Computational Biology and Chemistry 35 (2011) 341–346

Contents lists available at SciVerse ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Review article

New insights on gene regulation in archaea

Silvia Tenorio-Salgado[a], Alejandro Huerta-Saquero[b], Ernesto Perez-Rueda[a,*]

[a] Departamento de Ingeniería Celular y Biocatálisis, IBT-UNAM, AP. 565-A, Cuernavaca, Morelos 62210, Mexico
[b] Departamento de Microbiologia Molecular, IBT-UNAM, AP. 565-A, Cuernavaca, Morelos 62210, Mexico

ARTICLE INFO

Article history:
Received 30 August 2011
Accepted 1 October 2011

Keywords:
DNA-binding transcription factors
Protein families
Archaeal genomes
Evolution
Gene regulation

ABSTRACT

Archaea represent an important and vast domain of life. This cellular domain includes a large diversity of organisms characterized as prokaryotes with basal transcriptional machinery similar to eukarya. In this work we explore the most recent findings concerning the transcriptional regulatory organization in archaeal genomes since the perspective of the DNA-binding transcription factors (TFs), such as the high proportion of archaeal TFs homologous to bacteria, the apparent deficit of TFs, only comparable to the proportion of TFs in parasites or intracellular pathogenic bacteria, suggesting a deficit in this class of proteins. We discuss an appealing hypothesis to explain the apparent deficit of TFs in archaea, based on their characteristics, such as their small length sizes. The hypothesis suggests that a large fraction of these small-sized TFs could supply the deficit of TFs in archaea, by forming different combinations of monomers similar to that observed in eukaryotic transcriptional machinery, where a wide diversity of protein–protein interactions could act as mediators of regulatory feedback, indicating a chimera of bacterial and eukaryotic TFs' functionality. Finally, we discuss how global experiments can help to understand in a global context the role of TFs in these organisms.

Contents

1. Introduction

Archaea represent an important and vast domain of life, together with bacteria and eukarya (Auguet et al., 2010; Chaban et al., 2006; Clementino et al., 2007). This cellular domain includes four well-known divisions, Crenarchaeota, Euryarchaeota, Korarchaeota, and Nanoarchaeota that represent a large diversity of organisms, habitats and life styles. In terms of regulatory mechanisms, organisms included in this cellular domain possess basal transcription machinery resembling that of eukaryotes (Esser and Martin, 2007; Lopez-Garcia, 1999; Martin and Muller, 1998), such as a TATA box promoter sequence, a TATA-box-binding protein (TBP), a homologue of the transcription factor TFIIB (TFB), and a RNA polymerase (RNApol) containing between 8 and 13 subunits (Goede et al., 2006). Recently, Grohmann and Werner (2010), evidenced that diverse subunits of the RNApol are similar between archaea and eukarya, as the F/E (RPB4/7) subunit but not conserved in bacteria; whereas the subunit Spt4/5 is the only subunit conserved among bacteria, archaea and eukarya.

In addition, archaeal genes are organized in operons and co-transcribed in common mRNAs similar to bacterial mRNAs (Bell, 2005; Kyrpides and Woese, 1998). In broad terms, across almost all bacteria and even archaea, operons are characterized by close spacing of genes (intergenic regions), modest conservation, and modest function similarity (Price et al., 2006). In contrast in eukarya, operonic organization has not been documented. These observations

* Corresponding author. Tel.: +52 56227610; fax: +52 7773172388.
E-mail address: erueda@ibt.unam.mx (E. Perez-Rueda).

1476-9271/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiolchem.2011.10.006

raise different basic questions with regard to the mechanisms of transcriptional regulation and the manner by which bacterial-like transcription factors (TFs) may interact or interfere with the components of the eukaryotic-like basal transcriptional machinery within an archaeal cell. It is for this reason that archaeal DNA-binding TFs represent an important class of proteins in the understanding of the molecular mechanisms that underlie transcription regulation.

In this regard, the majority of identified TFs in archaeal organisms exhibit a large proportion of homologous to bacterial activators and repressors TFs (Bell, 2005; Kyrpides and Woese, 1998). Therefore, very few eukaryotic-like TFs have been identified in archaea (Kruger et al., 1998). Even though the ever-growing number of archaeal genome sequences reveals an increasing list of potential regulators (Coulson et al., 2007; Perez-Rueda et al., 2004; Perez-Rueda and Janga, 2010; Wu et al., 2008), most detailed and advanced studies have been performed with few TFs, mainly from the AsnC family (Bell, 2005; Leonard et al., 2001; Napoli et al., 1999). See Table 1. Therefore comparative genomic analysis of archaea represents an opportunity to understand how similar archaea and bacteria, and archaea and eukarya are, and to understand the evolution of gene regulation networks in prokaryotes and eukaryotes.

In the present review, we describe the most recent findings from studies on gene regulation in archaeal genomes from the perspective of DNA-binding TFs. In this regard, we break the subject into sections, covering the challenge to identify TFs in sequence genomes, the apparent under-representation of the number of TFs in archaea compared to bacterial genomes, a considerable number of small TFs with a significant fraction encoding for single domain proteins, and a high proportion of archaeal TFs homologous to bacteria. We finish with some conjectures that attempt to provide a comprehensive picture about how global experiments can help to understand in a global context the role of TFs in these organisms.

2. The identification of the repertoire of TFs in archaea genomes represents a big challenge

The identification of the TF repertoire in a genome sequence will allow to understand the regulation of gene expression and to elucidate its role in a global context.

Previous attempts to identify TFs in archaea using family-specific models from the bacterium *Escherichia coli* K12 TFs resulted in a low proportion of bacterial-like TFs (Perez-Rueda et al., 2004), probably because archaeal TF regulatory repertoire includes additional classes of DNA-binding motifs not observed in *E. coli*, suggesting that the repertoire of TFs in archaeal genomes is far from being complete. Recently, the distribution of TFs in archaeal genomes was evaluated by scanning diverse organisms from the four cellular divisions with a combination of diverse bioinformatics tools, such as Blast searches using as query sequences a well-known dataset of TFs of *Halobacterium* sp. NRC-1, a battery of family specific Hidden Markov Models (HMMs), a helix-turn-helix HMM that considers amino acid residue identity and solvent accessibility, constructed from a set of heterogeneous DNA-binding proteins with standard HTH motifs (Changhui, 2006); Cluster of Orthologous Genes (COGs) assignations, and DNA-Binding Domains database assignments characteristic of TFs. From these searches, 3,918 TFs were identified. Although an extensive survey in this work identified a large set of TFs widely distributed in archaea, it is still possible that some potential novel TFs, escaped the search criteria or are missing because of their lineage-specific nature, presumably due to *de novo* invention of TFs whose DNA-binding models are not included.

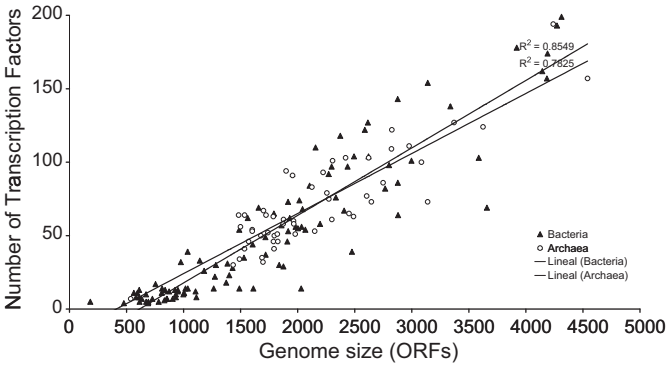

Fig. 1. Abundance and proportion of TFs identified in 52 archaeal and 100 bacterial genomes. Bacterial organisms correspond to intracellular pathogens, and extremophiles. On *x-axis* genomes are sorted from smallest to largest size. On *y*-axis is the number of TFs.

3. The deficit of TFs in archaeal genomes resembles bacterial parasites and/or intracellular pathogens

The abundance of regulatory complexes associated to organisms usually correlates with their complexity, i.e., large genomes containing an extense repertoire of genes could need more regulatory elements than small genomes. In this regard, the proportion of TFs in larger genomes is consistent with the hypothesis that an increase of genome complexity and physiological functionality is generally associated with a more complex regulation of gene expression (Bhardwaj et al., 2010). On the other hand, small archaeal genomes also contain a small amount of TFs in comparison to large ones, following a linear correlation ($r^2 = 0.85$), similar to bacteria TFs (Perez-Rueda et al., 2004) (Fig. 1). Therefore, large genomes might be harboring an ample repertoire of TFs to exploit diverse or more complex habitats. In counterpart small genomes containing fewer regulators might be associated with specific niches. For instance, *Haloarcula marismortui* and *Methanosarcina acetivorans* (with the largest genomes, described so far), shows the highest proportion of TFs among archaeal genomes sequenced so far, whereas the symbiotic hyperthermophile, *Nanoarchaeum equitans*, has both a reduced genome and a lower proportion of TFs than other archaea. Thus, complex life styles might require a higher proportion of genes and TFs to better orchestrate responses to changing environments, as in the case of *M. acetivorans* which can form aggregate multicellular structures when passing from anaerobiosis to aerobiosis (Oelgeschlager and Rother, 2008).

Although there is a correspondence between the genome size and TF proportion, less than 5% of the open reading frames (ORFs) in most archaeal genomes are devoted to gene regulation, similar to the intracellular pathogens, opportunistic pathogens and extremophiles, and in contrast to about 8–10% observed in bacterial genomes with similar number of ORFs (Perez-Rueda and Collado-Vides, 2000, 2001) (Fig. 1). Larger archaeal genomes, such as *M. acetivorans* and *H. marismortui*, with similar number of ORFs to the bacterium *E. coli* K12, encode a lesser proportion of TFs (4.8, 3.5 and 8%, respectively). In summary, the TF repertoire observed in archaea is much more similar to bacteria associated with gene loss events, such as intracellular-pathogens, extremophiles, and/or endosymbionts (3.9% in average). Notable exceptions are *Pyrococcus horikoshi* and *P. abyssi*, two small genomes containing 4.8% and 5.1% of TFs, respectively, comparable to the proportion of TFs in larger archaeal genomes. In contrast, *N. equitans* exhibited a clear deviation when proportion of genes coding for TFs was compared against genome size.

Diverse scenarios can be proposed to explain the apparent TF deficit in archaea, such as the inability to identify those lineage or