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

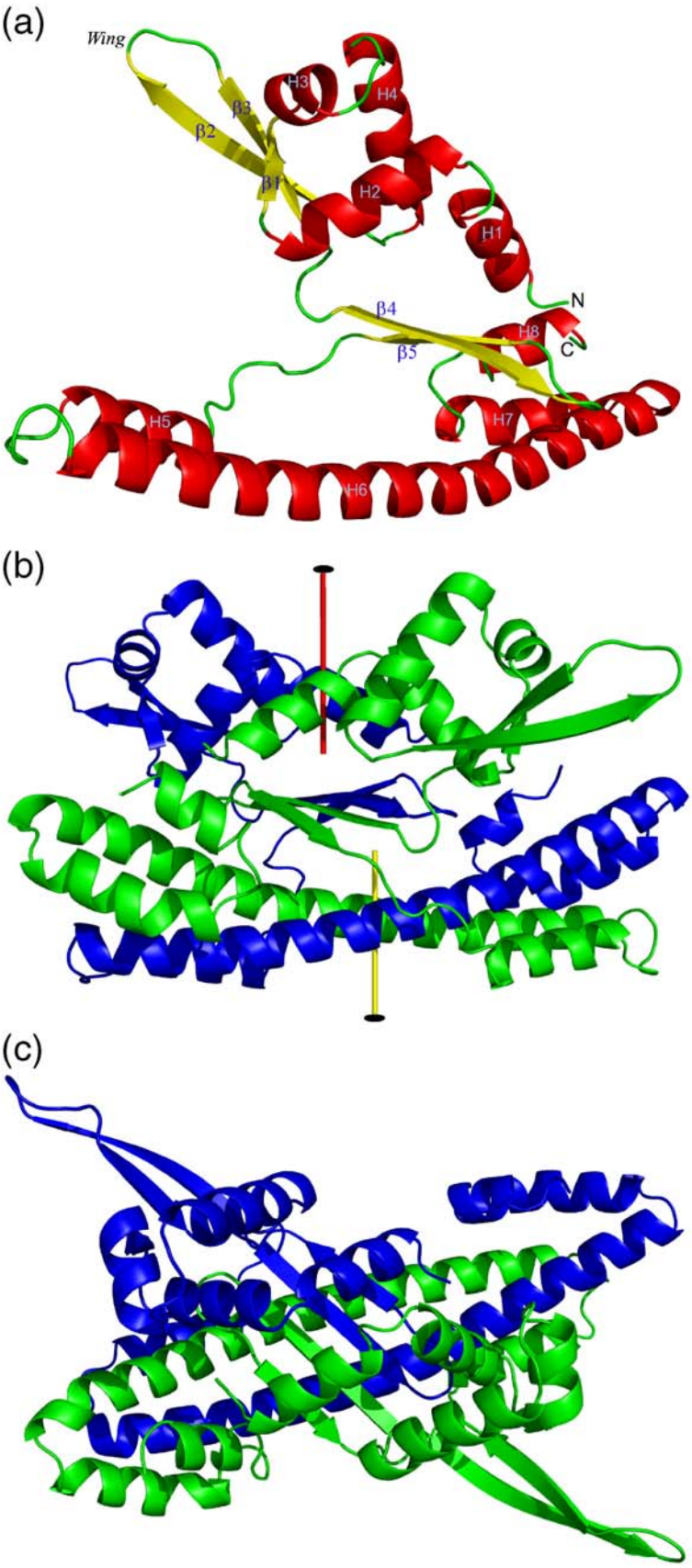

Figure 1. Ribbon model of Phr. (a) Subunit structure; α-Helices, β-strands, loops shown in red, yellow and green, respectively. Cartoon model of a Phr dimer, (b) side view and (c) top view; chains rendered in blue and green; local 2-fold axes are indicated.

wings on the wH domain protrude to opposite directions, offering a broadened DNA binding region covering a minimal length of 29 bp for the DNA fragment bound to Phr.[16] In the C-terminal domain, the long helices from both subunits, H6 and H6′, are wrapped around each other to form a left-

handed coiled coil. The middle one third of the coil is exposed to solvent with weak contacts from the remainder of the protein. A four-stranded antiparallel helical bundle is formed at either end of the coil, with the involvement of H5 from the same monomer and H7′ from the partner subunit.

High stability of the dimeric structure can be expected as more than one third of the amino acids from each subunit are involved in forming a large dimeric interface area (ca 4200 Å^2). In addition to hydrophobic interactions, 18 hydrogen bonds as well as two salt-bridges are also formed at the interface. The interactions held between the subunits can be divided into three main regions. Firstly, a hydrophobic core is formed at the DNA-binding domain with an antiparallel helix bundle held by H1 and H2 from both subunits. A hydrophobic cluster of four leucine residues, L9, L12, L9′ and L12′, is found at this region. Secondly, a hybrid β-sheet is present in the central part of the dimer. The formation of this four-stranded antiparallel sheet with the structure elements β4, β5, β4′ and β5′ contributes most hydrogen bonds between the subunits, thereby playing a key role for Phr dimerization. In fact, a truncated construct with only residues 1 to 97 still exists as a stable dimer in solution and the crystal (unpublished data). The third contact region is formed in the C-terminal domain, where the characteristic "knobs-into-holes" packing in the coiled coil represents dominating interactions between hydrophobic side-chains from both subunits. Charged residues at most *e* and *g* positions leading to rich inter-helical electrostatic interactions also contribute numerous interactions to the zipper-like coiled coil.[25,26]

Accessible surface and electrostatic potential

The Phr protein shows strong molecular polarity due to apparent uneven distribution of charged amino acids between the two domains. A basic N-terminal DNA binding domain and an acidic C-terminal domain can be easily distinguished by the calculated p*I* values, 9.62 and 4.99, respectively. This property is also clearly reflected by the electrostatic potential map on the Phr surface. The DBD is dominated by positive charges as expected, but shows some atypical properties with respect to canonical winged helix proteins (Figure 2(a)). The protein exhibits clustering of basic residues (R67, R69, R71 and K72) on the wing loop but displays a largely neutral surface of the helical core including the recognition helix (H4). These features are consistent with those found in cyanobacterial metallothionein repressor SmtB,[19] the eukaryotic histone protein GH5[27] and human regulatory factor X (hRFX1).[28] However Phr shows even stronger basicity on its wing than those proteins. Gajiwala *et al.* suggested to classify the winged helix proteins into two groups.[17] The first category is represented by typical wH proteins like HNF-3γ, DP2, E2F4 and genesis, which have a basic surface mainly on the recognition helix; the second group in contrast includes atypical wH proteins, e.g. hRFX1, GH5 and SmtB, all displaying a large number of basic residues on the wing surface. Apparently, Phr belongs to the latter class and hence may well use its wing region for DNA recognition as hypothesized by Gajiwala *et al.*[17] Since Phr is believed to bind DNA as a homodimer, like bacterial DNA-binding proteins, this assumption will explain well the consensus DNA binding sequence 5′-TTT..T..C…..G..A..AAA-3′ present in *Pyrococcus* heat shock promoters. The three consecutive thymine residues (adenine residues on the other strand) at the ends are assumed to be in contact with arginine side-chains on the wings allowing for the spatial orientation of the DBD in Phr (Figure 5). The determinant role of TTT in Phr binding has been revealed from the mutational studies on bound oligonucleotides (Figure 6(b)).

A very acidic region on the C-terminal electrostatic potential surface represents a striking feature of the Phr structure, as the majority of negatively charged residues are densely distributed on the coiled coil surface (Figure 2(b)). Such an unusual exposed surface may indicate itself to be a possible binding surface for other molecules, mostly likely for basic components involved in the regulation of Phr. This assumption is strengthened by the calculation of the surface curvature, which is defined as accessibility of water to a single water molecule placed in contact with a particular surface point. Clearly, big cavities denoted in grey color can be seen at the C-terminal region (Figure 2(c)). In fact the cavity at the interface of the two subunits on the coiled coil region forms the largest cleft as analyzed by the PDBsum server.[29]

N-terminal SmtB/ArsR DNA binding domain

The Phr structure was uploaded to Dali,[30] CE[31] and SSM[32] servers for comparison with all structures in the Protein Data Bank (PDB). Although there was no overall match with the entire structure of Phr, a number of hits were observed for the DNA-binding domain. The top two hits were the cyanobacterial SmtB (1SMT)[19] and the *Staphylococcus aureus* pI258 CadC (1U2W).[20] Both proteins belong to the bacterial SmtB/ArsR family, a major class of metal-sensing transcriptional repressors in prokaryotes. The proteins in this family are homodimeric repressors and all contain a wH DNA-binding domain and one or two potential metal-binding sites.[33] Comparison of these two structures with the Phr C^{α} model reveals a similar tertiary fold of the DBDs (Figure 3(a)), with an r.m.s.d of 1.2 Å for each. Sequence alignment further uncovers the evolutionary relations among these proteins (Figure 3(b)). Phr shares 29.9% and 27.6% sequence identity with SmtB and CadC, respectively. A certain degree of similarity was also found between Phr and other bacterial wH proteins. RTP from *Bacillus subtilis*[22] for example shows considerable similarity in its 3D fold with Phr (r.m.s.d. of 2.2 Å), but the sequence identity between them is only 12.3%. Compared to eukaryotic proteins containing a wH-DBD, like

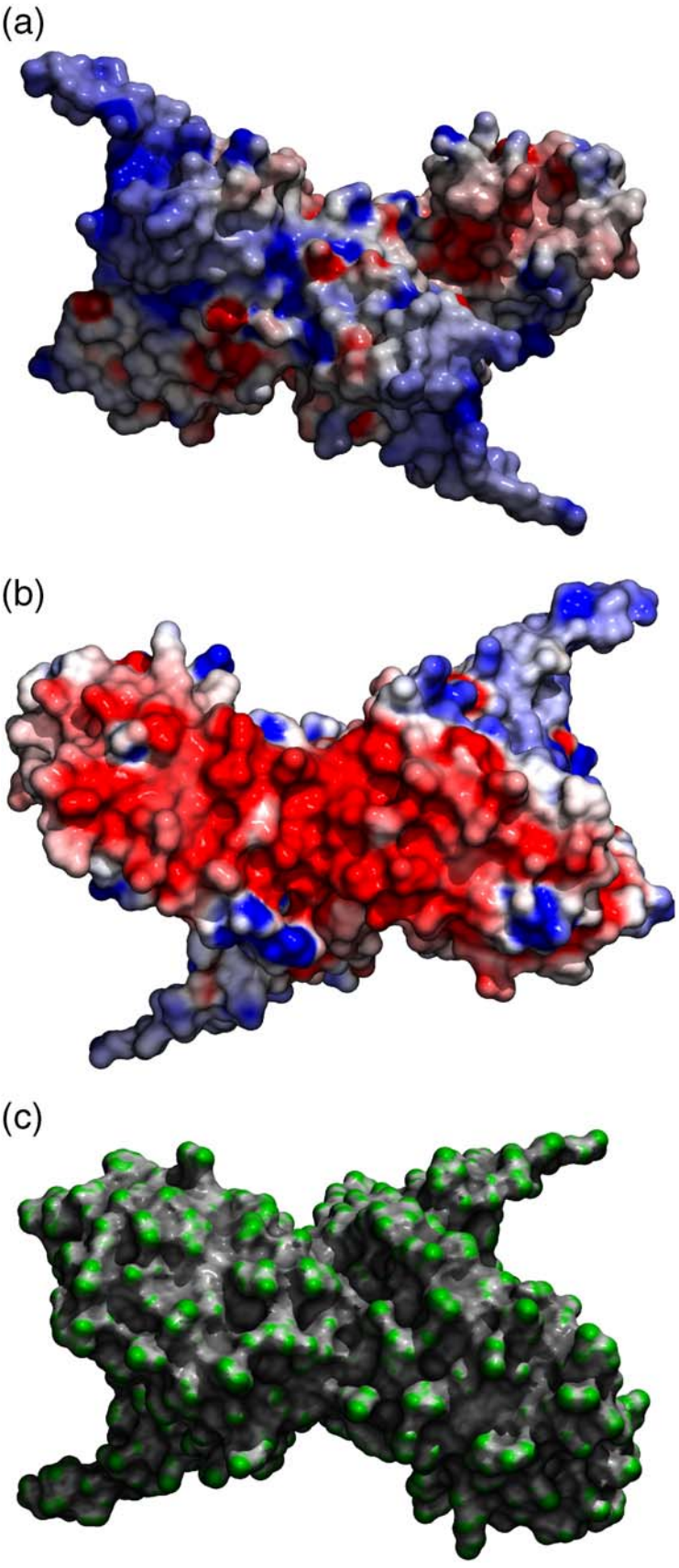

Figure 2. Electrostatic potential maps ((a) and (b)) and surface curvature (c) of a Phr dimer (program GRASP[58]). (a) Winged helix domain; blue, positive charges; red, negative charges. (b) The Phr C-terminal domain shows an exposed acidic surface; (c) surface curvature, cavities shown in grey, water accessible surface shown in green.

HNF-3γ, hRFX1, E2F4 and DP, Phr displays a large r.m.s.d (ca 5.0 Å) and poor sequence identity (2–7%). These data above clearly suggest that the DBD of Phr has a common molecular ancestor with the bacterial SmtB/ArsR protein family. Despite the above similarity to SmtB and CadC, a notable structural

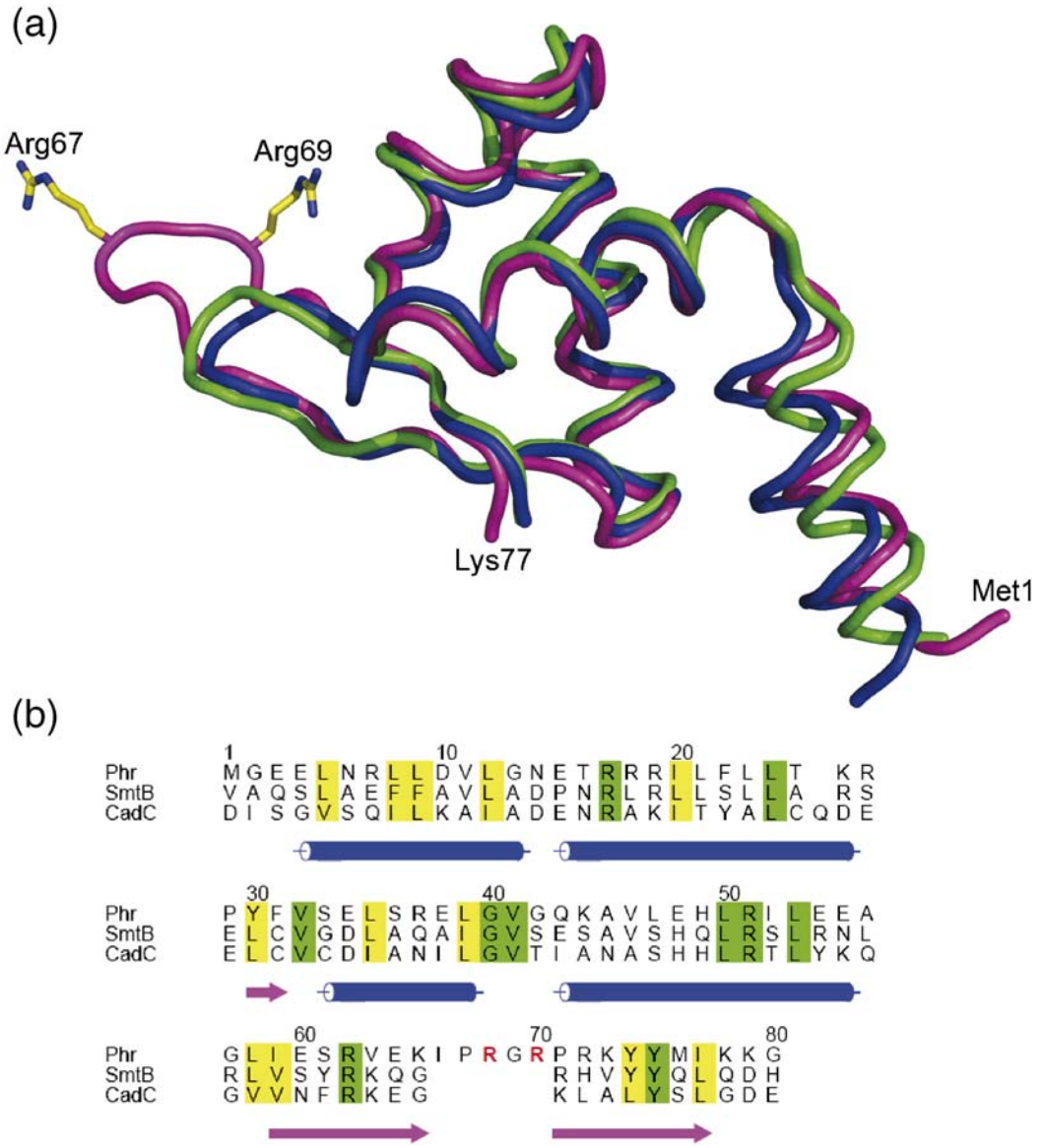

Figure 3. Structure and sequence comparison of Phr, SmtB (1SMT),[19] and CadC (1U2W).[20] (a) Superimposed C^{α} models of the winged helix DNA-binding domain of Phr (pink), SmtB (green) and CadC (blue). Side-chains of R67 and R69 on the wing loop in Phr displayed as ball-and-stick model. (b) Sequence alignment of the DNA-binding domains, α-helices shown as blue cylinders, β-strands as pink arrows; R67, R69 are shown in red color.

difference was found at the wing region. Obviously, Phr has a longer loop, as a result of a five-amino-acid insertion between β2 and β3 (Figure 3). Two additional arginine residues, Arg67 and Arg69, located on this insertion are noteworthy as the site-directed mutagenesis analysis has confirmed their crucial roles in contact with double-stranded DNA. The increased length of the wing accounts for the longer binding nucleotide sequences of Phr (29 bp) than those of SmtB or CadC (12+2+12).[33]

C-terminal BAG-like domain

No apparent structural neighbors have been found for the Phr C-terminal domain in the PDB. Surprisingly, however, both the SSM and Dali servers provided a hint that Phr may topologically be related to the eukaryotic BAG domain. The BAG domain, defined as Bcl-2 associated athanogene, is a conserved region at the carboxyl terminus of BAG proteins with approximately 110 amino acid residues in length and folds as an anti-parallel tri-α-helical bundle.[34,35] BAG family proteins bind through the BAG domain to the ATPase domain of heat shock proteins 70, both constitutive Hsc70 and inducible Hsp70.[36,37] In order to confirm the relationship between Phr and the BAG domain, a structure-based sequence analysis was done against the SUPERFAMILY database.[38] The C-terminal sequence of Phr was compared with all known SCOP domains. As shown in the sequence alignment (Figure 4(a)), Phr shares significant similarity with the human BAG domain, reflected in 23% and 44% of sequence identity and similarity. This is an

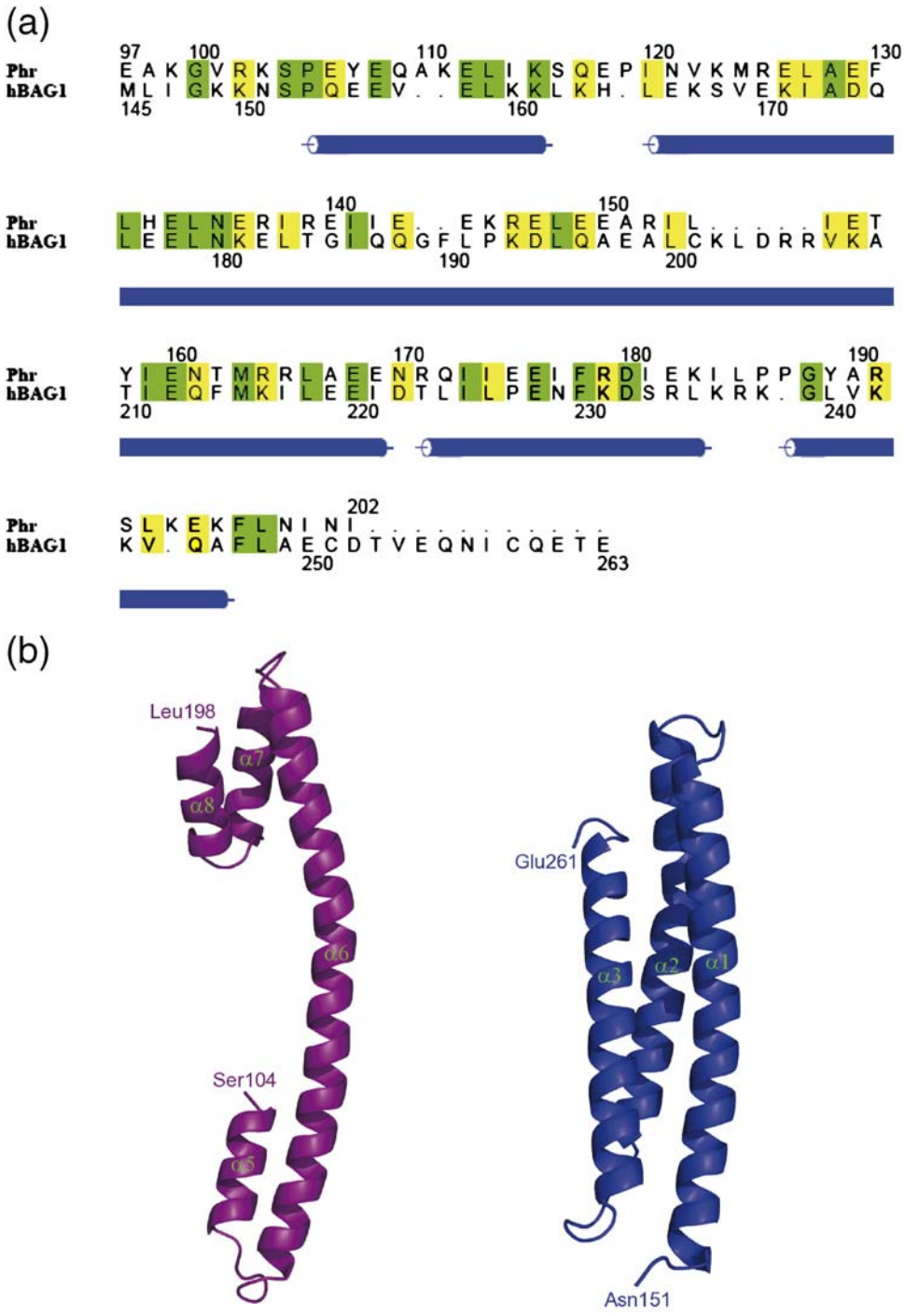

Figure 4. Sequence and topology comparison between Phr C-terminal domain and human BAG domain (1HX1_B).[35] (a) Sequence alignment between the two domains, identical and similar amino acids shown in green and yellow, respectively; structure elements drawn according to Figure 1. (b) Structure comparison between Phr (left) and human BAG domain (right).

interesting observation considering so diverse evolutionary positions of the two organisms, inferring the C-terminal part of Phr to be a remote homologue of the eukaryotic BAG domain. In the structural comparison, the two structures look topologically alike although unable to be superimposed perfectly. Both comprise anti-parallel α helices folding into similar helical bundles, but with different length (Figure 4(b)). In spite of the apparent structure variance, it is still suggestive to classify both structures into the same superfamily and thus establish the homology between them on account of the significant similarity in sequence.

DNA binding model

The sequence and structural similarity among the DNA-binding domain suggests a likewise pattern of

DNA–protein interaction among Phr, SmtB and CadC. The two recognition helices (H4) of the wH domain in the Phr dimer are separated by roughly 34 Å, corresponding well to the distance between adjacent major grooves in *B*-form DNA. Since the structures in complex with DNA are still unavailable so far, a docking model of Phr was built with its binding sequence as a prediction of the Phr-DNA binding mode (Figure 5). A straight 35 bp *B*-form DNA with the binding sequence of the *aaa*$^{+}$ *atpase* promoter was modeled and docked onto the surface of the N-terminal part of Phr using the information-driven docking program HADDOCK.[39] In the docking models the two recognition helices insert into the major grooves of adjacent turns. The best solutions with imposed 2-fold axes on both the protein dimer and the DNA were chosen as the initial models. Too close contacts were removed by rigid body energy minimization. Restrained flexibility on both DNA and the protein was introduced by subsequent semi-flexible and water refinement. In the final model with lowest energy (Figure 5), a number of amino acid residues on both the HTH motif and the wings are clearly involved in DNA binding. A striking residue is K43 located on H4 that directly contacts the sequence TNNC in the major grooves, forming hydrogen bonds with some bases. However, more significant contacts occur between the three arginine residues on the wing and the T-rich region at the end of the consensus nucleotide sequence. The side-chain of R69 completely inserts into the minor groove and contacts the thymine bases of TTT. R71 is located in the vicinity with its side-chain oriented toward the minor groove as well. Although R67 is present outside the minor groove, electrostatic interactions are formed through its guanidine group to the phosphates nearby the TTT sequence, thereby strengthening the local contacts. Other residues on the HTH motif, e.g. Q42, H48 (H4), T16, R17 (H2) and S32 (H3), are also involved in contacts with the phosphate and sugar groups on the DNA backbone. No apparent movement of the main chain in Phr has been observed in comparison between the free and the DNA-binding models, but conformational changes of side-chains close to the protein–DNA interface were considerable, especially in the wing region, which gave rise to an r.m.s.d of 2.5 Å in an all-atom calculation. Almost all the side-chains at the interface involved in favorable interactions are pointing to the bound DNA in the docking model. Since bent DNA conformation occurs quite often in complex structures upon protein binding, like the case of CAP-DNA complex,[40] our binding model however may not be the optimum because of the computational limit for DNA bending in the docking algorithm. In terms of interaction energy, a somewhat bent DNA and reasonable conformational changes on Phr, particularly at the wings that may move closer to the DNA, are likely to be required for favorable binding. Actually high flexibility of the wing region has been indicated by considerable movements observed in different crystal forms (data not shown).

Mutagenesis experiments

A list of candidate amino acid residues that may be involved in DNA binding was proposed from the above model and the corresponding mutants were prepared in order to test their roles. The EMSA results given in Figure 6(a) clearly show that most

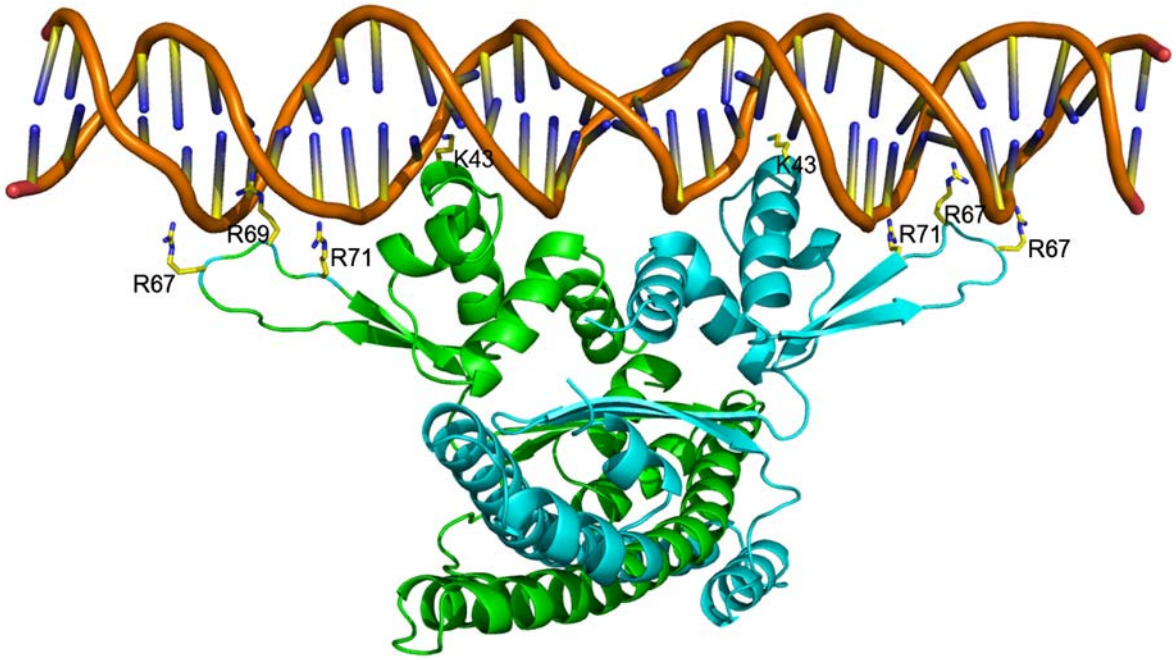

Figure 5. Phr-DNA docking model calculated by HADDOCK[39]; Phr dimer and the DNA duplex represented as cartoon model, all amino acids playing significant roles in DNA recognition are shown by a ball-and-stick model.

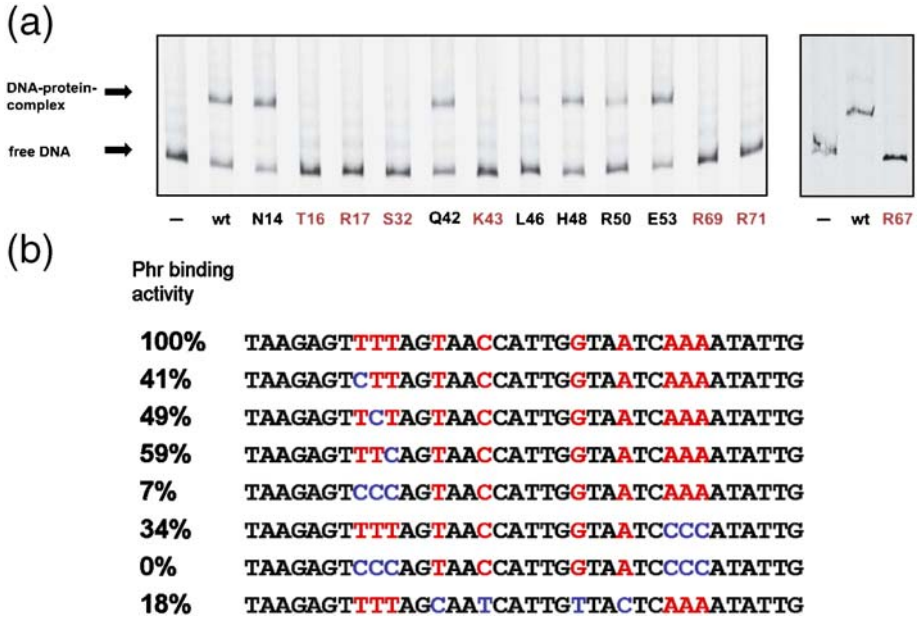

Figure 6. Mutational analyses on Phr and oligonuleotides. (a) EMSA results according to the mutation experiments on Phr, red letters representing amino acids leading to maximal loss of DNA binding activity. (b) Summarized results of mutations on the *aaa+ atpase* promoter; relative binding activity listed on the left side; nucleotides labeled in red form the palindromic DNA sequence, and those in blue are the mutation positions.

candidates are significantly involved in contact with DNA, which argues for the correctness of the docking model. The mutational studies indicate that the wing of Phr plays more decisive roles than the canonical recognition helix in DNA binding, since any single point mutants in the arginine cluster R67, R69 and R71 would lead to the complete loss of the DNA binding ability whereas only Lys43 on H4 had the equivalent effect. Moreover, Thr16, Arg17 (both on H2) and Ser32 (H3) produce null mutants as well, indicating their significant contributions to Phr-DNA affinity although in our model these residues contact the phosphate-sugar backbone rather than the bases of the bound DNA. Correlated to the mutational data on Phr, mutants on the oligonucleotide highlighted the determinant role played by the T-rich region at both ends of the Phr binding sequence (Figure 6(b)), pinpointing the significant contacts from the arginine cluster on the wing of Phr. According to the model, Phr uses particularly the wing region for DNA recognition similar to hRFX1.[28] In striking contrast to hRFX1 that dominantly binds to the major groove, however, the wing of Phr contacts the minor groove instead. Given that the major specific contacts occur in the minor groove of DNA, the docking model presented here has revealed an unprecedented way of DNA binding.

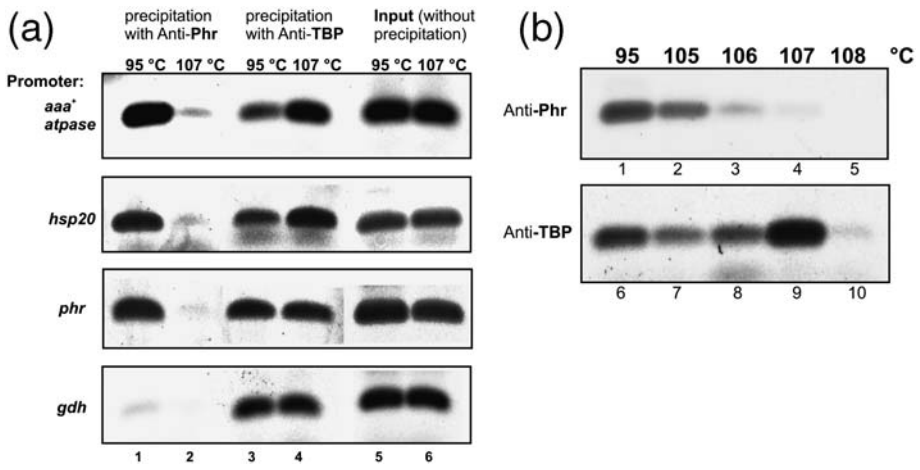

Figure 7. Temprature-dependent Phr-binding to heat shock promoters *in vivo*. (a) Chromatin immunoprecipitation experiments (Chip) with *P. furiosus* cells crosslinked at 95 °C and 107 °C. (b) Chip experiment with *P. furiosus* cells crosslinked at 95, 106, 107 and 108 °C, respectively, using antibodies against Phr and TBP.

In vivo evidence for the significance of Phr as heat shock transcriptional regulator

Phr binds specifically to heat shock promoters of the genes encoding Hsp20, the AAA^+ ATPase and of Phr itself *in vitro*.[16] To investigate the physiological role of Phr *in vivo*, we performed assays based on chromatin immunoprecipitation procedures, which have been successfully applied to the detection of intracellular DNA binding of an archaeal regulator.[41] As shown in Figure 7(a), Phr binding to all these three genes was clearly observed at 95 °C, in contrast to the weak binding to *gdh*, the control gene. These findings suggest that Phr binds specifically to heat shock promoters in *Pyrococcus* cells under normal growth condition. After exposure to heat shock, Phr binding was greatly reduced (Figure 7(a)). This result is consistent with experimental data obtained from *in vitro* DNA-binding and cell-free transcription experiments.[16] As a control protein, TBP strongly binds to heat shock promoters and *gdh* as well both at 95 and 107 °C (Figure 7(a)). This obviously indicates that Phr binding does not interfere with TBP-binding to the promoter as proposed from *in vitro* studies. More importantly, binding of TBP to the two heat shock promoters *aaa+ atpase* and *hsp20* was even increased after heat shock, indicating the activated expression of these proteins under stress as predicted. Consistent with this view, Phr binding to the *aaa+ atpase* promoter decreased with increasing temperature from 95 to 108 °C (Figure 7(b)). To study the physiological relevance of the heat shock reaction, the viability of cells was studied by exposure to 107 °C. No apparent difference between heat shocked and the control culture was observed, both showing growth in the 10^{-8} dilution indicating that *Pyrococcus* survives heat shock exposure for 30 min at 107 °C (data not shown). Higher temperatures turned out to be lethal, consistent with our observation that TBP-binding to the aaa^+ *atpase* promoter was disrupted after exposure to 108 °C.

Discussion

Phr shows a novel mode of DNA recognition

The work on Phr reported here describes the first structure of a heat shock transcription factor in the third domain, archaea. It reveals a stable dimeric structure of a novel DNA-binding protein that regulates the heat shock response in *Pyrococcus*, a hyperthermophilic microorganism with an optimal growth temperature around 100 °C. The winged helix DNA binding domain with an N-terminal helical extension confers a similar fold to proteins belonging to the bacterial SmtB/ArsR family. The similarity in structure and sequence to eukaryotic wH proteins, like hRFX1, HNF-3γ and E2F4, is poor however. From these structural comparisons, it seems more likely that Phr adopts a bacteria-like DNA binding mode. Many wH proteins involved in DNA binding as homodimers originate from bacteria, for example CAP,[42] RTP[22] and FadR,[22] while many eukaryotic proteins were observed to cooperatively bind DNA as monomers, such as hRFX1, HNF-3γ and HSF1.[43] Phr exists as a dimer and recognizes an imperfect palindromic nucleotide sequence with 29 bp in length, both of which strongly indicate that this protein binds DNA as a dimer, like SmtB and CadC.

The calculation of the electrostatic potential implies that the regulator may rely more on its wings than the recognition helix for DNA binding due to an imbalanced distribution of charged residues on these two regions (Figure 2(a)). This property is similar to bacterial SmtB, eukaryotic hRFX1 and GH5, and seemingly Phr can be classified into the group of atypical winged helix proteins as defined by Gajiwala *et al.*[17] However, according to our docking model based on biological information including the consensus recognition sequence of DNA and the invariant amino acid residues in the DBD of Phr, SmtB and CadC, the Phr-DNA binding mode is obviously very different from RFX1. Human RFX1 uses the wing (W1) for extensive contacts with the major groove, while only a single side-chain from the recognition helix interacts with the minor groove. By contrast, in our model the wing region of Phr interacts *via* a cluster of three arginine residues with the minor groove. At least one arginine side-chain (R69) inserts into the minor groove making direct contacts to the bases of the T-rich segment at the ends of the binding sequence (Figure 5). A similar contact occurs in the E2F4-DP-DNA complex,[21] in which an arginine binds in the minor groove near the TTT stretch at an extended binding site. The other two arginine residues (R67 and R71) in Phr partially contact the nucleosides in the minor groove or make electrostatic interactions with the backbone phosphate groups in the vicinity. As K43 in the recognition helix H4 is the only residue binding in the major groove, in the case of Phr the major specific protein–DNA contacts are formed at the minor groove. The significance of all these three arginine residues has been demonstrated by our mutagenesis experiments, which provide an important experimental evidence to support the above conclusion. Since such a binding mode has not yet been reported, the model presented here probably represents a novel mechanism for wH protein–DNA recognition.

Eukaryotic features of the C-terminal domain and evolutionary implications

One surprising feature deduced from our structural analysis is the striking similarity between the eukaryotic BAG domain and the Phr C-terminal domain. As shown in Figure 5, Phr shares 23% sequence identity and displays topological similarity to the human BAG domain. This finding renders an intriguing chimeric property on Phr as its N and C-terminal domains represent bacterial and

eukaryal features, respectively. Since no homologous sequence has been reported in prokaryotes, this finding may allow the establishment of a remote homology between Phr and the eukaryotic BAG domain. The amino acid sequences of the BAG domain are rather conserved among eukaryotic organisms,[36] suggesting their evolution from a common ancestor. A widely accepted hypothesis suggests that all life on Earth might have evolved from a high temperature origin and the emergence of the three domains of life was thermally inevitable due to the hot initial conditions of our planet.[1,44,45] In this sense, the molecular evolution of BAG domains may also follow the route from hyperthermophiles to mesothermophiles. A seemingly plausible scenario is that, as a remote homologue of the eukaryal BAG domain, Phr itself is probably a closer relative to the common ancestor. This amazing similarity may lead to a hypothetical conclusion that the gene encoding the BAG domain existed in an original gene pool before the divergence of archaea and eukarya. If this is true, Phr might thus represent a relict of the primordial form of the BAG domain, and be considered as a homologous variant functioning at high temperature. The physiological function of eukaryotic BAG proteins is to bind to the ATPase domain of Hsc70/Hsp70 through the helical bundle motif, and to regulate the chaperone activities. No homologues of Hsp70 are encoded in the gene pool from extremophiles, like *Pyrococcus*.[3,11] However, Hsp60-like chaperones (thermosomes) and/or AAA chaperones have been detected in the *Pyrococcus* genome and could be potential binding targets of the BAG-like domain in Phr. It will be of great interest to experimentally testify this assumption of a feedback mechanism that is triggered upon heat shock response by the interactions between chaperones and Phr *via* its C-terminal binding surface.

Biological implications

Although it has been known that Phr functions as a negative regulator of heat shock genes *in vitro* from previous studies,[16] we had little knowledge on how the cell works to derepress the modulated genes under stress in *Pyrococcus*. Our finding that Phr binds heat shock promoters *in vivo* during growth at 95 °C but dissociates at 107 °C provides direct evidence that Phr is indeed involved in repression of heat shock promoters during normal growth (Figure 7(a)). Apparently the observed stimulation of RNA synthesis from these promoters upon heat shock[3,16] results from the dissociation of Phr and its DNA-binding sequence. Both the *in vitro*[16] and *in vivo* data provided here are consistent with the following simple model for heat shock regulation. Phr binds under physiological growth conditions to the promoter region of heat shock genes, thereby inhibiting transcription by physically blocking RNAP recruitment. Subsequent release of Phr along with elevating temperature leads to activated expression of heat shock proteins.

It is presently unclear whether dissociation of Phr is mediated simply by temperature-induced conformation changes or by a more sophisticated mechanism. More experimental data are required to distinguish between these two possibilities. Our structural analysis suggests the possibility that the acidic C-terminal domain of Phr is involved in a signal transduction pathway sensing heat shock in *Pyrococcus* cells. A large repertoire of dimeric transcription factors have been observed to utilize antiparallel coiled coils for dimerization, for example the MerR protein family.[46] However, the case of Phr is different. The C-terminal coiled coil does to a certain degree contribute to intersubunit interactions but is not necessarily required for dimerization because a truncated construct containing only the N-terminal domain (residues 1–97) showed the same dimeric assembly as the full-length protein in solution (unpublished data). The acidic nature of the C-terminal domain is reminiscent of Tfx, a potential eukaryal-like transcriptional regulator found in *M. thermoautotrophicum*.[47] The N-terminal portion of Tfx contains a HTH motif, and like Phr, the C-terminal part contains an acidic domain, possibly involved in interactions with the basal transcription factors TBP and/or TFB.[47]

The heat shock response in living cells, both bacterial and eukaryotic, is mediated by a complex system with the participation of many cellular components, rather than a single event. Most likely therefore, cooperative regulation will exist also in archaea, where Phr plays a unique role to block RNAP recruitment. The calculated electrostatic potential map and molecular curvature (Figure 2) both suggest that the exposed surface on the Phr C-terminal domain might be a binding surface for other cellular factors, in particular those with basic components. Inducible interactions of some unknown cofactors, proteins and/or small molecules with the acidic C-terminal domain of Phr may represent a key step for heat shock regulation by affecting the DNA binding activity of Phr. Although no homology between Phr and eukaryotic HSFs was found, they seem to be linked in terms of domain arrangement. The N-terminal half of HSFs contains a winged helix DNA binding domain followed by a potential coiled coil (HR-A/B) region,[14,48] resembling the overall structure of Phr. In eukaryotic cells, feedback modulations have been proposed in the up-regulation of HSF1, at different levels from Hsps and co-chaperones, for instance the Hsp70 and Hsp90-containing complexes.[49,50] *Pyrococcus* lacks the homologues of Hsp70 and Hsp90 but the striking similarity between the Phr C-terminal domain and the BAG domain is reminiscent of eukaryotic BAG proteins to functions as co-chaperones in binding with Hsc70/Hsp70. If it can be demonstrated that Phr is able to interact with some heat shock factor existing in *Pyrococcus*, a feedback mechanism could be established accordingly.

Materials and Methods

Purification, crystallization and data collection

Expression and purification of the native Phr protein has been reported.[16] The seleno-methionine substituted protein was produced according to the standard protocol followed by mass spectrometry analysis to confirm full incorporation of seleno-methionine residues. Crystals of Phr were grown by the sitting-drop vapor diffusion method from buffered ammonium sulfate solutions with 20–25% of ethylene glycol as an additive. Crystal growth seemed very sensitive to pH. The seleno-methionine derivatives were obtained under similar conditions but with addition of DTT up to 10 mM. MAD data were collected at the EMBL beamline BW-7A, Hamburg, Germany using three wavelengths around the absorption edge of selenium (Table 1), while the collection of native data was done at the ESRF beamline ID 23-1, Grenoble, France. Data evaluation was performed by the HKL package or the HKL-2000 program suite.[51] The crystals belong to space group $P2_12_12$ with cell parameters $a=52.3$ Å, $b=82.8$ Å and $c=114.4$ Å. One dimer resides in the asymmetric unit with solvent content of 52%.

Structure solution and refinement

The selenium site substructure was solved by running the program SHELXD. Ten possible selenium sites were found in the top solution with a sharp drop distinctive from those peaks with lower height. The number of sites agrees well with the methionine content in the sequence and the MS results. All these sites were then used in the program SOLVE for site refinement and macromolecular phasing.[52] An interpretable electron density map was obtained after phase calculation at 3.0 Å resolution, with an overall figure of merit of 0.63. RESOLVE from the same program suite was subsequently used to improve the density map by means of statistical density modification,[53] and a partial model was built thanks to an autobuild algorithm implemented in RESOLVE.[52] The model was completed in cyclic way by manually building followed by phase combination and density modification. This approach seems quite helpful for density recognition at a resolution limit of 3 Å.[53] The final model was later used as a search model for molecular replacement in the native data measured at higher resolution (2.6 Å). Iterative cycles of model adjustment in O and positional refinement in the program REFMAC[54] were done to improve the model. Since TLS refinement has been proven to be important at moderate resolution,[55] ten cycles of TLS refinement were performed at the beginning of each round of REFMAC running before individual atomic B factor refinement. The TLS parameters were optimized by a recent established web server TLSMD.[56] NCS restraints were defined for the N and C-terminal domains separately and kept with high weights during refinement. Apart from the protein model, the highest peak in a F_o–F_c map was identified as a sulfate ion due to close contacts from two positive charged side-chains in the vicinity. Solvent molecules accounted for other peaks above 3.0σ. All data collection and refinement statistics are summarized in Table 1.

Structural analysis

The refined model was uploaded to two web servers, ProFunc[57] and PDBsum[29] for thorough structure analysis including model inspection, motif and structural neighbor search, binding cavity prediction and other analysis. An electrostatic potential map was calculated using the program GRASP.[58] The figures representing Phr ribbon models were prepared using the molecular visualization program PyMOL.[59]

Molecular docking

The coordinate file of a canonical B-form DNA segment used in docking calculations was generated using the program 3DNA.[60] Molecular docking was performed by the protein-DNA docking program HADDOCK[39] based on biological information. All the consensus bases from both strands in the Phr recognition sequence were

Table 1. Data collection and refinement statistics

Data sets	Native-Phr	SeMet-Phr peak	SeMet-Phr inflection	SeMet-Phr remote
Resolution range (Å)	38.12–2.60	47.62–2.97	47.62–2.98	41.52–2.80
Wavelength (Å)	0.9198	0.9829	0.9840	0.9299
Reflections[a]	269,525/14,921	205,000/10,846	207,779/10,797	419,029/13,010
Completeness (%)[a]	98.8/90.4	99.8/100	99.9/100	99.8/99.7
$I/\sigma(I)$[a]	17.87/5.79	16.61/3.77	18.18/3.95	25.49/2.67
R_{symm}[a,b]	0.056/0.325	0.055/0.394	0.052/0.395	0.063/0.655
Wilson B-factor	74.0			85.0
R_{cryst}/R_{free}[c]	0.252/0.305			
Residues in model	Chain A: 1–197 Chain B: 1–198			
Water molecules	50			
B-factors[d]	53.1 (protein) 53.7 (solvent)			
Ramachandran plot (%)[e]	87.6/10.1/2.0/0.3			
r.m.s.d. bond length (Å)	0.023			
r.m.s.d. bond angle (°)	2.128			

[a] Values on both sides of solidus are from the whole data set and the highest resolution shell, respectively.
[b] $R_{symm}=\sum|I-\langle I\rangle|/\sum I$, where I is the intensity of a given reflection.
[c] $R=\sum||F_o|-|F_c||/\sum|F_o|$. R_{cryst} and R_{free} are calculated from the working and the test reflection (5%) sets, respectively.
[d] Averaged B-factors calculated before TLS refinement using REFMAC.
[e] The values separated by solidus represent the percentage of dihedral angles located in the most favorable allowed, additional allowed, generously allowed and disallowed regions, respectively.

defined as active residues in the ambiguous interaction restraints (AIR) file. AIRs for Phr were defined on those positive charged residues with side-chains exposed on the protein surface of the DBD (Figure 2(a)), including R17, K43, H48, R50, R67, R69 and R71. DNA restraints and docking protocol were all set up according to the literature,[39] with additional restraints of *C*2 symmetry applied on both the protein and the DNA in order to keep the 2-fold symmetry at the binding interface. The model with the lowest energy at the interface was chosen as the final docking model.

Mutational studies and electrophoretic mobility shift analysis

Point mutants of Phr were made at the positions N14, T16, R17, S32, Q42, K43, L46, H48, R50, E53, R67, R69 and R71. Each substituted amino acid was replaced by alanine. The mutants were expressed and purified with the same profile as the native protein. The DNA probe for EMSA was constructed by PCR using the M13/pUC reverse primer and the Cy5 labelled M13/pUC forward primer. The plasmid upatpase,[16] which contains the *aaa*$^+$ *atpase* heat shock promoter, was used as DNA template. The mixture of the EMSA reaction containing 150 ng protein and equal molar oligonucletides was incubated at 70 °C for 10 min before loading onto a non-denaturing 5% (w/v) polyacrylamide gel.

Formaldehyde crosslinking of protein-DNA complexes *in vivo*

For *in vivo* experiments, *P. furiosus* cells were grown in a 45 l fermentor at 95 °C in marine medium. The cells were grown at 95 °C to a density of 1×10^8 cells per ml of medium and then the temperature was raised to 105, 106, 107 or 108 °C. At 30 min later formaldehyde was added to the cultures with the final concentration of 0.1% (v/v). The crosslinking reaction was quenched after 20 s by the addition of glycine to 125 mM. The cells were rapidly cooled to 4 °C using a heat exchange device and harvested by centrifugation before storing at −80 °C.

Immunoprecipitation of DNA-Phr complexes *in vivo*

Formaldehyde-treated cells were disrupted by sonification and insoluble particles were removed by centrifugation. The soluble fraction containing DNA was hydrolyzed by *Micrococcus* nuclease to obtain DNA fragments in the size range of 500–1000 bp. The precleared cell extracts were incubated with 1 mM PMSF and 15 ng anti-Phr IgG or anti-TBP IgG at 4 °C for 1 h till overnight, followed by the addition of protein A Sepharose particles to purify DNA-protein complexes. The immunprecipitate was resuspended in the elution buffer (10 mM EDTA, 1% SDS, 50 mM Tris-Cl (pH 8)) and incubated at 65 °C for 10 min. To reverse the crosslinking, the precipitates and the input sample were incubated overnight at 65 °C. Protein in the sample was degraded by proteinase K and the DNA purified from the precipitates was resuspended in TE buffer. DNA concentrations were measured with a fluorometer (TD700, Turner Designs). Specific promoter sequences in the precipitated DNA were detected using semi-quantitative PCR reactions. Primers for the *aaa*$^+$ *atpase* promoter were atp_new_F (5′-TTC AAA ATC CTT GGA TCA TAA CC- 3′) and atp_new-R (5′-TGC CCC TAC CAA CAT CTC TC- 3′), primers for the *hsp20* promoter were hsp20_new-F (5′-TGG AGT ATT TTT GAT TGT TCG GTA- 3′) and hsp20_new-R (5′-TTC CCT TAT TAG GTC GAA TGG A- 3′), primers for the *phr* promoter were phr_new-F (5′-TTG TGG GAA TTG GTG GAT CT- 3′) and phr_new-R (5′-TCC TTG TTT CAT TCC CCA ATA- 3′) and primers for the *gdh* promoter were gdh_new-F (5′-TTG AAA ATG TTT GAG GAA CAC C- 3′) and gdh_new-R (5′-TTG GGC AGC TCT TTC AAG TT- 3′). PCR products were analyzed on a 1% (w/v) agarose gel after staining with ethidium bromide by phosphoimaging.

Protein Data Bank accession codes

Atomic coordinates have been deposited with the RCSB Protein Data Bank and are available under accession code 2P4W.

Acknowledgement

We thank Dr Santosh Panjikar at the EMBL, Hamburg for his kind help in data collection and phasing.

References

1. Schwartzman, D. W. & Lineweaver, C. H. (2004). The hyperthermophilic origin of life revisited. *Biochem. Soc. Trans.* **32**, 168–171.
2. Bell, S. D. & Jackson, S. P. (1998). Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol.* **6**, 222–228.
3. Laksanalamai, P., Whitehead, T. A. & Robb, F. T. (2004). Minimal protein-folding systems in hyperthermophilic archaea. *Nature Rev. Microbiol.* **2**, 315–324.
4. Bartlett, M. S. (2005). Determinants of transcription initiation by archaeal RNA polymerase. *Curr. Opin. Microbiol.* **8**, 677–684.
5. Bell, S. D. (2005). Archaeal transcriptional regulation-variation on a bacterial theme? *Trends Microbiol.* **13**, 262–265.
6. Hickey, A. J., Conway de Macario, E. & Macario, A. J. (2002). Transcription in the archaea: basal factors, regulation, and stress-gene expression. *Crit. Rev. Biochem. Mol. Biol.* **37**, 537–599.
7. Bell, S. D. & Jackson, S. P. (2001). Mechanism and regulation of transcription in archaea. *Curr. Opin. Microbiol.* **4**, 208–213.
8. Geiduschek, E. P. & Ouhammouch, M. (2005). Archaeal transcription and its regulators. *Mol. Microbiol.* **56**, 1397–1407.
9. Brinkman, A. B., Ettema, T. J., de Vos, W. M. & van der Oost, J. (2003). The Lrp family of transcriptional regulators. *Mol. Microbiol.* **48**, 287–294.
10. Thaw, P., Sedelnikova, S. E., Muranova, T., Wiese, S., Ayora, S., Alonso, J. C. *et al.* (2006). Structural insight into gene transcriptional regulation and effector binding by the Lrp/AsnC family. *Nucl. Acids Res.* **34**, 1439–1449.
11. Shockley, K. R., Ward, D. E., Chhabra, S. R., Conners, S. B., Montero, C. I. & Kelly, R. M. (2003). Heat shock response by the hyperthermophilic archaeon *Pyrococcus furiosus*. *Appl. Environ. Microbiol.* **69**, 2365–2371.

12. Rosen, R. & Ron, E. Z. (2002). Proteome analysis in the study of the bacterial heat-shock response. *Mass Spectrom Rev.* **21**, 244–265.
13. Yura, T. & Nakahigashi, K. (1999). Regulation of the heat-shock response. *Curr. Opin. Microbiol.* **2**, 153–158.
14. Morano, K. A. & Thiele, D. J. (1999). Heat shock factor function and regulation in response to cellular stress, growth, and differentiation signals. *Gene Expr.* **7**, 271–282.
15. Morimoto, R. I. (1998). Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. *Genes Dev.* **12**, 3788–3796.
16. Vierke, G., Engelmann, A., Hebbeln, C. & Thomm, M. (2003). A novel archaeal transcriptional regulator of heat shock response. *J. Biol. Chem.* **278**, 18–26.
17. Gajiwala, K. S. & Burley, S. K. (2000). Winged helix proteins. *Curr. Opin. Struct. Biol.* **10**, 110–116.
18. Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. (2005). The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.* **29**, 231–262.
19. Cook, W. J., Kar, S. R., Taylor, K. B. & Hall, L. M. (1998). Crystal structure of the cyanobacterial metallothionein repressor SmtB: a model for metalloregulatory proteins. *J. Mol. Biol.* **275**, 337–346.
20. Ye, J., Kandegedara, A., Martin, P. & Rosen, B. P. (2005). Crystal structure of the *Staphylococcus aureus* pI258 CadC Cd(II)/Pb(II)/Zn(II)-responsive repressor. *J. Bacteriol.* **187**, 4214–4221.
21. Zheng, N., Fraenkel, E., Pabo, C. O. & Pavletich, N. P. (1999). Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. *Genes Dev.* **13**, 666–674.
22. Wilce, J. A., Vivian, J. P., Hastings, A. F., Otting, G., Folmer, R. H., Duggin, I. G. *et al.* (2001). Structure of the RTP-DNA complex and the mechanism of polar replication fork arrest. *Nature Struct. Biol.* **8**, 206–210.
23. Lupas, A., Van Dyke, M. & Stock, J. (1991). Predicting coiled coils from protein sequences. *Science*, **252**, 1162–1164.
24. Moll, J. R., Ruvinov, S. B., Pastan, I. & Vinson, C. (2001). Designed heterodimerizing leucine zippers with a ranger of pIs and stabilities up to 10(−15) M. *Protein Sci.* **10**, 649–655.
25. Mason, J. M. & Arndt, K. M. (2004). Coiled coil domains: stability, specificity, and biological implications. *ChemBioChem*, **5**, 170–176.
26. Arndt, K. M., Pelletier, J. N., Muller, K. M., Pluckthun, A. & Alber, T. (2002). Comparison of in vivo selection and rational design of heterodimeric coiled coils. *Structure*, **10**, 1235–1248.
27. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L. & Sweet, R. M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature*, **362**, 219–223.
28. Gajiwala, K. S., Chen, H., Cornille, F., Roques, B. P., Reith, W., Mach, B. & Burley, S. K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature*, **403**, 916–921.
29. Laskowski, R. A., Chistyakov, V. V. & Thornton, J. M. (2005). PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucl. Acids Res.* **33**, D266–D268.
30. Holm, L. & Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480.
31. Shindyalov, I. N. & Bourne, P. E. (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11**, 739–747.
32. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallog. sect. D*, **60**, 2256–2268.
33. Busenlehner, L. S., Pennella, M. A. & Giedroc, D. P. (2003). The SmtB/ArsR family of metalloregulatory transcriptional repressors: structural insights into prokaryotic metal resistance. *FEMS Microbiol. Rev.* **27**, 131–143.
34. Briknarova, K., Takayama, S., Brive, L., Havert, M. L., Knee, D. A., Velasco, J. *et al.* (2001). Structural analysis of BAG1 cochaperone and its interactions with Hsc70 heat shock protein. *Nature Struct. Biol.* **8**, 349–352.
35. Sondermann, H., Scheufler, C., Schneider, C., Hohfeld, J., Hartl, F. U. & Moarefi, I. (2001). Structure of a Bag/Hsc70 complex: convergent functional evolution of Hsp70 nucleotide exchange factors. *Science*, **291**, 1553–1557.
36. Doong, H., Vrailas, A. & Kohn, E. C. (2002). What's in the 'BAG'?–a functional domain analysis of the BAG-family proteins. *Cancer Letters*, **188**, 25–32.
37. Gehring, U. (2004). Biological activities of HAP46/BAG-1. The HAP46/BAG-1 protein: regulator of HSP70 chaperones, DNA-binding protein and stimulator of transcription. *EMBO Rep.* **5**, 148–153.
38. Madera, M., Vogel, C., Kummerfeld, S. K., Chothia, C. & Gough, J. (2004). The SUPERFAMILY database in 2004: additions and improvements. *Nucl. Acids Res.* **32**, D235–D239.
39. van Dijk, M., van Dijk, A. D., Hsu, V., Boelens, R. & Bonvin, A. M. (2006). Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility. *Nucl. Acids Res.* **34**, 3317–3325.
40. Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y. W., Ebright, R. E. & Berman, H. M. (1996). Structure of the CAP-DNA complex at 2.5 angstroms resolution: a complete picture of the protein-DNA interface. *J. Mol. Biol.* **260**, 395–408.
41. Bell, S. D., Cairns, S. S., Robson, R. L. & Jackson, S. P. (1999). Transcriptional regulation of an archaeal operon in vivo and in vitro. *Mol. Cell*, **4**, 971–982.
42. Chen, S., Gunasekera, A., Zhang, X., Kunkel, T. A., Ebright, R. H. & Berman, H. M. (2001). Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: alteration of DNA binding specificity through alteration of DNA kinking. *J. Mol. Biol.* **314**, 75–82.
43. Littlefield, O. & Nelson, H. C. (1999). A new use for the 'wing' of the 'winged' helix-turn-helix motif in the HSF-DNA cocrystal. *Nature Struct. Biol.* **6**, 464–470.
44. Wachtershauser, G. (2000). Origin of life. Life as we don't know it. *Science*, **289**, 1307–1308.
45. Wachtershauser, G. (2002). Discussing the origin of life. *Science*, **298**, 747–749.
46. Heldwein, E. E. & Brennan, R. G. (2001). Crystal structure of the transcription activator BmrR bound to DNA and a drug. *Nature*, **409**, 378–382.
47. Hochheimer, A., Hedderich, R. & Thauer, R. K. (1999). The DNA binding protein Tfx from *Methanobacterium thermoautotrophicum*: structure, DNA binding properties and transcriptional regulation. *Mol. Microbiol.* **31**, 641–650.
48. Pirkkala, L., Nykanen, P. & Sistonen, L. (2001). Roles of the heat shock transcription factors in regulation of the heat shock response and beyond. *Faseb J.* **15**, 1118–1131.
49. Voellmy, R. (2004). On mechanisms that control heat

shock transcription factor activity in metazoan cells. *Cell Stress Chaperones*, **9**, 122–133.
50. Voellmy, R. (2006). Feedback regulation of the heat shock response. *Handb. Exp. Pharmacol.* **172**, 43–68.
51. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. vols. Academic Press, New York.
52. Terwilliger, T. (2004). SOLVE and RESOLVE: automated structure solution, density modification and model building. *J. Synchrotron Radiat.* **11**, 49–52.
53. Terwilliger, T. C. (2003). Statistical density modification using local pattern matching. *Acta Crystallog. sect. D*, **59**, 1688–1701.
54. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
55. Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374**, 300–321.
56. Painter, J. & Merritt, E. A. (2006). TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallog.* **39**, 109–111.
57. Laskowski, R. A., Watson, J. D. & Thornton, J. M. (2005). ProFunc: a server for predicting protein function from 3D structure. *Nucl. Acids Res.* **33**, W89–W93.
58. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.
59. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System,* DeLano Scientific, San Carlos, CA, USA.
60. Lu, X. J. & Olson, W. K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucl. Acids Res.* **31**, 5108–5121.

Edited by K. Morikawa

(Received 8 February 2007; received in revised form 15 March 2007; accepted 18 March 2007)
Available online 24 March 2007