Mechanism of RNA Cleavage by VapC from *Pyrobaculum aerophilum*

A thesis submitted in partial fulfillment

of the requirements for the degree

of

Master of Science in Biological Sciences

at

The University of Waikato

by

JOHANNA MARIA DUYVESTYN

The University of Waikato

2012

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Abstract:

Proteins belonging to the PIN-domain family are widespread across bacteria, archaea and eukaryotes. PIN-domain structures are well characterised and contain three highly conserved acidic residues that are orientated to form the active site of the enzyme. VapC is the toxic component of a toxin-antitoxin (TA) complex, and contains a PIN-domain. The VapBC TAs are the largest TA family, and are found in expanded copy numbers in a variety of unrelated organism including the human pathogen *Mycobacterium tuberculosis*. It is proposed that TAs play a role in metabolic regulation, under conditions of stress. VapCs are metal dependent endoribonucleases that target specific sequences in single stranded RNA and inhibit translation by degrading mRNA transcripts (McKenzie, 2011). The mechanism by which VapC cleaves RNA and achieves this specificity is unknown

In this thesis, VapC 2754 from *Pyrobaculum aerophilum* ($VapC_{PAE2754}$) is analysed, and a mechanism for catalysis of RNA cleavage by PIN-domains is proposed. Mutations were made to conserved residues in the putative active site to test the hypothesized mechanism. A fluorometric kinetic assay was designed to allow differences in activity to be accurately determined, so that conclusions about the effect of the mutations could be drawn. Mutations made to the conserved acidic amino acids eliminated RNase activity, confirming that these residues are essential to catalysis. A structure for the wild-type $VapC_{PAE2754}$ was determined, containing cadmium ions and an acetate bound into the active site.

Knowledge of the enzymatic mechanism of PIN domains will enable a greater understanding of this important protein family, and serve as the basis for future drug design by transition state analogue inhibiters. This is especially important in VapC enzymes due to their proposed role in the pathology of *M. tuberculosis*.

Acknowledgements

Firstly and most importantly, I thank Associate Professor Vic Arcus, who has not only been supportive and encouraging but who has provided a lab in which it was a pleasure to study. The weekly meetings allowed me to receive feedback from everyone in the lab in a constructive environment, reassuring me that I was not alone in getting difficult results.

On that note I would also like to thank all of my colleagues that make up the team, both in our lab, and across the hall. Most especially Dr. Joanna McKenzie who showed me the ropes from day one and was always happy to help analyse results or decide what to do next. Also to the Genetics lab MSc girls, Rebecca, Nadine, Claire and Tania, I am so grateful to have had such wonderful classmates to be excited or disappointed along side me. To Judith for organizing everyone and everything, and making sure we never lacked anything (including the vitally important coffee and cake), to Jo H, Marisa, Emma L, Emma S, Ali, Tifany, Emily, Abbi, Erica, Mat, John, Mark, Joel, Ashley, Prof. Dave Musgrave, Prof. Dick Wilkins and of course Prof. Ray Cursons, thank you all for your input and encouragement.

To my 'non-science' friends and flat-mates, and particularly to my family and to Anthony, I am indebted to you all for your love and support, for showing interest (or pretending to) despite the fact that half of what I said was nerdish and most likely boring to anyone who doesn't wear a lab coat all day.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

SI (Systeme Internationale d'Unités) abbreviations for units and standard notations for chemical elements and formulae are used throughout this thesis. Other abbreviations are listed below.

A	adenosine
Amp	ampicillin
ATP	adenosine triphosphate
bp	base pair (s)
C	cytosine
C-terminal	carboxy terminus
Da	daltons
DEPC	diethyl pyruvate carbonate
DNase	deoxyribonuclease
ds	double-stranded
EDTA	ethylene diamine tetraacetic acid (disodium salt)
em.	emission
ex.	excitation
fit	fast intracellular trafficking
FEN	FLAP endonuclease
FPLC	fast protein liquid chromatography
G	guanine
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid
HGT	horizontal gene transfer
His-tag	poly-histidine tag
HIV	human immunodeficiency virus
HPA	hydroxypicolinic acid
HPLC	high performance liquid chromatography
IDT	integrated DNA technologies
IMAC	immobilised metal affinity chromatography

IPTG	isopropylthio-β-D-galactosidase
kb	kilobase
kDa	kilo dalton
kV	kilo volt
LB	luria bertani
MAD	multi-wavelength anomalous diffraction
MALDI	matrix assisted laser desorption ionisation
mAU	milli-absorbance units
MDR	multi-drug resistant
mRNA	messenger RNA
Mtb	*Mycobacterium tuberculosis*
MS	mass spectrometry
MW	molecular weight
Native-PAGE	non-denaturing PAGE
NMD	nonsense mediated decay
N-terminal	amino terminus
NYN	Nedd4-BP1, YacP nuclease
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCD	programmed cell death
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PIN	PilT N-terminal domain
PSK	post segregational killing
rpm	revolutions per minute
RNase	ribonuclease
rRNA	ribosomal RNA
SDS	sodium dodecyl sulphate

SOC	super optimal broth with catabolite repression
ss	single stranded
T	thymine
TA	toxin antitoxin
TAE	tris-acetate-EDTA
TB	tuberculosis
TE	tris EDTA buffer
TEMED	N, N, N, N,-tetramethylethylenediamine
TOF	time of flight
tRNA	transfer ribonucleic acid
tmRNA	transfer-messenger ribonucleic acid
Tyr	tyrosine
U	uracil
UV	ultra violet
Vap	virulence associated protein
$VapBC_{PAE2754}$	vapBC 2754 from *Pyrobaculum aerophilum*
$VapBC_{PAE0151}$	vapBC 0151 from *Pyrobaculum aerophilum*
$VapBC_{Mtb}$	vapBC from *Mycobacterium tuberculosis*
$VapBC_{Msmeg}$	vapBC from *Mycobacterium smegmatis*
$VapBC_{Sflex}$	vapBC from *Shigella flexnari*
v/v	volume per volume
WHO	world health organisation
WT	wild-type
w/v	weight per volume
w/w	weight per weight
XDR	extensively drug resistant

Chapter 1. Introduction:

In 1993 tuberculosis (TB) was deemed a global health emergency, with a third of the world's population is predicted to carry the disease. The most recent World Health Organisation (WHO) report states that mortality due to TB was still as high as 1.3 million deaths in 2010 (WHO, 2011). TB is caused by an infection of *Mycobacterium tuberculosis*, an intracellular pathogen that can exist as an asymptomatic infection over extended periods of time, and in a significant number of individuals progress to active infection. This bacteria has the ability to form dormant populations which can evade antibiotics and allow resistant strains to arise (Connolly, Edelstein, & Ramakrishnan, 2007). The current treatment regime requires multiple antibiotics to be taken for a period of six to twelve months. Due to non-compliance, multidrug resistant (MDR) and extensively drug resistant (XDR) cases are increasing in number. MDR cases require at least 20 months of treatment, and yet only a 60-75% cure rate is achieved with this treatment with recurrence remaining a serious risk. Therefore, new drugs are required to combat both infectious and latent forms of this disease if the treatment goal of halving the number of TB cases by 2015 is to be reached (WHO, 2011).

To survive within its host, the *M. tuberculosis* pathogen must be able to adapt to the environmental changes that occur within the host cells. They are subjected to such as nutritional, hypoxic, nitrosative and acidic stress conditions (Ahidjo et al). Toxin-antitoxin (TA) systems have been proposed to act as stress response elements that allow the bacteria to adapt rapidly, and to survive harmful conditions. They have also been linked to the development of persistent populations (Ramage, Connolly, & Cox, 2009). Of the classified TAs, the VapBC family is the largest. VapBC TA's are found in expanded numbers in *M. Tuberculosis*, and thought to allow rapid response within the bacteria through their ribonuclease activity (Gerdes, Christensen, & Lobner-Olesen, 2005). The expansion of these systems implies that they play an important role in the survival of the organisms. Therefore, TAs are of great interest to the study of the pathology of M. tuberculosis and other harmful bacteria, as potential candidates for new drug targets.

1.1 General Properties of Toxin-Antitoxin Systems

Toxin-antitoxin genes are found in a bicistronic operon that encodes a toxic protein and its specific antitoxin. The antitoxin has two functions; it binds to the toxin forming a benign complex, and it is a DNA binding protein that targets the promoter region of the operon functioning as a repressor of transcription in both its free and complexed states. The toxin is stable, but the antitoxin is susceptible to degradation. This means that without continual transcription, the antitoxin will be degraded and the toxin is free to exert it effects (Arcus, Rainey, & Turner, 2005). TA's are widely spread within all domains of life, and are found in high copy numbers in many bacterial pathogens (Georgiades & Raoult, 2011; Yamaguchi, Park, & Inouye, 2011). Nine TA families have been classified thus far; CcdB, RelBE, MazEF, ParDE, phd/doc, VapBC, HicA, HipA and ω-ε-ζ (the only three component family) (Gerdes, et al., 2005). Little is confirmed about what role TAs play within the cell, but they have been linked to antibiotic tolerance, persistence and bacteriostasis in a range of bacteria (Agarwal, Mishra, Bhatnagar, & Bhatnagar, 2010; Buts, Lah, Dao-Thi, Wyns, & Loris, 2005; Gerdes, et al., 2005; Hayes, 2003; Van Melderen & De Bast, 2009).

1.1.1 Regulation of Toxin-Antitoxin Systems

TA operons are auto-regulated through the antitoxin. The system is summarized schematically in Figure 1.1. The antitoxin has a N-terminal DNA binding motif that binds to palindromic or repeat sequences in the promoter region to repress transcription. In most cases, the toxin acts as a co-repressor, increasing the binding affinity (Eberl, Givskov, & Schwab, 1992).

For example in the RelBE system a heterotrimeric complex ($RelB_2$.RelE) binds to the promoter. However, when RelE is in excess, it destabilises the TA-DNA complex by binding to the C-terminal region of RelB (Overgaard, Borch, & Gerdes, 2009; Overgaard, Borch, Jorgensen, & Gerdes, 2008). CcdAB and phd/doc are also regulated by the ratio of toxin to antitoxin (Afif, Allali, Couturier, & Van Melderen, 2001; Garcia-Pino et al., 2010). MazE inhibits transcription in the form of a heterohexamer, made up of two toxin

homodimers connected by one antitoxin homodimer ($MazF_2.MazE_2.MazF_2$) (Kamada & Hanaoka, 2005).

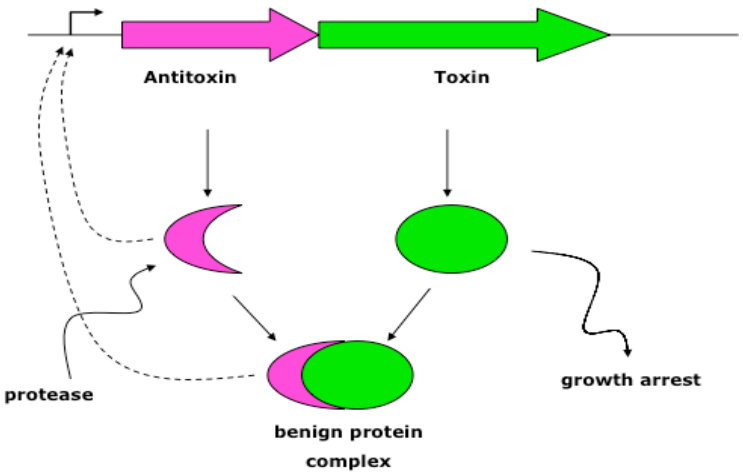


Figure 1.1. Schematic diagram of the toxin-antitoxin regulation system. Dotted arrows show antitoxin and the TA complex binding back to the promoter region (square arrow). The antitoxin (pink) is more susceptible to protease degradation than the toxin, which targets transcription or replication to induce 'toxic' effects such as growth arrest (Figure adapted from McKenzie 2011)

1.1.2 Modes of Action of the Toxin

Characterised toxins act by targeting free or ribosome bound RNA to impede translation, or by targeting DNA gyrases to inhibit replication (Buts, et al., 2005; Gerdes, et al., 2005; Van Melderen & De Bast, 2009). This is summarized in the table below (Table 1.1) for each of the nine identified TA systems.

TA Family	Toxin Target	Effect on cellular process
RelBE	Translating Ribosome	Translation through inducing mRNA cleavage
MazEF	RNA	Translation through endoribonuclease activity
VapBC	RNA	Translation through endoribonuclease activity
CcdAB	DNA gyrase	Replication through DNA gyrase
ParDE	DNA gyrase	Replication through DS breaks
Phd/doc	Translating ribosome	Translation through inducing mRNA cleavage
HipBA	EF-Tu	Translation through inducing mRNA cleavage
HicBA	RNA	Translation through inducing mRNA cleavage
ω-ε-ζ	ND	Phosphotransferase activity

Table 1.1. Targets and cellular activity of the toxins from the nine TA families currently described. Table adapted from Van Melderen and Bast (2009). ND- not determined, DS- double stranded.

CcdB and ParE both act as gyrase inhibitors. Gyrase is a class II topoisomerase required to introduce negative supercoils in DNA to relieve

the constraints of transcription and replication. The toxins act as direct inhibitors to form a complex with gyrase and impede its function. CcdB can also trap the gyrase in complex with DNA to stall the unraveling process (Gerdes, et al., 2005)

RelE and Doc bind at the A-site of the ribosome, inhibiting translation. RelE causes codon specific cleavage of both mRNA and tmRNA, but only when they are associated with ribosomes. (Christensen & Gerdes, 2003).

MazF and VapC target mRNA independently of the ribosome, and target specific sequences, that differs in different organisms (Condon, 2006; McKenzie, 2011). HicA also targets RNA however it is uncertain if it does so directly (Jørgensen, Pandey, Jaskolska, & Gerdes, 2009; Van Melderen, 2010).

1.1.3 Proposed Functions of TA modules

1.1.3.1 Discovery as Plasmid Stabilisation Genes

TA operons were first discovered in prokaryote plasmids, where they functioned to reduce the occurrence of plasmid free cells by post segregational killing (PSK). The plasmid that contains the TA gene is 'addictive' because it is required for continual inhibition of the toxin protein. During replication, daughter cells will inherit the benign TA protein complex. Over time the antitoxin is degraded away, but the stable toxin will remain. Without the plasmid that contains the TA operon, new antitoxin cannot be transcribed and the ratio of free toxin will increase. This gives a selective advantage to cells that retain the plasmid. In this scenario the genes therefore act as selfish entities (Gerdes, Rasmussen, & Molin, 1986; Hayes, 2003; Pandey & Gerdes, 2005).

1.1.3.2 Chromosomal Operons

Chromosomal TA genes are homologous to plasmid systems but have been implicated in a range of different roles (Buts, et al., 2005). It is proposed that the expansion of these elements occured by horizontal gene transfer. They are now widely distributed in chromosomes of diverse prokaryotes (Chim et al., 2011; Van Melderen, 2010). While there is an array of literature that

exists, the role that TA systems fulfill within the bacteria remains uncertain. Many functions have been proposed thus far, from selfish elements to growth regulation (Magnuson, 2007).

The addictive properties of TA systems suggest that they could play a role in the stabilisation of genetic parasites. By being in close proximity to the TA operon, the deletion rate of the chromosomal parasite would be reduced. Similarly, alleles that contained a TA operon would not be able to be replaced by a non-addictive allele, creating a slight benefit for the addictive element (Magnuson, 2007; Van Melderen & De Bast, 2009).

A number of roles related to the regulation of cell growth have been proposed for TA systems, as preventative measures or direct responses to environmental stress. TAs have been related to persistence, in which a subpopulation is insured against rapid environmental change. They could also play a role in controlled cell growth, where early growth is limited to conserve resources and improve long-term survival. Additionally TA systems have been implicated in the ability of a bacteria to respond to stress by bacteriostasis, programmed cell death and gene regulation (Magnuson, 2007). Finally, TA systems have also been linked to biofilm formation, where the regulation allows clusters of cells to survive in close proximity (X. X. Wang & Wood, 2011).

Much of the experimental evidence supports the regulation of cell growth as the most probable function of TA systems (Chim, et al., 2011). Evaluation of the function of the toxins provides evidence for roles related to the control of gene expression, cellular activity, and replication. Entering a dormant state or rapidly reprogramming the proteome can allow bacteria to tolerate environmental stresses such as hypoxia, low pH, macrophage infection, antibiotic infection and nutrient limitation (Ahidjo et al., 2011; Keren, Minami, Rubin, & Lewis, 2011).

1.1.3.3 Distribution of TA operons

The phylogeny of TA systems supports their function as elements involved in stress response. It seems likely that their ability to function as selfish

elements is the best explanation for their evolutionary origin, but that they have since evolved functional roles in many organisms. The fact that TA systems are maintained in high copy numbers in in some organisms but have been lost from others that are closely related lends support to this theory. For example, the obligate intracellular pathogen *Rickettsia prowzekii* contains no TA systems, while the facultative intracellular *Rickettsia conorii* contains 15 (Arcus, McKenzie, Robson, & Cook, 2011). Most parasitic or symbiotic prokaryotes lack TA loci and this has been linked to their stable environmental conditions (Makarova, Wolf, & Koonin, 2009). A comparative study of pathogenic bacteria found a higher number of TA systems in 'bad bugs' (defined as the 12 most dangerous pandemic causing bacteria for humans) in comparison to their most closely related non-pathogenic species (Georgiades & Raoult, 2011). The distribution of VapBC systems in the *Mycobacteria* genus is consistent with this pattern. *M. tuberculosis* and *Mycobacterium bovis* (mammalian pathogens) contain expanded numbers of VapBC operons, while the closely related pathogen *Mycobacterium leprae* contains none, and the relatively fast growing non-pathogenic *Mycobacterium smegmatis* contains only one, as does *Mycobacterium* avium susp. *paratuberculosis* (Arcus, et al., 2005).

1.1.3.4 Programmed Cell Death (PCD)

This theory proposes that TA modules induce cell suicide in a sub-population of altruistic cells as a response to environmental stress. However, there is little uncontested evidence that exists for PCD, and in light of evolution single celled organisms should not contain a large numbers of suicide genes (Pandey & Gerdes, 2005).

The Engleberg-Kulka group proposed that when the MazF toxin is activated during logarithmic growth phase, the bacteria can reach a point of no return at which stage PCD is activated and there is no recovery. Though these results have been controversial, more recent results confirmed cell death of most of the population, and also showed cell survival in a few select individuals (Amitai, Kolodkin-Gal, Hananya-Meltabashi, Sacher, & Engelberg-Kulka, 2009; Engelberg-Kulka, Amitai, Kolodkin-Gal, & Hazan, 2006; Hayes,

2003). Kenn Gerdes' group also revealed PCD mediated by RelE and ChpAK in *E. coli* but in later experiments concluded that the TAs actually induced a reversible bacteriostatic state (Kim Pedersen et al., 2003).

1.1.3.5 Persistence:

Persistent bacteria are small proportions of a population that enter a dormant state, able to survive stresses such as antibiotics. Therefore, they act as a bet hedging strategy. Persistence is not induced in response to an environmental stress, but is rather a stochastic mechanism related to population density. Unlike drug resistant strains that evolve by mutation, these persister cells are not genetically different and are only drug tolerant while they remain in a dormant state (Hayes, 2003; Lewis, 2007). While the majority of the population can benefit from maximized reproduction in growth promoting environments, the persistent individuals pay the price of proliferation to be protected against potential future stresses (Lewis, 2007).

Transcriptome analysis of persistent cells at different times in the growth curve found that many of the genes down-regulated were involved in energy metabolism pathways, and that within the up-regulated genes there were ten TA modules (Keren, et al., 2011). Another study successively deleted ten different TA loci in *E. coli* (MazEF, ChpB, RelBE, yefm/yoeB, dinJ/yafQ, HigAB, YafNO, MqsRA, YhaV, HicAB). This resulted in up to a 200 fold decrease in the number of persistent colonies formed, despite deletion of individual operons having no visible effect. It was concluded that a range of TA operons fulfill a similar role within the cell, implying redundancy (Maisonneuve, Shakespeare, Jorgensen, & Gerdes, 2011; Ramage, et al., 2009).

A HipBA mutation was discovered in an isolated *E. coli* strain that showed a 1000 fold increase in persistence (Moyed & Bertrand, 1983). When researched more extensively, it was established that HipBA regulated the onset and duration of the growth arrest, controlled by an expression threshold of the operon (Rotem et al., 2010). HipBA has also been shown to allow multidrug tolerance in *E. coli*, linking the persistence function to a biological response to stress (Korch & Hill, 2006).

TA operons are conserved in the genomes of many organisms with persistent phenotypes. In addition to *M. tuberculosis*, they can be found in *Haemophilus influenzae*, *Helicobacter pylori*, *Coxiella burnetti*, *Leptospira interrogans*, *Vibrio cholera* and *Salmonella enterica* serovars Typhi and Typherium (Daines, Wu, & Yuan, 2007).

1.1.3.6 Bacteriostasis:

TA systems have also been implicated in the induction of bacteriostasis, where the activity of the cell is shut down. In contrast to persistence, the TA systems play a role in triggering dormancy of the entire population (rather than a sub-population), in response to an environmental stress (rather than in case of an environmental stress) (Buts, et al., 2005).

As mentioned previously, Gerdes and colleagues showed that RelE and ChpAK both induced a static condition. The cells were viable but unable to proliferate, and this state was reversible through addition of the cognate antitoxins. Due to the stability of the toxin, the static state can be maintained without further transcription (K. Pedersen, Christensen, & Gerdes, 2002). Results that show bacteriostasis rather than cell death have also been obtained for MazEF, using a loci in *E. coli* which induced a dormant state from which the cells could be resuscitated (Christensen, Pedersen, Hansen, & Gerdes, 2003).

Support for the stasis hypothesis has also been obtained from other systems, including ParDE in *Caulobacter cresentus*, VapC and RelE homologues in *M. tuberculosis,* and RelE homologue systems in *E. coli* (Christensen-Dalsgaard, Jorgensen, & Gerdes, 2010; Fiebig, Rojas, Siegal-Gaskins, & Crosson, 2010; Ramage, et al., 2009)

1.1.3.7 Global Regulation Response

In this proposed role for TA systems, the mRNA is degraded to conserve energy, and allow a rapid change in the transcriptome of the cell. This is required for triggering bacteriostasis, but can also be part of a non-dormant stress response. The metabolic state of the cell is reprogrammed so that energy is only being used for expression of transcripts that are required for

response to the current stress (Condon, 2006; Gerdes, et al., 2005; Ramage, et al., 2009).

For example, RelE inhibits translation (incompletely) via a reduction in the number of charged tRNA molecules (Gerdes, et al., 2005). VapC-6 in *sulfolobus solfataricus* is required for normal rates of growth and for heat shock survival. Two other VapCs in *S. solfataricus* have also been related to the heat shock response (Maezato et al., 2011). In *M. tuberculosis,* different stresses induced different TA systems. Two systems were induced under hypoxia, and two others under macrophage infection (Ramage, et al., 2009). Differential expression was also shown for RelE homologues in *E. coli.* Three systems were induced in response to both amino acid starvation and inhibition of translation, but they were induced to different extents. Two systems were induced in response to carbon starvation, and one in response to DNA damage (Christensen-Dalsgaard, et al., 2010). This indicates that although the toxins may have similar functions within the cell, they are not redundant because they are activated in response to different environmental changes.

1.2 VapBC TA Systems

The VapBC TA family is grouped by their virtue of their toxin proteins, which contain a pilT N-terminal (PIN) motif. Like the systems described above, the VapB gene is located upstream of the VapC gene, and binds into the toxin to inhibit its activity. The benign VapBC complex auto-regulates the transcription of the operon by binding to repeat sequences in the promoter region. The toxins function as sequence specific endoribonucleases that regulate translation and induce a bacteriostatic state (Ahidjo, et al., 2011; Arcus, et al., 2011; Dienemann, Boggild, Winther, Gerdes, & Brodersen, 2011; Ramage, et al., 2009; Robson, McKenzie, Cursons, Cook, & Arcus, 2009). This section will review the characterised VapC proteins from *Mycobacteria*, *P. Aerophilum* and other organisms focusing on general features of their function, regulation, and any information relevant to the mechanism of cleavage.

1.2.1 PIN-domains

Overall, PIN-domains have low sequence conservation. They are defined by a triad of conserved acidic residues, and a less well conserved fourth acidic residue (Figure 1.2). The distribution of the conserved asparagines (D) and glutamine (E) is illustrated in the Hidden Markarov model below, in which the size of the residue represents the importance of its contribution to the PIN-domains. The four acidic residues are distant in the sequence, but arrange into a negatively charged pocket constituting the active site when the protein is folded (Arcus, et al., 2005).

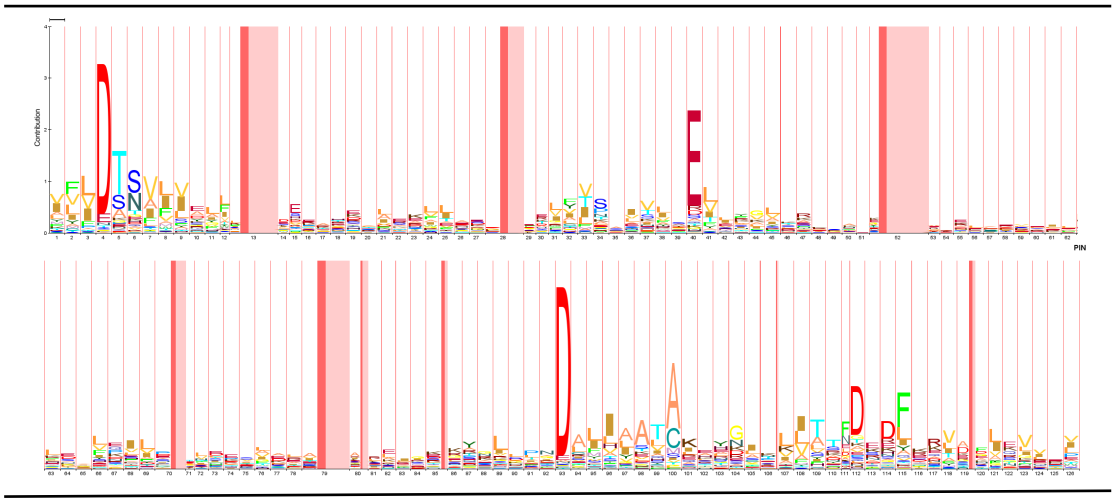


Figure 1.2. The Hidden Markov Model which defines the PIN-domain family of proteins. The height of each letter is proportional to the importance of the respective position in the PIN-domain family. The width of each column is also an indication of the importance of this position in defining the family. Sequence regions that contain insertions are shown as dark and light pink columns.

In the Pfam database there are 6253 PIN-domain sequences currently described in 1341 species from all three kingdoms of life (www.pfam.sanger.ac.uk). The majority of the sequences are found in bacteria, and *M. tuberculosis* has the highest number of PIN-domains out of any of the organisms. Structures have currently been determined for FitB from *Neisseria gonorrhoeae,* AF0951 and AF1683 from *Archeoglobus fulgidus*, PF0355 from *Pyrococcus furiosus*, PH0500 from *Pyrococcus horikoshii*, Est1A, SMG5 and SMG6 from human and PAE2754 and PAE0151 from *Pyrobaculum aerophilum*. The domains show structural similarity to RNase H enzymes and the 5'-3' flap endonucleases (FEN) (Arcus, Backbro, Roos, Daniel, & Baker,

2004). Therefore structural data and proposed mechanisms of RNA cleavage for RNase and FEN enzymes will be used to infer a mechanism for PIN-domain RNA cleavage.

The structures for SMG-5 and SMG-6 PIN-domain proteins show that the conserved acidic residues are essential for RNase activity (Figure 1.3). SMG-5 and SMG-6 are involved in nonsense mediated decay (NMD) of mRNA. SMG-6, which contains the acidic triad, shows metal dependent cleavage of ssRNA. SMG-5 is similar to SMG-6 in sequence and structure, but contains only one of the conserved residues, and shows very low RNase activity in vitro and in vivo (Glavan, Behm-Ansmant, Izaurralde, & Conti, 2006).

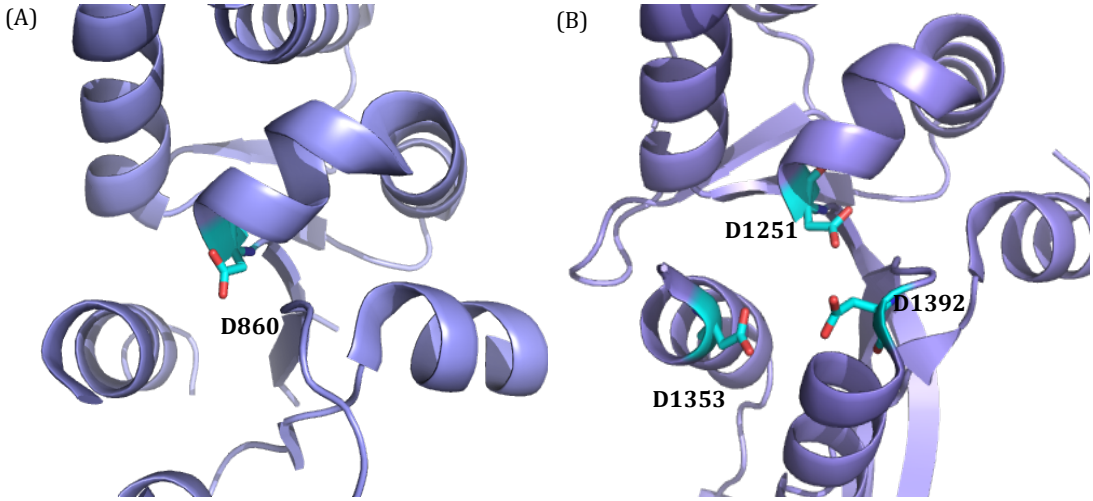


Figure 1.3. Cartoon structures of SMG-5 (A) and SMG-6 (B). Conserved PIN-domain residues are coloured by element in cyan (carbon) and red (oxygen) and labeled. This figure and subsequent structure figures were produced using PYMOL.

1.2.2 Characterised VapC Systems in Other Organisms

Though VapBC systems have only recently been discovered, they are now known to be found in a wide range of bacteria and archaea. The data collected thus far supports VapC toxins as stress response elements that function to inhibit translation via mRNA degradation (Arcus, et al., 2011; Ramage, et al., 2009).

1.2.2.1 Functions and Modes of Action

In *Sinorhizobium meliloti,* the VapBC system NtrPR regulates transcription of nitrogen fixation genes and a wide range of other metabolic processes under

a microoxic environment (Puskas et al., 2004). NtrPR also reduced colony formation and cell growth in *E. coli* (Bodogai et al., 2006). *Haemophilus influenzae* VapC-1 and VapC-2 cleave free RNA in vitro (but not DNA), and cause growth inhibition in vivo (Daines, et al., 2007). When expressed in *E. coli* the VapC from *Leptospira interogans* induced growth arrest that was reversed with the antitoxin. The cellular target of the toxin is unknown (Y. X. Zhang et al., 2004).

For *Neisseria gonorrhoeae* the *vapBC* locus (*fitAB*) is implicated in slowing bacterial growth. *N. gonorrhoeae* is a pathogenic bacteria that can avoid the host immune response by adopting an asymptomatic persistent state. FitB has been implicated in this mechanism by slowing intracellular trafficking and replication, however the mechanism of action for this response is unknown (Mattison, Wilbur, So, & Brennan, 2006). The heat shock response of VapBC-6 in *S. solfatarius* was mentioned previously. This gene is also required for normal growth rates in the organism, and the VapC toxin has been shown to play an essential ribonucleolytic role in this response (Maezato, et al., 2011). Growth inhibition and bacteriostasis in *E. coli* has also been shown for VapBC from Salmonella LT2 and Shigella plasmid pMYSH6000 (Winther & Gerdes, 2009). Transcription of VapBC from Salmonella LT2 is induced by amino acid starvation. (Winther & Gerdes, 2009).

VapBC operons were found in 10 out of 12 species analysed from the *Rickettsia* genus, and five of these contain two modules in their genome(Audoly et al., 2011). *Rickettsia* are intracellular bacteria that can cause fatal infections in humans. The VapBC from two *Rickettsia* species were shown to cause growth inhibition related to RNase activity, in *E. coli* and yeast cells. Furthermore, VapC was released from the bacteria into the host cell cytoplasm in response to an antibiotic (chloramphenicol), causing apoptotic cell death. This is the first evidence of VapC causing harm in host cells, and suggests a potentially neglected role for TA in the pathogenicity of intracellular bacteria (Audoly, et al., 2011).

1.2.2.2 Toxin Inhibition by the Antitoxin

The structures solved for VapBC from *Shigella flexnari* ($VapBC_{Sflex}$) and FitAB from *N. gonorrhoeae* provide insights into how the antitoxin ($VapB_{Sflex}$, FitA) inhibits the activity of the toxin ($VapC_{Sflex}$, FitB) (Dienemann, et al., 2011). For simplification toxin and antitoxin residues are denoted with subscript 'tox' and 'a.t' respectively.

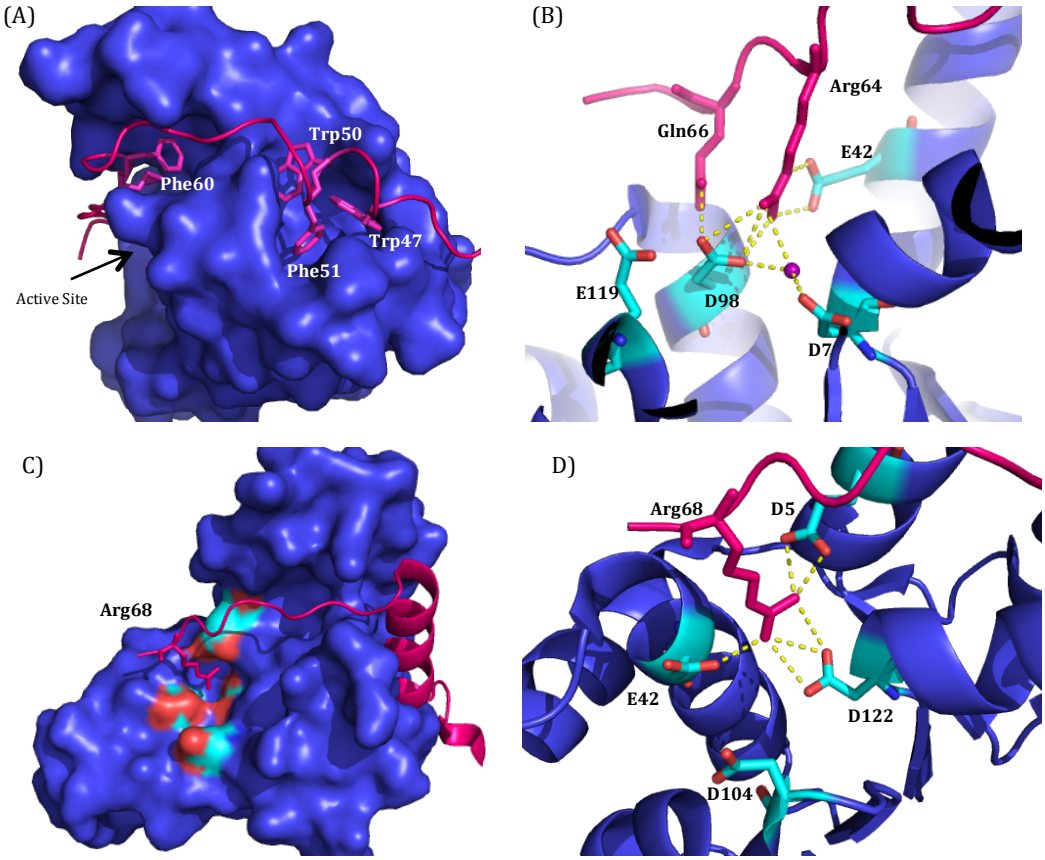


Figure 1.4. VapBC from *S. flexnari* and FitAB from *N. gonorrhoeae* showing interactions between the antitoxin and the toxin. In all figures the toxin is shown in blue and the antitoxin in pink, and conserved acidic residues (cyan) are shown as labeled sticks. **A)** VapC surface structure and VapB cartoon structure. The VapB residues that bind into the hydrophobic groove of the toxin are shown as sticks and labeled. The position of the active site pocket is also labeled. **B)** Active site of VapC with interactions to VapB residues shown as yellow dotted lines. A water molecule is also bound in the active site (purple sphere). **C)** Surface structure of FitB and cartoon structure of FitA binding. $Arg68_{a.t}$ is shown binding into the active site. **D)** Active site of FitA with the interactions to FitB residues shown as yellow dotted lines.

The extended C-terminal region of the antitoxin folds over a hydrophobic groove in the toxin and into the active site of the protein. The $VapB_{Sflex}$ residues Trp47, Trp50, Phe51 and Phe60 binding into the toxin groove (Figure 1.4A). $Arg64_{a.t}$ and $Gln66_{a.t}$ form close interactions to the conserved

acidic residues in the active site (Asp7 via a water molecule, and Glu42 and Asp98 directly) (Figure 1.4B). The FitAB structure also shows an arginine ($Arg68_{a.t}$) forming interactions with the carboxyl groups of the acidic active site residues (Asp5, Glu42, Asp104) (Mattison, et al., 2006). When the antitoxin residues are interacting with the active site residues the metal ion can no longer be coordinated, therefore catalysis is hindered (Dienemann, et al., 2011).

1.2.3 VapBC in Mycobacteria

As described earlier, there is large expansion of TA modules seen in the *Mycobacteria* genus. In *M. tuberculosis* there are 88 presumed TA systems in the chromosome and 45 of these are annotated as *vapBC* genes. The rest include MazEF, RelBE, ParDE, HigBA and a number of novel systems not found previously (Arcus, et al., 2005; Ramage, et al., 2009). Such high copy numbers in comparison with closely related species imply that the VapC TA complexes play a role that is important to the lifestyle and pathology of the bacteria (Ahidjo, et al., 2011).

1.2.3.1 Function and Mode of Action

M. smegmatis contains only three TA operons, MazEF, doc/phd and VapBC ($VapC_{Msmeg}$). This makes it a simplified candidate with which to characterise the VapBC operon in *Mycobacteria* (Robson, et al., 2009). Arcus and colleagues showed that $VapBC_{Msmeg}$ is transcribed as a single leaderless mRNA, and that the benign complex regulates the expression of the operon. Expression of $VapC_{Msmeg}$ in *M. smegmatis* reduces growth rate via inhibition of translation, and this is rescued by the cognate antitoxin (Robson, et al., 2009). The RNA cleavage was further characterised, and a specific cut-site AU*A(U/A) was determined for VapC-1284 (McKenzie, 2011). The enzyme seemed to have specificity for both sequence and secondary structure despite the fact that $VapC_{Msmeg}$ contains only three of the four conserved PIN-domain acidic residues.

Several VapBCs from *M. tuberculosis* ($VapC_{Mtb}$) were analysed in *M. smegmatis,* and showed inhibition of translation. RNase cleavage was shown

in vitro for $VapC_{Mtb}$ proteins rv0301 and rv1561, but with different levels of activity (Ramage, et al., 2009). The effect on growth rate was tested for 10 $VapC_{Mtb}$ operons in *M. smegmatis* and in *M. Tuberculosis.* The same 5 resulted in growth inhibition in both species (Rv0549c, Rv0595c, Rv2549c, Rv2829c and Rv3320c). Furthermore Rv0065-Rv0067 and Rv0616-Rv0617 cut a 150bp single stranded (ss) RNA strand in a sequence specific manner (Ahidjo, et al., 2011).

Non-cognate toxin-antitoxin interactions have been observed between different VapBCs, and between VapBC and MazEF proteins from *M. tuberculosis* (Zhu, Sharp, Kobayashi, Woychik, & Inouye, 2010). Binding of non-cognate antitoxins (both within and between different families) could increase the ability of the cell to tightly regulate metabolic processes in response to environmental cues. In contrast to this, two other studies in *M. tuberculosis* found no cross-talk between non-cognate toxin and antitoxin modules. Only the cognate antitoxins could rescue growth inhibition observed by Mizrahi and colleagues, and this was confirmed by yeast-two hybrid assays (Ahidjo, et al., 2011). Similarly only the cognate antitoxins could inhibit the VapBC activity in experiments carried out by Cox and colleagues (Ramage, et al., 2009).

1.2.3.2 VapBC Structures from M. tuberculosis

Two structures have been solved for $VapBC_{Mtb}$ proteins: rv0626/rv0627 (VapBC-5) and rv0300/rv0301 (VapBC-3) (Miallau et al., 2009). VapBC-5_{Mtb} was determined to 1.9Å using single anomalous dispersion (SAD). The acidic residues are distorted by the binding of VapB C-terminal and locked in an unfavourable position (Figure 1.5). This occurs by a less direct mechanism than that seen for $VapBC_{Sflex}$ and FitAB (Figure 1.4). $Arg75_{a.t}$ doesn't bind directly into the active site, but instead interacts with $Asp133_{tox}$ at the edge of the active site pocket. Additionally $Ala82_{a.t}$ orientates $Arg112_{tox}$ into a position where it locks the conserved residue $Glu57_{tox}$ in an unfavourable conformation (Miallau, et al., 2009).

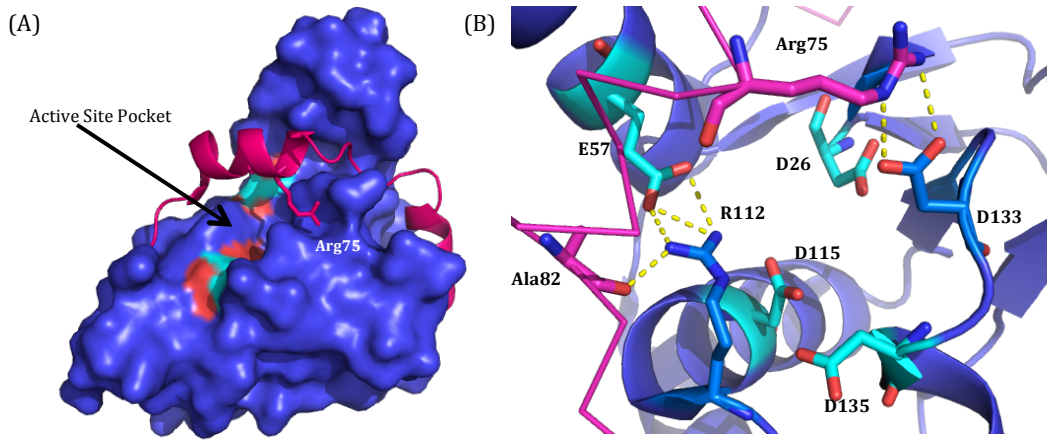


Figure 1.5. Structure of VapC-5 from *M. tuberculosis*. In both figures the toxin is shown in blue and the antitoxin in pink. **A)** VapC surface structure and VapB cartoon. The active site is labeled and conserved active site residues are shown as surface structures. $Arg75_{a.t}$ binds into a pocket near the active site. **B)** Active site of VapC. Residues interacting with the antitoxin and the conserved active site residues are shown as blue sticks, with interactions represented by yellow dotted lines.

The $VapBC\text{-}3_{Mtb}$ structure was solved to a resolution of 1.49Å by SAD. The toxin-antitoxin dimers are asymmetric. The C-terminal region of first antitoxin is well ordered and binds into the active site of the neighbouring dimer. $Arg73_{a.t}$ forces the active site into an unfavourable conformation, thus preventing the Mg^{2+} ion from binding (Figure 1.6a), similar to that seen for $VapC_{Sflex}$ and FitAB. However the C-terminal region of the second antitoxin is disordered, allowing an Mg^{2+} ion to bind into the active site of the neighbouring toxin. The metal is coordinated by three of the conserved active site residues (Asp114, Asp132, Asp134) (Figure 1.6b).

This has interesting implications for the catalytic mechanism, providing information on the orientation of the catalytic residues required to bind the Mg^{2+} (Chim, et al., 2011). Despite the fact that only a single Mg^{2+} is bound in the active site, a two metal mechanism cannot be ruled out. The antitoxin could still be constraining the residues, and without a substrate bound the true conformation of the active site cannot be confirmed. Eisenburg and colleagues have proposed Mg^{2+} dependence consistent with a two metal mechanism of cleavage, based on the similarity of $VapC\text{-}5_{Mtb}$ to FEN-1 (Miallau, et al., 2009).

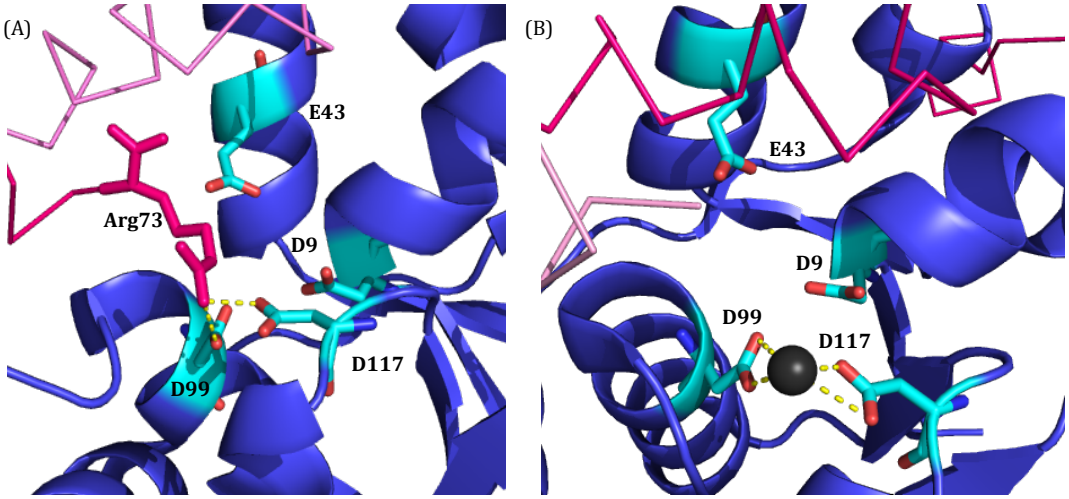


Figure 1.6. Active sites of neighbouring toxin-antitoxin dimers in VapBC-3 from *M. tuberculosis*. For each toxin the C-terminal region of the antitoxin of the neighbouring complex extends towards the active site pocket. **A)** Antitoxin II has an ordered C-terminal that binds into the active site of Toxin I. **B)** Antitoxin I has a disordered C-terminal allowing metal to bind into the toxin II active site. In both figures the toxin is shown in blue, and the antitoxins I and II are shown in pale pink and dark pink respectively. The ordered residue $Arg73_{a.t}$ II is shown as a labeled stick. Conserved acidic residues (cyan) are labeled. Waters (purple) and Mg^{2+} ions (grey) are shown as spheres, and potential hydrogen bonds are shown as yellow dotted lines.

1.2.4 VapBC in *Pyrobaculum aerophilum*

P. aerophilum contains 12 VapBC operons, and the structures for two of these ($VapBC_{PAE2754}$, $VapBC_{PAE0151}$) have been solved (Arcus, et al., 2004; Bunker, McKenzie, Baker, & Arcus, 2008). The toxin proteins can be purified without antitoxin and remain stable in solution over time. This makes *P. aerophilum* a good model for studying PIN-domain RNA cleavage. (Arcus, et al., 2011; Robson, et al., 2009).

Analysis of $VapC_{PAE2754}$ confirmed that like *M. smegmatis* and *M. tuberculosis* only ssRNA is cleaved, and specific sequences are targeted (McKenzie, 2011). Using pentaprobe RNA sequences and smaller pentaprobe oligos, J. L. McKenzie determined the consensus cut-sites of $VapC_{PAE2754}$ and $VapC_{PAE0151}$ to be GG*UG and the slightly less optimal GG*GG. Both proteins target the same sequence despite being from the same organism, but $VapC_{PAE2754}$ appears to cut the RNA more rapidly. It is possible that these proteins could be activated in response to different conditions. The determined cut-sites are frequent within the genome, however over half of the genes containing these recognition sites were not annotated. Because of this a metabolic process

that may be targeted by VapC$_{PAE2754}$ and VapC$_{PAE0151}$ has not yet been determined (McKenzie, 2011).

1.2.4.1 Structure of VapC$_{PAE2754}$ Protein

VapC$_{PAE2754}$ was the first VapC PIN-domain crystal structure obtained (Figure 1.7). The structure was solved to 2.5Å (Arcus, et al., 2004).

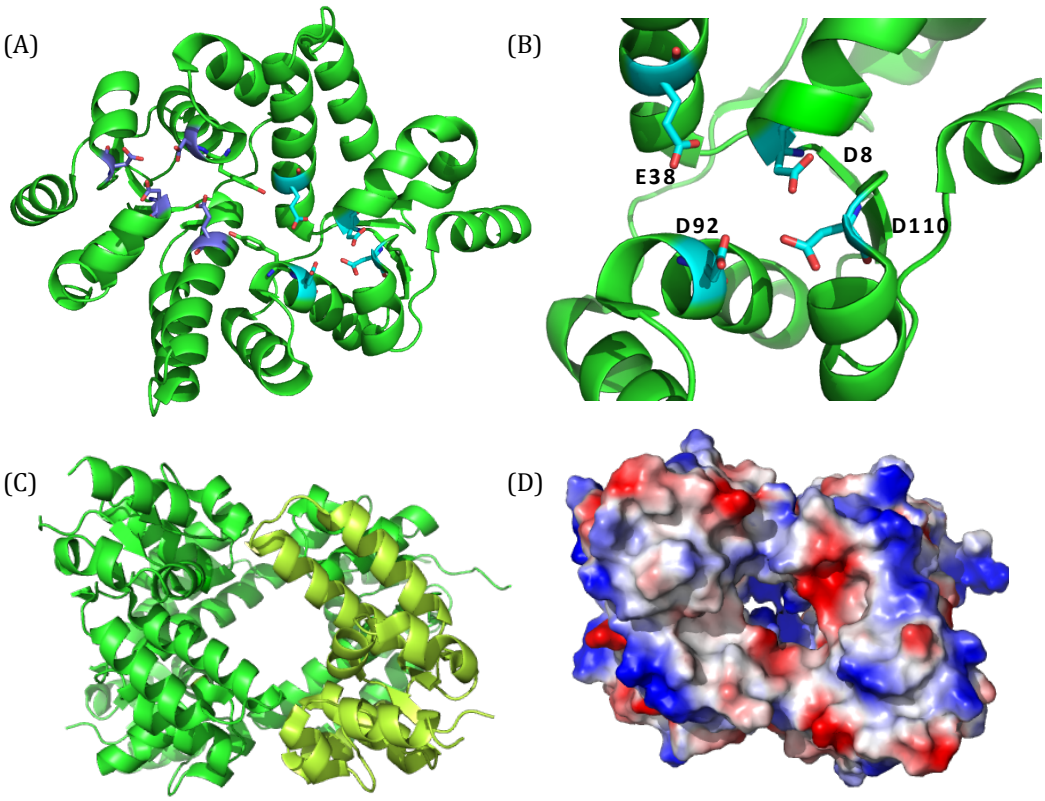


Figure 1.7. VapC$_{PAE2754}$ structure diagrams. A) Dimer showing both active sites (cyan and purple sticks). Tyrosines at the interface are shown as sticks. **B)** Toxin active site showing labeled conserved active site residues (cyan). **C)** Cartoon representation of tetramer (individual monomer in yellow, others in green). **D)** Surface structure diagram of tetramer coloured by electrostatic charge showing the tunnel formed. The active site is found within the tunnel.

The protein is folded as an α/β/α stack stabilized by hydrophobic cores above and below the central β-sheet. The β -sheet is made of 5 twisted parallel strands, and helices α1 and α2 pack in an anti-parallel manner with the strands, extending orthogonally out from the stack. The asymmetric unit is made up of three tetramer units, which are each a dimer of dimers (Figure 1.7A and C) The dimer interface is broad, and consists of hydrophobic interactions spanning helices α2, α3, and α4, burying 19% of the surface area of the monomer. At the center of the interface are stacked histidine rings that

form hydrogen bonds. The interface that forms the tetramer is less extensive (Figure 1.7C). It is formed by the C-terminus of helix α2 of one monomer and the N-terminus of helix α6 from the other, with few interactions between them (Arcus, et al., 2004). A tunnel is formed within which the active site is contained (Figure 1.7D). If the enzyme is a tetramer in solution the tunnel could provide selectivity for ssRNA strands.

The C-terminal end of the β-sheet and the N-termini of helices α2 and α6 bring together the four conserved acidic residues that constitute the active site, creating a negatively charged hole (Figure 1.7B). Thr-108 and Leu-112 flank the pocket, and Thr-108 is hydrogen bonded to Asp-8. In the dimer the active sites are separated by the interaction of the Tyr-91 residue from each monomer. The aromatic rings overlap parallel to one another, 6 Å apart (Arcus, et al., 2004).

1.2.4.2 Structure of $VapC_{PAE0151}$ Protein

$VapC_{PAE1051}$ structure forms a dimer very similar to $VapC_{PAE2754}$ (Figure 1.8A). An anomalous difference map was used to place a single Mn^{2+} ion in the active site (Figure 1.8B). The metal is incompletely coordinated by Asp118, and three water molecules (two of which are hydrogen bonded by Asp100) (Arcus, et al., 2004).

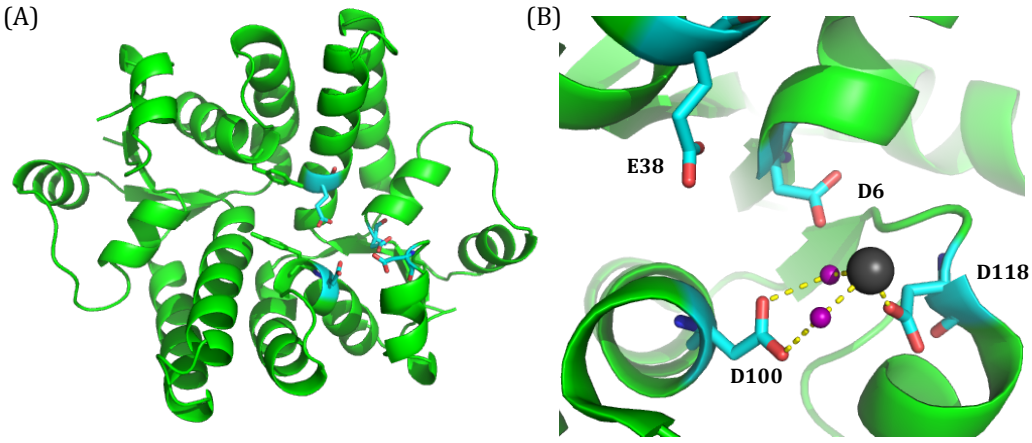


Figure 1.8. Structure of $VapC_{PAE0151}$. In both cartoons the conserved acidic site residues are shown in cyan. **A)** Dimer showing interacting tyrosine residues as sticks. **B)** Active site showing Mn^{2+} shown as a grey sphere and labeled acidic residues.

The collection of more structural data for $VapC_{PAE2754}$ would enable more accurate analysis active site, potentially confirming the position the metal ion

and a nucleophilic water. A substrate soak would provide further information about the interaction between the RNA and the active site molecules, potentially allowing the number of metal ions to be solved and providing insight as to why only RNA is cut by the protein.

1.3 RNase Mechanisms

Catalytic mechanisms have been proposed for a range of RNase proteins, but so far no mechanism for RNA cleavage by PIN-domains has been determined. The objective of this thesis was to propose a mechanism for $VapC_{PAE2754}$ by making comparisons of the substrate specificity, reaction products, metal cofactors and active site conformation of other nucleases. This will involve determining whether one or two metals are required for activity in PIN-domains proteins, and confirming which active site residues are essential to catalysis.

1.3.1 General Properties of Nucleolytic Reactions

Nucleases cleave DNA or RNA substrate at either the 3' or 5' bond of the phosphate binding two nucleotides together (Figure 1.9). This reaction occurs by general acid base catalysis, where a base activates a nucleophile by deprotonation, and an acid facilitates formation of the product by protonating the leaving group.

Most commonly the nucleophiles are deprotonated water molecules. The catalysis reaction usually occurs in a single step, divided into three stages. The nucleophile approaches from the opposite side of the bond to be cleaved and attacks the phosphate; a negatively charged pentacovalent phosphate transition state is formed; and the scissile bond is broken and the phosphate in the product has inverted stereochemistry. (Yang, 2011).

In RNases, the 2'OH of the ribose can also function as a nucleophile. The reaction proceeds via a 2'3' cyclic phosphate intermediate, and results in a 3' phosphate on the 5' product and a 5' OH on the 3' product (Figure 1.9 – Reaction B)(Thompson et al 94). RNase A and Barnase are examples of ribonucleases that utilize this mechanism (Mossakowska, Nyberg, & Fersht, 1989).

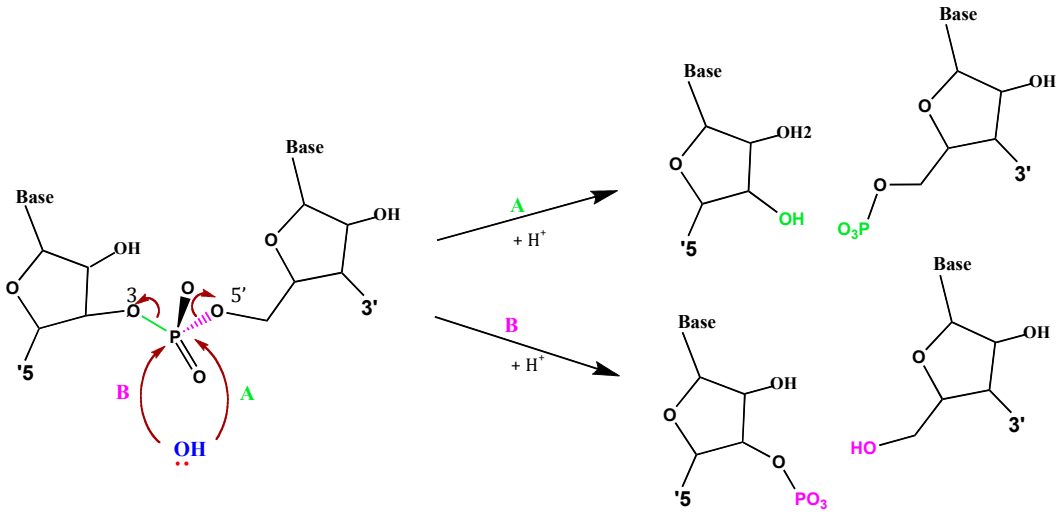


Figure 1.9. Products of general nucleolytic cleavage reaction of RNA. The substrate is labeled 5' and 3' and contains a general base that represents U, A, G or C. The movement of electrons from the nucleophile (blue) is denoted by red arrows. **Reaction A)** The 5' scissile phosphate bond is shown in green. The reaction products are a 5' product with a 3'OH, and a 3' product with a 5' phosphate. **Reaction B)** The 3' scissile phosphate bond is shown in pink. The products of this reaction are a 5' product with a 3' phosphate and a 3' product with a 5' phosphate.

1.3.2 Role of Divalent Cations

Nucleases often require one or more divalent metal cations as cofactors for activity. The metals increase substrate specificity and the efficiency of product release. The few reactions that are metal independent require the 2' OH of the ribose for nucleophilic attack, as described above. Metal independence is usually associated with single stranded unpaired substrate, and has positively charged residues in the active site to neutralize the negative charge of the pentacovalent transition state (Yang, 2011).

More commonly, two-metal catalysis mechanisms have been described. This was originally based upon structures of polymerases where the metal ions are bound by two conserved Asp residues. It is likely that Asp is preferable over Glu for this coordination because it is more rigid, a feature that is important for maintaining the correct alignment of the metals and the reactants in the active site (Yang, Lee, & Nowotny, 2006). Metal A is found on the nucleophile side of the scissile phosphate, where it is coordinated by non-bridging oxygens, waters and a conserved Asp residue. Metal ion B is on the 3'O leaving group side of the scissile phosphate, coordinated by non-bridging oxygens, water molecules and two acidic residues (Figure 1.10). It is this metal that is most often conserved in single metal mechanisms. The reaction

products of one and two metal mechanisms are always 5' phosphate and 3' OH (Yang, 2011).

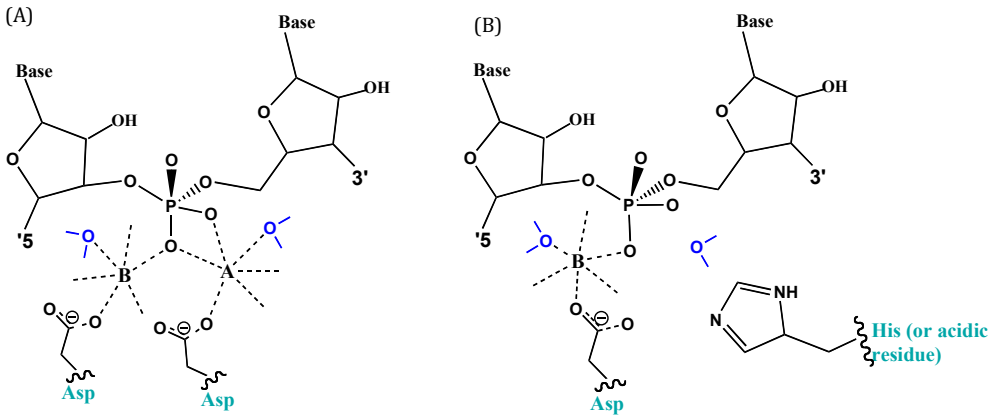


Figure 1.10. Positions of metals in two metal (A) and one metal (B) mechanisms. Metals A and B are labeled in with coordination shown as dotted lines. Conserved residues involved in the reaction mechanism are labeled in cyan. Water molecules are shown in blue. Modified from Yang (2011) and Dupureur (2010).

1.3.2.1 Metal Ion Co-factors

VapC enzymes are metal dependent and usually are most active in the presence of Mg^{2+} (Daines, et al., 2007; Dienemann, et al., 2011; Maezato, et al., 2011; McKenzie, 2011; Robson, et al., 2009). This metal dependency is due to the stability of the phosphate bonds, and the polyanionic substrates, which make breaking the phosphate bond difficult. Mg^{2+} has several key properties that make it well suited to function as the divalent cation in these proteins. Its dense positive charge acts to neutralize the negative charge of the nucleic acids, and the transition state. Mg^{2+} also has rigid coordination geometry of 6 inner-sphere ligands arranged octahedrally. The metal ion therefore can act to rigidly hold the key residues, nucleophile, and substrate in optimal conformation for catalysis to occur (Dupureur, 2010).

Manganese is chemically very similar to magnesium but has less stringent requirements for coordination. In the presence of Mn^{2+} many Mg^{2+} dependent enzymes are still active, but cleave non-ideal substrates with reduced specificity (Yang, 2011).

1.3.2.2 Arguments for One Metal or Two Metal Mechanisms.

It is controversial whether or not RNase enzymes require one or two metal ions. Crystallography data for many nucleases has shown one, two or even

three metal ions bound in the active site. However, difficulties in interpreting this information arise when considering factors such as whether or not a substrate was bound, and the concentration of the metal in the solution. It has been proposed that while the second metal ion increases specificity it is not vital for activity. In cases where the amino acids in the active site can function as a nucleophile or general base the second metal may not be vital to the mechanism (Dupureur, 2010). It will be important to determine whether or not the PIN-domain mechanism relies on one or two metals to help to solve the cleavage mechanism.

Investigation into the role of metal ions in FLAP endonuclease (FEN) reactions determined that there are at least two metal ions required for the activity of the enzymes and that at least one plays a catalytic role in the reaction mechanism (Tock, Frary, Sayers, & Grasby, 2003). For RNase H it was concluded that both metals are essential for catalysis and that they must be orientated correctly so that the metal ions shift closer together in the intermediate compared to the substrate structures (Nowotny, Gaidamakov, Crouch, & Yang, 2005). Mutations of residues involved in coordination of the 2^{nd} metal in T4 RNase H still resulted in an active enzyme, showing that the second metal is not directly required for catalysis (Bhagwat, Meara, & Nossal, 1997).

It is proposed that PIN-domains function by a mechanism that requires only one metal, and that the fourth acidic residue instead contributes directly to the active site, perhaps via a water molecule, rather than chelating a second metal (Anantharaman & Aravind, 2006). In contrast, the crystal structure of VapC-5_{Mtb} is likened to FEN-1, binding two metal ions, and a two metal mechanism has been proposed (Miallau, et al., 2009)

1.3.3 Metal Independent Mechanism

1.3.3.1 RNase A

RNase A has a cyclic mechanism as described above, utilizing the 2'OH as the nucleophile (Thompson, Venegas, & Raines, 1994). The catalytic mechanism still requires an acid and a base, and this is fulfilled by conserved histidines in the active site (Figure 1.11A).

1.3.3.2 Barnase

The key conserved active site residues in barnase were determined using site directed mutagenesis to be His102, Glu73 and Arg87 (Figure 1.11B). The mechanism has been shown to be consistent with the canonical cyclic mechanism, similar to RNase A (Mossakowska, et al., 1989).

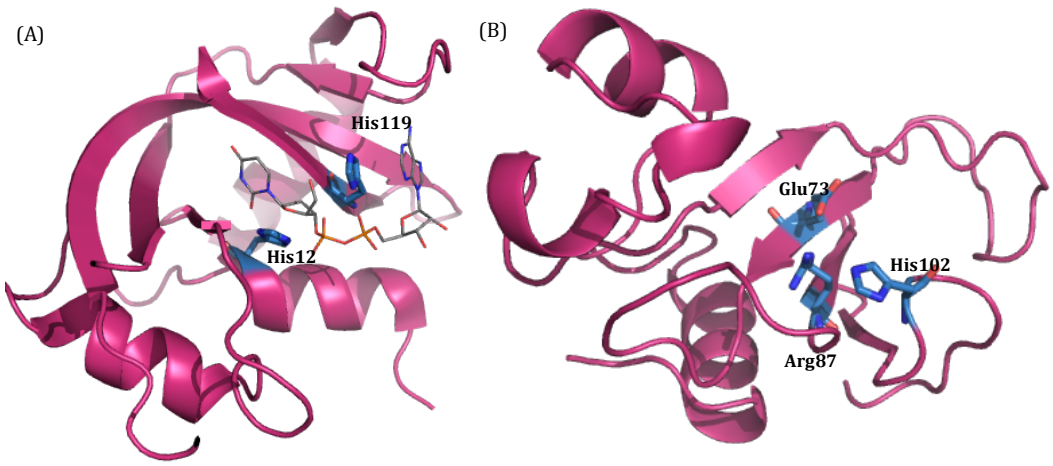


Figure 1.11. A) RNaseA and B) Barnase cartoon structures showing active site residues. Key amino acid residues are shown blue and labeled. Substrate soaked in RNase A structure is show as lines coloured by element in grey.

1.3.3.3 MazEF

The MazF toxin cleavage is also dependent on the 2'OH group of the RNA, and is Mg^{2+} independent (Y. L. Zhang, Zhang, Hara, Kato, & Inouye, 2005). It is yet to be determined which residues are involved in cleavage. There are some basic residues that are highly conserved, but no conserved histidines.

1.3.4 Homologous Metal-Dependent Enzymes

1.3.4.1 RNase H

RNase H functions to remove RNA primers from Okazaki fragments, to process R loops to modulate initiation of replication and to restore DNA topology. It targets the RNA strand in RNA/DNA hybrid substrates, with only slight specificity for the sequence. The enzyme cleaves the scissile phosphate via a water molecule to leave a 3' OH and a 5' phosphate (Nakamura, et al., 1991). In 2005, Yang and colleagues showed that the substrate requires a minimum of four ribose bases to be recognised, however, these can be flanked by DNA and the protein will remain active (Nowotny, et al., 2005). Figure 1.12 shows the RNase H active site with both metals and RNA substrate bound. In the RNA binding groove, the two 2'OH that are 5' to the scissile bond have H-bonds to the catalytic residues Glu109 and Asp132, as well as neighbouring residues Asn106 and Gln134. The two 2'OH groups 3' to the scissile bond contact Ser74 and Glu76. The DNA binding groove has a Trp which stops RNA from fitting due to steric clashing with the indole ring (Tadokoro & Kanaya, 2009).

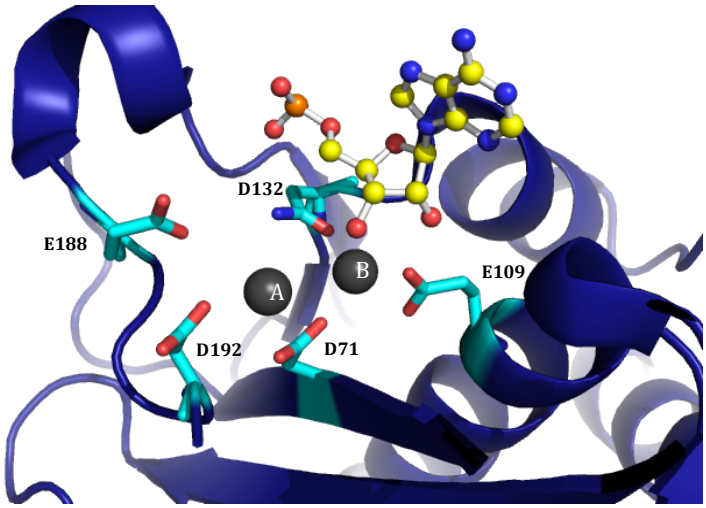


Figure 1.12. RNase H structure showing Mg^{2+} binding and RNA/DNA hybrid substrate in mutant protein D132N. RNA substrate (coloured by element in yellow) is shown as ball and stick structure. The rest of the RNA/DNA substrate is hidden for clarity. Mg^{2+} ions (grey spheres) are labeled. For simplicity, only the conserved acidic residues are shown (labeled cyan sticks).

Mutations of eight highly conserved residues to alanine found that only Asp10, Glu48 and Asp70 (*E. coli*) were essential to catalysis (Kanaya et al., 1990). Further mutagenesis and substrate titration lead to proposal of a single metal mechanism (Nakamura, et al., 1991). Yang and colleagues have

since proposed a two metal mechanism based on multiple high resolution structures of mutant proteins bound to substrate and containing two Mg^{2+} ions in the active site (Figure 1.12). In the intermediate state, metal A moves closer to metal B, and metal B obtains a more regular geometry (Nowotny & Yang, 2006).

1.3.4.2 FLAP endonucleases

FLAP endonucleases (FEN) function as both an exonuclease and structure specific endonuclease (Tock, et al., 2003). FEN-catalysed reactions are metal dependent and crystal structures show two metal ions bound in the active site (Feng et al., 2004) (Figure 1.13). However, it is proposed that the second metal ion does not play a catalytic role. Instead, it is thought to be involved in substrate recognition, required to stabilize the binding of the substrate to the enzyme (Anantharaman & Aravind, 2006; Devos, Tomanicek, Jones, Nossal, & Mueser, 2007; Tock, et al., 2003).

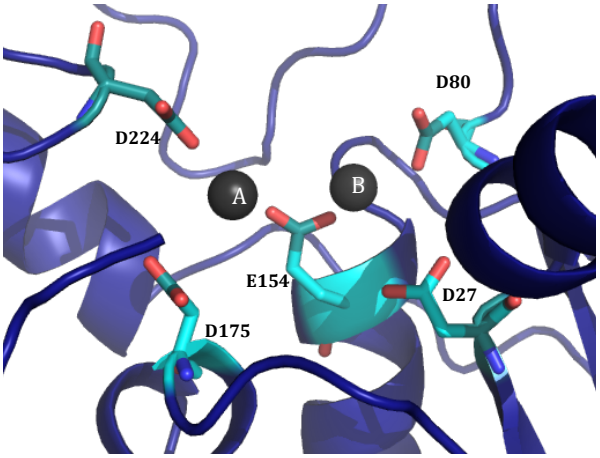


Figure 1.13. FEN structure diagram showing two Mg^{2+} ions in the active site. Metals are shown as grey spheres and conserved active site residues (cyan) are labeled.

1.3.4.3 NYN domains

Nedd4-BP1 and YacP Nucleases (NYN-domains) were identified by computational method. They share a common protein fold with PIN-domains and FEN superfamilies. The structure shows an α/β domain, with 5 core strands and 4 core helices. Alignment of NYN domains shows several key conserved acidic residues (two of the Asps are nearly invariant). The residues that are conserved between NYN, PIN and FEN domains coordinate metal B. The other residues, important for coordination of the metal ion A in the FEN enzymes, are not present in PIN or NYN domains. This implies that the principal role of the fourth conserved amino acid is not in coordinating a second metal, but is directly involved in catalysis. It was proposed that the

metal ion is involved in activating a water for nucleophilic attack of the phosphate bond (Anantharaman & Aravind, 2006).

1.3.4.4 RNase II

RNase II specifically targets ssRNA and functions in the maturation, turnover, and quality control of RNA within the cell. A structure has been solved with substrate bound into the enzyme (Figure 1.14). RNase II is selective for ssRNA through a cavity in the catalytic region which sterically hinders other substrates (Gan et al., 2006). The RNA residues surrounding the scissile bond are clamped by aromatic residues, and a threonine coordinates the 2'OH of the RNA 5' to the scissile bond. The mechanism is proposed to require only a single metal ion coordinated by Asp201 and Asp210, while Asp209 and Asp207 bind and orientate the substrate (Barbas et al., 2009; Frazao et al., 2006).

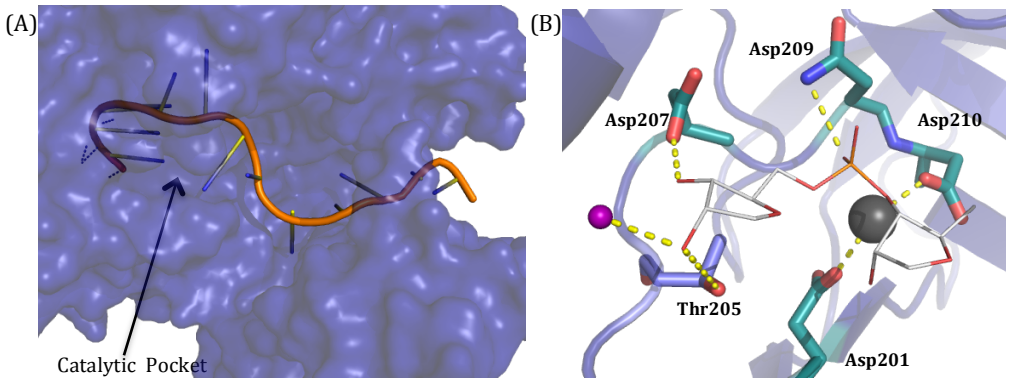


Figure 1.14. RNase II binding to ssRNA. A) The binding cavity of the protein (surface structure) envelops ssRNA (orange cartoon). The position of the active site is labeled. **B)** The conserved active site residues (cyan), and Thr residue (blue) are shown binding to the RNA substrate (grey lines) by yellow dotted lines. Bases have been removed from the substrate for clarity. The metal ion (grey) and water (purple) are shown as spheres.

1.4 Proposed Mechanism for $VapC_{PAE2754}$

Taking into account as much of the available information as possible, a mechanism for how $VapBC_{2754}$ may be cutting its ssRNA substrate was proposed. A structure of $VapBC_{PAE0151}$ with a dinucleotide substrate docked into the active site was used as a reference to determine the potential roles of the conserved acidic residues (Figure 1.15).

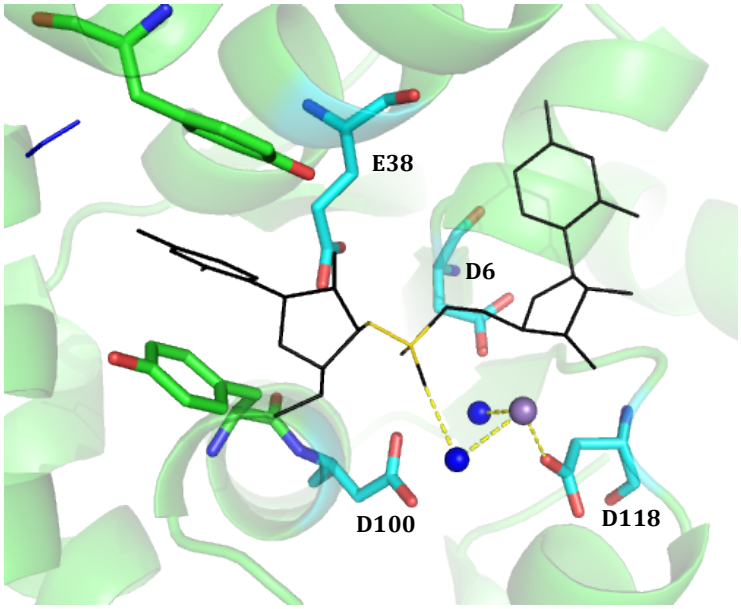


Figure 1.15. $VapC_{PAE0151}$ active site with docked substrate. Mn^{2+} (purple) and waters (blue) and the scissile bond (yellow line) are shown. The four conserved active site residues are labeled (cyan), and the two tyrosines (one from each monomer) are shown as sticks (green) due to their role in binding the substrate. Interactions with the waters are shown as yellow dotted lines.

It is known that the substrate is cleaved at the 3' phosphate bond leaving an OH group on the 5' product and the phosphate on the 3' product. This was shown using MALDI-TOF analysis of the products, which can determine the mass of the RNA accurately enough to show which product half the phosphate remains on (McKenzie, 2011). In light of this, the mechanism cannot be the same as barnase or RNase A in which a cyclic intermediate is formed and the phosphate is left on the 3' product. This initially seemed a likely method because it involves the 2'OH of the ribose, thus would explain the specificity of the protein for RNA over DNA.

As it remains unclear whether or not one or two metal ions are required for catalysis, the Mg^{2+} ions have not been shown in the proposed mechanism. However, at least one metal will be required for coordination of the waters and the active site residues for catalysis to occur.

In the proposed mechanism (Figure 1.16), Asp92 accepts a hydrogen and activates the water molecule. The nucleophile is a deprotonated water molecule that attacks the phosphate, cleaving the 3'O-P bond. E38 donates a proton to the 3' O of the 5' product, and the phosphate is retained on the 3' product. D8 and D110 are involved in orientation and coordination of the substrate and the metal ion.

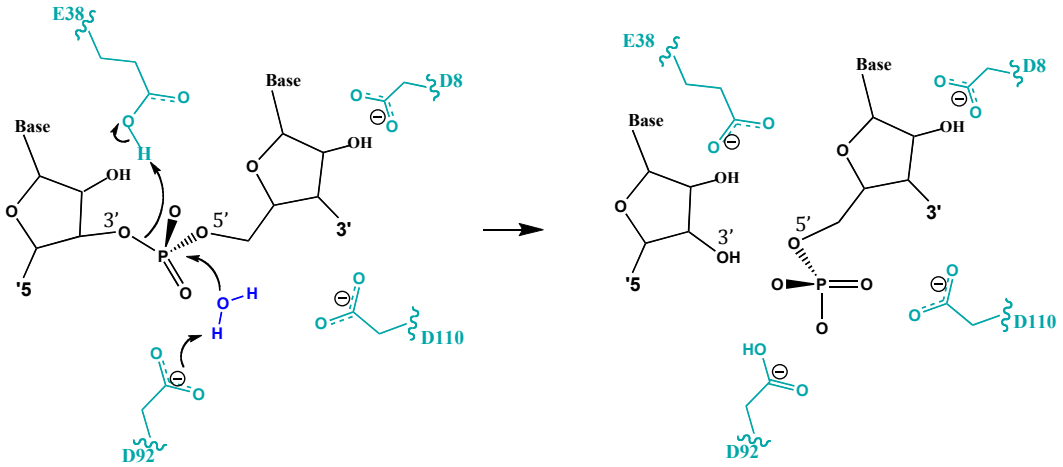


Figure 1.16. Proposed mechanism of RNA Cleavage for VapC$_{PAE2754}$. Conserved active site residues (cyan) are labeled. The nucleophilic water is shown in blue and the movement of electrons is denoted by arrows. After cleavage the phosphate is on the 3' product and the 5' product has a 3'OH.

1.5 Research Objectives

The aim of this thesis was to determine the mechanism of RNA cleavage by the PIN-domains using a member of the VapC family of toxins, $VapC_{PAE2754}$ from *Pyrobaculum aerophilum*. This mechanism will provide insightful information that may allow an inhibitor to be designed.

The main objectives to fulfill this aim are:

1) Propose a mechanism of action and make mutations to test this hypothesis.

2) Design a substrate to allow rapid and accurate testing of the activity by a fluorometric assay.

3) Solve a crystal structure that will provide more detail about the arrangement of the catalytic elements in the active site.

Chapter 2. Materials and Methods

2.1 General Methodology:

2.1.1 Agarose Gels

Gels were made up using TAE buffer (40 mM Tris-acetate, 2 mM EDTA) with either 1 % (w/v) or 2 % (w/v) agarose powder, depending on the size of the DNA fragments to be separated. The DNA sample (approximately 10 µl) was mixed with 5x DNA loading dye (approx. 2 µl) and loaded into wells, with 1kb-Plus DNA ladder (Invitrogen) in the outside well. Gels were run in electrophoresis tanks between 80 and 100 V for enough time to separate the bands of interest, usually 30-60 minutes.

The gels were stained in 0.5 µg/ml ethidium bromide (EtBr) for approximately 15 minutes and then de-stained by rinsing thoroughly in water and viewed on a UV light box. If the product was to be extracted from the gel then 1x SYBR safe stain (1:1000) was added to the agarose solution before the gels were poured and the gel was viewed on a safe blue light box so that the bands of interest could be excised.

2.1.2 Extraction of DNA from Agarose Gels

The bands were separated as described in section 2.1.1, and the desired band was cut out using a clean scalpel blade. The DNA was extracted from the agarose using a QIAquick Gel Extraction Kit (Qiagen, Netherlands). DNA was resuspended in either 30 or 50 µl of elution buffer depending on the concentration required.

2.1.3 Purification of DNA from Solution

DNA was extracted from polymerase chain reactions (PCR) using a QIAquick PCR Product Purification Kit (Qiagen, Netherlands), according to the manufacturers instructions. DNA was eluted in either 30 or 50 µl of elution buffer depending on the concentration of DNA required.

2.1.4 Transformation

2.1.4.1 Preparation of Electrocompetent Escherichia coli (E. coli) Cells

A glycerol stock of the *E. coli* strain (DH5α or BL21) was streaked out onto LB (Luria-Bertani media) agar plates and incubated at 37°C overnight. A 5 ml culture in LB broth was started using a single colony, and incubated at 37 °C overnight, shaking at 180 rpm. A 1 L LB culture was grown from the seeder until an OD_{600} between 0.5 and 0.7 was reached. The cells were chilled on ice for at least 15 minutes then spun down at 4500 rpm for 15 minutes in a sterile centrifuge bottle. The supernatant was discarded and the cell pellet was resuspended in 1 L of chilled sterile 10 % glycerol. The cells were spun down and resuspended in this manner using sequentially decreasing volumes of 10 % glycerol (500 ml, 20 ml and a final volume of 1-2 ml). 50 μl was aliquoted into sterile 1 ml microcentrifuge tubes and stored immediately at -80 °C.

2.1.4.2 Electroporation Method

Glycerol stocks of electrocompetent cells (DH5α or BL21) were defrosted slowly on ice. PCR amplified plasmid DNA was added to the electrocompetent cells. The volume added was dependent on the plasmid being transformed. 1 μl was added for wild-type plasmid, or 4-10 μl for mutant plasmid as well as 80-160 μl of 10 % glycerol to prevent arcing. The solution was transferred into a 0.2 cm electroporation cuvette (Bio-Rad Laboratories, USA). The cuvette was placed in a Bio-Rad Gene Pulser™ and electroporated with 2.5 kV at 25 μF capacitance and 200Ω resistance. 95 μl of SOC (Super Optimal broth with Catabolite repression) was added immediately and mixed well. Cells were kept at 37 °C for approximately 1 hour for cell recovery, before being plated onto LB-agar plates containing 100 μg/ml of ampicillin antibiotic.

2.1.4.3 Plating Transformed Cells

For wild-type transformations 200 μl of the SOC culture was spread evenly onto the plate. For mutants the culture was spun down for 5 min at 4600 rpm at 4 °C, the supernatant was removed, and the pellet was resuspended in remaining supernatant to be streaked out on a sterile loop. Plates were

grown overnight (approx. 16 hours) at 37 °C and either cultured immediately (for mutant transformations) or stored at 4 °C.

2.1.5 Plasmid DNA Extraction from E. coli

A single colony was selected from a plate streaked with a strain containing the desired plasmid. The colony was used to make a 5 ml culture which was grown overnight. Plasmid DNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen, Netherlands) according to the manufacturers instructions. Elution volumes were either 30 or 50 μl depending on the concentration of plasmid required.

2.1.6 Nucleic Acid Quantification

A Nanodrop ND-1000 Spectrometer (Nanodrop Technologies, USA) was used to measure the concentration of DNA and RNA samples.

2.1.7 Polymerase Chain Reaction Primers

Primers for PCR were designed using Geneious, and ordered from either Sigma Aldrich (USA) or Integrated DNA Technologies (IDT, USA). They were made up to 100 pmol stock solutions using TE buffer, and diluted further to 10 pmol to make working solutions.

2.2 Methods Used for Protein Expression and Purification

2.2.1 Expression and Purification of $VapC_{PAE2754}$ Wild-type and Mutants

A glycerol stock of BL21 cells containing the protein plasmid was streaked out onto a plate of LB agar containing ampicillin antibiotic (1:100) and grown overnight at 37 °C, shaking at 800 rpm. A 10 ml seeder culture in LB broth containing 1:100 ampicillin was created using a single colony, and was grown overnight at 37 °C. A 1L LB broth containing 1:100 ampicillin was inoculated with 5mls of this culture and grown at 37 °C. When an OD_{600} of close to 0.5 was reached 1 ml of 0.5 mM IPTG (isopropylthio-β-D-galactosidase) was

added to induce protein expression and the cells were incubated overnight at 37 °C.

The cells from the 1 L induced culture were spun down in sterile 500 μl centrifuge bottles at 4600 rpm for 25 minutes. Pellets were resuspended in 20 ml of lysis buffer (50 mM Tris, 100 mM salt, pH 9.2) and pooled together to be spun down again. The final pellet was then resuspended in approximately 20 ml of lysis buffer and an EDTA free protease inhibitor tablet (Roche Applied Science, Switzerland) was added. Cells were lysed by sonication on ice in 30 second bursts at setting 6 on a Misonix Sonicator (USA) until the cells were lysed, usually a total of 3 minutes. The lysate was spun down at 16000 rpm at 4 °C for 20 minutes to pellet out the insoluble material. The supernatant was collected and heated in a 80 °C water bath for 40 minutes to precipitate out the *E. coli* proteins but leave the thermostable $VapC_{PAE2754}$ in solution. The solution was spun down at 16000 rpm for 20 minutes at 4 °C to separate the soluble fraction from the insoluble fraction.

2.2.2 IMAC Purification of His-Tagged Proteins

Protein was expressed according to methods in 2.2.1. The supernatant from the final step was filtered through 0.2, 1.2, and 4.5 μl Minisart filters (Sartorius AG, Germany). The protein was loaded onto an IMAC (immobilized metal affinity chromatography) column, which had previously been flushed with 4 column volumes of water and 4 column volumes of lysis buffer. The column was then attached to either an ÄKTA Prime™ or ÄKTA basic™ FPLC (fast protein liquid chromatography) system. Lysis buffer was pumped though the column at a rate of 1 ml/min for approximately 20 minutes to flush out any unbound protein. A gradient of 0-50 % elution buffer (lysis buffer plus 1 M imidazole) was used to remove the bound protein at a rate of 1 ml/min over 50 mls. 2 ml fractions were collected over this period and analysed on SDS (sodium dodecyl sulphate) gels according to the method in section 2.2.8. Elution of proteins was monitored by absorbance at 280 nm.

2.2.3 Size Exclusion Chromatography

Size exclusion (SE) chromatography was used to further purify protein and to estimate whether the protein was behaving as a monomer, dimer, tetramer or an aggregate. Protein was concentrated to required volume according to the method in section 2.2.6. A HiLoad™ 16/60 or 10/300 superdex column (GE healthcare, UK) was equilibrated with buffer and protein was loaded through a 0.2 μm filter. Buffer was run through the column for 4-5 hours and 2 ml fractions were collected. Elution of the protein was monitored by absorbance at 280 nm.

Both columns had previously been calibrated, and a calibration curve had been constructed (Figure 2.1). This allowed the calculation of the molecular weight of the protein from the volume at which the protein eluted.

The elution volume was substituted into equation 2.1 to determine the K_{av} value. The K_{av} value was used to determine the molecular weight from the rearranged equation of the slope of the calibration curve (equations 2.2 and 2.3 below).

Equation 2.1 $K_{av} = (V_e - V_o) / (V_c - V_o)$

Where V_e = elution volume (peak volume-injection volume)

V_o =column void volume

V_c = geometric column volume

Rearranging the equation of the slope for the S200 10/300

Equation 2.2 $MW = e^{((Kav - 1.6833) / -0.115)}$

Rearranging the equation of the slope for the S200 16/60

Equation 2.3 $MW = e^{((Kav - 2.1739) / -0.159)}$

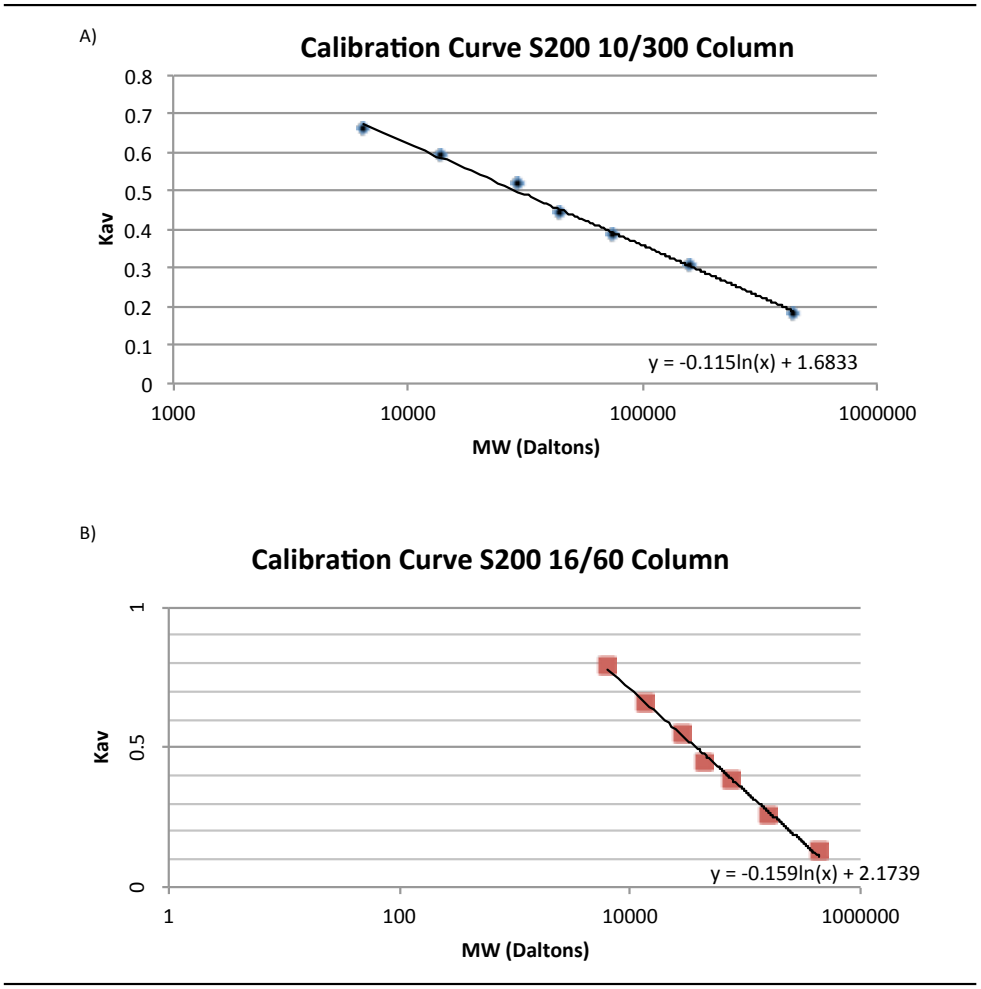


Figure 2.1. Calibration Curve for S200 size exclusion columns. A) S200 10/300 and **B)** S200 16/60 Plotted from a selection of calibration standards over a range of molecular weights using Excel from Microsoft office.

2.2.4 Glycerol Stocks

500 µl of cell culture (LB broth, 1:100 Ampicillin) was mixed well by pipetting up and down with 500 µl of 50 % glycerol in a sterilized 1.5 ml eppendorf tube and stored at -80 °C.

2.2.5 Protein Concentration Measurement

Protein concentration was determined using a nanodrop ND-1000 Spectrometer. (Nanodrop Technologies, USA)

2.2.6 Concentrating Protein

Protein was placed in 20 ml or 2 ml Vivaspin concentrators (Sartorius AG, Germany) with 5 kDa molecular weight cut offs. These were spun down at 3000 or 3600 rpm respectively at 4 °C for 15-20 minute intervals until the desired concentration or volume was reached.

2.2.7 Dialysis of Protein

For large volumes (>1 ml), protein was placed in Spectra Por® dialysis tubing (Spectrum Laboratories, USA) with 6-8 kDa weight cut off, which had been pre wet in the target buffer. The ends were sealed with dialysis clips and the tubing was placed in 1 L of the buffer with gentle stirring. The buffer was replaced after 3-4 hours and then the protein was left to dialyse overnight at room temperature. For smaller volumes (<1 ml) protein was added to 200 μl or 20 μl buttons, and pre-wet dialysis tubing was placed over top and sealed with a rubber ring. Buttons were placed in 50 ml of target buffer and shaken gently. The buffer was replaced after 2-3 hours, and either replaced again after a further 2-3 hours or left overnight to dialyse with gentle shaking.

2.2.8 SDS-PAGE Gels

Purified protein fractions were run on 15 % SDS polyacrylamide gels. The resolving layer (15 % v/v acrylamide) contained 15 ml 30 % acrylamide (Bio-Rad Laboratories, USA), 7.5 ml resolving buffer (1.5 M Tris pH 8.8), 300 μl 10 % SDS, 150 μl 10 % ammonium persulphate (APS), 15 μl tetramethylethylenediamine (TEMED), made up to 30 ml with double distilled (dd) H_2O. Hoeffer gel casts were filled with the resolving buffer to approximately 2 cm from the top, then overlaid with approximately 2 ml of isopropanol and left at room temperature to set (approx. 40 min). The stacking gel (5 % v/v acrylamide) contained 2.125 ml 30 % acrylamide, 1.6 ml stacking buffer (1.0 M Tris pH 6.8), 125 μl 10 % SDS, 63 μl APS, 6.3 μl TEMED, made up to 12.5 ml with dd H_2O. The isopropanol was fully removed from the casts and the stacking gel was poured over the resolving gel. Gel combs were inserted and the stacking layer was left to polymerise for

approximately 30 minutes or until set. Gels were stored at 4 °C in damp paper towels inside a sealed plastic bag for up to one month.

2.2.8.1 Running SDS-PAGE Gels

Approximately 15 μl of protein samples from collected fractions (from IMAC purification, method 2.2.2) was mixed with 4 x SDS loading buffer (QX4) and loaded into the wells. 10 μl of Precision Plus Protein™ Unstained Standard (Bio-Rad Laboratories, USA) was added to the outside well. Gels were run in a Bio-Rad PowerPac Basic (Bio-Rad Laboratories, USA) electrophoresis tank in constant voltage mode, at 70 V for approximately half an hour though the stacking gel and then at 150 V through the resolving gel till the dye front approached the bottom of the gel.

2.2.8.2 Coomasie blue staining

SDS-PAGE gels were stained using a succession of Fairbanks coomassie stain solutions and destained in a final destaining solution. Approximately 80 ml of the staining solution was poured over the gel and heated in a closed box on high for 30 seconds in an 1100 W microwave. The box was shaken at 160 rpm for 10-15 minutes before the stain was poured off and the next stain was poured over the gel. For gels in which the protein bands were small or harder to see the box was left shaking in the destain for a longer period of time, up to overnight to fully destain. The bands were visualised on a UV light box.

2.2.9 Whole Protein MALDI TOF Mass Spectrometry

2.2.9.1 Sample Preparation, Matrices, and Calibration

Protein samples were dialysed into milli-Q (MQ) water in 200 μl buttons after size exclusion. A matrix of 10 mg Super-DHB (9:1 mix of 2, 5-dihydroxybenzoic acid and 2-hydroxy-5-methoxybenzoic acid), plus 30 μl of 2:1 ACN and 0.1 % (v/v) TFA was sonicated for 10 minutes and spun down for 3 minutes at 13000 rpm. 0.5 μl of sample or calibration standard was mixed with 0.5 μl of the matrix and spotted onto an AnchorChip™ MALDI-TOF target plate (Bruker Daltronics, USA) and left to air-dry.

2.2.9.2 MALDI-TOF set-up

An Autoflex™ II MALDI-TOF mass spectrometer (Bruker Daltronics, USA) was set up using J. L. McKenzie's Whole Protein Method (McKenzie, 2011). Spectra for the whole protein calibration standard II (Bruker Daltronics, USA) were collected first. The spectrometer was calibrated with an automatic polynomial correction. Samples were analysed in linear mode, with the mass range selector set to 'medium range', pulsed ion extraction of 450 ns, gain to 2500 V, acceleration voltage to 20 kV and a range of 5-20 kDa data collected, (or extended out to 60 kDa if looking for dimer and tetrameric interactions). Laser power ranged from 60-90 percent.

2.2.9.3 Spectra Analysis

The spectra obtained were usually sums of two individual protein spots due to difficulties in getting such large samples to fly. These spectra were saved and exported to Flex Analysis™ software (Bruker Daltronics, USA), the baselines were removed and the peaks were identified and labeled.

2.3 Methods Used for $VapC_{PAE2754}$ Mutagenesis

2.3.1 Primer Design

Primers for mutagenesis were designed with the mutation in the middle of the strand (Figure 2.2) flanked by approximately 15 bases on either side, using Quick-change primer design software (Agilent Technologies, USA). The amino acid change was accomplished using the minimal amount of base changes possible (Table 2.1).

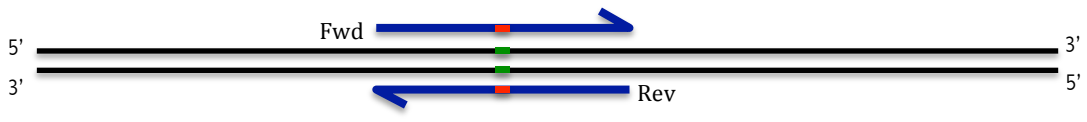


Figure 2.2. Schematic diagram of mutant primer design. The gene is shown in black, the introduced base mutation on the gene in green and the corresponding base in the primer in red.

Due to the need for the primers to cover the section of gene containing the mutation and not have any other changes to the sequence there were some restraints on the primer sequence. Primer sequences are found in appendix 7.1.1. Primers were ordered from Sigma-Aldrich (USA) or Integrated DNA Technologies (USA).

Mutation	Wild-type Amino Acid Codon	Base Change Made	Mutant Amino Acid Codon
Asp^{8}-Asn	GAC	g22a	AAC
Asp^{8}-Ala	GAC	a23c	GCC
Ser^{10}-Ala	TCC	t28g	GCC
Glu^{38}-Gln	GAG	g112c	CAG
Glu^{38}-His	GAG	g112c, g114c	CAC
Asp^{92}-Asn	GAC	g274a	AAC
Thr^{108}-Ala	ACG	a322g	GCG
Asp^{110}-Asn	GAC	g328a	AAC
Asp^{110}-Ala	GAC	a329c	GCC

Table 2.1 Mutagenesis amino acid and base changes. The base changes required to alter each amino acid to obtain the desired codon are listed for each mutation.

2.3.2 PCR Mutagenesis methods

Two different methods were used to introduce the desired mutation into the plasmid. Both use the same mutagenesis primers in a PCR amplification reaction so that the mutation is incorporated into the synthesised product.

2.3.2.1 Site-Directed Mutagenesis Method

For site-directed mutagenesis the primers were used in a PCR reaction with supercoiled 2754 plasmid. The original template plasmid was digested away using Dpn1, leaving only mutated plasmid as the product (Figure 2.3).

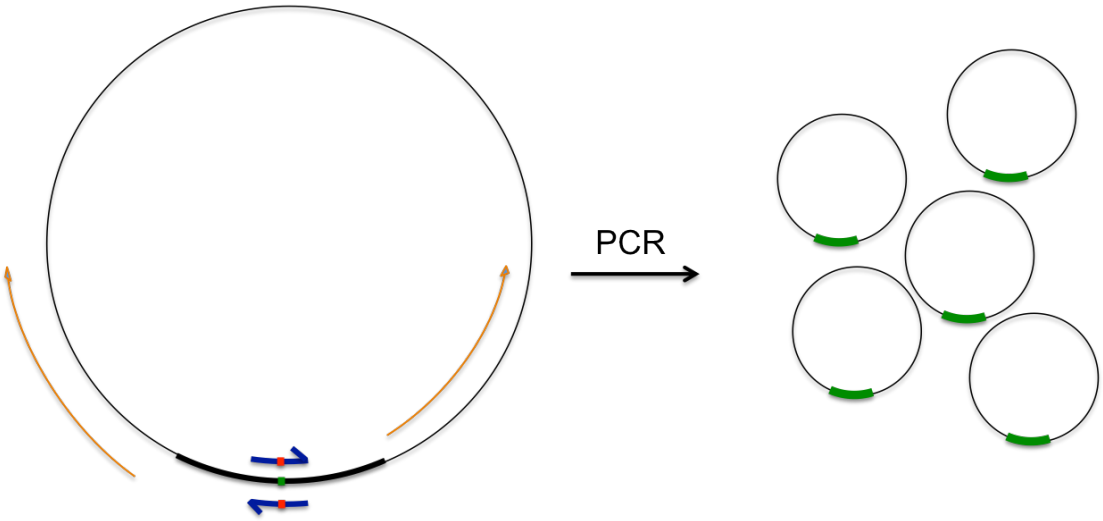


Figure 2.3. Schematic diagram of the site-directed mutagenesis method. The primers (blue) are shown binding to the template plasmid containing the gene to be mutated (bold). The mutated bases are shown in red, corresponding to the position of the intended mutation (green square). The direction of synthesis is shown in orange. The products of the PCR reaction are shown as plasmids containing the mutated gene (green).

Wild-type plasmid was purified according to method 2.1.5 and run on an agarose gel to ensure that the DNA was >80 % supercoiled. The PCR reaction contained 25 ng plasmid DNA, 125 ng of each primer, 0.2 mM dNTPs, 1 µl Pfu Ultra (a high fidelity DNA polymerase), 5 µl 10x Pfu Ultra reaction buffer and was made up to 50 µl with dH_2O.

Cycling conditions:	95 °C	0:30	
	95 °C	0:30	←
	55-60 °C*	1:00	x 18
	68 °C	5:00	
	68 °C	7:00	

* Initial PCRs were run as temperature gradients to determine optimum temperature for each set of mutagenesis primers, after that the annealing step was carried out at 59 °C.

The reactions were cooled on ice, and 0.5 µl (10 U/µl) Dpn1 was added, mixed and incubated for approximately 1 hour. To check the plasmid had been amplified 10 µl of reaction was run on a 1 % agarose gel. Once

amplification was confirmed 4 μl of the mutagenesis reaction was transformed into DH5α cells, according to method 2.1.4. If transformation was successful 4 individual colonies were selected and the plasmids were extracted as in method 2.1.5. Glycerol stocks were also made from these cultures (Section 2.2.4). The purified plasmids were sent for sequencing to confirm the presence of the intended mutation.

2.3.2.2 Two-Halves Mutagenesis Method

Description

Due to the constraints on the primers in the site-directed mutagenesis method, it was difficult to optimize the PCR conditions, and for this and other unknown reasons certain mutations could not be introduced into the plasmid using the site-directed method.

The two-halves method creates each half of the gene individually and once annealed this mutated gene is cut at restriction sites and inserted back into the plasmid (Figure 2.4). Two-halves mutagenesis involves using the mutagenesis primers but also primers that bind to either end of the 2754 gene. By adding the forward mutant primer with the reverse gene primer, and conversely the reverse mutant primer with the forward gene primer then each half of the mutant gene can be synthesized and two halves then joined together in a further PCR reaction. Restriction digests were carried out for the mutant gene and the plasmid, and the gene was ligated into the plasmid, which is then transformed into an electrocompetent growth strain of *E. coli*.

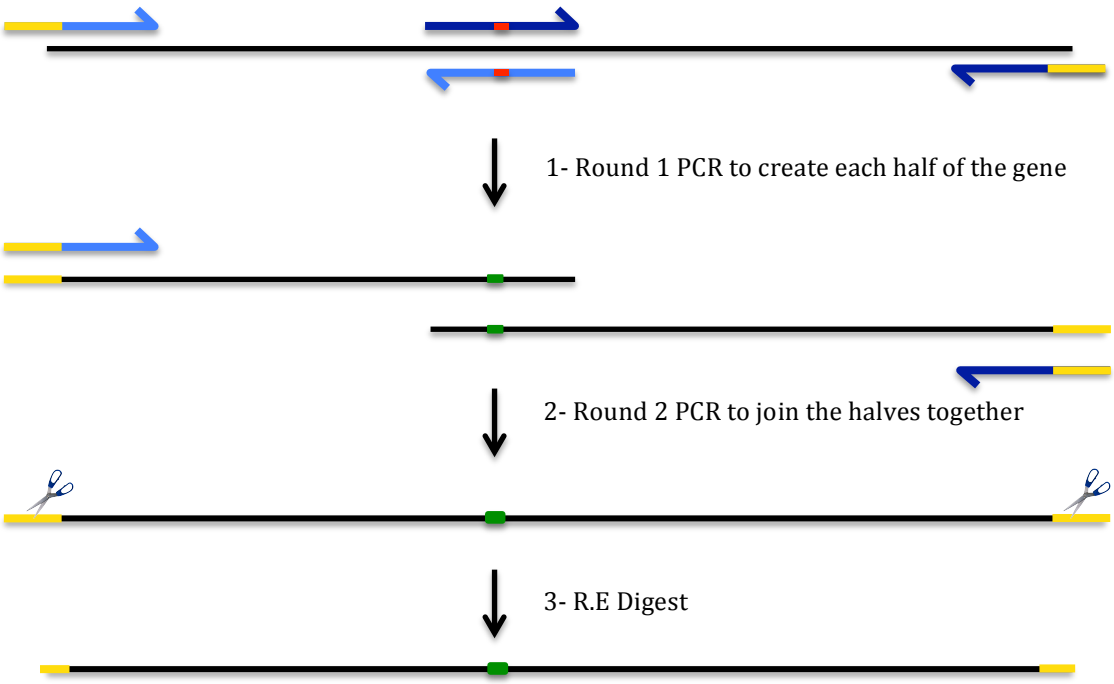


Figure 2.4. Schematic diagram of two-halves mutagenesis. Mutation in gene (black line) is shown in green. Mutagenesis primers contain the corresponding mutation in red. Primers to amplify the gene contain restriction sites shown in yellow. **1)** Fwd and Rev primers to amplify the first half of the gene are shown in light blue, and Fwd and Rev primers to amplify the second half of the gene are shown in dark blue. **2)** The products from the first round of PCR are used as templates in a second PCR reaction with the gene primers. **3)** The full gene product from the second round of PCR contains the intended mutation. Restriction Enzyme digest (R.E) creates the final product to be ligated into the plasmid.

Method

PCR reactions contained 2.5 µl 10x Amp buffer, 0.5 µl 50 mM MgS04, 0.75 ul 10 mM dNTPs, 2.5 µ 10 pmol/µl primer (first half; $VapC_{PAE2754}$ forward, and Mutagenesis reverse, Second half; $VapC_{PAE2754}$ reverse with Mutagenesis forward), 0.4 µl Pfx enzyme, 0.5 µl $VapC_{PAE2754}$ plasmid, and made up to 25 µl with ddH_20. Negatives had 0.5 µl ddH_20 added instead of plasmid DNA.

Cycling conditions:	95 °C	2:00	
	95 °C	0:30	
	55-65 °C*	0:30	x 30
	68 °C	0:30	
	68 °C	7:00	

* Initial PCRs were run as temperature gradients to determine optimum temperature for each set of mutagenesis primers, after that the annealing step was carried out at 59 °C.

10 μl of reactions were run on a 1 % agarose gel to determine the optimal temperature. If most temperatures resulted in bright bands then product was isolated by gel purification. Reactions were pooled, run on a deep 1 % agarose gel according to method. If only a few temperatures resulted in bands then the PCR reaction was repeated at the optimal temperature to obtain more product before gel extraction was carried out.

To add the two halves together PCR reactions were set up using equimolar concentrations of the two halves of the gene as the templates with the $VapC_{PAE2754}$ forward and reverse primers following the PCR protocol outlined above. These were run on a thick 1 % agarose gel and gel purified as in method 2.1.2 to obtain the total gene.

2.3.3 Restriction Enzyme Digestion

Enzymes for restriction (NcoI and HindIII) were purchased from Invitrogen (USA) and Roche Applied science (Switzerland). Digestion of the DNA was carried out according to the manufacturers instructions, and incubated for approximately 4 hours at 37 °C. Due to incompatibility of the buffers, reactions were carried out sequentially, the NcoI restriction digest first, and the HindIII digest after. After each restriction reaction the DNA was purified form solution according to the method in section 2.1.3

2.3.4 DNA ligation

Ligations were performed according to the manufacturers instructions using 5 U of T4 DNA ligase (Invitrogen, USA). Reactions were incubated at room temperature for 6 hours or at 16 °C overnight.

2.4 Methods used for RNase Assays and their Analysis

2.4.1 Pentaprobe RNA

2.4.1.1 Plasmid Preparation and Restriction Enzyme Digest of Pentaprobe Plasmids

Plasmids were prepared by J. L. McKenzie according to the methods in her PhD thesis (McKenzie, 2011). The plasmids were obtained from Joel McKay, school of molecular Bioscience, University of Sydney.

2.4.1.2 PCR Amplification of Pentaprobe Inserts

PCR reactions were made up as follows for each pentaprobe RNA; 2.5 μl 10x AMP buffer, 0.5 μl 50 mM $MgSO_4$, 0.75 μl 10 mM dNTPs, 2.5 μl forward pcDNA3 primer, 2.5 μl reverse pcDNA3 primer, 0.5 μl pentaprobe plasmid, 0.4 μl Pfx enzyme, made up to 25 μl with water.

Cycling conditions:	94 °C	2:00	
	94 °C	0:15	←
	64.8 °C	0:30	x 30
	68 °C	0:45	
	68 °C	7:00	

Products were extracted from an agarose gel as in section 2.1.1. If sufficient DNA has been purified (>800 ng) then the product was used in a transcription reaction to convert to RNA. If not then the product was used as a DNA template for further PCR amplification according to the same protocol. The re-amplified product was purified directly from the PCR reaction, after 2 μl had been run on an agarose gel to ensure that there was a single band.

2.4.1.3 In Vitro Transcription of Pentaprobe Inserts

Transcription of the PCR products to RNA was carried out using the T7 MEGAscript® kit (Ambion, USA). The reaction consisted of 8 μl rNTPs, 1 ng of the pentaprobe DNA product, 2 μl of T7 polymerase enzyme mix, 2 μl of 10 x enzyme buffer and made up with DEPC (diethyl pyruvate carbonate) water to 20 μl, and incubated overnight at room temperature. The reaction was

treated with 1 μl of TURBO DNase and incubated at 37°C, shaking at 800 rpm, for 15 minutes.

The RNA was precipitated using the sodium acetate and ethanol method. 2 μl of 3 M sodium acetate pH 5.2 and 60 μl of 100% ethanol was added and the reaction was left to precipitate on ice for approximately 30 minutes. The reaction was spun down at 13000 rpm at 4 °C for 20 minutes, and the supernatant was removed. The pellet was washed with 60 μl 70 % ethanol, spun down again and the supernatant removed. Pellet was left to air-dry before being resuspended in 50 μl of nuclease free water.

2.4.1.4 Design of 923 and 924 Pentaprobe Oligonucleotides

The 923 and 924 pentaprobe sequences were broken down into smaller overlapping RNA oligonucleotides of 28-36 bp long. They were designed using RNAfold (Vienna RNA webservers, TBI) to mimic the secondary structure of the parent oligo.

2.4.2 General Assay Method

Protein assays were carried out with different RNA substrates over a 60 minute time course to determine where the protein was cleaving the substrate and how rapidly the cleavage was occurring.

2.4.2.1 Assay Reactions

Precautionary measures were taken to reduce risk of RNase contamination. Individual reactions for each time point were set up as follows: 6 μl of assay buffer (20 mM Tris, 20 mM NaCl, 10 mM $MgCl_2$), 2 μl water (DEPC or Nuclease free), 1 μg of RNA, 1 μg of protein. Reactions were incubated at 37 °C for the appropriate time course and stopped with 8 μl formamide loading dye.

2.4.2.2 Negative Controls

RNA only 0 hour, and 1 hour controls were used to ensure the reactants were not RNase contaminated. Assay reaction was set up as above but no protein was added, instead 1 extra μl of nuclease free water was included. Zero hour controls were stopped immediately with formamide loading dye while 1 hour

controls were incubated at 37 °C for the assay duration before the dye was added.

EDTA controls were used to ensure that the protein sample was not RNase contaminated. The reaction was set out as above but the assay buffer was replaced with EDTA buffer (20 mM Tris, 20 mM NaCl, 10 mM $MgCl_2$, 20 mM EDTA). The reaction was incubated for one hour before stopping with formamide loading dye.

2.4.3 Urea Denaturing Gels

A solution of 4.4 ml 10 x TBE buffer, 12.6 g urea and 16.6 ml H2O and 10 ml of 30 % acrylamide was made up (giving a 10 % acrylamide gel), heating to 55 °C to dissolve. For 20 % gels 6.6 mls of water and 20 ml of acrylamide were added. 260 µl 10 % APS and 60 µl TEMED were added immediately prior to pouring into Hoeffer 5 gel casts.

2.4.3.1 Urea Denaturing Polyacrylamide Gel Electrophoresis

Gels were pre-run at 70 V for approximately 30 minutes and the wells were rinsed with running buffer (1 x TBE) to remove excess urea. 10 µl of sample, containing formamide loading dye was loaded into wells and 7 µl RNA ladder was loaded into the outside well. Gels were run at 150 V for approximately 1 hour or until bands visible from the dye were at the bottom of the gel.

2.4.4 MALDI-TOF Mass Spectrometry for RNA Oligonucleotides

Oligonucleotides for mass spectrometry were ordered HPLC (high performance liquid chromatography) purified from either Sigma Aldrich (USA) or Integrated DNA Technologies (USA). Care was taken to prevent RNase contamination.

2.4.4.1 Ammonium Acetate and Ethanol Precipitation

To clean up the RNA from the assays the reactions were stopped with 8 µl of 5 M ammonium acetate and 54 µl of 100 % ethanol instead of the formamide loading dye. and left to precipitate for approximately one hour on ice. Reactions were spun down at 13000 rpm, at 4 °C for 20 minutes and the

supernatant was removed. The pellet was washed in 60 μl of 70 % ethanol, re-spun, and the supernatant removed. Pellet was left to air dry before being resuspended in 5 μl of DEPC or nuclease free water.

2.4.4.2 Sample Preparation

The matrix solution consisted of 5 mg of 3-hydroxypicolinic acid, 10 μl of 2.5 M diammonium citrate, 125 μl of acetonitrile and 365 μl of DEPC or nuclease free water, and was vortexed well until fully dissolved, then centrifuged for 5 minutes at 13000 rpm.

0.6 μl of the matrix was spotted onto an Anchorchip™ (Bruker Daltronics, USA) and left to air dry. 0.6 μl of sample, or 0.6 μl of oligonucleotide calibration standard (Bruker Daltronics, USA) were spotted on top of the matrix and left to air dry.

2.4.4.3 MALDI TOF set-up

Samples were analysed by an Autoflex™ II MALDI-TOF mass spectrometer (Bruker Daltronics, USA) in linear mode, with mass range selector set to 'low range', pulsed ion extraction of 150 ns, gain to 2500 V, acceleration voltage to 20 kV and a range of 1-12 kDa collected. Laser power ranged between 60 and 90 %. (This method was determined by J. L. Mckenzie to obtain the best spectra for small RNA oligonucleotides). The spectrometer was calibrated using the spectra collected for the oligonucleotide calibration standard (Bruker Daltronics, USA) with an automatic polynomial correction.

2.4.4.4 Analysis of Mass Spectra

Spectra were analysed in the DataAnalysis program (Bruker Daltronics, USA). The baseline was subtracted and the list of m/z masses transferred into an excel file to be uploaded into the mass production program developed by Vic Arcus. This program determines all possible combinations of one or two cuts for the resulting peaks. For the results obtained before this program was developed the MONGO mass calculator was used to determine predicted masses of oligo fragments.

2.5 Methods Used For Fluorogenic Assays

A fast and accurate method was required for test the kinetic properties of the mutant enzymes so that the activity could be compared to the wild-type. The assay measures the increase in the fluorescence of a reaction as the quench is separated from the fluorophore by the cleavage action of the protein. This was designed based on an assay used to assess MazF activity, as well as A. D. Easters masters thesis (Easter, 2010; N. R. Wang & Hergenrother, 2007)

2.5.1 Substrate Design

A variety of substrates were designed and trialed to determine which sequence would provide rapid cleavage at a single cut-site so that the fluorogenic assay could be carried out rapidly and accurately. The basic design was a short (~16 bp) single stranded RNA or RNA/DNA hybrid molecule that contained an optimal cut-site flanked by sequence that would not be cleaved. Once designed and ordered from IDT (USA), the substrates were assayed with wild-type protein and analysed by mass spectrometry as in 2.4.4.

The substrate for the kinetic assay was designed by testing a range of oligonucleotides containing the RNA cut-site GGGG or GGUG. These four bases had previously been determined to be the consensus cut-sites using pentaprobe oligo and MALDI-MS analysis (J. L. McKenzie) and should therefore be cut rapidly by active enzymes. The substrates were tested using assays with $VapBC_{PAE2754,}$ and analysed by mass spectrometry. The results were used to optimise the oligo, by removing secondary cut-sites and selecting for rapid cleavage to ensure a fast and accurate fluorogenic assay. The process for the design of the substrate is described below and the sequences are listed in the appendix (Table A. 2).

The initial substrates designed (Oligos 1a and 1b) were two hybrid DNA/RNA oligonucleotides that contained the RNA cut-sites GGGG and GGUG in the middle of a sequence of DNA bases was designed based upon the chimeric substrate utilized by A. D. Easter.

The next substrates were designed with the optimal RNA sequences flanked by a triplet of RNA adenosines (Oligos 2a and 2b). The protein targets GC rich sites and therefore was unlikely to cleave in these flanking regions.

Substrates were then designed as RNA only versions of Oligos 1a and 1b (Oligos 3a and 3b), as it became apparent that the flanking sequence was important to the ability of the enzyme to cleave the target sequence.

The final substrate was designed using randomised A, C and U bases to replace the flanking RNA sequence (Oligos A, B and C). Secondary structure was also considered, and the flanks were designed so as not to fold and bind to one another. Because the previous results had shown that GGUG was cleaved faster than GGGG cut-site the new substrates were designed to contain only this cut-site.

The substrate selected (Oligo C) showed the most rapid degradation of the initial RNA peak out of the three RNA only substrates containing GGUG as the optimal cut-site and showed no secondary cleavage over the time course.

2.5.2 Fluorogenic Substrate

Substrate C (CUAAUCGGUGACUACA), labeled with 6-FAM™ (6-carboxyfluorscein fluorophore) at the 5' end and IBkFQ® (Iowa black flurophore quencher) at the 3' end was ordered from IDT (USA) at a synthesis scale of 1 μm. The reaction products (each half of the substrate either side of the GGUG cut site) were also ordered to make a standard curve for fluorescence. Both the first half (6-FAM.CUAAUCGG) and the second half (UGACUACA.IBkFQ) were ordered at 250 nm synthesis scale.

2.5.2.1 Stocks and Preparation

Substrate and products were shipped as lyophilized powders and made up using nuclease free water, with precautions taken to avoid any RNase contamination. The substrate was made up to 1 mM and diluted to 100 mM for the assay preparation. Both halves were made up to 100 mM.

2.5.2.2 Fluorometer Set-up

Fluorometer settings were determined by A. D. Easter (Easter, 2010). A Hitachi F-700 fluorescence spectrometer (Hitachi High Technologies Corp., Japan) was fitted with a Hitachi Thermostatted cell holder accessory. The cell was maintained at 37 °C by water fed from a Julabo 13 water bath and MP pump (Julabo, Germany) at 5.8 mL/minute. Time scans used a 0.5 data pitch, with PMT set to 450 V, excitation (ex.) 485 nm, emission (em.) 518 nm, 5 nm slits and 2 millisecond response time. Scans for assays and blanks were recorded for 100 seconds, once a 15 minute run had confirmed significant change would be seen over this time.

2.5.2.3 Standard Calibration Curve

A stock solution of 6 μm fluorophore and quench was made up, using nuclease free water, and aliquoted into a dilution series (3, 1.5, 1.25, 1, 0.5, 0.25, 0.1 μM) with 180 μl assay buffer, made up to a final volume of 300 μl with nuclease free water. Each sample was measured for 100 seconds and the average fluorescence was plotted against substrate concentration in Prism 5.02 (GraphPad Software, USA). The slope of the line of best fit was used to calculate velocity from fluorescence for the wild-type data.

2.5.2.4 Assay Method

Protein (wild-type and mutants) was prepared at 0.2 mg/ml in assay buffer and substrate was prepared at 100 mM stock in nuclease free water, to be diluted into final concentrations at 300 μl. Reaction mixture (180 μl assay buffer, x μl substrate & x μl nuclease free water – see table 2.2) were made up one at a time to reduce contamination, equilibrated to 37 °C in the fluorometer with a H1 thermocouple probe (Hanna Instuments, USA) in the cuvette solution. Once steady temperature was reached (36.2-36.1 °C) 2 μl of 0.2 mg/ml protein stock was added, mixed rapidly (total dead-time = 8-12 sec.) and the data collection started. Wild-type assays were performed over concentrations from 0.25-10 μM in order to plot data to determine K_M. Unfortunately the lack of substrate meant this could not be performed in duplicate. The activity for each run was plotted against substrate concentration in Prism 5.02 (GraphPad Software, USA) and fitted with a non-

linear least-squares analysis to the Michaelis-Menten equation. The velocity was calculated from the activity using the slope of the calibration curve (see section 2.5.2.4).

Equation 2.4 $\Delta F.min$ = activity/slope

= activity/783

(where ΔF is the change in the fluorescence)

Mutants were assayed at 5x wild-type K_M substrate concentrations in triplicate. Mutants were compared as a percentage of the wild-type. The activity of the wild-type at 5x K_m was calculated from the Michaelis-Menten equation of the graph.

Final Concentration (μM)	**100mM Substrate (μl)**	**Nuclease Free Water (μl)**
0.25	0.75	117.25
0.5	1.5	116.5
0.75	2.25	115.75
1.0	3	115
1.5	4.5	113.5
2.0	6	112
2.5	7.5	110.5
5.0	15	103
7.5	22.5	95.5
5x Km (5.87)	17.6	100.4

Table 2.2. Concentrations of flurophore made up for Michaelis-Menten data collection for VapC$_{PAE2754}$.

2.5.2.5 K_m, V_{max} and K_{cat} Determination

The kinetic data for VapC$_{PAE2754}$ was calculated using the Michaelis-Menten plot for previously determined as described above.

K_M calculated by Prism 5.02 (GraphPad Software, USA)

V_{max} calculated by Prism 5.02 (GraphPad Software, USA)

k_{cat} calculated from equation 2.5

Equation 2.5 k_{cat} = V_{max} (μM.L^{-1})/ [E] (μM.L^{-1})

2.6 Methods Used For Protein Crystallisation

2.6.1 Initial Trials

The crystallization trials were prepared by a liquid handling robotics system in an 18 °C controlled humidity room. 100 μl of PEGRx HT tm screens HR2-086, HR2-130, HR2-134 and HR2-136 were transferred into 4 x 96 well intelli-plates each (Hampton research, USA). 5 μl of ~8 mg/ml protein ($VapC_{PAE2754}$, E38Q, D92N, and S10A, E38H, D8N) was transferred into 8 wells and loaded onto a Mosquito® crystallisation robot (TTP LabTech Ltd, UK). The robot mixed protein with mother liquor and set the solution up as a sitting drop.

2.6.2 Fine Screening and Optimization of Conditions

The conditions that resulted in the best crystals for each protein from the robot screens were selected and optimized by fine variations of the conditions in 24-well VDX hanging drop plates. The tops of the wells were lined in grease and 500 μl of the mother liquor was added to each well. 2 μl of protein was mixed with 2 μl of mother liquor as a drop on a siliconised glass cover slip. The cover slip was placed drop side down on top off the greased will and pressed lightly to ensure the well would remain airtight.

In some cases further optimization was carried out, including sitting drops in micro-bridges, 1 μl volumes, decreased concentrations and streak seeding however no diffracting crystals were obtained using these methods.

2.6.3 Preparation of Crystals for Data Collection

2.6.3.1 Testing Cryo-protectants:

Cryo-protectants were made up by adding 10 % glycerol to the mother liquor conditions. However, testing at the home source showed that the conditions in which the optimized crystals formed contained a sufficiently high concentration of PEG400 (>24 %) that a cryo-protectant was not required to obtain diffraction.

2.6.3.2 Mounting and Freezing Crystals

The crystals were removed from the drop using a copper cryo-loop in a magnetic base (Hampton research, USA) and flash-frozen by rapid immersion in liquid nitrogen. The cassette was stored in a liquid nitrogen dry shipper for transport to the synchrotron.

2.6.4 Diffraction Data Collection

Crystals were taken to the School of Biological Sciences, The University of Auckland to be tested on the home-source before being sent to the synchrotron. Data was collected by Ass. Prof. V. Arcus (University of Waikato, School of Biological Sciences) using CuKα radiation (λ=1.5418Å). X-rays were generated by a Micromax™-007HF generator (Rigaku, Japan) operated at 50 kV and 100 mA with a rotating copper anode and Osmic VariMax-HF optics. The data were collected using the rotation method. The cryo-loop was mounted in the goniometer using a MAR345dtb goniometer setup with an *easymount*™ automatic sample mounting system (MAR Research GmbH, Germany). The crystals were kept at 110 K-113 K with a stream of cold liquid nitrogen gas produced by a Cobra cryosystem (Oxford Cryosystems, UK). Diffraction images were collected on a MAR345 image plate detector (MAR Research GmbH, Germany).

Final data was collected from $VapC_{PAE2754}$ wild-type protein crystals on beamline MX1, equipped with an ADSC Quantum 210 r detector, at the Australian synchrotron.

2.6.5 Data Processing and Structure Solution Methods

2.6.5.1 Integration of Data

The program MOSFLM was used to integrate data sets. The automatic spot finder was used to find spots. The diffraction patterns were indexed using the autoindex function and the mosaicity of the crystal was estimated. The parameters were refined using two sets of four consecutive images 90° apart. Integration was carried out by MOSFLM in one batch.

2.6.5.2 Scaling and Merging of Reflections

The SCALA program from the CCP4 suite of programs was used to scale and merge the integrated reflections. 5 % of reflections were marked as the R_{free} set to ensure validation during structural refinement.

2.6.5.3 Matthews Coefficient

The CCP4 program MATTHEWS_COEF was used to determine the Matthews Coefficient of the unit cell.

2.6.5.4 Structure Solutions using Molecular Replacement

The CCP4 program PHASER determined the structural solutions using molecular replacement methods.

2.6.5.5 Model Building and Refinement

Automatic model building was done by PHENIX AutoBuild wizard, which uses RESOLVE for iterative model building and phenix.refine for refinement of the model.

Manual building of the model was carried out using Coot. The model was built into SigmaA-weighted $2|F_o|-|F_c|$ electron density maps contoured at close to 1.0 σ, and helped by SigmaA-weighted $|F_o|-|F_c|$ contoured to around 3.0 σ.

Refinement was performed regularly throughout manual building using the CCP4 program Refmac5.

2.6.5.6 Structure Analysis

PYMOL was used to visualise the model and to create the structure diagrams used in this thesis.

Chapter 3. Experimental Results

3.1 Introduction

The aim of this research was to determine a mechanism of RNA catalysis for PIN-domain proteins. The hypothesized mechanism was tested by mutating the key amino acids proposed to be involved in the cleavage and specificity of the VapC2754 enzyme from *P. aerophilum* ($VapC_{PAE2754}$). The results have been separated into molecular biology, analysis of activity and crystallisation.

Molecular biology techniques were used to construct the mutant proteins. The purpose for the design of each mutation and the expected activity is explained and the process for expression and purification of both the wild-type $VapC_{PAE2754}$ and the mutants is also described.

Experiments were then carried out to analyse the activity of the mutants for comparison with the wild-type. Kinetic analysis was carried out using a fast and accurate fluorogenic assay and additional RNase assays were carried out and analysed by gel electrophoresis and mass spectrometry to supplement the kinetic data and provide additional information about the rate of cleavage and the specificity of the mutant proteins.

Finally, the crystallisation of the wild-type $VapC_{PAE2754}$ protein is described, including the data collection and refinement. A description of the final structure is given and comparisons are made to the $VapC_{PAE0151}$ structure.

3.2 Molecular Biology

3.2.1 Design of the Mutants

Single mutations were devised (Figure 3.1) so that the activity of the resulting mutant protein would provide information about the role of the targeted amino acid in the cleavage mechanism. The purpose of each of the 9 constructed mutations is described below and summarised in Table 3.1

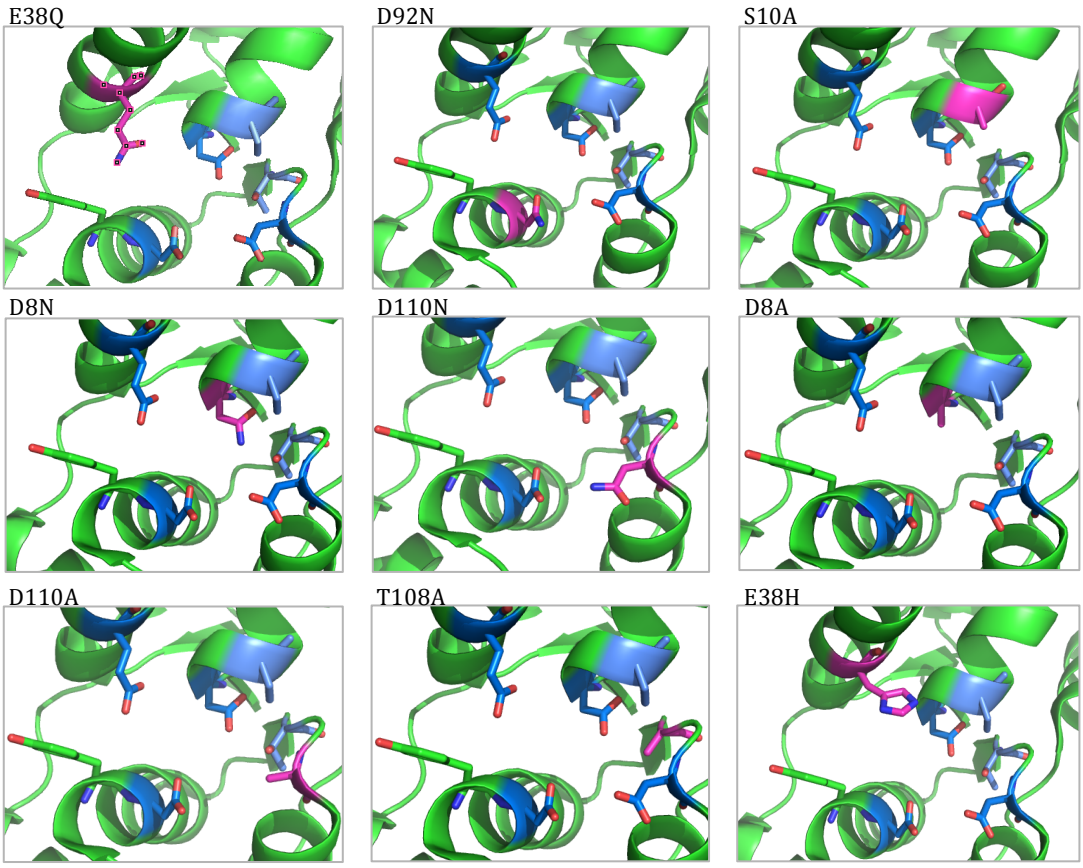


Figure 3.1. Cartoon structure diagrams of designed mutants. Mutations (pink) were made to the residues using PYMOL. The acidic residues (blue) and other conserved residues (pale blue) that have had mutations designed are shown as sticks, as well as Tyr99 (green) for orientation.

Mutants Asp^{92}-Asn (D92N) and Glu^{38}-Gln (E38Q) were designed because in the proposed mechanism (Figure 1.16) these residues act as the catalytic acid and base respectively. Glu-Gln and Asp-Asn mutations are isosteric, eliminating the carboxyl group with predicted minimal conformational disruption. Several past papers have designed similar mutations to determine if residues are acting as the acid or base, or whether a metal is

fulfilling this role (Barbas, et al., 2009; Bhagwat, et al., 1997; Devos, et al., 2007; Kanaya, et al., 1990; Yang, et al., 2006).

The S10 residue is partially conserved. Ser^{10}-Ala (S10A) mutation was designed because in the PAE1051 model S10A is positioned in a way that suggests it may interact with Asp8, a conserved active site residue.

The next round of mutants focused on the two other conserved active site residues, D8 and D110. In the proposed model these residues are involved in coordination of the metal ion and/or the substrate. Asp8-Asn (D8N) and Asp110-Asn (D110N) were designed on the basis that these residues were not acting as the catalytic acid or base and therefore isosteric mutations should not knock-out the activity completely. In contrast Asp8-Ala and Asp110-Ala were designed specifically as knockouts. In such highly conserved acidic residues a mutation to Ala should completely inhibit the activity of the protein.

The final mutants were designed based on results obtained from the first round of mutations. Thr108-Ala was created based on the partial activity shown by S10A, as T108A is also well positioned to coordinate residue D8. Glu38-His was designed to test more substantially whether or not this residue plays a role as the general acid or base. In related RNase enzymes there is often a conserved His in the active site which fulfills the role of the catalytic acid or base (Mossakowska, et al., 1989; Oda, Yoshida, & Kanaya, 1993).

	Residue description and role in proposed mechanism	Expected result
D92N	Highly conserved active-site residue, functions as acid activating the H_20 molecule by removal of a hydrogen	Highly reduced activity
E38Q	Highly conserved active-site residue, the catalytic base that donates a proton to the 5' oxygen.	Highly reduced activity
S10A	Conserved residue that coordinates acidic active site residue D8	Partially reduced activity
D8A	Highly conserved active-site residue, may be involved in coordination of the substrate, waters or metal	Knock-out
D8N	As above	Reduced activity
D110A	Least conserved of the four active site residues, involved in coordination of water, metal or substrate	Knock-out
D110N	As above	Reduced activity
T108A	Conserved residue that coordinates acidic active site residue D8	Reduced activity
E38H	Conserved active-site residue, the catalytic base that donates a proton to the 5' oxygen	Near wild-type activity

Table 3.1. Summary of the nine mutations designed. The role of each of the mutated residues in the proposed mechanism is outlined and the expected change in activity of the mutant is specified.

3.2.2 Synthesis of Mutant Plasmids

For each mutant PCR reactions were carried out using primers that contained the required base change (see section 2.3.1). Site-directed mutagenesis was attempted first. If this was unsuccessful after a repeat experiment the two-halves method was then used (see sections 2.3.2.1 and 2.3.2.2 respectively). The sequencing results for successful incorporation of each of the mutations are found in appendix C.2.

Mutated plasmid for D92N was obtained from the initial site-directed mutagenesis temperature gradient PCR (Figure 3.2)

For E38Q, no product was obtained using site-directed mutagenesis (Figure 3.2), therefore the two-halves method was attempted. Although the first half of the gene was successfully synthesised, the second half was not successfully amplified. A new forward primer was ordered to increase binding. Product was obtained by site-directed mutagenesis using the new primer with the original reverse primer (results not shown).

Product for S10A was obtained from site-directed mutagenesis PCR (Figure 3.2), however sequencing of the plasmid showed that the mutation had not

been incorporated. After extensive optimization of the PCR conditions the products were obtained from the two-halve mutagenesis method (results not shown).

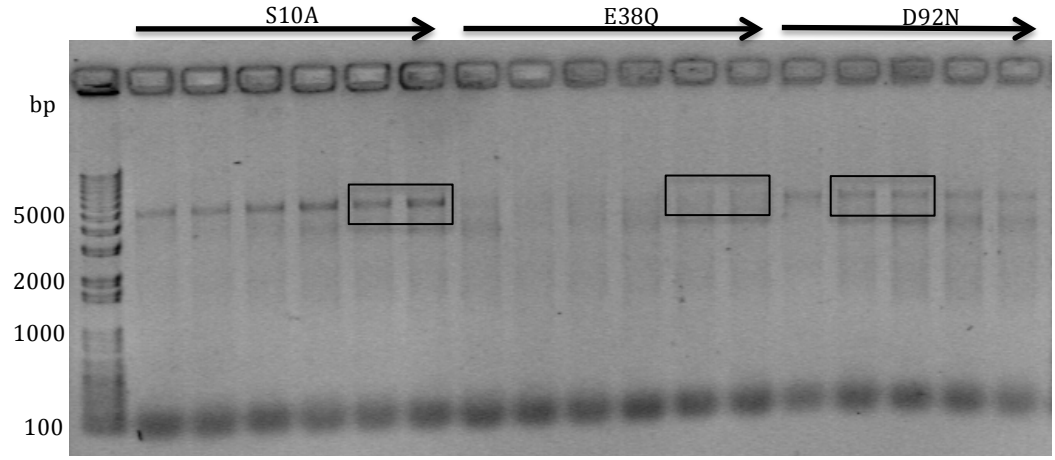


Figure 3.2. Agarose gel of temperature gradient site-directed mutagenesis PCR for S10A, E38Q and D92N. For each mutant the arrows show increasing temperature from 50-60 °C. Expected plasmid product was obtained for S10A (box 1) and D92N (box 3), but not E38Q (box 2).

Site-directed mutagenesis of D8A, D110A, D110N and T108A resulted in plasmid product without the need for optimization, and the transformations were successful. However, the sequencing results were either too messy to determine the true sequence, didn't align to the wild-type $VapC_{PAE2754}$ gene, or contained an insertion upstream of the mutation site that matched the primer sequence. Two-halves mutagenesis was attempted, and while this also resulted in correct PCR products throughout all the steps, problems arose during ligation. Trials of different reaction conditions and new reagents were unsuccessful (results not shown). New primers were ordered with the goal of increasing the specificity of the binding and eliminating the inserts seen in the sequencing results. Site-directed mutagenesis using these new primers produced faint bands on a gel (Figure 3.3). Sequencing confirmed that D8A, D110N and D110A contained the base mutation, while repeat PCR reactions were required to obtain mutated product for T108A, due to random insertions occurring as described above.

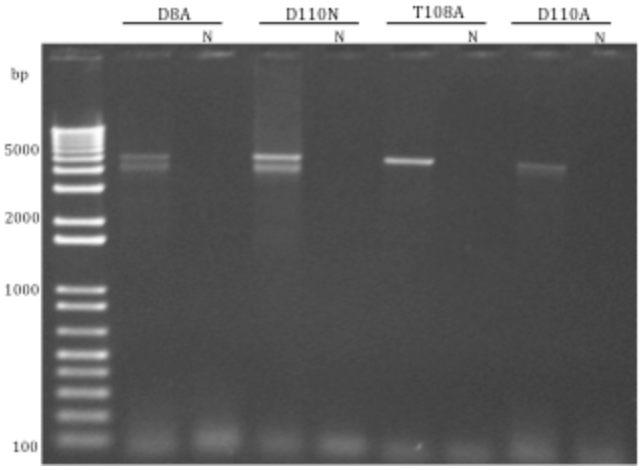


Figure 3.3. Agarose gel of site-directed mutagenesis PCR for D8A, D110N, T108A and D110A. For each mutant the PCR reaction was run next to a negative control (N) containing no plasmid. D8A and D110N have two product bands, likely corresponding to supercoiled and nicked plasmids.

D8N also produced the expected products using site directed mutagenesis but the sequencing results showed an indecipherable electropherogram. Unlike the above mutants however, the gene product from the two halves method was successfully ligated to create the mutant plasmid (results not shown).

Mutated plasmid for E38H was obtained from a temperature gradient site-directed mutagenesis PCR (results not shown).

3.2.3 Protein Expression and Purification

Previous studies have shown VapC to be difficult to express on its own due to its toxic effect on the cell (McKenzie, 2011; Miallau, et al., 2009; Ramage, et al., 2009). The two VapC toxins from *P. aerophilum* however, can be expressed and purified free from the antitoxin, and remain soluble and stable in solution. This method was developed by J. L. McKenzie (McKenzie, 2011)

3.2.3.1 $VapC_{PAE2754}$ Purification

$VapC_{PAE2754}$ protein was expressed and purified according to the methods in section 2.2. The result from the Ni^{2+} affinity column purification (Figure 3.4) shows a large peak (B) and a smaller peak (A) that eluted earlier. When run on an SDS gel, the fractions from peak A showed a weak band corresponding to the size of the VapC protein but also showed a larger product that is likely to be impurities, potentially from an *E. coli* protein with a high histidine content. A similar peak with the same impurities was observed previously with another protein. These earlier fractions from peak A were not pooled into the final protein solution. The fractions collected from peak B contained large amounts of extracted protein. The higher band is likely to be dimeric protein, while the pale lower bands are likely to be degradation of the protein. The collected protein solution was dialysed out of the imidazole buffer before the characterisation experiments were carried out.

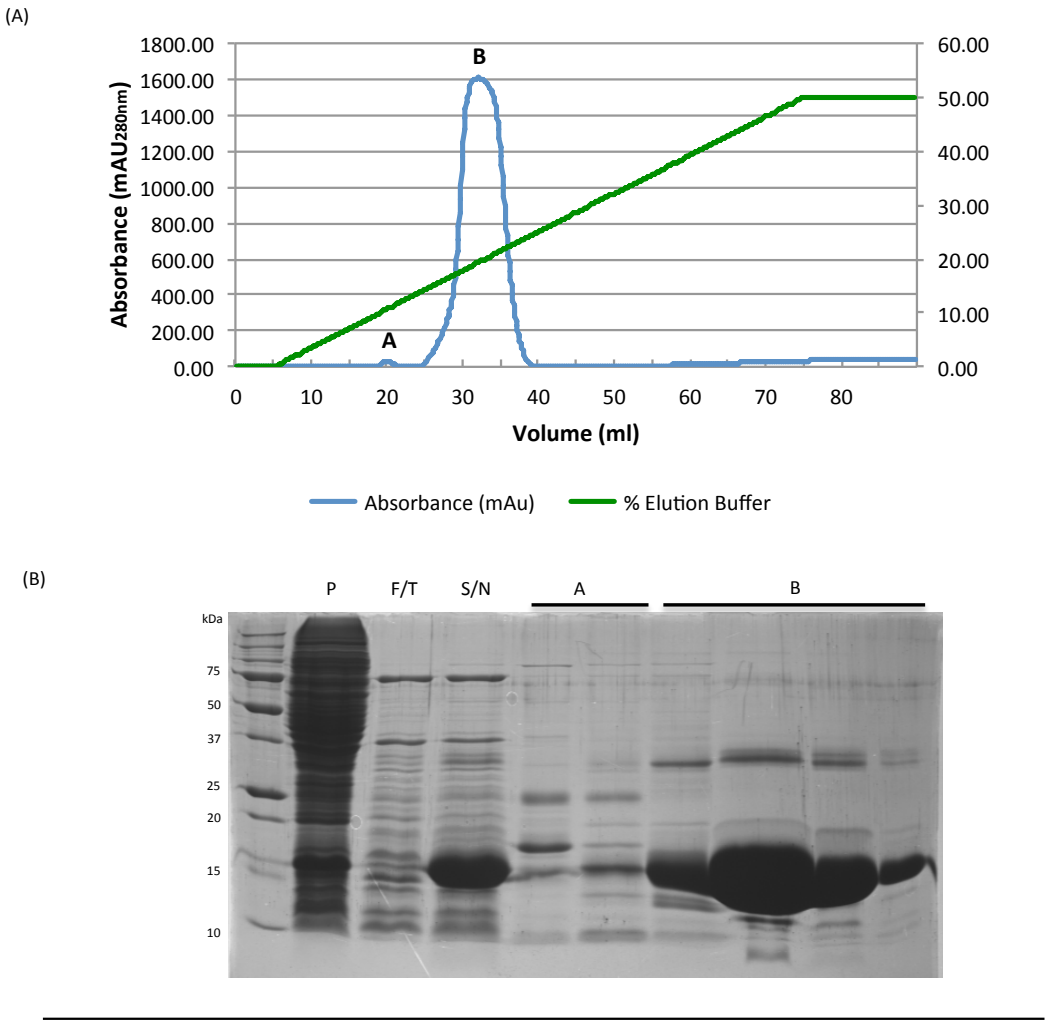


Figure 3.4. VapC$_{PAE2754}$ purification. A) IMAC Purification trace. Shows UV absorbance elution profile (blue) of the WT protein and percentage elution buffer (green). **B)** SDS-PAGE results. P-pellet, F/T – flow through, S/N – supernatant. Peak A eluted first, there fractions were not collected. Peak B corresponds to the fractions containing the protein that were pooled.

3.2.3.2 Mutant VapC$_{PAE2754}$ Purification

The mutants (D92N, E38Q, S10A, D8A, D8N, D110A, D110N, T108A, E38H) were all expressed and purified following the same method used for the wild-type. As a general rule, the mutant purification resulted in less protein than the wild-type purification, indicated by lower absorbance values in the IMAC purification chromatogram (not shown), smaller bands in the SDS-PAGE gels (Figure 3.5) and reduced concentrations shown by absorbance spectrometry once the protein fractions were pooled. As with the wild-type, a small peak (A) was observed, which eluted earlier than the protein peak (B). These fractions were not collected.

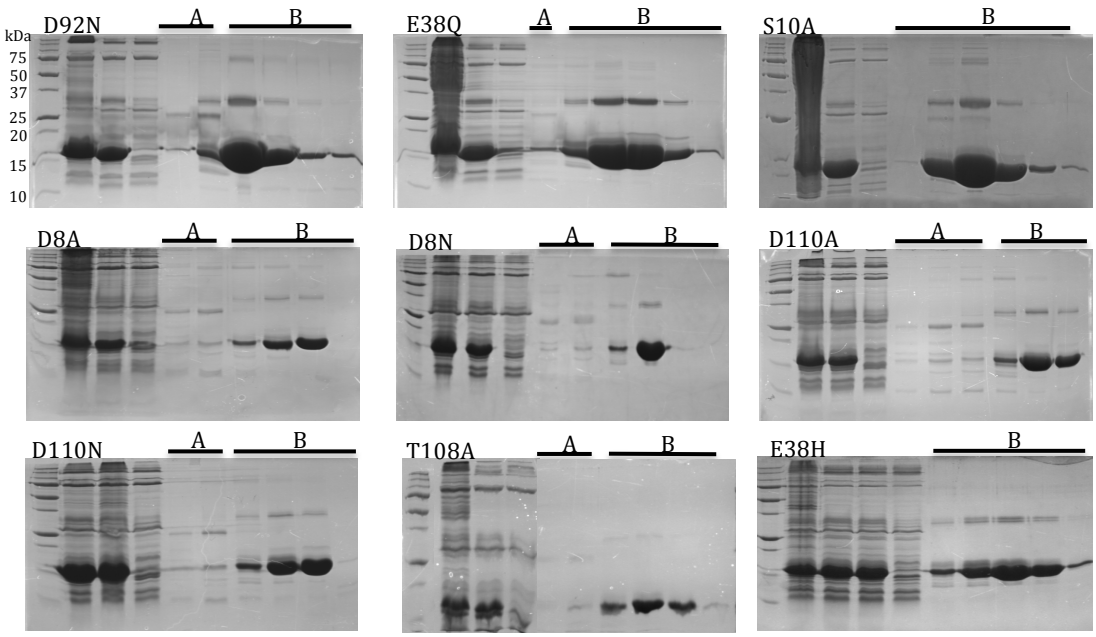


Figure 3.5. SDS-PAGE results for Ni purification of mutant proteins. Lanes labelled A represent the first eluted peak, and B represents the second. The first three lanes of each gel are the pellet, the load and the flow through respectively.

3.2.3.3 Size Exclusion Chromatography

The proteins were run on a size exclusion column to determine whether the protein was behaving as a dimer or a tetramer and for a further purification step to clean up the sample for mass spectrometry. A sample calculation is shown below using the wild-type $VapC_{PAE2754}$, which eluted off the column as two peaks, the first at 74.6 ml (A) and the latter at 80.5 ml (B). The equations used are found in methods section 2.2.3.

Calculation of K_{av} for $VapC_{PAE2754}$

A) V_e = 74.63 – 0.24
= 74.39 ml
K_{av} = (74.39 – 46.26) / (124 – 46.26)
= 0.36

B) V_e = 80.5 – 0.24
= 80.26 ml
K_{av} = (80.26 – 46.26) / (124 – 46.26)
= 0.44

Calculation of the molecular weight of $VapC_{PAE2754}$

A) MW $= e^{((0.36 - 2.1739) / -0.159)}$

$= 89.01$ kDa

B) MW $= e^{((0.44 - 2.1739) / -0.159)}$

$= 55.36$ kDa

The same calculations were carried out for the mutant proteins and tabulated below (Table 3.2). For each protein the calculated molecular weight was compared to the expected weight of a monomer (17.9 kDa).

Protein	Column Type	Ve	Kav	MW (kDa)	MW/monomer
2754 (A)	10/300	74.39	0.362	89.01	4.97
2754 (B)	10/300	80.26	0.437	53.36	3.09
D92N	16/60	15.86	0.485	33.46	1.87
E38Q	16/60	15.73	0.477	35.93	2.01
S10A	16/60	15.85	0.484	33.65	1.88
D8N	10/300	87.69	0.533	30.35	1.69
D110A	10/300	87.76	0.534	30.17	1.68
D110N	10/300	88.03	0.537	29.52	1.65
E38H	10/300	81.32	0.451	50.81	2.83

Table 3.2. Size exclusion data for calculation of molecular weight of wild-type and mutant proteins. The final column shows the calculated molecular weight divided by the expected molecular weight for a single monomer.

3.3 Characterisation of the Mutant Proteins

3.3.1 Analysis of Kinetics by Fluorogenic Assay

The kinetic assay was designed so that a change in fluorescence could be used as a measure of the protein activity. The criterion for the substrate is rapid cleavage at a single cut-site. The substrate has a flurophore attached to the 5' end and a quench attached to the 3' end. Once the substrate has been cleaved the quench no longer exerts its effect upon the fluorophore and the fluorescence increases. The fluorescence can be accurately measured in a fluorometer, so that each of the mutants can be assayed and differences in the rate of cleavage can be determined.

3.3.1.1 Substrate Analysis

Substrate oligonucleotides (oligos) were designed to contain a single cut-site that is cleaved rapidly by VapC protein, as described in method section 2.5.1. The substrates were analysed using MALDI-TOF mass spectrometry.

A range of substrates were trialed to find one that would meet the requirements for a fast and accurate assay: A chimeric oligonucleotide containing the cut-site flanked by DNA (Oligo1a and 1b); the cut-site flanked by a triplet of adenosines (Oligo2a and 2b); the hybrid sequence in all RNA (Oligo3a and 3b); and the cut-site surrounded by a random AT rich selection of bases (OligoA, B, and C).

Mass spectrometry results showed that the substrate peaks for Oligo1a, Oligo 1b, Oligo2a, 2b and Oligo 3a were not degraded by $VapC_{PAE2754}$ after a 60 minute assay (results not shown). Oligo3b however was cleaved. The MALDI-TOF spectra showed the relative intensity of the substrate peak was gradually decreased over the time course and product peaks appeared and increased over the time course. Comparison of the Oligo3 results show that despite GGGG being the consensus cut-site, GGUG is cut more rapidly. However for Oligo3b the early product peaks were further degraded over the rest of the time course indicating that there is a secondary cut-site present.

For the oligo to function effectively as a substrate for a fluorogenic assay this second cut site needed to be removed.

Based on the above results three new substrates were designed to minimize secondary cut-sites and contain only the GGUG cut-site. Mass spectrometry of these substrates with the wild-type revealed that all three had only the expected cut-site. No secondary cleavage was seen over the 60 minute time course. The mass spectrometry spectra for the time course assay of OligoC is shown in Figure 3.6. The initial peak at 5060 m/z seen in the negatives corresponds to the intact substrate. At one minute peaks that correspond to the product are already visible, and at five minutes the substrate peak is gone and two product peaks are observed at 2496 and 2560 m/z. These correspond to the expected masses of the two product halves of the substrate (Figure 3.6B).

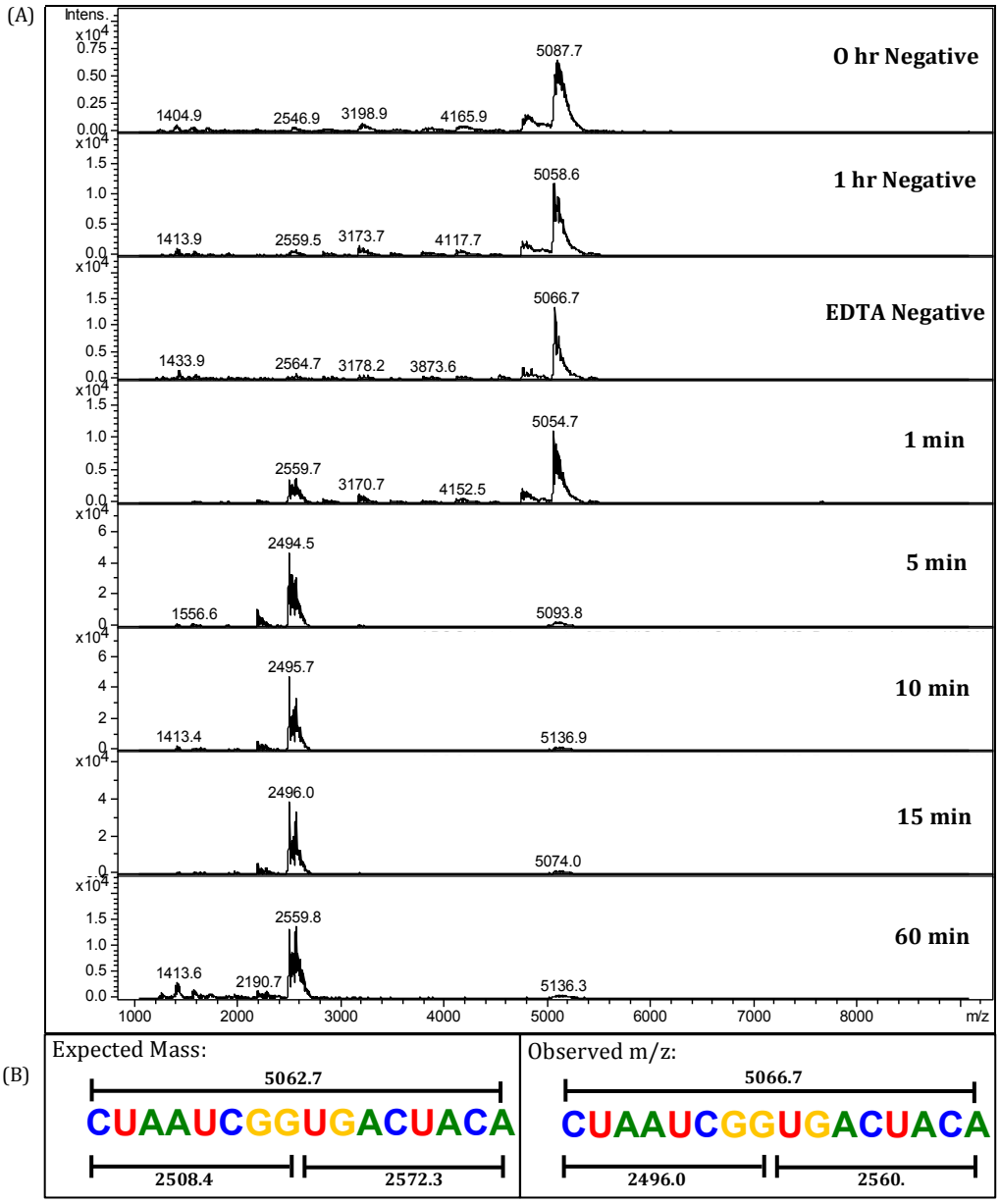


Figure 3.6. OligoC assay with VapC$_{PAE2754}$. A) Controls; O hr and 1 hr negatives contain no protein; EDTA sample contains 20 mM EDTA. Time Course; 1-60 minutes assay time points. Analysed by Bruker Data analysis and baseline subtracted. **B)** OligoC sequence showing expected and observed masses for substrate (top) and products (bottom). Expected masses were calculated using MONGO Mass Calculator and observed masses are taken from EDTA sample and 15 minute sample.

OligoC was selected for the fluorogenic substrate because it was cut the fastest by the protein, allowing for shorter assay times, thus more rapid analysis of the wild-type proteins and collection of kinetic data. The substrate was ordered with a fluorophore (6-FAM) on the 5' end and a quench (IBkQ) on the 3' end.

3.3.1.2 Standard Calibration

A standard calibration plot was created as in section 2.5.2.3 using the two halves of the substrate combined in equal molar ratios over a range of concentrations (0.1 μM - 9.0 μM). For each assay the average fluorescence over a 100 s time course was plotted against concentration and a line of best fit was determined for the points (Figure 3.7).

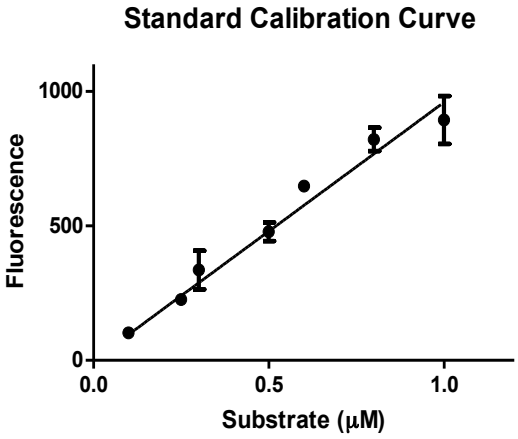


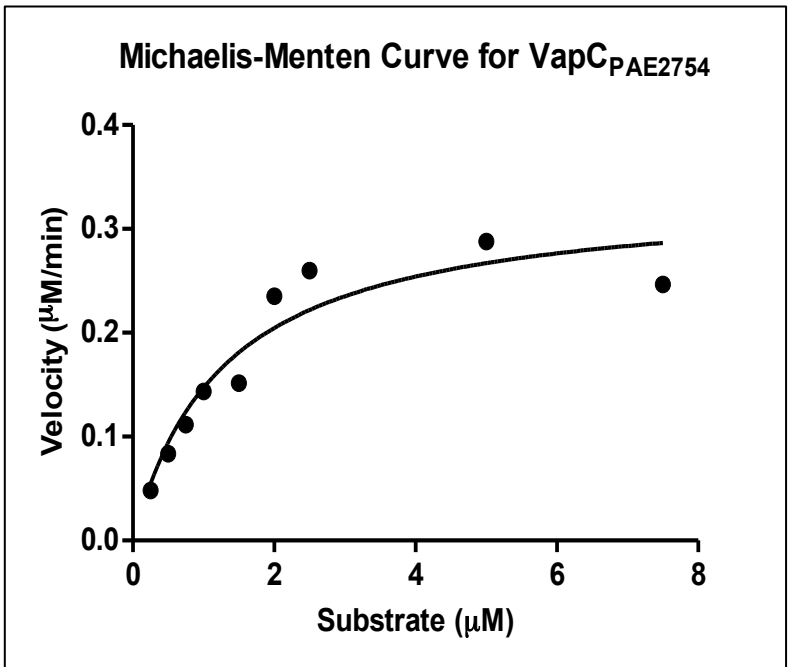
Figure 3.7. Standard calibration curve for fluorogenic product halves. Fluorescence is plotted against increasing concentrations of both the product halves (fluorophore and quench) of the substrate together. Values for 0.3, 0.5, 0.8 and 1μM show error bars, and are an average three values

At low concentrations (where fluorescence was below 1000 units) this line remained linear, however as the fluorescence increased the slope of the line started to decrease. Because of this the slope of the line was determined using points below 1.5 μM. Therefore the slope of the line of best fit could be used to calculate the rate of activity (velocity) from the change in fluorescence.

3.3.1.3 $VapC_{PAE2754}$ Wild-Type Michaelis-Menten Kinetics

The wild-type enzyme was assayed against increasing concentrations of the fluorogenic substrate (0.25 μM – 7.5 μM). The change in fluorescence was measured immediately for each reaction over 100s to determine the initial rate, as described in section 2.5.2.4. From this, the velocity at each concentration was calculated. The data was plotted on a graph and fitted with a Michaelis-Menten curve (Figure 3.8).

(A)



(B)

Vmax (μM s^{-1})	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (μM-1. s-1)
5.59 x 10^{-3}	1.35	0.015	0.5

Figure 3.8. Kinetic analysis results for VapC$_{PAE2754}$ Fluorogenic Assay. A) Michaelis-Menten plot for VapC$_{PAE2754}$. **B)** Table of kinetic data, See method 2.5.2.5 for calculations.

Above a substrate concentration of 5 μM, the fluorescence began to decrease rather than flattening out as expected for the Michaelis-Menten model. An explanation for this could be that the concentration of the products is no longer within the linear range determined by the standard calibration curve. Therefore, the quench remains in close proximity to the fluorophore even once the substrate is cut, and fluorescence no longer reflects the amount of cleavage product. Substrate inhibition was also proposed as explanation for the decreasing values and this model curve was fitted to the data. This resulted in an unexpectedly high K_M value, and so the model was dismissed.

Kinetic analysis was carried out as outlined in section 2.5.2.5. The values for V_{max} and K_M were obtained from the VapC$_{PAE2754}$ Michaelis-Menten graph using Prism 5.02 (GraphPad Software, USA). Using these data values, k_{cat} and k_{cat}/K_M were calculated for VapC$_{PAE2754}$, and tabulated in Figure 3.8 below.

3.3.1.4 Mutant Kinetic Assays

The mutants were assayed in triplicate using a concentration of substrate equal to 5 x the K_M of the wild-type. The initial rate was determined from the slope of the graph of absorbance over time (Table 3.3). A blank measurement was taken for each assay before the protein was added as described in section 2.5.2.4. Michaelis-Menten kinetic analysis was not carried out on the mutant proteins due to the low levels of activity observed and the expense of the fluorogenic substrate. The final values of activity for the mutants were determined as a percentage of the calculated rate of the wild-type (listed in Table 3.3 and plotted in Figure 3.9).

Sample	Rate (RFU/min) at 5x Km			Average rate (RFU/min)	Activity (% of WT)
	#1	#2	#3		
WT	ND	ND	ND	108.75	100
Blank	*	*	*	2.79 (5.71)	2.57 (5.25)
E38Q	12.19	9.96	12.50	11.55	10.62
D92N	10.05	11.9	10.9	10.78	9.91
S10A	26.25	30.57	33.57	30.13	27.71
E38H	7.91	6.77	6.47	7.05	6.48
D8N	6.34	5.14	8.28	6.59	6.06
D110A	4.09	4.98	4.28	4.45	4.09
D110N	9.32	4.20	5.45	6.32	5.81
D8A	7.27	6.07	ND	6.67	6.13
T108A	83.34	90.69	ND	87.02	80.01

Table 3.3. Fluorogenic assay data for mutant proteins. Activity is determined from assay results and given in triplicate. Rate of activity is calculated in relative fluorescent units per minute. These values were averaged to calculate the activity as a percentage of the wild-type. ND= value was not measured. Values in brackets for blanks are calculations plus 2 standard deviations. *See appendix for values obtained for blank measurements and for statistical analysis.

Statistical analysis was carried out on the data obtained for the blanks. Error was assumed to be 2 standard deviations (SD) above the mean (value in brackets in table above). The maximum value for the blank was determined to be 5.25 % of the wild-type, and this provides a baseline that needs to be considered when determining whether or not the mutants retained any activity. SD values were also calculated for the mutant data and these are presented in the appendix (Table A.6).

Error is most likely to arise from pipetting inaccuracy (due to small volumes of substrate and protein being used) and from RNA contamination due to the sensitivity of the fluorometer.

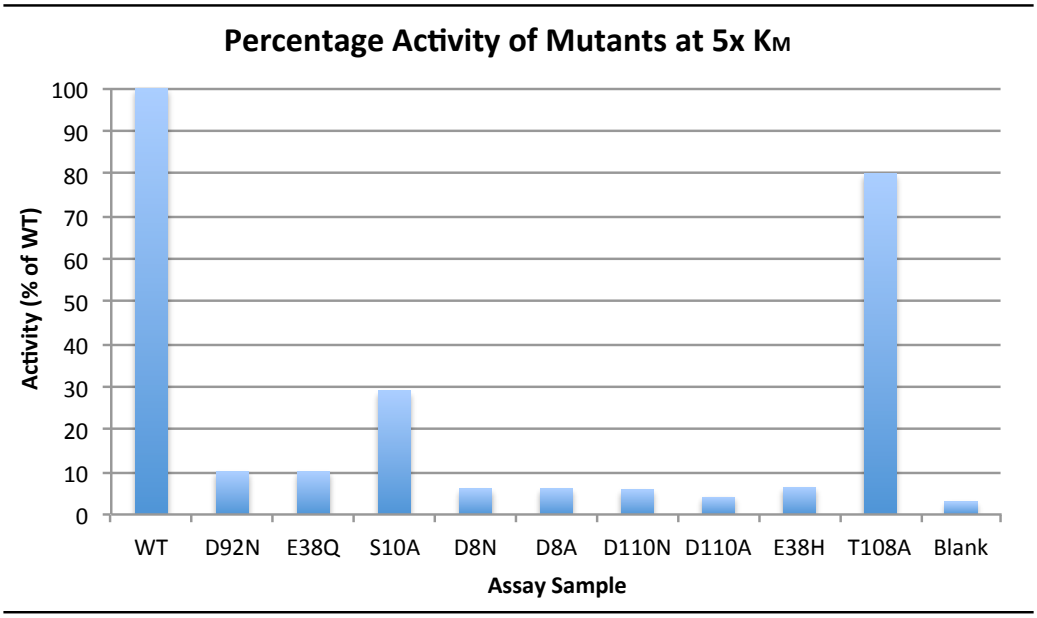


Figure 3.9. Graph of percentage wild-type activity for mutant proteins. Average activity of the mutants is plotted as a percentage of the WT (VapC$_{PAE2754}$) activity at 5x Km. The blank plotted is an average of the blank measurements taken for each assay at 5x K_M.

Mutants D8N, D8A, D100N, D110A, and E38H all demonstrate activity barely above the blank average, and therefore do not appear to be cleaving the substrate. D92N and E38Q appear to retain limited activity. The percentage activity is visibly higher than the blank measurements, but only around 10% of the wild-type. S10A and T108A retained 28% and 80% of the WT activity respectively.

3.3.2 Additional Characterisation of the Mutants

To verify the activity results, the proteins were assayed against pentaprobe RNA sequences, and the substrate without the fluorophore or quench attached. S10A was also assayed against 932 Oligo 5 for further characterisation. The aim of these additional assays was to provide more information about the activity and specificity of the mutant proteins.

3.3.2.1 Pentaprobe RNA Assays

Preliminary RNase assays were carried out using pentaprobe RNA to provide an initial screen for activity and specificity. These were analysed on urea denaturing gels (see section 2.4.2). Pentaprobes 922 and 924 are both known to be cut efficiently by VapC$_{PAE2754}$ (McKenzie, 2011) and were therefore selected for these assays.

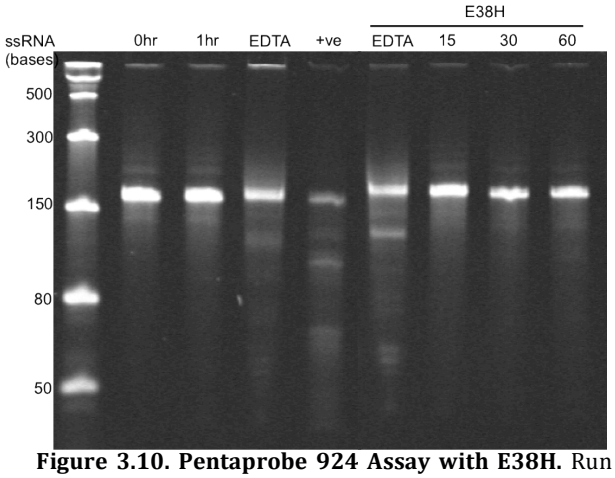


Figure 3.10. Pentaprobe 924 Assay with E38H. Run on a 10% urea denaturing gel. 0hr and 1hr are protein free negatives, EDTA negatives contain 20mM EDTA, and the +ve control contains VapC$_{PAE2754}$ protein. E38H protein was assayed over a time course 15-60 minutes.

No activity was seen against either of the pentaprobes for E38Q, D92N, S10A, E38H or D8N. An example of these results is shown in Figure 3.10. No cleavage of the substrate can be seen over the 60 minute time course. These assays provided an early indication that the mutant proteins were inactive.

The 924 pentaprobe assay for S10A is shown in Figure 3.11. This assay as and the 922 assay (not shown) do not show bands corresponding to cleavage products, however the substrate bands in the time course are weaker than those seen for the negatives and the presence of higher bands indicate that RNA is binding to the enzyme.

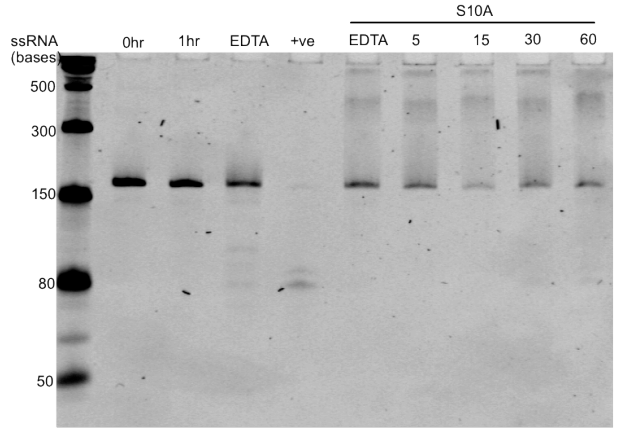


Figure 3.11. Pentaprobe 924 Assay with S10A. Run on a 10% urea denaturing gel. 0hr and 1hr are protein free negatives, EDTA negatives contain 20mM EDTA, and the +ve control contains VapC$_{PAE2754}$ protein. S10A protein was assayed over a time course 5-60 minutes.

D8A, D110N, D110A and T108A mutants were not tested against pentaprobe RNA due to time restraints.

3.3.2.2 OligoC Assays

RNase assays were carried out with OligoC (the optimised fluorogenic substrate without the fluorophore or quench attached) and analysed by mass spectrometry to validate activity results seen in the fluorogenic assays. Mass spectrometry was used because cleavage of the OligoC RNA produces fragments difficult to visualise on a urea-denaturing gel due to their short length. D8N, E38H, D92N, E38Q and S10A were all tested against this substrate. The results are shown below in Figure 3.12. After a 60 minute assay reaction no degradation of the substrate peak was seen for D92N, E38Q, D8N or E38H, however the peak was completely lost in the S10A 60 minute sample. D110A, D110N, D8A and T108A were not assayed with this substrate due to time constraints.

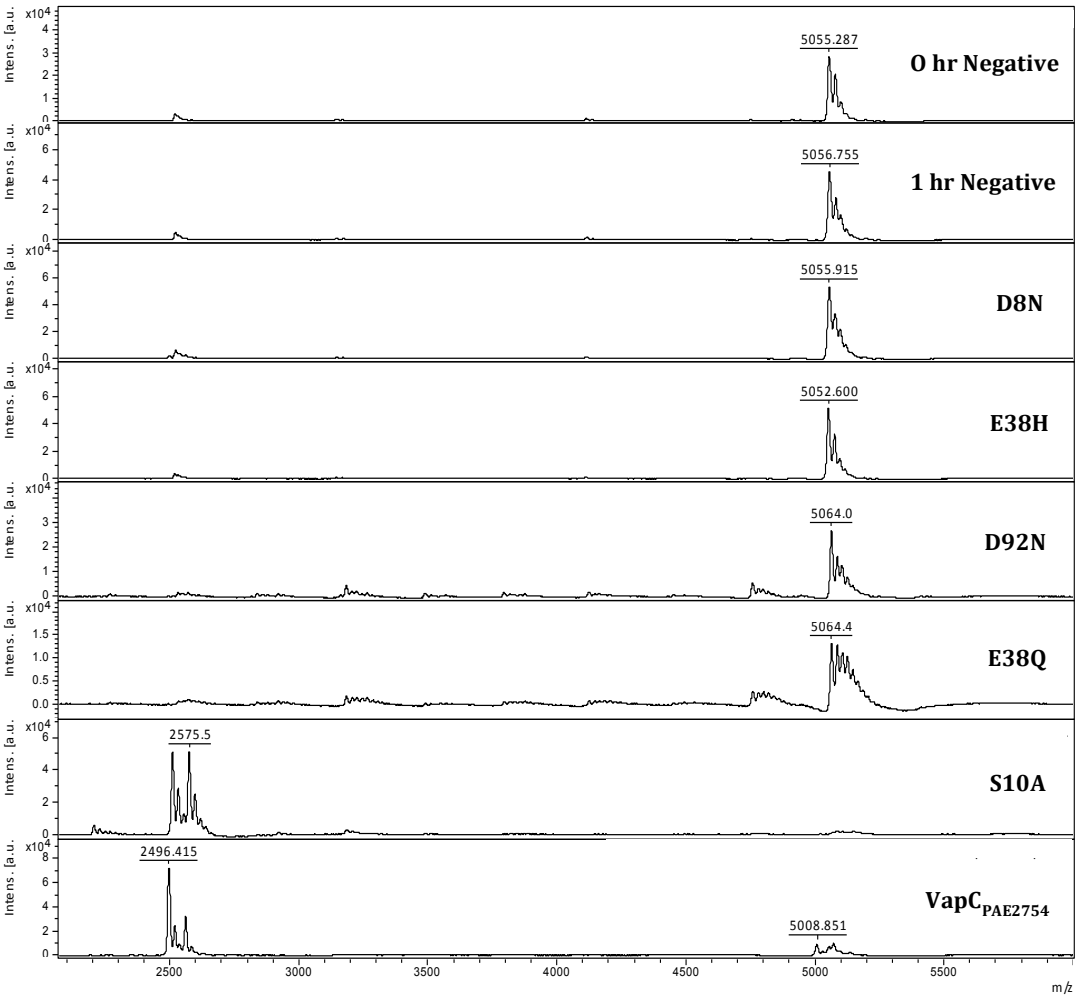


Figure 3.12. MS spectra of OligoC Assay with D8N, E38H, D92N, E38Q, S10A and Wild-Type. Controls; 0 hr and 1 hr negatives contain no protein, EDTA sample was unable to be collected. Mutant assays are shown at 60 minute time points, and WT is shown at 5 minute time point for comparison. Analysed by Bruker Data analysis and baseline subtracted.

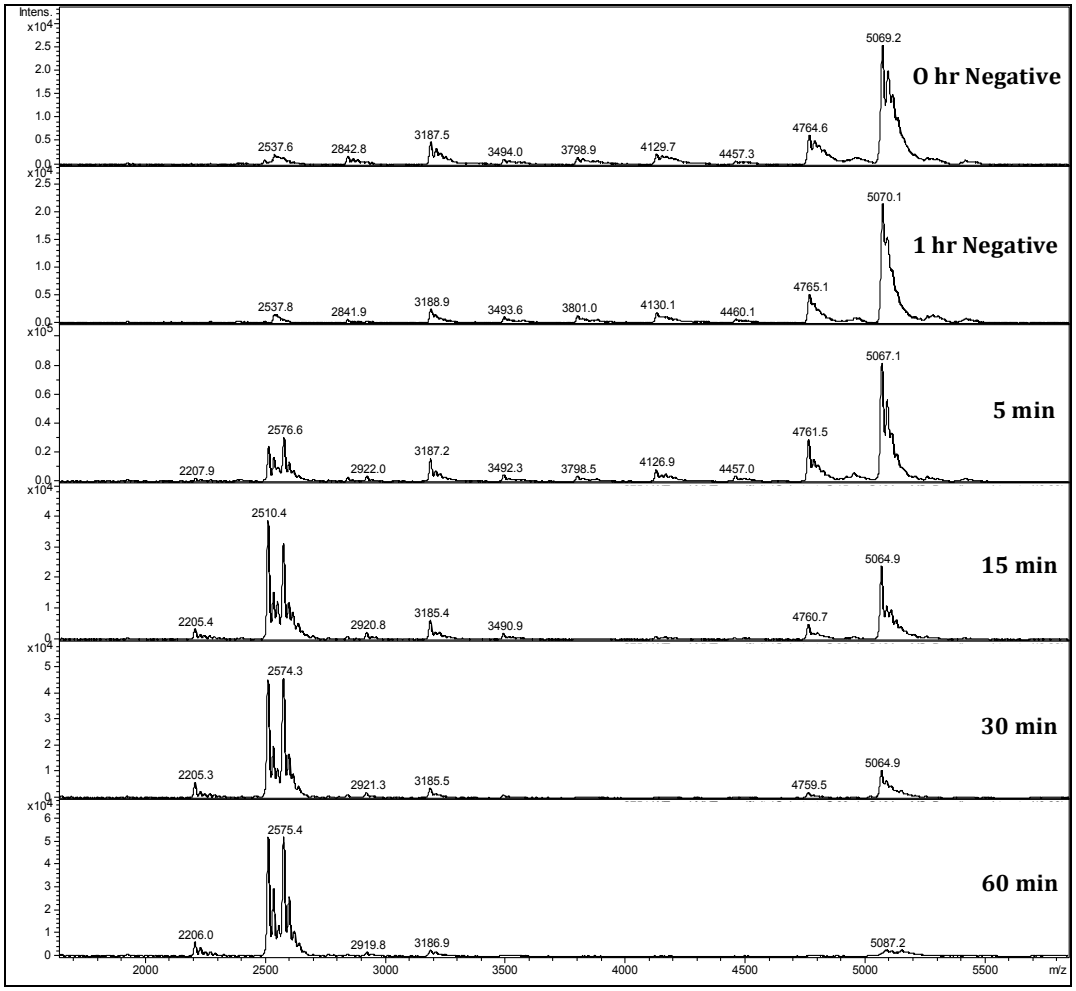


Figure 3.13. MS spectra of OligoC time course assay with S10A mutant. Controls; 0 hr and 1 hr negatives contain no protein, EDTA sample was unable to be collected. Time course; 5-60 minute assay samples with S10A mutant protein. Analysed by Bruker Data analysis and baseline subtracted.

S10A was assayed as a time course to determine how rapidly the substrate was being cut. Figure 3.13 shows the gradual loss of the substrate peak (5070 m/z) and an increase of the product peaks (2490 and 2575 m/z). When compared to the WT control, the same masses are seen, indicating that the protein is targeting the same cut-site. Analysis of the time points shows that S10A has not degraded all of the RNA after 30 minutes, while in the WT the substrate is fully degraded within 5 minutes. These results are at odds with those from the pentaprobe assay (Figure 3.11). This could indicate that small amounts of product are unable to be visualised on a urea-denaturing gel. The results could also reflect that OligoC is a more optimal substrate than the pentaprobe RNA, explaining why it was cut by the less active mutant S10A protein, while the longer pentaprobe sequences were not.

3.3.2.3 Specificity Assay

S10A was tested with 932 pentaprobe oligo 5 to check whether the mutation had any effect on the specificity of the protein. The nine pentaprobe oligonucleotides were designed by J. L. McKenzie as part of the method to determine the optimal cut-site sequence of RNase proteins (McKenzie, 2011). They are small overlapping fragments of the parent oligonucleotide that mimic the secondary structure and allow accurate analysis of the RNase assays using mass spectrometry. 932 Oligo 5 is cleaved rapidly by VapC$_{PAE2754}$ and contains multiple cut-sites (McKenzie, 2011).

The time course assay of S10A with 932 Oligo5 is shown in Figure 3.14. The peak masses match closely with those seen previously for VapC$_{PAE2754}$ (McKenzie, 2011) indicating that S10A is cleaving the substrate at the same cut-sites.

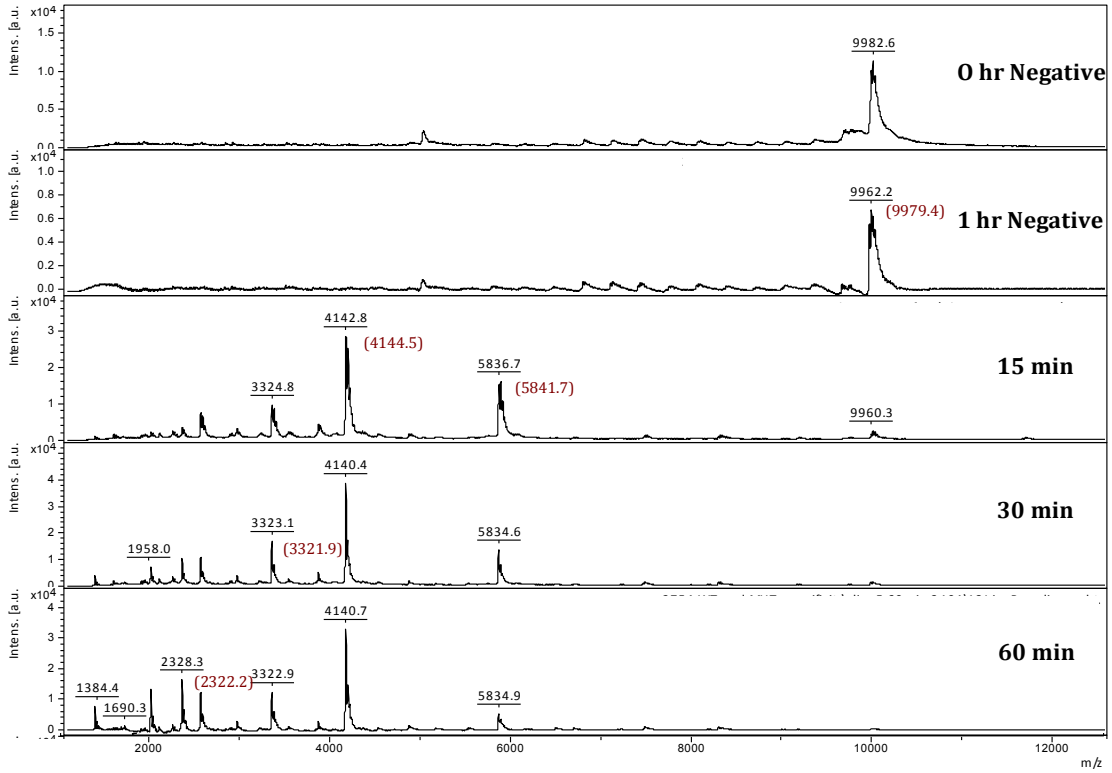


Figure 3.14. 932 Oligo5 Assay with S10A. Controls; 0 hr and 1 hr negatives contain no protein, EDTA sample was unable to be collected. Time course; 5-60 minute assay samples with S10A mutant protein. Analysed by Bruker Data analysis and baseline subtracted. Comparative masses (m/z) from VapC$_{PAE2754}$ assay are shown in red.

3.3.2.4 Manganese activity assays

As described earlier (see 1.3.2) magnesium is well suited to act as the metal cation in RNase enzymes, however RNase activity is also seen in the presence of manganese. Mn^{2+} is similar to Mg^{2+} in chemical nature but has relaxed coordination requirements. Yang and colleagues use this difference to help determine the catalysis process of a two-metal mechanism for RNase H (Nowotny, et al., 2005). A mutation of a conserved Asp to Asn prevented the direct coordination of metal B, inhibiting activity in the presence of Mg^{2+} and Mn^{2+}. The same mutation in a residue that coordinates metal A showed recovery of activity in the presence of Mn^{2+}. If $VapC_{PAE2754}$ also requires two metal ions for catalysis then a similar experiment should show recovery of activity in one of the corresponding mutations made to $VapC_{PAE2754}$ (D92N, E38Q, D8N, D110N). However, due to the high pH requirements for $VapC_{PAE2754}$ (pH9.2) a reaction buffer containing a high concentration (above 5mM) of Mn^{2+} resulted in precipitation of the metal out of solution. Activity assays with 5mM Mn^{2+} did not result in recovery of any activity for either E38Q or D92N, and could not be carried out with D110N or D8N due to precipitation.

3.4 VapC$_{PAE2754}$ Crystal Structure

3.4.1 Introduction

To determine the reaction mechanism of VapC$_{PAE2754}$ it is important to gain more information about the configuration of the active site. The number of metals required, and the interactions between the metals, the substrate, and the conserved amino acids all provide indications as to how the catalytic reaction unfolds. Molecular details of the mutant proteins also provide valuable information about the effects that an amino acid change may have on substrate binding, metal binding, and coordination. A high resolution crystal structure would provide this information and therefore crystallisation was attempted for VapC$_{PAE2754}$ and a selection of the mutant proteins. The VapC$_{PAE2754}$ structure was solved at 1.95Å resolution, however no diffracting crystals were obtained for any of the mutant proteins.

3.4.2 Crystallisation of VapC$_{PAE2754}$

The purified protein was screened over a wide range of conditions using the Mosquito® robot (see section 2.6.1). Several conditions showed promising crystals in the sitting drops. Fine screens around these were laid using the hanging drop method (see section 2.6.2). Most conditions did not result in repeat crystallization, but fine screens around 0.1 M $CdCl_2$ hydrate, 0.1 M NaAc and 30% v/v PEG400 resulted in crystallisation after approximately one week (

Figure 3.15).

(A) (B)

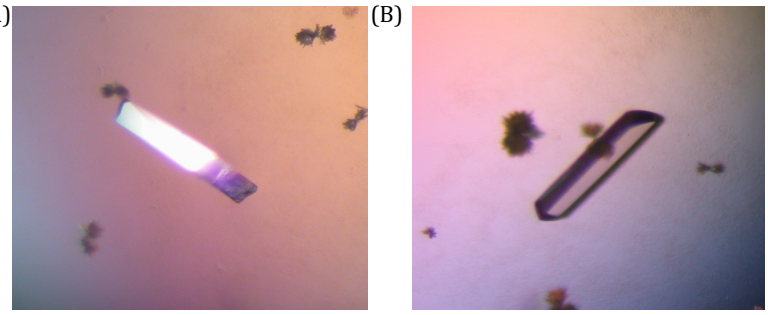


Figure 3.15. Wild-Type VapC$_{PAE2754}$ crystals for which data was collected. Conditions: **A)** 0.1 M $CdCl_2$ hydrate, 0.1 M NaAc and 32% v/v PEG400. **B)** 0.15 M CdCl hydrate, 0.1 M NaAc and 28% v/v PEG400.

3.4.3 X-Ray Diffraction

$VapC_{PAE2754}$ crystals from the fine screen were selected and tested at Auckland University on a rotating copper anode. They showed diffraction to a high resolution. Crystals from similar conditions in the fine screen were selected and sent to the Australian synchrotron (see section 2.6.4). The crystal structure was solved to a resolution of 1.95 Å using the methods in section 2.6.5.

3.4.4 Structure Determination

3.4.4.1 Data Collection for $VapC_{PAE2754}$

The data was scaled and integrated in space group $P2_12_12_1$. The Matthews coefficient predicted 4 molecules in the asymmetric unit. Complete data collection statistics are shown in table 3.4.

Data Collection	
Space Group	$p2_12_12_1$
Wavelength (Å)	0.9537
Cell Parameters	
a	42.41
b	94.72
c	132.96
α	90
β	90
γ	90
Resolution Range (Å)	23.68-1.95
R_{merge}	0.151 (0.554)
No. of Observations	649408
No. of Unique Reflections	39971
Mean I/σ(I)	13.2 (5.2)
Completeness	99.9 (99.8)
Multiplicity	16.2

Table 3.4. Data Collection Statistics for $VapC_{PAE2754}$

3.4.4.2 Structure Solution

The structure was solved using molecular replacement and the previous $VapC_{PAE2754}$ structure (1V80). The initial model was built using phenix autobuild. Towards the end of refinement the data was scaled with the anomalous flag on, allowing the calculation of an anomalous difference map.

3.4.4.3 Refinement of $VapC_{PAE2754}$

The CCP4 program Refmac5 was used to run an initial round of refinement on the output model from Phenix AutoBuild, using the scaled mtz file from the data collection. The output file from this refinement was used to carry out manual building of the model in COOT, as described in section 2.6.5.5. The refinement statistics are summarized below in Table 3.5.

Refinement Statistics	
R_{work} (R_{free})	23.1 % (27.96%)
Total No. of Atoms	4305
No. of missing atoms	146
No. of protein atoms	4085
Other molecules/ions	24
No. of waters	196
RMSD	
Bond lengths (Å)	0.0225
Bond angles (o)	1.9358
Average B-factors	
Chain A	15.0
Chain B	20.4
Chain C	24.6
Chain D	26.3
Waters	25.7

Table 3.5. Refinement statistics for $VapC_{PAE2754}$ crystal structure

The final model has an R_{work} of 23.1 % and an R_{free} of 27.96 %. The RMSD values for the bond length and bond angles are 0.225 Å and 1.9358 ° respectively. The Ramachandran plot showed 97 % of residues in the favoured region, 2.7 % in the allowed region and 0.4 % in the unfavoured region.

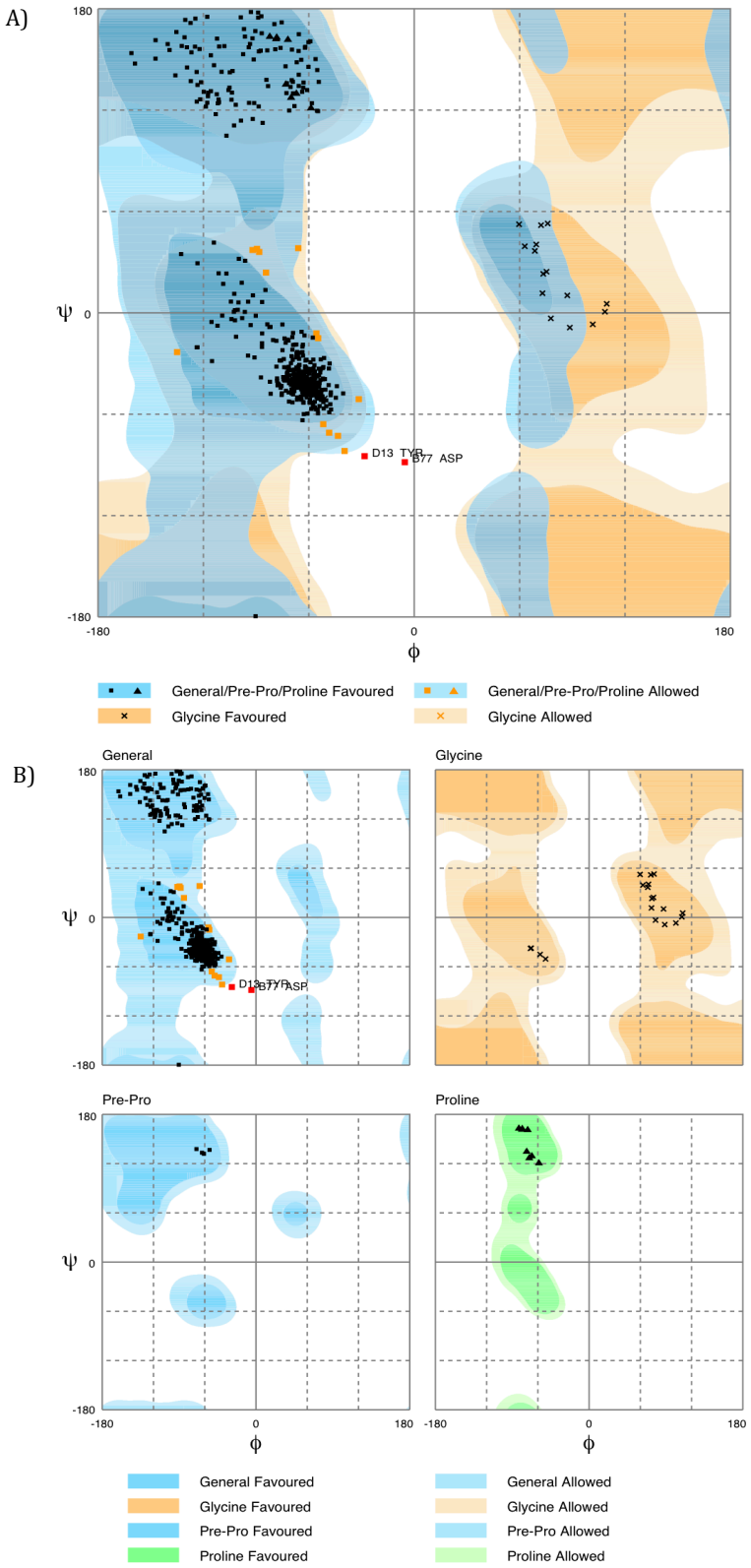


Figure 3.16. Ramachandran plot for VapC$_{PAE2754}$. Produced using Procheck in CCP4.

3.4.5 Structure of VapC$_{PAE2754}$

The *P. aerophilum* VapC$_{PAE2754}$ crystal structure has four copies of the monomer (chains A, B, C and D) in the asymmetric unit. The crystallographic symmetry shows that each of these monomers is part of a dimer, with the corresponding molecule found by crystallographic symmetry (Figure 3.16).

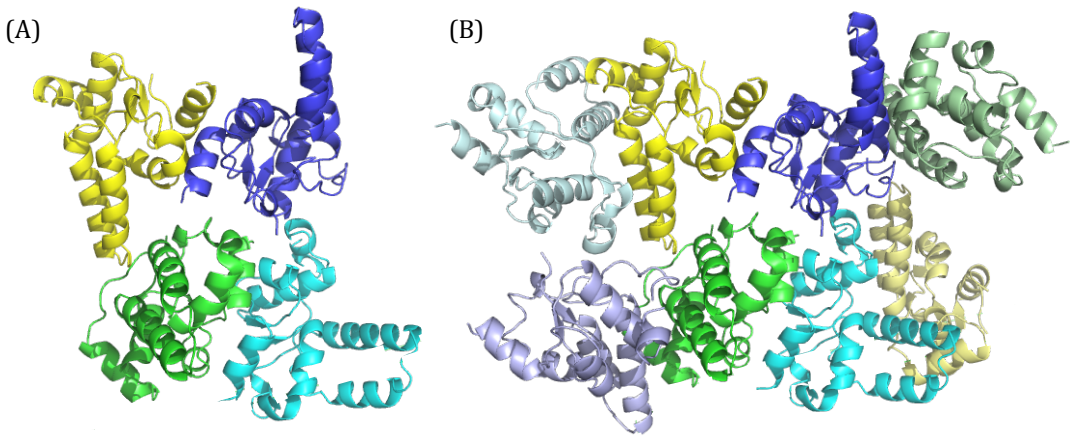


Figure 3.16. Cartoon representation of the unit cell. Chain A = green, Chain B = Cyan, Chain C = Blue, Chain D = yellow. **A)** The four monomers that make up the asymmetric unit. **B)** The dimer pairs found by crystallographic symmetry for each of the monomers are shown in pale colours corresponding to the chain.

The chains A, B and C are comprised of residues Ala2-Ala132. No interpretable electron density was seen for the C-terminal residue Gln-133, or for the side chain atoms beyond the backbone for chain A residues Asp20, Lys24, Lys63 and Lys118, chain B residues Tyr13, Asp20, Lys21, Lys24, Glu27 and Lys118, chain C residues Tyr13, Lys28, Val52, His60, Lys63 and Glu112. Chain D is incomplete. It is comprised of residues Ala2-Ala132 minus residues Glu112-Lys118 for which no interpretable electron density could be found (Figure 3.17). There was also insufficient density for the side chain atoms beyond the carbon backbone for residues Leu12, Tyr13, Tyr19, Lys21, Ile23, Lys28, and Arg111 for chain D.

The structure contains 196 water molecules, and 6 Cd^{2+} ions. For monomers A, B and C a single Cd^{2+} ion is contained within the active site. For the monomer D a second Cd^{2+} ion is present in the active site and a final Cd^{2+} ion is found between residue Glu27 of chain A, and helix 3 of chain D.

The folding of the protein is similar to the previously determined structure (see Figure 3.17). The RMS value for alignment of chain A and the previously determined $VapC_{PAE2754}$ structure is 0.40. Unlike the original structure however the crystal stacking is not symmetrical or organized into tetrameric units but instead is made up of 2 parallel strands of dimers as seen in figure 3.16.

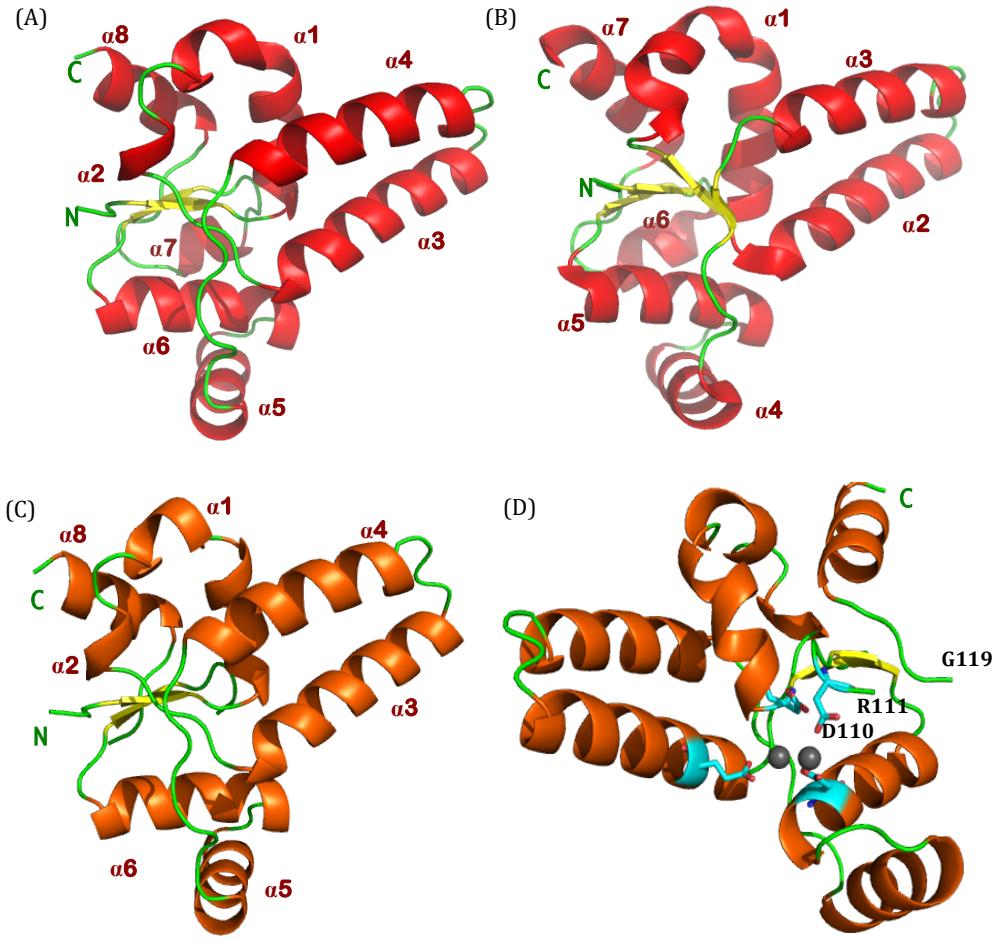


Figure 3.17. Comparison of chains A and D to previously solved structure for $VapC_{PAE2754}$ (1V80). **A)** Chain A monomer, **B)** Previous $VapC_{PAE2754}$ monomer: Helices (red) and end termini (green) are labeled. β-sheets are depicted in yellow. **C)** Chain D monomer. Helices (orange) and end termini (green) are labeled. **D)** Rotated view of chain D showing the conserved active site residues (cyan) and Cd^{2+} ions (grey). D110 and termini at each end of the missing electron density (R111 and G119) are labeled.

The RMS of the alignment of chain A to chain D is 0.48. Figure 3.17c shows that the missing residues constitute α helix 7 from chain A (or α helix 6 from the previously solved $VapC_{PAE2754}$). Other than the missing region the helices

align well, with only α3 and α4 showing slight distortion between chains A and D. The proximity of the missing residues to the active site and the Cd^{2+} ions is shown in figure 3.17d. Residues D110 and R111 were difficult to fit into the electron density, and the side chains could only be placed into the structure after the missing residues had been removed.

3.4.5.1 Binding of the Metal Ions

The Cd^{2+} ions are not coordinated in the same way in each chain. In chains A-C the Cd^{2+} ion is positioned between Asp92 and Asp110, in the opening to the pocket that constitutes the active site (Figure 3.18). The Cd^{2+} in chain A is coordinated to both oxygens of Asp92 (3.0 Å, 2.2 Å) and Asp110 (2.2 Å, 3.4 Å). It also shows coordination to two water molecules (1.9 Å and 3.9 Å). Although the metal is found in a similar position in chains B and C the interactions with the amino acids differ slightly. In chain B the metal is coordinated by a single oxygen of Asp92, and Asp110, and two water molecules (2.2 Å, 2.3 Å, 2 Å and 2.4 Å respectively). The Cd^{2+} in chain C is coordinated solely by the amino acid oxygen atoms Asp92 (2.9 Å and 2.4 Å) and Asp110 (2.2 Å and 3 Å)

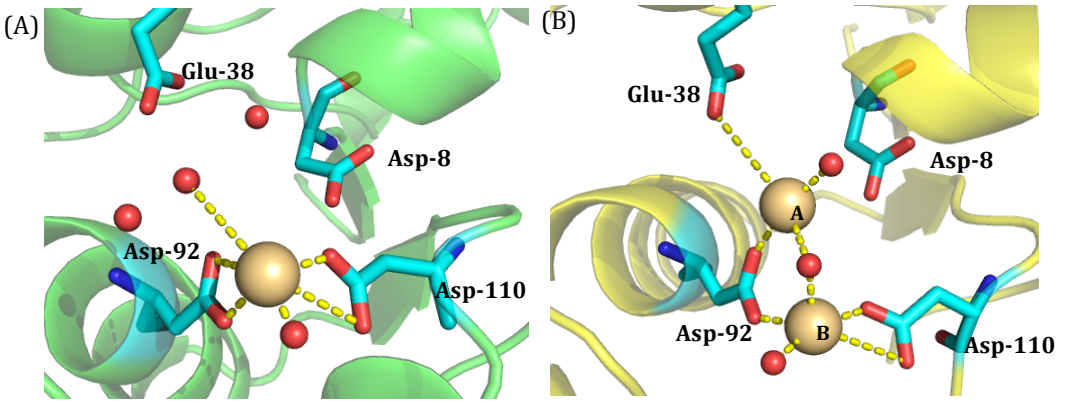


Figure 3.18. Active sites in $VapC_{PAE2754}$ crystal structure. A) ChainA (green) active site. **B)** ChainD (yellow) active site. Waters (red) and Cd^{2+} ions (pale yellow) are shown as spheres and potential interactions with the metal are shown as yellow dotted lines. Active site residues (cyan) are labeled.

In Chain D, Asp92 and Asp110 are slightly rotated away from each other in comparison to the other chains (Figure 3.18). Metal B is coordinated by a single oxygen of Asp92, both oxygens of Asp110, and two water molecules (2.4 Å, 2.4 Å, 3.8 Å, 2.1 Å, and 2.3 Å respectively) while metal A is coordinated by Asp92, Glu38 and two water molecules (2.2 Å, 3.8 Å, 2.9 Å, and 2.3 Å respectively). The two metal ions are 4 Å apart. The reason why Chain D contains a second metal, and appears to be less stable in the crystal structure than the other residues, is uncertain.

Alignment to the VapC$_{PAE0151}$ structure, which contains a single Mn^{2+} ion, gives an RMS value of 1.77 for chain A and 2.14 for chain D (Figure 3.19). The metals are in similar positions but do not overlay exactly. Alignment of chain A places the Cd^{2+} ion 1.9 Å from the Mn^{2+} ion. The Cd^{2+} is bound closer to the acidic active site residues than the Mn^{2+}. The alignment of chain D places the Cd^{2+} ion 1.7 Å from the Mn^{2+} ion. Residue D110 is in a similar orientation to VapC$_{PAE0151}$ in chain A but is distorted slightly in chain D.

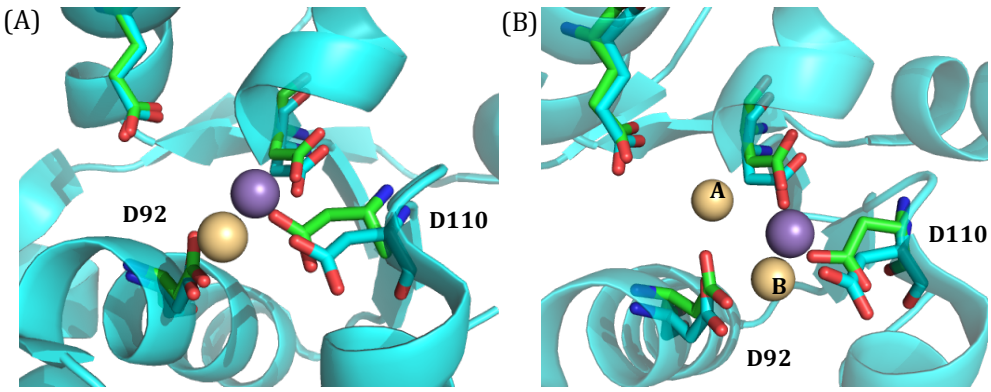


Figure 3.19. Alignment of VapC$_{PAE2754}$ and VapC$_{PAE0151}$ structure. A) Alignment of Chain A and VapC$_{PAE0151}$. **B)** Alignment of Chain D and VapC$_{PAE0151}$. In both figures VapC$_{PAE0151}$ is shown in cyan, conserved acidic residues are shown as sticks (cyan for VapC$_{PAE0151}$ and green for VapC$_{PAE2754}$). Cd^{2+} ions are shown as pale yellow spheres and Mn^{2+} are shown as purple spheres.

An acetate is bound into the active site of chains A, B and C. Chain D does not contain the acetate because the second Cd^{2+} ion is occupying the position instead. The acetate is placed centrally in the active site, between all four of the conserved active site residues and 3.4 Å away from the Cd^{2+} ion. When the structure is aligned with the $VapC_{PAE0151}$ containing the docked dinucleotide substrate the acetate and the phosphate are in similar positions (Figure 3.20). Thus the acetate could be acting as a pseudo-intermediate, mimicking the orientation of the three non-bound oxygens of the pentacovalent phosphate.

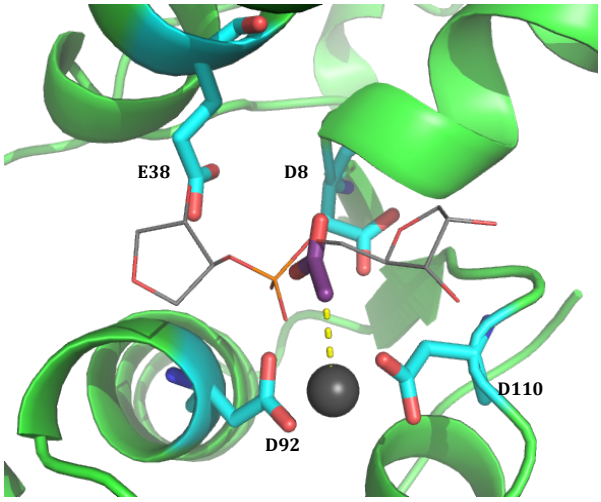


Figure 3.20. Substrate analogue in $VapC_{PAE2754}$ structure (Chain A). Conserved active site residues (cyan) are labeled. Coordination of Cd^{2+} (grey sphere) to the acetate molecule (coloured by element in purple) is shown as yellow dotted lines. Structure has been overlaid with the docked substrate (coloured by element in grey lines) from the $VapC_{PAE0151}$ structure.

3.5 Summary of Results

Nine mutant enzymes of $VapC_{PAE2754}$ were designed, expressed, and purified. They were compared to the wild-type using a fluorometric assay designed to accurately measure the rate of substrate cleavage. S10A and T108A retained unquestionable activity, while D8A, D8N, D110A, D110N and E38H were barely above baseline activity, and D92N and E38Q appeared to have greatly reduced activity. When analysed by mass spectrometry D92N and E38Q appeared to be inactive, while S10A cleaved the substrate after 30 minutes. Preliminary examination of specificity showed that S10A cut at the same site as the wild-type.

A crystal structure of the wild-type $VapC_{PAE2754}$ was solved to 1.95 Å, which is higher resolution than the previous structure. The structure also contains Cd^{2+} ions, and acetate bound in the active site. The structure aligned closely with $VapC_{PAE0151}$, however Chain D has some discrepancies. A region of electron density was missing and the conformation of D110 was distorted and there were two Cd^{2+} ions in the active site. The chain D monomer is the first time a VapC protein has been solved with two metal ions bound into the active site. The acetate bound in chains A, B and C provides information about the position of the RNA substrate, potentially mimicking the scissile phosphate.

Chapter 4. Discussion and Future Research

4.1 Discussion of Results

The results are discussed in relation to the implications for the proposed mechanism, considering both the catalysis reaction and specificity of $VapC_{PAE2754}$. The mutagenesis results and the solved crystal structure are compared to other PIN-domains, and RNase enzymes that show structural similarity to VapC enzymes. The results indicate that the reaction does not appear to proceed by the hypothesized mechanism. Therefore, revised mechanisms of cleavage are suggested, taking into account the debate between one or two metal involvement.

4.1.1 Mutant Activity

4.1.1.1 S10 and T108

Of the nine mutants made only two retained substantial activity. S10A displayed 28 % and T108A displayed 80 % activity, compared to the wild-type. Residues S10 and T108 are partially conserved, and are both involved in coordination of the D8 residue. S10 makes more contacts with the residue (Figure 4.1), and this is reflected in the lower activity seen for S10A than T108A. The corresponding residue S153 in T4 RNase also reduces catalytic activity (Bhagwat, et al., 1997).

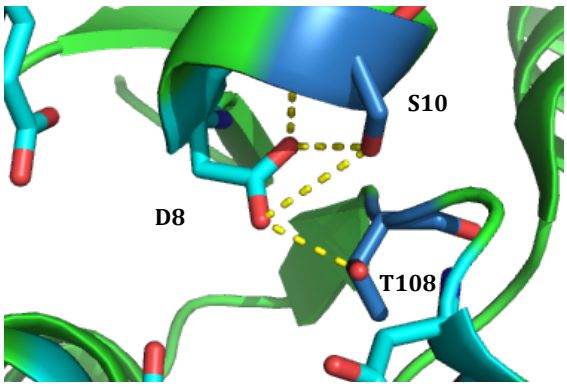


Figure 4.1. Coordination of residue D8 in the $VapC_{PAE2754}$ active site. Highly conserved acidic site residues that define the PIN-domain are shown in cyan and other conserved residues are shown in blue. Polar contacts between S10, T108, and D8 are shown as yellow dotted lines.

4.1.1.2 D8 and D110 (Conserved residues Asp(1) and Asp(3))

Mutation of D8 and D110 conserved active site residues to alanine was expected to completely knockout activity of the enzyme, and thus serve as a measure for total inhibition. D8A and D110A had 6.1 % and 4.1 % activity of the wild-type respectively. Therefore, compared to the blank D110A can conclusively be said to correspond to total inhibition of cleavage activity. Taking into account error, it is likely that D8A is also inactive.

The mutations of these residues to asparagine (D8N and D110N) were also inactive (6.1 % and 5.8 % respectively). This was not the expected result for these mutants, however previous studies have shown that isosteric mutations to conserved acidic residues inactivate the enzyme if the residue is directly involved in catalysis, or if the residue coordinates the essential metal ion (Bhagwat, et al., 1997; Devos, et al., 2007; Frazao, et al., 2006; Gan, et al., 2006). For example, in T4 RNase H the four Asp residues (Asp19, Asp71, Asp132 and Asp155) that coordinate metal one inactivate the enzyme when mutated to Asn. However, the residue coordinating metal two (D157) still retained both exonuclease and FLAP endonuclease activity, although this was reduced compared to the wild-type (Figure 4.2). Activity could be partially rescued in the D157N mutation by assay with Mn^{2+} instead of Mg^{2+} (Bhagwat, et al., 1997).

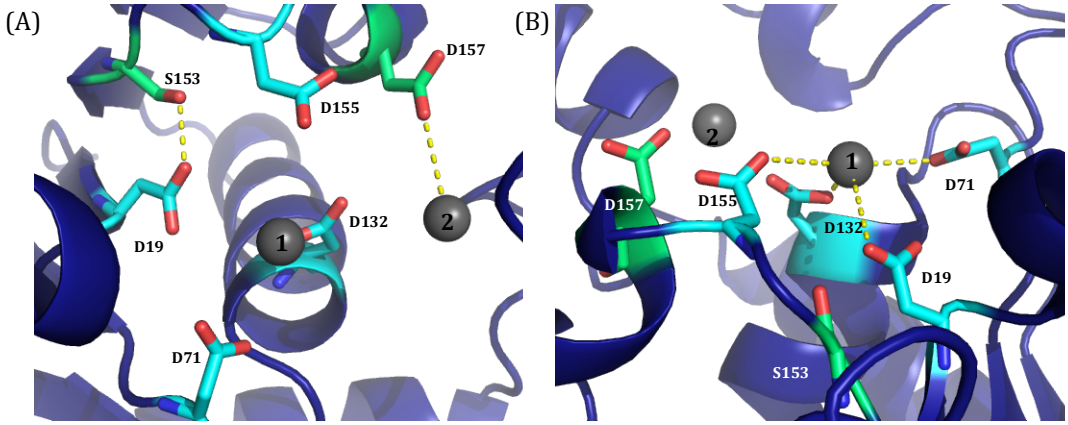


Figure 4.2. T4 RNase conserved active site residues. All conserved amino acids are shown as sticks and labeled. Mg^{2+} ions are shown as labeled grey spheres. **A)** Coordination (yellow dotted line) of residues that decrease but don't eliminate activity (green). **B)** Coordination of metal 1 to residues essential to catalytic activity (cyan).

Rescue of an Asp involved in binding the second metal ion was also seen in RNase H from *Bacillus halodurans* in the presence of Mn^{2+} (Nowotny, et al., 2005). Yang and colleagues propose that this is due to the reduced coordination requirements and larger size of Mn^{2+} compared to Mg^{2+}. Unfortunately, this method was unable to be used to test if any of the isosteric mutations in $VapC_{PAE2754}$ retained activity in the presence of Mn^{2+}. This was due to precipitation of the manganese out of solution at the pH of the buffer required for stable protein.

4.1.1.3 D92 and E38 (Conserved residues Asp(2) and Glu)

The D92 and E38 residues were proposed to act as the catalytic acid and base. Unexpectedly, D92N and E38Q both appear to be more active than D8N and D110N. The fluorogenic assays show that D92N and E38Q retain approximately 10 % of the wild-type activity, and when the baseline and measurement error is taken into account these values retain approximately 4 % of the wild-type activity. However, in assays with the non-fluorogenic substrate, analysed using mass spectrometry, neither mutant displayed any cleavage after 60 minutes at 37 °C. An overestimation of activity from the fluorometer could be explained by RNase contamination in the sample, or if the sample had not yet reached 37 °C and the temperature was still increasing over the time course of the assay. Lack of activity in the mass spectrometry assay could be explained by the limits of detection of the method. Whether these mutants are slightly active or inhibit activity completely, the results still highlight the importance of these residues to the catalytic mechanism.

In summary, all four conserved acidic residues that define PIN-domains were intolerant to the mutations made, and therefore must play an essential role in the catalytic mechanism. It is not clear from these results what roles the residues play or whether the reaction proceeds via a one metal or two metal catalysis mechanism.

4.1.2 Crystal Structure

4.1.2.1 Comparison to Other PIN-domain Proteins

Structures Containing Metal Ions - VapC$_{PAE0151}$ and VapC-3$_{Mtb}$

Alignment of VapC$_{PAE2754}$ to VapC-3$_{Mtb}$ shows more similar coordination of the metals than alignment with VapC$_{PAE0151}$ (Figures 1.19 and 4.3). In the VapC$_{PAE0151}$ structure the Mn^{2+} ion is coordinated to one oxygen of Asp$^{(3)}$ (2.2 Å) and two water molecules (2.4 Å and 2.5 Å). In VapC-3, the Mg^{2+} ion is coordinated by both oxygens of Asp$^{(2)}$ (3.6 Å and 2.3 Å) and Asp$^{(3)}$ (2.3 Å and 3.3 Å), and at least one water (2.2 Å). Coordination to the equivalent active site residues is seen for the Cd^{2+} ions in VapC$_{PAE2754}$ (Chain A - 3.4 Å, 2.2 Å, 2.3 Å, 3.4 Å, Chain D - 3.6 Å, 2.4 Å, 2.4 Å, 3.8 Å. The RMS values for alignment of the monomers with the VapC-3$_{Mtb}$ structure are 2.99 and 3.05 for Chain A and Chain D respectively (Figure 4.3).

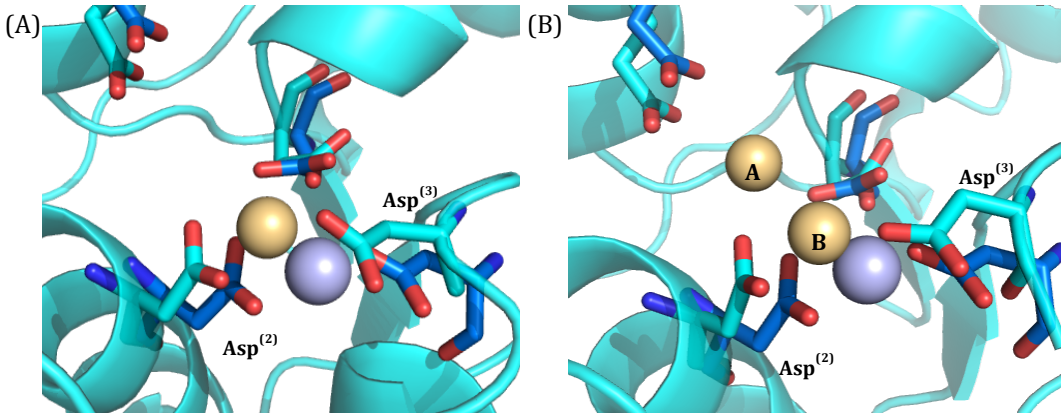


Figure 4.3. Alignment of VapC$_{PAE2754}$ with VapC-3$_{Mtb}$. A) Alignment of Chain A and VapC-$_{3Mtb}$. **B)** Alignment of Chain D and VapC-$_{3Mtb}$.In both figures VapC$_{PAE2754}$ is shown in cyan, conserved acidic residues are shown as sticks (blue for VapC-$_{3Mtb}$). Cd^{2+} ions are shown as pale yellow spheres and Mn^{2+} are shown as purple spheres.

Structures Containing Antitoxin - FitB and VapC$_{SFL}$

These structures were both solved with the antitoxin binding into the active site. In both enzymes, the first and second conserved Asp, and the conserved Glu interact with an Arg residue from the antitoxin. Metal B in VapC$_{PAE2754}$ binds to the second and third Asp. If the position of the metal is conserved,

possibly only one of the residues coordinating the metal needs to be sequestered to inhibit the binding of Mg^{2+} to the active site.

4.1.2.2 Comparison to Known RNase Mechanisms

From mutations made to the conserved acidic residues, and crystal structures containing Mg^{2+} ions and substrate, it was concluded that RNase H catalysis proceeds via a two metal mechanism. Structures resembling the intermediate state of the reaction mechanism were solved using a nicked substrate and vanadate ions. In the 'intermediate' active site, the metal ions were closer together. It was concluded from this that the metals may need to be less than 4 Å apart for catalysis to occur (Nowotny & Yang, 2006). In the proposed RNase H mechanism, metal B is required to destabilize the enzyme-substrate complex, and metal A is required for nucleophile activation and substrate release (Figure 4.4). It remains unclear whether one of the conserved acidic residues participates directly in catalysis as a base. The Cd^{2+} ions in $VapC_{PAE2754}$ are 4 Å apart therefore the second Cd^{2+} is in a position amenable to a two metal mechanism of catalysis.

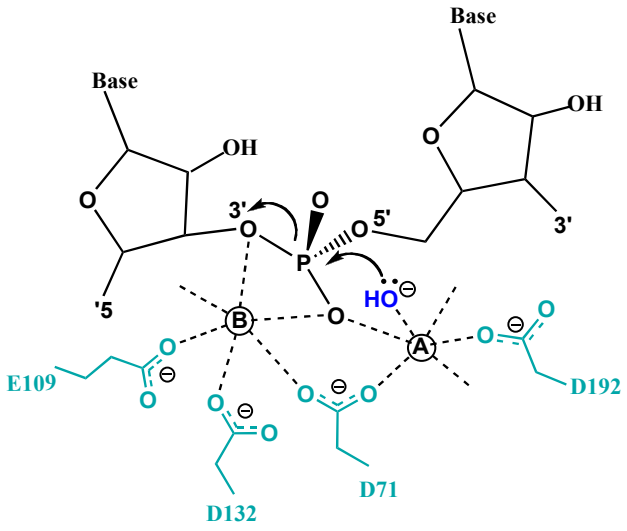


Figure 4.4. RNase H two metal mechanism of catalysis. The RNA substrate is shown in black and the conserved acidic residues are shown in cyan and labeled (Bha-RNase H1). The metals are shown as labeled circles with coordination shown as black dotted lines. The nucleophile is shown in blue and the movement of electrons is represented by arrows. (Nakamura et al., 1991; Tadokoro & Kanaya, 2009).

Two metal ions have also been shown in FLAP endonucleases, however whether the second is involved directly in catalysis is unknown. (Devos, et al., 2007; Feng, et al., 2004; Tock, et al., 2003). RNase H and FEN have more conserved acidic residues that make up the active site. The residues that bind the second metal in FEN are not present in PIN-domains, providing support for a single metal

mechanism (Anantharaman & Aravind, 2006). Although one metal mechanisms have been associated with decreased sequence specificity and a conserved histidine, a one metal mechanism has been proposed for EcoRI endonuclease. This enzyme contains an acidic active site, confirming that not all one metal mechanisms require a histidine, or have reduced specificity (Dupureur, 2010; Yang, 2011). If the reaction proceeds via a single metal mechanism then another mechanism for achieving specificity must exist.

The second metal ion bound in Chain D of $VapC_{PAE2754}$ does not conclusively indicate that catalysis is occurring via a two metal mechanism. It is possible that the binding of this metal is related to the unstructured region seen in Chain D, and without substrate bound it is difficult to confirm catalytically relevant binding (Bhagwat, et al., 1997; Yang, 2011). When compared to the chains A, B and C, the second Cd^{2+} ion occupies the position of the acetate molecule. This indicates that the placement of this metal is an anomaly and not a true representation of the arrangement of the active site.

4.1.3 Substrate and Sequence Specificity

How PIN-domains achieve specificity for sequence and ssRNA remains unclear. In RNase H, the RNA and DNA strands of the helix fit into different grooves, and residues recognise the 2'OH of four consecutive bases (Cerritelli & Crouch, 2009). In RNase II, and a homologue rrp44, the specificity is not for the scissile bond but for the surrounding residues. Catalysis requires 2-5 residues upstream of the cleavage site to be RNA (not DNA) for cleavage to occur (Cerritelli & Crouch, 2009; Frazao, et al., 2006; Gan, et al., 2006). Furthermore, the entrance to the binding groove is selective for ssRNA through steric hindrance. SMG-6 is inhibited by ssDNA, suggesting that the substrate can bind but that the 2'OH is required for catalysis (Glavan, et al., 2006).

The tetramer formed by $VapC_{PAE2754}$ (Figure 1.7) has a groove in the middle through which RNA could thread, and there are two tyrosine residues that stick out the side of each monomer to form a dimer interaction, potentially orientating the RNA ring as shown in the substrate docked in $VapC_{PAE0151}$

(Figure 1.15). Though the acetate bound in Chains A-C represents the likely position of the pentacovalent phosphate intermediate, a soaked substrate is required to determine what interactions are made between the RNA and the VapC residues.

4.1.4 Implications for the Mechanism

Based on the information above, the mechanism proposed for RNase catalysis by PIN-domains has been modified. In light of the debate around the involvement of the metal ions, a scheme has been proposed for both one and two metal catalysis (Figure 4.5).

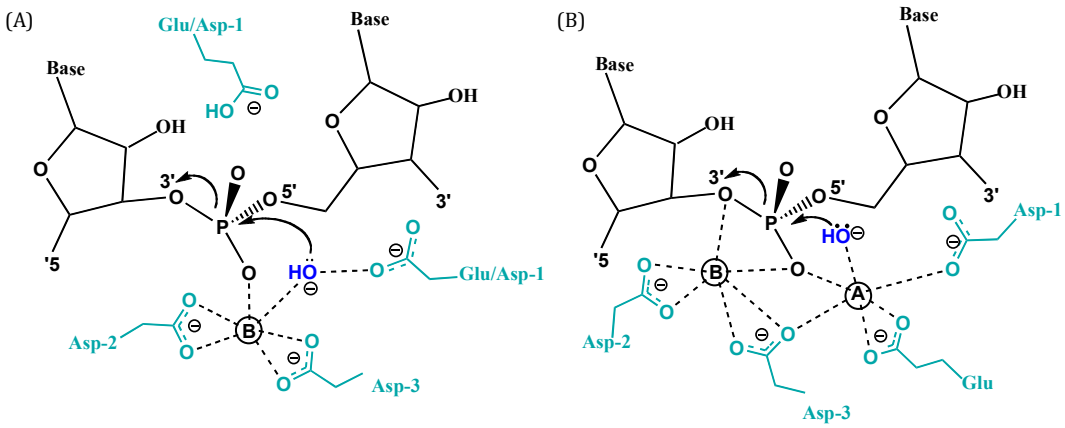


Figure 4.5. Proposed mechanisms of catalysis in PIN-domains. A) Two-metal mechanism. **B)** One metal mechanism. RNA substrate is shown in black, metal ion (labeled black circle) coordination is shown as dotted lines, conserved active site residues are shown in cyan and nucleophile is shown in blue with electron movement represented by arrows.

4.1.4.1 One Metal Mechanism

This mechanism has been proposed based on the evidence from the $VapC_{PAE2754}$ structure binding a single Cd^{2+} ion and an acetate molecule (Figure 3.20). The role of the single metal ion is to activate the nucleophilc water for attack. The nucleophile is also coordinated by $Asp^{(1)}$. $Asp^{(2)}$ and $Asp^{(3)}$ bind to the metal ion, which also coordinates an oxygen from the phosphate. Although a water is not seen in this position in the $Vap_{PAE2754}$ structure, this could be because Cd^{2+} is slightly larger than Mg^{2+}, and has less strict demand for octahedral coordination. D-N mutations in residues coordinating a metal involved in catalysis inhibit activity, as seen for D92N

and D110N (Bhagwat, et al., 1997; Kanaya, et al., 1990; Nakamura, et al., 1991). The slight activity seen in the D92N mutant may indicate that D110 is more important for coordination of the metal than D92N. In RNase H, it is uncertain if one of the residues is acting as the catalytic base. In $VapC_{PAE2754}$ this role may be played by E38. Mutant E38Q showed only very minimal activity, indicating that it is likely to be directly involved in catalysis. A proton could still be sequestered from somewhere else, but this would significantly reduce the rate of catalysis.

It is also important to consider that the four conserved acidic residues potentially interact with the substrate in a manner that allows specific bases of ssRNA to be recognised.

4.1.4.2 Two Metal Mechanism

This mechanism is based on the RNase H mechanism, and the assumption that a potential metal could be found deeper within the active site, coordinated by residue $Asp^{(1)}$. Residues $Asp^{(2)}$ and $Asp^{(3)}$ are still involved in binding metal B, but the role of activating the nucleophile is fulfilled by metal A. The conserved Glu residue also binds metal A, and possibly interacts with the substrate. The Glu residue could also still fulfill the role of the catalytic base. The previously mentioned properties of Glu compared to Asp, as well as the lack of bound substrate could justify why second metal ions are not seen in crystal structures.

4.2 Future Research

4.2.1 Structural Studies

A PIN-domain structure containing bound substrate and Mg^{2+} has yet to be determined. Binding of RNA substrate, or a transition state analogue such as vanadate, would show interactions of the amino acid residues and provide more information about how specificity is achieved. Crystallisation of the mutant proteins would provide more information about the changes occurring in the active site.

4.2.2 Mechanism

Elucidation of the mechanism will require knowledge of whether PIN-domains function by a one or two metal mechanism, as well as further analysis of the conserved active site residues to confirm whether they play a role as the catalytic acid or base. To determine how PIN-domains achieve their specificity for sequence and substrate, residues surrounding the active site and in the groove formed by the tetramer need to be identified. Mutations made to aromatic residues, such as the tyrosines between the dimer will confirm whether or not they are involved in locking the RNA ribose ring into place. In addition, it may be possible to find corresponding residues to those that hydrogen bond the 2'OHs of ribose rings surrounding the scissile bond in other RNases.

4.2.3 Transition State Analogue Inhibition

Current evidence implies that TA modules may play a role in the formation of persister colonies, or the regulation of metabolic processes. If VapC proteins are contributing to the ability of M. tuberculosis to survive the stresses encountered within the host, then inhibition of their action should enhance the ability of current antibiotics to eliminate infection, and reduce the number of recurrent cases. Effective inhibition of an enzyme requires an inhibitor that has higher binding affinity for the protein over any other enzymes that may be encountered.

Enzyme kinetics are such that the binding affinity toward the transition state is high, distorting the substrate into an optimal position for catalysis, and usually involving conformational change in the active site of the enzyme. If catalysis does not occur, the enzyme and substrate will remain as a tightly bound complex. This theory is applied to the development of inhibitors. Therefore, high affinity binding can be achieved without designing, synthesizing and testing a large range of potential molecules. Essentially, the dynamic potential involved in binding is transformed into thermodynamic energy forming enzyme-substrate complexes with long life times (Schramm, 2007; Schwartz & Schramm, 2009).

The ability to mimic a transition state analogue is dependent upon the geometry of the molecule, and the electrostatic charges. Therefore, by modeling molecular electrostatic potential surfaces (MEPS) of transition state analogues, chemical synthesis can be directed to form a tightly bound complex. This theory was used to develop an inhibitor for bovine purine nucleoside phosphorylase, which had a binding affinity (measured by K_M/K_d ratios) of 2400000 times the substrate (Schwartz & Schramm, 2009).

Knowledge of the transition state for VapC could allow synthesis of a transition state analogue that would bind to the active site with very high affinity, out competing the native substrate and being selective for VapC over any other enzymes.

Despite being unable to confirm the hypothesised mechanism, it may still be possible to design a transition state analogue based upon the knowledge of the target substrate (RNA), and the transition state of other RNase enzymes that produce similar cleavage products (i.e 3-OH and 5-phosphate).

4.3 Conclusion

The four conserved acidic residues that define PIN-domain proteins are essential for RNase activity in $VapC_{PAE2754}$. The reaction proceeds via a one metal, or two metal catalysis mechanism. The first metal (B) is coordinated by the second and third conserved aspartate residues, and is mostly likely involved in activation of the nucleophilic water.

Support for both one and two metal catalysis mechanisms is found in the data. The residues that coordinate the second metal in RNase and FEN are not present in PIN-domains, and the structural results show a single metal ion bound in all but Chain D of $VapC_{PAE2754}$. No magnesium has yet been shown in this second site. On the other hand, two metal mechanisms are associated with increased specificity, and no PIN-domain structure has been solved with substrate bound to confirm the position of the metal ions and the interactions of the conserved acidic residues. An acetate bound in the active site supported the placement of the docked dinucleotide substrate. This provided more accurate information on the orientation of the conserved active site residues and the metal ion in relation to the scissile bond.

Future research to obtain crystal structures will help to elucidate the mechanism and residues involved in sequence and substrate specificity. Mutagenesis of residues likely to be coordinating the RNA substrate will also provide vital information. Ultimately the mechanism would allow design of a transition state analogue inhibiter of VapC toxins. Disrupting the RNase activity of the toxin will knockout the ability of the TA operons to cleave their targets. This would in theory hinder the ability of the bacteria to regulate growth in response to the environmental stresses, or to form persistent colonies that can survive antibiotic attack. Knocking out these functions in *M. tuberculosis* may disrupt the ability of the bacteria to survive as latent infections, and increase the effectiveness of current antibiotic regimes, working towards the goal of decreasing the occurrence of both active and latent forms of tuberculosis.

Appendices

Appendix A: Reagents

A.1 Primers, Plasmids and Bacterial Strains Used

Primer	Sequence
$VapC_{PAE2754}$ Nco1 Fwd	ATATCCATGGCCCGTTGAGTACCTAGTGG
$VapC_{PAE2754}$ HindIII Rev	TGGTAAGGCTGGCGGCACAATAAG
D8A a23c Fwd	GAGTACCTAGTGGCCGCCTCCGCGCTA
D8A a23c #2 Fwd	CGTTGAGTACCTAGTGGCCGCCTCCGCGCTATACGCCC
D8A a23c Rev	TAGCGCGGAGGCGGCCACTAGGTACTC
D8A a23c #2 Rev	GGGCGTATAGCGCGGAGGCGGCCACTAGGTACTCAACG
D8N g22a fwd	GTTGAGTACCTAGTGAACGCCTCCGCGCTAT
D8N g22a rev	ATAGCGCGGAGGCGTTCACTAGGTACTCAAC
D8N g22a #2 fwd	GCCCGTTGAGTACCTAGTGGACGCCTCCGCGCTATACGC
D8N g22a #2 rev	GCGTATAGCGCGGAGGCGTCCACTAGGTACTCAACGGGC
S10A t28g Fwd	CCTAGTGGACGCCGCCGCGCTATACGC
S10A t28g Rev	GCGTATAGCGCGGCGGCGTCCACTAGG
E38Q g112c Fwd	TCTGCACTTGACCATATACCAGGCAGGCAACGC
E38Q Fwd-short	TCTGCACTTGACCATATACCAGGCAG
E38Q g112c Rev	GCGTTGCCTGCCTGGTATATGGTCAAGTGCAGA
E38H g112c g114c Fwd	TCTGCACTTGACCATATACCACGCAGGCAACGCGT
E38H g112c g114c Rev	ACGCGTTGCCTGCGTGGTATATGGTCAAGTGCAGA
D92N g274a Fwd	GGGGCTTGACCTTCTACAACGCCAGCTACG
D92N g274a Rev	CGTAGCTGGCGTTGTAGAAGGTCAAGCCCC
T108A a322g Fwd	GACTAGTCTTGGTGGCGCAAGACCGCGAG
T108A a322g Rev	CTCGCGGTCTTGCGCCACCAAGACTAGTC
T108A a322g #2 Fwd	CCGGACTAGTCTTGGTGGCGCAAGACCGCGAGCTACTGG
T108A a322g #2 Rev	CCAGTAGCTCGCGGTCTTGCGCCACCAAGACTAGTCCGG
D110A a329c Fwd	CTTGGTGACGCAAGCCCGCGAGCTACTGG
D110A a329c Rev	CCAGTAGCTCGCGGGCTTGCGTCACCAAG
D110A a329c #2 Fwd	CTAGTCTTGGTGACGCAAGCCCGCGAGCTACTGGCCAAG
D110A a329c #2 Rev	CTTGGCCAGTAGCTCGCGGGCTTGCGTCACCAAGACTAG
D110N g328a Fwd	GGACTAGTCTTGGTGACGCAAAACCGCGAGCTA
D110N g328a Rev	TAGCTCGCGGTTTTGCGTCACCAAGACTAGTCC
D110N g328a #2 Fwd	CCGGACTAGTCTTGGTGACGCAAAACCGCGAGCTACTGGC
D110N g328a #2 Rev	GCCAGTAGCTCGCGGTTTTGCGTCACCAAGACTAGTCCGG

Table A. 1. List of Primers used in the study

Substrate Name	Sequence
Oligo1 DNA seq GGGG	dTdAdAdGaTdCGGGGdAdCdAdTdCdAdG
Oligo1 DNA seq GGUG	dTdAdAdGaTdCGGUGdAdCdAdTdCdAdG
Oligo2 RNA triA GGGG	AAAGGGGAAA
Oligo2 RNA triA GGUG	AAAGGUGAAA
Oligo3 RNA seq GGGG	UAAUUCGGGGACUAUG
Oligo3 RNA seq GGUG	UAAUUCGGUGACUAUG
Oligo4 RNA rpt GGGG	TAATTCGGGGACTATG
Oligo 4 RNA rpt GGUG	TAATTCGGUGACTATG
GGUG Oligo A	AUCUCAGGUGACAUAG
GGUG Oligo B	UCUAACGGUGUCAUCA
GGUG Oligo C	CUAAUCGGUGACUACA
922 Pentaprobe	CGGAATTCTACGAATTTTTCTTTTGTTTATTTCCTTTCGCTTTGCTT CTCTTCCCTTCGGTTCTGTTCCGTTTTACCTTGTCTTGCCTTATCTT ACTTTA
924 Pentaprobe	CGCTCTATTCTACTGTCCTGTGCATTCAATCGTTGAGTTCGATCTAG TCTCGTCTAACCCTCCCCTGCTCCGCTGGTCTGGCCTCGCCTATCCT ACCCAT
932 Oligo5	GUGUUGUACUCUCCUCUGCGUUCACUUAGC

Table A. 2. Sequence of RNA substrates for fluorogenic assays and mutant analysis

Strain	Description
Escherichia coli	
DH5α	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80*
BL21 (DE3)	F- *ompT hsdS$_B$* (r_B-m_B-) *gal dcm* (DE3)
Plasmids	
pProEX-Hta/c	*E. coli* expression vector encoding a N-

Table A. 3. List of bacterial strains and plasmids used in the study

A.2 Buffers and Solutions

5 x DNA Loading Dye	0.05% (*w/v*) bromophenol blue, 0.25% (*w/v*) xylene cyanol, 30% (*v/v*) glycerol
Fairbanks Staining Solution A	0.05% (*w/v*) coomassie blue R-250, 25% (*v/v*) isopropanol, 10% (*v/v*) acetic acid
Fairbanks Staining Solution B	0.005% (*w/v*) coomassie blue R-250 10% (*v/v*) isopropanol, 10% (*v/v*) acetic acid
Fairbanks Staining Solution C	0.002% (*w/v*) coomassie blue,
10% (*v/v*) acetic acid	
Fairbanks Staining Solution D	10% (*v/v*) acetic acid
Formamide Loading Buffer	0% (*v/v*) formamide, 5 mM EDTA, 0.1% (*w/v*) bromophenol blue, 0.1% (*w/v*) xylene cyanol FF
Native Gel Loading Buffer	0.3 M Tris-HCl, 1% (*w/v*) bromophenol
blue pH 6.8	
Resolving Buffer	1.5 M Tris-HCl pH 8.8
4 x SDS loading buffer	200 mM Tris-HCl pH 6.8, 8% (*w/v*) SDS, 40% (*v/v*) glycerol, 0.4% (*w/v*) bromophenol blue, 400 mM β-mercaptoethanol
SDS running buffer	25 mM Tris-HCl pH 6.8, 0.1% (*w/v*) SDS 190 mM glycine
Stacking Buffer	0.5 M Tris-HCl pH 6.8
10 x TAE	400 mM Tris-acetate, 20 mM EDTA
1 x TAE	100 ml 10 x TAE + 900 ml H_2O
10 x TBE	0.89 M Tris-HCl, 0.89 M boric acid, 20mM EDTA
1 x TBE	100 ml 10 x TBE + 900 ml H_2O
10 x TBS	200 mM Tris-HC pH 7.6, 1.5 M NaCl
1 x TBS	100 ml 10 x TBS + 899 ml H_2O, 0.05% (*v/v*) tween 20
TE	10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0

A.3 Growth Media

A.3.1 Solid Media

LB-agar 1% (*w/v*) bactotryptone, 0.5% (*w/v*) yeast extract, 1% (*w/v*) NaCl, 15 $g.L^{-1}$ agar pH 8.0

4.3.1.1 A.3.2 Liquid Media

LB 1% (*w/v*) bactotryptone, 0.5%(*w/v*) yeast extract, 1% (*w/v*) NaCl pH 8.0

SOC 2% (*w/v*) bactotryptone or bactopeptone, 0.55% (*w/v*) yeast extract, 10 mM NaCl, 2.5 mM KCl 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose.

Appendix B: Gene and Protein Information

B.1 Pyrobaculum Protein Information

VapC$_{PAE0151}$ Protein Sequence With pProEX fusion tag

MSYYHHHHHHDYDIPTTENLYFQGAMKLVVDASAIAALYVPEERSEQAERAVSQAQELHT
LDLAAYEVANDLWKHARRGLLREDEASNMLEELWEFFKALKVHSYAEVLKDAFALALKHG
VTVYDAAYVALAEKIGGKLLTLDRQLAEKFPALVTP

156 amino acids
Molecular weight: 17721.09
Theoretical pI: 5.50

VapC$_{PAE2754}$ Sequence with pProEX fusion tag

MSYYHHHHHHDYDIPTTENLYFQGAMAVEYLVDASALYALAAHYDKWIKHREKLAILHLT
IYEAGNALWKEARLGRVDWAAASRHLKKVLSSFKVLEDPPLDEVLRVAVERGLTFYDASY
AYVAESSGLVLVTQDRELLAKTKGAIDVETLLVRLAAQ

158 amino acids
Molecular weight 17979.52
Theoretical pI: 6.18

B.2 Pentaprobe and Flanking Sequences

Sequences represent the RNA sequence that will be transcribed from the T7 promoter (excludes the T7 promoter). Yellow highlighting corresponds to the pentaprobe sequence (McKenzie, 2011).

922 + Flanking Sequences:

AGACCCAAGCTTGGTACCGGAATTCTACGAATTTTTCTTTTGTTTATTTCCTTTC
GCTTTGCTTCTCTTCCCTTCGGTTCTGTTCCGTTTTACCTTGTCTTGCCTTATCT
TACTTTATCTAGAGGGCCCTATTCTATAGTGTCACCTAAATGCTAGAGCTCGCT

924 + Flanking Sequences:

AGACCCAAGCTTGGTACCGCTCTATTCTACTGTCCTGTGCATTCAATCGTTGAGT
TCGATCTAGTCTCGTCTAACCCTCCCCTGCTCCGCTGGTCTGGCCTCGCCTATCC
TACCCATTCTAGAGGGCCCTATTCTATAGTGTCACCTAAATGCTAGAGCTCGCT

Appendix C: Raw Data

C.1 Fluorometric Assays

C.1.1 Calibration Data

[S] μM	Fluorescence (518)		
0.1	79.7613	131.388	95.38383
0.25	-	-	225.8391
0.3	407.4862	263.7419	-
0.5	450.0719	546.4808	437.5441
0.6	640.7263	654.813	-
0.8	777.3746	864.8207	-
1	785.7756	1070.303	824.1489

Table A. 4. Standard Calibration Measurements. Substrate concentration is of each of the two halves added (Fl. and Q.). Blank squares indicate where duplicates were not measured.

C.1.2 Wild-type VapC$_{PAE2754}$

[S] μM	RFU/min	Velocity μM/min
0.25	37.627	0.04807334
0.5	65.436	0.08360291
0.75	87.2862	0.1115194
1	112.294	0.14347
1.5	118.702	0.1516571
2	183.989	0.2350696
2.5	203.3	0.2597419
5	225.408	0.2879877
7.5	192.9	0.2464559

Table A. 5. VapC$_{PAE2754}$ Fluorogenic assay. RFU/min given by the by as activity was converted into velocity.

C.1.3 Mutant Proteins

Mutant	Rate (RFU/min)			Mean	% WT	SD
E38Q	12.19	9.96	12.5	11.55	10.62	1.13
D92N	10.05	11.9	10.39	10.78	9.91	0.80
S10A	26.25	30.57	33.57	30.13	27.71	3.00
E38H	7.91	6.77	6.47	7.05	6.48	0.62
D8N	6.34	5.14	8.28	6.59	6.06	1.29
D110A	4.09	4.98	4.28	4.45	4.09	0.38
D110N	9.32	4.2	5.45	6.32	5.81	2.18
D8A	7.27	6.07		6.67	6.13	0.6
T108A	83.34	90.69		87.02	80.01	3.68

Table A. 6. Statistical analysis of mutant protein. Rate was measured in duplicate or triplicate. The mean was calculated and used to determine a percentage of the wild-type activity (% WT). Standard deviation (SD) were calculated.

C.1.4 Blanks

Blanks at 5x K_M	Statistical Analysis
6.63	outlier
5.231	not outlier
4.8	not outlier
4.7817	not outlier
4.3	not outlier
4.23	not outlier
3.0519	not outlier
2.7	not outlier
2.6957	not outlier
2.6612	not outlier
2.4245	not outlier
2.34	not outlier
2.3078	not outlier
1.8279	not outlier
1.8	not outlier
1.74	not outlier
1.68	not outlier
1.63	not outlier
1.6	not outlier
1.2	not outlier
2.981585	**Mean**
1.458279306	**SD**
2.789563158	**Mean - outliers**
5.70612177	**Mean + 2x SD**
2.565115547	**Mean % of WT**
5.247008524	**Mean + 2xSD % of WT**

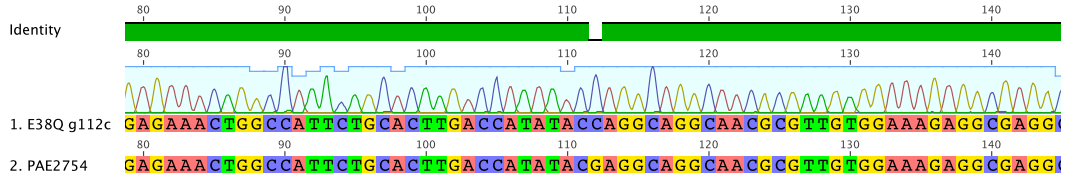
Table A. 7. Statistical Analysis of Blank Measurements at 5x K_M. Data was collected for all mutants and in duplicate or triplicate where possible. The values that were greater than 2 standard deviations (SD) from the mean were rejected as outliers and the mean was recalculated. The mean and the mean + 2xSD were calculated as percentages of the WT activity.

C.2 Sequencing Results

Mutant	Protein Sequence and Gene Sequence Alignment

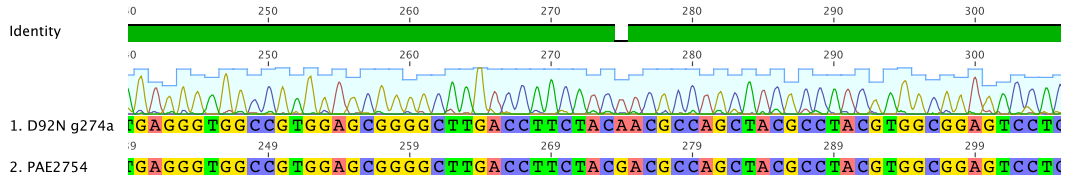
E38Q

```
1         10        20        30        40        50
|         |         |         |         |         |
MAVEYLVDASALYALAAHYDKWIKHREKLAILHLTIYQAGNALWKEARLG
RVDWAAASRHLKKVLSSFKVLEDPPLDEVLRVAVERGLTFYDASYAYVAE
SSGLVLVTQDRELLAKTKGAIDVETLLVRLAAQ
```



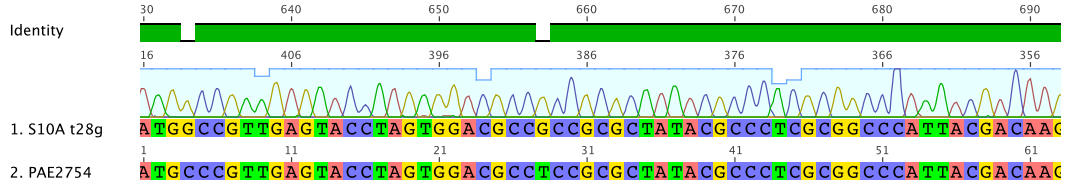
D92N

```
1         10        20        30        40        50
|         |         |         |         |         |
MAVEYLVDASALYALAAHYDKWIKHREKLAILHLTIYEAGNALWKEARLG
RVDWAAASRHLKKVLSSFKVLEDPPLDEVLRVAVERGLTFYNASYAYVAE
SSGLVLVTQDRELLAKTKGAIDVETLLVRLAAQ*
```



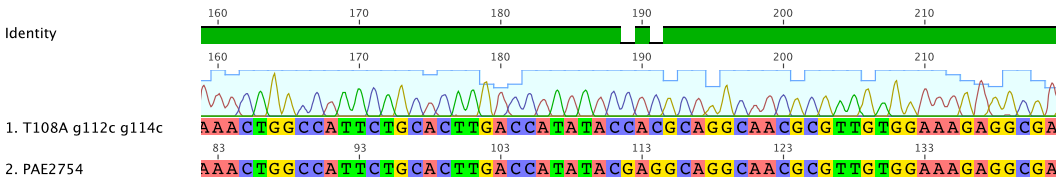
S10A

```
1         10        20        30        40        50
|         |         |         |         |         |
MAVEYLVDAAALYALAAHYDKWIKHREKLAILHLTIYEAGNALWKEARLG
RVDWAAASRHLKKVLSSFKVLEDPPLDEVLRVAVERGLTFYDASYAYVAE
SSGLVLVTQDRELLAKTKGAIDVETLLVRLAAQ*
```



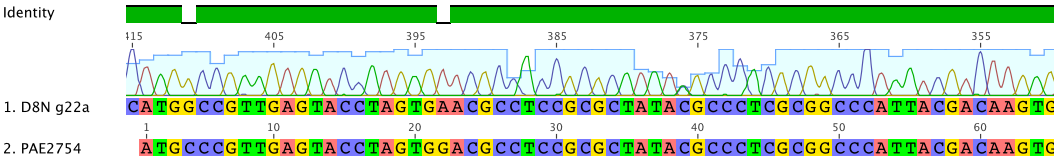
E38H

```
1         10        20        30        40        50
|         |         |         |         |         |
MAVEYLVDASALYALAAHYDKWIKHREKLAILHLTIYHAGNALWKEARLG
RVDWAAASRHLKKVLSSFKVLEDPPLDEVLRVAVERGLTFYDASYAYVAE
SSGLVLVTQDRELLAKTKGAIDVETLLVRLAAQ*
```



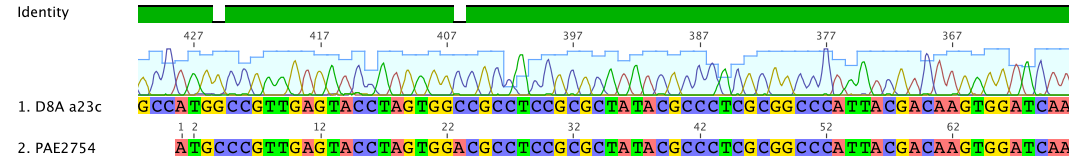
D8N

```
1         10        20        30        40        50
|         |         |         |         |         |
MAVEYLVNASALYALAAHYDKWIKHREKLAILHLTIYEAGNALWKEARLG
RVDWAAASRHLKKVLSSFKVLEDPPLDEVLRVAVERGLTFYDASYAYVAE
SSGLVLVTQDRELLAKTKGAIDVETLLVRLAAQ*
```



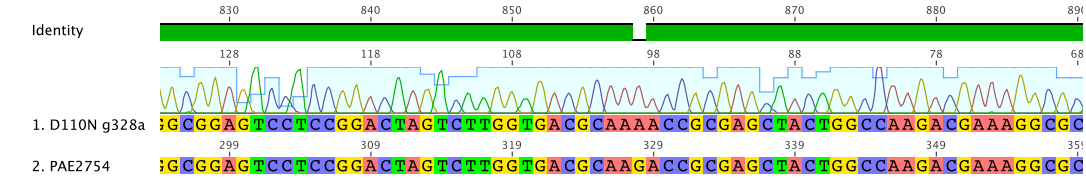
D8A

```
1         10        20        30        40        50
|         |         |         |         |         |
MAVEYLVAASALYALAAHYDKWIKHREKLAILHLTIYEAGNALWKEARLG
RVDWAAASRHLKKVLSSFKVLEDPPLDEVLRVAVERGLTFYDASYAYVAE
SSGLVLVTQDRELLAKTKGAIDVETLLVRLAAQ*
```



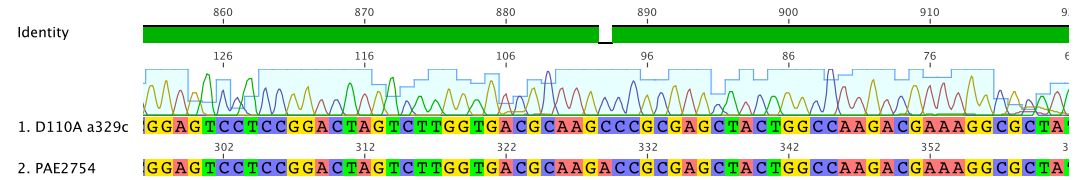
D110N

```
1         10        20        30        40        50
|         |         |         |         |         |
MAVEYLVDASALYALAAHYDKWIKHREKLAILHLTIYEAGNALWKEARLG
RVDWAAASRHLKKVLSSFKVLEDPPLDEVLRVAVERGLTFYDASYAYVAE
SSGLVLVTQNRELLAKTKGAIDVETLLVRLAAQ*
```



D110A

```
1         10        20        30        40        50
|         |         |         |         |         |
MAVEYLVDASALYALAAHYDKWIKHREKLAILHLTIYEAGNALWKEARLG
RVDWAAASRHLKKVLSSFKVLEDPPLDEVLRVAVERGLTFYDASYAYVAE
SSGLVLVTQARELLAKTKGAIDVETLLVRLAAQ*
```



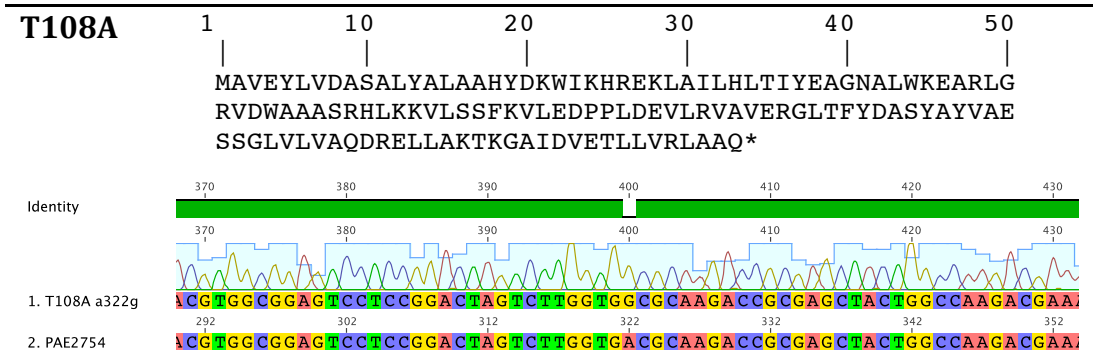


Table A. 8. Sequencing results for mutagenesis. Shown for each mutant: the amino acid sequence of the protein, labeled every 10 residues; a selection of the chromatogram results aligned with the wild-type $VapC_{PAE2754}$ gene sequence.

References

Afif, H., Allali, N., Couturier, M., & Van Melderen, L. (2001). The ratio between CcdA and CcdB modulates the transcriptional repression of the ccd poison-antidote system. *Molecular Microbiology, 41*(1), 73-82. doi: 10.1046/j.1365-2958.2001.02492.x

Agarwal, S., Mishra, N. K., Bhatnagar, S., & Bhatnagar, R. (2010). PemK Toxin of Bacillus anthracis Is a Ribonuclease AN INSIGHT INTO ITS ACTIVE SITE, STRUCTURE, AND FUNCTION. [Article]. *Journal of Biological Chemistry, 285*(10), 7254-7270. doi: 10.1074/jbc.M109.073387

Ahidjo, B. A., Kuhnert, D., McKenzie, J. L., Machowski, E. E., Gordhan, B. G., Arcus, V., . . . Mizrahi, V. (2011). VapC Toxins from Mycobacterium tuberculosis Are Ribonucleases that Differentially Inhibit Growth and Are Neutralized by Cognate VapB Antitoxins. [Article]. *Plos One, 6*(6).

Amitai, S., Kolodkin-Gal, I., Hananya-Meltabashi, M., Sacher, A., & Engelberg-Kulka, H. (2009). Escherichia coli MazF Leads to the Simultaneous Selective Synthesis of Both "Death Proteins'' and "Survival Proteins''. [Article]. *Plos Genetics, 5*(3).

Anantharaman, V., & Aravind, L. (2006). The NYN domains: Novel predicted RNAses with a PIN domain-like fold. *RNA Biology, 3*(1), 18-27.

Arcus, V. L., Backbro, K., Roos, A., Daniel, E. L., & Baker, E. N. (2004). Distant structural homology leads to the functional characterization of an archaeal PIN domain as an exonuclease. *Journal of Biological Chemistry, 279*(16), 16471-16478. doi: 10.1074/jbc.M313833200

Arcus, V. L., McKenzie, J. L., Robson, J., & Cook, G. M. (2011). The PIN-domain ribonucleases and the prokaryotic VapBC toxin-antitoxin array. [Review]. *Protein Engineering Design & Selection, 24*(1-2), 33-40. doi: 10.1093/protein/gzq081

Arcus, V. L., Rainey, P. B., & Turner, S. J. (2005). The PIN-domain toxin-antitoxin array in mycobacteria. *Trends in Microbiology, 13*(8), 360-365. doi: 10.1016/j.tim.2005.06.008

Audoly, G., Vincentelli, R., Edouard, S., Georgiades, K., Mediannikov, O., Gimenez, G., . . . Raoult, D. (2011). Effect of Rickettsial Toxin VapC on Its Eukaryotic Host. *Plos One, 6*(10).

Barbas, A., Matos, R. G., Amblar, M., Lopez-Vinas, E., Gomez-Puertas, P., & Arraiano, C. M. (2009). Determination of Key Residues for Catalysis and RNA Cleavage Specificity ONE MUTATION TURNS RNase II INTO A "SUPER-ENZYME". [Article]. *Journal of Biological Chemistry, 284*(31), 20486-20498. doi: 10.1074/jbc.M109.020693

Bhagwat, M., Meara, D., & Nossal, N. G. (1997). Identification of residues of T4 RNase H required for catalysis and DNA binding. *Journal of Biological Chemistry, 272*(45), 28531-28538.

Bodogai, M., Ferenczi, S., BashtoVyy, D., Miclea, P., Papp, P., & Dusha, I. (2006). The ntrPR operon of Sinorhimbium meliloti is organized and functions as a toxin-antitoxin module. *Molecular Plant-Microbe Interactions, 19*(7), 811-822. doi: 10.1094/mpmi-19-0811

Bunker, R. D., McKenzie, J. L., Baker, E. N., & Arcus, V. L. (2008). Crystal structure of PAE0151 from Pyrobaculum aerophilum, a PIN-domain (VapC) protein from a toxin-antitoxin operon. *Proteins-Structure Function and Bioinformatics, 72*(1), 510-518. doi: 10.1002/prot.22048

Buts, L., Lah, J., Dao-Thi, M. H., Wyns, L., & Loris, R. (2005). Toxin-antitoxin modules as bacterial metabolic stress managers. *Trends in Biochemical Sciences, 30*(12), 672-679.

Cerritelli, S. M., & Crouch, R. J. (2009). Ribonuclease H: the enzymes in eukaryotes. [Review]. *Febs Journal, 276*(6), 1494-1505. doi: 10.1111/j.1742-4658.2009.06908.x

Chim, N., Habel, J. E., Johnston, J. M., Krieger, I., Miallau, L., Sankaranarayanan, R., . . . Goulding, C. W. (2011). The TB Structural Genomics Consortium: A decade of progress. *Tuberculosis, 91*(2), 155-172. doi: 10.1016/j.tube.2010.11.009

Christensen-Dalsgaard, M., Jorgensen, M. G., & Gerdes, K. (2010). Three new RelE-homologous mRNA interferases of Escherichia coli differentially induced by environmental stresses. [Article]. *Molecular Microbiology, 75*(2), 333-348. doi: 10.1111/j.1365-2958.2009.06969.x

Christensen, S. K., & Gerdes, K. (2003). RelE toxins from Bacteria and Archaea cleave mRNAs on translating ribosomes, which are rescued by tmRNA. [Article]. *Molecular Microbiology, 48*(5), 1389-1400. doi: 10.1046/j.1365-2958.2003.03512.x

Christensen, S. K., Pedersen, K., Hansen, F. G., & Gerdes, K. (2003). Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA. *Journal of Molecular Biology, 332*(4), 809-819. doi: 10.1016/s0022-2836(03)00922-7

Condon, C. (2006). Shutdown decay of mRNA. *Molecular Microbiology, 61*(3), 573-583. doi: 10.1111/j.1365-2958.2006.05270.x

Connolly, L. E., Edelstein, P. H., & Ramakrishnan, L. (2007). Why Is Long-Term Therapy Required to Cure Tuberculosis? [doi:10.1371/journal.pmed.0040120]. *PLoS Med, 4*(3), e120.

Daines, D. A., Wu, M. H., & Yuan, S. Y. (2007). VapC-1 of nontypeable Haemophilus influenzae is a ribonuclease. *Journal of Bacteriology, 189*(14), 5041-5048. doi: 10.1128/jb.00290-07

Devos, J. M., Tomanicek, S. J., Jones, C. E., Nossal, N. G., & Mueser, T. C. (2007). Crystal structure of bacteriophage T4 5' nuclease in complex with a branched DNA reveals how flap endonuclease-1 family nucleases bind their substrates. [Article]. *Journal of Biological Chemistry, 282*(43), 31713-31724. doi: 10.1074/jbc.M703209200

Dienemann, C., Boggild, A., Winther, K. S., Gerdes, K., & Brodersen, D. E. (2011). Crystal Structure of the VapBC Toxin-Antitoxin Complex from Shigella flexneri Reveals a Hetero-Octameric DNA-Binding Assembly. *Journal of Molecular Biology, 414*(5), 713-722. doi: 10.1016/j.jmb.2011.10.024

Dupureur, C. M. (2010). One is enough: insights into the two-metal ion nuclease mechanism from global analysis and computational studies. [Review]. *Metallomics, 2*(9), 609-620. doi: 10.1039/c0mt00013b

Easter, A. D. (2010). *Decoupling Enzyme Catalysis from Thermal Denaturation.* Masters, University of Waikato, Hamilton.

Eberl, L., Givskov, M., & Schwab, H. (1992). THE DIVERGENT PROMOTERS MEDIATING TRANSCRIPTION OF THE PAR LOCUS OF PLASMID RP4 ARE SUBJECT TO AUTOREGULATION. *Molecular Microbiology, 6*(14), 1969-1979. doi: 10.1111/j.1365-2958.1992.tb01370.x

Engelberg-Kulka, H., Amitai, S., Kolodkin-Gal, I., & Hazan, R. (2006). Bacterial programmed cell death and multicellular behavior in bacteria. *Plos Genetics, 2*(10), 1518-1526.

Feng, M., Patel, D., Dervan, J. J., Ceska, T., Suck, D., Haq, I., & Sayers, J. R. (2004). Roles of divalent metal ions in flap endonuclease-substrate interactions. [Article]. *Nature Structural & Molecular Biology, 11*(5), 450-456. doi: 10.1038/nsmb754

Fiebig, A., Rojas, C. M. C., Siegal-Gaskins, D., & Crosson, S. (2010). Interaction specificity, toxicity and regulation of a paralogous set of ParE/RelE-family toxin-antitoxin systems. *Molecular Microbiology, 77*(1), 236-251. doi: 10.1111/j.1365-2958.2010.07207.x

Frazao, C., McVey, C. E., Amblar, M., Barbas, A., Vonrhein, C., Arraiano, C. M., & Carrondo, M. A. (2006). Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. [Article]. *Nature, 443*(7107), 110-114. doi: 10.1038/nature05080

Gan, J. H., Tropea, J. E., Austin, B. P., Court, D. L., Waugh, D. S., & Ji, X. H. (2006). Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. [Article]. *Cell, 124*(2), 355-366. doi: 10.1016/j.cell.2005.11.034

Garcia-Pino, A., Balasubramanian, S., Wyns, L., Gazit, E., De Greve, H., Magnuson, R. D., . . . Loris, R. (2010). Allostery and Intrinsic Disorder Mediate Transcription Regulation by Conditional Cooperativity. *Cell, 142*(1), 101-111. doi: 10.1016/j.cell.2010.05.039

Georgiades, K., & Raoult, D. (2011). Genomes of the Most Dangerous Epidemic Bacteria Have a Virulence Repertoire Characterized by Fewer Genes but More Toxin-Antitoxin Modules. [Article]. *Plos One, 6*(3).

Gerdes, K., Christensen, S. K., & Lobner-Olesen, A. (2005). Prokaryotic toxin-antitoxin stress response loci. [Review]. *Nature Reviews Microbiology, 3*(5), 371-382. doi: 10.1038/nrmicro1147

Gerdes, K., Rasmussen, P. B., & Molin, S. (1986). UNIQUE TYPE OF PLASMID MAINTENANCE FUNCTION - POSTSEGREGATIONAL KILLING OF PLASMID-FREE CELLS. [Article]. *Proceedings of the National Academy of Sciences of the United States of America, 83*(10), 3116-3120. doi: 10.1073/pnas.83.10.3116

Glavan, F., Behm-Ansmant, I., Izaurralde, E., & Conti, E. (2006). Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. [Article]. *Embo Journal, 25*(21), 5117-5125. doi: 10.1038/sj.emboj.7601377

Hayes, F. (2003). Toxins-antitoxins: Plasmid maintenance, programmed cell death, and cell cycle arrest. [Review]. *Science, 301*(5639), 1496-1499. doi: 10.1126/science.1088157

Jørgensen, M. G., Pandey, D. P., Jaskolska, M., & Gerdes, K. (2009). HicA of Escherichia coli defines a novel family of translation-independent mRNA interferases in bacteria and archaea. *Journal of Bacteriology, 191*(4), 1191-1199.

Kamada, K., & Hanaoka, F. (2005). Conformational change in the catalytic site of the ribonuclease YoeB toxin by YefM antitoxin. *Molecular Cell, 19*(4), 497-509. doi: 10.1016/j.molcel.2005.07.004

Kanaya, S., Kohara, A., Miura, Y., Sekiguchi, A., Iwai, S., Inoue, H., . . . Ikehara, M. (1990). IDENTIFICATION OF THE AMINO-ACID-RESIDUES INVOLVED IN AN ACTIVE-SITE OF ESCHERICHIA-COLI RIBONUCLEASE-H BY SITE-DIRECTED MUTAGENESIS. *Journal of Biological Chemistry, 265*(8), 4615-4621.

Keren, I., Minami, S., Rubin, E., & Lewis, K. (2011). Characterization and Transcriptome Analysis of Mycobacterium tuberculosis Persisters. [Article]. *Mbio, 2*(3).

Korch, S. B., & Hill, T. M. (2006). Ectopic overexpression of wild-type and mutant hipA genes in Escherichia coli: Effects on macromolecular synthesis and persister formation. *Journal of Bacteriology, 188*(11), 3826-3836. doi: 10.1128/jb.01740-05

Lewis, K. (2007). Persister cells, dormancy and infectious disease. *Nature Reviews Microbiology, 5*(1), 48-56. doi: 10.1038/nrmicro1557

Maezato, Y., Daugherty, A., Dana, K., Soo, E., Cooper, C., Tachdjian, S., . . . Blum, P. (2011). VapC6, a ribonucleolytic toxin regulates thermophilicity in the

crenarchaeote Sulfolobus solfataricus. [Article]. *Rna-a Publication of the Rna Society, 17*(7), 1381-1392. doi: 10.1261/rna.2679911

Magnuson, R. D. (2007). Hypothetical functions of toxin-antitoxin systems. *Journal of Bacteriology, 189*(17), 6089-6092. doi: 10.1128/jb.00958-07

Maisonneuve, E., Shakespeare, L. J., Jorgensen, M. G., & Gerdes, K. (2011). Bacterial persistence by RNA endonucleases. [Article]. *Proceedings of the National Academy of Sciences of the United States of America, 108*(32), 13206-13211. doi: 10.1073/pnas.1100186108

Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2009). Comprehensive comparative-genomic analysis of Type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. [Review]. *Biology Direct, 4*.

Mattison, K., Wilbur, J. S., So, M., & Brennan, R. G. (2006). Structure of FitAB from Neisseria gonorrhoeae bound to DNA reveals a tetramer of toxin-antitoxin heterodimers containing pin domains and ribbon-helix-helix motifs. *Journal of Biological Chemistry, 281*(49), 37942-37951. doi: 10.1074/jbc.M605198200

McKenzie, J. L. (2011). *The Biochemistry of VapBC Toxin-Antitoxins.* PhD, University of Waikato.

Miallau, L., Faller, M., Chiang, J., Arbing, M., Guo, F., Cascio, D., & Eisenberg, D. (2009). Structure and Proposed Activity of a Member of the VapBC Family of Toxin-Antitoxin Systems VapBC-5 FROM MYCOBACTERIUM TUBERCULOSIS. *Journal of Biological Chemistry, 284*(1), 276-283. doi: 10.1074/jbc.M805061200

Mossakowska, D. E., Nyberg, K., & Fersht, A. R. (1989). KINETIC CHARACTERIZATION OF THE RECOMBINANT RIBONUCLEASE FROM BACILLUS-AMYLOLIQUEFACIENS (BARNASE) AND INVESTIGATION OF KEY RESIDUES IN CATALYSIS BY SITE-DIRECTED MUTAGENESIS. [Article]. *Biochemistry, 28*(9), 3843-3850.

Moyed, H. S., & Bertrand, K. P. (1983). HIPA, A NEWLY RECOGNIZED GENE OF ESCHERICHIA-COLI K-12 THAT AFFECTS FREQUENCY OF PERSISTENCE AFTER INHIBITION OF MUREIN SYNTHESIS. *Journal of Bacteriology, 155*(2), 768-775.

Nakamura, H., Oda, Y., Iwai, S., Inoue, H., Ohtsuka, E., Kanaya, S., . . . Ikehara, M. (1991). How Does RNase H Recognize a DNAÂ·RNA Hybrid? *Proceedings of the National Academy of Sciences of the United States of America, 88*(24), 11535-11539.

Nowotny, M., Gaidamakov, S. A., Crouch, R. J., & Yang, W. (2005). Crystal structures of RNase H bound to an RNA/DNA hybrid: Substrate specificity and metal-dependent catalysis. [Article]. *Cell, 121*(7), 1005-1016. doi: 10.1016/j.cell.2005.04.024

Nowotny, M., & Yang, W. (2006). Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. *Embo Journal, 25*(9), 1924-1933. doi: 10.1038/sj.emboj.7601076

Oda, Y., Yoshida, M., & Kanaya, S. (1993). ROLE OF HISTIDINE-124 IN THE CATALYTIC FUNCTION OF RIBONUCLEASE-III FROM ESCHERICHIA-COLI. [Article]. *Journal of Biological Chemistry, 268*(1), 88-92.

Overgaard, M., Borch, J., & Gerdes, K. (2009). RelB and RelE of Escherichia coli Form a Tight Complex That Represses Transcription via the Ribbon-Helix-Helix Motif in RelB. *Journal of Molecular Biology, 394*(2), 183-196. doi: 10.1016/j.jmb.2009.09.006

Overgaard, M., Borch, J., Jorgensen, M. G., & Gerdes, K. (2008). Messenger RNA interferase RelE controls relBE transcription by conditional cooperativity. *Molecular Microbiology, 69*(4), 841-857. doi: 10.1111/j.1365-2958.2008.06313.x

Pandey, D. P., & Gerdes, K. (2005). Toxin–antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Research, 33*(3), 966-976. doi: 10.1093/nar/gki201

Pedersen, K., Christensen, S. K., & Gerdes, K. (2002). Rapid induction and reversal of a bacteriostatic condition by controlled expression of toxins and antitoxins. *Molecular Microbiology, 45*(2), 501-510. doi: 10.1046/j.1365-2958.2002.03027.x

Pedersen, K., Zavialov, A. V., Pavlov, M. Y., Elf, J., Gerdes, K., & Ehrenberg, M. (2003). The Bacterial Toxin RelE Displays Codon-Specific Cleavage of mRNAs in the Ribosomal A Site. *Cell, 112*(1), 131-140.

Puskas, L. G., Nagy, Z. B., Kelemen, J. Z., Ruberg, S., Bodogai, M., Becker, A., & Dusha, I. (2004). Wide-range transcriptional modulating effect of ntrR under microaerobiosis in Sinorhizobium meliloti. *Molecular Genetics and Genomics, 272*(3), 275-289. doi: 10.1007/s00438-004-1051-3

Ramage, H. R., Connolly, L. E., & Cox, J. S. (2009). Comprehensive Functional Analysis of Mycobacterium tuberculosis Toxin-Antitoxin Systems: Implications for Pathogenesis, Stress Responses, and Evolution. *Plos Genetics, 5*(12).

Robson, J., McKenzie, J. L., Cursons, R., Cook, G. M., & Arcus, V. L. (2009). The vapBC Operon from Mycobacterium smegmatis Is An Autoregulated Toxin-Antitoxin Module That Controls Growth via Inhibition of Translation. *Journal of Molecular Biology, 390*(3), 353-367. doi: 10.1016/j.jmb.2009.05.006

Rotem, E., Loinger, A., Ronin, I., Levin-Reisman, I., Gabay, C., Shoresh, N., . . . Balaban, N. Q. (2010). Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. [Article]. *Proceedings of the National Academy of Sciences of the United States of America, 107*(28), 12541-12546. doi: 10.1073/pnas.1004333107

Schramm, V. L. (2007). Enzymatic transition state theory and transition state analogue design. *Journal of Biological Chemistry, 282*(39), 28297-28300. doi: 10.1074/jbc.R700018200

Schwartz, S. D., & Schramm, V. L. (2009). Enzymatic transition states and dynamic motion in barrier crossing. *Nature Chemical Biology, 5*(8), 551-558.

Tadokoro, T., & Kanaya, S. (2009). Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *Febs Journal, 276*(6), 1482-1493. doi: 10.1111/j.1742-4658.2009.06907.x

Thompson, J. E., Venegas, F. D., & Raines, R. T. (1994). ENERGETICS OF CATALYSIS BY RIBONUCLEASES - FATE OF THE 2',3'-CYCLIC PHOSPHODIESTER INTERMEDIATE. *Biochemistry, 33*(23), 7408-7414. doi: 10.1021/bi00189a047

Tock, M. R., Frary, E., Sayers, J. R., & Grasby, J. A. (2003). Dynamic evidence for metal ion catalysis in the reaction mediated by a flap endonuclease. [Article]. *Embo Journal, 22*(5), 995-1004.

Van Melderen, L. (2010). Toxin-antitoxin systems: why so many, what for? [Review]. *Current Opinion in Microbiology, 13*(6), 781-785. doi: 10.1016/j.mib.2010.10.006

Van Melderen, L., & De Bast, M. S. (2009). Bacterial Toxin-Antitoxin Systems: More Than Selfish Entities? [Review]. *Plos Genetics, 5*(3).

Wang, N. R., & Hergenrother, P. J. (2007). A continuous fluorometric assay for the assessment of MazF ribonuclease activity. *Analytical Biochemistry, 371*(2), 173-183. doi: 10.1016/j.ab.2007.07.017

Wang, X. X., & Wood, T. K. (2011). Toxin-Antitoxin Systems Influence Biofilm and Persister Cell Formation and the General Stress Response. [Article]. *Applied and Environmental Microbiology, 77*(16), 5577-5583. doi: 10.1128/aem.05068-11

WHO. (2011). WHO Report 2011: Global Tuberculosis Control. Geneva.

Winther, K. S., & Gerdes, K. (2009). Ectopic production of VapCs from Enterobacteria inhibits translation and trans-activates YoeB mRNA interferase. [Article]. *Molecular Microbiology, 72*(4), 918-930. doi: 10.1111/j.1365-2958.2009.06694.x

Yamaguchi, Y., Park, J.-H., & Inouye, M. (2011). Toxin-Antitoxin Systems in Bacteria and Archaea. *Annual Review of Genetics, 45*(1), 61-79. doi: doi:10.1146/annurev-genet-110410-132412

Yang, W. (2011). Nucleases: diversity of structure, function and mechanism. [Review]. *Quarterly Reviews of Biophysics, 44*(1), 1-93. doi: 10.1017/s0033583510000181

Yang, W., Lee, J. Y., & Nowotny, M. (2006). Making and breaking nucleic acids: Two-Mg2+-ion catalysis and substrate specificity. [Review]. *Molecular Cell, 22*(1), 5-13. doi: 10.1016/j.molcel.2006.03.013

Zhang, Y. L., Zhang, J. J., Hara, H., Kato, I., & Inouye, M. (2005). Insights into the mRNA cleavage mechanism by MazF, an mRNA interferase. *Journal of Biological Chemistry, 280*(5), 3143-3150. doi: 10.1074/jbc.M411811200

Zhang, Y. X., Guo, X. K., Wu, C., Bi, B., Ren, S. X., Wu, C. F., & Zhao, G. P. (2004). Characterization of a novel toxin-antitoxin module, VapBC, encoded by Leptospira interrogans chromosome. *Cell Research, 14*(3), 208-216. doi: 10.1038/sj.cr.7290221

Zhu, L., Sharp, J. D., Kobayashi, H., Woychik, N. A., & Inouye, M. (2010). Noncognate Mycobacterium tuberculosis toxin-antitoxins can physically and functionally interact. *Journal of Biological Chemistry, 285*(51), 39732-39738.