

DISCOVERY OF NEW TRANSCRIPTION FACTORS IN PYROCOCCLUS FURIOSUS WITH DNA-AFFINITY PROTEIN CAPTURE AND MASS SPECTROMETRY

by

MEIYAO WANG

(Under the Direction of Robert A. Scott and I. Jonathan Amster)

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. First the method was tested by the protein capture and identification of LrpA, a known transcription factor. The native LrpA from *Pf* cell extract was successfully identified using the optimized protocol. Next, the *Pf* gene expression profile on the genome-wide scale has been described and compared between two different cell-growth conditions with maltose or peptides as the primary carbon source. Additional bioinformatics analysis identified a conserved palindromic sequence in the putative promoter region of several genes that are up-regulated by growth on maltose and all are involved in glycolysis. Based on these analyses, the DNA fragments upstream of the open reading frame of PF0272 and PF0464, which encode 4- α -

glucanotransferase and glyceraldehyde-3-phosphate ferredoxin oxidoreductase, were selected for DNA-affinity protein capture. A novel transcription regulatory protein PF0124 (PF0124p) was identified and its specific DNA binding ability at sites upstream of multiple genes was confirmed by the electrophoretic mobility shift assay. DNA footprinting showed that PF0124p binds to a region upstream of both PF0272 and PF0464, which contains the conserved palindromic motif. Furthermore, PF0124p was shown to repress both PF0272 and PF0464 gene transcription in cell-free transcription assays. These results suggest PF0124p is a transcriptional regulator recognizing a conserved palindromic motif and possibly regulating a group of genes in the glycolysis pathway in *Pyrococcus furiosus*.

INDEX WORDS: *Pyrococcus furiosus*, Archaea, Transcription regulation, Transcription factor, PF0124, TrmB, Mass spectrometry, DNA affinity, Environmental stress, Glycolysis

DISCOVERY OF NEW TRANSCRIPTION FACTORS IN PYROCOCCUS FURIOSUS WITH
DNA-AFFINITY PROTEIN CAPTURE AND MASS SPECTROMETRY

by

MEIYAO WANG

B.S., Jilin University, China, 2000

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

2006

© 2006

Meiyao Wang

All Rights Reserved

DISCOVERY OF NEW TRANSCRIPTION FACTORS IN PYROCOCCLUS FURIOSUS WITH
DNA-AFFINITY PROTEIN CAPTURE AND MASS SPECTROMETRY

by

MEIYAO WANG

Major Professors:	Robert A. Scott I. Jonathan Amster
Committee:	Marly K. Eidsness Michael W. W. Adams

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2006

DEDICATION

This work is dedicated to my beloved parents, for everything they have given me throughout my life. Also to my dear grandma, for the love and care she provided me in my childhood, which has become the most cherished memories in my life. 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, and to Dr. Amster, who inspired my interest in bioanalytical technology and taught me advanced mass spectrometry knowledge.

I want to also give special thanks to Dr. Eidsness and Dr. Adams who always encouraged me and helped me in face of any difficulties.

I am also very grateful to all the project collaborators including 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 from Dr. Michael Thomm's group at University Regensburg, Germany, who collaborated in the study of cell-free gene transcription, and people in Dr. Ying Xu's group in the department of Biochemistry and Molecular biology for bioinformatics analysis.

Thanks also to all my former and current fellows in Department of Chemistry for providing a good working atmosphere, especially Gina Lipscomb, my dear friend, who gave me great encouragement and help at work and to people in Dr. Amster's laboratory, Richard Wong and Chunyan Li who gave me the necessary time and help with mass spectrometry.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
CHAPTER	
1 INTRODUCTION	1
1.1. Gene Transcription and Transcriptional Regulation	1
1.2. Discovery of Transcriptional Regulators in <i>Pyrococcus furiosus</i>	10
1.3. Proteomics Mass Spectrometry	14
2 DEVELOPMENT OF DNA-AFFINITY PROTEIN CAPTURE TECHNIQUE FOR THE STRES APPROACH TO DISCOVERY OF TRANSCRIPTION FACTORS IN PYROCOCCUS FURIOSUS	19
2.1. Materials and Methods	19
2.2. Optimization of DNA-affinity Protein Capture Method	29
2.3. DNA-affinity Protein Capture Protocol	66
2.4. Capture of Native LrpA	70
3 IDENTIFICATION OF UOR BINDING PROTEINS WITH MALDI PSD MS	73
3.1. Introduction	73
3.2. Materials and Methods	77
3.3. Results and Conclusions	82
3.4. Discussion	89

4	CHARACTERIZATION OF NEW REGULATORY TRANSCRIPTION FACTOR, PF0124.....	96
4.1.	Introduction	96
4.2.	Materials and Methods	97
4.3.	Results and Conclusions.....	106
4.4.	Discussion	123
	REFERENCES	131
	APPENDICES	144
A	INDEX OF ABBREVIATIONS.....	144

CHAPTER 1

INTRODUCTION

1.1. Gene Transcription and Transcriptional Regulation

1.1.1. Transcription

DNA carries the genetic information for all the proteins essential to compose cellular structures and accomplish the physiological functions in cells. DNA serves as a template for synthesis of RNA catalyzed by RNA polymerase (RNAP) through the process called transcription. One transcription product, messenger RNA, supplies the templates for synthesis of proteins through translation. Appropriate transcription requires the participation of promoter (DNA sequence enabling a gene to be transcribed), RNAP, and multiple transcription factors (TFs) (proteins), as well as the proper order of all involved events. Gene transcription can be achieved in cell-free conditions by using the minimum necessary components called the basal transcriptional apparatus, including several conserved promoter elements, the general cognitive TFs and RNAP. The basal transcriptional apparatus is not promoter specific, and is capable of transcribing any gene without any cellular physiological requirements. However, in living cells, most genes are involved in cellular networks and must cooperate with other cellular reactions in responding to various conditions. Therefore, the basal transcriptional apparatus is not enough, and many regulatory proteins are involved in transcriptional regulation to either activate or repress expression of genes in response to the physiological requirements and environment of cells. The transcribed RNA and expressed proteins in a cell generally determine the nature of the cell. Therefore, the process of transcription plays a critical role in controlling the nature, identity,

and status of the organism. Current biomedical research indicates that transcriptional defects are associated with most kinds of cancers and many other diseases such as diabetes, hemophilia and developmental disorders (Latchman, 1996; Villard, 2004).

Clearly, it is important to investigate transcription and its regulation, the knowledge of which should facilitate the understanding of fundamental aspects of the nature of living organisms.

1.1.2. Differences in Transcription between Eukaryotes 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) has been summarized and discussed in several review articles (Asturias, 2004; Sims *et al.*, 2004; Woychik *et al.*, 2002). 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 TFIIIE is to recruit TFIIH to the promoter and stimulate TFIIH phosphorylating activity.

Compared to that of the eukaryotes, the bacterial basal transcription machinery is much simpler and comprised of fewer components (Gourse *et al.*, 2000; Soppa, 1999). 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 (Gross *et al.*, 1998). 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 (deHaseth *et al.*, 1998).

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. In addition to the distinctive features of the basal transcriptional apparatus, transcriptional regulation mechanisms in eukaryotes and bacteria are fundamentally different due to the difference in genomic DNA packing structure (Struhl, 1999). 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. Archaea were first documented in 1977 by their own distinctive genomic character (Woese *et al.*, 1977). These prokaryotes form the third domain of life on earth other than the eukaryotes and bacteria (Woese *et al.*, 1990). An interesting feature of archaea is that they share different characteristics with eukaryotes and bacteria. The morphology, behavior, and metabolic machinery of archaea are similar to those of bacteria, but the information processing systems (e.g., replication, transcription, and translation) of archaea are more eukaryotic-like (Soppa, 2001).

1.1.3. Transcription in Archaea

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 (Best *et al.*, 2001). The rest of the archaeal basal transcription machinery is also eukaryotic-like and can be considered a simplified version of eukaryotic RNAPII transcription machinery (Bell *et al.*, 1998a, 2001; Geiduschek *et al.*, 2005; Ouhammouch, 2004; Reeve, 2003). 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 (ca. –33 to –34), and an initiator element around the transcription start site (**Figure. 1.1**). But the eukaryotic DPE has not been found in archaea. Archaeal basal transcription factors include TBP and TFB, which are homologous to eukaryotic TBP and TFIIB, respectively (Hausner *et al.*, 1996). 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. Therefore the absence of TFIIH and TFIIIE homologs in archaea seems reasonable (Bell *et al.*, 1998b; Soppa, 1999). Archaeal gene transcription initiation can be activated in a cell-free system with only TBP, TFB, and RNAP (Hausner *et al.*, 1996), although some archaeal species contain a TFIIIE α subunit homolog that is suggested to facilitate TBP-promoter interactions under unfavorable conditions (Hanzelka *et al.*, 2001). 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 (Hausner *et al.*, 1996), the recognition of BRE by TFB defining the orientation of the complex and the direction of transcription (Bell *et al.*, 1999b; Littlefield *et al.*, 1999; Thomm, 1996). 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 (Kyrpides *et al.*, 1999). They predicted 280 transcription factors or transcription-associated proteins in the four archaeal genomes, 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.

Eukaryotic



Archaeal



Bacterial

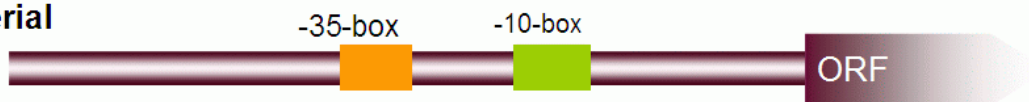


Figure 1.1 Conserved elements of promoter in three domains of life. Archaea and eukaryotes are more closely related in terms of promoter features. Both contain a TATA box element recognized by TBP, a TF(II)B recognition element (BRE) upstream of the TATA box, and a conserved transcription initiator element (INR). 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.

The most widely distributed archaeal rTFs with bacterial homologs are from the Lrp (Leucine-responsive Regulatory Protein)/AsnC (Asparagine synthase C) family (Brinkman *et al.*, 2003). In bacteria, Lrp can either stimulate or repress transcription of various genes, which are involved in amino acid synthesis and metabolism. In archaea, transcription factors of the bacterial Lrp/AsnC family have been found in various species, including LrpA from *Pyrococcus furiosus* (Brinkman *et al.*, 2000; Dahlke *et al.*, 2002), the negative autoregulator Lrs14 and lysine-modulated activator LysM from *Sulfolobus solfataricus* (Bell *et al.*, 2000; Brinkman *et al.*, 2002; Brinkman *et al.*, 2003), Ptr2 from *Methanococcus jannaschii* (Ouhammouch *et al.*, 2001; Ouhammouch *et al.*, 2003), and the recently identified transcriptional regulator SsLrpB in *S. solfataricus* (Peeters *et al.*, 2004). Another bacteria-like regulator, MDR1, from *Archaeoglobus fulgidus*, a homolog of bacterial DtxR (Diphtheria toxin repressor), represses transcription of its own gene in a metal-dependent manner (Bell *et al.*, 1999a). One eukaryotic-type transcriptional regulator found in archaea is GvpE, that positively regulates the gene encoding gas vesicle structural protein from *Halobacterium salinarum* and *Haloferax mediterranei* (Zimmermann *et al.*, 2003). Some rTFs are found only conserved among archaeal species. Phr is an inhibitor of heat shock response in *P. furiosus* (Vierke *et al.*, 2003) and it specifically represses the transcription of its own gene and two heat shock genes hsp20 and AAA+ ATPase. TrmB (transcriptional regulator of *mal* operon), from *Thermococcus litoralis* and *P. furiosus*, is a representative of a TrmB family, negatively regulating sugar transport cluster genes (Lee *et al.*, 2003; Lee *et al.*, 2005). NrpR, an archaeal regulator from *Methanococcus maripaludis*, represses nitrogen fixation genes, the *nif* operon, in response to nitrogen availability. 2-oxoglutarate acts as an intracellular indicator of nitrogen deficiency and induces *nif* genes via binding NrpR (Lie *et*

al., 2003; Lie *et al.*, 2005). The known archaeal rTFs are summarized in **Table 1.1**, and recent reviews have described them in detail (Bell, 2005; Geiduschek *et al.*, 2005; Ouhammouch, 2004).

Studies of archaeal repressors have revealed at least two types of negative transcriptional regulation pathways in archaea. One is mediated by an rTF binding a region overlapping the TATA/BRE promoter elements and blocking TBP/TFB interaction with the promoter. The rTFs known to utilize this pathway are TrmB (Lee *et al.*, 2003) and Lrs14 (Napoli *et al.*, 1999). The other negative regulation pathway involves regulatory repressors binding the promoter region close to the transcription start site and preventing recruitment of RNAP. This class of rTF includes LrpA (Dahlke *et al.*, 2002), Phr (Vierke *et al.*, 2003), and MDR1 (Bell *et al.*, 1999a). The stimulation of transcription by prokaryotic transcriptional activators is assumed to occur by mediating the recruitment of RNAP and basal transcription factors or by stabilizing the transcription PIC (Busby *et al.*, 1994). The mechanism of archaeal promoter-specific activation is even less well known than that of repression, though some regulatory activators have been identified. The only instance of transcriptional activation that has been demonstrated *in vitro* is by the Lrp family protein *Mja* Ptr2, which activates initiation of transcription by its conjugate RNAP at high temperature and recruits TBP to weak T/A-box-containing promoters (Ouhammouch *et al.*, 2005).

In recent years, genome-wide microarray assays have been used in archaea to study differential expression of genes in response to changing conditions of growth. Several stimulons (i.e., groups of transcription units that are differentially expressed in response to changing conditions of growth) have been identified (Kaur *et al.*, 2006; Rohlin *et al.*, 2005; Schut *et al.*, 2003; Shockley *et al.*, 2003; Weinberg *et al.*, 2005). However, few studies have been performed

Table 1.1 Summary of identified transcription factors in archaea.

Name	Species	Function	Ligand	Regulation pathway	Homologous type
MDR1	<i>A. fulgidus</i>	Repression	Metal ions	Abrogating RNAP recruitment	Bacterial DtxR family
LrpA	<i>P. furiosus</i>	Repression	Unknown	Abrogating RNAP recruitment	Bacterial Lrp/AsnC family
Phr	<i>P. furiosus</i>	Repression	Unknown	Abrogating RNAP recruitment	Conserved among archaea
Lrs14	<i>S. solfataricus</i>	Repression	Unknown	Preventing TBP/TFB recognition	Bacterial Lrp/AsnC family
TrmB	<i>T. litoralis</i> and <i>P. furiosus</i>	Repression	Multiple sugars	Prevent TBP/TFB recognition	Archaeal TrmB family
NrpR	<i>M. maripaludis</i>	Repression	2-oxoglutarate	Probably blocking components of transcription PIC	Archaea
GvpE	<i>H. salinarum</i> and <i>H. mediterranei</i>	Activation	Unknown	Unknown	Eukaryotic leucine-zipper family
LysM	<i>S. solfataricus</i>	Activation	Lysine	Unknown	Bacterial Lrp/AsnC family
Bat	<i>Halobacterium</i> sp. NRC-1	Activation	Unknown	Unknown	Bacterial AraC family
Ptr2	<i>M. jannaschii</i>	Activation	Unknown	Facilitating TBP recruitment	Bacterial Lrp/AsnC family
ssLrpB	<i>S. solfataricus</i>	Unknown	Unknown	Unknown	Bacterial Lrp family

to identify regulon members (i.e., transcription units that are controlled by a single regulatory protein) to elucidate co-regulation of genes.

Obviously, the understanding of transcriptional regulation in archaea is at an elementary level. Many questions remain open: How does the bacterial-like regulatory system cooperate with eukaryotic-like transcriptional machinery? How does archaeal transcriptional regulation manage to adjust to a wide range of extreme living conditions in response to environment stress? In particular, how are genes co-regulated by common regulators? The answers to such questions require advanced investigation and further exploration.

1.2. Discovery of Transcriptional Regulators in *Pyrococcus furiosus*

1.2.1. The Goal of the Present Study

We have developed an approach called STRES (Survey of Transcriptional Response to Environmental Stress) by integrating various techniques including microarray expression profiling, bioinformatics analysis, DNA-affinity protein capture, and different biophysical methods to identify novel transcription factors and investigate transcriptional regulatory mechanisms. As a part of this integrated investigation, the present study focused on discovery of regulatory transcription factor(s) in the archaeon *Pyrococcus furiosus* (*Pf*) responding to the carbon source (carbohydrates or peptides) in the cell growth condition.

P. furiosus is one of the best studied hyperthermophilic archaea. The *Pf* gene expression profile responding to different carbon source (maltose vs. peptides) has been described in a genome-wide microarray study (Schut *et al.*, 2003). Many genes showed differential expression levels between the two conditions, but microarray analyses alone cannot identify the rTFs or the transcriptional regulatory mechanism. Identification of the rTFs and investigation of the mechanism will be necessary and valuable to build a genome-wide regulatory network. As a part

of our STRES approach, we have developed the method of DNA-affinity protein capture for identification of rTFs using putative gene promoter DNA. Modern proteomics mass spectrometry contributes to this approach.

In general, the STRES approach will contribute to constructing the regulatory network of *Pf*, describing how transcriptional regulation mediates the corresponding adaptive response to environmental stress signals. In addition, the STRES technology could be suitable for targeting protein identification via affinity-capture from other organisms, organs, or tissues involved in a selected environmental stress or correlated with a specific disease state.

1.2.2. *Pyrococcus furiosus*

Pyrococcus furiosus, originally isolated from a shallow marine volcanic vent, is an anaerobic, heterotrophic, hyperthermophilic euryarchaeon, living optimally around 100°C (Fiala *et al.*, 1986). The *Pf* genome sequence of approximately 1.9 Mb in size encoding more than 2000 genes, has been completed (Poole *et al.*, 2005; Robb *et al.*, 2001). *Pf* ferments either peptides or carbohydrates with the production of organic acids, H₂ and CO₂. It can also utilize elemental sulfur (S⁰) as energy source, although the organism grows well in the absence of S⁰ (Fiala *et al.*, 1986). Several stimulons in *Pf* were investigated responding to various environmental stresses using microarray technology (Schut *et al.*, 2003; Shockley *et al.*, 2003). Sugar metabolism has been studied extensively in this organism. *Pf* contains standard archaeal basal transcriptional machinery in terms of both the transcriptional apparatus and the architecture of gene promoters, which as described earlier, resembles that of the eukaryotic RNAPII transcription system.

1.2.3. Transcriptional Regulators of *P. furiosus*

Although *Pf* is one of the best studied archaea, only a few transcriptional regulatory proteins are known to date. The two best known are LrpA and TrmB.

1.2.3.1. LrpA

LrpA is a representative archaeal rTF, homologous to the bacterial Lrp/AsnC transcriptional regulator family. LrpA is a helix-turn-helix DNA-binding protein with a molecular mass of 15.9 kDa (Brinkman *et al.*, 2000; Dahlke *et al.*, 2002). Among the Lrp family members, the best known protein is the bacterial Lrp protein in *Escherichia coli* (Calvo *et al.*, 1994; Newman *et al.*, 1995). Lrp is a global transcriptional regulator that can either activate or repress different genes. Most of the genes are related to amino acid synthesis and metabolism, and the regulation is in response to the effector leucine. Lrp is also a negative autoregulator of its own gene, but this auto-repression is independent of leucine. The paralog protein *E. coli* AsnC is an asparagine-dependent specific transcriptional activator of asparagine synthesis. Similar to Lrp, the negative auto-regulation of AsnC is effector-independent (de Wind *et al.*, 1985; Kolling *et al.*, 1985). *Pf* LrpA also appears to be a negative auto-regulator of its own gene independent of any effectors. The transcription start site of the *lrpA* gene is located at the position -14 bp upstream of the translation start site. The BRE/TATA region is at -35 to -23 bp upstream of the *lrpA* transcription start site (+1). Previous studies illustrated that LrpA binds promoter DNA from -22 to +24, which overlaps with the RNA polymerase binding region. Therefore, repression of the *lrpA* gene is via blocking the RNA polymerase binding site to abrogate recruitment of RNAP and prevent the completion of the pre-initiation complex (Brinkman *et al.*, 2002).

1.2.3.2. TrmB

TrmB is a substrate-specific transcriptional repressor for a maltose/trehalose ABC transport gene cluster (*malE*, *malF*, *malG*, *treT*, *trmB*, and *malk*) in *Thermococcus litoralis* (Lee *et al.*, 2003). This gene cluster repression by TrmB can be overcome by maltose/trehalose. As a result of lateral gene transfer, the identical *mal* operon including regulator TrmB is also

contained in *Pf* (TM operon) (DiRuggiero *et al.*, 2000). In addition, a secondary sugar transport operon MD encoding maltodextrin transport genes has been found in *Pf* (Koning *et al.*, 2002). The MD operon showed specific substrate recognition of maltodextrin. Recent studies indicate that TrmB can repress this gene cluster. However, neither maltose nor trehalose can induce the MD operon. The MD operon repression can be overcome by maltodextrin or sucrose (Lee *et al.*, 2005). Therefore, TrmB was suggested to be a bifunctional transcriptional regulator, binding two different promoters and responsive to multiple sugars. It is known that TrmB binds the promoter at two regions: one is an inverted repeat sequence upstream of the *mal* operon TATA/BRE; the other overlaps the BRE/TATA-box (Lee *et al.*, 2003). The core recognition site of TrmB is the sequence TACTnnnAGTA, a 4-3-4 palindrome (Lee *et al.*, 2005). Thus, the transcriptional repression of sugar ABC transporter genes by TrmB is mediated through preventing the recognition of the promoter by TBP.

1.2.4. The Targeted *Pf* Genes in the Present Study

From the results of the microarray assay for *Pf* gene differential expression between two conditions with maltose or peptides as the primary carbon source (Schut *et al.*, 2003), a group of genes involved in glycolytic metabolism were highly up-regulated in the presence of maltose. PF0272, α -amylase (*amy*) [or 4- α -glucanotransferase (*glt*) from a recent investigation (Lee *et al.*, 2006)] was up-regulated 26x; PF0464, glyceraldehyde-3-phosphate, ferredoxin oxidoreductase, GAPOR (*gor*), was up-regulated 5.7x; PF1959, phosphoglycerate mutase (*pgm*), was up-regulated 3.2x; and PF1920, triosephosphate isomerase (*tpi*), was up-regulated 4.7x. Whether members in this maltose-induced stimulon are regulated by common transcription factors is not known.

Moreover, a bioinformatics study using CUBIC (Olman *et al.*, 2003) to find repeating sequences in the promoter regions of the genes in this stimulon suggested the existence of a conserved palindromic motif, TCACnnnnnGTGA (n is any nucleotide), in several of the putative promoters (**Figure 1.2**, unpublished data from Dr. Ying Xu's group). There are seven genes from the maltose-induced glycolysis gene group containing this motif, which suggested a potential binding site for a common regulatory protein. Besides the genes mentioned above, this group includes PF0196, glucose-6-phosphate-isomerase (*pgi*); PF0312, ADP-specific glucokinase (*glk*); and PF1784, ADP-specific phosphofructokinase (*pfk*).

Based on the combination of both microarray assay and bioinformatics analysis, we chose to focus on the identification of a possible transcriptional co-regulatory protein(s) for the above genes from the maltose-induced glycolytic group such as PF0272 and PF0464. Modern proteomics mass spectrometry has been applied to facilitate the study.

1.3. Proteomics Mass Spectrometry

Mass spectrometry (MS) is a powerful analytical technique to identify or characterize compounds by measuring the mass-to-charge ratio of molecules. A mass spectrometer consists of three major components: an ion source to bring the charged molecule to the gas phase, an analyzer to sort the molecules according to their mass-to-charge ratio, and a detector to convert the ion flux to electronic signal. The difficulties of applying mass spectrometry to macromolecules such as proteins and peptides have been overcome by the discovery and development of two soft ionization methods during the last two decades, electrospray ionization (ESI) (Fenn *et al.*, 1989) and matrix-assisted laser desorption/ionization (MALDI) (Tanaka *et al.*, 1988), both of which were recognized by the 2002 Nobel prize in chemistry (Fenn, 2003; Tanaka, 2003). By combining ESI or MALDI with different MS analyzers, a variety of MS



Figure 1.2 Conserved palindromic sequence contained by maltose-induced stimulon involved in *Pf* glycolysis. The level of up-regulation (maltose over peptides) is indicated below each gene. Marked on the sequences are: the potential TATA element (bold), conserved palindromic sequence (red), predicted ribosome binding site (underlined), translation start codon (italic) and predicted transcription initiation site of some genes (grey shade). The logo (generated from <http://weblogo.berkeley.edu/>) results from sequence alignment of palindromic nucleotides from multiple genes. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each nucleic acid at that position. A sequence logo generally provides a richer and more precise description of, for example, a binding site, than would a consensus sequence.

instruments were designed and the choice depends on the application. Comparison of different mass spectrometers is the subject of recent reviews (Domon *et al.*, 2006; Lane, 2005).

Taking advantage of the development of MS instrumentation and genome sequencing, MS has replaced traditional Edman N-terminal degradation as a major method for protein sequencing. Protein identification by MS is commonly achieved by analyzing the peptides derived from protease digestion. There are two major ways to obtain the sequence identity of a protein from those peptides. In the first one, called peptide mass fingerprinting (PMF) (Cottrell, 1994), a peptide mixture from protease digestion of a protein is measured directly in MS and the resulting list of masses is compared to a theoretical digestion of all the proteins in a database. A statistically significant match will confirm the sequence identity of the protein. The second method of protein identification is to use tandem MS (MS/MS) to obtain sequence information of individual peptides. In a MS/MS measurement, a precursor ion (peptide) is selected from the first MS analyzer and fragmented in a collision cell. The masses of the product ions are then measured by the second analyzer and can be used to deduce the sequence of the parent peptide manually or by database searching.

Obviously both PMF and MS/MS database searching rely heavily on the availability of protein sequence, which can be deduced from genome sequence. There are 395 completed and 1602 ongoing genome projects as of July 3, 2006, as monitored by Genomes OnLine Database (Liolios *et al.*, 2006) (<http://www.genomesonline.org>). In the field of proteomics, that addresses the entire set or a subset of proteins in a living system, MS plays a central role in protein identification, not to mention its capability of quantitative analysis developed lately (Lane, 2005).

As proteomics samples can be extremely complicated, for example in some studies of the entire protein content from cells or tissues, powerful separation techniques are usually required before MS analysis. Two distinct approaches have been used widely, including separation of intact proteins followed by MS identification and the so-called shotgun proteomics method. Both methods have been broadly used, and are described and compared in several reviews (Domon *et al.*, 2006; Lane, 2005). In the former method, the intact protein separation is usually achieved by protein 2-dimensional electrophoresis (2DE), which is later followed by protease digestion of a single protein (2D gel spot) before MS analysis. The shotgun proteomics approach involves bulk digestion of a protein mixture, followed by separation of the resulting peptide mixture in one- or two-dimensional liquid chromatography (LC) coupled with MS/MS sequencing (Link *et al.*, 1999). The 2DE-based method is restricted by problems such as limited dynamic range, insufficient resolving power and relatively low throughput. Yet, 2DE has been used widely in the past and is still suitable for studies with samples of low complexity or with specifically targeted proteins. More recently, this method was adapted by replacing 2DE with capillary electrophoresis and capillary electrochromatography (Hu *et al.*, 2004; Shadpour *et al.*, 2006), which led to successful applications (Kasicka, 2006). On the other side, the shotgun proteomics approach offers higher proteomic resolution, experimental simplicity, and ease of automation. Currently the limitations of the method come from the data redundancy and the bioinformatics challenge to analyze the enormous amount of MS/MS spectra (Nesvizhskii *et al.*, 2004).

In contrast to profiling the entire proteome, many applications of proteomics mass spectrometry in the life sciences target a subset of proteins, such as proteins with post-translational modifications (PTMs) (e.g., phosphorylation, glycosylation, lipidation, etc.), proteins in subcellular locations (such as membrane proteins, nuclear proteins) or any subset of

proteins to address specific biological problems. In these cases, some fractionation/enrichment steps are normally introduced before the proteomics MS analysis, for example, immobilized metal-affinity chromatography (IMAC) for phosphorylated proteins (Ficarro *et al.*, 2002), lectin capture for glycoproteins (Zhang *et al.*, 2003), affinity-tag enrichment for ubiquitinated proteins (Peng *et al.*, 2003), etc. The affinity purification decreases the complexity of the samples greatly and allows increased resolution of separation and improved sensitivity of the following MS analysis (Lee *et al.*, 2004). In some cases, it is sufficient to separate the proteins using 1DE, avoiding the pitfalls of running 2DE (Guo *et al.*, 2002; Lee *et al.*, 2004; Rappsilber *et al.*, 2000). As will be introduced in the following chapters, the method of DNA-affinity protein capture developed in our group is a type of "affinity fractionation" capable of reducing the sample complexity for analysis with 1DE followed by mass spectrometry.

CHAPTER 2

DEVELOPMENT OF DNA-AFFINITY PROTEIN CAPTURE TECHNIQUE FOR THE STRES APPROACH TO DISCOVERY OF TRANSCRIPTION FACTORS IN PYROCOCCLUS FURIOSUS

A method for rTF discovery was developed using *Pyrococcus furiosus* as a model archaeon. The key step in this method consists of DNA-affinity protein capture using magnetic-bead immobilized DNA containing promoter and rTF binding site. To demonstrate this technique and to optimize protein capture conditions, a known *Pf* rTF was selected as a positive control.

The protein LrpA is known as an rTF for multiple genes in *Pyrococcus furiosus*. It also acts as a negative auto-regulator of its own (*lrpA*) gene (Kyrpides *et al.*, 1995). LrpA binding to promoter and the mechanism of regulation are well understood (Brinkman *et al.*, 2000; Brinkman *et al.*, 2003). Therefore the promoter DNA of *Pf lrpA* and the gene transcription regulator protein LrpA were a pair of good standards for method development. The dsDNA was prepared with Polymerase Chain Reaction (PCR) and recombinant protein LrpA was overexpressed and purified in our study.

2.1. Materials and Methods

2.1.1. Materials

Buffers

2x Binding & Washing (B&W) buffer: 1 mM EDTA, 2.0 M NaCl, 10 mM Tris, adjust the pH to ~7.0 using HCl.

5x Incubation buffer, stock: 5 mM EDTA, 0.5 M KCl, 20% glycerol, 0.5 % Triton X100, 1 mM DTT, 250 mM Tris, adjust pH to ~7.0 using HCl. (The final concentration used in DNA-protein incubation reaction was 1x Incubation buffer.)

4x Gel-loading buffer (SDS-PAGE sample mixing buffer): 50% glycerol (5 mL in 10 mL), 8% Sodium Dodecyl Sulfate (SDS) (0.8 g in 10 mL), 0.1% Bromophenol blue (10 mg in 10 mL), 160 mM Tris-HCl, pH~6.8 (3.2 mL 0.5 M Tris-HCl, pH 6.8 in 10 mL). Bring volume up to 10 mL with water. Freeze 4x gel-loading buffer in 0.5 mL aliquots at -20 °C. Add 0.1 mL β -mercaptoethanol to 0.5 mL aliquot before use.

DNA Oligonucleotides

All unlabeled or biotinylated DNA primers were synthesized by Integrated DNA Technologies, Inc.

DNA Competitors

Sodium heparin solution was made by dissolving 0.1 M sodium heparin in 1x incubation buffer. Poly-dIdC (Amersham) was dissolved in 50 mM Tris-HCl , pH 7.0, 1 mM EDTA, 0.1 M NaCl, at 500 ng/mL. Herring sperm, salmon sperm, and calf thymus DNA (Sigma) were dissolved in 10 mM Tris-HCl, pH 7.0, 1 mM EDTA at a concentration of 6.4 mg/mL.

Pf Cell Extract Preparation

15 L of *Pf* cell culture were grown in anaerobic condition at 95 °C in the presence of S^0 with peptides as the primary carbon source following a published protocol (Schut *et al.*, 2003). Membrane proteins were removed from cell lysis by ultra-high speed centrifugation at 100,000xg for at least 1 h. The concentration of protein in cell extract was measured with DC Protein Assay Kit II (BioRad).

Other Reagents and Materials

Dynabeads and MPC magnet were from DYNAL Inc. *Taq* DNA polymerase was from Fisher or Qiagen and dNTP mixture was from Sigma. PCR Purification kit, Gel Purification kit, and Plasmid Midi kit were from Qiagen. Porcine trypsin powder and isopropyl- β -D-thiogalactoside (IPTG) were from Promega. Sodium heparin, Coomassie Brilliant Blue (CBB), Bromophenol blue, silver nitrate, ammonium bicarbonate, dithiothreitol (DTT), Tris base, and Luria Broth powder were from Fisher Scientific Inc. Kanamycin and α -cyano-4-hydroxycinnamic acid was from Sigma. Hitrap Chelating HP column and Hoef 3275 fluorescent dye was from Amersham Biosciences. Turbo DNase was from Ambion. Gelcode Blue staining solution was obtained from Pierce Biotechnology Inc.

2.1.2. Overexpression and Purification of Recombinant *Pf* LrpA

For protein capture optimization, recombinant *Pf* LrpA with (His)₆ at the N-terminus was used. The gene for (His)₆-LrpA, contained in the modified plasmid PET24d, that is kanamycin-resistant, was provided by Dr. MWW Adams' laboratory. The plasmid was transformed in competent cells of *E. coli* strain BL21 GOLD (DE3) and grown in 1 L 2.5% LB media (25 g LB in 1 L water) containing 50 μ g/mL kanamycin. IPTG was added to a final concentration of 0.84 mM at OD 0.6. The cells continued to grow at 37 °C for 2 more hours and were harvested by centrifugation at 6115xg at 4 °C for 20 min with a Beckman JS-5.3 rotor. The cell pellet was resuspended in 15 mL Hitrap column binding buffer I (0.02 M sodium phosphate, 0.5 M NaCl, pH 7.4) containing 0.1 mM phenylmethylsulfonyl fluoride (PMSF). Cell lysis was performed by sonicating at 20% power intensity for 15 sec of every 25 sec for 3 min. The cell extract supernatant was collected after centrifugation at 48,384xg for 60 min with a Beckman JA-25.5 rotor, and loaded on Ni²⁺ charged Hitrap Chelating HP column (Amersham Biosciences)

equilibrated with column binding buffer 1. The column was washed with the same buffer, and protein was eluted with a linear gradient of imidazole (0-0.5 M) on FPLC. The elution fractions containing (His)₆-tagged protein were pooled and dialyzed with 20 mM Tris-HCl, pH 7.2, 0.5 M KCl. The final concentration of protein was about 0.5 mg/mL and protein was kept at -80 °C.

2.1.3. Preparation of Promoter dsDNA (UOR) by PCR

2.1.3.1. *lrpA* Promoter DNA (UOR)

Design of DNA Probe (UOR)

In our STRES approach, we assume that rTF binding sites lie near gene promoters. Archaeal promoters share characteristics of eukaryotic class II promoters, including TATA elements about 26 bp upstream of transcription start sites. For most genes in *Pf*, we know only the *translation* start, so to be confident that the DNA we use in our protein capture method contains rTF binding sites, we use a few hundred bp upstream of the translation start, referring to this DNA as UOR (Upstream of ORF (open reading frame) Region).

For example, we used 528 bp (-305 to +223, relative to ORF) as the UOR for gene *lrpA* (**Figure 2.1**). For comparison, a shorter *lrpA* UOR of 300 bp (from -200 to +100) was also amplified by PCR. Both UORs contain the TATA element, the BRE, and the LrpA binding site (Brinkman *et al.*, 2000).

For DNA immobilization on streptavidin-coated Dynabeads, the 5' end of the forward strand of the UOR was biotinylated. The UOR was cloned and amplified by PCR.

Gradient PCR

The PCR reaction conditions for each pair of primers are different and usually need to be optimized before the specific conditions are set. This was achieved by doing gradient PCR varying several reaction parameters. The most variable conditions are annealing temperature and

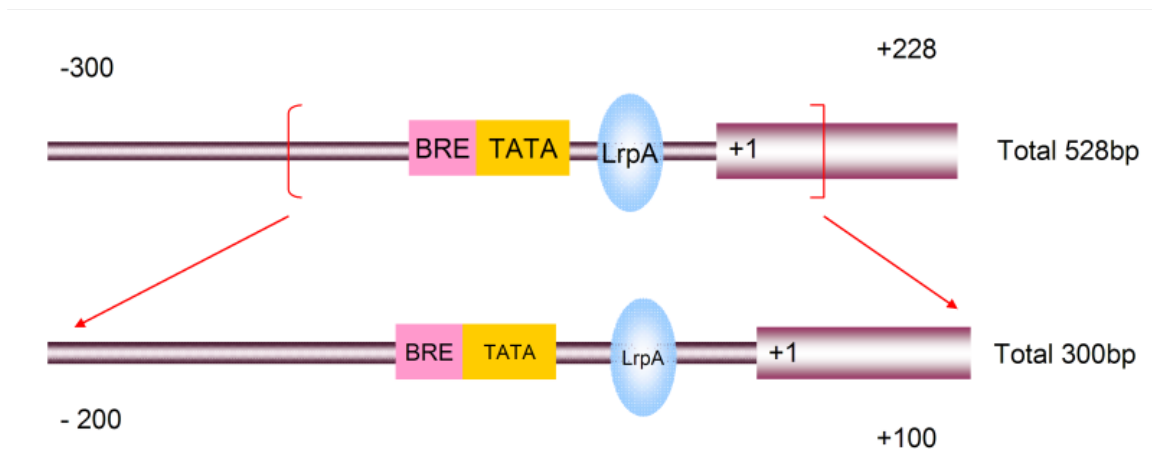


Figure 2.1 Schematic illustration of UOR of *lrpA*. The longer UOR (top) was 528 bp (-300 to +228, numbering relative to translation start site). The shorter UOR (bottom) was 300 bp (-200 to +100). Both UORs contain core elements of *lrpA* promoter including BRE, TATA element, and LrpA binding site.

concentration of Mg^{2+} , which might influence the oligonucleotide behavior and enzyme activity, affecting product purity and yield.

For example, in the gradient PCR for the 300-bp *lrpA* UOR, annealing temperatures were tested from 49 - 55 °C and Mg^{2+} concentrations were tested from 1-11 mM. It was found that the optimum PCR conditions were annealing at 53 °C with 2.6 mM $MgCl_2$ based on best yield and purity of the product DNA.

PCR for 300-bp *lrpA* UOR

The dsDNA was amplified by PCR from *Pf* genomic DNA. The forward primer sequence for *lrpA* UOR was 5'-biotin-TCTACAACATAGCAAAGGAG-3'; the reverse primer was 5'-CAGTCTCACTAATGCCTAAT-3'. The PCR reaction solution included 200 μ M dNTP mixture, 0.5 μ M of each primer, 10 ng genomic DNA as template, 1x PCR buffer, 2.6 mM $MgCl_2$, and 2.5 units *Taq* DNA polymerase (Qiagen) per 100 μ L reaction solution. The reaction solution was heated to 94 °C for 3 min, followed by 35 cycles of 1 min at 94 °C (denaturation), 1 min at 53 °C (annealing), and 1 min at 72 °C (extension) per cycle, and ended with final extension of 10 min at 72 °C.

PCR for 528-bp *lrpA* UOR

dsDNA was amplified from *Pf* genomic DNA with PCR, by *Taq* polymerase. The forward primer was 5'-biotin-CAGTCAGCTACTTCGAGTGG-3' and the reverse primer sequence was 5'-CAGGTTTTGTGTCAACTCCAG-3'. Gradient PCR was applied varying annealing temperature from 51-57 °C (at 2° intervals) at a constant $MgCl_2$ concentration of 1.5 mM. PCR was repeated for 35 cycles as described above. The PCR product was checked by DNA electrophoresis. From the gel, the optimum temperature was 55 °C considering both purity and yield.

2.1.3.2. Preparation of other dsDNA with PCR

PF0272 UOR

Control DNAs were used to confirm specificity of DNA binding by LrpA. The first control chosen was the UOR of a gene (PF0272) that is not known to be regulated by LrpA. We chose sequence from –223 to +79 relative to this ORF. The forward primer was 5'-biotin-CTCACTTGATTCTTTTTC-3'; the reverse primer was 5'-CCTCCTCAAACACCCATCCAAA-3'. PCR was performed using *Taq* polymerase with an annealing temperature of 53 °C and 1.5 mM MgCl₂.

Fragments of UOR of *lrpA*

In an experiment designed to confirm that LrpA binds only to its recognized DNA binding site in the *lrpA* UOR, we designed fragments of the UOR to also test in protein capture. These are described in detail later. For fragment F1R1, the forward (F1) primer sequence was 5'-biotin-CAGTCAGCTACTTCGAGTGG-3'; the reverse (R1) primer sequence was 5'-AACTCTCTGGACTGCAACTA-3'. For fragment F2R2, the forward (F2) primer sequence was 5'-biotin-TATCAAGCAATGCTTGACCG-3'; the reverse (R2) primer sequence was 5'-TCTTTTCCAATATTTCAG-3'. For fragment F3R3, the forward (F3) primer sequence was 5'-TGCGAGAACGCCCTTCACTG-3'; the reverse (R3) primer sequence was 5'-biotin-CAGGTTTTGTGTCAACTCCAG-3'. The PCR annealing temperature for all fragments was 53 °C and the MgCl₂ concentration was 1.5 mM.

2.1.3.3. dsDNA Purification

PCR products were purified by agarose gel extract with QIAquick Gel Extraction Kit for two purposes: (1) to remove extra biotinylated primers and side products; (2) to prepare pure DNA as supply for template DNA in subsequent PCR, as preparing *Pf* genomic DNA is time-

and labor-intensive. QIAquick PCR Purification Kit was also used to purify PCR product and has better yield, but it is not as good for purification, so it was used only when the PCR product was of high purity.

2.1.3.4. 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.

2.1.4. Electrophoresis and Gel Staining

2.1.4.1. 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.

2.1.4.2. DNA Electrophoresis

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.

2.1.5. In-gel Digestion and Mass Spectrometry

2.1.5.1. In-gel Digestion

Protein in-gel digestion was performed according to standard procedures (Gharahdaghi *et al.*, 1999; Shevchenko *et al.*, 1996). The gel bands were cut into 1-2 mm² particles and hydrated with 50 µL 25 mM NH₄HCO₃ for 10 min at room temperature. The gel spots were then dehydrated in 80 µL 50% acetonitrile in 25 mM NH₄HCO₃ for 15 min. The hydration and dehydration steps were repeated two more times. The gel spots were dried completely by vacuum centrifugation for about 10 min and then soaked in 10-15 µL of 10 ng/µL trypsin in 25

mM NH_4HCO_3 for 20-40 min on ice, until the gel spots swell back to original size. Then 10 μL of 25 mM NH_4HCO_3 was added to immerse the gel spots. The gel spots were incubated at 37 °C in an oven overnight. The tryptic peptides were extracted from the gel plugs by washing once with 15 μL 25 mM NH_4HCO_3 for 15 min, and twice with 15 μL 75% acetonitrile, 0.5% trifluoroacetic acid (TFA) in water for 15 min each. The supernatant from each wash was combined and concentrated to 4-10 μL by vacuum centrifugation for 40 min. Finally, 1-2 μL 5% TFA in H_2O was added to make final TFA concentration between 0.1-1%.

2.1.5.2. Mass Spectrometry Analysis

The tryptic peptides were desalted and concentrated with a ZipTip $_{\mu\text{-C18}}$ (Millipore) following the manufactures instructions. The peptide mixture was eluted from the Ziptip with 2 μL freshly made saturated matrix solution (α -cyano-4-hydroxycinnamic acid in 50% acetonitrile, 0.1% TFA), and spotted directly onto a MALDI plate. The sample was analyzed by MALDI-TOF (time-of-flight) MS (Autoflex, Bruker). The mass spectrum was acquired over a range of m/z 600 to 6000 and calibrated using trypsin autolysis products as internal standards. MALDI-TOF MS data were searched against the NCBI genome database of archaeal species using the MS-Fit program (<http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm>) or the machine built-in software, Biotool, which is based on the MASCOT searching program (http://www.matrixscience.com/search_form_select.html).

2.1.6. Strategy of DNA-affinity Protein Capture; Standard Operations

Separation of beads from supernatant. The Dynabeads were placed on a MPC magnet. After the beads all migrate to the side toward the magnet, the supernatant was aspirated with a pipette.

DNA immobilization. Biotinylated DNA was mixed with streptavidin-coated Dynabeads in 1x B&W buffer and incubated for 15 min on a vortex mixer.

Equilibration of bead-bound DNA. The bead-bound DNA was twice washed with 1x Incubation buffer at room temperature.

DNA-cell extract incubation. The bead-bound DNA was mixed with *Pf* cell extract and incubated for 30 min at selected temperature. The mixture was inverted every 1-2 min to prevent bead settling.

Standard washing. Bead-bound DNA-protein complexes were subjected to three quick washes at room temperature with 1x Incubation buffer.

Heparin challenge. Bead-bound DNA-protein complexes were washed in sodium heparin solution for 15 min.

DNA digestion with Turbo DNase. Turbo DNase was used for digestion of DNA in bead-bound DNA-protein complex. 1 unit of DNase was applied to 1 µg DNA in 1x Turbo buffer and incubated at 37 °C for 20 min. The beads were kept shaking on a shaker for efficient digestion.

Elution of proteins with Gel-loading buffer. The bead-bound DNA-protein complexes were incubated in 1x Gel-loading buffer for 10 min. The dissociated proteins were separated effectively by one-dimensional (1D) SDS Polyacrylamide Gel electrophoresis (SDS PAGE).

Conditions, amounts, and reaction parameters were further optimized as described in the following sections.

2.2. Optimization of DNA-affinity Protein Capture Method

2.2.1. Influence of UOR Length on DNA Immobilization

2.2.1.1. Dynabead Preparation

Dynabeads are covalently coated with streptavidin proteins (MW 66 kDa), which contains four subunits (16 kDa each) (DYNAL, 2002). 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% NaN₃ as a preservative (DYNAL, 2002). Before immobilizing biotinylated DNA on beads, we replaced the beads' preservative solution and washed beads with 2x B&W buffer. For example, 50 µL Dynabeads M-280 in a 1.5-mL microcentrifuge tube was washed with 100 µL 2x B&W buffer. (The buffer volume was twice the Dynabeads volume.) Washing was performed by suspending the Dynabeads in solution with gentle shaking or vortex mixing of the tube, followed by placing the tube in a MPC magnet for separation. The Dynabeads should remain in solution and cannot be left to dry in air for long.

50 µL Dynabeads is enough for good protein detection in gel by Coomassie blue stain or silver stain. But, for protein in-gel digestion, 100 µL Dynabeads was needed. With more Dynabeads, the applied DNA, protein cell extract, and washing solution should also be scaled up in the following steps.

2.2.1.2. Measurement of Amount of DNA Immobilized on Dynabeads

Before adding proteins to bead-bound DNA, it is necessary to know how much DNA is on the beads, because that decides the amount of protein needed for protein capture. The amount of dsDNA that is sufficient to saturate the bead-coated streptavidin is DNA length dependent. Specifically, the number of dsDNA molecules on the beads is approximately inversely correlated

to the length of DNA due to steric hindrance. To measure the amount of a given dsDNA that saturates the Dynabeads, we conducted a "titration", adding DNA to a known amount of beads until the beads can bind no more DNA. Quantification and titration of Dynabead-immobilized DNA followed these steps:

1. Shake the Dynabead container gently to form homogenous suspension, and take out 20 μL beads (10 mg/mL) in a 1.5- μL microcentrifuge tube for the following assay.
2. Wash beads twice with 40 μL 2x B&W buffer and remove the supernatant with a MPC magnet.
3. Add 20 μL biotinylated DNA (50-100 ng/ μL , quantified with DyNA Quant fluorometer) and 20 μL 2x B&W buffer to the Dynabeads (final salt concentration is 1 M NaCl). Incubate beads and DNA at room temperature for 15 min (for <1000 kb DNA) with gentle vortex mixing or gently tapping the tube to prevent bead precipitation. Separate the Dynabeads (binding biotinylated DNA now) from the supernatant. Save the supernatant as Sup1.
4. Quantify the DNA in Sup1 with DyNA Quant fluorometer. Calculate the decrease in DNA concentration after the immobilization, which is equal to the amount of DNA immobilized on the beads. (Note that the volume was doubled after incubation.)
5. If the amount of DNA immobilized is more than 10 ng/ μL , add more DNA to the Dynabeads, repeating steps 3 and 4, until less than 10 ng/ μL of DNA is lost after incubation.)
6. Sum the total amount of DNA immobilized on beads from repeated incubations.

2.2.1.3. Effect of DNA Length on Dynabead Binding

We found that the length of DNA influences the amount of DNA immobilized on the surface of Dynabeads (apparently due to steric hindrance). The DNA-binding capacity of Dynabeads M-280, 10 mg/mL ($6-7 \times 10^8$ Dynabeads/mL) is summarized in **Table 2.1**.

Table 2.1 The amount of DNA immobilized on Dynabeads

Length of dsDNA	Binding capacity (mass) for 1 μ L Dynabeads (10 mg/mL)	Binding capacity (molarity) for 1 μ L Dynabeads (10 mg/mL)
528 bp	64 ng ($\pm 8\%$) DNA	0.2 pmol ($\pm 8\%$) DNA
300 bp	73 ng ($\pm 7\%$) DNA	0.4 pmol ($\pm 7\%$) DNA
72 bp	75 ng ($\pm 7\%$) DNA	1.7 pmol ($\pm 7\%$) DNA

2.2.1.4. DNA Immobilization Procedure

The dsDNA was modified only at the 5' end of the forward oligonucleotide strand to assure all dsDNA bound to Dynabeads in the same orientation. Biotinylated dsDNA was mixed with Dynabeads in 1.5-mL microcentrifuge tubes and incubated in 1x B&W buffer at room temperature. The tube was kept shaking gently with a vortex mixer to prevent beads forming clumps in solution. After 15 min, (for <1 kb DNA; longer for >1 kb DNA), the suspension was placed in a MPC magnet. Once the beads migrated to the magnet side and separation was complete by observing clear supernatant, the supernatant containing non-immobilized DNA was aspirated.

Some DNA concentrating or diluting before application might be necessary to satisfy the solution volume used in DNA immobilization. Experience suggests that the reaction volume of each step should be about twice the volume of the original Dynabeads (at 10 mg/mL) (Chen, 2003). Either too little or too much solution will not let the beads and biomolecules mix thoroughly and efficiently. For example, if 50 μ L Dynabeads was used, 100 μ L DNA in 1x B&W buffer will be optimal for DNA immobilization. For 50 μ L Dynabeads, a reaction volume of 100 μ L for every step is suggested. However, if for any reason the concentration of DNA cannot be high enough, it is fine to just add enough DNA and an equal volume of 2x B&W buffer. The salt concentration has to be 1 M as in 1x B&W buffer. Actually for 50 μ L beads, reaction volumes of 30 to 200 μ L all worked fine as long as the salt concentration and other components in the buffer were proper.

2.2.2. Lack of Effect of BSA in Protein Capture

In earlier work by the Scott group, the use of BSA was suggested to block protein binding sites on the Dynabead surface preventing protein "capture" unrelated to DNA-affinity

(Chen, 2003). But, whether BSA successfully blocks the bead surface from other proteins had not been carefully studied. Since it is possible that BSA or proteins that contaminate commercial BSA would interfere with protein capture and identification, we decided to test the "blocking" effect of BSA.

2.2.2.1 Method and Experiment

Two pairs of experiments A, Ac (A-control), B, and Bc (B-control) were performed. In short, experiments A and Ac involved no cell extract, but A was mixed with BSA while Ac was not; B and Bc both involved cell extract, but B was mixed with BSA before addition of cell extract, while Bc was never mixed with BSA. Four tubes of 50 μ L Dynabeads were prepared following the standard procedures. The DNA immobilization step was omitted, but the beads still underwent a wash with 100 μ L of 1x B&W buffer. Dynabeads A and B were washed 3 times with 100 μ L Incubation buffer containing 0.2 mg/mL BSA, while Dynabeads Ac and Bc were just washed in the Incubation buffer without BSA. Incubation with *Pf* cell extract was omitted for A and Ac, so they were washed 3 times with 100 μ L Incubation buffer and the washing supernatant was saved for further analysis. Dynabeads B and Bc were incubated with 5 mg/mL *Pf* cell extract in 100 μ L Incubation buffer at 70 °C for 30 min, but 0.2 mg/mL BSA was mixed with the cell extract in experiment B, while no BSA was added for experiment Bc. Both Dynabeads B and Bc were then washed with 100 μ L Incubation buffer 3 times and the supernatants were saved. All Dynabeads were eluted with 100 μ L 1x Gel-loading buffer at 50 °C. The Dynabeads were then incubated for 5 min at 100 °C in 50 μ L 1x Gel-loading buffer. The results were analyzed by 10-20% protein SDS PAGE and visualized by silver staining.

2.2.2.2 Results

Figure 2.2 shows the results of the above experiments. In experiment B, BSA was applied to Dynabeads before cell extract, while in Bc, no BSA was added. Lanes 6 and 7 of **Figure 2.2b** were loaded with proteins that survived the washing, but were eluted with 50 °C incubation (these would be considered "captured"). Both these experiments showed almost the same protein mixture, indicating that the BSA did *not* block the Dynabead surface preventing access by cell extract proteins. **Figure 2.2a** shows very little BSA in lanes 5, 7, or 9 (the A experiment), showing that the Incubation buffer applied was enough to remove nearly all of the BSA from the beads. In summary, BSA does not act as a blocking protein, nor is it present in the final protein eluate as background. The DNA-affinity protein capture experiment does not have to involve the application of BSA; this will simplify the experimental procedures and reduce the reaction time.

2.2.3. Amount of Cell Extract Used for Protein Capture

The amount of cell extract protein applied to bead-bound DNA is an important factor that directly influenced experimental results. It is obvious that the applied cell extract has to contain a sufficient amount of transcription factors, so that they can be detected and identified by the applied analytical techniques. That amount is probably case-dependent, based on both the expression level and the DNA-binding affinity of the specific protein. More protein is preferred for the purpose of identification of unknown transcription factors. However, if too much protein is used, the quantitative information will be lost due to the saturation of the DNA binding sites when comparing different environmental stress conditions. Therefore the appropriate amount of applied protein mixtures will fall in a dynamic range that is dependent on multiple factors, such

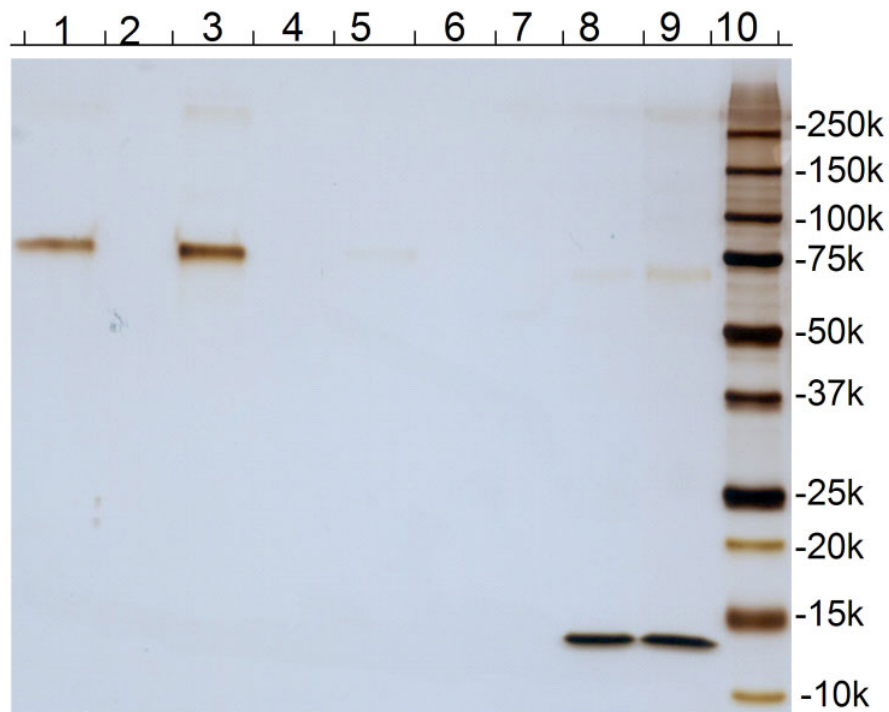


Figure 2.2a SDS PAGE for experiments A and Ac. Lane 1, BSA; Lane 2, supernatant of Washing buffer elution for Ac; Lane 3, supernatant of Washing buffer elution for A (after addition of BSA); Lanes 4/5, supernatant of 1xTurbo DNase buffer elution of Ac (4) and A (5); Lanes 6/7, supernatant of Gel-loading buffer elution of Ac (6) and A (7) after incubation at 50 °C; Lanes 8/9, supernatant of Gel-loading buffer elution of Ac (8) and A (9) after heating to 100 °C; Lane 10, protein molecular weight standard. Very little BSA is seen in lanes 5, 7, and 9, indicating that the Washing buffer was enough to remove BSA from beads.

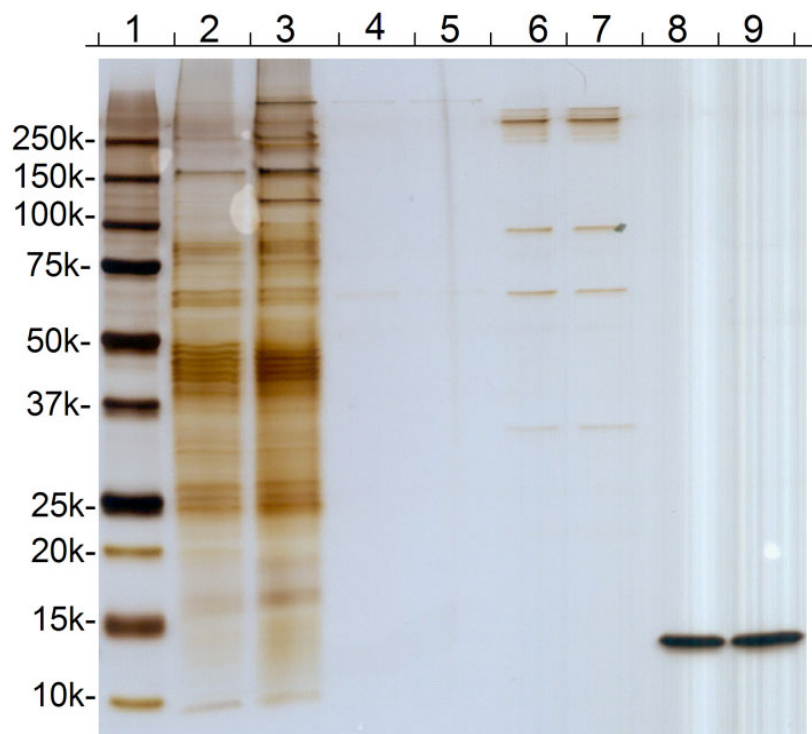


Figure 2.2b SDS PAGE for experiments B and Bc. Lane 1, Protein molecular weight standard; Lanes 2/3, supernatant of Washing buffer elution for Bc (2) and B (3) after incubation with the cell extract. Lanes 4/5, supernatant of 1xTurbo DNase buffer elution for Bc (4) and B (5); Lanes 6/7, supernatant of Gel-loading buffer elution for Bc (6) and B (7) after incubation at 50 °C; Lanes 8/9, supernatant of Gel-loading buffer elution of Bc (8) and B (9) after heating to 100 °C. No obvious difference was seen between B and Bc in the final protein eluate (Lanes 6/7), so the application of BSA did not influence the proteins associated with the Dynabead surface.

as the protein expression level, DNA-binding affinity, DNA molar amount, applied analytical techniques, etc.

2.2.3.1. Theoretical Estimation

Before any experiments were performed, a theoretical "back of the envelope" estimation was carried out to give a rough idea about the amount of protein to use. From the genome database in NCBI, *Pyrococcus furiosus* contains 2125 genes. Assuming that all genes are equally expressed, the average molecular weight of the proteins is 50 kD, and 1 pmol promoter DNA will bind 1 pmol rTF, then the total protein content in *Pf* cell extract for 1 pmol promoter dsDNA will be 10.6 μ g (assuming we want to capture 1 of the 2125 gene products). However, given the widely varying expression levels for different genes, we expect that the proper amount of cell extract to use has to be obtained by empirical experiments.

2.2.3.2. Empirical Estimation of Amount of Protein to Use

In this experiment, different amounts of cell extract were applied in DNA-affinity protein capture experiments. We chose final protein concentrations of 3.5, 14, and 28 mg/mL in the 160 μ L of bead-bound DNA-protein incubation suspension. The DNA used was 40 pmol of 300 bp *lrpA* UOR for each 80 μ L Dynabeads. The DNA-binding proteins were released by DNase digestion followed by eluting with 1x Gel-loading buffer and analysis by SDS PAGE using 10-20 % gel. The gel was visualized by GelCode Blue and quantitatively analyzed by densitometry (**Figure 2.3**). Proteins from the gel were subjected to in-gel digestion and mass spectrometry identification.

The gel densitometry analysis indicated that the amount of each DNA-binding protein in the 14 and 28 mg/mL cell extracts were essentially the same, higher than the amount of protein from the 3.5 mg/mL cell extract. Using MS for protein identification, the gel band labeled with

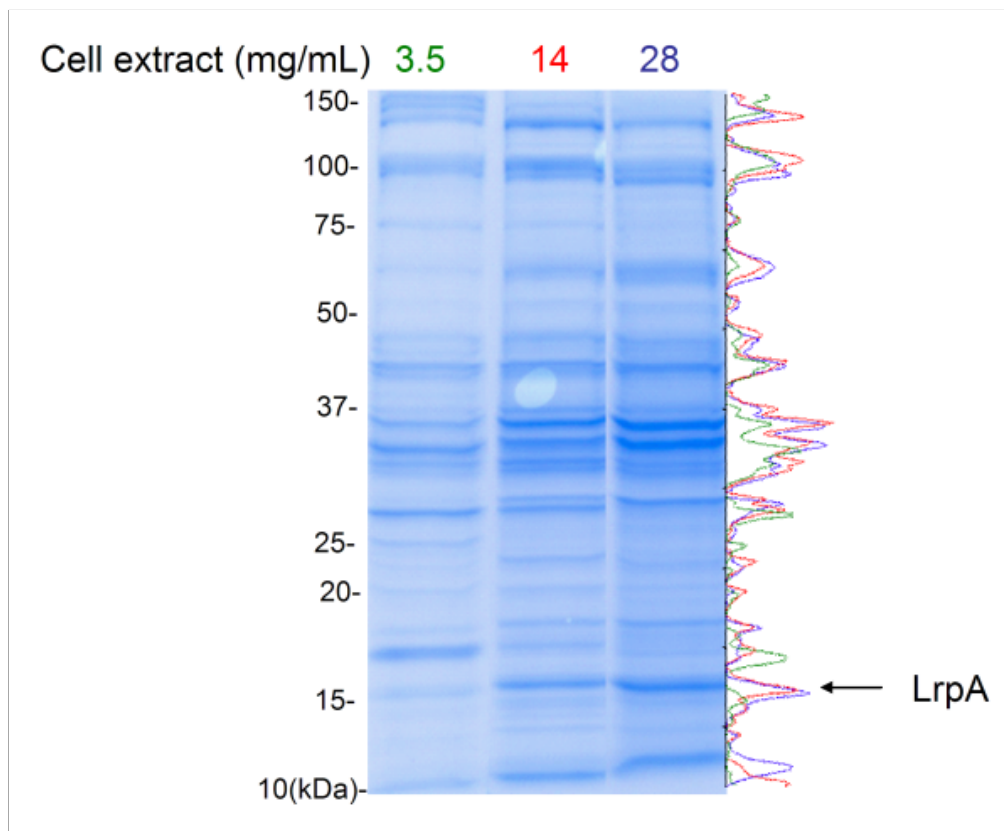


Figure 2.3 SDS PAGE analysis using different amounts of applied cell extract. Proteins bound to DNA from different applied concentrations: 3.5 (green); 14 (red); and 28 mg/mL (blue) were eluted and loaded into 10-20 % SDS PAGE. The gel densitometry analysis for each lane indicated generally the increase of protein intensity from the 3.5 mg/mL concentration sample to the 14mg/mL concentration sample, and about equal intensity of the proteins in the right two lanes (14 and 28 mg/mL). This demonstrated that the saturation of DNA binding was reached at least by the cell extract protein concentration of 14 mg/mL.

the arrow was identified as the known transcriptional regulator LrpA, which was about equally intense in the right two lanes and much fainter in the left lane. So in this experiment for 80 μ L beads with 40 pmol 300 bp DNA, a protein concentration of 14 mg/mL and beyond in the 160 μ L reaction solution was over the saturation point for DNA binding. Therefore the proper range of protein concentration in the DNA binding reaction is between 3 and 14 mg/mL for this experiment. A protein concentration of 5 mg/mL is a practicable value for both quantitative analysis and gel-based protein identification.

2.2.3.3. Other Factors

Assuming one rTF binding site per UOR, the amount of applied cell extract should be dependent on the molar amount of the immobilized UOR present in the reaction. Since we have already shown that the DNA loading is approximately inversely related to DNA length, the total binding capacity of a given amount of beads should be DNA length-dependent. For shorter DNA, the total number of rTF binding sites will be higher and more DNA-binding proteins are required to saturate the binding sites; for longer DNA, the protein amount for saturation of binding sites is lower.

Also, the required protein concentration for protein capture relies on the detection method. With more sensitive detection methods, we will require less protein. If the cell extract is fractionated, it would allow a relatively higher concentration of DNA-binding proteins, so the total amount of applied protein mixture could be less.

2.2.4. Protein-DNA Incubation Temperature

Incubation temperature is perhaps one of the most critical parameters for the study of protein-DNA interactions. The temperature should allow the protein to behave as it does naturally *in vivo*. Generally, protein molecules prefer low temperature *in vitro* (e.g., 4 °C) and are

easily denatured at high temperature. However, since *Pyrococcus furiosus* is a hyperthermophile naturally living around 100 °C, many *Pf* proteins are stable and functional at extremely high temperature (Adams *et al.*, 1994). Many *Pf* *in vitro* transcriptional assays and protein-DNA interactions were studied at temperatures of 70 °C or above (Dahlke *et al.*, 2002; Hethke *et al.*, 1996; Hethke *et al.*, 1999; Lee *et al.*, 2003). In our DNA-affinity protein capture studies, several additional factors must be considered in the practical experiments.

2.2.4.1. Thermal Stability of Dynabeads

First, since the magnetic Dynabeads are used in each step of the protein capture, we need to know the highest temperature at which Dynabeads are stable. The Dynabead surface is coated with streptavidin proteins, which will dissociate from the beads after certain amount of heating. The manufacturer claims that the beads can be heated to 95 °C for a short period of time (<5 min), but for a 30-min protein-DNA incubation step, the streptavidin-coated Dynabeads may not survive that temperature (**Figure 2.4**).

In this thermal stability experiment, each tube contained 20 µL Dynabeads and underwent the standard wash with 2x B&W buffer. The beads were resuspended in the Incubation buffer and heated for 30 min at 60, 70, 75, 80, 85, and 90 °C, respectively. The supernatants were analyzed by protein SDS PAGE. As shown in **Figure 2.5**, 70 °C seems to be the highest temperature allowing about half of the streptavidin to remain with the beads. Any temperature above 75 °C resulted in significant dissociation of streptavidin from beads. As the streptavidin is more stable after forming a complex with biotin molecules, the immobilization of enough biotinylated DNA should stabilize the streptavidin on the beads.

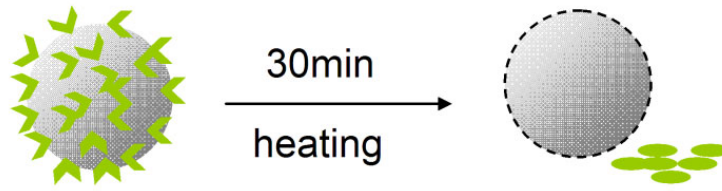


Figure 2.4 In our Dynabead thermal stability assay, streptavidin-coated Dynabeads underwent 30 min of heating to determine at what temperature the streptavidin is dissociated from the bead surface.

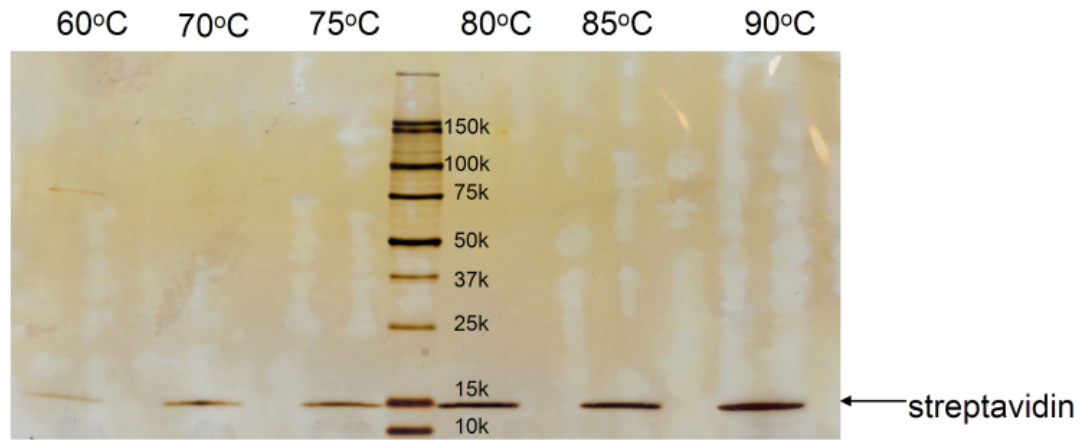


Figure 2.5 SDS PAGE of proteins eluted from streptavidin-coated Dynabeads incubated at indicated temperatures. 20 μ L Dynabeads in 40 μ L Incubation buffer were incubated for 30 min under the indicated temperatures. The supernatant from each reaction was separated from beads by a magnet (DynaL MPC) and loaded in 4-20% gradient SDS PAGE gel. The amount of dissociated streptavidin increased as the incubation temperature increased. 70 $^{\circ}$ C is considered the maximum incubation temperature for our protein capture experiments.

2.2.4.2 Incubation Temperature of DNA and Cell Extract

We compared two DNA-affinity protein capture experiments performed at two different incubation temperatures, 55 and 70 °C. Sufficient 528 bp *lrpA* UOR was immobilized on 100 µL Dynabeads in each reaction. After standard wash and equilibration, an equal amount of *Pf* cell extract (0.6 µg) was mixed with bead-bound DNA and incubated for 30 min at 55 or 70 °C. In addition, an equal amount of recombinant protein (His)₆-LrpA was added to the mixture and it was incubated an additional 30 min at the appropriate temperature. The DNA-binding proteins eluted in the final fractionation were analyzed by 10-20% 1D SDS PAGE. The gel was stained with silver nitrate and scanned for densitometry analysis of each protein band (**Figure 2.6**).

As shown by the gel densitometry analysis, the intensity of protein bands in the 70 °C experiment was generally lower than that in the 55 °C experiment. We interpret this to be the result of greater loss of DNA given the instability of streptavidin at 70 °C after heating for 1 h. The intensity of most of the background proteins from the 70 °C experiment (which we presume involve many nonspecific DNA-binding proteins) was much lower than that in the 55 °C experiment. We hypothesize that this indicates a relatively decreased protein-DNA-affinity as the temperature increases if there is no specific sequence recognition. At the same time, for LrpA the intensity ratio is reversed (higher at 70 °C compared to 55 °C). We hypothesize that this results from LrpA recognizing a specific DNA sequence and forming a tighter complex with DNA, which is even stable at high temperature.

This suggests that increasing the incubation temperature in our protein capture experiments will help to reduce the non-specific DNA-binding interactions.

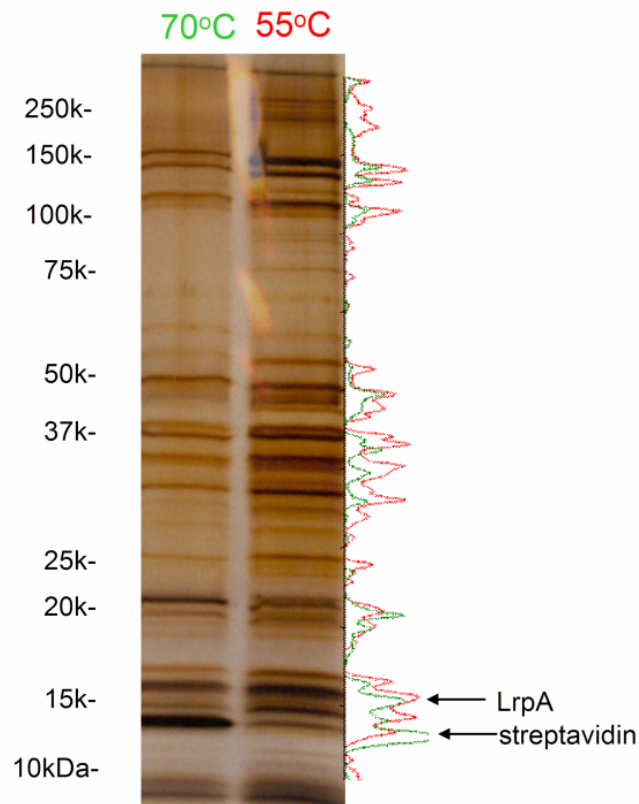


Figure 2.6 SDS PAGE for different incubation temperature. The gel image shows the results of two DNA-affinity protein capture experiments performed at different incubation temperatures: 70 °C (green) and 55 °C (red). The last eluting fractions in the two assays were loaded on the gel. Densitometric analysis indicated more loss of streptavidin from the bead surface in the 70 °C assay, which suggests the real DNA ratio in the experiment (less DNA remains on the beads in the 70 °C lane). Comparison of the protein intensity ratio suggests that a lower amount of background proteins are eluted in the 70 °C assay. Also, a relatively higher amount of LrpA was observed in the 70 °C assay.

2.2.5. Ionic Strength of Protein-DNA Incubation

Ionic strength (salt concentration) is another critical parameter for protein-DNA interactions. Both specific and non-specific protein-DNA binding involve electrostatic interactions that can be affected by ionic strength. *Pyrococcus furiosus* naturally lives in an extremely high salt environment. In published studies of *Pf* transcriptional regulators, the salt concentration was modest (Hethke *et al.*, 1996; Lee *et al.*, 2005), although it is known that the internal ionic strength of *Pf* cells is ca 0.8 M (K) (Scholz *et al.*, 1992). The effect of ionic strength on our protein capture experiment seemed worth investigating.

Two experiments (A and B) were performed in parallel comparing captured proteins from incubation under different salt concentration. Sufficient 528 bp *lrpA* UOR was immobilized on 40 μ L Dynabeads in each reaction. The bead-bound DNA was equilibrated with Incubation buffer containing different concentrations of salt: 0.1 M (A) or 0.5 M KCl (B). 0.2 μ g *Pf* cell extract and 0.5 μ g recombinant protein LrpA were mixed together with the bead-bound DNA. The mixtures underwent incubation at 55 °C, standard wash with Incubation buffer and heparin challenge. The DNA-binding proteins were released by DNase digestion. The washing supernatants and eluted proteins were analyzed with 1D 10-20% SDS PAGE and the gel was stained with silver nitrate.

Figure 2.7 shows that LrpA (labeled with arrow) was only observed in the eluate from experiment A (0.1 M KCl). No LrpA was detected in experiment B (0.5 M KCl). It is somewhat unexpected that the *in vitro* DNA-protein interaction was so sensitive to ionic strength in the DNA pull-down, given that *Pyrococcus furiosus* contains 0.8M intracellular KCl.

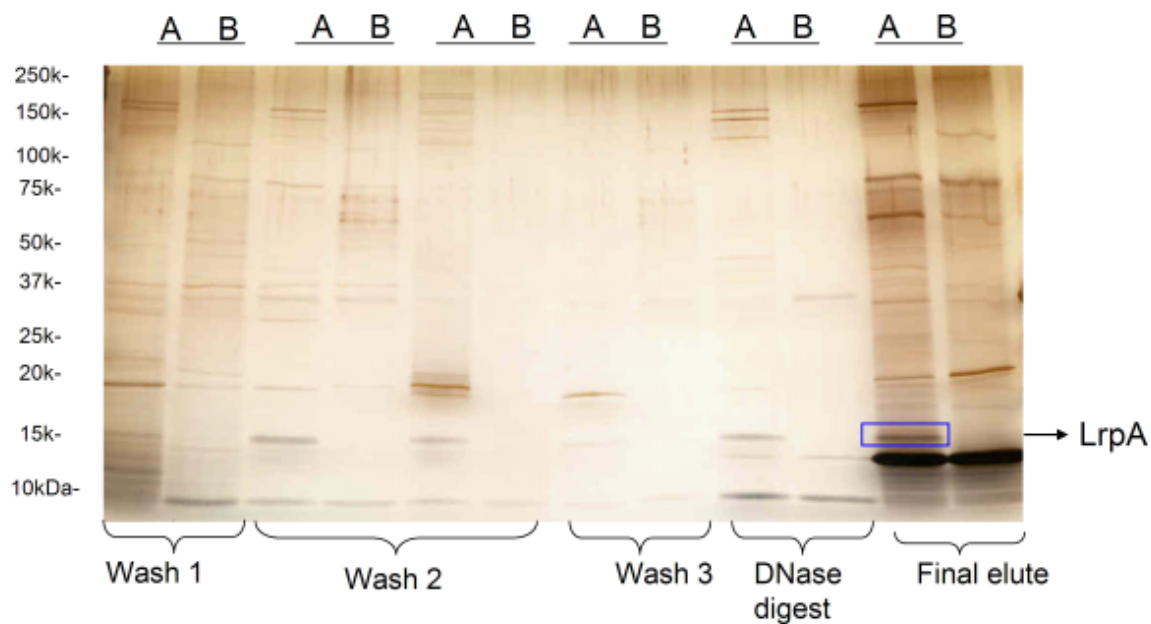


Figure 2.7 The effect of ionic strength in the protein capture reaction. The series of eluates from the two parallel protein capture experiments were analyzed by SDS PAGE. Each pair of lanes labeled with A and B were washing supernatants from experiment A (0.1 M KCl) or experiment B (0.5 M KCl). Our control rTF, recombinant LrpA, was only observed in experiment A. No LrpA was detected in experiment B. Higher ionic strength appears to disrupt DNA-protein interaction in these protein capture experiments.

2.2.6. DNA Competitors

Pf cell extract is expected to contain a mixture of sequence-specific and nonspecific DNA-binding proteins. The UORs used for our protein capture experiments are expected to capture both types. Therefore, to capture sequence-specific DNA-binding proteins, we should consider how to improve our ability to distinguish and separate these two types. It is believed that sequence-specific DNA-binding proteins bind DNA with much higher affinity than do nonspecific DNA-binding proteins (Gabrielsen *et al.*, 1993). This principle has been used as the basis of DNA-affinity chromatography for purification of transcription factors (Gadgil *et al.*, 2001; Kadonaga *et al.*, 1986). Various DNA competitors have been applied to selectively remove weaker nonspecific binding proteins for purification of specific transcription factors. Such competitors include sodium heparin, herring- and salmon-sperm DNA, and poly-dIdC (Gabrielsen *et al.*, 1993; Gadgil *et al.*, 2001; Yaneva *et al.*, 2003). We investigated some of these competitors in our protein capture experiments to determine whether they improved our ability to capture sequence-specific DNA-binding proteins.

2.2.6.1. Sodium Heparin

Heparin is an anionic polysaccharide, containing heterogeneous straight-chain sugars. Due to the chain structure of the sugar backbone, heparin is considered a DNA mimic, and was used as a DNA structure mimic in DNA-affinity transcription factor purification (Gabrielsen *et al.*, 1993). Sodium heparin was studied in our DNA-affinity protein capture experiment as a competitor to prevent access of nonspecific DNA-binding proteins to DNA.

Three dsDNA fragments of different lengths were designed based on the *lrpA* UOR (numbering with respect to *lrpA* translation start): –305 to –130 (F1R1, 175 bp), –305 to +47 (F1R2, 352 bp), and –305 to +223 (F1R3, 528 bp) (**Figure 2.8**). We used the same primers as

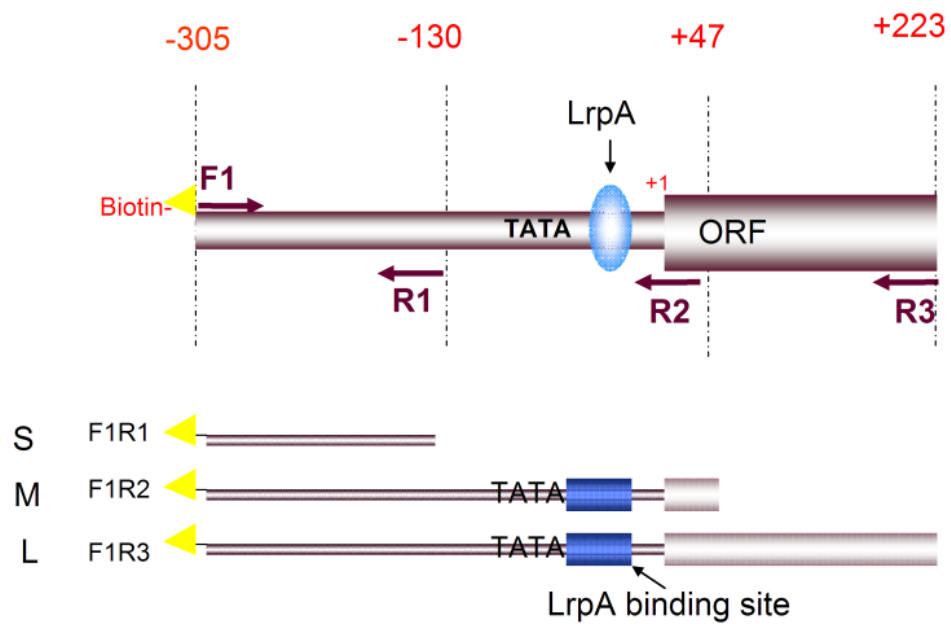


Figure 2.8 Three DNA fragments were amplified by PCR with only F1R1 and F1R3 containing the LrpA binding site.

discussed in 2.1.3.2. Only F1R2 and F1R3 contained the LrpA binding site. The dsDNA fragments were immobilized on 40 μ L Dynabeads for parallel protein capture experiments. The bead-bound DNA underwent standard equilibration, and each was incubated with 0.2 μ g *Pf* cell extract plus 0.5 μ g recombinant LrpA and *Pf* TBP at 70 °C. After standard washing with Incubation buffer, the heparin challenge step was performed by washing bead-bound complexes with Incubation buffer containing 0.1 M sodium heparin at 70 °C for 15 min. The proteins retained on the beads were dissociated by DNase digestion. Remaining beads were heated to 100 °C for dissociation of any complex molecules. The proteins eluted from each step were analyzed by 1D 4-20% SDS PAGE and visualized by silver staining. The proteins in gel underwent in-gel digestion and mass spectrometry identification (**Figure 2.9**).

From MS identification, recombinant protein LrpA and *Pf* TBP captured by *lrpA* UOR were identified as indicated by arrows in **Figure 2.9**. Many background proteins were released by heparin challenge (first three lanes, **Figure 2.9**), while only little LrpA protein was released by heparin. So the advantage of sodium heparin is that it is simple, cheap, and more importantly, able to release some apparently nonspecific DNA-binding proteins. The heparin challenge also can be performed at room temperature for even shorter time (data not shown). On the other hand, we do not know the identity of the proteins released by heparin challenge and it did not remove all "background" proteins. Heparin *did* release some LrpA and it is still theoretically capable of removing sequence-specific DNA-binding proteins whose properties are completely unknown.

We also found that LrpA was captured only by the DNA fragments containing the LrpA binding sequence (not by F1R1, the -305 to -130 UOR), indicating the feasibility of our protein capture approach.

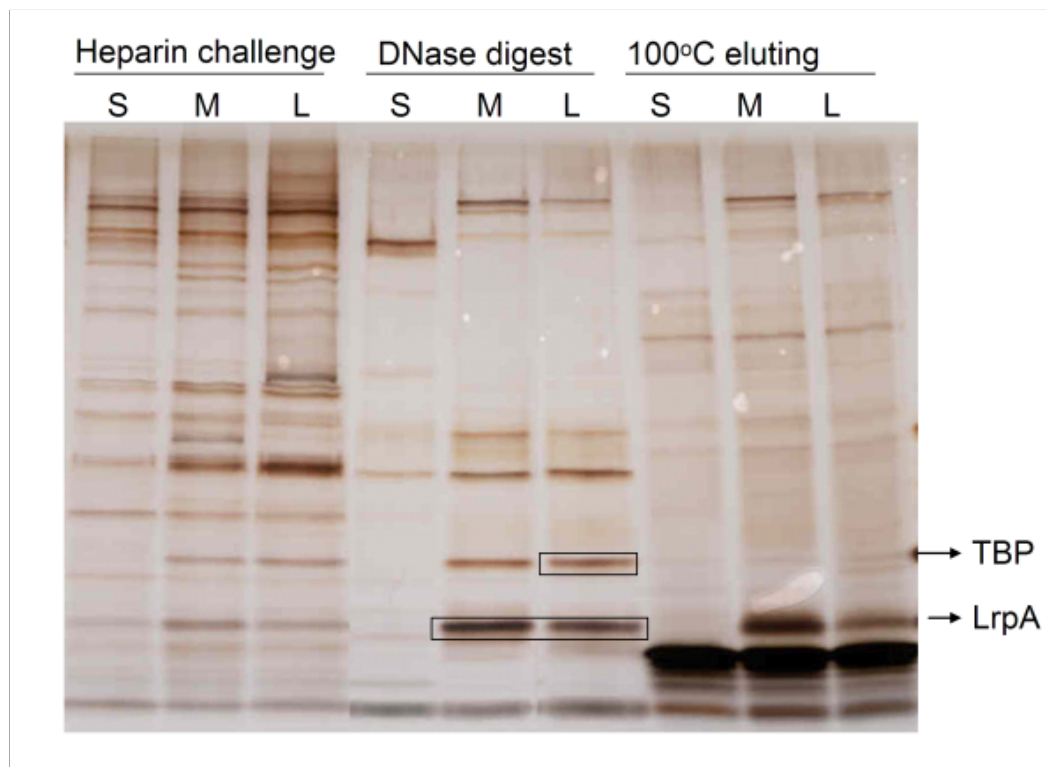


Figure 2.9 Parallel protein capture experiments with different lengths of DNA from *lrpA* UOR. 1D SDS PAGE separation of the protein supernatants from sodium heparin challenge, DNase digest, and 100 °C elution. Recombinant protein LrpA and *Pf* TBP were mixed with *Pf* cell extract and applied to the bead-bound DNA. As shown in the first three lanes, heparin challenge can release many "background" proteins but little LrpA and TBP. LrpA and TBP were captured only by the DNA containing their cognitive sequences (F1R2 and F1R3) and were eluted by DNase digestion. S, M, L are defined in **Figure 2.8**.

2.2.6.2. DNA Fragments

Regulatory transcription factors bind to their cognate DNA sequence somewhere within the UOR. One possible means of finding where this binding site is involves the use of UOR fragment constructs as soluble competitors. If different proteins can be dissociated with different UOR fragments, prepared from different regions of the same UOR, this method will provide us with additional information about protein location on the UOR. Based on that hypothesis, fragment DNA from different regions of the *lrpA* UOR were amplified by PCR and used as DNA competitors in a protein capture experiment targeting LrpA. Three UOR fragments of equal length were amplified from three regions of the *lrpA* UOR (**Figure 2.10**).

In the experiment, sufficient 528 bp biotinylated *lrpA* UOR was immobilized on 40 μ L Dynabeads in each reaction. The bead-bound DNA derived from standard equilibration was mixed with *Pf* cell extract plus 1 μ g recombinant LrpA. The mixture was incubated at 55 °C followed by the standard wash with the Incubation buffer. The derived bead-bound complexes then underwent two heparin challenges at 55 °C for 15 min each. The derived DNA-bound proteins were further washed with fragment F2R2 (6:1 mole ratio F2R2 to bead-bound DNA) at 55 °C for 15 min twice (F2R2-1 and F2R2-2). Finally, the remaining proteins on beads were eluted at 100 °C in Gel-loading buffer. The supernatant samples from each step were analyzed with 4-20% 1D SDS PAGE. In the second experiment, DNA-cell extract was incubated at 70 °C, while other procedures were the same.

UOR fragments F1R1 and F3R3, both the same length as F2R2, but not containing the *lrpA* core promoter (and also not containing the LrpA binding site), were also investigated as DNA competitors. The Dynabeads underwent the same DNA immobilization, standard equilibration, and cell extract incubation as above. The derived bead-bound complexes were

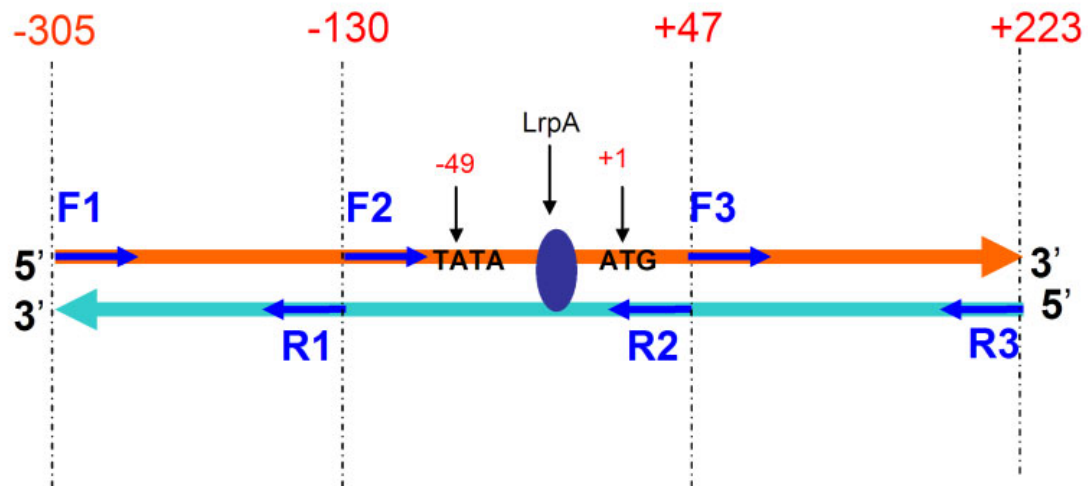


Figure 2.10 DNA competitors from *lrpA* UOR fragments. Three different fragments were amplified by PCR from 528 bp *lrpA* UOR: fragment F1R1 from -305 to -130, fragment F2R2 from -131 to +47, and fragment F3R3 from +48 to +223 (relative to translation start site). Fragment F2R2 contains the TATA element, BRE site, LrpA binding site, and the translation start region.

washed with sodium heparin at 55 °C for 15 min. Then the complexes were incubated sequentially once with DNA fragment F3R3, twice with F2R2, and once with F1R1. Each incubation reaction was performed at 55 °C for 15 min. The mole ratio of applied fragment DNA to bead-bound DNA was 5:1. The retained bead-bound complexes and each DNA-containing eluate were digested with Turbo DNase. The remaining beads were eluted at 100 °C in Gel-loading buffer. The protein sample from each step was analyzed with 10-20% 1D SDS PAGE and visualized by silver staining.

As shown in **Figure 2.11a**, the first wash with 6:1 DNA fragment F2R2 eluted the major fraction of protein LrpA efficiently; the second wash showed only minor additional release. Some LrpA was observed in the last elution step as expected since one-seventh the total amount of LrpA should be left attached to the bead-bound *lrpA* UOR (due to equilibration between these two identical binding sites). At the same time, some background proteins were dissociated by F2R2 as well. This is not surprising, given that F2R2 was 176 bp long. The experiment at 70 °C showed similar results (data not shown).

The experiment shown in **Figure 2.11b** confirms this result. LrpA was eluted by fragment F2R2, but not by fragment F1R1, though F1R1 was applied first. While the eluted proteins from fragment F3R3 also contained a small amount of LrpA protein, it was far less than eluted by fragment F2R2, and was probably due to overloading with LrpA.

This experiment is interpreted to show that the LrpA binding site is within the F2R2 fragment (we already know this to be true in this positive control). However, other proteins were also competed off the bead-bound DNA by F2R2. Possibly using shorter fragments could simplify this pattern, allowing a more detailed identification of binding site. On the other hand, for gel-based protein separation and identification, using fragment DNA as competitors will

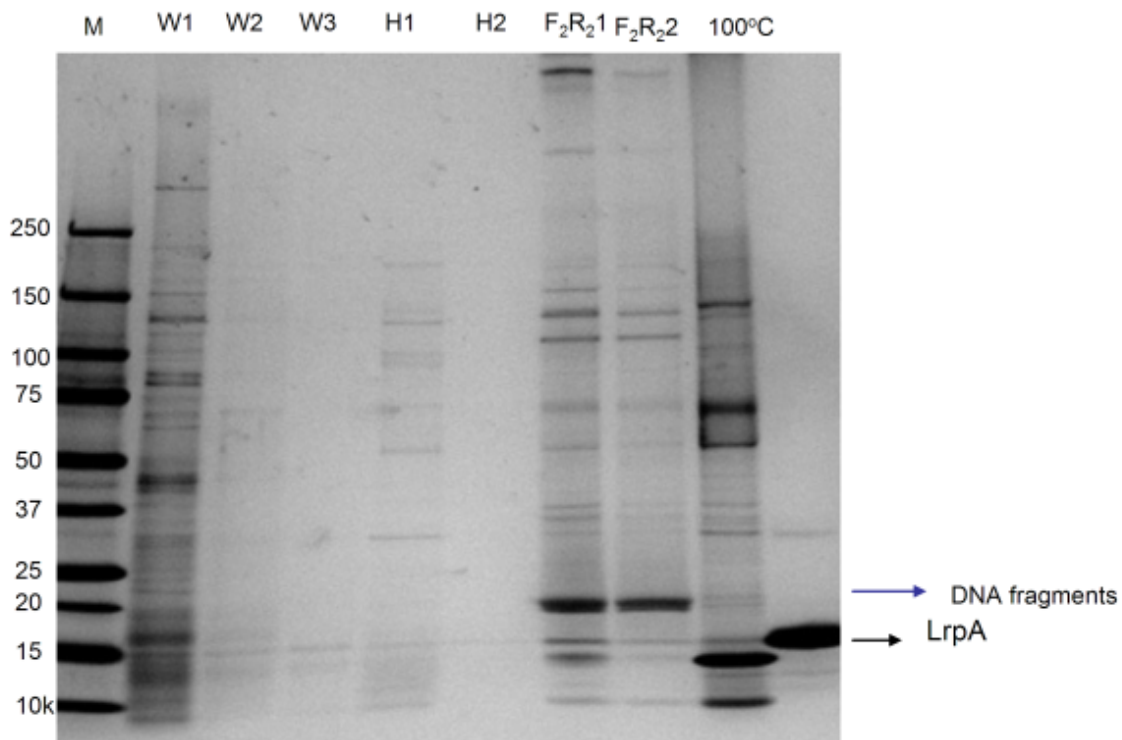


Figure 2.11a 1D SDS PAGE separation of eluted proteins from each step of protein capture with DNA fragment F2R2. The gel lanes from left to right are protein molecular weight marker (M), three Washing buffer eluates (W1, W2, W3), two sodium heparin challenges (H1, H2), two incubations with fragment F2R2 (F2R2-1, F2R2-2; 6:1 mole ratio), final elution with 100 °C heating of beads (100 °C), and pure recombinant LrpA. Sodium heparin did not elute much LrpA but some background proteins. Fragment F2R2 eluted most of LrpA (arrow), as well as some background proteins. The bands labeled with the blue arrow in lane F2R2-1 and -2 are DNA fragments stained by silver nitrate.

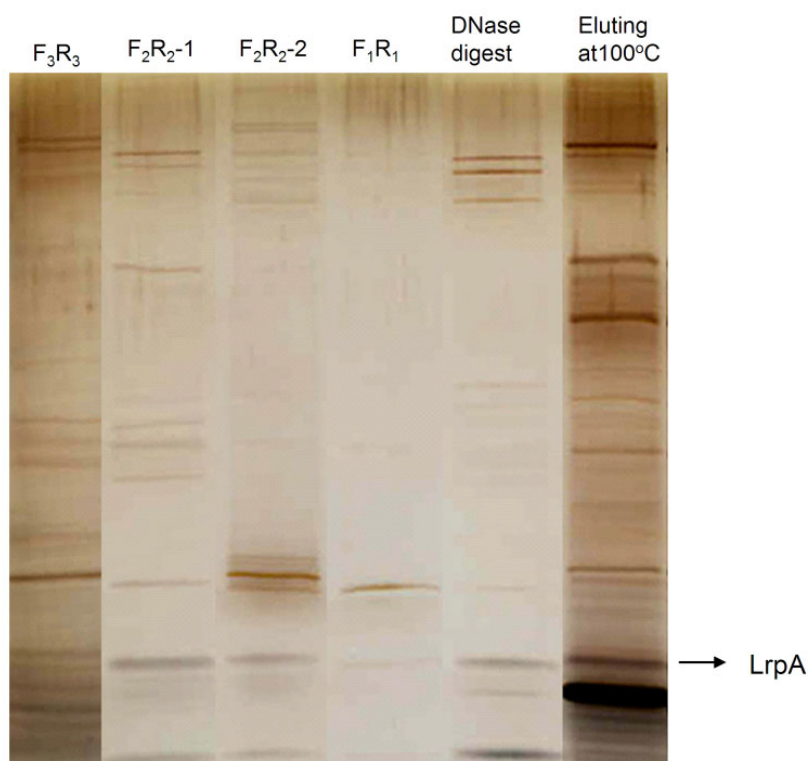


Figure 2.11b 1D SDS PAGE separation of eluted proteins from each step of protein capture with different DNA fragments. The samples in each lane are eluate with fragment F3R3 (F_3R_3), two elutions with F2R2 (F_2R_2-1 , F_2R_2-2), eluate with F1R1 (F_1R_1), eluate with DNase digest (DNase Digest), and final eluate after incubating at 100 °C (Eluting at 100 °C). Recombinant protein LrpA is indicated with an arrow. LrpA was mainly eluted by fragment F2R2. Little or no LrpA was observed in the eluate with DNA fragment F1R1 and F3R3. After the series of washing and eluting, some LrpA protein was still left on the immobilized promoter DNA and eluted by DNase digest and heat treatment.

involve extra steps of DNA preparation, and DNase digestion of each DNA-containing sample for silver staining gel.

2.2.6.3. Poly-dIdC

Poly-dIdC is an alternating copolymer duplex composed of deoxyinosinic-deoxycytidylic acid. Poly-dIdC displays a similar structure as native DNA but deoxyinosine is not a component of natural DNA. Therefore, poly-dIdC can be used as competitor for non-specific DNA-binding proteins, allowing the sequence-specific DNA-binding proteins to access their cognate sequences.

For better eluting efficiency, commercial poly-dIdC was sheared by sonication to an average length of 500 bp (**Figure 2.12**). In the experiment, 40 μ L Dynabeads were used to immobilize 2.5 μ g *lrpA* UOR (528 bp). After equilibration, bead-bound DNA was incubated with *Pf* cell extract plus 0.5 μ g recombinant LrpA and 0.3 μ g recombinant *Pf* TBP at 70 °C for 30 min. After the standard wash, the bead-bound complexes were incubated with 12.5 μ g poly-dIdC at 70 °C for 3 min. Then the bead-bound complexes were incubated with fragment DNA F2R2 (5:1 mole ratio to bead-bound DNA) at 70 °C for 30 min. The retained bead-bound DNA-protein complex was digested with 3 units Turbo DNase followed by protein dissociation in Gel-loading buffer at 100 °C. The eluted proteins were analyzed by 10-20% 1D SDS PAGE and visualized by silver staining.

As shown in **Figure 2.13**, Poly-dIdC was efficient at eluting many "background" DNA-binding proteins. The specific DNA-binding proteins LrpA and *Pf* TBP were not eluted by poly-dIdC but mostly eluted by DNA fragment F2R2. However, the background proteins were spread over several eluting fractions. This result may demonstrate DNA-binding protein equilibration between poly-dIdC and non-specific regions of bead-bound UOR, In conclusion, poly-dIdC was

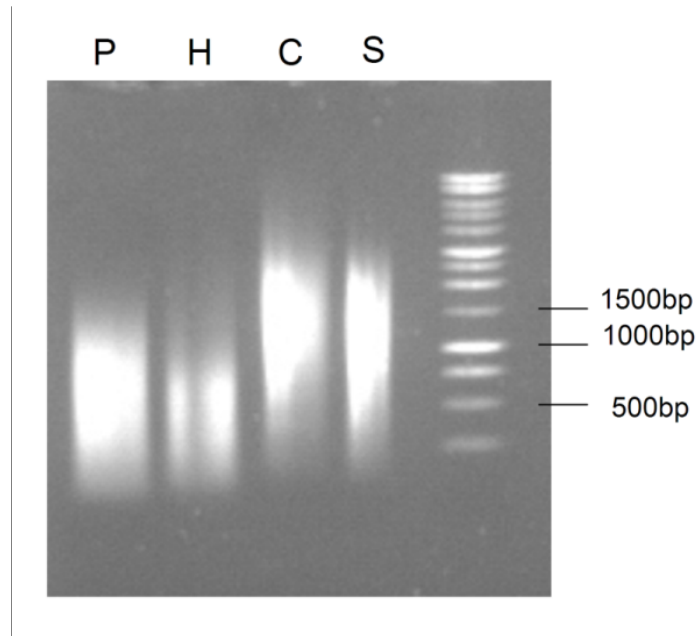


Figure 2.12 Sheared poly-dIdC (P), Herring-sperm genomic DNA (H), Calf thymus DNA (C), Salmon-sperm genomic DNA (S). The average length of sheared poly-dIdC and herring-sperm genomic DNA was about 500 bp, while the sheared Calf thymus and Salmon-sperm genomic DNA was about 1000 bp.

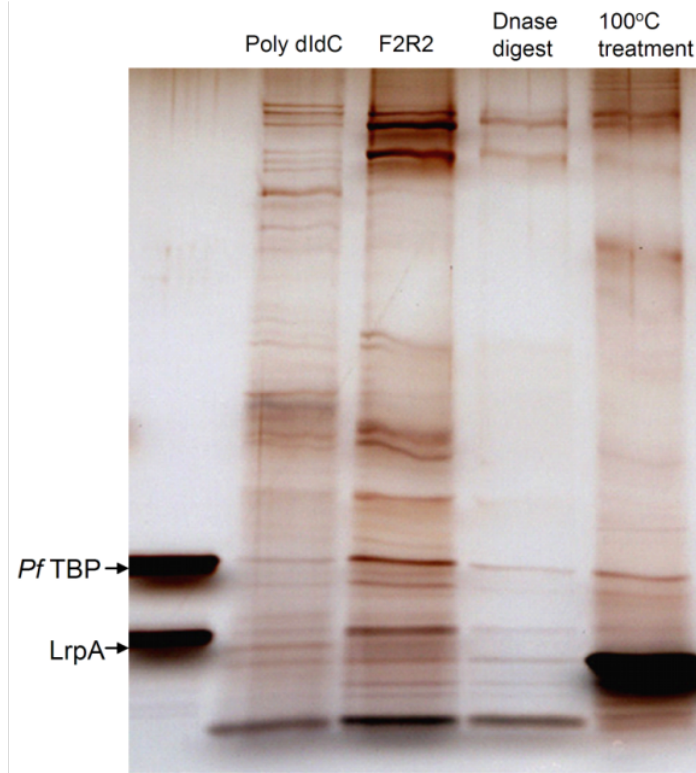


Figure 2.13 SDS PAGE of protein capture with competitor of poly-dIdC. Poly-dIdC eluted many apparently non-specific DNA-binding proteins. Sequence-specific DNA-binding proteins LrpA and *Pf* TBP were mostly eluted by fragment F2R2. LrpA and *Pf* TBP are indicated with arrows.

an efficient DNA competitor for non-specific DNA-binding proteins. Compared with fragment DNA, poly-dIdC was quick and single-step, but it does not provide any protein location information as fragment DNA does.

2.2.6.4. Other Genomic DNA as Competitors

For dissociation of non-specific DNA-binding proteins, native genomic DNA of other species may also serve as competitor to the bead-bound UOR. Herring-sperm, salmon-sperm, and calf-thymus DNA were sheared by sonicating for 2 min and the sheared DNA were checked by 1% agarose gel (**Figure 2.12**). For comparing different DNA competitors, five experiments were performed in parallel using the three genomic DNAs described above plus poly-dIdC and sodium heparin. 40 μ L Dynabeads were used in each experiment to immobilize 2.5 μ g 528 bp *lrpA* UOR. After equilibration, incubation of 0.25 μ g *Pf* cell extract at 55 °C, and our standard wash, 5 μ g (2:1 mole ratio to UOR) herring-sperm, salmon-sperm, calf-thymus DNA, poly-dIdC, and 0.1 mg/mL sodium heparin were applied to each experiment and incubated with bead-bound complexes at 55 °C for 15 min in parallel. Incubation with competitors was repeated with 7.5 μ g (for a total of 5:1 mole ratio to UOR) genomic DNA or poly-dIdC and 0.1 mg/mL sodium heparin. The supernatant was saved and the remaining bead-bound complexes were completely eluted at 100 °C in Gel-loading buffer. The protein samples were analyzed with SDS PAGE and stained with silver nitrate.

Figure 2.14 shows that the proteins eluted from herring-sperm DNA and that from poly-dIdC were very similar to one another, while the results from salmon-sperm and calf-thymus DNA were also similar to one another. The former pair of DNA competitors eluted more proteins than the latter two competitors, perhaps because the former pairs were both sheared to about 500 bp and shorter genomic DNA fragments are more efficient at capturing proteins. Heparin eluted

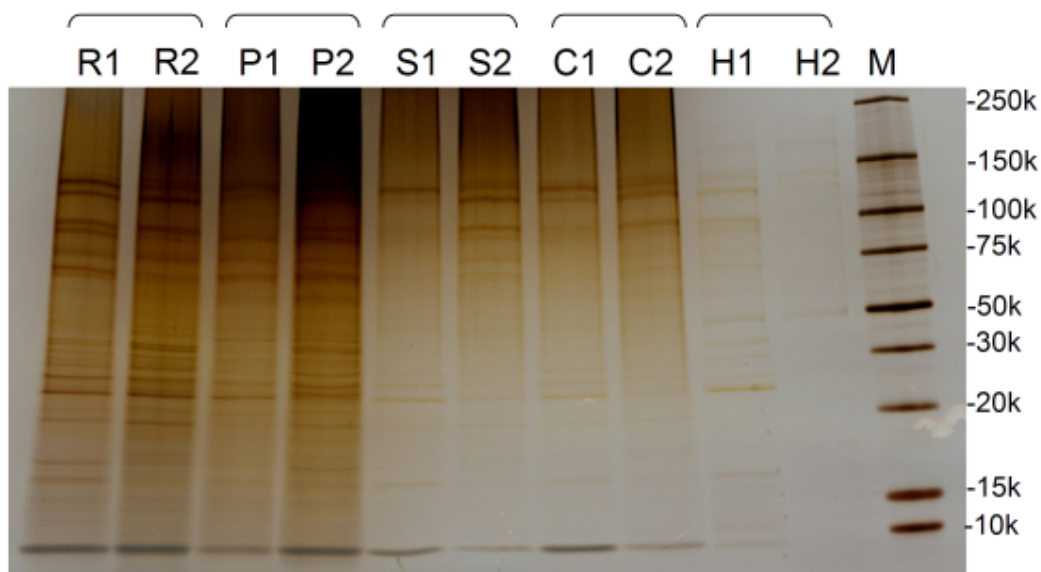


Figure 2.14 Comparison of different competitors for protein capture. The experiment underwent DNA immobilization, equilibration, cell extract incubation, and standard washing, competitors were applied twice consecutively: R1, 1st elution by 5 μ g herring-sperm DNA; R2, 2nd elution by 7.5 μ g herring-sperm DNA; P1, 1st elution by 5 μ g poly-dIdC; P2, 2nd elution by 7.5 μ g poly-dIdC; S1, 1st elution by 5 μ g salmon-sperm DNA; S2, 2nd elution by 7.5 μ g salmon-sperm DNA; C1, 1st elution by 5 μ g calf thymus DNA; C2, 2nd elution by 7.5 μ g calf thymus DNA; H1, 1st elution by 0.1 mg/mL sodium heparin; H2, 2nd elution by 0.1 mg/mL sodium heparin.

less total protein and perhaps a different distribution of proteins compare to the DNA competitors.

In summary, sheared genomic DNA can be used as competitor and it acts similarly to poly-dIdC of the same size. But it should be noted that the native eukaryotic genomic DNA contains TATA box elements that closely resemble archaeal *Pyrococcus furiosus* promoters, though probably account for only a small percentage of the entire genomic DNA sequence. Native genomic DNA was less expensive than poly-dIdC, but poly-dIdC was less relevant to transcription factors.

2.2.6.5. Ionic Strength

Since DNA-protein interactions generally involve electrostatic interactions, it may be possible to release DNA-binding proteins from bead-bound DNA using ionic strength. With the premise that non-specific DNA binding is dominated by electrostatics, we wondered whether non-specific DNA-binding proteins can be separated from sequence-specific DNA-binding proteins by the use of an ionic strength gradient.

We compared the results of protein elution using low- vs. high-salt buffer (0.2 M vs. 0.4 M KCl). Two protein capture experiments were performed in parallel using 120 μ L Dynabeads in each. 8.4 μ g biotinylated *lrpA* promoter (528 bp) was immobilized on beads in each experiment. After equilibration, bead-bound DNA was mixed with 120 μ L *Pf* cell extract (3 mg/mL) plus 1.2 μ g purified recombinant LrpA. The complexes were incubated at 70 °C for 30 min followed by the standard wash with Incubation buffer. Subsequently, the bead-bound complexes were washed by low- or high-salt buffer of 240 μ L in parallel. The beads were shaken gently in the solution for about 3 min at room temperature and separated from the supernatant with a magnet. The retained complexes on beads were released by DNase digestion. The eluted

proteins were analyzed with 10-20% 1D SDS PAGE, stained with Coomassie Brilliant Blue and analyzed by densitometry analysis. The proteins were subjected to standard in-gel digestion and MS identification.

Figure 2.15 shows that, after washing with 0.4 M KCl, the major portion of LrpA was dissociated from the beads, as were other proteins, but still many background proteins remained on the beads. In contrast, after washing with 0.2 M KCl, the major portion of LrpA and background proteins still remained on beads.

So increasing ionic strength did not separate non-specific from sequence-specific DNA-binding proteins, but it can be used to fractionate the bead-bound protein mixture. A KCl concentration of 0.4 M was sufficient to elute some transcription factors (LrpA). A salt concentration lower than that should be used in eluting buffer if the purpose is to wash off non-specific DNA-binding proteins. 0.1 M KCl in standard washing can be safely employed.

2.2.6.6. Summary

DNA competitors, including sodium heparin, UOR fragments, poly-dIdC and various native genomic DNAs, have been studied and are capable in general of selectively releasing non-specific DNA-binding proteins from bead-bound DNA. This conclusion applies only to the proteins studied: LrpA, a known rTF, and unknown proteins that appear to act like non-specific DNA-binding proteins. In comparison, ionic strength did not selectively dissociate non-specific DNA-binding proteins. The high salt concentration (>0.4 M) was only useful for general protein fractionation, eluting both non-specific DNA-binding proteins as well as known transcription factors.

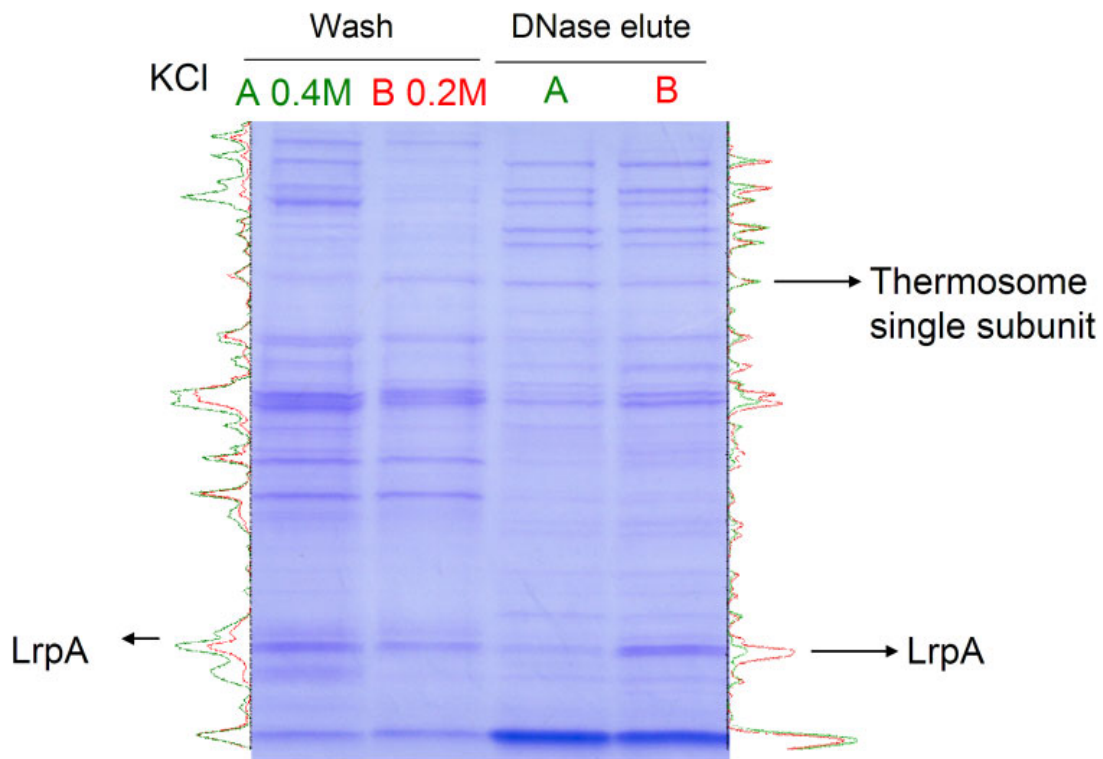


Figure 2.15 Protein elution using different salt concentrations. The first two gel lanes compare elutions with 0.4 M KCl (A, green) and 0.2 M KCl (B, red), while the second two lanes display proteins dissociated by DNase digestion. Protein LrpA and a thermosome subunit were identified by MS and are indicated by arrows. Transcription factor LrpA can be eluted by 0.4 M KCl, but is not eluted as much with 0.2 M KCl.

2.2.7. Influence of UOR Length on Affinity of Non-Specific DNA-Binding Proteins

Generally, the amount of DNA-bound proteins was proportional to the length of the UOR. Several things need to be considered about the length of UOR used in protein capture. First, the selection of UOR can be gene dependent based on known information about the transcription start site. For those genes for which only a translation start site is known, the UOR may have to involve a couple of hundred nucleotides before the gene translation start to cover possible transcription factor binding sites. But the UOR should not be too long, because the longer the UOR, the less DNA that can be immobilized on Dynabeads. More importantly, a longer UOR will probably involve more non-specific binding and lead to lower content of the rTFs in the protein mixture. We have studied UORs as long as 300 bp (-200 to +100) or as short as 72 bp (-68 to +3) and both led to successful identification of sequence-specific DNA-binding proteins.

2.2.8. Comparison of Methods for Complete Protein Dissociation

The last step of DNA-affinity protein capture is to dissociate all retained DNA-binding proteins from Dynabeads to provide a sample ready for subsequent analysis. We compared different methods for this last step.

2.2.8.1. Dissociation for Gel-based Analysis

The Dynabead-bound DNA-protein complexes can be eluted completely by incubating in Gel-loading buffer containing 2% SDS at 100 °C for 10 min (Chen, 2003). This method is simple and convenient for subsequent separation with SDS PAGE. But heating beads at boiling temperature can easily dissociate the streptavidin from the bead surface, which together with the broken pieces of beads may interfere with the separation of other proteins around the same molecular weight in gel electrophoresis. We tried reducing the SDS to as low as 0.1 % but observed the same problem. The high temperature is responsible for streptavidin dissociation.

Even without SDS, surface-coated streptavidin can be dissociated by 100 °C incubation for more than 5 min. So protein dissociation in Gel-loading buffer should be performed at lower temperature to avoid streptavidin interference. We found that the bead-bound DNA-protein complexes will dissociate completely by incubation in Gel-loading buffer at 55 °C for 10 min without any dissociation of streptavidin.

2.2.8.2. Dissociation for Subsequent LC/MS Analysis

We could use LC/MS for shotgun proteomics analysis of captured proteins (Domon *et al.*, 2006; Yates, 2004). For application of this method, we need to avoid SDS in the samples. So we studied other methods to dissociate the final DNA-binding proteins.

Turbo DNase successfully eluted most of the proteins from beads by digesting the bead-bound DNA without involving any detergent or dissociation of streptavidin. The DNase digestion takes 30 min and efficiently digests the DNA only when the reaction is mixed well during the incubation. The disadvantage of using DNase was that DNase protein was also digested by trypsin, generating background signals in the MS spectra, which interfered with the *Pf* peptide signals.

Another method studied was to insert a restriction site at the front of the UOR for cleavage by a restriction enzyme. In one experiment, a 64 bp UOR of PF0272 was used as a test. An EcoR I (GGAATTCC) restriction site was inserted at the 5'-biotinylated end of the UOR. The 72 bp dsDNA was prepared by oligonucleotide hybridization with single oligonucleotide strand. EcoR I was added up to 72 units to 750 ng bead-bound DNA and incubated up to 14 h. Digested DNA was quantified, but amounted to < 50% of the total DNA. We suspect that steric hindrance around the bead surface blocked access by EcoR I.

Rather than cleaving dsDNA, a parallel idea is to use a protease to cleave the proteins bound to the DNA. Trypsin is a protease which specifically cleaves after arginine and lysine residues in proteins. We performed a protein capture experiment following standard procedures and attempted protein digestion with trypsin at 37 °C overnight with a trypsin:protein mass ratio of 1:50. The supernatant was checked by SDS PAGE, and showed incomplete dissociation of proteins. This may have resulted from the beads precipitating during the overnight incubation. Trypsin on-bead digestion results can be improved by first heating the bead-bound complex at 100 °C before enzyme digestion to dissociate the DNA-bound proteins. This will produce background signals of streptavidin protein in the MS spectra, but since the streptavidin sequence is known, these background signals can be predicted and ignored.

2.2.8.3. Summary

For complete dissociation of all DNA-binding proteins, a Gel-loading buffer (2% SDS) combined with heating at moderate temperature (50-55 °C) is successful for gel-based analysis. For alternative analysis with direct LC/MS methods, the presence of SDS in the protein mixture needs to be avoided. In this case, either DNA digestion with Turbo DNase or protein digestion with trypsin was able to elute proteins. Trypsin on-bead digestion is a clean and quick method, adaptable to subsequent analysis with mass spectrometry. But elution of proteins was incomplete and needs to be improved.

2.3. DNA-affinity Protein Capture Protocol

Figure 2.16 schematically illustrates the detailed DNA-affinity protein capture protocol resulting from the experiments described in this chapter. Before the start of an experiment, frozen cell extract was thawed on ice, and fresh 100 mM DTT solution was prepared.

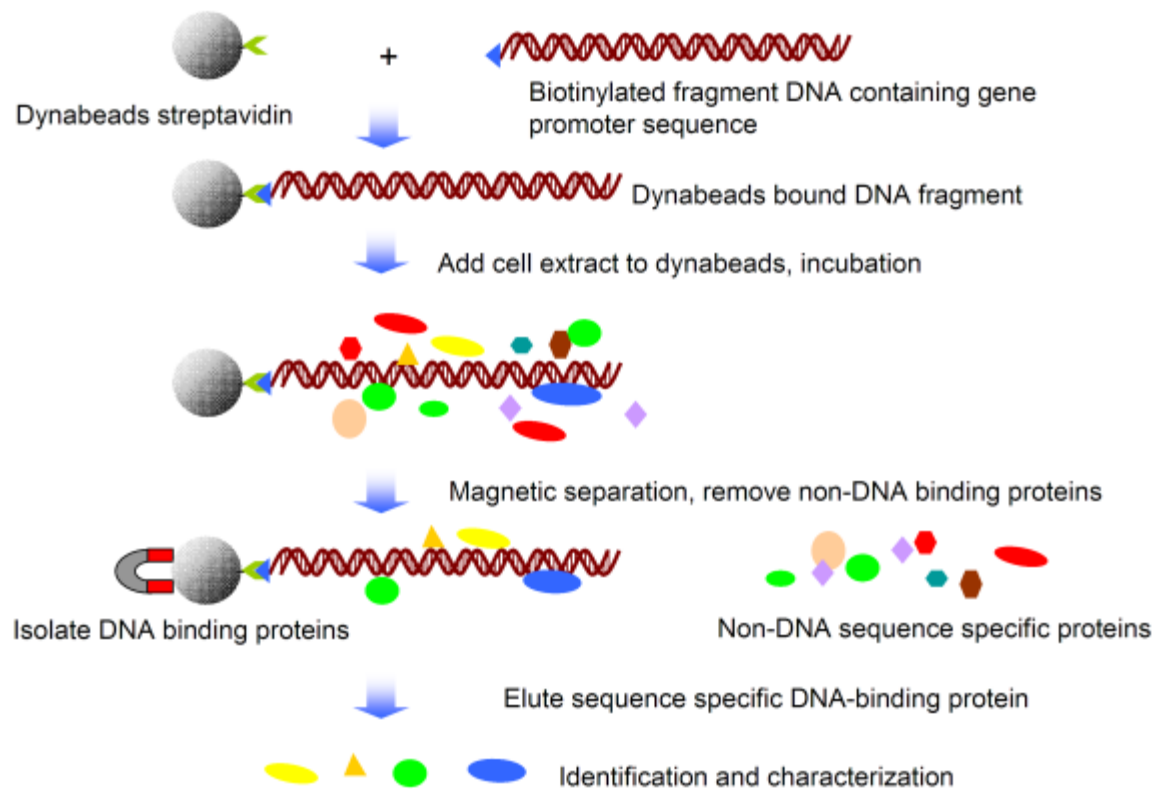


Figure 2.16 The schematic flow chart of DNA-affinity protein capture.

I. DNA Immobilization on DynaBeads

A. Dynabead Preparation

Place certain amount of Dynabeads M-280 (e.g., 50 μ L) into a 1.5-mL microcentrifuge tube and wash beads with 2x B&W buffer to replace the Dynabead storage buffer. The buffer volume used for washing is suggested to be twice the original Dynabead volume (100 μ L washing solution for 50 μ L Dynabeads). Washing beads means to suspend the Dynabeads in solution by gentle shaking or vortex mixing the tube, and placing the tube in a magnet. After the beads all migrate toward the magnet, aspirate the supernatant using a pipette. Repeat the process of washing beads one more time. Don't leave the Dynabeads drying in air for long.

(50 μ L of Dynabeads is enough for protein detection in gel by Coomassie Blue stain or silver stain. For in-gel digestion, 100 μ L of Dynabeads will be sufficient; if using more Dynabeads, also scale up the DNA, cell extract, washing solution, etc., in the following steps.)

B. DNA Immobilization

DNA is generally prepared by PCR amplification using 5'-biotinylated forward primers or by oligonucleotide hybridization with biotinylated forward strand DNA. For PCR amplified DNA, purification is necessary to separate PCR product from extra biotinylated primers. In oligonucleotide hybridization, apply more reverse-strand DNA than biotinylated forward-strand DNA.

Mix enough biotinylated dsDNA with beads in 1x B&W buffer. Keep shaking the beads gently on a vortex mixer to prevent beads forming clumps in solution. After 15 min (<1 kb DNA at room temperature), place the bead suspension in a magnet. Once the beads migrate to the magnet side and separation is complete (clear supernatant observed), aspirate the supernatant which contains extra DNA.

C. Wash Dynabeads

Dilute 2x B&W buffer with an equal amount of water, and use the 1x B&W buffer to wash Dynabeads two times. Avoid disturbing the beads by pipette tips in the process.

II. Protein-DNA Incubation

A. Equilibrate the bead-bound DNA by washing twice quickly with 1x Incubation buffer, 100 μ L for 50 μ L Dynabeads (each wash).

B. Incubate bead-bound DNA with cell extract in 1x Incubation buffer (containing 1 mM DTT) for 30 min at 70 °C in a heat block. The suggested protein concentration of the cell extract is 25-50 mg/ml. For 1 pmol DNA, 15-25 μ g protein is suggested. (If you do not have a good measure of the amount of DNA on your beads, use Table 1 to estimate the amount based on your DNA length.) During incubation, gently shake and suspend the beads every 2-3 min to avoid bead precipitation. After incubation, use a MPC magnet and a pipette to remove the supernatant containing extra cell extract (just like the operation to remove the extra DNA previously), also carefully avoid disturbing the beads which attach to the magnet).

C. After incubation, the Dynabead-bound DNA-protein complexes should undergo three quick washes with 1x Incubation buffer at room temperature. Resuspend the beads in washing solution completely, and make sure no bead aggregates remain at the bottom of the tube before the beads are placed on the magnet. It is a good idea to save the washing supernatant of each step for analysis with SDS PAGE.

III. Non-specific DNA-binding Protein Elution with Competitors

As an optional step, competitors (e.g., sodium heparin, poly-dIdC, or native genomic DNA) can be used (as detailed above) in an attempt to remove non-specific DNA-binding proteins.

IV. Complete Dissociation of DNA-binding Proteins

(If SDS PAGE will be used to separate the proteins, use step A; if direct LC-MS will be used, go to step B)

A. Gel-loading Buffer Elution

Add 50 μ L 1x Gel-loading buffer (per 50 μ L Dynabeads), and heat at 50-55 $^{\circ}$ C for 10 min to elute most DNA-binding proteins. After magnet separation, the supernatant solution containing the DNA-binding proteins should be heated at 100 $^{\circ}$ C for 1 min followed by analysis with SDS PAGE.

B. Elution by Turbo DNase Digestion

Turbo DNase digests immobilized DNA in 30 min at 37 $^{\circ}$ C. Typically, the reaction solution is composed of 4 μ L 10x Turbo DNase buffer, the required amount of DNase (2 units/ μ L, and 1-2 units DNase per μ g DNA), and Nanopure water to make a final volume of 40 μ L. Mix the beads frequently to avoid precipitation. Remove supernatant for gel analysis or LC-MS.

2.4. Capture of Native LrpA

2.4.1 Materials and Methods

80 μ L Dynabeads were used to immobilize 0.8 μ g *lrpA* UOR (300 bp) or PF0272 UOR (302 bp, from -223 to +79, relative to PF0272 translation start site, predicted by Dr. Ying Xu's group), the latter as a parallel control experiment. About 1.3 μ g of *Pf* cell extract was mixed with the bead-bound DNA and incubated at 70 $^{\circ}$ C for 30 min. Proteins were washed with the Incubation buffer as described previously.

Bead-bound DNA-protein complexes were digested with 14 units of Turbo DNase and eluted with Gel-loading buffer at 50 $^{\circ}$ C. The protein mixtures in supernatants were separated

from beads, mixed with 5 μ L β -mercaptoethanol, and heated to 100 °C for 10 min to denature the *Pf* proteins for subsequent SDS PAGE analysis. Protein mixtures were analyzed with 10-20% 1D SDS PAGE and the gel was stained overnight with 50 mL of GelCode Blue followed by destaining in water for at least 1 h. Densitometry was analyzed with Fluorchem 8000 (Alpha Innotech Corporation) in the Autogrid mode. In-gel digestion and mass spectrometry analysis were performed as described in 2.1.5.

2.4.2. Results

The proteins binding to *lrpA* UOR were separated by SDS PAGE (**Figure 2.17**). As shown by the gel densitometry, the *lrpA* UOR captures a different set of proteins than the control UOR. LrpA is a 15-kDa protein (arrow in Figure 17), and this gel band showed much stronger intensity with *lrpA* UOR than with the control UOR. Using MS, the protein band was identified unequivocally as LrpA based on a clear distinction of the first MASCOT score. The matched peptide masses covered 62% of the LrpA sequence. (The mass list was searched against a non-redundant genomic database of archaea species in NCBI by the machine-integrated software, Biotoool). The equivalent gel region of the control UOR protein capture was also excised and subjected to MS analysis, but LrpA was not identified in the ranked proteins. Therefore, the successful capture and identification of LrpA by the method of DNA-affinity protein capture coupled with mass spectrometry demonstrated the feasibility of using the method to identify new transcription factors.

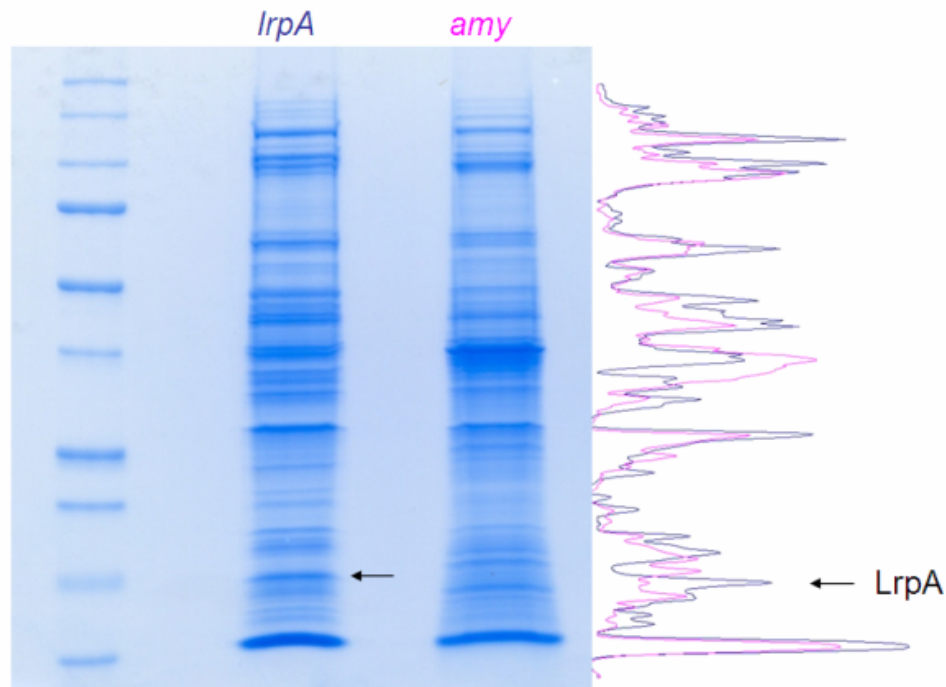


Figure 2.17 1D SDS PAGE separation of protein mixtures dissociated from *lrpA* UOR or control (PF0272) UOR. The gel was stained with GelCode Blue. The arrow on the densitometric scans indicates a protein that occurs at high intensity in the *lrpA* UOR lane (blue) compared to the control UOR (magenta). In-gel tryptic digestion and MS analysis identified this protein as LrpA. The equivalent region in the lane of PF0272 contained no identifiable LrpA.

CHAPTER 3

IDENTIFICATION OF UOR BINDING PROTEINS WITH MALDI PSD MS

3.1 Introduction

This chapter will illustrate the identification of discovered UOR-binding proteins (UBPs) in *Pf* using our STRES approach. Selected UORs correspond to genes involved in carbon source response (PF0272, PF0464) by genome-wide microarray study of gene expression profiles (Schut *et al.*, 2003). We also utilized bioinformatics analysis of gene promoter sequence (unpublished data from Ying Xu's group) to predict co-regulated genes.

PF0272, originally annotated as alpha-amylase, has recently been specified as 4- α -glucanotransferase, which is essential for cell growth on starch or maltose and as one of the enzymes responsible for metabolism of starch to generate glucose and glucose 1-phosphate (Lee *et al.*, 2006). PF0464, GAPOR, catalyzes the direct oxidation of glyceraldehyde-3-phosphate to 3-phosphoglycerate by using ferredoxin as an electron acceptor (Mukund *et al.*, 1995; van der Oost *et al.*, 1998).

Based on biochemical analyses of the purified proteins, PF0272 and PF0464 are both involved in starch conversion. The two genes were among the most up-regulated genes of this group under when maltose is used as the C source (Schut *et al.*, 2003). In addition, the motif-finding program CUBIC (Olman *et al.*, 2003) was used to identify repetitive sequences in the UORs of these glycolysis-related genes. From this study of the gene sequence, the UORs of both PF0272 and PF0464 and others contain a conserved motif comprised of a 5nt gapped palindromic repeat (TCACnnnnnGTGA), a potential binding site for a common regulator for

multiple genes. Therefore PF0272 and PF0464 were selected as the primary candidate genes for the discovery of common rTF(s).

3.1.1 Choice of Separation/Identification Methods

3.1.1.1 Protein Separation by Gel Electrophoresis

Protein gel electrophoresis coupled with proteomics mass spectrometry has been widely used in protein separation and identification. Both 1D and 2D gel electrophoresis are capable of resolving protein mixtures but both have advantages and disadvantages. 1D gel electrophoresis is fast (a couple of hours), convenient, and versatile in analysis of less complicated protein mixtures. 2D gel electrophoresis is able to resolve more complicated mixtures, and also can be used for quantitative analysis (Fujii *et al.*, 2006; Weeks *et al.*, 2005). However, a complete analysis by 2D gel electrophoresis usually takes a long time (several days). In our work, the complexity of the protein mixture was reduced greatly by the method of DNA-affinity protein capture. Therefore, we employ 1D gel electrophoresis in conjunction with current mass spectrometry techniques to identify the DNA-captured proteins.

The protein SDS PAGE gels can be stained by various methods, using silver nitrate, CBB, SYPRO Ruby, etc. Different gel-staining methods bear their own advantages for different purposes. Both detection sensitivity and influence on subsequent MS analyses are of concern for transcription factor identification. Silver nitrate staining is supersensitive and is ideal for detecting proteins in low amount. A drawback for silver staining is that not only proteins, but also oligonucleotides are stained, which could result in a smeared background or interference with protein detection. Moreover, samples stained by silver nitrate are difficult to recover and clean for mass spectrometry analysis. SYPRO Ruby has been used to stain 1D or 2D protein gels for wide dynamic range and good sensitivity, but it is expensive and inconvenient, since gel

excision requires fluorescent visualization. CBB is quick, inexpensive and the stain is easily removed from the sample, which is useful for MS analysis. There are two kinds of CBB dyes used in protein staining: R250 and G250. CBB-R250 is of lower sensitivity, and only suitable for detecting a relatively high amount of protein. Compared with the poor sensitivity of CBB-R250, CBB-G250 demonstrates higher sensitivity, ideal adaptability for subsequent sample preparation and wide dynamic range. So in this study, protein gels subjected to MS analysis were stained with CBB-G250 or staining reagent based on CBB-G250 (e.g., Gel-code Blue staining reagent from Pierce Inc.).

3.1.2. Principles of Peptide Mass Mapping

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 (Perkins *et al.*, 1999). 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, PMM is often sufficient for accurate protein identification according to statistical scores generated by the algorithm. When PMM 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. One can also use a peptide fragmentation technique such as MS/MS and post-source decay (PSD) to obtain protein sequence information. By integrating PMM with sequence analysis, the confidence of protein identification will be highly improved and the identification ambiguity will be reduced.

3.1.3. MALDI Post-Source Decay

PSD has been applied as a technique of peptide fragmentation for sequence analysis (Spengler *et al.*, 1992). Used in combination with standard PMM analysis, molecules have been successfully identified by a combination of PMM and PSD-MALDI-MS (Chaurand *et al.*, 1999; Gevaert *et al.*, 2001; Jungblut *et al.*, 1997). In our study, PSD was also employed for specific protein identification using a MALDI TOF mass spectrometer with PSD capability (Autoflex, Bruker).

In MALDI analysis, ions can gain high internal energy in the process of ion desorption. When flying in the first field-free drift path toward the linear detector, some metastable ions decay into fragments due to the high internal energy or collisions with the cloud of neutral molecules. The precursor ions and product ions still carry the same speed. When a reflection voltage is applied, the ions will be directed to the reflective direction, while the precursor and product ions are separated according to the molecular kinetic energy which is a function of the

ion mass (Spengler *et al.*, 1992). The smaller fragment ions will be reflected before the bigger ions and detected earlier by the reflective detector.

Peptide sequence information is obtained by MS measurement of the fragment ions produced by a certain peptide isolated by the ion gate during MS analysis. Peptide breakage can happen at C-C, C-N and N-C backbone bonds and produce a-, b-, c- (N-terminal ions) ions or x-, y-, z- (C-terminal ions) ions respectively. Most frequently observed fragment ions produced in PSD are a-, b-, and y- ions. The observed fragment ion mass list is compared with the peptide sequence derived from the database, and used for peptide/protein identification.

3.2. Materials and Methods

3.2.1. DNA-affinity Protein Capture with UOR PF0272 and PF0464

3.2.1.1. DNA Preparation

UOR DNA

The UOR DNA used for PF0272 (α -amylase) spanned bases -61 to +3 (relative to translation start site) plus a biotinylated linker of 8 base-pairs at the 5' end (**Figure 3.1a**). UOR DNA of PF0464 (GAPOR) spanned -69 to +3 (relative to the translation start site) (**Figure 3.2a**). Based on the bioinformatics analysis described in the Introduction, the UOR contained TATA Box element, BRE, (a promoter), a predicted 14 bp palindromic conserved motif, and the ribosome binding site (RBS) (unpublished data from Dr. Ying Xu's group).

Control DNA

Control DNA was applied in parallel experiment with the correlated UOR DNA for comparison purposes. For gene PF0272, the control DNA (control0272) contained the same sequence of the UOR PF0272 except 23 nucleotides (from the conserved motif to the predicted RBS) were replaced with random sequence preserving the same GC percentage (**Figure 3.1b**).

a.

5' biotin- GGAATTCCATTTTTGAGCCATTAAATATATAAGGAAGTATCACTCTT
CCTTAAGGTAAAACTCGGTAATTTATATATTCCTTCATAGTGAGAA

AGTGATTAATGGGTGGACGGAAGTG
TCACTAATTACCCACCTGCCTTCAC

b.

5' biotin- GGAATTCCATTTTTGAGCCATTAAATATATAAGGAAGTACTTACACT
CCTTAAGGTAAAACTCGGTAATTTATATATTCCTTCATGAATGTGA

TTGTGAATATGTCGCGACGGAAGTG
AACACTTATACAGCGCTGCCTTCAC

Figure 3.1 Sequence of UOR and control DNA of gene PF0272 **a.** Sequence of promoter DNA of PF0272. In blue font is the 8 bp linker; red font is the predicted TATA box; green font is the predicted conserved palindrome; underlined font is the predicted RBS; purple shade indicates sequence that was randomized in control DNA. **b.** Sequence of control DNA of PF0272, Orange shade, random nucleotides replacing the conserved motif and RBS in (a). The control DNA was made by mutating the conserved sequence (the palindrome and RBS) in PF0272 UOR with random nucleotide sequence while keeping the same GC percentage as that in the promoter DNA.

a.

5' biotin-ATG**CAAAAATTTTAA**GTAAGTTAAATGAAATCACGT**CACTG**
 TACGTTTTTAAAAATTCATTCAATTTACTTTAGTGCAGTGAC

GTAGTGGTTATAATCGAGGTGATGACGTATG
 CATCACCATATTAGCTCCACTACTGCATAC

b.



c.

5' biotin-ACATATGAAGTTGATCCCTACGATCCAAGGAATATAGTTATC
 TGTATACTTCAACTAGGGATGCTAGGTTTCCTTATATCAATAG

TTTGGAAATAGGACCTTTCGCAGGTTCTGTT
 AAACCTTATCCTGGAAAGCGTCCAAGACAA

Figure 3.2 Sequence of UOR and control DNA of gene PF0464. **a.** Sequence of PF0464 UOR, corresponding to the purple shade region in (b), consisting of 72 bp (-69 to +3, relative to translation start site), containing predicted TATA box (red); predicted 14-bp imperfect palindromic motif (green); ribosome binding site (underlined) and translation start (italics). **b.** Diagram of relationship of a and c. **c.** Sequence of control DNA of PF0464, corresponding to the green shade region in (b) consisting of 72 bp (+121 to +192, relative to translation start).

For UOR PF0464, the control DNA (control0464) for PF0464 was a 72 bp-fragment DNA from PF0464 ORF region (from +121 to +192, relative to the translation start) (**Figure 3.2b**).

Oligonucleotide Hybridization

The double-stranded UOR and control DNA were prepared by oligonucleotide hybridization. Each forward (top) ssDNA was biotinylated at the 5' end. 10 μ L of 60 mM forward ssDNA was mixed with 10 μ L of 70 mM of the complementary reverse ssDNA in a 0.5 mL micro centrifuge tube. The mixture was heated at 94 °C for 2 min, and cooled to room temperature over about 1 h. The hybridized dsDNA was checked with 2% agarose gel and quantified using DyNA Quant 200 Fluorometer with calf thymus DNA as standard.

3.2.1.2. Preparation of *Pf* Cell Extract

Pf cells used for STRES based on carbon source were grown under the same conditions with maltose/peptides described previously in chapter 1, and collection of cell extract followed the same protocol as that used in the correlated microarray gene expression study (Schut *et al.*, 2003).

3.2.1.3. Protein Capture by PF0272 and PF0464 UORs

The DNA-affinity protein capture experiments were performed following the procedures described in Chapter 2. For each UOR, two control experiments were performed in parallel: first, a “DNA control” experiment using the correlated control DNA (3.2.1.1) with extracts from the same cell growth condition; second, a “condition control” experiment with the UOR using extracts from different cell growth conditions.

The Dynabead-bound DNA complexes were mixed with cell extract diluted to a protein concentration of 7.6 μ g/ μ L. The bead-bound DNA-protein complexes were washed three times

with standard Washing buffer. The DNA binding proteins were dissociated by incubation at 50 °C in 1x Gel-loading buffer for 10 min. The protein mixture was analyzed using 10-20% gradient 1D SDS PAGE.

3.2.2. In-gel Digestion

Protein in-gel digestion was performed as described in Chapter 2. 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_μ-C18 before deposition with matrix (α -cyano-4-hydroxycinnamic acid) on a MALDI target plate.

3.2.3. Sample Analysis with MALDI TOF MS

The sample was analyzed by MALDI TOF MS. 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.

3.2.4. Sample Analysis with MALDI PSD

The precursor ions were selected from the first dimension mass spectrum of tryptic peptides. Only peptides with no other ions within the range of ± 5 mass units of the precursor ion were selected from the mass spectrum to avoid fragmentation of mixtures. So the precursor

ion window width was set to ± 5 Da. Because a single reflective voltage is not enough to cover the whole mass range, the MALDI-PSD was performed in the step mode to collect data at multiple mass segments. After data collection was completed, the data from all mass segment ranges were combined.

Data were processed by “subtract baseline” and “smooth spectra”, and a peak list was generated by the machine software automatically. The PSD-MALDI-MS spectra were interpreted with Biotoool software. Peptide sequence information and one dimensional MS data were combined to search against the database in searching mode of MS/MS in MASCOT.

3.3. Results and Conclusions

3.3.1. Identification of UOR-binding Proteins with Peptide Mass Mapping

The protein mixtures dissociated from DNA were separated by 1D SDS PAGE. The results of gel separation and protein quantitative analysis by densitometry are shown in **Figure 3.3** (for UOR PF0464) and **Figure 3.4** (for UOR PF0272). Using densitometry, protein intensity ratios between UOR and control DNA (e.g. Up/Cp for growth on peptides or Um/Cm for growth on maltose) between two cell conditions (Up/Um) were compared. The DNA-binding proteins were identified by MS and summarized in **Table 3.1** and **Table 3.2**.

After comparing the intensities of the DNA-bound proteins between the control DNA and the UOR in the same cell condition (i.e. Up/Cp or Um/Cm), the proteins showing higher intensity to PF0464 UOR were identified with the method of PMM. These proteins are the candidates of specific UOR-binding proteins. Band 5 was identified as a mixture of two proteins, PF0250 and PF0248. The MASCOT score of PF0248 was lower than the threshold score (60), but as the second candidate in the identification list, the amino acid sequence coverage is high and the molecular weight matches the information provided from the gel. Therefore, it was

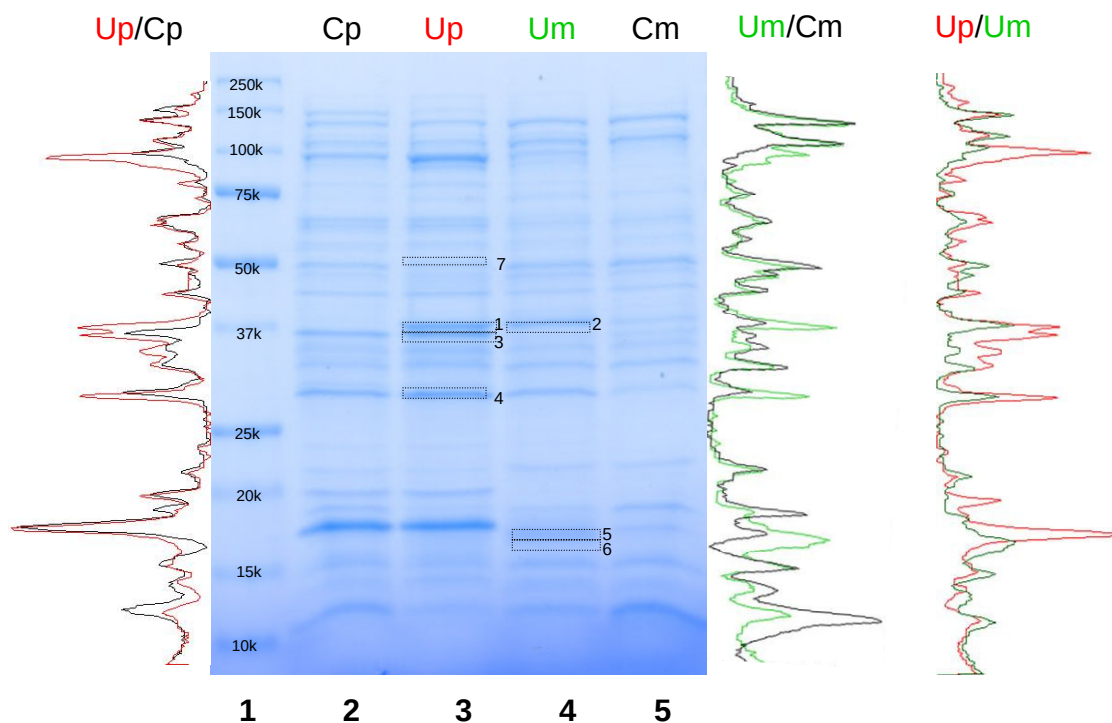


Figure 3.3 Separation of UBP mixtures dissociated from PF0464 UOR and control DNA by 1D SDS PAGE. Lane 1, protein marker; Lane 2, the proteins eluted from control DNA with cell extract of peptide condition (Cp); Lane 3, proteins from PF0464 UOR with peptide condition (Up, red); Lane 4, proteins from UOR with maltose condition (Um, green); Lane 5, proteins from control DNA with maltose condition (Cm). The abundance of proteins correlated to different DNA or cell condition was analyzed by gel densitometry. This analysis allowed selection of proteins captured selectively between UOR DNA and control DNA in peptide condition (Up/Cp, red/black) and in maltose condition (Um/Cm, green/black). The gel bands (numbered 1-7) showed higher intensity on UOR DNA than on the control DNA and were subjected to MS identification.

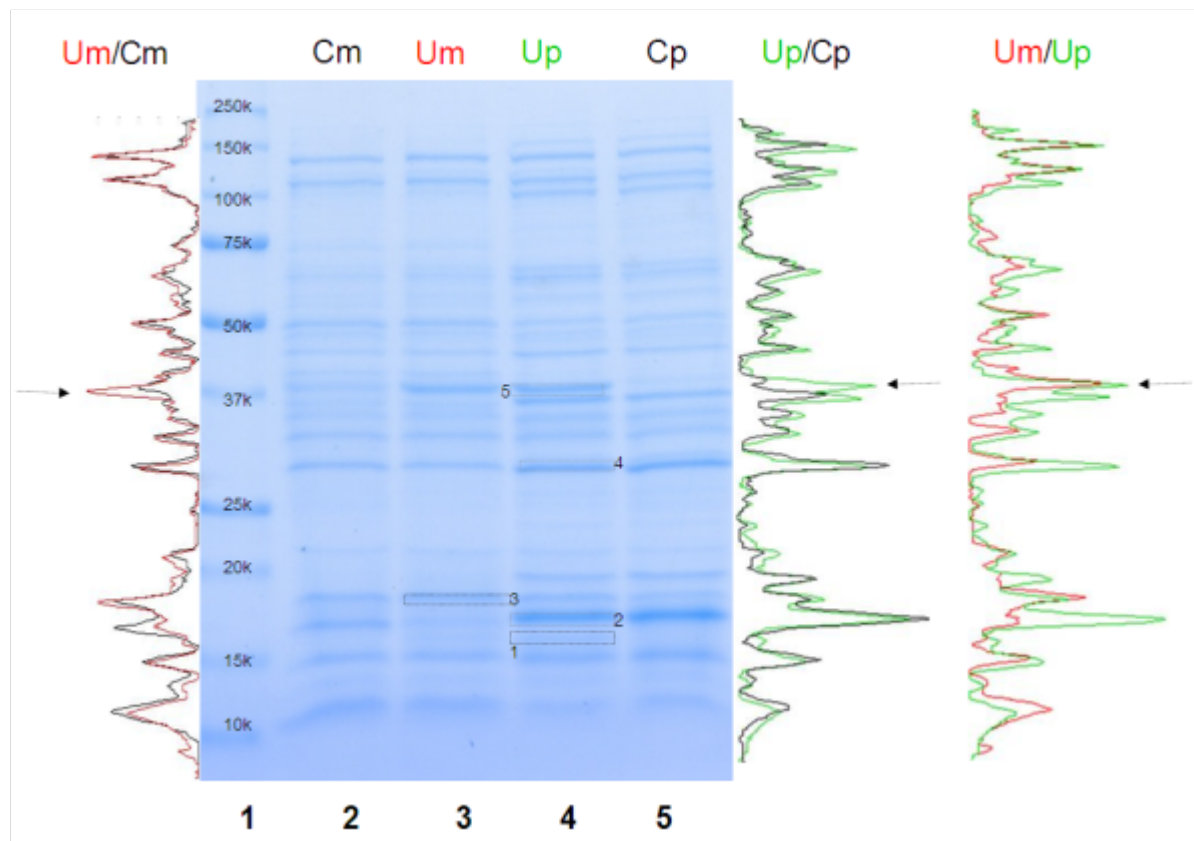


Figure 3.4 Separation of UBP mixtures dissociated from PF0272 UOR and control DNA by 1D SDS PAGE. Lane 1, protein mass standard; Lane 2, control DNA with maltose condition (Cm, black); Lane 3, UOR with maltose condition (Um, red); Lane 4, UOR with peptide condition (Up, green); Lane 5, control DNA with peptide condition (Cp, black). The abundance of proteins correlated to different DNA or cell condition was analyzed by gel densitometry. This analysis allowed selection of proteins captured selectively between UOR and control DNA in maltose condition (Um/Cm, red/black) and in peptide condition (Up/Cp, green/black). The gel bands (numbered 1-5) showed higher intensity on UOR DNA than on the control DNA and were subjected to MS identification. Arrow indicates position of PF0124.

Table 3.1 Protein identification from UOR PF0464

Gel Band	ORF #	Annotation ^a	MW(kDa)/pI	Score ^b	Seq% ^c
1	PF0124	Hypothetical protein	39.4/7.12	112	34%
2	PF0124	Hypothetical protein	39.4/7.12	98	48%
3	PF0062	Riboflavin-specific deaminase	39.7/9.59	187	66%
4	PF1931	Hypothetical protein	27.7/8.56	139	68.4%
5	PF0250	Transcriptional regulatory protein, lrp family	18.0/9.2	65	44%
	PF0248	Hypothetical protein	17.7/5.8	51	59%
6	PF0250	Transcriptional regulatory protein, lrp family	18.0/9.2	70	49.3%
7	PF1725	DNA primase	50.2/8.53	70	40%

^aThe protein annotations are based on the information from NCBI database.

^bMASCOT score. A score over 60 indicates a significant match.

^cSeq% is the sequence coverage of matched peptides from the identified protein

Table 3.2 Protein identification from UOR PF0272 (**Figure 3.4**)

Gel Band	ORF #	Annotation ^a	MW(kDa)/PI	Score ^b	Seq% ^c
1	*PF0322	Hypothetical protein	10.4/8.0	37	39%
2	PF0063	6,7-Dimethy-8-ribityllurazine synthase	17.5/6.9	155	70.1%
3	PF2053	Hypothetical protein	17.2/9.5	100	68%
4	PF1931	Hypothetical protein	27.7/8.5	187	89%
5	PF0124	Hypothetical protein	39.4/7.1	227	57%

^aThe protein annotations are based on the information from NCBI database.

^bMASCOT score. A score over 60 indicates a significant match.

^cSeq% is the sequence coverage of matched peptides from the identified protein.

*Combining with peptide sequence information from PSD, the score of PF0322 was improved to 103.

considered as a positive identification. Proteins selectively binding UOR DNA are PF0124, PF0250 and PF0248 (from bands 1, 2, 5, 6). PF0124 is present at higher level in peptide condition compared to maltose condition (see Up/Um in **Figure 3.3** and **3.4**). PF0250 and PF0248 were only captured by UOR in the maltose condition (band 5, 6 in Um). Protein PF1931 (band 4) was correlated to both UOR and control DNA, but showed higher intensity with UOR DNA for both cell conditions. Meanwhile, the protein in the nearby or counterpart regions of band 1 in the control gel was identified as the protein PF0062 (#3) as shown in **Figure 3.3**.

From the gel densitometry analysis, most of the proteins were of equal intensity on UOR and control DNA, except for protein PF0124, which clearly bound UOR selectively.

3.3.2. Protein Identification by Combination of MALDI MS and PSD

Identification of protein PF0124

As indicated in **Table 3.1** and **3.2**, UBP PF0124 was identified from UOR of both PF0464 and PF0272. In the peptide mass mapping analysis by MASCOT, it was ranked first and was the unique protein with identification score higher than the threshold score of 60.

To increase the confidence of the identification, the technique of MALDI PSD was applied to peptide fragmentation of the two most intense peptides, with m/z of 967 and m/z of 2053. Both peptides were identified as trypsin digest peptides from protein PF0124 (**Figure 3.5**). The trypsin digest data (PMM) were combined with the peptide sequence data (PSD) to search against the NCBI archaeal genome database with the mode of “search with combined mass spectra”, and improved the resulting MASCOT score to 208 as the unique matched protein from the database.

Figure 3.5a is a one-dimensional mass spectrum of trypsin digest of gel-extracted protein, which was identified as protein PF0124 by PMM. PSD sequencing was applied to peptide mass

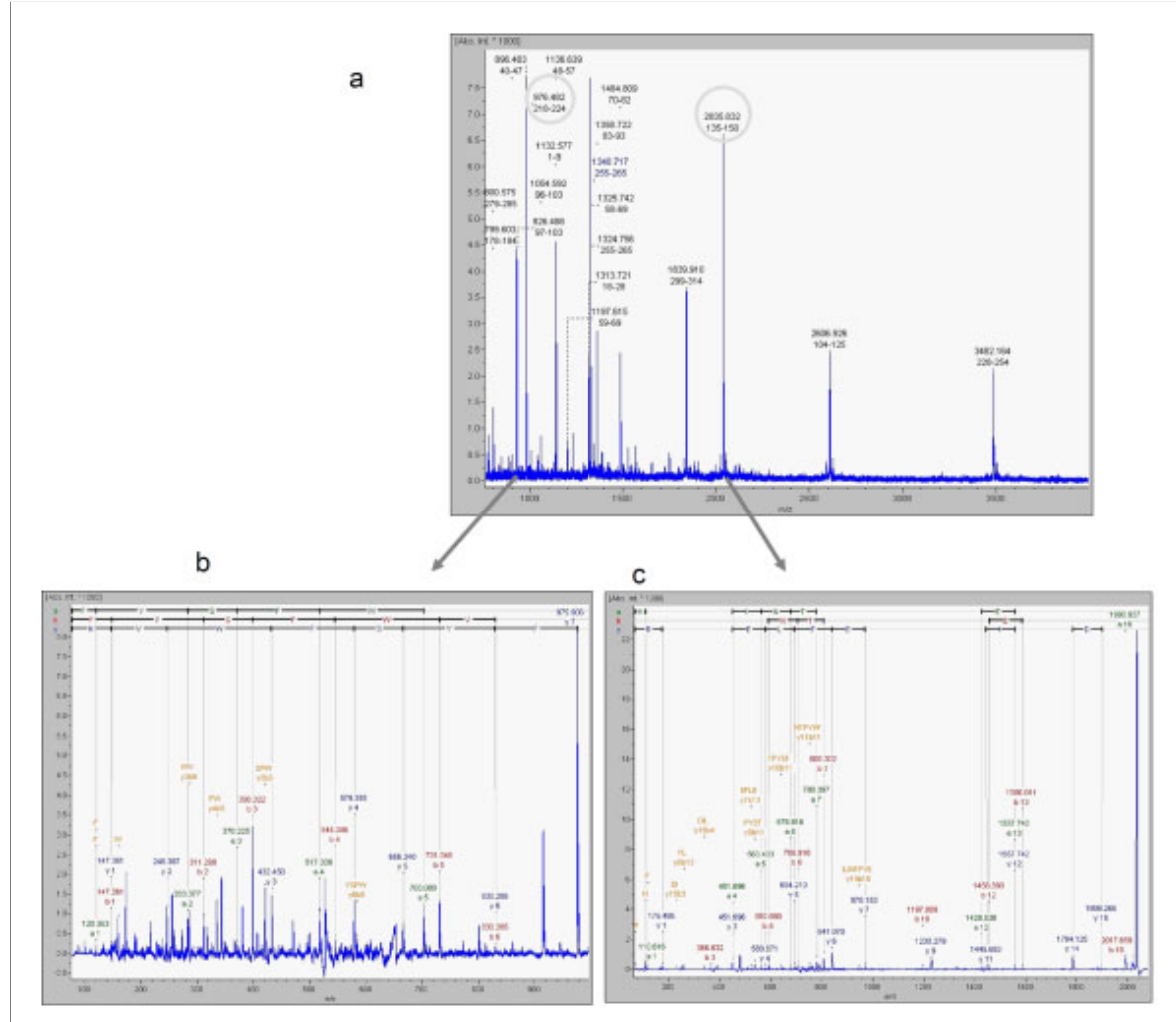


Figure 3.5 Mass spectra for the identification of protein PF0124. **a.** MALDI-TOF spectrum of gel-extracted tryptic peptides of band 1 in **Figure 3.3**. PMF using the peaks listed in the spectrum identified the protein as PF0124. **b.** PSD-MALDI-MS spectrum of peak 976.5. The precursor peptide was interpreted by PSD as FYSFWVK from PF0124. **c.** PSD-MALDI-MS spectrum of peak 2035.8. The precursor peptide was interpreted by PSD as HDILNTPYEFLEYLR from PF0124.

peaks of 976.5 and 2035.8, which were identified as peptide FYSFWVK (**Figure 3.5b, Table 3.3**) and HDILINTPYEFLEYLR (**Figure 3.5c, Table 3.4**), respectively. PF0124 was identified based solely on PSD sequencing of peptides.

The PF0124 protein was identified uniquely from the genome database with highly improved MASCOT score using peptide mass fingerprinting in combination with peptide PSD sequencing information. The summary of the PF0124 identification result is shown in **Figure 3.6**.

3.3.3. Bioinformatics analysis of identified UOR binding proteins

3.3.3.1. PF0124 Binding UOR of both PF0272 and PF0464

The sequence of protein PF0124 was aligned with all the sequences in NCBI non-redundant database, as well as the conserved domains in the Conserved Domain Database using the web-based NCBI BLAST program. The alignment result (**Figure 3.7**) indicated protein PF0124 is homologous to the TrmB protein with about 30% sequence identity at the protein N-terminus. The N-terminus also shows the characteristics of a helix-turn-helix (HTH) DNA-binding domain. The TrmB family was characterized as sugar-specific transcriptional regulators. One member of this family, TrmB has been shown to be a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter in *Thermococcus litoralis*. Other alignment results of the identified hypothetical proteins are summarized in the **Table 3.5**.

3.4. Discussion

3.4.1. Protein PF0124

As shown by the gel densitometry analysis, a higher quantity of PF0124 was captured from cell extract of peptide growth than from maltose growth by UORs of PF0272 and PF0464.

The identification of protein PF0124 as a specific UOR binding protein indicated a possible role of PF0124 as a common regulatory transcription factor for multiple genes,

Table 3.3 Fragment ions of peptide HDILINTPYEFLEYLR from peak 2035.8.

Predicted masses are listed and the matched ones are colored in red. Immonium

(Im), a, b, y ions are as described in 3.1.3. b* ions are b ions losing NH₃.

#	Im	a	b	b*	Seq	y	#
1	110.0713	110.0713	138.0662		H		16
2	88.0393	225.0982	253.0931		D	1898.9898	15
3	86.0964	338.1823	366.1772		I	1783.9629	14
4	86.0964	451.2663	479.2612		L	1670.8788	13
5	86.0964	564.3504	592.3453		I	1557.7948	12
6	87.0553	678.3933	706.3882	689.3617	N	1444.7107	11
7	74.0600	779.4410	807.4359	790.4094	T	1330.6678	10
8	70.0651	876.4938	904.4887	887.4621	P	1229.6201	9
9	136.0757	1039.5571	1067.5520	1050.5255	Y	1132.5673	8
10	102.0550	1168.5997	1196.5946	1179.5680	E	969.5040	7
11	120.0808	1315.6681	1343.6630	1326.6365	F	840.4614	6
12	86.0964	1428.7522	1456.7471	1439.7205	L	693.3930	5
13	102.0550	1557.7947	1585.7897	1568.7631	E	580.3089	4
14	136.0757	1720.8581	1748.8530	1731.8264	Y	451.2663	3
15	86.0964	1833.9421	1861.9370	1844.9105	L	288.2030	2
16	129.1135				R	175.1190	

Table 3.4 Fragment ions of peptide FYSFWVK from peak 976.5. Predicted masses are listed and the matched ones are colored in red. a^0 , b^0 ions are a, b ions losing H_2O .

#	Im	a	b	b^0	Seq.	y	#
1	120.0808	120.0808	148.0757		F		7
2	136.0757	283.1441	311.1390		Y	829.4243	6
3	60.0444	370.1761	398.1710	380.1605	S	666.3610	5
4	120.0808	517.2445	545.2395	527.2289	F	579.3289	4
5	159.0917	703.3239	731.3188	713.3082	W	432.2605	3
6	72.0808	802.3923	830.3872	812.3766	V	246.1812	2
7	101.1073				K	147.1128	1

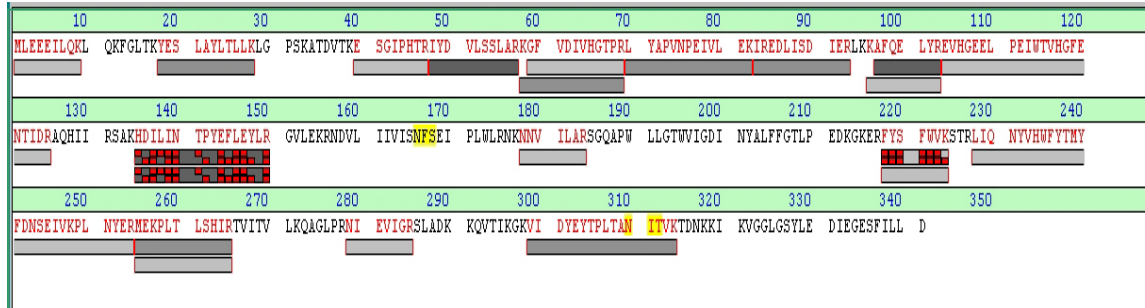


Figure 3.6 The chart shows protein PF0124 identification by combination of peptide mass fingerprinting and peptide fragment sequence information. The amino acid sequence coverage of the identified protein was generated by Biotoool software. The grey bars indicate the identified peptides and red squares indicated which peptides were fragmented with PSD.

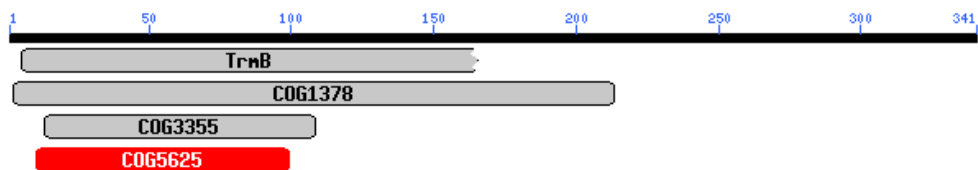


Figure 3.7 The conserved domain BLAST search result of PF0124. The black line represents the query sequence (PF0124) with scale showing residue numbers; the blocks below indicate the conserved domains found. Domains within similar functional group are colored the same. COG1378 and COG3355: predicted transcriptional regulators; COG5625: HTH domain.

Table 3.5 Summary of the sequence alignment results of the identified proteins.

ORF	BLAST, CD results	UOR affinity
PF0250	Transcription regulatory protein, lrp family; AsnC family an autogenously regulated activator of asparagine synthetase A transcription in <i>Escherichia coli</i>	Only present with UOR of PF0464 in peptide condition
PF0248	COG 1839, Uncharacterized conserved protein	Only present with UOR PF0464 in peptide condition
PF1725	DNA primase; Containing the topoisomerase- primase (TORPIM) nucleotidyl transferase/hydrolase domain	Only present with UOR PF0464 in peptide condition
PF2053	Lrp transcription regulators	Shown on control DNA and UOR of PF0272
PF0322	Predicted transcription regulators containing DNA binding domain and metal binding domain	Binding to PF0272 UOR in both cell conditions

especially those involved in glycolysis. Therefore, further investigation of this protein interacting with DNA and functional analysis in transcription will be necessary to verify the role of the protein.

3.4.2. Other identified UOR Binding Proteins

Several UOR-binding proteins were found as hypothetical proteins containing conserved transcriptional regulator domains (e.g., PF2053, PF0322, PF0250 and PF0248). However, PF2053 also bound to control DNA and was found to be associated with LrpA (PF1601) UOR DNA as well. Therefore PF2053 possibly non-specifically associates with DNA. PF0322 was only identified using PF0272 UOR, while PF0250 and PF0248 were only present with PF0464 UOR, indicating the possible specific DNA-binding properties of these proteins. Although they are potential transcription factor candidates, there is no evidence that these proteins are common transcriptional regulators for both PF0272 and PF0464. Proteins showing different intensities on UORs in different cell extracts are not necessarily regulatory proteins since the protein bound to the control DNA at the same level as with UOR. Therefore the control DNA is helpful for selection of candidate rTFs.

3.4.3. PSD

The hypothetical protein PF0124 from protein capture was identified with mass spectrometry analysis. The combination of peptide mass mapping and peptide sequencing using PSD allowed unique identification of protein from archaeal genomic database.

PSD has some limitations concerning low mass accuracy and low resolution. Due to relatively weak or insufficient fragmentation, only peaks with the highest intensity provide sufficient fragmentation information. Despite the limitations, PSD is still a very useful and complementary method for PMM.

CHAPTER 4

CHARACTERIZATION OF NEW REGULATORY TRANSCRIPTION FACTOR, PF0124

4.1. Introduction

Protein PF0124 (PF0124p) is annotated as a hypothetical protein with a calculated molecular weight of 34.8 kDa and pI of 7.1 in the genome database. As described in Chapter 3, PF0124p selectively binds to the UOR of both PF0272 and PF0464, which are genes involved in glycolysis and up-regulated during *Pf* growth on maltose. Protein sequence alignments indicate that PF0124p is homologous to protein TrmB, which is a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter in *Thermococcus litoralis* (Lee *et al.*, 2003; Lee *et al.*, 2005). The PF0124p sequence contains a helix-turn-helix DNA-binding domain at the N-terminus and shares about 30% sequence identity with protein TrmB. Based on the sequence homology and the STRES identification of PF0124p, we propose that PF0124p is a potential transcriptional regulator of multiple genes involved in glycolysis, which is worthy of further biochemical analysis.

The conserved motif TCACnnnnnGTGA (n represents any nucleotide), present in the PF0272 and PF0464 UORs, was found upstream of a group of genes that are also involved in *P. furiosus* glycolytic pathways and up-regulated by maltose, including PF1959 and PF1920. PF1959 codes for one of the phosphoglycerate mutases in archaea, catalyzing the interconversion of 3-phosphoglycerate and 2-phosphoglycerate and functioning independently of cofactors (van der Oost *et al.*, 2002). PF1920 codes for a triosephosphate isomerase, an enzyme that is well conserved among all three domains (Ronimus *et al.*, 2003). If PF0124p does have a regulatory

function on gene transcription by interacting with the conserved motif, it would be a common regulator of multiple genes and play an important role in archaeal glycolytic pathways. To confirm the interaction between PF0124p and the predicted conserved motif in UORs of these glycolytic genes, and to investigate the function of PF0124p on gene transcription, PF0124 was overexpressed, PF0124p was purified, and biophysical analyses of protein-DNA interactions by electrophoretic mobility shift assay (EMSA) and footprinting were performed.

4.2. Materials and Methods

4.2.1. Materials

4.2.1.1. DNA Sequences

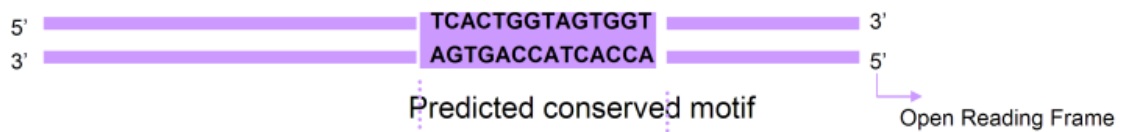
DNA for EMSA

UOR and Control DNA of PF0272 and PF0464 PF0272 UOR/control DNA and PF0464 UOR/ORF control DNA used in protein capture described in Chapter 3 were also employed for EMSA study. A new control DNA for PF0464 was applied as well to study the role of the conserved palindromic motif in protein-DNA interaction. The new control DNA was made by randomizing the sequence of the conserved motif (14 bp) in the PF0464 UOR while the GC content was kept the same. The PF0464 random control DNA is shown in **Figure 4.1**.

PF1959 UOR The UORs of PF1959 and PF1920 were used in EMSA and prepared by oligonucleotide hybridization (method was described in Chapter 3). PF1959 UOR from -47 to +3 (relative to translation start) contains TATA box and the palindromic repeated motif, as shown in **Figure 4.2**.

PF1920 UOR PF1920 UOR from -48 to +3 (relative to translation start site) contains TATA box and the palindromic repeated motif as shown in **Figure 4.3**.

PF0464 UOR



PF0464 random control DNA

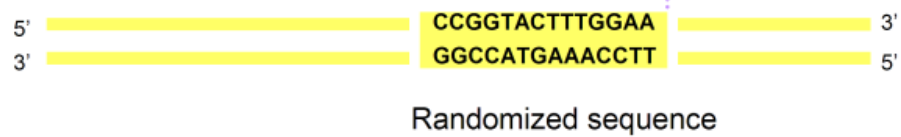


Figure 4.1 Schematic of PF0464 UOR and random control DNA. PF0464 control DNA (yellow), 72 bp, was made according to the UOR (purple) by replacing the conserved motif (14 bp) with randomized nucleotides with the same GC percentage.

PF1959 phosphoglycerate mutase

5'- CTTAAAGCTTATATTTTCCTTCTTCCACTCACTGTAGGTGATTAAAA***ATG***-3'
3'- GAATTTTCGAATATAAAAGGAAGAAGGTGAGTGACATCCACTAATTTTAC-5'

Figure 4.2 Sequence of PF1959 UOR, 50 bp from -47 to +3 (relative to translation start site), containing TATA box (red); conserved motif (underlined) comprising a palindromic repeated sequence (magenta); and translation start site (bold and italic).

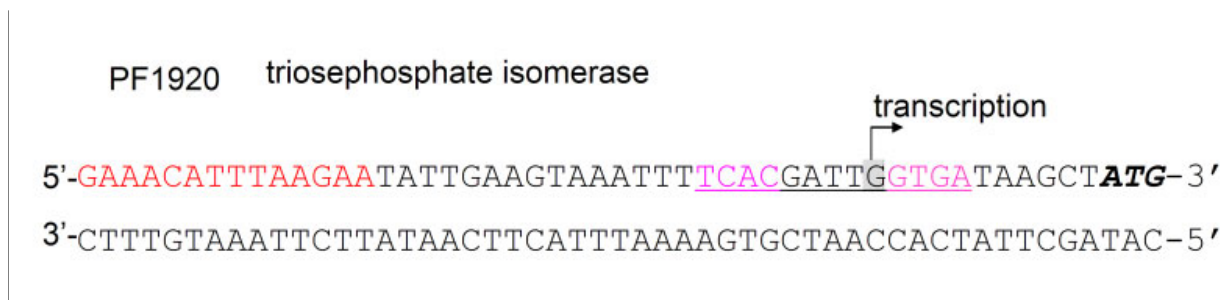


Figure 4.3 Sequence of PF1920 UOR, 51 bp from -48 to +3 (relative to translation start site), containing TATA box (red); 4-5-4 palindromic repeated sequence (magenta and underlined); translation start site (bold and italic); and predicted transcription site (grey) predicted based on CUBIC analysis.

DNA for Footprinting and *in vitro* Transcription

PF0272 Probe The PF0272 footprinting probe was 248-bp from -169 to +79 (relative to translation start) amplified by PCR with fluorescently labeled primers. The sequence of the 5'-6FAM-labeled forward PCR primer is 5'-GGCAAGCACTAGAAGGACTA-3' and the 5'-HEX-labeled reverse PCR primer is 5'-CCTCCTCAAACACCCATCCAAA-3'. The resulting DNA sequence is shown in **Figure 4.4** with highlighted sequence features.

The same DNA was also used as the template for PF0272 *in vitro* transcription. Another DNA template for PF0272 *in vitro* transcription was 302 bp (-223 to +79, relative to translation start). The PCR preparation was described in Chapter 2.

PF0464 Probe The PF0464 footprinting probe was 300-bp from -200 to +100 (relative to translation start) amplified by PCR using 5'-6FAM-labeled forward primer 5'-CTTA ACTATGCTGCCTCC-3' and 5'-HEX-labeled reverse primer 5'-TGATTCCGTAGTCTACA ACTCC-3'. The DNA sequence is shown in **Figure 4.5** with highlighted sequence features. The same DNA was used as the template for *in vitro* transcription of PF0464.

4.2.1.2. Reagents

Hitrap column binding buffer 1 (0.02 M sodium phosphate, 0.5 M NaCl, pH 7.4) and Hitrap Chelating HP column were from Amersham Biosciences. Polyacrylamide TBE gels were from Bio-Rad. 100 bp DNA ladder, Low mass DNA ladder and 6x Loading dye were from Promega. SYPRO Green dye was from Cambrex Bioscience. SYPRO Ruby dye was from Invitrogen. Footprinting binding buffer is composed of 25 mM Tris-HCl, 50 mM KCl, 6.25 mM MgCl₂, 10% glycerol, 0.5 mM DTT, pH 8.0. Footprinting stop solution is composed of 200 mM NaCl, 30 mM EDTA, 1% SDS. Ca²⁺/Mg²⁺ solution contains 5 mM CaCl₂ and 10 mM MgCl₂.

```

5'-6FAM- GGCAAGCACTAGGAAGGACTATAAAAATTATTGCTACAGCTGTTATTTTCT
          3'-CCGTTTCGTGATCTTCCTGATATTTTAAATAACGATGTCGACAATAAAAGA

TGTCCATGCTAACACCCTGTAATGAGATTTGGATTTTCCTATATAAAAAGCCTTAGT
ACAGGTACGATTGTGGGACATTACTCTAAACCTAAAAGGATATATTTTTCGGAATCA

TATTTTGTAGCCATTAAATATATAAGGAAGTATCACTCTTAGTGATTAATGGGTGGA
ATAAAAACCTCGGTAATTTATATATTCCTTCATAGTGAGAATCACTAATTACCCACCT

CGGAAGTGGGAGATAAAATTAACCTTCATATTTGGAATTCACAACCATCAGCCCCTGG
GCCTTCACCCTCTATTTTAATTGAAGTATAAACCTTAAGTGTTGGTAGTCGGGGACC

GCAACTTTGGATGGGTGTTTGAGGAGG -3'
CGTTGAAACCTACCCACAACTCCTCC -HEX-5'

```

Figure 4.4 The sequence of forward and reverse strands of the PF0272 footprinting probe.

TATA box (red); conserved motif (yellow); ribosome binding site (underlined); translation start site (grey).

5'-6FAM- CTTAACTATGCTGCCTCCAATAATTTTTGAAGGAGGATCAATAATTGAG
 3'- GAATTGATACGACGGAGGTTATTAAAACTTCCTCCTAGTTATTAAGTC

 GCACTTTCTTTTCCTTTTATCTTGATATTTCTGGCCATGTTTTATCTAGGCATTCA
 CGTGAAAGAAAAGGAAAATAGAAGCTATAAAGACCGGTACAAAATAGATCCGTAAGT

 CATTTACGGATATTTGACAAAATTAAATGCAAAAATTTTAAAGTAAGTTAAATGAA
 GTAAATGCCTATAAACTGTTTTAATTTACGTTTTTAAAAATTCATTCAATTTACTT

 ATCACGTCACTCGTAGTGGTATAATCGAGGTGATGACGTATGAAATTTTCAGTTCT
 TAGTGCAGTGACCATCACCATATTAGCTCCACTACTGCATACTTTAAAAGTCAAGA

 AAAACTAGATGTAGGTAAAAGGGAAGTAGAAGCTCAAGAGATAGAAAGGGAGGACA
 TTTTGATCTACATCCATTTTCCCTTCATCTTCGAGTTCTCTATCTTCCCTCCTGT

 TTTTCGGAGTTGTAGACTACGGAATCA-3'
 AAAAGCCTCAACATCTGATGCCTTAGT-HEX-5'

Figure 4.5 The sequence of forward and reverse strand of PF0464 footprinting probe. TATA box (red); conserved motif (yellow); transcription start site (magenta) (van der Oost *et al.*, 1998); translation start site (grey); ribosome binding site (underlined).

0.01 U/ μ L DNase I was freshly diluted from 10 U/ μ L (Roche). Phenol:chloroform solution is composed of phenol, chloroform and isoamyl alcohol at a ratio of 25:24:1.

4.2.2. Methods

4.2.2.1. Overexpression of PF0124 and Purification of PF0124p

The PF0124 gene fused with (His)₆ at the N-terminus was cloned into the modified kanamycin-resistant plasmid pET24d (by M.W.W. Adams' laboratory). The plasmid was transformed in *E.coli* CaCl₂ competent cells and grown in 1 L 2.5% LB media containing 50 μ g/mL kanamycin. IPTG was added to a final concentration of 0.84 mM at OD 0.6. The cells continued to grow at 37 °C for 2 more hours and were harvested after centrifugation at 6115xg at 4 °C for 20 min with a Beckman JS-5.3 rotor. The cell pellet was suspended in 15 mL Hitrap column binding buffer 1 (0.02 M sodium phosphate, 0.5 M NaCl, pH 7.4) containing 0.1 mM PMSF. Cell lysis was performed by sonicating at 20% power intensity for 15 sec in every 25 sec for 3 min. The cell extract supernatant was collected after centrifugation at 48,384xg for 60 min with a Beckman JA-25.5 rotor, and loaded on Ni²⁺ charged Hitrap Chelating HP column equilibrated with column binding buffer 1 containing 40 mM imidazole. The column was washed with the same equilibrating buffer, and protein was eluted with a linear gradient of imidazole (0.04-0.5 M) on FPLC. The eluted fractions were analyzed with SDS-PAGE. The fractions containing (His)₆-tagged protein were pooled and dialyzed with 20 mM Tris-HCl, pH 7.4, 0.4 M NaCl. The purified protein was confirmed by SDS PAGE and MS analyses. The final concentration of protein was about 0.2 mg/mL and protein was kept at -80 °C.

4.2.2.2. Electrophoretic Mobility Shift Assay

Standard EMSA methods (Lee *et al.*, 2005) were used to investigate protein–DNA interactions between PF0124p and various UORs. PF0124p was supplied as a gradient against a

constant concentration of dsDNA. The amounts of DNA and protein applied are detailed in the figures in the Results section. The protein and DNA were mixed in 15 μ L EMSA reaction buffer (50 mM Tris-HCl, pH 7.5; 1 mM EDTA; 0.1 M KCl; 5% glycerol). The mixture was incubated at 70 °C for 20 min, after which the 15 μ L products of reactions were loaded onto 5% polyacrylamide TBE gel. 2 μ L of 100-bp DNA ladder or Low mass DNA ladder mixed with 2 μ L 6x loading dye and 8 μ L of Nanopure water was loaded onto gel as the indicator of the sample front during the electrophoresis. The electrophoresis was run at room temperature under constant voltage of 200 V for 20 min. The TBE gel was stained in 50 mL 0.01% SYPRO Green for 30 min and washed with water before visualization of DNA. The gel was sometimes stained again with 50 mL SYPRO Ruby for 3 h and destained with 10% methanol, 7% acetic acid for 1 h before protein imaging. The gel visualization and image acquisition were performed with Fluorchem 8000 imaging system.

4.2.2.3. DNase I Footprinting

DNase I footprinting was performed using capillary electrophoresis (CE) (Wilson *et al.*, 2001) to explore the specific binding site of PF0124p on the PF0272 and PF0464 UORs. The DNA probes in the assay were amplified by PCR using fluorescently labeled primers with an annealing temperature of 56 °C. PCR products were gel-purified using QIAquick Gel Extraction Kit. 20 nM DNA probe was applied in 50 μ L footprinting binding buffer (buffers are specified in section 4.2.4.2) and incubated with 0.2 μ M PF0124p. After incubation for 20 min at 70 °C, the solution was mixed with 50 μ L $\text{Ca}^{2+}/\text{Mg}^{2+}$ solution and incubated at room temperature for 1 min. DNA cleavage was performed by addition of 3 μ L of DNase I solution (0.01 U/ μ L, RNase free) followed by incubation for 1 min at room temperature. The reaction was stopped by adding 99 μ L stop solution, after which the DNA was extracted with 200 μ L of phenol:chloroform. The

extracted DNA was precipitated by 0.3 M sodium acetate and 3x volume of ethanol. The precipitated DNA was washed with 75% ethanol and resuspended in 10 μ L formamide solution containing GeneScan-500-ROX size standard. The pattern of DNA fragment was analyzed by an auto sequencer.

4.3. Results and Conclusions

4.3.1. Interaction of PF0124p with PF0272 and PF0464

As shown by EMSA, PF0124p and PF0272 UOR formed a protein-DNA complex (**Figure 4.6**). The complex was detectable when the protein concentration reached 0.14 μ M (protein:DNA ratio of 1.5). Most DNA was shifted as the protein concentration reached 0.8 μ M. At a protein concentration of 0.4 μ M, a second complex started to form with much lower intensity than the major complex. On the other hand, no protein-DNA complexes were observed with the control DNA. Given the difference between the control DNA and PF0272 UOR, we conclude that the conserved sequence replaced in the control DNA contains the critical site(s) for protein-DNA interaction.

In the study of the interaction of PF0124p with PF0464 UOR by EMSA, a sequence from the PF0464 ORF was used as control DNA. Based on the detection of both DNA and protein, it is clear that only the PF0464 UOR was shifted by the proteins (**Figure 4.7a&b**). No control DNA was shifted by protein even at the highest protein concentration in the assay, indicating that the interaction between PF0124p and PF0464 UOR was sequence specific.

4.3.2. DNA binding Sites for PF0124p

4.3.2.1. The Conserved Palindromic Motif Is Involved in Protein-DNA Interaction

In the study of PF0124p with gene PF0272, the mutated conserved sequence in the UOR was shown to be critical for protein binding. To further investigate the specific protein-DNA

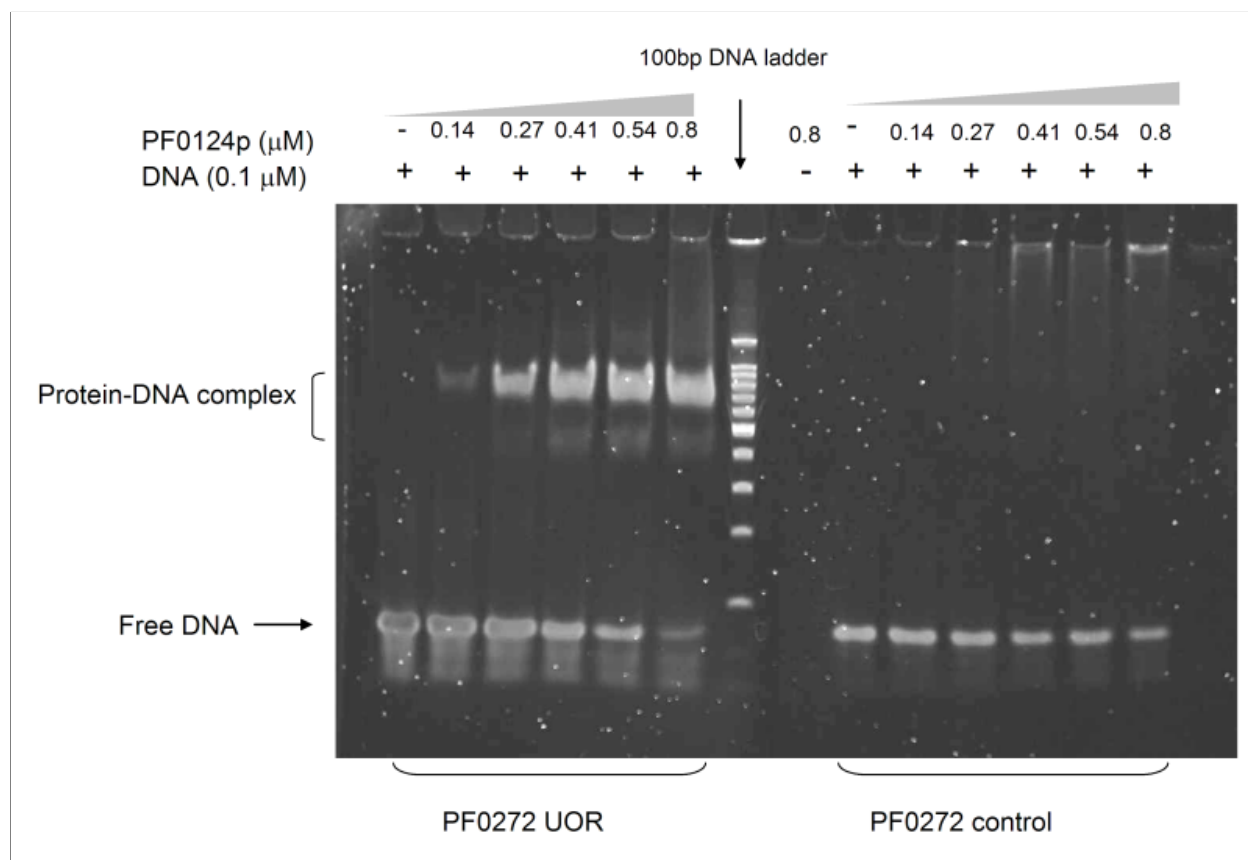
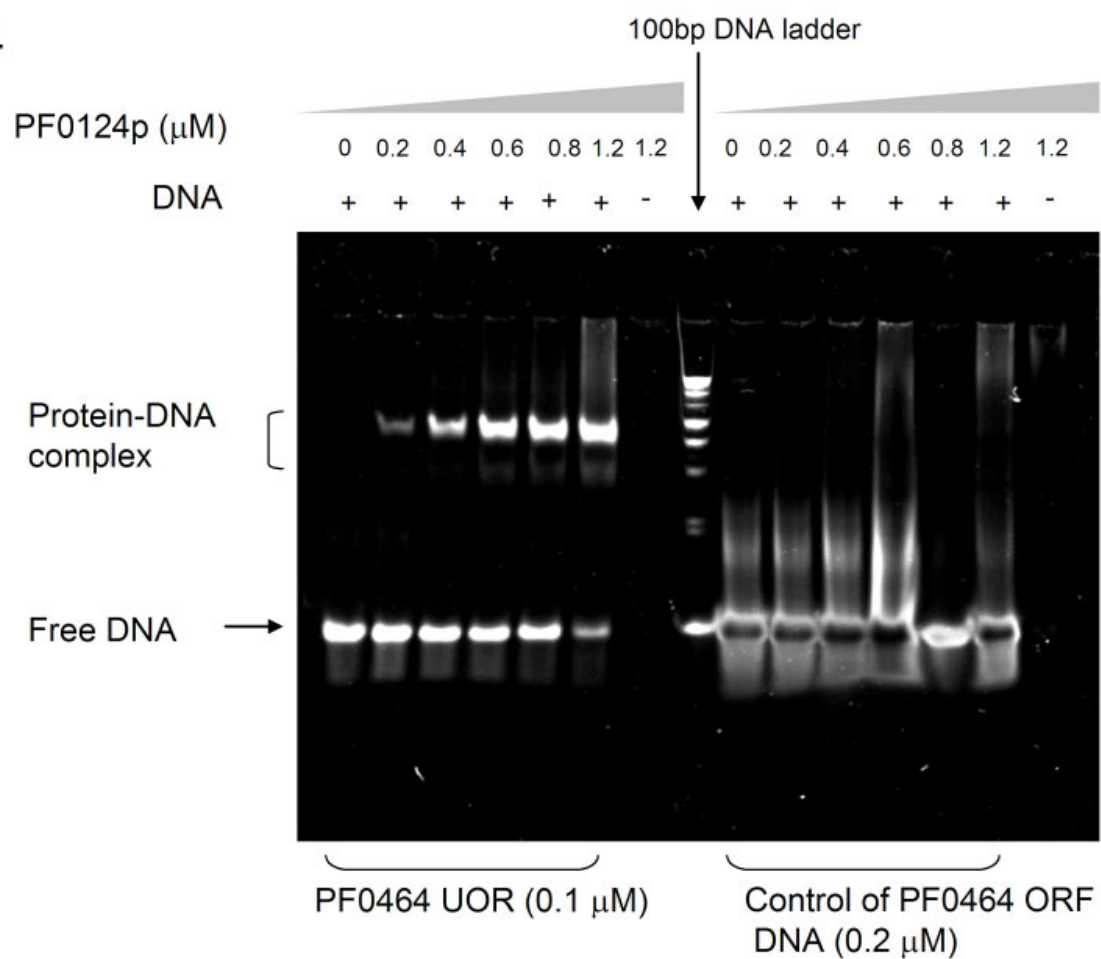


Figure 4.6 EMSA of PF0124p with PF0272 UOR and control DNA. 0.1 μ M of PF0272 UOR or control DNA was mixed with varying amounts (0-0.8 μ M) of PF0124 in a total volume of 15 μ L respectively, as indicated in the figure. The gel was stained with SYPRO Green for DNA visualization. Protein-DNA complex appeared in gel with PF0272 UOR DNA. As protein amount is increased, the protein and DNA formed secondary complex. On the right side, the control DNA did not form complex with the protein over the same protein concentration range, demonstrating the important role of the conserved sequence in the PF0272 UOR.

a.



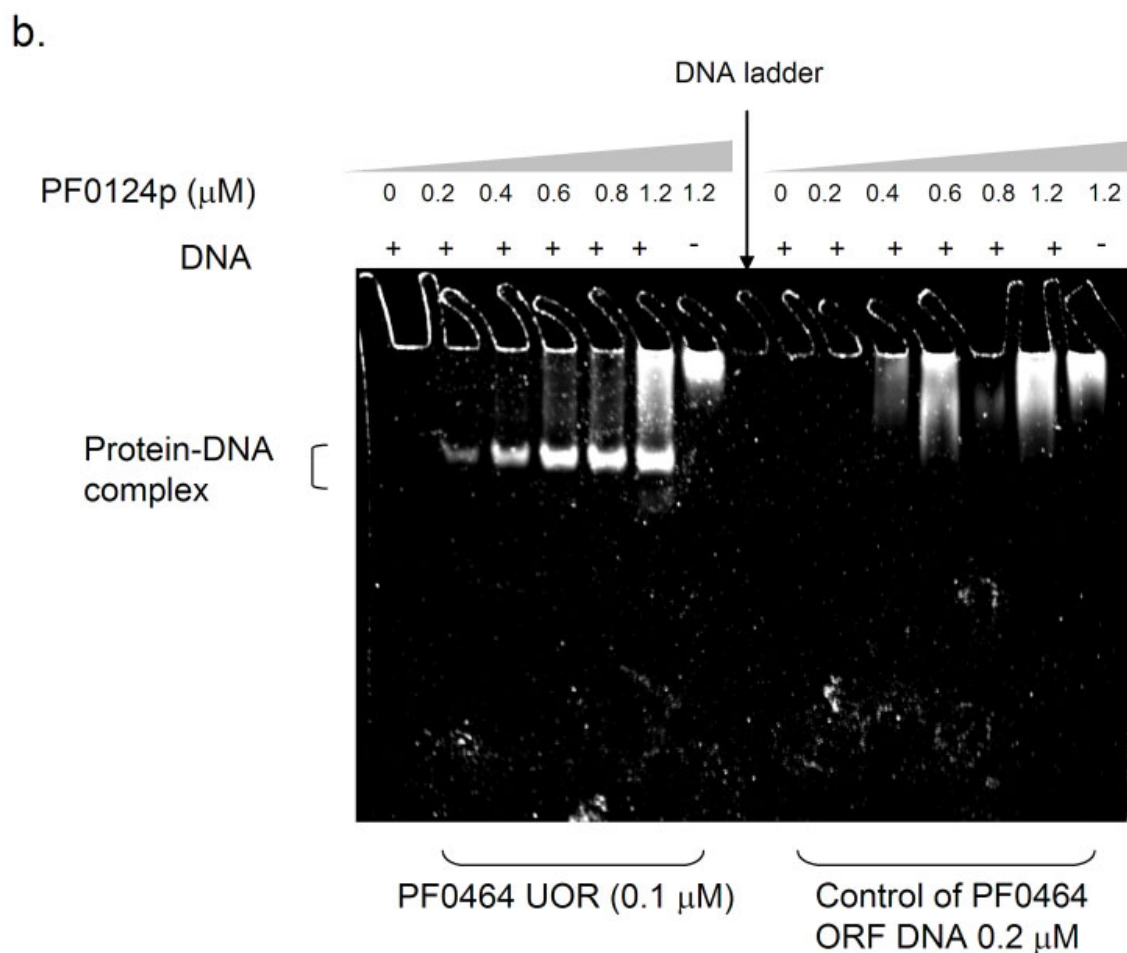


Figure 4.7 EMSA of PF0124p with PF0464 UOR and ORF control DNA. Constant concentration of PF0464 UOR (0.1 μM) or ORF control DNA (0.2 μM) was mixed with varying concentration of PF0124p (0-1.2 μM) in a total volume of 15 μL as indicated in the figure. **a.** DNA staining by SYPRO Green shows free UOR decreased as protein concentration increased in the reaction. The secondary protein-DNA complex was formed at high concentration of protein. No control DNA was shifted by the same gradient of PF0124p. **b.** Protein staining by SYPRO Ruby showed protein-UOR complex increased as protein amount increased. In the control experiments with ORF DNA, PF0124p appears at the top of the gel just as in the control lane of pure PF0124p without DNA.

interacting site(s), a new control DNA for PF0464 was constructed by mutating the PF0464 UOR in the palindromic sequence that is the potential PF0124p binding site. From the EMSA study, the UOR showed formation of complexes, but this new control DNA did not (**Figure 4.8**). The result illustrates again that the 14 bp palindromic sequence is important for PF0124p binding to PF0464 UOR, consistent with the results using the PF0272 UOR.

4.3.2.2. PF0124p Binding Site(s) Determined by Footprinting

The DNase I footprinting analysis of PF0124p protection of DNA indicated the specific binding site of PF0124p on DNA. Comparing with the DNase I digestion pattern of DNA without protein, the footprinting result of DNA with protein showed protection of a region of UOR against DNase I digestion, which is the consequence of protein-DNA interaction (**Figure 4.9**). The footprint on the forward strand DNA covers ca. -8 to -40 relative to PF0272 translation start site. The footprint on the DNA reverse strand was between ca. -5 to -36. The conserved sequence TCACtcttaGTGAt of PF0272 is from -17 to -30. The PF0124p footprint covered the conserved motif and supported the previous EMSA results. The nucleotide at the position of -18 on PF0272 UOR forward strand and the nucleotide at -27 on the reverse strand are hypersensitive sites for DNase I.

Footprinting analysis of PF0124p binding to PF0464 UOR was performed using recombinant protein and fragment DNA from -200 to +100 (relative to translation start site). PF0124p protects DNA on both strands (**Figure 4.10**). The conserved palindrome sequence is -20 to -33, which is over the transcription start site at -28 (van der Oost *et al.*, 1998) of PF0464. The binding sites of PF0124p are from -13 to -39 on the UOR forward strand and from -13 to -38 on the reverse strand. This result is in agreement with the EMSA study and confirms the

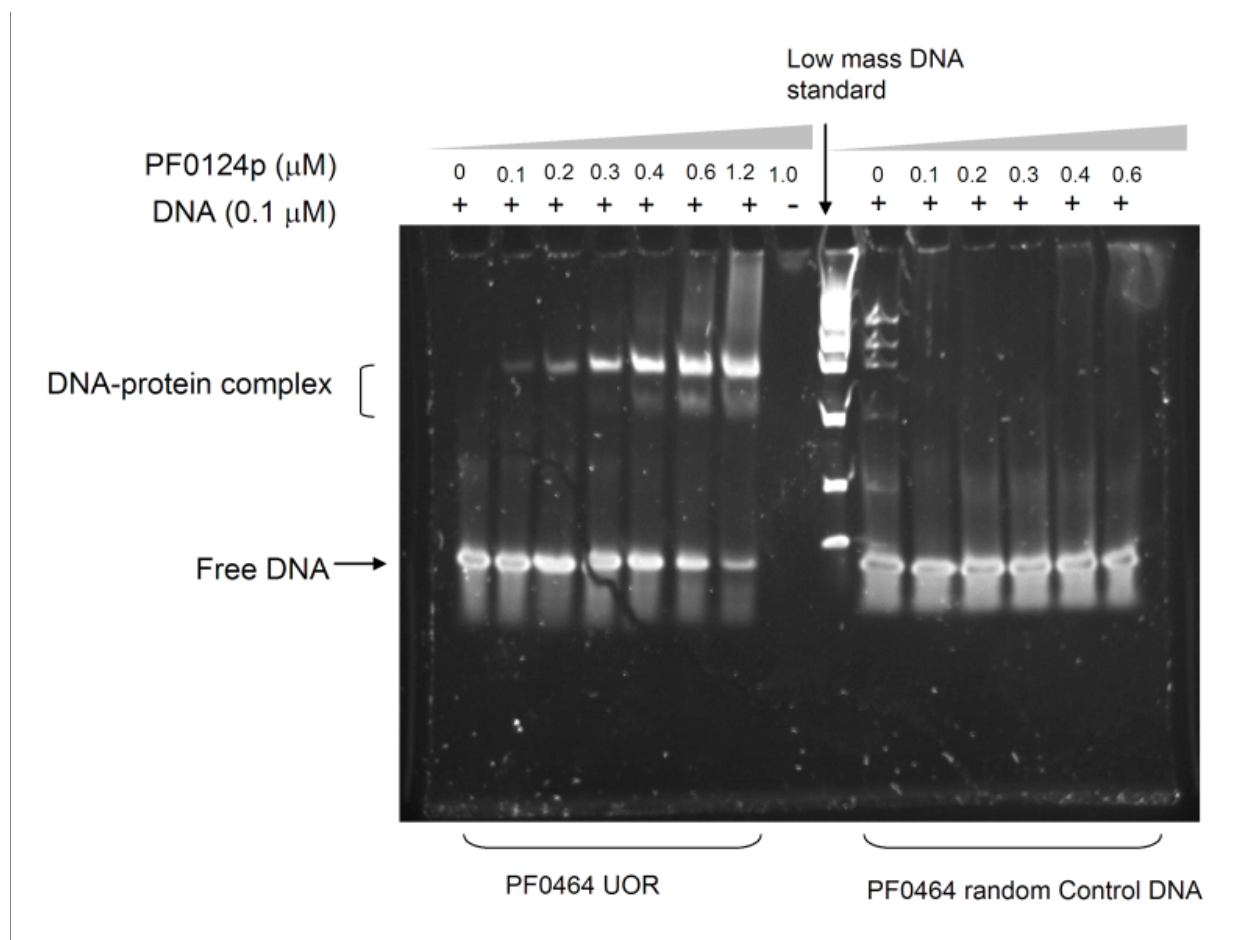


Figure 4.8 EMSA of PF0124p with PF0464 UOR and mutated control DNA. The UOR of PF0464 or control DNA was added with constant concentration of $0.1 \mu\text{M}$. The amount of added PF0124p in each reaction is shown in the figure. The gel shows the protein-DNA complex formed with PF0464 UOR. The control DNA with the conserved palindrome motif replaced with random nucleotides of the same GC content, did not form a complex with PF0124p.

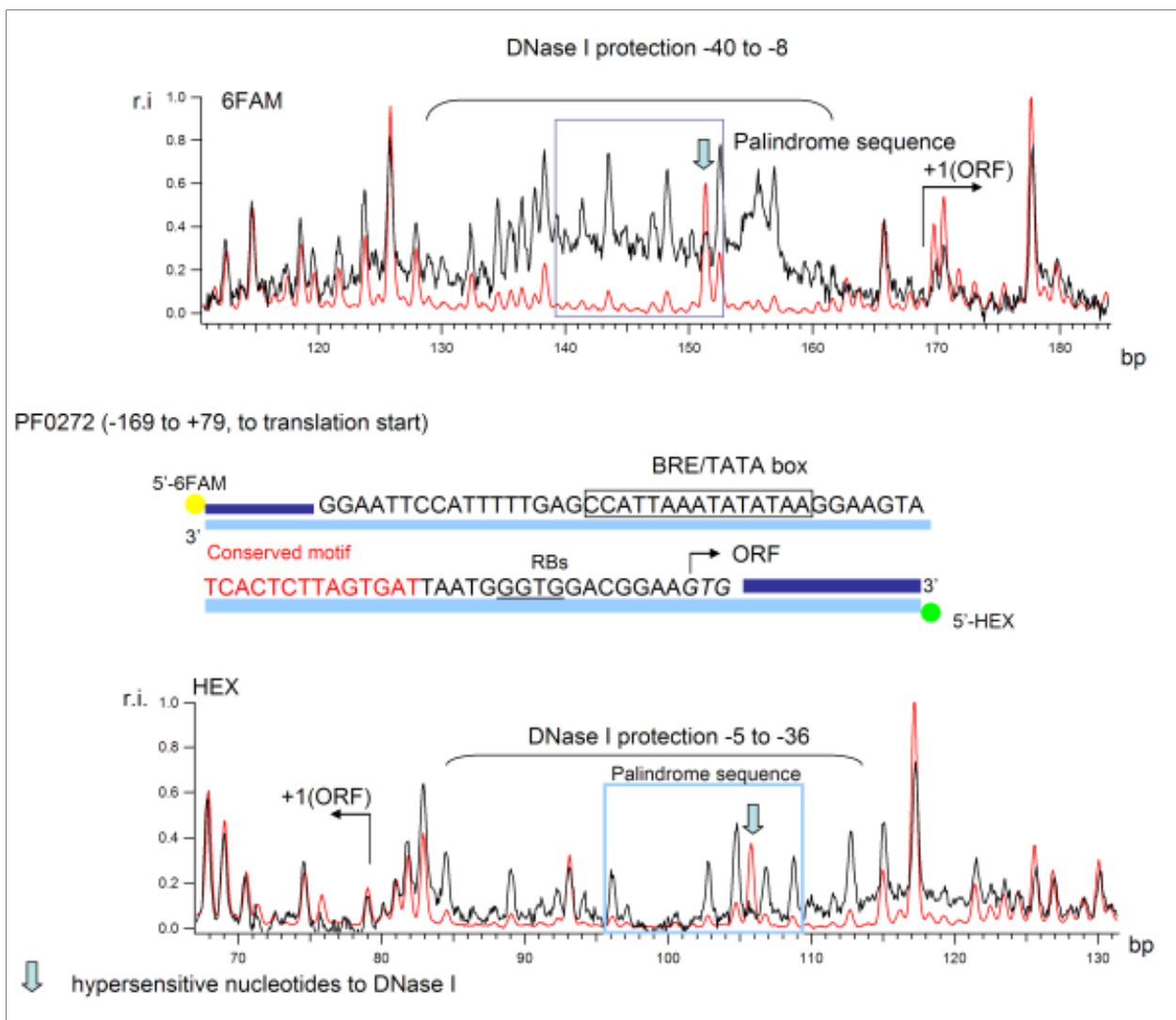


Figure 4.9 CE chromatograms showing footprint of PF0124p on PF0272 UOR. Each chromatogram compares control (black line, without protein) and sample (red line, with protein), with y axis relative intensity and x axis the DNA length (bp). The specific sites of PF0272 UOR protected by PF0124p from DNase I are indicated. The top chromatogram shows the protected region on PF0272 UOR forward strand (6FAM labeled) as from -40 to -8 (relative to PF0272 translation start). The bottom chromatogram shows the protected sites on the UOR reverse strand (HEX labeled) is from -5 to -36. The conserved palindromic sequence (-17 to -30) is indicated in the squares on the chromatograms. The sequence of PF0272 UOR is shown in the middle of the

figure with indication of BRE/TATA element (boxed), conserved palindromic sequence (red), ribosome binding site (RBs, underlined) and translation start site (italic).

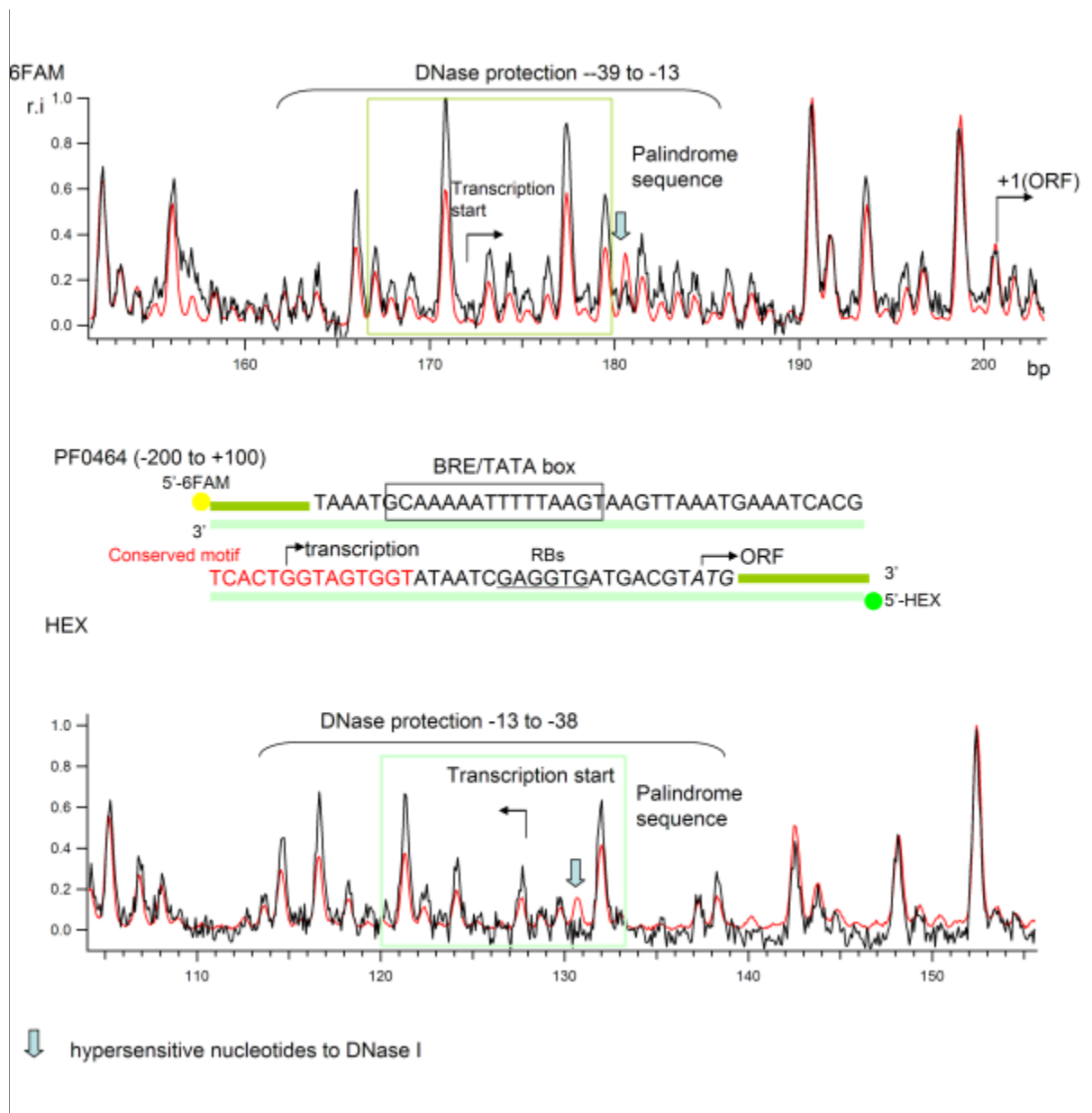


Figure 4.10 CE chromatograms showing footprint of PF0124p on PF0464 UOR. Each chromatogram compares control (black line, without protein) and sample (red line, with protein), with y axis relative intensity and x axis the DNA length (bp). The specific region of PF0464 UOR protected by PF0124p from DNase I is indicated. The top chromatogram indicates the protected region on PF0464 UOR forward strand (6FAM labeled) as -39 to -13 (relative to PF0464 translation start). The bottom plot shows the interaction sites on the UOR reverse strand

(HEX labeled) as -38 to -13. The conserved palindromic sequence was indicated in the squares on the chromatograms. The sequence of PF0464 UOR is shown in the middle of the figure with indication of BRE/TATA element (boxed), conserved palindromic sequence (red), RBS (underlined), translation start site (*italic*) and known transcription start site.

palindrome sequence as the PF0124p binding site. The DNase I hypersensitive sites are at -25 on the UOR forward strand and -31 on the reverse strand.

4.3.3. Other Genes Potentially Regulated by PF0124p

As the conserved palindromic sequence was considered probably the critical site for PF0124p interacting with both PF0272 and PF0464 UORs, we identified other genes containing the same conserved sequences to determine whether they also interact with PF0124p. PF1959 and PF1920 also contain the same conserved palindromic motif in their UORs. PF1959 and PF1920 are also both up-regulated during growth on maltose condition and are involved in glycolysis. So, the interaction of PF0124p with PF1959 and PF1920 UORs was studied with EMSA.

EMSA showed that the protein–DNA complexes between PF0124p and both PF1959 and PF1920 UOR (**Figure 4.11**). The reaction condition and protein-DNA ratio were similar to the previous analysis. Therefore PF0124p is probably a common binding protein for the group of genes that contain the conserved motif in their promoters.

4.3.4. Negative Regulation of PF0272 and PF0464 by PF0124p

To determine the regulatory function of PF0124p in the transcription of PF0272 and PF0464, *in vitro* transcription assays were performed in Dr. Michael Thomm's laboratory with an established cell-free transcription system (Hethke *et al.*, 1996) using the purified PF0124p and gel-purified DNA templates from PCR amplification. As a control, the glutamate dehydrogenase gene (*gdh*) of *P. furiosus* was used. The PF0464 templates (300bp, -200 to +100 relative to translation start site) produced a 128 nt (nucleotides) transcript as shown in **Figure 4.12**. As the amount of added PF0124p was increased, the amount of PF0464 run-off transcript was reduced. Two lengths of PF0272 templates were used in the *in vitro* transcription: (1) 248 bp, -169 to +79;

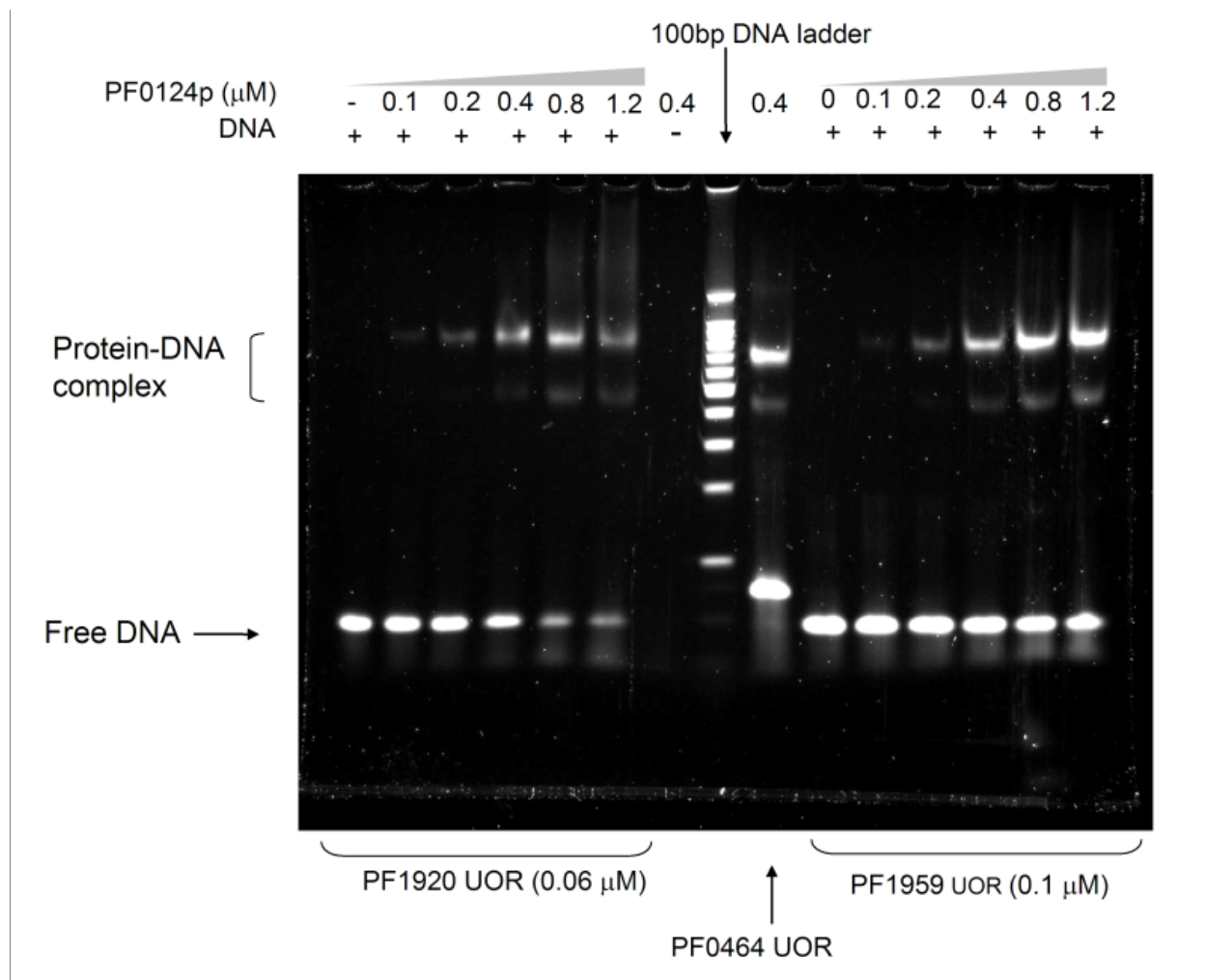


Figure 4.11 EMSA of PF0124p with UOR of PF1959 and PF1920. A constant concentration of PF1920 UOR (0.06μ M) and PF1959 UOR DNA (0.1μ M) was mixed with varying amounts of PF0124p (0.1 - 1.2μ M) as indicated in the figure. Both UORs of PF1920 and PF1959 formed specific DNA-protein complexes with PF0124p. A secondary multimer of protein-DNA complex was formed with both UORs as the amount of PF0124p increased.

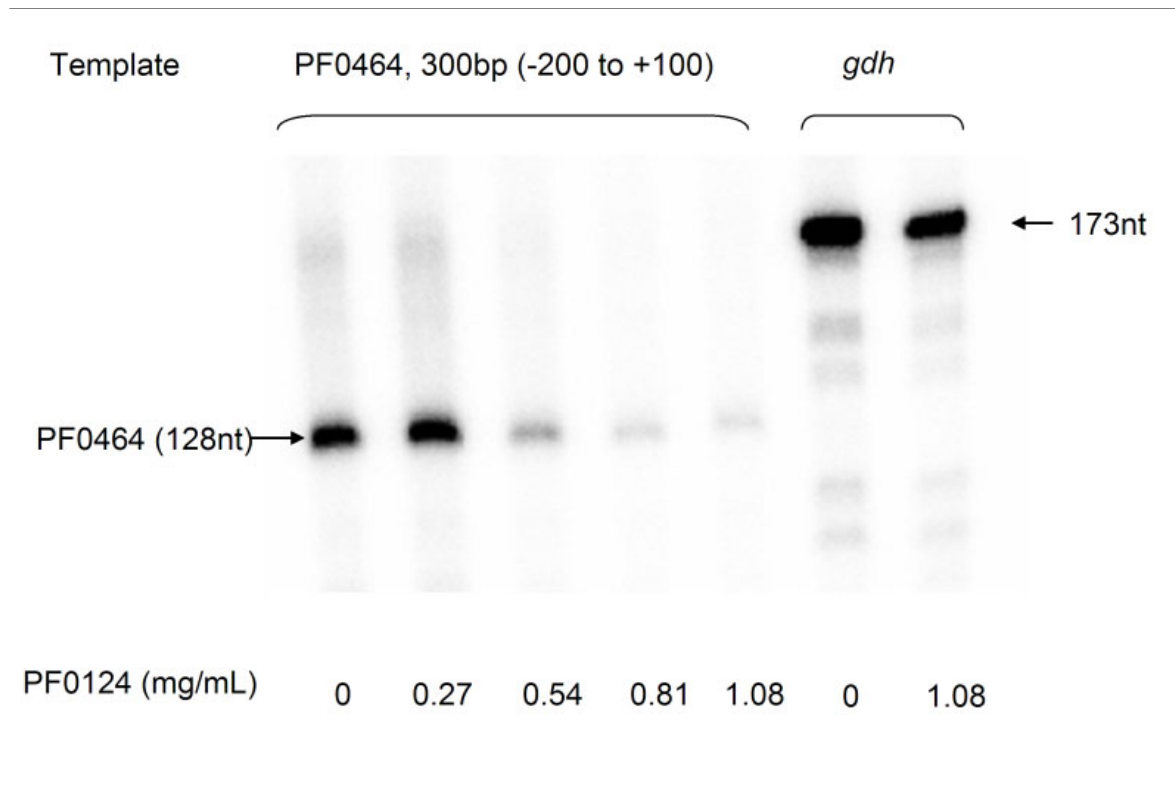


Figure 4.12 The repression of PF0464 transcription by PF0124p in *in vitro* transcription. Gel-purified PCR products of PF0464, 300bp from -200 to +100 (relative to translation start) were used as the template in *in vitro* transcription experiments. The run-off transcripts are indicated by arrows, and their respective sizes are shown. The added PF0124p was from 0-1.08 mg/mL as indicated at the bottom. The *gdh* transcripts from the control experiments are shown in the right most two lanes.

(2) 300 bp, -223 to +79 (both relative to PF0272 translation start site), ending at the same position in the PF0272 ORF. Therefore, the run-off transcripts of PF0272 from both templates should be of the same length. As we observed, a common transcript of identical size was produced from both templates as expected from PF0272. The amounts of the PF0272 transcript were decreased as more PF0124p was used (**Figure 4.13**), suggesting that the transcription was repressed by PF0124p. Interestingly, besides the PF0272 transcripts, an extra run-off transcript was observed from each template, and they are different in length. Considering that both templates also contain the UOR of PF0271, a hypothetical gene transcribed in the opposite direction of PF0272, and they end at different position in the PF0271 ORF, it is quite possible that the extra transcript from each template was the transcript of PF0271. In contrast to the repression of PF0272 transcription, PF0124p seemed to enhance the transcription of PF0271.

4.3.5. Search for Effectors of PF0124p

4.3.5.1. PF0124p Regulation Is Independent of Maltose or Sucrose

As PF0272 and PF0464 are both involved in glycolysis, we hypothesized that the transcriptional regulation by PF0124p is affected by some sugars such as maltose and sucrose. In the EMSA analysis, the intensity of PF0124p-PF0464 UOR complex did not show an obvious difference with addition of either maltose or sucrose (**Figure 4.14**). In the EMSA with PF0272, similar results were observed (**Figure 4.15**), which suggested that the PF0124p regulation of PF0272 and PF0464 is independent of maltose or sucrose, though both genes are responsive to maltose in the cell growth condition.

4.3.5.2. PF0124p Regulation Is Independent of Glucose

It is known that PF0272 exhibits 4- α -glucanotransferase activity in the presence of glucose, so an analysis was performed to study whether glucose is the effector in PF0124p

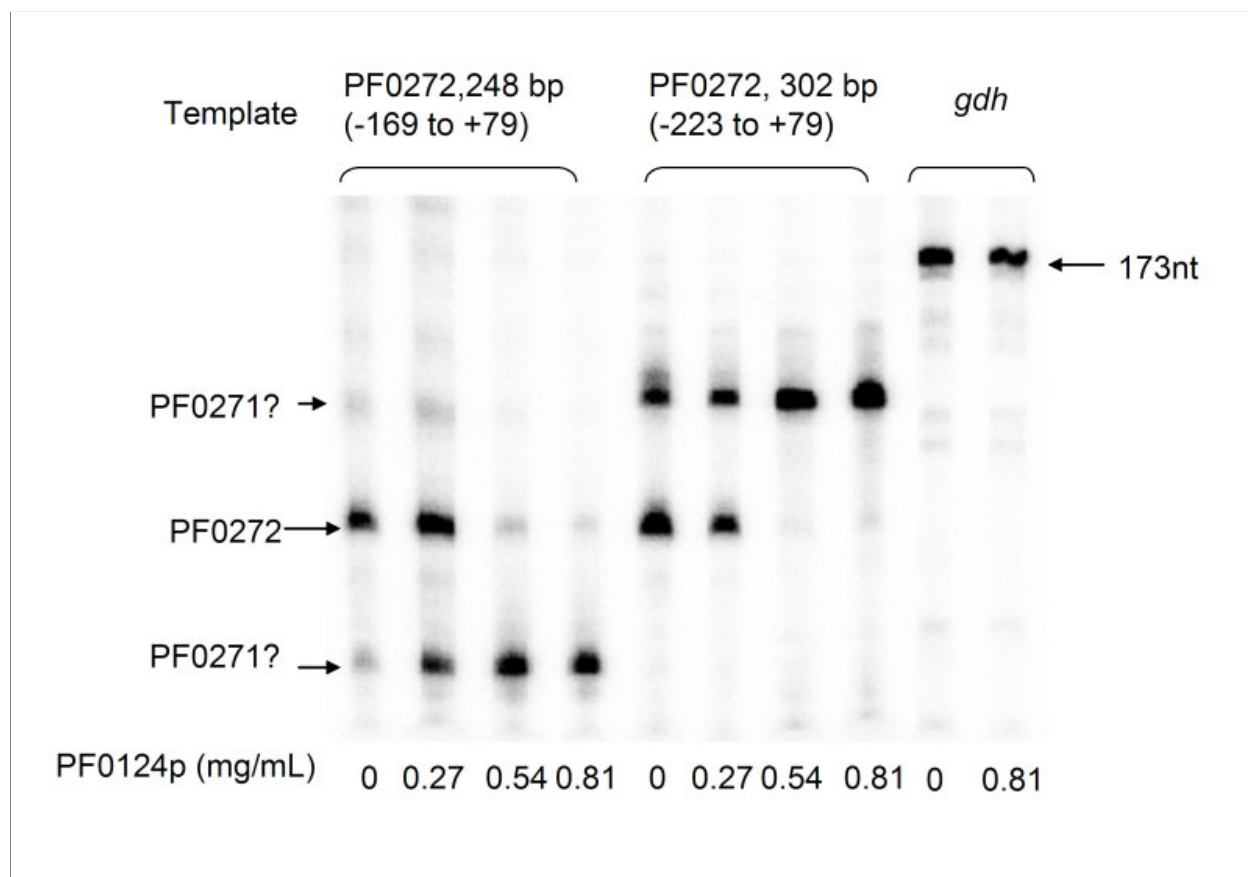


Figure 4.13 The repression of PF0272 transcription by PF0124p in *in vitro* transcription. Two lengths of PF0272 transcription templates: (1) 248 bp, -169 to +79 (2) 302 bp, -223 to +79, (both relative to translation start site) were used in *in vitro* transcription experiments. This template also contains the UOR of PF0271 that is transcribed in the opposite direction as PF0272. The run-off transcripts are indicated by arrows, and their respective sizes are shown. The concentrations of added PF0124p were as indicated at the bottom. The *gdh* transcripts from the control experiments are shown in the right most two lanes.

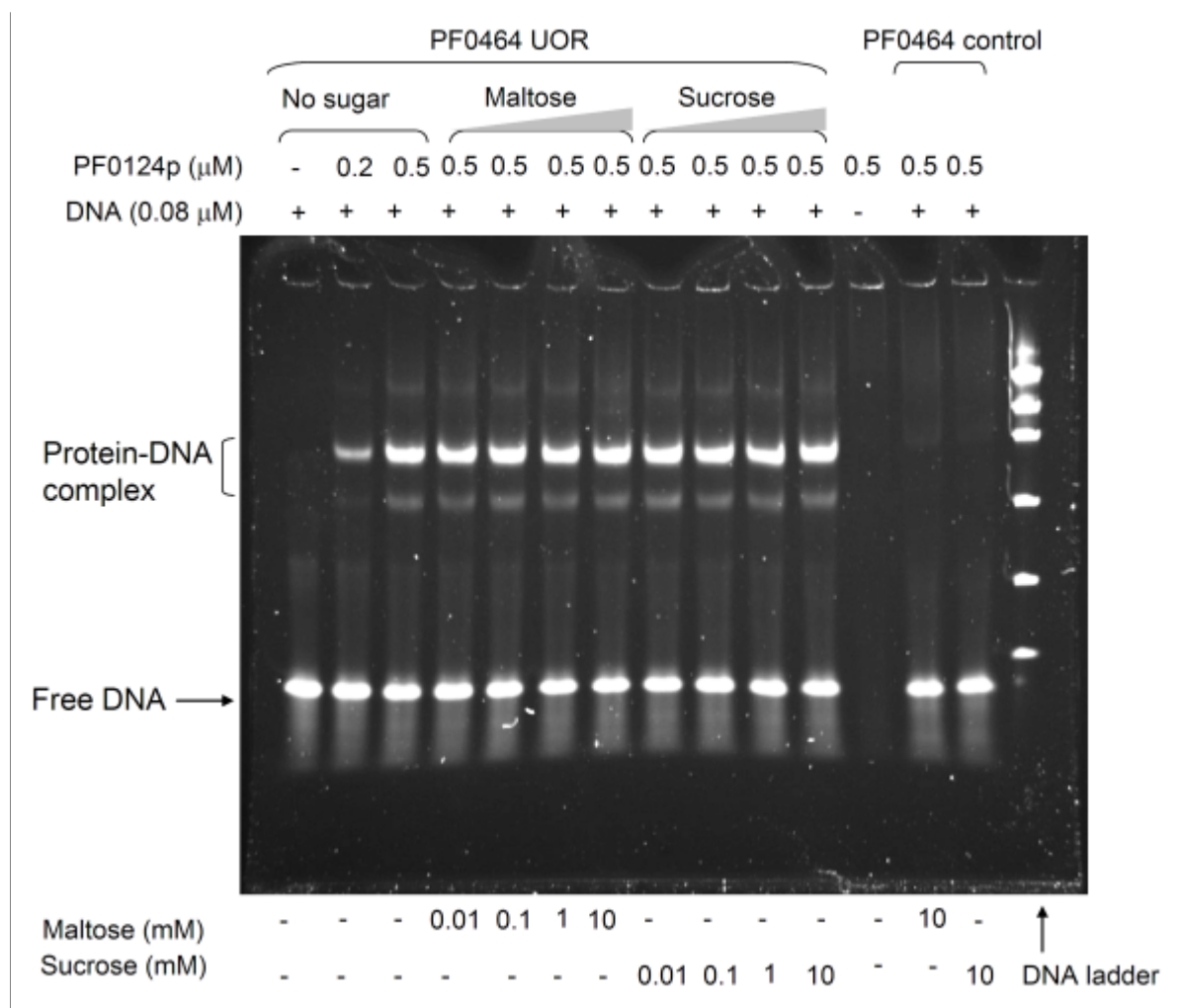


Figure 4.14 EMSA analysis of the effect of maltose and sucrose on PF0124p regulation of PF0464. Constant concentrations of PF0464 UOR (0.08 μM) and PF0124p (0.5 μM) were mixed with varying amounts of maltose or sucrose (0.01-10 mM) as indicated in the figure. No obvious difference of protein-DNA complex intensity is observed with either sugar. The reactions of only PF0124p and PF0464 UOR were performed as sugar negative controls. The reaction with PF0464 control DNA (with randomized sequences) was used as a DNA negative control.

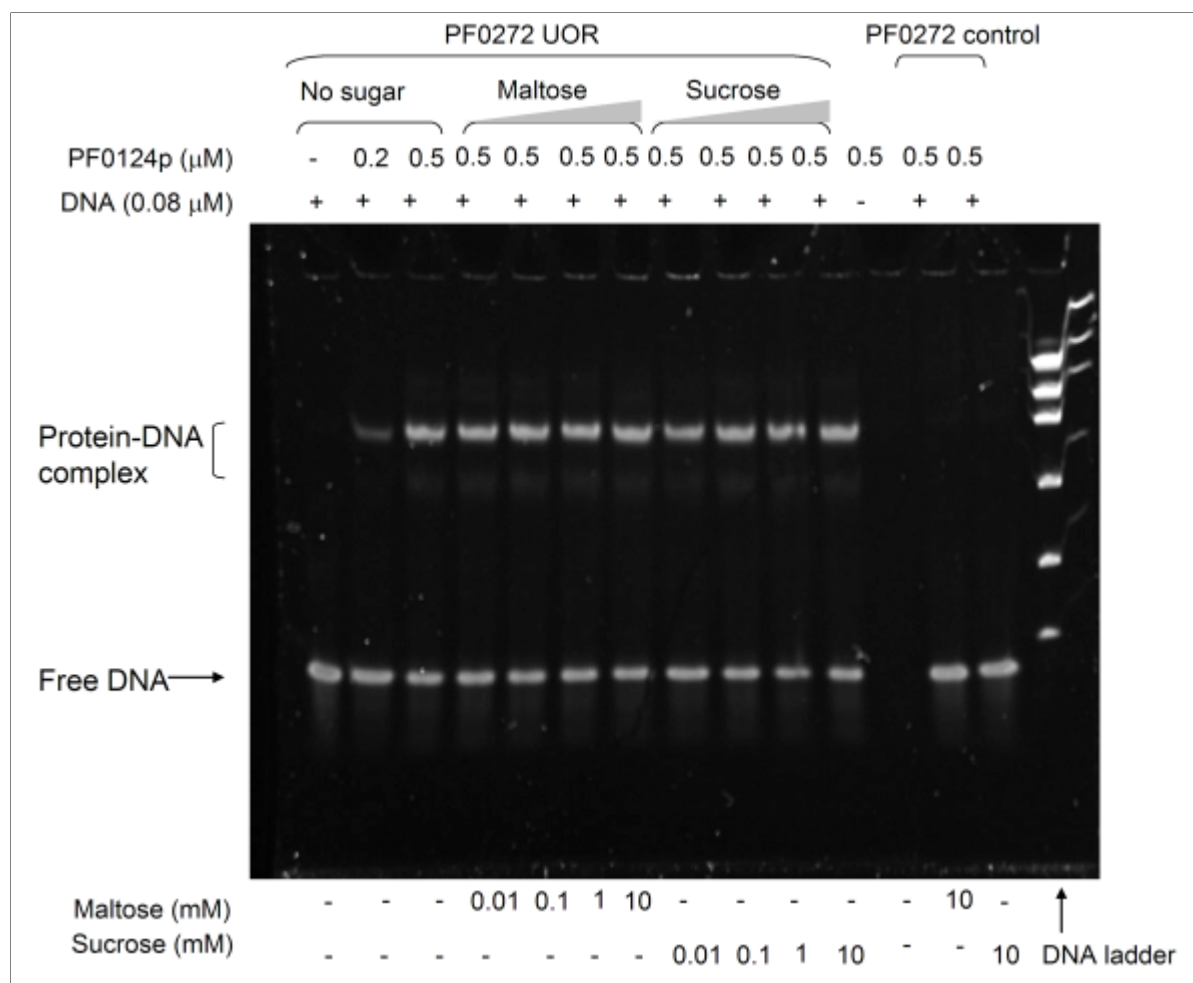


Figure 4.15 EMSA analysis of the effect of maltose and sucrose on PF0124p regulation of PF0272. A constant concentration of PF0272 UOR (0.08 μM) and PF0124p (0.5 μM) was mixed with varying amounts of maltose or sucrose (0.01-10 mM) as indicated in the figure. No obvious difference of protein-DNA complex intensity is observed with either sugar. The reactions of only PF0124p and PF0272 were performed as sugar negative controls. The reactions with PF0464 control DNA (with randomized sequence) was used as a DNA negative control.

regulation. However, EMSA did not show any obvious change caused by the presence of glucose to either PF0272 or PF0464 (**Figure 4.16**).

4.4. Discussion

4.4.1. PF0124p as Co-regulator for Multiple Genes

The biophysical study with EMSA and footprinting proved that PF0124p interacted specifically with PF0464 and PF0272 UORs. The predicted palindromic sequence was important for PF0124p binding to DNA. The *in vitro* transcription assay confirmed that PF0124p is a transcriptional repressor of both PF0272 and PF0464. In addition, the specific interaction of PF0124p with UORs of PF1959 and PF1920 is sensitive to the palindromic sequence, indicating that PF0124p is very likely a common transcriptional regulator of multiple genes. But the regulation by PF0124p of genes PF1959 and PF1920 requires confirmation.

Besides the genes studied here, more glycolytic genes carrying the palindromic motif are found from the maltose-induced stimulon, so it is of great interest to know whether PF0124p will interact with the promoters of other glycolytic genes. Moreover, in the *in vitro* transcription assay of gene PF0272 with PF0124p, we observed the likely activation of transcription of gene PF0271, a *Pf* hypothetical protein with unknown function, which is immediately upstream of PF0272 and transcribed in the opposite direction. Although it is still too early to conclude that PF0124p is also a transcriptional activator for genes such as PF0271 before further analysis, the role of PF0124p as a transcriptional repressor of genes PF0272 and PF0464 is clear.

4.4.2. Differential Interaction of PF0124p with Various Promoters

The binding site of PF0124p is a conserved palindromic sequence, suggesting that PF0124p binds DNA as a dimer. It was also noticed in the footprinting study that the protection by PF0124p of PF0272 UOR is stronger than that of PF0464 UOR. In addition, the EMSA study

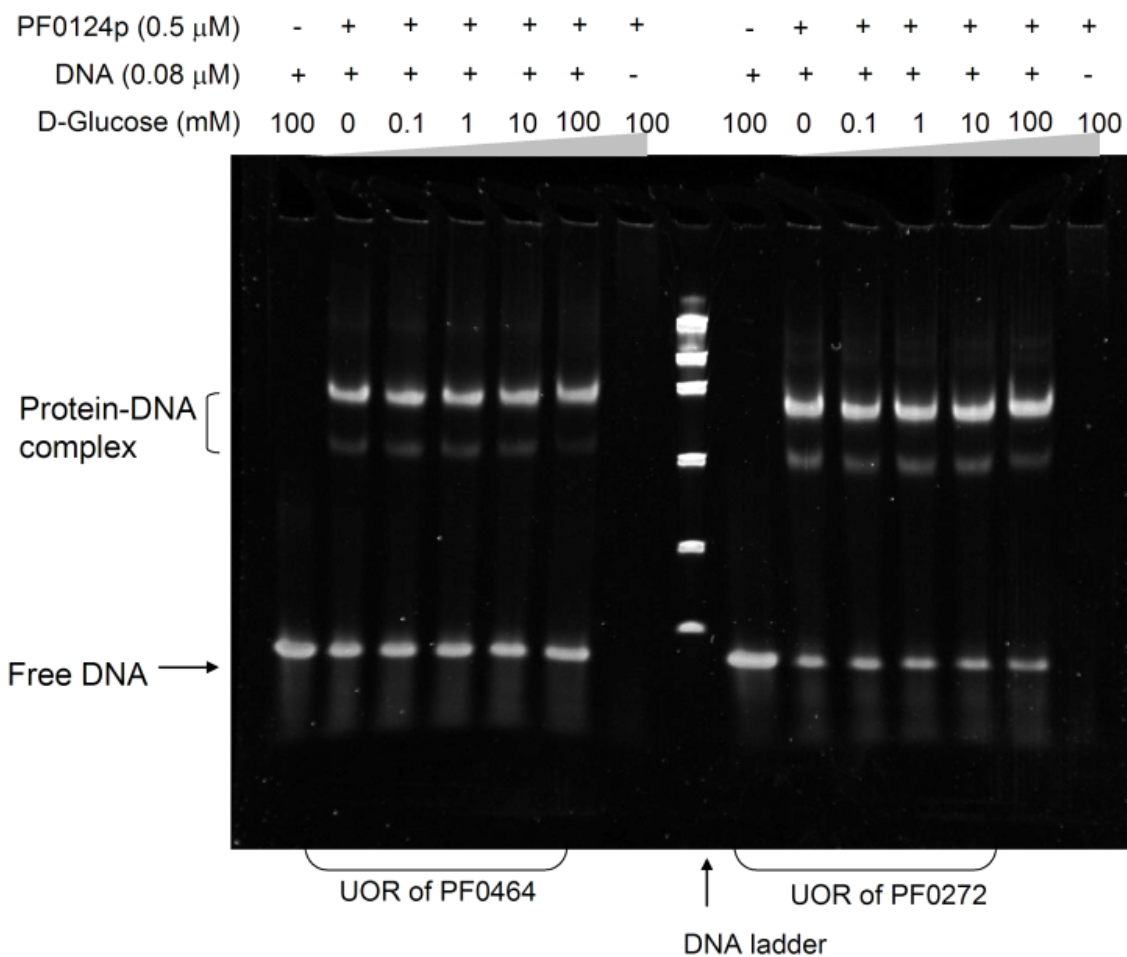


Figure 4.16 EMSA analysis of the effect of glucose on regulation of PF0272 and PF0464 by PF0124p. A constant concentration of UOR (0.08 μ M) and PF0124p (0.5 μ M) was mixed with varying amounts of glucose (0.01-100 mM) as indicated in the figure. No obvious difference of protein-DNA complex intensity is observed.

also showed a stronger interaction between PF0124p and PF0272 UOR than with other UORs including PF0464, PF1959 and PF1920. Comparing the palindromic sequence of TCAC-repeats in these UORs, PF0272 (TCAC/T/CTT/A/GTGA/T), PF0464 (TCAC/T/GGT/A/GTGG/T), PF1920 (TCAC/G/ATT/G/GTGA/T), and PF1959 (TCAC/T/GTA/G/GTGA/T), the second half of the palindromic sequence in PF0464 UOR (GTGG) is not a perfect invert of TCAC as in the other three. This could be the reason that PF0124p interacts with PF0464 UOR less strongly than with PF0272 UOR. Alternatively the PF0272 UOR sequence is also a perfect inverted repeat of “TCACT” with 3 nucleotides between (a 5-3-5 rather than 4-5-4 palindrome). PF0464, PF1959 and PF1920 all lack one or two nucleotides to have the 5-3-5 palindrome. This suggests the possibility that the 5-3-5 palindrome is the full recognition sequence for PF0124p binding. To find out the answer, further mutagenesis analysis or a SELEX (systematic evolution of ligands by exponential enrichment) analysis is necessary.

4.4.3. Suggested Regulation Mechanism of PF0124p

The two known primary mechanisms of transcription repression in archaeal are (1) interference with TBP/TFB recognition of TATA/BRE elements on the gene promoter (e.g. TrmB (Lee *et al.*, 2003)) and (2) obstructing RNA polymerase recruitment at the transcription initiation sites (e.g. LrpA (Dahlke *et al.*, 2002)). Our footprinting study indicated that the protection site on PF0272 by PF0124p is at about -5 to -40 (relative to the gene translation start site). A putative TATA/BRE of PF0272 is from -38 to -52, which is immediately upstream of the protected region. PF0124p does not seem to interfere with the TATA/BRE recognition by TBP/TFB for PF0272.

The transcription initiation site of PF0464 has been mapped at position of -28 (relative to the translation start site) by primer extension (van der Oost *et al.*, 1998). The TATA element of

the PF0464 promoter is from -51 to -67, 27 bp upstream of the transcription start site. In our DNase I footprinting study, the PF0464 UOR was mainly protected by PF0124p at -13 to -39, which encloses the transcription initiation site. Therefore, the PF0124p binding site does not overlap with the TATA/BRE elements in either PF0272 or PF0464. PF0124p binding covers the transcription initiation site in PF0464, suggesting that the negative regulation by PF0124p is probably mediated by abrogating the recruitment of RNAP to promoter.

From the bioinformatics analysis of all UORs of the glycolysis group (data from Ying Xu's group), the conserved palindromic sequence (putative binding site of PF0124p) is downstream of the predicted BRE/TATA elements and very close or overlapping the known or predicted transcription initiation site. It is quite possible that obstructing RNAP recruitment is a general repression mechanism of PF0124p. This is interesting since PF0124p is homologous to TrmB, but likely employs a different regulatory mechanism from TrmB. Therefore, PF0124p probably represents a new type of transcriptional regulator for archaeal glycolytic genes.

4.4.4. PF0124p Regulates in Response to Carbon Source

PF0124p was shown to be a common repressor of PF0272 and PF0464 in the *in vitro* transcription system. Specific interactions of PF0124p with PF1959 and PF1920 UORs are also observed. Other genes in this group are possibly regulated by PF0124p too, since they contain the same palindromic sequence. These genes are all involved in *Pf* glycolytic metabolism and highly up-regulated when *Pf* is grown on maltose rather than on peptides as the primary carbon source (Schut *et al.*, 2003). Therefore it is highly possible this co-regulation is in response to carbon source in the environment.

The densitometry analysis of our protein capture gels indicates higher intensity of the PF0124p band in the peptide growth condition compared to the maltose condition (**Figure 3.3**

and 3.4) with both PF0272 and PF0464 UORs, though the difference with PF0272 is not as dramatic as PF0464. However, microarray results do not show a significant change in expression of PF0124 between maltose and peptide growth conditions (Schut *et al.*, 2003). Considering that assays show that PF0124p represses transcription of PF0272 and PF0464, both PF0272 and PF0464 are proposed to be repressed by PF0124p when peptides are the primary carbon source. It is suspected that some factor(s) in the cell extract from maltose growth conditions may reduce the binding of PF0124p to UORs of PF0272 or PF0464. In the presence of maltose (and perhaps other sugars), the interaction of PF0124p with the regulated genes is reduced, derepressing gene transcription. This is in agreement with the results of the microarray assay since the transcription of PF0272 and PF0464 was up-regulated in maltose growth condition, apparently by derepression. But the identity of the factors that reduce the PF0124p binding to promoter DNA is unclear.

4.4.5. The Effector(s) of PF0124p

The N-terminus of PF0124p is a conserved helix-turn-helix DNA-binding domain, and the C-terminus shows similarity to the sugar-binding domain of TrmB. Therefore, PF0124p can possibly bind some effectors that affect the interaction of PF0124p and gene promoters. As both PF0272 and PF0464 were highly up-regulated by maltose, a simple assumption is that the transcriptional regulation by PF0124p is responsive to sugars, such as maltose. However, based on the EMSA analyses of PF0124p with PF0272 and PF0464 UORs, neither maltose nor sucrose induced an obvious intensity change of protein–DNA complex. PF0272 has recently been re-annotated as 4- α -glucanotransferase, a critical enzyme to produce glucose in maltose metabolism (Lee *et al.*, 2006). The direct product (glucose) of the reaction catalyzed by PF0272 could be the

effector of PF0124. However, the EMSA results with glucose did not show obvious effect on protein-DNA interaction under our experimental conditions.

The PF0124p homologous protein, TrmB, shows negative regulation function of multiple gene clusters in response to sugar effectors including trehalose, maltose, maltodextrins, etc. These effectors are not common inducers for TrmB-regulated genes, but are operon-specific (Lee *et al.*, 2003; Lee *et al.*, 2005). The structure of the sugar binding domain of *Pf* TrmB in complex with maltose was solved recently (Krug *et al.*, 2006). The residues interacting with maltose are displayed with structural details in **Figure 4.17** and are also highlighted in the alignment between PF0124 and TrmB in **Figure 4.18**. In the TrmB structure, except that S229 binds to the glucose part of maltose, all other residues in the binding sites bind to the non-reducing glucosyl part of maltose. The non-reducing end of maltose interacts with the sidechain of E326 and the main chain atoms of all other involved residues. E326 and G320 of TrmB are identical to the PF0124 counterparts and M321, V324, I325 are similar to the PF0124 counterparts. It is possible that the maltose binding sites at the non-reducing end are preserved in PF0124, though the effect of missing the N205 counterpart in PF0124 is hard to tell. However, S229 of TrmB is aligned with F241 of PF0124. The replacement could incur a severe steric hindrance to the glucose end of maltose. Therefore if PF0124 binds any sugar ligand, it could be different from those of TrmB.

Whether PF0124p is responsive to any factor(s) for regulation of multiple genes, or whether any other protein(s) are involved in the regulation need to be further explored. More questions about the regulation of *Pf* glycolytic genes by PF0124p are of great interest and require further investigation.

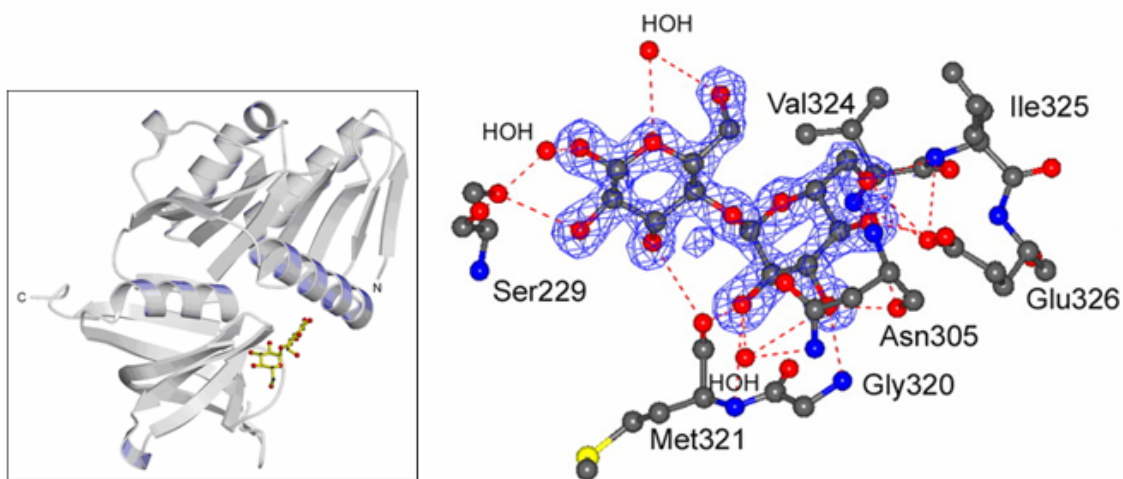


Figure 4.17 Structure of TrmB with maltose ligand. Left, overall structure of TrmB shown in cartoon with maltose shown in ball-and-stick. Right, the detailed structure of maltose and maltose interacting residues. Pictures are taken from the reference (Krug *et al.*, 2006).

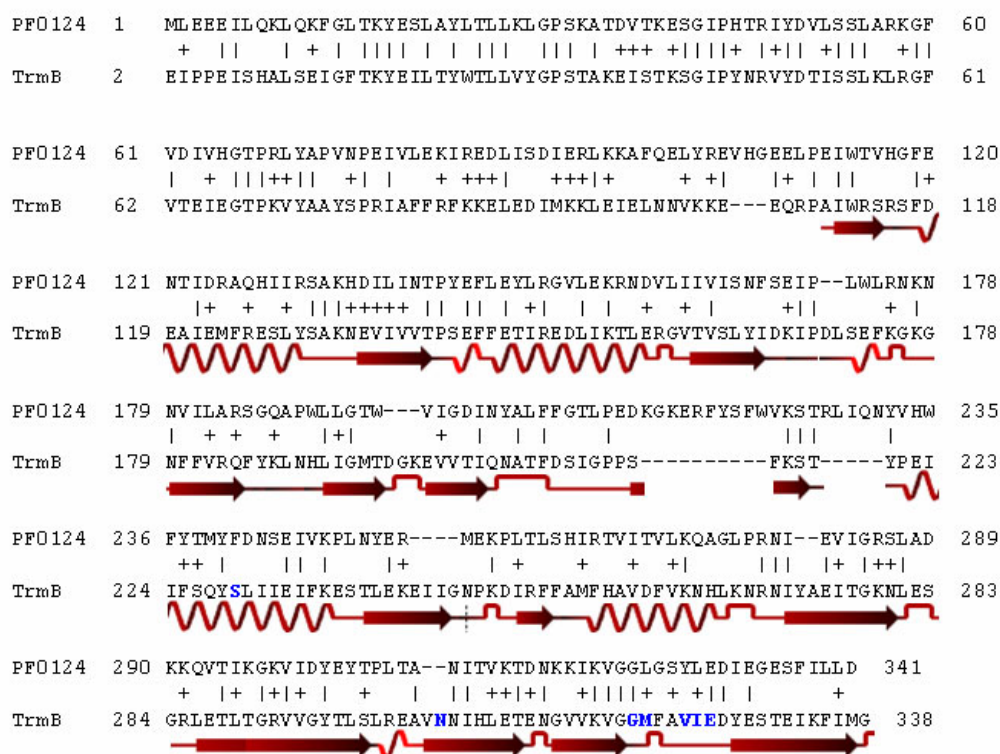


Figure 4.18 Sequence alignment between PF0124 and TrmB. The secondary structure of TrmB sugar binding domain was taken from the crystal structure and shown under the TrmB sequence. Residue S229, N305, G320, M321, V324, I325 and E326 in TrmB (shown in blue) are residues contacting maltose in the crystal structure.

REFERENCES

- Adams, M.W.W., and Kelly, R.M. (1994) Thermostability and thermoactivity of enzymes from hyperthermophilic archaea. *Bioorganic & Medicinal Chemistry* **2**: 659-667.
- Asturias, F.J. (2004) RNA polymerase II structure, and organization of the preinitiation complex. *Current Opinion in Structural Biology* **14**: 121-129.
- Bell, S.D., and Jackson, S.P. (1998a) Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol* **6**: 222-228.
- Bell, S.D., and Jackson, S.P. (1998b) Transcription in Archaea. *Cold Spring Harb Symp Quant Biol* **63**: 41-51.
- Bell, S.D., Cairns, S.S., Robson, R.L., and Jackson, S.P. (1999a) Transcriptional regulation of an archaeal operon in vivo and in vitro. *Mol Cell* **4**: 971-982.
- Bell, S.D., Kosa, P.L., Sigler, P.B., and Jackson, S.P. (1999b) Orientation of the transcription preinitiation complex in archaea. *Proc Natl Acad Sci U S A* **96**: 13662-13667.
- Bell, S.D., and Jackson, S.P. (2000) Mechanism of autoregulation by an archaeal transcriptional repressor. *J Biol Chem* **275**: 31624-31629.
- Bell, S.D., and Jackson, S.P. (2001) Mechanism and regulation of transcription in archaea. *Curr Opin Microbiol* **4**: 208-213.
- Bell, S.D. (2005) Archaeal transcriptional regulation - variation on a bacterial theme? *Trends in Microbiology* **13**: 262-265.
- Best, A.A., and Olsen, G.J. (2001) Similar subunit architecture of archaeal and eukaryal RNA polymerases. *FEMS Microbiol Lett* **195**: 85-90.

- Brinkman, A.B., Dahlke, I., Tuininga, J.E., Lammers, T., Dumay, V., de Heus, E., Lebbink, J.H., Thomm, M., de Vos, W.M., and van Der Oost, J. (2000) An Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus* is negatively autoregulated. *J Biol Chem* **275**: 38160-38169.
- Brinkman, A.B., Bell, S.D., Lebbink, R.J., de Vos, W.M., and van der Oost, J. (2002) The *Sulfolobus solfataricus* Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. *J Biol Chem* **277**: 29537-29549.
- Brinkman, A.B., Ettema, T.J., de Vos, W.M., and van der Oost, J. (2003) The Lrp family of transcriptional regulators. *Mol Microbiol* **48**: 287-294.
- Busby, S., and Ebright, R.H. (1994) Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* **79**: 743-746.
- Calvo, J.M., and Matthews, R.G. (1994) The leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* **58**: 466-490.
- Chaurand, P., Luetzenkirchen, F., and Spengler, B. (1999) Peptide and protein identification by matrix-assisted laser desorption ionization (MALDI) and MALDI-post-source decay time-of-flight mass spectrometry. *Journal of the American Society for Mass Spectrometry* **10**: 91-103.
- Chen, Y. (2003) The fishing expedition : searching for unidentified archaeal transcription factors. Thesis M S --University of Georgia 2003.
- Coligan, J.E. (1995) *Current protocols in protein science*. Brooklyn, N.Y.: Wiley.
- Cottrell, J.S. (1994) Protein identification by peptide mass fingerprinting. *Pept Res* **7**: 115-124.

- Dahlke, I., and Thomm, M. (2002) A *Pyrococcus* homolog of the leucine-responsive regulatory protein, LrpA, inhibits transcription by abrogating RNA polymerase recruitment. *Nucleic Acids Res* **30**: 701-710.
- de Wind, N., de Jong, M., Meijer, M., and Stuitje, A.R. (1985) Site-directed mutagenesis of the *Escherichia coli* chromosome near *oriC*: identification and characterization of *asnC*, a regulatory element in *E. coli* asparagine metabolism. *Nucleic Acids Res* **13**: 8797-8811.
- deHaseth, P.L., Zupancic, M.L., and Record, M.T., Jr. (1998) RNA Polymerase-Promoter Interactions: the Comings and Goings of RNA Polymerase. *J. Bacteriol.* **180**: 3019-3025.
- DiRuggiero, J., Dunn, D., Maeder, D.L., Holley-Shanks, R., Chatard, J., Horlacher, R., Robb, F.T., Boos, W., and Weiss, R.B. (2000) Evidence of recent lateral gene transfer among hyperthermophilic Archaea. *Molecular Microbiology* **38**: 684.
- Domon, B., and Aebersold, R. (2006) Mass Spectrometry and Protein Analysis. *Science* **312**: 212-217.
- DYNAL, I. (2002) *Biomagnetic Techniques in Molecular Biology, 3rd edition*.
- Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**: 64-71.
- Fenn, J.B. (2003) Electrospray wings for molecular elephants (Nobel lecture). *Angew Chem Int Ed Engl* **42**: 3871-3894.
- Fiala, G., and Stetter, K.O. (1986) *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100 °C. *Archives of Microbiology* **145**: 56-61.

- Ficarro, S.B., McClelland, M.L., Stukenberg, P.T., Burke, D.J., Ross, M.M., Shabanowitz, J., Hunt, D.F., and White, F.M. (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotech* **20**: 301-305.
- Fujii, K., Kondo, T., Yokoo, H., Okano, T., Yamada, M., Yamada, T., Iwatsuki, K., and Hirohashi, S. (2006) Database of two-dimensional polyacrylamide gel electrophoresis of proteins labeled with CyDye DIGE Fluor saturation dye. *Proteomics* **6**: 1640-1653.
- Gabrielsen, O.S., and Huet, J. (1993) Magnetic DNA affinity purification of yeast transcription factor. *Methods Enzymol* **218**: 508-525.
- Gadgil, H., Oak, S.A., and Jarrett, H.W. (2001) Affinity purification of DNA-binding proteins. *Journal of Biochemical and Biophysical Methods* **49**: 607-624.
- Geiduschek, E.P., and Ouhammouch, M. (2005) Archaeal transcription and its regulators. *Molecular Microbiology* **56**: 1397.
- Gevaert, K., Demol, H., Martens, L., Hoorelbeke, B., Puype, M., Goethals, M., Van Damme, J., De Boeck, S., and Vandekerckhove, J. (2001) Protein identification based on matrix assisted laser desorption/ionization-post source decay-mass spectrometry. *Electrophoresis* **22**: 1645-1651.
- Gharahdaghi, F., Weinberg, C.R., Meagher, D.A., Imai, B.S., and Mische, S.M. (1999) Mass spectrometric identification of proteins from silver-stained polyacrylamide gel: A method for the removal of silver ions to enhance sensitivity. *Electrophoresis* **20**: 601-605.
- Gourse, R.L., Ross, W., and Gaal, T. (2000) UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Molecular Microbiology* **37**: 687-695.

- Gross, C.A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J., and Young, B. (1998) The Functional and Regulatory Roles of Sigma Factors in Transcription. *Cold Spring Harbor Symposia on Quantitative Biology* **63**: 141-156.
- Guo, L., Eisenman, J.R., Mahimkar, R.M., Peschon, J.J., Paxton, R.J., Black, R.A., and Johnson, R.S. (2002) A proteomic approach for the identification of cell-surface proteins shed by metalloproteases. *Mol Cell Proteomics* **1**: 30-36.
- Hanzelka, B.L., Darcy, T.J., and Reeve, J.N. (2001) TFE, an archaeal transcription factor in *Methanobacterium thermoautotrophicum* related to eucaryal transcription factor TFIIEalpha. *J Bacteriol* **183**: 1813-1818.
- Hausner, W., Wettach, J., Hethke, C., and Thomm, M. (1996) 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* **271**: 30144-30148.
- Hethke, C., Geerling, A.C., Hausner, W., de Vos, W.M., and Thomm, M. (1996) A cell-free transcription system for the hyperthermophilic archaeon *Pyrococcus furiosus*. *Nucleic Acids Res* **24**: 2369-2376.
- Hethke, C., Bergerat, A., Hausner, W., Forterre, P., and Thomm, M. (1999) Cell-Free Transcription at 95{degrees}: Thermostability of Transcriptional Components and DNA Topology Requirements of *Pyrococcus* Transcription. *Genetics* **152**: 1325-1333.
- Hu, S., Michels, D.A., Fazal, M.A., Ratisoontorn, C., Cunningham, M.L., and Dovichi, N.J. (2004) Capillary sieving electrophoresis/micellar electrokinetic capillary chromatography for two-dimensional protein fingerprinting of single mammalian cells. *Analytical Chemistry* **76**: 4044-4049.

- Jungblut, P., and Thiede, B. (1997) Protein identification from 2-DE gels by MALDI mass spectrometry. *Mass Spectrom Rev* **16**: 145-162.
- Kadonaga, J.T., and Tjian, R. (1986) Affinity Purification of Sequence-Specific DNA Binding Proteins. *PNAS* **83**: 5889-5893.
- Kasicka, V. (2006) Recent developments in capillary electrophoresis and capillary electrochromatography of peptides. *Electrophoresis* **27**: 142-175.
- Kaur, A., Pan, M., Meislin, M., Facciotti, M.T., El-Gewely, R., and Baliga, N.S. (2006) A systems view of haloarchaeal strategies to withstand stress from transition metals. *Genome Res.* **16**: 841-854.
- Kolling, R., and Lotter, H. (1985) AsnC: an autogenously regulated activator of asparagine synthetase A transcription in Escherichia coli. *J Bacteriol* **164**: 310-315.
- Koning, S.M., Konings, W.N., and Driessen, A.J. (2002) Biochemical evidence for the presence of two alpha-glucoside ABC-transport systems in the hyperthermophilic archaeon Pyrococcus furiosus. *Archaea* **1**: 19-25.
- Krug, M., Lee, S.J., Diederichs, K., Boos, W., and Welte, W. (2006) Crystal structure of the sugar binding domain of the archaeal transcriptional regulator TrmB. *J Biol Chem* **281**: 10976-10982.
- Kyrpides, N.C., and Ouzounis, C.A. (1995) The eubacterial transcriptional activator Lrp is present in the archaeon Pyrococcus furiosus. *Trends Biochem Sci* **20**: 140-141.
- Kyrpides, N.C., and Ouzounis, C.A. (1999) Transcription in archaea. *Proc Natl Acad Sci U S A* **96**: 8545-8550.
- Lane, C.S. (2005) Mass spectrometry-based proteomics in the life sciences. *Cellular and Molecular Life Sciences (CMLS)* **62**: 848-869.

- Latchman, D.S. (1996) Transcription-factor mutations and disease. *N Engl J Med* **334**: 28-33.
- Lee, H.-S., Shockley, K.R., Schut, G.J., Connors, S.B., Montero, C.I., Johnson, M.R., Chou, C.-J., Bridger, S.L., Wigner, N., Brehm, S.D., Jenney, F.E., Jr., Comfort, D.A., Kelly, R.M., and Adams, M.W.W. (2006) Transcriptional and Biochemical Analysis of Starch Metabolism in the Hyperthermophilic Archaeon *Pyrococcus furiosus*. *J. Bacteriol.* **188**: 2115-2125.
- Lee, S.-J., Engelmann, A., Horlacher, R., Qu, Q., Vierke, G., Hebbeln, C., Thomm, M., and Boos, W. (2003) TrmB, a Sugar-specific Transcriptional Regulator of the Trehalose/Maltose ABC Transporter from the Hyperthermophilic Archaeon *Thermococcus litoralis*. *J. Biol. Chem.* **278**: 983-990.
- Lee, S.-J., Moulakakis, C., Koning, S.M., Hausner, W., Thomm, M., and Boos, W. (2005) TrmB, a sugar sensing regulator of ABC transporter genes in *Pyrococcus furiosus* exhibits dual promoter specificity and is controlled by different inducers. *Molecular Microbiology* **57**: 1797.
- Lee, W.-C., and Lee, K.H. (2004) Applications of affinity chromatography in proteomics. *Analytical Biochemistry* **324**: 1-10.
- Lie, T.J., and Leigh, J.A. (2003) A novel repressor of *nif* and *glnA* expression in the methanogenic archaeon *Methanococcus maripaludis*. *Molecular Microbiology* **47**: 235-246.
- Lie, T.J., Wood, G.E., and Leigh, J.A. (2005) Regulation of *nif* Expression in *Methanococcus maripaludis*: ROLES OF THE EURYARCHAEAL REPRESSOR NrpR, 2-OXOGLUTARATE, AND TWO OPERATORS. *J. Biol. Chem.* **280**: 5236-5241.

- Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., and Yates, J.R. (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotech* **17**: 676-682.
- Liolios, K., Tavernarakis, N., Hugenholtz, P., and Kyrpides, N.C. (2006) The Genomes On Line Database (GOLD) v.2: a monitor of genome projects worldwide. *Nucleic Acids Res* **34**: D332-334.
- Littlefield, O., Korkhin, Y., and Sigler, P.B. (1999) The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *PNAS* **96**: 13668-13673.
- Mukund, S., and Adams, M.W.W. (1995) Glyceraldehyde-3-phosphate Ferredoxin Oxidoreductase, a Novel Tungsten-containing Enzyme with a Potential Glycolytic Role in the Hyperthermophilic Archaeon *Pyrococcus furiosus*. *J. Biol. Chem.* **270**: 8389-8392.
- Napoli, A., van der Oost, J., Sensen, C.W., Charlebois, R.L., Rossi, M., and Ciaramella, M. (1999) An Lrp-Like Protein of the Hyperthermophilic Archaeon *Sulfolobus solfataricus* Which Binds to Its Own Promoter. *J. Bacteriol.* **181**: 1474-1480.
- Nesvizhskii, A.I., and Aebersold, R. (2004) Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS. *Drug Discovery Today* **9**: 173-181.
- Newman, E.B., and Lin, R. (1995) Leucine-Responsive Regulatory Protein: A Global Regulator of Gene Expression in *E. Coli*. *Annual Review of Microbiology* **49**: 747-775.
- Olman, V., Xu, D., and Xu, Y. (2003) CUBIC: identification of regulatory binding sites through data clustering. *J Bioinform Comput Biol* **1**: 21-40.

- Ouhammouch, M., and Geiduschek, E.P. (2001) A thermostable platform for transcriptional regulation: the DNA-binding properties of two Lrp homologs from the hyperthermophilic archaeon *Methanococcus jannaschii*. *The EMBO Journal* **20**: 146-156.
- Ouhammouch, M., Dewhurst, R.E., Hausner, W., Thomm, M., and Geiduschek, E.P. (2003) Activation of archaeal transcription by recruitment of the TATA-binding protein. *Proc Natl Acad Sci U S A* **100**: 5097-5102.
- Ouhammouch, M. (2004) Transcriptional regulation in Archaea. *Curr Opin Genet Dev* **14**: 133-138.
- Ouhammouch, M., and Geiduschek, E.P. (2005) An expanding family of archaeal transcriptional activators. *PNAS* **102**: 15423-15428.
- Peeters, E., Thia-Toong, T.-L., Gigot, D., Maes, D., and Charlier, D. (2004) Ss-LrpB, a novel Lrp-like regulator of *Sulfolobus solfataricus* P2, binds cooperatively to three conserved targets in its own control region. *Molecular Microbiology* **54**: 321-336.
- Peng, J., Schwartz, D., Elias, J.E., Thoreen, C.C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D., and Gygi, S.P. (2003) A proteomics approach to understanding protein ubiquitination. *Nat Biotech* **21**: 921-926.
- Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**: 3551-3567.
- Poole, F.L., II, Gerwe, B.A., Hopkins, R.C., Schut, G.J., Weinberg, M.V., Jenney, F.E., Jr., and Adams, M.W.W. (2005) Defining Genes in the Genome of the Hyperthermophilic Archaeon *Pyrococcus furiosus*: Implications for All Microbial Genomes. *J. Bacteriol.* **187**: 7325-7332.

- Rappsilber, J., Siniossoglou, S., Hurt, E.C., and Mann, M. (2000) A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Anal Chem* **72**: 267-275.
- Reeve, J.N. (2003) Archaeal chromatin and transcription. *Molecular Microbiology* **48**: 587-598.
- Robb, F.T., Maeder, D.L., Brown, J.R., DiRuggiero, J., Stump, M.D., Yeh, R.K., Weiss, R.B., and Dunn, D.M. (2001) Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology. *Methods Enzymol* **330**: 134-157.
- Rohlin, L., Trent, J.D., Salmon, K., Kim, U., Gunsalus, R.P., and Liao, J.C. (2005) Heat Shock Response of *Archaeoglobus fulgidus*. *J. Bacteriol.* **187**: 6046-6057.
- Ronimus, R.S., and Morgan, H.W. (2003) Distribution and phylogenies of enzymes of the Embden-Meyerhof-Parnas pathway from archaea and hyperthermophilic bacteria support a gluconeogenic origin of metabolism. *Archaea* **1**: 199-221.
- Scholz, S., Sonnenbichler, J., Schafer, W., and Hensel, R. (1992) Di-myo-inositol-1,1'-phosphate: a new inositol phosphate isolated from *Pyrococcus woesei*. *FEBS Lett* **306**: 239-242.
- Schut, G.J., Brehm, S.D., Datta, S., and Adams, M.W.W. (2003) Whole-Genome DNA Microarray Analysis of a Hyperthermophile and an Archaeon: *Pyrococcus furiosus* Grown on Carbohydrates or Peptides. *J. Bacteriol.* **185**: 3935-3947.
- Shadpour, H., and Soper, S.A. (2006) Two-dimensional electrophoretic separation of proteins using poly(methyl methacrylate) microchips. *Analytical Chemistry* **78**: 3519-3527.
- Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. *Anal. Chem.* **68**: 850-858.

- Shockley, K.R., Ward, D.E., Chhabra, S.R., Conners, S.B., Montero, C.I., and Kelly, R.M. (2003) Heat Shock Response by the Hyperthermophilic Archaeon *Pyrococcus furiosus*. *Appl. Environ. Microbiol.* **69**: 2365-2371.
- Sims, R.J., III, Mandal, S.S., and Reinberg, D. (2004) Recent highlights of RNA-polymerase-II-mediated transcription. *Current Opinion in Cell Biology* **16**: 263-271.
- Soppa, J. (1999) Transcription initiation in Archaea: facts, factors and future aspects. *Mol Microbiol* **31**: 1295-1305.
- Soppa, J. (2001) Basal and regulated transcription in archaea. *Adv Appl Microbiol* **50**: 171-217.
- Spengler, B., Kirsch, D., Kaufmann, R., and Jaeger, E. (1992) Peptide sequencing by matrix-assisted laser-desorption mass spectrometry. *Rapid Commun Mass Spectrom* **6**: 105-108.
- Struhl, K. (1999) Fundamentally Different Logic of Gene Regulation in Eukaryotes and Prokaryotes. *Cell* **98**: 1-4.
- Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., and Matsuo, T. (1988) Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **2**: 151-153.
- Tanaka, K. (2003) The origin of macromolecule ionization by laser irradiation (Nobel lecture). *Angew Chem Int Ed Engl* **42**: 3860-3870.
- Thomm, M. (1996) Archaeal transcription factors and their role in transcription initiation. *FEMS Microbiol Rev* **18**: 159-171.
- van der Oost, J., Schut, G., Kengen, S.W.M., Hagen, W.R., Thomm, M., and de Vos, W.M. (1998) The Ferredoxin-dependent Conversion of Glyceraldehyde-3-phosphate in the Hyperthermophilic Archaeon *Pyrococcus furiosus* Represents a Novel Site of Glycolytic Regulation. *J. Biol. Chem.* **273**: 28149-28154.

- van der Oost, J., Huynen, M.A., and Verhees, C.H. (2002) Molecular characterization of phosphoglycerate mutase in archaea. *FEMS Microbiology Letters* **212**: 111-120.
- Vierke, G., Engelmann, A., Hebbeln, C., and Thomm, M. (2003) A novel archaeal transcriptional regulator of heat shock response. *J Biol Chem* **278**: 18-26.
- Villard, J. (2004) Transcription regulation and human diseases. *Swiss Med Wkly* **134**: 571-579.
- Weeks, M.E., Sinclair, J., Jacob, R.J., Saxton, M.J., Kirby, S., Jones, J., Waterfield, M.D., Cramer, R., and Timms, J.F. (2005) Stress-induced changes in the *Schizosaccharomyces pombe* proteome using two-dimensional difference gel electrophoresis, mass spectrometry and a novel integrated robotics platform. *Proteomics* **5**: 1669-1685.
- Weinberg, M.V., Schut, G.J., Brehm, S., Datta, S., and Adams, M.W.W. (2005) Cold Shock of a Hyperthermophilic Archaeon: *Pyrococcus furiosus* Exhibits Multiple Responses to a Suboptimal Growth Temperature with a Key Role for Membrane-Bound Glycoproteins. *J. Bacteriol.* **187**: 336-348.
- Wilson, D.O., Johnson, P., and McCord, B.R. (2001) Nonradiochemical DNase I footprinting by capillary electrophoresis. *ELECTROPHORESIS* **22**: 1979-1986.
- Woese, C.R., and Fox, G.E. (1977) Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms. *PNAS* **74**: 5088-5090.
- Woese, C.R., Kandler, O., and Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**: 4576-4579.
- Woychik, N.A., and Hampsey, M. (2002) The RNA Polymerase II Machinery: Structure Illuminates Function. *Cell* **108**: 453-463.

- Yaneva, M., and Tempst, P. (2003) Affinity Capture of Specific DNA-Binding Proteins for Mass Spectrometric Identification. *Anal. Chem.* **75**: 6437-6448.
- Yates, J.R. (2004) MASS SPECTRAL ANALYSIS IN PROTEOMICS. *Annual Review of Biophysics and Biomolecular Structure* **33**: 297-316.
- Zhang, H., Li, X.-j., Martin, D.B., and Aebersold, R. (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotech* **21**: 660-666.
- Zimmermann, P., and Pfeifer, F. (2003) Regulation of the expression of gas vesicle genes in *Haloferax mediterranei*: interaction of the two regulatory proteins GvpD and GvpE. *Mol Microbiol* **49**: 783-794.

APPENDIX A

INDEX OF ABBREVIATIONS

Abbreviation	Full name	First appearance
1D	one-dimensional	2.1.6, Pg28
2DE	2-dimensional electrophoresis	1.3, Pg17
<i>amy</i>	α -amylase gene	1.2.4, Pg13
AsnC	Asparagine synthase C	1.1.3, Pg7
B&W buffer	Binding & Washing buffer	2.1.1, Pg19
BRE	transcription factor B recognition element	1.1.2, Pg2
CBB	Coomassie Brilliant Blue	2.1.1, Pg21
CE	capillary electrophoresis	4.2.2.3, Pg105
CTD	C-terminal domain	1.1.2, Pg3
DPE	downstream promoter element	1.1.2, Pg2
DTT	dithiothreitol	2.1.1, Pg21
EMSA	electrophoretic mobility shift assay	4.1, Pg97
ESI	electrospray ionization	1.3, Pg14
GAPOR	glyceraldehyde-3-phosphate ferredoxin oxidoreductase	1.2.4, Pg13
<i>gdh</i>	glutamate dehydrogenase gene	4.3.4, Pg116
<i>glt</i>	4- α -glucanotransferase gene	1.2.4, Pg13
HTH	helix-turn-helix	3.3.3.1, Pg89
IMAC	immobilized metal-affinity chromatography	1.3, Pg18

IPTG	isopropyl- β -D-thiogalactoside	2.1.1, Pg21
LC	liquid chromatography	1.3, Pg17
Lrp	Leucine-responsive regulatory protein	1.1.3, Pg7
MALDI	matrix-assisted laser desorption/ionization	1.3, Pg14
MS	mass spectrometry	1.3, Pg14
MS/MS	tandem MS	1.3, Pg16
ORF	open reading frame	2.3.1, Pg22
PCR	polymerase chain reaction	2, Pg19
<i>Pf</i>	<i>Pyrococcus furiosus</i>	1.2.1, Pg10
PF0124p	protein PF0124	4.1, Pg96
<i>pgm</i>	phosphoglycerate mutase gene	1.2.4, Pg13
PIC	pre-initiation complex	1.1.2, Pg2
PMF	peptide mass fingerprinting	1.3, Pg16
PMM	peptide mass mapping	3.1.2 Pg75
PMSF	phenylmethylsulfonyl fluoride	2.1.2, Pg21
PSD	post-source decay	3.1.2, Pg76
PTM	post-translational modification	1.3, Pg17
RBS	ribosome binding site	3.2.1.1, Pg77
RNAP	RNA polymerase	1.1.1, Pg1
rTF	regulatory transcription factor	1.1.2, Pg3
SDS	sodium dodecyl sulfate	2.1.1, Pg20
SDS-PAGE	SDS Polyacrylamide Gel electrophoresis	2.1.6, Pg28
SELEX	systematic evolution of ligands by exponential enrichment	4.4.2, Pg125

STRES	survey of transcriptional response to environmental stress	1.2.1, Pg10
TATA box	"T+A"-rich TATA element	1.1.2, Pg2
TBP	TATA box binding protein	1.1.2, Pg2
TF	transcription factor	1.1.1, Pg1
TFA	trifluoroacetic acid	2.1.5.1, Pg17
TOF	time-of-flight	2.1.5.2, Pg27
<i>tpi</i>	triosephosphate isomerase gene	1.2.4, Pg 13
TrmB	transcriptional regulator of <i>mal</i> operon	1.1.3, Pg7
UBP	UOR-binding protein	3.1, Pg73
UOR	Upstream of ORF Region	2.1.3.1, Pg22