SPECIFIC RESIDUES AND REGIONS WITHIN THE KH AND S1 DOMAINS OF PNPASE THAT ARE IMPORTANT FOR RNA BINDING AND AUTOREGULATION

by

Alexander Guanghua Wong

B.Sc., The University of British Columbia, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE STUDIES

(Biochemistry and Molecular Biology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2012

© Alexander Guanghua Wong, 2012
Abstract

The turnover of RNA is a critical process governing key aspects of cell growth, regulation, and survival. In particular, gene expression is largely mediated by the differential stabilities of mRNAs and their corresponding levels of translation. Polynucleotide phosphorylase (PNPase) is a phosphate-dependent 3’ exoribonuclease found in many different organisms. Although PNPase functions are diversified in nature, in *E. coli*, it functions primarily in mRNA degradation by 3’ to 5’ phosphorolysis. It is also a major component of the RNA degradosome – a macromolecular association of proteins involved in RNA decay. The six N-terminal RNase PH-like domains of PNPase assemble into a ring-shaped trimer forming the active core, through which the RNA substrate is threaded during degradation. The C-terminal KH and S1 RNA-binding domains are positioned at the entrance to this core structure and have been shown to facilitate substrate recognition and interaction. Towards developing a better understanding of the proposed KH-S1 binding platform, we have endeavoured to identify and investigate key residues in these domains responsible for RNA binding. In this study, we use a convenient *pnp::lacZ* fusion reporter strain to assess autoregulatory ability by both PNPase lacking the KH and S1 domains and containing point mutations in these domains. PNPase represses its own mRNA resulting in autoregulation and RNA binding has been shown to facilitate this. KH and S1 point mutants identified from a library and those predicted to contact RNA were expressed in the reporter strain. Mutants found to be deficient in autoregulation poorly repressed levels of *pnp::lacZ* transcript compared to WT PNPase, resulting in higher β-galactosidase levels. Interesting mutants were purified and studied using *in vitro* band-shift and phosphorolysis assays to assess binding and enzymatic activity. We show that reductions in substrate affinity accompanied impairment in PNPase autoregulation. A remarkably strong correlation was
observed between β-galactosidase levels reflecting ability to autoregulate, and apparent $K_{d}$ values reflecting \textit{in vitro} affinities for a model RNA substrate, implicating the KH and S1 domains of PNPase, and specific residues within them, in substrate binding and autoregulation. Our findings are discussed in the context of PNPase structure, binding architecture, and function.
Preface

Dr. George Mackie of the University of British Columbia constructed the reporter strain IBPC7322(λGF2) as described in section 2.1. The strain was used for a large part of the work reported in section 3.1. Also, Figures 1, 2, 6, and 15 contain elements from seminar presentation slides created by Dr. Mackie.

Work involving PNPase truncation, "triple-alanine", and "triple-arginine" mutants was a continuation of Leigh Stickney's work.

Work involving the PNPase F635A, F638A, and H650A mutants was substantially contributed to by Kristina McBurney.

Katharine Thompson performed a large portion of the many PNPase purifications as described in section 2.5.
Table of Contents

Abstract.......................................................................................................................................... ii

Preface........................................................................................................................................... iv

Table of Contents .......................................................................................................................... v

List of Tables ................................................................................................................................... vii

List of Figures ............................................................................................................................... viii

List of Symbols, Abbreviations, and Terms .............................................................................. ix

Acknowledgements ..................................................................................................................... xii

Dedication ................................................................................................................................... xiii

Chapter 1: Introduction ............................................................................................................... 1

1.1 mRNA destabilization in Escherichia coli .................................................................................. 1

1.1.1 The RNA degradosome ....................................................................................................... 3

1.2 Polynucleotide phosphorylase and related structures .............................................................. 5

1.2.1 A widely-distributed ribonuclease ....................................................................................... 6

1.2.2 The degrading exosomes .................................................................................................... 7

1.2.3 Escherichia coli PNPase .................................................................................................... 8

1.3 Distinct domains and their tertiary folding and organization in PNPase.............................. 9

1.3.1 The N-terminal catalytic core domains .............................................................................. 11

1.3.2 The KH and S1 RNA binding domains ............................................................................ 13

1.3.2.1 The K-Homology domain .......................................................................................... 15

1.3.2.2 The S1 domain ........................................................................................................... 17

1.4 Structure-function relationships of PNPase and related complexes ...................................... 19

1.4.1 PNPase autoregulation in E. coli ....................................................................................... 23

1.5 Overview of objectives and experimental scope ..................................................................... 27

Chapter 2: Materials and Methods ........................................................................................... 29

2.1 Bacterial strains and plasmids ............................................................................................... 29

2.2 Construction of mutagenic KH-S1 library .......................................................................... 30

2.3 β-galactosidase assay ........................................................................................................... 31

2.4 Analysis of PNPase levels in reporter strain IBPC7322(λGF2) by Western blot ............... 32
2.5 Purification of PNPase and PNPase variants ................................................................. 33
2.6 Preparation of SL9A substrate and enzyme assays ......................................................... 33

Chapter 3: Results ................................................................................................................. 35
3.1 Both the S1 and KH RNA-binding domains facilitate PNPase autoregulation ............... 35
3.2 RNA binding by S1 domain mutants .............................................................................. 44
3.3 RNA binding by KH domain mutants .............................................................................. 47
3.4 Enzymatic activity of KH, S1 domain mutants ............................................................... 48

Chapter 4: Discussion ............................................................................................................. 52
4.1 PNPase autoregulation activity strongly correlates with in vitro RNA binding activity ... 52
4.2 Plasticity of RNA contacts in the S1 domain ................................................................. 53
4.3 RNA binding regions within the KH domains were identified with novel insights into mechanisms of RNA binding impairment .......................................................... 55
4.4 Non-equivalency in RNA binding by the KH and S1 domains: a necessity for PNPase function? .................................................................................................................. 59

References ............................................................................................................................ 63

Appendices ............................................................................................................................ 74
Appendix A : Cloning, mutagenesis, and purification of mutants .......................................... 74
A.1 Primers and cloning strategy ....................................................................................... 74
A.2 Purification of PNPase I576T from over-expression in strain ENS134/pAW017 ............ 79
A.3 Partial proteolysis of PNPase F635A/F638A/H650A and F635R/F638R/H650R .......... 80
Appendix B : SWISS-MODEL Homology modeling parameters ........................................ 81
List of Tables

Table 1: Autoregulation of pnp::lacZ message by PNPass mutants ............................................. 36
Table 2: *In vitro* assays of PNPass and PNPass mutants.............................................................. 47

Appendix Tables

Table A1: Key primers used in initial PCR reactions (and Error-prone PCR).............................. 74
Table A2: Other primers for site-directed mutagenesis................................................................. 75
List of Figures

Figure 1: Current model for mRNA decay in *E. coli* ................................................................. 3
Figure 2: A structural theme adopted by RNase PH, PNPases, and the RNA-degrading exosomes. ......................................................................................................................................................... 5
Figure 3: Domain topology and quaternary structures of PNPase and *S. antibioticus* GPSI. ...... 10
Figure 4: Modeling of the KH and S1 RNA binding domains. .......................................................... 14
Figure 5: Relative positions of S1 RNA binding domains in *E. coli* ribonucleases. .................... 17
Figure 6: Model for the role of KH-S1 in substrate binding, retention, and release. ................... 22
Figure 7: Autoregulation of *pnp* mRNA ....................................................................................... 24
Figure 8: Use of reporter strain IBPC(λGF2) to assess PNPase autoregulation ......................... 26
Figure 9: Repression of *pnp::lacZ* message in the reporter strain by KH and S1 mutants ......... 38
Figure 10: *In vitro* RNA binding activity of WT PNPase and representative PNP mutants ....... 46
Figure 11: Assays of SL9A RNA processing by PNPase and its derivatives ................................. 50
Figure 12: Correlation between RNA binding and PNPase autoregulation. ............................... 52
Figure 13: Contributions to RNA binding by F635/F638/H650 from the S1 domain ............... 54
Figure 14: Comparison of RNA-bound Nova-2 KH3 to predicted structure of the KH domain in
*E. coli* PNP. ................................................................................................................................. 56
Figure 15: Model proposing greater role of KH in PNPase-RNA interaction and PNPase
function. ......................................................................................................................................... 60

Appendix Figures

Figure A1: Flowchart of the mutagenic KH/S1 and pAW101 cloning processes ....................... 77
Figure A2: The over-expression and purification of PNPase I576T. ........................................... 79
Figure A3: Trypsin sensitivity assay of FFH -> AAA and FFH -> RRR mutants ....................... 80
### List of Symbols, Abbreviations, and Terms

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-32P-CTP</td>
<td>[alpha-32P]-radiolabeled cytidine 5'-triphosphate nucleotide</td>
</tr>
<tr>
<td>3'-UTR</td>
<td>3'-untranslated region</td>
</tr>
<tr>
<td>5'-UTR</td>
<td>5'-untranslated region</td>
</tr>
<tr>
<td>~</td>
<td>approximately</td>
</tr>
<tr>
<td>Å</td>
<td>Angstrom, $10^{-10}$ meters</td>
</tr>
<tr>
<td>aa</td>
<td>amino acid</td>
</tr>
<tr>
<td>AAHD</td>
<td>all-α-helical domain</td>
</tr>
<tr>
<td>approx.</td>
<td>approximately</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine 5'-triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BL21</td>
<td><em>E. coli</em> strain for protein over-expression</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>B. subtilis</td>
<td><em>Bacillus subtilis</em></td>
</tr>
<tr>
<td>°C</td>
<td>degree Celsius</td>
</tr>
<tr>
<td>C. crescentus</td>
<td><em>Caulobacter crescentus</em></td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>c-di-GMP</td>
<td>cyclic diguanylic acid</td>
</tr>
<tr>
<td>Csp</td>
<td>cold-shock protein</td>
</tr>
<tr>
<td>C-terminal</td>
<td>carboxyl-terminal</td>
</tr>
<tr>
<td>CTP</td>
<td>cytidine 5'-triphosphate</td>
</tr>
<tr>
<td>cTP</td>
<td>chloroplast-transit peptide</td>
</tr>
<tr>
<td>DEAD</td>
<td>amino acid sequence (asp-glu-al-asp)</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleotide triphosphate</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>E. coli</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediamine tetraacetatic acid</td>
</tr>
<tr>
<td>fmoles</td>
<td>femto moles</td>
</tr>
<tr>
<td>GPSI</td>
<td>guanosine pentaphosphate synthetase I</td>
</tr>
</tbody>
</table>
GTP  guanosine 5'-triphosphate

*H. sapiens*  *Homo sapiens*

hPNPase  human PNPase

IPTG  isopropyl 1-thio-β-D-galactopyranoside

K<sub>d</sub>  dissociation constant

kDa  kiloDalton = 1000 gram/mole

KH  K-Homology RNA binding domain

lacZ  gene encoding β-galactosidase

LB  Luria-Bertani broth

M  molar

mg  milligram

μCi  microCurie

μg  microgram

μl  microliter

min  minutes

ml  milliliter

mM  milliMolar

mRNA  messenger RNA

MTS  mitochondrial-targeting sequence

MU  Miller units of β-galactosidase activity

NDP or ppX  nucleoside diphosphates

ng  nanogram

nm  nanometer (or in the context of mutagenic library construction, = not-mutated)

nM  nanoMolar

NMR  nuclear magnetic resonance

N-terminal  amino-terminal

ONPG  *ortho*-nitrophenyl-β-D-galactopyranoside

ORF  open reading frame

PAGE  poly-acrylamide gel electrophoresis

PAP I  poly(A) polymerase
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>protein data bank</td>
</tr>
<tr>
<td>P&lt;sub&gt;i&lt;/sub&gt;</td>
<td>inorganic phosphate</td>
</tr>
<tr>
<td>pmoles</td>
<td>pico moles</td>
</tr>
<tr>
<td><em>pnp</em></td>
<td>gene encoding polynucleotide phosphorylase in <em>E. coli</em></td>
</tr>
<tr>
<td>PNPase</td>
<td>polynucleotide phosphorylase</td>
</tr>
<tr>
<td>RhlB</td>
<td>ATP-dependent DEAD box RNA helicase</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>RNase E / Rne</td>
<td>ribonuclease E, an <em>E. coli</em> endonuclease</td>
</tr>
<tr>
<td>RNase PH</td>
<td>ribonuclease PH, 3' to 5' phosphorolytic exoribonuclease</td>
</tr>
<tr>
<td>RppH</td>
<td>5' pyrophosphohydrolase</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>S1; S1 domain</td>
<td>Ribosomal protein from which S1 RNA binding domain originally identified</td>
</tr>
<tr>
<td><em>S. antibioticus</em></td>
<td><em>Streptomyces antibioticus</em></td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SL9A</td>
<td>a synthetic radiolabeled RNA substrate used in this work</td>
</tr>
<tr>
<td><em>S. solfataricus</em></td>
<td><em>Sulfolobus solfataricus</em></td>
</tr>
<tr>
<td>sRNA</td>
<td>regulatory small RNA</td>
</tr>
<tr>
<td>TBE</td>
<td>Tris-borate-EDTA</td>
</tr>
<tr>
<td>TRAMP</td>
<td>Trf4/Air2/Mtr4p Polyadenylation complex</td>
</tr>
<tr>
<td>Tris</td>
<td>2-amino-2-hydroxymethyl-propane-1,3-diol</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>TTP</td>
<td>thymidine 5'-triphosphate</td>
</tr>
<tr>
<td>v/v</td>
<td>volume/volume</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-type</td>
</tr>
<tr>
<td>w/v</td>
<td>weight/volume</td>
</tr>
</tbody>
</table>
Acknowledgements

I wish to express my most sincere gratitude to Dr. George Mackie for allowing me to work on this important project, and for teaching me the basics of classic biochemistry and critical scientific thinking. Thank-you also for your guidance. I would also like to thank the members of my supervisory committee for their guidance, insight, and patience: Dr. Eric Jan and Dr. Lawrence McIntosh. Thank-you as well to Dr. Thomas J. Beatty for serving as University examiner and providing detailed feedback that was implemented prior to final submission.

I would also like to thank the past and present members, and visitors to the laboratory who have greatly assisted me: Janet Hankins, Dr. Stephen Garrey, Dr. George Jones, and Katharine Thompson.

A special thanks to Dr. Kush Dalal for his guidance and friendship.

In addition to my family, close friends have also contributed to this work through their generous support: Dr. Carson Chin; members of that particular friendship circle (you know who you are); Paul Moore and the Moore/Filan family; and Casey Wong. Thanks!

The work presented in this thesis was made possible by the generous funding provided by the Canadian Institutes of Health Research and The University of British Columbia.
Dedication

To my family. In particular, for my mother, who has sacrificed immensely for me and who, to this day, wishes she had the opportunity to receive a quality higher education. Love you, mom. I hope this work inspires you to pursue that dream. The gift of education truly is a blessing and should remain a right for all.

Also to science, the human spirit, and the less fortunate.
Chapter 1: Introduction

RNA metabolism is a tightly regulated process in all organisms and as such remains an increasingly significant focus of scientific research efforts. This vital life process mediates aspects of gene expression, development and growth, environmental adaptation, and reproduction. Many regulatory mechanisms are driven by RNA turnover including messenger RNA (mRNA) surveillance (29, 123) and RNA interference (132). Processing of RNAs by ribonucleases and related complexes is thus multifaceted and governs differential gene expression and ribonucleotide recycling in both pro- and eukaryotes [(42, 108); and reviewed in (6)]. Specifically, discrepancies in stability and translational efficiencies arising from differential susceptibilities of the various transcripts to ribonucleases contribute to changes in gene expression at the genome-wide scale. Rapid decay allows both the amplification of negative regulatory signals and, in some examples, the up-regulation of mRNA transcript via the destabilization of small regulatory RNA [ChiX sRNA, (38, 72)]. In E. coli, typical mRNA half-lives average under 10 minutes (11, 105) with a few exceptions like ompA mRNA exhibiting half-lives of up to 15 minutes (9).

1.1 mRNA destabilization in Escherichia coli

The organism of choice for studying mRNA metabolism and in particular, pathways for RNA processing, continues to be E. coli. Current paradigms of RNA decay in this organism [reviewed in (6) and (107)] evolved from earlier models based on data and interpretations from Kepes, Apirion, and Kennell [reviewed in (27)]. Key developments from the investigation of specific mRNAs in vivo, and the reconstitution of mRNA decay complexes with purified components in vitro (26, 28, 70, 71, 113), have resulted in a widely-accepted model of the
process as illustrated in Fig. 1. In this pathway, many endoribonucleases have been proposed to be instigators of nucleolytic decay, but mounting evidence implicates Ribonuclease E (RNase E) as the major player in the initiating cleavage step [reviewed in (6) and (8, 20)]. This essential, low-specificity endoribonuclease with a preference for single-stranded AU-rich regions also exhibits a kinetic bias for single-stranded 5'-monophosphorylated RNA species (54, 70, 120). Coincidentally, recent evidence suggests that for certain mRNAs, a step prior to initial cleavage is carried out by the pyrophosphatase RppH (23, 32). RppH catalyzes the rate-limiting conversion of the 5'-end triphosphate; after which the 5'-monophosphorylated end is subject to attack by RNase E. A portion of the resulting fragments contain unstructured 3'-ends, rendering them susceptible to the action of 3' exonucleases; namely polynucleotide phosphorylase (PNPase), RNase II, RNase R, and oligoribonuclease [reviewed in (6, 8)]. Among these enzymes, RNase II may account for approximately 90% of the activity, with PNPase the majority of the remaining 10% (34). *E. coli* strains lacking both PNPase and RNase II are inviable (35). The remaining fragments retaining 3'-end protection from exonucleases (Fig. 1, right side) are subject to further attack by RNase E and, importantly, to attack by the concerted efforts of an association of enzymes involved in RNA decay within which PNPase plays a critical role (see section 1.1.1). This macromolecular complex is known as the RNA degradosome [reviewed in (19) and (41)].
Degradation initiates with the rate-limiting conversion of 5'-triphosphate to 5'-monophosphate by RppH (23, 32). RNase E catalyzes an initiating endonucleolytic cleavage. Decay of resulting polynucleotides unprotected at the 3'-end is catalyzed by the exonucleases PNPase and RNase II. The 3'-terminal stem-loop is resistant to exoribonucleases, but can be “edited” by poly(A) polymerase, making it susceptible to the action of an RNA helicase coupled to PNPase, or the RNA degradosome.

### 1.1.1 The RNA degradosome

Under normal growth conditions in *E. coli*, the major components of the degradosome include PNPase, RNase E, the DEAD-box RNA helicase RhlB, and the glycolytic enzyme
enolase. The first three components are critical for the ATP-dependent degradation of highly structured 3'-terminal RNA fragments (28). Additionally, poly(A) polymerase (PAP I) can play a role in the decay process after initial cleavage by RNase E (Fig. 1): Here, PAP I catalyzes the ATP-dependent addition of an unstructured 3'-poly(A) tail and stimulates PNPase activity by potentiating its binding to the newly-formed single-stranded 3'-extension (13, 112, 128).

Subsequent to PAP I action on the RNA, further degradation is catalyzed by two components of RNA degradosome: PNPase and the helicase RhlB. The ATP-dependent helicase RhlB, activated by RNase E, assists PNPase by "melting" otherwise stable stem-loops (28, 64, 91). With the help of RhlB, the action of PNPase as part of the RNA degradosome is able to overcome structural barriers at the 3'-terminus.

The features of the degradosome components point to a particularly important and multifarious role for the complex in vivo. Indeed, electron microscopy has confirmed the existence of the multi-enzyme assembly associated with the cytoplasmic membrane in E. coli via "segment A" of the C-terminal half of RNase E (57). More recently, Taghbalout and Rothfield show through in vivo fluorescence tagging the organization of degradosome components into filamentous complexes of the E. coli cytoskeleton (118). Another interesting feature of the degradosome is its ability to dynamically alter its shape and composition in response to environmental changes like temperature drop and phosphosugar stress [reviewed in (41)]. In doing so, the RNA degradosome may lend its degradative functionalities to other pathways governing processes outside of native growth conditions. Not surprisingly, analogous macromolecular complexes involved in RNA decay have been described in other bacteria (1, 49, 90). As well, archaeal and eukaryotic renditions of the E. coli degradosome known as "RNA-degrading exosomes" have been studied and visualized [reviewed in (41, 50, 67)]. These
complexes also possess nuclease function coupled to RNA helicase activity provided by associating accessory multi-protein components [Yeast Mtr4p in TRAMP and Ski proteins; (63)]. The recurring architectural themes of these exosome structures are compared in Fig. 2, and exosome structure-function relationships are discussed in the context of PNPase in sections 1.2.2 and 1.4.

Figure 2: A structural theme adopted by RNase PH, PNPases, and the RNA-degrading exosomes.
The quaternary core structures of RNase PH, PNPase homologues, and the archaeal and eukaryotic exosomes are compared. Text on bottom line below structures describes domains and/or subunits that are attached and believed to confer RNA (or polynucleotide) binding. Exosome structures are reviewed in (41, 50, 67).

1.2 Polynucleotidase phosphorylase and related structures

Despite a storied history dating back to the 1950's that has facilitated the establishment of the genetic code and site-directed mutagenesis (40, 43), the full repertoire of the functions of polynucleotide phosphorylase (PNPase) remains to be elucidated. In *E. coli*, PNPase (EC 2.7.7.8) is a true enzyme of RNA metabolism in the sense that it can also synthesize oligonucleotides in the 5'-to-3' direction *in vivo* while primarily serving as a 3'-to-5' exoribonuclease [reviewed in (110)]. These two enzymatic activities can be summarized as
follows: \[ \text{ppX} + (\text{pX})_n \rightleftharpoons 5'-\text{(pX)}_n\text{pX-OH} + \text{P}_i; \] where \( \text{p} \) represents a phosphate group and \( \text{P}_i \), inorganic phosphate. The forward reaction uses nucleoside diphosphates (NDPs, or ppX) in the polymerization of both residual poly(A) tails in PAP I-deficient strains, and heteropolymeric tails in wild-type \textit{E. coli} (75). However, the reverse and processive phosphorolytic activity predominates \textit{in vivo} owing to high physiological concentrations of inorganic phosphate (\( \text{P}_i \)) (65). These activities allow the enzyme to function in vital cell processes that include, but are not limited to, mRNA degradation, ribosomal RNA (rRNA) quality control, transfer RNA (tRNA) processing, and the cold shock response (7, 25, 76, 86, 129). Recent evidence also suggests that PNPase is a key regulator of small non-coding RNA (sRNA) function in \textit{E. coli} (2, 3, 31). Additionally, PNPase appears to respond to changes in oxygen levels as the downstream target of the signal ligand cyclic diguanylic acid (c-di-GMP), and is also known to be modulated by ATP (33, 121). The Krebs cycle metabolite citrate has also been reported to modulate PNPase activity (83). Its regulation by c-di-GMP, ATP, and citrate in \textit{E. coli} hint at essential roles for PNPase in both environmental adaptation and cellular metabolism.

\subsection{1.2.1 A widely-distributed ribonuclease}

PNPase homologues have been studied in other bacteria and are also ubiquitously found in plant and eukaryotic organelles, and in higher mammals with both varied and specialized functions [reviewed in (63, 100, 103)]. To this effect, two genes encode PNPase in plants resulting in one type containing a chloroplast-transit peptide (cTP), while the other has a mitochondrial-targeting sequence (MTS) (85, 124, 130). Indeed, the differential localization of PNPase in plants results in one type functioning in both polyadenylation and phosphorolysis within chloroplasts whereas in plant mitochondria, PNPase appears to lack polyadenylation
activity [reviewed in (103)]. Human PNPase (hPNPase) was discovered during a screen for
genes implicated in cellular differentiation and senescence [reviewed in (30)]. Although a
typical MTS is found at the N-terminus of hPNPase, its functions are known to be diversified not
only within mitochondria, but also in the cytoplasm [reviewed in (24, 30)]. The over-expression
of hPNPase in the cytoplasm resulting in destabilization of *c-myc* mRNA along with a subset of
micro RNAs has been implicated in growth inhibition and age-related inflammation. In contrast,
hPNPase localized to the mitochondria resides primarily in the intermembrane space and
regulates aspects of mitochondrial RNA processing, homestostasis, and processing-independent
RNA import into the mitochondria (61, 125). Potential roles in antiviral responses and
tumorigenesis have also been proposed for hPNPase [reviewed in (24, 30)]. More complex
functions for PNase, however, are not limited to homologues in higher mammals; PNase has
been found to regulate virulence in *Salmonella enterica* and starvation response in the
extremophilic bacterium, *Deinococcus radiodurans* (126, 131).

### 1.2.2 The degrading exosomes

Interestingly, structures that are structurally and functionally analogous to PNase have
been found in yeast (and other eukaryotes) and archaea (59, 63). In these organisms, many of the
RNA degradation and quality control pathways involve the action of the archaeal or eukaryotic
exosomes [reviewed in (17, 67, 92)]. The archaeal and eukaryotic exosomes both share the ring-
shaped quaternary architecture conserved in the PNases (Fig. 2; discussed in section 1.4), but
the action of the eukaryotic exosome core is not phosphorolytic [reviewed in (30, 41, 50, 63,
67)].
1.2.3 *Escherichia coli* PNPase

PNPase in *E. coli* is encoded by the *pnp* gene which was first sequenced by Régnier *et al.* (95). The sequence cloned and used in this work is from the following strain: EcoGene Accession Number EG10743; *E. coli* str. *K-12 substr. MG1655*; (see Appendix A1 for details). The *pnp* gene is part of the *rpsO-pnp* operon encoding both PNPase and the *rpsO* gene product, ribosomal protein S15 (95). Transcription of *pnp* and *rpsO* is jointly regulated. Translation of the 711 amino-acid *pnp* product is initiated at a non-conventional 5'-UUG-3' start codon (replaced with 5'-AUG-3' by cloning in this work; see Appendix Table A1). Interestingly, the purified PNPase monomer runs anomalously as an ~85 kDa protein through denaturing SDS-polyacrylamide gels (111) (see Materials and Methods, section 2.5 for PNPase purification steps).

Crude extracts of *E. coli* contain a heterogeneous mixture of two active PNPase forms: Form A with a molecular mass of ~252 kDa, and form B with a molecular mass of ~365 kDa [reviewed in (63)]. The A-form is a homotrimer (α₃) of three identical catalytic (α) PNP monomers. The resulting quaternary structure of α₃ resembles that of the hexameric RNase PH complex and is an example of the recurring "doughnut-shaped" theme adopted by PNPase orthologues and RNA-degrading exosomes (Fig. 2). The B-form consists of PNPase α subunits assembled in an α₃β₂ arrangement with β subunits of ~48 kDa in size (87). It is generally thought that β is the 48-kDa protein enolase (21, 91, 94). Alternatively, recent Co-immunoprecipitation, *in vitro* reconstitution, and protein interaction studies implicate the DEAD-box RNA helicase RhlB as the β subunit of the active B-form of PNPase (63). Interestingly, in this putative B-form, PNPase α associates with RhlB (β) independent of the RNA degradosome (62).
Processive degradation utilizing the main \textit{in vivo} activity of PNPase, phosphorolysis (introduced in \textbf{section 1.2}), requires a minimal single-stranded overhang of 7-10 ribonucleotides (112). In addition, the enzyme requires the coordination of Mg$^{2+}$ or a closely related divalent cation (most commonly Mn$^{2+}$ experimentally) for its phosphorolytic activity [reviewed in (66)]. Though unresolved, the likely role of the coordinated metal ion in phosphorolytic catalysis is discussed in \textbf{section 1.3.1}. RNA substrates with tails containing \(> 10\) unpaired ribonucleotides are very efficiently degraded by PNPase (112). However, stem-loop constructions of \(\geq 7\) bp at the 3' terminus of these substrates stall the processive action of PNPase. After stalling, degradation may continue with the help of the RNA helicase, RhlB, either via a direct interaction with PNPase (explained in previous paragraph), or through association within the RNA degradosome (Fig. 1, right side; discussed in \textbf{section 1.1.1}).

\subsection*{1.3 Distinct domains and their tertiary folding and organization in PNPase}

The \textit{E. coli} PNPase monomer is 711 residues long and is organized into a 4-domain plus 1 linking sub-domain topology (5 domains altogether) that is widely duplicated in the bacterial, plant, and mammalian PNPase homologues (10, 59, 63). Although some inferences are made in regards to the folds and positioning of the oft-unresolved KH and S1 domains (\textbf{section 1.3.2}), recent crystal structures of PNPase from \textit{E. coli}, gram-negative \textit{Caulobacter crescentus}, and \textit{Homo sapiens} corroborate the conserved 5-domain organization (46, 61, 82, 106). This tertiary folding is also observed in the crystal structure of the PNPase homologue from \textit{Streptomyces antibioticus}, guanosine pentaphosphate synthetase I [GPSI, (55)], which serves as a convenient model for the \textit{E. coli} enzyme (see \textbf{Fig. 3}) (116).
Figure 3: Domain topology and quaternary structures of PNPase and _S. antibioticus_ GPSI.

**a.** The domain topology of the _E. coli_ PNPase and _S. antibioticus_ GPSI monomers are compared in an alignment. All panels are reproduced and modified from (116) and are based on their crystal structure of GPSI (PDB I.D. 1E3P). Coordinates for _E. coli_ PNP are those published in (114), but also align well based on modeling done as part of this work (see **Fig. 4acd**). Roman numerals represent numbering of the structured linkers.

**b.** Schematic drawing showing quaternary structure of GPSI as depicted and discussed in (116). Only one monomer is coloured: maroon for α helices, dark blue for β strands, yellow for 3_10 helices, and gray for loops. Conserved atoms of the important central-channel FFRR loop (GPSI residues 78-92) are also coloured.

**c.** View of single GPSI monomer from the outer surface of the trimer at right angles to the crystallographic 3-fold axis. Colouring scheme is as in (b).

**d.** A topology diagram for one GPSI monomer. Double-headed arrows indicate disjunctions made for display. Note the indication of proposed trimerization interfaces. Figure reproduced and modified with permission from (116).
1.3.1 The N-terminal catalytic core domains

In the crystal structures from *S. antibioticus* and *E. coli*, the N-terminal region of PNPase is organized into two duplicated "core" domains (named PH' and PH in *E. coli* PNPase; PH-1 and PH-2, respectively, in orthologues; Fig. 3a) that each closely resembles the phosphorolytic exoribonuclease, RNase PH (EC 2.7.7.56). In bacteria, PNPase and RNase PH catalyze phosphate-dependent RNA degradation and are thus members of the phosphate-dependent exoribonuclease family (PDX) of proteins [reviewed in (59, 63)]. However, in contrast to PNPase, RNase PH functions primarily in the 3'-end processing of tRNA precursors. Structure-based sequence and phylogenetic analyses suggest that the PH'-PH core of PNPase likely arose from a gene duplication event that also produced the RNase PH gene (10, 59). In this context, the second core domain, PH (residues 312-541 in *E. coli* PNPase, Fig. 3a), and RNase PH are more closely related than either is to the first core domain, PH' (residues 8-210 in *E. coli* PNPase). Indeed, the crystal structure of GPSI shows the phosphate (P$_i$) analog, tungstate, coordinated by T462 and S463 within the second PH domain (PHO2; residues 350-570 in GPSI, Fig. 3a), suggesting that the active site participating in phosphorolysis likely resides in the second core domain of *S. antibioticus* PNPase. In corroboration, the co-crystallized Mn$^{2+}$ ion of the *E. coli* PNPase core structure is observed in the active site coordinated by the conserved residues D486, D492, and K494 of the second core domain (PH, Fig. 3a) (82). The conserved equivalent metal-coordinating residues in *Bacillus subtilis* RNasePH and the *Sulfolobus solfataricus* exosome Rrp41 subunit (Fig. 2) have been implicated in catalysis, and a D492G substitution in *E. coli* PNPase abolishes both phosphorolysis and polymerization (47, 51, 69). By overlaying their Mn$^{2+}$-bound PNPase core structure with that of the Rrp41 subunit of the *Pyrococcus abyssi* exosome, Nurmohamed *et al.* deduced that D486 and D492 are likely
positioning Mg$^{2+}$ to activate the attacking phosphate, and to support the transition state (82). These residues contacting this proposed transition state are conserved in PNPase and RNase PH orthologues in addition to Rrp41 subunits of the archaeal exosome (Fig. 2), which supports the model of a metal-assisted catalytic mechanism originating from inside the second core domain (PH in *E. coli* PNPase; PH-2 in GPSI; Fig. 3).

An all-α-helical domain (AAHD) separates the two N-terminal RNase PH-like core domains in all the crystal structures of PNPase from the previously mentioned organisms (46, 61, 82, 106, 116). Symmons et al. reported that the AAHD likely plays primarily a structural role as the sequence in this region is not as well-conserved among PNPase orthologues compared to the core PH domains (116). Instead, Symmons et al. suggested that linkers I and II in GPSI (Fig. 3) flanking the AAHD are more ordered and that the latter provides structural coordination in the positioning of the core PH-2 and AAHD. Linker III is proposed to coordinate the positioning of the KH domain with respect to the second core domain (the C-terminal KH and S1 RNA-binding domains of PNPase are discussed in sections 1.3.2 and 1.4). However, in their mutational study of *E. coli* PNPase, Briani et al. reported that mutations to residues corresponding to L328 and V329 of GPSI in helix 8 of the AAHD (Fig. 3bcd) impair both phosphorolysis and PNPase autoregulation [(16); PNPase autoregulation is discussed in section 1.4.1]. Moreover, Nurmohamed et al. reported that the AAHD of *E. coli* PNPase is ordered in an RNA-bound state and disordered in the RNA-free form, suggesting that the AAHD dynamically responds to substrate binding. In this regard, the AAHD of *E. coli* PNPase may regulate the flow of NDPs and Pi in and out of the active site (82).
1.3.2 The KH and S1 RNA binding domains

Two distinct RNA-binding folds are appended at the C-terminus of PNPase: the K-Homology RNA binding domain (KH domain; residues 551-591; Fig. 4acd), followed C-terminally by the S1 RNA binding domain (residues 622-691; Fig. 4ab). Attempted crystallization of PNPase with clearly resolved and ordered electron densities for the KH and S1 domains has largely failed as the domains are highly dynamic relative to the PNPase core (PH'-PH). Only partial traces of these domains are modeled in the S. antibioticus crystal structure of GPSI [Fig. 3; (116)]. The KH and S1 domains are not visible in the Shi et al. crystal structure of full-length E. coli PNPase (106), and the E. coli structures from Nurmohamed et al. were crystallized with PNPase enzyme completely lacking the KH and S1 domains [PNPase ∆KH + ∆S1, or just the PNPase core; (82)]. Only the KH domains are defined in the recent crystal structure of human PNPase (61). The KH domains are resolved in both the apo- and RNA-bound C. crescentus PNPase structures (46). Although the S1 domain is resolved but poorly-ordered in the RNA-free structure, it could not be resolved in the RNA-bound structure of C. crescentus PNPase. Nevertheless, modeling of the conserved folds within these ancient RNA binding units is based on various structures of KH/S1 domain assemblies found in proteins whose functions depend on nucleic acid interaction [(12, 18, 60) and reviewed in (122)].
Figure 4: Modeling of the KH and S1 RNA binding domains.

**a.** Domain organization of *E. coli* PNPase. Boundary coordinates are those described in (114), but also align well based on homology modeling done as part of this work [see c and d].

**b.** NMR structure of the *E. coli* PNPase S1 domain showcasing the positions of aromatic residues predicted to contact RNA by aromatic stacking interactions (18). Figure reproduced and modified with permission from (18).

**c.** Model of the *E. coli* PNPase KH domain. Homology modeling was performed with SWISS-MODEL (5, 44, 58, 104) using the crystal structure of the Nova antigen-2 KH3 (type I) domain [(60); PDB I.D. 1ec6A]. Model statistics are included in Appendix B. Side-chains K571 and I576 are represented by green and blue sticks, respectively. The conserved GxxG loop is highlighted red.

**d.** Surface representation of the predicted electrostatic properties of the KH domain (by PyMOL), as modeled in (c). Positions of K571 and I576 are indicated. Blue represents electropositive regions; red, electronegative ones.
1.3.2.1 The K-Homology domain

KH domains are named for their homology to an RNA binding motif first identified in the human heterogeneous nuclear ribonucleoprotein K (hnRNP K) protein (109). The KH domain of *E. coli* PNPase comprises amino acid residues 551-591 (Fig. 4a), and with a much shorter variable loop region compared to other KH domains (discussed below), is smaller than the typical KH motif of ~70 amino acids [reviewed in (122)]. The domain coordinates presented in this work are well positioned based on a sequence alignment of the mammalian Nova [antigen-2; (60)] KH3 domain with the *E. coli* PNPase KH domain. This alignment with the third KH domain in Nova-2 and its corresponding crystal structure [PDB I.D. 1ec6; (60)] was used to generate a homology-based structure for the *E. coli* PNPase KH domain (Fig. 4cd). Both of these KH domains share the characteristic "type I KH" motif consisting of secondary structure arranged in an N-βα-GxxG-αβ...β'α'-C fold, where the GxxG loop intervenes between helix α1 and α2 and the three dots (...) represent a variable loop region of varying length [reviewed in (122)]. Type II KH domains are more characteristic for prokaryotic proteins, although the type I KH domain found in *E. coli* PNPase is one notable exception. Type II KH domains are arranged in an N-α'β'...βα-GxxG-αβ-C fold. Both types of KH folds contain the KH minimal motif, "βα-GxxG-αβ"; albeit in different positions relative to the remaining α' helix and β' sheet.

The GxxG sequence represents the conserved G-X-X-G loop which is a hallmark property of KH folds [red loop in Fig. 4c; reviewed in (122)]. Although the majority of KH domains contain this loop separating α1 and α2, a few structures of KH domains without canonical GxxG loops have been solved, and reveal the functioning of these domains in mediating protein-protein interactions (79, 84). In one example of this, none of the five type I KH domains in *Caenorhabditis elegans* GLD-3 contains the GxxG loop and the KH region of
the protein is unable to cross-link RNA \textit{in vitro} \cite{79}. Instead, the KH domains in GLD-3 are seen mediating extensive intra-molecular protein-protein contacts. Alternatively, the canonical GxxG loop is replaced by an exosome-specific "GxNG" sequence in the eukaryotic exosome Rrp40 subunit. Rrp40 contains a tandem-arrayed, N-terminal S1 domain, C-terminal type I KH domain configuration in which the unique GxNG sequence in KH is situated between two consecutive β-strands as opposed to the typical GxxG positioning between helix α1 and α2. The GxNG motif is buried at the interface between the S1 and KH domains and appears to play a structural role, making several hydrogen-bond contacts with residues of the S1 domain \cite{84}.

Typically, KH domains function in the recognition of, and binding to, RNA or ssDNA \cite{review(122),review(78)}. When present as part of a protein in combination with multiple KH and/or S1 domains, KH domains can function both independently or cooperatively. Type I domains are predominantly found in multiple copies of up to 15 KH domains in eukaryotic proteins [human vigilin protein, \cite{78}], whereas type II KH domains typically exist as single copies in prokaryotic proteins \cite{122}. Examples of tandem arrayed KH constructions in addition to vigilin include two type II KH domains trailing an S1 domain in bacterial transcription factor NusA \cite{12}, three type I KH domains in the RNA metabolism-regulating mammalian Nova antigen-2 protein \cite{60}, and two type I KH domains in the human fragile X mental retardation protein [FMR1, \cite{78}]. Although type I and type II KH domains differ in their fold ordering, a common RNA substrate binding mode is adopted by KH domains and is represented in the crystal structure of the Nova-2 KH3 domain bound to RNA \cite{60}. Typically, up to four nucleotides of the RNA or ssDNA is found in an extended, single-stranded conformation within a "binding cleft" pinioned by the GxxG loop, an α-helical/β-sheet platform of hydrophobic residues, and the variable loop region \cite{review(122)}.
1.3.2.2 The S1 domain

The first example of the conserved S1 RNA-binding fold was identified in, and thus named after, the *E. coli* ribosomal protein S1 (115). In gram negative bacteria, S1 protein facilitates the initiation of translation by binding to both the ribosome and the mRNA leader upstream of the Shine-Dalgarno sequence (14). The S1 fold was subsequently identified in many other proteins whose functions rely on interactions with single-stranded oligonucleotides. Notably, PNPase, RNase II, and RNase E all contain copies of the conserved S1 domain which belongs to a family of Oligonucleotide/oligosaccharide Binding (OB) folds [Fig. 5; (4, 18, 77, 119)]. The large OB-fold family spans all three domains of life and consists of compact nucleic acid recognition motifs exhibiting low sequence similarity while sharing prominent structural features. Although the S1 domain in RNase E has been closely investigated by NMR and X-ray crystallography by our lab and collaborators (102), many aspects of its interaction with RNA in general, and in the context of *E. coli* PNPase, remain unclear.

![Figure 5: Relative positions of S1 RNA binding domains in *E. coli* ribonucleases.](image)

The peptide chains of four *E. coli* ribonucleases are compared. Shaded regions with domain coordinates represent the relative positions of the S1 domains within each enzyme. The N-terminus is labeled and the peptide residue lengths are denoted by red numbers on the right.
In *E. coli* PNPase, the S1 domain spans residues 622-691 (Fig. 4ab and Fig. 5). The solution NMR structure of the *E. coli* PNPase S1 domain has been solved and is comprised of five antiparallel β-strands adopting a "Greek-key" topology while folding into a single β-barrel [Fig. 4b; (18)]. An alignment of known S1 sequences revealed five conserved residues within PNPase S1 (F635, F638, H650, D680, and R684) which were arrayed in close proximity to each other on the surface of one side of the β-barrel in such a way that suggested they were candidate substrate contact points (18). This cluster of residues is in a region topologically related to known or proposed binding sites as demonstrated structurally by other OB-fold proteins. Three of the aromatic residues (depicted in Fig. 4b) were subsequently targeted for mutagenesis in my study.

Interestingly, the expression of the isolated PNPase S1 domain suppresses the cold-sensitive phenotype of an *E. coli* strain containing a quadruple-deletion of four genes of the csp family (127). CspA is the major cold-shock protein in *E. coli* and functions as an RNA chaperone during the cold shock response (53). *E. coli* contains a large family of CspA homologues and the fold of the 70 amino-acid protein CspA classifies it as an OB-fold protein (36, 81, 101).

Accordingly, a fully active PNPase is essential for growth at low temperatures and during cold shock (7, 129, 133). Although the function of PNPase during cold acclimatization is not fully understood, expression of mutants containing both point substitutions and deletions in the KH and/or S1 domains are unable to rescue growth defects of cold-sensitive mutant *pnp* strains at low temperatures (37, 73).
1.4 Structure-function relationships of PNPase and related complexes

Functionally, PNPase adopts a homotrimeric quaternary structure that is conserved among PNPases and RNA-degrading exosomes across all three domains of life [Fig. 2; reviewed in (41, 50, 63, 67, 117)]. Its core PH domains are arranged into a "doughnut" ring forming a central channel-core that accommodates RNA substrates. There is substantial evidence that the KH and S1 RNA binding domains are positioned on top of the doughnut governing access into, and through a constricted aperture at the channel entrance enroute to the active site (46, 61, 82, 106, 116). The processivity of PNPase may be imparted by the ring-like structure of the core in combination with the entrance-proximal aperture, which work to "clamp" onto and retain the RNA substrate once engaged. In support of this, Nurmohamed et al. reported in their structure the presence of a conserved loop at the aperture of *E. coli* PNPase formed by residues F77-F78-R79-R80 (82). F77 was observed to make an aromatic stacking interaction with an RNA base (82). Nurmohamed et al. envisage the loop making successive contacts with the RNA strand as the strand is ratcheted through the aperture towards the active site. Other observations imply the functional importance of the PNPase central aperture residues, and mutational and biochemical studies in *E. coli* have supported this (51, 106). Equivalent residues within this aperture in other PNPases and the exosomes have also been predicted and observed to contact RNA (46, 68, 80, 116) (see Fig. 3b for the loop residues in GPSI that have been highlighted).

Interaction of PNPase with RNase E is a fundamental building-block of the RNA degradosome assembly. *C. crescentus* PNPase has been observed to interact with RNase E through surface interactions to form part of the *C. crescentus* degradosome complex (45, 46). As part of the RNA degradosome assembly in *E. coli*, RNase E forms, through a 20 residue micro-domain in its "C-terminal half", an extended β-sheet with an exposed edge of the PH' domain of
PNPase (82). It does not appear that the KH and S1 domains of PNPase are required for interaction with RNase E or degradosome formation (114). PNPase truncations lacking the KH, S1, or both these domains bind equally as well to RNase E as does full-length PNPase (114). However, the activities of the reconstituted minimal degradosomes containing these truncated PNPases are impaired.

Many biochemical and mutational studies have been designed to test the hypothesis that the KH and S1 domains of PNPase participate in RNA binding (16, 73, 106, 114). The dynamic nature of the domains, placement at the entrance to the catalytic core of PNPase, and RNA-binding properties of the individual domains all implicate KH and S1 in substrate interaction. Truncations in PNPase removing the KH and S1 domains have been made and these mutants have been shown to be impaired in RNA binding while retaining phosphorolytic activity (15, 16, 114). Shi et al. showed that the KH/S1 domains are not only involved in substrate binding, but also participate in conferring enzymatic activity in a way that is not yet fully understood (106).

In addition to being involved in RNA binding, it appears that the KH and S1 domains also function in stabilization of the doughnut-shaped trimer. Although these two domains do not appear to be necessary for degradosome formation (as discussed previously), biochemical and structural analysis suggests that they assist in the formation of the PNPase trimer and in regulation of central channel size (106).

These points in conjunction with structure-function observations discussed earlier are summarized succinctly in a model our lab proposed to account for the roles of the KH and S1 domains of E. coli PNPase (114). The model, detailed in Fig. 6, assumes that only one RNA strand is accommodated in the central channel of the trimer at a time even though theoretically, there are three core "binding" and active sites per PNPase trimer. The model proposes that two
RNA binding "platforms" exist within each monomer; one formed by the tandem KH/S1 domains, and one conferred by a site within the core channel. Conceivably, this core binding site in *E. coli* PNPase could be formed by, (or at least partly be attributed to), F77-F78-R79-R80, as supported by evidence presented earlier of the entrance-proximal aperture loop interacting with RNA (82). The initial step relies on the KH-S1 platform to initially engage and "capture" the substrate, allowing the RNA to form a second, stronger interaction with the core site which "retains" the interaction, conferring processivity and steering the substrate towards the active site. Phosphorolysis continues until stalling occurs, when it is proposed that another substrate must be engaged by an unoccupied KH-S1 site to displace the stalled strand.
Figure 6: Model for the role of KH-S1 in substrate binding, retention, and release.

**a.** Substrate recognition.  **b.** Substrate binding (weaker) to peripheral KH/S1 tandem binding site.  **c.** Initial interaction directs substrate entry into core channel where it forms second, (stronger) interaction with "core binding site"; phosphorolysis ensues.  **d.** Substrate stalls, but remains bound at core binding site.  **e.** Displacement of stalled substrate is promoted by binding of second substrate to KH-S1 site and subsequent threading of second substrate.

The PNPase trimer is simplified in this figure in its representation by only two subunits. Model is based on the work of Lorentzen & Conti, Büttner et al., and Stickney et al. (17, 68, 69, 114).
1.4.1 PNPase autoregulation in *E. coli*

PNPase post-transcriptionally regulates its own expression by destabilizing *pnp* mRNA (22, 52, 88, 89, 97, 98). The *pnp* gene is located downstream of *rpsO*, and transcription of *pnp* can be initiated at two different promoters: *rpsOp*; and downstream of this, *pnp*-p (89). Both primary transcripts are processed during PNPase autoregulation. The autoregulatory process is initiated by the action of RNase III, an endonuclease that cleaves at two sites (*RIII₁* and *RIII₂* in Fig. 7a) within a long stem-loop (SL1) in the 5'-UTR of the *pnp* transcript. The RNase III-dependent, staggered double-stranded cleavage results in a 37-nucleotide fragment (RNA37; Fig. 7) with an exposed 3'-OH overhang in a duplex with the rest of the processed *pnp* transcript (Fig. 7b, second drawing). Subsequent binding by PNPase to RNA37 and the resulting 3' to 5' degradation removes this protective 5' fragment of the duplex (Fig. 7b, third drawing), exposing the 5'-monophosphate of the remaining single-stranded *pnp* mRNA fragment to fast decay. It has been proposed that this last step in *pnp* mRNA autoregulation is carried out by RNase E, which may explain how the RNA37 of the unprocessed duplex protects *pnp* mRNA from fast decay; i.e. by obstructing access of the 5'-monophosphorylated end to RNase E (22). An alternative model for PNPase autoregulation postulates that PNPase binds determinants in the 5'-UTR of *pnp* mRNA to inhibit translation, thus rendering the transcript vulnerable to degradation (98). Although not ruling out the possibility that these non-mutually exclusive mechanisms both occur, recent data provide evidence that translational control by PNPase does not play a role in autoregulation (22).
Figure 7: Autoregulation of *pnp* mRNA.

**a. Map of the *pnp* gene with 5'- and 3'-untranslated regions in the context of *pnp* autoregulation.** 5'- and 3'-UTRs are drawn to scale, but the boxed *pnp* ORF is not (the TTG start codon is denoted by the "1" and the arrow below the vertical line). The Shine-Dalgarno AGGA sequence is shown. Coordinates above the vertical line represent nucleotide positions starting from the *pnp*-p transcription start point (bent arrow). Inverted arrows indicate the positions of the large and small stem-loops, SL1 and SL2, respectively. RIII1 and RIII2 mark the two RNase III cleavage points. The dashed brace defines the region of the 5'-UTR which corresponds to the RNA37 fragment of the processed *pnp* transcript.

**b. Proposed model for PNPase autoregulation.** Control of PNPase expression is initiated by endonucleolytic cleavage of *pnp* mRNA by RNase III (first, upper drawing). PNPase processes the 3' "RNA37" fragment of the resulting duplex (second drawing). RNase E is proposed to degrade the resulting single-stranded fragment (bottom drawing). Figures reproduced and modified with permission from (22).

The ability of PNPase to autoregulate its mRNA, thus repressing its own expression, is dependent on its enzymatic activity and RNA binding (16, 39, 51, 52, 96). PNPase mutants defective in either enzymatic activity or in RNA binding have been found to result in the over-expression of PNPase (39, 51, 96). This rationalizes the measurement of PNPase autoregulation as an assessment of both enzymatic and, in particular, RNA binding activity in PNPase.
Convenient reporter systems using chromosomal pnp-lacZ fusions have been developed and widely used for this purpose and allows for the convenient in vivo measurement of post-transcriptional auto-repression by PNPase (39, 51, 52, 73, 97, 98). The implications of KH and S1 domain involvement in PNPase autoregulation prompted us to construct a pnp::lacZ reporter strain [IBPC7332(λGF2); Materials and Methods, section 2.1] to be used in our mutagenic study (construction of mutagenic KH/S1 library detailed in Materials and Methods, section 2.2). Conditions were optimized for in vivo assessment of KH/S1 point mutants and we show that the pnp::lacZ reporter is an effective tool for evaluation of RNA binding activity in PNPase. The reporter system is detailed in Fig. 8.
Figure 8: Use of reporter strain IBPC(λGF2) to assess PNPase autoregulation.

The *pnp::lacZ* reporter strain used to assess PNPase autoregulation is *pnp* and *lacZ* null. It contains a chromosomal fusion gene encoding the entire *pnp* promoter, 5'-UTR, and the first 81 codons of *pnp* fused to *lacZ* expressing β-galactosidase. The resulting transcripts contain all the determinants (see Fig. 7) required for PNPase autoregulation. Thus, the expression of β-galactosidase is repressed by PNPase autoregulation. Vectors expressing WT or mutant PNPase are transformed into the reporter strain and β-galactosidase activities in transformant cells reflect autoregulatory abilities of respective plasmid-expressed PNP variants. Levels of β-galactosidase are quantified by assaying β-galactosidase activity using a modified Miller protocol [Materials and Methods, section 2.3; (74)]. The IBPC7332(λGF2) reporter strain was engineered by Dr. George Mackie with phage lambda provided by Dr. Matthias Springer of Institut de Biologie Physio-Chimique, Paris.
1.5 Overview of objectives and experimental scope.

Several aspects of KH and S1 domain function remain open to interpretation:

1. The functional implications of the KH and S1 domains in the context of its trimeric assembly with the rest of PNPase.
2. The impact of these domains on RNA degradosome function and RNA decay in general.
3. How structural elements and specific residues within these domains interact with RNA substrates.

These issues stem from the highly dynamic nature of the KH and S1 domains with respect to the RNase PH-like core rendering them difficult to visualize engaged with substrates. Some residues within these domains have been implicated in binding, but the mechanisms behind these interactions and whether or not additional residues participate in binding and contacting RNA remain important questions. To resolve these questions, we complemented in vitro activity assays assessing RNA binding and enzymatic activity with an in vivo assessment of autoregulation to identify and characterize putative RNA-binding residues and regions:

PNPase autoregulation has been shown to require both enzymatic and RNA binding activity (39, 51, 73). As such, we took advantage of a reporter strain expressing a pnp::lacZ fusion (Fig. 8). We hypothesized that mutations to key contact residues in the KH and S1 domains would interfere with PNPase autocontrol. Thus, measurement of resulting β-galactosidase activities in the presence of point mutations would allowed us to gauge their impact on RNA binding. After optimization, we tested this system on S1 domain mutants our lab had previously characterized in vitro, and found a striking relationship between RNA binding and autoregulation activity.
Next, we sought to use the reporter to identify additional residues in the KH and S1 regions which may be involved in substrate interactions. This was achieved by random mutagenesis and use of the reporter to screen for RNA binding-impaired mutants (Fig. 8; also see sections 2.2 and 2.3). We identified specific regions within the KH domain that appear to be involved in RNA binding, and targeted these regions with additional directed mutagenesis for further characterization.

Because a close relationship between PNPase enzymatic and RNA binding activity has been demonstrated from past studies, we also assayed phosphorolytic activity in our mutants of interest.

Together, our data present compelling evidence that the KH and S1 domains are important for RNA binding in the context of PNPase function, and that specifically, particular regions within these domains contribute substantially in ways that are discussed in this thesis. We discuss possible mechanisms governing these observations and present a model detailing the importance of KH-S1 cooperation with each other, and the rest of PNPase in light of recent crystallographic data (46, 61, 82, 106). Additionally, we confirmed that the widely-used pnp-lacZ reporter is indeed an excellent tool for assessment of both binding, and enzymatic activity in PNPase.
Chapter 2: Materials and Methods

2.1 Bacterial strains and plasmids

Strain IBPC7322 (thi-1, argE3, ΔlacX74, mtl-1, tss-29, rpsL, pnp::Tn5) from the collection of Dr. Claude Portier was obtained from Dr. Matthias Springer (Institut de Biologie Physico-chimique, Paris). It was lysogenized at 30°C with λGF2 containing a pnp::lacZ fusion (also from C. Portier via M. Springer) to create the β-galactosidase reporter strain IBPC7322(λGF2). Strain ENS134 [= BL21(DE3) pnp::Tn5] was provided by Dr. M. Dreyfus (Ecole Nomale Supérieure, Paris) and served as the host strain for purification of PNPase or its derivatives.

The plasmid pAW101 (Table 1) encoding WT PNPase under lac operator-promoter control, was constructed in several steps: With primers Kate-fprimer1 and -rprimer2 (See Appendix Table A1), the KH and S1 domains of WT pnp were cloned into pUC19 by way of engineered SalI and XbaI restriction enzyme recognition sites, creating plasmid p19khs1_nm (nm = not-mutated; Appendix Fig. A1, step 2a). The two core PH domains were cloned into pUC18 (p18phph; Appendix Fig. A1, step 1) and combined with the excised fragment from p19khs1_nm to regenerate full-length WT pnp in the correct orientation under the control of plac (pAW101; Table 1 and Fig. A1, step 2b). This plasmid served as the basis for all subsequent derivatives expressing the desired pnp mutant to be assayed in IBPC7322(λGF2) (Table 1, first column). The plasmid pGC400 [(26); Fig. A1, top] which encodes WT PNPase in the pETO11 backbone was the basis for all constructions for over-expression in strain ENS134 and subsequent purification (Table 2, first column; also see Materials and Methods, section 2.5). In this work, pGC400 was renamed pAW001 (encoding WT PNPase; Table 2, first column), and all subsequent derivatives of pAW001 (pAW013-pAW022, and pAW024; Table 2, first column)
encoded PNPase variants of interest. Site-directed mutants (sequences of primers listed in the Appendix, Table A2) were prepared by the "Quick-change" method (Invitrogen, Inc.) and were confirmed by DNA sequence analysis [Nucleic Acid Protein Service Unit (NAPS Unit), The University of British Columbia; and Genewiz, Inc.]. PCR cycling parameters used are detailed in Appendix, section A.1.

Many transformations were performed using "Library Efficiency" DH5α Competent Cells [Invitrogen Inc.; F− Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17(rk−, mk+) phoA supE44 thi-1 gyrA96relA1 λ−]. In addition, chemically competent "KCM" cells of desired strains were made in the lab using the Berglund KCM competent protocol: (http://openwetware.org/wiki/Berglund:KCM_competent; and http://openwetware.org/wiki/Transforming_chemically_competent_cells).

2.2 Construction of mutagenic KH-S1 library

DNA fragments containing the KH and S1 domains (amino acid residues 551-691; Fig. 4a) were amplified using the "Error-prone PCR with Tag DNA polymerase in mutagenic buffer" method [(93); see step 3a as depicted in Appendix Fig. A1]. The plasmid p19khs1_nm (see Materials and Methods, section 2.1) was used as a template along with primers Kate-fprimer1 and -rprimer2 (Appendix Table A1) to create a library of mutant KH-S1 fragments framed by SalI and XbaI sites. The concentration of MnCl2 was optimized to increase the frequency of fragments containing an ideal number of point mutations (between 1 and 3 substitutions). This was done by titrating the amount of MnCl2 added to the PCR reaction mix, cloning resulting PCR fragments into pUC19, and sequencing a few purified clones to manually count the number of mutations within the KH and S1 domains. The optimized reaction mix totaling 50 µL
combined 5U of *Taq* DNA polymerase, 0.3 μM each of the primers Kate-fprimer1 and -rprimer2, 200 μM each of the dNTPs, and approx. 90 ng of plasmid p19khs1_nm (see Appendix Fig. A1, step 2a) as template in regular PCR buffer (7.4 mM Tris-HCl [pH 8.3 at 25°C], 37 mM KCl, 1.1 mM MgCl₂, and 0.074% [v/v] Triton X-100) mixed with mutagenic PCR buffer (80 µM dTTP, 80 µM dCTP, 550 µM MgCl₂, and 50 µM MnCl₂) [all reagents listed are in final reaction mix concentrations]. Cycling parameters used are detailed in Appendix, section A.1. Resultant PCR products were digested with SalI and XbaI restriction enzymes and cloned into pUC19 creating a library of p19khs1* plasmids (Appendix Fig. A1, step 3b). These were transformed into competent DH5α and individual colonies were purified, grown to saturation, and plasmids extracted. DNA sequencing was performed to identify induced mutations in the KH and/or S1 domains [Nucleic Acid Protein Service Unit (NAPS Unit), The University of British Columbia; and Genewiz, Inc.]. Plasmids encoding interesting mutants were cleaved with SalI and NdeI and the resulting ~743 bp fragments containing mutated KH-S1 regions were subsequently cloned into p18phph (Appendix Fig. A1, step 3c) to generate plasmids in the pAW101 series (Table 1, first column). This reconstructed full-length *pnp* containing the respective mutation(s) in the KH and S1 domains. In all cases, following the transformation of ENS134, IBPC7322(λGF2), or DH5α strains, individual colonies were purified by re-streaking on selective media.

### 2.3 β-galactosidase assay

Cultures of reporter strain IBPC7322(λGF2) (Materials and Methods, section 2.1) were transformed with pAW101 or its derivatives (Table 1, first column), and individual colonies were purified on selective media. Cultures of purified transformants were grown at 30°C in Luria-Bertani (LB) medium supplemented with carbenicillin (100 μg/ml) and kanamycin (20
µg/ml). Samples were taken during mid-exponential growth (at an OD$_{600}$ of approx. 0.40-0.50) and assayed by Miller's method (74) with the following modifications: 0.3 ml samples were iced for 20 minutes then added to 0.7 ml Z buffer (74) with 2 drops of chloroform and 1 drop of 0.1% (w/v) SDS solution, and vortexed vigorously for 10 seconds. These mixtures were then incubated for 10 minutes at 30°C before the addition of 0.2 ml of 4 mg/ml o-nitrophenyl-β-D-galactopyranoside (ONPG) at time zero. The reactions were incubated for 15 minutes at 30°C before being quenched with 0.5 ml 1 M Na$_2$CO$_3$. Optical densities at 420 and 550 nm were subsequently recorded for the quenched reactions and used to determine Miller units (MU) of β-galactosidase activity (74). A minimum of three independent trials were performed for each assayed strain (Table 1, third column). The % repression values were calculated as follows: % repression = 

\[
\frac{[(\text{mutant pAW1xx MU} - \text{Empty Vector pUC18 MU})/(\text{WT pAW101 MU} - \text{Empty Vector pUC18 MU})]} \times 100.
\]

### 2.4 Analysis of PNPase levels in reporter strain IBPC7322(λGF2) by Western blot

Cultures of IBPC7322(λGF2) transformants that were assayed for β-galactosidase (Table 1) were also harvested for analysis by Western blot (Fig. 9c and d) once achieving an OD$_{600}$ of approx. 0.40-0.50. Volumes of 1 ml were collected by centrifugation and pellets re-suspended in 100 µl sample buffer [100 mM Tris-HCl, pH 6.8; 4% (w/v) SDS; 0.005% (w/v) bromophenol blue; 0.2 M DTT; and 20% (w/v) glycerol]. Cells were lysed by boiling for 5 minutes and portions of each extract (10 µl) were separated by electrophoresis, electroblotted onto nitrocellulose membrane, and probed with polyclonal rabbit anti-PNP antibodies (1:10,000) in a manner similar to that described in (99). Modifications include the use of IRDye 800CW Conjugated Goat (polyclonal) Anti-Rabbit IgG (LI-COR Biosciences; diluted 1:10,000) as the secondary antibody, and the use of a LI-COR Odyssey for detection.
2.5 Purification of PNPase and PNPase variants

All purifications were performed with untagged versions of PNPase over-expressed in host ENS134 and were performed as described previously (56) with the following changes: A French pressure cell was used for lysis and after dialysis and clarification, the S15 supernatant containing PNPase was applied to a column of Q Sepharose Fast Flow resin (GE Healthcare). Subsequent hydrophobic interaction chromatography was performed with Phenyl Sepharose 6 Fast Flow (high sub) matrix (GE Healthcare). SDS-PAGE of fractions eluted from each of the two chromatographic columns are shown in the Appendix, section A2.

2.6 Preparation of SL9A substrate and enzyme assays

Synthetic radiolabelled SL9A RNA [Fig. 11a; (112)] was utilized for all activity and band-shift assays. SL9A RNA was prepared by in vitro transcription of a modified derivative of pTZ18U linearized with XbaI (112). Transcription was performed with this template in the presence of α-\textsuperscript{32}P-CTP.

Assays of phosphorolytic activity were performed as previously described (114) except that the assay buffer contained 20 mM Tris-HCl, pH 7.5, 1.5 mM dithiothreitol, 1 mM MgCl\textsubscript{2}, 10 mM Na-phosphate (pH 7.5), 100 mM KCl, and 50 nM labeled SL9A RNA. Activity assays were initiated by the addition of the appropriate PNPase to give a final enzyme concentration of 2.5 nM.

RNA binding was assessed by electrophoretic mobility shifts largely as described by Stickney et al. (114) except that increasing concentrations of PNPase or its derivatives (0.25 nM to 150 nM) were incubated in 20 µl reaction buffer containing 10 mM Tris-HCl, pH 8.2, 0.1 mM EDTA, 80 mM NaCl, 1% glycerol and 0.01% dodecyl maltoside, and the labeled SL9A RNA.
concentration was lowered to 1 nM. Dilutions of purified PNPase and PNP mutants were made in the same buffer containing in addition 1 mM DTT and 25 µg/ml bovine serum albumin.
Chapter 3: Results

3.1 Both the S1 and KH RNA-binding domains facilitate PNPase autoregulation

PNPase autoregulation has been shown to require enzymatic activity and in particular, RNA binding, as mutants impaired in either result in de-repressed cellular levels of the enzyme (39, 51, 52, 96). We postulated that PNPase autoregulation would be affected by mutations to key RNA contact residues in the S1 and KH domains that would impact substrate binding. To this effect, we constructed the $pnp::lacZ$ reporter strain IBPC7322(λGF2) (Materials and Methods, section 2.1) to measure autoregulation by WT and various PNPase mutants in vivo (Materials and Methods, section 2.3). As detailed in Fig. 8, this strain contains a $pnp::Tn5$ insertion to inactivate the chromosomal $pnp$ gene. As a result, the expression of β-galactosidase from the chromosomal reporter was sensitive to autoregulation by the $pnp$ variant expressed from a vector of our choice (pAW1xx plasmids are listed in Table 1). The observed β-galactosidase activity was thus a measure of the ability of the plasmid-encoded $pnp$ mutant, with the desired allele, to exert autoregulation.
**Table 1: Autoregulation of pnp::lacZ message by PNPase mutants**

*E. coli* reporter strain IBPC7322(λGF2)

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>pnp Mutation</th>
<th>Activity (Miller units)</th>
<th>% Repression</th>
<th>Domain(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUC18</td>
<td>none</td>
<td>584 ± 44</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>pAW101</td>
<td>WT</td>
<td>24 ± 4</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>pAW102</td>
<td>ΔKH + ΔS1</td>
<td>535 ± 18</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>pAW103</td>
<td>ΔKH</td>
<td>479 ± 24</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>pAW104</td>
<td>ΔS1</td>
<td>427 ± 28</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>pAW110</td>
<td>I555T&lt;sup&gt;d&lt;/sup&gt;</td>
<td>44 ± 8</td>
<td>96</td>
<td>KH</td>
</tr>
<tr>
<td>pAW111</td>
<td>G570C/V679A</td>
<td>342 ± 28</td>
<td>43</td>
<td>KH, S1</td>
</tr>
<tr>
<td>pAW112</td>
<td>I576T/T585A</td>
<td>121 ± 9</td>
<td>83</td>
<td>KH</td>
</tr>
<tr>
<td>pAW113</td>
<td>G570C</td>
<td>467 ± 32</td>
<td>21</td>
<td>KH</td>
</tr>
<tr>
<td>pAW114</td>
<td>K571L</td>
<td>186 ± 23</td>
<td>71</td>
<td>KH</td>
</tr>
<tr>
<td>pAW115</td>
<td>K571Q</td>
<td>80 ± 3</td>
<td>90</td>
<td>KH</td>
</tr>
<tr>
<td>pAW116</td>
<td>I576A</td>
<td>119 ± 20</td>
<td>83</td>
<td>KH</td>
</tr>
<tr>
<td>pAW117</td>
<td>I576T</td>
<td>260 ± 25</td>
<td>58</td>
<td>KH</td>
</tr>
<tr>
<td>pAW118</td>
<td>F635A</td>
<td>80 ± 7</td>
<td>90</td>
<td>S1</td>
</tr>
<tr>
<td>pAW119</td>
<td>F638A</td>
<td>97 ± 11</td>
<td>87</td>
<td>S1</td>
</tr>
<tr>
<td>pAW120</td>
<td>H650A</td>
<td>136 ± 10</td>
<td>80</td>
<td>S1</td>
</tr>
<tr>
<td>pAW121</td>
<td>F635A/F638A/H650A</td>
<td>309 ± 24</td>
<td>49</td>
<td>S1</td>
</tr>
<tr>
<td>pAW122</td>
<td>F635R/F638R/H650R</td>
<td>135 ± 10</td>
<td>80</td>
<td>S1</td>
</tr>
<tr>
<td>pAW123</td>
<td>I576A/F638A</td>
<td>235 ± 40</td>
<td>62</td>
<td>KH, S1</td>
</tr>
<tr>
<td>pAW124</td>
<td>I576T/F638A</td>
<td>358 ± 18</td>
<td>40</td>
<td>KH, S1</td>
</tr>
</tbody>
</table>

**a.** β-Galactosidase activity expressed from a chromosomal pnp::lacZ fusion in IBPC7322(λGF2) was measured as described in Materials and Methods (section 2.3) and is the average of at least three independent trials.

**b.** % Repression was calculated as described in Materials and Methods (section 2.3) and is derived from the ratio of the difference in mutant and empty vector activity relative to the difference in WT and empty vector activity.

**c.** The domain(s) in which the pnp mutation(s) are found.

**d.** Blue text represents pnp mutants identified from mutagenic library.

In effect, the reporter gene is inversely sensing the relative activity of the PNPase variant expressed from the plasmid. A series of derivatives of IBPC7322(λGF2) was prepared by transformation with pUC18 or plasmids in the pAW1xx series [Table 1]. The resulting individual strains were grown individually from stock cultures and assayed for β-galactosidase activity. The measured activities are reported in Miller units [MU; (74)] in Figure 9a and Table
These MU values inversely reflect the activity and strength of autoregulation by the plasmid-encoded PNPase variant. To facilitate the comparison of repression efficiency, we also report data as calculated % repression in Table 1 (fourth column). These percentages were calculated by dividing the difference in measured activity (MU) between the strain in question and empty vector strain [IBPC7322(λGf2)/pUC18], by the difference in measured activity between the WT [IBPC7322(λGf2)/pAW101] and empty vector strain (see Materials and Methods, section 2.3 for the formula). The reporter strain transformed with empty-vector, pUC18 [IBPC7322(λGf2)/pUC18], generated the highest levels of β-galactosidase activity (584 ± 44 MU; Fig. 9a and Table 1). This reflects the complete de-repression of PNPase (0% repression). In contrast, strain IBPC7322(λGf2)/pAW101 expressing full-length, WT PNPase produced consistently low levels of β-galactosidase activity (24 ± 4 MU). This activity defines 100% repression as calculated by the process described above.
Figure 9: Repression of pnp::lacZ message in the reporter strain by KH and S1 mutants.

**a. Autoregulation activity by S1 Domain mutants** was assessed by β-galactosidase assays as described in Materials and Methods, section 2.3 and data is represented here as hashed bars denoting Miller units (MU) of activity (74). Higher activities corresponding to lower % repression (Table 1, fourth column) reflect a lower degree of autoregulation of the pnp::lacZ message with respect to the strain expressing WT PNPase (pAW101). Text in parentheses indicate the plasmid (pAW1xx) from which the mutant PNPase was expressed. Activities are the averages of at least three determinations.

**b. Autoregulation by KH Domain mutants.** Blue text represents mutants identified from the mutagenic library.

**c. A Western blot showing levels of expressed S1 Domain mutants in reporter strain IBPC7322(λGF2)** was performed as described in Materials and Methods, section 2.4. Typically, WT PNPase runs anomalously to the ~85 kDa position within SDS-PAGE gels (66). List of PNP mutants and corresponding plasmid numbers are shown in Table 1.

**d. Western Blot showing levels of expressed KH Domain mutants in reporter strain IBPC7322(λGF2).**
Having established a reliable assay for PNPase autoregulatory activity in vivo, we turned to assessing the role of the S1 domain in this process. The S1 domain is required for efficient PNPase autoregulation (51, 52). These observations together with findings previously obtained from our lab prompted us to propose a role for the PNPase S1 domain in initial substrate recognition and subsequent product displacement [Fig. 6; (114)]. Thus removal of the S1 domain from PNPase would likely impede the autocontrol process in the model of Carzaniga et al. [Fig. 7; (22, 52)]. Strain IBPC7322(λGF2)/pAW104 encoding PNPase completely lacking the S1 domain (ΔS1; Δ605-711) generated an average of 427 ± 28 MU (Table 1, third column and Fig. 9a), or 28% repression. This impairment in autoregulation parallels previous in vitro findings of loss in substrate affinity for the purified PNPaseΔS1 (114).

Although Bycroft et al. (18) predicted that F635 and F638 in the S1 domain were likely candidates for contacting RNA, the results obtained by Jarrige et al. (51) were equivocal. These authors reported that an F638G substitution reduced catalytic activity but had little effect on autoregulation of PNPase. Curiously, the double mutant, F635G/F638G, exhibited almost normal activity but with a partial loss of autoregulation (51). We endeavoured to clarify these findings by assessing autoregulation by S1 domain mutants with substitutions at these sites, and by investigating the properties of purified PNPase variants containing single or multiple point mutations in the S1 domain. Single mutations to alanine were introduced by site-directed mutagenesis at positions 635, 638, and 650 in the S1 domain creating pAW118, pAW119, and pAW120, respectively (Table 1). The relative positions of these three aromatic residues are shown in a modified image of the NMR structure of the S1 domain (Fig. 4b). Derivatives of IBPC7322(λGF2) containing plasmids pAW118-120 were assayed for β-galactosidase activity. Each of these strains exhibited similar levels of activity ranging from 80 to 136 MU (Table 1,
third column and Fig. 9a). These values represent small defects in autoregulation equivalent to 80 - 90% of the repression activity of WT PNPase.

Plasmid pAW121 (Table 1) encoding a pnp mutant with all three of the putative aromatic contact residues simultaneously mutated into alanine (F635A/F638A/H650A) was transformed into reporter strain IBPC7322(λGF2) and purified clones were assayed for β-galactosidase activity. This strain produced an average of 309 ± 24 MU of activity (Table 1, third column and Fig. 9a); over three times the activity of the singly-mutated F638A mutant in IBPC7322(λGF2)/pAW119 (97 MU, 87% repression). The activity of the F635A/F638A/H650A strain represents 49% the repression activity of WT PNPase. Interestingly, pAW122 encoding pnp with all three aromatic residues simultaneously mutated into arginine (F635R/F638R/H650R) produced an average of 135 ± 10 MU of activity in IBPC7322(λGF2) (Table 1 and Fig. 9a). This reflects relatively efficient autoregulation (80% of WT repression activity) and was in stark contrast to the triple-alanine mutant (135 MU versus 309 MU, respectively). At 80% repression, the triple-arginine mutant (pAW122; Table 1 and Fig. 9a) repressed the reporter to a level comparable to those of the single S1 point mutants (pAW118-120, 80 - 90% repression), effectively reversing the impact of the triple-alanine mutations on autoregulation. This reversal can be rationalized by considering that the substitutions at putative RNA contact sites (Fig. 4b) were converted from aliphatic residues into positively charged ones.

To assess the expression of each mutant pnp in reporter strain IBPC7322(λGF2), a Western blot of crude cell extracts was conducted (Fig. 9c). Almost identical levels of expressed WT or mutant PNPase were observed for each strain, confirming that relative PNPase abundance was not contributing to the differences in assayed β-galactosidase activity.
In order to sample the impact of additional residues in both the KH and S1 domains of PNPase, we turned to random mutagenesis PCR (see Materials and Methods, section 2.2 and Appendix, A1 for strategy). More than 60 independent clones from mutagenic PCR were examined by DNA sequencing. Three interesting KH domain mutants (reported in blue in Fig. 9b and Table 1) were identified and examined further. The first mutant investigated contained an Ile->Thr substitution within the KH domain at residue I555. Plasmid pAW110 encoding the I555T variant was transformed into reporter strain IBPC7322(λGF2) and assayed for β-galactosidase. This strain generated an average of 44 ± 8 MU, essentially autoregulating the pnp::lacZ message at WT levels (96% repression; Table 1, fourth column), and was used subsequently as an additional control. The second pnp mutant, encoded on plasmid pAW111, was transformed into strain IBPC7322(λGF2) and expressed full-length PNPase with substitutions at two residues: G570C; a Gly->Cys mutation in the first conserved G of the important GxxG loop (coloured red in Fig. 4c) within the KH domain; and V679A within the S1 domain. This transformant generated an average of 342 ± 28 MU of β-galactosidase activity (Fig. 9b and Table 1), equivalent to 43% WT repression activity. Interestingly, in another study, a similar, marked impairment in autoregulation was described for the pnp-71 mutant containing a G570D substitution (39). The third mutant identified from the mutagenic library, encoded on pAW112, contained substitutions at two residues within the KH domain (I576T and T585A; Table 1, blue text). The IBPC7322(λ.GF2)/pAW112 transformant produced an average of 121 ± 9 MU of β-galactosidase activity (Fig. 9b and Table 1), corresponding to 83% repression. This represents a small, but obvious defect in autoregulation similar to what was found for the S1 domain point mutants (pAW118-120; Fig. 9a and Table 1). Coincidentally, in their mutational analysis of PNPase, Jarrige et al. (51) found an I576N mutant to be impaired in autocontrol.
This, in conjunction with the aforementioned observations for \textit{pnp-71} \cite{39} and modeling of the KH domain (\textbf{Fig. 4cd}), prompted us to postulate that the G570C and I576T residues were responsible for the observed effects on autoregulation while V679A and T585A were likely benign. To test this idea, we used site-directed mutagenesis to eliminate the V679A and T585A mutations and construct single point mutants in full-length \textit{pnp} containing either G570C, I576A, or I576T (pAW113, pAW116 and pAW117, respectively; \textbf{Table 1}). Two of these KH domain substitutions, G570C and I576T (\textbf{Fig. 9b}, black text), resulted in further reductions in \% repression (pAW113 and pAW117; \textbf{Table 1}) compared to the respective original mutants identified from the mutagenic library (G670C/V679A and I576T/T585A; \textbf{Fig. 9b} and \textbf{Table 1}, blue text). The IBPC7322(\textit{λGF2})/pAW113 transformant encoding the G570C variant produced an average of $467 \pm 32$ MU equivalent to only 21\% WT repression activity. I576T encoded by the IBPC7322(\textit{λGF2})/pAW117 transformant assayed at $260 \pm 25$ MU, equivalent to 58\% repression. In contrast, I576A encoded by the IBPC7322(\textit{λGF2})/pAW116 transformant had a smaller effect on autoregulation, producing an average of $119 \pm 20$ MU comparable to the $121 \pm 9$ MU generated by the original I576T/T585A variant (blue text; \textbf{Fig. 9b} and \textbf{Table 1}).

The aforementioned effects of a cysteine substitution at G570 of the conserved GxxG loop (coloured red in \textbf{Fig. 4c}) can be rationalized in light of the structure of the KH domain in other proteins (discussed in \textbf{section 1.3.2.1}). An inherent feature of KH domains, the canonical GxxG sequence has been shown to interact with RNA in \textit{C.crescentus} PNPase \cite{46} and affect RNA affinity in \textit{E. coli} PNPase \cite{37, 39}. These observations together with the apparent positioning of the loop as demonstrated by the predicted structure of the KH domain (\textbf{Fig. 4cd}) suggested to us that the lysine in the second position of GxxG would also be important for RNA binding by \textit{E. coli} PNPase. To investigate this possibility, we used site-directed mutagenesis to
construct pAW114 and pAW115, respectively encoding the K571L and K571Q variants (Table 1). Both plasmids were transformed into reporter strain IBPC7322(λGF2) and assayed for β-galactosidase activity. The IBPC7322(λGF2)/pAW114 transformant encoding the K571L mutant produced an average of 186 ± 23 MU of activity, equivalent to 71% WT repression activity (Fig. 9b and Table 1). The IBPC7322(λGF2)/pAW115 transformant encoding the K571Q mutant was slightly less affected, generating an average of 80 ± 3 MU; equivalent to 90% repression. This small deficiency in autocontrol is similar to those seen for the three aromatic S1 point-mutation and I576A variants (ranging from 80 to 136 MU of activity; pAW118-120 & 116; Table 1 and Fig. 9ab).

Previously, our laboratory proposed a model to describe substrate recognition & processing, and product release as a concerted effort between the peripheral KH and S1 domains of E. coli PNPase [Fig. 6; (114)]. To further investigate the idea of a tandem KH-S1 RNA binding platform, we used site-directed mutagenesis to construct pAW123 and pAW124, each encoding simultaneous mutations in the KH and S1 domains (Table 1). Interestingly, IBPC7322(λGF2)/pAW124 encoding the I576T/F638A double-mutant generated an average of 358 ± 18 MU of β-galactosidase activity, nearly 1.5-fold the activity of the singly-mutated I576T variant (260 ± 25 MU; pAW117 in Fig. 9b and Table 1, third column). This was equivalent to autoregulation at 40% WT repression activity compared to 87% and 58 % for the single point mutants (F638A and I576T respectively; Table 1, fourth column). Likewise, IBPC7322(λGF2)/pAW123 encoding I576A/F638A, was also impaired in autoregulation, albeit less severe, repressing at 62% WT activity. This suggests that the two mutations (one in the KH, and one in the S1 domain) are acting synergistically to impair autoregulation.
As with the S1 domain variants, the expression of each KH variant in reporter strain IBPC7322(λGF2) was assessed by Western blot of crude cell extracts (Fig. 9d). Similarly, almost identical levels of expressed WT or mutant PNPase were observed for each of strain, confirming that relative PNPase abundance was not contributing to the differences in assayed β-galactosidase activity.

### 3.2 RNA binding by S1 domain mutants

Mutant PNPases warranting further investigation were purified to near homogeneity using two chromatographic steps and phosphorolysis to remove bound RNA (see Materials and Methods, section 2.5). We employed electrophoretic mobility shift assays using a radioactive model RNA, SL9A (see Materials and Methods, section 2.6, and Fig. 11a), to assess RNA binding by purified, full-length PNPase containing the various KH and S1 domain mutations. $K_d$ values were estimated from quantifying the fraction of shifted SL9A and are listed in Table 2 (fourth column). Radiographs of typical mobility shift experiments are shown in Figure 10. WT PNPase complexed with SL9A at relatively low concentrations with an apparent $K_d$ of 0.6 ± 0.07 nM [$n = 4$] (Table 2). In contrast, the singly-mutated PNPase F638A was roughly 4.5-fold deficient in RNA binding compared to WT (Table 2), with an apparent $K_d$ of 2.7 ± 0.05 nM [$n = 3$]. Consistent with the apparent defect in autoregulation represented by β-galactosidase activity (pAW121; Table 1), the triply-mutated PNPase F635A/F638A/H650A was also significantly impaired in its ability to shift SL9A compared to WT with an apparent $K_d$ of 7.7 ± 1.3 nM [$n = 3$]; almost three times the apparent $K_d$ calculated for the singly-mutated F638A (2.7 ± 0.05 nM), and over 12-fold the $K_d$ calculated for WT PNPase (0.6 ± 0.07 nM; Table 2). Also mirroring findings from the β-galactosidase assays (pAW122; Table 1), the triply-mutated PNPase F635R/F638R/H560R appeared to partially reverse the RNA binding defect observed for the
triple-alanine mutant with an apparent $K_d$ of $2.7 \pm 0.6$ nM [$n = 3$] (Table 2). This represents an affinity for SL9A nearly identical to that of PNPase F638A (pAW019; Table 2, fourth column). Interestingly, at concentrations of 7.5 nM or higher, WT PNPase was observed to form a second complex (asterisk in Fig. 10b) that has been reported for band shifts in other work, also at higher concentrations of PNPase or PNPase derivative [complex II-R in (37); (114)]. We believe this to be a mixture of protein and SL9A-protein aggregates of unclear composition. The same was not observed in the band-shift assays for any of the other PNPase derivatives.
Figure 10: *In vitro* RNA binding activity of WT PNPase and representative PNP mutants.

Electrophoretic mobility shift assays were performed with 1 nM SL9A as the substrate as described in Materials and Methods, section 2.6. **a.** Representative binding curves as determined from a minimum of three independent trials. $K_d$ values listed in Table 2, column 4 were calculated from plots of fraction SL9A shifted vs. increasing concentrations of PNPase or PNP mutant. Curves for WT and I576T appear truncated relative to G570C because higher concentrations of G570C were required to form a complex with SL9A. **b.** c. and d. Concentrations in nM of respective PNP assayed is shown below each lane. The positions of substrate (S) and retarded complex (C) after separation by non-denaturing electrophoresis are shown in the centre margin. The asterisk shows a putative aggregate in panel (b).
Table 2: *In vitro* assays of PNPase and PNPase mutants

*E. coli* expression strain ENS134

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Enzymea</th>
<th>Relative rates of phosphorolysisb</th>
<th>Apparent K&lt;sub&gt;d&lt;/sub&gt;d (fold WT K&lt;sub&gt;d&lt;/sub&gt;)</th>
<th>Domain&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAW001</td>
<td>WT PNPase</td>
<td>0.45 ± 0.04c (100%)</td>
<td>0.6 ± 0.07 nM (0)</td>
<td></td>
</tr>
<tr>
<td>pAW013</td>
<td>G570C</td>
<td>0.19 ± 0.05 (42 ± 11 %)</td>
<td>6.9 ± 2.6 nM (11.5)</td>
<td>KH</td>
</tr>
<tr>
<td>pAW014</td>
<td>K571L</td>
<td>0.47 ± 0.11 (104 ± 24 %)</td>
<td>2.4 ± 0.3 nM (4.0)</td>
<td>KH</td>
</tr>
<tr>
<td>pAW015</td>
<td>K571Q</td>
<td>0.40 ± 0.03 (88 ± 7 %)</td>
<td>1.6 ± 0.2 nM (2.7)</td>
<td>KH</td>
</tr>
<tr>
<td>pAW016</td>
<td>I576A</td>
<td>0.24 ± 0.03 (53 ± 7 %)</td>
<td>2.0 ± 0.2 nM (3.3)</td>
<td>KH</td>
</tr>
<tr>
<td>pAW017</td>
<td>I576T</td>
<td>0.27 ± 0.06 (60 ± 13 %)</td>
<td>7.0 ± 1.0 nM (11.7)</td>
<td>KH</td>
</tr>
<tr>
<td>pAW019</td>
<td>F638A</td>
<td>15%f</td>
<td>2.7 ± 0.05 nM (4.5)</td>
<td>S1</td>
</tr>
<tr>
<td>pAW021</td>
<td>F635A/F638A/H650A</td>
<td>0.13 ± 0.04 (29 ± 9 %)</td>
<td>7.7 ± 1.3 nM (12.8)</td>
<td>S1</td>
</tr>
<tr>
<td>pAW022</td>
<td>F635R/F638R/H650R</td>
<td>0.13 ± 0.03 (29 ± 7 %)</td>
<td>2.7 ± 0.6 nM (4.5)</td>
<td>S1</td>
</tr>
<tr>
<td>pAW024</td>
<td>I576T/F638A</td>
<td>0.14 ± 0.02 (31 ± 4 %)</td>
<td>6.2 ± 1.3 nM (10.3)</td>
<td>KH, S1</td>
</tr>
</tbody>
</table>

a. Each mutant PNPase was purified to near homogeneity after over-expression from respective plasmid (pAW0xx) as described in Materials and Methods, section 2.5.

b. Activity assays were performed as described in Materials and Methods, section 2.6. Data are reported as initial rates in units of pmoles/min SL9A processed by 100 fmoles enzyme, and as (% WT activity); (see footnote c).

c. 100% (WT) activity corresponds to 0.45 pmoles/min of the SL9A substrate processed by 100 fmoles PNPase.

d. Apparent dissociation constants (K<sub>d</sub>) were calculated by quantifying and plotting shifted/total SL9A against enzyme concentration and determining concentration required for a 50% shift. Electrophoretic mobility shift assays were performed as described in Materials and Methods, section 2.6.

e. The domain(s) in which the PNPase mutation(s) are found.

f. Data from Kristina McBurney; error not reported.

### 3.3 RNA binding by KH domain mutants

A similar pattern demonstrating loss in affinity for SL9A was observed for the purified KH domain mutants. Mutations at two KH residues, K571 and I576, were introduced by site-directed mutagenesis into full-length PNPase and the respective purified enzymes were assayed by band shift (Materials and Methods, section 2.6). Interestingly, different mutations at I576 resulted in significantly different apparent K<sub>d</sub>'s, consistent with β-galactosidase assays (pAW116 and pAW117; Fig. 9b and Table 1). PNPase I576T yielded an apparent K<sub>d</sub> of 7.0 ± 1.0 nM [n = 3] (Table 2), compared to an apparent K<sub>d</sub> of 2.0 ± 0.2 nM [n = 3] for PNPase I576A; a more than
3-fold difference. PNPase G570C exhibited an apparent $K_d$ of $6.9 \pm 2.6$ nM [$n = 3$], nearly 12-fold that of WT PNPase (Table 2). At the second position of the GxxG sequence, the introduction of a leucine mutation (PNPase K571L) resulted in an apparent $K_d$ of $2.4 \pm 0.3$ nM [$n = 3$], 4-fold that of WT PNPase (Table 2). A glutamine substitution at the same site (PNPase K571Q) was milder, resulting in an apparent $K_d$ of $1.6 \pm 0.2$ nM [$n = 3$].

To investigate the contributions of KH and S1 within the context of an extended binding platform, pAW024 (Table 2) was constructed. From this, full-length PNPase was purified containing simultaneous substitutions I576T in the KH domain, and F638A in the S1 domain. The double mutant PNPase I576T/F638A exhibited an apparent $K_d$ of $6.2 \pm 1.3$ nM [$n = 3$], over 10-fold that of WT PNPase. This reduction in affinity for SL9A for the double mutant and the other various KH mutants compared to WT PNPase, as demonstrated by the apparent $K_d$ values (Table 2), appear to correlate well with the loss in autoregulation as measured by the β-galactosidase assays discussed earlier (see Fig. 12).

### 3.4 Enzymatic activity of KH, S1 domain mutants

We assayed purified WT and mutant PNPase for phosphorolytic activity (see Materials and Methods, section 2.6) against a model substrate, SL9A, containing 3' unstructured tails [Fig. 11a; (112)]. Shortening of the poly(A) tail of 30 A residues 3' of a 9-bp stem-loop to the stalled, limit product results in a product of ~55 nt stalled 7 residues from the base of the stem-loop (red arrows in Fig. 11c-f). Conversion of substrate to this limit product was taken as a measure of PNPase enzymatic activity (112, 114). Typical time courses of digestion of SL9A are shown in Fig. 11c-f. Reaction times in all cases were 30 minutes as further incubation resulted in negligible additional product accumulation (data not shown). WT PNPase (100 fmoles)
shortened SL9A RNA at an average initial rate of 0.45 pmoles/min. This value is taken as 100%; the measured rates for mutant enzymes are expressed relative to WT (Table 2, column 3).

For S1 domain mutants, the effects on phosphorolytic activity were either moderate (greater than 50% loss in initial rates compared to WT), or severe (greater than 75% loss). The two triply-mutated S1 domain mutants (F635A/F638A/H650A and F635R/F638R/H650R; Table 2, third column) were only moderately impaired in phosphorolysis compared to WT (FFH->AAA; Fig. 11b and f). Interestingly, the only S1 domain mutant that was severely impaired in enzymatic activity was PNPase F638A (pAW019; Table 2). This PNPase mutant containing a single point substitution at one of the aromatic S1 residues proposed to contact RNA [Fig. 4b; (18)] exhibited 15% WT rates. A similar result was reported by Jarrige et al.(51) for a PNPase mutant substituted at F638. In that study, F638G exhibited only 20-30% WT activity levels.
Figure 11: Assays of SL9A RNA processing by PNPase and its derivatives.

a. Schematic diagram of synthetic SL9A with secondary structure as described in (112). "X" at the 5' end represents the RNA sequence 5'-pppGGGAAUUCGACUCGGUAC. Red arrow shows point at which processing by PNPase stalls. 
b. Product accumulation curves and typical time courses of activity represent pmoles of SL9A processed (total = 50 nM) by 100 fmols of indicated purified enzymes (added at time zero to 2.5 nM) - c, WT PNP; d, K571L; e, I576T; f, F635A/F638A/H650A. Activity assays were performed as described in Materials and Methods, section 2.6, and were repeated at least once. Aliquots (4 µl) of incubations were quenched at various times with 3 volumes 90% formamide and were loaded onto 7% polyacrylamide gels containing 8M urea in Tris-borate-EDTA buffer. Accumulated, ~55-nt SL9A product (red arrows) was quantified using a PhosphorImager (Molecular Dynamics) and normalized to total SL9A in the * lane (no-enzyme added control). Digestion incubation times in minutes are shown below each lane.
In contrast, mutations to residues in the KH domain were found to affect phosphorolytic activity either mildly (up to 50% loss in initial rates compared to WT) or moderately (between 50-75% loss), and in one case, resulted in activity at WT rates (K571L; Table 2, third column). PNPase K571L exhibited essentially WT activity (Fig. 11b and d). Both PNPase I576A and I576T exhibited initial rates and time courses of digestion that were mildly affected compared to WT [(Table 2, third column); (for I576T: Fig. 11b and e)]. PNPase G570C was moderately impaired compared to WT activity (42 ± 11% WT initial rates, Table 2; time course not shown). The doubly-mutated PNPase I576T/F638A also exhibited a moderately-impaired initial rate of phosphorolysis (31 ± 4% WT rates, Table 2) in addition to generally lowered rates of activity over a 30 minute time-course (Fig. 11b).
Chapter 4: Discussion

4.1 PNPase autoregulation activity strongly correlates with *in vitro* RNA binding activity

Past investigations have tied PNPase autocontrol to RNA binding (39, 51, 73). The presence of phosphorolytic activity coupled to a loss in autocontrol is taken to suggest that autoregulatory impairment is due to a loss in the ability of that particular PNPase mutant to bind RNA. In this study, we provide direct evidence that the affinity of a particular PNPase mutant for a synthetic RNA substrate (SL9A) *in vitro* correlates with its ability to autoregulate *in vivo*. We found a remarkably strong correlation between *in vitro* affinity for SL9A and repression (Fig. 12), with a Pearson r value of 0.892 for ten sets of data.

![Figure 12: Correlation between RNA binding and PNPase autoregulation.](image)

Repression by PNPase and PNP mutants represented by β-galactosidase activities measured as described in section 2.3 is plotted against apparent *in vitro* affinities for synthetic model substrate, SL9A. Statistics of measure of correlation are shown in the table to the right.

Moreover, the difference observed between full repression (WT PNPase, 100% with 24 ± 4 MU of activity) and near lack of autoregulation (ΔKH + ΔS1, 9% repression with 535 ± 18 MU activity) demonstrates a large dynamic range of activity allowing us to infer different degrees of
impairment in RNA binding conferred by different types of mutations. This is exemplified best by the differences in repression observed for the substitution by different residues at the same site (i.e. I576T and I576A; K571L and K571Q; see Table 1 and Fig. 9b). These apparently subtle changes may alter the local binding environment of a domain, or conceivably, affect the electrostatics of a base or ribonucleotide backbone interaction. Along the same lines, the reporter could distinguish differences in folding at the domain, or even, quaternary level. For example, a rather harmful mutation resulting in near complete unfolding or elimination of a domain resulted in a ≥ 90% decrease in repression activity (ΔKH + ΔS1), whereas a single mutation at a putative RNA contact residue resulted in a mere 10% decrease in activity (F635A; see pAW118, Table 1).

4.2 Plasticity of RNA contacts in the S1 domain

To that effect, the simultaneous mutation of all three of the putative RNA contact residues in S1 presented an interesting case in our study (F635A/F638A/H650A versus F635R/F638R/H650R). A rather significant difference in repression activity (49% WT repression for FFH->AAA versus 80% for FFH->RRR) was observed (see pAW121 and pAW122 in Table 1). These results suggest that these three aromatic residues on the surface of the S1 domain are directly contacting RNA, perhaps as part of an extended binding site. The elimination of aromatic contacts by the alanine substitutions is "rescued" by the introduction of electrostatic interactions with the RNA backbone afforded by the arginines (Fig. 13). If these residues were playing a strictly structural role, perturbation of S1 domain structure would intuitively be more pronounced in the bulkier FFH->RRR mutant compared to mutations to alanine. In corroboration, a trypsin sensitivity assay of these two mutants resulted in similar cleavage patterns: one band corresponding to the N-terminal core PNP fragment with the KH and
S1 domains removed and another, at higher excesses of each mutant PNP, corresponding to PNPase with the S1 domain removed, suggesting that at the very least, the triple mutations are not affecting the overall folded structures of the KH and PH'-PH domains (Appendix Fig. A3). Moreover, purification of FFH->RRR and FFH->AAA eluted large quantities of homogenous products at the expected fractions and of expected size, suggesting proper folding of these mutants and the absence of aggregation (data not shown).

Figure 13: Contributions to RNA binding by F635/F638/H650 from the S1 domain.
Schematic showing the plasticity of RNA binding contributed by surface residues within the S1 domain. The cognate aromatic binding mode can be replaced by an electrostatic binding mode along the same ligand-contacting path. Refer to Fig. 4b for the positioning of F635/F638/H650.
4.3 RNA binding regions within the KH domains were identified with novel insights into mechanisms of RNA binding impairment

In this study, we screened for additional mutations to PNPase RNA binding domains that affect RNA binding. We hypothesized that mutations in other key contact residues of the KH and S1 domains would impair PNPase autoregulation. Two candidate mutants from a mutagenic library were identified by sequencing. Indeed, with the convenient pnp::lacZ reporter fusion (Fig. 8), we show that reporter cells expressing these mutants were impaired in the autoregulation of the pnp::lacZ message compared to those expressing WT PNPase (blue text in Table 1 and Fig. 9b). Because each of these mutants contained two substitutions, we employed site-directed mutagenesis to construct PNP mutants containing substitutions at only one putative RNA contact residue. Our findings of substantial impairment in the autoregulatory ability of the G570C pnp mutant (pAW113, Table 1) corroborates with published data describing the effects of the autoregulation-deficient PNPase G570D expressed in the pnp-71 mutant (37, 39). As well, our data for the autoregulation-impaired I576T and I576A mutants agree with the observations for the autocontrol-impaired I576N mutant investigated by Jarrige et al. in their mutational study of E. coli PNPase (51). Also, the residue equivalent to I576 in the KH3 domain of Nova-2 (Leu-28) was observed to provide an aliphatic stacking interaction with a base of the bound RNA and was accordingly designated a contact residue as part of an "aliphatic" binding platform (60). Moreover, the equivalent mutation in the second KH domain of FMRP (I304N) results in a particularly devastating form of fragile X mental retardation syndrome [reviewed in (122)]. This suggests that I576 and equivalent residues in the KH domains of other proteins play an important role either in direct substrate contact, or by domain fold stabilization (see Fig. 14)
for comparison of an RNA-bound KH domain to the homology model performed as part of this work and first presented in Fig. 4cd).

Figure 14: Comparison of RNA-bound Nova-2 KH3 to predicted structure of the KH domain in *E. coli* PNP. Figure on right depicting structure of Nova-2 KH3 bound to SELEX RNA [PDB I.D. 1EC6 originally published by (60); reproduced with permission from (122)]. Protein side-chains providing polar and hydrophobic contacts with RNA are shown in cyan and green, respectively. Structure is positioned in same orientation as the predicted structures from Fig. 4cd as shown on the left. GxxG loop is red in the predicted structure (left) and cyan in the structure on the right.
The identification and subsequent analysis of the G570C mutant offered more insight into the dynamics of the KH domain, and in particular, the conserved GxxG loop [discussed in section 1.3.2.1, reviewed in (122), and depicted in Fig. 4c]. This mutation to the first glycine of the conserved motif may interfere with the binding-interface between the KH domain and the RNA substrate (see Fig. 4cd and Fig. 14). As reported in a crystal structure of the Nova antigen-2 KH3 domain, four bases of a tetra-loop are accommodated in a binding cleft formed by the GxxG motif in conjunction with a platform of hydrophobic residues and the variable loop region [observed in (60) and reviewed in (122)]. We found the G570C mutation to have a pronounced effect on in vitro RNA binding activity and autoregulation (see pAWx13, Tables 1 and 2). Moreover, the effects of the two different mutations at the same site within GxxG, K571, confirmed that the loop indeed plays an important role in RNA binding. Although the effects were less pronounced than those observed for G570C, the results found for our K571 mutants can nonetheless be compared to the findings of a study by Hollingworth et al. These authors reported that a GDDG double mutant impairs nucleic acid binding without compromising KH domain stability (48). Thus, they proposed that GDDG can be used as an efficient biophysical tool to assess KH domain function. As discussed in section 1.3.2.1 and demonstrated by our data, the absence or impairment of the GxxG loop is accompanied by loss in RNA binding [(79, 84) and pAWx13, Tables 1 and 2].

Additionally, these results confirm that the KH domain most likely plays a role in direct contact with the RNA rather than just a structural role in positioning of the S1 domain appropriately. This is exemplified by the work of Lewis et al. (Fig. 14, right), in which they described a "molecular vise" comprising of the GxxG loop, a platform of aliphatic residues, and the variable loop region clamping onto the RNA substrate (60). One could also envision that
once the RNA is engaged (or clamped onto), the KH domain rigidifies to stabilize the interaction.

The same could perhaps be said for the roles of individual regions and residues within the KH domain. While Valverde et al. argued that mutations equivalent to I576T in other proteins (i.e. FMRP, or Fragile X Mental Retardation Protein) result in destabilization of the KH domain, they conceded that in other contexts, it is just as likely that the particular residue directly contacts RNA substrates (122). Lewis et al. maintained that the I576 equivalent in Nova-2 KH3 (Leu-28) indeed contacts RNA via an aliphatic stacking interaction [(60) and see Fig. 14]. In our study, the differences in effects seen for I576 by substitution at the same site with two different residues suggests the same. Moreover, a third I->T substitution within the KH domain identified from our mutagenic library, I555T, behaves like WT PNPase (blue text; Fig. 9b and Table 1). Also, it appears that I576T, more impaired than I576A, folds natively based on two observations. First, the level of impairment in auto-repression activity (58% WT levels for I576T; Table 1, fourth column) is not so drastic as to suggest unfolding of the KH domain. Secondly, figures in Appendix, section A.2 show that I576T elutes from each of two chromatographic columns (section 2.5) in high yield with the native WT PNPase profile, suggesting the absence of aggregates as a result of improper folding. Still, we cannot say for certain from this study whether substitution at I576 of the KH domain in E. coli PNPase affects secondary and tertiary folding. CD spectroscopy of the KH domain containing I576T may offer insight into this question and native-gel electrophoresis would probe the ability of PNPase I576T to trimerize.
4.4 Non-equivalency in RNA binding by the KH and S1 domains: a necessity for PNPase function?

Past work from this lab resulted in the proposal of a model detailed in Fig. 6 in which two binding sites exist in PNPase; one of which is a binding platform created by the KH and S1 domains cooperatively, with weaker affinity for RNA substrates compared to the (second) "core" binding site (114). Based on data gathered from this study in conjunction with a recent crystal structure of *C. crescentus* PNPase with the KH domains clearly resolved (46), we propose an improvement of this model (Fig. 15).

The combination of KH and S1 in PNPase as the arrayed extended binding platform may have become established evolutionarily as a result of binding affinity disparity. This model precludes the notion that fixed "contact" residues exist in the S1 and particularly, the KH domains of PNPase. Instead, it is because of a weaker, more generalized binding that PNPase is able to function in processive degradation. Hardwick *et al.* proposed that the KH domains of *C. crescentus* PNPase form non-equivalent interactions with its RNA substrate as the polynucleotide threads through the central core (46). One could imagine the core binding site binding with higher affinity to the strand as it propagates through the channel, facilitated by a "ratcheting" action of the three KH domains guarding the entrance to the core (Fig. 15, right schematic). This "hands gripping a rope" model would likely require binding by the KH and S1 domains in a "slippery" fashion (i.e., reversible); a binding mode in which there is just enough initial contact to capture the RNA substrate and direct it to the core of PNPase. Once engaged, the reversible low-affinity binding afforded by KH and S1 and the mechanical-ratcheting motion assists the propagation of the strand through the enzyme core.
Figure 15: Model proposing greater role of KH in PNPase-RNA interaction and PNPase function.
An evolution of the model proposed by Stickney et al. [(114); Fig. 6] (left schematic) showcasing importance of the KH domain (right schematic). Asterisks denote presence of putative substrate contact points within PNPase. Ratcheting by the KH domains and conformational changes at the core aperture regulated by KH-substrate interactions are represented by the ( and bi-directional arrows, respectively.

Support for this model stems from a few different, but corroborating observations. First, many different combinations of KH and S1 assemblies exist in nature primarily within proteins that interact with RNA [reviewed in (122)]. Although the difference in repression activity of ΔKH versus ΔS1 is minimal as demonstrated by our pnp::lacZ reporter (19% WT repression activity versus 28%, respectively), our results do suggest that ΔKH is more affected. Moreover, out of all the point mutations tested, KH domain point mutants generally resulted in greater impairments in both autoregulatory and RNA binding activity. Perhaps PNPase acquired the S1
domain from natural selection, as an intermediate binding affinity was required; two KH
domains in succession would result in "gripping" of the RNA strand too tightly resulting in
reduced activity and prevention of the ratcheting mechanism. NusA [with one S1 domain linked
to two KH domains; (12) and reviewed in (122)] was observed contacting its RNA substrate only
through the two KH domains. The S1 domain of NusA was not observed contacting the
substrate in the crystal structure (12), suggesting that it either has a lower affinity compared to its
KH counterparts, or that the substrate prefers a tandem KH site over a KH-S1 tandem site.
(Although we should also not rule out positioning affects in the context of the entire protein).

This model could conceivably be tested by artificially linking a second KH domain to
PNPase effectively "swapping" out S1 for another KH domain. For such a fusion, we could use
the convenient β-galactosidase reporter system optimized in this study to assess autoregulation in
addition to characterization by established in vitro assays. A scenario in which the KH-swapped
mutant is minimally, or not affected in in vitro binding activity coupled to a decrease in
autoregulatory ability (invoking the requirement of both enzymatic and binding activity for
PNPase autoregulation) would suggest that this model may be appropriate. Our model of
disparity in RNA affinities for the KH and S1 domains in addition to the ratcheting mechanism
proposed by Hardwick et al.(46) fit well with the data from this study and those from recent
structure determinations in addition to past findings. Observations from these studies suggest
that the KH and S1 domains communicate with the conformationally dynamic core and also
implicate the KH domain in playing a larger role than S1 (61, 82, 106). Intriguingly, Shi et al.
observed that the KH and S1 domains were necessary for "tightening" of the E. coli PNPase core
(106). The same notion was proposed by Nurmohamed et al. who surmise that the KH and S1
domains may contribute to structural changes at the core aperture (82). Lin et al. reported that
conformational changes to the narrowest region of the central channel in human PNPase (hPNPase) were expected upon RNA binding as the apo-core channel is not wide enough to accommodate single-stranded RNA (61). In addition, they observed the KH domains of hPNPase forming a pore, effectively extending the RNA funnel formed by the core channel, and implicated the KH domains in trimer formation and stability. Their biochemical experiments suggested that the S1 domain of hPNPase is not critical for RNA binding and cleavage, and that instead, the conserved GxxG motif of the KH domain directly participates in RNA binding in hPNPase (61). Another finding which may support the ratcheting mechanism of RNA propagation through PNPase was the observation that a sufficiently long 3' overhang was necessary for the substrate to be threaded into the RNA binding pore and cleavage channel to be digested by hPNPase, as RNAs that were only transported by hPNPase possessed short or no 3' overhangs (61). It should be noted that superimposition of ∆S1 hPNPase with the core of S. antibioticus PNPase (PDB I.D. 1E3P) and the core of E. coli PNPase (PDB I.D. 3CDI) gave an average RMSD of 1.38Å for 374 Ca atoms (61).

My thesis research showed that RNA binding can accurately be assessed by use of a pnp::lacZ reporter fusion. RNA binding activity correlates well with PNPase autoregulation activity. In addition, our study has provided insight towards a more complete picture of how the KH and S1 domains of PNPase function and interact with RNA substrates. Specific residues directly involved in RNA-contacting may function in a heterogeneous manner in a mechanism of engagement by KH and S1 that warrants further investigation. However, it is clear that particular regions within these domains impact PNPase-RNA substrate interactions, and may affect tertiary and quaternary folding.
References


2. Andrade, JM, Arraiano, CM. 2008. PNase is a key player in the regulation of small RNAs that control the expression of outer membrane proteins. RNA. 14:543-551.

3. Andrade, JM, Pobre, V, Matos, AM, Arraiano, CM. 2012. The crucial role of PNase in the degradation of small RNAs that are not associated with Hfq. RNA. 18:844-855.


64. Liou, GG, Chang, HY, Lin, CS, Lin-Chao, S. 2002. DEAD box RhlB RNA helicase physically associates with exoribonuclease PNPase to degrade double-stranded RNA
independent of the degradosome-assembling region of RNase E. J. Biol. Chem. 277:41157-41162.


120. **Tock, MR, Walsh, AP, Carroll, G, McDowall, KJ.** 2000. The CafA protein required for the 5'-maturation of 16 S rRNA is a 5'-end-dependent ribonuclease that has context-dependent broad sequence specificity. J. Biol. Chem. **275:**8726-8732.


124. **Walter, M, Kilian, J, Kudla, J.** 2002. PNPase activity determines the efficiency of mRNA 3'-end processing, the degradation of tRNA and the extent of polyadenylation in chloroplasts. EMBO J. **21:**6905-6914.


Appendices

Appendix A: Cloning, mutagenesis, and purification of mutants

A.1 Primers and cloning strategy

<table>
<thead>
<tr>
<th>Primer Sequence</th>
<th>Number of nucleotides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kate-fprimer2</td>
<td>43</td>
</tr>
<tr>
<td>Kate-fprimer1</td>
<td>28</td>
</tr>
<tr>
<td>Kate-rprimer1</td>
<td>28</td>
</tr>
<tr>
<td>Kate-rprimer2</td>
<td>39</td>
</tr>
</tbody>
</table>

Table A1: Key primers used in initial PCR reactions (and Error-prone PCR).

Primers were designed to amplify sequences within the predicted coding region of the \emph{pnp} gene (EcoGene Accession Number EG10743; \emph{E. coli} str. K-12 substr. MG1655). In this work, pGC400 (26) encoding the WT \emph{pnp} gene was used as the template for all initial PCR amplifications requiring the WT \emph{pnp} sequence. The Shine-Dalgaro ribosome binding sequence is italicized. Four changes to the WT sequence were introduced by the two sets of primers: 1. As part of Kate-fprimer2; a \emph{Bam}HI restriction enzyme recognition sequence (bolded). 2. As part of Kate-fprimer2; a change in the WT start codon sequence from TTG to ATG (underlined). 3. As a result of the introduction of a \emph{Sal}I restriction enzyme recognition site (highlighted yellow) by Katefprimer1 and Katerprimer1; a change from Gly\textsubscript{546} to Val\textsubscript{546}. 4. As part of Kate-rprimer2; an \emph{Xba}I restriction enzyme recognition site (highlighted red) immediately 3' of the TAA stop sequence.

\textit{N.B.}: A previously unidentified and unexpected single, but benign, point mutation in WT \emph{pnp} and all subsequent derivatives was discovered during this work at Thr\textsuperscript{261} $\rightarrow$ Ala\textsuperscript{261}. Site-directed mutagenesis was used to reverse this mutation and through experimentation with described assays (Materials and Methods, section 2), we confirmed that the "fixed" WT PNP behaves like the WT PNP Thr\textsuperscript{261} $\rightarrow$ Ala\textsuperscript{261} mutant (Data not shown).

**PCR mix and cycling conditions for site-directed mutagenesis.** [final] optimized PCR concentrations of components that resulted in highest transformation success of desired product were as follows: Mix totaling 50 µL combined 2.5 U of Fermentas \textit{Pfu} DNA polymerase, 0.3 µM
each of the forward and reverse set of primers listed in Table A2, 200 µM each of the dNTPs, and approx. 20 ng of plasmid p19khs1_nm or pGC400 [pAW001] as template in PCR buffer (1X Fermentas Pfu buffer with 3 mM MgSO₄). Optimized PCR cycling parameters for these reactions, programmed as "LSPNP1" in the Eppendorf Mastercycler (Personal) Thermo-cycler, were as follows: LID = 95°C; 1. = 95°C for 0:00:30  2. = 95°C for 0:00:30  3. = 55°C for 0:00:30  4. = 68°C for 0:14:00  5. = go to step 2.; Repeat 12X; Hold at 4.0°C.

Optimized PCR cycling parameters for Error-prone PCR (see section 2.2 also): "EPPCRII" programmed into the Eppendorf Mastercycler (Personal) Thermocycler: LID = 99°C; 1. = 95°C for 0:10:00  2. = 95°C for 0:00:30  3. = 52°C for 0:00:30  4. = 72°C for 0:02:15  5. = go to step 2.; Repeat 29X; 6. = 72°C for 0:10:00; Hold at 10.0°C.

Other primers used for site-directed mutagenesis:

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence</th>
<th>Number of nucleotides</th>
<th>Pnp Mutant</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS01F</td>
<td>5'-GTGGCGATATCTCTTTAGTTCGACCCCGG-3'</td>
<td>28</td>
<td>∆KH + ∆S1</td>
</tr>
<tr>
<td>LS01R</td>
<td>5'-CGCGGTCCGACTAGAGATATCGCCAC-3'</td>
<td>28</td>
<td>∆KH + ∆S1</td>
</tr>
<tr>
<td>LS02F</td>
<td>5'-GACGGCCGAGAAAGCTGAACATGCTATTCG-3'</td>
<td>29</td>
<td>∆S1</td>
</tr>
<tr>
<td>LS02R</td>
<td>5'-CGAATAGCATGTACGCTTTCTGCACCACGTC-3'</td>
<td>29</td>
<td>∆S1</td>
</tr>
<tr>
<td>AWFG22C1</td>
<td>5'-GGACAAAGATCAAAGATGTATCTTGTAAAGCGGCTTCTG-3'</td>
<td>38</td>
<td>G570C</td>
</tr>
<tr>
<td>AWRG22C1</td>
<td>5'-CAGAACCAGCCTTTACAGATAACATCTTTGATCTTGCC-3'</td>
<td>38</td>
<td>G570C</td>
</tr>
<tr>
<td>AWFK23L1</td>
<td>5'-GATCAAAAGATTTATCGGTCTACCCGTTCTGTAATCTGAC-3'</td>
<td>39</td>
<td>K571L</td>
</tr>
<tr>
<td>AWRK23L1</td>
<td>5'-CGATTACAGAAGCCCTTACAGACACAGATAACATCTTTGATC-3'</td>
<td>39</td>
<td>K571L</td>
</tr>
<tr>
<td>AWFI28A1</td>
<td>5'-GGCCGTTTCTGTAGCCCGTGCTCTGACCGC-3'</td>
<td>28</td>
<td>I576A</td>
</tr>
<tr>
<td>AWRIL28A1</td>
<td>5'-CGGTACAGACCCGCTAGTACAGAACCAGC-3'</td>
<td>28</td>
<td>I576A</td>
</tr>
<tr>
<td>AWFI28T1</td>
<td>5'-GGCCGTTTCTGTAGCCCGTGCTCTGACCGC-3'</td>
<td>28</td>
<td>I576T</td>
</tr>
<tr>
<td>AWRIL28T1</td>
<td>5'-CGGTCAGACCGCCGTACAGAACCAGC-3'</td>
<td>28</td>
<td>I576T</td>
</tr>
<tr>
<td>KMF1F</td>
<td>5'-CCCGTATCGTGGACGCTGCGCATTGTTGCACATCGGC-3'</td>
<td>38</td>
<td>F635A</td>
</tr>
<tr>
<td>KMF1R</td>
<td>5'-GCCGATGGCAACAAATGCGCAAGCTGCAACGATACGGG-3'</td>
<td>38</td>
<td>F635A</td>
</tr>
</tbody>
</table>
Table A2: Other primers for site-directed mutagenesis

A list of all primer sets used in this work including the primer names (F=forward; R=reverse), along with the corresponding sequences and resulting mutants. See Tables 1 and 2 for more information. The mutated residues are bolded and underlined. A261T is not a mutant but a "fix" (see note after Table A1); a codon in *pnp* coding for Thr at amino acid position 261 of PNPase is the correct WT sequence.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer Sequence</th>
<th>Number of nucleotides</th>
<th>Pnp Mutant</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KMF22F</strong></td>
<td>5'-GACTTTGGGCGCAGCTTTGCCATCGGCGGC-3'</td>
<td>30</td>
<td>F638A</td>
</tr>
<tr>
<td><strong>KMF22R</strong></td>
<td>5'-GCCGCCGATGGCAACAGCTTTGCGCCAAAGTC-3'</td>
<td>30</td>
<td>F638A</td>
</tr>
<tr>
<td><strong>KMH1F</strong></td>
<td>5'-GAAGGTCTGGTCAGGCTTTGCTTTAATCGCTGACAACACGC-3'</td>
<td>39</td>
<td>H650A</td>
</tr>
<tr>
<td><strong>KMH2R</strong></td>
<td>5'-GCCGCCGATGGCAACAGCTTTGCTTTAATCGCTGACACACCTC-3'</td>
<td>39</td>
<td>H650A</td>
</tr>
<tr>
<td><strong>PNPFIX3F</strong></td>
<td>5'-GCATCTCAGACAAACAGAGC-3'</td>
<td>21</td>
<td>WT; A261T</td>
</tr>
<tr>
<td><strong>PNPFIX3R</strong></td>
<td>5'-GCTCTGTTTTCGCCTGATGC-3'</td>
<td>21</td>
<td>WT; A261T</td>
</tr>
</tbody>
</table>
Figure A1: Flowchart of the mutagenic KH/S1 and pAW101 cloning processes.

Step-wise generation of the mutagenic KH/S1 library through error-prone PCR as described in Materials and Methods section 2.2. Cloning steps ultimately result in plasmid pAW101 or derivatives (Table 1, first column) expressing WT or desired pnp variant for measurement of β-galactosidase activity in reporter strain IBPC7322(λGF2) (see Materials and Methods, sections 2.1 and 2.3). Step 1. Plasmid pGC400 encoding the WT pnp gene (26) was used as a template to amplify the N-terminal PH'-PH region independent of the KH-S1 C-terminal region. Kate-fprimer2 and k-primer1 (black text and Table A1) were used to engineer fragments containing
the N-terminal PH' and PH domains framed by BamHI and SalI sites. The amplified fragments were digested with BamHI and SalI restriction enzymes and cloned into the high-copy pUC18 plasmid, creating p18phph. **Step 2. a)** Plasmid pGC400 was used as a template to amplify the C-terminal KH and S1 domains independent of the PH-PH core. Kate-fprimer1 and -rprimer2 (red text and Table A1) were used to engineer fragments containing the KH and S1 domains framed by SalI and XbaI sites that were subsequently cleaved by SalI and XbaI and cloned into pUC19, creating p19khs1_nm (nm = not-mutated). **Step 2. b)** Plasmid p19khs1_nm was cleaved with SalI and NdeI and cloned into the correspondingly cleaved p18phph to create pAW101 encoding full-length, WT pnp. **Step 3. a)** Error-prone PCR ([93]; also see Materials and Methods, section 2.2) was performed with p19khs1_nm as the template and Kate-fprimer1 and -rprimer2 (red text) as primers; creating a library of mutant KH/S1 fragments with the majority containing between 1 and 3 point mutations spanning the two RNA-binding domains. **Step 3. b)** These PCR fragments framed by engineered SalI and XbaI sites were cleaved with SalI and XbaI and subsequently cloned into pUC 19, creating p19khs1* (* denoting a library of mutants). **Step 3. c)** Plasmids encoding pnp containing interesting mutants as determined by sequencing analysis were digested with SalI and NdeI and subsequently cloned into p18phph, reconstructing the full-length pnp variant.
A.2 Purification of PNPase I576T from over-expression in strain ENS134/pAW017

Figure A2: The over-expression and purification of PNPase I576T. a. Cultures of ENS134/pAW017 (Table 2) were induced with 0.5 mM IPTG at an OD_{600} of ~0.3. Culture volumes of 0.5 mL were extracted at 0, 1, 2, and 3 hours after induction, lysed by boiling, and separated by 7.5% SDS-PAGE. Proteins were visualized by Coomassie staining. b. and c. Q-Sepharose anion exchange chromatography of untagged PNPase I576T. Analysis by 7.5% SDS-PAGE and Coomassie staining of every-other fraction eluted from the column (b), or every fraction eluted (c). d. Additional purification of PNPase I576T by hydrophobic interaction chromatography. Analysis by 7.5% SDS-PAGE and Coomassie staining of every-other fraction eluted.
A.3 Partial proteolysis of PNPase F635A/F638A/H650A and F635R/F638R/H650R

Figure A3: Trypsin sensitivity assay of FFH -> AAA and FFH -> RRR mutants.

Purified PNP mutant enzymes were treated with no trypsin (- lanes) or increasing molar dilutions of ~10 µg trypsin. Protein size markers are in the first lane. Control lanes on the left represent untreated, purified WT PNPase or PNPase with the KH and S1 domains truncated. Digestions were incubated for between 1.5 to 2 hours at room temperature in a buffer containing 25 mM Tris-HCl, 50 mM KCl, 0.1 mM EDTA, pH 7.9. Digestions were quenched with 10 µl 100 mM phenylmethylsulfonyl fluoride in ethanol. Volumes of quenched mix (20 µl) were separated by electrophoresis on 7% polyacrylamide gels which were subsequently stained with Coomassie stain.
Appendix B : SWISS-MODEL Homology modeling parameters

Workunit: P000774 Title:KH domain ecoli to nova KH3

Model Summary:
Model information:
Modelled residue range: 7 to 67
Based on template: 1ec6A (2.40 A)
Remark: No search for template was performed. Only user specified template was used for modelling.

Sequence Identity [%]: 20

Evalue: 4.8E-12

Quality information:
QMEAN Z-Score: 2.17

Quaternary structure information:
Template (1ec6): MONOMER
Model built: SINGLE CHAIN

Ligand information:
Ligands in the template: none.
Ligands in the model: none.

Global Model Quality Estimation:
QMEAN4 global scores: Local scores:

QMEANscore4:
Estimated absolute model quality:
Score components: Coloring by residue error:
Residue error plot: 1.138
Z-Score: 2.17

QMEAN4 global scores: The QMEAN4 score is a composite score consisting of a linear combination of 4 statistical potential terms (estimated model reliability between 0-1). The pseudo-energies of the contributing terms are given below together with their Z-scores with respect to scores obtained for high-resolution experimental structures of similar size solved by X-ray crystallography:

Scoring function term Raw score Z-score
C_beta interaction energy -58.93 2.70
All-atom pairwise energy -2644.90 1.77
Solvation energy -13.55 2.04
Torsion angle energy -24.02 1.22
QMEAN4 score 1.138 2.17
Local Model Quality Estimation:

Alignment:
TARGET 1 IHTIK INPDKIKDVI GKGGSVIRAL TEETGTTIEI EDDG------
1ec6A 4 mke--lveia vpenlvgail gkggktlvey qeltgariqi skkgeflpgt
TARGET sss hhhhh hhhhhhh hhhh sss s
1ec6A s sssss sss hhhhh hhhhh sss s
TARGET 40 ---TVKIAAT DGEEKAKHAIR RIEEI ---- ---------
1ec6A 52 mrrvititgO spaatqaay lisqrvtyeq gvrasnpqkv
TARGET sssss hhhhhhhhh hhhhh
1ec6A sssssssss hhhhhhhhhh hhhhhhhhhh hhh

Modeling Log:
3.70 (SP3)
Loading Template: 1ec6A.pdb
Loading Raw Sequence
Loading Alignment: ./NXXX.align.submit.fasta
Removing HET groups from template structure
Refining Raw Sequence Alignment
ProModII: doing simple assignment of backbone
ProModII: adding blocking groups
Adding Missing Sidechains
AddPolar H
BuildDeletetedLoopsModel
Trying Ligating with anchor residues ASP 37 and THR 40
Trying Ligating with anchor residues GLU 36 and THR 40
Number of Ligations found: 30
ACCEPTING loop 18: clash= 0 FF= -8.9 PP= -5.00
Building CSP loop with anchor residues ALA 45 and GLY 48
Building CSP loop with anchor residues ALA 44 and GLY 48
Number of Ligations found: 41
ACCEPTING loop 31: clash= 0 FF= 126.5 PP= -1.00
Optimizing Sidechains
Adding Hydrogens
Optimizing loops and OXT (nb = 9)
Final Total Energy: -982.155 KJ/mol
Dumping Sequence Alignment

Template Selection Log:
- Start SMR-Pipeline in automated mode on BC2-cluster at Wed Feb 1 04:42:29 2012
- User specified template structure by a PDB identifier (1ec6A), entering user template mode
- Aligning sequence of the user template structure with the target sequence using BLAST
- Alignment quality between target and specified template is too low
- Aligning sequence of the user template structure with the target sequence using HHSearch
- Send 1 target-template alignments for modeling

- building model based on 1ec6A (7-67) was successful
- Workspace Pipeline parameters
  
  Cut-off parameters to model the target based on a BLAST target-template alignment
  
  Evalue : 0.0001
  Minimum Template size (aa) for ranking : 25
  Minimum Sequence identity : 60

  Cut-off parameters to model the target based on a HHSearch target-template alignment
  
  Evalue : 0.0001
  Probability : 50
  MAC : 0.3

  Parameters for model selection
  
  Minimal number of uncovered target residues after BLAST to run HHSEARCH : 50
  Minimal number of uncovered target residues to model an additional template : 25

- Finish SMR-Pipeline in automated mode on BC2-cluster at Wed Feb 1 04:52:12 2012

**Quaternary Structure Annotation of the Template**

1ec6 is annotated as MONOMER

The oligomeric state of the structure was assigned by the authors of the corresponding PDB entry

The following biological unit was used to build the template structure: 1ec6.pdb1.gz

**Quaternary Structure Modelling of the Target Protein**

The target and template sequences are too diverse (seqid: 20.000) to infer a conservation of the oligomeric state

Please use the advanced features of the SwissModel Project Mode

The final model was calculated as single chain

**Ligand Modeling Log: Template’s ligands section**

Template without ligands.

The template contains ligands that are not yet part of the pipeline. Ligands which are currently assessed are listed in the help page.

No ligands were included in the model.