# IDENTIFICATION AND CHARACTERIZATION OF A NEW TRANSCRIPTION FACTOR INVOLVED IN PYROCOCCUS FURIOSUS RESPONSE TO OXIDATIVE STRESS

by

SUNGGUAN HONG

(Under the Direction of Robert A. Scott)

## ABSTRACT

Gene expression is often under regulation at the transcriptional level, controlled by transcription factors. The primary goal of this research project is to discover and characterize new transcription factors in *Pyrococcus furiosus* (*Pf*), a model archaeal organism. The transcription system of archaea has both bacterial-like and eukaryotic-like features. Only a very limited number of archaeal transcription regulatory proteins have been identified to date, and transcription regulatory pathways are far from being well known. In this project, transcription factors were discovered using DNA affinity protein capture followed by mass spectrometry based protein identification. The *Pf* gene expression profile on the genome-wide scale has been described and compared between two different cell-growth conditions with cumene hydroperoxide and without cumene hydroperoxide. Additional bioinformatics analysis identified a conserved palindromic sequence in the putative promoter region of several genes that are upregulated by growth with cumene hydroperoxide and all are involved in oxidative stress. Based on these analyses, the DNA fragments upstream of the open reading frame of PF1983 and

PF1513, which both encode hypothetical protein, were selected for DNA-affinity protein capture. A novel transcription regulatory protein PF0230p was identified and its specific DNA binding ability at sites upstream of multiple genes was confirmed by the electrophoretic mobility shift assay (EMSA). DNase I footprinting showed that PF0230p binds to a region upstream of both PF1983 and PF1513, which contains the conserved palindromic motif, ATTAAT. Furthermore, PF0230p was shown to repress PF1983 but enhance PF1513 gene transcription in cell free transcription assays. These results suggest PF0230p is a transcriptional regulator recognizing a conserved palindromic motif and possibly regulating a group of genes in the oxidative stress pathway in *Pyrococcus furiosus*.

INDEX WORDS: Archaea, *Pyrococcus furiosus* , transcription factor, transcriptional regulation, DNA microarray, Oxidative stress response, SELEX, fluorescence footprinting, winged helix-turn-helix motif, PF0230p.

# IDENTIFICATION AND CHARACTERIZATION OF A NEW TRANSCRIPTION FACTOR INVOLVED IN PYROCOCCUS FURIOSUS RESPONSE TO OXIDATIVE STRESS

by

SUNGGUAN HONG

B.S., Chung-Ang University, South Korea, 2001

M.S., Chung-Ang University, South Korea, 2003

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

# IDENTIFICATION AND CHARACTERIZATION OF A NEW TRANSCRIPTION FACTOR INVOLVED IN PYROCOCCUS FURIOSUS RESPONSE TO OXIDATIVE STRESS

by

SUNGGUAN HONG

Major Professor: Robert A. Scott

Committee: I. Jonathan Amster
John Rose

Electronic Version Approved:

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

# DEDICATION

This work is dedicated to my beloved parents and sisters, for everything they have given me throughout my whole life. Also to my wife and best friend who always support me, I will always appreciate their love and support, get strength from them and benefit from their wisdoms and advices in my entire life.

ACKNOWLEDGEMENTS

First of all, I would like to express my acknowledgment to my supervisors Dr. Scott, who guided me to a deeper understanding of biological science.

I am also very grateful to all the project collaborators including Priyanka Srivastava and Gerrit Schut from Dr. Adams' group in the Department of Biochemistry and Molecular Biology, who provided the microarray data and many valuable scientific suggestions; Annette Keese and Antonia Gindner from Dr. Michael Thomm's group at University Regensburg, Germany, who collaborated in the study of cell-free gene transcription, Quentin Florence from Dr. John Rose's group in the Department of Biochemistry and Molecular Biology, who's trying to get crystal for the protein.

Thanks also to all my former and current fellows in Department of Chemistry for providing a good working atmosphere, especially Gina Lipscomb, who gave me great encouragement and help at this project.

Lastly, but perhaps most importantly, I express my gratitude to those who have played an indirect yet vital role in motivating me and keeping me going: all my lab mates past and present and my friends on and off campus. I especially thank my family and particularly my parents and my wife whose support and love have made this work possible.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

## 1.1 Transcription in Archaea

Transcription is a fundamental biological process that uses the DNA-dependent enzyme RNA polymerase (RNAP) to convert the genetic information stored in DNA to a usable code in the form of RNA. RNAP can modulate its catalytic activity and gene specificity by associating with various regulatory proteins. For eukaryal systems, these include activators, repressors and general transcription factors (GTFs). For transcription of eukaryal class II genes (protein encoding genes), RNAP associates specifically with class II GTFs (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH) for the synthesis of messenger RNA (mRNA) [16]. The mRNA is then translated to amino acids for the formation of proteins. Since proteins control the phenotypic expression of the organism, the selection of the amount and time at which an RNA molecule is transcribed from a particular gene is tightly regulated. In particular, transcription initiation events play a key role.

Transcription factors are generally divided into two groups [17]. The first group is the basal transcription factors which are ubiquitous and recruit the RNA polymerase II multi-protein complex to the minimal promoter; the second group is gene-specific transcription factors that activate or repress basal transcription. These proteins bind to regulatory sequences organized in a series of regulatory modules along the DNA [18]. Although archaea contain only one RNA polymerase as in bacteria, the archaeal RNAP contains 8-13 subunits, which are homologous to

eukaryotes in both structure and function [19]. The rest of the archaeal basal transcription machinery is also eukaryotic-like and can be considered a simplified version of eukaryotic RNAPII transcription machinery [12, 20-22]. Accordingly, the archaeal gene promoter contains a eukaryotic TATA box from –22 to –30 bp upstream of the transcription start site, a BRE element immediately upstream of the TATA box (–33 to –34), and an initiator element around the transcription start site (**Figure 1.1) [23]**. But the eukaryotic downstream promoter element (DPE) has not been found in archaea.

Archaeal basal transcription factors include TBP and TFB, which are homologous to eukaryotic TBP and TFIIB, respectively [23]. Since the large subunit of the archaeal RNA polymerase does not contain the CTD that is the substrate for phosphorylation in eukaryotes, the polymerase phosphorylation between initiation and elongation is not expected in archaeal transcription [24]. Therefore the absence of TFIIH and TFIIE homologs in archaea seems reasonable [17, 24]. Archaeal gene transcription initiation can be activated in a cell-free system with only TBP, TFB, and RNAP [23], although some archaeal species contain a TFIIE α subunit homolog that is suggested to facilitate TBP-promoter interactions under unfavorable conditions [25]. Transcription initiation starts with TBP recognizing the TATA element and binding to promoter DNA. This complex is stabilized by TFB site-specific interactions with both promoter DNA and TBP [23], the recognition of BRE by TFB defining the orientation of the complex and the direction of transcription [26-28]. Subsequently, the promoter-TBP-TFB complex recruits RNA polymerase.

Despite recent progress in the study of archaeal basal transcription machinery, the regulation of archaeal transcription is poorly understood. A bioinformatics study was performed using the known bacterial and eukaryotic transcription-associated proteins to identify transcription factors

in four archaeal species by sequence analysis [29]. They predicted 280 transcription factors or transcription-associated proteins in the four archaeal genomes known at that time, of which 168 have homologs only in bacteria, 51 have homologs only in eukaryotes, and the remaining 61 have homologs in both phylogenetic domains. This result suggests that archaeal rTFs are more like those of bacteria than those of eukaryotes. Since then, an increased number of archaeal rTFs have been discovered, most of which are found to be bacterial homologs.

### 1.1.1 Comparison between Archaea, Eukarya and Bacteria

Eukaryotic transcription is carried out by three different RNA polymerases: RNAPI for rRNA, RNAPII for mRNA and snRNA, and RNAPIII for tRNA and small RNA. The structure and function of RNAPII subunits as well as the machinery and organization of the transcription pre-initiation complex (PIC) have been summarized and discussed in several review articles [30-32]. The eukaryotic promoter DNA contains a conserved sequence just upstream of the transcription initiation site, consisting of a "T+A" rich TATA element (TATA box), a purine-rich BRE (transcription factor B recognition element) immediately upstream of the TATA box, and a downstream promoter element (DPE) after the transcription start site (**Figure 1.1**). Besides the promoter and a multi-subunit RNA polymerase, the basal apparatus of RNAPII transcription also includes transcription factor TFIID (consisting of the TATA box binding protein, TBP, and TBP-associated factors, TAFs), TFIIB (binding BRE and TBP-TATA), TFIIA, TFIIE, TFIIF and TFIIH. In the first step of PIC formation, the TATA box is recognized by TBP in TFIID with the facilitation of TFIIA. Secondly, TFIIB recognizes BRE and binds with both the TBP protein of TFIID and the promoter. Then, the complex of RNAPII and TFIIF is recruited to the promoter, followed by binding of TFIIE and TFIIH. The function of TFIIH is to phosphorylate the C-

terminal domain (CTD) of the largest subunit of RNAPII, which is prerequisite in the transition from transcription initiation to the elongation phase. The function of TFIIE is to recruit TFIIH to the promoter and stimulate TFIIH phosphorylating activity (**Figure 1.2**).

Compared to that of the eukaryotes, the bacterial basal transcription machinery is much simpler and comprised of fewer components [17, 33]. The bacterial promoter contains two conserved sequence elements located at –10 bp and –35 bp upstream of the transcription start site (**Figure. 1.1**). Bacteria have only one RNA polymerase system. The holoenzyme contains a four-subunit core polymerase ($\alpha_2\beta\beta'$) and a σ factor. The σ factor plays an important role for promoter recognition through binding at –10 and –35 promoter elements [34]. 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 [35]. As mentioned above, gene transcription *in vivo* is usually activated or repressed by regulatory transcription factors (rTFs) in response to certain environmental stresses and physiological conditions. The activators or repressors function through interaction with basal transcription machinery via either protein-protein interaction or protein–DNA interaction [12].

In addition to the distinctive features of the basal transcriptional apparatus, transcription regulation mechanisms in eukaryotes and bacteria are fundamentally different due to the difference in genomic DNA packing structure [36]. Without a tightly packed structure of genomic DNA, the unregulated state of bacterial promoters is active, and generally repressors are needed to keep the gene transcription level low. Activators are not always necessary for every bacterial gene. For eukaryotes, genomic DNA is packaged tightly in chromosomes with histone proteins and other chromosomal proteins. The tightly packed DNA prevents access by other

molecules such as TBP. In other words, the unregulated state of eukaryotic promoters is repressed, so activators are necessary for eukaryotic gene transcription.Although eukaryotes and bacteria share little in transcription initiation and transcriptional regulation, the discovery of archaea "bridges the gap" by displaying a hybrid transcription system [37].

### 1.2 Transcriptional Regulation in Archaea

Regulatory factors influencing transcription efficiency have to interact, directly or indirectly, with the components of the basal transcription apparatus and modulate their activity. Although the work on the basal archaeal transcription apparatus has progressed well, characterization of regulatory processes in Archaea is still in its infancy. Nevertheless, a wide variety of approaches are emerging. A few different elements possibly involved in global regulatory processes have been detected, although none of them has been studied in detail. As already mentioned, duplication of a gene for a basal transcription factor and subsequent specialization of one of the proteins for a subset of genes is one possibility that could be used by *H. salinarum* (TBP) and *P. horikishii* (TFB) [38]. The influence of DNA topology on transcription efficiency could be a second mode of global regulation. More than 20 Z-DNA-containing genomic fragments have been isolated from *H. salinarum* making use of an anti-Z-DNA antibody [39]. If these were located in promoter regions, conditional Z-DNA formation could influence transcription. For one gene, the bacterioopsin gene, it has been shown that transcription is influenced by the superhelicity of the DNA, and that a non-B-DNA conformation, possibly Z-DNA, in the promoter is involved in regulation [40]. A third mechanism could be a differential packaging of the chromosome. For *H. salinarum*, it has been shown that the chromosome is protein free in the early exponential growth phase, and that it is packaged into regular nucleosome-like structures in

stationary phase [41]. It is reasonable to postulate that this growth phase-dependent DNA packaging will have a global influence on transcription. The hyperthermophilic archaeon *Methanothermus fervidus* produces histone-like proteins throughout the growth phase, and archaeal histone-like proteins have been most thoroughly studied in this organism [42]. However, two variants exist, HmfA and HmfB, for which differences in DNA binding and compaction has been shown [43]. During exponential growth phase, predominantly HmfA is produced, whereas HmfB production equals HmfA production during stationary phase, again a possible mechanism for global transcriptional regulation [43].

A variety of systems to study gene-specific transcription regulation in different Archaea have been established, which will expand our understanding in the near future. It is emerging that there will be no common 'archaeal' mechanism of gene regulation. Three examples should testify to this prediction: (i) The protein GvpE has been characterized as an activator of gas vesicle gene transcription in *H. salinarum*. It contains a leucine zipper motif, which was shown to be important for function and probably is necessary for protein-protein interaction [44]. Thus, it is a 'Eukarya-like' activator of transcription; (ii) The first archaeal gene-specific regulator to be described was the repressor T6 of the haloarchaeal phage FH. It contains a helix-turn-helix DNA-binding domain, and a protein dimer interacts with a DNA motif of dyad symmetry [45], and thus is a protein that follows a typical 'bacterial' paradigm; (iii) The ArcR protein is involved in regulation of the genes for arginine fermentation in *H. salinarum* [46]. ArcR has similarities to repressors containing helix-turn-helix motifs, but ArcR is shorter by 100 amino acids and is lacking this DNA-binding domain [46]. Thus, this 'bacteria-like' protein has a different mode of action from its homologues. The archaeal genomes also indicate variability in regulatory mechanisms. The *A. fulgidus* genome contains multiple members of 'bacteria-like' two-

component regulatory systems (at least 15 histidine kinases and nine response regulators), whereas the *M. jannaschii* genome contains none at all. Even if the necessity for regulation and concomitantly the abundance of regulators are different [47], the genome sequences indicate that several paradigms for regulatory mechanisms are used in Archaea (e.g. helix- turn-helix and coiled-coil), and that group-specific differences exist [48].

### 1.2.1 Activation, Repression by Transcription factors in Archaea

There can be two kinds of regulators in archaeal transcriptional regulation, responsible for negative and positive regulation (repression and activation, respectively). Several archaeal repressors and putative repressors exert their effects on transcription according to simple bacterial rules: they bind to DNA sites that overlap with a promoter, occluding the TATA box and BRE or blocking RNAP recruitment [21]. Regulatory mechanisms have been examined in vivo as well as in vitro in two instances. The *Archaeoglobus fulgidus* MDR1 (metal-dependent repressor) gene, which encodes a homologue of the bacterial DtxR family of transcription regulators [49], is the first gene of a cotranscribed four-gene cluster; the three downstream genes encode a presumed ABC transporter for metal ions. Metal ion ($Fe^{2+}$, $Mn^{2+}$ or $Ni^{2+}$) dependent MDR1 binding to multiple sites that overlap with, and extend downstream of the start site of the MDR1 promoter blocks polymerase recruitment in vitro [26]. Sequestering metal ions in the growth medium releases MDR1 from (the vicinity of) its promoter, with concomitant restoration of MDR1 transcription [50].

The *Methanococcus maripaludis* nitrogen regulator NrpR is a tetrameric helix–turn–helix DNA-binding protein with only euryarchaeal homologues [51]. It controls the expression of the nitrogen fixation (*nif*) operon by binding cooperatively to tandem operators ($OR_1$ and $OR_2$)

located directly downstream of the transcriptional start site. 2-Oxoglutarate, the inducer of the *nif* operon, generates derepression by binding to NrpR and reducing its DNA affinity [52]. $OR_1$, the transcription start-proximal, stronger NrpR binding site suffices for repression when $NH_4^+$ is available. $OR_2$, the transcription start-distal and intrinsically weaker binding site, is cooperatively bound by NrpR together with $OR_1$. $OR_2$ alone is entirely ineffective, presumably because it is located too far downstream for bound NrpR to block RNAP entry to the transcriptional start site and prevent initial steps of RNA chain elongation [51]. However, $OR_2$ and $OR_1$ together generate a differential response to 2-oxoglutarate that allows the establishment of intermediate levels of derepression of the *nif* operon for growth on other nitrogen sources, probably by making $OR_1$ occupancy by NrpR more resistant to the inducer 2-oxoglutarate and making the response of operator occupancy to inducer concentration more graded [52].

Three Lrp family proteins, respectively, from *P. furiosus*, *S. solfataricus* and *M. jannaschii*, have also been shown to repress transcription of their own genes in vitro [53-55]: *P. furiosus* LrpA blocks polymerase entry to the promoter complex and *S. solfataricus* Lrs-14 blocks access to the TATA box and BRE [53, 56]. A different arrangement of cis-acting sites has been found for *S. solfataricus* LrpB, which binds cooperatively to three sites upstream of the TATA box and BRE of its promoter [53]. The presumption that this array of sites generates autoregulation leads to interesting regulatory models that have not yet been put to experimental test [57].

Other proteins with transcription-repressing activities in vitro include the maltose-binding *Thermococcus litoralis* TrmB, whose DNA binding and transcription blocking activity is suppressed by maltose, and *P. furiosus* PhR, which represses transcription of its own gene by blocking polymerase recruitment [58, 59]. Other repressors have been identified through genetic and physiological approaches. Notable among these is the haloarchaeal repressor GvpD [60],

because it may act in a different manner. *Haloferax mediterranei* 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 homologue of the large family of bZIP eukaryotic transcription regulators, and GvpD interact directly. Thus, GvpD may be an anti-activator [61].

Two Lrp family proteins have been proposed as positive regulators of transcription. *S. solfataricus* lysM is part of a large gene cluster encoding lysine biosynthetic enzymes [54]. Transcription of a four-gene segment of this cluster, *lys*WXJK, is induced in a medium lacking lysine. LysM binds immediately upstream of the BRE and TATA box of the *lys*WXJK transcription unit, and lysine diminishes this binding. Thus, LysM might be the activator of this transcription unit and lysine its negative small molecule effector [54]. However, no effect of LysM on transcription of this gene cluster was detected in vitro, either with purified RNAP and transcription factors or in a crude cell extract. A requirement for a missing component was postulated [15].

On the other hand, Ptr2, one of the two Lrp family proteins of *M. jannaschii*, has been shown to activate transcription in vitro at a small number of transcription units encoding proteins that participate in electron transfer processes [55]. The *rb2* (rubredoxin 2 gene) promoter has been the most extensively studied. Ptr2 binds specifically to two adjacent sites upstream of the BRE and TATA box of this weak promoter. Each site deviates from the Ptr2 consensus [62], the promoter-proximal site more so than the distal site (designated sites 1 and 2 respectively, see **Figure 1.3**) [36]. Knocking out either site (essentially) eliminates activation. The Ptr2 sites in *rb2* constitute a bona fide UAS (upstream activating site): placement upstream of a heterologous weak promoter conveys Ptr2-dependent transcriptional activation. Ptr2 generates its transcriptional activation of the *rb2* promoter by facilitating the recruitment of TBP to a weak

TATA box. Transcriptional activation by regulatory proteins in eukaryotes similarly culminates in the recruitment of TBP [36]. 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 [21]. The geometry of the *rb2* UAS is highly constrained, both with regard to the placement of the start site-proximal site 1 (restricted to a center-to-center separation from the TATA box of 21 bp), and the center-to-center separation of sites 1 and 2 corresponding to three or two DNA helical repeats (the two-repeat separation providing a lower level of activation) [63] (**Figure. 1.3**). A single consensus site 1 suffices for strong activation at the *rb2* promoter; another Ptr2-activated promoter, *rbr*, has a Ptr2 UAS that is comparably simple, consisting of a strong, but not consensus, site 1 and a weak site 2 (with a center-to-center site 2–site 1 separation of two DNA turns). In this case also, site 1 suffices for substantial UAS activity [63]. The simplicity of structure of the Ptr2 UAS is remarkable, particularly in view of the potential of the Lrp family proteins for forming higher order arrays on DNA. Even the *E. coli ilvIH* promoter, with its six non-consensus sites for LrpA, is much more complex [64, 65]. In their economy of construction, the *rb2* and *rbr* promoters and activation sites closely resemble the simplest bacterial promoters, at which a single transcriptional effector produces activation by direct interaction with the upstream-facing structure domain 4 of the RNAP holoenzyme's σ-subunit [66]. Indeed, one way of thinking about the action of Ptr2 is to propose that, in working on its eukaryotic-type transcription apparatus, this bacterial family transcriptional activator follows simple bacterial rules [67].

The production of two specialized organelles in haloarchaea is under positive regulatory control: the gas vesicle regulating GvpD has already been referred to; Bat, the positive regulator of the photosynthetic purple membrane has been the subject of genomic and genetic analysis [68]. Bat activates the promoters of a cluster of four transcription units. Transcription initiating at these four promoters is highly (perhaps absolutely) dependent on Bat [68]. Mutagenesis of the promoter of the bacterioopsin gene (*bop*) defines a site upstream of, and close to, the TATA box, comparable with the placement of Ptr2 site 1 at the *rb2* promoter and of LysM at the *lysW* promoter [67]. A scanning mutagenesis of one of the GvpE-activated *Halobacterium salinarum* promoters also identifies a site just upstream of the BRE that is required for activated transcription [69]. It should be recalled that many archaeal genomes encode multiple TBPs and TFBs; *Halobacteria* species hold the current record at six *tbp* and seven *tfb* genes [70]. Paralogous TBPs and TFBs also exist in eukaryotes, but it is the bacterial $\sigma_{70}$ family proteins that promise to be the closer functional equivalent. Accessory $\sigma$ factors provide promoter recognition for transcription of genes associated with specialized functions, such as production of flagella, cell differentiation, adaptation to stationary phase, iron uptake, heat shock and response to diverse extracellular signalling [71]. Upregulation of *tfb* and *tbp* genes in response to UV irradiation [72] and of *tfb* genes in heat shock [73, 74] has been noted. How accessory archaeal TFBs and TBPs are integrated into differential gene expression and regulation, whether they are directed to different promoters and how their presumably differential interactions with transcription activators are generated, is currently unknown [20]. **Table 1.1** summarizes the functionally characterized archaeal transcriptional regulators and **Figure 1.4** illustrates what is known about the modes of archaeal transcriptional activation and repression. Table 1.1 Summary of identified transcription factors in archaea.

| Name | Species | Effector(s) | Mode of regulation |
|---|---|---|---|
| GvpE | *Halobacterium salinarium, Halobacterium mediterranei* | Unknown | Putative activator; mechanism unknown |
| LrpA | *P. furiosus* | Unknown | Repressor; blocks RNAP recruitment |
| Lrs14 | *S. solfataricus* | Unkown | Repressor; blocks TBP/TFB binding |
| LysM | *S. solfataricus* | Lysine | Putative activator; mechanism unknown |
| MDR1 | *A. fulgidus* | Metals ions: $Fe^{2+}$, $Mn^{2+}$, $Ni^{2+}$ | Repressor; blocks RNAP recruitment |
| NrpR | *Methanococcus maripaludis* | 2-oxoglutarate | Repressor; probably blocks RNAP recruitment |
| Phr | *P. furiosus* | Unknown | Repressor, blocks RNAP recruitment |
| Ptr2 | *M. jannaschii, M. thermolithotrophicus* | Unknown | Activator: recruits TBP |
| Sta1 | *Sulfolobus islandicus* | Unknown | Activator (from viral promoter), mechanism unknown |
| Tgr | *T. kodakaraensis* | Unknown | Repressor; probably blocks TBP/TFB binding. Putative activator, mechanism unknown |
| TrmB | *T. litoralis, P. furiosus* | Maltose, trehalose, sucrose, maltodextrins | Repressor; probably blocks TBP/TFB binding |
| TrmBL1 | *P. furiosus* | Unknown | Repressor; probably blocks TBP/TFB binding |
| TrpY | *M.thermoautotrophicus* | Tryptophan | Repressor, blocks TBP/TFB binding |

### 1.3 The model archaeon *Pyrococcus furiosus*

One archaeon that has received a lot of attention is the hyperthermophile *Pyrococcus furiosus.* *P. furiosus* was originally isolated from a shallow marine volcanic vent, and is an anaerobic, heterotrophic, hyperthermophilic euryarchaeon, living optimally around 100 °C [75]. The *P. furiosus* genome sequence of approximately 1.9 Mb in size encoding more than 2000 genes, has been completed [76, 77]. *P. furiosus* ferments either peptides or carbohydrates with the production of organic acids, $H_2$ and $CO_2$. It can also utilize elemental sulfur ($S^0$) as energy source, although the organism grows well in the absence of $S^0$ [75]. Several stimulons in *P. furiosus* have been investigated responding to various environmental stresses using microarray expression profile (transcriptomic) technology [74, 78].

### 1.3.1 What is known about oxidative stress in Archaea

A wealth of genes encoding $H_2O$-producing NADH oxidase (NOX) homologues have been discovered in the genomes of the hyperthermophilic Archaea [79-82], including two homologues in the genome of *P. furiosus* which have been designated as NOX1 and NOX2 [83]. In order to investigate the function of NOX1, the structural gene encoding NOX1 was cloned from the genome of *P. furiosus* and expressed in *Escherichia coli* [79]. NOX1 catalyzes the oxidation of NADH, producing both $H_2O_2$ and $H_2O$ as reduction products of $O_2$ [84]. This is the first NADH oxidase found to produce both $H_2O_2$ and $H_2O$. Transcriptional analysis demonstrated that NOX1 is constitutively expressed regardless of the carbon source and a single promoter was identified 25 bp upstream of the *nox1* gene by primer extension [83, 85]. From this study, *P. furiosus* may tolerate oxygen to some extent and NOX1 may play a key role in the response to oxygen at high temperatures. Another example of oxidative stress is ionizing radiation [86]. *radA*, encoding the archaeal homolog of the RecA/Rad51 recombinant, was moderately up-regulated by irradiation and putative DNA-repair gene cluster was specifically induced by exposure to ionizing radiation [87, 88]. This novel repair system appears to be unique to thermophilic archaea and bacteria and is suspected to be involved in translational synthesis [84].

### 1.3.2 Oxidative stress in *Pyrococcus furiosus*

Oxidative stress is a universal phenomenon experienced by organisms in all domains of life. Proteins like those in the ferritin-like di-iron carboxylate superfamily have evolved to manage oxidative stress. Dps-like protein from the hyperthermophilic archaeon *Pyrococcus furiosus* (PfDps-like) is characterized now [89]. Phylogenetic analysis, primary structure alignments and higher order structural predictions all suggest that the *P. furiosus* protein is related to proteins

within the broad superfamily of ferritin-like di-iron carboxylate proteins. The recombinant PfDps protein self-assembles into a 12 subunit quaternary structure with an outer shell diameter of approximately 10nm and an interior diameter of approximately 5 nm. Dps proteins functionally manage the toxicity of oxidative stress by sequestering intracellular ferrous iron and using it to reduce $H_2O_2$ [90] in a two electron process to form water [89]. The iron is converted to a benign form as Fe(III) within the protein cage [91-94]. This Dps-mediated reduction of hydrogen peroxide, coupled with the protein's capacity to sequester iron, contributes to its service as a multifunctional antioxidant.

**Figure 1.1 Conserved elements of promoter in three domains of life.** Archaea and eukaryotes are more closely related in terms of promoter features. Both contain a TATA box element recognized by TBP and a TF(II)B recognition element (BRE) upstream of the TATA box and RBS (Ribosome binding site). Eukaryotic promoters in addition carry a downstream promoter element (DPE) that is not observed in archaea. Bacterial gene promoters contain two conserved elements located at –35 and –10 nucleotides upstream of the transcription start site. Figure was taken from the reference [95]

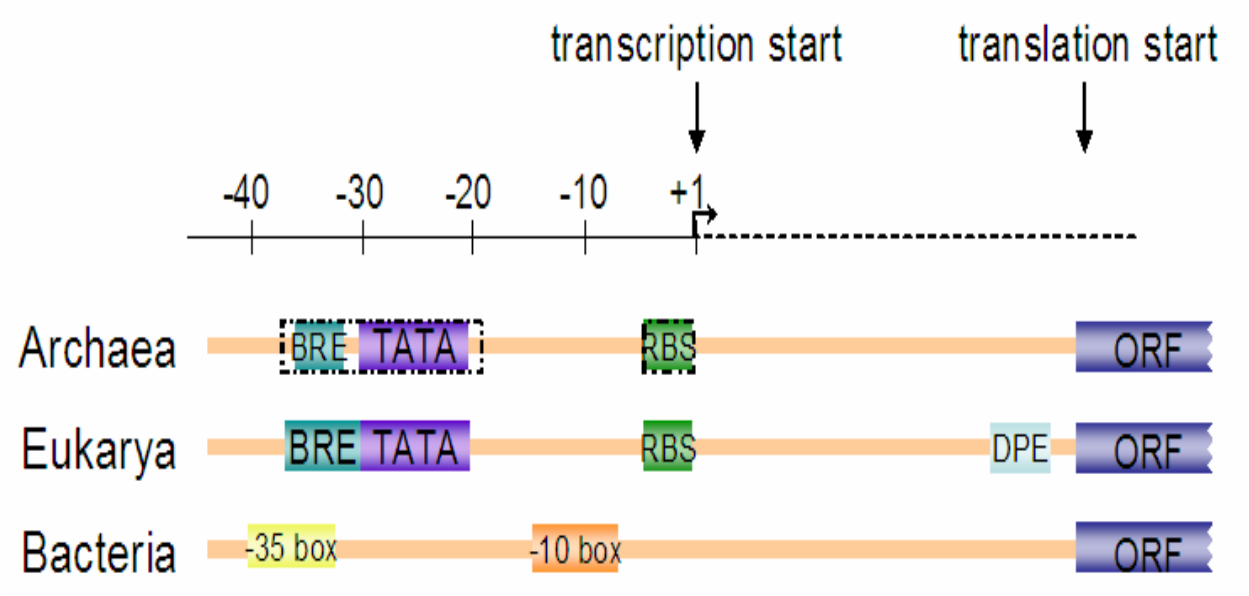

transcription start
translation start
-40
-30
-20
-10
+1
Archaea
BRE
TATA
RBS
ORF
Eukarya
BRE
TATA
RBS
DPE
ORF
Bacteria
-35 box
-10 box
ORF

**Figure 1.2 Basal transcriptional machinery comparison of the three domains.** Archaea, Eukarya and Bacteria. Figure was taken from the reference [96]

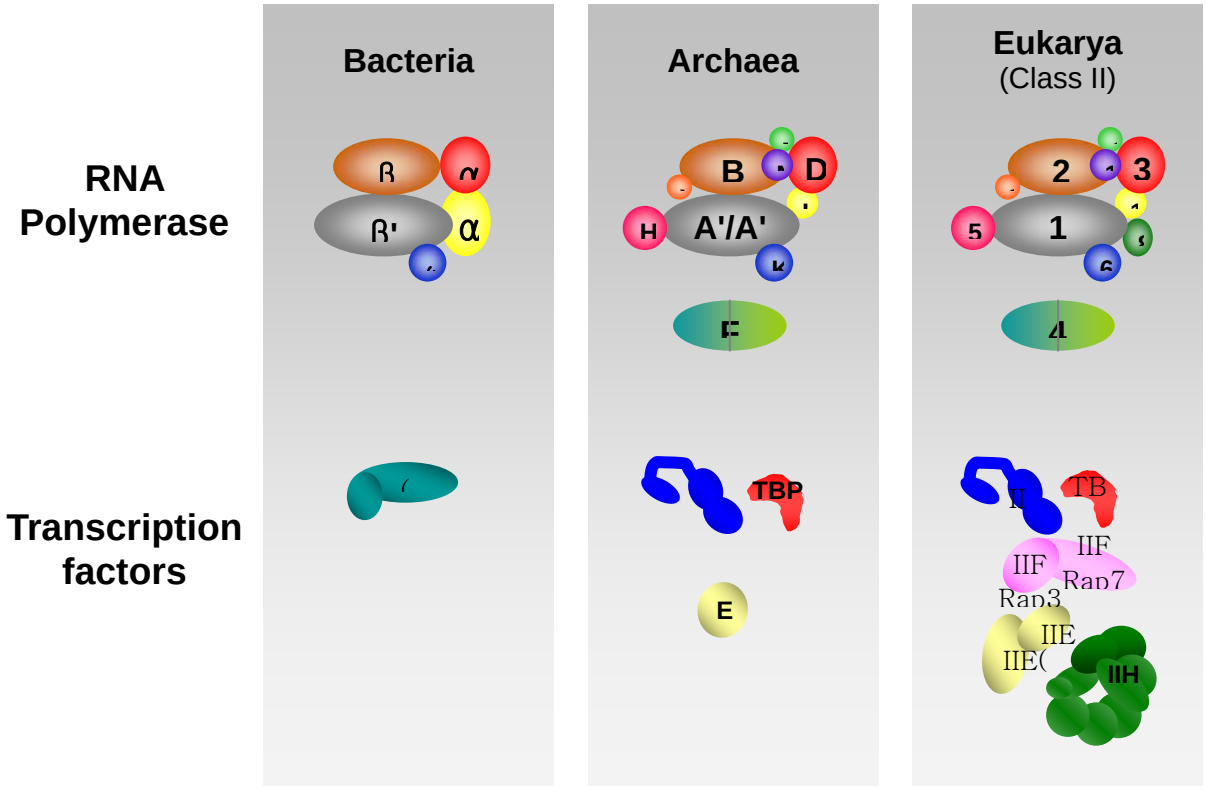

Bacteria
Archaea
Eukarya
(Class II)
RNA
Polymerase
β
α
β'
B
D
H
A'/A'
2
3
5
1
Transcription
factors
TBP
E
IIF
IIF
Rap7
Rap3
IIE
IIE(
IIH

**Figure 1.3 The architecture of the Ptr2 UAS.** The individual Ptr2-binding sites are 6 bp inverted repeats (green) with three central A:T base pairs. All spacings are specified center-to-center. The possibility that Ptr2 may also exert a direct, although site 1 occupancy-dependent, effect on activation of transcription from site 2 has not been excluded and is indicated by a broken green line Figure was taken from the reference [63].

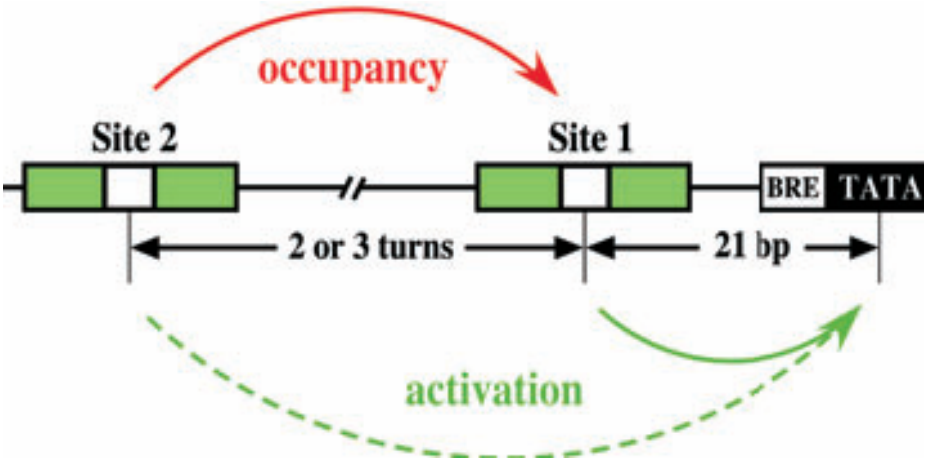
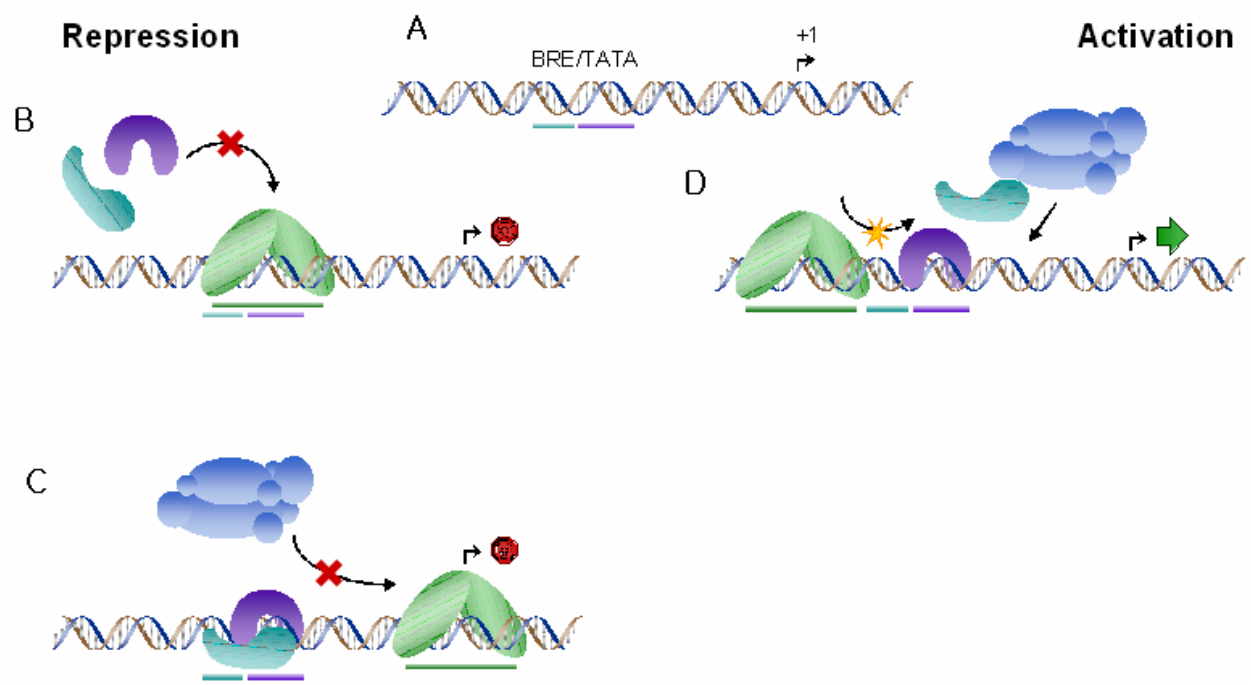

occupancy
Site 2
Site 1
BRE
TATA
2 or 3 turns
21 bp
activation

**Figure 1.4 Archaeal mechanisms of transcriptional repression and activation. A.** An archaeal core promoter consisting of a BRE (teal bar), TATA box (purple bar), and transcription initiation site (bent arrow) **B-D.** Diagrams depicting archaeal transcriptional regulatory mechanisms with the regulatory transcription factor shaded in green (here depicted as a dimer), TBP shaded in purple, TFB shaded in teal, and their corresponding binding sties indicated by bars of the same color below the DNA. The 12-subunit RNAP is shaded in blue. Transcription from a core promoter can be repressed (represented by a red octagon) by two mechanisms: binding of the regulatory transcription factor to a binding site overlapping (**B**) the BRE/TATA region thereby preventing the binding of TBP and subsequently TFB and (**C**) the transcription initiation site thereby preventing the recruitment of RNAP by the TBP/TFB/DNA ternary complex. Transcription activation (represented by a green arrow) has been shown to occur through at least one mechanism, (**D**) facilitating the recruitment of TBP to the core promoter from a binding position upstream of the TATA box. This facilitation speeds up the entire process of transcription initiation, thereby increasing the rate of transcription from the promoter. Note that the bend caused by the binding of TBP to the TATA box is not illustrated in this figure. Activation mechanism (Right); Repression mechanism of Archaea (Left). Figures adapted from Gina lipscomb's dissertation


Repression
A
BRE/TATA
+1
Activation
B
D
C

CHAPTER 2

MATERIALS AND METHOD

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

Two 15-L cultures of *P. furiosus* (DSM 3638) were grown anaerobically at 95 °C essentially as in [97] , using maltose as the carbon source with addition of 5 μM cumene hydroperoxide to only one of the cultures. After approximately 2.5 h of growth, cells were cooled by pumping the culture through a coiled tube in ice water, concentrated by ultrafiltration to 2-3 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 20-30 min on ice. The sonication time was extended to account for not having added DNase I to aid in digesting the genomic DNA; addition of DNase I would have interfered with the downstream application of the cell extract in the DNA affinity protein capture experiment. 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 to remove remaining insoluble and membrane materials. Soluble cell extract was aliquoted into anaerobic vials and stored at -80 °C.

### 2.1.1 Electrophoresis and Gel Staining

2.1.1.1 Protein Electrophoresis

Protein samples were mixed with 1x gel-loading buffer and denatured by heating at 100 °C for 10 min. The sample was loaded into linear gradient SDS polyacrylamide gel (Tris-HCl Criterion

gel, BioRad). The gel was run at 200 V for 60 min and stained with silver staining or Coomassie Brilliant Blue following standard methods (Coligan, 1995) or GelCode Blue.

2.1.1.2 DNA Elctrophoresis

DNA was loaded onto 1.2-1.5 % agarose gel and run at constant 85 V (BioRad power supply) for 55 min and stained and visualized by ethidium bromide.

### 2.2 DNA Quantification

Purified DNA was quantified using a Hoefer DyNA Quant 200 fluorometer (Amersham Pharmacia Biotech) or by comparison with a DNA low mass ladder. Also, the amplified DNAs from PCR (Polymerase Chain Reaction) were quantified using a DyNA Quant 200 Fluorometer with calf thymus DNA as standard.

### 2.3 DNA affinity protein capture

The DNA affinity protein approach is outlined in **Figure 2.1** and **Figure 2.2**. The DNA probes used for protein capture were PCR-amplified from *P. furiosus* genomic DNA using the primers listed in **Table 2.1**. The biotinylated probe was bound to magnetic DynaBeads M-280 Streptavidin (Invitrogen, Carlsbad, CA) per the manufacturer's protocol. Dynabeads are covalently coated with streptavidin proteins (MW 66 kDa), which contains four subunits (16 kDa each ) [98]. Each subunit of streptavidin has a high affinity for biotin or biotinylated molecules. The Dynabeads Streptavidin are supplied in phosphate-buffered saline solution, pH 7.4, containing 0.1% bovine serum albumin (BSA) and 0.02% $NaN_3$ as a preservative [98]. The bead-bound DNA was then mixed with 2.5 mg/mL *P. furiosus* soluble cell extract from cells grown either in the presence or absence of cumene hydroperoxide and incubated at 55 °C for 30 min

with intermittent mixing to keep the beads in suspension. Unbound proteins were eluted with three washes of Buffer B (50 mM EPPs, 100 mM KCl, 1 mM EDTA, 5% glycerol, 0.1% triton-X, 1 mM DTT, pH 7.5). DNA-bound proteins were eluted at 55 °C for 5 min with 1x Laemmli buffer containing no β-mercaptoethanol as this tended to strip the streptavidin from the bead surface. Eluted proteins were analyzed by SDS-PAGE with silver staining. In this study, the biotinylated PF1983 UOR DNA was bound to magnetic DynaBeads M-280 Streptavidin and bead-bound DNA complex was then mixed with 8.0 mg/mL *P. furiosus* soluble cell extract from cells grown either in the presence or absence of cumene hydroperoxide. As a control, DNA from the PF1983 ORF was used in a parallel protein capture experiment. The proteins bound to bead-bound DNA were eluted and analyzed using 10-20% gradient 1D SDS PAGE. In this study, the biotinylated PF1983 UOR DNA was bound to magnetic DynaBeads M-280 Streptavidin and bead-bound DNA complex was then mixed with 8.0 mg/mL *P. furiosus* soluble cell extract from cells grown either in the presence or absence of cumene hydroperoxide. As a control, DNA from the PF1983 ORF was used in a parallel protein capture experiment. The proteins bound to bead-bound DNA were eluted and analyzed using 10-20% gradient 1D SDS PAGE.

**Table 2.1 Probes used in DNA affinity protein capture**

| Probe name | Genome coordinates | Forward primer[a] (5' biotinylated) | Reverse primer[a] |
|---|---|---|---|
| 1983 | 1832907-1833117 | gggagagactagacaactagc | gatttgtggtgaggaggagt |
| 0514 | 535524 - 535724 | tatatgagcacacaaggttttttattttgaactacc | aggcaagaaaaattaaggatattagggacatcc |
| 0357 | 369939-370181 | ctccccagtaaaattccataaacttcccc | gcatgtcacctcattataatttag |

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

### 2.4 In-gel tryptic digestion and peptide mass mapping

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 slices were subjected to three cycles of hydration and dehydration: a 10-min incubation in 25 mM ammonium bicarbonate followed by a 15-min incubation in 50% acetonitrile in 25 mM ammonium bicarbonate. The dehydrated gel slices were then completely dried in a vacuum centrifuge (~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. After rehydration in trypsin a small amount 25 mM ammonium bicarbonate was added to cover the gel slices. Digested peptides were extracted with one ~10-µL wash of 25 mM ammonium bicarbonate followed by two successive ~10-µL washes with 75% acetonitrile, 0.5% trifluoroacetic acid (TFA). The pooled, extracted solution was concentrated to 4-5 µL by vacuum centrifugation, and 1 µL 5% TFA was added to make a final volume of 5-6 µL and a final TFA concentration of 0.1-1%. A $NuTip_{\mu\text{-}C18}$ (Glygen Corp., Columbia, MD) was used to concentrate and purify the trypsin-digested samples prior to depositing on the MALDI target. In this study, the gel plugs were destained and cleaned by repeated washing with alternating acetonitrile and aqueous ammonium bicarbonate solution. After drying by vacuum centrifugation, the gel plugs were subjected to trypsin digestion at 37 °C overnight. The digested peptides were extracted from gel the next day and concentrated with $ZipTip_{\mu\text{-}C18}$ before deposition with matrix (α-cyano-4-hydroxycinnamic acid) on a MALDI target plate.Peptide mass mapping was performed 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). Proteins were identified by peptide mass fingerprinting using MS-Fit program (http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm) and the MASCOT online search

engine (www.matrixscience.com, [99]) to search the NCBI database of archaeal genomes and a local server hosting ProteinProspector [100] to search a *P. furiosus* genome database. In this study, the eluted samples from in-gel tryptic digestion experiment were analyzed by MALDI TOF MS as described before. The MS spectrum was acquired over a range of m/z 600 to 6000 and calibrated using trypsin autolysis products as internal standards. The protein monoisotopic mass (M+H) list was searched against the archaeal genome database in NCBI using the MS Fit program or the machine built-in software, Biotool, which is based on MASCOT searching program. The applied searching mass accuracy was 100 ppm. The number of missed cleavages tolerated was 1. Protein identification required at least 4 matched peptides. Oxidation of methionine was considered as possible modification of peptides during the sample manipulation. Briefly, to explain the principle of peptide mass analysis, Peptide mass mapping (PMM) is a fast, convenient method for protein identification which is widely coupled with gel-based protein separation and in-gel digestion. Every protein with a sequence available in a genome or protein database can be subjected to theoretical digestion by a specific protease. Each protein produces a unique list of peptide fragments, the collection of which is called the "peptide fingerprint" of the protein. The peptide sample derived from in-gel proteolytic digestion by the same protease can be measured by mass spectrometry with high accuracy (now as good as 10 ppm). The protein can be identified when the observed "peptide fingerprint" matches the theoretical one from the database. Therefore, protein identification by PMM largely depends on the protein purity, accurate mass measurement by MS and availability of the gene sequence in the genome database. Obviously, good software with an accurate autoanalysis algorithm is necessary. Several web-based searching programs have been widely used, such as MS Fit and MASCOT [99]. The identification result also depends on the database searching parameters and restrictions in term of

a genome search scale, maximum number of missed cleavages by the protease, allowed peptide modifications, minimum number of required matched peptides, etc. Many times, Peptide mass mapping is sufficient for accurate protein identification according to statistical scores generated by the algorithm. When Peptide mass mapping cannot provide unequivocal identification, other information or techniques can be combined for improving the confidence of the identification result. For example, using multiple searching programs and comparing the results, or referring to molecular weight information of the protein provided from the gel electrophoresis can help.

### 2.5 Sequence analysis for selection of target protein for characterization

Sequence analysis of identified proteins was performed using NCBI BLAST [101, 102] and Conserved Domain searches [103-105] against the NCBI non-redundant protein database.

### 2.6 Overexpression and purification of recombinant his-tagged PF0230p

The vector containing the PF0230 sequence (genome coordinates 242646-243224) was from Francis Jenney (from the laboratory of Michael Adams, University of Georgia). The pET24dBAM vector harboring the clone was a derivative of pET24d, modified to incorporate an N-terminal hexahistidine tag (his-tag) on the expressed protein (e.g., as in [76]).

The pET24d vector is selectable for kanamycin resistance and is designed for use in combination with a host containing a T7 lysogen under control of the *lac* promoter. The pET24d vector also contains a copy of the *lac* repressor which represses expression of the endogenous T7 RNA polymerase (RNAP) except in the presence of the chemical inducer IPTG (isopropyl-β-D-thiogalactopyranoside) which causes release of the repressor from the *lac* promoter, thereby permitting expression of T7 RNAP. Recombinant protein expression is therefore inducible with

IPTG and is driven from the T7 promoter by T7 RNAP under control of the *lac* promoter. For expression of the his-tagged recombinant protein, the clone was transformed into BL21-CodonPlus(DE3)-RIPL cells (Stratagene, La Jolla, CA) using the manufacturer's protocol. Protein expression from a 1-L culture of LB media was grown to an $OD_{600}$ of ~0.7-0.8, and protein expression was induced with final 0.4 mM IPTG. Cells were harvested 5 h after induction and resuspended in ~20 mL Binding Buffer (20 mM sodium phosphate, 0.5 M NaCl, pH 7.4) containing ~10 μL of protease inhibitor cocktail (Sigma, St. Louis, MO). Cells were sonicated on ice using a large horn at 40% power for 6 pulses of 15 s, with capping and mixing of the solution between pulses. Soluble cell extract was obtained after centrifugation for 60 min at 21,000 rpm 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 FPLC system (GE Healthcare, Piscataway, NJ), the soluble cell extract was loaded onto a 1-mL HiTrapFF metal affinity column (GE Healthcare, Piscataway, NJ) preloaded with nickel-sulfate per the manufacturer's instructions. The column was washed with 5 mL Binding Buffer containing 25 mM imidazole followed by a gradient elution with Eluting Buffer (20 mM sodium phosphate, 0.5 M NaCl, 0.5 M imidazole, pH 7.4) first with a 10-mL 0-20% gradient and then with a 10-mL 20-100% gradient. Protein-containing fractions that were relatively pure were pooled, and a 5-mL desalting column (GE Healthcare, Piscataway, NJ) was used for buffer exchange into 20 mM HEPES, 100 mM NaCl, pH 7.6. The resulting protein was estimated to be >98% pure. Protein concentration was determined using a Bio-Rad DC Protein Assay kit, and aliquots of $his_6$-PF0230p were stored at -80 °C.

## 2.7 Electromobility shift assay with PF0230p

Electromobility shift assay (EMSA) was performed as a modification from that originally described in [106] . DNA probes for EMSA were PCR-amplified from *P. furiosus* genomic DNA using primers listed in **Table 2.2**, followed by either PCR purification using a PCR Purification Kit (Qiagen, Valencia, CA), gel purification using a Qiaquick Gel Purification Kit (Qiagen, Valencia, CA), or ethanol precipitation. EMSA reactions of DNA with various amounts of protein were set up in 10-μL volumes in EMSA buffer (20 mM HEPES, 200 mM KCl, 5% glycerol, 1 mM EDTA, pH 7.5) using a 5x stock. DNA concentration was typically 75-100 ng/μL in each reaction, and protein was adjusted according to the molar amount of DNA. The EMSA reactions were assembled as follows. A master mix of water, 5x EMSA buffer and DNA was made according to the number of reactions in the experiment (typically 14 total, to be loaded into a 15-well gel with one gel lane reserved for a DNA marker), then distributed to 0.5-mL microcentrifuge tubes on ice. Protein dilutions were made in a final concentration of 1x EMSA buffer, and 2 μL of the appropriate protein dilution was added to each EMSA reaction (with 2 μL of 1x EMSA buffer added instead of protein for the DNA-only lane). In cases where an extra reagent was added to the reaction (e.g. cumene hydroperoxide or hydrogen peroxide or metals), volumes of water and/or 5x EMSA buffer were adjusted accordingly such that the final buffer concentration of each reaction was always 1x. Reactions were incubated at 55 °C for 20 min and immediately loaded onto a BioRad 5% TBE gel; 15-well Ready gels were typically run at 200 V for 20-30 min while the 26-well Criterion gels (BioRad, Hercules, CA) were typically run at 100 V for 60-110 min. The gel was then stained with SYBR Green Nucleic Acid Gel Stain (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. SYPRO Ruby protein gel stain (BioRad, Hercules, CA) was used according to the manufacturer's instructions to verify

the presence of protein as necessary. 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) |
|---|---|---|---|---|
| 1983 | 1832907-1833117 | gggagagactagacaactagc | gatttgtggtgaggaggagt | 211 |
| 1983 ORF | 1833208-1833441 | tgggaaatactttggagggaggct | ataaaggctaaaggcctcccgcaa | 234 |

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

### 2.8 Footprinting for PF0230p with DNase I

Fluorescence footprinting was performed as a modification from (**Figure 2.3**) [107] based on the DNase I footprinting method [108]. Footprinting probes were PCR-amplified from genomic DNA using 5' 6FAM and HEX labeled modified primers for analysis on a 3730x1 automated DNA sequencer (Applied Biosystems, Foster City, CA). The forward primer had to be extended at the 5' end to ensure that the base adjacent to the fluorophore would not be a guanine, as guanine can quench the fluorescence of fluorophores at that proximity. Probes were amplified using Taq polymerase (Sigma-Aldrich, Piscataway, NJ) for a maximum of 25 cycles to minimize the amount of non-specific PCR products. Typically, six 50-µL PCR reactions were performed for each probe, and these were then concentrated by ethanol precipitation prior to gel-purification. To purify away primers and truncated PCR products, probes were separated on BioRad 5% TBE gels (18-well Criterion gels run at 100V for 60-110 min were optimal). Gels were stained with SYBR Green I Nucleic Acid Gel Stain (Lonza, Basel, Switzerland) for 10 min prior to visualization under long-range UV light. Bands were carefully excised, with special attention to cutting the lower side of the band as close as possible to the bulk of DNA thereby eliminating shorter DNAs which would interfere with footprinting results. Probe DNA was eluted from the polyacrylamide gel slices using the crush-and-soak method [109]. Resulting probe DNA was concentrated by ethanol precipitation and quantified prior to use in footprinting reactions.

Footprinting reactions were composed of two parts, the protein-DNA binding reaction and the cleavage reaction. The protein-DNA incubations were set up similar to the EMSA reactions except the reactions were set up in 50-µL volumes with ~150-200 ng DNA probe, incubations of individual reactions were separated by 2- or 3-min intervals, and the buffers were different. The following buffers were used: for DNase I footprinting, 20 mM HEPES, 100 mM KCl, 15 mM $MgCl_2$, 5 mM $CaCl_2$, 1 mM EDTA, 1 mM DTT, 5% glycerol, pH 8. Protein-DNA solutions were incubated for 20 min at 55 °C. For the DNase I cleavage reaction, 0.03-0.05 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 vortexing. After DNase I cleavage reactions, 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 Applied Biosystems automated DNA sequencer at the Sequencing and Synthesis Facility (University of Georgia). 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, Inc., Lake Oswego, OR).

### 2.9 SELEX of PF0230p consensus DNA binding motif

A modification of the artificial selection method termed SELEX (Systematic Evolution of Ligands by Exponential Enrichment) [110, 111] was applied to determine the consensus DNA recognition sequence of PF0230p and the scheme is shown in **Figure 2.4**. This method involves the use of an artificial library of DNA containing random sequences to allow for elucidation of the PF0230p DNA binding site through successive cycles of selection with PF0230p. The single-stranded SELEX probe from which the artificial library was generated was designed with a 30-nucleotide randomized region flanked by constant primer regions, each containing three restriction sites (XbaI, EcoRI, HindIII on the 5' side and BamHI, EcoRI, SalI on the 3' side). The double-stranded SELEX probe was PCR-amplified from the synthetic single-stranded oligonucleotide with primers that slightly extended the original SELEX probe length (thereby eliminating 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 2.3. To create the dsDNA probe, 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 [109]. Selection rounds were set up essentially as for the EMSA reactions, except for the amount of SELEX probe used and the protein-DNA ratios. For the first round of selection, 0.6 μM of SELEX probe was used, and for all succeeding selection rounds, 0.1 μM was used; protein concentrations ranged from 0.6-1.2 μM. After each selection round, DNA was purified from shifted protein-DNA complexes, amplified with the SELEX primers using 15 cycles of PCR, and polyacrylamide gel-purified before proceeding to the next selection round. A total of 6 selection rounds were performed in this manner. The selected DNA was

digested with EcoRI, concatemerized, and cloned into the pUC18 standard cloning vector. Blue/white color screening and colony PCR were used to identify colonies that contained plasmids with the largest concatemers, and plasmid was isolated from these colonies for sequencing. A total of 21 sequences were obtained from the round 6 selected DNA, and a total of 6 sequences were obtained from round 5 selected DNA. These sequences were input into MEME online motif searching software [112] to elucidate a common motif among the selected DNA, and a graphical representation of the motif was generated using WebLogo [113].

**Table 2.3 DNA probe and primers used for SELEX**

| Name | Sequence[a] |
|---|---|
| SELEX single-stranded probe | ggtctagagaattcaagcttc(n)$_{30}$ggatccgaattcgtcgac |
| SELEX primer F | gctcaggtctagagaattcaa |
| SELEX primer R | actactgtcgacgaattcgga |

[a]DNA primers are listed from 5' to 3'. HindIII, EcoRI sites used in cloning are colored blue and red

### 2.10 Generation of a his-tag cleavable PF0230p gene construct

A modified version of the pET24dBAM 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). Protein expression from this vector, termed pET24dBAM-TEV, generated a recombinant protein with a cleavable his-tag. The pET24dBAM vector harboring the PF0230 sequence was used to subclone the PF0230 sequence into pET24dBAM-TEV since the same pair of restriction sites could be used to transfer the insert from one vector to the other. The pET24dBAM-*PF0230* plasmid was amplified in XL1-Blue cells and purified using a Plasmid Miniprep Kit (Qiagen, Valencia, CA). The parent plasmid and destination vector were digested with BamHI and NotI restriction enzymes for 3 h at 37 °C: Digestion products were then separated on a 1% agarose gel, and bands of insert and linear

destination vector were excised and gel-purified using and a QIAEX II Gel Extraction Kit (Qiagen, Valencia, CA). Gel-purified insert and linearized vector were quantified and ligation reactions were set up as follows: 100 ng linearized vector was combined with varying mole ratios of insert (1:1, 1:2, 1:3), together with buffer and 0.5 Weiss units of T4 DNA ligase in a total volume of 20 μL. Ligation reactions were incubated at room temperature for 3 h and immediately transformed (5 μL) by heat-shock into $CaCl_2$-competent XL1-Blue cells. After the 1-h incubation of the transformation culture, volumes of 50 and 200 μL were spread on agar plates containing 30 μg/mL kanamycin. Plates were incubated at 37 °C for 18 h, and resulting colonies were picked and streaked onto fresh selective plates. These streaks were used to inoculate 5-mL cultures for plasmid production. Plasmid was verified to contain the insert by restriction mapping and was then transformed by heat-shock into $CaCl_2$-competent BL21-CodonPlus(DE3)-RIPL cells (made from a stock obtained from Stratagene, La Jolla, CA) for protein expression.

### 2.11 Site-directed mutagenesis of PF0230p ATTAAT binding motif

In order to determine whether the ATTAAT motif in the PF0230p footprinting region and SELEX plays a role in PF0230p DNA-binding activity, single strand 60-bp nucleotides were created in which ATTAAT were mutated to GCGCGC. These single strand nucleotides were ordered from Integrated DNA Technologies (Athens, GA) and dissolved in STE Buffer (10 mM Tris, 50 mM NaCl, 1 mM EDTA, pH 8.0). The presence of some salt is necessary for the oligonucleotides to hybridize and dissolve at high concentration (1 - 10 $OD_{260}$ units / 100 uL). After this, the two strands were mixed together in equal molar amounts and these mixtures were heated to 94 °C and gradually cooled. Usually, for many oligonucleotides this can be as simple

as transferring to the benchtop at room temperature. This is easily done by placing the oligonucleotides in a water bath or constant-temperature block and "unplugging the machine". The resulting product was in stable, double-stranded form and stored at 4 °C or frozen [114].

The primers and their complements are listed in Table 2.4.

**Table 2.4 Oligonucleotides used for site-directed mutagenesis of PF1983 ATTAAT motif**

| Name | Sequence[a] |
|---|---|
| ATTAAT Oligo F | atgagctggggcttgaatttataaaggattaatttcatctgaatagtttatttctttatgc |
| ATTAAT Oligo R | tactcgaccccgaacttaaatatttcctaattaaagtagacttatcaaataaagaaatacg |
| GCGCGC Oligo F | atgagctggggcttgaatttataaagggcgcgcttcatctgaatagtttatttctttatgc |
| GCGCGC Oligo R | tactcgaccccgaacttaaatatttccgcgcgaagtagacttatcaaataaagaaatacg |

[a]DNA primers are listed from 5' to 3'. Mutated nucleotides are colored red.

## 2.12 EMSA for PF0230p binding ability with TATA box

To determine whether PF0230p binds to the TATA box DNA sequence, two different probes as described in Table 2.4 were used. Hybridization of ATTAAT Oligo F and ATTAAT Oligo R generated dsDNA named ATTAAT-PF1983; hybridization of GCGCGC Oligo F and GCGCGC Oligo R generated dsDNA named GCGCGC-PF1983. Four EMSA lanes for each dsDNA contained TBP, TFB, TBP+TFB, and TBP+TFB+PF0230p. EMSA was performed for 20 min at 55 °C based on these conditions

## 2.13 Analytical gel filtration to determine PF0230p quaternary structure

Analytical gel filtration using a Superdex 75 or Superdex 200 10/30 GL size exclusion column (GE Healthcare, Piscataway, NJ) was performed on various PF0230p samples to determine their quaternary structure. A 200-µL sample of 2.5 mg/mL protein was prepared. The running buffer used was 20 mM HEPES, 200 mM NaCl, pH 7.6. Molecular weight standards (200 µL of each) were run through the column individually and the elution volume ($V_e$) of each was noted. The

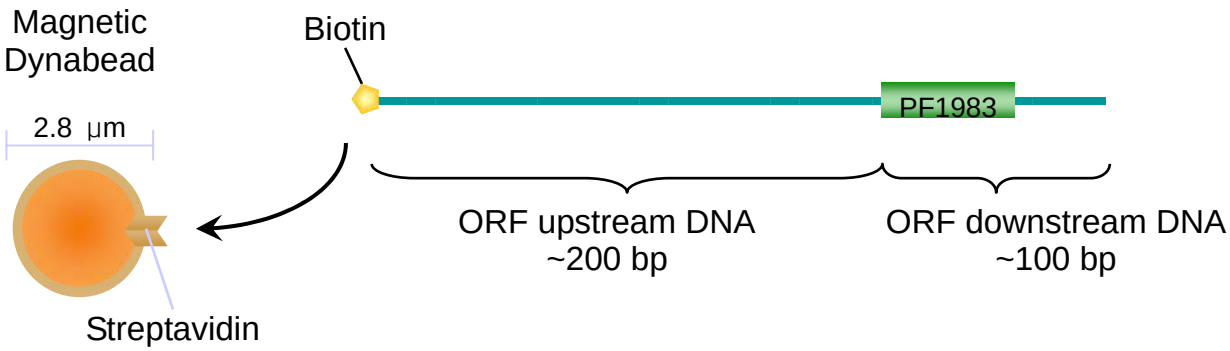
following standards (Sigma, St. Louis, MO) were used: blue dextran (2 mg/mL), bovine serum albumin (10 mg/mL), carbonic anhydrase (3 mg/mL), cytochrome *c* (2 mg/mL), and vitamin $B_{12}$ (0.1 mg/mL). Blue dextran was used to determine the column void volume ($V_0$), and a standard curve of molecular weight versus $V_e/V_0$ was used to determine the corresponding approximate molecular weights of the sample peaks.

### 2.14 Expression and purification of his-tag cleavable PF0230p for crystallization

The autoinduction protein expression protocol established by Studier [115] was followed for expression of his-tag cleavable PF0230p for crystallization trials. Briefly, one 50-mL culture grown for 8-10 h at 37 °C in P-0.5G media was used to inoculate 1 L of ZYP-5052 media 20 [116] which was then divided equally into two 2-L flasks. The large-scale cultures were grown for 18-20 h at 37 °C before harvesting at 6,000 × g at 4 °C for 15 min. Cell extract was prepared as described in Section 2.5, and nickel affinity chromatographic separation of the his-tagged protein using a HisTrapFF column (GE Healthcare, Piscataway, NJ) was accomplished using an ÄKTA system (GE Healthcare, Piscataway, NJ) with essentially the same method and buffers listed in section 2.5. Fractions contained $his_6$-TEV-PF0230p were pooled for concentration and buffer exchange into 20 mM HEPES, 200 mM NaCl, pH 7.6, using an Amicon Ultra-15 centrifugal filter device with a 30 kDa molecular weight cut-off (Millipore, Temecula, CA). The partially purified his-tagged protein was subjected to his-tag cleavage using *Ac* TEV protease (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions, except that less protease was used in conjunction with a longer incubation time. Approximately 30 mg of his-tagged protein (obtained from 1 L of culture) was digested with 500 units of *Ac* TEV protease in a volume of 2-3 mL using reaction buffer and DTT supplied with the protease according to

manufacturer instructions. The cleavage reaction was incubated at 30 °C with shaking at ~65 rpm for 10-14 h. Following his-tag digestion, the protein sample was applied directly to a fresh nickel affinity HisTrapFF column (GE Healthcare, Piscataway, NJ) equilibrated with 20 mM HEPES, 200 mM NaCl, pH 7.6, using a 0.2 mL/min flow rate. The column flow-through containing the tagless protein was collected, while the column-bound uncleaved protein, his-tagged TEV protease, and *E. coli* proteins were eluted with Elution Buffer. SDS-PAGE was used throughout the purification process to monitor purification and his-tag cleavage and to check for protein stability and purity. For protein purified for use in crystallization trials, a gel-filtration polishing step was performed mainly to remove protein aggregates. A HiPrep 26/60 Sephacryl S-100 High Resolution (GE Healthcare, Piscataway, NJ) gel filtration column was equilibrated with two column volumes of 20 mM HEPES, 200 mM NaCl, pH 7.6, and protein sample was injected onto the column at a flow rate of 0.7 mL/min. Protein-containing fractions were pooled and concentrated to 21.1 mg/mL using an Amicon Ultra-15 centrifugal filter device with a 30 kDa molecular weight cutoff (Millipore). Protein concentration was determined using a Bio-Rad DC Protein Assay kit. Quentin Florence (John Rose's Laboratory, University of Georgia) used this protein for crystallization trials.

**Figure 2.1 Design and immobilization of probe DNA.** The probe DNA is designed to have approximately 200 bp upstream from translation start site and probe DNA is amplified from genomic DNA using one biotinylated labeled primer and one unlabeled primer such that the PCR-amplified probe contains a biotin on 5' end so that the DNA can be bound to streptavidin-coated magnetic beads (Dynabeads M-280 Streptavidin, Invitrogen). The dissociation constant between biotin and streptavidin is approximately $10^{-14}$. 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.


Magnetic
Dynabead
2.8 µm
Biotin
PF1983
ORF upstream DNA
~200 bp
ORF downstream DNA
~100 bp
Streptavidin

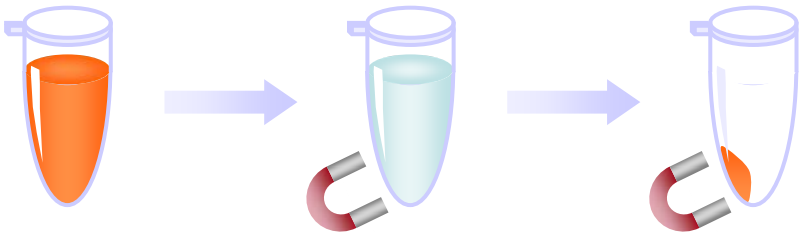


**Figure 2.2 The DNA affinity protein capture experiment.** Biotinylated DNA is bound to magnetic streptavidin-coated beads. The bead-DNA complex is then incubated with soluble cell extract, and some proteins associate with the DNA including basal transcriptional machinery (RNAP subunits, etc), non-specific DNA-binding proteins, and other transcription factors. Proteins which do not bind DNA are removed, and finally the DNA-binding proteins which remain are eluted and analyzed.

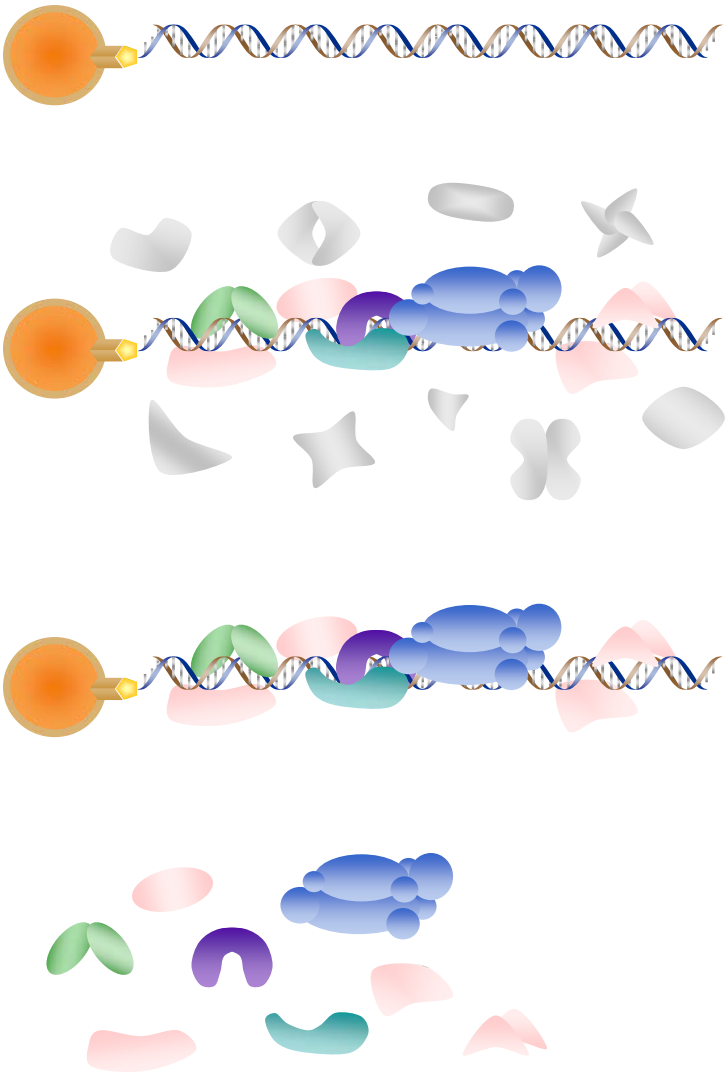

Incubate with cell extract
Wash away unbound proteins
Elute bound proteins

**Figure 2.3 Fluorescence-detected DNase I footprinting scheme.** A protein-bound fluorescently-tagged probe (6FAM) is nicked with DNase I, and the region of DNA bound by the PF0230p is protected from cleavage. Strands are denatured prior to undergoing capillary electrophoresis on a capillary sequencer (ABI 3730x1, Applied Biosystems). Fluorescently-labeled fragments are detected, and resulting electropherograms of samples with and without added protein are overlaid to determine footprint position.

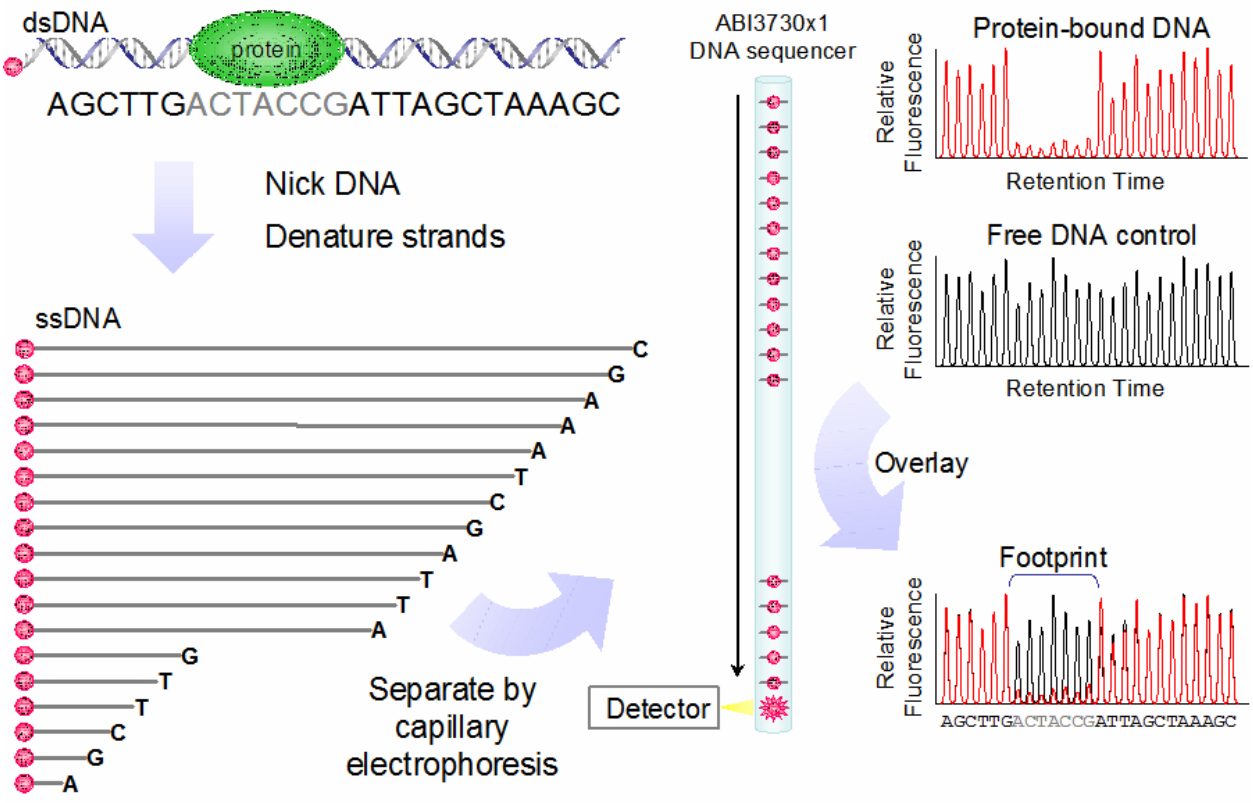
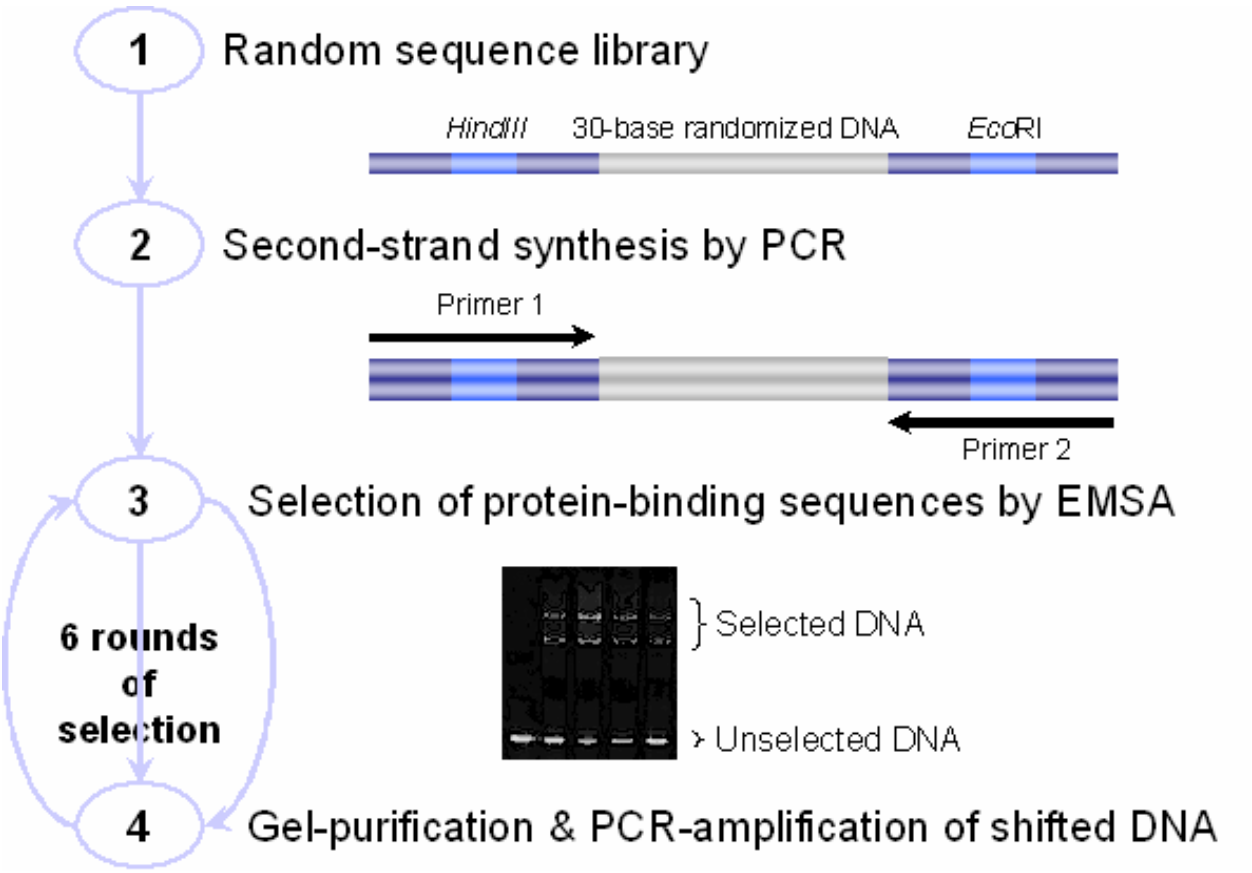
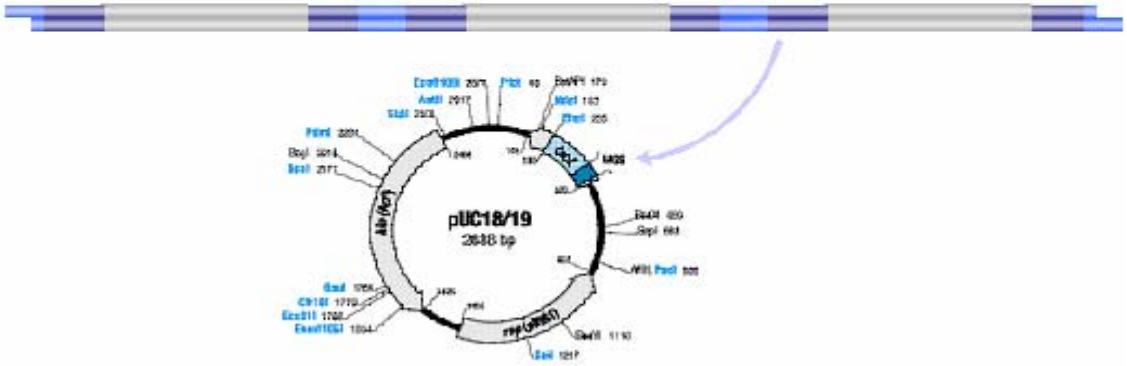

dsDNA
protein
AGCTTGACTACCGATTAGCTAAAGC
Nick DNA
Denature strands
ssDNA
C
G
A
A
A
T
C
G
A
T
T
A
G
T
T
C
G
A
Separate by capillary electrophoresis
ABI3730x1
DNA sequencer
Detector
Protein-bound DNA
Relative Fluorescence
Retention Time
Free DNA control
Relative Fluorescence
Retention Time
Overlay
Footprint
Relative Fluorescence
AGCTTGACTACCGATTAGCTAAAGC

**Figure 2.4 SELEX scheme. (1)** The SELEX library was made up of single-stranded oligonucleotides designed to contain 30 bases of random DNA flanked with two ~20 bp sites each containing an *Eco*RI and HindIII restriction site. **(2)** Second-strand synthesis and amplification of the library was performed using primers complementary to the two priming sites. **(3)** SELEX probes with higher-affinity sequences to PF0230p were selected from the library pool using EMSA. **(4)** Shifted DNA was gel-purified and PCR-amplified for an additional selection round. 6 rounds of selection via EMSA, gel-purification, and PCR-amplification were carried out. **(5)** Resulting selected DNA was digested with *Eco*RI and Hind III **(6)** cloning into pUC18. Cloned plasmids were transformed into XL1-Blue cells, and insert-containing plasmids were identified via blue/white colony screening. **(7)** Clones were then sequenced to identify the binding motif.


1
Random sequence library
HindIII
30-base randomized DNA
EcoRI
2
Second-strand synthesis by PCR
Primer 1
Primer 2
3
Selection of protein-binding sequences by EMSA
6 rounds of selection
Selected DNA
Unselected DNA
4
Gel-purification & PCR-amplification of shifted DNA

5 Digestion with HindIII and *Eco*RI

6 Cloning




7 Sequencing of clones & motif identification

CHAPTER 3

DISCOVERY OF A NEW REGULATORY TRANSCRIPTION FACTOR PF0230P

### 3.1 DNA microarray expression profiles identify genes up-regulated under oxidative stress

Whole-genome microarray analyses of mRNA transcript levels (transcriptomics) were investigated in this work to study the genes and regulatory pathways involved during oxidative stress (the presence of cumene hydroperoxide, CHP) in *P. furiosus*. **Table 3.1** lists the microarray expression profile obtained for *P. furiosus* batch cultures grown in the presence or absence of CHP. A total of 11 ORFs showed greater than 5-fold up-regulation when CHP was added to the culture medium. PF1983, encoding a hypothetical protein, and PF0514, encoding D-alanine glycine permease were strongly induced under oxidative stress. As a result, these two ORFs were selected as targets for transcription factor discovery. Proteins that bound to the DNA upstream of these ORFs were identified from cell extracts of cultures grown with and without cumene hydroperoxide using the DNA affinity protein capture method described in Chapter 2. One of the proteins identified by this method was PF0230p, the subject of this dissertation.

**Table 3.1 DNA microarray expression profiles of *P. furiosus* grown +/–CHP**[a]

| ORF | ORF description / operon function | Expression +CHP/–CHP |
|---|---|---|
| PF0514 | D-alanine glycine permease | 14.9 |
| PF1983 | Conserved hypothetical protein | 10.1 |
| PF0571 | Conserved hypothetical protein | 10.0 |
| PF0581 | Hypothetical protein | 9.4 |

| | |
|---|---|
| PF0401 Methyltransferase | 9.1 |
| PF1975 Conserved hypothetical protein | 8.8 |
| PF0744 abc transporter (msbA subfamily) | 7.9 |
| PF0935 [Acetolactate synthase] | 7.5 |
| PF0938 3-Isopropylmalate dehydrogenase | 6.2 |
| PF1528 Imidazoleglycerol-phosphate synthase | 5.6 |
| PF1592 Tryptophan synthase, beta | 5.5 |

[a]ORFs up-regulated by ≥ 5.5-fold

3.1.1 Design, amplification, and purification of DNA probes

With the premise that archaeal transcriptional regulation has similarities with bacterial mechanisms, many transcription factor binding sites should exist just upstream of the ORF (Open Reading Frame) translation start site, near promoter elements. Based on this, we generally design UOR (Upstream of ORF Region) DNA probes spanning ca. –200 to +100 bp relative to the ORF start. The PF1983 UOR DNA spanned -200 to +11 (relative to the PF1983 translation start site) with a biotin linker of 8 bp at the 5' end (**Figure 3.1**). This UOR contained a TATA Box, BRE (TFB-Recognition Element), which together make up an archaeal promoter, and a ribosome binding site (RBS). As a control, DNA from the PF1983 ORF was amplified from +106 to +340 relative to translation start, also with a biotin linker at the 5' end (**Figure 3.1**). The PF0514 UOR DNA probe was amplified from -200 to +1 relative to the PF0514 translation start site with a biotin linker at the 5' end and the PF0514 control DNA was chosen from +200 to +349 relative to translation start site, also with a biotin linker at the 5' end (**Figure 3.2**). Each DNA (PF1983, PF0514 UOR and control DNAs) was amplified by PCR from *P. furiosus* genomic DNA and synthesized primers.

### 3.2 DNA affinity protein capture identifies transcriptional regulator PF0230p

Each of the UOR and control ORF DNA probes were attached to magnetic beads and used in the DNA affinity protein capture experiment described in Chapter 2 to separate proteins from cell extract (either growth with or without cumene hydroperoxide) that bind to these DNA probes. The mixtures of proteins retained by the immobilized DNA were separated by one-dimensional SDS PAGE. The results of gel separation and densitometry are shown in **Figure 3.3** (PF1983) and **Figure 3.4** (PF0514). Using densitometry, protein intensity ratios between UOR and control DNA in the two growth conditions (+CHP/-CHP) were compared. Generally, we expect that a putative transcription factor will have different DNA affinity or be present in different amounts in the two growth conditions. We also compared the amounts of the DNA-bound proteins between the control ORF and the UOR DNA using cell extract from the same growth condition. Putative transcription factors should show sequence-specific binding to sites on UOR DNA, but not on control DNA.

Proteins of interest were identified by the method of peptide mass mapping, as described in Chapter 2, and the results are summarized in **Tables 3.2 and 3.3**. These proteins were the candidates for specific UOR-binding proteins. From Figure 3.3, bands 2, 3, 5, 6, 9, appear to bind more strongly to PF1983 UOR from cell extract from growth in the presence of cumene hydroperoxide, compared to growth in the absence. Of these, the only ones that resemble a transcriptional regulator were PF0230p, band 2, and PF1473p, band 3. We conducted the same DNA affinity protein capture experiments with PF0514 UOR and control DNA and identified two proteins (Figures 3.4, Table 3.3). Band 1 appear to bind more strongly to PF0514 UOR from cell extract from growth in the absence of CHP, while band 2 appears to bind more strongly to PF0514 UOR from cell extract from growth in the prsence of CHP. However, these proteins have

no recognizable homology to transcription factors and were not considered further.

**Table 3.2 Identification of protein bands of Figure 3.3 using peptide mass mapping**

| Band number[a] | Annotation*b* | Locus | Molecular weight (kDa) | Score[b] | Sequence coverage[c] |
|---|---|---|---|---|---|
| 1 | conserved hypothetical protein | PF0977 | 21.3 | 65 | 36.4 % |
| 2 | transcriptional regulatory Protein | PF0230 | 23 | 145 | 98.2 % |
| 3 | hypothetical protein | PF1473 | 27 | 86 | 80.9 % |
| 4 | conserved hypothetical protein | PF0496 | 30.6 | 75 | 66.7 % |
| 5 | quinolinate synthetase | PF1977 | 34.4 | 80 | 77.3 % |
| 6 | cell division protein ftsZ homolog | PF1507 | 39.9 | 77 | 81.5 % |
| 7 | hypothetical protein | PF1967 | 47 | 90 | 72.7 % |
| 8 | hypothetical protein | PF1725 | 50.2 | 60 | 55.4 % |
| 9 | thermosome | PF1974 | 59.9 | 78 | 76.3 % |
| 10 | hypothetical protein | PF0961 | 82.1 | 112 | 89.2 % |

[a]Numbers correspond to numbered bands in Figure 3.3.
[b]MASCOT score. A score over 60 indicates a significant match.
[c]Seq% is the sequence coverage of matched peptides from the identified protein

**Table 3.3 Identification of protein bands of Figure 3.4 using peptide mass mapping**

| Band number[a] | Annotation[b] | Locus | Molecular weight (kDa) | Score[b] | Sequence coverage[c] |
|---|---|---|---|---|---|
| 1 | transposase | PF0536 | 27.7 | 62 | 35.2 % |
| 2 | hypothetical protein | PF0917 | 19.6 | 81 | 78.2 % |

[a]Numbers correspond to numbered bands in Figure 3.4.
[b]MASCOT score. A score over 60 indicates a significant match.
[c]Seq% is the sequence coverage of matched peptides from the identified protein

### 3.3 Other identified UOR Binding Proteins

Several PF1983 UOR-binding proteins were annotated as hypothetical proteins but also contained conserved transcriptional regulator domains [PF0977 (band 1), PF1473 (band 3), PF0496 (band 4), PF1967 (band 7), PF1725 (band 8), and PF0961 (band 10)]. However, all of these proteins either bound to control PF1983 ORF DNA as well as UOR DNA, or bound equally from the two cell extracts in the presence or absence of cumene hydroperoxide (Figure 3.3). Given that PF0230p was the only differentially captured transcription-related protein that did not bind to the control DNA, we chose to focus our efforts on this putative transcriptional regulator.

### 3.4 Overexpression and purification of PF0230p

The PF0230 gene fused with a hexahistidine tag at the N-terminus was cloned into the modified kanamycin-resistant plasmid pET24d (by M.W.W. Adams' laboratory). The cloned gene was transformed into *E.coli* and overexpressed in 2.5% LB media containing 50 μg/mL kanamycin. In this experiment, 0.4 mM IPTG was added to induce protein expression when $OD_{600}$ reached 0.8 and the protein was overexpressed for 5 h at 37 °C after adding IPTG. The cells were harvested by centrifugation and suspended in 20 mL Hitrap FF column binding buffer 1 (0.02 M sodium phosphate, 0.5 M NaCl, pH 7.6) containing 0.1 mM PMSF. Cells were lysed by sonication and the lysate supernatant was loaded on $Ni^{2+}$-charged Histrap FF column equilibrated with column binding buffer 1 containing 25 mM imidazole. The his-tagged PF0230p was eluted by a gradient of imidazole (Figure 3.5).

Only the fractions which were acceptable purity were collected and pooled for buffer exchange using a desalting column with 20 mM HEPES, pH 7.6, 0.2 M NaCl. The final purified

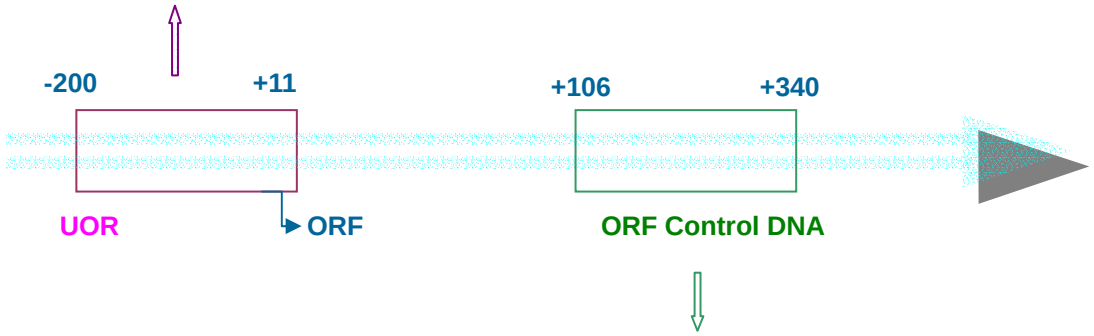
protein was confirmed by SDS-PAGE (Figure 3.5) and MS analyses. PF0230p protein sequence is 192 amino acids in length and has a calculated molecular weight of 22,859 Da; however, with the addition of a his-tag (Met-Ala-$His_6$-Gly-Ser-) at the N-terminus, the new molecular weight is 23,897 Da. After the protein was exchanged into a suitable salt-containing buffer, small aliquots of the protein were stored at -80 °C to be used for characterization.

**Figure 3.1 Sequence of UOR and control DNA for PF1983.** Top, partial sequence of PF1983 UOR, corresponding to the purple shaded region, consisting of 211 bp (-200 to +11, relative to translation start site), containing a predicted ribosome binding site (red) and translation start site (bold). Bottom, sequence of PF1983ORF control DNA, corresponding to the green shaded region consisting of 234 bp (+106 to +340, relative to translation start).

UOR -200 to +11 relative to translation start site, 211bp

RBS

**5' biotin—**GTAGGGGAGAGACTAAGACAA----------TTGGTGATTT**GTG--**



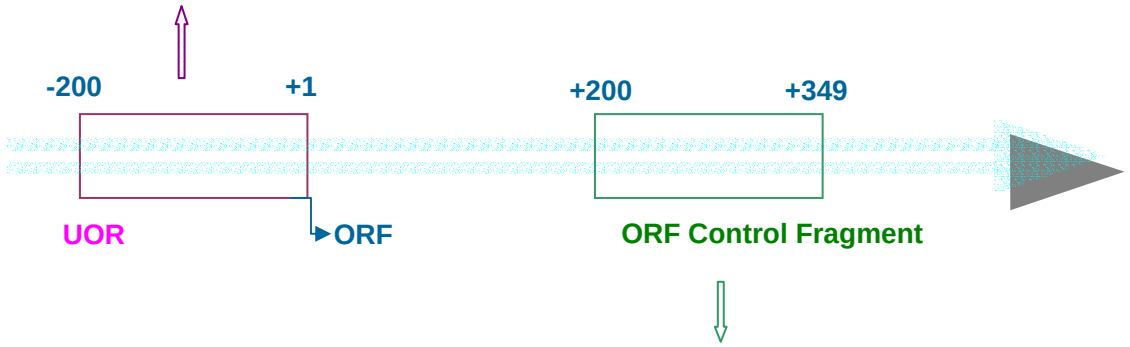

Control DNA +105 to +338 relative to translation start site, 234bp

**5' biotin–** TGGGAAATACTTTGGAGGGAGGCT----------ATAAAGGCTAAAGGCCTCCCGCAA

**Figure 3.2 Sequence of UOR and control DNA for PF0514.** Top, sequence of PF0514 UOR, corresponding to the purple shaded region, consisting of 201 bp (-200 to +1, relative to translation start site). Bottom, sequence of PF0514ORF control DNA, corresponding to the green shaded region consisting of 250 bp (+100 to +349, relative to translation start).

UOR -200 to +1 relative to translation start site, 201bp

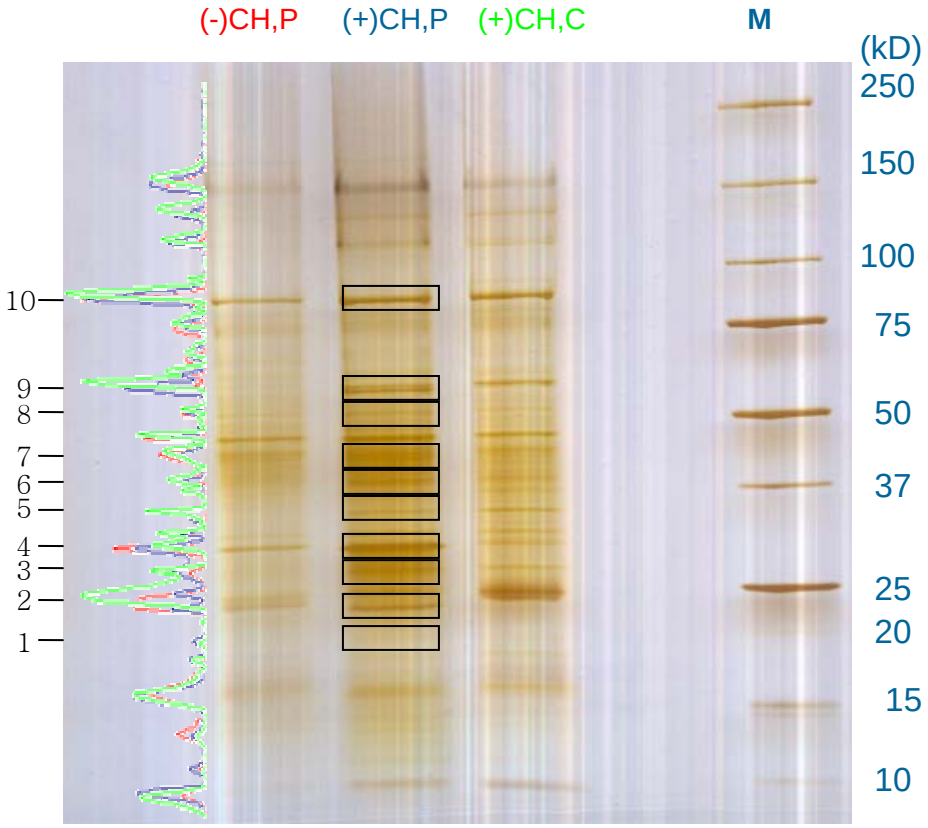
**5’ biotin—TATATGAGCACACAAGG----------GAAAAATTAAGGATATTAGGGACATCC**




Control +100 to +349 relative to translation start site, 150bp

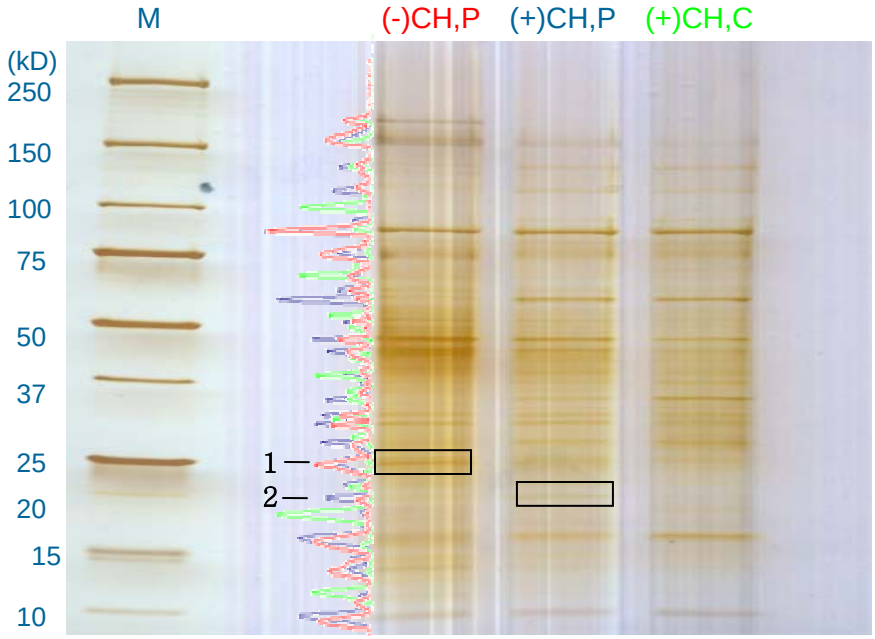
**5’ biotin—GGACTTACTGAGATTTAAAGTTG----------GCCCGATCCATACAGCGTCTATTA**

**Figure 3.3 1D SDS PAGE separation of protein mixtures eluted from PF1983 UOR and control (PF1983) ORF DNA.** The gel was stained with silver. The right side peaks are bands intensity detected by the densitometry. The arrow on the band indicates a protein that occurs at high intensity in the PF1983 UOR lane (blue) compared to the control ORF DNA (green). In-gel tryptic digestion and MS analysis identified this protein as PF0230. The equivalent region in the lane of PF1983 contained no identifiable PF0230 protein.



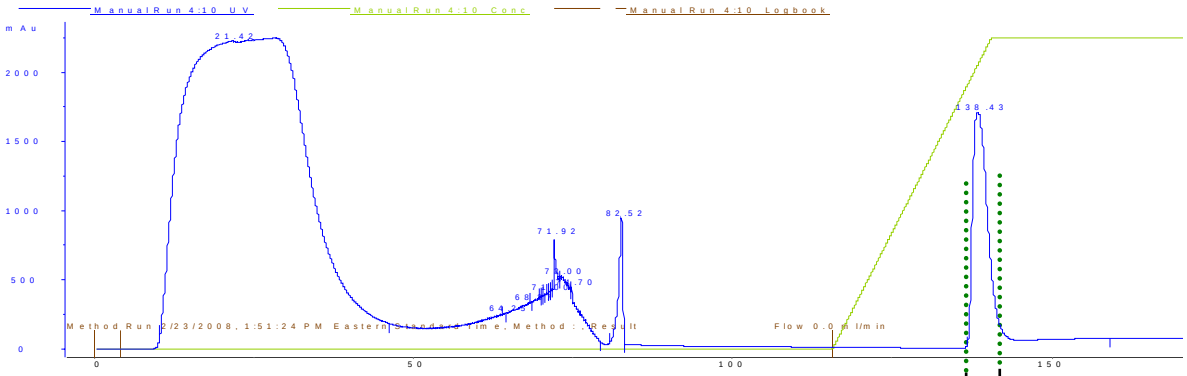

(−) CH, P : Without Cumene Hydroperoxide, Promoter
(+) CH, P : With Cumene Hydroperoxide, Promoter
(+) CH, C : With Cumene Hydroperoxide, Control DNA

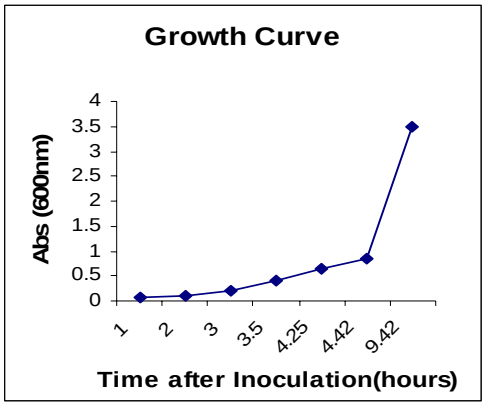
**Figure 3.4 1D SDS PAGE separation of protein mixtures eluted from PF0514 UOR and control (PF514) ORF DNA.** The gel was stained with silver. The right side peaks are bands intensity using the densitometry. The arrows on the bands indicate a protein that occurs at high intensity in the PF0514 UOR lane (red, blue) compared to the control ORF DNA (green).

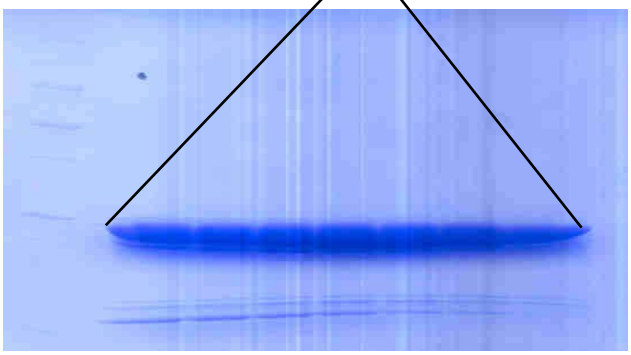



(−) CH, P : Without Cumene Hydroperoxide, Promoter
(+) CH, P : With Cumene Hydroperoxide, Promoter
(+) CH, C : With Cumene Hydroperoxide, Control DNA

**Figure 3.5 Expression and purification of $his_6$-PF0230p.** expression. Protein production was induced with IPTG, and the culture was harvested after four hours of growth. For the growth curve, cell culture samples were diluted to obtain $OD_{600}$ readings in the range of 0.1 to 0.8, and the readings were multiplied by the dilution factor to obtain the actual $OD_{600}$ of the culture. SDS-PAGE of eluted samples from nickel affinity purification.


ManualRun 4:10 UV
ManualRun 4:10 Conc
ManualRun 4:10 Logbook
mAu
2000
1500
1000
500
0
214.2
71.92
82.52
138.43
0
50
100
150


Growth Curve
Abs (600nm)
4
3.5
3
2.5
2
1.5
1
0.5
0
1
2
3
3.5
4.25
4.42
9.42
Time after Inoculation(hours)



CHAPTER 4

# CHARACTERIZATION OF NEW REGULATORY TRANSCRIPTION FACTOR, PF0230P

## 4.1 PF0230p binds specifically to the PF1983 UOR, but not to PF1983 ORF DNA

Standard EMSA methods [58] were used to investigate protein–DNA interactions between PF0230p and PF1983 UOR DNA. The PF0230p was supplied as a gradient against a constant concentration of PF1983 dsDNA. The amounts of DNA and protein applied are detailed in the **Figure 4.1**. To test the specific binding of PF0230p for PF1983 UOR DNA, we also carried out EMSA using PF1983 ORF control DNA. The competitor heparin, which was used in the DNA affinity protein capture experiments, was also used as a competitor in EMSA to compete with DNA nonspecifically bound to PF0230p (standard DNA competition was not possible with our non-radioactive EMSA method). Heparin's electrostatic properties mimic DNA, allowing it to be used as a nonspecific competitor in EMSA [117]. Base-specific contacts of a protein with DNA have much higher affinity than pure electrostatic association with the phosphate-sugar backbone; therefore, a protein binding nonspecifically to DNA solely through electrostatic associations should be able to be challenged off with heparin.

The use of heparin as a non-specific DNA competitor revealed that the association of PF0230p with PF1983 UOR DNA was indeed specific while the binding to PF1983 ORF DNA was nonspecific (**Figure 4.1**). PF0230p binding to the PF1983 ORF DNA was completely blocked at heparin concentrations between 10 and 100 μg/mL. Given this finding, it is not surprising that at high protein/DNA mole ratios, PF0230p associates nonspecifically even with

UOR DNA, as evidenced by the shift to very low mobility complexes at protein/DNA mole ratios of 16 and 32 and the corresponding shift back to the position of a specific complex with the addition of heparin at 10 to 1000 μg/mL (**Figure 4.1**). It is interesting that the position of these low mobility bands in the absence of heparin matches closely with the position of the shifted bands for the ORF DNA at protein/DNA mole ratios of 6, 8 and 12 (**Figure 4.1**). The implication is that many proteins are associating on the DNA in a nonspecific manner, thereby creating a protein-DNA complex with very low mobility and less band definition. It is also worth noting that heparin at sufficiently high concentrations can also challenge off sequence-specific DNA binding as evidenced by its effect on PF0230p binding to PF1983 UOR at a concentration of 1 mg/mL (over 270-fold higher concentration than the DNA probe); this observation has been noted [117]. These data conclusively show that PF0230p is a sequence-specific DNA binding protein, and furthermore, that one of its binding sites appears to be in DNA sequence upstream of the PF1983 ORF.

### 4.2 PF0230p has 80% sequence identity compared with PH1932p

PF0230p is annotated as a transcriptional regulatory protein with a calculated molecular weight of 22.8 kDa and pI of 6.5 in the genome database (TIGR, www.tigr.org). As described in Chapter 3, PF0230p selectively binds to the UOR of PF1983, which is a gene involved in oxidative stress, up-regulated during *P. furiosus* cell growth in the presence of 5 μM cumene hydroperoxide. A conserved domain search with PF0230p sequence places it in a COG with other transcriptional regulators and indicates an ArsR-like HTH domain in the N-terminal region (**Figure 4.2**). A BLAST search indicates that the best matches are orthologs from Ph, Pab, Tkod with high sequence identity (**Figure 4.3**). Protein sequence alignments indicate that PF0230p is

homologous to PH1932 protein, which is annotated as a hypothetical protein but has transcriptional regulator homology in *Pyrococcus horikoshii* OT3 [118]. PH1932p has been crystallographically characterized and shown to have a wHTH domain in its N-terminus [118]. PH1932p consists of 192 amino acids, and is composed of two domains. The N-terminal domain contains a winged helix-turn-helix (HTH) motif similar to that from a bacterial metal-sensing transcription factor, ArsR protein, while no functional information is available for the C-terminal domain. ArsR homologs are widespread in bacteria, and form a class of prokaryotic metal-sensor proteins from the SmtB/ArsR family [119]. These metalloregulatory transcription factors repress expression of operons linked to detoxification of several toxic metal ions such as Zn(II), Co(II), and Ni(II) [119]. Therefore, their derepression results in enhancement of bacterial tolerance for survival under harsher environments. Although proteins that are homologous to the PH1932 protein are found in various archaeal species, none of their biological functions have yet been characterized. Surprisingly, the N-terminal sequences of PF0230p and PH1932p are virtually identical and a PSIPRED prediction of secondary structure in this region identifies helices nearly precisely aligned with the structurally characterized helices in PH1932p (**Figure 4.4**). In result, PF0230p shares about 80% identity with PH1932p amino acid sequences.

### 4.2.1 PF0230p contains an HTH DNA binding domain at the beginning of N-terminus

Analysis of the PF0230p sequence revealed homology with two conserved protein domains that are related to transcription (**Figure 4.2**). The closest match was to a family of conserved proteins, COG1777, annotated as "predicted transcriptional regulators" comprising eight proteins from eight Euryarchaeal species. The N-terminal region of PF0230p was aligned with an HTH/ArsR family of bacterial and archaeal regulators of which the well characterized arsenical

resistance operon repressor (ArsR) was described in the previous section. Secondary structure prediction also predicted the presence of four helices in the N-terminal region (residues 12-81) which showed homology with the HTH/ArsR domain, a further indication that PF0230p contained an HTH DNA-binding domain (**Figure 4.2**). A BLAST search of the sequence resulted in approximately 29 hits having an e-value less than 1, with the majority of high-scoring hits falling within the archaea as predicted transcriptional regulators or conserved hypothetical proteins. As expected, the best hits were the orthologs from three other members of the *Thermococcaceae* family: PH1932 (80% sequence identity), PAB1227 (81% sequence identity), and TK1881 (80% sequence identity). A tree view of the highest scoring BLAST hits can be seen in **Figure 4.3.**

## 4.3 Verification of PF0230p binding sites by DNase I footprinting

### 4.3.1 DNase I footprinting

The PF1983 footprinting probe was 413 bp spanning -200 to +213 relative to translation start and amplified by PCR with fluorescently labeled primers at their 5' ends, one with HEX and the other with 6FAM. The primer sequences are given in **Table 4.1**, the full probe sequences for both strands are given in **Figure 4.5**, and a gel image of the purified PCR amplicon is shown in **Figure 4.6**. The same DNA was used as the template for *in vitro t*ranscription of PF1983 (vide infra).

Table 4.1 Probe used in DNase I footprinting

| Probe name | Genome coordinates | Forward primer[a] | Reverse primer[a] | Probe length (bp) |
|---|---|---|---|---|
| 1983 | 1832907-1833319 | 6FAM/gggagagactagacaact | HEX/ttatggcattctcagccccag | 413 |

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

4.3.2 PF0230p forms a 60-bp footprint on PF1983 promoter

The DNase I footprinting analysis of PF0230p on PF1983 UOR is shown in **Figure 4.7**. The footprint region on the 6FAM strand covers ca. -164 to -104 relative to PF1983 translation start site. The PF0230p footprint supported the previous EMSA results that PF0230p binds sequence-speficially to the PF1983 promoter region.

## 4.4 SELEX shows that PF0230p recognition site contains the palindrome ATTAAT

4.4.1 Identification of PF0230p binding motif by SELEX

The DNase I footprint does not show us the exact recognition sequence. So, to confirm and define the consensus DNA motif recognized by PF0230p, the artificial selection method SELEX (**S**ystematic **E**volution of **L**igands by **EX**ponential enrichment) was performed [110]. First, we synthesized an 80-bp long SELEX probe with the central 30 bp randomized sequence (**Figure 4.8**). EMSA with PF0230p was used to select sequences bound specifically by PF0230p; shifted DNA was excised from an EMSA gel. Each pool of selected sequences was used as template in PCR amplification and the amplified DNA was purified to be used in EMSA again for further rounds of selection. After six selection rounds were completed, the total selected DNAs were cloned into pUC18 and sequenced. MEME analysis identified the consensus palindromic DNA sequence ATTAAT in 21 total sequenced SELEX DNAs, with the motif displaying perfect palindrome in 16 of 21 (**Figure 4.9**). This putative binding motif sequence ATTAAT occurs in the PF1983 UOR footprint from -137 to -132 relative to translation start site (Figure 4.8).

4.4.2 EMSA validation of ATTAAT as the PF0230p binding motif

In order to validate whether the ATTAAT motif in the PF0230p footprinting plays a role in PF0230p DNA-binding activity, 60-bp DNA oligonucleotides were synthesized based on the sequence surrounding the ATTAAT sequence in the PF1983 UOR footprint (**Figure. 4.7**). Another pair of complementary oligonucleotides containing the sequence GCGCGC in place of ATTAAT were prepared for comparison (**Figure 4.10 a**). EMSA analysis confirmed that the ATTAAT-containing DNA showed formation of complexes, while the mutated DNA did not (**Figure 4.10 b**). This result illustrates that the 6-bp palindromic sequence ATTAAT is very important for PF0230p binding to PF1983 UOR.

## 4.5 Search for additional potential PF0230p binding sites in the *P. furiosus* genome

The identified ATTAAT binding sequence was used to search for other potential binding sites in the UORs of other genes within the genome. A *P. furiosus* genome database of UORs has been created by a coworker, Darin Cowart (University of Georgia). Searching all UORs in the genome for the motif ATTAAT found 192 UORs. We presume that not all of the ATTAAT motifs identified from the search are real PF0230p binding sites as it seems that the SELEX motif is too short to be the only essential feature for specific binding of PF0230p to DNA; not enough is known about the DNA sequence recognition by PF0230p to be able to identify false positives. However, we should exclude the UORs of genes that are not the first gene of the operon. Doing this reduces the hit list to 15 UORs. Of this set only 15 UORs containing the ATTAAT motif also appeared in the list of genes highly regulated (≥ 3.6-fold) in the cumene hydroperoxide microarray expression profiles (**Table 4.1**). 6 genes among these 15 (PF1983, PF0105, PF0581, PF1080, PF1513, PF1659) were up-regulated at 1 h after stress and 5 genes (PF0340, PF0972,

PF1264, PF1483, PF1774) were down-regulated at 1 h after stress. The remaining 4 genes (PF0289, PF0422, PF1767, PF1772) were down-regulated at 2 h after stress.

**Table 4.1 UOR database search results[a] for the motif ATTAAT**

| UORs[b] | ORF Annotation[c] | Start[d] | Stop[d] |
|---|---|---|---|
| **Up-regulated at 1 h after adding cumene hydroperoxide** | | | |
| PF1983 | Hypothetical protein | -137 | -132 |
| PF0105 | Hypothetical protein | -52 | -47 |
| PF0581 | Hypothetical protein | -81 | -76 |
| PF1080 | Hypothetical protein | -176 | -171 |
| PF1513 | Hypothetical protein | -21 | -16 |
| PF1659 | Histidinol dehydrogenase | -37 | -32 |
| **Down-regulated at 1 h after adding cumene hydroperoxide** | | | |
| PF0340 | HTH transcription regulator | -68 | -63 |
| PF0972 | Acyl carrier protein synthase | -113 | -108 |
| PF1264 | TIF eIF-5a | -140 | -135 |
| PF1483 | Hypothetical protein | -60 | -55 |
| PF1774 | Iron III ABC transporter | -22 | -17 |
| **Down-regulated at 2 h after adding cumene hydroperoxide** | | | |
| PF0289 | GTP-hydrolyzing phosphoenolpyruvate caboxykinase | -43 | -38 |
| PF0422 | Phosphoribosylamine-glycine ligase | -39 | -34 |
| PF1767 | 2-keto acid ferredoxin oxidoreductase subunit delta | -96 | -91 |
| PF1772 | 2-keto acid ferredoxin oxidoreductase subunit beta | -183 | -178 |

[a] The *P. furiosus* UOR database and corresponding motif searching software was created by Darin Cowart (University of Georgia).
[b] UOR (Upstream of ORF Region) designation corresponds to the locus of the ORF from which the upstream sequence was taken.
[c] Characterized ORFs/proteins are listed with the corresponding reference, and ORF annotations in italics are from REFSEQ or TIGR.
[d] Start and stop positions are relative to the UOR sequence where -1 corresponds to the first nucleotide upstream from the ORF start.

### 4.6 EMSA shows that PF0230p binds to other promoters that have ATTAAT sequence

EMSA was used initially as an indicator of whether PF0230p bound these identified UOR DNAs. We chose seven different UORs listed in **Table 4.1** to test the binding affinity with PF0230p. All of these UORs were synthesized from -200 to +100 relative to translation start site (**Table 4.2**). EMSA results for all seven UORs are shown in **Figure 4.11**. It is clear that PF0230p binds to all of these UORs containing the ATTAAT binding motif. Therefore PF0230p is a common DNA binding protein for the group of genes whose promoters contain this binding motif.

**Table 4.2 Probes used in EMSA**

| Probe name | Genome coordinates | Forward primer[a] | Reverse primer[a] | length (bp) |
|---|---|---|---|---|
| 0340 | 353959-353659 | cattccccctcaactcttagaaaat | ctttacgagatttatccttctgccac | 300 |
| 1774 | 1649241-1649541 | gccgttttagactccgtccc | caacaattccaacaactccatcagcacc | 300 |
| 0581 | 601554-601254 | gggatcataaacgattggatagtc | catggacattgaacgagcgtac | 300 |
| 1080 | 1030201-1030501 | tataataatttccagctacag | gcagcacgtccagtagatgacac | 300 |
| 0289 | 302236-302536 | ttctacaatgttttcttttcc | ccctacctacactcatttttggc | 300 |
| 0422 | 432330-432030 | ccacaatgtcttgaacctttttc | gctctatgtagtttcgaaac | 300 |
| 1513 | 1412070-1412370 | caaaccctactaagaaatgctatagctcc | gcattagctgtccttctgaagtac | 300 |

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

### 4.7 PF0230p binds to the TATA box and blocks TBP binding

The discovered ATTAAT binding motif has sequence similarities to a TATA box and could actually overlap a functioning TATA box. We hypothesized that PF0230p binds to TATA and blocks TBP (TATA binding protein) binding. To test this hypothesis, we performed EMSA with

PF0230p, TBP, and TFB together as described in Chapter 2. Seven EMSA experiments were performed, each with the ATTAAT-containing probe and with the GCGCGC-containing probe. The seven EMSA experiments were: (1) PF0230p only; (2) TBP only; (3) PF0230p + TBP; (4) TFB only; (5) PF0230p + TFB; (6) TBP + TFB; (7) PF0230p + TBP + TFB (Figure 4.13 A, B). TBP is known to have very low binding affinity to the TATA box by itself and no shift is observed in experiment (2). However, experiment (6) confirms that the TFB/TBP combination shifts DNA by binding to the BRE/TATA combination. but we wanted to show the fact in condition 3 and TBP could have binding affinity with DNA probe when it existed with TFB. EMSA gel results indicated that PF0230p has binding ability to TATA box. We could observe band shift of lane 1, 3, 5, 7 which we added PF0230p to the reaction in ATTAAT binding motif condition (**Figure 4.12 A**) but we couldn't see band shift of these lane in GCGCGC mutated DNA condition (**Figure 4.12 B**).

### 4.8 PF0230p exists as a dimer, determined by analytical gel filtration

Transcription factors that bind to palindromic DNA sequences often do so as dimmers. To determine the quaternary structure of PF2030p in solution, we carried out analytical gel filtration using a Superdex200 10/30 gel filtration column (Amersham Pharmacia Biotech, Sweden) with a 2000 kDa exclusion limit. The approximate corresponding molecular weight of the PF0230p sample was calculated using a calibration curve (**Figure 4.13 B**), generated from a set of standards run individually through the column with the same buffer. The resulting chromatogram for PF0230p sshown in **Figure 4.13 A**, and the calculated molecular weight corresponding to the elution volume (15.22 mL) was 46,500 Da.

According to the PF0230p amino acid sequence, the molecular weights for a $His_6$-tagged PF0230p monomer, dimer and trimer are 22,859, 45,718 and 68,577 Da, respectively. These results suggest that PF0230p is a dimer in solution.

### 4.9 *In vitro* transcription suggests PF0230p is both a repressor and activator

To determine the regulatory function of PF0230p in the transcription of DNA probes which have DNA binding motif, *in vitro* transcription assays were performed in Dr. Michael Thomm's laboratory (University of Regensburg) with an established cell-free transcription system [120] using the purified PF0230p and gel-purified DNA templates from PCR amplification. As a control, the glutamate dehydrogenase gene (*gdh*) of *P. furiosus* was used; this is assume to be unregulated by PF0203p. We chose three promoters for *in vitro* transcription (**Table 4.3**). The PF1983 template (300bp, -200 to +100 relative to translation start site) showed transcript (**Figure 4.14)** and as the amount of PF0230p was increased, the amount of PF1983 run-off transcript was reduced, indicating repression and they showed us that the transcription start site is -119 relative to the translation start site. The transcript from the PF1513 template (300bp, -200 to +100 relative to translation start site) can be seen in **Figure 4.15**. In contrast to the PF1983 result, PF0230p seemed to enhance slightly the transcription of PF1513 (**Figure 4.15**). *In vitro* transcription of PF0340 (300bp, -200 to 100 relative to translation start site) results in a transcript as shown in **Figure 4.16**. However, little effect on this transcription by PF0230p was evident, even though it has a binding motif on this promoter (**Figure 4.16**).

**Table 4.3 Probes used in *in-vitro* transcription**

| Probe name | Genome coordinates | Forward primer[a] | Reverse primer[a] | length (bp) |
|---|---|---|---|---|
| 1983 | 1832907-1833207 | gggagagactagacaactagctaagg | ctagtactcaggatccctatgaagagttt | 300 |

| | | | | |
|---|---|---|---|---|
| 1513 | 1412070-1412370 | caaaccctactaagaaatgctatagctcc | gcattagctgtccttctgaagtac | 300 |
| 0340 | 353959-353659 | cattccccctcaactcttagaaaat | ctttacgagatttatccttctgccac | 300 |

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

### 4.10 PF0230p binding is independent of cumene hydroperoxide or hydrogen peroxide

In vitro transcription results showed that PF0230p is a transcriptional repressor of PF1983 and a transcriptional activator of PF1513, both of which were shown by expression profiles to be involved in oxidative stress (PF1983 is up-regulated by 10.1-fold, while PF1513 is also up-regulated by 8.5-fold stressing the presence of cumene hydroperoxide). We therefore hypothesized that the transcriptional regulation by PF0230p is induced by some peroxide such as cumene hydroperoxide or hydrogen peroxide. This was tested by investigating the effect of these peroxides on EMSA. However, the intensity of the shifted PF0230p-PF1983 UOR complex was not significantly different in the presence of either cumene hydroperoxide or hydrogen peroxide (**Figure 4.17**). A similar result was observed for EMSA with PF1513 UOR (**Figure 4.18**). This suggests that the relationship of PF0230p regulation of both PF1983 and PF1513 with oxidative stress is not the direct effect of interaction with these peroxides influencing the DNA binding affinity of PF0230p.

**Figure 4.1 PF0230p binds specifically to PF1983 promoter DNA in the presence of heparin up to 1000ug/ml.** Top; Diagrams of the DNA probes are shown. The probes used are indicated at the top of each gel image with corresponding protein concentration listed above each lane. PF1983 promoter region is shaded in yellow and PF1983 ORF region is colored in pink. Bottom; DNA (100 nM or ~4.0 μg/mL) was incubated with protein in buffer (50 mM HEPES pH 7.5, 200 mM KCl, 5% glycerol, 1 mM EDTA) with and without heparin (concentrations indicated) for 20 min at 55 °C. Gel was stained with SYBR Green I nucleic acid gel stain.

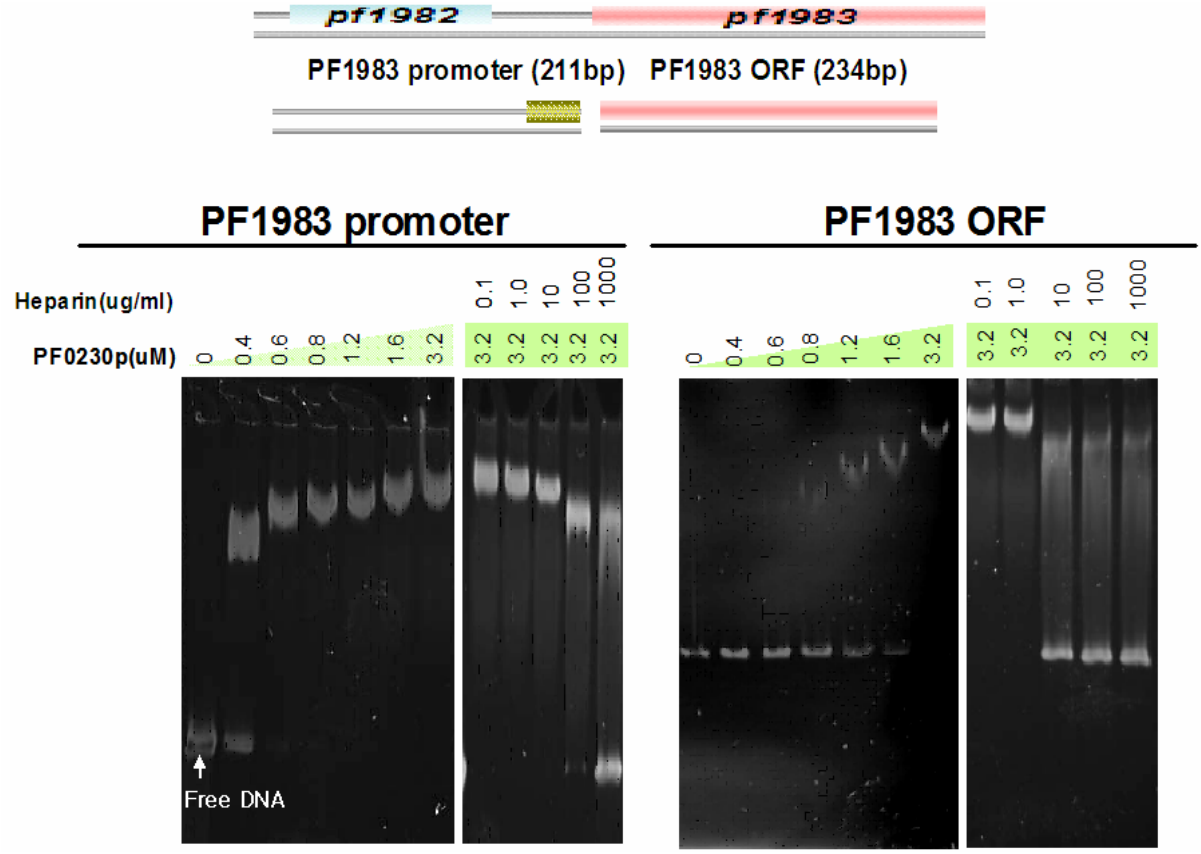
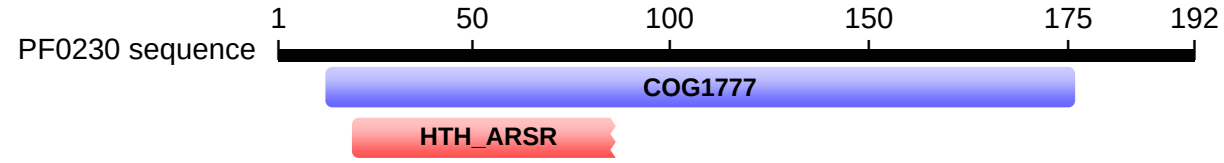
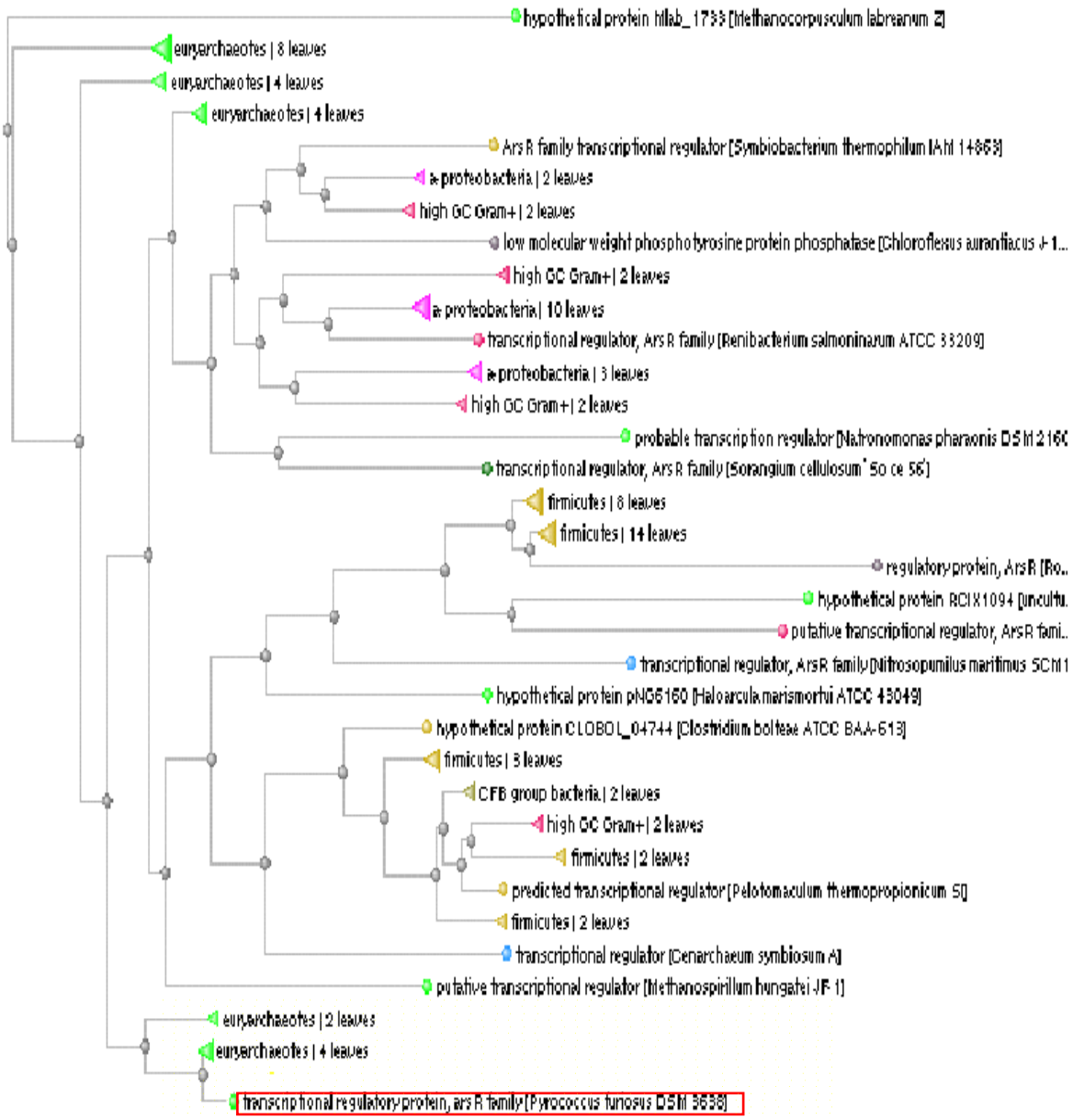

pf1982
pf1983
PF1983 promoter (211bp)
PF1983 ORF (234bp)
PF1983 promoter
PF1983 ORF
Heparin(ug/ml)
PF0230p(uM)
0
0.4
0.6
0.8
1.2
1.6
3.2
0.1
1.0
10
100
1000
Free DNA

**Figure 4.2 Conserved domain search results for PF0230 protein.** Sequence from online tools available at NCBI. PF0230p 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




```
10        20        30        40        50        60        70
     ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*...
PF0230p  14 IKVMLEDTRRKILQLLRNREMTISQLSEILGKMPQTIYHHIEKLKEAGLVEVKRTEMKgnIveKYYGRTADVFYINLY 91
HTH_ARSR  1 LKALSDPTRLRILRLLLEGPLTVSELAERLGLSQSTVSRHLKKLEEAGLVESRREGRR-----VYYSLTDAERLLALL 73
```

**Figure 4.3. Tree-view for results of a BLAST search of the PF0230p sequence against the NCBI non-redundant database and PF, PAB, PH and TKOD sequence blast analysis.** PF0230 is shown boxed in red, proteins of the Thermococcaceae family are highlighted in yellow. A.PF versus PAB, B. PF versus PH, PF versus TKOD Amino acid sequence analysis.


hypothetical protein Mlab_1733 [Methanocorpusculum labreanum Z]
euryarchaeotes | 8 leaves
euryarchaeotes | 4 leaves
euryarchaeotes | 4 leaves
ArsR family transcriptional regulator [Symbiobacterium thermophilum IAM 14863]
α-proteobacteria | 2 leaves
high GC Gram+ | 2 leaves
low molecular weight phosphotyrosine protein phosphatase [Chloroflexus aurantiacus J-1...
high GC Gram+ | 2 leaves
α-proteobacteria | 10 leaves
transcriptional regulator, ArsR family [Renibacterium salmoninarum ATCC 33209]
α-proteobacteria | 3 leaves
high GC Gram+ | 2 leaves
probable transcription regulator [Natronomonas pharaonis DSM 2160
transcriptional regulator, ArsR family [Sorangium cellulosum 'So ce 56']
firmicutes | 8 leaves
firmicutes | 14 leaves
regulatory protein, ArsR [Ro...
hypothetical protein RCIX1094 [unculti...
putative transcriptional regulator, ArsR fami...
transcriptional regulator, ArsR family [Nitrosopumilus maritimus SCM1
hypothetical protein pNG6160 [Haloarcula marismortui ATCC 43049]
hypothetical protein CLOBOL_04744 [Clostridium bolteae ATCC BAA-613]
firmicutes | 3 leaves
CFB group bacteria | 2 leaves
high GC Gram+ | 2 leaves
firmicutes | 2 leaves
predicted transcriptional regulator [Pelotomaculum thermopropionicum SI]
firmicutes | 2 leaves
transcriptional regulator [Cenarchaeum symbiosum A]
putative transcriptional regulator [Methanospirillum hungatei JF-1]
euryarchaeotes | 2 leaves
euryarchaeotes | 4 leaves
transcriptional regulatory protein, arsR family [Pyrococcus furiosus DSM 3638]

A

```
 Score =  320 bits (819),  Expect = 6e-86
 Identities = 157/192 (81%), Positives = 180/192 (93%), Gaps = 0/192 (0%)

Query  1    MTKAVKVITDPEVIKVMLEDTAAKILQLLANAEMTISQLSEILGKMPQTIVHHIEKLKEA  60
            M ++VKVITDPEVIKVMLEDTAA IL+LLANAEMTISQLSEILGK PQTIVHHIEKLKEA
Sbjct  1    MAAKVKVITDPEVIKVMLEDTAAAILKLLANAEMTISQLSEILGKTPQTIVHHIEKLKEA  60

Query  61   GLVEVKATEMKGNLVEKVYGATADVFYINLYMGDEELAYLARSALKTKLDIFKKLGYKFD  120
            GLVEVKATEMKGNLVEKVYGATADVFYINLY+GDEELAY+ARSALKTK+DIFK+LGYKF+
Sbjct  61   GLVEVKATEMKGNLVEKVYGATADVFYINLYLGDEELAYIARSALKTKIDIFKALGYKFE  120

Query  121  EEELLNVMDRILEKEHEVKVEISKEIENVEDSLKEFSNEDIIHAIEWLAMARLAQDEEVI  180
            EEELL+VMD++ +KE +  V +SK+IE +ED+LKEFSNEDIIHAIEWL+ A L++DEEV+
Sbjct  121  EEELLDVMDKMSQKEFDATVRVSKQIEEMEDALKEFSNEDIIHAIEWLSTAELSADEEYM  180

Query  181  ELVKKLGQILKA  192
             L+KKLG+ILKA
Sbjct  181  SLLKKLGEILKR  192
```

B

```
 Score =  315 bits (808),  Expect = 1e-84
 Identities = 156/189 (82%), Positives = 176/189 (93%), Gaps = 0/189 (0%)

Query  4    AVKVITDPEVIKVMLEDTAAKILQLLANAEMTISQLSEILGKMPQTIVHHIEKLKEAGLV  63
            +VKVITDPEVIK+MLEDTAAKILQLLANAEMTISQLSEILGK PQTIVHHIEKLKEAGLV
Sbjct  3    KVKVITDPEVIKLMLEDTAAKILQLLANAEMTISQLSEILGKTPQTIVHHIEKLKEAGLV  62

Query  64   EVKATEMKGNLVEKVYGATADVFYINLYMGDEELAYLARSALKTKLDIFKKLGYKFDEEE  123
            EVKATEMKGNLVEKVYGATAD FYINLY+GDEELAY ARSALKTKL+IFK LGY+F++EE
Sbjct  63   EVKRTEMKGNLVEKVYGRTADAFYINLYLGDEELRYFARSRLKTKLEIFKALGYEFNDEE  122

Query  124  LLNVMDAILEKEHEVKVEISKEIENVEDSLKEFSNEDIIHAIEWLAMARLAQDEEVIELV  183
            LLNVMD +L+KEH+ K EISKEIE  E+ LK+FSNEDIIHAIEWLAMA++ +DEE + L+
Sbjct  123  LLNVMDELLKKEHDVKTEISKEIEVNEEKLKDFSNEDIIHAIEWLAMAKMGRDEEVLNLL  182

Query  184  KKLGQILKA  192
            +KLG+ILK+
Sbjct  183  RKLGEILKK  191
```

C

```
 Score =  319 bits (818),  Expect = 8e-86
 Identities = 158/192 (82%), Positives = 178/192 (92%), Gaps = 0/192 (0%)

Query  1    MTKAVKVITDPEVIKVMLEDTAAKILQLLANAEMTISQLSEILGKMPQTIVHHIEKLKEA  60
            M K+VKVITDPEVIKVMLEDTAAKIL+LLAN+EMTISQLSEILGK PQTIVHHIEKLKEA
Sbjct  1    MAKKVKVITDPEVIKVMLEDTAAKILKLLANKEMTISQLSEILGKTPQTIVHHIEKLKEA  60

Query  61   GLVEVKATEMKGNLVEKVYGATADVFYINLYMGDEELAYLARSALKTKLDIFKKLGYKFD  120
            GLVEVKATEMKGNLVEKVYGATADVFYINLY+GDEELAY+ARSALKTK+DIFK+LGY+F+
Sbjct  61   GLVEVKATEMKGNLVEKVYGATADVFYINLYLGDEELAYIARSRLKTKIDIFKRLGYQFE  120

Query  121  EEELLNVMDRILEKEHEVKVEISKEIENVEDSLKEFSNEDIIHAIEWLAMARLAQDEEVI  180
            E ELLN+MDR+ +KE +  V ISK IE  ED+LK+FSNEDIIHAIEWL+ A LA+DEEV+
Sbjct  121  ENELLNIMDRMSQKEFDATVRISKVIEEKEDALKDFSNEDIIHAIEWLSTAELARDEEYL  180

Query  181  ELVKKLGQILKA  192
            EL+K+LG ILKA
Sbjct  181  ELLKALGSILKA  192
```

**Figure 4.4 Sequence comparison between PF0230p and PH1932p.** PF0230p ("query") has 80% sequence identity compared with PH1932p ("1ULY_A"). The N-terminal sequence of PF0230p is nearly identical to that of PH1932 and the predicted secondary structure of the former matches the helices found in the structure of PH1932p [118].

**E-value = 4e-66, Bit score = 253, Aligned length = 190, Sequence Identity = 80%**

```
                 10        20        30        40        50        60        70        80
               ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
query     3 KRVKVITDPEVIKVMLEDTRRKILQLLRNREMTISQLSEILGKMPQTIYHHIEKLKEAGLVEVKRTEMKGNLVEKYVGRT 82
1ULY_A    3 KKVKVITDPEVIKVXLEDTRRKILKLLRNKEXTISQLSEILGKTPQTIYHHIEKLKEAGLVEVKRTEXKGNLVEKYVGRT 82

                 90        100       110       120       130       140       150       160
               ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
query    83 ADVFYINLVMGDEELRYLARSRLKTKLDIFKKLGVKFDEEELLNVMDRILEKEHEVKVEISKEIENVEDSLKEFSNEDII 162
1ULY_A   83 ADVFYINLVLGDEELRYIARSRLKTKIDIFKRLGVQFEENELLNIXDRXSQKEFDATVRISKYIEEKEDALKDFSNEDII 162

                 170       180       190
               ....*....|....*....|....*....|
query   163 HAIEWLAMARLAQDEEYIELVKKLGQILKR 192
1ULY_A  163 HAIEWLSTAELARDEEYLELLKRLGSILKR 192
```

**Figure 4.5 The sequence of forward and reverse strands of the PF1983 footprinting probe.** The ribosome binding site is boxed in red and the translation start site of the PF1983 ORF in black.


6FAM /
GGGAGAGACTAGACAACTAGCTAAGGAATTAGCAAATGAGCTGGGGCTTGAATTTATAAA
CCCTCTCTGATCTGTTGATCGATTCCTTAATCGTTTACTCGACCCCGAACTTAAATATTT
GGATTAATTTCATCTGAATAGTTTATTTCTTTATGCGTGCTCAAAGTAATAAGTATGAGT
CCTAATTAAAGTAGACTTATCAAATAAAGAAATACGCACGAGTTTCATTATTCATACTCA
TCAATTTTATCACTATGTTTGTTTTTTATATAGAAATGTTTTTTAGTACTATTTTGGGTT
AGTTAAAATAGTGATACAAACAAAAAATATATCTTTACAAAAAATCATGATAAAACCCAA
GTAATTTTGGTGATTTGTGGTGAGGAGGAGTATTGTATTCTTTGTTATAAGTTTGTTATT
CATTAAAACCACTAAACACCACTCCTCCTCATAACATAAGAAACAATATTCAAACAATAA
GGGAGGCCTAGTTAATCCTAGTATTGTCCAAGCTAGTACTCAGGATCCCTATGAAGAGTT
CCCTCCGGATCAATTAGGATCATAACAGGTTCGATCATGAGTCCTAGGGATACTTCTCAA
TTGGGAAATACTTTGGAGGGAGGCTAGGCTCGTTGGGGAGGCAGAATCTGGAAACACTAC
AACCCTTTATGAAACCTCCCTCCGATCCGAGCAACCCCTCCGTCTTAGACCTTTGTGATG
AGCAATAAACGAGTTAATAGAGAACTCCAAGGCTGGGGCTGAGAATGCCATAA
TCGTTATTTGCTCAATTATCTCTTGAGGTTCCGACCCCGACTCTTACGGTATT/ HEX

**Figure 4.6 Gel-purified PCR product of fluorescently labeled probe for DNase I footprinting.** Total size of this PF1983 probe is 413 bp and concentration was 200ng/μL. DNA ladder marker was used (right side of gel)

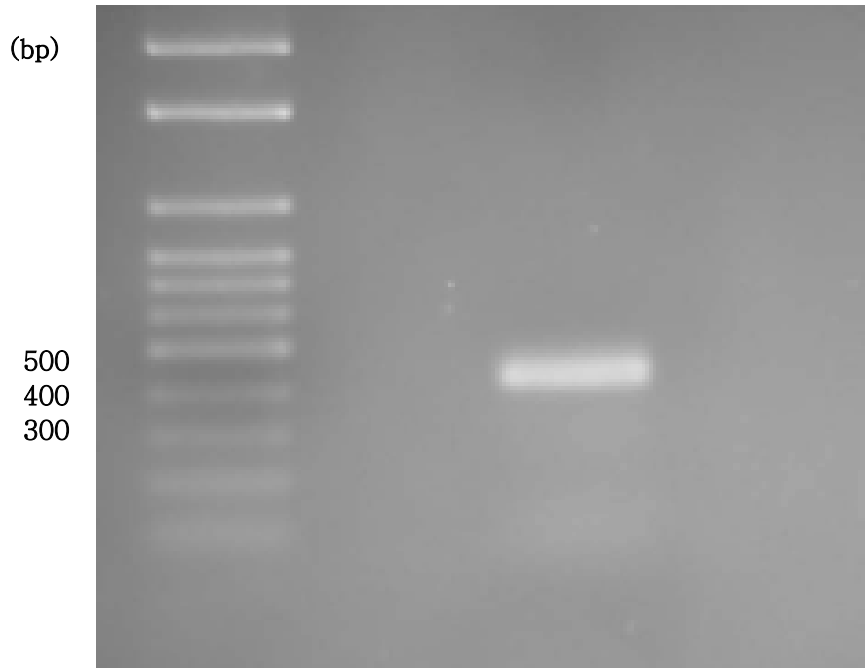

(bp)
500
400
300

**Figure 4.7 Footprinting electropherograms showing footprint of PF0230p on PF1983 UOR.** The electropherogram compares control (red line, without protein) and sample (black line, with protein), with y axis relative intensity and x axis the DNA length (bp). The specific sites of PF1983 UOR protected by PF0230p from DNase I are indicated. The top electropherogram shows undigested probe (blue) and standard marker (ROX, in black). The bottom sequence shows the protected region on PF1983 UOR forward strand (6FAM labeled) from -164 to -104 (relative to PF1983 translation start). The palindromic sequence (-137 to -132) is indicated in red and transcription start site in purple.

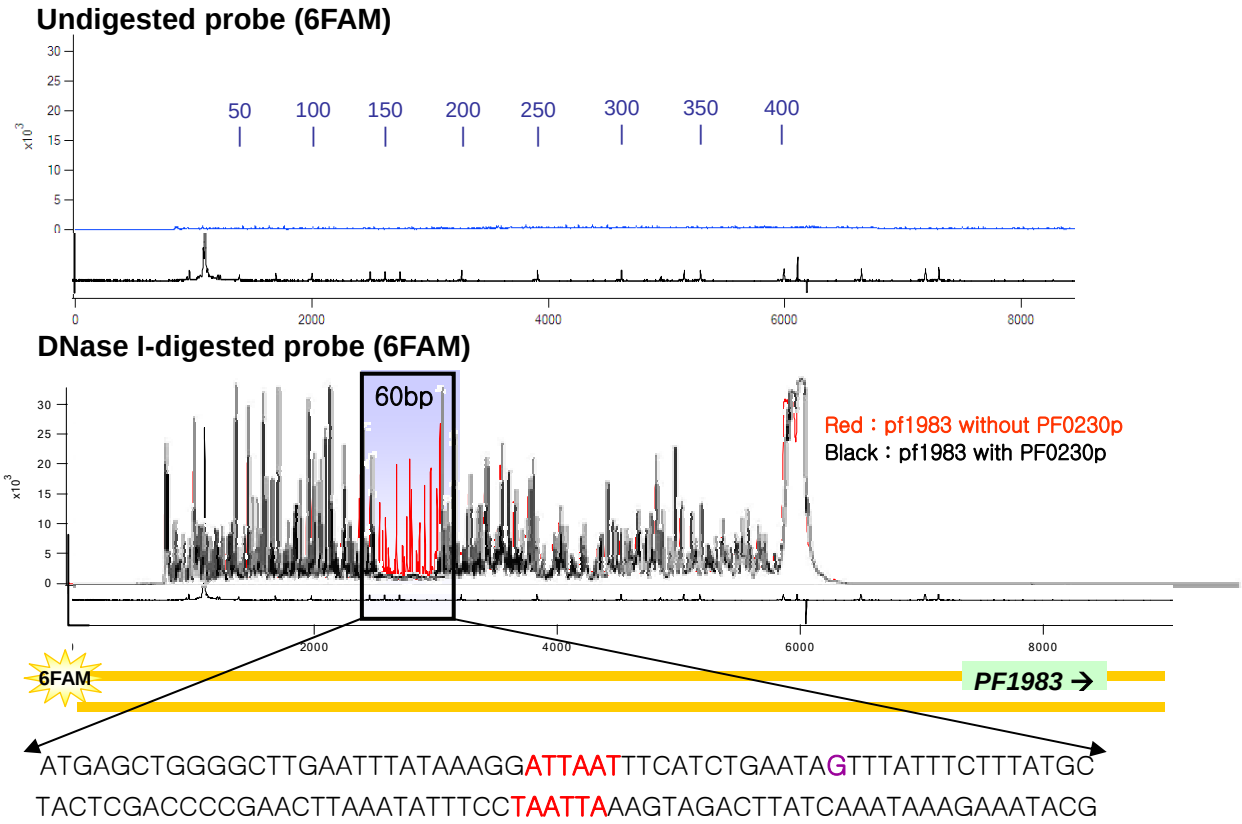

Undigested probe (6FAM)
50
100
150
200
250
300
350
400
DNase I-digested probe (6FAM)
60bp
Red : pf1983 without PF0230p
Black : pf1983 with PF0230p
6FAM
PF1983 →
ATGAGCTGGGGCTTGAATTTATAAAGGATTAATTTCATCTGAATAGTTTATTTCTTTATGC
TACTCGACCCCGAACTTAAATATTTCCTAATTAAAGTAGACTTATCAAATAAAGAAATACG

**Figure 4.8 Design of SELEX probe.** 30bp randomized probe with Hind III and *EcoR* I at the end of randomized probe (top). Used *EcoR* I and HindIII to digest SELEX probe (bottom)

## SELEX probe

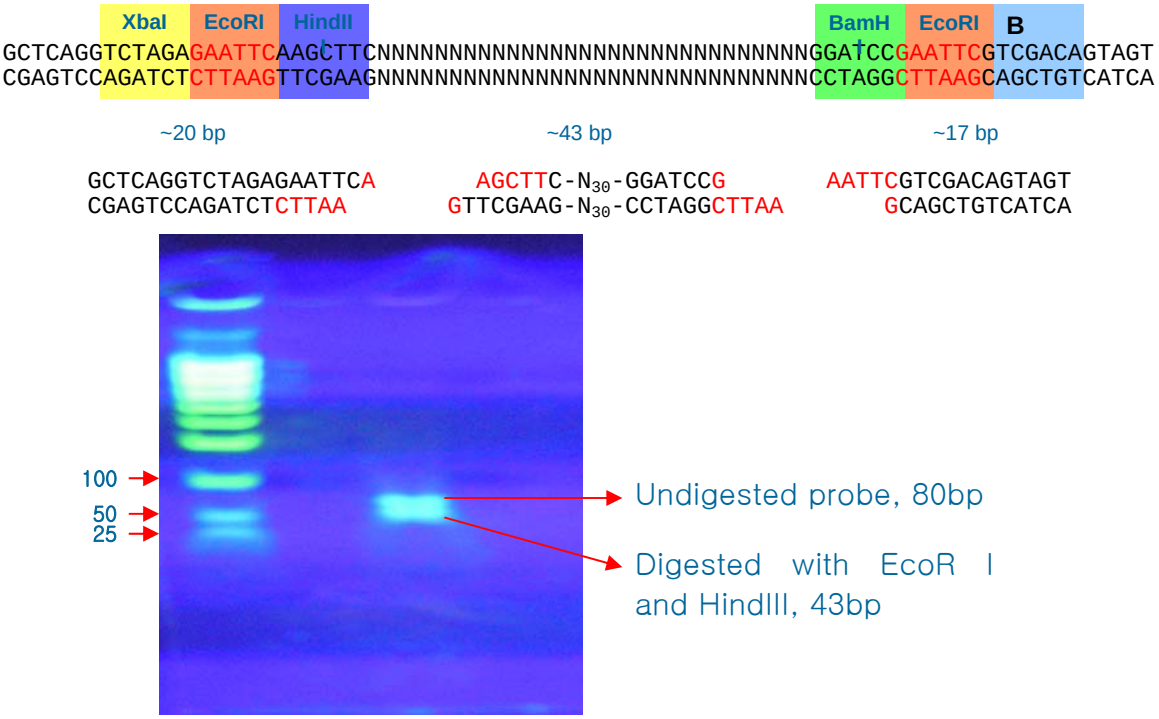

XbaI
EcoRI
HindIII
BamH
EcoRI
B
GCTCAGGTCTAGAGAATTCAAGCTTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGATCCGAATTCGTCGACAGTAGT
CGAGTCCAGATCTCTTAAGTTCGAAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCTAGGCTTAAGCAGCTGTCATCA
~20 bp
~43 bp
~17 bp
GCTCAGGTCTAGAGAATTCA
CGAGTCCAGATCTCTTAA
AGCTTC-N30-GGATCCG
GTTCGAAG-N30-CCTAGGCTTAA
AATTCGTCGACAGTAGT
GCAGCTGTCATCA
100
50
25
Undigested probe, 80bp
Digested with EcoR I and HindIII, 43bp

**Figure 4.9 SELEX result from MEME.** 21 sequenced SELEX results were searched for repetitive sequence using MEME [112]. The results of this search are displayed as a WebLogo at the bottom [113].

| NAME | STRAND | START | P-VALUE | SITES | |
|---|---|---|---|---|---|
| 16 | + | 7 | 3.10e-06 | GTCCTA | ATTAATGCCTC TTTACGCATT |
| 7 | - | 3 | 3.10e-06 | CGGGAATGAG | ATTAATAGCAG CA |
| 4 | - | 3 | 3.10e-06 | CGGGAATGAG | ATTAATAGCAG CA |
| 6 | + | 8 | 5.58e-06 | AIAICAI | ATTAATAGCAC CGICGAGCCG |
| 19 | - | 8 | 1.36e-05 | GGCACTAGGG | ATTAATGCATG GAAGCGT |
| 9 | - | 2 | 2.37e-05 | GGGCCTAGAG | ATTAATGCGGG G |
| 26 | + | 9 | 3.06e-05 | GACCGTII | ATTAATGCTGG AAGGGTGTAC |
| 23 | - | 5 | 3.06e-05 | GACGCTACAI | ATTAATAGCCC GTAA |
| 27 | - | 10 | 4.08e-05 | TCTGGACTGA | ATTAATCGCTC AGTGACAAG |
| 20 | - | 3 | 6.44e-05 | CTCGATAGGA | ATTAATGAAAC CC |
| 17 | - | 2 | 6.44e-05 | AGGICTCACG | ATTAATGGAAA T |
| 5 | + | 9 | 6.44e-05 | CICICGGA | ATTAATGGCCA TAAICICCGG |
| 10 | + | 8 | 6.86e-05 | CGTCCAG | ATTAATAGGCC ATAATCTCCC |
| 21 | - | 9 | 7.42e-05 | GCGGAAGCCG | ATTAATAACGG CGTGGGTT |
| 25 | + | 5 | 7.95e-05 | CCTA | ATTAATCACAC CTTCAGACCT |
| 12 | - | 3 | 8.33e-05 | CGTAAAGAIG | ATTCATGCTAG GA |
| 8 | + | 8 | 1.03e-04 | AGICCGA | ATTAATACTTA GCCACCTGCA |
| 13 | - | 3 | 1.44e-04 | CGTAAAGATG | ACTAATGCTAG AC |
| 22 | + | 8 | 1.80e-04 | AACCGGA | ATTCATCCGAG ACCATGAACA |
| 3 | + | 5 | 1.92e-04 | TCGC | ATCAATGCATG CACICCIAIC |
| 24 | + | 3 | 2.71e-04 | IG | ATTACTACGTC GAGIATTTCG |

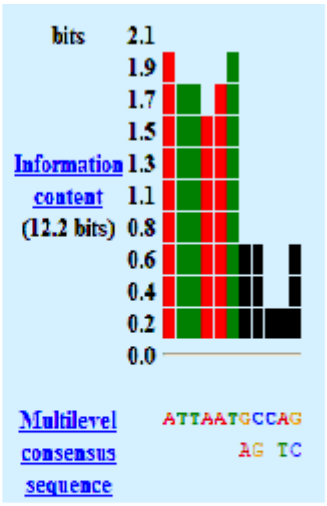
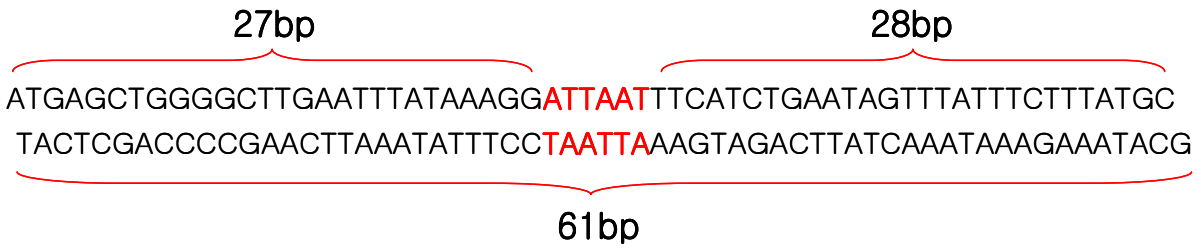



**Figure 4.10 EMSA of PF0230p with PF1983 UOR and mutated UOR DNA.**

a. Sequences showing footprinting region of PF1983 and mutated binding motif from ATTAAT to GCGCGC b. DNA staining by SYBR Green I shows free UOR decreased as protein concentration increased in the reaction. The secondary protein-DNA complex was formed at high concentration of protein. No mutated PF1983 UOR DNA was shifted by the same gradient of PF0230p.

a

**Binding motif**




**Mutation DNA**

b

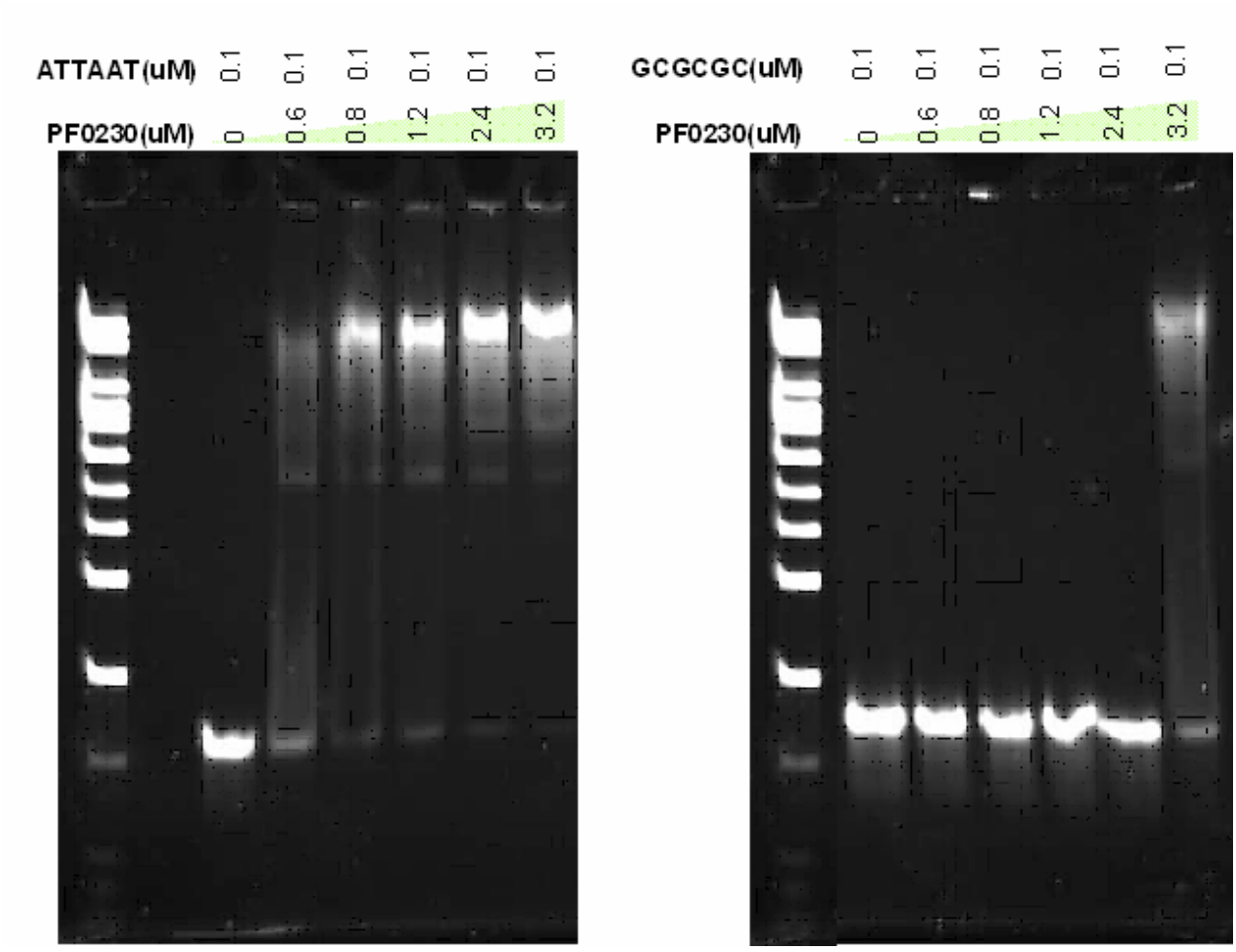
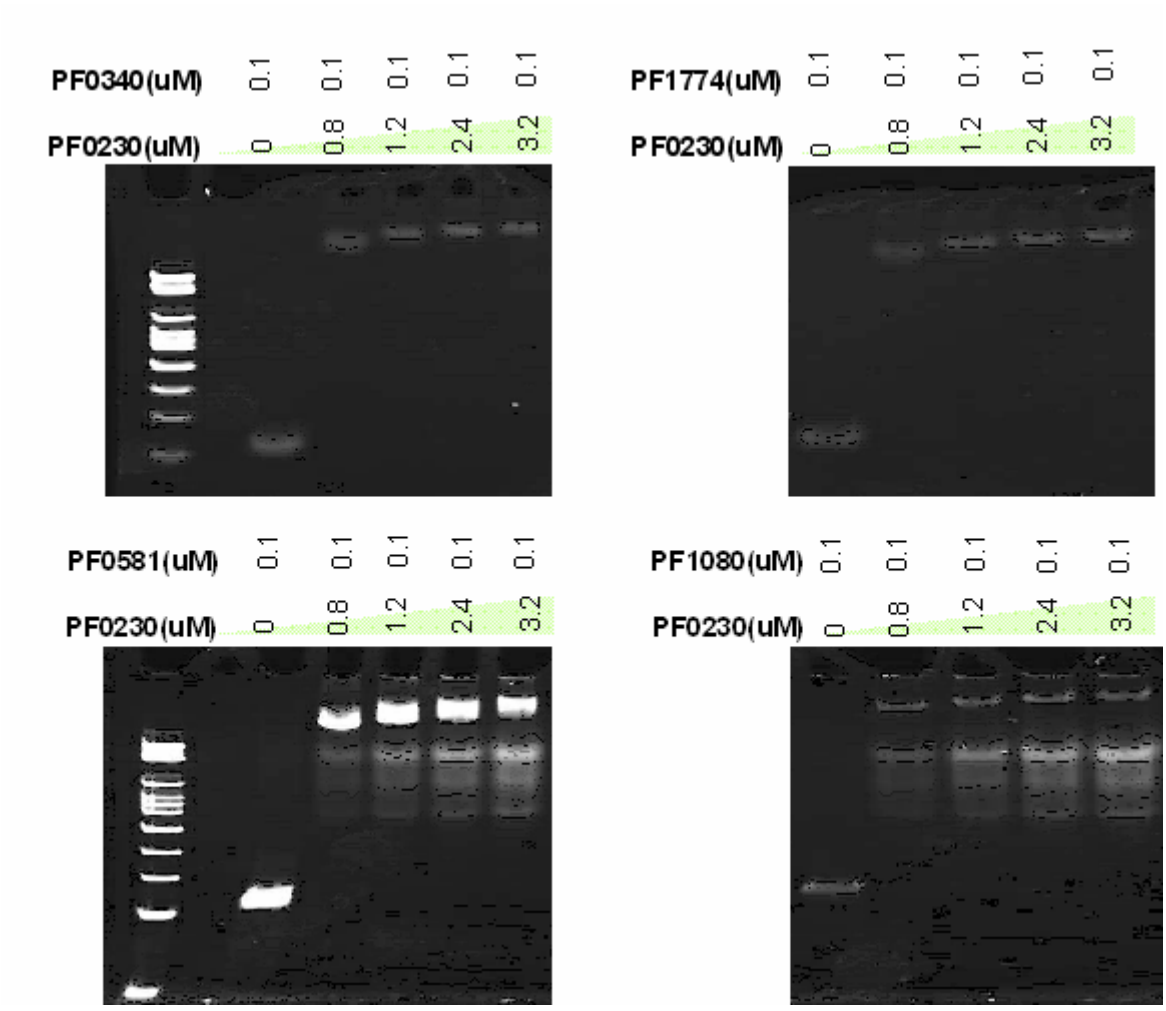

ATTAAT(uM) 0.1 0.1 0.1 0.1 0.1 0.1
PF0230(uM) 0 0.6 0.8 1.2 2.4 3.2
GCGCGC(uM) 0.1 0.1 0.1 0.1 0.1 0.1
PF0230(uM) 0 0.6 0.8 1.2 2.4 3.2

**Figure 4.11 EMSA of PF0230p with other promoters having binding motif, ATTAAT.**

PF0340, 1774, 0581, 1080, 0289, 0422 and PF1513 having ATTAAT binding motifs shows binding affinity with PF0230p


PF0340(uM) 0.1 0.1 0.1 0.1 0.1
PF0230(uM) 0 0.8 1.2 2.4 3.2
PF1774(uM) 0.1 0.1 0.1 0.1 0.1
PF0230(uM) 0 0.8 1.2 2.4 3.2
PF0581(uM) 0.1 0.1 0.1 0.1 0.1
PF0230(uM) 0 0.8 1.2 2.4 3.2
PF1080(uM) 0.1 0.1 0.1 0.1 0.1
PF0230(uM) 0 0.8 1.2 2.4 3.2

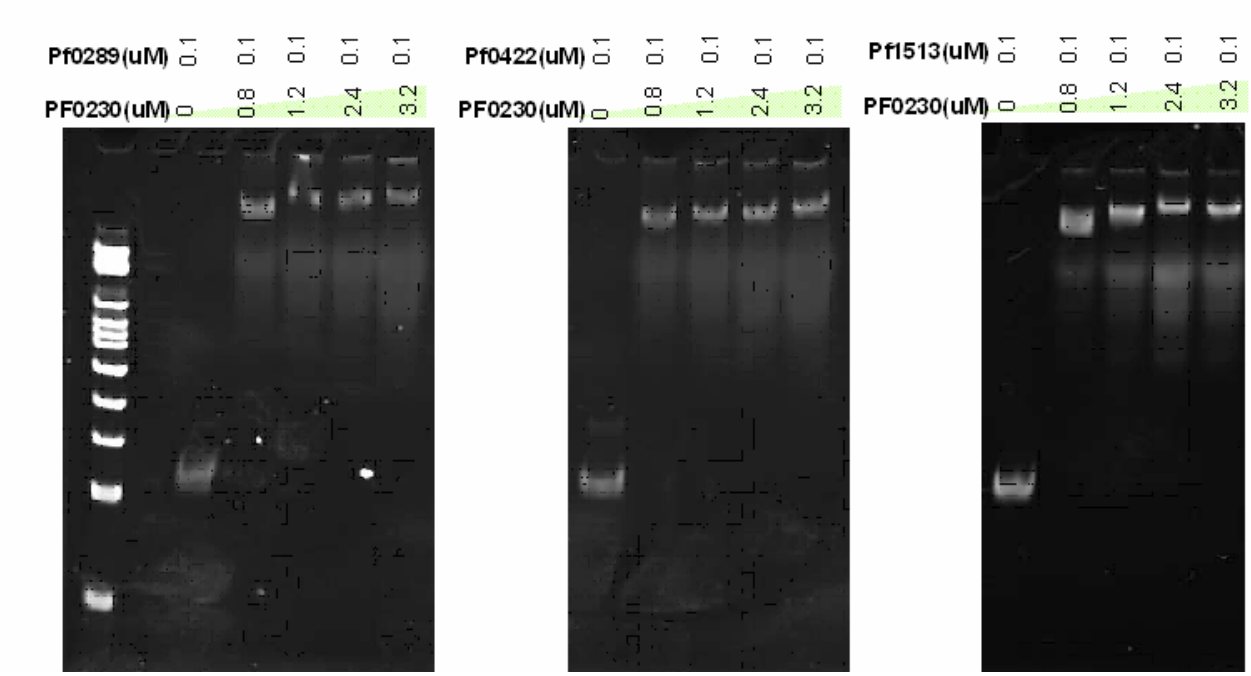
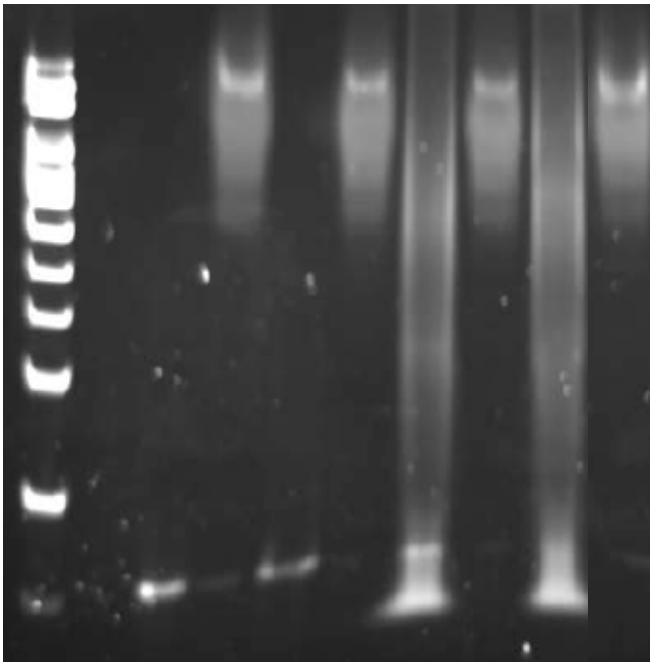

Pf0289(uM) 0.1 0.1 0.1 0.1 0.1
PF0230(uM) 0 0.8 1.2 2.4 3.2
Pf0422(uM) 0.1 0.1 0.1 0.1 0.1
PF0230(uM) 0 0.8 1.2 2.4 3.2
Pf1513(uM) 0.1 0.1 0.1 0.1 0.1
PF0230(uM) 0 0.8 1.2 2.4 3.2

**Figure 4.12 EMSA for PF0230p with TBP and TFB using between ATTAAT binding motif and mutated GCGCGC DNA. A**. Lane 1: 100bp standard DNA marker ; 2 : ATTAAT free DNA only; 3 : ATTAAT + PF0230p ; 4 : ATTAAT + TBP ; 5 : ATTAAT + PF0230p + TBP ; 6 : ATTAAT + TFB ; 7 : ATTAAT + PF0230p + TFB ; 8 : ATTAAT + TBP + TFB ; 9 : ATTAAT + PF0230p + TBP + TFB **B.** Lane 1: GCGCGC free DNA only ; 2 : GCGCGC + PF0230p ; 3 : GCGCGC + TBP ; 4 : GCGCGC + TFB ; 5 : GCGCGC + TBP + TFB; 6 : GCGCGC + PF0230p + TBP + TFB

**A**



**B**

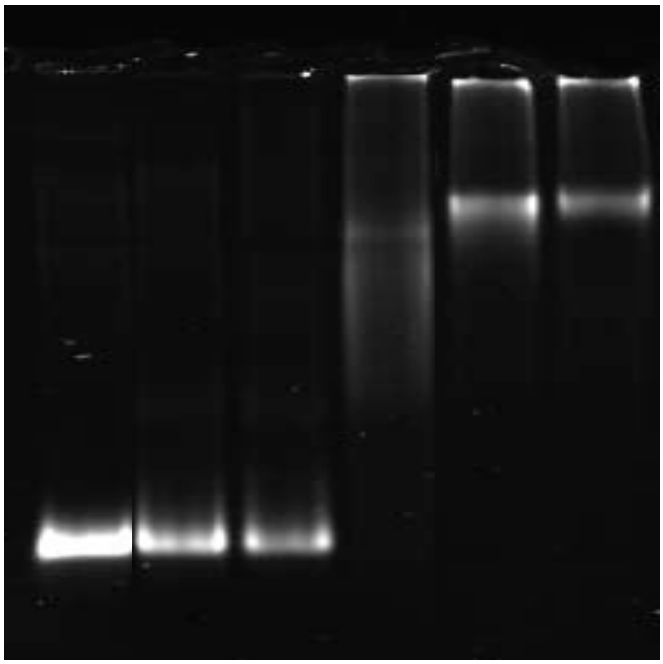
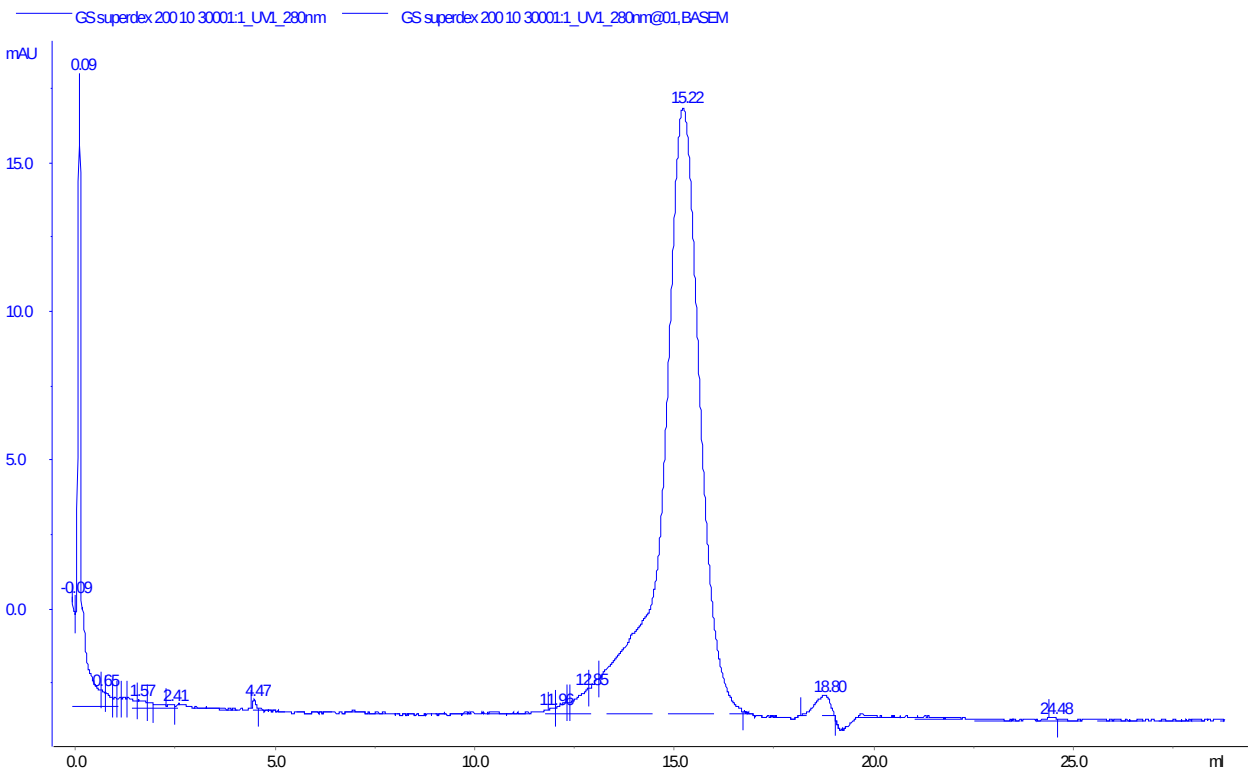


**Figure 4.13 Gel filtration result and calibration curve using Superdex200 10/30.** According to the gel filtration curve, PF0230p exist as dimer. A. PF0230p gel filtration curve. B. Superdex 200 10/30 standard calibration done by Aleksander Cvetcovic in Adams lab, University of Georgia.

**A**


GS superdex 200 10 30001:1_UV1_280nm
GS superdex 200 10 30001:1_UV1_280nm@01,BASEM
mAU
15.0
10.0
5.0
0.0
0.09
15.22
-0.09
0.65
1.57
2.41
4.47
11.96
12.85
18.80
24.48
0.0
5.0
10.0
15.0
20.0
25.0
ml

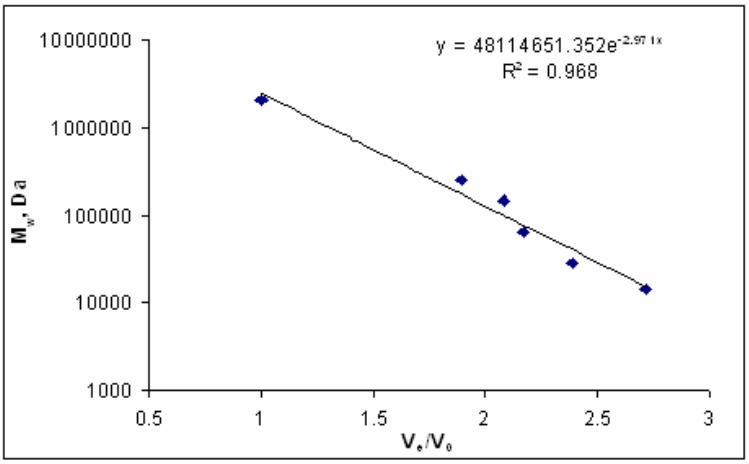
**B**

## Superdex 200 pg 10/30 calibration (ID 9416010)
## (done by Aleksander Cvetcovic)

Sample size: 100 ul

Flow rate: 0.75 ml/min

Buffer: 50mM Tris HCl (pH 8.05) 300 mM KCl




| Standards | Conc mg/ml | Ve (ml) | Vo (ml) | MW (Da) | Ve/V0 |
|---|---|---|---|---|---|
| dextran blue | 1 | 6.612 | 6.612 | 2000000 | 1 |
| catalase | 4 | 12.51 | 6.612 | 250000 | 1.89201 |
| ADH | 5 | 13.762 | 6.612 | 150000 | 2.08137 |
| albumin | 10 | 14.32 | 6.612 | 66000 | 2.16576 |
| carbonic anhydrase | 3 | 15.752 | 6.612 | 29000 | 2.38234 |
| lysozyme | 2 | 17.92 | 6.612 | 14300 | 2.71022 |

**Figure 4.14 The repression of PF1983 transcription by PF0230p in *in vitro* transcription**. Gel-purified PCR products of the PF1983 UOR, 300bp from -200 to +100 (relative to translation start) were used as the template in *in vitro* transcription experiments. The run-off transcripts are indicated by arrows. The added PF0230p ranged from 0-0.6 μM as indicated at the top. The *gdh* transcripts from the control experiments are shown in the rightmost two lanes. *In vitro* transcription assays with $his_6$-PF0230p performed by Annette Keese (unpublished data from the laboratory of Michael Thomm, University of Regensberg, Germany)

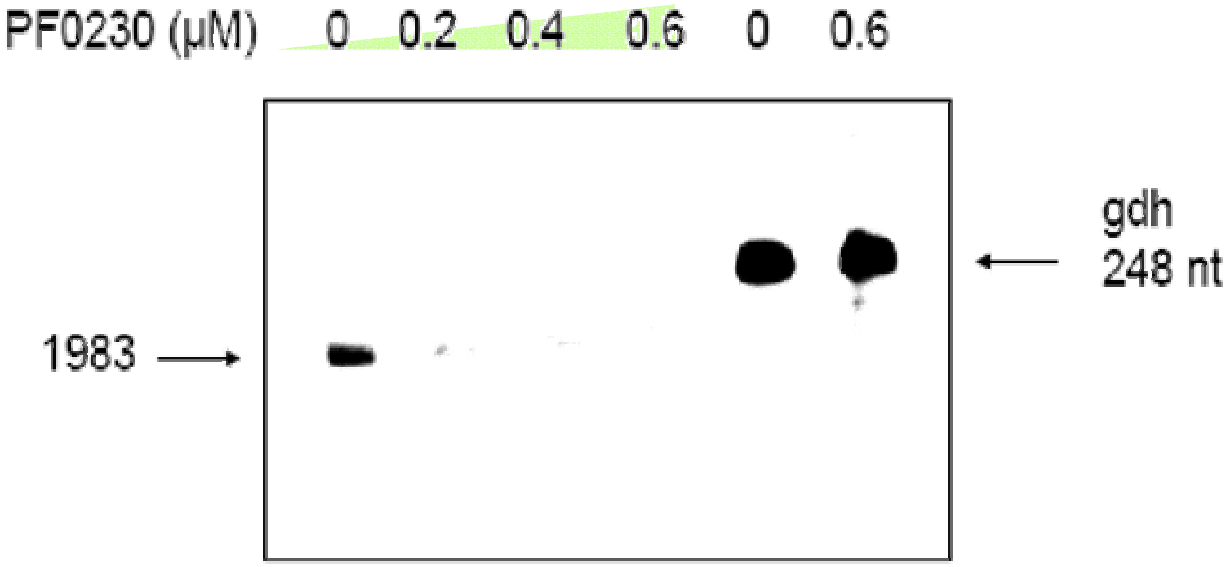

PF0230 (μM) 0 0.2 0.4 0.6 0 0.6
gdh
248 nt
1983

**Figure 4.15 The enhancement of PF1513 transcription by PF0230p in *in vitro* transcription.** Gel-purified PCR products of the PF1513 UOR, 300bp from -200 to +100 (relative to translation start) were used as the template in *in vitro* transcription experiments. The run-off transcripts are indicated by arrows. The added PF0230p ranged from 0-0.6 μM as indicated at the top. The *gdh* transcripts from the control experiments are shown in the rightmost two lanes. *In vitro* transcription assays with $his_6$-PF0230p performed by Antonia.Gindner (unpublished data from the laboratory of Michael Thomm, University of Regensberg, Germany)

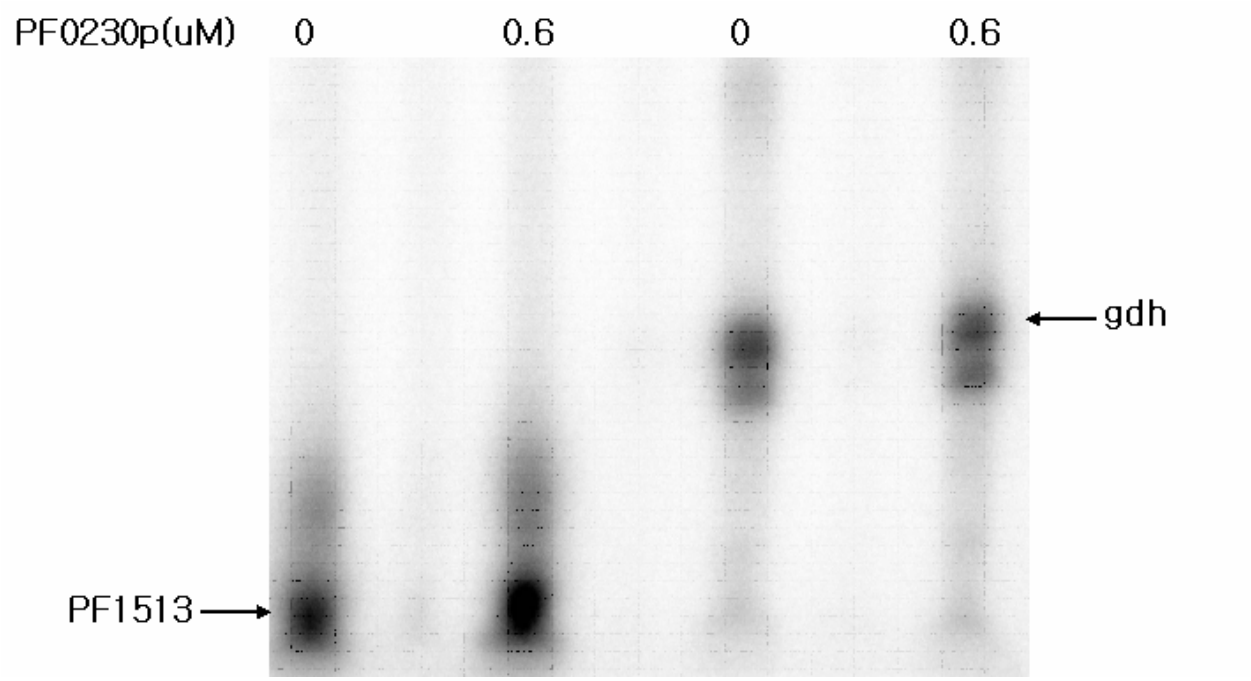

PF0230p(uM)
0
0.6
0
0.6
gdh
PF1513

**Figure 4.16 PF0340 transcription is unaffected by PF0230p in *in vitro* transcription.** Gel-purified PCR products of the PF0340 and PF1513 UOR, 300bp from -200 to +100 (relative to translation start) were used as the template in *in vitro* transcription experiments. The run-off transcripts are indicated by arrows. The added PF0230p ranged from 0 and 0.6 μM. The *gdh* transcripts from the control experiments are shown in the two rightmost lanes. *In vitro* transcription assays with $his_6$-PF0230p performed by Antonia Gindner (unpublished data from the laboratory of Michael Thomm, University of Regensberg, Germany)

| Lane | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Template-DNA | pf0340 | pf0340 | pf0340 | pf0340 | pf1513 | pf1513 | pf1513 | pf1513 | gdh-C20 | gdh-C20 | gdh-C20 | gdh-C20 |
| 0,2 µM TBP | - | + | - | + | - | + | - | + | - | + | - | + |
| 1 µl PF0230 | - | - | + | + | - | - | + | + | - | - | + | + |

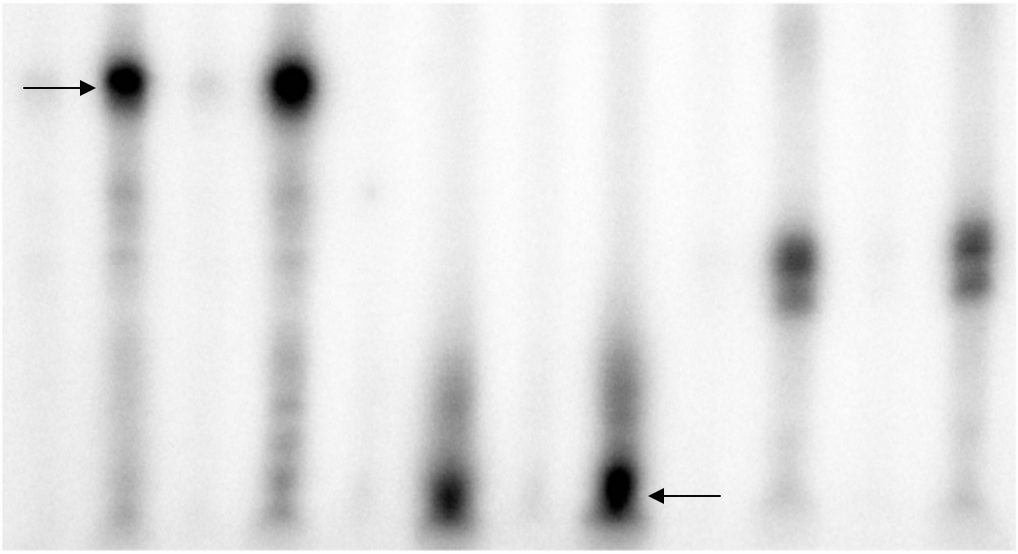


**Figure 4.17 EMSA analysis of the effect of cumene hydroperoxide and hydrogen peroxide on PF0230p binding to PF1983 UOR.** 0.1 μM PF1983 UOR and 1.6 and 3.2 μM PF0230p were mixed with constant amounts of cumene hydroperoxide or hydrogen peroxide (5 μM) as indicated in the figure. No obvious change in gel shift is observed with either peroxide.

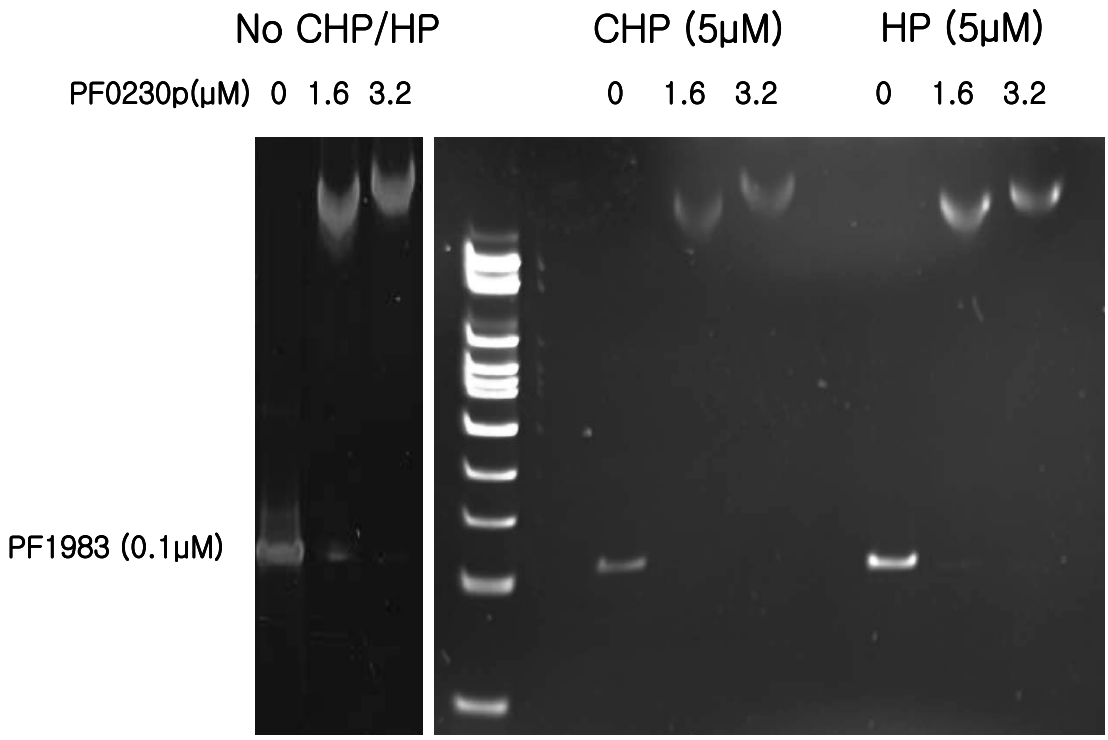
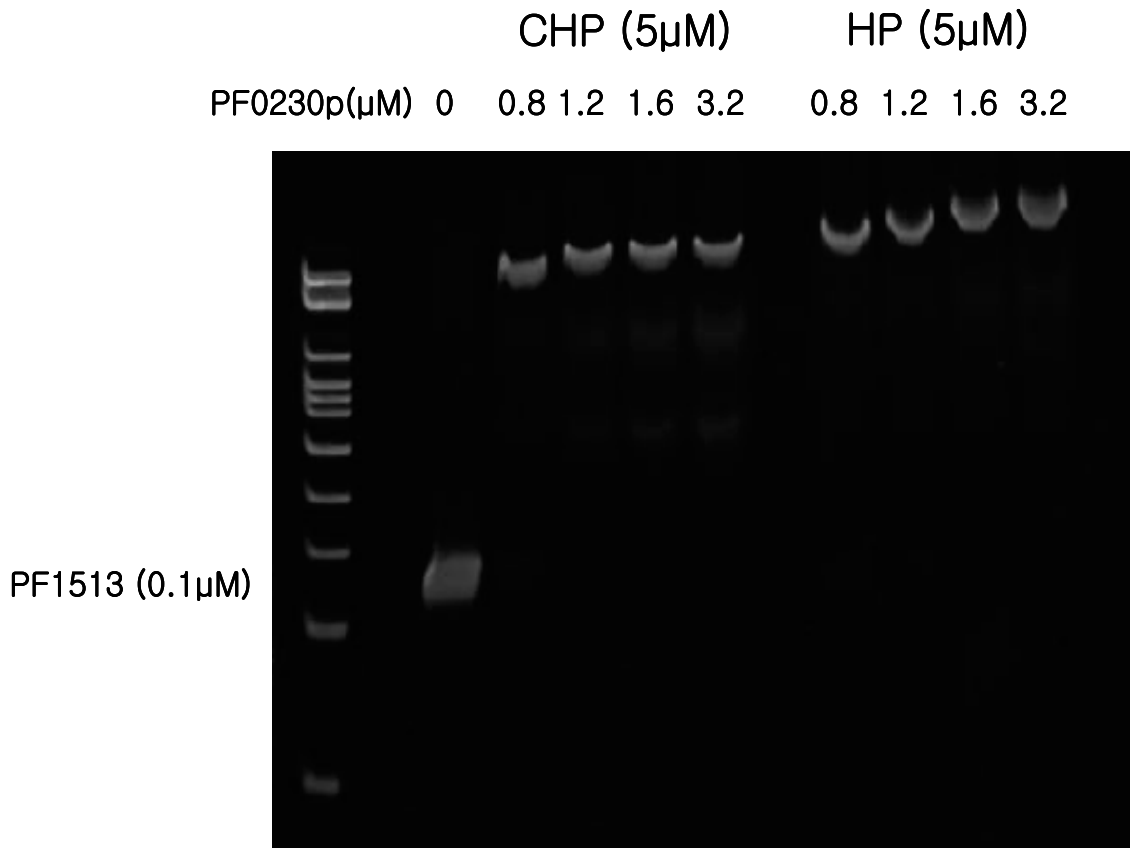

No CHP/HP
CHP (5μM)
HP (5μM)
PF0230p(μM) 0 1.6 3.2
0 1.6 3.2
0 1.6 3.2
PF1983 (0.1μM)

**Figure 4.18 EMSA analysis of the effect of cumene hydroperoxide and hydrogen peroxide on PF0230p binding to PF1513 UOR.** 0.1 μM PF1513 UOR and various amounts of PF0230p (0 – 3.2 μM) were mixed with constant amounts of cumene hydroperoxide or hydrogen peroxide (5 μM) as indicated in the figure. No obvious change in gel shift is observed with either peroxide.


CHP (5μM)
HP (5μM)
PF0230p(μM) 0 0.8 1.2 1.6 3.2 0.8 1.2 1.6 3.2
PF1513 (0.1μM)

CHAPTER 5

CONCLUSION AND DISCUSSION

## 5.1 PF0230p is probably a regulator of multiple genes

EMSA, DNase I footprinting, and *in vitro* transcription results show that PF0230p interacts specifically with PF1983 and PF1513 promoters and regulates the expression of these genes. Mutational analysis demonstrated that the SELEX-predicted ATTAAT binding motif sequence was important for PF0230p binding to DNA. In addition, EMSA demonstrated that PF0230p binds to the UORs of PF0340, PF1774, PF0581, PF1080, PF0289, and PF0422, all of which contain the ATTAAT binding motif. These results suggest that PF0230p is very likely a common transcriptional regulator of multiple genes.

Further investigation is necessary to prove or disprove this hypothesis. Regulation of PF0340, PF1774, PF0581, PF1080, PF0289, and PF0422 genes by PF0230p requires confirmation by *in vitro* transcription. Also, although we observed some evidence for activation of transcription of PF1513, a hypothetical protein with unknown function, it is still too early to conclude that PF0230p is a true transcriptional activator for PF1513 given the minor effect on transcript levels. However, the role of PF0230p as a transcriptional repressor of gene PF1983 has been clearly demonstrated in this work.

## 5.2 Predicted PF0230p tertiary structure based on the PH1932p crystal structure

As described in Chapter 4, PF0230p has 80% sequence identity with PH1932p. The N-terminal sequence of PF0230p is almost identical with PH1932p and is predicted to contain the same helices as are present in the winged helix-turn-helix motif of PH1932p. The crystal structure (**Figure 5.3**) of this protein has been published as part of a structural genomics effort, but its function is still unknown [118]. Structural analogs of PH1932p were determined by a DALI search [121]. These proteins are SmtB [122], MecI [123, 124], and MarR [125]. SmtB and MecI are transcription factors responsive to $Zn^{2+}$ ions and the antibiotic methicillin, respectively, while MarR is a transcription factor responsive to a wide range of compounds, including antibiotics and organic solvents.

The tertiary structure of PH1932p (**Figure 5.3**) shows that each protein monomer of the dimer has a winged-HTH motif structurally similar to those of bacterial transcription factors, as described above, known to act as transcriptional repressors in regulating resistance systems against cytotoxic compounds, such as antibiotics or heavy metal ions. The C-terminal domain of PH1932p is responsible for dimerization and forms a unique "hat-shaped" helical bundle. The inside of the hat is a possible effector binding site of this protein [118] (**Figure 5.4**). Gao et al. predicted a set of binding site residues that they suggested could possibly bind with other proteins or metals. These residues are Glu135, Glu166, Thr170, Tyr179 Trp167, Arg104, Lys108, and His163 (**Figure 5.4**), nearly all of which are conserved in PF0230p except Thr170, replaced by Met170 in PF0230p. Whether PF0230p is responsive to any factor(s) for regulation of multiple genes, or whether any other protein(s) are involved in the regulation needs to be further investigated.

### 5.3 Suggested regulation mechanism of PF0230p

The two known primary mechanisms of transcription repression in archaea are (1) interference with TBP/TFB recognition of TATA/BRE elements on the gene promoter (e.g. TrmB [59]) and (2) obstructing RNA polymerase recruitment at the transcription initiation sites (e.g. LrpA [56]). Our footprinting study indicated that the ATTAAT recognition motif lies between -137 and -132 (relative to the translation start site), in the middle of the footprint on the PF1983 promoter (see **Figure 4.8**) and this sequence was confirmed as important for PF0230p binding by mutational analysis. Using the archaeal consensus sequences for TATA and BRE (**Table 5.1**), a putative TATA/BRE of PF1983 can be identified from -143 to -148 relative to the translation start site, which also lies within the protected region determined by DNase I footprinting (–164 to –104). This TATA/BRE positioning is supported by the location of the PF1983 transcription start site mapped at position -119 (relative to the translation start site) by primer extension [126] by Annette Keese. This transcription start site also lies within the PF0230p footprint.

The extended footprint of PF0230p on the PF1983 promoter encompasses the BRE/TATA promoter sequence, the SELEX-identified PF0230p binding sequence, and the PF1983 transcription start site. Given this arrangement of DNA sequence elements, it is not possible to choose between the two common repression mechanisms introduced above. We have some evidence from EMSA that PF0230p and TBP cannot both occupy their respective binding sites at the same time, which makes sense regardless of whether these sites are the same (i.e., that PF0230p also binds to TATA). If this is true, then blocking the binding of TBP/TFB to the promoter is likely involved in the mechanism of repression of PF1983 transcription by PF0230p (Figure 5.7), although it need not be the only contributor.

**Table 5.1 Archaeal BRE and TATA consensus sequences**

| Archaeal class / group | TATA box[a] | BRE[a] |
|---|---|---|
| Halophiles | -29 TTTWWW -24 | none predicted |
| Methanogens | -30 YTTATATA -23 | none predicted |
| *Sulfolobus* | -30 YTTTTAAA -23 | -36 RNWAAW -31 |
| *Pyrococcus* | -29 TTWWWAW -23 | -36 VRAAA -32 |

[a]Numbers indicate the position of the motifs relative to the transcription start site. 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. [b]This table was taken from [116].

### 5.4 The Effector(s) of PF0230p

The N-terminus of PF0230p is a conserved helix-turn-helix DNA-binding domain, and we found that there is homology between SmtB and PF0230p by Conserved Domain Search (NCBI, www.ncbi.nlm.nih.gov). SmtB is a structurally well characterized metalloregulatory protein which can bind metals such as Zn, Co, and Ni [119]. His105 and His106 have been suggested as a Zn(II) binding site in SmtB by mutational analysis [119]. PF0230p also has sequential His50 and His51 residues. Thus, we chose to test whether PF0230p can bind Zn(II) and affect DNA binding (no Zn was detected in heterologously expressed PF0230p). However, the EMSA results with Zn(II) did not show any obvious effect on protein-DNA interaction under our experimental conditions (**Figure 5.6**).

**Figure 5.1 Crystal structure of PH1932**. Ribbon representation of the PH1932 protein dimer. The subunits are colored blue and red, respectively. The biological unit of this protein was also supposed to be a dimer based on the estimated molecular weight using size exclusion chromatography (SEC). The second helix in the winged-HTH motif, a putative DNA-recognition helix, is labeled red. Pictures are taken from the reference [118].

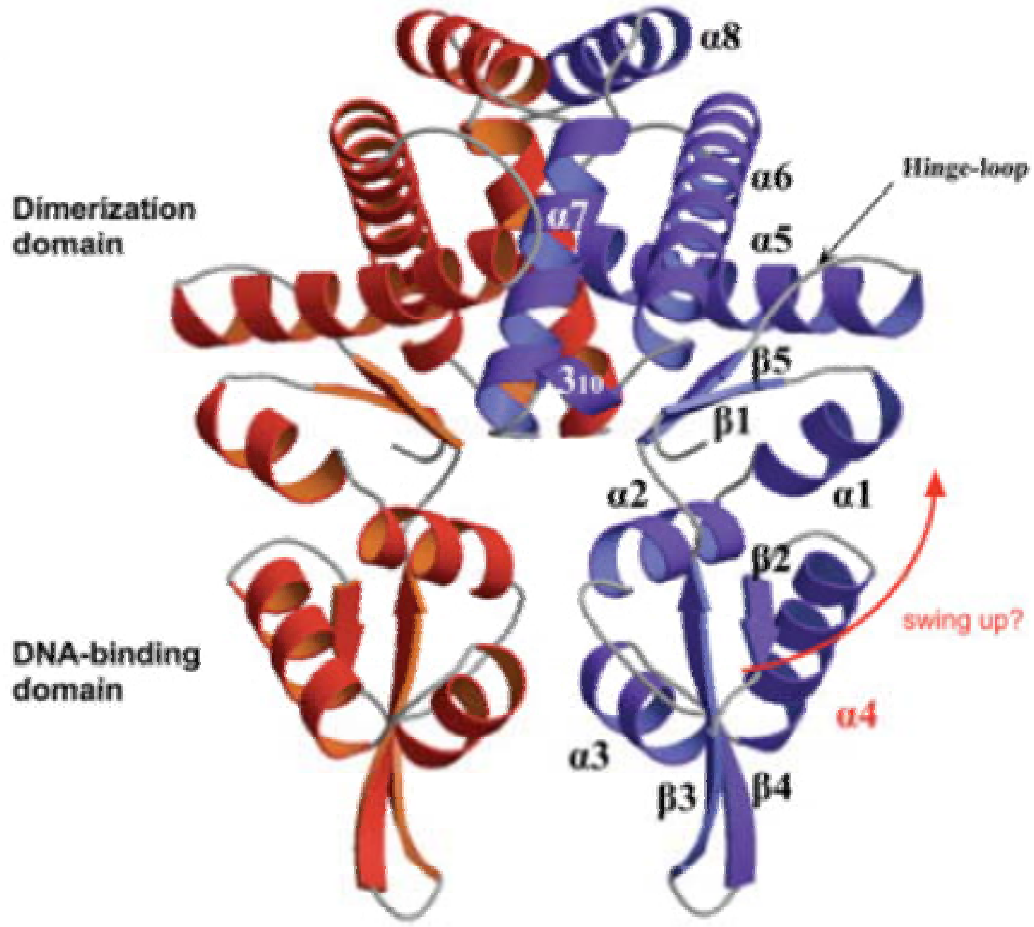

α8
α6
Hinge-loop
Dimerization domain
α7
α5
β5
310
β1
α2
α1
β2
swing up?
DNA-binding domain
α4
α3
β3
β4

**Figure 5.2 Close-up molecular surface representation around the putative ligand-binding pocket observed on the inside of the dimerization domains.** Red and blue on the molecular surface represent negatively or positively charged potentials, respectively. As shown here, the pocket has a significant negative charge. Bound water molecules are shown as yellow balls, and candidate side chains for ligand binding are represented as white stick models. Symmetry-related mates are also indicated as lime-green sticks with dashed labels. Each yellow dotted line shows hydrogen bond between the water and the ligated side chain. Pictures are taken from the reference [118].

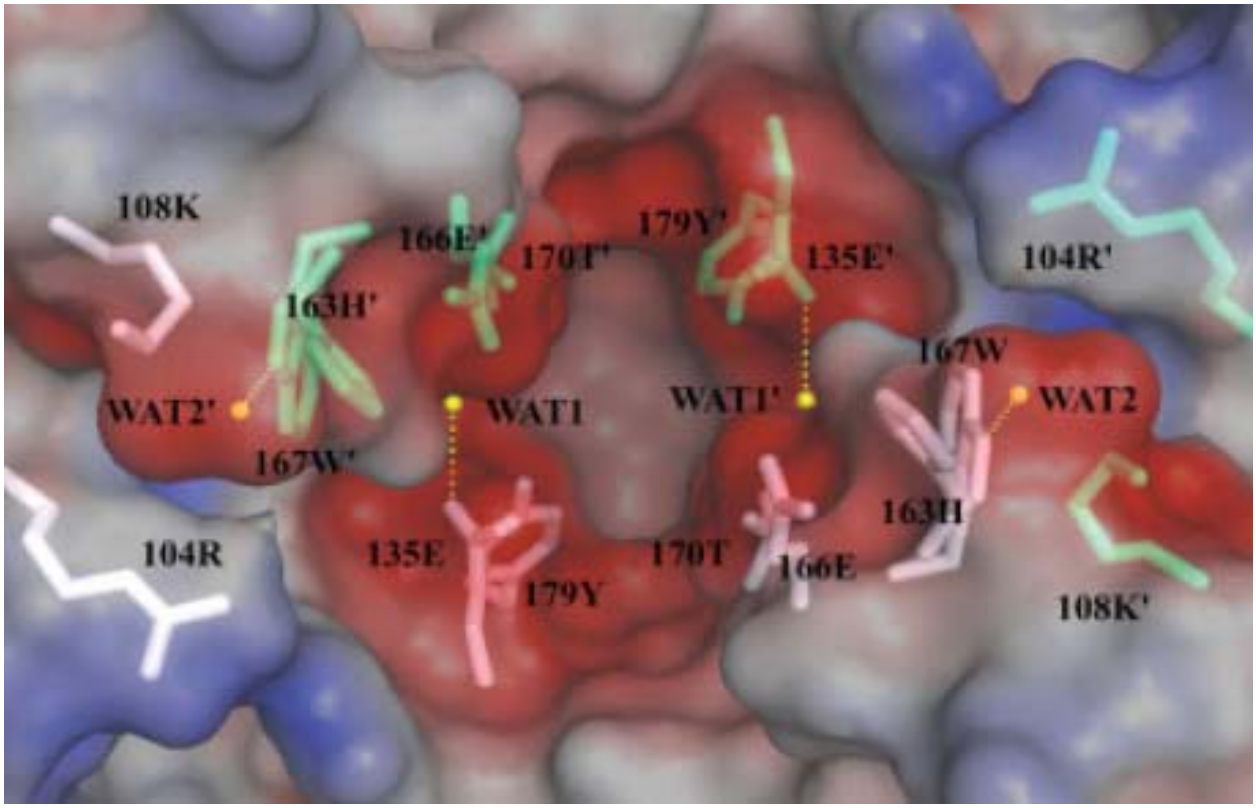

108K
166E'
170T'
179Y'
135E'
104R'
163H'
167W
WAT2'
WAT1
WAT1'
WAT2
167W'
104R
135E
170T
166E
163H
179Y
108K'

**Figure 5.3 Sequence alignment between PF0230p and PH1932p.** Red arrows show binding residues with water. Possibly can replace with metals and other proteins and they all exist both PH1932p and PF0230p except threonine170 residue (blue arrow).

E-value = 4e-66, Bit score = 253, Aligned length = 190, Sequence Identity = 80%

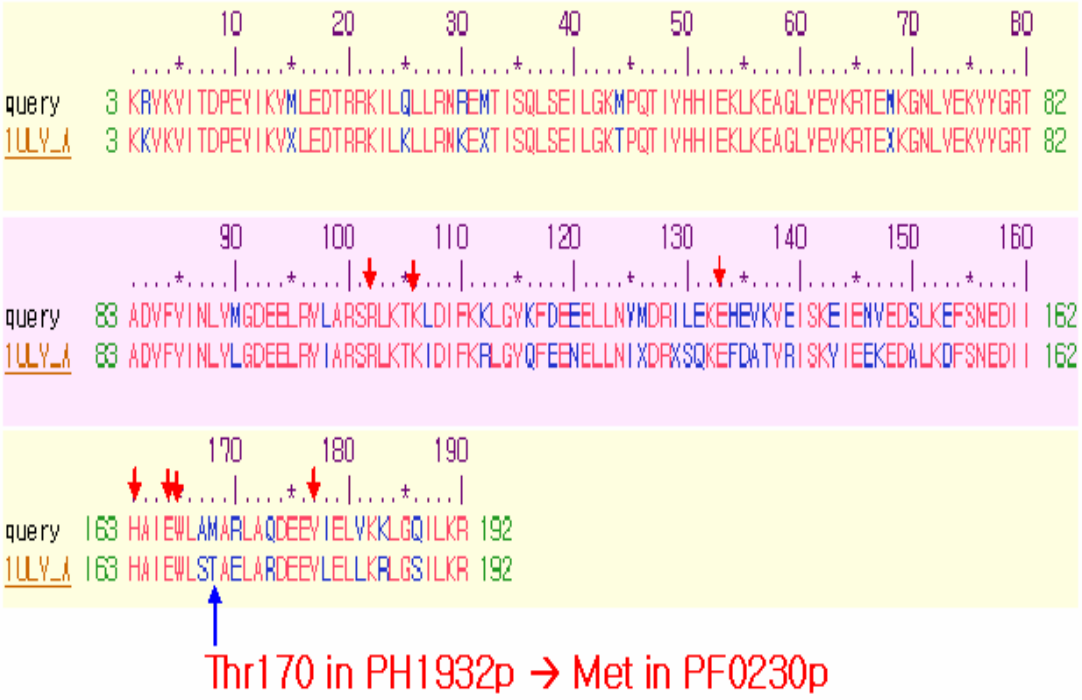
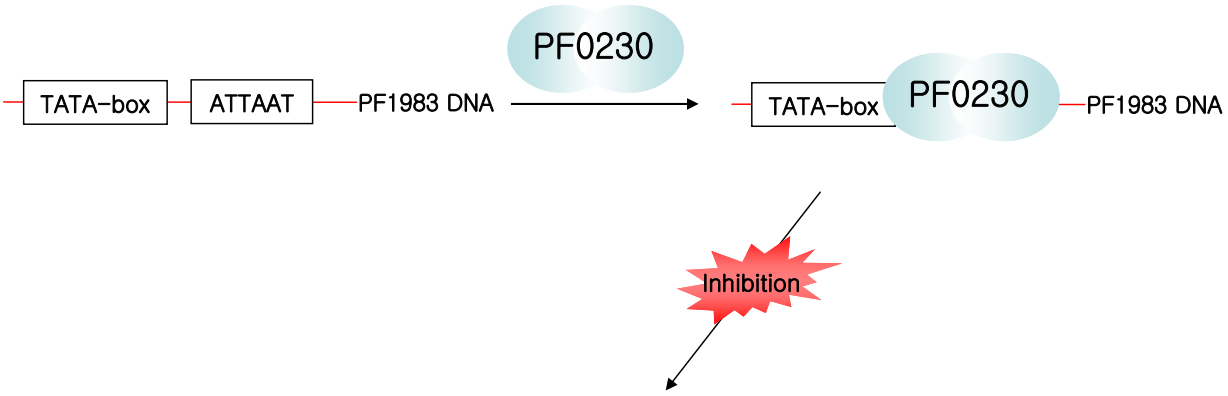


**Figure 5.4 Proposed regulation mechanism of PF0230p with PF1983.** PF0230p dimer binds to ATTAAT near the TATA box and block TBP recognition of TATA box




**Block TBP recognition of TATA box**

**Figure 5.5 Conserved domain search and sequence analysis for the metal binding site.**

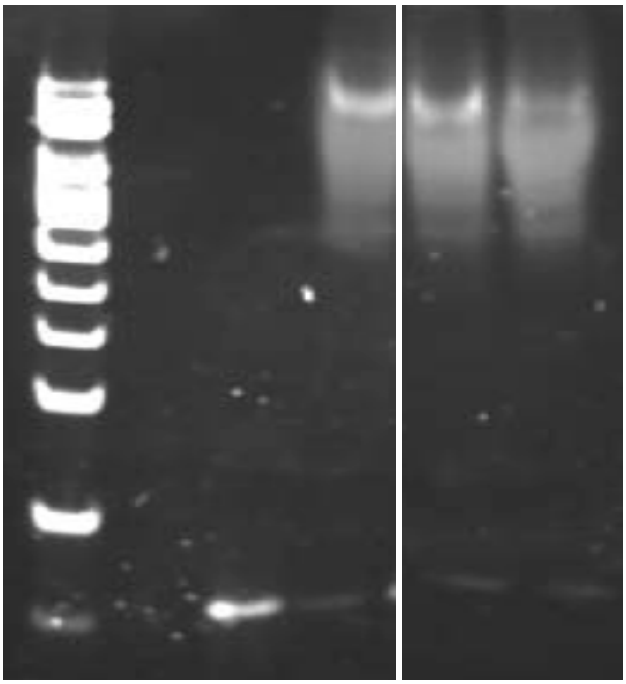
Metal binding site for SmtB is predicted LCVGD [119], but there is no metal binding site in PF0230p ("query") except H(Histidine)H(Histidine) metal binding site.


1SMT_B 40 FAVL.[2].PNRLRLLSLLARS.[1].LCVGDLAQAIG.[1].SESAVSHQLRSLRNLRLVSYRK.[4].YYYQL.[3]. 105
query 14 IKVW.[2].DTRRKILQLLRNR.[1].WTISQLSEILG.[1].MPQTIVHHIEKLKEAGLVEVKR.[9].KYYGR.[4]. 85
1SMT_B 106 HIVALY.[5]. 116
query 86 FYINLY.[5]. 96

**Figure 5.6 EMSA for the metal binding site.**

PF0230p has His50-His51 residues which is responsible for Zn(II) binding motif.

Lane 1 : ATTAAT binding motif only ; 2 : ATTAAT + PF0230p ; 3 : ATTAAT + PF0230p+100μM $ZnSO_4$ ; 4 : ATTAAT + PF0230p+200μM $ZnSO_4$



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. Winker, S. and C.R. Woese, *A definition of the domains Archaea, Bacteria and Eucarya in terms of small subunit ribosomal RNA characteristics.* Syst Appl Microbiol, 1991. **14**(4): p. 305-10.
3. Woese, C.R., *Evolutionary questions: the "progenote".* Science, 1990. **247**(4944): p. 789.
4. Rothschild, L.J. and R.L. Mancinelli, *Life in extreme environments.* Nature, 2001. **409**(6823): p. 1092-101.
5. Karner, M.B., E.F. DeLong, and D.M. Karl, *Archaeal dominance in the mesopelagic zone of the Pacific Ocean.* Nature, 2001. **409**(6819): p. 507-10.
6. Schleper, C., G. Jurgens, and M. Jonuscheit, *Genomic studies of uncultivated archaea.* Nat Rev Microbiol, 2005. **3**(6): p. 479-88.
7. Olsen, G.J. and C.R. Woese, *Lessons from an Archaeal genome: what are we learning from Methanococcus jannaschii?* Trends Genet, 1996. **12**(10): p. 377-9.
8. Lecompte, O., et al., *Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale.* Nucleic Acids Res, 2002. **30**(24): p. 5382-90.
9. Hampsey, M., *Molecular genetics of the RNA polymerase II general transcriptional machinery.* Microbiol Mol Biol Rev, 1998. **62**(2): p. 465-503.
10. Londei, P., *Evolution of translational initiation: new insights from the archaea.* FEMS Microbiol Rev, 2005. **29**(2): p. 185-200.
11. Myllykallio, H., et al., *Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon.* Science, 2000. **288**(5474): p. 2212-5.
12. Bell, S.D. and S.P. Jackson, *Mechanism and regulation of transcription in archaea.* Curr Opin Microbiol, 2001. **4**(2): p. 208-13.
13. Grabowski, B. and Z. Kelman, *Archeal DNA replication: eukaryal proteins in a bacterial context.* Annu Rev Microbiol, 2003. **57**: p. 487-516.
14. Robinson, N.P. and S.D. Bell, *Origins of DNA replication in the three domains of life.* Febs J, 2005. **272**(15): p. 3757-66.
15. 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.
16. Reinberg, D., et al., *The RNA polymerase II general transcription factors: past, present, and future.* Cold Spring Harb Symp Quant Biol, 1998. **63**: p. 83-103.
17. Soppa, J., *Transcription initiation in Archaea: facts, factors and future aspects.* Mol Microbiol, 1999. **31**(5): p. 1295-305.
18. Villard, J., *Transcription regulation and human diseases.* Swiss Med Wkly, 2004. **134**(39-40): p. 571-9.
19. 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.
20. Geiduschek, E.P. and M. Ouhammouch, *Archaeal transcription and its regulators.* Mol

Microbiol, 2005. **56**(6): p. 1397-407.
21. Ouhammouch, M., *Transcriptional regulation in Archaea.* Curr Opin Genet Dev, 2004. **14**(2): p. 133-8.
22. Reeves, T.M., et al., *Matrix metalloproteinase inhibition alters functional and structural correlates of deafferentation-induced sprouting in the dentate gyrus.* J Neurosci, 2003. **23**(32): p. 10182-9.
23. 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.
24. Bell, S.D. and S.P. Jackson, *Transcription in Archaea.* Cold Spring Harb Symp Quant Biol, 1998. **63**: p. 41-51.
25. Hanzelka, B.L., T.J. Darcy, and J.N. Reeve, *TFE, an archaeal transcription factor in Methanobacterium thermoautotrophicum related to eucaryal transcription factor TFIIEalpha.* J Bacteriol, 2001. **183**(5): p. 1813-8.
26. 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.
27. Littlefield, O., Y. Korkhin, and P.B. Sigler, *The structural basis for the oriented assembly of a TBP/TFB/promoter complex.* Proc Natl Acad Sci U S A, 1999. **96**(24): p. 13668-73.
28. Thomm, M., *Archaeal transcription factors and their role in transcription initiation.* FEMS Microbiol Rev, 1996. **18**(2-3): p. 159-71.
29. Kyrpides, N.C. and C.A. Ouzounis, *Whole-genome sequence annotation: 'Going wrong with confidence'.* Mol Microbiol, 1999. **32**(4): p. 886-7.
30. Asturias, F.J., *RNA polymerase II structure, and organization of the preinitiation complex.* Curr Opin Struct Biol, 2004. **14**(2): p. 121-9.
31. 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.
32. Woychik, N.A. and M. Hampsey, *The RNA polymerase II machinery: structure illuminates function.* Cell, 2002. **108**(4): p. 453-63.
33. Gourse, R.L., W. Ross, and T. Gaal, *UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition.* Mol Microbiol, 2000. **37**(4): p. 687-95.
34. 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.
35. 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.
36. Struhl, K., *Fundamentally different logic of gene regulation in eukaryotes and prokaryotes.* Cell, 1999. **98**(1): p. 1-4.
37. Brown, J.R. and W.F. Doolittle, *Archaea and the prokaryote-to-eukaryote transition.* Microbiol Mol Biol Rev, 1997. **61**(4): p. 456-502.
38. Takayanagi, K. and T. Mizuno, *Activation of the osmoregulated ompF and ompC genes by the OmpR protein in Escherichia coli: a study involving chimeric promoters.* J Biochem, 1992. **112**(1): p. 1-6.
39. Kim, J. and S. DasSarma, *Isolation and chromosomal distribution of natural Z-DNA-forming sequences in Halobacterium halobium.* J Biol Chem, 1996. **271**(33): p. 19724-31.
40. Yang, C.F., et al., *Genetic and topological analyses of the bop promoter of*

*Halobacterium halobium: stimulation by DNA supercoiling and non-B-DNA structure.* J Bacteriol, 1996. **178**(3): p. 840-5.
41. Takayanagi, S., et al., *Chromosomal structure of the halophilic archaebacterium Halobacterium salinarium.* J Bacteriol, 1992. **174**(22): p. 7207-16.
42. Reeves, W.H., et al., *Initiation of autoimmunity to self-proteins complexed with viral antigens.* Ann N Y Acad Sci, 1997. **815**: p. 139-54.
43. Grayling, R.A., K. Sandman, and J.N. Reeve, *Histones and chromatin structure in hyperthermophilic Archaea.* FEMS Microbiol Rev, 1996. **18**(2-3): p. 203-13.
44. Kruger, K., et al., *The transcriptional activator GvpE for the halobacterial gas vesicle genes resembles a basic region leucine-zipper regulatory protein.* J Mol Biol, 1998. **279**(4): p. 761-71.
45. Ken, R. and N.R. Hackett, *Halobacterium halobium strains lysogenic for phage phi H contain a protein resembling coliphage repressors.* J Bacteriol, 1991. **173**(3): p. 955-60.
46. Ruepp, A. and J. Soppa, *Fermentative arginine degradation in Halobacterium salinarium (formerly Halobacterium halobium): genes, gene products, and transcripts of the arcRACB gene cluster.* J Bacteriol, 1996. **178**(16): p. 4942-7.
47. Gelfand, M.S., et al., *Comparative analysis of regulatory patterns in bacterial genomes.* Brief Bioinform, 2000. **1**(4): p. 357-71.
48. Klenk, H.P., et al., *The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus.* Nature, 1997. **390**(6658): p. 364-70.
49. Schmitt, M.P., E.M. Twiddy, and R.K. Holmes, *Purification and characterization of the diphtheria toxin repressor.* Proc Natl Acad Sci U S A, 1992. **89**(16): p. 7576-80.
50. Bell, S.D., *Use of Passy-Muir tracheostomy speaking valve in mechanically ventilated neurological patients.* Crit Care Nurse, 1996. **16**(1): p. 63-8.
51. 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.
52. 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.
53. Bell, S.D. and S.P. Jackson, *Mechanism of autoregulation by an archaeal transcriptional repressor.* J Biol Chem, 2000. **275**(41): p. 31624-9.
54. 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.
55. 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.
56. Dahlke, I. and M. Thomm, *A Pyrococcus homolog of the leucine-responsive regulatory protein, LrpA, inhibits transcription by abrogating RNA polymerase recruitment.* Nucleic Acids Res, 2002. **30**(3): p. 701-10.
57. 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.
58. 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.
59. 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.
60. Offner, S. and F. Pfeifer, *Complementation studies with the gas vesicle-encoding p-vac region of Halobacterium salinarium PHH1 reveal a regulatory role for the p-gvpDE genes.* Mol Microbiol, 1995. **16**(1): p. 9-19.
61. 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.
62. 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.
63. Ouhammouch, M., et al., *Promoter architecture and response to a positive regulator of archaeal transcription.* Mol Microbiol, 2005. **56**(3): p. 625-37.
64. Cui, Y., et al., *A consensus sequence for binding of Lrp to DNA.* J Bacteriol, 1995. **177**(17): p. 4872-80.
65. Jafri, S., S. Chen, and J.M. Calvo, *ilvIH operon expression in Escherichia coli requires Lrp binding to two distinct regions of DNA.* J Bacteriol, 2002. **184**(19): p. 5293-300.
66. Jain, D., et al., *Structure of a ternary transcription activation complex.* Mol Cell, 2004. **13**(1): p. 45-53.
67. Baliga, N.S. and S. DasSarma, *Saturation mutagenesis of the TATA box and upstream activator sequence in the haloarchaeal bop gene promoter.* J Bacteriol, 1999. **181**(8): p. 2513-8.
68. Baliga, N.S., et al., *Genomic and genetic dissection of an archaeal regulon.* Proc Natl Acad Sci U S A, 2001. **98**(5): p. 2521-5.
69. 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.
70. Baliga, N.S., et al., *Is gene expression in Halobacterium NRC-1 regulated by multiple TBP and TFB transcription factors?* Mol Microbiol, 2000. **36**(5): p. 1184-5.
71. Paget, M.S. and J.D. Helmann, *The sigma70 family of sigma factors.* Genome Biol, 2003. **4**(1): p. 203.
72. Baliga, N.S., et al., *Systems level insights into the stress response to UV radiation in the halophilic archaeon Halobacterium NRC-1.* Genome Res, 2004. **14**(6): p. 1025-35.
73. Thompson, D.K., J.R. Palmer, and C.J. Daniels, *Expression and heat-responsive regulation of a TFIIB homologue from the archaeon Haloferax volcanii.* Mol Microbiol, 1999. **33**(5): p. 1081-92.
74. Shockley, K.R., et al., *Heat shock response by the hyperthermophilic archaeon Pyrococcus furiosus.* Appl Environ Microbiol, 2003. **69**(4): p. 2365-71.
75. Fiala, J., *[Use of the Fenwal CS-3000 automatic blood cell separator for cytopheresis (review)].* Vnitr Lek, 1986. **32**(2): p. 173-9.
76. Poole, F.L., 2nd, et al., *Defining genes in the genome of the hyperthermophilic archaeon Pyrococcus furiosus: implications for all microbial genomes.* J Bacteriol, 2005. **187**(21): p. 7325-32.
77. Robb, A. and J.D. Brown, *Protein transport: two translocons are better than one.* Mol Cell, 2001. **8**(3): p. 484-6.
78. Schut, G.J., et al., *Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: Pyrococcus furiosus grown on carbohydrates or peptides.* J Bacteriol, 2003.

**185**(13): p. 3935-47.
79. Brown, D.M., J.A. Upcroft, and P. Upcroft, *A H2O-producing NADH oxidase from the protozoan parasite Giardia duodenalis.* Eur J Biochem, 1996. **241**(1): p. 155-61.
80. Hecht, H.J., et al., *Crystal structure of NADH oxidase from Thermus thermophilus.* Nat Struct Biol, 1995. **2**(12): p. 1109-14.
81. Gomes, C.M. and M. Teixeira, *The NADH oxidase from the thermoacidophilic archaea Acidianus ambivalens: isolation and physicochemical characterisation.* Biochem Biophys Res Commun, 1998. **243**(2): p. 412-5.
82. Liu, X.L. and R.K. Scopes, *Cloning, sequencing and expression of the gene encoding NADH oxidase from the extreme anaerobic thermophile Thermoanaerobium brockii.* Biochim Biophys Acta, 1993. **1174**(2): p. 187-90.
83. Claiborne, A., et al., *Protein-sulfenic acids: diverse roles for an unlikely player in enzyme catalysis and redox regulation.* Biochemistry, 1999. **38**(47): p. 15407-16.
84. Ward, D.E., et al., *The NADH oxidase from Pyrococcus furiosus. Implications for the protection of anaerobic hyperthermophiles against oxidative stress.* Eur J Biochem, 2001. **268**(22): p. 5816-23.
85. Ward, D.E., et al., *Purification and characterization of the alanine aminotransferase from the hyperthermophilic Archaeon pyrococcus furiosus and its role in alanine production.* J Bacteriol, 2000. **182**(9): p. 2559-66.
86. Lipman, R.M., B.J. Tripathi, and R.C. Tripathi, *Cataracts induced by microwave and ionizing radiation.* Surv Ophthalmol, 1988. **33**(3): p. 200-10.
87. Goldoni, J., *[Non-ionizing radiation protection in medicine].* Arh Hig Rada Toksikol, 1994. **45**(2): p. 175-87.
88. Shalguev, V.I., et al., *Rad51 protein from the thermotolerant yeast Pichia angusta as a typical but thermodependent member of the Rad51 family.* Eukaryot Cell, 2004. **3**(6): p. 1567-73.
89. Ramsay, B., et al., *Dps-like protein from the hyperthermophilic archaeon Pyrococcus furiosus.* J Inorg Biochem, 2006. **100**(5-6): p. 1061-8.
90. Grant, R.A., et al., *The crystal structure of Dps, a ferritin homolog that binds and protects DNA.* Nat Struct Biol, 1998. **5**(4): p. 294-303.
91. Yamamoto, Y., et al., *An iron-binding protein, Dpr, from Streptococcus mutans prevents iron-dependent hydroxyl radical formation in vitro.* J Bacteriol, 2002. **184**(11): p. 2931-9.
92. Zhao, G., et al., *Iron and hydrogen peroxide detoxification properties of DNA-binding protein from starved cells. A ferritin-like DNA-binding protein of Escherichia coli.* J Biol Chem, 2002. **277**(31): p. 27689-96.
93. Su, M., et al., *The so-called Listeria innocua ferritin is a Dps protein. Iron incorporation, detoxification, and DNA protection properties.* Biochemistry, 2005. **44**(15): p. 5572-8.
94. Kauko, A., et al., *Iron incorporation in Streptococcus suis Dps-like peroxide resistance protein Dpr requires mobility in the ferroxidase center and leads to the formation of a ferrihydrite-like core.* J Mol Biol, 2006. **364**(1): p. 97-109.
95. Bell, S.D., *Archaeal transcriptional regulation--variation on a bacterial theme?* Trends Microbiol, 2005. **13**(6): p. 262-5.
96. Cramer, P., *Multisubunit RNA polymerases.* Curr Opin Struct Biol, 2002. **12**(1): p. 89-97.
97. Schut, G.J., J. Zhou, and M.W. Adams, *DNA microarray analysis of the hyperthermophilic archaeon Pyrococcus furiosus: evidence for anNew type of sulfur-reducing enzyme complex.* J Bacteriol, 2001. **183**(24): p. 7027-36.

98. Weagant, S.D. and A.J. Bound, *Evaluation of techniques for enrichment and isolation of Escherichia coli O157:H7 from artificially contaminated sprouts.* Int J Food Microbiol, 2001. **71**(1): p. 87-92.
99. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.
100. Clauser, K.R., P. Baker, and A.L. Burlingame, *Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching.* Anal Chem, 1999. **71**(14): p. 2871-82.
101. 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.
102. 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.
103. Marchler-Bauer, A., et al., *CDD: a curated Entrez database of conserved domain alignments.* Nucleic Acids Res, 2003. **31**(1): p. 383-7.
104. Marchler-Bauer, A. and S.H. Bryant, *CD-Search: protein domain annotations on the fly.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W327-31.
105. Marchler-Bauer, A., et al., *CDD: a Conserved Domain Database for protein classification.* Nucleic Acids Res, 2005. **33**(Database issue): p. D192-6.
106. 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.
107. Wilson, D.O., P. Johnson, and B.R. McCord, *Nonradiochemical DNase I footprinting by capillary electrophoresis.* Electrophoresis, 2001. **22**(10): p. 1979-86.
108. 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.
109. Andersson, S., et al., *Molecular genetics and pathophysiology of 17 beta-hydroxysteroid dehydrogenase 3 deficiency.* J Clin Endocrinol Metab, 1996. **81**(1): p. 130-6.
110. 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.
111. 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.
112. 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.
113. Crooks, G.E., et al., *WebLogo: a sequence logo generator.* Genome Res, 2004. **14**(6): p. 1188-90.
114. Zheng, L., U. Baumann, and J.L. Reymond, *An efficient one-step site-directed and site-saturation mutagenesis protocol.* Nucleic Acids Res, 2004. **32**(14): p. e115.
115. Studier, F.W., *Protein production by auto-induction in high density shaking cultures.* Protein Expr Purif, 2005. **41**(1): p. 207-34.
116. 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.
117. Moxley, R.A. and H.W. Jarrett, *Oligonucleotide trapping method for transcription factor purification systematic optimization using electrophoretic mobility shift assay.* J Chromatogr A, 2005. **1070**(1-2): p. 23-34.

118. Itou H, Yao M, Watanabe N, and Tanaka I, Crystal structure of the PH1932 protein, a unique archaeal ArsR type winged-HTH transcription factor from *Pyrococcus horikoshii* OT3. Proteins, 2008. 70(4): 1631-4
119. Busenlehner, L.S., M.A. Pennella, and D.P. Giedroc, *The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance.* FEMS Microbiol Rev, 2003. **27**(2-3): p. 131-43.
120. Hethke, C., et al., *A cell-free transcription system for the hyperthermophilic archaeon Pyrococcus furiosus.* Nucleic Acids Res, 1996. **24**(12): p. 2369-76.
121. Holm, L. and C. Sander, *Dali: a network tool for protein structure comparison.* Trends Biochem Sci, 1995. **20**(11): p. 478-80.
122. 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.
123. Garcia-Castellanos, R., et al., *Three-dimensional structure of MecI. Molecular basis for transcriptional regulation of staphylococcal methicillin resistance.* J Biol Chem, 2003. **278**(41): p. 39897-905.
124. Clark, K.L., et al., *Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5.* Nature, 1993. **364**(6436): p. 412-20.
125. Alekshun, M.N., et al., *The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 A resolution.* Nat Struct Biol, 2001. **8**(8): p. 710-4.
126. van der Oost, J., et al., *The ferredoxin-dependent conversion of glyceraldehyde-3-phosphate in the hyperthermophilic archaeon Pyrococcus furiosus represents a novel site of glycolytic regulation.* J Biol Chem, 1998. **273**(43): p. 28149-54.