DISCOVERY OF A POTENTIAL REGULATORY TRANSCRIPTION FACTOR UPON REDOX ALTERATIONS IN *SHEWANELLA ONEIDENSIS* MR-1

by

XIAYU HUANG

(Under the Direction of Robert A. Scott)

ABSTRACT

Shewanella oneidensis is a model organism for bioremediation studies because of its diverse respiratory capabilities and the regulatory mechanisms underlying the ability of *S. oneidensis* to survive and adapt to various environmental stresses is poorly understood. This work describes the discovery and characterization of a potential γ-proteobacerial regulatory transcription factor, SO3363p, of the mesophile *Shewanella oneidensis*. An integrated approach was used including genomics, proteomics and bioinformatics. A DNA affinity protein capture experiment was employed to capture transcription factors responsible for the observed gene regulation in DNA microarray expression profiles of the transcriptional response of *S. oneidensis* to growth in the presence or absence of oxygen (O_2). The recombinant SO3363p was verified to bind specifically to promoters by electrophoretic mobility shift assay (EMSA) and the recognized DNA-binding motif was determined by SELEX. In order to understand the role of SO3363 in *S. oneidensis*, SO3363 deletion strain was constructed and subjected to both physiological characterization and microarray analysis. The phenotypic microarray revealed that the mutant exhibited a defect in utilizing uridine, 2-deoxy adenosine, pyruvic acid, etc. compared with the wild type MR-1; Growth assays showed that the deletion mutant is more resistant to H_2O_2; Whole genome DNA microarray analyses showed that a number of genes' expression was altered in mutant strain under the conditions tested, including *nqr* regulon involved in Na^+ translocation, three chemotaxis proteins responding to the change in chemical concentration with anaerobic growth condition; *cys* regulon related to cysteine biosynthesis, *pur* regulon involved in de novo synthesis of purine with H_2O_2 growth condition.

INDEX WORDS: *Shewanella oneidensis*, regulatory transcription factor, DNA microarray, EMSA, SELEX, phenotypic microarray, *nqr* regulon, Na+ translocating, chemotaxis, *cys*, *pur*, *de novo*.

DISCOVERY OF A POTENTIAL REGULATORY TRANSCRIPTION FACTOR UPON REDOX ALTERATIONS IN *SHEWANELLA ONEIDENSIS* MR-1

by

XIAYU HUANG

B.S., East China University of Science and Technology, China, 1999

M.S., East China University of Science and Technology, China, 2002

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2008

DISCOVERY OF A POTENTIAL REGULATORY TRANSCRIPTION FACTOR UPON REDOX ALTERATIONS IN *SHEWANELLA ONEIDENSIS* MR-1

by

XIAYU HUANG

Major Professor: Robert A. Scott

Committee: Jonathan I. Amster
Robert S. Phillips

Electronic Version Approved:

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

DEDICATION

To my dear parents Yong Huang and Lirong Yu, to my brother Chunyu Huang for their encouragement and unconditional love, to my sweet husband Yanghui Zhang who has been my biggest support and has made my life so colorful.

ACKNOWLEDGEMENTS

Many people provided valuable guidance and advice through the course of my graduate studies. First, I want to thank my major advisor Professor Robert A. Scott for the opportunity to work in his research group and for being very supportive of my scientific research.

Next, I want to thank Dr. Amster and Dr. Phillips for their useful advice.

I am very grateful to our project collaborator Dr. Zhou's laboratory at the University of Oklahoma. Many thanks to the two guys in his group including Zhili He, who designed the *Shewanella oneidensis* whole genome DAN microarray and Haichun Gao, who helped me with all the microarray experiments; Ken Jones from Travis Glenn's lab, who contributed all the microarray data analysis; Gerti Schut from Dr. Adams' group in the Department of Biochemistry and Molecular Biology, who also helped me with the miroarray data analysis, and Frank Jenney from this group too, who provided me the vectors for cloning; Frank Sugar from Bioexpress lab helped me with protein purification and Dennis Phillips from Mass Spec Facility helped me with protein identification with Mass Spectrometry, Ed Johnson from IBL helped me arrange some of the microarray data.

Many thanks to all my current and former group members in Department of Chemistry for being a team and caring for each other. Particularly, I want to thank Gina Lipscomb, who is so kind and always being there to help us not only at research but at other aspects as well.

I also want to thank my dear parents for their endless love and encouragement, and my dear husband Yanghui Zhang for his love and support, which makes my life so sweet.

TABLE OF CONTENTS

Page

CHAPTER 1

TRANSCRIPTIONAL REGULATION IN γ-PROTEOBACTERIA AND APPROACH TO TRANSCRIPTION FACTORS DISCOVERY

1.1 Introduction to γ –proteobacteria

Prokaryotes are the most abundant and diverse forms of life on earth and bacteria are one of the major groups [1]. The Proteobacteria are a major phylum of bacteria and they include a wide variety of pathogens, such as *Escherichia*, *Salmonella*, *Vibrio*, *Helicobactera* and many other notable genera. Others are free-living, and include many of the bacteria responsible for nitrogen fixation [2]. The group is defined primarily in terms of ribosomal RNA (rRNA) sequences [3].

All Proteobacteria are Gram-negative, with an outer membrane mainly composed of lipopolysaccharides and most members are facultative anaerobes and heterotrophs. Many move using flagella, but some are non-motile or rely on bacterial gliding [4].

The proteobacteria consist of four subdivisions, α, β, δ/ξ and γ. The γ -proteobacteria are of particular interest because this group of bacteria comprise several medically and scientifically relevant members, such as the *Enterobacteriaceae*, *Vibrionaceae* and *Pseudomonadaceae* and an expanding number of pathogens also belong to this class, e.g. *Salmonella* (enteritis), *Escherichia coli O157:H7* (food poisoning), etc. Among all the γ –proteobacteria, *E. coli* has been most intensively studied in every aspect, including the transcription and transcriptional regulation processes associated with different environmental stimuli: it is a model system for studying other organisms.

Since more and more bacterial genomes including some γ-proteobacterial genomes are being sequenced and far less is known about the transcriptional regulatory elements in these bacterial genomes, it is of great interest to decipher the transcriptional regulation in these less studied organisms based on the knowledge from *E. coli*.

The work presented here is primarily for the purpose of learning more about these characteristic organisms, particularly their transcriptional apparatus and regulatory process in order to further understand the evolution of the transcriptional mechanism of these microorganisms.

1.2 Bacterial basal transcription and transcriptional regulation

1.2.1 Bacterial basal transcription

Transcription is the process of RNA synthesis from DNA template by DNA-dependent RNA polymerases within a cell. The messenger RNA synthesized from the transcription process will be used as template for the purpose of protein synthesis via the translation process, wherein the instructions encoded in the sequence of bases in mRNA are translated into a specific amino acid sequence by ribosomes, the "workbenches" of polypeptide synthesis. In prokaryotes, virtually all RNA is synthesized by a single of DNA-dependent RNA polymerase, whereas eukaryotes have three RNAPs, responsible mainly for the synthesis of ribosomal RNA, messenger RNA, and transfer small RNA [5]. The RNA polymerase in *E. coli*, so called core RNAP, contains five subunits: a dimer of α (329 amino acids), β' (1407 amino acids), β (1342 amino acids) and the ω subunit (91 amino acids). The core RNAPs of eukaryotes and archaea share the basic $\alpha_2\beta'\beta\omega$ subunit composition (with other names) but contain 5-8 additional subunits that are absent from bacterial core enzymes [6] (see Figure 1.1).

Bacterial core RNAP is catalytically proficient for synthesizing an RNA copy from a DNA template but is unable to initiate transcription from promoters. In bacteria, transcription initiation requires the binding of one of the specificity σ subunits to the core RNAP to form the RNAP holoenzyme that can recognize DNA promoters and initiate transcription. There are more than one type of holoenzymes and they all contain the core RNAP $\alpha_2\beta'\beta\omega$ subunits but differ in the σ subunit. Most bacteria have several different σ subunits, with *E. coli* having seven: σ^{70}, σ^{32}, σ^{54}, σ^{S}, σ^{F}, σ^{E}, σ^{FecI}. σ^{70} and σ^{54} are the most important ones in bacterial transcriptional regulation. The availability of different holoenzymes is controlled by various environmental stimuli [7] and they transcribe different classes of genes, depending on the specificity of the promoter-binding sequence in different promoters [8].

Bacterial basal transcription is relatively simpler compared to that of eukaryotes due to the complexity of RNA polymerase and the presence of multiple general transcription factors in eukaryotes [9]. Bacterial basal transcription machinery only requires RNAP holoenzyme and a promoter which is a region of DNA on the genome where RNAP holoenzyme binds to initiate transcription (see Figure 1.2). The promoter contains two consensus DNA elements, -10 hexamer TATAAT and -35 hexamer TTGACA, located at –10 bp and –35 bp upstream of the transcription start site (+1). It also contains an extended -10 and UP (upstream promoter) element located ~20 bp upstream of the -35 element. The -10, -35 and extended -10 elements are recognized by different domains of the RNAP σ subunit [10, 11] and the UP element is recognized by the C-terminal domains of the RNAP α subunit.

The bacterial transcription process includes transcription initiation, elongation and termination. Transcription initiation is a multistep process [12] and is the most important regulated stage of the transcription cycle. First, the RNAP holoenzyme binds to the promoter

region of an ORF to form an closed binary complex; second, DNA within the binding region melts, leading to the formation of an open complex; third, the open complex leads to the formation of transcription initiated complex in the presence of nucleotide phosphates and generates a short RNA chain; fourth, RNAP breaks contact with the promoter by releasing the σ factor and escapes as an elongation complex. This transcription initiation mechanism is not conserved among bacteria, eukaryotes and archaea, unlike the RNAP catalytic mechanism [6]. Bacterial transcription initiation needs σ factor to recruit RNAP to promoters while eukaryotes and archaea rely on protein-protein interactions to initiate transcription.

1.2.2 Bacterial transcriptional regulation

Bacteria live in habitats of frequently changing conditions, such as the fluctuation of temperature, pH, oxygen, salinity, light, etc., and have therefore developed sophisticated response to adapt to environmental stimuli. These responses often lead to the activation and /or repression of a number of genes to change cell physiology or metabolism for adaptation [12]. Bacterial have developed a variety of mechanisms to regulate gene expression affecting every step from transcription initiation to protein inactivation or degradation. Gene expressions can be turned on or off by various factors, such as by the action of regulatory proteins activating or inhibiting transcription initiation [13], or by modulation of transcription elongation and termination at specific sites [14, 15], by RNA polymerase (RNAP) modifications [16], by DNA rearrangements connecting a gene to or disconnecting it from a particular promoter [17], etc. Among all the factors affecting gene expressions, activating and inhibiting transcription initiation by regulatory proteins is of great interest as a primary response.

Regulation of gene transcription generally involves the DNA-binding transcription factors (TF) as the key component. In general, TFs regulate specific genes transcription by

recognizing a binding site (an operator) in the specific promoter region. This binding is mediated by specific protein motifs on the TF called DNA-binding motifs including the most characterized ones: the helix-turn-helix, the winged helix, leucine zipper, zinc finger and β ribbons [18]. TFs can function as activators, repressors, or dual regulators. In *E. coli* K-12 genome, there are about 314 genes annotated as TFs including ~110 activators, 135 repressors and 69 dual regulators [19]. Seven of them (CRP, FNR, IHF, Fis, ArcA, NarL and Lrp) are global transcription factors which control 50% of all the regulated genes. Bacterial transcription factors can be grouped into different families on the basis of sequence analysis and the best characterized ones include the LacI, AraC, LysR [41], CRP and OmpR families [20].

The mechanism of bacterial TFs to activate and repress gene transcription varies widely. There are two major types of mechanisms of activating transcription initiation in *E. coli*, depending on the type of RNAP holoenzyme involved. The first type of activating mechanism [21] uses α and σ^{70} subunits of RNAP as the contact points for activators, such as CRP at the *lac* promoter in *E. coli*. The σ^{70} subunit binds -10 and -35 elements and the α subunit binds the AT-rich UP element on the promoter region. CRP [22] binds to the *lac* promoter near -60 and has an direct interaction with the carboxy-terminal domain of the α subunit (αCTD), which helps recruit the RNA polymerase to the promoter to initiate transcription (see Figure 1.4a). This is called Class I activation. Besides that, CRP can also bind the *proP P2* at -41 position that overlaps the -35 element on the promoter region to make a direct contact with the σ^{70} subunit to activate the transcription [23] (Figure 1.4b). This is called Class II activation. It is unusual that some activators, e.g. such as MerR-type activators, can alter the conformation of the target promoter to enable a productive interaction of RNA polymerase with the promoter –10 and/or –35 elements. They bind the promoter DNA between -10 and -35 elements and twist the DNA to reorient the -

10 and -35 elements so that they can be bound by the RNA polymerase σ subunit (Figure 1.4c) to initiate the transcription [24].

The second type of activating mechanism occurs through enhancer elements to deliver the activator by looping out the DNA or tracking along the DNA and σ^{54} subunit, as exemplified by some bacteriophage T4 TFs. *NtrC* is an enhancer protein in *E. coli* that can activate the transcription of the gene controlling nitrogen metabolism by binding to an enhancer located -100 bp upstream of the promoter for *glnA*. The enhancer keeps *NtrC* in high local concentration near the promoter such that it can have frequent interaction with the σ^{54} subunit. The formation of the NtrC complexes can function to convert the polymerase from the inactive form to one that can melt the start site of DNA to initiate transcription [25].

The mechanisms of TFs to repress transcription initiation can be divided into three types, depending on when it occurs during the initiation process [12]. The first type of repression is the most common one and is caused by binding of a repressor protein to the promoter in a way that hinders RNAP binding. It can be achieved by binding a TF to the promoter region that overlaps the RNAP binding site to impede RNAP binding (see Figure 1.5a), such as the LexA repressor on the *uvrA* promoter [26] or bend the DNA, causing steric hindrance for RNAP binding, such as the *B. subtilis* p4 repressor on the *A2b* promoter [27]. Transcription initiation can be repressed when multiple repressor molecules bind to promoter-distal sites and cause DNA looping, which shuts off transcription initiation in the looped domain [28] (see Figure 1.5b). Another mechanism involves binding DNA with low sequence specificity but covering large DNA regions flanking the promoter region [29]. A more complicated case is that in which the repressor binds to an activator on the promoter and prevents the activator from functioning by blocking promoter

recognition by the RNA polymerase holoenzyme; in this case the repressor functions as an anti-activator (Figure 1.5c), such as CRP (repressor) and CytR (activator) on the *CytR* promoter [30].

The second type of repression is caused by blocking the transition of the closed RNAP-DNA complex to the open complex. This happens when repressors and RNAP bind promoter DNA simultaneously to form a ternary complex which will prevent RNAP from forming the open complex [31], or the repressor binds to the promoter region between the RNAP and the activator and causes DNA bending that inhibits the interaction between the activator and RNAP [32].

The third type of repression is caused by inhibiting promoter clearance. This occurs when the binding between the TFs and promoter DNA is too tight so that the RNAP cannot break contact with the promoter, such as for the *B. subtilis* p4 repressor on the *A2c* promoter [33, 34, 35] or when the TFs bind downstream from RNAP, such as the LacI repressor on phage T7 late promoter [36].

1.3 Redox stress and transcription factors mediating its response

Cells undergo significant alterations in their physiology when they are exposed to reactive oxygen and nitrogen intermediates including hydroxyl radical (OH·), hydrogen peroxide (H_2O_2), superoxide anion ($O_2^{\cdot -}$), and NO·, which can cause damage to proteins, nucleic acid and cell membranes, or when their growth condition shifts from aerobiosis to anaerobiosis (or vice versa). To handle the redox stress, cells utilize different transcription factors to control the expression of enzymes to detoxify the reactive oxygen species and/or repair the damage [37]. These transcription factors use different mechanisms to sense the redox stress [38].

In *E.coli*, SoxR/SoxS and FNR are the two best characterized TFs using iron sulfur centers to sense the redox stress and all of them have helix-turn-helix DNA-binding motifs [38, 39, 40]. SoxR is a homodimer in solution, containing two [2Fe-2S] centers, each bound to 4 cysteines, and the oxidized SoxR binds the *soxS* promoter region between -10 and -35 elements which is also recognized by the σ^{70} subunit to activate the transcription of *soxS* whose product will then activate the transcription of 10 proteins, such as endonuclease IV, Mn-superoxide dimutase and NADH:ferredoxin oxidoreductase, etc., to provides protection against superoxide ($O_2^{\cdot -}$).

FNR is a global TF in *E. coli* controlling ca. 120 genes' aerobic-anaerobic regulation, including succinate dehydrogenase, fumarase, etc. FNR activates gene transcription by direct interact with the RNAP α subunit. FNR contains a ferredoxin-like four-cysteines cluster and it senses redox through an iron center coordinated by four cysteines. When the growth shifts from aerobic to anaerobic, the [2Fe-2S] center is converted into [4Fe-4S] center converting FNR into a homodimer with high DNA-binding affinity.

In *E. coli*, OxyR is the TF that is responsible for the adaptation to H_2O_2 challenge and it induces approximately 30 proteins including those encoding catalase, alkyl hydroperoxide reductase, glutathione reductase, *Dps*, etc. [41, 42]. OxyR exists as a tetramer in solution and it activates transcription initiation of genes by interacting with RNAP α subunit. Upon oxidation, a disulfide bond is formed between two cysteines, C_{199} and C_{208}, which stabilizes the conformational change and favors DNA-binding.

There are some two-component regulatory systems that sense redox alterations such as ArcA/ArcB, NarX/NarL in *E. coli*, RegA/RegB in *Rhodobacter capsulatus*, and FixL/FixJ in *Rhizobium meliloti* [43, 44, 45, 46]. Usually, one of the components is a sensor kinase which

undergoes autophosphorylation on one of its histidines in response to decreased redox potential such as anaerobic or microaerobic respiratory conditions and then transfers a phosphate to an aspartate site on the other member which is a DNA-binding response regulator containing the helix-turn-helix DNA binding motif. The phosphorylated response regulator activates or represses the transcription of certain genes under certain conditions. ArcA is known as a global transcriptional regulator in *E. coli* and other γ-proteobacteria [47] and the ArcA/ArcB regulon in *E. coli* consists of over 30 loci, including flavoprotein-type dehydrogenase, cytochrome oxidase, etc. ArcB is autophosphorylated when growth conditions become anaerobic by sensing the redox state of the quinone pool [48, 49]. Under these conditions, ArcB stimulates changes in gene expression by transferring a phosphate group to ArcA [50], which then becomes competent to bind DNA [51, 52]. For FixL/FixJ [38, 53] which controls the expression of the *nif* genes under oxygen- and nitrogen- limiting conditions, oxygen is sensed through the bound heme which is attached to His_{194} in the regulatory domain close to the autophosphorylation histidine site in FixL. Phosphorylated FixJ by FixL-P, is the response regulator that mediates the transcription activation of *nifA* via interaction with σ^{70} of RNAP. How the other two-component regulatory systems sense the redox stress needs to be further investigated.

NifL is another redox response regulator in γ-proteobacteria [38, 54]. It has an N-terminal PAS domain, which is similar to the heme-binding domain in FixL, and a C-terminal domain similar to the transmitter domain of a histidine sensor kinase. However, NifL does not undergo autophosphorylation as most of the two-component regulatory systems do and instead, it senses redox stress by binding flavin at the N-terminal domain and it binds the promoter DNA *nifA* through the carboxy-terminal domain.

1.4 The model *Shewanella oneidensis* MR-1

One of the important model γ-proteobacterial organisms is the mesophile *Shewanella oneidensis* (formerly *Shewanella putrefaciens*). There are tens of Shewanella species isolated and these bacteria inhabit diverse environments, including spoiled food [55], infected animals [56], deep-sea [57], oil-field waste sites [58], and fresh water lake sediments [59]. *Shewanella oneidensis* MR-1 is a facultative anaerobe with a maximum growth temperature of ~35 °C, but the ability to grow over a wide range of temperatures, including temperatures near zero [60]. It was first isolated from sediments of Lake Oneida in New York State [61]. *Shewanella oneidensis* is unique among the γ-proteobacteria because of its diverse respiration capabilities. It can utilize different organic and inorganic substrates as terminal electron acceptors under anaerobic respiration. These include fumarate, nitrate, TMAO, Fe(III), Mn(III) and (IV), Cr(VI), and U(VI) [61,62,63], as well as using oxygen as terminal electon acceptors under aerobic growth conditions [64]. Its ability to anaerobically respire, causing metal ion reduction, has been exploited for the bioremediation of metal contaminants in the environment [65]. It is also of interest to the food industry and medicine since it can spoil food and act as an opportunistic pathogen[66]. For the past 30 years, lots of effort has been done on Shewanella based on the knowledge of the related *E. coli*. However, so far, little is known about the genetic basis and regulatory mechanisms underlying the microorgansim's response to diverse enviromental stresses such as limiting oxygen and nutrient, H_2O_2 challenge, extreme pH, etc.

Currently, the study of *Shewanella oneidensis* MR-1 is part of DOE's Genomics: GTL program under the direction of Shewanella Federation which is a cross-institution consortium consisting of teams of scientists from academia, national laboratories, and private industry

working collaboratively to achieve a systems-level understanding of how *S. oneidensis* senses and responds to its environment (www.shewanella.org).

1.5 Method of transcription factor discovery and characterization

An integrated approach coupling genomics, proteomics, and bioinformatics, has been developed to discover and predict regulatory transcription factors (rTFs), their DNA binding sites (TFBS), and the genes they regulate, associated with a given environmental stress. For any given (prokaryotic) organism, first a stress or metabolite of interest is chosen defining two growth conditions A (control) and B (stress), used to generate cell extracts from which RNA pools are analyzed by microarray expression profiling which can demonstrate gene expression changes in transcript levels between two growth conditions. Presumably, observed up or down-regulation of a transcript for a given gene in response to a particular growth condition as compared to a control growth condition is the result of a regulatory transcription factor (rTF) exerting its effect in response to the environmental change or stress and the upstream DNA of the these genes or operons should contain the binding site of the transcription factors to initiate the transcription. Experimentally, the upstream DNA (promoter/operator regions) of these highly up- or down-regulated genes selected is amplified and used to capture sequence-specific DNA-binding proteins as prospective rTFs. One-dimensional SDS-PAGE gels are used to identify prospective rTFs that bind differentially from cell extract A compared to B and protein bands of interest are in-gel digested and identified by mass spectrometry. Identified proteins are prioritized for characterization based on sequence similarity to known transcription factor families and then the recombinant prospective TFs are cloned, expressed and purified. The protein-DNA interactions are determined by electrophoretic mobility shift assay (EMSA) and the

protein binding site can be determined by footprinting and SELEX. For predicted rTFs, DIP-chip is used to determine the genome-wide localization of transcription factor binding site and deletion mutagenesis is used for functional characterization followed by comparing the phenotype of the mutant with wild type by phenotype microarray, as well as the genotype by standard expression profile microarray. That is, the genes regulated by this potential rTF can be determined by comparing the gene expression patterns between the mutant and the wild type under specific growth conditions known to be important for the rTF to regulate transcription. The work described in this dissertation focused on the identification and characterization of a potential regulatory transcription factor, SO3363p, involved in *Shewanella oneidensis* MR-1 redox response. This work took advantage of a collaboration with the group of Jizhong Zhou, originally at the Oak Ridge National Laboratory and a member of the Shewanella Federation. Some of the work was done in Dr. Zhou's laboratory after his move to the University of Oklahoma.

Figure 1.1 Basal transcriptional machinery of Bacteria, Archaea and Eukarya. RNA polymerase subunits (top) and general transcription factors (bottom) are color-coded according to their homologies across domains and protein/subunit names are indicated. Figure adapted from [5].

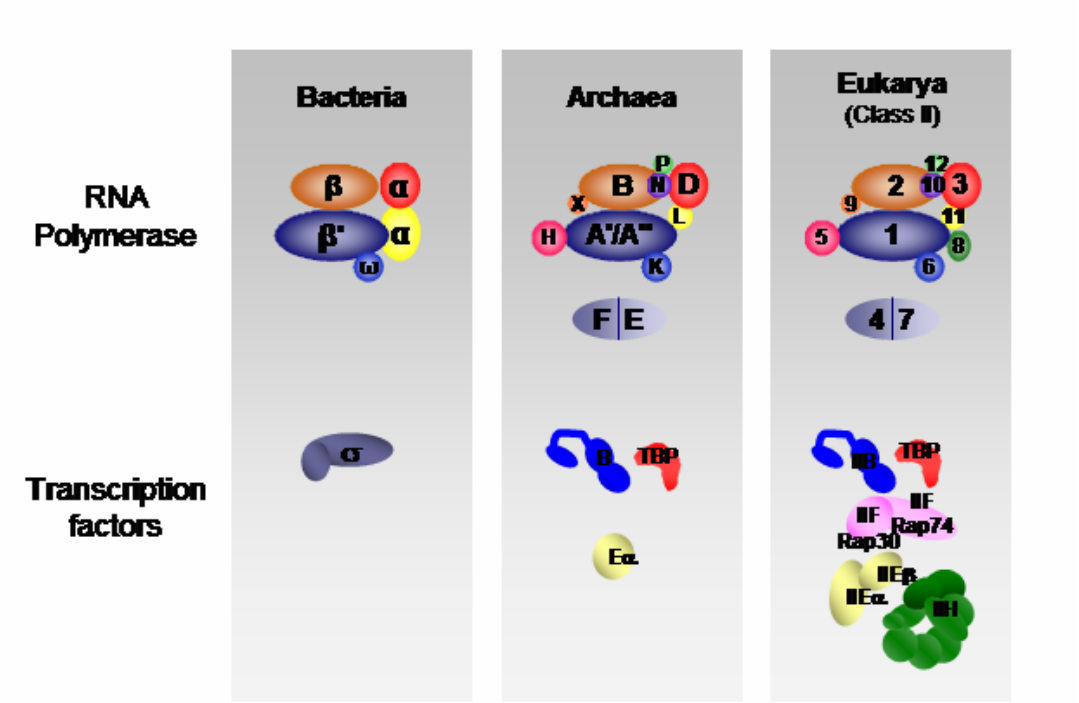
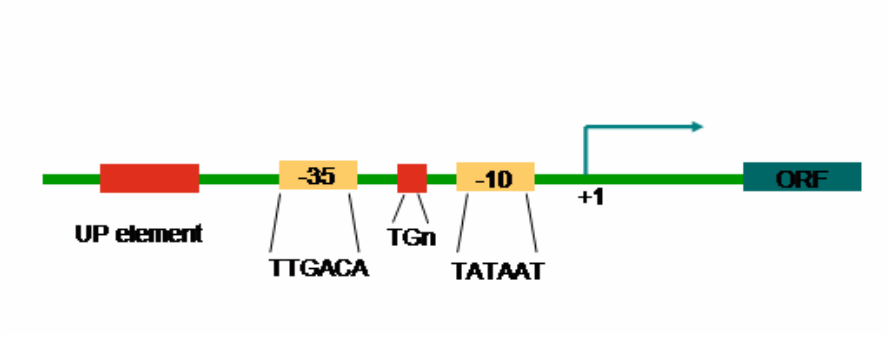

Bacteria
Archaea
Eukarya
(Class II)
RNA
Polymerase
Transcription
factors
β
α
β'
α
ω
B
N
D
P
X
L
H
A'/A"
K
F
E
2
12
10
3
9
11
5
1
8
6
4
7
σ
B
TBP
Eα
IIB
TBP
IIF
IIF
Rap74
Rap30
IIEβ
IIEα
IIH

Figure 1.2 Structure of bacterial core promoter and bacterial basal transcription. Promoter is the region of DNA on the genome where RNA polymerase and TF bind to initiate transcription. +1 indicates the transcription start point. Base pairs upstream of the transcription start point are assigned positive numbers. The core promoter consists of –10, –35 and extended –10 (TGn) and UP elements. The –10, –35 and extended –10 elements are recognized by domains 2, 4, and 3 of the RNA polymerase σ subunit, respectively. The UP element, located upstream of the –35 element, is recognized by the C-terminal domains of the RNA polymerase α subunits. Figure adapted from [10, 11].


-35
-10
+1
ORF
UP element
TGn
TTGACA
TATAAT

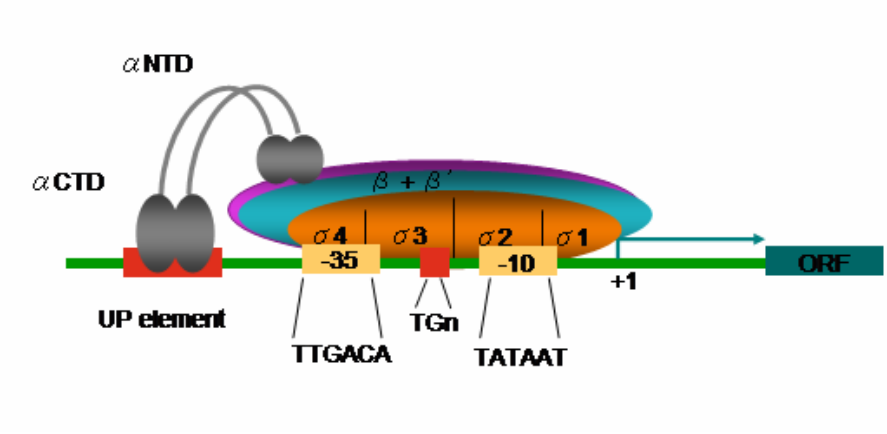

αNTD
αCTD
β + β´
σ4
σ3
σ2
σ1
-35
-10
+1
ORF
UP element
TGn
TTGACA
TATAAT

Figure 1.3 Transcription initiation process at bacterial promoters [12]. First, the RNA polymerase (R) interacts with promoter DNA (P) to form the closed complex (RP_C). Dashed lines show the position on the promoter DNA that is bound by the RNA polymerase holoenzyme. Second, the duplex DNA around the transcript start site is unwound to form the open complex (RP_O). Third, the initiation complex (RP_{INIT}) is formed and synthesis of the DNA-template-directed RNA chain (shown as a dashed red line) begins. Fourth, RNA polymerase breaks contact with promoter DNA, releasing σ factor, and escapes as an elongation complex as the RNA chain length increases, shown as a solid red line. Figure adapted from [11].

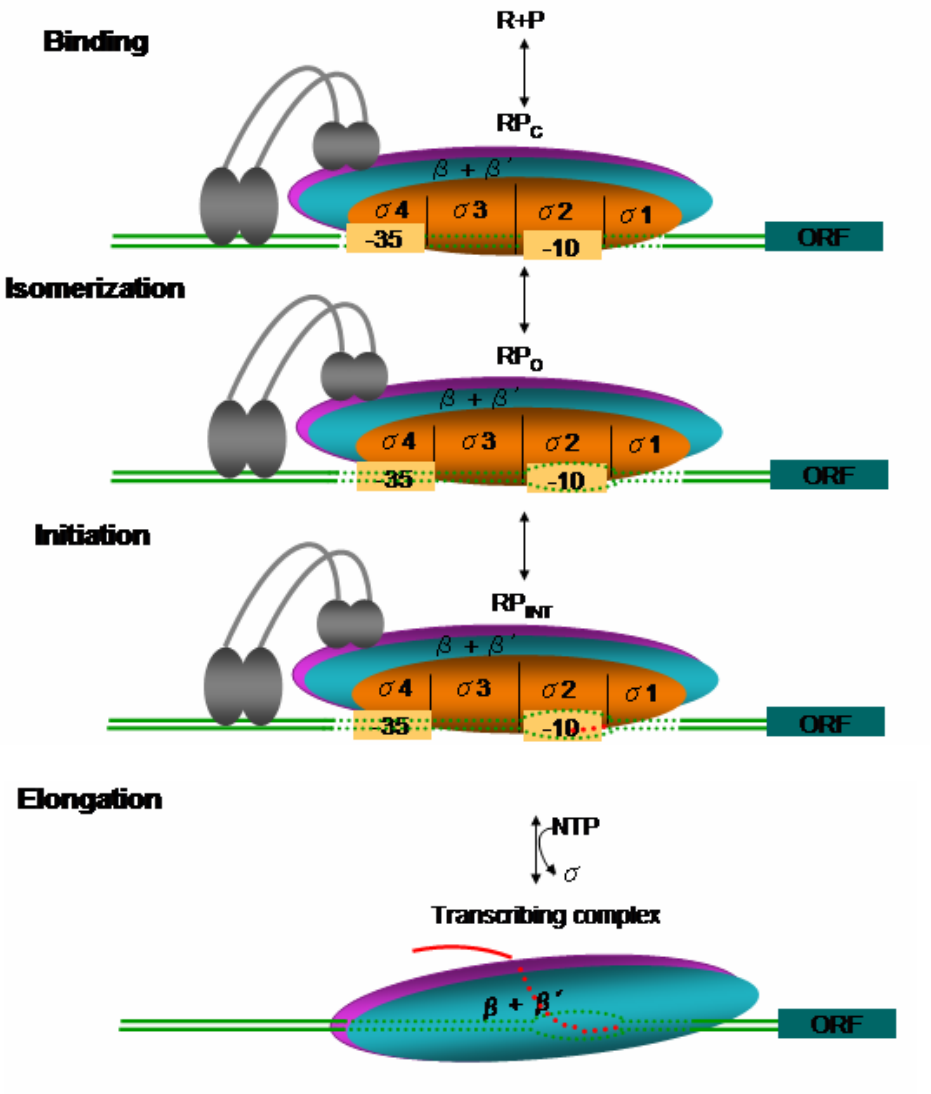
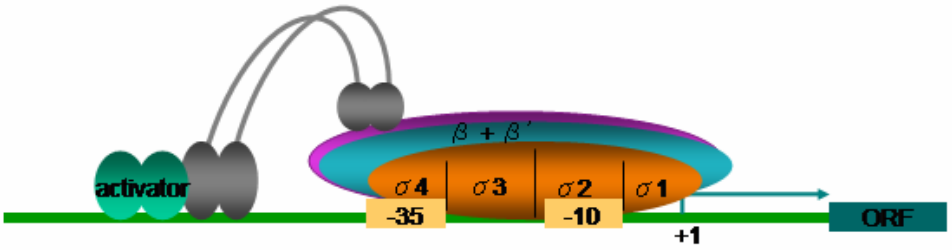
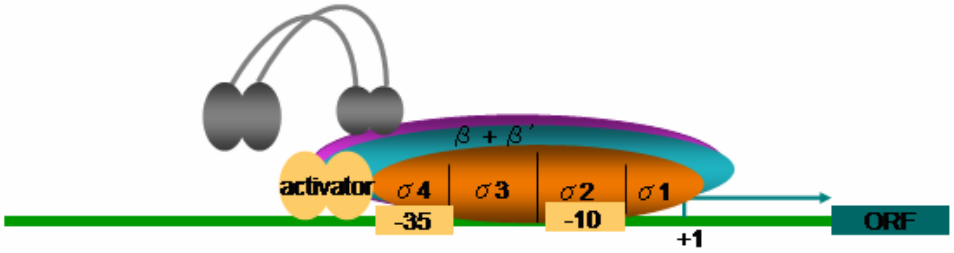

Binding
R+P
RP$_C$
β + β′
σ4
σ3
σ2
σ1
-35
-10
ORF
Isomerization
RP$_O$
β + β′
σ4
σ3
σ2
σ1
-35
-10
ORF
Initiation
RP$_{INT}$
β + β′
σ4
σ3
σ2
σ1
-35
-10
ORF
Elongation
NTP
σ
Transcribing complex
β + β′
ORF

Figure 1.4 Simple transcriptional activation modes. The figure illustrates the organization of RNA polymerase subunits and activators during activation at simple promoters. **a.** Class I activation. The activator (shown in dark green) is bound to an upstream site and contacts the αCTD of RNA polymerase, thereby recruiting the RNA polymerase to the promoter. **b.** Class II activation. The activator (shown in orange) binds to the promoter with binding site partially overlapping with the –35 promoter element, which is recognized by domain 4 of σ^{70} of RNA polymerase. **c.** Activation by conformation changes. The activator (shown in blue) binds between the –10 and the –35 promoter elements, which is the recognition site of σ subunit and realigns them so that the RNA polymerase holoenzyme can bind to the promoter to initiate the transcription. Figure adapted from [11].

a. Class I activation



b. Class II activation



c. Activation by conformational change

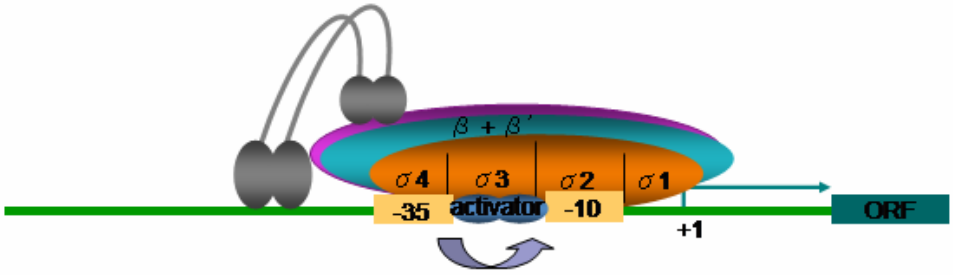
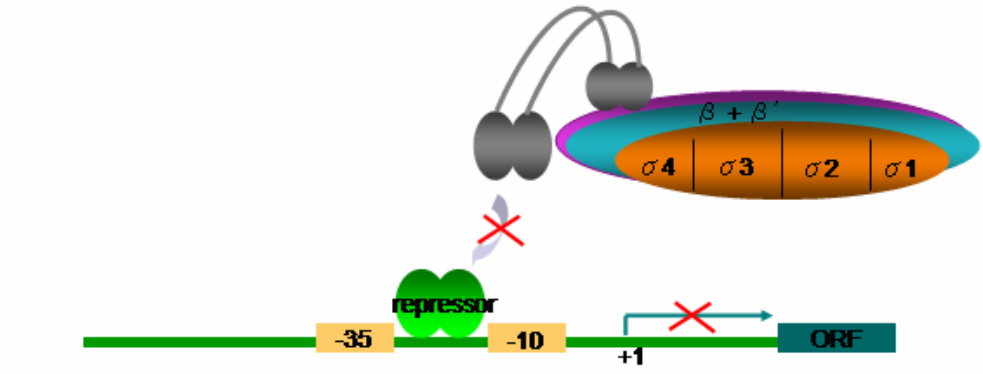
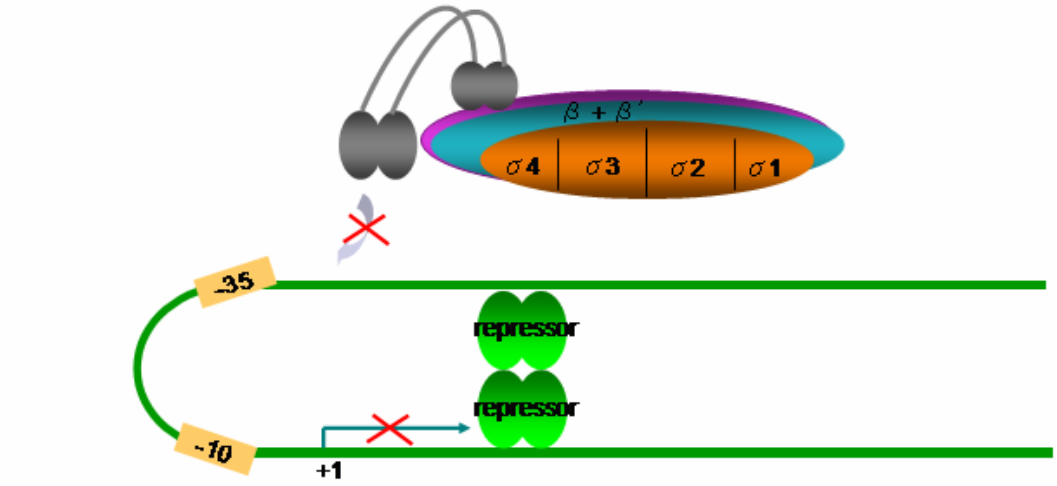


Figure 1.5 Mechanisms of repression at transcription initiation binding step. a. Repression by steric hindrance. The repressor binding site overlaps core promoter elements and blocks recognition of the promoter by the RNA polymerase holoenzyme. **b.** Repression by DNA looping. Multiple repressors bind to distal sites and interact by looping the promoter DNA in a way to impede RNA polymerase binding the promoter elements. **c.** Repression by the modulation of an activator protein. The repressor binds to an activator on the promoter and prevents the activator from functioning by blocking promoter recognition by the α subunit of RNA polymerase holoenzyme. Figure adapted from [11]

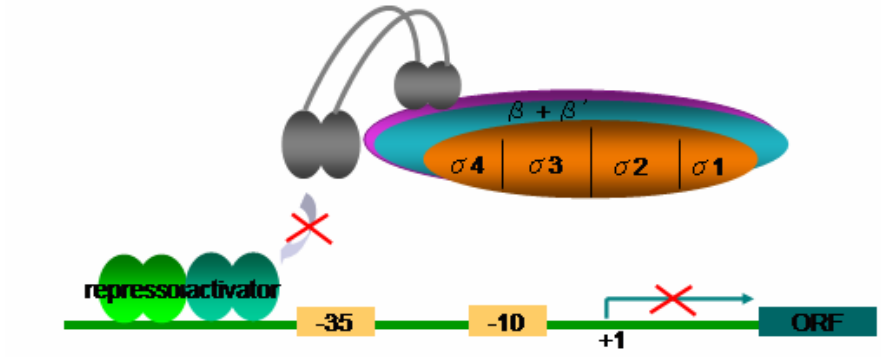
a. Repression by steric hindrance



b. Repression by looping



c. Repression by modulation of an activator



CHAPTER 2

MATERIALS AND METHODS

2.1 *Shewanella oneidensis* MR-1 culture growth and processing of soluble cell extract

S. oneidensis MR-1 cell extracts (from aerobic growth and from anaerobic growth) were obtained from Dr. Zhou's laboratory at the University of Oklahoma.

2.2 Identifying operons of interest based on microarray expression profiling analysis

DNA microarray expression profiling (transcriptomics) is a common technique used to determine changes in mRNA levels simultaneously for all the genes in a cell [10, 67, 68, 69]. We are particularly interested in identifying stimulons, groups of genes or operons that are differentially expressed in response to a given environmental stimulus [70, 71]. Usually, the cDNA microarray is used to compare wild type (WT) strain gene expression patterns under a stimulating condition with an expression patterns in an unperturbed control. In our specific case, we are comparing these patterns under anaerobic vs. aerobic growth. This DNA microarray-based expression profiles will provide important information about how specific genes or operons respond and adapt to environmental changes.

In prokaryotes, an operon consists of one or more adjacent genes that are transcribed to a single polycistronic RNA transcript, in combination with upstream the regulatory elements. For a two-condition microarray experiment, an operon can be identified as a cluster of adjacent genes that have short intergenic regions in length and are putatively transcribed in the same direction

(on the same strand)[72, 73], and that show the same tendency of up- or down-regulation determined by microarray.

In this project, the microarray expression profile data of *S. oneidensis* MR-1 grown under aerobic and anaerobic conditions were provided by Dr. Zhou's laboratory at the University of Oklahoma Based on that, we identified those operons whose expression is significantly up- or down- regulated (e.g. >2 fold) in aerobic growth condition compared to anaerobic growth condition and the DNA upstream of these operons (presumably containing the promoter for the operon) will be used for further experiments.

2.3 DNA affinity protein capture

Our DNA affinity protein capture approach is outlined in Fig 2.1 & 2.2. The DNA probes used for protein capture were PCR-amplified from *S. oneidensis* MR-1 genomic DNA using the primers listed in Table 2.1. The biotinylated ds DNA probe was immobilized on magnetic DynaBeads M-280 Streptavidin (Invitrogen, Carlsbad, California) using 2x B & W buffer (1 mM EDTA, 2.0 mM NaCl, 10 mM Tris, pH7.0). The bead-bound DNA was then mixed with 2.5 mg/mL *S. oneidensis* MR-1 soluble cell extract from cells grown either aerobically or anaerobically and incubated at 37 °C for 30 min with gentle shaking to keep the beads in suspension. Unbound proteins were eluted with three washes of incubation buffer (50 mM Tris, 1 mM EDTA, 100 mM KCl, 5% glycerol, 0.1% triton-X, 1 mM DTT, pH7.5). DNA bound proteins were eluted by incubation at 55 °C for 5 min with 1x Laemmli buffer containing no β-mercaptoethanol, as this tended to strip the streptavidin from the bead surface. Eluted proteins were analyzed by SDS-PAGE with silver-blue staining [74].

Table 2.1 Probes used in DNA affinity protein capture

Probe name	Genome coordinates	Forward primer[a] (biotinylated)	Reverse primer[a]	Probe length (bp)
SO0060	68131-68327	ggcaaatgtcataacggctaaga	gtttgactatctcctcacccaaaat	197
SO0275	280546-280324	tagccagcagttgagatgagttaca	ccgataatggcgatgtttttca	223
SO1427	1488019-1488219	ccttatggtgtttttgagaatg	cgccatctcatatttaatgcc	201
SO2093	2189271-2189469	catcttgtcattctccttgttctc	tgacgctctaaccaaataacc	199
SO3921	4069059-4068851	aagatgtgaccgtcaatgtcgc	cttgttcatttctgtttcccgt	209

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

2.4 In-gel tryptic digestion and peptide mass mapping

Bands of interest in SDS-PAGE lanes of eluted proteins from DNA affinity protein capture were excised and subjected to in-gel tryptic digestion. The gel slices were subjected to three cycles of hydration and dehydration: a 10-min incubation in 50 µL 25 mM ammonium bicarbonate followed by a 15-min incubation in 80 µL 50% acetonitrile in 25 mM ammonium bicarbonate. The dehydrated gel slices were then completely dried in a vacuum centrifuge (~10 min), after which they were rehydrated in a ~10-µL solution of 10 ng/µL trypsin for 30 min and kept on ice. 5-10 µL 25 mM ammonium bicarbonate was added to cover the gel surface and the solution was incubated overnight at 37 °C. Digested peptides were extracted with one ~10-µL wash of 25 mM ammonium bicarbonate followed by three successive ~10-µL washes with 75% acetonitrile, 0.5% trifluoroacetic acid (TFA). The pooled, extracted solution was concentrated to 4-5 µL by vacuum centrifugation, and 1 µL 5% TFA was added to make a final volume of 5-6 µL and a final TFA concentration of 0.1-1%. A NuTip C-18 (Glygen Corp., Columbia) was used to concentrate and purify the trypsin-digested samples prior to depositing on the MALDI target.

Peptide mass mapping was performed by the Chemical and Biological Sciences Mass Spectrometry Facility (University of Georgia, Athens, GA) on a Bruker Autoflex (TOF) mass spectrometer (Bruker Daltonics Inc., Billerica, MA). Proteins were identified by peptide mass fingerprinting using the MASCOT online search engine (*www.matrixscience.com*, [75]) to search the NCBI database of bacterial genomes and a local server hosting ProteinProspector [76] to search a *S. oneidensis* MR-1 genome database.

2.5 Sequence analysis of target protein for characterization

Sequence analysis of identified proteins was performed using NCBI BLAST [77,78] and Conserved Domain searches [79, 80, 81] against the NCBI non-redundant protein database.

2.6 Cloning, expression and purification of recombinant his-tagged target protein

The vectors used for cloning are either pET24dBAM, pET24dBAMTEV or pET24bMI-2 vector, which were kind gifts from the laboratory of Michael Adams, University of Georgia. The pET24dBAM vector harboring the clone was a derivative of pET24d, modified to incorporate an N-terminal hexahistidine tag (his-tag) on the expressed protein. pET24dBAMTEV is the standard pET24dBAM vector with a TEV protease site inserted just upstream of the BamHI cloning site and pET24bMI-2 is a modified version of pET24b (Novagen, Gibbstown, NJ) containing *Nde*I and *Bam*HI restriction sites and a C-terminal hexahistidine tag. All these vectors are selectable for kanamycin resistance and are designed for use in combination with a host containing a T7 lysogen under control of the *lac* promoter. The vectors also contain a copy of the *lac* repressor which represses expression of the endogenous T7 RNAP except in the presence of the chemical inducer IPTG (isopropyl-β-D-thiogalactopyranoside) which causes release of the

repressor from the *lac* promoter, thereby permitting expression of T7 RNAP. Recombinant protein expression is therefore inducible with IPTG and is driven from the T7 promoter by T7 RNAP under control of the *lac* promoter.

For *SO3363p* and *SO3988p* cloning into corresponding vector, first, their open reading frame regions were PCR amplified from the *S. oneidensis* MR-1 genome using a high-fidelity Pfu DNA polymerase (Stratagene, Kirkland, WA) and primers containing BamHI and NotI or restriction sites listed in Table 2.2 and the PCR products were purified using a PCR purification kit (Qiagen, Valencia, CA). Second, digestion reactions were set up with ~4 μg of vectors using the appropriate restriction enzymes (BamHI/NotI or BamHI/NdeI) and incubated at 37°C for 1-3 h followed by dephosphorylation by shrimp alkaline phosphatase and 1% agarose gel purification. Meanwhile, ~100 ng of each PCR-amplified cloning fragment was digested with NotI or NdeI overnight at 37 °C followed by a 3 h digestion with *BamHI* at 37 °C. Third, the ligation reactions were set up in 20 μL volumes using 100 ng of linearized, gel-purified pET vector with various molar ratios of vector to insert (3:1, 1:1, 1:3) together with buffer and 0.5 Weiss units of T4 DNA Ligase (Fermentas, Glen Burnie, MD). Ligation reactions were incubated at room temperature for 3 h or at 16 °C overnight and immediately transformed by heat-shock into $CaCl_2$-competent XL1-Blue cells (made from a stock obtained from Stratagene, Kirkland, WA). Transformation cultures were plated onto LB-agar plates containing 100 μg/mL kanamycin followed by incubation at 37°C overnight. Clones were selected for plasmid amplification in 5-mL liquid cultures, purified by Qiaquick Plasmid MiniPrep Kit (Qiagen, Valencia, CA) and clones with correct inserts were identified by restriction digestion followed by sequence verification by primer extension (Retrogen, San Diego, CA).

Table 2.2 ORF probe DNA for cloning into pET vectors

Probe name	Genome coordinates	Forward primer[a,b]	Reverse primer[a,b]	Probe length (bp)
SO3363_1	3506998-3506060	AAAGGATCCTTAGAACTATTA GAACCTATTGCG	AAAAGCGGCCGCCTAACG TTCACTGAAAAGCTC	939
SO3363_2	3506998-3506060	AAACATATGTTAGAACTATTA GAACCTATTGCG	AAAAGGATCCACGTTCACT GAAAAGCTCA	939
SO3988	4126001-4126717	AAAGGATCCCAAAATCCGCAC ATTCTGATCG	AAAAGCGGCCGCTTAGTC TTCTAAGTTACCGCAGA	717

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

[b] BamHI restriction site is colored in red, NotI restriction site is colored in blue, and NdeI restriction site is colored in orange.

For expression of the his-tagged recombinant protein, the selected clone was transformed into BL21-CodonPlus (DE3)-RIPL cells (Stratagene, Kirkland, WA) using the manufacturer's protocol. Protein expression from a 1-L culture of LB media was grown to an OD_{600} of ~0.5-0.8, and protein expression was induced with 0.1 mM IPTG and then grown overnight at 20 °C. Cells were harvested from the 1-L culture and resuspended in ~20 mL Binding Buffer (20 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole pH 8.0) containing ~20 µl 100 mM PMSF and 15.6 µl β-mercaptoethanol. Cells were sonicated on ice using a large horn at 20% power for 2 minutes, 10 s on and 10 s off, with capping and mixing of the solution between pulses. Soluble cell extract was obtained after centrifugation for 60 min at 21,000 rpm with a Beckman JA25.5 rotor. The supernatant was centrifuged at 21,000 rpm for an additional 15 min and filtered through 0.2 µm polyethersulfone membrane (Schleicher & Schuell) prior to purification of the protein by column chromatography.

Using an automated FPLC system (AKTA, GE Healthcare, Piscataway, NJ), the soluble cell extract was loaded onto a 1-mL HiTrap metal affinity column (GE Healthcare, Piscataway, NJ) preloaded with nickel-sulfate per the manufacturer's instructions. The column was washed with 5 mL Binding Buffer followed by a gradient elution with Eluting Buffer (20 mM sodium phosphate, 0.5 M NaCl, 0.5 M imidazole, pH 8.0), first with a 10-mL 0-20% gradient and then with a 25-mL 20- 80% gradient. Protein-containing fractions that were relatively pure were pooled, and a 5-mL desalting column (GE Healthcare, Piscataway, NJ) or 15ml Amicon Ultrafiltration cell (Millipore, Billerica, MA) was used for buffer exchange into 20 mM Tris, 100 mM NaCl, pH 8.0. Protein concentration was determined using a Protein Assay kit (Biorad, Hercules, CA), and aliquots of purified recombinant proteins were stored at -80 °C.

2.7 Eletrophoretic mobility shift assay (EMSA)

DNA probes for EMSA [82] were PCR-amplified from *S.oneidensis* MR-1 genomic DNA using primers listed in Table 2.2 followed by PCR-purification using a PCR Purification Kit (Qiagen, Valencia, CA). EMSA was performed by incubating DNA with various amounts of protein in 10-μL volumes in 1x EMSA buffer (50 mM Tris, 100 mM KCl, 5% glycerol, 1 mM EDTA, pH 8.0) using a 5x stock. DNA concentration was typically 50-100 nM per reaction, and protein was adjusted according to the molar amount of DNA. For reaction set up, a master mix of water, 5x EMSA buffer and DNA was made according to the number of reactions in the experiment, then distributed to 0.5-mL microcentrifuge tubes on ice. Protein dilutions were made in a final concentration of 1x EMSA buffer, and 2 μL of the appropriate protein dilution was added to each EMSA reaction (with 2 μL of 1x EMSA buffer added instead of protein for the DNA-only lane). In cases where an extra reagent was added to the reaction (e.g. DNA competitor

heparin, H_2O_2, etc), volumes of water and/or 5x EMSA buffer were adjusted accordingly such that the final buffer concentration of each reaction was always 1x.

Reactions were incubated at 37 °C for ~20 min and immediately loaded onto a 5% TBE gel (BioRad, Hercules, CA) that was typically run at 100V for 45-60 min. The gel was then stained with SYBR Green Nucleic Acid Gel Stain (Invitrogen, Carlsbad, California) or SYPRO Ruby protein gel stain (BioRad, Hercules, CA) according to the manufacturer's instructions to visualize the presence of DNA and protein. Gels were imaged via UV transillumination.

Table 2.3 DNA Probes used in EMSA

Probe name	Genome coordinates	Forward primer[a] (biotinylated)	Reverse primer[a]	Probe length (bp)
SO0060	68131-68327	ggcaaatgtcataacggctaaga	gtttgactatctcctcacccaaaat	197
SO0275	280546-280324	tagccagcagttgagatgagttaca	ccgataatggcgatgtttttca	223
SO1427	1488019-1488219	ccttatggtgtttttgagaatg	cgccatctcatatttaatgcc	201
SO2093	2189271-2189469	catcttgtcattctccttgttctc	tgacgctctaaccaaataacc	199
SO3921	4069059-4068851	aagatgtgaccgtcaatgtcgc	cttgttcatttctgtttcccgt	209

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

2.8 Identification of consensus DNA recognition sequence for *SO3363p*

SELEX (Systematic Evolution of Ligands by Exponential Enrichment) [83, 84] was applied to determine the consensus DNA recognition sequence of SO3363p (Fig 2.3). This method involves the use of an artificial library of DNA containing random sequences to allow for enrichment of the SO3363p DNA-binding sequence through successive cycles of selection with SO3363p. The single-stranded SELEX probe from which the artificial library was generated was designed with a 25-nt randomized region flanked by constant primer regions, each

containing an EcoRI restriction site. The double-stranded SELEX probe was PCR-amplified from the synthetic single-stranded oligonucleotide; the SELEX probe and primers are listed in Table 2.4. To create the dsDNA probe, 100 pmol of single-stranded SELEX probe was amplified with 2 nmol of each primer for a total of 5 PCR cycles. The PCR-amplified double-stranded SELEX probe was polyacrylamide gel-purified according to the crush-and-soak method [85]. Selection rounds were set up essentially as for the EMSA reactions, except for the amount of SELEX probe used and the protein-DNA ratios. 0.5 μM of SELEX probe was used in the first round of selection and 0.2μM was used in all the other succeeding selection rounds; the maximum protein-DNA ratios decreased from 40 in the first round to 4 in the final round. After each selection round, DNA was purified from shifted protein-DNA complexes, amplified with the SELEX primers using 10 cycles of PCR followed by 3% Nusieve 3:1 agarose (Cambrex Bio Science Rockland, Inc. South San Francisco, CA) gel purification before proceeding to the next selection round. A total of 8 selection rounds were performed in this manner. The selected DNA was digested with EcoRI (Promega, Madison, WI), concatemerized, and cloned into the pUC18 standard cloning vector (Fermentas, Glen Burnie, MD). Blue/white color screening and colony PCR were used to identify colonies that contained plasmids with different number of insertions and plasmid was isolated from these colonies for sequencing. A total of 9 sequences were obtained from the round 8 selected DNA, and a total of 22 sequences were obtained from round 7 selected DNA. These sequences were input into MEME online motif searching software [86] to generate a common motif among the selected DNA, and a graphical representation of the motif was created using WebLogo [87].

Table 2.4 DNA probe and primers used for SELEX

Name	Sequence[a]
SELEX single-stranded probe	gtaaaacgatgaattctgccagt(n)$_{25}$gacatagcgaatccacttagcag
SELEX forward primer	gtaaaacgatgaattctgccagt
SELEX reverse primer	ctgctaagtgaattcgctatgtc

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

2.9 Mutant construction by in-frame deletion

In-frame deletion of the SO3363 gene was accomplished by the method that has been developed and successfully utilized in *Shewanella oneidensis* MR-1 strain [88, 89]. Primers used for generating PCR products in mutant construction are listed in Table 2.5 and the schematic presentation of the method is shown in Figure 2.4. The deletion process was carried out as follows. In the first step (Fig 2.4A), primers SO3363-5o and SO3363-5i were used for the 5'-end fragment, and SO3363-3o and SO3363-3i were used for the 3'-end fragment. The outside primers (5o and 3o) harbor a *Sac*I restriction site and the inside primers (5i and 3i) contain complementary 20-nt tags at their respective 5' termini [90, 91, 92]. PCR amplification was performed in a 100-µL volume containing 1 µL easy A enzyme PCR cloning (Invitrogen, Carlsbad, CA), 10µl 10x PCR buffer (Sigma, St. Louis, MO), 2 µL 5o or 3o (50 pmol/µL), 2 µL 5i or 3i (10 pmol/µL), 25 mM MgCl2, 1 µL genomic DNA of *S. oneidensis* MR-1 as template (80 ng/µL), 2 µL 10 mM dNTP Mix (Roche, Florence, SC) and 82 µL ddH2O. The PCR mixture was denatured at 94 °C for 5min, followed by 30 cycles of 94 °C for 30 s, 55 °C for 1 min, and 72 °C for 1 min, followed by 7 min at 72 °C. The PCR products were then purified from agarose gel using QIAquick Gel Purification Kit (Qiagen Inc. Valencia, CA).

In the second step (Fig. 2.4B), the 5'- and 3'-end fragments were annealed at their overlapping region and amplified by PCR as a single fragment, using the outer primers (SO3363-5o and SO3363- 3o). The PCR amplification conditions were as described essentially as above except that the genomic DNA template was replaced by adding 1:1 mixture of 5o-5i and 3o-3i fragments at ~80 ng/uL each and 0.5 μL Hi-Fi taq polymerase was added in each 100 μL PCR reaction . The fusion product was purified from agarose gel using QIAquick Gel Purification Kit (Qiagen Inc. Valencia, CA), resuspended in 50 μL of 1x *Sac*I restriction buffer containing 5 U of *Sac*I restriction enzyme, and digested for 3-4 h at 37 °C followed by inactivation at 65 °C for 20 min. The digested fragment was gel-purified, ligated into *Sac*I-digested and phosphatase-treated suicide vector (pDS3.0), electroporated into *E. coli* (WM3064/λpir), spread on LB+ Am at 50 ug/mL + DAP (2,6-diaminopimelic acid, Sigma, St. Louis, MO) at 50 ug/mL agar plates and incubated at 37 °C overnight. Plasmids were extracted using Qiaquick Plasmid MiniPrep Kit (Qiagen, Valencia, CA) and digested with SacI to verify the insertion.

In the third step, the suicide plasmid construct was integrated into *S. oneidensis* MR-1 chromosome by homologous recombination (Fig 2.4B) [93]. This process was facilitated by conjugal transfer between *E. coli* WM3064/λpir cells harboring the suicide plasmid construct (donor) with strain MR-1. *E. coli* transformants and MR-1 cells were grown separately in LB medium with proper antibiotics overnight, washed in fresh medium, and mixed in a 1:1 ratio (donor: recipient) by spotting onto 0.2 μm Millipore membrane filter paper (Millipore,). Following 16-18 h incubation at 30 °C, the cells were removed from the filter paper, resuspended in LB medium, and plated onto LB agar supplemented with Gentamycin (Gm) (15μg/ml) and incubated at 30 °C for 24 h. Transfer colonies to both Gentamycin and NaCl-less LB+10% sucrose plates followed by incubation at 30 °C for 12-16 h and colonies that grew on Gentamycin plates

but not on sucrose plates were picked. Correct integration of the suicide vector was verified by PCR amplification. This was accomplished by comparing the sizes of the products amplified from the wild-type and mutant DNA using primers *SO3363*-SF/LR and *SO3363*-LF/SR; the colony with the correct insertion should produce DNA fragments with different sizes.

In the fourth step, to obtain the deletion mutation of *SO3363* gene, the integrated suicide plasmid was resolved from chromosome through recombination (Fig 2.4C). This was accomplished by growing the strain carrying the integrated suicide vector in NaCl-less LB liquid overnight, followed by 10x, 100x and 1000x dilutions of the culture, and plating 100 µL of each dilution onto LB agar containing 5% sucrose. The plates were then incubated at 30 °C for 16-18 h and colonies were transferred onto LB/Gm and LB/sucrose at the same time. Colonies grown only on LB/sucrose but not on LB/Gm were selected for further verification. Screening for deletion was accomplished by comparing the sizes of products amplified from wild type and mutant DNA using the SO3363-LF/LR primers.

Table 2.5 DNA primers used for SO3363 mutant construction

Primer name	Sequence[a,b]
SO3363-3o	gaagagctcgttaatcaatgg
SO3363-3i	gtgagctaatactcgatcagaattgacccatcatctaccg
SO3363-5o	ttgagctccaagttgaaactc
SO3363-5i	ctgatcgagtattagctcacagttctaacatgaaagtcgcc
SO3363-SF	ttgctgtgagatattaaggc
SO3363-SR	cgctaaatgcttacgacgtg
SO3363-LF	atgcaaaactccatcggtac
SO3363-LR	cgaactaagacacaaagctga

[a] DNA primers are listed from 5' to 3'. SacI sites used in cloning are colored red.

[b] Linkers used in crossover PCR are colored in blue.

2.10 Phenotype microarray

Phenotype microarrays (PM) provide a high-throughput technology for simultaneous testing of a large number of cellular phenotypes [94]. This was employed to determine the phenotypic differences between the MR-1 parental strain and the SO3363 mutant strain in this project.

At Biolog (Biolog, Inc., Hayward, California), they employed testing of cellular phenotypes using cell respiration as a reporter system [95, 96] and a tetrazolium violet dye is used to detect colorimetrically the respiration of cells by monitoring the color change of the dye due to the electron flow during cell respiration, which causes the reduction of the dye and a color change from colorless to purple. As part of this technology, the OmniLog instrument has been developed for the purpose of reading and recording the color change in PM assays. The instrument cycles 96-wells microplates in front of a color CCD camera to read 50 wells in as little as 5 min, providing quantitative and kinetic information about the response of cells.

PM assays were performed according to the manufacturer's protocol. First, both mutant and wild type cell suspensions were prepared by swabbing cells from agar plates to 16ml IF-0 (Biolog, Inc, Hayward, California) inoculating fluids to achieve 42% T (measured by Biolog Turbidimeter) cell suspension and then 15 mL of 42% T cell suspension were added into 75 mL of IF-0+dye solution to make the final cell density 85% T. For PM 1-2 microplates, 100 µL of 85% T cell suspensions above were added into each well; For PM 3-8 microplates, 100 µL of cell suspensions made by adding 680 µL sodium succinate/200 µM ferric citrate to 68 mL 85%

cell suspensions were added into each well; For PM 9-20 microplates, 100 uL of cell suspension made by transferring 600 μL of the 85%T cell suspension into 120 mL IF-10+dye solution, were added into each well. All PMs were incubated in the Omnilog instrument for 24-48 h at 37 °C.

The OmniLog reports a 96-box chart corresponding to the 96 wells in the PM for each strain and for each PM. The cellular response is recorded showing the amount of purple color in each well (vertical axis) throughout the time course (horizontal axis) of the assay. The *SO3363* mutant strain is recorded as a green tracing and the wild type MR-1 control strain is recorded as a red tracing. These tracings were overlaid to look for the differences between strains. For most phenotypes, the mutant and parental strain give virtually identical phenotypes reported as yellow color. If the mutant is respiring more slowly in one well, its corresponding box in the chart will be red, whereas if the mutant is respiring more rapidly the color will be green.

2.11 Bioscreening

Bioscreening is a technique used for physiological characterization of a mutant strain. Growths of the deletion strain under aerobic, anaerobic and hydrogen peroxide conditions were characterized by recording growth curves in triplicate with a Bioscreen C microbiology reader (Labsystems Oy, Helsinki, Finland) with wild type MR-1 as the control. For aerobic growth, exponential phase cultures were diluted to approximately ~1×10^5 cells/mL in fresh medium, and 400 μL was transferred to the honeycomb plate wells of the Bioscreen C reader. The cultures were shaken at medium intensity continuously for 24h, and the turbidity was measured every 30 min at 600 nm. For anaerobic growth, exponential phase cultures grown aerobically were centrifuged, purged in nitrogen and suspended in fresh medium containing 20 mM sodium lactate to approximately ~1×10^5 cells/mL in an anaerobic glove box. Electron acceptors tested

in this study included fumarate (10 mM), nitrate (10 mM), sulfate (1-5%), TMAO (10 mM), and DMSO (10 mM). For the hydrogen peroxide challenge, exponential phase cultures gown aerobically were diluted to approximately ~$1x10^5$ cells/mL in fresh medium and then varying amounts of H_2O_2 were added to produce the desired concentration (0.2-1.0 mM).

2.12 Transcriptomic Analysis of *SO3363* mutant strain

S. oneidensis MR-1 whole genome DNA microarray chips were designed and printed by Dr. Zhou's lab at the University of Oklahoma. We used these microarrays to determine the regulons affected by the deletion of *SO3363* gene. These arrays contain 95% PCR-amplicons that cover 95.6 % of the whole genome of 4921 ORFs; generally there were two replicates of each gene on a single slide. In addition to the duplicates of arrays on the same slide, two or three biological cell samples were obtained (biological replicates) and for each biological sample, two slides were used (technical replicates).

***S. oneidensis* MR-1 and *SO3363* mutant strain growth under different conditions.** In this part of experiment, different growth condition were tested for wild type MR-1 strain and *SO3363* mutant strain, which include aerobic growth, anaerobic growth under 20 mM sodium lactate and aerobic growth under 0.6 mM H_2O_2. For extraction of RNA under different conditions, *S. oneidensis* MR-1 and *SO3363* mutant cells were grown in LB medium at 30 °C overnight in a shaker (180 rpm). For aerobic growth experiments, *S. oneidensis* MR-1 and mutant cells were grown in LB medium at 30 °C overnight in a shaker (180 rpm), and 20 µL of the cells was inoculated into 10 mL of new LB medium (pH 7) and grown for ~5 h (at the mid-logarithmic phase with an optical density of 0.6-0.8). For anaerobic growth, 200 µL overnight grown cells under aerobic condition were inoculated into 45 mL LB medium containing 20 mM

lactate (which was purged with nitrogen and autoclaved) in glove box; cells were grown for about 8-10 h; For aerobic growth under 0.6 mM H_2O_2, 400 µL overnight grown cells under aerobic condition were inoculated into 20 mL fresh LB medium and then H_2O_2 was added. OD values were monitored till they reach 0.6-0.8. After growth, all cell types were harvested by centrifugation at 8000 x g for 5 min at room temperature. Cell pellets were immediately frozen in liquid nitrogen and stored at –80 °C until RNA was extracted.

RNA extraction, purification, and labeling [97, 98]. For RNA extraction, total cellular RNA of each strain was first isolated using TRIzol reagent (Invitrogen Life Technologies, California) and chloroform extraction according to the manufacturer's protocol. RNA samples were purified using a Mini RNeasy kit (Qiagen, Chatsworth, CA) and eluted into RNease-free H_2O. The concentration and purity of RNA samples were estimated with a Nanodrop spectrophotometer (Thermo Fisher Scientific, Miami, OK) using A_{260}/A_{230} and A_{260}/A_{280} ratios and good quality RNA will have an A_{260}/A_{280} ratio of 1.8 to 2 and an A_{260}/A_{230} of 1.8 or greater. For cDNA synthesis and labeling, total cellular RNA (10 µg) was incubated at 70 °C for 10 min in the presence of 10 µg of random primers (Invitrogen Life Technologies, California) in 16.5 µL volumes and kept on ice. A master mix of 6 µL 5x buffer, 3 µL 0.1 M DTT, 1.5 µL of 10 mM dA/G/CNTP's and 0.5 mM dTTP mixture, 1 µL Rnase inhibitor, 1 µL 1mM Cy3- or Cy5-dUTP (Amersham BioSciences, United Kingdom) was made according to the number of reactions in this experiment; 12.5 µL of this master mix was distributed into each 16.5 µL reaction; Then 1 µL Superscript II RNase H^- reverse transcriptase (Invitrogen, Invitrogen Life Technologies, California) was added and the reverse transcription reaction was allowed to proceed for 2 h at 42 °C followed by RNA hydrolysis in 1 N NaOH at 37 °C for 10 min. The labeled cDNA probe was purified immediately using a QIAquick PCR purification column

(Qiagen. Inc, Valencia, CA) and was concentrated in a Savant Speedvac centrifuge (Savant Instruments Inc, Holbrook, NY).

cDNA hybridization [97]. Hybridization of cDNA was performed according to the manufacturer's instruction manual (Corning Life Sciences, Lowell, MA), using two or three biological replicates, with each biological replicate having two technical replicates and each slide containing two gene replicates. Thus, for each gene there was at least a total of six data points. The whole procedure mainly involves pre-hybridization, hybridization and post-hybridization washes. Pre-hybridization was performed by adding 37-38 µL pre-hybridization solution containing 5x SSC (1x SSC is 0.15 M NaCl, 0.015 M sodium citrate), 40% formamide, 0.1% SDS, 0.1 mg/mL BSA onto each slide. The arrays were sealed in a waterproof hybridization chamber (Corning life Sciences, Lowell, MA) put in 45 °C water bath for 1 h. The arrays were then transferred to 0.1x SSC twice for 5 min, purified water once for 30 s and dried by duster. For the hybridization step, 400 ng of each labeled cDNA was resuspended in 40-45 µL of hybridization solution that contained 5x SSC, 40% formamide, 0.1% SDS, and 1% herring sperm DNA (Invitrogen, California). The cDNA suspension was incubated at 98 °C for 5 min, centrifuged to collect condensation, and kept at 50-60 °C. 37-38 µL of the sample was immediately applied onto a microarray slide, and hybridization was carried out in a hybridization chamber (Corning Life Science, Lowell, MA) submerged in a 45 °C water bath in the dark for 16 h. After hybridization, the coverslips were immediately removed, and the slides were washed in a buffer containing 2x SSC and 0.1% SDS for 5 min at 42 °C. The microarrays were washed in a new buffer with 0.1x SSC and 0.1% SDS for another 5 min at room temperature twice, with 0.1x SSC at room temperature for 1 min four times, 0.01x SSC for 10 s. Finally, the microarrays were washed with distilled water for 30 s at room temperature and dried with compressed air.

Microarrays were scanned using the ScanArray 5000 microarray analysis system (Packard BioChip Technologies, Massachusetts). Normally, 95-100% laser power and 70-80% photomultiplier tube efficiency were selected for scanning and scanned TIFF images were analyzed using the software ImaGene 5.5 (Biodiscovery Inc, California).

2.13 Analytical gel filtration to determine *SO3363p* quaternary structure

Analytical gel filtration using a Superdex 75 10/300 GL size exclusion column (GE Healthcare) was performed to determine N-terminal and C-terminal his-tagged SO3363 protein (with a monomer MW of ~34 kDa) quaternary structure. A mixture of molecular weight standards (Biorad,) containing thyroglobulin (670 kDa), globulin (158 kDa), albumin (66 kDa), myoglobin (17 kDa), vitamin B12 (1.35 kDa) was dissolved in 0.5 mL H_2O or in 0.5 mL buffer containing 20 mM sodium phosphate, 500 mM NaCl, 250 mM imidazole and another standard, ovalbumin (44 kDa) was prepared separately. The running buffer used was 20 mM sodium phosphate, 200 mM NaCl, 250 mM imidazole, pH 8.0. The mixture of molecular weight standards (200 µL of each) and ovalbumin were run through the column separately and the elution volume (V_e) of each was noted. Thyroglobulin was used to determine the column void volume (V_o), and a standard curve of molecular weight versus V_e/V_o was used to determine the corresponding approximate molecular weights of the sample peaks.

Figure 2.1 Design and immobilization of probe DNA. The probe DNA is designed to have approximately 200 bp DNA sequence upstream from the target ORF start. Probe DNA is amplified from genomic DNA using one biotinylated primer and one unlabeled primer such that the PCR-amplified probe contains a biotin group on one 5' end so that the DNA can be bound to streptavidin-coated magnetic beads (Dynabeads M-280 Streptavidin, Invitrogen,). The magnetic properties of the beads allow them, and correspondingly whatever is attached to them, to be easily separated from solution with the use of a magnet.

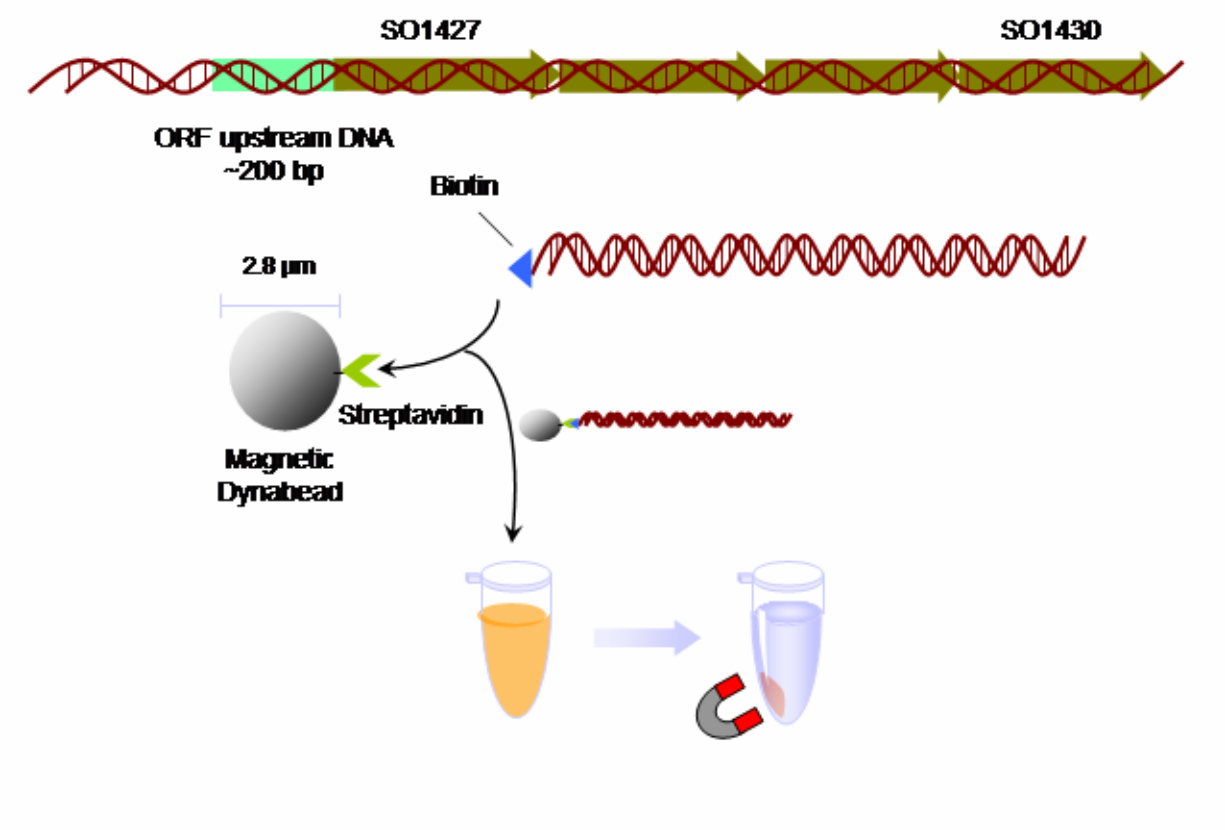

SO1427
SO1430
ORF upstream DNA
~200 bp
Biotin
2.8 µm
Streptavidin
Magnetic
Dynabead

Figure 2.2 The DNA affinity protein capture experiment. Biotinylated DNA is bound to magnetic streptavidin-coated beads. The bead-DNA complex is then incubated with soluble cell extract, and some proteins associate with the DNA including basal transcriptional machinery (RNAP subunits, etc), nonspecific DNA-binding proteins, and other transcription factors. Proteins which do not bind DNA are removed, and finally the DNA-binding proteins which remain are eluted and analyzed. Figure adapted from [99].

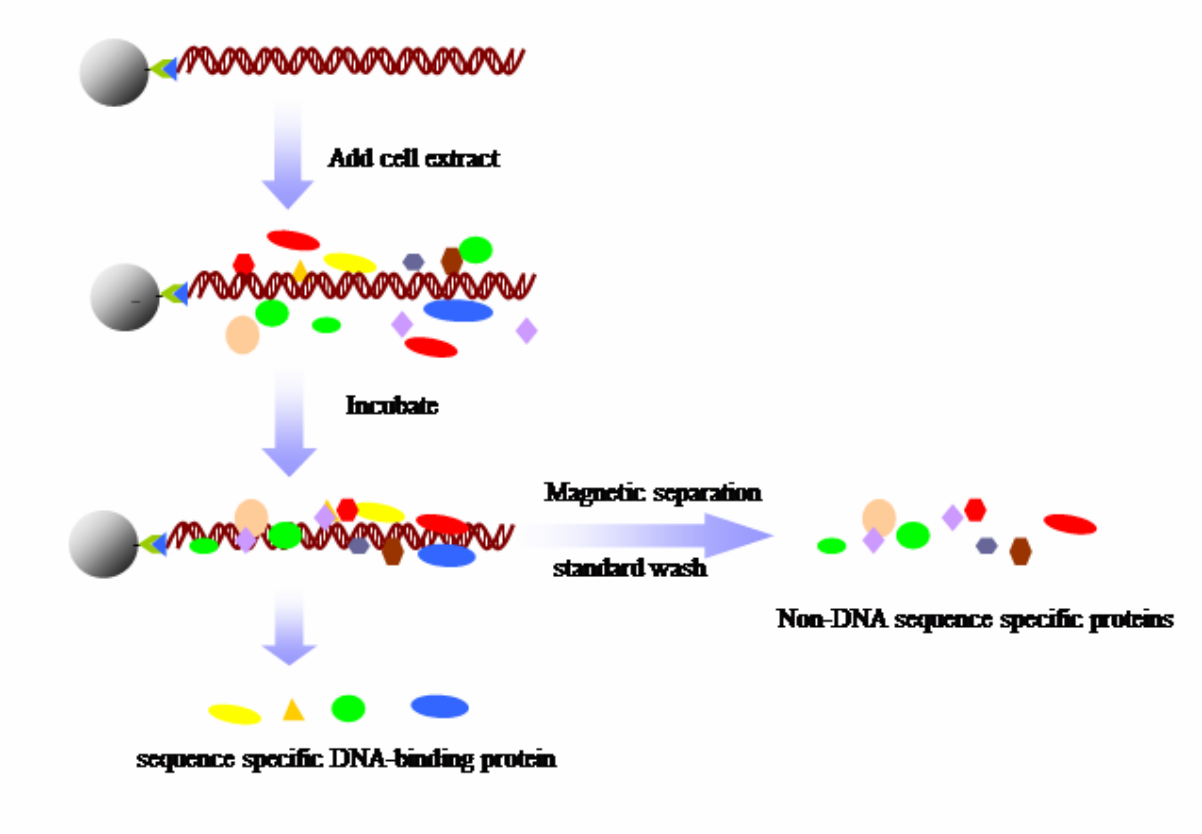
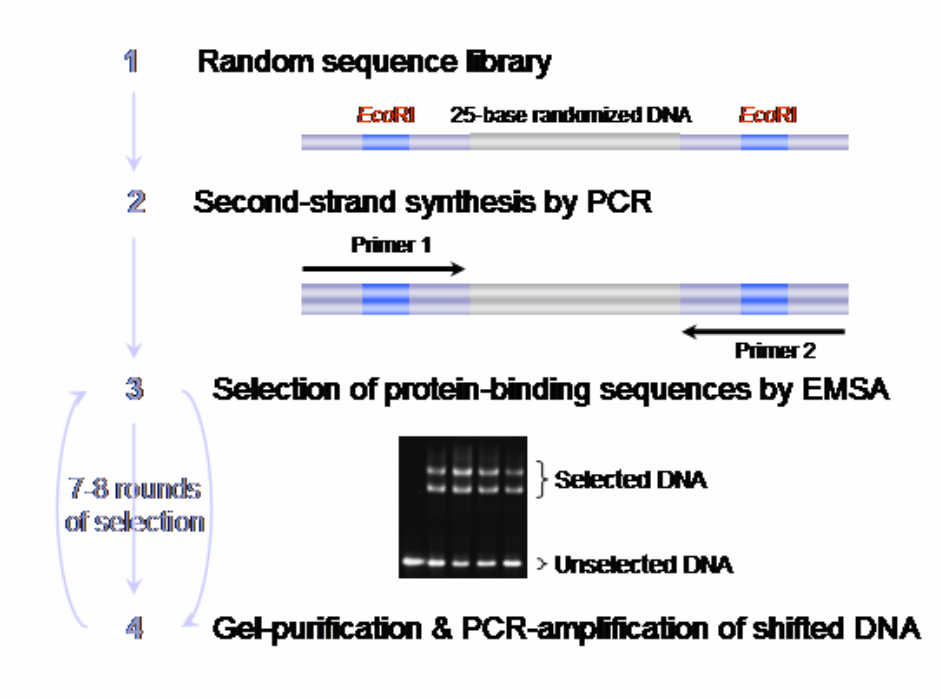

Add cell extract
Incubate
Magnetic separation
standard wash
Non-DNA sequence specific proteins
sequence specific DNA-binding protein

Figure 2.3 SELEX scheme. (1) The SELEX library was made up of single-stranded oligonucleotides designed to contain 25 bases of random DNA flanked with two ~20-base primer sites each containing an *Eco*RI restriction site. **(2)** Second-strand synthesis and amplification of the library was performed using primers complementary to the two priming sites. **(3)** SELEX probes with sequences that have higher-affinity to SO3363p were selected from the library pool using EMSA. **(4)** Shifted DNA was gel-purified and PCR-amplified for an additional selection round. Seven rounds of selection via EMSA, gel-purification, and PCR-amplification were carried out. **(5)** The Resulting selected DNA was digested with *Eco*RI and **(6)** concatemerized for cloning into pUC18. Cloned plasmids were transformed into XL1-Blue cells, and insert-containing plasmids were identified via blue/white colony screening. **(7)** Clones were then sequenced to identify the binding motif. Figure adapted from [100].


1
Random sequence library
EcoRI
25-base randomized DNA
EcoRI
2
Second-strand synthesis by PCR
Primer 1
Primer 2
3
Selection of protein-binding sequences by EMSA
7-8 rounds of selection
Selected DNA
Unselected DNA
4
Gel-purification & PCR-amplification of shifted DNA

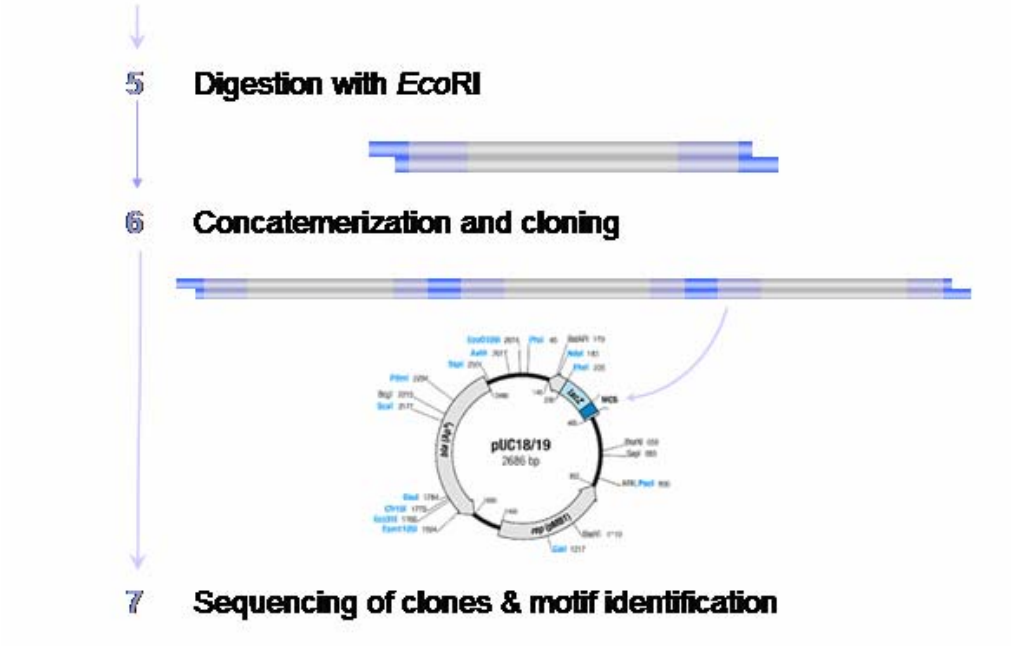
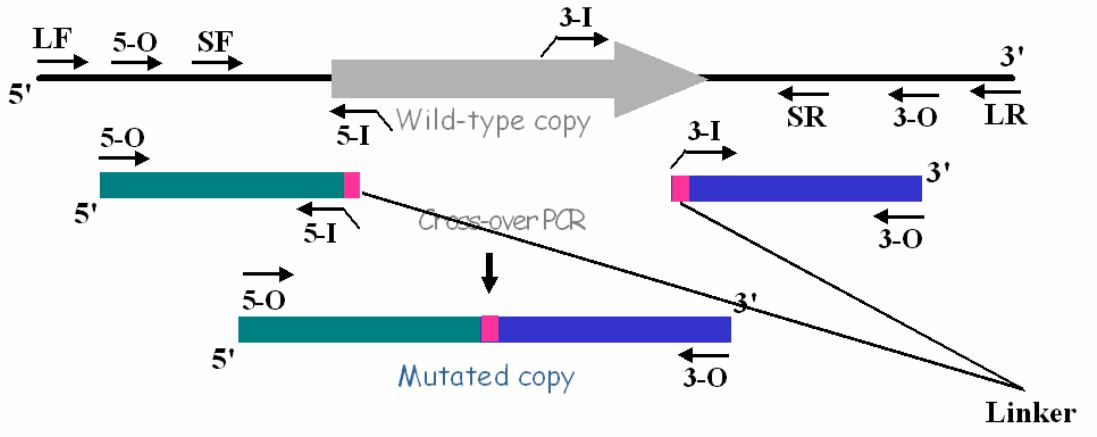
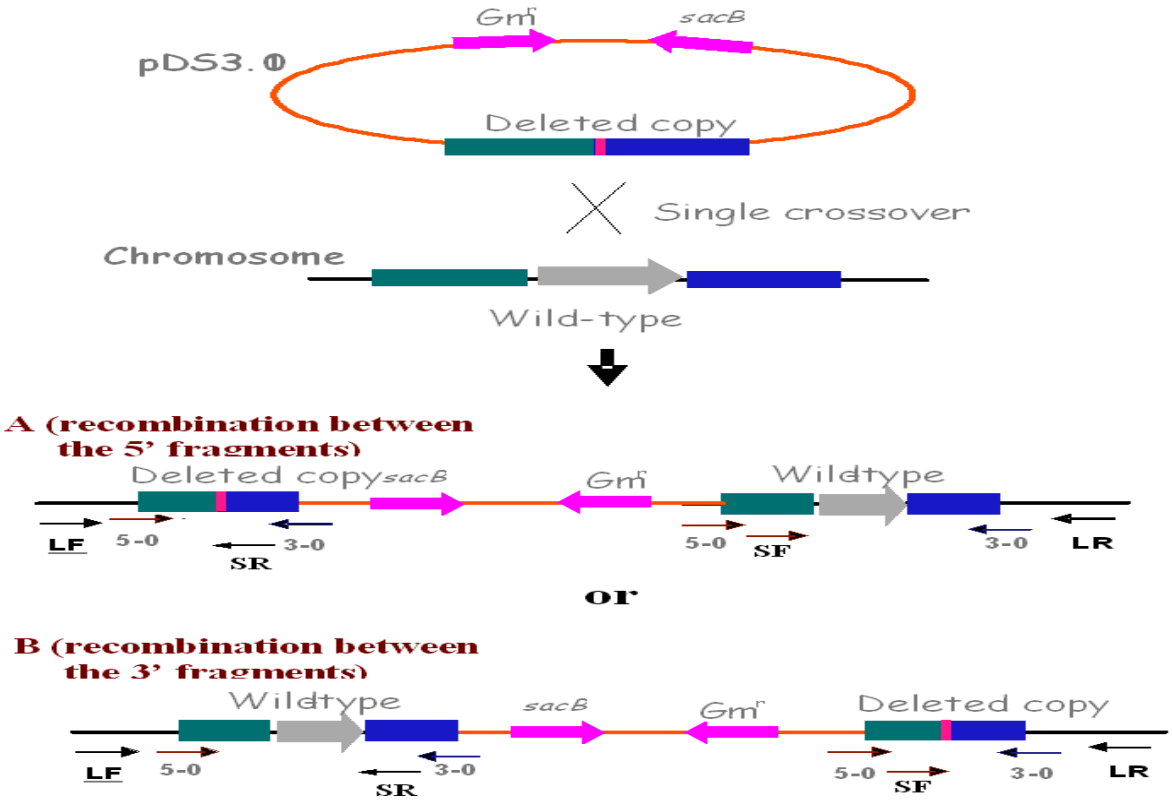

5
Digestion with EcoRI
6
Concatemerization and cloning
pUC18/19
2686 bp
MCS
7
Sequencing of clones & motif identification

Figure 2.4 Schematic presentation of generation of in-frame deletion mutant of SO3363 gene. **A.** Generation of a deletion construct using asymmetric crossover PCR; **B.** Integration of the suicide plasmid during a single crossover event; **C.** Generation of in-frame deletion during the excision of the suicide plasmid. Figure adapted from [88].

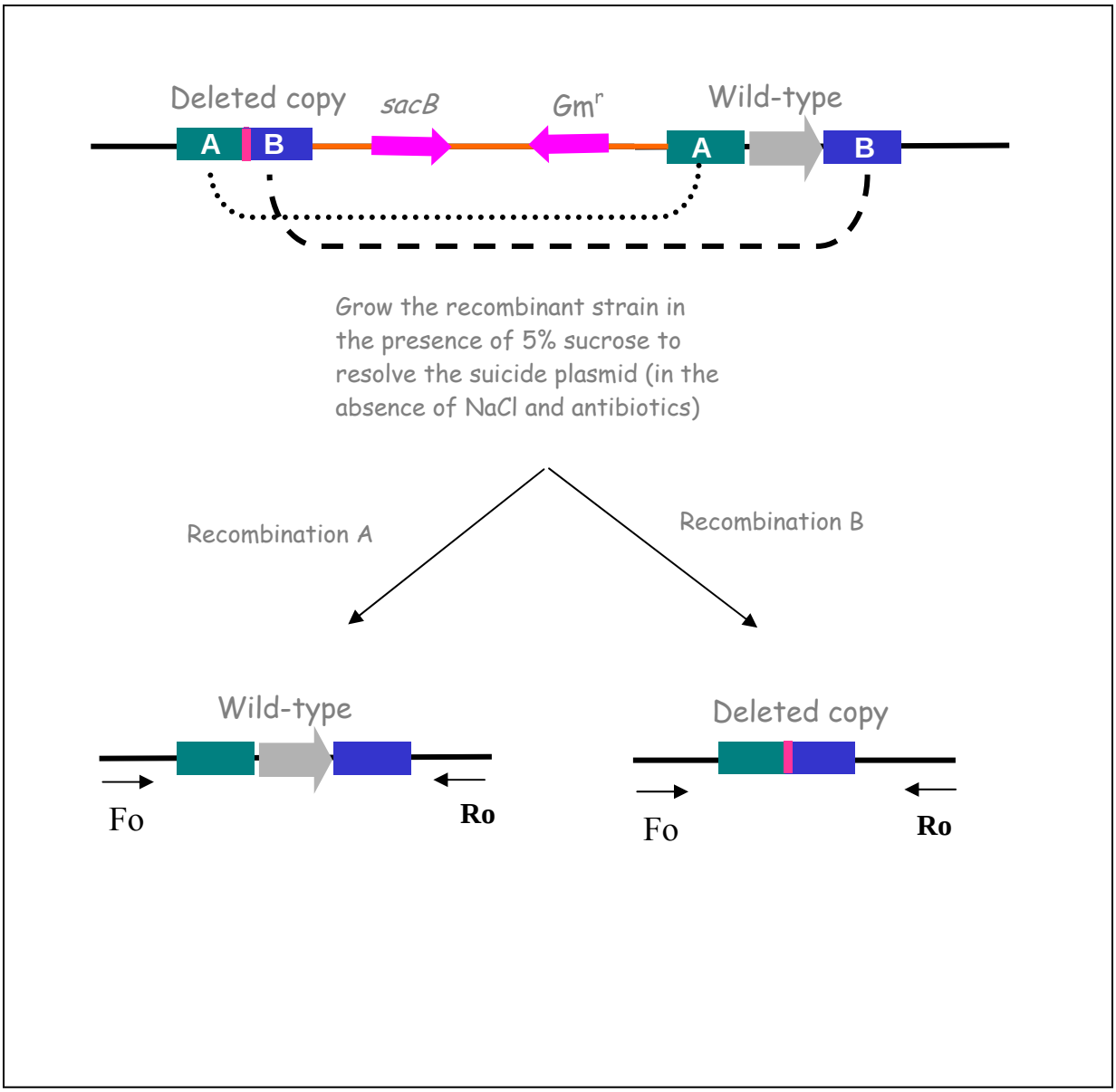
A.


LF
5-O
SF
3-I
5'
3'
5-I
Wild-type copy
SR
3-O
LR
5-O
3-I
5'
3'
5-I
Cross-over PCR
3-O
5-O
3'
5'
Mutated copy
3-O
Linker

B.


Gmr
sacB
pDS3.0
Deleted copy
Single crossover
Chromosome
Wild-type
A (recombination between the 5' fragments)
Deleted copy
sacB
Gmr
Wildtype
LF
5-0
SR
3-0
5-0
SF
3-0
LR
or
B (recombination between the 3' fragments)
Wildtype
sacB
Gmr
Deleted copy
LF
5-0
SR
3-0
5-0
SF
3-0
LR

C.


Deleted copy
sacB
Gm^r
Wild-type
A
B
A
B
Grow the recombinant strain in the presence of 5% sucrose to resolve the suicide plasmid (in the absence of NaCl and antibiotics)
Recombination A
Recombination B
Wild-type
Fo
Ro
Deleted copy
Fo
Ro

CHAPTER 3

TRANSCRIPTION FACTOR DISCOVERY AND INITIAL CHARACTERIZATION

3.1 Targeted transcription factor discovery by DNA affinity protein capture

3.1.1 Selection of target ORFs and operons for transcription factor discovery

As discussed in Chapter 2, *S. oneidensis* MR-1 whole-genome microarray expression profiling experiments were used to test the comparative response of *S.oneidensis* MR-1 under aerobic (control) and anaerobic (experiment) growth conditions by identifying genes with significant up and down-regulation. These criterions were used for the selection of target ORFs for further DNA affinity protein capture to identify transcription factors that associate with particular promoters under specific growth conditions.

Table 3.1 summarizes the expression profile obtained for a subset of *S. oneidensis* MR-1 genes comparing cultures grown either aerobically or anaerobically. A total of 303 ORFs displayed greater than 2-fold up-regulation with the anaerobic growth condition. 34 ORFs showed more than 5-fold up-regulation and 9 of them encode peroxidase, catalase, Ni/Fe hydrogenase, formate dehydrogenase and reductase which are known to be involved in redox responses in cells [101-104]. The most up-regulated ORF was *SO0957* encoding hypothetical protein and ORF *SO0956* and ORF *SO0958*, which encode alkyl hydroperoxide reductase F and C subunit, are in the same operon as *SO0957* and were also up-regulated more than 5-fold. Thus, this operon was selected as a target for transcription factor discovery. Operons *SO0057-60*, *SO0725*, *SO1427-30*, *SO2178* and *SO4512-15* were also selected as targets either because they

displayed dramatic up-regulation or these ORFs were predicted to play a significant role in anaerobic respiration or oxidative stress based on the annotation in *S. oneidensis* MR-1[105].

There were 562 ORFs that were down-regulated more than 2-fold and 199 of them were down-regulated more than 5-fold. Moreover, most of the 199 ORFs were annotated as hypothetical proteins and 68 of them were down-regulated 100-fold under anaerobic growth condition. I did not choose any down-regulated ORFs for transcription factor discovery except the one containing *SO2089-94* encoding hydrogenase (*hypA-F*) which was recommended by our collaborator, Jingzhong Zhou at the University of Oklahoma even though *hypA-F* ORFs did not display dramatic up- or down-regulation in the microarray experiment. However, it was reported in *E. coli* that *hypA-F* are involved in maturation of all three Ni/Fe hydrogenases and they are regulated by FNR anaerobic transcription factor [106].

Table 3.1 DNA microarray expression profiles of *S. oneidensis* MR-1 for growth +/–O_2

ORF[a]	ORF description / operon function[b]	Fold change from – O_2 to +O_2
Up-regulated >5-fold[c]		
SO0060	sensor histidine kinase	4.0
SO0403	hypothetical protein	6.1
SO0471	conserved hypothetical protein	16.7
SO0724	hypothetical protein	5.4
SO0725	catalase/peroxidase HPI	10.4
SO0827	L-lactate permease	8.3
SO0858	sodium:alanine symporter family protein	15.5
SO0901	hypothetical protein	8.0
SO0958	alkyl hydroperoxide reductase, C subunit	10.0

SO1070	catalase	65.0
SO1158	Dps family protein	5.4
SO1235	hypothetical protein	17.2
SO1427	decaheme cytochrome c	8.8
SO1532	hypothetical protein	6.7
SO1580	TonB-dependent heme receptor	6.0
SO1835	conserved hypothetical protein	23.5
SO1984	hypothetical protein	5.6
SO2099	quinone-reactive Ni/Fe hydrogenase, small subunit precursor	3.2
SO2178	cytochrome c551 peroxidase	69.2
SO2299	threonyl-tRNA synthetase	1.1
SO2716	hypothetical protein	8.9
SO2870	hypothetical protein	6.6
SO3565	2',3'-cyclic-nucleotide 2'-phosphodiesterase	6.7
SO3704	hypothetical protein	8.2
SO3734	hypothetical protein	8.4
SO3785	hypothetical protein	26.0
SO4512	conserved hypothetical protein	11.8
SO4630	hypothetical protein	5.1

[a] For ORFs which are at the beginning of operons, only the first ORF in the operon is listed with its corresponding fold change.

[b] ORF descriptions and operon functions are essentially derived from TIGR database.

[c] One of the genes within the listed operon displayed up regulation more than 5-fold under anaerobic growth condition while the fist gene of the operon may not be up-regulated more than five fold.

3.1.2 Design of the DNA probes used in protein capture

Most bacterial transcription factor binding sites should lie in close proximity upstream of the ORF translation start site; therefore, we selected a probe designed covering 200 bp upstream, relative to the translation start site of the first ORF in a target operon. Such a probe should contain the promoter for that operon as well as potential transcription factor binding sites (operators).

3.1.3 Protein capture on SO0060 and SO1427 probe DNA

The DNA affinity protein capture experiment relies on the differences in the population of proteins captured from the cell extracts of the two growth conditions (anaerobic vs. aerobic) using promoters designed above. The premise of this experiment is that the factor responsible for differential expression observed in DNA microarray data is either differentially present or shows differential DNA-binding affinity between the two growth conditions being compared. Also, it is assumed to be able to bind probe DNA containing some specific sequence (its operator, which should be present near the promoter). In either case, there should be an observable difference in the proteins captured by DNA affinity either from two different cell extracts using the same probe DNA or from different probe DNAs using the same cell extract.

The *SO0060* and *SO1427* probe DNA were bound to magnetic beads and incubated in cell extracts from cultures grown in the presence and absence of oxygen at 37 °C for 30 min. Eluted proteins were separated using SDS-PAGE and visualized by silver-blue staining. The resulting gel image can be seen in Figure 3.1. Many of the protein bands were the same between the two cell extracts and even among different DNA probes. Those that could be identified fell into four general categories, based on their identification: translation elongation factor and ribosomal proteins (bands 3, 4, 6, 9, 10) involved in protein synthesis, acetyl-coA carboxylase

(band 1) and ATP-dependent RNA helicase SrmB (band 2) and two regulatory proteins (bands 5 and 7). See Table 3.2 for protein band identification results.

Table 3.2 Identification of protein bands using peptide mass fingerprinting (cf. Fig. 3.2)

Band Number[a]	Locus	Annotation[b]	Molecular weight (kDa)	Score[c]	Sequence coverage
1	SO1894	Acetyl-coA carboxylase	75.6	4.4E+13	64%
2	SO0947	ATP-dependent RNA helicase SrmB	46.8	1.7E+05	51%
3	SO0217	Translation elongation factor Tu	43.3	2.1E+06	51%
4	SO0229	Translation elongation factor Tu	43.2	3.9E+09	58%
5	SO3363	Transcriptional regulator LysR family	34.7	4.0E+03	37%
6	SO2270	Ribosomal protein S6 modification protein	32.1	2.2E+06	62%
7	SO3988	Aerobic respiration control protein arcA	27.2	1.0E+04	55%
8	No match				
9	SO0255	Ribosomal protein S4	23.4	2.3E+03	69%
10	SO0248	Ribosomal protein S5	17.7	850	44

[a] Numbers correspond to numbered bands in Figure 3.2.

[b] Protein annotations are derived from the TIGR database.

[c] Score listed is from the MSFIT search against *S. oneidensis* MR-1 database.

Among these proteins, two of them were identified as translation elongation factor and three of them were identified as ribosomal proteins or ribosomal modification protein, which are all involved in protein/amino acid biosynthesis [107]. Acetyl-coA carboxylase (band 1), using its biotin prosthetic group as a carrier, transferring CO_2 from bicarbonate to the acetyl group [108], probably involved in the synthesis of fatty acids, was captured at a higher level in aerobically grown cell extract. ATP-dependent RNA helicase SrmB (band 2) is involved in ribosomal protein synthesis [109] and was captured more in anaerobically grown cell extract. Sequence analysis of the remaining two proteins (bands 5 and 7) indicated that they were predicted transcriptional regulators: protein products of *SO3363* and *SO3988*. These proteins were apparent in both cell extracts for both DNA probes but in different amounts. Analysis of the *SO3363* protein sequence revealed homology with three conserved protein domains that are transcription-factor related (Fig. 3.2). The closest match was to a LysR family of transcriptional regulators comprised of 9 proteins from proteobacterial species and this protein also has the conserved N-terminal HTH DNA binding domain (residues 5-63) and C-terminal substrate binding domain (residues 89-294). A BLAST search of the sequence resulted in approximately 90 hits having an e-value less than 1E-51, with the majority of high-scoring hits falling within other Shewanella species as transcriptional regulators and with high identities. *SO3988* protein sequence analysis (Fig. 3.3) reveals that it has a conserved N-terminal signal receiver domain, a C-terminal effector domain containing an HTH DNA binding motif; this protein is an OmpR type response regulator [110] which is one component of the two-component ArcA/ArcB system involved in signal transduction and transcription in *E. coli*. *SO3988* protein is being studied by our collaborator at the University at Oklahoma and it is identified as a global regulator, which regulates expression of over 50 operons [47]. It is functional in the oxygen response. From the

differential representation in the DNA affinity protein capture assay and sequence analysis, SO3363 was chosen for further research studies.

3.2 Validation of *SO3363p* as a DNA-binding protein

3.2.1 Expression and purification of recombinant his-tagged SO3363p

To verify that *SO3363p* (locus number followed by ´p´ refers to the protein product of that ORF) was a sequence specific DNA-binding protein, the protein was overexpressed as a his-tagged recombinant protein in *E. coli* for *in vitro* studies. The recombinant protein was produced using a protein-expression vector that allowed for the expression of *SO3363p* with an N-terminal or C-terminal his-tag which facilitates protein purification. The *SO3363p* sequence is 312 amino acids in length and has a calculated molecular weight of 34753 Da.

The protein was expressed in an *E. coli* BL21-CodonPlus (DE3)-RIPL strain compatible with the host expression vector, yielding an inducible system for protein production. Expression of *SO3363p* at 37 °C did not appear to be toxic to the cells and did not significantly slow culture growth. However, the protein was expressed predominantly in an insoluble though abundant form and thus lower temperature (20 °C) was used for the culture growth after the culture was induced to slow the formation of the protein.

The his-tagged protein was purified from cell extract using a nickel-affinity column, and a gradient of imidazole was used to elute the protein from the column. The fractions with relatively high purity were collected for further buffer exchange to an appropriate buffer system for protein storage. Two gels showing protein purified in this way are shown in Figure 3.4.

3.2.2 SO3363p is not a dimer

According to literature papers, LysR family proteins are either a dimer or tetramer in structure [41]. Thus analytical gel filtration was used to determine the oligomeric state of N-terminal his-tagged *SO3363p* as well as C-terminal his-tagged *SO3363*. Approximately 2 mg of protein sample was loaded onto an analytical gel filtration column with a 70 kDa exclusion limit. The approximate corresponding molecular weight was calculated for each eluted peak, using a calibration curve generated from a mixture of standards run through the column with the same buffer. The resulting traces for each sample can be seen in Figure 3.5, and the corresponding calculated molecular weights are shown in Table 3.3. Each of the *SO3363* samples displayed a small peak that eluted at the column void volume, as determined by thyroglobulin.

The calculated molecular weights of N-terminal and C-terminal his-tagged *SO3363p* were around 116 kDa which indicated that this protein is not a dimer in solution since the molecular weight of the monomer is about 34.7 KDa. So, it is possible that this protein is a tetramer in solution, which needs to be further determined by S-200 gel filtration column which has much higher molecular weight exclusion limit than S-75.

Table 3.3 Molecular weight of SO3363 samples as determined by S-75 gel filtration column

Standards & Samples	Elution volume V_e (ml)	V_e/V_o	Molecular weight (g/mol)[a]
Thyroglobulin	7.68		670,000
Globulin	8.33	1.08	158,000
Albumin	9.34	1.22	66,000
Ovalbumin	10.23	1.33	44,000
Myoglobin	12.11	1.58	17,000

Vitamin B12	17.65	2.30	1,350
His_6-SO3363	8.47	1.10	116,376
SO3363-His_6	8.50	1.11	114,646

[a] Molecular weight of each peak was calculated from the ratio of peak elution volume to column void volume using a calibration curve generated from known molecular weight standards.

3.2.3 His-tagged SO3363p shows low binding affinity to SO0060, SO1427and SO3921 promoter DNA

In order to show that *SO3363p* binds to the DNA probes from which it was identified in the DNA affinity protein capture experiment, EMSA experiments were performed using the same DNA probes to test for binding. *SO3363p* was found to shift both the *SO0060* and *SO1427* DNA probes completely at a protein/DNA mole ratio of around 40 and *SO1427* seems to have higher binding affinity to SO3363*p* since it was almost completely shifted even at a ratio of 20 (Fig. 3.6). The *SO3921* DNA probe has the weakest binding affinity to *SO3363p* since it can not be completely shifted even at a protein-DNA ratio of 40 (Fig. 3.6). Also, in the *SO3363p* sequence, there are three cysteines present that could form disulfide bonds that have an effect on protein-DNA interaction [111]. Furthermore, it was predicted that the binding site on *SO1427* promoter could be around 270 bp upstream relative to the translation start site [112]. Thus, EMSA experiments were also performed using *SO1427* containing ~300bp upstream promoter DNA with *SO3363p* under the presence of oxidizing reagents and reducing reagents. The oxidizing reagent used in EMSA experiments was 10 mM H_2O_2 or 10 mM diamide and the reducing reagent used was 10 mM DTT or 10 mM β-mercaptoethanol. From Figure 3.7, the DNA was almost completely shifted at a protein-DNA ratio of 32 under all conditions and there

was only subtle difference between EMSA in the presence of either the oxidizing or reducing reagent. The probe DNA seems to have better binding affinity toward the *SO3363p* under reducing reagent DTT or β-mercaptoethanol and there were two distinct bands of protein-DNA complexes at low protein-DNA mole ratios (around 4, 8).

3.3 Consensus DNA sequence GAACn$_4$GTT was discovered by SELEX

Promoter DNAs chosen for the EMSA experiments do not have very strong binding affinity to the *SO3363p* and there is no obvious way to choose promoters that would have much higher affinity than those used above. A better way to solve this problem would be to determine and define the DNA sequence more favored by *SO3363p* recognition using the artificial selection method SELEX. Starting with a pool of synthetic DNA containing a central 25-bp randomized sequence (see Table 2.4). EMSA with *SO3363p* was used to select sequences bound specifically by the protein. Each pool of selected sequences was PCR-amplified and purified to be used in EMSA again for further selection. After the final selection round was complete, the resulting DNA was cloned and sequenced. After seven rounds of selection, the consensus DNA sequence GAACn$_4$GTT (see Fig. 3.8A) was uncovered in all 31 sequenced SELEX DNAs; 18 of them showed a perfect match. With a different parameter (zero or one per sequence) for the MEME motif search, the consensus DNA sequence can be extended containing two possible overlapping palindromes: GTTCn$_8$GAAC and GAACn$_4$GTTC (see Fig. 3.8B). The second palindrome is much stronger than the first and it is included in 22 out of 27 of the DNA sequences (see Fig. 3.8C).

3.4 Verification of the importance of GTTCn_8GAACn_4GTTC for *SO3363p* binding

3.4.1 SO336p binds specifically to DNA containing GTTCn_8GAACn_4GTTC motif

To further confirm that the presence of palindromes are critical for *SO3363p* specifically interacting with DNA, a 71-bp control DNA having the same length with the synthetic DNA (referring to selected DNA) containing the binding motif (the 25 bp underlined DNA sequence in selected DNA in Table 3.4 is the SELEX result from one of MEME motif searches), was designed by mutating the palindromes but keeping the same GC content. The sequences of the two 71-bp DNA probes are listed in Table 3.4.

Table 3.4 71-bp DNA probes

Name	Sequence[a]
Selected DNA[b]	gtaaaacgatgaattctgccagttgttcggtggagcgaacgatcgttcgacatagcgaattcacttagcag
Control DNA[c]	gtaaaacgatgaattctgccagttttttcaggtggagcccgtgatcttgggacatagcgaattcacttagcag

[a] DNA Sequences are listed from 5′ to 3′.

[b] EcoRI restriction site is colored in red. The consensus DNA sequence is underlined and the palindromes are colored in green. The sequence other than the consensus DNA sequence in selected DNA is from the original SELEX single strand template design.

[c] The mutated palindromes are colored in purple.

The second strand DNA was PCR amplified and purified as we described in Materials and Methods and the same EMSA experiment was performed as we did in the SELEX experiment except that an excess amount of heparin was used as DNA competitor to distinguish specific protein-DNA binding from the nonspecific protein-DNA binding. Heparin can be used

as a DNA competitor in EMSA is because it binds protein through electrostatic associations which is relatively weak binding force and it won't be visualized under UV after gel staining [113]. Thus, heparin can challenge off the weak binding by nonspecifically bound DNA (Figure 3.9).

For selected DNA probes containing the palindromes, protein-DNA complexes with lower mobility can be clearly seen at a protein-DNA mole ratio of 4 and free DNA probe is completely shifted around ratio of 20. Furthermore, the protein-DNA complexes can still be seen when 50 μg/mL of heparin was added in the experiment, which indicated the specific binding between the protein and selected DNA probe. For the control DNA probe without the palindromes, a faint band of protein-DNA complex can also be seen at a protein-DNA mole ratio of 20, however the shift disappeared when only 5 μg/mL heparin was added, suggesting that the binding between the protein and the control DNA was nonspecific. Therefore, we can reach the conclusion that the presence of the SELEX-identified palindromes is very important for DNA specifically binding *SO3363p.* Thus, *SO3363p* is a sequence-specific binding protein.

3.4.2 Search for additional SO3363p binding sites in S. oneidensis MR-1 genome

The SELEX-discovered abbreviated motifs $AACn_4GTT$ and $GAACn_4GTT$ were used to search for other potential binding sites in the DNA regions upstream of all genes within the whole *S. oneidensis* MR-1 genome. We refer to these DNA regions as UORs (Upstream of ORF Regions) and a UOR database of *S. oneidensis* MR-1 was created by Darin Cowart, a Scott group member. We searched for all occurrences of the specified motifs in *S. oneidensis* MR-1 UORs using the 'search by motif' function in the software.

Searching with $AACn_4GTT$ motif gave 301 UORs containing the motif with a total of 281 non-redundant motifs. (Some UORs contain upstream DNA sequence for more than one

ORF and consequently the identical motif can occur in more than one UOR in the database. Also, some UORs contain this motif at two different positions.) (see Appendix, Table A). Searching with GAACn$_4$GTT motif gave 82 UORs containing the motif with a total of 78 non-redundant motifs (only *SO1859* UOR contains this motif at two different positions).

A large number of the ORFs were discovered from the motif searches. Obviously, not all of the AACn$_4$GTT or GAACn$_4$GTT motifs identified are genuine *SO3363p* binding sites as it seems that the SELEX motif is too short to be the only prerequisite for specific binding of *SO3363p* to DNA. Therefore, a comparison was performed between the up-regulated genes under anaerobic growth condition from the microarray expression profiles and the UORs identified in this section.

We found that many of the UORs of the ORFs which are up-regulated in the microarray experiment contain the AACn$_4$GTT motif (see Table 3.5). More importantly, the UORs of operon *SO0096-98* encoding alkyl hydroperoxide reductase subunits, *SO0725* encoding catalase/peroxidase HPI, *SO1070* encoding catalase, which are usually involved in oxidative stress, contain this motif. Also, UORs of DNA *SO0060* and *SO1427*used in the DNA-affinity capture experiment also contain this motif. This suggests that the AACn$_4$GTT may play an important role in the *SO3363p* specifically binding to DNA. An EMSA experiment was performed between *SO3363p* and *SO1427* promoter in the presence of heparin (Figure 3.10). Two shifted bands were observed at ratios of 2, 4, 8, and 16 and they still can be seen when 50 and 100 μg/mL of heparin were added. This suggested that *SO3363p* binds to *SO1427* promoter DNA containing AACn$_4$GTT motif specifically and the presence of AACn$_4$GTT in promoter DNA is important for specific binding between *SO3363p* and DNA.

Table 3.5 Overlap between up-regulated ORFs under anaerobic ($-O_2$) growth condition and 200 UOR database search results for the motif $AACn_4GTT$

ORF	Fold change from $-O_2$ to $+O_2$	UOR[c]	Motif	ORF description / operon function[d]
SO0060	4.0	SO0060	AACCTGAGTT	Sensor histidine kinase
SO0403	6.1	SO0403	AACAGTAGTT	Hypothetical protein
SO0724	5.4	SO0724	AACCTCGGTT	Hypothetical protein
SO0725	10.4	SO0725	AACCTCGGTT	Catalase/peroxidase HPI
SO0958	10.0	SO0958	AACAATAGTT	Alkyl hydroperoxide reductase, C subunit
SO1070	65.0	SO1070	AACTCTGGTT	Catalase
SO1427	8.8	SO1427	AACATTCGTT	Decaheme cytochrome c
SO2870	6.6	SO2870	AACGTGGGTT	Hypothetical protein

[a]For ORFs which are at the beginning of operons, only the first ORF in the operon is listed with its corresponding change.

[b]ORF descriptions and operon functions are essentially derived from Tigr database.

[c]UOR (Upstream of ORF Region) designation corresponds to the locus of the ORF from which the upstream sequence was taken and the length of UORs used in the search is 200 bp relative to the translation start site.

[d]ORF annotations are from NCBI

3.4.3 UOR database search using longer $GTTCn_8GAAC$ and $GAACn_4GTTC$ motifs

In order to obtain more information from SELEX on the consensus DNA-binding site recognized by *SO3363p*, the extended motifs $GTTCn_8GAAC$ and $GAACn_4GTTC$ were used to search for other potential binding sites in the upstream regions of other genes (UORs) within the whole *S. oneidensis* MR-1 genome. The search result using these motifs showed that only 6

upstream regions contain the first motif and 14 upstream regions contain the second motif (one UOR contains this motif at two positions) (Table 3.6).

Table 3.6 200UOR database search results for the motif $GTTCn_8GAAC$ and $GAACn_4GTTC$

UOR[a]	Motif	Start[b]	Stop[b]	ORF Annotation[c]
SO1288	gttcaattgctcgaac	-48	-32	ribosomal protein S21
SO1289	gttcaattgctcgaac	-185	-169	O-sialoglycoprotein endopeptidase
SO1542	gttcttttagtggaac	-149	-133	ISSod4, transposase, interruption
SO3103	gttcgcgtcactgaac	-149	-133	AcrB/AcrD/AcrF family protein
SO3186	gttcagagtgaagaac	-76	-60	glucose-1-phosphate-thymidylyltransferase
SO3218	gttcaggttggcgaac	-138	-122	flagellar biosynthetic protein FliP
SO0322	gaacgatagttc	-191	-179	hypothetical protein
SO0328	gaaccttagttc	-23	-11	hypothetical protein
SO0329	gaactaaggttc	-177	-165	hypothetical protein
SO0620	gaacctcggttc	-156	-144	conserved hypothetical protein
SO0656	gaacctatgttc	-104	-92	ISSodl, transposase OrfA
SO1859	gaacccgggttc	-64	-52	hypothetical protein
SO1859	gaacccgggttc	-153	-141	hypothetical protein
SO2108	gaactcaagttc	-85	-73	periplasmic glucans biosynthesis protein MdoH
SO2133	gaactagggttc	-39	-27	hypothetical protein
SO3292	gaactatcgttc	-139	-127	GMP synthase
SO3607	gaacaactgttc	-75	-63	ISSod1, transposase OrfA
SO3896	gaacctaagttc	-29	-17	outer membrane porin, putative

SO3943	gaacacgcgttc	-75	-63	protease DegS
SO4583	gaacttttgttc	-183	-171	RNA polymerase sigma-32 factor
SO4746	gaacaagcgttc	-103	-91	ATP synthase F1, epsilon subunit

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

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

[c]ORF annotations are from TIGR.

This number seems reasonable even though it is not likely that all of the motifs are genuine *SO3363p* binding sites since some of them are not in the promoter region of an operon.

3.4.4 SO3363p tends to bind shorter DNA containing the SELEX motif in vitro

I attempted to verify that the identified motifs are real *SO3363p* binding sites by conducting EMSA on three selected putative promoter regions. Of those ORFs listed in Table 3.6 obtained from the 200 UOR database search, 11 of them were the first gene in their operons based on operon prediction (*www.Microbesonline.org*) and therefore more likely to contain a promoter with potential regulatory elements in their upstream DNA. I chose *SO1288* since *SO1288* and *SO1289* are divergently transcribed genes that share an intergenic region of 214 bp containing the GTTCn_8GAAC motif. The other two I chose were *SO1859* and *SO3607* containing GAACn_4GTTC motif. The DNA fragments used in EMSA experiment are about 400 bp in length containing ~300 bp upstream region and ~100 bp downstream relative to the translation start site; the primers used are listed in Table 3.7.

Table 3.7 Primers for selected DNA probes used in EMSA

Probe name	Genome coordinates	Forward primer[a]	Reverse primer[a]	Probe length (bp)
SO1288	1340331-1339908	tggcttctcgtagaattcac	taaagcatgggaaagtaacc	424
SO1859	1951092-1951516	tgtatattatctcggtcgttcg	aatgcaaaaagccgctaattg	425
SO3607	3769189-3768794	caacgctataaccttgctccag	tcctcagaatcagatgaatatggtg	396
SO4583[c]	4778981-4779360	cagtctcagcgtctaacatg	tgtcggtaaggcaacatttaag	380
SO0523[c]	546961-547430	aacaactgagtacccgta tc	ataaactggttatcttcgagctg	470
SO2052[c]	2151041-2151530	tggcttgtggttgtatttcc	ttcagacaattcaaccgctg	490
SO2633[c]	2768441-2768800	agcaacaaggcgtgatagaa	ataggctgaaacggatgaat	360

[a]primers are listed from 5′to 3′ are modified with HEX fluorophore at 5′-.

[b]primers are listed from 5′ to 3′ and are modified with 6FAM fluorophore at 5′-.

[c]primers are not modified with either HEX or 6FAM.

EMSA was performed between only *SO3363p* and DNA probes of *SO1288*, *SO3607* due to the failure of PCR amplification of *SO1859* DNA probe. The result showed that there was no obvious binding for *SO1288* containing GTTCn_8GAAC and there was weak binding for *SO3607* probe containing GAACn_4GTTC motif (Figure 3.11). This may suggest that the presence of palindrome GTTCn_8GAAC is less important than GAACn_4GTTC for *SO3363p*-DNA interaction. I also noticed that *SO3363p* seems to have much better binding affinity to shorter DNA probe (Fig. 3.9, Fig. 3.10) instead of longer DNA probe containing the SELEX binding motif (Fig. 3.12). This result was out of my expectation and I think probably the presence of the palindrome GTTCn_8GAAC is not required for protein-DNA binding and the motif GTn_{12}ACn_4GTTC may

favor the binding better since only 'GT', 'AC' and 'GTTC' are more conserved in this multilevel consensus DNA sequence(Fig. 3.8B).

Based on this assumption, the extended motif $GTn_{10}GAAC$ and $GTn_{12}ACn_4GTTC$ was used in *S. oneidensis* MR-1 UOR database search. The search result with the first extended motif gave 253 hit while the search result with the second extended motif gave 14 hits and the search result of the second extended motif was shown in Table 3.8.

Table 3.8 200UOR database search results for the motif $GTn_{12}ACn_4GTTC$

UOR [a]	Motif	Start [b]	Stop [b]	ORF Annotation [c]
SO0523	GTTATTACTGGTGGACTTAAGTTC	-182	-159	Transcriptional regulator, LysR family
SO0524	GTTATTACTGGTGGACTTAAGTTC	-87	-64	HlyD family secretion protein
SO1298	GTAGAATACGACACACTGTTGTTC	-68	-45	Hypothetical Na+/H+ antiporter
SO1571	GTGTAAGGTAGCTAACCTCAGTTC	-51	-28	Hypothetical protein
SO1844	GTAGCCAATTTGTAACCTTTGTTC	-123	-100	Extracellular nuclease, putative
SO1910	GTTTGAATTGGCGCACAGATGTTC	-125	-102	1,4-dihydroxy-2-naphthoate octaprenyltrasferase, putative
SO2051	GTTTTATATTTTTTACAATCGTTC	-97	-74	Hypothetical protein
SO2052	GTTTTATATTTTTTACAATCGTTC	-60	-37	Hypothetical oxidoreductase
SO2371	GTCCTAGCTTGCAAACCATGGTTC	-99	-76	Hypothetical protein
SO2372	GTCCTAGCTTGCAAACCATGGTTC	-88	-65	Hypothetical protein
SO2422	GTCATGGCCATAAAACTTTGGTTC	-174	-151	Hypothetical protein
SO2633	GTCATGCGCGAATGACGATGGTTC	-70	-47	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase
SO2927	GTCTGTGTGTTACCACGCCAGTTC	-110	-87	ABC transporter, ATP-binding protein
SO3202	GTGCAAAATGCCACACATTTGTTC	-91	-68	Purine-binding chemotaxi protein Chew

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

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

[c] ORF annotations are from NCBI.

Among the 14 UORs containing the $GTn_{12}ACn4GTTC$ motif, I chose *SO0523*, *SO2052* and *SO2633* to study the interactions of their promoter regions with *SO3363p* by EMSA. Among these three ORFs, *SO0523* is divergently transcribed with *SO0524* and they shared the promoter region containing the motif. *SO2052* also shares the promoter region with *SO2051* and *SO2633* is the first gene in operon of *SO2633-35*. The primers for PCR amplification are listed in Table 3.5. EMSA results showed that there was a weak binding between SO3363p and DNA probe SO0523 and SO2652 (Figure 3.12) while no binding was observed between *SO3363p* and *SO2633* promoter.

So, SELEX experiment identified *SO3363p*'s recognition motif containing consensus DNA sequences and EMSA experiment verified that the presence of palindromes is important for the specific interactions between *SO3363p* and DNA. However, the weak binding between the *SO3363p* and the promoters containing those palindromes, makes it difficult to know what the real binding motif is for *SO3363p* specific binding without footprinting data.

Figure 3.1 DNA affinity protein capture with *SO0060* and *SO3921* probes. Blue-silver stained SDS-PAGE gel of eluted proteins from DNA affinity capture with *SO0060* and *SO3921* probes incubated in soluble cell extracts from cells grown in the presence (red and green) and absence (blue and purple) of oxygen with the corresponding band intensities (traces to the left are for SO3921 lanes and traces to the right are for SO0060 lanes). Arrows indicate identified proteins (Table 3.2).

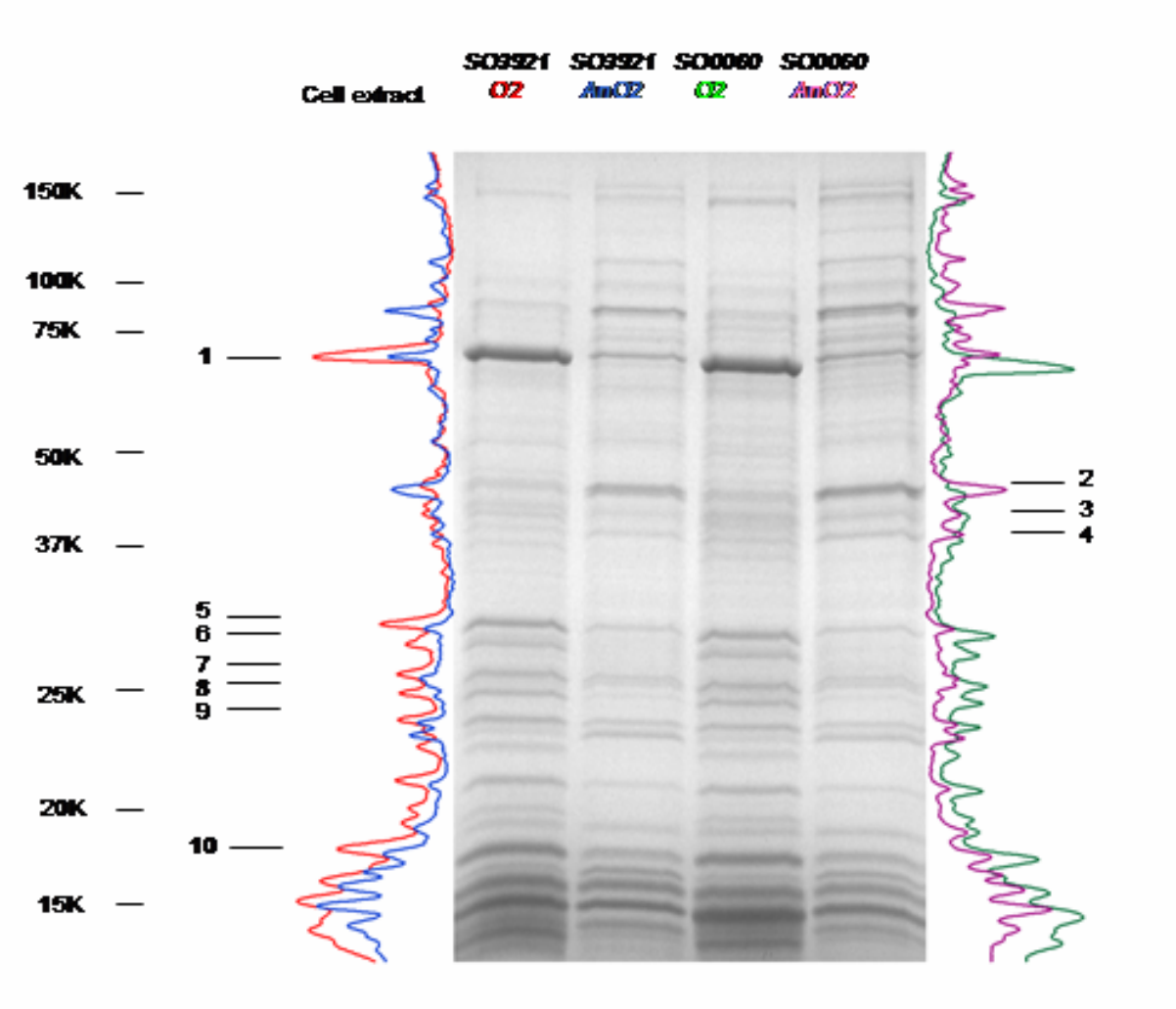
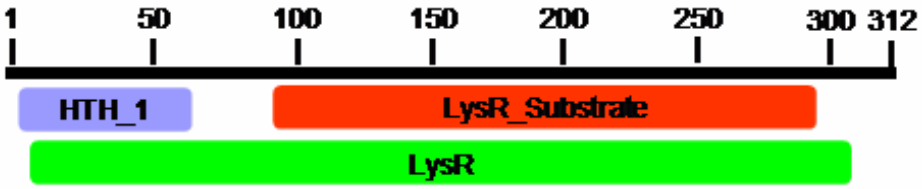

Cell extract
SO3921 O2
SO3921 AnO2
SO0060 O2
SO0060 AnO2
150K
100K
75K
50K
37K
25K
20K
15K
1
2
3
4
5
6
7
8
9
10

Figure 3.2 *SO3363p* sequence analysis. Conserved domain search results for SO3363 protein sequence from online tools available at NCBI. SO3363 sequence is represented by a black line with matching conserved domains in different colors.



HTH_1, pfam00126, HTH_1, Bacterial regulatory helix-turn-helix protein, lysR family..

CD-Length = 60 residues, 100% aligned, Score = 55.8 bits (81), Expect = 8e-09

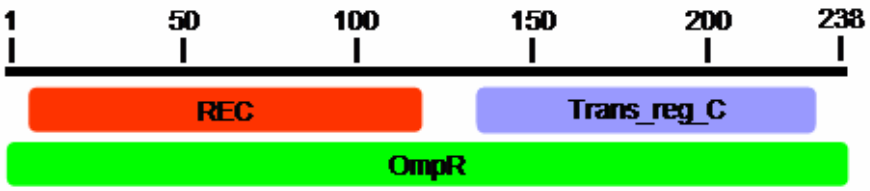
LysR_Substrate, pfam03466, LysR_substrate, LysR substrate binding domain.

CD-Length = 208 residues, 100% aligned, Score = 117 bits (368), Expect = 3e-27

LysR, Transcriptional regulator [Transcription]

CD-Length = 297 residues, 99.3% aligned, Score = 142 bits (368), Expect = 8e-35

Figure 3.3 *SO3988p* sequence analysis. Conserved domain search results for SO3988 protein sequence from online tools available at NCBI.



REC, Signal receiver domain

CD-Length = 113residues, 100% aligned , Expect = 1e-26

trans_reg_C, Effector domain of response regulator

CD-Length = 95residues, 100% aligned, Expect = 2e-18

OmpR, Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain [Signal transduction mechanisms / Transcription]

CD-Length = 229 residues, 99.5% aligned , Expect = 1e-51

Figure 3.4 Expression and purification of his_6-*SO3363p* and *SO3363p*-his_6. Left, SDS-PAGE of samples of time points after 0.1 mM IPTG was added at 37°C when OD=0.6~0.8 and the culture was harvested after 10-12 h of growth at 20°C. Right**,** final purified protein after nickel-affinity purification.

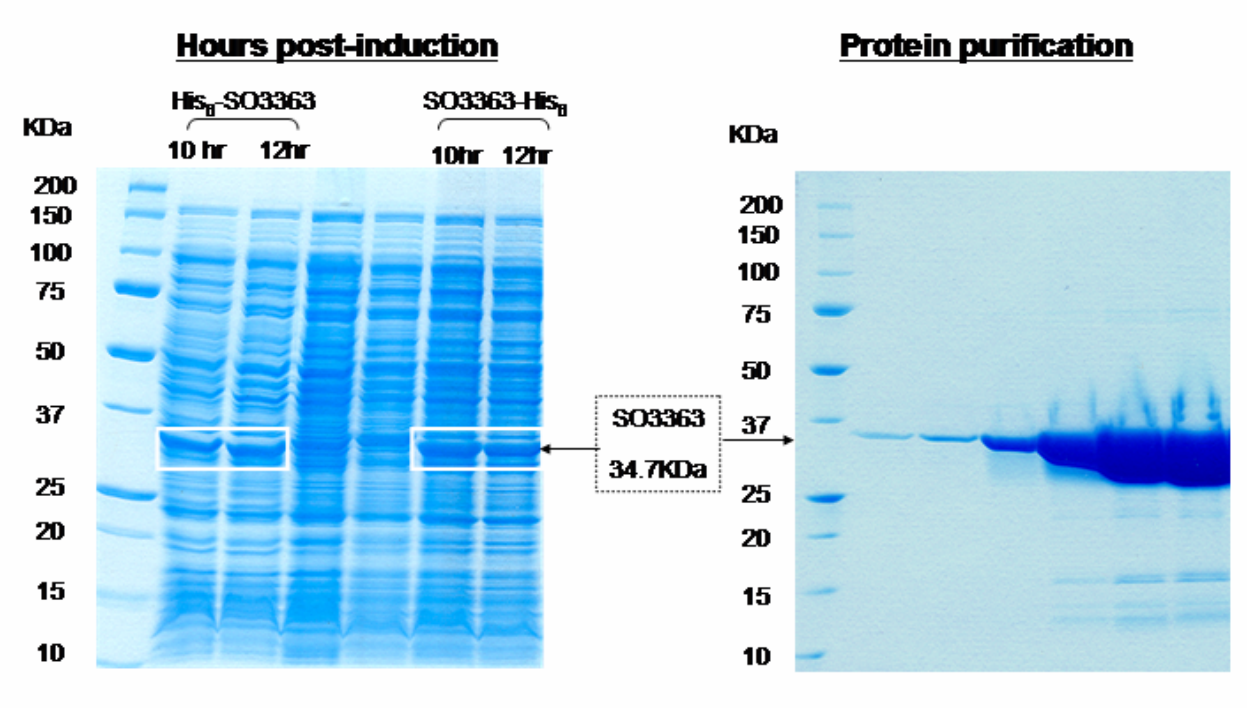

Hours post-induction
Protein purification
KDa
His6-SO3363
SO3363-His6
10 hr
12hr
10hr
12hr
200
150
100
75
50
37
25
20
15
10
SO3363
34.7KDa
KDa
200
150
100
75
50
37
25
20
15
10

Figure 3.5 Quaternary structure of C-terminal his-tagged *SO3363p*. The elution profiles of molecular weight standards and C-terminal his-tagged SO3363 protein from a Superdex 75 10/300 GL column (GE Healthcare) are shown.

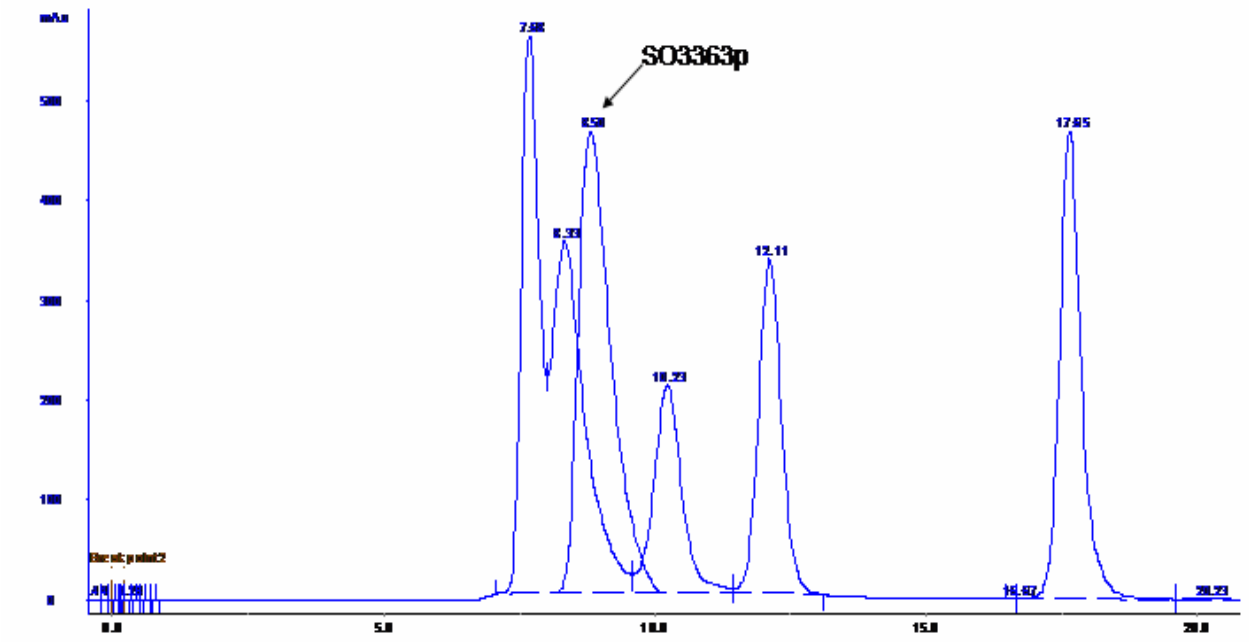

SO3363p
7.88
8.58
8.33
12.11
18.23
17.85
16.97
20.23
5.0
10.0
15.0

Figure 3.6 *His$_6$-SO3363*p binds to *SO0060, SO1427* and *SO3921* probes. EMSAs with *his$_6$-SO3363p* and *SO0060*, *SO1427*and *SO3921* probes were performed. The DNA probes used are indicated at the top of each gel image with corresponding protein/DNA mole ratios listed above each lane. UOR=Upstream of ORF Region. DNA (100 nM) was incubated with protein in buffer (50 mM Tris·HCl, 100 mM KCl, 5% glycerol, 1 mM DTT, 1 mM EDTA, pH 8.0) for 20 min at 37 °C. Gels were stained with SYBR Green I nucleic acid gel stain.

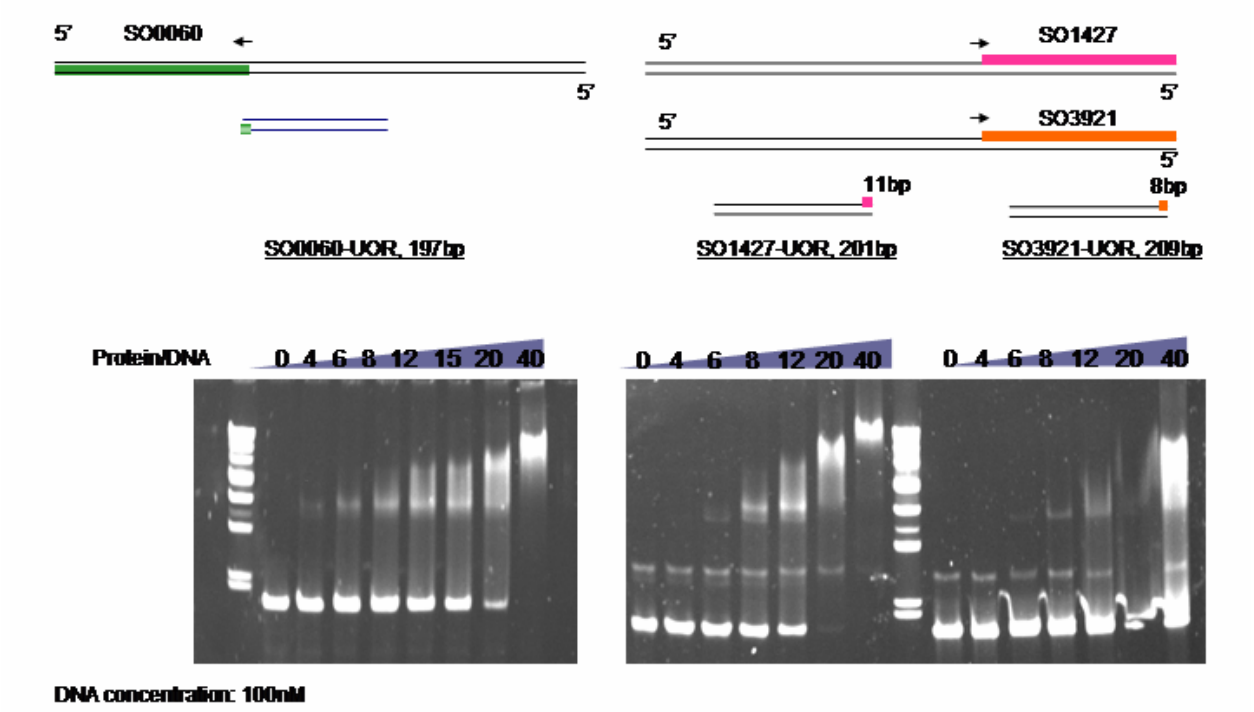

5'
SO0060
5'
5'
SO1427
5'
5'
SO3921
5'
11bp
8bp
SO0060-UOR, 197bp
SO1427-UOR, 201bp
SO3921-UOR, 209bp
Protein/DNA
0 4 6 8 12 15 20 40
0 4 6 8 12 20 40
0 4 6 8 12 20 40
DNA concentration: 100nM

Figure 3.7 His_6-*SO3363p* binds to *SO1427* probe in the presence of 10 mM DTT, diamide, β -ME and H_2O_2. EMSAs with *his_6-SO3363p* and *SO1427* (323 bp) probe under different reducing and oxidizing reagents were performed. The DNA probe used is indicated at the top of the gel images with corresponding protein/DNA mole ratios listed above each lane. DNA (100 nM) was incubated with protein in buffer (50 mM Tris·HCl, 100 mM KCl, 5% glycerol, 1 mM DTT, 1 mM EDTA, pH 8.0) for 20 min at 37°C. Gels were stained with SYBR Green I nucleic acid gel stain.

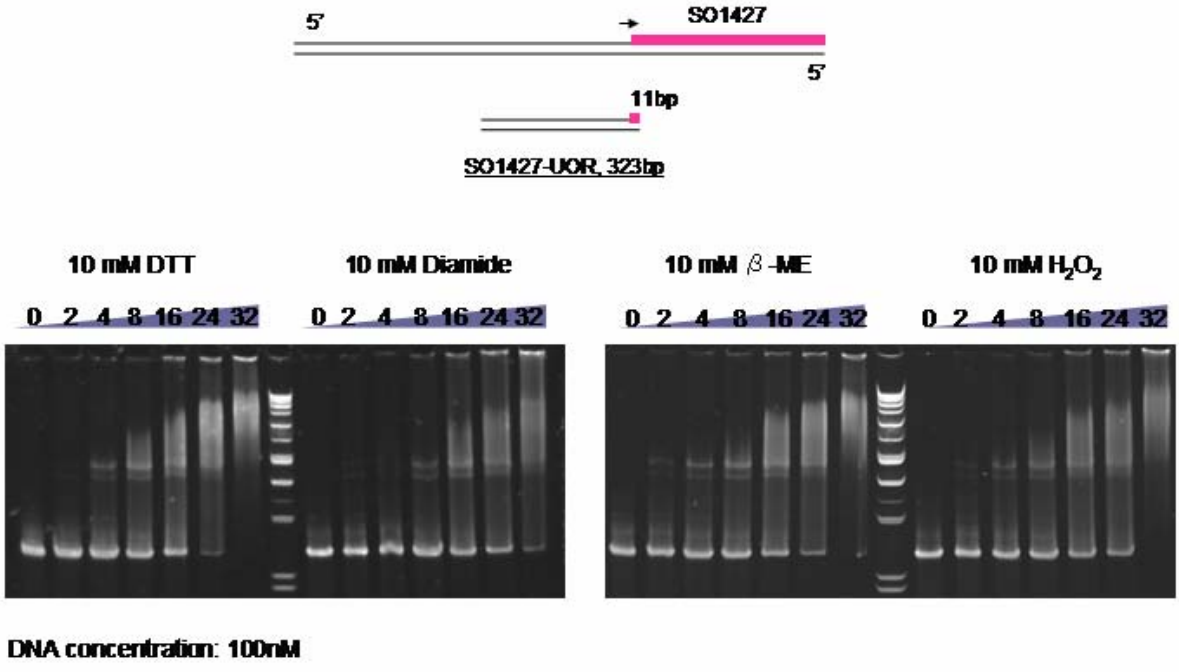

5'
SO1427
5'
11bp
SO1427-UOR, 323bp
10 mM DTT
10 mM Diamide
10 mM β-ME
10 mM H2O2
0 2 4 8 16 24 32
0 2 4 8 16 24 32
0 2 4 8 16 24 32
0 2 4 8 16 24 32
DNA concentration: 100nM

Figure 3.8 SELEX-determined *SO3363p* DNA-binding motif. A. Motif logo (generated using WebLogo for motif uncovered from all 31 selected SELEX sequences as determined by MEME motif-finding software blind search. **B.** Motif logo for 19 out of 27 SELEX sequences. **C.** Motif logo for 22 out of 27 SELEX sequences containing the palindrome.

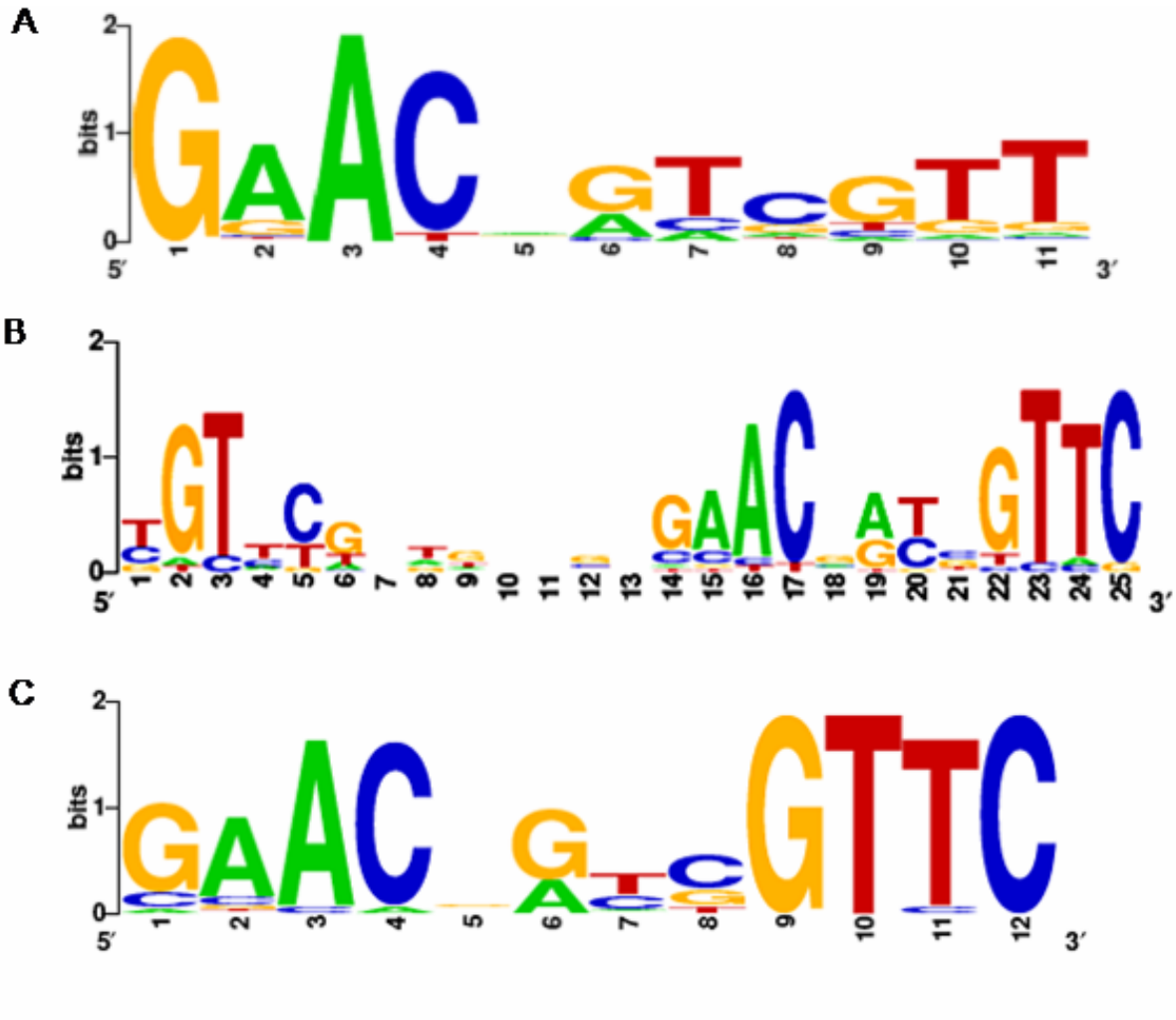

A
B
C
bits
5′
3′

Figure 3.9 *SO3363p* binds specifically to selected DNA probe in the presence of heparin. 200 nM selected and control DNA probes were incubated with protein in buffer (50 mM Tris, 20 mM KCl, 5% glycerol, 1 mM EDTA, pH 8.0) with and without heparin (concentrations indicated) for 20 min at 37° C. Gel was stained with SYBR Green I nucleic acid gel stain. Note that much higher heparin concentration is required to compete off the specific binding of *SO3363p* to selected DNA containing the extended palindrome.

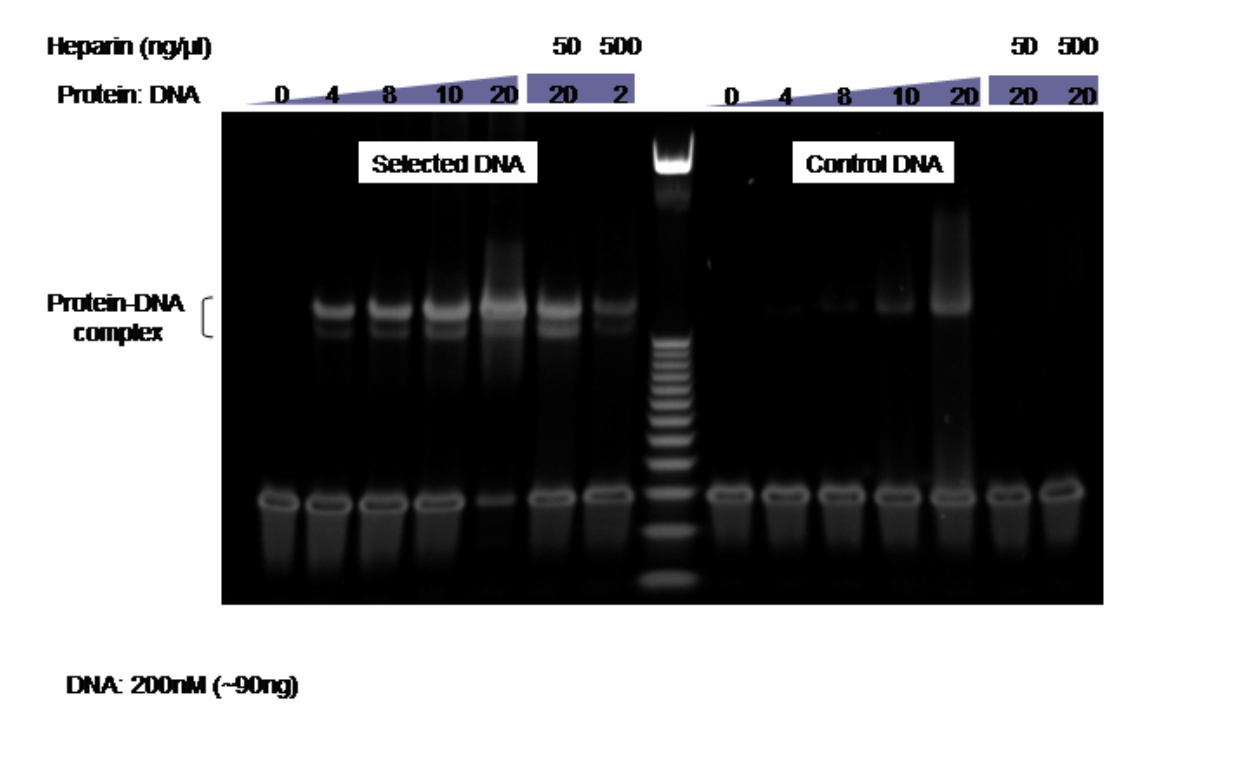
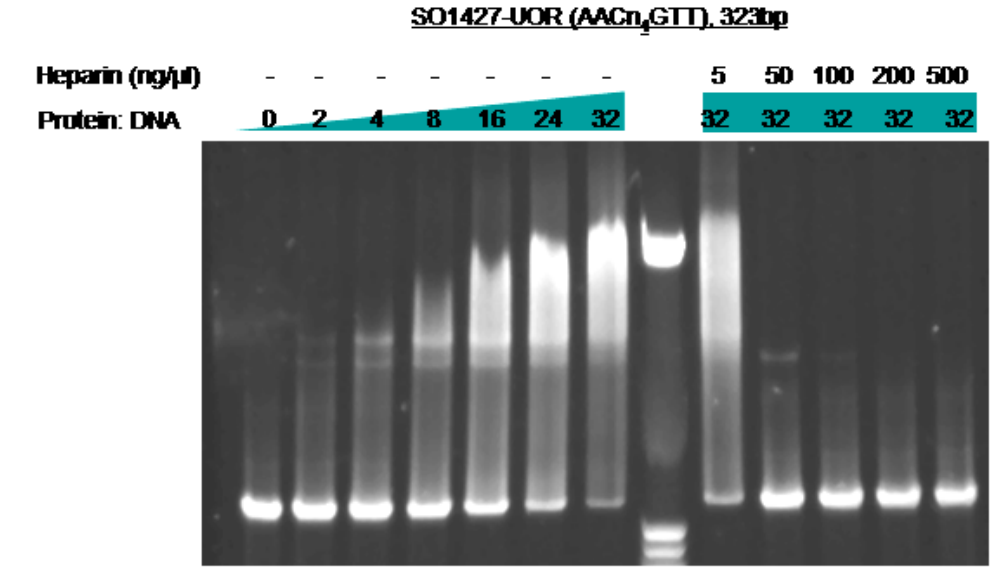
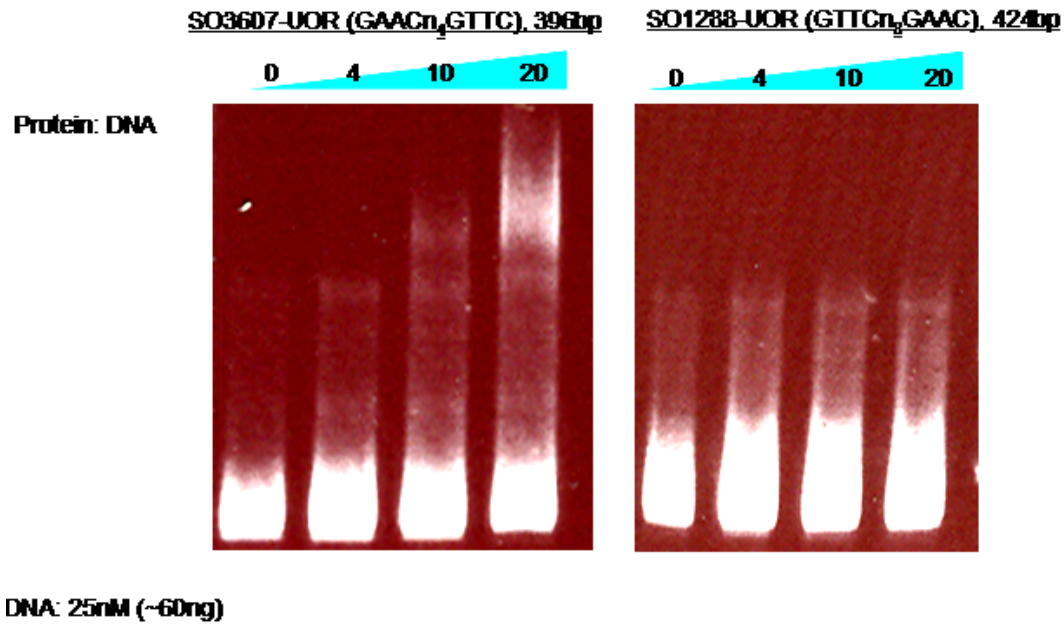

Heparin (ng/µl)
50 500
50 500
Protein: DNA
0 4 8 10 20 20 2
0 4 8 10 20 20 20
Selected DNA
Control DNA
Protein-DNA complex
DNA: 200nM (~90ng)

Figure 3.10 *SO3363p* binds specifically to *SO1427* probe in the presence of heparin. 50 nM *SO1427* probe (327 bp) was incubated with protein in buffer (50 mM Tris, 20 mM KCl, 5% glycerol, 1 mM EDTA, pH 8.0) with and without heparin (concentrations indicated) for 20 min at 37° C. Gel was stained with SYBR Green I nucleic acid gel stain.



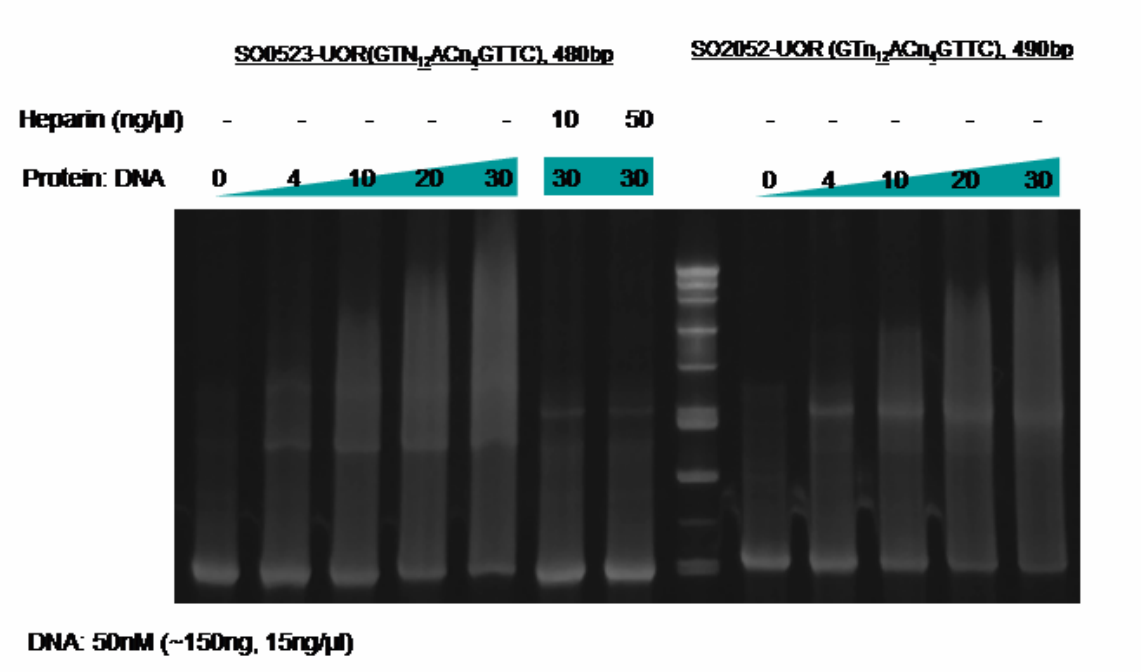
DNA: 50nM (~100ng, 10ng/μl)

Figure 3.11 EMSA between *SO3363p* and *SO1288*, and *SO3067* DNA probes. 50 nM *SO3067* and *SO1288* DNA probes were incubated with protein in buffer (50 mM Tris, 20 mM KCl, 5% glycerol, 1 mM EDTA, pH 8.0) with and without heparin (concentrations indicated) for 20 min at 37° C. Gel was stained with SYBR Green I nucleic acid gel stain.



DNA: 25nM (~60ng)

Figure 3.12 *SO3363p* binds *SO0523* and *SO2052*. 50 nM *SO0523* and *SO2052* DNA probes were incubated with protein in buffer (50 mM Tris, 20 mM KCl, 5% glycerol, 1 mM EDTA, pH 8.0) with and without heparin (concentrations indicated) for 20 min at 37° C. Gel was stained with SYBR Green I nucleic acid gel stain.


SO0523-UOR(GTN12ACn4GTTC), 480bp
SO2052-UOR (GTn12ACn4GTTC), 490bp
Heparin (ng/µl) - - - - - 10 50 - - - - -
Protein: DNA 0 4 10 20 30 30 30 0 4 10 20 30
DNA: 50nM (~150ng, 15ng/µl)

CHAPTER 4

FUNCTIONAL CHARACTERIZATION OF SO3363

4.1 *SO3363*-knockout mutant construction

In order to elucidate the functions of *SO3363*, an in-frame deletion mutant of this gene was constructed, followed by characterization of its phenotypic and physiological changes compared with the wild type strain. The mutant was constructed according to a protocol from Jizhong Zhou's laboratory at the University of Oklahoma and was performed in that laboratory. In the first step, we generated a deleted copy of the target gene *SO3363* using a two-step asymmetric/crossover PCR amplification. The products from the asymmetric PCR amplification are 3o-3i, 845 bp in length, 5o-5i, 673 bp in length and the crossover PCR product is 5o-3o, 776 bp in length (see Table 4.1). In a second step, we cloned 5o-3o into the suicide vector pDS3.0 and the resulting suicide plasmid was integrated into the *S. oneidensis* MR-1 chromosome by single crossover. The correct insertion was confirmed by colony PCR with LF/SR and LR/SF primers, which should produce two strong DNA bands of 1608 bp for the wild type strain and 1066 bp for the mutant strain or 1822 bp for the wild type strain and 852 bp for the mutant strain (see Table 4.1 and Fig. 4.1A). In a third step, we resolved the integrated plasmid from the chromosome by sucrose counter-selection and the result was also confirmed by colony PCR with LF/LR primers. The colonies of the mutant should give a strong DNA band of 1703 bp and wild type colonies give a DNA band of 2479 bp (see Table 4.1 and Fig. 4.1B).

Table 4.1 PCR products obtained from SO3363 mutant construction

PCR products	Length (bp)
3o-3i	845
5o-5i	673
5o-3o	776
Wild type LF-SR	1608
Wild type LR-SF	1822
Mutant LF-SR	852
Mutant LR-SF	1066
Wild-type LF-LR	2479
Mutant LF-LR	1703

4.2 Characterization of phenotype of mutant strain by phenotype miroarray

In order to identify the function of the *SO3363* gene, the cellular phenotype of the mutant strain (compared to wild type) was carried out by using Phenotype MicroArrays (PMs) [114]. PM tests are performed in 96-well microplates containing different nutrients or inhibitors in which cell respiration is measured with a redox indicator, tetrazolium violet dye. Mutants were expected to display one or more altered phenotypes if the mutated gene has a role under the condition(s) tested.

The current PM platform (also used in Jizhong Zhou's laboratory at the University of Oklahoma) for microbial cells contains 20 microplates, providing ~2000 phenotype tests designed to cover known metabolic pathways [115]. The PM1 and PM2 plates test different carbon sources, the PM3 to PM8 plates test various N, P, or S sources, the PM9 plate tests

different osmotic stress factors and ion effects and the PM10 plate investigates different pH growth range and pH regulation with different amino acids. The rest of the ten plates test the sensitivity of cells to a number of chemicals such as antibiotics, antimetabolites, membrane-active agents, respiratory inhibitors, and toxic metals. PM tests were performed essentially as described in section 2.11. IF-0 inoculating fluid was used for PM1 to PM3 and IF-10 was used for PM9 to PM10 in this experiment. The OmniLog records a 96-box chart corresponding to the 96 wells in the PM for each strain and for each PM. The cellular response is recorded showing the amount of purple color (from redox dye) in each well (vertical axis) throughout the time course (horizontal axis) of the assay. The *SO3363* mutant strain was recorded as a green tracing and the MR-1 wild type strain was recorded as a red tracing. These tracings were overlaid to identify quickly the phenotype differences between the two strains (Fig. 4.2).

Both the mutant and the control strain did not grow well in PM3 (data not shown) and they grew better on PM9 and PM10 compared to PM1 and PM2. Only one phenotype was detected in PM2 under gelatin condition (Fig. 4.2). For most of the growth detected, the mutant and control strain gave virtually identical phenotypes resulting in yellow color. The mutant strain lost phenotype (grew more slowly than wild type) under 7 test conditions (uridine, 2-deoxy adenosine, adenosine, and pyruvic acid in the PM1 carbon source microplate; pH 8.5, pH 9, pH 9.5+ tryptamine and X-caprylate in PM10 microplate) and gained phenotype under only three conditions (formic acid, α-hydroxy butyric acid in the PM1microplate; sodium lactate in the PM9 microplate). The major phenotype differences in PM1, 2, 9, 10 are also summarized in Table 4.2.

Table 4.2 Major phenotype differences between the mutant and wild type strains

Test[a]	Differences[b]	Mode of action[c]
N-Acetyl-D-Glucosamine	+ -	Carbon source
L-Lactic Acid	+ -	Carbon source
Formic Acid	+	Carbon source
Tween 20	+ -	Carbon source
Tween 40	+ -	Carbon source
Uridine	-	Carbon source
α-hydroxy butyric acid	+	Carbon source
2-Deoxy Adenosine	-	Carbon source
Adenosine	-	Carbon source
Inosine	+ -	Carbon source
Glycyl-L-Glutamic acid	-	Carbon source
Methyl pyruvate	+ -	Carbon source
Gelatin	-	Carbon source
Pyruvic acid	-	Carbon source
Sodium sulfate 5%	-	Osmolarity
Urea 3%	+ -	Osmolarity
Sodium lactate 6%	+	Osmolarity
Ammonium sulfate pH8 100mM	-	Toxicity, ammonia [c]
Sodium nitrite 10mM	-	Transport, toxic anion [c]
pH8.5	-	Alkaline pH
pH9	-	Alkaline pH
pH9.5 + Tryptamine	-	Alkaline pH
X-Caprylate	-	Alkaline pH

[a] Chemicals were tested in 96-well PMs

[b] The OmniLog-PM software generates time course curves for respiration (tetrazolium color formation) and calculates differences in the areas for mutant and control cells. "+" indicates that the mutant showed greater rates of respiration than the control through the time monitored. "-" indicates that the control showed greater rates of respiration than the mutant through the time monitored. "+ -" indicate that the rates of respiration for mutant are sometimes higher and sometimes lower than the control through the time and thus we can see both red and green area under the test conditions.

[c] The mode of action is from reference [114].

It is important to understand the relationship between the phenotype changes and the deleted gene SO3363 in order to determine the function of SO3363. Some questions remain: why did the mutant show defective growth under 2-deoxy adenosine, adenosine, higher pH (8.5-9+) and X-caprylate? More importantly, why did the mutant gain phenotype under formic acid and sodium lactate conditions? We do not have complete answers to these questions at this point.

4.3 *SO3363* deletion mutant is more resistant to hydrogen peroxide

Bacteria may behave differently upon environmental changes and differences in the bacterial growth curves observed under varying physiological conditions, or from the growth of different organisms, which indicates biological differences. In order to determine the biological differences between the *S. oneidensis* MR-1 wild type strain and the *SO3363* knock out mutant strain, their growth curves were recorded under different conditions. *SO3363* was identified by DNA affinity protein capture using the promoters of genes with expression levels that responded

to a change from anaerobic to aerobic growth; a reasonable hypothesis is that it is a regulator responding to oxidative conditions. So, the growth conditions chosen were based on the oxidative conditions we have been studying and the results just described from phenotype microarrays. First, the growth of the wild type and the mutant strains under aerobic and anaerobic conditions was recorded (Fig. 4.3A, E). For aerobic growth, only LB medium was used; for anaerobic growth, modified LB medium (tryptone 10: yeast 5: NaCl 1), 20mM sodium lactate was used as electron donor and 10 mM sodium fumarate, 10 mM sodium nitrate, 10mM DMSO, and 10 mM TMAO, were tested as electron acceptors. Insignificant differences were observed between mutant and wild type growth under either condition (Table 4.3).

In section 4.2, we also detected the phenotype differences under varying physiological conditions between the two strains, so we chose to test those physiological conditions for the growth of the two strains under normal aerobic growth condition. These conditions included pH ranging from 8.0-9.5, sodium sulfate at levels of 1-5 wt % (see Fig. 4.3B-C, Table 4.3). Furthermore, hydrogen peroxide was tested as another type of oxidative condition because a mutant constructed by our collaborators showed resistance to hydrogen peroxide (see Fig. 4.3D, Table 4.3). All the physiological conditions used and the comparison results are summarized in Table 4.3.

Table 4.3 Growth difference between *SO3363* mutant and wild type under tested conditions

Tested conditions[a]	Aerobic or anaerobic	OD[b]	Differences[d]
aerobic	aerobic	(1.29, 1.35)	no
pH 8.5	aerobic	(0.94, 0.94)	no
pH 9.0	aerobic	(0.78, 0.76)	no

Sodium sulfate 4%	aerobic	(0.54, 0.41)	yes
Sodium sulfate 5%	aerobic	(0.48, 0.36)	yes
H_2O_2 0.6 mM	aerobic	(1.22, 1.10)	yes
H_2O_2 0.8 mM	aerobic	(1.25, 0.37)[c]	yes
Sodium lactate 20 mM[e]	anaerobic	(0.08, 0.08)	yes
Sodium lactate 20 mM Sodium fumarate 10 mM	anaerobic	(0.29, 0.29)	no
Sodium lactate 20 mM[e] Sodium nitrate 10 mM	anaerobic	(0.45, 0.45)	no
Sodium lactate 20 mM[e] DMSO 10 mM	anaerobic	(0.26, 0.25)	no
Sodium lactate 20 mM[e] TMAO 10 mM	anaerobic	(0.28, 0.28)	no

[a] The tested conditions here meant the chemicals or pH used

[b] The optical density (OD) value the when the stationary phase begins was used here and the OD of the wild type and the mutant is enclosed in a bracket and separated by comma.

[c] For this experiment, the mutant strain did not reach the stationary phase during the recording time and the maximum OD was used here.

[d] "yes" indicates there is growth difference between the wild type and mutant strain under the tested condition if the OD difference between the two strains is more than 0.1 and "no" means the growth of the two strains under the tested conditions are essentially identical.

[e] Growth curves under these tested conditions are not shown in Fig 4.3.

Comparison of all the growth curves in Fig 4.3 illustrates that both the wild type and mutant strains grew much better under aerobic conditions than anaerobic conditions based on the OD values recorded. Among the tested conditions under aerobic growth, both the wild type and mutant strain showed growth deficiency in the presence of sodium sulfate and pH 9.0 compared to the other aerobic conditions tested. For aerobic growth with 4% and 5% sodium sulfate, the mutant strain showed some growth deficiency compared to the wild type strain and this result is somewhat consistent with the result from the phenotype microarray that the wild type cells grew better under these two conditions (see Fig. 4.2). More importantly, for aerobic growth with H_2O_2, the wild type strain began to show growth deficiency with 0.6 mM H_2O_2 and the mutant strain reached the stationary phase one hour earlier with a higher OD value than the wild type strain. Furthermore, this trend was more obvious in the presence of 0.8 mM H_2O_2 since the wild type strain failed to reach the stationary phase while the mutant strain did with an OD value of 1.25. This is a clear indication that the mutant strain is more resistant to hydrogen peroxide and *SO3363* gene may play an important role in hydrogen peroxide response.

4.4 Comparison of transcriptional profiles of the mutant and wild type strains

In order to identify genes whose transcription was affected by the deletion of the *SO3363* gene, *Shewanella oneidensis* MR-1whole genome microarrays were employed to dissect the transcriptomic differences elicited by the mutation in *SO3363* during aerobiosis and anaerobiosis. Gene expression in the mutant cells, recovered from log phase growth in aerobic medium with or without hydrogen peroxide, was compared with the expression of wild type cells using the same two growth conditions. Although little difference in physiology was observed between the wild-

type and SO3363 deletion strain during anaerobiosis with fumarate as the sole electron acceptor, gene expression between the two strains was also compared to elucidate the genotypic response.

Standard statistical analysis ($p < 0.05$ as cut-off value in *t*-test) was used to select genes whose expression has significantly changed due to the deletion of *SO3363* under tested conditions. For the aerobic growth condition, 123 genes in total with at least 1.5-fold change (and up to 11-fold change) in expression passed this statistical analysis and genes displaying significant differences in expression were observed in almost every category. Nearly all genes were up-regulated in the mutant strain; only 6 of them were down-regulated. Of the 117 up-regulated genes, 13, 10, and 2 were involved in 'energy metabolism', 'cellular processes', and ´transport and binding proteins´, respectively and significant changes of transcription occurred in 42 hypothetical proteins (see Figure 4.4A).

For the anaerobic growth condition, 172 genes passed the same statistical analysis and only 9 of them were down-regulated. 14, 14, and 11 of the up-regulated ones were involved in 'energy metabolism', 'cellular processes' and ‘transport and binding proteins’, respectively, and 57 hypothetical proteins showed significant up-regulation in expression (see Figure 4.4B). Moreover, there were 14 genes involved in ‘regulatory function’ under anaerobic growth condition compared to 6 in aerobic growth condition. For aerobic growth with H_2O_2, only 48 genes passed the analysis due to insufficient data and among these genes, 24 of them were hypothetical proteins which displayed significant up-regulation (See Figure 4.4C). As expected, the transcription of the *SO3363* gene was not detected in the mutant cells but expressed in wild type strain for all three growth conditions.

If we take a detailed look at all the up- or down-regulated genes under aerobic and anaerobic growth conditions, these genes can be grouped into several expression patterns based on their log ratio of expression.

Expression pattern I: induced in the mutant strain under both aerobic and anaerobic growth

Among all the genes up-regulated under either aerobic or anaerobic growth conditions, only 9 of them were induced under both tested condition and they are *SO0016* (DNA-3-methyladenine glycosidase I, *tag*), *SO0215* (pantothenate kinase *panK*), *SO0576* (*PhoH* family protein), *SO0758* (hypothetical protein), *SO1111* (bacterioferritin subunit 2 *brf2*), *SO1120* (xanthine/uracil permease family protein), *SO1726* (phosphate transport system regulatory protein *phoU*), *SO3045* (hypothetical protein) and *SO3507* (conserved hypothetical protein). In E. coli, DNA-3-methyladenine glycosidase I is involved in the hydrolysis of alkylated DNA, releasing 3-methyladenine [116, 117]; *panK* is the key enzyme in CoA biosynthetic pathway, transferring a phosphoryl group from ATP to pantothenate (Pan) [118]; *brf* is an iron-storage protein, containing multiple subunits [119]; *phoU* participates in phosphate transport and in the regulatory role of the phosphate-specific transport system [120, 121].

Expression pattern II: repressed in the mutant strain under both aerobic and anaerobic growth

Only 6 genes passed statistical analysis were repressed under aerobic condition and 9 were repressed under anaerobic condition. Among these repressed genes, 5 genes were repressed under both conditions: *SO1066* (extracellular nuclease), *SO3363* (due to deletion), *SO3545* (*OmpA* family protein), *SO3705* (5-methylthioadenosine nucleosidase) and *SO3706* (*NupC* family protein) were repressed under both tested condition. *SO3064*

(amidophosphoribosyltransferase *purF*) was only repressed under aerobic condition and *SO2301* (ribosomal protein L35 *rpml*), *SO2348* (hypothetical protein), *SO4343* (aminotransferase, class V) and *SO3613* (phosphoribosylglycinamide formyltransferase 2 *purT*) were only repressed under anaerobic condition. *purF* and *purT* belong to Pur regulon in E. coli, involving in purine nucleotide de novo biosynthesis [122, 123]; 5'-methylthioadenosine nucleosidase/S-adenosylhomocysteine nucleosidase (*Pfs*) catalyzes the hydrolysis of 5'-methylthioadenosine (MTA) to 5'-methylthioribose (MTR) and S-adenosylhomocysteine (SAH) to S-ribosylhomocysteine (SRH) in prokaryotes[124].

Expression pattern III: induced in the mutant strain under aerobic growth but unaffected under anaerobic growth

There were 108 genes induced in the mutant strain under aerobic growth but unaffected under anaerobic growth. Among the 13 genes involved in energy metabolism under aerobic growth, *SO4357* (anaerobic dimethyl sulfoxide reductase, B subunit, *dmsb-2*), *SO4514* (formate dehydrogenase, iron-sulfur subunit, *fdhB-2*), *SO4620* (fumarate reductase, flavoprotein subunit precursor, *ifcA-2*) are induced in mutant strain. In E. coli, both *dmsB* and *fdhB* belong to the family of bacterial complex [Fe-S]-molybdoenzymes and they function under anaerobic growth condition [125, 126, 127]; *ifcA* is synthesized when Fe(III) is present as the sole respiratory electron acceptor or when it is present in combination with oxygen, fumarate, or nitrate [128]. *SO3363* appears to repress expression of a number of genes encoding regulatory proteins. These included *SO0490* (transcriptional regulator), *SO0529* (*trpBA* operon transcriptional activator, *trpi*), *SO0622* (DNA-binding response regulator), *SO1687* (transcriptional regulator, *MerR* family) and *SO4487* (DNA-binding response regulator). Among these genes, only *trpi* has been

defined. In *Pseudomonas aeruginosa*, the product of the *trpi* gene positively regulates *trpB* and *trpA* which encode the subunits of tryptophan synthase and negatively regulates itself [129]

Expression pattern IV: induced in the mutant strain under anaerobic growth but unaffected under aerobic growth

A total of 153 genes were grouped into this pattern. Among the 14 energy metabolism genes, the most notable observation was that four members of *nqrA-F* operon encoding NADH:ubiquinone oxidoreductase, Na^+ translocating were induced in the mutant strain. Studies in *V. cholerae* and some other marine bacteria revealed that Na^+-translocating NADH:quinone oxidoreductase (Na^+-NQR) [130], with similar role as the NADH:quinone oxidoreductase complex I comprising 14 subunits *nuoA-N* in *E. coli* [131], is the primary respiratory complex that couples the exergonic oxidation of NADH by quinone to the transport of Na^+ across the membrane and the Na^+-NQR in these bacteria maintains an electrochemical Na^+ gradient across the inner bacterial membrane. So, these genes are of the utmost importance for energy conservation in cells. It was also reported that this Na^+ pump system is dependent on respiration and has high activity at alkaline pH [132] and thus, the deletion of *SO3363* gene may cause the cell living in an alkaline environment which induces the expression of *nqrA-F* operon to extrude Na^+ out of the membrane. Among those induced genes belonging to this expression pattern involved in cellular processes, three of them encode members of methyl-accepting chemotaxis group (*SO2083*, *SO2023* and *SO4454*); one encodes the chemotaxis proein (*SO2326* (*cheD*-2)) and one encodes protein-glutamate methylesterase *che*B-3 (*SO3206*). It is known that the methyl-accepting chemotaxis protein (MCP) is required for cells to respond to changes in chemical concentration such as serine (MCP-I), aspartate and maltose (MCP-II), and ribose and galactose (MCP-III), etc. [133]. Regulation of the level of MCP methylation is required for cells

to adapt to those changes and they are associated with the control of the direction of flagellar rotation with cell's migration [134]. In *E.coli*, the methylation reactions can be catalyzed by a specific methyltransferase *cheR* [135], which transfers the methyl group from S-adenosylmethionine to glutamyl residues to form glutamyl methyl esters [136] by a demethylase *che*B [137]. *SO2567* encoding S-adenosylmethionine methyltransferase involved in the biosynthesis of cofactors was also induced. All of this information suggests that deletion of *SO3363* may cause the change of some biochemical concentration under anaerobic growth which induced the expression of these genes in order for cells to adapt to the changes. Another noticeable observation was that some genes encoding transport binding proteins (14) and regulatory proteins (14) were induced under anaerobic growth and most of them are not defined. One of the regulatory proteins, *TyrR* is identified in *Escherichia coli* where it regulates the expression of genes for aromatic amino acid uptake and biosynthesis [138]. Fold changes of *nqr* operon and the methyl-accepting chemotaxis proteins (*MCP*) are listed in Table 4.4

Table 4.4 Expression changes of *nqr* operon genes and *MCP* genes in *SO3363* mutant compared to MR-1 wild type under –O2 growth

Gene symbol[a]	Gene ID	Fold change[b]	P value[c]	Annotation[d]
nqrA-1	SO0902	2.5	0.01	NADH:ubiquinone oxidoreductase, Na translocating, alpha subunit
nqrB-1	SO0903	1.8	0.02	NADH:ubiquinone oxidoreductase, Na translocating, hydrophobic membrane protein NqrB
nqrC-1	SO0904	1.8	0.007	NADH:ubiquinone oxidoreductase, Na translocating, gamma subunit

nqrF-1	SO0907	1.8	0.006	NADH:ubiquinone oxidoreductase, Na translocating, beta subunit
cheB-3	SO3206	1.8	0.05	protein-glutamate methylesterase CheB
cheD-2	SO2326	1.6	0.03	chemotaxis protein CheD, putative
	SO2083	1.8	0.004	methyl-accepting chemotaxis protein
	SO2323	1.7	0.04	methyl-accepting chemotaxis protein
	SO4454	4.5	0.01	methyl-accepting chemotaxis protein

[a] Gene symbol were from TIGR .

[b] Fold change was calculated based on the Log_2 ratio of expression between *SO3363* mutant and wild type strains.

[c] P-value was generated from SAS program.

[d] Protein annotations are derived from TIGR *S. oneidensis* MR-1 genome database.

Expression pattern V: induced in the mutant strain under H_2O_2 growth

For all the selected up-regulated genes in the mutant under H_2O_2 growth condition, three of them were involved in central intermediary metabolism and they encode adenylylsulfate kinase (*SO3723*, *cysC*), sulfate adenylyltransferase, subunit 1 (*SO3726*, *cysD*) and sulfate adenylyltransferase, subunit 2 (*SO3727*, *cysN*). There were another three genes involved in the transport and binding proteins category and they include sulfate ABC transporter, periplasmic sulfate-binding protein (*SO4652*, *sbp*), sulfate ABC transporter, permease protein, sulfate ABC transporter (*SO4654*, *cysW*) , and ATP-binding protein (*SO4655*, *cysA*). In *E. coli*, all these genes belong to the *cys* regulon directing cysteine biosynthesis. The *cys* regulon in *E. coli* consists of over 10 genes which are positively regulated by the *cysB* transcriptional regulator which negatively auto-regulates its own promoter under sulfur limitation [139] (see Fig. 4.5) and the

other *cys* genes other than those we mentioned above were mildly induced under H_2O_2 condition in *S. oneidensis* MR-1, which included *SO3736* (phosphoadenosine phosphosulfate reductase, *cysH)*, *SO3737* (sulfite reductase (NADPH) hemoprotein beta-component*, cysI)*, SO3738 (sulfite reductase (NADPH) flavoprotein alpha-component ,*cysJ*), *SO2903* (cysteine synthase A, *cysK*), *SO3598* (cysteine synthase B, *cysM*) (see Table 4.3). The expression of *cysK*, which synthesizes cysteine form O-acetylserine and sulfide, has been reported to be a general bacterial response to hydrogen peroxide [140]. As expected, *SO2649* (*cys* regulon transcriptional activator, *cysB*) was repressed under tested condition in the mutant and the expression of *SO2262* (serine acetyltransferase, *cysE*) (see Table 4.5) did not change much in the mutant since its expression is not affected by sulfur availability. Another noticeable observation was that some hypothetical proteins were induced in deletion mutant under H_2O_2 growth, including *SO2997-98* and *SO2999-3003*. *SO2997-98* are in the same operon and *SO2999-3006* are in the same operon.

My hypothesis is that the mutant strain grown under H_2O_2 condition probably enhanced the cellular demand for cysteine which may be needed to repair oxidatively damaged iron-sulfur cluster proteins with crucial role in metabolism.

Table 4.5 Expression changes of *cys* regulon genes in *SO3363* mutant compared to MR-1 wild type under H_2O_2 growth

Gene symbol[a]	Gene ID	Fold change[b]	P value[c]	Annotation[d]
cysA	SO4655	2.2	0.03	sulfate ABC transporter, ATP-binding protein
cysB	SO2649	-0.2	0.3	cys regulon transcriptional activator
cysC	SO3723	1.7	0.008	adenylylsulfate kinase

cysE	SO2262	0.3	0.04	serine acetyltransferase
cysD	SO3726	2.0	0.006	sulfate adenylyltransferase, subunit 1
cysN	SO3727	1.7	0.001	sulfate adenylyltransferase, subunit 2
cysH	SO3736	1.2	0.007	phosphoadenosine phosphosulfate reductase
cysI	SO3737	1.3	0.008	sulfite reductase (NADPH) hemoprotein beta-component
cysJ	SO3738	1.5	0.004	sulfite reductase (NADPH) flavoprotein alpha-component
cysK	SO2903	0.9	0.08	cysteine synthase A
cysM	SO3598	1.1	0.01	cysteine synthase B
sbp	SO4652	2.3	0.03	sulfate ABC transporter, periplasmic sulfate-binding protein
cysT	SO4653	0.98	0.27	sulfate ABC transporter, permease protein
cysW	SO4654	2.2	0.03	sulfate ABC transporter, permease protein

[a] Gene symbol were from TIGR .

[b] Fold change was calculated based on the Log_2 ratio of expression between SO3363 mutant and wild type strains.

[c] P-value was generated from SAS program.

[d] Protein annotations are derived from TIGR *S. oneidensis* MR-1 genome database.

Of the five down-regulated genes under H_2O_2 growth, which passed the statistical analysis ($p<0.05$), two genes involved in purine, pyrimidine, nucleoside, and nucleotide

metabolism were *SO2001* (5-nucleotidase, *ushA*) and SO*0442* (phosphoribosylaminoimidazolecarboxamide formyltransferase, *purH*) and the other three down-regulated genes were *SO3705*, *SO3706* and *SO3363*, which were also down-regulated under aerobic and anaerobic growth condition. In *E. coli*, *purH* belongs to the *pur* regulon and it is involved in de novo purine biosynthesis; *ushA* can hydrolyse UDP-sugar into uridine, sugar-1 phosphate, and orthophosphate and release inorganic phosphate (Pi) from UDP (uridine 59-diphosphate)-sugar [141, 142]. Besides the down-regulation of *purH*, the other pur regulon members, *purD*, *purE*, *purH*, *purK* were slightly repressed in mutant under H_2O_2 growth (Table 4.6).

Table 4.6 Expression changes of *pur* regulon genes in *SO3363* mutant compared to MR-1 wild type under H_2O_2 growth

Gene symbol[a]	Gene ID	Fold change[b]	P value[c]	Annotation[d]
purD	SO0441	-1.4	0.01	phosphoribosylamine--glycine ligase
purE	SO3554	-1.5	0.006	phosphoribosylaminoimidazole carboxylase, catalytic subunit
purF	SO3064	-1.4	0.03	amidophosphoribosyltransferase
purH	SO0442	-1.6	0.003	phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase
purK	SO3555	-1.2	0.03	phosphoribosylaminoimidazole carboxylase, ATPase subunit
purT	SO3613	-1.3	0.009	phosphoribosylglycinamide formyltransferase 2

[a]Gene symbol were from TIGR .

[b]Fold change was calculated based on the Log_2 ratio of expression between SO3363 mutant and wild type.

[c]P-value was generated from SAS program.

[d]Protein annotations are derived from TIGR *S. oneidensis* MR-1 genome database.

4.5 Some induced or repressed genes containing the SELEX motif possibly directly regulated by *SO3363*

In section 4.4, microarray data analyses showed that a number of genes were significantly induced or repressed under three tested conditions and it is important to know whether the upstream region of these genes contain the SELEX binding motif or not. Under aerobic growth condition, there were about 347 ORFs showed more than 2-fold up-regulation ($p<0.05$) and 20 of them contain the $AACn_4GTT$ motif; under anaerobic growth condition, there were 621 ORFs displayed more 2-fold up-regulation and 48 of them contain the $AACn_4GTT$ motif; under H_2O_2 growth condition, 105 ORFs showed more than 2-fold up-regulation and 14 of them contain the $AACn_4GTT$ motif (see Fig. 4.6 and Appendix, Table B).

Among those induced genes under anaerobic growth condition, *SO4404*, an iron sulfur binding protein, containing the $AACn_4GTT$ binding motif, was up-regulated 3.5-fold and *SO4454*, an methyl-accepting chemotaxis protein, containing $AACn_4GTT$ binding motif, was up-regulated 22.7-fold; among those induced genes under H_2O_2 growth condition, *SO4652* (*sbp* in *cys* regulon), an sulfate ABC transporter, containing the extended motif $GTn_{10}GAAC$, was up-regulated 4.9-fold and five hypothetical proteins, *SO2995*, *SO2997*, *SO2998*, *SO3003* and *SO3005*, containing the $AACn_4GTT$ motif, were up-regulated 4.0-fold, 4.0-fold, 3.2-fold, 3.2-

fold, 2.8-fold, and 2.8-fold, respectively. Among those repressed genes, *SO3555* (*purK*), encoding phosphoribosylaminoimidazole carboxylase, ATPase subunit, containing the $AACn_4GTT$ motif, was down-regulated 2.3-fold under H_2O_2 growth condition and *SO3705*, encoding 5-methylthioadenosine nucleosidase, containing the extended motif $GTn_{10}GAAC$, was down-regulated 10-fold under aerobic growth, 24-fold under anaerobic growth and 16-fold under H_2O_2 growth condition. The fold change of these selected genes is listed in Table 4.7.

Table 4.7 Induced or repressed genes in *SO3363* mutant compared to MR-1 wild type containing SELEX motif under tested conditions

Gene ID	SELEX motif[a]	Fold change[b]	Annotation[c]
anaerobic growth, up-regulation			
SO4404	AACATGAGTT	3.5	iron sulfur cluster binding protein
SO4454	AACCAAGGTT	22.7	methyl-accepting chemotaxis protein
H_2O_2 growth, up-regulation			
SO2995	AACGATGGTT	4.0	hypothetical protein
SO2997	AACAGCAGTT	4.0	hypothetical protein
SO2998	AACTGCTGTT	3.2	hypothetical protein
SO3003	AACATGCGTT	2.8	hypothetical protein
SO3005	AACCGATGTT	2.8	hypothetical protein
SO4652	GTATTGAGTTTGGAAC	4.9	sulfate ABC transporter, periplasmic sulfate-binding protein
H_2O_2 growth, down-regulation			
SO3555	AACGCCGGTT	2.3	phosphoribosylaminoimidazole carboxylase, ATPase subunit

SO3705	GTCCATCATCTTGAAC	16	5-methylthioadenosine nucleosidase/S-adenosylhomocysteine nucleosidase

[a] $AACn_4GTT$ or $GTn_{10}GAAC$ motif.

[b] Fold change was calculated based on the absolute ratio of expression between SO3363 mutant and wild type.

[c] Protein annotations are derived from TIGR *S. oneidensis* MR-1 genome database.

Figure 4.1 Confirmation of mutant construction by colony PCR. A. Colony PCR of conjugation result. Left: colony PCR with LF/SR primers; Right: colony PCR with LR/SF primers; correct insertions were labeled with numbers in red. B. Colony PCR of sucrose-counter selection with LF/LR primers. PCR from the correct colonies were labeled with numbers in red.

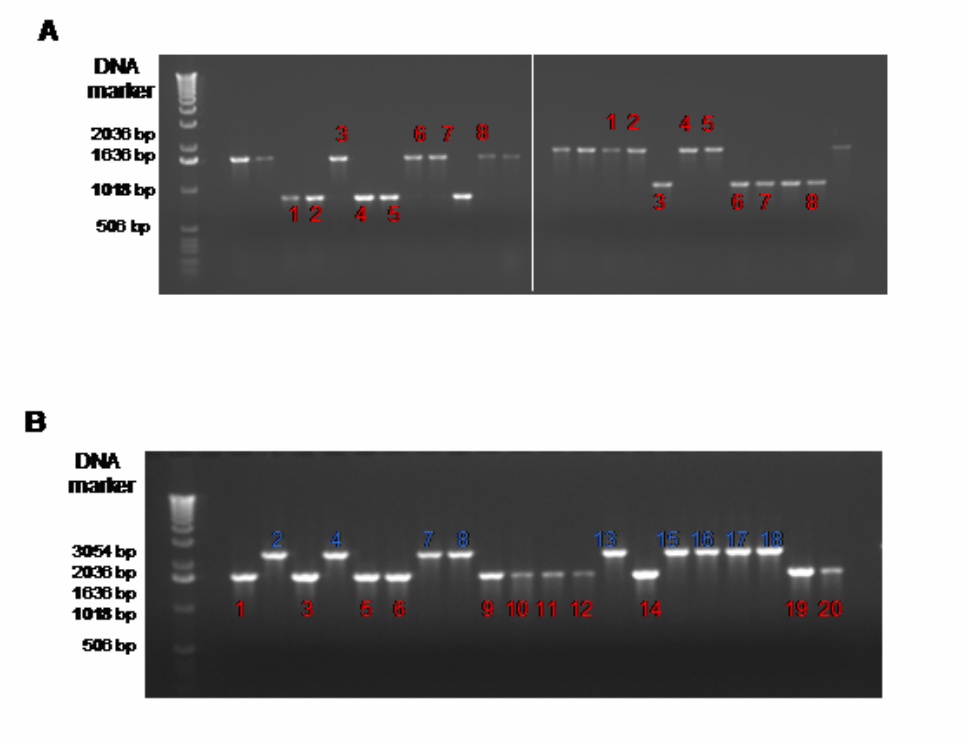
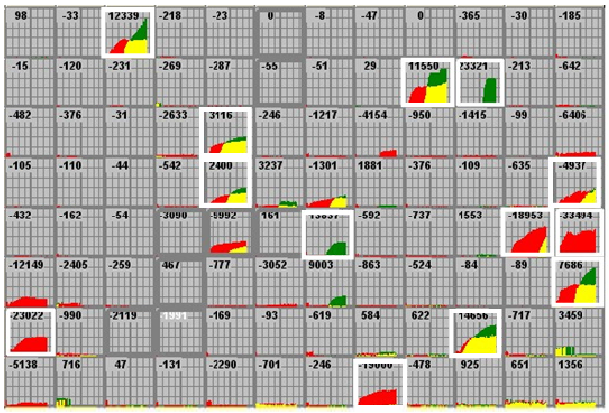
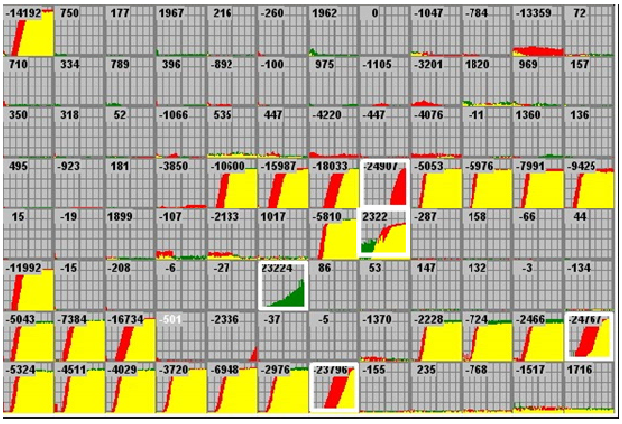

A
DNA marker
2036 bp
1636 bp
1018 bp
506 bp
1 2 3 4 5 6 7 8
1 2 3 4 5 6 7 8
B
DNA marker
3054 bp
2036 bp
1636 bp
1018 bp
506 bp
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Figure 4.2 Phenotypic changes in PM assays. Each box represents development of dye color as a function of growth time for a specific growth condition. Significant changes are enclosed in boxes. Yellow indicates that growth of the wild type and growth of the mutant were similar. Red indicates faster growth of the wild type. Green indicates faster growth of the mutant.

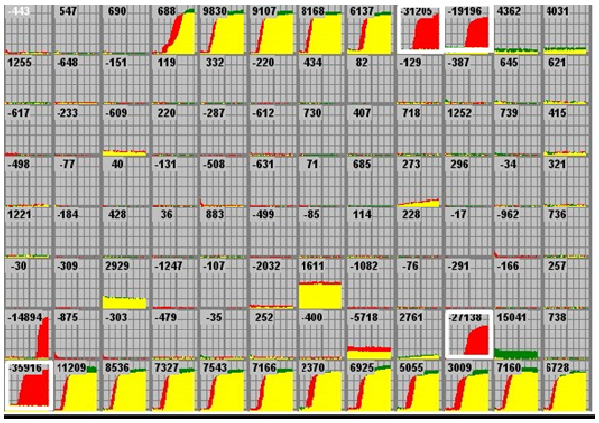


PMM1 Carbon Source

PMM2 Carbon Source

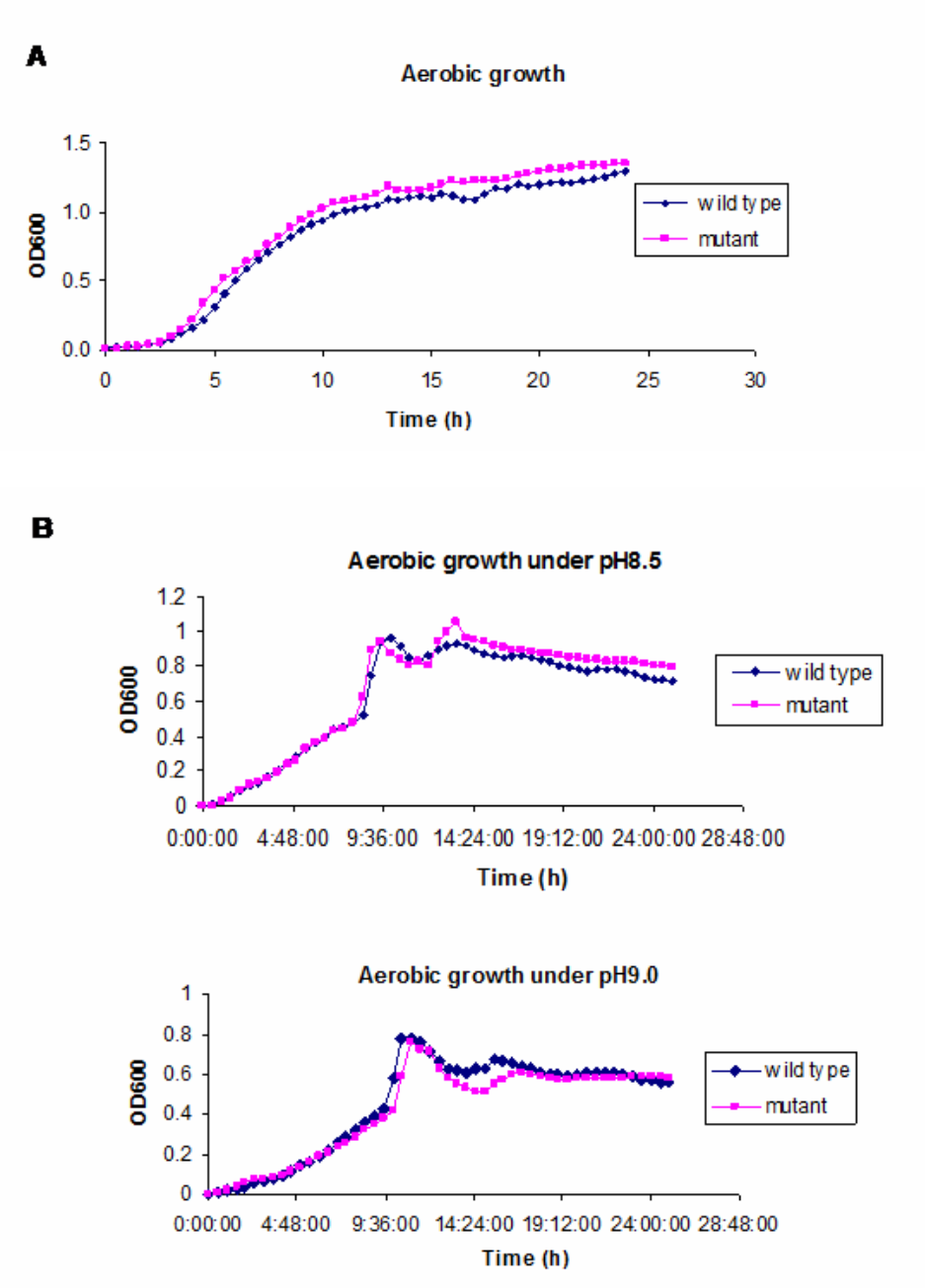


PMM9 Osmolytes

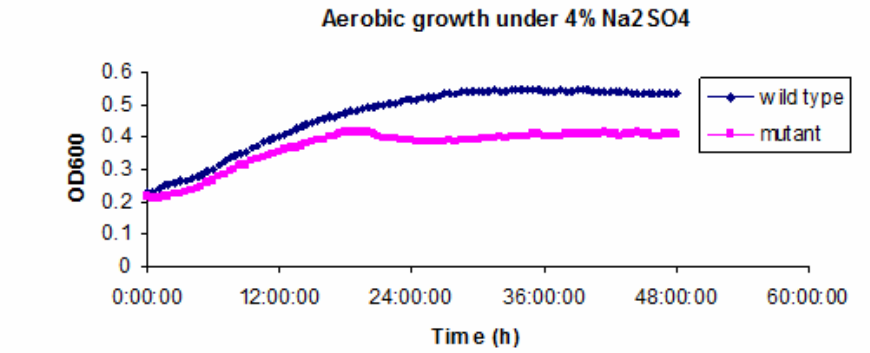


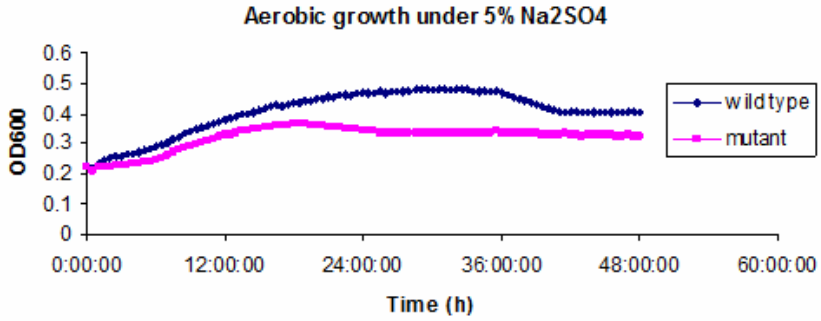
PMM10 PH

Figure 4.3 Comparisons of growth of the SO3363 mutant and wild type strains under selected conditions. A. Aerobic growth; B. Alkaline growth at pH 8.5 and 9.0; C. Aerobic growth under 4% and 5% sodium sulfate; D. Aerobic growth under 0.6 mM and 0.8 mM H_2O_2; E. Anaerobic growth under 20 mM sodium lactate and 10 mM sodium fumarate. Wild type growth curve was recorded in blue (diamonds) and the mutant growth curve was recorded in pink (squares).

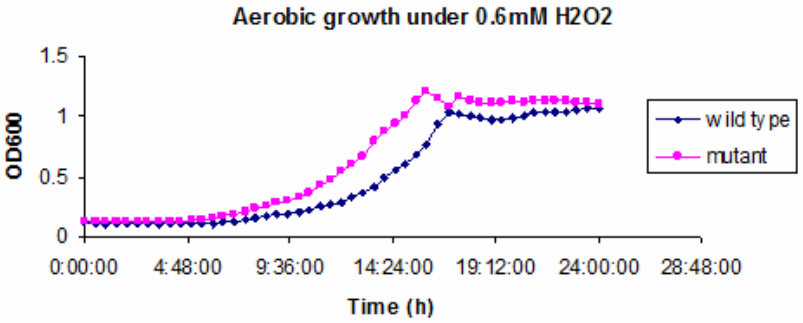

A
Aerobic growth
OD600
1.5
1.0
0.5
0.0
0
5
10
15
20
25
30
Time (h)
wild type
mutant
B
Aerobic growth under pH8.5
OD600
1.2
1
0.8
0.6
0.4
0.2
0
0:00:00
4:48:00
9:36:00
14:24:00
19:12:00
24:00:00
28:48:00
Time (h)
wild type
mutant
Aerobic growth under pH9.0
OD600
1
0.8
0.6
0.4
0.2
0
0:00:00
4:48:00
9:36:00
14:24:00
19:12:00
24:00:00
28:48:00
Time (h)
wild type
mutant

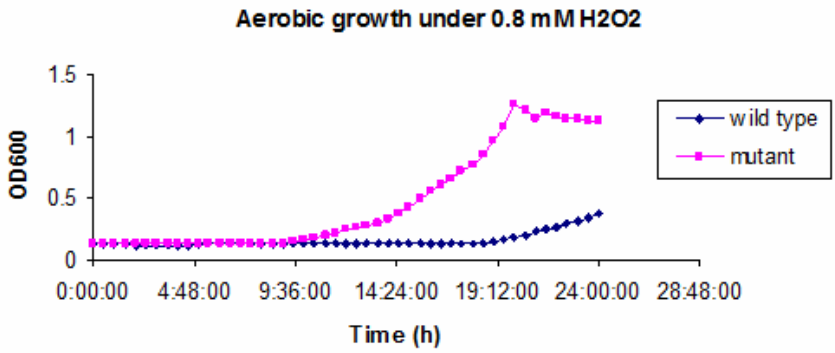
C


Aerobic growth under 4% Na2SO4
OD600
0.6
0.5
0.4
0.3
0.2
0.1
0
0:00:00
12:00:00
24:00:00
36:00:00
48:00:00
60:00:00
Time (h)
wild type
mutant


Aerobic growth under 5% Na2SO4
OD600
0.6
0.5
0.4
0.3
0.2
0.1
0
0:00:00
12:00:00
24:00:00
36:00:00
48:00:00
60:00:00
Time (h)
wild type
mutant

D


Aerobic growth under 0.6mM H2O2
OD600
1.5
1
0.5
0
0:00:00
4:48:00
9:36:00
14:24:00
19:12:00
24:00:00
28:48:00
Time (h)
wild type
mutant


Aerobic growth under 0.8 mM H2O2
OD600
1.5
1
0.5
0
0:00:00
4:48:00
9:36:00
14:24:00
19:12:00
24:00:00
28:48:00
Time (h)
wild type
mutant

E

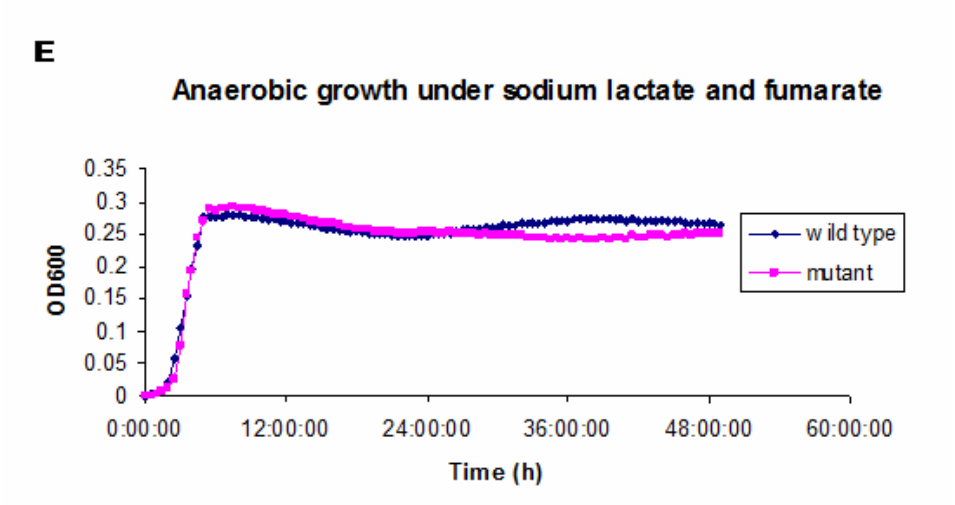
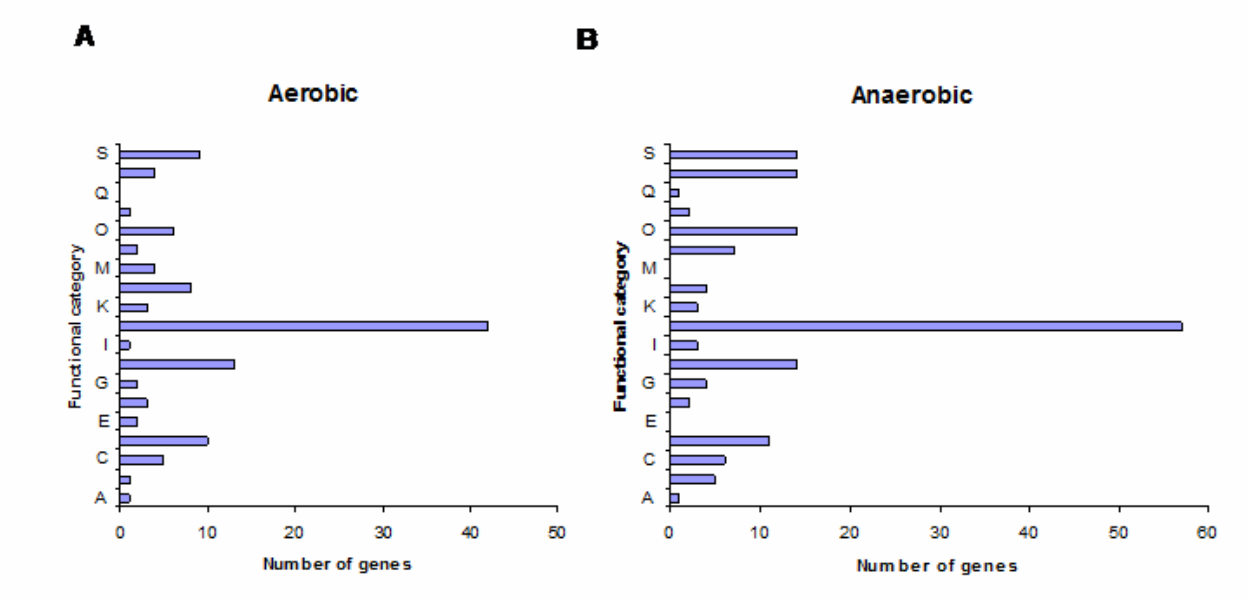

Anaerobic growth under sodium lactate and fumarate
OD600
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0:00:00
12:00:00
24:00:00
36:00:00
48:00:00
60:00:00
Time (h)
wild type
mutant

Figure 4.4 Differentially expressed genes under tested conditions grouped by functional classification according to the TIGR *S. oneidensis* MR-1 genome database. A. Aerobic growth. B. Anareobic growth. C. H_2O_2 growth. (Categories are defined as: A, Amino acid biosynthesis; B, Biosynthesis of cofactors, prosthetic groups, and carriers; C, Cell envelope; D. Cellular processes; E, Central intermediary metabolism; F, Disrupted reading frame; G, DNA metabolism; H, Energy metabolism; I, Fatty acid and phospholipids metabolism; J, Hypothetical proteins; K, Mobile and extrachromosomal element functions; L, Protein fate; M, Protein synthesis; N, Purines, pyrimidines, nucleosides, and nucleotides; O, Regulatory functions; P, Signal transduction; Q, Transcription; R, Transport and binding proteins; S, Unknown function.).


A
Aerobic
Functional category
S
Q
O
M
K
I
G
E
C
A
0
10
20
30
40
50
Number of genes
B
Anaerobic
Functional category
S
Q
O
M
K
I
G
E
C
A
0
10
20
30
40
50
60
Number of genes

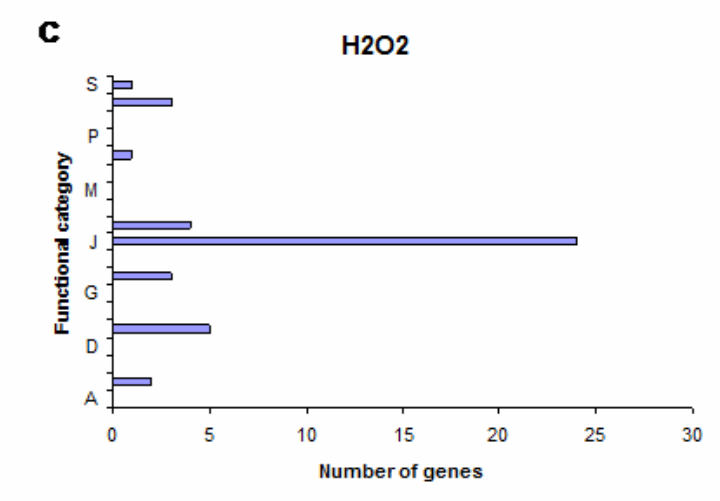

C
H2O2
Functional category
S
P
M
J
G
D
A
0
5
10
15
20
25
30
Number of genes

Figure 4.5 The pathway of cysteine biosynthesis in *E. coli* (adapted from [139]). Cysteine can be synthesized from either sulfate (left) or thiosulphate (right). Sulphate is transported into the cell by periplasmic protein *sbp* and reduced to sulphide. Then, it reacts with O-acetylserine to form cysteine. Thiosulphate is transported into the cell by another periplasmic protein *cysP* and reacts with O-acetylserine to form S-sulphocysteine, which is then reduced to cysteine. The sulphate and thiosulphate transport systems share three components encoded by *cysT.* cysW and *cysA.* O-Acetylserine (thiol)-lyases A and B are encoded by *cysK and cysM*, respectiveiy, and can both catalyze the reaction with sulphide but only the B isoenzyme catalyzes the reaction with thiosulphate. The abbreviations are: APS, adenosine 5'-phosphosulphate; PAPS, 3'-phosphoadenosine 5'-phosphosulphate.

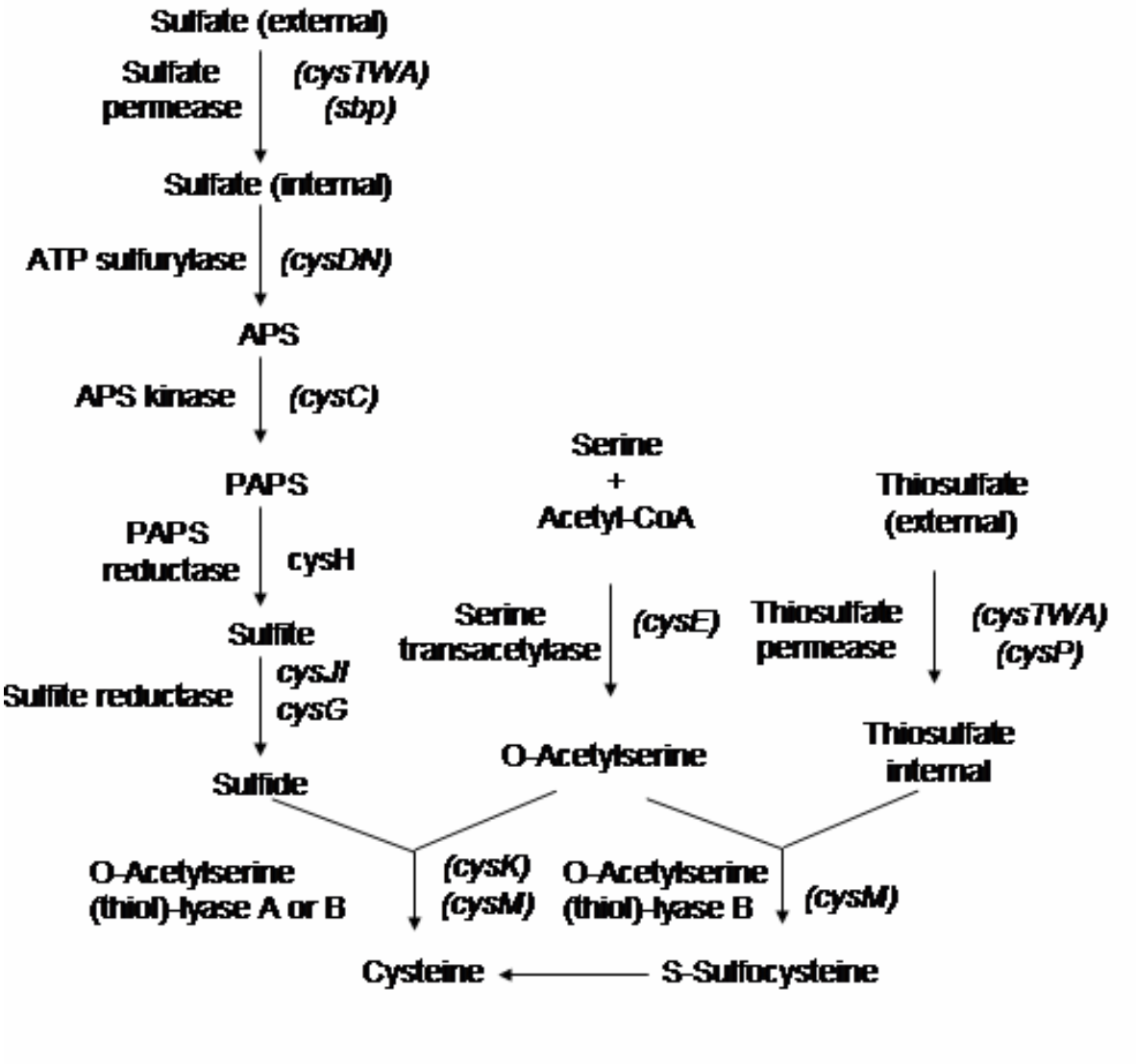
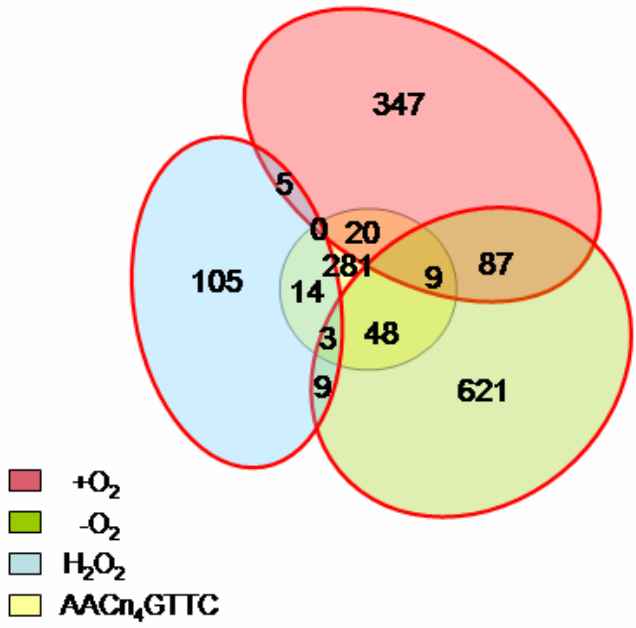

Sulfate (external)
Sulfate permease
(cysTWA)
(sbp)
Sulfate (internal)
ATP sulfurylase
(cysDN)
APS
APS kinase
(cysC)
PAPS
PAPS reductase
cysH
Sulfite
Sulfite reductase
cysJI
cysG
Sulfide
Serine
+
Acetyl-CoA
Serine transacetylase
(cysE)
O-Acetylserine
Thiosulfate (external)
Thiosulfate permease
(cysTWA)
(cysP)
Thiosulfate internal
O-Acetylserine (thiol)-lyase A or B
(cysK)
(cysM)
O-Acetylserine (thiol)-lyase B
(cysM)
Cysteine
S-Sulfocysteine

Figure 4.6 Venn diagram of the number of induced genes under the three tested conditions containing SELEX AACn$_4$GTT motif. The yellow circle in the center represents 281 UORs in Shewanell oneidensis MR-1 genome contain AACn$_4$GTT motif; the other circles represent the genes induced under different growth conditions and the intersection of these circles and the yellow circle represents the number of induced gene under each growth condition containing the AACn$_4$GTT SELEX motif.


347
5
0
20
281
87
105
14
9
3
48
9
621
+O2
-O2
H2O2
AACn4GTTC

CHAPTER 5

CONCLUSION

Shewanella oneidensis MR-1 is a facultative gram-negative γ-proteobacterium. It is capable of coupling the oxidation of organic carbon to a wide range of electron acceptors such as oxygen, nitrate and metal oxides, and has potential for bioremediation of heavy metal contaminated sites. However, the genetic basis and regulatory mechanisms underlying the ability of *S. oneidensis* to survive and adapt to various environmentally relevant stresses is poorly understood. With the availability of the sequenced 5-Mb genome of *S. oneidensis* MR-1, it is possible to study how it utilizes different transcription factors to control gene expression to adapt to a given environmental alteration.

In this study, an integrated approach coupling genomics, proteomics, and bioinformatics was used to discover and predict regulatory transcription factors (rTFs), their DNA binding sites, and the genes they regulate, associated with a given environmental alteration (e.g. aerobic growth shifted to anaerobic growth in this case).

SO3363p, encoding a LysR family transcription factor, was discovered in the DNA-affinity capture experiment and it showed different binding affinity with *SO0060*, *SO1427* promoters that were selected, based on the microarray expression profiling of the two growth conditions (aerobic and anaerobic growth), as the bait in the capture experiment. *SO3363p*'s recognition motif was determined by SELEX using a pool of synthetic DNA containing a central 25-bp randomized sequence. The consensus DNA sequence $GAACn_4GTT$ (Fig. 3.8A) was discovered in all 31 sequences and $GTn10GAACn_4GTTC$ (Fig. 3.8B) was present in the majority

of sequences. A 71-bp synthetic DNA with or without the consensus sequence was used to determine the importance of the presence of these palindromic sequences for the specific binding to *SO3363p.* The result showed that *SO3363p* binds specifically to the 71-bp probe containing the consensus sequence.

In order understand the role of *SO3363* in *S. oneidensis*, an *SO3363* deletion strain was constructed and subjected to both physiological characterization and microarray analysis. Compared to the wild-type MR-1, the mutant exhibited a defect in utilizing uridine, 2-deoxy adenosine, adenosine, gelatin, pyruvic acid, as well as higher pH. The mutant grows slightly better in α-hydroxy butyric acid; Bioscreening experiments showed that the mutant is more resistant to H_2O_2 while there were no growth differences between the wild type and the mutant strains under aerobic growth or anaerobic growth; Microarray analyses showed that a significant number of genes (>2-fold, $p<0.05$) were induced in the mutant under the three tested conditions. For anaerobic growth, it is noticeable that four members of the *nqrA-F* operon encoding NADH:ubiquinone oxidoreductase, Na^+ translocating were induced in the mutant strain and this observation may indicated that the deletion strain grown under anaerobic condition may cause a change in pH, which will lead to the expression of *nqr* operon to transport Na^+. Three chemotaxis proteins were also induced under anaerobic condition, including the highly up-regulated *SO4454* which contains $AACn_4GTT$ motif. This suggested that the deletion of *SO3363* may cause a change in some chemical concentration under anaerobic growth condition. For aerobic growth in the presence of H_2O_2 condition, the *cys* regulon members was induced, including *SO4652* (*sbp*) containing (the extended $GTn_{10}GAAC$ motif). This implied that the deletion of *SO3363* may demand the synthesis of cysteine controlled by the *cys* regulon to counteract the H_2O_2. Five hypothetical proteins containing the $AACn_4GTT$ motif were also

induced with anaerobic growth and the relationships of these up-regulated hypothetical proteins and *SO3363* need to be further investigated. For all the down-regulated genes under the conditions tested, *SO3705* and *SO3706* are in the same operon and both of them were highly down-regulated under the three tested conditions. *SO3705* contains the extended $GTn_{10}GAAC$motif and therefore, it is possible that this gene is directly regulated by *SO3363p*. *pur* regulon members involved in *de novo* synthesis of purine were repressed in the presence of H_2O_2 and one of the members, *SO3555*, contains the $AACn_4GTT$ motif.

The repression of *SO3705* encoding nucleosidase in the mutant provides a reasonable explanation of one of the PM results that the mutant could not grow in the presence of adenosine, 2-deoxy adenosine, and uridine. When the nucleosidase is repressed, nucleosides cannot be hydrolyzed and then utilized by cells and thus the presence of extra nucleoside including adenosine and uridine will inhibit cell growth.

There is still a lot that needs to be done in order to better understand this protein. First, the recognition motif of *SO3363p* was determined by SELEX, however, it is still not known which palindrome is responsible for the specific binding. Thus, another two mutated DNA probes in which only one palindrome is mutated should be used to perform EMSA experiments in the presence of heparin; second, in order to verify whether this protein tends to bind shorter DNA containing the motif *in vitro*, shorter DNA probes selected from the UOR database search results should be designed to perform EMSA experiments in the presence of heparin; *In vitro* transcription experiments need to be performed in order to verify which induced or repressed gene containing the SELEX motif is regulated directly by this protein. Further investigation of those induced or repressed genes under tested conditions is needed in order to understand the regulation pathway associated with the specific growth condition and *SO3363p*.

REFERENCES

1. Brown, J. R., and Volker C. (2004). Phylogeny of g-proteobacteria: resolution of one branch of the universal tree?. *BioEssays* **26**:463–468.
2. Madigan, M; and Martinko, J (2005). *Brock Biology of Microorganisms*, 11th ed.
3. Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America* **74** (11), 5088-5090.
4. Discover life: tree of life.
5. Cramer, P. (2002). Multisubunit RNA polymerase. *Current opinion in Structural Biology* **12**, 89-97.
6. Borukhov, S., and Severinov, K. (2002). Role of the RNA polymerase sigma subunit in transcription initiation. *Research in Microbiology* **153**, 557-562.
7. Mitchell D., Probasco, N. E., and Thompson, R. B. (2007). Immunoaffinity purification and characterization of RNA polymerase from *Shewanella oneidensis*. *Protein expression and purification* **55**, 23-30.
8. Zhao, K., Liu, M., and Burges, R. R. (2005). The global transcriptional response of Escherichia coli to induced σ^{32} protein involves σ^{32} regulon activation followed by inactivation and degradation of σ^{32} in vivo. *Journal of Biological Chemistry* **280**, 17758-17768.

9. Kornberg, R. D. (2007). The molecular basis of eukaryotic transcription. . *Proceedings of the National Academy of Sciences of the United States of America* **104**, 12955-12961.
10. Zhou, D., and Yang, R. (2006). Global analysis of gene transcription regulation in prokaryotes. *Cellular and Molecular Life Sciences*, **63**, 2260-2290.
11. Browning, D. F., and Busby, S. W. (2004). The regulation of bacterial transcription initiation. *Nature Reviews Microbiology* **2**, 1-9.
12. Rojo, F. (1999). Repression of transcription initiation in bacteria. *Journal of Bacteriology* **181** (10), 2987-2991.
13. Busby, S., and Ebright, R. E. (1994). Promoter structure, promoter recognition and transcription activation in prokaryotes. *Cell* **79,** 743–746.
14. Mooney, R. A., I. Artsimovitch, and R. Landick. (1997). Information processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation. *Journal of Bacteriology* **180,** 3265–3275.
15. Henkin, T. M. (1997). Control of transcription termination in prokaryotes. *Annual Review of Genetics* **30,** 35–57.
16. Goldfarb, A., and Palm, P. (1981). Control of promoter utilization by bacteriophage T4-induced modification of RNA polymerase α subunit. *Nucleic Acids Research* **9** (19), 4863-4878.
17. Wang, X. Z., Zeng, W. K., Murakawa, M., and Freeman, M.W., Seed, B. (2000). Episomal segregation of the adenovirus enhancer sequence by conditional genome rearrangement abrogates late viral gene expression. *Journal of Virology* **74** (23), 11296-11303.

18. Huffman, J. L., and Brennan, R. G. (2002). Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Current Opinion in Structural Biology* **12**, 98-106.

19. Perez-Rueda, E., and Collado-Vides, J. (2002). The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Research* **28**, 1838-1847.

20. Martinez-Antonio, A., and Collado-Vides, J. (2003). Identifying global regulators in transcriptional regulatory networks in bacteria. *Current Opinion in Microbiology* **6**, 482–489.

21. Gralla, J. D. (1996). Activation and repression of E. coli promoters. *Current Opinion in Genetics and Development* **6**, 526-530.

22. Barnard, A., Wolfe, A., and Busby, S. (2004). Regulation at complex bacterial promoters: how bacterial use different promoter organizations to produce different regulatory outcomes. *Current Opinion in Microbiology* **7**, 102-108.

23. Dove S., Darst, S., and Hochschild A. (2003). Region 4 of σ as a target for transcription regulation. *Molecular Microbiology* **48**, 863-874.

24. B rown, N. L., Stoyanov, J. V., Kidd, S. P., and Hobman, J. L. (2003). The MerR family of transcriptional regulators. *FEMS Microbiology Reviews* **27**, 145–163.

25. Porter, S. C., North, A. K., Wedel, A. B., and Kustu, S. (1993). Oligomerization of NTRC at the glnA enhancer is required for transcriptional activation. *Genes and Development* **7**, 2258-2273.

26. Ber trand-Burggraf, E., S. Hurstel, M. Daune., and M. Schnarr. (1987). Promoter properties and negative regulation of the *uvrA* gene by LexA repressor and its amino terminal DNA binding domain. *Journal of Molecular Biology* **193,** 293–302.

27. Ross, W., K. K. Gosink, J. Salomon, K. Igarashi, C. Zou, A. Ishihama, K. Severinov, and R. L. Gourse. (1993). A third recognition element in bacterial promoters: DNA binding by the α subunit of RNA polymerase. *Science* **262,** 1407–1413.

28. Choy, H., and Adhya, S. (1996). *In Escherichia coli and Salmonella: Cellular and Molecular Biology* 1287–1299.

29. Serrano, M., M. Salas., and J. M. Hermoso. (1990). A novel nucleoprotein complex at a replication origin. *Science* **248,** 1012–1016.

30. Valentin-Hansen, P., L. Søgaard-Andersen., and H. Pedersen. (1996). A flexible partnership: the CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control. *Molecular Microbiology*. **20,** 461–466.

31. Summers, A. O. (1992). Untwist and shout: a heavy metal-responsive transcriptional regulator. *Journal of Bacteriology* **174,** 3097–3101.

32. Shin, M. (2001). Repression of *deo*P2 in *Escherichia coli* by CytR: conversion of a transcription activator into a repressor. *The EMBO Journal* **20**, 5392–5399 (2001).

33. Monsalve, M., B. Calles, M. Mencı´a, F. Rojo., and M. Salas**.** (1998). Binding of phage f29 protein p4 to the early A2c promoter: recruitment of a repressor by the RNA polymerase. *Journal of Molecular Biology* **283,** 559–569.

34. Monsalve, M., M. Mencı´a, F. Rojo., and M. Salas. (1996). Activation and repression of transcription at two different phage f29 promoters are mediated by interaction of the same residues of regulatory protein p4 with the RNA polymerase. *The EMBO Journal* **15,** 383–391.

35. Monsalve, M., M. Mencı´a, M. Salas., and F. Rojo. 1996. Protein p4 represses phage f29 A2c promoter by interacting with the α subunit of *Bacillus subtilis* RNA

polymerase. . *Proceedings of the National Academy of Sciences of the United States of America* **93,** 8913–8918.

36. Lopez, P. J., Guillerez, J., Sousa, R., and Dreyfus. M. (1988). On the mechanism of inhibition of phage T7 RNA polymerase by *lac* repressor. *Journal of Molecular Biology* **276,** 861–875.
37. Storz, G., and Imlay, J. A. (1999). Oxidative stress. *Current opinion in Microbiology* **2**, 188-194.
38. Bauer, C. E., Elsen, S., and Bird, T. H. (1999). Mechanism for redox control of gene expression. *Annual Review of Microbiology* 53, 495-523.
39. Hidalgo, E., and Demple, B. (1996). Adaptive responses to oxidative stress: The *soxRS* and *oxyR* regulons. In *Regulation of Gene Expression in Escherichia coli*, ed. ECC Lin, A. Simon-Lynch, pp. 433–50.
40. Sawers, RG., Zalelein, E., and B¨ock A. (1988). Two-dimensional gel electrophoretic analysis of *Escherichia coli* proteins: influence of various anaerobic growth conditions and the *fnr* gene product on cellular protein composition. *Arch. Microbiol* **149**, 240–44.
41. Schell MA. 1993. Molecular biology of LysR family of transcriptional regulators. *Annual Review of Microbiology* **47**, 597-626.
42. Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L., and Storz, G. (1997). A small, stable RNA induced by oxidative stress: Role as a pleiotropic regulator and antimutator. *Cell* **90**, 45-53.

43. Mukhopadhyay, S., and Shellhorn, HE. (1997). Identification and characterization of hydrogen peroxide-sensitive mutants of Escherichia coli; genes that require OxyR for expression. *Journal of Bacteriology* **179**, 330-338.

44. Stewart, V., Parales, J., and Merkel, S. M. (1989). Structure of genes narL and narX of the nar (nitrate reductase) locus in Escherichia coli K-12.*Journal of Bacteriology* **171**, 2229-2234.

45. Allen, J. F. (1993). Redox control of transcription: sensors, response regulators, activators and repressors. *Federation of European Biochemical Societies* **332** (3): 203-207.

46. Sganga, M.W., and Bauer, C. E. (1992). Regulatory factors controlling photosynthetic reaction center and light-harvesting gene expression in *Rhodobacter capsulatus*. *Cell* **68**, 945-954.

47. Gao, H. C., Wang, X. H., Yang, Z. K., Palzkill, T., and Zhou, J.Z. (2008). Probing regulon of ArcA in *Shewanella oneidensis* MR-1 by integrated genomic analyses. *Biomed Central Genomics* **9**, 42-58.

48. Georgellis, D., Kwon, O., and Lin, E.C. (2001) Quinones as the redox signal for the *arc* two- component system of bacteria. *Science* **292,** 2314–2316.

49. Malpica, R., Franco, B., Rodriguez, C., Kwon, O., and Georgellis, D. (2004) Identification of a quinone-sensitive redox switch in the ArcB sensor kinase. . *Proceedings of the National Academy of Sciences of the United States of America* **101,** 13318–13323.

50. Kwon, O., Georgellis, D., and Lin, E.C. (2000) Phosphorelay as the sole physiological route of signal transmission by the arc two-component system of *Escherichia coli*. *Journal of Bacteriology* **182,** 3858–3862.

51. Lynch, A.S., and Lin, E.C.C. (1996a) Responses to molecular oxygen. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*. Neidhardt, F.C., Curtiss, R., III, Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., *et al.* (eds). Washington, DC: American Society for Microbiology, pp. 1526–1538.

52. Jeon, Y., Lee, Y.S., Han, J.S., Kim, J.B., and Hwang, D.S. (2001) Multimerization of phosphorylated and non-phosphorylated ArcA is necessary for the response regulator function of the Arc two-component signal transduction system. *Journal of Biological Chemistry* **276,** 40873–40879.

53. Gilles-Gonzalez, M. A., Ditta, G. S., and Helinski, D. R. (1991). *Nature* **350**, 315-319.

54. Siddavattam, D., Steibl, H-D., Kreutzer, R., Klingmuller, W. (1995). Regulation of *nif* gene expression in *Enterobacter agglomerans*: nucleotide sequence of the *nifLA* operon and influence of temperature and ammonium on its transcription. *Molecular and General Genetics* **249**, 629-636.

55. Derby, H. A., and B. W. Hammer. (1931). Bacteriaology of butter. IV. Bacteriological studies of surface taint butter. *Iowa Agric. Exp. Stn. Res. Bull.* **145,** 387–416.

56. Pagani, L., A. Lang, C. Vedovelli, O. Moling, G. Rimenti, R. Pristera., and P. Mian. (2003). Soft tissue infection and bacteremia caused by *Shewanella putrefaciens.* *Journal of Clinical Microbiology* **41,** 2240–2241.

57. Skerratt, J. H., J. P. Bowman, and P. D. Nichols. (2002). *Shewanella olleyana* sp. nov., a marine species isolate from temperate estuary which produces high levels of polyunsaturated fatty acids. *International Journal of Systematic and Evolutionary Microbiology*. **52,** 2101–2106.

58. Semple, K., D. W. S. Westlake. (1987). Characterization of iron-reducing *Alteromonas putrefaciens* strains from oil field fluids. *Canadian Journal of Microbiology* **33,** 366–371

59. Yamada, M., K. Nakasoni, H. Tamegai, C. Kato, R. Usami., and K. Horikoshi. (2000). Pressure regulation of soluble cytochromes *c* in a deep-sea piezophilic bacterium, *Shewanella violacea*. *Journal of Bacteriology* **182,** 2945–2952.

60. Abboud, R., Popa, R., Souza-Egipsy, V., Giometti, C. S., Tollaksen, S. Mosher, J. J., Findlay, R. H., and Nealson, K. H. (2005). Low-Temperature Growth of *Shewanella oneidensis* MR-1. *Applied and Environmental Microbiology* **71** (2), 811–816.

61. Myers, C. R., and Nealson, K. H. (19880. Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor. *Science* **240,** 1319–1321.

62. Venkateswaran, K., Moser, D. P., Dollhopf, M. E., Lies, D. P., Saffarini, B. J., Ringerlberg, D. B., Nishijima, M., Sano, H., Burghardt, J., Stachebrandt, E., and Nealson, K. H. (1999). Polyphasic taxonomy of the genus Shewanella and description of *Shewanella oneidensis* sp. Nov. *International Journal of Systematic Bacteriology* **49**, 705-724.

63. Liu, C., Y. Gorby, J. M. Zachara, J. K. Fredrickson, and C. F. Brown. (2002). Reduction kinetics of Fe (III), Co (III), U (VI), Cr (VI), and Tc (VII) in cultures of

dissimilatory metal- reducing bacteria. *Biotechnology and Bioengineering* **80,** 637–649.

64. Tang, Y.J., Hwang, J. S., Wemmer, D. E., and keasling, J. D. (2007). Shewanella oneidensis MR-1 fluxome under various oxygen conditions. *Applied and Environmental Micorbiology* **73** (3), 718-729.
65. Tiedje, J. M. (2002). Shewanella-the environmentally versatile genome. *Nature Biotechnology* 20, 1093-1094.
66. Jorens, P. G., Goovaerts, K., and Leven, M. (2004). *Shewanella putrefaciens* isolated in a case of ventilator-associated pneumonia. *Respiration* **71**, 199-201.
67. DeRisi, J. L., V. R. Iyer, and P. O. Brown. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278,** 680–686.
68. Richmond, C. S., J. D. Glasner, R. Mau, H. Jin, and F. R. Blattner. (1999). Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Research* **27,** 3821–3835.
69. Wodicka, L., H. Dong, M. Mittmann, M. H. Ho, and D. J. Lockhart. (1997). Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nature Biotechnology* **15:**1359–1367.
70. Rhodius, V. A., and LaRossa, R. A. (2003). Uses and pitfalls of microarrays for studying transcriptional regulation. *Current Opinion in Microbiology* **6**, 114–119.
71. Cases, I., and de Lorenzo, V. (2005). Promoters in the environment: transcriptional regulation in its natural context. *Nature Review Microbiology* **3**, 105–118.

72. Merrell, D. S., Thompson, L. J., Kim, C. C., Mitchell, H., Tompkins, L. S., and Lee, A., Falkow, S. (2003). Growth phasedependent response of *Helicobacter pylori* to iron starvation. *Infection and Immunity* **71**, 6510–6525.

73. Schuster, M., Hawkins, A. C., Harwood, C. S. and Greenberg, E. P. (2004). The *Pseudomonas aeruginosa* RpoS regulon and its relationship to quorum sensing. *Molecular Microbiology* **51**, 973– 985.

74. Candiano, G., Bruschi, M., Musante, L., Santucci, L., Ghiggeri, GM., Carnemolla, B., Orecchia, P., Zardi, and L., Righetti, PG. (2004). Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **25**(9), 1327-33.

75. 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.

76. Clauser, K.R., Baker, P., and Burlingame, A.L. (1999). Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Analytical Chemistry* **71**, 2871-2882.

77. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402.

78. Schaffer, A.A., Aravind, L., Madden, T.L., Shavirin, S., Spouge, J.L., Wolf, Y.I., Koonin, E.V., and Altschul, S.F. (2001). Improving the accuracy of PSI-BLAST protein 231 database searches with composition-based statistics and other refinements. *Nucleic Acids Research* **29**, 2994- 3005.

79. Marchler-Bauer, A., Anderson, J.B., DeWeese-Scott, C., Fedorova, N.D., Geer, L.Y., He, S., Hurwitz, D.I., Jackson, J.D., Jacobs, A.R., Lanczycki, C.J., Liebert, C.A., Liu, C., Madej, T., Marchler, G.H., Mazumder, R., Nikolskaya, A.N., Panchenko, A.R., Rao, B.S., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Vasudevan, S., Wang, Y., Yamashita, R.A., Yin, J.J., and Bryant, S.H. (2003). CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Research* **31**, 383-387.

80. Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Yamashita, R.A., Yin, J.J., Zhang, D., and Bryant, S.H. (2005). CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Research* **33**, D192-196.

81. Marchler-Bauer, A., and Bryant, S.H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Research* **32**, W327-331.

82. Fried, M., and Crothers, D.M. (1981). Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Research* **9**, 6505-6525.

83. Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510.

84. Oliphant, A.R., Brandl, C.J., and Struhl, K. (1989). Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence

oligonucleotides: analysis of yeast GCN4 protein. *Molecular and Cellular Biology* **9**, 2944-2949.

85. Sambrook, J., and Russel, D.W. (2001). *Molecular Cloning*, 3rd Edition (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
86. Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology* **2**, 28-36.
87. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Research* **14**, 1188-1190.
88. Gao, W., Liu, Y., Giometti, C., Tollaksen, S., Khare, T., Wu, L., Klingeman, D., Fields, M., and Zhou, J. (2006). Knock-out of SO1377 gene, which encodes the member of a conserved hypothetical bacterial protein family COG2268, results in alteration of iron metabolism, increased spontaneous mutation and hydrogen peroxide sensitivity in *Shewanella oneidensis* MR-1. *BMC Genomics* **7**(1)**,** 76.
89. Wan, X-F., VerBerkmoes, NC., McCue, LA., Stanek, D., Connelly, H., Hauser, LJ., Wu, L., Liu, X., Yan, T., Leaphart, A., Hettich, RL., Zhou, J., and Thompson, DK. Transcriptomic and proteomic characterization of the Fur modulon in the metal-reducing bacterium *Shewanella oneidensis*. *Journal of Bacteriology* **186**(24)**,** 8385-8400.
90. Alexeyev, M. F., Shokolenko, I. N., and Croughan, T. P. (1995). Improved antibiotic-resistance gene cassettes and omega elements for *Escherichia coli* vector construction and in vitro deletion/insertion mutagenesis. *Gene* **160 (1)**: 63-67.

91. Link, A.J., Phillips, D., and Church, M. (1997). Methods for Generating Precise Deletions and Insertions in the Genome of Wild-Type *Escherichia coli*: Application to Open Reading Frame Characterization. *Journal of Bacteriology* **179** (20), 6228.

92. Saltikov, C. W., and Newman, D. K. (2003). Genetic identification of a respiratory arsenate reductase. *Proceedings of the National Academy of Sciences of the United States of America* **100** (19), 10983-10988.

93. Lorenzo, V. D., Herrero, M., and Timmis, K. N. (1990). Mini-Tn5 transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative eubacteria. *Journal of Bacteriology* **172**, 6568-6572.

94. Bochner, B. R., Gadzinski, P., and Panomitros, E. (2001). Phenotype MicroArrays for High- throughput Phenotypic Testing and Assay of Gene Function. *Genome Research.* **11**,1246-1255.

95. Bochner, B.R. (1988). New methods aid microbial identification. *Bio/Technology* **6,** 756

96. Bochner, B.R. (1989). Sleuthing out bacterial identities. *Nature* **339,** 157–158.

97. He, Z., Wu L., Fields, M. W., and Zhou J. (2005) Use of Microarrays with Different Probe Sizes for Monitoring Gene Expression. *Applied and Environmental Microbiology* **71**(9), 5154–5162.

98. Wang, X., Verberkmoes, Nathan, C., McCue, L. A., Stanek, D., Connelly, H., Hauser, L. J., Wu, L., Liu, X., Yan, T., Leaphart, A., Hettich, R. L., Zhou, J., and Thompson. K. (2004). Transcriptomic and proeromic characterization of the FUR modulon in the metal-reducing Bactrium *Shewanella oneidensis*. *Journal of Bacteriology* **186**(24), 8385-8400.

99. Wang, M. (2006). Discovery of new transcription factors in *Pyrococcus furiosus* with DNA-affinity protein capture and mass spectrometry. Ph.D. Dissertation thesis, University of Georgia, Athens.

100. Lipscomb, G. (2007). A Novel Redox-Sensitive Transcriptional Regulator Involved in *Pyrococcus furiosus* Sulfur Response. Ph.D. Dissertation thesis, University of Georgia, Athens.

101. Christman, M. F., Storz, G., and Ames, B. N, (1989). OxyR, a Positive Regulator of Hydrogen Peroxide-Inducible Genes in Escherichia coli and Salmonella typhimurium, is Homologous to a Family of Bacterial Regulatory Proteins. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 3484-3488.

102. Jormakka, M., Byrne B., Iwata S. (2003). Formate dehydrogenase – a versatile enzyme in changing environments. *Current Opinion in Structural Biology* **13**, 418–423.

103. Bagramyan, K., Trchounian, A. (2003). Structural and functional features of formate hydrogen lyase, an Enzyme of mixed-acid fermentation from Escherichia coli. *Biochemistry* (Moscow), **68**(11), 1159-1170.

104. Casalot, L., and Rousset, M. (2005). Maturation of the [NiFe] hydrogenases. *Biochemical Society Transaction* **9** (5), 228-237.

105. Jeffrey, A., Gralnick, C., Brown, T., and Newman. D. K. (2005). Anaerobic regulation by an atypical Arc system in *Shewanella oneidensis*. *Molecular Microbiology* **56**(5), 1347-1357.

106. Sarah, L., Messenger, J. G. (2003). FNR-mediated regulation of *hyp* expression in *Escherichia coli*. *FEMS Microbiology Letters* **228,** 81-86.

107. Rudinger-Thirion, J., Giege, R., Felden, B. (1999). Aminoacylated tmRNA from Escherichia coli interacts with prokaryotic enlongation factor Tu. *RNA* **5**, 989-992.

108. Kondo, H., Shiratsuchi, K., Yoshimoto, T., Masuda, T., Kitazono, A., Tsuru, D., Anai, M., Sekiguchi, M., and Tanabe, T. (1991). Acetyl-CoA carboxylase from Escherichia coli: gene organization and nucleotide sequence of the biotin carboxylase subunit. *Proceedings of the National Academy of Sciences of the United States of America* **88**(21), 9730-9733.

109. Iost, I., Dreyfus, M. (2006). DEAD-box RNA helicases in Escherichia coli. *Nucleic Acids Research* **34**(15), 4189–4197.

110. Tsung, K., Brissette, RE., and Inouye M. (1989). Identification of the DNA-binding domain of the OmpR protein required for transcriptional activation of the ompF and ompC genes of Escherichia coli by in vivo DNA footprinting. *Journal of Biological Chemistry* **264**(17), 10104-9.

111. Lee, C., Lee, S. M., Mukhopadhyay, P., Kim, S, J., Lee, A. C., Ahn, W. S., Yu, M. H., Storz, G., and Ryu, S. E. (2004). Redox regulation of OxyR requires specific disulfide bond formation involving a rapid kinetic reaction path. *Nature Structural & Molecular Biology* **11**, 1179-1185.

112. Gralnick, J. A., Brown, C., and Newman, D. K. (2005). Anaerobic regulation by an atypical Arc system in *Shewanella oneidensis*. *Molecular Microbiology* **56** (5), 1347–1357.

113. Moxley, R.A., and Jarrett, H.W. (2005). Oligonucleotide trapping method for transcription factor purification systematic optimization using electrophoretic mobility shift assay. *Journal of Chromatography A* **1070**, 23-34.

114. Zhou, L., Lei, X., Bochner, B. R., and Wanner, B. L. (2003). Phenotype MicroArray Analysis of *Escherichia coli* K-12 Mutants with Deletions of All Two-Component Systems *Journal of Bacteriolog* 4956–4972.

115. Fodor, IK., Holtz-Morris, AE., and McCutchen-Maloney SL. (2005). Growth curve models for the analysis of phenotype arrays for a systems biology overview of *Yersinia pestis* (Joint statistical meetings), 1-7.

116. Evensen, G. and Seeberg, E. (1982). Adaptation to alkylation resistance involves the induction of a DNA glycosylase. *Nature* **296**, 773-775.

117. Thomas, L., Yang, C.-H. and Goldthwait, D.A. (1982). Two DNA glycosylases in *Escherichia coli* which release primarily 3-methyladenine. *Biochemistry* **21**, 1162-1169.

118. Yang, K., Huerta, C., Zhang, H. (2008). Structural Basis for Substrate Binding and the Catalytic Mechanism of Type III Pantothenate Kinase. *Biochemistry* **47**, 1369-1380.

119. Cheesman, M. R.,Le Brun, N. E., Kadir F. H. A., Thomson, A. J., Moore, G. R., Andrew, S. C., Guest, J. R., Harrison, P. M., Smith, J. M. A., and Yewdall, S. J. (1993). Haem and non-haem iron sites in Escherichia coli bacterioferritin : spectroscopic and model building studies. *Biochemical Journal* **292** (1), 47-56.

120. Rao, N. N., and Torriani, A. (1990). Molecular aspects of phosphate transport in *Escherichia coli. Molecular Microbiology* **4**(7), 1083-1090.

121. Muda, M., Rao, N. N., and Torriani, A. (1992). Role of PhoU in Phosphate Transport and Alkaline Phosphatase Regulation. Journal of Bacteriology 8057-8064.

122. Atsu Aiba and Kiyoshi Mizobuchiz. (1989). Nucleotide Sequence Analysis of Genes *purH* and *purD* Involved in the *de Novo* Purine Nucleotide Biosynthesis of *Escherichia coli.* The Journal of Biological Cheimistry **264**(35), 21239-21246.

123. He, B., Shiau, A., Choi, K. Y., Zalkin, H., and Smith, J. M. (1990). Genes of the *Escherichia coli* pur regulon are negatively controlled by a repressor-operator Interaction. *Journal of Bacteriology* **172**(8), 4555-4562.

124. Brett, PJ., Vasu, SK., Grant, CC., Levin, JC., and Mckenzie DT. (2002). Characterization of 5'-methylthioadenosine mucleosidase/S-mdenosylhomocysteine mucleosidase (Pfs) mutant phenotypes in pathogenic and non-Pathogenic bacteria. *Interscience Conference on Antimicrobial Agents and Chemotherapy*.

125. Rothery, R. A., Trieber, C. A., Weiner, J. H. (1999). Interactions between the Molybdenum Cofactor and Iron-Sulfur Clusters of *Escherichia coli* Dimethylsulfoxide Reductase. *The Journal of Biological Chemistry* **274**(19), 13002-13009.

126. Berg, B. L., Li, J., Heider, J., and Stewart, V. (1991) *Journal of Biological Chemistry* **266,** 22380–22385.

127. Plunkett, G., Burland, V., Daniels, D. L., and Blattner, F. R. (1993) *Nucleic Acids Research* **21,** 3391–3398.

128. Reyes-Ramirez, F., Dobbin, P., Sawers, Gary., and Richardson, D. J. (2003). Characterization of Transcriptional Regulation of *Shewanella frigidimarina* Fe(III)-Induced Flavocytochrome *c* Reveals a Novel Iron-Responsive Gene Regulation System. *Journal of Bacteriology* **185**(15), 4564-4571.

129. Gao, J., and Gussin, G. N. (1991). Activation of the trpBA Promoter of Pseudomonas aeruginosa by TrpI Protein In Vitro. *Journal of Bacteriology* 3763-3769.

130. Tao, M., Tu¨rk, K., Diez, J., Gru¨tter, M. G., Fritzd, G., and Steubera, J. (2006).Crystallization of the NADH-oxidizing domain of the Na+-translocating NADH:ubiquinone oxidoreductase from Vibrio cholerae. Acta crystallographica section F **62**, 110-112.

131. Stolpe, S., and Friedrich, T. (2004). The *Escherichia coli* NADH:Ubiquinone Oxidoreductase (Complex I) Is a Primary Proton Pump but May Be Capable of Secondary Sodium Antiport. *The Journal of Biological Chemistry* **279**(18), 18377-18383.

132. Hajime Tokudat and TsutomuU nemoto. (1982). Characterization of the Respiration-dependent Na' Pump in the Marine Bacterium *Vibrio alginolyticus. The Journal of Biological Chemistry* **257**(17), 10007-10014.

133. Callahan, A. M., and Parkinson, J. S. (1985). Genetics of Methyl-Accepting Chemotaxis Proteins in Escherichia coli: *che*D Mutations Affect the Structure and Function of the Tsr Transducer. *Journal of Bacteriology* 96-104.

134. Kondoh, H., Ball, C. B., and Adler, J. (1979). Identification of a methyl-accepting chemotaxis protein for the ribose and galactose chemoreceptors of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **76**(1), 260-264.

135. Springer, W. R., and D. E. Koshland, Jr. 1977. Identification of a protein methyltransferase as the *cheR* gene product in the bacterial sensing system. Proc. Natl. Acad. Sci. U.S.A. **74**, 533-537.

136. Van Der Werf, P., and D. E. Koshland, Jr. 1977. Identification of a y-glutamyl methyl ester in bacterial membrane protein involved in chemotaxis. J. Biol. Chem. 252:2793-2795.

137. Stock, J. B., and D. E. Koshland, Jr. 1978. A protein methylesterase involved in bacterial sensing. Proc. Natl. Acad. Sci. U.S.A. 75:3659-3663.

138. Dixon, M. P., Pau, R. N., Howlett, G. J., Dunstan, D. E., Sawyer, W. H., and Davidson, B. E. (2002). The Central Domain of *Escherichia coli* TyrR Is Responsible for Hexamerization Associated with Tyrosine-mediated Repression of Gene Expression. *The Journal of Biological Chemistry* **277**(26), 23186-23192.

139. Kredich, N. M. (1992). The molecular basis for positive regulation of *cys* promoters in *Salmonella typhimurium and Escherichia coli*. *Molecular Microbiology* **6**(19), 2747-2753.

140. Lithgow, J. K., E. J. Hayhurst, G. Cohen, Y. Aharonowitz, and S. J. Foster. (2004). Role of a cysteine synthase in *Streptococcus aureus. Journal of Bacteriology* **186**, 1579-1590.

141. Lee, K., Song, S., Kim, K., Kim, Y., Kim, S., Kho, B., Ko, D., Choi, Y., Lee, Y., Kim, C., Kim, Y., Lim, J., Kim, Y., Min, K., Wanner, B. (2000). Cloning and Characterization of the UDP-Sugar Hydrolase Gene (*ushA*) of *Enterobacter aerogenes* IFO 12010. *Biochemical and Biophysical Research Communications* **269**, 526-531.

142. Edwards, C. J., Innes, D. J., Burns, D. M., and Beacham, I. R. **(**1993). *FEMS Microbiol. Lett.* **114,** 293–298.

APPENDIX

Table A. UOR database search results for the motif AACn$_4$GTT

UOR Name	Motif	UOR Start	UOR Stop	Genome Start	Genome Stop	Annotation
SO0021	**AACAGTCGTT**	-157	-148	26791	26800	multifunctional fatty acid oxidation complex subunit alpha
SO0024	**AACAGCGGTT**	-149	-140	28957	28948	potassium uptake protein TrkH
SO0035	**AACTTGAGTT**	-68	-59	40205	40196	smg protein
SO0041	**AACTACAGTT**	-46	-37	44570	44561	hypothetical protein
SO0057	**AACCAACGTT**	-170	-161	65821	65830	potassium uptake protein, Trk family
SO0060	**AACCTGAGTT**	-165	-156	68298	68307	sesor histidine kinase
SO0061	**AACTCAGGTT**	-110	-101	68307	68298	lipoprotein, NLP/P60 family
SO0069	**AACGCCTGTT**	-53	-44	81850	81859	transport protein, putative
SO0080	**AACGAGCGTT**	-65	-56	93669	93678	hypothetical 4-hydroxybenzoyl-CoA thioesterase
SO0080	**AACTAGAGTT**	-53	-44	93657	93666	hypothetical 4-hydroxybenzoyl-CoA thioesterase
SO0081	**AACTCTAGTT**	-171	-162	93666	93657	hypothetical protein
SO0081	**AACGCTCGTT**	-159	-150	93678	93669	hypothetical protein
SO0082	**AACATTGGTT**	-91	-82	94079	94070	transcriptional regulator, MerR family
SO0099	**AACCATAGTT**	-16	-7	108394	108385	hypothetical protein
SO0100	**AACCATAGTT**	-183	-174	108394	108385	hypothetical protein
SO0157	**AACCAAGGTT**	-159	-150	165410	165419	proton/glutamate symporter
SO0166	**AACTGAAGTT**	-95	-86	172216	172207	general secretion pathway protein D
SO0166	**AACACGCGTT**	-27	-18	172284	172275	general secretion pathway protein D
SO0169	**AACCCATGTT**	-173	-164	177116	177107	general secretion pathway protein G

SO0218	**AACGTGCGTT**	-117	-108	228634	228625	preprotein translocase subunit SecE
SO0223	**AACATTTGTT**	-11	-2	231755	231746	50S ribosomal protein L7/L12
SO0224	**AACTTAAGTT**	-53	-44	232324	232315	DNA-directed RNA polymerase, beta subunit
SO0248	**AACGAATGTT**	-107	-98	253498	253489	30S ribosomal protein S5
SO0262	**AACAATTGTT**	-133	-124	262781	262790	heme exporter protein CcmB
SO0273	**AACACAGGTT**	-11	-2	277182	277191	hypothetical protein
SO0299	**AACGCCTGTT**	-178	-169	304802	304811	hypothetical protein
SO0319	**AACAGGCGTT**	-180	-171	325056	325047	hypothetical protein
SO0322	**AACGATAGTT**	-189	-180	329162	329171	hypothetical protein
SO0328	**AACTAAGGTT**	-21	-12	332313	332304	hypothetical protein
SO0329	**AACTAAGGTT**	-175	-166	332313	332304	hypothetical protein
SO0360	**AACTGGAGTT**	-17	-8	370690	370699	DNA-directed RNA polymerase subunit omega
SO0379	**AACACAAGTT**	-25	-16	391539	391548	hypothetical protein
SO0395	**AACCAGAGTT**	-115	-106	409351	409360	ribosomal protein L11 methyltransferase
SO0397	**AACTCAGGTT**	-99	-90	410555	410546	fumarate reductase cytochrome B subunit
SO0404	**AACAGTAGTT**	-22	-13	418191	418182	hypothetical protein
SO0414	**AACATCCGTT**	-41	-32	429951	429960	type 4 prepilin-like proteins leader peptide processing enzyme
SO0416	**AACCTTCGTT**	-29	-20	433337	433346	type IV pilus biogenesis protein PilB
SO0418	**AACCTTTGTT**	-70	-61	434073	434082	hypothetical protein
SO0426	**AACACAGGTT**	-97	-88	443175	443166	dihydrolipoamide dehydrogenase
SO0445	**AACAACCGTT**	-88	-79	474799	474808	hflC protein, putative
SO0494	**AACGACTGTT**	-74	-65	518716	518707	ISSod1, transposase OrfA
SO0498	**AACAATGGTT**	-117	-108	523488	523479	hypothetical protein
SO0516	**AACTTATGTT**	-67	-58	538274	538265	hypothetical protein
SO0528	**AACAGGAGTT**	-15	-6	550868	550859	hypothetical protein

SO0556	**AACGGCGGTT**	-146	-137	577099	577108	hypothetical protein
SO0580	**AACTGTTGTT**	-178	-169	607226	607217	hypothetical protein
SO0583	**AACTTATGTT**	-23	-14	609104	609113	bacterioferritin-associated ferredoxin
SO0588	**AACCTAAGTT**	-36	-27	618201	618210	hypothetical phosphatidylserine decarboxylase
SO0590	**AACCTAAGTT**	-36	-27	618201	618210	phosphatidylserine decarboxylase, authentic frameshift
SO0613	**AACCTTAGTT**	-171	-162	637455	637446	para-aminobenzoate synthase glutamine amidotransferase, component II
SO0620	**AACCGAGGTT**	-154	-145	647313	647304	hypothetical protein
SO0656	**AACCTATGTT**	-102	-93	679015	679006	ISSod1, transposase OrfA
SO0673	**AACTTCTGTT**	-17	-8	690897	690888	hypothetical protein
SO0723	**AACCGAGGTT**	-94	-85	738349	738358	hypothetical protein
SO0724	**AACCTCGGTT**	-61	-52	738358	738349	hypothetical protein
SO0725	**AACCTCGGTT**	-146	-137	738358	738349	catalase/peroxidase HPI
SO0758	**AACTCCAGTT**	-111	-102	775000	775009	hypothetical protein
SO0798	**AACGAGGGTT**	-113	-104	812753	812762	hypothetical protein
SO0799	**AACCCTCGTT**	-47	-38	812762	812753	hypothetical protein
SO0813	**AACTCACGTT**	-36	-27	825213	825204	hypothetical protein
SO0868	**AACATATGTT**	-131	-122	899072	899081	hypothetical protein
SO0875	**AACCTAAGTT**	-85	-76	905519	905528	sugar fermentation stimulation protein
SO0876	**AACTTAGGTT**	-94	-85	905528	905519	aminopeptidase B
SO0880	**AACACACGTT**	-48	-39	910225	910234	hypothetical protein
SO0881	**AACGTGTGTT**	-74	-65	910234	910225	hypothetical protein
SO0884	**AACTTGAGTT**	-106	-97	913437	913446	hypothetical protein
SO0949	**AACATAGGTT**	-117	-108	984905	984896	branched-chain amino acid transport system II carrier protein BrnQ

SO0958	**AACAATAGTT**	-121	-112	993714	993723	alkyl hydroperoxide reductase, C subunit
SO0964	**AACTTCGGTT**	-49	-40	998552	998561	hypothetical protein
SO0970	**AACAATCGTT**	-162	-153	1003477	1003486	fumarate reductase flavoprotein subunit precursor
SO0996	**AACCATTGTT**	-189	-180	1028762	1028771	glyoxalase family protein
SO1016	**AACACAGGTT**	-23	-14	1055354	1055363	NADH dehydrogenase subunit G
SO1025	**AACATCCGTT**	-109	-100	1063323	1063332	ISSod1, transposase OrfA
SO1059	**AACATTGGTT**	-164	-155	1100120	1100129	aminopeptidase N
SO1070	**AACTCTGGTT**	-40	-31	1109951	1109942	catalase
SO1094	**AACCTTCGTT**	-49	-40	1133858	1133867	hypothetical Transcription elongation factor
SO1094	**AACTGGGGTT**	-13	-4	1133822	1133831	hypothetical Transcription elongation factor
SO1112	**AACACTTGTT**	-149	-140	1154867	1154858	hypothetical chemotactic transducer
SO1113	**AACACTTGTT**	-122	-113	1154867	1154858	methyl-accepting chemotaxis protein, authentic point mutation
SO1127	**AACGCCCGTT**	-41	-32	1170750	1170741	chaperone protein DnaJ
SO1142	**AACCTGCGTT**	-144	-135	1182812	1182803	carbamoyl phosphate synthase large subunit
SO1167	**AACGCCGGTT**	-74	-65	1210786	1210795	rod shape-determining protein RodA
SO1169	**AACCCATGTT**	-195	-186	1213279	1213288	hypothetical protein
SO1204	**AACATCTGTT**	-56	-47	1247920	1247911	translation initiation factor IF-2
SO1206	**AACAGCAGTT**	-80	-71	1251060	1251051	tRNA pseudouridine synthase B
SO1210	**AACCGCTGTT**	-38	-29	1257139	1257130	TPR domain protein
SO1221	**AACAGGGGTT**	-17	-8	1267391	1267382	purine nucleoside phosphorylase
SO1224	**AACTCTTGTT**	-63	-54	1270620	1270611	hypothetical protein
SO1249	**AACAATTGTT**	-179	-170	1300820	1300829	peptidase, U32 family
SO1254	**AACCTCTGTT**	-181	-172	1303323	1303314	hypothetical protein
SO1298.4	**AACGGCAGTT**	-152	-143	1351222	1351231	hypothetical Na+/H+ antiporter

SO1299	**AACGGCAGTT**	-152	-143	1351222	1351231	Na+/H+ antiporter, degenerate
SO1330	**AACACTCGTT**	-138	-129	1388010	1388019	DNA mismatch repair protein
SO1330	**AACCTTGGTT**	-42	-33	1387914	1387923	DNA mismatch repair protein
SO1331	**AACAATTGTT**	-128	-119	1388454	1388445	dinucleoside polyphosphate hydrolase
SO1347	**AACTAAAGTT**	-78	-69	1403939	1403930	signal peptidase I
SO1366	**AACTTGTGTT**	-46	-37	1421518	1421509	sodium/hydrogen exchanger family protein
SO1373	**AACTTAGGTT**	-152	-143	1430670	1430679	hypothetical protein
SO1393	**AACGAATGTT**	-31	-22	1450425	1450416	transcriptional regulator, TetR family
SO1397	**AACTTAAGTT**	-70	-61	1454328	1454337	cytosine deaminase
SO1411	**AACTTTGGTT**	-129	-120	1468658	1468667	hypothetical protein
SO1427	**AACATTCGTT**	-86	-77	1488133	1488124	decaheme cytochrome c
SO1464	**AACCTGAGTT**	-90	-81	1529934	1529925	ISSod1, transposase OrfA
SO1504	**AACGCAAGTT**	-65	-56	1577298	1577289	hypothetical protein
SO1516	**AACGGCAGTT**	-18	-9	1590499	1590508	hypothetical protein
SO1560	**AACCTAAGTT**	-43	-34	1636745	1636736	phosphate-binding protein
SO1571	**AACTGAGGTT**	-50	-41	1650831	1650840	hypothetical protein
SO1583	**AACATGCGTT**	-162	-153	1660176	1660167	hypothetical protein
SO1583	**AACACTGGTT**	-102	-93	1660236	1660227	hypothetical protein
SO1587	**AACTTAAGTT**	-112	-103	1662948	1662939	hypothetical protein
SO1597	**AACCACAGTT**	-148	-139	1670279	1670288	hypothetical protein
SO1603	**AACACGGGTT**	-77	-68	1687135	1687144	transcriptional regulator, putative
SO1606	**AACTCGAGTT**	-79	-70	1691515	1691524	metallo-beta-lactamase superfamily protein
SO1607	**AACTCGAGTT**	-55	-46	1691524	1691515	transcriptional regulator, LysR family
SO1641	**AACGCGCGTT**	-165	-156	1723347	1723338	UDP-N-acetylglucosamine acyltransferase
SO1655	**AACGCCTGTT**	-186	-177	1742565	1742556	cysQ protein

SO1657	**AACACCAGTT**	-170	-161	1744422	1744413	hypothetical protein
SO1673	**AACCAATGTT**	-30	-21	1759698	1759689	outer membrane protein W
SO1691	**AACATTGGTT**	-59	-50	1779429	1779438	lipoprotein Blc
SO1715	**AACACGAGTT**	-20	-11	1803789	1803798	hypothetical protein
SO1758	**AACTGCAGTT**	-113	-104	1839125	1839116	hypothetical protein
SO1770	**AACGACAGTT**	-199	-190	1850962	1850971	glycerate kinase, putative
SO1792	**AACAATTGTT**	-45	-36	1879698	1879707	methylenetetrahydrofolate dehydrogenase/methylenetetrahydrofolate cyclohydrolase
SO1796	**AACACAGGTT**	-161	-152	1883971	1883962	ATP-dependent protease La
SO1805	**AACAGCCGTT**	-16	-7	1896448	1896457	peptide ABC transporter, periplasmic peptide-binding protein
SO1830	**AACATTTGTT**	-158	-149	1923166	1923175	GGDEF domain protein
SO1831	**AACCTGAGTT**	-109	-100	1922999	1922990	hypothetical protein
SO1844	**AACCTTTGTT**	-110	-101	1932898	1932889	extracellular nuclease, putative
SO1859	**AACCCGGGTT**	-151	-142	1951265	1951256	hypothetical protein
SO1859	**AACCCGGGTT**	-62	-53	1951354	1951345	hypothetical protein
SO1898	**AACCAAAGTT**	-126	-117	1996984	1996993	transcriptional regulator, putative
SO1908	**AACAGTTGTT**	-170	-161	2006990	2006981	hypothetical protein
SO1937	**AACAAAGGTT**	-90	-81	2042474	2042483	ferric uptake regulator
SO1938	**AACCTTTGTT**	-177	-168	2042483	2042474	NAD-dependent deacetylase
SO1956	**AACTCAGGTT**	-51	-42	2062618	2062609	hypothetical protein
SO1987	**AACTAAAGTT**	-24	-15	2090456	2090465	ATP-dependent protease La (LON) domain protein
SO1988	**AACTTTAGTT**	-83	-74	2090465	2090456	hypothetical methyltransferase
SO1996	**AACTTAAGTT**	-171	-162	2096080	2096071	hypothetical protein
SO2041	**AACCTGTGTT**	-193	-184	2140220	2140229	hypothetical protein
SO2042	**AACACAGGTT**	-81	-72	2140229	2140220	putative sulfite oxidase subunit YedY

SO2070	**AACCTGAGTT**	-134	-125	2167200	2167209	imidazole glycerol phosphate synthase subunit HisH
SO2108	**AACAATGGTT**	-150	-141	2205668	2205659	glucosyltransferase MdoH
SO2108	**AACTCAAGTT**	-83	-74	2205735	2205726	glucosyltransferase MdoH
SO2133	**AACTAGGGTT**	-37	-28	2239107	2239098	hypothetical protein
SO2151	**AACGCGTGTT**	-44	-35	2266668	2266677	hypothetical protein
SO2154	**AACTGATGTT**	-42	-33	2267992	2267983	transposase, IS110 family, degenerate
SO2156	**AACCATAGTT**	-139	-130	2271933	2271942	alpha amylase family protein
SO2156	**AACCTAGGTT**	-107	-98	2271901	2271910	alpha amylase family protein
SO2159	**AACTTCGGTT**	-46	-37	2273469	2273478	TonB domain protein
SO2199	**AACTTTTGTT**	-138	-129	2314692	2314701	hypothetical protein
SO2202	**AACCCTTGTT**	-146	-137	2316401	2316392	transcriptional regulator, LysR family
SO2234	**AACATTTGTT**	-106	-97	2352051	2352042	hypothetical protein
SO2267	**AACGAAGGTT**	-88	-79	2380862	2380853	co-chaperone HscB
SO2302	**AACAAGTGTT**	-80	-71	2413157	2413148	50S ribosomal protein L20
SO2337	**AACACAAGTT**	-176	-167	2445063	2445072	peptide methionine sulfoxide reductase
SO2338	**AACTTGTGTT**	-113	-104	2445072	2445063	succinylglutamate desuccinylase
SO2371	**AACCATGGTT**	-86	-77	2477392	2477401	hypothetical protein
SO2372	**AACCATGGTT**	-87	-78	2477401	2477392	hypothetical protein
SO2373	**AACGCTTGTT**	-135	-126	2479138	2479147	inner membrane transport protein YdhC
SO2375	**AACCACGGTT**	-119	-110	2480180	2480171	hypothetical protein
SO2394	**AACTTGGGTT**	-130	-121	2497268	2497259	penicillin-binding protein 4
SO2395	**AACAACTGTT**	-36	-27	2500756	2500765	acyl-CoA dehydrogenase family protein
SO2422	**AACCAAAGTT**	-173	-164	2532432	2532441	hypothetical protein
SO2447	**AACCAGAGTT**	-25	-16	2562531	2562522	channel protein, hemolysin III family subfamily
SO2455	**AACGGGAGTT**	-127	-118	2572872	2572881	transcriptional regulator, LysR family

SO2456	**AACTCCCGTT**	-149	-140	2572881	2572872	hypothetical protein
SO2466	**AACATTTGTT**	-125	-116	2587946	2587955	transposase, degenerate
SO2466	**AACTTAAGTT**	-10	-1	2587831	2587840	transposase, degenerate
SO2477	**AACGGGAGTT**	-17	-8	2598196	2598187	alcohol dehydrogenase, iron-containing
SO2487	**AACGCCCGTT**	-43	-34	2610426	2610435	phosphogluconate dehydratase
SO2494	**AACACTCGTT**	-196	-187	2618739	2618730	zinc-dependent metallopeptidase
SO2499	**AACATGAGTT**	-29	-20	2625730	2625739	hypothetical protein
SO2514	**AACGGCGGTT**	-126	-117	2641616	2641607	endonuclease III
SO2535	**AACAATTGTT**	-102	-93	2663235	2663244	hypothetical amidotransferase
SO2536	**AACAATTGTT**	-153	-144	2663244	2663235	acyl-CoA dehydrogenase
SO2537	**AACAGCAGTT**	-117	-108	2668090	2668099	sodium/hydrogen exchanger family protein
SO2570	**AACTTAAGTT**	-148	-139	2701116	2701125	lipoprotein, putative
SO2590	**AACGAGCGTT**	-128	-119	2717663	2717654	ribosomal biogenesis GTPase
SO2602	**AACTTTAGTT**	-134	-125	2739114	2739123	putative solute/DNA competence effector
SO2626	**AACATCCGTT**	-77	-68	2763620	2763629	ATP-dependent Clp protease, ATP-binding subunit ClpA
SO2655	**AACACCCGTT**	-47	-38	2791459	2791450	prophage MuSo2, DNA transposition protein, putative
SO2673	**AACCACCGTT**	-152	-143	2799611	2799602	hypothetical protein
SO2678	**AACTAAGGTT**	-88	-79	2801690	2801681	hypothetical protein
SO2680	**AACAACCGTT**	-57	-48	2803474	2803465	hypothetical protein
SO2684	**AACCGAGGTT**	-130	-121	2806573	2806564	prophage MuSo2, protein Gp32, putative
SO2690	**AACGAGCGTT**	-99	-90	2810547	2810538	prophage MuSo2, virion morphogenesis protein, putative
SO2701	**AACAGCAGTT**	-130	-121	2818842	2818833	hypothetical protein
SO2701	**AACATTAGTT**	-69	-60	2818903	2818894	hypothetical protein
SO2725	**AACGGATGTT**	-78	-69	2842915	2842924	transcriptional regulator, LuxR family

SO2748	**AACCAACGTT**	-26	-17	2870536	2870545	tolB protein
SO2762	**AACCATTGTT**	-80	-71	2884532	2884523	UMP phosphatase
SO2810	**AACTTTGGTT**	-127	-118	2931853	2931844	transposase family protein, authentic frameshift
SO2859	**AACTCAGGTT**	-64	-55	2984361	2984370	hypothetical protein
SO2865	**AACATTAGTT**	-67	-58	2988940	2988949	L-lysine exporter, putative
SO2866	**AACTAATGTT**	-56	-47	2988949	2988940	chromosome replication initiation inhibitor protein
SO2870	**AACGTGGGTT**	-43	-34	2993244	2993253	hypothetical protein
SO2893	**AACAGCTGTT**	-68	-59	3016361	3016370	hypothetical protein
SO2907	**AACAAGTGTT**	-24	-15	3034779	3034770	TonB-dependent receptor domain protein
SO2935	**AACATGCGTT**	-69	-60	3071749	3071740	short chain dehydrogenase
SO2949	**AACAGCCGTT**	-86	-77	3087413	3087422	prophage LambdaSo, minor tail protein L
SO2995	**AACGATGGTT**	-174	-165	3116841	3116832	hypothetical protein
SO2996	**AACAGCAGTT**	-29	-20	3117357	3117366	hypothetical protein
SO2997	**AACTGCTGTT**	-54	-45	3117366	3117357	hypothetical protein
SO2998	**AACCGCAGTT**	-53	-44	3117665	3117656	hypothetical protein
SO3003	**AACATGCGTT**	-107	-98	3119788	3119779	hypothetical protein
SO3005	**AACCGATGTT**	-130	-121	3121071	3121062	hypothetical protein
SO3029	**AACCTAGGTT**	-191	-182	3138964	3138955	hypothetical protein
SO3089	**AACAGGTGTT**	-158	-149	3214273	3214282	3-ketoacyl-CoA thiolase
SO3090	**AACACCTGTT**	-127	-118	3214282	3214273	MoxR domain protein
SO3091	**AACTAACGTT**	-65	-56	3215408	3215399	hypothetical protein
SO3093	**AACGTTGGTT**	-119	-110	3217000	3216991	von Willebrand factor type A domain protein
SO3129	**AACAGGGGTT**	-187	-178	3258481	3258490	hypothetical protein
SO3129	**AACTAATGTT**	-152	-143	3258446	3258455	hypothetical protein
SO3145	**AACAGGCGTT**	-116	-107	3282535	3282544	electron transfer flavoprotein, beta subunit

SO3160	**AACAGGTGTT**	-12	-3	3298089	3298098	dTDP-4-dehydrorhamnose 3,5-epimerase
SO3178	**AACTTTTGTT**	-122	-113	3317274	3317283	hypothetical protein
SO3189	**AACATGTGTT**	-148	-139	3328203	3328212	polysaccharide biosynthesis protein
SO3211	**AACAATCGTT**	-90	-81	3351793	3351802	flagellar biosynthetic protein FlhG
SO3215	**AACTCCCGTT**	-124	-115	3356857	3356866	flagellar biosynthesis protein FlhB
SO3281	**AACCTAAGTT**	-136	-127	3422491	3422500	hypothetical protein
SO3282	**AACTTAGGTT**	-47	-38	3422500	3422491	methyl-accepting chemotaxis protein
SO3284	**AACGTGAGTT**	-195	-186	3425153	3425162	conserved hypothetical protein
SO3292	**AACTATCGTT**	-137	-128	3437380	3437389	bifunctional GMP synthase/glutamine amidotransferase protein
SO3301	**AACAATTGTT**	-195	-186	3447575	3447584	flavocytochrome c flavin subunit
SO3324	**AACTGTTGTT**	-114	-105	3476069	3476078	acetyltransferase, GNAT family
SO3355	**AACCCACGTT**	-75	-66	3498832	3498823	hypothetical protein
SO3358	**AACAAAAGTT**	-141	-132	3500224	3500215	HAM1 protein
SO3419	**AACTGCCGTT**	-41	-32	3559895	3559904	trp operon repressor
SO3420	**AACGGCAGTT**	-159	-150	3559904	3559895	cytochrome c'
SO3422	**AACTCTCGTT**	-68	-59	3561520	3561511	ribosomal subunit interface protein
SO3436	**AACTCGGGTT**	-96	-87	3581851	3581860	hypothetical protein
SO3442	**AACCGCAGTT**	-43	-34	3587579	3587588	MazG family protein
SO3442	**AACCGATGTT**	-28	-19	3587564	3587573	MazG family protein
SO3465	**AACAACAGTT**	-28	-19	3611725	3611734	transcription antitermination protein NusB
SO3479	**AACAGTTGTT**	-55	-46	3627151	3627160	hypothetical protein
SO3502	**AACTCACGTT**	-67	-58	3652538	3652529	hypothetical protein
SO3507	**AACTAAAGTT**	-78	-69	3659749	3659758	hypothetical protein
SO3508	**AACAGTTGTT**	-157	-148	3659655	3659646	hypothetical protein

SO3508	**AACTTTAGTT**	-54	-45	3659758	3659749	hypothetical protein
SO3510	**AACTAGCGTT**	-138	-129	3663726	3663735	methyl-accepting chemotaxis protein, putative
SO3513	**AACTTAGGTT**	-31	-22	3667020	3667029	tryptophan halogenase, putative
SO3540	**AACTCAAGTT**	-107	-98	3694587	3694596	hypothetical protein
SO3555	**AACGCCGGTT**	-126	-117	3712815	3712806	phosphoribosylaminoimidazole carboxylase ATPase subunit
SO3607	**AACAGTTGTT**	-73	-64	3768965	3768974	ISSod1, transposase OrfA
SO3619	**AACGTCCGTT**	-147	-138	3781436	3781427	hypothetical protein
SO3621	**AACGATGGTT**	-103	-94	3782500	3782509	hypothetical protein
SO3651	**AACTATTGTT**	-111	-102	3807427	3807436	50S ribosomal protein L27
SO3663	**AACGGGTGTT**	-174	-165	3816502	3816511	hypothetical protein
SO3664	**AACACCCGTT**	-59	-50	3816511	3816502	long-chain-fatty-acid--CoA ligase
SO3665	**AACTTAAGTT**	-197	-188	3820191	3820200	ABC transporter, ATP-binding/permease protein, putative
SO3676	**AACCTGAGTT**	-137	-128	3828818	3828809	hypothetical protein
SO3684	**AACGCATGTT**	-55	-46	3834753	3834744	transcriptional regulator, TetR family
SO3695	**AACTTGCGTT**	-118	-109	3845902	3845893	dihydroorotase
SO3729	**AACGAATGTT**	-67	-58	3873004	3872995	hypothetical phospholipase A
SO3753	**AACCCGCGTT**	-190	-181	3902060	3902069	ISSod11, transposase, truncation
SO3833	**AACTAAGGTT**	-20	-11	3983502	3983511	peptide chain release factor 1
SO3844	**AACATACGTT**	-90	-81	3993020	3993011	peptidase, M13 family
SO3861	**AACATAGGTT**	-167	-158	4007690	4007681	iron-sulfur cluster-binding protein
SO3876	**AACTTCGGTT**	-97	-88	4023523	4023514	ISSod4, transposase
SO3879	**AACGAGTGTT**	-134	-125	4026806	4026797	hypothetical protein
SO3881	**AACGCCAGTT**	-178	-169	4028283	4028274	hypothetical protein
SO3884	**AACTACAGTT**	-181	-172	4030466	4030457	site-specific recombinase, phage integrase family

SO3896	**AACTTAGGTT**	-27	-18	4042038	4042029	outer membrane porin, putative
SO3908	**AACATCTGTT**	-76	-67	4055117	4055108	enoyl-CoA hydratase/isomerase family protein
SO3940	**AACTCACGTT**	-167	-158	4085410	4085419	50S ribosomal protein L13
SO3942	**AACCTTTGTT**	-158	-149	4086962	4086953	serine protease, HtrA/DegQ/DegS family
SO3943	**AACACGCGTT**	-73	-64	4088525	4088516	protease DegS
SO4078	**AACCTCAGTT**	-153	-144	4233379	4233388	pmba protein
SO4079	**AACTGAGGTT**	-19	-10	4233388	4233379	hypothetical protein
SO4082	**AACTATGGTT**	-185	-176	4238090	4238099	hypothetical protein
SO4084	**AACCATAGTT**	-199	-190	4238099	4238090	hypothetical protein
SO4086	**AACAGTGGTT**	-109	-100	4240563	4240554	hypothetical protein
SO4100	**AACCCGTGTT**	-35	-26	4261319	4261328	MSHA biogenesis protein MshQ
SO4171	**AACGCAAGTT**	-81	-72	4344706	4344715	hypothetical protein
SO4196	**AACAGGTGTT**	-180	-171	4364728	4364737	hypothetical protein
SO4197	**AACACCTGTT**	-154	-145	4364737	4364728	ribonuclease activity regulator protein RraA
SO4213	**AACCAGCGTT**	-96	-87	4386704	4386695	hypothetical protein
SO4214	**AACTTATGTT**	-95	-86	4388256	4388265	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase
SO4246	**AACAACCGTT**	-198	-189	4425100	4425109	50S ribosomal protein L33
SO4246	**AACCACCGTT**	-141	-132	4425043	4425052	50S ribosomal protein L33
SO4250	**AACTCACGTT**	-57	-48	4427256	4427247	deoxyuridine 5'-triphosphate nucleotidohydrolase
SO4251	**AACTGGTGTT**	-136	-127	4427667	4427658	nucleoid occlusion protein
SO4262	**AACTCAAGTT**	-109	-100	4439014	4439023	hypothetical protein
SO4268	**AACTAAAGTT**	-181	-172	4447795	4447786	ISSod2, transposase OrfA
SO4270	**AACTTCGGTT**	-168	-159	4449062	4449053	hypothetical protein

SO4273	**AACAGTTGTT**	-66	-57	4453092	4453101	hypothetical protein
SO4280	**AACAACAGTT**	-111	-102	4459826	4459835	hypothetical protein
SO4282	**AACTCAAGTT**	-116	-107	4461123	4461114	potassium uptake protein KtrB
SO4298	**AACAAACGTT**	-173	-164	4476591	4476582	cyclic nucleotide binding protein, putative
SO4316	**AACCCCTGTT**	-62	-53	4493551	4493542	hemY protein, putative
SO4319	**AACAACAGTT**	-121	-112	4505554	4505545	HlyD family secretion protein
SO4320	**AACAACCGTT**	-124	-115	4507032	4507023	agglutination protein
SO4392	**AACGGCCGTT**	-160	-151	4588801	4588810	hypothetical protein
SO4404	**AACATGAGTT**	-91	-82	4595868	4595859	iron-sulfur cluster-binding protein
SO4410	**AACTTGAGTT**	-58	-49	4603606	4603597	glutamine synthetase
SO4411	**AACAGTCGTT**	-52	-43	4605643	4605652	hypothetical protein
SO4433	**AACAGTGGTT**	-86	-77	4623247	4623256	hypothetical protein
SO4434	**AACCACTGTT**	-133	-124	4623256	4623247	hypothetical protein
SO4451	**AACTTAGGTT**	-110	-101	4640121	4640130	molybdenum cofactor biosynthesis protein C
SO4453	**AACGCCCGTT**	-66	-57	4641428	4641419	electron transfer flavoprotein-ubiquinone oxidoreductase, putative
SO4454	**AACCAAGGTT**	-87	-78	4643346	4643337	methyl-accepting chemotaxis protein
SO4485	**AACAATCGTT**	-180	-171	4673195	4673186	diheme cytochrome c
SO4506	**AACGTCAGTT**	-173	-164	4697572	4697563	iron-sulfur cluster-binding protein
SO4583	**AACTTTTGTT**	-181	-172	4779261	4779270	RNA polymerase factor sigma-32
SO4606	**AACGCACGTT**	-118	-109	4800608	4800599	cytochrome c oxidase, subunit II
SO4617	**AACGCACGTT**	-40	-31	4810735	4810726	DNA-damage-inducible protein F
SO4633	**AACATGAGTT**	-187	-178	4828582	4828573	osmolarity response regulator
SO4656	**AACTAGGGTT**	-117	-108	4852618	4852627	hypothetical protein
SO4656	**AACCGTGGTT**	-107	-98	4852608	4852617	hypothetical protein
SO4668	**AACGATTGTT**	-20	-11	4865499	4865490	hypothetical protein

SO4679	**AACCACTGTT**	-55	-46	4877599	4877590	glycosyl transferase, group 1 family protein
SO4691	**AACCTACGTT**	-24	-15	4891374	4891383	hypothetical protein
SO4709	**AACACGGGTT**	-114	-105	4912235	4912244	hypothetical protein
SO4724	**AACCAAGGTT**	-143	-134	4928721	4928712	molybdenum cofactor biosynthesis protein A
SO4734	**AACCAATGTT**	-161	-152	4942072	4942081	sensory box protein
SO4746	**AACAAGCGTT**	-101	-92	4958681	4958690	ATP synthase F1, epsilon subunit
SOA0073	**AACAAAGGTT**	-179	-170	58866	58875	transposase, IS5 family, truncation
SOA0084	**AACGCGGGTT**	-154	-145	65105	65114	transposase, IS6 family, truncation
SOA0097	**AACTGCCGTT**	-52	-43	76538	76547	ISSod6 transposase, TnpA_ISSod6_8
SOA0110	**AACTATGGTT**	-35	-26	90060	90051	expressed lipoprotein
SOA0123	**AACGGATGTT**	-36	-27	109957	109966	conserved domain protein
SOA0151	**AACCCAAGTT**	-20	-11	139341	139350	ISSod13 transposase, TnpA_ISSod13_5

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

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

[c] Annotations are from TIGR.

Table B. Induced or repressed genes containing SELEX motif in mutant under the three conditions tested.

Gene ID/ UOR[a]	Motif	Fold change [b]	P-value [c]	Annotation
+O_2				
SO0035	**AACTTGAGTT**	1.4	0.02	smg protein
SO0273	**AACACAGGTT**	1.5	0.002	hypothetical protein
SO0414	**AACATCCGTT**	1.7	0.03	type 4 prepilin-like proteins leader peptide processing enzyme
SO0758	**AACTCCAGTT**	1.6	0.03	hypothetical protein
SO0881	**AACGTGTGTT**	1.2	0.04	hypothetical protein
SO1221	**AACAGGGGTT**	1.2	0.03	purine nucleoside phosphorylase
SO1330	**AACACTCGTT AACCTTGGTT**	1.1	0.03	DNA mismatch repair protein
SO1331	**AACAATTGTT**	1.0	0.02	dinucleoside polyphosphate hydrolase
SO1427	**AACATTCGTT**	1.2	0.05	decaheme cytochrome c
SO1691	**AACATTGGTT**	1.2	0.04	lipoprotein Blc
SO2042	**AACACAGGTT**	2.7	0.02	putative sulfite oxidase subunit YedY
SO2866	**AACTAATGTT**	1.2	0.05	chromosome replication initiation inhibitor protein
SO2907	**AACAAGTGTT**	1.5	0.04	TonB-dependent receptor domain protein
SO3419	**AACTGCCGTT**	1.2	0.03	trp operon repressor
SO3479	**AACAGTTGTT**	1.6	0.04	hypothetical protein
SO3507	**AACTAAAGTT**	1.5	0.02	hypothetical protein
SO3705	**GTCCATCATCT TGAAC**	-3.5	0.002	5-methylthioadenosine nucleosidase/S-adenosylhomocysteine nucleosidase, putative
SO3729	**AACGAATGTT**	1.7	0.02	hypothetical phospholipase A
SO3844	**AACATACGTT**	1.3	0.03	peptidase, M13 family
SO4485	**AACAATCGTT**	1.6	0.03	diheme cytochrome c
SO4617	**AACGCACGTT**	1.1	0.008	DNA-damage-inducible protein F
-O_2				
SO0021	**AACAGTCGTT**	1.3	0.001	multifunctional fatty acid oxidation complex subunit alpha
SO0157	**AACCAAGGTT**	1.4	0.02	proton/glutamate symporter

SO0273	**AACACAGGTT**	1.2	0.02	hypothetical protein
SO0360	**AACTGGAGTT**	1.2	0.005	DNA-directed RNA polymerase subunit omega
SO0397	**AACTCAGGTT**	1.1	0.004	fumarate reductase cytochrome B subunit
SO0556	**AACGGCGGTT**	1.0	0.02	hypothetical protein
SO0673	**AACTTCTGTT**	1.2	0.002	hypothetical protein
SO0758	**AACTCCAGTT**	1.9	0.05	hypothetical protein
SO0875	**AACCTAAGTT**	2.4	0.005	sugar fermentation stimulation protein
SO0880	**AACACACGTT**	3.9	0.05	hypothetical protein
SO0881	**AACGTGTGTT**	1.9	0.01	hypothetical protein
SO0949	**AACATAGGTT**	1.3	0.04	branched-chain amino acid transport system II carrier protein BrnQ
SO1112	**AACACTTGTT**	1.0	0.01	hypothetical chemotactic transducer
SO1221	**AACAGGGGTT**	1.6	0.04	purine nucleoside phosphorylase
SO1249	**AACAATTGTT**	1.4	0.03	peptidase, U32 family
SO1397	**AACTTAAGTT**	1.4	0.03	cytosine deaminase
SO1504	**AACGCAAGTT**	1.3	0.003	hypothetical protein
SO1603	**AACACGGGTT**	1.6	0.02	transcriptional regulator, putative
SO1641	**AACGCGCGTT**	1.0	0.05	UDP-N-acetylglucosamine acyltransferase
SO1691	**AACATTGGTT**	1.5	0.0005	lipoprotein Blc
SO1758	**AACTGCAGTT**	1.1	0.03	hypothetical protein
SO1770	**AACGACAGTT**	1.4	0.02	glycerate kinase, putative
SO1898	**AACCAAAGTT**	1.9	0.05	transcriptional regulator, putative
SO2041	**AACCTGTGTT**	1.2	0.003	hypothetical protein
SO2373	**AACGCTTGTT**	1.4	0.04	inner membrane transport protein YdhC
SO2375	**AACCACGGTT**	1.5	0.04	hypothetical protein
SO2514	**AACGGCGGTT**	1.2	0.040	endonuclease III
SO2570	**AACTTAAGTT**	1.1	0.01	lipoprotein, putative
SO2590	**AACGAGCGTT**	1.4	0.02	ribosomal biogenesis GTPase
SO2673	**AACCACCGTT**	1.5	0.003	hypothetical protein
SO2701	**AACAGCAGTT** **AACATTAGTT**	1.1	0.007	hypothetical protein
SO2762	**AACCATTGTT**	1.3	0.02	UMP phosphatase
SO2810	**AACTTTGGTT**	1.2	0.02	transposase family protein, authentic frameshift
SO2866	**AACTAATGTT**	1.5	0.0002	chromosome replication initiation inhibitor protein
SO3355	**AACCCACGTT**	1.9	0.02	hypothetical protein
SO3442	**AACCGATGTT** **AACCGATGTT**	1.4	0.001	MazG family protein
SO3479	**AACAGTTGTT**	1.3	0.002	hypothetical protein

SO3507	**AACTAAAGTT**	2.5	0.03	hypothetical protein
SO3664	**AACACCCGTT**	1.0	0.01	long-chain-fatty-acid—CoA ligase
SO3705	**GTCCATCATCT TGAAC**	-4.6	0.0001	5-methylthioadenosine nucleosidase/S-adenosylhomocysteine nucleosidase, putative
SO3884	**AACTACAGTT**	2.0	0.03	site-specific recombinase, phage integrase family
SO3942	**AACCTTTGTT**	1.0	0.04	serine protease, HtrA/DegQ/DegS family
SO4214	**AACTTATGTT**	1.9	0.04	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase
SO4392	**AACGGCCGTT**	1.8	0.007	hypothetical protein
SO4404	**AACATGAGTT**	1.8	0.01	iron-sulfur cluster-binding protein
SO4434	**AACCACTGTT**	1.2	0.03	hypothetical protein
SO4454	**AACCAAGGTT**	4.5	0.01	methyl-accepting chemotaxis protein
H_2O_2				
SO0580	**AACTGTTGTT**	1.0	0.04	hypothetical protein
SO2456	**AACTCCCGTT**	1.2	0.02	hypothetical protein
SO2655	**AACACCCGTT**	2.1	0.008	prophage MuSo2, DNA transposition protein, putative
SO2684	**AACCGAGGTT**	1.1	0.05	prophage MuSo2, protein Gp32, putative
SO2762	**AACCATTGTT**	1.7	0.04	UMP phosphatase
SO2995	**AACGATGGTT**	2.0	0.01	hypothetical protein
SO2997	**AACTGCTGTT**	2.1	0.002	hypothetical protein
SO2998	**AACCGCAGTT**	1.7	0.03	hypothetical protein
SO3003	**AACATGCGTT**	1.5	0.02	hypothetical protein
SO3005	**AACCGATGTT**	1.5	0.01	hypothetical protein
SO3281	**AACCTAAGTT**	3.2	0.04	hypothetical protein
SO3555	**AACGCCGGTT**	-1.2	0.03	phosphoribosylaminoimidazole carboxylase ATPase subunit
SO3705	**GTCCATCATCT TGAAC**	-4.0	0.008	5-methylthioadenosine nucleosidase/S-adenosylhomocysteine nucleosidase, putative
SO3942	**AACCTTTGTT**	1.3	0.03	serine protease, HtrA/DegQ/DegS family
SO4454	**AACCAAGGTT**	1.4	0.04	methyl-accepting chemotaxis protein
SO4652	**GTATTGAGTTT GGAAC**	2.6	0.02	sulfate ABC transporter, periplasmic sulfate-binding protein
SO4709	**AACACGGGTT**	1.3	0.03	hypothetical protein

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

[b] Fold change was calculated based on the absolute ratio of expression between SO3363 mutant and wild type.

[c] Annotations are from TIGR.