Contents lists available at ScienceDirect

Gene

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

Prevalent structural disorder carries signature of prokaryotic adaptation to oxic atmosphere

Arup Panda, Tapash Chandra Ghosh *

Bioinformatics Centre, Bose Institute, P 1/12, C.I.T. Scheme VII M, Kolkata 700 054, India

ARTICLE INFO

ABSTRACT

Microbes have adopted efficient mechanisms to contend with environmental changes. The emergence of oxygen was a major event that led to an abrupt change in Earth's atmosphere. To adjust with this shift in environmental condition ancient microbes must have undergone several modifications. Although some proteomic and genomic attributes were proposed to facilitate survival of microorganisms in the presence of oxygen, the process of adaptation still remains elusive. Recent studies have focused that intrinsically disordered proteins play crucial roles in adaptation to a wide range of ecological conditions. Therefore, it is likely that disordered proteins could also play indispensable roles in microbial adaptation to the aerobic environment. To test this hypothesis we measured the disorder content of 679 prokaryotes from four oxygen requirement groups. Our result revealed that aerobic proteomes are endowed with the highest protein disorder followed by facultative microbes. Minimal disorder was observed in anaerobic and microaerophilic microbes with no significant difference in their disorder content. Considering all the potential confounding factors that can modulate protein disorder, here we established that the high protein disorder in aerobic microbe is not a by-product of adaptation to any other selective pressure. On the functional level, we found that the high disorder in aerobic proteomes has been utilized for processes that are important for their aerobic lifestyle. Moreover, aerobic proteomes were found to be enriched with disordered binding sites and to contain transcription factors with high disorder propensity. Based on our results, here we proposed that the high protein disorder is an adaptive opportunity for aerobic microbes to fit with the genomic and functional complexities of the aerobic lifestyle.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Environments play crucial roles in shaping and maintaining the biochemical, morphological and physiological diversities of prokaryotes. The close interaction between microorganism and its environment leaves footprint on its nucleotide and amino acid composition (Dutta and Paul, 2012; Foerstner et al., 2005; Mann and Chen, 2010; Tekaia et al., 2002). Thus, closely related microbes from different environmental communities are often found to have different genomic and proteomic compositions than the distantly related microbes inhabiting the same environment (Foerstner et al., 2005). These genomic footprints provide important clues about the mechanisms of microbial adaptations across different environmental communities. The simplest compositional parameter that has been most extensively studied to assess the impact of the environment on genome evolution is genomic GC composition (Foerstner et al., 2005; Mann and Chen, 2010; Musto et al., 2004). GC composition was shown to be influenced by environmental variables such as temperature, aerobiosis, niche complexity, nitrogen utilization, etc. (Foerstner et al., 2005; McEwan et al., 1998; Musto et al., 2004; Naya et al., 2002; Zheng and Wu, 2010). Among the other factors that have been widely studied to address different adaptive strategies include codon usage, di-nucleotide frequencies, protein composition, protein stability, enzymatic activity, etc. (Burra et al., 2010; Dutta and Paul, 2012; Tekaia et al., 2002; Willenbrock et al., 2006). Studies of these genomic footprints now become an instrumental tool to understand the processes of environmental adaptation. Until very recently all these studies were mainly concentrated on the proteins having well defined 3-D structure. Emerging evidences suggest that disordered residues could also play an important role in such adaptations (Burra et al., 2010; Pazos et al., 2013; Sun et al., 2013). However, these studies are limited to few cases, while a diverse area of ecological adaptation remains unexplored. Photosynthetic release of oxygen by some cyanobacteria incurred a massive change in the anoxic condition of the early atmosphere. In order to survive this challenging environment microorganisms that existed by that time have adopted various changes (Cabiscol et al., 2000; McEwan et al., 1998; Naya et al., 2002; Vieira-Silva and Rocha, 2008). Considering the complexity of such phenotypic changes it could be hypothesized that disordered proteins may confer significant advantages for evolutionary adaptations. Therefore, in

---

this study, we systematically analyzed whether disordered proteins have been exploited for microbial adaptation to the oxic condition.

Due to their extensive adaptability to different environmental conditions special interest has been paid to find the attributes of disordered proteins contributing to ecological adaptations. Intrinsically disordered proteins (IDPs) are a class of proteins that lack stable 3-D structure under physiological condition (reviewed in references Babu et al., 2011; Gsponer and Babu, 2009; Midic et al., 2009; Uversky, 2011). These proteins are biased for charged and polar amino acids and are devoid of hydrophobic and aromatic amino acids (Uversky, 2011). Due to their conformational plasticity, IDPs can participate in a large number of interactions with multiple partners and can mediate new interactions when necessary (Gsponer and Babu, 2009; Panda et al., 2012; Uversky, 2011). Therefore, it was proposed that IDPs could be efficiently used for a quick response against environmental stimuli (Pazos et al., 2013; Pietrosemoli et al., 2013). In higher eukaryotes like plants, adaptations mainly come from the complex interactions between different environmental variables. Thus, plants were found to use higher fraction of disordered residues to facilitate complex interactions for their adaptation to changing environments (Pazos et al., 2013; Pietrosemoli et al., 2013). In *Arabidopsis thaliana* genome, IDPs were found to be specifically related to functions responsible for their environmental adaptation (Pietrosemoli et al., 2013). Since, disordered proteins are resistant to extreme temperature, these proteins were supposed to contribute for the thermal stability of prokaryotes (Burra et al., 2010; Tantos et al., 2009; Wang et al., 2013). However, prokaryotes living in extreme temperature were found to selectively reduce their disorder content to fit with their reduced genomic, proteomic size and functional complexity in extreme conditions (Burra et al., 2010). This diminution of structural disorder was shown to be a genuine adaptation for the survival in extreme temperature condition (Burra et al., 2010). IDPs were described as ensembles of alternative conformers (Uversky and Dunker, 2010). Extensive research on these conformational variants have revealed that the transition from fully disordered state to semi-disordered state can also confer thermal adaptation (Wang et al., 2013). Disordered residues were also implicated in intracellular lifestyle of parasitic eukaryotes and in viral pathogenicity (Goh et al., 2009; Mohan et al., 2008; Pushker et al., 2013; Xue et al., 2012). Altogether, these studies open a new avenue of research on the functional roles of disordered proteins in adaptation to various environmental conditions.

Following the rise of atmospheric oxygen level, aerobes have harnessed efficient mechanisms to mitigate the harmful effects oxygen derivatives. However, anaerobes that lack such protective mechanisms gradually become sensitive to oxygen and retreat to oxygen-limited environments. Therefore, aerobes and anaerobes differ considerably in their aerotolerance. Several physiological and biochemical factors were described to account for the microbial aerotolerance which include defense against reactive oxygenic species, the presence of the cellular components resistant to oxidative damage, etc. (Cabiscol et al., 2000). In line with these, several genomic and proteomic changes were proposed (McEwan et al., 1998; Naya et al., 2002; Vieira-Silva and Rocha, 2008). Comparative studies between aerobe and anaerobes have established that aerobes prefer higher genomic GC content than the anaerobes (McEwan et al., 1998; Naya et al., 2002). High GC value could be advantageous for aerobes in the sense that GC pair is more stable than the AT pairs and increased use of GC nucleotide may also render amino acids less susceptible to oxidation (Naya et al., 2002). At protein level, aerobes were proposed to prefer amino acids rich in oxygen and to avoid amino acids susceptible to oxidative damage (Acquisti et al., 2007; Vieira-Silva and Rocha, 2008). However, detailed investigation of the impact of oxygen on protein evolution found no evidence for such biased amino acid composition in aerobic microbes (Vieira-Silva and Rocha, 2008). The presence of oxygen is not only a source of massive environmental stress but also an adaptive opportunity for the evolution of complex multicellular life (Berner et al., 2007; Koch and Britton, 2008; Raymond and Segre, 2006). Aerobic metabolism was supposed to

facilitate the development of multicellular organisms through establishment of nuclear and epigenetic signaling systems (Jeltsch, 2013; Jiang et al., 2010). IDPs are integral to signal transduction systems and have been implicated in epigenetic modifications (Babu et al., 2011; Gsponer and Babu, 2009; Mahmoudabadi et al., 2013; Sandhu, 2009; Uversky, 2011). Moreover, availability of oxygen was proposed to instigate network complexity and to underwrite capacity for complex biological functions (Koch and Britton, 2008; Raymond and Segre, 2006). IDPs are crucial for network plasticity and are associated with myriad of biological functions (Gsponer and Babu, 2009; Pazos et al., 2013; Pietrosemoli et al., 2013; Uversky, 2011). Therefore, it could be assumed that aerobic genomes would need high amount of disordered proteins to sustain their functional complexity. Almost all eukaryotic genomes were found to a have higher fraction of disordered residues than most of the prokaryotes (Pancsa and Tompa, 2012; Schad et al., 2011; Xue et al., 2012). This sharp increase of proteome disorder content upon transition from prokaryotes to eukaryotes (correlated with the availability of oxygen (Koch and Britton, 2008; Raymond and Segre, 2006; Thannickal, 2009)), and not beyond that clearly suggests that the abundance of disordered proteins may have a strong link with the time scale of oxygen evolution (Schad et al., 2011). Our main objective is to unveil the probable link between these two variables over a wide range of genomes. However, for an initial assessment of the impact of oxygen on protein disorder content, in this study, we mainly concentrated on prokaryotic genomes and set out a comparative analysis of their disorder content in four major oxygen requirement groups. Some anaerobic genomes are also found among the protists (Hug et al., 2010). Therefore, to test our hypothesis in the eukaryotic lineage we collected available genome sequences of known anaerobic protists and compared their disorder content with that of protist genomes obtained from NCBI database (aerobic).

We noticed a wide variability in the extent of protein disorder among the different groups of prokaryotic and eukaryotic genomes. Most interestingly, we found that aerobic prokaryotes have high protein disorder irrespective of their selection for any other attributes. Considering the functional implications of disorder residues in aerobic prokaryotes, here we proposed that the high protein disorder in aerobic genomes is an adaptive trait that has assisted microorganisms to survive under oxic environment.

2. Methods

2.1. Collection of dataset

Complete sets of protein coding sequences for 679 prokaryotic genomes (with available information on oxygen requirement) were retrieved from NCBI database (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/). Related information (genome size, phylum, temperature range, habitat, etc.) was collected from NCBI genome information site (ftp://ftp.ncbi.nlm.nih.gov/genomes/genomeprj_archive/). We excluded genomes with multiple chromosomes and when multiple strains are available for a species we choose one strain arbitrarily. List of the genomes included in this study was given in the Supplementary dataset. Protein coding sequences of protist genomes were retrieved primarily from NCBI database. Ten protist genomes are available in this database: *Cryptosporidium parvum*, *Dictyostelium discoideum*, *Leishmania braziliensis*, *Leishmania donovani*, *Leishmania infantum*, *Leishmania major*, *Plasmodium falciparum*, *Theileria annulata*, *Theileria parva* and *Trypanosoma brucei*. Genome sequences of five anaerobic protists (with available genome sequences) were obtained from different resources. CDS sequences of *Giardia intestinalis AssemblageA* were obtained from GiardiaDB (v-4.0), and that of *Trichomonas vaginalis* were obtained from TrichDB (v-1.3) (Aurrecoechea et al., 2009). Entire proteome of *Entamoeba histolytica* was downloaded from ftp site of Wellcome Trust Sanger Institute (ftp://ftp.sanger.ac.uk/pub/pathogens/Entamoeba/histolytica/Proteome) (Aslett et al., 2005). Protein sequences of