of aMBF1

Plasmids pWUR298, pWUR300 and pWUR557 were transformed into *Escherichia coli* Rosetta (DE3) cells (Novagen) and heterologous expression was carried out using standard procedures. Full-length aMBF1, aMBF1-N and aMBF1-C were produced as follows. Cells were resuspended in buffer TK300 (20 mM Tris/HCl, pH 7.5, 300 mM KCl and 1 mM DTT) and passed through a French pressure cell (Aminco) three times at 16 000 psi (1 psi = 6.9 kPa). Cell debris was removed by centrifugation at 37 000 g for 30 min at 4 °C. The cell-free extract was incubated at 75 °C for 15 min and the heat-unstable proteins from the expression host were then removed by centrifugation. The resulting supernatant was then purified further by nickel-affinity chromatography using His-Select Nickel affinity gel (Sigma) and size-exclusion chromatography using a HiLoad 16/60 Superdex 75 column (GE Healthcare) equilibrated in TK300. Elution fractions containing aMBF1 were combined and the buffer was exchanged by ultrafiltration to 20 mM Tris/HCl (pH 7.2), 50 mM NH_4Cl, 10 mM magnesium acetate, 1 mM DTT and 10% (v/v) glycerol. Aliquots of the proteins were snap-frozen in liquid nitrogen and stored at -80 °C. All proteins were quantified on the basis of their absorption at 280 nm using calculated molar absorption coefficients for the respective protein [17].

For isotopic labelling of recombinant proteins for NMR studies, the heterologous expression was carried out in M9 medium (6.78 g/l Na_2HPO_4, 3 g/l KH_2PO_4, 0.5 g/l NaCl, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 1× MEM vitamin solution), with $^{15}NH_4Cl$ (1 g/l) and 0.4% [^{13}C]glucose (Cambridge Isotope laboratories) as the sole sources of nitrogen and carbon respectively.

Immunodetection of aMBF1, aMBF1-C, aMBF1-N and Alba

Rabbit antiserum against recombinant aMBF1-C was produced at Eurogentec. Antiserum from the final bleed was purified over Protein A-agarose (Sigma-Aldrich), and antibodies were allowed tp react with digoxigenin-3-*O*-methylcarbonyl-ε-aminocaproic acid-*N*-hydroxy-succinimide ester (Roche) at a 1:10 molar ratio according to the manufacturer's protocol. Immunodetection of aMBF1 and aMBF1-C was performed as described previously for *S. solfataricus* HflX [18] using digoxigenin-labelled primary antibodies and alkaline-phosphatase-conjugated anti-digoxigenin Fab fragments (Roche) as secondary antibodies. For immunodetection of aMBF1–N, samples were resolved by Tris/Tricine SDS/PAGE [19]. Immunodetection was performed with anti-His_6 antibody (Roche) following the manufacturer's

recommendation. For the immunodetection of Alba, a 1:3000 dilution of Alba antiserum (a gift of Malcolm White, University of St. Andrews, St. Andrews, U.K.) was used with alkaline-phophatase-conjugated donkey anti-(goat) IgG (Promega) (1:10000 dilution). For the detection, CDP-Star® reagent (NEB) was used and signals were captured on Kodak Biomax films.

Cell growth of S. solfataricus strains

S. solfataricus strain P2 was obtained from DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen; Braunschweig, Germany), and strain PBL2025 was a gift from Paul Blum (University of Nebraska-Lincoln, Lincoln, NE, U.S.A.). The generation of the Δ*mbf1* strain was achieved generally according to [20] (exact details available from B.d.K. on request). For the preparation of cell lysates for *in vitro* translation experiments and ribosome co-purification, cell cultures were grown in shake flasks at 80 °C on modified Brock medium [21] supplemented with 0.4% sucrose and 0.1% tryptone to an OD_{600} of 0.7-1.2 and 0.4 respectively. For all other cell cultures of *S. solfataricus* P2 strain, modified Brock medium was supplemented with 0.3% glucose. For isotope labelling of proteins, [14]NH_4Cl was replaced by [15]NH_4Cl (Cambridge Isotope Laboratories) as described previously [22].

Affinity purification and identification of interacting proteins

Recombinant aMBF1 was purified as described above, except that Tris was replaced by triethanolamine and DTT was omitted. A 600 µg amount of aMBF1 or His_6 peptide (Innovagen) was coupled to 200 µl of pre-washed *N*-hydroxysuccinimide-activated Sepharose resin (GE Healthcare) according to manufacturer's protocol. *S. solfataricus* S30 cell lysates were produced as described previously [23] in 20 mM Tris/HCl (pH 7.4), 50 mM KCl, 10 mM magnesium acetate, 5 mM $CaCl_2$, 20% (v/v) glycerol, 0.5% Nonidet P40 and 1 mM DTT, and were diluted to 5 mg/ml. Then, 25 mg of lysate was treated with 300 units of micrococcal nuclease (Fermentas) at 30 °C for 10 min to degrade nucleic acids that can lead to bridging effects and overall increased background [24].

Volumes of 25 µl of aMBF1-coupled beads were added to 2 ml of further diluted cell lysate (0.5-4 mg of protein/ml) and incubated for 2 h at 4 °C. The cell lysates were then transferred to a spin column, and beads were washed four times with 500 µl of 20 mM Tris/HCl (pH 7.4), 105 mM KCl, 20% (v/v) glycerol and 0.1% Nonidet P40. Proteins were eluted with 100 µl of 2× SDS/PAGE sample buffer. As a negative control, the same experiment was performed using His_6 peptide-coupled beads and ^{15}N-labelled cell lysate. Equal volumes of [14]N- and ^{15}N-labelled elution fractions from aMBF1-coupled or His_6 peptide-coupled beads respectively were mixed and resolved by SDS/PAGE [8% (30:2) gel].

Gel lanes were cut into eight blocks, destained and treated with trypsin for in-gel digestion. Peptides were analysed by LC-MS/MS using an Ultimate 3000 HPLC and a MaXis UHR-Q-TOF (ultra-high-resolution quadrupole time-of-flight) tandem mass spectrometer. All MS data were acquired in profile mode. Bruker .baf files were converted to mzXML files by CompassExport. Mascot Distiller then used mzXML files for peak detection and quantification. Mascot version 2.1 was used to search the peak lists against a database containing the *S. solfataricus* P2 proteome sequences in fasta format concatenated with a randomized version of the same database. The Mascot Distiller precursor quantification protocol was performed using a ^{15}N metabolic incorporation percentage of 98%.

In vitro translation and formaldehyde cross-linking

For the template used in the *in vitro* translation reactions, a synthetic gene was generated by GENEART (Germany) (pWUR560) based on the *orf104* gene [25] (encoding ribosomal protein L30ae) that has been well-characterized in the *S. solfataricus in vitro* translation systems [23, 26, 27] (Supplementary Figure S1). *In vitro* transcription with the Megascript T7 kit (Ambion) was carried out on SacI-linearized plasmid pWUR560 and transcripts were purified on RNeasy spin columns (Qiagen).

The a*IF6* (archaeal translation initiation factor 6) ORF was PCR-amplified from a pET-based expression plasmid [26] using forward primer 5'-GCGCGC<u>GGTACC</u>GCCTAATATGAGGTG-AAATGTAA**ATG**AATCTGCAAAGGTTATCAGTTTTTGG-3' and reverse primer 5'-GCG-CGC<u>TCTAGA</u>TCATTCACCTAATGCTTTTTGAATTC-3' (start codon in bold) and cloned via KpnI and XbaI restriction sites (underlined). The forward primer contains a leader sequence identical to that found in the *orf104* template [27].The a*IF6* mRNA was generated by *in vitro* transcription similar as described above.

The preparation of *S. solfataricus* cell lysates and *in vitro* translation reactions as well as cross-linking of 70S ribosomes were conducted as described previously [18], with a pre-incubation of the cell lysates for 10 min at 73°C to uncharge all ribosomes.

To localize aMBF1 in the cell lysates, samples were loaded on 10.5 ml 10-30% sucrose gradients in buffer A (20 mM Tris/HCl, pH 7.4, 40 mM NH_4Cl, 10 mM magnesium acetate and 1 mM DTT) and centrifuged for 4 h at 220 000 *g* in a TST41.14 rotor (Kontron Instruments) at 4°C. Fractions of 500 µl were subjected to sodium deoxycholate/trichloroacetic acid precipitation. Protein pellets were resuspended in 20 µl 3× SDS/PAGE sample buffer, resolved by Bis-Tris SDS/PAGE [8% (30:2) gel] using Mes running buffer, and aMBF1-containing fractions were identified by immunodetection.

For detection of newly synthesized proteins during *in vitro* translation, the assays were modified as follows. The lysates were supplemented with micrococcal nuclease and 1 mM $CaCl_2$ for the degradation of endogenous mRNA for 20 min at 20 °C. The reaction was stopped by the addition of 2 mM sodium EGTA (pH 7.4). A 25 µl volume of *in vitro* translation assay mixture contained 17.3 µCi L-[^{35}S]methionine (specific radioactivity >1000 Ci/mmol) (PerkinElmer) and 700 ng of mRNA. After 50 min of incubation at 72 °C, the reaction was stopped by the addition of 0.5 volume of 3× SDS/PAGE sample buffer. Then, 15 µl of each sample was separated by Tris/Tricine SDS/PAGE (15% gel). Gels were dried and exposed to autoradiography.

Determination of aMBF1 expression during different growth phases

Cells were grown as described above. Aliquots of 50 µl were withdrawn at the given time points. Cells were harvested by centrifugation at 4 °C. After resuspension in buffer A, cells were lysed by sonication. After removal of cell debris by centrifugation at 16 100 *g* for 40 min at 4 °C, the protein concentration of the lysates was determined using the Bradford assay and adjusted to 4 mg/ml for all lysates. Lysates were supplemented with 0.5 volumes of 3× SDS/PAGE sample buffer and proteins were resolved by Bis-Tris SDS/PAGE using Mes running buffer. Immunodetection of aMBF1 and Alba was carried out as described above.

Isolation of ribosomes and ribosomal subunits

Ribosomes were purified in buffer A as described previously [23, 28] and quantified on the basis of absorbance measurements at 260 nm (1 A_{260} unit corresponds to 40 pmol). Isolated 30S ribosomal subunits were recovered from sucrose gradients by ultrafiltration. Each A_{260} unit was assumed to correspond to 70 pmol of 30S ribosomal subunit on the basis of a concentration of 70 µg per A_{260} unit and a calculated molecular mass of ~1 MDa on the basis of the genome sequence.

Ribosome binding assays

For ribosome-binding assays, 100 pmol of recombinant aMBF1 or aMBF1-C was incubated with 100 pmol of ribosomes in 100 µl of buffer A for 30 min on ice. Samples were loaded on 10.5 ml 10-30% sucrose gradients in buffer A and processed further as described above.

NMR spectroscopy

All aMBF1 and ribosome preparations were buffer-exchanged to 10 mM Hepes/KOH (pH 7.3), 40 mM NH_4Cl, 10 mM $MgCl_2$, 1 mM DTT. A 190 µM solution of ^{15}N-labelled aMBF1 was supplemented with 10% $^{2}H_2O$ and 0.033% 3-(trimethylsilyl)propane-1-sulfonic acid. NMR spectra were recorded on a Bruker Avance III 700 MHz spectrometer equipped with a TXI cryogenic probe. The Fast-HSQC pulse sequence [29] was used for heteronuclear spectra of aMBF1 constructs, with a spectral width of 30 p.p.m. in the indirect dimension and 128 complex points acquired. The direct dimension was recorded with 2048 complex points and a spectral width of 12 p.p.m. All spectra were recorded at 25 °C. ^{15}N-heteronuclear relaxation rates were measured using standard procedures [30], and were analysed using the Lipari-Szabo formalism [31]. PFG (pulsed-field gradient) diffusion experiments were performed using the stimulated echo-PFG [32] and the heteronuclear Xste-PFG sequence [33] with a bipolar 1 ms square gradient for isolated protein and 2 ms square gradient for ribosomal subunits with the gradient strength varying from 5% to 95% of the maximum strength (0.563 T·m^{-1}) together with a 200 ms diffusion delay. The PFG diffusion data were analyzed using the Stejskal-Tanner equation [32]. The diffusion coefficient obtained from the PFG data and the rotational correlation time obtained from ^{15}N-relaxation data were used to approximate an apparent molecular mass using the Stokes-Einstein and Stokes-Einstein-Debye equations respectively, using a sphere model and a 2 Å (0.2 nm) hydratation layer [34]. All spectra were processed in nmrPipe [35]. The backbone resonance assignment of the C-terminal aMBF1 resonances was performed via standard triple-resonance (HNCACB, HNCOCACB, HNCO and HNCACO) experiments, processed in nmrPipe and analysed using CCPN software [36]. The secondary-structure populations were calculated from the chemical shifts using the δ2D method [37]. A structural model was determined using CA, CB, CO, N and HN chemical shift with the CHESHIRE fragment replacement protocol [38]. The ensemble has been generated using the chemical shifts as restraints in a replica-averaged MD simulation [39, 40] using the Amber03W force field [41] (see Supplementary text for details). Residual dipolar couplings were measured using the difference in the ^{15}N-coupling of HSQC-IPAP (in-phase/antiphase) spectra [42] recorded on an anisotropic sample and a sample of MBF1 partially aligned in 5% penta(ethylene glycol) dodecyl ether/hexanol [43]. The correlation spectra of [^{15}N]aMBF1 in complex with archaeal ribosomal subunits were detected via a SOFAST-HMQC acquisition scheme [44].

Results and discussion

### Heterologous production of S. solfataricus aMBF1

An aMBF1 orthologue was identified in *S. solfataricus*, comprising the archaea-specific N-terminal putative zinc-ribbon domain and the C-terminal HTH domain [2, 15]. Recombinant *S. solfataricus* aMBF1 was successfully produced as His-tag fusion protein in *E. coli* Rosetta

Table 1. Proteins identified in the aMBF1 affinity purification from S. solfataricus strain P2 cell lysate. Mascot probability scores are calculated as $-10 \times \log_{10}(P)$, where *P* is the probability that the match was a random event. Specific interactors (sp.) were defined as those proteins being at least 10-fold enriched in the aMBF1 affinity purification compared with the control experiment. non-sp. indicates that the enrichment was below threshold, n.q. indicates that a quantification was not possible.

Protein name	Identifier	Mascot score	Specificity
Proteins of the 30S ribosomal subunit			
S2p	15897033	206	sp.
S3p	15897617	47	n.q.
S4p	15897040	116	sp.
S5p	15897604	306	sp.
S7p	15897165	94	n.q.
S8p	15897609	76	sp.
S9p	15897035	149	sp.
S10p	15897163	71	non-sp.
S11p	15897039	67	n.q.
S12p	15897167	104	n.q.
S13p	15897041	84	n.q.
S17p	15897615	32	n.q.
S19p	15897619	34	n.q.
S4e	15897612	130	sp.
S6e	15897344	269	sp.
S8e	15897114	80	n.q.
S19e	15897290	48	sp.
S24e	15897365	174	sp.
S25e	15897354	48	n.q.
Proteins of the 50S ribosomal subunit			
L1p	15897279	51	n.q.
L4p	15897622	140	sp.
L5p	15897611	31	n.q.
L6p	15897608	99	non-sp.
L11p	15897280	38	n.q.
L18p	15897605	30	n.q.
L30	15897603	39	n.q.
Non-ribosomal proteins			
Thermosome *β*	15897225	69	non-sp.
Alba	15897841	91	non-sp.
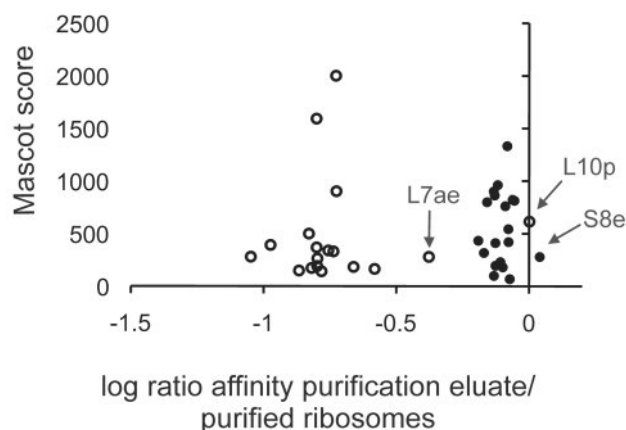
Figure 1. Relative enrichment of individual ribosomal proteins in the aMBF1 affinity purification experiment from S. solfataricus strain P2 cell lysate. Half of the ^{15}N-labelled eluate from an aMBF1 affinity purification experiment was mixed with 4 pmol of purified ^{14}N-labelled ribosomes. Proteins of the 30S ribosomal subunits are represented by closed circles, and proteins of the 50S ribosomal subunits by open circles. Mascot probability scores are calculated as $-10 \times \log 10(P)$ where P is the probability that the match was a random event.

(DE3) cells, but expression in *E. coli* strain BL21(DE3) was found to yield an apparently uniformly truncated product; analysis of this truncated product by tryptic digest and MS identified peptides covering the entire C-terminal domain from Lys[59] onwards (results not shown).

Constructs for the expression of the isolated N-terminal (amino acids 1-58) and C-terminal (amino acids 57-165) domains (aMBF1-N and aMBF1-C respectively) were then designed according to these partial proteolysis data and the mutant proteins were stably produced in *E. coli* Rosetta (DE3) cells (Supplementary Figure S2). MS determined the mass of aMBF1-C to be 13 161 ± 1 Da, corresponding precisely to the calculated mass of the protein with an N-terminal proteolytic processing of the first three amino acids (the start methionine residue, Arg[57] and Lys[58]; numbering according to full-length MBF1) (Supplementary Figure S3). N-terminal processing of aMBF1-C was also observed by NMR spectroscopy (see below).

Affinity purification of interacting proteins

High-throughput screens for protein complexes by tandem affinity purification have not previously identified any complex that included MBF1 in yeast [45]. We screened for possible protein interactors of aMBF1 in an attempt to gain information on its possible physiological function. To identify transient protein-protein interactions in our screen, we combined a single-step affinity chromatography procedure with quantitative MS. *S. solfataricus* P2 cells were grown in [14]N- or ^{15}N-containing medium yielding isotopically labelled proteins. P2 cell lysate was mixed with immobilized aMBF1 and affinity-purified proteins were compared with a negative control experiment carried out using cell lysate with different isotopic labelling and immobilized His$_6$ peptide instead of aMBF1. The eluate fractions from the two different experiments were mixed, resolved by SDS/PAGE and subjected to MS, allowing the determination of the relative levels of each identified protein in the two experiments. Proteins that were enriched at least 10-fold using immobilized aMBF1 when compared with the control experiment were considered to be possible interactors of aMBF1. Intriguingly, the majority of

the potential interactors (nine of the ten) were proteins from the 30S ribosomal subunit, and, in addition, a single protein from the 50S ribosomal subunit could be confirmed to bind to aMBF1 (Table 1). Comparison between experiments using the reverse ^{14}N- and ^{15}N-labeling of *S. solfataricus* P2 cells confirmed the overall trend of the experiment, with a S.D. of ± 10% for the normalized ratios.

The affinity purification indicated binding of aMBF1 to ribosomes or specific ribosomal proteins. In order to investigate further whether specific ribosomal proteins were enriched during the aMBF1 affinity purification, we compared the level of individual ribosomal proteins in the aMBF1 affinity purification with those within purified *S. solfataricus* ribosomes. To do this, the ^{15}N-labelled affinity purification eluate was mixed with 4 pmol of ribosomes and analysed by MS. Overall, the results suggested a strong preference of aMBF1 for the 30S ribosomal subunit over the 50S ribosomal subunit (Figure 1). The relative proportions of ribosomal proteins of each individual ribosomal subunit were similar for both the affinity-purified material and the purified ribosomes, with three exceptions. Ribosomal proteins S8e, L10p, and L7ae were significantly overrepresented in the affinity-purification fractions compared with the purified ribosomes. L7ae has been shown to be part of not only the 50S, but also of the 30S ribosomal subunits in Archaea [46, 47] explaining why it was enriched to a similar extent as the other 30S ribosomal proteins. The data suggested that, in the experiment, intact 30S ribosomal subunits interacted directly with the immobilized aMBF1. There appeared to be no evidence to suggest that an additional extrinsic factor mediates this interaction nor that there was a subset of ribosomal proteins enriched that maps to a specific part of the ribosome. It should be noted that the experiment was carried out under buffer conditions in which the ribosomal particles are intact. The C-terminal HTH of yeast MBF1 has been shown to interact with the general transcription factor TBP in GST-pull-down assays with purified recombinant proteins [4, 5]. Interestingly, we did not detect TBP as possible interactor for aMBF1 in our screen.

Expression and cellular localization of aMBF1

The expression levels of aMBF1 during different growth phases of *S. solfataricus* were determined by immunodetection. aMBF1 was expressed during all growth phases, but expression was observed to be at its highest during exponential growth (Figure 2A). On the basis of a concentration calibration using recombinant aMBF1, the expression levels were calculated to range from 320 ± 60 ng of aMBF1/mg of cytosolic proteins in the exponential growth phase to 90 ± 30 ng of aMBF1/mg of cytosolic proteins in the stationary growth phase.

We also investigated the interaction of endogenous aMBF1 with ribosomal subunits by isolating ribosomes and testing co-purification of aMBF1. When ribosomes were pelleted by ultracentrifugation from an S30 extract prepared from cells in late-exponential growth phase, a significant fraction of aMBF1 was depleted from the S100 supernatant (Figure 2B), suggesting that it co-purifies with the ribosomal pellet. Further purification of the crude ribosome fraction on sucrose cushions under low (100 mM) and high (500 mM) salt conditions revealed that the association of MBF1 is salt-sensitive. Loading of crude ribosomes on sucrose density gradients revealed that aMBF1 co-migrated specifically with the 30S, but not with the 50S ribosomal subunit.

Another question was whether aMBF1 also associates with 70S ribosomes. 70S ribosomes of *S. solfataricus* are relatively unstable and the intact particle readily dissociates into 30S and 50S ribosomal subunits during sucrose density gradient centrifugation. The entire ribosomal complex can be reassembled by programming a cell lysate for translation in an *in vitro*

translation assay using the well-characterized *orf104* mRNA template (Supplementary Figure S1) and subsequently stabilizing the 70S ribosomes by chemical cross-linking [23]. This procedure can potentially also cross-link aMBF1 directly to the ribosomes. Under these conditions, only a minor fraction of free aMBF1 was found, and most aMBF1 co-migrated with 70S ribosomes and isolated 30S ribosomal subunits (Figure 2C). Similar results were obtained using a*IF6* mRNA as template, indicating that aMBF1-30S ribosomal subunit interaction is not limited to specific mRNAs (results not shown).

The cell lysates used in these *in vitro* translation assays were pre-incubated at high temperature before their use in the translation assay in order to unload the ribosomes from endogenous mRNA and to increase the specificity of the translation reaction for the
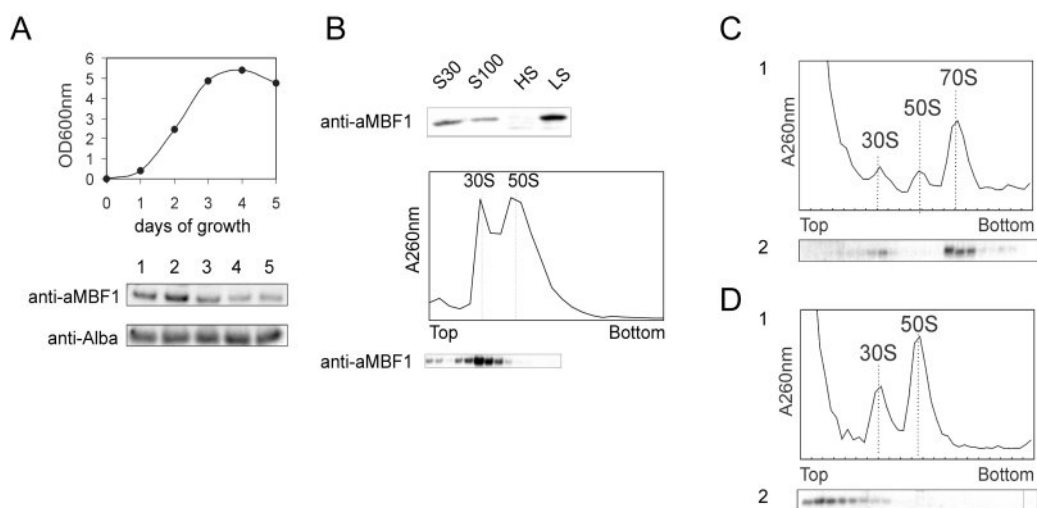
Figure 2. Expression of endogenous aMBF1 during different growth phases of S. solfataricus strain P2, co-purification with ribosomal subunits and binding of aMBF1 to 70S ribosomes in cross-linked cell lysates after activation for translation. (A) Expression of endogenous aMBF1 during different growth phases of *S. solfataricus* P2 as detected by immunodetection. Upper panel: the representative growth curve of a *S. solfataricus* P2 culture. Lower panel: immunodetection of aMBF1 and the abundant nucleic acid-binding protein Alba as control for samples taken at the given time points. Equal total soluble protein content was loaded on each lane. (B) Co-purification of aMBF1 with *S. solfataricus* P2 ribosomes. Upper panel: immunodetection of aMBF1 in different fractions obtained from ribosome isolation. S30 and S100 extracts (5% of total fraction respectively) and ribosomes purified on sucrose cushions under low (100 mM) (LS) and high (500 mM) (HS) salt conditions were tested. Lower panel: fractionation of 400 pmol crude ribosomes (on the basis of absorption at 260 nm) on a 15-30% sucrose gradient (30 mM KCl) to separate the ribosomal subunits and to verify the co-fractionation of aMBF1 with ribosomes. (C and D) Localization of endogenous aMBF1 in *S. solfataricus* P2 cell lysate. (C) Cell lysate (480 µg of protein content) was pre-incubated at 73 °C to unload the ribosomes from any endogenous mRNA and subsequently the cell lysate was programmed for translation with *orf104* mRNA as template and followed by formaldehyde cross-linking to stabilize 70S ribosomes. The cross-linked *in vitro* translation reaction was fractionated on a 10-30% sucrose density gradient, and immunodetection was used to localize aMBF1. (D) A control experiment with the same amount of cell lysate, but without subsequent *in vitro* translation and cross-linking. (C and D) Upper panels (1): absorption at 260 nm of the fractions after sucrose density gradient centrifugation to identify the position of the ribosomal subunits. Lower panels (2): immunodetection of aMBF1.

recombinant mRNA provided. We also tested whether aMBF1 is associated with ribosomal subunits in the pre-incubated lysates without the subsequent *in vitro* translation reaction and chemical cross-linking. Under these conditions, aMBF1 was spread over many fractions in the upper third of the gradient down to the position of the 30S ribosomal subunit. This possibly indicates that aMBF1 dissociated during centrifugation from the 30S ribosomal subunit (Figure 2D).

Similarly to endogenous aMBF1, recombinant aMBF1 co-migrated with 70S ribosomes and free 30S ribosomal subunit when 50 pmol of recombinant aMBF1 was added to a cell lysate prepared from a Δ*mbf1* strain before programming for translation (Figure 3A). The *S. solfataricus* Δ*mbf1* strain did not reveal any significant differences in growth kinetics from wild-type strain P2 under standard laboratory conditions (chapter 4). Immunodetection of recombinant aMBF1 gave the expected signals at ~20 kDa corresponding to aMBF1 monomer and a minor fraction at ~40 kDa. The reason for this apparent dimerization observed only for recombinant full-length aMBF1 is unknown, but it might involve (re-)oxidation of the cysteine residues in the N-terminal domain despite the presence of reducing agents throughout all experiments. Supplementation of up to 300 nM recombinant aMBF1 to cell lysate from *S. solfataricus* strain P2 (exceeding more than three times the concentration of endogenous
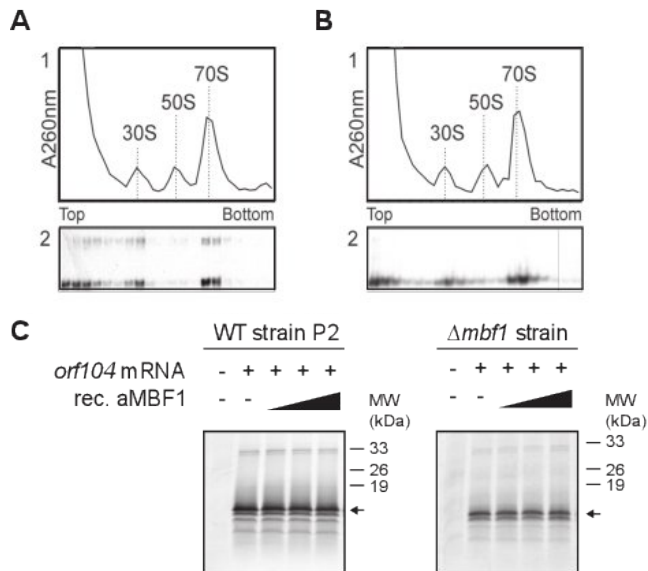
Figure 3. Localization of recombinant aMBF1 and aMBF1-C in cell lysates programmed for translation. (A and B) *In vitro* translation assays with cell lysate from the Δ*mbf1* strain (480 µg of protein content) were supplemented with 50 pmol recombinant aMBF1(A) and aMBF1-C (B). The lysates were pre-incubated, programmed for translation, chemically cross-linked and fractionated on sucrose density gradient as described in Figure 2. Upper panels (1): absorption profile at 260 nm of the sucrose density gradient was measured to identify the position of the ribosomal subunits. Lower panels (2): immunodetection of recombinant aMBF1 in the fractions obtained from the sucrose density gradient. (C) Effect of increasing amounts of recombinant aMBF1 (0, 3, 30 and 300 nM) on cell-free translation. *In vitro* translation was carried out in the presence of [^{35}S]methionine using lysate from *S. solfataricus* P2 or the Δ*mbf1* strain and production of ORF104 was detected by autoradiography after resolving the samples by Tris/Tricine SDS/PAGE. The first lane shows an *in vitro* translation reaction without an mRNA template. Molecular masses (MW) are indicated in kDa.

aMBF1 being present in the assay) or the Δ*mbf1* strain in *in vitro* translation assays did not affect protein synthesis from the *orf104* mRNA template (Figure 3C).

To investigate which domain of aMBF1 is responsible for the binding to ribosomes, we complemented a cell lysate of a Δ*mbf1* strain with the domain deletion variants aMBF1-C and aMBF1-N. Deletion of the predicted N-terminal zinc-ribbon domain of aMBF1 did not affect the ribosome interaction (Figure 3B), suggesting that the HTH domain (aMBF1-C) is sufficient to mediate the interaction with the 30S ribosomal subunit. In experiments carried out under the same conditions with aMBF1-N, the protein was not detectable (results not shown), probably due to degradation during the high-temperature *in vitro* translation reaction.

Reconstituted complexes

To determine whether aMBF1 directly interacts with the small ribosomal subunit, we purified ribosomes at different salt concentrations from cells grown to early stationary phase. Levels of co-purified endogenous aMBF1 were below the detection limit for ribosomes purified both under low (100 mM NH_4Cl) and high (500 mM NH_4Cl) salt conditions. Recombinant aMBF1 was added to the purified 30S subunits and the formation of the 30S-aMBF1 complex was observed using sucrose density gradient and immunodetection of aMBF1 in the 30S elution fractions. When incubated with both low- and high-salt-washed ribosomes, recombinant aMBF1 co-fractionated with 30S, indicating that no cofactor was apparently required to form the
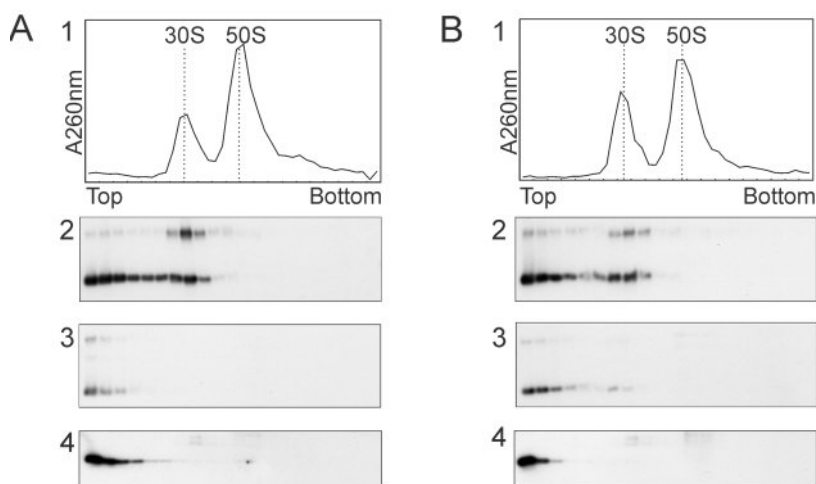
Figure 4. Binding of aMBF1, aMBF1-N and aMBF1-C to purified ribosomal subunits. Ribosomes (100 pmol) purified under low-salt (100 mM) (A) or high salt (500 mM) (B) conditions were incubated with 100 pmol of recombinant aMBF1 or mutant proteins and incubated on ice for 30 min. The samples were subsequently resolved on 10-30% sucrose gradients and SDS/PAGE, and immunodetection was used to localize aMBF1 and the mutant proteins. Panels (1): representative A_{260} profile with the position of the ribosomal subunits. Panels 2-4: immunodetection of binding assays for aMBF1 (panels 2), aMBF1-C (panels 3), and aMBF1-N (panels 4). The immunodetection of recombinant aMBF1 gave a second minor band approximately twice the apparent molecular mass of monomeric aMBF1. The reason for this apparent aMBF1 dimerization observed only for recombinant full-length aMBF1 is unknown, but it might involve (re-)oxidation of the cysteine residues in the N-terminal domain of aMBF1 despite the presence of reducing agents throughout all experiments.

complex (Figure 4). A significant proportion of aMBF1 eluted in the low-molecular-mass fractions of the sucrose gradient, suggesting that a fraction of the aMBF1-30S ribosomal subunit complex dissociated during the centrifugation (Figures 4A and 4B, panels 2). Surprisingly, when the experiment was repeated using only the C-terminal HTH domain of aMBF1 (aMBF1-C), only a small fraction of aMBF1-C remained bound to the 30S ribosomal subunits after sucrose density gradient centrifugation (Figure 4A and 4B, panels 3). This might be due to dissociation of the aMBF1-C small ribosomal subunits complex during sucrose density gradient centrifugation. The lower stability of the aMBF1-C small ribosomal subunits complex compared with that formed with full-length aMBF1 might indicate a role for the predicted zinc-ribbon domain or parts of the linker region to increase the complex affinity.

Structural analysis of isolated aMBF1 and aMBF1-C

Having established that aMBF1 binds directly to the small ribosomal subunits without cofactor, we aimed to characterize the structural and dynamic features of this interaction using NMR spectroscopy. A structural analysis of aMBF1 resulted in a well-resolved 2D ^{1}H-^{15}N HSQC spectrum in which almost all of the expected 169 resonances are clearly discerned: 91 well-dispersed resonances as well as ~78 in the central region of the spectrum (^{1}H frequency lying between 7.5 and 8.5 p.p.m.) were observed (Figure 5A).

Since the aMBF1-30S interaction seems to be mediated mainly by the C-terminal HTH domain, the NMR spectrum of aMBF1-C was compared with that of the full-length protein; its spectrum (Figure 5B) was found to overlay very well with the spectrum of the full-length aMBF1 (Figure 5C) indicating that the HTH and zinc-ribbon domains are structurally independent. The backbone resonances of aMBF1-C were assigned via standard triple-resonance strategies (see the Experimental section) to ~82% completion (Figure 6A). The secondary-structure populations calculated by the δ2D method using the CA, CB, C', N and HN chemical shifts (Figure 6, see the Experimental section) indicate the presence of four α-helices at residues Ile^{81}-Gln^{88}, Gln^{93}-Lys^{99}, Glu^{104}-Glu^{111} and Ile^{119}-Gly^{130} in line with bioinformatics prediction of a tetrahelical bundle HTH domain [1]. Leu^{133} and Val^{134} in the fourth α-helix as well as residues that are structurally close to that α-helix (Leu^{125} and Glu^{126}) give rise to two set of resonances in NMR spectra. The intensity ratio for those two sets of resonances is ~ 70%/30% and probably results from the *cis-trans* isomerization of Pro^{136} on a long timescale for NMR spectroscopy (> 100ms). Interestingly, Pro^{136} is widely conserved among crenarchaeal aMBF1 sequences (Supplementary Figure S4). The consequence of this putative *cis-trans* isomerization has not been investigated in detail, but, as shown below, does not seem to affect ribosome binding.

The boundaries of the HTH motif within aMBF1 were investigated further via a series of ^{15}N NMR measurements of the relaxation parameters (T_1, T_2 and {^{1}H-^{15}N}-NOE, see the Experimental section). Relaxation NMR data were obtained for 82 of the 100 residues in aMBF1-C (Figure 6). The data are characterized by uniform R_1, R_2 and {^{1}H-^{15}N}-NOE values from residues Ile^{81} to Thr^{135}. The absence of regions of the polypeptide showing higher R_2 values indicates the likely absence of significant conformational exchange processes on the millisecond timescale. aMBF1-C appears to be a compact domain with a rotational correlation time (τ_c) of 5.8 ± 0.2 ns. Flanking this domain are regions (before Ile^{81} and after Thr^{135}) characterized by reduced {^{1}H-^{15}N}-NOE values (<0.5), indicating that these regions undergo sub-nanosecond motions. This is characteristic of structurally intrinsically disordered regions. Notably, the C-terminal disordered region is highly conserved in sequence and specific to aMBF1 orthologues [15].
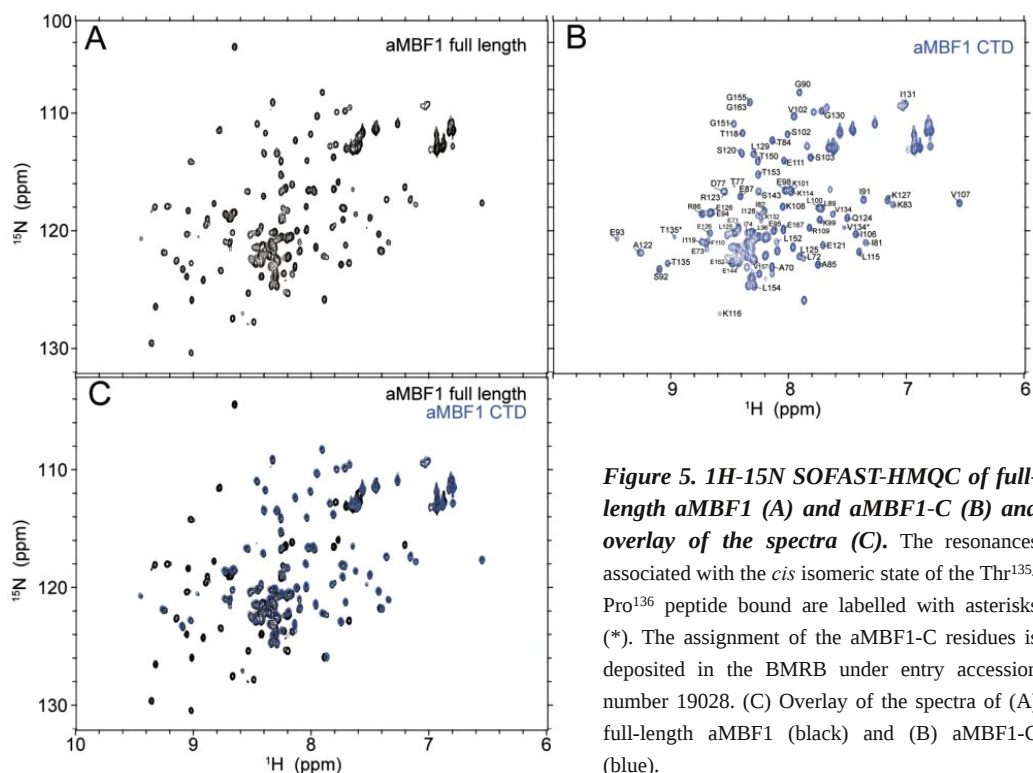
Figure 5. 1H-15N SOFAST-HMQC of full-length aMBF1 (A) and aMBF1-C (B) and overlay of the spectra (C). The resonances associated with the *cis* isomeric state of the Thr[135]-Pro[136] peptide bound are labelled with asterisks (*). The assignment of the aMBF1-C residues is deposited in the BMRB under entry accession number 19028. (C) Overlay of the spectra of (A) full-length aMBF1 (black) and (B) aMBF1-C (blue).

An ensemble of structures representing the structure and dynamics of the HTH domain was then determined using NMR chemical shifts as restraints in MD simulations using the CamShift-MD approach [39, 40] (Supplementary Figure S5). The resulting structures (PDB code 2MEZ) show the disordered regions flanking the HTH domain, with the electrostatic surfaces indicating the presence of a positively charged region at the N-terminus of the HTH domain (Figure 7, lower panel). In comparison, eukaryotic MBF1 exhibits a significantly less pronounced positively charged surface [48] (Supplementary Figure S6).

Structural investigation of the interaction of aMBF1 and 30S

In order to clarify the role of the HTH domain in the interaction of aMBF1 with the 30S ribosomal subunit, ^{1}H-^{15}N correlation NMR spectra of ^{15}N-labelled aMBF1 in the absence and presence of unlabelled 30S ribosomal subunit were recorded. These showed the selective broadening of a highly discrete subset of 38 cross-peaks occurring in the presence of the 30S ribosomal subunit (Figure 8A and 8B), and further analysis showed that these resonances arose from the C-terminal HTH motif (Figure 8C). More specifically, all resonances assigned to residues within Ala[70] to Ile[135] were broadened beyond detection, indicating that the ribosomal interaction is specific and mediated by the HTH domain. The doubled set of resonances for residues neighbouring Pro[136] that were assigned to two isoforms of the isomerization of Pro[136] were both broadened beyond detection, indicating that both isomers bind to the ribosome to the same extent. The remaining resonances observable in the complex were probed using X-STE diffusion NMR methods. A diffusion coefficient of $(4 \pm 0.2) \times 10^{-11}$ m$^2 \cdot$s^{-1} was determined from these data, a value significantly lower than that observed for the isolated protein
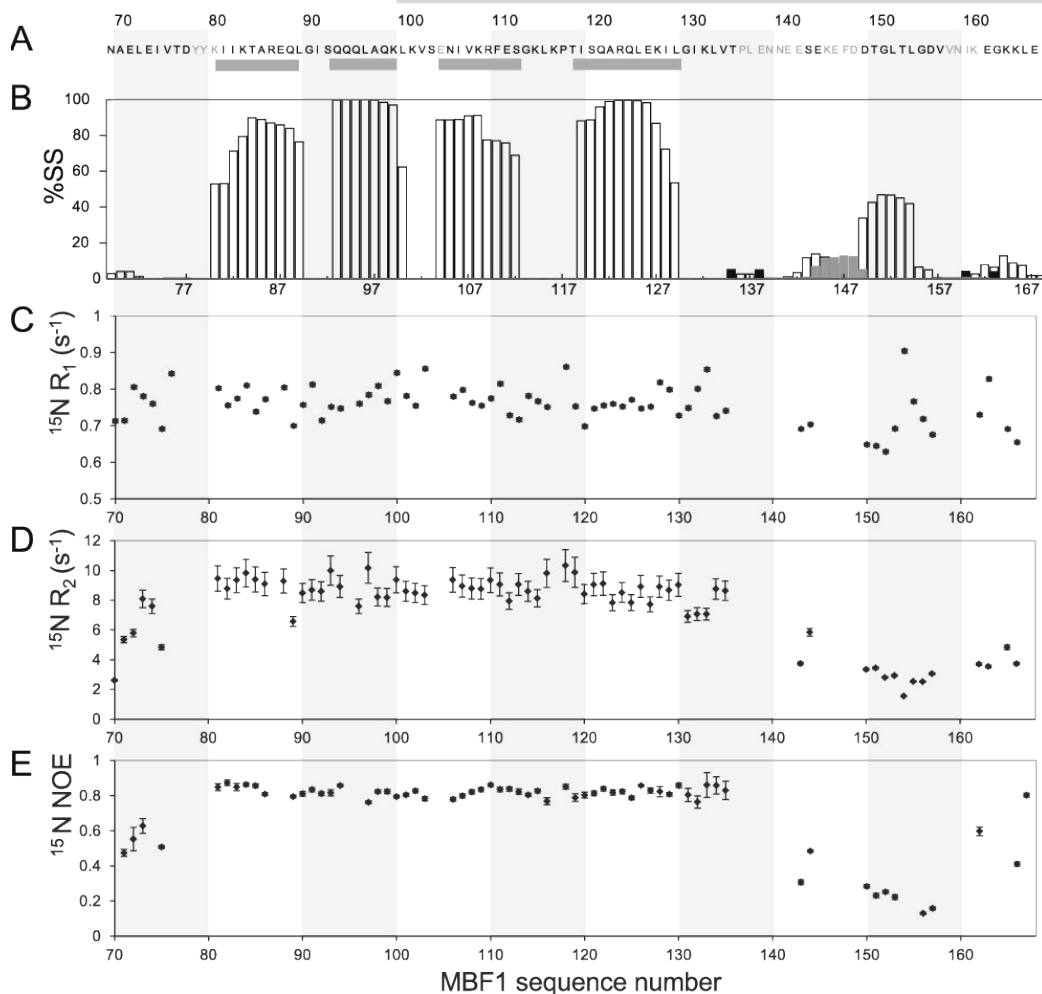
Figure 6. NMR analysis of the C-terminal HTH domain of aMBF1. (A) Sequence of aMBF1 with the α-helices as defined by the Cα and C' chemical shift shown by the grey bars. (B) Secondary-structure populations estimated from the chemical shifts using the δ2D method [37] (α-helices are shown as white bars, β-sheet as black bars and polyproline II as grey bars). (C-E) ^{15}N relaxation parameters: R_1 (C), R_2 (D) and the heteronuclear NOE (E) (see the Experimental section for details).

[(1.3 ± 0.05)×10^{-10} m^2·s^{-1}]. Moreover, as the former value is identical with that measured for the ^{1}H resonances observed for the 30S ribosomal subunit alone (Figure 8D), this appears to reflect the association of aMBF1 to the 30S ribosomal subunit. An NMR titration of ^{15}N-labelled aMBF1-C to the 30S subunit was then undertaken (Supplementary Figure S7) and showed that resonances of free aMBF1-C could be detected only with a 10-fold excess of aMBF1 to 30S (at a concentration of 5 µM), This finding suggests that the complex lifetime is at least 10-fold lower than the NMR acquisition time (50 ms), confirming a highly transient complex (Supplementary Figure S7).

The flanking disordered C-terminal region (from Ser[143] onwards) is, however, observed in the spectrum of 30S-bound aMBF1-C, whereas the linker region at the N-terminal end of the HTH domain is broadened beyond detection. The positively charged surface at the N-terminal hemisphere of the model structure of the HTH domain (Figure 7B, blue) could mediate the interaction with rRNA.

The resonances of the predicted zinc-ribbon domain are only partially broadened in the presence of 30S ribosomal subunit, indicating that the flexible linker remains disordered, resulting in sufficient mobility of the N-terminal domain to tumble independently from the MDa ribosomal complex (Figure 8E). Although the zinc-ribbon domain of *S. solfataricus* aMBF1 does not participate directly in the interaction with 30S ribosomal subunit, sucrose density gradient centrifugation of reconstituted complexes alluded to a contribution of the zinc-ribbon to 30S ribosomal subunit binding (Figure 4). The effect may be indirect through the influence of the zinc-ribbon domain on the conformation of the linker region; however, we cannot rule out some bridging function with an unidentified partner (in analogy to eukaryotic MBF1). The interaction of aMBF1 with 30S ribosomal subunits is thus restricted to the HTH domain and the adjacent part of the linker that are shared with eukaryotic MBF1, whereas the Archaea-specific N-terminal zinc-ribbon domain and C-terminal extension are not involved.

aMBF1 deletion does not influence misreading in translation fidelity

Deletion of *mbf1* in yeast affects the rate of ribosomal frame shifting, as well as the sensitivity to the antibiotic paramomycin that targets the 30S ribosomal subunit. The molecular basis for this phenomenon is unknown, but, given that all other identified suppressors of frameshift mutations in yeast map to core factors of the translation apparatus [tRNAs, EF1α (translation elongation factor 1α) and ribosomal protein S3] [8, 49–52], we reasoned that possibly yeast MBF1 interacts directly with the translation machinery as well. To test whether aMBF1 might influence translation fidelity in a similar manner, we made use of the fact that paramomycin induces misreading in *S. solfataricus* cell-free translation systems, whereas, in general, the archaeal translation apparatus is rather insensitive to antibiotics [53–55]. Misreading was measured using *in vitro* translation assays with a synthetic poly(U) RNA template coding for polyphenylalanine. The use of near-cognate tRNA[Leu] instead of cognate tRNA[Phe] was measured as the misincorporation rate of leucine into polyphenylalanine using radiolabelled amino acids. We observed no significant difference in the basic rate of misreading for lysate of the $\Delta mbf1$ strain or its parental strain PBL2025, and misreading increased to the same extent in response to paromomycin for both strains (Supplementary Figure S8A). Furthermore, when a cell lysate programmed for translation was supplemented with 100 μM paromomycin, aMBF1 was still mostly associated with 70S ribosomes, suggesting that paromomycin does not compete with aMBF1 for its ribosome-binding site (Supplementary Figure S8B).

Conclusions

We have shown that aMBF1 from *S. solfataricus* interacts with the 30S ribosomal subunit. aMBF1 is structurally composed of two domains, which have independent mobility. The ribosome-binding interface is potentially the positively charged surface that we identified at the N-terminal hemisphere of the HTH domain of aMBF1. This binding interface is likely to be affected by the conformational sampling of the linker region in the presence of the zinc-ribbon
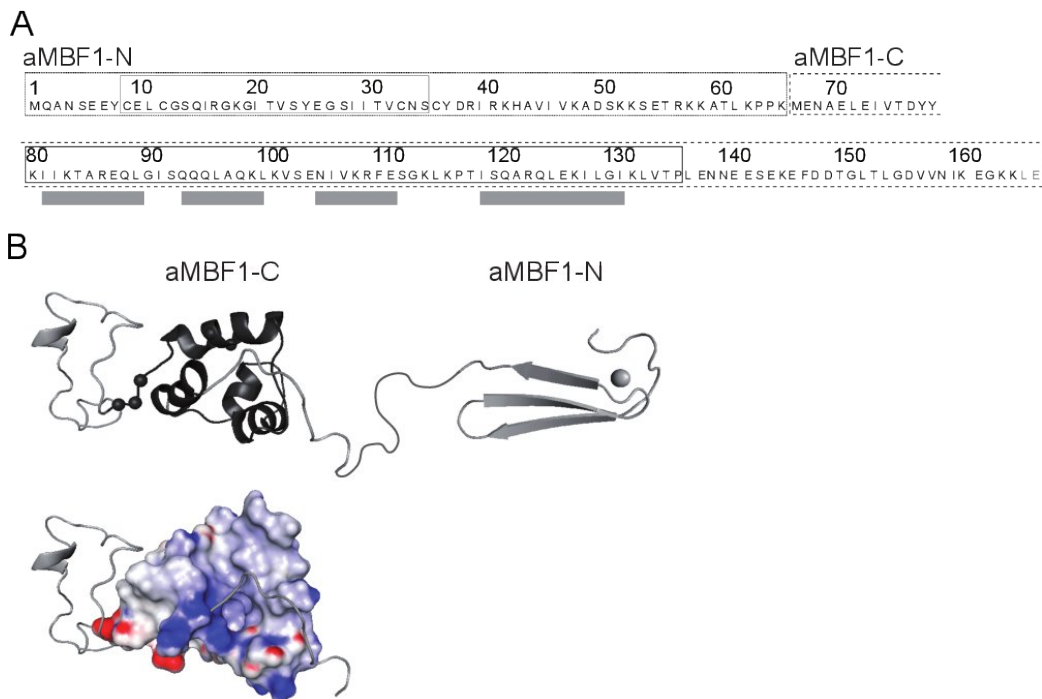
aMBF1-N domain. Within the aMBF1-ribosome complex, the aMBF1 zinc-ribbon-binding domain maintains very high mobility, suggesting that it remains accessible for potential interaction with a third binding partner. Our data suggest a function for aMBF1 related to translation, perhaps as a recruitment factor of the translation apparatus by bridging different ligands, analogous to the function of eukaryotic MBF1 in transcription regulation. aMBF1 does not inhibit translation and its expression is highest during exponential growth. This may indicate that the role of aMBF1 in protein synthesis is not related to the stress response. Our understanding of the molecular function of aMBF1 will be advanced further by the identification of potential ligands that may bind to the N-terminal zinc-ribbon domain. Archaeal MBF1 and eukaryotic MBF1 are orthologous proteins and hence it has been proposed that aMBF1 functions as core transcription factor in archaea as well [2]. We provide evidence that aMBF1 interacts physically with the translation machinery and it is likely that aMBF1 carries out a function related to the translation process. Interestingly, there is accumulating evidence that also the eukaryotic MBF1 carries out additional function(s) beyond transcription initiation. It is worth noting that, since the conserved HTH of aMBF1 mediates the interaction with the 30S ribosomal subunit, eukaryotic MBF1 might bind similarly to the 40S ribosomal subunit. The recent findings that MBF1 associates with polyadenylated mRNA in different eukaryotic species [11–13] cannot be explained by the known function of MBF1 in transcription initiation, but it would be compatible with a conserved interaction of MBF1 and ribosomes in archaea and eukaryotes.

Acknowledgements

Funding

A

aMBF1-N aMBF1-C

| 1 | 10 | 20 | 30 | 40 | 50 | 60 | 70 |

MQAN SE EY CE L CGSQI RGKGI TV S YE G SI I TVCN SCY DRI RKHAV I VKAD SKKS ET RKK ATL KPPK MENA E LE I VT DYY

| 80 | 90 | 100 | 110 | 120 | 130 | 140 | 150 | 160 |

KI I KTA REQL GI SQQQL AQKL KVS ENI VKRF ESGKL KPT I SQARQL EKI LGI KLVTP L ENNE ESEKE FD DTGL TL GDVVN IK EGKKL E

B

aMBF1-C   aMBF1-N

Figure 7. Structural characterization of S. solfataricus aMBF1. (A) Amino acid sequence with the aMBF1-N and aMBF1-C domain boundaries shown as broken-line boxes, the structured domains shown by boxes and the four α-helices of the HTH domain as defined by the δ2D method [37] shown by grey bars (see Figure 2). The C-terminal two residues in grey constitute the linker to the His$_6$ tag. (B) Top: structural model of aMBF1. A model for the zinc-ribbon motif of aMBF1 was built using homology modelling and the zinc-ribbon structure of *Methanococcus jannaschii* translation initiation factor 2β (PDB code 1K81) as template. The zinc atom is represented as a sphere. A structural model for the HTH domain of aMBF1 was built on the basis of the information contained in the chemical shifts following the procedure implemented in [38]. The intrinsically disordered sequences are represented as lines; an ensemble of structures representing the dynamics of the protein determined using chemical shift retrained MD [39, 40] is shown in Supplementary Figure S5 (see Supplementary Methods). The residues that give two sets of resonances associated with the isomerization of the Thr^{135}-Pro^{136} peptide bound are shown as spheres. Bottom: the electrostatic surface is shown ranging from -20 kT/e (red) to +20 kT/e (blue).
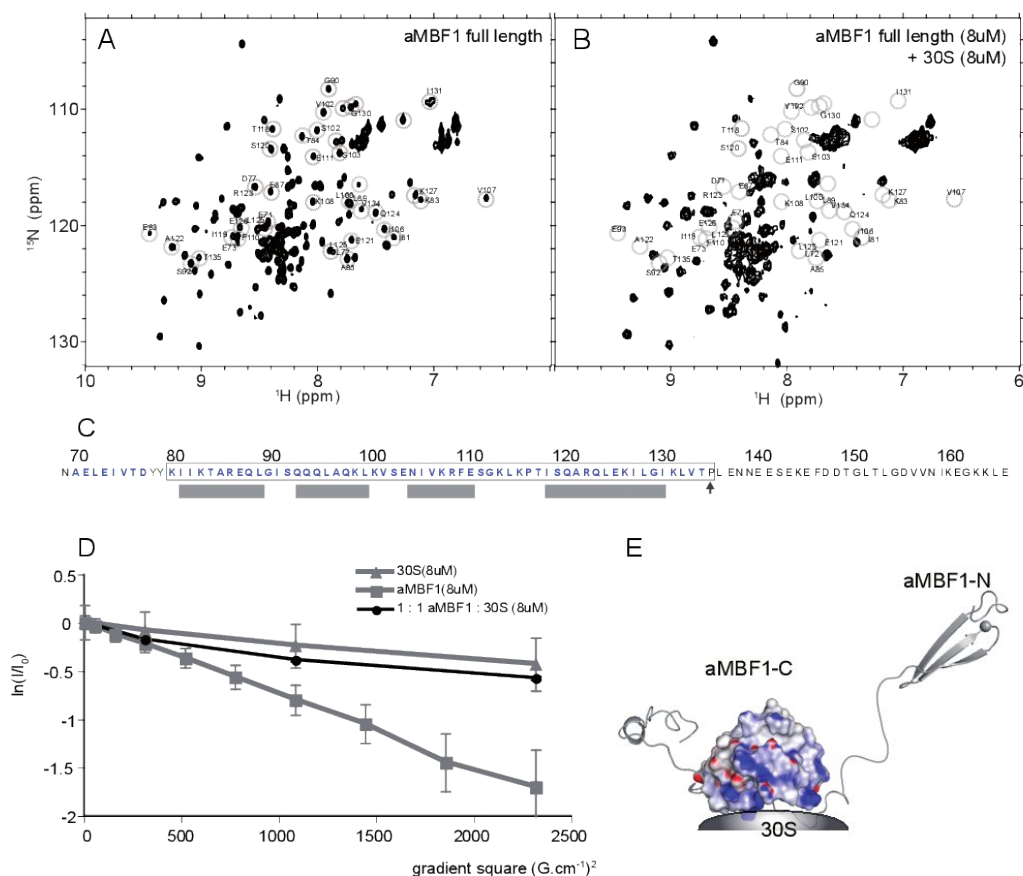
Figure 8. 1H-15N heteronuclear NMR spectra of aMBF1 in the absence and presence of 30S ribosomal subunits. (A) Full-length aMBF1 (^{1}H-^{15}N HSQC). (B) Full-length aMBF1 and 30S ribosomal subunits in a 1:1 molar ratio (8 µM) (^{1}H-^{15}N SOFAST-HMQC). The cross-peaks marked with open circles were broadened in the presence of 30S ribosomal subunits; these were all assignable to aMBF1-C. (C) Sequence of the C-terminal domain of aMBF1. The folded region determined by NMR is boxed and the residues whose resonances are broadened due to the ribosome interaction are shown in blue. The grey residues are not assigned. (D) Translational diffusion NMR measurements of the interactions. Stejskal-Tanner plot [relative NMR signal intersities of the aMBF1-C resonances gradient strengths ($G^2 \cdot cm^{-2}$)] for aMBF1 (grey squares), 30S (grey circles) and aMBF1 in the presence of 30S (black circles). In the presence of 30S particle, the resonances of aMBF1-N are associated with a diffusion coefficient identical with that of the ribosomal subunit, indicating an interaction with the complex. (E) Schematic diagram of the aMBF1-30S complex, in which the interaction is mediated by the positively charged surface of the aMBF1-C domain, and both the C-terminal disordered end of aMBF1-C and the N-terminal domain remains flexible enough to tumble independently from the aMBF1-30S complex.

References

1. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM: **The many faces of the helix-turn-helix domain: transcription regulation and beyond.** *FEMS Microbiol Rev* 2005, **29**:231–62.

2. Aravind L, Koonin E V: **DNA-binding proteins and evolution of transcription regulation in the archaea.** *Nucleic Acids Res* 1999, **27**:4658–70.

3. Blombach F, Makarova KS, Marrero J, Siebers B, Koonin E V, van der Oost J: **Identification of an ortholog of the eukaryotic RNA polymerase III subunit RPC34 in Crenarchaeota and Thaumarchaeota suggests specialization of RNA polymerases for coding and non-coding RNAs in Archaea.** *Biol Direct* 2009, **4**:39.

4. Takemaru K, Harashima S, Ueda H, Hirose S: **Yeast coactivator MBF1 mediates GCN4-dependent transcriptional activation.** *Mol Cell Biol* 1998, **18**:4971–6.

5. Liu Q-X, Nakashima-Kamimura N, Ikeo K, Hirose S, Gojobori T: **Compensatory change of interacting amino acids in the coevolution of transcriptional coactivator MBF1 and TATA-box-binding protein.** *Mol Biol Evol* 2007, **24**:1458–63.

6. Li FQ, Ueda H, Hirose S: **Mediators of activation of fushi tarazu gene transcription by BmFTZ-F1.** *Mol Cell Biol* 1994, **14**:3013–21.

7. Kabe Y, Goto M, Shima D, Imai T, Wada T, Morohashi K i, Shirakawa M, Hirose S, Handa H: **The role of human MBF1 as a transcriptional coactivator.** *J Biol Chem* 1999, **274**:34196–202.

8. Hendrick JL, Wilson PG, Edelman II, Sandbaken MG, Ursic D, Culbertson MR: **Yeast frameshift suppressor mutations in the genes coding for transcription factor Mbf1p and ribosomal protein S3: evidence for autoregulation of S3 synthesis.** *Genetics* 2001, **157**:1141–58.

9. Costanzo MC, Mueller PP, Strick CA, Fox TD: **Primary structure of wild-type and mutant alleles of the PET494 gene of Saccharomyces cerevisiae.** *Mol Gen Genet* 1986, **202**:294–301.

10. Culbertson MR, Gaber RF, Cummins CM: **Frameshift suppression in Saccharomyces cerevisiae. V. Isolation and genetic properties of nongroup-specific suppressors.** *Genetics* 1982, **102**:361–78.

11. Klass DM, Scheibe M, Butter F, Hogan GJ, Mann M, Brown PO: **Quantitative proteomic analysis reveals concurrent RNA-protein interactions and identifies new RNA-binding proteins in Saccharomyces cerevisiae.** *Genome Res* 2013, **23**:1028–38.

12. Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, Bonneau R, Selbach M, Dieterich C, Landthaler M: **The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts.** *Mol Cell* 2012, **46**:674–90.

13. Kwon SC, Yi H, Eichelbaum K, Föhr S, Fischer B, You KT, Castello A, Krijgsveld J, Hentze MW, Kim VN: **The RNA-binding protein repertoire of embryonic stem cells.** *Nat Struct Mol Biol* 2013, **20**:1122–30.

14. Marrero Coto J, Ehrenhofer-Murray AE, Pons T, Siebers B: **Functional analysis of archaeal MBF1 by complementation studies in yeast.** *Biol Direct* 2011, **6**:18.

15. De Koning B, Blombach F, Wu H, Brouns SJJ, Van Der Oost J: **Role of multiprotein bridging factor 1 in archaea: bridging the domains?** *Biochem Soc Trans* 2009, **37**(Pt 1):52–57.

16. She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CC, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PM, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, et al.: **The complete genome of the crenarchaeon Sulfolobus solfataricus P2.** *Proc Natl Acad Sci U S A* 2001, **98**:7835–40.

17. Gill SC, von Hippel PH: **Calculation of protein extinction coefficients from amino acid sequence data.** *Anal Biochem* 1989, **182**:319–26.

18. Blombach F, Launay H, Zorraquino V, Swarts DC, Cabrita LD, Benelli D, Christodoulou J, Londei P, van der Oost J: **An HflX-type GTPase from Sulfolobus solfataricus binds to the 50S ribosomal subunit in all nucleotide-bound states.** *J Bacteriol* 2011, **193**:2861–7.

19. Schägger H, von Jagow G: **Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa.** *Anal Biochem* 1987, **166**:368–79.

20. Schelert J, Dixit V, Hoang V, Simbahan J, Drozda M, Blum P: **Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon Sulfolobus solfataricus by use of gene disruption.** *J Bacteriol* 2004, **186**:427–37.

21. Zaparty M, Esser D, Gertig S, Haferkamp P, Kouril T, Manica A, Pham TK, Reimann J, Schreiber K, Sierocinski P, Teichmann D, van Wolferen M, von Jan M, Wieloch P, Albers S V, Driessen AJM, Klenk H-P, Schleper C, Schomburg D, van der Oost J, Wright PC, Siebers B: **"Hot standards" for the thermoacidophilic archaeon Sulfolobus solfataricus.** *Extremophiles* 2010, **14**:119–42.

22. Snijders APL, Walther J, Peter S, Kinnman I, de Vos MGJ, van de Werken HJG, Brouns SJJ, van der Oost J, Wright PC: **Reconstruction of central carbon metabolism in Sulfolobus solfataricus using a two-dimensional gel electrophoresis map, stable isotope labelling and DNA microarray analysis.** *Proteomics* 2006, **6**:1518–29.

23. Benelli D, Londei P: **In vitro studies of archaeal translational initiation.** *Methods Enzymol* 2007, **430**:79–109.

24. Nguyen TN, Goodrich JA: **Protein-protein interaction assays: eliminating false positive interactions.** *Nat Methods* 2006, **3**:135–9.

25. Pühler G, Lottspeich F, Zillig W: **Organization and nucleotide sequence of the genes encoding the large subunits A, B and C of the DNA-dependent RNA polymerase of the archaebacterium Sulfolobus acidocaldarius.** *Nucleic Acids Res* 1989, **17**:4517–34.

26. Benelli D, Marzi S, Mancone C, Alonzi T, la Teana A, Londei P: **Function and ribosomal localization of aIF6, a translational regulator shared by archaea and eukarya.** *Nucleic Acids Res* 2009, **37**:256–67.

27. Condò I, Ciammaruconi A, Benelli D, Ruggero D, Londei P: **Cis-acting signals controlling translational initiation in the thermophilic archaeon Sulfolobus solfataricus.** *Mol Microbiol* 1999, **34**:377–84.

28. Londei P, Teixidò J, Acca M, Cammarano P, Amils R: **Total reconstitution of active large ribosomal subunits of the thermoacidophilic archaebacterium Sulfolobus solfataricus.** *Nucleic Acids Res* 1986, **14**:2269–85.

29. Mori S, Abeygunawardana C, Johnson MO, van Zijl PC: **Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation.** *J Magn Reson B* 1995, **108**:94–8.

30. Peng JW, Wagner G: **Investigation of protein motions via relaxation measurements.** *Methods Enzymol* 1994, **239**:563–96.

31. Lipari G, Szabo A: **Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results**. *J Am Chem Soc* 1982, **104**:4559–4570.

32. Stejskal EO, Tanner JE: **Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient**. *J Chem Phys* 1965, **42**:288.

33. Ferrage F, Zoonens M, Warschawski DE, Popot J-L, Bodenhausen G: **Slow diffusion of macromolecular assemblies by a new pulsed field gradient NMR method.** *J Am Chem Soc* 2003, **125**:2541–5.

34. Cavanagh J, Fairbrother WJ, Palmer III AG, Rance M, Skelton N: *Protein NMR Spectroscopy: Principles and Practice*. Second Edi. London: Academic Press; 2005.

35. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A: **NMRPipe: a multidimensional spectral processing system based on UNIX pipes.** *J Biomol NMR* 1995, **6**:277–93.

36. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED: **The CCPN data model for NMR spectroscopy: development of a software pipeline.** *Proteins* 2005, **59**:687–96.

37. Camilloni C, De Simone A, Vranken WF, Vendruscolo M: **Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts.** *Biochemistry* 2012, **51**:2224–31.

38. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M: **Protein structure determination from NMR chemical shifts.** *Proc Natl Acad Sci U S A* 2007, **104**:9615–20.

39. Camilloni C, Robustelli P, De Simone A, Cavalli A, Vendruscolo M: **Characterization of the conformational equilibrium between the two major substates of RNase A using NMR chemical shifts.** *J Am Chem Soc* 2012, **134**:3968–71.

40. Camilloni C, Cavalli A, Vendruscolo M: **Assessment of the use of NMR chemical shifts as replica-averaged structural restraints in molecular dynamics simulations to characterize the dynamics of proteins.** *J Phys Chem B* 2013, **117**:1838–43.

41. Best RB, Mittal J: **Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse.** *J Phys Chem B* 2010, **114**:14916–23.

42. Ottiger M, Delaglio F, Bax A: **Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra.** *J Magn Reson* 1998, **131**:373–8.

43. Rückert M, Otting G: **Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments**. *J Am Chem Soc* 2000, **122**:7793–7797.

44. Schanda P, Kupce E, Brutscher B: **SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds.** *J Biomol NMR* 2005, **33**:199–211.

45. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrín-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, et al.: **Global landscape of protein complexes in the yeast Saccharomyces cerevisiae.** *Nature* 2006, **440**:637–43.

46. Armache J-P, Anger AM, Márquez V, Franckenberg S, Fröhlich T, Villa E, Berninghausen O, Thomm M, Arnold GJ, Beckmann R, Wilson DN: **Promiscuous behaviour of archaeal ribosomal proteins: implications for eukaryotic ribosome evolution.** *Nucleic Acids Res* 2013, **41**:1284–93.

47. Márquez V, Fröhlich T, Armache J-P, Sohmen D, Dönhöfer A, Mikolajka A, Berninghausen O, Thomm M, Beckmann R, Arnold GJ, Wilson DN: **Proteomic characterization of archaeal ribosomes reveals the presence of novel archaeal-specific ribosomal proteins.** *J Mol Biol* 2011, **405**:1215–32.

48. Salinas RK, Camilo CM, Tomaselli S, Valencia EY, Farah CS, El-Dorry H, Chambergo FS: **Solution structure of the C-terminal domain of multiprotein bridging factor 1 (MBF1) of Trichoderma reesei.** *Proteins* 2009, **75**:518–23.

49. Winey M, Mathison L, Soref CM, Culbertson MR: **Distribution of introns in frameshift-suppressor proline-tRNA genes of Saccharomyces cerevisiae.** *Gene* 1989, **76**:89–97.

50. Sandbaken MG, Culbertson MR: **Mutations in elongation factor EF-1 alpha affect the frequency of frameshifting and amino acid misincorporation in Saccharomyces cerevisiae.** *Genetics* 1988, **120**:923–34.

51. Mendenhall MD, Culbertson MR: **The yeast SUF3 frameshift suppressor encodes a mutant glycine tRNA(CCC).** *Nucleic Acids Res* 1988, **16**:8713.

52. Ball CB, Mendenhall MD, Sandbaken MG, Culbertson MR: **The yeast SUF5 frameshift suppressor encodes a mutant glycine tRNA(CCC).** *Nucleic Acids Res* 1988, **16**:8712.

53. Cammarano P, Teichner A, Londei P, Acca M, Nicolaus B, Sanz JL, Amils R: **Insensitivity of archaebacterial ribosomes to protein synthesis inhibitors. Evolutionary implications.** *EMBO J* 1985, **4**:811–6.

54. Ruggero D, Londei P: **Differential antibiotic sensitivity determined by the large ribosomal subunit in thermophilic archaea.** *J Bacteriol* 1996, **178**:3396–8.

55. Londei P, Altamura S, Sanz JL, Amils R: **Aminoglycoside-induced mistranslation in thermophilic archaebacteria.** *Mol Gen Genet* 1988, **214**:48–54.

Supplementary Data

Structural modelling of the aMBF1 HTH domain

An initial model of the aMBF1-C HTH structure was derived from the chemical shift data using the CHESHIRE protocol [1] as follows. Fragments of the protein (from three to nine residues) are generated with main-chain dihedral angles and secondary structure compatible with the information contained in the chemical shifts. The fragments are then assembled in a combinatorial manner (molecular fragment replacement) to produce an ensemble of trial structures that are subsequently refined by exploiting the information about tertiary structure contained in the chemical shifts. In order to obtain the ensemble of aMBF1-C structures, replica-averaged chemical shift-restrained MD simulations were performed using GROMACS and PLUMED as described previously [2, 3] using the Amber03W force field [4]. The starting conformation was built starting from the initial HTH structures derived from chemical shifts by adding the disordered segments using PyMOL (http://www.pymol.org) [5].

The structure was protonated and solvated with 21 000 water molecules in a dodecahedron box of 666 nm^3 volume. The final structure from four 1 ns simulations were selected as starting structures for four replicas. Each replica was evolved through a series of annealing cycles between 300 and 380 K, each cycle being composed of 100 ps at 300 K, 100 ps of linear increase in the temperature to 380 K, 100 ps of constant temperature at 380 K and 300 ps of linear decrease in the temperature to 300 K. Each replica was evolved for 100 ns. Only structures from the 300 K portions of the simulations were taken into account for analysis. The resulting ensemble is composed by all the structures sampled at 300 K in the four replicas after discarding the first 10 ns.

```
taatacgactcactatagggagaGAGGTGAAATATAAATGTCTCAAAGCTTTGAGGGAGA
    T7 promoter        SD motif      M  S  Q  S  F  E  G  E

ATTAAAAACAATTCTCAGAAGTGGCAAAGTTATTTTGGGAACAAGGAAGACATTGAAATT
  L  K  T  I  L  R  S  G  K  V  I  L  G  T  R  K  T  L  K  L

ATTAAAGACAGGAAAAGGTAAAGGGAGTAGTAGTTTCTTCTACATTAAGGCAGGATCTAAA
  L  K  T  G  K  V  K  G  V  V  V  S  S  T  L  R  Q  D  L  K

AGACGATATAATGACATTTTCAAAATTTTCTGATATTCCAATTTATCTCTATAAAGGTAG
  D  D  I  M  T  F  S  K  F  S  D  I  P  I  Y  L  Y  K  G  S

TGGATATGAATTAGGGACATTATGCGGTAAACCTTTTATGGTATCTGTTATAGGTATAGT
  G  Y  E  L  G  T  L  C  G  K  P  F  M  V  S  V  I  G  I  V

TGATGAAGGGGAATCAAAAATTTTGGAGTTTATTAAAGAGGTGAAGCAATGAGTGCCAGA
  D  E  G  E  S  K  I  L  E  F  I  K  E  V  K  Q  *

AATTAAATTAAC
```

Figure S1. Sequence of the in vitro translation template used in the present study based on the orf104 mRNA. Mutations in the nucleotide and protein sequence are indicated as bold letters. The T7 promoter sequence used for in vitro transcription and the Shine–Dalgarno (SD) motif are underlined. Synthetic DNA of this sequence was cloned via KpnI and SacI into a pBluescript-derived vector and the plasmid was linearized with SacI before *in vitro* transcription.
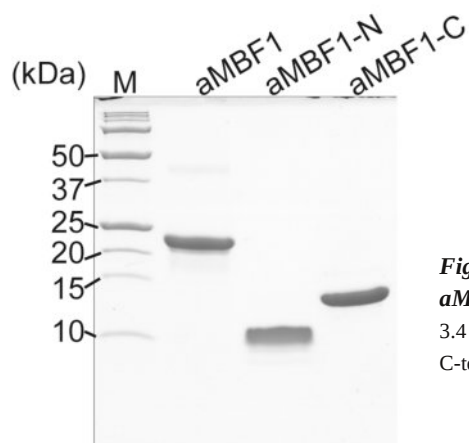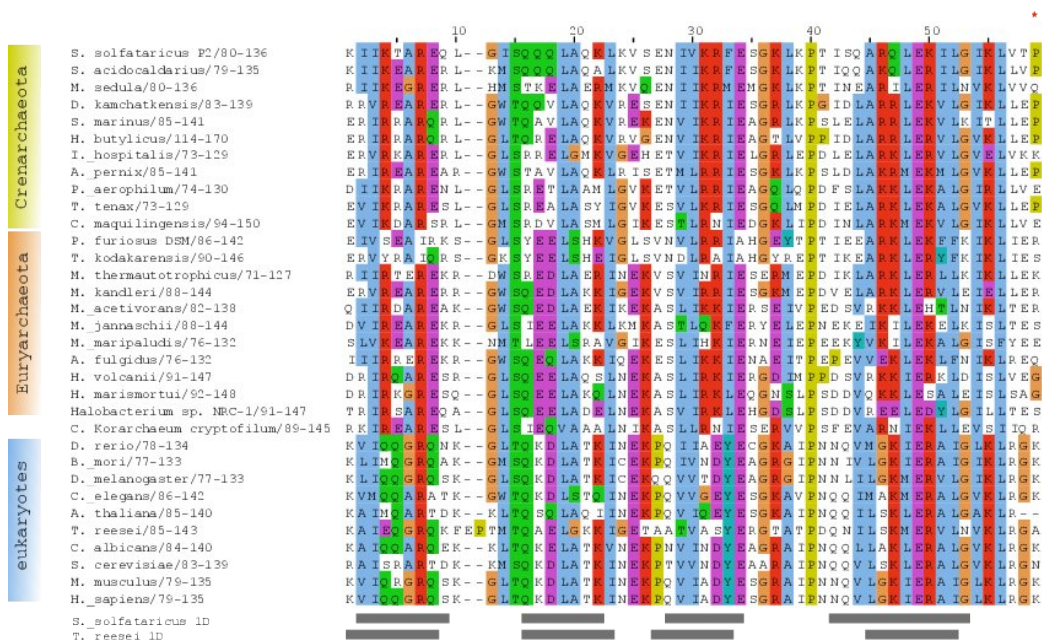
Figure S2. Heterologous expression of S. solfataricus aMBF1. Coomassie Blue-stained Tris/Tricine SDS/PAGE gel with 3.4 µg of purified recombinant aMBF1 and the isolated N- and C-terminal domains. Molecular masses (M) are indicate in kDa.
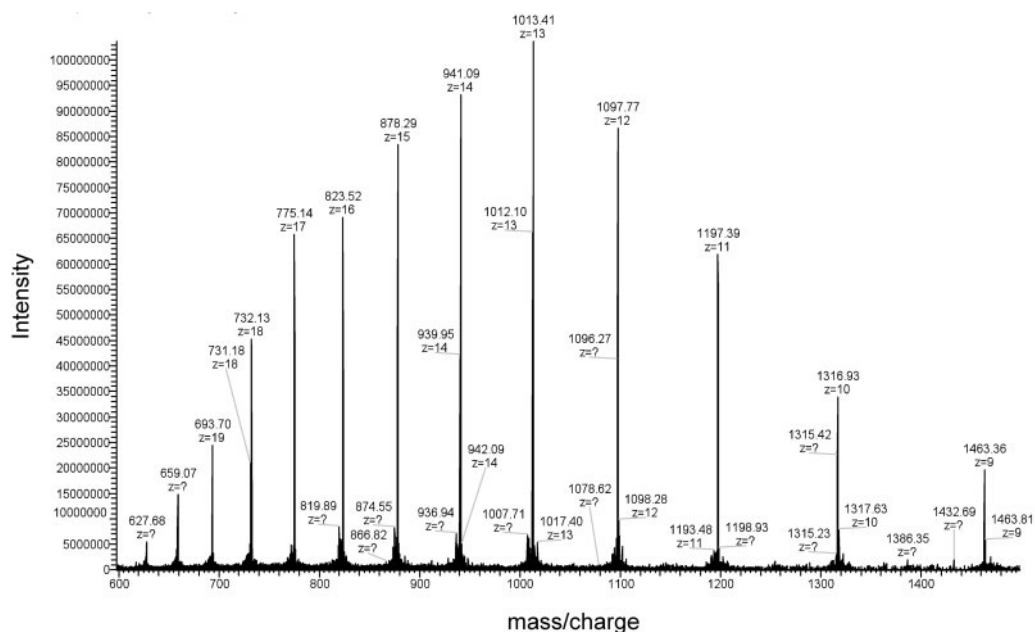

***Figure S3. Mass spectrum of intact MBF1-C indicates a molecular mass of 13.161 kDa.***

Figure S4. Sequence alignment of the HTH domains of archaeal and eukaryotic MBF1 orthologues. The red asterisk denotes the position of Pro[136] that undergoes *cis–trans* isomerization. Key to species names and GI numbers of sequences: *Sulfolobus solfataricus* strain P2, 15896971; *Sulfolobus acidocaldarius*, 70605853; *Metallosphaera sedula*, 146302785; *Desulfurococcus kamchatkensis*, 218883314; *Staphylothermus marinus*, 126464913; *Hyperthermus butylicus*, 124026906; *Ignicoccus hospitalis*, 156936795; *Aeropyrum pernix*, 118430835; *Pyrobaculum aerophilum*, 18311643; *Thermoproteus tenax*, 352681234; *Caldivirga maquilingensis*, 159040592; *Pyrococcus furiosus*, 18976372; *Thermococcus kodakaraensis*, 57639935; *Methanothermobacter thermoautotrophicus*, 15678031; *Methanopyrus kandleri*, 20093440; *Methanosarcina acetivorans*, 20088899; *Methanocaldococcus jannaschii*, 15668172; *Methanococcus maripaludis*, 45357563; *Archaeoglobus fulgidus*, 11497621; *Haloferax volcanii*, 292654178; *Haloarcula marismortui*, 55376942; *Halobacterium* sp. NRC-1, 15789340; 'Candidatus Korarchaeum cryptofilum', 170289627; *Danio rerio*, 312144725; *Bombyx mori*, 112984061; *Drosophila melanogaster*, 116010443; *Caenorhabditis elegans*, 392973747; *Arabidopsis thaliana*, 240254678; *Trichoderma reesei*, 340517347; *Saccharomyces cerevisiae*, 330443743; *Mus musculus*, 372099108; *Homo sapiens*, 224589821. The location of α-helices in *T. reesei* MBF1 and *S. solfataricus* aMBF1 are according to Salinas et al. [6] and the present study respectively.

Figure S5. Overlay of the ensemble of aMBF1-C structures determined using the CamShift-MD approach. See the Experimental section of the main text for a detailed description. The structures are aligned on the HTH motif (red). The flexible C-terminus is depicted in grey; the truncated part of the flexible linker is depicted in green.
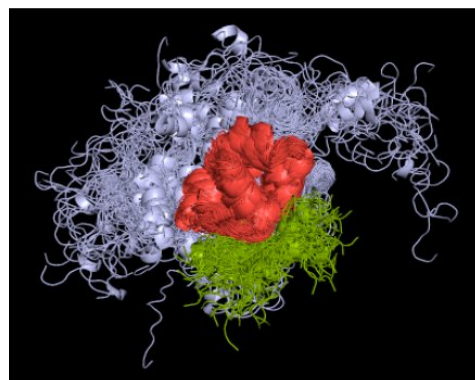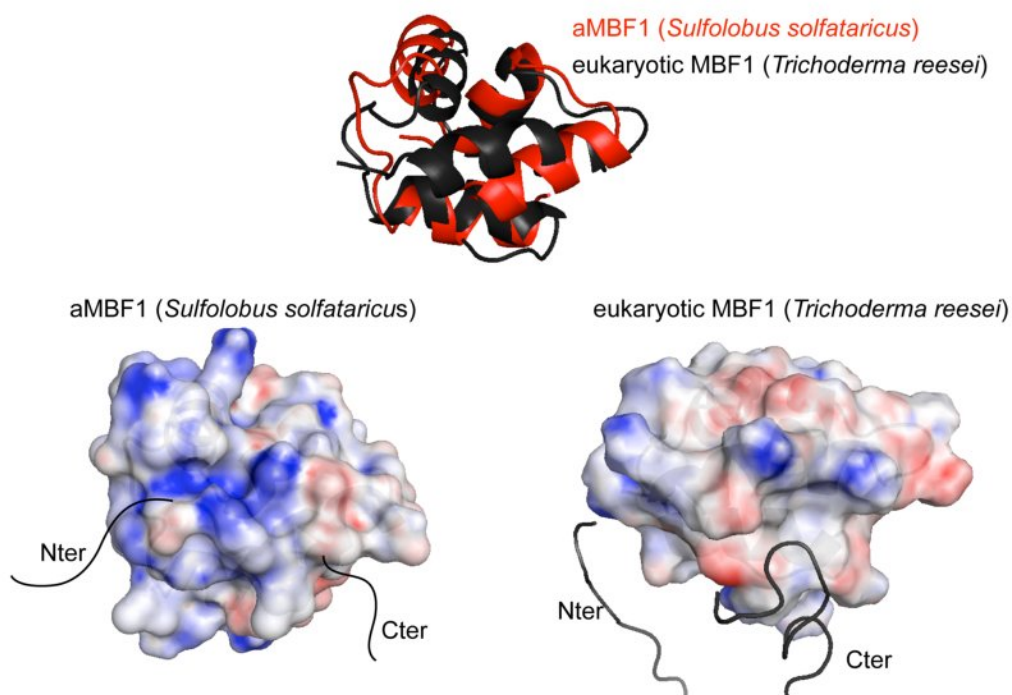
Figure S6. Comparison of the electrostatic surfaces of the HTH domains of aMBF1 from S. solfataricus and eukaryotic MBF1 from Trichoderma reesei (PDB code 2JVL). Upper panel: overlay of the two HTH domains. Lower panel: comparison of the electrostatic surfaces with the proteins in the same orientation as above. A range from −20 kT/*e* (red) to +20 kT/*e* (blue) is shown.
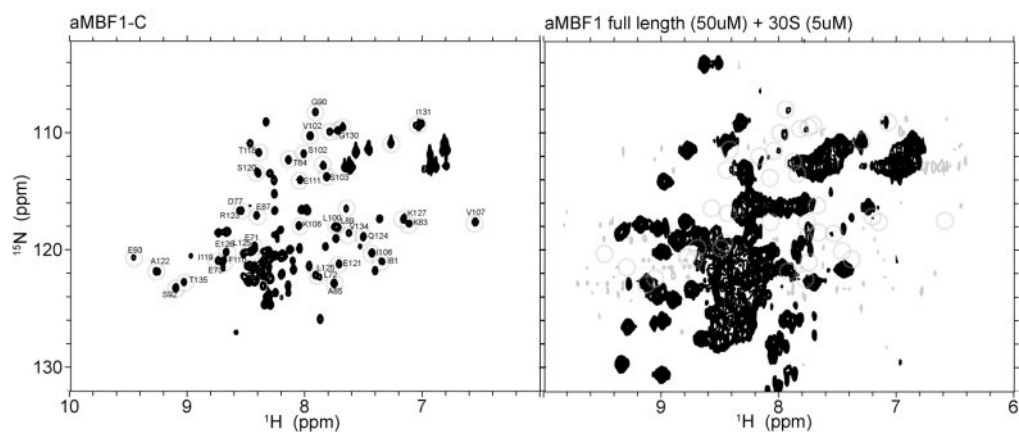
Figure S7 . 1H-15N NMR spectra of an excess of aMBF1 in the presence of 30S ribosomal subunits. (A) ^{1}H-^{15}N HSQC of isolated aMBF1-C. The cross-peaks marked with open circles were broadened in the presence of 30S ribosomal subunits. (B) ^{1}H-^{15}N SOFAST-HMQC of full-length aMBF1 and 30S ribosomal subunits at a 10:1 molar ratio (50 to 5 µM). The cross-peaks marked with open circles were broadened in the presence of 30S ribosomal subunits.
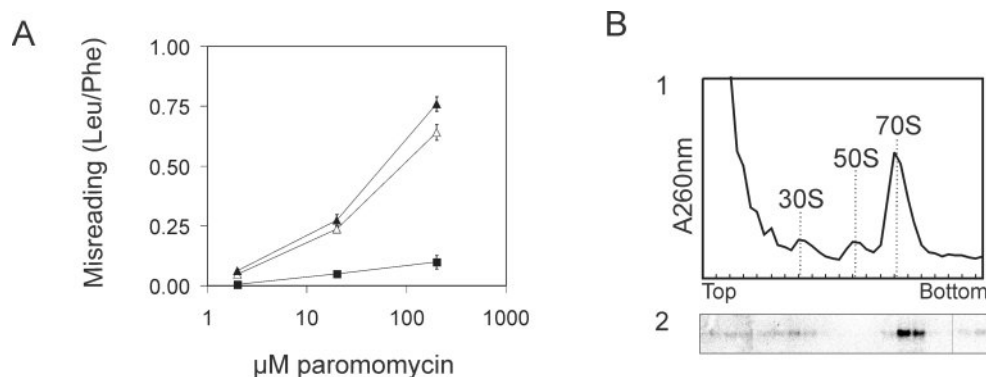
Figure S8. Effect of aMBF1 on paromomycin-induced misreading in poly(U)-directed translation assays. (A) *In vitro* misreading rate as detected by the incorporation of leucine into polyphenylalanine by poly(U)-programmed ribosomes. Poly(U) encodes polyphenylalanine; upon misreading of tRNA[leu] by the ribosomes, leucine will be incorporated. The ratio between phenylalanine and leucine incorporated into the polypeptide is measured. *In vitro* translation reactions were set up as in the other experiments, but the mRNA template was replaced by 20 μg of poly(U) per 25 μl assay volume. In addition, assays contained 3 mM spermine to increase translation fidelity [7] and 20 μM phenylalanine and leucine replacing the amino acid mixture. In each assay, one of the amino acids was replaced by either L-[U-14C]phenylalanine or L-[U-14C]leucine (PerkinElmer) respectively. After incubation at 70 °C for 30 min, 18 μl was spotted on a 1 cm×1.5 cm sheet of 3 mm Whatman chromatography filter paper and precipitated overnight at 4 °C in 10 % trichloroacetic acid. Filters were washed at 95 °C in 5 % trichloroacetic acid for 5 min, three times in ice-cold 5 % trichloroacetic acid and finally, filters in 96 % ethanol before scintillation counting. Each assay was carried out in duplicate plus two additional samples lacking poly(U) for background subtraction. Different cell lysates were tested: Δ*mbf1* (open triangle), parental strain PBL2025 (closed triangle) and wild-type strain P2 (closed square). The rate of leucine residues incorporated per phenylalanine residue was approximately 0.004 for the wild-type strain P2 in the absence of antibiotic, in good agreement with the published value of 0.003 [7]. Both the Δ*mbf1* strain and its parental strain PBL2025 exhibited misreading rates of approximately 0.020 leucine residues incorporated per phenylalanine residue in the absence of antibiotic. The ~5-fold higher misreading rates observed in PBL2025 and the mbf1 strain might be due to the deletion of ~50 genes in the parental PBL2025 strain [8]. Alternatively, it could be a phenotype of the S. solfataricus 98/2 strain from which PBL2025 was derived. (B) Effect of 100 μM paromomycin on the co-migration of endogenous aMBF1 with 30S ribosomal subunits and 70S ribosomes in cell lysate programmed for translation.

References

1. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M: **Protein structure determination from NMR chemical shifts.** *Proc Natl Acad Sci U S A* 2007, **104**:9615–20.
2. Camilloni C, Cavalli A, Vendruscolo M: **Assessment of the use of NMR chemical shifts as replica-averaged structural restraints in molecular dynamics simulations to characterize the dynamics of proteins.** *J Phys Chem B* 2013, **117**:1838–43.
3. Best RB, Mittal J: **Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse.** *J Phys Chem B* 2010, **114**:14916–23.
4. Ottiger M, Delaglio F, Bax A: **Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra.** *J Magn Reson* 1998, **131**:373–8.
5. Schrödinger L: **The PyMOL Molecular Graphics System, Version 1.3**. 2010.
6. Salinas RK, Camilo CM, Tomaselli S, Valencia EY, Farah CS, El-Dorry H, Chambergo FS: **Solution structure of the C-terminal domain of multiprotein bridging factor 1 (MBF1) of Trichoderma reesei.** *Proteins* 2009, **75**:518–23.
7. Londei P, Teixidò J, Acca M, Cammarano P, Amils R: **Total reconstitution of active large ribosomal subunits of the thermoacidophilic archaebacterium Sulfolobus solfataricus.** *Nucleic Acids Res* 1986, **14**:2269–85.
8. Schelert J, Dixit V, Hoang V, Simbahan J, Drozda M, Blum P: **Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon Sulfolobus solfataricus by use of gene disruption.** *J Bacteriol* 2004, **186**:427–37.

Molecular characterization of tRNA-guanine transglycosylase of *Sulfolobus solfataricus*

6

Bart de Koning, Rie Matsumi, Sakharam. P. Waghmare, Olga Pougovkina, Ambrosius P. L. Snijders, Mark Dickman, Stan J.J. Brouns, John van der Oost

Abstract

The fidelity level of the translation of ribonucleotides to amino acid sequences is a trade off between velocity and precision. To increase the performance of translation, cells have evolved a multitude of tRNA nucleotide modifications that can increase the selective power to discriminate between highly alike tRNAs or that can increase the rigidity of tRNAs to make them more stable. One of those stability increasing modifications is the conversion of G15 to archaeosine (G*15). This modification is restricted to the archaeal domain and is present in almost every archaeon discovered today. While this modification itself is completely absent in the bacterial and eukaryotic domains, a very similar one adjusts selectivity in those: the queuosine modification of the G at the wobble position within the anti-codon of some tRNAs. It is assumed that archaeosine conversion is necessary to increase the stability of tRNAs, and it was proposed that this could be especially beneficial for the numerous (hyper)thermophiles present in the archaeal domain. To test this hypothesis *in vivo*, the gene that encodes a tRNA-guanine transglycosylase, which is an essential component of the modification pathway, was knocked out in the hyperthermophilic crenarchaeon *S. solfataricus*. Using Mass Spectrometry it was shown that archaeosine incorporation is abolished in this mutant. To test stress tolerance, growth was compared between the parental strain and the knockout mutant, while being exposed to stress conditions. The only difference observed was a small decrease in growth under normal laboratory conditions. Similar performance was observed in comparison to the parental strain if it was exposed to one of the stress conditions. We therefore postulate that the increase of tRNA stability by the archaeosine modification has a slight effect on the growth rate under optimal growth conditions, and therefore a tiny increase in fitness, but does not increase tolerance levels against heat, antibiotic, and nutrient stresses.

Introduction

Translation and Fidelity

Translation is the final step during protein synthesis as was already formulated by Francis Crick in his famous central dogma of Molecular Biology [1, 2]. During this step RNA transcripts of DNA, stretches of nucleotides, are translated into peptides, stretches of amino acids. This translation of genetic information occurs in ribosome ribonucleoprotein complexes, and involves a tight interplay between a messenger RNA transcript (mRNA), aminoacylated transfer RNAs (tRNAs), ribosomal RNAs, ribosomal proteins, and translation factors. Translation proceeds in such manner that a fidelity rate is maintained within strict borders. When fidelity would be too low, energy would be wasted by the formation of erroneous proteins, but when it is too high, translation velocity would drop below the point necessary to survive in the struggle for life [3].

Within the translation process the actual decoding takes place at the interface between a codon of a mRNA transcript and the complementary anti-codon of a aminoacylated-tRNA. These tRNAs are initially loaded with amino-acids in a separate process called aminoacylation, which is catalysed by specific aminoacyl-tRNA synthetases [4]. The ribosomes themselves are unable to verify whether the combination between the amino acid and the tRNA nucleotides are correct or not, therefore this acylation is a highly important step with respect to translation fidelity. To decrease the level of misacylated tRNAs several checkpoints exists, including (i) discriminating, by selection and editing, between highly alike amino acids for specific coupling of the appropriate amino acid to the corresponding tRNA by each specific aminoacyl-tRNA synthetase [5–7], (ii) subsequent proof-reading of coupled tRNAs by a special class of aminoacyl-tRNA synthetases that lack the synthesis domain, but kept their editing domain [8–10], and (iii) verification by EF-Tu (Elongation Factor-Thermo Unstable) before delivery to the ribosomes by binding misacylated tRNAs differently than the otherwise uniform binding to correctly acylated tRNAs [11].

Modifications

Cells take thorough measurements to increase the fidelity of both key steps in protein synthesis: tRNA loading and codon/anti-codon pairing. To check and correct the aminoacyl-tRNA coupling in a highly processive manner it is necessary to discriminate fast between highly similar tRNAs as well as similar amino acids. Specific synthetase binding regions within the tRNA stem and loops have been found, but the nucleotide divergence between the possible combinations is too small to explain fast and reliable discrimination. Also the flexibility of the nucleotide stretches is too high to create a beneficial micro-environment for this discrimination to take place. Cells have evolved therefore a multitude of nucleotide modifications to further differentiate between tRNA molecules and to increase their rigidity [12, 13].

tRNAs are modified at a relatively high frequency as compared to other RNAs. Typically around 10% of the nucleosides in tRNA are modified, and this level can increase up to 25%. These modifications range from simple methylations to very complex multistep transformations, often referred as hypermodifications ([14, 15] and references herein). In general, tRNA modification occurs in all domains of life, but specific modifications have evolved in different phylogenetic lineages, showing that there is a strong selective pressure on adjusting modifications [16]. The advantage of these modifications range from lowering of conformational flexibility, improving (thermal) stability, increasing the aminoacylation rate and specificity, altering the recognition appearance of tRNAs, or even adjusting the codon

recognition of the anti-codon [17]. The nucleotides that are modified, mainly concern nucleotides that are located within the core of the tRNA molecule as well as in its anti-codon arm. An important modification in the latter is commonly found in all three domains of life, i.e. in the wobble position (N34) of the anti-codon. The modifications of the wobble position are species specific and vary among tRNAs, and occur primarily in codons belonging to codon boxes that are shared by more amino acids. These modifications are known to contribute to the accuracy and efficacy of translation: a clear indication that there is a necessity to increase the discriminative power for these codons. Interestingly these modifications are often not identical in phylogenetically unrelated organisms, suggesting that this adaptation occurred independently in different lineages and therefore relatively late in the early evolution of life [18].

The relatively high abundance in the structural core of tRNAs, which is less accessible to the external factors, suggests that stabilization of the 3D structure is a key purpose of many of these modifications. This notion is supported by results obtained in *in vitro* systems where the lack of modification to some extent can be compensated by addition of magnesium ions (reviewed in [19]). In archaea, a general modification occurs at position 15 in the D-loop: the well conserved guanine is replaced by a positively charged archaeosine (G*) (figure 1, figure 2 and table 1) [20–23]. Interestingly, an almost identical process modifies the guanine at wobble position 34 to queuosine in bacteria and eukaryotes for specific tRNAs [24–28] (figure 2). This shows that related modifications with a common origin can be used to increase both the discriminative power of codon/anti-codon interactions, as the rigidity of tRNA structure. Nevertheless, although both pathways show a remarkable similarity, which suggests common ancestry, there is no species known to have both the archaeosine as the queosine pathways acting
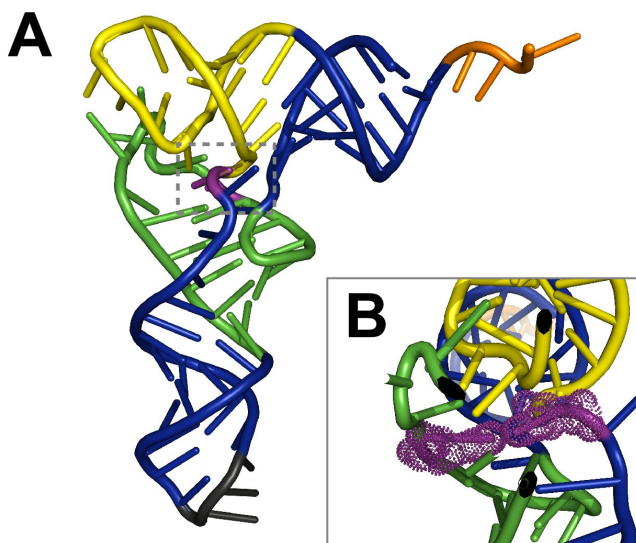
Figure 1. tRNA structure. (A) Complete tertiary structure of a yeast tRNA[Phe] where the location of the Levitt base pair is located in the grey dotted square and (B) a detail, focussed on the Levitt base pair, taken from the left side of the figure in B in the direction of the aminoacyl binding site. Purple dots indicate the surface of the interacting Levitt bases, showing the interaction between both. Colours in both structures match: the Levitt base pair (purple), the D-loop (green), the T-loop (yellow), the anticodon (black), and the CCA aminoacyl binding site (orange). Figures are rendered using PyMOL [44] from data deposited in the Protein Data Bank (1ehz) [45].
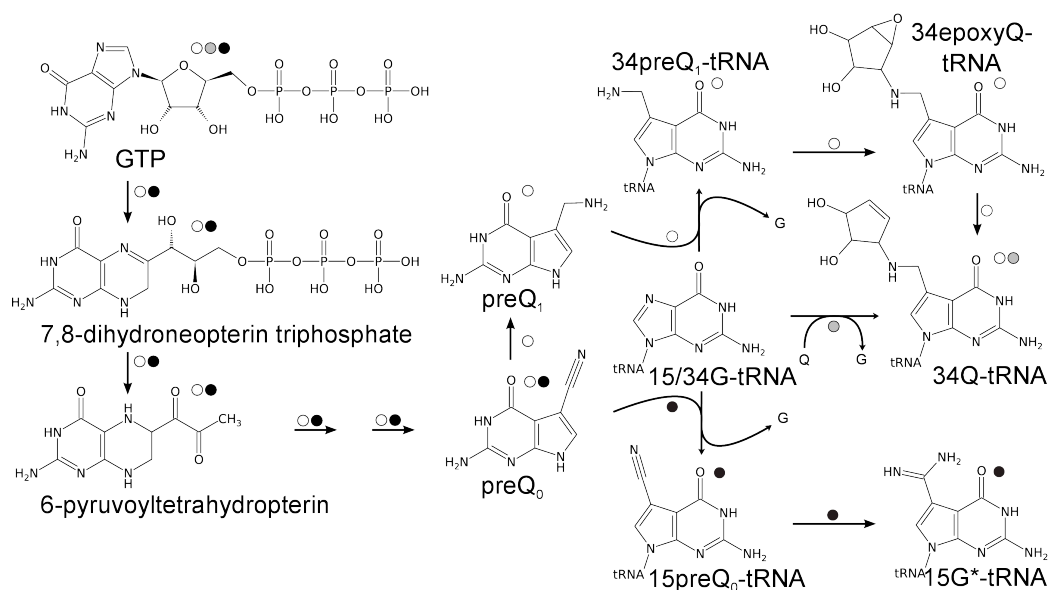
Figure 2. Queuosine (Q) and Archaeosine (G*) modification pathways. Molecular representation of the queuosine and archaeosine pathways in bacteria (white circles), eukaryotes (grey circles) and archaea (black circles). In both bacteria and eukaryotes, guanine (G) is replaced at position 34 with a queuosine. However, in bacteria queosine TGT exchanges G with preQ$_1$, which is converted into queuosine. Whereas eukaryotes cannot metabolize queosine and have to rely on queosine intake. In these organisms, TGT replaces the guanine directly by a queuosine. In archaea, TGT replaces the guanine at position 15 with a preQ$_0$ molecule, which is converted into archaeosine by a TGT paralog.

simultaneously. This could imply that the divergence between the pathways originates from after the split between the archaea and the bacteria, and might indicate that both pathways interfere with each other.

tRNA modifications are introduced by specific tRNA modifying enzymes. To support the impressive number of different modifications a cell is equipped with a huge array of these enzymes. A bacterium like *Salmonella typhimurium* is estimated to spend 1% of its genome to tRNA modifying enzymes ([15] and references herein). Despite their abundance and diversity, only a small portion of these enzymes has been studied in detail [29]. Amongst these well studied enzymes is TGT (transfer-RNA guanine transglycosylase), which is responsible for the replacement of guanine with preQ$_1$ in bacterial tRNAs, with preQ$_0$ in archaeal tRNAs, and directly with queuosine in eukaryotes [15]. Despite the differences in substrate or target sequence, this enzyme is highly conserved in the three domains. Although this conservation does not span the whole protein: it is restricted to the entire bacterial protein, the catalytic subunit of the eukaryotic dimer, and the N-terminal region of the archaeal protein [15, 30, 31].

The differences in substrate correlates with small substitutions in the active site. Most restricted is the archaeal enzyme that allows only utilisation of preQ$_0$, which appears to be caused by a size reduction of the active site. The bacterial enzyme has a slightly bigger pocket that is allows preQ$_1$ and analogues. Two substitutions (Val233Gly and Cys158Val), allow the eukaryotic enzyme to utilize both queuosine and guanine itself. Both the bacterial and the eukaryotic enzymes have a minimal sequence requirement of a minihelix with a 7-base loop containing the UGU sequence, although the eukaryotic enzyme additionally requires an intact

Table 1. tRNAs present in Sulfolobus solfataricus. G15 is shown in bold. The anticodon is underlined and introns are depecited in lower case letters. Source: the Genomic tRNA Database (GtRNAdb) [43].

Name	AminoAcid	Anticodon	Intron	Length	Sequence
trna19	Ala	CGC		74	GGGCCGGTAGCTCAGCCtgGAAGAGTGCTCGGTTCGCACCCGAGAGGtCCCGGGTTCAAATCCCGGCCGGTCCA
trna7	Ala	GGC		74	GGGCCGGTAGCTCAGCCtgGAAGAGTGCTCGGTTCGCATCCGAGAGGtCCCGGGTTCAAATCCCGGCCGGTCCA
trna4	Ala	TGC		74	GGGCCGGTAGCTCAGCCtgGAAGAGTGCTCGGTTTGCACCCGAGAGGtCCCGGGTTCAAATCCCGGCCGGTCCA
trna5	Arg	CCG		75	GGACCCGGTAGCTCAGCCaGGAtGGAGCGCGGCCTCCCGAGGtCCCGGGTTCAAATCCCGGCGGGTCCG
trna34	Arg	CCT	40-52	88	GGACCCGGTAGCTCAGCCaGGAtGGAGCGCGGCCTCCTAacttcggaagaaAGCCGGTGtCGCGGGTTCGAATCCCGCCGGGTCCG
trna31	Arg	GCG		75	GGACCCGGTAGCTCAGCCaGGAcGGAGCGCGGCCTTGCGGAGCCGGAGtCCCGGGTTCAAGTCCCGGCGGGTCCG
trna25	Arg	TCG		75	GGACCCGGTAGCTCAGCTaGGAtGGAGCGCGGCCTTCGGAGCCGGTGtCCCGGGTTCAAATCCCGGCGGGTCCG
trna35	Arg	TCT	40-54	90	GGACCCGGTAGCTCAGCCaGGAcAGGAGCGCGGCCCTTCTAgcgtctaaggaggatAGCCGGCGGtCCCGGGTTCAAATCCCGGCGGGTCCG
trna2	Asn	GTT	38-51	87	GCCGGAGTAGCTCAGTTGGCAGAGCCCCGGCTGTTAtgccggtgcagaaACCGGGTGGtCGGAGGTTCAGTCCTCCCTCCGGCG
trna40	Asp	GTC		73	GCCGGGTAGTCTAGTTGGTctAGGATGGGGGCTCGTCGAGCCCTGACCCGGGTTCAAATCCCGGCCCGGCG
trna42	Cys	GCA	39-52	89	GCCGGGTGGCCGAGTGGTtaAGGCGGCGGGCCCAGtgtactagttgacACCCGTTATtCGCGGGTTCGAATCCCGCCCCCCGGCT
trna17	Gln	CTG		73	AGCCGGGTAGTCTAGTtGGtcaAGGATCCAGGGCTTTGGCCCCTGGGACCAGGGTTCGAATCCCTGCCCGGCTA
trna9	Gln	TTG		73	AGCCGGGTAGTCTAGTtGGtcaAGGATCCAGGGCTTTGGCCCCTGGGACCAGGGTTCGAATCCCTGCCCGGCTA
trna26	Glu	CTC		75	GCCGGGTGGTGTAGCCcGGtCaAGAATCGGGCCTCTGACAAAtCGGGCGTTCGACCCGTGACCCGGCGCGGCA
trna32	Glu	TTC		75	GCCGGGTGGTtAGCCcGGTCaAGAATCGGGCCTTTCGACCCCGTGACCCGGGTTCAAATCCCGGCGCGGCA
trna30	Gly	CCC		76	GCGGGCGTAGTCTAGCCtGGAttAGGACGCCGCCCTCCCAAGCGGGTGAtCCCGGGTTCAAATCCCGGCGCCCGCCA
trna10	Gly	GCC		76	GCGGGCGTAGTCTAGCCcGGAttAGGACGCCGGCCTCCCAGCGAGGGCCCCGGGTTCAAATCCCGGCGCCCGCCA
trna8	Gly	TCC		76	GCGGGCGTAGTCTAGCCtGGAtAGGACGCCGGCCTTCCAAGCCGGTGAtCCCGGGTTCAAATCCCGGCGCCGCCA
trna41	His	GTG		74	GCTCGGGTAGCTCAGCCtGGAAGAGCGCCGGGCTGTGAACCCGGAAGtCCCGGGTTCAAATCCCGGCCCCAGCC
trna22	Ile	GAT	39-50	86	GGGCCCGGTAGCTTAGCTtGGTGGAGCGCCCGGCTGATAGggtgtcagaaACCGGGAGtCACGGGTTCAAATCCCGTCGGGCCCA
trna18	Leu	CAA	40-55	100	GCGGGGGTGCCCGAGCtGGTcaAGGGGGCGGACTCAAGaccctaagtgagaAtCCGCTGGCGAAGGCCTtCGCGGGTTCAAATCCCGTCCCCGCA
trna13	Leu	CAG		85	GCGGGGGTGCCCGAGCtaGGTcgAAGGGGGtAGGCTCAGGCCCTACtGGTGAAGGCCttCGCGGGTTCAAATCCCGCCCCCCGCA
trna21	Leu	GAG		85	GCGGGGGTGCCCGAGCcaGGAGGGGGCGACTAGGCTTCGCCTGGCCAAGGCCTgGCGAAGGCCTgCACGGGTTCAAATCCCGTCCCCGCA
trna12	Leu	TAA	41-55	100	GCGGGGGTGCCCGAGCtcGTcaAAGGGGGGGACTTAAGaccccaagtaagagAtCCGCTGGCGAAGGCCTgCACGGGTTCAAATCCCGTCCCCGCA
trna16	Leu	TAG		85	GCGGGGGTGCCCGAGCtaGGTcaAAGGGGCAGGGTTAGGCTTTAAccatgtaaggaaagcggaatACCGTAGtCCCGGGTTCAAATCCCGTCCCCGCA
trna43	Lys	CTT	39-60	95	GGGCCCGTAGCTTAGCCaGTAGAGCGAGGGGCTCTTTAaccatgtaagaaaagcggataAGGCGCCTgGCGAAGGCCTtGCGGGTTCAATCCCGGCGGGCCCG
trna44	Lys	TTT	39-61	97	GGGCCCGTAGCTCAGCCaGTAGAGCGAGGGGCTTTTAacgcaaagggttgcggaaacaACCTAGtCCCGGGTTCGAATCCCGGCGGGCCCG
trna14	Met	CAT		74	GGGCCCGGTAGCTTAGTCGGTAGAGCCTCCGCTCATAACCCTCGGCCCA
trna39	Met	CAT	40-64	103	AGCGGGGTGGGCTTAGCCtGGtaAGGCCCGGGGGTCCATAtactcttaaagcacgcgttggaaACCCCGAGGtCCCTGGTTCAAATCCAGGCCCCGCTACCA
trna1	Met	CAT	40-56	92	GCCCGGGTCAGCCtGGtAaGGtacggcaATCCGGTTGtCGGGGTTCAAATCCCGGCGGGCA
trna29	Phe	GAA		74	GCCCGGGTAGCTCAGCCcGGAGAGCGCCCGGCTGAAGACCGGGTGtCCCGGGTTCAAGTCCCGGCGGGCCA
trna6	Pro	CGG		76	GGGTCCGTGGTCTAGCCtGGAcTAGGATGCGACGTTCGGGTCGTCGTGGTCCCGGGTTCAAATCCCGGCGGACCCA
trna23	Pro	GGG	41-61	97	GGGGCCGTAGTTAGCTtGGcAcTAGGATGCCAGCCTGGGataccaccagaaCGCTGGTGtCCCGGGTTCAAATCCCGGCGGCCCCA
trna1	Pro	TGG		76	GGGGCCGTAGCTtGGTaAGGCGCCCTGGGCTTGGGCCTCCAGacctttccoacgttactgcgaggaaATCCGGTTGtCGGGGTTCAAATCCCGGCGGGCCCCA
trna28	Ser	CGA	38-61	108	GCCGGGGTCCCCGAGCGGtCcAAGGGGtTAGGCTCGAaccttccoacgttactcgaggaAGCTACtGCCTCTCTGAGGCACACGGGTTCGAATCCCGTCCCCGG
trna46	Ser	GCT		85	GCcGAGGTCCCCAGCtaGGAtCTCCCGGGATCCTCCGGGCGCCTAAGCCGCTCAAGCGCTAAGCCGCCTCCGGCG
trna38	Ser	GGA		84	GCCGGGGTCCCCGAGCGacLAAGGGGGCTGGCCTGGCAGGGCATGGGGTTTcCCCGCGGCGGGTTCAAATCCCGCCCCCGCG
trna33	Ser	TGA		85	GCCGGGGTCCCCGAGCgGtGCTcCAAGGGGCTGGCCTTGAGCAGCCGATTAGGTGTCTaGCCTgCGGCGGGTtCAAGTCCCGGCCCGGCG
trna37	Thr	CGT	40-52	88	GCCGGGGTAGCTtCAGCCtcGtTLAGAGCGGCCTGTAAGGCAAGTCCCGGGTTCAAATCCCGGCCCGGCT
trna20	Thr	GGT		75	GCCGGGGTAGCTCAGCCtgGTcTAGGACGTGGCCTCCCGGGTAAGGCAAGTCCCGGGTTCAAATCCCGCGCGGCT
trna24	Thr	TGT	40-54	90	GCCGGGGTAGCTtCAGCCtcGtTGAGAGCGCCGCCTTGTAagttactgcctagatAGCCGGCGGtCGCGGGTTCGAATCCCGGCGGGCCCT
trna45	Trp	CCA	39-103	139	GGGCCGGGTAGCTTAGCCaGGTAGAGtACTGGGCTCCAGatattaatgctagacattggtcttagacccttatgggatgaagtgtatctggaggaggagACCCAGTGtCGCGGGTTCGAGTCCCGCCCGGCCCA
trna27	Tyr	GTA	41-53	89	CCCGGGCTAGCTtCAGGCtcGtTtAGAGCGCCGGCTGtAGataggagcagtaCCGGGTGGtCCGGGGTTCAAGTCCCGGCGGGGGA
trna3	Val	CAC		75	GGGCCCGGTCGTCTAGTLtGtTAGGACGTCGCCTCACGCGGAGATCCTGGGTTCAAGTCCCAGCCGGGCCCA

103

tRNA. The archaeal enzyme, on the other hand, specifically recognizes always a guanosine at position 15 both on type I and type II tRNAs irrespectively of the size of the D-loop. However it appears to act very inefficiently on eukaryotic tRNAs in comparison to archaeal or bacterial tRNAs, suggesting an additional, unidentified structural feature. The actual reaction mechanism was revealed by co-crystallization and mutation studies, showing that the totally conserved Asp[89] (*E. coli* numbering) acts as the nucleophile in a double-displacement reaction, followed by a deprotonation event to release the substrate again ([15] and references herein).

Archaeosine formation

Archaeosine modification occurs always at G15 of the Levitt basepair, which is the pairing between N15 located in the D-loop and N48 in the variable loop near the start of the T-loop (figure 1) [32], and is purely restricted to the archaeal domain. Its presence in tRNAs was discovered soon after the discovery of the distinction between bacteria and archaea [21], and this unique feature turned out to be widespread amongst the archaea [16, 33]. Because archaeosine has an unusual structure it took a while before the structure of this non-purine, non-pyrimidine base was solved using (collision-induced dissociation) mass spectrometry to be 2-amino-4,7-dihydro-4-oxo-7-*β*-D-ribofuranosyl-1*H*-pyrrolo[2,3-*d*]pyrimidine-5-carboximidamide (7-formamidino-7-deazaguanosine) (figure 2) [20]. Archaeosine modification is a two-step process that starts with the exchange of guanine at position 15 with the archaeosine precursor $preQ_0$ (7-cyano-7-deazaguanine) by the archaeal TGT (arcTGT) [15, 22, 29], followed by its conversion to archaeosine by addition of ammonia in an ATP independent manner. This step is catalysed by a novel amidotransferase called ArcS, which is a paralog of arcTGT and often annotated as *tgtA2* (figure 2) [13, 31]. Noteworthy is the fact that the SEED database [34] mentions only one species amongst the euryarchaeotes that misses an arcTGT orthologue: *Haloquadratum walsbyi*. This species also lacks archaeosine in their tRNAs as well [13].

Table 2. Primers used in this study.

Primer	Name	Sequence 5'→3' (with restriction site underlined)	Restr. site
Primers used for construct construction			
BG2009	*lacS* fw	CCGGCCGGATCCCTATATCAATCTCTTTTTGAAAGTGC	BamHI
BG2010	*lacS* rev	GCGCGCGGTACCGTAACAGTATTAAATCTAAATGAC	KpnI
BG2502	u. flank *tgt* fw	GGCCGGCCATGGATTAATAACAAAGGCGATAAAAGCA	NcoI
BG2503	u. flank *tgt* rev	CGCGCGGCGGCCGCCAGTCATAGACCTAATTCCTTTTTCTT	NotI
BG2504	d. flank *tgt* fw	GGCCGGTCTAGAGGCTTACATATAGACTCTATCAGCACT	XbaI
BG2630	d. flank *tgt rev*	CGCGCGCTGCAGAAAGATTTAAGTGAATTAAAAGGTTCTACA	PstI
Primers used in PCR analysis and Southern blot hybridization			
BG2637	*lacS* probe fw	GGGGGCCATGGACTCATTTCCAAATAGCTTTAGG	-
BG2638	*lacS* probe rev	CCATAGAGGTAATGGCCAATGATACATG	-
BG2724	*tgt* probe fw	AGGGTCAAATTTCTCTAGATAATCC	-
BG2725	*tgt* probe rev	TTCATCTATTTGGAGGAGGGTT	-

### Disruption of the archaeosine modification in vivo

The precise function of archaeosine modification is still unknown. Theoretical models indicated a role in tRNA stability [32]. To investigate the function of archaeosine *in vivo*, we constructed an arcTGT disruption mutant in *Sulfolobus solfataricus*, one of the first completely sequenced crenarchaeotes [35]. *S. solfataricus* is one of the few creanarchaeotes for which a genetic system has been developed, although studying a disruption mutant within this species can be quite a daunting task considering the relatively high genomic instability because of the presence of numerous transposons. Here we report the successful disruption of arcTGT in *S. solfataricus*. The effect of this disruption was confirmed by analyzing tRNAs for the absence of archaeosine by mass spectrometry. Growth comparisons revealed only minor differences, even under heat stress conditions. Against the backdrop of these results the role of this unique archaeal modification is discussed.

Material and Methods

Genetic manipulation of Sulfolobus solfataricus

A schematic overview of the disruption process is represented in figure 3. Strains used in this study were *Sulfolobus solfataricus* PBL2025 (PBL2025), a derivative of strain 98/2 with a deletion that spans the ORFs SSO3004 to SSO3050, including SSO3019 that encodes a beta-galactosidase (LacS) necessary for growth on lactose as carbon source [36], and *S. solfataricus* Δ*tgt* disruption mutant (Δ*tgt*).

Δ*tgt* was produced as follows. Flanking regions of *tgt* (SSO0274) and *lacS*, including promoter and terminator, were PCR amplified from genomic DNA of PBL2025 using the following primers: BG2009 and BG2010 for *lacS* amplification, BG2502 and BG2503 for the upstream flank, and BG2504 and BG2630 for the downstream flank (Table 1). The suicide
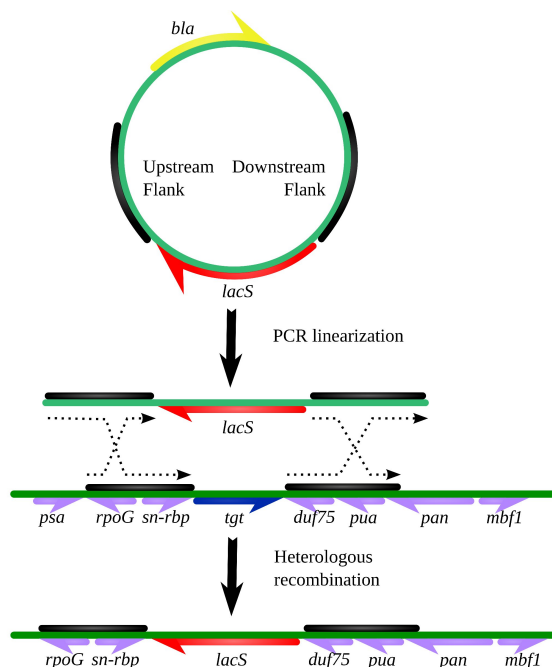
Figure 3. Schematic overview of the disruption process. Regions of heterologous recombination between flanks are indicated by dotted arrows, while the upstream and downstream flanks themselves are indicated with black bars. For clarity purposes the genes present in these flanks are not shown within the plasmid. *bla*: beta-lactamase (ampicillin resistance), *his-ts*: histidyl-tRNA synthetase, *psa*: Proteasome endopeptidase complex, beta subunit, *rpoG*: RNA-polymerase subunit G, *sn-rbp*: snRNA binding protein, *tgt*: tRNA guanine transglycolyase, *duf75*: protein of unknown function belonging to the PAC2 (Proteasome Assembly Chaperone) family, *pua*: pua domain of TGT, *pan*: proteasome activating nucleotidase, *mbf1*: multi-protein bridging factor 1, *hflX*: GTP-binding protein belonging to HflX family.

recombination plasmid was made by introducing the *lacS* gene placed between the upstream and downstream flanks into the multiple cloning site of pUC29. The final plasmid was checked by sequencing. To avoid single recombination events, linear PCR fragments were obtained by propagating the inserted region using M13-Forward and M13-Reverse primers. To confirm the absence of unintended errors the resulting construct was checked using Sanger sequencing.

Electroporation of *S. solfataricus* was essentially performed as described previously [37]. Electro-competent *S. solfataricus* PBL2025 cells were prepared from cultures grown on Brock's medium with 0.1% tryptone and 0.4% sucrose by washing the cells twice with a 20 mM sucrose solution. Linear PCR fragments were used to transform 50 µl of the competent cells by electroporation (2 mm, 1.5 kV, 400 Ω, and 25 µF). After electroporation, cells were immediately transferred to mQ water and placed for 1 minute on ice, followed by an incubation step of 10 minutes at 75°C, cells were transferred to pre-warmed medium containing 0.4% lactose. After initial growth cells were transferred to fresh 0.4% lactose containing medium, grown again to OD_{600} of approximately 1.0, and plated on gelrite plates containing 0.1% tryptone and 0.4% lactose. A number of colonies were picked, screened using colony PCR, grown in tryptone containing medium, and analysed for the presence of *lacS* and *tgt* by PCR. Colony PCR was performed using RedTaq (Invitrogen) according to the manual with the addition of 2.5% DMSO.

S. solfataricus growth

S. solfataricus strains were grown in modified Brock's medium at 75°C using an incubator (New Brunswick) shaking at 120 rpm, or an oil bath (New Brunswick) at 180 rpm [38]. Cells were grown with 0.4% sucrose and 0.1% tryptone as carbon source. To avoid evaporation Erlenmeyer flasks with elongated necks were used. In addition *S. solfataricus* was grown in tubes (30 mm diameter) containing 20 ml medium and closed with Silicosen T-32 plugs (Hirschmann Laborgeräte) to reduce water evaporation. Growth was monitored by measuring the optical density at 600 nm (OD_{600}) in a spectrophotometer U-1500 (Hitachi).

For sulphate limitation experiments, the medium composition was altered as follows. $MgSO_4$ was replaced by $MgCl_2$ (0.21 gl^{-1}) and the pH of the medium was adjusted with hydrochloric acid instead of sulphuric acid. In this way, the sulphate concentration was lowered to 1 µM compared to 11 mM in the original medium. Cultures were first grown in duplicate in 50 ml normal medium containing 0.4% sucrose, washed by low centrifugation (5,000×g, 3 min, room temperature) and resuspended in sulphate-limited medium, and then transferred in duplicate to tubes containing either sulphate limited medium with 0.4% sucrose in duplicate or standard medium with 0.4% sucrose serving as control. Growth rates and yields were calculated using the R statistical environment. Graphs are plotted using the ggplot2 R package (http://had.co.nz/ggplot2/).

Southern blot hybridization

Southern blots were performed according to a standard protocol [39]. Genomic DNA was extracted from *Sulfolobus* strains by phenol extraction as described elsewhere [38]. Genomic DNA was treated with RNase A (10 µg/ml) overnight at 4°C. Extracted DNA was digested with AflII (New England Biolabs) and with HindIII (New England Biolabs), and transferred to a nytran membrane (Perkin Elmer) via capillary transfer. DNA was immobilized by incubating the membranes for 2 hours at 90°C. PCR-generated probes against *lacS* (Table 1) were labelled with Digoxigenin using DIG High Prime according to the manufacturer's protocol (Roche). Prehybridization of the membranes was performed by incubating the membranes 4 hours in

hybridization buffer (50% freshly deionized formamide, 5×SSC, 2% blocking reagent (Roche), 0.1% Na-lauroylsarcosine, 0.02% SDS) at 60°C. Hybridization with the probes was done overnight at 60°C in fresh hybridization buffer. Membranes were washed twice in 2×SSC, 0.1% SDS, twice in 0.2×SSC, 0.1% SDS at 60°C, once in maleic acid buffer (0.1 M maleic acid, pH 7.5, 0.15 M NaCl) with 0.3% Tween20, incubated for 30 minutes in maleic acid buffer with 2% blocking reagent followed by an incubation for 30 minutes in the same solution with Anti-Digoxigenin-AP (Roche). Then they were washed again twice in maleic acid buffer with 0.3% Tween20. After two final washes and an additional wash in Assay buffer (100 mM diethanolamine, pH 10, 1 mM $MgCl_2$, 100 mM NaCl), membranes were incubated with 1:100 diluted CDP-star in the supplied buffer (New England Biolabs). Signals were captured on BioMax light films (Kodak) and films were scanned with a GS800 densitometer (BioRad).

tRNA extraction

Cells were disrupted by grinding with an equal mass of Al_2SO_3 on ice. After disruption, the ground cells were dissolved in 1.5 volumes of cold buffer A (20 mM Tris-HCl, pH 7.4, 10 mM MgOAc, 40 mM NH_4Cl, 1mM DTT). To degrade DNA, DNase I (Invitrogen) was added, incubated for five minutes at room temperature, and spun down 2 times for 30 minutes at 30,000 RCF to obtain a clear S30 extract. Ultracentrifugation tubes were filled up to approximately 8 ml using buffer A. Samples were centrifuged for 2 hours at 100,000 RCF (Sorvall ultracentrifuge, 38,000 rpm, TFT 65.13 rotor), and the supernatant was divided in 500 µl aliquots. RNAs were extracted using multiple Rotiaqua (Roth) extractions (1:1) until the white interphase disappeared, and then subsequently extracted using chloroform/IAA (25:1, Roth). RNAs were ethanol precipitated overnight, centrifuged for 15 minutes and resuspended in a combined volume of 1.0 ml 10 mM glycine (pH 9.0). Eventually 1/10th volume of 3 M Na-acetate pH 5.2 was added, and RNAs were again ethanol precipitated using 2.5 volumes of pure ethanol. RNAs were resuspended in standard QIAgen elution buffer.

tRNA Mass Spectrometry

S. solfataricus mutant and wild-type cells were grown in the presence of ^{14}N as well as in the presence of [15]N. tRNAs were extracted as described above. Two mixtures were prepared in a 1:1 ratio: one consists of wild-type [14]N- and wild-type ^{15}N-tRNAs, while the other was a mixture of mutant [14]N- and wild-type ^{15}N-tRNAs. Both mixtures were digested by RNase T1 and analysed using an ultra-high resolution time-of-flight (UHR-TOF) mass spectrometer in conjunction with online HPLC (figure 5A). Observed spectra were compared with theoretical masses, to indicate whether guanosine or archaeosine was present at the 15th position.

Theoretical monoisotopic masses of oligonucleotides after digestion with RNase T1 were predicted using the the 'mongo oligo mass calculator' (http://library.med.utah.edu/masspec/mongo.htm). Only those oligonucleotides that included the 15th base position, a guanosine or an archaeosine, were selected for analysis.

A Strain **WT** **Δtgt**

PCR *lacS tgt df uf 1kb lacS tgt df uf*

B **Strain** **WT** **Δtgt**

Digestion MWM PCR HindIII AflII HindIII AflII MWM
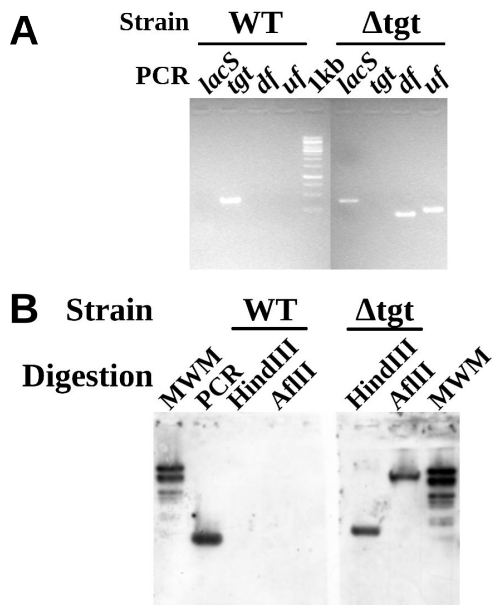
Figure 4. Validation of the tgt disruption. (A) PCR results: *lacS* and *tgt* indicate the respective set of primers that are located within the indicated gene. *df* and *uf* indicate primersets that are located within respectively the downstream and upstream flanks and within the *tgt* gene. (B) Southern analysis of digested genomic DNA, using a *lacS* targeted dig-labeled probe. PCR: positive control. MWM: DNA Molecular Weight Marker III (Roche). WT indicates no presence of *lacS*, while Δ*tgt* indicates its presence. Product sizes are in agreement with the expected values (HindIII: 718bp, AflII: >8000bp).

Results

Disruption mutagenesis

For the disruption, the complete *tgt* gene (SSO0274) was replaced by the selection marker *lacS*, which restores the capability for growing on lactose as carbon and energy source. PCR analysis (figure 4A) and Southern analysis (figure 4B) show that disruption mutagenesis of the *tgt* gene of *S. solfataricus* was successfully accomplished. The selection marker, *lacS* in *S. solfataricus*, was present, whereas the *tgt* gene could not be detected after mutagenesis.

During sequencing of the disruption plasmid, we noticed a reproducible alteration (T>G) in respect to the published genome sequence in the ORF adjacent to the *tgt* gene [35]. This alteration changes the stop-codon (TAG) of SSO0273 to a Glutamine (GAG), and results in a read through into ORF SSO5544. Both ORFs code for truncated halves of an enzyme belonging to COG1938 [PAC2 (proteasome assembly chaperone 2) family protein]. Most likely this corresponds to a sequencing error in the published genome of P2, as this read through was already suggested during the annotation process. The finding is supported by the recent deposited genome of *S. solfataricus* 98/2, in which also a glutamine encoding GAG-codon is present.

Archaeosine incorporation

To estimate the effect of the disruption of the tRNA modification, the amount of archaeosine incorporation was compared between the disruption mutant Δ*tgt* and the parental strain PBL2025 using mass spectrometry analysis. To produce the oligoribonucleotide mixture for mass spectrometry analysis, tRNA extracts were digested with RNase T1. Digestion with RNase T1 cleaves oligoribonucleotides at the 3' side of G residues but not of archaeosine residues. Oligoribonucleotides were analysed on an ultra-high resolution time-of-flight (UHR-TOF) mass spectrometer in conjunction with an online HPLC. The theoretical RNase T1 digestion of 28
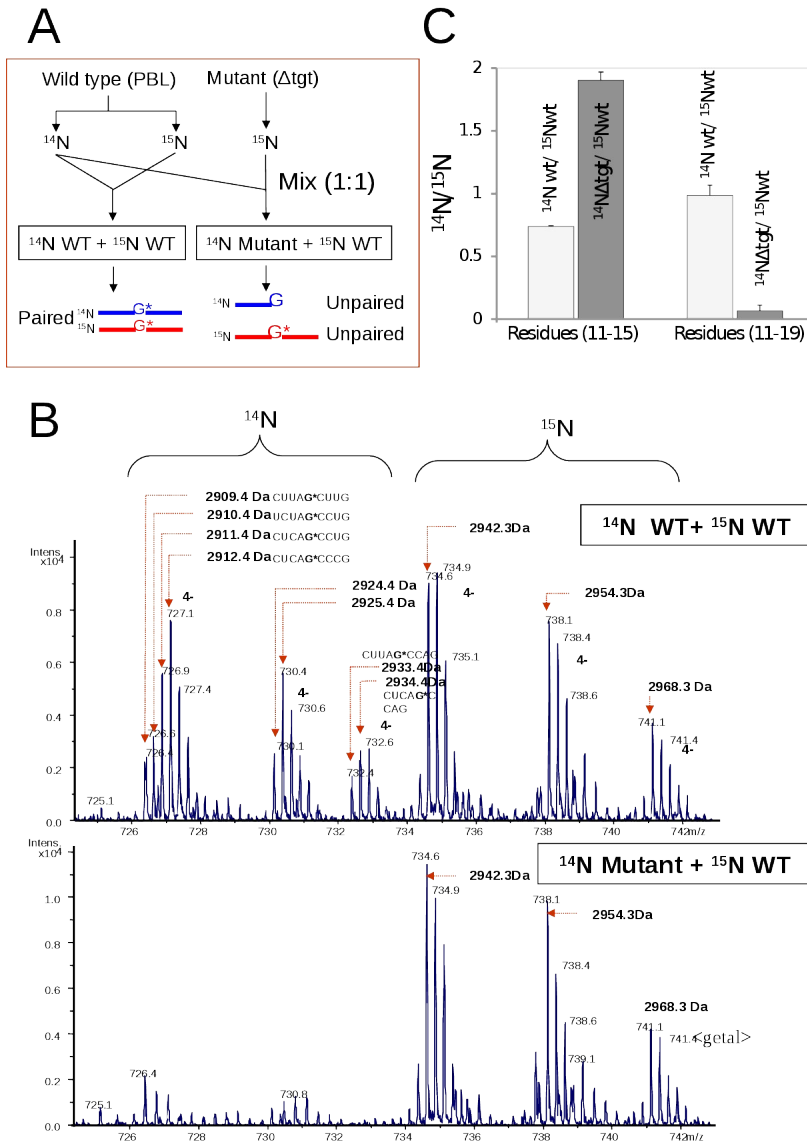
Figure 5. Archaeosine content measurements using mass spectrometry. (A) Schematic overview of experimental workflow of the mass spectrometry experiments showing the difference in expected outcome between the two pools of samples. RNase T1 normally digest the RNA molecule after guanine, but not after archaeosine, creating larger fragments from tRNAs that contain the modification than from tRNAs that do not. (B) UHR-TOF MS analysis of long oligoribonucleotides. The MS spectra show 3 doublet peaks from the two pools. In the upper spectrum one can see that if both the ¹⁴N sample and the ¹⁵N labelled sample contain archaeosine there is a duplication of the mass pattern. In the lower spectrum one can see that the unlabelled ¹⁴N pattern is hardly present, indicating the loss of archaeosine in the mutant. (C) Quantitative LC MS analysis. The relative abundance of oligoribonucleotides masses corresponding to (residues 11-15) and (residues 11-19) were calculated using the ratios represent the relative peak areas corresponding to the presence and absence of archaeosine in wild type and Δ*tgt* mutant tRNAs. The data represent the pooled results of two technical replicates.
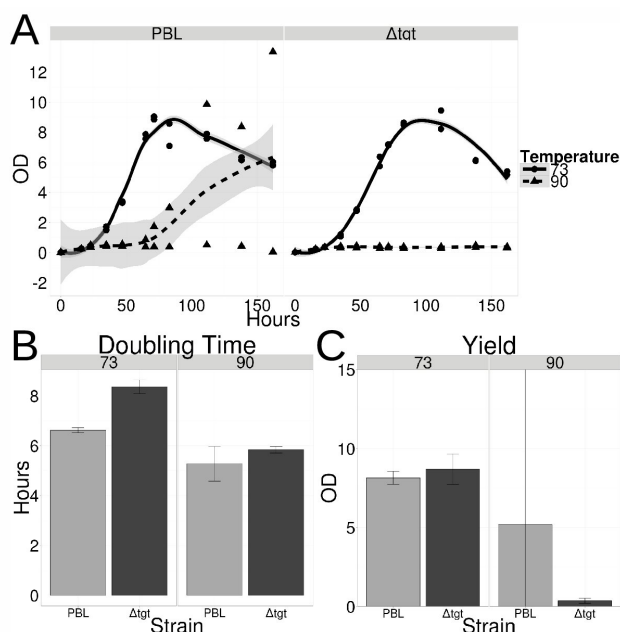
Figure 6. Growth characteristics of the Δtgt and PBL2025 strains. Under normal growing conditions (0.1% Trypton, 0.4% Sucrose, 73°C) at a constant temperature and under heat shock conditions using an elevated temperature (90°C), between 15 and 35 hours. (A) Growth curve (lines indicate the fitted model by local polynomial regression (loess), grey area surrounding the lines indicates 95% confidence interval, n = 2), (B) Doubling time in hours, (C) Yield (in OD_{600}) of the cultures.

tRNAs, using the mongo-oligo software, produced about 895 nucleotides ranging from Gp monomers to 16 mers. This *in silico* method showed that without the modification, shorter sequences, for example CUUAG and CUCAG, are present, while they are absent when archaosine is present, leading to longer fragments like CUUAG*CCAG and CUCAG*CCAG. The *in silico* method demonstrated that, although there is an additional level of complexity, it is possible to discriminate tRNAs that contain the modification from tRNAs that do not, in a total tRNA mixture digested with RNase T1.

Stable isotope labelling can be used to detect differences in the number of macromolecules between cell cultures in a qualitative manner [40]. Different cell cultures, grown in the presence of different nitrogen isotopes, can be mixed into one sample. Subsequent digestion and mass spectrometry analyses will therefore be performed in exactly the same manner under the same circumstances, eliminating technical variation. Molecules that are labelled with the slightly heavier ^{15}N have similar physico-chemical properties, except from a small increase in mass, and therefore behave similar, except from a small decrease in velocity, during the Mass Spectrometry experiment. Therefore during analyses metabolites from the different cell cultures are separated by exactly half the number of nitrogen atoms m/z units. Comparing peak height after normalization gives the relative difference in abundance of molecules present. A summary of the quantitative Mass Spectrometry analysis used in this study is shown in figure 5A. The relative abundance of oligoribonucleotides masses corresponding to residues 11-15 and residues 11-19 were calculated using the ^{14}N WT/^{15}N WT and the [14]N Δ*tgt*/^{15}N WT ratios, corresponding to respectively the presence and the absence of archaeosine in wildtype and Δ*tgt* mutant tRNAs.
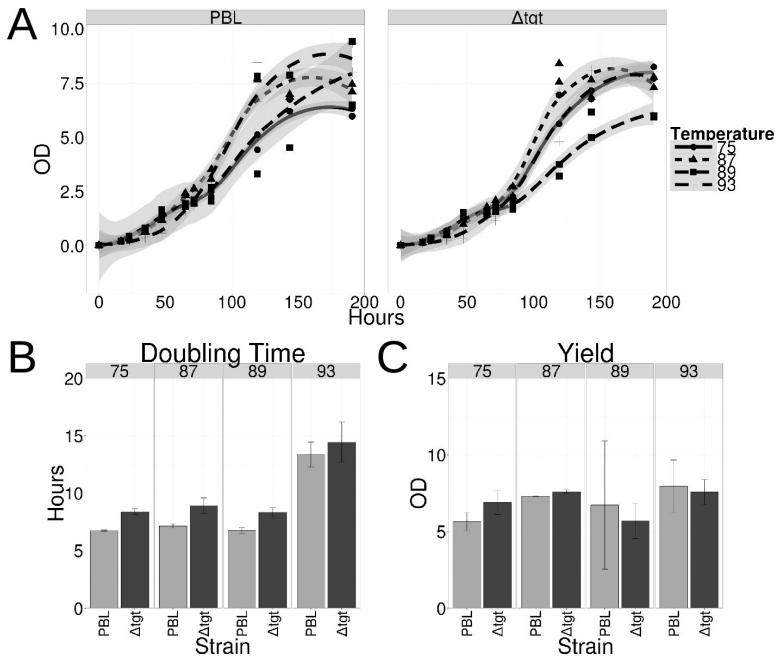
Figure 7. Growth characteristics of used strains. Similar to figure 6, however grown in (∅30mm) tubes (n = 2). Cultures are heat shocked for four hours at 5 different temperatures.

The increase in abundance of the ratio of the oligoribunucleotides masses corresponding to residues 11-15 of ¹⁴N Δ*tgt*/¹⁵N WT compared to the ones of ¹⁴N WT/¹⁵N WT in conjunction with the decrease in abundance of the ratio of oligoribonucleotides masses corresponding to residues 11-19 of ¹⁴N Δ*tgt*/¹⁵N WT compared to the ones of ¹⁴N WT/¹⁵N WT, demonstrate that the absence of archaeosine in the Δ*tgt* mutant tRNAs.

Growth characteristics

Having indicated the absence of archaeosine in our disruption mutant, the effect of this absence was investigated by comparing growth characteristics between Δ*tgt* and its parental strain PBL2025. At first sight growth of Δ*tgt* was not strikingly different from PBL2025 under standard laboratory growth conditions (figure 6ABC). The only major difference in growth was the ability of Δ*tgt* to grow on media containing lactose, unlike PBL2025, due the introduction of *lacS* as a selection marker during the disruption process. However a more detailed look at the doubling time revealed that it is repeatedly significantly higher in the mutant strain in comparison to PBL2025 (figure 6B, left panel). Despite the slightly slower growth, and as such the delayed entry into the stationary phase, the yield is similar to the parental strain.

The proposed effect on stability of archaeosine modification is tested under several stress conditions: nutrient stress, delayed transfer, antibiotic addition, and heat shock stress. Nutrient stress presses the cells to avoid any spillage of sparse nutrients, and as such might affect cells with a reduced tRNA stability more than cells with wildtype tRNA stability. Growth was monitored during nutrient and sulphate limitation (1.0 µM versus 11 mM). Under both

circumstances conditions no significant alteration in doubling time neither in yield was observed (data not shown), indicating that lack of nutrients does not affect the mutant strain more than the parental strain. Also survival during the stationary growth phase was tested by transfer of aliquots of a culture growing on 0.4% sucrose at different time points to new medium (of similar composition) in duplicate (data not shown) to test if reduced tRNA stability might affect fitness of cells in stationary phase. During stationary phase no reduction in fitness was monitored in comparison to PBL2025. This is unlike the Δ*mbf1* mutant, which showed a clear extension in lag phase that correlates to the time in stationary phase (see chapter 4). Also addition of antibiotics (paromomycin, and tetracycline) that interfere with the translation process did not reveal significant alterations in sensitivity to these antibiotics (data not shown). Under these conditions the mutation appears not to inhibit the strain more than the parental strain.

It was proposed that alterations on tRNA stability might affect temperature sensitivity. Because it is expected that tRNAs with decreased stability are more vulnerable to loss of conformation at higher temperatures, cells were subjected to heat shock treatments at several different shock temperatures. Heat shock stress was applied when growing strains reached an OD_{600} of approximately 1.0, by dividing tubes over water baths with different temperatures, ranging from 75°C to 95°C. They were incubated without shaking for four hours, after which they were transferred back to their original shaker at 75°C (figure 7ABC). This experiment was done in parallel to a similar experiment in which cultures where kept in culture flasks and incubated at 90°C for 20 hours in duplicate (figure 6BC, right panels). From these latter cultures only one PBL2025 culture was able to recover. Within the tubes, up to at 87°C and 89°C a slightly higher doubling time and equal yield levels were observed similar to the standard laboratory growth conditions described above. At 93°C both characteristics of the mutant were similar to PBL2025, showing that, like for the other stress conditions, tRNA stability is not the major rate limiting factor for the cells under those conditions. At 95°C growth in all cultures was severely impaired by the shock showing hardly any survival in all cultures.

Discussion

The generation of a *tgt* knockout mutant by homologous recombination in *S. solfataricus*, was successful. Despite its conservation throughout the three domains, and despite the overall presence of *tgt* among the archaea, *tgt* is not an essential gene in *S. solfataricus*, at least not under laboratory growth conditions. This result was in agreement with a study that recently reported on a *tgt* knockout in the Halophilic archaeal branch [41], and therefore we expect this gene to be non-essential - in the sense that a null mutant does not interfere colony formation [42] - throughout the archaeal domain.

The archaeal *tgt* gene has an essential role in the modification of tRNA G15 to archaeosine, of which the presence in tRNAs is considered as a hallmark for the archaeal domain. To check the effect of the knockout, and to investigate whether redundancy could explain the survival of the knockout mutant, we compared the presence of archaeosine in wildtype to its presence in the knockout using mass spectrometry. The results show clearly that the presence of archaeosine is completely diminished in this knockout, showing that arcTGT plays an essential role in the modification of tRNAs, apparently with no alternative TGT-like enzyme present.

It is believed that the archaeosine modification is important for the correct folding and stability of the tertiary structure of tRNAs in archaea. However, because cells are viable without

the modification, we carefully compared growth characteristics between the wildtype and knockout mutant. A reduced stability of tRNAs could impose a growth defect, as the abundance of correctly folded mature tRNAs is expected to be lower than in the wildtype, potentially impairing translation efficiency. It has been suggested that this impairment would be of particular importance for (hyper) thermophiles, like *S. solfataricus*, as a reduced tRNA stability is proposed to be especially devastating at higher temperatures [20]. During our tests the only difference was observed under standard laboratory conditions, and at only slightly elevated temperatures. Stress by nutrient limitation, extended stationary phase, antibiotic treatment, and heat shock treatment, showed no specific extra hindrance compared to the parental PBL2025 strain. One could imagine that a decrease in stability due to the absence of the archaeosine modification, leads to a higher abundance of erroneous folded tRNAs than in cells that have modification. When growth is optimal, and as such when translation is occurring at full speed, this situation could inhibit translation to a certain degree. It could hinder the aminoacyl-tRNA synthetases in their selection for the right tRNAs, or lower their processing rate. Or it could hinder the ribosomes in their selection of aminoacylated tRNAs, or any combination of these. Interestingly recent findings in a *H. volcanii* knockout mutant screen, an arcTGT knockout turned out to be cold-sensitive [41]. As molecule movements and conformation changes occur at lower levels at lower temperatures this might actually indicate that the chance of the ribosome of aaRS enzymes hitting a correctly folded tRNAs might be even lower than under standard conditions. This might indicate that archaeosine modification is actually not beneficial for growth at elevated temperatures, because the higher molecule movements might compensate for the instability of the molecule, but instead might be benefical to endure periods in which the temperatures is lower than optimal, because it might fix tRNAs in their correct conformation.

Acknowledgements

References

1. Crick FHC: **On protein synthesis.** *Symp Soc Exp Biol* 1956, **XII**:139–163.

2. Crick FHC: **Central dogma of molecular biology.** *Nature* 1970, **227**:561–3.

3. De Koning B, Blombach F, Brouns SJJ, Van Der Oost J: **Fidelity in Archaeal Information Processing**. *Archaea Vancouver BC* 2010, **2010**:1–34.

4. Ibba M, Söll D: **Aminoacyl-tRNAs: setting the limits of the genetic code**. *Genes Dev* 2004, **18**:731–738.

5. Fersht AR: **Enzymic editing mechanisms and the genetic code.** *Proc R Soc Lond B Biol Sci* 1981, **212**:351–79.

6. Jakubowski H, Goldman E: **Editing of errors in selection of amino acids for protein synthesis.** *Microbiol Rev* 1992, **56**:412–29.

7. Francklyn CS: **DNA polymerases and aminoacyl-tRNA synthetases: shared mechanisms for ensuring the fidelity of gene expression.** *Biochemistry* 2008, **47**:11695–703.

8. Korencic D, Ahel I, Schelert J, Sacher M, Ruan B, Stathopoulos C, Blum P, Ibba M, Söll D: **A freestanding proofreading domain is required for protein synthesis quality control in Archaea.** *Proc Natl Acad Sci U S A* 2004, **101**:10260–5.

9. Wong F-C, Beuning PJ, Silvers C, Musier-Forsyth K: **An isolated class II aminoacyl-tRNA synthetase insertion domain is functional in amino acid editing.** *J Biol Chem* 2003, **278**:52857–64.

10. Ahel I, Korencic D, Ibba M, Söll D: **Trans-editing of mischarged tRNAs.** *Proc Natl Acad Sci U S A* 2003, **100**:15422–7.

11. LaRiviere F, Wolfson A, Uhlenbeck O: **Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation**. *Science (80-)* 2001, **294**:165–168.

12. Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, Vendeix FAP, Fabris D, Agris PF: **The RNA Modification Database, RNAMDB: 2011 update.** *Nucleic Acids Res* 2011, **39**(Database issue):D195–201.

13. Phillips G, Crécy-Lagard V de: **Biosynthesis and function of tRNA modifications in Archaea**. *Curr Opin Microbiol* 2011, **14**:335–341.

14. Ibba M, Francklyn C: **Turning tRNA upside down: When aminoacylation is not a prerequisite to protein synthesis.** *Proc Natl Acad Sci U S A* 2004, **101**:7493–4.

15. Iwata-Reuyl D: **Biosynthesis of the 7-deazaguanosine hypermodified nucleosides of transfer RNA**. *Bioorg Chem* 2003, **31**:24–43.

16. Edmonds CG, Crain PF, Gupta R, Hashizume T, Hocart CH, Kowalak JA, Pomerantz SC, Stetter KO, McCloskey JA: **Posttranscriptional modification of tRNA in thermophilic archaea (Archaebacteria).** *J Bacteriol* 1991, **173**:3138–48.

17. Grosjean H, Björk GR: **Enzymatic conversion of cytidine to lysidine in anticodon of bacterial isoleucyl-tRNA--an alternative way of RNA editing.** *Trends Biochem Sci* 2004, **29**:165–8.

18. Grosjean H, de Crécy-Lagard V, Marck C: **Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes.** *FEBS Lett* 2010, **584**:252–64.

19. Helm M: **Post-transcriptional nucleotide modification and alternative folding of RNA.** *Nucleic Acids Res* 2006, **34**:721–33.

20. Gregson JM, Crain PF, Edmonds CG, Gupta R, Hashizume T, Phillipson DW, McCloskey JA: **Structure of the archaeal transfer RNA nucleoside G*-15 (2-amino-4,7-dihydro- 4-oxo-7-beta-D-ribofuranosyl-1H-pyrrolo[2,3-d]pyrimidine-5-carboximi dam ide (archaeosine)).** *J Biol Chem* 1993, **268**:10076–86.

21. Kilpatrick MW, Walker RT: **The nucleotide sequence of the tRNAMMet from the archaebacterium Thermoplasma acidophilum.** *Nucleic Acids Res* 1981, **9**:4387–90.

22. Watanabe M, Matsuo M, Tanaka S, Akimoto H, Asahi S, Nishimura S, Katze JR, Hashizume T, Crain PF, McCloskey JA, Okada N: **Biosynthesis of archaeosine, a novel derivative of 7-deazaguanosine specific to archaeal tRNA, proceeds via a pathway involving base replacement on the tRNA polynucleotide chain.** *J Biol Chem* 1997, **272**:20146–51.

23. Ishitani R, Nureki O, Nameki N, Okada N, Nishimura S, Yokoyama S: **Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme**. *Cell* 2003, **113**:383–394.

24. Kuchino Y, Kasai H, Nihei K, Nishimura S: **Biosynthesis of the modified nucleoside Q in transfer RNA.** *Nucleic Acids Res* 1976, **3**:393–8.

25. Okada N, Harada F, Nishimura S: **Specific replacement of Q base in the anticodon of tRNA by guanine catalyzed by a cell-free extract of rabbit reticulocytes.** *Nucleic Acids Res* 1976, **3**:2593–603.

26. Okada N, Noguchi S, Kasai H, Shindo-Okada N, Ohgi T, Goto T, Nishimura S: **Novel mechanism of post-transcriptional modification of tRNA. Insertion of bases of Q precursors into tRNA by a specific tRNA transglycosylase reaction.** *J Biol Chem* 1979, **254**:3067–73.

27. Chen Y-C, Kelly VP, Stachura S V, Garcia GA: **Characterization of the human tRNA-guanine transglycosylase: confirmation of the heterodimeric subunit structure.** *RNA* 2010, **16**:958–68.

28. Romier C, Reuter K, Suck D, Ficner R: **Crystal structure of tRNA-guanine transglycosylase: RNA modification by base exchange.** *EMBO J* 1996, **15**:2850–7.

29. Byrne RT, Waterman DG, Antson A a: **Enzyme-RNA Substrate Recognition in RNA-Modifying Enzymes**. In *DNA RNA Modif Enzym Struct Mech Funct Evol*. Edited by Grosjean H. Landes Bioscience; 2009:303–327.

30. Sabina J, Söll D: **The RNA-binding PUA domain of archaeal tRNA-guanine transglycosylase is not required for archaeosine formation**. *J Biol Chem* 2006, **281**:6993–7001.

31. Phillips G, Chikwana VM, Maxwell A, El-Yacoubi B, Swairjo MA, Iwata-Reuyl D, de Crécy-Lagard V: **Discovery and characterization of an amidinotransferase involved in the modification of archaeal tRNA.** *J Biol Chem* 2010, **285**:12706–13.

32. Oliva R, Tramontano A, Cavallo L: **Mg2+ binding and archaeosine modification stabilize the G15 C48 Levitt base pair in tRNAs.** *RNA* 2007, **13**:1427–36.

33. Sprinzl M, Hartmann T, Weber J, Blank J, Zeidler R: **Compilation of tRNA sequences and sequences of tRNA genes.** *Nucleic Acids Res* 1989, **17 Suppl**:r1–172.

34. Overbeek R, Begley T, Butler RM, Choudhuri J V, Chuang H-Y, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, et al.: **The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes.** *Nucleic Acids Res* 2005, **33**:5691–702.

35. She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CC, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PM, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, et al.: **The complete genome of the crenarchaeon Sulfolobus solfataricus P2.** *Proc Natl Acad Sci U S A* 2001, **98**:7835–40.

36. Schelert J, Dixit V, Hoang V, Simbahan J, Drozda M, Blum P: **Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon Sulfolobus solfataricus by use of gene disruption.** *J Bacteriol* 2004, **186**:427–37.

37. Albers S-V, Driessen AJM: **Conditions for gene disruption by homologous recombination of exogenous DNA into the Sulfolobus solfataricus genome.** *Archaea* 2008, **2**:145–9.

38. Zaparty M, Esser D, Gertig S, Haferkamp P, Kouril T, Manica A, Pham TK, Reimann J, Schreiber K, Sierocinski P, Teichmann D, van Wolferen M, von Jan M, Wieloch P, Albers S V, Driessen AJM, Klenk H-P, Schleper C, Schomburg D, van der Oost J, Wright PC, Siebers B: "**Hot standards" for the thermoacidophilic archaeon Sulfolobus solfataricus.** *Extremophiles* 2010, **14**:119–42.

39. Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning: A Laboratory Manual*. Second. New York: Cold Spring Harbor Laboratory Press; 1989.

40. Snijders APL, De Koning B, Wright PC: **Relative quantification of proteins across the species boundary through the use of shared peptides.** *J Proteome Res* 2007, **6**:97–104.

41. Blaby IK, Phillips G, Blaby-Haas CE, Gulig KS, El Yacoubi B, de Crécy-Lagard V: **Towards a systems approach in the genetic analysis of archaea: Accelerating mutant construction and phenotypic analysis in Haloferax volcanii.** *Archaea* 2010, **2010**:426239.

42. Jordan IK, Rogozin IB, Wolf YI, Koonin E V: **Essential genes are more evolutionarily conserved than are nonessential genes in bacteria.** *Genome Res* 2002, **12**:962–8.

43. Chan PP, Lowe TM: **GtRNAdb: a database of transfer RNA genes detected in genomic sequence.** *Nucleic Acids Res* 2009, **37**(Database issue):D93–7.

44. Schrödinger L: **The PyMOL Molecular Graphics System, Version 1.3**. 2010.

45. Shi H, Moore PB: **The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited.** *RNA* 2000, **6**:1091–105.

RNA Polymerase complex tagging *in vivo* in *S. solfataricus* without insertion of foreign elements

7

Bart de Koning, Ruben Engelhart, Stan J.J. Brouns, John van der Oost

Abstract

Last decade, the RNAP (RNA polymerase) from some archaea could be crystallized with reasonable resolution. Comparisons between these archaeal structures with their eukaryotic RNAP counterparts revealed that the archaeal RNAP is remarkably similar to the eukaryotic RNAP. This similarity makes the archaeal RNAP an interesting target to act as a model for the far more complex RNAPs of eukaryotes. Archaeal model systems could provide the opportunity to reconstitute active polymerases *in vitro* from homologously expressed subunits. Being assembled in their original manner in their native environment, they may provide more reliable insights into their native function and the effects of specific mutations thereupon than current heterologously expressed models do. This requires the availability of purification methods for RNAPs that can be used with reasonable ease. A solution to improve the efficiency is to tag one of the subunits to permit purification of the whole complex by a single affinity chromatography step. Therefore, the possibility of an additional tagging method was explored that would permit homologous protein purification and proteomic studies *in vivo* without introducing foreign vector sequences or replacing promoters. A his-tagged Rpo7 subunit mutant of *S. solfataricus* was successfully created by homologous recombination with a linearized suicide plasmid. To show potential usage, conventional histidine affinity purification was used to enrich a protein extract for RNAP complexes, and evidence for the purification of RNAP was provided by immuno detection of enrichment of the untagged Rpo2 subunit.

Introduction

Archaeal Polymerases

During the last decade great progress has been made in understanding the molecular mechanisms behind archaeal and eukaryotic transcription. Not in the least because the responsible multi-protein complex, the DNA directed RNA polymerase (RNAP), of some archaea could be crystallized with reasonable resolution. Comparisons between the archaeal structures with their eukaryotic RNAP counterparts revealed clearly what biochemists had postulated decades before: the archaeal RNAP is remarkably similar to the eukaryotic RNAP [1–10].

This similarity makes the archaeal RNAP an interesting target to act as a model for the more elaborate RNAPs of eukaryotes. Not only for general understanding about the molecular mechanisms of these complexes, but also, for example, in research for molecular behaviour of antibiotic compounds that act on the transcription machinery. The major advancement of using archaeal model systems lies within the possibility of *in vitro* reconstitution of heterologously expressed subunits to create active polymerases. This turned out to be possible for archaeal components or archaeal-eukaryotic chimeras, but is still not possible with subunits from purely eukaryotic origin. *In vitro* reconstitution for eukaryotic RNAPs can currently only be achieved with isolated RNAP complexes from the original eukaryotic cells, and as such are limited in their capabilities for molecular manipulation [11]. Using *in vitro* reconstituted polymerases lead to renewed understanding of the biochemical characteristics of these complexes, especially for the role of several subparts. A nice example is the Rpo4/7 stalk like structure, that is present in both eukaryotic and archaeal RNAPs [4, 6, 8, 12], and of special interest within the rest of this article. The Rpo4/7 stalk protrudes from the central structure and is required for RNAP activity at lower temperatures, probably by playing a central role in the transcription bubble formation [13].

The generation of high resolution structures allows for prediction of functional roles for specific parts and even for specific amino-acids of RNAPs, which can be verified by phenotypic effects of mutations in biological model systems. These high resolution structures enabled for example *in silico* prediction and development of antibiotic sensitivity [14]. Additionally, recent advances in reconstitution of the archaeal RNAPs *in vitro* from heterologously expressed subunits, made it possible to do mutation analyses of RNAPs in a high throughput manner [15]. These advantages already resulted in new insights in particular parts of the RNAP, like the highly conserved Bridge Helix. This structure seems to act as a nano-mechanical switchboard to coordinate catalysis and substrate movement [15]. This highly important feature was largely neglected before, as in the static crystal structures the bridge helix seemed a rigid structure with "not enough space to move". Mutagenesis revealed that in reality this structure used conserved hinges controlled by electrostatic forces to control the speed of catalysis [15, 16].

These studies were performed on reconstituted RNAPs from heterologously expressed subunits. Heterologous expression of subunits has a number of disadvantages over homologously expressed or natively purified complexes. Complexes that are homologously expressed or purified from organisms without genetic modification will be assembled in their original manner in their native environment and may provide more reliable insights about its native function and the effects of specific mutations thereupon. However, working with homologously expressed archaeal RNAPs, requires the availability of purification methods for RNAPs that can be used with reasonable ease, unlike the available approaches today. Providing easier methods for purification would enable homologous expression in a model system with prokaryotic advantages on cell culturing, scale of complexity, and genetic manipulation.

Purification

Purification is the bottleneck in archaeal RNAP research. The recent crystal structures were possible only after isolation procedures involving numerous purification steps [17]. That means that isolating RNAPs from archaeal cells requires not only large cell masses to obtain a reasonable quantity, the standard procedure involves at the moment a 115 L culture (500 g cells) and has a yield of about 0.2 mg per liter culture (23 mg total) [17], but also requires precious time, not to mention the high chance of failure.

Protein purification can be enhanced in several ways. The most popular method is to use affinity purification. Affinity purification is a method during which a protein extract is applied to a column that contains a matrix that binds the protein of interest. After washing, the eluate will contain an enriched fraction of this protein. Some proteins have affinities for certain materials themselves, but most often, an affinity tag has to be attached to the protein of interest, which can be achieved for example by genetic modification. A number of popular vectors for heterologous expression, both commercial as academic, already contain sequences for affinity purification tags, and the only step for successful use of those is to pay attention during construct design to keep the protein in frame with the affinity tag, and to avoid a stop codon in between.
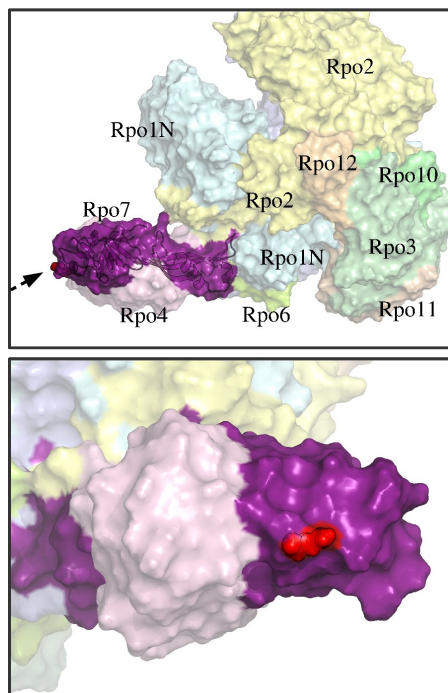
Figure 1. Structure of the archaeal RNA polymerase. In the upper panel is a structure shown from the complete RNA polymerase, in which the claw like core complex consisting of Rpo1 and Rpo2 is clearly visible, as well as the stalk like extrusion that consists of Rpo4 and Rpo7. The backbone of Rpo7 is visualized using a cartoon representation. The anchoring point of the histag is indicated with the arrow pointing at the red sphere representation of the last residue present in the structure (Q^{177}). This residue is followed by an additional TKKLHHHHHH peptide in our construct (not shown). The bottom panel shows a close-up of the Rpo4/7 stalk, using a view angle represented by the arrow in upper panel. Color representation is identical to the upper panel. The structure was originally obtained by Wojtas et al. [9]. Picture composition and rendering were done using PyMOL [39].

Several affinity tags exists and are commonly used (as reviewed in [18, 19]). Roughly one can divide them in four categories. The first category consist of the large affinity tags that target small ligands that can be fixated on a column. Within this category MBP (maltose binding protein) and GST (glutathione S-transferase) are most common. MBP is a protein that binds strongly to maltose or amylose, whereas GST binds to glutathione. These affinity tags are large (396 and 211 amino acids respectively). The main advantage of both tags, besides the relatively low costs for the resins, are their positive effects on solubility – they 'drag' their target into a soluble conformation –, on expression levels, and on binding capacities. The main disadvantage is that, due to their sizes, they often negatively influence the functionality of the protein, they might disturb complex formation, and are a considerable burden to the host when the tagged protein is overexpressed [18]. Nevertheless, both tags are regularly used to express proteins from *Sulfolobus* spp. heterologously in *E. coli*. Both tags cannot be used in the thermophilic organisms themselves, due to the poor thermostability of both proteins. Another tag, worth mentioning, within this category is the thioredoxin tag, which is originally the *trxA* gene from *E. coli*. The protein, that it encodes, keeps cysteines of substrate proteins, like ribonucleotide reductase, which is used in DNA metabolism, in their reduced form, and is conserved in all living species [20]. Although, strictly speaking, it is not direct an affinity purification tag, because it does not increase the affinity of the target to a certain resin, it is smaller than MBP or GST (109 amino-acids), and like MBP and GST it has a positive effect on the solubility and expression level of the target. Enrichment of the target protein is achieved by using a thermo-shock or osmotic shock, against which the target protein is protected by the tag [21]. Because this tag was known to be highly thermostable, it has recently successfully been used in *Sulfolobus islandicus* [22]. However, use in thermophilic species is of limited use, as most of the host proteins have increased thermostability as well. There exists also a modified version of this tag that exhibit binding properties to metal chelate resins, similar to the his-tag below [21].

The his-tag based system is the second category, and probably the most common tag used in protein purification. Like the MBP and GST tags it binds strongly to a small ligand: a double charged metal ion, especially Ni^{2+} or Co^{2+} . However, unlike the first category, the tag is very small, only six to eight histidine residues in a row, and therefore it hardly influences the chemo-physical properties of the target, and is less of a burden to the host during overexpression. Because of its strong binding to metals, this affinity purification is also known as immobilized metal affinity chromatography (IMAC). Resins that can be used to enrich his-tagged proteins, are readily available, reusable, and affordable. Elution from the resin can be achieved by supplementing the buffer with imidazole that out-competes the his-tag, strong chelating agents like EDTA, or a low pH. The first method is often the best choice as the other two might have a negative effect on the condition of the purified protein. The histidine tag is regularly used in *Sulfolobus solfataricus*, and is known to lead to high purification levels, as hardly any endogenous protein co-purifies [23].

The third category consists of only specific binding peptides (size ranges from 15 to 51 amino acids) that bind to protein-binding partners which can be immobilized on a resin. Popular examples include CBP (calmodulin binding peptide) and SBP (streptavidin binding peptide). CBP was originally obtained from the C-terminal fragment of the human Myosin Light-Chain Kinase (MYLK), and binds to calmodulin affinity resin in the presence of calcium. Addition of a chelator like EGTA, which removes calcium from the medium, induces a conformational change in the calmodulin that inhibits its binding to the peptide. This will elute the target protein from the resin. A noteworthy point though, is that calmodulin domains play an important role in eukaryotic cell signaling, so this affinity tag is largely unusable in eukaryotic cells. SBP was

obtained after a mRNA screen for streptavidin binding peptides, and was truncated to 38 amino acids, which turned out essential for binding. The binding peptide binds strongly to streptavidin, a protein purified from the bacterium *Streptomyces avidinii* with a very high affinity for biotin, that can be used eluate target proteins from the resin [24]. Due to the existence of streptavidin-fluorophores this tag can also be used for single step fluorescence detection, comparable to GFP-tagged proteins, but without the need to attach the larger GFP tag (38 versus 238 amino acids) [25]. Another similar tag that binds to streptavidin is the Strep II tag. The original 8 amino acid Strep II was selected from random sequences to bind strongly to streptavidin. More recently streptavidin was engineerd to Strep-Tactin to have a higher affinity for the Strep II tag, but still is the binding not as strong as the original SBP tag. Nevertheless the purity of this tag system is high, while the costs are moderate. Strep II turned out to be usable in *Sulfolobus solfataricus*, but unfortunately co-purifies the biotinylated acetyl-CoA/propionyl-CoA carboxylase, which is used for CO_2 fixation, but is also present under oxidative conditions [23].

Small peptides (8-11 amino acids) that belong to the fourth category are epitopes of common antibodies. The FLAG-tag, the HA-tag, and the c-Myc tag are most common examples. Main advantage of this category is the purity of the eluates, which is the highest amongst the four categories. However, the main disadvantage is the high cost. The resins, which contain antibodies against the used epitope, are very costly, and have a rather low binding capacity, certainly in comparison with the yields achieved with MBP, GST or the his-tag based systems [19].

A solution to improve the efficiency of RNAP purification is to tag one of the protein subunits. Tagging single subunits to purify complete complexes is a proven strategy, and even used to identify elements not associated to the complex before [25, 26]. Because the RNAP forms a quite stable core complex, tagging one of the proteins might therefore permit purification of the whole complex by a single affinity chromatography step. The tag chosen for our proof of principle experiment is the his-tag. This tag is easy to use, has a long list of achievements, uses a resin that is readily available and affordable, and is known to work in *S. solfataricus* with only a low amount of endogenous protein to co-purify.

Genetic Modification in Sulfolobus solfataricus

S. solfataricus is an aerobic hyperthermophilic, acidophilic crenarchaeote, which complete genome is already known since 2001 [27]. It grows at an optimum temperature of 80°C at a pH optimum of 4. Culturing *S. solfataricus* is therefore relatively easy: culture contamination is almost neglectable – although this advantage is lost when culturing multiple different strains of *S. solfataricus* simultaneously –, whereas there is no need to take measurements against the presence of oxygen in the environment. Genetic modification of this species turned out to be possible by the use of electroporation, however the genetic toolbox of this species is still far from being established. One cause is the lack of selection strains, what makes the number of genetic possibilities very limited [28]. Another cause is the abundant presence of transposons [27, 29], what makes the genome of *S. solfataricus* instable to work with [30, 31]. Both issues make that genetic manipulation of *S. solfataricus* is lacking behind its euryarchaeotic nephews like *Haloferax volcanii* or *Thermococcus kodakarensis*, which have a much more elaborate genetic system [32–34]. A recent project to establish a systems biology approach for *S. solfataricus* changed things for the better, by establishing default laboratory conditions for culturing, manipulating, and doing transcriptomics and proteomics experiments in this organism [30]. Additionally, the project described in this paper, aimed to explore the possibility of an additional tagging method that would permit homologous protein purification and proteomic studies *in vivo* without introducing foreign vector sequences or replacing promoters.

Tagging

To tag a single subunit from a stable complex from *S. solfataricus*, like RNAP, with the purpose to purify the complex as a whole, three strategies could be used: (1) heterologous expression of a single tagged subunit, and subsequently use it as a bait by mixing it with a wildtype *S. solfataricus* extract, (2) homologous expression of a tagged subunit using the SSV1 virus overexpression system, or (3) insertion of a tag sequence in the host genome using homologous recombination.

(1) The subunit is expressed in a conventional overexpression strain like *E. coli*. After purification the purified subunit is mixed with a cell extract from *S. solfataricus* to catch the whole complex using affinity chromatography. The advantage of this method is that no genetic manipulation of *S. solfataricus* is necessary. However, overexpression could fail due to a multitude of reasons ranging from inefficient codon usage to protein stability at moderate temperatures. It has also been found that mesophylic expression of proteins from thermophilic origin leads to underperforming proteins [22]. So even when it is successful, the protein might be misfolded and unable to integrate into the native complex during the mixing with the cell extracts.

(2) The subunit is overexpressed in *S. solfataricus* using the SSV1 overexpression system. This overexpression system has been established in last decade and appears to be relatively easy to use [23, 34]. It makes use of a vector that is based on *S. solfataricus* virus SSV1 . This virus integrates into the genome, and will transcribe, amongst some other viral proteins, the protein of interest. The virus spreads through the culture, infecting the majority of the present *S. solfataricus* cells, and thereby spreading expression of the tagged protein throughout the culture. A frequently used system uses the *ara* promoter and the unidentified corresponding arabinose dependent transcriptional regulator, which responds to the arabinose content of the medium, and is tunable [23, 34, 35]. A major advantage is that overexpression is accomplished in *S. solfataricus* cells itself, and thereby circumventing the disadvantages of using heterologous expression systems. The major disadvantage is the use of a viral vector that integrates into the host genome [35]. The culture will be affected by the viral invasion, making monitoring for phenotypic changes by overexpression less ideal. Additionally, the original untagged protein is still present too and might outcompete the tagged protein in complexes.

(3) The third method is the fusion of a tag at C-terminus or N-terminus of the protein at its original location in the genome. Because there is no viral invasion the culture will be less affected by the integration procedure, there will be no untagged proteins present that potentially could affect the efficiency, and the protein will remain in its genomic context under control of its own promoter. When desired, promoter exchange can be achieved by exchanging the promoter within the suicide plasmid with a high expression or inducible promoter, like the *ara* promoter used in the viral overexpression system. On the other hand, in comparison with the other methods, the manipulation procedure is the most difficult, being less established in comparison with the systems regularly used for the first method and more tedious in comparison to the second. Further more, without promoter exchange, there is no overexpression, and as such expression will be at endogenous levels. The effects on the culture are therefore minimal, but expression could be extremely low or might even be absent under certain circumstances, and as such might be hard to detect. Additionally, if the tag itself is harmful or located on a position that interferes with the biological function of the target, the colony might react similar as a knockdown strain. Last, but not least, because markerless integration is still not established in *S. solfataricus*, a marker will have to be integrated in the genomic neighbourhood of the gene of interest. This could interfere with other genes located in

this neightbourhood, and the expression of the marker, often LacS with a strong promoter, might affect the culture as well.

This project was aimed to provide a proof of principle for the third method. We have created a his-tagged Rpo7 subunit mutant of *S. solfataricus* by homologous recombination with a linearized suicide plasmid. To show potential usage a simple conventional bench top histidine affinity purification step was used to enrich a protein extract for RNAP complexes. Evidence for the purification of RNAP is provided by immuno detection of another, untagged, subunit (Rpo2), which showed enrichment for Rpo2 in the mutated strain in comparison with its parental strain..

Material and Methods

Homologous recombination of Sulfolobus solfataricus

Strains used in this study were *Sulfolobus solfataricus* PBL2025 (PBL2025), a derivative of strain 98/2 with a deletion that spans the ORFs SSO3004 to SSO3050, including SSO3019 that encodes a beta-galactosidase (LacS) necessary for growth on lactose as carbon source [36], and *S. solfataricus rpoE*-his (*rpoE*-his).

RpoE-his was produced as follows. *RpoE1* (SSO0415, coding for Rpo7), containing an additional histidine tag, and *rpoE2* (SSO5798, coding for a Spt4 like protein), a downstream flanking region, and *lacS* (SSO3019, coding for the LacS marker), including promotor and
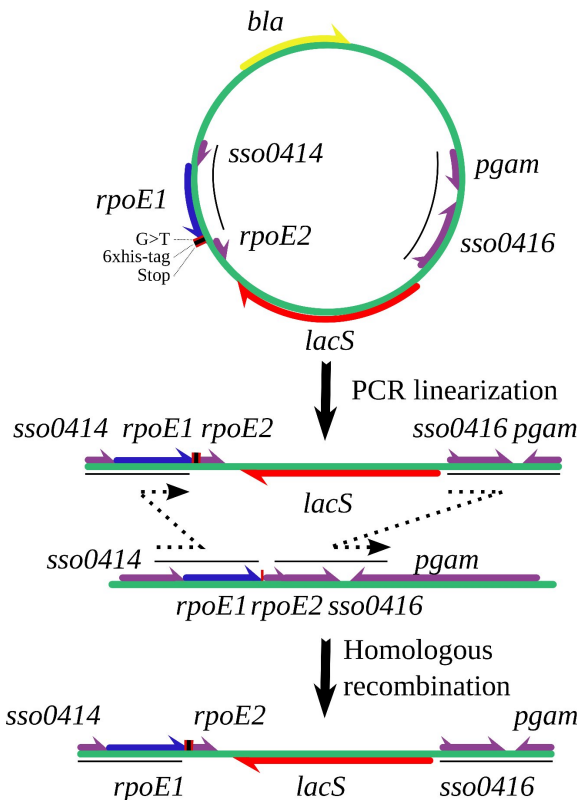
Figure 2. Schematic overview of the his-tag integration procedure. Genes are depicted by horizontal arrows. Regions of homologous recombination are indicated by dotted arrows. The former stop codon (G>T) and the newly inserted stop codon are represented by vertical red bars. The inserted 6xhis-tag is represented by a vertical black bar. Upstream and downstream flanks are indicated with horizontal black bars. Represented genes are as follows: *bla*: beta-lactamase (ampicillin resistance, yellow arrow), *rpoE1*: RNA-polymerase subunit E/Rpo7 (blue arrow), *rpoE2*: Transcription elongation Spt4 like protein, *lacS*: beta-galactosidase (selection marker, red arrow), *sso0414*: putative Fcf1 protein, involved in pre-rRNA processing, *sso0416*: Protein of unknown function DUF359, *pgam*: Phosphoglycerate phosphomutase.

terminator, were amplified by PCR from genomic DNA using the following primers: for *lacS* amplification BG2009 and BG2010, for *rpoE1* BG2506 and BG2507, which contains a G>T mutation to suppress the natural stop codon, a tagging sequence that codes for 6 histidines, and a novel downstream stop codon, for *rpoE2* BG2508 and BG2509, and for the downstream flank BG2510 and BG2511 (table 1). The suicide recombination plasmid was made by introducing the *lacS* gene placed between the upstream and downstream flanks into the multiple cloning site of pUC29. The final insert was checked by Sanger sequencing. To avoid single recombination events linear PCR fragments were generated by propagating the inserted region using M13-forward and M13-reverse primers.

Electroporation of *S. solfataricus* was performed as described previously [37]. Electrocompetent *S. solfataricus* PBL2025 cells were prepared from cultures grown on Brock's medium with 0.1% tryptone and 0.4% sucrose by washing the cells twice with 20 mM sucrose solution. Linear PCR fragments were used to transform 50 µl of the competent cells by electroporation (2 mm, 1.5 kV, 400 Ω, and 25 µF). After electroporation, cells were immediately transferred to mQ water and placed for 1 minute on ice, followed by an incubation step of 10 minutes at 75°C, cells were transferred to prewarmed medium containing 0.4% lactose. After initial growth, cells were transferred to fresh 0.4% lactose-containing medium (1:100), grown again to OD_{600} of approximately 1.0, and plated on gelrite plates containing 0.1% tryptone and 0.4% lactose. A number of colonies were picked, screened using colony PCR, grown in tryptone-containing medium, and analysed for the presence of *lacS* and *rpoE-his* by PCR. Colony PCR was performed using RedTaq (Invitrogen) according to the manual, with the addition of 2.5% DMSO.

Protein extraction and purification

Cells were disrupted in 6.5 ml Buffer A [50 mM Tris/HCl, pH 7.8, 25 mM $MgCl_2$, 150 mM KCl, 10% (v/v) glycerol, cOmplete Mini EDTA-free Protease Inhibitor Cocktail (1 tablet / 30 ml, Roche)] using French Press three times on ice. To prepare cell free extracts cells were spun down at 16.1 kRCF for 30 minutes.

Table 1. Primers used in this study. u: upstream, d: downstream. Restriction sites are underlined within the primer sequence. The 6×his sequence is shown in italic.

Primer	Name	Sequence 5'→3' (with restriction site underlined)	Restr. site
Primers used for construct construction			
BG2009	*lacS* fw	CCGGCC<u>GGATCC</u>CTATATCAATCTCTTTTTGAAAGTGC	BamHI
BG2010	*lacS* rev	GCGCGC<u>GGTACC</u>GTAACAGTATTAAATCTAAATGAC	KpnI
BG2506	u. flank *rpoE1* fw	GGCCGG<u>CCATGG</u>ATTAAGATAGCCAGAAAATATTTGGAA	NcoI
BG2507	u. flank *rpoE1*-TTA-6×his-stop rev	GCGCGC<u>ATGCAT</u>TCA*GTGATGATGGTGGTGATG*TAACTTCTTAGTCTGTGTTATCCACTC	NsiI
BG2508	*rpoE2 fw*	GCGCGC<u>ATGCAT</u>CAGACTAAGAAGTGAGTATAAATGGCT	NsiI
BG2509	*rpoE2 rev*	GGCCGG<u>GCGGCCGC</u>TTACTTTATTATTATTGCGTACTTCCA	NotI
BG2510	d. flank *rpoE1* fw	GCGCGC<u>TCTAGA</u>ATCATTAATTCAGAGTCTGAAATAGCC	XbaI
BG2511	d. flank *rpoE1 rev*	GGCCGG<u>CTGCAG</u>GATGCAGCATCACATGATGG	PstI
Primers used in PCR analysis			
BG2699	*rpoE1*-his	GCTTTAACAATGAGGCAACCCTAC	-
BG2509	*rpoE2 rev*	GGCCGG<u>GCGGCCGC</u>TTACTTTATTATTATTGCGTACTTCCA	NotI

Benchtop columns (BioRad) were prepared using 1 ml HIS-Select nickel affinity gel resin (Sigma), leaving 500 µl resin in the column (CV). Columns were prewashed using Buffer A + 20 mM Imidazole. Cell free extracts were applied to the column, washed using 6 CV Buffer A + 20 mM Imidazole and eluted 2 times 2 CV Buffer A + 500 mM Imidazole.

Immunodetection

PBL2025 and *rpoE-his* cell extracts were separated by bis-Tris SDS-PAGE and transferred overnight to Nitrocellulose membranes (0.2 µm pore size, BioRad) in 10 mM CAPS (pH 11.0), 10% methanol at 10 V using a tank transfer system. Efficient transfer was verified by PonceauS staining. Filters were washed twice in TBS (20mM Tris-HCl, pH 7.9; 150 mM NaCl) and incubated for 1hr in blocking solution [0.2 % i-block (Applied Biosystems), 0.1 % Tween-20 in TBS]. Filters were then incubated with anti-Rpo2 antibody, which was kindly provided by Stephen Bell (Oxford, UK), diluted 1:10000 in blocking solution for 1 hr. After three 15 min washes in 50-100 ml TTBS (TBS, 0.2 % Tween-20), filters were incubated with Anti-Rabbit-AP, Fab fragments (Roche) diluted 1:1500 in blocking solution for 30 min. After three more TTBS washes and an additional 5 min TBS wash detection of Rpo was performed by chemiluminescence using 1:100 diluted CDP-Star reagent (New England Biolabs) and Kodak Biomax films for signal capture.

Results

Genetic manipulation of S. solfataricus

To create the suicide plasmid a progressive approach was used, in which we integrated part for part into the plasmid. Although this approach takes longer to complete, the chance of success is higher than when using a multiple ligation step. The single exception, in which we used a triple ligation, was to integrate the fragments that contained *rpoE1*, with the additional his-tag, and *rpoE2*. As can be seen in figure 3, this final plasmid was created successfully, and was checked using Sanger sequencing. After this confirmation, the plasmid was linearized by PCR and used as a recombination fragment in *S. solfataricus* PBL2025.

Several colonies were obtained that showed a restored capacity to grow on lactose. They were screened using colony PCR for having the intended size of the fragment containing *rpoE2*. In the plasmid design, a small duplication containing the ribosome binding site was introduced to avoid interruption of translation of RpoE2. The resulting small increase in size is detectable on a high percentage agarose gel. All cultures contain the *lacS* gene, however not all cells did integrate the tag properly, which was visible by the dual bands on agarose gels in these test (figure 3). Further selection rounds by selecting colonies that where enriched for the larger, correct, fragment, were necessary to obtain a pure culture with the intended insert.

Protein extraction and purification

Cell free extract preparation was achieved by standard techniques and showed a huge quantity of soluble proteins. Although most of these soluble proteins did not bind to the nickel column, the remaining enriched RNAPs were nevertheless below detection levels for coomassie blue staining. Silver staining showed that proteins were present at sizes that correspond to subunits of RNAPs, but at detection levels only slightly above background levels (data not shown).
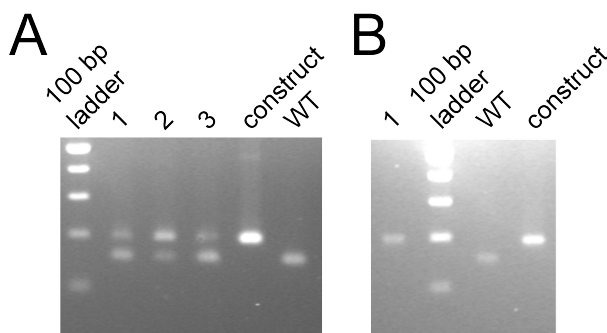
Figure 3. PCR analysis results from the mutant construction in *S. solfataricus*. (A) After first round of selection by growing on lactose containing medium, two separate insert sizes can be detected differing only 20 basepairs. The largest covers the G>T mutation, the 6xhistag, and the newly inserted stop codon, while the smallest has the wildtype configuration. The creation of this chimera is most probably caused by different sites of crossing-over occurring during the selection. (B) After several rounds of growing and transmission to fresh medium, only the intended construct is left. 100 bp Ladder: 100 bp size marker; WT: parental strain; construct: plasmid construct used for genetic manipulation; 1, 2, and 3: mutant strains.

Subsequent western blot analysis comparing both the *rpoE*-his and its parental strains, showed the presence of the untagged Rpo2 subunit in the eluate obtained using the *rpoE*-his strain, whereas it was absent in the eluate obtained from purifications of the parental strain (figure 4).

Discussion

Insertion of an in-frame his-tag behind the gene that encodes the Rpo7 subunit, the archaeal ortholog of the eukaryotic RNAP II Rpb7, was achieved using homologous recombination. Histag affinity purification of cell extracts from the resulting *rpoE*-his strain resulted in a strong enrichment of the large Rpo2 subunit in comparison to a similar procedure done on cell extracts obtained from its parental strain. This enrichment of the untagged Rpo2 subunit is most likely due to enrichment of the complex as a whole by anchoring of the Rpo7 subunit to the HIS-Select resin during the purification process.

Together with Rpo4, Rpo7 forms the stalk like structure protruding from the complex core [4, 6, 7, 9]. Therefore the his-tag is located at the furthermost outward place possible, and still strongly associated to the complex, making as much contact to the resin as possible while interfering with the complex catalytic function as less as possible (figure 2). To see if the his-tag insertion does not indeed affect the physiological function of the RNAP, *in vitro* transcription still has to be tested on purified RNAP complexes. To achieve this successfully, purification in larger quantities and to higher purity would be necessary, which was outside the scope of this initial proof of principle study. On the other hand, growth impairment of the mutant strain in comparison to the parental strain is not observed (data not shown). This indicates the presence of functional RNAPs in quantities at least above the minimal levels for abundant growth. As our insertion method abolishes the presence of wildtype – untagged – Rpo7 subunits, the physiological function cannot be severely affected by the insertion of the tag.
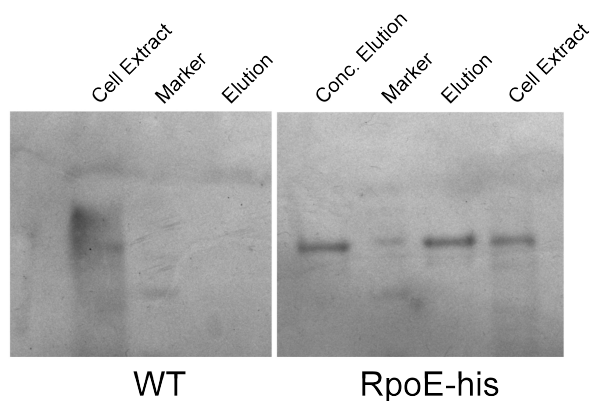
Figure 4. Western blot analysis comparing the outcomes of a bench top histidine affinity purification of the mutant strain in comparison to its parental strain. The primary antibody was directed against RNA polymerase subunit B, which is present in the elutions obtained with RpoE tagged strain, and absent in the other.

Due to the absence of markerless integration, obtaining a mutant in such a way as described above is not straightforward. Suicide plasmid design construction is complicated, because of the genomic neighbourhood of the gene. *S. solfataricus*, like other prokaryotic species, has a tightly packed genome, and, therefore, genes overlap on numerous occasions. Directly adjacent to *rpoE1*, is the small *rpoE2* gene located. This little gene encodes for a small protein (RpoE2), and, unlike what the name suggests, this protein is not in any way associated to RNAP. Due to historical reasons it was once believed to be part of Rpo7, which was formerly known as RpoE1 [3]. Closer inspection revealed, that it was a separate gene, and further investigation showed that the gene was, in fact, not even associated with transcription at all, but with translation instead, as it was found to act on ribosomes [4, 6, 38]. To avoid any unintended side-effects on the translational level, the upstream region, in which the putative ribosomal binding site of the *rpoE2* gene is located, was kept intact by duplication of this region, what added to the complexity of the procedure (figure 1).

This duplication could cause the occurrence of crossing-over between the suicide plasmid and the chromosome in the *rpoE2* gene located after the insertion of the his-tag, and as such would avoid integration of the his-tag into the genome. Such an event could explain the dual insertion band observed after the initial selection rounds. During both recombination events, the intended, where recombination takes place before the tag, and the unintended, where recombination takes place after the tag, the selection marker (*lacS*) is integrated into the genome. Therefore, both recombinants are capable to grow on lactose, but show a small product size difference in PCR results, which is visible on high percentage agarose gels. Recombination after the tag results in fragments with a smaller, wildtype, size, whereas recombination before the tag – which includes the tag in the exchange – results in the intended, larger, size. During our initial selection a chimeric culture arose that did not only consist of cells harboring the selection marker and the inserted his-tag but also of cells harboring the selection marker without insertion of the his-tag. Multiple rounds of selection were necessary to obtain a pure culture containing both the selection marker and the inserted his-tag.

Eventually this resulted in a stable integration of the his-tag within a mutant cell culture. The resulting culture did not show any observable growth hindrance in comparison with its parental strain PBL2025. Natural expression levels of RNAPs are low. Even though a dense protein mixture was extracted out of our cells and the maximum capacity of our bench top purification system was used, protein levels were too low to detect immediately. Western Blot nevertheless showed a clear enrichment for the untagged Rpo2 subunit. Enrichment of the Rpo2 subunit by affinity purification of a different histagged subunit, shows that we in fact enriched the RNAP as a whole.

The goal of this project was to see if it was possible to enable easy purification of protein complexes, but leaving the cell as close to the parental strains as possible. Although, it was possible to detect enrichment of the untagged Rpo2 subunit and as such provide a proof or principle of this technique within *S. solfataricus*, it is not yet of a very particular use to *in vitro* RNAP research, as the yield of the purification is still too low. Larger quantities should increase the yield easily, but will introduce the necessity to implement further purification methods to decrease the impurities that will increase as well. A solution to make it applicable for *in vitro* research, could be promoter exchange. If the natural promoter of RNAP is exchanged with a promoter known to have high yields of protein, this system could potentially be used for protein overexpression. This overexpression could have the benefit over the already established viral overexpression system that no virus particles are used to integrate foreign DNA into the organism. Two remarks are worth mentioning against such an approach. The first one, is that creation of the suicide plasmid and the modification process are likely to be much more complicated than doing something similar using the viral system. And secondly, for complexes like RNAP, for which the encoding genes are not neatly clustered together but scattered over the genome, one of the other components of the complex might become lacking, reducing the yield of complex purification, with even the danger to create imbalances in RNAP subunit stoichiometry, which could result in experimental artefacts.

Nevertheless, this project showed that *in vivo* tagging of proteins by incorporation of tags into the genome of *S. solfataricus* using homologous recombination is possible. Tagging stable protein complexes in such manner, is likely to enhance easy purification without the danger of huge unintended effects caused by the manipulation procedure, or the danger of competition between tagged and non-tagged proteins, and as such losing yield to untagged wildtype remnants.

References

1. Bartlett MS: **Determinants of transcription initiation by archaeal RNA polymerase.** *Curr Opin Microbiol* 2005, **8**:677–84.
2. Zillig W, Stetter KO, Janeković D: **DNA-dependent RNA polymerase from the archaebacterium Sulfolobus acidocaldarius.** *Eur J Biochem* 1979, **96**:597–604.

3. Korkhin Y, Unligil UM, Littlefield O, Nelson PJ, Stuart DI, Sigler PB, Bell SD, Abrescia NGA: **Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure.** *PLoS Biol* 2009, **7**:e1000102.

4. Werner F: **Structure and function of archaeal RNA polymerases.** *Mol Microbiol* 2007, **65**:1395–404.

5. Pühler G, Leffers H, Gropp F, Palm P, Klenk HP, Lottspeich F, Garrett RA, Zillig W: **Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome.** *Proc Natl Acad Sci U S A* 1989, **86**:4569–73.

6. Goede B, Naji S, von Kampen O, Ilg K, Thomm M: **Protein-protein interactions in the archaeal transcriptional machinery: binding studies of isolated RNA polymerase subunits and transcription factors.** *J Biol Chem* 2006, **281**:30581–92.

7. Hirata A, Klein BJ, Murakami KS: **The X-ray crystal structure of RNA polymerase from Archaea.** *Nature* 2008, **451**:851–4.

8. Hirata A, Murakami KS: **Archaeal RNA polymerase.** *Curr Opin Struct Biol* 2009, **19**:724–31.

9. Wojtas MN, Mogni M, Millet O, Bell SD, Abrescia NGA: **Structural and functional analyses of the interaction of archaeal RNA polymerase with DNA.** *Nucleic Acids Res* 2012, **40**:9941–52.

10. Albers S-V, Forterre P, Prangishvili D, Schleper C: **The legacy of Carl Woese and Wolfram Zillig: from phylogeny to landmark discoveries.** *Nat Rev Microbiol* 2013, **11**:713–9.

11. Palangat M, Larson MH, Hu X, Gnatt A, Block SM, Landick R: **Efficient reconstitution of transcription elongation complexes for single-molecule studies of eukaryotic RNA polymerase II.** *Transcription* 2012, **3**:146–53.

12. Werner F, Weinzierl RO.: **A Recombinant RNA Polymerase II-like Enzyme Capable of Promoter-Specific Transcription**. *Mol Cell* 2002, **10**:635–646.

13. Naji S, Grünberg S, Thomm M: **The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation.** *J Biol Chem* 2007, **282**:11047–57.

14. Trinh V, Langelier M-F, Archambault J, Coulombe B: **Structural perspective on mutations affecting the function of multisubunit RNA polymerases.** *Microbiol Mol Biol Rev* 2006, **70**:12–36.

15. Nottebaum S, Tan L, Trzaska D, Carney HC, Weinzierl ROJ: **The RNA polymerase factory: a robotic in vitro assembly platform for high-throughput production of recombinant protein complexes.** *Nucleic Acids Res* 2008, **36**:245–52.

16. Weinzierl ROJ: **The Bridge Helix of RNA polymerase acts as a central nanomechanical switchboard for coordinating catalysis and substrate movement.** *Archaea* 2011, **2011**:608385.

17. Korkhin Y, Littlefield O, Nelson PJ, Bell SD, Sigler PB: **Preparation of components of archaeal transcription preinitiation complex.** *Methods Enzymol* 2001, **334**:227–39.

18. Zhao X, Li G, Liang S: **Several Affinity Tags Commonly Used in Chromatographic Purification.** *J Anal Methods Chem* 2013, **2013**:581093.

19. Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S: **Comparison of affinity tags for protein purification.** *Protein Expr Purif* 2005, **41**:98–105.

20. Ritz D, Beckwith J: **Roles of thiol-redox pathways in bacteria.** *Annu Rev Microbiol* 2001, **55**:21–48.

21. McCoy J, Lavallie E: **Expression and purification of thioredoxin fusion proteins.** *Curr Protoc Mol Biol* 2001, **Chapter 16**:Unit16.8.

22. Mei Y, Peng N, Zhao S, Hu Y, Wang H, Liang Y, She Q: **Exceptional thermal stability and organic solvent tolerance of an esterase expressed from a thermophilic host.** *Appl Microbiol Biotechnol* 2012, **93**:1965–74.

23. Albers S-V, Jonuscheit M, Dinkelaker S, Urich T, Kletzin A, Tampé R, Driessen AJM, Schleper C: **Production of recombinant and tagged proteins in the hyperthermophilic archaeon Sulfolobus solfataricus.** *Appl Environ Microbiol* 2006, **72**:102–11.

24. Keefe AD, Wilson DS, Seelig B, Szostak JW: **One-step purification of recombinant proteins using a nanomolar-affinity streptavidin-binding peptide, the SBP-Tag.** *Protein Expr Purif* 2001, **23**:440–6.

25. Kim JH, Chang TM, Graham AN, Choo KHA, Kalitsis P, Hudson DF: **Streptavidin-Binding Peptide (SBP)-tagged SMC2 allows single-step affinity fluorescence, blotting or purification of the condensin complex.** *BMC Biochem* 2010, **11**:50.

26. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B: **A generic protein purification method for protein complex characterization and proteome exploration.** *Nat Biotechnol* 1999, **17**:1030–2.

27. She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CC, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PM, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, et al.: **The complete genome of the crenarchaeon Sulfolobus solfataricus P2.** *Proc Natl Acad Sci U S A* 2001, **98**:7835–40.

28. Wagner M, Berkner S, Ajon M, Driessen AJM, Lipps G, Albers S-V: **Expanding and understanding the genetic toolbox of the hyperthermophilic genus Sulfolobus.** *Biochem Soc Trans* 2009, **37**(Pt 1):97–101.

29. Redder P, She Q, Garrett RA: **Non-autonomous mobile elements in the crenarchaeon Sulfolobus solfataricus.** *J Mol Biol* 2001, **306**:1–6.

30. Zaparty M, Esser D, Gertig S, Haferkamp P, Kouril T, Manica A, Pham TK, Reimann J, Schreiber K, Sierocinski P, Teichmann D, van Wolferen M, von Jan M, Wieloch P, Albers S V, Driessen AJM, Klenk H-P, Schleper C, Schomburg D, van der Oost J, Wright PC, Siebers B: "**Hot standards" for the thermoacidophilic archaeon Sulfolobus solfataricus.** *Extremophiles* 2010, **14**:119–42.

31. Martusewitsch E, Sensen CW, Schleper C: **High spontaneous mutation rate in the hyperthermophilic archaeon Sulfolobus solfataricus is mediated by transposable elements.** *J Bacteriol* 2000, **182**:2574–81.

32. Soppa J: **Functional genomic and advanced genetic studies reveal novel insights into the metabolism, regulation, and biology of Haloferax volcanii.** *Archaea* 2011, **2011**:602408.

33. Imanaka T: **Molecular bases of thermophily in hyperthermophiles.** *Proc Jpn Acad Ser B Phys Biol Sci* 2011, **87**:587–602.

34. Leigh JA, Albers S-V, Atomi H, Allers T: **Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales**. *FEMS Microbiol Rev* 2011, **35**:577–608.

35. Schleper C, Kubo K, Zillig W: **The particle SSV1 from the extremely thermophilic archaeon Sulfolobus is a virus: demonstration of infectivity and of transfection with viral DNA.** *Proc Natl Acad Sci U S A* 1992, **89**:7645–9.

36. Schelert J, Dixit V, Hoang V, Simbahan J, Drozda M, Blum P: **Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon Sulfolobus solfataricus by use of gene disruption.** *J Bacteriol* 2004, **186**:427–37.

37. Albers S-V, Driessen AJM: **Conditions for gene disruption by homologous recombination of exogenous DNA into the Sulfolobus solfataricus genome.** *Archaea* 2008, **2**:145–9.

38. Ponting CP: **Novel domains and orthologues of eukaryotic transcription elongation factors.** *Nucleic Acids Res* 2002, **30**:3643–52.

39. Schrödinger L: **The PyMOL Molecular Graphics System, Version 1.3**. 2010.

8 Summary and General Discussion

Bart de Koning

Studying Information Processing

In his central dogma of molecular biology, Francis Crick stated that the flow of information within living cells starts with DNA that is transcribed into messenger RNA molecules, which in the end are translated into proteins [1, 2]. In his view, the proteins are the functional entities that perform all physiological tasks. After more than 50 years of investigating the molecular mechanisms of living cells, this theory remains unchallenged, although additional roles of (small) regulatory and catalytic RNA have been added more recently [3]. Key element in cellular information processing is fidelity. Fidelity is tightly balanced against processivity: lower fidelity means higher error rates, which at a certain point leads to an 'error catastrophe', meaning that fitness has dropped below the minimum for survival. On the other hand, fidelity comes at a cost: more accurate systems are generally slower, consume more energy, and as such have a lower processivity, which increases the danger of being outcompeted.

Therefore, living cells evolved complex systems to handle the flow of information both accurately and efficiently. Interestingly, these systems are highly comparable between the three domains of life. The central components of replication, transcription, aminoacylation, and translation are found in every living cell known today, with only relatively small deviations, despite a separation of billions of years of evolution. Archaea are unicellular, do not contain organelles, and have relatively small genomes, so are, at first sight, quite similar to their far better known prokaryotic cousins: the bacteria. Nevertheless, if it comes down to information processing, archaea are, surprisingly, more related to eukaryotes than to bacteria, both at the sequence level of RNA and proteins, and at the architecture level of key complexes as well [Chapter 2].

Archaea as model organisms to study highly conserved factors

The fact that archaea share more characteristics with eukaryotes than bacteria, but are still unicellular, 'simple' prokaryotic beings, make them excellent model systems to study eukaryote-like information processing. The absence of cell specialization, less cell organization, less or even no intracellular compartmentalization, and less intensive regulation, have proven to give a clearer picture of the function of conserved key elements within these complex systems [Chapter 2].

Most model archaea are not difficult to cultivate, certainly not with respect to some of the higher developed eukaryotes. However, a lag of well established biomolecular techniques for this domain of life hamper widespread usage. Altough methanogens were already discovered during the first decade of the last century and the haloarchaea halfway that century [4–7], the recognition that they belong to a very distinct group of prokaryotes took until 1977, when Carl Woese and George Fox presented the outcomes of their laborious categorisation of 16S/18S rRNAs [8]. They described it as follows: "*A phylogenetic analys based upon ribosomal RNA sequence characterization reveals that living systems represent one of three aboriginal lines of descent: (i) the eubacteria, comprising all typical bacteria; (ii) the archaebacteria, containing methanogenic bacteria; and (iii) the urkaryotes, now represented in the cytoplasmic component of eukaryotic cells.*" [8]. This finding was supported by distinct molecular characteristics, not in the least by the finding that the DNA-dependent RNA polymerase of archaea was more similar to eukaryotic ones than to bacterial ones [9]. In addition, lipid composition analyses of archaea displayed another surprising feature. Unlike the normal ester linkage based composition known

from the textbooks, archaea contain lipids based on ether linkages. These ether based lipids do even occur as bipolar entities, enabling even the existence of monolayer membrane arrangements [10]. Archaea appeared to be so different from the rest of the living world, that taxonomy needed an additional organisational layer above the kingdom level – historically designed to distinct animals from plants – to reflect the primary tripartite division of the living world that appeared from molecular analyses. Therefore, in 1990, Woese, Kandler and Wheelis proposed the Domain level, which is situated above the kingdom level, dividing life into Archaea, Bacteria and Eucarya [11]. Separation into a separate domain truly enhanced attention, but already by then bacterial and eukaryotic genetics gained a head start that proved difficult to catch up for the archaeal field. Nevertheless, during the last decade joined efforts provided numerous solutions for molecular modification of this remarkable group of organisms. J. Leigh, S. Albers, H. Atomi, and T. Allers, each a pioneer in establishing genetic systems for each of the main archaeal branches (methogens, hyperthermophilic crenarchaeotes, hyperthermophilic eurarchaeotes, and halophiles respectively) wrote together a compelling review article that thoroughly describes the current state of the art [12]. An impressive list of methodologies that has been developed during the last decade.

To establish archaea as model systems, widespread availability and adoption of best practices and standardized techniques is of utmost importance. This has been accomplished for some archaeal species. For example, *S. solfataricus* was a target of a systems biology approach by a European consortium during which best practices have been standardized throughout the participating labs [13]. Another example of the availability of standardized practices is the Halohandbook, an impressive collection of best practices from a number of contributors [14]. So, although the number of possibilities might seem somewhat more limited, and commercial support and kits are, with a few exceptions, not available, the use of archaea as model organisms to study eukaryotic (or bacterial) information processing is mainly limited by invention skills and perseverance.

In this thesis, we report several attempts to elucidate functional details of some very conserved factors in information processing in *S. solfataricus* using recently established genetic modification techniques. *S. solfataricus* is a thermoacidophilic crenarchaeote that grows optimally at temperatures between 70°C and 85°C and at pH values between 2 and 3. Its genome sequence is known since 2001 [15]. As mentioned above, best practices have become standardized between laboratories, which not only concerns widespread molecular genetics techniques, but also more elaborate ~omics approaches [13]. Already in 1992, the first transformation of *S. solfataricus* was reported [16]. Since then, the genomic toolbox was expanded to include gene knockout, and overexpression systems, the availability of reporter genes, and tunable promoters [12].

MBF1: a highly conserved activator

MBF1 (multi-protein bridging factor 1) was believed to be a transcriptional co-activator. It was shown to cross the gap between transcription regulators and the transcriptional machinery itself. In *Bombyx mori* (silkworm), it was shown to bind to a regulation complex, and subsequently to recruit the RNA polymerase by recruiting the TATA-box Binding Protein first [17]. MBF1 was found to be highly conserved within archaea, being present in almost all species with the key exception of marine thaumarchaeotes. However, none of the associated transcription regulators were known to be present within this domain, raising the question

whether a class of other regulators was overlooked, or that archaeal MBF1 might be a transcriptional activator itself, binding to DNA directly instead of indirectly via a binding partner. One study revealed a surprising dual role of this protein: in yeast it was not only associated with transcription but contributed to translation fidelity as well [18]. Neighbourhood analysis across the archaeal domain revealed no clear preference for either transcription or translation. Elements of both systems are equally present, especially in the well conserved neighbourhood within the crenarchaeotes [Chapter 3].

To investigate this, a *mbf1* disruption mutant of the thermoacidophilic crenarchaeote *S. solfataricus* was made using heterozygous recombination with a suicide plasmid. Under standard laboratory growth conditions *mbf1* appears to be not essential for growth, and comparing growth characteristics with its parental strain did not reveal striking differences between the two. Although this is not unusual, it is still striking that disrupting a protein, that is so well conserved during evolution, has so little effect on its phenotype under standard growth conditions. These observations were confirmed by similar analyses of a *mbf1* disruption mutant of the euryarchaeote *Thermococcus kodakariensis* (Matsumi, De Koning & Van der Oost, unpublished results). It was observed, that the *Sulfolobus mbf1* disruption mutant is much more sensitive during cultivation than its parental strain, showing sudden death during growth much more often. Being hard to quantify, this behaviour was especially observed when cultures were transferred at later stages during stationary phase or unfrozen from long term storage. But the largest difference was observed in the increased sensitivity of the *mbf1* disruption mutant towards paromomycin. Paromomycin is an aminoglycoside-type antibiotic that interferes with the recognition of cognate codon-anti-codon binding within the ribosomes during translation [Chapter 4].

A more detailed study to the molecular characteristics of the archaeal MBF1 from *S. solfataricus* revealed hardly any associations to the transcription machinery, but strengthened the assumed association to the translation apparatus. It was found that archaeal MBF1 consists of two domains that are structurally independent: an N-terminal zinc-ribbon, which is not conserved beyond the archaeal MBF1s, and the well conserved C-terminal HTH-domain (helix-turn-helix domain). This C-terminal HTH domain was shown to bind to the small ribosomal subunit by affinity purification, using immobilized archaeal MBF1 as a bait, and in co-purification experiments, in which we detected the presence of archaeal MBF1 in ribosomal purifications. NMR structure comparisons confirmed that archaeal MBF1 binds to the small ribosomal subunit using its C-terminal HTH domain, whereas the N-terminal zinc-ribbon might only contribute to this interaction, but does not participate directly in binding. The N-terminal domain remains independent during binding, and might therefore be involved in binding of other, yet unidentified, partners [Chapter 5].

Altogether, these findings made us believe that MBF1s in archaea are not associated with transcription but rather with translation. Based on the observations in yeast [18], and more recently its binding to polyadenylated mRNAs in different eukaryotic species [19–21], and, against the backdrop that the protein domain that binds to the small ribosomal subunit in *S. solfataricus* is highly conserved across the archaeao-eukaryotic lineage, it is tempting to speculate that the eukaryotic MBF1 plays a comparable role in the translation process, and that the observed association with transcription could be a more recently gained additional functionality. During eukaryotic evolution of MBF1, after the adoption of its role as a transcriptional co-activator, it may have lost its former role in translation in some eukaryotic lineages, and instead, only some species, like yeast, might have retained the participation of MBF1 in translation fidelity.

TGT, a conserved dichotomy

Another well conserved element within all three domains of life, which is involved in information processing, is the TGT (tRNA-guanine transglycosylase) family of proteins. This family of proteins shows a clear dichotomy between the archaeal domain of life on the one hand and the eukaryotic and bacterial domains on the other. In the latter two domains, TGT is responsible for the exchange of guanine at the wobble position (position 34) of the anti-codon of tRNAs with either queuosine in eukaryotes or its precursor $preQ_1$ in bacteria. Only a limited number of tRNAs have a guanine present at the wobble position, therefore in bacteria and eukaryotes TGT acts on specific tRNAs of four aminoacids [22]. In archaea, TGT is responsible for the exchange of guanine with $preQ_0$ at position 15 in allmost, if not all, archaeal tRNAs. $PreQ_0$ is in a later stage converted to archaeosine by a protein that belongs to the TGT family as well [23].

The role of queuosine in bacteria and eukaryotes is closely related to fidelity, increasing the discriminative power between highly similar tRNAs that have almost identical anti-codon boxes, but are linked to different amino acids nevertheless. Archaeosine in archaea, on the other hand, is exclusively present at position 15, which is part of the Levitt base-pair: a structural element that connects the D-loop to the start of the T-loop, and helps stabilization of the tRNA's tertiary structure. Although it is arguable that this influences fidelity of translation as well, restricting the flexibility of tRNAs, it is foremost a structural modification that is thought to be equivalent to magnesium binding in the other two domains of life [24]. Disruption of the *tgt* gene, which encodes the TGT protein in *S. solfataricus*, revealed that it was solely responsible for this process without any redundancy present. Like *mbf1*, this gene appeared to be non-essential, as this mutant was also as viable as its parental strain, and showed hardly any changes in growth characteristics. In comparison to the *mbf1* disruption mutant, the *tgt* disruption mutant was much more stable and did not reveal the sensitivity to stationary phase. It grew slightly slower than the parental strain, especially at normal temperatures (75°C), but when temperature levels were raised (87-93°C) growth returned to almost wild-type levels [Chapter 6]. It was long believed that the archaeosine modification was present in the archaeal domain because of the abundance of thermophilic organisms in this domain. It was believed that the extra stabilization caused by archaeosine was especially beneficial at high temperatures, despite the presence of the archaeosine modification among the numerous mesophilic archaea. This contradicting observation was explained as an indication of the thermophilic origin of the archaea [25]. Our findings, in agreement with a recent study to several disruption mutants in the halophilic *Haloferax volcanii,* in which a *tgt* disruption mutant was found to be cold sensitive [26], suggest that, on the contrary, it could be more important at lower temperatures. This might be explained because at lower temperatures fixation might actually be more important than at higher temperatures. At lower temperatures molecular movements occur less frequent than at higher temperatures. The chance that an aminoacyl tRNA acquires the right conformation during its contact with the ribosome or aminoacyl-tRNA synthetase is therefore increased at higher temperatures in comparison to lower temperatures. This would explain the necessity to fix a molecule in its correct conformation at lower temperatures. In analogy with magnesium binding in bacteria and eukaryotes, it is intriguing to speculate whether organisms within these domains also have increasing demands for magnesium ions at lower temperatures similar to the increasing demands for archaeosine as observed in archaea.
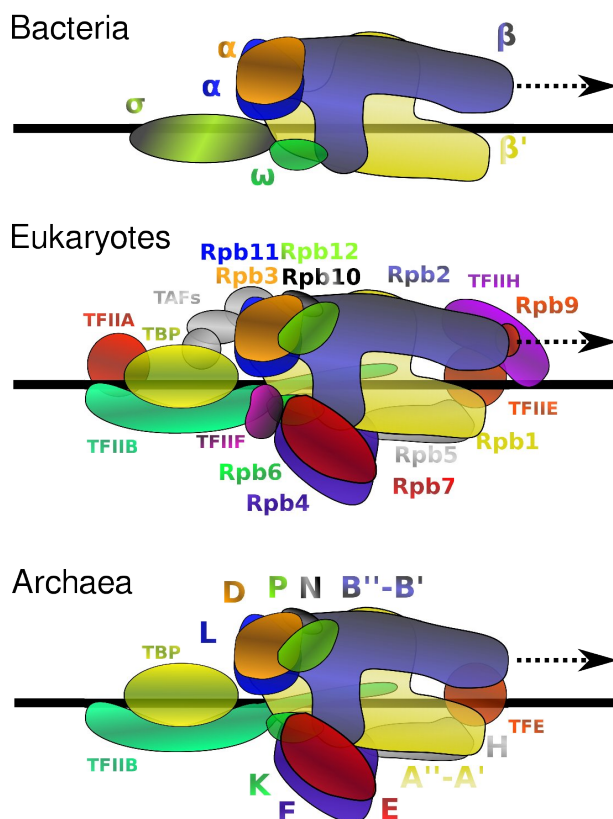
Figure 1. Schematic comparison between the transcription pre-initiation complexes of the three domains. Orthologues proteins between the three domains have the same colour. Although the core complex is shared between all three domains, the degree of similarity is much larger between archaea and eukaryotes. After [28].

Aiding research to the basal machinery of RNAP

Beyond doubt, the best studied, element of information processing systems is the RNAP (RNA polymerase) complex. Its basal core is present in all known life forms, and is highly conserved. The surrounding, auxiliary, and regulatory elements are less conserved, but, interestingly, a number of them is shared between the archaea and the eukaryotes. In fact, the archaeal RNAP is almost identical to the eukaryotic RNAP II complex [27]. This high resemblance already proved beneficial, as the archaeal RNAP could be reconstituted from heterologously expressed subunits in popular protein expression systems, in contrast to its eukaryotic orthologues. The heterologous expression of the archaeal RNAP revealed numerous functional details about the molecular characteristics of the complex as a whole, and, in addition, revealed also an unprecedented insight in the separate subunits as this RNAP expression platform provided opportunities to tamper with the subunit composition and to modify the separate subunits themselves by introducing genetic variations.

Unfortunately, purification of homologously expressed complexes, which are expressed in archaeal systems itself, are, in contrast to ones heterologously expressed in bacterial hosts, hard to obtain, and involve a number of purification steps and therefore a substantial amount of

biomass. To enable easier purification, a method was developed in which a purification tag was inserted in the genome of *S. solfataricus* after a gene that encodes an RNAP subunit, avoiding artificial overproduction by viral infections or heterologous expression in other less adapted hosts [Chapter 7]. In a proof of principle experiment, the enrichment an RNAP core component was proven, whereas an auxiliary element was tagged using this novel method. This showed the potential of the method, and further optimization, and up scaling might be relatively straight forward.

Cultivation of *Sulfolobus*

There are clear advantages of working with aerobic thermophiles, like the neglectable danger of contamination (other than from simultaneously growing thermophilic cultures), and the ease of cultivation: the only necessity is an efficient stove that can handle temperatures between 70°C and 85°C. However, a major disadvantage is the higher rate of evaporation. Liquid cultures are continuously threatened by drying out. The default solution to maintain cultures for prolonged periods is the use of erlenmeyer culture flasks with prolonged necks. The second solution, mainly used in ~omics approaches to keep cultures in steady condition, is the use of fermentors. Handling of fermentors is difficult, and much more costly, and therefore not suited for common practice or large scale usage. Comparing phenotypic traits for microbes involves comparisons between large numbers of cultures to raise the statistical power necessary, especially if differences in doubling times or yields are expected to be small. Using growth reactors would therefore require either a high number of reactors for parallel cultivation, or a long time to grow them sequentially, what would be a bridge too far for most labs (and students). On the other hand, growing these cultures in prolonged erlenmeyers poses logistic issues as the number of stoves needed for a good comparison, will reach surrealistic numbers for even a small experimental design.

Figure 2. Two types of culturing S. solfataricus. Upper row: erlenmeyers with elongated necks. The most common way to culture thermophilic aerobic organisms like *S. solfataricus.* Bottom row: tubes with a diameter of 25 mm, using silicosen T-32 stoppers (Hirschmann Laborgeräte).

To do the growth experiments described in this thesis, growing thermophiles in growth tubes were explored with kind help of Rie Matsumi. Aerobic growth in tubes is limited by two major factors: oxygen transfer rate and evaporation rate. The interface between air and liquid has to be sufficient for efficient oxygen transfer. At elevated temperatures gas exchange drops to a minimum, which increases the minimal tube diameter necessary for aerobic growth. An increased diameter, on the other hand, has a higher evaporation rate, which limits the number of days a culture can grow before it will dry out, or imposes extra addition of water during the experiment, which increases the variation in measurements. Several approaches were tested, but best results were obtained using glass tubes with a diameter of 25 mm, filled with 20 ml medium, and stirred continuously at a low rpm. To decrease the evaporation rate and extend the culture period, Silicosen T-32 stoppers (Hirschmann Laborgeräte) were used. These silicon plugs are very porous, and as such they are rather permeable to air, but limited for transfer of evaporated water, and therefore limit the dehydration of the cultures. Using these tubes we could grow cultures for approximately two weeks, taking samples (0.2 ml) for measurements several times a day. And increased our growth capacity to 100 simultaneously growing cultures, making some of the experiments described in this thesis possible [Chapter 3 & 5].

A disadvantage of using *S. solfataricus*, is the chromosomal instability. Chromosomal rearrangements have been reported and are known to happen, which is a continuous threat to experiments, and limit the growing possibilities to a great extent. It has to be noted therefore, that its far more stable nephews *S. acicaldarius* or *S. islandicus*. turned out to be better suited for genetic modification and phenotypological experiments than *S. solfataricus* itself. Futhermore, the genetic toolbox of these nephews becomes at least as extensive as, or might even already surpassed, the toolbox of *S. solfataricus*. Therefore, the use of *S. acicaldarius* or *S. islandicus* should actually be preferred over the use of *S. solfataricus*.

Extrapolating Results

At a certain point, with any model organism, the question is raised if it is possible to translate the results found into theorems applicable to vertebrate species in general and to humans in particular. For archaeal species this questions might even be more relevant. There is a border of billions of years of evolution that separates these species from our own. Is it even sensible to assume that bridging this is possible?

It is astonishing to realise that, despite the separation of an almost infinite amount of time, an amount of time from which even a percent cannot be comprehended by any human in any way, such a high level of conservation occurs in systems related to information processing as described in chapter 2. For information processing, which is as essential as life itself, all three domains of life rely on systems that are in principle the same. Whether it is a human, a platypus, a bug, a slime mold, a grass, a yeast, a gut living bacterium, or an organism living at approximately 80°C in a yellowish hot 'tub' in the middle of Yellow Stone park, they all have a DNA replication machinery based upon a DNA-dependent DNA polymerase, a transcription device that at least has a composition of 4 highly conserved DNA-dependent RNA polymerase subunits, approximately 18 aminoacyl tRNA synthetases, and an rRNA based mega complex consisting of two parts that makes up the ribosome. Within the whole tree of life, these elements are always present, highly alike, and of identical origin. The only exceptions to this rule are organisms from which most people are convinced they should not be entitled 'alive' at all: viruses. But even these 'creatures' parasite on other organisms, by using the information

processing systems from their hosts to 'build' their own offspring.

When one restricts the comparison to only include information processing systems in archaea and eukaryotes, the similarities are even clearer. To extrapolate the well known 'Darwinian' picture of the human that was transformed out of an ape even further down to even include our hot 'tub' growing, yellowish, sweaty feet smelling *S. solfataricus* into the picture, the only major difference in, for example, the RNAP complex concerns the composition of the auxiliary proteins. Most probably, this affects the regulation capabilities of the complex, suiting the requirements of more complex organisms for more elaborate regulation schemes.

The explorations towards the function of the core RNAP elements, which were performed in archaea during the last decade, have proven that archaeal systems indeed are a suitable model. Without these archaeal structural models our knowledge about the biochemical function of the transcription complex would be far more limited than it is today. However, there are still a number of proteins floating around this central core machinery of which we have, at most, a vague description of what they might do. This group of proteins, MBF1 amongst them, certainly is a subject for studies in archaea. In analogy to the well conserved core machinery, these conserved auxiliary units might have shared characteristics as well. In addition, knowing the evolutionary history of an object could reveal new characteristics that were overlooked or not understood before. An example could include MBF1, which is now known to bind to the small ribosomal subunit, and as such has a role in translation, at least in archaea. This poses the question, could eukaryotic MBF1 have a moonlighting function in translation as well?

But even well studied proteins, of which it is known that in eukaryotes the function is different from the function in archaea, like TGT, are worth investigating in archaea. Proteins are not random collections of amino acids, but are composed of well-structured orderings of subparts. These subparts, known as structural domains, have functions that are well conserved throughout the domains of life. Therefore, studying their archaeal counterparts can reveal functional elements or mechanisms that were kept hidden thus far, or cannot be understood from eukaryotic models alone. In addition, from advancements throughout human history, and science in particular, it has to be acknowledged that numerous findings were conceived focussing on seemingly contradicting behaviour.

Thus, to come back to our question, whether it is even sensible to use archaeal systems as models to study eukaryotic complexes: I am convinced it is even the only option to reveal characteristics of well conserved elements that might be kept hidden within the complexity of "higher" organisms.

References

1. Crick FHC: **On protein synthesis.** *Symp. Soc. Exp. Biol.* 1956, **XII**:139–163.
2. Crick FHC: **Central dogma of molecular biology.** *Nature* 1970, **227**:561–3.
3. Mattick JS: **Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms.** *Bioessays* 2003, **25**:930–9.
4. Schnellen CGTP: **Onderzoekingen over de methaangisting**. 1947.
5. STADTMAN TC, BARKER HA: **Studies on the methane fermentation. X. A new formate-decomposing bacterium, Methanococcus vannielii.** *J. Bacteriol.* 1951, **62**:269–80.
6. Lochhead AG: **NOTE ON THE TAXONOMIC POSITION OF THE RED CHROMOGENIC HALOPHILIC BACTERIA.** *J. Bacteriol.* 1943, **45**:574–5.
7. Söhngen NL: **Het Ontstaan en Verdwijnen van Waterstof en Methaan**. 1906.
8. Woese CR, Fox GE: **Phylogenetic structure of the prokaryotic domain: the primary kingdoms.** *Proc. Natl. Acad. Sci. U. S. A.* 1977, **74**:5088–90.
9. Zillig W, Stetter KO, Janeković D: **DNA-dependent RNA polymerase from the archaebacterium Sulfolobus acidocaldarius.** *Eur. J. Biochem.* 1979, **96**:597–604.
10. De Rosa M, Gambacorta A, Gliozzi A: **Structure, biosynthesis, and physicochemical properties of archaebacterial lipids.** *Microbiol. Rev.* 1986, **50**:70–80.
11. Woese CR, Kandler O, Wheelis ML: **Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.** *Proc. Natl. Acad. Sci. U. S. A.* 1990, **87**:4576–9.
12. Leigh JA, Albers S-V, Atomi H, Allers T: **Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales**. *FEMS Microbiol. Rev.* 2011, **35**:577–608.
13. Zaparty M, Esser D, Gertig S, Haferkamp P, Kouril T, Manica A, Pham TK, Reimann J, Schreiber K, Sierocinski P, Teichmann D, van Wolferen M, von Jan M, Wieloch P, Albers S V, Driessen AJM, Klenk H-P, Schleper C, Schomburg D, van der Oost J, Wright PC, Siebers B: "**Hot standards" for the thermoacidophilic archaeon Sulfolobus solfataricus.** *Extremophiles* 2010, **14**:119–42.
14. Dyall-Smith: *The Halohandbook - Protocols for haloarchaeal genetics*. 2009.
15. She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CC, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PM, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, Van der Oost J: **The complete genome of the crenarchaeon Sulfolobus solfataricus P2.** *Proc. Natl. Acad. Sci. U. S. A.* 2001, **98**:7835–40.
16. Schleper C, Kubo K, Zillig W: **The particle SSV1 from the extremely thermophilic archaeon Sulfolobus is a virus: demonstration of infectivity and of transfection with viral DNA.** *Proc. Natl. Acad. Sci. U. S. A.* 1992, **89**:7645–9.
17. Takemaru K i, Li FQ, Ueda H, Hirose S: **Multiprotein bridging factor 1 (MBF1) is an evolutionarily conserved transcriptional coactivator that connects a regulatory factor and TATA element-binding protein.** *Proc. Natl. Acad. Sci. U. S. A.* 1997, **94**:7251–6.
18. Hendrick JL, Wilson PG, Edelman II, Sandbaken MG, Ursic D, Culbertson MR: **Yeast frameshift suppressor mutations in the genes coding for transcription factor Mbf1p and ribosomal protein S3: evidence for autoregulation of S3 synthesis.** *Genetics* 2001, **157**:1141–58.
19. Klass DM, Scheibe M, Butter F, Hogan GJ, Mann M, Brown PO: **Quantitative proteomic analysis reveals concurrent RNA-protein interactions and identifies new RNA-binding proteins in Saccharomyces cerevisiae.** *Genome Res.* 2013, **23**:1028–38.

20. Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, Bonneau R, Selbach M, Dieterich C, Landthaler M: **The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts.** *Mol. Cell* 2012, **46**:674–90.

21. Kwon SC, Yi H, Eichelbaum K, Föhr S, Fischer B, You KT, Castello A, Krijgsveld J, Hentze MW, Kim VN: **The RNA-binding protein repertoire of embryonic stem cells.** *Nat. Struct. Mol. Biol.* 2013, **20**:1122–30.

22. Grosjean H, de Crécy-Lagard V, Marck C: **Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes.** *FEBS Lett.* 2010, **584**:252–64.

23. Phillips G, Chikwana VM, Maxwell A, El-Yacoubi B, Swairjo MA, Iwata-Reuyl D, de Crécy-Lagard V: **Discovery and characterization of an amidinotransferase involved in the modification of archaeal tRNA.** *J. Biol. Chem.* 2010, **285**:12706–13.

24. Oliva R, Tramontano A, Cavallo L: **Mg2+ binding and archaeosine modification stabilize the G15 C48 Levitt base pair in tRNAs.** *RNA* 2007, **13**:1427–36.

25. Gregson JM, Crain PF, Edmonds CG, Gupta R, Hashizume T, Phillipson DW, McCloskey JA: **Structure of the archaeal transfer RNA nucleoside G*-15 (2-amino-4,7-dihydro- 4-oxo-7-beta-D-ribofuranosyl-1H-pyrrolo[2,3-d]pyrimidine-5-carboximi dam ide (archaeosine)).** *J. Biol. Chem.* 1993, **268**:10076–86.

26. Blaby IK, Phillips G, Blaby-Haas CE, Gulig KS, El Yacoubi B, de Crécy-Lagard V: **Towards a systems approach in the genetic analysis of archaea: Accelerating mutant construction and phenotypic analysis in Haloferax volcanii.** *Archaea* 2010, **2010**:426239.

27. Werner F, Grohmann D: **Evolution of multisubunit RNA polymerases in the three domains of life.** *Nat. Rev. Microbiol.* 2011, **9**:85–98.

28. Werner F: **Structure and function of archaeal RNA polymerases.** *Mol. Microbiol.* 2007, **65**:1395–404.

Samenvatting

In 1956 postuleerde Francis Crick zijn "Centrale Dogma" voor de moleculaire biologie. In dit dogma stelde hij dat erfelijke informatie van levende organismen wordt opgeslagen in DNA. Uit dit DNA worden de genen overgeschreven als mRNA transcripten, welke uiteindelijk worden vertaald naar eiwitten. In zijn visie zijn deze eiwitten de functionele entiteiten die alle fysiologische taken verrichten binnen een cel. Meer dan 50 jaar moleculair onderzoek later, staat dit model nog als een huis. Wel is het model iets uitgebreid door de ontdekking dat RNA zelf ook regulerende en katalyserende functies kan hebben. Voor deze informatieverwerking hebben levende organismen tal van systemen in huis en bij al deze systemen speelt betrouwbaarheid een grote rol. Een lagere betrouwbaarheid betekent meer fouten, wat leidt tot een opstapeling van fouten waardoor een cel uiteindelijk niet meer kan functioneren. Aan de andere kant gaat betrouwbaarheid ten koste van de snelheid en ten koste van energie, wat weer kan leiden tot het verliezen van de concurrentiestrijd met andere organismen. Hierdoor bestaat er een nauwe balans tussen enerzijds betrouwbaarheid en anderzijds snelheid.

De levende wereld om ons heen kent veel meer organismen dan de planten en dieren waar de meeste mensen direct aan denken. In feite wordt de wereld gedomineerd door organismen die een magnitude kleiner zijn dan deze eukaryoten: de prokaryoten. Deze prokaryote groep bestaat uit twee heel verschillende takken: de bacteriën en de archaea. De laatste groep werd vooral bekend door de extreme leefomstandigheden waarin ze voorkomen. Zo zijn er archaea aangetroffen die leven bij temperaturen van meer dan 100 °C, onder zure omstandigheden met een pH van minder dan 2, bij extreem hoge druk, of in heel zoute habitats. Pas in deze eeuw kwam men erachter dat deze vrij onbekende groep organismen ook in grote getale voorkomt in normale biotopen, en zelfs een belangrijke rol speelt in globale stikstof- en carbonkringlopen. Archaea bezitten een paar opmerkelijke kenmerken: ze bezitten afwijkend ribosomaal RNA, hebben afwijkende celmembranen (ether in plaats van ester bindingen, L- in plaats van D-glycerol, isoprenoid ketens in plaats van vetzuren en zijn in sommige gevallen enkellaags in plaats van altijd dubbellaags) en archaea maken voor hun informatieverwerking gebruik van systemen die veel meer lijken op systemen uit eukaryoten dan op de systemen uit bacteriën. Hierom heeft men, eind vorige eeuw, de levende wereld opgedeeld in drie domeinen: de bacteriën, de archaea en de eukaryoten.

In hoofdstuk 2 zijn de informatiesystemen uit deze drie domeinen met elkaar vergeleken en is de huidige kennis over informatieverwerking in archaea geïnventariseerd. Hierin wordt duidelijk dat informatieverwerking in alle drie de domeinen op een vergelijkbare manier gebeurt. De kerncomponenten van de complexen voor replicatie, transcriptie en translatie vertonen maar kleine verschillen, ondanks de afstand van miljarden jaren aan evolutie. De meer perifiëre elementen vertonen wel grotere verschillen. Maar juist de archaeële systemen lijken hierin veel meer op de eukaryote systemen, dan op de bacteriële [hoofdstuk 2].

Deze sterke verwantschap maakt het aantrekkelijk archaeële systemen te gebruiken als model voor de eukaryotische. Net als bacteriën zijn archaea unicellulair, kennen geen cel specialisatie, niet of nauwelijks cel compartimentalisatie, hebben een vrij klein, circulair genoom, en zijn sommige makkelijk te cultiveren. Alleen de afwezigheid van uitgebreide moleculaire mogelijkheden staat veelvuldige toepassing als model voor eukaryotische systemen in de weg. In deze thesis wordt verslag gedaan van een studie om functionele eigenschappen van enkele, tijdens de evolutie sterk geconserveerde factoren binnen informatiesystemen op te helderen, waarbij we als modelsysteem *Sulfolobus solfataricus* hebben gebruikt. *S. solfataricus* is een crenarchaeoot, oorspronkelijk geïsoleerd uit thermale bronnen, onder andere in Yellow Stone park (VS) en bij de Solfatara vulkaan dichtbij Napels (Italië). Het groeit optimaal bij een temperatuur tussen de 70 °C en 85 °C en een pH waarde tussen de 2 en de 3 in een oxidatieve

omgeving. Het complete genoom is al beschikbaar sinds 2001 en de moleculaire toolbox is, sinds de eerste transformatie in 1992, ondertussen vrij uitgebreid, met onder andere een genknockout systeem, overexpressie systemen, reporter genen en regelbare promotoren.

MBF1 (multi-protein bridging factor 1) is voor het eerst aangetroffen in de zijderups, en bleek een brug te kunnen slaan tussen enkele transcriptieregulatoren en het transcriptiesysteem zelf. In 1999 bleek dat MBF1 een HTH (helix-turn-helix) domein bevat die sterk geconserveerd voorkomt bij alle tot dan toe bekende archaea en eukaryoten. Ondertussen blijkt dat alleen de mariene tak van de recentelijk ontdekte thaumarchaeoten dit gen mist. Of MBF1 eenzelfde functie heeft in archaea is de vraag, omdat geen van de bekende geassocieerde regulatoren in archaea is aangetroffen. Daarnaast bleek uit een studie in gist dat MBF1 ook een effect heeft op de betrouwbaarheid van het translatiesysteem. Dit is niet helemaal te rijmen met de bekende functie als co-regulator tijdens transcriptie.

In het verleden bleek dat het voor archaea mogelijk is om functionele voorspellingen te doen door te kijken naar welke genen tijdens de evolutie in de buurt van een bepaald gen zijn gebleven. Een uitgebreide genoom-neighbourhood-analyse onder archaea gaf geen uitsluitsel over een voorkeur van het archaeële MBF1 voor transcriptie of translatie. Beide systemen werden ongeveer in gelijke mate aangetroffen, vooral in het sterk geconserveerde operon van *mbf1* in crenarchaeoten [hoofdstuk 3].

Om meer te kunnen vinden over de functionele rol van MBF1 in archaea, is er een gen-knockout gemaakt. Onder normale laboratorium condities bleek er geen verschil in groei te bestaan tussen de wildtype culturen en de knockout mutant. Wel bleek de knockout mutant sensitiever te zijn tijdens het cultiveren dan het wildtype, waarbij vooral de gevoeligheid voor de stationaire fase opvallend is. Omdat het eukaryote MBF1, vooral in planten, ook een rol heeft bij de adaptatie aan stress, lijkt het archaeële MBF1 hierin eenzelfde rol te vervullen. Ook bleek de knockout, net als bij gist, gevoeliger voor het aminoglycoside antibioticum paromomycine. Paromomycine is een antibioticum dat van invloed is op de codon/anti-codon herkenning door het ribosoom en is daardoor van invloed op de betrouwbaarheid van translatie [hoofdstuk 4].

In een meer gedetailleerdere studie naar de moleculaire eigenschappen van MBF1 uit *S. solfataricus* kwamen geen associaties met het transcriptiesysteem naar boven, maar wel een erg sterke associatie met het translatie systeem. Tijdens een eiwitinteractie studie en uit NMR experimenten blijkt dat het sterk geconserveerde HTH-domein in archaea sterk bindt aan de kleine ribosomale subunit. Tijdens deze studies werd geen enkele binding met het transcriptiecomplex of met bekende transcriptieregulatoren aangetroffen. Ook tijdens ribosoomisolaties blijkt MBF1 mee te liften met het ribosoom door een sucrosegradient. Alleen is er geen invloed op de betrouwbaarheid van *in vitro* translaties gevonden, zoals bij gist [hoofdstuk 5].

Op grond van deze bevindingen concluderen wij dat in archaea MBF1 geen transcriptie regulator is, maar een factor die betrokken is bij translatie. De overeenkomsten met de bevindingen in gist en meer recentelijk de vondst dat MBF1 bindt aan gepolyadenylyseerde mRNA's in verschillende eukaryoten (waaronder muisstamcellen en humane HEK293 cellen) en de sterke conservering van het HTH domein tussen archaea en eukaryoten, maakt het aantrekkelijk te speculeren of deze rol ook van toepassing kan zijn in eukaryoten. Dit zou erop kunnen wijzen dat het eukaryote MBF1, naast zijn rol in transcriptie, ook nog de oorspronkelijke nevenfunctie heeft in translatie. Deze rol kan weliswaar ook verloren zijn gegaan in de evolutie.

Een ander geconserveerd element binnen alle drie de domeinen is de TGT- (tRNA-guanine transglycosylase) familie van eiwitten. In deze familie is er wel sprake van een duidelijke dichotomie. TGT in bacteriën en eukaryoten zorgt voor de uitwisseling van guanine met queuosine of de precursor $preQ_1$ daarvan. Deze guanine bevindt zich op de wobble positie (positie 34) van het anti-codon van tRNA's. Omdat niet alle tRNA's hier een guanine bevatten, gebeurt dit alleen bij specifieke tRNA's van 4 aminozuren. In archaea is TGT verantwoordelijk voor een heel andere uitwisseling: namelijk de uitwisseling van guanine op positie 15 van alle tRNA's met $preQ_0$. Deze base wordt hierna weer omgezet in archaeosine, een base die alleen in archaea voorkomt, door een ander lid uit de TGT familie.

Uit de analyse van een *tgt* deletie mutant, blijkt dat TGT inderdaad een centrale rol vervult in deze tRNA modificatie. In deze mutant werd archaeosine niet meer ingebouwd. Deze mutant bleek nauwelijks te lijden onder het wegvallen van de modificatie, aangezien er nauwelijks verschillen in groei zijn aangetroffen. Alleen onder normale groeicondities leek de mutant iets langzamer te groeien dan het wildtype, een verschil dat onder verhoogde temperaturen niet meer meetbaar was. Eenzelfde mutatie in *Haloferax volcanii*, een ander archaeon, leidde tot kou sensitiviteit, wat erop zou kunnen wijzen dat het inbouwen van archaeosine vooral nuttig is bij verlaagde temperaturen. Wellicht omdat het onder deze omstandigheden belangrijker is dat het tRNA zich in de juiste conformatie bevindt en blijft, dan onder hogere temperaturen [hoofdstuk 6].

Het best bestudeerde informatieverwerkingssysteem in levende cellen is zonder twijfel het RNA polymerase complex dat verantwoordelijk is voor transcriptie. De overeenkomsten in dit systeem tussen eukaryoten en archaea bleken heel waardevol, aangezien het archaeële systeem *in vitro* in elkaar gezet kan worden uit heteroloog tot expressie gebrachte onderdelen. Iets wat nog niet is gelukt bij eukaryote systemen. Hierdoor is het mogelijk de samenstelling van het polymerase complex te veranderen en mutaties aan te brengen op specifieke locaties. Dit heeft tot nieuwe inzichten geleid. Toch heeft het heteroloog tot expressie brengen nadelen ten opzichte van het rechtstreeks isoleren uit organismen zelf. Zo is de vorming van onderdelen suboptimaal in een vreemde omgeving, hetgeen kan leiden tot artefacten tijdens *in vitro* experimenten. Echter, het rechtstreeks isoleren uit organismen is momenteel heel inefficiënt. Er is erg veel celmassa en een groot aantal isolatie stappen nodig om gezuiverd RNA polymerase te krijgen. Als mogelijke oplossing hiervoor hebben we een methode bedacht waarbij we een zuiverings-"tag" introduceren in het genoom achter het *rpo7* gen, dat codeert voor het Rpo7 eiwit, welke uit de RNA polymerase steekt als een soort steel. Als proof of principle, hebben we na een simpele reguliere één-stap-isolatie een verrijking aangetoond van een kernonderdeel van het RNA polymerase complex. Omdat we dit onderdeel niet zelf hadden "getagged", kunnen we stellen dat we het complex als geheel hebben verrijkt [hoofdstuk 7].

Het werken met *Sulfolobus* als model heeft veel voordelen. Zo is er nauwelijks gevaar voor contaminatie van de culturen, doordat deze meestal de hoge temperatuur en de lage pH niet overleven. Eigenlijk is het enige wat nodig is om ze te cultiveren een efficiënte stoof of een oliebad. Een nadeel van de hoge temperatuur is de verhoogde verdamping. Om dit te beperken wordt vaak gebruik gemaakt van erlenmeyers met verlengde nekken. Maar deze conventionele methode is niet efficiënt. Zelfs met een klein experimenteel design begint het ruimtegebruik aardig op te lopen. Om toch de groeiexperimenten, die beschreven zijn in dit manuscript, mogelijk te maken, is samen met Rie Matsumi een efficiëntere methode met het groeien van *S. solfataricus* in normale glazen buizen onderzocht. De beste resultaten werden opgedaan met buizen van een diameter van 25 mm, een vloeistof volume van 20 ml en afgesloten met Silicosen T-32 stoppen (Hirschmann Laborgeräte). Deze stoppen zijn erg poreus en laten

daardoor wel gassen door maar nauwelijks waterdamp. Hiermee was het mogelijk een groot aantal culturen minstens 2 weken lang simultaan te laten groeien [hoofdstuk 4, 6 en 8].

Is het mogelijk resultaten van archaea te extrapoleren naar mensen? Ondanks de enorme evolutionaire afstand, blijkt uit de opsomming in hoofdstuk 3, dat alle levensvormen voor informatieverwerking gebruik maken van systemen die in essentie hetzelfde zijn. Of het nu gaat om een mens, een vogelbekdier, een insect, een slijmzwam, een gras, een gist, een bacterie in onze maag of *S. solfataricus*, het heeft een replicatie systeem met een DNA polymerase, een transcriptie systeem dat minimaal uit 4 hoog geconserveerde RNA polymerase subunits bestaat, ongeveer 18 aminoacyl tRNA synthetases en een op ribosomaal RNA gebaseerd mega complex bestaande uit 2 helften, die samen het ribosoom vormen voor de translatie. Als we in de vergelijking alleen kijken naar archaea en eukaryoten, worden de overeenkomsten nog veel groter. De experimenten, die zijn uitgevoerd in de afgelopen 10 jaar, naar de functie van RNA polymerase subunits, tonen aan dat archaea bruikbare modelsystemen vormen. Zonder deze modelsystemen zou onze kennis van deze systemen een stuk beperkter zijn. Maar nog steeds zijn er tal van eiwitten die rond deze systemen zweven, waarvan zelfs geen vage beschrijving beschikbaar is van wat ze doen. In deze groep bevinden zich zeker kandidaten voor studies in archaea. Het kennen van de evolutionaire achtergrond, kan nieuwe inzichten opleveren die voorheen zijn gemist. Een voorbeeld daarvan kan MBF1 zijn, waarvan we nu weten dat het in archaea aan de kleine ribosomale subunit bindt en een rol vervult in translatie.

Dus, om terug te keren naar de oorspronkelijke vraag of het zelfs wel denkbaar is om archaeële systemen als modelsystemen voor de eukaryote te kunnen gebruiken: ik ben ervan overtuigd dat het zelfs de enige optie is om alle karakteristieken van geconserveerde elementen boven tafel te halen.

About the author

Bart de Koning was born on the 19th of May 1980 in a little village, Wouw, in the South-West of the Netherlands, right between Antwerpen and Rotterdam. During his secondary education at the Norbertus College in Roosendaal, he developed a keen interest in natural sciences, informatics and astronomy. And, although his father, who was a Biology teacher at a secondary school, never persuaded him to do so, he basically followed his genes, and started studying Biology at Wageningen University in 1998. He specialised himself in cellular and molecular Biology, doing a minor thesis at the Laboratory of Molecular Biology on the infection of nitrogen fixing bacteria in burclover, a major thesis at the Laboratory for Genetics on linear plasmids in mitochondria of filamentous fungi, and his final thesis at the Laboratory of Microbiology on homologues of eukaryotic co-activators in archaea, an initial study of which the follow up resulted in the work that is described within this thesis. He did his intership in the Biological and Environmental Systems Group at the University of Sheffield doing mass spectrometry experiments within the archaeon *S. solfataricus*.

Beside a thorough interest in biological affairs, he became active within social and political affairs by joining educational committees within Wageningen, becoming a student member of the student council of Wageningen University for the Progressive Student Party faction, and a founder of the student association of the Euroleague for Life Sciences. During his studies he even managed to discover the most important subject of his attention, Lisette S. Weissman, with whom he married the 18th of May 2004 in Wageningen.

He started his PhD in 2005 at the Bacterial Genetics group at the Laboratory of Microbiology, under the supervision of Prof. dr. John van der Oost on a VICI grant awarded project to characterise conserved proteins in Information Processing Systems in Archaea. The results of this study are described within this thesis.

In 2010, he decided to shift gears and joined the Clinical Genomics group at Maastricht University headed by Prof. dr. Bert Smeets to become a bioinformatician in Next Generation Sequencing data handling. Currently, he is a bioinformatician within the Laboratory for Clinical Genetics at the MUMC+ (Maastricht) and the section of Genome Diagnostics at the Radboudumc (Nijmegen) both headed by Prof. dr. Han Brunner, and responsible for the implementation of bioinformatic analyses pipelines for the use of novel NGS techniques in genetic diagnostics at the MUMC+. He currently lives in Maastricht with his wife Lisette, and their three kids: Olaf, Myrthe and Annelie.

List of publications

Snijders APL, de Vos MGJ, de Koning B, Wright PC: **A fast method for quantitative proteomics based on a combination between two-dimensional electrophoresis and 15N-metabolic labelling.** *Electrophoresis* 2005, **26**:3191–3199.

Snijders APL, de Koning B, Wright PC: **Perturbation and interpretation of nitrogen isotope distribution patterns in proteomics.** *J Proteome Res* 2005, **4**:2185–2191.

Snijders APL, de Koning B, Wright PC: **Relative quantification of proteins across the species boundary through the use of shared peptides.** *J Proteome Res* 2007, **6**:97–104.

de Koning B, Blombach F, Wu H, Brouns SJJ, van der Oost J: **Role of multiprotein bridging factor 1 in archaea: bridging the domains?** *Biochem Soc Trans* 2009, **37**(Pt 1):52–57.

de Koning B, Blombach F, Brouns SJJ, van der Oost J: **Fidelity in Archaeal Information Processing**. *Archaea Vancouver BC* 2010, **2010**:1–34.

Gerards M, Kamps R, van Oevelen J, Boesten I, Jongen E, de Koning B, Scholte HR, de Angst I, Schoonderwoerd K, Sefiani A, Ratbi I, Coppieters W, Karim L, de Coo R, van den Bosch B, Smeets H: **Exome sequencing reveals a novel Moroccan founder mutation in SLC19A3 as a new cause of early-childhood fatal Leigh syndrome.** *Brain* 2013, **136**(Pt 3):882–90.

Blombach F, Launay H, Snijders AP, Zorraquino V, Wu H, de Koning B, Brouns SJJ, Ettema T, Camilloni C, Cavalli A, Vendruscolo M, Dickman MJ, Cabrita LD, La Teana A, Benelli D, Londei P, Christodoulou J, van der Oost J: **Archaeal MBF1 binds to 30S and 70S ribosomes via its helix-turn-helix domain.** *Biochem J* 2014, **462**:373–84.

Vrijenhoek T, Kraaijeveld K, Elferink M, de Ligt J, Kranendonk E, Santen G, Nijman I, Butler D, Claes G, Costessi A, Dorlijn W, van Eyndhoven W, Halley D, van den Hout M, van Hove S, Johansson L, Jongbloed J, Kamps R, Kockx C, de Koning B, Kriek M, dit Deprez RL, Lunstroo H, Mannens M, Nelen M, Ploem C, Rijnen M, Saris J, Sinke R, Sistermans E, van Slegtenhorst M, Sleutels F, van der Stoep N, van Tienhoven M, Vermaat M, Vogel M, Waisfisz Q, Weiss M, van den Wijngaard A, van Workum W, IJntema H, van der Zwaag B, van IJcken WFJ, Den Dunnen J, Veltman J, Hennekam R, and Cuppen E **Next-genereation sequencing-based genome diagnostics across clinical genetics centers - implementation choices and their effects.** *Eur J Hum Genet* 2015, *manuscript accepted for publication*

de Koning B, Blombach F, Matsumi R, Albers SV, Sierocinski P, Pougovkina O, Ettema TJG, Atomi H, Roovers M, Brouns SJJ, van der Oost J: **Phenotypic analysis of MBF1 from the archaeaon** *Sulfolobus solfataricus* **reveals a regulatory function beyond the level of transcription.** *manuscript in preparation*

de Koning B, Matsumi R, Waghmare SP, Pougovkina O, Snijders APL, Dickman M, Brouns SJJ, van der Oost J: **Molecular characterization of tRNA-guanine transglycosylase of** *Sulfolobus solfataricus***.** *manuscript in preparation*

de Koning B, Engelhart R, Brouns SJJ, van der Oost J: **RNA Polymerase complex tagging in vivo in** *S. solfataricus* **without insertion of foreign elements.** *manuscript in preparation*

Jóri B, Kamps R, Xanthoulea S, Delvoux B, Blok MJ, van de Vijver K, de Koning B, Oei TF, Carli Tops, Speel EJ, van Gorp T, Kruitwagen R, Gomez-Garcia EB, Romano A: **Novel germ-line mutations modify the penetrance of Lynch syndrome mutations for endometrial cancer.** *manuscript in preparation*

Dankwoord

De laatste woorden... Ik moet bekennen: er zijn momenten geweest dat ik dacht dat ik deze drie woorden nooit zou schrijven, maar zie: ze staan er nu toch. En dat is niet in het minst door de gigantische hulp van een groot aantal mensen. Al deze mensen, van wie ik de afgelopen periode hulp heb gehad, hier afzonderlijk bedanken, dat gaat jammer genoeg niet. Maar iedereen, die hier mist, was zeker niet minder belangrijk voor mij, en mijn dank is groot aan eenieder, die wel weet waarom, en waarvoor geen letters op papier nodig zijn!

Allereerst wil ik mijn promotor John bedanken. Zonder zijn mental support was dit proefschrift zeker niet tot stand gekomen. Hij gaf mij een kans, en op momenten dat ik echt het einde niet meer zag, wist hij toch weer een lichtpunt te creëren. Al was het maar door stokken achter deuren te plaatsen, en in discussies de leemtes in mijn kennis en resultaten te openbaren. Makkelijk kan het niet geweest zijn, ook het einde niet, en bijzonder dat er toch altijd dat vertrouwen bleef bestaan. En al dekt het woord de lading bij lange na niet: Bedankt!

Stan, co-promotor, in het begin zelfs zonder het te weten: ook enorm bedankt. De CRISPRs vergden je aandacht, maar je was daar als ik je nodig had, en altijd bereid een helpende hand uit te steken. In de kamer creëerde je een wetenschappelijk uitdagende en prettige sfeer, en het was een eer getuige te mogen zijn van de doortastende stappen om het CRISPRsysteem bloot te leggen.

Special thanks to Rie. Your knowledge about and experience with archaeal modification techniques really turned the tide. And your help with the sulfolobus and thermococcus mutants was of great importance for the contents of this thesis.

Ook vielen dank aan mijn "Partner in Crime" Fabian. Faab, uiteraard in het Nederlands: enorm goede herinneringen heb ik aan onze samenwerking. Op avontuur in het buitenland, of het nu ging om het zoeken naar de Goethe boom, of om afgevoerd worden door de marechausee, in een Japans antiquariaat overnachten, een bankpas kwijtraken, of in het midden van de nacht Into the Wild over door god verlaten wegen rijden door New Hampshire om een dorpje te vinden van 4 huizen, een benzinepomp, en een gigantisch summerschoolcomplex, en daarna in WallMart kwaliteit t-shirts te testen, omdat onze bagage lost was, of juist hoe pikant de "Instant death sauce" van een Amerikaans wrap concern wel niet was. Er was altijd een extra dimensie aan iets wat toch recht-toe-recht-aan had moeten gaan. Maar ook voor de discussies die we gehad hebben, en je enorme parate kennis van biomoleculaire (en andere) zaken, en je voorkeur voor het oprakelen van artikelen die in vergetelheid dreigden te verzakken. Nu ja veel van dit boekje is daar in meer of mindere mate dan ook het eindproduct van!

And to the last member of our mbf1 team: Hao. Thanks for the great discussions we always had, and your perseverence and help in finding binding partners for mbf1. Especially your optimism, and your faith in our great results that would undoubtly come, were inspiring.

Sonja, dank voor de hulp bij het creëren van de mbf1 mutant, en de cursus sulfolobus modificeren die ik bij jullie heb mogen volgen. De mutanten die daaruit voortkwamen zijn zeker de basis van het hier gepresenteerde onderzoek. Ook de extra cursus pizza bakken is zijn doel niet voorbij geschoten.

Thijs, jouw vooronderzoek was de aanleiding voor de beurs waarbinnen dit onderzoek werd verricht. De link naar transcriptie hebben wij nooit kunnen vinden, maar translatie is toch zeker ook mooi. Dank nog voor de hulp bij het afronden van de laatste hoofdstukken. Jammer dat Zweden toch een tikkie ver is, maar we komen zeker nog een keer binnenvallen.

Bram, Mark and Sakh, during my internship I had absolutely a wonderful time in Sheffield. Thanks for all the extra analysis to confirm the absence of archaeosine, this was an excellent follow-up of that time in Sheffield.

Mijn paranimfen Mark en Marcel, tja, waar moet ik beginnen. Vanaf mijn stages en afstudeervakken, was Mark er ook, en Marcel begon vlak na mij aan zijn PhD. De spelavonden en BBQs, ook met John en Eline, waren legendarisch. En Marcel je ongenuanceerd relativeren van mijn linkse gedachtegoed is iets wat ik absoluut mis. Tegen wie moet ik nu ageren? Mijn dank is groot voor jullie (mental) support en kameraadschap tijdens de goede, maar niet altijd even makkelijke tijd. En natuurlijk ook tijdens deze laatste loodjes!

John (Raedts), Marco, Mark (Musters), Christian, Bram, Vincent: gelachen hebben we, en lachen zullen we. Dank voor de enorm mooie tijd: de koffiepauzes, de lunchwandelingen, de BBQs, het rondcrossen door de states, de beach party aan een meer, het sprookjesfeest in een kasteel, het rondhangen in Haarlem, etc, etc, etc, etc....

Overige kamergenoten: Pawel, Magnus, Edze, Matthijs. Thanks for the scientific atmosphere in our room, the lively discussions, the conversations, the jokes... Pawel, other sulfolobian, thanks for your help with the transcriptomics, and if I can finish my thesis, you will manage it too! Matthijs, sinds een mini-afstudeervak bleven onze wegen elkaar kruisen, tot aan dezelfde labtafel aan toe, ben benieuwd waar we elkaar de volgende keer weer kruisen: hoop spoedig!

De overigen van de "oude" garde Suzanne, Harmen en Jaapie. Dank voor de goede tijd op Microbiologie, en de rake kritiekpuntjes. Harmen, de koffie discussies over resultaten, eiwitten, sequencen, software, politiek en de rest van de wereld, blijf ik missen. Jaapie, een betere labsergeant heb ik niet meer gezien, en het bezoek aan Janelia en Washington blijft in mijn geheugen gegrift.

Ronnie, de in between gardist, en Liesbeth, bedankt! Bij tijd en wijlen oppassen op de babyfoon, en het spelen van een spelletje, hebben zeker bijgedragen aan het voltooien van mijn boekje. Al was het alleen maar om de broodnodige ontspanning te creëren. Liesbeth, ik hoop dat je plan lukt, en dat wij ooit ook een boekje van je mogen ontvangen.

Other microbiologists: thanks for the good times, the discussions, the travels, the committees, working together, drinking together, and playing football. Thomas: thanks for FC-KAAS and the philosophical afterthoughts. An en Sjon: dank voor al het werk om het lab draaiende en de groep bijeen te houden. Wim, ondertussen weet ik iets meer over ICT en netwerkaangelegenheden: dank dat ik altijd moeilijk mocht doen. Jannie, de hoeveelheid glaswerk die je voor mij door de afwassers en de autoclaven heb moeten halen is ongekend, maar altijd stond het er later weer klaar voor gebruik.

Mijn studenten Corné, Olga, Ruben. Dank voor alle inzet, en energie die jullie in de projecten hebben gestoken. Sommige van jullie resultaten hebben zich weten te vertalen in delen van dit boekje. Jullie bijdrage, alleen al in het uitproberen van sommige ideeën, was zeker van onschatbare waarde! Corné, het nachten doorhalen in het lab op energy drankjes om te kunnen samplen, was toch wel één van de apartere laberervaringen die ik gehad heb.

Bert Smeets and Patrick Lindsey: although my PhD was not finished, and was focussed on archaeal genetics instead of human genetics, and that my experience with programming was limited, you gave me the opportunity to dive deep into the world of bioinformatics and next generation sequencing. An opportunity that was life changing, and for which my gratitude will always remain. In the same time you both stimulated me and gave me time to finish this book, although it took somewhat longer than expected.

Overview of completed training activities

Discipline specific activities

- NWO molecular genetics meetings *-poster presentation-* (Lunteren, 2005-2008)
- NWO protein research meetings *-poster presentation-* (Lunteren & Veldhoven, 2005-2008)
- General meeting of the Association for General and Applied Microbiology (VAAM, Jena, Germany, 2006)
- Gordon research conference - Archaea: Ecology, Metabolism, and Molecular Biology *-poster presentation-* (Proctor Academy, NH, USA, 2007)
- Molecular Biology of Archaea *-poster presentation-* (St. Andrews, UK, 2008)
- 8th Annual UK meeting on Genetics & Molecular Mechanisms in Archaea *-oral presentation-* (Nottingham, UK, 2009)
- Systems biology course: Principles of ~omics data analysis (EPS-NBIC, 2006)
- Advanced course on applied genomics of industrial fermentation (BODL, 2006)
- Hands on training molecular manipulation of *Sulfolobus solfataricus* (RUG, 2007)

General activities

- VLAG PhD Introweek (VLAG, 2006)
- Techniques for Writing and Presenting a Scientific Paper (WGS, 2007)
- Afstudeervak organiseren en begeleiden (OWU, 2007)
- Eén op één begeleiding (OWU, 2007)
- Safe handling with radioactive materials and sources (radiation expert 5B) (WUR, 2007)
- Career Perspectives (WGS, 2009)

Optional activities

- Biweekly PhD/PostDoc meetings Microbiology (2005-2009)
- Weekly group meetings Bacterial Genetics (2006-2009)
- VLAG PhD excursion California, USA *-poster and oral presentations-* (2006)
- Organization of VLAG PhD excursion California (2006)
- VLAG PhD council (2006-2009)
- Chair VLAG PhD council (2007-2009)