doi:10.1016/j.jmb.2007.03.044 J. Mol. Biol. (2007) 369, 474–488

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Crystal Structure of the Archaeal Heat Shock Regulator from *Pyrococcus furiosus*: A Molecular Chimera Representing Eukaryal and Bacterial Features

Wei Liu[1]*, Gudrun Vierke[2], Ann-Kathrin Wenke[2], Michael Thomm[2] and Rudolf Ladenstein[1]

[1]*Karolinska Institutet NOVUM Center for Structural Biochemistry, 141 57 Huddinge, Sweden*

[2]*Archaea Center, University of Regensburg, 93053 Regensburg, Germany*

**Corresponding author*

We report here the crystal structure of a protein from *Pyrococcus furiosus* (Phr) that represents the first characterized heat shock transcription factor in archaea. Phr specifically represses the expression of heat shock genes at physiological temperature *in vitro* and *in vivo* but is released from the promoters upon heat shock response. Structure analysis revealed a stable homodimer, each subunit consisting of an N-terminal winged helix DNA-binding domain (wH-DBD) and a C-terminal antiparallel coiled coil helical domain. The overall structure shows as a molecular chimera with significant folding similarity of its DBD to the bacterial SmtB/ArsR family, while its C-terminal part was found to be a remote homologue of the eukaryotic BAG domain. The dimeric protein recognizes a palindromic DNA sequence. Molecular docking and mutational analyses suggested a novel binding mode in which the major specific contacts occur at the minor groove interacting with the strongly basic wing containing a cluster of three arginine residues.

Keywords: archaea; Phr; heat shock; transcriptional regulation; winged helix

Introduction

It has been suggested that the last common ancestor of all life on Earth may have been a hyperthermophile before divergence of the three domains.[1] Most hyperthermophiles belong to the kingdom of archaea, which constitute a fundamental domain of life distinct from Bacteria and Eukarya. Genetic and biochemical work have shown striking parallels of archaeal basic biochemistry with the other two domains. Many cellular components of the primary information-processing systems, such as DNA replication, transcription and translation in archaea, work in similar ways to those in eukaryotes.[2] In contrast, many biosynthetic or metabolic processes in archaea are more similar to the same bacterial systems than to the eukaryotic counterparts.[3]

Archaea possess a basic transcriptional machinery resembling the eukaryal RNA polymerase II (RNAP II) apparatus, but with simplified features which only requires two basic transcription factors, TBP and TFB, for transcription initiation.[4–6] Surprisingly, most specific transcriptional regulators identified in archaea so far seem to be more bacterial-like as a plethora of homologues of bacterial regulatory factors were found in the archaeal genome.[5,7,8] The most widely represented archaeal DNA-binding proteins with known or surmised gene-regulatory potential are related to members of the bacterial Lrp/AsnC family,[9] which influence cellular metabolism in both a global (Lrp) and specific (AsnC) manner.[10]

All living organisms share a common molecular stress response upon rapidly up-shifted environmental temperature. The heat shock response is characterized by a dramatic change in gene expression patterns and elevated syntheses of a family of heat shock proteins (Hsps), most of which function as molecular chaperones in preventing the aggregation of denatured proteins and/or helping protein refolding. Hyperthermophilic archaea lack the Hsps Hsp90, Hsp70/DnaK, DnaJ, GrpE, Hsp33 and Hsp10 homologues.[3,11] The subset of chaperones

Abbreviations used: wH, winged helix; DBD, DNA-binding domain; PDB, Protein Data Bank; HSF, heat shock transcription factor.

E-mail address of the corresponding author: wei.liu@biosci.ki.se

found in extremophiles thus consists of a minimal protein-folding machinery,[3] which may represent a prototype of anti-stress systems in early life.

The expression of most heat shock genes is strictly repressed under normal conditions, but activated once stress response is triggered. The mechanism of heat shock regulation differs between bacteria and eukaryotes. Bacteria utilize alternative sigma factors to enhance the transcription of Hsps,[12,13] while in eukaryotic cells, the enhanced synthesis of heat shock proteins upon stress stimulation is regulated by the heat shock transcription factors (HSFs). Inducible trimerization is required for HSF activation before it is transferred from the cytoplasm into the nucleus and specifically bind to heat shock elements (HSEs) upstream of the promoter regions in heat shock genes.[14,15]

Compared to bacteria and eukaryotes little is known on heat shock regulation in archaea. No homologues of bacterial sigma factors and eukaryotic heat shock factors have been detected in archaeal genomes, indicating that archaea may have developed a machinery different from the other two domains to cope with stress conditions. The first transcriptional regulator selectively inhibiting cell-free transcription of archaeal heat shock promoters has been recently identified from the hyperthermophilic archaeon *Pyrococcus furiosus* (Phr) (29). The 24 kDa protein forms a homodimer and specifically inhibits the transcription *in vitro* by binding to a 29 bp DNA sequence overlapping the transcription start site in heat shock promoters. Phr does not affect binding of TBP and TFB to the promoter, but abrogates RNA polymerase recruitment to the TBP/TFB complex at optimal temperature. EMSA and DNaseI footprinting analyses showed that Phr recognizes a palindromic nucleotide sequence: 5′-TTT..T..C…..G..A..AAA-3′, which is conserved in heat shock promoters in *P. furiosus* and *Pyrococcus abyssi*.[3,16] Here, we report the crystal structure of Phr refined at 2.6 Å resolution using the multiwavelength anomalous diffraction (MAD) technique. The structure showed some surprising features. It contains two distinctive domains, which are remotely homologous and structurally related to two different superfamilies from bacteria and eukaryotes, respectively. We also derived a model of the Phr-DNA complex by molecular docking that agrees well with site-directed mutations on both the protein and the palindromic DNA recognition sequence. Our studies identify Phr as novel type of transcriptional regulator mediating modulation of the heat shock response in archaeal cells.

Results

Subunit structure

The secondary structure of Phr is dominated by α-helices (56.7%) while only 14.8% of residues form β-strands. The 202 residue-long subunit comprises two distinctive domains (Figure 1(a)). The N-terminal domain (residues 1–97) conforms to the fold of the winged helix DNA binding domain (wH-DBD), a subclass belonging to the large ensemble of helix-turn-helix (HTH) proteins.[17,18] As the characteristics of members from this subfamily, three α-helices (H2–H4 in Phr) form a right-handed helical bundle resembling the basic three-helical core, followed by a C-terminal β-strand hairpin (the wing) opposite to the helical core (Figure 1(a)). Compared with the orthodox wH motif, Phr contains an N-terminal helical extension (H1), as observed in bacterial metal repressors such as SmtB[19] and CadC,[20] and some eukaryotic transcription factors like E2F4.[21] Correspondingly, H2, H3 and H4 form the tri-helical core where H4 is the recognition helix potentially inserting into the major groove upon DNA binding. A short β-strand (β1) between H2 and H3 forms the first antiparallel sheet together with β2 and β3 following the tri-helical core. The β-hairpin connecting β2 and β3 represents the wing showing high flexibility as reflected by the *B*-factor profile. As a variant of the bacterial wH proteins, like SmtB, CadC or RTP,[19,20,22] Phr lacks a C-terminal helical extension. Instead, two antiparallel strands, β4 and β5, following the wing compose the second sheet in Phr (Figure 1(a)).

The C-terminal domain (residues 105–202) comprises four α-helices: H5 to H8, which are antiparallel to each other. The most striking feature of this domain is the presence of a long helix (H6) consisting of 46 amino acid residues (121–168). Notably, the segment (120–172) was predicted as a potential coiled coil region by the COILS server,[23] with greater than 90% probability in a 28 residue window. It spans 7.4 heptad repeats where isoleucine residues outnumber leucine residues at positions *a* or *d*, which could be a potential property of hyperthermophilic proteins.[24] The two domains in the Phr structure are linked by a highly flexible peptide (98–104) on which a 3_{10} helical turn is present. The C terminus remains unstructured in the refined model and the last four (chain B) or five (chain A) residues are missing in the electron density map.

Dimeric structure

Phr forms a stable homodimer with overall dimensions of 80 Å×52 Å×86 Å in the crystal (Figure 1(b) and (c)). The two subunits in the asymmetric unit are related by 2-fold local symmetries, but the symmetry axes for both domains are notably separated by 6.4 Å although they are parallel to the *y* axis in the orthogonal system (Figure 1(b)). The distance is large enough to cause considerable asymmetry between the two subunits. Such a large spatial shift is probably attributed to crystalline packing; however, a natural conformation like this cannot be excluded. Whatsoever it indicates the high flexibility between the two domains. In the top view, both domains have an elongated dimeric shape but with an intersecting angle of approximately 60°, conferring an intriguing "X" figure on the overall structure (Figure 1(c)). The