A NOVEL REDOX-SENSITIVE TRANSCRIPTIONAL REGULATOR INVOLVED IN

PYROCOCCUS FURIOSUS SULFUR RESPONSE

by

GINA LYNETTE PRIES LIPSCOMB

Under the Direction of ROBERT A. SCOTT

ABSTRACT

The archaeal basal transcription apparatus is eukaryotic in nature while transcriptional regulation appears to occur primarily through bacteria-like mechanisms; however, only a small number of archaeal transcriptional regulators have been characterized in detail. This work describes the discovery and characterization of an additional archaeal transcriptional regulator, SurR, of the hyperthermophile *Pyrococcus furiosus*. Much of the gene regulation which occurs at the transcriptional level is the result of regulatory proteins; therefore, 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 *P. furiosus* to growth in the presence or absence of elemental sulfur (S^0). The upstream DNA from the first ORF in the S^0 down-regulated membrane-bound hydrogenase operon (*mbh1*) was used as "bait" to capture and identify SurR from *P. furiosus* cell extracts derived from cultures grown in the presence and absence of S^0. The recombinant protein was verified to bind specifically to the *mbh1* promoter by electromobility shift assay and fluorescence-detected DNase I footprinting. Elucidation of its recognized DNA-binding motif by SELEX allowed the detection and verification of other binding sites in the genome, many of which were found upstream from the ORFs which are up- and down-regulated within 10 min of S^0 addition to a growing *P. furiosus* culture as determined by DNA microarray. *In vitro* transcription with two of these ORFs demonstrated that SurR is both a transcriptional activator and repressor. The X-ray crystal structure revealed the presence of a disulfide bond at the CxxC motif in the N-terminal DNA binding domain. This motif was found to act as a redox switch sensitive to colloidal sulfur, such that oxidation of SurR abolishes its sequence specific DNA binding affinity. The work presented here demonstrates that this novel transcriptional regulator, termed SurR for Sulfur-response regulator, is likely a relevant participant in transcriptional regulation pathways related to *P. furiosus* S^0 metabolism.

INDEX WORDS: Archaea, *Pyrococcus furiosus*, transcription factor, transcriptional regulation, DNA microarray, elemental sulfur response, SELEX, fluorescence footprinting, redox switch, disulfide switch, CxxC motif, colloidal sulfur, hydrogenase, winged helix-turn-helix motif, SurR.

A NOVEL REDOX-SENSITIVE TRANSCRIPTIONAL REGULATOR INVOLVED IN *PYROCOCCUS FURIOSUS* SULFUR RESPONSE

by

GINA LYNETTE PRIES LIPSCOMB

B.S., Clemson University, 2001

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

2007

A NOVEL REDOX-SENSITIVE TRANSCRIPTIONAL REGULATOR INVOLVED IN *PYROCOCCUS FURIOSUS* SULFUR RESPONSE

by

GINA LYNETTE PRIES LIPSCOMB

Major Professor: Robert A. Scott

Committee: Michael W. Adams
Timothy R. Hoover

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2007

DEDICATION

To my parents Stan and Charlotte Pries and sister Andrea for their unwavering support and unconditional love, to my husband and best friend David who has made my life so sweet and taught me so much, and most of all, to my creator God who has given me my life and abilities; He is worthy of all glory:

For the LORD gives wisdom, and from His mouth come knowledge and understanding.

Proverbs 2:6

By wisdom the LORD laid the earth's foundations, by understanding He set the heavens in place; by His knowledge the deeps were divided, and the clouds let drop the dew.

Proverbs 3:19-20

ACKNOWLEDGEMENTS

There are so many who have contributed to my personal development as a scientist and to the research presented in this finished work in so many ways. As my good friend and fellow scientist Meiyao once pointed out to me, scientific research is a never-ending process: when at first you don't succeed, re-search, re-search, re-search! And without the help and guidance of people much more knowledgeable and more experienced than I, this work would not have been possible.

First I thank the faculty of the Biochemistry and Molecular Biology Department and especially those who gave particular unforgettable words of encouragement in my first year of graduate school: Dr. Schmidt, Dr. Lanzilotta, Dr. Terns, and Dr. Puett; and Dr. Westpheling of the Genetics Department. Thanks to Matt Renfrow who originally inspired me to join the Scott lab and gave much useful advice and direction during my first few months of research. Thanks to my major professor Dr. Scott who has been a very dependable mentor throughout my years of research, who encouraged me in my ideas and gave me space to exercise them and to spread my wings, so to speak. Thanks to Dr. Eidsness who was there from the beginning with much needed encouragement during the particular times I most needed it. Thanks to both my committee members Dr. Adams and Dr. Hoover who have taken the time to challenge and mentor me in my scientific development. I also thank the Chemistry Department which in many ways adopted me as one of their own.

To those who through their help have made so much of my research possible, many thanks: to Meiyao Wang my good friend and lab mate who trained me in the lab early on, to

Frank Jenney and Gerti Schut who gave me invaluable advice and guidance at the very outset of my project and continued to support me throughout my work, to Dr. Eidsness who taught me much about protein expression and purification, to Jake Shokes who showed me how to use the FPLC for protein purification, to Dr. Phillips of the mass-spectrometry facility who spent much time helping me run my many samples, to Mohammed Ouhammouch and the Geiduschek lab of UCSD who greatly motivated me and gave me valuable training and protocols without which I would not have been able to accomplish the SELEX and footprinting work, to Darin Cowart who helped me in sorting through the SELEX data and provided the database to search for other potential SurR binding sites, to Annette Keese of the University of Regensburg for collaborating with me on this project and providing valuable functional data on SurR, to Frank Sugar for training me how to use the ÄKTA and giving me very useful advice and help with protein purification and column chromatography, to Hua Yang for a very enjoyable collaboration which resulted in getting crystals and structures for SurR and AxxA-SurR so quickly, and to the many others who I have not mentioned but have helped me so much!

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

TABLE OF CONTENTS

Page

CHAPTER 1

TRANSCRIPTIONAL REGULATION IN ARCHAEA AND APPROACH TO TRANSCRIPTION FACTOR DISCOVERY

1.1 Introduction to archaea

The many unique features of archaea have earned them the distinction of their own domain in the classification system of all known life on earth [1, 2]. Within the archaeal domain, however, a vast diversity of morphologies, habitats, and metabolic features exist. Archaea have often been called the 'extremophiles', and rightly so, since they are often found in the world's aforethought uninhabitable environments: deep-sea volcanic vents, solfataric fields, boiling springs, saturated salt seas, Antarctic lakes, acid and alkaline waters, etc. [3-5]. However, they are also one of the more common life forms on earth, inhabiting the soil, seawater, and even the gut of humans and animals [3, 4]. Based on this variety of characteristics alone, archaea are becoming increasingly studied, but there is an even more appealing reason to learn more about these organisms. They were formerly classified as bacteria, but although morphologically and metabolically they largely resemble bacteria, their information-processing machinery—translation, transcription, and replication pathways—has more similarities with eukaryotes [6]. Research in this area is valuable for two reasons. The first is for pure discovery science: how can an organism combine and mesh both prokaryotic and eukaryotic elements into a functional unit? The second is that investigating comparable but simpler pathways in archaea will lend insight to the more complex eukaryotic systems which are inherently more difficult to study.

The undertaking of the work presented here is primarily for the purpose of learning more about these unique organisms, particularly their transcriptional apparatus and the regulation of this process, with the long-term application that this knowledge will further the understanding of more complex life forms in the eukaryotic domain (although admittedly, the function of a single-celled organism has seemingly infinite complexity!).

1.2 Archaeal transcription

Transcription is the process by which DNA within a cell is transcribed into RNA, predominantly for the end-purpose of protein synthesis via the process of translation: messenger RNA as the protein synthesis template, transfer RNA for use in template codon recognition, ribosomal RNA for the ribosome—the workhorse in protein synthesis. In eukaryotes, there are three multisubunit DNA-dependent RNA polymerase (RNAP)[1] enzymes which catalyze each of these classes of transcription; however, archaea have only one RNAP to fulfill the diverse transcription needs of the cell, not unlike bacteria in this respect [7]. The archaeal RNAP is most like eukaryal RNA Polymerase II (Pol II) which is responsible for transcription of messenger RNA. Archaeal RNAPs are made up of 11 to 12 subunits (B'/B'', A', A'', D, E', F, L, H, N, K, P), each with sequence homology to the corresponding eukaryotic counterparts in the Pol II complex [7]. The bacterial RNAP is simpler in composition, comprised of five subunits: αI, αII, β', β'', and ω [8]. These five bacterial RNAP subunits have paralogs in both archaeal and eukaryal RNAPs, and together, these constitute a core in the overall architecture of RNAPs, though the archaeal and eukaryal RNAP cores are more homologous to each other than each is to the bacterial RNAP [7, 9]. Bacterial RNAP also requires a σ factor, which together with its five-subunit core constitutes the 'holoenzyme' capable of promoter specificity and transcription initiation [8].

[1] Appendix Table A1 lists abbreviations used in this work.

The fully functional eukaryotic transcriptional apparatus is complex, requiring the 12-subunit Pol II enzyme, at least five general transcription factors having 26 total subunits (TFIIB, -D, -E, -F, and -H), and the Mediator complex composed of at least 21 subunits [10]. In yeast, at least 44 of these proteins/subunits must be present for functional *in vitro* transcription [11]. The archaeal system, by comparison, is much more simplified in that it requires only two general transcription factors coupled with RNAP to achieve basal transcription *in vitro*; these are TBP (TATA-binding protein) and TFB (Transcription Factor B), homologs of eukaryotic TBP—part of the larger complex of TFIID also containing at least 10 TBP-associated factors—and TFIIB, respectively [12]. A third general transcription factor TFE (Transcription Factor E), homologous to the α-subunit of eukaryal TFIIE, is sometimes necessary to drive transcription from weak promoters [13]. Figure 1.1 illustrates the similarities and differences in RNAP subunits and general transcription factors among the three domains of life.

Not surprisingly, archaeal promoters contain the same major core elements as eukarya necessary to direct the basal transcriptional machinery: the AT-rich TATA box centered at 25 bases upstream from the translation start site and the 7 bp purine-rich BRE (TFB Recognition Element) just upstream of the TATA box [14]. These DNA elements serve as the binding sites for TBP and TFB, respectively, causing them to be correctly positioned and oriented to recruit RNAP to the transcription start site. The eukaryal TATA box has some sequence variation, but in general, can be described with the consensus sequence of TATA(A/T)AA(G/A) [15]. The archaeal element is less well-defined, with variations of the consensus sequence existing in different classes of archaea (Table 1.1). Early studies on TFB from *Sulfolobus shibatae* determined that its sequence-specific contacts upstream from the TATA box influenced the efficiency of transcription from the core promoter, and an *in vitro* selection strategy revealed

that, in particular, purines at positions -3 and -6 relative to the TATA box start were important for TFB recognition and promoter strength [16]. This study also concluded that base positions downstream from the TATA box may also be contacted by TFB. A summary of the suggested consensus sequences for the archaeal BRE and TATA boxes is listed in Table 1.1. In contrast to these eukaryal-type elements, the bacterial core promoter consists of 5 to 6 bp sequences centered at -10 and -35 relative to the transcription start site, and specificity to these elements is directed by a σ factor [17]. Figure 1.2 summarizes the similarities and differences among core promoters of the three domains of life.

Table 1.1 Archaeal BRE and TATA consensus sequences

Archaeal class / group	TATA box[a]	BRE[a]	References
Halophiles	-29 TTTWWW -24	none predicted	[18]
Methanogens	-30 YTTATATA -23	none predicted	[18]
Sulfolobus	-30 YTTTTAAA -23	-36 RNWAAW -31	[18, 19]
Pyrococcus	-29 TTWWWAW -23	-36 VRAAA -32	[20][b]

[a] Numbers indicate the position of the motifs relative to the transcription start site. Ambiguous nucleotides are represented according to the IUPAC code as follows: W = T/A, Y = C/T, R = A/G, V = A/C/G, N = any base.
[b] This table was essentially reproduced from [20].

Archaeal transcription initiation is similar to that in eukarya; TBP and TFB bind cooperatively to the core promoter and together recruit RNA polymerase [14]. TBP locates the TATA box and binds to the minor groove, causing a ~65° bend in the DNA and allowing the cooperative binding of TFB [9, 21]. TFB binds to the BRE element via its C-terminal helix-turn-helix (HTH) domain and also interacts with TBP, thus stabilizing the TBP-DNA complex and determining the direction of transcription [19, 22]. This ternary complex of promoter, TBP and TFB then recruit RNAP to the transcription start site via contacts between the N-terminal zinc finger domain of TFB and RNAP, specifically subunit B in *Pyrococcus* or in the case of

Solfolobus, possibly subunit K [23-25] (Fig. 1.3). TFB and RNAP together define the transcription start site position, 25 to 30 bp downstream of the TATA box [10, 15]. TFE interacts with RNAP subunits E and F and possibly associates with DNA between the TATA box and transcription start site, thereby stabilizing the initiation complex [23].

One significant difference in the archaeal versus eukaryal RNAP is the lack of the C-terminal repeat extension found on Pol II which serves as an assembly platform for complexes which carry out transcription-related tasks including chromatin remodeling [26]. While there is no known singly-conserved class of chromatin proteins in the archaeal domain suggestive of a common genome compacting mode, there are at least seven families of chromatin proteins, one of which is homologous to the eukaryal histones [27]. Bacteria also contain chromatin proteins, but chromatin remodeling does not appear to function in bacteria; however, there is increasing evidence for post-translational modification of archaeal chromatin proteins which results in potential architectural rearrangement of chromatin, similar to that which occurs in eukaryotes [28]. In eukaryotes, transcription and chromatin remodeling are interconnected [29], and the presence of archaeal histone-like proteins and chromatin-modifying proteins points to an even closer relationship between archaeal and eukaryal transcriptional mechanisms.

1.3 Archaeal transcriptional regulation and regulatory transcription factors

Though basal transcription in archaea is most similar to the eukaryotic system, regulation of transcription is typified for the most part by the bacterial system, both in the sequence similarities of the regulators themselves to bacterial families and in their mode of action [30]. By far the most prevalent DNA-binding domain found in prokaryotes is the HTH fold with its many variations and sub-classes (e.g. winged HTH) [31], and sequence analyses have demonstrated

that the same is true for predicted archaeal regulators (at least for those that can be identified bioinformatically) [32]. Indeed, many bacterial families of transcriptional regulators have homologs in the archaea [33]. Of the handful of regulators that have been characterized in detail, most have been demonstrated to be transcriptional repressors (many of which are auto-regulatory, down-regulating transcription from their own gene) [26].

The governing factor for negative regulation of archaeal transcription appears to be the placement of the regulatory transcription factor's cognate DNA-binding site. Thus far, there are two mechanisms which characterize transcriptional repression in archaea: binding of a regulatory transcription factor can block either the binding of TBP and TFB to the TATA region or the recruitment of RNAP to the transcription initiation site [34]. It remains to be seen whether bacterial repression mechanisms that involve an already-bound RNAP, such as preventing open-complex formation or inhibiting promoter clearance [35], are also employed in archaea.

One of the first archaeal regulators to be characterized was MDR1 of *Archaeoglobus fulgidus* [36], a homolog of bacterial metal-dependent repressors DtxR and SirR. In the presence of certain metal ions, MDR1 binds to its recognition sequence which overlaps the transcription initiation site and represses transcription from its own gene by blocking recruitment of RNAP [36]. Another regulator, LrpA from *Pyrococcus furiosus*, utilizes the same mechanism to repress transcription from its own gene [37, 38]; it is a member of the bacterial Lrp/AsnC family of transcriptional regulators which appear to predominantly function as regulators of amino acid metabolism [39]. Other archaeal regulators from this protein family have also been characterized, but their modes of regulation and mechanisms vary. Lrs14 from *Sulfolobus solfataricus* represses transcription from its own gene by obstructing the binding of TBP and TFB because its binding site overlaps the BRE and TATA box [40].

Three other Lrp family putative regulators have been described in some detail; however, functional data are lacking to classify these as bona fide transcriptional repressors or activators: Sa-Lrp from *Sulfolobus acidocaldarius* [41], Ss-Lrp from *S. solfataricus* [42], and Ss-LrpB from *S. solfataricus* [43, 44]. Interestingly, Ss-LrpB in complex with its cognate DNA has been shown to form higher-order nucleoprotein structures dependent on the stoichiometries of protein and DNA used [44], and it appears that other archaeal members of this family, also known as feast/famine regulatory proteins, have the potential to form similar higher-order complexes according to their determined crystal structures [45, 46]. Although the modeled protein-DNA structure is reminiscent of the eukaryotic nucleosome [45], the DNA-binding specificity of the archaeal Lrp family regulators suggests that their potential to form large nucleoprotein assemblies (even hetero-complexes with different Lrp family regulators [46]) is for versatility in gene regulation and not DNA condensation. It remains to be seen how such regulatory structures may effect chromatin reorganization and if this serves as a means to exert transcriptional control [30].

Bacterial regulators of the Lrp family have been shown to respond to small-molecule effectors, typically amino acids [39], but the archaeal Lrp-like regulators studied thus far have unknown effectors, with the exception of LysM from *S. solfataricus*. In the absence of lysine, LysM binds to a site upstream of the BRE/TATA box in the promoter of a lysine-biosynthesis gene cluster and most likely activates transcription; in the presence of lysine, LysM has lower DNA-binding affinity *in vitro*, and correspondingly, the gene cluster has reduced expression *in vivo* [47].

Not all archaeal regulators characterized thus far are representatives of bacterial families; there are a handful which are archaeal-specific and do not appear to have any homologs either in

bacteria or eukarya. The most notable example is that of TrmB, first characterized from *T. litoralis* as a repressor of the trehalose/maltose ABC transporter gene cluster as well as its own gene [48]. The nearly identical system also exists in *P. furiosus*, presumably as a result of lateral gene transfer [49]. Multiple sugars have been shown to differentially affect the DNA-binding affinity of TrmB to two different promoters [50], and the recently determined structure of the C-terminal sugar-binding domain reveals a novel sugar-binding fold [51].

Since the discovery of TrmB, two related archaeal transcriptional regulators have also been described, TrmBL1 from *P. furiosus* [52] and Tgr from *Thermococcus kodakaraensis* [53]. TrmBL1 is especially intriguing, as it appears to be a global transcriptional regulator of genes involved in glycolysis and sugar transport [52]; interestingly, these genes were previously identified as part of a putative regulon due to the presence of a common motif in their promoter regions, now known to be recognized by TrmBL1 [20].

A concurrent study on Tgr, an ortholog of TrmBL1 recognizing the same DNA-binding motif, complements the *in vitro* work done on *P. furiosus* TrmBL1 with *in vivo* data obtained from the genetic system available in *T. kodakaraensis*. From microarray expression profiling experiments comparing a wild-type strain of *T. kodakaraensis* with a Tgr-deletion strain under glycolytic and gluconeogenic conditions, a model for Tgr (and correspondingly TrmBL1) regulation emerges in which Tgr both represses and activates target genes depending on the position of the recognition sequence in the promoter and on the needs of the cell under glycolytic or gluconeogenic growth conditions [53]. Although it remains to be shown experimentally that this protein can indeed activate transcription, this is the first instance that an archaeal transcriptional regulator has been shown indirectly to exert both positive and negative control in transcription. This phenomenon is quite common in bacterial systems, and is mainly dependent

on the position and orientation of the *cis* regulatory sequence relative to the core promoter elements or the conformation of the transcription factor under the influence of some effector [54]. It will be interesting to explore whether this dual functionality of regulatory transcription factors is a general rule rather than the exception for archaea as it appears to be in bacteria.

A second example of an archaeal-specific transcriptional regulator is NrpR from *Methanococcus maripaludis* which represses transcription from the nitrogen fixation (*nif*) operon, probably by blocking RNAP recruitment, and is released by the effector 2-oxoglutarate, an intracellular indicator of nitrogen shortage [55, 56]. NrpR homologs exist only in select members of the archaeal domain, particularly in methanogens which contain this nitrogen fixation system.

Two transcription factors from halophilic archaea have been studied in some detail *in vivo* using the *Haloferax volcanii* genetic system; *in vitro* systems in halophiles are difficult considering the high salt concentrations necessary (e.g., 2 M). GvpE and GvpD act in concert to control the transcription of gas-vesicle formation genes; genetic evidence suggests that GvpE activates transcription while GvpD may repress transcription, perhaps through an interaction with GvpE [57, 58]. In contrast to all the archaeal regulators that have been described thus far, GvpE is the first archaeal transcriptional regulator to show homology to a class of eukaryotic transcription factors having leucine-zipper (bzip) DNA-binding domains [59, 60].

One archaeal regulator for which the structure is known appears to have domain homologies to both bacteria and eukarya. Phr from *P. furiosus* has been shown to be a regulator of heat-shock response, repressing transcription from heat shock genes and its own gene by preventing the recruitment of RNAP [61]. Sequence analysis shows that the most similar homologs are from other archaeal species; however, the recently solved crystal structure shows

homology to both bacterial and eukaryal protein domains. The N-terminal DNA-binding domain has strong homology to the bacterial ArsR family of transcriptional regulators, and the C-terminal domain has a distant relationship with the eukaryotic BAG domain which in eukaryotic proteins is utilized to associate with certain heat shock proteins [62].

Transcriptional activation in archaea is poorly understood, for the main reason that the large majority of transcription factors studied thus far have been repressors; however, there is one activator whose mechanism has been studied in detail using an *in vitro* transcription system available for the thermophilic methanogen *Methanocaldococcus jannaschii*. This activator is Ptr2 from *M. jannaschii*, another Lrp homolog, which stimulates transcription of the *rb2* and *fdxA* genes by facilitating the recruitment of TBP to the promoter [63-65]. Close homologs to Ptr2 have also been analyzed in *Methanothermococcus thermolithotrophicus* and *Methanococcus maripaludis*, with only Ptr2 from the former showing comparable activation; however, mutating some of the residues of the *M. maripaludis* protein to match those of *M. jannaschii* Ptr2 conferred on it the ability to activate transcription [66]. A detailed report of the promoter architecture requirements for Ptr2-mediated transcriptional activation showed that the spacing between the upstream activating site bound by Ptr2 and the TATA box was critical for activation [67].

Similarly, bacterial activation can also be accomplished through recruitment of basal transcriptional machinery, but other modes of activation also exist, such as stabilizing the pre-initiation complex or influencing promoter melting and/or clearance [54]. These functions are typically carried out from a binding site upstream from the core promoter, and are mediated most often through interactions with either the σ subunit (most commonly σ^{70}) or the C-terminal domain of RNAP [68]. Considering the many activation mechanisms utilized by bacterial

promoters, it is of special interest to analyze additional archaeal activation mechanisms in detail for further comparisons. At this point, it is clear that the mechanism of archaeal activation favors those of bacterial systems and not the complexities of eukaryotic activation involving multisubunit chromatin remodeling machines and a concert of activators which many times must be first activated themselves [29, 69].

Table 1.2 summarizes the functionally-characterized archaeal transcriptional regulators and Figure 1.4 illustrates what is known about the modes of archaeal transcriptional activation and repression.

Table 1.2 Summary of functionally-characterized archaeal transcriptional regulators

Name	Species	Effector(s)	Mode of regulation	Reference(s)
GvpE	*Halobacterium salinarium, Halobacterium mediterranei*	Unknown	Putative activator; mechanism unknown	[58, 70]
LrpA	*P. furiosus*	Unknown	Repressor; blocks RNAP recruitment	[37, 38]
Lrs14	*S. solfataricus*	Unkown	Repressor; blocks TBP/TFB binding	[40]
LysM	*S. solfataricus*	Lysine	Putative activator; mechanism unknown	[47]
MDR1	*A. fulgidus*	Metals ions: Fe^{2+}, Mn^{2+}, Ni^{2+}	Repressor; blocks RNAP recruitment	[36]
NrpR	*Methanococcus maripaludis*	2-oxoglutarate	Repressor; probably blocks RNAP recruitment	[55, 56]
Phr	*P. furiosus*	Unknown	Repressor, blocks RNAP recruitment	[61, 62]
Ptr2	*M. jannaschii, M. thermolithotrophicus*	Unknown	Activator: recruits TBP	[63, 66]
Sta1	*Sulfolobus islandicus*	Unknown	Activator (from viral promoter), mechanism unknown	[71]
Tgr	*T. kodakaraensis*	Unknown	Repressor; probably blocks TBP/TFB binding. Putative activator, mechanism unknown.	[53]
TrmB	*T. litoralis, P. furiosus*	Maltose, trehalose, sucrose, maltodextrins	Repressor; probably blocks TBP/TFB binding	[48, 50]
TrmBL1	*P. furiosus*	Unknown	Repressor; probably blocks TBP/TFB binding	[52]
TrpY	*M.thermoautotrophicus*	Tryptophan	Repressor, blocks TBP/TFB binding	[72]

1.4 The model archaeon *Pyrococcus furiosus*

One of the model archaeal organisms that has emerged for focused study is the hyperthermophile *Pyrococcus furiosus*. Hyperthermophiles comprise a group of organisms that grow optimally at or above 80°C (and do not grow below 50°C, in general), including 34 genera from archaeal and and bacterial domains [73]. *P. furiosus*, originally isolated by Fiala and Stetter from a hydrothermal vent community off the coast of Italy [74], is a member of the genus *Thermococcus* which currently contains 24 species of obligately organotrophic fermentative anaerobes [75]. In cell culture, *P. furiosus* has an optimal growth temperature of 100°C and can utilize both carbohydrates and peptides as carbon sources, via fermentation to CO_2 and H_2 [74]. Most hyperthermophiles are strictly anaerobic and chemolithoautotrophic, and many are dependent on the availability of sulfur for growth [73, 75]. *P. furiosus* is somewhat unique among other archaeal hyperthermophiles in that it can grow in the presence or absence of elemental sulfur (S^0), depending on the available carbon source [74, 76, 77]. This trait presents an opportunity to study the regulation of sulfur metabolism by investigating the changes which occur in the absence or presence of sulfur in a growing culture.

1.5 Method of transcription factor discovery and characterization

The high-throughput approaches made possible by DNA microarray technology coupled with the availability of whole-genome sequences have provided the ability to monitor the changes in gene expression for an organism as a result of a response to some change or stress in its environment—in the laboratory, this would correspond to a controlled change in culture growth. For prokaryotes, much regulation of gene expression occurs at the level of transcription,

and therefore, gene expression profiles which demonstrate changes in transcript levels between two growth conditions can be utilized in targeted transcription factor discovery.

The working hypothesis for this approach is that the 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 is the result of a regulatory transcription factor exerting its effect in response to the environmental change or stress. If this is the case, then the upstream DNA of the gene or operon, which contains the core promoter necessary for basal transcription, should also contain the binding sequence of a particular transcription factor responsible for regulating transcription. An affinity fractionation method can then be employed using this DNA sequence as 'bait' to capture the transcription factor from soluble protein extract derived from the growth condition for which the change in transcript expression was initially observed. The fractionated proteins are then visualized and separated by denaturing gel electrophoresis, and protein bands of interest are identified by mass spectrometry.

Identified proteins are prioritized for characterization based on the degree of sequence homology to transcription factor families and the presence of DNA-binding domains. The recombinant protein is then obtained, and the same upstream/promoter DNA used in the protein-capture experiment is used in an electromobility shift assay (EMSA) with the protein to verify sequence specific DNA binding as compared to a control 'non-promoter' or ORF DNA. Once a protein is verified to be specific to a particular gene promoter region, the putative transcription factor is then further characterized. This work details the identification and characterization of a regulatory transcription factor involved in *P. furiosus* sulfur response, termed SurR for Sulfur-response Regulator.

Figure 1.1 Basal transcriptional machinery of the three domains. RNA polymerase subunits (top) and general transcription factors (bottom) are color-coded according to their homologies across domains and protein/subunit names are indicated. Only those transcription factors which are necessary for *in vitro* basal transcription are shown. Figure adapted from [7].

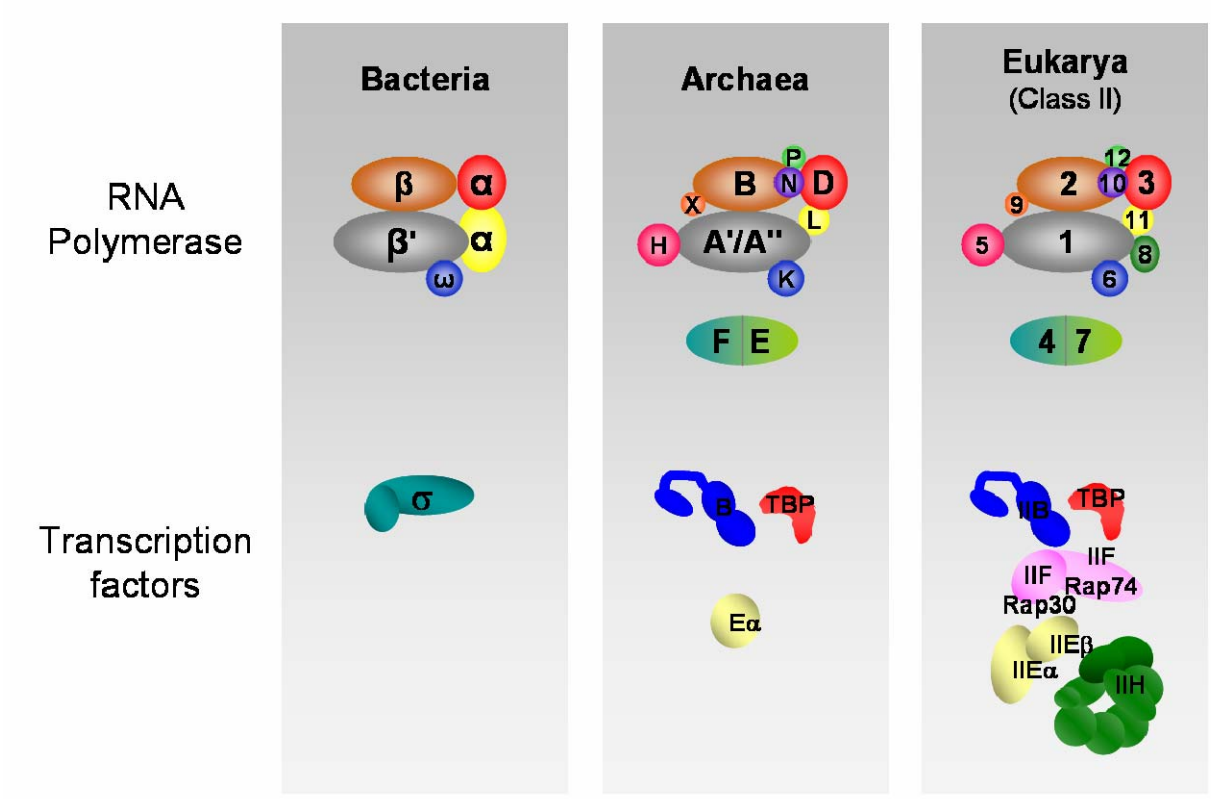

RNA Polymerase
Transcription factors
Bacteria
β
α
β'
α
ω
σ
Archaea
P
B
N
D
X
L
H
A'/A"
K
F
E
B
TBP
Eα
Eukarya
(Class II)
12
2
10
3
9
11
5
1
8
6
4
7
IIB
TBP
IIF
IIF
Rap74
Rap30
IIEβ
IIEα
IIH

Figure 1.2 Core promoter elements of the three domains. The position of each element relative to the transcription start site is indicated by the scale at the top of the figure. There are no distance indicators between the transcription start site and the translation start site because this distance varies according to gene context. Note that there is no zero position; instead, the nucleotide immediately upstream of the transcription start site (+1) begins at -1. Sequence elements are shown with boxes and are indicated by their abbreviations: BRE, TFB Recognition Element; TATA Box, TBP Binding Element; -10 and -35 boxes, bacterial promoter elements. Archaeal promoter elements are boxed with dotted lines to indicate minor differences in sequence consensuses as compared to the related elements in eukarya which are shaded with the same colors. Figure adapted from [18].

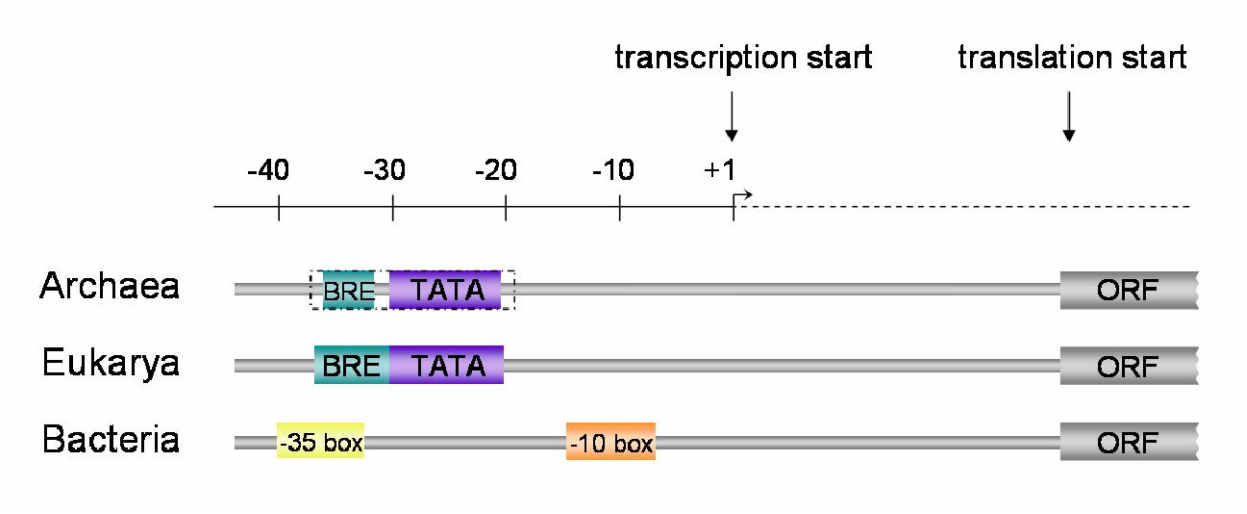

transcription start
translation start
-40
-30
-20
-10
+1
Archaea
BRE
TATA
ORF
Eukarya
BRE
TATA
ORF
Bacteria
-35 box
-10 box
ORF

Figure 1.3 Process of archaeal transcription initiation. An archaeal core promoter (top) consists of a BRE (teal box), TATA box (purple box), and transcription initiation site (bent arrow). TBP (purple) first binds to the TATA box, causing a sharp bend in the DNA, and TFB (teal) stabilizes the ternary complex (middle). This ternary complex of TBP, TFB and promoter DNA recruits the 12-subunit RNAP to the transcription start site (bottom), forming the pre-initiation complex. Adapted from [30].

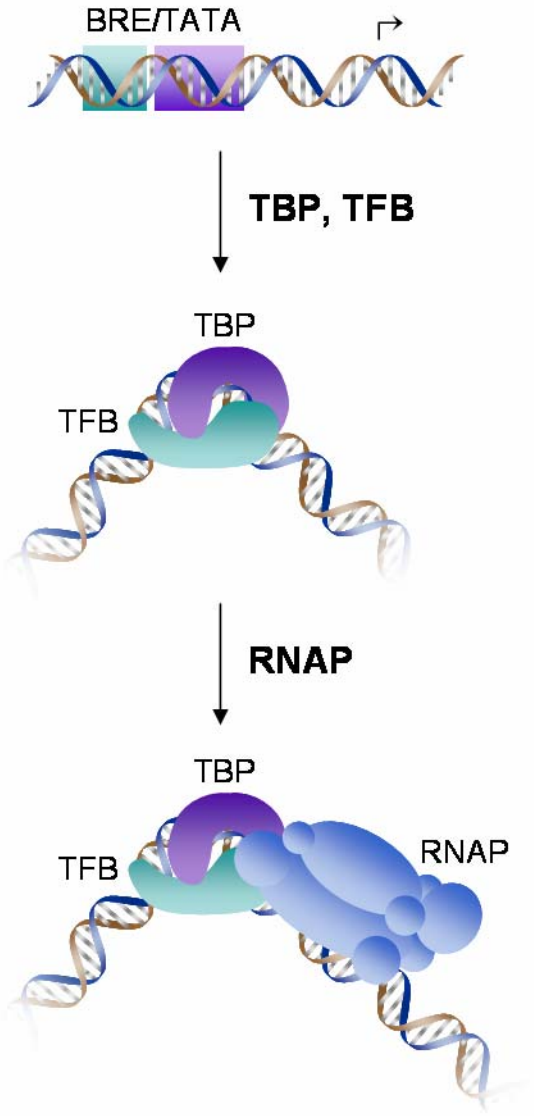
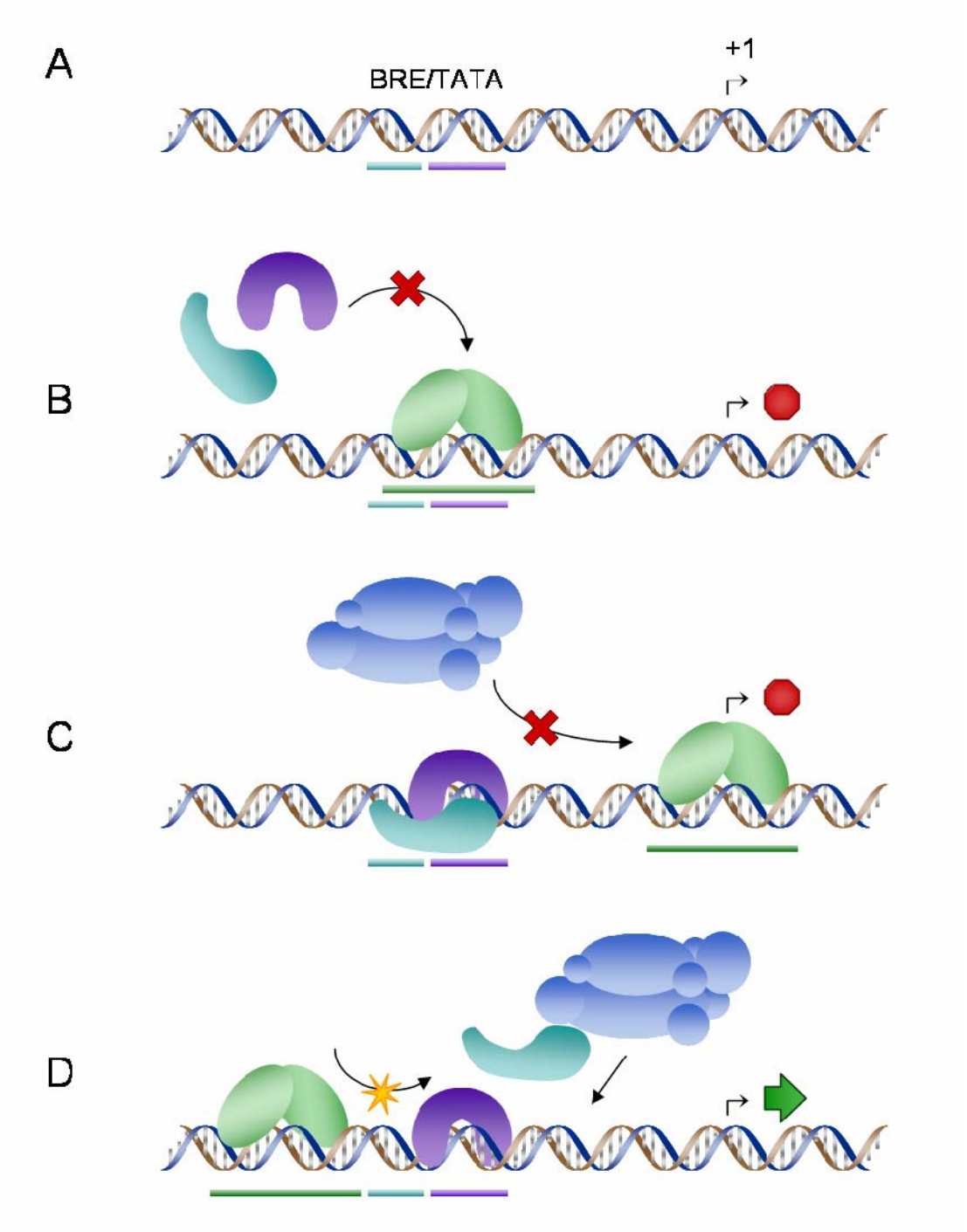

BRE/TATA
TBP, TFB
TBP
TFB
RNAP
TBP
TFB
RNAP

Figure 1.4 Archaeal mechanisms of transcriptional repression and activation. A. An archaeal core promoter consisting of a BRE (teal bar), TATA box (purple bar), and transcription initiation site (bent arrow with +1 designation). **B-D.** Diagrams depicting archaeal transcriptional regulatory mechanisms with the regulatory transcription factor shaded in green (here depicted as a dimer), TBP shaded in purple, TFB shaded in teal, and their corresponding binding sties indicated by bars of the same color below the DNA. The 12-subunit RNAP is shaded in blue. Transcription from a core promoter can be repressed (represented by a red octagon) by two mechanisms: binding of the regulatory transcription factor to a binding site overlapping (**B**) the BRE/TATA region thereby preventing the binding of TBP and subsequently TFB and (**C**) the transcription initiation site thereby preventing the recruitment of RNAP by the TBP/TFB/DNA ternary complex. Transcription activation (represented by a green arrow) has been shown to occur through at least one mechanism, (**D**) facilitating the recruitment of TBP to the core promoter from a binding position upstream of the TATA box. This facilitation speeds up the entire process of transcription initiation, thereby increasing the rate of transcription from the promoter. Note that the bend caused by the binding of TBP to the TATA box is not illustrated in this figure.


A
BRE/TATA
+1
B
C
D

CHAPTER 2

MATERIALS AND METHODS

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

Two 15-L cultures of *P. furiosus* (DSM 3638) were grown anaerobically at 95°C essentially as in [78], using maltose as the carbon source with addition of S^0 to only one of the cultures. After approximately 11 hours of growth, cells were cooled by pumping the culture through a coiled tube in ice water, concentrated by ultrafiltration to 2-3 L, and harvested by centrifugation at 10,000 g for 15 min. The cell pellets were resuspended in 10 mM EPPS buffer (pH 8.0) and lysed by sonication for 20-30 min on ice. The sonication time was extended to account for not having added DNase I to aid in digesting the genomic DNA; addition of DNase I would have interfered with the down-stream application of the cell extract in the DNA affinity protein capture experiment (Section 2.2). Lysate was centrifuged for 15 min at 10,000 g to remove cell debris. Soluble cell extract was then obtained after centrifugation at 100,000 g to remove remaining insoluble and membrane materials. Soluble cell extract was aliquoted into anaerobic vials and stored at -80°C.

2.2 DNA affinity protein capture

The DNA probes used for protein capture were PCR-amplified from *P. furiosus* genomic DNA using the primers[1] listed in Table 2.1. The biotinylated probe was bound to magnetic

[1] All primers were purchased from Integrated DNA Technologies with the exception of the NED-labeled primers described in Section 2.8.

DynaBeads M-280 Streptavidin (Invitrogen) per the manufacturer's protocol. The bead-bound DNA was then mixed with 2.5 mg/mL *P. furiosus* soluble cell extract from cells grown either in the presence or absence of S^0 and incubated at 55°C for 30 min with intermittent mixing to keep the beads in suspension. Unbound proteins were eluted with three washes of Buffer B (50 mM EPPs, 100 mM KCl, 1 mM EDTA, 5% glycerol, 0.1% triton-X, 1 mM DTT, pH 7.5). DNA-bound proteins were eluted at 55°C for 5 min with 1x Laemmli buffer containing no β-mercaptoethanol, as this tended to strip the streptavidin from the bead surface. Eluted proteins were analyzed by SDS-PAGE with silver staining.

Table 2.1 Probes used in DNA affinity protein capture

Probe name	Genome coordinates	Forward primer[a] (5' biotinylated)	Reverse primer[a]	Probe length (bp)
mbh1	1337201-1337513	ccctaacttggtggtgccta	ccagctatgagttcctgggta	314
sipB	1872258-1872623	tctgcaagtgtgcgctttac	ccctttcatcgacttcagc	366

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

2.3 In-gel tryptic digestion and peptide mass mapping

Bands of interest in SDS-PAGE lanes of eluted proteins from DNA affinity protein capture were excised and subjected to in-gel tryptic digestion. The gel slices were subjected to three cycles of hydration and dehydration: a 10-min incubation in 25 mM ammonium bicarbonate followed by a 15-min incubation in 50% acetonitrile in 25 mM ammonium bicarbonate. The dehydrated gel slices were then completely dried in a vacuum centrifuge (~10 min), after which they were rehydrated in a ~10-µL solution of 10 ng/µL trypsin in 25 mM ammonium bicarbonate for overnight digestion at 37°C. After rehydration in trypsin a small amount 25 mM ammonium bicarbonate was added to cover the gel slices. Digested peptides

were extracted with one ~10-µL wash of 25mM ammonium bicarbonate followed by two successive ~10-µL washes with 75% acetonitrile, 0.5% trifluoroacetic acid (TFA). The pooled, extracted solution was concentrated to 4-5 µL by vacuum centrifugation, and 1 µL 5% TFA was added to make a final volume of 5-6 µL and a final TFA concentration of 0.1-1%. A ZipTip (Millipore) 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, [79]) to search the NCBI database of archaeal genomes and a local server hosting ProteinProspector [80] to search a *P. furiosus* genome database.

2.4 Sequence analysis for selection of target protein for characterization

Sequence analysis of identified proteins was performed using NCBI BLAST [81, 82] and Conserved Domain searches [83-85] against the NCBI non-redundant protein database.

2.5 Expression and purification of recombinant his-tagged SurR

The vector containing the PF0095 sequence encoding SurR (genome coordinates 103332-104030) was a kind gift from Francis Jenney (from the laboratory of Michael Adams, University of Georgia). The pET24dBAM vector harboring the clone was a derivative of pET24d, modified to incorporate an N-terminal hexahistidine tag (his-tag) on the expressed protein (e.g., as in [86]). The pET24d vector is selectable for kanamycin resistance and is designed for use in combination

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

For expression of the his-tagged recombinant protein, the clone was transformed into BL21-CodonPlus(DE3)-RIPL cells (Stratagene) using the manufacturer's protocol. Protein expression from a 1-L culture of LB media was grown to an OD_{600} of ~0.9-1, and protein expression was induced with 0.4 mM IPTG. Cells were harvested 4 hours after induction and resuspended in ~20 mL Binding Buffer (20 mM sodium phosphate, 0.5 M NaCl, pH 7.4) containing ~10 μL of protease inhibitor cocktail (Sigma). Cells were sonicated on ice using a large horn at 40% power for 6 pulses of 30 seconds, 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 prior to purification of the protein by column chromatography.

Using an automated FPLC system (GE Healthcare), the soluble cell extract was loaded onto a 1-mL HiTrap metal affinity column (GE Healthcare) preloaded with nickel-sulfate per the manufacturer's instructions. The column was washed with 5 mL Binding Buffer containing 20 mM imidazole followed by a gradient elution with Eluting Buffer (20 mM sodium phosphate, 0.5 M NaCl, 0.5 M imidazole, pH 7.4) first with a 10-mL 0-20% gradient and then with a 10-mL 20-100% gradient. Protein-containing fractions that were relatively pure were pooled, and a 5-mL desalting column (GE Healthcare) was used for buffer exchange into 20 mM HEPES, 100 mM

NaCl, pH 7.6. Resulting protein was estimated to be >95% pure. Protein concentration was determined using a Bio-Rad DC Protein Assay kit, and aliquots of his_6-SurR[1] were stored at -80°C.

2.6 Electromobility shift assay with SurR

Electromobility shift assay (EMSA) was performed as a modification from that originally described in [87]. DNA probes for EMSA were PCR-amplified from *P. furiosus* genomic DNA using primers listed in Table 2.2 or from pUC18-cloned promoter-ORF DNA using M13 Forward and Reverse sequencing primers (Section 2.7), followed by either PCR-purification using a Qiagen PCR Purification Kit, gel-purification using a Qiagen Qiaquick Gel Purification Kit, or ethanol-precipitation. EMSA reactions of DNA with various amounts of protein were set up in 10-µL volumes in EMSA buffer (20 mM HEPES, 200 mM KCl, 5% glycerol, 1 mM EDTA, pH 7.5) using a 5x stock. DNA concentration was typically 25-75 ng/µL in each reaction, and protein was adjusted according to the molar amount of DNA. The EMSA reactions were assembled as follows. A master mix of water, 5x EMSA buffer and DNA was made according to the number of reactions in the experiment (typically 14 total, to be loaded into a 15-well gel with one gel lane reserved for a DNA marker), then distributed to 0.5-mL microcentrifuge tubes on ice. Protein dilutions were made in a final concentration of 1x EMSA buffer, and 2 µL of the appropriate protein dilution was added to each EMSA reaction (with 2 µL of 1x EMSA buffer added instead of protein for the DNA-only lane). In cases where an extra reagent was added to the reaction (e.g. diamide, colloidal sulfur), volumes of water and/or 5x EMSA buffer were adjusted accordingly such that the final buffer concentration of each reaction was always 1x.

[1] Throughout this work, the recombinant SurR protein which contains a his-tag will be noted as 'his_6-SurR' to distinguish it from the cleaved-tag construct described in section 2.10 which will be denoted as 'SurR'.

Reactions were incubated at 55 or 70°C for 20 min and immediately loaded onto a BioRad 5% TBE gel; 15-well Ready gels were typically run at 200V for 20-30 min while the 26-well Criterion gels were typically run at 100V for 60-110 min. The gel was then stained with SYBR Green Nucleic Acid Gel Stain (Invitrogen) according to the manufacturer's instructions. SYPRO Ruby protein gel stain (BioRad) was used according to the manufacturer's instructions to verify the presence of protein as necessary. Gels were imaged via UV transillumination.

Table 2.2 Probes used in EMSA

Probe name	Genome coordinates	Forward primer[a]	Reverse primer[a]	Probe length (bp)
mbh1	1337201-1337513	ccctaacttggtggtgccta	ccagctatgagttcctgggta	314
sipB	1872258-1872623	tctgcaagtgtgcgctttac	ccctttttcatcgacttcagc	366
lrpA	1493485-1493784	tctacaacatagcaaaggag	cagtctcactaatgcctaat	300
mbh1 promoter	1337338-1337418	gccaatacgaatttggagagagg	gacattcgccaaacctcctt	81
mbh1 ORF	1337435-1337514	atatgggcttacttcctctggc	ccagctatgagttcctgggta	80
sur-pdo intergenic	103197-103334	cattttcctcacccctatc	catattcatcacctacccatatc	138

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

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

Promoter-ORF DNA fragments were cloned from *P. furiosus* genomic DNA into pUC18[1] for versatile PCR-amplification of probes for EMSA, footprinting and *in vitro* transcription experiments. The design of the DNA fragment to be cloned was primarily optimized for use as a transcription template, and secondarily as a footprinting probe. For a transcription template, it was necessary to include at least ~100 bp of the ORF to allow for an adequate transcript in an *in vitro* transcription assay; for footprinting, it was necessary to center the promoter region in the DNA probe and have a total length of ~300 to 500 bp to allow for satisfactory visualization of a

[1] GenBank/EMBL accession number: L09136

protein footprint. Also, many of the DNA regions to be cloned contained divergently-transcribed ORFs with intergenic spaces of 260 bp or less, leading to the possibility of two resulting transcripts in an *in vitro* transcription assay. This being the case, it was then necessary to design the promoter-ORF DNA fragment in such a way that resulting transcripts from two divergently transcribed ORFs on the same template would be of sufficiently different sizes in order to be distinguishable on a transcription gel. In general, this meant that the length of each ORF in the clone had to differ by at least 50 bp.

Sequences of *P. furiosus* genomic DNA to be cloned were obtained from TIGR[1]. Primers were designed to include approximately 20 bp of DNA complementary to the region to be cloned and also the appropriate restriction sites at the 5' ends with an additional ~5 random bases to ensure complete restriction digestion of the PCR product ends before cloning. Restriction maps of the genomic DNA sequences to be cloned were obtained using online software available from the Sequence Manipulation Suite[2] to determine compatible restriction sites for cloning into the standard cloning vector pUC18 (Stratagene). Restriction enzyme pairs were chosen for their compatibility in reaction and buffer requirements since both insert and vector would require digestion with two different restriction enzymes. Cloning primers for each promoter-ORF fragment are listed in Table 2.3.

The multiple cloning site for pUC18 lies within the reading frame of the *lacZ* gene encoding β-galactosidase which can cleave the synthetic substrate analogue X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside), yielding galactose and the blue chromophore 4-chloro-3-brom-indigo which causes colonies that contain the plasmid to turn blue in the presence of X-gal. Insertion of DNA into the multiple cloning site will disrupt the gene reading frame and

[1] The Institute for Genomic Research (TIGR) website: www.tigr.org
[2] Sequence Manipulation Suite website: http://www.ualberta.ca/~stothard/javascript/

cause colonies which contain the plasmid to remain white even when X-gal is present in the agar plates. Therefore, the selection process for clones that contain inserts consists of blue/white color screening of colonies grown on plates containing X-gal, along with IPTG which drives expression of the plasmid-encoded *lacZ* gene.

DNA for cloning was PCR-amplified from *P. furiosus* genomic DNA using a high-fidelity Pfu DNA polymerase (Stratagene) and purified using a Qiagen Qiaquick PCR Purification Kit to remove excess dNTPs, primers and proteins. pUC18 plasmid DNA was amplified and purified from XL1-Blue cells (Stratagene) to obtain a sufficient amount for cloning. Digestion reactions were set up with ~1 μg of pUC18 DNA per clone to be produced using the appropriate restriction enzymes (EcoRI/HindIII or SacI/SphI) and incubated at 37°C for 1-3 hours. The digestion reaction was then run on a 1% agarose gel, bands of linearized vector were excised, and DNA was purified using a Qiagen QIAEX II Gel Extraction Kit. Meanwhile, ~100 ng of each PCR-amplified cloning fragment was digested with the appropriate restriction enzymes for 2-3 hours at 37°C. Restriction enzymes were inactivated following the digestion reaction by incubation at 65°C for 20 min. The ligation reactions were set up in 20-μL volumes using 100 ng of linearized, gel-purified pUC18 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 (Promega). Ligation reactions were incubated at room temperature for 3 hours and immediately transformed (using 1 μL of the ligation reaction) by heat-shock into $CaCl_2$-competent XL1-Blue cells (made from a stock obtained from Stratagene). Transformation cultures were plated onto LB-agar plates containing 100 μg/mL ampicillin, 0.5 mM IPTG, and 80 μg/mL X-gal. Plates were incubated at 37°C for 18 hours followed by chilling at 4°C to intensify the blue color for blue/white colony screening.

White colonies were tested for the presence of the insert using colony PCR. Briefly, 10-µL PCR reactions were set up in 0.2-mL PCR tubes with M13 Forward and Reverse sequencing primers, complementary to the DNA on either side of the multiple cloning site (to amplify the insert): 5'-gtaaaacgacggccagt and 5'-caggaaacagctatgac, respectively. A sterile 10-µL pipette tip was used to introduce a tiny amount of cells from the colony into the 10-µL PCR reaction; in this way, the colonies could be tested for the presence of a positive clone without first having to amplify and purify plasmid from liquid culture. Colony PCR reactions were then run on 1% agarose gels to check the insert size. Clones with correctly sized inserts were selected for plasmid amplification in 5-mL liquid cultures (purified by Qiagen Qiaquick Plasmid MiniPrep Kit) and for sequence verification by primer extension (Sequetech, Inc, Mountain View, CA).

Table 2.3 Promoter-ORF probe DNA for cloning into pUC18

Probe name[a]	Genome coordinates	Forward primer[b]	Reverse primer[b]	Length (bp)	Restriction site pair[c]
PF0094	103080-103499	cgtagaattcctgaagaac accttctctttcc	tagtcaagcttctggtcaca gtattgacagtg	441	EcoRI/HindIII
PF0531	547577-547992	cgtagagctcgatttctaag tgagtacgctag	tagtcgcatgcttgaacttct tcctccctaac	437	SacI/SphI
PF0559	575393-575712	cgtagaattcgtgaggccc agcaaataatag	tagtcaagcttaacgtatcc cctcaagttg	341	EcoRI/HindIII
PF0891	863554-863873	cgtagaattcactatattcta tcctacttcaac	tagtcaagcttaatctccct gaagtcatagaac	341	EcoRI/HindIII
PF1186	1132759-1133078	cgtagaattcagcaaagga agctgctcagg	tagtcaagcttcacccactc cgttgcctcg	341	EcoRI/HindIII
PF1423	1337173-1337562	cgtagaattccctaaactga acatgtcc	tagtcaagctttccgataat gtttctggttg	411	EcoRI/HindIII
PF1453	1359941-1360395	cgtgaattctcagttcttga ggatcaagac	tagtcaagctttccgtaacc tcgatttgc	475	EcoRI/HindIII
PF1911	1762713-1763180	aaggaattctttctccgttcg	atctcaagcttaactctctta ttctcctcgc	480	EcoRI/HindIII
PF2051	1892585-1892936	cgtagagctcagaggaac atcgctatattctc	tagtcgcatgcaacctactt atctccttggag	373	SacI/SphI

[a] Probes contain ORF and ORF-upstream DNA but are named for the SurR target ORF, although some contain DNA from divergently transcribed ORFs and their corresponding upstream DNA.
[b] DNA primers are listed from 5' to 3'.
[c] Restriction sites included in forward/reverse primers used for cloning of PCR-amplified DNA.

2.8 Fluorescence-detected DNase I and hydroxyl radical footprinting

Fluorescence footprinting was performed as a modification from [88] based on the DNase I footprinting method [89] and the hydroxyl radical (OH•) footprinting method [90, 91]. Footprinting probes were PCR-amplified from pUC18-cloned promoter-ORF DNA (Section 2.7) using 5' 6FAM and HEX labeled modified M13 primers for analysis on a 3730x1 automated DNA sequencer (Applied Biosystems). The standard M13 Forward primer had to be extended at the 5' end to ensure that the base adjacent to the fluorophore would not be a guanine, as guanine can quench the fluorescence of fluorophores at that proximity. Consequently, the standard M13 Reverse primer had to be lengthened to adjust the melting temperature (T_m) to be closer to the T_m of the modified M13 Forward primer. These modified M13 primers are listed in Table 2.4.

Since the footprinting probes were each amplified using the same primers, the annealing temperature used for the probe-amplifying PCR reactions was the same (55 °C). Probes were amplified using Taq or JumpStart Taq DNA polymerase (Sigma-Aldrich) for a maximum of 25 cycles to minimize the amount of non-specific PCR products. Typically, six 50-µL PCR reactions were performed for each probe, and these were then concentrated by ethanol precipitation prior to gel-purification. To purify away primers and truncated PCR products, probes were separated on BioRad 5% TBE gels (18-well Criterion gels run at 100V for 60-110 min were optimal). Gels were stained with SYBR Green I Nucleic Acid Gel Stain (Lonza) for 10 min prior to visualization under long-range UV light. Bands were carefully excised, with special attention to cutting the lower side of the band as close as possible to the bulk of DNA thereby eliminating shorter DNAs which would interfere with footprinting results. Probe DNA was eluted from the polyacrylamide gel slices using the crush-and-soak method [92]. Resulting probe

DNA was concentrated by ethanol precipitation and quantified prior to use in footprinting reactions.

Probe-specific 'A+G' and 'C+T' ladders were PCR-amplified from each pUC18 clone with NED-labeled modified M13 primers[1] using the Fermentas CycleReader™ Auto DNA Sequencing Kit, modifying the protocol to obtain ladders instead of sequencing products. Briefly, reactions were set up as follows. For each dual-labeled probe, ladders for each strand were PCR-amplified separately. Two mixes were assembled according to the number of probes for which ladders were to be amplified: one containing forward primer with A and G termination mixes and the other containing reverse primer with C and T termination mixes; the volumes per probe were 1 μL of each termination mix and 0.25 μL of 2 μM primer. A PCR mix of buffer (1.25 μL), nuclease-free water (5.25 μL), template (1 μL of ~200 ng/μL) and Taq polymerase (0.5 μL) was made for each probe and divided into two 0.2-mL PCR tubes. To each tube, 2.25 μL of the primer/termination mixes was added such that for each probe, an A+G ladder would be synthesized with the forward primer and a C+T ladder would be synthesized with the reverse primer. Following a 3-min incubation at 94°C, ladders were PCR-amplified according to the manufacturer's protocol using an annealing temperature of 56°C. Ladders were stored at -20°C until used in the footprinting reactions.

Footprinting reactions were composed of two parts, the protein-DNA binding reaction and the cleavage reaction. The protein-DNA incubations were set up similar to the EMSA reactions except the reactions were set up in 50-μL volumes with ~150-200 ng DNA probe, incubations of individual reactions were separated by 2 or 3-min intervals, and the buffers were different. The following buffers were used: for DNase I footprinting, 20 mM HEPES, 100 mM KCl, 15 mM $MgCl_2$, 5 mM $CaCl_2$, 1 mM EDTA, 1 mM DTT, 5% glycerol, pH 8; and for OH•

[1] NED-labeled primers for A+G and C+T ladder amplification were purchased from Applied Biosystems.

footprinting, 20 mM HEPES, 10 mM $MgCl_2$, 200 mM KCl, pH 7.5. Protein-DNA solutions were incubated for 20 min at 55°C or 70°C.

For the DNase I cleavage reaction, 0.03-0.05 U of DNase I (from a 0.01 U/μL dilution in 10 mM Tris, pH 8.0) was added to each 50-μL protein-DNA mixture and the solution was incubated for 1 min at room temperature. The reaction was stopped by adding 145 μL of Stop Solution (130 mM NaCl, 20 mM EDTA, 0.6% SDS) followed immediately by 200 μL of buffered phenol:chloroform:isoamyl alcohol (25:24:1) with vigorous vortexing. A+G or C+T ladder reactions were added to the appropriate digested DNA control tubes which were then vortexed again. For the OH• cleavage reaction, the cutting reagent was made as follows: a bead of 2 μL of 4% H_2O_2 was applied to the wall of the tube containing the protein-DNA mixture, 2 μL of 56 mM sodium ascorbate was added and mixed with it, and finally 2 μL of 0.5 mM Fe-EDTA (ethylenediaminetetraacetic acid, iron(III) sodium salt, Sigma-Aldrich) was mixed into the bead. The bead was then tapped into the protein-DNA solution to start the cutting reaction which proceeded for 2 min at 55°C. The OH• cleavage reaction was stopped by adding 150 μL of Stop Solution (13 mM Tris, 4 mM EDTA, 0.3% SDS, 1.3% glycerol, 1.3 mM thiourea, pH 8) followed immediately by 200 μL of buffered phenol:chloroform:isoamyl alcohol (25:24:1) with vigorous vortexing. A+G or C+T ladder reactions were added to the appropriate OH• cleavage treated DNA control tubes which were then vortexed again.

After both DNase I and OH• cleavage reactions, 180 μL of the aqueous phase was removed from the phenol:chloroform extraction and ethanol precipitated with 18 μL 3 M sodium acetate (pH 5.2), 1 μL glycogen (20 mg/mL, Roche), and 500 μL 100% ethanol. Samples were stored at -20°C in precipitation solution until preparation and assembly of all accumulated samples into a 96-well reaction plate (Applied Biosystems) for sample submission. Precipitated

DNA samples were resuspended in 10 μL of HiDi deionized formamide (Applied Biosystems) premixed with GS-500 ROX internal size standard (Applied Biosystems) (0.2 μL per sample) and analyzed on a 3730x1 Applied Biosystems automated DNA sequencer at the Sequencing and Synthesis Facility (University of Georgia). Raw peak data were extracted from the ABI result files (fsa file extension) using the BatchExtract program available from NCBI[1]. Electropherograms from the raw peak data were viewed and analyzed using the graphing and analysis software IGOR Pro (Wavemetrics, Inc.).

Table 2.4 Primers used for footprinting probe amplification from pUC18 clones

Primer	5' label	Sequence[a]
pUC18-M13m-F-FAM	6FAM	ttgtaaaacgacggccagt
pUC18-M13m-R-HEX	HEX	caggaaacagctatgaccatg
pUC18-M13m-F-NED	NED	ttgtaaaacgacggccagt
pUC18-M13m-R-NED	NED	caggaaacagctatgaccatg
pUC18-M13m-F	-	ttgtaaaacgacggccagt
pUC18-M13m-R	-	caggaaacagctatgaccatg

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

2.9 Artificial selection of SurR consensus DNA recognition sequence

A modification of the artificial selection method termed SELEX (Systematic Evolution of Ligands by Exponential Enrichment) [93, 94] was applied to determine the consensus DNA recognition sequence of SurR. This method involves the use of an artificial library of DNA containing random sequences to allow for elucidation of the SurR DNA-binding site through successive cycles of selection with SurR.

The single-stranded SELEX probe from which the artificial library was generated was designed with a 30-nt randomized region flanked by constant primer regions, each containing

[1] Web address: http://www.ncbi.nlm.nih.gov/IEB/Research/GVWG/OSIRIS/index.htm

three restriction sites (XbaI, EcoRI, HindIII on the 5' side and BamHI, EcoRI, SalI on the 3' side). The double-stranded SELEX probe was PCR-amplified from the synthetic single-stranded oligonucleotide with primers that slightly extended the original SELEX probe length (thereby eliminating the 3' self-complementary region at the SalI site which caused unwanted PCR products using primers that exactly matched the SELEX probe priming sites). The SELEX probe and primers are listed in Table 2.5. 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 [92]. Selection rounds were set up essentially as for the EMSA reactions, except for the amount of SELEX probe used and the protein-DNA ratios. For the first round of selection, 0.6 μM of SELEX probe was used, and for all succeeding selection rounds, 0.1 μM was used; protein concentrations ranged from 0.6-1.2 μM. After each selection round, DNA was purified from shifted protein-DNA complexes, amplified with the SELEX primers using 15 cycles of PCR, and polyacrylamide gel-purified before proceeding to the next selection round. A total of 6 selection rounds were performed in this manner. The selected DNA was digested with EcoRI, concatemerized, and cloned into the pUC18 standard cloning vector. Blue/white color screening and colony PCR were used to identify colonies that contained plasmids with the largest concatemers, and plasmid was isolated from these colonies for sequencing. A total of 14 sequences were obtained from the round 6 selected DNA, and a total of 5 sequences were obtained from round 5 selected DNA. These sequences were input into MEME online motif searching software [95] to elucidate a common motif among the selected DNA, and a graphical representation of the motif was generated using WebLogo [96].

Table 2.5 DNA probe and primers used for SELEX

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

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

2.10 Generation of a his-tag cleavable SurR protein construct

A modified version of the pET24dBAM vector adapted to include a TEV protease site between the N-terminal his-tag and the insert site was a kind gift from Francis Jenney (from the laboratory of Michael Adams, University of Georgia). Protein expression from this vector, termed pET24dBAM-TEV, generated a recombinant protein with a cleavable his-tag. The pET24dBAM vector harboring the PF0095 sequence encoding SurR was used to subclone the PF0095 sequence into pET24dBAM-TEV since the same pair of restriction sites could be used to transfer the insert from one vector to the other.

The pET24dBAM-*surr* plasmid was amplified in XL1-Blue cells and purified using a Qiagen Plasmid Miniprep Kit. The parent plasmid and destination vector were digested with BamHI and NotI restriction enzymes for 3 hours at 37°C: Digestion products were then separated on a 1% agarose gel, and bands of insert and linear destination vector were excised and gel-purified using and a QIAEX II Gel Extraction Kit. Gel-purified insert and linearized vector were quantified and ligation reactions were set up as follows: 100 ng linearized vector was combined with varying mole ratios of insert (1:1, 1:2, 1:3), together with buffer and 0.5 Weiss units of T4 DNA Ligase in a total volume of 20 µL. Ligation reactions were incubated at room temperature for 3 hours and immediately transformed (5 µL) by heat-shock into $CaCl_2$-competent XL1-Blue cells. After the 1-hour incubation of the transformation culture, volumes of

50 and 200 µL were spread on agar plates containing 30 µg/mL kanamycin. Plates were incubated at 37°C for 18 hours, and resulting colonies were picked and streaked onto fresh selective plates. These streaks were used to inoculate 5-mL cultures for plasmid production. Plasmid was verified to contain the insert by restriction mapping and was then transformed by heat-shock into $CaCl_2$-competent BL21-CodonPlus(DE3)-RIPL cells (made from a stock obtained from Stratagene) for protein expression.

2.11 Site-directed mutagenesis of SurR CxxC motif

In order to determine whether the cysteines in the SurR N-terminal CxxC motif (Cys23 and Cys26) play a role in redox regulation of SurR DNA-binding activity, a SurR construct was created in which both cysteines were mutated to alanine (C23A and C26A). The QuikChange® Site-Directed Mutagenesis Kit from Stratagene was used. A primer complementary to the region surrounding the cysteine codons was designed in such a way to mutate the two cysteine codons (tgt and tgc) to alanine codons (gct and gcc). The four alanine codons available for use in *E. coli* have a frequency from 17 to 34%, and none are considered rare; therefore, alanine codons which would require the least number of nucleotide changes were selected. The length of the primer extending on either side of the mutated nucleotides was determined by a calculation of the T_m using a formula supplied by the kit. The primer and its complement are listed in Table 2.6.

Table 2.6 Primers used for site-directed mutagenesis of SurR CxxC motif

Name	Sequence[a]
AxxA Primer F	ggatttgttatctcacctaactgctatggaagcctactttagccttttaagtagc
AxxA Primer R	gctacttaaaaggctaaagtaggcttccatagcagttaggtgagataacaaatcc

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

Site-directed mutagenesis on pET24dBAM-TEV-*surr* was performed using the primers listed in Table 2.6 according to the kit protocol (Stratagene). Briefly, the plasmid was PCR-amplified with the mutagenesis primers using a high-fidelity DNA polymerase, the nonmutated parental template plasmid was digested with *Dpn* I (specific for methylated DNA, as is produced by bacterial plasmid amplification), and the resulting amplified plasmid was transformed into XL1-Blue Supercompetent cells. Some of the resulting colonies were selected and streaked onto fresh kanamycin-containing LB-agar plates, and these streaks were used to inoculate 5-mL cultures for plasmid production. Plasmid was purified from the cultures using a Qiagen Plasmid MiniPrep Kit, resulting DNA was quantified, and samples were sent to Sequetech, Inc. (Mountain View, CA) for primer extension analysis to verify the correct insertion of the mutations. The resulting plasmid was pET24dBAM-TEV-*AxxA-surr* for expression of the mutant protein AxxA-Sur.

2.12 Expression and purification of his-tag cleavable SurR for crystallization

The auto-induction protein expression protocol established by Studier [97] was followed for expression of his-tag cleavable SurR and AxxA-SurR for crystallization trials, as well as selenomethionine-labeled SurR and AxxA-SurR. Briefly, one 50-mL culture grown for 8-10 hours at 37°C in P-0.5G media was used to inoculate 1 L of ZYP-5052 media [97] which was then divided equally into two 2-L flasks. The large-scale cultures were grown for 18-20 hours at 37°C before harvesting at 6,000 x g at 4°C for 15 min. For selenomethionine-labeling of expressed protein, minimal media supplemented with selenomethionine was used (PA-0.5G for the 50-mL culture and PASM-5052 for the large-scale protein expression culture [97]).

Cell extract was prepared as described in Section 2.5, and nickel affinity chromatographic separation of the his-tagged protein (using a GE Healthcare HisTrap column) was accomplished using an ÄKTA system (GE Healthcare) with essentially the same method and buffers listed in section 2.5. Fractions contained his_6-SurR were pooled for concentration and buffer exchange into 20 mM HEPES, 200 mM NaCl, pH 7.6, using an Amicon Ultra-15 centrifugal filter device with a 10 kDa molecular weight cut-off (Millipore). The partially-purified his-tagged protein was subjected to his-tag cleavage using AcTEV protease (Invitrogen) according to the manufacturer's instructions, except that less protease was used in conjunction with a longer incubation time. Approximately 30 mg of his-tagged protein (obtained from 1-L of culture) was digested with 500 units of AcTEV protease (Invitrogen) in a volume of 2-3 mL using reaction buffer and DTT supplied with the protease according to manufacturer instructions. The cleavage reaction was incubated at 30°C with shaking at ~65 rpm for 10-14 hours.

Following his-tag digestion, the protein sample was applied directly to a fresh nickel-affinity HisTrap column (GE Healthcare) equilibrated with 20 mM HEPES, 200 mM NaCl, pH 7.6, using a 0.2 mL/min flow rate. The column flow-through containing the tagless protein was collected, while the column-bound uncleaved protein, his-tagged TEV protease, and *E. coli* proteins were eluted with Elution Buffer (see Section 2.5). SDS-PAGE was used throughout the purification process to monitor purification and his-tag cleavage and to check for protein stability and purity.

For protein purified for use in crystallization trials, a gel-filtration polishing step was performed mainly to remove protein aggregates. A HiPrep 26/60 Sephacryl S-100 High Resolution (GE Healthcare) gel filtration column was equilibrated with two column volumes of 20 mM HEPES, 200 mM NaCl, pH 7.6, and protein sample was injected onto the column at a

flow rate of 0.7 mL/min. Protein-containing fractions were pooled and concentrated to 8-12 mg/mL using an Amicon Ultra-15 centrifugal filter device with a 10 kDa molecular weight cut-off (Millipore). Protein concentration was determined using a Bio-Rad DC Protein Assay kit. Protein samples were given to Hua Yang (Bi-Cheng Wang's Laboratory, University of Georgia) for crystallization trials.

2.13 Ellman's assay for quantification of SurR free thiols

In order to determine what percentage of the cysteines in aerobically purified recombinant SurR were in the oxidized and reduced state, Ellman's assay was employed. Ellman's reagent, also known as DTNB (5,5′-Dithiobis(2-nitrobenzoic acid)), reacts with free thiols resulting in the release of a yellow-colored thiolate in a one-to-one correspondence; this color change is quantifiable when absorbance is measured at 412 nm. The assay was also used to test whether the thiol-specific oxidant diamide [98, 99] was effective at oxidation of the free thiols remaining in the protein. Diamide is compatible with the Ellman's assay because it contains no free thiols which could react with the DTNB reagent.

Two standards, DTT (dithiothreitol) and L-cysteine, were used to create a calibration curve of absorbance versus thiol concentration. Fresh 0.1 M stock solutions of each standard were made using degassed 0.1 M sodium phosphate buffer (pH 8.0), and dilutions were made from each stock using the same buffer. Protein samples were prepared from a frozen stock of SurR by making 10, 25 and 50 mM dilutions with degassed 20 mM HEPES, 200 mM NaCl (pH 7.6). A second set of samples with the same concentrations was prepared in the presence of 10 mM diamide. All protein samples were incubated at room temperature for 15 min prior to preparation of the 96-well microplate.

The protocol for the Ellman's assay[1] was modified such that the volumes of standards, samples and reagents were suitable for use in a 96-well microplate: 25 µL of each standard and sample was aliquoted in triplicate into wells of the microplate, then 5 µL of 4 mg/mL DTNB solution was added using a repeator pipette, followed by 250 µL of 0.1 M sodium phosphate buffer (pH 8.0) using a multi-channel pipette. After 5 to 10 min, the absorbance of each well was read at 412 nm using a microplate reader.

A calibration curve for each standard was created by plotting absorbance at 412 nm versus thiol concentration (1 thiol per molecule of cysteine and 2 thiols per molecule of DTT). Absorbance of each protein sample was converted to a concentration of thiols using the equation of the line derived from each calibration curve. The free thiol concentration for each sample was divided by the corresponding protein concentration to obtain the number of free thiols per protein in each sample.

2.14 Oxidation and reduction EMSA experiments with SurR

The thiol-specific oxidant diamide was shown to completely oxidize SurR, and so it was of interest to determine the effect of SurR oxidation on its DNA-binding affinity and specificity. The anaerobic portions of the experiment were performed with the help of Gerrit Schut in Michael Adams' laboratory (University of Georgia). An EMSA experiment similar to that described in Section 2.6 was set up such that incubations were carried out in the presence of 10 mM diamide. After incubation of the EMSA reactions at 70°C for 20 min, 8-µL aliquots of the 'oxidized' protein-DNA mix were made into anaerobic half-dram vials containing 2 µL of a reductant such that the final concentration of the reductant would be 20 mM: DTT, cysteine, sodium dithioninte, and sodium sulfide. A second incubation was carried out for 5 min at 70°C

[1] The protocol was modified from Support protocol 2 in Current Protocols in Protein Science, section 15.1.13.

after which the anaerobic vials were opened so that the samples could be loaded and run on a gel (aerobically), as for EMSA (Section 2.6).

The effect of the oxidant colloidal sulfur, a compound more likely to be an *in vivo* effector, was also tested and compared with that of diamide. The colloidal sulfur solution was made by Gerrit Schut by diluting a stock polysulfide solution into EMSA buffer after which the yellow color characteristic of polysulfide faded and a suspension of colloidal sulfur became visible. The colloidal sulfur solution was used in EMSA as soon as possible, as it was not stable for more than a few hours. Oxidation/reduction experiments with colloidal sulfur were performed as for those with diamide.

2.15 Analytical gel filtration to determine SurR quaternary structure

Analytical gel filtration using a Superdex 75 10/300 GL size exclusion column (GE Healthcare) was performed on various SurR samples to determine their quaternary structure. Four 200-uL samples of 2.5 mg/mL protein were prepared: untreated SurR, untreated AxxA-SurR, SurR treated with 10 mM diamide for 10 min at room temperature, and SurR treated with ~2 mM colloidal sulfur for 10 min at room temperature. The running buffer used was 20 mM HEPES, 200 mM NaCl, pH 7.6. Molecular weight standards (200 μL of each) were run through the column individually and the elution volume (V_e) of each was noted. The following standards (Sigma) were used: blue dextran (2 mg/mL), bovine serum albumin (10 mg/mL), carbonic anhydrase (3 mg/mL), cytochrome C (2 mg/mL), and vitamin B_{12} (0.1 mg/mL). Blue dextran was used to determine the column void volume (V_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.

CHAPTER 3

TRANSCRIPTION FACTOR DISCOVERY APPROACH AND INITIAL FINDINGS

3.1 Targeted transcription factor discovery by DNA affinity protein capture

3.1.1 Selection of target ORFs for transcription factor discovery

Whole-genome microarray expression profiling experiments testing the response of *P. furiosus* under growth conditions in the presence and absence of elemental sulfur (S^0) reveals changes in the expression levels of many genes [78, 100]. The premise that gene up and down-regulation seen in the microarray results occurs mainly at the transcriptional level allows for the use of the DNA affinity protein capture approach to identify transcription factors that associate with particular promoters under specific growth conditions. The first step in this approach is to select target ORFs that show significant regulation in microarray expression profiles comparing two growth conditions.

Table 3.1 lists the microarray expression profile obtained for a subset of *P. furiosus* genes comparing batch cultures grown either in the presence or absence of S^0 [78]. A total of 12 ORFs displayed greater than 5-fold up-regulation in the $+S^0$ condition. The two ORFs showing the highest expression were PF2025 and PF2026 which, as it turns out, are divergently transcribed and share an intergenic space of only 164 bp. This ORF region was selected as a target for transcription factor discovery because of the dramatic up-regulation, but also because these ORFs were originally hypothesized to play a significant role in S^0 metabolism [78], although

recent kinetic DNA microarray experiments have shown that they are involved in a secondary response to S^0 [100].

There were 21 ORFs that were down-regulated more than 5-fold, and surprisingly, 18 of these were members of three hydrogenase operons, encoding the membrane bound hydrogenase and two cytosolic hydrogenases (I and II), all of which have been characterized [101-104]. Moreover, the most down-regulated ORF was PF0559 which encodes a protein most likely involved in hydrogenase maturation. The second most down-regulated ORF was the first ORF in the 14-ORF membrane-bound hydrogenase operon, *mbh1* (PF1423) [101]. It has been confirmed that this down-regulation extends beyond the transcriptional level since the activity of all three of the hydrogenases is also considerably reduced in the presence of S^0 [77]. This indicates that none of these hydrogenases are involved in hydrogen sulfide (H_2S) production during S^0-related metabolism. Because of this interesting metabolic change and the dramatic down-regulation of the membrane-bound hydrogenase operon, the *mbh1* ORF was chosen as a second target for discovery of transcription factors relevant to S^0-related metabolism. Proteins that bound to these DNA targets were identified from cell extracts of cultures grown with and without S^0 using the DNA affinity protein capture method described below.

Table 3.1 DNA microarray expression profiles of *P. furiosus* for growth +/–S^0 [78]

ORF	ORF description / operon function[a]	Fold change from – S^0 to +S^0
Up-regulated >5-fold with S^0 [78]		
PF2025	Sulfur-induced protein A, *sipA* [78]	61.4
PF2026	Sulfur-induced protein B, *sipB* [78]	25.1
PF1592	[Tryptophan synthase, subunit beta homolog]	7.9
PF1954	[Conserved hypothetical protein]	7.6
PF1053	[Aspartokinase II alpha subunit]	7.6
PF1186	NAD(P)H sulfur reductase (NSR) [100]	7.4
PF1974	[Thermosome, single subunit]	7.1
PF0935	[Acetolactate synthase]	6.9
PF0059	[Fibrillarin-like pre-rRNA processing protein]	6.5

PF0191	[Oligopeptide transport system permease protein]	6.4
PF0094	Protein disulfide oxidoreductase, *pdo* [105]	6.1
PF1443	[NADH dehydrogenase subunit]	6
Down-regulated >5-fold with with S^0 [78]		
PF0044	[Conserved hypothetical protein]	-6
PF0559	[Hydrogenase expression/formation regulatory protein, *hypF*]	-39
PF0594	Ornithine carbamoyltransferase, *argF* [106]	-5.4
PF0891[b]	Hydrogenase I beta, *hydB1* (4-ORF operon) [102]	-6.3
PF1329[b]	Hydrogenase II beta, *hydB2* (4-ORF operon) [103]	-6.3
PF1423[b]	Membrane bound hydrogenase, *mbh1* (14-ORF operon) [101]	-17.6

[a] ORF descriptions and operon functions are essentially derived from [78]. Brackets indicate predicted function, while those genes that have been experimentally studied are listed without brackets followed by the corresponding reference.
[b] For ORFs which are at the beginning of operons, only the first ORF in the operon is listed with its corresponding fold change.

3.1.2 Optimization of the DNA affinity protein capture method

The method for transcription factor discovery presented in this work, designated as DNA affinity protein capture, employs magnetic streptavidin-coated beads to immobilize a biotinylated DNA fragment containing the upstream DNA of a given gene containing its putative promoter region. The bead-bound DNA is then used as 'bait' to capture and isolate DNA-binding proteins from soluble cell extract; all proteins which do not bind to the DNA can simply be removed after the beads are separated from the solution by a magnet (Fig. 3.1). This procedure was described with regard to capture of *P. furiosus* DNA-binding proteins by Yu Chen [107], and further development of the *P. furiosus*-specific protocol was done by Meiyao Wang [108].

The DNA affinity protein capture experiment basically consists of four steps: 1) the biotinylated probe DNA is bound to the magnetic beads, 2) bead-bound DNA is incubated in cell extract solution, 3) beads and bound proteins are magnetically separated from solution, and 4) DNA-bound proteins are eluted from the beads and visualized with SDS-PAGE (Fig. 3.2). There are multiple variables which influence the results obtained from this procedure, and therefore, some parameter optimization was undertaken to prepare for the work presented here.

Specifically, the following parameters were varied and optimized: buffer composition (buffer type, salt concentration, DTT concentration, and presence of BSA), cell extract concentration, incubation temperature, and protein elution conditions.

Parameters chosen for the protein-DNA incubation step were the most critical since these defined the conditions which would dictate protein-DNA binding *in vitro*. The removal of proteins from their intracellular environment will inevitably change their behavior in an *in vitro* system; however, in general, incubation conditions were designed to best mimic *in vivo* conditions, although these can only be approximated at best. Since the cell extract incubation occurred at high temperatures suitable for thermophilic proteins, the buffer EPPS was chosen to minimize the pH change with increasing temperatures. EPPS buffer has a negligible change in pH with increasing temperatures compared with Tris (the buffer which had been used in previous protocols). Salt concentration was also a concern since *P. furiosus* is a marine archaeon and presumably has high intracellular potassium ion concentrations [109]. Accordingly, the concentration of potassium chloride (KCl) in the binding buffer was varied from 100 to 300 mM; however, 100 mM KCl was chosen since higher concentrations inhibited the binding of many proteins. Not surprisingly, increased incubation temperatures (from 55 to 75°C) also dramatically reduced the amount of proteins eluted from the DNA, and therefore the 55°C incubation temperature was typically used. It should be noted, however, that *P. furiosus* cell-free transcription is performed optimally at 70°C and has been shown to be functional even at 95°C [110, 111].

Heparin was also used as a competitor for non-specific DNA-binding proteins since it resembles DNA with its elongated structure and highly negatively charged surface. It is therefore a good competitor for non-specific proteins interacting electrostatically with the DNA backbone.

The proteins which were eluted from the bead-DNA after a wash with heparin were considered to have lower affinity to the DNA; however, heparin was used with caution since at sufficiently high concentrations, it can challenge sequence-specific DNA-binding proteins from their cognate DNA [112]. This was especially a concern since transcription factors can be present in low concentrations within the cell.

The elution step was another challenge, since previous feasibility studies [107] found that the thermophilic proteins would not elute from the beads except under harsh conditions, in particular, incubation of the beads in Laemmli gel-loading buffer at 99°C for a defined period of time. However, this harsh elution step tended to strip everything from the bead surface, including the covalently-bound streptavidin subunits which anchored the DNA to the beads. Much of this background noise was eliminated without reduction in recovered protein by simply eluting the protein from the bead-DNA complex using Laemmli buffer without β-mercaptoethanol and a 5-min incubation at 55°C. A different elution strategy which aided in identification of proteins which were retained on the beads but did not appear to bind to the DNA was the utilization of DNase to elute DNA-bound proteins from the bead-DNA complex. Although parameter-space is nearly infinite for such an experiment, the range and variety of variables tested aided in analysis of identified eluted proteins.

3.1.3 Design of the DNA probes used in protein capture

Since archaeal transcriptional regulation follows a bacterial pattern, most transcription factor binding sites should lie in close proximity to the ORF translation start site; therefore, a length of 200 bp was selected for inclusion of upstream, putative promoter DNA, and approximately 100 bp of DNA downstream from the ORF was included to ensure complete coverage of possible transcription factor binding sites. With this in mind, the *mbh1* probe for the

PF1423 target region was designed to contain approximately 200 bp upstream and 100 bp downstream from the ORF start site with the biotin label anchoring the probe to the bead on the side of the DNA farthest from the ORF. Incidentally, PF1423 shares a 149-bp intergenic region with a divergently-transcribed ORF, PF1422; therefore part of this ORF and its putative promoter region were also included in the *mbh1* probe. The *sipB* probe for the PF2025-PF2026 target region was designed such that approximately 100 bp of DNA downstream from each divergent translation start site was included, along with the 164-bp intergenic region; the biotin label was placed on the PF2026 (*sipB*) side of the probe (Fig. 3.3 A).

3.1.4 Protein capture on sipB and mbh1 probe DNA

The DNA affinity protein capture experiment presented here relies on differences between two growth conditions, and thereby the corresponding cell extracts, to elucidate the protein factor which is presumed to cause or contribute to the differential expression profiles observed in the DNA microarray data utilized to select the target DNA used for protein capture. One of the assumptions of this experiment is that the factor responsible for differential expression observed in DNA microarray data is either present in different amounts between the two growth conditions being compared or is only able to bind DNA in one condition as compared to the other due to the presence or absence of some effector. Therefore, in either case, there should be an observable difference, however minor, in the DNA affinity captured proteins from two different cell extracts using the same probe DNA.

The *sipB* and *mbh1* probe DNA were bound to magnetic beads and incubated in cell extracts from cultures grown in the presence and absence of S^0 at 55°C for 30 min. Eluted proteins were separated using SDS-PAGE and visualized by silver staining. The resulting gel image can be seen in Figure 3.3. Many of the protein bands were the same between the two cell

extracts and even between both *sipB* and *mbh1* probes. Most of the bands which appeared differentially were hardly discernible with silver staining, and therefore the proteins were not present in high enough concentrations to be identified by mass spectrometry. Those that could be identified fell into four general categories, based on their identification and on previous experiments: basal transcription proteins (bands 2, 3, 11), non-specific DNA-binding proteins that could be challenged off with heparin (bands 1, 6, 7), bead-binding proteins not removed by DNase digestion (bands 4, 5), and several conserved hypothetical proteins (bands 8-10,12). See Table 3.2 for protein band identification results.

Table 3.2 Identification of protein binds using peptide mass fingerprinting

Band number[a]	Apparent gel molecular weight (kDa)	Annotation[b]	Locus	Molecular weight (kDa)	High score, error[c]	Masses matched	Sequence coverage
1	145	reverse gyrase (rgy)	PF0495	139.9	87 (59), 200 ppm	46/73	36%
2	130	DNA-directed RNA polymerase subunit b	PF1564	126.9	173 (74), 200 ppm	59/100	45%
3	105	DNA-directed RNA polymerase subunit a'	PF1563	103	39 (59), 200 ppm	18/51	27%
4	95	cell division control protein 48, AAA family	PF0963	94.1	328 (59), 200 ppm	57/98	57%
5	55	methylmalonyl-CoA decarboxylase, subunit alpha	PF0671	57.1	84 (59), 200 ppm	40/95	50%
6	38	methionine synthase vitamin B12-independent isozyme	PF1269	39.3	184 (60), 100 ppm	21/38	52%
7	35	conserved hypothetical protein	PF1268	35.5	72 (60), 100 ppm	8/24	28%
8	30	conserved hypothetical protein	PF1827	31.3	21 (59)	5/30	21%
9	29	conserved hypothetical protein	PF0496	30.6	76 (59), 100 ppm[d]	18/57	43%
10	27	conserved hypothetical protein	PF0095	26.9	88 (59), 100 ppm	17/42	46%

11	20	transcription initiation factor TFIID chain a (TBP)	PF1295	21.3	50 (59), 100 ppm[e]	6/42	34%
12[f]	18	conserved hypothetical protein	PF1572[g]	20	44 (59), 200 ppm[d]	9/34	54%
12[f]	18	conserved hypothetical protein	PF0284	22.1	48 (59), 200 ppm	11/51	62%

[a] Numbers correspond to numbered bands in Figure 3.3.
[b] Protein annotations are derived from the NCBI database.
[c] Score listed is the high score for the MASCOT search against a database including all archaea. Cut-off for significant scores is listed in parenthesis, and the error used for each search is listed after each score, separated by a comma.
[d] Search was performed without masses from blank.
[e] Search was performed without any variable modifications.
[f] Two proteins were identified in sample 12, and the matched masses were unique for each hit.
[g] This hit was the listed fourth on the MASCOT search against all archaea.

Previous experiments using DNase to elute proteins from the beads were utilized for comparison to determine which protein bands were binding to the bead surface and not to DNA since Laemmli buffer was used to elute the proteins from the bead-DNA complex in this experiment. Two of the identified proteins were found to associate with the bead itself: these were annotated as ‘cell division control protein 48, AAA+-type ATPase family’ (PF0963, band 4) and ‘methylmalonyl-CoA decarboxylase, subunit alpha’ (PF0671, band 5). It is possible that ‘methylmalonyl-CoA decarboxylase’ associates with the biotin used to attach the DNA to the bead surface since this enzyme requires a biotin cofactor; however, it is not clear why ‘cell division control protein 48’ associates with the bead. As would be expected for bead-binding proteins, both of these bands were evident in both cell extracts for both DNA probes.

Protein components of the basal transcriptional machinery were also identified, and this finding helped to validate the parameters chosen for the experiment. Two of the largest RNAP subunits, b and a' (PF1564 and PF1563, bands 2 and 3, respectively), were evident in both cell extracts for both DNA probes. Interestingly, TBP (PF1295, band 11) was only identified in the –

S^0 lane from the *sipB* probe. A faint band at the same position for the $-S^0$ lane from the *mbh1* probe may also be TBP; however, it was not positively identified as such.

Other experiments using heparin as a competitor for non-specific DNA-binding proteins indicated that proteins in bands 1, 6, and 7 were probably associating with the DNA probes in a non-specific manner. Band 1 was identified as reverse gyrase, a protein which has been studied in *P. furiosus*, and non-specific affinity for DNA is a necessary part of this protein's function [113]. Bands 6 and 7 were two of the few examples of notable differences in protein abundances between the cell extracts. These were identified as 'methionine synthase vitamin B12-independent isozyme' (PF1269) and 'conserved hypothetical protein' (PF1268), respectively, and were present in the $+S^0$ lane for both *sipB* and *mbh1* probes. It is interesting that their ORFs may share the same operon and perhaps the proteins form a complex; however, their apparent weak DNA-binding affinity disqualifies them as potential transcription factors. Also, they likely function as methionine-synthase-related proteins according to sequence analysis, and perhaps the need to utilize excess sulfur in the synthesis of methionine causes these proteins to be present in high abundance in cells grown in the presence of S^0.

The remaining identified proteins fell into the category of 'conserved hypothetical proteins' (bands 8-10, 12), and sequence analysis was used to determine their probable functions. The protein sequences were subjected to BLAST searches of the NCBI non-redundant database [81] and Conserved Domain searches of the Conserved Domain Database of NCBI [84, 85]. The two proteins identified in band 12, PF0284 and PF1572, were both predicted to be involved in regulation of amino acid metabolism through a conserved ACT domain [114] (COG2150 from the Cluster of Orthologous Genes database [115]) and are in fact homologous to one another, having 30% sequence identity. Although these proteins may serve as metabolic regulators, there

is no indication that they are DNA-binding proteins. It is possible that their association with the bead-DNA complex occured as a result of protein-protein contacts with DNA-bound proteins.

Sequence analysis of the remaining three conserved hypothetical proteins (bands 8-10) indicated that they were predicted transcriptional regulators: protein products of PF1827, PF0496, and PF0095. PF1827 and PF0496 proteins were apparent in both cell extracts for both DNA probes. PF1827 has homology with families containing C-terminal cystathionine beta synthase (CBS) domains, predicted to play a role in signal transduction, and also with a family of transcriptional regulators (COG2524 [115]) containing a C-terminal CBS domain. Computational analysis of archaeal genomes has revealed an abundance of CBS domains which can also be found in conjunction with HTH domains, suggestive of a role in transcription regulation [116]. PF0496 matched highly with a conserved domain of which TrmB from *T. litoralis* and *P. furiosus* is the characterized representative [48]. Curiously, the PF0496 ORF is potentially in an operon with reverse gyrase (PF0495), which was also identified in the DNA affinity protein capture experiments. PF0496 has recently been investigated by Lee and coworkers, who have named it TrmBL2 since it also recognizes the TrmBL1 binding site, although with lower affinity, and may be involved in regulation of sugar metabolism [52].

Of these putative transcriptional regulators, only the PF0095 protein was represented differentially between the two cell extracts, and more importantly, this protein band was only present on the *mbh1* probe, suggesting that it binds specifically to the upstream DNA of this ORF. Analysis of the PF0095 protein sequence revealed homology with two conserved protein domains that are transcription-factor related (Fig. 3.4 A). The closest match was to a family of conserved proteins, COG1777 [115], annotated as ‘predicted transcriptional regulators’ and comprised of ten proteins from eight Euryarchaeal species. The N-terminal region of PF0095

was aligned with an HTH/ArsR family of bacterial and archaeal regulators of which the well-characterized arsenical resistance operon repressor (ArsR) is a member [117]. Secondary structure prediction revealed the presence of four helices in the N-terminal region (residues 1-60) which contained homology to the HTH/ArsR domain, a further indication that PF0095 contained an HTH DNA-binding domain (Fig. 3.4 B). A BLAST search of the sequence resulted in approximately 30 hits having an e-value less than 1, with the majority of high-scoring hits falling within the archaea as either predicted transcriptional regulators or conserved hypothetical proteins. The three closest hits (having e-values of less than or equal to $2x10^{-74}$) were from the three other members of the Thermococcaceae family and are presumably PF0095 orthologs: PH0180 (72% sequence identity), PAB0108 (72% sequence identity), and TK1086 (65% sequence identity). A tree view of the highest scoring BLAST hits can be seen in Figure 3.4 C.

From the differential representation in the DNA affinity protein capture assay and sequence analysis indicating the presence of a DNA-binding domain, the protein product of PF0095, identified from the *mbh1* probe, emerged as the best potential transcription factor candidate and was chosen for expression and characterization. The protein product of ORF PF0095 will hereafter be referred to as SurR[1], a name given this protein as a result of its characterization which is described in this work.

3.2 Validation of SurR as a DNA-binding protein with sequence specificity

3.2.1. Expression and purification of recombinant his-tagged SurR

To verify that SurR was a sequence specific DNA-binding protein, the protein was over-expressed as a his-tagged recombinant protein in *E. coli* for *in vitro* studies. The recombinant protein was produced using a protein-expression vector which allowed for the expression of

[1] SurR stands for Sulfur-response Regulator

SurR with an N-terminal his-tag which facilitates protein purification. The SurR protein sequence is 232 amino acids in length and has a calculated molecular weight of 26,886 Da; however, with the addition of a his-tag which produces the construct Met-Ala-His_6-Gly-Ser-protein, the new molecular weight is 27,924 Da. The protein was expressed in an *E. coli* strain compatible with the host expression vector, yielding an inducible system for protein production. The first protein expression strain selected did not compensate for rare codon usage, so truncated N-terminal his-tagged protein products were observed during the purification steps since the SurR sequence contained some rare codons as compared to codon frequency in *E. coli* proteins. The BL21-CodonPlus(DE3)-RIPL strain which compensates for rare codons by expressing supplementary corresponding tRNAs (for Arg, Ile, Pro, and Leu rare codons) was then utilized, and this change effectively prevented expression of the truncated protein products. Expression of the protein did not appear to be toxic to the cells and did not significantly slow culture growth, as evidenced by the growth curve for protein expression in Figure 3.5. Moreover, the protein was expressed in abundance, predominantly in soluble form, which greatly simplified the purification procedure.

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. Only the fractions which were of acceptable purity were collected and pooled for buffer exchange using a desalting column. An example gel showing protein purified in this manner can be seen in Figure 3.5. Once the protein was exchanged into a suitable salt-containing buffer, small aliquots were stored at -80°C to be thawed individually for use in *in vitro* assays.

3.2.2. His-tagged SurR binds to the mbh1 promoter region in vitro

In order to show that SurR binds specifically to the *mbh1* probe DNA from which it was identified in the DNA affinity protein capture experiment, an EMSA was performed using this same probe DNA to test for binding. SurR was found to shift the *mbh1* probe DNA completely at a protein/DNA mole ratio of around 20. Also, more than one distinct protein-DNA complex was evident in the range of protein-DNA ratios tested, with two bands of protein-DNA complexes predominating in particular (Fig. 3.6).

For comparison, the *sipB* probe was also tested in EMSA with SurR, as well as another probe, *lrpA* (a kind gift from Meiyao Wang), which contained approximately 100 bp of *lrpA* ORF DNA and 200 bp of upstream DNA containing the characterized *lrpA* promoter [37]. SurR also shifted both of these probes, but to a lesser extent (Fig. 3.6). Also, the well-defined protein-DNA complexes which were so prominent for the *mbh1* probe were not as evident for either of these probes. For the *sipB* and *lrpA* probes, only approximately half of the probe DNA was shifted by SurR at protein/DNA mole ratios of ~20. These data indicated that SurR has higher affinity for the *mbh1* probe DNA; however, it was still of interest to further demonstrate specificity of SurR binding, and in particular to compare the difference in binding of SurR to ORF upstream DNA[1] versus ORF.

To further confirm the specificity of SurR binding to the *mbh1* upstream/promoter DNA using EMSA, an 81-bp DNA probe was designed that included 76 bases of DNA upstream from the *mbh1* translation start and extended 5 bases into the *mbh1* ORF. Two 80-bp DNA probes were selected as controls; the first contained only *mbh1* ORF DNA, and the second was a synthetic DNA consisting of dAdG repeating units. SurR bound to the *mbh1* promoter fragment

[1] For ORFs mentioned in this work that do not have characterized promoters, the ORF upstream DNA is assumed to include the promoter region, and the terms ‘upstream DNA’ and ‘promoter’ are used somewhat interchangeably.

with higher affinity than the full-length *mbh1* probe, completely shifting the DNA at a protein/DNA mole ratio between 10 and 15 (Fig. 3.7, left). Once again, the distinct protein-DNA complexes were observed at various protein concentrations tested, displaying a gradual shift of protein-DNA complexes from higher to lower electrophoretic mobility with increasing protein concentration. In contrast, SurR did not completely shift ORF DNA even at a protein/DNA mole ratio of 25, although some shifting of the DNA was apparent. A minimal amount of DNA began shifting at a ratio of 10; however, the band was less well-defined and the appearance of the shift pattern was distinctly different than that observed for the promoter fragment. The synthetic dAdG probe, by comparison, did not bind SurR at all; perhaps the repeating nature of the bases caused the formation of a DNA structure that was not conducive even to non-specific DNA binding. These results suggested that the association of SurR with the *mbh1* promoter was indeed of a specific nature.

3.2.3. Expression and purification of recombinant SurR with a cleavable his-tag

The putative HTH DNA-binding domain of SurR is located at the N-terminus where the his-tag used for protein purification was also placed, and in order to test if the his-tag had any effect on DNA binding, a SurR construct with a cleavable his-tag was required. A modified version of the host vector from which his_6-SurR was expressed, containing an inserted TEV protease site between the his-tag and the cloning site (pET24dBAM-TEV), was obtained from the laboratory of Michael Adams (University of Georgia). Since the his_6-SurR vector and the pET24dBAM-TEV vector originated from the same parent plasmid, it was relatively straightforward to subclone the SurR sequence from one vector to the other.

This construct allowed for better purification of the protein with the end result being a protein having a nearly native N-terminus. The protein was expressed with a his-tag and purified

from cell extract using a nickel-affinity column (Fig. 3.8). A his-tagged TEV protease was then used to cleave the tag after which the tagless protein was purified away from uncleaved protein and TEV protease over a second nickel-affinity column (Fig. 3.9). This process of purification was effective at removing most contaminants from cell extract, and having a tagless protein is more desirable, since any possible interference of the his-tag in *in vitro* assays with the protein is eliminated.

3.2.4. Confirmation of specific binding of SurR to the mbh1 promoter

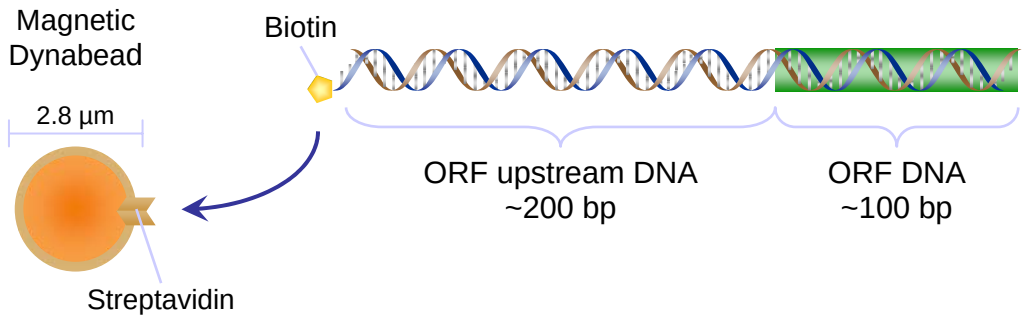
To test the effect of his-tag removal on the affinity of SurR for DNA, the same experiment comparing *mbh1* upstream/promoter versus ORF DNA was carried out. The non-specific competitor heparin used in the DNA affinity protein capture experiments was used as a competitor in EMSA for non-specific DNA binding. Because the EMSA experiments were not performed in the traditional way using radioactively-labeled DNA probes, and instead relied on a DNA-staining agent to visualize the DNA, it was impossible to use standard competitor DNA in the assays without the possibility of interfering with the visualization of the probe DNA and shifts in the resulting EMSA gel. Heparin is a good substitute because it is not bound by DNA-staining agents, and yet its structure and electrostatic properties mimic DNA enough so that it has been used as a non-specific competitor in EMSA [112]. Base-specific contacts of a protein with DNA have much higher affinity than pure electrostatic association with the phosphate-sugar backbone; therefore, a protein binding non-specifically to DNA solely through electrostatic associations should be able to be challenged off with heparin since it too has a highly negative overall charge and resembles DNA in this regard.

As can be seen in Figure 3.10, the cleaved-tag SurR protein (denoted as ‘SurR’) is able to shift *mbh1* promoter and ORF DNA to the same extent as his_6-SurR (compare with Fig. 3.7), so

there appears to be no interference of the his-tag with the DNA-binding activity of SurR. Furthermore, the use of heparin as a non-specific DNA competitor revealed that the association of SurR with promoter DNA was indeed specific while the binding to ORF DNA was of a non-specific nature. At heparin concentrations between 10 and 100 μg/mL, SurR binding to the *mbh1* ORF was completely blocked. Given this finding, it is not surprising that at high protein/DNA mole ratios, SurR associates non-specifically even with promoter DNA, as evidenced by the shift to very low mobility complexes at protein/DNA mole ratios of 24 and 32 and the corresponding shift back to the position of a specific complex with the addition of heparin at 10 to 100 μg/mL (this corresponds to approximately 3 to 30-fold excess over the DNA probe). It is interesting that the position of these low mobility bands in the absence of heparin matches closely with the position of the shifted bands for the ORF DNA at protein/DNA mole ratios of 16, 24 and 32. The implication is that many proteins are associating on the DNA in a non-specific manner, thereby creating a protein-DNA complex with very low mobility and less band definition. Once the *mbh1* promoter probe is completely shifted by the protein at a protein/DNA mole ratio of 16, non-specific binding begins to occur with higher protein/DNA mole ratios, forcing the resulting protein-DNA complexes to have lower and lower electrophoretic mobility. It is also worth noting that heparin at sufficiently high concentrations can also challenge off sequence specific DNA binding as evidenced by its effect on SurR binding of *mbh1* promoter at a concentration of 1 mg/mL (over 270-fold higher concentration than the DNA probe), and this observation has been noted elsewhere [112]. These data conclusively show that SurR is a sequence specific DNA-binding protein, and furthermore, that one of its targets appears to be DNA upstream of the *mbh1* ORF (and thereby presumably, the entire membrane-bound hydrogenase operon).

Figure 3.1 Design and immobilization of probe DNA. A. The probe DNA is designed to have approximately 200 bp upstream and 100 bp downstream from the target ORF start, resulting in a probe DNA of length ~300 bp. 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). **B.** The magnetic properties of the beads allow them, and correspondingly whatever is attached to them, to be easily separated from solution with the use of a magnet.

A


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

B

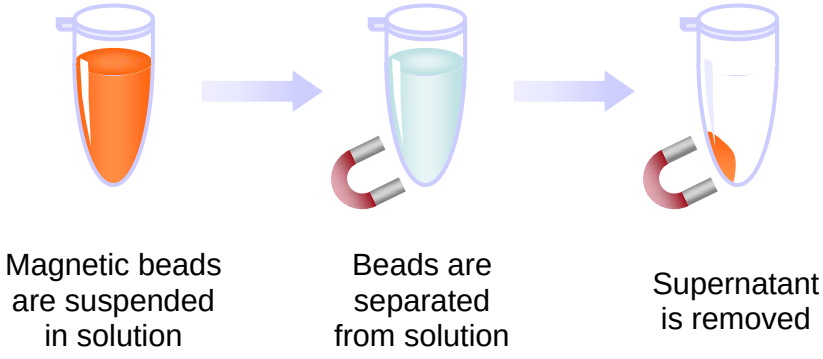

Magnetic beads
are suspended
in solution
Beads are
separated
from solution
Supernatant
is removed

Figure 3.2 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: general transcriptional machinery (TBP, purple; TFB, teal; RNAP, blue), non-specific DNA-binding proteins (pink), and other transcription factors (green). Proteins which do not bind DNA (grey) are removed, and finally the DNA-binding proteins which remain are eluted and analyzed.

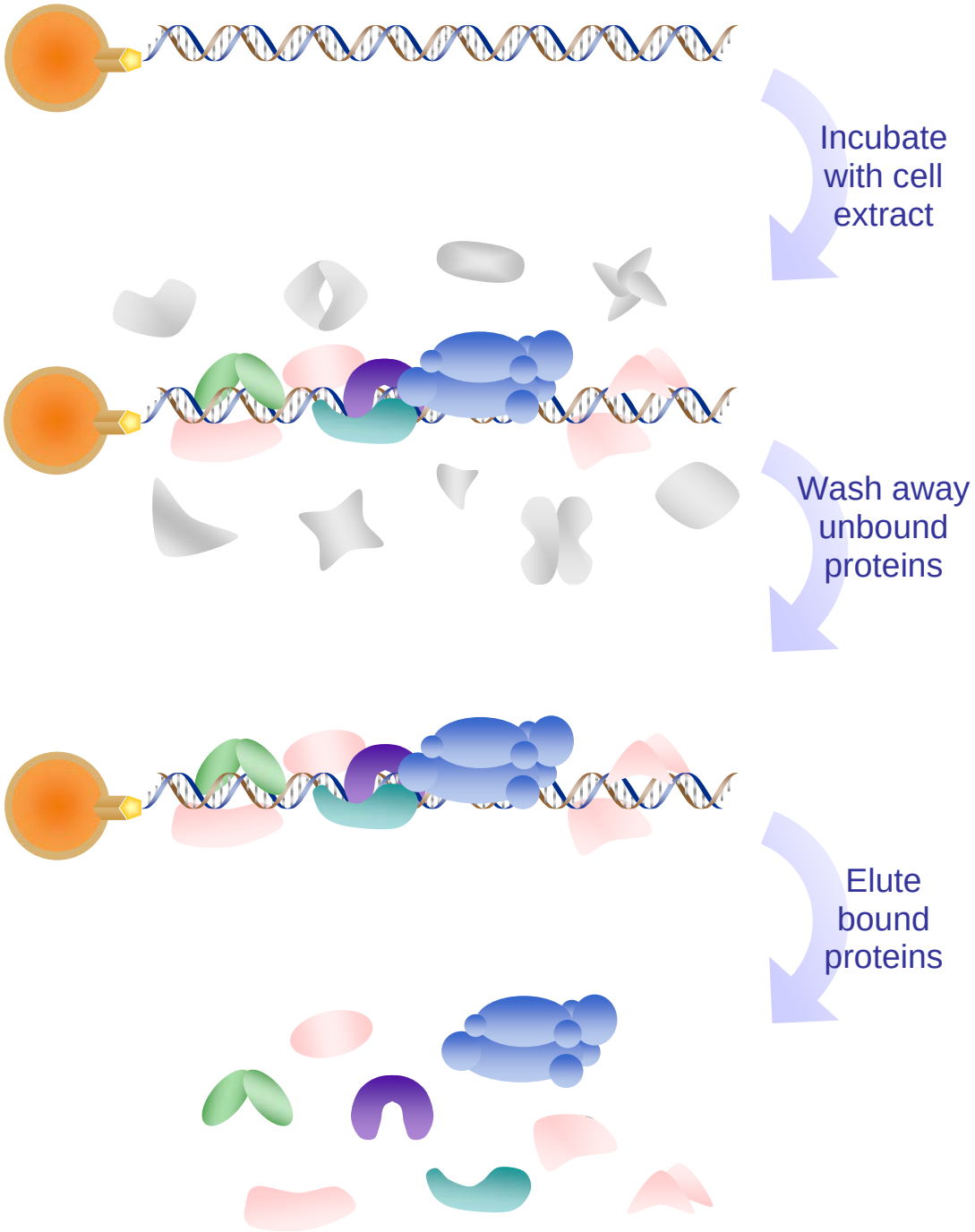

Incubate
with cell
extract
Wash away
unbound
proteins
Elute
bound
proteins

Figure 3.3 DNA affinity protein capture with *sipB* and *mbh1* probes. A. Diagrams of the probes used in DNA affinity protein capture. The position of the biotin group is represented by a yellow pentagon. **B.** Silver-stained denaturing gel of eluted proteins from DNA affinity capture with *sipB* and *mbh1* probes incubated in soluble cell extracts from cells grown in the presence (red) and absence (black) of S^0 with the corresponding band intensities (traces to the left are for *sipB* lanes and traces to the right are for *mbh1* lanes). Arrows indicate identified proteins: 1, reverse gyrase, PF0495; 2, DNA-directed RNA polymerase subunit b, PF1564; 3, DNA-directed RNA polymerase subunit a', PF1563; 4, Cell division control protein 48, AAA family, PF0963; 5, methylmalonyl-CoAdecarboxylase, subunit alpha, PF0671; 6, methionine synthase vitamin B12-independent isozyme, PF1269; 7, conserved hypothetical protein, PF1268; 8, conserved hypothetical protein, PF1827; 9, conserved hypothetical protein, PF0496; 10, conserved hypothetical protein, PF0095; 11, transcription initiation factor TFIID chain a (TBP), PF1295; 12, conserved hypothetical proteins, PF0284 and PF1572.

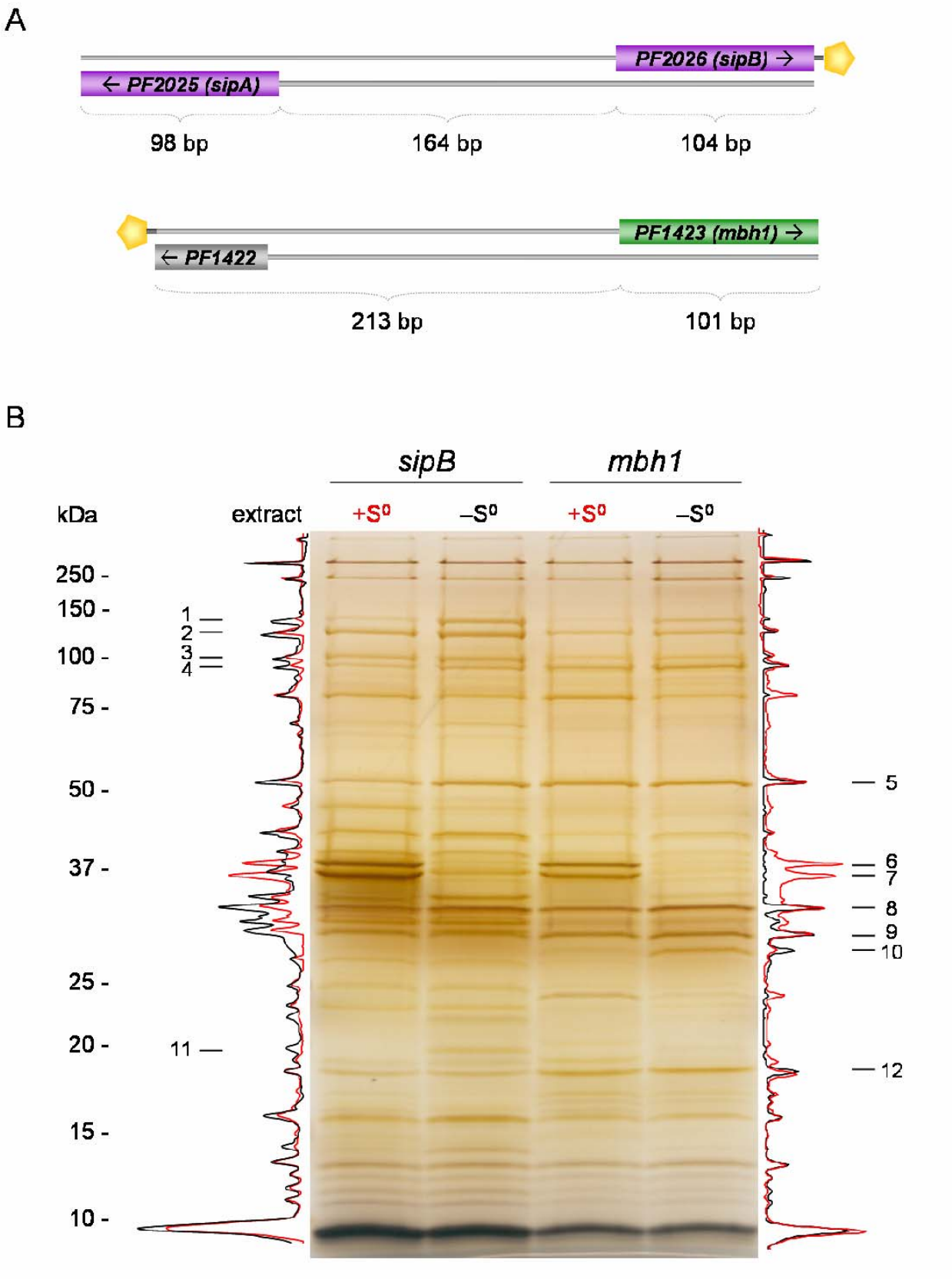
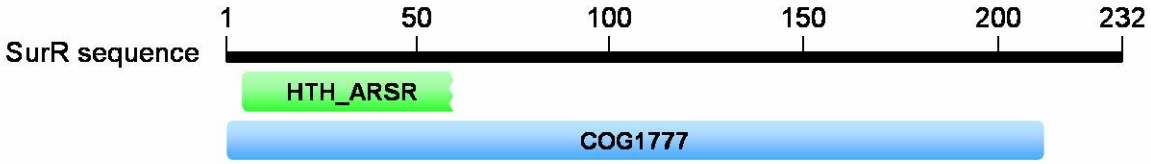

A
PF2026 (sipB) →
← PF2025 (sipA)
98 bp
164 bp
104 bp
PF1423 (mbh1) →
← PF1422
213 bp
101 bp
B
sipB
mbh1
kDa
extract
+S⁰
–S⁰
+S⁰
–S⁰
250 -
150 -
100 -
75 -
50 -
37 -
25 -
20 -
15 -
10 -
1
2
3
4
5
6
7
8
9
10
11
12

Figure 3.4 SurR (PF0095) protein sequence analysis. A. Conserved domain search results for SurR protein sequence from online tools available at NCBI [84, 85]. SurR sequence is represented by a black line with matching conserved domains indicated below. Conserved domain descriptions and sequence alignments are shown. For the sequence alignments, identical residues are colored red and similar residues are colored blue. **B.** PSIPRED [118, 119] secondary structure prediction of N-terminal HTH_ARSR domain of SurR. Helices are indicated in yellow according to their predicted positions in the sequence. **C.** Tree-view for results of a BLAST search of the SurR protein sequence against the NCBI non-redundant database [81]. SurR (PF0095) is shown boxed in red, proteins of the Thermococcaceae family are highlighted in yellow, and a homolog from *P. furiosus* is underlined (PF1744).

A



COG1777, Predicted transcriptional regulators
CD-Length = 217 residues, 98.2% aligned, Expect = $6x10^{-36}$

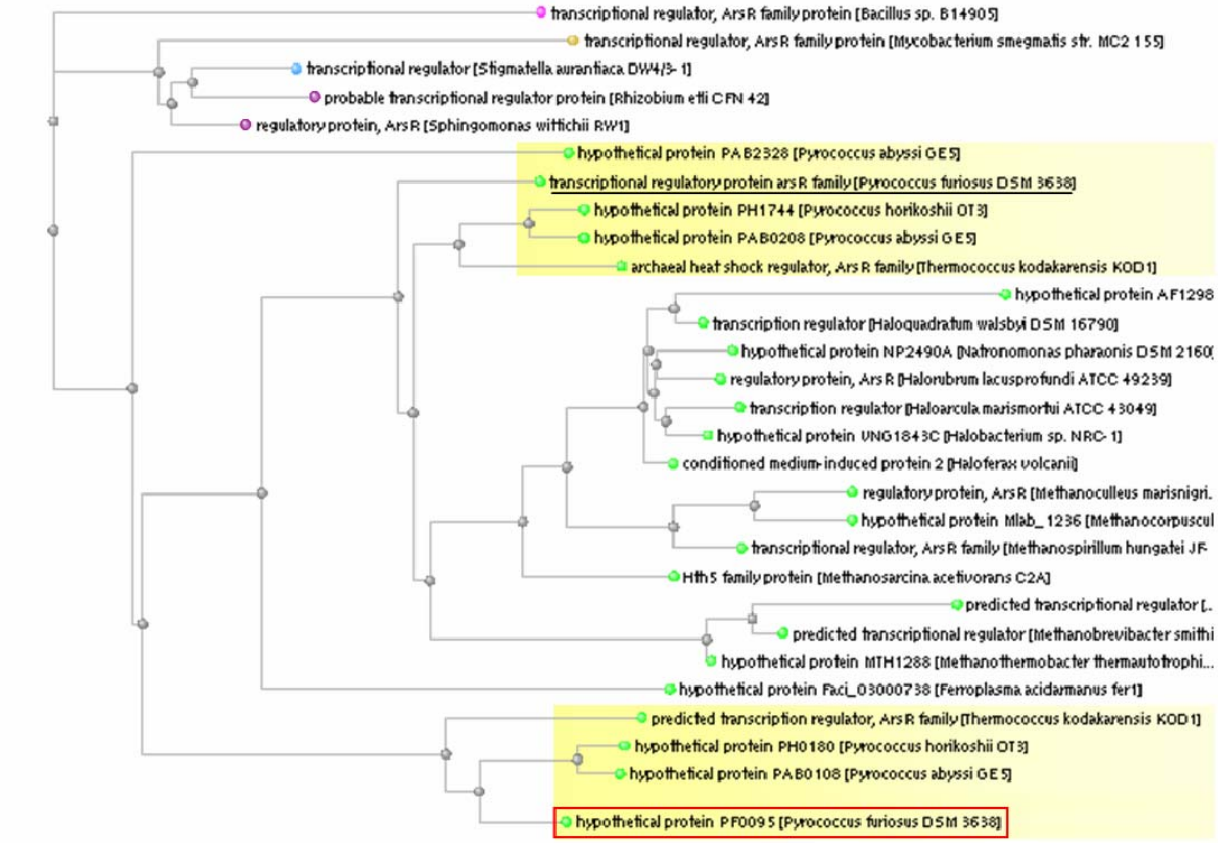
```
   SurR:  1   MMNMEPDLFYILGNKVRRDLLSHLTCMECYFSLLSSKVSVSSTAVAKHLKIMEREGVLQS  60
COG1777:  1   MVMDIDKILDVLGNETRRRILQLLTRRPCYVSEISRELGVSQKAVLKHLRILERAGLVES  60

   SurR: 61   -YEKEERFIGPTKKYYKISIAKSYVFTLTPEMFWYKGLDLGDAELRDFEISLSGLDTEPS  119
COG1777: 61   RIEKIPR--GRPRKYYMISRNLRLEVTLSPNFFGAERFDLEEDDLESERSEVSKLFKSPE  118

   SurR: 120  TLKEMITDFIKANKELEKVLEAFKTIESYRSSLMRKIKEAYLKEIGDMTQLAILHYLLLN  179
COG1777: 119  GISELISRLLEINREIEELSRAQTELQKQLNELMDRIKEEIEDKDGDMTERIVLEYLLKN  178

   SurR: 180  GRATVEELSDRLNLKEREVREKISEMARFVPVKIIN  215
COG1777: 179  GAADVEETSRRTVLKIEEVLEILAEKG-FVEIKEKN  213
```

HTH_ARSR, helix_turn_helix, Arsenical Resistance Operon Repressor
CD-Length = 79 residues, 70.9% aligned, Expect = 0.006

```
    SurR:  5   LFYILGNKVRRDLLSHLTCMECYFSLLSSKVSVSSTAVAKHLKIMEREGVLQSYEK  60
HTH_ARSR:  2   ILKALSDPTRLKILKLLAEGELSVCELAEILGLSQSTVSHHLKKLREAGLVESRRE  57
```

B

C

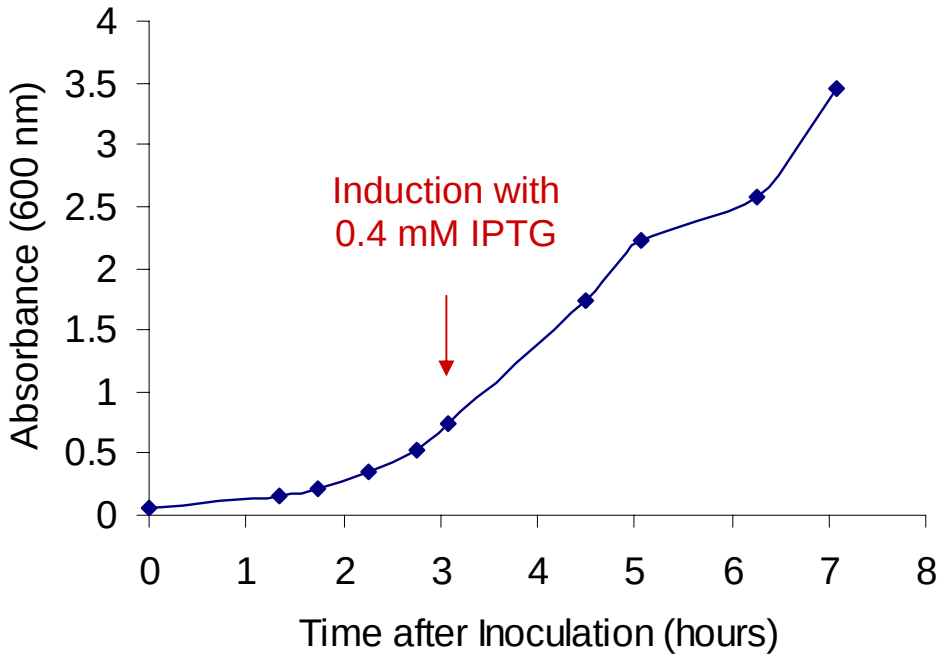


Blast names color map

- euryarchaeotes
- high GC Gram+
- δ-proteobacteria
- bacteria
- α-proteobacteria

Figure 3.5 Expression and purification of his_6-SurR. A. Growth curve for SurR expression. Protein production was induced with IPTG, and the culture was harvested after four hours of growth. For the growth curve, cell culture samples were diluted to obtain OD_{600} readings in the range of 0.1 to 0.8, and the readings were multiplied by the dilution factor to obtain the actual OD_{600} of the culture. **B.** SDS-PAGE of samples of each time point prior to and after induction which were normalized according to OD_{600} (left) and final purified protein after nickel-affinity purification (right).

A


Absorbance (600 nm)
4
3.5
3
2.5
2
1.5
1
0.5
0
Induction with
0.4 mM IPTG
0
1
2
3
4
5
6
7
8
Time after Inoculation (hours)

B

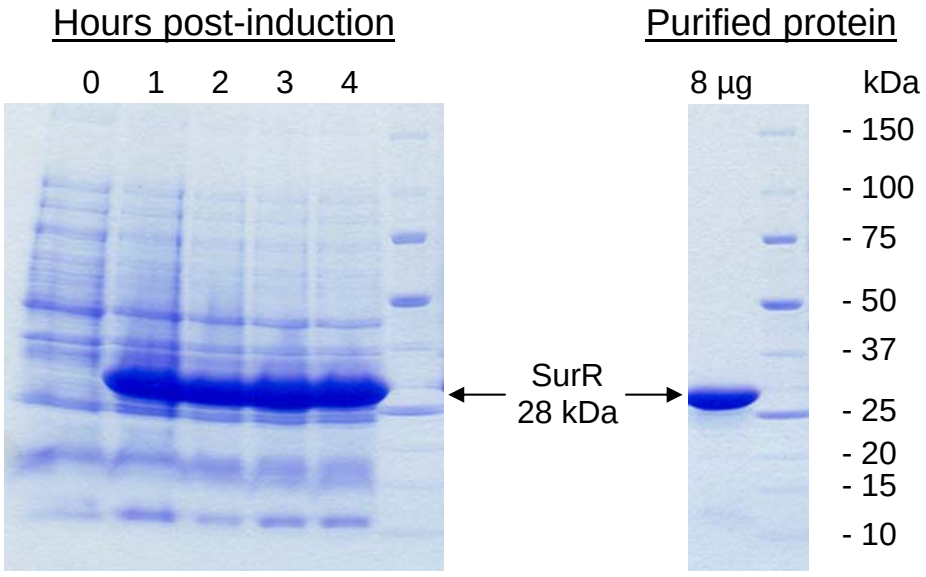

Hours post-induction
0
1
2
3
4
Purified protein
8 µg
kDa
- 150
- 100
- 75
- 50
- 37
- 25
- 20
- 15
- 10
SurR
28 kDa

Figure 3.6 SurR is specific to the *mbh1* probe. EMSA with his_6-SurR and *mbh1*, *sipB* and *lrpA* probes was performed. The DNA probes used are indicated at the top of each gel image with corresponding protein/DNA mole ratios listed above each lane. DNA (80 nM) was incubated with protein in buffer (50 mM Tris pH 7.5, 100 mM KCl, 5% glycerol, 1 mM DTT, 1 mM EDTA) for 20 min at 55°C. Gel was stained with SYBR Green I nucleic acid gel stain.

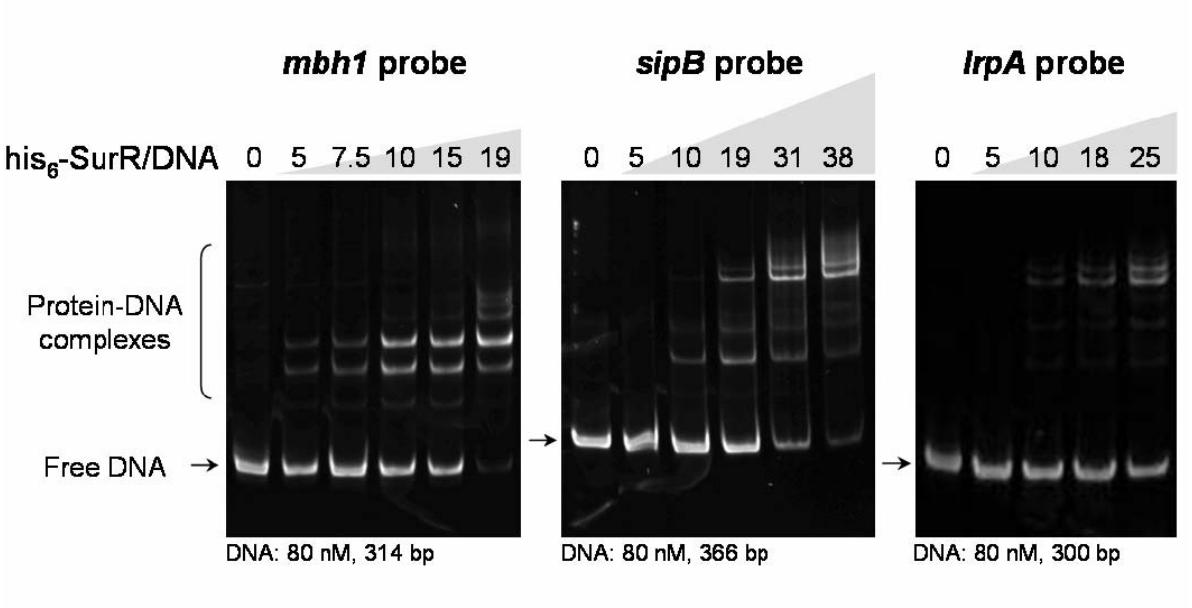

mbh1 probe
sipB probe
lrpA probe
his6-SurR/DNA
0 5 7.5 10 15 19
0 5 10 19 31 38
0 5 10 18 25
Protein-DNA complexes
Free DNA
DNA: 80 nM, 314 bp
DNA: 80 nM, 366 bp
DNA: 80 nM, 300 bp

Figure 3.7 His_6-SurR binds specifically to an *mbh1* promoter fragment. Diagrams of the DNA probes are shown (top). The probes used are indicated at the top of each gel image with corresponding protein/DNA mole ratios listed above each lane. DNA (80 nM) was incubated with protein in buffer (50 mM Tris pH 7.5, 100 mM KCl, 5% glycerol, 1 mM DTT, 1 mM EDTA) for 20 min at 55°C. Gel was stained with SYBR Green I nucleic acid gel stain.

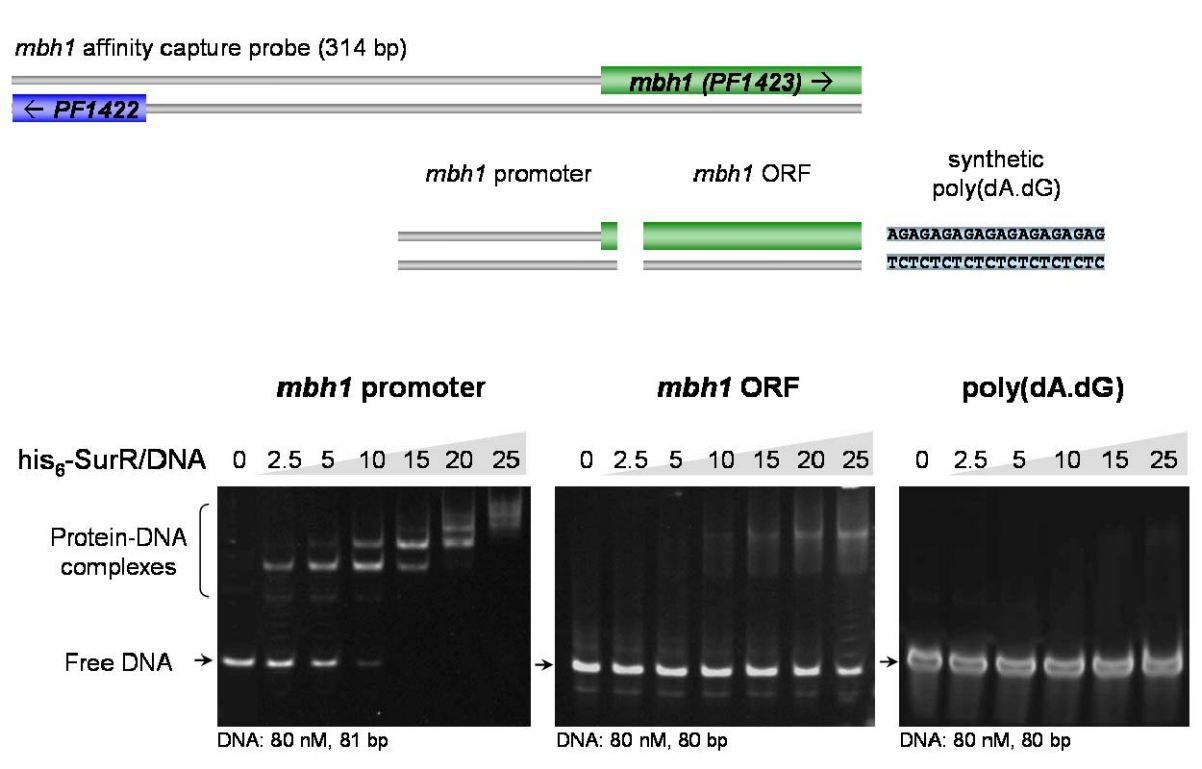

mbh1 affinity capture probe (314 bp)
mbh1 (PF1423) →
← PF1422
mbh1 promoter
mbh1 ORF
synthetic poly(dA.dG)
AGAGAGAGAGAGAGAGAG
TCTCTCTCTCTCTCTCTC
mbh1 promoter
mbh1 ORF
poly(dA.dG)
his6-SurR/DNA
0 2.5 5 10 15 20 25
0 2.5 5 10 15 20 25
0 2.5 5 10 15 25
Protein-DNA complexes
Free DNA
DNA: 80 nM, 81 bp
DNA: 80 nM, 80 bp
DNA: 80 nM, 80 bp

Figure 3.8 Expression and purification of his_6-TEV-SurR. A. Results for automated chromatographic purification of his_6-TEV-SurR from soluble cell extract using a nickel-affinity column with Binding Buffer (20 mM sodium phosphate pH 7.4, 0.5 M NaCl, 20 mM imidazole) and Eluting Buffer (20 mM sodium phosphate pH 7.4, 0.5 M NaCl, 0.5 M imidazole). UV trace is indicated in blue, and trace for percentage of Eluting Buffer is indicated in green. Fractions collected are indicated in red with corresponding numbers. Soluble cell extract (~20 mL) was loaded at 1 mL/min followed by a ~10-mL wash with binding buffer. Protein was eluted with a 10-mL 0-20% gradient of Eluting Buffer followed by a 10-mL 20-100% gradient. **B.** SDS-PAGE gel indicating expression of his_6-TEV-SurR before and four hours after induction with IPTG. Sample loading volume was normalized according to culture OD_{600} readings (1.204 and 5.082, respectively). **C.** SDS-PAGE gel of fractions collected from elution step of nickel-affinity chromatography. The indicated fractions were pooled for buffer-exchange into 20 mM HEPES, 100 mM NaCl, pH 7.6.

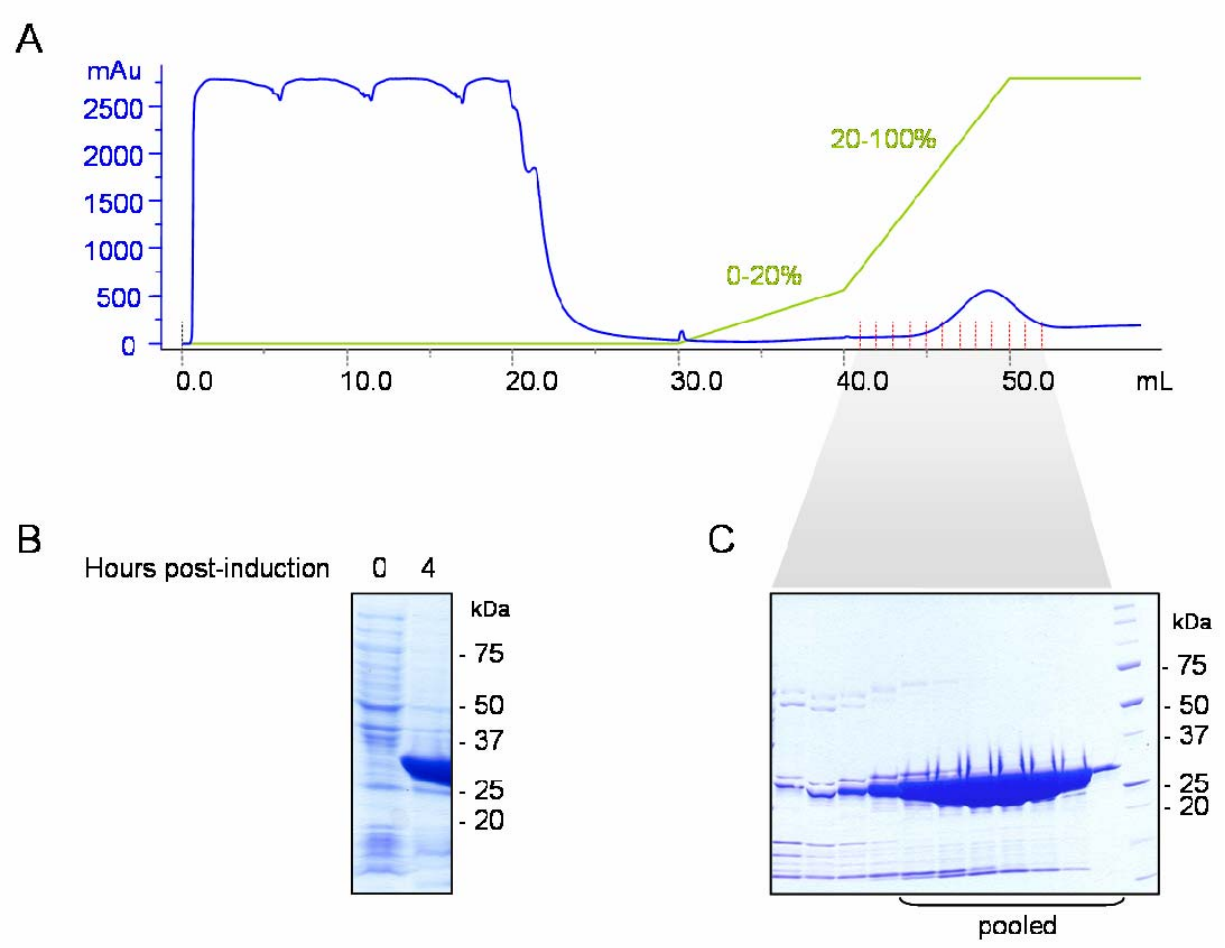
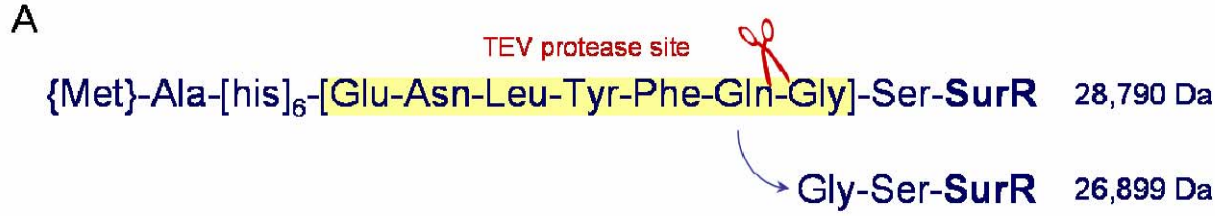
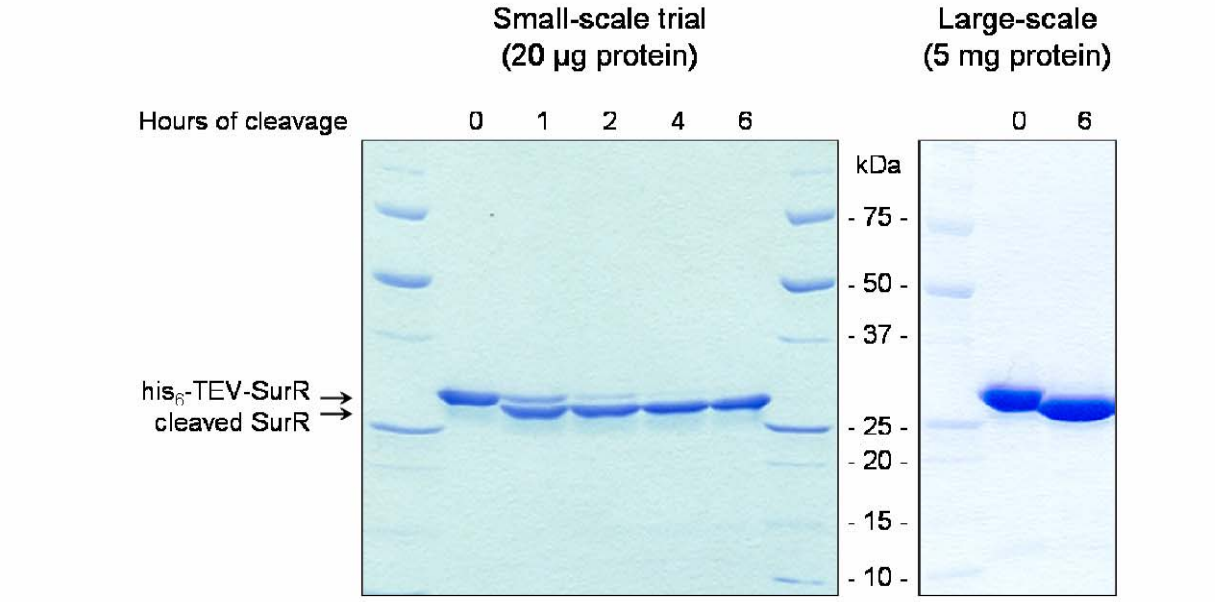

A
mAu
2500
2000
1500
1000
500
0
20-100%
0-20%
0.0
10.0
20.0
30.0
40.0
50.0
mL
B
Hours post-induction
0
4
kDa
75
50
37
25
20
C
kDa
75
50
37
25
20
pooled

Figure 3.9 His-tag cleavage and final purification of SurR. A. Amino acid sequence of N-terminal addition to SurR protein sequence is shown before and after cleavage with TEV protease (red scissors) at the recognition site (yellow highlight). Corresponding protein molecular weights are shown adjacent to each sequence. **B.** SDS-PAGE gel showing cleavage of his_6-TEV-SurR at various time points for small-scale (left) and large-scale (right) cleavage reactions. The cleaved product is easily distinguishable from the uncleaved product on SDS-PAGE. **C.** Results for automated chromatographic purification of SurR from contaminating proteins, uncleaved his_6-TEV-SurR, and his_6-AcTEV protease (Invitrogen) using a second nickel-affinity column with Binding Buffer (20 mM HEPES pH 7.6, 100 mM NaCl) and Eluting Buffer (20 mM sodium phosphate pH 7.4, 0.5 M NaCl, 0.5 M imidazole). UV trace is indicated in blue and trace for percentage of Eluting Buffer is indicated in green. After the TEV protease cleavage reaction, sample was loaded onto the second nickel affinity column at 0.2 mL/min and the flow-through containing the tagless protein was collected.

A


TEV protease site
{Met}-Ala-[his]6-[Glu-Asn-Leu-Tyr-Phe-Gln-Gly]-Ser-SurR
28,790 Da
Gly-Ser-SurR
26,899 Da

B


Small-scale trial
(20 µg protein)
Large-scale
(5 mg protein)
Hours of cleavage
0 1 2 4 6
0 6
kDa
- 75 -
- 50 -
- 37 -
- 25 -
- 20 -
- 15 -
- 10 -
his6-TEV-SurR
cleaved SurR

C

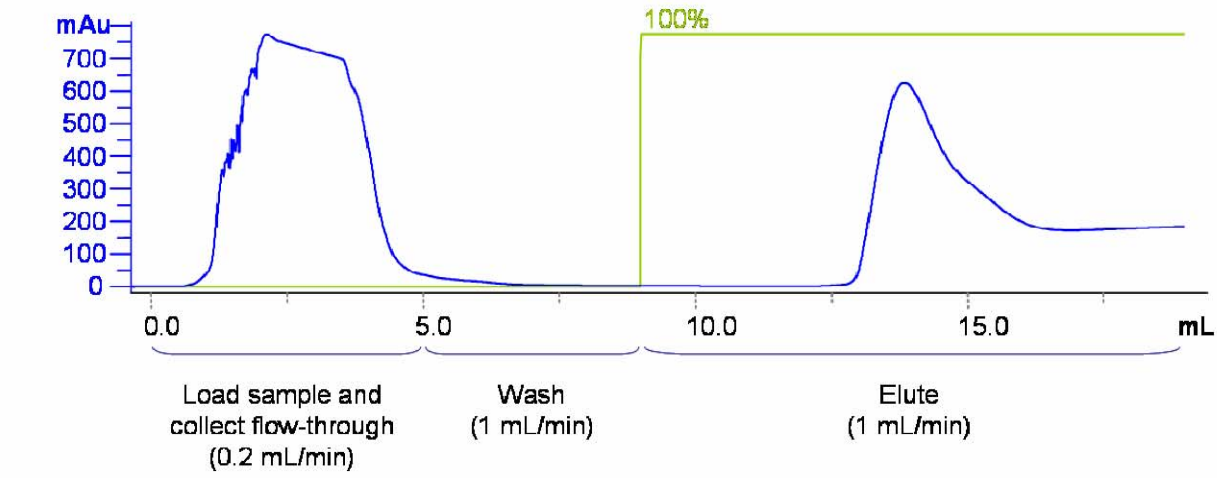
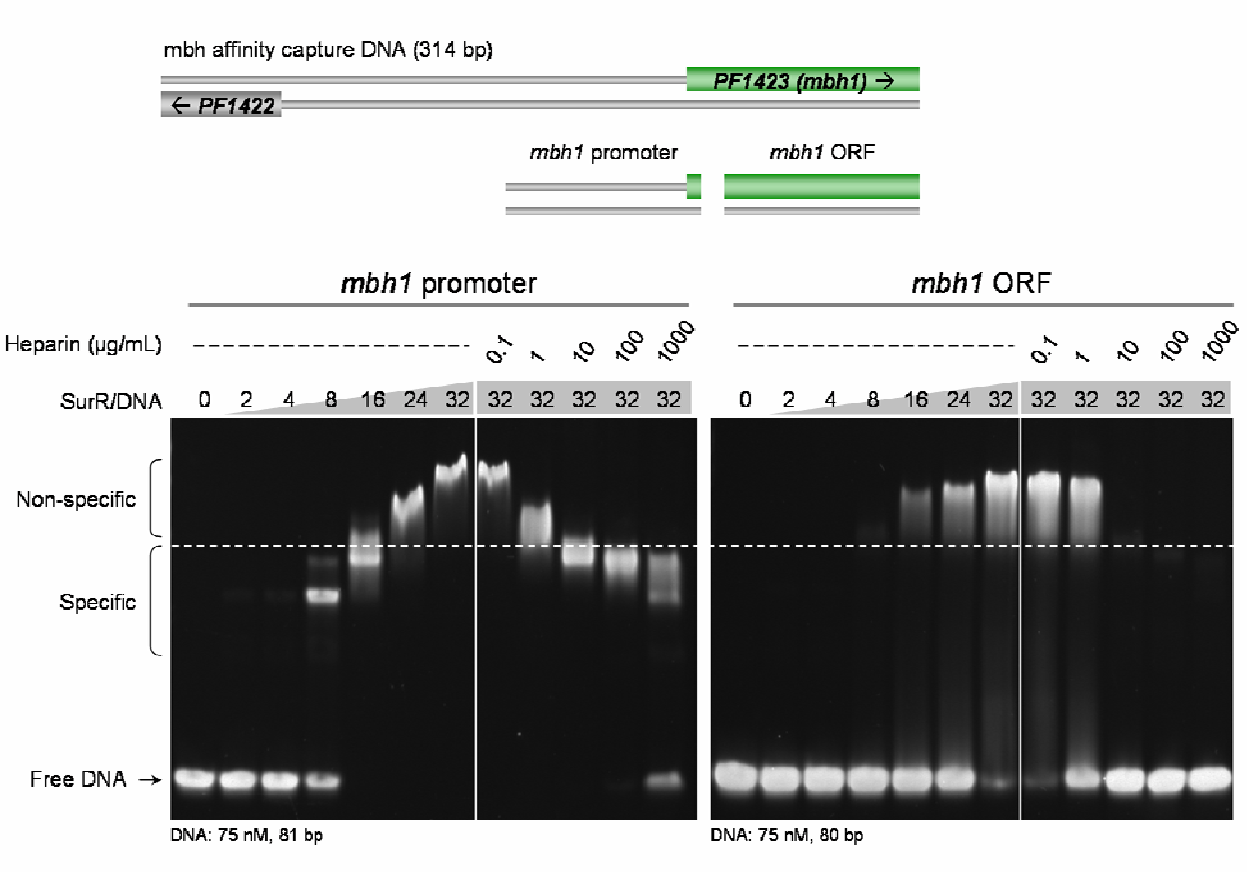

mAu
700
600
500
400
300
200
100
0
100%
0.0
5.0
10.0
15.0
mL
Load sample and
collect flow-through
(0.2 mL/min)
Wash
(1 mL/min)
Elute
(1 mL/min)

Figure 3.10 SurR binds specifically to *mbh1* promoter fragment in the presence of heparin. Diagrams of the DNA probes are shown (top). The probes used are indicated at the top of each gel image with corresponding protein/DNA mole ratios listed above each lane. DNA (75 nM or ~3.7 μg/mL) was incubated with protein in buffer (50 mM HEPES pH 7.5, 200 mM KCl, 5% glycerol, 1 mM EDTA) with and without heparin (concentrations indicated) for 20 min at 70°C. The dotted line across the gel separates nonspecific from sequence-specific protein-DNA complexes. Gel was stained with SYBR Green I nucleic acid gel stain.


mbh affinity capture DNA (314 bp)
PF1423 (mbh1) →
← PF1422
mbh1 promoter
mbh1 ORF
mbh1 promoter
mbh1 ORF
Heparin (µg/mL)
0.1
1
10
100
1000
SurR/DNA
0 2 4 8 16 24 32 32 32 32 32 32
0 2 4 8 16 24 32 32 32 32 32 32
Non-specific
Specific
Free DNA →
DNA: 75 nM, 81 bp
DNA: 75 nM, 80 bp

CHAPTER 4

SURR IS A REGULATORY TRANSCRIPTION FACTOR RELATED TO PYROCOCCUS FURIOSUS SULFUR RESPONSE

4.1 Identification of SurR binding sites on *mbh1* promoter

To define the site in the *mbh1* promoter region bound by SurR, fluorescence-based DNase I footprinting was performed (Fig. 4.1) [88]. The *mbh1* probe DNA was cloned into pUC18 for ease of probe amplification and synthesis of probe-specific A+G and C+T sequence ladders[1]. Primers used to amplify the footprinting probe from pUC18 were fluorescently-labeled at their 5' ends, one with HEX and the other with 6FAM. The PCR-amplified dual-labeled probe enabled both strands from one footprinting experiment to be analyzed simultaneously using a capillary electrophoreses automated DNA sequencer. Internal standards GS-500 ROX (labeled with ROX, Applied Biosystems) and probe specific A+G and C+T ladders (labeled with NED) were used as size standards.

Footprinting with SurR was carried out for the PF1423 probe following binding conditions determined by EMSA. Surprisingly, the results indicated the presence of two distinct footprints in the intergenic region between *mbh1*and PF1422: a proximal site extending from -5 to -43 (length ~39 bases) and a distal site extending from -107 to -159 (length ~53 bases) relative to the *mbh1* translation start site on the *mbh1* coding strand (Fig. 4.2 A, B). The footprints on the opposite strand were located at similar positions, -10 to -39 (length ~30 bases) and -112 to -164

[1] All pUC18-derived probes mentioned in this work will be referred to by the locus name of the target ORF region they contain.

(length ~53 bases) from the *mbh1* start (Fig. 4.2 C, D). The *mbh1*-distal site extends into the divergent PF1422 ORF on both strands with 63 bases separating the distal site from the proximal on the *mbh1* coding strand and 72 bases on the opposite strand. The fact that the protein can produce two footprints of differing sizes suggests that multiple proteins may be binding to adjacent motifs within each footprint, in effect oligomerizing[1] onto the DNA to different extents. For footprinting experiments in which SurR was titrated, the larger *mbh1*-distal site began to be occupied first (site 1), followed by the smaller site immediately adjacent to the *mbh1* start (site 2), as can be seen in Figure 4.3 comparing SurR/DNA mole ratios of 12 and 24. This result indicates that site 1 is a higher-affinity binding site than site 2 in the *mbh1* promoter region. There is no indication of a gradual enlargement of either footprint which might suggest 'nucleation' at a higher-affinity motif within the footprint followed by binding at adjacent sites.

Some peaks within and surrounding the footprints displayed noticeably increased fluorescence intensity, signifying that nucleotides at these positions were hypersensitive to DNase I when the protein was bound to the DNA (see arrows, Fig. 4.2). These hypersensitive sites could be indicative of DNA bending by the protein, which would make nucleotides at the bend positions more susceptible to DNase I cleavage. The footprints were most evident at protein/DNA mole ratios of 20 to 30, comparable to the ratios required for complete shifting of the *mbh1* probes in EMSA. Increasing the protein/DNA ratio up to 80 resulted in complete protection of the DNA from digestion (not shown), and this is further evidence of the nature of the low-mobility non-specific complexes observed in EMSA.

[1] The term 'oligomerizing' as used throughout this work does not imply that the SurR proteins themselves are forming oligomers, but simply that the proteins in conjunction with target DNA are forming various 'oligomeric' complexes.

4.2 SurR binds to its own promoter region shared with S^0-regulated ORF PF0094

During the course of experiments, it became evident that the gene which encodes SurR, PF0095, lies immediately downstream and is divergently oriented from PF0094, one among the original list of ORFs which showed significant regulation in the DNA microarray gene expression profiling experiments for *P. furiosus* cultures grown in the presence versus the absence of S^0. PF0094 encodes a biochemically and structurally characterized enzyme, protein disulfide oxidoreductase (PDO) [105, 120]. In contrast to the down-regulation of the membrane-bound hydrogenase operon whose promoter was the target DNA from which SurR was discovered, *pdo* expression is up-regulated in the presence of S^0 for both batch and kinetic DNA microarray experiments [78, 100]. With only 134 bp of intergenic space between the *pdo* and *surr* genes, it became immediately appealing to test this DNA as a probe in EMSA with SurR. Most of the archaeal regulators characterized thus far have been shown to be auto-regulatory, and this case was even more interesting because of the S^0-induced up-regulation of *pdo* gene expression.

To determine if SurR bound to this promoter region, a probe DNA was designed that included the full 134-bp *pdo-surr* intergenic region and was tested in EMSA with varying concentrations of SurR. The probe DNA was completely shifted at a protein/DNA mole ratio of 24, displaying multiple protein-DNA complexes of varying electrophoretic mobility (Fig. 4.4). Interestingly, five distinct complexes were apparent for this probe DNA as compared to the three complexes observed for the *mbh1* probes (compare with Fig. 3.7). With increasing protein/DNA mole ratios, there was a gradual shift toward the lower mobility complex ("Complex 5" in Figure 4.4), followed by a recognizable shift to excess non-specific binding at a protein/DNA mole ratio of 32.

In order to investigate the differences observed in EMSA between the *mbh1* probes and the *pdo-surr* intergenic probe, a larger DNA fragment encompassing the *pdo-surr* intergenic region was cloned into pUC18 in order to produce a probe for use in footprinting experiments with SurR. Footprinting results for this promoter region were surprisingly different than those for the PF1422-*mbh1* promoter region, with the presence of only one footprint which extended over an 83-bp region of DNA immediately upstream from the *pdo* start site, as can be seen in Figure 4.5. Specifically, SurR protected bases -4 to -86 from the *pdo* translation start on the *pdo* coding strand and -8 to -90 from the *pdo* translation start on the opposite strand. The total size of this footprint equals the combined length of the PF1422-*mbh1* footprints, and perhaps this extended region can somehow accommodate more oligomerization of the protein onto the DNA at specific binding motifs, causing the appearance of an increased number of distinct complexes in EMSA. Another interesting difference with this footprint is the notable lack of DNase I hypersensitive sites, with the exception of one possible site in the footprint on the *surr* coding strand (see arrow, Fig. 4.5 D). Also, SurR binding within the footprint region appears to occur simultaneously across the entire footprint as evidenced by the reduction in peaks across the whole footprint region at a protein/DNA mole ratio of 12 (Fig. 4.6).

4.3 The SurR recognition site contains the palindrome GTTn$_3$AAC

To determine and define the consensus DNA motif recognized by SurR, the artificial selection method SELEX was employed [93, 94]. Starting with a pool of synthetic DNA containing central 30-bp randomized sequences, EMSA with SurR was used to isolate 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. A diagram of this selection scheme is displayed in Figure 4.7. After six rounds of selection, the consensus palindromic DNA sequence $GTTn_3AAC$ was uncovered in 19 total sequenced SELEX DNAs, with the motif occurring most strongly in 11 out of 19 of the sequences (Fig. 4.8). Only 6 out of 19 sequences displayed the perfect palindrome $GTTn_3AAC$. Had further rounds of selection been carried out, a stronger motif may have been elucidated; however, it was of interest to obtain a motif which displayed some of the variability allowed at each position, not the highest affinity motif.

Each DNase footprint site in the *mbh1* promoter was verified to contain one perfect $GTTn_3AAC$ motif, and the 83-bp footprint in the *pdo-surr* intergenic region contained two perfect SELEX consensus motifs (Fig. 4.9). The presence of the $GTTn_3AAC$ motifs in the SurR footprint regions confirmed that the consensus derived from the synthetic SELEX library had some relationship to the *in vivo* recognition site for SurR. It is likely that additional sequence elements play a role in SurR binding; however, the $GTTn_3AAC$ motif is very likely an essential part of the SurR recognition site. For three of the $GTTn_3AAC$ consensus motifs in these two promoter regions, an additional palindrome half-site was positioned 5 bp away from the consensus, indicating the possibility that adjacent motifs degenerate from the $GTTn_3AAC$ motif could exist in series at defined intervals. The different sized footprints in the *mbh1*and *pdo-surr* promoter regions together with the multiple protein-DNA complexes observed in EMSA for these two promoter regions are a good indication of multiple SurR proteins binding to adjacent motifs.

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

In order to take full advantage of the information gained from SELEX on the consensus DNA-binding site recognized by SurR, the $GTTn_3AAC$ sequence was used to search for other potential binding sites in the upstream regions of other genes within the genome. A colleague, Darin Cowart, created a database of ORF upstream sequences (UORs) extracted from the *P. furiosus* genome which could be used in conjunction with his "Search by Motif" software to search for all occurrences of the motif in *P. furiosus* ORF upstream DNA. Searching with the string $GTTn_3AAC$ returned 148 UORs containing the motif with a total of 123 nonredundant 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) (see Appendix Table A2). This number, although high, is in sharp contrast to the 944 hits returned from a search for $GTTn_3AAC$ in the entire *P. furiosus* genome.

Presumably, not all of the $GTTn_3AAC$ motifs identified are genuine SurR binding sites as it seems that the SELEX motif is too short to be the only prerequisite for specific binding of SurR to DNA; not enough is known about the mechanism of SurR binding specifically to DNA to be able to bioinformatically separate out false positives. However, one more piece of information allowed the identification of a subset of UORs which likely contained authentic SurR binding sites. Analysis of the sequence surrounding the SELEX consensus motifs within the known SurR DNA footprints gave rise to the possibility of searching the UOR database with an extended motif that included a second palindrome 'half-site' with a 5-bp gap: $GTTn_3AACn_5GTT$. This search returned only 24 UORs containing the motif with a total of 16 non-redundant motifs (Table 4.1). Interestingly, over half (15 out of 24) of the corresponding ORFs were part of divergently oriented gene pairs sharing intergenic regions shorter than 260 bp.

For those ORFs that were predicted to be part of operons [121], all but one (PF0913) were first in their operons and therefore more likely to contain a promoter with potential regulatory elements in their upstream DNA. Furthermore, the ORFs found to contain this sequence in their upstream DNA overlapped to a high degree with S^0-regulated ORFs, particularly with the ORFs affected within 10 minutes after S^0 addition in the kinetic DNA microarray experiment [100]. As can be seen in Table 4.2, ten out of twelve of the 10-min down-regulated ORFs/operons have the extended motif in their upstream DNA, and four out of five of the up-regulated ORFs/operons contain the SELEX consensus motif, though only two contain the extended motif, in their upstream DNA. It is worth noting that none of the ORFs which show up or down-regulation only after 30 min of S^0 addition, an effect which constitutes the secondary response to S^0 [100], have the $GTTn_3AAC$ motif in their upstream DNA. These results clearly point to a relationship between SurR target genes and the primary response of *P. furiosus* to the presence of S^0.

Table 4.1 UOR database search results[1] for the motif $GTT(n)_3AAC(n)_5GTT$

UOR[a]	Motif	Start[b]	Stop[b]	ORF Annotation[c]
PF0094	GTTGTCAACCTTAGGTT	-82	-66	Protein disulfide oxidoreductase, *pdo* [105]
PF0095	GTTGTCAACCTTAGGTT	-67	-51	Sulfur-response regulator, *surr* [this work]
PF0531	GTTGGTAACAAAATGTT	-84	-68	*cobalamin biosynthesis protein m*
PF0532	GTTACCAACGTAAGGTT	-100	-84	*hypothetical protein*
PF0547	GTTGAAAACTTCAAGTT	-190	-174	*hypothetical protein*
PF0548	GTTGAAAACTTCAAGTT	-67	-51	*hydrogenase expression/formation protein*
PF0559	GTTTCTAACTTTTGGTT	-48	-32	*hydrogenase expression/formation regulatory protein*
PF0568	GTTTATAACGCTATGTT	-24	-8	*hypothetical protein*
PF0569	GTTTATAACGCTATGTT	-34	-18	*hypothetical protein*
PF0891	GTTTTTAACCTTTGGTT	-142	-126	Hydrogenase I beta, *hydB1* [102]
PF0913	GTTCAAAACCAAAGGTT	-197	-181	*Mo formylmethanofuran dehydrogenase related*
PF0914	GTTCAAAACCAAAGGTT	-81	-65	*Mo formylmethanofuran dehydrogenase related*
PF0915	GTTAGAAACCTTAGGTT	-58	-42	*hypothetical protein*

[1] The *P. furiosus* UOR database and corresponding motif searching software was created by Darin Cowart (University of Georgia).

PF0926	GTTTGGAACCTATTGTT	-81	-65	*hypothetical protein*
PF1100	GTTTGCAACTCGTAGTT	-108	-92	*hypothetical protein, putative hydrogenase*
PF1186	GTTTAAAACCTTTAGTT	-34	-18	NAD(P)H sulfur reductase (NSR) [100]
PF1328	GTTCAAAACCTAAGGTT	-87	-71	*hydrogenase gamma subunit*
PF1329	GTTCAAAACCTAAGGTT	-88	-72	Hydrogenase II beta, *hydB2* [103]
PF1422	GTTATAAACCAAAAGTT	-19	-3	*thioredoxin reductase*
PF1423	GTTATAAACCAAAAGTT	-146	-130	Membrane bound hydrogenase, *mbh1* [101]
PF1423	GTTTCAAACCAAATGTT	-31	-15	Membrane bound hydrogenase, *mbh1* [101]
PF1516	GTTTACAACTTTATGTT	-58	-42	*GMP synthase subunit B*
PF1621	GTTAGTAACTAAAAGTT	-78	-62	*hypothetical protein*
PF1622	GTTAGTAACTAAAAGTT	-196	-180	*n-type ATP pyrophosphatase superfamily*

[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] Characterized ORFs/proteins are listed with the corresponding reference, and ORF annotations in italics are from REFSEQ or TIGR.

Table 4.2. Summary of motifs found upstream of 10-minute S^0-regulated ORFs

ORF / operon[a]	ORF description / operon function[b]	$[GTT(n)_3AAC(n)_5GTT]$[c]	$[GTT(n)_3AAC]$[c]
Up-regulated >3-fold 10 minutes after S^0-addition [100]			
PF0094	Protein disulfide oxidoreductase [105]	YES	YES (2)
PF0261-0262	*unknown transporter*	NO	NO
PF1186	NAD(P)H sulfur reductase (NSR) [100]	YES	YES
PF1453-1441	Membrane bound oxidoreductase, MbxA-N [122]	NO	YES (2)
PF2051-2052	*putative transcriptional regulators*	NO	YES
Down-regulated >3-fold 10 minutes after S^0-addition [100]			
PF0450	*Glutamine synthetase, catalytic region*	NO	NO
PF0531-0529	*Cobalt transport*	YES	YES
PF0559	*Hydrogenase maturation protein, HypF*	YES	YES
PF0736	*Conserved hypothetical protein*	YES (300 bp)	YES (300 bp)
PF0736.1	*Conserved hypothetical protein*	NO	NO
PF0891-0894	Hydrogenase I [102]	YES	YES
PF0913	*Formylmethanofuran dehydrogenase, subunit E*	YES	YES (2)
PF0915	*Cytochrome c biogenesis protein*	YES	YES
PF0926-0925	*Unknown*	YES	YES
PF1329-1332	Hydrogenase II [103]	YES	YES
PF1423-1436	Membrane Bound Hydrogenase, MbhA-N [101]	YES	YES (2)
PF1621	*Fibronectin, type III-like fold*	YES	YES

[a] For predicted operons, the first ORF of the operon (from which the motif was found upstream) is listed first, followed by the last ORF of the operon.
[b] ORF descriptions and operon functions are essentially derived from [100]. Italics indicate predicted function, while characterized ORFs/proteins are listed with the corresponding reference.
[c] Motifs were identified in 200 bp UORs except for PF0736 in which the motif was found in a 300 bp UOR. Some UORs had two $GTTn_3AAC$ motifs as indicated by the (2).

This new connection between 10 min S^0-regulated ORFs and apparent SurR targets yielded an opportunity to better understand the SurR binding motif. The same motif-finding software that was used to elucidate the SurR $GTTn_3AAC$ motif from the selected SELEX library was applied to the $GTTn_3AAC$-containing 200-bp UORs of the S^0-regulated ORFs (see Table 4.2). Since the $GTTn_3AAC$ motif was derived from a synthetic library, it was of interest to see if the motif could be extended to better represent the *in vivo* binding motif of SurR. MEME motif finding software [95] was used to search forward and reverse strands of (1) all thirteen of the UORs, (2) the nine UORs of the up-regulated ORFs, and (3) the four UORs of the down-regulated ORFs for a common motif. The resulting motif diagrams can be seen in Figure 4.10. The consensus motif for all thirteen UORs of both up and down-regulated ORFs combined clearly showed the $GTTn_3AAC$ consensus motif. Not surprisingly, adjacent palindrome 'half-sites' were observed with the five-base spacing from the $GTTn_3AAC$ motif. The motif for the UORs of the down-regulated ORFs was very similar, and it was essentially the reverse-complement of it. The motif found in the UORs of the four up-regulated ORFs was shorter; whether this is due to a small sample size or to a difference in the way SurR binds to these UORs versus UORs of the down-regulated ORFs is not known. The adjacent half-sites were not as apparent, and there is an additional conserved guanine in front of the SELEX motif, making the consensus motif for these four UORs to be '$GGTTn_3AAC$'. One consideration with the new genomic motif information is the possibility that the recognized palindrome for SurR is not

$GTTn_3AAC$ but instead $AACn_5GTT$ and that perhaps the motifs must occur in pairs at defined intervals in order to direct specific high-affinity binding of SurR. Furthermore, the genomic motifs indicated the relevance of A/T tracts in the 5-bp spaces (Fig. 4.10 A, C).

4.5 Validation of additional SurR binding sites in promoters of S^0-regulated ORFs

4.5.1 Selection of putative promoter regions to test

The presence of the extended SELEX motif in the putative promoter regions of the ORFs which were both up and down-regulated within 10 min of S^0-addition gave good indication that SurR is a regulator involved in *P. furiosus* S^0 response. In order to verify that the identified motifs represented bona fide SurR binding sites, several of the putative promoter regions were selected for further study in conjunction with SurR. To streamline the process of testing several new DNA probes in EMSA and footprinting experiments, selected promoter-ORF DNA was cloned into pUC18 for amplification with the M13 sequencing primers. Since many of the S^0-regulated ORFs were part of divergently oriented gene pairs with relatively short intergenic regions, promoter-ORF DNA was designed such that the length of each ORF in the probe was different enough so as to be able to distinguish RNA transcripts from either ORF if the probe were used as an *in vitro* transcription template.

Besides PF1423 (*mbh1*) the upstream DNA of three additional S^0 down-regulated ORFs were chosen for EMSA and footprinting analysis with SurR: PF0559, PF0891, and PF0531. The PF0559 ORF was of interest because it encodes a potential hydrogenase maturation protein, and also because this was the most highly down-regulated ORF in the batch microarray experiment, with 39-fold less expression in a culture grown in the absence versus the presence of S^0 [78]. PF0891 is the start of a four-ORF operon encoding one of the soluble hydrogenases of *P.*

furiosus [102, 104], and was selected since the primary response to S^0 appears to be down-regulation of hydrogenases. PF0531 encodes a 'cobalt transporter' according to annotation, and it is likely the start of a four-ORF operon, two ORFs of which are also down-regulated with S^0 (PF0530 and PF0529). This ORF/operon has no apparent connection with S^0-metabolism as of yet.

With the exception of PF0094 which has already been shown to have a binding site, all of the remaining up-regulated ORFs that contained the $GTTn_3AAC$ motif were selected for study: PF1186, PF1453, and PF2051. All three of these ORF regions are very interesting in terms of their relationship to S^0 response. The protein product of PF1186 has recently been identified as the sulfur reductase of *P. furiosus*, termed NSR for NAD(P)H Sulfur Reductase [100]. PF1453 is the start of a 14-ORF operon (PF1453-PF1441) encoding a putative hydrogenase which likely serves as a replacement for the membrane-bound hydrogenase when S^0 is present [122, Schut, 2007 #144]. PF2051 and PF2052 encode two probable transcriptional regulators which may function in the secondary response to sulfur which occurs within 30 min after S^0 addition and consists of regulation of genes involved primarily in amino acid metabolism [100].

4.5.2 Confirmation of specific SurR binding by EMSA

EMSA was used initially as an indicator of whether SurR bound these selected promoter-ORF DNAs. From previous EMSAs, the presence of multiple distinct bands of protein-DNA complexes and complete shifting of the DNA at reasonable protein/DNA mole ratios constituted specific binding of DNA by SurR. EMSA confirmed that SurR bound to all promoter-ORF DNAs that were tested, as can be seen in Figure 4.11; however, the number of protein-DNA complexes formed with each probe were notably different. Incidentally, there appeared to be no

correlation between the number of protein-DNA complexes observed in EMSA and the S^0-mediated up- or down-regulation of the corresponding ORFs. The presence of multiple bands reinforces the notion that multiple proteins may bind to adjacent recognition sites, and that the various promoter regions may have different numbers of these sites resulting in variability in the number of protein-DNA complexes observed in EMSA.

For the probes of the down-regulated genes PF0531, PF0891 and PF0559, the number of protein-DNA complexes observable from the addition of protein until the DNA was completely shifted was two, three and possibly five, respectively. The EMSA for the PF0559 probe was not carried out with high enough protein-DNA ratios to cause a complete shift of the DNA; however, the multiple defined bands observed at a ratio of 24 is enough to indicate specific binding. Approximately half of the free DNA is shifted at this ratio, and therefore it is reasonable to assume that a complete shift might occur at a ratio of 40. The probes of the up-regulated ORFs PF2051, PF1453 and PF1186 displayed three, six and two distinct protein-DNA complexes, respectively. The intensity of the shifted bands varied, implying that certain oligomer conformations are more stable on the DNA than others. There were clear affinity differences, as judged by the protein/DNA mole ratio required for a complete shift of the probe DNA; however, for a more accurate gauge, the extent of the SurR footprint on each probe must be determined. The PF0891 and PF1186 probes shifted the DNA completely between protein/DNA mole ratios of 16 and ~20; however only two to three protein-DNA complexes were observable. On the other hand, PF0531 and PF2051 probes also had between two and three complexes and these shifted at protein/DNA mole ratios of 32 and 40, respectively. PF1453 and PF0559 displayed the highest number of complexes (five and six), and therefore it is not surprising that these might require higher amounts of protein in order to cause a complete shift of the DNA (a ratio of 40 or greater).

The upstream DNA of some of the S^0-regulated ORFs contained two conserved $GTTn_3AAC$ motifs, and the probes for these ORFs displayed the higher number of protein-DNA complexes (above three), for the most part: *mbh*, *pdo-surr*, and PF1453. Promoter DNA for *mbh1* shifts completely through formation of at least four protein-DNA complexes; however, since there are two footprints, the total shift of free DNA may only indicate complete binding at the higher-affinity *mbh1*-distal site (site 1). Bands of even lower mobility are evident above the predominant complex formed when almost all of the free DNA has been shifted (see Fig. 3.6), indicating that higher-order oligomerization may take place on this promoter-ORF probe. By contrast, the shorter 81-bp mbh-promoter probe which contains most of the *mbh1*-proximal footprint (site 2) shows a maximum of three complexes before the shift to a non-specific complex (compare figures 3.7 and 3.10). The EMSA with the probe for the *pdo-surr* promoter region clearly displays the presence of five protein-DNA complexes (see Fig. 4.4). Similarly, EMSA with the PF1453 probe reveals a total of six complexes formed. The PF0559 probe, showing up to five complexes in EMSA, does not seem to follow this trend, as it only has one conserved $GTTn_3AAC$ motif; however, inspection of the PF0559 upstream DNA revealed that it contains an additional $AACn_5GTT$ motif besides the known $GTTn_3AACn_5GTT$ extended motif. This could indicate the presence of degenerate SELEX motifs which may help direct the oligomerization of SurR on this putative promoter.

4.6 Detailed analysis of SurR binding sites with DNase I footprinting

4.6.1 SurR footprints are present on each EMSA-tested probe

In order to verify the location of the SurR binding sites on these various promoter-ORF DNAs, fluorescence DNase I footprinting was performed. Probe-specific ‘A + G’ or ‘C + T’

ladders were used to identify the footprint locations within the probe sequences, and for the PF0559 probe, the internal size marker was used. Probes were purified by either agarose or polyacrylamide gel electrophoresis prior to use in the footprinting reaction, and although the agarose-purified probes had significantly more background noise due to impurities (as can be seen in the electropherogram of the undigested probe shown above each footprint), a clear footprint which correlated with the positions of the conserved $GTTn_3AAC$ motifs within the probe sequences was still discernible. Not surprisingly, a SurR footprint was detected on each probe tested, and the resulting footprint diagrams for one strand of each probe are shown in Figures 4.12 through 4.17. One footprint was visible on each probe; however, the sizes and positions of each varied greatly. Upon close inspection of these data, the footprint size together with sequence analysis and the EMSA data provide a picture of the SurR DNA-binding pattern in relation to sequence elements, and the footprint positions give good indications of the likely mode of regulation imposed by SurR on transcription of each ORF.

4.6.2 SurR footprint lengths indicate the DNA-binding pattern of SurR

The lengths of the SurR footprint on each probe correspond rather nicely with the number of protein-DNA complexes observed in EMSA for the same probe: essentially, the larger the footprint, the more protein-DNA complexes observed. The locations of the conserved $GTTn_3$-AAC motif varied among the footprints, and this recognition sequence alone does not account for the size variation of the footprints, ranging from 32 to 95 bases. The sequences within the footprints were therefore analyzed for degenerate repetitive sequences which would explain both the appearance of multiple protein-DNA complexes in EMSA and the variable footprint sizes. The spacing of the extended motif, three sequence blocks of 3 bp separated by 3 and 5-bp non-conserved blocks ($3\text{-}_3\text{-}3\text{-}_5\text{-}3$), was used as a guide to look for degenerate sequences comparable

to the $GTTn_3AACn_5GTT$ motif. What was discovered in each footprint, for the most part, was not degenerate $GTTn_3AAC$ SELEX motifs repeating at 5-bp intervals, but $AACn_5GTT$ motifs repeating at 3-bp intervals (see annotated footprint sequences in Figs 4.12 to 4.17). This was also the case for the footprint regions upstream of *mbh1* and *pdo* (Fig. 4.18). This seems to indicate that the actual palindromic recognition sequence for SurR is $AACn_5GTT$; however, perhaps only one of these palindromes will not suffice for high-affinity binding of SurR, and this is why they occur in pairs or degenerate series at 3-bp intervals. Moreover, the number of protein-DNA complexes observed in EMSA for each probe corresponds closely to the number of $AACn_5GTT$ and degenerate repeats within each footprint, as summarized in Table 4.3.

Table 4.3 Number of SurR-DNA complexes in EMSA corresponds to footprint length

Probe name	ORF/promoter DNA in probe[a]	Number of SurR-DNA complexes in EMSA	Footprint length(s)[b]	Number of 3-$_5$-3 motifs in footprint[c]	Total length of motifs[d]
PF0094	PF0094, PF0095	5	83	5	70
PF1186	PF1186	2	53	3	42
PF1453	PF1453, PF1454	6	95	6	84
PF2051	PF2050, PF2051	3	54	3	42
PF1423	PF1422, PF1423	4-5	53, 39	5	70
PF0891	PF0891	3	42	3?	42
PF0531	PF0531, PF0532	2	32	2	28
PF0559	PF0559	4-5	73	4?	56

[a] Some probes include DNA regions with closely-spaced divergently transcribed ORFs; therefore, these probes will contain putative promoter DNA of more than one ORF.
[b] Footprint lengths are derived from the SurR footprint on one strand, and this is a good estimate of overall footprint length since for those probes where the footprint on both strands has been determined, the have footprints of similar lengths on each strand.
[c] The 3-$_5$-3 motif corresponds to motifs matching and degenerate from the $AACn_5GTT$ consensus.
[d] Motifs repeat at 14-bp intervals, therefore the number of motifs multiplied by 14 gives the total length.

The $AACn_5GTT$ motif makes more physiological sense since the spacing between each half-site is 9 bp, center-to-center, which is almost one full turn (10 bp) in B-DNA. The 3-bp

spacing in between each repeating motif indicates a 14-bp spacing, center-to-center for each repeating motif. This means that the protein does not bind on only one side of the DNA, but perhaps binds to adjacent sites in a spiraling manner along the DNA axis. A new search of the UOR database of 200-bp UORs using this AACn$_5$GTT palindrome returns a total of 132 hits with 110 non-redundant motifs (see Appendix Table A3). This is 16 fewer motifs than are found with the search for GTTn$_3$AAC (see Appendix Table A2); however, only 27 UORs contain both GTTn$_3$AAC and AACn$_5$GTT motifs. These 27 UORs include the 12 occurences of the extended SELEX motif upstream of ORFs/operons significantly regulated within 10 min of S^0 addition, as well as the PF1453 UOR which contained two separate GTTn$_3$AAC motifs (see Table 4.2). Taken together, these data suggest that the most likely true SurR binding sites are those that have the extended motif, more than one of the same motif, or a combination of both GTTn$_3$AAC and AACn$_5$GTT motifs, and these UORs are listed in Table 4.4. Notably, only one UOR from the 10-min S^0-regulated ORFs is not included in Table 4.4, and that is PF2051 which contains only one GTTn$_3$AAC motif.

Table 4.4 UORs which contain multiple GTTn$_3$AAC and/or AACn$_5$GTT motifs[1]

UOR Name[a]	Motif[b]	UOR start[c]	UOR stop[c]	Genome start	Genome stop	Annotation[d]	Operon Prediction[e]
PF0094	AACCTAAGGTT	-82	-72	103271	103281	Protein disulfide oxidoreductase, PDO [105]	**PF0094**
	AACTCTAGGTT	-52	-42	103241	103251		
	GTTATAAAC	-30	-22	103221	103229		
	GTTGACAAC	-74	-66	103265	103273		
PF0095	AACCTAGAGTT	-91	-81	103251	103241	Sulfur response regulator, SurR [this work]	**PF0095**
	AACCTTAGGTT	-61	-51	103281	103271		
	GTTGTCAAC	-67	-59	103273	103265		
	GTTTATAAC	-111	-103	103229	103221		
PF0443	AACCTCAAGTT	-200	-190	461003	461013	*putative membrane transport*	PF0446 -

[1] The *P. furiosus* UOR database and corresponding motif searching software was created by Darin Cowart (University of Georgia).

UOR Name[a]	Motif[b]	UOR start[c]	UOR stop[c]	Genome start	Genome stop	Annotation[d]	Operon Prediction[e]
	GTTGGGAAC	-69	-61	460874	460882	*protein*	**PF0443**
PF0531	AACAAAATGTT	-78	-68	547735	547745	*cobalamin biosynthesis protein m*	**PF0531** - PF0528
	AACCTTACGTT	-92	-82	547749	547759		
	GTTGGTAAC	-84	-76	547743	547751		
PF0532	AACATTTTGTT	-108	-98	547745	547735	*hypothetical protein*	**PF0532** - PF0534
	AACGTAAGGTT	-94	-84	547759	547749		
	GTTACCAAC	-100	-92	547751	547743		
PF0547	AACTTGAAGTT	-190	-180	566581	566591	*hypothetical protein*	**PF0547**
	GTTTTCAAC	-182	-174	566575	566583		
PF0548	AACTTCAAGTT	-61	-51	566591	566581	*hydrogenase expression / formation protein*	**PF0548** - PF0549
	GTTGAAAAC	-67	-59	566583	566575		
PF0559	AACCAAAGGTT	-74	-64	575529	575519	*hydrogenase expression/formation regulatory protein*	**PF0559**
	AACTTTTGGTT	-42	-32	575561	575551		
	GTTTCTAAC	-48	-40	575553	575545		
PF0568	AACGCTATGTT	-18	-8	590839	590849	*hypothetical protein*	**PF0568**
	GTTTATAAC	-24	-16	590847	590855		
PF0569	AACATAGCGTT	-34	-24	590849	590839	*hypothetical protein*	**PF0569**
	GTTATAAAC	-26	-18	590855	590847		
PF0576	GTTTCTAAC	-134	-126	596314	596322	*hypothetical protein*	**PF0576** - PF0575
	GTTTTCAAC	-173	-165	596353	596361		
PF0632	AACGGTTGGTT	-128	-118	645533	645543	*hypothetical protein*	PF0633 - PF0630
	AACTTAGCGTT	-25	-15	645430	645440		
PF0799	GTTCATAAC	-195	-187	785146	785154	*hypothetical protein*	**PF0799** - PF0794
	GTTGACAAC	-65	-57	785016	785024		
PF0891	AACCTTTGGTT	-136	-126	863628	863618	Hydrogenase I, beta subunit, HydB1 [102]	**PF0891** - PF0894
	GTTTTTAAC	-142	-134	863620	863612		
PF0913	AACCAAAGGTT	-191	-181	884566	884576	*Formylmethanofuran dehydrogenase, subunit E*	PF0914 - PF0909
	GTTAATAAC	-183	-175	884560	884568		
	GTTCAAAAC	-197	-189	884574	884582		
PF0914	AACCAAAGGTT	-75	-65	884566	884576	*hypothetical protein*	**PF0914** - PF0909
	GTTAATAAC	-67	-59	884560	884568		
	GTTCAAAAC	-81	-73	884574	884582		
PF0915	AACATCTAGTT	-66	-56	885392	885402	*Cytochrome c biogenesis protein*	**PF0915**
	AACCTTAGGTT	-52	-42	885378	885388		
	GTTAGAAAC	-58	-50	885386	885394		
PF0926	AACCTATTGTT	-75	-65	892090	892100	*hypothetical protein*	**PF0926**
	AACTTATTGTT	-89	-79	892104	892114		
	GTTTGGAAC	-81	-73	892098	892106		
PF1063	GTTATAAAC	-82	-74	1016376	1016368	*hypothetical protein*	PF1062 - PF1064
	GTTTTTAAC	-30	-22	1016428	1016420		
PF1100	AACTCGTAGTT	-102	-92	1047039	1047029	*hypothetical protein*	**PF1100**
	GTTTGCAAC	-108	-100	1047031	1047023		

UOR Name[a]	Motif[b]	UOR start[c]	UOR stop[c]	Genome start	Genome stop	Annotation[d]	Operon Prediction[e]
PF1186	AACTAAAGGTT	-34	-24	1132903	1132913	NAD(P)H sulfur reductase, NSR [100]	**PF1186**
	GTTTTAAAC	-26	-18	1132897	1132905		
PF1328	AACCTTAGGTT	-87	-77	1249939	1249949	Hydrogenase II, gamma subunit [103]	**PF1328** - PF1326
	GTTTTGAAC	-79	-71	1249933	1249941		
PF1329	AACCTAAGGTT	-82	-72	1249949	1249939	*H-II hydrogenase subunit beta*	**PF1329** - PF1332
	GTTCAAAAC	-88	-80	1249941	1249933		
PF1412	GTTCCAAAC	-199	-191	1327897	1327889	*dipeptide transport ATP-binding protein*	PF1409 - PF1413
	GTTCCAAAC	-148	-140	1327948	1327940		
PF1422	AACATTTGGTT	-134	-124	1337388	1337398	*thioredoxin reductase*	**PF1422**
	AACTTTTGGTT	-19	-9	1337273	1337283		
	GTTTATAAC	-11	-3	1337267	1337275		
	GTTTGAAAC	-126	-118	1337382	1337390		
PF1423	AACCAAAAGTT	-140	-130	1337283	1337273	Membrane Bound Hydrogenase, subunit A, MbhA [101]	**PF1423** - PF1436
	AACCAAATGTT	-25	-15	1337398	1337388		
	GTTATAAAC	-146	-138	1337275	1337267		
	GTTTCAAAC	-31	-23	1337390	1337382		
PF1453	GTTGTTAAC	-61	-53	1360093	1360101	Membrane bound oxidoreductase, subunit A, MbxA [122]	**PF1453** - PF1441
	GTTTAAAAC	-103	-95	1360135	1360143		
PF1516	AACTTTATGTT	-52	-42	1415014	1415024	*GMP synthase subunit B*	**PF1516** - PF1515
	GTTTACAAC	-58	-50	1415022	1415030		
PF1621	AACTTTTAGTT	-78	-68	1513974	1513984	*hypothetical protein*	**PF1621**
	GTTACTAAC	-70	-62	1513968	1513976		
PF1622	AACTAAAAGTT	-190	-180	1513984	1513974	*n-type ATP pyrophosphatase superfamily*	**PF1622**
	GTTAGTAAC	-196	-188	1513976	1513968		
PF1684	AACATAATGTT	-73	-63	1565645	1565635	*acetylglutamate kinase*	PF1681 - PF1686
	GTTCAAAAC	-85	-77	1565631	1565623		
PF1911	AACCTTTGGTT	-90	-80	1762931	1762941	*hypothetical protein*	**PF1911** - PF1910
	GTTCACAAC	-54	-46	1762897	1762905		
PF1912	AACCAAAGGTT	-60	-50	1762941	1762931	*hypothetical protein*	**PF1912** - PF1913
	GTTGTGAAC	-94	-86	1762905	1762897		

[a] UOR (Upstream of ORF Region) designation corresponds to the locus of the ORF from which the upstream sequence was taken. UOR names in bold indicate those which have verified SurR footprints.
[b] Motifs listed are not non-redundant. Separate UORs can have the identical motif if they are from divergently oriented gene pairs.
[c] Start and stop positions are relative to the UOR sequence where -1 corresponds to the first nucleotide upstream from the ORF start.
[d] Annotations in italics are from REFSEQ or TIGR, and characterized ORFs/proteins are listed with the corresponding reference.
[e] Operon predictions are derived from [121]. The ORFs are listed in the direction of the operon, with the UOR ORF listed in bold when applicable.

4.6.3 Hydroxyl radical footprinting reveals SurR binding pattern

Hydroxyl radical (OH•) footprinting was used to probe the specific contacts that SurR made with its cognate DNA. OH• is a much smaller DNA cleavage agent and can therefore yield higher-resolution footprints which can be interpreted in light of DNase I footprinting data. OH• footprints were obtained for four probes: PF1423, PF0094, PF0891, and PF1186 (containing *mbh1*, *pdo*, *hyd1B*, and *nsr* promoter-ORF DNA, respectively) (Fig. 4.19). The OH• footprinting results revealed that SurR binds to the central regions of the corresponding DNase I footprints in a 6-$_8$-6 pattern, where two clusters of 6 bases are significantly protected from cleavage and are separated by 8 bases. The fringes of the footprints show smaller 3-base regions of protection, and these regions are separated from the central 6-base regions by a spacing of 7 or 8 bases. When these results are compared with the DNase I footprints for the same probes, it becomes evident that the internal DNase I hypersensitive sites in the outer edges of the DNase I footprint correspond to the 7 to 8-base spacing in between OH•-protected blocks of 3 or 6 bases. This result indicates that a distortion of the DNA takes place at the edges of the footprint regions, making DNase I cleavage at these regions more favorable.

The OH• footprints for both strands of the PF1186 probe (containing *nsr* ORF/promoter DNA) are shown in Figure 4.20, with corresponding SurR sequence contacts indicated. An A+G sequence ladder was used to locate the DNA sequences protected from OH• cleavage on the 6FAM strand, and the HEX strand was aligned accordingly using the ROX internal size standard. Clear regions of protection can be seen in a pattern which is repeated on both strands: protected blocks of 3, 6, 6, and 3 bases each separated by 8 bases. Interestingly, the 6-base blocks contain the 'GTT' of the GTTn$_3$AAC motif at their 3' ends. This is apparently the major sequence element which is recognized by SurR since the complementary 'AAC' bases are not protected.

As was expected given the 14-bp spacing, center-to-center, of adjacent motifs, the binding pattern elucidated by OH• footprinting seems to wind along the DNA axis, suggesting that SurR binds to adjacent DNA recognition sites in a spiraling manner along the DNA (Fig. 4.20 B).

4.6.4 SurR footprint positions hint at the mode of transcriptional regulation of SurR

For probes containing the S^0 down-regulated ORFs and their putative promoter DNA, the footprints were, on average, positioned farther from the ORF start. The SurR binding site on the PF0531 probe was located at -65 from the ORF start and spanned 32 bases (Fig. 4.12). The PF0891 probe footprint spanned 42 bases starting far upstream from the ORF at position -107 (Fig. 4.13), and that for the PF0559 probe footprint spanned 73 bases starting at -26 relative to the ORF start (Fig. 4.14). Data presented earlier demonstrated that the *mbh1* (PF1423) upstream DNA had two SurR binding sites, with the strongest affinity *mbh1*-distal site starting at -107 from the ORF start and extending over 53 bases; the *mbh1*-proximal site covered 39 bases beginning at -5 from the ORF start (Fig. 4.2). Footprints for probes containing the S^0 up-regulated ORFs together with their putative promoter regions were positioned, for the most part, very close to the ORF start. The SurR binding site on the PF1453 probe was the largest, spanning 95 bases, beginning at base -40 relative to the PF1453 start site (Fig. 4.15). The PF2051 probe footprint began at +1 of the ORF start and extended 54 bases upstream from it (Fig. 4.16) while the PF1186 probe footprint spanned 53 bases starting at position -5 from the ORF start (Fig. 4.17). Recall that the *pdo-surr* (PF0094 probe) footprint was similarly positioned very close to the ORF start, extending from -4 to -86 and spanning a total of 83 bases (Fig. 4.5).

As was discussed in Chapter 1, the position of an archaeal transcription factor's binding site is important in dictating its role in transcriptional regulation: the binding sites of repressors tend to overlap either the BRE/TATA elements or the transcription start site while activators tend

to bind upstream from the BRE/TATA elements (see Fig 1.4). Therefore, once a footprint position has been determined, its location relative to the core promoter elements will give a clue as to its mechanism of regulation for the corresponding ORF/operon. Knowing the position of the transcription start site allows the location of the BRE/TATA elements, since the spacing between these core promoter elements is critical for productive recruitment of RNAP to the transcription start site: the TATA box is centered at approximately -25 from the transcription start site, with the BRE located directly upstream [9]. The *in vitro* transcription start sites for PF1423 and PF0891, as specified in Figures 4.18 and 4.13, were determined by Annette Keese (personal communication, unpublished data) through collaboration with Michael Thomm (University of Regensburg, Germany). This allowed the exact location of the BRE/TATA elements, and for both of these ORFs, the SurR footprints covered none of the core promoter elements. Moreover, the PF0891 footprint was positioned upstream from the BRE/TATA box and minimally overlapped only two bases on the opposite strand (Fig 4.13 C). The high-affinity *mbh1*-distal footprint also was located upstream from the BRE/TATA box; although the *mbh1*-proximal site is positioned directly upstream from the *mbh1* ORF, it does not interfere with any of the core promoter elements (Fig. 4.18 A). The binding site pattern on the *mbh1* promoter is reminiscent of that of the well-characterized archaeal activator Ptr2 of *M. jannaschii* which activates transcription *in vitro* at the *rb2* promoter from binding sites located both upstream and downstream of the TATA box, even though the downstream site overlaps the transcription start site [67]. Together, these data implicate SurR as a transcriptional activator for both *mbh1* (PF1423) and *hydB1* (PF0891).

Given that the BRE/TATA elements have some sequence conservation (see Table 1.1), their positions upstream of a given ORF can sometimes be deduced based on sequence analysis.

A recent analysis of characterized glycolytic gene promoters of *Pyrococcus* species determined the BRE/TATA consensus sequences to be VRAAANN/TTWWWAW[1] [20]. Therefore, sequences similar to this consensus were in the ORF upstream regions of each fooprinted probe. The probable positions of the BRE/TATA boxes are indicated in part C of Figures 4.12 through 4.17. For the remaining S^0 down-regulated ORFs, PF0531 and PF0559, possible positions for their BRE/TATA boxes can be located just downstream of the SurR footprints. The BRE/TATA signal for PF0531 is strong and the position is relatively certain; however, a slight movement of the BRE/TATA upstream is also possible. The positioning of the PF0559 BRE/TATA box is less certain, and two possible positions are indicated in Figure 4.13 C. The PF0559-proximal BRE/TATA box is positioned too close to the translation start site to allow for adequate transcription initiation upstream from the ORF; however, a second in-frame start codon (ATG) is positioned only 18 bases downstream of the TIGR-predicted PF0559 ORF start, and this is the PF0559 start site predicted by NCBI. This alternate start site would accommodate the BRE/TATA box located just downstream of the SurR footprint (Fig. 4.14 C). Interestingly, when MEME was used to search for a common motif on only the coding strand of $GTTn_3AAC$-containing 200-bp UORS of all the 10-min S^0 down-regulated ORFs, a BRE/TATA-like conserved sequence was evident at the 3' end of the common motif, implying that the SurR binding site may be located upstream from the BRE/TATA for all of these ORFs (Fig. 4.21).

For the S^0 up-regulated ORFs, the putative BRE/TATA boxes appear to be positioned either within or upstream from the SurR footprints. The most likely position for the BRE/TATA box of the PF1453 ORF is located well inside the SurR footprint (Fig. 4.15 C). There is a BRE/TATA-like sequence just downstream from the footprint; however, it is positioned too

[1] Ambiguous nucleotides are represented according to the IUPAC code as follows: V = A/C/G, R = A/G, N = any base, and W = T/A.

close to the ORF start, and no alternate start codons are found within the first 100 bases of the PF1453 ORF. For the PF2051 ORF, there are two possible locations for the BRE/TATA box, as shown in Figure 4.15 C: one is situated well inside the footprint while the other is centered 22 bases upstream. Both of these configurations suggest that SurR would repress transcription from PF2051 by either blocking the binding of TBP and TFB to the TATA and BRE elements or by obstructing recruitment of RNAP to the transcription initiation site. It should be noted that NCBI predicts the PF2051 start site to begin 54 bases downstream of the TIGR-predicted start site; however, the only possible BRE/TATA-like sequence that can be found in this region is positioned slightly too close to the alternate start site. For PF1186 (*nsr*) and PF0094 (*pdo*), the putative BRE/TATA elements also lay either within or just upstream from the SurR footprint (Figs. 4.17 C and 4.18 B). Table 4.5 summarizes the sequence information for the known and predicted BRE/TATA boxes for ORFs containing upstream SurR footprints.

Table 4.5 Predicted and known BRE/TATA boxes for SurR target ORFs

ORF	BRE[a]	TATA Box[a]
Consensus [20]		
	VRAAANN	TTWWWAW
Predicted[b]		
PF0531	AAAAATA	TTTAAAA
PF0559	AGAAACC	TATAAAG
PF1454 (*mbx1*)	GCAAACA	TTTATAA
PF2051	ATAAATT	TTTTATA
	AAAAAGC	TTTTTCA
PF1186 (*nsr*)	CTAAAAT	TTTTTAA
	CTAAATG	TTTGAAT
PF0094 (*pdo*)	ATTAAAG	TTATAAA
	GCAAAGT	TTATATA
Known[c]		
PF0891 (*hydB1*)	AAAAGGC	TTTTTTA
PF1423 (*mbh1*)	AAAAACC	TTTTTAA

[a] Ambiguous nucleotides are represented according to the IUPAC code as follows: W = T/A, Y = C/T, R = A/G, V = A/C/G, N = any base. Nucleotides which deviate from the proposed consensus are shown in red.

[b] Some ORFs have two potential BRE/TATA sequences, and both are listed.
[c] The positions of the known BRE/TATA sequences were derived from the transcription start sites for the corresponding ORFs which were determined by Annette Keese (personal communication, unpublished data).

4.7 SurR is a transcriptional activator and repressor

In order to functionally validate SurR as a bona fide regulatory transcription factor, its direct effect on basal transcription needed to be established. There is no genetic system for *P. furiosus* as of yet, so *in vivo* functional validation was not possible; however, an *in vitro* transcription system for this organism has been well-established by Michael Thomm's research group [110, 111]. Through collaboration with Michael Thomm (University of Regensburg, Germany), the *in vitro* transcriptional regulation profile of SurR was determined by Annette Keese for three ORFs: *mbh1* (PF1423) and *hydB1* (PF0891) which are down-regulated with S^0 according to DNA microarray and *pdo* (PF0094) which is up-regulated with S^0. ORFs both up and down-regulated with S^0 were selected for *in vitro* transcription experiments to determine if SurR could function as both an activator and repressor, as was suggested by the DNase I footprinting data. The same pUC18-cloned probe DNAs used for EMSA and footprinting were provided to Annette for use as *in vitro* transcription templates, along with recombinant his-tagged and untagged SurR protein.

Basal transcription from the PF1423 probe produced a transcript only from *mbh1* (PF1423) and not from the divergent ORF, PF1422. Addition of increasing concentrations of SurR showed no change in transcription from the basal level; however, *mbh1* appears to have a relatively strong promoter *in vitro* which may mask any regulation conferred by SurR (Fig. 4.22 A). Transcription from the PF0891 probe, however, showed a clear increase in the PF0891 (*hydB1*) transcript with increasing SurR concentration (Fig. 4.22 B). The basal level of

transcription for this ORF appeared to be much lower, compared to *mbh1* and the *gdh* control. In both cases, SurR did not affect the basal transcription level from the *gdh* promoter, indicating that SurR is a specific transcriptional activator for *hydB1*, and presumably the entire four-ORF operon that encodes the soluble hydrogenase I. Since the SurR binding site lies directly upstream from the BRE/TATA box of the *hydB1* ORF, it is reasonable to suggest that SurR may activate transcription by facilitating recruitment of TBP, perhaps similar to the activation mechanism determined for Ptr2 [63]. It is somewhat surprising that no regulation was apparent for *mbh1*; however, the high promoter strength which is observed *in vitro* could be masking any regulation which SurR might confer on transcription from *mbh1*. The position of the SurR footprints upstream from *mbh1* do not overlap with any of the core promoter elements, so transcriptional repression of *mbh1* by SurR would not be expected, unless it proceeds through some unknown mechanism. Therefore, it is likely that SurR is also a transcriptional activator for *mbh1* given the results for *hydB1* since these ORFs both contain SurR binding sites in their promoter regions and are both among the ORFs down-regulated by S^0.

The PF0094 probe contained divergent ORFs *surr* (PF0095) and *pdo* (PF0094), and transcription with this template produced transcripts from both ORFs, although the *pdo* transcript was much more abundant. Addition of SurR completely repressed transcription from both ORFs, while transcription of the *gdh* template was unaffected (Fig. 4.23). Repression of *pdo* transcription may occur through SurR blocking the binding of TBP and TFB or recruitment of RNAP. It is not clear where the *surr* BRE/TATA is located in the *pdo-surr* intergenic region, and perhaps the sequence is somewhat deviant considering that the transcription signal for *pdo* was much higher than *surr*. The SurR footprint covers 83 bases of the 123-bp *pdo-surr* intergenic

space, so it is possible that it overlaps with the core promoter elements of its own gene as well as those for *pdo*.

Taken together, the *in vitro* transcription results provided by Annette Keese show that SurR can both activate and repress transcription, depending on the gene context. This is the first direct evidence of an archaeal transcriptional regulator that has dual regulatory functions. The somewhat paradoxical implication from the footprinting results that SurR may activate S^0 down-regulated ORFs and repress S^0 up-regulated ORFs was corroborated by the *in vitro* transcription results showing activation of *hydB1* and repression of *pdo*.

4.8 Validation of a SurR binding site in the promoter of a non S^0-regulated ORF

In order to validate at least one of the UOR database search results[1] that did not include one of the S^0-regulated ORFs (see Table 4.4), the PF1911 ORF upstream DNA was chosen for EMSA and footprinting studies since it contained both an $AACn_5GTT$ motif centered at 85 bp and a $GTTn_3AAC$ motif centered at 50 bp upstream of the ORF start. PF1911 is annotated as a putative hydrogenase and has 43% sequence identity with PF1328, a putative hydrogenase gamma subunit. Incidently, PF1328 is divergently transcribed from PF1329, the start of the operon encoding the soluble Hydrogenase I [102], which is down-regulated in the presence of S^0 and contains a putative SurR binding site in its upstream DNA. BLAST results with the PF1911 sequence reveals homology with proteins annotated as 'hydrogenase gamma subunits' as well as 'FAD/NAD(P)-binding oxidoreductases'. A conserved domain, 'ferredoxin-NADP reductase subunit alpha,' extends almost completely across the protein sequence (E-value of $5x10^{-109}$). The role of PF1911 protein product is not clear; however, if it is a hydrogenase subunit, it should be

[1] The *P. furiosus* UOR database and corresponding motif searching software was created by Darin Cowart (University of Georgia).

noted that it most likely does not contain a catalytic center. Regardless, its upstream DNA was considered a likely candidate to contain a genuine SurR binding site due to the known relationship of SurR to the membrane-bound hydrogenase and hydrogenase I operon promoters. Incidentally, PF1911 is one of another pair of closely-spaced divergently transcribed ORFs: PF1912 is separated from PF1911 by only 139 bp. Therefore, PF1912 is also present in the UOR search results for the SELEX-derived motifs. BLAST and conserved domain searches of PF1912 indicate a strong relationship to 2-methylthioadenine synthetase (MiaB), a tRNA-modification enzyme.

EMSA results confirmed that SurR does indeed bind to the PF1911 putative promoter region, completely shifting the DNA probe at a protein/DNA mole ratio of 40, and forming four to five distinct protein-DNA complexes in the process (Fig. 4.24). DNase I footprinting confirmed that the binding site overlapped both the conserved $AACn_5GTT$ and $GTTn_3AAC$ motifs and extended over approximately 65 bases, from -33 to -97 from the PF1911 start, almost centered between the divergent PF1911 and PF1912 ORFs (Fig. 4.24). The footprint contains four $3\text{-}_5\text{-}3$ motifs (degenerate from $AACn_5GTT$), corresponding relatively well with the four to five protein-DNA complexes observed in EMSA. The putative BRE/TATA elements lie within the SurR footprint, indicating that SurR may function to repress transcription from PF1911.

PF1911 is not significantly regulated by S^0 according to the DNA microarray data; however, there is a possibility that any up or down-regulation of the ORF was not detected in DNA microarray experiments, and this would have to be verified by quantitative PCR. If PF1911 is not regulated by S^0, then demonstration of the binding of SurR to the upstream DNA of this ORF implies that SurR may not be a transcriptional regulator exclusive to mediating events in *P. furiosus* S^0-response. Indeed, the transcriptional regulation of genes by SurR may be extensive

considering the number of UOR database hits for the conserved sequences $GTTn_3AAC$ and $AACn_5GTT$ which are not related to S^0-regulated ORFs as defined by DNA microarray experiments. More of these UOR database hits unrelated to S^0 response will need to be verified in order to confirm this hypothesis.

Figure 4.1 Fluorescence-detected DNase I footprinting scheme. A protein-bound fluorescently-tagged probe is nicked with DNase I, and the region of DNA bound by the protein is protected from cleavage. Strands are denatured prior to undergoing capillary electrophoresis on a capillary sequencer (ABI 3730x1, Applied Biosystems). Fluorescently-labeled fragments are detected, and resulting electropherograms of samples with and without added protein are overlaid to determine footprint position.

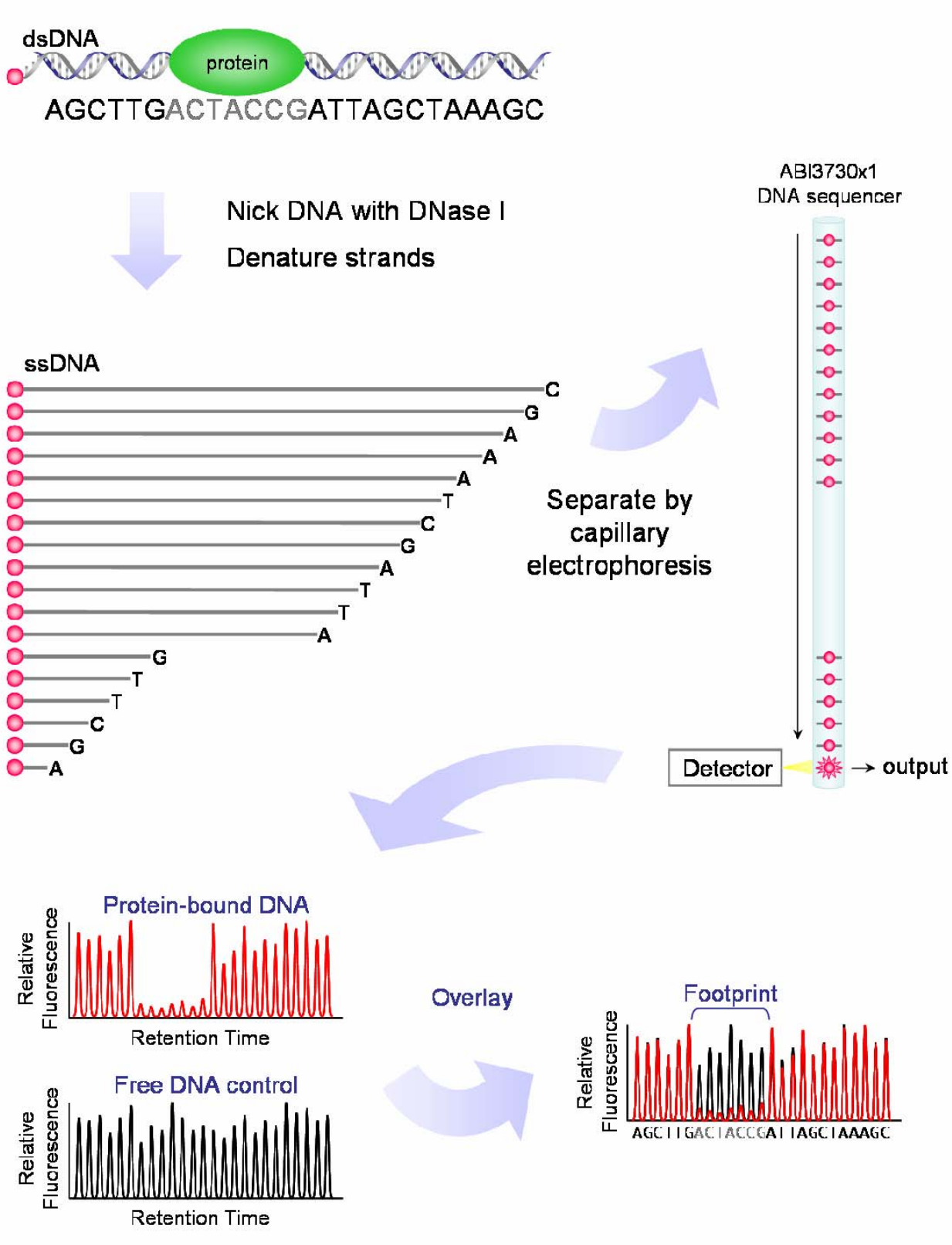
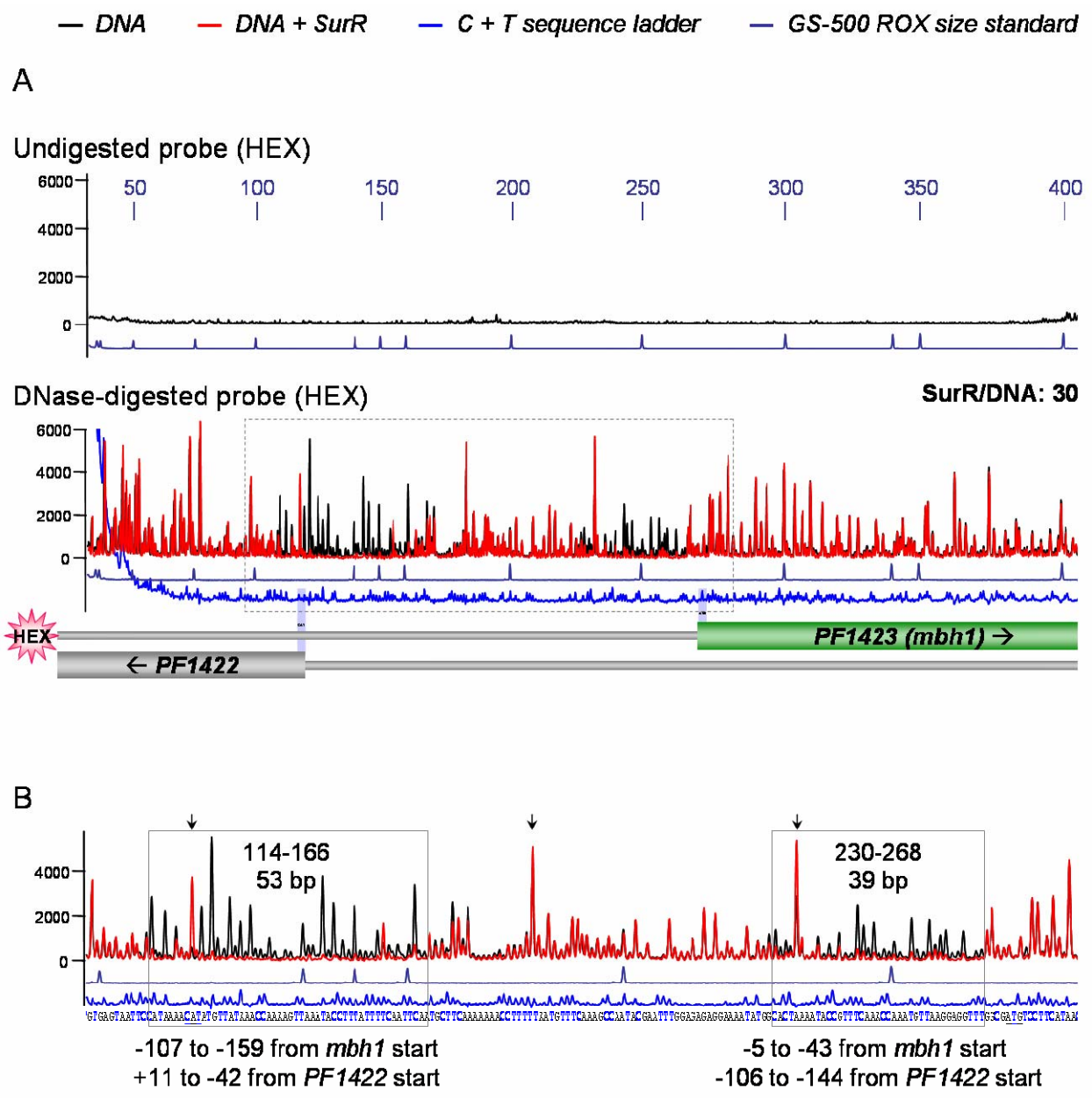

dsDNA
protein
AGCTTGACTACCGATTAGCTAAAGC
Nick DNA with DNase I
Denature strands
ssDNA
C
G
A
A
A
T
C
G
A
T
T
A
G
T
T
C
G
A
ABI3730x1
DNA sequencer
Separate by
capillary
electrophoresis
Detector
output
Protein-bound DNA
Relative
Fluorescence
Retention Time
Free DNA control
Relative
Fluorescence
Retention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

Figure 4.2 DNase I footprint of SurR on PF1423 probe. A. Electropherogram traces of HEX-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Internal standards GS-500 ROX (peak sizes in bases are labeled) and C+T probe-specific sequence ladder were used to determine positions of electropherogram peaks relative to probe sequence. Probe diagram indicates positions of ORFs relative to electropherogram traces and the SurR footprint region is boxed. **B.** Enlarged view of footprint region boxed in (A) with corresponding sequence as determined by sequence ladder. SurR footprints are boxed and arrows indicate DNase I hypersensitive sites. **C.** Electropherogram traces of 6FAM-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Internal standards GS-500 ROX (peak sizes in bases are labeled) and A+G probe-specific sequence ladder were used to determine positions of electropherogram peaks relative to probe sequence. Probe diagram indicates positions of ORFs relative to electropherogram traces and the SurR footprint region is boxed. **D.** Enlarged view of footprint region boxed in (C) with corresponding sequence as determined by sequence ladder. SurR footprints are boxed and arrows indicate DNase I hypersensitive sites.


— DNA
— DNA + SurR
— C + T sequence ladder
— GS-500 ROX size standard
A
Undigested probe (HEX)
6000
4000
2000
0
50
100
150
200
250
300
350
400
DNase-digested probe (HEX)
SurR/DNA: 30
HEX
PF1423 (mbh1) →
← PF1422
B
114-166
53 bp
230-268
39 bp
-107 to -159 from mbh1 start
+11 to -42 from PF1422 start
-5 to -43 from mbh1 start
-106 to -144 from PF1422 start

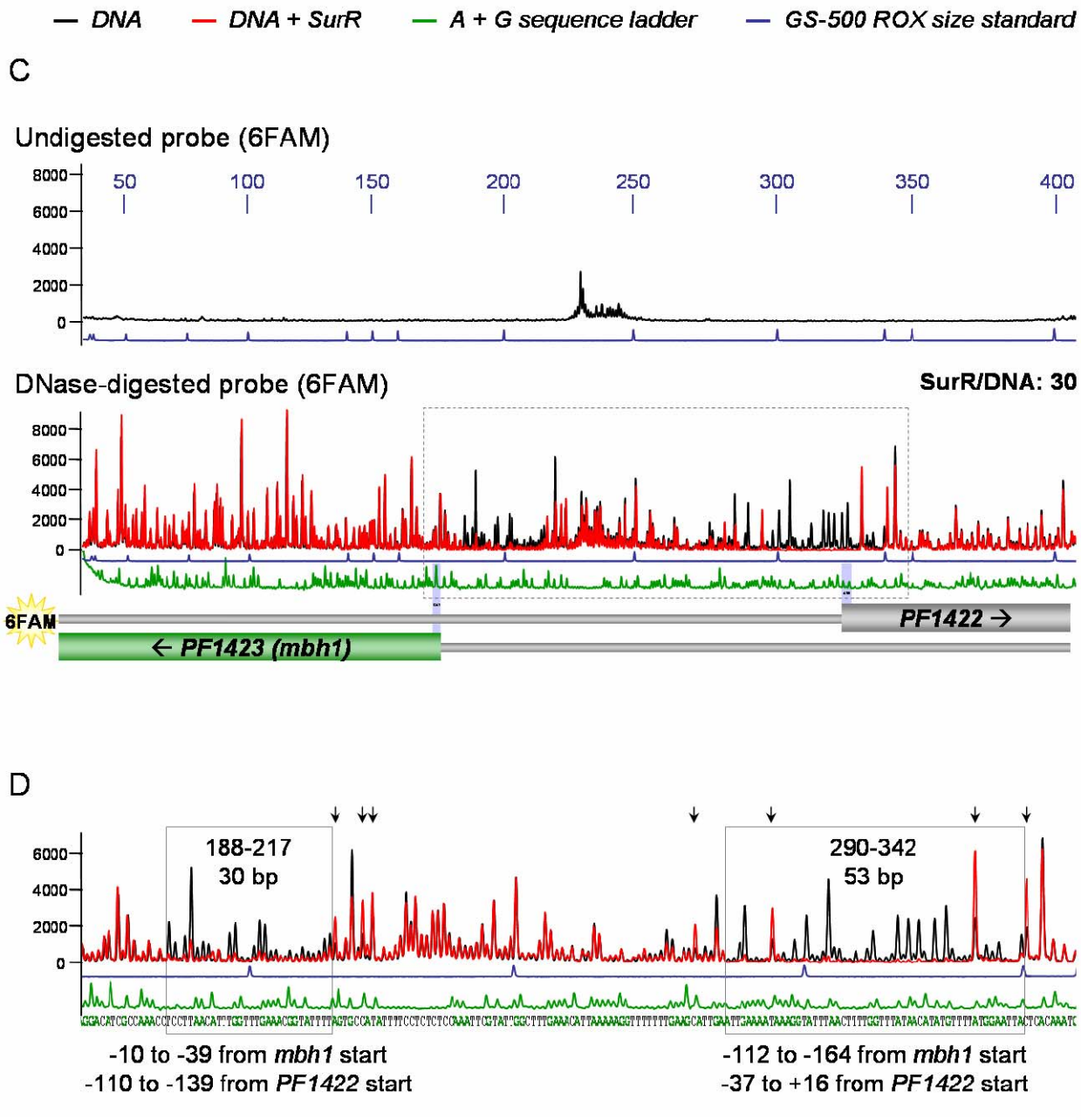

DNA
DNA + SurR
A + G sequence ladder
GS-500 ROX size standard
C
Undigested probe (6FAM)
8000
6000
4000
2000
0
50
100
150
200
250
300
350
400
DNase-digested probe (6FAM)
SurR/DNA: 30
6FAM
PF1422 →
← PF1423 (mbh1)
D
188-217
30 bp
290-342
53 bp
-10 to -39 from mbh1 start
-110 to -139 from PF1422 start
-112 to -164 from mbh1 start
-37 to +16 from PF1422 start

Figure 4.3 SurR shows differential affinity for the binding sites upstream of *mbh1*. Electropherogram traces for HEX-labeled strand of DNase-digested probe with and without protein for protein/DNA mole ratios of 12 (top) and 24 (bottom). Peak sizes of internal GS-500 ROX standard are indicated. Probe diagram indicates positions of ORFs relative to electropherogram, and footprint regions are boxed.

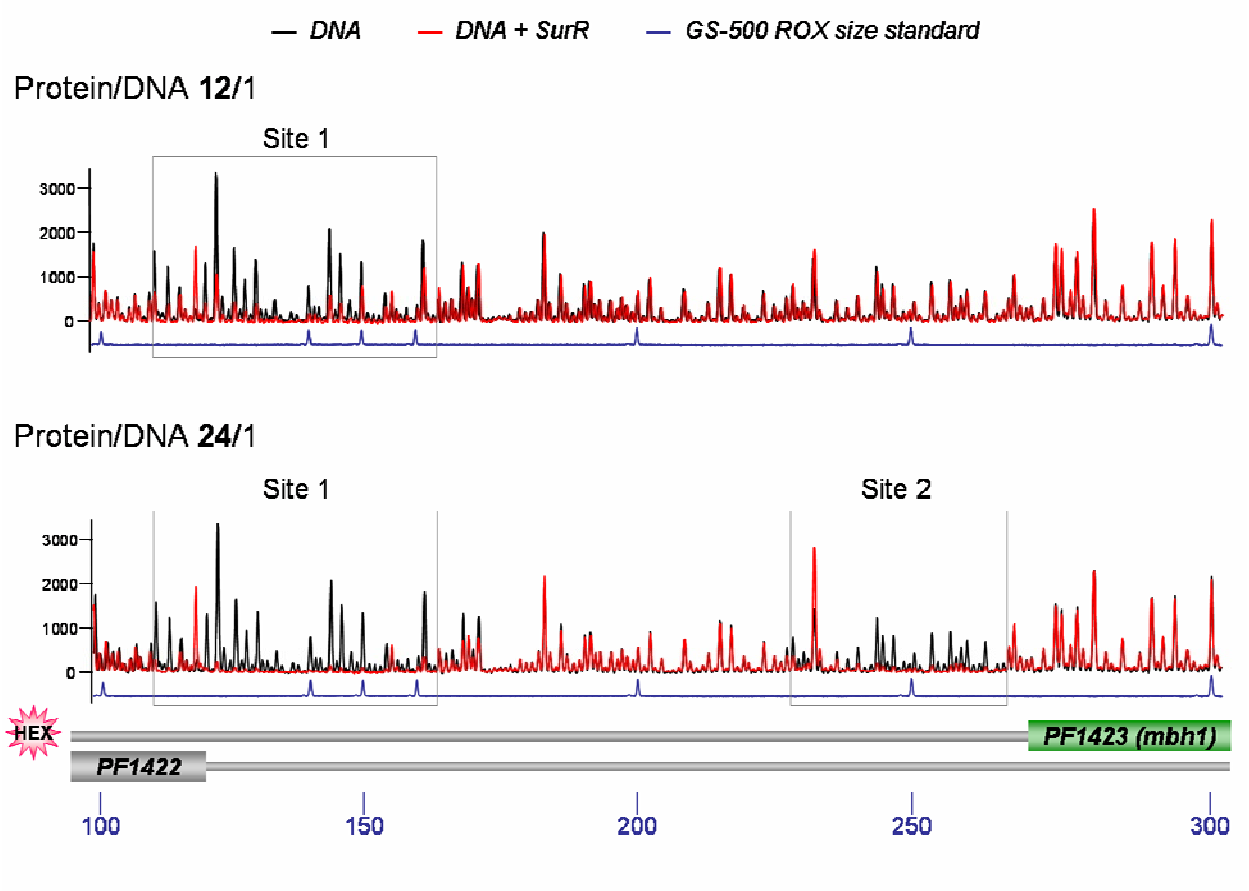
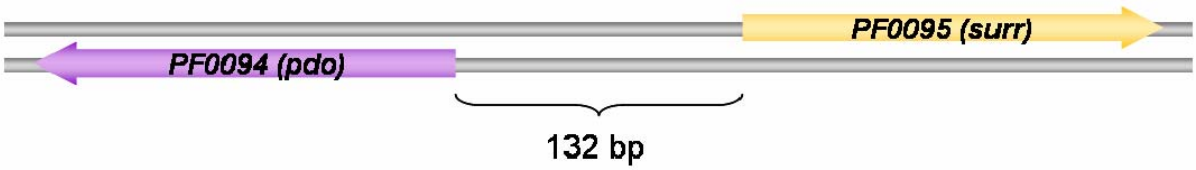

DNA
DNA + SurR
GS-500 ROX size standard
Protein/DNA 12/1
Site 1
3000
2000
1000
0
Protein/DNA 24/1
Site 1
Site 2
3000
2000
1000
0
HEX
PF1423 (mbh1)
PF1422
100
150
200
250
300

Figure 4.4 SurR binds to *pdo-surr* intergenic DNA. A. Diagram of the *pdo-surr* DNA region showing the 123-bp intergenic space. The probe DNA used in EMSA consisted of the intergenic space, plus 3 bp on either end to include each ORF start codon. **B.** The EMSA gel lists corresponding protein/DNA mole ratios above each lane. Samples of DNA (50 nM) with or without protein were assembled in EMSA buffer (20 mM HEPES pH 7.5, 200 mM KCl, 5% glycerol, 1 mM EDTA) and incubated for 20 min at 70°C. Gel was stained with SYBR Green I nucleic acid gel stain.

A


PF0095 (surr)
PF0094 (pdo)
132 bp

B

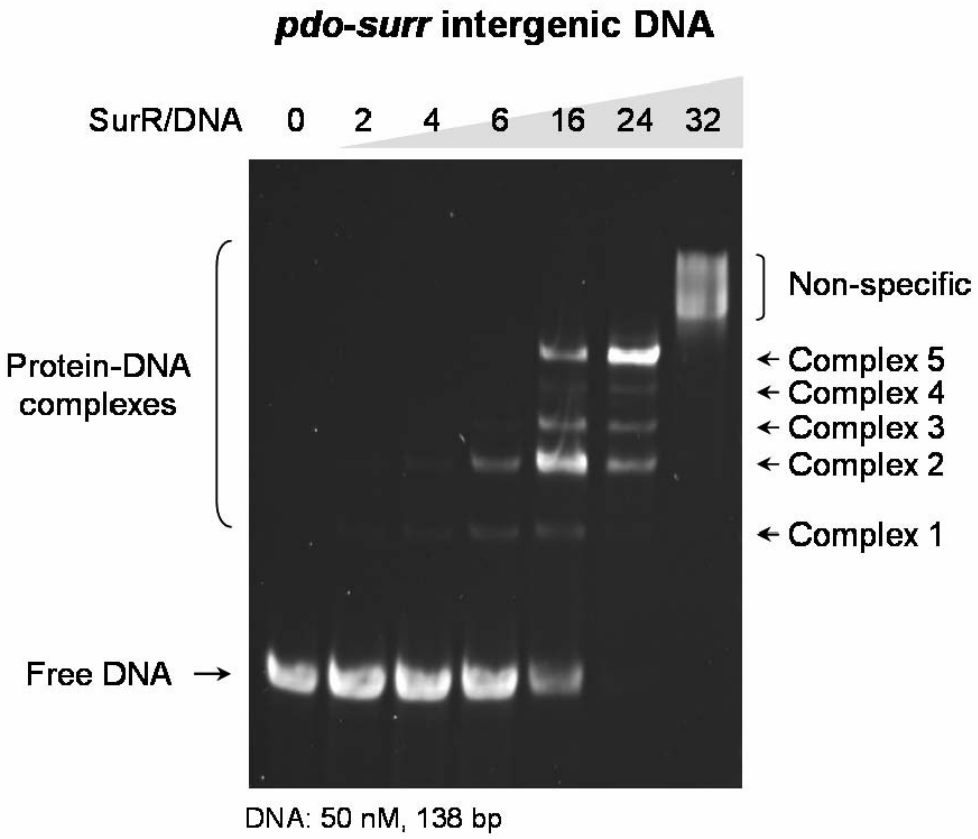
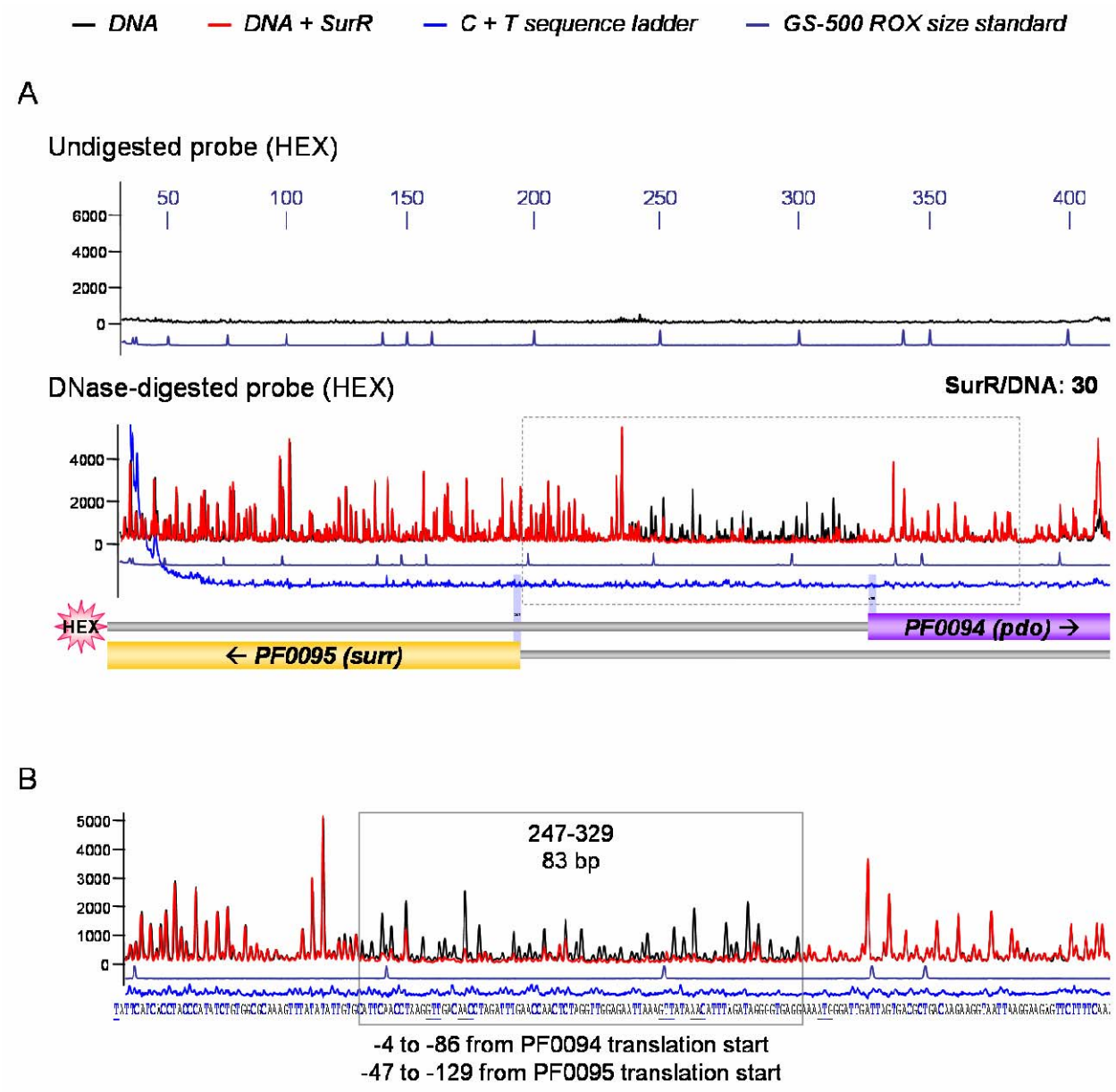

pdo-surr intergenic DNA
SurR/DNA 0 2 4 6 16 24 32
Protein-DNA complexes
Free DNA
Non-specific
Complex 5
Complex 4
Complex 3
Complex 2
Complex 1
DNA: 50 nM, 138 bp

Figure 4.5 DNase I footprint of SurR on PF0094 probe. A. Electropherogram traces of HEX-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Internal standards GS-500 ROX (peak sizes in bases are labeled) and C+T probe-specific sequence ladder were used to determine positions of electropherogram peaks relative to probe sequence. Probe diagram indicates positions of ORFs relative to electropherogram traces and the SurR footprint region is boxed. **B.** Enlarged view of footprint region boxed in (A) with corresponding sequence as determined by sequence ladder. SurR footprints are boxed and arrows indicate DNase I hypersensitive sites. **C.** Electropherogram traces of 6FAM-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Internal standards GS-500 ROX (peak sizes in bases are labeled) and A+G probe-specific sequence ladder were used to determine positions of electropherogram peaks relative to probe sequence. Probe diagram indicates positions of ORFs relative to electropherogram traces and the SurR footprint region is boxed. **D.** Enlarged view of footprint region boxed in (C) with corresponding sequence as determined by sequence ladder. SurR footprints are boxed and arrows indicate DNase I hypersensitive sites.


— DNA
— DNA + SurR
— C + T sequence ladder
— GS-500 ROX size standard
A
Undigested probe (HEX)
50
100
150
200
250
300
350
400
6000
4000
2000
0
DNase-digested probe (HEX)
SurR/DNA: 30
4000
2000
0
HEX
PF0094 (pdo) →
← PF0095 (surr)
B
5000
4000
3000
2000
1000
0
247-329
83 bp
-4 to -86 from PF0094 translation start
-47 to -129 from PF0095 translation start

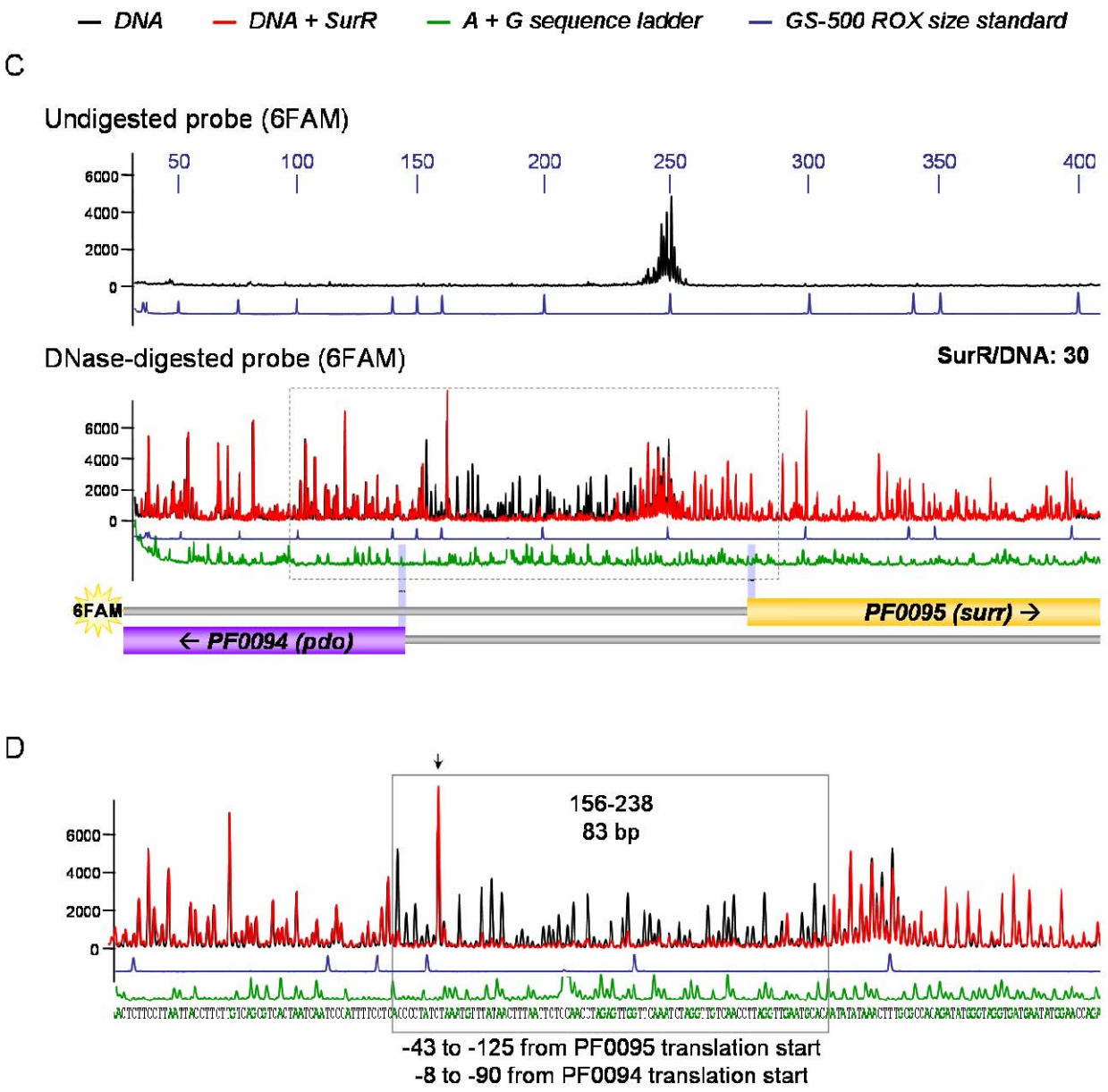

DNA
DNA + SurR
A + G sequence ladder
GS-500 ROX size standard
C
Undigested probe (6FAM)
50
100
150
200
250
300
350
400
6000
4000
2000
0
DNase-digested probe (6FAM)
SurR/DNA: 30
6FAM
PF0095 (surr) →
← PF0094 (pdo)
D
156-238
83 bp
-43 to -125 from PF0095 translation start
-8 to -90 from PF0094 translation start

Figure 4.6 SurR binding within the *surr-pdo* footprint occurs simultaneously.

Electropherogram traces for HEX-labeled strand of DNase-digested probe with and without protein for protein/DNA mole ratios of 12 (top) and 24 (bottom). Peak sizes of internal GS-500 ROX standard are indicated. Probe diagram indicates positions of ORFs relative to electropherogram, and footprint regions are boxed.

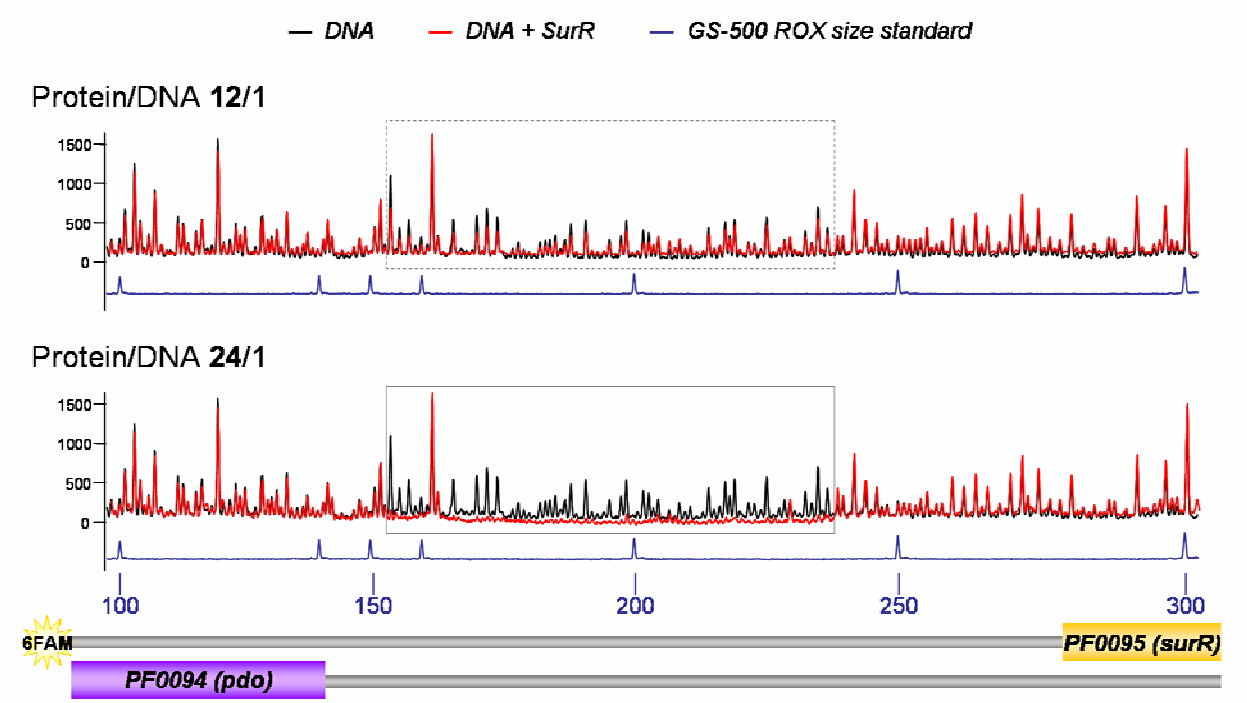

DNA
DNA + SurR
GS-500 ROX size standard
Protein/DNA 12/1
1500
1000
500
0
Protein/DNA 24/1
1500
1000
500
0
100
150
200
250
300
6FAM
PF0095 (surR)
PF0094 (pdo)

Figure 4.7 SELEX scheme. (1) The SELEX library was made up of single-stranded oligonucleotides designed to contain 30 bases of random DNA flanked with two ~20-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 higher-affinity sequences to SurR were selected from the library pool using EMSA. **(4)** Shifted DNA was gel-purified and PCR-amplified for an additional selection round. Six rounds of selection via EMSA, gel-purification, and PCR-amplification were carried out. **(5)** 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 shared motif.

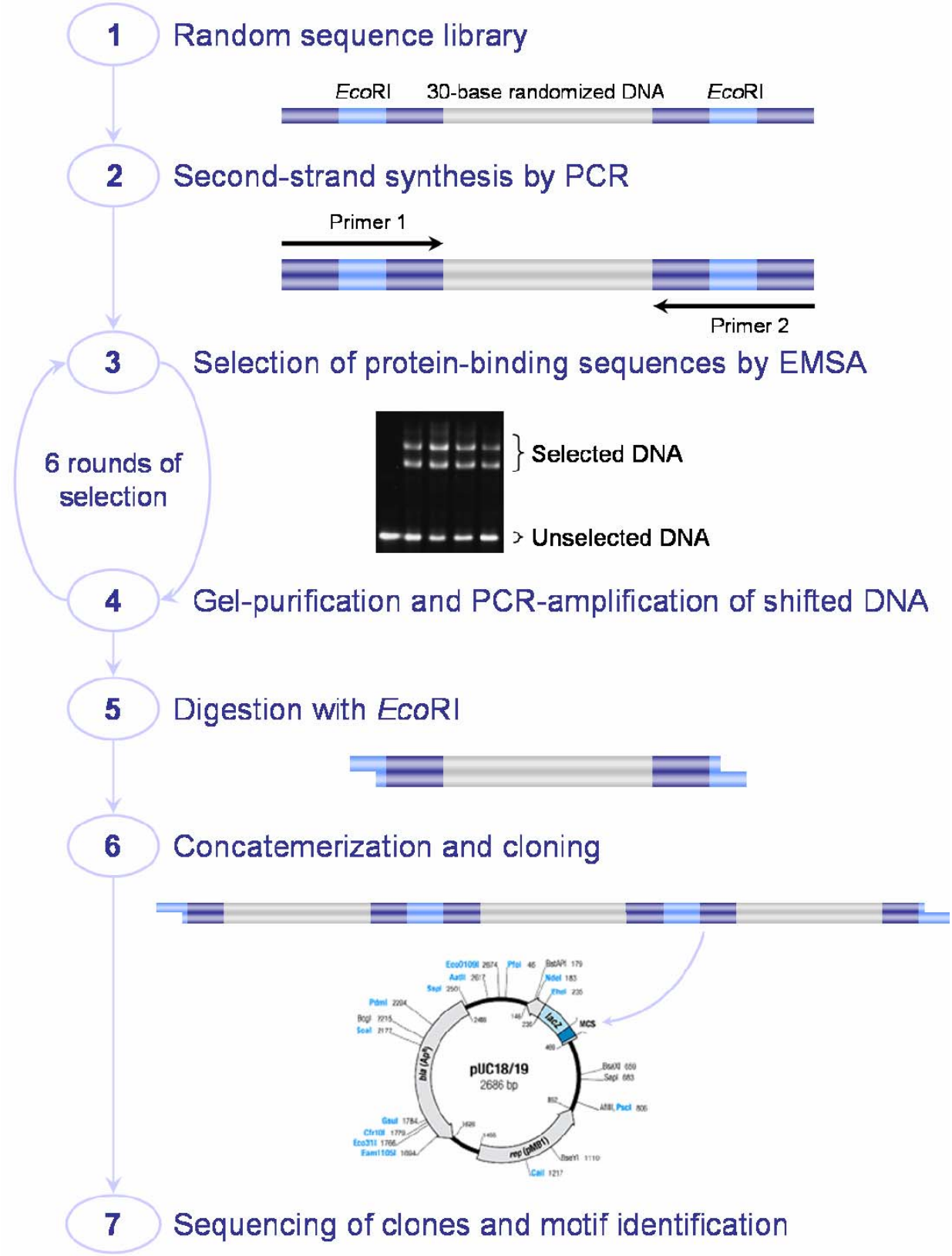
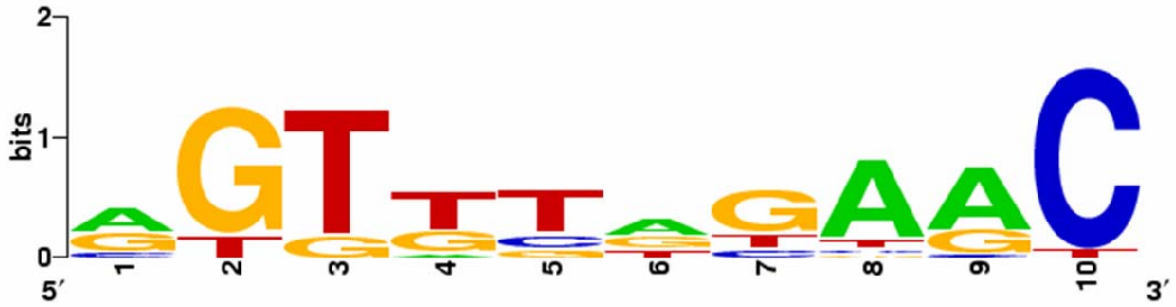
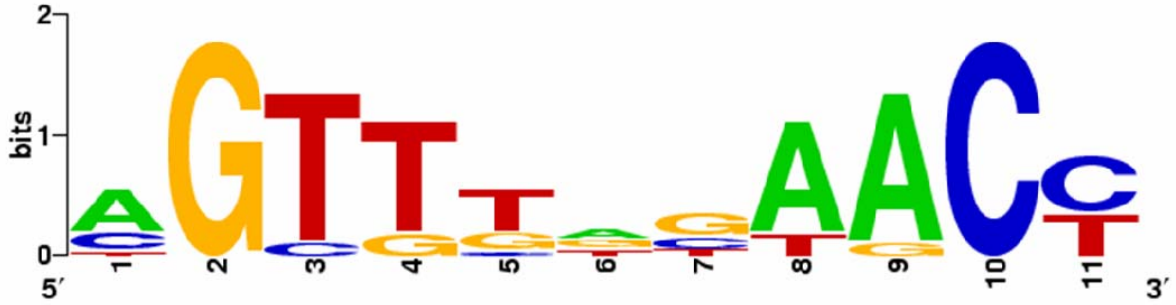

1
Random sequence library
EcoRI
30-base randomized DNA
EcoRI
2
Second-strand synthesis by PCR
Primer 1
Primer 2
3
Selection of protein-binding sequences by EMSA
6 rounds of selection
Selected DNA
Unselected DNA
4
Gel-purification and PCR-amplification of shifted DNA
5
Digestion with EcoRI
6
Concatemerization and cloning
pUC18/19
2686 bp
lacZ
MCS
bla (Ap^R)
rep (pMB1)
7
Sequencing of clones and motif identification

Figure 4.8 SELEX-determined SurR DNA-binding motif. A. Motif logo (generated using WebLogo [96]) for motif uncovered from all 19 selected SELEX sequences as determined by MEME motif-finding software [95]. **B.** Motif logo for 11 out of 19 SELEX sequences with corresponding sequence pile-up.

A



B



tccccatatat **AGTTTTGAACC** ggtttgta
cgaatctatgttgacctgt **AGTTTAGAACT**
aggccat **AGTTGGCAACC** catgggaaccctt
AGTGTACAACC ccttgttcccataatatag
tcatagacataacacccta **CGTTTGTAACT**
tgaaggatg **AGTTTAGAGCT** tacatgtcta
ggtaccgttggtatga **CGTTGGTAACT** gga
taa **AGCTGTGAACC** agaagacactttcaat
atccgtga **CGTTCACAACC** aaatagattga
actgg **TGTTTGGTACC** gcacgatgagttct
aagacga **AGTGTTGTACT** cctaggtgcaat

Figure 4.9 SurR footprints contain the GTTn$_3$AAC SELEX motif. Relevant DNA sequence of PF1423 (top) and PF0094 (bottom) probes are shown with SurR DNase footprint sites boxed in red. The positions of the SELEX consensus are highlighted in yellow and indicated in brackets. 'GTT' sites spaced 5-bp from SELEX consensus sites are also highlighted in yellow.

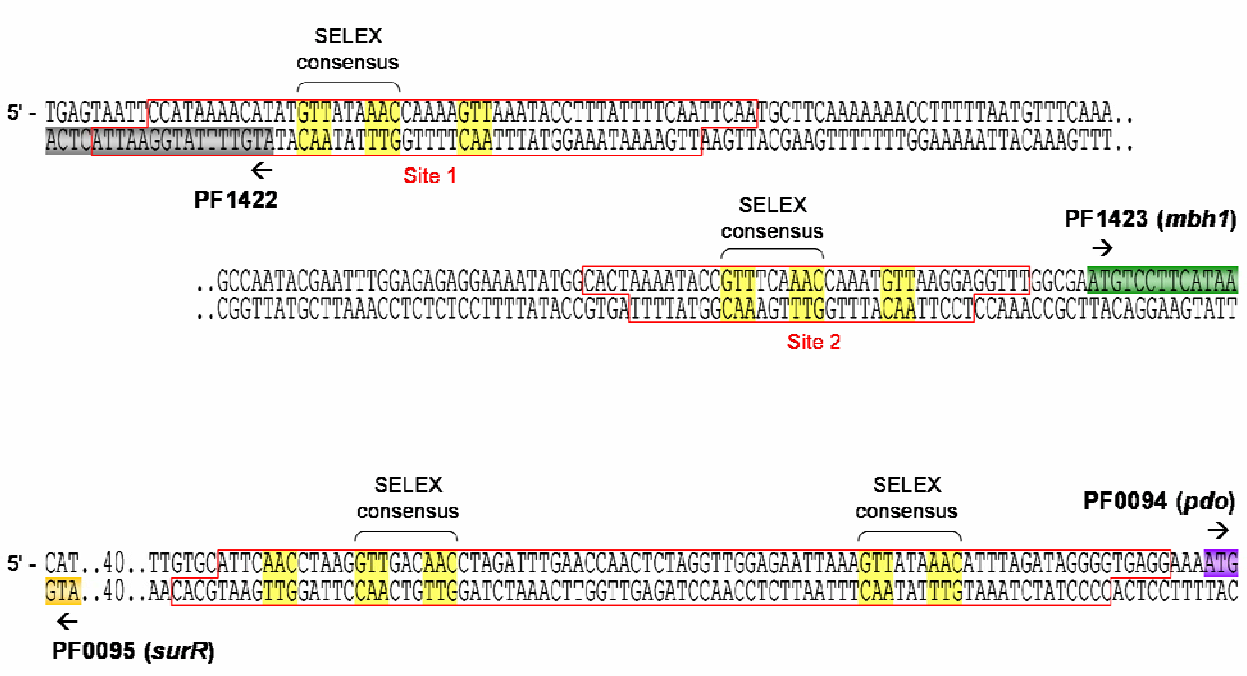

SELEX
consensus
5' - TGAGTAATTCCATAAAACATATGTTATAAACCAAAAGTTAAATACCTTATTTTCAATTCAATGCTTCAAAAAAACCTTTTTAATGTTTCAAA..
ACTCATTAAGGTATTTTGTATACAATATTGGTTTTCAATTTATGGAAATAAAAGTTAAGTTACGAAGTTTTTTTGGAAAAATTACAAAGTTT..
PF1422
Site 1
SELEX
consensus
PF1423 (mbh1)
..GCCAATACGAATTTGGAGAGAGGAAAATATGGCACTAAAATACCGTTTCAAACCAAATGTTAAGGAGGTTTGGCGAATGTCCTTCATAA
..CGGTTATGCTTAAACCTCTCTCCTTTTATACCGTGATTTTATGGCAAAGTTTGGTTTACAATTCCTCCAAACCGCTTACAGGAAGTATT
Site 2
SELEX
consensus
SELEX
consensus
PF0094 (pdo)
5' - CAT..40..TTGTGCATTCAACCTAAGGTTGACAACCTAGATTTGAACCAACTCTAGGTTGGAGAATTAAAGTTATAAACATTTAGATAGGGGTGAGGAAAATG
GTA..40..AACACGTAAGTTGGATTCCAACTGTTGGATCTAAACTTGGTTGAGATCCAACCTCTTAATTTCAATATTTGTAAATCTATCCCCACTCCTTTTAC
PF0095 (surR)

Figure 4.10 Upstream DNA of S^0-regulated ORFs contains the $GTTn_3AAC$ motif.

MEME motif finding software [95] was used to search 200-bp upstream DNA from 10-min S^0 up and down-regulated ORFs/operons for a common motif. Motif sequence logos were generated using WebLogo [96]. Motif logo for 200-bp UORs of $GTTn_3AAC$-containing ORFs which are regulated with S^0: up and down-regulated (**A**), up-regulated only (**B**), and down-regulated only (**C**).

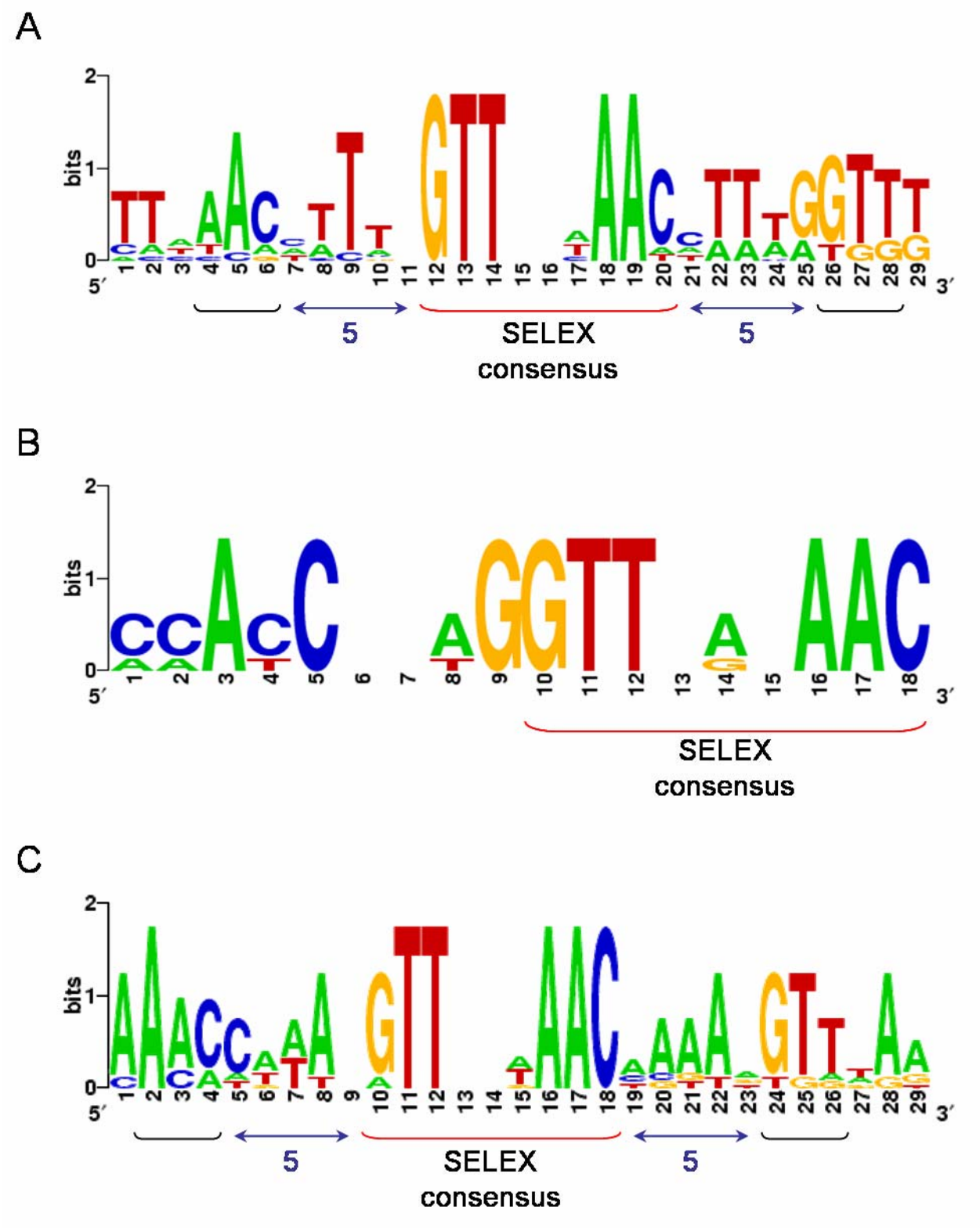
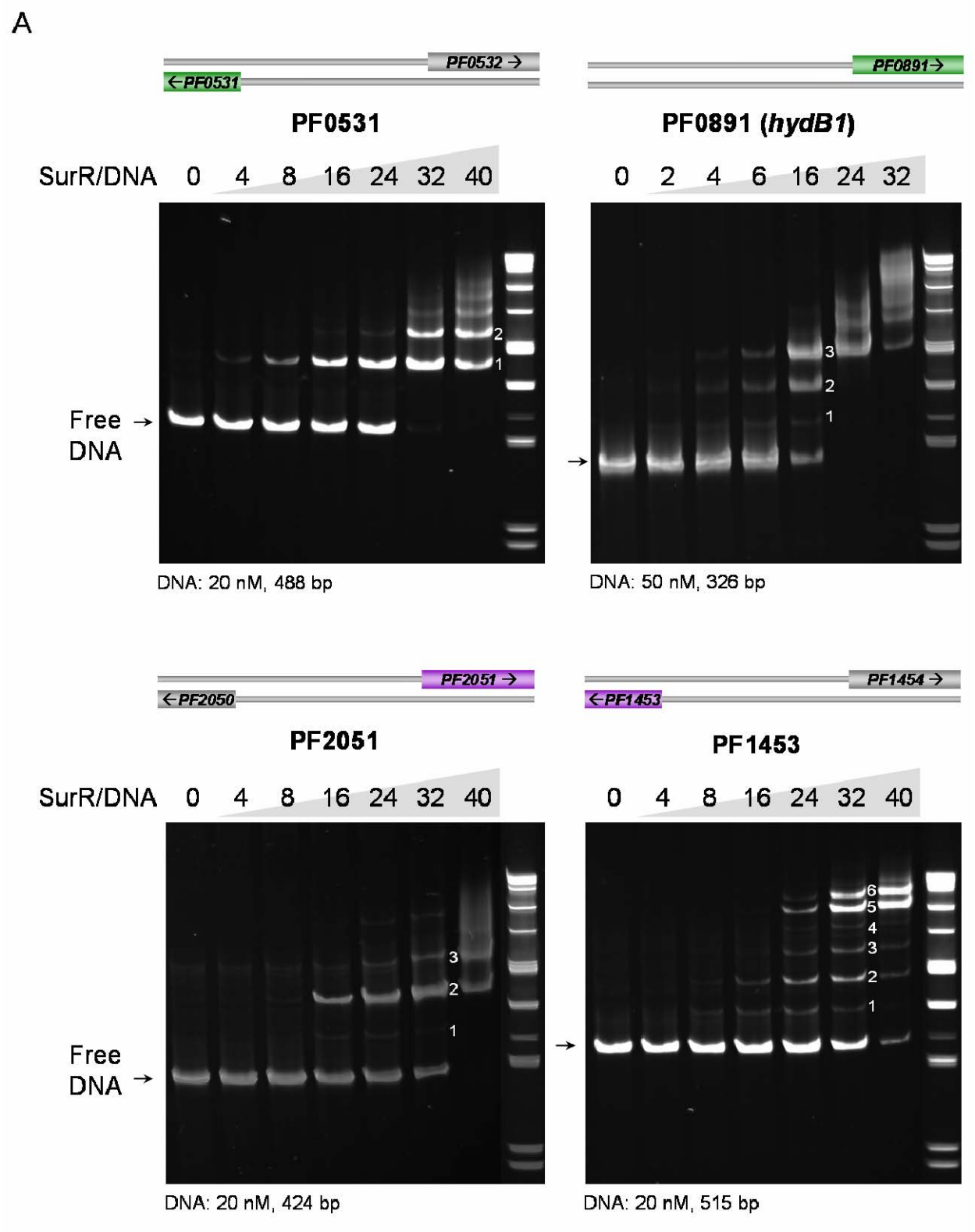

A
bits
5′
3′
5
SELEX
consensus
5
B
bits
5′
3′
SELEX
consensus
C
bits
5′
3′
5
SELEX
consensus
5

Figure 4.11 SurR binds to upstream DNA of additional S^0-regulated ORFs. EMSA with probe DNA from S^0-regulated ORFs containing upstream $GTTn_3AAC$ motifs. Probe diagrams are indicated at the top of each gel image with corresponding protein/DNA mole ratios listed above each lane. Protein-DNA complexes are numbered within the gel images. **A.** EMSAs for SurR with probes containing ORFs down-regulated (purple, top) and up-regulated (green, bottom) within 10 min of S^0 addition to a growing culture. **B.** EMSAs for his_6-SurR with probes containing ORFs down-regulated (purple, top) and up-regulated (green, bottom) within 10 min of S^0 addition to a growing culture.


A
PF0532 →
←PF0531
PF0531
SurR/DNA 0 4 8 16 24 32 40
Free DNA →
2
1
DNA: 20 nM, 488 bp
PF0891→
PF0891 (hydB1)
0 2 4 6 16 24 32
3
2
1
DNA: 50 nM, 326 bp
PF2051 →
←PF2050
PF2051
SurR/DNA 0 4 8 16 24 32 40
Free DNA →
3
2
1
DNA: 20 nM, 424 bp
PF1454 →
←PF1453
PF1453
0 4 8 16 24 32 40
6
5
4
3
2
1
DNA: 20 nM, 515 bp

B

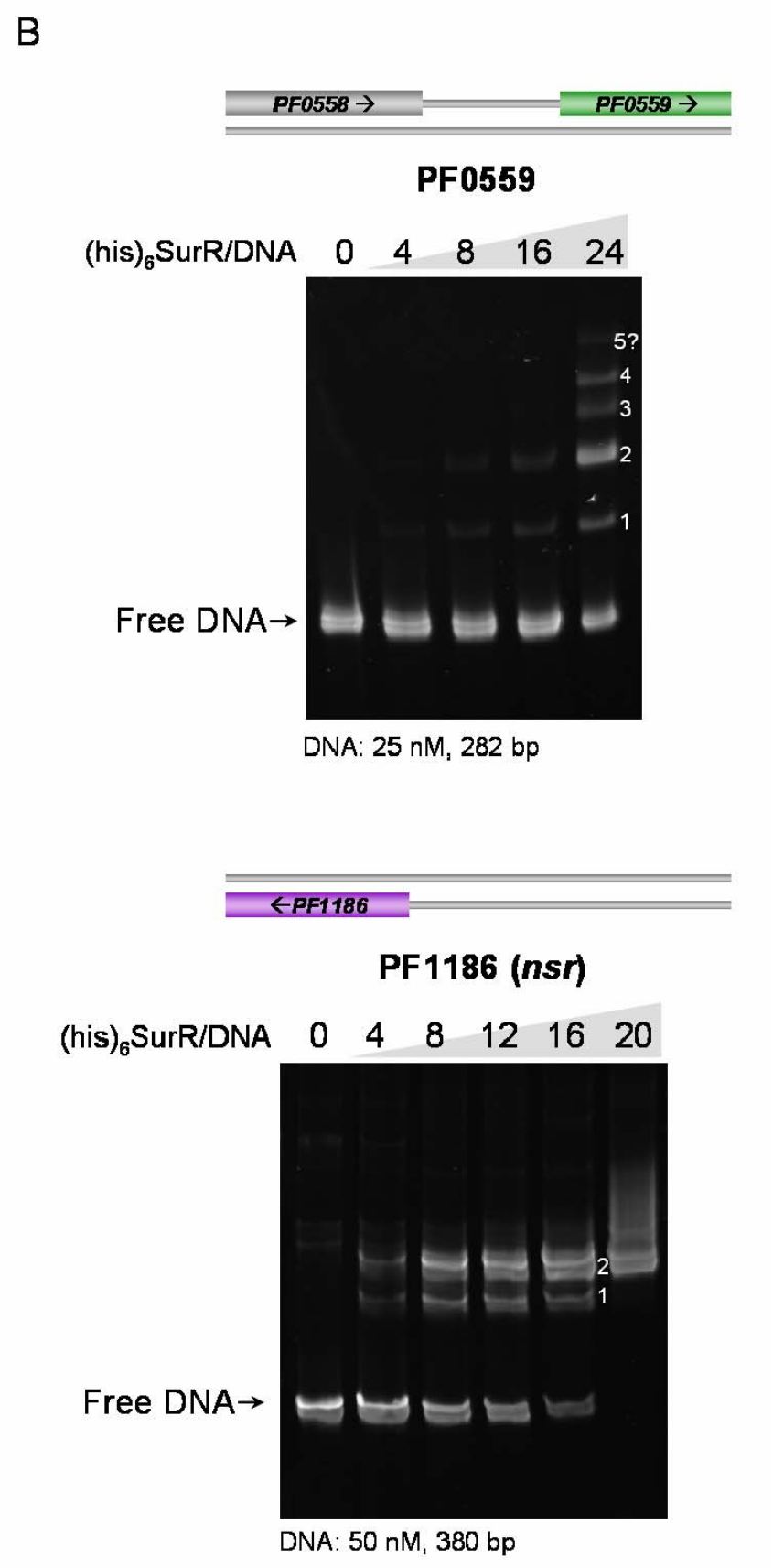

PF0558 →
PF0559 →
PF0559
(his)6SurR/DNA 0 4 8 16 24
5?
4
3
2
1
Free DNA→
DNA: 25 nM, 282 bp
←PF1186
PF1186 (nsr)
(his)6SurR/DNA 0 4 8 12 16 20
2
1
Free DNA→
DNA: 50 nM, 380 bp

Figure 4.12 DNase I footprint of SurR on PF0531 probe. A. Diagram of PF0531 probe used for footprinting assay, indicating ORFs present. **B.** Electropherogram traces of 6FAM-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Internal standards GS-500 ROX (peak sizes in bases are labeled) and A+G probe-specific sequence ladder were used to determine position of electropherogram peaks relative to probe sequence. Probe diagram indicates positions of ORFs relative to electropherogram traces. The SurR footprint region is boxed, and the DNA sequence as determined by the ladder is shown expanded below. SELEX consensus (red bracket) and degenerate sites are highlighted in yellow. Asterisks denote bases that are consistent with the $GTTn_3AAC$ consensus. **C.** Relevant DNA sequence of PF0531 probe is shown with SurR DNase footprint site boxed in red. The putative positions of the BRE/TATA elements are indicated.

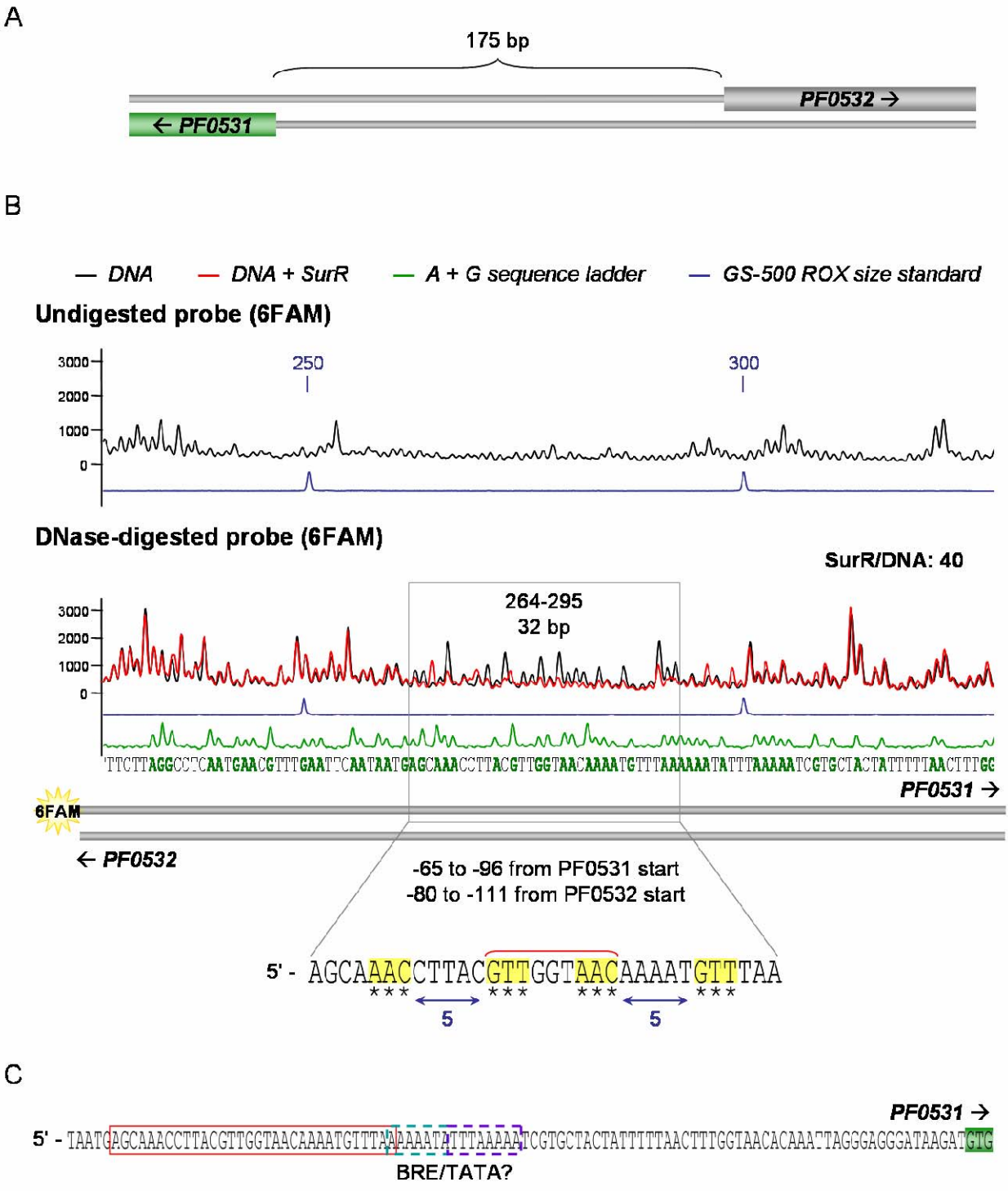
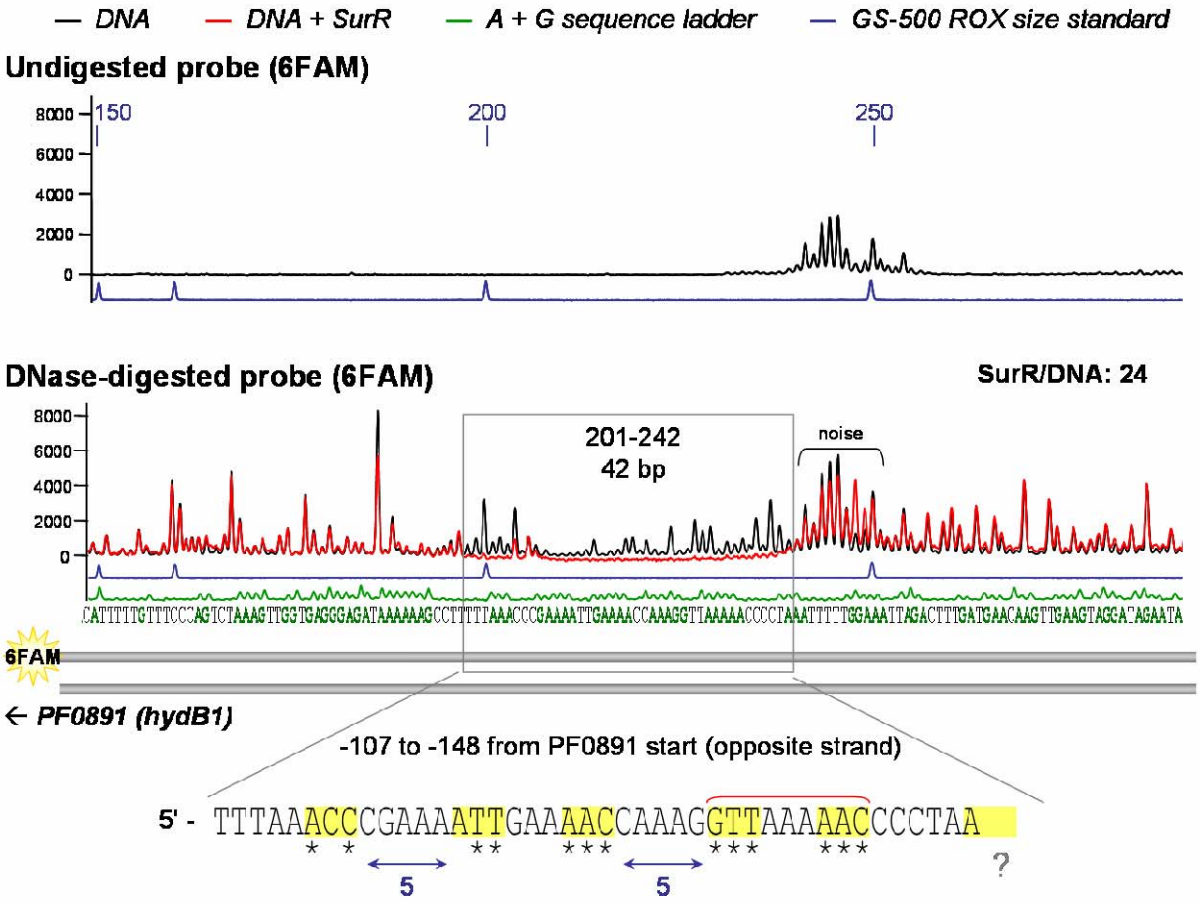

A
175 bp
PF0532 →
← PF0531
B
— DNA
— DNA + SurR
— A + G sequence ladder
— GS-500 ROX size standard
Undigested probe (6FAM)
250
300
DNase-digested probe (6FAM)
SurR/DNA: 40
264-295
32 bp
PF0531 →
6FAM
← PF0532
-65 to -96 from PF0531 start
-80 to -111 from PF0532 start
5' - AGCAAACCTTACGTTGGTAACAAAATGTTTAA
5
5
C
PF0531 →
5' - TAATGAGCAAACCTTACGTTGGTAACAAAATGTTTAAAAATATTTAAAAATCGTGCTACTATTTTTAACTTTGGTAACACAAACTAGGGAGGGATAAGATGTG
BRE/TATA?

Figure 4.13 DNase I footprint of SurR on PF0891 probe. A. Diagram of PF0891 probe used for footprinting assay. **B.** Electropherogram traces of 6FAM-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Internal standards GS-500 ROX (peak sizes in bases are labeled) and A+G probe-specific sequence ladder were used to determine position of electropherogram peaks relative to probe sequence. Probe diagram indicates positions of ORFs relative to electropherogram traces. The SurR footprint region is boxed, and the DNA sequence as determined by the ladder is shown expanded below. SELEX consensus (red bracket) and degenerate sites are highlighted in yellow. Asterisks denote bases that are consistent with the $GTTn_3AAC$ consensus. **C.** Relevant DNA sequence of PF0891 probe is shown with SurR DNase footprint site boxed in red. The positions of the BRE/TATA elements and transcription initiation site are indicated.

A


PF0891 (hydB1) →

B


— DNA
— DNA + SurR
— A + G sequence ladder
— GS-500 ROX size standard
Undigested probe (6FAM)
8000
6000
4000
2000
0
150
200
250
DNase-digested probe (6FAM)
SurR/DNA: 24
201-242
42 bp
noise
6FAM
← PF0891 (hydB1)
-107 to -148 from PF0891 start (opposite strand)
5' - TTTAAACCCGAAAATTGAAAACCAAAGGTTAAAAACCCCTAA
5
5
?

C

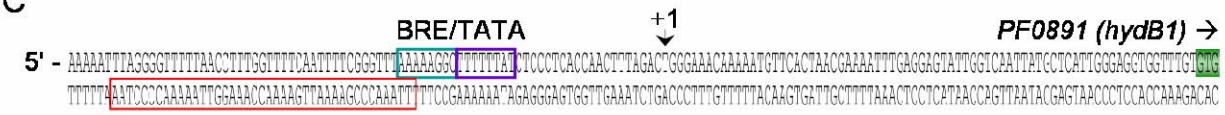
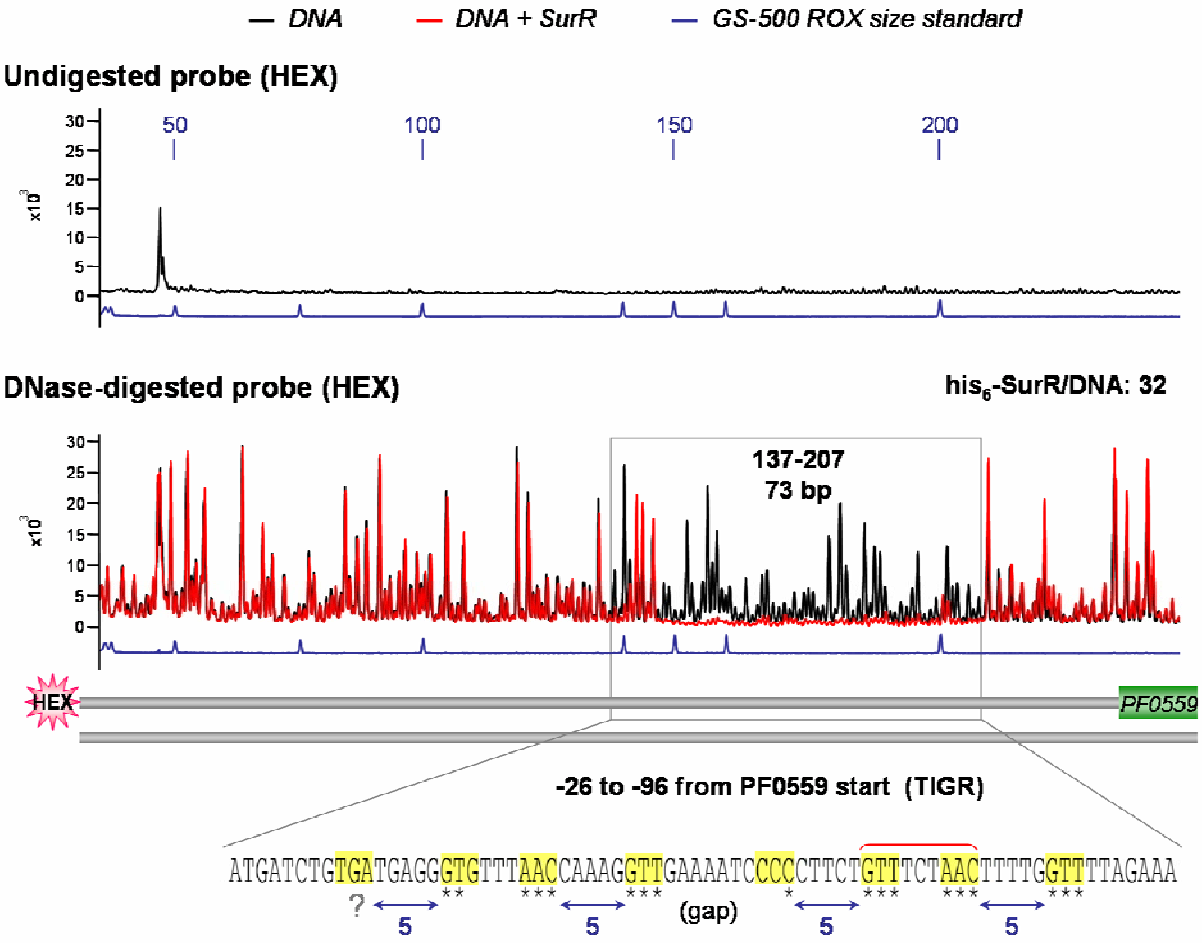

BRE/TATA
+1
PF0891 (hydB1) →
5' -

Figure 4.14 DNase I footprint of SurR on PF0559 probe. A. Diagram of PF0559 probe used for footprinting assay. **B.** Electropherogram traces of HEX-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). The internal standard GS-500 ROX (peak sizes in bases are labeled) was used to determine position of electropherogram peaks relative to the probe sequence. Probe diagram indicates positions of ORFs relative to electropherogram traces. The SurR footprint region is boxed, and the DNA sequence as determined by the standard is shown expanded below. SELEX consensus (red bracket) and degenerate sites are highlighted in yellow. Asterisks denote bases that are consistent with the $GTTn_3AAC$ consensus. **C.** Relevant DNA sequence of PF0559 probe is shown with SurR DNase footprint site boxed in red. The putative positions of the BRE/TATA elements are indicated, as well as two potential start sites as defined by TIGR and NCBI.

A


PF0559 →

B


DNA
DNA + SurR
GS-500 ROX size standard
Undigested probe (HEX)
DNase-digested probe (HEX)
his6-SurR/DNA: 32
137-207
73 bp
HEX
PF0559
-26 to -96 from PF0559 start (TIGR)
ATGATCTGTGATGAGGGTGTTTAACCAAAGGTTGAAAATCCCCCTTCTGTTTCTAACTTTTGGTTTTAGAAACCTATAAAGGCTCCAAATTCGGAG
(gap)
5
5
5
5

C

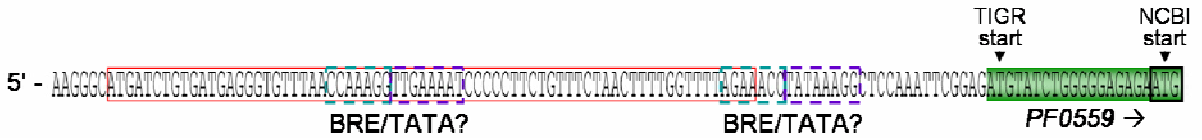

TIGR start
NCBI start
5' - AAGGGCATGATCTGTGATGAGGGTGTTTAACCAAAGGTTGAAAATCCCCCTTCTGTTTCTAACTTTTGGTTTTAGAAACCTATAAAGGCTCCAAATTCGGAGATGTATCTGGGGGAGAGAATG
BRE/TATA?
BRE/TATA?
PF0559 →

Figure 4.15 DNase I footprint of SurR on PF1453 probe. A. Diagram of PF1453 probe used for footprinting assay, indicating ORFs present. **B.** Electropherogram traces of HEX-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Internal standards GS-500 ROX (peak sizes in bases are labeled) and C+T probe-specific sequence ladder were used to determine position of electropherogram peaks relative to probe sequence. Probe diagram indicates positions of ORFs relative to electropherogram traces. The SurR footprint region is boxed, and the DNA sequence as determined by the ladder is shown expanded below. SELEX consensus (red bracket) and degenerate sites are highlighted in yellow. Asterisks denote bases that are consistent with the $GTTn_3AAC$ consensus. **C.** Relevant DNA sequence of PF1453 probe is shown with SurR DNase footprint site boxed in red. The putative positions of the BRE/TATA elements are indicated.

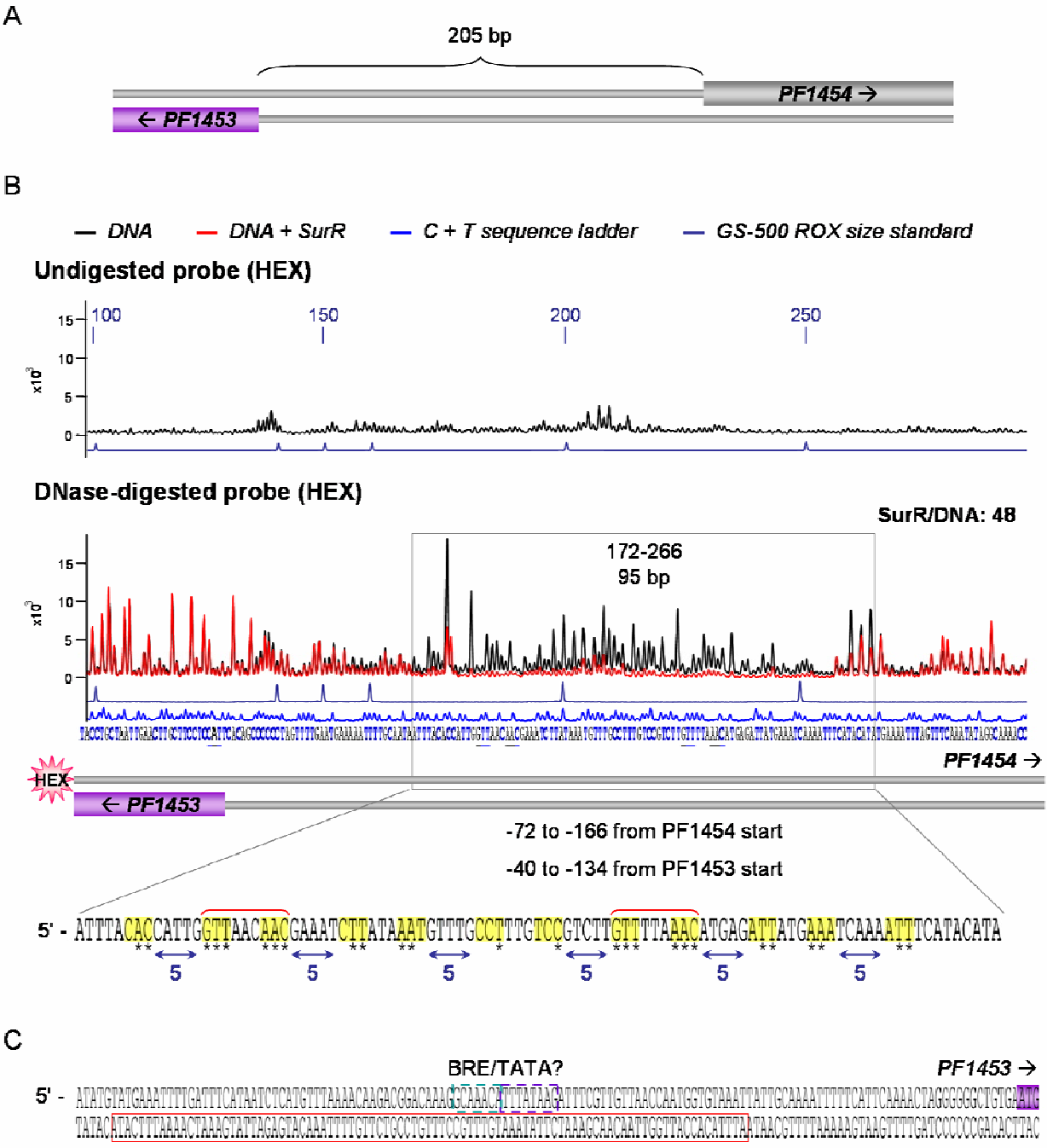

A
205 bp
PF1454 →
← PF1453
B
DNA
DNA + SurR
C + T sequence ladder
GS-500 ROX size standard
Undigested probe (HEX)
100
150
200
250
x10³
DNase-digested probe (HEX)
SurR/DNA: 48
172-266
95 bp
PF1454 →
HEX
← PF1453
-72 to -166 from PF1454 start
-40 to -134 from PF1453 start
5' - ATTTACACCATTGGTTAACAACGAAATCTTATAAATGTTTGCCTTGTCCGTCTTGTTTTAAACATGAGATTATGAAATCAAAATTTCATACATA
5
C
BRE/TATA?
PF1453 →
5' -

Figure 4.16 DNase I footprint of SurR on PF2051 probe. A. Diagram of PF2051 probe used for footprinting assay, indicating ORFs present. **B.** Electropherogram traces of HEX-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Internal standards GS-500 ROX (peak sizes in bases are labeled) and C+T probe-specific sequence ladder were used to determine position of electropherogram peaks relative to probe sequence. Probe diagram indicates positions of ORFs relative to electropherogram traces. The SurR footprint region is boxed, and the DNA sequence as determined by the ladder is shown expanded below. SELEX consensus (red bracket) and degenerate sites are highlighted in yellow. Asterisks denote bases that are consistent with the $GTTn_3AAC$ consensus. **C.** Relevant DNA sequence of PF2051 probe is shown with SurR DNase footprint site boxed in red. The putative positions of the BRE/TATA elements are indicated.

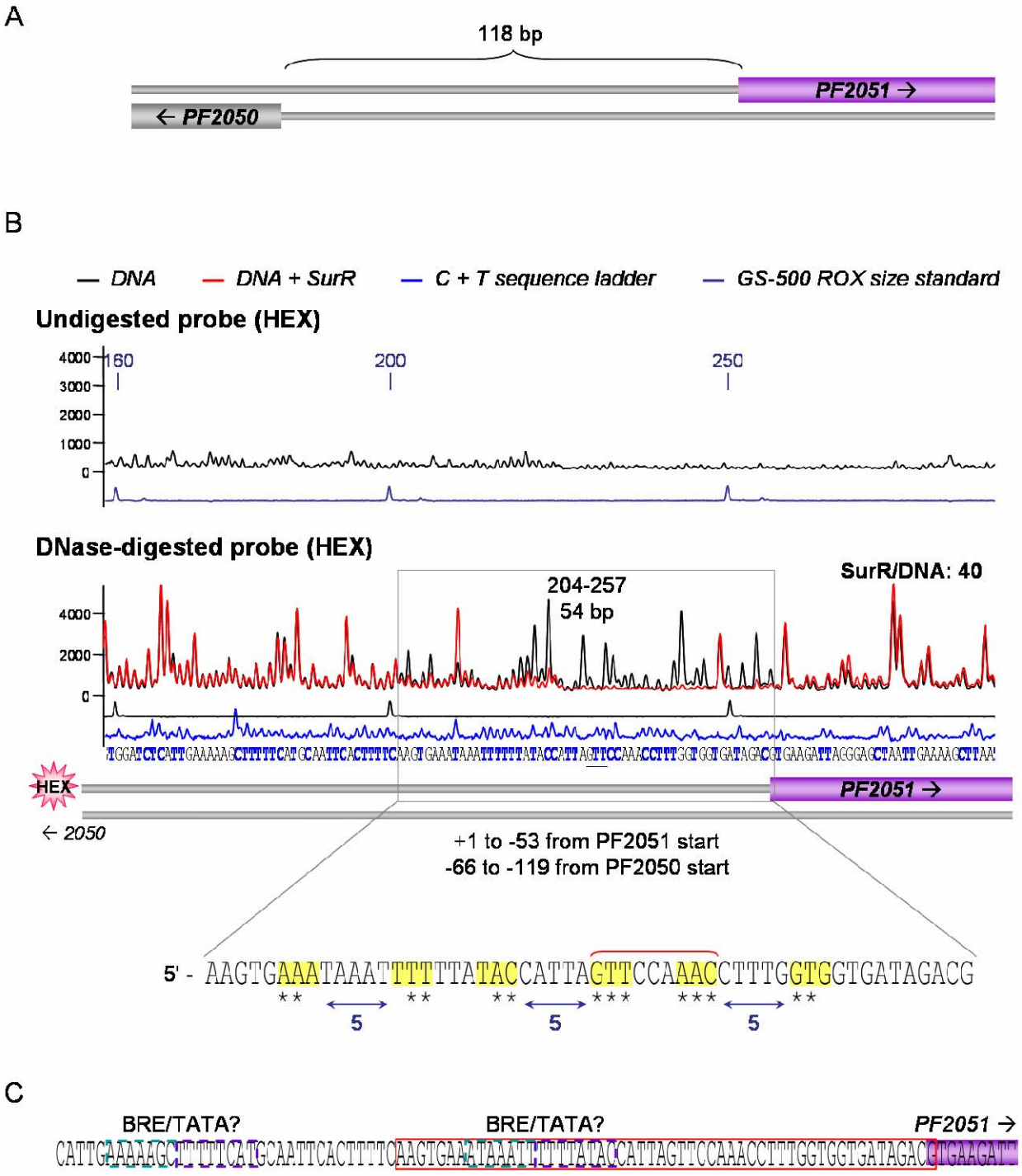
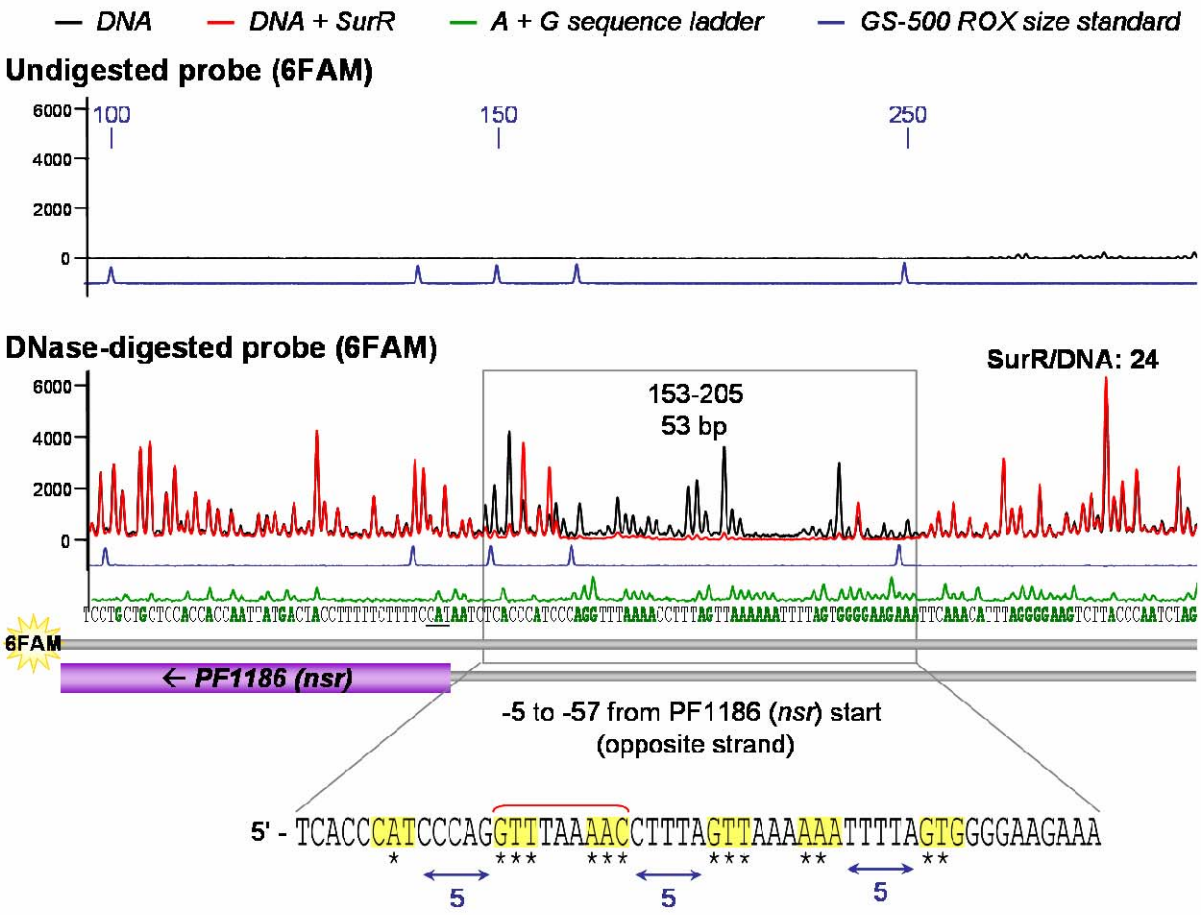

A
118 bp
PF2051 →
← PF2050
B
DNA
DNA + SurR
C + T sequence ladder
GS-500 ROX size standard
Undigested probe (HEX)
4000
3000
2000
1000
0
160
200
250
DNase-digested probe (HEX)
SurR/DNA: 40
204-257
54 bp
4000
2000
0
HEX
PF2051 →
← 2050
+1 to -53 from PF2051 start
-66 to -119 from PF2050 start
5' - AAGTGAAATAAATTTTTATACCATTAGTTCCAAACCTTTGGTGGTGATAGACG
5
5
5
C
BRE/TATA?
BRE/TATA?
PF2051 →

Figure 4.17 DNase I footprint of SurR on PF1186 probe. A. Diagram of PF1186 probe used for footprinting assay, indicating ORFs present. **B.** Electropherogram traces of 6FAM-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Internal standards GS-500 ROX (peak sizes in bases are labeled) and A+G probe-specific sequence ladder are used to determine position of electropherogram peaks relative to probe sequence. Probe diagram indicates positions of ORFs relative to electropherogram traces. The SurR footprint region is boxed, and the DNA sequence as determined by the ladder is shown expanded below. SELEX consensus (red bracket) and degenerate sites are highlighted in yellow. Asterisks denote bases that are consistent with the $GTTn_3AAC$ consensus. **C.** Relevant DNA sequence of PF1186 probe is shown with SurR DNase footprint site boxed in red. The putative positions of the BRE/TATA elements are indicated.

A


← PF1186 (nsr)

B


DNA
DNA + SurR
A + G sequence ladder
GS-500 ROX size standard
Undigested probe (6FAM)
6000
4000
2000
0
100
150
250
DNase-digested probe (6FAM)
SurR/DNA: 24
153-205
53 bp
6FAM
← PF1186 (nsr)
-5 to -57 from PF1186 (nsr) start
(opposite strand)
5' - TCACCCATCCCAGGTTTAAAACCTTTAGTTAAAAAATTTTAGTGGGGAAGAAA
5
5
5

C


BRE/TATA?
BRE/TATA?
PF1186 (nsr) →
5' - ACTTCCCCTAAATGTTTGAATTTCTTCCCCACTAAAATTTTTTAACTAAAGGTTTTAAACCTGGGATGGGTGAGATTATG
TGAAGGGGATTTACAAACTTAAAGAAGGGGTCATTTTAAAAAATTGATTTCCAAAATTTGGACCCTACCCACTCTAATAC

Figure 4.18 SurR recognition motifs and BRE/TATA elements of *mbh1* and *pdo*. A. Relevant DNA sequence of PF1423 probe is shown with SurR DNase footprint sites boxed in red. SELEX consensus (red bracket) and degenerate sites are highlighted in yellow. Asterisks denote bases that are consistent with the $GTTn_3AAC$ consensus. The position of the BRE/TATA elements and the PF1423 transcription start site are indicated. **B.** Relevant DNA sequence of PF0094 probe is shown with SurR DNase footprint site boxed in red. SELEX consensus (red bracket) and degenerate sites are highlighted in yellow. Asterisks denote bases that are consistent with the $GTTn_3AAC$ consensus. Possible positions of the BRE/TATA elements are indicated.

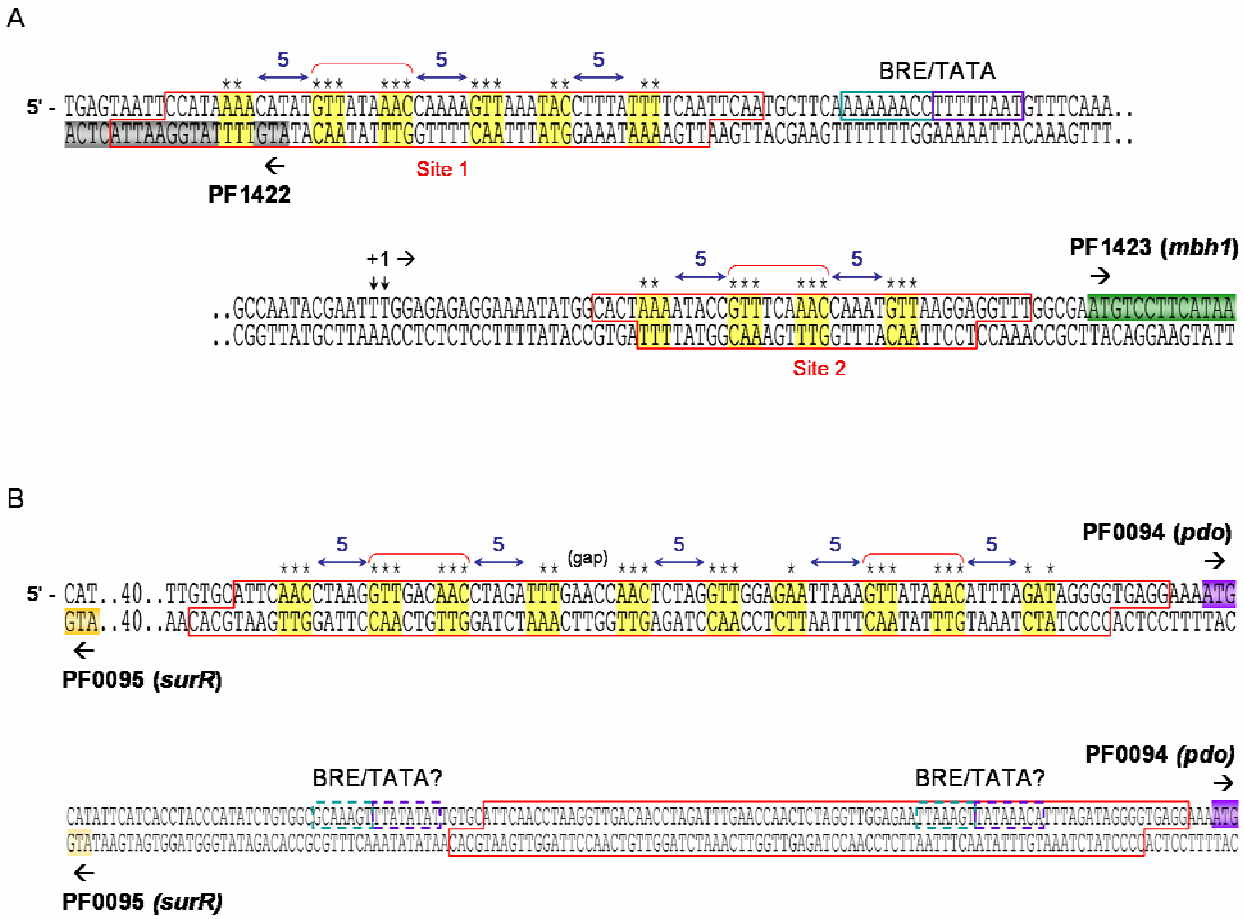
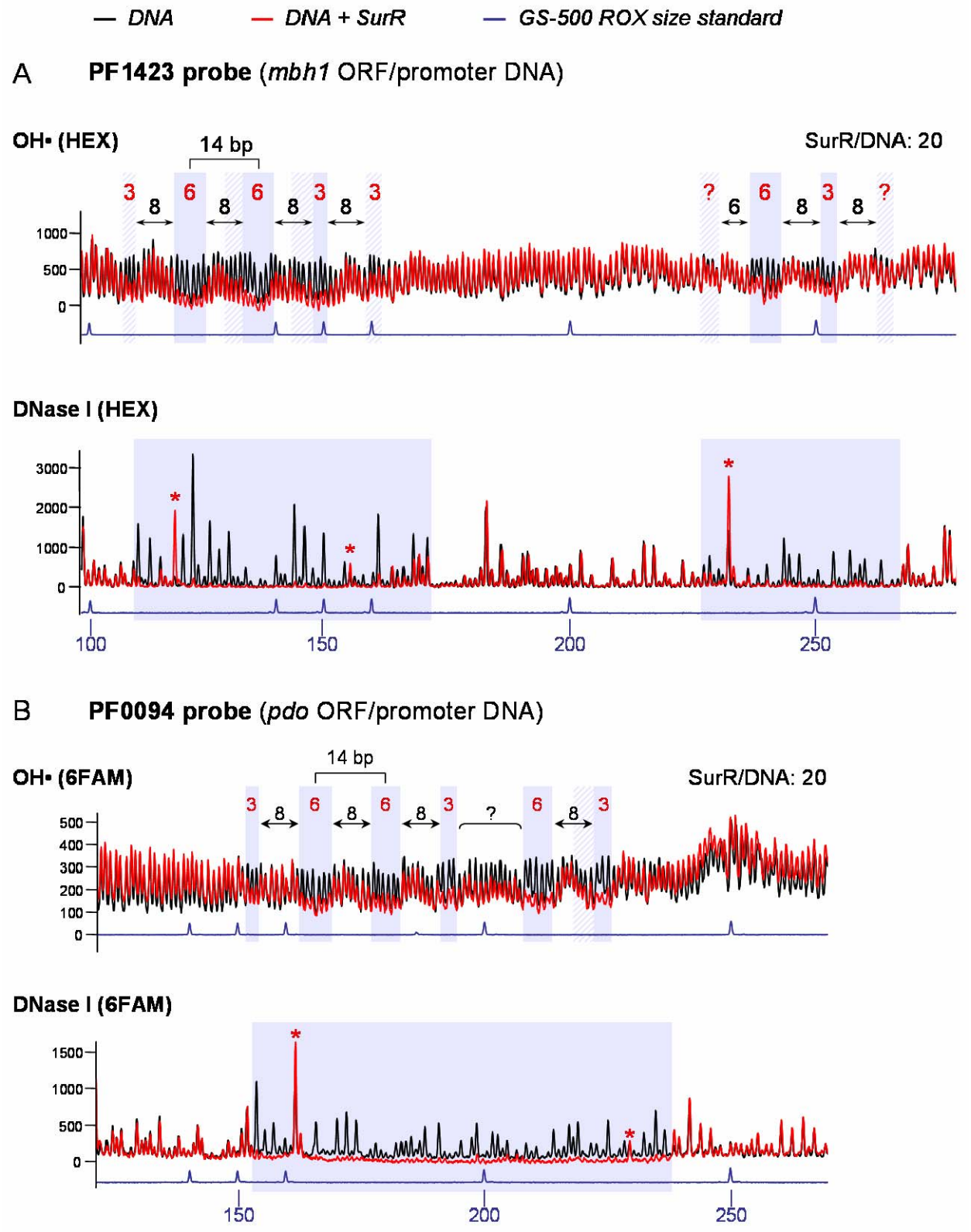

A
5
5
5
BRE/TATA
5' - TGAGTAATTCCATAAAACATATGTTATAAACCAAAAGTTAAATACCTTTATTTTCAATTCAATGCTTCAAAAAAACCTTTTTAATGTTTCAAA..
ACTCATTAAGGTATTTTGTATACAATATTTGGTTTTCAATTTATGGAAATAAAAGTTAAGTTACGAAGTTTTTTTGGAAAAATTACAAAGTTT..
PF1422
Site 1
+1
PF1423 (mbh1)
..GCCAATACGAATTTGGAGAGAGGAAAATATGGCACTAAAATACCGTTTCAAACCAAATGTTAAGGAGGTTTGGCGAATGTCCTTCATAA
..CGGTTATGCTTAAACCTCTCTCCTTTTATACCGTGATTTTATGGCAAAGTTTGGTTTACAATTCCTCCAAACCGCTTACAGGAAGTATT
Site 2
B
PF0094 (pdo)
5
5
(gap)
5
5
5
5' - CAT..40..TTGTGCATTCAACCTAAGGTTGACAACCTAGATTTGAACCAACTCTAGGTTGGAGAATTAAAGTTATAAACATTTAGATAGGGGTGAGGAAAATG
GTA..40..AACACGTAAGTTGGATTCCAACTGTTGGATCTAAACTTGGTTGAGATCCAACCTCTTAATTTCAATATTTGTAAATCTATCCCCACTCCTTTTAC
PF0095 (surR)
PF0094 (pdo)
BRE/TATA?
BRE/TATA?
CATATTCATCACCTACCCATATCTGTGGCGCAAAGTTTATATATTGTGCATTCAACCTAAGGTTGACAACCTAGATTTGAACCAACTCTAGGTTGGAGAATTAAAGTTATAAACATTTAGATAGGGGTGAGGAAAATG
GTATAAGTAGTGGATGGGTATAGACACCGCGTTTCAAATATATAACACGTAAGTTGGATTCCAACTGTTGGATCTAAACTTGGTTGAGATCCAACCTCTTAATTTCAATATTTGTAAATCTATCCCCACTCCTTTTAC
PF0095 (surR)

Figure 4.19 OH• footprints show SurR DNA-binding patterns. OH• footprint results for PF1423 (A), PF0094 (B), PF0891 (C), and PF1186 (D) probes compared with their corresponding DNase I footprints. The peaks which show the most protection from OH• cleavage are highlighted in yellow with length and separation for each highlighted region indicated (numbers are in base pairs). Evidence of slight protection is highlighted in yellow with crosshatching. For DNase I footprints, DNase I hypersensitive sites are denoted with an asterisk.


— DNA
— DNA + SurR
— GS-500 ROX size standard
A PF1423 probe (mbh1 ORF/promoter DNA)
OH• (HEX)
14 bp
SurR/DNA: 20
3 8 6 8 6 8 3 8 3
? 6 6 8 3 8 ?
1000
500
0
DNase I (HEX)
3000
2000
1000
0
100
150
200
250
B PF0094 probe (pdo ORF/promoter DNA)
OH• (6FAM)
14 bp
SurR/DNA: 20
3 8 6 8 6 8 3 ? 6 8 3
500
400
300
200
100
0
DNase I (6FAM)
1500
1000
500
0
150
200
250

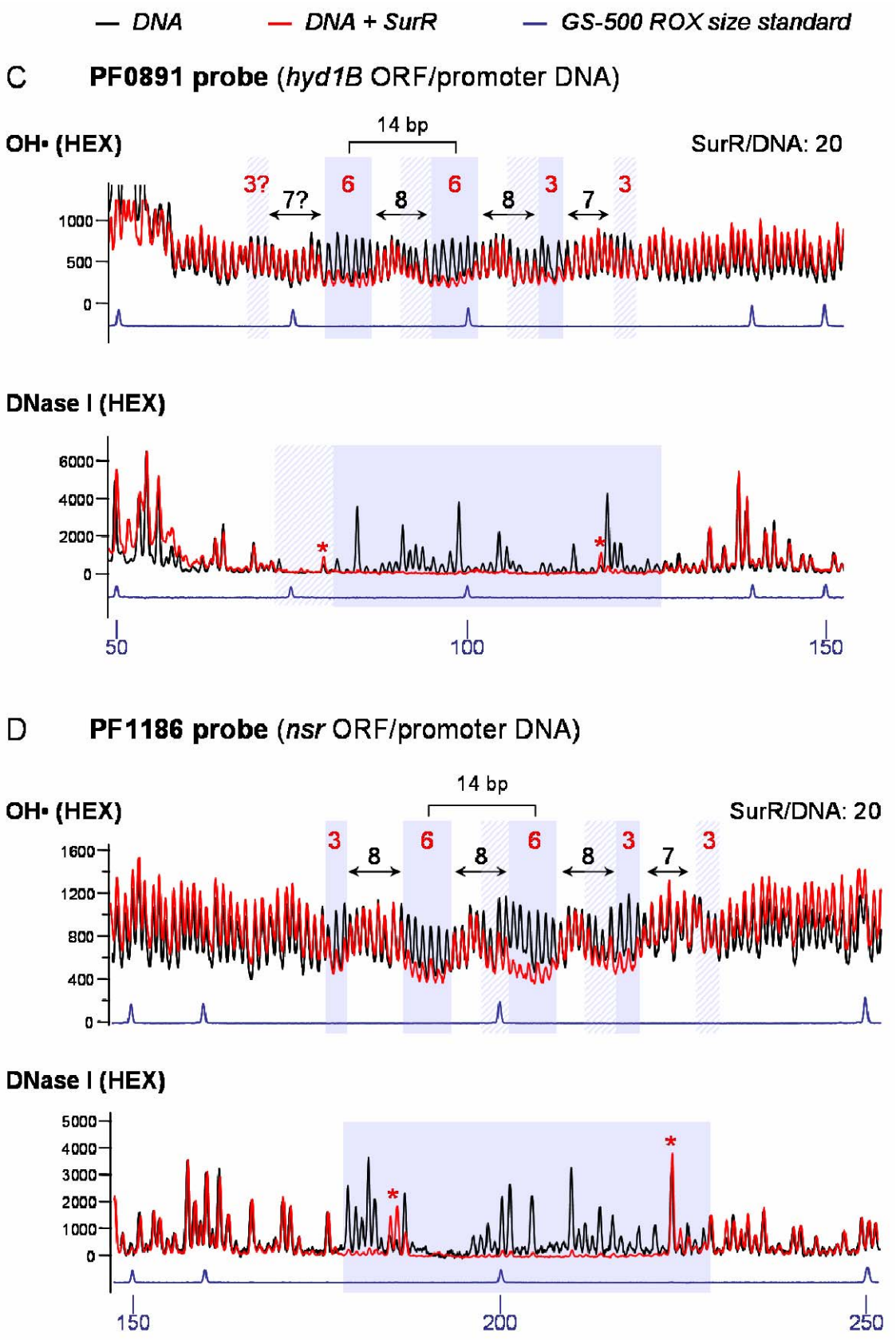
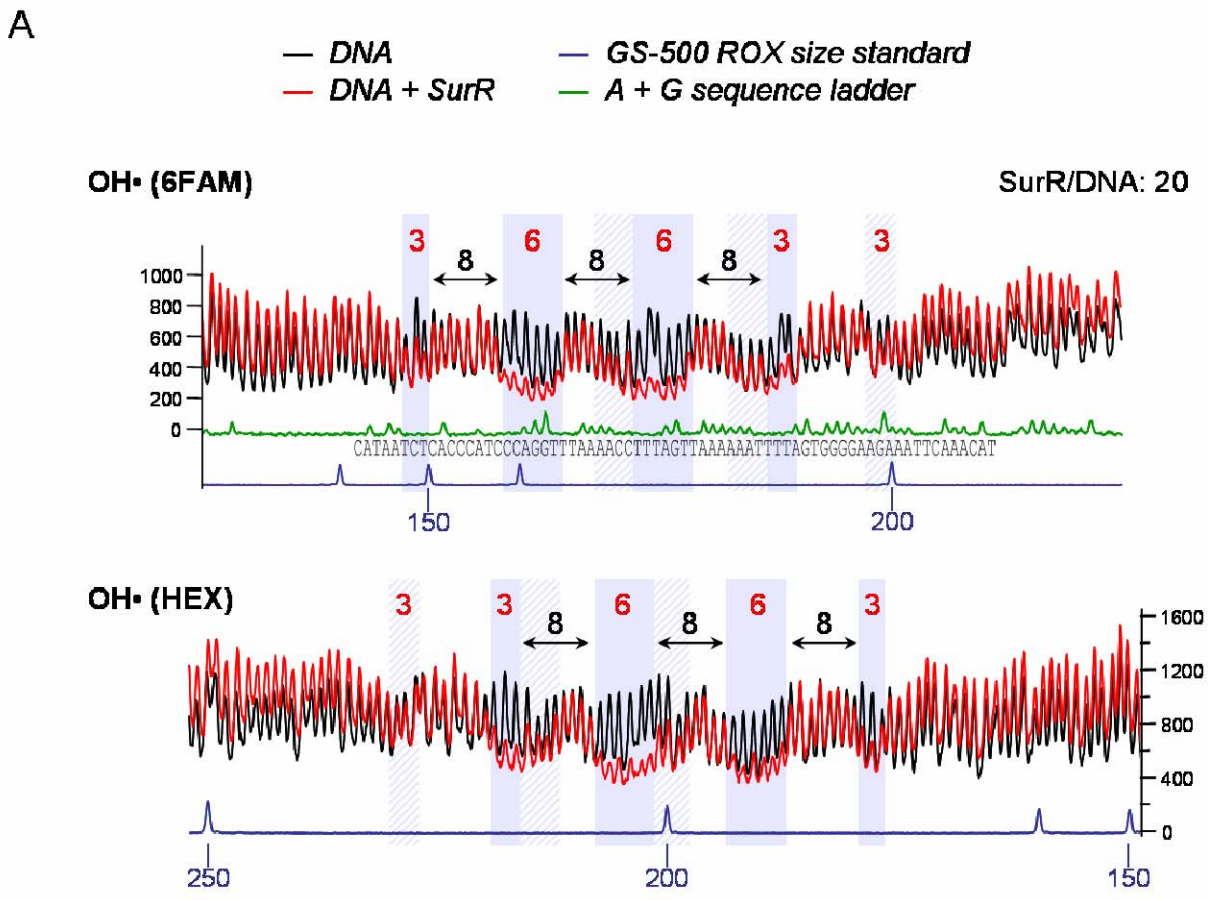

— DNA
— DNA + SurR
— GS-500 ROX size standard
C PF0891 probe (hyd1B ORF/promoter DNA)
OH• (HEX)
14 bp
SurR/DNA: 20
3?
7?
6
8
6
8
3
7
3
1000
500
0
DNase I (HEX)
6000
4000
2000
0
50
100
150
D PF1186 probe (nsr ORF/promoter DNA)
OH• (HEX)
14 bp
SurR/DNA: 20
3
8
6
8
6
8
3
7
3
1600
1200
800
400
0
DNase I (HEX)
5000
4000
3000
2000
1000
0
150
200
250

Figure 4.20 OH• footprint contacts of SurR on the *nsr* promoter. A. OH• footprint results for PF1186 probe (containing *nsr* ORF/promoter DNA). The peaks which show the most protection from OH• cleavage are highlighted in yellow with length and separation for each highlighted region indicated (numbers are in base pairs). Evidence of slight protection is highlighted in yellow with crosshatching. **B.** Corresponding protected bases within the DNA sequence are highlighted according to strand (red for 6FAM and blue for HEX). GTT sequences are boxed in yellow. Corresponding three-dimensional B-DNA helix is shown below with bases highlighted according to the sequence. B-DNA of PF1186 upstream sequence fragment was created using the Make-NA server [123], and DNA structure image was produced using PyMOL [124].

A


DNA
DNA + SurR
GS-500 ROX size standard
A + G sequence ladder
OH• (6FAM)
SurR/DNA: 20
3
8
6
8
6
8
3
3
1000
800
600
400
200
0
CATAATCTCACCCATCCCAGGTTTAAAACCTTTAGTTAAAAAATTTTAGTGGGGAAGAAATTCAAACAT
150
200
OH• (HEX)
3
3
8
6
8
6
8
3
1600
1200
800
400
0
250
200
150

B

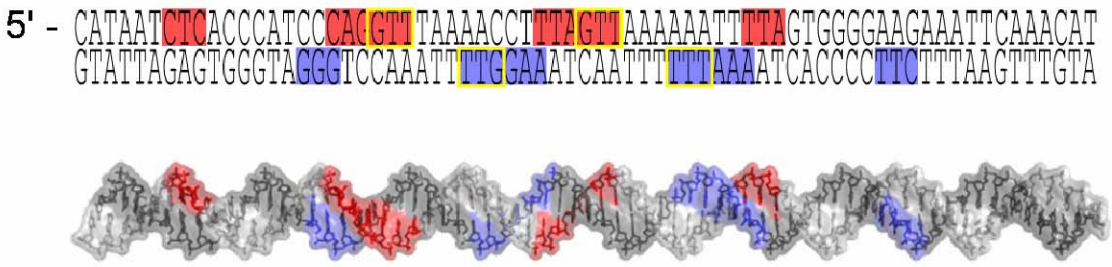

5' - CATAATCTCACCCATCCCAGGTTTAAAACCTTTAGTTAAAAAATTTTAGTGGGGAAGAAATTCAAACAT
GTATTAGAGTGGGTAGGGTCCAAATTTTGGAAATCAATTTTTTAAAATCACCCCTTCTTTAAGTTTGTA

Figure 4.21 SurR DNA-binding motif for S^0 up-regulated ORFs contains a possible downstream BRE/TATA. MEME motif finding software [95] was used to search the coding strand of 200-bp upstream DNA from 10-min S^0 up-regulated ORFs/operons for a common motif (PF0094, PF1186, PF1453, PF2051). Motif sequence logos were generated using WebLogo [96]. The 3' end of the motif may be the beginning of a conserved BRE/TATA element.

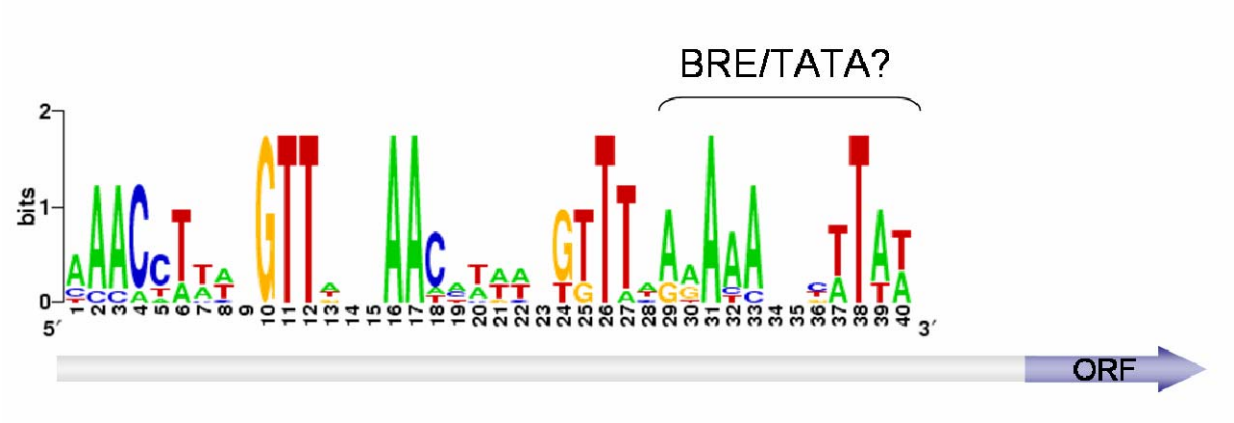

BRE/TATA?
bits
2
1
0
5′
3′
ORF

Figure 4.22 SurR is a transcriptional activator. *In vitro* transcription assays with his_6-SurR performed by Annette Keese (unpublished data from the laboratory of Michael Thomm, University of Regensberg, Germany, shown with permission) for PF1423 (**A**) and PF0891 (**B**) probe templates. Probe diagrams are indicated at the top of each gel image with corresponding protein/DNA mole ratios listed above each lane. The *gdh* promoter is used as a control. Position and expected size of run-off transcripts are indicated.

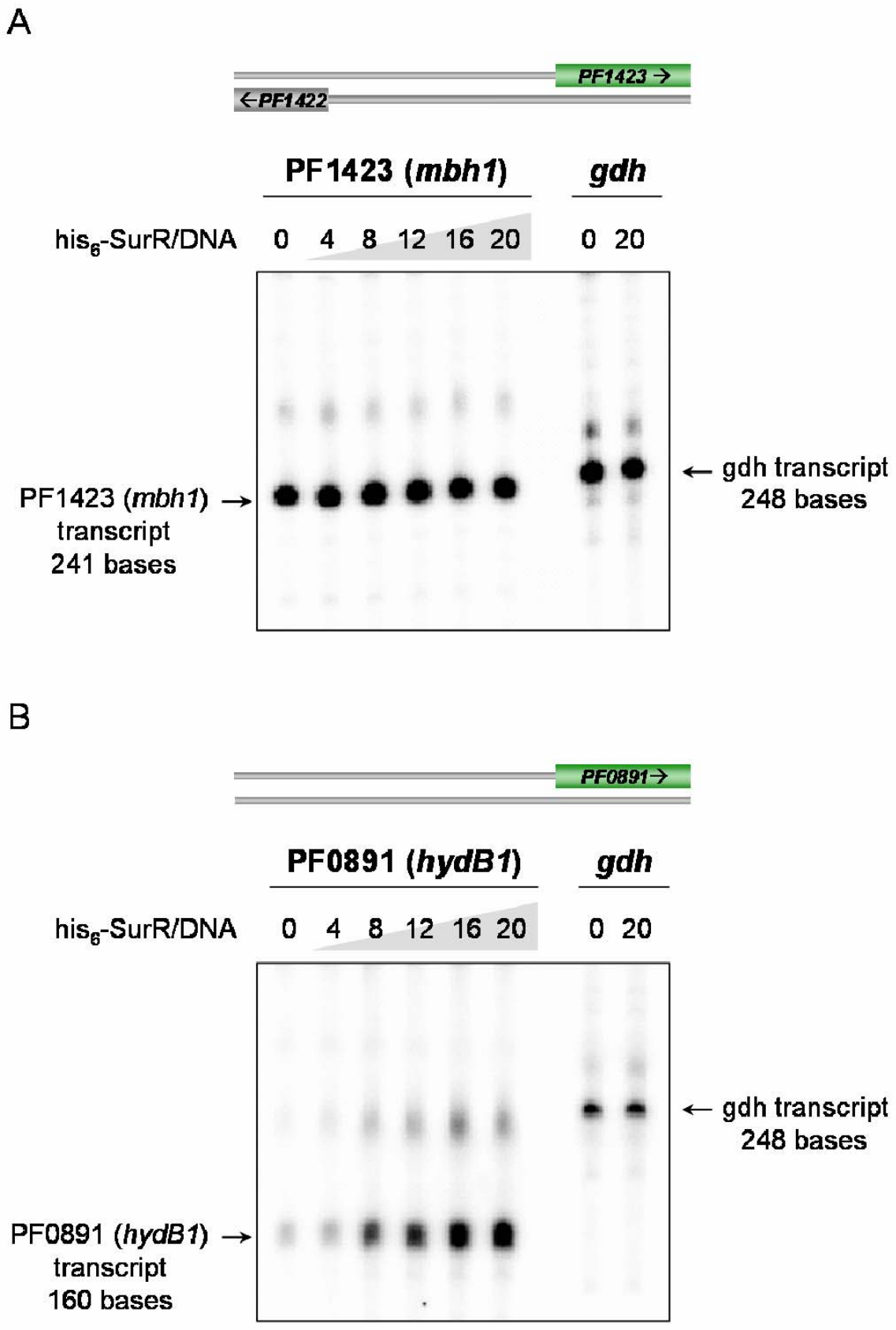

A
PF1423 →
←PF1422
PF1423 (mbh1)
gdh
his6-SurR/DNA
0 4 8 12 16 20
0 20
PF1423 (mbh1) → transcript 241 bases
← gdh transcript 248 bases
B
PF0891→
PF0891 (hydB1)
gdh
his6-SurR/DNA
0 4 8 12 16 20
0 20
← gdh transcript 248 bases
PF0891 (hydB1) → transcript 160 bases

Figure 4.23 SurR is a transcriptional repressor. *In vitro* transcription assays with SurR performed by Annette Keese (unpublished data from the laboratory of Michael Thomm, University of Regensberg, Germany, shown with permission) for PF0094 probe template. Probe diagrams are indicated at the top of the gel image with corresponding protein/DNA mole ratios listed above each lane. The *gdh* promoter is used as a control. Position and expected size of run-off transcripts for both PF0095 and PF0094 are indicated.

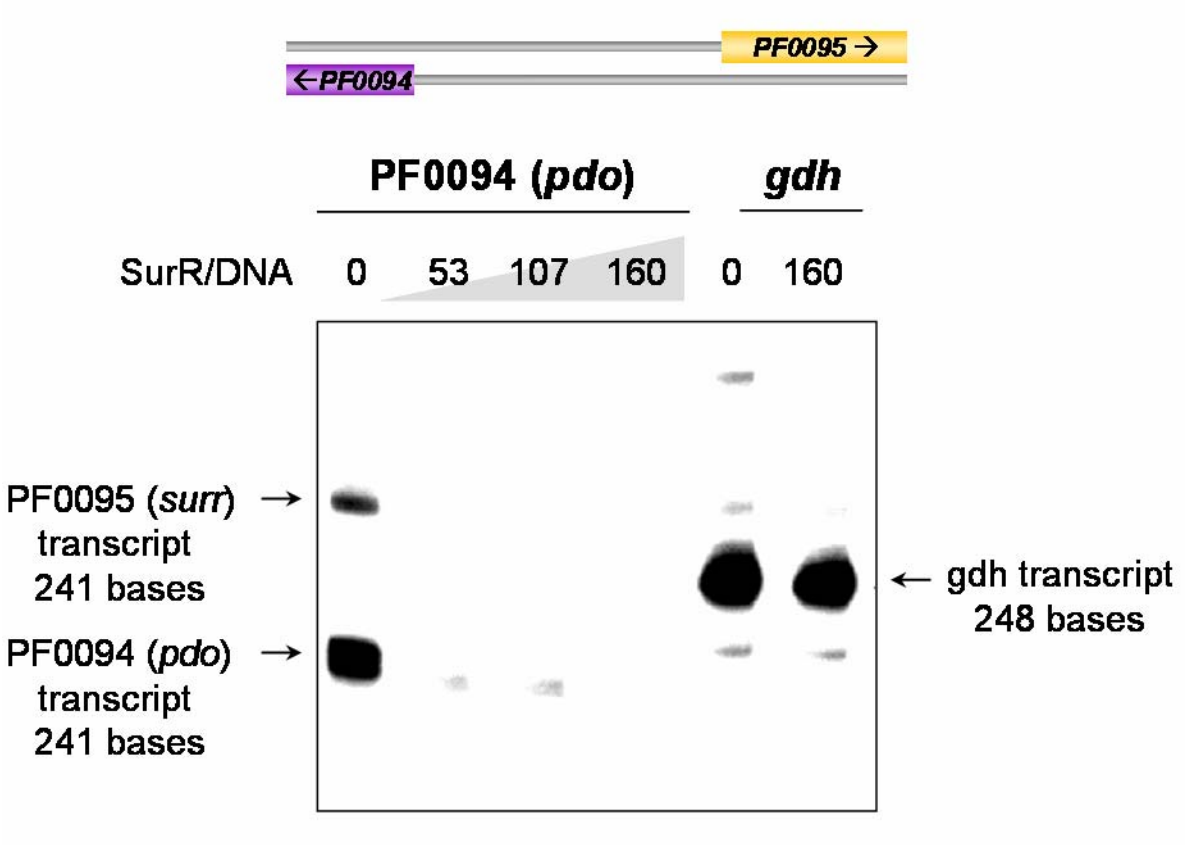
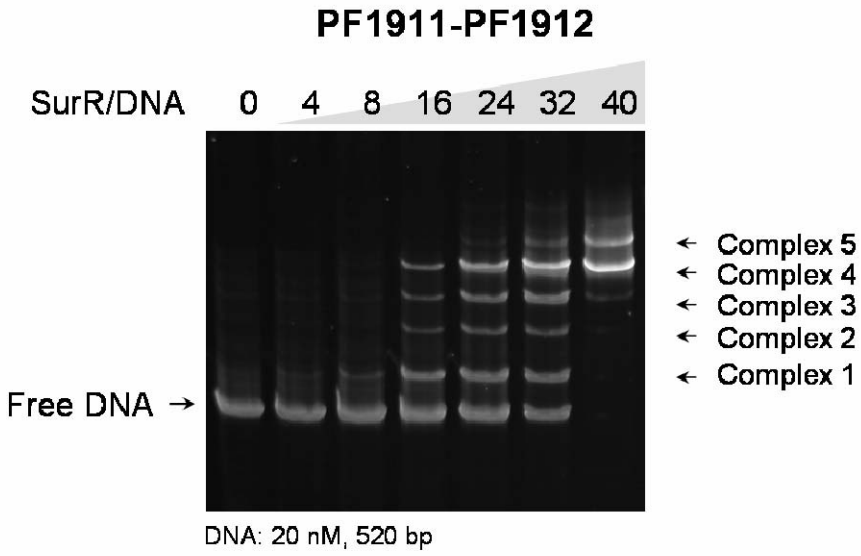

PF0095 →
←PF0094
PF0094 (pdo)
gdh
SurR/DNA
0
53
107
160
0
160
PF0095 (surr)
transcript
241 bases
PF0094 (pdo)
transcript
241 bases
← gdh transcript
248 bases

Figure 4.24 SurR binds specifically to upstream DNA of a non-S^0-regulated ORF. A. Diagram of PF1911 probe used for EMSA and footprinting assay, indicating ORFs present along with intergenic distance. **B.** EMSA with PF1911 probe DNA containing upstream $GTTn_3AAC$ site. Corresponding protein/DNA mole ratios are listed above each lane, and protein-DNA complexes are numbered. **C.** Electropherogram traces of 6FAM-labeled strand for undigested probe and DNase-digested probe with and without protein (protein/DNA mole ratio is indicated). Internal standards GS-500 ROX (peak sizes in bases are labeled) and A+G probe-specific sequence ladder were used to determine position of electropherogram peaks relative to probe sequence. Probe diagram indicates positions of ORFs relative to electropherogram traces. The SurR footprint region is boxed, and the DNA sequence as determined by the ladder is shown expanded below. SELEX consensus (red bracket) and degenerate sites are highlighted in yellow. Asterisks denote bases that are consistent with the $GTTn_3AAC$ consensus. **C.** Relevant DNA sequence of PF1911 probe is shown with SurR DNase footprint site boxed in red. The putative positions of the BRE/TATA elements are indicated.

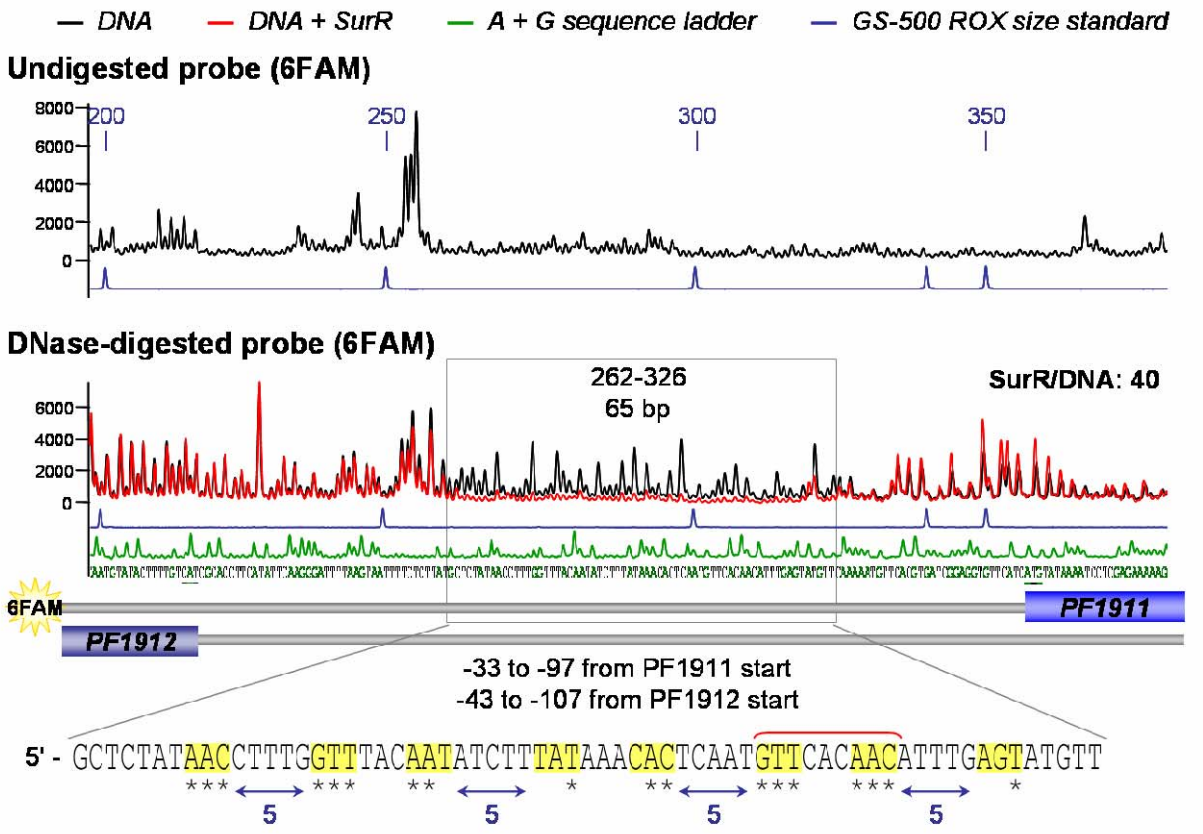
A


139 bp
PF1912 →
← PF1911

B


PF1911-PF1912
SurR/DNA 0 4 8 16 24 32 40
Free DNA →
← Complex 5
← Complex 4
← Complex 3
← Complex 2
← Complex 1
DNA: 20 nM, 520 bp

C


DNA
DNA + SurR
A + G sequence ladder
GS-500 ROX size standard
Undigested probe (6FAM)
8000
6000
4000
2000
0
200
250
300
350
DNase-digested probe (6FAM)
262-326
65 bp
SurR/DNA: 40
6000
4000
2000
0
6FAM
PF1911
PF1912
-33 to -97 from PF1911 start
-43 to -107 from PF1912 start
5' - GCTCTATAACCTTTGGTTTACAATATCTTTATAAACACTCAATGTTCACAACATTTGAGTATGTT
*** *** ** * ** *** *** *
5 5 5 5

D


5' - TCTTATGCTCTATAACCTTTGGTTTACAATATCTTTATAAACACTCAATGTTCACAACATTTGAGTATGTTCAAAAATGTTCACGTGATCGGAGGTGTTCATCATG
BRE/TATA?
PF1911 →

CHAPTER 5

SURR STRUCTURE AND SULFUR-RESPONSIVE SWITCH

5.1 The SurR structure reveals an HTH DNA-binding domain with a nearby disulfide bond

The *in vitro* function of SurR in the transcriptional regulation of S^0-responsive ORFs versus the type of transcriptional regulation that these ORFs undergo *in vivo* in the presence of S^0 as demonstrated by DNA microarray appeared to present a paradox. It is often said that structure leads to function, and therefore the structure of SurR could lead to some clues to synergize the *in vitro* knowledge gained about SurR with the *in vivo* sulfur response data to illuminate its physiological role inside the cell. To this end, a collaboration was formed with Bi-Cheng Wang (University of Georgia) to obtain the X-ray crystal structure of SurR. Concentrated, purified tagless SurR protein was supplied for crystallization trials. Selenomethionine-phasing was used to solve the crystal structure at 2.5 Å resolution. Crystallization trials and structure determination were performed by Hua Yang (unpublished data, structure used with permission).

The crystal structure of SurR reveals a dimeric configuration with an N-terminal domain matching that of a winged HTH (wHTH) DNA-binding fold, as expected from sequence analysis. The structure is displayed in Figure 5.1. The wHTH fold comprises the first ~75 residues; however, the β-strand hairpin which forms the 'wing' is not well-resolved, indicating that it is probably somewhat flexible. Three helices were visible in the SurR HTH domain: the first two N-terminal helices were joined at a sharp turn, and these were connected with the third helix via a random coil. In order to define which helix was most likely the canonical recognition

helix commonly responsible for base-specific contacts, the sequence of SurR was aligned with those of structurally characterized proteins SmtB, CadC and Phr using ClustalW (Fig. 5.2) [125, 126]. SmtB and CadC are two members of the HTH/ArsR protein family which have been well-characterized and for which the position of the recognition helix is known [117]. Phr is a heat shock regulator of *P. furiosus* [61] and has sequence homology to SurR (see Fig. 3.5).

The helices in the SurR dimer structure which corresponds to the SmtB recognition helix are depicted in Fig. 5.3. The striking difference in the SurR structure is that it contains no helix immediately preceding the recognition helix within the HTH motif; in its place, there is simply a long random coil. More interesting still is the presence of a disulfide bond between two closely-spaced cysteine residues, forming a CxxC motif, immediately to the N-terminal side of the coil (Fig. 5.3 B). A secondary structure prediction by PSIPRED [119, 127] based on the SurR sequence alone predicted the presence of four—not three—helices in the N-terminal domain: the random coil observed in the structure was actually predicted to be a helix, although the confidence interval was somewhat lower than for the other predicted helices (see underlined portions of sequence in Fig. 5.3 B). It was hypothesized that the presence or absence of the disulfide bond in the CxxC motif could alter the conformation of the DNA-binding domain and thereby modulate the DNA-binding ability of SurR.

5.2 Oxidation of the SurR CxxC motif affects DNA binding

5.2.1 Aerobically purified SurR is predominantly in the reduced state

All previous experiments that had been performed with recombinant SurR had been carried out aerobically. It had been determined that the reductants DTT and sodium sulfide did not significantly affect the DNA-binding affinity of SurR; therefore, the cysteines in the protein

were assumed to be predominantly in the reduced state. However, with the new information obtained from the structure showing the presence of the disulfide bond in the CxxC motif, it was of great importance to determine the redox state of the cysteines in the aerobically purified protein which had been used in all previous experiments. If the disulfide bond was present in the aerobically purified protein, then this would explain its presence in the structure. Otherwise, if the disulfide form was not predominant in the aerobically purified protein initially, then the crystallization conditions may have favored disulfide-bond formation, or the disulfide form of SurR may have selectively crystallized from a mixture of oxidized and reduced SurR proteins.

In order to determine if aerobically purified SurR was primarily in the oxidized or reduced state, DTNB (Ellman's reagent) was used to test the redox state of the cysteines in the CxxC motif since these were the only cysteines in the protein sequence. DTNB reacts with free thiols resulting in the release of a yellow-colored thiolate in a one-to-one correspondence; this color change is quantifiable when absorbance is measured at 412 nm (Fig. 5.4). Using DTT as a thiol standard and degassed buffer solutions, aerobically purified SurR was found to be approximately 80% in the reduced state according to Ellman's Assay results (Table 5.1); therefore, the crystal structure of SurR containing the disulfide bond did not necessarily represent the form of the protein which was shown to bind to DNA in EMSA and footprinting experiments.

Table 5.1 Ellman's assay results for untreated SurR

Sample	Concentration of thiols (nM)[a]	Number of thiols[b]	Number of thiols with buffer correction[c]
buffer[d]	3.8	-	-
10 nM SurR	20.1	2.0	1.6
25 nM SurR	44.1	1.8	1.6
50 nM SurR	88.7	1.8	1.7

[a] Concentration of thiols was determined using a standard curve of known DTT concentrations, taking into account that DTT has two thiols per molecule.
[b] Number of thiols were determined by dividing the concentration of thiols by the concentration of protein in each sample.
[c] The concentration of thiols in each sample was adjusted to account for the fact that the buffer in each protein sample registered a small amount of thiols.
[d] The buffer consisted of degassed 20 mM HEPES, 200 mM NaCl, pH 7.5.

5.2.2 Diamide oxidizes the SurR CxxC thiols

With this new piece of evidence indicating that the reduced form of the protein containing free thiols in its CxxC motif bound to DNA, it was of interest to determine the DNA-binding affinity of the oxidized form of the protein which was represented in the crystal structure. To this end, a chemical capable of protein thiol oxidation with subsequent disulfide bond formation was required; therefore, the thiol-specific oxidant diazenedicarboxylic acid, otherwise known by the simpler name diamide, was determined to be suitable [98, 99, 128]. Diamide reacts with thiols through its diazene double bond in two steps: the first reaction entails the addition of one thiolate anion to the diazene double bond and is followed by reaction of the bound sulfur with a second thiolate anion, forming a disulfide and leaving behind a hydrazide (Fig. 5.5) [98].

Diamide proved to be a good tool for promoting disulfide formation in SurR, especially since the oxidizable cysteines of SurR were in close proximity (within the CxxC motif), reducing the chances for formation of mixed disulfides with either the diamide reagent or thiols of other SurR proteins. Since diamide has no thiol groups (see Fig. 5.5), it is compatible for use with Ellman's reagent and did not need to be removed from the reaction prior to free thiol quantification. Ellman's assay was therefore used to confirm the effectiveness of diamide in reducing the number of free thiols in SurR. Indeed, the number of free thiols in SurR protein

preparations was reduced to zero after a 10-min treatment at room temperature with 10 mM diamide, and presumably this elimination of free thiols implies the formation of intramolecular disulfide bonds in SurR monomers (Table 5.2).

Table 5.2 Ellman's assay results for diamide-treated SurR

Sample treated with 10 mM diamide [a]	Concentration of thiols (nM) [b]	Number of thiols [b]	Number of thiols with buffer correction [b]
buffer [c]	6.8	-	-
10 nM SurR	5.8	0.6	-0.1
25 nM SurR	6.0	0.2	0.0
50 nM SurR	6.1	0.1	0.0

[a] Each sample contained 10 mM diamide. Protein samples were incubated for 10 min in the presence of diamide prior to free thiol count.
[b] See footnotes for table 5.1.
[c] The buffer consisted of degassed 20 mM HEPES, 200 mM NaCl, pH 7.5, containing 10 mM diamide.

5.2.3 Oxidation of SurR CxxC motif alters its DNA-binding ability

EMSA was used to test the effect of diamide on the DNA-binding affinity of SurR. The *pdo-surr* intergenic probe was shifted with increasing concentrations of SurR in the presence of either the reductant DTT or the thiol-specific oxidant diamide. The results indicated that the presence of diamide affected the DNA-binding ability of SurR (Fig. 5.6). The shift pattern in the presence of DTT was nearly indistinguishable from that with the same probe in the absence of DTT (compare with Fig. 4.4). In comparison, the affect of diamide on the observed pattern of shifting was dramatic. At high concentrations of SurR, the apparent DNA affinity remained the same since in the presence of either DTT or diamide, a complete shift occurred at a protein/DNA mole ratio of approximately 24. At lower protein/DNA ratios of SurR, shifted bands were not observed for the diamide-treated samples, indicating lower affinity of the protein for DNA. In the presence of diamide, the protein-DNA smear indicative of non-specific DNA binding

predominated. The evidence thus far indicated that diamide-oxidized SurR bound to the DNA in a nonspecific manner; however, the possibility remained that the chemical additive diamide itself interfered with either the DNA-binding of SurR or the movement of the protein-DNA complexes through the gel.

In order to show it was the oxidizing capacity of diamide alone that was affecting the DNA-binding ability of SurR, EMSA was carried out in the presence of diamide with a mutant version of SurR in which both cysteines in the CxxC motif were replaced with alanine residues. These C23A and C26A mutations altered the CxxC motif to 'AxxA', and therefore, this recombinant protein mutant is referred to as AxxA-SurR. The mutant protein was incapable of forming a disulfide bond since both cysteines had been removed, causing it to be perpetually in the 'reduced' state. EMSA with *pdo-surr* intergenic DNA comparing SurR with AxxA-SurR in the presence of diamide revealed that it is indeed the oxidation of the cysteines which is responsible for the change in DNA-binding ability; the mutant protein displayed a characteristic shift, while SurR exhibited weaker binding reminiscent of the non-specific SurR DNA-binding mode (Fig. 5.7). In fact, the mutant displayed an even higher affinity for the DNA, almost completely shifting it at a protein/DNA mole ratio of 16 (compare AxxA-SurR panel in Fig. 5.7 to DTT panel in Fig. 5.6). This is not entirely surprising considering that the aerobically purified SurR protein is only around 80% in the reduced state; if oxidation reduces DNA-binding affinity, then it would be expected that the AxxA-SurR mutant which cannot be oxidized would display a higher overall DNA affinity than the aerobically purified SurR.

5.3 Colloidal sulfur reversibly alters the DNA-binding affinity of SurR

5.3.1 Colloidal sulfur affects SurR DNA-binding ability in EMSA

The change in DNA-binding ability of SurR produced by oxidation with diamide suggested the possibility that this effect might represent an *in vivo* mechanism for regulation of SurR activity. If this were the case, then there had to be a naturally occurring oxidant present within *P. furiosus* cells which could produce the same effect. Since SurR had already been shown to be associated with ORFs that were involved in the response of *P. furiosus* to S^0, oxidizing sulfur species were a natural choice to test as oxidant effectors of SurR. Inorganic sulfur is not water-soluble; therefore, in order to test the effect of sulfur species on SurR in EMSA, a colloidal sulfur suspension was used. The colloidal sulfur was a kind gift of Gerrit Schut (Michael Adams Laboratory, University of Georgia) and was made by aerobic dilution of an anaerobic polysulfide solution into EMSA reaction buffer. The resulting colloidal sulfur suspension was used immediately in EMSA before the suspended sulfur particles precipitated out of solution.

A comparison of the effect of colloidal sulfur on AxxA-SurR versus SurR in EMSA with *pdo-surr* intergenic DNA is dramatic; while it does not affect the shifting pattern of AxxA-SurR, colloidal sulfur is effective at completely eliminating the specific distinct banding pattern for SurR (Fig. 5.8). A nearly complete shift of the DNA is observed at the highest protein/DNA mole ratio tested (32); however, the smeared shift appears to be a result of non-specific DNA binding due to its very low electrophoretic mobility and undefined nature. Although an approximate concentration of 10 mM colloidal sulfur was used in these experiments, a titration of colloidal sulfur with SurR and the same probe DNA in EMSA showed that colloidal sulfur

concentrations between 100 and 500 μM are sufficient to convert the specific banding pattern to a smear (Fig. 5.9).

5.3.2 Colloidal sulfur abolishes SurR DNA footprints

In order to establish that the DNA-binding specificity of SurR was indeed eliminated as a result of oxidation by colloidal sulfur, fluorescence-detected DNase I footprinting of SurR on both the PF1423 (*mbh1*) and PF0094 (*pdo*) probes was performed in the presence and absence of colloidal sulfur. For both of the *mbh1* and *pdo-surr* promoter regions, the SurR footprints which are readily visible at protein/DNA mole ratios of 30 under non-oxidizing conditions completely disappeared with the addition of 1 mM colloidal sulfur (Fig. 5.10). This result confirms that, while the DNA is still completely shifted by SurR at a protein/DNA ratio of 32 in the presence of colloidal sulfur (Fig. 5.8), the affinity of SurR for DNA is reduced to non-specific association and can no longer bind with sequence-specificity to its DNA recognition sites. This implies that oxidation of the cysteines in the CxxC motif registers a conformational change in the DNA-binding wHTH domain which prevents SurR from recognizing and binding specifically to DNA.

It is unclear which chemical components in the colloidal sulfur suspension are effecting the oxidation of SurR to prevent specific recognition of its DNA recognition sequence; however, it seems clear that this oxidant is most likely physiologically relevant, considering the relationship of SurR to ORFs involved in sulfur response. The colloidal sulfur used was formed from the natural oxidation of a polysulfide solution in a neutral aerobic buffer. According to a recent publication studying the oxidation of polysulfide in aqueous phosphate buffer solutions, the products of polysulfide oxidation are almost exclusively thiosulfate and sulfur at a pH ranging from 9 to 13.5 and at temperatures from 23 to 40°C [129]. The ‘biologically produced’ sulfur is colloidal and differs from inorganic sulfur in that it is considerably more soluble in

aqueous solutions and tends to form S_8 rings or ionic chains with the main species being S_6^{2-}, S_5^{2-} and S_4^{2-} [130]. The mechanism of polysulfide oxidation is unknown; however, it has been suggested that radical polysulfide anions such as $S_2^{\bullet-}$ and $S_3^{\bullet-}$ are the reactive species responsible for the oxidation [131, 132]. Whether or not these are the same species responsible for SurR oxidation is not known at this point; however, it would be interesting if the oxidation of the SurR CxxC motif resulted not in the formation of a disulfide bond, but of a polysulfide bond, which would presumably force an even more radical conformational change of the HTH domain.

5.3.3 SurR oxidation by colloidal sulfur is reversible

If the oxidation of SurR represents a physiologically relevant "switch" for regulation of SurR activity, then the oxidation must be reversible. In other words, the DNA-binding affinity of SurR must be able to be restored with reducing agents following SurR oxidation. EMSA was used to test the reversibility of SurR oxidation by colloidal sulfur with the following reductants: cysteine, dithiothreitol, dithionite and sulfide. SurR was first incubated with the *pdo-surr* intergenic DNA probe for 20 min at 70°C in the presence of ~10 mM colloidal sulfur; aliquots of this solution were then introduced into anaerobic vials containing each reductant such that the final concentration of the reductants in each reaction was 20 mM. These vials were then incubated another 5 min at 70°C prior to the samples being loaded and run aerobically on a non-denaturing gel. The results can be seen in Fig. 5.11, where the only reductant capable of reversing the oxidation effect of colloidal sulfur is dithiothreitol (DTT). This result showed that the oxidation of SurR by colloidal sulfur appears to be completely reversible depending on the redox conditions, demonstrating that the CxxC motif represents a redox switch capable of modulating the DNA-binding activity of SurR *in vitro*.

5.4 Modulation of SurR DNA-binding activity is due to a conformational change

5.4.1 Quaternary structure of SurR does not change upon oxidation

Since oxidation of SurR resulted in such a dramatic effect on its DNA-binding specificity, the possibility that the quaternary structure of SurR changes upon oxidation was investigated. Analytical gel filtration was used to determine the oligomeric state of untreated SurR, SurR treated with diamide or colloidal sulfur, as well as AxxA-SurR, which appears to behave like the reduced form of the protein in EMSAs. 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 set of standards run individually through the column with the same buffer. The resulting traces for each sample can be seen in Figure 5.12, and the corresponding calculated molecular weights are shown in Table 5.3.

Each of the SurR samples displayed a small peak that eluted at the column void volume, as determined by blue dextran. The AxxA-SurR sample did not display this void-volume peak because this protein was purified previously using gel filtration as the final step; therefore, the void-volume peak observed in the SurR samples is most likely due to protein aggregates. A few low-intensity peaks of less than 3 kDa and of unknown composition were apparent in each of the oxidant-treated samples. The predominant peak in all four protein samples was of similar intensity (around 50 mAu) and eluted at essentially the same volume (range 9.21 to 9.41 mL). The calculated molecular weights of each peak are listed in Table 5.3, and together, the average is approximately 62 kDa. According to the SurR amino acid sequence, the molecular weights for a SurR monomer, dimer and trimer are 26,899, 53,798 and 80,697 Da, respectively. These results suggest that SurR is a dimer in solution, for both oxidized and reduced forms of the protein, in

agreement with the crystal structure of the dimeric form of SurR. However, since there is no change in quaternary structure upon oxidation, there must be a conformational change in the protein dimer caused by the CxxC redox switch which in turn modulates the DNA-binding activity of SurR.

Table 5.3 Molecular weight of SurR samples as determined by gel filtration

Sample	Peak elution volume (mL)	Molecular weight (g/mol)[a]
SurR, untreated	9.41	**61,500**
AxxA-SurR, untreated	9.41	**61,500**
SurR + diamide (10 mM)	9.38	**62,400**
	16.16	2,600
	18.07	1,100
	19.35	600
SurR + colloidal sulfur (2 mM)	9.21	**67,500**
	16.43	2,300
	17.64	1,300

[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. Molecular weights of major peaks are listed in bold.

5.4.2 The structure of mutant AxxA-SurR hints at the mechanism of the switch

Reduction or oxidation of SurR is not accompanied by any change in quaternary configuration; therefore, a structural change in the dimer itself must be responsible for modulation of DNA-binding activity. The previously determined structure of SurR displayed the oxidized form, having a disulfide bond at the CxxC motif, and therefore was most likely not the form capable of sequence specific DNA binding. In order to determine the structural changes in SurR in the absence of a disulfide bond at the CxxC motif, crystallization and structure determination of the mutant AxxA-SurR was undertaken. Purified protein was once again

supplied to Hua Yang who performed crystallization trials and structure determination at 2.8 Å resolution (unpublished data, structure used with permission).

Comparison of the mutant dimer structure with that of SurR revealed only minor differences (overall RMS deviation of 0.815 Å) between the presumed 'reduced'[1] and oxidized states of SurR, and curiously, only in chain A of the dimer and not chain B (Fig. 5.13). As can be seen in Figure 5.14, the 'reduced' and oxidized SurR monomers (chain A) overlay nicely except at the region beginning with the CxxC motif and extending through to the start of the recognition helix (Fig. 5.14 A, B) . Whereas the C_α atoms of Ala23 and Cys23 are only separated by 2 Å between the 'reduced' and oxidized structures, those of Cys26 and Ala26 are separated by a distance of 8.1 Å. (Fig. 5.14 C). Correspondingly, the Cys C_α atoms of oxidized SurR are 5.2 Å apart, while in 'reduced' SurR the Ala C_α atoms are 8.8 Å apart (chain A). In one monomer of the 'reduced' state, the coil separating helix 2 from the recognition helix is repositioned slightly away from the protein core and becomes more helical. Because of this shift, chain A of the AxxA-SurR structure aligns better with the Phr structure than does either chain of the oxidized SurR dimer (Fig. 5.15). Since Phr contains a canonical HTH motif with a helix at the position where SurR has a random coil, this conformational change from oxidized to reduced states could indicate a shift toward a form that more readily binds to DNA in a specific manner.

Along with this shift to a more canonical HTH fold in one monomer, the amino acid side-chains which have altered positions between oxidized SurR and 'reduced' AxxA-SurR chain A structures may give some clues to the mechanism of the CxxC redox switch. Residues 23 through 36 in the coil region show the most change between the two structures, and these are shown in Figure 5.16. As can be seen, some residue side chains move significantly, with several

[1] The AxxA-SurR structure is assumed to be equivalent to the structure of reduced SurR and will therefore be referred to as the 'reduced' form; however, the single quotes will be used to designate this assumption.

becoming more solvent-exposed in the 'reduced' form. Four residues in particular appear to be more solvent-exposed and positioned closer to the recognition helix in the 'reduced' form: Ser32, Ser29, Phe28, and Tyr 27 (Fig. 5.16). Both Ser29 and Ser32 may contribute to sequence specific DNA binding as the amino acid serine has a somewhat higher propensity to be involved in protein-DNA contacts, particularly with the DNA backbone [133, 134]. Phe28 flips toward the recognition helix and the wing of the winged HTH motif, and Tyr27 moves out from a position close to the protein core to a region near the wing. Perhaps the shift in the position of the Phe and Tyr residues could change or stabilize the position of the wing so that it can make necessary contacts with the DNA. For Phr, the wing appears to play an important role in sequence specific DNA recognition [62]; however, DNA binding involves three arginine residues in the hairpin loop, none of which are conserved in SurR. Nevertheless, there are some charged residues, including one Arg, in the wing hairpin of SurR which could play important roles in DNA binding, as charged residues in this region in particular can be important in DNA binding by wHTH proteins [32].

In order to identify residues which might play a role in DNA contacts, detailed sequence analysis of the N-terminal 75 amino acids which constitute the ArsR/HTH domain was performed. A BLAST of the SurR N-terminal sequence (residues 1 to 75) did not reveal many close homologs, so a second Position-Specific Iterated BLAST (PSI-BLAST) [81, 82] was carried out using the first BLAST hits with e-values better than 0.001. The sequences from the second round which had e-values higher than 10^{-10} were aligned to identify conserved residues. Interestingly, these 25 sequences were all from organisms of the Euryarchaeota and did not include Phr. The alignment of SurR with these sequences can be seen in Fig. 5.17. The first three sequences listed under SurR represent close homologs from the Thermococcales (*P. abyssi*, *P.*

horikoshii, *T. kodakaraensis*), and only these have the conserved CxxC motif. The rest are more distantly related; however, there is still significant residue conservation.

There are eight residues throughout the domain which are strictly conserved among these 26 protein sequences; these and several of the other highly conserved residues are structurally internal to the HTH motif, undoubtedly playing important roles in stabilizing the HTH fold. Notably, in the recognition helix, highly conserved residues Ala41, Val42, His45 and Leu46 occur on the underside of the helix while the charged residues of low conservation Thr40, Lys44, Lys47 and Arg51 lie on the outer surface, and are probably involved in making DNA contacts. Arg and Lys have the highest likelihood of base-specific interactions in protein-DNA complexes, most often forming hydrogen-bonds with guanine [135]. Threonine is generally involved in making contacts with the DNA backbone due to its shorter side chain [134]. Finally, the three Lys residues in the second strand of the wing, one of which is conserved (Lys41), may also be important in stabilization of the protein-DNA structure through contacts with the DNA backbone. It should be noted that these numerous non-specific contacts play a critical role in stable protein-DNA complex formation; indeed most interactions of the protein with the DNA involve the backbone, and in many cases there are a surprisingly small number of base-specific contacts [134].

The structure of the 'reduced' form of SurR also lends some insight to the mechanics of the redox switch. It is interesting that for redox sensitive proteins with CxxC motifs, it is common to find an α-helix immediately following the motif [136], and this appears to be the case in SurR, at least for the 'reduced' form (chain A) where the random coil becomes distinctly more helical. Another aspect which might promote disulfide bond formation is the presence of a

histidine residue (His171) near the N-terminal cysteine (Cys23) which could stabilize the thiolate form giving the cysteine residue a higher propensity to be oxidized [136, 137].

It is unclear why there are slightly different conformations adopted by each monomer in the AxxA-SurR structure; however, it is suggestive of an allosteric mechanism for SurR: perhaps the binding of SurR to its cognate DNA begins with one monomer binding to a palindromic half-site, inducing a conformational change which registers at the other monomer HTH fold, and in turn allows it to bind specifically to the other palindrome half site. Though the mechanism of the redox switch is not fully understood, the structural differences between the oxidized and 'reduced' forms of SurR confirm the switch hypothesis and validate that oxidation of the cysteines in the CxxC motif plays an important role in modulating the DNA-binding activity of SurR.

Figure 5.1 Structure of SurR. A. Dimer structure of SurR with monomers in blue and red. **B.** view from the top **C**. View from the side. Structure images were produced using PyMOL [124].

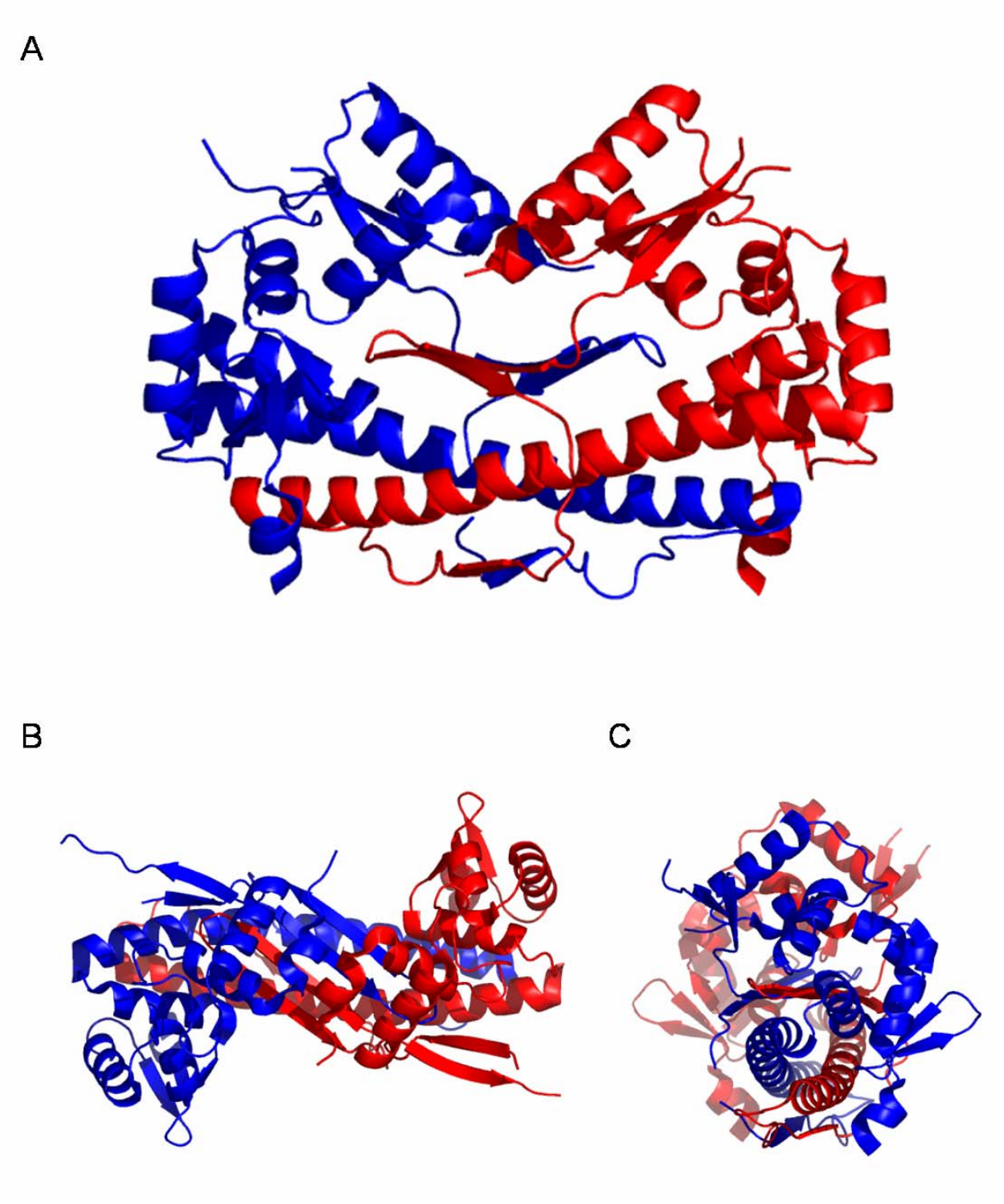

A
B
C

Figure 5.2 Placement of the recognition helix in the N-terminal wHTH domain. A sequence pile-up (ClustalW, [125, 126]) of SurR N-terminal region (residues 1-75) is shown with that of Phr, SmtB, and CadC (top). Secondary structure elements known from protein structures are shown with colored font: helices one and two are shown in blue, the recognition helix in red, and the wing in yellow. Residues that are conserved are highlighted in yellow. The structure of the SurR wHTH domain (residues 1-75) is shown (bottom) using the same color code with a diagram of the fold to the right. Structure image was produced using PyMOL [124].

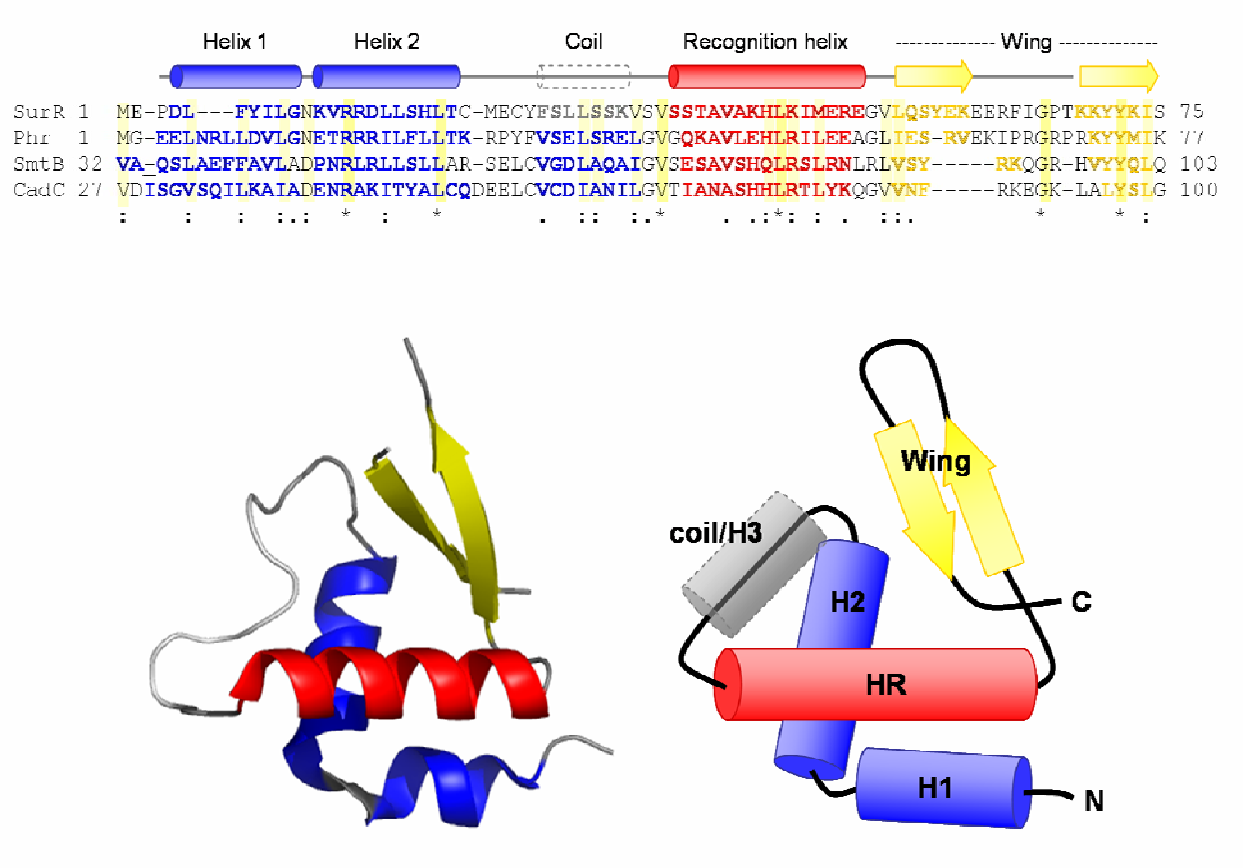
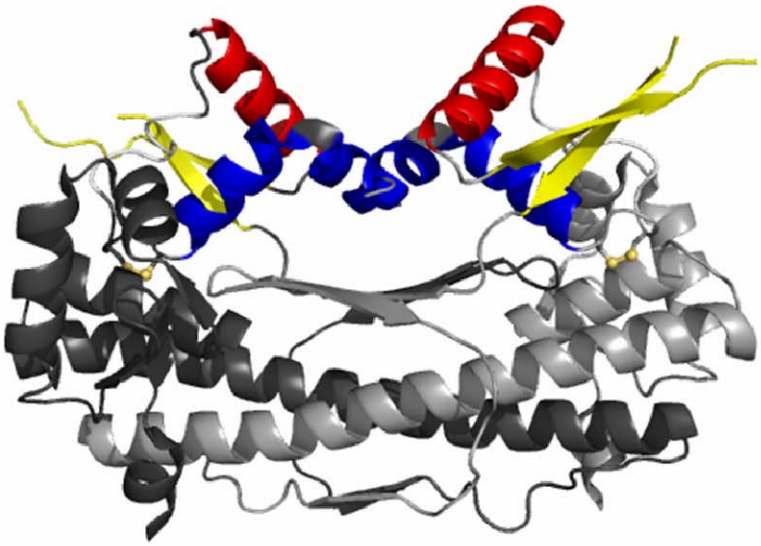
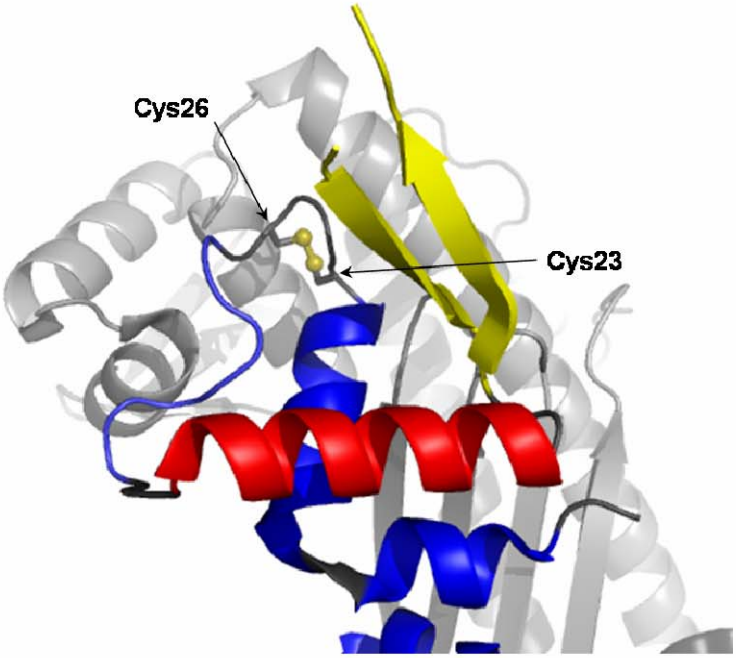

Helix 1
Helix 2
Coil
Recognition helix
Wing
SurR 1 ME-PDL---FYILGNKVRRDLLSHLTC-MECYFSLLSSKVSVSSTAVAKHLKIMEREGVLQSYEKEERFIGPTKKYYKIS 75
Phr 1 MG-EELNRLLDVLGNETRRRILFLLTK-RPYFVSELSRELGVGQKAVLEHLRILEEAGLIES-RVEKIPRGRPRKYYMIK 77
SmtB 32 VA-QSLAEFFAVLADPNRLRLLSLLAR-SELCVGDLAQAIGVSESAVSHQLRSLRNLRLVSY-----RKQGR-HVYYQLQ 103
CadC 27 VDISGVSQILKAIADENRAKITYALCQDEELCVCDIANILGVTIANASHHLRTLYKQGVVNF-----RKEGK-LALYSLG 100
Wing
coil/H3
H2
C
HR
H1
N

Figure 5.3 The SurR wHTH contains a disulfide bond within a CxxC motif. A. Structure of the SurR dimer with wHTH elements highlighted: helices one and two are shown in blue, the recognition helix in red, and the wing in yellow. The disulfide bond sulfurs are colored gold. **B.** Top view of one wHTH domain showing the disulfide bond between Cys23 and Cys26. The corresponding sequence of residues 1-75 is shown below with secondary structure prediction of helices underlined (PSIPRED [118, 119]). Structure images were produced using PyMOL [124].

A



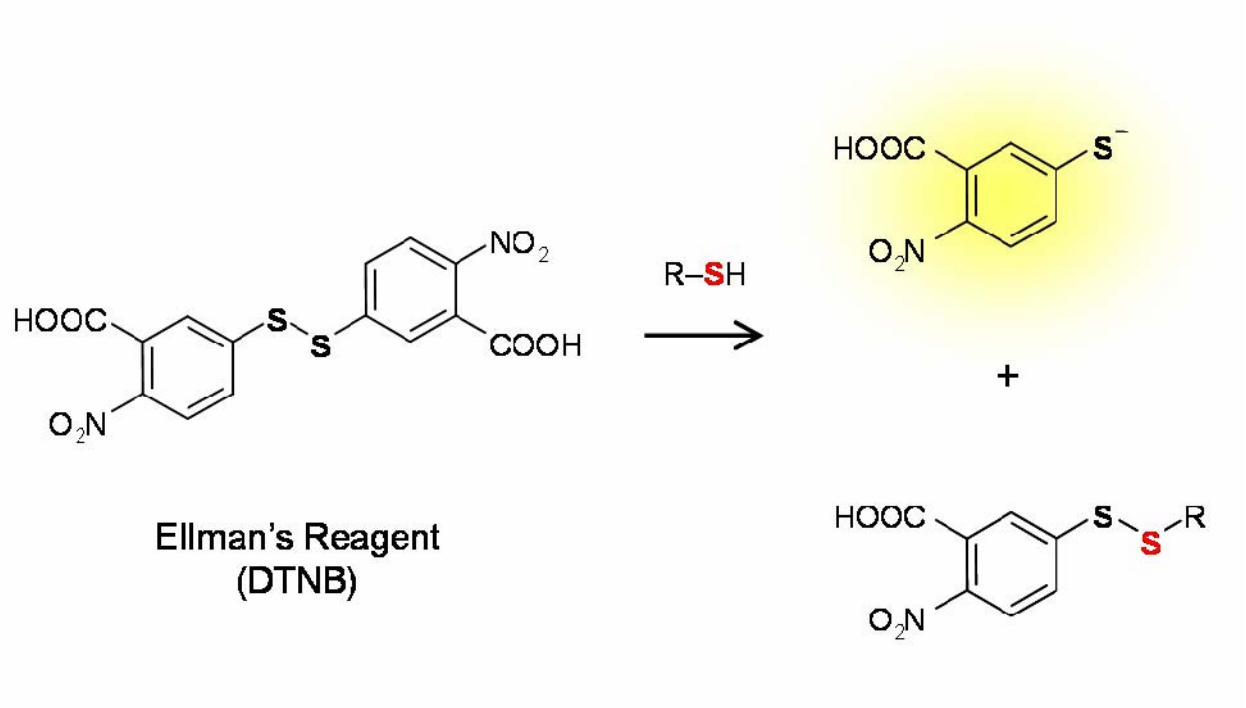
B



1 75

MEPDLFYILGNKVRRDLLSHLTCMECYFSLLSSKVSVSSTAVAKHLKIMEREGVLQSYEKEERFIGPTKKYYKIS

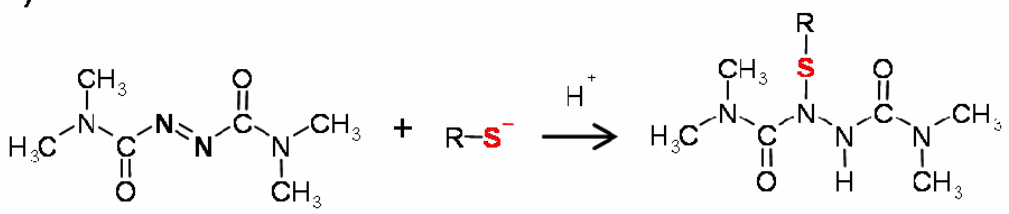
Figure 5.4 Ellman's assay for quantification of free thiols. Ellman's reagent reacts with free thiols (-SH) in proteins to produce a yellow chromophore which is detectable at 412 nm.



Ellman's Reagent
(DTNB)

Figure 5.5 Diamide is a thiol-specific oxidant that promotes disulfide-bond formation. Diamide reacts with cysteine thiolates in a two-step mechanism: (1) the first thiolate anion is added to the diazene double bond and (2) the second thiolate anion reacts with the bound sulfur to form a disulfide bond and leave behind hydrazide [98].

1)



2)

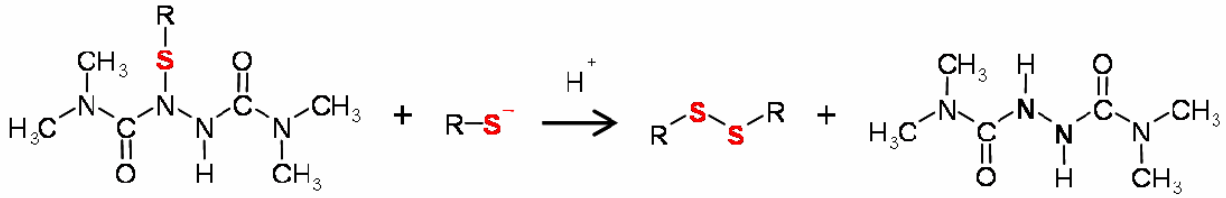


Figure 5.6 Diamide affects DNA binding of SurR. The effect of DTT versus diamide on the DNA-binding ability of SurR with *pdo-surr* intergenic DNA probe was determined via EMSA. The reductant/oxidant is indicated at the top of each gel image with corresponding protein/DNA mole ratios listed above each lane. *Pdo-surr* intergenic DNA probe (50 nM) was incubated with protein in buffer (20 mM HEPES pH 7.5, 200 mM KCl, 5% glycerol, 1 mM EDTA) plus DTT or diamide for 20 min at 70°C. Gel was stained with SYBR Green I nucleic acid gel stain.

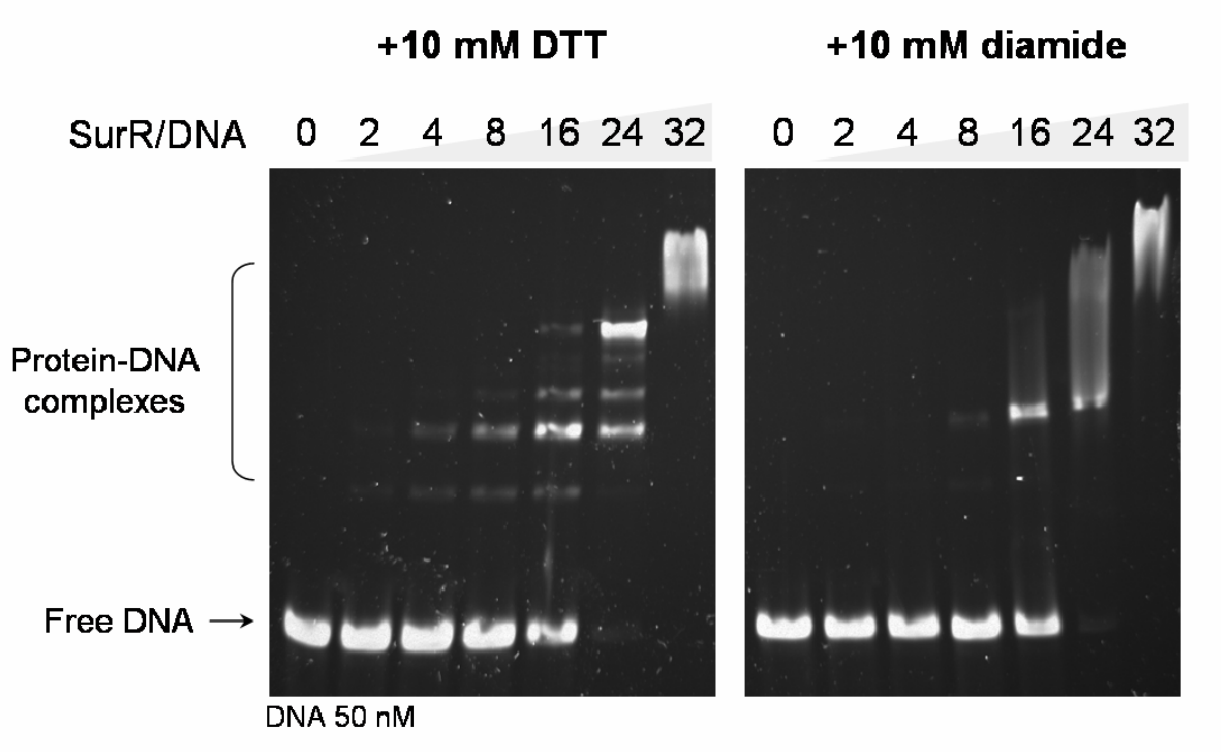

+10 mM DTT
+10 mM diamide
SurR/DNA
0 2 4 8 16 24 32
0 2 4 8 16 24 32
Protein-DNA complexes
Free DNA
DNA 50 nM

Figure 5.7 Diamide oxidation of the CxxC motif affects SurR DNA binding. EMSA of SurR versus mutant AxxA-SurR (in which both cysteines had been replaced with alanines) with *pdo-surr* intergenic DNA probe in the presence of 10 mM diamide shows that the effect of diamide on SurR DNA-binding affinity is at the CxxC motif. The protein used is indicated at the top of each gel image with corresponding protein/DNA mole ratios listed above each lane. *Pdo-surr* intergenic DNA probe (50 nM) was incubated with protein in buffer (20 mM HEPES pH 7.5, 200 mM KCl, 5% glycerol, 1 mM EDTA) plus diamide for 20 min at 70°C. Gel was stained with SYBR Green I nucleic acid gel stain.

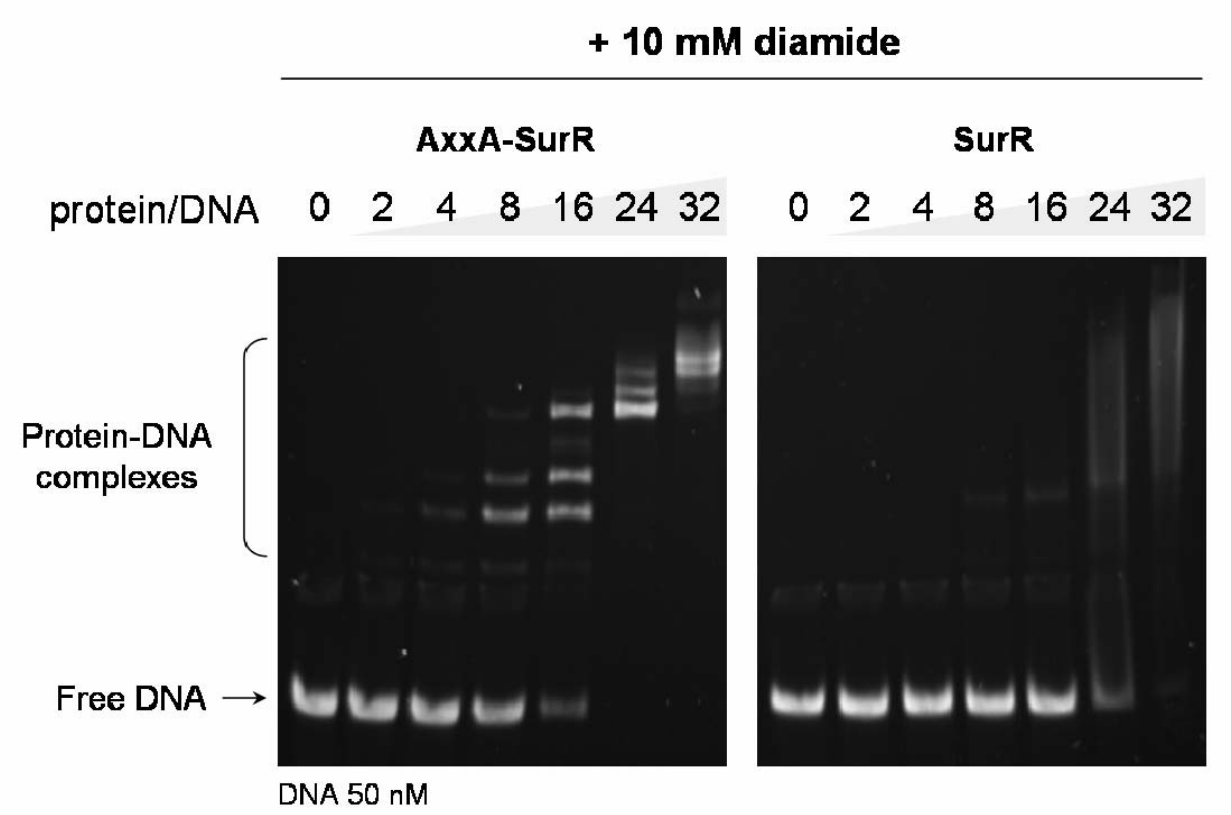

+ 10 mM diamide
AxxA-SurR
SurR
protein/DNA
0 2 4 8 16 24 32
0 2 4 8 16 24 32
Protein-DNA complexes
Free DNA
DNA 50 nM

Figure 5.8 Colloidal sulfur oxidation of the CxxC motif affects SurR DNA binding. EMSA of SurR versus mutant AxxA-SurR with *pdo-surr* intergenic DNA probe in the presence of ~10 mM colloidal sulfur shows that colloidal sulfur oxidizes SurR at the CxxC motif and affects DNA binding. The protein used is indicated at the top of each gel image with corresponding protein/DNA mole ratios listed above each lane. *Pdo-surr* intergenic DNA probe (50 nM) was incubated with protein in buffer (20 mM HEPES pH 7.5, 200 mM KCl, 5% glycerol, 1 mM EDTA) plus colloidal sulfur for 20 min at 70°C. Gel was stained with SYBR Green I nucleic acid gel stain.

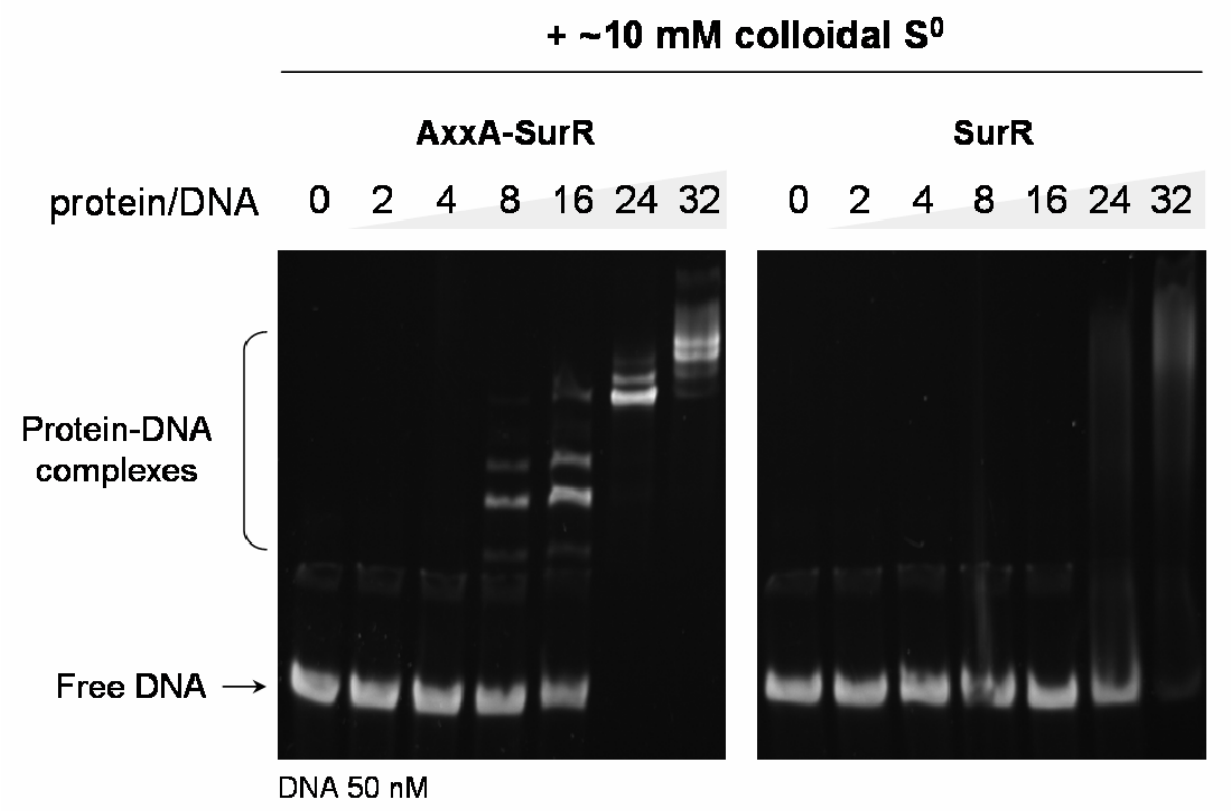

+ ~10 mM colloidal S0
AxxA-SurR
SurR
protein/DNA
0 2 4 8 16 24 32
0 2 4 8 16 24 32
Protein-DNA complexes
Free DNA
DNA 50 nM

Figure 5.9 Colloidal sulfur between 100-500 µM is sufficient to reduce SurR DNA binding. An EMSA of SurR with *pdo-surr* intergenic probe DNA is shown with a titration of colloidal sulfur. The first five lanes show the typical shift observed with increasing concentrations of SurR in the absence of any oxidant. The next nine lanes show the effect of increasing concentrations of colloidal sulfur while maintaining the protein concentration constant. Final concentrations of colloidal sulfur in each reaction are listed above each lane along with corresponding protein/DNA mole ratios. *Pdo-surr* intergenic DNA probe (50 nM) was incubated with protein in buffer (20 mM HEPES pH 7.5, 200 mM KCl, 5% glycerol, 1 mM EDTA) plus various concentrations of colloidal sulfur for 20 min at 70°C. Gel was stained with SYBR Green I nucleic acid gel stain. It is noteworthy that the colloidal sulfur dilution added to the sample in the final lane still had a distinct yellow color characteristic of the presence of polysulfide. Also note that the reactions used for the colloidal sulfur titration were aliquoted from a master mix of buffer, DNA and protein, while the non-oxidized samples each had protein added to them separately.

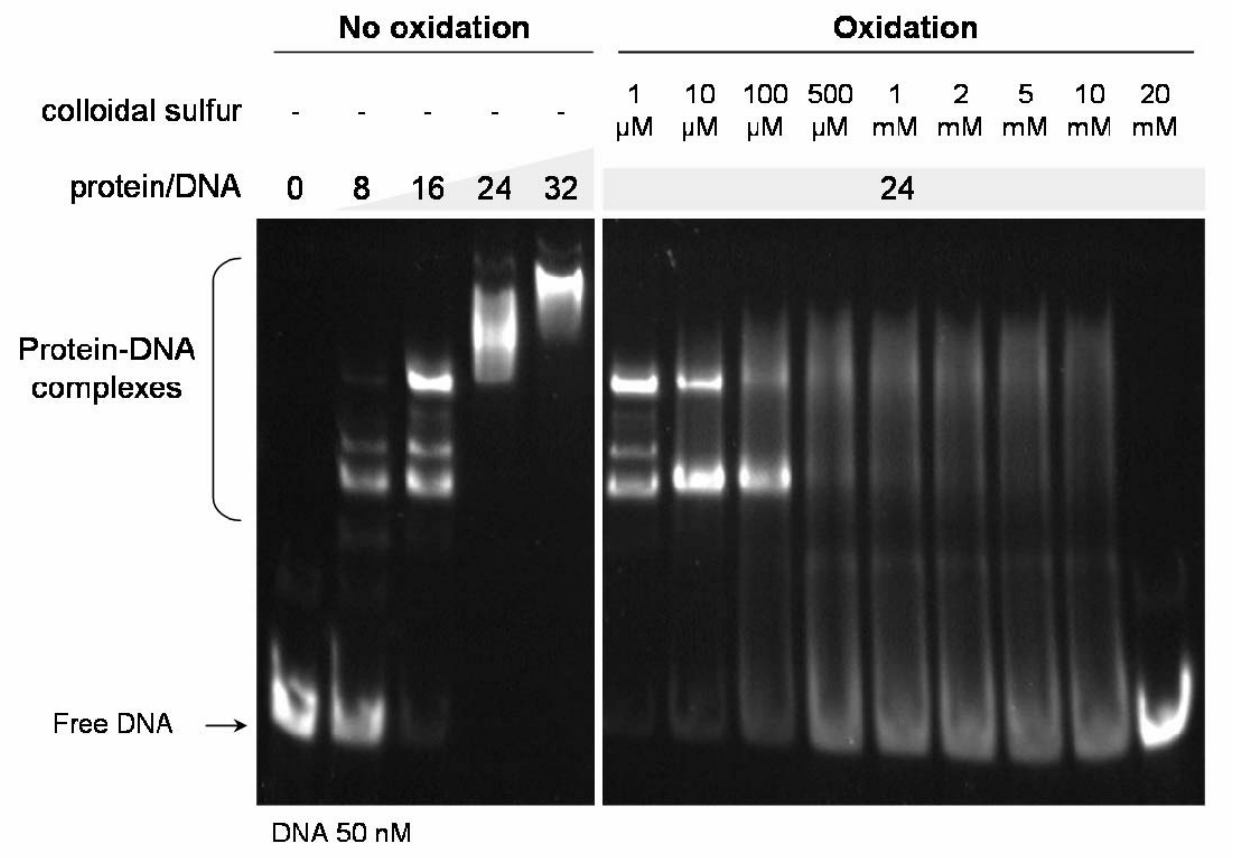

No oxidation
Oxidation
colloidal sulfur
- - - - -
1 µM
10 µM
100 µM
500 µM
1 mM
2 mM
5 mM
10 mM
20 mM
protein/DNA
0 8 16 24 32
24
Protein-DNA complexes
Free DNA
DNA 50 nM

Figure 5.10 Colloidal sulfur abolishes the SurR footprint on PF1423 and PF0094 probe DNA. Electropherogram traces of HEX-labeled strand DNase-digested PF1423 (**A**) and PF0094 (**B**) probes with and without protein for untreated samples and those treated with ~1 mM colloidal sulfur, as indicated. Protein/DNA mole ratio is indicated. Peak sizes of internal size standard GS-500 ROX are indicated below each trace with a probe diagram which indicates positions of ORFs relative to electropherogram. The footprint regions are boxed in yellow.

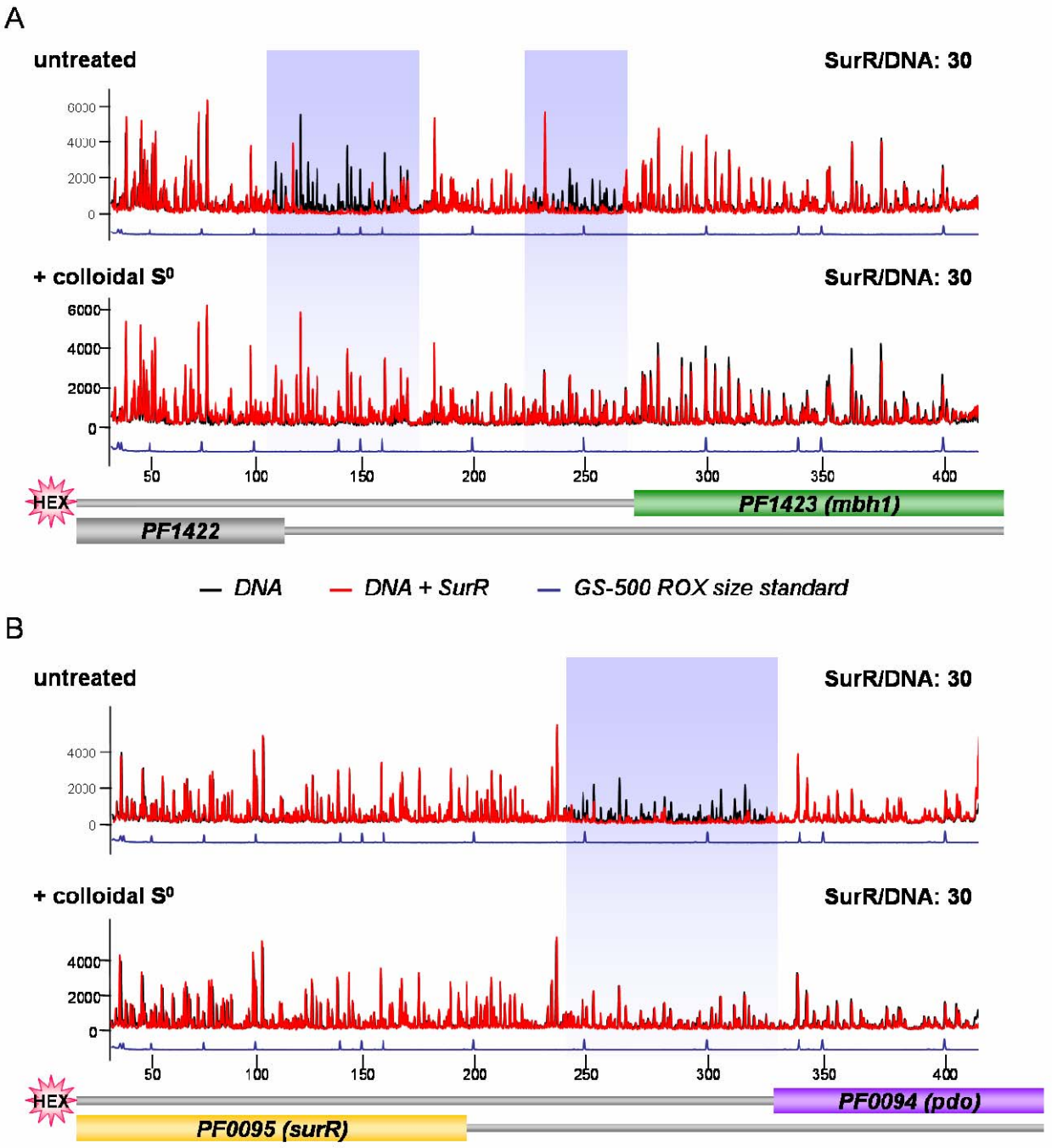
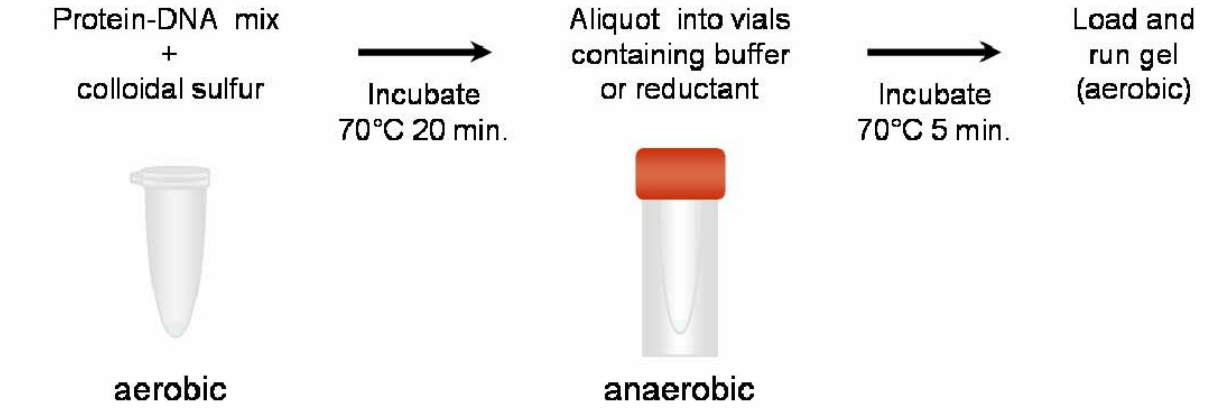

A
untreated
SurR/DNA: 30
6000
4000
2000
0
+ colloidal S⁰
SurR/DNA: 30
6000
4000
2000
0
50
100
150
200
250
300
350
400
HEX
PF1423 (mbh1)
PF1422
DNA
DNA + SurR
GS-500 ROX size standard
B
untreated
SurR/DNA: 30
4000
2000
0
+ colloidal S⁰
SurR/DNA: 30
4000
2000
0
50
100
150
200
250
300
350
400
HEX
PF0094 (pdo)
PF0095 (surR)

Figure 5.11 Effect of SurR oxidation with colloidal sulfur is reversible. An EMSA of SurR with *pdo-surr* intergenic probe DNA is shown with colloidal sulfur treatment followed by incubation with one of four reductants. **A.** The scheme for the oxidation/reduction experiment is as follows: the protein-DNA mix is incubated with colloidal sulfur, and the reaction is then transferred into an anaerobic vial containing a reductant followed by an additional incubation. **B.** EMSA showing the effects of reduction after oxidation of colloidal sulfur. The reductant used is indicated at the top of each lane along with corresponding protein/DNA mole ratios. Gel was stained with SYBR Green I nucleic acid gel stain.

A


Protein-DNA mix
+
colloidal sulfur
Incubate
70°C 20 min.
Aliquot into vials
containing buffer
or reductant
Incubate
70°C 5 min.
Load and
run gel
(aerobic)
aerobic
anaerobic

B

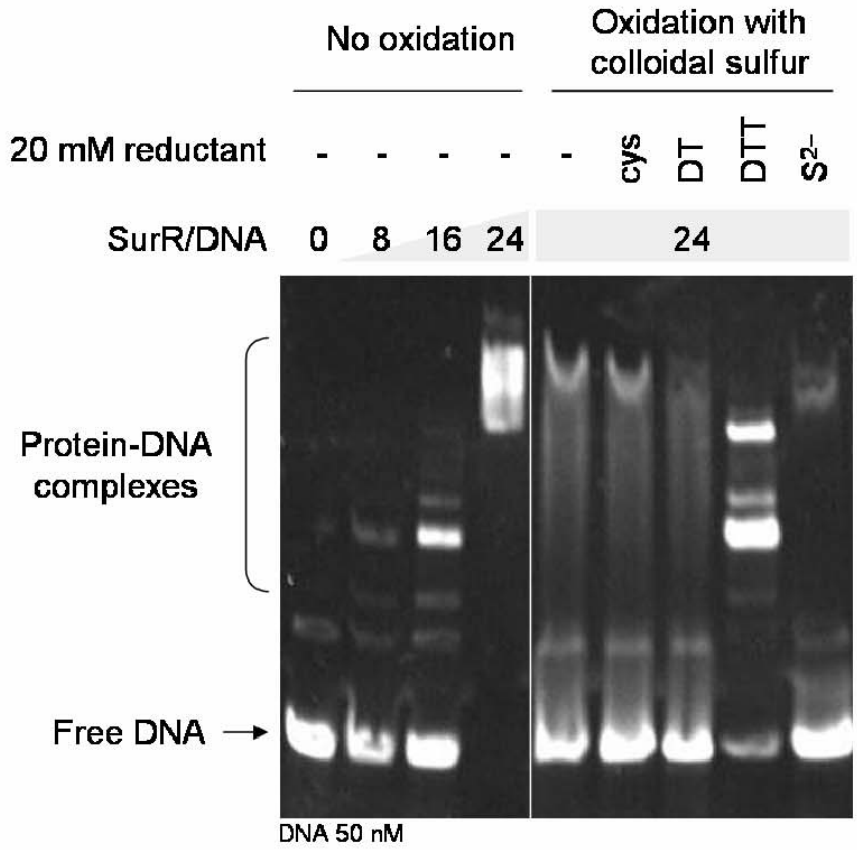

No oxidation
Oxidation with
colloidal sulfur
20 mM reductant
-
-
-
-
-
cys
DT
DTT
S2-
SurR/DNA
0
8
16
24
24
Protein-DNA
complexes
Free DNA
DNA 50 nM

Figure 5.12 Quaternary structure of SurR does not change upon oxidation. The elution profiles of four different protein samples from a Superdex 75 10/300 GL column (GE Healthcare) are shown. The protein samples run were SurR, AxxA-SurR, SurR treated with 10 mM diamide, and SurR treated with ~2 mM colloidal sulfur. The intensity of each major peak ranged from ~40 to 60 mAu; however, for this figure, the intensity of each sample was scaled to the SurR sample peak.

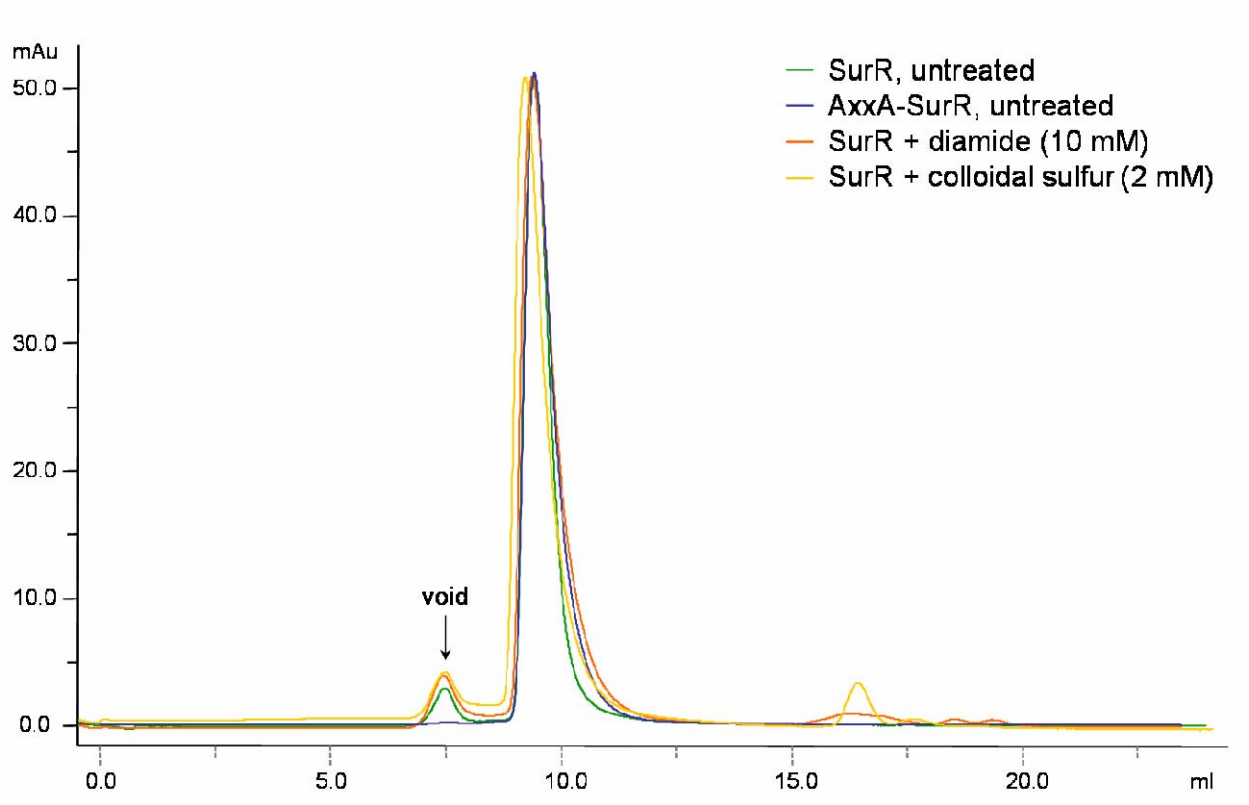

mAu
50.0
40.0
30.0
20.0
10.0
0.0
void
0.0
5.0
10.0
15.0
20.0
ml
SurR, untreated
AxxA-SurR, untreated
SurR + diamide (10 mM)
SurR + colloidal sulfur (2 mM)

Figure 5.13 Comparison of SurR with AxxA-SurR structures reveals a change in only one monomer. Overlay of SurR dimer (blue) with AxxA-SurR dimer (red) in ribbon (**A**) and cartoon (**B**) display, with the structural difference outlined by a broken circle. Chain A of the structures is located on the left while chain B is to the right. **C.** Dimer structure of AxxA-SurR N-terminal residues 1-75 showing the difference in the two monomers around the region of the mutant 'AxxA' motif. Helices 1 and 2 are in blue, random coil/predicted helix is in light blue, recognition helices are in red, and wings are in yellow. AxxA alanine residues in each monomer are shown in ball-and-stick form. Structure images were produced using PyMOL [124].

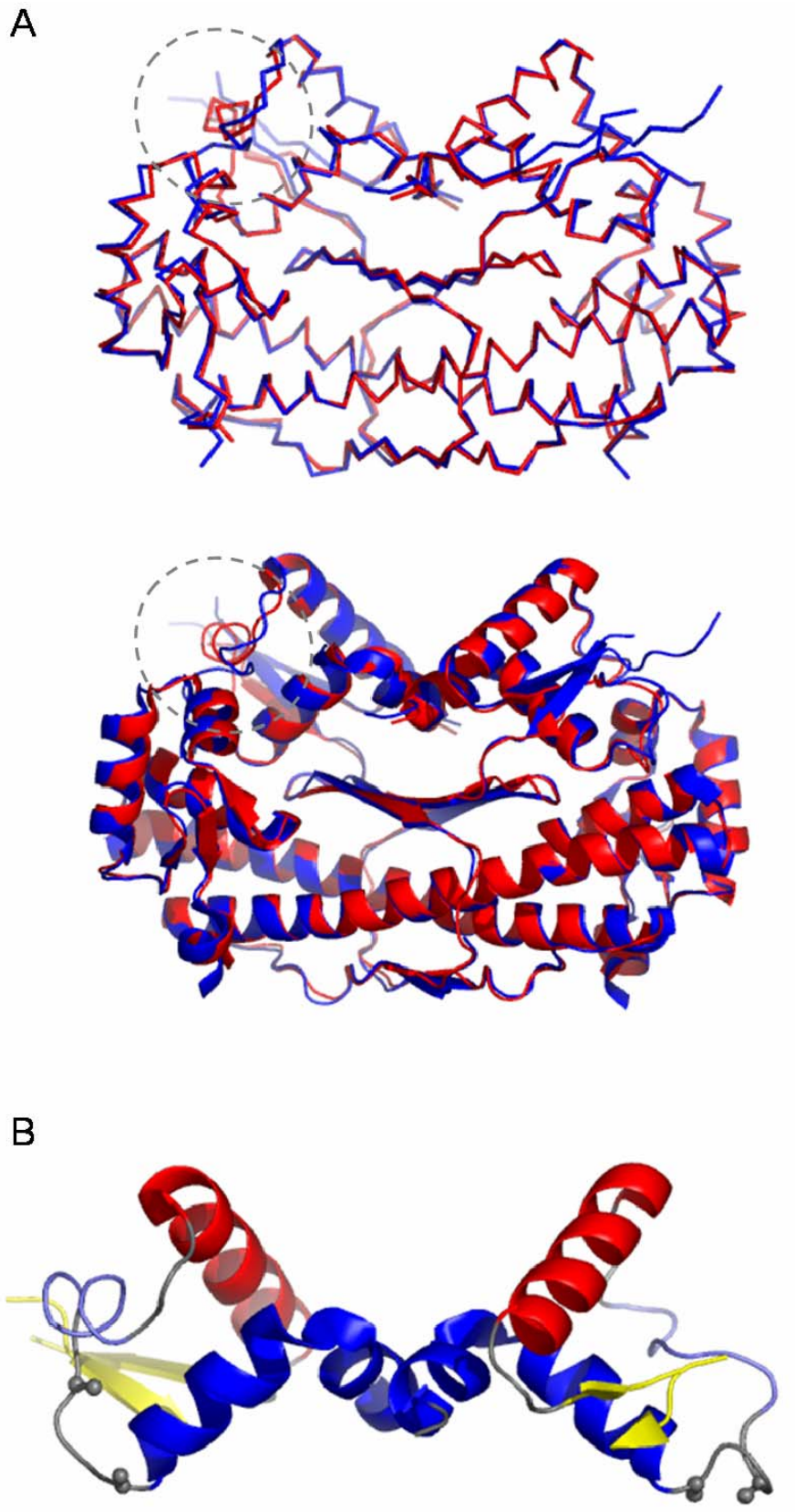

A
B

Figure 5.14 Comparison of the structure of the wHTH domain of SurR versus AxxA-SurR. A. Ribbon overlay of chain A from SurR (blue) and AxxA-SurR (red). **B.** Detailed comparison of wHTH domain (residues 1-75) for SurR representing the oxidized form (top) and AxxA-SurR representing the reduced form of the protein (bottom). Helices one and two are shown in blue, the recognition helix in red, and the wing in yellow. **C.** Close-up view of the overlay of chain A from SurR (blue) and AxxA-SurR (red) showing the positions of the cysteine residues in the CxxC motif and the alanine residues in the mutant 'AxxA' motif. Structure images were produced using PyMOL [124].

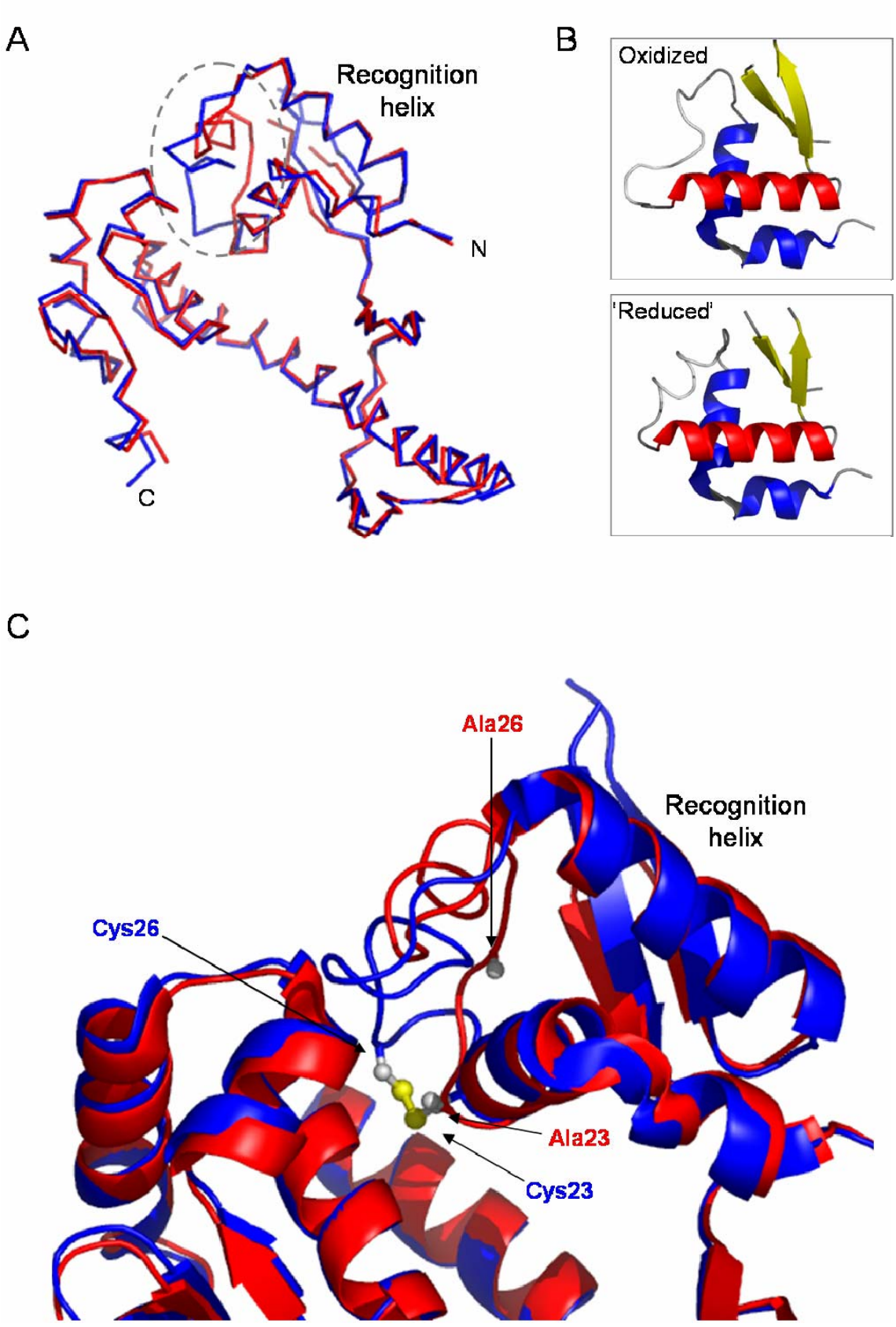

A
Recognition
helix
N
C
B
Oxidized
'Reduced'
C
Ala26
Recognition
helix
Cys26
Ala23
Cys23

Figure 5.15 Comparison of SurR and Phr DNA-binding domains. A. Residues 1-75 of SurR (chain A) are overlaid with residues 1-77 of Phr (chain A, PDB ID 2P4W) shown in ribbon (left) and cartoon (right). **B.** Residues 1-75 of AxxA-SurR (chain A) are overlaid with residues 1-77 of Phr (chain A) shown in ribbon (left) and cartoon (right). **C.** Sequence alignment (ClustalW [125, 126]) of the N-terminal wHTH domains of SurR and Phr showing secondary structure elements. Identical residues are highlighted in yellow and the cysteines in the SurR sequence are in red font. Structure images were produced using PyMOL [124].

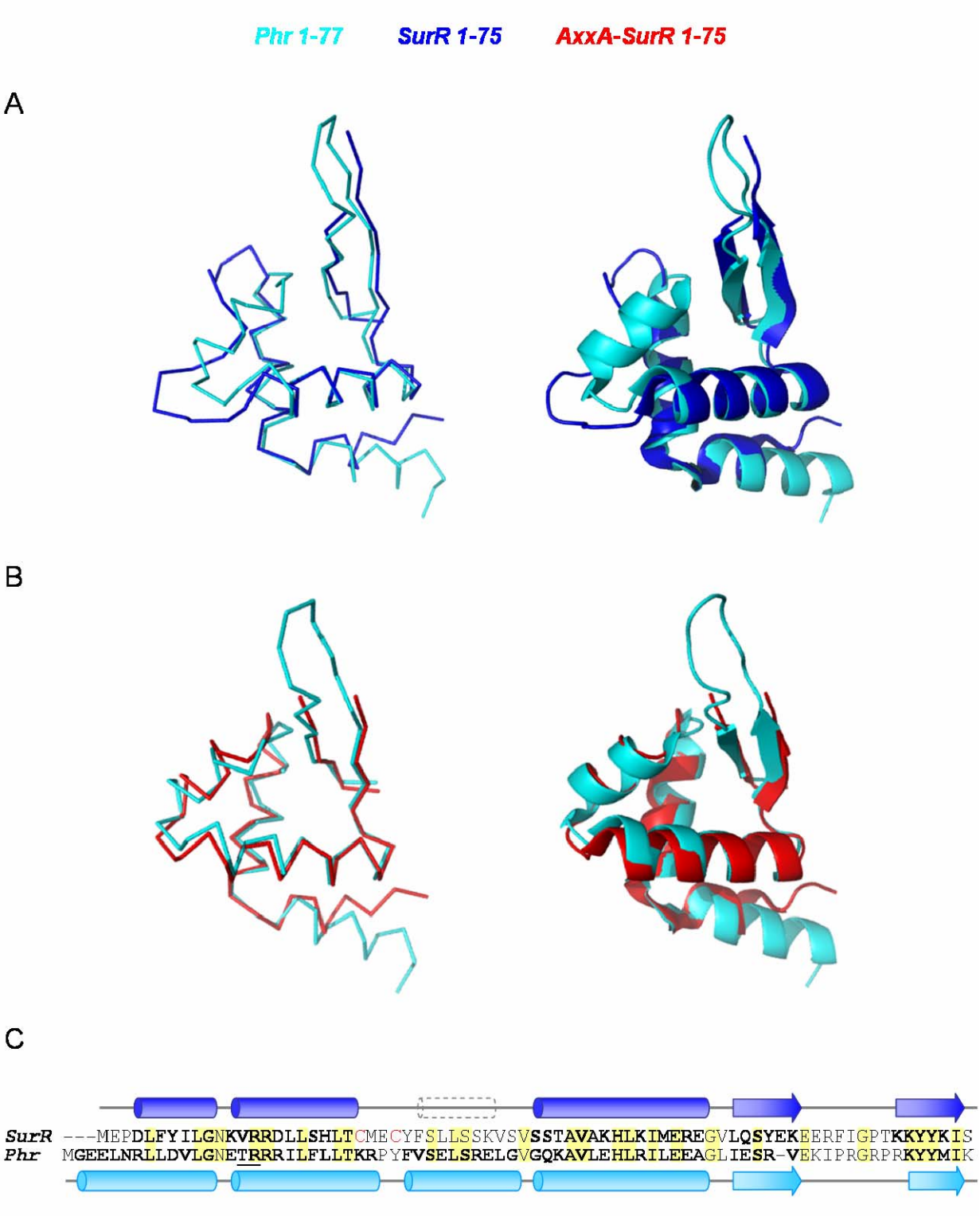

Phr 1-77
SurR 1-75
AxxA-SurR 1-75
A
B
C
SurR ---MEPDLFYILGNKVRRDLLSHLTCMECYFSLLSSKVSVSSTAVAKHLKIMEREGVLQSYEKEERFIGPTKKYYKIS
Phr MGEELNRLLDVLGNETRRRILFLLTKRPYFVSELSRELGVGQKAVLEHLRILEEAGLIESR-VEKIPRGRPRKYYMIK

Figure 5.16 Residues in coil region have altered positions in ‘reduced' versus oxidized form of SurR. Overlay of chain A of SurR with chain A of AxxA-SurR highlighting the coil residues 23 to 36 in the ‘reduced’ form (red) versus the oxidized form (blue): top view with recognition helix (H_R) oriented at the bottom (**A**) and side view with key residues labeled (**B**). Note that the side chain of Tyr27 for the AxxA-SurR structure was not able to be resolved from the electron density. Structure images were produced using PyMOL [124].

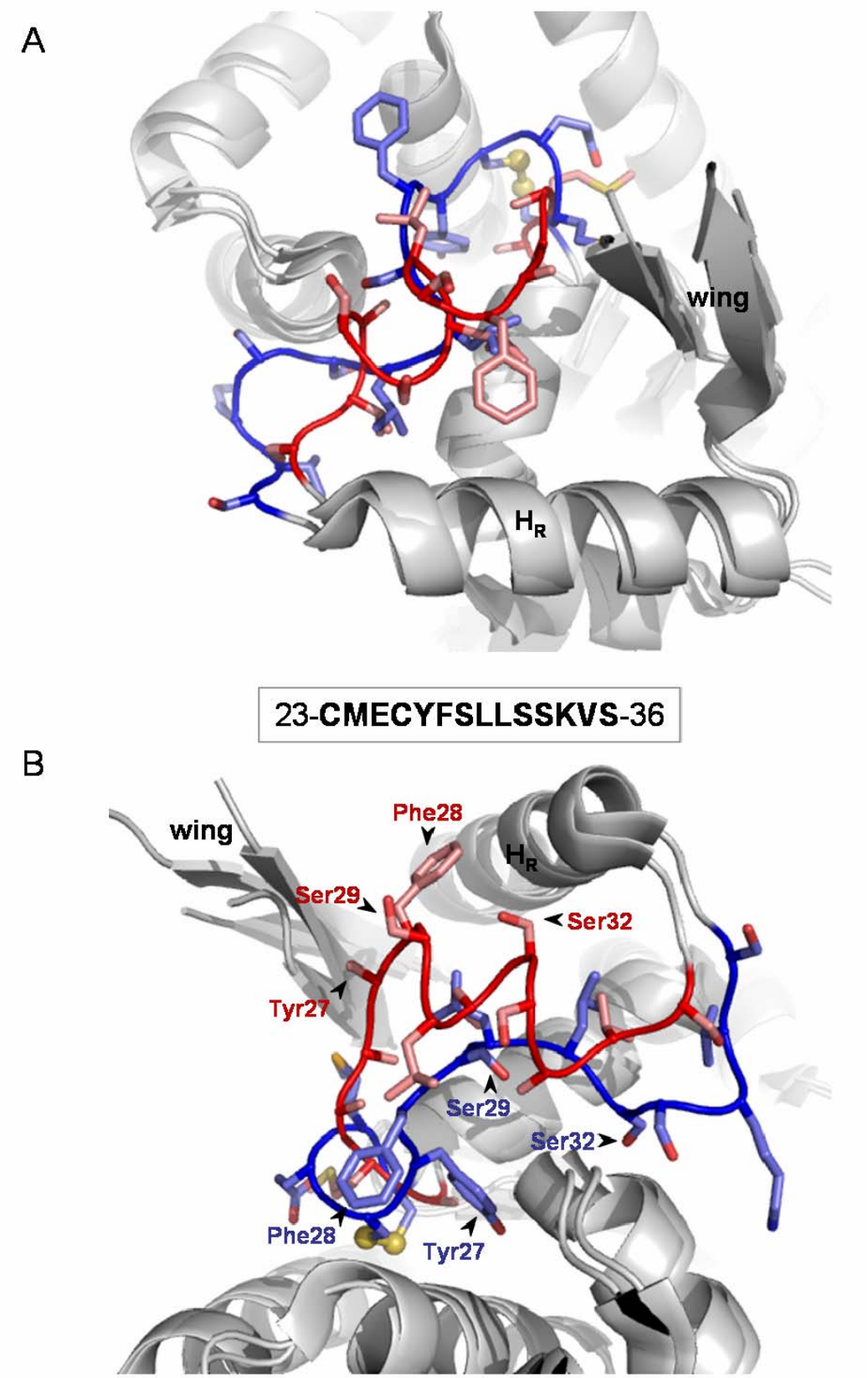
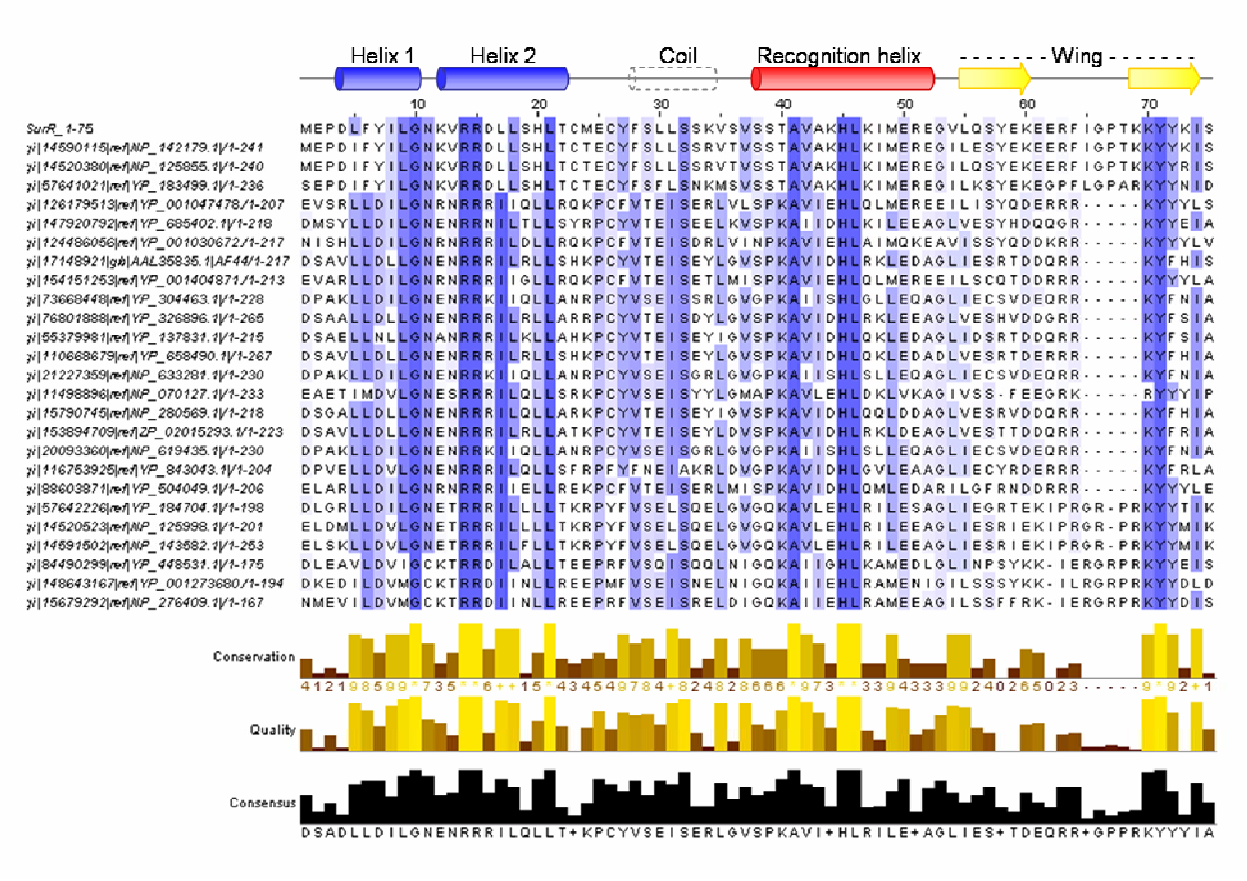

A
wing
H_R
23-CMECYFSLLSSKVS-36
B
wing
Phe28
H_R
Ser29
Ser32
Tyr27
Ser29
Ser32
Phe28
Tyr27

Figure 5.17 Sequence pile-up of SurR with close relatives shows conserved residues. PSI-BLAST [81, 82] of SurR wHTH domain (residues 1-75) was used to find close relatives with conserved residues. The cut-off for inclusion in the second PSI-BLAST round was an E-value of 0.001 or better. A sequence alignment of the resulting retrieved sequences with E-values better than 10^{-10} is shown. ClustalW [125, 126] was use to perform the sequence alignment and JalView [138] was used to format it as shown.


Helix 1
Helix 2
Coil
Recognition helix
Wing
SurR_1-75
Conservation
Quality
Consensus

CHAPTER 6

THE BIG PICTURE: THE ROLE OF SURR IN *PYROCOCCUS FURIOSUS* SULFUR RESPONSE

6.1 *P. furiosus* sulfur response and transcriptional regulation

In volcanic communities where hyperthermophilic species such as *P. furiosus* thrive, sulfur and its compounds are some of the most plentiful sources of electron acceptors and donors [139]. It is not surprising, therefore that most hyperthermophiles are sulfur-dependent [73, 75]. *P. furiosus*, however, is one exception in that it can grow equally well in the presence or absence of elemental sulfur (S^0) [74]. This being the case, the regulatory activity of transcription factors is one means to orchestrate the metabolic switch necessary to accommodate either the presence or absence of sulfur. In its natural environments near shallow and abyssal marine hydrothermal vents, *P. furiosus* must be able to respond quickly to environmental changes, with the appearance of sulfur being one such change. The ability of *P. furiosus* to sense and respond to S^0 in culture has already been shown by DNA microarray expression analyses [78, 100]. In particular, H_2S production begins only 10 min after the addition of S^0 to a growing culture, concomitant with the significant up and down-regulation of several genes, demonstrating the rapidity of the response of *P. furiosus* to this environmental change [100]. This work has described the characterization of SurR[1], a regulatory transcription factor which appears to be an integral part of this response and is itself regulated in response to the presence of sulfur through its internal redox switch.

[1] This is an abbreviation for sulfur-response regulator.

SurR has been shown to bind to the putative promoters of several of the genes which are both up and down-regulated within 10 min of S^0 addition to a growing culture according to DNA microarray studies; moreover, its DNA-binding motif elucidated by SELEX was found in 14 out of 17 of these ORF/operon promoters (see Table 4.2). At first, this result was puzzling since the relationship of SurR transcriptional regulation to the DNA microarray results was not understood. With the discovery of the CxxC redox switch and demonstration of SurR's dual regulatory activity, however, the physiological role of SurR began to come into focus. From the experimental data that have been presented, the relationship of SurR to the intracellular response of *P. furiosus* to sulfur can now be defined as such: SurR is capable of binding specifically to its cognate DNA and exerting transcriptional control on its target ORFs when the reduced form of the protein predominates in the absence of sulfur, but when sulfur is present, the CxxC motif becomes oxidized and functions as a redox switch causing SurR to release from the DNA. SurR activates transcription of the hydrogenase operons and seven other genes while at the same time it represses transcription from at least four ORFs/operons which encode proteins necessary for sulfur metabolism, most notably, NSR (NAD(P)H sulfur reductase). The DNA microarray data mapping sulfur response show exactly the opposite regulation [100], but this can be explained by the redox switch of SurR, since when sulfur is present SurR no longer binds to its cognate DNA. Therefore, the effect that is observed in microarray is at least in part a result of 'de-activation' and 'de-repression' when SurR is no longer exerting any transcriptional control due to its loss of DNA-binding ability. This explanation for the physiological role of SurR is summarized in Figure 6.1.

It is notable that at least one of the target operons of SurR encodes two putative transcriptional regulators whose expression is up-regulated in response to sulfur according to

DNA microarray data (PF2051, PF2052) [100]. It is possible that these regulators are necessary for the secondary response to sulfur which occurs 30 min after S^0 addition to a growing culture [100].

The ORF encoding the SurR transcription factor does not display significant regulation in response to sulfur in DNA microarray experiments, but it is likely that SurR DNA-binding activity can be regulated at the protein level *in vivo* via its redox switch. *In vitro* transcription assays suggest that SurR represses transcription of its own gene, but the range of concentrations of SurR that were tested in the assay were high and the basal transcription level of the *surr* transcript was somewhat low (see Fig. 4.23). The SurR footprint is located approximately 40 bp upstream from the *surr* ORF, so it is conceivable that SurR may not block core promoter elements for its own gene. Whether or not SurR represses transcription from its own ORF *in vivo*, it seems likely that other factors may also be involved in orchestrating the transcriptional regulation of *surr*.

It is likely that SurR is not the only transcriptional regulator involved in early sulfur response. Simple 'de-activation' and 'de-repression' does not adequately explain the dramatic transcriptional regulation observed in the DNA microarray results, particularly in the case of *mbh1*, which appears to have a very strong promoter in the absence of any regulator as evidenced by *in vitro* transcription (see Fig. 4.22 A). Accordingly, there must be a few other regulatory transcription factors which function in synchrony with SurR to elicit a rapid coordinated response to the presence of S^0. Now that so much is known about the role of SurR in S^0 response, it would be fascinating to understand the complete picture of *P. furiosus* S^0-mediated transcriptional response.

6.2 Possibilities for intracellular redox effectors of SurR

6.2.1 Chemical effectors

Sulfur in its elemental state is not water-soluble; however, *P. furiosus* and many other S^0-dependent bacteria and archaea are able to internalize solubilized sulfur for use in downstream metabolic processes [140]. The end result of anaerobic respiration with S^0 is the production of H_2S which takes the place of the H_2 that is normally produced along with CO_2 in the absence of sulfur [141]. The process by which elemental sulfur is solubilized begins by natural diffusion of polysulfide from the solid sulfur and is probably accelerated by sulfur-solubilizing agents such as sulfide produced abiotically and from *P. furiosus* sulfur reduction [140, 142]. At temperatures above 80°C, S^0 undergoes spontaneous reduction to polysulfide [143, 144], and correspondingly, *P. furiosus* growth is sustainable on either S^0 or the appropriate concentration of polysulfide (elevated concentrations raise the pH to a point inhibitory to growth) [142]. Polysulfide is stable only at high pH and quickly dissociates in the neutral pH range [145], and colloidal sulfur in the form of stable S_8 rings and anionic chains is a significant product of polysulfide oxidation [129, 130, 132]. Therefore, while the elevated temperatures at which *P. furiosus* cultures are grown promotes the reduction of S^0 to polysulfide, the neutral pH of the culture does not promote polysulfide stability. It is likely that the soluble S_8 species of sulfur is the predominant form which finds its way inside the cell where it can be metabolized [146]. In support of this notion is the recent identification and characterization of the *P. furiosus* sulfur-reductase, NSR (NAD(P)H sulfur reductase), a cytosolic protein which appears to utilize colloidal sulfur as a substrate for coenzyme A-dependent evolution of sulfide *in vitro* [100]. The sulfur which enters the cell is used primarily for dissimilatory reduction in energy-yielding processes and not for incorporation into cellular proteins and other metabolites [139, 142]. It is reasonable to assume that sulfur in

some form must also serve to signal the cell in some manner in order to mobilize it to put the sulfur to use in energy-yielding pathways.

In vitro studies demonstrated that SurR was able to be oxidized by the thiol-specific oxidant diamide, with concomitant loss of its DNA binding specificity. According to the mechanism of diamide, the resulting oxidation state at the CxxC motif of SurR is most likely a disulfide bond. Therefore, it can be assumed that the CxxC motif represents a disulfide-switch such as those described for the bacterial regulator OxyR and the yeast regulator Yap1 [147], and more recently for the bacterial activator CprK [148].

As has been demonstrated, colloidal sulfur appears to have the same and perhaps more complete effect on SurR DNA-binding affinity *in vitro*, and this represents a more biological redox effector. The oxidation state in the CxxC motif caused by the colloidal sulfur is not known, but the possibility of polysulfide bond formation cannot be ignored since the oxidizing species may well be anionic polysulfide radicals, as discussed in Chapter 5. This may explain why colloidal sulfur appears to be a better redox effector of SurR, since the presence of a polysulfide bond would presumably drive an even greater conformational change in the DNA-binding domain than that which is observed with a simple disulfide bond. It is quite possible that colloidal sulfur is a redox effector responsible for oxidation of the SurR CxxC motif *in vivo*, for the very reason that colloidal sulfur is the probable *in vivo* substrate for NSR. The increasing availability of colloidal sulfur (and potential oxidizing free radical derivatives) could force the oxidation of SurR, causing derepression of *nsr*, allowing it to be transcribed and produced to metabolize the available sulfur.

If reactive sulfur species generated within the cell cause oxidative stress enough to modulate the DNA-binding activity of SurR, then a robust system must exist to keep the

oxidation under control. It has been suggested that CoA might serve as a redox buffer to combat oxidative stress instead of glutathione, particularly within *P. furiosus* and closely related species, because of the intrinsic stability of CoA under harsh conditions such as the elevated temperature required for growth of hyperthermophiles [149]. Also, intracellular levels of CoA in *P. furiosus* are relatively high, and interestingly, these levels nearly double when cultures are grown in the presence of S^0, with nearly half of the CoA being in the disulfide form [149]. If CoA is indeed a redox buffer in hyperthermophiles like glutathione is in mesophiles, then it makes sense for CoA levels to increase if sulfur metabolism causes excess oxidative stress. On the other hand, perhaps the requirement of NSR for CoA in its reduction of colloidal sulfur to sulfide necessitates the higher availability of CoA. If these suppositions are correct, it could be possible that CoA buildup within the cell as a result of sulfur unavailability might trigger redox activation of SurR through reduction at its CxxC motif, and SurR would in turn bind to its cognate DNA, repressing NSR and other genes necessary for sulfur metabolism while activating the hydrogenases, etc., which are essential to energy production in the absence of sulfur.

6.2.2 Enzymatic effectors

Interestingly, one of the genes which SurR has been shown to repress *in vitro* is *pdo* (PF0094) which encodes protein disulfide oxidoreductase (PDO) and is divergently transcribed from the gene encoding SurR (PF0095). PDO contains two thioredoxin folds with two CxxC active site motifs and has the capability to both reduce disulfide bonds and oxidize dithiols [105, 120]. The possibility that PDO could be responsible for oxidizing or reducing the SurR CxxC motif is intriguing, and if this were the case, then the fact that PDO is regulated by SurR suggests a kind of feedback mechanism for control of SurR activity. In support of this notion is the fact that OxyR relies on a glutaredoxin to reverse its own oxidation, and OxyR itself regulates the

gene which encodes the glutaredoxin [150]. Also, the redox sensitivity of Yap1 is dependent upon the presence of a thioredoxin whose gene it induces upon being oxidized [151].

PDO is thought to be part of a system for maintenance of intracellular disulfide bonds, since hyperthermophiles in particular have the highest occurrence of intracellular protein disulfide bonds which presumably are critical to the stability of these thermophilic proteins [152, 153]. It is also possible that PDO functions like eukaryotic PDI (protein disulfide isomerase) which can form, break or isomerize disulfide bonds depending on the intracellular redox conditions [154]. This seems probable since in function, PDO resembles eukaryal PDI, even though it structurally resembles bacterial glutaredoxin; moreover, CxxC is the principal motif for disulfide bond formation, reduction and isomerization [136]. It is possible that within the cell, SurR requires enzymatic oxidation and does not possess intrinsic redox sensitivity, as is the case for Yap1 which can only be oxidized by enzymatic means via the action of a glutathione-peroxidase-like enzyme [147]. Regardless of the poorly defined specific intracellular activity of PDO, the implication for its involvement in the regulation of SurR activity is strong.

One other possible enzymatic effector is the protein encoded by PF1422, divergently transcribed from the membrane-bound hydrogenase operon. It is annotated as a 'thioredoxin reductase,' and sequence analyses with BLAST and conserved domain searches are in agreement. Interestingly, a PF1422 ortholog named *Ph*TrxR exists in *P. horikoshii* (PH1426, 89% identity) which can reduce thioredoxin I from *E. coli* [155], providing evidence that PF1422 is likely a redox-active enzyme. Moreover, *Ph*TrxR was found to reduce the protein product of PH0178 (named *Ph*RP), an ortholog of *P. furiosus* PDO (PF0094) [155]. This NAD(P)H-dependent reduction of *Ph*RP by *Ph*TrxR *in vitro* constitutes a general disulfide reductase system [152, 153, 155] which would presumably function for the orthologs PDO and PF1422 protein.

Although the *in vivo* function of PDO and PF1422 protein may differ, the possibility that they function in concert in a thioredoxin-like system is especially intriguing considering their potential dual regulation by SurR. PF1422 is not listed among those that are significantly regulated in DNA microarray expression analyses comparing *P. furiosus* growth in the presence and absence of S^0 [78, 100]. However, it is likely to be transcriptionally regulated in some manner by SurR due to the position of the SurR footprints in the PF1422-*mbh1* intergenic region, although no transcript from the PF1422 ORF was detectable in *in vitro* transcription assays in the presence or absence of SurR. If this protein is indeed a redox enzyme, it may represent a second feedback regulation mechanism which can exert control over the activity of SurR, perhaps in concert with PDO.

6.3 SurR as a global transcriptional regulator

Whereas this work has focused on the role of SurR in regulating ORFs involved in sulfur response, a word must be said with regard to other as-of-yet unknown targets for SurR regulation. Genome-wide searches of ORF upstream DNA with the SurR SELEX consensus motif resulted in over 100 hits, indicating that SurR may be involved in regulation of genes not necessarily involved in S^0 response. Even if some or many of these hits are false positives, the binding footprint of SurR on at least one of these upstream regions has been shown (PF1911-PF1912 intergenic region), indicating that some database hits are indeed representative of true binding sites. It is significant that the extended SELEX motif, $GTTn_3AACn_5GTT$, was present in upstream DNA regions of some ORFs which were not involved in S^0 response as determined by DNA microarray. Due to the presence of the extended motif in their upstream DNA, these ORFs represent very likely targets for SurR transcriptional regulation.

Since this regulator appears to have a redox-responsive switch, it is reasonable to assume that it would be sensitive to oxidative stress in general. This being said, however, there are no putative SurR bindins sites upstream of at least four genes thought to play a role against oxidative stress in *P. furiosus*: superoxide reductase (SOR, PF1281), rubredoxin (Rd, PF1281), NAD(P)H:rubredoxin oxidoreductase (NROR, PF1197), and NADH oxidase (NOX1, PF1532) [156, 157]. A recent DNA microarray study testing the effect of gamma radiation on *P. furiosus*, the result of which is predominantly hydroxyl radical production via radiolysis of water, found that many oxygen detoxification response genes appear to be constituatively expressed [158]. The only notable SurR-targets that showed regulation in this study were the membrane bound hydrogenase operon (PF1423-PF1434), the soluble hydrogenase I operon (PF0891-PF0894) and the MBX operon (PF1453-PF1444), all of which exhibited the same regulation pattern observed in *P. furiosus* S^0 response. It is very likely that the regulation observed for these operons may be orchestrated at least in part by SurR in response to oxidative stress.

Since oxidation of the CxxC motif appears to be a negative regulator of SurR DNA binding, then in order for SurR to exert the appropriate physiological response to oxidative stress it would have to serve as a repressor for genes necessary to combat oxidative stress, unlike both oxyR and Yap1 which activate target 'antioxidant' genes upon oxidation [151].

For proteins with redox-active cysteine residues, formation of a disulfide bond is not the only means of oxidation, and it is not known whether or not disulfide formation is the only way whereby the action of SurR is modulated. There are a host of possible redox modifications, both reversible and irreversible, which can be applied to cysteine residues through chemical or enzymatic means: thiyl radicals, sulfenic, sulfinic, and sulfonic acids, sulfenyl-amides, thiosulfinates, and polysulfides, for example [159]. An interesting question that arises is: could

SurR generate different intracellular responses based on different oxidation states of its CxxC motif?

Under non-oxidizing conditions, the DNA-binding ability of SurR may be regulated by other means. The possibility of the existence of other protein cofactors or DNA-binding partners for SurR has not been addressed in this work, but would indeed be interesting. Even the formation of heterodimers which recognize different sequences is a possibility, especially considering the structural similarity of the heat-shock regulator Phr from *P. furiosus* to SurR. It has been suggested that Lrp-family proteins, or "Feast/Famine Regulatory Proteins" as they have also been named, have the capacity to form heterodimers capable of exerting different regulatory effects [46].

There is another intriguing possibility with regard to alternate regulation pathways involving SurR. The two cysteine residues in the CxxC motif are positioned in close proximity to two histidine residues, and together, these could potentially coordinate a metal ion (Fig. 6.2). Because of the relationship of the SurR sequence to bacterial metal-responsive regulator SmtB and ArsR, early experiments tested the effect of various metals on the ability of his_6-SurR to shift DNA, and no obvious effect was observed. However, if SurR contains a true metal binding site, then the presence of the his-tag could have interfered with metal binding. It is intriguing that the *E. coli* ferric uptake regulator, Fur, contains a CxxC motif which can alternate between disulfide formation and zinc binding *in vitro* as a way of regulating dimer formation [160]. Also, two other bacterial redox-sensing proteins, Hsp33 chaperone and anti-sigma factor RsrA, are sensitive to cellular disulfide stress through their cysteine-containing redox centers which also coordinate zinc ions as part of the sensing mechanism [161].

6.4 Conclusions and outlook

SurR can now be added to the list of known archaeal transcriptional regulators which is increasingly expanding. To date, SurR is the only archaeal regulator that has been directly shown *in vitro* to both activate and repress transcription, and it is the first archaeal regulator shown to contain a redox-sensitive switch. SurR is especially interesting since it appears to function as a global transcriptional regulator, acting upon an entire regulon of S^0-responsive ORFs, and potentially other ORFs which contain its binding site in their upstream DNA. It has been shown to be versatile in its DNA-binding ability, with its binding mode dependent on the number of adjacent motifs present within its binding site, and this no doubt contributes to its global function in transcriptional regulation of many ORFs.

Even though much has been discovered regarding this novel archaeal transcriptional regulator, there is much more work that could be done to elucidate the complete regulatory network and *in vivo* functions of SurR within the cell. Since a genetic system is not available for *P. furiosus*, work is limited to *in vitro* applications; however, genetic capabilities are possible for the family member *T. kodakaraensis* which contains a SurR ortholog. Using this ortholog which also contains the CxxC motif, *in vivo* experiments could be conducted in a similar manner to the recent work done for Tgr mentioned in Chapter 1 [53]. Structural information of the DNA-bound form of SurR would help to fully understand the mechanism behind the CxxC redox switch, particularly why only one monomer of the dimer shows a structural change between the oxidized and 'reduced' states. It would also help to sort out the manner in which the protein binds to adjacent motifs. The fact that this protein is an activator is also intriguing; only one archaeal activator, Ptr2, has been mechanistically characterized to date [63]. More information regarding

the particular mechanism of activation carried out by SurR would contribute to the overall knowledge of archaeal transcriptional regulatory mechanisms.

As the field of archaeal research continues to expand, there will certainly be many advances in the area of transcriptional regulation, already evidenced by an increasing number of publications on this topic. The work that has been done for SurR will undoubtedly contribute to this expanding field and help to shed light on novel mechanisms and processes involving transcription within the archaeal domain.

Figure 6.1 Model for SurR involvement in *P. furiosus* S^0 response. In the absence of sulfur, SurR remains in the reduced state with free thiols in its CxxC motif. This allows the binding of SurR to its *cis* regulatory elements where it exerts transcriptional control. In the presence of sulfur, the SurR CxxC motif becomes oxidized (possibly to a disulfide or polysulfide bond), whereupon SurR can no longer bind specifically to its DNA recognition sequences.

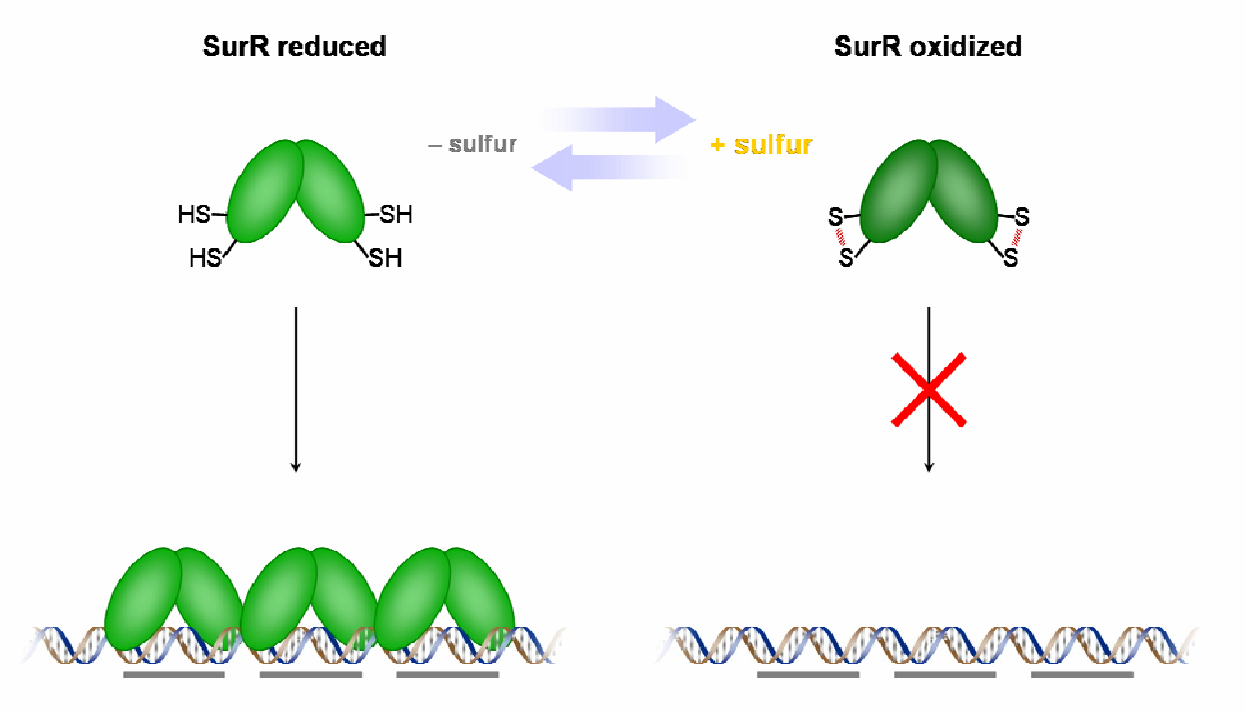
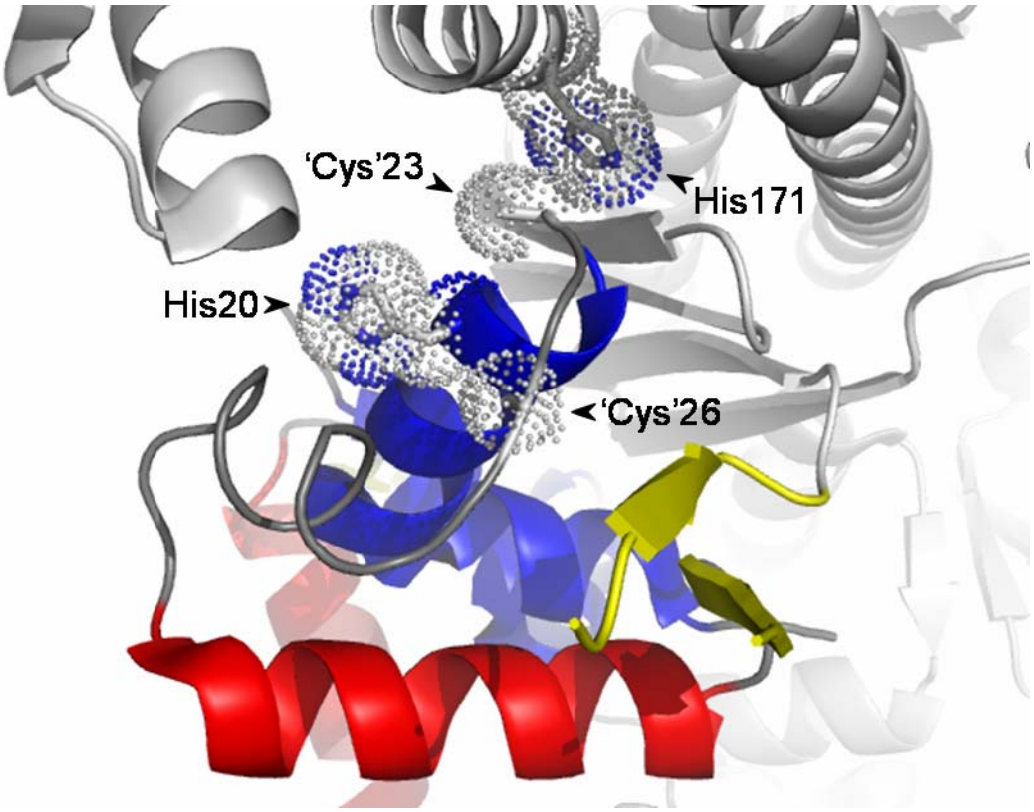

SurR reduced
SurR oxidized
– sulfur
+ sulfur
HS
SH
HS
SH
S
S
S
S

Figure 6.2 A potential metal binding site? View of the wHTH domain with residues His20, Cys23, Cys26, and His171 indicated. Helices 1 and 2 are in blue, coil is in dark grey, recognition helices are in red, and wing is in yellow. Structure image was produced using PyMOL [124].


'Cys'23
His171
His20
'Cys'26

REFERENCES

1. Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A *74*, 5088-5090.
2. Woese, C.R., Kandler, O., and Wheelis, M.L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A *87*, 4576-4579.
3. Chaban, B., Ng, S.Y.M., and Jarrell, K.F. (2006). Archaeal habitats - from the extreme to the ordinary. Canadian Journal of Microbiology *52*, 73-116.
4. Robertson, C.E., Harris, J.K., Spear, J.R., and Pace, N.R. (2005). Phylogenetic diversity and ecology of environmental Archaea. Current Opinion in Microbiology *8*, 638-642.
5. Stein, J.L., and Simon, M.I. (1996). Archaeal ubiquity. Proceedings of the National Academy of Sciences of the United States of America *93*, 6228-6230.
6. Bell, S.D., and Jackson, S.P. (1998). Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. Trends Microbiol *6*, 222-228.
7. Cramer, P. (2002). Multisubunit RNA polymerases. Curr Opin Struct Biol *12*, 89-97.
8. Borukhov, S., and Nudler, E. (2003). RNA polymerase holoenzyme: structure, function and biological implications. Current Opinion in Microbiology *6*, 93-100.
9. Thomm, M. (2007). Transcription: Mechanism and Regulation. In Archaea : molecular and cellular biology, R. Cavicchioli, ed. (Washington, DC: ASM Press), pp. xii, 523 , [518] of plates.
10. Kornberg, R.D. (2007). The molecular basis of eukaryotic transcription. Proc Natl Acad Sci U S A *104*, 12955-12961.
11. Myers, L.C., Leuther, K., Bushnell, D.A., Gustafsson, C.M., and Kornberg, R.D. (1997). Yeast RNA polymerase II transcription reconstituted with purified proteins. Methods *12*, 212-216.
12. Hausner, W., Wettach, J., Hethke, C., and Thomm, M. (1996). Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase. J Biol Chem *271*, 30144-30148.
13. Bell, S.D., Brinkman, A.B., van der Oost, J., and Jackson, S.P. (2001). The archaeal TFIIE alpha homologue facilitates transcription initiation by enhancing TATA-box recognition. Embo Reports *2*, 133-138.
14. Bartlett, M.S. (2005). Determinants of transcription initiation by archaeal RNA polymerase. Curr Opin Microbiol *8*, 677-684.
15. Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annual Review of Biochemistry *72*, 449-479.
16. Qureshi, S.A., and Jackson, S.P. (1998). Sequence-specific DNA binding by the S-shibatae TFIIB homolog, TFB, and its effect on promoter strength. Molecular Cell *1*, 389-400.
17. Paget, M.S.B., and Helmann, J.D. (2003). Protein family review - The sigma(70) family of sigma factors. Genome Biology *4*, -.

18. Soppa, J. (1999). Transcription initiation in Archaea: facts, factors and future aspects. Mol Microbiol *31*, 1295-1305.
19. Bell, S.D., Kosa, P.L., Sigler, P.B., and Jackson, S.P. (1999). Orientation of the transcription preinitiation complex in archaea. Proc Natl Acad Sci U S A *96*, 13662-13667.
20. van de Werken, H.J., Verhees, C.H., Akerboom, J., de Vos, W.M., and van der Oost, J. (2006). Identification of a glycolytic regulon in the archaea Pyrococcus and Thermococcus. FEMS Microbiol Lett *260*, 69-76.
21. Kosa, P.F., Ghosh, G., DeDecker, B.S., and Sigler, P.B. (1997). The 2.1-A crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. Proc Natl Acad Sci U S A *94*, 6042-6047.
22. Littlefield, O., Korkhin, Y., and Sigler, P.B. (1999). The structural basis for the oriented assembly of a TBP/TFB/promoter complex. Proc Natl Acad Sci U S A *96*, 13668-13673.
23. Werner, F., and Weinzierl, R.O. (2005). Direct modulation of RNA polymerase core functions by basal transcription factors. Mol Cell Biol *25*, 8344-8355.
24. Goede, B., Naji, S., von Kampen, O., Ilg, K., and Thomm, M. (2006). Protein-protein interactions in the archaeal transcriptional machinery: binding studies of isolated RNA polymerase subunits and transcription factors. J Biol Chem *281*, 30581-30592.
25. Magill, C.P., Jackson, S.P., and Bell, S.D. (2001). Identification of a conserved archaeal RNA polymerase subunit contacted by the basal transcription factor TFB. Journal of Biological Chemistry *276*, 46693-46696.
26. Geiduschek, E.P., and Ouhammouch, M. (2005). Archaeal transcription and its regulators. Mol Microbiol *56*, 1397-1407.
27. Sandman, K., and Reeve, J.N. (2006). Archaeal histones and the origin of the histone fold. Current Opinion in Microbiology *9*, 520-525.
28. White, M.F., and Bell, S.D. (2002). Holding it together: chromatin in the Archaea. Trends Genet *18*, 621-626.
29. Kadonaga, J.T. (1998). Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. Cell *92*, 307-313.
30. Bell, S.D. (2005). Archaeal transcriptional regulation--variation on a bacterial theme? Trends Microbiol *13*, 262-265.
31. Koonin, E.V., Tatusov, R.L., and Rudd, K.E. (1995). Sequence Similarity Analysis of Escherichia-Coli Proteins - Functional and Evolutionary Implications. Proceedings of the National Academy of Sciences of the United States of America *92*, 11921-11925.
32. Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M., and Iyer, L.M. (2005). The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol Rev *29*, 231-262.
33. Aravind, L., and Koonin, E.V. (1999). DNA-binding proteins and evolution of transcription regulation in the archaea. Nucleic Acids Res *27*, 4658-4670.
34. Ouhammouch, M. (2004). Transcriptional regulation in Archaea. Curr Opin Genet Dev *14*, 133-138.
35. Rojo, F. (2001). Mechanisms of transcriptional repression. Curr Opin Microbiol *4*, 145-151.
36. Bell, S.D., Cairns, S.S., Robson, R.L., and Jackson, S.P. (1999). Transcriptional regulation of an archaeal operon in vivo and in vitro. Mol Cell *4*, 971-982.

37. Brinkman, A.B., Dahlke, I., Tuininga, J.E., Lammers, T., Dumay, V., de Heus, E., Lebbink, J.H., Thomm, M., de Vos, W.M., and van Der Oost, J. (2000). An Lrp-like transcriptional regulator from the archaeon Pyrococcus furiosus is negatively autoregulated. J Biol Chem *275*, 38160-38169.
38. Dahlke, I., and Thomm, M. (2002). A Pyrococcus homolog of the leucine-responsive regulatory protein, LrpA, inhibits transcription by abrogating RNA polymerase recruitment. Nucleic Acids Res *30*, 701-710.
39. Brinkman, A.B., Ettema, T.J., de Vos, W.M., and van der Oost, J. (2003). The Lrp family of transcriptional regulators. Mol Microbiol *48*, 287-294.
40. Bell, S.D., and Jackson, S.P. (2000). Mechanism of autoregulation by an archaeal transcriptional repressor. Journal of Biological Chemistry *275*, 31624-31629.
41. Enoru-Eta, J., Gigot, D., Thia-Toong, T.L., Glansdorff, N., and Charlier, D. (2000). Purification and characterization of Sa-lrp, a DNA-binding protein from the extreme thermoacidophilic archaeon Sulfolobus acidocaldarius homologous to the bacterial global transcriptional regulator Lrp. Journal of Bacteriology *182*, 3661-3672.
42. Enoru-Eta, J., Gigot, D., Glansdorff, N., and Charlier, D. (2002). High resolution contact probing of the Lrp-like DNA-binding protein Ss-Lrp from the hyperthermoacidophilic crenarchaeote Sulfolobus solfataricus P2. Molecular Microbiology *45*, 1541-1555.
43. Peeters, E., Thia-Toong, T.L., Gigot, D., Maes, D., and Charlier, D. (2004). Ss-LrpB, a novel Lrp-like regulator of Sulfolobus solfataricus P2, binds cooperatively to three conserved targets in its own control region. Molecular Microbiology *54*, 321-336.
44. Peeters, E., Willaert, R., Maes, D., and Charlier, D. (2006). Ss-LrpB from Sulfolobus solfataricus condenses about 100 base pairs of its own operator DNA into globular nucleoprotein complexes. Journal of Biological Chemistry *281*, 11721-11728.
45. Koike, H., Ishijima, S.A., Clowney, L., and Suzuki, M. (2004). The archaeal feast/famine regulatory protein: Potential roles of its assembly forms for regulating transcription. Proceedings of the National Academy of Sciences of the United States of America *101*, 2840-2845.
46. Yokoyama, K., Ishijima, S.A., Clowney, L., Koike, H., Aramaki, H., Tanaka, C., Makino, K., and Suzuki, M. (2006). Feast/famine regulatory proteins (FFRPs): Escherichia coli Lrp, AsnC and related archaeal transcription factors. FEMS Microbiol Rev *30*, 89-108.
47. Brinkman, A.B., Bell, S.D., Lebbink, R.J., de Vos, W.M., and van der Oost, J. (2002). The Sulfolobus solfataricus Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. Journal of Biological Chemistry *277*, 29537-29549.
48. Lee, S.J., Engelmann, A., Horlacher, R., Qu, Q., Vierke, G., Hebbeln, C., Thomm, M., and Boos, W. (2003). TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon Thermococcus litoralis. J Biol Chem *278*, 983-990.
49. DiRuggiero, J., Dunn, D., Maeder, D.L., Holley-Shanks, R., Chatard, J., Horlacher, R., Robb, F.T., Boos, W., and Weiss, R.B. (2000). Evidence of recent lateral gene transfer among hyperthermophilic Archaea. Molecular Microbiology *38*, 684-693.
50. Lee, S.J., Moulakakis, C., Koning, S.M., Hausner, W., Thomm, M., and Boos, W. (2005). TrmB, a sugar sensing regulator of ABC transporter genes in Pyrococcus furiosus exhibits dual promoter specificity and is controlled by different inducers. Molecular Microbiology *57*, 1797-1807.

51. Krug, M., Lee, S.J., Diederichs, K., Boos, W., and Welte, W. (2006). Crystal structure of the sugar binding domain of the archaeal transcriptional regulator TrmB. Journal of Biological Chemistry *281*, 10976-10982.
52. Lee, S.J., Surma, M., Seitz, S., Hausner, W., Thomm, M., and Boos, W. (2007). Characterization of the TrmB-like protein, PF0124, a TGM-recognizing global transcriptional regulator of the hyperthermophilic archaeon Pyrococcus furiosus. Mol Microbiol *65*, 305-318.
53. Kanai, T., Akerboom, J., Takedomi, S., van de Werken, H.J., Blombach, F., van der Oost, J., Murakami, T., Atomi, H., and Imanaka, T. (2007). A global transcriptional regulator in Thermococcus kodakaraensis controls the expression levels of both glycolytic and gluconeogenic enzyme-encoding genes. J Biol Chem.
54. Roy, S., Garges, S., and Adhya, S. (1998). Activation and repression of transcription by differential contact: two sides of a coin. J Biol Chem *273*, 14059-14062.
55. Lie, T.J., and Leigh, J.A. (2003). A novel repressor of nif and glnA expression in the methanogenic archaeon Methanococcus maripaludis. Mol Microbiol *47*, 235-246.
56. Lie, T.J., Wood, G.E., and Leigh, J.A. (2005). Regulation of nif expression in Methanococcus maripaludis - Roles of the euryarchaeal repressor NrpR, 2-oxoglutarate, and two operators. Journal of Biological Chemistry *280*, 5236-5241.
57. Zimmermann, P., and Pfeifer, F. (2003). Regulation of the expression of gas vesicle genes in Haloferax mediterranei: interaction of the two regulatory proteins GvpD and GvpE. Molecular Microbiology *49*, 783-794.
58. Hofacker, A., Schmitz, K.M., Cichonczyk, A., Sartorius-Neef, S., and Pfeifer, F. (2004). GvpE- and GvpD-mediated transcription regulation of the p-gvp genes encoding gas vesicles in Halobacterium salinarum. Microbiology-Sgm *150*, 1829-1838.
59. Pfeifer, F., Gregor, D., Hofacker, A., Plosser, P., and Zimmermann, P. (2002). Regulation of gas vesicle formation in halophilic archaea. Journal of Molecular Microbiology and Biotechnology *4*, 175-181.
60. Plosser, P., and Pfeifer, F. (2002). A bZIP protein from halophilic archaea: structural features and dimer formation of cGvpE from Halobacterium salinarum. Mol Microbiol *45*, 511-520.
61. Vierke, G., Engelmann, A., Hebbeln, C., and Thomm, M. (2003). A novel archaeal transcriptional regulator of heat shock response. J Biol Chem *278*, 18-26.
62. Liu, W., Vierke, G., Wenke, A.K., Thomm, M., and Ladenstein, R. (2007). Crystal structure of the archaeal heat shock regulator from Pyrococcus furiosus: A molecular chimera representing eukaryal and bacterial features. Journal of Molecular Biology *369*, 474-488.
63. Ouhammouch, M., Dewhurst, R.E., Hausner, W., Thomm, M., and Geiduschek, E.P. (2003). Activation of archaeal transcription by recruitment of the TATA-binding protein. Proc Natl Acad Sci U S A *100*, 5097-5102.
64. Ouhammouch, M., and Geiduschek, E.P. (2001). A thermostable platform for transcriptional regulation: the DNA-binding properties of two Lrp homologs from the hyperthermophilic archaeon Methanococcus jannaschii. Embo J *20*, 146-156.
65. Ouhammouch, M., Werner, F., Weinzierl, R.O., and Geiduschek, E.P. (2004). A fully recombinant system for activator-dependent archaeal transcription. J Biol Chem *279*, 51719-51721.

66. Ouhammouch, M., and Geiduschek, E.P. (2005). An expanding family of archaeal transcriptional activators. Proc Natl Acad Sci U S A *102*, 15423-15428.
67. Ouhammouch, M., Langham, G.E., Hausner, W., Simpson, A.J., El-Sayed, N.M., and Geiduschek, E.P. (2005). Promoter architecture and response to a positive regulator of archaeal transcription. Mol Microbiol *56*, 625-637.
68. Hochschild, A., and Dove, S.L. (1998). Protein-protein contacts that activate and repress prokaryotic transcription. Cell *92*, 597-600.
69. Herrera, F.J., Shooltz, D.D., and Triezenberg, S.J. (2004). Mechanisms of transcriptional activation in eukaryotes. Handbook Exp Pharm *166*, 3-31.
70. Kruger, K., Hermann, T., Armbruster, V., and Pfeifer, F. (1998). The transcriptional activator GvpE for the halobacterial gas vesicle genes resembles a basic region leucine-zipper regulatory protein. Journal of Molecular Biology *279*, 761-771.
71. Kessler, A., Sezonov, G., Guijarro, J.I., Desnoues, N., Rose, T., Delepierre, M., Bell, S.D., and Prangishvili, D. (2006). A novel archaeal regulatory protein, Sta1, activates transcription from viral promoters. Nucleic Acids Res *34*, 4837-4845.
72. Xie, Y., and Reeve, J.N. (2005). Regulation of tryptophan operon expression in the archaeon Methanothermobacter thermautotrophicus. J Bacteriol *187*, 6419-6429.
73. Stetter, K.O. (2006). Hyperthermophiles in the history of life. Philos Trans R Soc Lond B Biol Sci *361*, 1837-1842; discussion 1842-1833.
74. Fiala, G., and Stetter, K.O. (1986). Pyrococcus-Furiosus Sp-Nov Represents a Novel Genus of Marine Heterotrophic Archaebacteria Growing Optimally at 100-Degrees C. Archives of Microbiology *145*, 56-61.
75. Miroshnichenko, M.L., and Bonch-Osmolovskaya, E.A. (2006). Recent developments in the thermophilic microbiology of deep-sea hydrothermal vents. Extremophiles *10*, 85-96.
76. Adams, M.W. (1994). Biochemical diversity among sulfur-dependent, hyperthermophilic microorganisms. FEMS Microbiol Rev *15*, 261-277.
77. Adams, M.W., Holden, J.F., Menon, A.L., Schut, G.J., Grunden, A.M., Hou, C., Hutchins, A.M., Jenney, F.E., Jr., Kim, C., Ma, K., Pan, G., Roy, R., Sapra, R., Story, S.V., and Verhagen, M.F. (2001). Key role for sulfur in peptide metabolism and in regulation of three hydrogenases in the hyperthermophilic archaeon Pyrococcus furiosus. J Bacteriol *183*, 716-724.
78. Schut, G.J., Zhou, J., and Adams, M.W. (2001). DNA microarray analysis of the hyperthermophilic archaeon Pyrococcus furiosus: evidence for anNew type of sulfur-reducing enzyme complex. J Bacteriol *183*, 7027-7036.
79. 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.
80. 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. Anal Chem *71*, 2871-2882.
81. 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 Res *25*, 3389-3402.
82. 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

database searches with composition-based statistics and other refinements. Nucleic Acids Res *29*, 2994-3005.
83. 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 Res *31*, 383-387.
84. 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 Res *33*, D192-196.
85. Marchler-Bauer, A., and Bryant, S.H. (2004). CD-Search: protein domain annotations on the fly. Nucleic Acids Res *32*, W327-331.
86. Poole, F.L., 2nd, Gerwe, B.A., Hopkins, R.C., Schut, G.J., Weinberg, M.V., Jenney, F.E., Jr., and Adams, M.W. (2005). Defining genes in the genome of the hyperthermophilic archaeon Pyrococcus furiosus: implications for all microbial genomes. J Bacteriol *187*, 7325-7332.
87. Fried, M., and Crothers, D.M. (1981). Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res *9*, 6505-6525.
88. Wilson, D.O., Johnson, P., and McCord, B.R. (2001). Nonradiochemical DNase I footprinting by capillary electrophoresis. Electrophoresis *22*, 1979-1986.
89. Galas, D.J., and Schmitz, A. (1978). DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res *5*, 3157-3170.
90. Tullius, T.D., and Dombroski, B.A. (1986). Hydroxyl radical "footprinting": high-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein. Proc Natl Acad Sci U S A *83*, 5469-5473.
91. Tullius, T.D., Dombroski, B.A., Churchill, M.E., and Kam, L. (1987). Hydroxyl radical footprinting: a high-resolution method for mapping protein-DNA contacts. Methods Enzymol *155*, 537-558.
92. Sambrook, J., and Russel, D.W. (2001). Molecular Cloning, 3rd Edition (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
93. Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505-510.
94. 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. Mol Cell Biol *9*, 2944-2949.
95. Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol *2*, 28-36.
96. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.
97. Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr Purif *41*, 207-234.

98. Kosower, N.S., and Kosower, E.M. (1995). Diamide: an oxidant probe for thiols. Methods Enzymol *251*, 123-133.
99. Kosower, N.S., Kosower, E.M., Wertheim, B., and Correa, W.S. (1969). Diamide, a new reagent for the intracellular oxidation of glutathione to the disulfide. Biochem Biophys Res Commun *37*, 593-596.
100. Schut, G.J., Bridger, S.L., and Adams, M.W. (2007). Insights into the Metabolism of Elemental Sulfur by the Hyperthermophilic Archaeon Pyrococcus furiosus: Characterization of a Coenzyme A- Dependent NAD(P)H Sulfur Oxidoreductase. J Bacteriol *189*, 4431-4441.
101. Sapra, R., Verhagen, M.F., and Adams, M.W. (2000). Purification and characterization of a membrane-bound hydrogenase from the hyperthermophilic archaeon Pyrococcus furiosus. J Bacteriol *182*, 3423-3428.
102. Ma, K., Schicho, R.N., Kelly, R.M., and Adams, M.W. (1993). Hydrogenase of the hyperthermophile Pyrococcus furiosus is an elemental sulfur reductase or sulfhydrogenase: evidence for a sulfur-reducing hydrogenase ancestor. Proc Natl Acad Sci U S A *90*, 5341-5344.
103. Ma, K., Weiss, R., and Adams, M.W. (2000). Characterization of hydrogenase II from the hyperthermophilic archaeon Pyrococcus furiosus and assessment of its role in sulfur reduction. J Bacteriol *182*, 1864-1871.
104. Ma, K., and Adams, M.W. (2001). Hydrogenases I and II from Pyrococcus furiosus. Methods Enzymol *331*, 208-216.
105. Pedone, E., Ren, B., Ladenstein, R., Rossi, M., and Bartolucci, S. (2004). Functional properties of the protein disulfide oxidoreductase from the archaeon Pyrococcus furiosus: a member of a novel protein family related to protein disulfide-isomerase. Eur J Biochem *271*, 3437-3448.
106. Legrain, C., Villeret, V., Roovers, M., Gigot, D., Dideberg, O., Pierard, A., and Glansdorff, N. (1997). Biochemical characterisation of ornithine carbamoyltransferase from Pyrococcus furiosus. Eur J Biochem *247*, 1046-1055.
107. Chen, Y. (2003). The Fishing Expedition: Searching for Unidentified Archaeal Transcription Factors. M.S. Thesis thesis, University of Georgia, Athens.
108. 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.
109. Adams, M.W. (1993). Enzymes and proteins from organisms that grow near and above 100 degrees C. Annu Rev Microbiol *47*, 627-658.
110. Hethke, C., Bergerat, A., Hausner, W., Forterre, P., and Thomm, M. (1999). Cell-free transcription at 95 degrees: thermostability of transcriptional components and DNA topology requirements of Pyrococcus transcription. Genetics *152*, 1325-1333.
111. Hethke, C., Geerling, A.C., Hausner, W., de Vos, W.M., and Thomm, M. (1996). A cell-free transcription system for the hyperthermophilic archaeon Pyrococcus furiosus. Nucleic Acids Res *24*, 2369-2376.
112. Moxley, R.A., and Jarrett, H.W. (2005). Oligonucleotide trapping method for transcription factor purification systematic optimization using electrophoretic mobility shift assay. J Chromatogr A *1070*, 23-34.

113. Borges, K.M., Bergerat, A., Bogert, A.M., DiRuggiero, J., Forterre, P., and Robb, F.T. (1997). Characterization of the reverse gyrase from the hyperthermophilic Archaeon Pyrococcus furiosus. Journal of Bacteriology *179*, 1721-1726.
114. Chipman, D.M., and Shaanan, B. (2001). The ACT domain family. Curr Opin Struct Biol *11*, 694-700.
115. Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J., and Natale, D.A. (2003). The COG database: an updated version includes eukaryotes. BMC Bioinformatics *4*, 41.
116. Makarova, K.S., Aravind, L., Galperin, M.Y., Grishin, N.V., Tatusov, R.L., Wolf, Y.I., and Koonin, E.V. (1999). Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell. Genome Res *9*, 608-628.
117. Busenlehner, L.S., Pennella, M.A., and Giedroc, D.P. (2003). The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. FEMS Microbiol Rev *27*, 131-143.
118. McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics *16*, 404-405.
119. Bryson, K., McGuffin, L.J., Marsden, R.L., Ward, J.J., Sodhi, J.S., and Jones, D.T. (2005). Protein structure prediction servers at University College London. Nucleic Acids Res *33*, W36-38.
120. Ren, B., Tibbelin, G., de Pascale, D., Rossi, M., Bartolucci, S., and Ladenstein, R. (1998). A protein disulfide oxidoreductase from the archaeon Pyrococcus furiosus contains two thioredoxin fold units. Nat Struct Biol *5*, 602-611.
121. Tran, T.T., Dam, P., Su, Z., Poole, F.L., 2nd, Adams, M.W., Zhou, G.T., and Xu, Y. (2007). Operon prediction in Pyrococcus furiosus. Nucleic Acids Res *35*, 11-20.
122. Silva, P.J., van den Ban, E.C.D., Wassink, H., Haaker, H., de Castro, B., Robb, F.T., and Hagen, W.R. (2000). Enzymes of hydrogen metabolism in Pyrococcus furiosus. European Journal of Biochemistry *267*, 6541-6551.
123. Stroud, J. (2004). Make-NA Server. http://structure.usc.edu/make-na/server.html.
124. DeLano, W.L. (2002). The PyMOL Molecular Graphics System. http://www.pymol.org.
125. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-4680.
126. Lopez, R., Programme, S., and Lloyd, A. (1997). ClustalW WWW Service at the European Bioinformatics Institute. http://www.ebi.ac.uk/Tools/clustalw/.
127. Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol *292*, 195-202.
128. Kosower, N.S., and Kosower, E.M. (1987). Formation of disulfides with diamide. Methods Enzymol *143*, 264-270.
129. Kleinjan, W.E., de Keizer, A., and Janssen, A.J. (2005). Kinetics of the chemical oxidation of polysulfide anions in aqueous solution. Water Res *39*, 4093-4100.
130. Giggenba.W (1972). Optical-Spectra and Equilibrium Distribution of Polysulfide Ions in Aqueous-Solution at 20 Degrees. Inorganic Chemistry *11*, 1201-&.

131. Steudel, R., and Eckert, B. (2003). Solid Sulfur Allotropes. In Elemental Sulfur and Sulfur-Rich Compounds I. (Berlin / Heidelberg, Germany: Springer), pp. 1-80.
132. Kleinjan, W.E., de Keizer, A., and Janssen, J.H. (2003). Biologically Produced Sulfur. In Elemental Sulfur and Sulfur-Rich Compounds I. (Berlin / Heidelberg, Germany: Springer), pp. 167-188.
133. Jones, S., van Heyningen, P., Berman, H.M., and Thornton, J.M. (1999). Protein-DNA interactions: A structural analysis. J Mol Biol *287*, 877-896.
134. Coulocheri, S.A., Pigis, D.G., Papavassiliou, K.A., and Papavassiliou, A.G. (2007). Hydrogen bonds in protein-DNA complexes: Where geometry meets plasticity. Biochimie *89*, 1291-1303.
135. Luscombe, N.M., Laskowski, R.A., and Thornton, J.M. (2001). Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res *29*, 2860-2874.
136. Fomenko, D.E., and Gladyshev, V.N. (2003). Identity and functions of CxxC-derived motifs. Biochemistry *42*, 11214-11225.
137. Declercq, J.P., Evrard, C., Clippe, A., Stricht, D.V., Bernard, A., and Knoops, B. (2001). Crystal structure of human peroxiredoxin 5, a novel type of mammalian peroxiredoxin at 1.5 A resolution. J Mol Biol *311*, 751-759.
138. Clamp, M., Cuff, J., Searle, S.M., and Barton, G.J. (2004). The Jalview Java alignment editor. Bioinformatics *20*, 426-427.
139. Kletzin, A., Urich, T., Muller, F., Bandeiras, T.M., and Gomes, C.M. (2004). Dissimilatory oxidation and reduction of elemental sulfur in thermophilic archaea. J Bioenerg Biomembr *36*, 77-91.
140. Hedderich, R., Klimmek, O., Kroger, A., Dirmeier, R., Keller, M., and Stetter, K.O. (1998). Anaerobic respiration with elemental sulfur and with disulfides. Fems Microbiology Reviews *22*, 353-381.
141. Malik, B., Su, W.W., Wald, H.L., Blumentals, I.I., and Kelly, R.M. (1989). Growth and Gas-Production for Hyperthermophilic Archaebacterium, Pyrococcus-Furiosus. Biotechnology and Bioengineering *34*, 1050-1057.
142. Blumentals, II, Itoh, M., Olson, G.J., and Kelly, R.M. (1990). Role of Polysulfides in Reduction of Elemental Sulfur by the Hyperthermophilic Archaebacterium Pyrococcus furiosus. Appl Environ Microbiol *56*, 1255-1262.
143. Kamyshny, A., Jr., Gun, J., Rizkov, D., Voitsekovski, T., and Lev, O. (2007). Equilibrium distribution of polysulfide ions in aqueous solutions at different temperatures by rapid single phase derivatization. Environ Sci Technol *41*, 2395-2400.
144. Belkin, S., Wirsen, C.O., and Jannasch, H.W. (1985). Biological and Abiological Sulfur Reduction at High Temperatures. Appl Environ Microbiol *49*, 1057-1061.
145. Gun, J., Modestow, A.D., Kamyshny, A., Jr., Ryzkov, D., Gitis, V., Goifman, A., Lev, O., Hultsch, V., Grischek, T., and Worch, E. (2004). Electrospray Ionization Mass Spectrometric Analysis of Aqueous Polysulfide Solutions. Microchimica Acta *146*, 229-237.
146. Le Faou, A., Rajagopal, B.S., Daniels, L., and Fauque, G. (1990). Thiosulfate, polythionates and elemental sulfur assimilation and reduction in the bacterial world. FEMS Microbiol Rev *6*, 351-381.
147. Barford, D. (2004). The role of cysteine residues as redox-sensitive regulatory switches. Curr Opin Struct Biol *14*, 679-686.

148. Pop, S.M., Gupta, N., Raza, A.S., and Ragsdale, S.W. (2006). Transcriptional activation of dehalorespiration. Identification of redox-active cysteines regulating dimerization and DNA binding. J Biol Chem *281*, 26382-26390.
149. Hummel, C.S., Lancaster, K.M., and Crane, E.J. (2005). Determination of coenzyme A levels in Pyrococcus furiosus and other Archaea: implications for a general role for coenzyme A in thermophiles. Fems Microbiology Letters *252*, 229-234.
150. Zheng, M., and Storz, G. (2000). Redox sensing by prokaryotic transcription factors. Biochem Pharmacol *59*, 1-6.
151. Paget, M.S., and Buttner, M.J. (2003). Thiol-based regulatory switches. Annu Rev Genet *37*, 91-121.
152. Beeby, M., O'Connor, B.D., Ryttersgaard, C., Boutz, D.R., Perry, L.J., and Yeates, T.O. (2005). The genomics of disulfide bonding and protein stabilization in thermophiles. PLoS Biol *3*, e309.
153. Ladenstein, R., and Ren, B. (2007). Reconsideration of an early dogma, saying "there is no evidence for disulfide bonds in proteins from archaea". Extremophiles.
154. Ladenstein, R., and Ren, B. (2006). Protein disulfides and protein disulfide oxidoreductases in hyperthermophiles. Febs J *273*, 4170-4185.
155. Kashima, Y., and Ishikawa, K. (2003). A hyperthermostable novel protein-disulfide oxidoreductase is reduced by thioredoxin reductase from hyperthermophilic archaeon Pyrococcus horikoshii. Arch Biochem Biophys *418*, 179-185.
156. Grunden, A.M., Jenney, F.E., Jr., Ma, K., Ji, M., Weinberg, M.V., and Adams, M.W. (2005). In vitro reconstitution of an NADPH-dependent superoxide reduction pathway from Pyrococcus furiosus. Appl Environ Microbiol *71*, 1522-1530.
157. Ward, D.E., Donnelly, C.J., Mullendore, M.E., van der Oost, J., de Vos, W.M., and Crane, E.J., 3rd (2001). The NADH oxidase from Pyrococcus furiosus. Implications for the protection of anaerobic hyperthermophiles against oxidative stress. Eur J Biochem *268*, 5816-5823.
158. Williams, E., Lowe, T.M., Savas, J., and DiRuggiero, J. (2007). Microarray analysis of the hyperthermophilic archaeon Pyrococcus furiosus exposed to gamma irradiation. Extremophiles *11*, 19-29.
159. Jacob, C., Knight, I., and Winyard, P.G. (2006). Aspects of the biological redox chemistry of cysteine: from simple redox responses to sophisticated signalling pathways. Biol Chem *387*, 1385-1397.
160. D'Autreaux, B., Pecqueur, L., de Peredo, A.G., Diederix, R.E., Caux-Thang, C., Tabet, L., Bersch, B., Forest, E., and Michaud-Soret, I. (2007). Reversible redox- and zinc-dependent dimerization of the Escherichia coli fur protein. Biochemistry *46*, 1329-1342.
161. Ilbert, M., Graf, P.C.F., and Jakob, U. (2006). Zinc center as redox switch - New function for an old motif. Antioxidants & Redox Signaling *8*, 835-846.

APPENDIX

Table A1. Table of abbreviations

Abbreviation	Definition
Basal transcription protein and DNA elements	
BRE	TFB recognition element, purine-rich core promoter element
ORF	Open reading frame
RNAP	RNA polymerase
Pol II	Eukaryotic RNA polymerase II
TATA box	TBP recognition element, T/A-rich core promoter element
TBP	TATA binding protein
TFB	Transcription factor B
Protein names	
NSR	NAD(P)H sulfur reductase of *P. furiosus*
PDO	Protein disulfide oxidoreductase of *P. furiosus*
Phr	Heat shock regulator of *P. furiosus*
SurR	Sulfur response regulator of *P. furiosus*
Gene names	
surr	Gene encoding sulfur response regulator of *P. furiosus* (PF0095)
hydB1	First ORF in *P. furiosus* hydrogenase I operon (PF0891)
mbh1	First ORF in *P. furiosus* membrane-bound hydrogenase operon (PF1423)
pdo	Gene encoding protein disulfide oxidoreductase (PF0094)
Miscellaneous terms	
cys	cysteine
his-tag	Hexahistidine or his_6 tag
HTH	Helix-turn-helix DNA-binding domain
LB	Lauria-Bertani broth
OD_{600}	Optical density at 600 nm
OH•	Hydroxyl radical
RMS	Root mean square
S^0	Elemental sulfur
T_m	Melting temperature
UOR	Upstream of ORF region
UV	Ultraviolet
wHTH	Winged helix-turn-helix DNA-binding domain
Methods	
EMSA	Electromobility shift assay
PCR	Polymerase chain reaction
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SELEX	Systematic evolution of ligands by exponential enrichment

Chemical names	
DTNB	5,5′-Dithiobis(2-nitrobenzoic acid), Ellman's Reagent
DTT	Dithiothreitol
IPTG	Isopropyl-β-D-thiogalactopyranoside
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
Unit definitions	
Å	Angstrom
µg	microgram
µL	microliter
µM	micromolar
bp	base pair
Da	Dalton
kDa	kiloDalton
L	Liter
mg	milligram
min	minute
mL	milliliter
nm	nanometer
rpm	revolutions per minute
U	unit
V	volt

Table A2. UOR database search results[1] for the motif GTTn$_3$AAC

UOR Name[a]	Motif	UOR Start[b]	UOR Stop[b]	Genome Start	Genome Stop	Annotation[c]
PF0003	GTTAGTAAC	-112	-104	3444	3452	moeB-like protein
PF0034	GTTGAGAAC	-169	-161	41144	41152	hypothetical protein
PF0038	GTTTAGAAC	-106	-98	45607	45615	glyoxalase II family member
PF0047	GTTTGTAAC	-190	-182	55974	55982	hypothetical protein
PF0055	GTTTTTAAC	-58	-50	62333	62341	putative HTH transcription regulator
PF0094	GTTATAAAC	-30	-22	103221	103229	glutaredoxin-like protein
PF0095	GTTTATAAC	-111	-103	103229	103221	hypothetical protein
PF0094	GTTGACAAC	-74	-66	103265	103273	glutaredoxin-like protein
PF0095	GTTGTCAAC	-67	-59	103273	103265	hypothetical protein
PF0147	GTTCATAAC	-199	-191	156648	156656	potassium channel related protein
PF0157	GTTTTTAAC	-41	-33	168839	168847	D-fructose-6-phosphate amidotransferase
PF0177	GTTTGAAAC	-50	-42	186065	186057	ATP synthase subunit I
PF0207	GTTGTGAAC	-180	-172	222673	222665	argininosuccinate synthase
PF0254	GTTGAGAAC	-71	-63	263369	263377	hypothetical protein
PF0283	GTTGTAAAC	-118	-110	293740	293732	putative acylphosphatase
PF0315	GTTTTCAAC	-148	-140	330297	330305	hypothetical protein
PF0335	GTTAAAAAC	-84	-76	350529	350537	flagella-related protein d, putative
PF0339	GTTAATAAC	-111	-103	353523	353531	methyltransferase
PF0342	GTTGGGAAC	-123	-115	355692	355700	hypothetical protein
PF0350	GTTTGCAAC	-171	-163	364077	364085	putative cation antiporter
PF0360	GTTTATAAC	-81	-73	373902	373894	oligopeptide ABC transporter
PF0369	GTTGAAAAC	-115	-107	384503	384511	putative endoglucanase
PF0388	GTTGTAAAC	-54	-46	397599	397607	hypothetical protein
PF0389	GTTTACAAC	-101	-93	397607	397599	transposase
PF0393	GTTTTCAAC	-89	-81	401354	401346	hypothetical protein
PF0439.3n	GTTGATAAC	-154	-146	451153	451145	hypothetical protein
PF0443	GTTGGGAAC	-69	-61	460874	460882	putative membrane transport protein
PF0470	GTTAAAAAC	-31	-23	485912	485920	hypothetical protein
PF0529	GTTTGGAAC	-48	-40	546842	546850	hypothetical protein
PF0531	GTTGGTAAC	-84	-76	547743	547751	cobalamin biosynthesis protein m
PF0532	GTTACCAAC	-100	-92	547751	547743	hypothetical protein
PF0538	GTTCATAAC	-121	-113	558421	558429	hypothetical protein
PF0547	GTTTTCAAC	-182	-174	566575	566583	hypothetical protein
PF0548	GTTGAAAAC	-67	-59	566583	566575	hydrogenase expression/formation protein
PF0557	GTTCGGAAC	-79	-71	574517	574525	3-ketoacyl-(acyl-carrier-protein) reductase
PF0558	GTTCCGAAC	-116	-108	574525	574517	cation efflux system protein
PF0559	GTTTCTAAC	-48	-40	575553	575545	hydrogenase expression/formation regulatory protein
PF0568	GTTTATAAC	-24	-16	590847	590855	hypothetical protein
PF0569	GTTATAAAC	-26	-18	590855	590847	hypothetical protein
PF0576	GTTTCTAAC	-134	-126	596314	596322	hypothetical protein
PF0576	GTTTTCAAC	-173	-165	596353	596361	hypothetical protein

[1] The *P. furiosus* UOR database and corresponding motif searching software was created by Darin Cowart (University of Georgia).

UOR Name[a]	Motif	UOR Start[b]	UOR Stop[b]	Genome Start	Genome Stop	Annotation[c]
PF0583	GTTGAAAAC	-106	-98	603286	603294	daunorubicin resistance ATP-binding protein
PF0588	GTTTTTAAC	-40	-32	608204	608212	phospho-sugar mutase
PF0597	GTTCTTAAC	-176	-168	617545	617537	hypothetical iaa-amino acid hydrolase 1 precursor
PF0606	GTTCATAAC	-59	-51	628097	628089	hypothetical protein
PF0656	GTTTTCAAC	-102	-94	669051	669059	hypothetical protein
PF0660	GTTTTCAAC	-109	-101	672080	672088	iron-sulfur cluster binding protein
PF0673	GTTGAAAAC	-38	-30	683729	683721	methylmalonyl-CoA decarboxylase gamma chain
PF0675	GTTAGTAAC	-24	-16	685443	685435	hypothetical protein
PF0694	GTTTCAAAC	-138	-130	702563	702555	flavoprotein
PF0768	GTTAAAAAC	-99	-91	761500	761508	acetyl / acyl transferase related protein
PF0781	GTTGATAAC	-53	-45	768401	768393	hypothetical protein
PF0799	GTTGACAAC	-65	-57	785016	785024	hypothetical protein
PF0800	GTTGTCAAC	-124	-116	785024	785016	transposase
PF0801	GTTGTCAAC	-133	-125	785024	785016	hypothetical protein
PF0799	GTTCATAAC	-195	-187	785146	785154	hypothetical protein
PF0804	GTTATAAAC	-145	-137	788735	788743	methanol dehydrogenase regulator
PF0840	GTTGCAAAC	-188	-180	813851	813843	hypothetical nucleotidyltransferase
PF0865	GTTAAAAAC	-44	-36	840293	840301	3-octaprenyl-4-hydroxybenzoate carboxy-lyase
PF0866	GTTTTTAAC	-74	-66	840301	840293	hypothetical protein
PF0891	GTTTTTAAC	-142	-134	863620	863612	sulfhydrogenase beta subunit
PF0913	GTTAATAAC	-183	-175	884560	884568	hypothetical protein
PF0914	GTTAATAAC	-67	-59	884560	884568	hypothetical protein
PF0913	GTTCAAAAC	-197	-189	884574	884582	hypothetical protein
PF0914	GTTCAAAAC	-81	-73	884574	884582	hypothetical protein
PF0915	GTTAGAAAC	-58	-50	885386	885394	hypothetical protein
PF0920	GTTTCCAAC	-198	-190	888142	888150	nucleotidyltransferase
PF0926	GTTTGGAAC	-81	-73	892098	892106	hypothetical protein
PF0947	GTTAGCAAC	-176	-168	911549	911557	hypothetical protein
PF1027	GTTACTAAC	-138	-130	983677	983685	hypothetical protein
PF1063	GTTATAAAC	-82	-74	1016376	1016368	hypothetical protein
PF1063	GTTTTTAAC	-30	-22	1016428	1016420	hypothetical protein
PF1065	GTTTATAAC	-176	-168	1018769	1018777	hypothetical protein
PF1087	GTTGTTAAC	-31	-23	1036410	1036418	hypothetical protein
PF1098	GTTTCAAAC	-147	-139	1045905	1045913	hypothetical protein
PF1100	GTTTGCAAC	-108	-100	1047031	1047023	hypothetical protein
PF1126	GTTCGTAAC	-152	-144	1075478	1075486	hypothetical protein
PF1135	GTTGTGAAC	-119	-111	1084791	1084783	heat shock protein HtpX
PF1147	GTTGACAAC	-72	-64	1101031	1101039	NADH dehydrogenase subunit N
PF1164	GTTTGGAAC	-98	-90	1110574	1110566	putative Na+/H+ antiporter
PF1186	GTTTTAAAC	-26	-18	1132897	1132905	NADH oxidase
PF1204	GTTACAAAC	-90	-82	1145394	1145386	seryl-tRNA synthetase
PF1227	GTTAGAAAC	-135	-127	1161999	1161991	transposase
PF1238	GTTGAAAAC	-152	-144	1171247	1171239	putative ABC transporter
PF1253	GTTTATAAC	-42	-34	1188453	1188461	aspartate aminotransferase
PF1255	GTTCGGAAC	-170	-162	1189998	1189990	hypothetical protein
PF1269	GTTTGAAAC	-104	-96	1201641	1201633	methionine synthase
PF1316	GTTCATAAC	-103	-95	1239717	1239709	hypothetical protein
PF1319	GTTTTGAAC	-134	-126	1242031	1242023	hypothetical protein

UOR Name[a]	Motif	UOR Start[b]	UOR Stop[b]	Genome Start	Genome Stop	Annotation[c]
PF1324	GTTAGTAAC	-70	-62	1245432	1245424	hypothetical protein
PF1328	GTTTTGAAC	-79	-71	1249933	1249941	hydrogenase gamma subunit
PF1329	GTTCAAAAC	-88	-80	1249941	1249933	H-II hydrogenase subunit beta
PF1366	GTTGAAAAC	-34	-26	1285122	1285130	transposase
PF1367	GTTTTCAAC	-47	-39	1285130	1285122	50S ribosomal protein L7Ae
PF1412	GTTCCAAAC	-199	-191	1327897	1327889	dipeptide transport ATP-binding protein
PF1412	GTTCCAAAC	-148	-140	1327948	1327940	dipeptide transport ATP-binding protein
PF1414	GTTGGAAAC	-112	-104	1329475	1329467	flap endonuclease-1
PF1422	GTTTATAAC	-11	-3	1337267	1337275	thioredoxin reductase
PF1423	GTTATAAAC	-146	-138	1337275	1337267	hypothetical protein
PF1422	GTTTGAAAC	-126	-118	1337382	1337390	thioredoxin reductase
PF1423	GTTTCAAAC	-31	-23	1337390	1337382	hypothetical protein
PF1453	GTTGTTAAC	-61	-53	1360093	1360101	hypothetical protein
PF1454	GTTAACAAC	-153	-145	1360101	1360093	hypothetical protein
PF1453	GTTTAAAAC	-103	-95	1360135	1360143	hypothetical protein
PF1454	GTTTTAAAC	-111	-103	1360143	1360135	hypothetical protein
PF1467	GTTGGCAAC	-106	-98	1373568	1373560	putative ABC transporter
PF1479	GTTAAAAAC	-95	-87	1382832	1382824	putative oxidoreductase, Fe-S subunit
PF1490	GTTTTAAAC	-34	-26	1391897	1391905	hypothetical protein
PF1491	GTTTAAAAC	-54	-46	1391905	1391897	30S ribosomal protein S17
PF1516	GTTTACAAC	-58	-50	1415022	1415030	GMP synthase subunit B
PF1537	GTTTTTAAC	-57	-49	1436963	1436971	hypothetical protein
PF1561	GTTGACAAC	-129	-121	1455816	1455824	50S ribosomal protein L30
PF1569	GTTTGGAAC	-153	-145	1466829	1466837	exosome complex RNA-binding protein 1
PF1617	GTTTGGAAC	-176	-168	1509509	1509501	hypothetical protein
PF1621	GTTACTAAC	-70	-62	1513968	1513976	hypothetical protein
PF1622	GTTAGTAAC	-196	-188	1513976	1513968	n-type ATP pyrophosphatase superfamily
PF1652	GTTATAAAC	-123	-115	1537640	1537632	hypothetical protein
PF1665	GTTATGAAC	-29	-21	1548857	1548849	histidinol-phosphate aminotransferase
PF1679	GTTAGCAAC	-9	-1	1561854	1561846	3-isopropylmalate dehydratase
PF1684	GTTCAAAAC	-85	-77	1565631	1565623	acetylglutamate kinase
PF1698	GTTAAAAAC	-116	-108	1578254	1578246	putative ribose ABC transporter
PF1715	GTTATAAAC	-43	-35	1595766	1595774	pyrroline-5-carboxylate reductase
PF1749	GTTTATAAC	-80	-72	1626536	1626528	putative sulfate transport integral membrane protein
PF1752	GTTTAGAAC	-159	-151	1629141	1629133	transposase
PF1795	GTTGAGAAC	-79	-71	1667079	1667071	sarcosine oxidase subunit alpha
PF1815	GTTTGAAAC	-90	-82	1679926	1679934	30S ribosomal protein S17
NT01PF2016	GTTATCAAC	-195	-187	1701873	1701881	
PF1864	GTTAGAAAC	-137	-129	1721104	1721112	transposase
PF1878	GTTAAAAAC	-196	-188	1730945	1730937	methylated-DNA--protein-cysteine methyltransferase
PF1896	GTTAGAAAC	-32	-24	1746846	1746838	l-isoaspartyl protein carboxyl methyltransferase
PF1898	GTTTTAAAC	-179	-171	1748875	1748883	hypothetical protein
PF1906	GTTAATAAC	-27	-19	1757229	1757221	adenosylmethionine-8-amino-7-oxononanoate aminotransferase
PF1911	GTTCACAAC	-54	-46	1762897	1762905	hypothetical protein

UOR Name[a]	Motif	UOR Start[b]	UOR Stop[b]	Genome Start	Genome Stop	Annotation[c]
PF1912	GTTGTGAAC	-94	-86	1762905	1762897	hypothetical protein
PF1920	GTTAGAAAC	-52	-44	1771957	1771965	triosephosphate isomerase
PF1927	GTTCTTAAC	-41	-33	1778257	1778265	hypothetical protein
PF1928	GTTAAGAAC	-28	-20	1778265	1778257	ATP-binding protein phnp
PF1940.1n	GTTTATAAC	-42	-34	1793571	1793579	hypothetical protein
PF1975	GTTTAAAAC	-60	-52	1825360	1825352	hypothetical protein
PF1981	GTTAAAAAC	-40	-32	1832121	1832129	hypothetical protein
PF1982	GTTTTTAAC	-40	-32	1832129	1832121	hypothetical protein
PF2011	GTTATAAAC	-142	-134	1858487	1858479	hypothetical protein
PF2047.1n	GTTAAAAAC	-122	-114	1889574	1889582	hypothetical protein
PF2048	GTTTTTAAC	-50	-42	1889582	1889574	hypothetical protein
PF2050	GTTTGGAAC	-101	-93	1892777	1892785	hypothetical protein
PF2051	GTTCCAAAC	-26	-18	1892785	1892777	arsr family transcriptional regulator
PF2060.1n	GTTTAAAAC	-12	-4	1901228	1901236	hypothetical arylsulfatase regulatory protein
PF2064	GTTATTAAC	-67	-59	1905977	1905985	arylsulfatase regulatory protein, putative

[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 REFSEQ, GENBANK or TIGR.

Table A3. UOR database search results[1] for the motif AACn$_5$GTT

UOR Name[a]	Motif	UOR Start[b]	UOR Stop[b]	Genome Start	Genome Stop	Annotation[c]
PF0023	AACTTCAAGTT	-99	-89	26727	26737	hypothetical protein
PF0041	AACTTCAGGTT	-63	-53	48431	48441	putative oxidative cyclase
PF0074	AACATGAAGTT	-150	-140	79422	79412	alcohol dehydrogenase, short chain
PF0079	AACTTTTGGTT	-94	-84	85409	85419	cyclic 2,3-diphospoglycerate-synthetase
PF0080	AACCAAAAGTT	-40	-30	85419	85409	hypothetical protein
PF0094	AACTCTAGGTT	-52	-42	103241	103251	glutaredoxin-like protein
PF0095	AACCTAGAGTT	-91	-81	103251	103241	hypothetical protein
PF0094	AACCTAAGGTT	-82	-72	103271	103281	glutaredoxin-like protein
PF0095	AACCTTAGGTT	-61	-51	103281	103271	hypothetical protein
PF0119	AACGGATTGTT	-53	-43	125178	125188	putative perplasmic sugar binding protein
PF0134	AACCTATAGTT	-62	-52	144544	144554	hypothetical protein
PF0135	AACTATAGGTT	-63	-53	144554	144544	lysyl-tRNA synthetase
PF0144	AACAATGGGTT	-176	-166	153555	153545	hypothetical protein
PF0155	AACCGGGGGTT	-165	-155	163948	163958	asparaginyl-tRNA synthetase
PF0238	AACCAGCGGTT	-176	-166	249992	250002	hypothetical protein
PF0275	AACAAAGCGTT	-194	-184	285599	285589	Na+/H+ antiporter
PF0279	AACGTCAAGTT	-124	-114	292413	292423	putative ABC transporter
PF0296	AACTTTAAGTT	-154	-144	312041	312051	cobalamin biosynthesis protein
PF0306	AACGTAAAGTT	-98	-88	323019	323029	sua5 superfamily-related protein
PF0312	AACGAAAAGTT	-47	-37	329099	329109	ADP-specific phosphofructokinase
PF0322	AACTTAACGTT	-49	-39	337801	337811	nickel responsive regulator
PF0432.3n	AACTCAAAGTT	-74	-64	443332	443342	hypothetical protein
PF0443	AACCTCAAGTT	-200	-190	461003	461013	putative membrane transport protein
PF0485	AACTTAACGTT	-86	-76	503277	503287	cell division inhibitor minD - like protein
PF0486	AACGTTAAGTT	-38	-28	503287	503277	hypothetical protein
PF0487	AACTTTGTGTT	-76	-66	505223	505233	related to bacterial autolysins
PF0488	AACACAAAGTT	-124	-114	505233	505223	30S ribosomal protein S6
PF0531	AACAAAATGTT	-78	-68	547735	547745	cobalamin biosynthesis protein m
PF0532	AACATTTTGTT	-108	-98	547745	547735	hypothetical protein
PF0531	AACCTTACGTT	-92	-82	547749	547759	cobalamin biosynthesis protein m
PF0532	AACGTAAGGTT	-94	-84	547759	547749	hypothetical protein
PF0545	AACAGGCAGTT	-195	-185	563768	563778	hypothetical protein
PF0546	AACCATTAGTT	-125	-115	564950	564960	hypothetical protein
PF0547	AACTTGAAGTT	-190	-180	566581	566591	hypothetical protein
PF0548	AACTTCAAGTT	-61	-51	566591	566581	hydrogenase expression/formation protein
PF0552	AACCATTGGTT	-59	-49	570056	570066	arsenical-resistance protein acr3

[1] The *P. furiosus* UOR database and corresponding motif searching software was created by Darin Cowart (University of Georgia).

UOR Name[a]	Motif	UOR Start[b]	UOR Stop[b]	Genome Start	Genome Stop	Annotation[c]
PF0553	AACCAATGGTT	-92	-82	570066	570056	arsenate reductase
PF0559	AACCAAAGGTT	-74	-64	575529	575519	hydrogenase expression/formation regulatory protein
PF0559	AACTTTTGGTT	-42	-32	575561	575551	hydrogenase expression/formation regulatory protein
PF0563	AACAAAGCGTT	-176	-166	580071	580061	hypothetical protein
PF0568	AACGCTATGTT	-18	-8	590839	590849	hypothetical protein
PF0569	AACATAGCGTT	-34	-24	590849	590839	hypothetical protein
PF0578	AACGAAGAGTT	-107	-97	597050	597040	hypothetical protein
PF0594	AACTTCACGTT	-189	-179	616280	616290	ornithine carbamoyltransferase
PF0603	AACTTTGAGTT	-185	-175	625539	625529	hypothetical protein
PF0632	AACTTAGCGTT	-25	-15	645430	645440	hypothetical protein
PF0632	AACGGTTGGTT	-128	-118	645533	645543	hypothetical protein
PF0636	AACGAAAGGTT	-175	-165	650300	650310	DEXX-box atpase
PF0646	AACAACAAGTT	-83	-73	659203	659213	hypothetical protein
PF0687.1n	AACTCAAAGTT	-121	-111	695583	695573	hypothetical protein
PF0688	AACTAAGTGTT	-124	-114	698965	698955	subtilisin-like protease
PF0704	AACAGGAAGTT	-178	-168	709163	709173	hypothetical protein
PF0715	AACAAGAAGTT	-139	-129	715746	715756	NAD(P)H oxidase
PF0736	AACAAGTTGTT	-25	-15	732381	732391	hypothetical protein
PF0740	AACATTTGGTT	-78	-68	735957	735947	heavy-metal transporting cpx-type atpase
PF0758	AACAAAATGTT	-49	-39	753429	753439	hypothetical protein
PF0783	AACTCTCGGTT	-114	-104	769447	769437	hypothetical protein
PF0806	AACCTAATGTT	-35	-25	789610	789620	hypothetical protein
PF0807	AACCAAAAGTT	-82	-72	791342	791352	hypothetical protein
PF0859	AACTCCAAGTT	-35	-25	834813	834823	hypothetical protein
PF0860	AACTTGGAGTT	-15	-5	834823	834813	btpA family protein
PF0891	AACCTTTGGTT	-136	-126	863628	863618	sulfhydrogenase beta subunit
PF0913	AACCAAAGGTT	-191	-181	884566	884576	hypothetical protein
PF0914	AACCAAAGGTT	-75	-65	884566	884576	hypothetical protein
PF0915	AACCTTAGGTT	-52	-42	885378	885388	hypothetical protein
PF0915	AACATCTAGTT	-66	-56	885392	885402	hypothetical protein
PF0926	AACCTATTGTT	-75	-65	892090	892100	hypothetical protein
PF0926	AACTTATTGTT	-89	-79	892104	892114	hypothetical protein
PF0933	AACTTTTCGTT	-125	-115	897405	897415	DNA repair helicase rad3, putative
PF0952	AACCGTACGTT	-75	-65	913874	913864	hypothetical protein
PF0958	AACTTCATGTT	-127	-117	919394	919404	hypothetical protein
PF0971	AACCAAGGGTT	-119	-109	931105	931115	indolepyruvate ferredoxin oxidoreductase
PF0972	AACCCTTGGTT	-130	-120	931115	931105	hypothetical protein
PF0981	AACTTTTAGTT	-87	-77	937534	937524	hypothetical efflux transport protein
PF0984	AACAAGAAGTT	-142	-132	940468	940478	hypothetical protein
PF1031	AACCAGGGGTT	-97	-87	988107	988117	divalent cation tolerance protein
PF1047	AACGATGAGTT	-148	-138	1000260	1000250	hypothetical protein

UOR Name[a]	Motif	UOR Start[b]	UOR Stop[b]	Genome Start	Genome Stop	Annotation[c]
PF1057	AACAACATGTT	-87	-77	1011386	1011376	phosphoglycerate kinase
PF1088	AACAGCATGTT	-134	-124	1036847	1036857	hypothetical protein
PF1100	AACTCGTAGTT	-102	-92	1047039	1047029	hypothetical protein
PF1117	AACTTTGAGTT	-119	-109	1065907	1065897	hypothetical protein
PF1129	AACATATAGTT	-87	-77	1080421	1080431	hypothetical protein
PF1161	AACAGAGGGTT	-40	-30	1110387	1110397	hypothetical protein
PF1163	AACCCTCTGTT	-78	-68	1110397	1110387	hypothetical protein
PF1186	AACTAAAGGTT	-34	-24	1132903	1132913	NADH oxidase
PF1188	AACTCTTAGTT	-92	-82	1133979	1133969	pyruvate kinase
NT01PF1351	AACATAGAGTT	-190	-180	1143352	1143342	
PF1202	AACATAGAGTT	-108	-98	1143352	1143342	hypothetical protein
PF1247.1n	AACTTTTTGTT	-154	-144	1183168	1183178	hypothetical protein
PF1248	AACTTTTTGTT	-157	-147	1183168	1183178	transposase
PF1266	AACTCCATGTT	-149	-139	1198529	1198519	cystathionine gamma-synthase
PF1268	AACGAGGTGTT	-48	-38	1200768	1200758	5-methyltetrahydropteroyltriglutamate- - homocysteine methyltransferase
PF1321	AACGATAGGTT	-25	-15	1242871	1242861	DEXX-box atpase
PF1328	AACCTTAGGTT	-87	-77	1249939	1249949	hydrogenase gamma subunit
PF1329	AACCTAAGGTT	-82	-72	1249949	1249939	H-II hydrogenase subunit beta
PF1365	AACTTCGTGTT	-104	-94	1283735	1283725	hypothetical protein
PF1364	AACATAAGGTT	-43	-33	1283874	1283884	hypothetical protein
PF1377	AACTAGCAGTT	-52	-42	1292942	1292952	transcription initiation factor IIB
PF1378	AACTGCTAGTT	-106	-96	1292952	1292942	ribonuclease P protein component 2
PF1408	AACTTTTTGTT	-153	-143	1322564	1322554	putative dipeptide-binding protein
PF1420	AACATTAAGTT	-154	-144	1334098	1334088	hypothetical protein
PF1422	AACTTTTGGTT	-19	-9	1337273	1337283	thioredoxin reductase
PF1423	AACCAAAAGTT	-140	-130	1337283	1337273	hypothetical protein
PF1422	AACATTTGGTT	-134	-124	1337388	1337398	thioredoxin reductase
PF1423	AACCAAATGTT	-25	-15	1337398	1337388	hypothetical protein
PF1477	AACTTTAGGTT	-61	-51	1382311	1382321	methylmalonyl-CoA mutase
PF1478	AACCTAAAGTT	-96	-86	1382321	1382311	hypothetical protein
PF1481	AACCTTTGGTT	-60	-50	1385346	1385336	hypothetical protein
PF1516	AACTTTATGTT	-52	-42	1415014	1415024	GMP synthase subunit B
PF1520	AACGATGCGTT	-94	-84	1419855	1419865	ABC transporter integral membrane protein
PF1568	AACGGGGAGTT	-66	-56	1465971	1465981	exosome complex exonuclease 1
PF1621	AACTTTTAGTT	-78	-68	1513974	1513984	hypothetical protein
PF1622	AACTAAAAGTT	-190	-180	1513984	1513974	n-type ATP pyrophosphatase superfamily
PF1684	AACATAATGTT	-73	-63	1565645	1565635	acetylglutamate kinase
PF1716	AACAATGGGTT	-156	-146	1596360	1596370	hypothetical protein
PF1747	AACTAACTGTT	-78	-68	1624581	1624571	hypothetical protein
PF1792	AACCAAAAGTT	-146	-136	1664951	1664941	hypothetical 5'-cyclic-nucleotide phosphodiesterase cpda
PF1794	AACTGGAGGTT	-19	-9	1666511	1666501	inosine-5'-monophosphate

UOR Name[a]	Motif	UOR Start[b]	UOR Stop[b]	Genome Start	Genome Stop	Annotation[c]
						dehydrogenase related protein I
PF1831	AACAAAAGGTT	-25	-15	1689350	1689340	archaeal histone a1
PF1837	AACCGATAGTT	-164	-154	1693366	1693376	hypothetical protein
PF1838	AACCTTTCGTT	-102	-92	1694729	1694739	hypothetical protein
NT01PF2012	AACGAAAGGTT	-71	-61	1694739	1694729	
PF1851	AACTCAGTGTT	-47	-37	1709264	1709274	hypothetical protein
PF1881	AACATTTCGTT	-38	-28	1733860	1733870	chromatin protein
PF1899	AACTAATCGTT	-178	-168	1748565	1748555	putative 2-deoxy-d-gluconate 3-dehydrogenase
PF1911	AACCTTTGGTT	-90	-80	1762931	1762941	hypothetical protein
PF1912	AACCAAAGGTT	-60	-50	1762941	1762931	hypothetical protein
PF1959	AACAAAGCGTT	-111	-101	1808800	1808790	cofactor-independent phosphoglycerate mutase
PF1960	AACTTAGCGTT	-63	-53	1810191	1810181	aldose reductase
PF1978	AACAAGAAGTT	-180	-170	1829066	1829056	nicotinate-nucleotide pyrophosphorylase
PF1988	AACGGGGGGTT	-17	-7	1839097	1839087	cell division protein FtsZ
PF2034	AACAAAGCGTT	-135	-125	1877906	1877916	hypothetical protein

[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 REFSEQ, GENBANK or TIGR.