"Not a toxin but a regulator:" Structure guided conclusions

by

Peter S. Horanyi

(Under the guidance of Bi-Cheng Wang)

ABSTRACT

Toxin antitoxin (TA) operons are encoded on almost all freeliving microorganisms' chromosome. By far the most abundant TA operon family is VapBC, which represents over 42% of total identified TA operons. These proteins were first identified on the bacterial plasmids' fall-back maintenance mechanism, also known as Post-Segregational Killing (PSK). All TA operons to date are known to be bacteriostatic or bacteriocidal when expressed. These attributes together make the TA family a potential alternative as a bacteriocide to address the multi-drug resistance of bacteria. Because of the relative large size of this family, a considerable amount of sequence divergence has been observed, but all known structures are very similar. Many of these toxin antitoxin proteins have been considered to be regulators instead of killers by Gerdes and Hayes, but all TA proteins found to date have some role in inhibiting one of the housekeeping functions of the cell, such as replication or translation. The potential regulatory function of these proteins has not been addressed and must be well understood before their use in medicine.

We report the characterization and first scientific evidence for a unique non-toxic VapBC-like operon in *Pyrococcus furiosus*. The operon investigated here encodes one extra gene besides the TA proteins, a Glycohydrolase Superfamily 1 enzyme. The novel

antitoxin protein, PF0355.1p, of this operon binds DNA as shown by DNaseI footprinting. The protein PF0355p that is annotated as the toxin, has no toxic function, but binds adenosine-diphosphate as demonstrated by ITC. PF0355p also protected the PF0355.1p from proteolysis and PF0355.1p inhibited PF0355p's ADP binding. We propose that this operon is proof that these chromosomal TA proteins regulate cellular functions other than death, which is the ultimate way of regulation.

The structure of PF0355p was determined here using Single Wavelength Anomalous Scattering to 2.3Å. The structure of the protein contains eight α helices and five β sheets as seen in the previous VapC homolog structures solved. The carboxy terminal α helix is the main difference among those structures. The other notable differences are the divergent residues that make-up the predicted active sites.

The structure of PF0355 allowed us to demonstrate why this protein lacks enzymatic function. The lack of killer activity demonstrated by PF0355 is also presented here based on the structure of PF0355 and the functional characterization of the proteins. The structure also gave insights as to what makes those plasmid maintenance-associated PIN domains killers. These data suggests that the hundreds of VapBC homologs found in microbes contain divergent functions that are not always killing.

Index Words: Toxin antitoxin (TA), Post-Segregational Killing (PSK), VapBC-like operon in *Pyrococcus furiosus*, PIN domain, Single Wavelength Anomalous Scattering,

"Not a toxin but a regulator:" Structure guided conclusions

by

Peter S. Horanyi

B.S. The University of Georgia, 1998

A Dissertation submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment for the Requirements for the degree

DOCTOR OF PHILOSOPHY

Athens, GA 2005

"Not a toxin but a regulator:" Structure guided conclusions

by

Peter S. Horanyi

Major Professor : Bi-Cheng Wang
Committee: Harry Dailey
Claiborne Glover
John Rose
Robert Woods

Electronic Version Approved

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December, 2005

DEDICATION

I dedicate my hard earned accomplishment to those three women who made it possible: My dearest Mother, who has worked very hard from my birth, by herself, so I always try to be the best in anything I try, so I can represent both her and our country, Hungary, with pride.

Wonderful Tammy Dailey, my teacher, the person that took the time to teach me how to do science in the lab. The person that all too often does not get the recognition that she deserves.

My greatest wife, Lisa Edge Horanyi, the wonder who made sure I ate and slept and had clean clothes throughout graduate school. She always let me share my frustrations with her.

ACKNOWLEDGEMENTS

I would like to extend the warmest thanks to my Major Professor Dr. Bi-Cheng Wang for the all the great opportunity that he has given me. From the time I joined the lab six years ago, only my imagination limited me in my research. Dr. Wang showed great support in all of my ideas and has allowed me to excel at my science. The influence that I received in this world-renowned group is unforgettable, and I will cherish the memories as long as I live.

I would also like to thank, from the bottom of my heart, Drs. Harry Dailey, Claiborne Glover, John Rose, and Robert Woods, for serving on my advisory committee. They have identified all of my scientific shortcomings and have helped me in correcting them. Their guidance and leadership has aided me to become a scientist from a student. I also feel that they provided me with a way of thinking that nobody can take away from me.

I am grateful to Dr. John Rose, for helping me with the day-to-day workings of science. He has checked on me regularly throughout my long stay in the lab to make sure I am OK. He has also cheered me on many times, when I needed some encouragement.

My heartfelt thanks are extended to Dr. Zhi-Jie (James) Liu for teaching me crystallography. He has spent numerous hours explaining and demonstrating this difficult science to me, which has made a huge difference in my understanding of crystallography. He has also allowed me to “bend his ear” for hours at a time about ideas I had. His friendship and patience made this arduous learning process fun and unforgettable.

Last but not least, I would like to thank the Department of Biochemistry and Molecular Biology's faculty, Dr Puett especially, and graduate students, for all their support and belief in me over the years. I believe that the most memorable six years of my life were spent here.

GO DAWGS!

LIST OF TABLES

TABLE 1 The listing of all known TA operons with information pertinent to our discussion 4

TABLE 2 The VapBC homologs indentified by BLASTP searching of the *Pyrococcus furiosus* genome 15-16

TABLE 3 The Statistics of data collection and final data processing of PF0355 23-24

LIST OF ABBREVIATIONS

ATP; Adenosine Triphosphate

ADP; Adenosine Diphosphate

AMP; Adenosine Monophosphate

cAMP; 2'-3' cyclic Adenosine Monophosphate

CTP; Cytosine Triphosphate

cGMP; 2'-3' cyclic Guanosine Monophosphate

GTP; Guanosine Triphosphate

GDP; Guanosine Diphosphate

GlcNAC; N-Acetyl Glucosamine

Bp; basepairs

Nt; nucleotides

RQ1; RNAse Qualified

AIDS; Autoimmune Deficiency Syndrome

ANL; Argonne National Lab, Argonne, IL

APS; Advanced Photon Source

DNA; Deoxyribonucleic acid

RNA; Ribonucleic acid

DNAse; DNA degrading enzyme, also nuclease

LIST OF FIGURES

FIGURE 1 Model of toxin-antitoxin operons with the *Repressor Regulator* operons contrasted 5-7

FIGURE 2 The proposed model of the revised nomenclature of the TA operons. The *Repressor Regulator* operons. 17

FIGURE 3 SECSG Sca2Structure pipeline input screen 26

FIGURE 4 The refinement history of 1Y82 27-29

FIGURE 5 The region of PF0355p that is conserved with its closest homologs 32-33

FIGURE 6 The investigation of DNA binding by PF0355.1p 38-39

FUGIRE 7 Temperature factors plotted versus residue number 44-45

FIGURE 8. The structural evaluation of PF0355 46-49

FIGURE 9 Sample electron density from the final map used during refinement 50

FIGURE 10 The binding of ADP by PF0355p, the *Regulator,* demonstrated by Isothermal Titration Calorimetry (ITC) 53-55

FIGURE 10 The interaction of PF0355p, the *Regulator,* and the *Repressor* PF0355.1p 60

TABLE OF CONTENTS

ACKNOWLEDGMENTS vi

LIST OF TABLES vii

LIST OF ABBREVIATIONS viii

LIST OF FIGURES ix

1.0 Literature review 1

1.1 Background 1

1.2 The history of proteic PSK 3

1.3 Chromosomally encoded toxin antitoxin operons 5

1.4 The function of the toxins 8

1.5 The function of the antitoxin 10

2.0 Introduction 11

3.0 Materials and Methods 18

3.1 Purification of PF0355p and PF0355.1p 18

3.2 DNA binding by PF0355.1p 19

3.3 DNase I footprinting of PF0355.1p binding the promoter 20

3.4 Crystallization/data collection of PF0355p 21

3.5 Phasing of PF0355 data 25

3.6 Structural comparison 30

3.7 Structure prediction for alignment 30

3.8 Isothermal Titration Calorimetry 34

3.9 Interaction of PF0355p and PF0355.1p 34

4.0 Results 36

4.1 DNA binding by PF0355.1p 36

4.2 DNAseI footprinting of PF0355.1p 40

4.3 PF0355 structure and comparison to structural homologs 41

4.4 ITC of PF0355 51

4.5 Predicted structure-based sequence alignment of PF0355p with select PIN domain-containing proteins. 56

4.6 Interaction of PF0355p and PF0355.1p 57

4.7 Prediction of β- glycosidase structure to infer function 58

5.0 Discussion 61

6.0 References 67

1.0 Literature review

1.1 Background

In medicine one of the most pressing problems is the emergence of antibiotic resistant strains of bacteria [1]. These organisms can cause acute and chronic infections in humans with potentially lethal consequences. They are especially dangerous to those individuals immune-compromised from age and auto-immune disorders like Lupus or AIDS. The emergence of antibiotic resistance by bacteria is known to occur through the development of novel enzymatic functions or mutations of current enzymes like acetyl transferases and others[1]. It seems that the constant development of new antibiotics is not an effective way of achieving the desired bacteriocidal effect [1].

It has been proposed by Edelberg-Kulka and others that the solution of killing antibiotic resistant bacteria may reside within the microbes themselves [1]. Microorganisms propagate by asexual reproduction, which uses division as the method of propagation. This process does not allow the merging of genetic material in the offspring. One way the microbes exchange genetic material is the use of extra-chromosomal self-maintaining DNA fragments. A class of these DNA molecules is called plasmids. Plasmids are 10-80 kilobase (kb) extrachromosomal self-propagating DNA molecules that encode specific genes that have a variety of roles such as antibiotic resistance, virulence and other controlling functions. The naturally occurring plasmids are maintained in the cells by three mechanisms: 1. Centromere like systems that prior to cell division actively secure segregation of replicons [2], 2. Site-specific recombination systems that separate tandem plasmid multimers into monomers [2],

and 3. addiction modules that mediate the killing of newborn plasmid-free cells. The third system mentioned is the one under evaluation here.

This review on the plasmid maintenance literature focuses on a family of genes related to the proteic plasmid retention mechanism called Post Segregational Killing (PSK) [3]. The specific purpose of this section is to review diversity of the PSK subfamily called Toxin-Antitoxin (TA) operons. We will also point out the ambiguity of the nomenclature of this class of proteins and emphasize a need for new uniform terminology.

The family of TA operons has been found in almost all sequenced free-living bacterial chromosomes, and because of its potential association with bacterial cell death, it has been investigated widely [1, 4-7]. Since this family of TA operons is present in both Gram negative and Gram positive bacteria and has the potential to kill the cell, it presents itself as a viable target of antimicrobial research. The potential that these inherent toxins can be turned on to kill the cell, instead of using "improved" antibiotics, can help in the development of drugs as an alternative to current methods. Understanding the potential role of these TA operons is essential in identifying ways to specifically activate certain inherent toxins of bacteria to achieve the desired bacteriocidal effect [1]. On the other hand, if all classes of TA operons are not identified / characterized, the activation of one class of toxins may trigger undesired effects, such as non-lethal outcomes or improved stress response to current antibiotics.

The TA proteins considered herein represent hundreds of proteins; still only one half of the plasmid addiction modules are discussed. Those PSK systems that work through RNA-level inhibition are a separate family and will not be reviewed here [8]. The body of this thesis deals with a PSK sub-family called VapBC related proteins. This operon was first identified on low-

copy number Virulence Associated Plasmids (VAP) in the genera *Salmonella* and *Shigelia* [9-11].

1.2 The history of proteic PSK

The first PSK genes names CcdA and CcdB were identified in 1983 by Ogura et. al [12]. These genes are the foundation for the model of TA operons today (Figure 1). These two genes ensure the propagation of the Escherichia coli F plasmid [12]. They perform this role through the killing function of the toxin, CcdB [13]. The antitoxin protein CcdA inhibits the function of the toxin CcdB [14]. Another function of CcdA is to repress the operon. The antitoxin has a short lifetime but CcdB is resistant to proteolysis [15]. In the cell these proteins auto repress their expression but are leaky [16]. This behavior allows them to be present in the cytoplasm at all times. When the cell divides, the daughter cell that did not receive the plasmid encoding the TA operon retains the proteins CcdA and CcdB in its cytoplasm. In the absence of their plasmid, over time, CcdA degrades and CcdB is free to perform its function, to kill the cell, see Figure 1.

These CcdAB attributes are conserved among all TA operons. CcdB's lethal function works through stabilization of DNA gyrase covalent intermediates, which in turn causes double-stranded breaks in DNA resulting cell death [17, 18].

Six other TA operon families have been identified (Table 1). ParD/pem of R1/R100 [19], VapBC of a *Salmonella dublin* virulence plasmid [10], phd/doc of P1 [20, 21], parDE of RK2 [21], higBA of Rts1 [22], and relBE of P307 [23]. The TA loci that belong to these seven families have similar gene organization and overall regulation and phenotypes [24]. The only exception is higBA, which has a reversed gene order (higB toxin gene is located upstream of higA that encodes the antitoxin) [7]. It is also important to note that although the phenotypes of

these seven families are similar, the toxins work through a variety of mechanisms and targets [25].

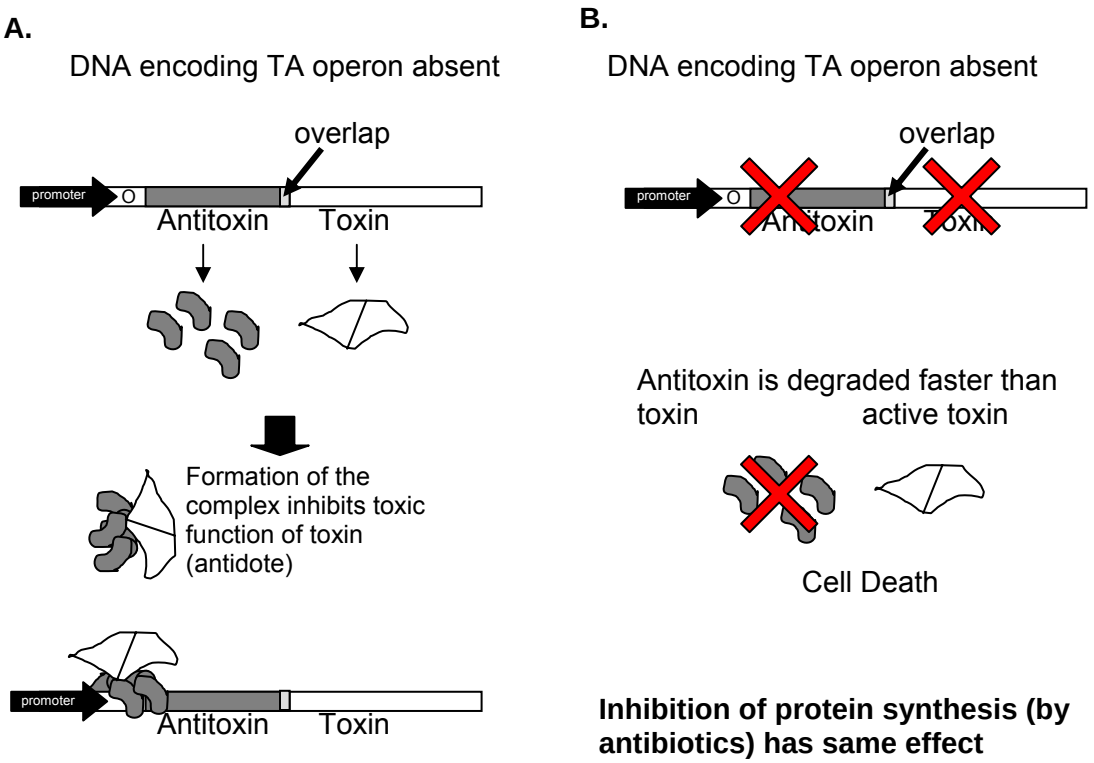
Table 1.

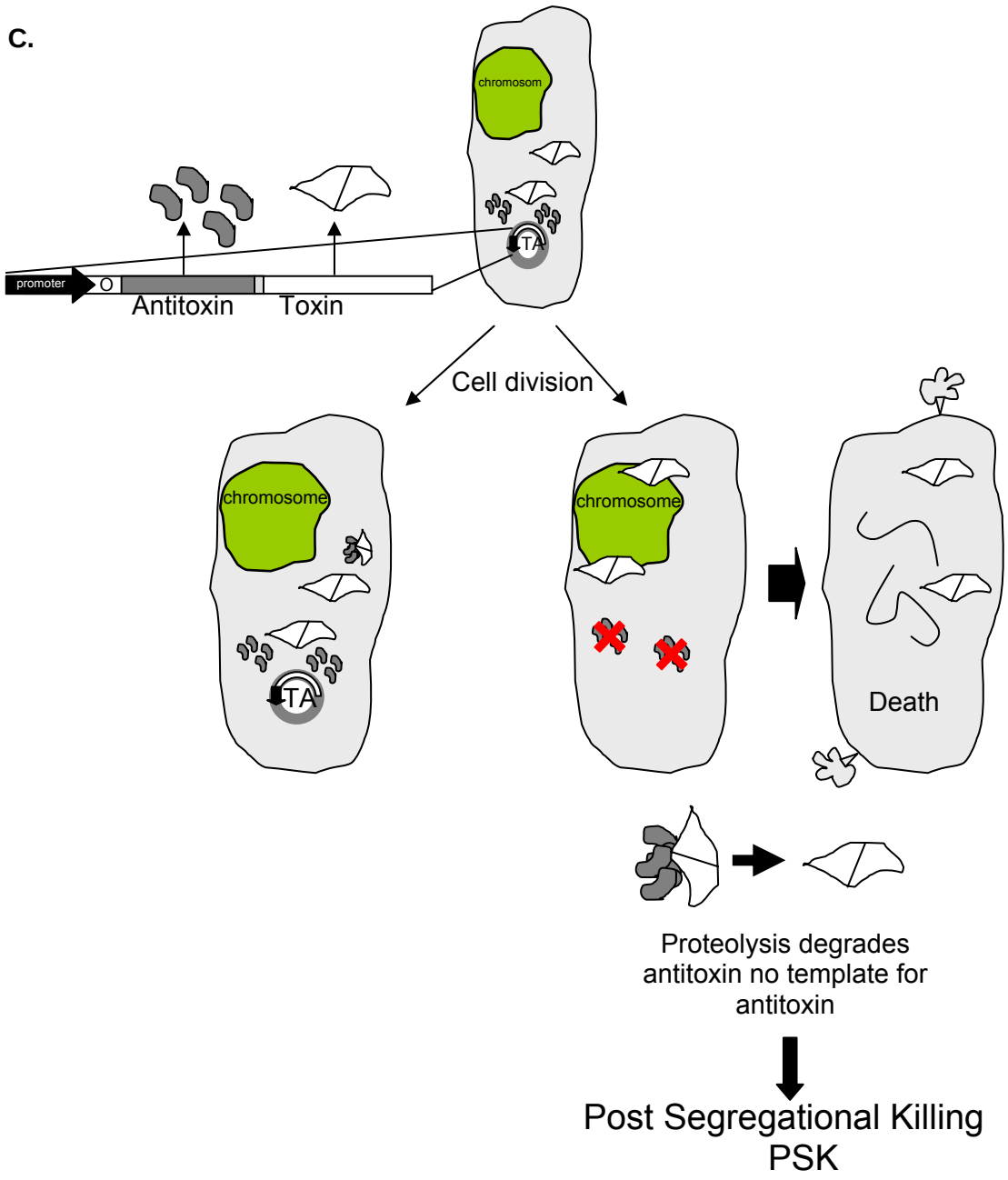
TA family	*ccd*	*relBE*	*parDE*	*higBA*	*phd/ doc*	*VapBC*	*MazEF*
Toxin	CcdB	RelE	ParE	HigB	Doc	VapC	MazF
Target	DNA gyrase	mRNA in ribosome dependent fashion	DNA gyrase	Unknown	Translation	Unknown	Unknown
Anti toxin	CcdA	RelB	ParD	HigA	Phd	VapB	MazE
Protease	Lon	Lon	Unknown	Unknown	ClpAP	Unknown	Unknown
number	5	156	59	74	25	285	67

Table 1. The listing of all known TA operons with information pertinent to our discussion. Data sourced from Pandey *et al.* and Gerdes *et al.* [24, 25].

Figure 1. The current hypothesis on the toxicity of the antitoxin-toxin systems. A. The classical TA operon model with the 2 genes depicted in gray for antitoxin and white for the toxin. The toxin is not toxic in the presence of its own DNA which allows for the production of both the toxin and the antitoxin. **B.** The toxin is toxic in the absence of its DNA, as no more antitoxin is produced and antitoxin remaining in the cell is degraded. This is termed as Post Segregational Killing (PSK). Inhibition of protein synthesis by antibiotics has same effect.
C. The cell always contains both the toxin and antitoxin, due to the mechanism that controls expression. When the cell divides, both daughter cells contain toxin and antitoxin. The daughter cell that does not contain the DNA (plasmid) encoding the TA operon, will be killed by activation of the toxin. The toxin is activated by the degradation of the antitoxin in the absence of the TA plasmid.

Figure 1


A.
DNA encoding TA operon absent
overlap
promoter
O
Antitoxin
Toxin
Formation of the complex inhibits toxic function of toxin (antidote)
promoter
Antitoxin
Toxin
B.
DNA encoding TA operon absent
overlap
promoter
O
Antitoxin is degraded faster than toxin
active toxin
Cell Death
Inhibition of protein synthesis (by antibiotics) has same effect


C.
chromosom
TA
promoter
Antitoxin
Toxin
Cell division
chromosome
TA
chromosome
Death
Proteolysis degrades antitoxin no template for antitoxin
Post Segregational Killing
PSK

1.3 Chromosomally encoded toxin antitoxin operons

Although these genes were initially found on plasmids, they have been shown to be widespread on chromosomes of all free-living microbial genomes [24]. Initially, the role of these genes was proposed to be restricted to the killing of the plasmid free daughter cells. However, their widespread chromosomal presence makes this plasmid based hypothesis questionable. The first identified chromosomal TA operon was the *maz*EF operon from *E. coli* [26], where these genes were found to regulate bacterial programmed cell death. Several recent publications have highlighted the potential stress regulating characteristics of these chromosomally encoded TA (cTA) operons [5, 36]. The 126 microbial genome sequences available in 2005 were inspected for the presence of cTA operons [24]. The survey revealed that only obligate, host-associated prokaryotes and only five free-living bacteria had no cTA operons. Over one hundred evaluated free-living microbes had one or more of these cTA operons. Every archaea evaluated had cTA operons.

All seven classes of TA operons were found in chromosomal copies with the class of VapBC operons representing the majority with over 42% of the total 671 cTA operons identified [24]. This sub-class of cTA operons represented over 80% of all cTA operons found in the 16 sequenced archaeal genomes [24].

1.4 The function of the toxins

All proteic TA operons reported to date follow the model depicted in Figure 1A. The only noticeable divergence among these proteins is in the function of the toxin, see Table 1. The only TA operons' "toxins" with known functions are CcdB, ParE, RelE, MazF/PemK and Doc [25]. As discussed previously, CcdB acts on DNA gyrase and in turn on DNA replication [18]. Another toxin that works on the same substrate is ParE. RelE and MazF both target translation

[21, 26], since they both act through mRNA cleavage. RelE was proposed to be a ribosome-specific RNAse that cleaves at specific sites: stop codons (UAG > UAA > UGA) and sense codons (UCG and CAG). MazF's action was shown to be independent of ribosomes [27]. Doc is thought to inhibit translation [20]. These toxin proteins are mostly bacteriostatic, two of the seven (*ccd*B, and parE) kill the host, but the others are not lethal to the cell [21]. This shows that the role of these non-toxic "toxins" is to REGULATE cellular response. Killing the cell is also a way of regulation, since death is the ultimate regulation.

It is essential to have a technique for testing bacteriostatic or bacteriocidal effects of these proteins. The Gerdes lab has reported data suggesting that the RelBE class of cTA operons that inhibit translation [28]. This is important, since these genes are encoded on the chromosome of the organism and they are cytotoxic to the host while induced. Recently the crystal structure of the RelBE complex was published from the Archaeon *Pyrococcus horikoshii ot3* [29]. Since the exact method of action or target of this RelBE complex is not known, the structure was insightful. It is important to mention that this hyperthermophilic protein complex had to be expressed in as a complex in *E. coli* because of its cytotoxicity. This complex inhibits translation in an *in vitro* S30 extract-based translation system at 37°C [29]. These data show that testing for inhibition of cell growth during induction of recombinant expression in *E. coli* could be used to identify the toxic function of these proteins. In fact, this method has been used in all publications of TA module characterization of toxic function [29-31].

1.5 The function of the antitoxin

The antitoxin in each of these examples is a labile protein that binds the toxin and inhibits its function [32]. These protein complexes, through the function of the antitoxin, are also known to bind DNA and repress their own operon [32]. The evaluated TA promoters, through their interactions with the antitoxin, have been shown to support basal expression in most steady-state growth conditions. In some cases the regulation of these operons seems more complex than just constitutive expression. Numerous reports show stress-induced expression of these genes beyond those basal levels [15, 23, 35-38].

RelBE and MazEF are induced during starvation, which makes them potential stress regulators [5]. *E. coli* RelBE has been thought to be induced by the SOS response of the organism [14, 20]. These operons have also been predicted to be induced by glucose or amino acid starvation [14, 20]. All of these antitoxins share the characteristics of repression of the operon and repression of the "toxin's" activity. Although the function of the antitoxins are conserved, the sequences of all known antitoxins are diverse [33, 34].

The nomenclature of these proteins makes them an antitoxin, but not all of these proteins are ready to remove the toxicity of other proteins. This is because only two out of those seven named toxin listed in Table 1 are toxic. The confusing nomenclature of these proteins will be addressed.

2.0 Introduction

Antibiotic resistance is an emerging problem associated with modern medical progress. Bacteria in order to prevent their own extinction have come up with novel ways of counteracting all antimicrobial agents known to date. In the past development of new antibiotics has shown no real success in preventing this trend, instead multi-drug-resistant bacteria have started emerging. One solution to this problem may be to use a family of naturally occurring bacteriocidal agents, the plasmid-encoded toxins of Post Segregational Killing (PSK) (Figure 1C) [39]. Initially, these proteins' function was shown to be the killing of the plasmid-cured daughter cells to ensure plasmid retention. These are called toxin and antitoxin proteins and they represent a vast protein family that is found on the chromosomes of almost all known microbes. Chromosomal TA operons (cTA) encode a bacteriocidal or bacteriostatic agent generically called a toxin and its antidote, the antitoxin (Figure 1AB). The sub-family of TA operons that was found on the virulence-associated plasmid (Vap) of infectious bacteria like *Shigellia* and *Salmonella* [9-11] is the focus of this document. These Vap genes are only one of the seven known TA subfamilies, but they represents over 42%, of all known cTA proteins [24].

This subfamily is named VapBC after the two proteins VapB and VapC. VapB (also known as VagC or MvpA) is the antitoxin, and VapC (also known as VagD, MvpT) is the toxin [9, 10]. VapC's domain annotation is the PilT-N-terminus (PIN) domain and is present in all three kingdoms of life [33]. One member of this domain has been shown to be a DNA dependent nuclease [41]. The PIN domain is the conserved protein sequence in this cTA sub-family. In these operons the PIN domain overlaps, and therefore shows transcriptional association with a

very diverse set of VapB genes [33, 34]. VapB proteins amongst all microbes have divergent sequences and have been shown to have members from protein families like Phd (See Table 1) or proteins with conserved motifs like HTH (helix turn helix) [33]. These genes always precede the PIN domain in gene organization, resembling the TA operons of PSK [3]. VapB proteins' function is conserved in spite of their large sequence divergence. These proteins are characterized by performing as the antidote to their respective VapCs. They also regulate their operon's expression by binding their regulatory sequences. The binding is influenced by VapB's short half-life due to protease sensitivity [40]. The VapBC complex not only extends the life of VapB [40] but also affect the regulation of their own expression by binding to the promoter of their genes.

These chromosomal TA (cTA) operons have been postulated by Gerdes and Hayes to be operons encoding regulatory proteins and not killers (Figure 1C) [5, 36]. Although a regulatory function for these proteins has been shown indirectly [10], no direct evidence has validated this hypothesis [10, 11]. The identification of a non-toxic cTA operon would prove that these proteins have a regulatory role in the cell. Our goal was to identify one of these non-toxic VapBC operons.

The VapBC operons are most abundant in Archaea. Evaluation of the sixteen sequenced archaea identifies 133 VapBC homologs [24]. In our model organism *Pyrococcus furiosus (P. furiosus)* fourteen VapBC type operons were found [24]. These VapBC operons can be classified into seven sub-classes, based on their VapB sequences (Table 2A). Subclass 1 has a unique feature. Both of the operons in this subclass contain a third, non-PSK associated gene. These extra genes show transcriptional association with those cTA genes, based on intergenic-distance (<14 nt). In both cases in sub-class 1, these extra proteins are glycosidic enzymes. These

cases are only a fraction of the total number of cTA operons, but this subclass of VapBC proteins gives clues to non-lethal function of this operon. These identified unique cTA proteins have great potential to provide the first direct evidence that cTA operons do not kill the cell, but regulate cellular functions. In order to fully understand the potential bacteriocidal actions of the VapBC family of TA proteins, the non-toxic VapC proteins need to be investigated. Out of those chromosomal VapC homologs published to date, no chromosomally encoded TA operons have been characterized. Moreover no cTA operons have been identified that have only regulatory function and no killing activity.

Those cTA operons predicted by Gerdes and Hayes that lack killing activity, do not fit the description of Toxin and Antitoxin [5, 36]. Here we report the first example of those predicted non-toxic cTA operon. Based on these observations these non-toxic cTA genes should be re-named a *Repressor* instead of antitoxin and *Regulator* instead of toxin, see Figure 2A. The revised nomenclature would allow for the lack of killing activity of these proteins to be recognized. This would overcome the problem of the current nomenclature that implies killing functions to these unique operons, which they do not have.

Closer evaluation of those above mentioned *P. furiosus Repressor Regulator* (RR) operons demonstrates, the extra enzyme seemingly has no functional relationship to PSK (Figure 2A). The classical toxin-antitoxin architecture is maintained in these unique cTA operons, with the enzyme's sequence inserted between the start site of the antitoxin and the self-regulatory sequence bound by it (Figure 2A). This means that the *Repressor Regulator* complex seems to be regulating the expression of this enzyme along with the cTA players (Figure 2A). In these cases the role of those RR proteins seems to be the regulation of the respective enzyme's expression instead of killing.

Herein we report the characterization of a TA operon in *P. furiosus* that also contains an enzyme. In addition, we report the characterization of an antitoxin protein that is unique to *Thermococcales*. We show here that the proteins in these archaeal cTA operons follow the classical TA model but do not have the capability to kill the host or inhibit growth. This information along with the structure of our and other VapC homologs are the first step towards udenrstanding what makes some PIN domains toxic and some non-toxic. Due to the lack of killing activity exhibited by the toxin, the only function of the so called toxin encoded by this operon is to comprise one half of the regulator complex. This makes the PF0356-PF0355 operon not only the potential first identified non-killer cTA operon, but a true regulator of cellular metabolism through constitutive expression of an enzyme necessary for breakdown of the most common carbohydrate linkage on earth. Based on previous predictions and the current physical evidence of regulation and not killing, we propose to name this family of cTA operons *Repressor-Regulator* operons. After comparing previous predictions of the Gerdes and Hayes labs with in-house data, the renaming of this gene sub-family is appropriate [5, 36].

The mechanism of regulation where all the parts to repress the expression of the operon are encoded within the unit itself supports a simple, auto-repressor role. Based upon its widespread and diverse presence of these genes, this may have been the original role of these operons that later evolved as killers. Also the role of this family is undeniable in the regulation of cellular response to specific stimuli like inhibition of translation or cellular stress-induced increase in protease concentration [5, 36]. Death is only one of those responses that can be induced by these cTA operons, and their divergent role is not well represented by the names “toxin” and “antitoxin,” which are fitting for plasmid-encoded TA operons.

Table 2.A. The VapBC homologs identified by BLASTP searching the *Pyrococcus furiosus* genome. The template for searching was the PIN domain. Each of those domains identified were then checked to see if a smaller, sometimes overlapping, upstream gene encodes either a protein with unknown function, a predicted DNA binding protein or known antitoxin is present. Those that fit the model were chosen and further evaluated amongst other VapBC homologs. The subfamilies shown in the table are based on sequence homology of 30% or higher. These 14 VapBC homolog operons can be classified into seven subfamilies. Only one of these subfamilies encodes a protein with a known Pfam family prediction, while the others have no predicted information on them at all, with only archaeal homologs. Each gene from *P.furiosus* listed in the table was searched with BLASTP against their respective genomes to identify potential homologs. Abbreviations used: PAB, *Pyrococcus abysii;* PH, *Pyrococcus horikoshii ot3;* TK, *Thermococcus kodakarensis (KOD1).* **B.** Comparison of all *Repressor-Regulator* domains identified. Based on sequence alignment through NCBI the genes were identified. When comparing *Repressor* proteins it is notable that five out of seven genes were not annotated initially as indicated by the decimal in the Open Reading Frame (ORF) number. The location of the *Repressor's* putative recognition sequence is 11 or less nucleotides away from the start site. The mean distance is 5.4 nt and median of 5nt from the start codon of the first gene in the operon. All but one of these operons contains three genes. The PIN domains are the most divergent genes of these operons based on sequence homology, with potential alternate activities, such as interacting with different proteins. The *Repressor* is most conserved in each of these operons, indicating conserved activities. The first genes, where present, are also analyzed here. The organisms tested only have these cytoplasmic GH-1 family β-glycosidases in their genomes; others are membrane-bound versions of the same enzyme. Membrane bound versions of β-glycosidases in archaea have been shown to have alternate substrate (bond and carbohydrate-species) specificities than the soluble enzymes.

A

Sub-class	*VapB protein*	*PFAM*	*PAB*	*PH*	*TK*	*Pfu homologs*
1	PF0355.1	NO	PAB2376.1n	PH0500.1n	TK1762	PF1207
2	PF0573	COG2002	PAB3187	PHS027	TK0457	
3	PF0576	NO	NO	NO	NO	PF2058 PF1308
4	PF0774		PAB3103	NO	TK1752	PF1222
5	PF0775	NO	NO	NO	TK0373	PF1353
6	PF0780.1	NO	NO	NO	NO	PF0813
6	PF0813	NO	PAB1672.1n	NO	TK1291	PF0780.1
7	PF0838	NO	PAB7298	NO	NO	PF1224
1	PF1207	NO	PAB1741.1n	PH0707.2n	NO	PF0355.1
4	PF1222	NO	PAB3103	NO	TK1752	PF0774
7	PF1224	NO	PAB7298	PH0403.1n	NO	PF0383
3	PF1308	NO	NO	NO	NO	PF0576 PF2058

5	PF1353	NO	PAB1598.2n	PH0389.1n	TK0973	PF0775
3	PF2058	NO	NO	NO	TK0733	PF1308 PF0576

B

Repressors	*Repressor* Homology	Predicted recognition sequence	Distance from ATG	Function of 1st encoded protein	Homology of 1st gene	*Regulator* homology
PF0355.1	100%	aaggggaa	-33	β-glycosidase	100%	100%
PH0500.1	97%	gagggggag	-5	β-glycosidase	95%	93%
PAB2376.1	95%	aggtggga	-6	β-glycosidase	92%	90%
TK1762	93%	aggtgga	-3	β-glycosidase	71%	70%
PH0707.2	83%	aaggggat	-5	none	NA	66%
PF1207	80%	aagaggaa	-11	β-glycosidase	74%	68%
PAB1741.1	79%	aggggga	-5	β-glycosidase	66%	64%

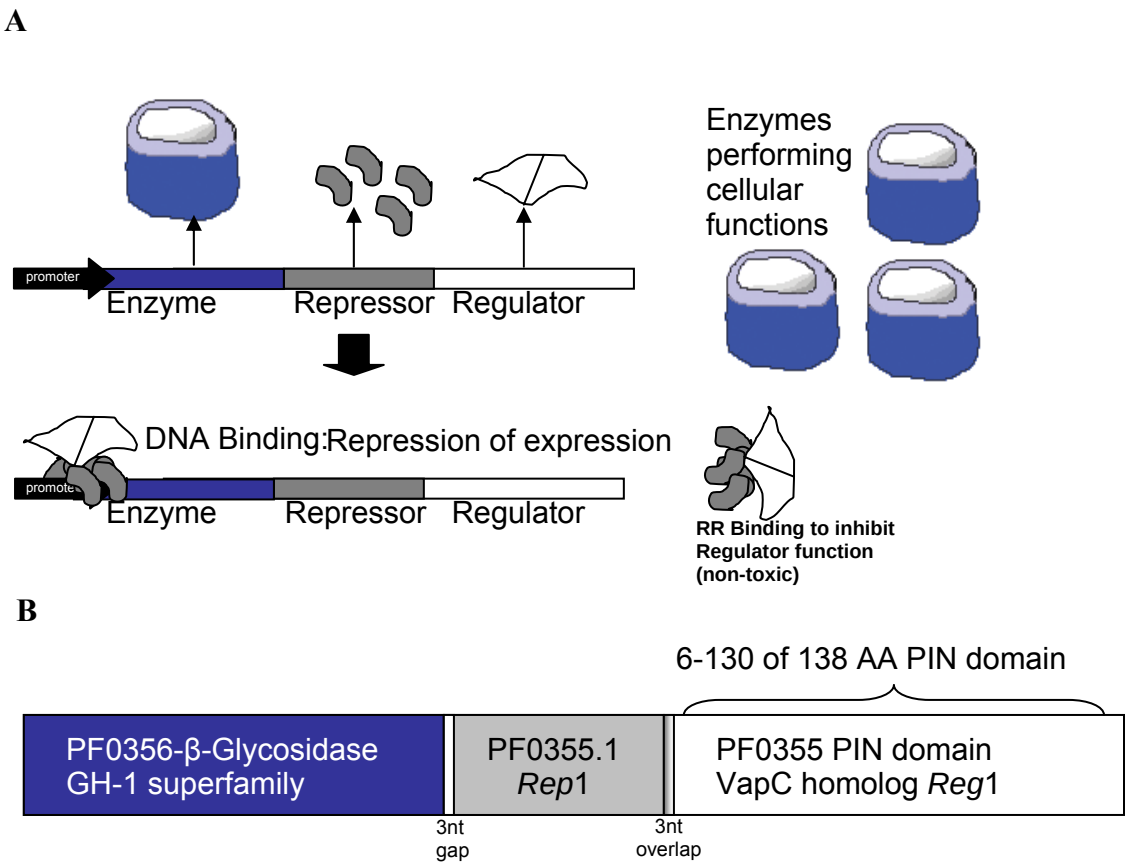


Figure 2. The proposed model of the revised nomenclature of the TA operons. The *Repressor Regulator* operons. A. The *Repressor Regulator* operon's model. The function of these operons is not killing, but the regulation of the expression of the enzyme encoded in the operon. The killing function of the toxin, or *Regulator,* is abolished through naturally occurring mutations. The protection of the antitoxin or *Repressor* is the *Regulator's* sole function. **B.** The structure of the operon under investigation. The three genes are annotated from left to right in order of expression. The gaps and overlaps between DNA sequences are noted. The length of the PIN domain is noted inside of the VapC protein, or *Reg*1.

3.0 Materials and Methods

3.1 Purification of PF0355p and PF0355.1p

PF0355 and PF0355.1 were cloned as published in Liu *et. al.* at the Southeast Collaboratory for Structural Genomics into vector pET24 dBam [42]. For characterization experiments BL21-DE3 (Stratagene, LaJolla, CA) cells were grown in ZYP 5052 self-inducing media [43] containing proper antibiotics. The same cells were grown in PASM 5052 self-inducing media, supplemented with Seleno-methionine for structure determination. The cells were incubated using a 5L fermenter at the University of Georgia Fermentation Facility at 37°C for 20 hr at 300 RPM agitation and with maintenance of 50% dissolved oxygen. The media was self-inducing - no inducer was added. The cells were lysed by sonication in 20 mM HEPES pH 7.6 100 mM NaCl (Buffer A) with 1 mM PMSF. The solution was supplemented with 1 μg/50ml Hen Egg White lysozyme (Sigma, St Louis, MO) and 1 μg/50 ml RQ1 DNaseI (Promega, Madison, WI). After lysis by sonication the cell-debris was sedimented at 10,000 x*g* for 30 min, and the supernatant was passed over NiNTA resin (Qiagen, Valencia, CA). After washing the column with 10 column volumes of Buffer A, the sample was then eluted with 400 mM Imidazole (Sigma, St Louis, MO). The yielding eluate's buffer was exchanged to Buffer A with 1 mM dithiothreitol (DTT) (Buffer B), and the protein was then heated to 100°C for 1 hour. The precipitate was cleared by spinning at 6,000 x*g* for 15 minutes. The purity was checked with SDS-PAGE. The supernatant was then used for further experiments. Maximal protein concentrations were 3.6 mg/ml for PF0355p and 1.7 mg/ml for PF0355.1p based on absorbance

at 280nm and the calculated extinction coefficient. Protein concentration over those listed caused irreversible precipitation of both samples.

3.2 DNA binding by PF0355.1p

The DNA probes used in the DNA binding experiments were prepared for as follows. The DNA fragments IG1 and IG2 were designed using sequence information from genome sequence of *Pyrococcus furiosus* (NCBI) between nucleotides 369910-370200 (NCBI). The DNA fragments used as bait in the DNA binding experiments were amplified from genomic DNA graciously provided by Dr. Gerrit Schut. The following primers: IG1F 5'-ATTTCAAACTGGAATGCCGA-3', IG1R 5'-AAGAACTACTCCAAGATTG-3', IG2R 5'-TGGTTATAACTAGGTAAGGA-3', IG3F 5'-TCCTTACCTAGTTATAACCA-3' were ordered from Integrated DNA Technologies (IDT), Coralville, Iowa. The PCR conditions used were those required by the PCR kit (Expand High Fidelity, Roche Indianapolis, IN), except the 50°C annealing temp for 20 cycles. When using IG1F and IG2R, the 160 bp IG1 fragment is amplified. The 130 bp IG2 was made with primers IG3F and IG1R. A small amount of each PCR reaction was run out on a 4-20% Criterion Tris-HCl acrylamide gel to ensure proper size of the fragment and single product. The PCR reaction was cleaned up with a Qiagen PCR cleanup kit, and the DNA samples were used in further experiments. PF0355p and PF0355.1p were in Buffer B (above) with 1 mM $MgCl_2$ for the DNA-binding experiments. All binding experiments were carried out at 78°C for 30 minutes. The amount of protein used in each experiment was 17μg PF0355.1p and 36 μg PF0355.1p, and the amount of DNA was standardized at 1 μg/μL for IG1 and IG2. One half of the DNA-protein mixture was then loaded onto a native 4-20% gradient Criterion Tris-HCl acrylamide gel purchased from BioRad (Hercules, CA). The gel was then stained with ethidium bromide to detect DNA (BioRad Hercules, CA).

3.3 DNase I footprinting of PF0355.1p binding the promoter

The fluorescent labeling of the DNA was done to perform the DNaseI footprinting. The DNA used were synthesized based on the publication from Wilson *et al.* [44]. The modifications to that protocol were only the DNA fragments and the proteins used. The DNA fragment called IG1 was amplified with a new set of fluorescently labeled primers made by IDT. The dyes were chosen based on directions from the user's manual for the ABI 3700 DNA Analyzer. When testing the antisense strand of PF0356 on IG1, the primers used were 6-FAM-labeled IG1F and non-labeled IG2R. These primers along with same protocol as used for amplification of DNA for the gel shift assays were used for the synthesis of these fluorescent DNA fragments. When analyzing the respective sense strand, the primers used were non-labeled IG1F and NED-labeled IG2R. The proteins assayed in each experiment were recombinant PF0355p and PF0355.1p with directions and reagents from the Core Footprinting Kit (Promega, Madison, WI). The binding reaction was carried out as before at 78°C for 30 minutes. The solutions were then chloroform/isoamyl alcohol-extracted and ethanol-precipitated to remove all protein and prepare DNA for later run in the sequencer instrument. The DNA pellet was then re-suspended in a solution made up of 20:1 HiDi Formamide and GeneScan-350ROX (ABI), the molecular weight standard labeled with ROX dye. These samples were then incubated at 95°C for 5 minutes. The samples were then loaded on an ABI 3700 DNA Analyzer. The molecular weight of the fractions was determined using the included GeneScan-350 ROX molecular weight standard labeled with a third fluorescent dye. Each sample was then passed through the instrument separately. The presence individual fluorescent dyes attached to DNA were then detected by the laser. The output was converted to Microsoft Excel format and analyzed. The background intensities were

identified based on intensities found in the data for fragments of size 165-170 nt where the size of the loaded sample was 160 basepairs.

3.4 Crystallization/data collection of PF0355p

PF0355p crystallization experiments were performed using the modified microbatch method as described previously [45, 46]. The protein was screened with seven commercially available crystallization kits: Hampton I, Hampton II, MembFrac (Hampton Research, Aliso Viejo, CA), Wizard I, Wizard II, Cryo I (Emerald Biosystems, Bainbridge Island, WA) and a custom screen designed by Shah [44]. For the setups the protein was concentrated to 3.6 mg/ml in Buffer B (above). Precipitants were screened using 1 μL drops containing equal amounts of protein and precipitant solution. Crystals were observed in wells containing 0.1 M citrate, 0.1 M NaCl, 0.1 M Lithium Sulfate, 30% V/V PEG 400, pH 5.5. Although these crystals were further optimized, optimization did not improve the diffraction quality over the initial crystals. For all experiments, the crystals were harvested using a 0.2 mm cryo-loop mounted on an 18 mm CrystalCap Copper magnetic pin (Hampton Research, Aliso Viejo, CA) and quickly immersed in LN_2. The crystals were mounted and flash cooled without cryo-protectant and shipped to beamline 22ID, Southeast Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source (APS), at Argonne National Lab (ANL). Using 22ID an X-ray fluorescence scan was made to confirm that seleno-methionine was incorporated into the protein. Phasing data were collected on the SER-CAT bending magnet beam line 22BM using 0.9763 Å (12660 eV) X-rays. A single SAS (Single wavelength Anomalous Scattering) dataset to 2.3Å resolution was collected on a MAR165 CCD detector. For data collection, the crystal was rotated 360 degrees in the beam in 0.5-degree steps. Refinement data were collected on a smaller crystal, taken from the same well, on 22ID using a MAR225 CCD detector. Both data sets were indexed, integrated

and scaled using HKL2000 [47] keeping the Bijvoet pairs separate during scaling. The data collection and processing statistics are shown in Tables 3A and B.

The data showed that the crystal was orthorhombic with a = 57.6 Å, b = 103.67 Å and c = 104.25 Å. The systematic absences (h00, h≠2n, 0k0, k≠2n, 00l, l≠2n) uniquely determined the space groups as $P2_12_12_1$ (#19). Assuming 4 molecules per asymmetric unit the Matthews Coefficient was calculated to be V_M=2.2 $Å^3$/Da corresponding to a solvent content of 42.5% [52].

Table 3. The Statistics of data collection and final data processing of PF0355

Dataset	*Phasing Dataset*	*Refinement Dataset*
Name of crystal	**Pfu-367848_2-dt_1**	**Pfu-367848-001_dt-1**
Beamline	**APS 22BM**	**APS 22ID**
Wavelength	**0.9763Å**	**0.9763Å**
Crystal size μm	**100 X 150 X 400**	**100 X 100 X 100**
Rotation of crystal (ω) / 2 θ	**360/0**	**235/0**
Image width (°)/ exposure time (s)	**0.5/7**	**0.5/3**
Detector used	**MAR165**	**MAR225**
Detector distance	**180**	**190**
Resolution		
Overall/outer (Å)	**50-2.3/2.38-2.30**	**40.0-2.3/2.38-2.30**
Unit cell (Å)		
a (Å)	**57.587**	**57.601**
b (Å)	**103.459**	**103.667**
c (Å)	**104.186**	**104.253**
Space group	**$P2_12_12_1$**	**$P2_12_12_1$**
No. of reflection observations/unique	**377235/27435**	**265366/28211**
Overall Completeness /outer (%)	**96.0/59.8**	**98.7/99.7**
Redundancy/outer	**13.8/5.9**	**9.4/9.6**
R_{sym}*	**8.7/23.6**	**10.9/34.9**
I/σ_I overall	**39.14/6.65**	**30.32/6.33**

Structure Identifier	*1Y82*
R/R_{free}**	**22.3/27.8**
Torsion angles	
period 1,	**5.796,**
period 2,	**33.407,**
period 3,	**15.829,**
period 4 (°)	**19.953**
R.M.S deviations	
bond length (Å)	**0.017**
angle (°)	**2.008**
dihedral (°)	**0.005**
Average B factor ($Å^2$)	**21.48**

*** Rsym(I) = Σ_{hkl} [Σ_i|I_i(hkl) – I‹hkl›|] / $\Sigma_{hkl}\Sigma_i$ I_i‹hkl› where $I_{hkl,i}$ is the measured intensity an individual Miller indices h, k, l, and I‹hkl› is the average intensity of that reflection**

**** R = Σ ||F_o| - | F_c|| / Σ |F_o|**

R_{free} = Σ_{test} ||F_o| - | F_c|| / Σ_{test} |F_o| test represents the 10% of the dataset set-aside before refinement

Table 3B. The final data processing statistics of **Pfu-367848-001_dt-1**.

Shell limit	Lower Angstrom	Upper Angstrom	Average I	Average error	Average stat.	Norm. Chi**2	Linear R-fac	Square R-fac
	40.00	4.95	4339.1	68.7	44.5	5.056	0.072	0.075
	4.95	3.93	4564.1	80.0	59.7	3.550	0.076	0.080
	3.93	3.44	3559.2	81.4	67.6	3.471	0.096	0.121
	3.44	3.12	2084.7	57.7	50.8	2.370	0.102	0.104
	3.12	2.90	1325.3	48.7	45.3	1.653	0.110	0.109
	2.90	2.73	908.5	47.5	45.5	1.297	0.139	0.128
	2.73	2.59	834.5	50.4	48.6	1.225	0.155	0.154
	2.59	2.48	606.4	52.7	51.6	1.042	0.199	0.178
	2.48	2.38	631.4	72.2	71.2	1.001	0.265	0.248
	2.38	2.30	542.6	85.7	85.0	0.883	0.349	0.320
All reflections			1953.0	64.4	56.7	2.127	0.109	0.096

Table 3C. The Root Mean Square Deviation of the main chain atoms of each polypeptide visible in the crystal structure. The residues evaluated were those amino acids 10-141, the stable core of all chains.

	Chain A	***Chain B***	***Chain C***	***Chain D***
Chain A	**0**	**.737**	**.284**	**.609**
Chain B		**0**	**.682**	**.296**
Chain C			**0**	**.545**
Chain D				**0**

3.5 Phasing and Refinement

The 22BM SAS data, after scaling, was input to the SCA2structure pipeline [47] using the parameters illustrated in see Figure 3. The pipeline uses an array of separate programs (SOLVE [49], RESOLVE [50-51], Arp/Warp [52]) to fine screen parameter space in order to achieve the optimal parameter combination for phasing of the data. The program SOLVE identified the anomalous substructure and was used to produce SAS phases. RESOLVE carries out the phase improvements and produces an initial sequence fit. Based on the RESOLVE phases, ARP/WARP traces the initial model. In the pipeline SOLVE/RESOLVE were used to phase the structure to 3.0 Å. Although the Bijvoet Patterson map was unimpressive, SOLVE identified 20 of the 32 selenium sites. Based on these sites, ARP/WARP produced an initial structure (93.9% complete) with an R value of 28%.

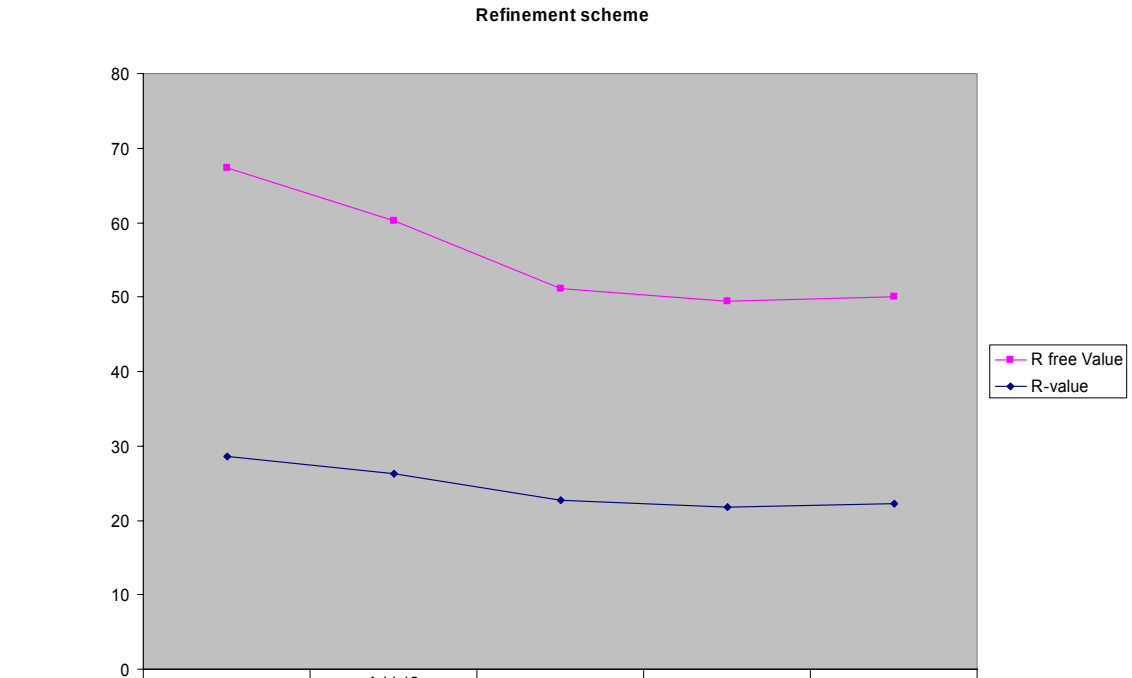
Using the more complete refinement data set collected on 22ID several rounds of positional and B-factor refinement were carried out using REFMAC 5 (See Figure 4) [48]. A random 3.11% set of the reflection data were excluded from the refinement and used to estimate the free R (R_{free}) as a monitor of over-fitting. XFIT was used for model building and manual fittings during the refinement [49] as needed. The final R- and R_{free}–values for the structure after refinement were 22.2% and 27.8%, respectively. The structure was validated using MolProbity [50] and those amino acids that were found to be in energetically non-favorable orientation were adjusted as required (See Figure 4). The atomic coordinates have been deposited in the Protein Data Bank (www.rcsb.org/pdb/) under identifier 1Y82. It should be noted that six days after the public release of the PF0355 structure, the corresponding structure from *Pyrococcus horikoshii* (92% sequence homology and a backbone Root Mean Square Deviation (RMSD) of 0.8Å to the PF0355 structure) was released (PDB entry 1V96).

Figure 3

Figure 3. SECSG Sca2Structure pipeline input screen. The pipeline takes a Scalepack file as the input and screens parameter space to find the best solution to the data.

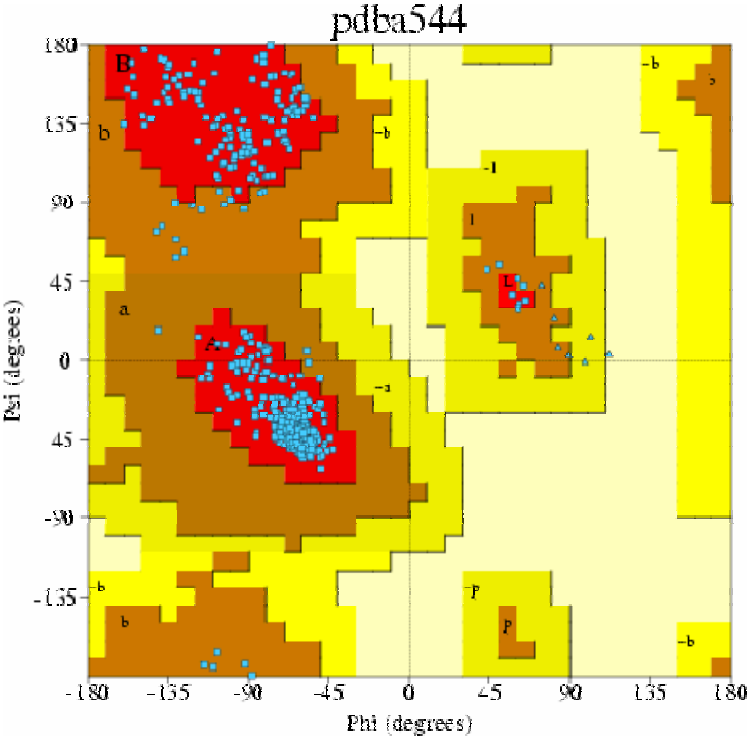
Figure 4. The refinement history of 1Y82. A. Three rounds of Maximum-likelihood refinement with REFMAC 5 were done in each case. The R values are depicted in the bottom of the figure with the modifications performed in between each step also listed in the table. **B.** The Ramachandran plot showing the analysis of phi and psi angles of the PF0355p structure. The figure shows all amino acids are in allowed regions, with glycines as only exceptions.

A.



	ARP/WARP	Add 12 residues/chain, except D	Add 81 waters	MOLprobity and corrections	1Y82
R free Value	38.7	33.98	28.4	27.64	27.8
R-value	28.6	26.3	22.7	21.75	22.3

B.


pdba544
Psi (degrees)
Phi (degrees)
180
135
90
45
0
45
-90
-135
-180
-90
-45
0
45
90
135
180
B
A
L
P

3.6 Structural comparison

Structural comparisons of 1Y82 to other PDB structures was performed with DALI (http://www.ebi.ac.uk/dali/) [51]. The top structural matches (Z-scores of greater than three) were used for the structural overlays. Translation/rotation of the molecules, if indeed, was done, using PDBset from the CCP4 program suite [48]. The atomic models were visualized using Pymol [52] for evaluation.

3.7 Structure prediction for alignment

Since the sequence homology amongst the toxin/regulator proteins was less than 40% active site identification based on sequence proved problematic and a structure based approach was used.

FUGUE v2.0 (http://www-cryst.bioc.cam.ac.uk/~fugue/prfsearch.html) [53] was used for the structure predictions. Predictions were evaluated based on the Z-score, and only those models with Z-scores of three or more were used. For modeling the two plasmid encoded toxic VapC homologs *Salmonella dublin* VapC and *Shigelia flexneri* mvpT, the *Pyrococcus horikoshii ot3* VapC homolog structure 1V96 was used.

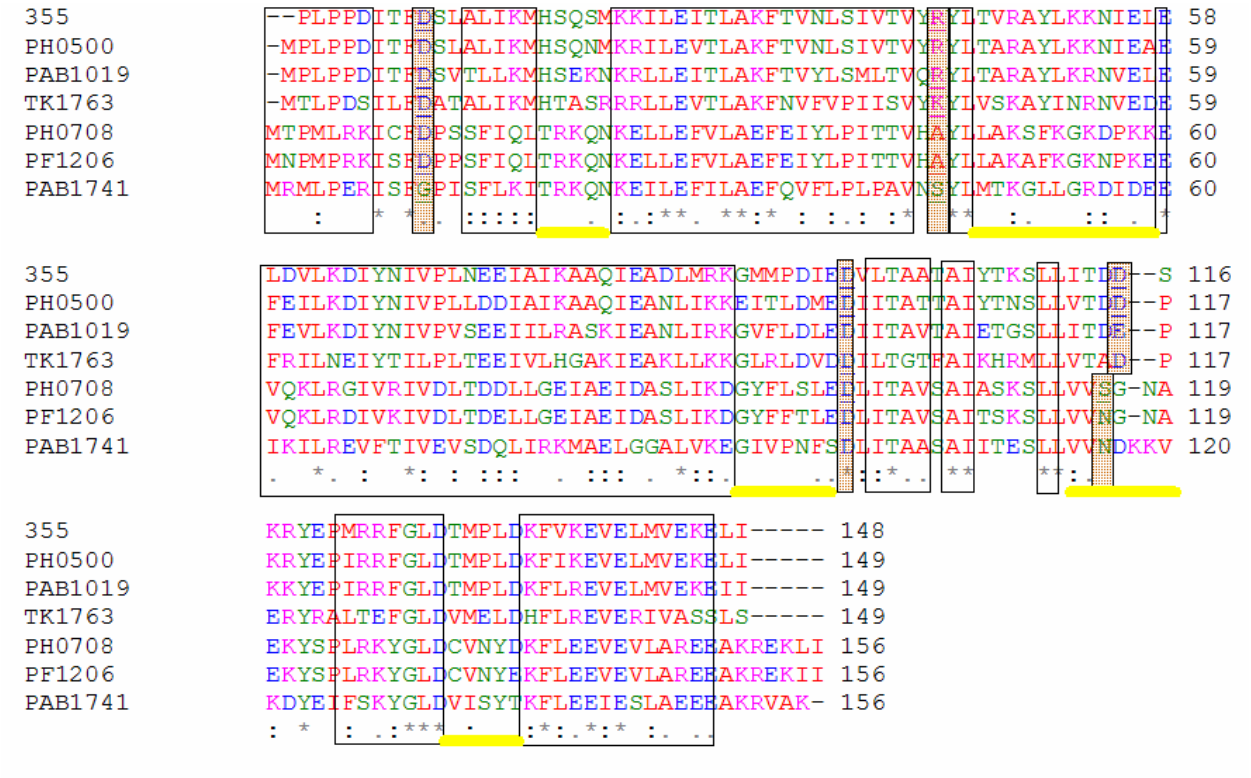
Threading of the VapB homolog PF0355.1 did not produce an acceptable structure based on the pre-set Z-score limit.

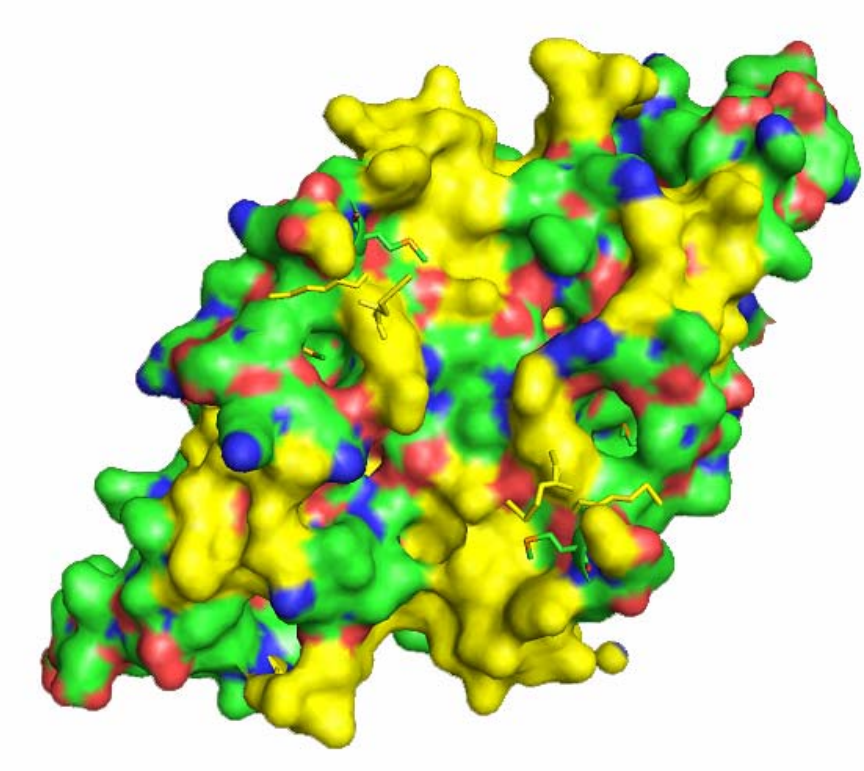
When evaluating the third component of the operon PF0356 (in this case a β-glycosidase) a similar approach was employed. The program FUGUE V2.0 used structures of multiple Glycohydrolase Superfamily 1 members (1CBG, 1E73, 1HXJ, 1UG6, 1GNX, 1E4I, 1QOX,

1QVB) as a template for the structure prediction of PF0356p. The β-galactosidase structural prediction allowed the visualization of the active site in this protein as well (Data not shown).

The structures from the output were visualized with Pymol. All protein sequence alignments were done using clustalW [55] and the sequences for each toxic VapC genes was retrieved from EXPASY (http://us.expasy.org/). All PF0355, PF0355.1, and β-galactosidase gene sequences were retrieved from NCBI.

Figure 5. The region of PF0355p that is conserved with its closest homologs. Based on the multiple sequence alignment, all homologous regions are highlighted with brackets. The regions outside of the brackets represent 51 amino acids. 37 of the total 51 divergent amino acids are located in the large grove in the center of PF0355p dimer. These amino acids are highlighted in yellow in the structure and in the alignment. The residues highlighted in red are those that structurally align with the 4 acidic amino acids indicated as the active site in 1V8O structure with exonuclease function. In the position of the last amino acid, Asp 116 in PF0355p, the register of the helix shifts. Therefore, the sequence alignment does not correspond to the structural alignment.





3.8 Isothermal Titration Calorimetry

The PF0355p was initially characterized in a CSC 4000 Isothermal Titration Calorimeter (CSC, Lindon UT). In these experiments, PF0355p was in Buffer B (above) in 0.095 mM concentration with 4 mM substrate with 10 μl injections with an initial delay of 400 seconds and automated injections every 400 sec. In an additional experiment an ITC from Microcalorimetry (Microcalorimetry, Northampton, MA) was employed. In this instrument, when identifying potential PF0355p and PF0355.1p substrates, the protein concentrations were 0.095 mM in Buffer B (above). 3 μl injections of 4mM substrates were used. When testing the action of PF0355.1p on the 0.095 mM PF0355 binding of adenosine-diphosphate (ADP), the concentration of PF0355.1p was 0.025 mM and 0.095 mM. ADP (Sigma) was 4 mM in Buffer B. All of the experiments for calorimetry were carried out at 65°C. The experiments were performed by an automated algorithm with a 400 second initial delay and 200 second relaxation time between each injection for a minimum of 30 injections. The data in each case were evaluated with the software supplied with the instruments. The software supplied with the Microcalorimetry instrument was Origin and the software from the CSC machine was called Bindworks. The data was fitted by each type of fitting available in the software, and the best fit was obtained, based on Chi-squared values.

3.9 Interaction of PF0355p and PF0355.1p

Based on the model from the TA operons, the two proteins are expected to form a protective complex that inhibits protease degradation of the antitoxin. This assay shows the interaction of the toxin and antitoxin in this protective role. In the protease protection assay, PF0355p and PF0355.1p proteins were in the same concentration of 0.095 mM. The proteins by

themselves were 0.2 mM and 0.19 mM, respectively. The whole-cell extracts were from *P. furiosus* grown on a variety of media (maltose and/or peptides and/or sulfur) lysed in a solution containing 50 mM Tris-HCl pH 7.8, 2 mM DTT, and 2 mM sodium dithionite (buffer C). In the assays the whole cell extract to pure protein concentration ratio was 1:12. The reactions were incubated at 70°C and 90°C for 180 minutes. The samples were then analyzed on SDS-PAGE. In the reactions where cell-extract was not used, buffer C replaced the whole cell extract.

When evaluating these proteins' interactions on a native polyacrylamide gel electrophoresis, the solution of PF0355p and PF0355.1p was incubated at 70°C and 90°C for 60 min in 25 mM HEPES (pH7.5), 1 mM DTT and 1 mM $MgCl_2$. ADP concentrations were 0.5 mM and 1 mM. Electrophoresis running buffer was TBE (pH 8.0). Both gel types were stained with Comassie brilliant blue (Sigma).

4.0 Results

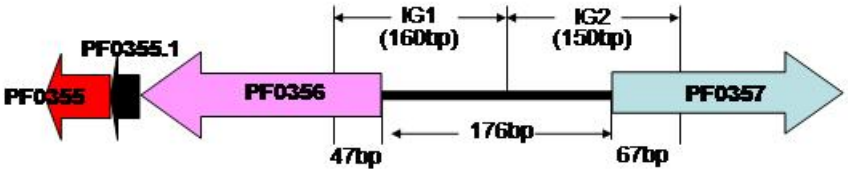
4.1 DNA binding by PF0355.1p

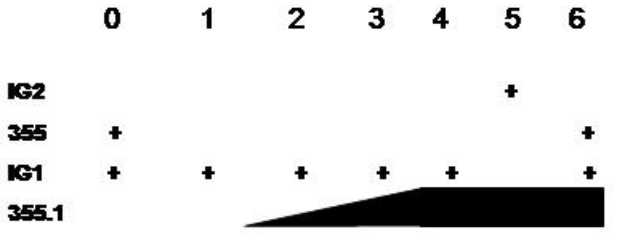
PF0355.1p is a newly annotated gene that is currently not present in all databases, including TIGR [56]. Since there was no previous information on PF0355.1p's activity, Pandey *et al.*'s prediction of PF0355.1p's potential VapB-like function was tested [24]. In this model the antitoxin binds DNA and auto-regulates the expression of proteins. PF0355.1 was used to bind the promoter of the operon. When the promoter of PF0356 was used, the protein bound it specifically. It was earlier determined by Poole *et al.* that the transcriptional unit contains PF0356 through PF0355 [56]. Based on these data, the intergenic region between PF0357 and PF0356 was the bait in these experiments. In order to investigate the sequence specificity of PF0355.1p DNA binding, the intergenic sequence was divided in two at the potential regulatory sequences for the PF0357-PF0363 operon and PF0356-PF0355 operon as shown in Figure 6A [56]. The DNA next to the ORF PF0356, the intergenic, potential regulatory sequence of the operon was named IG1. This DNA fragment contained 160 basepairs (with 47 bp of intragenic sequence from PF0356). The remainder of the intergenic region, flanking PF0357 and 67 bp of PF0357 make up IG2. Figure 6 shows that when the PF0356-PF0355 operon's promoter region is presented to PF0355.1p, the protein binds it specifically. It does not show affinity to the promoter region of the PF0357-PF0363 operon. The presence of PF0355p did not inhibit the interaction between PF0355.1p and IG1. The DNA fragments included some of the intragenic sequences, so the binding of the protein to the intergenic region would not be inhibited by the close proximity of the end of the DNA fragment. PF355.1p binds the putative promoter region of

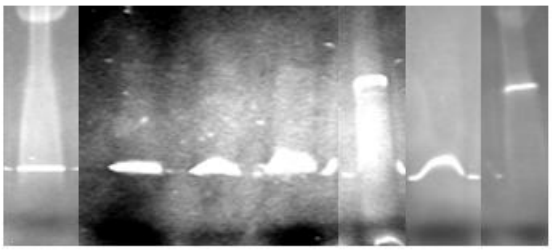
PF356, which is also the regulatory region of the operon. Using DNase I footprinting, we have narrowed down the sequence of PF355.1p binding as discussed below.

Figure 6. The investigation of DNA binding by PF0355.1p. A. The depiction of the promoter architecture tested in gel shift assay. The DNA sequence was amplified by PCR from genomic DNA as described in Methods. The 160 bp IG1 fragment is the potential promoter of the PF0356-PF0355 operon. The 150 bp IG2 fragment is the putative promoter of the opposing operon containing the ABC transporter encoded by PF0357-PF0363. The electrophoretic mobility shift assay was performed as described in Methods. Lane 0 was the IG1 fragment incubated with PF0355. Lane 1 was the control, IG1 incubated by itself with buffer replacing PF0355.1. Lanes 2,3,4 were IG1 incubated with 1-9 µg of PF0355.1 with lane 4 having 17 µg PF0355.1. Lane 5 was the incubation of IG2 with 17 µg PF0355.1. The DNA was standardized at 2 µg. **B**. The sequence encoding IG1. The pink region is PF0356 with the translation start site indicated with +1. The binding region of PF0355.1 is highlighted in red. The regions underlined are a series of inverted repeats spaced 14 nt apart. Although this seemed a likely place for binding, it was not the case. **C**. The outputs from the ABL sequencers. The DNAseI fooprinting results show the protected region of DNA by PF0355.1. The protection pattern was observed when PF0355 was also included in the experiment. PF0355 did not show any direct interaction with DNA. The length of the protected sequence on the sense strand was larger than on the antisense strand, probably due to the conformation PF0355.1 takes on DNA.

A.





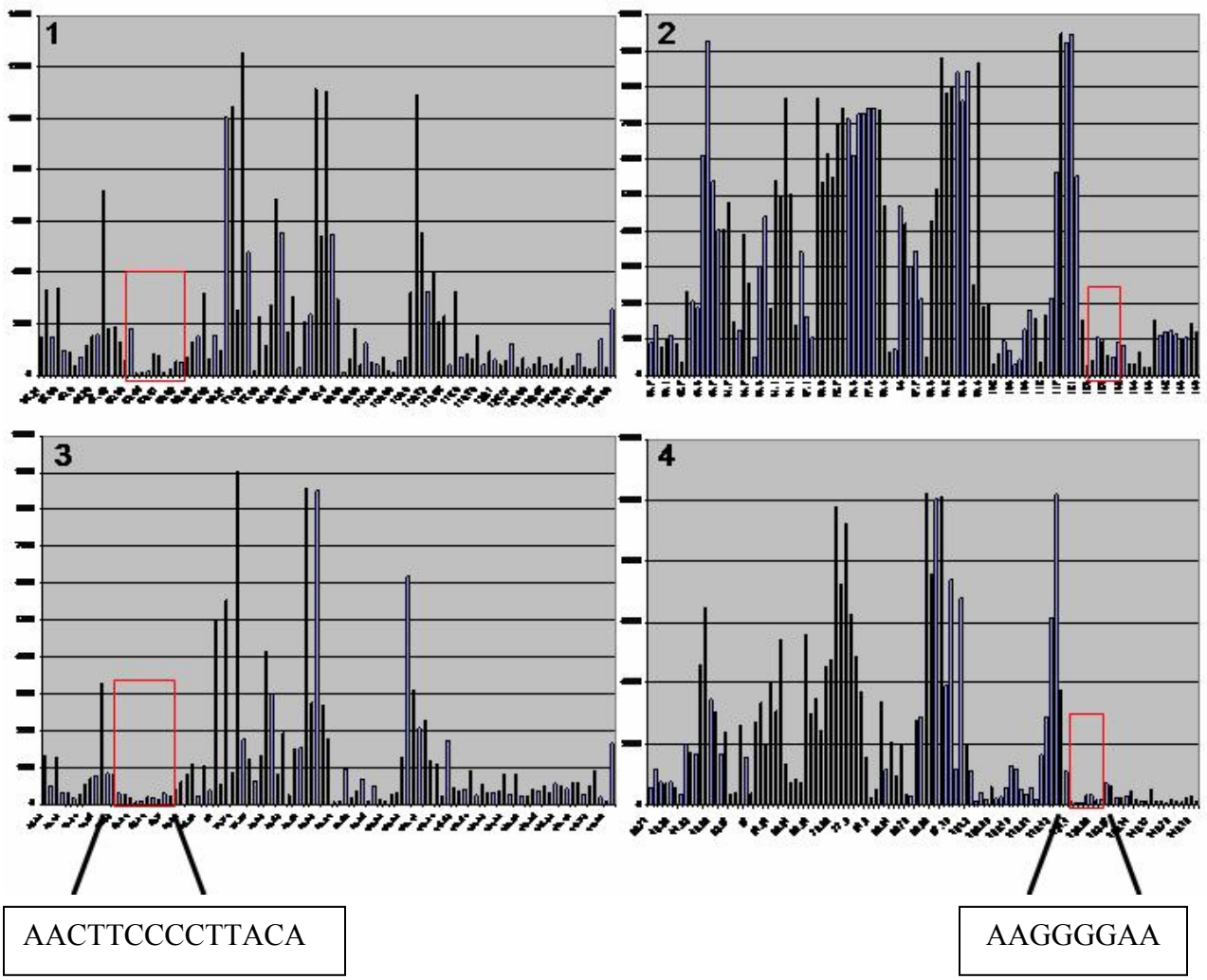


B.

atttcaaact ggaatgccga ttgaacgact
ccccagtaaa attccataaa cttcccctta
cagaattata tgtttcagtg ttttttataat
tttgcatctt ccaccatatt aaatctagct
ataggacttg gcacgttcaa tccttaccta
gttataacca

+1

C.



4.2 DNAseI footprinting of PF0355.1p

Determining the exact DNA sequence to which PF0355.1p binds was determined using a recently published a protocol for non-radioactive DNAseI footprinting, which served as the template for our assay [44]. This experiment used a fluorescently labeled DNA fragment where only a single strand is labeled instead of the classical radioactive 5' phosphate labeling. The 6-FAM labeled IG1 fragment was used to test the binding of the antisense strand of PF0356 by PF0355.1p (Figure 6, C.1 and C.3). The NED labeled IG1 was used to test sense strand binding by PF0355.1p (Figure 6, C.2 and C.4). Figure 6C shows the resulting chromatograms produced by the ABI 3700 sequencer (Applied Biosystems, Foster City, CA). These instruments have an internal standard that allows precise, single nucleotide resolution of the bases and do not require parallel sequencing [44]. Figure 6C shows that PF0355.1 binds the sense strand between bases 102-109 nt (AAGGGGAA). The octa-nucleotide sequence was protected only in the presence of PF0355.1p. The presence of PF0355p did not alter the binding pattern (data not shown). When the anti-sense strand of PF0356 operon was evaluated, the DNA showed protection on nucleotides 49-61 (AACTTCCCCTTACA). The presence of PF0355p did not alter the binding pattern by PF0355.1p in this case either (data not shown). These experiments showed a unique DNA sequence bound by PF0355.1p, whose location is predicted to be outside of the PF0356 translation start site (bases -5 through -12 from translation start) [56]. On the anti-sense strand for this operon, the protein protected the sequence -2 through -15 from the PF0356 translation start. According to gel filtration studies on an analytical Superdex 75 column (GE Healthcare-Amersham, Piscataway, NJ), the protein binds the DNA as a tetramer (data not shown).

4.3 PF0355 structure and comparison to structural homologs

The structure of PF0355p has been solved here by the Single Wavelength Anomalous Scattering method using seleno-methionine labeled crystals (eight methionines per molecule). Selenium incorporation of PF0355's crystal was confirmed by X-ray fluorescence before data collection (data not shown). The initial phasing data was collected using 0.9763 Å X-rays (12660 eV) on 22BM (SER-CAT) using a MAR165 CCD detector. Size limitations of the detector limited the resolution to 2.3 Å. (See Table 3). Based on unit cell and space group parameters the Matthews coefficient calculation by CCP4 suggested 4 molecules in the asymmetric unit [48]. This meant that 32 Selenium atoms should be observed in the Bijvoet Difference Patterson Map. The phasing dataset was evaluated with SOLVE and 20 of the 32 Se sites were identified [57]. The 20 Se sites found correspond to the 20 non-surface accessible methionine residues. The missing Se sites correspond to methionines 88, 92 and 93 that lie on a solvent exposed loop and are probably disordered since no peaks corresponding to these sites were observed in the Bijvoet Patterson map or were identified in the SOLVE output (Figure 7). The 22BM data from the first crystal showed very good statistics (Table 3) and was optimal for phasing. Using the SCA2Structure pipeline phases to 3Å were generated. Phase extension using the more complete 22ID data allowed for an initial trace of 556 of the 592 residues present in the crystallographic asymmetric unit (93.9% complete). After 15 rounds (see Figure 4) of Maximum-likelihood refinement (REFMAC) the structure was validated using MolProbity (http://kinemage.biochem.duke.edu/molprobity/main.php?use_king) that evaluates all-atom contacts and the geometry of the model. In addition rotamer analysis, and C-beta deviations were analyzed and the outliers corrected. The structure was then subjected to a last round of Maximum-likelihood refinement (REFMAC) and converged to give an R value of 22.8 (R_{free}

27.8). This structure was then subjected to a clash-score test by MolProbity, which gave a value of 6.69 [50], indicating that the structure is better than 98% of comparable structures in the PDB (resolution between 2.05 Å and 2.55 Å, 123 samples) [50]. The final refined model (PDB entry 1Y82) consists of four molecules: Residue 2-148 in Chain A, Residue 3-148 in Chain B, Residue 3-148 in Chain C and Residue 3-141 in Chain D. The histidine purification tag and missing N- and C- terminal residues are assumed to be disordered, since they were not observed in any electron density map. The refined model also contains 39 solvent molecules modeled as water. There are 42 additional solvent molecules (14 per molecule) located in clefts on the protein surface, which we believe mimic ligand binding. These molecules were also refined as water.

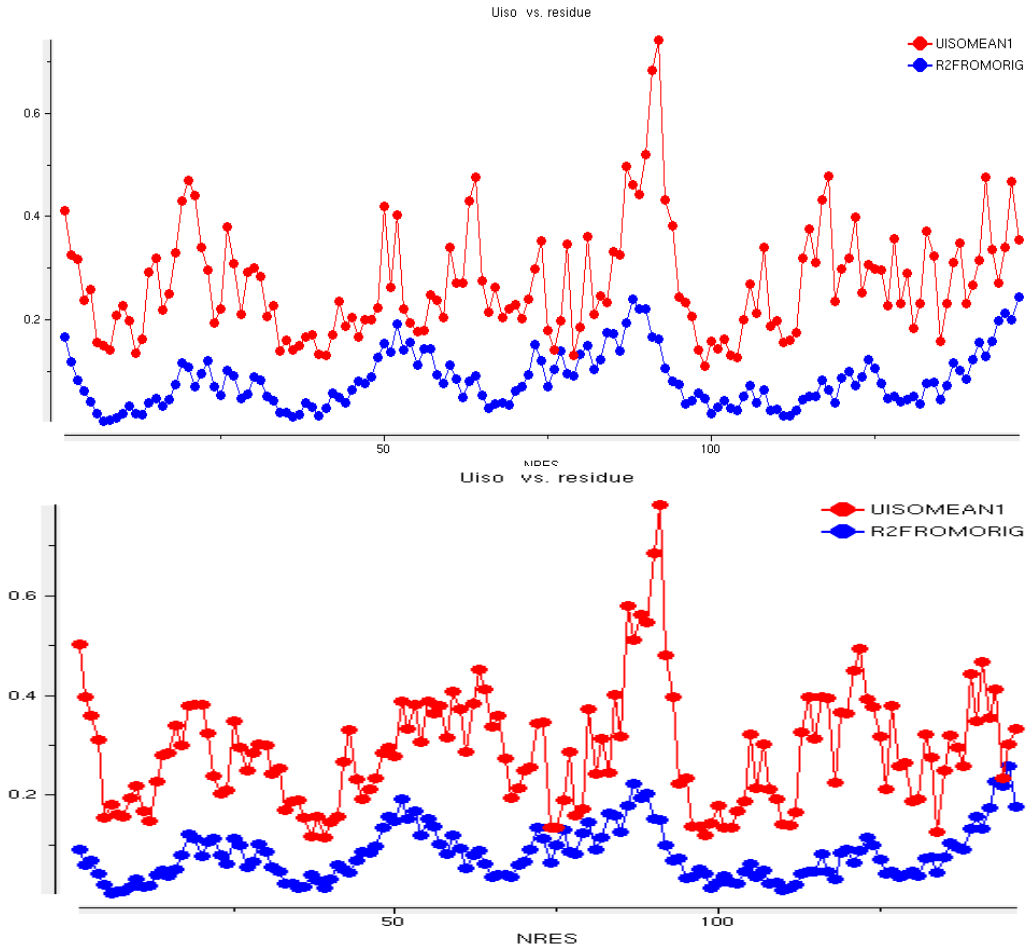
The protein has an α/β structure with five parallel β-sheets in the center of the structure and nine α-helices surrounding the central β-core (Figure 8B). The asymmetric unit contains two homo-dimers, consistent with results from analytical gel filtration (data not shown). During structure evaluation and refinement, a prominent solvent channel was identified in the center of each monomer, which contained seven to nine ordered solvent molecules (Figure 9). These solvent molecules form an ordered water network with an average inter peak distance of 2.5 Å and may represent a possible substrate-binding site. Although these sites were refined as water they are denoted by UNX atoms in the 1Y82 PDB entry.

The four molecules that make up the crystallographic asymmetric unit are very similar and can be superimposed [52] with RMSD's ranging from 0.28 to0.74 Å see Table 3C. The AB dimer has 0.33Å RMSD to the CD dimer.

A search for similar structures in the PDB using DALI [51], yielded only one structural homolog with a characterized enzymatic function, the VapC toxin from *Pyrobaculum aerophylum* Pae2754 (PDB code 1V8O, 11% sequence homology). Pae2754 is the first structure

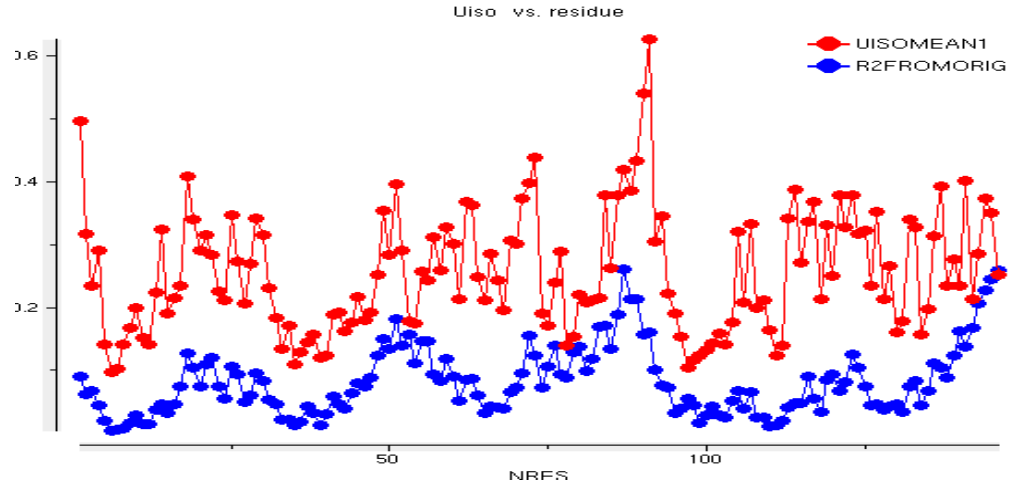
of a VapC toxin reported [24] and can be superimposed onto PF0355 structure with an RMSD of 3.1Å (backbone), see Figure 8. Overall the structures are very similar with core residues of the proteins having almost identical conformations. The Pae2754 protein has been shown to be an exonuclease [41] with an active site consisting of 4 acidic amino acids (Asp 8, Glu 38, Asp 92 and Asp 110) that bind the Mg^{2+} necessary for the cleavage [41]. These residues correspond to Asp 10, Arg 44, Asp 92, and Asp 116 in the PF0355 structure [41]. Although three of the four amino acids are identical between the two structures, Arg 44 in the putative PF0355 active site is a basic amino acid and would disrupt Mg^{2+} binding. This substitution may explain why exonuclease activity was not observed for PF0355 as discussed below.

Figure 7. Temperature factors plotted versus residue number. The program ANISOANL from CCP4 was used to determine the temperature factor and the deviation from origin per amino acid. The spike of the temperature factors around residues 89-93 shows the extreme thermal motion of those residues causing them to not be visible in the SOLVE output.

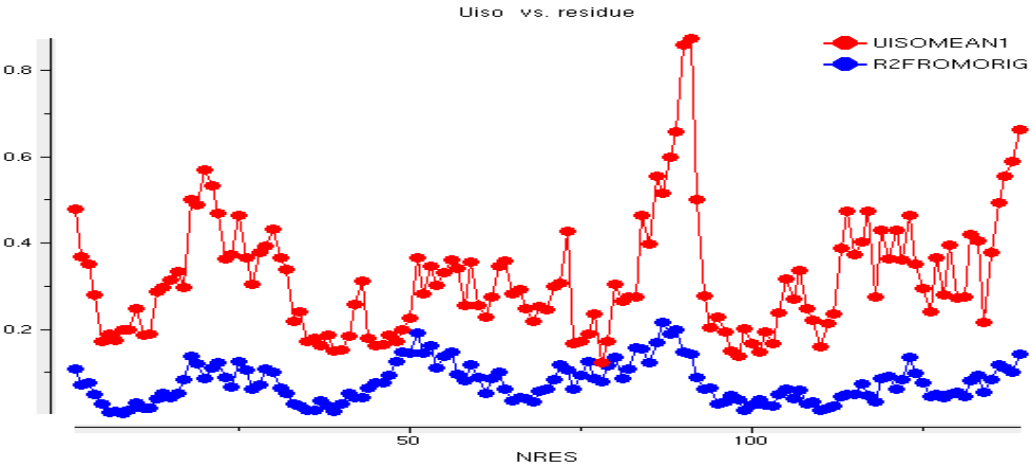


Chain A

Chain B



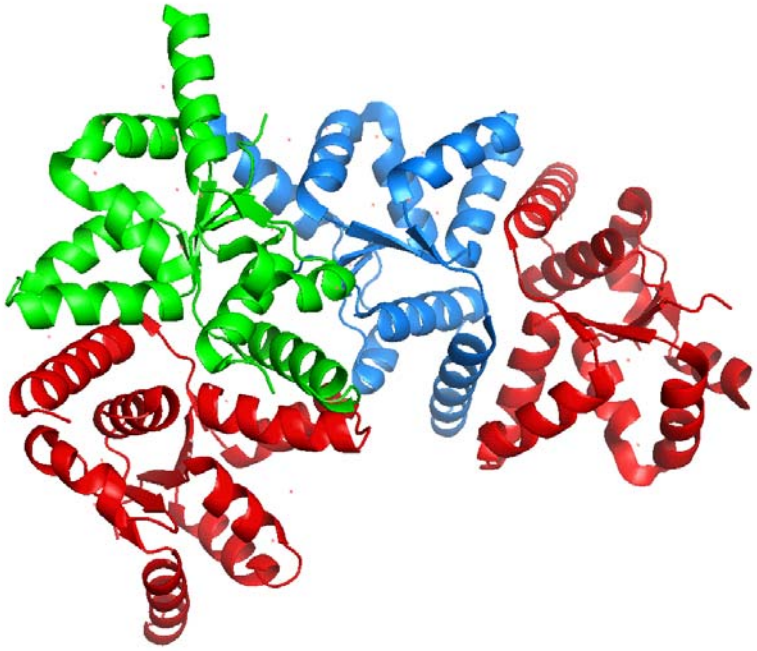
Chain C



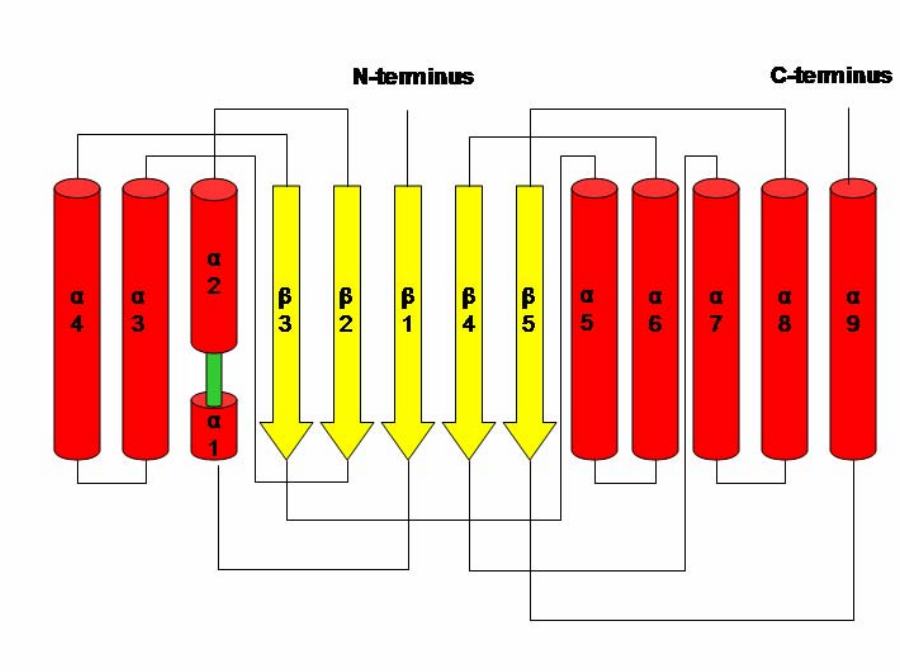
Chain D

Figure 8. The structural evaluation of PF0355. A. The structure of PF0355 represented as a cartoon. The entire asymmetric unit is represented in this figure. The four subunits are colored based on conformations. The major difference amongst the structures is the difference between the red subunits and the other two. The red subunits are chains B and D. These two molecules are identical in conformations. What differentiates those chains B and D from A and C is that the solvent channels in the center of B/D are better formed with more ordered solvent. The green and blue subunits are different only in the orientation of their N-terminus. **B.** The ribbon diagram of one subunit of PF0355p. There are five parallel beta strands in yellow surrounded by nine alpha helices in red. There is one loop region that connects helices one and two indicated by a green connector. **C.** The structural alignment of 1V8O and 1Y82. The overall alignment of the structures based on DALI output (1V8O in cyan, PF0355 or 1Y82 in green). The deviation of the main chains is less than 3.1 Å [51]. The structures' cores strongly resemble each other with the largest deviation in structures at the C-termini. **D.** The structural representation on the putative active sites. **I.** The functional amino acids of 1V8O, the active exonuclease are represented. These are Asp 8, Glu 38, Asp 92, Asp 110. **II.** The structural alignment of the functional enzyme 1V8O and 1O4W, another PIN domain from the PDB. Both of these proteins contain the putative active site's 4 acidic amino acids. **III.** The structural alignment of our structure with 1V8O, the active exonuclease. PF0355's putative active site is depicted with the arrow pointing to the arginine in PF0355 that replaced E38 from 1V8O. **IV.** The structural comparison of the two 92% homologous structures that are non-functional PIN domains. Three of the four amino acids are identical to those found in the active enzyme, but R44 replaces E38. This charge difference is the reason why our protein showed no exonuclease activity. **E.** The structural overlay of 1V96 and the FUGUE V 2.0 predicted structure of the *Salmonella Dublin* VapC*, Shigelia flexneri* mvpT genes [53]. The yellow coloring is the model of those predicted structures. The red structure is 1V96. The space-filled segment of atoms in the center shows those active site residues discussed in C. The black arrow points to the region where the toxic proteins end, but all archaeal proteins continue into a large secondary structure, stabilizing and blocking the active site from the substrates of those toxic PIN proteins.

Figure 8
A.

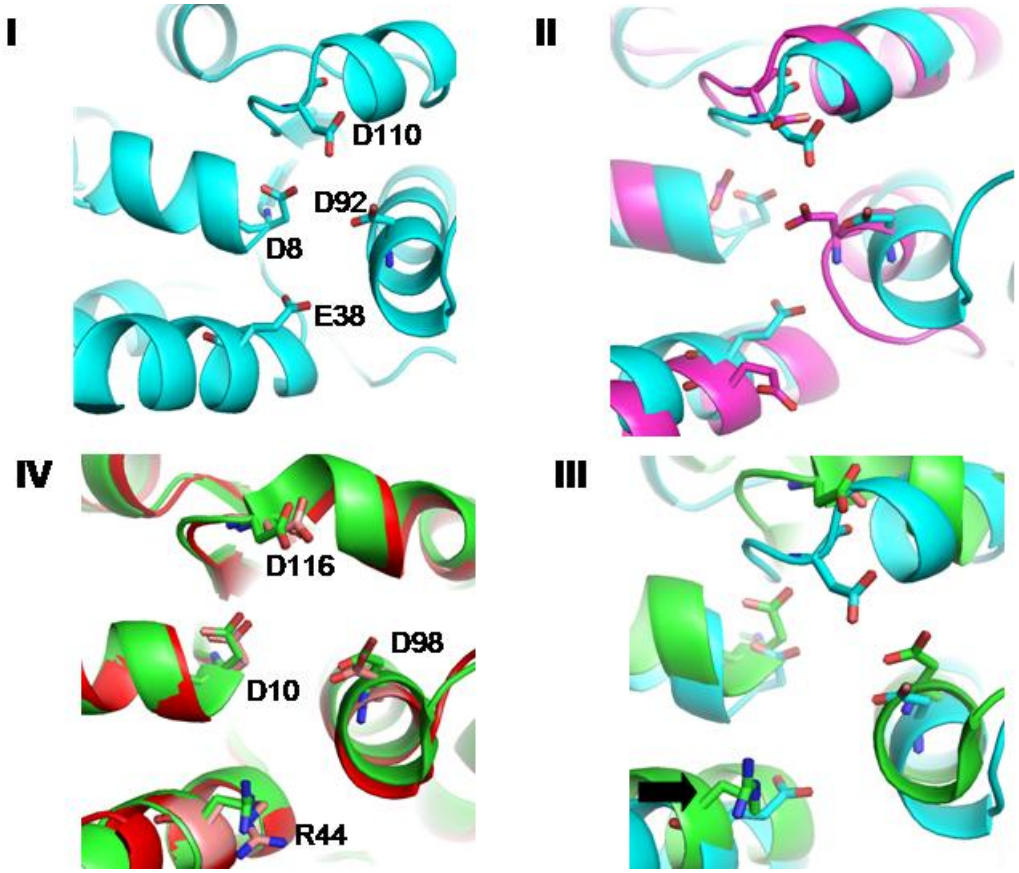


B.



C.

D.



E.

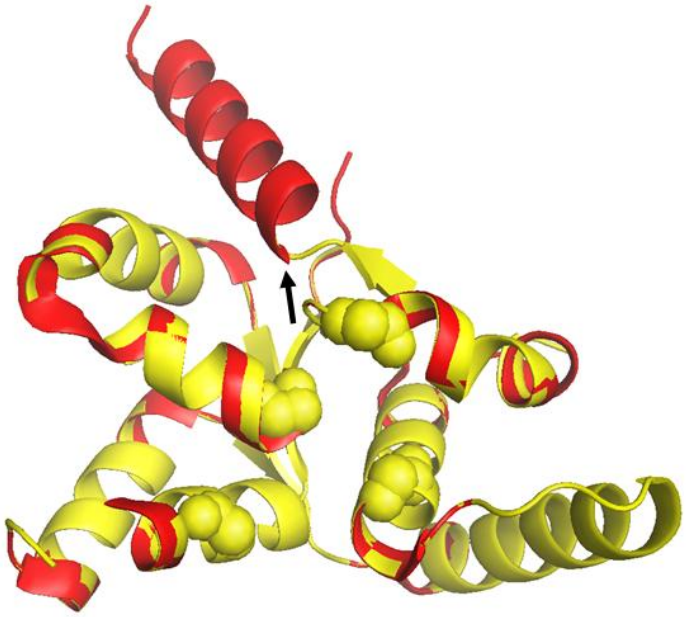
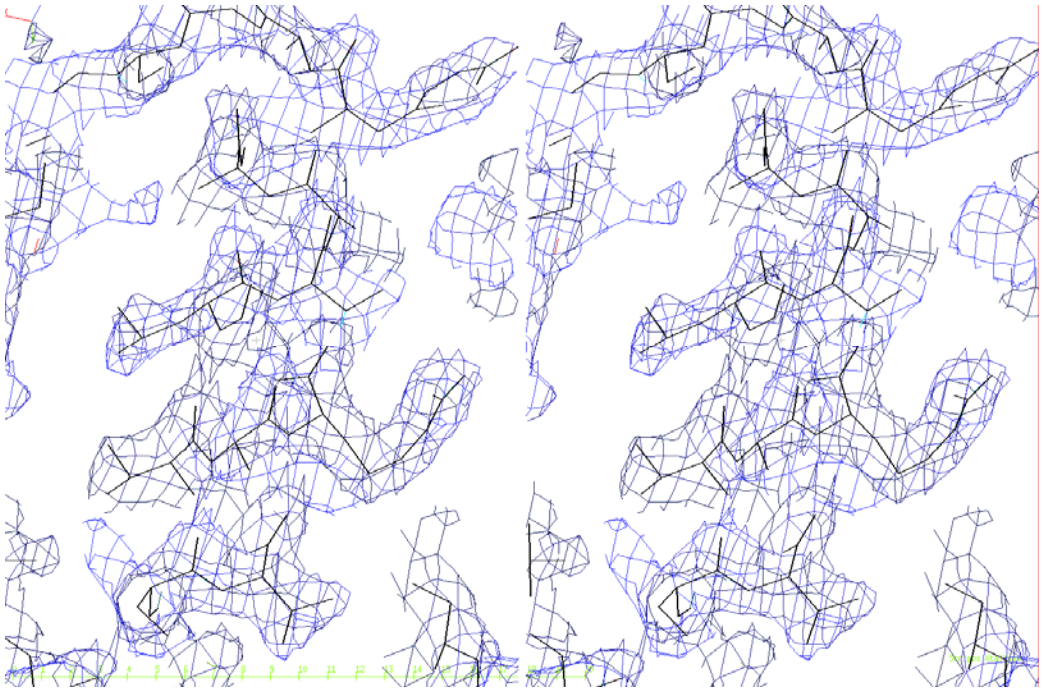


Figure 9
A



B

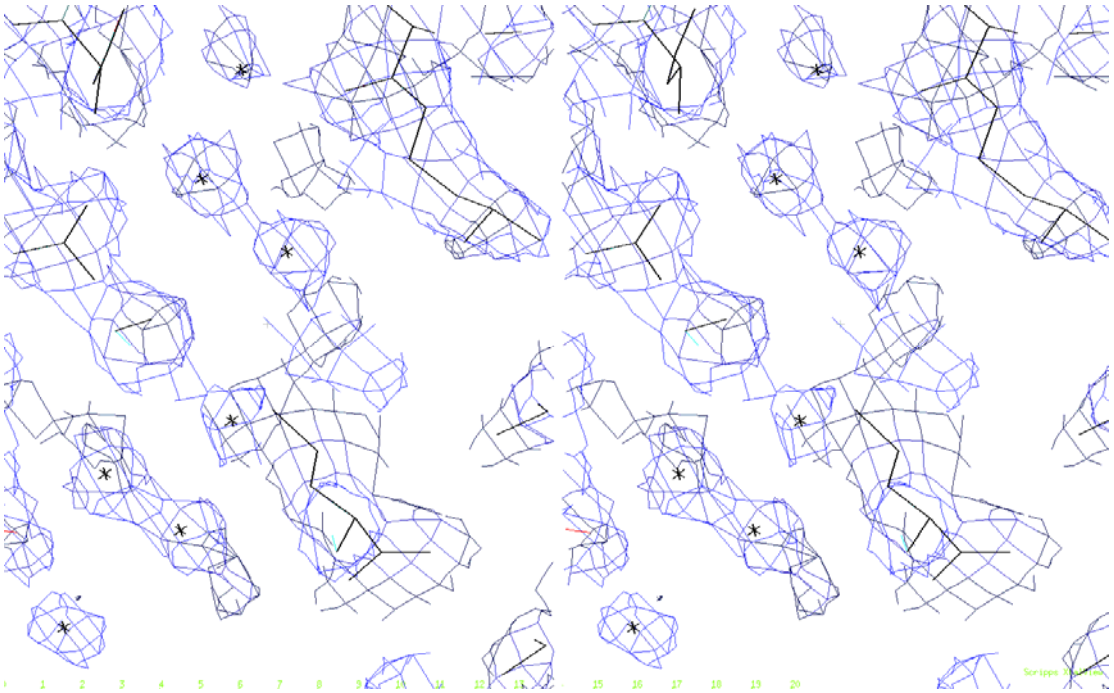


Figure 9. Sample electron density from the final map used during refinement. A. A section of the electron density residues Val 40- Tyr 51 Showing an α helix with the refined structure superimposed in cross-eyed stereo **B.** A section of the electron density map showing ordered solvent in the potential substrate binding sites of the PF0355 structure (1Y82) in stereo.

4.4 ITC of PF0355

The PIN domain was exhaustively investigated on a wide variety of potential substrates, based on structural alignment by DALI [51]: FAD, NAD(H), NADP(H), ATP, GTP, CTP, TTP, ADP, GDP, AMP, cAMP, cGMP, dATP, threitol, GlcNac and Chitobiose. Each of these substrates was tested in an isothermal titration calorimeter from CSC and Microcalorimetry. The only substrate PF0355p bound with a significant binding constant was ADP. The average binding constant of ADP by PF0355p was $3X10^4$ M^{-1} or a K_d=33 μM. (Figure 10A and B). These statistics show weak binding of ADP, the only detectable substrate of PF0355p. According to the mathematical fit of the data, two molecules of ADP with a total of four per dimer of PF0355p is the predicted stoichoimetry of binding. The kinetics demonstrated by the data show two non-interacting sites, based on evaluation by the respective manufacturer's software (Figure 10A and B).

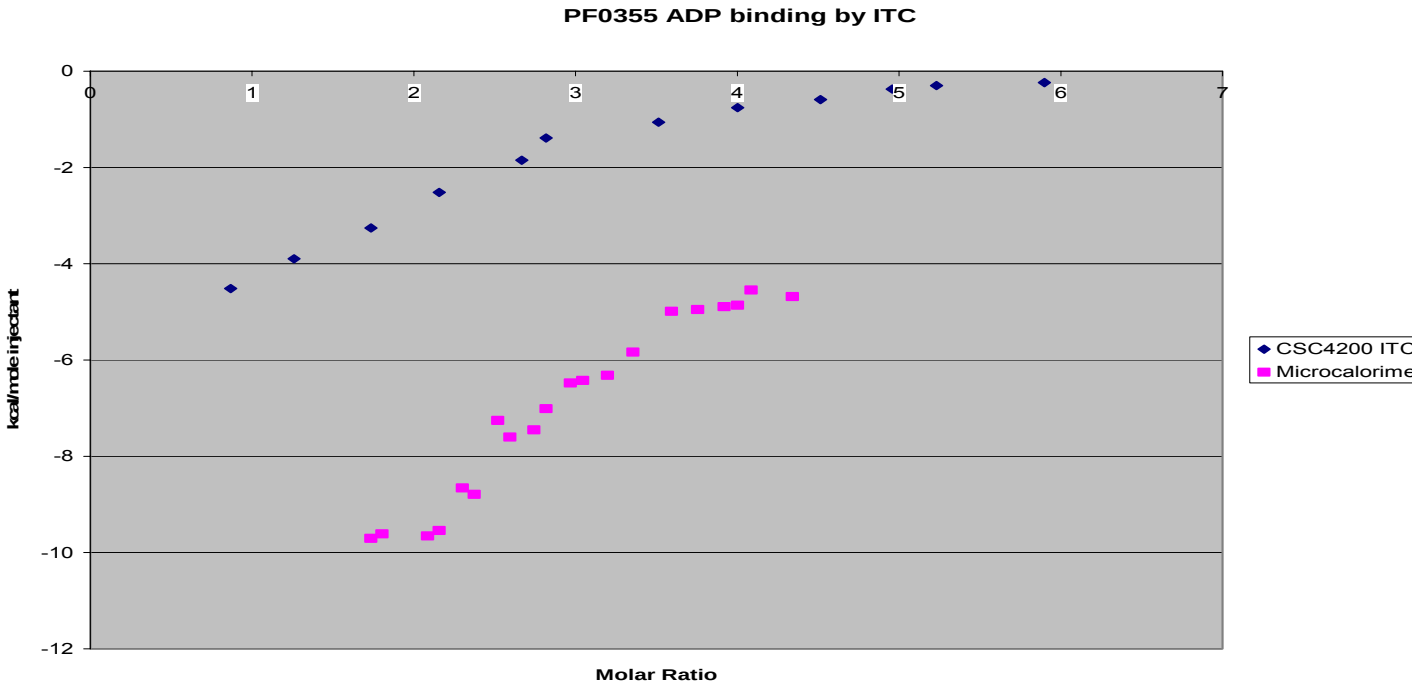
If these proteins follow the TA model, these the toxin and antitoxin proteins should form a complex. In order to see if the TA model is true here, the interaction of PF0355.1p and PF0355p was examined. When PF0355.1p was titrated into PF0355 or *vice versa*, no detectable binding was observed (data not shown). This was due to the low solubility of both proteins, since the injected substrate's manufacturer-recommended concentration in the syringe should be seven times more in the syringe than in the cell [58]. The only interpretable data were collected at higher than 0.05 mM PF0355p concentration. This meant that PF0355.1p's minimal concentration should logically have been 0.35 mM, which is almost two times what we could achieve.

Since the direct interaction of the protein was not detectable an alternate method was employed. Based on the TA model the antitoxin interacts with the toxin and inhibits its function. Here the binding of ADP by PF0355p would be inhibited in the presence of PF0355.1p if these proteins follow the TA model. When PF0355.1p was in equal concentration with PF0355p, it was not possible to detect ADP binding (Figure 10C). When 4mM ADP was titrated into the mixture of 0.025 mM PF0355.1p and 0.095 mM PF0355p, some binding of ADP by PF0355p was detected. The latter data was very noisy showing potent inhibition of PF0355p function by PF0355.1p (Figure 10D). The inhibition of the weak binding was difficult to further characterize, but the potent inhibitory effect of PF0355.1p was demonstrated on PF0355p binding of ADP.

Figure 10. The binding of ADP by PF0355p, the *Regulator,* demonstrated by Isothermal Titration Calorimetry (ITC). A. The ITC experiment performed on a CSC 4200 ITC. The protein showed specific binding of ADP. PF0355p was 0.095 mM in the cell and a 4 mM ADP solution was titrated into the cell in 10 µL injections. The second line is from an ITC experiment in a Microcalorimetry ITC. PF0355p was 0.095 mM in the cell and a 4 mM ADP solution was titrated into the cell in 3 µL injections. The average of these datasets show a K_d of 33 µM. **B.** The statistics of the two experiments in a table format. The experiment was conducted at 65°C. The enthalpy and stoichiometry of the reactions is depicted here. **C.** ADP binding by PF0355 was inhibited by the addition of PF0355.1. Equal molar amounts of 0.095 mM final concentration PF0355 and PF0355.1 solution were used in the cell of the instrument. A 4 mM ADP solution was titrated into the cell of the instrument. **D.** The titration of 4 mM ADP into 0.025 mM PF0355.1 and 0.095 mM PF0355 solution. The data are very noisy and the correct determination of the binding is not possible. The data were fitted with 1 binding site, which could explain the way PF0355.1 inhibits PF0355 binding.

Figure 10.

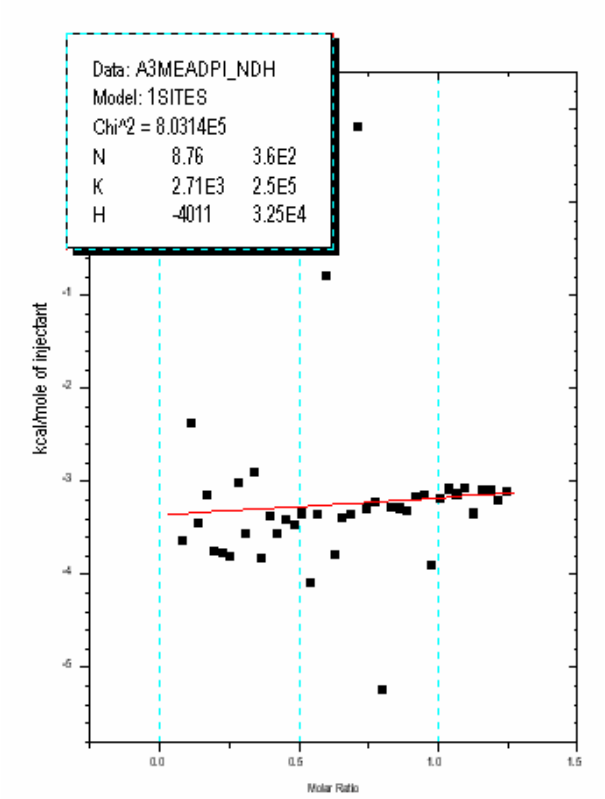
A.



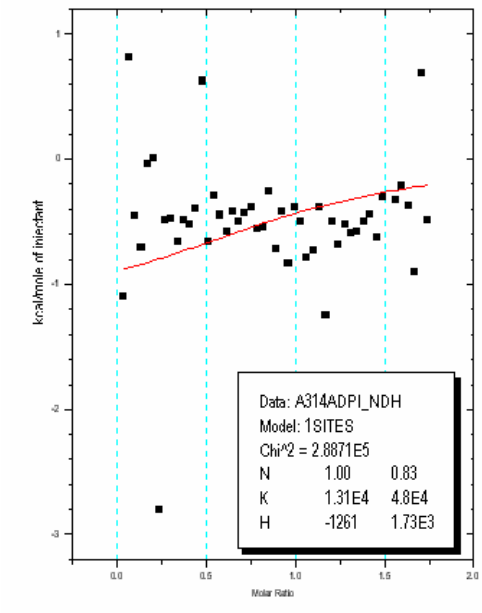
B.

Instrument	protein	substrate	N	K	ΔH
CSC 4200 ITC	PF0355p	ADP	1.97 ± 0.39	2.4X104 ± 2X104	-37 ± 10 kcal/ mole
Microcal ITC	PF0355p	ADP	2.49 ± .12	5 X104 ± 2 X104	-1 X104 ± 1 X103 kcal/ mole

C.


Data: A3MEADPI_NDH
Model: 1SITES
Chi^2 = 8.0314E5
N 8.76 3.6E2
K 2.71E3 2.5E5
H -4011 3.25E4
kcal/mole of injectant
Molar Ratio

D.


Data: A314ADPI_NDH
Model: 1SITES
Chi^2 = 2.8871E5
N 1.00 0.83
K 1.31E4 4.8E4
H -1261 1.73E3
kcal/mole of injectant
Molar Ratio

4.5 Predicted structure-based sequence alignment of PF0355p with select PIN domain-containing proteins.

In order to evaluate whether the PF0355p homologs shown in Table 2 also lack the (Asp, Glu, Asp and Asp) active site predicted by Arcus *et al.*, the structures of the Table 2 homologues were predicted with FUGUE V2.0 [41, 53]. Based on the best Z-scores, PDB entry 1V96, (92% sequence homolog with PF0355) was chosen as the modeling the template. The modeled structures were then overlaid on the 1V96 structure and residues corresponding to the four (Asp 10, Arg 44, Asp 98, and Asp 116) putative active site residues for each structure were noted. These amino acids are highlighted in red in (Figure 5). The residues making up the putative active site are quite divergent amongst these sequences. None of the PF0355 homologs encode the proper type (acidic) of residues to create a functional exonulcease active site, based on the 1V96 structure [41]. In the putative active site the most divergent position corresponds to the Arg 44 position in PF0355 while the residue (Asp) corresponding to position Asp 98 in the PF0355 structure was found to be the most conserved. Based on these structural alignments, none of these proteins in Table 2 is predicted to have exonulcease activity.

When this same method was used to determine why the archaeal PIN domains are not toxic an interesting phenomenon was observed. Based on the 1V96 structure the *Salmonella dublin* VapC, and *Shigelia flexneri* mvpT sequences were modeled, and their resulting structures superimposed on the 1V96 structure (see Figure 8E). The three proteins all have the acidic amino acids in the active site (side chains depicted as space-filling models). Figure 8E also shows the C-terminal helix (depicted in red) that is missing from the toxic PIN domain proteins, but is present in all archaeal non-toxic PIN domains tested. In addition all archaeal structures found in

the PDB have this extra C-terminal feature (either a helix or sheet) that appears to stabilize Asp 116, one of the amino acids that make up the putative active site in the PF0355 structure (Asp 110 in the Pae2754 structure, 1V8O). In the PF0355 structure, this helix is kept in place through interactions with a hydrophobic pocket formed by the residues Ile 7, Phe 9, Met 17, Ile 25, Ile 28, Phe 33, Leu 111, Ile 113, Met 131, Phe 136, Val 140, Met 144 and Val 145. The pocket is capped by the interaction of Glu 147 and Lys 32 (data not shown). The PF0355 C-terminal helix has high temperature factors (Figure 7) similar to all other homologous structures (data not shown), indicating that it may be unstructured in solution. This suggests that a possible role for the C-terminal helix is in influencing or shielding the active site thus blocking the killer action of those archaeal proteins. Conversely, those proteins in this family that lack this C-terminal extension allow open access to the putative catalytic site by solvent or substrate so that a diverse set of substrates may be affected.

4.6 Interaction of PF0355p and PF0355.1p

Analysis of PF0355p and PF0355.1p binding was done in two separate experiments. The interaction of PF0355p with PF0355.1p was first characterized using isothermal titration calorimetry where the binding of ADP by PF0355p was shown to be inhibited (Figure 10C and D). The interaction of PF0355p and PF0355.1p was also examined by native gel electrophoresis, however due to the pH of the buffer system; PF0355.1p was not visible in the gel (Figure 11, lane 8). Upon addition of PF0355.1p to PF0355p the electrophoretic banding pattern exhibited by PF0355p was altered, showing some interaction between the two proteins. The incubation of the two proteins at different temperatures and in the presence and absence of ADP seemed not to

alter the banding patterns in the gel, showing that the interaction between the proteins is not inhibited by ADP.

Based on the previous models of TA operons, the antitoxin, also known as the transcriptional repressor or the smaller upstream gene that overlaps with the PIN domain, is supposed to be protease sensitive, and the PIN domain's role is to prevent this degradation.

Thus, an experiment was designed to see whether PF0355.1p proves to be more susceptible to degradation than PF0355p. This experiment also tested if the association between PF0355p and PF0355.1p could offer protection from proteolysis to the PF0355.1p. Upon incubation of PF0355.1p with cellular extract the protein showed significant degradation as seen in (Figure 11 lanes 23, 24, 31, 32). Protection of PF0355.1p from proteolytic degradation in *P. furiosus* cellular extract was exhibited when the PIN protein, PF0355p, was present (Figure 11 lanes 20, 21, 28, 29). The protection PF0355.1p by PF0355p was not affected by the presence of ADP at 70°C. When incubated at 90°C (the nominal growth temperature) some degradation of PF0355.1p was observed (Fugre 10 Lane 28). The addition of ADP to the mixture at 90°C restored protection (Figure 11 Lane 29).

4.7 Prediction of β- glycosidase structure to infer function

Based upon numerous structures of the archaeal and bacterial proteins of the Glyco-Hydrolase-1 (GH-1) superfamily, a putative structure of PF0356p was predicted by FUGUE v 2.0 [53]. The amino acids ensuring the β-glycosidase activity were located, and the protein showed 38% sequence identity to the *Thermus thermophilus* structure [59]. The Gln and Asp amino acids ensuring glycosidic activity have been identified in this family through extensive

mutagenesis and are encoded in PF0356p by Gln 76 and Asp 205 [59]. The activity of the 95% homologous protein to PF0356p from *P. horikoshii* has been characterized and possessed glycosidase activity on substrates such as mannose, galactose, and glucose [60]. Those proteins listed in Table 2B are the only cytoplasmic version of GH-1 enzymes in these organisms, based on extensive BLASTP searches. The only similar proteins in these archaea are the membrane-bound counterparts of these enzymes, which have been shown to have alternate substrate and bond cleavage specificities, compared to their cytoplasmic counterparts [61].

Figure 11.

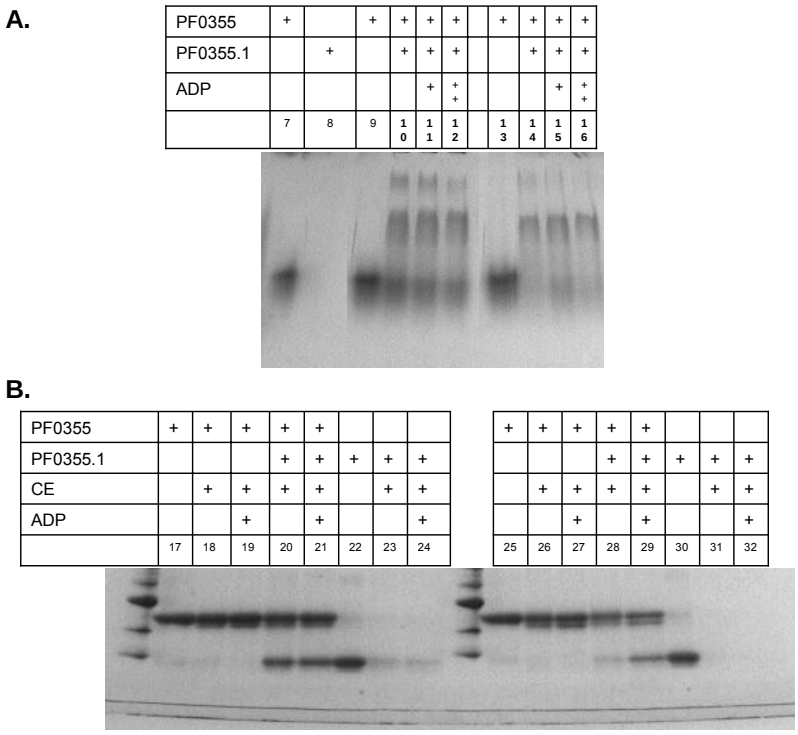


Figure 11. The **interaction of PF0355p the *Regulator* and PF0355.1p, the *Repressor.* A.** The native PAGE of the individual proteins and together. When PF0355.1p is run on a native gel with tris-glycine buffer, the protein does not run in the well as shown in lane 8. Lane 7 is PF0355p by itself. Lanes 9-12 were incubated at 70°C for 30 minutes. Lanes 13-16 were incubated at 90°C for the same time. Lanes 9-16 all have PF0355p present with the addition of PF0355.1p in wells 10-12 and 14-16. Every time PF0355.1p and PF0355p are incubated together, the banding pattern of PF0355p changes. Analytical gel filtration is not feasible to address this problem, since the proteins are 9 and 18kDa, which makes their identification not possible. **B.** The protease protection assay of PF0355.1p. An SDS-PAGE image of the protection assay is shown. Molecular weight markers are present in each unlabeled part of the gel. Lanes 17 and 25 have PF0355p in them alone, after incubation. Lanes 22 and 30 have PF0355.1p alone in them, also after respective incubation. *P. furiosus* extract was added to rest of the lanes to test if there is some cellular component that degrades either PF0355p or PF0355.1p. Lane 18 and 26 has PF0355p incubated with cell extract from the native organism. Lanes 19 and 27 has PF0355p plus its substrate, ADP incubated with cell extract. Lanes 20 and 28 has PF0355p plus PF0355.1p together incubated with cell extract to test protection. Lanes 21 and 29 has PF0355p and PF0355.1p incubated with ADP. Lanes 23 and 31 is PF0355.1p incubated with cell extract, erasing the band. Lanes 24 and 32 are PF0355.1p plus ADP incubated with cellular extract. Lanes 17-24 were incubated at 70°C while lanes 25-32 were incubated at 90°C for 3 hours.

5.0 Discussion

The toxin antitoxin family of proteins is a conserved gene family of microbes [24, 33]. They are found both on plasmids as well as chromosomes of these organisms. The most abundant toxin, as shown by Pandey *et al*., is VapC and it is annotated as a PIN domain [24**]**. These toxic proteins' substrate is not known. VapC's plasmid encoded counterparts in bacteria have been shown to kill the host in the absence of their antidotes [30]. It has been suggested by Gerdes and others that chromosomal TA operons are actually stress regulators and cannot kill the cell [36]. The goal of our project was to find a non-toxic VapC homolog to aid in the understanding of plasmid encoded VapC toxicity. Our search for such a protein took us to archeae, the most ancient group of organisms on Earth. Among these organisms we found that several structures have been solved of those archaeal VapC toxins. These PIN domain-proteins represented by three structures in the PDB showed no-toxic activity, since they were expressed individually. The VapC homologs' sequence varies greatly but their structure is almost identical with backbone RMSD of less than 4 Å amongst all four known PIN domain structures [41] (Figure 8C). Although they showed no toxicity, the DNA dependent nuclease function of the PIN domain has been established through these structures [41]. We chose one PIN protein from *P. furiosus,* PF0355p, which showed promise in not-killing the host, based on replacement mutation in its predicted active site of Glutamate to Arginine in position 44, Figure 5. This protein also can be found in a unique operon that gave additional promise of a non-toxic protein. Our protein did not show the predicted nuclease activity and that is why we named this protein as Regulator. Because of these data we called this protein Reg1.

A function of Reg1 was determined to be association with ADP, which is not lethal to *E. coli* during over-expression, passing the standard test of toxicity [10, 11, 29]. Nucleotide binding is a characteristic of nucleases [41], but no modification of DNA by PF0355p was observed. This lack of activity can be explained by the architecture of the active site of this protein when compared to that proposed by Arcus *et. al.* [41] by the substitution of arginine in place of glutamate in position 44 our protein's sequence, Figure 5, Figure 8E. Based on the solvent density inside of the protein two potential ADP binding pockets were identified Figure 9. The first pocket outlined by the residues backbone oxygen of Phe 9 and Ala 105 along with the side chain oxygens of Thr 8, Ser 11, Asn 36, Ser 38, Thr 41, Glu 97, Thr 101, Thr 108, and Ser 110. The second binding pocket is lined with backbone oxygens of Leu 37, Ile39, Val 70, Pro 71 and Leu 100. The side chain oxygen of Ser 38 and Thr 104 also contribute to the second pocket.

When comparing all archaeal non-toxic PIN domains with determined structures to those known toxic proteins' predicted structure, a structural reason beyond the in-tact active site for this cytotoxicity can be seen, Figure 8E. The C-terminus of all archaeal PIN domains is a longer, secondary structure-containing segment than those found in toxic VapC proteins. The Arcus predicted active site's Asp 110 is located at the base of this C-terminal helix in 1V8O [41]. This aspartic acid's flexibility would be enhanced by the absence of the twenty-eight amino acid C-terminal structure, helix in 1V8O, 1V96 and 1Y82, while it is a β-strand in 1O4W. The absence of this C-terminal, so far archaea specific, structure may uncover the nuclease active site to increase active site accessibility and could therefore kill the cell. Our structure has excessive thermal motion in this C-terminal region, as demonstrated in Figure 7 and by the absence of the C-terminal seven amino acids in chain D. This thermal motion indicated by the temperature factors deposited with the structures is present in all other structures (data not shown).

In order to try and understand where the *Regulator* interacts with the *Repressor*, the sequence divergence of all predicted Regulators listed in Table 2B were evaluated in Figure 5. Since the role of these proteins is to interact with their *Repressor* protein, in the case of PF1206p and PF0355p in the same organism, then the place of interaction influences the sequence divergence of each protein. Sequence divergence of up to 36% can be noted on amongst those *Regulator* proteins listed in Table 2B. The region of the protein that is the most divergent is highlighted in yellow in Figure 5. These yellow labeled regions localize to one surface of the protein into a large 27Å wide grove, where the interaction of these proteins may happen. This region was predicted as the DNA-interaction region by Arcus *et. al.*[41]. This wide grove in the center of the protein with the section of highest divergence is the region we predict as the *Repressor Regulator* interaction grove. These inactive toxins or *Regulator* proteins have one known activity that may influence the cell, which is the regulation of the enzyme's expression encoded in their operon.

PF0355.1p is a novel archaeal protein with sequence homologs only in *Thermococcales*. This protein has never been shown as a VapB homolog, or a potential antitoxin, but based on our data it follows the model in every aspect tested. This protein is next to a non-toxic toxin, therefore we named this gene *Rep*1. The competition of Rep1 away from the promoter was not successful with a variety of substrates such as chitobiose, GlcNAC, or ADP, products of the surrounding genes. This protein has no structural data available. The presence of the *Regulator* did not inhibit or enhance the DNA binding as shown by gel mobility shift and DNAseI footprinting. This protein is the *Repressor* and by the location of its DNA binding, it potentially regulates the expression of the operon. The protein binds a mirror repeat that makes it a unique DNA binding protein. This unique characteristic of the protein may explain why its function was

not predicted. The putative recognition sequences of the related *RR* operons are listed in Table 2B.

PF0356 is a glycosyil hydrolase family-1 enzyme with 95% sequence homology to the characterized enzyme from *Pyrococcus horikoshii ot3* and 71% to the characterized enzyme from *Thermococcus kodakarensis KOD1*. These enzymes work on substrate such as chitosan cellobiose or even lactose-6 Phosphate [60]. This family of archaeal enzymes has a preference for fucose or glucose over galactose [60]. The structural prediction allowed for us to locate the active site that is predicted by functional analyses of the homologs. The GH-1 family of enzymes present in these *RR* operons is the only copy of this cytoplasmic enzyme in the respective organisms. Other membrane bound glycosyl hydrolases are also present in these genomes, but their substrate specificity is different from those cytoplasmic proteins [61]. These facts together show that the enzyme encoded in these operons is essential for the cell. Since the organisms are Saccharolytic archaea, they require these enzymes to help degrade the disaccharides inside the cell. Other potential functions may be present with these proteins and should be investigated later.

One of the proposed models of activation of this *RR* operon is through the same protease-induced short-induction mechanism that is well characterized amongst all lethal-phenotype TA proteins [31]. The protein's expression remains at a basal level through all conditions tested by the Adams group with microarray in *P. furiosus* (data not shown). Here, the degradation of the *Repressor* and the momentary induction of the operon allow the "extra" enzyme, a β-glycosidase, to be expressed along with the *Repressor*. This mechanism of leaky expression of the genes in this operon facilitates the partial breakdown of a variety of substrates and the induction of the more tightly regulated carbohydrate-degrading operons in the cell. This pattern

of behavior with the only homologous gene (lacS) in *Sulfolobus solfataricus* has been noted by Haseltine *et al.* [62].

If these *Repressor Regulator* proteins' function is not to perform their own enzymatic role in the cell, then they are reduced to the still retained auto-regulatory function and they only act as expression regulators of the "extra" enzyme, here PF0356. The mechanism of expression-regulation, as predicted by the model, is through Rep1 stability. The *Repressor* binds the promoter of PF0356. The *Regulator's* presence does not inhibit this interaction between *Repressor* and the DNA, but it protects the labile *Repressor* from degradation. These interactions influence transcriptional regulation of PF0356 and the rest of the operon therefore making this an *RR* operon.

These *RR* operons are predicted to be self-regulating in nature and they contain both the functional enzyme and the regulators of expression for these genes. This mode of regulation is an effective way of constitutive expression and can be the original role of cTA operons. This may be a form on an ancient switching mechanism that later found a role in PSK.

The above findings support the prediction of Gerdes and others [5, 36]. These genes are clearly non-toxic and regulate the expression of other genes. It was essential to identify those proteins that are part of the toxin family, but cannot function as effective bacteriocidal agents. The role of these operons is not to kill the cell, but to regulate cellular responses inside the cell based on stimuli to be identified, but potentially the same, as other TA operons. The identification of this unique TA operon (PF0356-PF0355) shows that these proteins are *Regulators*, with some performing regulation through killing. The identification of the first non-toxic toxin and establishing these proteins as *Regulators* is another step towards understanding the full potential of this family of toxins. The further comparison of the toxic and non-toxic

proteins can aid science in designing a new era of bacteriocidal agents where the real potent toxins encoded in the organisms' chromosome can be specifically activated. These data also give clues to the possibility that the *Repressor* of these operons may inhibit the expression of other genes, which still remains to be seen. Beyond those already-mentioned benefits, this system of hundreds of complimentary interacting proteins gives a good example for studying protein-protein interactions.

6.0 References

1. Engelberg-Kulka, H., et al., *Bacterial programmed cell death systems as targets for antibiotics.* Trends Microbiol, 2004. **12**(2): p. 66-71.
2. Jensen, R.B., R. Lurz, and K. Gerdes, *Mechanism of DNA segregation in prokaryotes: replicon pairing by parC of plasmid R1.* Proc Natl Acad Sci U S A, 1998. **95**(15): p. 8550-5.
3. Gerdes, K., P.B. Rasmussen, and S. Molin, *Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells.* Proc Natl Acad Sci U S A, 1986. **83**(10): p. 3116-20.
4. Sat, B., et al., *Programmed cell death in Escherichia coli: some antibiotics can trigger mazEF lethality.* J Bacteriol, 2001. **183**(6): p. 2041-5.
5. Hayes, C.S. and R.T. Sauer, *Toxin-antitoxin pairs in bacteria: killers or stress regulators?* Cell, 2003. **112**(1): p. 2-4.
6. Franch, T. and K. Gerdes, *Programmed cell death in bacteria: translational repression by mRNA end-pairing.* Mol Microbiol, 1996. **21**(5): p. 1049-60.
7. Gerdes, K., *Toxin-antitoxin modules may regulate synthesis of macromolecules during nutritional stress.* J Bacteriol, 2000. **182**(3): p. 561-72.
8. Gerdes, K., et al., *The hok killer gene family in gram-negative bacteria.* New Biol, 1990. **2**(11): p. 946-56.
9. Sayeed, S., et al., *Surprising dependence on postsegregational killing of host cells for maintenance of the large virulence plasmid of Shigella flexneri.* J Bacteriol, 2005. **187**(8): p. 2768-73.
10. Pullinger, G.D. and A.J. Lax, *A Salmonella dublin virulence plasmid locus that affects bacterial growth under nutrient-limited conditions.* Mol Microbiol, 1992. **6**(12): p. 1631-43.
11. Sayeed, S., et al., *The stability region of the large virulence plasmid of Shigella flexneri encodes an efficient postsegregational killing system.* J Bacteriol, 2000. **182**(9): p. 2416-21.
12. Ogura, T. and S. Hiraga, *Mini-F plasmid genes that couple host cell division to plasmid proliferation.* Proc Natl Acad Sci U S A, 1983. **80**(15): p. 4784-8.
13. Jaffe, A., T. Ogura, and S. Hiraga, *Effects of the ccd function of the F plasmid on bacterial growth.* J Bacteriol, 1985. **163**(3): p. 841-9.
14. Karoui, H., et al., *Ham22, a mini-F mutation which is lethal to host cell and promotes recA-dependent induction of lambdoid prophage.* Embo J, 1983. **2**(11): p. 1863-8.
15. Van Melderen, L., P. Bernard, and M. Couturier, *Lon-dependent proteolysis of CcdA is the key control for activation of CcdB in plasmid-free segregant bacteria.* Mol Microbiol, 1994. **11**(6): p. 1151-7.
16. Engelberg-Kulka, H. and G. Glaser, *Addiction modules and programmed cell death and antideath in bacterial cultures.* Annu Rev Microbiol, 1999. **53**: p. 43-70.
17. Maki, S., et al., *Partner switching mechanisms in inactivation and rejuvenation of Escherichia coli DNA gyrase by F plasmid proteins LetD (CcdB) and LetA (CcdA).* J Mol Biol, 1996. **256**(3): p. 473-82.

18. Maki, S., et al., *Modulation of DNA supercoiling activity of Escherichia coli DNA gyrase by F plasmid proteins. Antagonistic actions of LetA (CcdA) and LetD (CcdB) proteins.* J Biol Chem, 1992. **267**(17): p. 12244-51.
19. Bravo, A., G. de Torrontegui, and R. Diaz, *Identification of components of a new stability system of plasmid R1, ParD, that is close to the origin of replication of this plasmid.* Mol Gen Genet, 1987. **210**(1): p. 101-10.
20. Lehnherr, H., et al., *Plasmid addiction genes of bacteriophage P1: doc, which causes cell death on curing of prophage, and phd, which prevents host death when prophage is retained.* J Mol Biol, 1993. **233**(3): p. 414-28.
21. Roberts, R.C., A.R. Strom, and D.R. Helinski, *The parDE operon of the broad-host-range plasmid RK2 specifies growth inhibition associated with plasmid loss.* J Mol Biol, 1994. **237**(1): p. 35-51.
22. Tian, Q.B., et al., *A new plasmid-encoded proteic killer gene system: cloning, sequencing, and analyzing hig locus of plasmid Rts1.* Biochem Biophys Res Commun, 1996. **220**(2): p. 280-4.
23. Gronlund, H. and K. Gerdes, *Toxin-antitoxin systems homologous with relBE of Escherichia coli plasmid P307 are ubiquitous in prokaryotes.* J Mol Biol, 1999. **285**(4): p. 1401-15.
24. Pandey, D.P. and K. Gerdes, *Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes.* Nucleic Acids Res, 2005. **33**(3): p. 966-76.
25. Gerdes, K., S.K. Christensen, and A. Lobner-Olesen, *Prokaryotic toxin-antitoxin stress response loci.* Nat Rev Microbiol, 2005. **3**(5): p. 371-82.
26. Aizenman, E., H. Engelberg-Kulka, and G. Glaser, *An Escherichia coli chromosomal "addiction module" regulated by guanosine [corrected] 3',5'-bispyrophosphate: a model for programmed bacterial cell death.* Proc Natl Acad Sci U S A, 1996. **93**(12): p. 6059-63.
27. Gotfredsen, M. and K. Gerdes, *The Escherichia coli relBE genes belong to a new toxin-antitoxin gene family.* Mol Microbiol, 1998. **29**(4): p. 1065-76.
28. Christensen, S.K. and K. Gerdes, *RelE toxins from bacteria and Archaea cleave mRNAs on translating ribosomes, which are rescued by tmRNA.* Mol Microbiol, 2003. **48**(5): p. 1389-400.
29. Takagi, H., et al., *Crystal structure of archaeal toxin-antitoxin RelE-RelB complex with implications for toxin activity and antitoxin effects.* Nat Struct Mol Biol, 2005. **12**(4): p. 327-31.
30. Zhang, Y.X., et al., *Characterization of a novel toxin-antitoxin module, VapBC, encoded by Leptospira interrogans chromosome.* Cell Res, 2004. **14**(3): p. 208-16.
31. Jensen, R.B., et al., *Comparison of ccd of F, parDE of RP4, and parD of R1 using a novel conditional replication control system of plasmid R1.* Mol Microbiol, 1995. **17**(2): p. 211-20.
32. Jensen, R.B. and K. Gerdes, *Programmed cell death in bacteria: proteic plasmid stabilization systems.* Mol Microbiol, 1995. **17**(2): p. 205-10.
33. Anantharaman, V. and L. Aravind, *New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system.* Genome Biol, 2003. **4**(12): p. R81.

34. Makarova, K.S., et al., *Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell.* Genome Res, 1999. **9**(7): p. 608-28.
35. Christensen, S.K., et al., *RelE, a global inhibitor of translation, is activated during nutritional stress.* Proc Natl Acad Sci U S A, 2001. **98**(25): p. 14328-33.
36. Christensen, S.K., et al., *Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA.* J Mol Biol, 2003. **332**(4): p. 809-19.
37. Christensen, S.K. and K. Gerdes, *Delayed-relaxed response explained by hyperactivation of RelE.* Mol Microbiol, 2004. **53**(2): p. 587-97.
38. Christensen, S.K., et al., *Overproduction of the Lon protease triggers inhibition of translation in Escherichia coli: involvement of the yefM-yoeB toxin-antitoxin system.* Mol Microbiol, 2004. **51**(6): p. 1705-17.
39. Gerdes, K., et al., *Mechanism of postsegregational killing by the hok gene product of the parB system of plasmid R1 and its homology with the relF gene product of the E. coli relB operon.* Embo J, 1986. **5**(8): p. 2023-9.
40. Fatica, A., D. Tollervey, and M. Dlakic, *PIN domain of Nob1p is required for D-site cleavage in 20S pre-rRNA.* Rna, 2004. **10**(11): p. 1698-701.
41. Arcus, V.L., et al., *Distant structural homology leads to the functional characterization of an archaeal PIN domain as an exonuclease.* J Biol Chem, 2004. **279**(16): p. 16471-8.
42. Liu, Z.J., et al., *The high-throughput protein-to-structure pipeline at SECSG.* Acta Crystallogr D Biol Crystallogr, 2005. **61**(Pt 6): p. 679-84.
43. Studier, F.W., *Protein production by auto-induction in high density shaking cultures.* Protein Expr Purif, 2005. **41**(1): p. 207-34.
44. Wilson, D.O., P. Johnson, and B.R. McCord, *Nonradiochemical DNase I footprinting by capillary electrophoresis.* Electrophoresis, 2001. **22**(10): p. 1979-86.
45. Shah, A.K., et al., *On increasing protein-crystallization throughput for X-ray diffraction studies.* Acta Crystallogr D Biol Crystallogr, 2005. **61**(Pt 2): p. 123-9.
46. Chayen, N.E., *The role of oil in macromolecular crystallization.* Structure, 1997. **5**(10): p. 1269-74.
47. Liu, Z.J., et al., *Parameter-space screening: a powerful tool for high-throughput crystal structure determination.* Acta Crystallogr D Biol Crystallogr, 2005. **61**(Pt 5): p. 520-7.
48. *The CCP4 suite: programs for protein crystallography.* Acta Crystallogr D Biol Crystallogr, 1994. **50**(Pt 5): p. 760-3.
49. McRee, D.E., *XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density.* J Struct Biol, 1999. **125**(2-3): p. 156-65.
50. Word, J.M., et al., *Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms.* J Mol Biol, 1999. **285**(4): p. 1711-33.
51. Holm, L. and C. Sander, *Mapping the protein universe.* Science, 1996. **273**(5275): p. 595-603.
52. DeLano, W.L., *The case for open-source software in drug discovery.* Drug Discov Today, 2005. **10**(3): p. 213-7.
53. Shi, J., T.L. Blundell, and K. Mizuguchi, *FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties.* J Mol Biol, 2001. **310**(1): p. 243-57.

54. Bateman, A., et al., *The Pfam protein families database.* Nucleic Acids Res, 2004. **32**(Database issue): p. D138-41.
55. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic Acids Res, 1994. **22**(22): p. 4673-80.
56. Poole, F.A., M. W. W., *TBA.* TBA, 2005.
57. Terwilliger, T.C. and J. Berendzen, *Automated MAD and MIR structure solution.* Acta Crystallogr D Biol Crystallogr, 1999. **55 (Pt 4)**: p. 849-61.
58. Personal communications with CSC, D.T., and Microcal Tech Support, Verna Frasca, Ph.D. February 18th 2005. April 15th 2005.
59. Lokanath, N.K., et al., *Structure of aldolase from Thermus thermophilus HB8 showing the contribution of oligomeric state to thermostability.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 10): p. 1816-23.
60. Kaper, T., et al., *Substrate specificity engineering of beta-mannosidase and beta-glucosidase from Pyrococcus by exchange of unique active site residues.* Biochemistry, 2002. **41**(12): p. 4147-55.
61. Matsui, I., et al., *Novel substrate specificity of a membrane-bound beta-glycosidase from the hyperthermophilic archaeon Pyrococcus horikoshii.* FEBS Lett, 2000. **467**(2-3): p. 195-200.
62. Haseltine, C., et al., *Coordinate transcriptional control in the hyperthermophilic archaeon Sulfolobus solfataricus.* J Bacteriol, 1999. **181**(13): p. 3920-7.
63. Zivanovic, Y., et al., *Pyrococcus genome comparison evidences chromosome shuffling-driven evolution.* Nucleic Acids Res, 2002. **30**(9): p. 1902-10.