THE FISHING EXPEDITION: SEARCHING FOR UNIDENTIFIED ARCHAEAL TRANSCRIPTION FACTORS

by

YU CHEN

(Under the Direction of Robert A. Scott)

ABSTRACT

The archaeal transcription machinery shows high similarity with the eucaryal RNAP II system. Three archaeal homologues of eucaryal transcription factors, TBP (TATA box Binding Protein), TFB (Transcription Factor B) and the α subunit of TFE (Transcription Factor E) have been found. Other transcription factors are believed to participate in archaeal transcription initiation. A protocol has been developed to selectively retrieve transcription related proteins from cell extract of the archaeon *Pyrococcus furiosus* (*Pf*). A number of proteins including the *Pf* TFB were retrieved from the cell extract and identified.

INDEX WORDS: Archaea, *Pyrococcus furiosus*, transcription, transcription factors, pre-initiation complex, PIC, RNA polymerase, RNAP, Promoter DNA pull-down assay, tryptic digestion, mass spectrometry, conserved domain, CD

THE FISHING EXPEDITION: SEARCHING FOR UNIDENTIFIED ARCHAEAL TRANSCRIPTION FACTORS

by

YU CHEN

B.S., Peking University, Beijing, China 1999

A Thesis Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirement for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2003

THE FISHING EXPEDITION: SEARCHING FOR UNIDENTIFIED ARCHAEAL TRANSCRIPTION FACTORS

by

YU CHEN

Major Professor: Robert A. Scott

Committee: I. Jonathan Amster
Marly K. Eidsness

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2003

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1. Archaea

With the recognition of Archaea, cellular organisms on earth are now divided into three domains of life: Bacteria, Eucarya and Archaea [Woese *et al*., 1977, 1990; Winker *et al*., 1991].

Although archaea are now found to thrive in a variety of habitats, they are recognized mostly by their ability to survive extreme conditions such as hot springs and extremely acidic or salty aqueous environments. Archaea are single-celled organisms, sharing some characteristics with both bacteria and eucarya. Like bacteria, archaea do not have a nucleus and their genetic material is a single circular DNA molecule. Together they were once referred to as prokaryotes. However, genetic studies have distinguished archaea from bacteria and revealed a closer relationship between archaea with eucarya. Fig. 1.1 is the phylogenetic tree constructed on the basis of comparative sequence analysis of 16S rRNA genes [Woese *et al*., 1990; Winker *et al*., 1991].

1.2. Transcription in Archaea

Transcription is the synthesis of RNA under the direction of DNA used by all three domains of life (Fig. 1.2).

The synthesis of RNA is catalyzed by a group of enzymes called RNA polymerase (RNAP). Bacteria have one type of RNAP that catalyzes the synthesis of all types of RNA.

Fig. 1.1 The phylogenetic tree of cellular organisms [adopted from Doolittle, 1999; Woese, 2000]. The red highlighted branch is the group to which the organism studied in this paper belongs.

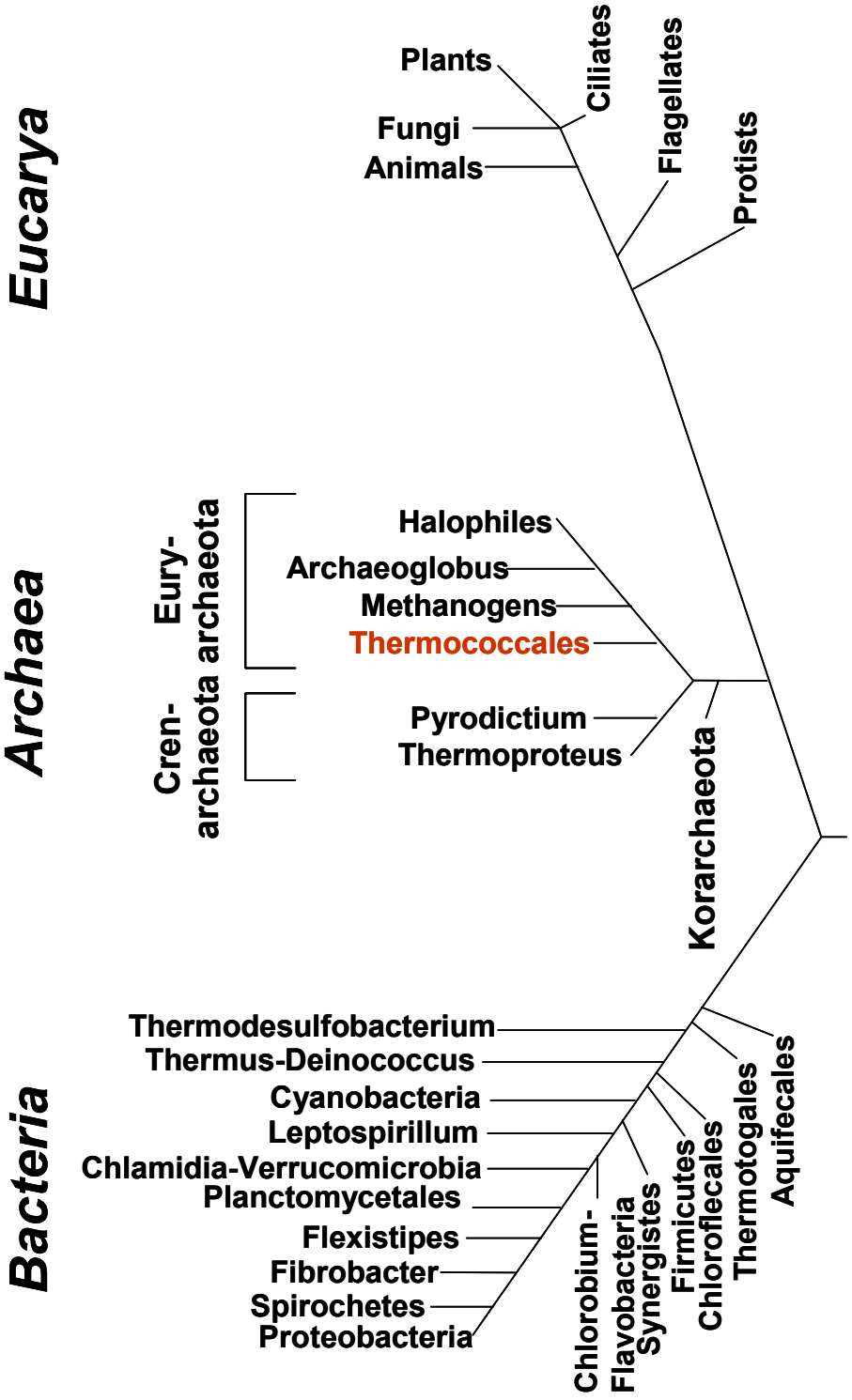
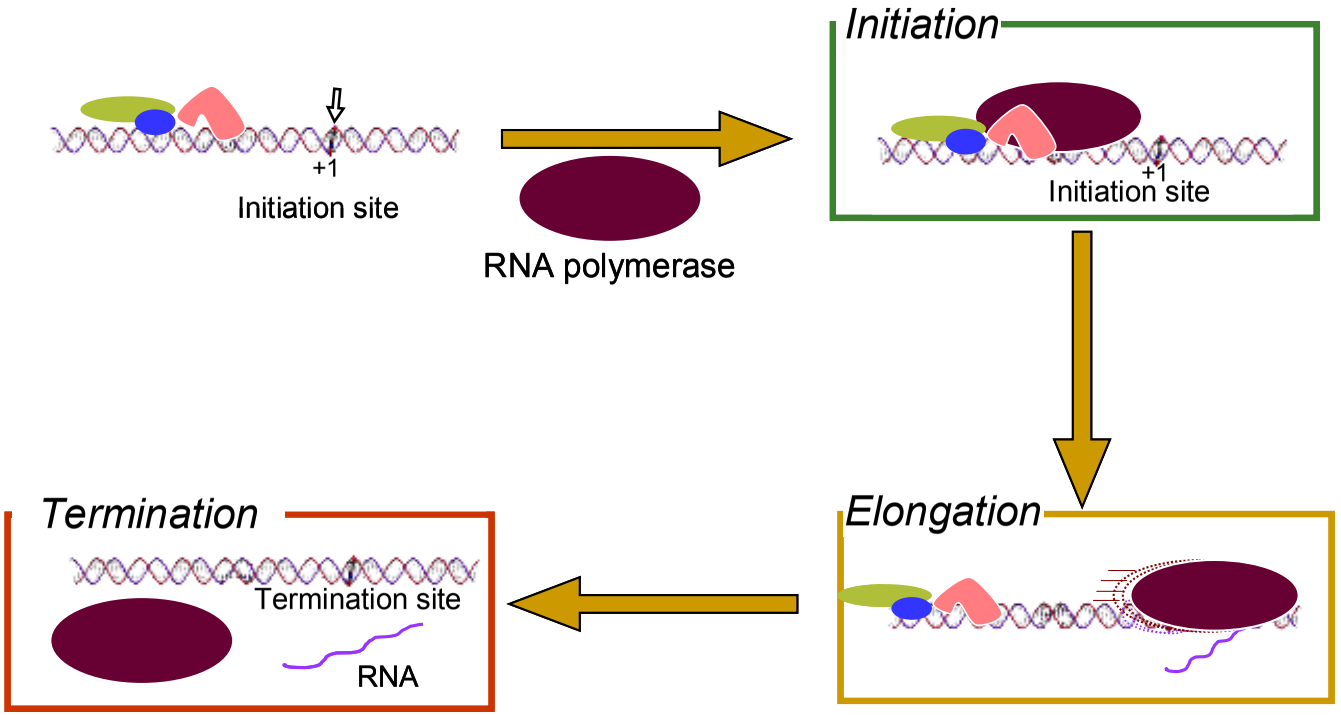

Eucarya
Archaea
Bacteria
Plants
Fungi
Animals
Ciliates
Flagellates
Protists
Halophiles
Archaeoglobus
Methanogens
Thermococcales
Eury-
archaeota
Cren-
archaeota
Pyrodictium
Thermoproteus
Korarchaeota
Thermodesulfobacterium
Thermus-Deinococcus
Cyanobacteria
Leptospirillum
Chlamidia-Verrucomicrobia
Planctomycetales
Flexistipes
Fibrobacter
Spirochetes
Proteobacteria
Chlorobium-
Flavobacteria
Synergistes
Firmicutes
Chloroflecales
Thermotogales
Aquifecales

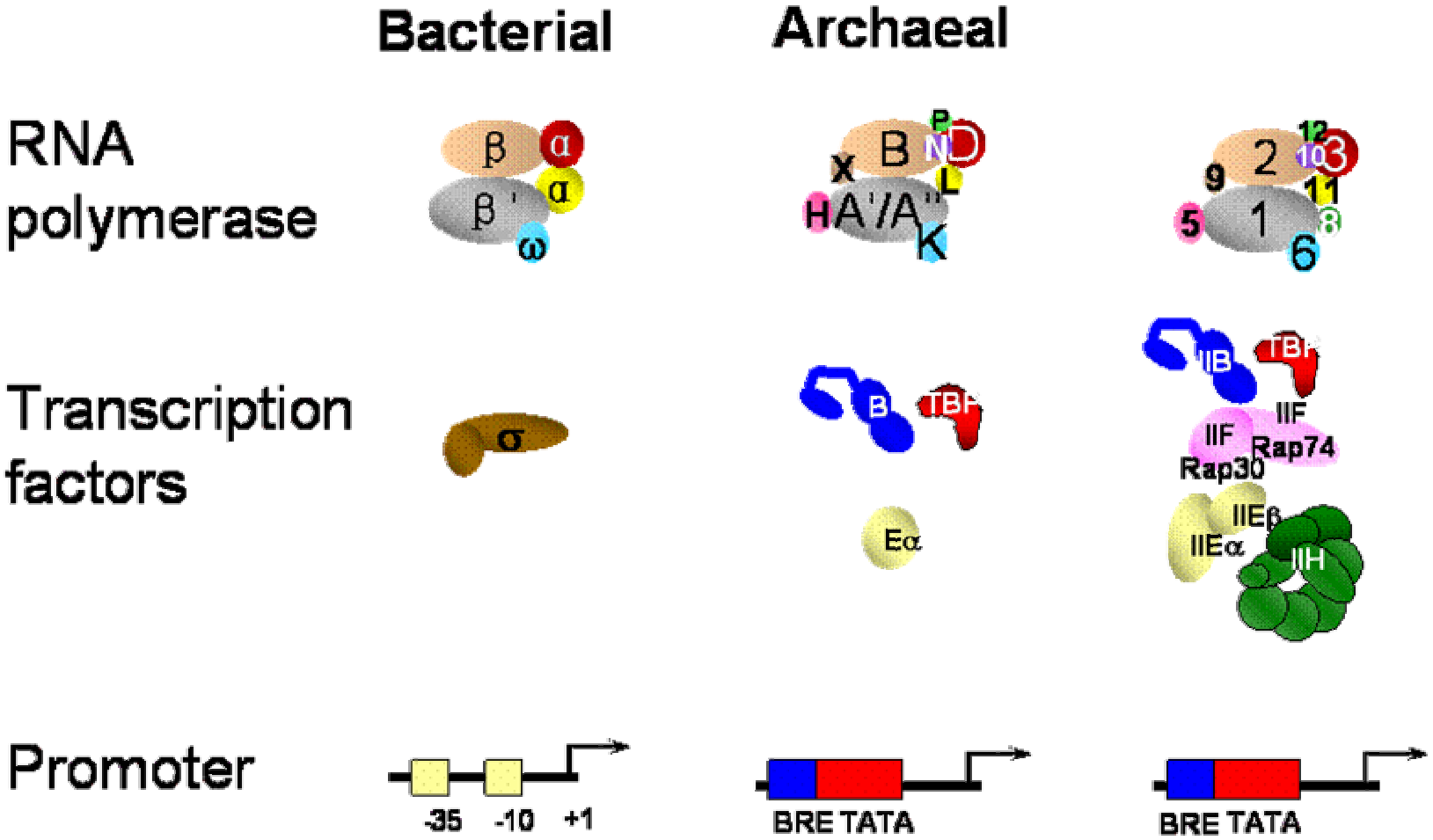
Fig. 1.2 Schematic illustration of eucaryal transcription mechanism


Initiation
+1
Initiation site
RNA polymerase
Initiation
+1
Initiation site
Elongation
Termination
Termination site
RNA

Eucarya have three types of RNAP: RNAP I, II and III for different types of RNAs (rRNA, mRNA, tRNA, respectively). Archaea, like bacteria, are equipped with only one type of RNAP. However, the archaeal transcription machinery shows high similarity with the eucaryal RNAP II system [Reviewed by Soppa, 1999, Bell *et al.*, 2001]. Fig 1.3 is a comparison of transcription systems of the three domains. The bacterial RNAP consists of four subunits ($\alpha_2\beta\beta'$). Both archaeal and eucaryal RNAP have 10-14 essential subunits including two largest subunits homologous to the β and β' subunits of bacterial RNAP. The promoter elements in archaea also resemble eucarya. A consensus TA-rich sequence similar to the TATA box in eucaryal promoters has been recognized at -25 to -30 bp relative to the transcription start site in archaeal genes. Another consensus sequence upstream of the TATA box, the transcription factor B recognition element (BRE), was also identified in archaeal promoters [Soppa, 1999; Bell *et al*., 2001]. In contrast, the bacterial promoter has two consensus sequences centered at -10 and -35 bp upstream of the transcription start site [Soppa, 1999]. These two sequences are recognized by a family of proteins called σ factor.

Proteins that assist or inhibit RNA polymerase in initiation and maintenance of transcription are called transcription factors. In bacteria, the σ factor binds to the RNAP first then the complex will recognize the promoter elements and initiate transcription [Soppa, 1999; Bell *et al*., 2001]. In archaea as well as eucarya, transcription factors bind to the promoter stepwise to form a pre-initiation complex (PIC). This PIC will then recruit the RNAP to the transcription start site [Soppa, 1999; Bell *et al*., 2001]. Although there are as many as six transcription factors in the eucaryal PIC, only three archaeal homologues of these eucarya transcription factors have been identified: TBP (TATA-box Binding Protein), TFB

Fig 1.3 Comparison of transcription machinery in bacteria, archaea and eucarya class II RNAP system. The color coding of the subunits of RNA polymerase, transcription factors and promoter elements represents their sequence homology. The same color is used to highlight homologous components in each transcription system.


Bacterial
Archaeal
RNA polymerase
Transcription factors
Promoter
β
α
β'
ω
σ
B
TBP
Eα
IIB
IIF
Rap74
Rap30
IIEβ
IIEα
IIH
-35
-10
+1
BRE TATA
BRE TATA

(Transcription Factor B) and the α subunit of TFE (Transcription Factor E). The reconstructed archaeal PIC looks like a simplified version of the eucaryal PIC, containing only TBP and TFB [Thomm, 1996]. TFE does not directly interact with the DNA; its possible function is to facilitate transcription initiation *in vivo* under sub-optimal conditions [Bell *et al*., 2000].

1.3. The Fishing Expedition

In archaea*, in vitro* transcription can be initiated with TBP, TFB, and RNAP alone from a TATA-box promoter for archaea [Soppa, 1999]. TBP, TF(II)B and RNAP are also sufficient to initiate *in vitro* transcription for eucarya [Qureshi *et al*., 1997]. However, there are other transcription factors involved in eucaryal transcription *in vivo* (TF(II)A, TF(II)F, TF(II)H, etc.). Given the facts above, we believe there are additional transcription factors in archaea. We have developed a pull-down analysis to "fish" out possible transcription factors from the cell extract of *Pyrococcus furiosus* (*Pf*) based on the technique of magnetic bead DNA affinity purification of proteins [Gabrielsen *et al*., 1989]. An immobilized promoter DNA fragment is incubated with cell extract solution. By protein-DNA or protein-protein interactions, transcription factors and other transcription related proteins are expected to bind to form a complete PIC, which is then separated from the cell extract. A number of proteins have been retrieved from the cell extract by this way, and identified by mass spectrometry. Among the identified proteins is the transcription factor B. Sequence analysis has been carried out for other identified proteins to identify their possible function during transcription.

CHAPTER 2

ELECTROPHORETIC MOBILITY-SHIFT ASSAY (EMSA) OF RECOMBINANT *PYROCOCCUS FURIOSUS* TRANSCRIPTION FACTORS WITH PROMOTER DNA FRAGMENT

2.1. Materials and Methods

DNA oligonucleotides were purchased from IDT (Integrated DNA Technology, Coralville, IA). Recombinant *Pyrococcus furiosus* (*Pf*) transcription factors (TBP, TFB, TFEα) and were expressed and purified by former Scott group members [Chen, 2000]. All other chemicals were analytical grade.

A 95-base pair double stranded DNA fragment with the sequence from -65 to +30 (relative to the transcription start site) of the *Pf* glutamate dehydrogenase gene (*Pf* gdh), which contains the TATA box and BRE, was chosen as the template. The sequence of the coding strand is:

(-65)-5'-AAACAAAAGGATTTCCACTCTTGTTTACCGAAAGCTTTATATAGGCTATTG

ACCCAAAAATGTATC|GCCAATCACCTAATTTGGAGGGATGAACAT-3'-(+30)

The promoter elements, TATA box (blue) and BRE (brown) are underlined. | marks the transcription start site.

EMSA experiments were performed by mixing together TBP, TFB and DNA fragment in EMSA reaction buffer (50 mM Tris·HCl/ 100 mM KCl/5% glycerol/ 1 mM DTT/ 1 mM EDTA/ 5 $\mu g \cdot mL^{-1}$ poly(dG·dC) / 0.2 $mg \cdot mL^{-1}$ BSA, pH=7.5). After incubation at 37ºC, 55ºC or 75ºC for 30 min, the mixtures were loaded onto 5% TBE gels (BIO-RAD, Hercules, CA)

and run in TBE buffer (89 mM Tris·HCl/ 89 mM boric acid/ 2 mM EDTA, pH 8.0). Gels were stained with ethidium bromide [Chen, 2000].

2.2. Results and Discussion

Fig. 2.1 shows the EMSA results at three different temperatures. The presence of an additional ethidium bromide stained band at an effective molecular weight of >300 bp indicates the formation of a protein-DNA complex in the presence of TBP and TFB. No shift can be observed in reactions performed at room temperature (results not shown). A shift can be seen in the reactions at 37ºC, 55ºC and 75ºC. 55ºC was thus chosen as the temperature for later experiments, a temperature high enough to activate the recombinant proteins without damaging the integrity of TFB or denaturing the DNA [Bartlett *et al*., 2000]. The effect of protein concentrations was also studied and the results are shown in Fig. 2.2. No shift was observed in reactions containing only TBP or only TFB. Neither protein can form a stable protein-DNA complex alone under the studied conditions. Complete formation of the protein-DNA complex requires protein/DNA ratios well above 1:1. For TFB, when TFB: DNA≤1, only a very faint shifted band can be seen even with an overwhelmingly high TBP concentration (lanes 3 and 4, Fig 2.2a. TBP:DNA=171:1, TFB:DNA=0.5:1 and 1:1, respectively). At TFB:DNA=3, the reaction is still not complete; unbound DNA can still be seen (TBP:DNA=171:1, lane 8, Fig. 2.2 a). When the TFB:DNA was fixed at 2:1, to see an obvious shifted band, TBP:DNA needed to be at least 1 and intensity of the shift band began to appear stronger than the unbound1 DNA band only when TBP:DNA≥5(lanes 5, 6, 7 and 8, Fig 2.2b. TFB:DNA=2:1, TBP:DNA=5:1, 10:1, 50:1 and

Fig. 2.1 Electrophoretic Mobility-Shift Assay of *Pf* TBP, *Pf* TFB and promoter DNA fragment at three different temperatures. 0.136 μM promoter DNA, 24 μM *Pf* TBP, 0.48 μM *Pf* TFB were mixed in 50 μL EMSA reaction buffer (50 mM Tris·HCl (pH 7.5 at 25ºC) / 100 mM KCl / 5% glycerol / 1mM DTT / 1mM EDTA / 5 $\mu g \cdot mL^{-1}$ poly(dG·dC) / 0.2 $mg \cdot mL^{-1}$ BSA) in each experiments. The mixtures were incubated at the indicated temperature for 30 min. The first lane shows the position of the unbound DNA at 37ºC.

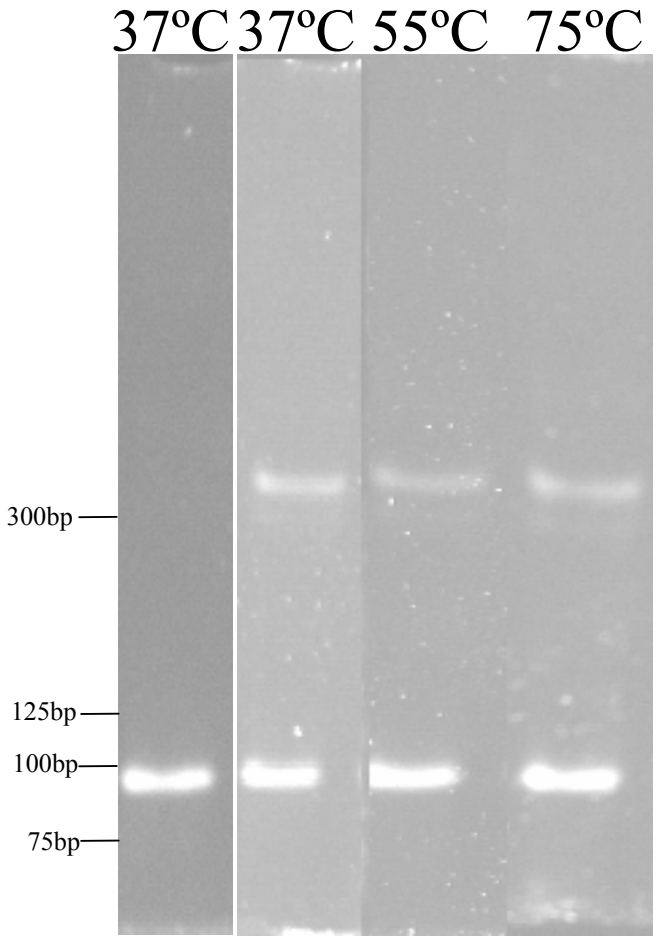
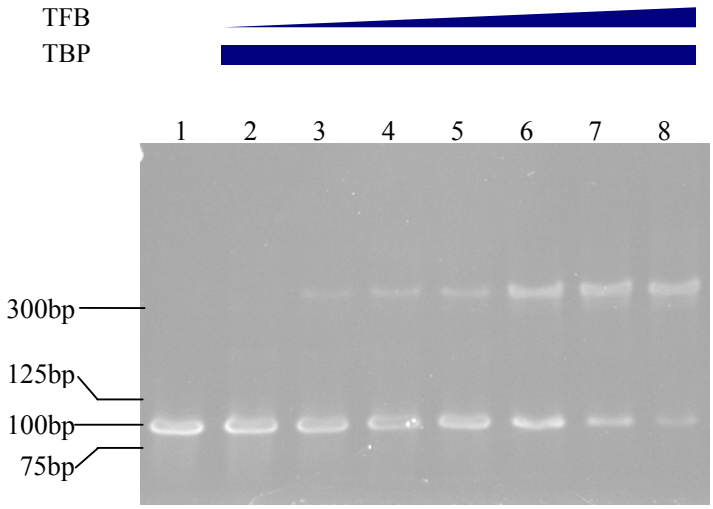
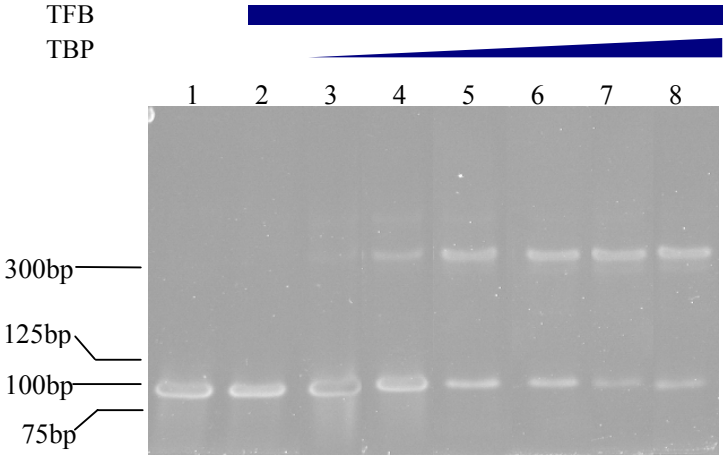

37ºC
37ºC
55ºC
75ºC
300bp
125bp
100bp
75bp

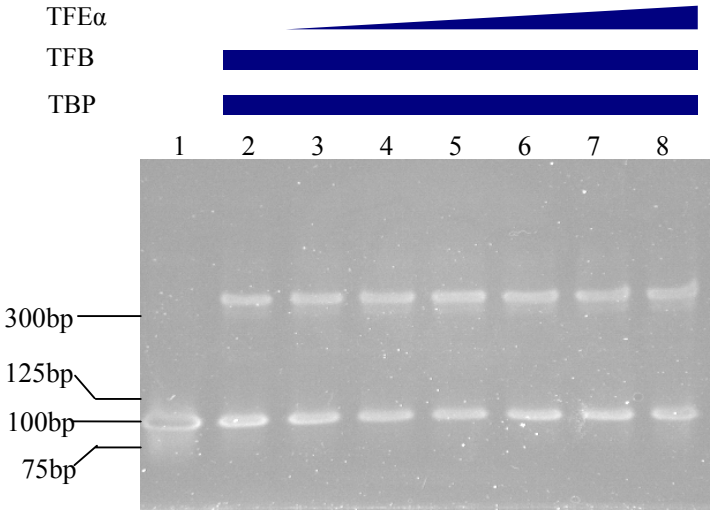
Fig. 2.2 Electrophoretic Mobility-Shift Assay of *Pf* TBP, *Pf* TFB, and *Pf* TFEα with promoter DNA fragment at different protein/DNA ratios. C_{DNA}=0.113 μM in (a) and (b), C_{DNA}=0.085 μM in (c). From lane 1 through lane 8, C_{DNA}:C_{TBP}:C_{TFB} are : (a) 1:0:0, 1:177:0, 1:177:0.5, 1:177:1.0, 1:177:1.5, 1:177:2, 1:177:2.5, 1:177:3.0; (b) 1:0:0, 1:0:2, 1:0.5:2, 1:1:2, 1:5:2, 1:10:2, 1:50:2, 1:100:2. In (c), C_{DNA}:C_{TBP}:C_{TFB}:$C_{TFE\alpha}$ from lane 1 through lane 8 are: 1:0:0:0, 1:10:2:0, 1:10:2:1, 1:10:2:2, 1:10:2:5, 1:10:2:10, 1:10:2:20, 1:10:2:35.


TFB
TBP
1 2 3 4 5 6 7 8
300bp
125bp
100bp
75bp

a


TFB
TBP
1 2 3 4 5 6 7 8
300bp
125bp
100bp
75bp

b


TFEα
TFB
TBP
1 2 3 4 5 6 7 8
300bp
125bp
100bp
75bp

c

100:1, respectively). No additional shifted band was detected, nor did the intensity of either the original shifted band or the unbound DNA band change after TFEα was introduced into the system (Fig. 2.2c). This result is consistent with previous studies on TFEα suggesting that it does not interact with DNA directly [Bell *et al*., 2000].

CHAPTER3

FISHING EXPEDITION

3.1. Experimental

3.1.1. Materials

All DNA oligonucleotides were purchased from IDT (Integrated DNA Technology, Coralville, IA). DNA oligonucleotides used for magnetic bead DNA affinity purification are all modified at the 5' end of the coding strand with biotin. Recombinant *Pyrococcus furiosus* (*Pf*) transcription factors (TBP, TFB, TFEα) and RNAP were expressed and purified by former group members [Lewis, 2000; Chen, 2000]. Dynabeads M-280 Streptavidin were purchased from Dynal Biotech (Lake Success, NY). *Pf* cells were obtained from Dr. Michael W. W. Adams' laboratory in the Department of Biochemistry and Molecular Biology, University of Georgia.

Table 3.1 lists the buffers used for the following experiments.

3.1.2. Promoter DNA Pull-down Assay

The promoter DNA pull-down assay is based on the technique of the Magnetic bead DNA affinity purification (Fig. 3.1). The whole process includes the immobilization of DNA, the binding of proteins, heparin challenge and elution of retrieved proteins from the beads.

Immobilization of DNA. Dynabeads M-280 streptavidin (referred to now as 'beads') are uniform, superparamagnetic, polystyrene beads with Streptavidin covalently attached to the bead surface. The beads were supplied as a 10 $mg \cdot mL^{-1}$ suspension containing 0.1% BSA

Fig. 3.1 Schematic illustration of magnetic DNA affinity purification of proteins. (I) The beads with immobilized DNA are incubated with protein mixture. DNA binding proteins (blue rectangles) will form a protein/DNA complex with the immobilized DNA. (II) When the binding reaction is complete, the beads together with the protein/DNA complex are attracted to a magnet. Proteins that do not bind to the DNA (ovals and triangles) remain in the supernatant and can be removed using a pipette. (III) The protein/DNA complex can be dissociated to release the proteins back into the solution. (IV) Again the beads are separated from the solution using a magnet and the DNA binding proteins in the supernatant can be collected with a pipette.

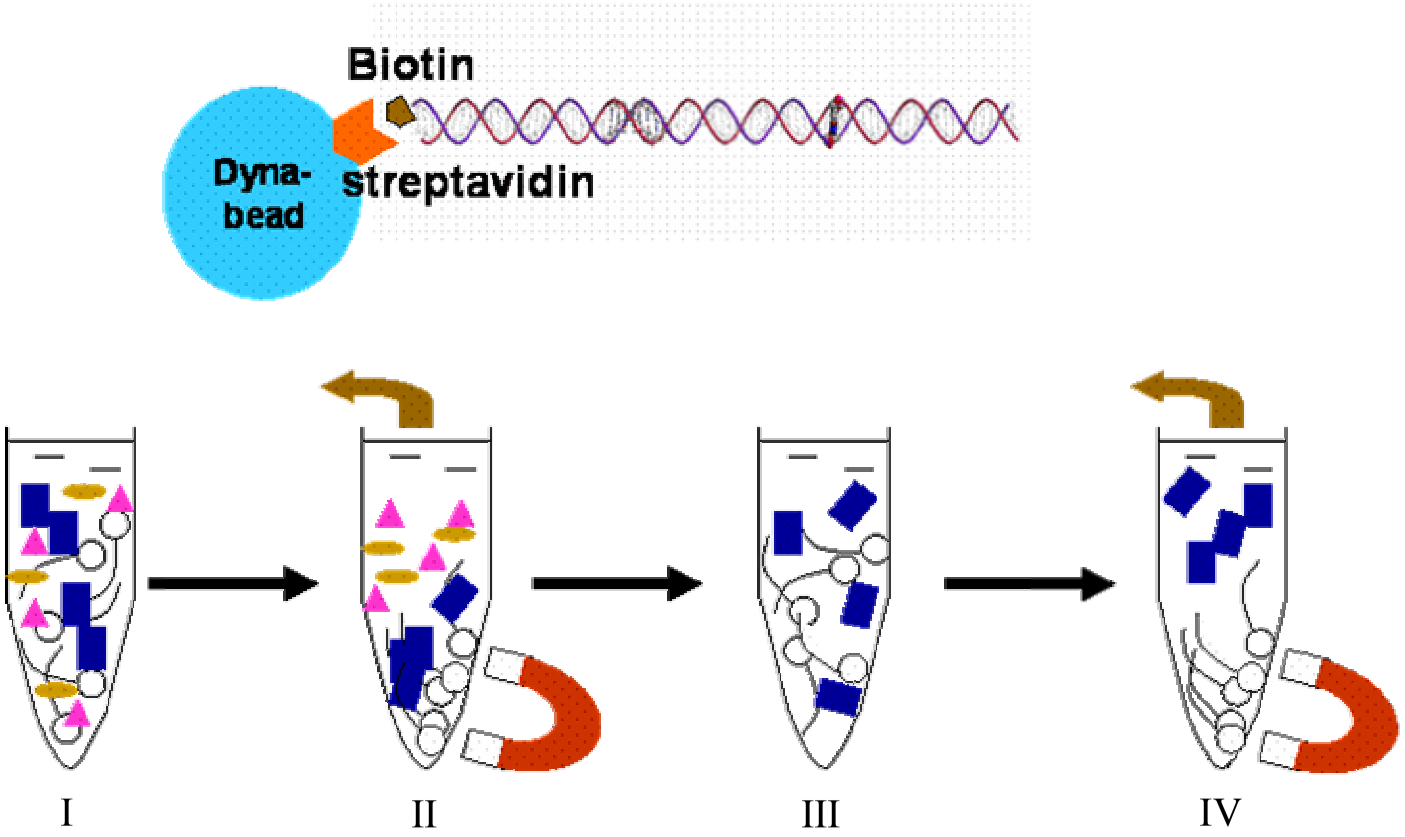

Biotin
Dyna-
bead
streptavidin
I
II
III
IV

Table 3.1 Buffer conditions. All chemicals were analytical grade.

Name	Use	Composition/Procedure
B & W	immobilization of DNA	(2x) 10 mM Tris·HCl / 1 mM EDTA / 2.0 M NaCl (pH 7.5)
Reaction Buffer	protein binding to the immobilized DNA	50 mM Tris·HCl / 100 mM KCl / 5% glycerol / 1 mM DTT / 1 mM EDTA / 0.2 mg/mL BSA / 0.1% Triton X- 100 (pH 7.5)
Washing Buffer	after the binding of proteins, washing off the excess proteins trapped on the beads	50 mM Tris·HCl / 100 mM KCl / 5% glycerol / 1 mM DTT / 1 mM EDTA / 0.1% Triton X-100 (pH 7.5)
Laemmli Buffer	gel loading buffer. Used here to dissociate the protein DNA complex	4 ml glycerol / 0.8g SDS (dissolve in 1 mL hot water) / 2.5 ml stacking gel buffer (0.5 mM Tris pH 6.8) / 10 mg Bromophenol blue, Bring volume up to 8 mL with water. Store in 1ml aliquots at -20ºC, add 0.1 mL beta-mercaptoethanol to 1 mL aliquot before use.

and 0.02% NaN_3. Biotin-modified DNA molecules and antibodies can be immobilized on the beads and separated from solutions using a magnetic particle concentrator (MPC). The following steps were used to immobilize biotin-modified dsDNA to these beads.

1. Following the manufacturer's instructions, transfer appropriate volume (normally 25 μL unless otherwise specified) of shaken bead suspension into a 1.7 mL Ependorf tube.
2. Place the tube in the MPC and wait until the beads are attracted to the back wall of the tube by the magnet (approx. 30 sec to 1 min). Remove the preserving buffer by aspiration with a pipette, avoiding touching the inside wall of the tube with the pipette tip.

3. Remove the tube from the MPC and resuspend the beads in 1x B&W buffer to a final "concentration" of 5 $\mu g \cdot \mu L^{-1}$ (half the original concentration; bead concentration is maintained at this level through out the rest of the protocol). Repeat step 2.
4. Repeat step 3 two more times to wash off the residual preserving buffer.
5. Resuspend beads in 0.125 μM biotinylated DNA solution in 1x B&W buffer, incubate at room temperature for 20 min, remove the DNA solution and wash the beads one more time with 1x B&W buffer.

Binding of the proteins. The following protocol was used to test whether proteins in solution will bind to the dsDNA immobilized on the beads.

1. After immobilization, wash the beads three times with Reaction buffer.
2. Resuspend the beads in protein solutions (of varying concentration) in Reaction buffer and incubate at 55ºC for 30 min. Agitate the tube constantly during incubation to keep the beads from gathering at the bottom.
3. Place the tube in the MPC and wait until all beads are attracted to the back wall of the tube. Remove the reaction solution, wash the beads with Washing buffer three times.

Heparin challenge. Heparin is a heterogeneous group of straight-chain anionic mucopolysaccharides composed of repeating disaccharides of uronic acid → glucosamine. It is an analog of DNA and was used here to serve as a competitor for proteins bound non-specifically or weakly to the immobilized dsDNA. The following protocol describes the process of heparin challenge.

1. Resuspend beads in 100 μg·mL^{-1} sodium heparin in Washing buffer. Incubate at 55ºC for 15 min.
2. Place the tube in the MPC and wait until all beads are attracted to the back wall of the tube. Remove the heparin solution and wash the beads one more time with the Washing buffer.

Elution of retrieved proteins from beads. Proteins can be eluted from beads by any of the following methods.

A. Resuspend beads in 1x Laemmli buffer (volume of suspension can vary according to the need of different experiments), boil at 95-100ºC for 5 min, and load the supernatant on gel.
B. Resuspend beads in 0.1% SDS water solution at desired volumes; boil at 95º-100ºC for 5-10 min.
C. Resuspend beads in guanidine isothiocyanate solution (≥2 M) in washing buffer, incubate at room temperature for 15 min.
D. Resuspend beads in solutions containing DNase I and miccrococcul nuclease. Incubate at 37ºC for 1 hr. The dsDNA will be digested and the proteins will be released into the solution.

3.1.3 Control Experiments

Several control experiments were carried out to test whether the "fishing hook" (immobilized DNA) and the "bait" (promoter elements on the DNA or the DNA-TBP-TFB complex) will effectively retrieve transcription factors from solution, while minimizing the

number of non-specifically bound proteins retrieved. The control experiments were also used to optimize the fishing protocol, probing for the best conditions for binding and dissociating the proteins.

To eliminate interference from the beads (streptavidin, etc.), an experiment was first performed on "naked beads", i.e. no DNA molecules were immobilized on beads, and the beads were incubated in Reaction buffer only; no proteins were added. The experiment includes steps 1 through 4 of the immobilization protocol and steps 1 and 2 of the protein binding protocol. Beads were finally boiled in 1x Laemmli Buffer for 5 min, and the supernatant was loaded onto a 15% gel (BIO-RAD, Hercules, CA).

To study the possible interference of DNA in the detection of retrieved proteins, experiments were carried out on beads with immobilized DNA fragments but in the absence of any added proteins (the immobilization protocol plus steps 1 and 3 of the protein binding protocol). Two dsDNA sequence were used for the "fishing" experiments: the 95bp gdh promoter (-65~+30), the same promoter fragment used in EMSA experiments but was biotin modified at the 5' end of the coding strand; and a 95bp control DNA, poly dTdC.

gdh promoter:

-65-5'-biotin-AAACAAAAGGATTTCCACTCTTGTTTACCGAAAGCTTTATATAGGCTA

TTGACCCAAAAATGTATCGCCAATCACCTAATTTGGAGGGATGAACAT-+30

Control DNA:

5'-biotin-TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCTCTCTCTCTCTCTCTTC

TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT

Recombinant TBP and TFB were used to test the formation of a protein-DNA complex immobilized on the beads. This process included the complete immobilization protocol and protein binding protocol. In step 2 of the protein binding protocol, beads with immobilized promoter DNA or control DNA were incubated with s TBP, TFB mixture in reaction buffer at 55ºC for 30 min, the concentration of the proteins was ~0.64 μM. The DNA concentration in the solution for immobilization was 0.125 μM before incubation with the beads. Assuming all DNA molecules in solution were immobilized, the mole ratio of DNA to protein during the binding reaction would be DNA:protein=1:5.1.

RNAP was added in another set of experiment. After incubation with TBP, TFB mixtures (step 2 of the protein binding protocol), the beads were then exchanged into RNAP solution in Reaction buffer, and incubated at 55ºC for another 30 min before continuing to step 3 of the protein binding protocol. Archaeal RNAP alone cannot recognize the promoter elements, although it might bind to the DNA in a non-specific manner. A heparin challenge was added to the process after the immobilization and protein binding protocol.

In the above experiment, TBP, TFB and RNAP were incubated with the beads stepwise. In cell extract, the transcription factors and RNAP co-exist with other proteins as a complicated mixture. To better simulate the actual fishing condition, the beads with immobilized DNA were incubated with TBP, TFB, TFEα and RNAP mixture in the reaction buffer in step 2 of the protein binding protocol. A heparin challenge was also included in this experiment.

For all the control experiments above, proteins were eluted by boiling the beads in 1x Laemmli buffer and were visualized with silver stain after SDS-PAGE.

3.1.4. Go Fishing (search for unidentified archaeal transcription factors using immobilized promoter DNA)

Pf cells were stored at -80ºC. To prepare cell extract solution, 3.2 g cells were mixed with 10 mL 50 mM Tris·HCl (pH 8.0). After the cells were completely thawed, the cell suspension was sonicated on ice for 5 min, and then centrifuged at 10 krpm for 30 min; the pellet was discarded. The supernatant was stored in 1 mL aliquots at -80ºC. In each experiment, 5 µL thawed supernatant was diluted to 50 µL in Reaction buffer to make the cell extract solution for binding reactions.

In one set of experiments, beads with immobilized promoter or control DNA were first incubated with recombinant TBP and TFB solution and then the cell extract solution. After heparin challenge, the beads were boiled in 1x Laemmli buffer at 95-100ºC for 5 min.

In another set of experiments, the reaction with recombinant transcription factors was omitted. Beads with immobilized promoter and control DNA were incubated with cell extract solution directly. Proteins were eluted by boiling the beads in 1x Laemli buffer at 95-100ºC for 5 min.

Eluted proteins were visualized by silver stain after SDS-PAGE.

3.1.5 Protein Identification

Proteins detected by SDS-PAGE were identified using mass spectrometry following in-gel tryptic digestion. The experiments were scaled up to get a sufficient amount of proteins for mass spectrometry analysis. 100 µL of beads was used for each experiment,

making the total reaction volume 200 μL. After heparin challenge, the beads were boiled in 40 μL 1x Laemmli buffer, and all supernatant was loaded onto the gel. Instead of silver stain, gels were stained with Sypro Ruby, which is more compatible with tryptic digestion and mass spectrometry [Lopez *et al*., 2000]. Gel plugs of 1.5 mm in diameter were cored from protein bands or unstained portions of gels for control. The gel plugs were treated with the following protocol.

1. Place gel plugs in 1.7 mL clear polypropylene tubes.
2. Incubate gel plugs with 100 μL 50 mM ammonium bicarbonate in 50% methanol for 20 min.
3. Remove methanol solution. Repeat step 2.
4. Incubate gel plugs with 100 μL 75% acetonitrile for 20 min.
5. Remove acetonitrile solution. Dry gel plugs at 40ºC for 15-20 min.

Dried gel plugs were then sent to the Proteomics Resource Facility of the Molecular Genetics Instrumentation Facility of the University of Georgia for in-gel tryptic digestion and mass spectrometric analysis.

Mass spectrometry will provide a list of fragment masses of the protein studied after tryptic digestion. The protein can then be identified using MS-Fit, which searches the protein database for matches of the experimental mass.

Fig. 3.2 summarizes the procedure of the fishing expedition.

3.2. Results and Discussion

3.2.1. The Control Experiments

Fig 3.2 Schematic illustration of the fishing experiment. The biotin modified promoter fragment is immobilized on the streptavidin coated magnetic beads. The beads are then incubated with the cell extract solution. Alternatively, the beads are first incubated with recombinant *Pf* transcription factors to form a DNA/TBP/TFB complex, and then incubated with cell extract. Transcription related proteins are expected to be attracted to the immobilized DNA or the DNA/protein complex due to protein-DNA or protein-protein interactions. These proteins will be separated from rest of the cell extract and eluted from the beads. After electrophoresis, the proteins can be identified by mass spectrometry following in-gel tryptic digestion

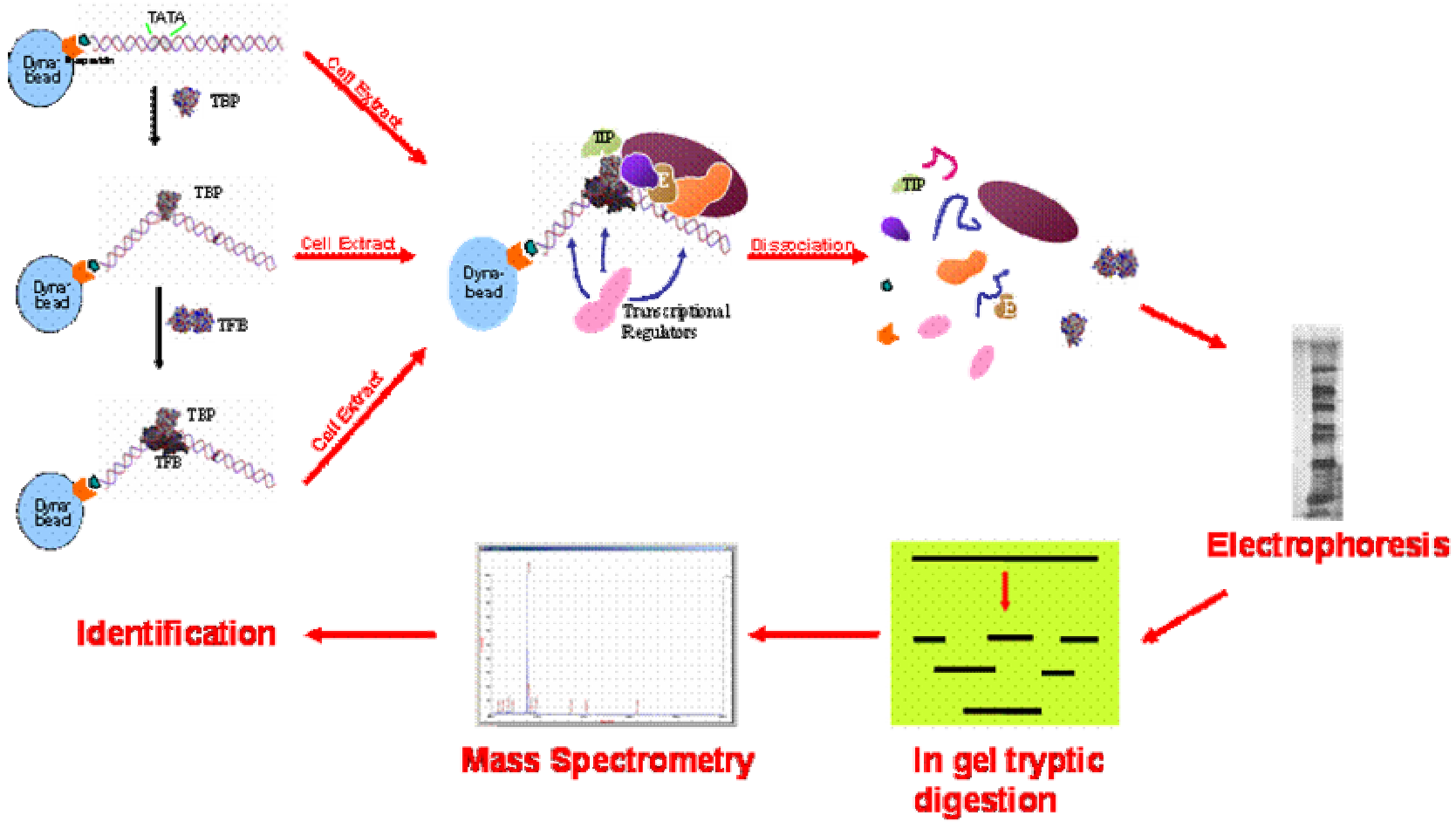
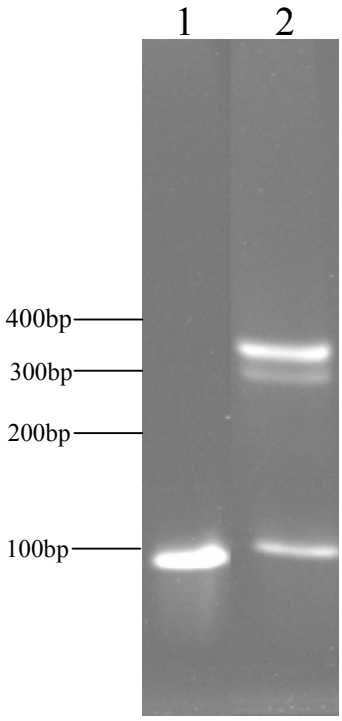

TATA
Dyna-bead
TBP
TFB
Cell Extract
Cell Extract
Cell Extract
Dyna-bead
Transcriptional Regulators
Dissociation
Electrophoresis
Identification
Mass Spectrometry
In gel tryptic digestion

Fig. 3.3A shows the EMSA results of *Pf* TBP and *Pf* TFB with the two DNA fragments selected for the fishing experiments: the *gdh* promoter fragment and the control fragment (poly dTdC). A shift is observed in the experiment with the promoter fragment but not in the one with the control fragment. This shows that the transcription factors have a much stronger affinity for the promoter sequence. Fig. 3.3B shows the immobilization of DNA molecules on Dynabeads M280 streptavidin. 1 is 0.125 μM DNA solution before incubation with beads and 2 is the same solution recovered after incubation with beads. Most of the DNA molecules in solutions were immobilized onto the beads.

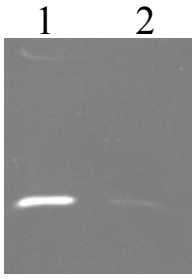
Even without incubation with any proteins, several protein bands were observed by SDS-PAGE upon the beads (Fig. 3.4A). These bands could be the streptavidin, DNA, or BSA in the storage solution or Reaction buffer. The beads have a very strong affinity to proteins in the solution. BSA is recommended by the manufacturer as the blocker for non-specific binding sites on the beads, and it is also included in the reaction buffer to minimize non-specific binding of proteins to the DNA. If BSA is not included in the reaction buffer, TBP and TFB can be seen binding in the experiment with control DNA as well as the one with promoter DNA. The protein bands in promoter DNA experiment are stronger in intensity (Fig. 3.4A.3). The amount of proteins eluted from beads with control DNA is not negligible (Fig. 3.4B). Once BSA is included, no TBP or TFB can be detected in the control DNA experiments. They can still be seen clearly in the promoter DNA experiments (Fig. 3.4C).

Fig. 3.4D shows the results of incubating bead-immobilized DNA with RNAP solution following the reaction with TBP and TFB. A number of RNAP subunits were eluted

Fig. 3.3 (A) Electrophoretic mobility shift assay of control DNA vs. promoter DNA. Experiments were carried out at 55ºC, [DNA]= 0.125 μM, DNA:TBP:TFB=1:5:4 (mole ratio). (1) Control DNA; (2) Promoter DNA. (B) Immobilization of DNA. 50 μL 0.125 μM DNA in 1x B&W buffer (1) before; (2) after incubation with Dynabeads M-280

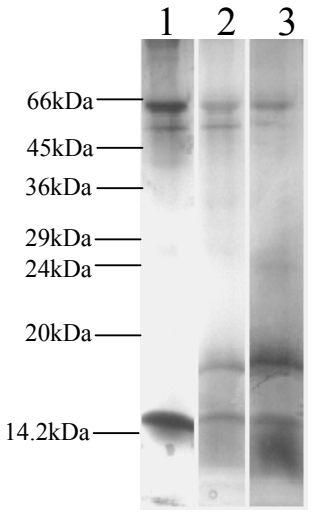


A

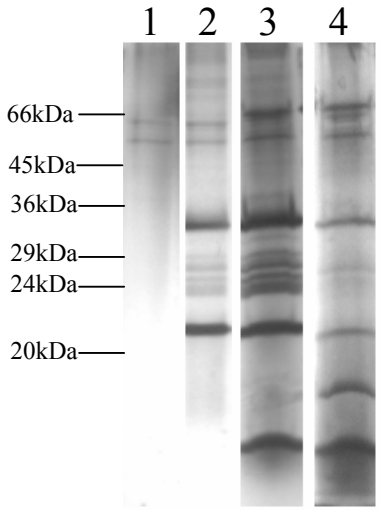


B

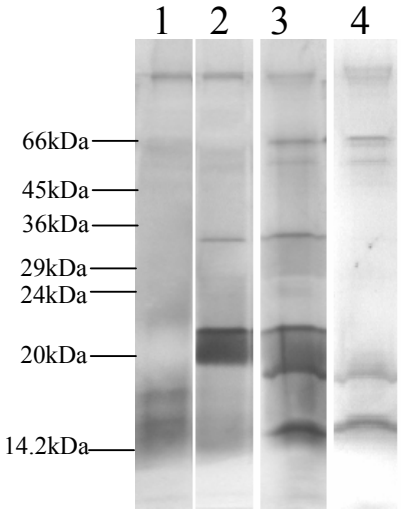
Fig. 3.4 Fishing experiments with recombinant *Pf* proteins. A: (1) "Naked" beads (no DNA molecules immobilized) boiled in 1x Laemmli buffer after incubation with the reaction buffer. (2) Beads with control DNA immobilized. (3) Beads with promoter DNA immobilized. B: (1) Washing buffer. (2) TBP and TFB in Washing buffer. (3) Proteins eluted from promoter DNA after incubation with (2). (4) Proteins eluted from control DNA after incubation with (2). C: (1) Reaction buffer. (2) TBP, TFB in Reaction buffer. (3) Proteins eluted from beads with promoter DNA immobilized after incubation with (2). (4) Proteins eluted from beads with control DNA immobilized after incubation with (2). D: (1) TBP and TFB in Reaction buffer. (2) RNAP in Reaction buffer. (3) Proteins eluted from beads with promoter DNA after incubation with TFB and TBP (1) first, then with RNAP (2). (4) Proteins eluted after two incubations from control DNA.


1 2 3
66kDa
45kDa
36kDa
29kDa
24kDa
20kDa
14.2kDa

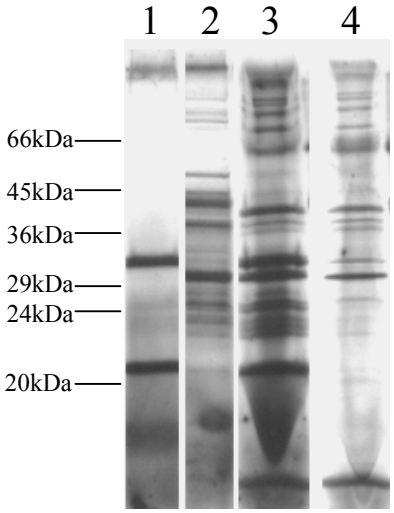
A


1 2 3 4
66kDa
45kDa
36kDa
29kDa
24kDa
20kDa

B


1 2 3 4
66kDa
45kDa
36kDa
29kDa
24kDa
20kDa
14.2kDa

C


1 2 3 4
66kDa
45kDa
36kDa
29kDa
24kDa
20kDa

D

together with TBP and TFB in the experiment with promoter DNA. However, several RNAP subunits were also detected in the elution of the incubation with control DNA. No TBP and only insignificant amounts of TFB were detected in the control DNA experiment (Fig. 3.4D.4) compared to the promoter DNA experiment (Fig.3.4D.3). RNAP may bind to the DNA without sequence specificity. A heparin challenge was then added into the protocol to strip off the non-specific binding proteins. Fig. 3.5 shows the results of the heparin challenge. Beads and DNA were incubated with TBP, TFB, TFE and RNAP in one solution together. Most of the non-specific binding proteins were challenged off the control DNA after incubation with heparin sodium solution. In the experiment with promoter DNA, besides the non-specific binding proteins, TBP and TFB were also challenged off the DNA (Fig. 3.5. 2 & 4). However, plenty of them survived to be seen clearly on gel (Fig. 3.5. 3 & 5). Although TFE does not appear to bind to the promoter or the TBP-TFB-DNA complex in the EMSA experiments, it was found among the proteins challenged off from the promoter Fig. 3.5. 2 & 4). Unlike TBP and TFB, little TFE survived after incubation with heparin solution for 20 min (Fig. 3.5. 5). The likely function of TFE is to facilitate the interaction between TBP and the TATA box thus stimulating transcription *in vivo* and there is evidence that TFE interacts with RNAP and TBP [Bell *et al*., 2000]. This result indicates that some of the proteins involved in transcription initiation may not be tightly attached to the PIC. Proteins challenged off by heparin can be of interest, too.

3.2.2. Retrieving Transcription Factors from Cell Extract.

Fig. 3.5 Heparin challenge: after the binding of proteins, the beads were incubated with 250 μg· ml^{-1} heparin sodium in Washing buffer at 55ºC for 10 min or 20 min. After the heparin solution was collected, the beads were washed one more time in Washing buffer and was finally boiled in 1x Laemmli buffer. (1) TBP, TFB, RNAP and TFE in reaction buffer. (2) Promoter DNA after incubation with proteins challenged by 250 μg· mL^{-1} heparin sodium in washing buffer. Beads were incubated with heparin solution at 55ºC for 10 min. (3) Boil beads in 1x Laemmli buffer after (2). (4) Promoter DNA challenged by 250 μg· mL^{-1} heparin sodium at 55ºC for 20 min. (5).Boil beads in 1x Laemmli buffer after (4). (6) Control DNA after incubation with proteins challenged by 250 μg· mL^{-1} heparin sodium in washing buffer. Beads were incubated with heparin solution at 55ºC for 10 min. (7) Boil beads in 1x Laemmli buffer after (6). (8) Control DNA challenged by 250 μg· mL^{-1} heparin sodium at 55ºC for 20 min. (9).Boil beads in 1x Laemmli buffer after (8).

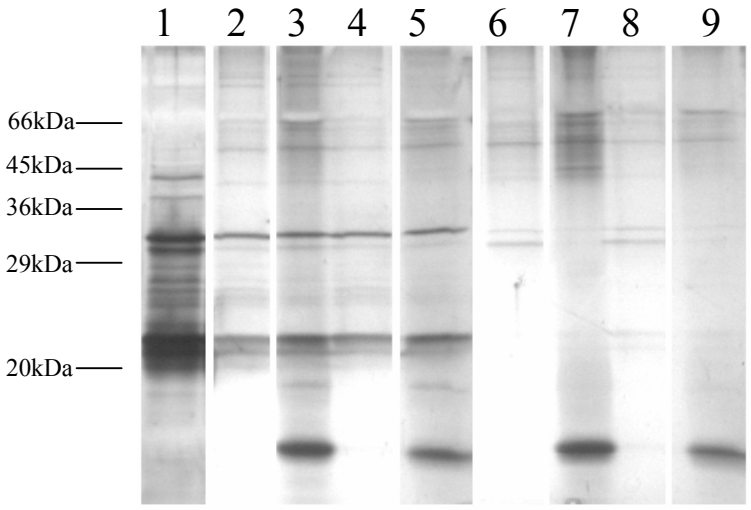
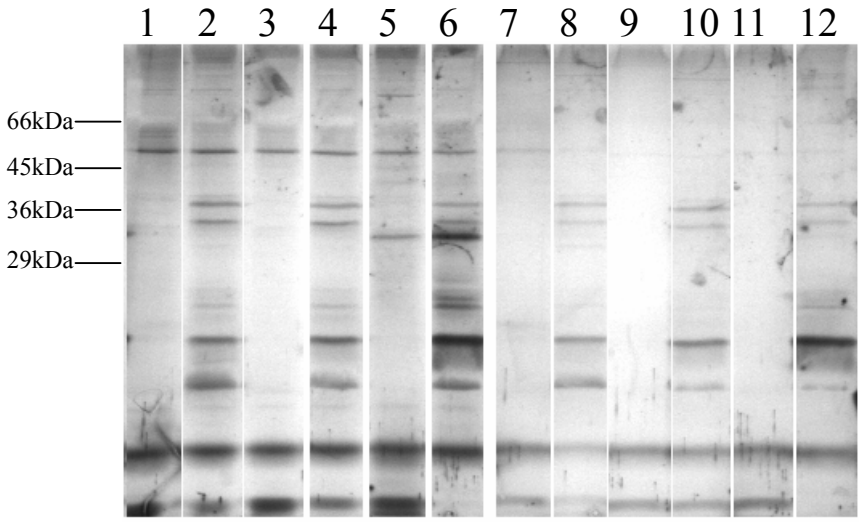

1
2
3
4
5
6
7
8
9
66kDa
45kDa
36kDa
29kDa
20kDa

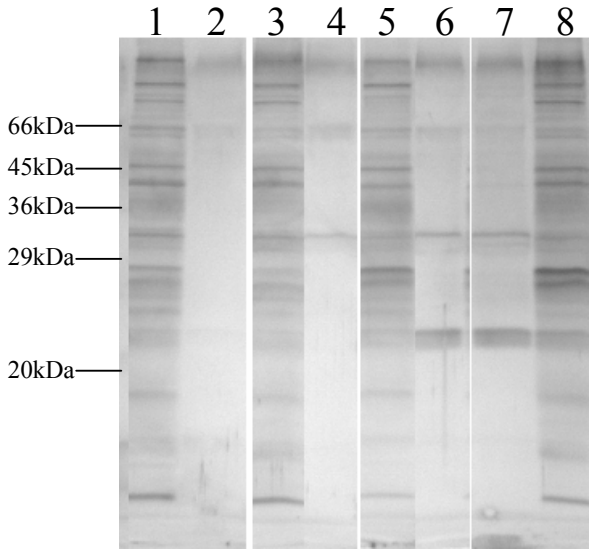
Fig. 3.6 Fishing in *Pf* cell extract: proteins eluted from control and promoter DNA after incubation with cell extract, elution was done after a heparin challenge, 250 μg· mL^{-1} heparin sodium at 55ºC for 15 min. For lanes 1 through 6, elution of the proteins was done by boiling beads in 1x Laemmli buffer. For lanes 7 through 12, elution was done by boiling beads in washing buffer at 95ºC for 5-10 min. (1) & (7) Control DNA incubated with cell extract without incubation with any of the recombinant *Pf* proteins. (2) & (8) Promoter DNA incubated with cell extract without incubation with any of the recombinant *Pf* proteins. (3) & (9) Control DNA incubated with TBP first, then cell extract. (4) & (10) Promoter DNA incubated with TBP first, then cell extract. (5) & (11) Control DNA incubated with both TBP and TFB first, then cell extract. (6) & (12) Promoter DNA incubated with both TBP and TFB first, then cell extract.


1 2 3 4 5 6 7 8 9 10 11 12
66kDa
45kDa
36kDa
29kDa

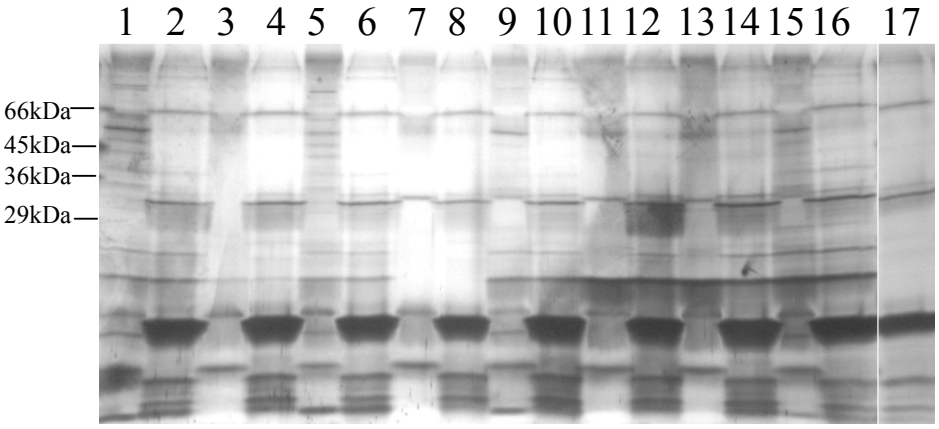
The immobilized DNA or TBP-TFB-DNA complex was incubated with whole cell extract solution (Fig. 3.6). Dozens of bands with molecular weights ranging from less than 10 kDa to over 150 kDa were observed on a gel in experiments using promoter DNA with or without recombinant transcription factors binding to the DNA prior to incubation with cell extract. The bands detected on gel are likely to be the same in the three sets of experiments, except that they appear to be slightly stronger when beads and DNA were incubated with both TBP and TFB before incubation with cell extract. Many fewer bands were detected in experiments using control DNA. The protocol works fairly well with high selectivity (compared with the abundance of proteins present in the cell extract, only a small number of proteins were retrieved).

Different elution methods have been tried (Fig 3.6, 3.7 and 3.8). Boiling beads in 1x Laemmli buffer can completely dissociate the complex. However, this method can only be used to prepare samples for 1-D electrophoresis. Replacing the Laemmli buffer with aqueous SDS can dissociate the complex with the same efficiency. SDS concentration can be kept as low as 0.1% to lower its interference in further processing. Heating up the beads in washing buffer without adding denaturing reagent cannot disscociate all the complexes formed on the beads. Elution were also done under room temperature by incubating the beads in guanidine isothiocyanate (GuSNC) solution. The guanidine isothiocyanate concentration needs to be at least 2M to achieve a complete dissociation. Using enzymes to digest DNA into pieces can dissociate the complex at relatively mild conditions without denaturing the proteins. The disadvantage is that alien proteins are introduced into the system (DNase I and micrococcal nuclease).

Fig. 3.7. A, Heparin challenge of (1) promoter DNA after incubation with cell extract; (2) Promoter DNA after incubation with TBP; (3) Promoter DNA after incubation with TBP then cell extract; (4) Promoter DNA after incubation with TFB; (5) Promoter DNA after incubation with TFB then cell extract; (6) Promoter DNA after incubation with TBP, TFB mixture; (7) Promoter DNA after incubation with TBP, TFB and RNAP mixture; (8) Promoter DNA after incubation with TBP, TFB mixture then cell extract. B: Elution of proteins after heparin challenge (A). For lanes 1, 3, 5, 7, 9, 11, 13, 15, elution was done by boiling beads in 1x Laemmli buffer. For lanes 2, 4, 6, 8, 10, 12, 14, elution was done by adding enzymes into the system to digest the DNA of the beads and release the proteins. For each experiment, 25 μL beads were used, reaction volume 50 μL during the experiment. After heparin challenge, beads were washed by washing buffer twice then resuspended in 20 μl washing buffer, then add 2.5 μL 50 mM $CaCl_2$, 0.5 μL DNase I and 0.5 μL micrococcul nuclease, incubate at 37ºC for 1 hr. (1)& (2) Promoter DNA after incubation with cell extract. (3)& (4) Promoter DNA after incubation with TBP. (5)& (6) Promoter DNA after incubation with TBP then cell extract. (7) & (8) Promoter DNA after incubation with TFB. (9) & (10) Promoter DNA after incubation with TFB then cell extract. (11) & (12) Promoter DNA after incubation with TBP, TFB mixture. (13) & (14) Promoter DNA after incubation with TBP, TFB and RNAP mixture. (15) & (16) Promoter DNA after incubation with TBP, TFB mixture then cell extract. (17) DNase I and miccrococcul nuclease in washing buffer.

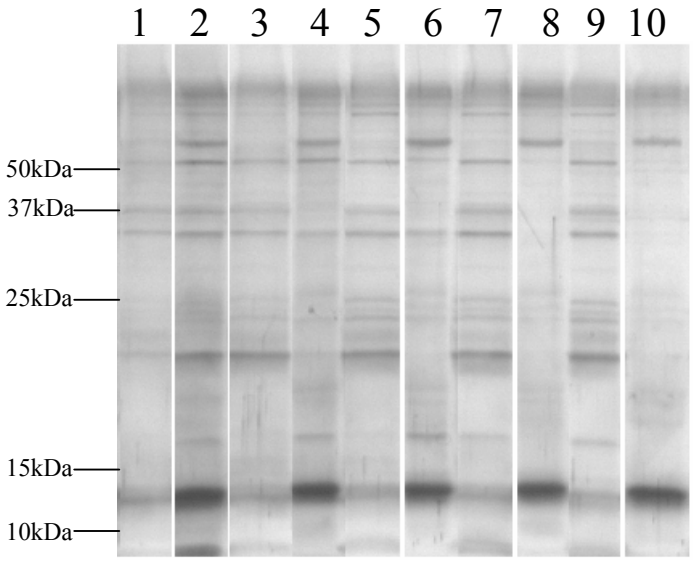

1 2 3 4 5 6 7 8
66kDa
45kDa
36kDa
29kDa
20kDa

A


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
66kDa
45kDa
36kDa
29kDa

B

Fig.3.8 Elution of proteins using guanidine isothiocyanate. Beads were resuspended in guanidine isothiocyanate solution in Washing buffer, incubated at room temperature for 15 min. Supernatant were treated with 2-D gel cleanup kit to get rid of guanidine isothiocyanate. The beads after incubation with guanidine isothiocyanate were boiled again in 1x Laemmli buffer to test if dissociation is complete. (1) 0.1 M guanidine isothiocyanate. (2) beads boiled in 1x Laemmli buffer after (1). (3) 0.5 M guanidine isothiocyanate. (4) beads boiled in 1x Laemmli buffer after (3). (5) 1.0 M guanidine isothiocyanate. (6) beads boiled in 1x Laemmli buffer after (5). (7) 2.0 M guanidine isothiocyanate. (8) beads boiled in 1x Laemmli buffer after (7). (9) 3.25 M guanidine isothiocyanate. (10) beads boiled in 1x Laemmli buffer after (9)


1 2 3 4 5 6 7 8 9 10
50kDa
37kDa
25kDa
15kDa
10kDa

The fishing experiment without recombinant transcription factors has provided enough proteins for further analysis. Fig. 3.9 is the Sypro Ruby stained gel of the proteins retrieved by control and promoter DNA from cell extract. Nine bands from both experiments with control and promoter DNA were randomly selected (marked by blue rectangles in Fig. 3.9) for in-gel tryptic digestion and mass spectrometry analysis. Fig. 3.10 though Fig. 3.17 are the mass spectra of the selected samples. Table 3.2 contains the masses selected for database search. These were selected by eliminating contaminant peaks identified in blank experiments from all well resolved peaks in the spectra of the samples. The database search results and brief description of identified proteins are listed in Table 3.3 through 3.11. In the MS-Fit results of Band VI (Table 3.8), although the first hit is a conserved hypothetical protein, the second hit, GTP cyclohydrolase II, has more masses matched and a more reasonable molecular weight compared with the results of electrophoresis. So GTP cyclohydrolase II is listed as the identity of Band VI in Table 3.11.

3.3. The “Fish". Proteins Retrieved from Promoter Pull-down Experiments.

One of the previously known archaeal transcription factors, the transcription factor B (TFB), is among the proteins retrieved by the promoter DNA fragment. This provides a positive control to suggest that retrieved proteins may be relevant.

Sequence homology studies have been carried out on the identified proteins to further study their functions and possible involvement in archaeal transcription. Conserved domains (CD) in the sequences of all the “fish” were recognized using the conserved domain search tool on the NCBI website (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

Fig. 3.9 Protein bands randomly selected for in-gel tryptic digest and mass spectrometry. Heparin challenge was done in 100 μg· mL^{-1} heparin sodium at 55ºC for 10 min. (1) Control DNA after incubation with cell extract. (2) Promoter DNA after incubation with cell extract.

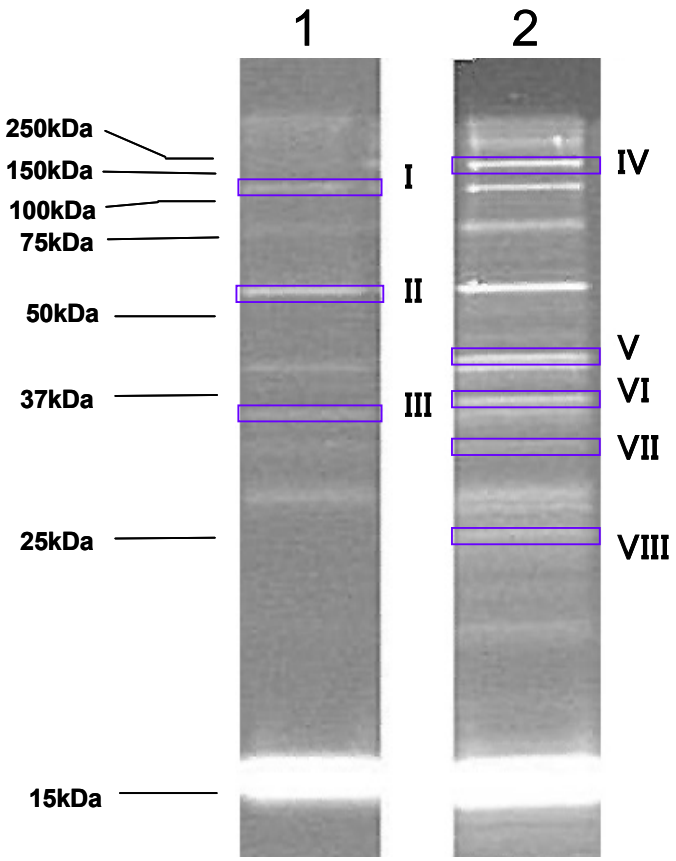

1
2
250kDa
150kDa
100kDa
75kDa
50kDa
37kDa
25kDa
15kDa
I
II
III
IV
V
VI
VII
VIII

Fig. 3.10 MS spectrum of Band I. MS experiments were performed by the Proteomics Resource Facility of the Molecular Genetics Instrumentation Facility of the University of Geargia. The instrument used is ABI 4700 TOF/TOF. The experiments performed were MOLDI-TOF mass spectrometry.

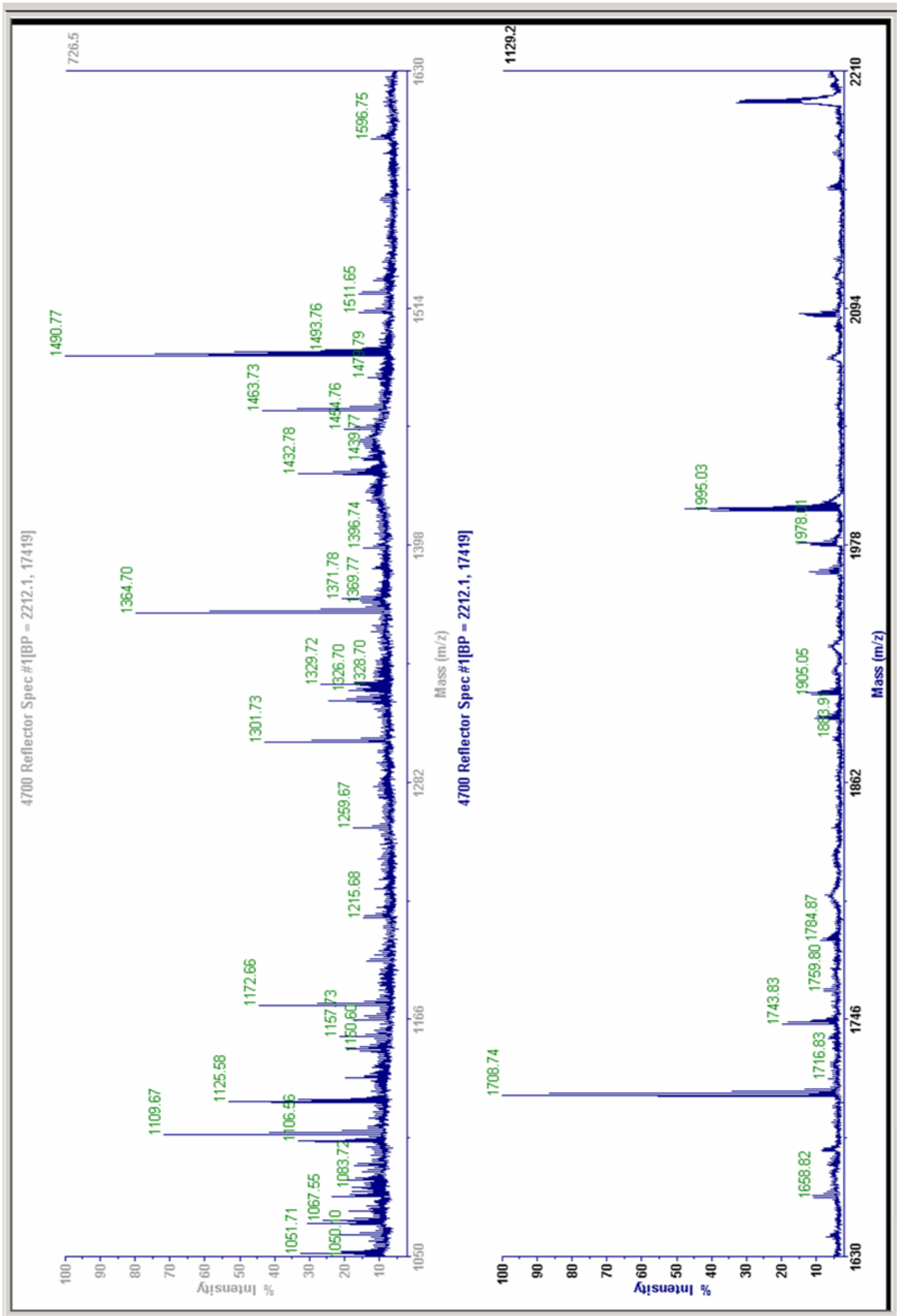

4700 Reflector Spec #1[BP = 2212.1, 17419]
% Intensity
Mass (m/z)
726.5
1490.77
1364.70
1109.67
1125.58
1172.66
1301.73
1463.73
1432.78
1596.75
1511.65
1493.76
1329.72
1371.78
1396.74
1215.68
1259.67
1051.71
1067.55
1083.72
4700 Reflector Spec #1[BP = 2212.1, 17419]
% Intensity
Mass (m/z)
1129.2
1708.74
1995.03
1743.83
1716.83
1658.82
1759.80
1784.87
1905.05
2210

Fig. 3.11 MS spectrum of Band II. MS experiments were performed by the Proteomics Resource Facility of the Molecular Genetics Instrumentation Facility of the University of Geargia. The instrument used is ABI 4700 TOF/TOF. The experiments performed were MOLDI-TOF mass spectrometry.

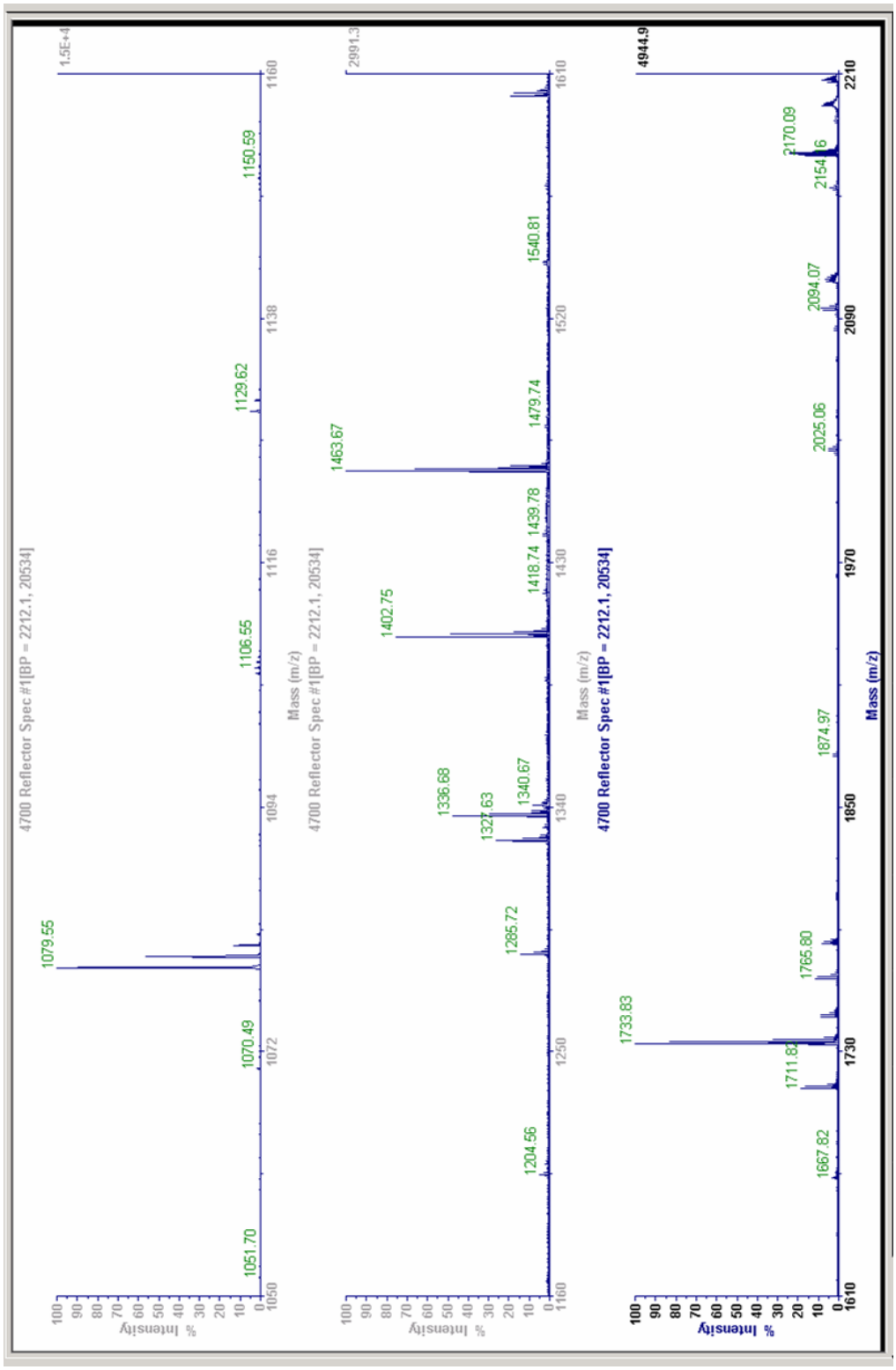

4700 Reflector Spec #1[BP = 2212.1, 20534]
1.5E+4
1079.55
1051.70
1070.49
1106.55
1129.62
1150.59
% Intensity
Mass (m/z)
4700 Reflector Spec #1[BP = 2212.1, 20534]
2991.3
1463.67
1402.75
1336.68
1327.63
1340.67
1285.72
1204.56
1418.74
1439.78
1479.74
1540.81
% Intensity
Mass (m/z)
4700 Reflector Spec #1[BP = 2212.1, 20534]
4944.9
1733.83
1765.80
1711.83
1667.82
1874.97
2025.06
2094.07
2154.6
2170.09
% Intensity
Mass (m/z)

Fig. 3.12 MS spectrum of Band III. MS experiments were performed by the Proteomics Resource Facility of the Molecular Genetics Instrumentation Facility of the University of Geargia. The instrument used is ABI 4700 TOF/TOF. The experiments performed were MOLDI-TOF mass spectrometry.

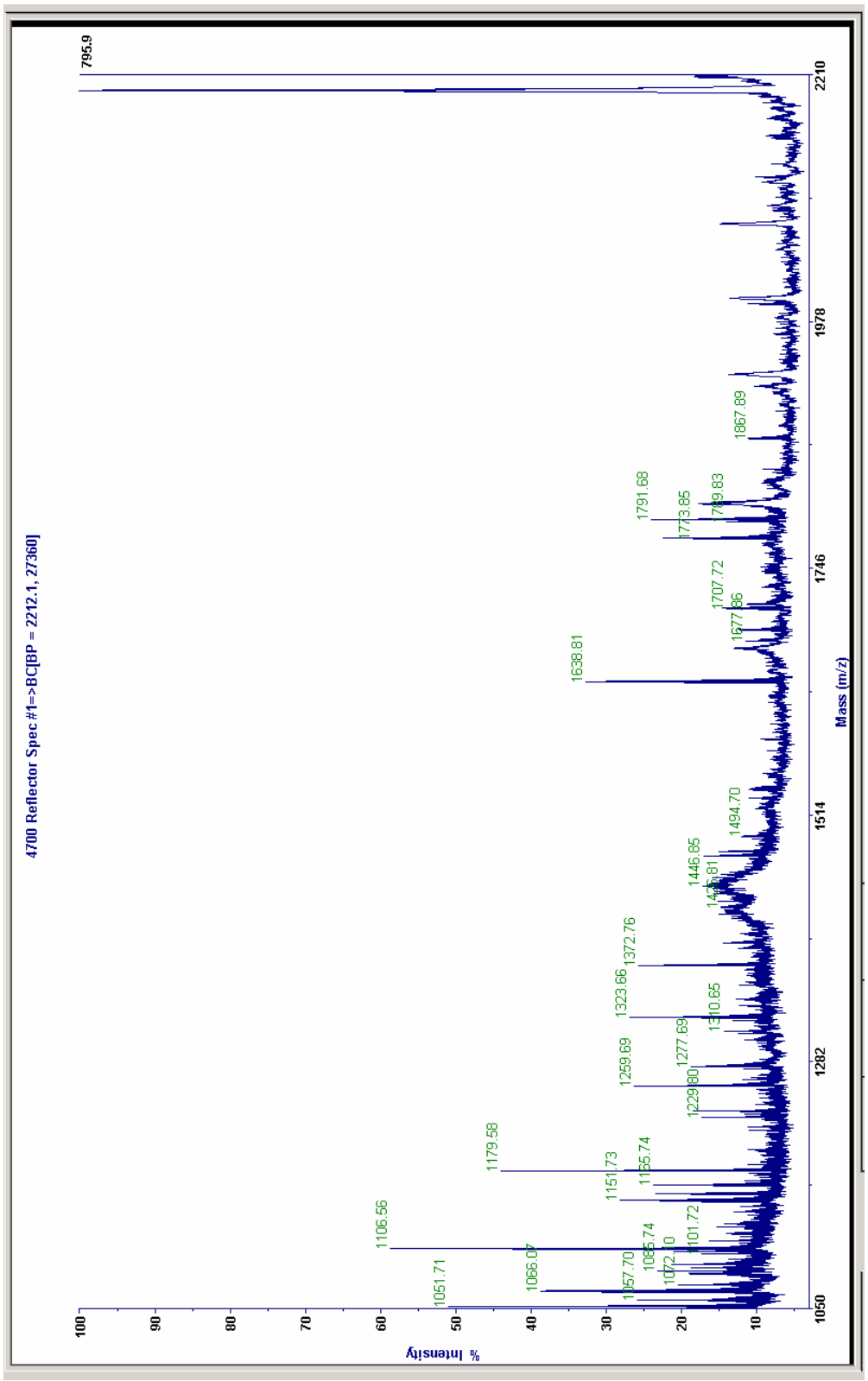

4700 Reflector Spec #1=>BC[BP = 2212.1, 27360]
% Intensity
Mass (m/z)
795.9
1638.81
1106.56
1179.58
1372.76
1446.85
1494.70
1791.68
1867.89
1707.72
1773.85
1789.83
1677.86
1323.66
1259.69
1277.69
1310.65
1151.73
1165.74
1051.71
1066.07
1101.72
1157.70
1229.80
1100
1050
1282
1514
1746
1978
2210
10
20
30
40
50
60
70
80
90

Fig. 3.13 MS spectrum of Band IV. MS experiments were performed by the Proteomics Resource Facility of the Molecular Genetics Instrumentation Facility of the University of Geargia. The instrument used is ABI 4700 TOF/TOF. The experiments performed were MOLDI-TOF mass spectrometry.

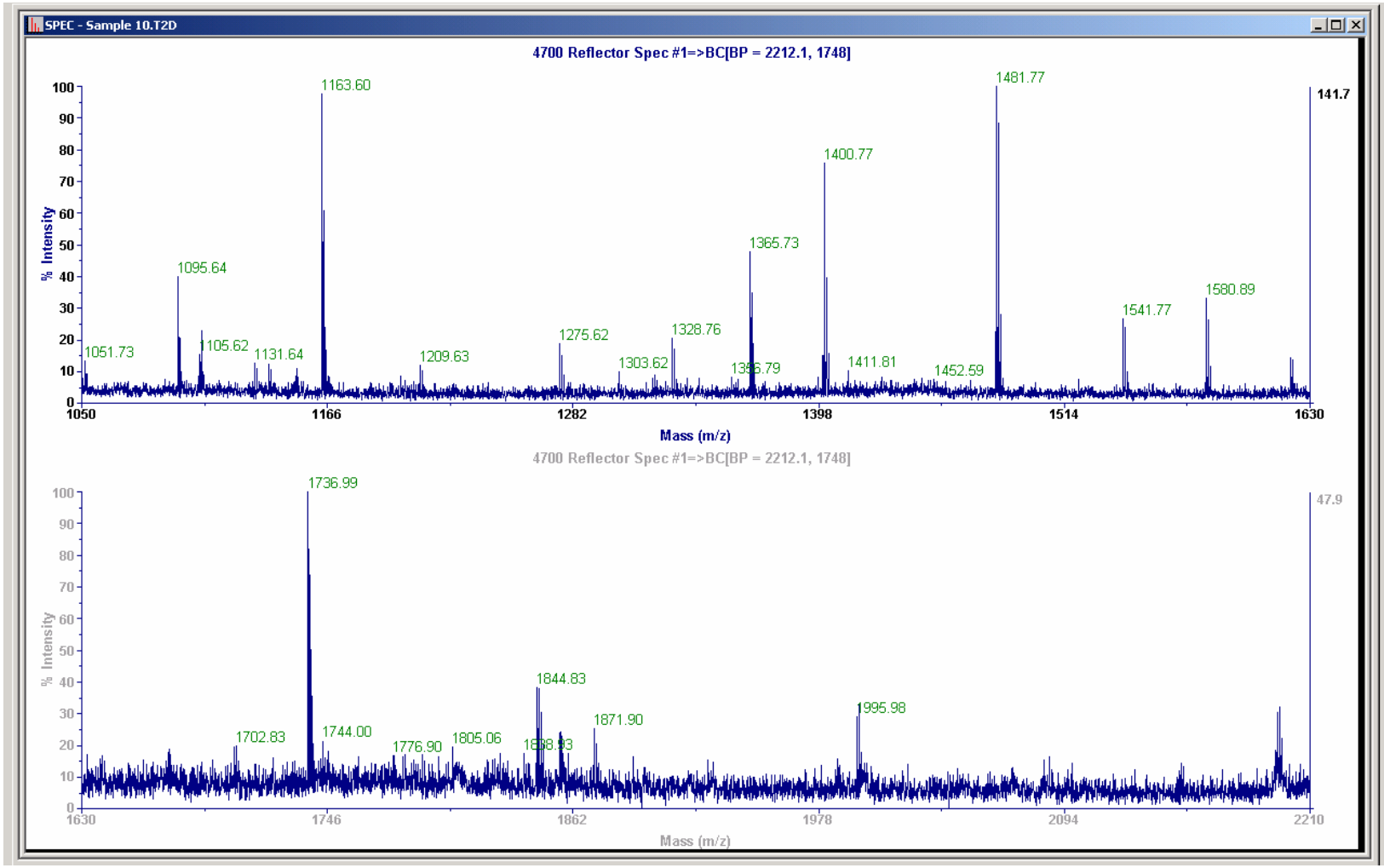

SPEC - Sample 10.T2D
4700 Reflector Spec #1=>BC[BP = 2212.1, 1748]
% Intensity
Mass (m/z)
1163.60
1481.77
1400.77
1365.73
1095.64
1580.89
1541.77
1051.73
1105.62
1131.64
1209.63
1275.62
1328.76
1303.62
1411.81
1452.59
4700 Reflector Spec #1=>BC[BP = 2212.1, 1748]
% Intensity
Mass (m/z)
1736.99
1844.83
1995.98
1871.90
1702.83
1744.00
1776.90
1805.06

Fig. 3.14 MS spectrum of Band V. MS experiments were performed by the Proteomics Resource Facility of the Molecular Genetics Instrumentation Facility of the University of Geargia. The instrument used is ABI 4700 TOF/TOF. The experiments performed were MOLDI-TOF mass spectrometry.

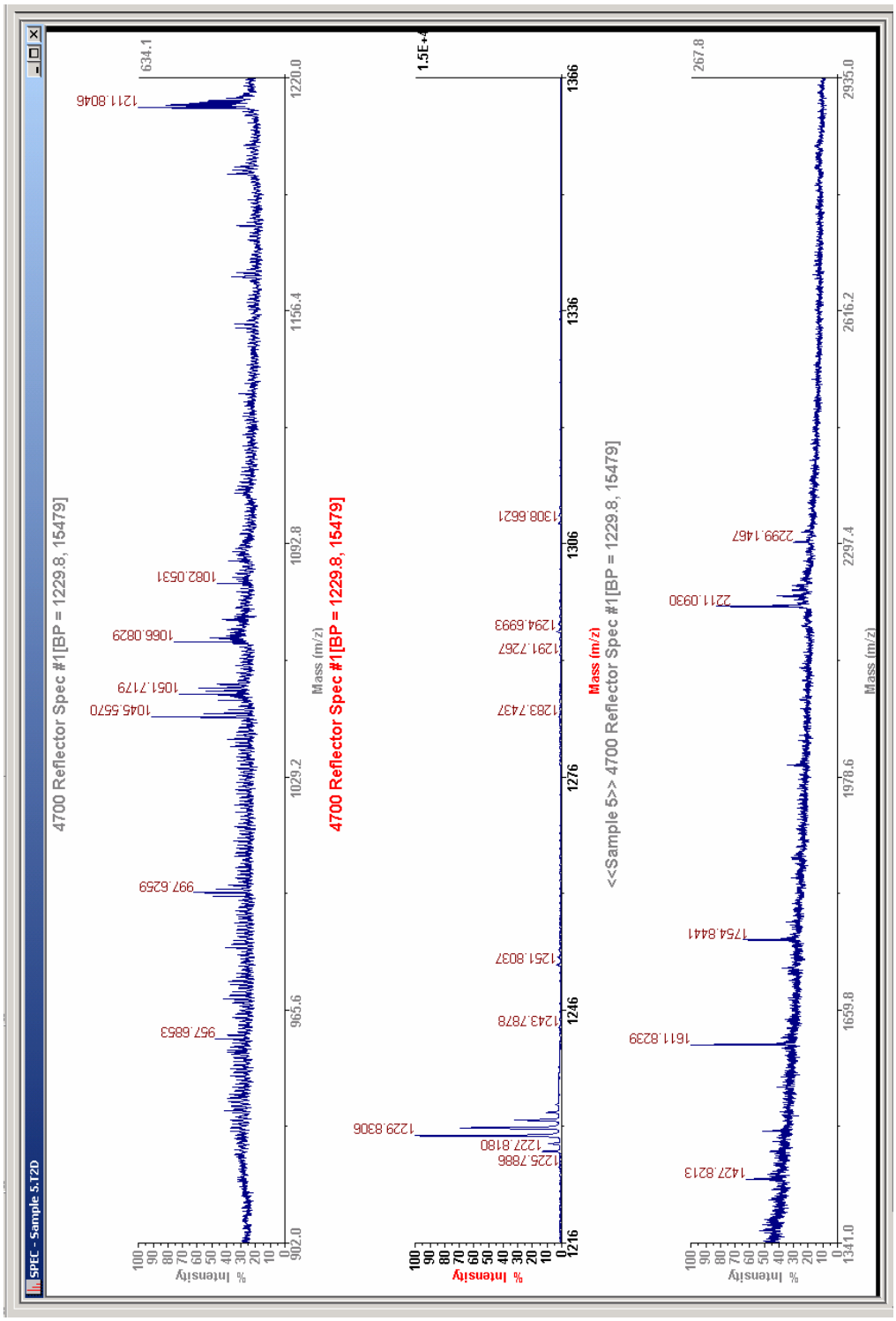

SPEC - Sample 5.12D
4700 Reflector Spec #1[BP = 1229.8, 15479]
1211.8046
634.1
1082.0531
1066.0829
1051.7179
1045.5570
997.6259
957.6853
902.0
965.6
1029.2
1092.8
1156.4
1220.0
Mass (m/z)
% Intensity
4700 Reflector Spec #1[BP = 1229.8, 15479]
1.5E+4
1308.6621
1294.6993
1291.7267
1283.7437
1251.8037
1243.7878
1229.8306
1227.8180
1225.7886
1216
1246
1276
1306
1336
1366
Mass (m/z)
% Intensity
<<Sample 5>> 4700 Reflector Spec #1[BP = 1229.8, 15479]
2299.1467
2211.0930
1754.8441
1611.8239
1427.8213
267.8
1341.0
1659.8
1978.6
2297.4
2616.2
2935.0
Mass (m/z)
% Intensity

Fig. 3.15 MS spectrum of Band VI. MS experiments were performed by the Proteomics Resource Facility of the Molecular Genetics Instrumentation Facility of the University of Geargia. The instrument used is ABI 4700 TOF/TOF. The experiments performed were MOLDI-TOF mass spectrometry.

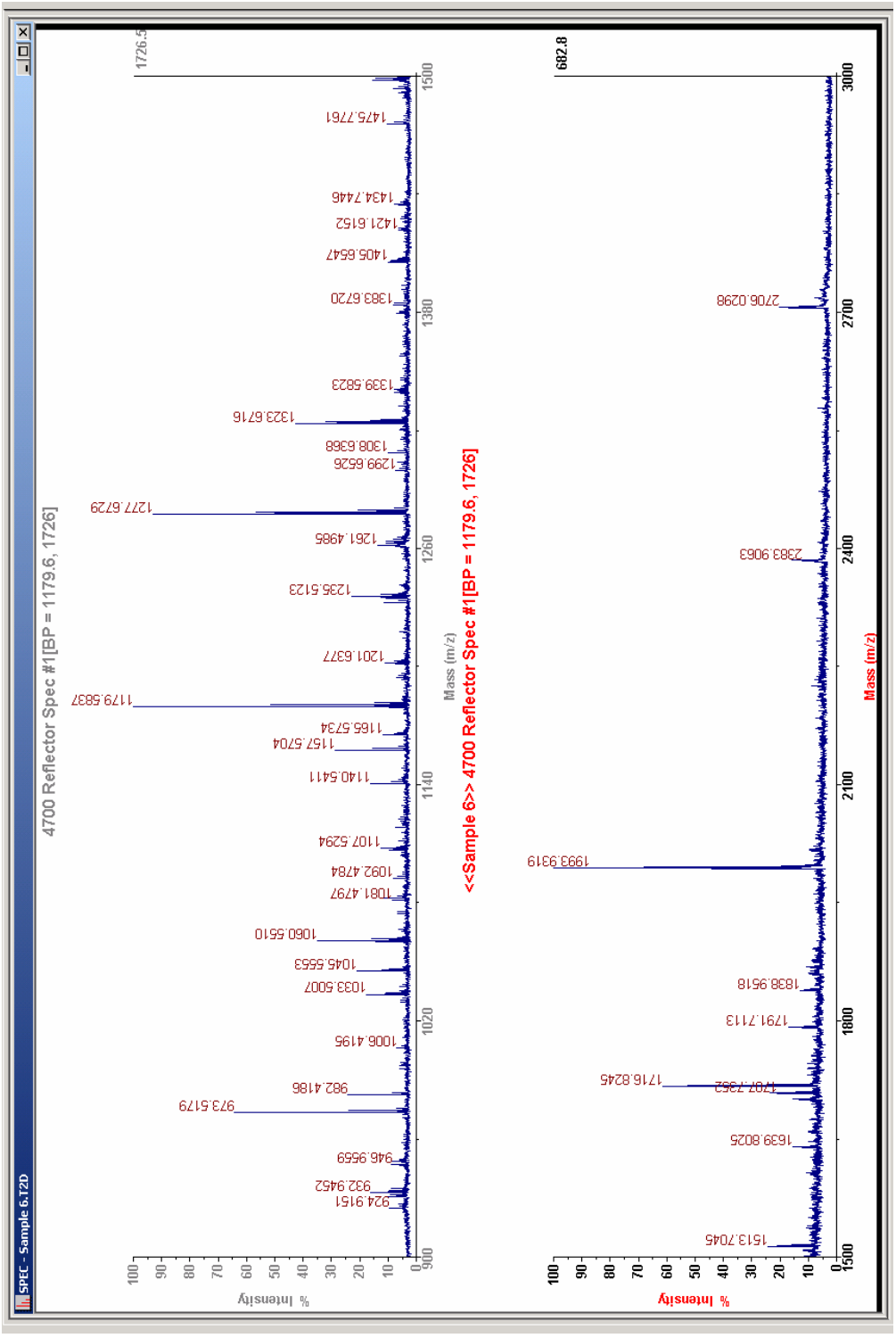

SPEC - Sample 6.12D
4700 Reflector Spec #1[BP = 1179.6, 1726]
<<Sample 6>> 4700 Reflector Spec #1[BP = 1179.6, 1726]
% Intensity
Mass (m/z)
1726.5
682.8
1475.7761
1434.7446
1421.6152
1405.6547
1383.6720
1339.5823
1323.6716
1308.6368
1299.6526
1277.6729
1261.4986
1235.5123
1201.6377
1179.5837
1165.5734
1157.5704
1140.5411
1107.5294
1092.4784
1081.4797
1060.5510
1045.5553
1033.5007
1006.4195
982.4186
973.5179
946.9559
932.9452
924.9151
2706.0298
2383.9063
1993.9319
1838.9518
1791.7113
1797.7352
1716.8245
1639.8025
1513.7045

Fig. 3.16 MS spectrum of Band VII. MS experiments were performed by the Proteomics Resource Facility of the Molecular Genetics Instrumentation Facility of the University of Geargia. The instrument used is ABI 4700 TOF/TOF. The experiments performed were MOLDI-TOF mass spectrometry.

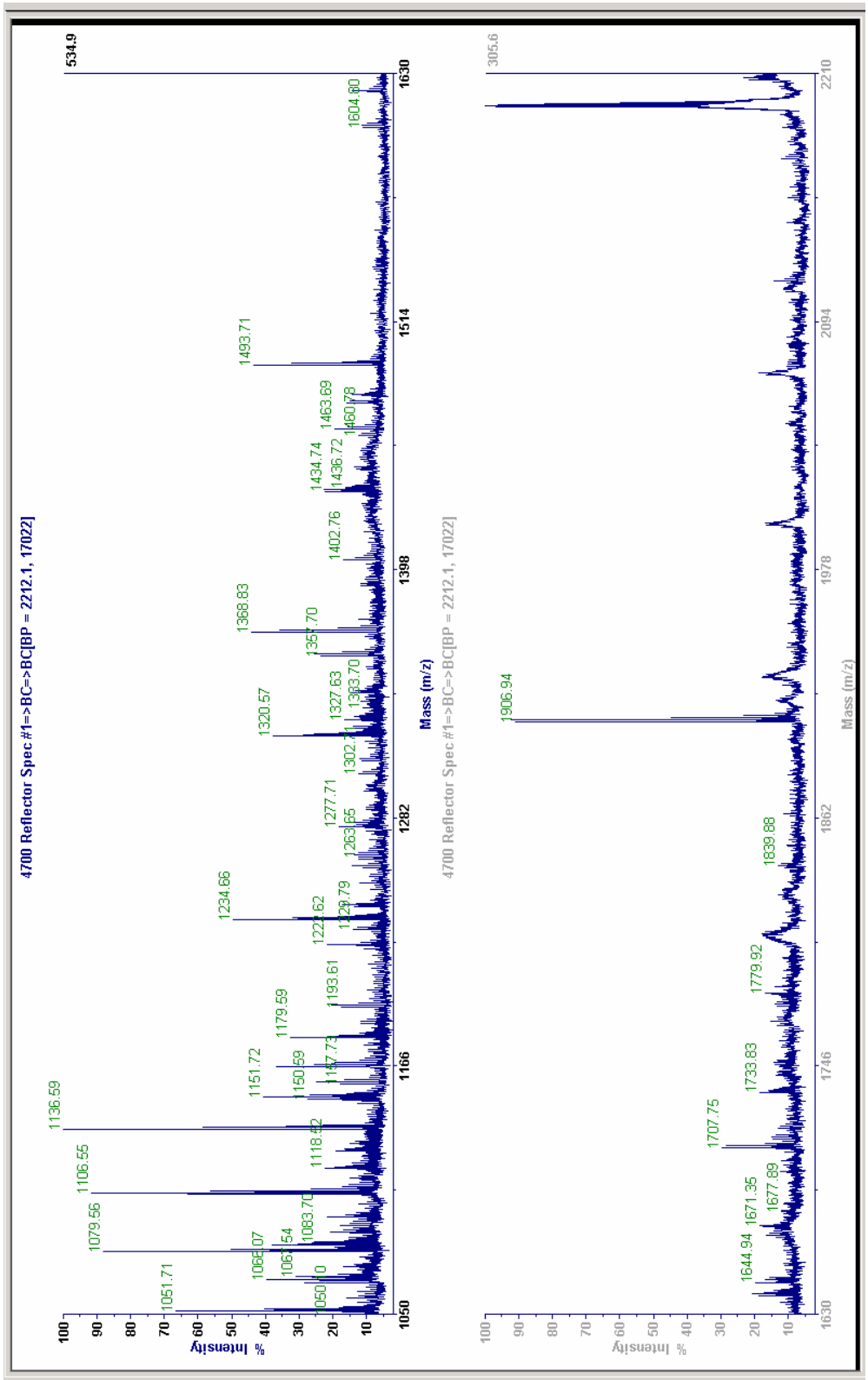

4700 Reflector Spec #1=>BC=>BC[BP = 2212.1, 17022]
% Intensity
Mass (m/z)
534.9
1604.80
1493.71
1463.69
1460.78
1434.74
1436.72
1402.76
1368.83
1357.70
1327.63
1383.70
1320.57
1302.71
1277.71
1263.65
1234.66
1221.62
1193.61
1179.59
1151.72
1150.59
1136.59
1118.52
1106.55
1083.70
1079.56
1066.07
1061.54
1051.71
1050.10
100
90
80
70
60
50
40
30
20
10
1050
1166
1282
1398
1514
1630
4700 Reflector Spec #1=>BC=>BC[BP = 2212.1, 17022]
% Intensity
Mass (m/z)
305.6
1906.94
1839.88
1779.92
1733.83
1707.75
1677.89
1671.35
1644.94
1630
1746
1862
1978
2094
2210

Fig. 3.17 MS spectrum of band VIII MS experiments were performed by the Proteomics Resource Facility of the Molecular Genetics Instrumentation Facility of the University of Geargia. The instrument used is ABI 4700 TOF/TOF. The experiments performed were MOLDI-TOF mass spectrometry.

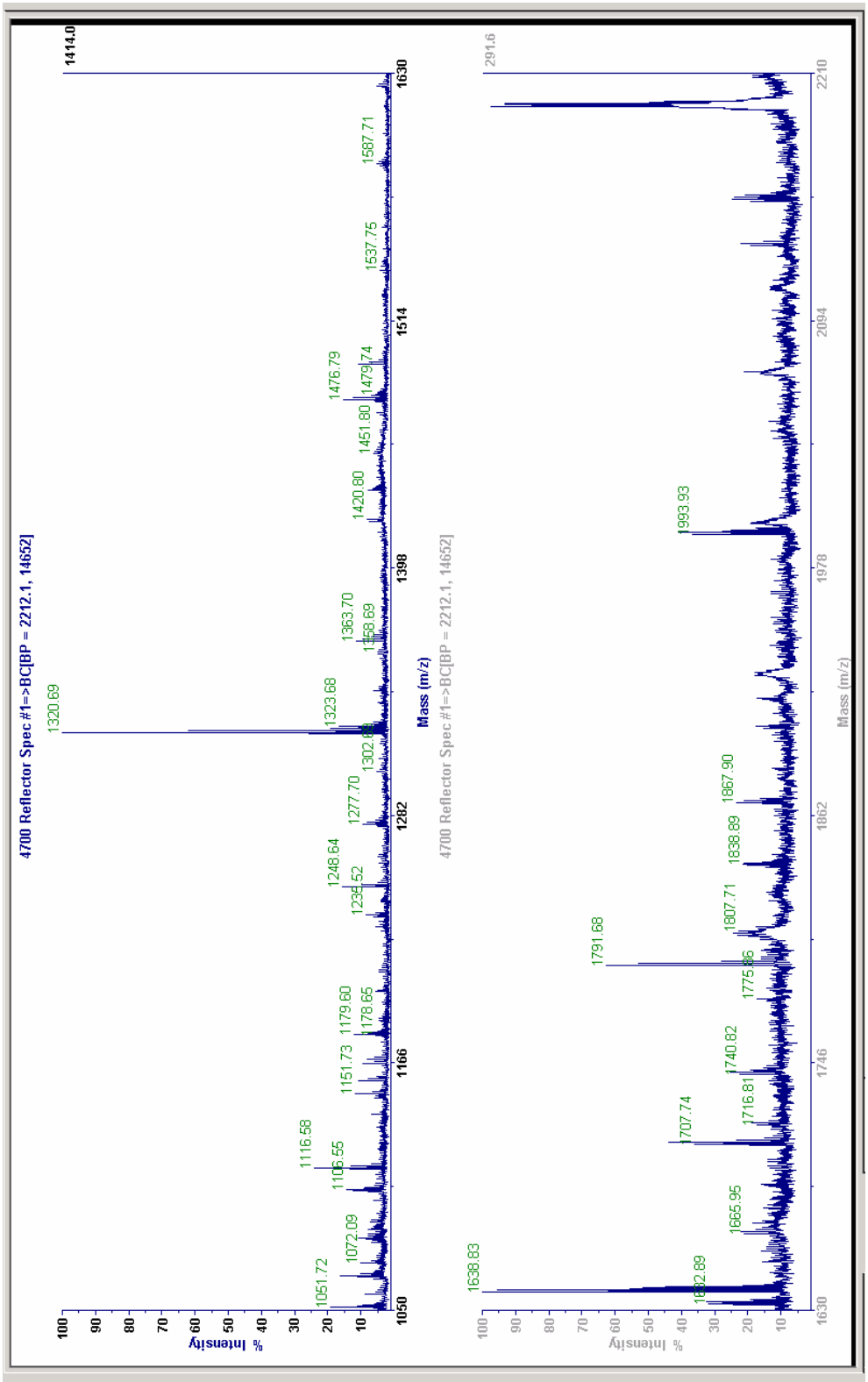

4700 Reflector Spec #1=>BC[BP = 2212.1, 14652]
% Intensity
Mass (m/z)
1320.69
1323.68
1363.70
1420.80
1451.80
1476.79
1537.75
1587.71
1248.64
1277.70
1179.60
1151.73
1116.58
1106.55
1072.09
1051.72
1707.74
1740.82
1716.81
1665.95
1632.89
1638.83
1791.68
1807.71
1838.89
1867.90
1993.93
291.6
1414.0
1630
1514
1398
1282
1166
1050
2210
2094
1978
1862
1746
1630

Table 3.2 Masses selected for identification of each sample.

I		II		III		IV		V	VI		VII		VIII	
1109.67	1490.77	1070.49	1746.87	1057.70	1536.68	1095.64	1623.87	957.685	924.915	1379.674	1067.54	1666.88	1065.69	1623.90
1172.66	1511.65	1079.55	1765.80	1066.07	1539.72	1131.64	1702.83	996.608	930.948	1405.655	1082.59	1670.91	1083.73	1632.89
1194.65	1520.80	1129.62	1782.88	1067.55	1670.85	1132.64	1736.99	997.626	932.945	1421.615	1136.59	1671.35	1116.58	1665.95
1215.68	1527.65	1148.60	1874.97	1085.74	1671.35	1138.56	1744.00	1211.805	948.940	1434.745	1158.65	1677.89	1158.62	1667.92
1259.67	1596.75	1150.08	2025.06	1129.80	1677.37	1151.11	1746.26	1212.376	982.419	1493.723	1222.62	1733.83	1167.73	1670.84
1301.73	1658.82	1204.56	2094.07	1259.69	1677.86	1163.60	1776.90	1227.818	1060.551	1497.737	1234.66	1734.86	1248.64	1707.74
1321.76	1674.80	1208.61	2108.02	1308.67	1687.94	1200.68	1790.91	1229.831	1081.480	1513.705	1279.75	1764.90	1259.72	1740.82
1324.73	1736.85	1285.72	2154.16	1310.65	1707.72	1209.63	1805.06	1243.788	1106.521	1639.803	1320.57	1766.80	1320.69	1775.86
1326.70	1743.83	1327.63	2170.09	1320.58	1711.85	1275.62	1838.93	1251.804	1107.529	1699.805	1327.63	1839.88	1340.66	1807.71
1328.70	1759.80	1332.64	2206.06	1334.88	1773.85	1319.72	1844.83	1283.744	1140.541	1707.735	1329.73	1906.94	1355.73	1838.89
1329.72	1770.87	1336.68		1340.66	1789.83	1328.76	1855.97	1294.699	1157.570	1791.711	1340.68		1358.69	
1364.70	1784.87	1340.67		1372.76	1805.86	1365.73	1871.90	1308.662	1165.573	1838.952	1368.83		1363.70	
1369.77	1806.85	1402.75		1393.70		1372.67	1995.98	1611.823	1201.638	2705.128	1402.76		1419.82	
1371.78	1883.91	1418.74		1426.81		1397.72		1754.842	1232.586	2706.030	1434.74		1434.75	
1396.74	1893.05	1439.78		1432.69		1400.77		2235.085	1235.512		1436.72		1451.80	
1423.72	1905.05	1463.67		1446.85		1411.81			1261.499		1444.90		1470.83	
1432.78	1964.09	1540.81		1448.61		1452.59			1263.647		1460.78		1476.79	
1439.77	1978.01	1601.87		1454.77		1481.77			1299.653		1463.69		1537.75	
1448.78	1995.03	1667.82		1458.82		1541.77			1308.637		1604.80		1539.76	
1454.76		1711.82		1494.70		1580.89			1323.672		1621.84		1557.84	
1463.73		1733.83		1529.85		1621.01			1339.582		1644.94		1587.71	

Table 3.3 Output of MS-Fit search results for I. Matched proteins were ranked according to their Probability Based Mowse Score. Mowse score=-10*Log (P), where P is the probability that the observed match is a random event. Proteins with scores greater than 72 are significant (P<0.05).

Database searched: **PN_pyrococcus**
Full Molecular Weight range: **2196** entries.
Full pI range: **2196** entries.
MS-Fit search selects **37** entries (results displayed for top **5** matches).

Considered modifications: | **Peptide N-terminal Gln to pyroGlu** | **Oxidation of M** | **Protein N-terminus Acetylated** | **Acrylamide Modified Cys** |

Min. # Peptides to Match	Peptide Mass Tolerance (+/-)	Peptide Masses are	Digest used	Max. # Missed Cleavages	Cysteines Modified by	Peptide N terminus	Peptide C terminus	Input # Peptide Masses
4	50.000 ppm	monoisotopic	Trypsin	1	unmodified	Hydrogen (H)	Free Acid (O H)	40

Result Summary

Rank	MOWSE Score	# (%) Masses Matched	Protein MW (Da)/pI	Species	Protein Name
1	1.01E+09	23/40 (57%)	94118.2 / 5.56	Pyrococcus furiosus	cell division control protein 48, aaa family
2	681	4/40 (10%)	41174.7 / 5.50	Pyrococcus furiosus	putative 3-isopropylmalate dehydratase largesubunit
3	416	5/40 (12%)	73801.1 / 8.51	Pyrococcus furiosus	dehydrogenase subunit alpha
4	316	5/40 (12%)	55806.0 / 5.83	Pyrococcus furiosus	conserved hypothetical protein
5	240	4/40 (10%)	36481.8 / 9.12	Pyrococcus furiosus	conserved hypothetical protein

Table 3.4 Output of MS-Fit search results for II

Database searched: **PN_pyrococcus**
Molecular weight search (**1000** - **100000** Da) selects **2163** entries.
Full pI range: **2196** entries.
Combined molecular weight and pI searches select **2163** entries.
MS-Fit search selects **14** entries (results displayed for top **5** matches).
Considered modifications: | **Peptide N-terminal Gln to pyroGlu** | **Oxidation of M** | **Protein N-terminus Acetylated** | **Acrylamide Modified Cys** |

Min. # Peptides to Match	Peptide Mass Tolerance (+/-)	Peptide Masses are	Digest used	Max. # Missed Cleavages	Cysteines Modified by	Peptide N terminus	Peptide C terminus	Input # Peptide Masses
4	50.000 ppm	monoisotopic	Trypsin	1	unmodified	Hydrogen (H)	Free Acid (O H)	31

Result Summary

Rank	MOWSE Score	# (%) Masses Matched	Protein MW (Da)/pI	Species	Protein Name
1	4.50E+06	19/31 (61%)	57112.6 / 6.21	Pyrococcus furiosus	methylmalonyl-CoAdecarboxylase, subunit alpha
2	347	4/31 (12%)	47540.1 / 9.44	Pyrococcus furiosus	conserved hypothetical protein
3	274	4/31 (12%)	37382.4 / 9.18	Pyrococcus furiosus	conserved hypothetical protein
4	260	4/31 (12%)	46224.7 / 6.03	Pyrococcus furiosus	phosphoglycerate kinase
5	112	4/31 (12%)	48316.2 / 5.67	Pyrococcus furiosus	sulfhydrogenase alpha subunit

Table 3.5 Output of MS-Fit search results for III

Database searched: **PN_pyrococcus**
Molecular weight search (**1000** - **100000 Da**) selects **2163** entries.
Full pI range: **2196** entries.
Combined molecular weight and pI searches select **2163** entries.
MS-Fit search selects **5** entries.
Considered modifications: | **Peptide N-terminal Gln to pyroGlu** | **Oxidation of M** | **Protein N-terminus Acetylated** | **Acrylamide Modified Cys** |

Min. # Peptides to Match	Peptide Mass Tolerance (+/-)	Peptide Masses are	Digest used	Max. # Missed Cleavages	Cysteines Modified by	Peptide N terminus	Peptide C terminus	Input # Peptide Masses
4	50.000 ppm	monoisotopic	Trypsin	1	unmodified	Hydrogen (H)	Free Acid (O H)	33

Result Summary

Rank	MOWSE Score	# (%) Masses Matched	Protein MW (Da)/pI	Species	Protein Name
1	1.98E+03	7/33 (21%)	39722.7 / 9.10	Pyrococcus furiosus	glutamate synthase small subunit
2	139	4/33 (12%)	30773.7 / 8.33	Pyrococcus furiosus	probable tRNA pseudouridine synthase a
3	21.9	4/33 (12%)	82255.9 / 6.27	Pyrococcus furiosus	elongation factor
4	7.94	4/33 (12%)	40848.4 / 6.21	Pyrococcus furiosus	pleiotropic regulatory protein degT
5	5.98	4/33 (12%)	49270.5 / 6.00	Pyrococcus furiosus	tbp-interacting protein tip49

Table 3.6 Output of MS-Fit search results for IV

Database searched: **PN_pyrococcus**
Full Molecular Weight range: **2196** entries.
Full pI range: **2196** entries.
MS-Fit search selects **15** entries (results displayed for top 5 matches).

Considered modifications: | **Peptide N-terminal Gln to pyroGlu** | **Oxidation of M** | **Protein N-terminus Acetylated** | **Acrylamide Modified Cys** |

Min. # Peptides to Match	Peptide Mass Tolerance (+/-)	Peptide Masses are	Digest used	Max. # Missed Cleavages	Cysteines Modified by	Peptide N terminus	Peptide C terminus	Input # Peptide Masses
4	50.000 ppm	monoisotopic	Trypsin	1	unmodified	Hydrogen (H)	Free Acid (O H)	34

Result Summary

Rank	MOWSE Score	# (%) Masses Matched	Protein MW (Da)/pI	Species	Protein Name
1	1.85E+12	21/34 (61%)	139985.0 / 7.51	Pyrococcus furiosus	reverse gyrase (rgy)
2	533	4/34 (11%)	31803.7 / 8.48	Pyrococcus furiosus	conserved hypothetical protein
3	184	4/34 (11%)	44680.1 / 8.96	Pyrococcus furiosus	conserved hypothetical protein
4	105	4/34 (11%)	90390.6 / 8.94	Pyrococcus furiosus	conserved hypothetical protein
5	85.9	4/34 (11%)	69679.4 / 6.02	Pyrococcus furiosus	formaldehyde:ferredoxin oxidoreductase

Table 3.7 Output of MS-Fit search results for V

Database searched: **PN_pyrococcus**
Molecular weight search (**1000 - 100000 Da**) selects **2163** entries.
Full pI range: **2196** entries.
Combined molecular weight and pI searches select **2163** entries.
MS-Fit search selects **1** entries.
Considered modifications: | **Peptide N-terminal Gln to pyroGlu** | **Oxidation of M** | **Protein N-terminus Acetylated** | **Acrylamide Modified Cys** |

Min. # Peptides to Match	Peptide Mass Tolerance (+/-)	Peptide Masses are	Digest used	Max. # Missed Cleavages	Cysteines Modified by	Peptide N terminus	Peptide C terminus	Input # Peptide Masses
4	50.000 ppm	monoisotopic	Trypsin	1	unmodified	Hydrogen (H)	Free Acid (O H)	33

Result Summary

Rank	MOWSE Score	# (%) Masses Matched	Protein MW (Da)/pI	Species	Protein Name
1	1.36E+03	4/15 (26%)	41045.5 / 4.94	Pyrococcus furiosus	replication factor A related protein

Table 3.8 Output of MS-Fit search results for VI

Database searched: **PN_pyrococcus**

Molecular weight search (**1000** - **100000** Da) selects **2163** entries.

Full pI range: **2196** entries.

Combined molecular weight and pI searches select **2163** entries.

MS-Fit search selects **31** entries (results displayed for top **5** matches).

Considered modifications: | **Peptide N-terminal Gln to pyroGlu** | **Oxidation of M** | **Protein N-terminus Acetylated** | **Acrylamide Modified Cys** |

Min. # Peptides to Match	Peptide Mass Tolerance (+/-)	Peptide Masses are	Digest used	Max. # Missed Cleavages	Cysteines Modified by	Peptide N terminus	Peptide C terminus	Input # Peptide Masses
4	100.000 ppm	monoisotopic	Trypsin	1	unmodified	Hydrogen (H)	Free Acid (O H)	35

Result Summary

Rank	MOWSE Score	# (%) Masses Matched	Protein MW (Da)/pI	Species	Protein Name
1	1.48E+03	4/35 (11%)	23355.0 / 5.75	Pyrococcus furiosus	conserved hypothetical protein
2	1.31E+03	7/35 (20%)	44155.3 / 7.13	Pyrococcus furiosus	GTP cyclohydrolase II
3	766	4/35 (11%)	50415.4 / 5.94	Pyrococcus furiosus	conserved hypothetical protein
4	456	4/35 (11%)	12463.7 / 9.00	Pyrococcus furiosus	conserved hypothetical protein
5	406	4/35 (11%)	35962.6 / 4.97	Pyrococcus furiosus	agmatinase

Table 3.9 Output of MS-Fit search results for VII

Database searched: **PN_pyrococcus**
Molecular weight search (**1000** - **100000 Da**) selects **2163** entries.
Full pI range: **2196** entries.
Combined molecular weight and pI searches select **2163** entries.
MS-Fit search selects **16** entries (results displayed for top **5** matches).
Considered modifications: | **Peptide N-terminal Gln to pyroGlu** | **Oxidation of M** | **Protein N-terminus Acetylated** |

Min. # Peptides to Match	Peptide Mass Tolerance (+/-)	Peptide Masses are	Digest used	Max. # Missed Cleavages	Cysteines Modified by	Peptide N terminus	Peptide C terminus	Input # Peptide Masses
4	50.000 ppm	monoisotopic	Trypsin	1	unmodified	Hydrogen (H)	Free Acid (O H)	31

Result Summary

Rank	MOWSE Score	# (%) Masses Matched	Protein MW (Da)/pI	Species	Protein Name
1	354	4/31 (12%)	28005.4 / 4.56	Pyrococcus furiosus	pcna sliding clamp (proliferating-cell nuclearantigen)
2	318	6/31 (19%)	57112.6 / 6.21	Pyrococcus furiosus	methylmalonyl-CoAdecarboxylase, subunit alpha
3	154	4/31 (12%)	102510.4 / 8.40	Pyrococcus furiosus	conserved hypothetical protein
4	33.9	4/31 (12%)	89208.6 / 5.54	Pyrococcus furiosus	cell division control protein 48, aaa family
5	32.2	6/31 (19%)	104933.5 / 5.72	Pyrococcus furiosus	valyl-tRNA synthetase

Table 3.10 Output of MS-fit search results for VIII

Database searched: **PN_pyrococcus**
Molecular weight search (**1000 - 100000 Da**) selects **2163** entries.
Full pI range: **2196** entries.
Combined molecular weight and pI searches select **2163** entries.
MS-Fit search selects **12** entries (results displayed for top **5** matches).
Considered modifications: **| Peptide N-terminal Gln to pyroGlu | Oxidation of M | Protein N-terminus Acetylated | Acrylamide Modified Cys |**

Min. # Peptides to Match	Peptide Mass Tolerance (+/-)	Peptide Masses are	Digest used	Max. # Missed Cleavages	Cysteines Modified by	Peptide N terminus	Peptide C terminus	Input # Peptide Masses
4	50.000 ppm	monoisotopic	Trypsin	1	unmodified	Hydrogen (H)	Free Acid (O H)	31

Result Summary

Rank	MOWSE Score	# (%) Masses Matched	Protein MW (Da)/pI	Species	Protein Name
1	892	7/31 (22%)	34105.8 / 9.52	Pyrococcus furiosus	transcription initiation factor IIB chain b
2	369	4/31 (12%)	47347.8 / 5.81	Pyrococcus furiosus	ribulose-1,5-bisphosphate carboxylase
3	110	4/31 (12%)	27243.2 / 9.41	Pyrococcus furiosus	hypothetical protein
4	44.8	5/31 (16%)	97706.0 / 5.83	Pyrococcus furiosus	alpha-glucan phosphorylase
5	44.2	4/31 (12%)	38214.3 / 5.26	Pyrococcus furiosus	protein similar to endo-1,4-beta-glucanase

Table 3.11 Identities of I-V. VII and VIII. For I, II, IV, and V, search was performed using the NCBInr.11.08.02 database. For III, VII and VIII, search was performed using a database for *Pyrococcus furiosus*.

	Name	MW/Da	MOWSE score	Description
I	cell division control protein 48, aaa family, (cdc48-2)	94118.2	1.01E+09	
II	Methylmalonyl-CoAdecarboxylase, subunit α	57112.6	4.50E+06	
III	Glutamate synthase small subunit	39722.7	1.98E+03	
IV	Reverse gyrase	139985	1.85E+12	an enzyme unique to hyperthermophiles, it catalyzes the ATP-dependant introduction of positive supercoils into DNA
V	Replication factor A related protein	41045.5	1.36E+03	replication factor A related protein: single stranded DNA-binding protein, its counterparts in bacteria and eukarya play important roles during DNA replication, repair and recombination. Studies show that Pfu replication factor A is involved in homologous DNA recombination
VI	GTP cyclohydrolase II	44155.3	1.31E+03	in *E. coli* this enzyme catalyzes the formation of a pyrimic ı derivative that serves as the first commtted intermediate in the biosynthetic pathway of riboflavin.
VII	PCNA sliding clamp (proliferating cell nuclear antigen)	28005.4	354	PCNA sliding clamp (proliferating cell nuclear antigen): interacts directly with DNA polymerase and as well as a wide variety of proteins related to important cellular processes
VIII	Transcription initiation factor B (TFB)	34105.8	892	transcription factor B

The database used for CD search is CDD v1.62-11088 PSSMs, which contains domains from two popular collections: Smart (Simple modular architecture research tool); and Pfam (Protein family database of alignments and HMMs) as well as several other databases such as COG (Clusters of Orthologous Groups of proteins) [Marchler-Bauer *et al*., 2002, 2003]. CDs found in the submitted proteins are ranked according to their bit score (S'), which is calculated from the raw score (S) by normalizing with the statistical variables that define a given scoring system. The E (expect) value indicates the expected number of chance alignments with a score of S or better. In CD search, results with E-values ≥ 1 should be considered putative false positives [Altschul *et al*., 1997]. Below, each band that was identified (Fig. 3.9) is discussed in terms of its conserved domains. Band I through III were retrieved from experiments using control DNA. Band IV through VIII were retrieved from experiments using promoter DNA.

Band I. cell division control protein 48, aaa family; (cdc48-2). Most conserved domain found in this protein belongs to the AAA family ATPase. Proteins in this family have similar structure and highly diversified functions [Patel *et al*., 1998; Ogura *et al*., 2001]. Several CDs recognized are related to DNA replication, recombination, and repair and may contribute to the protein's affinity for DNA (Table 3.12). Although poorly aligned, a CD found in both archaeal and eucaryal TIP49, TBP-interacting protein, was also identified in a couple of positions in this protein [Kanemaki *et al*., 1997]

Band II. Methylmalonyl-CoAdecarboxylase, subunit α. Two DNA binding CDs, pfam04931, DNA polymerase V and pfam03871, RNA polymerase Rpb5, N-terminal domain,

Table 3.12 CDs with possible DNA affinity found in I, cell division control protein 48, aaa family

CD name	% aligned	Score (bits)	E value	sequence alignment
COG2256, MGS1, ATPase related to the helicase subunit of the Holliday junction resolvase[DNA replication, recombination, and repair] . 436 residues	11.2	48.7	3.00E-06	I 246 VLLYGPPGTGKTLLAKAVANEANAYFIAINGPEIMSKYYGESEERLREIFKEAEEN 301 COG2256 51 MILWGPPGTGKTTLARLIAGTTNAAFEALSAVT-------SGVKDLREIIEEARKN 99
	10.8	47.9	4.00E-06	I 582 LLYGPPGTGKTLLAKAVATESQANFIAIRGPEVLSKWVGESEKRIREIFRKARQ 635 COG2256 52 ILWGPPGTGKTTLARLIAGTTNAAFEALSA-------VTSGVKDLREIIEEARK 98
cd00009, AAA, AAA-superfamily of ATPases associated with a wide variety of cellular activities, including membrane fusion, proteolysis, and DNA replication. 129 residues	99.2	45.3	3.00E-05	I 581 ILLYGPPGTGKTLLAKAVATESQ-----ANFIAIRGPEVLSKWV--GESEKRIREIFRKA 633 cd00009 2 VLIVGPPGSGKTTLARAIARELLPTGLGKRVIYVNQESLLFNGGSSLSGGQKQRLLLARA 61 I 634 RQAAPAIIFIDEIDAIAPARGGYEGERVTDRLINQLLTEMDGIQENSGVVVIGATNRPDI 693 cd00009 62 LEAAGEEGKPDVLILDEITSLLDSETR------EELLEALLELLEEEGVTVILITHDLSL 115 I 694 IDPALLRPGRFDRLILVP 711 cd00009 116 LEL----RDRLDRRLVLD 129
	97.7	45	3.00E-05	I 246 VLLYGPPGTGKTLLAKAVANEA-----NAYFIAINGPEIMSK--YYGESEERLREIFKEA 298 cd0009 2 VLIVGPPGSGKTTLARAIARELLPTGLGKRVIYVNQESLLFNGGSSLSGGQKQRLLLARA 61 I 299 EENAPAIIFIDEIDAIAPKREEVVGEVEKRVVSQLLTLMDGLKSRGKVIVIAATNRPDAI 358 cd0009 62 LEAAGEEGKPDVLILDEITSLLDSETR-----EELLEALLELLEEEGVTVILITHDLSLL 116 I 359 DPALRRPGRFD 369 cd0009 117 ELRDRLDRRLV 127
COG0593, DnaA, ATPase involved in DNA replication initiation [DNA replication, recombination, and repair]. 408 residues	36.8	45.3	3.00E-05	I 246 VLLYGPPGTGKTLLAKAVANEANAYFIAINGPEIMSKYYGESEERLREIFKEAEENAPAI 305 COG0593 116 LFIYGGVGLGKTHLLQAIGNEALA-----NGPNARVVYL-TSEDFTNDFVKALRDNEMEK 169 I 306 IFIDEIDAIAPKREEVVGEVEKRVVSQLLTLMDGLKSRGKVIVIAATNRPDAI---DPAL 362 COG0593 170 FKEKYSLDLLLIDDIQFLAGKERTQEEFFHTFNALLENGKQIVLTSDRPPKELNGLEDRL 229 I 363 RRPGRFDR--EIEVGVPDKQGRKEILQIHTRGMPIEPD 398 COG0593 230 RS--RLEWGLVVEIEPPDDETRLALLRKKAEDRGIEIP 265
COG2255, RuvB, Holliday junction resolvasome, helicase subunit [DNA replication, recombination, and repair] . 332 residues	18.7	45.1	3.00E-05	I 543 WDDIGGLEEVKQQLREAVEWPLKYPKAFKRLGISPPKGILLYGPPGTGKTLLAKAVATES 602 COG2255 25 LDEFIGQEKVKEQLQIFI-------KAAKKRG-EALDHVLLFGPPGLGKTTLAHIIANEL 76 I 603 QANFIAIRGP 612 COG2255 77 GVNLKITSGP 86
	11.7	44.4	3.00E-05	I 241 EPPKGVLLYGPPGTGKTLLAKAVANEANAYFIAINGPEI 279 COG2255 50 EALDHVLLFGPPGLGKTTLAHIIANELGVNLKITSGPAL 88
COG1224, TIP49, DNA helicase TIP49, TBP-interacting protein [Transcription]. 450 residues	16.4	44.1	6.00E-05	I 557 REAVEWPLKYPKAFKRLGISPPKGILLYGPPGTGKTLLAKAVATE--SQANFIAIRGPEV 614 COG1224 48 REAAGVIVKMIKQGKMAG----RGILIVGPPGTGKTALAMGIARELGEDVPFVAISGSEI 103 I 615 LSKWVGESEKRIREIFRKA 633 COG1224 104 YSLEVKKTEA-LTQALRRA 121
	10.4	42.2	3.00E-04	I 244 KGVLLYGPPGTGKTLLAKAVANE--ANAYFIAINGPEIMSKYYGESE 288 COG1224 66 RGILIVGPPGTGKTALAMGIARELGEDVPFVAISGSEIYSLEVKKTE 112
COG1484, DnaC, DNA replication protein [DNA replication, recombination, and repair]. 254 residues.	26.8	44.2	7.00E-05	I 240 IEPPKGVLLYGPPGTGKTLLAKAVANEANAY---FIAINGPEIMSK-----YYGESEERL 291 COG1484 102 FERGENLVLLGPPGVGKTHLAIAIGNELLKAGISVLFITAPDLLSKLKAAFDEGRLEEKL 161 I 292 REIFKEAE 299 COG1484 162 LRELKKVD 169
	20.5	37.3	8.00E-03	I 578 PKGILLYGPPGTGKTLLAKAVATESQANFIAIRGPEVLSKWVGESEKRIREIFRKAR 634 COG1484 105 GENLVLLGPPGVGKTHLAIA-----IGNELLKAGISVLFITAPDLLSKLKAAFDEGR 156

were recognized [Shimizu *et al*., 2002; Yee *et al*., 2000]. Both have high E values and relatively low percentage of alignment (Table 3.13).

These two proteins were also seen in the experiments using promoter DNA fragment, their possible involvement in transcription can not be immediately eliminated despite its inability of sequence discrimination.

Band III. Glutamate synthase small subunit. No DNA-binding CDs were recognized in this protein. This retrieved protein might be a false match or contain DNA binding structure that has not been identified so far, or bind to another protein non-specifically bound to DNA.

Band IV. Reverse gyrase. This protein is an enzyme unique to all hyperthermophiles [Borges *et al*., 1997]. It introduces positive supercoil into dsDNA, making it more stable under high temperatures. *Pf* reverse gyrase has been observed to inhibit transcription *in vitro*. This may be due to the fact that transcription favors negative supercoiled DNA, which is easier to unwind and more readily accessible for transcription factors and RNAP [Borges *et al*., 1997]. The sequence of this protein matches the 1187 residue reverse gyrase CD with a bit score of 1490 and E value of 0. This CD is involved in DNA replication, recombination, and repair. No CDs with possible function in transcription were found.

Band V. Replication factor A related protein. Replication factor A (RPA) is a single-stranded DNA-binding protein; it is among the key proteins in DNA replication and recombination. The *Pf* RPA was reported to interact with a number of recombination and replication proteins and may be involved in homologous DNA recombination [Komori *et al*., 2001]. A CD characteristic to single-stranded DNA-binding RPA (RFA1) has been recognized. Another CD common to single-stranded DNA-binding protein in bacteria and

Table 3.13 CDs with DNA affinity found in II, Methylmalonyl-CoAdecarboxylase, subunit α

CD name	% aligned	Score (bits)	E value	sequence alignment
pfam04931, DNA_pol_V, DNA polymerase V. This family includes the fifth essential DNA polymerase in yeast EC:2.7.7.7. Pol5p is localised exclusively to the nucleolus and binds near or at the enhancer region of rRNA-encoding DNA repeating units. 781 residues	4.1	26.9	5.9	II 1 MSMEEKVKDLYERKKKIMQMGGEEAIKKQHDK 32 pfam04931 732 MALDERLAEIFKERKERIQAGKEEKKKAQSEK 763
pfam03871, RNA_pol_Rpb5_N, RNA polymerase Rpb5, N-terminal domain. Rpb5 has a bipartite structure which includes a eukaryote-specific N-terminal domain and a C-terminal domain resembling the archaeal RNAP subunit H. The N-terminal domain is involved in DNA binding and is part of the jaw module in the RNA pol II structure. This module is important for positioning the downstream DNA. 93 residues	40.9	26	10	II 1 MSMEEKVKDLYERKKKIMQMGGEEAIKKQHDKGKLTARERIELLLD 46 pfam03871 1 MDDEEEIKRLFLARRTVMQM--------LRDRGYLVSQEELNLTLE 38

eucaryal RPA, the oligonucleotide/oligosaccharide binding (OB) fold nucleic acid binding domain was also found in this protein (Table 3.14) [Murzin, 1993].

Band VII. Proliferating cell nuclear antigen (PCNA sliding clamp). The *Pf* PCNA sliding clamp has a structure similar to the eucaryal PCNA sliding clamp [Mazumiya *et al*., 2001]. Three protein molecules form a ring-shaped trimer which binds to the DNA. Eucaryal PCNA sliding clamp works as the elongation factor during DNA replication. In bacteria, the PCNA sliding clamp is incorporated into DNA polymerase. *Pf* PCNA sliding clamp interacts with DNA polymerase directly and was observed to stimulate *in vitro* DNA polymerization. This protein has also been found to interact with various proteins that participate in several important cellular activities [Cann *et al*., 1999; Mazumiya *et al*., 2001]. In addition to CDs of PCNA sliding clamp and DNA polymerase, one DNA binding domain that recognized a specific DNA sequence was also found in this protein (Table 3.15). This domain is previous found in a variety of bacterial transcription regulators (pcnaref3) [Nikolskaya *et al*., 2002]. However, its low percentage of aligned amino acid residues and high E value may suggest a false hit.

In the MS-fit result of VI (Table 3.8), the first hit is a conserved hypothetical protein with a molecular weight of 23355.0 Da. The second hit, which also has a very high Mowse score, is the GTP cyclohydrolase II with a molecular weight of 44155.3 Da. The VI band on the gel is at approximately 37 kDa. Among the top five hits of the results, agmatinase has the closest molecular weight of 35962.6 kDa. However, proteins may run abnormally on gel and their position on the gel may not be consistent with their molecular weight in all cases.

Table 3.14 DNA binding CDs found in V, Replication factor A related protein

CD name	% aligned	Score (bits)	E value	sequence alignment
COG1599, RFA1, Single-stranded DNA-binding replication protein A (RPA), large (70 kD) subunit and related ssDNA-binding proteins [DNA replication, recombination, and repair]	68.8	133	3.00E-32	V 74 DPHEVNIVGRILKKYPPREYTKKDGSIGRVASLVIYDD-TGRARVVLWDSKVLEYYS--- 129 COG1599 111 EKLEPGDVIRIRNAYTSLYRGGKRLSVGRVGSVADVDDEEDEARESEDAREIGEESLLSP 170 V 130 ----KLEVGDVIKVLDAQVRESLSGLPELHINFRARIIKNPDDP-RVQDIPPLEEVRVAT 184 COG1599 171 YQKARVVVGSEIKTFDNQGGESKVFSNELEDEERGVIVFTDWDPSQDGDVYRIEGARVKT 230 V 185 YTRKK-ISEVEPGDRFVELRGTIAKVYRVLVYDACPECKKKVDYDPG----MDVWICPEH 239 COG1599 231 KNKQPEENLAEELVLRVEVRVAIEKAEREEFVDEVKESVSLVEADGAVVDVTRVPECERV 290 V 240 GEVEPIKITILDFGLDDGSGYIRITLFGDDAEELLGVGPEEIAQKLKEMESMGMT-LKEA 298 COG1599 291 VRKGGCKGHGKDIGLDDLTGKIRVTLWGDATEVLINEESVEALKGINVEDASGIALSALD 350 V 299 ARKLAEEEFYNIIGKEIIVRGNVIEDRFLGLILRASSWEE 338 COG1599 351 TEAVALEIWYDILGKYLRVTGDAREDRYLIENLVESSTWD 390
	34.2	44.7	1.00E-05	V 63 LMKISDLYPGMDPHEVNIVGRILKKYPPREYTKKDGSIGRVASLVIYDDTGRARVV-LWD 121 COG1599 49 IGKISDISEASS--RVNVTGRVLSIGEKKTFDRKRGAEGKLAEVLVGDETGSVKTVTLWN 106 V 122 SKVLEYYSKLEVGDVIKVLDAQVRESLSGLPELHINFRARIIKNPDD----PRVQDIPPL 177 COG1599 107 IAALE---KLEPGDVIRIRNAYTSLY-RGGKRLSVGRVGSVADVDDEEDEARESEDAREI 162 V 178 EEVRVATYTRKKISEVEPGDRFVELRG 204 COG1599 163 GEES--LLSPYQKARVVVGSEIKTFDN 187
pfam01336, tRNA_anti, OB-fold nucleic acid binding domain. This family contains OB-fold domains that bind to nucleic acids.	93.4	41.7	1.00E-04	V 78 VNIVGRILKKYPPREYTKKDGSIGRVASLVIYDDTGRARVVLW-DSKVLEYYSKLEVGDV 136 pfam01336 1 VTVAGRV---------HSVRDSRGKLIFLTLRDGTGRIQVVVFNDEELEKLAKKLKEGDV 51 V 137 IKVLDAQVRESLSGLPELHIN 157 pfam01336 52 VGVR-GKVKKRPTGELELVVE 71
	97.4	30.6	0.27	V 233 VWICPEHGEVEPIKITILDFGLDDGSGYIRITLFGDDAEELLGVGPEEIAQKLKEMESMG 292 pfam01336 1 VTVAGRVHSVRDSRGKLIFLTLRDGTGRIQVVVFNDEEL-------EKLAKKLKE----- 48 V 293 MTLKEAARKLAEEEFYNIIGKEIIVRGNVIEDRFLGLILRASSWE 337 pfam01336 49 --------------------GDVVGVRGKVKKRPTGELELVVEEIE 74

Table 3.15 A DNA binding domain that may exit in VII Proliferating cell nuclear antigen (PCNA sliding clamp).

CD name	% aligned	Score (bits)	E value	sequence alignment
pfam04397, LytTR, LytTr DNA-binding domain. This domain is found in a variety of bacterial transcriptional regulators. The domain binds to a specific DNA sequence pattern. 105 residues.	24.8	23.6	21	VII 71 VNMDHLKKILKRGKAKDTLILKKGEE 96 pfam04397 64 VNLAKIKEIDRDFNGKRRLYLKNGEK 89

CD searches were performed on all five proteins identified by MS-Fit (Table 3.8). the identified CDs are summarized in Table 3.16.

VI-1. Conserved hypothetical protein. The best matched CDs are from phosphoesterase. Several CDs from DNA polymerase and one from eucaryal RNAP II and TFIIH were also found. All are poorly aligned and have high E values.

VI-2. GTP cyclohydrolase II. In *E. coli* this protein caralyzes the formation of a pyrimidine derivative that is used as the first intermediate for the biosynthesis of riboflavin. In addition to CDs related to this activity, one DNA binding domain in bacterial DNA topoisomerase I was recognized [Feinberg *et al.*, 1999]. Again, the low number of aligned amino acid residues and high E value suggest a possible false hit.

Table 3.16 DNA binding CDs found in proteins of the MS-fit result of VI

	CD name	% aligned	Score (bits)	Evalue	sequence alignment
VI1	COG1311, HYS2, Archaeal DNA polymerase II, small subunit/DNA polymerase delta, subunit B [DNA replication, recombination, and repair]. 481 residues	6	24.5	8.8	VI1 159 FIEEKQPKLCICGHIHE-GRGVEKIGNTI 186 COG1311 413 LVIEEVPDVFHTGHVHKFGTGVYEGVNLV 441
	COG5151, SSL1, RNA polymerase II transcription initiation/nucleotide excision repair factor TFIIH, subunit SSL1 [Transcription / DNA replication, recombination, and repair]. 421 residues	10.7	24.3	11	VI1 132 LTHAPPYGTKLDKTHSGIHAGSMGLRKFIEEKQPKLCICGHIHEGRGVE 180 COG5151 266 LSHPTDF----NGTKTDLSLVKMGFPSPMMEQLPSVCACHSEVKGGGYE 310
	COG5214, POL12, DNA polymerase alpha-primase complex, polymerase-associated subunit B [DNA replication, recombination, and repair]. 581 residues.	10.2	23.9	15	VI1 137 PYGTKLDKTHSGIHAGSMGLRKFIEEKQPKLCICG----HIHEGRGVEKIGNTIVVNPGP 192 COG5214 481 PGGSLEKCNPSSLDVVSLSLPEFMSMTAPDIYIVPSKLKHF-----CRDVGNVVVVNPGL 535 VI1 193 LARG 196 COG5214 536 QAKE 539
VI2	smart00437, TOP1Ac, Bacterial DNA topoisomerase I DNA-binding domain; Bacterial DNA topoisomerase I and III, Eukaryotic DNA topoisomeraes III, reverse gyrase alpha subunit. 260 residues.	12.7	24.9	16	VI2 134 HTEASL-ELVEMAGFKR---YALIIEILDDKGY 162 smart00437 210 YTEASLIKLMEKRGIGRPSTYAEIIETLLERGY 242
VI3	COG2344, AT-rich DNA-binding protein [General function prediction only]. 211 residues.	13.3	29.7	2.9	VI3 71 EVDVAVIAVPAKFVPDVIDECGEKGVKG 98 COG2344 146 DVEIAILTVPAEHAQEVADRLVKAGVKG 173
VI4	pfam00136, DNA_pol_B, DNA polymerase family B. This region of DNA polymerase B appears to consist of more than one structural domain, possibly including elongation, DNA-binding and dNTP binding activities. 458 residues.	15.3	23.4	6.6	VI4 19 YKIIEVKGIARGGIVMATHLGRDILALLRNIKGGEVKEYTQMMAEAREEALRRMIEH--- 75 pfam00136 97 LDVLHIVVGSLGEIFVKDGVRPGILPELLEDRLAKRKEIKKAMKESKDPEKRRLLLDKQQ 156 VI4 76 -AKKLGANAV 84 pfam00136 157 LALKVTANSV 166
VI5	LOAD_ku, ku, A family of DNA binding domains including those in Ku70/80. 521 residues.	16.9	28.7	0.92	VI5 107 DAGDVALPVSIEEALKVAEETVR-------ELREINPKALPIFLGGEHSMTLAPVRVLKP 159 LOAD_ku 274 DTGSSVLKEDTKKSYRYGSDDIVLSKDQLEQVKYFSEPGL-FLLGFKPSSMLKRHHYVRP 332 VI5 160 KSYVVFDAHLDLRDSYGGSKFNHACVARR 188 LOAD_ku 333 SVFIYPDESSIEGSTRLFSALLKRCLERD 361

VI-3. Conserved hypothetical protein. A CD predicted as AT-rich DNA-binding protein was recognized [Omura *et al*., 2001]. Only 13.3% of the sequence was aligned. The E value is 2.6.

VI-4. Conserved hypothetical protein. A CD in family B DNA polymerase that may include a DNA-binding domain was found in this protein [Edgell *et al*., 1997]. Another CD in ATP-dependent DNA ligase was also found. Both hits have low percentage of alignment and high E value.

VI-5. Agmatinase. Agmatinase is the enzyme that converts agmatine to putrescine [Wu *et al*., 1998]. One DNA binding domain was recognized in this enzyme. This domain includes the DNA binding domains in Ku70/80, which binds DNA double-strand breaks and facilitates repair by the non-homologous end-joining pathway [Aravind *et al*., 2001; Walker *et al*., 2001].

MS-MS analysis may be able to determine which of the five proteins above is the correct identity for VI.

Summary

The identification of TFB proved the fishing protocol is effective at selectively retrieving transcription-related proteins from *Pf* cell extract and can be a useful tool for probing for unidentified archaeal transcription factors. The CD search results above indicate diverse functions including DNA replication, repair, recombination, translation and possibly transcription in the proteins identified. Further characterization of these proteins is needed to confirm or eliminate their role in *Pf* transcription.

REFERENCES:

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997 Sep 1; **25**(17):3389-402.

Aravind L, Koonin EV. Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res*. 2001 Aug; **11**(8):1365-74.

Bartlett MS, Thomm M, Geiduschek EP. The orientation of DNA in an archaeal transcription initiation complex. *Nat Struct Biol.* 2000 Sep; **7**(9):782-5.

Bell SD, Brinkman AB, van der Oost J, Jackson SP. The archaeal TFIIEalpha homologue facilitates transcription initiation by enhancing TATA-box recognition. *EMBO Rep*. 2001 Feb; **2**(2):133-8.

Bell SD, Magill CP, Jackson SP. Basal and regulated transcription in Archaea. *Biochem Soc Trans.* 2001 Aug; **29**(Pt 4):392-5.

Berggren KN, Schulenberg B, Lopez MF, Steinberg TH, Bogdanova A, Smejkal G, Wang A, Patton WF. An improved formulation of SYPRO Ruby protein gel stain: comparison with the original formulation and with a ruthenium II tris (bathophenanthroline disulfonate) formulation. *Proteomics.* 2002 May;**2** (5):486-98.

Borges KM, Bergerat A, Bogert AM, DiRuggiero J, Forterre P, Robb FT. Characterization of the reverse gyrase from the hyperthermophilic archaeon Pyrococcus furiosus. *J Bacteriol*. 1997 Mar; **179**(5):1721-6.

Cann IK, Ishino S, Hayashi I, Komori K, Toh H, Morikawa K, Ishino Y. Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea. *J Bacteriol*. 1999 Nov; **181**(21):6591-9.

Chen HT. Structural study of eucaryal and archaeal transcription factors TF(II)B. Ph.D. Dissertation, University of Georgia, Athens, 2000.

Doolittle WF. Phylogenetic classification and the universal tree. *Science*. 1999 Jun 25; **284**(5423):2124-9.

Edgell DR, Klenk HP, Doolittle WF. Gene duplications in evolution of archaeal family B DNA polymerases. *J. Bacteriol.* 1997 Apr; **179**(8):2632-40.

Feinberg H, Changela A, Mondragon A. Protein-nucleotide interactions in E. coli DNA topoisomerase I. *Nat Struct Biol.* 1999 Oct; **6**(10):961-8.

Gabrielsen OS, Hornes E, Korsnes L, Ruet A, Oyen TB. Magnetic DNA affinity purification of yeast transcription factor tau--a new purification principle for the ultrarapid isolation of near homogeneous factor. *Nucleic Acids Res*. 1989 Aug 11; **17**(15):6253-67.

Kanemaki M, Makino Y, Yoshida T, Kishimoto T, Koga A, Yamamoto K, Yamamoto M, Moncollin V, Egly JM, Muramatsu M, Tamura T. Molecular cloning of a rat 49-kDa TBP-interacting protein (TIP49) that is highly homologous to the bacterial RuvB. *Biochem Biophys Res Commun*. 1997 Jun 9; **235**(1):64-8.

Komori K, Ishino Y. Replication protein A in Pyrococcus furiosus is involved in homologous DNA recombination. *J Biol Chem*. 2001 Jul 13; **276**(28):25654-60.

Lewis, L. M. Structural and Functional Characterization of the Transcription

Preinitiation Complex from Pyrococcus furiosus, Ph.D. Dissertation, University of Georgia, Athens, 2000.

Lopez MF, Berggren K, Chernokalskaya E, Lazarev A, Robinson M, Patton WF. A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis*. 2000 Nov; **21**(17):3673-83.

Marchler-Bauer A, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S, Hurwitz DI, Jackson JD, Jacobs AR, Lanczycki CJ, Liebert CA, Liu C, Madej T, Marchler GH, Mazumder R, Nikolskaya AN, Panchenko AR, Rao BS, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Vasudevan S, Wang Y, Yamashita RA, Yin JJ, Bryant SH. CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res*. 2003 Jan 1; **31**(1):383-7.

Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res*. 2002 Jan 1; **30**(1):281-3

Matsumiya S, Ishino Y, Morikawa K**.** Crystal structure of an archaeal DNA sliding clamp: proliferating cell nuclear antigen from Pyrococcus furiosus. *Protein Sci*. 2001 Jan; **10**(1):17-23.

Murzin AG. OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J*. 1993 Mar; **12**(3):861-7.

Nikolskaya AN, Galperin MY. A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. *Nucleic Acids Res.* 2002 Jun 1; **30**(11):2453-9.

Ogura T, Wilkinson AJ. AAA+ superfamily ATPases: common structure--diverse function. *Genes Cells*. 2001 Jul; **6**(7):575-97.

Omura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, Shinose M, Takahashi Y, Horikawa H, Nakazawa H, Osonoe T, Kikuchi H, Shiba T, Sakaki Y, Hattori M. Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites. *Proc Natl Acad Sci* U S A. 2001 Oct 9; **98**(21):12215-20.

Patel S, Latterich M. The AAA team: related ATPases with diverse functions. *Trends Cell Biol*. 1998 Feb; **8**(2):65-71.

Qureshi SA, Bell SD, Jackson SP. Factor requirements for transcription in the Archaeon Sulfolobus shibatae. *EMBO J*. 1997 May 15; **16**(10):2927-36.

Shimizu K, Kawasaki Y, Hiraga S, Tawaramoto M, Nakashima N, Sugino A. The fifth essential DNA polymerase phi in Saccharomyces cerevisiae is localized to the nucleolus and plays an important role in synthesis of rRNA. *Proc Natl Acad Sci* U S A. 2002 Jul 9; **99**(14):9133-8.

Soppa J. Transcription initiation in Archaea: facts, factors and future aspects. *Mol Microbiol*. 1999 Mar; **31**(5):1295-305.

Thomm M. Archaeal transcription factors and their role in transcription initiation. *FEMS Microbiol Rev*. 1996 May; **18**(2-3):159-71.

Walker JR, Corpina RA, Goldberg J. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature*. 2001 Aug 9; **412**(6847):607-14.

Winker S, Woese CR. A definition of the domains Archaea, Bacteria and Eucarya in terms of small subunit ribosomal RNA characteristics. *Syst Appl Microbiol*. 1991; **14**(4):305-10.

Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci* U S A. 1977 Nov; **74**(11):5088-90.

Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci* U S A. 1990 Jun; **87**(12):4576-9.

Woese CR. The universal ancestor. *Proc Natl Acad Sci* U S A. 1998 Jun 9; **95**(12):6854-9.

Woese CR. Interpreting the universal phylogenetic tree. *Proc Natl Acad Sci* U S A. 2000 Jul 18; **97**(15):8392-6.

Wu G, Morris SM Jr. Arginine metabolism: nitric oxide and beyond. *Biochem J*. 1998 Nov 15; **336** (Pt 1):1-17.

Yee A, Booth V, Dharamsi A, Engel A, Edwards AM, Arrowsmith CH. Solution structure of the RNA polymerase subunit RPB5 from Methanobacterium thermoautotrophicum. *Proc Natl Acad Sci* U S A. 2000 Jun 6; **97**(12):6311-5.