A NOVEL TRANSCRIPTIONAL REGULATOR INVOLVED IN

PYROCOCCUS FURIOSUS COLD SHOCK RESPONSE

by

XIN LI

(Under the Direction of ROBERT A. SCOTT)

ABSTRACT

The transcription system of archaea has both bacterial-like and eukaryotic-like features; however, only a small number of archaeal transcriptional regulators have been characterized in detail. This work describes the discovery and characterization of archaeal transcriptional regulators of the hyperthermophile *Pyrococcus furiosus*. A DNA affinity protein capture experiment was employed to pull specific DNA-binding proteins out of *P. furiosus* cell extract with promoter DNA. These proteins are responsible for the observed gene regulation in DNA microarray expression profiles of the transcriptional response of *P. furiosus* to cold shock. A novel transcription regulatory protein PF1072 was identified and its specific DNA binding ability at sites upstream of multiple genes was confirmed by the electrophoretic mobility shift assay and fluorescence-detected DNase I footprinting. Systematic evolution of ligands by exponential enrichment (SELEX) was used to determine the consensus binding sequence, so other binding sites in the genome can be detected and verified. The work presented here demonstrates that the novel transcriptional regulator PF1072 is likely a relevant participant in transcriptional regulation pathways related to *P. furiosus* cold shock adaptation.

INDEX WORDS: Archaea, *Pyrococcus furiosus*, transcription factor, transcriptional regulation, DNA microarray, cold shock, fluorescence footprinting, SELEX, PF1072

A NOVEL TRANSCRIPTIONAL REGULATOR INVOLVED IN *PYROCOCCUS FURIOSUS* COLD SHOCK RESPONSE

by

XIN LI

B.S., University of Science and Technology of China, China, 2002

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2008

© 2008

Xin Li

All Rights Reserved

A NOVEL TRANSCRIPTIONAL REGULATOR INVOLVED IN *PYROCOCCUS FURIOSUS* COLD SHOCK RESPONSE

by

XIN LI

Major Professor: Robert A. Scott

Committee: I. Jonathan Amster
Jeffrey L. Urbauer

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2008

DEDICATION

To my grandparents and parents for everything they have given me.

ACKNOWLEDGEMENTS

Thanks to my major professor Dr. Scott for his guidance and support. Thanks to my committee members Dr. Amster, Dr. Urbauer and Dr. Eidsness who have taken the time to mentor me in my scientific development. To those who through their help have made so much of my research possible, many thanks: to Meiyao Wang and Gina Lipscomb who trained me in the lab early on, to Frank Jenney, Gerti Schut and Michael Weinberg who provided the microarray data and many valuable scientific suggestions, to Dr. Phillips of the mass-spectrometry facility who spent much time helping me run my many samples, to Darin Cowart who provided the database, to Antonia Gindner and Dr. Winfried Hausner for collaborating with me on this project and providing valuable functional data, to Guanqun Yuan for a very enjoyable collaboration, and to the many others who I have not mentioned but have helped me so much. Thanks to all my lab mates past and present and my friends on and off campus.

TABLE OF CONTENTS

Page

CHAPTER 1

TRANSCRIPTIONAL REGULATION IN ARCHAEA AND TRANSCRIPTION FACTOR DISCOVERY APPROACH

1.1 Introduction to Archaea and *Pyrococcus furiosus*

1.1.1 Introduction to Archaea

Archaea form a third domain of organisms apart from Eukarya and Bacteria because of their unique features [1, 2]. Archaea are common life forms on earth, inhabiting ocean waters, lake waters, and soil. However, they are also found in the world's least inhabitable environments: deep-sea volcanic vents, solfataric fields, boiling springs, saturated salt seas, Antarctic lakes, acid and alkaline waters, etc. [3, 4]. As a deeply rooted branch of the universal tree of life, they are the closest relatives to the original prokaryotic organisms from which all cellular life forms are thought to have evolved. A great deal of research has been devoted to learning more about these prokaryotes. It has become clear that most of the organisms that can thrive under the most diverse conditions (including extremes of temperature, pressure, salinity and pH) belong to the Archaea. The organisms of this domain are of particular interest because their information-processing machinery—translation, transcription, and replication pathways—has more similarities with eukaryotes [5], whereas, most of their metabolic and morphologic aspects are similar to those of Bacteria. In particular, the promoter elements and basal transcription machinery are analogous to those in Eukarya, although they appear to be somewhat simpler. Investigating comparable but simpler pathways in Archaea will lend insight to the more complex

eukaryotic systems which are inherently more difficult to study. A better understanding of transcriptional apparatus and regulation of transcription in Archaea will help to integrate the body of knowledge regarding the regulatory mechanisms in all three domains of life.

1.1.2 Introduction to Pyrococcus furiosus

Pyrococcus furiosus has emerged as a useful model archaeon for focused study. It is one of the hyperthermophiles which comprise a group of organisms that grow optimally at or above 80 °C, including 34 genera from archaeal and bacterial domains [6]. *P. furiosus* is a member of the family *Thermococcaceae*. The optimal growth temperature for *P. furiosus* is 100 °C. Both carbohydrates and peptides can be used as carbon sources by *P. furiosus* [7]. It was discovered by Fiala and Stetter from a hydrothermal vent community off the coast of Italy [7]. The vent fluids are expelled into oxygenated seawater at ambient temperature (20 °C), conditions under which *P. furiosus* can survive but not grow [8]. Growth resumes when the organism is subsequently exposed to temperatures above 70 °C under anaerobic conditions. It would therefore seem likely that *P. furiosus* has mechanisms to help it survive temporary exposure to both oxygen and lower temperatures. This trait presents an opportunity to study the regulation of metabolism of *P. furiosus* as a function of temperature by investigating the changes that occur at lower growth temperature in a growing culture.

1.2 Archaeal transcription

Transcription is the process by which RNA polymerase (RNAP) synthesizes RNA using DNA as a template. One transcription product, messenger RNA, supplies the templates for synthesis of proteins through translation. Gene transcription can be achieved in cell-free conditions by using the minimum necessary components called the basal transcriptional

apparatus, including several conserved promoter elements (DNA sequences), the general cognitive transcription factors (TFs) and RNAP. However, the basal transcriptional apparatus is not enough for all cellular responses to various conditions. Many regulatory proteins are involved in transcriptional regulation to either activate or repress expression of genes in response to environmental stimuli. The transcribed RNA and expressed proteins in a cell generally determine the nature of the cell. The process of transcription plays a critical role in controlling the nature and development of the organism. Therefore, it is important to investigate transcription and its regulation, the knowledge of which should facilitate the understanding of fundamental aspects of the nature of living organisms.

In eukaryotes, there are three multisubunit DNA-dependent RNA polymerase (RNAP) enzymes which catalyze three classes of transcription: RNAPI for rRNA (ribosomal RNA for the ribosome), RNAPII for mRNA (messenger RNA as the protein synthesis template), and RNAPIII for tRNA (transfer RNA for use in template codon recognition). The structure and function of RNAPII subunits has been summarized and discussed in several review articles [9-11]. Archaea have only one RNAP to fulfill the diverse transcriptional needs of the cell. The archaeal RNAP is most like eukaryal RNA Polymerase II (Pol II) which is responsible for transcription of messenger RNA. Although archaea contain only one RNA polymerase, the archaeal RNAP contains 11 to 12 subunits (B'/B", A', A", D, E', F, L, H, N, K, P), which are homologous to eukaryotic versions in both structure and function [12]. The most notable difference between archaeal and eukaryal RNA polymerase is that archaeal RNA polymerase lacks an acidic C-terminal tail found on Pol II which in eukarya plays a major role in transcriptional regulation [13].

Bacteria have only one RNA polymerase system which is simpler in composition, comprised of five subunits: αI, αII, β', β'', and ω [14]. These five bacterial RNAP subunits have paralogs in both archaeal and eukaryal RNAPs, and together, these constitute a core in the overall architecture of RNAPs, though the archaeal and eukaryal RNAP cores are more homologous to each other than each is to the bacterial RNAP [15]. Bacterial RNAP also requires a σ factor, which plays an important role for promoter recognition through binding at –10 and –35 promoter elements [16]. It also associates with the core polymerase weakly and reversibly, recruiting RNAP to the promoter to form the transcription pre-initiation complex and releasing from the core polymerase upon the start of transcription elongation [17].

The fully functional eukaryotic transcriptional apparatus is a complicated system, including the 12-subunit Pol II enzyme and five general transcription factors having 26 total subunits (TFIIB, -D, -E, -F, and -H). The archaeal system is much simpler. There are only two general transcription factors coupled with RNAP in archaea to achieve basal transcription *in vitro*: TBP (TATA-binding protein) and TFB (Transcription Factor B). TBP is the homolog of eukaryotic TBP which is part of the larger complex of TFIID. TFB is the homolog of eukaryotic TFIIB [18]. Another general transcription factor TFE (Transcription Factor E), homologous to the α-subunit of eukaryal TFIIE, is sometimes necessary to drive transcription from weak promoters [19]. The similarities and differences in RNAP subunits and general transcription factors among the three domains of life are illustrated in Figure 1.1.

The eukaryal promoters consist of the AT-rich TATA box centered at 25 bp upstream from the transcription start site and the 7-bp purine-rich BRE (TFB Recognition Element) just upstream of the TATA box [20]. TBP and TFB interact with the TATA box and BRE, respectively, for transcription initiation. The eukaryal TATA box has some sequence variation,

but in general, can be described with the consensus sequence of TATA(A/T)AA(G/A) [21]. The consensus archaeal TATA sequence is somewhat vague, with variations of the consensus sequence existing in different classes of archaea [22]. A Ribosome Binding Site (RBS) has also been identified, located approximately 6 bp upstream of the translation start site. The sequence of the RBS is not conserved among archaeal species or among genes within a single species. Despite this lack of conservation, there are often some common characteristics shared by different but related initiator elements. The same is true for the archaeal TFB recognition element (BRE) that is usually located 2 bp upstream of the TATA box. The binding of TFB to BRE is critical for determining the orientation of transcription [23]. In some eukaryotic systems, the TFIIB-TBP-promoter complex is stabilized through direct interaction between BRE and TFIIB, which enhances transcription. However, the BRE appeared to play a role in suppressing basal transcription under different conditions. Thus, the BRE sequence and function are poorly conserved and the role of the BRE and the BRE-binding factors in transcription initiation has followed different evolutionary pathways. In contrast to these eukaryal-type elements, the bacterial core promoter consists of 5 to 6 bp sequences centered at -10 and -35 relative to the transcription start site, and specificity to these elements is directed by a σ factor [24]. The similarities and differences among core promoters of the three domains of life are illustrated in Figure 1.2.

Transcription initiation in archaea is similar to that in eukarya. TBP binds to the TATA box in the minor groove and bends the DNA, allowing the cooperative binding of TFB [25]. TFB binds to TBP and the BRE, stabilizing the TBP-DNA complex and determining the direction of transcription [23]. TBP and TFB then recruit RNAP to the transcription start site. TFB and RNAP together define the transcription start site position, 25 to 30 bp downstream of the TATA

box [21]. TFE interacts with RNAP subunits E and F and possibly associates with DNA between the TATA box and transcription start site, thereby stabilizing the initiation complex [26].

1.3 Archaeal transcriptional regulation and regulatory transcription factors

Although archaea possess a basal transcription machinery resembling that of eukaryotes, their regulators are more homologous to bacterial activators and repressors [27]. To date, only a few archaeal regulatory systems have been characterized at the molecular level. Among the regulators that have been characterized in detail, most of them have been demonstrated to be transcriptional repressors [13].

MDR1 is one of the first archaeal regulators to be characterized. It is a metal-dependent regulator of *Archaeoglobus fulgidus* [28], a homolog of bacterial metal-dependent repressors DtxR and SirR. In the presence of certain metal ions, MDR1 binds to its recognition sequence which overlaps the transcription initiation site and represses transcription from its own gene by blocking recruitment of RNAP [28].

LrpA, a member of the bacterial Lrp/AsnC family from *P. furiosus*, negatively regulates its own transcription via binding to a 46-bp sequence that overlaps the transcription start site of its own promoter and preventing recruitment of RNAP [29]. Several other Lrp family putative regulators have been characterized: Sa-Lrp from *Sulfolobus acidocaldarius* [30], Ss-Lrp from *S. solfataricus* [31], Ss-LrpB from *S. solfataricus* [32, 33], and Lrs14 from *Sulfolobus solfataricus* [34]. Lrs14 represses transcription from its own gene by obstructing the binding of TBP and TFB through overlapping the BRE and TATA box.

The members of the Lrp family are also known as feast/famine regulatory proteins that have the potential to form higher-order nucleoprotein structures with DNA [35]. Bacterial

regulators of the Lrp family have been shown to respond to small-molecule ligands, such as amino acids. An archaeal Lrp-like regulator, LysM from *S. solfataricus*, was found to respond to lysine. In the absence of lysine, LysM binds to a site upstream of the BRE/TATA box in the promoter of a lysine-biosynthesis gene cluster and most likely activates transcription; in the presence of lysine, LysM has lower DNA-binding affinity, and the expression of the gene cluster was reduced [36].

Another characterized archaeal repressor was TrmB, which does not appear to have any homologs either in bacteria or eukarya. TrmB is the transcriptional repressor for the gene cluster of the trehalose/maltose ABC transporter of the hyperthermophilic archaea *Thermococcus litoralis* and *P. furiosus* (the gene is identical in both organisms), with maltose and trehalose acting as inducers [37]. TrmB also regulated the transcription of genes encoding a separate maltodextrin ABC transporter in *P. furiosus* with maltotriose, longer maltodextrins and sucrose acting as inducers, but not maltose or trehalose [38]. The recently determined structure of the C-terminal sugar-binding domain reveals a novel sugar-binding fold of Trmb [39].

Two related archaeal transcriptional regulators have also been described since the discovery of TrmB, TrmBL1 from *P. furiosus* [40] and Tgr from *Thermococcus kodakaraensis* [41]. TrmBL1 appears to be a global transcriptional regulator. It regulates the expression of the genes encoding enzymes of the glycolytic pathway as well as the maltodextrin ABC transporter [40]. These genes were previously identified as part of a putative regulon due to the presence of a common motif in their promoter regions, now known to be recognized by TrmBL1 [22]. Tgr is an ortholog of TrmBL1, recognizing the same DNA-binding motif, functioning as both an activator and repressor of transcription in the hyperthermophilic archaeon *Thermococcus kodakaraensis.* Growth of a Tgr disruption strain displayed a significant decrease in growth rate

under gluconeogenic conditions compared with the wild-type strain, whereas comparable growth rates were observed under glycolytic conditions. A whole genome microarray analysis revealed that transcript levels of almost all genes related to glycolysis and maltodextrin metabolism were at relatively high levels with the Tgr disruption strain even under gluconeogenic conditions. The Tgr disruption strain also displayed defects in the transcriptional activation of gluconeogenic genes under these conditions [41]. This is the first instance that an archaeal transcriptional regulator has been shown indirectly to exert both positive and negative control in transcription, although mechanistic details are lacking.

NrpR, from *Methanococcus maripaludis*, is another archaeal-specific transcriptional regulator. The euryarchaeal nitrogen repressor NrpR controls the expression of the nitrogen fixation (*nif*) operon, resulting in full repression with ammonia, intermediate repression with alanine, and derepression with dinitrogen. NrpR binds to two tandem operators in the *nif* promoter region, $nifOR_1$ and $nifOR_2$ [42, 43]. NrpR homologues are widely distributed in Euryarchaeota and present in a few bacterial species. They exist in three different domain configurations and phylogenetic analysis indicated that the NrpR family forms five distinct groups [44].

Other transcription factors have been identified through genetic and physiological approaches. The haloarchaeal repressor GvpD has been studied in some detail *in vivo* using the *Haloferax volcanii* genetic system. GvpD is the negative partner of a positive–negative regulator pair controlling gas vesicle production in this extreme halophile. The positive-regulator partner, GvpE, a homolog of the large family of leucine-zipper (bzip) eukaryotic transcription regulators [45], and GvpD interact directly. Thus, GvpD may be an anti-activator [46, 47]. Although the relationship between GvpD and GvpE is reminiscent of bacterial σ and anti-σ pairs [24], GvpE is

the first archaeal transcriptional regulator to show homology to a class of eukaryotic transcription factors.

Phr, a regulator of heat-shock response from *P. furiosus*, appears to have domain homologies to both bacteria and eukarya. This protein specifically inhibited cell free transcription of its own gene and from promoters of a small heat shock protein, Hsp20, and of an ATPase. Inhibition of transcription was brought about by abrogating RNA polymerase recruitment to the BRE and TATA boxes [48]. The recently solved crystal structure revealed a stable homodimer, each subunit consisting of an N-terminal winged helix DNA binding domain and a C-terminal antiparallel coiled coil helical domain. The overall structure appears as a molecular chimera with significant folding similarity of its DNA binding domain to the bacterial SmtB/ArsR family, while its C-terminal part was found to be a remote homolog of the eukaryotic BAG domain which in eukaryotic proteins is utilized to associate with certain heat shock proteins [49].

An Lrp family protein Ptr2, has been proposed as positive regulator of transcription. Ptr2 is one of the two Lrp family proteins of *M. jannaschii* and has been shown to activate transcription *in vitro* at a small number of transcription units encoding proteins that participate in electron transfer processes. Ptr2 stimulates transcription of the *rb2* and *fdxA* genes by facilitating the recruitment of TBP to a weak TATA box [50-52]. This is the first direct demonstration of positive regulation of eukaryal-type archaeal RNAPs by a bacterial-type DNA-binding protein. Recent experiments demonstrate the retention of activator function in an entirely recombinant transcription system that utilizes *M. jannaschii* RNAP assembled *in vitro* from its recombinant subunits. Close homologs to Ptr2 have also been analyzed in *Methanothermococcus thermolithotrophicus* (*Mth*) and *Methanococcus maripaludis* (*Mma*). The homolog from *Mth*

activates transcription in conjunction with its cognate TATA box binding proteins (TBPs) and with heterologous TBPs. The homolog from *Mma* is nearly inert as a transcriptional activator, but a cluster of mutations that converts a surface patch of this homolog to identity with Ptr2 confers transcriptional activity [53].

Another activator, Sta1, was isolated from *Sulfolobus islandicus*. Its activating effect on transcription initiation was demonstrated in *in vitro* transcription experiments. Most activation was observed at low concentrations of either of the two transcription factors, TBP or TFB. Sta1 was able to bind promoters independently of any component of the pre-initiation complex [54].

There is now an increasing body of data available on the DNA binding properties of several putative transcriptional regulators in the archaeal domain of life. Repressors and activators prevent or facilitate recruitment of basal machinery to the core elements, respectively. However, little is known about the cofactor requirements, *in vivo* mechanisms, and targets of many of these regulators. The answers to these questions require advanced investigation and further exploration.

Table 1.1 Summary of functionally characterized archaeal transcriptional regulators

Name	Species	Type	Ligand(s)	Mode of action	Reference(s)
MDR1	*A. fulgidus*	Repressor	Fe^{2+}, Mn^{2+}, Ni^{2+}	Blocks RNAP recruitment	[28]
LrpA	*P. furiosus*	Repressor	Unknown	Blocks RNAP recruitment	[29]
Lrs14	*S. solfataricus*	Repressor	Unkown	Blocks TBP/TFB binding	[34]
LysM	*S. solfataricus*	Activator	Lysine	Unknown	[36]
TrmB	*T. litoralis, P. furiosus*	Repressor	Maltose, trehalose, sucrose, maltodextrins	Probably blocks TBP/TFB binding	[37, 38]
TrmBL1	*P. furiosus*	Repressor	Unknown	Probably blocks TBP/TFB binding	[40]
NrpR	*M .maripaludis*	Repressor	2-oxoglutarate	Probably blocks RNAP recruitment	[42, 43]

GvpE	*H. salinarium, H. mediterranei*	Activator	Unknown	Unknown	[46]
Phr	*P. furiosus*	Repressor	Unknown	Blocks RNAP recruitment	[48, 49]
Ptr2	*M. jannaschii, M. thermolithotrophicus*	Activator	Unknown	recruits TBP	[50, 53]
Sta1	*S. islandicus*	Activator	Unknown	Unknown	[54]

1.4 Introduction to cold shock

Archaea are capable of growth from 113 °C [55] to below 0 °C [56], and they are numerically abundant in diverse low temperature environments throughout the globe. In addition, non-low-temperature-adapted Archaea are commonly exposed to sudden decreases in temperature, as are other microorganisms, animals, and plants. However, little is known about how members of the archaeal domain respond to suboptimal growth temperatures, particularly in comparison to what is known about archaeal thermophiles and hyperthermophiles and responses to heat shock. The study of the cold shock response and low-temperature adaptation in Archaea will have broad implications to the understanding of cellular biology and evolution.

How bacteria respond to suboptimal growth temperatures has been well characterized [57]. A decrease in temperature results in a temporary halt in protein synthesis. However, a number of proteins are found to be induced under cold shock conditions [58, 59]. These are the so-called cold shock proteins (Csp). Expression of the cold shock proteins reaches a maximum level during a phase of acclimation. After this, synthesis of cold shock proteins declines and a new steady state level of protein expression is established which is lower than before the cold shock. The first major cold shock protein to be characterized was CspA of *Escherichia coli* [60]. Eight more proteins named from CspB to CspI that are homologous to CspA are identified, but only four (CspA, -B, -G, and -I) are cold inducible [61]. The function of these proteins appears to be to prevent the formation of inhibitory mRNA secondary structures, which are stabilized at the

lower temperatures. Other cold-inducible proteins in *E. coli* include initiation factor 2 (IF2) [62], ribosomal binding factor A (RbfA) [63], and DEAD-box RNA helicase [64]. These proteins are thought to associate with the ribosome and play a role in protein synthesis.

On the other hand, little is known about how members of the archaeal domain respond to suboptimal growth temperatures. A cold-inducible DEAD-box RNA helicase was identified from *Methanococcoides burtonii* [65]. This is the only evidence for a link between the archaeal and bacterial cold shock responses. Homologs of the cold shock proteins CspA and RbfA are not found in mesophilic and thermophilic members of the archaea. Therefore, archaea must possess a distinct cold shock response, different from the cold shock response of bacteria. In the archaeal domain, organisms with an optimal growth temperature of at least 80 °C are defined as hyperthermophiles [66]. The cold shock responses of these organisms are of particular interest. A prolyl isomerase from *Thermococcus sp.* was proven to be cold induced when the temperature of the culture was shifted from its optimal growth temperature of 85 °C to 60 °C [67]. To date, this is the only hyperthermophilic cold shock response protein that has been studied.

Recently, a genome-wide approach to studying the response of hyperthermophilic archaea to suboptimal growth temperatures was reported [68]. DNA microarrays were used to measure transcription profiles in response to cold stress of all 2,065 Open Reading Frames (ORFs) of *Pyrococcus furiosus*, whose optimal growth temperature is around 100 °C. The results show that the organism has three very different cellular responses to the suboptimal growth temperature of 72 °C. The high-throughput approaches made possible by DNA microarray technology coupled with the availability of whole-genome sequences have provided the ability to monitor the changes in gene expression for *Pyrococcus furiosus* as a result of a response to cold shock in its environment.

1.5 Method of transcription factor discovery and characterization

With the increasing amount of microarray expression profiling being done with *P. furiosus*, the stage is being set for mapping of transcriptional regulatory networks. For prokaryotes, much regulation of gene expression occurs at the level of transcription, and therefore, gene expression profiles which demonstrate changes in transcript levels between two growth conditions can be utilized in targeted transcription factor discovery. A method of utilizing the expression data for targeted discovery of transcription factors by pull-down from cell extract can be performed with *P. furiosus*. The discovered putative transcription factor is then overexpressed, and promoter DNA is used in an electromobility shift assay (EMSA) with the transcription factor to verify sequence-specific DNA binding. Once the transcription factor is verified to be specific to a particular gene promoter region, the protein is then further characterized with DNase I footprinting and SELEX (Systematic Evolution of Ligands by Exponential Enrichment). Furthermore, the transcription factor is characterized functionally by utilizing *in vitro* transcription. This work details the identification and characterization of a regulatory transcription factor, PF1072 protein, involved in *P. furiosus* low temperature response.

Figure 1.1 Basal transcriptional machinery of the three domains.

RNA polymerase subunits (top) and general transcription factors (bottom) are color-coded according to their homologies across domains and protein/subunit names are indicated. Only those transcription factors which are necessary for *in vitro* basal transcription are shown. Figure adapted from [15].

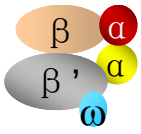
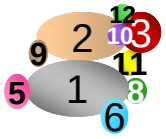
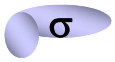
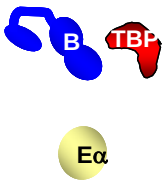
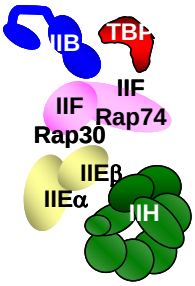
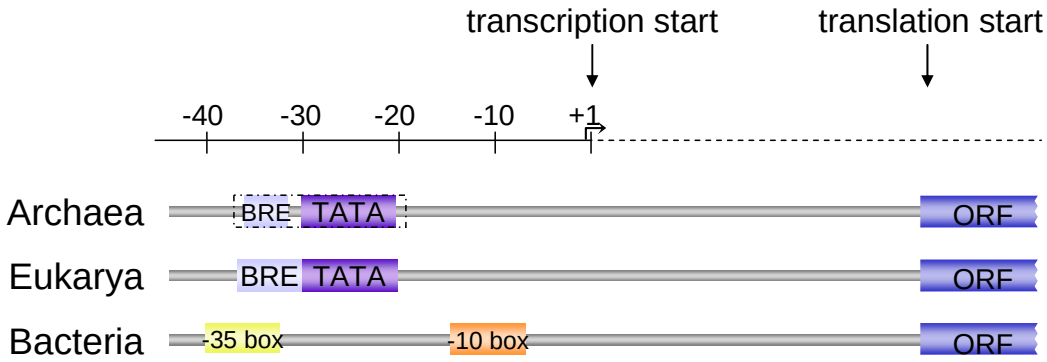
	Bacterial	Archaeal	Eucaryal class II
RNA polymerase			
Transcription factors			

Figure 1.2 Core promoter elements of the three domains.

The position of each element relative to the transcription start site is indicated by the scale at the top of the figure. Sequence elements are shown with boxes and are indicated by their abbreviations: BRE, TFB Recognition Element; TATA Box, TBP Binding Element; -10 and -35 boxes, bacterial promoter elements. Archaeal promoter elements are boxed with dotted lines to indicate minor differences in consensus sequence as compared to the related elements in eukarya, which are shaded with the same colors. Figure adapted from [69].


transcription start
translation start
-40
-30
-20
-10
+1
Archaea
BRE
TATA
ORF
Eukarya
BRE
TATA
ORF
Bacteria
-35 box
-10 box
ORF

CHAPTER 2

MATERIALS AND METHODS

2.1 *P. furiosus* culture growth and processing of soluble cell extract

P. furiosus (DSM 3638) was grown at 95 °C with maltose as the primary carbon source in a 20-L custom fermentor. The medium was the same as previously described [68]. After approximately 5 h of growth, cultures growing at 95 °C were rapidly cooled to 72 °C when they reached a cell density of ~3×10^7 cells/mL by pumping them through a glass cooling coil maintained at 20 °C. The complete culture (15 L) was cooled from 95 °C to 72 °C within 15 min. Cells were harvested before, at 1 h, and at 5 h after the cold shock. Cell cultures were cooled by pumping the culture through a coiled tube in ice water, concentrated by ultrafiltration to 2 L, and harvested by centrifugation at 10,000 × g for 15 min.

The cell pellets were resuspended in 10 mM EPPS buffer (pH 8.0) and lysed by sonication for 30 min on ice. DNase I was not added to prevent interfering with the downstream application of the cell extract in the DNA affinity protein capture experiment (Section 2.3). Lysate was centrifuged for 15 min at 10,000 × g to remove cell debris. Soluble cell extract was then obtained after centrifugation at 100,000 × g for at least 1 h to remove remaining insoluble and membrane materials. The concentration of protein in cell extract was measured with DC Protein Assay Kit II (BioRad, Hercules, CA). Soluble cell extract was saved as 20-mL aliquots in anaerobic vials and stored at -80 °C.

2.2 Primer design and PCR amplification of DNA probes

DNA oligonucleotides were purchased from IDT (Integrated DNA Technology，Coralville, IA). Primers were designed to cover a ~300 bp region (from -200 to +100) with respect to translation start site.

dsDNA was amplified from *P. furiosus* genomic DNA using PCR with *Taq* DNA polymerase (Sigma, St. Louis, MO). The PCR reaction solution included 200 μM dNTP mixture, 0.5 μM of each primer, 10 ng genomic DNA as template, 1x PCR buffer (Sigma, St. Louis, MO), and 2.5 units *Taq* DNA polymerase (Sigma, St. Louis, MO) per 50 μL reaction solution. The reaction solution was heated to 94 °C for 3 min, followed by 30 cycles of 1 min at 94 °C (denaturation), 2 min at 55 °C (annealing), and 3 min at 72 °C (extension) per cycle, and ended with final extension for 10 min at 72 °C.

PCR products were purified by agarose gel extract using QIAquick Gel Extraction Kit (Qiagen, Valencia, CA). Purified DNA was quantified using a Hoefer DyNA Quant 200 fluorometer (Amersham Pharmacia Biotech, Uppsala, Sweden) using standards.

2.3 DNA affinity protein capture

The biotinylated probe dsDNA was bound to magnetic DynaBeads M-280 Streptavidin (Invitrogen, Carlsbad, CA) in Binding&Washing buffer (0.5 mM EDTA, 1.0 M NaCl, 5 mM Tris, adjust the pH to ~7.0 using HCl) and incubated for 15 min on a vortex mixer (Fisher, Pittsburgh, PA). The bead-bound DNA was then mixed with 2.5 mg/mL *P. furiosus* soluble cell extract from cells grown either before or after cold shock and incubated at 55 °C for 30 min with intermittent mixing to keep the beads in suspension. Then bead-bound DNA-protein complexes were subjected to three quick washes at room temperature with 1x Incubation buffer (50 mM EPPs,

100 mM KCl, 1 mM EDTA, 5% glycerol, 0.1% triton-X, 1 mM DTT, pH 7.5). Bead-bound DNA-protein complexes were washed in 0.1 M sodium heparin solution for 15 min. Then DNA-bound proteins were eluted at 55 °C for 5 min with 1x Laemmli buffer. Eluted proteins were analyzed by SDS-PAGE gel (BioRad, Hercules, CA). The gel was run at 200 V for 60 min and stained with silver staining [70].

Table 2.1 Probes used in DNA affinity protein capture

Probe name	Genome coordinates	Forward primer[a] (5' biotinylated)	Reverse primer[a]	Probe length (bp)
PF0324	338566-338880	CGGTTTCTTTTCTGTTCCTT	GCTTCTGCTTGCGTTGTAGT	315
PF1072	1022704-1022972	GTCTTAGAGAATTCGCCCAGAAG	GGAGAGCATTACTTCATCTACAGG	269

[a]DNA primers are listed from 5' to 3'.

2.4 In-gel tryptic digestion and mass spectrometry identification

Bands of interest in SDS-PAGE lanes of eluted proteins from DNA affinity protein capture were excised and subjected to in-gel tryptic digestion. The gel bands were cut into small particles (ca. 1 × 1 mm) and hydrated with 50 μL 25 mM NH_4HCO_3 for 10 min at room temperature. The gel spots were then dehydrated in 80 μL 50% acetonitrile in 25 mM NH_4HCO_3 for 15 min. The hydration and dehydration steps were repeated two more times. The dehydrated gel slices were then completely dried by vacuum centrifugation for about 10 min, after which they were rehydrated in a 10-μL solution of 10 ng/μL trypsin in 25 mM ammonium bicarbonate for overnight digestion at 37 °C. The tryptic peptides were extracted from the gel plugs by washing once with 15 μL 25 mM NH_4HCO_3 for 15 min, and twice with 15 μL 75% acetonitrile, 0.5% trifluoroacetic acid (TFA) in water for 15 min each. The supernatant from each wash was

combined and concentrated to 5 μL by vacuum centrifugation for 40 min. Finally, 1 μL 5% TFA in H_2O was added to make final TFA concentration between 0.1-1%.

The tryptic peptides were desalted and concentrated with a ZipTipμ-C18 (Millipore) following the manufactures instructions. The peptide mixture was eluted from the Ziptip with 2 μL freshly made saturated matrix solution (α-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 0.1% TFA), and spotted directly onto a MALDI plate. The sample was analyzed by the Chemical and Biological Sciences Mass Spectrometry Facility (University of Georgia, Athens, GA) on a Bruker Autoflex (TOF) mass spectrometer (Bruker Daltonics Inc., Billerica, MA). MALDI-TOF MS data were searched against the NCBI genome database of archaeal species using the MS-Fit program (http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm).

2.5 Sequence analysis for selection of target protein for characterization

The protein sequences of identified proteins were analyzed using NCBI BLAST [71, 72] and the conserved domain searches [73] against the NCBI non-redundant protein database (www.ncbi.nlm.nih.gov).

2.6 Expression and purification of recombinant his-tagged protein

The ORF encoding PF1072 was amplified from *P. furiosus* genomic DNA with PCR, by Platinum PfxDNA Polymerase (Invitrogen, Carlsbad, CA), using primers that included NotI and EcoRI restriction sites for cloning into the pET24dBAM-TEV vector. This vector, adapted to include a TEV protease site between the N-terminal his-tag and the insert site was a kind gift from Francis Jenney (from the laboratory of Michael Adams, University of Georgia). The PCR product was purified by agarose gel extract with QIAquick Gel Extraction Kit (Qiagen, Valencia,

CA). The PCR product and vector were digested using NotI and EcoRI, and the 5' ends of the vector were dephosphorylated using shrimp alkaline phosphatase. Restriction enzymes were inactivated following the digestion reaction by incubation at 65 °C for 20 min.

The ligation reactions were set up in 20-µL volumes using 200 ng of linearized, gel-purified pET24dBAM-TEV vector with various molar ratios of vector to insert (1:3, 1:6) together with buffer and 1 Weiss units of T4 DNA Ligase (Promega, Madison, WI). Ligation reactions were incubated at room temperature for 3 h and immediately transformed by heat-shock into CaCl2-competent XL1-Blue cells (Stratagene, La Jolla, CA). A 3-mL culture of the cells was allowed to grow for 16 h at 37 °C with shaking at 225 rpm in LB media containing 50 µg/mL kanamycin and plasmid was extracted using a Qiagen Qiaquick Plasmid Miniprep kit (Qiagen, Valencia, CA). The insert was verified first by digestion with several restriction enzymes (NotI, BamHI, EcoRI, AvaI), each of which yielded a unique digestion pattern easily identified on a DNA gel. The plasmid was then sent to Integrated Biotech Laboratories (University of Georgia, Athens, GA) where DNA sequencing in both 5' and 3' directions verified the success of the ligation.

The plasmid was then transformed into BL21 RIPL Codon Plus (DE3) *E. Coli* cells (Stratagene, La Jolla, CA, La Jolla, CA). These cells were plated and allowed to grow overnight, then transferred and allowed to grow in 3 mL, 50 mL, and finally 1 L of 2.5% LB media containing 50 µg/mL kanamycin. IPTG was added to a final concentration of 0.5 mM at OD 0.6. The cells continued to grow at 37 °C for 4 more h and were harvested after centrifugation at 5000 rpm at 4 °C for 20 min with a Beckman JS-5.3 rotor. The cell pellet was suspended in 30 mL Hitrap column binding buffer (0.02 M sodium phosphate, 0.5 M NaCl, pH 7.4) containing 0.1 mM PMSF. Cell lysis was performed by sonicating at 40% power intensity for 10 s in every

20 s for 3 min. The cell extract supernatant was collected after centrifugation at 21,000 rpm for 60 min with a Beckman JA-25.5 rotor. The supernatant was centrifuged at 21,000 rpm for an additional 15 min prior to purification of the protein by column chromatography.

Using an automated AKTA prime FPLC system (GE Healthcare, Pittsburgh, PA), the soluble cell extract was loaded onto a 1-mL HisTrap FF metal affinity column (GE Healthcare, Pittsburgh, PA) preloaded with nickel sulfate. Once applied to the nickel column, the sample was washed with 5 mL binding buffer and the His6-tagged protein was eluted with a linear gradient of buffer containing 20-500 mM imidazole. The fraction containing the protein of interest was collected and concentrated to 500 μL using Amicon (MWCO: 10,000) centrifugal membrane filter unit (Amicon, Houston, TX) and then run through a HiTrap 5-mL desalting column using an exchange buffer (20 mM Tris, 100 mM sodium chloride, pH 7.6). The purified protein was confirmed by SDS PAGE and MS analyses. Protein concentration was determined using a DC Protein Assay kit (BioRad, Hercules, CA), and 2-mL aliquots of his-tagged protein were stored at -80 °C.

2.7 Electromobility shift assay

Electromobility shift assay (EMSA) [74] was used to investigate protein–DNA interactions between proteins and various promoter-ORFs. DNA probes for EMSA were PCR-amplified from *P. furiosus* genomic DNA using primers listed in Table 2.2 or from pUC18-cloned promoter-ORF DNA using M13 Forward and Reverse sequencing primers (Section 2.8), followed by either PCR-purification using a PCR Purification Kit (Qiagen, Valencia, CA), or gel-purification using a Qiaquick Gel Purification Kit (Qiagen, Valencia, CA). DNA concentration was typically 20 nM in each reaction, and protein was adjusted according to

the molar amount of DNA. The protein and DNA were mixed in 15 μL EMSA reaction buffer (50 mM Tris-HCl, pH 7.5; 1 mM EDTA; 0.1 M KCl; 5% glycerol).

The mixture was incubated at 55 °C for 20 min, after which the 15-μL products of reactions were loaded onto a 5% TBE gel (BioRad, Hercules, CA) and run at room temperature under constant voltage of 200 V for 30 min. The TBE gel was then stained in 50 mL 0.01% SYBR Green Nucleic Acid Gel Stain (Invitrogen, Carlsbad, CA) for 30 min and washed with water before visualization of DNA. Gels were imaged via UV transillumination.

Table 2.2 Probes used in EMSA

Probe name	Genome coordinates	Forward primer[a]	Reverse primer[a]	Probe length (bp)
PF0190	200824-201180	CCATTACTAACTTGCTTACACGG	GCACCCATAAATAGAGAACCAGT	357
PF0324	338566-338880	CGGTTTCTTTTCTGTTCCTT	GCTTCTGCTTGCGTTGTAGT	315
PF1072	1022704-1022972	GTCTTAGAGAATTCGCCCAGAAG	GGAGAGCATTACTTCATCTACAGG	269
PF1072 ORF	1022978-1023177	AGCTAGATTAAGGAAGAAGATAGA	TCGTGAGAATCTTCAAATGGTGG	200
PF1076	1025870-1026136	TTGATAGTGGGAGCCTATGTTGGG	ACCTTGCTTCTGCATAATCGGCAC	267
PF2051	1892585-1892936	AGAGGAACATCGCTATATTCTCAT	AACCTACTTATCTCCTTGGAGGGG	352

[a]DNA primers are listed from 5' to 3'.

2.8 Cloning of *P. furiosus* promoter-ORF DNA

Promoter-ORF DNA fragments were cloned from *P. furiosus* genomic DNA into pUC18 for versatile PCR-amplification of probes for EMSA, footprinting and in vitro transcription experiments. DNA fragments were designed to cover a ~300 bp region (from -200 to +100) with respect to translation start site, since at least ~100 bp of the ORF was required for an adequate

transcript in an *in vitro* transcription assay and a total length of ~300 bp DNA was required to assure that the promoter region occurs within the probe for footprinting.

Sequences of *P. furiosus* genomic DNAs to be cloned were obtained from NCBI. Primers were designed to be approximately 30 bp including about 20 bases of DNA complementary to the region to be cloned and the respective restriction sites at the 5' ends with 4 bases overhang to improve the efficiency of restriction digestion of the PCR product. Restriction maps of the genomic DNA sequences to be cloned were obtained using online software available from the New England BioLabs NEBcutter (tools.neb.com) to determine appropriate restriction sites.

pUC18 vector is a small, high copy number, *E.coli* plasmid with the reading frame of the *lacZ* gene encoding β-galactosidase which can cleave the synthetic substrate analogue X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside). The multiple cloning site for pUC18 lies within the *lacZ* gene. In the presence of IPTG, bacteria synthesize both fragments of the enzyme and form blue colonies on media with X-gal. Insertion of DNA into the multiple cloning sites located within the *lacZ* gene inactivates the N-terminal fragment of β-galactosidase. Bacteria carrying recombinant plasmids therefore give rise to white colonies. The selection process for clones that contain inserts consists of blue/white color screening of colonies grown on plates containing X-gal, along with IPTG [75].

DNA for cloning was PCR-amplified from *P. furiosus* genomic DNA using Platinum Pfx DNA Polymerase (Invitrogen, Carlsbad, CA) and purified by agarose gel extract with QIAquick Gel Extraction Kit, then digested with the appropriate restriction enzymes for 3 h at 37 °C. pUC18 plasmid DNA was amplified and purified from XL1-Blue cells (Stratagene, La Jolla, CA). The vector was digested using the appropriate restriction enzymes (BamHI/EcoRI or HindIII/BamHI), and the 5' ends of the vector were dephosphorylated using shrimp alkaline

phosphatase. Restriction enzymes were inactivated following the digestion reaction by incubation at 65 °C for 20 min. The digestion reaction was then run on a 1% agarose gel, bands of linearized vector were excised, and DNA was purified using QIAquick Gel Extraction Kit (Qiagen, Valencia, CA).

The ligation reactions were set up in 20-µL volumes using 200 ng of linearized, gel-purified pUC18 vector with various molar ratios of vector to insert (1:3, 1:6) together with buffer and 1 Weiss units of T4 DNA Ligase (Promega, Madison, WI). Ligation reactions were incubated at room temperature for 3 h. Then the ligase was deactivated by incubation at 65 °C for 20 min and immediately transformed by heat-shock into CaCl2-competent XL1-Blue cells (Stratagene, La Jolla, CA). Transformation cultures were plated onto LB-agar plates containing 100 µg/mL ampicillin, 0.5 mM IPTG, and 80 µg/mL X-gal. Plates were incubated at 37 °C for 18 h followed by chilling at 4 °C to intensify the blue color for blue/white colony screening.

White colonies were tested for the presence of the insert. A 3-mL culture of the cells was allowed to grow for 16 h at 37 °C with shaking at 225 rpm in LB media containing 100 µg/mL ampicillin, and plasmid was extracted using a Qiaquick Plasmid Miniprep kit (Qiagen, Valencia, CA). The insert was verified first by digestion with several restriction enzymes (HindIII, BamHI, EcoRI), each of which yielded a unique digestion pattern. Restricted digestion reactions were then run on 1% agarose gels to check the insert size. The plasmid was sent to Integrated Biotechnology Laboratories (University of Georgia, Athens, GA) where DNA sequencing in both 5' and 3' directions verified the success of the ligation.

Table 2.3 Promoter-ORF probe DNA for cloning into pUC18

Probe name	Genome coordinates	Forward primer[a]	Reverse primer[a]	Length (bp)	Restriction site pair
PF0190	200824-201180	AAAAGGATCCCC ATTACTAACTTGC TTACACGG	AAAAGAATTCGC ACCCATAAATAGA GAACCAT	377	BamHI/EcoRI
PF0324	338566-338880	AAAAGGATCCCG GTTTCTTTTCTGTT CCTT	AAAAGAATTCGCT TCTGCTTGCGTTG TAGT	335	BamHI/EcoRI
PF0347	360221-360513	AAAAGGATCCCCC AGTTGCCATACAT AGAG	AAAAGAATTCGC GTATCCTAAAACT GTGTC	313	BamHI/EcoRI
PF0666	676257-676573	AAAAGGATCCAG ACTTTAGAGTGAG CCTCC	AAAAGAATTCTTC TCCAAGTCTAACA GCCT	337	BamHI/EcoRI
PF0897.1n	870199-870465	AAAAGGATCCCTG CTACAGCGACATA CTAC	AAAAGAATTCGCT AAGGGTATTGCCG AAAC	287	BamHI/EcoRI
PF0947	911292-911614	AAAAGGATCCGA GAACCCTGTCTAG TGCTT	AAAAGAATTCCCA TACTATGAGCCCA GGAGC	343	BamHI/EcoRI
PF1072	1022704-1022972	AAAAAAGCTTGTC TTAGAGAATTCGC CCAGAAG	AAAAGGATCCGG AGAGCATTACTTC ATCTACAGG	289	HindIII/BamH I
PF2036	1879085-1879351	AAAAGGATCCTCA AGCCATTTATCCA ACGG	AAAAGAATTCGA CCGTTATTCTAGG TTCTTCC	287	BamHI/EcoRI

[a]DNA primers are listed from 5' to 3'.

2.9 Fluorescence-detected DNase I footprinting

DNase I footprinting was performed using capillary electrophoresis (CE) [76] based on traditional method [77] (Fig. 2.1). Footprinting probes were PCR-amplified from pUC18-cloned promoter-ORF DNA (Section 2.8) using 5' 6FAM- and HEX- labeled modified M13 primers (6FAM/TTGTAAAACGACGGCCAGT and HEX/CAGGAAACAGCTATGACCATG) for analysis on a 3730x1 automated DNA sequencer (Applied Biosystems, Foster City, CA).

Probes were amplified using *Taq* DNA polymerase (Sigma, St. Louis, MO). The PCR reaction solution included 200 μM dNTP mixture, 0.5 μM of each primer, 50 ng pUC18-cloned promoter-ORF DNA (Section 2.8) as template, 1x PCR buffer (Sigma, St. Louis, MO), and 2.5

units *Taq* DNA polymerase (Sigma, St. Louis, MO) per 50 µL reaction solution. The reaction solution was heated to 94 °C for 3 min, followed by 25 cycles of 1 min at 94 °C (denaturation), 2 min at 55 °C (annealing), and 3 min at 72 °C (extension) per cycle, and ended with final extension of 10 min at 72 °C. PCR products were then concentrated by ethanol precipitation prior to gel-purification. Probes were loaded onto a 5% TBE gel (BioRad, Hercules, CA) and were run at room temperature under constant voltage of 200 V for 30 min. Gels were stained with SYBR Green I Nucleic Acid Gel Stain (Lonza, Allendale, NJ) for 10 min prior to visualization under long-wavelength UV light. Bands were carefully excised. Probe DNA was eluted from the polyacrylamide gel slices using the crush-and-soak method [78]. Resulting probe DNA was concentrated by ethanol precipitation and quantified using a Hoefer DyNA Quant 200 fluorometer (Amersham Pharmacia Biotech, Uppsala, Sweden) prior to use in footprinting reactions. 250 ng DNA probe with various molar ratios of protein were mixed in 50 µL footprinting reaction buffer (20 mM HEPES, 100 mM KCl, 15 mM $MgCl_2$, 5 mM CaCl2, 1 mM EDTA, 1 mM DTT, 5% glycerol, pH 8.0). The mixture was incubated at 55 °C for 20 min. 0.05 U of DNase I (from a 0.01 U/µL dilution in 10 mM Tris, pH 8.0) was added to each 50 µL protein-DNA mixture and the solution was incubated for 1 min at room temperature. The reaction was stopped by adding 145 µL of Stop Solution (130 mM NaCl, 20 mM EDTA, 0.6% SDS), followed immediately by 200 µL of buffered phenol:chloroform:isoamyl alcohol (25:24:1) with vigorous shaking on a vortex mixer (Fisher, Pittsburgh, PA) followed by centrifuging at maximum speed in a microcentrifuge for 5 min. Then 180 µL of the aqueous phase was removed from the phenol:chloroform extraction and ethanol precipitated with 18 µL 3 M sodium acetate (pH 5.2), 1 µL glycogen (20 mg/mL, Roche), and 500 µL 100% ethanol. Samples were stored at -20 °C in precipitation solution until preparation and assembly of all accumulated samples into a

96-well reaction plate (Applied Biosystems, Foster City, CA) for sample submission. Precipitated DNA samples were resuspended in 10 μL of HiDi deionized formamide (Applied Biosystems, Foster City, CA) premixed with GS-500 ROX internal size standard (Applied Biosystems, Foster City, CA) (0.2 μL per sample) and analyzed on a 3730x1 automated DNA sequencer (Applied Biosystems, Foster City, CA) at the Sequencing and Synthesis Facility (University of Georgia, Athens, GA). Raw peak data were extracted from the ABI result files (.fsa file extension) using the BatchExtract program available from NCBI. Electropherograms from the raw peak data were viewed and analyzed using the graphing and analysis software IGOR Pro (Wavemetrics, Portland, OR).

2.10 Artificial selection of consensus DNA recognition sequence

A modification of the artificial selection method termed SELEX (Systematic Evolution of Ligands by Exponential Enrichment) [79, 80] was applied to determine the consensus DNA recognition sequence of the protein (Fig. 2.2). An artificial library of DNA containing random sequences was created for elucidation of the DNA-binding site through successive cycles of selection with protein. The single-stranded SELEX probe was designed with a 30 bp randomized region in the middle and three restriction sites at the end (XbaI, EcoRI, HindIII on the 5' side and BamHI, EcoRI, SalI on the 3' side). Then the synthetic single-stranded oligonucleotide was used as a template to PCR-amplify a double-stranded SELEX probe. Two primers were designed to match the 5' and 3' ends of the template, and a 5 or 6 bp overhang was added to the 5' end of the primers to eliminate the 3' self-complementary region at the SalI site which caused unwanted PCR products using primers that exactly matched the SELEX probe priming sites. The SELEX probe and primers are listed in Table 2.4. For the PCR amplification, 100 pmol of

single-stranded SELEX probe was amplified with 2 nmol of each primer for a total of 5 PCR cycles. The PCR-amplified double-stranded SELEX probe was polyacrylamide gel-purified according to the crush-and-soak method [78]. Selection rounds were set up as for the EMSA reactions (Section 2.7). The SELEX probe was present at 0.5 μM in each reaction, and protein was adjusted according to the molar amount of SELEX probe from 0.5 to 2 μM. After each selection round, shifted protein-DNA complexes bands were excised and extracted using crush-and-soak method [78], then amplified with the SELEX primers using 15 cycles of PCR, and polyacrylamide gel-purified before proceeding to the next selection round. After 8 selection rounds were performed, the selected DNA was digested with EcoRI, concatemerized by ligation with each other, and cloned into the pUC18 standard cloning vector. The ligation reaction was then transformed by heat-shock into $CaCl_2$-competent XL1-Blue cells (Stratagene, La Jolla, CA). Transformation cultures were plated onto LB-agar plates containing 100 μg/mL ampicillin, 0.5 mM IPTG, and 80 μg/mL X-gal. Plates were incubated at 37 °C for 18 h followed by chilling at 4 °C to intensify the blue color for blue/white colony screening. Plasmid with the concatemers was isolated from these colonies for sequencing. A total of 20 sequences were obtained from the round 8 selected DNA, and these sequences were input into MEME online motif searching software [81] to find a common motif among the selected DNA, and a graphical representation of the motif was generated using WebLogo [82].

Table 2.4 DNA probe and primers used for SELEX

Name	Sequence[a]
SELEX single-stranded probe	XbaI EcoRI HindIII BamHI EcoRI SalI GGTCTAGAGAATTCAAGCTTC$(N)_{30}$GGATCCGAATTCGTCGAC
SELEX primer F	GCTCAGGTCTAGAGAATTCAA
SELEX primer R	ACTACTGTCGACGAATTCGGA

[a]DNA primers are listed from 5' to 3'.

2.11 Protein-protein pull-down assay

The his-tagged protein was bound to Ni-NTA magnetic agarose beads (Qiagen, Valencia, CA) in binding buffer (0.02 M sodium phosphate, 0.5 M NaCl, pH 7.4) and incubated at room temperature for 30 min on a vortex mixer (Fisher, Pittsburgh, PA). The bead-bound protein was then mixed with 2.5 mg/mL *P. furiosus* soluble cell extract from cells grown either before or after cold shock and incubated at 55 °C for 20min or 60 min with intermittent mixing to keep the beads in suspension. Then bead-bound protein-protein complexes were subjected to three quick washes at room temperature with incubation buffer (50 mM EPPs, 100 mM KCl, 1 mM EDTA, 5% glycerol, 0.1% Triton-X, 1 mM DTT, pH 7.5). Then proteins were eluted at room temperature for 10 min with elution buffer (500 mM imidazole, 0.02 M sodium phosphate, 0.5 M NaCl, pH 7.4). Eluted proteins were analyzed by SDS-PAGE (BioRad, Hercules, CA). The gel was run at 200 V for 60 min and stained with silver staining [70].

2.12 Analytical gel filtration to determine protein quaternary structure

A Superdex 75 10/300 GL size exclusion column (GE Healthcare, Pittsburgh, PA) was used for analytical gel filtration to determine the quaternary structure of proteins. The running buffer used was 20 mM Na_2HPO_4, 500 mM NaCl, pH 7.5. Molecular weight standards were run through the column simultaneously and the elution volume (V_e) of each was noted. The standard mixture contains bovine thyroglobulin (10 mg/mL), bovine γ-globulin (10 mg/mL), chicken ovalbumin (10 mg/mL), horse myoglobin (5 mg/mL), and vitamin B_{12} (1 mg/mL). Blue dextran was used to determine the column void volume (V_0), and a standard curve of molecular weight versus V_e/V_0 was used to determine the corresponding approximate molecular weights of the sample peaks.

Figure 2.1 Fluorescence-detected DNase I footprinting scheme.

A protein-bound fluorescently tagged probe is nicked with DNase I, and the region of DNA bound by the protein is protected from cleavage. Strands are denatured prior to undergoing capillary electrophoresis on a capillary sequencer (ABI 3730x1, Applied Biosystems, Foster City, CA). Fluorescently labeled fragments are detected, and resulting electropherograms of samples with and without added protein are overlaid to determine footprint position.

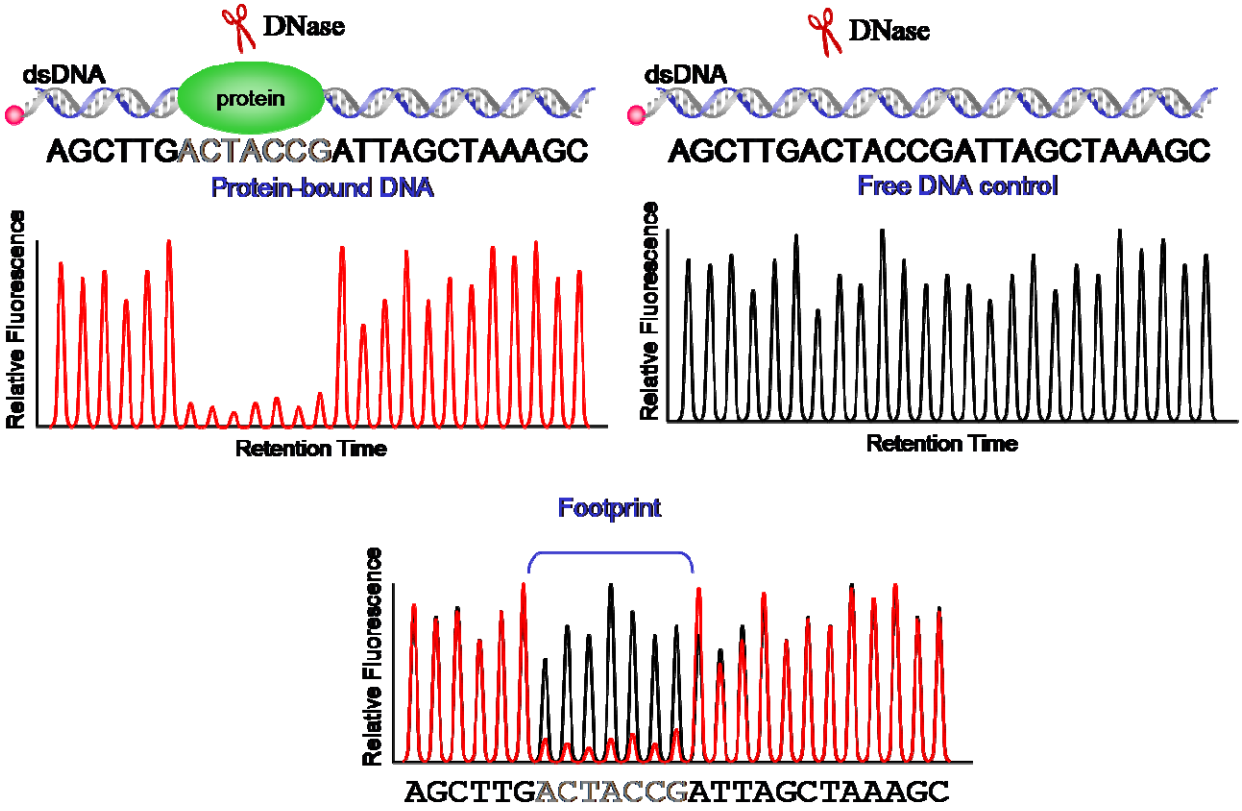

DNase
dsDNA
protein
AGCTTGACTACCGATTAGCTAAAGC
Protein-bound DNA
Relative Fluorescence
Retention Time
DNase
dsDNA
AGCTTGACTACCGATTAGCTAAAGC
Free DNA control
Relative Fluorescence
Retention Time
Footprint
Relative Fluorescence
AGCTTGACTACCGATTAGCTAAAGC

Figure 2.2 SELEX scheme.

The SELEX library was made up of single-stranded oligonucleotides designed to contain 30 bases of random DNA flanked with two ~20-base primer sites each containing an *Eco*RI restriction site. Second-strand synthesis and amplification of the library was performed using primers complementary to the two priming sites. SELEX probes with sequences that had higher affinity to protein were selected from the library pool using EMSA. Shifted DNA was gel-purified and PCR-amplified for an additional selection round. Six rounds of selection via EMSA, gel-purification, and PCR-amplification were carried out. The resulting selected DNA was digested with *Eco*RI and concatemerized for cloning into pUC18. Cloned plasmids were transformed into XL1-Blue cells, and insert-containing plasmids were identified via blue/white colony screening. Clones were then sequenced to identify the shared motif.

Random sequence library

XbaI EcoRI HindIII BamHI EcoRI SalI

GGTCTAGAGAATTCAAGCTTC(N)$_{30}$GGATCCGAATTCGTCGAC

Second strand synthesis by PCR

Primer 1

5′ ggTCTAGAGAATTCAAGCTTc-**N$_{30}$**-GGATCCGAATTCGTCGAC 3′

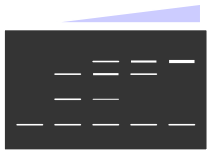
3′ ccAGATCTCTTAAGTTCGAAg-**N$_{30}$**-CCTAGGCTTAAGCAGCTG 5′

Primer 2

Selection of binding sequences by gel-shift

5 - 10 x

Protein:



Selected DNA

Unselected DNA

Gel-purification & PCR amplification of shifted DNA

Final selection of DNA

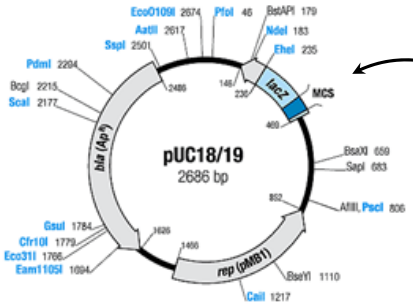
Digestion (*EcoRI*)

5′ AATTCAAGCTTc-**N$_{30}$**-GGATCCG 3′

3′ GTTCGAAg-**N$_{30}$**-CCTAGGCTTAA 5′

Concatamerization

Cloning into pUC18



5′ AATTCAAGCTTc-**N$_{30}$**-GGATCCG 3′

3′ GTTCGAAg-**N$_{30}$**-CCTAGGCTTAA 5′

Sequencing

CHAPTER 3

TRANSCRIPTION FACTOR DISCOVERY AND CHARACTERIZATION

3.1 Targeted transcription factor discovery by DNA affinity protein capture

3.1.1 Selection of target ORFs for transcription factor discovery

P. furiosus exhibits a cold shock response when the temperature of a growing culture is shifted from its optimal growth temperature, 95 °C to the nearly minimum growth temperature, 72 °C. Transcriptional analyses using whole-genome DNA microarrays representing 2,065 open reading frames (ORFs) in the *P. furiosus* genome showed that there were significant changes in the expression levels of many genes after the cold shock [68]. That these changes occur mainly at the transcriptional level makes it feasible to identify transcription factors that associate with particular promoters under different growth temperatures using a DNA affinity protein capture approach. The first step of this approach is to select target ORFs that show significant regulation in microarray expression profiles after cold shock.

In the cold shock experiment, *P. furiosus* was grown on maltose at 95 °C, and then cultures were shocked by rapidly dropping the temperature from 95 to 72 °C. This resulted in a 5 h lag phase, during which little growth occurred. Transcriptional analyses showed that cells undergo three very different responses at 72 °C: an early shock (1 to 2 h), a late shock (5 h), and an adapted response. In addition, *P. furiosus* was grown at 72 °C in a cold-adapted experiment. The results of this cold shock experiment revealed a significant down-regulation of gene expression. There were 171, 69, and 152 ORFs down-regulated by more than 2.5 fold after 1, 2,

and 5 h, respectively. Similar microarray analyses were performed in the cold-adapted experiment and 189 genes were down-regulated after adaptation to the lower growth temperature [68]. On the other hand, some ORFs were significantly up-regulated in response to the temperature decrease, and we focused on these ORFs for transcription factor discovery.

After 1, 2, and 5 h of shock at 72 °C, and in cells adapted to 72 °C, there were 49, 35, 30, and 59 ORFs, respectively, that were significantly up-regulated (>2.5 fold; $p < 0.01$). Table 3.1 lists the expression profile obtained for a subset of *P. furiosus* genes comparing cultures grown before and after cold shock [68]. An operon (PF0324 to PF0326) encoding three conserved hypothetical proteins is the most dramatically up-regulated of all ORFs involved in the response to the temperature decrease. It was up-regulated by 23-fold after 5 h and by 57-fold in the adapted cells. Besides this operon, there were a number of other ORFs encoding conserved hypothetical proteins that were up-regulated in the late shock phase that increased further in the adapted cells such as the operon PF1072-1074. Although their functions were unclear, these ORFs were selected as targets for transcription factor discovery because of the dramatic up-regulation. In fact, the responses to the late shock phase and in the adapted cells were dominated by "conserved hypothetical proteins" with 42 of 76 up-regulated ORFs falling in this category. For the annotated ORFs, several of them are only given general names.

These up-regulated ORFs were chosen as targets for discovery of transcription factors relevant to low temperature response. Proteins that bound to the DNA upstream of these ORF targets were identified from cell extracts of cultures grown before and after cold shock using the DNA affinity protein capture method described below.

Table 3.1 ORFs whose expression is up-regulated more than 5-fold in cells adapted to 72 °C compared to cells adapted to 95 °C [68]

ORF	Description	Fold change
PF0101	Conserved hypothetical protein	5.5
PF0190	Conserved hypothetical protein	7.4
PF0323	Conserved hypothetical protein	5.3
PF0324	Conserved hypothetical protein	57.4
PF0325	Conserved hypothetical protein	39.0
PF0327	Conserved hypothetical protein	5.3
PF0429	Putative proline permease	14.6
PF0719	Hypothetical protein	5.5
PF0934	Conserved hypothetical protein	6.9
PF0935	Acetolactate synthase	9.9
PF1062	Conserved hypothetical protein	6.2
PF1072	Conserved hypothetical protein	12.7
PF1073	Conserved hypothetical protein	9.8
PF1074	Conserved hypothetical protein	11.3
PF1528	Imidazoleglycerol Pi synthase	5.2
PF1677	Conserved hypothetical protein	6.8
PF1678	2-Isopropyl malate synthase	9.4
PF1679	3-Isopropyl malate dehydratase I	8.8
PF1701	Chorismate mutase	14.5
PF1702	Aspartate aminotransferase	7.1
PF1703	Prephenate dehydrogenase	8.5

3.1.2 Protein capture on PF0324 and PF1072 promoter DNA

Since archaeal transcriptional regulation appears to be similar to bacteria, it seems reasonable to expect that regulatory transcription factor binding sites (operators) may occur just upstream of ORFs, in the promoter region. Thus, the DNA probes used in DNA affinity protein capture experiments were designed to cover approximately 200 bp DNA upstream of the ORF translation start site; approximately 100 bp of DNA downstream from the ORF start were included to ensure complete coverage of possible transcription factor binding sites. The DNA probe for the PF0324-PF0326 target region was designed to contain 207 bp upstream and 108 bp downstream from the PF0324 ORF start site with the biotin label anchoring the probe to the bead on the side of the DNA farthest from the ORF. The DNA probe for the PF1072-P1074 target region was designed such that 176 bp of DNA upstream from translation start site was included,

along with 93 bp downstream of the PF1072 ORF start; the biotin label was placed on the upstream side of the probe (Fig. 3.1).

In the DNA affinity protein capture experiment [83, 84], the biotinylated probe DNA was bound to the magnetic beads (Fig. 3.2) and then bead-bound DNA was incubated in cell extract from cells grown either before or after cold shock. Then heparin was used as a competitor to remove nonspecific DNA-binding proteins since it resembles DNA polyanionic character. Then the beads and bound proteins were magnetically separated from solution, and DNA-bound proteins were eluted from the beads and separated using SDS-PAGE (Fig. 3.3). The gel was then visualized by silver staining (Fig. 3.4). Many of the protein bands were the same between the two cell extracts and between different DNA probes. Bands which appeared differentially were excised and subjected to in-gel tryptic digestion, followed by mass spectrometry identification. Mass spectral data were searched against a database including all archaea, and the error allowed in mass matching was 200 ppm. Table 3.2 contains protein band identification results.

Table 3.2 Identification of protein binds using peptide mass fingerprinting

Band number[a]	Annotation[b]	Locus	Molecular weight (kDa)	Sequence coverage
1	DNA topoisomerase VI subunit b	PF1579	64.43	40%
2	hypothetical protein	PF0128	46.01	54%
3	DNA directed RNA polymerase subunit a	PF1562	44.40	27%
4	DNA topoisomerase VI subunit a	PF1578	44.07	52%
5	Replication factor A	PF2020	41.05	35%
6	hypothetical protein	PF1931	27.65	60%
7	transcription initiation factor TFIID chain a (TBP)	PF1295	21.32	28%
8	hypothetical protein	PF0396	21.13	55%
9	conserved hypothetical protein	PF1476	19.08	45%
10	transcriptional regulatory protein, asnC family	PF2053	17.21	43%
11	transcriptional regulatory protein, lrp family	PF0250	16.98	32%
12	hypothetical protein	PF1072	15.80	51%
13	hypothetical protein	PF0574	15.63	37%

[a] Numbers correspond to numbered bands in Figure 3.4.
[b] Protein annotations are derived from the NCBI database.

Identified bands fell into four general categories: basal transcription factors (bands 3, 7), nonspecific DNA-binding proteins that could be challenged off with heparin (bands 1, 4, 5), several hypothetical proteins of unknown function (bands 2, 6, 8, 13), and potential transcriptional proteins (bands 9, 10, 11, 12). Detailed analysis of all identified proteins follows.

DNA-directed RNA polymerase subunit a (PF1562, band 3) and transcription initiation factor TFIID chain a (TBP) (PF1295, band 7) were identified in both cell extracts for both DNA probes, which helped to validate the experiment. A faint band between band 3 and band 6 for both lanes from PF1072 probe may be transcription initiation factor IIB (TFB); however, it was not positively identified from the PF0324 probe.

The proteins which were eluted from the bead-DNA after a wash with heparin were considered to have lower affinity to the DNA. Experiments using heparin as a competitor for nonspecific DNA-binding proteins indicated that proteins in bands 1, 4, and 5 were probably associating with the DNA probes in a nonspecific manner.

Band 1 and band 4 were identified as two subunits of DNA topoisomerase. The function of these proteins is to catalyze and guide the unknotting of DNA by creating transient breaks in the DNA [85], and nonspecific affinity for DNA is a necessary part of this function. Band 5 was identified as replication factor A, a protein which has been studied [86]. This protein forms a complex with ssDNA. A bioinformatics approach was taken to determine if replication factor A was a potential dsDNA binding protein. Homology was found with *Sulfolobus solfataricus* Single Stranded DNA-binding Protein, proven to have an active role in transcription [87]. A conserved domain search showed that replication factor A also contained the OB nucleic acid binding fold domain. However, its weak DNA-binding affinity disqualifies it as a potential transcription factor.

Four of the identified proteins fell into the category of hypothetical proteins (bands 2, 6, 8, 13), and sequence analysis was used to determine their probable functions. The protein sequences were subjected to BLAST searches of the NCBI nonredundant database [71] and Conserved Domain searches of the Conserved Domain Database of NCBI [73]. The protein identified in band 13, PF0574, was predicted to be a nucleic acid-binding protein with a conserved PIN (PilT N terminus) domain. Although the function of this protein is unknown, it seems that PF0574 is captured by the DNA probe because of its nucleic acid-binding ability. Sequence analysis of the remaining three hypothetical proteins (bands 2, 6, 8) indicated that they were a predicted ATPase: PF0128, PF1931 and PF0396. There is no indication that they are DNA-binding proteins. It is possible that their association with the bead-DNA complex occurred as a result of protein-protein contacts with DNA-bound proteins. It is worth noting that the apparent gel molecular weight of PF0396 is less than its real molecular weight. This band may be a fragment of the PF0396 protein.

The remaining bands (9-12) were identified as potential transcriptional regulators. PF0250 protein and PF1476 protein were apparent in both cell extracts for both DNA probes. PF0250 has homology with families containing helix-turn-helix asnC-type DNA-binding domains, predicted to play a role in transcriptional regulation, and also with a family of Lrp (leucine response protein) transcriptional regulators (COG1522 [88]). PF1476 matched the transcriptional regulator PadR-like family which is named after PadR, a protein that is involved in negative regulation of phenolic acid metabolism. PF1476 also matched with a family of transcriptional regulators (COG 1695 [88]) containing a helix-turn-helix DNA-binding domain. PF0250 protein and PF1476 protein were identified from both DNA probes showing that the binding between DNA and these proteins may be nonspecific. PF2053 protein was apparent in both cell extracts for the

PF0324 DNA probe. This protein was predicted to be a member of the Lrp (leucine response protein) family as was the PF0250 protein (Fig. 3.5). It also highly homologous with helix-turn-helix AsnC type transcriptional regulators, which are homologs of an autogenously regulated activator of asparagine synthetase A transcription in *Escherichia coli*. PF2053 protein was represented differentially between the two cell extracts, and this protein band was only present on the PF0324 probe. However, in other experiments, PF2053 protein bound to PF0272 promoter-ORF DNA and was found to be associated with LrpA (PF1601) promoter-ORF DNA as well [84]. Therefore the PF2053 protein possibly nonspecifically associates with DNA. Secondary structure prediction revealed the presence of three helices in the N-terminal (residues 3-46) (Fig. 3.6) which contained homology to the HTH_AsnC domain, an indication that PF2053 contained an HTH DNA-binding domain. A BLAST search of the sequence resulted in approximately 100 hits having an e-value less than 0.0001, with the majority of high-scoring hits falling within the archaea as either predicted transcriptional regulators or conserved hypothetical proteins. A tree view of the highest scoring BLAST hits can be seen in Figure 3.7.

Like PF2053 protein, the PF1072 protein was represented differentially between the two cell extracts, and this protein band was only present on the PF1072 probe, suggesting that it binds specifically to the upstream DNA of its own ORF. This may indicate that it is a self-regulated repressor protein. Analysis of the PF1072 protein sequence revealed homology with five conserved protein domains that are transcription-factor related (Fig. 3.8). The closest match was to a family of conserved proteins, helix-turn-helix Arsenical Resistance Operon Repressor (HTH_ArsR). Other matches were annotated as predicted transcriptional regulators containing helix-turn-helix DNA binding motifs. The middle region (residues 55-72) of PF1072 protein was aligned with a dimerization interface which is similar to the dimer interface of the

SmtB repressor [89], and residues 77 and 80 were predicted to be putative zinc binding ligands while the residues from 55 to 100 might be involved in DNA interactions (Fig. 3.8). The overlap between the putative zinc binding site and the DNA binding domain may suggest that the PF1072 protein tends to dissociate from DNA in the presence of zinc ions.

Secondary structure prediction revealed the presence of four helices in the middle region (residues 45-100) (Fig. 3.9) that contained homology to the HTH_ArsR domain, a further indication that the PF1072 protein contained an HTH DNA-binding motif. A BLAST search of the sequence resulted in approximately 100 hits having an e-value less than 0.001, with the majority of high-scoring hits falling within the archaea as either predicted transcriptional regulators or conserved hypothetical proteins. The three closest hits (having e-values of less than or equal to $2x10^{-74}$) were from the three other members of the Thermococcaceae family and are presumably PF1072 orthologs: PAB0751 (79% sequence identity), PH1101 (77% sequence identity), and TK1041 (72% sequence identity). A tree view of the highest scoring BLAST hits can be seen in Figure 3.10.

The PF1072 protein, identified by capture on its own promoter DNA, emerged as the best potential transcription factor candidate and was chosen for expression and characterization since it presented in the DNA affinity protein capture assay differentially and sequence analysis indicated that this protein contains a DNA-binding domain. The PF2053 protein was also chosen for expression in case the PF0324, PF0272, and PF1601 promoter-ORF DNA happen to contain the specific binding sites of this protein, although it seems that the PF2053 protein binds to these probes nonspecifically.

3.2 Validation of target protein as a potential transcription factor

3.2.1. Expression and purification of recombinant his-tagged target proteins

In order to show that the target proteins were sequence specific DNA-binding proteins, they were overexpressed as his-tagged recombinant proteins in *E. coli* for *in vitro* studies. A vector containing a TEV protease site between the his-tag and the cloning site (pET24dBAM-TEV) was obtained from the laboratory of Michael Adams (University of Georgia). Use of this vector allowed expression of target proteins with an N-terminal his-tag that facilitates protein purification. The his-tag can then be removed by TEV protease in case the his-tag has an effect on DNA binding.

The PF1072 protein (Fig. 3.11) is 137 amino acids in length and has a calculated molecular weight of 15,804 Da. With the addition of the his-tag and TEV cleavage site, the new molecular weight is 17,914 Da. The PF2053 protein (Fig. 3.12) is 148 amino acids in length and has a calculated molecular weight of 17,215 Da. With the addition of the his-tag and TEV cleavage site, the new molecular weight is 19,324 Da. Since the target protein sequences contained some rare codons (Arg mostly) as compared to the codon frequency in *E. coli* proteins, BL21-CodonPlus(DE3)-RIPL strain, which compensates for rare codons by expressing supplementary corresponding tRNAs, was used for expression. Target proteins were expressed in abundance, predominantly in soluble form.

His-tagged proteins were purified from cell extract using a nickel-affinity column, and a gradient of imidazole was used to elute the protein from the column. Purified fractions were collected and pooled for buffer exchange using a desalting column. Once the protein was exchanged into a suitable salt-containing buffer, small aliquots were stored at -80 °C to be thawed individually for use in *in vitro* assays.

3.2.2. His-tagged target proteins bind to promoter DNA in vitro

An electrophoretic mobility shift assay (EMSA) was performed using PF1072 promoter DNA to verify that the PF1072 protein binds specifically to its own promoter DNA, used in the DNA affinity protein capture experiment (Fig. 3.13). Four distinct protein-DNA complexes were evident in the range of protein-DNA ratios tested. The PF1072 promoter DNA was shifted completely at a protein/DNA mole ratio of around 24.

For comparison, a probe covering 200 bp of the PF1072 ORF was also tested in EMSA, as well as another promoter, of PF1076, which was up-regulated up to 4-fold after cold shock. The PF1072 protein shifted these probes only at very high protein/DNA mole ratios of around 48. Also, the distinct protein-DNA complexes observed with the PF1072 promoter were not as obvious for either of these probes. These data indicated that the PF1072 protein has higher affinity for its own promoter DNA.

The PF2053 protein was found to bind to the PF0324 promoter DNA from which it was identified (Fig. 3.14). Two bands of protein-DNA complexes were evident in the range of protein-DNA ratios tested, and the PF2053 protein shifted the PF0324 promoter DNA completely at a protein/DNA mole ratio of around 16.

Two other DNA probes were used to test the DNA binding ability of the PF2053 protein. PF0190 was up-regulated up to 7-fold after cold shock and PF2051 was not induced by the cold shock but is predicted to be the first gene in the same operon with PF2053. The promoter DNA of both genes were shifted by the PF2053 protein. The difference was that there were four bands of protein-DNA complexes for the PF0190 promoter but only two for PF2051. The PF0190 promoter DNA was shifted completely at a protein/DNA mole ratio of around 24, whereas only about half of the PF2051 promoter DNA was shifted at the same ratio. To further confirm the

specificity of the PF2053 protein binding to the PF0324 promoter DNA, a synthetic DNA consisting of dAdC repeating units was used as a control in EMSA. The PF2053 protein did not completely shift this synthetic DNA even at a protein/DNA mole ratio of 48, although some shifting of the DNA was apparent. A minimal amount of DNA began shifting at a ratio of 24; however, the band was less well-defined and the appearance of the shift pattern was distinctly different than that observed for other promoter fragments.

3.2.3. Confirmation of specific binding of target protein to the promoter DNA

The nonspecific competitor heparin used in the DNA affinity protein capture experiments was used as a competitor in EMSA for nonspecific DNA binding. The structure and electrostatic properties of heparin are very similar to DNA and it is not bound by DNA-staining agents, so heparin can be used as a nonspecific competitor in EMSA. The binding between protein and specific bases is much stronger than the solely electrostatic association between the protein and the phosphate-sugar backbone. So a protein binding nonspecifically to DNA through electrostatic association should be challenged off DNA by heparin since heparin has a highly negative overall charge. As can be seen in Figure 3.15, the association of the PF1072 protein with its own promoter DNA survived heparin challenge, suggesting that this binding is sequence specific. The binding at heparin concentrations 0.1 and 1 μg/mL was the same as in the EMSA experiment without heparin challenge while the binding was blocked at heparin concentrations 10 to 1000 μg/mL. This result indicated that the protein bound to DNA specifically at low protein/DNA ratio and formed protein-DNA complexes with high mobility. But when protein/DNA mole ratios rose to 24, extra proteins began to associate on the DNA in a nonspecific manner, thereby creating a protein-DNA complex with low mobility and less band definition. These data show that the PF1072 protein is a sequence specific DNA-binding protein.

For the PF2053 protein, the binding at heparin concentrations between 0.1 and 1.0 μg/mL was the same as in the EMSA experiment without heparin challenge while the binding was blocked completely at heparin concentration 10 μg/mL (Fig. 3.16). This result suggested that the PF2053 protein binds to the PF0324 promoter DNA nonspecifically.

3.3 Biophysical characterization of target protein

3.3.1. Identification of target protein binding sites on promoter DNA

To define the target protein's specific binding site in the promoter region, fluorescence-based DNase I footprinting was performed [76]. The promoter DNA was cloned into the pUC18 vector to make a plasmid used as template for probe amplification. Primers labeled with fluorescent tags were used to PCR amplify the footprinting probe. Both primers were labeled at 5' ends, one with 6FAM (6-carboxy-fluorescein) and the other with HEX (hexachlorofluorescein). The dual-labeled probe enabled both strands from one footprinting experiment to be analyzed simultaneously using a capillary electrophoreses automated DNA sequencer. Internal standards GS-500 ROX (labeled with ROX, Applied Biosystems, Foster City, CA) were used as size standards.

The footprinting experiment was performed using binding conditions determined by EMSA. For the PF1072 protein, there was an evident footprint in the promoter region of its own gene. The footprint covered about 50 bases, extending from -5 to -55 relative to the PF1072 translation start site on the coding strand (Fig. 3.17). The footprint on the opposite strand was located at an equivalent position, -5 to -55 from the translation start site.

The footprint of the PF2053 protein on the PF0324 promoter region was not as distinct. The footprint covered about 50 bases, extended from -70 to -120 relative to the PF0324

translation start site on the coding strand (Fig. 3.18). The footprint on the opposite strand was located at an equivalent position, -70 to -120 from the translation start site. The intensity of some peaks within and surrounding the footprints increased, indicating that nucleotides at these positions were hypersensitive to DNase I when the protein was bound to the DNA (see arrows, Fig. 3.18). This may suggest DNA bending, the nucleotides at the bend positions becoming more susceptible to DNase I cleavage. The footprint was most evident at protein/DNA mole ratios of 40, comparable to the ratios required for complete shifting of the probes in EMSA. Increasing the protein/DNA ratio up to 80 resulted in complete protection of the DNA from digestion, which made the fluorescence signal disappear completely and this seems to correspond to the low-mobility nonspecific complexes observed in EMSA.

3.3.1. The target protein recognizes a consensus binding sequence

The artificial selection method SELEX was used to define the consensus DNA motif recognized by the target protein [79, 80]. A pool of synthetic DNA containing a central 30-bp randomized sequence was amplified as the library. EMSA with the target protein was used to select sequences bound specifically by the protein. The pool of selected sequences from the last round was PCR-amplified and purified for the next round selection. After the final selection round was complete, the resulting DNA was cloned and sequenced.

After eight rounds of selection, the consensus palindromic DNA sequence $GTACAn_6TGTAC$ was uncovered as the preferred binding sequence for the PF1072 protein. Nine DNAs contained this motif in 20 total sequenced SELEX DNAs (Fig. 3.19). Only 5 out of 20 sequences displayed the perfect palindrome $GTACAn_6TGTAC$. The DNase footprint site in the PF1072 promoter was verified to contain two $GTACAn_6TGTAC$ motifs; one of them was perfect and the other had one mismatch. The presence of the $GTACAn_6TGTAC$ motifs in the

PF1072 protein footprint regions confirmed that the consensus derived from the synthetic SELEX library had some relationship to the *in vitro* recognition site for this protein. This result also explained why there was more than one protein-DNA complex observed in EMSA. This suggests that one of the motifs began to be occupied first, followed by the other one, forcing the resulting protein-DNA complexes to have lower and lower electrophoretic mobility.

After five rounds of selection, the consensus DNA sequence ASCTGC was uncovered as the preferred binding sequence for the PF2053 protein. 7 DNAs contained this motif in 9 total sequenced SELEX DNAs (Fig. 3.20). 6 of them displayed the motif ASCTGC. This motif was not found in the PF0324 promoter DNase footprint but it had a similar motif AGATGC (from -89 to -94 relative to the translation start site). This motif could be the binding site of the PF2053 protein. The fact that the ~50-base footprinting region is much bigger than this 6-base motif suggests that multiple proteins may be binding to adjacent positions within the footprint.

3.4 Verification of additional potential binding sites in the *P. furiosus* genome

The target protein DNA-binding site identified from SELEX was used to search for other potential binding sites in the upstream regions of other genes within the genome**.** The DNA sequence was searched against a database of upstream of ORF regions (UORs) extracted from the *P. furiosus* genome, which was created by a colleague Darin Cowart.

Searching with the PF1072 protein binding palindrome GTACAn_6TGTAC returned 39 UORs containing the motif with a total of 32 nonredundant motifs considering some ORFs next to each other share the same UOR (Table 3.3). Only the UOR of PF1072 has a perfect GTACAn_6TGTAC palindrome, 3 sequences have one mismatch and the other 28 sequences have two mismatches. There were more UORs found with more than two mismatches, and these

UORs were ignored since the binding between PF1072 protein and these UORs is expected to be very weak because of too many mismatches. Only a few ORFs found to contain the binding sequence in their upstream DNA overlapped with low-temperature regulated ORFs, most of them were down-regulated in the DNA microarray experiment. As can be seen in Table 3.4, 11 out of 32 of the ORFs that have the motif in their upstream DNA were regulated by low temperature more than two-fold, and only the ORF coding PF1072 protein is up-regulated.

Table 3.3 UOR database search results for the motif GTACAn_3TGTAC

UOR[a]	Motif	Start[b]	Stop[b]	ORF Annotation[c]
PF0163	GTTCATCGTGGAGTAC	-86	-71	hypothetical protein
PF0173	ATACATTTGCAAGTAC	-181	-166	protein export protein SecF
PF0195	TTACAGCGCTTGGTAC	-125	-110	hypothetical protein
PF0289	ATACACAAAATTGTCC	-87	-72	phosphoenolpyruvate carboxykinase (GTP)
PF0338	GTAGAAGTTCAAGTAC	-81	-66	flagellin
PF0346	GTACAAAAATTTGTAG	-141	-126	aldehyde:ferredoxin oxidoreductase
PF0347	GTACAAAAATTTGTAG	-69	-54	hypothetical protein
PF0489	TTACATTTACTTGAAC	-126	-111	hypothetical protein
PF0666	GTAAATTTAATTATAC	-172	-157	nol1-nop2-sun family putative nucleolar protein IV
PF0702	TTACATTCTCTTCTAC	-179	-164	x-pro dipeptidase
PF0758	TTACAGGTAACGGTAC	-144	-129	hypothetical protein
PF0759	TTACAGGTAACGGTAC	-114	-99	hypothetical protein
PF0785.3n	GTGCATAATTATATAC	-145	-130	hypothetical protein
PF0786	GTGCATAATTATATAC	-148	-133	transposase
PF0821	GAACAGTTTTTTGCAC	-181	-166	50S ribosomal protein L34
PF0897.1n	ATACCTCCACCTGTAC	-142	-127	hypothetical protein
PF0897.1n	GCACATCTTTGTGTAG	-192	-177	hypothetical protein
PF0947	GAACAGACTAATGTAC	-198	-183	hypothetical protein
PF0961	CTACATACCTATGTCC	-136	-121	hypothetical protein
PF0989	GGACATGTGAATATAC	-64	-49	phenylalanyl-tRNA synthetase alpha subunit
PF1003	CTACATACAAGTGAAC	-118	-103	phosphate-binding periplasmic protein precursor
PF1067	GTGCATCAAATTTTAC	-24	-9	hypothetical protein
PF1068	GTGCATCAAATTTTAC	-175	-160	hypothetical protein
PF1072	GTACAGAAAAATGTAC	-42	-27	hypothetical protein
PF1072	CTACAAAAATTTGTAC	-27	-12	hypothetical protein
PF1089	GTATAAATACCTCTAC	-41	-26	hypothetical protein

PF1123	GGACATCCTTTGGTAC	-97	-82	hypothetical protein
PF1145	TTACACACATTTGTTC	-79	-64	mrp/nbp35 family nucleotide-binding protein
PF1146	TTACACACATTTGTTC	-64	-49	hypothetical protein
PF1322	GTGCACTGTATAGTAC	-85	-70	hypothetical protein
PF1377	ATACATTTAAAAGTAC	-127	-112	transcription initiation factor IIB
PF1378	ATACATTTAAAAGTAC	-36	-21	ribonuclease P protein component 2
PF1472	GTAAAATAACATTTAC	-31	-16	aspartate/serine transaminase
PF1512	TTACATAAAATTCTAC	-112	-97	hypothetical protein
PF1513	TTACATAAAATTCTAC	-59	-44	hypothetical protein
PF1832	GTACCTCGAAGAGTAC	-198	-183	regulatory protein
PF2036	GTCCAGAACGTTGTAC	-136	-121	magnesium and cobalt transporter
PF2057	GAACAGAAGAATCTAC	-151	-136	hypothetical protein
PF2060.1n	ATACATCATGCTGCAC	-31	-16	hypothetical arylsulfatase regulatory protein

[a] UOR (Upstream of ORF Region) designation corresponds to the locus of the ORF from which the upstream sequence was taken.
[b] Start and stop positions are relative to the UOR sequence where -1 corresponds to the first nucleotide upstream from the ORF start.
[c] Protein annotations are derived from the NCBI database.

Table 3.4 Summary of motifs found upstream of low temperature regulated ORFs

UOR[a]	Motif	Start[b]	Stop[b]	ORF Annotation[c]	Fold
PF0195	TTACAGCGCTTGGTAC	-125	-110	conserved hypothetical protein	-2.21
PF0289	ATACACAAAATTGTCC	-87	-72	phosphoenolpyruvate carboxykinase (gtp)	-5.53
PF0758	TTACAGGTAACGGTAC	-144	-129	conserved hypothetical protein	-3.12
PF0947	GAACAGACTAATGTAC	-198	-183	conserved hypothetical protein	-2.26
PF0989	GGACATGTGAATATAC	-64	-49	phenylalanyl-tRNA synthetase	-2.21
PF1072	GTACAGAAAAATGTAC	-42	-27	conserved hypothetical protein	12.71
PF1072	CTACAAAAATTTGTAC	-27	-12	conserved hypothetical protein	12.71
PF1123	GGACATCCTTTGGTAC	-97	-82	conserved hypothetical protein	-2.52
PF1513	TTACATAAAATTCTAC	-59	-44	hypothetical protein	-2.93
PF1832	GTACCTCGAAGAGTAC	-198	-183	regulatory protein	-2.26

[a] UOR (Upstream of ORF Region) designation corresponds to the locus of the ORF from which the upstream sequence was taken.
[b] Start and stop positions are relative to the UOR sequence where -1 corresponds to the first nucleotide upstream from the ORF start.
[c] Protein annotations are derived from the NCBI database.

Searching with the PF2053 protein binding motif ASCTGC (S=G/C) returned 99 UORs containing the motif with a total of 91 nonredundant motifs considering some ORFs next to each

other share the same UOR (Appendix Table A2). As can be seen in Table 3.5, 12 out of 99 ORFs were down-regulated more than two-fold at low temperature, and only 5 ORFs were up-regulated more than two-fold. These results suggested that the PF2053 protein was not closely related to the response of *P. furiosus* to low temperature.

Table 3.5 UOR database search results for the motif ASCTGC

UOR[a]	Motif	Start[b]	Stop[b]	ORF Annotation[c]	Fold
PF0190	ACCTGC	-15	-10	hypothetical protein	7.37
PF0430	ACCTGC	-159	-154	phosphoribosylglycinamide formyltransferase 2	-5.68
PF0448	AGCTGC	-162	-157	putative mannose-1-phosphate guanyltransferase	-2.19
PF0621	ACCTGC	-144	-139	hypothetical protein	-2.64
PF0629	ACCTGC	-67	-62	hypothetical protein	-2.17
PF0942	AGCTGC	-148	-143	dihydroxy-acid dehydratase	2.22
PF1032	ACCTGC	-165	-160	hypothetical protein	-2.75
PF1033	AGCTGC	-159	-154	peroxiredoxin	-4.86
PF1108	ACCTGC	-59	-54	putative alpha-dextrin endo-1, 6-alpha-glucosidase	-2.07
PF1186	AGCTGC	-191	-186	NADH oxidase	-2.86
PF1390	ACCTGC	-103	-98	hypothetical protein	-7.53
PF1400	AGCTGC	-88	-83	hypothetical protein	-2.27
PF1400	AGCTGC	-90	-85	hypothetical protein	-2.27
PF1400	ACCTGC	-117	-112	hypothetical protein	-2.27
PF1529	AGCTGC	-137	-132	pyridoxine biosynthesis protein	3.54
PF1592	AGCTGC	-165	-160	tryptophan synthase subunit beta	2.10
PF1832	ACCTGC	-141	-136	regulatory protein	-2.26
PF1863	AGCTGC	-65	-60	dimethyladenosine transferase	-4.14
PF1961	AGCTGC	-154	-149	tungsten-containing formaldehyde ferredoxin oxidoreductase	3.77
PF2048	AGCTGC	-99	-94	peptidase	-2.43

[a] UOR (Upstream of ORF Region) designation corresponds to the locus of the ORF from which the upstream sequence was taken.
[b] Start and stop positions are relative to the UOR sequence where -1 corresponds to the first nucleotide upstream from the ORF start.
[c] Protein annotations are derived from the NCBI database.

3.5 Validation of additional binding sites in the *P. furiosus* genome

3.5.1 Selection of putative promoter regions to test

In order to verify that the identified motifs represented true binding sites, several of the putative promoter regions were selected for further study in conjunction with the PF1072 protein.

Promoter-ORF DNA fragments were cloned from *P. furiosus* genomic DNA into pUC18 for versatile PCR amplification of probes for EMSA, footprinting and *in vitro* transcription experiments. DNA fragments were designed to cover a ~300 bp region (from -200 to +100) with respect to translation start site, since at least ~100 bp of the ORF was required for an adequate transcript in an *in vitro* transcription assay and a total length of ~300 bp DNA was required to assure that the promoter region occurs within the probe for footprinting.

Besides its own promoter, the upstream DNA of two other slightly up-regulated ORFs were chosen for EMSA and footprinting analysis with the PF1072 protein: PF0666 and PF2036. These two ORFs were of interest because they were the only two genes with a predicted function among nine up-regulated ORFs that have the PF1072 protein binding motif in their promoters. PF0666 is annotated as a nucleolar nol1-nop2-sun family protein and was up-regulated 1.26-fold at low temperature. It was used as a positive control to ensure that PF1072 protein was able to bind to the motif even though there were two mismatches. PF2036 encodes a magnesium and cobalt transporter according to annotation, and it is likely the start of a two-ORF operon, both of which are up-regulated in the early shock response (PF2036 and PF2037). Both PF0666 ORF and PF2036 ORF have no apparent functional connection with low temperature response as of yet.

Three down-regulated ORFs were chosen for EMSA and footprinting analysis with the PF1072 protein: PF0347, PF0897.1n and PF0947. PF0347 shares a 195-bp intergenic region with a divergently transcribed ORF, PF0346; both ORFs were down-regulated about 1.5-fold. PF0347 is annotated as a conserved hypothetical protein and PF0346 is annotated as a ferredoxin oxidoreductase. PF0897.1n encodes a hypothetical protein, and it is the start of a four-ORF operon, three ORFs of which are down-regulated up to 3.5-fold with low temperature

(PF0895-0897). This ORF was of interest also because there were two PF1072 protein binding motifs present in its promoter. PF0947 is annotated as a conserved hypothetical protein. A BLAST of the PF0947 sequence indicates this protein may function as a permease.

Table 3.6 Promoter-ORFs used in validation of binding sites in the *P. furiosus* genome

UOR[a]	Motif	Start[b]	Stop[b]	ORF Annotation[c]	Fold
PF0346	GTACAAAAATTTGTAG	-141	-126	aldehyde:ferredoxin oxidoreductase	-1.41
PF0347	GTACAAAAATTTGTAG	-69	-54	hypothetical protein	-1.59
PF0666	GTAAATTTAATTATAC	-172	-157	nol1-nop2-sun family nucleolar protein IV	1.26
PF0897.1n	ATACCTCCACCTGTAC	-142	-127	hypothetical protein	-1.01
PF0897.1n	GCACATCTTTGTGTAG	-192	-177	hypothetical protein	-1.01
PF0897				hypothetical protein	-2.36
PF0896				hypothetical protein	-1.65
PF0895				ABC transporter (ATP-binding protein)	-3.54
PF0947	GAACAGACTAATGTAC	-198	-183	hypothetical protein	-2.26
PF2036	GTCCAGAACGTTGTAC	-136	-121	magnesium and cobalt transporter	1.45[d]
PF2037				hypothetical protein	2.08[d]

[a] UOR (Upstream of ORF Region) designation corresponds to the locus of the ORF from which the upstream sequence was taken.
[b] Start and stop positions are relative to the UOR sequence where -1 corresponds to the first nucleotide upstream from the ORF start.
[c] Protein annotations are derived from the NCBI database.
[d] PF2036 & PF2037 were up-regulated 2.4 & 2.8 fold respectively in the early shock response.

3.5.2 Confirmation of specific PF1072 protein binding by EMSA

EMSA was used to test whether the PF1072 protein bound these selected promoter-ORF DNAs. It was confirmed previously that the PF1072 protein bound to its own promoter. There were four distinct protein-DNA complexes evident in the range of protein-DNA ratios tested. The PF1072 probe DNA was shifted completely at a protein/DNA mole ratio of around 24. EMSA confirmed that the PF1072 protein bound to all promoter-ORF DNAs that were tested, as can be seen in Figure 3.21 and 3.22. However, there was only one well-defined protein-DNA complex formed with these probes except with the PF0897.1n promoter. It seems that there is a correlation between the number of protein-DNA complexes observed in EMSA and the number of binding

motifs present in the promoter region. Only one protein-DNA complex was observed with PF0347, PF0666, PF0947 and PF2036 promoters, while there were two with PF0897.1n promoter because of the presence of two binding motifs.

For the promoter-ORF probes of the genes PF0347, PF0666, PF0947 and PF2036, the number of protein-DNA complexes observable from the addition of protein to a ratio of 16 was one. The EMSA for the PF0666 promoter was not carried out with high enough protein-DNA ratios to cause a complete shift of the DNA. Approximately half of the free DNA is shifted at a ratio of 24, and therefore it is reasonable to assume that a complete shift might occur at a ratio of 48. Considering there are two mismatches in the binding motif, it makes sense that the PF0666 promoter was shifted completely at a higher protein-DNA ratio compared to other probes. The probe of the PF0897.1n displayed two distinct protein-DNA complexes. It is worth noting that these protein-DNA complexes have very low electrophoretic mobility. Although the mobility is similar to nonspecific protein-DNA complexes, these bands are well defined, indicating that they are specific protein-DNA complexes with similar volume to the nonspecific complexes. It is also worth noting that the two protein-DNA complexes formed simultaneously. The second complex with the PF0897.1n promoter appeared on the gel at a low protein/DNA ratio of 2, while the second complex formed with the PF1072 promoter was not observable until the protein/DNA ratio increased to 4. This result suggests that the protein binding affinities of the two motifs in the PF0897.1n promoter region are similar to each other.

3.5.3 PF1072 protein footprints are present on each EMSA-tested probe

In order to verify the location of the PF1072 protein binding sites on these various promoter-ORF DNAs, fluorescence-based DNase I footprinting was performed. Probes were purified by PAGE prior to use in the footprinting reaction. Footprints were detected on all

promoters tested, and the results for one strand of each probe are shown in Figures 3.23 through 3.27. The footprint size, together with the EMSA data, provide a picture of the DNA-binding pattern in relation to sequence elements and the footprint positions suggest the likely mode of regulation imposed by the PF1072 protein on transcription of each ORF.

The $GTACAn_6TGTAC$ palindrome was discovered in each footprint (Fig. 3.28). The size of the footprint on each probe is ~26 bases, and the 16-base palindrome was located very near the middle of the footprint with ~5 extra bases on both ends. The locations of the footprints, relative to the translation start sites, varied among the promoter DNAs. The PF0666 probe footprint began at -157 (Fig. 3.23) while the PF2036 probe footprint started at position -121 from the ORF start (Fig. 3.24). The PF1072 protein binding site on the PF0347 probe was located at -54 from the ORF start (Fig. 3.25). The PF0947 probe footprint spanned 26 bases starting far upstream from the ORF at position -183 (Fig. 3.26), and the two footprints on the PF897.1n probe started at -127 and -177 relative to the ORF start (Fig. 3.27). For these probes, the footprints were positioned farther from the ORF start than for the PF1072 probe.

The position of the protein binding site is important in dictating its role in transcriptional regulation. Repressors tend to overlap either the BRE/TATA elements or the transcription start site while activators tend to bind upstream from the BRE/TATA elements [50]. The location of the PF1072 protein footprint relative to the core promoter elements will give a clue as to its possible mechanism of regulation for the corresponding ORF. Although the transcription start sites of these ORFs are unknown, they can be deduced from the positions of BRE/TATA elements, since the spacing between them is critical for productive recruitment of RNAP to the transcription start site. The TATA box usually locates at approximately -25 from the transcription start site, with the BRE located directly upstream [23].

A recent analysis of characterized glycolytic gene promoters of *Pyrococcus* species determined the BRE/TATA consensus sequences to be VRAAANN/TTWWWAW [22]. Based on this conserved sequence, positions of these essential promoter motifs upstream of a given ORF can be deduced. For the down-regulated ORFs, possible positions for their BRE/TATA boxes can be located downstream of the PF1072 protein footprints except for two ORFs (PF0195 and PF1123) whose BRE/TATA boxes cannot be found in the promoter region. For the up-regulated PF1072 ORF, the putative BRE/TATA box appears to be positioned upstream of the footprints. This configuration suggests that the PF1072 protein would repress transcription from its own ORF by either blocking the binding of TBP and TFB to the TATA and BRE elements or by obstructing recruitment of RNAP to the transcription initiation site. Table 3.7 summarizes the sequence information for the known and predicted BRE/TATA boxes for ORFs containing PF1072 protein footprints.

Table 3.7 Predicted BRE/TATA boxes for cold-regulated PF1072 protein target ORFs

UOR[a]	Footprint position	BRE/TATA position	BRE (VRAAANN)[b]	TATA (TTWWWAW)[b]	Fold
PF0195	-125 to -110	N/A	N/A	N/A	-2.21
PF0289	-87 to -72	-50 to -37	AACAAAA	TTAATAA	-5.53
PF0758	-144 to -129	-47 to -34	CAAAATG	TTTAATT	-3.12
PF0947	-198 to -183	-50 to -37	AAAAATC	ATTTAAA	-2.26
PF0989	-64 to -49	-47 to -34	AAAAGAA	TTTTTAA	-2.21
PF1072	-42 to -27	-75 to -62	ATAAATT	TTTATAT	12.71
PF1072	-27 to -12	-75 to -62	ATAAATT	TTTATAT	12.71
PF1123	-97 to -82	N/A	N/A	N/A	-2.52
PF1513	-59 to -44	-36 to -23	CAAAAGA	TTTTTAA	-2.93
PF1832	-198 to -183	-49 to -36	CAAATAA	TTTTTAA	-2.26

[a] UOR (Upstream of ORF Region) designation corresponds to the locus of the ORF from which the upstream sequence was taken.
[b] Ambiguous nucleotides are represented according to the IUPAC code as follows: W = T/A, Y = C/T, R = A/G, V = A/C/G, N = any base. Nucleotides which deviate from the proposed consensus are shown in red.

3.6 PF1072 protein is a transcriptional activator and repressor

In order to validate the PF1072 protein as a functional regulatory transcription factor, its direct effect on basal transcription was tested. An *in vitro* transcription system for *P. furiosus* has been established by Michael Thomm's research group at the University of Regensburg, Germany [90, 91]. Through collaboration with this group, the *in vitro* transcriptional regulation profile of PF1072 protein was determined (by Antonia Gindner) for two ORFs: PF0947, which is down-regulated with low temperature according to DNA microarray and PF1072, which is up-regulated with low temperature. ORFs both up- and down-regulated with low temperature were selected for *in vitro* transcription experiments to determine if PF1072 protein could function as both an activator and repressor, as was suggested by the DNase I footprinting data. The same pUC18-cloned probe DNAs used for EMSA and footprinting were provided to Antonia Gindner for use as *in vitro* transcription templates, along with recombinant his-tagged PF1072 protein.

The PF1072 probe covered the promoter region of PF1072 ORF. Addition of PF1072 protein completely repressed transcription from its own ORF (Fig. 3.29). Repression of PF1072 transcription may occur through PF1072 protein blocking the binding of TBP and TFB or blocking the recruitment of RNAP. The PF1072 protein footprint covers about 50 bases between the translation start site and BRE/TATA, so it is likely that it overlaps with the transcription start site of its own gene.

Transcription from the PF0947 promoter template showed an increase in the PF0947 transcript with increasing PF1072 protein concentration (Fig. 3.30). The basal level of transcription for this ORF appeared to be a little higher, compared to the PF1072 promoter. Since the PF1072 protein binding site lies upstream from the BRE/TATA box of the PF0947 ORF, it is

reasonable to hypothesize that the PF1072 protein may activate transcription by facilitating recruitment of TBP.

The *in vitro* transcription results provided by Antonia Gindner showed that PF1072 protein can both activate and repress transcription, depending on the gene context.

3.7 The PF1072 protein is a dimer in cell extract

We performed a protein-protein pull-down experiment to discover any protein partners of the PF1072 protein. His-tagged PF1072 protein was bound to Ni-NTA magnetic agarose beads. The bead-bound protein was then mixed with *P. furiosus* soluble cell extract from cells grown either before or after cold shock and incubated for 1 h, and the same amount of his-tagged PF1072 protein was incubated without cell extract as a control. Then proteins were eluted and analyzed by SDS-PAGE, followed by in-gel tryptic digestion and mass spectrometry identification (Fig. 3.31).

Four protein bands were identified from this experiment (Fig. 3.32). For the his-tagged PF1072 protein incubated with the cell extract before or 1 h after cold shock, bands 1 and 2 were identified as his-tagged and wild type PF1072 protein, respectively. The migration on the gel suggests that these are both dimeric, unusual for SDS-PAGE. Bands 3 and 4 were his-tagged and wild type PF1072 protein also, presumably the monomer. For the protein incubated with cell extract 5 h after cold shock, only the wild type PF1072 protein monomer was identified. It is worth noting that not only that the his-tagged protein did not catch the dimer from the cell extract 5 h after cold shock, but also that the band was weaker than the control while the band with cell extract before or 1 h after cold shock were much stronger. This result suggests that the concentration of PF1072 protein in cells decreased significantly 5 h after cold shock.

It is also worth noting that the bands are different when the incubation time is changed. For the experiment in which the PF1072 protein was incubated with cell extracts for 20 min (Fig. 3.32A), the bands from cell extract 5 h after cold shock were as strong as the without-extract control. On the other hand, for the experiment in which the PF1072 protein was incubated with cell extract for 60 min (Fig. 3.32B), the bands from cell extract 5 h after cold shock were much weaker than the control. This result suggests that more PF1072 protein is destroyed by cell extract 5 h after cold shock when the incubation time is extended.

It is interesting that the dimer of the PF1072 protein appeared on the SDS-PAGE gel. The protein sample was denatured at 100 °C for 10 min before loading onto the gel. Therefore, all quaternary structure of the protein should have been destroyed. A covalent disulfide bond between cysteines on two monomers can result in a dimer stable enough to survive the denaturation procedure [92]. However, PF1072 protein does not contain any cysteine. In the absence of cell extract, the PF1072 protein dimer formation was not observed on SDS-PAGE, even though the concentration was ten times higher than that used in the protein-protein pull-down experiment.

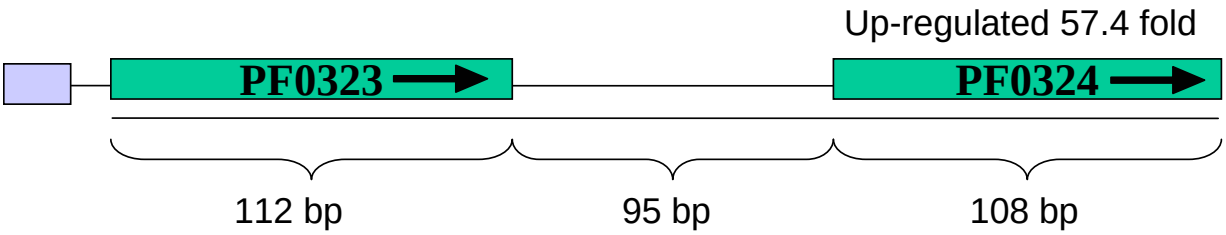
Analytical gel filtration was also used to determine the oligomeric state of the PF1072 protein. 2 mg of protein sample was loaded onto an analytical gel filtration column with a 70 kDa exclusion limit. The corresponding molecular weight was calculated for each eluted peak, using a calibration curve generated from a set of standards (Fig. 3.33). The calculated molecular weight of the peak is approximately 38 kDa. According to the PF1072 protein amino acid sequence, the molecular weights for his-tagged monomer, dimer, and trimer are 17.913, 35.826 and 53.739 kDa, respectively. These results suggest that PF1072 protein is a dimer in solution.

3.8 The PF1072 protein tends to dissociate from DNA in the presence of metal ions

Using magnesium, cobalt and zinc, as representative metals, the metal ions responsive DNA-binding affinity of the PF1072 protein was investigated by EMSA. It is anticipated that a small amount of imidazole contained in the protein solution will compete for metal ion binding with the PF1072 protein. Thus, imidazole molecules were carefully removed from the protein solution by gel filtration. Figure 3.34 shows EMSA results of the PF1072 protein in the absence and in the presence of metal ions. The results indicate that metal ions decrease the affinity between the PF1072 protein and its own promoter. The decrease of DNA-binding affinity for metal ion-bound PF1072 protein is thought to occur as a result of overlap between putative metal ion binding site and the DNA binding domain. The middle region (residues 55-72) of PF1072 protein sequence is similar to the dimer interface of the SmtB repressor [89], and residues Ser77 and Glu80 were predicted to be putative metal ion binding ligands while the residues from 55 to 100 might be involved in DNA interactions. Thus, metal ions might inhibit the binding between protein and DNA through blocking the putative DNA binding domain directly or preventing the dimerization. Structural information of the DNA/metal ion-bound form of the PF1072 protein would help to understand the mechanism behind this result.

Figure 3.1 Probes used in DNA affinity protein capture.

The DNA probe for the PF0324 promoter region was designed to contain 207 bp upstream and 108 bp downstream of the ORF start site. The DNA probe for the PF1072 promoter region was designed to contain 176 bp of DNA upstream from translation start site, along with 93 bp downstream of the ORF start. The position of the biotin group is represented by a square.


Up-regulated 57.4 fold
PF0323
PF0324
112 bp
95 bp
108 bp

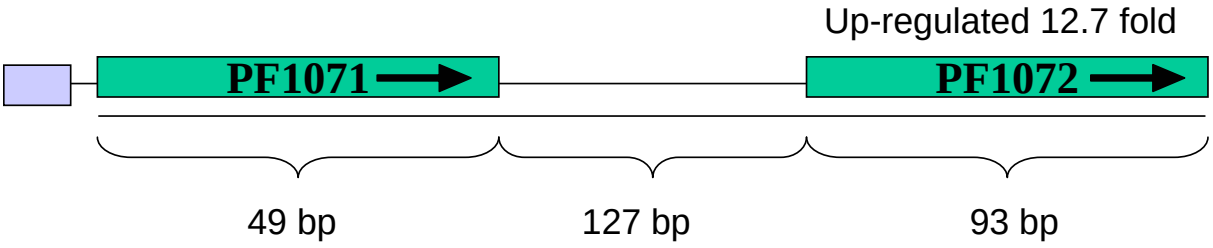
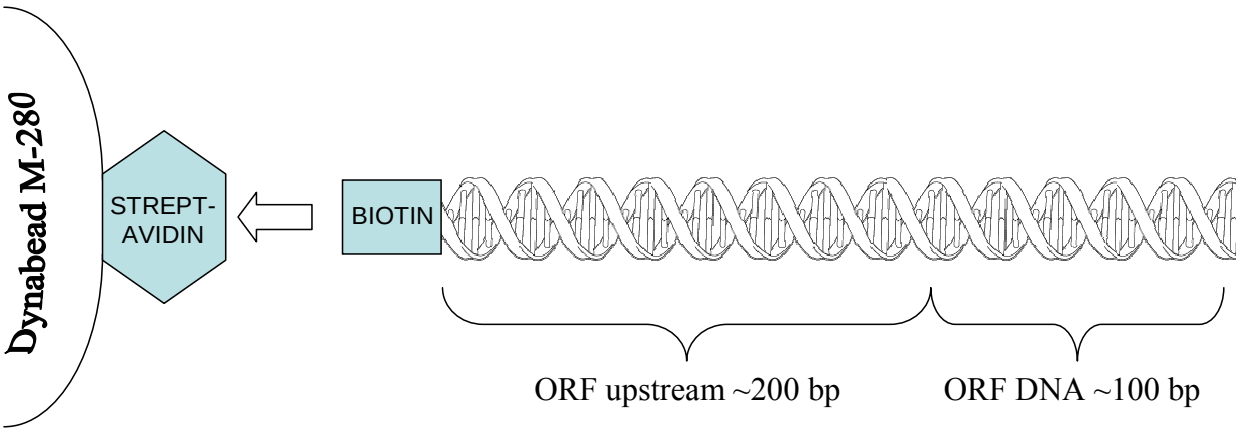

Up-regulated 12.7 fold
PF1071
PF1072
49 bp
127 bp
93 bp

Figure 3.2 Design and immobilization of probe DNA.

A. The probe DNA is designed to cover approximately 200 bp upstream and 100 bp downstream from the target ORF start. Probe DNA is amplified from genomic DNA using one biotinylated primer and one unlabeled primer such that the PCR-amplified probe contains a biotin group on one 5' end so that the DNA can be bound to streptavidin-coated magnetic beads (Dynabeads M-280 Streptavidin, Invitrogen, Carlsbad, CA). **B.** The magnetic properties of the beads allow them, and correspondingly whatever is attached to them, to be easily separated from solution with the use of a magnet.

A


Dynabead M-280
STREPT-
AVIDIN
BIOTIN
ORF upstream ~200 bp
ORF DNA ~100 bp

B

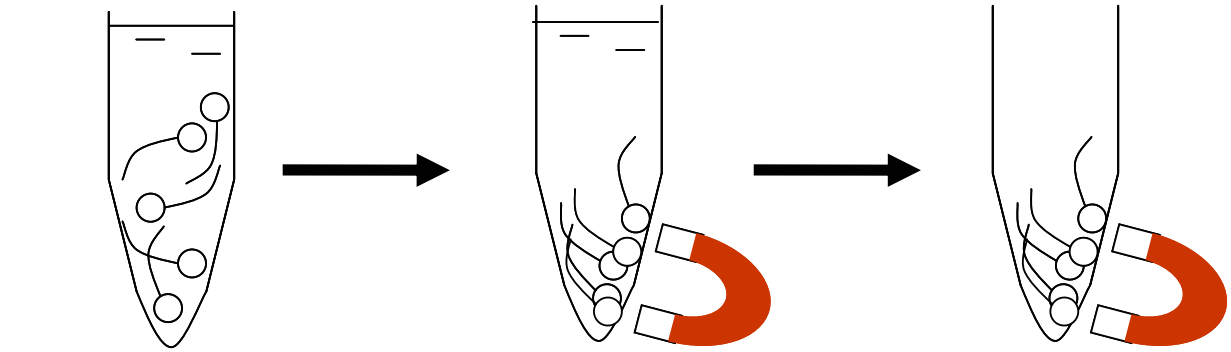

Magnetic beads are suspended in solution
Beads are separated from solution
Supernatant is removed

Figure 3.3 The DNA affinity protein capture experiment.

Biotinylated DNA is bound to magnetic streptavidin-coated beads. The bead-DNA complex is then incubated with soluble cell extract, and some proteins associate with the DNA: general transcriptional machinery (TBP, green; TFB, orange; RNAP, blue), non-specific DNA-binding proteins (pink and cyan), and other transcription factors (brown). Proteins which do not bind DNA are removed, and finally the DNA-binding proteins which remain are eluted and analyzed.

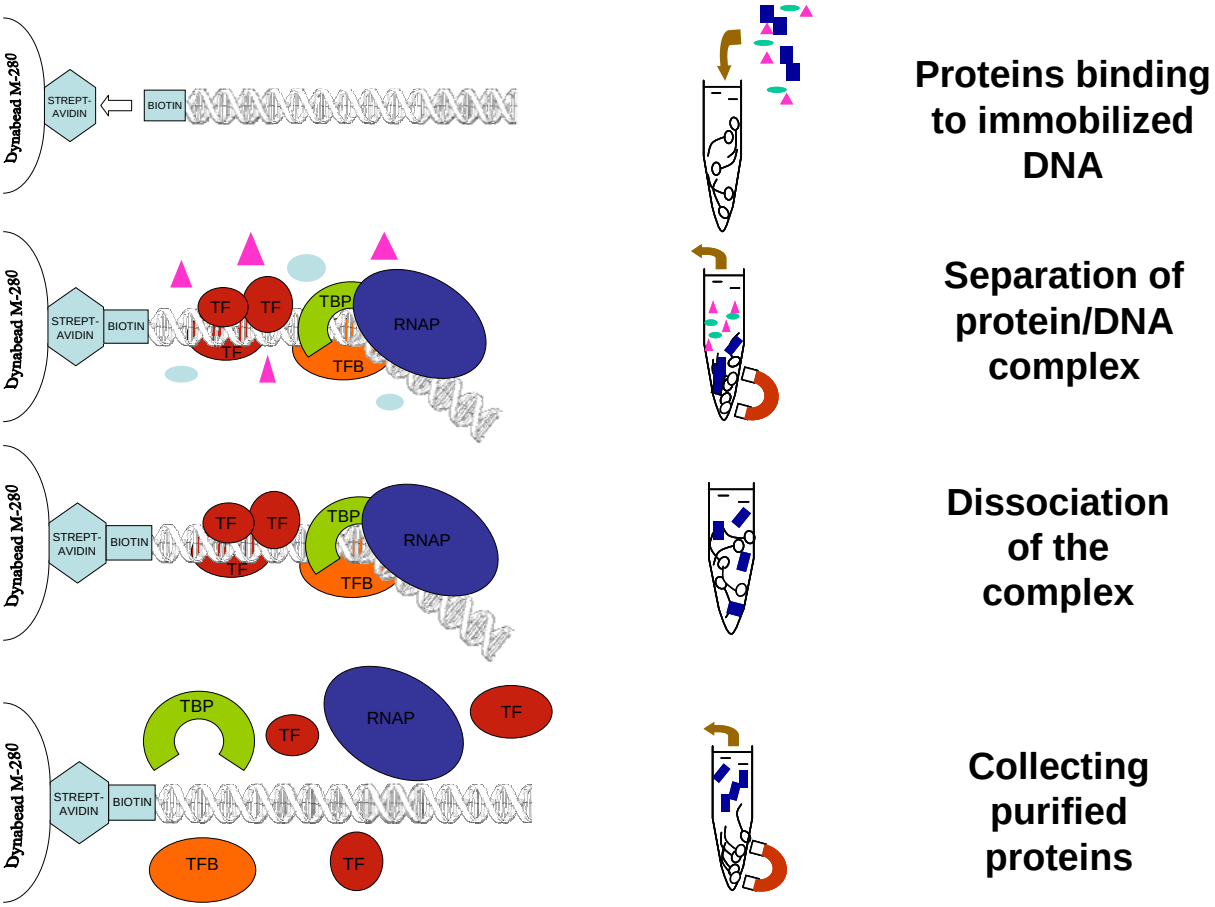

Dynabead M-280
STREPT-AVIDIN
BIOTIN
TF
TBP
RNAP
TFB
Proteins binding to immobilized DNA
Separation of protein/DNA complex
Dissociation of the complex
Collecting purified proteins

Figure 3.4 DNA affinity protein capture with PF0324 and PF1072 probes.

Silver-stained denaturing gel of eluted proteins from DNA affinity capture with PF0324 and PF1072 probes incubated with soluble cell extracts from cells grown before (black) and after (red) cold shock with the corresponding densitometric intensities. Arrows indicate identified proteins (Table 3.2)

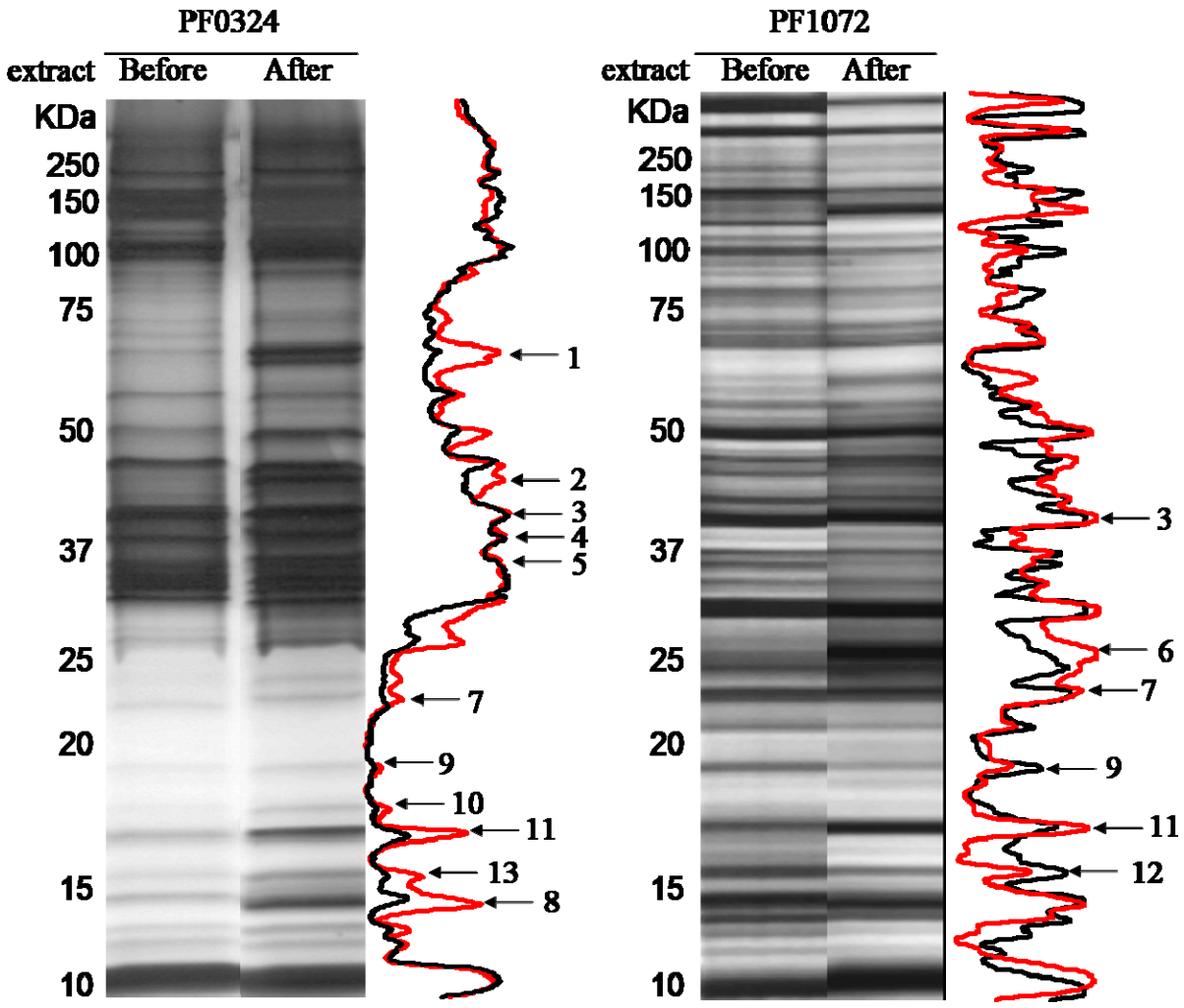
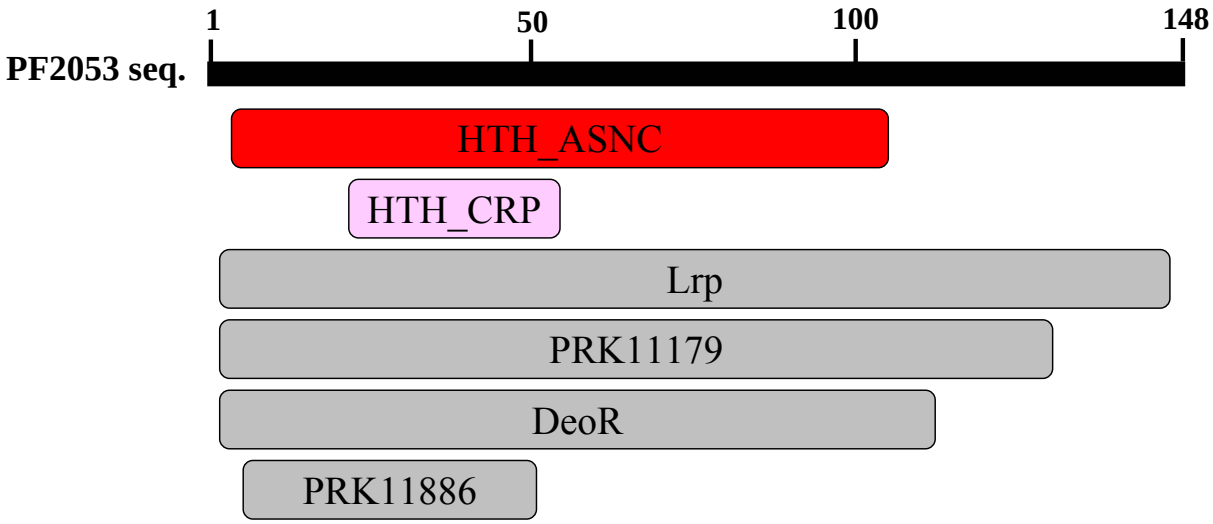

PF0324
extract
Before
After
KDa
250
150
100
75
50
37
25
20
15
10
1
2
3
4
5
7
9
10
11
13
8
PF1072
extract
Before
After
KDa
250
150
100
75
50
37
25
20
15
10
3
6
7
9
11
12

Figure 3.5 PF2053 protein sequence analysis.

Conserved domain search results for PF2053 protein sequence from online tools available at NCBI [73, 93]. PF2053 sequence is represented by a black line with matching conserved domains indicated below. Conserved domain descriptions and sequence alignments are shown. For the sequence alignments, identical residues are colored red and similar residues are colored blue.



smart00344, HTH_ASNC, helix_turn_helix ASNC type

CD Length: 104 **Bit Score:** 71.31 **E-value:** 7e-14

```
                    10        20        30        40        50        60        70        80
                    ....*.... |....*.... |....*.... |....*.... |....*.... |....*.... |....*.... |....*.... |
PF2053          4 MDKVDLQLIKILSQNSRLTYRELAEMLGTTRQRVARKVDKLKKLGIIRKFTIIPNLEKLNY-MYAILLIKVKATENIYQV 82
smart00344      1 LDEIDRKILEELQKDARISLAELAKKVGLSPSTVHNRVKRLEEEGVIKGYTAVLNPKKLGLsVTAFVGVTLESPDKLEEF 80

                    90        100
                    ....*.... |....*.... |....
PF2053         83 AKVLKDHEDVKILELGVGKYNIIA 106
smart00344     81 LEKLAKLPEVVEVYLVTGDYDYLL 104
```

smart00419, HTH_CRP, helix_turn_helix, cAMP Regulatory protein

CD Length: 49 **Bit Score:** 34.61 **E-value:** 0.007

```
                              10        20        30
                              ....*.... |....*.... |....*.... |...
PF2053         20 RLTYRELAEMLGTTRQRVARKVDKLKKLGIIRK 52
smart00419      8 PLTRQEIAELLGLTRETVSRTLKRLEKEGLISI 40
```

COG1522, Lrp, Transcriptional regulators

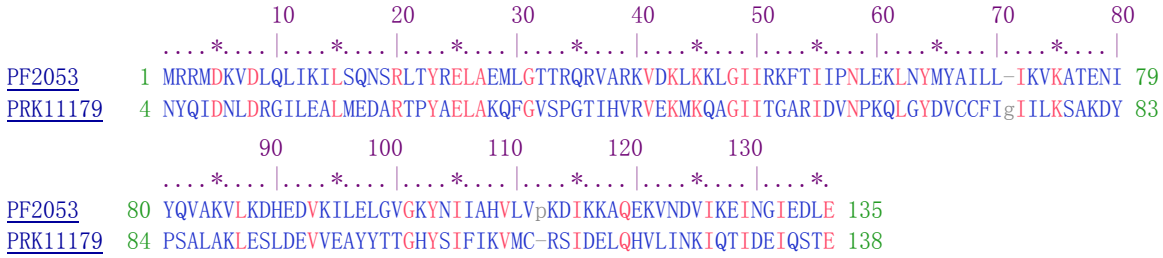
CD Length: 154 **Bit Score:** 78.56 **E-value:** 4e-16

```
                    10        20        30        40        50        60        70        80
                    ....*.... |....*.... |....*.... |....*.... |....*.... |....*.... |....*.... |....*.... |
PF2053          1 MRRMDKVDLQLIKILSQNSRLTYRELAEMLGTTRQRVARKVDKLKKLGIIRKFTIIPNLEKLNYMY-AILLIKVKAT-EN 78
COG1522         3 MMKLDDIDRRILRLLQEDARISNAELAERVGLSPSTVLRRIKRLEEEGVIKGYTAVLDPEKLGLDLtAFVEVKLERSlED 82

                    90        100       110       120       130       140       150
                    ....*.... |....*.... |....*.... |....*.... |....*.... |....*.... |....*.... |
PF2053         79 IYQVAKVLKDHEDVKILELGVGKYNIIAHVLVpKDIKKAQEKVNDVIKEINGIEDLEVEFVSDIPKFELL 148
COG1522        83 LEEFAEALAKLPEVVECYRVTGDYDYLLKVRV-RDLEELERFLGELIRAIPGVESTETLVVLETVKDTTR 151
```

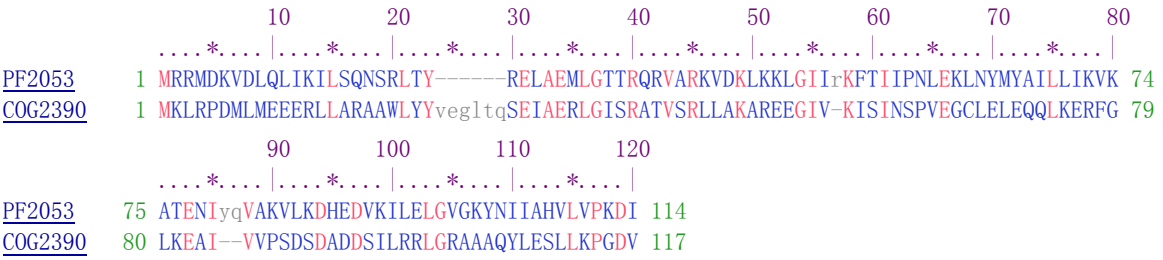
PRK11179, PRK11179, DNA-binding transcriptional regulator AsnC

CD Length: 153 **Bit Score:** 47.23 **E-value:** 1e-06



COG2390, DeoR, Transcriptional regulator

CD Length: 321 **Bit Score:** 35.56 **E-value:** 0.005



PRK11886, PRK11886, biotin--protein ligase

CD Length: 319 **Bit Score:** 35.12 **E-value:** 0.005

Figure 3.6 PF2053 protein secondary structure prediction.

PSIPRED [94, 95] secondary structure prediction of PF2053 protein. Helices are indicated as green rods and strands are indicated as yellow arrows.

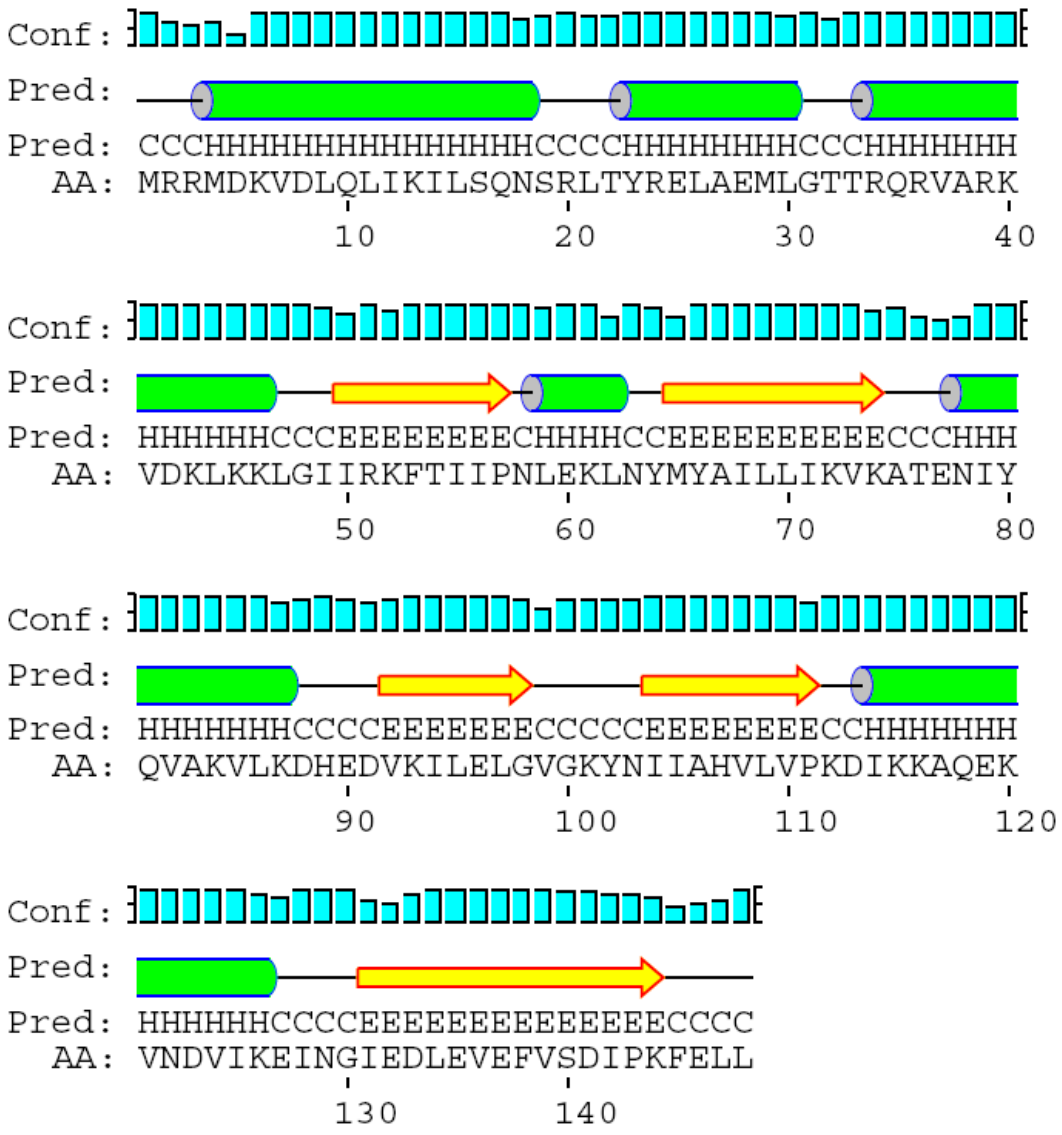

Conf:
Pred:
Pred: CCCHHHHHHHHHHHHHHHCCCCHHHHHHHHCCCHHHHHHH
AA: MRRMDKVDLQLIKILSQNSRLTYRELAEMLGTTRQRVARK
10
20
30
40
Conf:
Pred:
Pred: HHHHHHCCCEEEEEEECHHHHCCEEEEEEEEECCCHHH
AA: VDKLKKLGIIRKFTIIPNLEKLNYMYAILLIKVKATENIY
50
60
70
80
Conf:
Pred:
Pred: HHHHHHHCCCCEEEEEEECCCCCEEEEEEECCHHHHHHH
AA: QVAKVLKDHEDVKILELGVGKYNIIAHVLVPKDIKKAQEK
90
100
110
120
Conf:
Pred:
Pred: HHHHHHCCCCEEEEEEEEEEEEECCCC
AA: VNDVIKEINGIEDLEVEFVSDIPKFELL
130
140
Legend:
= helix
Conf:
= confidence of prediction
-
+
= strand
Pred: predicted secondary structure
= coil
AA: target sequence

Figure 3.7 PF2053 protein sequence BLAST.

Tree-view for results of a BLAST search of the PF2053 protein sequence against the NCBI non-redundant database [71]. PF2053 is shown boxed in red, proteins of the Thermococcaceae family are highlighted in yellow, and homologs from *P. furiosus* are underlined (PF1231 and PF1543).

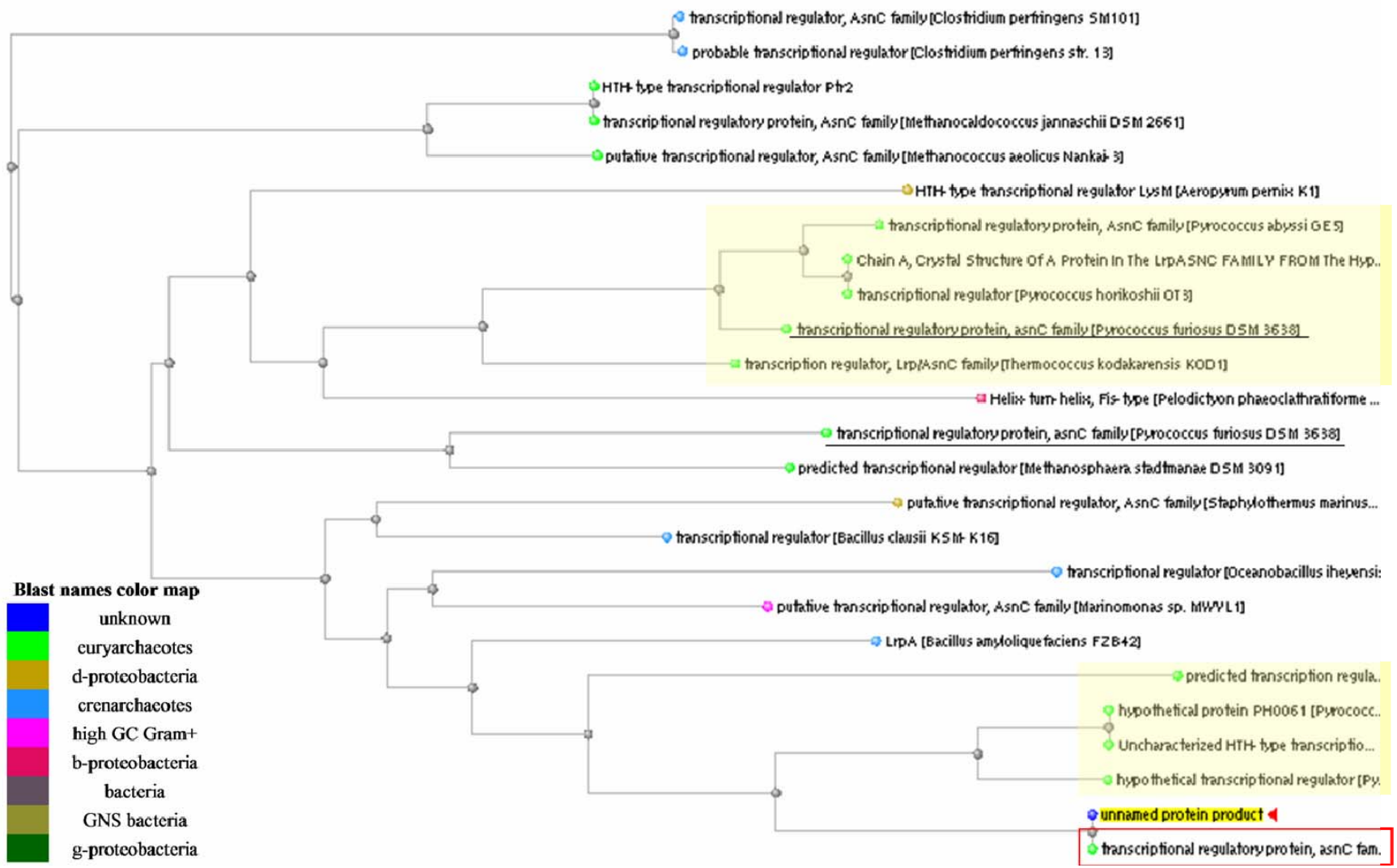
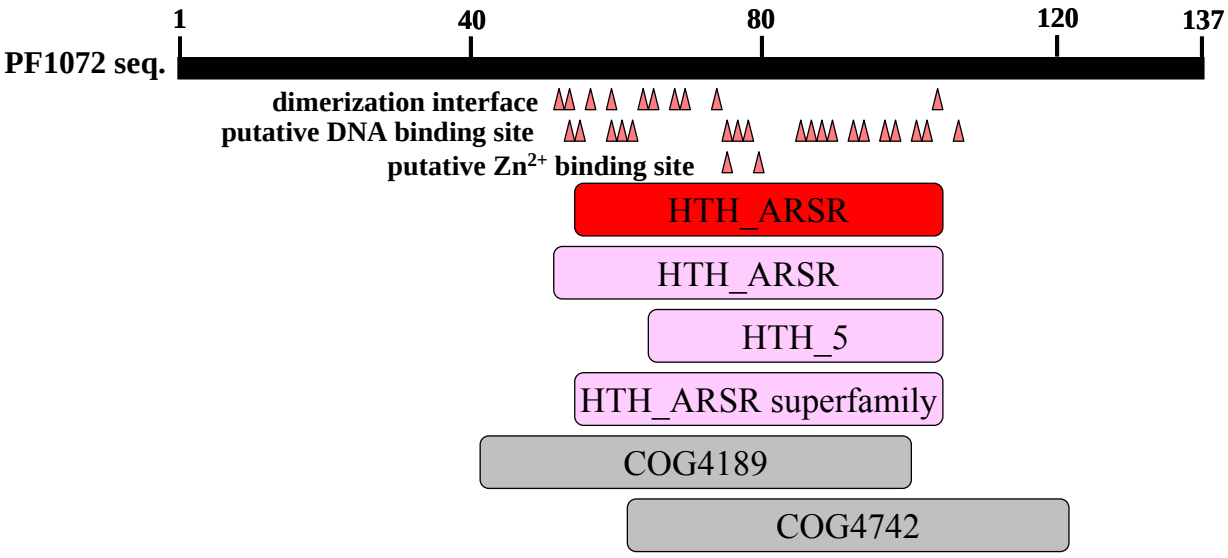

transcriptional regulator, AsnC family [Clostridium perfringens SM101]
probable transcriptional regulator [Clostridium perfringens str. 13]
HTH-type transcriptional regulator Ptr2
transcriptional regulatory protein, AsnC family [Methanocaldococcus jannaschii DSM 2661]
putative transcriptional regulator, AsnC family [Methanococcus aeolicus Nankai-3]
HTH-type transcriptional regulator LysM [Aeropyrum pernix K1]
transcriptional regulatory protein, AsnC family [Pyrococcus abyssi GE5]
Chain A, Crystal Structure Of A Protein In The LrpASNC FAMILY FROM The Hyp..
transcriptional regulator [Pyrococcus horikoshii OT3]
transcriptional regulatory protein, asnC family [Pyrococcus furiosus DSM 3638]
transcription regulator, Lrp/AsnC family [Thermococcus kodakarensis KOD1]
Helix-turn-helix, Fis-type [Pelodictyon phaeoclathratiforme ...
transcriptional regulatory protein, asnC family [Pyrococcus furiosus DSM 3638]
predicted transcriptional regulator [Methanosphaera stadtmanae DSM 3091]
putative transcriptional regulator, AsnC family [Staphylothermus marinus...
transcriptional regulator [Bacillus clausii KSM-K16]
transcriptional regulator [Oceanobacillus iheyensis
putative transcriptional regulator, AsnC family [Marinomonas sp. MWYL1]
LrpA [Bacillus amyloliquefaciens FZB42]
predicted transcription regula..
hypothetical protein PH0061 [Pyrococc..
Uncharacterized HTH-type transcriptio...
hypothetical transcriptional regulator [Py.
unnamed protein product
transcriptional regulatory protein, asnC fam.
Blast names color map
unknown
curyarchaeotes
d-proteobacteria
crenarchaeotes
high GC Gram+
b-proteobacteria
bacteria
GNS bacteria
g-proteobacteria

Figure 3.8 PF1072 protein sequence analysis.

Conserved domain search results for PF1072 protein sequence from online tools available at NCBI [73, 93]. PF1072 sequence is represented by a black line with matching conserved domains indicated below. The pink triangle indicates putative sites for dimerization, DNA binding and Zn^{2+} binding. Conserved domain descriptions and sequence alignments are shown. For the sequence alignments, identical residues are colored red and similar residues are colored blue. The pound indicates putative sites for dimerization, DNA binding and Zn^{2+} binding aligned with SmtB [89].




```
dimerization interface
               ## #     #  ##  ##  #                                      #
1SMT_B      40 FAVL.[2].PNRLRLLSLLARS.[1].LCVGDLAQAIG.[1].SESAVSHQLRSLRNLRLVSY.[25]. 116
PF1072      55 LKAL.[2].PDRIRIMKMLSER.[1].MSFKEIKEALG.[1].ESPTVSHHLKILTKTKMVRK.[31]. 137


putative DNA binding site
               ##   ###             ###        #### ##  ## ##     ###       ###
1SMT_B      40 FAVLadPNRLRLLSLLARSe-LCVGDLAQAIGvSESAVSHQLRSLRNLRLVSYRKqg---rhVYYQLqd 104
PF1072      55 LKALanPDRIRIMKMLSERp-MSFKEIKEALGvESPTVSHHLKILTKTKMVRKgekyeitqdgmlflri 122

putative Zn2+ binding site
                                      #  #
1SMT_B      40 FAVL.[2].PNRLRLLSLLARS.[1].LCVGDLAQAIG.[1].SESAVSHQLRSLRNLRLVSY.[25]. 116
PF1072      55 LKAL.[2].PDRIRIMKMLSER.[1].MSFKEIKEALG.[1].ESPTVSHHLKILTKTKMVRK.[31]. 137
```

cd00090, HTH_ARSR, Arsenical Resistance Operon Repressor and similar prokaryotic

CD Length: 78 **Bit Score:** 52.65 **E-value:** 2e-08

```
                10        20        30        40        50
            ....*....|....*....|....*....|....*....|....*....|..
PF1072   55 LKALANPDRIRIMKMLSERPMSFKEIKEALGVESPTVSHHLKILTKTKMVRK 106
cd00090   1 LKALSDPTRLRILRLLLEGPLTVSELAERLGLSQSTVSRHLKKLEEAGLVES 52
```

smart00418, HTH_ARSR, helix_turn_helix, Arsenical Resistance Operon Repressor

CD Length: 79 **Bit Score:** 59.08 **E-value:** 3e-10

```
                10        20        30        40        50
            ....*....|....*....|....*....|....*....|....*....|....
PF1072   53 KVLKALANPDRIRIMKMLSERPMSFKEIKEALGVESPTVSHHLKILTKTKMVRK 106
smart00418  1 EILKALSDPTRLKILKLLAEGELSVCELAEILGLSQSTVSHHLKKLREAGLVES 54
```

pfam01022, HTH_5, Bacterial regulatory protein, arsR family

CD Length: 47 **Bit Score:** 48.42 **E-value:** 5e-07

```
                10        20        30        40
            ....*....|....*....|....*....|....*....|....*..
PF1072   60 NPDRIRIMKMLSERPMSFKEIKEALGVESPTVSHHLKILTKTKMVRK 106
pfam01022  1 DPTRLKILYLLSEGELCVCELAEILGLSQSTVSHHLKKLREAGLVEK 47
```

COG4189, Predicted transcriptional regulator

CD Length: 308 **Bit Score:** 48.43 **E-value:** 5e-07

```
                10        20        30        40        50        60
            ....*....|....*....|....*....|....*....|....*....|....*....|....
PF1072   43 ISTINEENAAKVLKALANPDRIRIMKMLSER-PMSFKEIKEALGVESPTVSHHLKILTKTKMVR 105
COG4189   5 ILTVDPNESLDVLKALASKVRVAILQLLHRKgPLNVNEIAEALGLPQSTMSANIKVLEKAGLIR 68
```

COG4742, Predicted transcriptional regulator

CD Length: 260 **Bit Score:** 45.15 **E-value:** 4e-06

```
                10        20        30        40        50
            ....*....|....*....|....*....|....*....|....*....|....*....
PF1072   63 RIRIMKMLSERPMSFKEIKEALGVESPTVSHHLKILTKTKMV-RKGEKYEITQDGMLFL 120
COG4742  15 RKDLLLLLKEGPKTIEEIKNELNVSSSAILPQIKKLKDKGLVvQEGDRYSLSSLGKIIV 73
```

Figure 3.9 PF1072 protein secondary structure prediction.

PSIPRED [94, 95] secondary structure prediction of PF1072 protein. Helices are indicated as green rods and strands are indicated as yellow arrows.

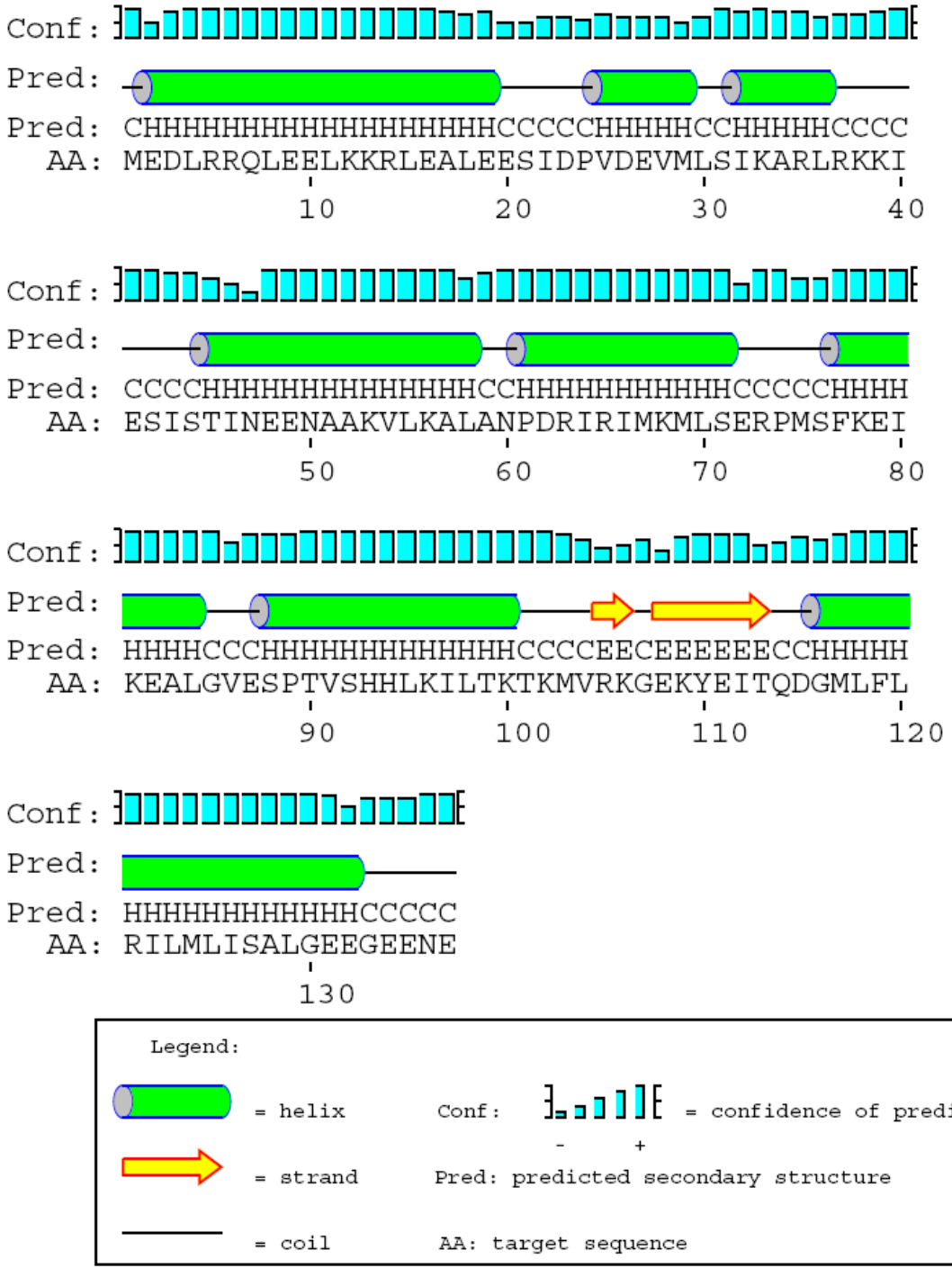

Conf:
Pred:
Pred: CHHHHHHHHHHHHHHHHHHCCCCCHHHHHCCHHHHHCCCC
AA: MEDLRRQLEELKKRLEALEESIDPVDEVMLSIKARLRKKI
10
20
30
40
Conf:
Pred:
Pred: CCCCHHHHHHHHHHHHHHCCHHHHHHHHHHHCCCCCHHHH
AA: ESISTINEENAAKVLKALANPDRIRIMKMLSERPMSFKEI
50
60
70
80
Conf:
Pred:
Pred: HHHHCCCHHHHHHHHHHHHCCCCEECEEEEECCHHHHH
AA: KEALGVESPTVSHHLKILTKTKMVRKGEKYEITQDGMLFL
90
100
110
120
Conf:
Pred:
Pred: HHHHHHHHHHHCCCCC
AA: RILMLISALGEEGEENE
130
Legend:
= helix
= strand
= coil
Conf:
- +
= confidence of prediction
Pred: predicted secondary structure
AA: target sequence

Figure 3.10 PF1072 protein sequence BLAST.

Tree-view for results of a BLAST search of the PF1072 protein sequence against the NCBI non-redundant database [71]. PF1072 is shown boxed in red, proteins of the Thermococcaceae family are highlighted in yellow.

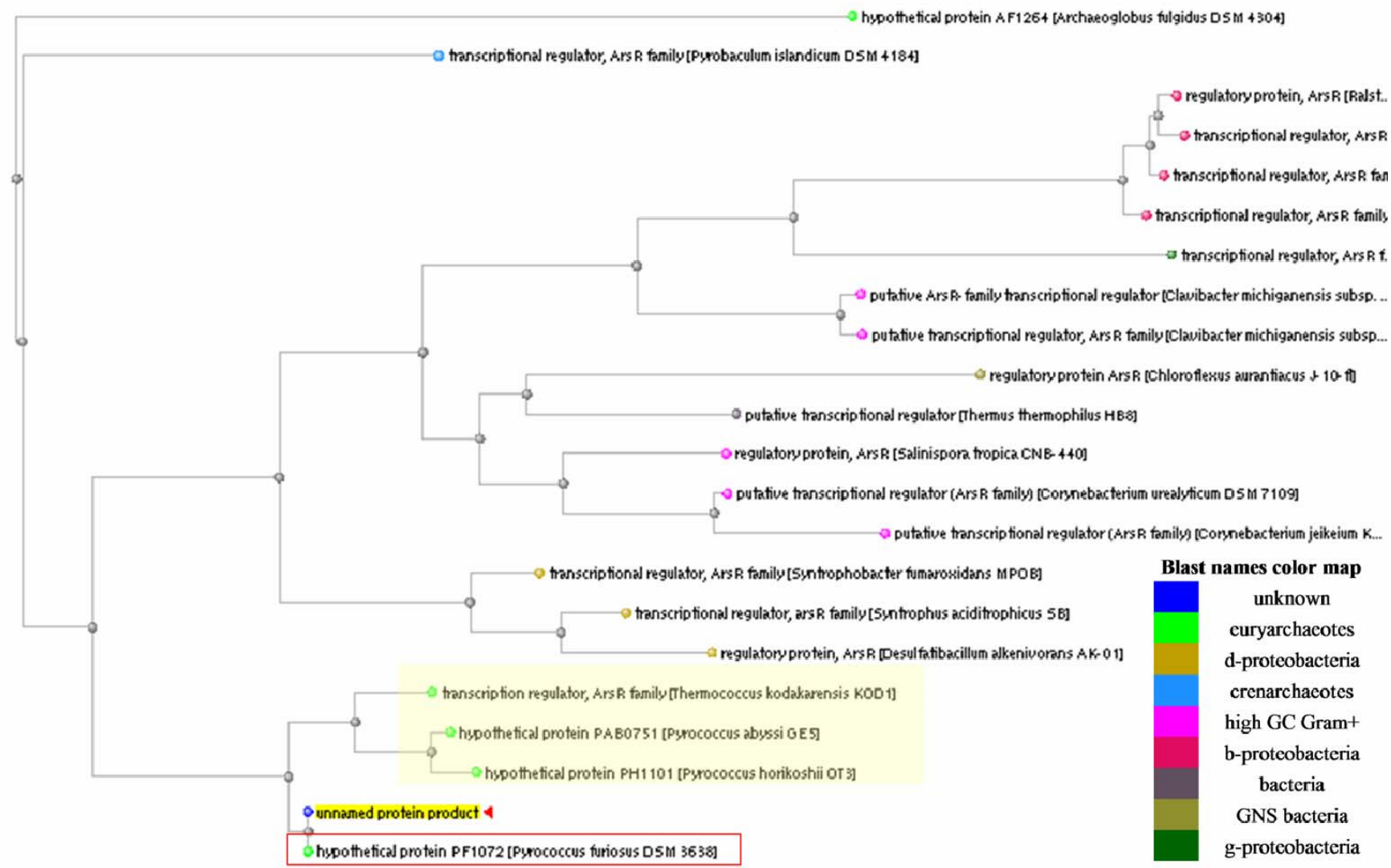

hypothetical protein AF1264 [Archaeoglobus fulgidus DSM 4304]
transcriptional regulator, ArsR family [Pyrobaculum islandicum DSM 4184]
regulatory protein, ArsR [Ralst..
transcriptional regulator, ArsR
transcriptional regulator, ArsR fa
transcriptional regulator, ArsR family
transcriptional regulator, ArsR f.
putative ArsR-family transcriptional regulator [Clavibacter michiganensis subsp. ...
putative transcriptional regulator, ArsR family [Clavibacter michiganensis subsp...
regulatory protein ArsR [Chloroflexus aurantiacus J-10-fl]
putative transcriptional regulator [Thermus thermophilus HB8]
regulatory protein, ArsR [Salinispora tropica CNB-440]
putative transcriptional regulator (ArsR family) [Corynebacterium urealyticum DSM 7109]
putative transcriptional regulator (ArsR family) [Corynebacterium jeikeium K...
transcriptional regulator, ArsR family [Syntrophobacter fumaroxidans MPOB]
transcriptional regulator, arsR family [Syntrophus aciditrophicus SB]
regulatory protein, ArsR [Desulfatibacillum alkenivorans AK-01]
transcription regulator, ArsR family [Thermococcus kodakarensis KOD1]
hypothetical protein PAB0751 [Pyrococcus abyssi GE5]
hypothetical protein PH1101 [Pyrococcus horikoshii OT3]
unnamed protein product
hypothetical protein PF1072 [Pyrococcus furiosus DSM 3638]
Blast names color map
unknown
curyarchaeotes
d-proteobacteria
crenarchaeotes
high GC Gram+
b-proteobacteria
bacteria
GNS bacteria
g-proteobacteria

Figure 3.11 Expression and purification of PF1072 protein.

SDS-PAGE of PF1072 protein samples at each time point after induction. Protein concentrations were normalized according to OD_{600} (left). The final purified protein after nickel-affinity purification is shown on the right.

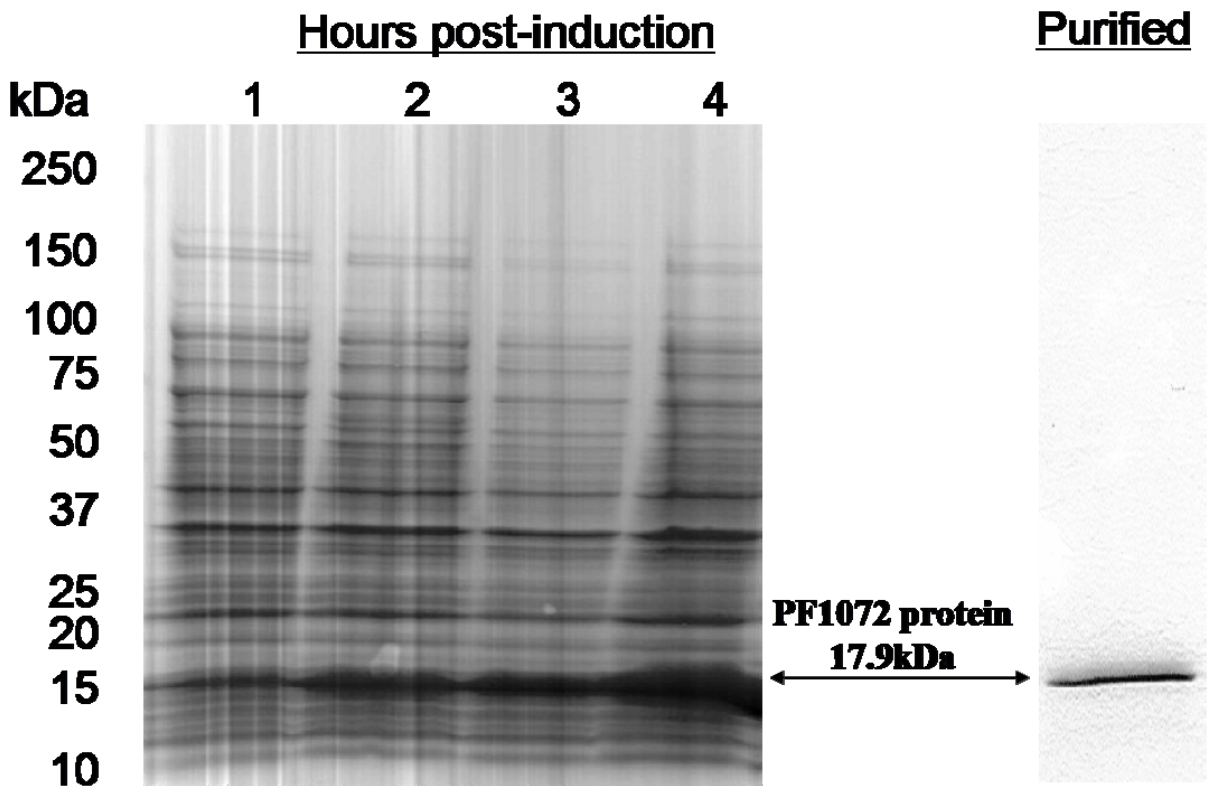

Hours post-induction
Purified
kDa
1
2
3
4
250
150
100
75
50
37
25
20
15
10
PF1072 protein
17.9kDa

Figure 3.12 Expression and purification of PF2053 protein.

SDS-PAGE of PF2053 protein samples at each time point after induction. Protein concentrations were normalized according to OD_{600} (left). The final purified protein after nickel-affinity purification is shown on the right.

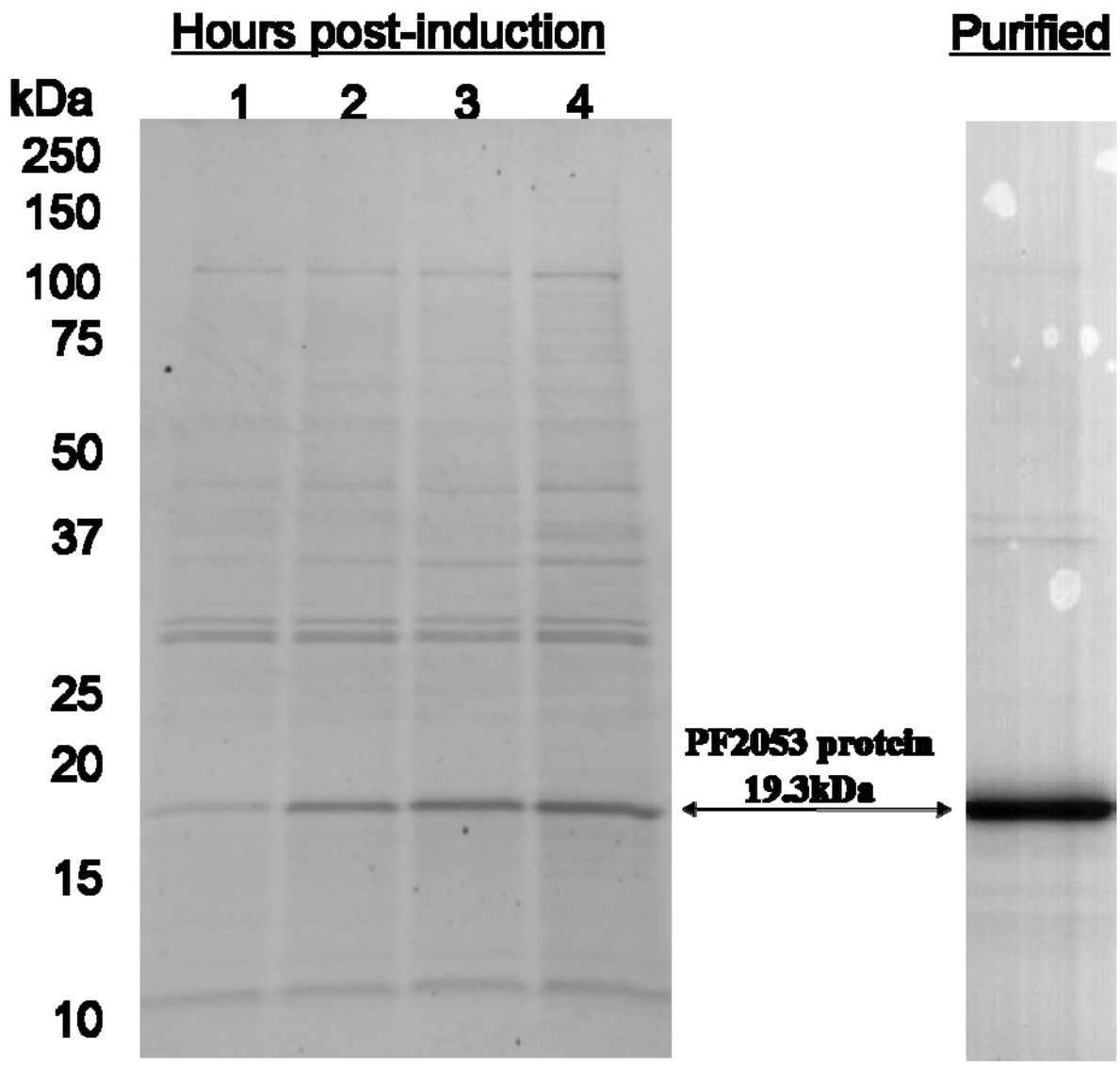
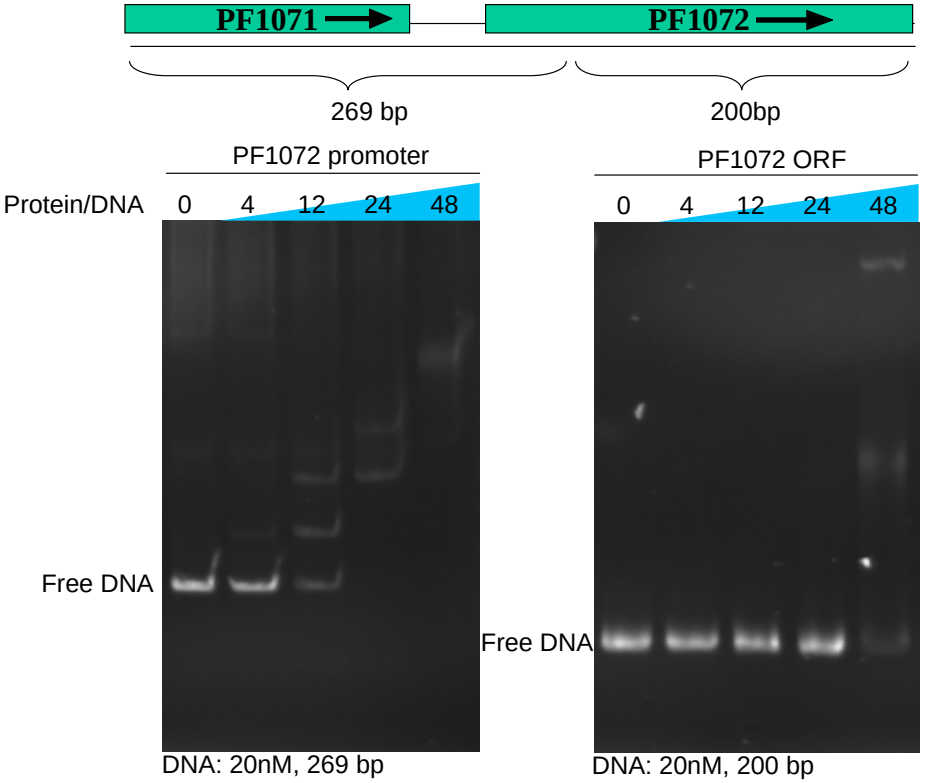

Hours post-induction
Purified
kDa
1
2
3
4
250
150
100
75
50
37
25
20
15
10
PF2053 protein
19.3kDa

Figure 3.13 PF1072 protein binds to its own promoter DNA.

Diagrams of the DNA probes are shown at the top. The probes used are indicated at the top of each gel image with corresponding protein/DNA mole ratios listed above each lane. DNA (20 nM) was incubated with protein in buffer (50 mM Tris-HCl, pH 7.5; 1 mM EDTA; 0.1 M KCl; 5% glycerol) for 20 min at 55 °C. The gel was stained with SYBR Green I nucleic acid gel stain.


PF1071
PF1072
269 bp
200bp
PF1072 promoter
PF1072 ORF
Protein/DNA
0
4
12
24
48
0
4
12
24
48
Free DNA
Free DNA
DNA: 20nM, 269 bp
DNA: 20nM, 200 bp

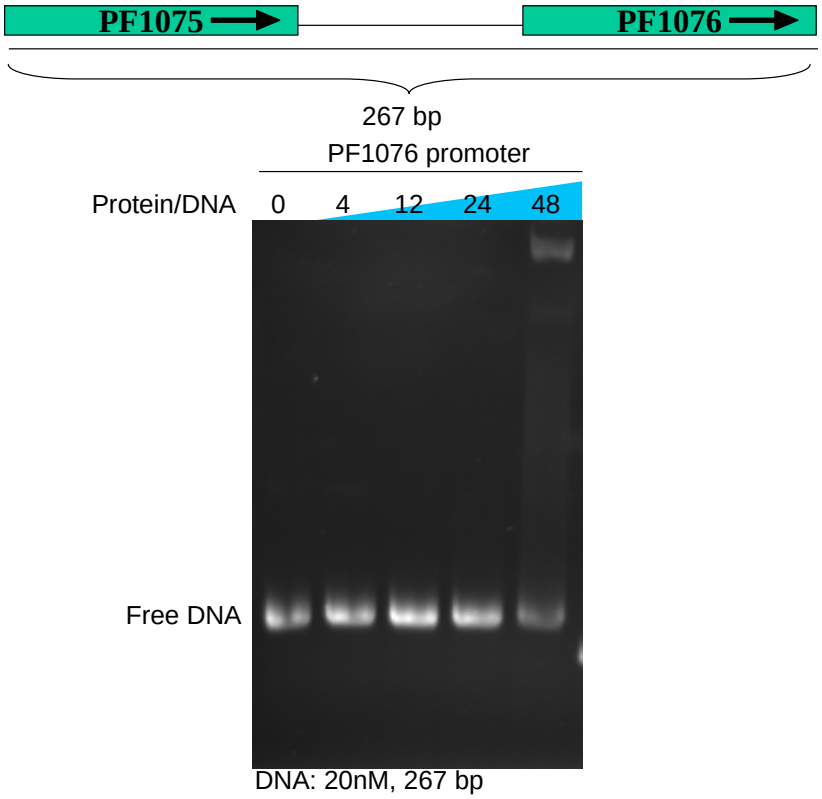
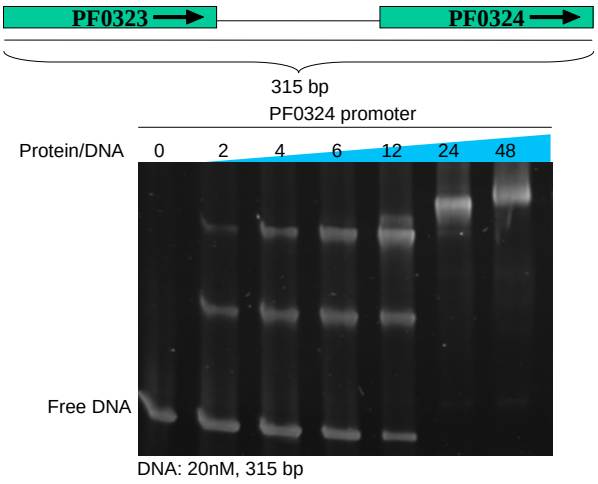
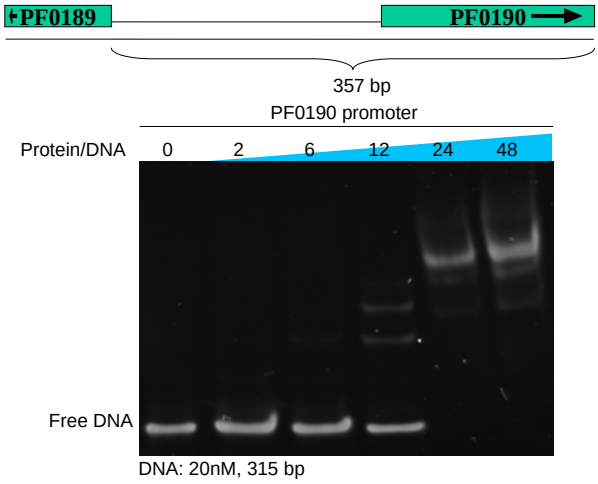
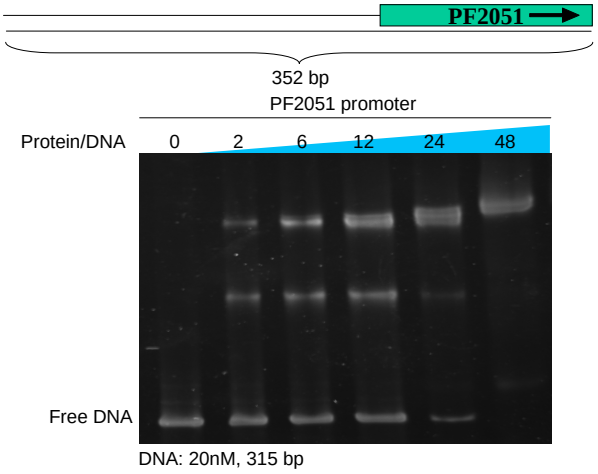

PF1075
PF1076
267 bp
PF1076 promoter
Protein/DNA
0
4
12
24
48
Free DNA
DNA: 20nM, 267 bp

Figure 3.14 PF2053 protein binds to three promoter fragments.

Diagrams of the DNA probes are shown at the top. The probes used are indicated at the top of each gel image with corresponding protein/DNA mole ratios listed above each lane. DNA (20 nM) was incubated with protein in buffer (50 mM Tris-HCl, pH 7.5; 1 mM EDTA; 0.1 M KCl; 5% glycerol) for 20 min at 55 °C. The gel was stained with SYBR Green I nucleic acid gel stain.


PF0323
PF0324
315 bp
PF0324 promoter
Protein/DNA
0
2
4
6
12
24
48
Free DNA
DNA: 20nM, 315 bp


PF0189
PF0190
357 bp
PF0190 promoter
Protein/DNA
0
2
6
12
24
48
Free DNA
DNA: 20nM, 315 bp


PF2051
352 bp
PF2051 promoter
Protein/DNA
0
2
6
12
24
48
Free DNA
DNA: 20nM, 315 bp

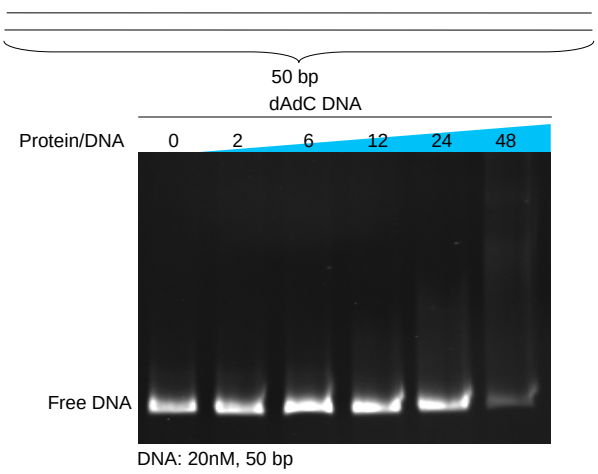
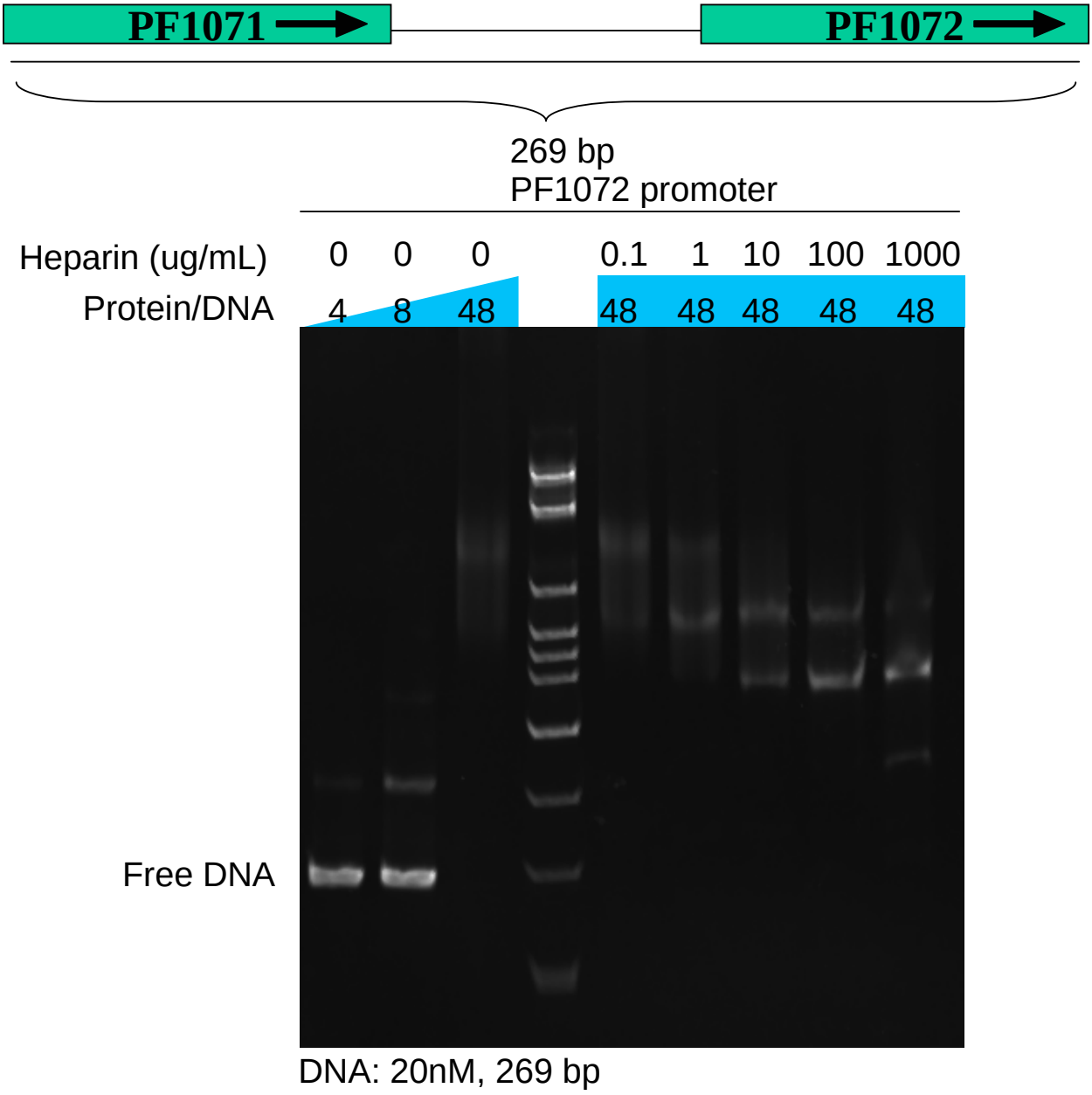

50 bp
dAdC DNA
Protein/DNA
0
2
6
12
24
48
Free DNA
DNA: 20nM, 50 bp

Figure 3.15 PF1072 protein binds specifically to its own promoter fragment in the presence of heparin.

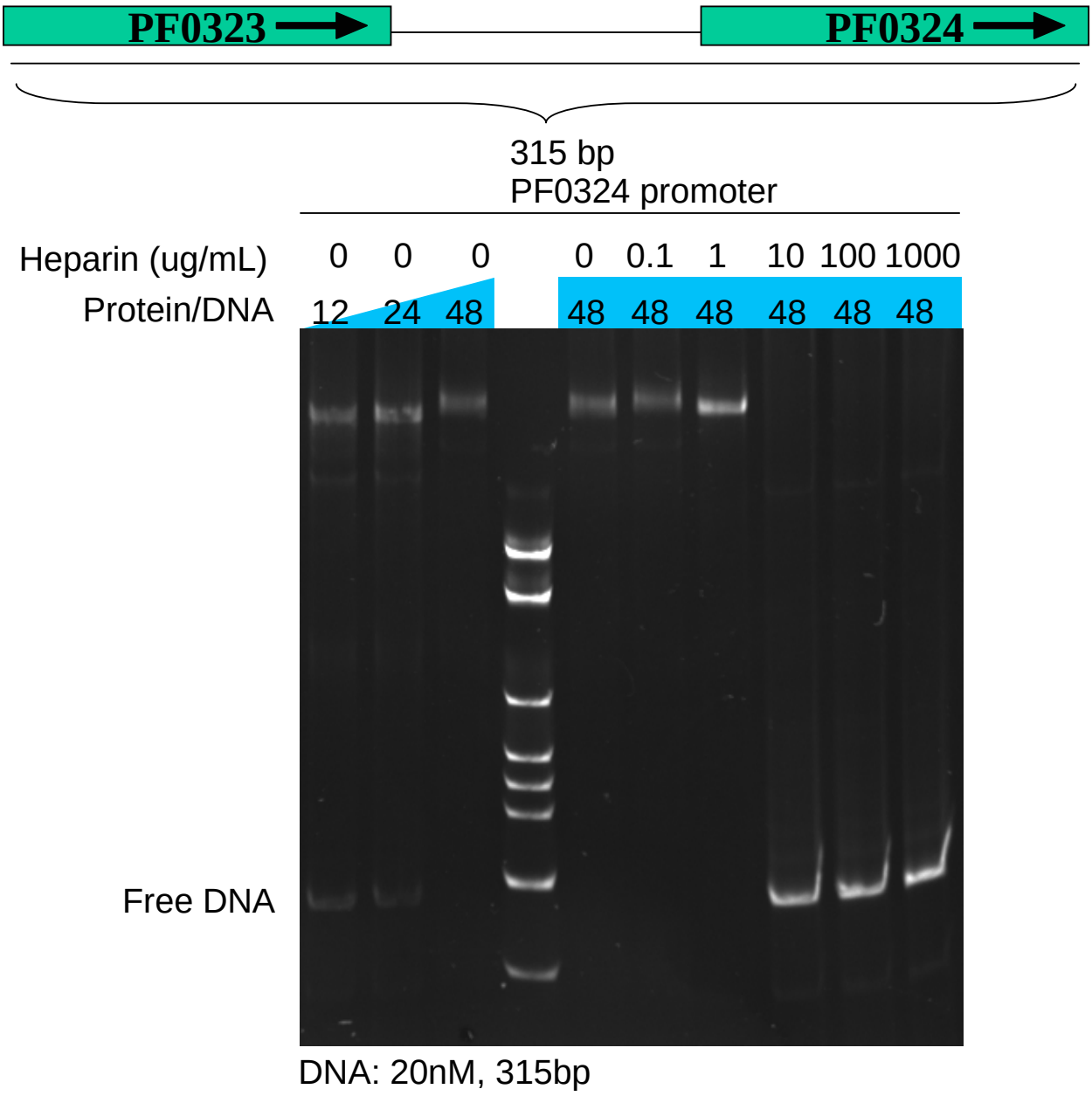
Diagram of the DNA probe is shown at the top. The probe used is indicated at the top of each gel image with corresponding protein/DNA mole ratios listed above each lane. DNA (20 nM) was incubated with protein in buffer (50 mM Tris-HCl, pH 7.5; 1 mM EDTA; 0.1 M KCl; 5% glycerol) with and without heparin (concentrations indicated) for 20 min at 55 °C. Gel was stained with SYBR Green I nucleic acid gel stain.



DNA: 20nM, 269 bp

Figure 3.16 PF2053 protein binds nonspecifically to PF0324 promoter fragment as determined by heparin challenge.

Diagram of the DNA probe is shown at the top. The probe used is indicated at the top of each gel image with corresponding protein/DNA mole ratios listed above each lane. DNA (20 nM) was incubated with protein in buffer (50 mM Tris-HCl, pH 7.5; 1 mM EDTA; 0.1 M KCl; 5% glycerol) with and without heparin (concentrations indicated) for 20 min at 55 °C. Gel was stained with SYBR Green I nucleic acid gel stain.

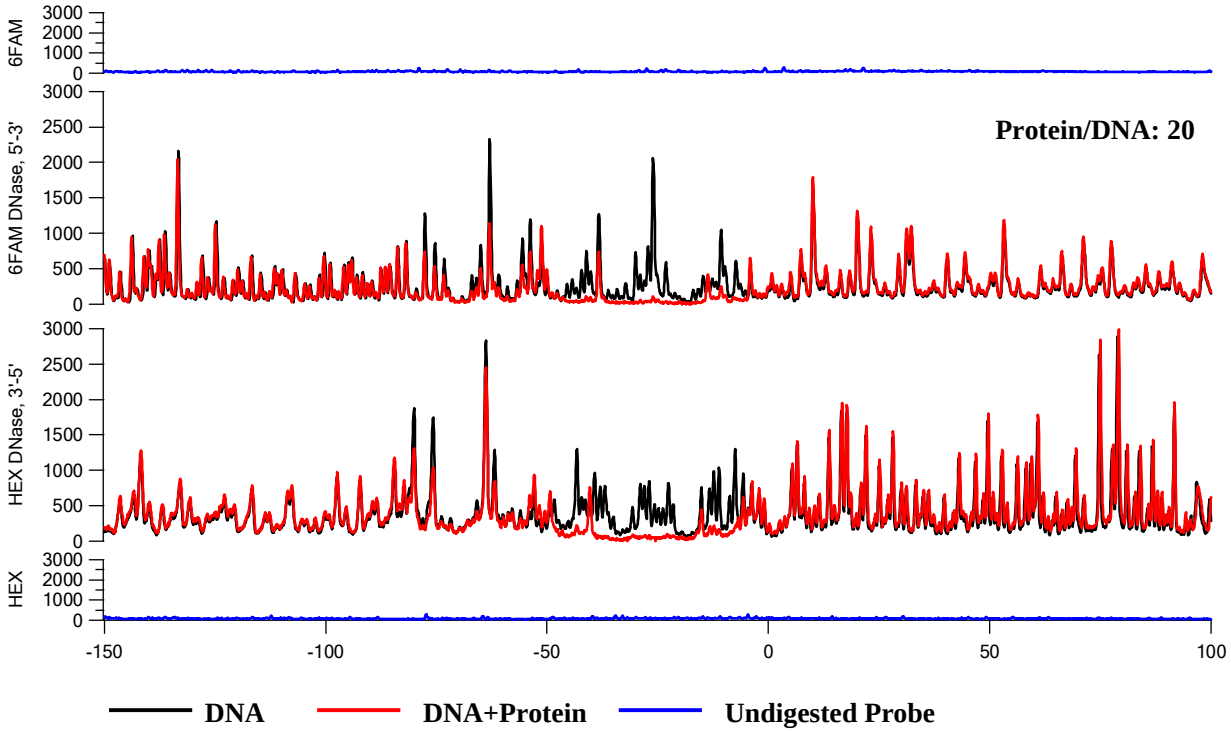


DNA: 20nM, 315bp

Figure 3.17 DNase I footprint of the PF1072 protein on the PF1072 promoter.

A. Electropherograms of 6FAM/HEX-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Numbers on bottom indicate positions relative to translation start site of the PF1072 ORF. **B.** Enlarged view of footprint region with corresponding sequence.

A


6FAM
6FAM DNase, 5'-3'
HEX DNase, 3'-5'
HEX
Protein/DNA: 20
-150
-100
-50
0
50
100
DNA
DNA+Protein
Undigested Probe

B

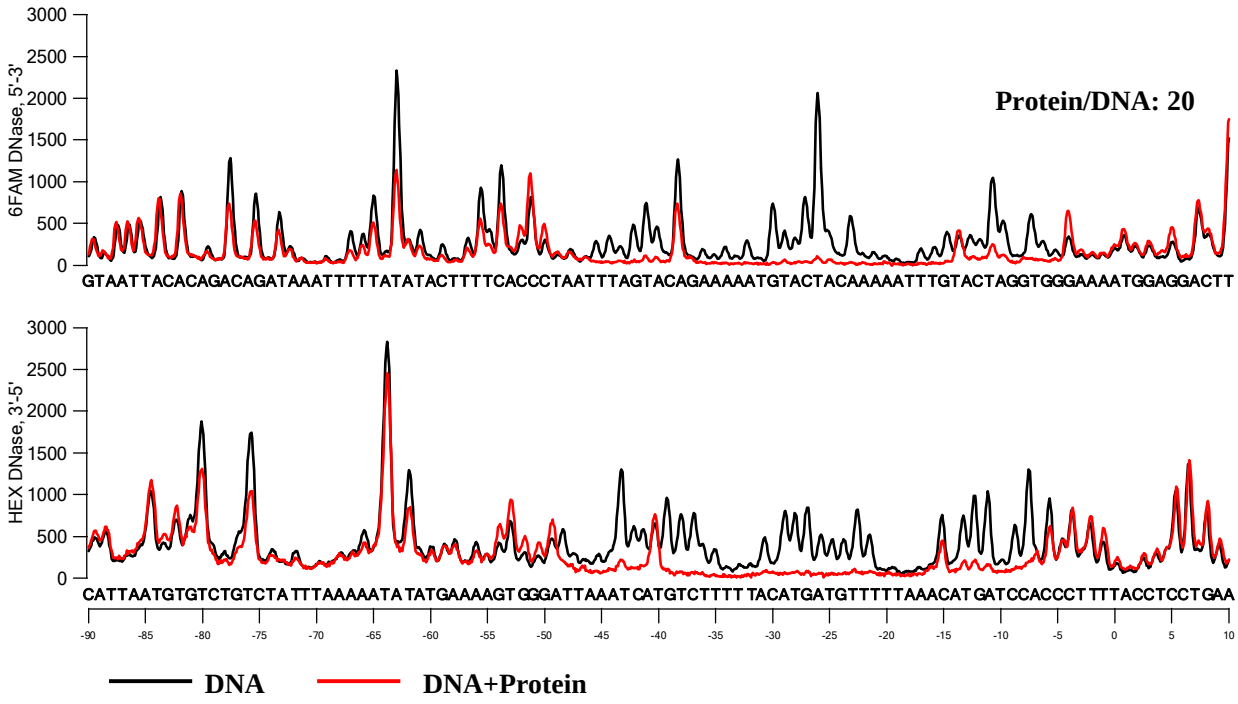
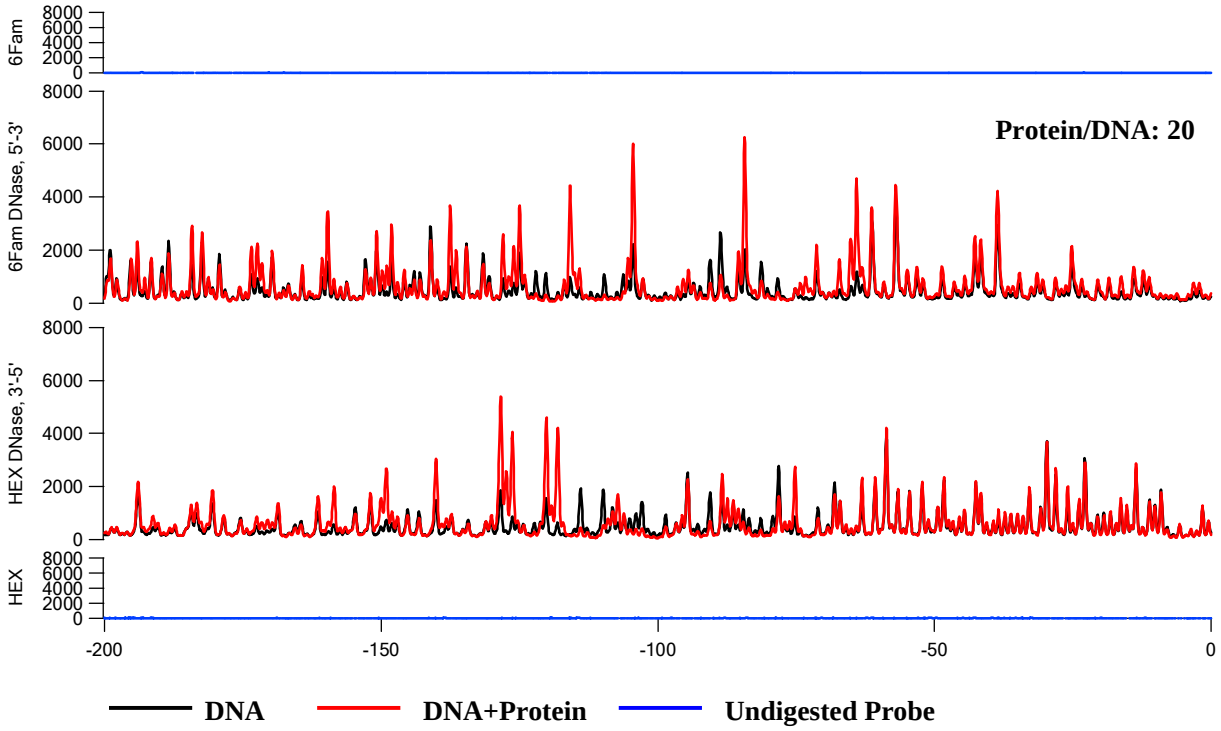

6FAM DNase, 5'-3'
HEX DNase, 3'-5'
Protein/DNA: 20
GTAATTACACAGACAGATAAATTTTTATATACTTTTCACCCTAATTTAGTACAGAAAAATGTACTACAAAAATTTGTACTAGGTGGGAAAATGGAGGACTT
CATTAATGTGTCTGTCTATTTAAAAATATATGAAAAGTGGGATTAAATCATGTCTTTTTACATGATGTTTTTAAACATGATCCACCCTTTTACCTCCTGAA
-90
-85
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5
0
5
10
DNA
DNA+Protein

Figure 3.18 DNase I footprint of the PF2053 protein on the PF0324 promoter.

A. Electropherograms of 6FAM/HEX-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Numbers on bottom indicate positions relative to translation start site of PF0324 ORF. **B.** Enlarged view of footprint region with corresponding sequence. Arrows indicate DNase I hypersensitive sites.

A


6Fam
6Fam DNase, 5'-3'
HEX DNase, 3'-5'
HEX
8000
6000
4000
2000
0
Protein/DNA: 20
-200
-150
-100
-50
0
DNA
DNA+Protein
Undigested Probe

B

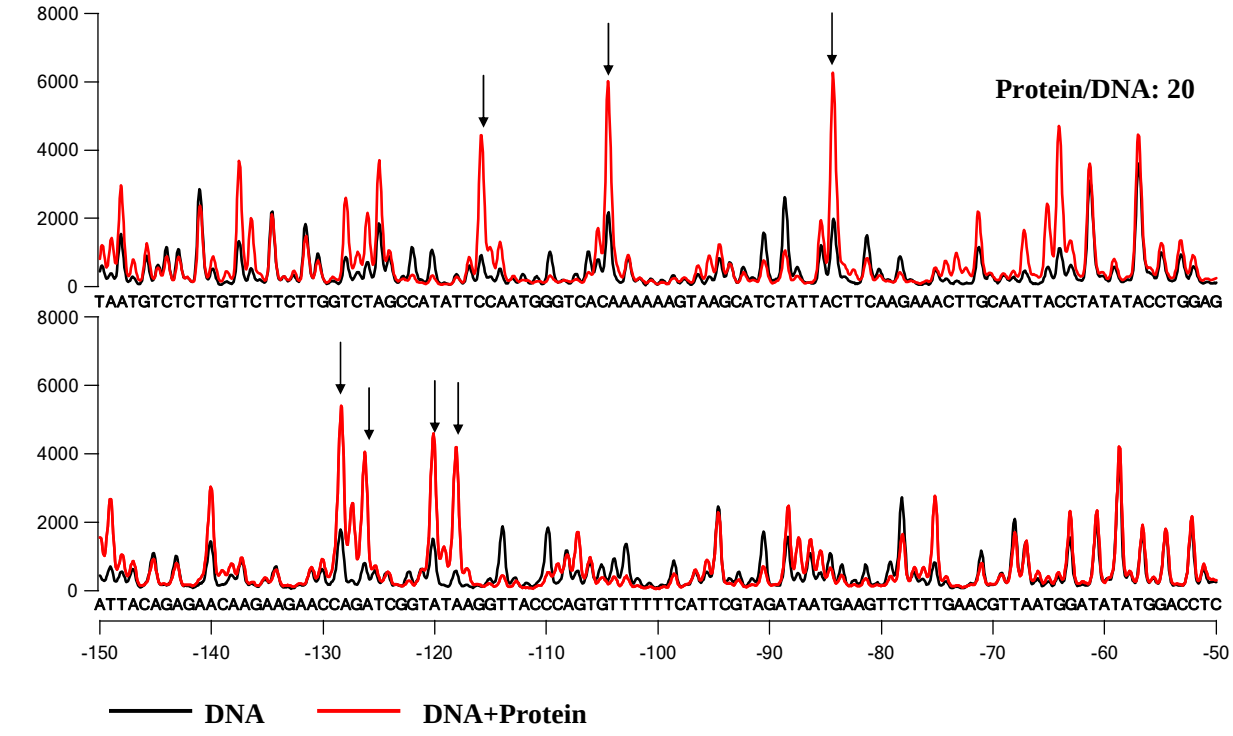

8000
6000
4000
2000
0
Protein/DNA: 20
TAATGTCTCTTGTTCTTCTTGGTCTAGCCATATTCCAATGGGTCACAAAAAAGTAAGCATCTATTACTTCAAGAAACTTGCAATTACCTATATACCTGGAG
ATTACAGAGAACAAGAAGAACCAGATCGGTATAAGGTTACCCAGTGTTTTTTCATTCGTAGATAATGAAGTTCTTTGAACGTTAATGGATATATGGACCTC
-150
-140
-130
-120
-110
-100
-90
-80
-70
-60
-50
DNA
DNA+Protein

Figure 3.19 SELEX-determined DNA-binding motif of the PF1072 protein.

Motif logo (generated using WebLogo [82]) for 9 out of 20 SELEX sequences as determined by MEME motif-finding software [81] with corresponding sequence pileup.

GTACATGACCATGTACTTTGCTCGTC
ACCCGCGTACACGTGGTTGTACTGTCGAAC
CTCGACCGTTCATGACAGTGTACTTGCGGG
AGTACATATGGTTGTACGGGGACTAGG
TCTAGACCTGTACAGGTACTTGTACCGACG
GGTACAGGACAGTGTATCAGGTCGACG
CATGCAAGGTACTTGACCCTGTACAGTGTT
TAGTACACAGTAGTGTACGGAGTGGCCG
AAGCGTACACACCTAAGTACTACTGTCGAC

Figure 3.20 SELEX-determined DNA-binding motif of the PF2053 protein.

Motif logo (generated using WebLogo [82]) for 7 out of 9 SELEX sequences as determined by MEME motif-finding software [81] with corresponding sequence pile-up.

ATTCAAGCTGCTGATCGGCTT

AATAGCTGCGGACTCTGGC

TTACCTCCCCACCTGCCCGGTT

AGCCAACGTTACCTGCTGCTATT

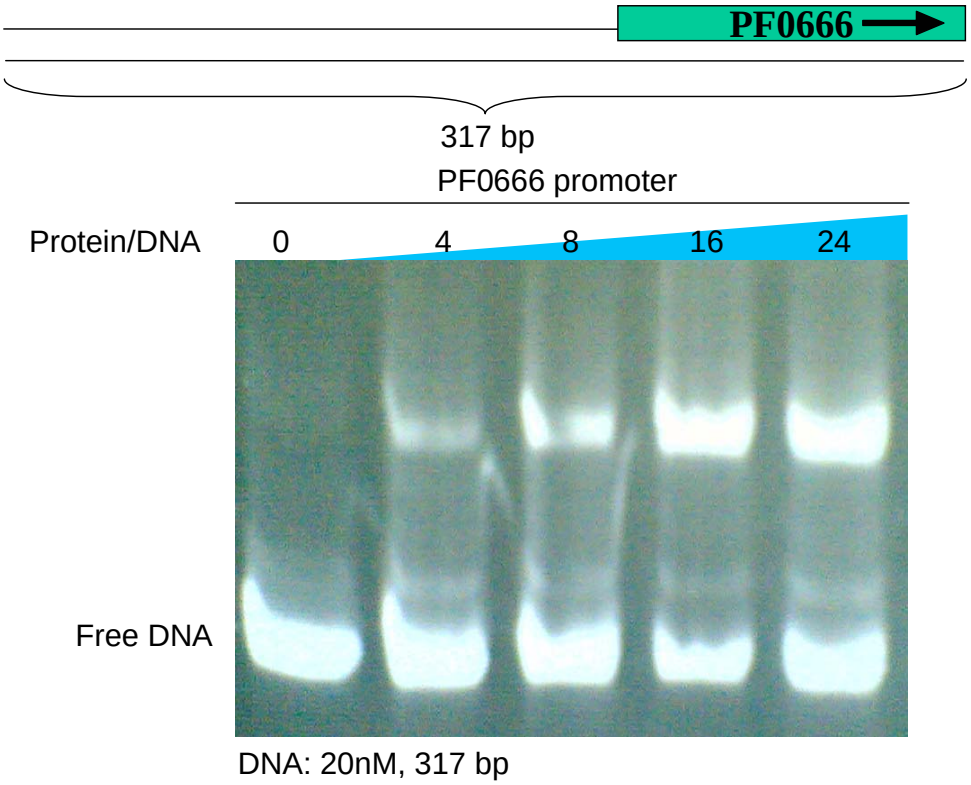
CGTTATACAAAGGTGCCCCG

CACCACAGAGCTCCGCTGTACCCC

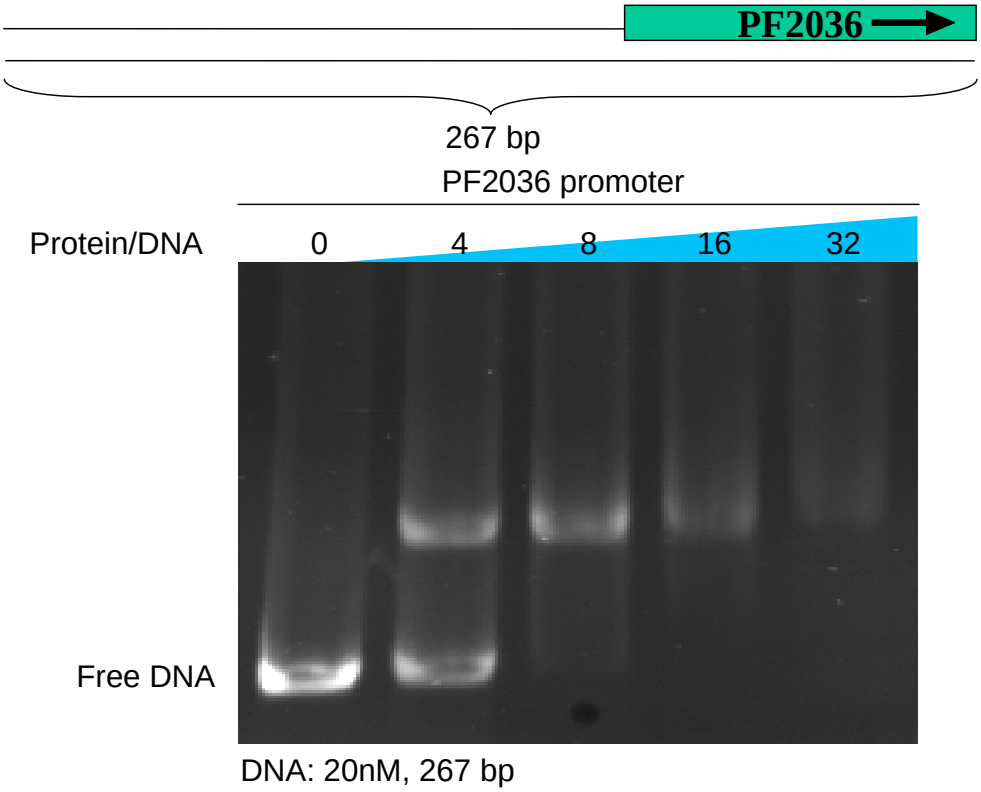
CTAGTACACTGCCTGCGGTGTAGTCC

Figure 3.21 PF1072 protein binds to DNA upstream of up-regulated ORFs.

EMSA with promoter DNA from up-regulated ORFs containing upstream GTACAn_6TGTAC motifs. Probe diagrams are indicated at the top of each gel image with corresponding protein/DNA mole ratios listed above each lane.



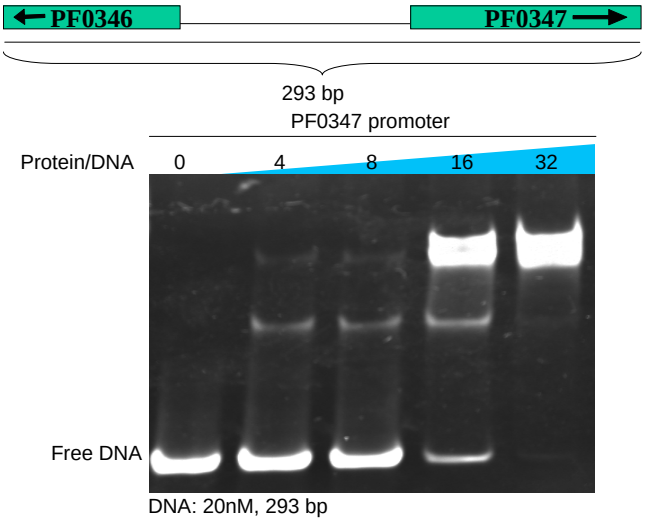
DNA: 20nM, 317 bp

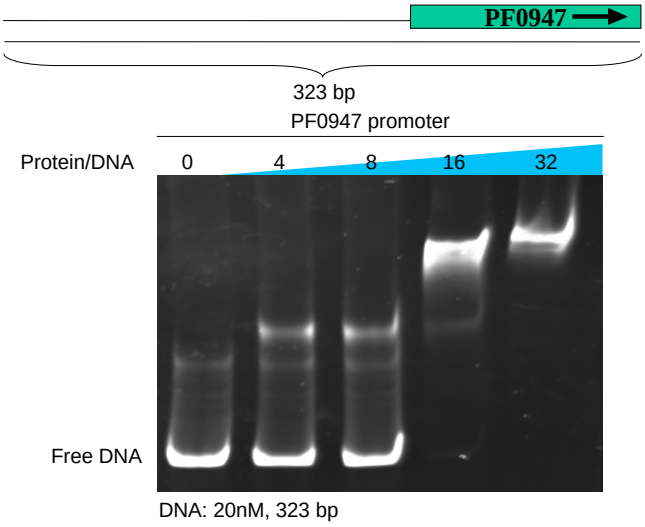


DNA: 20nM, 267 bp

Figure 3.22 PF1072 protein binds to DNA upstream of down-regulated ORFs.

EMSA with promoter DNA from down-regulated ORFs containing upstream GTACAn_6TGTAC motifs. Probe diagrams are indicated at the top of each gel image with corresponding protein/DNA mole ratios listed above each lane.


PF0346
PF0347
293 bp
PF0347 promoter
Protein/DNA
0
4
8
16
32
Free DNA
DNA: 20nM, 293 bp


PF0947
323 bp
PF0947 promoter
Protein/DNA
0
4
8
16
32
Free DNA
DNA: 20nM, 323 bp

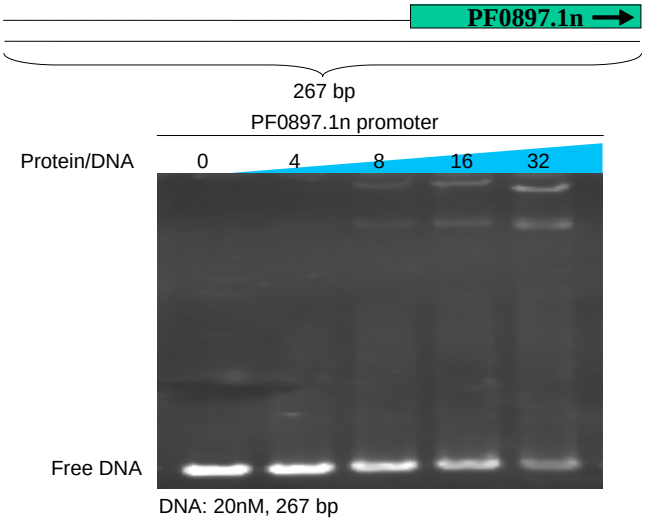
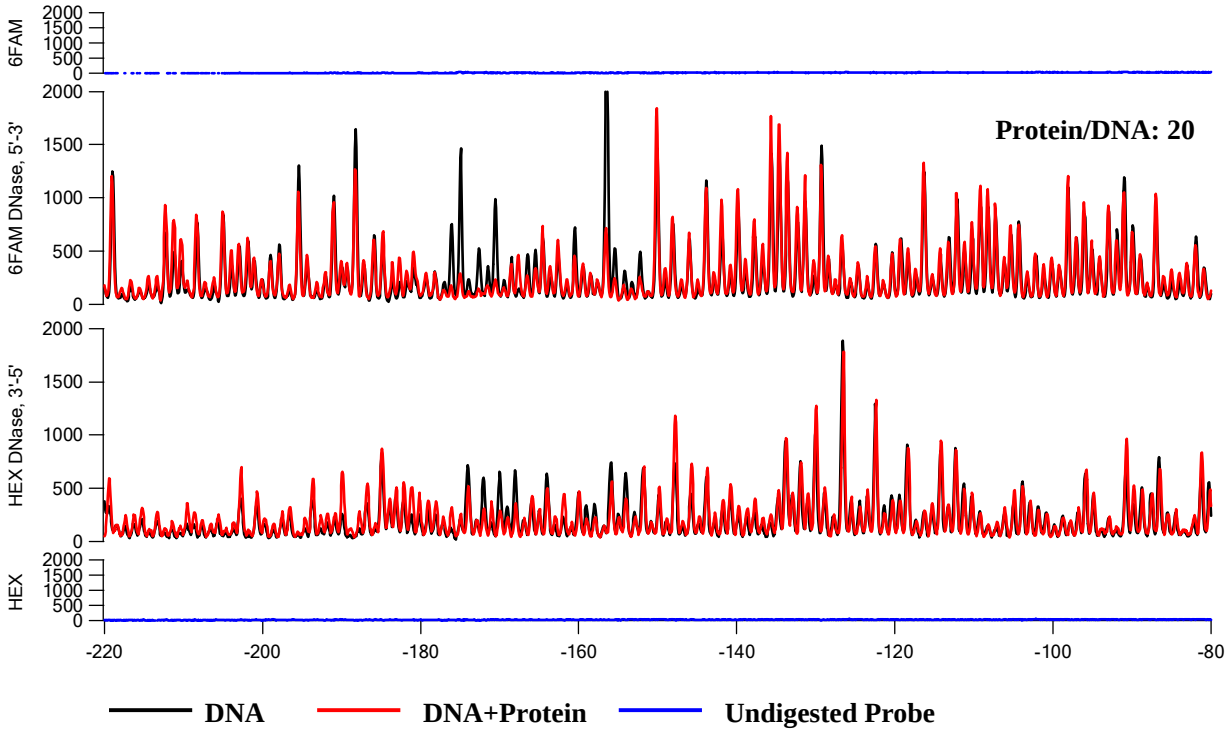

PF0897.1n
267 bp
PF0897.1n promoter
Protein/DNA
0
4
8
16
32
Free DNA
DNA: 20nM, 267 bp

Figure 3.23 DNase I footprint of the PF1072 protein on the PF0666 promoter.

A. Electropherograms of 6FAM/HEX-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Numbers on bottom indicates positions relative to translation start site of PF0666 ORF. **B.** Enlarged view of footprint region with corresponding sequence.

A


6FAM
6FAM DNase, 5'-3'
HEX DNase, 3'-5'
HEX
Protein/DNA: 20
-220
-200
-180
-160
-140
-120
-100
-80
DNA
DNA+Protein
Undigested Probe

B

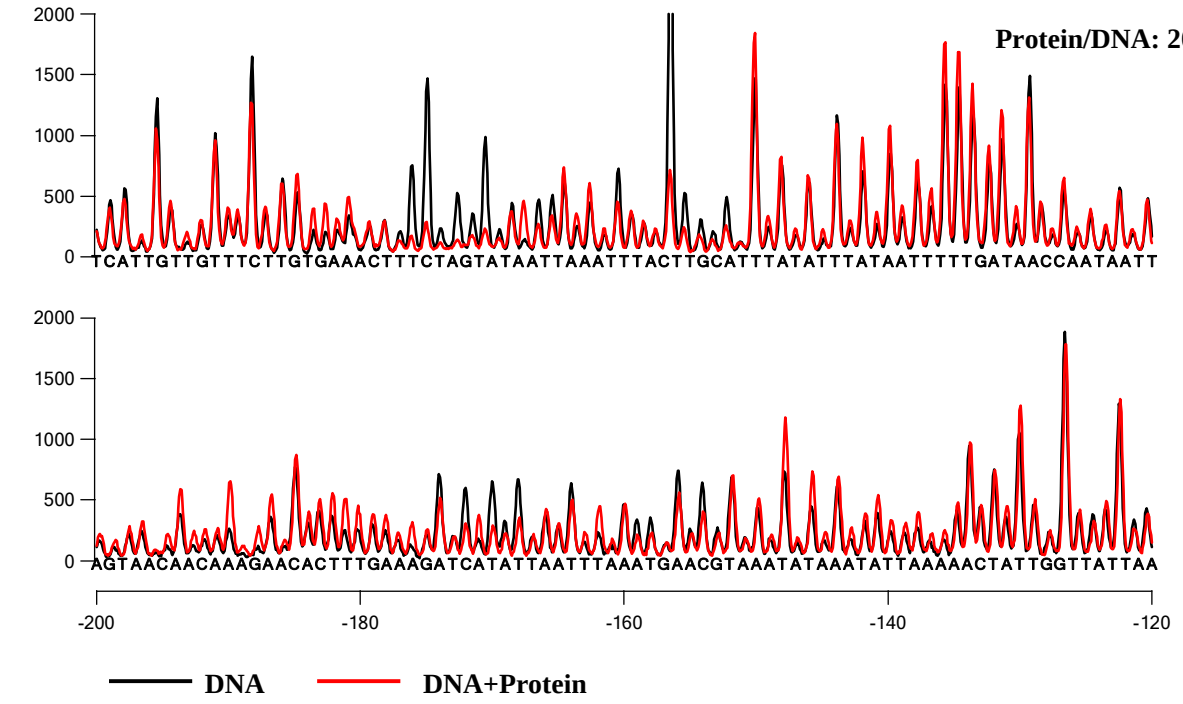
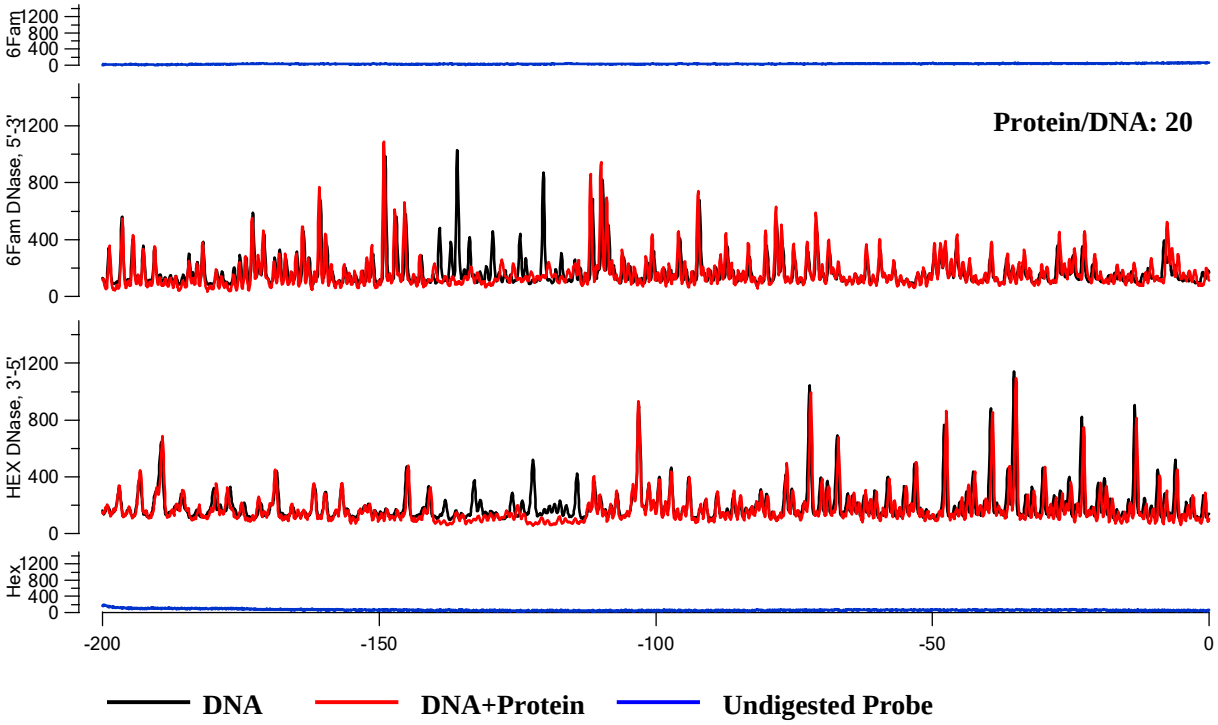

Protein/DNA: 20
TCATTGTTGTTTCTTGTGAAACTTTCTAGTATAATTAAATTTACTTGCATTTATATTTATAATTTTTGATAACCAATAATT
AGTAACAACAAAGAACACTTTGAAAGATCATATTAATTTAAATGAACGTAAATATAAATATTAAAAACTATTGGTTATTAA
-200
-180
-160
-140
-120
DNA
DNA+Protein

Figure 3.24 DNase I footprint of the PF1072 protein on the PF2036 promoter.

A. Electropherograms of 6FAM/HEX-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Numbers on bottom indicates positions relative to translation start site of PF2036 ORF. **B.** Enlarged view of footprint region with corresponding sequence.

A


6Fam
1200
800
400
0
6Fam DNase, 5'-3'
Protein/DNA: 20
HEX DNase, 3'-5'
Hex
-200
-150
-100
-50
0
DNA
DNA+Protein
Undigested Probe

B

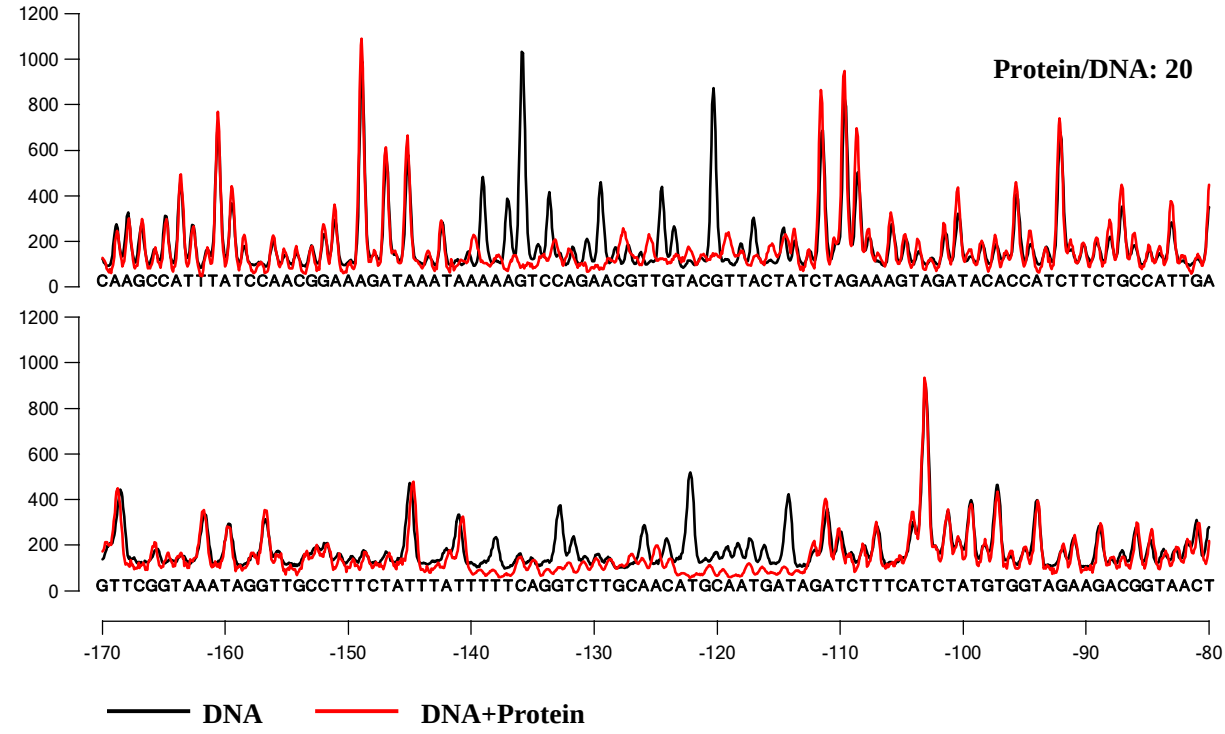
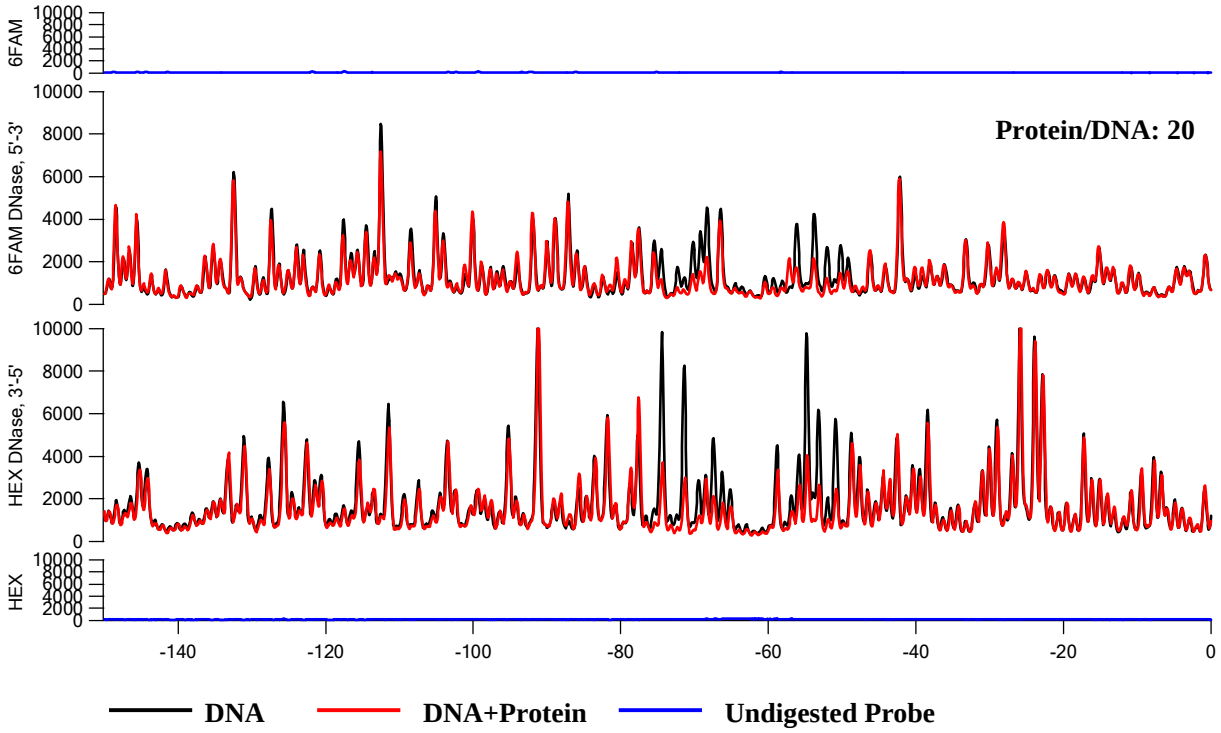

1200
1000
800
600
400
200
0
Protein/DNA: 20
CAAGCCATTTATCCAACGGAAAGATAAATAAAAAGTCCAGAACGTTGTACGTTACTATCTAGAAAGTAGATACACCATCTTCTGCCATTGA
GTTCGGTAAATAGGTTGCCTTTCTATTTATTTTTCAGGTCTTGCAACATGCAATGATAGATCTTTCATCTATGTGGTAGAAGACGGTAACT
-170
-160
-150
-140
-130
-120
-110
-100
-90
-80
DNA
DNA+Protein

Figure 3.25 DNase I footprint of the PF1072 protein on the PF0347 promoter.

A. Electropherograms of 6FAM/HEX-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Numbers on bottom indicates positions relative to translation start site of PF0347 ORF. **B.** Enlarged view of footprint region with corresponding sequence.

A


6FAM
6FAM DNase, 5'-3'
HEX DNase, 3'-5'
HEX
Protein/DNA: 20
-140
-120
-100
-80
-60
-40
-20
0
DNA
DNA+Protein
Undigested Probe

B

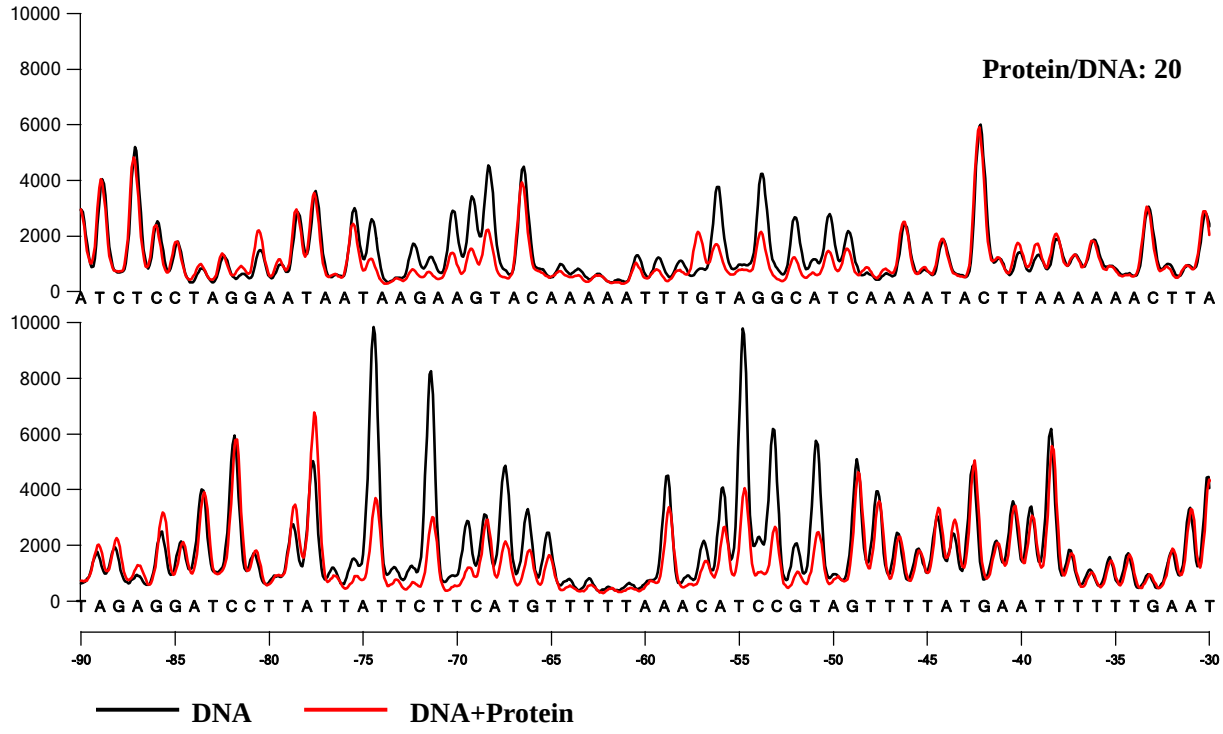
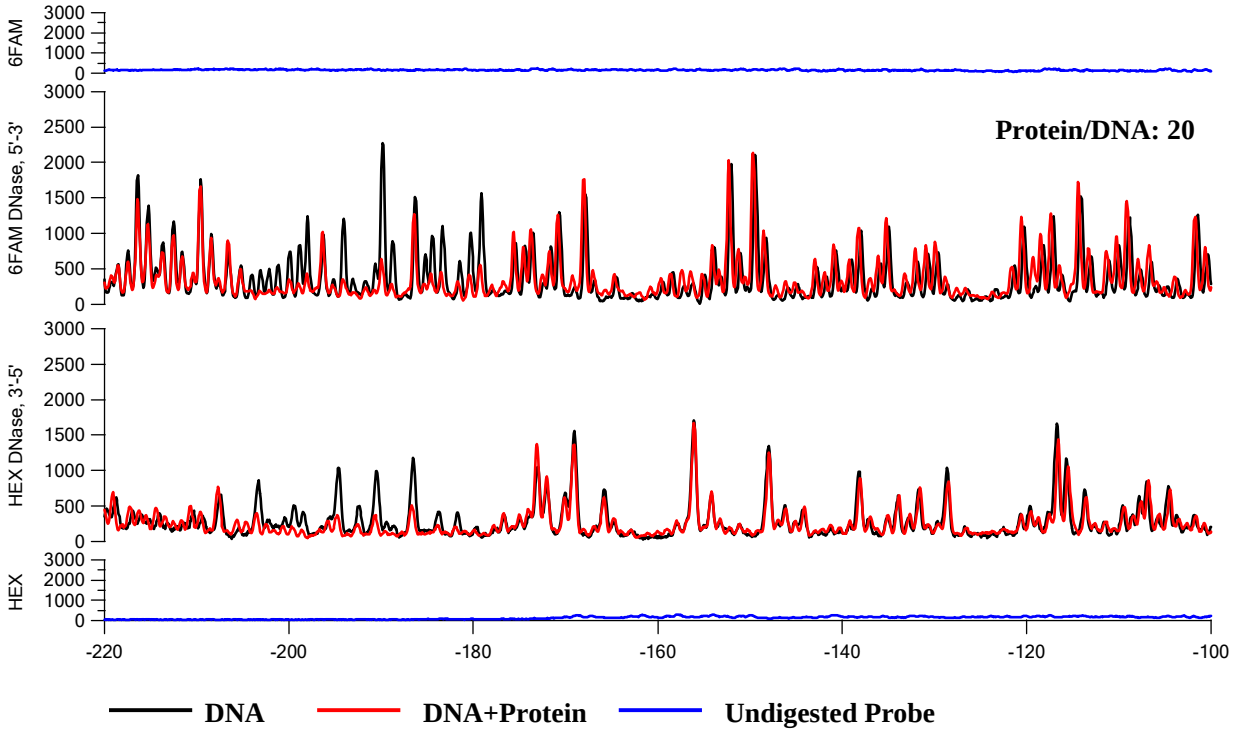

Protein/DNA: 20
A T C T C C T A G G A A T A A T A A G A A G T A C A A A A A T T T G T A G G C A T C A A A A T A C T T A A A A A A C T T A
T A G A G G A T C C T T A T T A T T C T T C A T G T T T T T A A A C A T C C G T A G T T T T A T G A A T T T T T T T G A A T
-90
-85
-80
-75
-70
-65
-60
-55
-50
-45
-40
-35
-30
DNA
DNA+Protein

Figure 3.26 DNase I footprint of the PF1072 protein on the PF0947 promoter.

A. Electropherograms of 6FAM/HEX-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Numbers on bottom indicates positions relative to translation start site of PF0947 ORF. **B.** Enlarged view of footprint region with corresponding sequence.

A


6FAM
6FAM DNase, 5'-3'
HEX DNase, 3'-5'
HEX
Protein/DNA: 20
-220
-200
-180
-160
-140
-120
-100
DNA
DNA+Protein
Undigested Probe

B

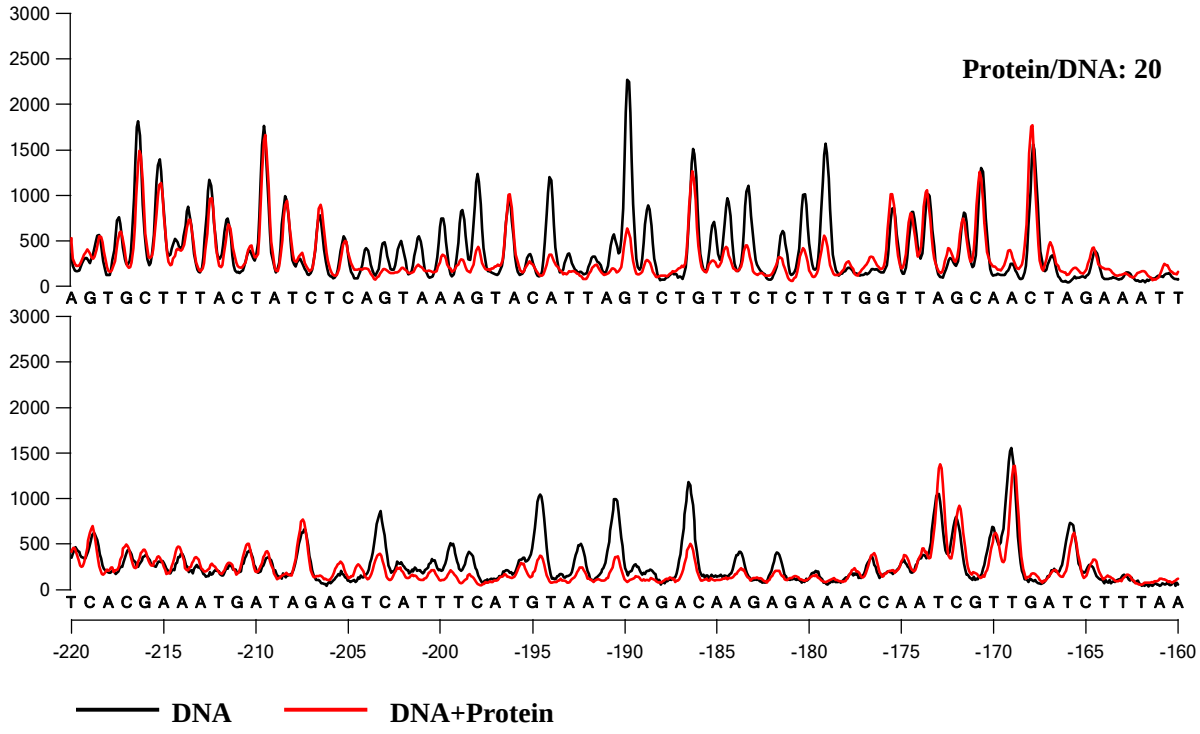
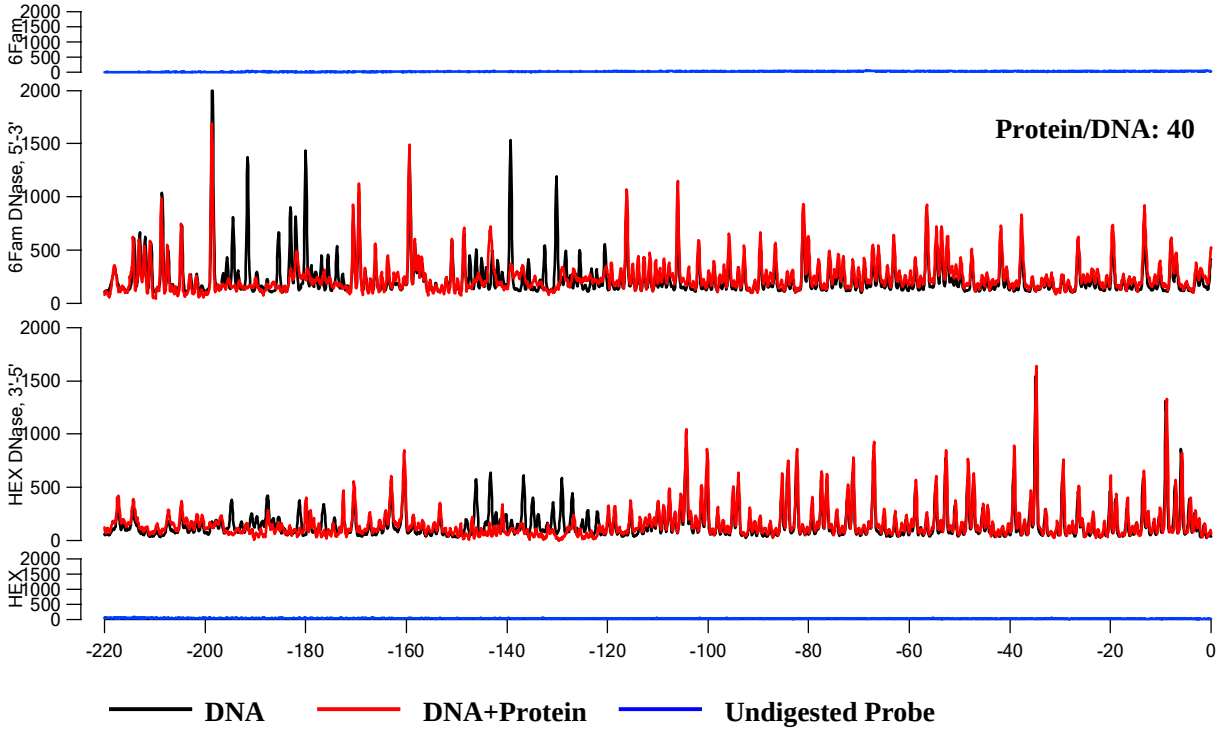

Protein/DNA: 20
AGTGCTTTACTATCTCAGTAAAGTACATTAGTCTGTTCTCTTTGGTTAGCAACTAGAAATT
TCACGAAATGATAGAGTCATTTCATGTAATCAGACAAGAGAAACCAATCGTTGATCTTTAA
-220
-215
-210
-205
-200
-195
-190
-185
-180
-175
-170
-165
-160
DNA
DNA+Protein

Figure 3.27 DNase I footprint of the PF1072 protein on the PF0897.1n promoter.

A. Electropherograms of 6FAM/HEX-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Numbers on bottom indicates positions relative to translation start site of PF0897.1n ORF. **B.** Enlarged view of footprint region with corresponding sequence.

A


6Fam
6Fam DNase, 5'-3'
HEX DNase, 3'-5'
HEX
Protein/DNA: 40
-220 -200 -180 -160 -140 -120 -100 -80 -60 -40 -20 0
DNA
DNA+Protein
Undigested Probe

B

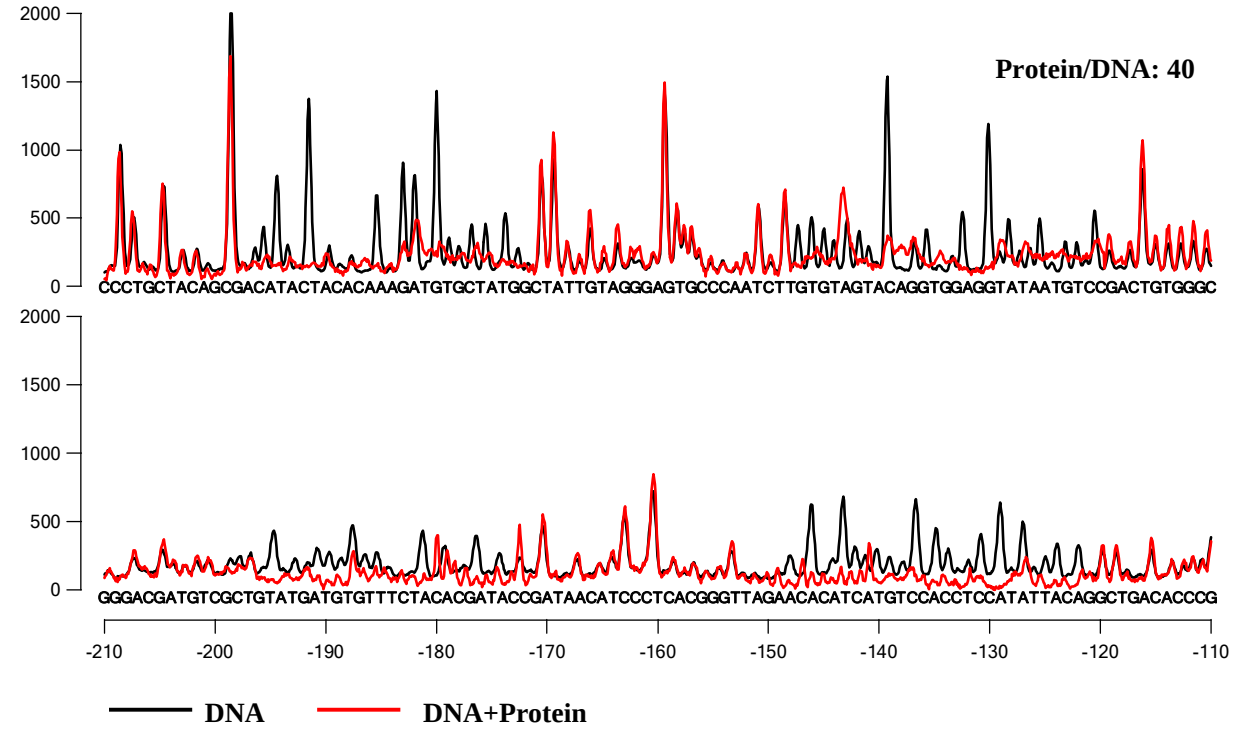

Protein/DNA: 40
CCCTGCTACAGCGACATACTACACAAAGATGTGCTATGGCTATTGTAGGGAGTGCCCAATCTTGTGTAGTACAGGTGGAGGTATAATGTCCGACTGTGGGC
GGGACGATGTCGCTGTATGATGTGTTTCTACACGATACCGATAACATCCCTCACGGGTTAGAACACATCATGTCCACCTCCATATTACAGGCTGACACCCG
-210 -200 -190 -180 -170 -160 -150 -140 -130 -120 -110
DNA
DNA+Protein

Figure 3.28 PF1072 protein recognition motifs.

Relevant DNA sequences of probes are shown with DNase footprint sites boxed in red. SELEX consensuses are highlighted in red. Mismatches are underlined. The numbers on both ends indicate the position relative to the translation start site.

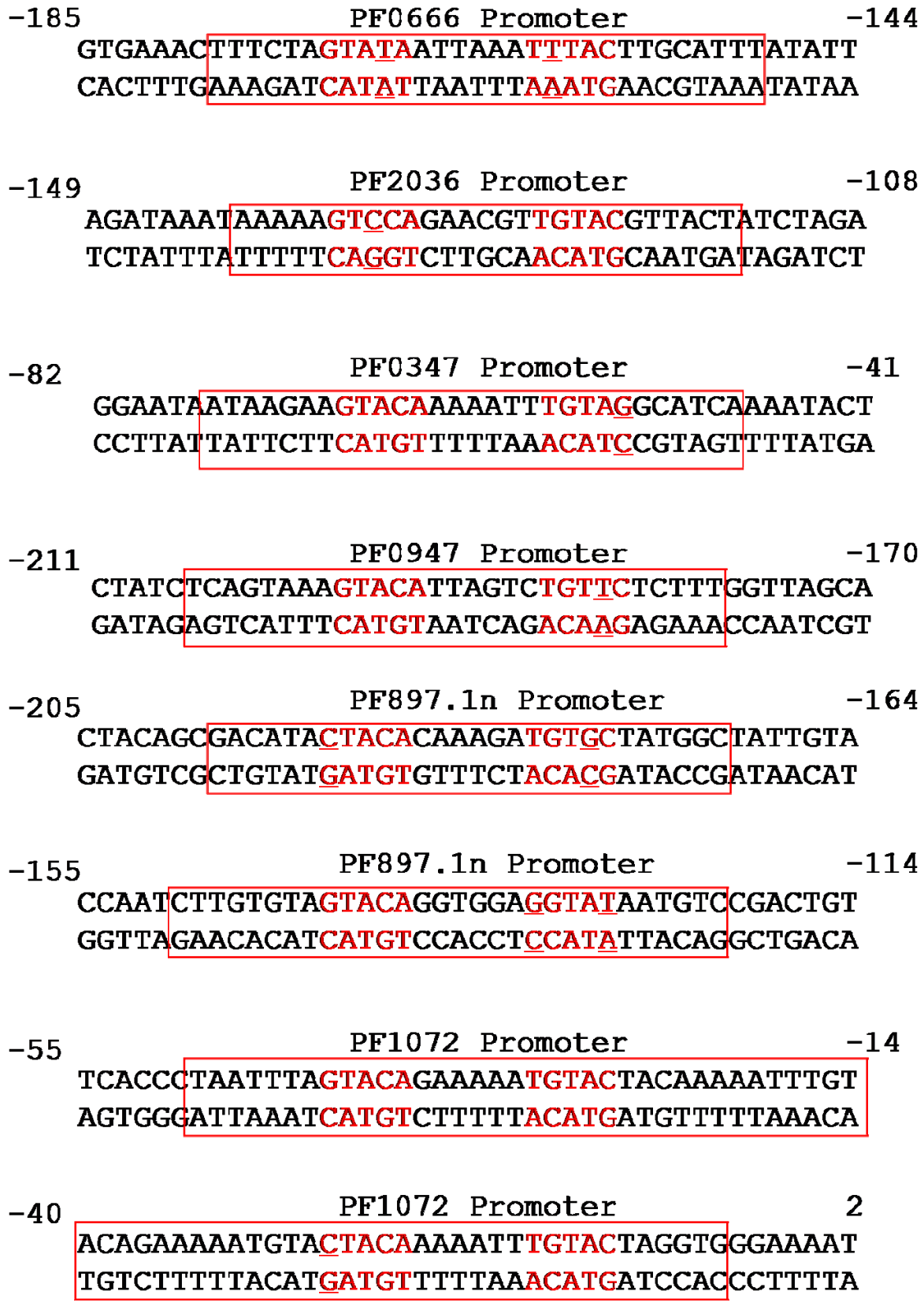

-185
PF0666 Promoter
-144
GTGAAACTTTCTAGTATAATTAAATTTACTTGCATTTATATT
CACTTTGAAAGATCATATTAATTTAAATGAACGTAAATATAA
-149
PF2036 Promoter
-108
AGATAAATAAAAAGTCCAGAACGTTGTACGTTACTATCTAGA
TCTATTTATTTTTCAGGTCTTGCAACATGCAATGATAGATCT
-82
PF0347 Promoter
-41
GGAATAATAAGAAGTACAAAAATTTGTAGGCATCAAAATACT
CCTTATTATTCTTCATGTTTTTAAACATCCGTAGTTTTATGA
-211
PF0947 Promoter
-170
CTATCTCAGTAAAGTACATTAGTCTGTTCTCTTTGGTTAGCA
GATAGAGTCATTTCATGTAATCAGACAAGAGAAACCAATCGT
-205
PF897.1n Promoter
-164
CTACAGCGACATACTACACAAAGATGTGCTATGGCTATTGTA
GATGTCGCTGTATGATGTGTTTCTACACGATACCGATAACAT
-155
PF897.1n Promoter
-114
CCAATCTTGTGTAGTACAGGTGGAGGTATAATGTCCGACTGT
GGTTAGAACACATCATGTCCACCTCCATATTACAGGCTGACA
-55
PF1072 Promoter
-14
TCACCCTAATTTAGTACAGAAAAATGTACTACAAAAATTTGT
AGTGGGATTAAATCATGTCTTTTTACATGATGTTTTTAAACA
-40
PF1072 Promoter
2
ACAGAAAAATGTACTACAAAATTTGTACTAGGTGGGAAAAT
TGTCTTTTTACATGATGTTTTTAAACATGATCCACCCTTTTA

Figure 3.29 PF1072 protein is a transcriptional repressor.

In vitro transcription assays with PF1072 protein performed by Antonia Gindner for PF1072 promoter template. A schematic diagram of the template is indicated at the top of the gel image with corresponding protein/DNA mole ratio listed above each lane. The relative amount of transcript in percentage is listed below each lane.

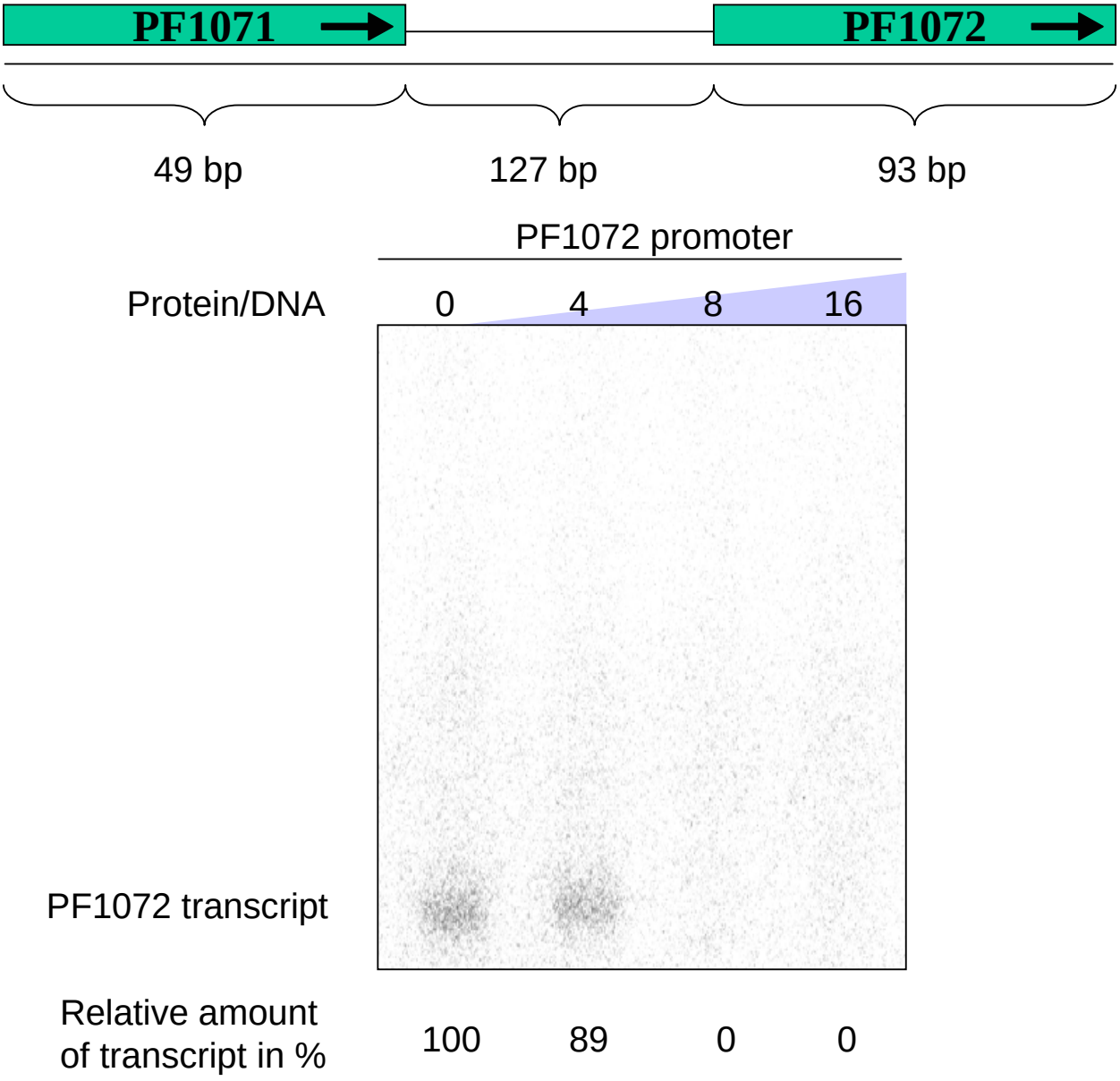

PF1071
PF1072
49 bp
127 bp
93 bp
PF1072 promoter
Protein/DNA
0
4
8
16
PF1072 transcript
Relative amount
of transcript in %
100
89
0
0

Figure 3.30 PF1072 protein is a transcriptional activator.

In vitro transcription assays with PF1072 protein performed by Antonia Gindner for PF10947 promoter template. A schemiatic diagram of the template is indicated at the top of the gel image with corresponding protein/DNA mole ratio listed above each lane. The relative amount of transcript in percentage is listed below each lane.

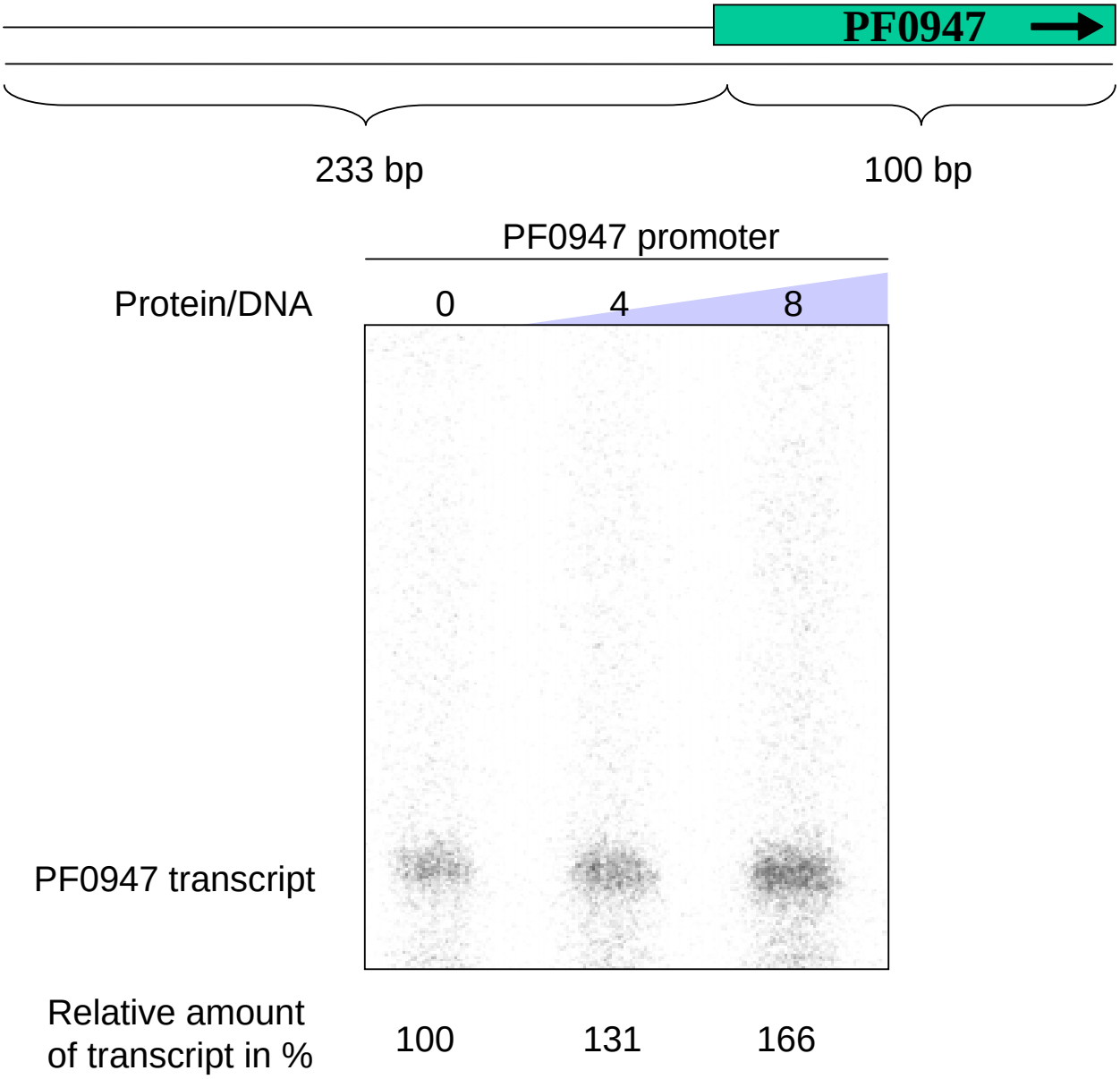

PF0947
233 bp
100 bp
PF0947 promoter
Protein/DNA
0
4
8
PF0947 transcript
Relative amount
of transcript in %
100
131
166

Figure 3.31 The protein-protein pull-down experiment.

His-tagged PF1072 protein was bound to Ni-NTA magnetic agarose beads. The bead-bound protein was then incubated with soluble cell extract. Proteins which did not bind his-tagged protein are removed, and finally the proteins which remain were eluted and analyzed.

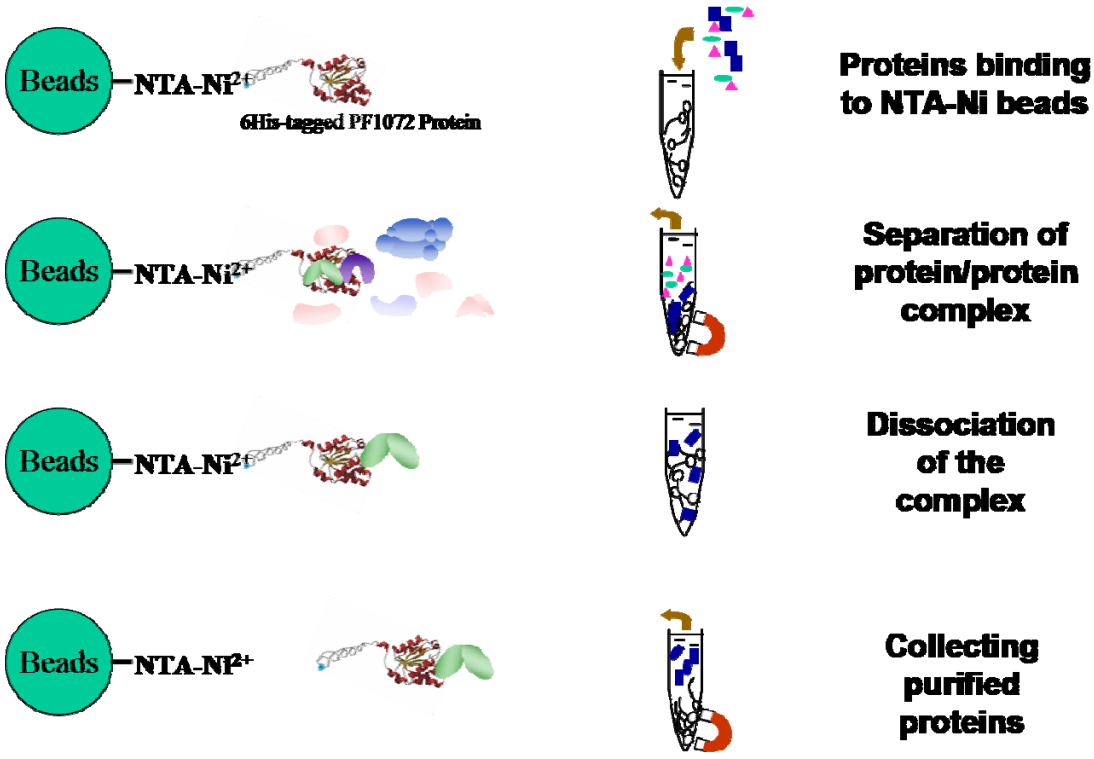
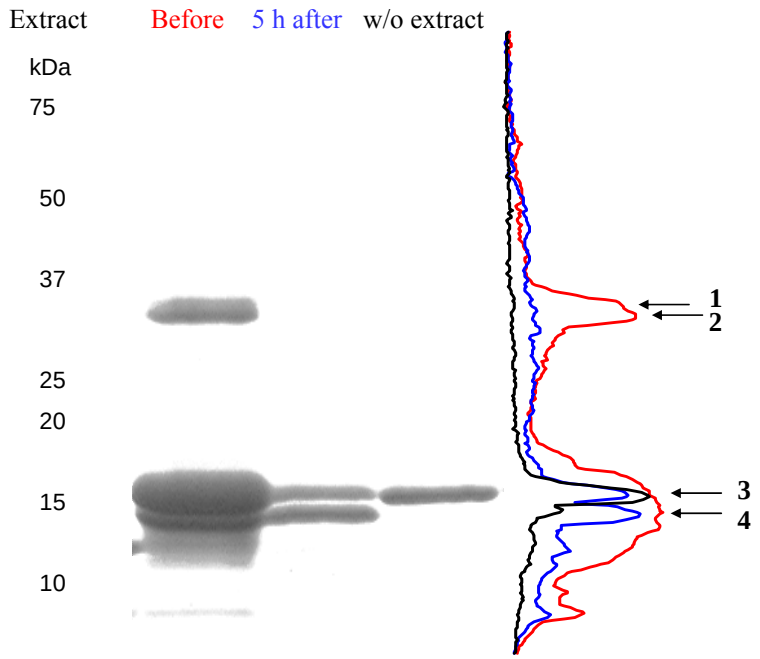

Beads
NTA-Ni2+
6His-tagged PF1072 Protein
Proteins binding to NTA-Ni beads
Beads
NTA-Ni2+
Separation of protein/protein complex
Beads
NTA-Ni2+
Dissociation of the complex
Beads
NTA-Ni2+
Collecting purified proteins

Figure 3.32 Protein capture with his-tagged PF1072 protein.

Stained denaturing gel of eluted proteins from protein-protein pull-down assay with his-tagged PF1072 protein incubated in soluble cell extracts from cells grown before or 1 h after cold shock (red), 5 h after cold shock (blue) and a control incubated without cell extract (black). Arrows indicate identified proteins: 1, his-tagged PF1072 protein (presumably a dimer); 2, wild type PF1072 protein (presumably a dimer); 3, his-tagged PF1072 protein monomer; 4, wild type PF1072 protein monomer. **A.** Incubated in soluble cell extracts for 20 min. **B.** Incubated in soluble cell extracts for 60 min.

A


Extract
Before
5 h after
w/o extract
kDa
75
50
37
25
20
15
10
1
2
3
4

B

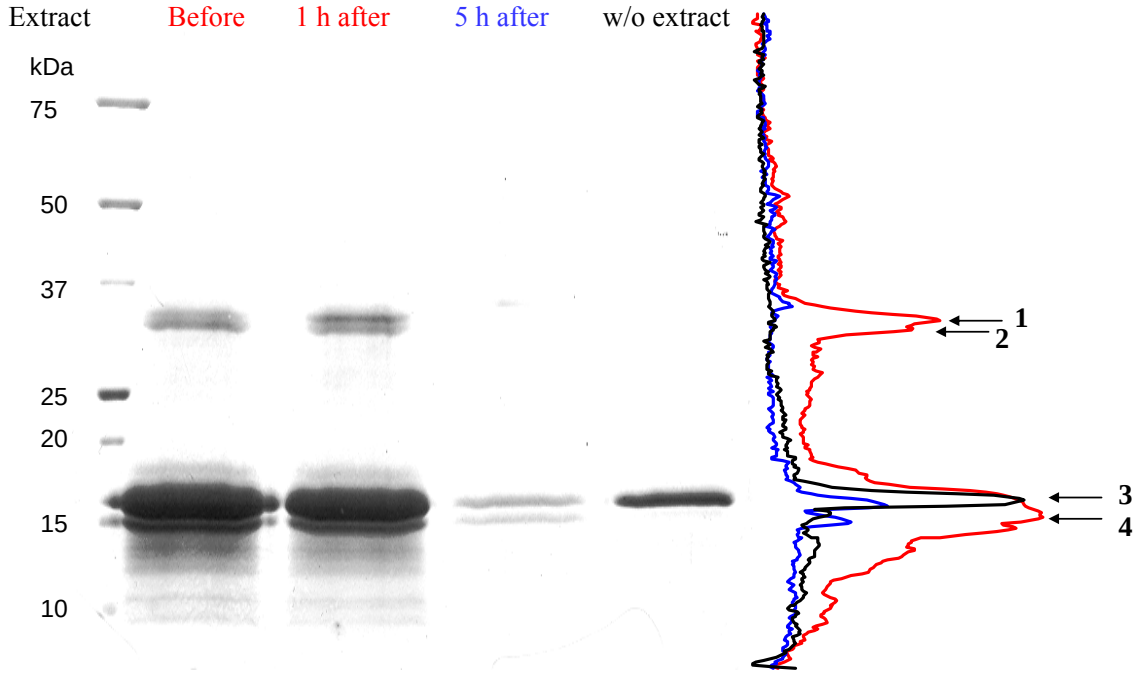

Extract
Before
1 h after
5 h after
w/o extract
kDa
75
50
37
25
20
15
10
1
2
3
4

Figure 3.33 Quaternary structure of the PF1072 protein.

The elution profile of the PF1072 protein sample from a Superdex 75 10/300 GL column (GE Healthcare, Pittsburgh, PA) is shown. The corresponding molecular weight was calculated using a calibration curve (blue) generated from a set of standards run through the column. The calculated molecular weight of the PF1072 protein peak (red) is approximately 38 kDa.

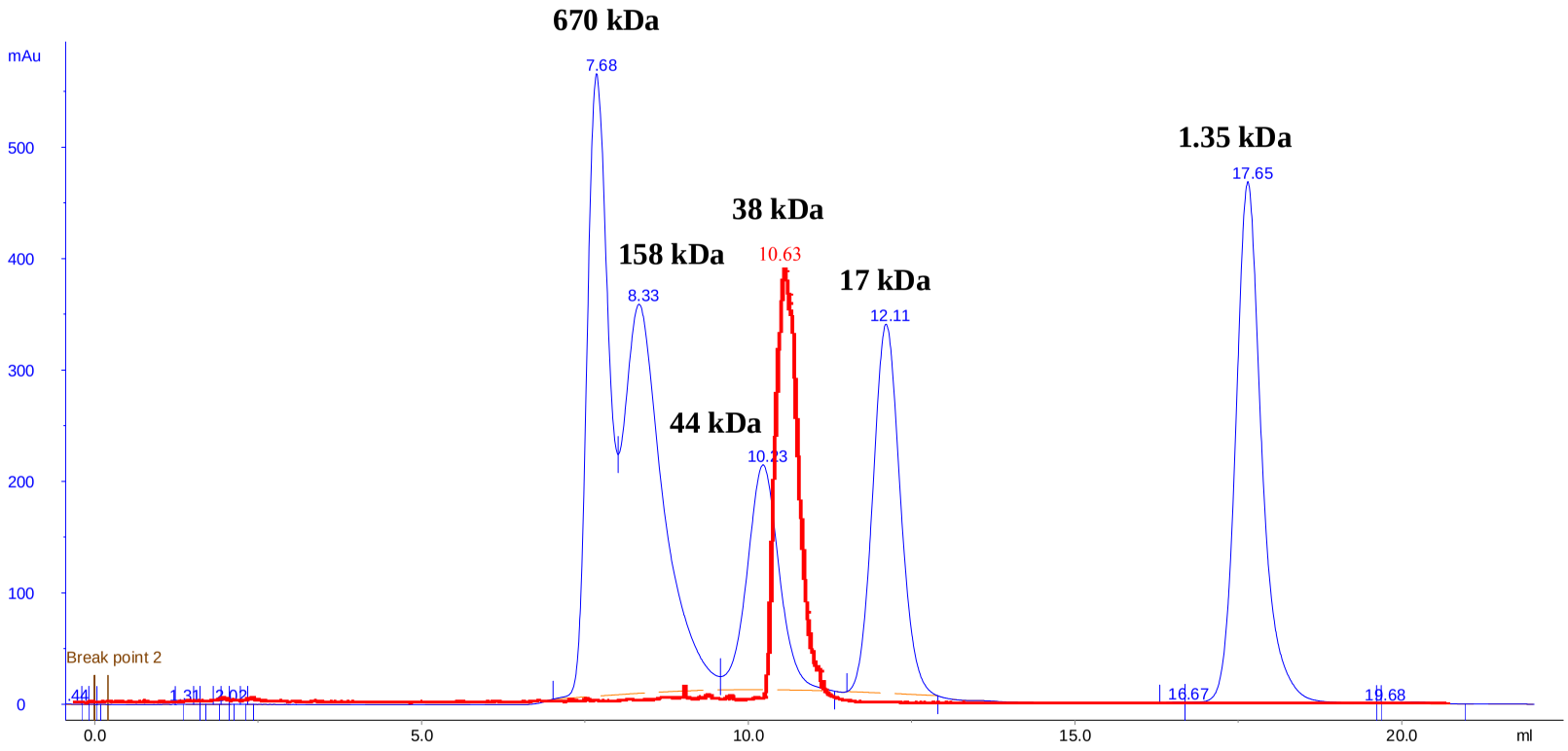
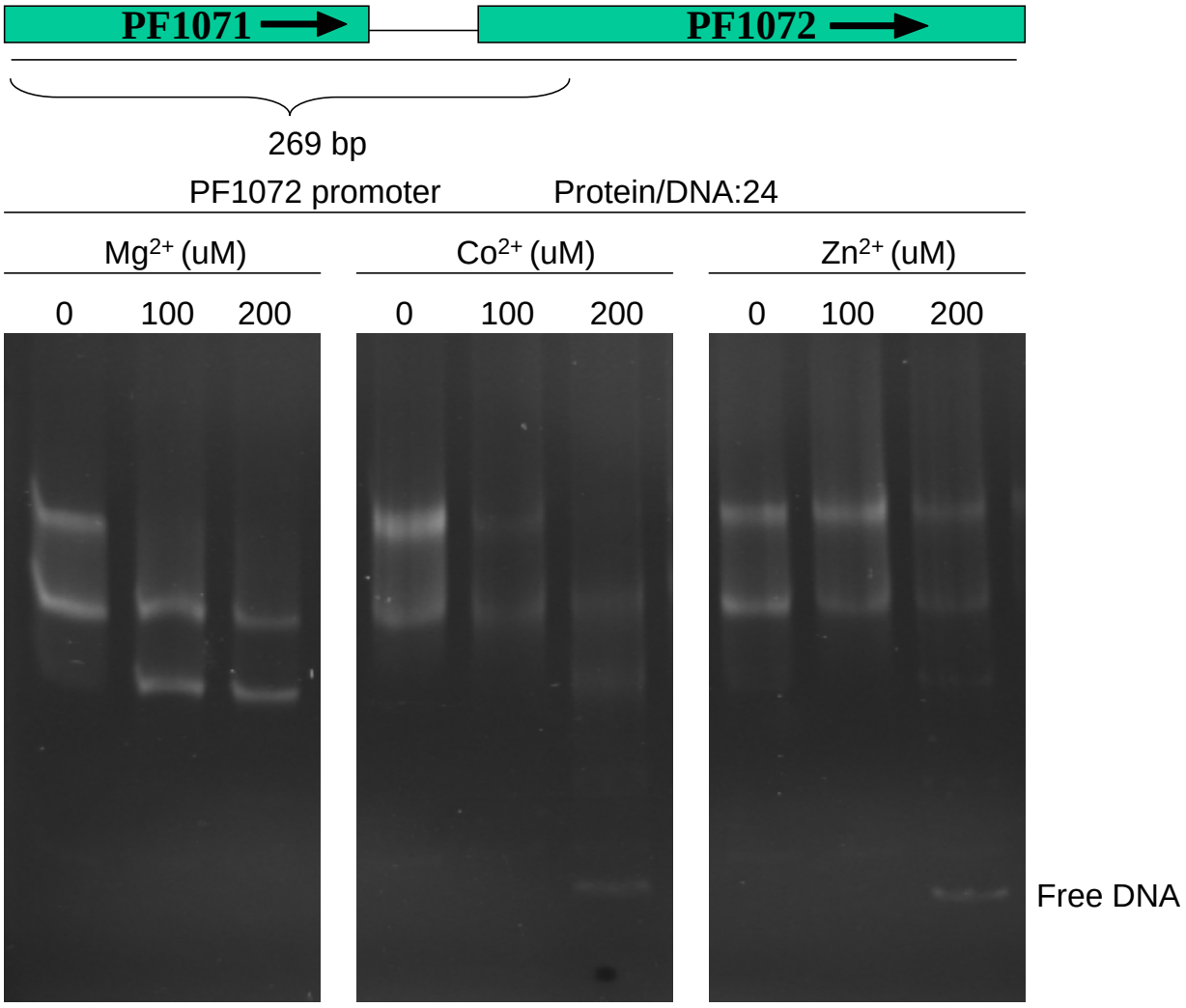

670 kDa
7.68
158 kDa
8.33
44 kDa
10.23
38 kDa
10.63
17 kDa
12.11
1.35 kDa
17.65
mAu
500
400
300
200
100
0
Break point 2
16.67
19.68
0.0
5.0
10.0
15.0
20.0
ml

Figure 3.34 Metal ions have effect on protein-DNA affinity.

Diagrams of the DNA probes are shown at the top. The probes used are indicated at the top of gel image with corresponding protein/DNA mole ratios listed above. DNA (20 nM) was incubated with protein and metal ions in buffer (50 mM Tris-HCl, pH 7.5; 1 mM EDTA; 0.1 M KCl; 5% glycerol) for 20 min at 55 °C. The gel was stained with SYBR Green I nucleic acid gel stain.



DNA: 20nM, 269 bp

CHAPTER 4

THE ROLE OF PF1072 IN *PYROCOCCUS FURIOSUS* COLD SHOCK RESPONSE

4.1 *P. furiosus* cold shock response and transcriptional regulation

In the cold shock experiment, *P. furiosus* was grown on maltose at 95 °C, and then cultures were shocked by rapidly dropping the temperature from 95 to 72 °C. This resulted in a 5 h lag phase, during which little growth occurred. Transcriptional analyses showed that cells undergo three very different responses at 72 °C: an early shock (1 to 2 h), a late shock (5 h), and an adapted response. In addition, *P. furiosus* was grown at 72 °C in a cold-adapted experiment.

In the early shock phase, a number of conserved hypothetical proteins were up-regulated. These proteins appear to be involved in the translation process, in amino acid and primary carbohydrate metabolism, in oxidoreductase-type reactions, and in solute transport. In the late shock phase, multiple ORFs involved in the biosynthesis of branched-chain amino acids and methionine were up-regulated, but none of these ORFs were up-regulated even after 2 h. The main response in the late shock phase therefore seems to be an increase in the biosynthesis of certain amino acids that utilize the intermediates generated in the early shock phase from the primary carbon source maltose, as well as the up-regulation of a number of proteins of unknown function.

Once the cells are fully adapted to growth at the lower temperature, only a few conserved hypothetical proteins were up-regulated. A significant number of these proteins are predicted to contain transmembrane domains. On the other hand, the expression of most proteins was

down-regulated. 152 ORFs were found to be significantly down-regulated for cold adapted *P. furiosus*. These included proteins with predicted involvement in energy metabolism and in central intermediary metabolism, proteins active in translation, amino acid biosynthesis, those necessary for binding and transport, and transcription.

All life forms must contend with the crises caused by cold shock; however, it is unclear which of the specific cold shock response systems are universal. None of the previously defined bacterial cold shock proteins, such as the CspA family, CsdA, or RbfA, are present in the genomes of any sequenced hyperthermophilic archaea. Therefore, these organisms must possess a novel and distinct cold shock response, which is different from the cold shock response of bacteria. Overall, however, a universal response to cold shock is metabolic rate depression. As mentioned before, the expression of most proteins was down-regulated for cold adapted *P. furiosus*. The depression of protein synthesis will lead to the accumulation of sources and intermediates which might inhibit cellular growth. The response to this is to repress the membrane transport proteins which import ions and small molecules into the cell. On the other hand, some other membrane transport proteins are activated to import necessary compounds for low temperature adaptation or to export toxic compounds. Therefore, it seems that the regulation of transporters and permeases play an important role in the cold shock response, which correlates to the up-/down-regulation of many ORFs containing transmembrane domains.

4.2 PF1072 protein as transcriptional regulator

PF1072 protein has been shown to bind to the putative promoters of several of the genes which are both up- and down-regulated after cold shock according to DNA microarray studies; moreover, its DNA-binding motif elucidated by SELEX was found in 32 ORF promoters; 11 of

these that have the motif in their upstream DNA were regulated by low temperature more than two-fold (Tables 3.3, 3.4). This result was puzzling since most of the products of these ORFs are hypothetical proteins.

PF1072 protein has been shown to repress transcription from its own gene that is part of an operon with two other genes, PF1073 and PF1074. A BLAST search of these proteins shows that they are highly conserved among the Thermococcaceae family, but the functions remain unknown. PF1073 contains four transmembrane domains, suggesting it may be a membrane transporter. Another possible transporter, PF0947, was proved to be activated by PF1072 protein. A BLAST search shows PF0947 is predicted to be a permease of major facilitator superfamily, but the solute bound by this protein is not known. For other genes regulated by PF1072 protein, they fall into a number of different categories. No conclusions can be drawn from this result.

There are two known primary mechanisms of transcription repression in archaea: interference with TBP/TFB recognition of TATA/BRE elements on the gene promoter or obstructing RNA polymerase recruitment at the transcription initiation sites. The footprinting study indicated that the protection site on its own promoter by PF1072 protein is at about -5 to -55. A putative TATA/BRE of PF1072 is from -62 to -75, which is upstream of the protected region. PF1072 protein does not seem to interfere with the TATA/BRE recognition by TBP/TFB for its own promoter, suggesting that the negative regulation by PF1072 protein is probably mediated by abrogating the recruitment of RNAP to promoter.

PF1072 protein was shown to be a repressor of its own gene in the *in vitro* transcription system. The protein-protein pull-down experiment indicates higher concentration of PF1072 protein before cold shock compared to after cold shock. Furthermore, microarray results show a significant increase in expression of PF1072 after the cold shock. It is suspected that some

factor(s) in the cell extract from low temperature growth conditions may reduce the concentration and dimer formation of PF1072 protein. Under the low temperature growth conditions, the interaction of PF1072 protein with its own gene is reduced, de-repressing gene transcription. For the PF0947 ORF, microarray results show a significant decrease in expression after cold shock. It is reasonable to assume that the interaction of PF1072 protein with PF0947 promoter is reduced therefore de-activating the gene transcription.

The middle region (residues 55-72) of the PF1072 protein is similar to the dimer interface of the SmtB repressor [89], and residues Ser77 and Glu80 were predicted to be putative zinc binding ligands while the residues from 55 to 100 might be involved in DNA interactions (Fig. 3.8). The overlap between the putative zinc binding site and the DNA binding domain may predict that the PF1072 protein tends to dissociate from DNA in the presence of metal ions. Experiments were performed to test the effect of various metals on the ability of the PF1072 protein to shift DNA. Magnesium, cobalt and zinc appear to decrease the affinity between PF1072 protein and its own promoter.

4.3 Conclusion and outlook

PF1072 protein has been directly shown *in vitro* to both activate and repress transcription, and it is the first archaeal regulator shown to respond to the low temperature growth conditions. This protein is interesting since it appears to function as an autoregulator, acting upon some cold shock responsive ORFs, and potentially other ORFs which contain its binding site in their upstream DNA. There is much more work that could be done to elucidate the complete regulatory network and *in vivo* functions of PF1072 protein within the cell. Structural information of the PF1072 protein might help to elucidate the mechanism behind cold shock

response. It might also help to identify the effectors of this protein by defining potential binding sites for them. Whether PF1072 protein is responsive to any effectors for regulation of multiple genes, or whether any other proteins are involved in the regulation need to be further explored. The fact that PF1072 protein is both an activator and a repressor is interesting. More information regarding the particular mechanism of activation carried out by this protein would contribute to the overall knowledge of archaeal transcriptional regulatory mechanisms.

Examining global gene expression using microarrays for mRNA levels has enhanced the understanding of the genetics and physiology of archaeal responses to the cold. This knowledge should be coupled with biophysical and biochemical studies to determine structural and functional properties of proteins important for activity at low temperature. The work that has been done for PF1072 protein will contribute to this field and help to shed light on novel mechanisms and processes involving transcription within the archaeal domain.

REFERENCES

1. Woese, C.R. and G.E. Fox, *Phylogenetic structure of the prokaryotic domain: the primary kingdoms.* Proc Natl Acad Sci U S A, 1977. **74**(11): p. 5088-90.
2. Woese, C.R., O. Kandler, and M.L. Wheelis, *Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.* Proc Natl Acad Sci U S A, 1990. **87**(12): p. 4576-9.
3. Chaban, B., S.Y. Ng, and K.F. Jarrell, *Archaeal habitats--from the extreme to the ordinary.* Can J Microbiol, 2006. **52**(2): p. 73-116.
4. Robertson, C.E., et al., *Phylogenetic diversity and ecology of environmental Archaea.* Curr Opin Microbiol, 2005. **8**(6): p. 638-42.
5. Bell, S.D. and S.P. Jackson, *Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features.* Trends Microbiol, 1998. **6**(6): p. 222-8.
6. Stetter, K.O., *Hyperthermophiles in the history of life.* Philos Trans R Soc Lond B Biol Sci, 2006. **361**(1474): p. 1837-42; discussion 1842-3.
7. Fiala, G. and K.O. Stetter, *Pyrococcus-Furiosus Sp-Nov Represents a Novel Genus of Marine Heterotrophic Archaebacteria Growing Optimally at 100-Degrees C.* Archives of Microbiology, 1986. **145**(1): p. 56-61.
8. Kelly, R.M. and M.W. Adams, *Metabolism in hyperthermophilic microorganisms.* Antonie Van Leeuwenhoek, 1994. **66**(1-3): p. 247-70.
9. Asturias, F.J., *RNA polymerase II structure, and organization of the preinitiation complex.* Curr Opin Struct Biol, 2004. **14**(2): p. 121-9.
10. Sims, R.J., 3rd, S.S. Mandal, and D. Reinberg, *Recent highlights of RNA-polymerase-II-mediated transcription.* Curr Opin Cell Biol, 2004. **16**(3): p. 263-71.
11. Woychik, N.A. and M. Hampsey, *The RNA polymerase II machinery: structure illuminates function.* Cell, 2002. **108**(4): p. 453-63.
12. Best, A.A. and G.J. Olsen, *Similar subunit architecture of archaeal and eukaryal RNA polymerases.* FEMS Microbiol Lett, 2001. **195**(1): p. 85-90.
13. Geiduschek, E.P. and M. Ouhammouch, *Archaeal transcription and its regulators.* Mol Microbiol, 2005. **56**(6): p. 1397-407.
14. Borukhov, S. and E. Nudler, *RNA polymerase holoenzyme: structure, function and biological implications.* Curr Opin Microbiol, 2003. **6**(2): p. 93-100.
15. Cramer, P., *Multisubunit RNA polymerases.* Curr Opin Struct Biol, 2002. **12**(1): p. 89-97.
16. Gross, C.A., et al., *The functional and regulatory roles of sigma factors in transcription.* Cold Spring Harb Symp Quant Biol, 1998. **63**: p. 141-55.
17. deHaseth, P.L., M.L. Zupancic, and M.T. Record, Jr., *RNA polymerase-promoter interactions: the comings and goings of RNA polymerase.* J Bacteriol, 1998. **180**(12): p. 3019-25.
18. Hausner, W., et al., *Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase.* J Biol Chem, 1996. **271**(47): p. 30144-8.
19. Bell, S.D., et al., *The archaeal TFIIEalpha homologue facilitates transcription initiation*

by enhancing TATA-box recognition. EMBO Rep, 2001. **2**(2): p. 133-8.
20. Bartlett, M.S., *Determinants of transcription initiation by archaeal RNA polymerase.* Curr Opin Microbiol, 2005. **8**(6): p. 677-84.
21. Smale, S.T. and J.T. Kadonaga, *The RNA polymerase II core promoter.* Annu Rev Biochem, 2003. **72**: p. 449-79.
22. van de Werken, H.J., et al., *Identification of a glycolytic regulon in the archaea Pyrococcus and Thermococcus.* FEMS Microbiol Lett, 2006. **260**(1): p. 69-76.
23. Bell, S.D., et al., *Orientation of the transcription preinitiation complex in archaea.* Proc Natl Acad Sci U S A, 1999. **96**(24): p. 13662-7.
24. Paget, M.S. and J.D. Helmann, *The sigma70 family of sigma factors.* Genome Biol, 2003. **4**(1): p. 203.
25. Kosa, P.F., et al., *The 2.1-A crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box.* Proc Natl Acad Sci U S A, 1997. **94**(12): p. 6042-7.
26. Werner, F. and R.O. Weinzierl, *Direct modulation of RNA polymerase core functions by basal transcription factors.* Mol Cell Biol, 2005. **25**(18): p. 8344-55.
27. Bell, S.D., *Archaeal transcriptional regulation--variation on a bacterial theme?* Trends Microbiol, 2005. **13**(6): p. 262-5.
28. Bell, S.D., et al., *Transcriptional regulation of an archaeal operon in vivo and in vitro.* Mol Cell, 1999. **4**(6): p. 971-82.
29. Brinkman, A.B., et al., *An Lrp-like transcriptional regulator from the archaeon Pyrococcus furiosus is negatively autoregulated.* J Biol Chem, 2000. **275**(49): p. 38160-9.
30. Enoru-Eta, J., et al., *Purification and characterization of Sa-lrp, a DNA-binding protein from the extreme thermoacidophilic archaeon Sulfolobus acidocaldarius homologous to the bacterial global transcriptional regulator Lrp.* J Bacteriol, 2000. **182**(13): p. 3661-72.
31. Enoru-Eta, J., et al., *High resolution contact probing of the Lrp-like DNA-binding protein Ss-Lrp from the hyperthermoacidophilic crenarchaeote Sulfolobus solfataricus P2.* Mol Microbiol, 2002. **45**(6): p. 1541-55.
32. Peeters, E., et al., *Ss-LrpB, a novel Lrp-like regulator of Sulfolobus solfataricus P2, binds cooperatively to three conserved targets in its own control region.* Mol Microbiol, 2004. **54**(2): p. 321-36.
33. Peeters, E., et al., *Ss-LrpB from Sulfolobus solfataricus condenses about 100 base pairs of its own operator DNA into globular nucleoprotein complexes.* J Biol Chem, 2006. **281**(17): p. 11721-8.
34. Bell, S.D. and S.P. Jackson, *Mechanism of autoregulation by an archaeal transcriptional repressor.* J Biol Chem, 2000. **275**(41): p. 31624-9.
35. Yokoyama, K., et al., *Feast/famine regulatory proteins (FFRPs): Escherichia coli Lrp, AsnC and related archaeal transcription factors.* FEMS Microbiol Rev, 2006. **30**(1): p. 89-108.
36. Brinkman, A.B., et al., *The Sulfolobus solfataricus Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability.* J Biol Chem, 2002. **277**(33): p. 29537-49.
37. Lee, S.J., et al., *TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon Thermococcus litoralis.* J Biol Chem, 2003. **278**(2): p. 983-90.
38. Lee, S.J., et al., *TrmB, a sugar sensing regulator of ABC transporter genes in Pyrococcus*

furiosus exhibits dual promoter specificity and is controlled by different inducers. Mol Microbiol, 2005. **57**(6): p. 1797-807.
39. Krug, M., et al., *Crystal structure of the sugar binding domain of the archaeal transcriptional regulator TrmB.* J Biol Chem, 2006. **281**(16): p. 10976-82.
40. Lee, S.J., et al., *Characterization of the TrmB-like protein, PF0124, a TGM-recognizing global transcriptional regulator of the hyperthermophilic archaeon Pyrococcus furiosus.* Mol Microbiol, 2007. **65**(2): p. 305-18.
41. Kanai, T., et al., *A global transcriptional regulator in Thermococcus kodakaraensis controls the expression levels of both glycolytic and gluconeogenic enzyme-encoding genes.* J Biol Chem, 2007. **282**(46): p. 33659-70.
42. Lie, T.J. and J.A. Leigh, *A novel repressor of nif and glnA expression in the methanogenic archaeon Methanococcus maripaludis.* Mol Microbiol, 2003. **47**(1): p. 235-46.
43. Lie, T.J., G.E. Wood, and J.A. Leigh, *Regulation of nif expression in Methanococcus maripaludis: roles of the euryarchaeal repressor NrpR, 2-oxoglutarate, and two operators.* J Biol Chem, 2005. **280**(7): p. 5236-41.
44. Lie, T.J., et al., *Diverse homologues of the archaeal repressor NrpR function similarly in nitrogen regulation.* FEMS Microbiol Lett, 2007. **271**(2): p. 281-8.
45. Plosser, P. and F. Pfeifer, *A bZIP protein from halophilic archaea: structural features and dimer formation of cGvpE from Halobacterium salinarum.* Mol Microbiol, 2002. **45**(2): p. 511-20.
46. Hofacker, A., et al., *GvpE- and GvpD-mediated transcription regulation of the p-gvp genes encoding gas vesicles in Halobacterium salinarum.* Microbiology, 2004. **150**(Pt 6): p. 1829-38.
47. Zimmermann, P. and F. Pfeifer, *Regulation of the expression of gas vesicle genes in Haloferax mediterranei: interaction of the two regulatory proteins GvpD and GvpE.* Mol Microbiol, 2003. **49**(3): p. 783-94.
48. Vierke, G., et al., *A novel archaeal transcriptional regulator of heat shock response.* J Biol Chem, 2003. **278**(1): p. 18-26.
49. Liu, W., et al., *Crystal structure of the archaeal heat shock regulator from Pyrococcus furiosus: a molecular chimera representing eukaryal and bacterial features.* J Mol Biol, 2007. **369**(2): p. 474-88.
50. Ouhammouch, M., et al., *Activation of archaeal transcription by recruitment of the TATA-binding protein.* Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5097-102.
51. Ouhammouch, M. and E.P. Geiduschek, *A thermostable platform for transcriptional regulation: the DNA-binding properties of two Lrp homologs from the hyperthermophilic archaeon Methanococcus jannaschii.* Embo J, 2001. **20**(1-2): p. 146-56.
52. Ouhammouch, M., et al., *A fully recombinant system for activator-dependent archaeal transcription.* J Biol Chem, 2004. **279**(50): p. 51719-21.
53. Ouhammouch, M. and E.P. Geiduschek, *An expanding family of archaeal transcriptional activators.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15423-8.
54. Kessler, A., et al., *A novel archaeal regulatory protein, Sta1, activates transcription from viral promoters.* Nucleic Acids Res, 2006. **34**(17): p. 4837-45.
55. Forterre, P., et al., *Speculations on the origin of life and thermophily: review of available information on reverse gyrase suggests that hyperthermophilic procaryotes are not so primitive.* Orig Life Evol Biosph, 1995. **25**(1-3): p. 235-49.

56. Cavicchioli, R., et al., *Low-temperature extremophiles and their applications.* Curr Opin Biotechnol, 2002. **13**(3): p. 253-61.
57. Ermolenko, D.N. and G.I. Makhatadze, *Bacterial cold-shock proteins.* Cell Mol Life Sci, 2002. **59**(11): p. 1902-13.
58. Graumann, P.L. and M.A. Marahiel, *A superfamily of proteins that contain the cold-shock domain.* Trends Biochem Sci, 1998. **23**(8): p. 286-90.
59. Graumann, P.L. and M.A. Marahiel, *Cold shock response in Bacillus subtilis.* J Mol Microbiol Biotechnol, 1999. **1**(2): p. 203-9.
60. Jiang, W., Y. Hou, and M. Inouye, *CspA, the major cold-shock protein of Escherichia coli, is an RNA chaperone.* J Biol Chem, 1997. **272**(1): p. 196-202.
61. Wang, N., K. Yamanaka, and M. Inouye, *CspI, the ninth member of the CspA family of Escherichia coli, is induced upon cold shock.* J Bacteriol, 1999. **181**(5): p. 1603-9.
62. Gualerzi, C.O., A.M. Giuliodori, and C.L. Pon, *Transcriptional and post-transcriptional control of cold-shock genes.* J Mol Biol, 2003. **331**(3): p. 527-39.
63. Jones, P.G. and M. Inouye, *RbfA, a 30S ribosomal binding factor, is a cold-shock protein whose absence triggers the cold-shock response.* Mol Microbiol, 1996. **21**(6): p. 1207-18.
64. Moll, I., et al., *Effects of ribosomal proteins S1, S2 and the DeaD/CsdA DEAD-box helicase on translation of leaderless and canonical mRNAs in Escherichia coli.* Mol Microbiol, 2002. **44**(5): p. 1387-96.
65. Lim, J., T. Thomas, and R. Cavicchioli, *Low temperature regulated DEAD-box RNA helicase from the Antarctic archaeon, Methanococcoides burtonii.* J Mol Biol, 2000. **297**(3): p. 553-67.
66. Ishoy, T., et al., *An improved method for single cell isolation of prokaryotes from meso-, thermo- and hyperthermophilic environments using micromanipulation.* Appl Microbiol Biotechnol, 2006. **69**(5): p. 510-4.
67. Ideno, A., et al., *FK506-binding protein of the hyperthermophilic archaeum, Thermococcus sp. KS-1, a cold-shock-inducible peptidyl-prolyl cis-trans isomerase with activities to trap and refold denatured proteins.* Biochem J, 2001. **357**(Pt 2): p. 465-71.
68. Weinberg, M.V., et al., *Cold shock of a hyperthermophilic archaeon: Pyrococcus furiosus exhibits multiple responses to a suboptimal growth temperature with a key role for membrane-bound glycoproteins.* J Bacteriol, 2005. **187**(1): p. 336-48.
69. Soppa, J., *Transcription initiation in Archaea: facts, factors and future aspects.* Mol Microbiol, 1999. **31**(5): p. 1295-305.
70. Coligan, J.E., *Current protocols in protein science*. 1995, Brooklyn, N.Y.: Wiley. v.
71. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res, 1997. **25**(17): p. 3389-402.
72. Schaffer, A.A., et al., *Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements.* Nucleic Acids Res, 2001. **29**(14): p. 2994-3005.
73. Marchler-Bauer, A., et al., *CDD: a Conserved Domain Database for protein classification.* Nucleic Acids Res, 2005. **33**(Database issue): p. D192-6.
74. Fried, M. and D.M. Crothers, *Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis.* Nucleic Acids Res, 1981. **9**(23): p. 6505-25.
75. Yanisch-Perron, C., J. Vieira, and J. Messing, *Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors.* Gene, 1985.

33(1): p. 103-19.
76. Wilson, D.O., P. Johnson, and B.R. McCord, *Nonradiochemical DNase I footprinting by capillary electrophoresis.* Electrophoresis, 2001. **22**(10): p. 1979-86.
77. Galas, D.J. and A. Schmitz, *DNAse footprinting: a simple method for the detection of protein-DNA binding specificity.* Nucleic Acids Res, 1978. **5**(9): p. 3157-70.
78. Sambrook, J. and D.W. Russell, *Molecular cloning : a laboratory manual*. 3rd ed. 2001, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.
79. Oliphant, A.R., C.J. Brandl, and K. Struhl, *Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein.* Mol Cell Biol, 1989. **9**(7): p. 2944-9.
80. Tuerk, C. and L. Gold, *Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.* Science, 1990. **249**(4968): p. 505-10.
81. Bailey, T.L. and C. Elkan, *Fitting a mixture model by expectation maximization to discover motifs in biopolymers.* Proc Int Conf Intell Syst Mol Biol, 1994. **2**: p. 28-36.
82. Crooks, G.E., et al., *WebLogo: a sequence logo generator.* Genome Res, 2004. **14**(6): p. 1188-90.
83. Chen, Y. and University of Georgia., *The fishing expedition : searching for unidentified archaeal transcription factors*. 2003. p. iv, 86 leaves.
84. Wang, M. and University of Georgia., *Discovery of new transcription factors in Pyrococcus furiosus with DNA-affinity protein capture and mass spectrometry*. 2006. p. vii, 146 leaves.
85. Champoux, J.J., *DNA topoisomerases: structure, function, and mechanism.* Annu Rev Biochem, 2001. **70**: p. 369-413.
86. Komori, K. and Y. Ishino, *Replication protein A in Pyrococcus furiosus is involved in homologous DNA recombination.* J Biol Chem, 2001. **276**(28): p. 25654-60.
87. Richard, D.J., S.D. Bell, and M.F. White, *Physical and functional interaction of the archaeal single-stranded DNA-binding protein SSB with RNA polymerase.* Nucleic Acids Res, 2004. **32**(3): p. 1065-74.
88. Tatusov, R.L., et al., *The COG database: an updated version includes eukaryotes.* BMC Bioinformatics, 2003. **4**: p. 41.
89. Cook, W.J., et al., *Crystal structure of the cyanobacterial metallothionein repressor SmtB: a model for metalloregulatory proteins.* J Mol Biol, 1998. **275**(2): p. 337-46.
90. Hethke, C., et al., *Cell-free transcription at 95 degrees: thermostability of transcriptional components and DNA topology requirements of Pyrococcus transcription.* Genetics, 1999. **152**(4): p. 1325-33.
91. Hethke, C., et al., *A cell-free transcription system for the hyperthermophilic archaeon Pyrococcus furiosus.* Nucleic Acids Res, 1996. **24**(12): p. 2369-76.
92. Darby, N. and T.E. Creighton, *Disulfide bonds in protein folding and stability.* Methods Mol Biol, 1995. **40**: p. 219-52.
93. Marchler-Bauer, A., et al., *CDD: a curated Entrez database of conserved domain alignments.* Nucleic Acids Res, 2003. **31**(1): p. 383-7.
94. Bryson, K., et al., *Protein structure prediction servers at University College London.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W36-8.
95. McGuffin, L.J., K. Bryson, and D.T. Jones, *The PSIPRED protein structure prediction server.* Bioinformatics, 2000. **16**(4): p. 404-5.

APPENDIX

Table A1. Table of abbreviations

Abbreviation	Definition
Basal transcription protein and DNA elements	
BRE	TFB recognition element, purine-rich core promoter element
ORF	Open reading frame
RNAP	RNA polymerase
Pol II	Eukaryotic RNA polymerase II
RBS	ribosome binding site
TATA box	TBP recognition element, T/A-rich core promoter element
TBP	TATA binding protein
TFB	Transcription factor B
Miscellaneous terms	
B&W buffer	Binding & Washing buffer
his-tag	Hexahistidine or his6 tag
HTH	Helix-turn-helix DNA-binding domain
LB	Lauria-Bertani broth
OD_{600}	Optical density at 600 nm
UOR	Upstream of ORF region
UV	Ultraviolet

Methods

EMSA	Electromobility shift assay
PCR	Polymerase chain reaction
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SELEX	Systematic evolution of ligands by exponential enrichment

Chemical names

DTT	Dithiothreitol
IPTG	Isopropyl-β-D-thiogalactopyranoside
PMSF	phenylmethylsulfonyl fluoride
TFA	trifluoroacetic acid
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

Unit definitions

µg	microgram
µL	microliter
µM	micromolar
bp	base pair
kDa	kiloDalton
L	Liter
mg	milligram
min	minute
mL	milliliter
rpm	revolutions per minute
V	volt

Table A2. UOR database search results for the motif ASCTGC

UOR[a]	Motif	UOR Start[b]	UOR Stop[b]	Genome Start	Genome Stop	ORF Annotation[c]
PF0065	ACCTGC	-164	-159	70680	70675	hypothetical protein
PF0091	ACCTGC	-177	-172	97427	97422	hypothetical protein
PF0136	AGCTGC	-58	-53	146132	146127	hypothetical protein
PF0139	ACCTGC	-199	-194	148659	148654	indolepyruvate oxidoreductase subunit B
PF0145	ACCTGC	-179	-174	154479	154474	hypothetical protein
PF0189	ACCTGC	-199	-194	201021	201026	dihydroorotase
PF0190	ACCTGC	-15	-10	201026	201021	hypothetical protein
PF0218	ACCTGC	-188	-183	234752	234747	30S ribosomal protein S27e
PF0219	AGCTGC	-145	-140	235695	235700	putative 6-pyruvoyl tetrahydrobiopterin synthase
PF0366	ACCTGC	-48	-43	381885	381890	hypothetical protein
PF0408	ACCTGC	-151	-146	416268	416273	transposase
PF0430	ACCTGC	-159	-154	439856	439861	phosphoribosylglycinamide formyltransferase 2
PF0435	ACCTGC	-161	-156	443114	443109	DEXX-box atpase
PF0448	AGCTGC	-162	-157	465146	465151	putative mannose-1-phosphate guanyltransferase
PF0554	ACCTGC	-175	-170	570715	570710	hypothetical protein
PF0594	ACCTGC	-151	-146	616247	616252	ornithine carbamoyltransferase
PF0621	ACCTGC	-144	-139	638498	638493	hypothetical protein
PF0629	ACCTGC	-67	-62	642658	642653	hypothetical protein
PF0636	AGCTGC	-98	-93	650228	650233	DEXX-box atpase
PF0655.1n	AGCTGC	-20	-15	668259	668254	hypothetical protein
PF0726	ACCTGC	-173	-168	723104	723099	hypothetical protein
PF0729	ACCTGC	-108	-103	724617	724612	multi domain protein containing corrinoid/iron-sulfur region
PF0735	ACCTGC	-66	-61	731495	731500	hypothetical protein
PF0764	ACCTGC	-162	-157	757485	757490	DEXX-box atpase
PF0774	ACCTGC	-125	-120	765617	765622	hypothetical protein
PF0774	ACCTGC	-121	-116	765613	765618	hypothetical protein
PF0777	ACCTGC	-93	-88	766269	766264	hypothetical protein
PF0779	ACCTGC	-182	-177	767668	767673	hypothetical protein
PF0784	ACCTGC	-171	-166	771492	771497	hypothetical protein
PF0821	ACCTGC	-125	-120	800432	800437	50S ribosomal protein L34e
PF0832	AGCTGC	-95	-90	810412	810417	hypothetical protein
PF0888	ACCTGC	-14	-9	858521	858516	hypothetical protein
PF0900	ACCTGC	-17	-12	872099	872094	hypothetical protein
PF0900.1n	ACCTGC	-82	-77	873487	873492	hypothetical protein
PF0942	AGCTGC	-148	-143	906800	906795	dihydroxy-acid dehydratase
PF0944	ACCTGC	-33	-28	909030	909025	hypothetical protein
PF0978	AGCTGC	-18	-13	935193	935188	hypothetical protein
PF0984	ACCTGC	-107	-102	940438	940443	hypothetical protein
PF1021	AGCTGC	-63	-58	976636	976631	hypothetical protein
PF1029	ACCTGC	-85	-80	984539	984534	hypothetical protein
PF1032	ACCTGC	-165	-160	989355	989360	hypothetical protein
PF1033	AGCTGC	-159	-154	990062	990067	peroxiredoxin
PF1056	ACCTGC	-187	-182	1010071	1010066	aspartate-semialdehyde dehydrogenase

PF1090	AGCTGC	-182	-177	1038204	1038209	putative ABC transporter
PF1107	ACCTGC	-141	-136	1054606	1054611	hit family protein
PF1108	ACCTGC	-59	-54	1054611	1054606	putative alpha-dextrin endo-1, 6-alpha-glucosidase
PF1186	AGCTGC	-191	-186	1133065	1133070	NADH oxidase
PF1216	AGCTGC	-31	-26	1157221	1157226	hypothetical protein
PF1217	AGCTGC	-187	-182	1157737	1157742	hypothetical protein
PF1222	AGCTGC	-69	-64	1159630	1159625	hypothetical protein
PF1234	ACCTGC	-63	-58	1169900	1169905	putative chitinase
PF1241	ACCTGC	-21	-16	1175089	1175084	uracil phosphoribosyltransferase
PF1252	ACCTGC	-179	-174	1186095	1186090	hypothetical protein
PF1313	ACCTGC	-158	-153	1237914	1237909	hypothetical DNA processing smf protein
PF1342	ACCTGC	-194	-189	1261779	1261774	hypothetical protein
PF1380	AGCTGC	-42	-37	1294461	1294456	arginyl-tRNA synthetase
PF1390	ACCTGC	-103	-98	1303754	1303759	hypothetical protein
PF1400	AGCTGC	-88	-83	1314543	1314538	hypothetical protein
PF1400	AGCTGC	-90	-85	1314541	1314536	hypothetical protein
PF1400	ACCTGC	-117	-112	1314514	1314509	hypothetical protein
PF1421	AGCTGC	-165	-160	1336246	1336251	4-aminobutyrate aminotransferase
PF1439	AGCTGC	-166	-161	1349278	1349283	hypothetical protein
PF1440	AGCTGC	-125	-120	1349072	1349067	bifunctional phosphopantothenoylcysteine decarboxylase/phosphopantothenate synthase
PF1459	AGCTGC	-120	-115	1364136	1364131	hypothetical protein
PF1476	AGCTGC	-103	-98	1380698	1380703	hypothetical protein
PF1484	ACCTGC	-58	-53	1388123	1388128	hypothetical protein
PF1494	AGCTGC	-11	-6	1393148	1393143	hypothetical protein
PF1496	ACCTGC	-136	-131	1395940	1395945	hypothetical protein
PF1526	AGCTGC	-119	-114	1424912	1424917	hypothetical protein
PF1526	AGCTGC	-121	-116	1424914	1424919	hypothetical protein
PF1529	AGCTGC	-137	-132	1427067	1427072	pyridoxine biosynthesis protein
PF1557	AGCTGC	-145	-140	1453914	1453919	hypothetical protein
PF1565	ACCTGC	-138	-133	1463394	1463399	DNA-directed RNA polymerase subunit H
PF1587	ACCTGC	-51	-46	1482108	1482103	hypothetical protein
PF1592	AGCTGC	-165	-160	1485872	1485867	tryptophan synthase subunit beta
PF1601	AGCTGC	-159	-154	1493738	1493743	transcriptional regulatory protein, asnC family
PF1624	AGCTGC	-188	-183	1515367	1515362	hypothetical protein
PF1629	ACCTGC	-123	-118	1520288	1520283	hypothetical protein
PF1668	AGCTGC	-81	-76	1550975	1550970	hypothetical protein
PF1680	AGCTGC	-51	-46	1562949	1562944	3-isopropylmalate dehydratase small subunit
PF1723	ACCTGC	-134	-129	1601284	1601279	hypothetical protein
PF1763	AGCTGC	-141	-136	1641957	1641962	GTP-binding protein, gtp1/obg family
PF1832	ACCTGC	-141	-136	1689520	1689515	regulatory protein
PF1863	AGCTGC	-65	-60	1719353	1719348	dimethyladenosine transferase
PF1865	ACCTGC	-185	-180	1720875	1720870	putative regulatory protein
PF1879	AGCTGC	-123	-118	1731528	1731523	o-linked glcnac transferase
PF1922	AGCTGC	-57	-52	1773432	1773437	protein-L-isoaspartate O-methyltransferase
PF1950	ACCTGC	-61	-56	1801974	1801979	xanthine-guanine phosphoribosyltransferase
PF1961	AGCTGC	-154	-149	1810932	1810927	tungsten-containing formaldehyde ferredoxin oxidoreductase wor4

PF1970	ACCTGC	-139	-134	1818743	1818738	putative multiple sugar-binding transport ATP-binding protein
PF1970	AGCTGC	-96	-91	1818786	1818781	putative multiple sugar-binding transport ATP-binding protein
PF2000	AGCTGC	-70	-65	1847871	1847866	glycine dehydrogenase subunit 2
PF2007	ACCTGC	-15	-10	1857110	1857105	hypothetical protein
PF2015	AGCTGC	-163	-158	1862643	1862638	ATP-dependent RNA helicase, putative
PF2047.1n	AGCTGC	-71	-66	1889526	1889531	hypothetical protein
PF2048	AGCTGC	-99	-94	1889531	1889526	peptidase
PF2054.1n	AGCTGC	-86	-81	1894906	1894911	hypothetical protein
PF2054.1n	ACCTGC	-82	-77	1894902	1894907	hypothetical protein
PF2056	ACCTGC	-164	-159	1896903	1896908	30S ribosomal protein S15P

[a] UOR (Upstream of ORF Region) designation corresponds to the locus of the ORF from which the upstream sequence was taken.

[b] Start and stop positions are relative to the UOR sequence where -1 corresponds to the first nucleotide upstream from the ORF start.

[c] Protein annotations are derived from the NCBI database.