Characterization of tRNA healing enzymes from yeast and lancelet

Dissertation

submitted to the
Department of Chemistry,
Faculty of Mathematics, Informatics and Natural Sciences of the
University of Hamburg
for the award of the degree of
Doctor of Science

Gopinath Muruganandam

Hamburg

March 2014

The research work reported in this dissertation was carried out from October 2009 until November 2013 in the research laboratory of Dr. Inari Kursula at the Centre for Structural Systems Biology – Helmholtz Centre for Infection Research and University of Hamburg, Hamburg, Germany.

Reviewers of the dissertation

Asst. Prof. Dr. Inari Kursula

Prof. Dr. Andrew Torda

Date of disputation: 25 April 2014, Friday

Examiners

Asst. Prof. Dr. Inari Kursula

Prof. Dr. Wolfgang Maison

JProf. Dr. Henning Tidow

Abbreviations and symbols

Å	Ångström (10^{-10} m)
ADP	adenosine diphosphate
AI	auto-induction
AKAP18	protein kinase A anchoring protein 18
AMP	adenosine monophosphate
ANKA	Angstromquelle Karlsruhe
Appr>p	ADP-ribose 1’’-2’’cyclic phosphate
ASCC1	activating signal cointegrator 1 complex subunit 1
ATP	adenosine triphosphate
β-ME	beta-mercaptoethanol
Bf or *B. floridae*	*Branchiostoma floridae*
BLAST	Basic Local Alignment Search Tool
BSA	bovine serum albumin
C	carboxyl
c	concentration
°C	degree Celsius
CaM	calmodulin
CCD	charge-coupled device
CD	circular dichroism
CNPase	2’,3’-cyclic nucleotide 3’-phosphodiesterase
CNS	central nervous system
CPDase	cyclic phosphodiesterase
CV	column volume
ddH_2O	double-distilled water
DEPC	diethylpyrocarbonate
DESY	Deutsches Elecktronen-Synchrotron
DLS	dynamic light scattering
D_{max}	maximum dimension of a particle
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	2’-deoxynucleoside-5’-triphosphate
DTT	dithiothreitol
E. coli	*Eschericia coli*
EDTA	ethylenediaminetetraacetic acid
et al.	*et alii*
EtBr	ethidium bromide
exoG	exogenous guanosine

ExPASy	Expert Protein Analysis System
FRET	fluorescence resonance energy transfer
Fwd	forward
g	gram
g	relative centrifugal force (rcf)
GTP	guanosine triphosphate
h	hour(s)
HEPES	hydroxyethyl piperazineethanesulfonic acid
HIV	human immunodeficiency virus
Hs	*Homo sapiens*
I	intensity
IMAC	immobilized-metal affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
k	kilo- (multiplied by 10^3)
K_m	Michaelis constant
k_{cat}	turnover number (of an enzyme)
kDa	kilodalton
l	litre
LB	lysogeny broth
LigT	2'-5' RNA ligase
µ	micro- (multiplied by 10^{-6})
M	molar
m	milli- (multiplied by 10^{-3})
MALS	multi-angle static light scattering
MES	2-(N-morpholino)ethanesulfonic acid
min	minute(s)
MM	molecular mass
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MWCO	molecular weight cut-off
m/z	mass-to-charge ratio
N	amino
N	nano- (multiplied by 10^{-9})
NAD	nicotinamide adenine dinucleotide
$NADP^+$	oxidized form of nicotinamide adenine dinucleotide phosphate
NADPH	reduced form of nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NMR	nuclear magnetic resonance

NRMSD	normalized root mean square deviation
NTA	nitrilotriacteic acid
OD	optical density
p	pico- (multiplied by 10^{-12})
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
Phyre	protein homology/analogy recognition engine
PE	phosphoesterase
PEG	polyethylene glycol
PETRA	Positron-Electron Tandem Ring Accelerator
PIPES	Piperazine-N,N′-bis(2-ethanesulfonic acid)
PMSF	phenylmethylsulfonyl fluoride
PNK	polynucleotide kinase
PNS	peripheral nervous system
Poly-A	polyadenylate
Rev	reverse
R_g	radius of gyration
RICH	regeneration-induced CNPase homologue
RNA	ribonucleic acid
RNase	ribonuclease
Rnl	RNA ligase
rRNA	ribosomal ribonucleic acid
s	second(s)
SAXS	small-angle X-ray scattering
Sc or *S. cerevisiae*	*Saccharomyces cerevisiae*
SCOP	Structural Classification of Proteins
SDS	sodium dodecyl sulphate
SEC	size exclusion chromatography
SEN	splicing endonuclease
SLIC	sequence and ligation independent cloning
SMART	Simple Modular Architecture Research Tool
SRCD	synchrotron radiation CD
T	temperature
T_m	melting temperature
TAE	Tris-acetate-EDTA
TB	transformation buffer
TBE	Tris-borate-EDTA
TCEP	tris(2-carboxyethyl)phosphine
T-Coffee	Tree-based Consistency Objective Function for alignment Evaluation

Tris	tris(hydroxymethyl)aminomethane
Trl1	tRNA ligase 1
tRNA	transfer ribonucleic acid
U	unit (enzyme activity)
UV	ultraviolet
V_{max}	maximum velocity (of an enzymatic reaction)
V_P	Porod volume
v/v	volume/volume
w/v	weight/volume
2’,3’-cAMP	adenosine 2’,3’-cyclic monophosphate
2’,3’-cCMP	cytidine 2’,3’-cyclic monophosphate
2’,3’-cNADP$^+$	nicotinamide adenine dinucleotide 2′,3′-cyclic monophosphate (oxidized)

Nucleotide bases

A	adenine
C	cytosine
G	guanine
T	thymine
U	uracil

Amino acids

A	Ala	alanine	C	Cys	cysteine
D	Asp	aspartate	E	Glu	glutamate
F	Phe	phenylalanine	G	Gly	glycine
H	His	histidine	I	Ile	isoleucine
K	Lys	lysine	L	Leu	leucine
M	Met	methionine	N	Asn	asparagine
P	Pro	proline	Q	Gln	glutamine
R	Arg	arginine	S	Ser	serine
T	Thr	threonine	V	Val	valine
W	Trp	tryptophan	Y	Tyr	tyrosine

Contents

1. Introduction

1.1. Transfer RNA

Transfer RNA (tRNA) plays a crucial role as a molecular adaptor in the translation of genetic information from nucleic acid to protein. The function of tRNA is to carry amino acids to a growing polypeptide chain within the ribosome. Each tRNA molecule consists of a site for the attachment of an amino acid and a site, the anticodon, which recognizes the corresponding three-base codon on the messenger RNA (mRNA). The anticodon interacts with an mRNA codon at one end of the tRNA, and enables an attached amino acid at the other end of the tRNA to fuse with the peptide chain on a second tRNA through the action of the ribosome. This essential role of tRNA requires mature tRNA molecules to be recognized by aminoacyl tRNA synthetase for the addition of an appropriate amino acid to the 3'-end (Crick 1968, Haselkorn and Rothman-Denes 1973, Ibba and Soll 2000). The primary product of transcription of a tRNA gene is a precursor molecule. Prior to its function in protein synthesis, the pre-tRNA transcript undergoes extensive processing to generate a mature functioning tRNA [Figure 1].

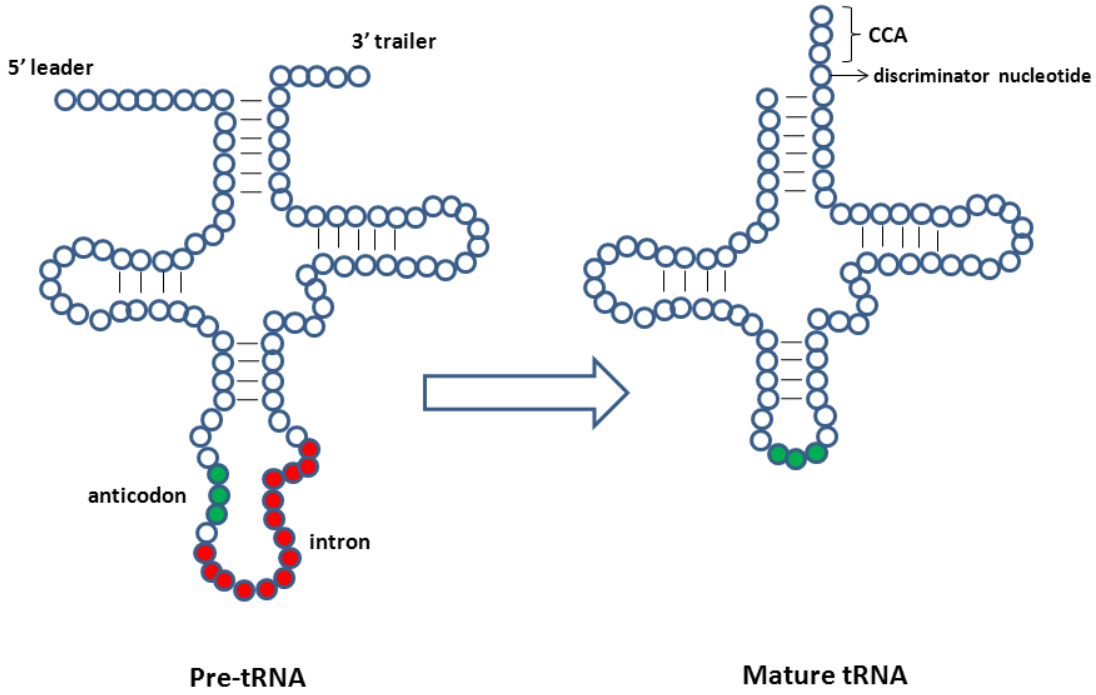


Figure 1. Processing of pre-tRNA. A schematic representation of a pre-tRNA and a mature tRNA. Each circle represents a nucleotide. Intron and anticodon are colored in red and green, respectively.

Maturation of tRNA is a collection of enzymatic reactions and involves five major steps: removal of the 5'-leader sequence by RNase P, removal of the 3'-trailer sequence by certain combinations of endonucleases and exonucleases, addition of CCA to the 3'-end of the molecule (in eukaryotes, several eubacteria and some archaea), splicing of introns and several residue-level modifications (Phizicky and Hopper 2010). Among these steps, tRNA splicing is universally conserved and the enzymes involved in this process are essential for cell growth (Phizicky, Consaul *et al.* 1992, Culver, McCraith *et al.* 1997, Trotta, Miao *et al.* 1997, Abelson, Trotta *et al.* 1998, Phizicky and Hopper 2010). The work reported in this dissertation focused specifically on the enzymes involved in tRNA splicing. The available literature on tRNA splicing is reviewed in the following sections.

1.2. Transfer RNA splicing

In many eukaryotic genes, the presence of noncoding sequences, or introns, is common and so is their removal by splicing. All three major classes of RNA - transfer RNA, ribosomal RNA and messenger RNA - contain introns. Four different splicing mechanisms have evolved: self-splicing of group I introns, self-splicing of group II introns, mRNA splicing and tRNA splicing. The autocatalytic self-splicing of group I and group II introns represents the most primitive splicing mechanism and involves two phosphotransfer reactions (Thomas 1993). The introns present in nuclear pre-mRNA are removed by an extensive splicing pathway, which requires a large and dynamic RNA-protein complex, the spliceosome (Madhani and Guthrie 1994). Pre-mRNA splicing is mechanistically related to the two self-splicing mechanisms of group I and group II introns. It differs from the group II intron splicing mechanism only in the requirement for specific proteins and co-factors and, thus, the mechanism of group II intron self-splicing and nuclear pre-mRNA splicing may have diverged from a common ancestor in the course of evolution (Valadkhan 2005). The evolutionary conservation of intron boundaries in nuclear pre-mRNA found in yeast and mammals indicates a conserved mechanism of nuclear mRNA splicing (Sharp 1981, Mount 1982). However, the intron boundaries in pre-tRNA, pre-rRNA and mitochondrial pre-mRNA have been observed to be different from those found in nuclear mRNA (Peebles, Gegenheimer *et al.* 1983). All three splicing methods discussed so far involve an RNA catalyst.

Unlike other splicing methods, the splicing mechanism of pre-tRNA requires three enzymatic activities in addition to adenosine or guanosine triphosphate (ATP/GTP), but no RNA catalysts (Culbertson and Winey 1989, McCraith and Phizicky 1991). Bacterial pre-tRNA introns undergo the most primitive, autocatalytic self-splicing, whereas archaeal and eukaryal tRNA splicing reactions require specific sets of enzymes (Abelson, Trotta *et al.* 1998). Until the identification of the evolutionary

conservation of the first enzyme involved in the tRNA splicing pathway, tRNA endonuclease, in both archaea and eukarya, the splicing pathways of the two domains had been considered to be unrelated to each other (Fabbri, Fruscoloni *et al*. 1998). Although tRNA splicing has been found to occur in all three domains of life, the bacteria, the archaea and the eukarya, the mechanism of reaction is not conserved. Since the first discovery of tRNA introns in 1977, a wealth of information has been generated by genetic, biochemical and structural investigations (Goodman, Olson et al. 1977, Valenzuela, Venegas et al. 1978).

1.2.1. Bacterial tRNA splicing

tRNA introns are less common in bacteria than in eukaryotes (Raghavan and Minnick 2009). Splicing of bacterial tRNA introns, known as group I introns, involves the primitive autocatalytic mechanism of self-splicing. Group I introns are present immediately 3' of the anticodons of $tRNA^{Leu}$ and $tRNA^{fMet}$ in various cyanobacterial species, and $tRNA^{Arg}$ and $tRNA^{Ile}$ in α- and β-proteobacteria, respectively (Kuhsel, Strickland *et al*. 1990, Xu, Kathe *et al*. 1990, Reinhold-Hurek and Shub 1992, Biniszkiewicz, Cesnaviciene *et al*. 1994, Paquin, Kathe *et al*. 1997). Group I introns are not found in protein-coding genes of bacterial chromosomes, and only interrupt tRNA genes (Paquin, Kathe *et al*. 1997, Rudi and Jakobsen 1997, Paquin, Heinfling *et al*. 1999). Out of 74 known bacterial chromosomal group I introns, 73 have been found to interrupt tRNA genes (Edgell, Belfort *et al*. 2000). The self-splicing group I introns have been detected in crude RNA extracts by end-labeling with the cofactor, [^{32}P]GTP (Garriga and Lambowitz 1984).

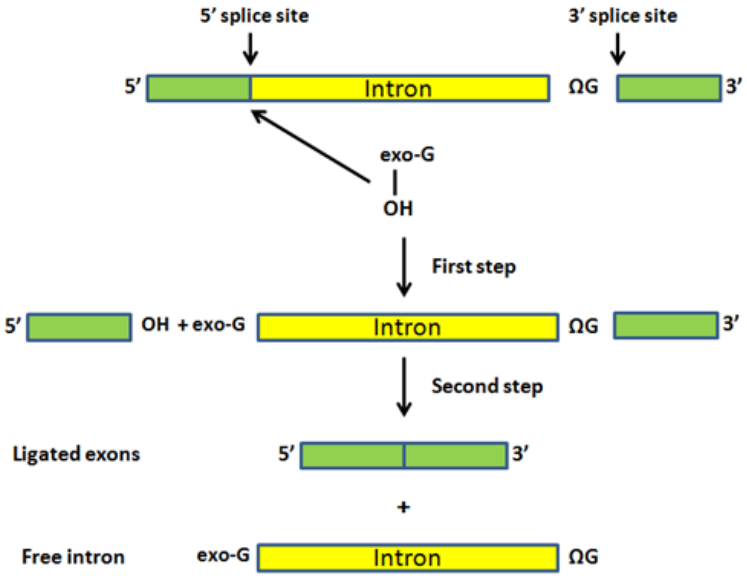


Figure 2. Bacterial tRNA splicing. A scheme of the reactions involved in self-splicing of group I introns.

The group I splicing reaction is GTP dependent and involves two steps [Figure 2]. A pre-requisite for splicing is the binding of an exogenous guanosine (exoG) to the catalytic core of the intron, called the G-binding site. In the first step, the hydroxyl group at the 3'-terminus of GTP carries out a nucleophilic attack on the 5'-splice site and gets covalently attached to the 5'-end of the intron during its excision. Thus the GTP end-labeling assay has been used to demonstrate the self-splicing of introns (Xu, Kathe *et al*. 1990, Reinhold-Hurek and Shub 1992, Biniszkiewicz, Cesnaviciene *et al*. 1994). The exoG leaves the G-binding site and is replaced by the last nucleotide of the intron (ΩG). During the second part of the reaction, the 3'-end of the released 5'-exon attacks at the 3'-splice site and leads to ligation of the exons and removal of the intron. The catalysis depends primarily on correct folding of the intron (Cech 2002, Westhof 2002, Adams, Stahley *et al*. 2004, Adams, Stahley *et al*. 2004).

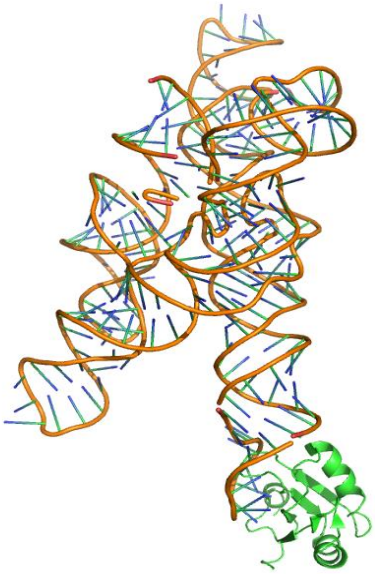


Figure 3. Structure of a bacterial tRNA intron. The crystal structure of a group I self-splicing intron from *Azoarcus sp*. BH72 [PDB ID: 1U6B] (Adams, Stahley *et al*. 2004). The U1 small nuclear ribonucleoprotein A (UIA), bound to the intron is shown as cartoon (green). The figure was generated using the molecular graphics program, PyMOL (www.pymol.org).

The crystal structure of a group I self-splicing intron [Figure 3] from the pre-tRNA^{Ile} of the purple bacterium, *Azoarcus sp*. BH72 was the first structure of the splicing complex including a complete intron, both exons and an active site with metal ions (Adams, Stahley *et al*. 2004). The structure shows that the selection of the 3'-splice site depends primarily on interactions with ΩG, whereas the selection of the 5'-splice site involves an extensive network of tertiary interactions between the 5'-exon and the intron (Adams, Stahley *et al*. 2004). The structure also reveals the coordination of three magnesium ions within the active site and provides structural evidence that the group I intron is a metalloenzyme. Based on its ability to precisely align the

substrates, select the splice sites and coordinate the metal-ions, the self-splicing group I intron has demonstrated that "not all enzymes are proteins" (Adams, Stahley *et al*. 2004). Although bacteria do not require tRNA splicing enzymes, a 2'-5' RNA ligase activity purified from extracts of *Eschericia coli* has been found to catalyze the ligation of eukaryotic tRNA halves generated by a tRNA splicing endonuclease (Greer, Javor *et al*. 1983, Arn and Abelson 1996). Most recently, a candidate approach has identified the *E. coli* RtcB protein as a 3'-5' RNA ligase (Tanaka and Shuman 2011). Surprisingly, a functional homolog of the yeast tRNA splicing enzyme, 2'-phosphotransferase, has also been identified in *E. coli*, suggesting the possibility of a unique class of RNA processing reactions in bacteria (Spinelli, Malik *et al*. 1998).

1.2.2. Archaeal tRNA splicing

The archaea contain the most intron insertions in tRNA genes. An estimated 15% of archaeal tRNA genes contain introns, and as many as 70% of tRNA genes of members of the archaeal order, thermoproteales are interrupted by introns (Marck and Grosjean 2003, Sugahara, Kikuta *et al*. 2008). The archaeal genes contain the 3'- and 5'-splice sites in two 3-nucleotide bulges separated by a 4 base-pair helix, called the bulge-helix-bulge (BHB) motif (Thompson and Daniels 1988). The BHB is a highly conserved structural motif of archaeal tRNA introns. The length of most archaeal introns is between 16 and 44 nucleotides. However, the length of the intron present in the $tRNA^{Trp}$ gene of *Haloferax volcanii* is 105 nucleotides (Daniels, Gupta *et al*. 1985). Some archaeal tRNA genes, especially those of thermoproteales contain two or three introns inserted at various loci (Sugahara, Yachie *et al*. 2007). Independent of its length, the intron base-pairs with the 5'-exon to form the BHB motif (Thompson and Daniels 1988). Despite their common location that is one base 3' to the anticodon, introns have also been reported elsewhere in the tRNA molecule (Marck and Grosjean 2003). This is an indication that the archaeal splicing machinery is independent of the pre-tRNA mature domain. Similarly to eukaryotic tRNA splicing, the archaeal pathway requires the activities of an endonuclease and a ligase.

The archaeal endonuclease recognizes substrates with the consensus BHB motif and cleaves at the 5'- and 3'-splice sites to produce 2',3'-cyclic phosphate and 5'-hydroxyl ends (Thompson and Daniels 1988, Thompson, Brandon *et al*. 1989, Lykke-Andersen, Aagaard *et al*. 1997). High-resolution crystal structures of archaeal endonucleases, from different organisms, have been determined (Li, Trotta *et al*. 1998, Li and Abelson 2000, Tocchini-Valentini, Fruscoloni *et al*. 2005, Mitchell, Xue *et al*. 2009, Yoshinari, Shiba *et al*. 2009). The structure of a homodimeric endonuclease, bound to its substrate RNA containing the BHB motif [Figure 4], reveals the catalytic mechanism of the enzyme in atomic detail (Xue, Calvin *et al*. 2006). The enzyme forms a four-subunit quaternary structure. The active site of the enzyme contains

conserved tyrosine, histidine and lysine residues, and the cleavage mechanism appears to be similar to that of RNase A-catalyzed cleavage (Raines 1998, Xue, Calvin *et al*. 2006). The tyrosine residue is involved in deprotonation of the 2'-nucleophilic oxygen, the histidine donates a proton to the 5' leaving group and the lysine stabilizes the electrostatic interactions of the transition state (Xue, Calvin *et al*. 2006).

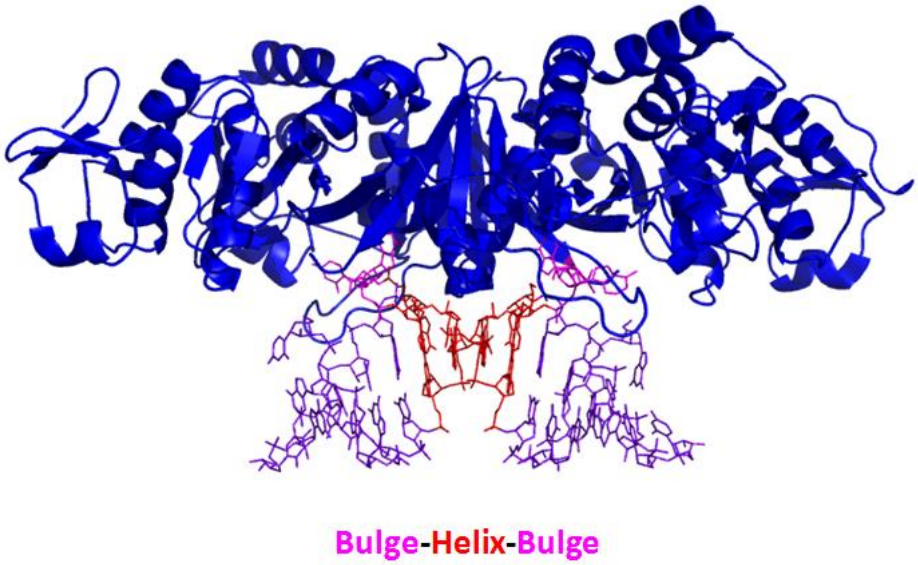


Figure 4. Substrate recognition by archaeal tRNA splicing endonuclease. The crystal structure of the homodimeric endonuclease from *Archaeoglobus fulgidus* in complex with a Bulge-Helix-Bulge RNA substrate [PDB ID: 2GJW] (Xue, Calvin *et al*. 2006). The bulges are indicated in magenta and the central helix is indicated in red. The figure was generated using PyMOL.

Upon cleavage of the intron, the tRNA halves need to be ligated. The tRNA ligase activity found in *Haloferax volcanii* involves the formation of a 3'-5'-phosphodiester bond and does not require ATP/ GTP or divalent cations (Gomes and Gupta 1997, Zofallova, Guo *et al*. 2000). The possibility of T4 Rnl2 family of ligases (part of 5'-3' RNA ligases), present in viral, bacterial and archaeal genomes, to be part of the archaeal RNA splicing has also been proposed and confirmed by the finding of a homolog of T4 Rnl2 in the thermophilic archaeon, *Methanobacterium thermoautotrophicum* (Ho and Shuman 2002, Torchia, Takagi *et al*. 2008). This enzyme joins a 3'-hydroxyl to a 5'-phosphate in an ATP-dependent manner and is a 5'-3' RNA ligase. A GTP-dependent RNA ligase (PF0027) has been identified in the hyperthermophilic archaeon, *Pyrococcus furiosus*, with ~27% sequence identity to 2'-5' RNA ligase from *E. coli* (ligT) (Kanai, Sato *et al*. 2009). This enzyme incorporates the cyclic phosphate into a 2',5'-phosphodiester bond. Recently, another 3'-5' RNA ligase activity has been purified from the extracts of *Methanopyrus kandleri* and the enzyme involved has been identified as archaeal RtcB (Englert, Sheppard *et al*. 2011).

Hence, three different RNA ligases (3'-5' Rnl, 2'-5' Rnl and 5'-3' Rnl) have been identified in archaea.

1.2.3. Eukaryotic tRNA splicing

Available knowledge on eukaryotic tRNA splicing has emerged from studies conducted in *Saccharomyces cerevisiae*, *Xenopus laevis*, *Arabidopsis thaliana*, wheat germ, and humans. Eukaryotes, similarly to archaea, contain numerous intron-containing tRNA genes. The length of eukaryal tRNA introns ranges between 12 and 104 nucleotides (Chan and Lowe 2009). Eukaryotes possess less tRNA introns than archaea. Only 6% and 20% of tRNA genes are interrupted by introns in humans and yeast, respectively (Chan and Lowe 2009). The eukaryotic tRNA introns do not contain any conserved structural motifs, such as the BHB motif of archaea. However, almost all eukaryotic tRNA introns studied so far interrupt the anticodon loop invariably one base 3' to the anticodon (Zillmann, Gorovsky *et al*. 1991, Abelson, Trotta *et al*. 1998, Phizicky and Hopper 2010). One exception to this rule is exemplified by the non-canonical introns found in the circularly permuted tRNA genes of the red alga, *Cyanidioschyzon merolae* (Soma, Onodera *et al*. 2007).

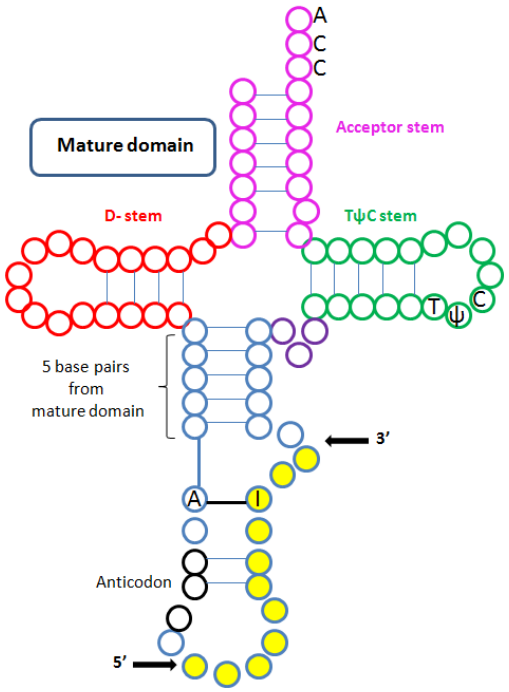


Figure 5. Endonucleolytic cleavage of eukaryotic pre-tRNA. Scheme of a eukaryotic pre-tRNA substrate with its components involved in the ruler mechanism used by the splicing endonuclease for substrate recognition and cleavage. The mature domain consists of the acceptor stem, the D-stem and the TψC stem. Each circle represents a nucleotide, and yellow circles represent the intron. The arrows labeled 5' and 3' indicate the intron boundaries. Labels A and I indicate the A-I pair.

Eukaryal tRNA splicing involves stepwise activities of three enzymes; an endonuclease, a ligase and a phosphotransferase (Abelson, Trotta *et al*. 1998, Phizicky and Hopper 2010). The eukaryal tRNA endonuclease recognizes the pre-tRNA mature domain [Figure 5] consisting of the acceptor stem, the D-stem and the TψC stem, and applies a so-called "ruler mechanism" in its native substrates to position the splice sites (Reyes and Abelson 1988). According to the ruler mechanism, the endonuclease measures five base pairs from the anchoring mature domain to locate the 5'-splice site.

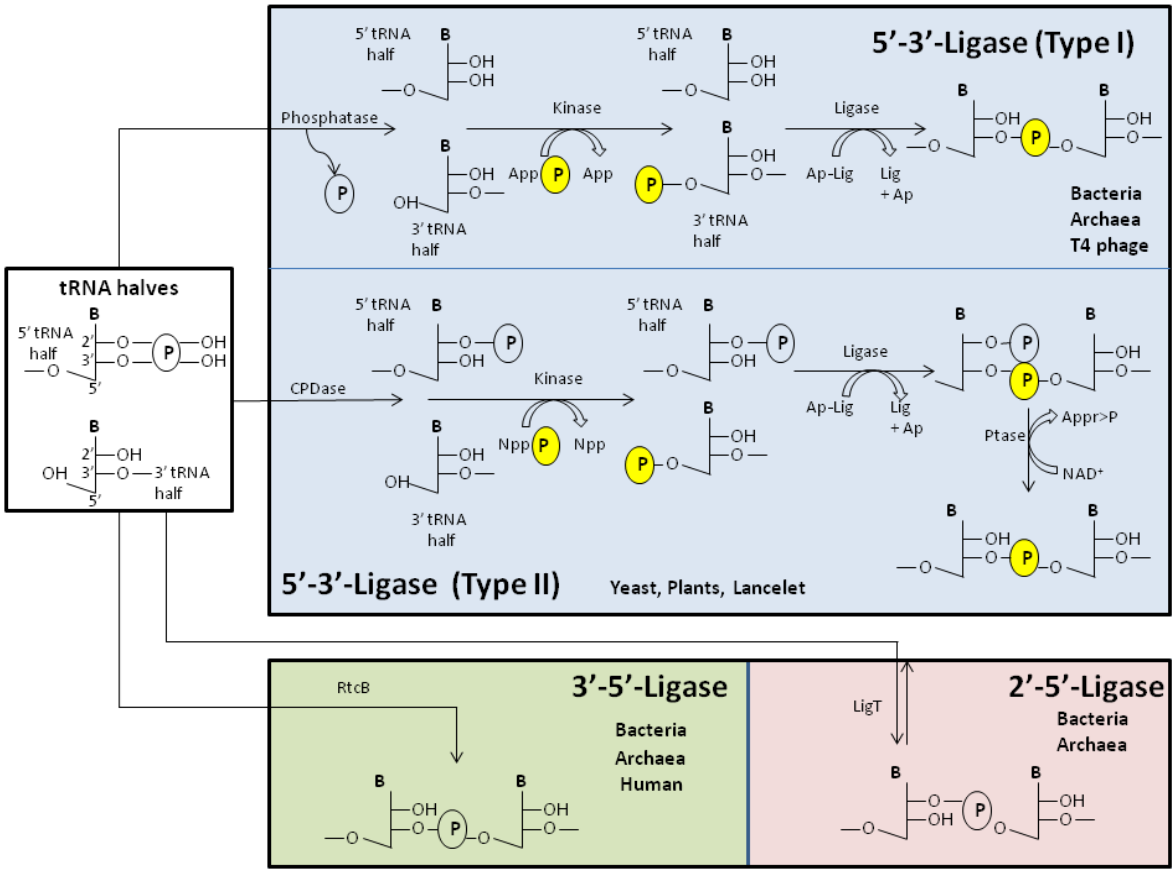


Figure 6. An overview of different RNA ligation mechanisms. The 5'-3' RNA ligases (blue background) are divided into type I and type II based on the difference in the second half of the ligation pathway. The 3'-5' RNA ligase (green background) and 2'-5' RNA ligase (pink background) catalyze ligation in a single step. B - base; Appp - adenosine 5'-triphosphate; App - adenosine 5'-diphosphate; Ap - adenosine 5'-monophosphate; Nppp - unspecified nucleoside 5'-triphosphate; Np - unspecified nucleoside 5'-monophosphate; Lig+Ap - adenylated ligase protein; NAD^+ - nicotinamide adenine dinucleotide; Ptase - 2'-phosphotransferase; Appr>p - ADP-ribose-1'',2''-cyclic phosphate. [Figure modified from (Popow, Schleiffer *et al*. 2012)].

To locate the 3'-splice site, the enzyme depends upon a base pair called the A-I pair or anticodon-intron pair that is formed between a nucleotide two bases upstream of the anticodon and a nucleotide in the intron, three bases upstream of the 3'-splice site

(Mattoccia, Baldi *et al*. 1988, Reyes and Abelson 1988, Baldi, Mattoccia *et al*. 1992, Trotta, Miao *et al*. 1997, Fabbri, Fruscoloni *et al*. 1998). Except for the strictly conserved A:I base pair, mutations in the intron do not seem to affect its recognition by eukaryal endonuclease (Johnson, Ogden *et al*. 1980, Baldi, Mattoccia *et al*. 1992).

The eukaryal and archaeal tRNA endonucleases are phylogenetically related (Lykke-Andersen and Garrett 1997). Although the modes of splice site recognition by eukaryal and archaeal endonucleases appear to be different, at least one eukaryal enzyme has been found to retain the ability to recognize archeal pre-tRNA substrates (Fabbri, Fruscoloni *et al*. 1998). The eukaryal endonuclease is a tetrameric enzyme (αβγδ) with two structural subunits and two catalytic subunits, whereas the archaeal endonuclease is composed of fewer subunits, with α2, α2β2, or α4 configurations (Trotta, Miao *et al*. 1997, Trotta, Paushkin *et al*. 2006, Xue, Calvin *et al*. 2006). Both eukaryal and archeal endonucleases, through catalysis of pre-tRNA cleavage, generate a 2',3'-cyclic phosphate at the 3'-end of the 5'-exon and a 5'-hydroxyl at the 5'-end of the 3'-exon in addition to a linear intron with 2',3'-cyclic phosphate and 5'-hydroxyl termini (Abelson, Trotta *et al*. 1998). Thus, the mechanism of endonucleolytic cleavage is conserved among archaea, and lower and higher eukaryotes (Filipowicz and Shatkin 1983, Peebles, Gegenheimer *et al*. 1983, Stange, Gross *et al*. 1988, Thompson, Brandon *et al*. 1989, Baldi, Mattoccia *et al*. 1992, Abelson, Trotta *et al*. 1998).

The mechanism of ligation of tRNA halves [Figure 6] in archaea and eukarya is not as conserved as the endonucleolytic cleavage (Konarska, Filipowicz *et al*. 1981, Filipowicz and Shatkin 1983, Englert, Sheppard *et al*. 2011). Yeast and plants use a 5'-phosphate ligation pathway, in which the 5'-phosphate of the 3'-tRNA half functions as the junction phosphate of the new phosphodiester linkage (Greer, Peebles *et al*. 1983, Englert and Beier 2005). The 5'-phosphate ligation requires three enzymatic activities, a cyclic phosphodiesterase (CPDase), a polynucleotide kinase (PNK) and a ligase (Abelson, Trotta *et al*. 1998, Englert and Beier 2005, Wang and Shuman 2005, Wang, Schwer *et al*. 2006). In contrast to yeast and plants, animal cells employ two different ligation pathways. The 3'-phosphate ligation pathway of animal cells utilizes the 3'-phosphate of the 5'-tRNA half as the junction phosphate and has been detected for the first time in HeLa cell extracts (Filipowicz and Shatkin 1983). The second pathway used by animal cells is the yeast-type 5'-phosphate ligation pathway, which has been detected in HeLa cell extracts and in the lancelet, *Branchiostoma floridae* (Zillmann, Gorovsky *et al*. 1991, Englert, Sheppard *et al*. 2010).

1.2.3.1. Yeast tRNA splicing

Since intron-containing tRNA genes were first discovered in the yeast *Saccharomyces cerevisiae* in 1977, the tRNA splicing mechanism of yeast has become the earliest and the most extensively studied (Goodman, Olson *et al*. 1977, Valenzuela, Venegas *et al*. 1978). By 1997, with the availability of the sequenced *S. cerevisiae* genome, 274 tRNA genes had been identified in yeast, and 61 of them (~20%) that encode ten different tRNAs contain introns (Trotta, Miao *et al*. 1997). The introns were found to be 14-60 nucleotides in length and to interrupt the anticodon loop one base to the 3' side of the anticodon (Ogden, Lee *et al*. 1984). Despite the invariable location of the 3'-splice site in a bulged loop, no sequence conservation at the splice sites has been found among the ten different yeast pre-tRNAs (Baldi, Mattoccia *et al*. 1992). The introns in the same family of tRNA genes (such as $tRNA^{Tyr}$ or $tRNA^{Phe}$) are completely or nearly identical (Peebles, Gegenheimer *et al*. 1983). Limited nucleolysis-guided structure probing of yeast tRNA precursors has revealed the presence of the conventional "cloverleaf" tertiary structure and that the intron with the splice sites is exposed the most, as evidenced by its sensitivity to nucleolytic cleavage (Swerdlow and Guthrie 1984, Lee and Knapp 1985).

Interestingly, all the enzymes involved in yeast tRNA splicing have been found to be essential for the growth of yeast cells, confirming that tRNA splicing is a mandatory process *en route* protein synthesis (Phizicky, Consaul *et al*. 1992, Culver, McCraith *et al*. 1997, Trotta, Miao *et al*. 1997). Although the splicing of nuclear tRNA genes has long been considered to be an event confined to the nucleus, one of the tRNA splicing enzymes, the endonuclease, localizes to the outer surface of mitochondria in yeast, suggesting the possibility of cytoplasmic tRNA splicing (Yoshihisa, Yunoki-Esaki *et al*. 2003, Yoshihisa, Ohshima *et al*. 2007). A chimeric pre-tRNA (pre-$tRNA^{ArchEuka}$), consisting of an intron of the archaeal type (with a BHB motif) and the mature domain of *S. cerevisiae* suppressor SUP4 $tRNA^{Tyr}$, has been found to be correctly processed, cleaved and ligated in *S. cerevisiae* cells (Di Segni, Borghese *et al*. 2005). Despite the extensive characterization of tRNA splicing in yeast, the essentiality of tRNA introns is not yet fully understood. Although the introns of some tRNAs have been found to be crucial for their proper nucleotide modifications, such as pseudouridylation, a recent study has indicated that deletion of all introns of a particular tRNA isodecoder family does not affect the growth or translation of the mutants (Szweykowska-Kulinska, Senger *et al*. 1994, Mori, Kajita *et al*. 2011).

1.2.3.2. Lancelet tRNA splicing

The tRNA splicing mechanism of the lancelet *Branchiostoma floridae* has been elucidated recently and found to involve the yeast-type 5'-P RNA ligation pathway

(Englert, Sheppard *et al*. 2010). The enzymes involved in lancelet tRNA splicing have been identified as (i) an RNA ligase (*Bf* Rnl) that contains the same substrate specificity as yeast and plant tRNA ligases, (ii) a cyclic phosphodiesterase protein with CPDase and PNK activity (*Bf* PNK/CPDase), and (iii) another enzyme with exclusive PNK activity (*Bf* Clp1) (Englert, Sheppard *et al*. 2010). The combined activities of the bi-functional *Bf* PNK/CPDase and *Bf* Rnl have been shown to be sufficient for healing and sealing of the tRNA halves *in vitro* and *in vivo* (Englert, Sheppard *et al*. 2010). The yeast (Class I) and plant (Class II) 5'-P RNA ligases contain PNK, CPDase and ligase domains in a single polypeptide, whereas the newly identified *Bf* Rnl does not encode PNK and CPDase domains, forming another class of 5'-P RNA ligases (Class III). The Class III 5'-P RNA ligases are present in the arthropod *Homarus americanus*, the mollusk *Aplysia californica*, the rotifer *Brachionus plicatilis*, in addition to the cephalochordate *Branchiostoma floridae* and absent in vertebrates, insects, plants and fungi (Englert, Sheppard *et al*. 2010). In Class IV 5'-P RNA ligation pathway of vertebrates, the enzymes kinase (Clp1), cyclic nucleotide phosphodiesterase (CNPase) and phosphotransferase (TRPT1) have been identified, whereas the enzyme involved in 5'-3' RNA ligation is yet unidentified (Hu, Lu *et al*. 2003, Ramirez, Shuman *et al*. 2008, Schwer, Aronova *et al*. 2008).

Owing to its homology to vertebrate CNPase, *Bf* CNP was the name assigned to the bi-functional (PNK/CPDase) polypeptide of *B. floridae*; however, following the verification of its PNK activity, it has been renamed as *Bf* PNK/CPDase (Englert, Sheppard *et al*. 2010). The ligation of tRNA halves that were prepared by the activity of splicing endonuclease and *Sc* PNK/CPDase or T4 PNK has revealed a strict requirement of a 2'-phosphate at the end of the 5'-tRNA half for subsequent ligation by *Bf* Rnl (Englert, Sheppard *et al*. 2010). Upon addition of excessive calf intestinal phosphatase to the spliced tRNA, the 2'-phosphate at the splice junction is removed, and the adjacent phosphodiester junction becomes cleavable by RNase T2 (Englert, Sheppard *et al*. 2010). This finding has revealed that *Bf* Rnl forms a 2'-phosphomonoester-3',5'-phosphodiester linkage at the end of ligation, and not a 2'-5' phosphodiester linkage, as this linkage is resistant to cleavage by RNase T1 and T2 (Englert, Sheppard *et al*. 2010). *Bf* Rnl and *Sc* PNK/CPDase have been found to complement the *Sc* Trl1-deficient strain (Trl1Δ) of the yeast *S. cerevisiae in vivo* (Englert, Sheppard *et al*. 2010). However, the individual enzymes *Bf* Rnl or *Sc* PNK/CPDase could not complement the Trl1Δ strain, confirming that both the healing and sealing enzymes are essential for tRNA splicing (Englert, Sheppard *et al*. 2010).

ATP-dependent RNA ligases react with ATP and form a ligase-adenylyl intermediate, in which adenosine monophosphate (AMP) is joined to the ligase through a phosphoamide linkage to an N-terminal lysine residue (Thøgersen, Morris *et al*. 1985, Xu, Teplow *et al*. 1990, Odell, Sriskanda *et al*. 2000). The lysine residue is present as

part of the conserved K-x-(D/N)-G motif that is found in mRNA capping enzymes and DNA/RNA ligases (Sawaya, Schwer *et al*. 2003). The adenylyl transferase activity of *Bf* Rnl has been assayed by using [α^{32}P] ATP. The wild-type *Bf* Rnl reacts with ATP, whereas the lysine mutant (K184A) does not, confirming that Lys184, located within the KANG motif, is the site of adenylylation (Englert, Sheppard *et al*. 2010).

Following the verification of the PNK and CPDase activities of *Bf* PNK/CPDase, co-expression of *Bf* Rnl and *Bf* PNK/CPDase has been found to complement a tRNA ligase-deficient strain of *S. cerevisiae* (Englert, Sheppard *et al*. 2010). This result has indicated that *Bf* Rnl and *Bf* PNK/CPDase together could carry out healing and sealing activities of tRNA ligation. The second PNK of *B. floridae*, *Bf* Clp1, has been found to be more efficient than *Bf* PNK/CPDase in RNA-kinase reactions. However, adding both PNKs in the reaction showed no difference in the rate of overall ligation (Englert, Sheppard *et al*. 2010). Although Clp1 seems to have replaced PNK in vertebrates, *B. floridae* tRNA maturation does not require *Bf* Clp1 and can be performed by *Bf* Rnl and *Bf* PNK/CPDase (Englert, Sheppard *et al*. 2010).

Since vertebrates additionally employ 3'-P RNA ligation of tRNA halves, it has been suggested that they might use the 5'-P Rnl for an alternative RNA ligation pathway that requires only Clp1 and not PNK; as a result, PNK might have become dispensable (Filipowicz and Shatkin 1983, Englert, Sheppard *et al*. 2010). However, in *B. floridae*, both tRNA maturation and this putative alternative function might use the 5'-P RNA ligation pathway and that may necessitate the requirement of both Clp1 and PNK activities (Englert, Sheppard *et al*. 2010). The alternative functions of 5'-P RNA ligase have also been cited as the possible explanation for the sequence divergence among the three known classes of 5'-P RNA ligases (Englert, Sheppard *et al*. 2010). The recent finding that human Clp1 and mammalian CNPase could complement the corresponding yeast Trl1 activities *in vivo* has suggested that the healing and sealing activities, in mammalian cells, might be encoded in separate polypeptides as in *B. floridae* (Ramirez, Shuman *et al*. 2008, Schwer, Aronova *et al*. 2008, Englert, Sheppard *et al*. 2010).

1.2.4. Mechanism of yeast tRNA splicing

The accumulation of pre-tRNAs in a yeast temperature-sensitive mutant (at the *rna1* locus*)* was the essential clue for the source of pre-tRNA substrates and paved the way for the development of the first tRNA splicing system *in vitro* (Hopper and Banks 1978, Knapp, Beckmann *et al*. 1978, O'Farrell, Cordell *et al*. 1978). The *in vitro* system has been efficiently applied to deduce the pathway of tRNA splicing (Knapp, Ogden *et al*. 1979, Peebles, Ogden *et al*. 1979, Greer, Peebles *et al*. 1983, Peebles, Gegenheimer *et al*. 1983). Pre-tRNA^{Tyr} and pre-tRNA^{Phe} have been isolated from the

yeast mutant and used as substrates in a wild-type strain to detect an enzymatic activity that specifically excises the introns and splices the ends to generate mature-sized tRNAs (Knapp, Beckmann *et al*. 1978).

Subsequent studies have proposed that yeast tRNA splicing could be a two-step process that involves ATP-independent endonucleolytic scission of the introns from pre-tRNA and ATP-dependent ligation of the broken tRNA halves (Peebles, Ogden *et al*. 1979). The absence of ATP in the splicing reaction leads to accumulation of tRNA half molecules. Similar tRNA halves have appeared as reaction intermediates, as the ligation reaction was inhibited with mature $tRNA^{Phe}$ (Peebles, Ogden *et al*. 1979). Both the endonucleolytic cleavage and the ligation reactions are inhibited by the presence of mature tRNA, but differently. Thus, the reactions are independent of each other *in vitro* (Peebles, Ogden *et al*. 1979).

The intermediates of the yeast tRNA splicing reaction have been identified (Knapp, Ogden *et al*. 1979). The cleaved intron exists as a linear molecule containing a 5'-hydroxyl terminus and a 3'-phosphorylated terminus. The site-specific cleavage of the intron also leaves two broken tRNA halves: a 5'-half tRNA molecule with a phosphorylated 3'-terminus and a 3'-tRNA half molecule with a 5'-hydroxyl terminus. (Knapp, Ogden *et al*. 1979). Furthermore, the two steps of the yeast tRNA splicing reaction are catalyzed by two different enzymes that act independently (Greer, Peebles *et al*. 1983, Peebles, Gegenheimer *et al*. 1983). In the first step, a site-specific endonuclease cleaves the intron off the 5'- and 3'-splice sites in the pre-tRNA, in the absence of ATP. In the second stage of splicing, the tRNA halves are joined by an ATP-dependent RNA ligase. The endonuclease, purified from membrane fractions by solubilization with a non-ionic detergent, has been regarded as an integral membrane protein (Peebles, Gegenheimer *et al*. 1983). The yeast tRNA ligase has been purified from soluble fractions of yeast extract prepared in high-salt concentrations. The endonuclease and ligase activities have been shown to be physically separated (Greer, Peebles *et al*. 1983, Peebles, Gegenheimer *et al*. 1983).

The structures of the substrates and products of the yeast tRNA ligase reaction are identical to those reported for wheat germ ligase (Konarska, Filipowicz *et al*. 1982, Greer, Peebles *et al*. 1983). Similarly to T4 RNA ligase and wheat germ RNA ligase, the yeast ligase reaction has been predicted to occur through an adenylated RNA intermediate, formed by the transfer of the enzyme-bound AMP moiety to the 5'-phosphate of the 3'-tRNA half molecule in a 5'-5' phosphoanhydride linkage. AMP is released upon ligation of the two tRNA halves (Cranston, Silber *et al*. 1974, Greer, Peebles *et al*. 1983, Schwartz, Greer *et al*. 1983). The T4 RNA ligase together with T4 polynucleotide kinase has been shown to replace the yeast tRNA ligase and complete the process of joining the tRNA halves generated by the yeast endonuclease (Greer,

Peebles *et al*. 1983). In the first step, the initial products of yeast endonucleolytic cleavage are converted to tRNA halves with 3'-hydroxyl and 5'-phosphate termini by the action of T4 PNK that contains cyclic phosphodiesterase, 3'-phosphatase and kinase activities (Becker and Hurwitz 1967, Cameron and Uhlenbeck 1977). In the second step, those tRNA halves are joined by the T4 ligase (Greer, Peebles *et al*. 1983). Hence, it has been suggested that the mechanism of ligation of the yeast RNA ligase could be similar to that of T4 RNA ligase and both mechanisms could have diverged from a common ancestor (Cranston, Silber *et al*. 1974, Greer, Peebles *et al*. 1983). However, the two enzymes show differences in substrate specificity; while the T4 ligase ligates a broad range of tRNA substrates and synthetic oligonucleotides, the yeast ligase specifically acts on cognate tRNA halves, which are products of the yeast endonucleolytic cleavage with hydrogen bonds (Greer, Peebles *et al*. 1983). The exclusive substrate specificity of the yeast enzyme could be linked to its possible role in tRNA splicing *in vivo.*

The distinct physical properties of the yeast endonuclease and the ligase have raised the possibility that the excision and the ligation steps could be independent rather than concerted *in vivo*. In that case, the two enzymatic activities could even be present in two different subcellular compartments. An *in vitro* tRNA splicing assay, constituted by coupling the excision and ligation steps of the splicing reaction, has been used to examine whether the excision and ligation reactions are concerted or independent (Greer 1986). It was basically a competition assay, using yeast tRNA ligase and T4 PNK+ligase. In the T4 splicing pathway, the phosphomonoester that is formed as a result of the cyclic phosphodiesterase activity is cleared by the phosphatase activity of the associated T4 PNK (Cameron and Uhlenbeck 1977). This activity is absent in the yeast RNA ligase and so the yeast ligation product contains a 2'-phosphate. This differential property of the ligation products has been used in this competition assay to distinguish between the yeast tRNA ligase and T4 ligase products. The results indicated that the yeast tRNA ligase has preferential access to the products of the endonuclease and suggested that a multienzyme complex, possibly associated with the nuclear envelope, might carry out the tRNA splicing *in vivo* (Greer 1986). This study was the first evidence for concerted reactions of tRNA splicing through the assembly of a tRNA splicing complex.

Initial studies to examine the substrate requirements for accurate splicing of pre-tRNA have been carried out by introducing mutations at different segments of the pre-tRNA including the mature domain (tRNA-like) and the intron. Several of the mutations in the mature domain have been found to affect the rate of splicing (Kurjan, Hall *et al*. 1980, Willis, Hottinger *et al*. 1984, Pearson, Willis *et al*. 1985). Single-base mutations in the D-stem, the intron and the extra stem affect the efficiency of splicing, whereas mutations in the anticodon stem alter the accuracy of the process (Willis, Hottinger *et*

al. 1984, Mathison, Winey *et al*. 1989). The primary recognition sites for both the endonuclease and the ligase are present within the mature domain of the pre-tRNA, and the splice sites might be determined by their location relative to the primary recognition sites (Greer, Söll *et al*. 1987). Experiments with pre-tRNA variants have led to several conclusions: i) the sequences in the mature domain of pre-tRNA affect substrate recognition by endonuclease, ii) the specificity of cleavage at both 5' and 3'-splice sites is determined by the length of the anticodon stem [Figure 7A], and iii) the recognition of the 5'-splice site depends on the length of the anticodon loop [Figure 7A]. These conclusions are collectively referred to as "the ruler mechanism", the widely-accepted model of substrate recognition and cleavage by the yeast tRNA endonuclease (Reyes and Abelson 1988, Mathison, Winey *et al*. 1989, Trotta, Miao *et al*. 1997, Calvin and Li 2008). The recognition of the 3'-splice site relies upon the base pair (A-I pair) formed between a conserved purine base three nucleotides upstream of the 3'-splice site and a pyrimidine at position 32 in the anticodon loop (Baldi, Mattoccia *et al*. 1992, Di Nicola Negri, Fabbri *et al*. 1997).

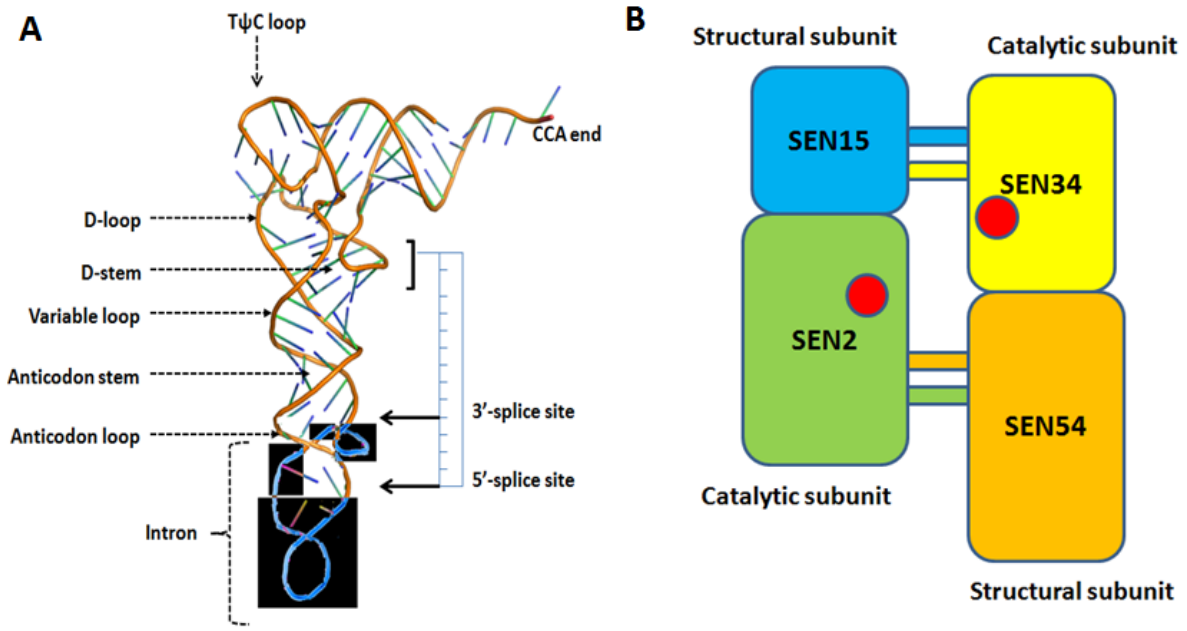


Figure 7. Structural models of yeast pre-tRNA and tRNA splicing endonuclease. A) Model of the tertiary structure of pre-tRNA built on the crystal structure of yeast $tRNA^{Phe}$. The intron is indicated in blue on a black background. The figure shows the ruler applied in splice site recognition, and the respective locations of 5' and 3'-splice sites in pre-tRNA. The figure was generated using PyMOL. **B)** Structural model of a eukaryal αβγδ tRNA-splicing endonuclease. Splicing endonuclease (SEN) subunits are named after their molecular weight. Differently colored boxes indicate distinct and separate subunits. Red circles indicate the active sites.

Successful cloning and purification of the yeast tRNA splicing endonuclease has revealed that the enzyme is an αβγδ heterotetramer [Figure 7B] with two catalytic subunits (*Sc* SEN2 and *Sc* SEN34) and two structural subunits (*Sc* SEN15 and *Sc* SEN54) (Rauhut, Green *et al*. 1990, Trotta, Miao *et al*. 1997). The yeast subunits are homologous to the human proteins, *Hs* TSEN2, *Hs* TSEN34, *Hs* TSEN15 and *Hs* TSEN54 (Paushkin, Patel *et al*. 2004). The catalytic subunits SEN2 and SEN34 are homologous to the archaeal α2 or α4 subunits (Kleman-Leyer, Armbruster *et al*. 1997). All four subunits are essential for cell viability in yeast. Interactions between SEN2-SEN54 and SEN15-SEN34 have been detected in two-hybrid analysis (Trotta, Miao *et al*. 1997).

The yeast tRNA ligase has been purified to near homogeneity and shown to ligate an artificial substrate, [oligo (A_{16})], in addition to yeast tRNA halves (Phizicky, Schwartz *et al*. 1986). This finding has demonstrated the possible existence of an activated adenylyl-RNA intermediate [A(5')pp(5')A_{16}], as suggested by Greer *et al*. (Greer, Peebles *et al*. 1983, Phizicky, Schwartz *et al*. 1986). The DNA sequence of the entire coding region of the *S. cerevisiae* tRNA ligase gene has revealed that the enzyme is a basic protein of 827 amino acids with a molecular weight of 95.4 kDa (Westaway, Phizicky *et al*. 1988). Partial proteolysis of the enzyme has identified three fragments containing three independent and non-overlapping enzymatic activities (Xu, Teplow *et al*. 1990). The fragment responsible for cyclic phosphodiesterase activity resides in the carboxyl terminus of the enzyme, and the site of adenylylation lies in the amino terminus, specifically, at Lys114. Sequence comparisons have shown that Lys114 is an equivalent of the active lysine of T4 RNA ligase (Xu, Teplow *et al*. 1990, Apostol, Westaway *et al*. 1991). Each of the three fragments is separated from the others by protease-sensitive regions (Xu, Teplow *et al*. 1990).

The yeast tRNA ligase has been found to contain binding sites for both ATP and GTP (Belford, Westaway *et al*. 1993). GTP is associated with the polynucleotide kinase activity, while ATP is involved in the formation of the activated adenylyl-RNA intermediate. It has been suggested that such a complex NTP requirement could couple the splicing reaction to transcription and/or translation (Belford, Westaway *et al*. 1993). The yeast tRNA ligase is structurally and mechanistically related to the RNA ligase and polynucleotide kinase of T4 bacteriophage (Apostol, Westaway *et al*. 1991, Sawaya, Schwer *et al*. 2003, Wang, Ho *et al*. 2003). The two T4 enzymes together possess the three activities encoded in the yeast tRNA ligase, and the T4 polynucleotide kinase additionally contains a 3'-phosphatase activity that removes the phosphate at the splice junction (Cameron and Uhlenbeck 1977).

Investigations into the activity responsible for the removal of the 2'-phosphate at the splice junction have identified two separate components in the yeast extracts. The first

component is the cellular cofactor, nicotinamide adenine dinucleotide (NAD^+) (McCraith and Phizicky 1991). NAD^+ functions as the receptor in a reaction that involves the transfer of the 2'-phosphate to the 2'' position of the ribose of NAD^+ to produce ADP-ribose 1'',2''cyclic phosphate [Appr>p] (Culver, McCraith *et al*. 1993). The second component is the 2'-phosphotransferase enzyme encoded by the TPT1 gene (Culver, McCraith *et al*. 1997). The enzyme catalyzes the transfer of a 2'-phosphate to NAD^+. TPT1 has been suggested to play an essential role either in the removal of 2'-phosphate from all intron-containing tRNA molecules or in the generation of the novel molecule, Appr>p (Culver, McCraith *et al*. 1997). Further experiments with a phosphotransferase mutant have revealed that the removal of 2'-phosphate is essential for the correct modification of residues near the splice junction, as a tRNA containing the 2'-phosphate is not recognized as a substrate by the modification enzyme and is functionally inactive in translation (Spinelli, Consaul *et al*. 1997). In the temporal order of the tRNA maturation process, splicing and removal of the 2'-phosphate occur just prior to the modification of certain positions in the tRNA molecule.

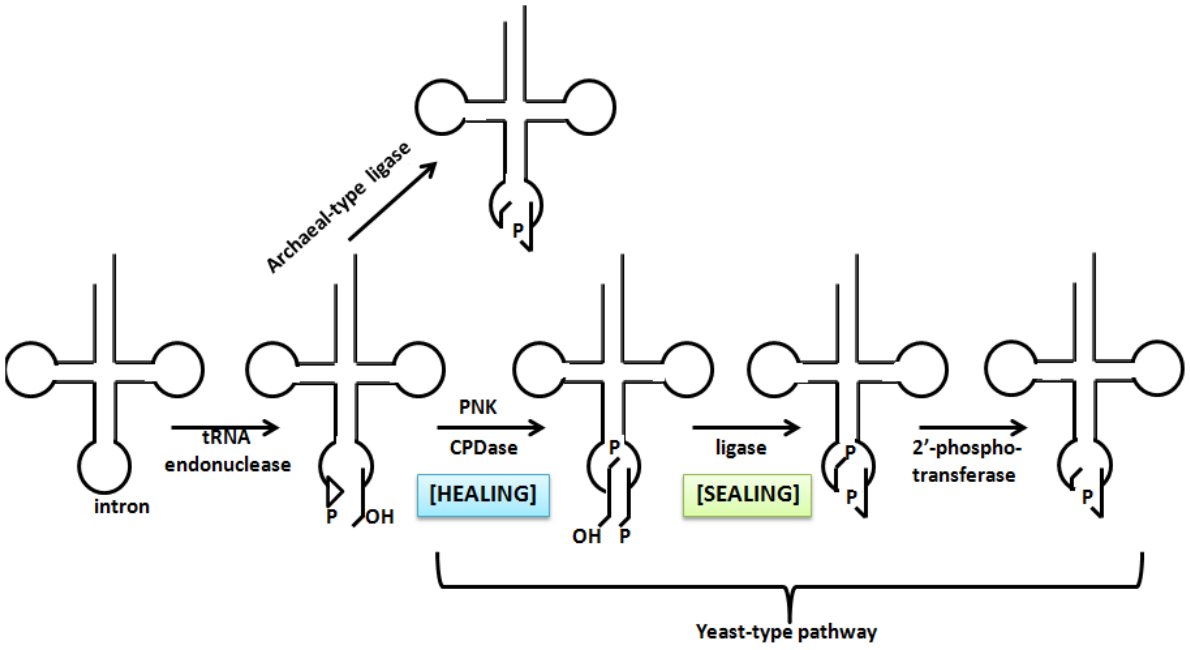


Figure 8. Overview of the archaeal-type and the yeast-type tRNA splicing pathways. In the yeast-type pathway, tRNA splicing occurs through a healing step, where the tRNA halves generated by the splicing endonuclease are modified by the PNK and CPDase activities prior to ligation. The archaeal splicing pathway does not require the healing phase, and the tRNA halves generated by the endonuclease are directly ligated.

As a result of all the above-mentioned crucial findings, the pathway of yeast tRNA splicing has been deduced [Figure 8]. To summarize the findings, yeast tRNA splicing occurs in three steps. In the first step, the intron-containing pre-tRNA is cleaved at the

5’ and 3’ splice sites by a site-specific tRNA splicing endonuclease. The products of the endonucleolytic cleavage are two tRNA half molecules and a linear intron with 5’-OH and 3’-cyclic PO_4 termini. In the second step, the two tRNA halves become substrates for tRNA ligase that catalyzes three reactions: i) hydrolysis of 2’,3’-cyclic phosphate by the CPDase domain of yeast tRNA ligase yields a 2’-PO_4 and a 3’-OH group [Figure 8]; ii) phosphorylation of the 5’-OH group by the GTP-dependent activity of the PNK domain of yeast tRNA ligase [Figure 8]. Reactions i and ii are collectively called “the healing step” of tRNA splicing [Figure 8]. iii) The ligase reacts with ATP to form a covalent ligase-(lysyl-N)-AMP intermediate and pyrophosphate. The AMP moiety is transferred from ligase-adenylate to the 5’-phosphate of the 3’-exon forming an activated RNA-adenylate intermediate (AppRNA) *via* a 5’-5’ phosphoanhydride linkage. The tRNA exons are joined by the ATP-dependent activity of the ligase that catalyses an attack of the 3’-OH on the activated donor phosphoanhydride to form a 2’-phosphomonoester-3’,5’-phosphodiester bond with the release of AMP.

In the third step of tRNA splicing, the 2’-phosphate, that originally comes from the γ-phosphate of GTP and remains at the splice junction, is removed from the spliced tRNA molecules by the activity of a NAD^+-dependent 2’-phosphotransferase that catalyses the transfer of the 2’-phosphate to NAD^+ to form Appr>p and mature tRNA with the release of nicotinamide [Figure 8]. Thus, the tRNA splicing reaction involves three nucleotidyl transfer steps.

1.3. The tRNA healing enzymes, PNK and CPDase

The healing of tRNA halves is catalyzed by the PNK and CPDase domains of Trl1 in the yeast, *Saccharomyces cerevisiae,* and by the bi-functional polypeptide PNK/CPDase in the lancelet, *Branchiostoma floridae*. Cloning and sequence analysis of the yeast Trl1 have identified the segment constituted by amino acid residues 389 - 827 as the end-healing domain with distinct PNK and CPDase activities (Apostol, Westaway *et al*. 1991, Sawaya, Schwer *et al*. 2003).

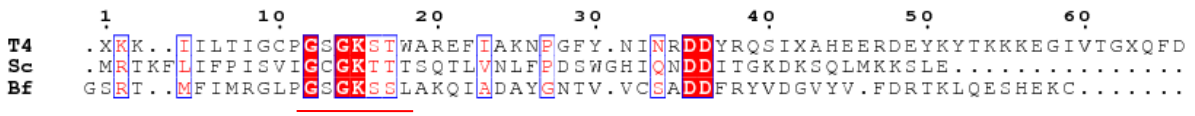


Figure 9. NTP-binding motif of the PNK domain. Conservation of the P-loop motif (underlined in red) among T4 PNK (T4), *Sc* Trl1 (Sc) and *Bf* PNK/CPDase (Bf).

The central kinase module of yeast Trl1 and the N-terminal PNK domain of lancelet PNK/CPDase show resemblance to the bacteriophage T4 PNK and belong to the P-loop phosphotransferase superfamily. They contain the signature Walker A motif “G-

x-GK(T/S)" (P-loop) which is an NTP-binding site in most of the NTP-dependent phosphotransferases [Figure 9] (Walker, Saraste *et al*. 1982, Wang and Shuman 2001, Galburt, Pelletier *et al*. 2002, Wang, Lima *et al*. 2002, Wang and Shuman 2002, Englert, Sheppard *et al*. 2010).

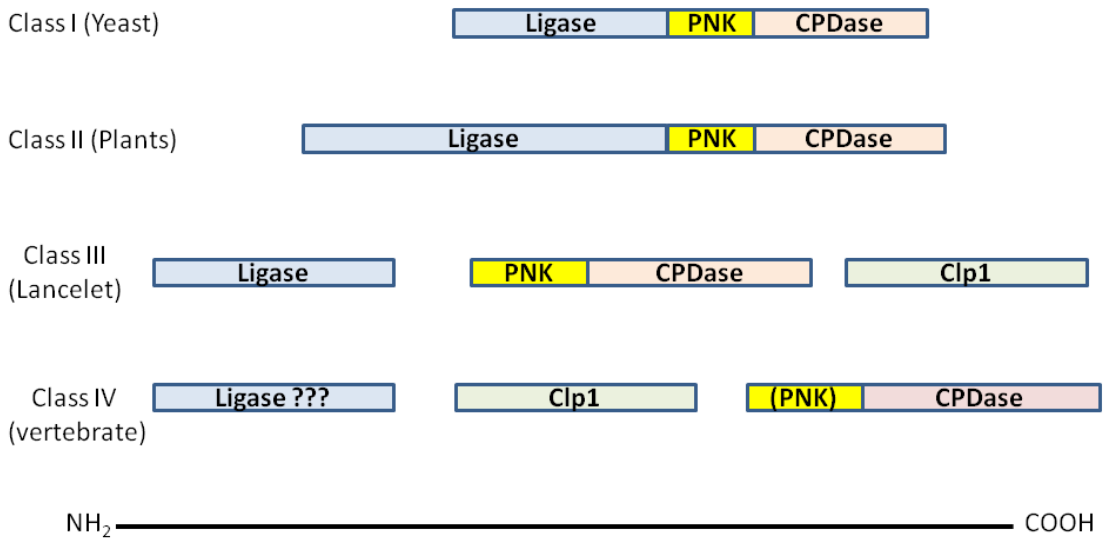


Figure 10. Classification of 5'-P RNA ligases and domain organization. Class I and Class II enzymes contain the ligase, PNK and CPDase activities in one polypeptide. The recently identified class III enzyme contains the PNK and CPDase domains in a single polypeptide and the ligase domain in another polypeptide. The class IV RNA ligase is not known yet. The class IV PNK (Clp1) and CPDase enzymes have been identified. [Scheme modified from (Englert, Sheppard *et al*. 2010)].

Initial functional analysis of the kinase module of yeast Trl1, using single alanine mutations in the P-loop, has revealed that the P-loop motif (^{401}GCGKT^{405}) is a determinant of NTP binding (Sawaya, Schwer *et al*. 2003). GTP is the preferred *in vivo* physiological substrate, and the yeast Trl1 contains a single NTP binding site (Sawaya, Schwer *et al*. 2003). The GTP dependence of the yeast Trl1 kinase domain has also been verified *in vitro* (Westaway, Belford *et al*. 1993).

The CPDase domain of both yeast Trl1 and *Bf* PNK/CPDase bears resemblance to the so-called "2H" superfamily of phosphoesterases, which is defined by the presence of two conserved "H-x-(T/S)-x" (where "x" is a hydrophobic residue) motifs (Nasr and Filipowicz 2000, Mazumder, Iyer *et al*. 2002, Englert, Sheppard *et al*. 2010). The presence of the P-loop motif in the N-terminal domain and the two "H-x-(T/S)-x" motifs in the C-terminal domain of vertebrate CNPase indicates that the PNK/CPDase domains of yeast Trl1 (*Sc* PNK/CPDase) and the *Bf* PNK/CPDase are homologs of vertebrate CNPase (Englert, Sheppard *et al*. 2010). Although 5'-P RNA ligases of

fungi, plants and animals [Figure 10] share essential mechanistic features and key residues required for their PNK and CPDase activities, their overall sequence similarity is low.

1.4. 2H phosphoesterase superfamily

Members of the 2H phosphoesterase superfamily exhibit extreme sequence divergence, although the active site motifs are conserved. This explains the difficulty in identification of their homologs through sequence analysis and the requirement for a combination of a variety of sensitive sequence analysis methods (Mazumder, Iyer *et al*. 2002). Multiple alignment of the entire 2H phosphoesterase family members has revealed that both the "H-x-(T/S)-x" motifs are almost absolutely conserved and that threonine is found in 86% of the motifs (Mazumder, Iyer *et al*. 2002). The only variants of the "H-x-(T/S)-x" motifs have been detected in the C-terminus of *Arabidopsis* At5g40190-like family and *Agrobacterium* AGR_C_4233 protein, where the histidine is replaced by phenylalanine and glutamine, respectively (Mazumder, Iyer *et al*. 2002).

Prior to clustering of the members under the 2H phosphoesterase (PE) superfamily, several proteins had been found to contain the two conserved "H-x-(T/S)-x" motifs and to play crucial roles in nucleic acid metabolism and cellular signaling with 2',3'-cyclic nucleotide phosphodiesterases being a typical example (Culver, McCraith *et al*. 1993, Phizicky and Greer 1993). The CPDase activity has been described in yeast, where it is implicated in the metabolism of Appr>p (Culver, Consaul *et al*. 1994). Another example is the bacterial ligase, LigT that contains the "2H" conserved motifs. Although bacterial tRNA splicing does not involve any cyclic phosphate RNA intermediate, the *E. coli* LigT has been found to ligate eukaryotic tRNA halves through cleavage of the 2',3'-cyclic phosphate (Greer, Javor *et al*. 1983, Arn and Abelson 1996). The vertebrate brain CNPase is yet another enzyme that catalyzes the hydrolysis of similar phosphodiester bonds in cyclic nucleotides, oligonucleotides and nicotinamide adenine dinucleotide 2′,3′-cyclic monophosphate (2',3'-cNADP$^+$) (Olafson, Drummond *et al*. 1969, Sprinkle 1989, Myllykoski and Kursula 2010). All detected members of the 2H phosphoesterase superfamily are classified into four groups with a few unclassified and more divergent members (Mazumder, Iyer *et al*. 2002).

1.4.1. Group 1: The archaeo-bacterial LigT-like group

Members of the archaeo-bacterial LigT-like group bear similarity to the *E. coli* LigT (Arn and Abelson 1996). Orthologs of LigT are spread across all archaeal genomes and more sporadically in bacteria (Mazumder, Iyer *et al*. 2002). Enzymes of this group

catalyze the hydrolysis of 2',3'-cyclic phosphate in the 5'-tRNA half to generate a 2'-phosphate that is directly joined to the 5'-OH group of the 3'-tRNA half through an atypical 2'-5' phosphodiester linkage (Arn and Abelson 1996, Kanai, Sato *et al*. 2009). The fungal tRNA ligases are the most closely related 2H phosphoesterases to the archaeo-bacterial LigT-like group (Phizicky, Schwartz *et al*. 1986, Mazumder, Iyer *et al*. 2002). Unlike LigT, the fungal tRNA ligases form 3'-5' phosphodiester linkages during tRNA splicing (Phizicky, Schwartz *et al*. 1986, Wang and Shuman 2005). Plant RNA ligases with similar functions and broader substrate specificity have been detected; however, despite the presence of the 2H motifs, the plant enzymes show no sequence similarity to the four known groups of the 2H phosphoesterase family (Englert and Beier 2005).

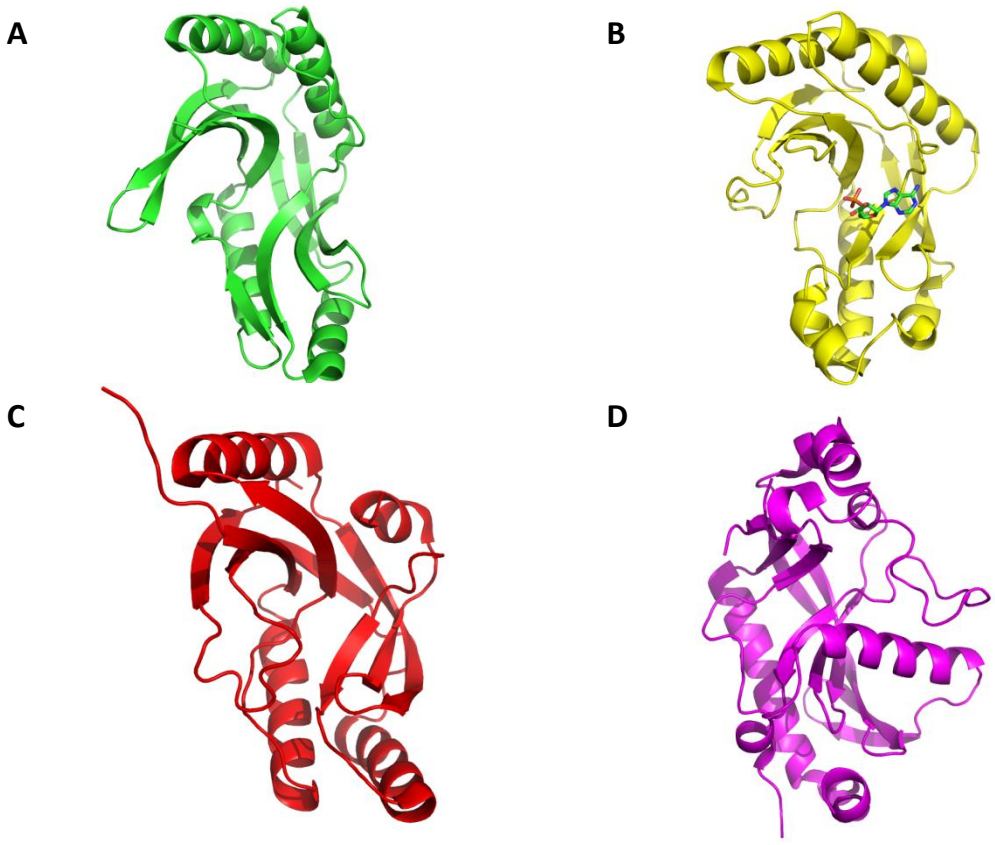


Figure 11. Structures of the members of 2H phosphoesterase superfamily. A) 2'-5' RNA ligase from *Pyrococcus horikoshii* (PDB ID: 1VGJ) (Gao, Yao et al. 2006). **B)** The AKAP18 central domain with adenosine (stick) bound (PDB ID: 2VFK) (Gold, Smith et al. 2008). **C)** *Arabidopsis thaliana* CPDase (PDB ID: 1FSI) (Hofmann, Zdanov et al. 2000). **D)** The C-terminal domain of human CNP (PDB ID: 1WOJ) (Sakamoto, Tanaka et al. 2005). The figures were generated using PyMOL (www.pymol.org).

The archaeo-bacterial LigT-like group contains three families of RNA ligases: archaeal 2'-5' RNA ligases, bacterial 2'-5' RNA ligases and fungal tRNA ligases. Atomic structures of 2'-5' RNA ligase from the archaea, *T. thermophilus, P. furiosus* and *P.horikoshii* [Figure 11A] have been determined and used to describe the active site containing the two "2H" motifs (Kato, Shirouzu *et al*. 2003, Rehse and Tahirov 2005, Gao, Yao *et al*. 2006, Kanai, Sato *et al*. 2009).

1.4.2. Group 2: The eukaryotic-viral LigT-like group

The eukaryotic-viral LigT-like group consists of three families: eukaryotic LigT-like proteins, RNA virus LigT-like proteins and *Drosophila melanogaster* CG16790-like proteins (Mazumder, Iyer *et al*. 2002). The eukaryotic LigT-like family is prototyped by the human CGI-18 gene and spread across plants, animals, fungi and *Cryptosporidium parvum* (Mazumder, Iyer et al. 2002). The human proteins in the eukaryotic LigT-like family are all constituents of multiprotein complexes. These human proteins are the activating signal cointegrator 1 complex subunit 1 (ASCC1), the protein kinase A anchoring protein 18 (AKAP18) and leukocyte receptor cluster member 9 (Mazumder, Iyer et al. 2002, Myllykoski 2013). The crystal structure of AKAP 18 δ-isoform has been determined [Figure 11B] (Gold, Smith et al. 2008). AKAP18 does not react with 2',3'- or 3',5'-cyclic nucleotides, indicating that the protein does not share the catalytic activity of the 2H PE superfamily (Gold, Smith *et al*. 2008).

The viral LigT-like family includes proteins from the type C rotaviruses (VP3 protein) and the coronaviruses (NS2 protein) (Mazumder, Iyer *et al*. 2002). The type C rotaviruses are double-stranded multipartite RNA viruses, whereas the coronaviruses are positive strand RNA viruses. The rotaviral protein VP3 plays a crucial role in capping of the transcribed viral mRNA during proliferation of the virus (Chen, Luongo *et al*. 1999). The 2H motif of the coronaviral non-structural protein (NS2) is essential for organ-specific viral replication in the liver (Roth-Cross, Stokes *et al*. 2009). It has been suggested that the coronoviral proteins might take part in the ligation process that most likely generates the subgenomic RNAs containing the same 5'-terminus of the genomic RNA (Sawicki and Sawicki 1998). The *Drosophila melanogaster* CG16790-like family has a similar phylogenetic profile as the eukaryotic LigT-like family and is typified by the CG16790 gene of *Drosophila* (Mazumder, Iyer *et al*. 2002). Members of this family are present in animals, plants, fungi (except *S. cerevisiae*) and *Entamoeba histolytica* (Mazumder, Iyer *et al*. 2002).

1.4.3. Group 3: The YjcG-like group

Members of the YjcG-like group are similar to the *Bacillus subtilis* YjcG protein and grouped under four families: the *B. subtilis* YjcG-like protein family, the mll4975-like protein family, the UBASH3 family and the At5g40190 family (Mazumder, Iyer *et al*. 2002). The members of *B. subtilis* YjcG-like family are found in Gram-positive bacteria, actinomycetes and cyanobacteria, and *Deinococcus* (Mazumder, Iyer *et al*. 2002). The members of the mll4975-like protein are related to the *Mesorhizobium loti* protein mll4975 and restricted to α-proteobacteria (Mazumder, Iyer *et al*. 2002). These proteins co-occur in operons that are implicated in the metabolism of phosphonates, which contain a carbon-phosphate bond and get catabolized by a carbon-phosphate lyase pathway (Metcalf and Wanner 1993, Mazumder, Iyer *et al*. 2002). The product of the cleavage pathway is a 1',2'-cyclic 5'-ribose, which could be a possible substrate for the 2H phosphoesterase activity (Kamat, Williams *et al*. 2011).

Members of the UBASH3 family are restricted to coelomates, and contain multiple domains with the 2H domain fused to a ubiquitin-associated (UBA) domain, an SH3 domain and a phosphoglyceromutase domain (Wattenhofer, Shibuya *et al*. 2001, Mazumder, Iyer *et al*. 2002). The phosphoglyceromutase generates 2,3 phosphate containing polyols, and the combination of 2H and phosphoglyceromutase domains has been suggested to play a role in processing RNA substrates (Mazumder, Iyer et al. 2002). Members of the plant-specific At5g40190 family lack the C-terminal "H-x-(T/S)-x" motif and have been suggested to be either inactive or to contain an alternative function with the requirement of a single "H-x-(T/S)-x" motif (Mazumder, Iyer *et al*. 2002).

1.4.4. Group 4: The mlr3352-like group

The mlr3352-like group of 2H phosphoesterases contains a single family typified by the mlr3352 protein from *Mesorhizobium loti* (Mazumder, Iyer *et al*. 2002). Members of this family are found in various α-proteobacteria. The crystal structure of one of the closest homologues of mlr3352, the hypothetical protein ATU0111 from *Agrobacterium tumefaciens* has been determined (PDB ID: 2FSQ). A member of this group has been found in a large eukaryotic DNA virus, suggesting that 2H proteins might play a role in viral tRNA metabolism (Mazumder, Iyer *et al*. 2002).

1.4.5. Divergent members of the 2H superfamily

Several proteins of the 2H phosphoesterase superfamily could not be included in the above-described major families due to sequence divergence (Mazumder, Iyer *et al*. 2002). However, the more divergent forms have been the more extensively studied members of the superfamily. The vertebrate versions of the divergents, such as the C-

terminal catalytic domain of vertebrate CNPases, have been fused to the N-terminal domain that contains the P-loop in the course of evolution (Koonin and Gorbalenya 1990). Thus, the vertebrate CNPases are homologues of the PNK/CPDase domains of the yeast Trl1 and the *Bf* PNK/CPDase (Englert, Sheppard *et al.* 2010).

The CPDase enzymes from yeast, wheat germ and *Arabidopsis thaliana* are another set of divergent members of the 2H PE superfamily (Culver, Consaul *et al*. 1994, Hofmann, Zdanov *et al*. 2000, Nasr and Filipowicz 2000, Mazumder, Iyer *et al*. 2002). As discussed previously, the yeast and plant CPDases are involved in tRNA splicing and catalyze the hydrolysis of Appr>p and 2',3'-cyclic nucleotides to generate ADP-ribose 1''-phosphate (Appr-1''p) and 2'-nucleotides, respectively (Culver, Consaul *et al*. 1994, Hofmann, Zdanov *et al*. 2000, Nasr and Filipowicz 2000). Despite differences in substrate specificity, the yeast and plant CPDases share several biochemical properties (Hofmann, Zdanov *et al*. 2000). Thus, the yeast and plant CPDases constitute one group within the divergent forms of the 2H PE superfamily. Interestingly, the first crystal structure determined for a member of the 2H phosphoesterase family is that of CPDase from *A. thaliana* [Figure 11C] (Hofmann, Zdanov *et al*. 2000). The structure shows that the enzyme exists as a single compact domain with two similar lobes and that the active site "2H" motifs are located between the two lobes in a water-filled cavity (Hofmann, Zdanov *et al*. 2000).

The regeneration-induced CNPase homolog (RICH) proteins constitute another family of divergent members. The RICH proteins are close homologues of the vertebrate CNPases and found in the axons of regenerating optic nerves in goldfish and zebrafish (Ballestero, Wilmot *et al*. 1995, Ballestero, Dybowski *et al*. 1999). Structural analysis has revealed that the RICH proteins consist of three regions: a glutamate- and aspartate-rich N-terminal domain, a cyclic phosphodiesterase domain, and a C-terminal isoprenylation site. The catalytic domain alone is sufficient for the activity of RICH *in vitro* (Challa, Chapa *et al*. 2006).

The FPV025 protein of the fowlpox virus, Cpd1p of the yeast, Cgl1020 protein of *Corynebacterium*, Faci_p1766 protein of *Ferroplasma*, the polyproteins of piscine retroviruses and the plant 2',3'-CPDases are other known divergent members of the 2H superfamily (Mazumder, Iyer *et al*. 2002). The *M. mazei* MM1887 protein and the WSV147 protein of the shrimp white spot syndrome virus bear no affinity to any group (Mazumder, Iyer *et al*. 2002).

1.5. 2',3'-cyclic nucleotide 3'-phosphodiesterase

The vertebrate CNPase is abundantly expressed in the myelin sheath, a multilayered proteolipid membrane system. CNPase constitutes 4% of total myelin protein in the

central nervous system (CNS) and about 0.4% in the peripheral nervous system (PNS) (Bifulco, Laezza *et al*. 2002). CNPase is expressed as two isoforms, CNP1 (401 amino acids) and CNP2 (421 amino acids), CNP2 being longer at the N-terminus by 20 amino acid residues (Kurihara, Takahashi *et al*. 1988, Kurihara, Monoh *et al*. 1990, Gravel, DeAngelis *et al*. 1994). The N-terminal extension of CNP2 functions as a mitochondrial targeting signal and is regulated by phosphorylation of N-terminal serine residues by protein kinase C (O'Neill and Braun 2000, Lee, O'Neill *et al*. 2006). Both isoforms are enzymatically active and isoprenylated at the C-terminus (De Angelis and Braun 1994). The isoprenylated cysteine residue at the C-terminus mediates membrane binding (Braun, De Angelis *et al*. 1991, De Angelis and Braun 1996). Mutational studies have shown that CNPase is involved in interactions between axons and glia at the nodes of Ranvier in the CNS, and that the protein facilitates proper myelination of axons with a small diameter (Rasband, Tayler *et al*. 2005, Edgar, McLaughlin *et al*. 2009).

Although myelin in CNPase-deficient mice has been found to be normal in ultra-structure and protein composition, they have developed progressive motor deficits and died prematurely due to diffuse brain axonal swelling and neurodegeneration (Lappe-Siefke, Goebbels *et al*. 2003). However, it is not known whether the late onset of neurodegeneration is caused by the absence of the CNPase protein as a structural component of the myelin sheath or the absence of CNPase enzymatic activity. CNPase has been implicated in neurodegenerative disorders, such as multiple sclerosis and schizophrenia (Muraro, Kalbus *et al*. 2002, Peirce, Bray *et al*. 2006). A recent study shows that CNPase inhibits the assembly of infectious particles of several primate lentiviruses, including the human immunodeficiency viruses HIV-1 and HIV-2, by binding to the structural protein Gag (Wilson, Schoggins *et al*. 2012). CNPase from human, macaque, grivet, cow and sheep has been found to possess antiviral activity, whereas mouse CNPase has not shown any such activity (Wilson, Schoggins *et al*. 2012).

The enzymatic activity of CNPase, the hydrolysis of 2',3'-cyclic nucleotides to 2'-nucleotides, was detected in the brain tissue in the 1960s (Drummond, Iyer *et al*. 1962). The activity of CNPase is specific to 2',3'-cyclic nucleotides, as the enzyme does not hydrolyze ribonucleoside 3',5'-cyclic phosphate (Drummond, Iyer *et al*. 1962). Atomic structures of the C-terminal catalytic domain of human [Figure 11D], rat and mouse CNPase have been determined (Kozlov, Lee *et al*. 2003, Sakamoto, Tanaka *et al*. 2005, Myllykoski, Raasakka *et al*. 2012, Myllykoski, Raasakka *et al*. 2013). These structures have illuminated the possible catalytic roles of the "H-x-(T/S)-x" motifs and other functionally important groups in the active site. The two "H-x-(T/S)-x" motifs are located in the active site. It has been proposed that the histidines and threonines coordinate the phosphodiester to be cleaved and facilitate its hydrolysis

by stabilization of the transition state and general acid-base catalysis. Although sequence similarity among the different families of 2H PE superfamily is remarkably low, the active sites of members of different families are quite similar, suggesting a possible conservation of the catalytic mechanism.

Since the *in vitro* substrate of CNPase, the 2',3'-cyclic nucleotide, is not found in considerable amounts *in vivo*, the actual physiological substrate of CNPase activity is not known (Gravel, Robert *et al*. 2009, Myllykoski and Kursula 2010). However, a mammalian CNPase has been found to carry out the essential 3'-end healing steps of tRNA splicing in yeast, suggesting an RNA processing function *in vivo* for the mammalian CNPase (Schwer, Aronova *et al*. 2008). CNPase has also been shown to interact with mRNA *in vivo* and inhibit translation *in vitro* (Gravel, Robert *et al*. 2009). Since CNPase interacts with RNA and tubulin simultaneously *in vitro*, a possible role in RNA trafficking has been suggested (Gravel, Robert *et al*. 2009).

The structure and function of the N-terminal domain of CNPase have not been characterized yet. The expression and purification of the N-terminal domain has been reported to be difficult as compared to the C-terminal domain, limiting the availability and characterization of the molecule (Myllykoski and Kursula 2010). The N-terminal domain is not essential for the phosphodiesterase activity *in vitro* (Lee, Gravel *et al*. 2001). CNPase has been shown to bind purine nucleotides and weakly hydrolyze purine triphosphates (Stingo, Masullo *et al*. 2007). Although the N-terminal domain of CNPase contains the NTP-binding P-loop motif and the study by Stingo *et al*. claims that nucleotide binding occurs in the N-terminal domain, it has not been experimentally proven, whether the interaction and hydrolysis take place exclusively in the N-terminal domain, as only the full-length CNPase has been used in the experiments (Stingo, Masullo *et al*. 2007). Recent studies have shown that the N-terminal domain of CNPase binds RNA, mediates dimerization and directly interacts with calmodulin (CaM) in a calcium-dependent manner (Myllykoski, Itoh *et al*. 2012, Myllykoski, Raasakka *et al*. 2012). However, CaM binding to the N-terminal domain does not affect the cyclic phosphodiesterase activity or the RNA-binding property of CNPase (Myllykoski, Itoh *et al*. 2012).

The available knowledge on the structure and function of the homologues of tRNA healing enzymes, such as the different members of the 2H phosphoesterase superfamily and certain members of the P-loop containing phosphotransferases, could be used to characterize the PNK/CPDase enzymes. Since the vertebrate CNPase proteins are among the extensively studied molecules of 2H phosphoesterase superfamily, the available findings could aid in designing experiments to unravel the structural and functional properties of PNK/CPDase. The conserved motifs, such as the two "H-x-(T/S)-x" phosphoesterase motifs and the NTP-binding P-loop motif, in

PNK/CPDase homologues have been analyzed both structurally and functionally, and their importance for the catalytic activity of the enzymes have been elucidated.

Although the structure of C-terminal phosphodiesterase domain of vertebrate CNPase has been determined, no high-resolution structure of either the full-length CNPase or its N-terminal domain is available. The functional properties of vertebrate CNPase, such as interactions with membranes, cytoskeletal proteins and RNA have been studied; however its *in vivo* physiological function remains enigmatic. Hence, structural and functional characterization of PNK/CPDase would benefit, not only the field of tRNA splicing, but also the homologues in the 2H phosphoesterase family. Since the newly identified *Bf* PNK/CPDase is a participant in a new class of 5'-P RNA ligation, characterization of this enzyme might improve our understanding of the growing complexity of animal tRNA splicing. Despite genetic and biochemical analysis of PNK and CPDase domains of yeast Trl1, no structural information is available for any domain of the enzyme. Therefore, detailed biophysical and biochemical analysis of *Sc* PNK/CPDase and *Bf* PNK/CPDase is warranted.

2. Objectives of the study

Despite their significance in tRNA metabolism, the structure and function of tRNA healing enzymes have not yet been fully understood. This study aims to investigate the structural and functional properties of the PNK and CPDase domains of the tri-functional yeast Trl1, and the lancelet PNK/CPDase, by applying a combination of different biophysical and biochemical methods. The objectives of this study were:

i) To optimize the recombinant production, purification of the PNK/CPDase domains of yeast Trl1 and lancelet PNK/CPDase and to analyze their folding state and thermal stability.

ii) To determine their oligomeric state, shape and conformation in solution.

iii) To study their enzymatic activity, ligand specificity and changes in conformation upon ligand binding.

iv) To crystallize both PNK and CPDase domains and each of them separately, in the presence and absence of their substrates.

v) To compare the structural and functional properties of PNK/CPDase enzymes to their homologues from the 2H phosphoesterase superfamily, especially the vertebrate CNPase.

3. Materials

3.1. Laboratory equipment

Equipment	Manufacturer
ÄKTAexplorer	GE Healthcare, Sweden
ÄKTApurifier	GE Healthcare, Sweden
Analytical balance	Sartorius, Germany
Astacus distillation unit	MembraPure, Germany
Avanti J26-XP centrifuge	Beckman Coulter, USA
Biacore T100	GE Healthcare, Sweden
CERTOMAT^{R} IS benchtop incubator	Sartorius, Germany
CFX96 RealTime System	Bio-Rad, Germany
ChiraScan Plus spectrophotometer	Applied Photophysics, UK
DynaPro NanoStar^{TM}	Wyatt, Germany
Electrophoresis unit	Bio-Rad, Germany
Gel documentation system	PEQ Lab, Germany
Genie vortex	Scientific Industries, Germany
Gyromini^{TM} nutating mixer	Labnet International, Germany
Heraeus^{TM} FRESCO21^{TM} centrifuge	Thermo Scientific, Germany
Laminar air flow chamber	Kojair Tech Oy, Finland
Mastercycler^{R} gradient	Eppendorf, Germany
MiniDAWN^{TM} TREOS detector	Wyatt, Germany
Multitron Pro shaker	Infors, Germany
Nanodrop 2000 spectrophotometer	Thermo Scientific, Germany
Optilab^{R} T-rEX refractometer	Wyatt, Germany
pH meter	Mettler Toledo, Germany
Sonopuls Sonifier	Bandelin, Germany
Systec VX150 autoclave	Systec, Germany
Tabletop centrifuge 5810-R	Eppendorf, Germany
TECAN infinite M200 fluorometer	TECAN, Germany
Thermomixer comfort	Eppendorf, Germany
Tube rotator	Stuart, UK
VARIOMAG^{R} magnetic shaker	Thermo Scientific, Germany

3.2. Laboratory consumables

Consumable	Supplier
Amicon-Ultra centrifugal filter units	Millipore, Ireland
Assay plates (96-well)	Greiner Bio-One, Germany
Dialysis membranes	Carl Roth, Germany
Disposable plastic cuvettes	Carl Roth, Germany
Erlenmeyer flasks	Schott Duran, Germany
Falcon tubes	Greiner Bio-One
Gravity-flow columns	Bio-Rad, Germany
Inoculation loops	Greiner Bio-One
Low 96-well clear plate	Bio-Rad, Germany
Microcentrifuge tubes	Eppendorf, Germany
Microseal adhesive films	Bio-Rad, Germany
Mini-PROTEAN^{R} TGX^{TM} precast gels	Bio-Rad, Germany
Polymerase chain reaction (PCR) tubes	Brand, Germany
Pipette tips	Sartorius, Germany
Plastic Petri dishes	Sarstedt, Germany
Plastic syringes	Braun Melsungen, Germany
Serological pipettes	Greiner Bio-One, Germany
Slide-A-Lyzer mini dialysis units	Thermo Scientific, Germany
Syringe filters	Millipore, Germany
Vivaspin 20 concentrators (MWCO: 10 & 30K)	Sartorius Stedim Biotech, Germany
PD-10 columns	GE Healthcare, UK

3.3. Chemicals

Chemicals used in this study were of analytical grade and were purchased from Carl Roth (Germany), Sigma Aldrich (Germany), Roche Diagnostics (Germany), Calbiochem (Germany), GE Healthcare (Sweden) and AppliChem (Germany), unless stated otherwise.

3.4. Kits, spin columns and reagents

QIAprep^{R} Spin mini-prep kit	Qiagen, Germany
QIAquick^{R} Gel Extraction Kit	Qiagen, Germany
QIAquick^{R} PCR Purification Kit	Qiagen, Germany
Quick-Load DNA Ladder	New England Biolabs, Germany
PageRuler prestained protein ladder	Thermo Scientific, Lithuania

3.5. Vectors

pET20b-*Bf*PNK/CPDase	Prof. Dieter Söll's Group, Yale University, USA
pET28a-*Sc*Trl1 (389-827 residues)	Prof. Dieter Söll's Group, Yale University, USA
pETNKI-his3C-LIC-amp	Protein facility of the Netherlands Cancer Institute, Amsterdam, The Netherlands

3.6. Growth media and antibiotics

LB medium	Carl Roth, Germany
Ampicillin	Carl Roth, Germany
Chloramphenicol	Carl Roth, Germany
Kanamycin	Carl Roth, Germany

3.7. Strains of *Eschericia coli*

Cloning strain:	
NEB5α	New England Biolabs, Germany
Expression strains:	
BL21 (DE3)	New England Biolabs, Germany
BL21 CodonPlus (DE3) RIPL	Agilent Technologies, Germany
Rosetta (DE3)	Novagen, Germany

3.8. Enzymes, substrates and nucleotides

Ambion^{R} RNase cocktail	Life Technologies, Germany
EDTA-free protease inhibitor cocktail tablets	Roche Applied Science, Germany
2',3'-cNADP^{+}	Sigma-Aldrich, USA
Glucose-6-phosphate	Sigma-Aldrich, USA
Glucose-6-phosphate dehydrogenase	Sigma-Aldrich, USA
KpnI	New England Biolabs, Germany
Mant-AMP, Mant-ATP & Mant-GTP	Jena Bioscience, Germany
PCR primers	Eurofins MWG, Germany
Phusion^{R} high-fidelity PCR master mix	Thermo Fisher Scientific, Germany
SYPRO Orange dye	Molecular Probes, Germany
T4 DNA polymerase	New England Biolabs, Germany

3.9. Materials for chromatography

Resins:	
Ni-NTA agarose	GE Healthcare, Sweden
Columns:	
HisTrapFF (1ml) column	GE Healthcare, Sweden
HiLoad 16/60 Superdex 200 column	GE Healthcare, Sweden
HiLoad 16/60 Superdex 75 column	GE Healthcare, Sweden
Superdex 200 10/300 GL	GE Healthcare, Sweden
Superdex 75 10/300 GL	GE Healthcare, Sweden

3.10. Growth media

3.10.1. Lysogeny broth medium

The lysogeny broth (LB) medium is the most widely used, nutritionally-rich medium for the growth of bacteria. The composition of LB medium (Bertani 1951), used in this study, is given below:

Tryptone	1.0% (w/v)
Yeast extract	0.5% (w/v)
NaCl	1.0% (w/v)

The medium was prepared in double-distilled water (ddH_2O) and sterilized using an autoclave at 121°C for 15 min, and stored at 4°C until used.

3.10.2. Auto-induction medium

The auto-induction (AI) medium, introduced by William Studier (Studier 2005), allows spontaneous induction of protein expression, in isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible *E. coli* strains, when the cells reach high density close to saturation phase. The medium contains a limited amount of glucose, which prevents uptake of lactose until it is depleted. The glucose is metabolized during the initial phase of bacterial cell growth and depleted in mid to late log phase. At this stage, lactose is taken up by the cells and converted to the natural inducer, allolactose, by β-galactosidase. Allolactose causes the release of lactose repressor from its binding sites in DNA and induces the expression of T7 polymerase, which in turn induces the expression of target proteins. The ZYM-5052 auto-induction medium was used in this study and its final composition is given below:

N-Z-amine AS	1.0%
Yeast extract	0.5%
Na_2HPO_4	25 mM
KH_2PO_4	25 mM
NH_4Cl	50 mM
Na_2SO_4	5 mM
$MgSO_4$	2 mM
1000 X Trace metals	0.2 X
Glycerol	54 mM
Glucose	2.8 mM
α-Lactose	5.6 mM

3.11. Buffers and solutions

3.11.1. Transformation buffer

$MnCl_2.4H_2O$	55 mM
$CaCl_2.2H_2O$	15 mM
KCl	250 mM
Piperazine-N,N′-bis(2-ethanesulfonic acid) (PIPES) (0.5M, pH 6.7)	10 mM
H_2O	to 1 litre

The buffer was filter sterilized and stored at 4°C until use.

3.11.2. Buffers for agarose gel electrophoresis

1X Tris-acetate-EDTA (TAE) buffer	40 mM tris(hydroxymethyl)aminomethane (Tris), 1 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0
6X DNA loading buffer	10 mM Tris-HCl (pH 7.6) 0.03% (w/v) bromophenol blue 0.03% (w/v) xylene cyanol 60 mM EDTA 60% glycerol

3.11.3. Buffers and solutions for sodium dodecyl sulphate polyacrylamide gel electrophoresis

1X sodium dodecyl sulphate (SDS) running buffer	25 mM Tris, 192 mM glycine and 0.1% (w/v) SDS

6X SDS loading buffer	375 mM Tris (pH 6.8), 12% (w/v) SDS, 60% glycerol, 600 mM dithiothreitol (DTT), 0.06% (w/v) bromophenol blue
Coomassie staining solution	0.05% (w/v) Coomassie brilliant blue, 10% (v/v) acetic acid and 25% (v/v) ethanol
Destaining solution	10% (v/v) acetic acid

3.11.4. Buffers for immobilized-metal affinity chromatography

Buffer A (lysis buffer)	50 mM hydroxyethyl piperazineethanesulfonic acid (HEPES) (pH 7.5), 250 mM NaCl, 10 mM β-mercaptoethanol (β-ME), 1 mg/ml lysozyme, 20 µg/ml deoxyribonuclease (DNase), 100 µl Ambion^{R} ribonuclease (RNase) cocktail, 10 mM imidazole and EDTA-free protease inhibitor cocktail
Buffer B (equilibration buffer)	50 mM HEPES (pH 7.5), 250 mM NaCl, 10 mM β-ME and 10 mM imidazole
Buffer C (wash buffer 1)	50 mM HEPES (pH 7.5), 250 mM NaCl, 10 mM β-ME and 50 mM imidazole
Buffer D (wash buffer 2)	50 mM HEPES (pH 7.5), 1000 mM NaCl, 10 mM β-ME and 50 mM imidazole
Buffer E (elution buffer)	50 mM HEPES (pH 7.5), 250 mM NaCl, 10 mM β-ME and 500 mM imidazole
Buffer F (dialysis buffer)	50 mM HEPES (pH 7.5), 250 mM NaCl and 10 mM β-ME

3.11.5. Buffers for size exclusion chromatography

Buffer 1	50 mM HEPES (pH 7.5), 200 mM NaCl and 10 mM β-ME
Buffer 2	50 mM HEPES (pH 7.5), 200 mM NaCl and 5 mM β-ME
Buffer 3	50 mM HEPES (pH 7.5), 200 mM NaCl and 5 mM DTT

3.12. Bioinformatics tools used

The following bioinformatics tools were used to extract information from both nucleotide and amino acid sequences of the expression constructs studied. Properties, such as sequence similarity, structural and functional homology to families of proteins, conserved sequence patterns in the course of evolution, secondary and tertiary structural features, regions of flexibility and disorderness, were predicted from the sequence using the programs described below.

3.12.1. T-Coffee

T-Coffee (Tree-based Consistency Objective Function For alignment Evaluation) is a tool for fast and accurate alignment of multiple sequences (Notredame, Higgins *et al*. 2000). The T-Coffee algorithm has two main features: a simple and flexible multiple sequence alignment using data from heterogeneous sources, and the optimization method that finds the multiple sequence alignment with the best fit to the pair-wise alignments in the input library (Notredame, Higgins *et al*. 2000). The alignment accuracy is remarkably improved by the use of a combination of local and global alignments (Notredame, Higgins *et al*. 2000).

3.12.2. SMART

SMART (Simple Modular Architecture Research Tool) is a web-based tool for the identification of genetically mobile domains and for the analysis of domain architectures (Schultz, Copley *et al*. 2000). SMART is able to detect more than 400 domain families that are present in signaling, extra cellular and chromatin associated proteins (Schultz, Copley *et al*. 2000). Proteins with specific combinations of domains can be found through a user interface to a relational database system (Schultz, Copley *et al*. 2000).

3.12.3. PSIPRED

PSIPRED is a web-based tool for the prediction of protein secondary structure based on position-specific scoring matrices generated by PSI-BLAST (Jones 1999). PSIPRED uses a two-stage neural network to predict the secondary structure, and the results have been found to be superior to those predicted by other tools/ methods (Jones 1999).

3.12.4. IUPred

IUPred is a web server for the prediction of intrinsically disordered regions of proteins based on estimated energy content (Dosztányi, Csizmok et al. 2005). The IUPred

applies an algorithm that estimates the total pairwise interresidue interaction energy to predict the intrinsically unstructured sequences, assuming that such regions do not fold due to their inability to form sufficient stabilizing interresidue interactions (Dosztányi, Csizmók et al. 2005). Slightly different parameters are employed to predict long disorder, short disorder, and structured domains (Dosztányi, Csizmok et al. 2005).

3.12.5. Phyre^2

The protein homology/analogy recognition engine (Phyre) version 2.0, available at http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index, is a server for homology modeling (Bennett-Lovsey, Herbert *et al*. 2008, Kelley and Sternberg 2009). The server uses the available protein structures from the Structural Classification of Proteins (SCOP) and the Protein Data Bank (PDB) databases. It first creates a profile by scanning the query sequence against a nonredundant sequence database and predicts the secondary structure of the query protein by using three independent programs, Psi-Pred (McGuffin, Bryson *et al*. 2000), SSPro (Pollastri, Przybylski *et al*. 2002) and JNet (Cole, Barber *et al*. 2008). The profile and predicted secondary structure are then scanned against a fold library, which contains the profiles of all available structures in SCOP and PDB, by using a profile-profile alignment algorithm. The resulting alignments are ranked, and the top ten alignments are used to generate 3D models of the query. Side chains are included into the model with the help of a fast graph-based algorithm and a library of side-chain rotamers. In this study, the sequences of different expression constructs were entered as input and the server was instructed to perform modeling in normal mode.

3.12.6. BLAST

The basic local alignment search tools (Altschul, Gish *et al*. 1990), Blastp and Blastn, were used to search for sequences similar to the given sequence of amino acids and nucleotides respectively. The BLAST tools were mainly used to check the similarity between the sequences of the newly constructed recombinant plasmids and those of the same proteins available at the nucleotide database of the National Center for Biotechnology Information (NCBI).

3.12.7. ProteinCCD

The Protein Crystallization Construct Designer (ProteinCCD) is a tool, available at http://xtal.nki.nl/ccd, used to design multiple truncation constructs to be used for protein expression and crystallization (Mooij, Mitsiki *et al*. 2009). ProteinCCD is a meta-server that runs several prediction programs from other web servers and gathers information about secondary structure, disordered regions, coiled coils, flexible

linkers and transmembrane domains. The output can be analyzed and used by the user to choose the possible start and end of appropriate protein constructs on an interactive screen. In this study, the ProteinCCD tool was used for the identification of suitable segments of the PNK and CPDase domains of yeast Trl1 for automated design of oligonucleotides for PCR amplification of those segments.

3.12.8. ParCrys

The ParCrys tool, available at www.compbio.dundee.ac.uk/parcrys, was used to assess the propensity of the different constructs to produce diffraction-quality crystals. The program uses data from PDB, TargetDB and PepcDB to estimate the crystallization propensity by applying hydrophobicity, isoelectric point and standard amino acid frequencies as predictive features (Overton, Padovani *et al*. 2008).

3.12.9. XtalPred

XtalPred is another web server, available at http://ffas.burnham.org/XtalPred, which was used to predict the crystallizability of proteins. The prediction is based on the comparison of the protein's features with those available in TargetDB (Slabinski, Jaroszewski *et al*. 2007). This server also suggests possible ligands that might aid in increasing the crystallizability and also lists proteins of the same family from different organisms and their predicted ability to form crystals.

3.12.10. ExPASy tools

A few tools of the Expert Protein Analysis System (Artimo, Jonnalagedda *et al*. 2012), such as ExPASy Translate and Protparam, were routinely used in this study. The Translate tool was used to translate a nucleotide sequence into a protein sequence. ProtParam was used to compute parameters, such as molecular weight, theoretical isoelectric point, hydrophobicity and extinction co-efficient, from an amino acid sequence.

4. Methods

This section provides a general description of the experimental procedures used in the study. Case-specific modifications are reported wherever necessary.

4.1. Preparation of chemically competent *Eschericia coli* cells

Competence of bacteria, defined as the ability to incorporate exogenous genetic material from the environment, can be induced by causing the bacterial cell membrane to become transiently permeable through exposure either to suitable chemicals, most commonly calcium chloride, or to an electric field. In this study, high-efficiency chemically competent *E. coli* cells were prepared by the Inoue method (Inoue, Nojima *et al*. 1990) and used for cloning and expression studies. Briefly, cells were grown on an LB agar plate at 37°C overnight. A single colony was picked up and used to inoculate 5 ml LB medium containing the appropriate antibiotic(s) and grown overnight at 37°C with shaking. The overnight pre-culture was used to inoculate 250 ml LB medium and grown at 18°C with vigorous shaking until the optical density (OD) at 600 nm reached 0.4. This process took between 24 and 36 h, depending on the cell line and its antibiotic resistance. Once the desired OD was reached, the flasks were transferred to the cold room and kept on ice for 10 min. The cells were harvested by centrifugation at 4000 *g* for 10 min at 4°C. The pellets were gently resuspended in 80 ml ice-cold transformation buffer (TB) and stored on ice for 10 min. The cells were harvested at 4000 *g* for 10 min at 4°C. All further steps were carried out under sterile conditions. The cells were again resuspended in 20 ml ice-cold TB solution and 1.4 ml dimethyl sulfoxide (DMSO), and the suspension was placed on ice for 10 min. Aliquots of 100 µl were dispensed into pre-chilled microcentrifuge tubes, frozen instantaneously in liquid nitrogen and stored at -80°C. The prepared cells were tested for their transformation efficiency using a plasmid DNA of known concentration, and the efficiency was determined by applying the following formula:

Transformation Efficiency (Transformants/µg) = no. of colonies on plate / ng of DNA plated X 1000 ng/µg

4.2. Sequence and ligation independent cloning

Multiple expression constructs encoding different lengths of the PNK and CPDase domains of yeast tRNA ligase, and lancelet PNK/CPDase were assembled by

sequence and ligation independent cloning (SLIC), which involves a single-step assembly of DNA fragments applying the principles of *in vitro* homologous recombination and single strand annealing (Li and Elledge 2007). SLIC mimics *in vivo* homologous recombination by generating single-stranded DNA overhangs both in the insert and in the vector sequences with the use of the 3'→5' exonuclease activity of T4 DNA polymerase. The length of the generated single-stranded DNA fragment depends on the duration of the T4 DNA polymerase treatment. The single-stranded overhangs are annealed *in vitro,* and the reaction is used to transform *E. coli* cells, where the non-homologous segments at the ends of the annealed DNA molecules are excised, and a recombinant plasmid DNA is generated. The following sub-sections detail the steps followed in this study to clone target DNA fragments into the vector of choice by using the SLIC method.

4.2.1. Choice of constructs

As shown in Table 1, a total of nine constructs, including eight for the PNK/CPDase domains of yeast Trl1 and one construct for the full-length lancelet PNK/CPDase were selected, after consideration of properties such as the location of PNK and CPDase domains and the predicted secondary structural features, flexible regions and intrinsically disordered segments. The output of ProteinCCD contained a summary of all the results from different prediction programs and was used for the selection of multiple constructs.

Table 1. Constructs prepared by SLIC method

Name of the construct	**Residues in yeast Trl1**
*Sc*Trl1 PNK/CPDase	389 - 827
*Sc*Trl1 PNK	393 - 562
*Sc*Trl1 (N+) PNK	389 - 562
*Sc*Trl1 (N+) PNK (C+)	389 - 582
*Sc*Trl1 PNK (C+)	393 - 582
*Sc*Trl1 CPDase	566 - 827
*Sc*Trl1 (N+) CPDase	564 - 827
*Sc*Trl1 (N-) CPDase	590 - 827
	Residues in lancelet PNK/CPDase
Bf PNK/CPDase	1 - 520

4.2.2. Design of primers

Primers were designed [Table 2] and properties, such as the G+C content and melting temperatures, were calculated using ProteinCCD.

Table 2. Primers used for PCR amplification of target DNA fragments

Construct	Primers designed
*Sc*Trl1 PNK/CPDase	**Fwd:** 5'-cagggacccggtCGCACTAAATTTTTAATCTT-3' **Rev:** 5'-cgaggagaagcccggttaAAAATTTAAATATACACTGC-3'
*Sc*Trl1 PNK	**Fwd:** 5'-cagggacccggtTTAATCTTTCCAATATCAGT-3' **Rev**: 5'-cgaggagaagcccggttaTCCAAAATCTAACTCTATCA-3'
*Sc*Trl1 (N+) PNK	**Fwd**: 5'-cagggacccggtCGCACTAAATTTTTAATCTT-3' **Rev:** 5'-cgaggagaagcccggttaTCCAAAATCTAACTCTATCA-3'
*Sc*Trl1 (N+) PNK (C+)	**Fwd:** 5'-cagggacccggtCGCACTAAATTTTTAATCTT-3' **Rev:** 5'-cgaggagaagcccggttaATAAGCTTTATGAATTTCAT-3'
*Sc*Trl1 PNK (C+)	**Fwd:** 5'-cagggacccggtTTAATCTTTCCAATATCAGT-3' **Rev:** 5'-cgaggagaagcccggttaATAAGCTTTATGAATTTCAT-3'
*Sc*Trl1 CPDase	**Fwd:** 5'-cagggacccggtTCTTCATTAACCAATGCGAA-3' **Rev:** 5'-cgaggagaagcccggttaAAAATTTAAATATACACTGC-3'
*Sc*Trl1 (N+) CPDase	**Fwd:** 5'-cagggacccggtGCAGACTCTTCATTAACCAA-3' **Rev:** 5'-cgaggagaagcccggttaAAAATTTAAATATACACTGC-3'
*Sc*Trl1 (N-) CPDase	**Fwd:** 5'-cagggacccggtCCGAAAGATGATGAAATTGA-3' **Rev:** 5'-cgaggagaagcccggttaAAAATTTAAATATACACTGC-3'
Bf PNK/CPDase	**Fwd:** 5'-cagggacccggtATGCCAGGAAGGAAGAAAAA-3' **Rev:** 5'-cgaggagaagcccggttaATAGAAGCCTGAAAACATTG-3'

The primers were synthesized and delivered by Eurofins MWG GmbH, Ebersberg (Germany) in lyophilized form with high purity salt free (HPSF)-purified quality. The primers were dissolved in ddH_2O to a final concentration of 100 pmol/µl and frozen at -20°C until needed.

The pETNKI_his3C_LIC_amp vector, kindly provided by the Netherlands Cancer Institute (NKI), was used for SLIC-based sub-cloning. As indicated by its name, the vector includes an N-terminal hexa-histidine tag and a 3C protease-cleavable site. The 3C protease cleavage site enables removal of the affinity tag following initial purification of the protein from a soluble cell lysate. The vector also contains an ampicillin resistance marker.

4.2.3. Amplification of target genes

The DNA fragments to be cloned were amplified *in vitro* by PCR (Mullis, Faloona *et al*. 1986) using appropriate sets of primers. The recombinant plasmids, pET28a that harbours PNK/CPDase domains of yeast Trl1 and pET20b that contains the lancelet PNK/CPDase domains, were used as templates.

Following is the mixture of reagents used for each PCR reaction:

Template plasmid	0.1 µg
Forward insert primer	0.5 µM
Reverse insert primer	0.5 µM

35 µl of Phusion^{R} High-Fidelity Mastermix was added to the above and the reaction volume was adjusted to 75 µl. The Phusion^{R} High-Fidelity Mastermix is a mixture of 0.04 U/µl Phusion^{R} DNA polymerase, 200 µM of each deoxynucleotide and an optimized reaction buffer. Following are the conditions used in the thermal cycle:

Initial denaturation	98°C for 2 min	
Annealing	98°C for 20 s 60°C for 30 s 72°C for 2 min	X 30 cycles
Extension	72°C for 5 min	

4.2.4. Plasmid purification

The template plasmids and the cloning vector were purified from overnight bacterial cultures by the alkaline lysis method using the QIAprep spin mini-prep kit, according to the instructions of the manufacturer.

4.2.5. Agarose gel electrophoresis

The size and purity of the amplified DNA fragments or purified plasmid DNA molecules were analyzed by agarose gel electrophoresis (Sharp, Sugden *et al*. 1973). This method uses electric current to separate charged molecules on a porous matrix called agarose. Agarose is a polysaccharide derivative of agar and is highly purified to remove impurities and charge. When electric current is applied, the strongly negatively-charged DNA molecules migrate towards the positive pole of the electrophoresis chamber, and the speed of migration depends on the size of the molecule, as smaller fragments move faster than larger molecules. The addition of a fluorescent dye, ethidium bromide (EtBr), enables visualization of the DNA bands under a short-wavelength UV light source (transilluminator). 1.2 % agarose was used for optimal resolution. To detail, the agarose was first dissolved in TAE buffer by boiling, and cooled down prior to the addition of 12 µl of ethidium bromide. The solution was poured in a casting tray that contained a sample comb, and left in the tray to solidify. A few minutes later, the gel was transferred to a chamber filled with 1X TAE buffer and was immersed in the buffer. DNA samples mixed with 6X gel loading

buffer were loaded into the slots on the gel. Electrophoresis was carried out at 100 V. The gel was visualized under an UV transilluminator and photographed using a gel documentation system.

4.2.6. Gel extraction of DNA fragments

The PCR-amplified DNA fragments and purified plasmid DNA molecules were electrophoresed and subsequently extracted using the QIAquick gel extraction kit. The gel pieces containing the DNA of interest were excised from the low-melt agarose gel using a clean and sharp scalpel. The excised gel fragments were processed according to the instructions provided in the kit.

4.2.7. Linearization of the vector by KpnI digestion

0.5 µg of the vector (pETNKI_his3C_LIC_amp) was linearized by restriction digestion with the KpnI enzyme using enzyme concentration, incubation temperature and duration as prescribed by the manufacturer.

4.2.8. T4 DNA polymerase treatment of insert and vector

0.05 µg of both the amplified insert and linearized vector were treated separately with T4 DNA polymerase using the enzyme concentration as prescribed by the manufacturer. 1 µl of 100 mM DTT and 2 µl of 2 M urea were added to the reaction mixture, followed by incubation at 23°C for 20 min. The reaction was ceased by the addition of 1 µl of 500 mM EDTA. T4 DNA polymerase was inactivated by heating the reaction at 75°C for 20 min.

4.2.9. Annealing

For the annealing reaction, the T4 DNA polymerase-treated insert and vector were mixed in a 5:1 molar ratio and incubated at 65°C for 10 min. The reaction was slowly cooled down to 25°C in a PCR machine to improve the efficiency of the annealing reaction.

4.2.10. Transformation of *E. coli* cells

1 µl of the annealing product was added to 10 µl of NEB5α cells and kept on ice for 30 min. For transformation of expression plasmids, a frozen aliquot of 100 µl of competent *E. coli* cells was thawn on ice, and 2 µl of the recombinant plasmid were added to the cells. The cells were exposed to a heat-shock at 42°C for 45 s and immediately placed on ice for 5 min. 500 µl of LB medium were added to the cells, and the mixture was incubated at 37°C for 45 min with shaking. The cells were plated

out on an LB agar plate supplemented with appropriate antibiotic(s) and incubated overnight at 37°C.

4.2.11. Colony PCR

Single colonies were used as DNA templates for the identification of positive transformants by colony PCR, using the reaction mixture described in 4.2.3. The amplified products of colony PCR were analyzed by agarose gel electrophoresis, and the bacterial cultures corresponding to the positive clones were used for the isolation of plasmids.

4.2.12. Plasmid sequencing

The purified plasmids were sent to Eurofins MWG GmbH (Ebersberg, Germany) for DNA sequencing. The standard pET forward and pET reverse primers were used in sequencing. The sequences were compared to the original gene sequences available on the NCBI database by using Blastn. The DNA sequences were translated to protein sequences by the use of the ExPASy translate tool, and the comparison of protein sequences to the database was performed by using Blastp. Only the plasmids that showed 100% sequence identity to both reference sequences were considered for further experiments.

4.3. Recombinant expression, and purification of proteins

4.3.1. *Eschericia coli* strains used for expression screening

The following *E. coli* strains were used to test the heterologous overexpression of newly-prepared constructs.

BL21 (DE3)

Genotype: fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5

Properties: BL21 (DE3) is an *E. coli* B strain that carries the λ DE3 lysogen. It is deficient in proteases lon and ompT and minimizes proteasomal degradation. This strain is widely used for high-level expression of recombinant proteins that are not toxic to *E. coli.*

BL21 CodonPlus (DE3) RIPL

Genetype: F–ompT hsdS(rB– mB–) dcm+ Tetr gal λ(DE3) endA Hte [argU proL Camr] [argU ileY leuW Strep/Specr]

Properties: This strain of *E. coli* is engineered to include extra copies of the tRNA genes, argU, ileY, leuW and ProL. These genes encode tRNAs that often limit the translation of heterologous proteins in *E. coli*. The addition of such tRNAs improves the expression levels of heterologous proteins that are otherwise poorly expressed in conventional BL21 strains.

Rosetta (DE3)

Genotype: F^-ompT $hsdS_B(r_B^- m_B^-)$ gal dcm(DE3) pRARE (Cam^R)

Properties: Rosetta (DE3) is a derivative of the conventional BL21 strain and is specifically constructed to improve the expression of eukaryotic proteins that contain rare *E. coli* codons. This strain expresses six rare tRNAs for the codons AUA, AGG, AGA, CUA, CCC and GGA.

4.3.2. Optimization of growth conditions

The expression plasmids were initially transformed into all the three above-mentioned *E. coli* strains. A single colony from each construct was inoculated into 1 ml LB medium supplemented with respective antibiotics and grown overnight at 37°C. In the following steps, in order to monitor the levels of overexpression and the solubility of the expressed constructs, a range of different growth conditions were applied to smaller volumes of *E. coli* cells (typically, 10 ml). Firstly, equal volumes of the two different growth media, the LB medium (Bertani 1951) and the AI medium (Studier 2005) were chosen for screening. 2% of the overnight pre-culture was used to inoculate 10 ml of both LB medium and AI medium separately.

Since the cells grown in LB medium produce proteins through an IPTG-inducible expression system, another variant, the concentration of the inducer, IPTG, was introduced. The cells in LB medium were cultured at 37°C until the OD at 600 nm reached 0.6. Then, each culture was divided into four sub-cultures and induced with four different concentrations of the inducer IPTG (0.15, 0.25, 0.5 and 1 mM). The trials were continued by growing the sub-cultures at 18°C for 15 h post-induction. The cells were harvested by centrifugation. The pellets were lysed as described in 4.3.3, and the resulting yields of soluble protein were compared to find out the optimal IPTG concentration. In the second round of expression screening, variations in post-induction growth temperature and duration were applied. One of the three induced cultures was grown at 37°C for 6 h and harvested. The other two cultures were grown at 20°C or 15°C for 15 h prior to harvesting. For auto-induced cultures, the incubation temperature and duration were altered. The cultures were grown either at 25°C for 36 h or at 17°C for 48 h, and harvested by centrifugation. The harvested cell pellets were flash-frozen and stored at -80°C.

4.3.3. Optimization of cell lysis

Frozen bacterial pellets were suspended in selected lysis buffers with varying pH. The buffers used at different pH ranges were MES (pH 5.5 and 6.0), sodium phosphate (pH 6.0 and 6.5), HEPES (pH 7.0 and 7.5) and Tris (pH 8.0 and 8.5). Other components used in the buffer were 20 µg/ml DNase, 20 µg/ml RNase A, 5 mM $MgCl_2$, 2 mM TCEP, 10 mM imidazole, 200 µg/ml lysozyme, and a Complete mini EDTA-free protease inhibitor cocktail tablet. The suspension was incubated for 20 min at 4°C on a tube-rotator. Two different modes of cell disruption, namely sonication and freeze-thawing, were tested. Using sonication, the cells were lysed by five 3-min pulse cycles (0.5 s on + 0.5 s off) in a Bandelin Sonopuls Sonifier equipped with an MS72 probe operating at 30% power. For freeze-thawing, the cell suspensions were immersed in liquid N_2 for 40 s, and the frozen cells were transferred immediately to a water bath at 30°C for thawing. Following lysis, the cell debris was removed by centrifugation at 35000 *g* for 30 min at 4°C. Samples were taken from both soluble and insoluble fractions of the cell lysate and analyzed on SDS PAGE.

4.3.4. Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SDS PAGE is used for the electrophoretic separation of proteins (Laemmli 1970, Matsudaira and Burgess 1978). This method is based on the principle that a charged molecule in an electric field migrates towards the electrode with an opposite charge. Since the mobility of proteins on the gel depends on both charge and size, natively folded proteins are treated with SDS to acquire uniform charge, so that electrophoretic mobility depends primarily on size, and the molecular weight can be determined. SDS binds proteins at a ratio of one SDS molecule per two amino acid residues; proteins become denatured, and negatively charged in proportion to their mass. The SDS-treated proteins migrate towards the positive electrode, and are separated based on their size, applying the effect of molecular sieving. The protein bands can be visualized by staining with Coomassie brilliant blue, and the sizes can be determined by the distance of migration compared with that of known molecular weight standards.

In this study, protein samples were analyzed by SDS-PAGE using mini-PROTEAN^R precast gels and a Bio-Rad electrophoresis system. The proteins were mixed with SDS sample buffer, heated at 95°C for 5 min and electrophoresed at a constant voltage (150-180 V). The gels were stained with Coomassie blue stain for 20 min on an orbital shaking platform and placed in the destaining solution for 20 min before visualization of the bands.

In this study, protein samples were analyzed by SDS-PAGE using mini-PROTEAN^{R} precast gels and a Bio-Rad electrophoresis system. The proteins were mixed with SDS sample buffer, heated at 95°C for 5 min and electrophoresed at a constant voltage (150-180 V). The gels were stained with Coomassie blue stain for 20 min on an orbital shaking platform and placed in the destaining solution for 20 min before visualization of the bands.

4.3.5. Quantification of proteins

Protein concentration was determined by measuring the light absorption at 280 nm using a Nanodrop spectrophotometer and by applying the working principle of the Beer-Lambert law (Beer, 1852):

$$c = A / (\varepsilon.l)$$

Where, c is the protein molar concentration ($mol.l^{-1}$)

l is the pathlength of light (cm)

A is the absorption

ε is the molar extinction co-efficient ($l.mol^{-1}.cm^{-1}$), which is the sum of the extinction co-efficients of the absorbing species (Tyr, Trp and disulphides)

4.3.6. Confirmation of identity by mass spectrometry

The identity of the purified proteins was verified by tryptic peptide mapping using mass spectrometry. This method determines the masses of biomolecules by measuring the mass-to-charge (m/z) ratios of the respective ions (Fenn *et al*. 1989). In this study, the mass spectrometric analyses of samples were carried out at the Biocenter Oulu Proteomics Core Facility, Department of Biochemistry, University of Oulu (Oulu, Finland).

4.3.7. Large-scale expression

Optimal conditions, identified by small-scale expression trials, were applied for large-scale soluble expression of target proteins in Rosetta (DE3) cells. An overnight primary culture, grown from a single colony of cells containing the expression construct, was used to inoculate 1 l of LB medium supplemented with appropriate antibiotic(s) in a baffled 3-litre Erlenmeyer flask. Similarly, 4 l of LB media were inoculated with overnight pre-cultures. The cultures were incubated at 37°C, with shaking, until an OD_{600} of 0.8 was reached. The cultures were cooled down to 18°C prior to addition of IPTG to a final concentration of 0.15 mM. The incubation was

continued at 18°C for another 16-18 h. The cells were harvested by centrifugation, and the cell pellets were frozen at -70°C until further use. Cell pellets were thawn and resuspended in lysis buffer containing 50 mM HEPES, pH 7.5, 200 mM NaCl and other components as mentioned in 4.3.3. The cells were disrupted by sonication and clarified by centrifugation at 35000 *g* for 30 min at 4°C. The supernatant containing the protein of interest was used for purification.

4.3.8. Affinity purification of hexa-histidine tagged proteins

Immobilized-metal affinity chromatography (IMAC) is a method of protein purification and is based on the ability of histidine and cysteine residues to form stable interactions with divalent metal cations such as Ni^{2+}, $Co^{2+,}$ Zn^{2+} and Cu^{2+} (Porath, Carlsson *et al*. 1975). Polyhistidine-tagged proteins, expressed recombinantly, can be purified from other non-tagged proteins present in the soluble cell lysate by passing the supernatant through a solid matrix loaded with a divalent cation. In this study, Ni^{2+} immobilized to nitrilotriacetic acid (NTA) on agarose beads was used. The protein bound to the matrix can be eluted competitively by the use of high concentrations of imidazole. As all of the expressed constructs in this study contained an N-terminal hexa-histidine tag, Ni-NTA chromatography was effectively used as the initial purification step. Either a HisTrapFF (1 ml) column pre-packed with Ni-NTA matrix or a gravity-flow column loaded with 1 ml of Ni-NTA agarose was first washed with 5 - 10 column volumes (CV) of filtered and degassed ddH_2O to remove ethanol in which the matrix was stored. The matrix was equilibrated with 5 CV of buffer B containing 10 mM imidazole, which was found to prevent most of the non-specific interactions of non-tagged proteins with the matrix. The supernatant of the cell lysate was applied onto the HisTrapFF column with the use of a pump that allows the liquid to flow in and out of the column at a constant flow rate; in the case of gravity-flow column, the effect of gravitational force was utilized to load the sample onto the matrix. The protein was flown through the HisTrapFF column at a fixed flow rate of 0.5 ml/min; the gravity-flow tubes were first horizontally rotated for approximately 1 h to ensure binding of the protein to the matrix. The matrix was then washed with buffer C that contained 50 mM imidazole to wash off any impurities. The bound protein was eluted with buffer E with 500 mM imidazole. The protein was eluted in 1.5 ml microcentrifuge tubes as 1 ml fractions in order to monitor the extent of purity at different stages of elution. The eluted fractions were resolved by SDS PAGE, and the fractions that contained the protein of the expected size were pooled. The protein was dialysed against buffer D without imidazole overnight at 4°C.

4.3.9. Cleavage of the hexa-histidine tag

As the expression constructs contain an N-terminal 3C protease cleavage site between the polyhistidine tag and the target protein, recombinantly expressed His-tagged 3C protease was used to cleave the tag off the target protein. The fusion proteins were incubated with different concentrations of 3C protease at room temperature and at 4°C. Samples were taken at different time intervals and analyzed by SDS PAGE to monitor the progress of the reaction. Thus, the concentration of 3C protease, temperature and duration of the enzyme treatment were optimized. The cleavage mixture was passed through the Ni-NTA matrix; the cleaved hexa-histidine tag, His-tagged 3C protease and any uncleaved His-tagged target protein were bound to the matrix, whilst the cleaved target protein remained in the flow-through.

4.3.10. Size exclusion chromatography

The protein was concentrated to 500 µl (100 µl for analytical gel filtration) by using Vivaspin 20 concentrators and centrifugation at 3220 *g* (4°C). PNK/CPDase and CPDase proteins were concentrated using concentrators with a molecular weight cut-off (MWCO) of 30 kDa and 10 kDa, respectively. The concentrated protein was further purified by size exclusion chromatography (SEC), which was used as the final step of purification and also to assess the oligomeric state of the protein purified. 500 µl of the concentrated protein were applied onto a HiLoad 16/60 Superdex 200 preparative grade column, pre-equilibrated with the SEC buffer, using an ÄKTApurifier or an ÄKTAexplorer (GE Healthcare Life Sciences). A constant flow rate of 1 ml/min was applied at equilibration, sample injection and fractionation. SEC separates molecules based on their size or hydrodynamic volume (Porath and Flodin 1959). Absorbance at 280 and 260 nm was recorded in order to monitor the elution of the protein and the presence of any nucleic acid impurities. The gel filtered protein was collected as 1 ml fractions. Based on the shape of the peak and comparison of the elution volume to that of the calibration standards, the oligomeric state of the proteins was estimated. The collected fractions were resolved by SDS PAGE.

4.4. Analysis of binding of RNA from expression host

Since the absorbance of the proteins eluted from Ni-NTA matrix peaked at 260 nm, which is characteristic of nucleic acids, non-specific binding of nucleic acids from the expression host, *E. coli*, was suspected. In order to detect the type of nucleic acid co-purified, the samples were processed further using the following methods.

4.4.1. Size-exclusion chromatography

The proteins, purified by Ni-NTA affinity chromatography, were subjected to size exclusion chromatography. The fractions that eluted around the void volume were pooled, concentrated and their absorbance was measured. The samples were analyzed by SDS-PAGE and agarose gel electrophoresis to check the presence of any protein-nucleic acid complex.

4.4.2. Nuclease treatment and agarose gel electrophoresis

The concentrated gel-filtered protein was treated with 0.5 mg/ml DNase A or 0.5 mg/ml RNase A for 20 min at room temperature. Following the treatment with nucleases, the samples were analyzed by agarose gel electrophoresis with EtBr staining to determine the nature of the nucleic acid found in PNK/CPDase preparations.

4.5. Analysis of folding by synchrotron radiation circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy (Adler, Greenfield *et al*. 1973) was used to analyze the folding state of the purified proteins. This method applies the principle of CD, defined as the differential absorption of left and right circularly polarized light by optically active or chiral molecules, to probe the secondary structural composition of peptides and proteins in solution. The phenomenon of CD occurs as the chromophores in a chiral environment interact with polarized light. The amide bonds of the peptide backbone and the aromatic side chains are the optically most active groups present in a protein molecule. When the amides of the peptide backbone are present in highly ordered arrays, such as α-helices and β-strands, the optical transitions of the amide bond are split into multiple transitions. As a result of such transitions, different structural contents exhibit characteristic CD spectra.

4.5.1. Preparation of samples

Proteins purified by size exclusion chromatography were used for CD sample preparation. One of the crucial factors was to eliminate any material, except the protein, that is optically active and can interfere with the signal. The proteins purified in 50 mM HEPES, pH 7.5 and 200 mM NaCl were dialyzed into CD buffer (50 mM potassium phosphate, pH 6.5). For SRCD measurements, the range of sample concentrations used was between 3.5 and 5 mg/ml.

4.5.2. SRCD measurements

SRCD measurements were carried out on UV-CD12 beamline at the Angstromquelle Karlsruhe (ANKA) synchrotron facility (Karlsruhe, Germany). Three spectra were collected at a scan rate of 14 nm/min, in a demountable 13 µm CaF_2 cuvette at a wavelength range between 260 and 170 nm at 1 nm intervals. A spectrum from the corresponding buffer was measured for all the samples and used for background correction.

4.5.3. Analysis of SRCD spectra

The spectra were averaged and the spectrum measured from the corresponding buffer was subtracted. The CD units (mdeg) were converted to Δε (M^{-1} cm^{-1}). The data were analysed and processed with CDtool (Lees, Smith *et al*. 2004). The spectra were deconvoluted at the Dichroweb server (Lobley, Whitmore *et al*. 2002) using the CDSSTR algorithm (Compton and Johnson 1986) and the SP175 reference dataset (Lees, Miles *et al*. 2006).

4.6. Polynucleotide kinase activity assay

PNK catalyses the transfer of the γ phosphate from ATP to the 5'-OH group of polynucleotides in a reversible reaction. In this study, the PNK activity of the N-terminal PNK domain of yeast Trl1 and lancelet PNK/CPDase was assayed using the synthetic oligonucleotides, ribo A_{20} (rA_{20}), deoxy A_{20} (dA_{20}), a mixture of r+dA_{20}, and [γ^{32}P] ATP. The assay was performed by Mr. Arne Raasakka at the radio-isotope laboratory of the Department of Biochemistry, University of Oulu (Oulu, Finland). The enzymes, *Bf* PNK/CPDase, *Sc*Trl1 PNK/CPDase and *Sc*Trl1 CPDase were purified, diluted to 1mg/ml and sent to Oulu on ice. T4 PNK and *Sc*Trl1 CPDase were used as positive and negative controls, respectively. The reaction mixture contained the PNK buffer (500 mM Tris, pH 7.5, 100 mM $MgCl_2$ and 50 mM DTT), 2 µg of the enzyme, and 100 pmol rA_{20} or dA_{20} or both. 55 pmol of active radiolabeled ATP were added to each reaction mixture, and the volume of the mixtures was adjusted to 40 µl with diethylpyrocarbonate (DEPC)-treated water. The samples were incubated at 37°C, and 8 µl aliquots were removed at different time points (1, 15, 30 and 60 min). Each sample was quenched with 8 µl of 2X urea sample loading buffer (Invitrogen Novex), heated at 90°C for 4 min and stored on ice until electrophoresis. The samples were electrophoresed, using a 15% Mini-PROTEAN^R TBE-Urea gel (Bio-Rad), at 150 V for 85 min, and visualized by PhosphorImager analysis.

4.7. Analysis of nucleotide binding

The N-terminal domain of *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase contains an NTP-binding P-loop and catalyzes phosphorylation of tRNA halves in an NTP-dependent manner (Wang, Schwer *et al*. 2006, Englert, Sheppard *et al*. 2010). The nucleotide-binding properties of the PNK domain were analyzed using N-methylanthranilolyl (mant) labeled fluorescent analogues of AMP, ATP and GTP (mant-AMP, mant-ATP, and mant-GTP). Fluorescence resonance energy transfer (FRET) resulting from tryptophan fluorescence excitation (at 280 nm) of mant fluorescence indicates that the donor fluorophore (Trp) and the acceptor fluorophore (mant) are in close proximity.

4.7.1. Preparation of samples

The samples were prepared as described previously (Wang, Neugebauer *et al*. 2011). The nucleotide analogues, mant-ATP, mant-GTP and mant-AMP were prepared at three different concentrations (5, 10 and 25 µM) in 10 mM HEPES, pH 7.5. Enzymes at a concentration of 0.5 µM were mixed with the nucleotides, and the volume was made up to 200 µl in a 96-well black flat bottom plate.

4.7.2. Measurements

Fluorescence measurements were carried out using the Tecan Infinite M200 fluorescence spectrometer. Excitation was carried out at 280 nm and emission was collected between 310 and 500 nm, with the peak of mant-fluorescence expected at 440 nm.

4.7.3. Analysis of results

To analyze the results, the fluorescence from the mant-labeled nucleotide analogues was collected separately and subtracted from the signal obtained from the protein-nucleotide complex.

4.8. Analysis of conformational changes upon ligand binding

SRCD spectroscopy was applied to detect changes in conformation of yeast and lancelet PNK/CPDase enzymes upon binding of ligands, such as GTP and 2',3'-cNADP$^+$. The ligands at two different concentrations (500 µM and 1 mM) were prepared in the buffer containing 50 mM potassium phosphate, pH 6.5 and 1 mM $MgCl_2$. The proteins, at a concentration of 10 µM, were mixed with different concentrations of each ligand. SRCD measurements were carried out as described in 4.5.2.

4.9. Cyclic phosphodiesterase activity assay

CPDase catalyses the irreversible hydrolysis of 3'-phosphodiester bonds in 2',3'-cyclic nucleotides to produce 2'-nucleotides. In the assay, CPDase hydrolyses the phosphodiester bond in β-nicotinamide adenine dinucleotide 2′,3′-cyclic monophosphate ($cNADP^+$), and the resulting nicotinamide adenine dinucleotide phosphate ($NADP^+$) is reduced to NADPH by glucose-6-phosphate dehydrogenase and used to transform glucose-6-phosphate to 6-phosphoglucanolactone. The quantity of NADPH formed during the coupled enzymatic reaction is a direct measure of the activity of CPDase (Sogin 1976). Since CPDase also catalyses the conversion of 2',3'-cyclic CMP to 2'-CMP (Hugli, Bustin *et al*. 1973), 2',3'-cCMP was used as another substrate to measure the activity of CPDase.

4.9.1. Preparation of samples

*Sc*Trl1 PNK/CPDase, *Bf* PNK/CPDase and *Sc*Trl1 CPDase were used in the activity assay. The assay mixture, as described previously (Lee, Gravel *et al*. 2001, Myllykoski and Kursula 2010) contained 100 mM MES, pH 6.0, 3 mM $MgCl_2$, 5 mM glucose-6-phosphate and 0.6 U glucose-6-phosphate dehydrogenase. Initially, the assay was performed by mixing 0.5 mM of the substrate with different amounts of each of the constructs (0, 100, 250, 500, 1000, 5000 ng). Once the suitable amount of enzyme was determined, the assay was repeated in triplicates by using 500 ng of each of the enzymes with varying concentrations of the substrate, $cNADP^+$ (0, 0.05, 0.1, 0.2, 0.5 and 1.0 mM). The same concentrations (same as those used for $cNADP^+$) of the substrate 2',3'-cCMP were prepared in 50 mM MES, pH 6.0 and mixed with 500 ng of each enzyme.

4.9.2. Measurements

Kinetic measurements were performed using a 96-well flat-bottom transparent plate at 25°C using the TECAN infinite M200 fluorescence spectrophotometer and the iControl software. The activity was measured for 1 h, and the samples were shaken linearly for 1 s before each of the 120 kinetic cycles performed. The amount of produced NADPH formed was measured spectrophotometrically at 340 nm using the absorption co-efficient 6.22 $cm^{-1}mM^{-1}$. The formation of 2'-CMP was measured by the increase in absorbance at 286 nm.

4.9.3. Analysis of results

The kinetic parameters K_m and V_{max} were determined by fitting the mean velocities versus substrate concentration to the Michaelis-Menten equation of enzyme kinetics

(Michaelis, Menten *et al*. 2011) using non-linear regression analysis in Prism 5 (GraphPad Software™).

4.10. High-throughput thermal stability analysis

The fluorescence-based thermal stability assay monitors the effects of buffer conditions on thermally-induced protein unfolding (Pantoliano, Petrella *et al*. 2001, Ericsson, Hallberg *et al*. 2006). Since properties, such as stability, solubility and conformational homogeneity of proteins, are crucial requirements for structural studies, it is important to identify stabilizing buffers and ligands. This high-throughput assay uses a screen of different buffer formulations and additives, and determines the thermal melting points of the protein in the conditions applied and allows the selection of the most suitable buffer and additives (Ericsson, Hallberg *et al*. 2006). The method involves the binding of a hydrophobic fluoroprobe to the exposed hydrophobic core of an unfolding protein resulting in an increase in the fluorescence emission. The increase in fluorescence is measured as a function of temperature. Since the assay can be performed in high-throughput mode, several variants, such as pH, salt type and concentration, and the presence of ligands, can all be screened at the same time.

4.10.1. Preparation of samples

A series of different buffer conditions were chosen and prepared in a 96-well deep well block. Each well in a 96-well thin-wall PCR plate was filled with 20 µl of a buffer solution. The working solution of the fluoroprobe, SYPRO Orange, was prepared by adding 5 µl of the 5000X gel stain into 1000 µl of ddH_2O and by vigorous shaking on a vortex. The proteins were prepared at a concentration of 1 mg/ml. Based on the number of wells to be used, the required amount of protein and dye (2.5 µl each per well) were estimated and used to prepare the [protein+dye] mixture. 5 µl of the [protein+dye] mixture were added to each well, and the solution was mixed by gently pipetting it up and down.

4.10.2. Measurements

The plate was sealed with optical-quality adhesive film (Bio-Rad) and placed inside the CFX96 Real-time PCR system (BioRad). The plate was heated from 25°C to 99°C at 0.5°C increments. The wavelengths for excitation and emission were 490 and 575 nm, respectively.

4.10.3. Analysis of results

The results were analyzed using the data analysis program of Bio-Rad CFX Manager. The melting points calculated from each condition were compared, which can be used for the identification of the stabilizing buffers and additives for each construct.

4.11. Analysis of solution shape and oligomeric state using small-angle X-ray scattering

Small-angle X-ray scattering (SAXS) has emerged as a powerful tool to determine the low-resolution three-dimensional structures of biological macromolecules in terms of averaged particle size and shape. SAXS is based on the principle of elastic scattering of X-ray photons by particles in an illuminated volume (Koch, Vachette *et al*. 2003). Basically, a monochromatic focused beam of X-rays hits the solution containing the molecule of interest, which is a protein in this study, and the scattered intensity [I(s)] is recorded as a function of momentum transfer, s = $4\pi sin\theta/\lambda$, where 2θ is the angle between incident and scattered radiation (Koch, Vachette *et al*. 2003). Provided that the scattering is isotropic, the recorded image is radially averaged to obtain the scattering curve. The scattering curve can be used to estimate the global structure and conformation of the protein and also to extract parameters, such as the molecular mass (MM), radius of gyration (R_g), hydrated particle volume (V_P) and maximum intramolecular distance of the particle (D_{max}).

The forward scattering intensity I(0) and R_g can be obtained from the scattering curve by using the Guinier approximation (Guinier1939). The Guinier equation for a monodisperse solution of a globular protein, at very small angles ($s<1.3/R_g$), is represented as:

$$\text{Scattering intensity} = I(s) = I(0) \exp [-(sR_g)^2/3]$$

Non-linearity of the Guinier plot indicates that the quality of the sample is poor. Factors that can affect the linearity of the plot are the presence of any attractive or repulsive inter-molecular interactions and the polydispersity of the sample. The molecular mass of the protein can be determined by using the following formula, which requires I(0) and concentrations of both the protein sample and the standard, bovine serum albumin (BSA) (Mylonas and Svergun 2007).

$$\text{Molecular mass (MM)} = [I(0)_{sample}/ I(0)_{BSA}] \text{ x } MM_{BSA}$$

Independent of the Guinier analysis, the volume of the hydrated particle can be extracted from the small angle region of the scattering data using Porod's equation (Porod 1982):

$$\text{Volume of the particle} = V_P = 2\pi^2 I(0) / \int_0^\infty s^2 \, I(s) \, ds$$

The particle volume (V_P) can also be used to estimate the molecular mass of the protein, and this approach is independent of any possible errors in the concentration. The typical estimate for globular proteins is that the V_P (nm^3) is 1.7 x MM (kDa) (Petoukhov and Svergun 2013).

The limitations of the Guinier approximation prompt the use of indirect Fourier transformation of the scattering intensity (Glatter 1977, Svergun 1992), which introduces another function called the distance distribution function [p(r)]. The distance distribution function represents the scattering data in real space and displays graphically the geometric shape of the particle. The Kratky plot, which represents the scattering intensity weighted by the momentum transfer squared [$s^2I(s)$ *vs* s], is an indicator of the folding state of the protein (Doniach 2001).

4.11.1. Preparation of samples

Similarly to any other method for structural analysis of proteins in solution, the primary requirements of a SAXS sample are purity, monodispersity, and solubility at high concentration. Protein samples used for SAXS measurements were all purified by size exclusion chromatography. The *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase were eluted both as a monomer and as a dimer. The fractions were separately collected, concentrated and used for SAXS studies. Each protein was prepared in three different concentrations and the concentrations were determined using the Nanodrop spectrophotometer immediately prior to the measurements. The samples were kept on ice during and after preparation.

4.11.2. Beamlines used

DORIS X33 BioSAXS beamline

The X33 BioSAXS beamline, at the DORIS-III storage ring (shut down permanently since October 2012) in German Electron Synchrotron (DESY), Hamburg (Germany), was used for SAXS measurements of *Bf* PNK/CPDase. Following are the optical features of the beamline:

Source of synchrotron radiation: DORIS bending magnet

Wavelength: 1.5 Å
Focal spot at the detector: 2 mm x 0.6 mm
Detectors: Photon counting Pilatus 1M-W pixel detector

PETRA-III P12 BioSAXS beamline

The P12 BioSAXS beamline at the PETRA-III storage ring in DESY was used for SAXS measurements of *Sc*Trl1 PNK/CPDase and *Sc*Trl1 CPDase. Following are the important beamline properties:

Source of synchrotron radiation: PETRA undulator U29
Wavelength: 1.24 Å
Focal spot at the detector: 0.2 mm x 0.12 mm
Detectors: 2D photon counting Pilatus 2M pixel X-ray detector
Resolution: 10000 - 8 Å

P12 uses an automated sample changer and an automated data acquisition and processing pipeline that helps users to evaluate the quality of data instantly after the measurements and enables to prepare alternative samples for further measurements.

I911-4 SAXS beamline of MAX-Lab

The I911-4 SAXS beamline at MAX-II, Lund (Sweden) was used for SAXS measurements of *Sc*Trl1 PNK/CPDase, *Sc*Trl1 (N+) CPDase and *Sc*Trl1 (N-) CPDase, and the protein-RNA complexes of *Bf* PNK/CPDase and *Sc*Trl1 PNK/CPDase. The beamline properties are given below:

Source of synchrotron radiation: Multi-pole wiggler
Wavelength: 0.91 Å
Focal spot at the detector: 0.3 mm x 0.3 mm
Detectors: Photon counting Pilatus 1M pixel and Mar 165 CCD

4.11.3. Measurements

Measurements were carried out at the three beamlines described above. Firstly, SAXS from the standard, BSA (5 mg/ml) was measured. The scattering from the corresponding buffer was measured and subtracted from that of BSA. The calculated scattering intensity of BSA was used to evaluate the molecular mass of the protein samples. For each protein sample, three different concentrations were measured (Table 3). At X33 and at P12, the setup was such that the SAXS measurement of each sample was always preceded and succeeded by a SAXS measurement of the corresponding buffer.

At the X33 beamline, the data were recorded at a sample-detector distance of 2.7 m and at the range of momentum transfer 0.1 nm^{-1} < s <5 nm^{-1} at 10°C; eight successive images of 15-second exposures each were acquired and compared to check for possible radiation damage. In the majority of the cases, all the images were usable, and no significant damage was observed.

At the P12 beamline, 20 successive frames of 50 ms exposure each were collected within a second, and the automated software at the beamline selected the images based on their quality, and used them for initial processing. The measurements were carried out at 20°C. The sample-to-detector distance was 3.1 m. The sample volume was 25 µl.

Table 3. Concentrations of different proteins used in SAXS measurements

Beamline	Protein sample	Concentration (mg/ml)
X33	*Bf* PNK/CPDase (monomer)	7.5
		5.0
		3.5
	Bf PNK/CPDase (dimer)	3.0
		1.7
		1.0
P12 and I911-4	*Sc*Trl1 PNK/CPDase (monomer)	7.5
		5.0
		3.5
	*Sc*Trl1 PNK/CPDase (dimer)	5.5
		3.5
		1.5
P12 and I911-4	*Sc*Trl1 CPDase	8.0
		5.5
		2.5
	*Sc*Trl1 (N+) CPDase	5.7
		3.8
		1.3
	*Sc*Trl1 (N-) CPDase	7.5
		4.5
		1.7

At the I911-4 beamline, the measurements were performed in a flow-through capillary setup at 20°C. For each sample, five images were collected in 5 min with an exposure of 60 s each. The data reductions were performed using the beamline software, BLi911-4. The range of momentum transfer was 0.1 nm^{-1}< s <5 nm^{-1}.

4.11.4. Analysis of results

The scattering data were normalized to the intensity of the transmitted beam and radially averaged; the experimental scattering profiles were corrected for the background by the solvent and analyzed using standard procedures and the programs from the ATSAS suite (Petoukhov, Konarev *et al*. 2007). The R_g was determined using Guinier approximation in PRIMUS (Konarev, Volkov et al. 2003) and using the indirect transform package GNOM (Svergun 1992). The molecular mass of the protein was calculated using the I(0) estimates calibrated against BSA (66 kDa). The excluded volume of the hydrated particle was calculated by using the Porod invariant (Porod 1982). GNOM was also used to compute the distance distribution function of the protein [p(r)]. This function was applied to determine the maximum dimension of the protein (D_{max}) in solution. *Ab initio* restoration of low-resolution models was done by the programs DAMMIN (Svergun 1999) and DAMMIF (Franke and Svergun 2009). In total, ten *ab initio* models were created and averaged by DAMAVER (Volkov and Svergun 2003) and the resulting structures were compared in PyMOL (http://www.pymol.org).

4.12. Analysis of oligomeric state using multi-angle static light scattering

Multi-angle static light scattering (MALS) is a non-invasive method used for the determination of absolute molecular mass of macromolecules in solution. It involves exposure of the sample to low intensity laser light of wavelength 690 nm (Debye, 1947). When the laser hits the sample, the intensity of the light scattered by the protein is recorded as a function of the scattering angle and used to calculate parameters such as, molar mass, root mean square radius and the second virial coefficient (A_2). A_2 is a thermodynamic property that describes the strength of interaction between the molecule and the solvent, and thus a good indicator of any aggregation.

4.12.1. Preparation of samples

The proteins were purified by size exclusion chromatography in two different buffers, SEC buffer 1 with and without 10 mM β-ME. The samples were diluted to 1 mg/ml and kept on ice until injection into the column.

4.12.2. Measurements

MALS measurements were carried out in the chromatography mode using a mini-DAWN TREOS multi-angle static light scattering (MALS) detector (Wyatt) and an Optilab rEX differential refractometer (Wyatt) in the flow path of an ÄKTApurifier (GE Healthcare). The system was equilibrated with running buffer (SEC buffer 3 +

5 mM β-ME or SEC buffer 3). For calibration purposes, BSA was used as a standard, since it forms different oligomeric states such as monomer, dimer and tetramer of 66 kDa subunits in solution. 100 µg of BSA were injected into a Superdex 200 10/300GL column attached to an ÄKTA purifier, and the data collection was initiated. 100 µg of target protein were injected into the column and the scattering data collection was carried out. The MALS detector is designed to obtain scattering data from three different angles.

4.12.3. Analysis of results

The data obtained from the MALS detector were analyzed using the ASTRA software - version 5.3.4.11 (Wyatt) according to the manufacturer's instructions. Molecular weight was determined based on the measured light scattering and refractive index or UV absorbance.

4.13. Crystallization trials

Proteins in supersaturated solutions undergo a phase transition and form crystallization nuclei, from which protein crystals can grow. In the sitting drop vapor diffusion method, aqueous protein solution is mixed with a precipitant solution in an airtight container. The drop is positioned as a sitting drop separated from the precipitant reservoir. Due to the difference in the precipitant concentrations in the drop and the reservoir, an equilibrium process sets in, during which the concentration of both protein and precipitant increase, and might reach the nucleation zone, as a result of which protein crystals can form. Since structural characterization of the PNK and CPDase domains was one of the main objectives of the study, the proteins purified by size exclusion chromatography were subjected to crystallization trials using the sitting drop vapor diffusion method. Commercial screens such as JCSG+, Morpheus, and Proplex (Molecular Dimensions, UK) and PEG ion screen (Jena Bioscience, Germany) were used for screening optimal conditions for crystallization. Sitting drops were prepared by mixing 0.5 µl of protein and reservoir solution. The reservoir solution volume was 50 µl. Screens were set up both at room temperature and at 4°C. *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase were used at concentrations ranging from 3 to 8 mg/ml. The concentration range used for *Sc*Trl1 CPDase was between 5 and 18 mg/ml. Screens were also set up following the addition of the substrates, 2',3'-cNADP$^+$ and 2', 3'-cCMP, to the proteins in different stoichiometric ratios (protein: ligand = 1:1 or 1:2).

5. Results and Discussion

5.1. *In silico* analysis of protein sequences

A series of bioinformatics tools were used to predict possible structural and functional properties, such as the elements of secondary structure, presence of intrinsically disordered regions, organization of structural and functional domains, homology to proteins of different families and crystallizability, from the amino acid sequence of both *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase. The results are summarized in the following sub-sections.

5.1.1. Multiple sequence alignment

The sequence alignment of *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase with the homologues, mouse and human CNP proteins, using T-Coffee web server, shows the presence of the conserved P-loop motif [GXXGXGK(T/S)] in the N-terminal domain and the two conserved "H-x-(T/S)-x" motifs in the C-terminal domain [Figure 12]. The alignment indicates that the *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase are divergent members of the 2H phosphoesterase superfamily.

Although the yeast and lancelet tRNA healing enzymes share the conserved motifs at the N-terminal and C-terminal domains with those of their vertebrate homologues, their sequence similarity is very low. Similarity searches using BLAST show that the identity between the sequences of *Bf* PNK/CPDase and mouse CNP is about 34% at a sequence coverage of 15%. Similarly, the identity between the sequences of *Sc*Trl1 PNK/CPDase and mouse CNP is 26% at a sequence coverage of 28%.

The N-terminal domain of the proteins resembles T4 polynucleotide kinase, and the CPDase domain resembles the members of the 2H phosphoesterase superfamily. The residues that constitute the P-loop motif and the two "2H" motifs are given in Table 4. Another functionally important region in the N-terminal domain is the "R-x-x-x-R" motif of yeast Trl1, located approximately 100 amino acids downstream of the P-loop. In addition to the P-loop, the R-x-x-x-R motif is also involved in NTP binding (Wang, Schwer *et al*. 2006). The putative sequence that follows the R-x-x-x-R motif (GNNHQSIK) is conserved among fungal tRNA ligases (Wang, Schwer *et al*. 2006).

Comparison of the sequences identifies the biggest variation between PNK/CPDase proteins and their vertebrate homologues as the long inserts in the yeast and lancelet PNK/CPDase proteins. *Sc*Trl1 PNK/CPDase contains an insertion of 105 residues between PNK and CPDase domains, and *Bf* PNK/CPDase contains an insertion of 96

residues in the CPDase domain [Figure 12]. These insertions are absent in vertebrate CNP proteins [Figure 12].

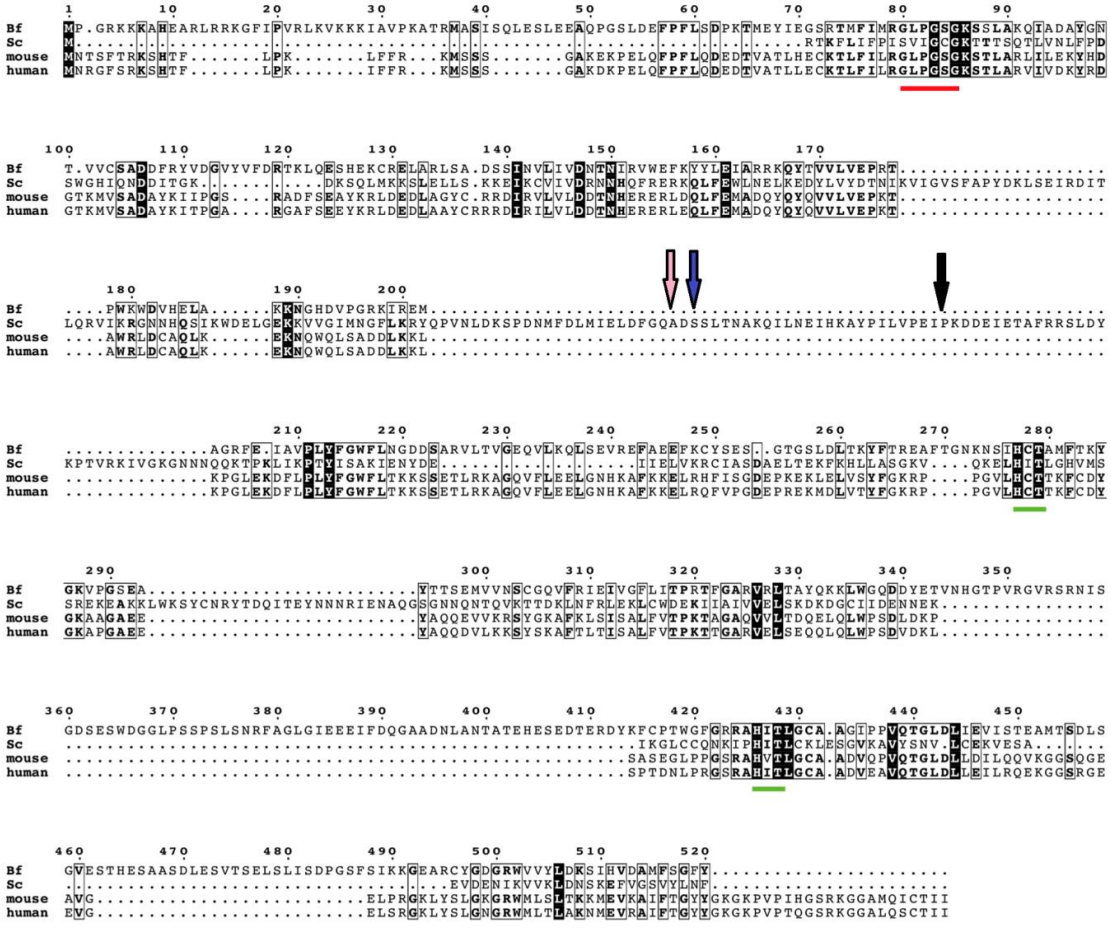


Figure 12. Multiple sequence alignment using T-Coffee. Aligned sequences of *Sc*Trl1 PNK/CPDase (*Sc*), *Bf* PNK/CPDase (*Bf*), mouse CNP (mouse) and human CNP (human). The N-terminal P-loop motif and two C-terminal "H-x-(T/S)-x" motifs are underlined in red and green, respectively. The N-termini of the yeast CPDase constructs, *Sc*Trl1 CPDase, (N+) *Sc*Trl1 CPDase and (N-) *Sc*Trl1 CPDase are denoted by pink, blue and black arrows, respectively. The numbers on the top of the sequences indicate the positions of the residues of *Bf* PNK/CPDase.

Table 4. Positions of residues that form the conserved P-loop and "2H" motifs in the divergent members of 2H phosphoesterase superfamily

Protein	P-loop residues	"H-x-(T/S)-x" motif I	"H-x-(T/S)-x" motif II
Yeast Trl1	^{401}GCGKT^{405}	^{673}HITL^{676}	^{777}HITL^{780}
Lancelet PNK/CPDase	^{80}GLPGSGKS^{87}	^{277}HCTA^{280}	^{426}HITL^{429}
Mouse CNP	^{37}GLPGSGKS^{44}	^{230}HCTT^{233}	^{309}HVTL^{312}
Human CNP	^{37}GLPGSGKS^{44}	^{231}HCTT^{234}	^{310}HITL^{313}

5.1.2. Secondary structure prediction

According to PSIPRED, both *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase are predicted to contain both alpha helices and beta strands. The PSIPRED predicts that the "P-loop" motif of *Sc*Trl1 PNK/CPDase is located between the first β-strand and the first α-helix, and that the last three residues of the motif are part of the α-helix [Figure 13]. It is also predicted that both "H-x-(T/S)-x" motifs in *Sc*Trl1 PNK/CPDase form β-strands.

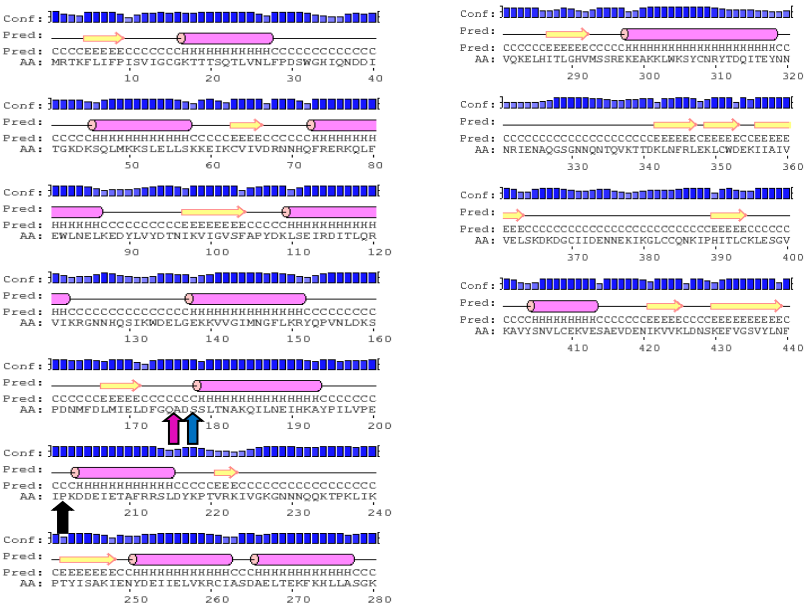


Figure 13. The PSIPRED prediction of secondary structural features of *Sc*Trl1 PNK/CPDase. The cylinders represent α-helices, the arrows indicate β-strands, and the straight lines indicate random coil structure. Conf: confidence of prediction, Pred: predicted secondary structure, and AA: target amino acid sequence. The N-termini of *Sc*Trl1 CPDase, *Sc*Trl1 (N+) CPDase and *Sc*Trl1 (N-) CPDase are denoted by pink, blue and black arrows, respectively.

Similarly to *Sc*Trl1 PNK/CPDase, the P-loop of *Bf* PNK/CPDase is predicted to be present between the first β-strand and the second α-helix and the last three residues are part of the α-helix [Figure 14]. However, the secondary structural features of the "H-x-(T/S)-x" motifs in *Bf* PNK/CPDase are predicted to be different from *Sc*Trl1 PNK/CPDase. The "H-x-(T/S)-x" motif II (HITL) is predicted to form a β-strand, whereas the "H-x-(T/S)-x" motif I (HCTA) is part of a random coil structure.

Further prediction of the amount of secondary structures, using the SOPMA tool, indicates that the *Sc*Trl1 PNK/CPDase contains 36% of α-helices and 22 % of β-strands and the *Bf* PNK/CPDase contains more α (38%) and less β (20%) structures. The full-length mouse CNP is predicted to contain 28% of α-helices and 23% of β-

strands, which is consistent with the results obtained using CD spectroscopy (Myllykoski and Kursula 2010).

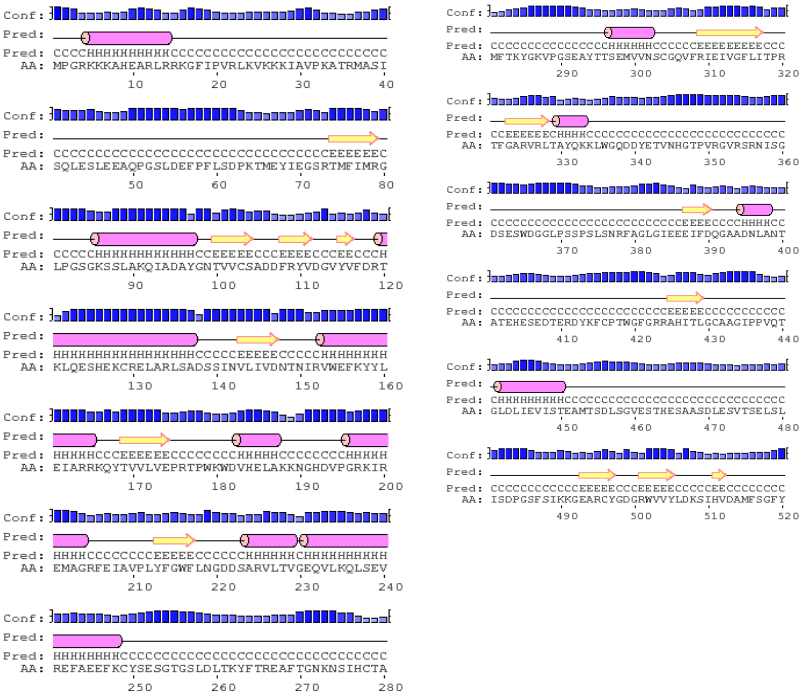


Figure 14. The PSIPRED prediction of secondary structural features of *Bf* PNK/CPDase. The features that represent different secondary structures are the same as mentioned in Figure 13.

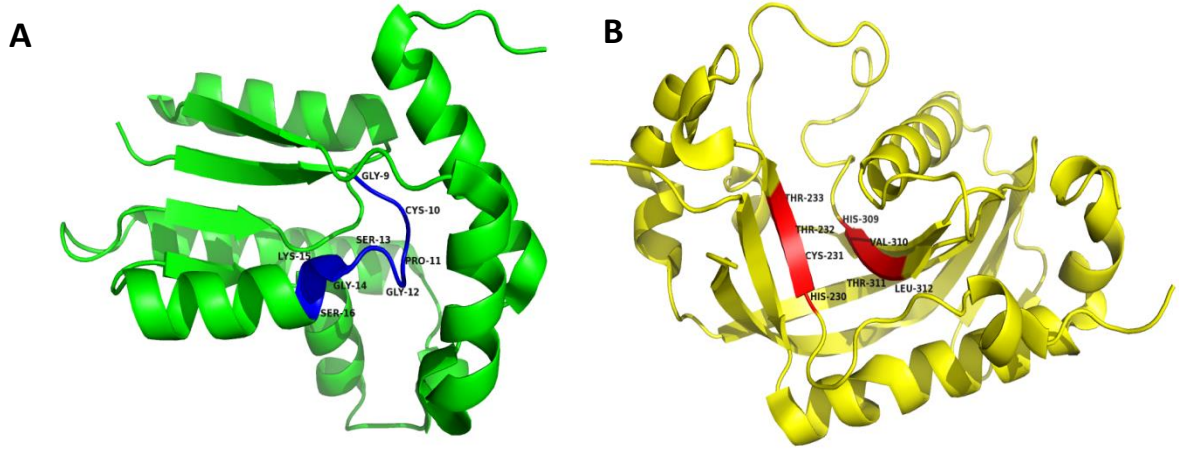


Figure 15. Crystal structures of homologous proteins with the location of the residues that constitute "P-loop" and "2H" motifs. A) T4 PNK (green) with the P-loop (blue) residues (^{9}GCPGSGKS^{16}) labeled [PDB ID: 1LY1] (Wang, Lima et al. 2002). **B)** The catalytic domain of mouse CNPase (yellow) with the residues of "2H" motifs (^{230}HCTT^{233} and ^{309}HVTL^{312}; red) labeled [PDB ID: 2YDB] (Myllykoski, Raasakka et al. 2012). The figures were generated using PyMOL.

In the crystal structure of T4 PNK, the P-loop (^{9}GCPGSGKS^{16}) residues, Gly9, Cys10, Pro11, Gly12 and Ser13 are part of a random coil and Gly14, Lys15 and Ser16 lie on the first α-helix [Figure 15 A] (Wang, Lima et al. 2002). The predicted secondary structural elements of the PNK domain of both *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase are in agreement with the crystal structure of the homologue, T4 PNK (Wang, Lima et al. 2002). The crystal structure of the catalytic domain of mouse CNP reveals that both the "2H" motifs form β-strands [Figure 15 B] (Myllykoski, Raasakka et al. 2012). The predicted secondary structure of the "2H" motifs in the CPDase domain of *Sc*Trl1 PNK/CPDase is in agreement with the crystal structure of the catalytic domain of mouse CNP (Myllykoski, Raasakka et al. 2012). However, in *Bf* PNK/CPDase, only the "H-x-(T/S)-x" motif II (HITL) is predicted to form a secondary structure similar to that in the crystal structure of the catalytic domain of mouse CNP.

5.1.3. Domain prediction

The analysis of protein sequences, using SMART, predicts the presence of kinase and CPDase domains in both *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase, and defines the boundaries of the two domains [Figure 16].

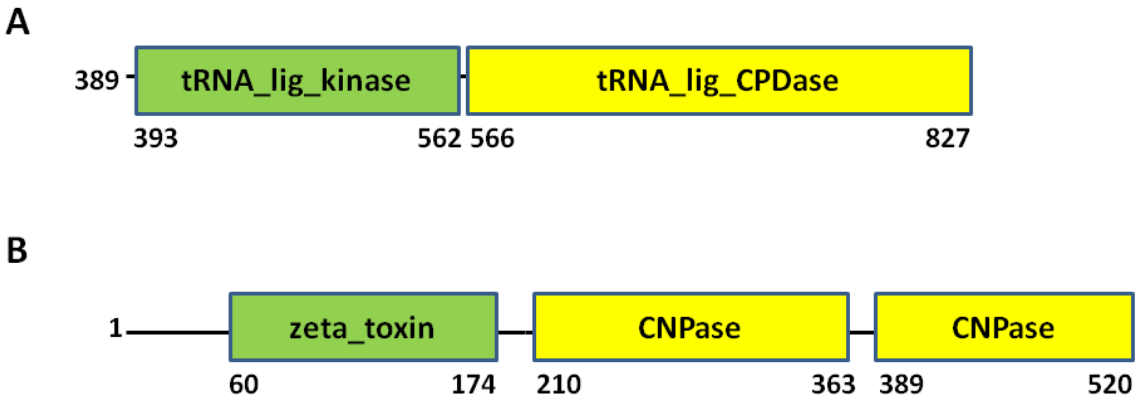


Figure 16. Domain organization as predicted by SMART. **A)** *Sc*Trl1 PNK/CPDase, **B)** *Bf* PNK/CPDase. Green - predicted PNK domain, Yellow - predicted CPDase domain.

The N-terminal domain of *Bf* PNK/CPDase, identified as a zeta toxin domain, is the polynucleotide kinase domain. The name, zeta toxin, comes from bacterial zeta toxin proteins, in which the domain was originally identified. The zeta toxin domain has been verified to function as a kinase domain (Galburt, Pelletier *et al*. 2002).

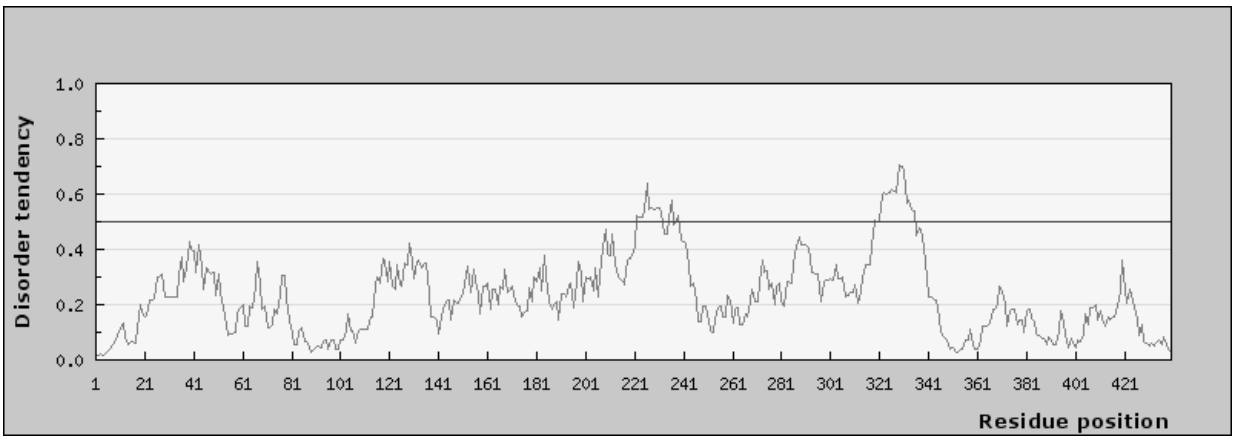
Both the kinase domain of *Sc*Trl1 PNK/CPDase and the zeta toxin (kinase) domain of *Bf* PNK/CPDase contain the P-loop motif, characteristic of PNK family of proteins [Figure 12]. The very N terminus of *Bf* PNK/CPDase, upstream of the predicted zeta

toxin domain is predicted to be mostly a random coil structure [Figure 14]. The "H-x-(T/S)-x" motifs that define the "2H" phosphoesterase superfamily are present in the predicted CPDase domain of *Sc*Trl1 PNK/CPDase and CNPase domain of *Bf* PNK/CPDase [Figure 12]. The CPDase domain of *Sc*Trl1 PNK/CPDase is predicted to be a close homologue of plant CPDase, whereas the CPDase domain of *Bf* PNK/CPDase is predicted to be a relatively close homologue of vertebrate CNPase, as indicated by their names used in the prediction [Figure 16]. The prediction of domain boundaries is useful to prepare constructs of different lengths of the individual domains.

5.1.4. Disorder prediction

The intrinsically disordered regions of *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase were predicted using the IUPred tool [Figure 17].

A



B

Figure 17. Prediction of intrinsically disordered regions using IUPred tool. The predicted disordered regions of **A)** *Sc*Trl1 PNK/CPDase (sequence number from 1 to 439 is equal to sequence number from 389 to 827 in *Sc*Trl1). **B)** *Bf* PNK/CPDase. Prediction threshold is indicated by a straight line **[A & B]**.

The IUPred predicts that intrinsically disordered regions of both *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase are present in the CPDase domain. In *Sc*Trl1 PNK/CPDase, a portion of the C-terminal insertion [Figure 12], which is located between residues 608 and 628 of *Sc*Trl1 (residues 220 and 240 in Figure 17 A), is predicted to be a disordered region, however with a low disorder probability. Another segment, which lies between residues 709 and 729 of *Sc*Trl1 (residues 320 and 340 in Figure 17 A), shows a higher tendency to be unstructured [Figure 16 A]. This segment is located between the two "H-x-(T/S)-x" motifs of *Sc*Trl1 PNK/CPDase [Table 4]. The intrinsically unstructured regions of *Bf* PNK/CPDase are predicted to be present in the C-terminal insertion located between the two "H-x-(T/S)-x" motifs [Figure 17 B, Table 4].

5.1.5. Homology-based models

Sequence-based homology models of *Sc*Trl1 PNK, *Sc*Trl1 CPDase, and the CPDase domain of *Bf* PNK/CPDase were built using Phyre[2], based on the available high-resolution structure of a close homologue of the proteins. The homology model of *Sc*Trl1 PNK [Figure 18A] was built based on the crystal structure of mammalian bifunctional polynucleotide kinase 3'-phosphatase (PNK) [PDB ID: 3ZVL] (Garces, Pearl et al. 2011). The mammalian PNK plays a crucial role in DNA repair either by removal of 3'-phosphates or by phosphorylation of 5'-hydroxyl groups on the ribose sugar of the DNA backbone. The crystal structure of murine PNK, used as a template in homology-modeling, has been determined with DNA bound to both the PNK and phosphatase domains (Garces, Pearl et al. 2011). 156 residues (35% of sequence) in the PNK domain of *Sc*Trl1 PNK/CPDase were used in the modeling with 99.8% confidence. The overall structure and the P-loop motif of *Sc*Trl1 PNK are structurally similar to the crystal structure of T4 PNK, another close homologue of *Sc*Trl1 PNK [Figure 18 A] (Wang, Lima et al. 2002).

The homology model of *Sc*Trl1 CPDase [Figure 18 B] was built using the crystal structure of YjcG protein from *Bacillus subtilis* [PDB ID: 2D4G] as a template (Li, Liu et al. 2008). 143 residues out of 265 residues (54%) in the CPDase domain of ScTrl1 PNK/CPDase were modeled with 96.3% confidence. The YjcG protein contains the two conserved tetrapeptide motifs (H-x(T/S)-x) that define the 2H phosphoesterase superfamily, and exhibits sequence homology to bacterial and archaeal 2'-5' RNA ligases (Mazumder, Iyer et al. 2002, Kato, Shirouzu et al. 2003, Li, Liu et al. 2008). The side chains of His and Thr/Ser residues of the "H-x-(T/S)-x" motifs are structurally conserved among the members of 2H phosphoesterase superfamily (Li, Liu et al. 2008).

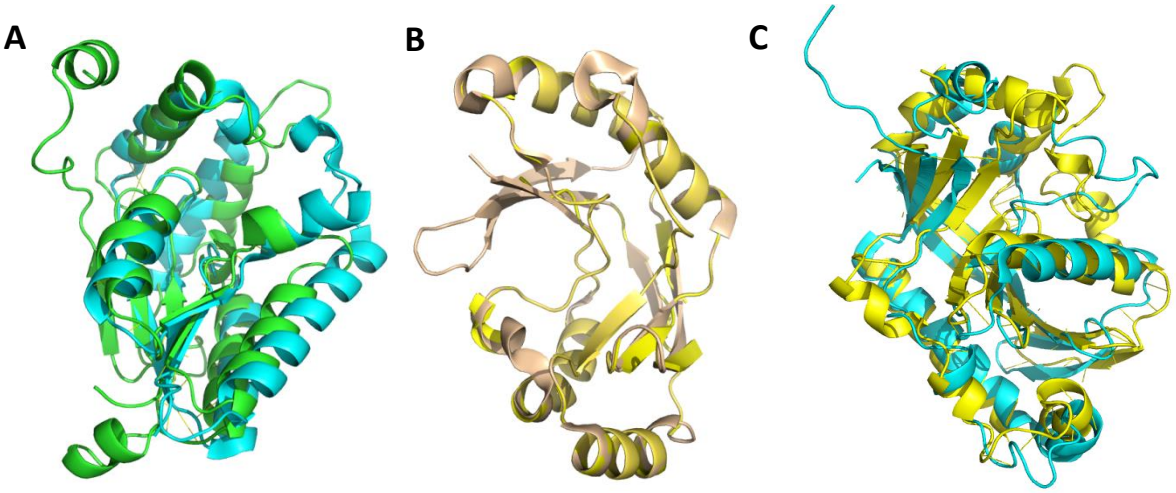


Figure 18. Homology models built using Phyre[2]. A) Superposition of the homology model of *Sc*Trl1 PNK (green) and the crystal structure of T4 PNK [PDB ID: 1LY1] (Wang, Lima et al. 2002). **B)** Superposition of the homology model of *Sc*Trl1 CPDase (yellow) and the crystal structure of YjcG protein from *Bacillus subtilis* (tint) [PDB ID: 2D4G] (Li, Liu et al. 2008). **C)** Superposition of the homology model of the CPDase domain of *Bf* PNK/CPDase (yellow) and the solution structure of the catalytic domain of goldfish RICH protein (cyan) [PDB ID: 2I3E] (Kozlov, Denisov et al. 2007). The figures were generated using PyMOL molecular graphics program.

The homology model of the CPDase domain of *Bf* PNK/CPDase was built based on the solution structure of the catalytic domain of goldfish RICH protein (Kozlov, Denisov et al. 2007). 215 residues (41% of sequence) of *Bf* PNK/CPDase were modeled with 100% confidence. The superposition of the homology model on the solution structure of the RICH protein indicates that the active site consisting of the "2H" motifs is structurally conserved [Figure 18 C]. The homology model of *Bf* PNK/CPDase contains a symmetrical bilobed tertiary structure, which is similar to the crystal structures of the catalytic domain of human and mouse CNPase proteins (Sakamoto, Tanaka et al. 2005, Myllykoski, Raasakka et al. 2012).

5.1.6. Prediction of crystallizability

As one of the main goals of the study was to determine the three-dimensional structures of the proteins using X-ray crystallography, production of diffraction-quality crystals is an inevitable requirement. The crystallizability of different constructs of yeast Trl1 and lancelet PNK/CPDase was predicted using the crystallization propensity predictors, XtalPred and ParCrys. The results from the XtalPred tool are summarized in Table 5.

Table 5. Results of crystallizability prediction by XtalPred

Protein	Expert Pool Crystallizability score	Instability index	Longest disorder region (residues)
ScTrl1 PNK/CPDase	4	39 %	26
*Sc*Trl1 CPDase	4	30 %	22
*Sc*Trl1 (N+) CPDase	4	31 %	21
*Sc*Trl1 (N-) CPDase	4	26 %	20
Bf PNK/CPDase	4	39 %	26

The expert pool crystallizability score, also known as crystallization feasibility score, is a combination of the probability distributions calculated for several individual features of the protein, such as protein length, isoelectric point, instability index, extinction co-efficient, secondary structure, percentage of coiled-coil regions, disordered regions, transmembrane helices and signal peptides (Slabinski, Jaroszewski et al. 2007). The score assigns proteins to one of the five crystallization classes (1-optimal; 2 - suboptimal; 3 -average; 4 - difficult, and 5 - very difficult). All the constructs used in this study are predicted to be difficult to crystallize [Table 5]. The longest disorder regions predicted by XtalPred are consistent with those predicted by IUPred [Figures 17 A & B].

Table 6. Results of ParCrys crystallizability prediction

Protein	ParCrys prediction
ScTrl1 PNK/CPDase	Recalcitrant
*Sc*Trl1 CPDase	Recalcitrant
*Sc*Trl1 (N+) CPDase	Recalcitrant
*Sc*Trl1 (N-) CPDase	Recalcitrant
Bf PNK/CPDase	Recalcitrant
Mouse CNP	Amenable

The results of ParCrys prediction are given in Table 6. The proteins investigated in this study are predicted to be “recalcitrant”, whereas the homologue, mouse CNPase is predicted to be “responsive” [Table 6]. In agreement with the ParCrys prediction, the catalytic domain of mouse CNPase has been crystallized, and its structure has been determined (Myllykoski, Raasakka et al. 2012). ParCrys also predicts a high level of instability for the PNK/CPDase and CPDase proteins, which could affect the crystallizability significantly. According to the predictions, crystallization of these proteins could be a challenging task.

5.2. Preparation of expression constructs

In order to study, in detail, the structural and functional properties of PNK/CPDase domains of *Sc*Trl1 and *Bf* PNK/CPDase, a total of eight constructs were prepared for the PNK and CPDase domains of yeast Trl1, in addition to another construct encoding the lancelet PNK/CPDase. The scheme of the constructs is given in Figure 19. One construct was prepared for both PNK and CPDase domains of yeast Trl1, four for the N-terminal PNK domain and three for the C-terminal CPDase domain. The identity of the constructs was verified by plasmid sequencing. The constructs can be used to express histidine-tagged proteins with a 3C protease-cleavable site at the N-terminus.

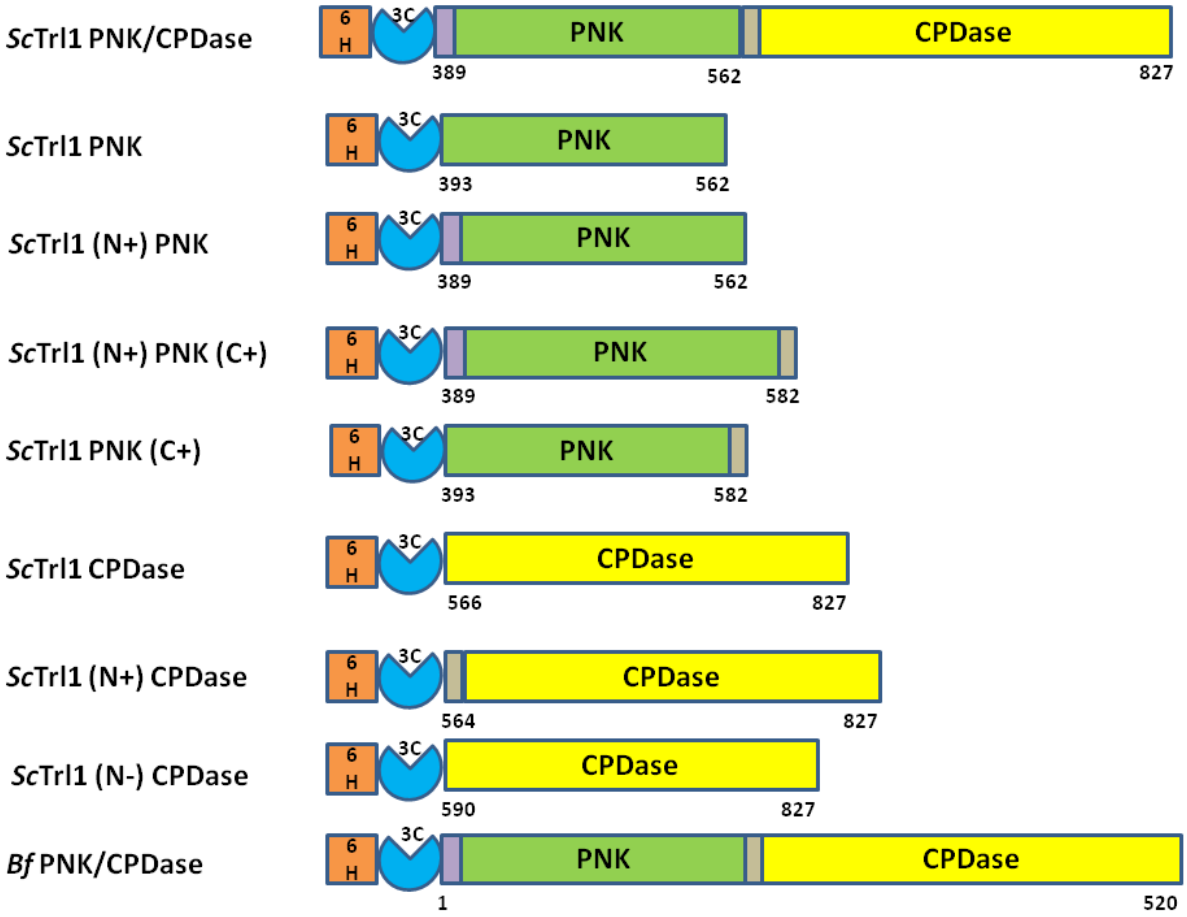


Figure 19. Preparation of expression constructs. Scheme of the constructs prepared for the kinase and phosphodiesterase domains of yeast Trl1 and for lancelet PNK/CPDase. 6H indicates the hexa-histidine tag and 3C indicates the 3C protease cleavable site.

5.3. Expression screening in bacterial cells

The expression level of the different constructs was screened by applying a variety of different expression conditions. The expression levels were initially tested in three different *E. coli* cell lines, BL21 (DE3), BL21 CodonPlus (DE3) RIPL and Rosetta

(DE3) grown in LB medium, and by applying four different inducer concentrations [Table 7].

The PNK/CPDase and different CPDase constructs of yeast Trl1 and full-length lancelet PNK/CPDase were successfully expressed in both BL21 CodonPlus (DE3) RIPL and Rosetta (DE3) cells. These constructs showed little expression in BL21 (DE3) cells and only at higher inducer concentrations. It could be related to the availability of certain additional tRNA genes in BL21 CodonPlus (DE3) RIPL and Rosetta (DE3) that enhances the translation of heterologous proteins in *E. coli*. Those additional tRNA genes are not present in BL21 (DE3) cells.

Table 7. Relative expression levels of the constructs in different growth conditions

Construct	**Relative level of expression**											
	BL21 DE3 IPTG (mM)				**BL21 (DE3) CodonPlus RIPL** IPTG (mM)				**Rosetta (DE3)** IPTG (mM)			
	0.15	0.25	0.5	1	0.15	0.25	0.5	1	0.15	0.25	0.5	1
*Sc*Trl1 PNK/CPDase	--	+	+	+	++	++	++	++	+++	+++	++	+++
*Sc*Trl1 PNK	--	--	--	--	--	--	--	--	--	--	--	--
*Sc*Trl1 (N+) PNK	--	--	--	--	--	--	--	--	--	--	--	--
*Sc*Trl1 (N+) PNK (C+)	--	--	--	--	--	--	--	--	--	--	--	--
*Sc*Trl1 PNK C+	--	--	--	--	--	--	--	--	--	--	--	--
*Sc*Trl1 CPDase	--	--	+	+	++	++	++	++	+++	+++	++	+++
*Sc*Trl1 (N+) CPDase	--	--	+	+	++	++	++	++	+++	+++	++	+++
*Sc*Trl1 (N-) CPDase	--	--	+	+	++	++	++	++	+++	+++	++	+++
Bf PNK/CPDase	--	+	+	+	++	++	++	++	+++	+++	++	+++

--- indicates no expression and +++ indicates the highest soluble expression

The proteins encoding different regions of the PNK domain of yeast Trl1 did not show any expression in the three cell lines tested. The same pattern was observed previously for the constructs encoding the N-terminal domain of mouse 2',3'-cyclic nucleotide 3'-phosphodiesterase (Myllykoski and Kursula 2010). The attempts made by others previously to express the PNK domain of yeast Trl1 in bacterial cells were also unsuccessful (Sawaya, Schwer *et al*. 2003). Since the solubility of the proteins expressed in Rosetta (DE3) cells was high, the same cell line was used for further large-scale protein production. The level of solubility of the proteins at three different growth conditions post-induction was tested, and longer incubation at lower temperatures was found to be the optimal growth condition.

The conditions used for the growth of bacterial cells were incubation at 37°C until the optical density at 600 nm reached 0.6 and post-induction growth at 18°C for 16 h. IPTG was used at a concentration of 0.15 mM. The expressed proteins were initially purified by Ni-NTA affinity chromatography and further by size exclusion chromatography. The amount of soluble *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase produced was always comparatively less than that of the shorter CPDase constructs.

Table 8. Relative solubility levels at different post-induction growth conditions

Construct	**Amount of solubility**		
	Post-induction growth		
	37°C for 6 h	**20°C for 16 h**	**15°C for 16 h**
*Sc*Trl1 PNK/CPDase	+	++	++
*Sc*Trl1 CPDase	+	++	+++
*Sc*Trl1 (N+) CPDase	+	+++	+++
*Sc*Trl1 (N-) CPDase	+	+++	+++
Bf PNK/CPDase	+	++	+++

+: medium, ++: high and +++: the highest solubility obtained

Expression levels were also tested using LB and AI media, and no significant difference was observed. Hence, LB medium was routinely used for the growth of *E. coli* cells. As a result of the expression and solubility screens, optimal conditions for increased expression and solubility were established. For large-scale protein production of *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase, up to 8 l of LB medium were used to obtain about 5 mg of protein, as the typical yield from one liter of culture was less than 1 mg. For the shorter constructs, since the yields were higher, 2 l of culture was normally enough to achieve similar yields. The harvested cells were stored at -70°C until further use.

5.4. Purification of the yeast proteins

The expressed proteins were initially purified by Ni-NTA affinity chromatography [Figure 20]. The proteins bound to the Ni-NTA matrix were eluted by using 500 mM imidazole after the matrix was briefly washed with 50 mM imidazole to remove any non-specific interactants. The proteins encoding the CPDase domain were comparatively more pure than those containing both the PNK and CPDase domains. The absorbance of the eluates, monitored at both 260 and 280 nm, indicated the presence of nucleic acid impurities. Two different molecular populations of the purified PNK/CPDase proteins were detected. Upon elution with 500 mM imidazole,

the nucleic acid-binding PNK/CPDase was eluted at first, and the fractions collected afterwards were found to be nucleic-acid free (data not shown).

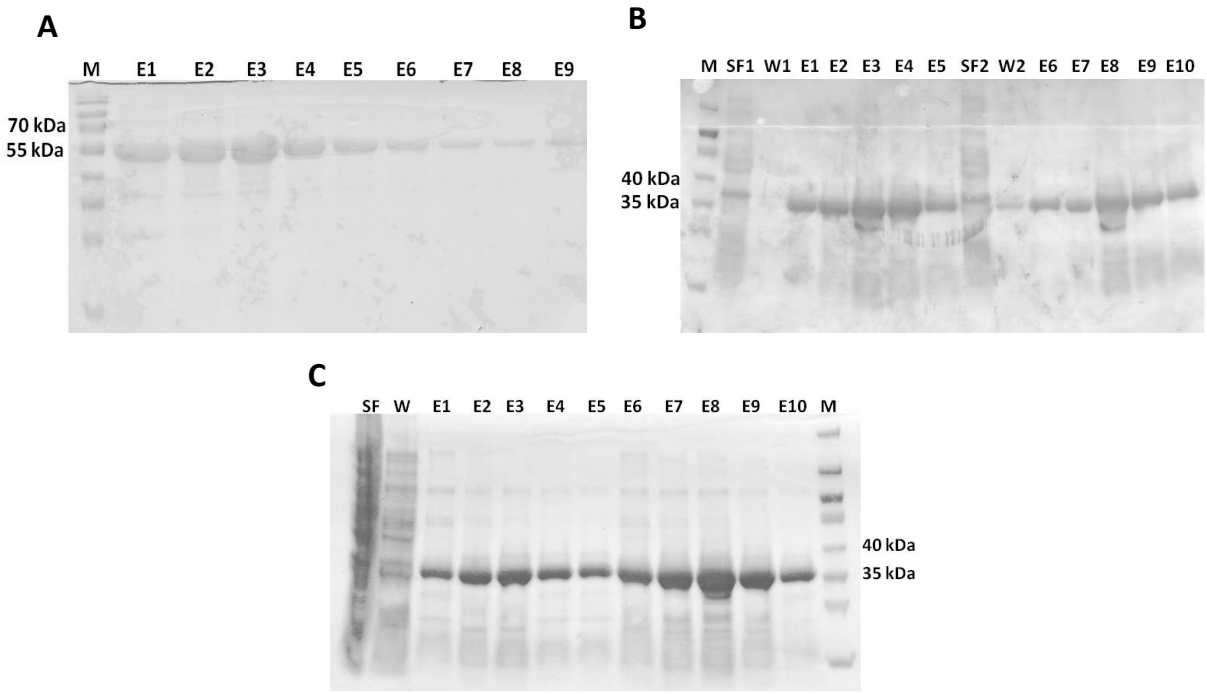


Figure 20. The yeast PNK/CPDase and CPDase proteins purified by Ni-NTA affinity chromatography. A) SDS PAGE from a Ni-NTA purification of *Sc*Trl1 PNK/CPDase. M: marker, E1-E8: fractions eluted with 500 mM imidazole. The calculated size of the protein was 52.7 kDa. **B)** SDS PAGE from Ni-NTA purifications of *Sc*Trl1 CPDase and *Sc*Trl1 (N+) CPDase. M: marker, SF1 and W1 are supernatant and wash flow-through samples from the purification of *Sc*Trl1 CPDase, E1-E5: *Sc*Trl1 CPDase fractions eluted with 500 mM imidazole. The calculated size of the protein was 32.2 kDa. SF2 and W2 are supernatant and wash flow-through samples from the purification of *Sc*Trl1 (N+) CPDase, E6-E10: *Sc*Trl1 (N+) CPDase fractions eluted with 500 mM imidazole. The calculated size of the protein was 32.4 kDa. **C)** SDS PAGE from a Ni-NTA purification of *Sc*Trl1 (N-) CPDase. SF: supernatant flow-through, W: wash flow-through, E1-E10: fractions eluted with 500 mM imidazole, M: marker. The calculated size of the protein was 29.6 kDa.

The purified His-tagged proteins were dialyzed overnight with His-3C protease for cleavage of the hexa-histidine tag. The tag-cleaved protein was separated from the mixture of His-3C protease, cleaved tag and any uncleaved protein by another step of Ni-NTA affinity chromatography. The proteins were subjected to size exclusion chromatography [Figure 21] using either HiLoad 16/60 Superdex 200 (120 ml) or Superdex 200 10/300 GL (25 ml). The *Sc*Trl1 PNK/CPDase eluted as double peaks with the elution volumes corresponding to dimeric and monomeric forms of the protein. The fractions from each of the two peaks were collected separately, concentrated and analyzed by SDS PAGE. Both peaks were found to contain pure

protein. The homologue, mouse CNPase exhibits similar oligomeric behavior and the N-terminal domain of the enzyme has been shown to mediate dimerization (Myllykoski, Raasakka et al. 2012).

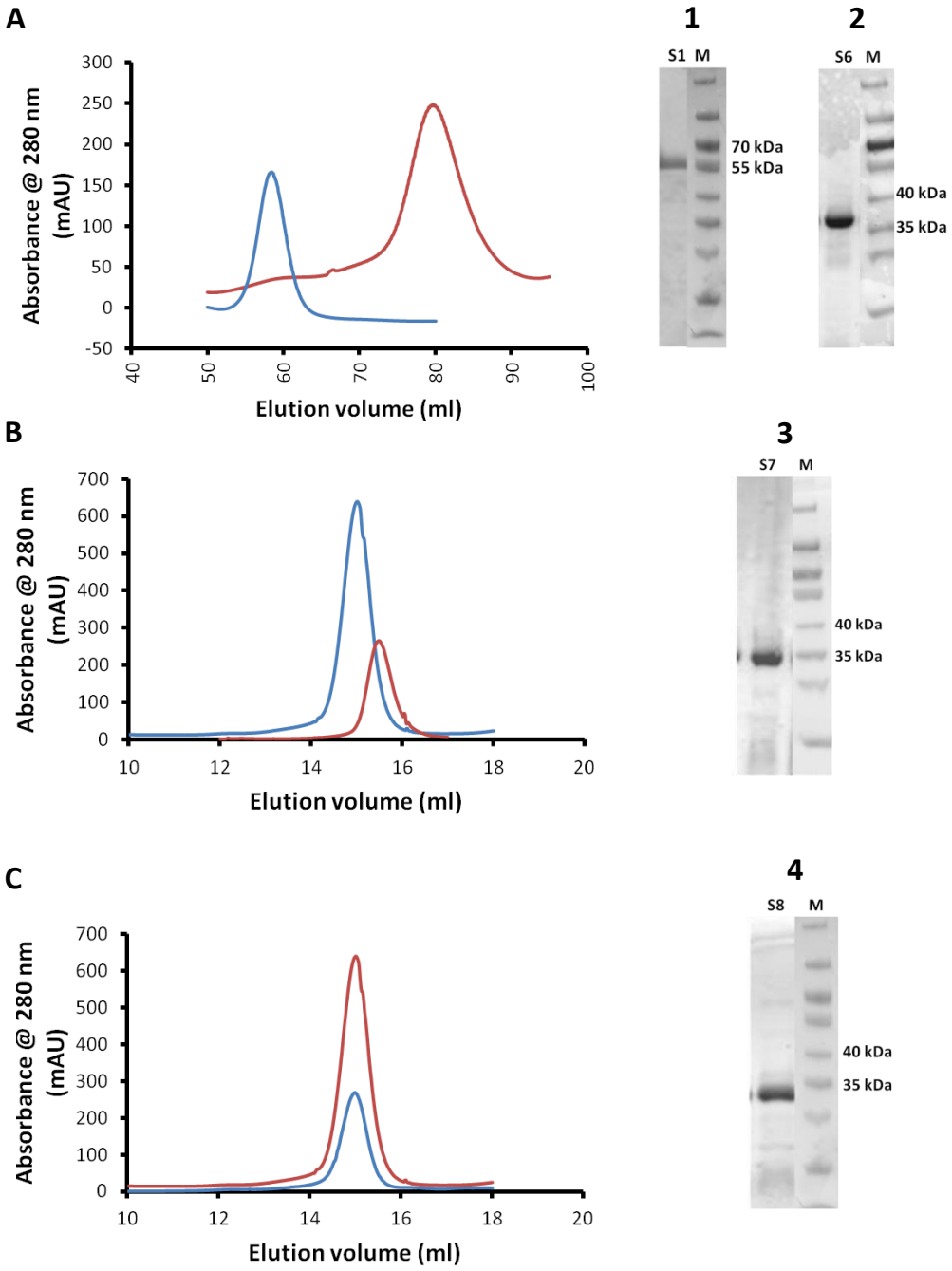


Figure 21. The yeast PNK/CPDase and CPDase proteins purified by size exclusion chromatography. A) Superdex200 SEC profile of *Sc*Trl1 PNK/CPDase monomer (blue) and *Sc*Trl1 CPDase (red). **B)** Superdex75 SEC profile of *Sc*Trl1 CPDase (blue) and *Sc*Trl1 (N-) CPDase (red). **C)** Superdex75 SEC profile of *Sc*Trl1 CPDase (red) and *Sc*Trl1 (N+) CPDase (blue). On the right: SDS PAGE of purified proteins **1)** S1: *Sc*Trl1 PNK/CPDase **2)** *Sc*Trl1 CPDase **3)** *Sc*Trl1 (N-) CPDase **4)** N+ *Sc* CPDase *Sc*Trl1 (N+) CPDase. M: marker (1, 2, 3 and 4). The CPDase proteins elute around the same elution volume and exist as monomers.

The C-terminal extended catalytic domain of mouse CNPase also forms a dimer, which could be dissociated into monomers by using DTT (Myllykoski and Kursula 2010). Similarly, the addition of reducing agents, either 10 mM β-ME or 5 mM DTT, in the lysis and purification buffers was found to inhibit dimerization of *Sc*Trl1 PNK/CPDase, indicating that the monomers are linked by a disulphide bridge (discussed in sub-section 5.15).

The elution profiles of the CPDase constructs contained mostly a single symmetric peak [Figures 21 B & 21 C]. During the purification of *Sc*Trl1 PNK/CPDase, a large peak was always noticed at the void volume of the size exclusion column [Figure 23 A]. Since the absorbance at 260 nm was very high for this peak, co-purification of a nucleic acid, possibly RNA from the expression host, was suspected. When a combination of RNase A and RNase T1 was added in the lysis buffer, the A_{260} peak height remarkably decreased.

The purification of the CPDase domain was relatively easier as there was no oligomerization or RNA binding detected. The absence of dimerization and RNA binding in the CPDase domain suggest that these processes require the N-terminal PNK domain. Similarly, the N-terminal PNK-like domain of mouse CNPase is involved not only in dimerization, but also in RNA binding (Myllykoski, Raasakka et al. 2012). All three CPDase proteins eluted around the same elution volume, and the final yields were almost the same. The purified proteins were easily degraded upon storage at 4°C. Hence, immediate flash-freezing and storage of the purified proteins in the freezer was essential. Since the stability of *Sc*Trl1 CPDase was found to be comparatively better than the other two CPDase proteins, it was used in other biophysical and biochemical experiments. The identity of all purified proteins was verified by mass spectrometry (data not shown).

5.5. Purification of lancelet PNK/CPDase

Similarly to the yeast proteins, *Bf* PNK/CPDase was purified in three chromatographic steps; initial affinity purification from soluble lysate [Figure 22 A], purification of the tag-cleaved protein by a second affinity chromatography and size exclusion chromatography [Figures 22 B & 22 C]. Like *Sc*Trl1 PNK/CPDase, the lancelet protein also contained RNA that was carried over from the expression host [discussed in sub-section 5.6].

Similarly to *Sc*Trl1 PNK/CPDase and mouse CNPase, the *Bf* PNK/CPDase eluted from the Superdex 200 10/300 GL column at 13 ml and formed a broad peak with a shoulder indicating the presence of a dimer and a monomer. However, in the presence of 10 mM β-ME, the protein eluted as a single peak around 15 ml on the Superdex

200 10/300 GL column, indicating that the purified protein exists as monomer [Figure 22 B]. Like the purified yeast proteins, *Bf* PNK/CPDase was also found to undergo rapid degradation. The addition of a combination of 2 mM phenylmethylsulfonyl fluoride (PMSF) and EDTA-free protease inhibitors in the lysis buffer significantly reduced the fragmentation of the protein by proteases.

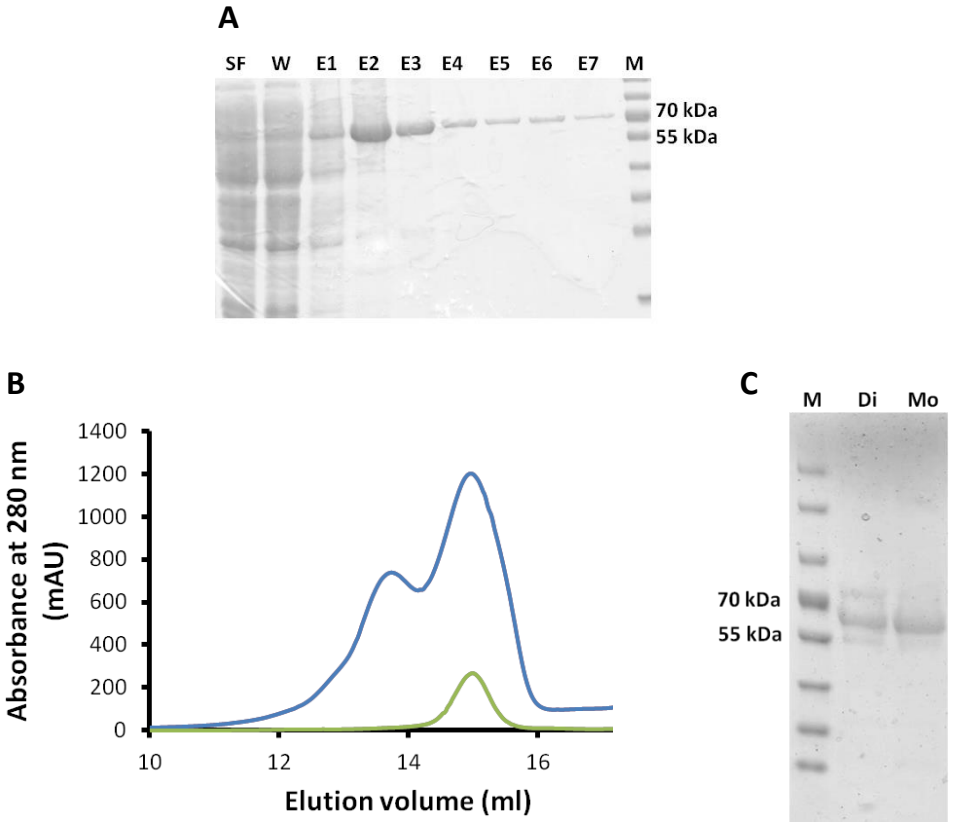


Figure 22. Purification of *Bf* PNK/CPDase. **A)** SDS PAGE from a Ni-NTA purification of *Bf* PNK/CPDase. SF: supernatant flow-through, W: wash flow-through, E1-E7: fractions eluted with 500 mM imidazole, M: marker. The calculated size of the protein was 60.3 kDa. **B)** Superdex75 SEC profile of *Bf* PNK/CPDase. The blue curve shows a mixture of monomer and dimer in the absence of any reducing agent and the green curve shows the monomer in the presence of 5 mM β-ME. The single monomeric peak (green) indicates that the dimer formation is inhibited by β-ME. **C)** SDS PAGE from a SEC purification of *Bf* PNK/CPDase in the absence of a reducing agent. M: marker, Di: *Bf* PNK/CPDase dimer, Mo: *Bf* PNK/CPDase monomer.

The purification of monomeric PNK/CPDase proteins was relatively difficult due to the presence of non-specifically associated RNA, the formation of dimers in a dynamic equilibrium and a rapid degradation of the purified proteins. The proteins purified from size exclusion chromatography were to be used instantly for any biophysical or biochemical analyses due to the instability of the purified proteins.

5.6. PNK/CPDase enzymes interact with *Eschericia coli* RNA

The presence of nucleic acid in the purified *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase was detected at first in the IMAC step. The large void volume peak in size exclusion chromatography also indicated that the nucleic acid gets co-purified [Figure 23 A]. To identify the nucleic acid bound to the protein, the fractions eluted around the void volume in SEC were treated with the nucleases DNase and RNase A. The agarose gels of nuclease-treated samples show that the bright smear that appears in the void volume sample and in the sample treated with DNase is not found in the sample treated with RNase A [Figure 23 B]. This finding confirms that the co-purified nucleic acid impurity is RNA.

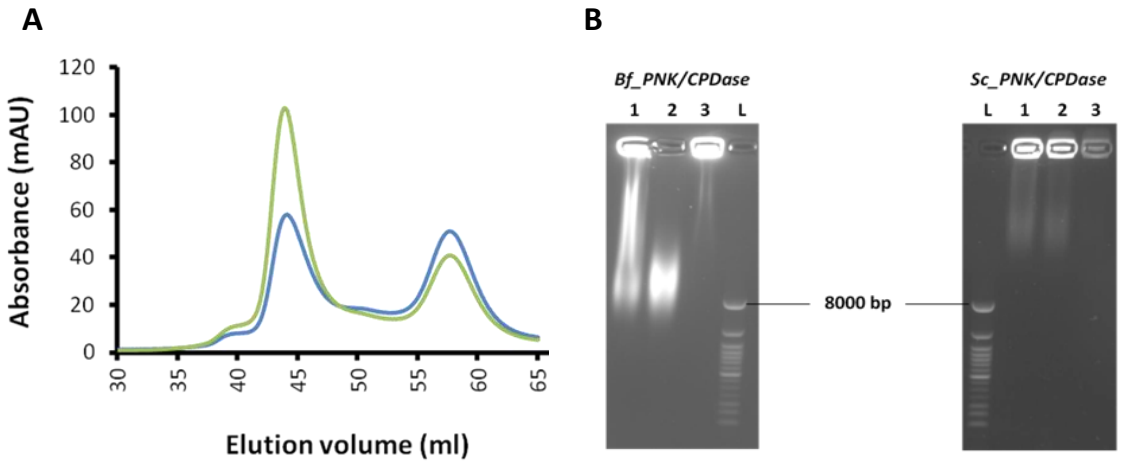


Figure 23. PNK/CPDase interacts with *E. coli* RNA. A) Superdex 200 SEC profile of *Sc*Trl1 PNK/CPDase. The void volume peak with A_{260} (green) is higher than A_{280} (blue), indicating the presence of a nucleic acid impurity. **B)** Agarose gels of *Bf* PNK/CPDase/ *Sc*Trl1 PNK/CPDase + RNA. L: DNA ladder, 1: The protein-nucleic acid complex from SEC void volume, 2: DNase-treated protein-nucleic acid complex and 3: RNase-treated protein-nucleic acid complex. RNase degrades the nucleic acid, confirming that the co-purified molecule is, in fact, RNA.

The behaviour of PNK/CPDase proteins in these assays is highly similar to the uracil-DNA degrading factor and mouse CNPase (Bekesi, Pukancsik et al. 2011, Myllykoski, Raasakka et al. 2012). CNPase has also been shown to interact with RNA *in vitro* and co-purify with poly(A)$^+$ RNA; the catalytic domain of CNPase has been claimed to be sufficient for binding with single stranded RNA homopolymers (Gravel, Robert et al. 2009). However, a pulldown assay with mouse CNPase and poly(A)-sepharose has indicated that the N-terminal domain binds RNA more efficiently than the C-terminal domain (Myllykoski, Raasakka et al. 2012). The absence of the void volume peak during SEC purification of the yeast CPDase domain indicates the N-

terminal PNK domain is required for RNA binding, at least in the case of *E. coli* RNA.

5.7. Bacterially expressed PNK/CPDase and CPDase proteins are folded

The folding state of the purified proteins was analyzed by SRCD spectroscopy. Visual inspection of the SRCD spectra, averaged from three consecutive scans and buffer-subtracted, indicates that the PNK/CPDase and CPDase proteins are folded, containing both α-helices and β-strands [Figure 24]. Following are the conformation-driven features applied while examining the spectra [Figure 24]: a strong negative peak at 200 nm for an unfolded peptide, the characteristic negative peaks at 208 and 220 nm and a positive peak at 191-193 nm for an α-helical peptide, and a broad negative peak around 218 nm for a β-stranded peptide. The SRCD spectra from *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase show a similar shape [Figure 24 A], an indication of the presence of a similar overall secondary structure.

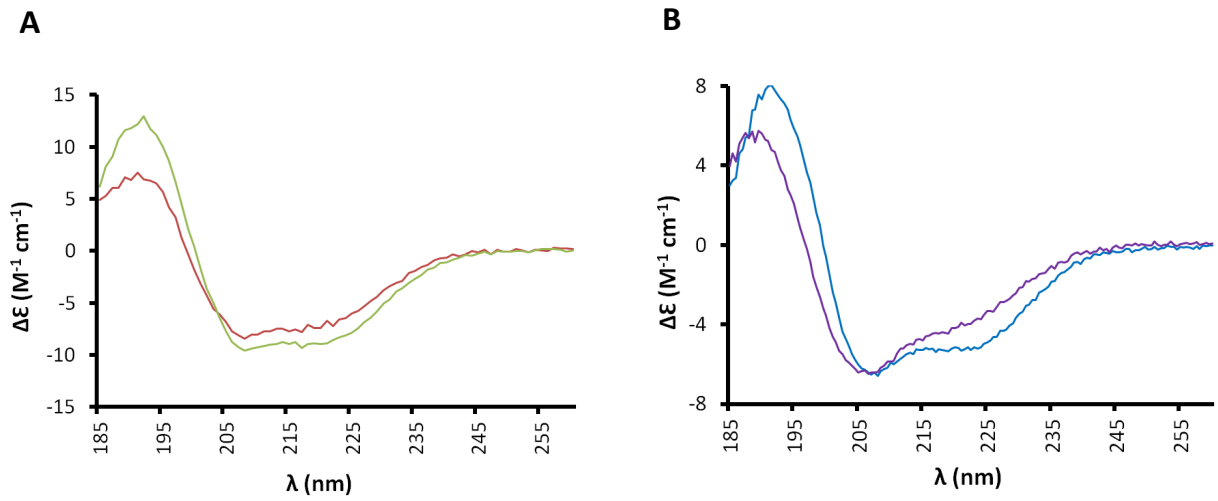


Figure 24. SRCD spectra of purified proteins. A) *Sc*Trl1 PNK/CPDase (red) and *Bf* PNK/CPDase (green). **B)** *Sc*Trl1 CPDase (blue) and *Sc*Trl1 (N-) CPDase (purple). The SRCD spectra indicate that all tested proteins are folded.

The percentages of secondary structural elements were calculated from the SRCD spectra using Dichroweb, and are listed in Table 9. All normalized root-mean-square deviations (NRMSDs) from the deconvolutions were between 0.003 and 0.009 [Table 9]. From spectral deconvolution, it becomes clear that the proteins contain a mixture of alpha and beta structure [Table 9]. Upon comparison to the available SRCD spectra, crystal and solution structures of the homologous proteins of the 2H phosphoesterase superfamily, the yeast and lancelet proteins have higher alpha helical contents and almost the same amount of beta strands [Table 9]. The secondary structure from SRCD analysis of the mouse CNPase constructs (Myllykoski and

Kursula 2010), the crystal structure of human CNPase (Sakamoto, Tanaka *et al*. 2004, Kozlov, Denisov *et al*. 2007)) and the solution structure of the catalytic domain of gold fish RICH protein were compared (Kozlov, Denisov *et al*. 2007) [Table 9]. The results also indicate that the α-helical content calculated from SRCD analysis of PNK/CPDase and CPDase proteins is higher than the amount predicted by the SOPMA tool [sub-section 5.1.2].

Table 9. Deconvolution of CD spectra

Protein	**α-helix** (%)	**β-strands** (%)	**NRMSD** (%)
Protein used in the study			
*Sc*Trl1 PNK/CPDase	52	27	0.003
Bf PNK/CPDase	61	20	0.007
*Sc*Trl1 CPDase	53	29	0.004
*Sc*Trl1 (N-) CPDase	43	26	0.008
Homologous proteins			
Full-length mouse CNPase[1]	28	23	-
Mouse CNPase N-terminal domain[1]	8	39	-
Mouse CNPase C-terminal domain[1]	16	33	-
Crystal structure of C-terminal domain[2]	30	32	-
NMR structure of C-terminal domain[3]	28	27	-

[1](Myllykoski and Kursula 2010), [2](Sakamoto, Tanaka *et al*. 2004), [3](Kozlov, Denisov *et al*. 2007)

However, the presence of higher α-helical content in both PNK/CPDase and CPDase proteins is in agreement with the secondary structure predicted from the amino acid sequence of these proteins using PSIPRED [Figures 13 and 14]. The additional α-helices, according to PSIPRED prediction, are mostly formed by the residues that are present in the insertions of yeast Trl1 and lancelet PNK/CPDase proteins [Figures 12, 13 and 14]. The crystal structure of the catalytic domain of mouse CNPase reveals the presence of 2 long helices and 5 short helices in the C-terminal domain (Myllykoski, Raasakka et al. 2012). Based on PSIPRED prediction, the C-terminal domain of *Sc*Trl1 PNK/CPDase contains 9 long helices (long helix - formed by more than 10 residues) and 1 short helix (short helix - formed by less than 10 residues), out of which 5 long helices and the short helix are present in the C-terminal CPDase domain [Figure 13]. The same trend is predicted for *Bf* PNK/CPDase, which contains 5 long helices and 7 short helices [Figure 14]. Despite significant differences in the amount of α-helical contents, the amount of β-strands calculated from the SRCD spectra is in the same range as predicted by PSIPRED and SOPMA tools [sub-section 5.1.2].

5.8. Thermal stability analysis

Since the stability of the monomers of both *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase during and after purification was low, the identification of a stabilizing buffer was necessary. A thermal stability assay was used to identify suitable buffer conditions that would offer increased stability of the protein. A total of 80 different conditions, varying in pH, salt type and concentration, and the presence of ligands, were screened using the high throughput thermal stability assay. The apparent thermally induced melting points (T_m) were measured for each construct under all 80 conditions over a temperature range between 25°C and 99°C. The melting curves and corresponding T_m values of *Sc*Trl1 PNK/CPDase, *Sc*Trl1 CPDase and *Bf* PNK/CPDase indicate that all three proteins tested are most stable around pH 7.5 [Fig 25, Table 10].

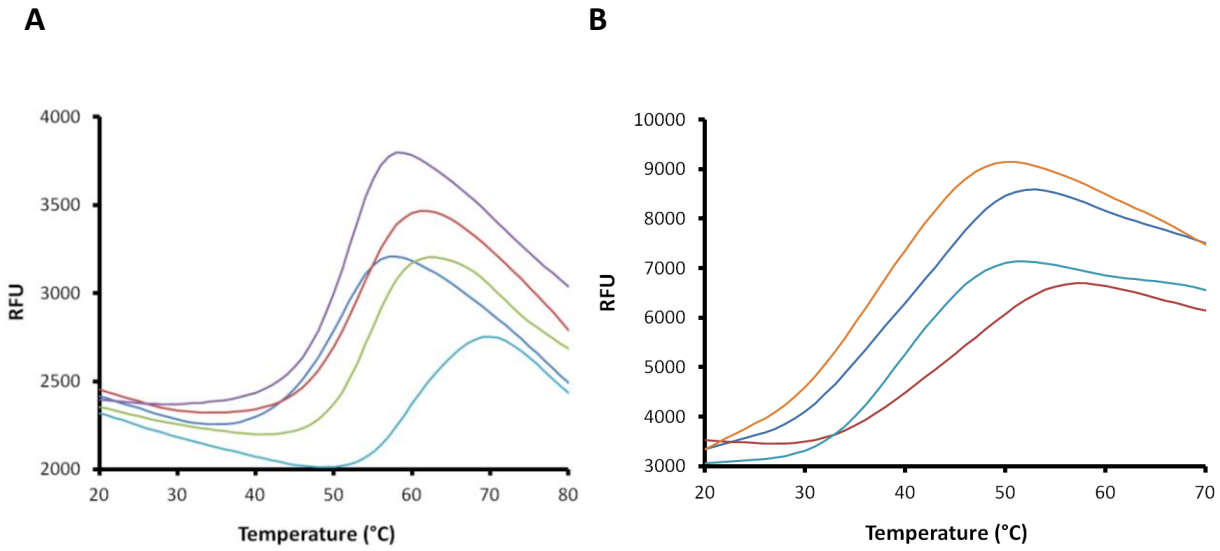


Figure 25. Thermal stability assay. A) A selection of thermal shift assay melting curves for *Sc*Trl1 PNK/CPDase. Color code: Violet - 50 mM citric acid, pH 5.5, 150 mM NaCl; Blue - 50 mM MOPS, pH 6.5, 500 mM NaCl; Red - 50 mM imidazole, pH 8.0, 150 mM NaCl; Green - 50 mM HEPES, pH 7.5, 150 mM NaCl; Light blue - 50 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM cNADP$^+$. **B)** A selection of thermal shift assay melting curves for *Bf* PNK/CPDase. Color code: Orange - 50 mM MOPS, pH 6.5, 500 mM NaCl; Blue - 50 mM MES, pH 7.0, 150 mM NaCl; Light blue - 50 mM HEPES, pH 7.5,150 mM NaCl; Red - 50 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM cNADP$^+$. A clear increase in the melting temperature can be observed from the graphs, induced by buffer conditions with pH 7.5, low salt concentration and cNADP$^+$.

The combination of 50 mM HEPES, pH 7.5 and 150 mM NaCl was identified as an optimal stabilizing buffer. Another buffer combination that was found to stabilize the proteins was 50 mM MES, pH 7.0, 150 mM NaCl. The buffers with high T_m values, as shown in Table 10, were later used in lysis and purification steps and found to

improve the stability of the proteins. In general, the T_m values obtained in the assay and stability during purification are both higher for yeast CPDase than the PNK/CPDase proteins. Similarly, the thermal stability of the catalytic domain of mouse CNPase is higher than the full-length enzyme (Myllykoski and Kursula 2010).

Table 10. Comparison of T_m values obtained for *Sc*Trl1 PNK/CPDase, *Sc*Trl1 CPDase and *Bf* PNK/CPDase under similar buffer conditions

Protein	Buffer condition	T_m (°C)
***Sc*Trl1 PNK/CPDase**	50 mM HEPES, pH 7.5, 150 mM NaCl	55
	50 mM imidazole, pH 8.0, 150 mM NaCl	50
	50 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM cNADP$^+$	**63**
***Sc*Trl1 CPDase**	50 mM HEPES, pH 7.5, 150 mM NaCl	58
	50 mM imidazole, pH 8.0, 150 mM NaCl	55
	50 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM cNADP$^+$	**72**
***Bf* PNK/CPDase**	50 mM HEPES, pH 7.5, 150 mM NaCl	43
	50 mM MES, pH 7.0, 500 150 mM NaCl	43
	50 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM cNADP$^+$	**48**

The highest T_m (°C) values are indicated in bold.

To analyze the effect of ligands on the thermal stability of the proteins, two known substrates of CPDase, 2',3'-cNADP$^+$ and 2',3'-cCMP, and a dinucleotide product, NADP$^+$ derived from 2',3'-cNADP$^+$, were included in the screen. The T_m values indicate that 2',3'-cNADP$^+$ improves the thermal stability of all three proteins tested [Figure 25, Table 10]; however, 2',3'-cCMP and NADP$^+$ do not influence the stability of any of them. The highest melting points obtained for *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase were 63°C and 48°C, respectively, both in the presence of the ligand 2',3'-cNADP$^+$. *Sc*Trl1 CPDase was found to be the most stable among the proteins tested, with a melting temperature of 72°C in the buffer containing 50 mM HEPES,

pH 7.5, 150 mM NaCl and 1 mM 2',3'-cNADP$^+$ [Table 10]. As a result of the thermal stability assay, 50 mM HEPES, pH 7.5 and 150 mM NaCl were identified and used as suitable components of the lysis and purification buffers.

The homologous mouse CNPase is more stable at pH 5.5 with higher salt than at neutral pH with lower salt, which stabilizes the yeast and lancelet proteins (Myllykoski and Kursula 2010). The thermal stability of *Sc*Trl1 CPDase is also higher than the wild type and mutants of the mouse CNPase catalytic domain (Myllykoski, Raasakka *et al*. 2013).

5.9. PNK/CPDase enzymes possess polynucleotide kinase activity

PNK catalyzes the transfer of the phosphate from γ position of ATP or other nucleoside triphosphates to the 5'-hydroxyl terminus of polynucleotides, such as double- and single-stranded DNA and RNA, to form a 5'-monophosphate polynucleotide and a nucleoside diphosphate; the enzyme also phosphorylates the 5-hydroxyl group of nucleoside 3'-monophosphates (Richardson 1965, Novogrodsky and Hurwitz 1966, Novogrodsky, Tal et al. 1966). The assay to test the PNK activity of the purified *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase involves transfer of ^{32}P label from [γ^{32}P] ATP to a 20-mer 5'-OH synthetic RNA oligonucleotide (rA_{20}) or DNA oligonucleotide (dA_{20}). To study the difference in preference of the enzymes on RNA and DNA oligonucleotide substrates, the transfer of the radio-labeled phosphate to a mixture of RNA and DNA synthetic oligonucleotides ($r+dA_{20}$) was also analyzed. The time-dependent transfer of the phosphate from [γ^{32}P] ATP to the 5'-OH groups of the synthetic oligonucleotides clearly indicates that both *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase possess PNK activity [Figures 26 A & B]. *Sc*Trl1 CPDase, used as a negative control, did not catalyze the phosphoryl transfer [Figure 26 C]. The results also indicate that *Sc*Trl1 PNK/CPDase prefers rA_{20} [Figure 26 A] and *Bf* PNK/CPDase strongly prefers dA_{20} [Figure 26 B].

In the process to identify the essential constituents of the PNK domain of yeast Trl1, the kinase activity of wild type yeast Trl1 and its P-loop mutants has been assayed earlier using either [γ^{32}P] ATP or [γ^{32}P] GTP and an 18-mer 5'-OH RNA oligonucleotide (Sawaya, Schwer *et al*. 2003). The single-residue mutations K404A and T405A in the putative NTP binding P-loop (^{401}GCGKT^{405}) have no effect on the kinase activity of yeast Trl1 *in vivo*; the two mutants abolish ATP-dependent kinase activity, but preserve GTP-dependent kinase activity (Sawaya, Schwer et al. 2003). However, the double alanine mutant in the P-loop (K404A+T405A) eliminates GTP-dependent kinase activity and is lethal *in vivo* (Sawaya, Schwer et al. 2003). Single alanine mutations at the equivalent P-loop residues of T4 PNK, Lys15 and Ser16 abolish the kinase activity (Wang and Shuman 2001).

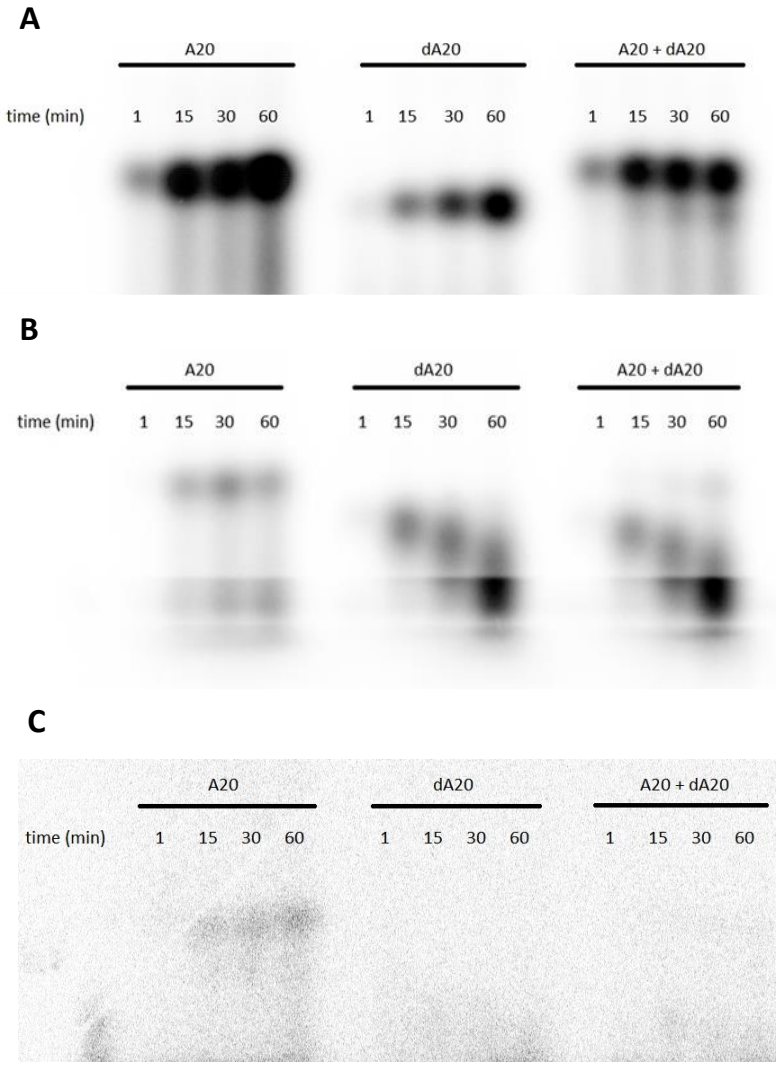


Figure 26. Polynucleotide kinase activity assay. PNK reaction mixtures of **A)** *Sc*Trl1 PNK/CPDase, **B)** *Bf* PNK/CPDase and **C)** *Sc*Trl1 CPDase. A20 - synthetic RNA oligonucleotide [rA_{20}]; dA20 - synthetic DNA oligonucleotide [dA_{20}] and A20+dA20 - a mixture of both RNA and DNA oligonucleotides [rA_{20}+dA_{20}]. The DNA or RNA or a mixture of both oligonucleotides were incubated with [γ^{32}P] ATP and the indicated enzyme. 8 µl aliquots were taken after 1, 15, 30 and 60 min, quenched with a urea dye mixture, and loaded on a 15% TBE-Urea gel. The radiolabeled nucleic acids were visualized by PhosphorImager analysis. The prominent background in Fig. C is caused by automated exposure settings executed by the imaging software. The seemingly little activity in the A20 reaction seen on the gel (Fig. C) was not present as the gel was checked with a Geiger-Müller counter.

The mutational studies have addressed the possibility of specific contacts between yeast Trl1 and GTP, which might not exist for ATP, that could compensate for the loss of either Lys404 or Thr405, but not for the loss of both residues (Sawaya, Schwer *et al*. 2003). The same is evidenced by the K_m of yeast Trl1 PNK for ATP, which is

about 800-fold greater than the K_m for GTP (Belford, Westaway *et al*. 1993, Westaway, Belford *et al*. 1993). Thus, it has been suggested that the *in vivo* physiological substrate is GTP and that the kinase domain of yeast Trl1 possesses a single NTP binding site containing the P-loop.

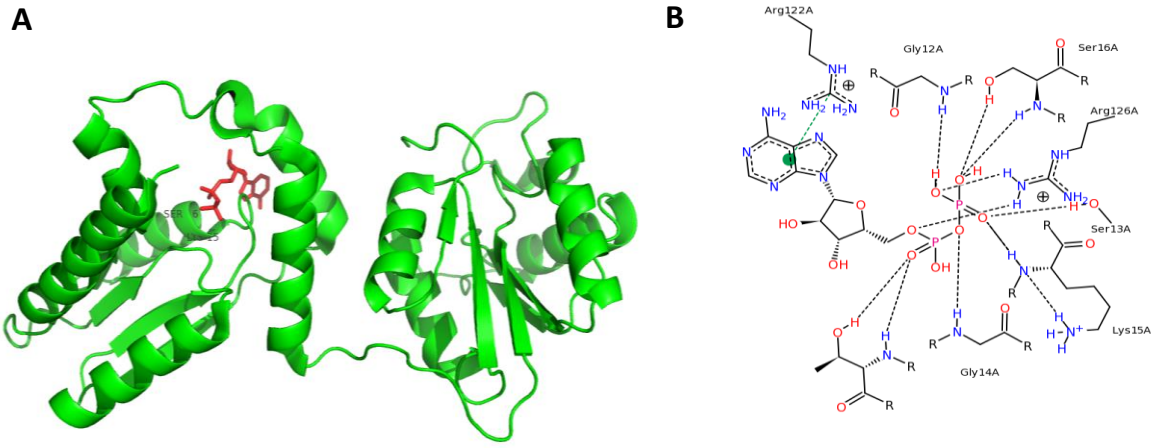


Figure 27. The structure of T4 polynucleotide kinase. A) A monomer of T4 polynucleotide kinase with both kinase and phosphatase domains is shown as a ribbon diagram, with an ADP molecule (red stick) bound to the kinase domain [PDB ID: 1LTQ] (Galburt, Pelletier et al. 2002). The figure was generated using PyMOL. **B)** A map of interactions at the kinase active site shows the residues involved in nucleotide binding. Black-dashed lines: hydrogen bonds, salt bridges and metal interactions, green solid lines: hydrophobic interactions and green dashed lines: π-cation interaction.

Since the PNK domain of yeast Trl1 has been shown to function similarly to the T4 polynucleotide kinase (Wang and Shuman 2002, Wang, Schwer *et al*. 2006), understanding the active site of the T4 PNK is important to further probe into the structure and activity relationships of yeast Trl1. The crystal structure of T4 polynucleotide kinase with bound ADP [Figures 27 A & B] describes the active site of the kinase domain (Galburt, Pelletier *et al*. 2002). The kinase active site in the nucleotide-bound PNK complex consists of a tunnel formed through contacts between two enzyme surface loops that form a lid over the phosphate tail of bound ADP (Galburt, Pelletier *et al*. 2002). The P-loop (^{12}GSGKS^{16}) residues of T4 PNK, Lys15 and Ser16 constitute the surface of the tunnel; ADP binds on the side of the tunnel that opens into a channel in the middle of the enzyme tetramer (Galburt, Pelletier *et al*. 2002).

The "R-x-x-x-R" motif (^{122}RxxxR^{126}) of T4 PNK is located within an alpha helix and loop that form a 'lid' over the NTP binding pocket (Wang, Lima et al. 2002).The distal arginine of this motif (Arg126) interacts with the α and β phosphates of the

nucleotide substrate. The nucleotide binding pocket is formed through electrostatic interactions between the diphosphate tail of ADP and the residues of the P-loop motif (Lys15 and Ser16) and the "R-x-x-x-R" motif (Arg126) in the kinase domain (Galburt, Pelletier *et al*. 2002). The proximal arginine (Arg122) forms a π-cation stack on the purine base of the nucleotide substrate [Figure 27 B] (Galburt, Pelletier et al. 2002). However, the mutation of arginine 122 to alanine has no effect on kinase activity (Wang and Shuman 2002). Similarly, the equivalent mutation of Arg507 in yeast Trl1 does not affect the PNK activity (Wang, Schwer et al. 2006). The P-loop component of T4 PNK joins a β-strand to an α-helix and forms an oxyanion hole for interactions with the β phosphate of the NTP substrate (Wang, Lima *et al*. 2002). The binding involves multiple contacts of the phosphate oxygens to the backbone amides of the P-loop, in addition to interactions with lysine and serine side chains (Wang, Lima *et al*. 2002). Hence, it is possible that the P-loop motifs of yeast Trl1 and lancelet PNK/CPDase are required for NTP binding. The P-loop motif of the N-terminal PNK-like domain of mouse CNPase could also be crucial for nucleotide binding, as evidenced by the requirement of the N-terminal domain for RNA binding *in vitro* (Myllykoski, Raasakka et al. 2012).

Bf PNK/CPDase has been previously shown to prefer DNA over RNA, whereas another lancelet PNK, *Bf* Clp1, has been reported to act exclusively on RNA (Englert, Sheppard *et al*. 2010). This finding has suggested a possible role for *Bf* PNK/CPDase in DNA repair; however, a role of *Bf* PNK/CPDase as a DNA-healing enzyme has not yet been proven.

The homologue of yeast and lancelet PNK/CPDase, mammalian CNPase, has been found to rescue a yeast strain with an inactivating mutation in the CPDase domain of yeast tRNA ligase, but not a strain with a mutated kinase domain (Schwer, Aronova *et al*. 2008). It indicates that the N-terminal PNK-like domain of vertebrate CNPase is not functionally a PNK. The same is evidenced by our recent results from a PNK activity assay, in which the full-length mouse CNPase was found to be inactive (Arne Raasakka and Petri Kursula, unpublished results), while the homologous yeast and lancelet PNK/CPDase proteins were active [Figures 26 A & B]. This finding also supports the existing hypothesis that vertebrates might employ the 5'-P RNA ligation pathway only as an alternate to 3'-P RNA ligation and thus require only the exclusive polynucleotide kinase (Clp1) for the healing reaction, and not the N-terminal PNK-like domain of vertebrate CNPase.

5.10. PNK/CPDase enzymes interact with nucleotide analogues

Since *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase contain the NTP-binding P-loop motif in the N-terminal PNK domain, their interaction with different fluorescently labeled nucleotide analogues (mant-ATP, mant-GTP and mant-AMP) was studied. The FRET-based assay used in this study relies upon the non-radiative transfer of energy between UV light-excited tryptophans within the protein under investigation and a nearby fluorophore (mant-NTP in this case) having overlapping energy spectra. The energy transfer is sensitive to the distance between the two fluorophores.

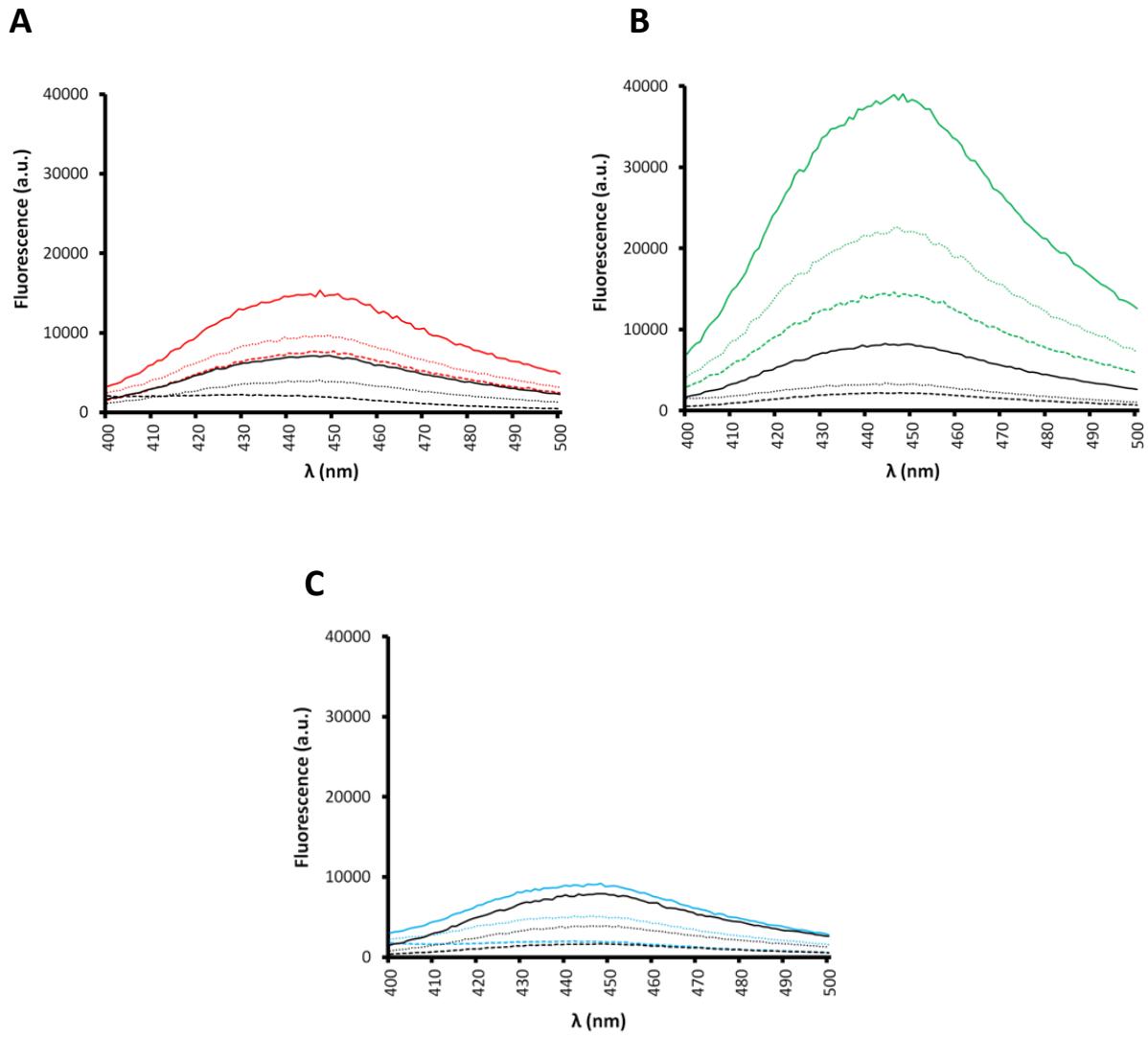


Figure 28. Interactions between *Sc*Trl1 PNK/CPDase and mant-labeled nucleotides studied using FRET. Interaction of *Sc*Trl1 PNK/CPDase with **A)** mant-ATP, **B)** mant-GTP and **(C)** mant-AMP. The curves correspond to different concentrations of mant-labeled nucleotides (solid line: 25 µM, round-dotted line: 10 µM and square-dotted line: 5 µM). The signal from the complex is given in red (ATP), green (GTP), blue (AMP) and the signal from the corresponding mant-labeled nucleotide alone is given in black.

In this assay, upon excitation at 280 nm (λ_{ex}=280 nm), the fluorescence energy from a donor tryptophan is used to excite the mant fluorophore resulting in an increase in mant fluorescence (λ_{em}=440 nm), which indicates the protein's close vicinity to the bound nucleotide. The curves shown in Figure 28 represent the mant-fluorescence detected for the protein-nucleotide complex and the signal from the mant-nucleotide alone. For the protein-ATP and protein-GTP complexes, the difference between the fluorescence from the complex and that from the mant-nucleotide alone was high, whereas the same for the protein-AMP complex was negligibly low [Figure 28]. By using mant-labeled nucleotides, it became clear that while mant-ATP and mant-GTP interact with both *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase [Figures 28 A & B], mant-AMP does not [Figure 28 C]. The behavior of yeast and lancelet PNK/CPDase proteins in this assay resembles the binding of fluorescent nucleotide analogues by myelin basic protein, one of the most abundant proteins in the myelin sheath of vertebrate nervous system (Wang, Neugebauer et al. 2011).

The homologue, vertebrate CNPase, has also been found to interact with guanosine or adenosine triphosphates as well as to possess purine triphosphatase activity (Stingo, Masullo *et al*. 2007). Molecular modeling and docking have revealed that the binding of the two nucleotides is similar to that observed in other P-loop containing proteins (Stingo, Masullo *et al*. 2007). Docking simulations with either GTP or ATP on the model of the N-terminal domain of CNPase have shown that the phosphate and sugar groups of the purine nucleotide triphosphates adopt identical canonical conformations with the phosphate groups binding to the P-loop (Stingo, Masullo *et al*. 2007). This is yet another indication that the P-loop residues are involved in nucleotide binding.

5.11. Conformational changes upon nucleotide binding of PNK/CPDase enzymes

Since both *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase were found to interact with GTP and 2',3'-cNADP$^+$, SRCD spectroscopy was applied to examine whether the binding of these ligands alter the conformation of the proteins. Two different concentrations of the ligands (500 µM and 1 mM) were used. Visual comparison of the conformation-driven features of the SRCD spectra from proteins alone and from the complex indicate that the conformation of both *Bf* PNK/CPDase and *Sc*Trl1 PNK/CPDase is altered upon binding to different concentrations of the nucleotide substrates [Figure 29 A and B]. Deconvolution of the spectra from *Sc*Trl1 PNK/CPDase and ligand complexes show that the addition of cNADP$^+$ and GTP significantly reduced the α-helical content, and increased the amount of β-strands [Table 11]. Conversely, the results indicate that the amount of β-strands of *Bf* PNK/CPDase is almost three times lower upon binding to cNADP$^+$, while the amount α-helices declined about 10% [Table 11]. The enzyme becomes predominantly α-helical through this conformational transition. The spectra show that the broad

negative peak at 218 nm, which is a characteristic feature of β-strands, is more prominent for *Bf* PNK/CPDase than for its complexes with cNADP$^+$, indicating the reduction of β-strands in the latter; the shift of the negative peak at 208 nm (in protein alone) towards 205 nm (in the presence of ligands) shows a decline in the α-helical content [Figure 29 B].

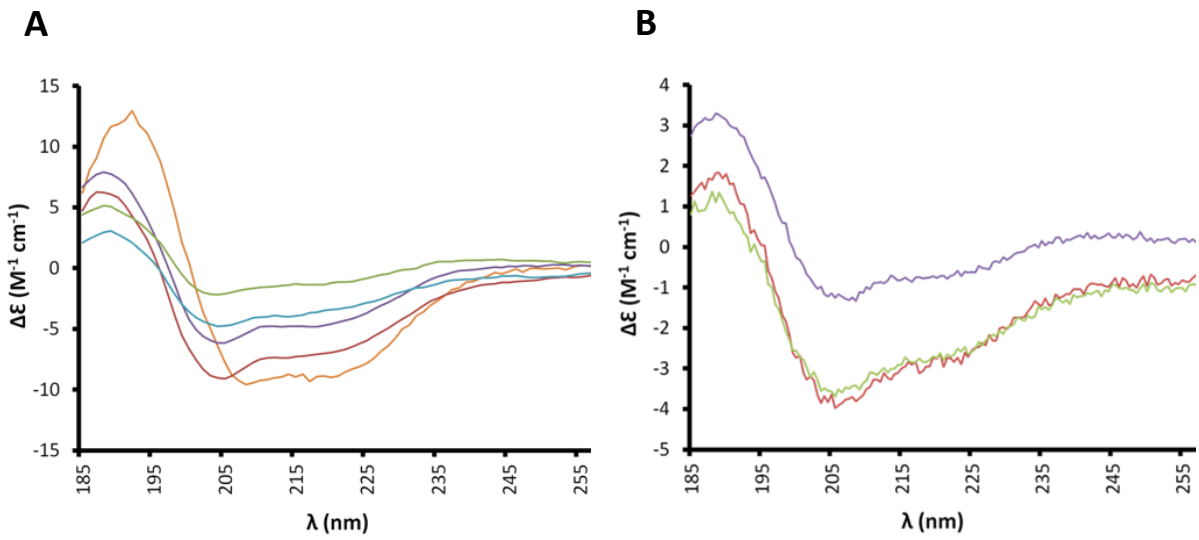


Figure 29. Conformational changes upon nucleotide binding of *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase as probed by SRCD spectroscopy. **A)** SRCD spectra from *Sc*Trl1 PNK/CPDase (orange), *Sc*Trl1 PNK/CPDase + 500 µM cNADP$^+$ (purple), *Sc*Trl1 PNK/CPDase + 1 mM cNADP$^+$ (red), *Sc*Trl1 PNK/CPDase + 500 µM GTP (green) and *Sc*Trl1 PNK/CPDase + 1 mM GTP (blue). **B)** SRCD spectra from *Bf* PNK/CPDase (blue), *Bf* PNK/CPDase+500 µM cNADP$^+$ (red) and *Bf* PNK/CPDase + 1 mM cNADP$^+$ (green).

However, the spectra from protein-ligand complexes show similar shapes, despite the differences in the molar ellipticity, which could also arise from inaccurate determination of the protein concentration, especially when it is bound to nucleotides that show strong UV absorbance and CD signal. Even formation of small bubbles in the sample during the measurement could lead to such differences. The presence of nucleotides in the buffers used as reference might also explain the shift in the baseline at higher wavelengths. The results of this experiment have to be verified by high-resolution structural methods prior to drawing conclusions. Since the N-terminal PNK domain is involved in GTP binding, and the C-terminal CPDase domain is involved in binding cNADP$^+$, high-resolution structures of PNK/CPDase enzymes with bound active site ligands are required to further understand the conformational changes at the atomic level.

Table 11. Deconvolution of CD spectra

Protein ± Ligand	**α-helix** (%)	**β-sheet** (%)	**NRMSD** (%)
*Sc*Trl1 PNK/CPDase	56	25	0.002
*Sc*Trl1 PNK/CPDase + 500 µM cNADP	10	47	0.007
*Sc*Trl1 PNK/CPDase + 1 mM cNADP	14	45	0.002
*Sc*Trl1 PNK/CPDase + 500 µM GTP	17	49	0.038
*Sc*Trl1 PNK/CPDase + 1 mM GTP	8	34	0.028
Bf PNK/CPDase	63	21	0.023
Bf PNK/CPDase + 500 µM cNADP	44	8	0.020
Bf PNK/CPDase + 1 mM cNADP	44	9	0.018

The effect of GTP binding on the molecular properties of the homologue, mouse CNPase has been reported (Stingo, Masullo et al. 2007). Fluorescence experiments indicated a change in the microenvironment of certain aromatic residues of the protein, as observed through a significant quenching of the fluorescent signal at the two excitation wavelengths, 280 nm (fluorescence from all aromatic residues) and 295 nm (mainly tryptophan fluorescence) (Stingo, Masullo et al. 2007).

5.12. PNK/CPDase and CPDase enzymes possess cyclic nucleotide phosphodiesterase activity

CPDase catalyzes *in vitro* hydrolysis of 2',3'-cyclic phosphates to produce 2'-phosphates exclusively [Figure 30]. The enzyme also acts on other molecules that contain this cyclic moiety, such as 2',3'-cyclic mononucleotides, 2',3'-cyclic oligonucleotides and 2',3'-cyclic $NADP^+$ (Olafson, Drummond et al. 1969, Sogin 1976).

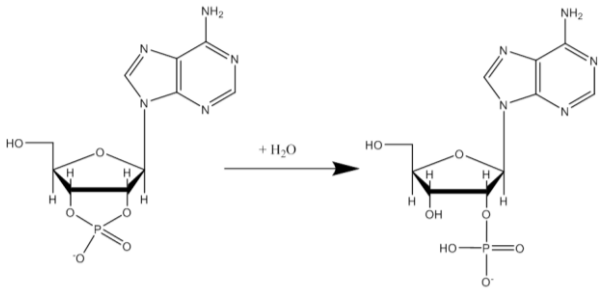


Figure 30. Reaction catalyzed by cyclic phosphodiesterase. CPDase catalyzes the hydrolysis of 2',3'-cyclic nucleotide to 2'-nucleotide.

The CPDase activity of *Sc*Trl1 PNK/CPDase, *Bf* PNK/CPDase and *Sc*Trl1 CPDase was assayed using 2',3'-cNADP$^+$ and 2',3'-cCMP as substrates. The CPDase activity of yeast Trl1 and lancelet PNK/CPDase on cNADP$^+$ was assayed using a coupled enzyme assay (Sogin 1976, Lee, Gravel et al. 2001). The activity measurements were carried out for the PNK/CPDase and CPDase domains of yeast Trl1, and lancelet PNK/CPDase.

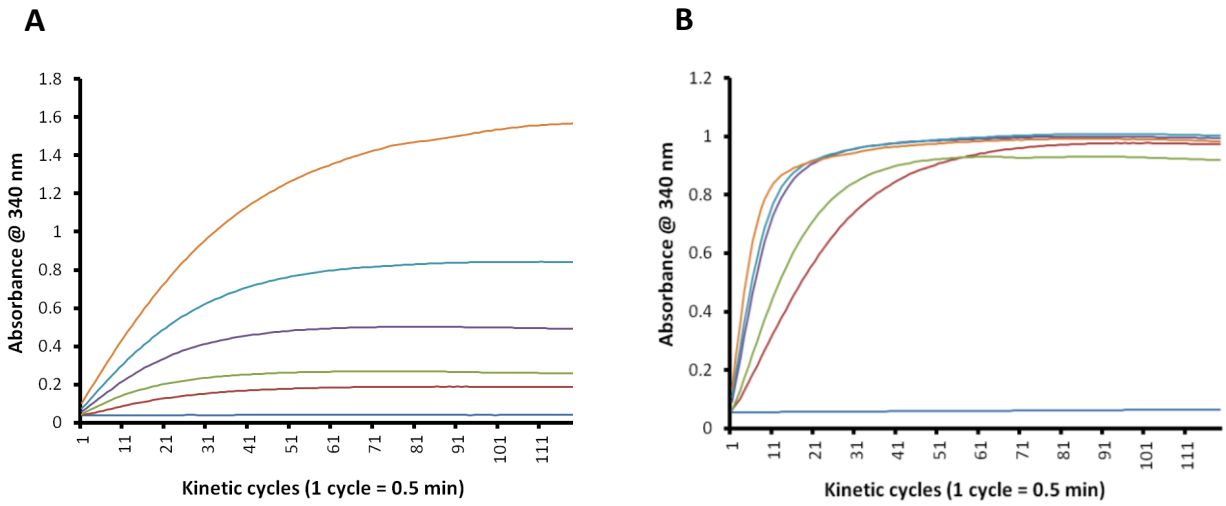


Figure 31. CPDase activity of *Bf* PNK/CPDase measured using a spectrophotometric coupled enzyme assay. Curves from absorbance measurements with **A)** different amounts of *Bf* PNK/CPDase in complex with 2 mM cNADP$^+$; blue - 0 ng, red - 100 ng, green - 250 ng, purple - 500 ng, light blue - 1000 ng and orange - 5000 ng of *Bf* PNK/CPDase. **B)** different concentrations of cNADP$^+$ in complex with 500 ng *Bf* PNK/CPDase; blue - 0 mM, red - 0.05 mM, green - 0.1 mM, purple - 0.2 mM, light blue - 0.5 mM and orange - 1 mM of 2',3'-cNADP$^+$.

Initially, the amount of enzyme to be used in the assay was estimated by performing the measurements with different amounts of *Bf* PNK/CPDase [Figure 31 A], *Sc*Trl1 PNK/CPDase, *Sc*Trl1 CPDase from 0 to 5000 ng, and a constant substrate concentration of 2 mM. 500 ng of the enzyme was found to be optimal based on the initial velocity of the reaction, and used for further measurements. Measurements for all the constructs were carried out with a series of substrate concentrations ranging from 0 to 1 mM [Figure 31 B]. The experiments were carried out in triplicate. The results clearly indicate that the all tested constructs are active in the phosphodiesterase reaction [Figures 32 A & B] and that the activity is comparable to that of the different constructs of mouse and rat CNPase enzymes [Table 12] (Lee, Gravel et al. 2001, Myllykoski and Kursula 2010). The K_m values range from 230 to 440 µM, and k_{cat} values from 275 to 510 s^{-1} [Table 12].

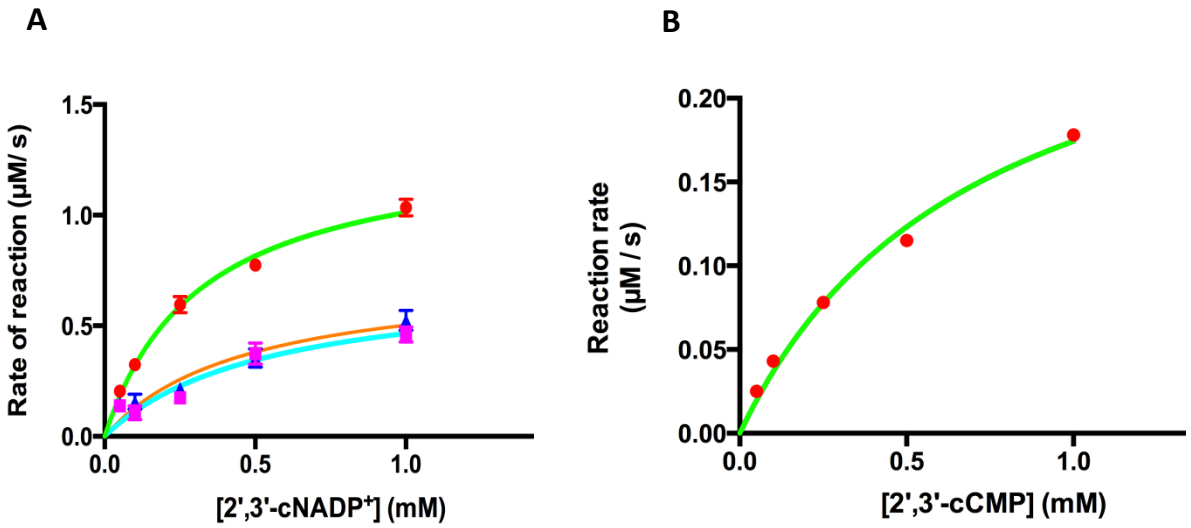


Figure 32. Michaelis-Menten curves from CPDase activity assays. A) Curves from enzymatic activity measurements of *Sc*Trl1 PNK/CPDase, *Sc*Trl1 CPDase and *Bf* PNK/CPDase with 2',3'-cNADP$^+$. The measurements were carried out in triplicate. For each enzyme, the mean value for each velocity point as well as standard error of the mean (*Sc*Trl1 PNK/CPDase - blue, *Sc*Trl CPDase - pink and *Bf* PNK/CPDase - red) is shown along with the respective Michaelis-Menten fit (*Sc*Trl1 PNK/CPDase - orange, *Sc*Trl CPDase - light blue and *Bf* PNK/CPDase - green). **B)** A Michaelis-Menten curve from a single measurement of the activity of ScTrl1 CPDase on 2',3'-cCMP as substrate. The red dots indicate the velocity at different substrate concentrations and the green curve shows the Michaelis-Menten fit.

The ability of *Sc*Trl1 PNK/CPDase, *Sc*Trl1 CPDase and *Bf* PNK/CPDase to catalyze the hydrolysis of the phosphodiester bond in 2',3'-cyclic CMP was assayed spectrophotometrically by measuring the amount of formed 2'-CMP through the change in absorbance at 286 nm. The steady increase in absorbance indicated that the enzymes are active, and that 2',3'-cCMP is a potential substrate. The measurements have to be carried out in duplicate or triplicate in order to obtain accurate kinetic parameters. The spectrophotometric method used in this assay has originally been applied to measure the conversion of 2',3'-cyclic cytidylate to 3'-monophosphate by pancreatic ribonuclease (Hugli, Bustin *et al*. 1973).

2',3'-cNADP$^+$ was identified as a substrate for CNPase in 1976 (Sogin 1976). A highly active diesterase (CNPase) that hydrolyses the ribonucleoside 2',3'-cyclic phosphates to the corresponding ribonucleoside 2'-phosphate has been prepared from bovine brain and found to be most active in the brain and spinal cord (Drummond, Iyer *et al*. 1962). The enzyme, later identified as the myelin-associated enzyme CNPase acts, *in vitro,* on substrates, such as 2',3'-cyclic mononucleotides and 2',3'-cyclic oligonucleotides, in addition to 2',3'-cNADP$^+$ (Olafson, Drummond *et al*. 1969). CNPase does not hydrolyze 3',5'-cyclic nucleotides or internucleotide bonds (Drummond, Iyer *et al*. 1962).

The mechanism of CPDase catalytic activity has been revealed by the crystal structures available for several members of the 2H-phosphoesterase superfamily, including human, rat and mouse CNPase, 2'-5' RNA ligase from *Thermus thermophilus* and CPDase from *Arabidopsis thaliana*. The structures of the plant CPDase with sulfate or cyclic vanadate inhibitor bound, the human CNPase with phosphate bound and the mouse CNPase bound to inhibitors, nucleotide substrates and products elucidate the possible catalytic roles of the "H-x-T-x" motifs and other functional groups at the active site (Hofmann, Zdanov et al. 2000, Kato, Shirouzu et al. 2003, Kozlov, Lee et al. 2003, Sakamoto, Tanaka et al. 2005, Myllykoski, Raasakka et al. 2012, Myllykoski, Raasakka et al. 2013). The structures show that the members of the "2H" phosphoesterase superfamily share a symmetrical bilobed tertiary structure. Each lobe consists of a three or four-stranded β-sheet at the lobar interface and α-helices on the periphery [Figure 33].

Table 12. Kinetic parameters obtained for the CPDase activity of different constructs on the substrate, cNADP$^+$

Protein	**K_m** (μM)	**k_{cat}** (s^{-1})	**k_{cat}/ K_m** (μM^{-1} s^{-1})
Enzymes used in the study			
Bf PNK/CPDase (500 ng)	440 ± 50	510 ± 20	1.15
*Sc*Trl1 PNK/CPDase (500 ng)	420 ± 35	317 ± 15	0.75
*Sc*Trl1 CPDase (500 ng)	237 ± 27	275 ± 18	1.16
Homologous enzymes			
Mouse CNPase - full length (500 ng)[1]	193±26	270±10	1.40
Mouse CNPase - catalytic domain (500 ng)[1]	445±41	570±19	1.28
Rat CNPase - full length[2]	263±12	836±11	3.2
Rat CNPase - catalytic fragment[2]	295±22	1690±39	5.7

[1](Myllykoski and Kursula 2010), [2](Lee, Gravel *et al*. 2001)

The catalytic mechanism proposed by Sakamoto *et al*. based on the crystal structure of human CNPase bound to phosphate resembles the second part of the well-characterized reaction mechanism of RNase A, in which the RNA 2',3'-cyclic end is processed to generate a 3'-phosphate (Raines 1998, Sakamoto, Tanaka *et al*. 2005). According to the proposed mechanism, the histidine of the distal "H-x-T-x" motif is activated by the carbonyl group of an adjacent arginine. The histidine deprotonates

and activates a water molecule, which carries out a nucleophilic attack on the phosphorous of the cyclic phosphate and forms a transient pentavalent intermediate. The histidine of the proximal "H-x-T-x" motif protonates the 3'-oxygen and releases the product. Mutational analysis of rat brain, zebrafish and yeast 2H family enzymes have indicated that both histidines are required for catalysis (Ballestero, Dybowski *et al*. 1999, Nasr and Filipowicz 2000, Lee, Gravel *et al*. 2001, Kozlov, Lee *et al*. 2003). However, only the distal histidine is strictly required for the CPDase activity of yeast and plant tRNA ligases (Wang, Schwer et al. 2006). Mutational analysis of brain CNPase has shown that both threonines are essential for the CPDase reaction, whereas the yeast and plant tRNA ligases require only the proximal threonine (Kozlov, Lee *et al*. 2003, Wang, Schwer *et al*. 2006). Thus, the differences in the contributions of the individual amino acids at the CPDase active site are apparent between the tRNA ligases and other 2H family members. Neither of the two hydroxyl amino acids of the "H-x-(S/)T-x" motifs of yeast CPDase are essential for its activity on the physiological substrate, Appr>p (Nasr and Filipowicz 2000). However, similarly to yeast and plant tRNA ligases, the plant CPDase requires the proximal threonine for its activity on 2',3'-cAMP (Nasr and Filipowicz 2000) .

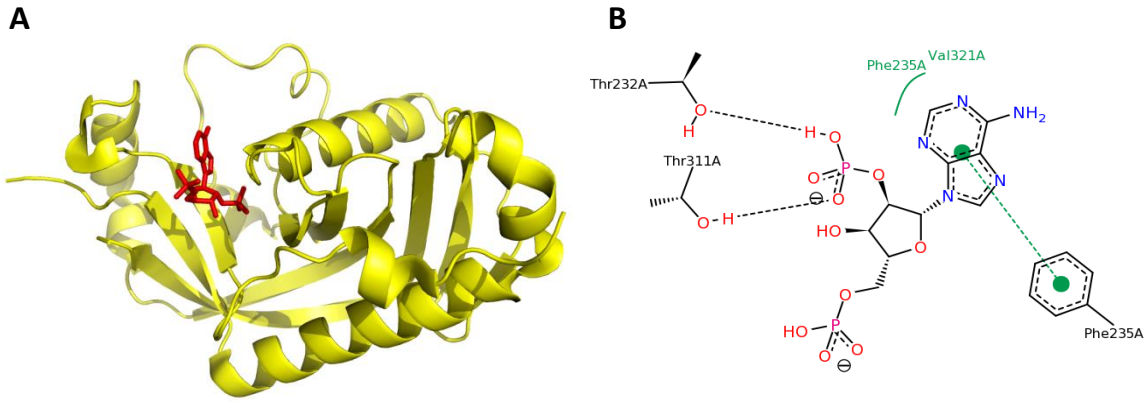


Figure 33. Structure of the CNPase active site. A) An overall view of the active site of mouse CNPase (cartoon; yellow) bound to $NADP^+$ (stick; red) [PDB ID: 2YDB] (Myllykoski, Raasakka et al. 2012). The figure was generated using PyMOL. **B)** Interaction map of the CNPase active site showing the active-site residues. Black dashed lines: hydrogen bonds, salt bridges and metal interactions, green solid lines: hydrophobic interactions and green dashed lines: π- π interaction.

The binding mode of nucleotide ligands at the active site of mouse CNPase has been elucidated in the recent crystal structure of the enzyme determined in the presence of a 2'-nucleotide product in the active site, after soaking the crystals with the *in vitro* substrates 2',3'-$cNADP^+$ and 2',3'-cAMP (Myllykoski, Raasakka *et al*. 2012). This is the first high-resolution structure available of a CNPase-nucleotide complex, in which the 2'-AMP resulting from the substrate, 2',3'-cAMP, is tightly bound in the active-

site cavity [Figure 34]. The reaction mechanism is similar to the previously proposed mechanism of human CNPase (Sakamoto, Tanaka et al. 2005). The core structural features of the mouse CNPase resemble those of the human enzyme and gold fish RICH (Sakamoto, Tanaka et al. 2005, Kozlov, Denisov et al. 2007, Myllykoski, Raasakka et al. 2012).

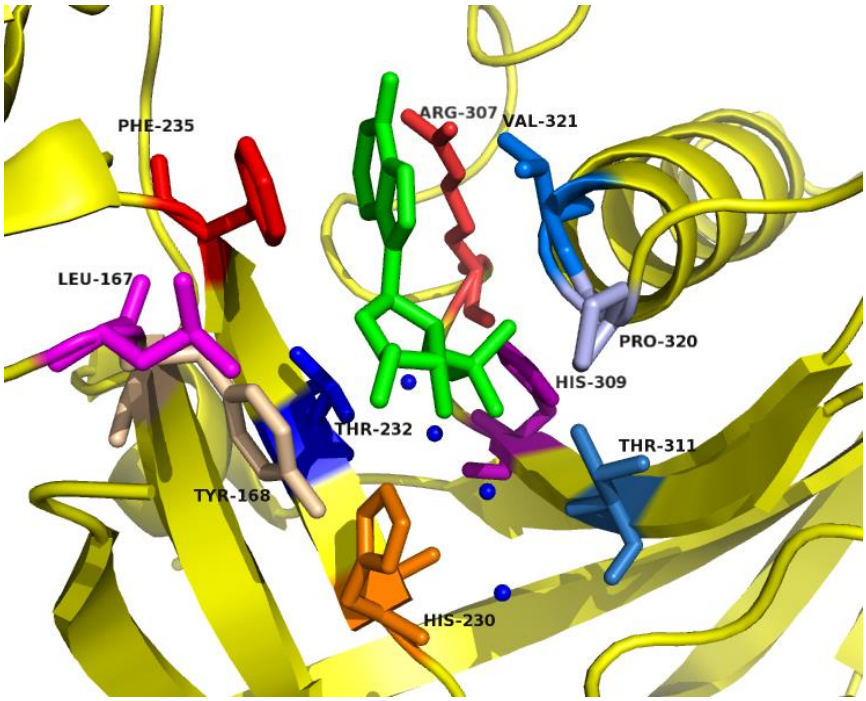


Figure 34. A detailed view of the mouse CNPase active site with 2'-AMP (shown as green stick) bound [PDB ID: 2YDD] (Myllykoski, Raasakka et al. 2012). Side chains of the active site residues that participate in the enzymatic reaction or envelop the active site are displayed in different colors and labeled. The four ordered water molecules at the bottom of the active site are also displayed. The figure was generated using PyMOL.

The crystal structure of the first substrate complex of an enzyme in the 2H family has recently been determined for a mutant of the mouse CNPase catalytic domain (Myllykoski, Raasakka *et al*. 2013). The 2',3'-cAMP substrate has been co-crystallized with the H309S mutant of the mouse CNPase catalytic domain, and the position of the substrate and reactive water molecule in the active site is well-defined; the position of the cyclic nucleotide is in agreement with the previously proposed reaction mechanisms (Sakamoto, Tanaka et al. 2005, Myllykoski, Raasakka et al. 2012, Myllykoski, Raasakka et al. 2013). The structure presents a more detailed mechanism for the catalytic activity of CNPase. To initiate the phosphodiesterase reaction, a 2',3'-cyclic nucleotide substrate enters the active site, but does not induce large-scale conformational changes in the active site. The reactive cyclic phosphate group is positioned at the bottom of the active site by the two "H-x-T-x" motifs and

two water molecules. The optimal orientation of the nucleophilic water molecule is coordinated by hydrogen bonds. The reaction starts with the removal of a proton from the nucleophilic water molecule by the side chain of the distal histidine residue.

The activated water molecule attacks the cyclic phosphate group and then forms a pentavalent intermediate. In the second part of the reaction, the pentavalent intermediate is dissociated and a 2'-phosphate group is formed. Then the proximal histidine residue protonates the 3'-oxygen and the product is released. The active site is regenerated by reversal of the protonation sites of the active side histidine side chain for a new reaction to occur. Despite the available knowledge on the mechanism of action of vertebrate CNPase, questions concerning substrate specificity and conformational changes upon substrate-binding of yeast and lancelet PNK/CPDase enzymes need to be answered by high-resolution structures of the enzymes in the presence of the substrates.

5.13. PNK/CPDase enzymes form dimers in solution

SAXS measurements were carried out to determine the low-resolution three-dimensional shape of the purified proteins and to analyze their oligomeric state in solution. Three different concentrations ranging from 1 to 8 mg/ml of each protein were measured. The best datasets were collected between 3 and 4 mg/ml, with strong scattering signal and no aggregation or repulsion. Although SAXS data were collected for all the five expressed constructs, only three of them, the *Sc*Trl1 PNK/CPDase, the *Sc*Trl1 CPDase and the *Bf* PNK/CPDase yielded good-quality data and only those were used for further processing and modeling.

The dimeric and monomeric forms of *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase resulting from size exclusion chromatography were used in SAXS measurements separately [Figures 35-38]. The sizes of the molecules obtained from SAXS data indicated that both proteins exist in both monomeric and dimeric forms. The results resemble the oligomeric state of the archaeal translation initiation factor (aIF5B) determined by SAXS (Rasmussen, Oliveira *et al*. 2011). The dimeric form of *Sc*Trl1 PNK/CPDase is an elongated molecule with a maximum dimension of 27.35 nm and a radius of gyration of 5.3 nm [Figure 36]. The determined molecular mass based on comparison of the *Sc*Trl1 PNK/CPDase forward scattering intensity [I(0)] to that of the standard, bovine serum albumin was 104 kDa, very close to the expected molecular mass of a dimer (101 kDa) indicating a dimeric state for *Sc*Trl1 PNK/CPDase in solution. The estimated Porod volume [Table 13] is also in line with the presence of dimeric *Sc*Trl1 PNK/CPDase.

The dimeric *Bf* PNK/CPDase displays a solution structural behavior similar to that of the *Sc*Trl1 PNK/CPDase dimer. The molecule is elongated with a maximum dimension of 24.63 nm and a radius of gyration of 6.99 nm [Figure 38]. The molecular mass determined was 118 kDa, while the expected molar mass of a dimer is 116 kDa. Thus, the SAXS data reveals dimerization of *Bf* PNK/CPDase. The monomeric *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase display similarly elongated conformation with the values determined of maximum dimension of the particle and radius of gyration being approximately half of those determined for the dimeric forms [Figures 35 & 37, Table 13]. The estimated Porod volume is also in agreement with a monomeric form of the proteins.

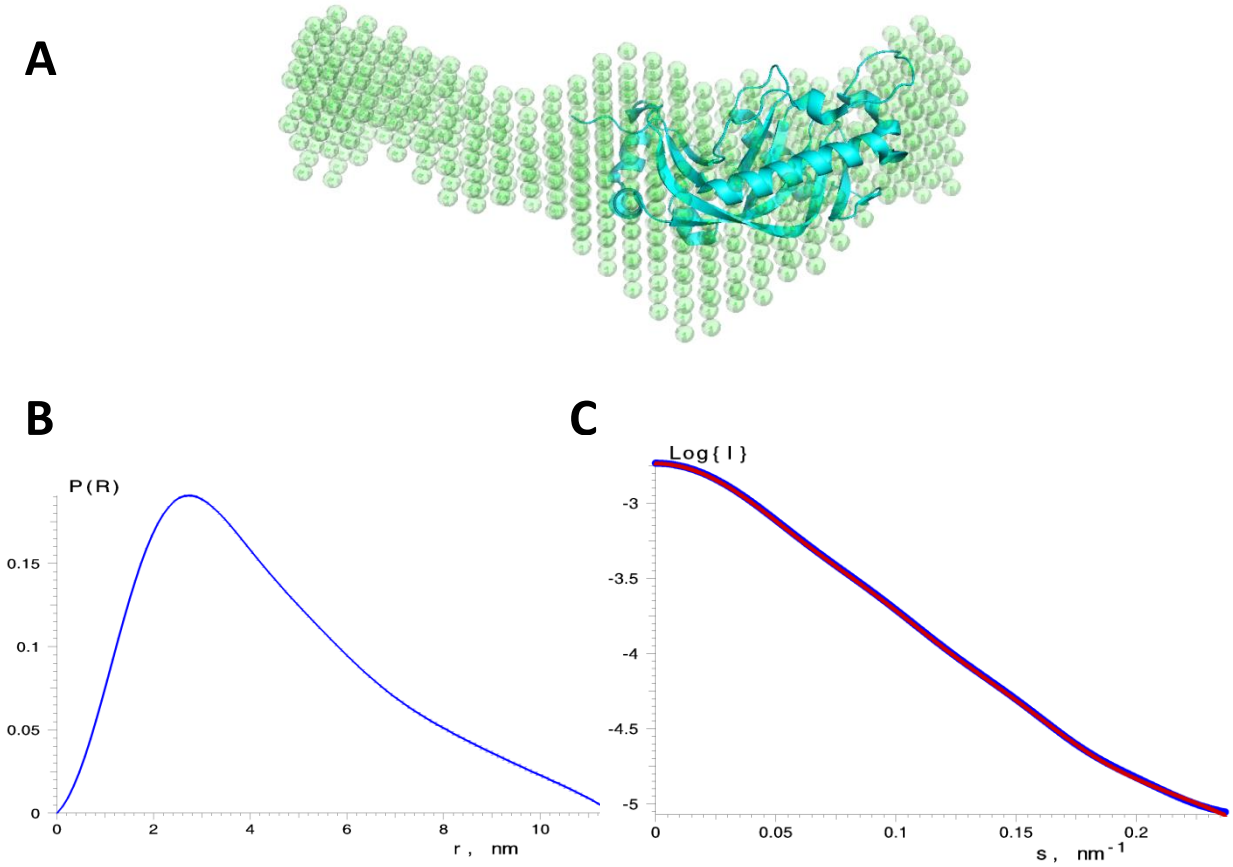


Figure 35. SAXS structure of *Sc*Trl1 PNK/CPDase monomer. A) Superposition of the *ab initio* SAXS model (green) of the *Sc*Trl1 PNK/CPDase monomer and the high-resolution crystal structure (blue) of the mouse CNPase catalytic domain. **B)** Distance-distribution function of the *Sc*Trl1 PNK/CPDase monomer. **C)** Scattering curve of the monomeric *Sc*Trl1 PNK/CPDase. The raw data are indicated as blue dots, and the curve in red color represents the fit of the *ab initio* model to the raw data.

The GNOM analysis computes the distance distribution function [*p*(r)] that provides direct information about the distances between electrons in the scattering molecules. The shapes of the curves already suggest an elongated conformation for both *Sc*Trl1

PNK/CPDase and *Bf* PNK/CPDase. The *p*(r) profile displays a single peak with two shoulders for both monomeric and dimeric fractions of the PNK/CPDase proteins [Figures 35-38], indicating an asymmetric scattering particle possibly containing two domains bridged by a flexible linker.

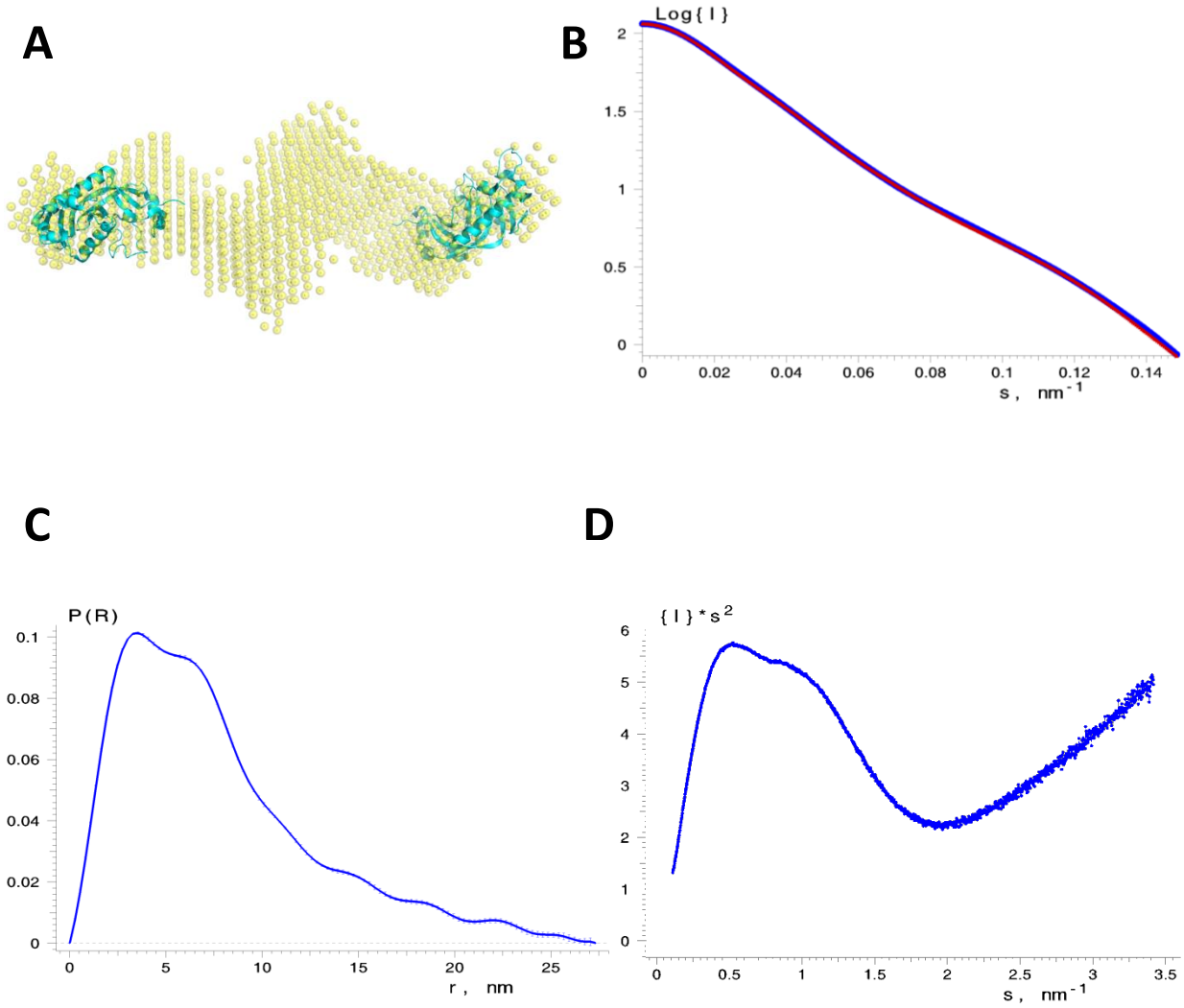


Figure 36. SAXS structure of *Sc*Trl1 PNK/CPDase dimer. A) Superposition of the *ab initio* SAXS model (yellow) of the *Sc*Trl1 PNK/CPDase dimer and the high-resolution crystal structure (blue) of the mouse CNPase catalytic domain. **B)** Scattering curve of the dimeric *Sc*Trl1 PNK/CPDase. The raw data are indicated in blue, and the curve in red represents the fit of the *ab initio* model to the raw data. **C)** Distance-distribution function of the *Sc*Trl1 PNK/CPDase dimer. **D)** The Kratky plot displaying a peak characteristic of folded, multi-domain proteins.

The initial processing by GNOM was followed by *ab initio* modeling using DAMMIF (Franke and Svergun 2009). This program represents the particle as a collection of densely packed beads within an unrestricted search volume (Franke and Svergun

2009). Initially, models were built without any symmetrical constraint and later a two-fold symmetry was applied for the molecules that had a molecular mass of a dimer. Both kinds of procedures resulted in similar elongated models. The models fit to the raw data well. To evaluate the accuracy of the reconstructed model, DAMAVER was applied. DAMAVER aligns all reconstructed models and removes outliers below a given cutoff volume. The model suggests that the active sites of the PNK domain and CPDase domain are accessible and not being wrapped by each other. The open active site might facilitate substrate binding without any large conformational changes. If the domains would cover each other, the shape of the molecule would be more globular.

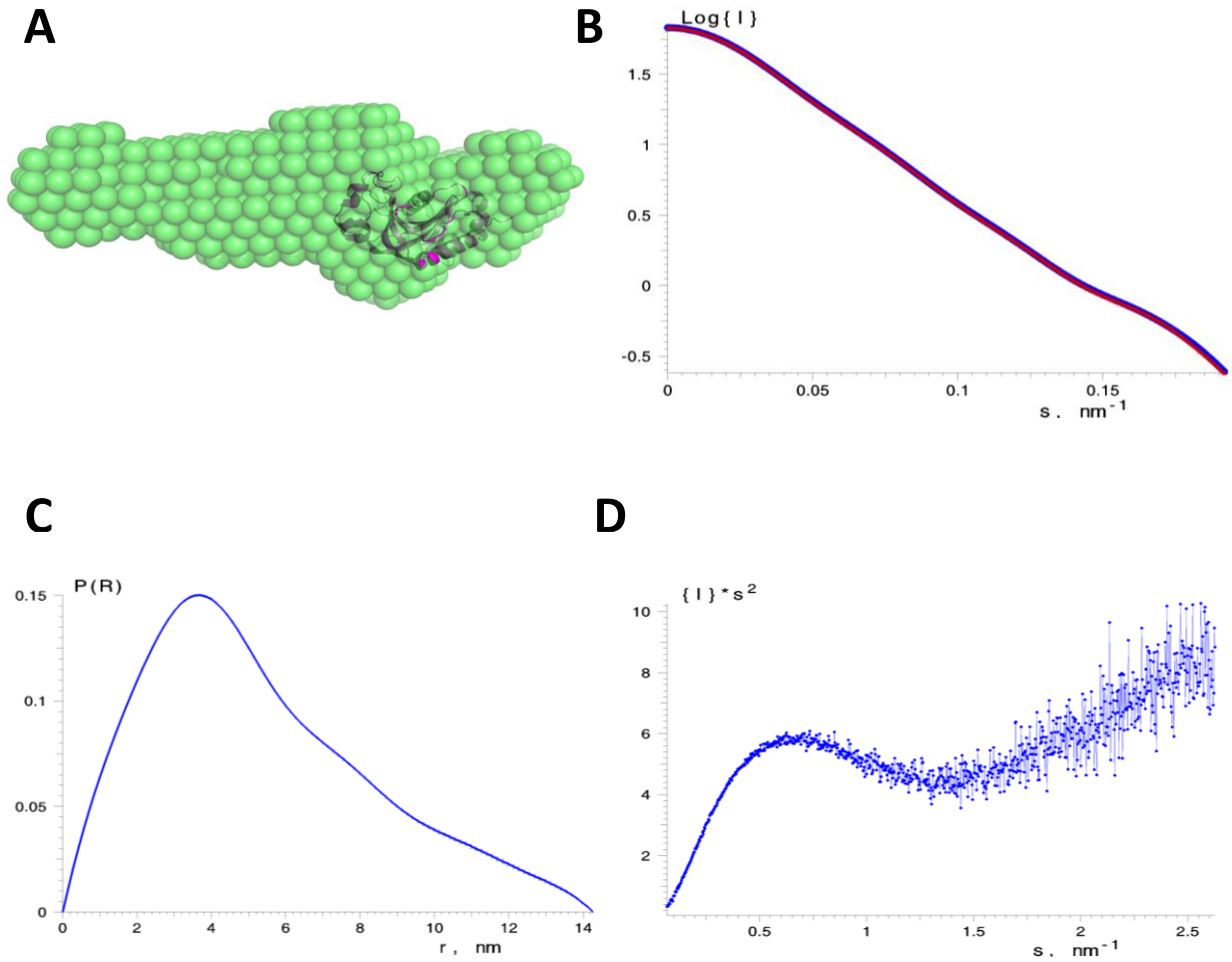


Figure 37. SAXS structure of *Bf* PNK/CPDase monomer. A) Superposition of the *ab initio* SAXS model (green) of the *Bf* PNK/CPDase monomer and the high-resolution crystal structure (pink) of the mouse CNPase catalytic domain. **B)** Scattering curve of the monomeric *Bf* PNK/CPDase. The raw data are indicated in blue, and the curve in red represents the fit of the *ab initio* model to the raw data. **C)** Distance-distribution function of the *Bf* PNK/CPDase monomer. **D)** The Kratky plot displaying a peak characteristic of a multi-domain protein.

The solution structure of the homologue, mouse CNPase, is similar to that of the monomeric *Sc*Trl1 PNK/CPDase and *Bf* PNK/CPDase (Myllykoski, Raasakka *et al.*

2013). The full-length mouse CNPase is a monomer with an elongated conformation in solution (Myllykoski, Raasakka *et al*. 2013). Although the molecular masses of monomeric *Sc*Trl1 PNK/CPDase [50.52 kDa] and monomeric mouse CNPase [44.8 kDa] are close to each other, the D_{max} and R_g values are comparatively higher for the *Sc*Trl1 PNK/CPDase. It could be due to the presence of the flexible linker region between the N-terminal PNK and C-terminal CPDase domains of *Sc*Trl1 PNK/CPDase: this segment is absent in mouse CNPase. It has been suggested that the C-terminal membrane-anchoring tail of the mouse CNPase could be present in the middle region of the molecule to enable the association of the active site to the lipid bilayer (Myllykoski, Raasakka *et al*. 2013).

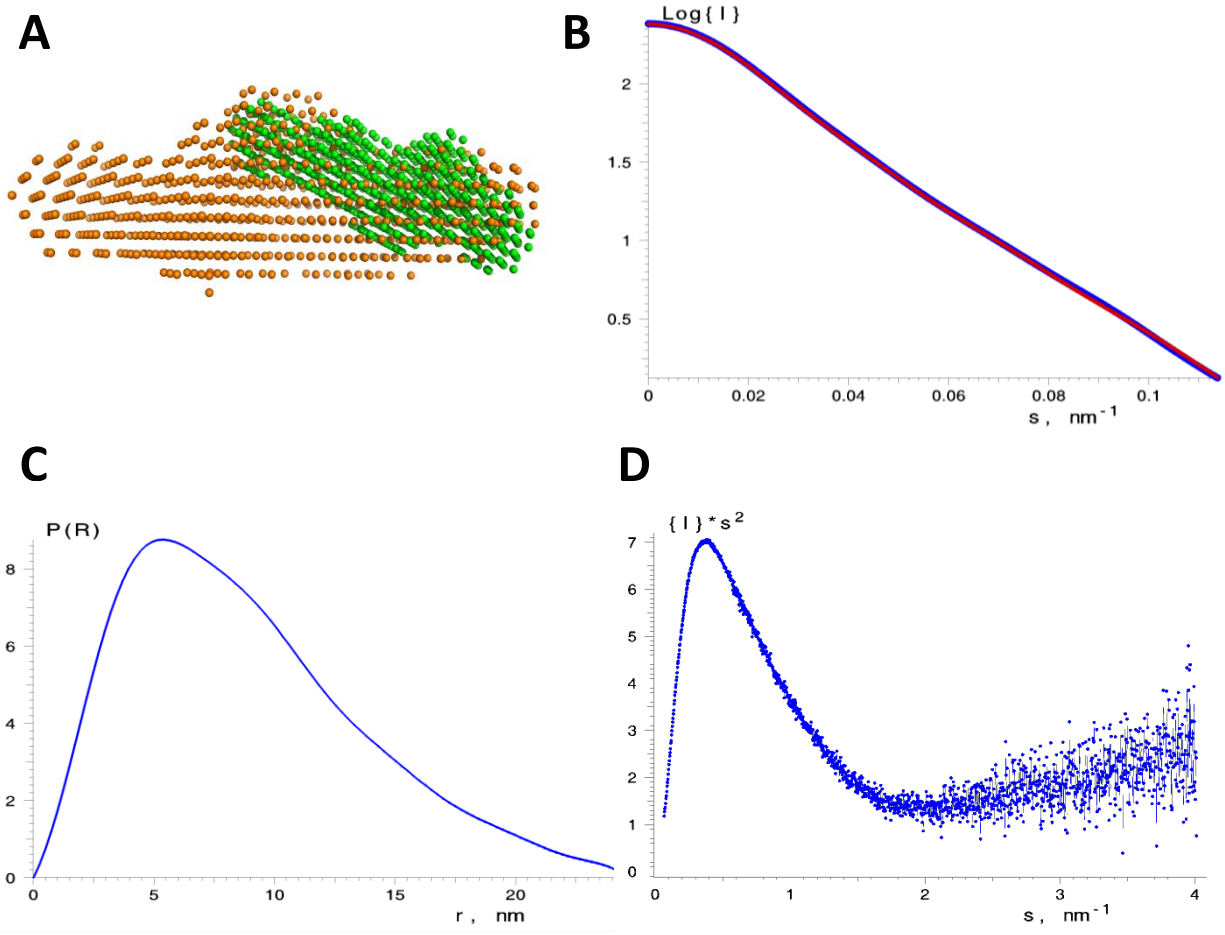


Figure 38. SAXS structure of *Bf* PNK/CPDase dimer. A) Superposition of the *ab initio* models of the dimeric (orange) and monomeric (green) *Bf* PNK/CPDase. **B)** Scattering curve of the dimeric *Bf* PNK/CPDase. The raw data are indicated as blue dots, and the curve in red color represents the fit of the *ab initio* model to the raw data. **C)** Distance-distribution function of the *Bf* PNK/CPDase dimer. **D)** The Kratky plot displaying a peak characteristic of a folded multi-domain protein.

The Kratky plot indicates the folding state of proteins. In a Kratky plot, a prominent peak at low angles is indicative of folded globular proteins, whereas a continuous increase in s^2 I(s) indicates the unfolded state of proteins. However, multi-domain proteins with flexible linkers often display a Kratky plot that is a mixture of the characteristic features of both folded and unfolded proteins and the same is detected for a partially unfolded state. Thus, the Kratky plots of PNK/CPDase proteins displayed peaks characteristic of partially unfolded proteins [Figures 36D, 37D, 38D].

5.14. CPDase is a monomer in solution

The molecular size determined from the SAXS profile of *Sc*Trl1 CPDase indicates that the molecule is more compact in solution than *Sc*Trl1 PNK/CPDase and exists as a monomer [Figure 39]. The maximum dimension of the molecule is 8.51 nm, and the radius of gyration is 2.48 nm. Molecular mass determination yields a mass of 32 kDa, which is close to the theoretical molecular mass (30 kDa). The Porod volume is also in line with the presence of monomeric *Sc*Trl1 CPDase. The Kratky plot indicates that the protein is folded [Figure 39D]. The result indicates that the PNK domain is required for dimerization. Comparison of the SAXS models of PNK/CPDase and CPDase also suggest that the PNK domain should lie in the middle of the dimer. The SAXS model shows the presence of an extension presumably in the N-terminus of the protein; this tail possibly arises from the insertion in the N-terminus of the CPDase domain, suggesting the requirement for a shorter construct of the protein without the N-terminal insertion to facilitate crystallization.

The solution structure of *Sc*Trl1 CPDase resembles that of the homologue, the catalytic domain of mouse CNPase (Myllykoski, Raasakka *et al*. 2012). The mouse CNPase catalytic domain is also monomeric in solution. Both *Sc*Trl1 CPDase and the catalytic domain of mouse CNPase adopt an elongated conformation possibly due to the opening of the active-site cleft, as found in the NMR structure of rat CNPase catalytic domain (Kozlov, Lee *et al*. 2003, Myllykoski, Raasakka *et al*. 2012). The mouse catalytic domain displays a more compact structure in the presence of citrate and the ligand also reduces aggregation of the protein (Myllykoski, Raasakka *et al*. 2012). It has been suggested that the overall solution conformation of fully active CNPase might be more open than that seen in the crystal structures as the crystals were obtained only in the presence of citrate or sulphate (Myllykoski, Raasakka *et al*. 2012). Different lengths of the mouse CNPase catalytic domain have been studied by SAXS (Myllykoski, Raasakka *et al*. 2012). The presence of the C-terminal tail does not alter the conformation of the catalytic domain in solution, whereas the N-terminally extended catalytic domain has a higher radius of gyration. Both N- and C-terminally extended protein is remarkably more elongated (Myllykoski, Raasakka *et al*. 2012). The symmetric peak of the p(r) for *Sc*Trl1 CPDase [Figure 39 C] suggests

that the protein, similarly to the mouse CNPase catalytic domain, is more compact than the proteins that also contain the N-terminal domain.

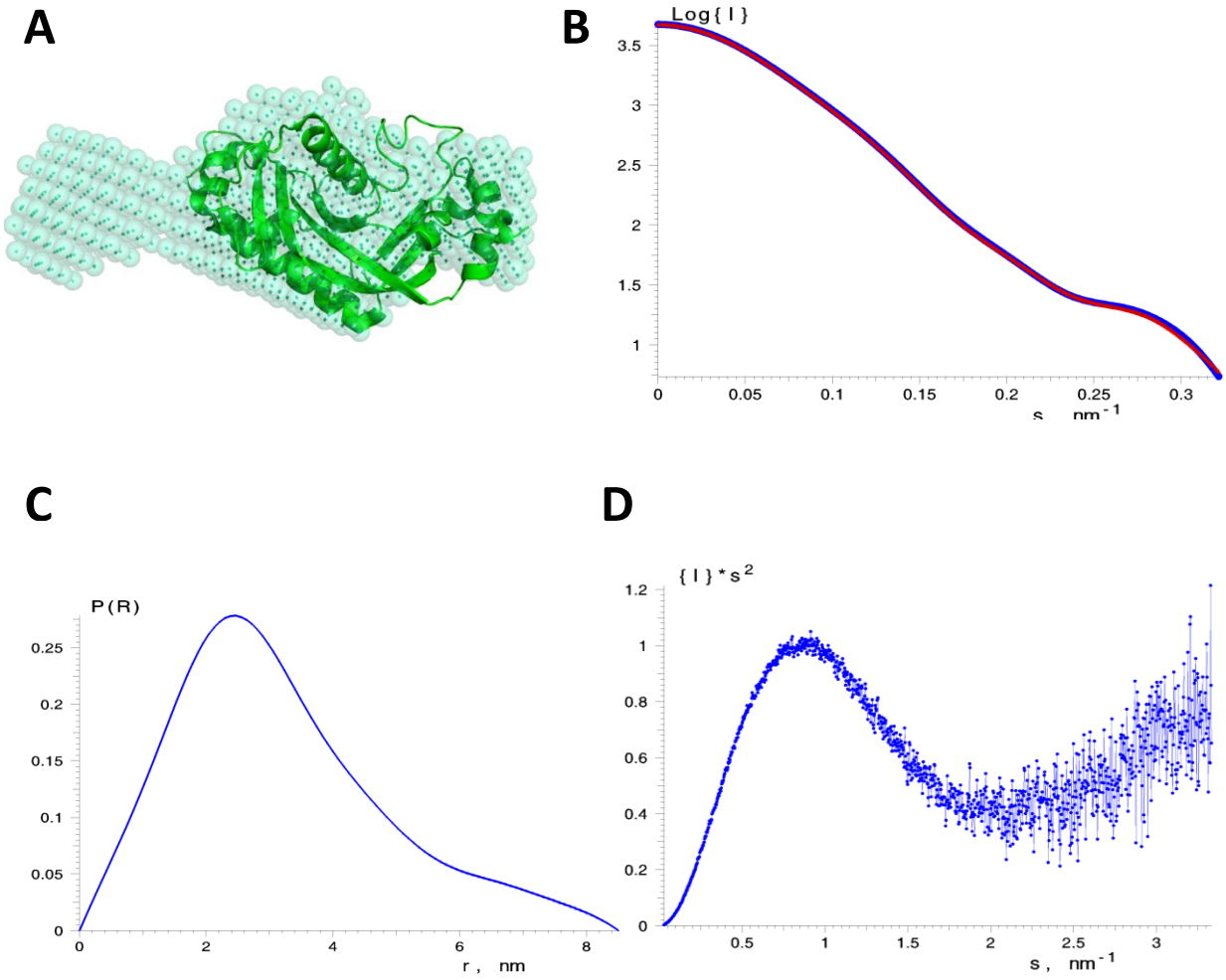


Figure 39. SAXS structure of *Sc*Trl1 CPDase. A) Superposition of the *ab initio* SAXS model (light blue) of the *Sc*Trl1 CPDase and the high-resolution crystal structure (green) of the mouse CNPase catalytic domain. **B)** Scattering curve of the monomeric *Sc*Trl1 CPDase. The raw data are indicated as blue dots, and the curve in red color represents the fit of the *ab initio* model to the raw data. **C)** Distance-distribution function of the *Sc*Trl1 CPDase. **D)** The Kratky plot displaying a peak characteristic of a folded protein.

The solution shapes and behavior of the PNK/CPDases and the CPDase are in many aspects comparable to those of the full-length mouse CNPase and its catalytic domain, respectively. The crystal structure of the mouse CNPase catalytic domain is superimposed on the monomeric and dimeric models constructed. The SAXS model of full-length mouse CNPase suggests that the N-terminal PNK-like and the C-terminal phosphodiesterase domains of mouse CNPase form an elongated assembly (Myllykoski, Raasakka *et al.* 2013), and the same is apparently true for the yeast and

lancelet enzymes. It has been suggested that an open conformation of the full-length protein could be crucial for substrate binding (Myllykoski, Raasakka *et al*. 2013). The domain arrangements of the PNK/CPDase in solution resemble that of the CNPase quite well. The crystal structure of the catalytic domain of mouse CNPase has shown that the elongated arrangement contains an open active site in the phosphodiesterase domain (Myllykoski, Raasakka *et al*. 2012).

Table 13. Structural parameters defined from experimental SAXS data and comparison to known parameters of mouse CNPase

Protein	**D_{max}** (nm)	**R_g** (nm)	**V_P** (nm^3)
Proteins used in the study			
*Sc*Trl1 PNK/CPDase (dimer)	27.35	5.30	206.95
*Sc*Trl1 PNK/CPDase (monomer)	11.56	3.37	78.56
Bf PNK/CPDase (dimer)	24.63	6.99	545.84
Bf PNK/CPDase (monomer)	14.3	4.16	137.65
*Sc*Trl1 CPDase	8.51	2.48	50.07
Homologous proteins			
Mouse CNPase[1]	8.8	2.6	-
Mouse CNPase catalytic domain[2]	7.1	2.13	44.4

[1] (Myllykoski, Raasakka et al. 2013), [2] (Myllykoski, Raasakka et al. 2012)

To test the effect of ligands on the conformation of the proteins in solution, the known substrates, ATP, GTP, $cNADP^+$ and yeast tRNA were complexed with the proteins and the scattering was measured, but data of usable quality could not be obtained (data not shown). Comparison of the SAXS models of *Sc*Trl1 PNK/CPDase, *Bf* PNK/CPDase and *Sc*Trl1 CPDase with the SAXS structural parameters of the homologous mouse CNPase indicates that the structural, oligomeric and hydrodynamic properties of the different constructs presented in this study are reliable. A structural model of the PNK/CPDase proteins [Figure 40] can be proposed based on the three-dimensional shapes determined in this study. Both yeast and lancelet PNK/CPDase exhibit an elongated conformation with the presence of a flexible region between the N-terminal PNK and C-terminal CPDase domains. Removal of the flexible region, which is not present in mouse CNPase, might aid in crystallization of the domains separately. The model also indicates that the flexible

region in lancelet PNK/CPDase is more elongated than yeast PNK/CPDase, although the length of the insertion present in the amino acid sequences of yeast and lancelet PNK/CPDase is almost the same.

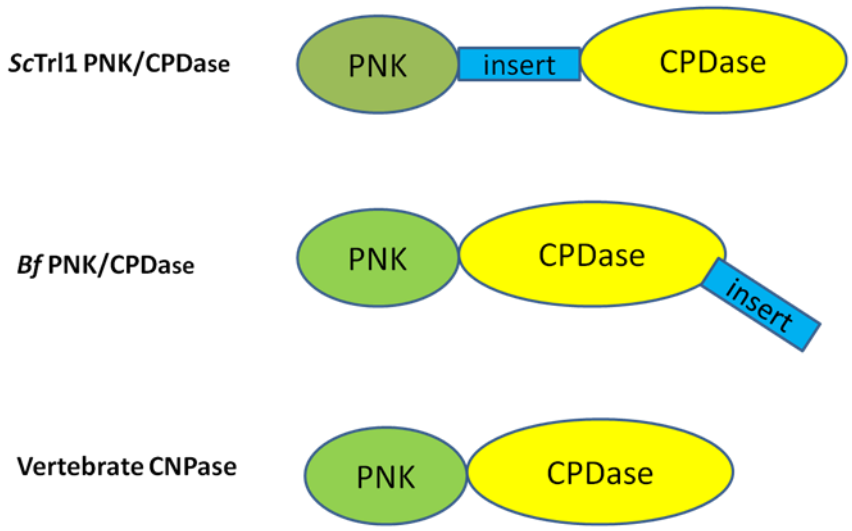


Figure 40. Structural models of PNK/CPDase proteins. The overall three-dimensional shapes of PNK/CPDase proteins determined in this study are used to propose a possible structural arrangement of the domains in solution.

The open and more elongated conformation of the PNK/CPDases might play a role in substrate binding; however, this needs to be confirmed by high-resolution structures of the enzymes with bound substrates. Based on the structural data presented here, we can also conclude that both the enzymes display structural similarities to the members of the 2H-phosphoesterase family. Upon availability of high-resolution structures, different methods such as CORAL (Petoukhov, Franke *et al*. 2012) and BUNCH (Petoukhov and Svergun 2005) can be used for rigid body SAXS modeling to obtain a detailed picture of the structural arrangement of the PNK and CPDase domains.

5.15. Dimerization of PNK/CPDase enzymes is inhibited by reducing agents

MALS experiments were carried out to determine the molecular weight and the oligomeric state of the purified proteins. The light scattering profiles from *Sc*Trl1 PNK/CPDase, *Sc*Trl1 CPDase and *Bf* PNK/CPDase indicate that the proteins are monomers in the presence of the reducing agent, 5 mM β-ME [Figures 41 & 42, Table 14]. This result confirms the monomeric nature of the proteins in the presence of reducing agents as observed during SEC purification, and thereby indicates the involvement of a disulphide bridge in dimerization. However, due to poor light scattering signal from the proteins in the absence of reducing agents, the molecular mass of the dimer could not be determined. Refractive indices were used as the source of protein concentration, to calculate the molar mass of the proteins. The

concentration of proteins measured from UV absorbance at 280 nm was not used in data analysis due to the possibility of errors in the presence of nucleic acid impurities. The estimated molecular mass values have high percentage of errors which might arise from the low scattering signal, and the high background from the buffer, possibly due to the presence of β-ME.

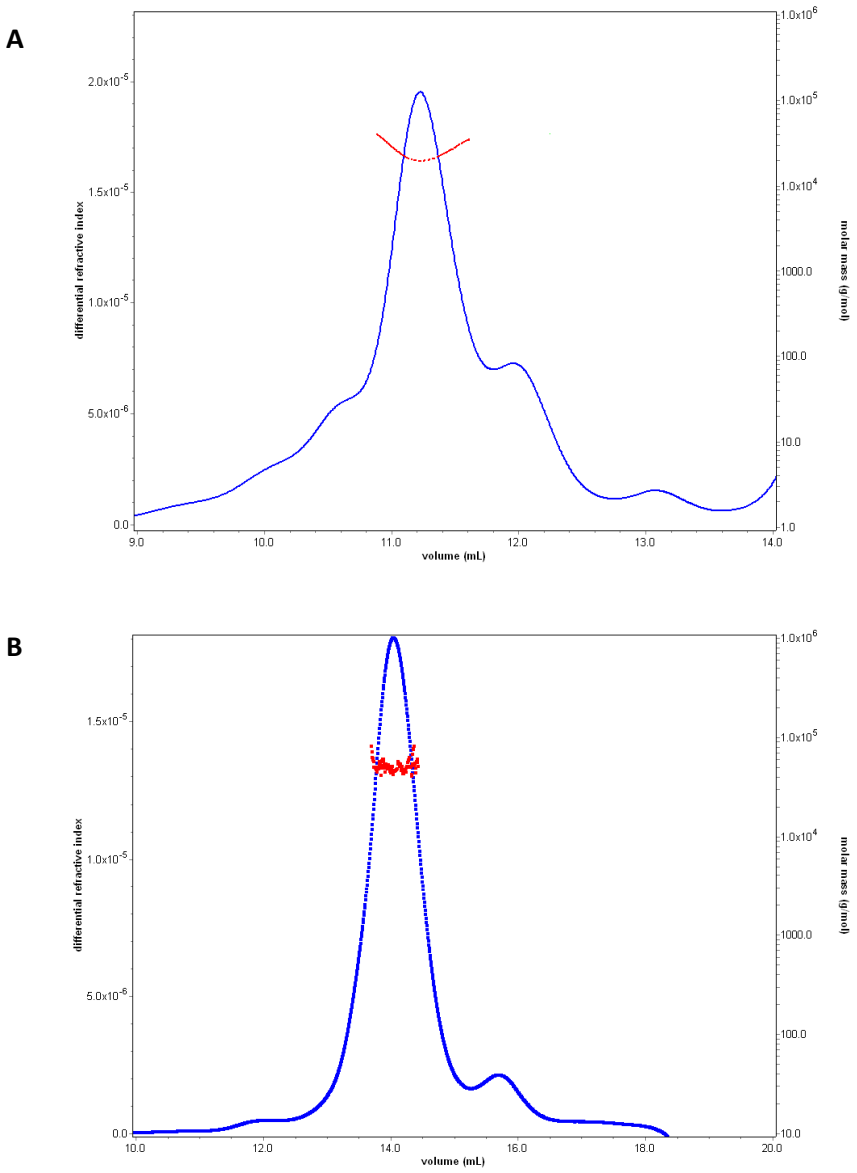


Figure 41. MALS analysis of PNK/CPDase and CPDase. MALS profiles of **A)** *Sc*Trl1 PNK/CPDase and **B)** *Bf* PNK/CPDase. The vertical lines (red) indicate mass distribution (molar mass as a function of elution volume) of SEC peak fractions.

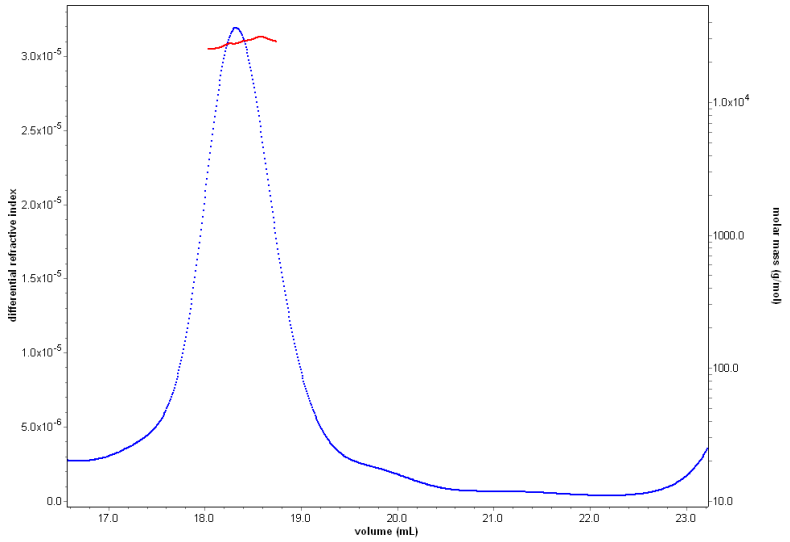


Figure 42. MALS analysis of *Sc*Trl1 CPDase. MALS profile of *Sc*Trl1 CPDase indicates that the protein is a monomer in solution. The color codes are the same as mentioned in Figure 41.

Table 14. Determination of molecular mass of the proteins from MALS analysis

Protein	**Monomer molecular mass** (kDa)	**Molecular mass from MALS** (kDa)
*Sc*Trl1 PNK/CPDase	52.7	55.7 (±6%)
*Sc*Trl1 CPDase	32.2	33.8 (±3 %)
Bf PNK/CPDase	60.3	65 (±5 %)
BSA	66.4	68 (±5 %)

It has been shown previously that the CNPase dimer could be dissociated into catalytically active monomers by the addition of 1 mM DTT and 5 M urea (Müller 1982). Although 5 M urea completely inhibits the enzymatic activity of the CNPase, the activity could be restored without any significant difference (Müller 1982). The tendency of dimerization has also been observed in the homologous full-length and the N-terminal domain of mouse CNPase; the N-terminal domain has been suggested to mediate self-association of full-length CNPase (Myllykoski, Raasakka *et al.* 2012). Our SAXS results also suggest the requirement of the N-terminal PNK domain for dimerization of yeast Trl1 PNK/CPDase. One of the two cysteine residues in the PNK domain of yeast Trl1 and lancelet PNK/CPDase might be involved in the formation of a disulphide bond that leads to dimerization of the enzymes. The role of the N-terminal domain of PNK/CPDase in dimerization has to be confirmed by MALS

analysis of the PNK/CPDase, PNK, and CPDase domains in the presence and absence of reducing agents.

5.16. Crystallization trials

As one of the main goals of the current study was to determine high-resolution structures of yeast and lancelet PNK/CPDase proteins with and without substrates in order to probe into the overall structure and the molecular determinants of substrate specificity, pure and stable *Sc*Trl1 PNK/CPDase, *Sc*Trl1 CPDase and *Bf* PNK/CPDase were subjected to extensive screening for crystallization hits. The results from thermal stability analysis and other interaction studies were all considered, while preparing samples for crystallization screens. Routine screening of conditions was carried out to explore the crystallization parametric space by applying variations in a number of parameters such as protein concentration, storage temperature, crystallization drop volume and the presence of ligands, such as 2',3'-cNADP$^+$, 2',3'-cCMP, ATP and GTP in various stoichiometric proportions. The different oligomeric states of the PNK/CPDase domains were used separately. The maximum concentration used in the screening for yeast and lancelet PNK/CPDases was approximately 8 mg/ml, and most of the drops in different conditions appeared clear even after six months from preparation of the screens. Thus, higher protein concentrations might be required to facilitate crystallization. Yeast CPDase was used in the screens up to a maximum concentration of 18 mg/ml, and concentrations above 15 mg/ml resulted in heavy precipitation in almost 90% of the drops, suggesting to lower the protein concentration. No clear difference or improvement in drops was observed between storage at 20°C and at 4°C. Despite extensive screening of conditions, no crystals of PNK/CPDase and CPDase proteins have appeared thus far.

Although a few high-resolution structures of the catalytic domain (CPDase) of 2H phosphoesterase family members have been determined, a structure of the full-length protein with both the N-terminal PNK-like domain and the CPDase domain is not available yet. The lack of a crystal structure for the PNK/CPDase domains and the crystallization probability predicted from amino acid sequence both indicate that crystallization of these proteins is difficult. Factors such as variable oligomerization in a slow equilibrium, local flexibility in a multi-domain protein and non-specific interactions with nucleic acids from the expression host might affect the crystallization of PNK/CPDase proteins. The unfavorable crystallization properties are not limited to the PNK/CPDase proteins from yeast and lancelet as the behavior of mouse CNPase is also the same.

As indicated by SAXS results, the removal of the long insertions in the yeast and lancelet PNK/CPDase proteins might aid in crystallization. Thus, further optimization

of constructs might be helpful. It is known that protein motions due to solvent-exposed, flexible amino acid side chains could be disruptive to the formation of a three-dimensional crystal lattice. In such cases, surface entropy reduction mutagenesis techniques, such as lysine methylation have been proven as rescue strategies for effective protein crystallization (Walter, Meier et al. 2006). The primary structures of *Sc*Trl1 PNK/CPDase, *Bf* PNK/CPDase and *Sc*Trl1 CPDase contain 12%, 6.2% and 12.9% lysine residues, respectively. Thus, methylation of the lysine residues in PNK/CPDase and CPDase proteins prior to crystallization could be attempted. As the SAXS models and the secondary structure predicted from the sequence indicate the presence of flexible regions in both PNK/CPDase and CPDase proteins, *in situ* limited proteolysis using a serial dilution of different proteases in the drops could be tried as it might cleave off the flexible fragments and promote crystallization.

6. Conclusions and Future Perspectives

The main objective of this study was to investigate structural and functional properties of the tRNA healing enzymes, PNK and CPDase, from yeast and lancelet. The results of the study show that the bacterially expressed yeast and lancelet PNK/CPDase and yeast CPDase proteins are folded and contain a mixture of α-helices and β-strands. The thermal stability of the proteins is high around neutral pH, and the stability increases in the presence of a substrate. The PNK/CPDase proteins are active in both polynucleotide kinase and cyclic phosphodiesterase activity assays, while the yeast CPDase is an active cyclic phosphodiesterase. The PNK/CPDase proteins, containing the NTP-binding P-loop or Walker A motif, interact with nucleotide analogues. The interaction of PNK/CPDase proteins with nucleotides and cyclic nucleotide substrates induces conformational transitions in the proteins. The PNK/CPDase proteins bind RNA from the expression host, *E. coli*, and the nucleic acid could be removed upon addition of a combination of RNase enzymes. The PNK/CPDase proteins form dimers and the CPDase exists as a monomer in solution. However, dimerization of PNK/CPDase could be inhibited by addition of a reducing agent, indicating the possible involvement of a disulphide bridge. Solution structures indicate that the PNK/CPDase proteins adopt an elongated conformation, and the CPDase protein exhibits a compact structure. The initial structural and functional properties investigated in this study would aid researchers to design experiments to further understand the molecular architecture and mechanism of reaction of the enzymes.

In this study, recombinant expression and purification of the CPDase and PNK/CPDase domains of yeast Trl1, and the lancelet PNK/CPDase has been optimized. Although the key residues involved in the PNK and CPDase activities of yeast Trl1 were identified previously by applying site-directed mutagenesis, the active-site residues of lancelet PNK/CPDase are not yet explored. Mutational analyses of the "P-loop" residues in the PNK domain and the "H-x-(T/S)-x" motifs in the CPDase domain of lancelet PNK/CPDase would be helpful to identify the essential residues. The current study has identified two cyclic nucleotide compounds as potential substrates for CPDase activity of the enzymes and screening of more substrates could be used for detailed functional characterization. The oligomeric state of the proteins in the presence and absence of reducing agents, nucleotides and RNA substrates needs to be analyzed. Changes in the overall shape and conformation of the proteins upon ligand binding could be probed using a combination of techniques, such as DLS and SAXS. Efforts to analyze the preliminary SAXS data from PNK/CPDase-RNA complexes are ongoing; however more measurements with different concentrations of RNA are required to confirm the binding and validate the SAXS

model. Since none of the constructs encoding the PNK domain of yeast Trl1 show expression in bacterial cells, expression in insect cells could be tried. Although our results provide clues about the possible role of the PNK domain in dimerization and RNA binding, further analyses using different constructs and RNA substrates are necessary. Since crystallization of the proteins has not been successful yet, strategies such as reductive methylation and *in situ* proteolysis could be tried. As the SAXS models from this study indicate the presence of an extended tail in the CPDase domain, a new set of constructs without the insertion at the N-terminus could be prepared and used for crystallization. Hence, new constructs, strategies and methods have to be applied to further explore the structure and function of these enzymes. Although the current study provides novel structural and functional information on the 2H phosphoesterase superfamily, further experiments are essential to understand the evolution of structure and catalytic activity of its members.

7A. Abstract

The primary product of transcription of a tRNA gene, in all systems studied thus far, is a pre-tRNA molecule. In order to become a mature and functional tRNA, the primary transcript undergoes extensive processing. tRNA processing involves a collection of enzymatic reactions that cleave, trim and splice the precursor molecule. tRNA splicing is an essential step in the processing of pre-tRNA. In eukaryotes, splicing of a pre-tRNA involves three steps: the cleavage of an intron-containing tRNA by a site-specific endonuclease, healing of the broken tRNA halves and sealing of the healed ends.

The work reported in this dissertation focused specifically on the enzymes polynucleotide kinase (PNK) and 2',3'-cyclic nucleotide 3'-phophodiesterase (CPDase) that catalyze the healing step of tRNA splicing in the yeast, *Saccharomyces cerevisiae,* and the lancelet, *Branchiostoma floridae*. The NTP-dependent PNK catalyzes the phosphorylation of the 5'-OH group of the 3'-tRNA half, and the CPDase catalyzes the hydrolysis of the 2',3'-cyclic phosphate at the 3'-end of the 5'-tRNA half. In yeast, both activities exist, together with a ligase activity, in a single tri-functional polypeptide (tRNA ligase), whereas the lancelet enzyme contains only the PNK and CPDase domains in a bi-functional polypeptide.

To study the biophysical and biochemical properties of the PNK/CPDase domains of the yeast tRNA ligase and the lancelet PNK/CPDase, multiple expression constructs encoding different versions of the full-length proteins and separate domains of the enzymes were prepared for expression in *Eschericia coli* cells. Conditions for soluble expression were screened in different strains of *E. coli*. The PNK/CPDase domains and different versions of the CPDase domain of the yeast Trl1, and the lancelet PNK/CPDase were successfully expressed and used for large-scale protein production. The proteins were purified by immobilized-metal affinity and size exclusion chromatography. The secondary structure of the purified proteins was examined by synchrotron radiation circular dichroism (SRCD) spectroscopy, and all purified proteins were folded. A thermal stability assay was used to find a stabilizing buffer and to study the effect of ligands on the thermal stability of the proteins.

The PNK activity of the N-terminal kinase domain was confirmed using DNA and RNA oligonucleotides with radiolabeled ATP. The CPDase activity assays using either 2',3'-cNADP$^+$ or 2',3'-cCMP as a substrate, confirmed the CPDase activity of the C-terminal domains. Since the N-terminal kinase domain of the yeast Trl1 and lancelet PNK/CPDase contain a putative NTP-binding P-loop or Walker-A motif, the

binding of nucleotides to the proteins was tested using fluorescently labeled nucleotide analogues of AMP, ATP and GTP and applying fluorescence resonance energy transfer (FRET). Both proteins bound ATP and GTP, but not AMP. Changes in the conformation of the proteins upon nucleotide binding were probed by SRCD spectroscopy.

The proteins containing both PNK and CPDase domains appeared as dimers during purification by size exclusion chromatography. Small-angle X-ray scattering was used to demonstrate that the proteins exist both as a dimer and a monomer in solution. SAXS also provided low-resolution shapes of the proteins and indicated that the proteins containing the PNK/CPDase domains are elongated and that the CPDase domain alone is more compact in solution. Reducing agents decreased the extent of dimerization, and static light scattering in the presence of a reducing agent showed, indeed, a monomer in solution. The presence of a nucleic acid impurity in the expressed proteins was suspected based on size exclusion chromatography, and the nucleic acid was identified as RNA by nuclease treatment of the complex and subsequent analysis by agarose gel electrophoresis. Crystallization of the enzymes was attempted, but has not yielded crystals yet. Thus, preliminary structural and functional information were obtained for the yeast and lancelet tRNA healing enzymes and could be used for further characterization of the molecules.

7B. Zusammenfassung

In allen bisher untersuchten Systemen wird das tRNA-Gen zunächst in eine prä-tRNA transkribiert. Die reife, funktionale tRNA entsteht aus dem prä-tRNA-Vorläufer in Folge einer umfangreichen Prozessierung. Während der tRNA-Prozessierung finden eine Reihe von enzymatischen Reaktionen statt, die zur Spaltung, dem Trimmen und dem Spleißen des Vorläufermoleküls führen. Das tRNA-Spleißen stellt dabei einen essentiellen Bestandteil dieses Prozesses dar. In Eukaryoten lässt sich dieser Vorgang in folgende Schritte unterteilen: die Spaltung der prä-tRNA durch eine spezifische Endonuklease, die daraus folgende Freisetzung des Introns und sogenannter tRNA-Halbmoleküle und eine anschließende Ligation der Halbmoleküle.

Die vorliegende Arbeit befasst sich im Besonderen mit den Enzymen Polynukleotidkinase (PNK) und 2',3'-zyklische-Nukleotid-3'-Phosphodiesterase (CPDase), die den Reparaturschritt des tRNA-Spleißens in der Hefe *Saccharomyces cerevisiae* und im Lanzettfischchen *Branchiostoma floridae* katalysieren. Die NTP-abhängige PNK katalysiert die Phosphorylierung der 5'-OH Gruppe der 3'-tRNA-Hälfte, und die CPDase katalysiert die Hydrolyse des 2',3'-zyklischen Phosphats am 3' Ende der 5'-tRNA-Hälfte. In der Hefe bestehen die PNK- und CPDaseaktivitäten zusammen mit der Ligaseaktivität in einem tri-funktionalem Polypeptid. Im Gegensatz dazu, enthält das Enzym des Lanzettfischchens die PNK und CPDasedomänen in einem bi-funktionalem Polypeptid.

Um die biophysikalischen und biochemischen Eigenschaften der PNK/CPDasedomänen der Hefe tRNA-Ligase und der Lanzettfischchen PNK/CPDase zu untersuchen, wurden unterschiedliche Expressionskonstrukte für Expression in *Eschericia coli* generiert. Diese Konstrukte kodieren für Domänen von unterschiedlichen Längen, und deren Expression wurde in verschiedenen *E. coli*-Stämmen getestet. Die PNK/CPDasedomänen und verschiedene Längen der CPDasedomäne aus der Hefe Trl1 und derLanzettfischchen PNK/CPDase konnten erfolgreich exprimiert werden und wurden weiter für die Proteinproduktion im großen Maßstab verwendet. Die Proteine wurden mit Hilfe von Affinitäts- und Gelpermeationschromatographie aufgereinigt.

Die Sekundärstrukturen der aufgereinigten Proteine wurden mit Hilfe von Synchrotronstrahlungcirculardichroismusspektroskopie (SRCD) analysiert, und eine korrekte Faltung konnte nachgewiesen werden. Mit Hilfe eines 'thermal shift assay' wurden die optimale Pufferlösung und der Effekt von Liganden auf die Stabilität der Proteine getestet. Die PNKaktivität der N-terminalen Kinasedomäne konnte unter Verwendung von DNA- und RNA-Olikonukleotiden, mit radioaktiv markiertem ATP,

bestätigt werden. Die CPDaseaktivität der C-terminalen Domäne wurde mit den Substraten 2',3'-cNADP$^+$ oder 2',3'-cCMP nachgewiesen. Da die N-terminale Kinasedomäne der Hefe Trl1 und die Lanzettfischchen PNK/CPDase den mutmaßlichen NTP-Bindungs-P-loop oder Walker-A-Motif enthält, wurde die Bindung von Nukleotiden and die Proteine mit Hilfe von fluoreszenz-markierten Nukleotidanalogen von AMP, ATP und GTP und unter Anwendung des Förster-Resonanzenergietransfers (FRET) getestet. Demanch binden beide Proteine ATP und GTP, jedoch nicht AMP. Konformationsänderungen der Proteine in Folge der Nukleotidbindung konnten anhand von SRCD-Messungen nachvollzogen werden.

Die Proteine, welche sowohl die PNK-, als auch die CPDasedomänen enthielten, erschienen während der Gelpermeationschromatographie als Dimere. Anhand von Röntgenkleinwinkelstreuung (SAXS) konnte demonstriert werden, dass in Lösung sowohl Dimere, als auch als Monomere der Proteine existieren. Außerdem konnten mit Hilfe von SAXS die Strukturen der Proteine in niedriger Auflösung aufgeklärt werden. Diese Ergebnisse ließen erkennen, dass die Proteine, welche die PNK/CPDasedomänen enthielten, eine ausgedehnte Form in Lösung aufweisen, während die CPDasedomäne eher kompakt strukturiert sind. Reduktionsmitteln führten zu einer reduzierten Dimerisierung, und nach static light scattering Messungen waren die Proteine tatsächlich mit Reduktionsmitteln in Lösung Monomere. Mit Hilfe von Gelpermeationschromatographie wurden Nukleinsäureunreinheiten in den exprimierten Proteinen detektiert, welche in Folge einer Nukleasebehandlung des Komplexes und anschließender Agarosegelelektrophorese als RNA identifiziert werden konnten. Zusätzlich wurden Kristallisationsversuche durchgeführt, die jedoch bisher keine erfolgreichen Ergebnisse erbrachten.

Im Zuge dieser Arbeit konnten strukturelle und funktionelle Informationen für die Hefe und Lanzettfischchen tRNA-Reparaturenzyme erbracht werden, welche für weitere Charakterisierungen der Moleküle verwendet werden können.

8. References

Abelson, J., C. R. Trotta and H. Li (1998). "tRNA splicing." J Biol Chem **273**(21): 12685-12688.

Adams, P. L., M. R. Stahley, M. L. Gill, A. B. Kosek, J. Wang and S. A. Strobel (2004). "Crystal structure of a group I intron splicing intermediate." RNA **10**(12): 1867-1887.

Adams, P. L., M. R. Stahley, A. B. Kosek, J. Wang and S. A. Strobel (2004). "Crystal structure of a self-splicing group I intron with both exons." Nature **430**(6995): 45-50.

Adler, A. J., N. J. Greenfield and G. D. Fasman (1973). Circular dichroism and optical rotatory dispersion of proteins and polypeptides. Methods Enzymol. Academic Press. **Volume 27:** 675-735.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." J Mol Biol **215**(3): 403-410.

Apostol, B. L., S. K. Westaway, J. Abelson and C. L. Greer (1991). "Deletion analysis of a multifunctional yeast tRNA ligase polypeptide. Identification of essential and dispensable functional domains." J Biol Chem **266**(12): 7445-7455.

Arn, E. A. and J. N. Abelson (1996). "The 2'-5' RNA ligase of Escherichia coli. Purification, cloning, and genomic disruption." J Biol Chem **271**(49): 31145-31153.

Artimo, P., M. Jonnalagedda, K. Arnold, D. Baratin, G. Csardi, E. de Castro, S. Duvaud, V. Flegel, A. Fortier, E. Gasteiger, A. Grosdidier, C. Hernandez, V. Ioannidis, D. Kuznetsov, R. Liechti, S. Moretti, K. Mostaguir, N. Redaschi, G. Rossier, I. Xenarios and H. Stockinger (2012). "ExPASy: SIB bioinformatics resource portal." Nucleic Acids Res **40**(Web Server issue): W597-603.

Baldi, M. I., E. Mattoccia, E. Bufardeci, S. Fabbri and G. P. Tocchini-Valentini (1992). "Participation of the intron in the reaction catalyzed by the Xenopus tRNA splicing endonuclease." Science **255**(5050): 1404-1408.

Ballestero, R. P., J. A. Dybowski, G. Levy, B. W. Agranoff and M. D. Uhler (1999). "Cloning and characterization of zRICH, a 2',3'-cyclic-nucleotide 3'-phosphodiesterase induced during zebrafish optic nerve regeneration." J Neurochem **72**(4): 1362-1371.

Ballestero, R. P., G. R. Wilmot, M. L. Leski, M. D. Uhler and B. W. Agranoff (1995). "Isolation of cDNA clones encoding RICH: a protein induced during goldfish optic

nerve regeneration with homology to mammalian 2',3'-cyclic-nucleotide 3'-phosphodiesterases." Proc Natl Acad Sci U S A **92**(19): 8621-8625.

Becker, A. and J. Hurwitz (1967). "The enzymatic cleavage of phosphate termini from polynucleotides." J Biol Chem **242**(5): 936-950.

Beer, A (1852). Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. Annalen der Physik und Chemie. 86:78-88.

Bekesi, A., M. Pukancsik, P. Haasz, L. Felfoldi, I. Leveles, V. Muha, E. Hunyadi-Gulyas, A. Erdei, K. F. Medzihradszky and B. G. Vertessy (2011). "Association of RNA with the uracil-DNA-degrading factor has major conformational effects and is potentially involved in protein folding." FEBS J **278**(2): 295-315.

Belford, H. G., S. K. Westaway, J. Abelson and C. L. Greer (1993). "Multiple nucleotide cofactor use by yeast ligase in tRNA splicing. Evidence for independent ATP- and GTP-binding sites." J Biol Chem **268**(4): 2444-2450.

Bennett-Lovsey, R. M., A. D. Herbert, M. J. Sternberg and L. A. Kelley (2008). "Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre." Proteins **70**(3): 611-625.

Bertani, G. (1951). "Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli." J Bacteriol **62**(3): 293-300.

Bifulco, M., C. Laezza, S. Stingo and J. Wolff (2002). "2',3'-Cyclic nucleotide 3'-phosphodiesterase: a membrane-bound, microtubule-associated protein and membrane anchor for tubulin." Proc Natl Acad Sci U S A **99**(4): 1807-1812.

Biniszkiewicz, D., E. Cesnaviciene and D. A. Shub (1994). "Self-splicing group I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria." EMBO J **13**(19): 4629-4635.

Braun, P. E., D. De Angelis, W. W. Shtybel and L. Bernier (1991). "Isoprenoid modification permits 2',3'-cyclic nucleotide 3'-phosphodiesterase to bind to membranes." J Neurosci Res **30**(3): 540-544.

Calvin, K. and H. Li (2008). "RNA-splicing endonuclease structure and function." Cell Mol Life Sci **65**(7-8): 1176-1185.

Cameron, V. and O. C. Uhlenbeck (1977). "3'-Phosphatase activity in T4 polynucleotide kinase." Biochemistry **16**(23): 5120-5126.

Cech, T. R. (2002). "Ribozymes, the first 20 years." Biochem Soc Trans **30**(Pt 6): 1162-1166.

Challa, M., G. R. Chapa, S. Govindaraju, M. González-García and R. P. Ballestero (2006). "Characterization of the domains of zRICH, a protein induced during optic nerve regeneration in zebrafish." Brain Res **1100**(1): 42-54.

Chan, P. P. and T. M. Lowe (2009). "GtRNAdb: a database of transfer RNA genes detected in genomic sequence." Nucleic Acids Res **37**(Database issue): D93-97.

Chen, D., C. L. Luongo, M. L. Nibert and J. T. Patton (1999). "Rotavirus open cores catalyze 5'-capping and methylation of exogenous RNA: evidence that VP3 is a methyltransferase." Virology **265**(1): 120-130.

Cole, C., J. D. Barber and G. J. Barton (2008). "The Jpred 3 secondary structure prediction server." Nucleic Acids Res **36**(Web Server issue): W197-201.

Compton, L. A. and W. C. Johnson (1986). "Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication." Anal Biochem **155**(1): 155-167.

Cranston, J. W., R. Silber, V. G. Malathi and J. Hurwitz (1974). "Studies on ribonucleic acid ligase. Characterization of an adenosine triphosphate-inorganic pyrophosphate exchange reaction and demonstration of an enzyme-adenylate complex with T4 bacteriophage-induced enzyme." J Biol Chem **249**(23): 7447-7456.

Crick, F. H. (1968). "The origin of the genetic code." J Mol Biol **38**(3): 367-379.
Culbertson, M. R. and M. Winey (1989). "Split tRNA genes and their products: a paradigm for the study of cell function and evolution." Yeast **5**(6): 405-427.

Culver, G. M., S. A. Consaul, K. T. Tycowski, W. Filipowicz and E. M. Phizicky (1994). "tRNA splicing in yeast and wheat germ. A cyclic phosphodiesterase implicated in the metabolism of ADP-ribose 1",2"-cyclic phosphate." J Biol Chem **269**(40): 24928-24934.

Culver, G. M., S. M. McCraith, S. A. Consaul, D. R. Stanford and E. M. Phizicky (1997). "A 2'-phosphotransferase implicated in tRNA splicing is essential in Saccharomyces cerevisiae." J Biol Chem **272**(20): 13203-13210.

Culver, G. M., S. M. McCraith, M. Zillmann, R. Kierzek, N. Michaud, R. D. LaReau, D. H. Turner and E. M. Phizicky (1993). "An NAD derivative produced during transfer RNA splicing: ADP-ribose 1"-2" cyclic phosphate." Science **261**(5118): 206-208.

Daniels, C. J., R. Gupta and W. F. Doolittle (1985). "Transcription and excision of a large intron in the tRNATrp gene of an archaebacterium, Halobacterium volcanii." J Biol Chem **260**(5): 3132-3134.

De Angelis, D. A. and P. E. Braun (1994). "Isoprenylation of brain 2',3'-cyclic nucleotide 3'-phosphodiesterase modulates cell morphology." J Neurosci Res **39**(4): 386-397.

De Angelis, D. A. and P. E. Braun (1996). "Binding of 2',3'-cyclic nucleotide 3'-phosphodiesterase to myelin: an in vitro study." J Neurochem **66**(6): 2523-2531.

Debye, P (1947). Molecular weight determination by light scattering. J Phys Colloid Chem 51:18-32.

Di Nicola Negri, E., S. Fabbri, E. Bufardeci, M. I. Baldi, D. Gandini Attardi, E. Mattoccia and G. P. Tocchini-Valentini (1997). "The eucaryal tRNA splicing endonuclease recognizes a tripartite set of RNA elements." Cell **89**(6): 859-866.

Di Segni, G., L. Borghese, S. Sebastiani and G. P. Tocchini-Valentini (2005). "A pre-tRNA carrying intron features typical of Archaea is spliced in yeast." RNA **11**(1): 70-76.

Doniach, S. (2001). "Changes in biomolecular conformation seen by small angle X-ray scattering." Chem Rev **101**(6): 1763-1778.

Dosztányi, Z., V. Csizmok, P. Tompa and I. Simon (2005). "IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content." Bioinformatics **21**(16): 3433-3434.

Dosztányi, Z., V. Csizmók, P. Tompa and I. Simon (2005). "The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins." J Mol Biol **347**(4): 827-839.

Drummond, G. I., N. T. Iyer and J. Keith (1962). "Hydrolysis of Ribonucleoside 2',3'-Cyclic Phosphates by a Diesterase from Brain." J Biol Chem **237**(11): 3535-3539.

Edgar, J. M., M. McLaughlin, H. B. Werner, M. C. McCulloch, J. A. Barrie, A. Brown, A. B. Faichney, N. Snaidero, K. A. Nave and I. R. Griffiths (2009). "Early ultrastructural defects of axons and axon-glia junctions in mice lacking expression of Cnp1." Glia **57**(16): 1815-1824.

Edgell, D. R., M. Belfort and D. A. Shub (2000). "Barriers to intron promiscuity in bacteria." J Bacteriol **182**(19): 5281-5289.

Englert, M. and H. Beier (2005). "Plant tRNA ligases are multifunctional enzymes that have diverged in sequence and substrate specificity from RNA ligases of other phylogenetic origins." Nucleic Acids Res **33**(1): 388-399.

Englert, M., K. Sheppard, A. Aslanian, J. R. Yates and D. Söll (2011). "Archaeal 3'-phosphate RNA splicing ligase characterization identifies the missing component in tRNA maturation." Proc Natl Acad Sci U S A **108**(4): 1290-1295.

Englert, M., K. Sheppard, S. Gundllapalli, H. Beier and D. Söll (2010). "Branchiostoma floridae has separate healing and sealing enzymes for 5'-phosphate RNA ligation." Proc Natl Acad Sci U S A **107**(39): 16834-16839.

Ericsson, U. B., B. M. Hallberg, G. T. Detitta, N. Dekker and P. Nordlund (2006). "Thermofluor-based high-throughput stability optimization of proteins for structural studies." Anal Biochem **357**(2): 289-298.

Fabbri, S., P. Fruscoloni, E. Bufardeci, E. Di Nicola Negri, M. I. Baldi, D. G. Attardi, E. Mattoccia and G. P. Tocchini-Valentini (1998). "Conservation of substrate recognition mechanisms by tRNA splicing endonucleases." Science **280**(5361): 284-286.

Filipowicz, W. and A. J. Shatkin (1983). "Origin of splice junction phosphate in tRNAs processed by HeLa cell extract." Cell **32**(2): 547-557.

Franke, D. and D. I. Svergun (2009). "DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering." J Appl Crystallogr **42**(2): 342-346.

Galburt, E. A., J. Pelletier, G. Wilson and B. L. Stoddard (2002). "Structure of a tRNA repair enzyme and molecular biology workhorse: T4 polynucleotide kinase." Structure **10**(9): 1249-1260.

Gao, Y. G., M. Yao, A. Okada and I. Tanaka (2006). "The structure of Pyrococcus horikoshii 2'-5' RNA ligase at 1.94 A resolution reveals a possible open form with a wider active-site cleft." Acta Crystallogr Sect F Struct Biol Cryst Commun **62**(Pt 12): 1196-1200.

Garces, F., L. H. Pearl and A. W. Oliver (2011). "The structural basis for substrate recognition by mammalian polynucleotide kinase 3' phosphatase." Mol Cell **44**(3): 385-396.

Garriga, G. and A. M. Lambowitz (1984). "RNA splicing in neurospora mitochondria: self-splicing of a mitochondrial intron in vitro." Cell **39**(3 Pt 2): 631-641.

Glatter, O. (1977). "A new method for the evaluation of small-angle scattering data." J Appl Crystallogr **10**(5): 415-421.

Gold, M. G., F. D. Smith, J. D. Scott and D. Barford (2008). "AKAP18 contains a phosphoesterase domain that binds AMP." J Mol Biol **375**(5): 1329-1343.

Gomes, I. and R. Gupta (1997). "RNA splicing ligase activity in the archaeon Haloferax volcanii." Biochem Biophys Res Commun **237**(3): 588-594.

Goodman, H. M., M. V. Olson and B. D. Hall (1977). "Nucleotide sequence of a mutant eukaryotic gene: the yeast tyrosine-inserting ochre suppressor SUP4-o." Proc Natl Acad Sci U S A **74**(12): 5453-5457.

Gravel, M., D. DeAngelis and P. E. Braun (1994). "Molecular cloning and characterization of rat brain 2',3'-cyclic nucleotide 3'-phosphodiesterase isoform 2." J Neurosci Res **38**(3): 243-247.

Gravel, M., F. Robert, V. Kottis, I. E. Gallouzi, J. Pelletier and P. E. Braun (2009). "2',3'-Cyclic nucleotide 3'-phosphodiesterase: a novel RNA-binding protein that inhibits protein synthesis." J Neurosci Res **87**(5): 1069-1079.

Greer, C. L. (1986). "Assembly of a tRNA splicing complex: evidence for concerted excision and joining steps in splicing in vitro." Mol Cell Biol **6**(2): 635-644.

Greer, C. L., B. Javor and J. Abelson (1983). "RNA ligase in bacteria: formation of a 2',5' linkage by an E. coli extract." Cell **33**(3): 899-906.

Greer, C. L., C. L. Peebles, P. Gegenheimer and J. Abelson (1983). "Mechanism of action of a yeast RNA ligase in tRNA splicing." Cell **32**(2): 537-546.

Greer, C. L., D. Söll and I. Willis (1987). "Substrate recognition and identification of splice sites by the tRNA-splicing endonuclease and ligase from Saccharomyces cerevisiae." Mol Cell Biol **7**(1): 76-84.

Guinier, A (1939). La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques. Ann Phys (Paris) 12: 161-237.

Haselkorn, R. and L. B. Rothman-Denes (1973). "Protein synthesis." Annu Rev Biochem **42**: 397-438.

Ho, C. K. and S. Shuman (2002). "Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains." Proc Natl Acad Sci U S A **99**(20): 12709-12714.

Hofmann, A., A. Zdanov, P. Genschik, S. Ruvinov, W. Filipowicz and A. Wlodawer (2000). "Structure and mechanism of activity of the cyclic phosphodiesterase of Appr>p, a product of the tRNA splicing reaction." EMBO J **19**(22): 6207-6217.

Hopper, A. K. and F. Banks (1978). "A yeast mutant which accumulates precursor tRNAs." Cell **14**(2): 211-219.

Hu, Q. D., H. Lu, K. Huo, K. Ying, J. Li, Y. Xie, Y. Mao and Y. Y. Li (2003). "A human homolog of the yeast gene encoding tRNA 2'-phosphotransferase: cloning, characterization and complementation analysis." Cell Mol Life Sci **60**(8): 1725-1732.

Hugli, T. E., M. Bustin and S. Moore (1973). "Spectrophotometric assay of 2',3'-cyclic nucleotide 3'-phosphohydrolase: application to the enzyme in bovine brain." Brain Res **58**(1): 191-203.

Ibba, M. and D. Soll (2000). "Aminoacyl-tRNA synthesis." Annu Rev Biochem **69**: 617-650.

Inoue, H., H. Nojima and H. Okayama (1990). "High efficiency transformation of Escherichia coli with plasmids." Gene **96**(1): 23-28.

Johnson, J. D., R. Ogden, P. Johnson, J. Abelson, P. Dembeck and K. Itakura (1980). "Transcription and processing of a yeast tRNA gene containing a modified intervening sequence." Proc Natl Acad Sci U S A **77**(5): 2564-2568.

Jones, D. T. (1999). "Protein secondary structure prediction based on position-specific scoring matrices." J Mol Biol **292**(2): 195-202.

Kamat, S. S., H. J. Williams and F. M. Raushel (2011). "Intermediates in the transformation of phosphonates to phosphate by bacteria." Nature **480**(7378): 570-573.

Kanai, A., A. Sato, Y. Fukuda, K. Okada, T. Matsuda, T. Sakamoto, Y. Muto, S. Yokoyama, G. Kawai and M. Tomita (2009). "Characterization of a heat-stable enzyme possessing GTP-dependent RNA ligase activity from a hyperthermophilic archaeon, Pyrococcus furiosus." RNA **15**(3): 420-431.

Kato, M., M. Shirouzu, T. Terada, H. Yamaguchi, K. Murayama, H. Sakai, S. Kuramitsu and S. Yokoyama (2003). "Crystal structure of the 2'-5' RNA ligase from Thermus thermophilus HB8." J Mol Biol **329**(5): 903-911.

Kelley, L. A. and M. J. Sternberg (2009). "Protein structure prediction on the Web: a case study using the Phyre server." Nat Protoc **4**(3): 363-371.

Kleman-Leyer, K., D. W. Armbruster and C. J. Daniels (1997). "Properties of H. volcanii tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems." Cell **89**(6): 839-847.

Knapp, G., J. S. Beckmann, P. F. Johnson, S. A. Fuhrman and J. Abelson (1978). "Transcription and processing of intervening sequences in yeast tRNA genes." Cell **14**(2): 221-236.

Knapp, G., R. C. Ogden, C. L. Peebles and J. Abelson (1979). "Splicing of yeast tRNA precursors: structure of the reaction intermediates." Cell **18**(1): 37-45.

Koch, M. H., P. Vachette and D. I. Svergun (2003). "Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution." Q Rev Biophys **36**(2): 147-227.

Konarev, P. V., V. V. Volkov, A. V. Sokolova, M. H. J. Koch and D. I. Svergun (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." J Appl Crystallogr **36**(5): 1277-1282.

Konarska, M., W. Filipowicz, H. Domdey and H. J. Gross (1981). "Formation of a 2'-phosphomonoester, 3',5'-phosphodiester linkage by a novel RNA ligase in wheat germ." Nature **293**(5828): 112-116.

Konarska, M., W. Filipowicz and H. J. Gross (1982). "RNA ligation via 2'-phosphomonoester, 3'5'-phosphodiester linkage: requirement of 2',3'-cyclic phosphate termini and involvement of a 5'-hydroxyl polynucleotide kinase." Proc Natl Acad Sci U S A **79**(5): 1474-1478.

Koonin, E. V. and A. E. Gorbalenya (1990). "Related domains in yeast tRNA ligase, bacteriophage T4 polynucleotide kinase and RNA ligase, and mammalian myelin 2',3'-cyclic nucleotide phosphohydrolase revealed by amino acid sequence comparison." FEBS Lett **268**(1): 231-234.

Kozlov, G., A. Y. Denisov, E. Pomerantseva, M. Gravel, P. E. Braun and K. Gehring (2007). "Solution structure of the catalytic domain of RICH protein from goldfish." FEBS J **274**(6): 1600-1609.

Kozlov, G., J. Lee, D. Elias, M. Gravel, P. Gutierrez, I. Ekiel, P. E. Braun and K. Gehring (2003). "Structural evidence that brain cyclic nucleotide phosphodiesterase is a member of the 2H phosphodiesterase superfamily." J Biol Chem **278**(46): 46021-46028.

Kuhsel, M. G., R. Strickland and J. D. Palmer (1990). "An ancient group I intron shared by eubacteria and chloroplasts." Science **250**(4987): 1570-1573.

Kurihara, T., K. Monoh, K. Sakimura and Y. Takahashi (1990). "Alternative splicing of mouse brain 2',3'-cyclic-nucleotide 3'-phosphodiesterase mRNA." Biochem Biophys Res Commun **170**(3): 1074-1081.

Kurihara, T., Y. Takahashi, A. Nishiyama and T. Kumanishi (1988). "cDNA cloning and amino acid sequence of human brain 2',3'-cyclic-nucleotide 3'-phosphodiesterase." Biochem Biophys Res Commun **152**(2): 837-842.

Kurjan, J., B. D. Hall, S. Gillam and M. Smith (1980). "Mutations at the yeast SUP4 tRNATyr locus: DNA sequence changes in mutants lacking suppressor activity." Cell **20**(3): 701-709.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**(5259): 680-685.

Lappe-Siefke, C., S. Goebbels, M. Gravel, E. Nicksch, J. Lee, P. E. Braun, I. R. Griffiths and K. A. Nave (2003). "Disruption of Cnp1 uncouples oligodendroglial functions in axonal support and myelination." Nat Genet **33**(3): 366-374.

Lee, J., M. Gravel, E. Gao, R. C. O'Neill and P. E. Braun (2001). "Identification of essential residues in 2',3'-cyclic nucleotide 3'-phosphodiesterase. Chemical modification and site-directed mutagenesis to investigate the role of cysteine and histidine residues in enzymatic activity." J Biol Chem **276**(18): 14804-14813.

Lee, J., R. C. O'Neill, M. W. Park, M. Gravel and P. E. Braun (2006). "Mitochondrial localization of CNP2 is regulated by phosphorylation of the N-terminal targeting signal by PKC: implications of a mitochondrial function for CNP2 in glial and non-glial cells." Mol Cell Neurosci **31**(3): 446-462.

Lee, M. C. and G. Knapp (1985). "Transfer RNA splicing in Saccharomyces cerevisiae. Secondary and tertiary structures of the substrates." J Biol Chem **260**(5): 3108-3115.

Lees, J. G., A. J. Miles, F. Wien and B. A. Wallace (2006). "A reference database for circular dichroism spectroscopy covering fold and secondary structure space." Bioinformatics **22**(16): 1955-1962.

Lees, J. G., B. R. Smith, F. Wien, A. J. Miles and B. A. Wallace (2004). "CDtool-an integrated software package for circular dichroism spectroscopic data processing, analysis, and archiving." Anal Biochem **332**(2): 285-289.

Li, D., C. Liu, Y. H. Liang, L. F. Li and X. D. Su (2008). "Crystal structure of B. subtilis YjcG characterizing the YjcG-like group of 2H phosphoesterase superfamily." Proteins **72**(3): 1071-1076.

Li, H. and J. Abelson (2000). "Crystal structure of a dimeric archaeal splicing endonuclease." J Mol Biol **302**(3): 639-648.

Li, H., C. R. Trotta and J. Abelson (1998). "Crystal structure and evolution of a transfer RNA splicing enzyme." Science **280**(5361): 279-284.

Li, M. Z. and S. J. Elledge (2007). "Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC." Nat Methods **4**(3): 251-256.

Lobley, A., L. Whitmore and B. A. Wallace (2002). "DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra." Bioinformatics **18**(1): 211-212.

Lykke-Andersen, J., C. Aagaard, M. Semionenkov and R. A. Garrett (1997). "Archaeal introns: splicing, intercellular mobility and evolution." Trends Biochem Sci **22**(9): 326-331.

Lykke-Andersen, J. and R. A. Garrett (1997). "RNA-protein interactions of an archaeal homotetrameric splicing endoribonuclease with an exceptional evolutionary history." EMBO J **16**(20): 6290-6300.
Madhani, H. D. and C. Guthrie (1994). "Dynamic RNA-RNA interactions in the spliceosome." Annu Rev Genet **28**: 1-26.

Marck, C. and H. Grosjean (2003). "Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: evolutionary implications." RNA **9**(12): 1516-1531.

Mathison, L., M. Winey, C. Soref, M. R. Culbertson and G. Knapp (1989). "Mutations in the anticodon stem affect removal of introns from pre-tRNA in Saccharomyces cerevisiae." Mol Cell Biol **9**(10): 4220-4228.

Matsudaira, P. T. and D. R. Burgess (1978). "SDS microslab linear gradient polyacrylamide gel electrophoresis." Anal Biochem **87**(2): 386-396.

Mattoccia, E., I. M. Baldi, D. Gandini-Attardi, S. Ciafrè and G. P. Tocchini-Valentini (1988). "Site selection by the tRNA splicing endonuclease of Xenopus laevis." Cell **55**(4): 731-738.

Mazumder, R., L. M. Iyer, S. Vasudevan and L. Aravind (2002). "Detection of novel members, structure-function analysis and evolutionary classification of the 2H phosphoesterase superfamily." Nucleic Acids Res **30**(23): 5229-5243.

McCraith, S. M. and E. M. Phizicky (1991). "An enzyme from Saccharomyces cerevisiae uses NAD+ to transfer the splice junction 2'-phosphate from ligated tRNA to an acceptor molecule." J Biol Chem **266**(18): 11986-11992.

McGuffin, L. J., K. Bryson and D. T. Jones (2000). "The PSIPRED protein structure prediction server." Bioinformatics **16**(4): 404-405.

Metcalf, W. W. and B. L. Wanner (1993). "Evidence for a fourteen-gene, phnC to phnP locus for phosphonate metabolism in Escherichia coli." Gene **129**(1): 27-32.

Michaelis, L., M. L. Menten, K. A. Johnson and R. S. Goody (2011). "The original Michaelis constant: translation of the 1913 Michaelis-Menten paper." Biochemistry **50**(39): 8264-8269.

Mitchell, M., S. Xue, R. Erdman, L. Randau, D. Söll and H. Li (2009). "Crystal structure and assembly of the functional Nanoarchaeum equitans tRNA splicing endonuclease." Nucleic Acids Res **37**(17): 5793-5802.

Mooij, W. T., E. Mitsiki and A. Perrakis (2009). "ProteinCCD: enabling the design of protein truncation constructs for expression and crystallization experiments." Nucleic Acids Res **37**(Web Server issue): W402-405.

Mori, S., T. Kajita, T. Endo and T. Yoshihisa (2011). "The intron of tRNA-TrpCCA is dispensable for growth and translation of Saccharomyces cerevisiae." RNA **17**(9): 1760-1769.

Mount, S. M. (1982). "A catalogue of splice junction sequences." Nucleic Acids Res **10**(2): 459-472.

Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn and H. Erlich (1986). "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction." Cold Spring Harb Symp Quant Biol **51 Pt 1**: 263-273.

Muraro, P. A., M. Kalbus, G. Afshar, H. F. McFarland and R. Martin (2002). "T cell response to 2',3'-cyclic nucleotide 3'-phosphodiesterase (CNPase) in multiple sclerosis patients." J Neuroimmunol **130**(1-2): 233-242.

Myllykoski, M. (2013). Structure and function of the myelin enzyme 2',3'-cyclic nucleotide 3'-phosphodiesterase. Acta Universitatis Ouluensis A609.

Myllykoski, M., K. Itoh, S. M. Kangas, A. M. Heape, S. U. Kang, G. Lubec, I. Kursula and P. Kursula (2012). "The N-terminal domain of the myelin enzyme 2',3'-cyclic nucleotide 3'-phosphodiesterase: direct molecular interaction with the calcium sensor calmodulin." J Neurochem **123**(4): 515-524.

Myllykoski, M. and P. Kursula (2010). "Expression, purification, and initial characterization of different domains of recombinant mouse 2',3'-cyclic nucleotide 3'-phosphodiesterase, an enigmatic enzyme from the myelin sheath." BMC Res Notes **3**: 12.

Myllykoski, M., A. Raasakka, H. Han and P. Kursula (2012). "Myelin 2',3'-cyclic nucleotide 3'-phosphodiesterase: active-site ligand binding and molecular conformation." PLoS One **7**(2): e32336.

Myllykoski, M., A. Raasakka, M. Lehtimäki, H. Han, I. Kursula and P. Kursula (2013). "Crystallographic Analysis of the Reaction Cycle of 2',3'-Cyclic Nucleotide 3'-Phosphodiesterase, a Unique Member of the 2H Phosphoesterase Family." J Mol Biol.

Mylonas, E. and D. I. Svergun (2007). "Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering." J Appl Crystallogr **40**: s245-s249.

Müller, H. W. (1982). "Dissociation of CNPase dimer into catalytically active monomers by 1,4-dithiothreitol and urea. Protein-promoted reassembly." FEBS Lett **144**(1): 77-80.

Nasr, F. and W. Filipowicz (2000). "Characterization of the Saccharomyces cerevisiae cyclic nucleotide phosphodiesterase involved in the metabolism of ADP-ribose 1",2"-cyclic phosphate." Nucleic Acids Res **28**(8): 1676-1683.

Notredame, C., D. G. Higgins and J. Heringa (2000). "T-Coffee: A novel method for fast and accurate multiple sequence alignment." J Mol Biol **302**(1): 205-217.

Novogrodsky, A. and J. Hurwitz (1966). "The enzymatic phosphorylation of ribonucleic acid and deoxyribonucleic acid. I. Phosphorylation at 5'-hydroxyl termini." J Biol Chem **241**(12): 2923-2932.

Novogrodsky, A., M. Tal, A. Traub and J. Hurwitz (1966). "The enzymatic phosphorylation of ribonucleic acid and deoxyribonucleic acid. II. Further properties of the 5'-hydroxyl polynucleotide kinase." J Biol Chem **241**(12): 2933-2943.

O'Farrell, P. Z., B. Cordell, P. Valenzuela, W. J. Rutter and H. M. Goodman (1978). "Structure and processing of yeast precursor tRNAs containing intervening sequences." Nature **274**(5670): 438-445.

O'Neill, R. C. and P. E. Braun (2000). "Selective synthesis of 2',3'-cyclic nucleotide 3'-phosphodiesterase isoform 2 and identification of specifically phosphorylated serine residues." J Neurochem **74**(2): 540-546.

Odell, M., V. Sriskanda, S. Shuman and D. B. Nikolov (2000). "Crystal structure of eukaryotic DNA ligase-adenylate illuminates the mechanism of nick sensing and strand joining." Mol Cell **6**(5): 1183-1193.

Ogden, R. C., M. C. Lee and G. Knapp (1984). "Transfer RNA splicing in Saccharomyces cerevisiae: defining the substrates." Nucleic Acids Res **12**(24): 9367-9382.

Olafson, R. W., G. I. Drummond and J. F. Lee (1969). "Studies on 2',3'-cyclic nucleotide-3'-phosphohydrolase from brain." Can J Biochem **47**(10): 961-966.

Overton, I. M., G. Padovani, M. A. Girolami and G. J. Barton (2008). "ParCrys: a Parzen window density estimation approach to protein crystallization propensity prediction." Bioinformatics **24**(7): 901-907.

Pantoliano, M. W., E. C. Petrella, J. D. Kwasnoski, V. S. Lobanov, J. Myslik, E. Graf, T. Carver, E. Asel, B. A. Springer, P. Lane and F. R. Salemme (2001). "High-density miniaturized thermal shift assays as a general strategy for drug discovery." J Biomol Screen **6**(6): 429-440.

Paquin, B., A. Heinfling and D. A. Shub (1999). "Sporadic distribution of tRNA(Arg)CCU introns among alpha-purple bacteria: evidence for horizontal transmission and transposition of a group I intron." J Bacteriol **181**(3): 1049-1053.

Paquin, B., S. D. Kathe, S. A. Nierzwicki-Bauer and D. A. Shub (1997). "Origin and evolution of group I introns in cyanobacterial tRNA genes." J Bacteriol **179**(21): 6798-6806.

Paushkin, S. V., M. Patel, B. S. Furia, S. W. Peltz and C. R. Trotta (2004). "Identification of a human endonuclease complex reveals a link between tRNA splicing and pre-mRNA 3' end formation." Cell **117**(3): 311-321.

Pearson, D., I. Willis, H. Hottinger, J. Bell, A. Kumar, U. Leupold and D. Söll (1985). "Mutations preventing expression of sup3 tRNASer nonsense suppressors of Schizosaccharomyces pombe." Mol Cell Biol **5**(4): 808-815.

Peebles, C. L., P. Gegenheimer and J. Abelson (1983). "Precise excision of intervening sequences from precursor tRNAs by a membrane-associated yeast endonuclease." Cell **32**(2): 525-536.

Peebles, C. L., R. C. Ogden, G. Knapp and J. Abelson (1979). "Splicing of yeast tRNA precursors: a two-stage reaction." Cell **18**(1): 27-35.

Peirce, T. R., N. J. Bray, N. M. Williams, N. Norton, V. Moskvina, A. Preece, V. Haroutunian, J. D. Buxbaum, M. J. Owen and M. C. O'Donovan (2006). "Convergent evidence for 2',3'-cyclic nucleotide 3'-phosphodiesterase as a possible susceptibility gene for schizophrenia." Arch Gen Psychiatry **63**(1): 18-24.

Petoukhov, M. V., D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V. Konarev and D. I. Svergun (2012). "New developments in the ATSAS program package for small-angle scattering data analysis." J Appl Crystallogr **45**(2): 342-350.

Petoukhov, M. V., P. V. Konarev, A. G. Kikhney and D. I. Svergun (2007). "ATSAS 2.1 – towards automated and web-supported small-angle scattering data analysis." J Appl Crystallogr **40**: s223-s228.

Petoukhov, M. V. and D. I. Svergun (2005). "Global rigid body modeling of macromolecular complexes against small-angle scattering data." Biophys J **89**(2): 1237-1250.

Petoukhov, M. V. and D. I. Svergun (2013). "Applications of small-angle X-ray scattering to biomacromolecular solutions." Int J Biochem Cell Biol **45**(2): 429-437.

Phizicky, E. M., S. A. Consaul, K. W. Nehrke and J. Abelson (1992). "Yeast tRNA ligase mutants are nonviable and accumulate tRNA splicing intermediates." J Biol Chem **267**(7): 4577-4582.

Phizicky, E. M. and C. L. Greer (1993). "Pre-tRNA splicing: variation on a theme or exception to the rule?" Trends Biochem Sci **18**(1): 31-34.

Phizicky, E. M. and A. K. Hopper (2010). "tRNA biology charges to the front." Genes Dev **24**(17): 1832-1860.

Phizicky, E. M., R. C. Schwartz and J. Abelson (1986). "Saccharomyces cerevisiae tRNA ligase. Purification of the protein and isolation of the structural gene." J Biol Chem **261**(6): 2978-2986.

Pollastri, G., D. Przybylski, B. Rost and P. Baldi (2002). "Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles." Proteins **47**(2): 228-235.

Popow, J., A. Schleiffer and J. Martinez (2012). "Diversity and roles of (t)RNA ligases." Cell Mol Life Sci **69**(16): 2657-2670.

Porod, G (1982). General Theory. In Small Angle X-ray Scattering (ed. O. Glatter and O. Kratky) Academic Press, London 17-51.

Porath, J., J. Carlsson, I. Olsson and G. Belfrage (1975). "Metal chelate affinity chromatography, a new approach to protein fractionation." Nature **258**(5536): 598-599.

Porath, J. and P. Flodin (1959). "Gel filtration: a method for desalting and group separation." Nature **183**(4676): 1657-1659.

Raghavan, R. and M. F. Minnick (2009). "Group I introns and inteins: disparate origins but convergent parasitic strategies." J Bacteriol **191**(20): 6193-6202.

Raines, R. T. (1998). "Ribonuclease A." Chem Rev **98**(3): 1045-1066.

Ramirez, A., S. Shuman and B. Schwer (2008). "Human RNA 5'-kinase (hClp1) can function as a tRNA splicing enzyme in vivo." RNA **14**(9): 1737-1745.

Rasband, M. N., J. Tayler, Y. Kaga, Y. Yang, C. Lappe-Siefke, K. A. Nave and R. Bansal (2005). "CNP is required for maintenance of axon-glia interactions at nodes of Ranvier in the CNS." Glia **50**(1): 86-90.

Rasmussen, L. C., C. L. Oliveira, O. Byron, J. M. Jensen, J. S. Pedersen, H. U. Sperling-Petersen and K. K. Mortensen (2011). "Structure and dimerization of translation initiation factor aIF5B in solution." Biochem Biophys Res Commun **416**(1-2): 140-145.

Rauhut, R., P. R. Green and J. Abelson (1990). "Yeast tRNA-splicing endonuclease is a heterotrimeric enzyme." J Biol Chem **265**(30): 18180-18184.

Rehse, P. H. and T. H. Tahirov (2005). "Structure of a putative 2'-5' RNA ligase from Pyrococcus horikoshii." Acta Crystallogr D Biol Crystallogr **61**(Pt 9): 1207-1212.

Reinhold-Hurek, B. and D. A. Shub (1992). "Self-splicing introns in tRNA genes of widely divergent bacteria." Nature **357**(6374): 173-176.

Reyes, V. M. and J. Abelson (1988). "Substrate recognition and splice site determination in yeast tRNA splicing." Cell **55**(4): 719-730.

Richardson, C. C. (1965). "Phosphorylation of nucleic acid by an enzyme from T4 bacteriophage-infected Escherichia coli." Proc Natl Acad Sci U S A **54**(1): 158-165.

Roth-Cross, J. K., H. Stokes, G. Chang, M. M. Chua, V. Thiel, S. R. Weiss, A. E. Gorbalenya and S. G. Siddell (2009). "Organ-specific attenuation of murine hepatitis virus strain A59 by replacement of catalytic residues in the putative viral cyclic phosphodiesterase ns2." J Virol **83**(8): 3743-3753.

Rudi, K. and K. S. Jakobsen (1997). "Cyanobacterial tRNA(Leu)(UAA) group I introns have polyphyletic origin." FEMS Microbiol Lett **156**(2): 293-298.

Sakamoto, Y., N. Tanaka, T. Ichimiya, T. Kurihara and K. T. Nakamura (2004). "Three-dimensional structure of a cyclic-nucleotide phosphodiesterase from human brain." Nucleic Acids Symp Ser (Oxf)(48): 157-158.

Sakamoto, Y., N. Tanaka, T. Ichimiya, T. Kurihara and K. T. Nakamura (2005). "Crystal structure of the catalytic fragment of human brain 2',3'-cyclic-nucleotide 3'-phosphodiesterase." J Mol Biol **346**(3): 789-800.

Sawaya, R., B. Schwer and S. Shuman (2003). "Genetic and biochemical analysis of the functional domains of yeast tRNA ligase." J Biol Chem **278**(45): 43928-43938.

Sawicki, S. G. and D. L. Sawicki (1998). "A new model for coronavirus transcription." Adv Exp Med Biol **440**: 215-219.

Schultz, J., R. R. Copley, T. Doerks, C. P. Ponting and P. Bork (2000). "SMART: a web-based tool for the study of genetically mobile domains." Nucleic Acids Res **28**(1): 231-234.

Schwartz, R. C., C. L. Greer, P. Gegenheimer and J. Abelson (1983). "Enzymatic mechanism of an RNA ligase from wheat germ." J Biol Chem **258**(13): 8374-8383.

Schwer, B., A. Aronova, A. Ramirez, P. Braun and S. Shuman (2008). "Mammalian 2',3' cyclic nucleotide phosphodiesterase (CNP) can function as a tRNA splicing enzyme in vivo." RNA **14**(2): 204-210.

Sharp, P. A. (1981). "Speculations on RNA splicing." Cell **23**(3): 643-646.

Sharp, P. A., B. Sugden and J. Sambrook (1973). "Detection of two restriction endonuclease activities in Haemophilus parainfluenzae using analytical agarose--ethidium bromide electrophoresis." Biochemistry **12**(16): 3055-3063.

Slabinski, L., L. Jaroszewski, L. Rychlewski, I. A. Wilson, S. A. Lesley and A. Godzik (2007). "XtalPred: a web server for prediction of protein crystallizability." Bioinformatics **23**(24): 3403-3405.

Sogin, D. C. (1976). "2',3'-Cyclic NADP as a substrate for 2',3'-cyclic nucleotide 3'-phosphohydrolase." J Neurochem **27**(6): 1333-1337.

Soma, A., A. Onodera, J. Sugahara, A. Kanai, N. Yachie, M. Tomita, F. Kawamura and Y. Sekine (2007). "Permuted tRNA genes expressed via a circular RNA intermediate in Cyanidioschyzon merolae." Science **318**(5849): 450-453.

Spinelli, S. L., S. A. Consaul and E. M. Phizicky (1997). "A conditional lethal yeast phosphotransferase (tpt1) mutant accumulates tRNAs with a 2'-phosphate and an undermodified base at the splice junction." RNA **3**(12): 1388-1400.

Spinelli, S. L., H. S. Malik, S. A. Consaul and E. M. Phizicky (1998). "A functional homolog of a yeast tRNA splicing enzyme is conserved in higher eukaryotes and in Escherichia coli." Proc Natl Acad Sci U S A **95**(24): 14136-14141.

Sprinkle, T. J. (1989). "2',3'-cyclic nucleotide 3'-phosphodiesterase, an oligodendrocyte-Schwann cell and myelin-associated enzyme of the nervous system." Crit Rev Neurobiol **4**(3): 235-301.

Stange, N., H. J. Gross and H. Beier (1988). "Wheat germ splicing endonuclease is highly specific for plant pre-tRNAs." EMBO J **7**(12): 3823-3828.

Stingo, S., M. Masullo, E. Polverini, C. Laezza, I. Ruggiero, R. Arcone, E. Ruozi, F. Dal Piaz, A. M. Malfitano, A. M. D'Ursi and M. Bifulco (2007). "The N-terminal domain of 2',3'-cyclic nucleotide 3'-phosphodiesterase harbors a GTP/ATP binding site." Chem Biol Drug Des **70**(6): 502-510.

Studier, F. W. (2005). "Protein production by auto-induction in high density shaking cultures." Protein Expr Purif **41**(1): 207-234.

Sugahara, J., K. Kikuta, K. Fujishima, N. Yachie, M. Tomita and A. Kanai (2008). "Comprehensive analysis of archaeal tRNA genes reveals rapid increase of tRNA introns in the order thermoproteales." Mol Biol Evol **25**(12): 2709-2716.

Sugahara, J., N. Yachie, K. Arakawa and M. Tomita (2007). "In silico screening of archaeal tRNA-encoding genes having multiple introns with bulge-helix-bulge splicing motifs." RNA **13**(5): 671-681.

Svergun, D. I. (1992). "Determination of the regularization parameter in indirect-transform methods using perceptual criteria." J Appl Crystallogr **25**(4): 495-503.

Svergun, D. I. (1999). "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing." Biophys J **76**(6): 2879-2886.

Swerdlow, H. and C. Guthrie (1984). "Structure of intron-containing tRNA precursors. Analysis of solution conformation using chemical and enzymatic probes." J Biol Chem **259**(8): 5197-5207.

Szweykowska-Kulinska, Z., B. Senger, G. Keith, F. Fasiolo and H. Grosjean (1994). "Intron-dependent formation of pseudouridines in the anticodon of Saccharomyces cerevisiae minor tRNA(Ile)." EMBO J **13**(19): 4636-4644.

Tanaka, N. and S. Shuman (2011). "RtcB is the RNA ligase component of an Escherichia coli RNA repair operon." J Biol Chem **286**(10): 7727-7731.

Thomas, R. C. (1993). "11 Structure and Mechanism of the Large Catalytic RNAs: Group I and Group II Introns and Ribonuclease P." Cold Spring Harbor Monograph Archive; Volume 24 (1993): The RNA World.

Thompson, L. D., L. D. Brandon, D. T. Nieuwlandt and C. J. Daniels (1989). "Transfer RNA intron processing in the halophilic archaebacteria." Can J Microbiol **35**(1): 36-42.

Thompson, L. D. and C. J. Daniels (1988). "A tRNA(Trp) intron endonuclease from Halobacterium volcanii. Unique substrate recognition properties." J Biol Chem **263**(34): 17951-17959.

Thøgersen, H. C., H. R. Morris, K. N. Rand and M. J. Gait (1985). "Location of the adenylylation site in T4 RNA ligase." Eur J Biochem **147**(2): 325-329.

Tocchini-Valentini, G. D., P. Fruscoloni and G. P. Tocchini-Valentini (2005). "Coevolution of tRNA intron motifs and tRNA endonuclease architecture in Archaea." Proc Natl Acad Sci U S A **102**(43): 15418-15422.

Torchia, C., Y. Takagi and C. K. Ho (2008). "Archaeal RNA ligase is a homodimeric protein that catalyzes intramolecular ligation of single-stranded RNA and DNA." Nucleic Acids Res **36**(19): 6218-6227.

Trotta, C. R., F. Miao, E. A. Arn, S. W. Stevens, C. K. Ho, R. Rauhut and J. N. Abelson (1997). "The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases." Cell **89**(6): 849-858.

Trotta, C. R., S. V. Paushkin, M. Patel, H. Li and S. W. Peltz (2006). "Cleavage of pre-tRNAs by the splicing endonuclease requires a composite active site." Nature **441**(7091): 375-377.

Valadkhan, S. (2005). "snRNAs as the catalysts of pre-mRNA splicing." Curr Opin Chem Biol **9**(6): 603-608.

Valenzuela, P., A. Venegas, F. Weinberg, R. Bishop and W. J. Rutter (1978). "Structure of yeast phenylalanine-tRNA genes: an intervening DNA segment within the region coding for the tRNA." Proc Natl Acad Sci U S A **75**(1): 190-194.

Volkov, V. V. and D. I. Svergun (2003). "Uniqueness of ab initio shape determination in small-angle scattering." J Appl Crystallogr **36**(3-1): 860-864.

Walker, J. E., M. Saraste, M. J. Runswick and N. J. Gay (1982). "Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold." EMBO J **1**(8): 945-951.

Walter, T. S., C. Meier, R. Assenberg, K. F. Au, J. Ren, A. Verma, J. E. Nettleship, R. J. Owens, D. I. Stuart and J. M. Grimes (2006). "Lysine methylation as a routine rescue strategy for protein crystallization." Structure **14**(11): 1617-1622.

Wang, C., U. Neugebauer, J. Bürck, M. Myllykoski, P. Baumgärtel, J. Popp and P. Kursula (2011). "Charge isomers of myelin basic protein: structure and interactions with membranes, nucleotide analogues, and calmodulin." PLoS One **6**(5): e19915.

Wang, L. K., C. K. Ho, Y. Pei and S. Shuman (2003). "Mutational analysis of bacteriophage T4 RNA ligase 1. Different functional groups are required for the nucleotidyl transfer and phosphodiester bond formation steps of the ligation reaction." J Biol Chem **278**(32): 29454-29462.

Wang, L. K., C. D. Lima and S. Shuman (2002). "Structure and mechanism of T4 polynucleotide kinase: an RNA repair enzyme." EMBO J **21**(14): 3873-3880.

Wang, L. K., B. Schwer, M. Englert, H. Beier and S. Shuman (2006). "Structure-function analysis of the kinase-CPD domain of yeast tRNA ligase (Trl1) and

requirements for complementation of tRNA splicing by a plant Trl1 homolog." Nucleic Acids Res **34**(2): 517-527.

Wang, L. K. and S. Shuman (2001). "Domain structure and mutational analysis of T4 polynucleotide kinase." J Biol Chem **276**(29): 26868-26874.

Wang, L. K. and S. Shuman (2002). "Mutational analysis defines the 5'-kinase and 3'-phosphatase active sites of T4 polynucleotide kinase." Nucleic Acids Res **30**(4): 1073-1080.

Wang, L. K. and S. Shuman (2005). "Structure-function analysis of yeast tRNA ligase." RNA **11**(6): 966-975.

Wattenhofer, M., K. Shibuya, J. Kudoh, R. Lyle, J. Michaud, C. Rossier, K. Kawasaki, S. Asakawa, S. Minoshima, A. Berry, B. Bonne-Tamir, N. Shimizu, S. E. Antonarakis and H. S. Scott (2001). "Isolation and characterization of the UBASH3A gene on 21q22.3 encoding a potential nuclear protein with a novel combination of domains." Hum Genet **108**(2): 140-147.

Westaway, S. K., H. G. Belford, B. L. Apostol, J. Abelson and C. L. Greer (1993). "Novel activity of a yeast ligase deletion polypeptide. Evidence for GTP-dependent tRNA splicing." J Biol Chem **268**(4): 2435-2443.

Westaway, S. K., E. M. Phizicky and J. Abelson (1988). "Structure and function of the yeast tRNA ligase gene." J Biol Chem **263**(7): 3171-3176.

Westhof, E. (2002). "Group I introns and RNA folding." Biochem Soc Trans **30**(Pt 6): 1149-1152.

Willis, I., H. Hottinger, D. Pearson, V. Chisholm, U. Leupold and D. Söll (1984). "Mutations affecting excision of the intron from a eukaryotic dimeric tRNA precursor." EMBO J **3**(7): 1573-1580.

Wilson, S. J., J. W. Schoggins, T. Zang, S. B. Kutluay, N. Jouvenet, M. A. Alim, J. Bitzegeio, C. M. Rice and P. D. Bieniasz (2012). "Inhibition of HIV-1 particle assembly by 2',3'-cyclic-nucleotide 3'-phosphodiesterase." Cell Host Microbe **12**(4): 585-597.

Xu, M. Q., S. D. Kathe, H. Goodrich-Blair, S. A. Nierzwicki-Bauer and D. A. Shub (1990). "Bacterial origin of a chloroplast intron: conserved self-splicing group I introns in cyanobacteria." Science **250**(4987): 1566-1570.

Xu, Q., D. Teplow, T. D. Lee and J. Abelson (1990). "Domain structure in yeast tRNA ligase." Biochemistry **29**(26): 6132-6138.

Xue, S., K. Calvin and H. Li (2006). "RNA recognition and cleavage by a splicing endonuclease." Science **312**(5775): 906-910.

Yoshihisa, T., C. Ohshima, K. Yunoki-Esaki and T. Endo (2007). "Cytoplasmic splicing of tRNA in Saccharomyces cerevisiae." Genes Cells **12**(3): 285-297.

Yoshihisa, T., K. Yunoki-Esaki, C. Ohshima, N. Tanaka and T. Endo (2003). "Possibility of cytoplasmic pre-tRNA splicing: the yeast tRNA splicing endonuclease mainly localizes on the mitochondria." Mol Biol Cell **14**(8): 3266-3279.

Yoshinari, S., T. Shiba, D. K. Inaoka, T. Itoh, G. Kurisu, S. Harada, K. Kita and Y. Watanabe (2009). "Functional importance of crenarchaea-specific extra-loop revealed by an X-ray structure of a heterotetrameric crenarchaeal splicing endonuclease." Nucleic Acids Res **37**(14): 4787-4798.

Zillmann, M., M. A. Gorovsky and E. M. Phizicky (1991). "Conserved mechanism of tRNA splicing in eukaryotes." Mol Cell Biol **11**(11): 5410-5416.

Zofallova, L., Y. Guo and R. Gupta (2000). "Junction phosphate is derived from the precursor in the tRNA spliced by the archaeon Haloferax volcanii cell extract." RNA **6**(7): 1019-1030.

9. Appendix

9.1. Risk and Safety Statements

Following is the list of potentially hazardous materials, and respective hazard and precautionary statements as introduced by the Globally Harmonised System of Classification and Labeling of Chemicals (GHS).

Compound	Chemical Abstracts Service No.	Hazard statements	GHS hazard	Precautionary statements
Ampicillin	69-52-3	H334, H317	GHS08	P280, P261, P302+P352, P342+P311
β -mercaptoethanol	60-24-2	H301, H310, H330, H315, H318, H410	GHS05, GHS06, GHS09	P280, P273, P302+P352, P304+P340, P305+P351+P338, P310
$CaCl_2$	10043-52-4	H319	GHS07	P280, P305+P351+P338
Chloramphenicol	56-75-7	H350, H361	GHS08	P281, P308+P313, P501
Citric acid	5949-29-1	H319	GHS07	P305+P351+P338
DTT	3483-12-3	H302, H315, H319, H335	GHS07	P280, P301+P312, P302+P352, P403+P233
EDTA	60-00-4	H319	GHS07	P264, P280, P305+P351+P338
Ethanol	64-17-5	H225	GHS02	P210
Ethidium bromide	1239-45-8	H302, H330, H341	GHS06, GHS08	P260, P281, P284, P310
Hydrocholoric acid > 25%	7647-01-0	H314, H335	GHS05, GHS07	P261, P280, P310, P305+P351+P338
Imidazole	288-32-4	H301, H314, H361	GHS05, GHS06, GHS08	P260, P281, P303+P361+P353, P301+P330+P331, P305+P351+P338, P308+P313

Isopropanol	67-63-0	H225, H319, H336	GHS02, GHS07	P210, P233, P305+P351+P338
Kanamycin	25389-94-0	H360	GHS08	P281, P260, P308+P313
NaOH	1310-73-2	H290, H314	GHS05	P280, P303+P361+P353, P301+P330+P331, P305+P351+P338, P310, P406
$NiSO_4$	10101-97-0	H302, H332, H315, H334, H317, H341, H350, H360, H372, H410	GHS07, GHS08, GHS09	P280, P201, P302+P352,P308+P313, P342+P311
PMSF	329-98-6	H301, H314	GHS05, GHS06	P280, P305+P351+P338, P310
RNase A	9001-99-4	H334	GHS08	P261, P304+P341
SDS	151-21-3	H228, H302, H311, H315, H319, H335	GHS02, GHS06	P210, P261, P280, P312, P305+P351+P338
TCEP	51805-45-9	H314		P280, P305+P351+P338, P310
Tris	1185-53-1	H315, H319, H335	GHS07	P261, P305+P351+P338

9.1.1. GHS hazard statements

H225	Highly flammable liquid and vapor
H228	Flammable solid
H290	May be corrosive to metals
H301	Toxic if swallowed
H302	Harmful if swallowed
H310	Fatal in contact with skin
H311	Toxic in contact with skin
H314	Causes severe skin burns and eye damage
H315	Causes skin irritation
H317	May cause an allergic skin reaction
H318	Causes serious eye damage
H319	Causes serious eye irritation
H330	Fatal if inhaled
H332	Harmful if inhaled
H334	May cause allergy or asthma symptoms or breathing difficulties if inhaled
H335	May cause respiratory irritation
H336	May cause drowsiness or dizziness
H341	Suspected of causing genetic defects
H350	May cause cancer
H360	May damage fertility or the unborn child
H361	Suspected of damaging fertility or the unborn child
H372	Causes damage to organs through prolonged or repeated exposure
H410	Very toxic to aquatic life with long lasting effects

9.1.2. GHS precautionary statements

P201	Obtain special instructions before use
P210	Keep away from heat/sparks/open flames/hot surfaces – No smoking
P233	Keep container tightly closed
P260	Do not breathe dust/fume/gas/mist/vapors/spray
P261	Avoid breathing dust/fume/gas/mist/vapors/spray
P264	Wash….. thoroughly after handling
P273	Avoid release to the environment
P280	Wear protective gloves/protective clothing/eye protection/face protection
P281	Use personal protective equipment as required
P284	Wear respiratory protection
P310	Immediately call a POISON CENTER or doctor/physician
P312	Call a POISON CENTER or doctor/physician if you feel unwell
P406	Store in a corrosive resistant/…..container with a resistant inner liner
P501	Dispose of contents/container to ….
P301+P312	IF SWALLOWED: Call a POISON CENTER or doctor/ physician if you feel unwell
P302+P352	IF ON SKIN: Wash with soap and water
P304+P340	IF INHALED: Remove victim to fresh air and keep at rest in a position comfortable for breathing
P304+P341	IF INHALED: If breathing is difficult, remove victim to fresh air and keep at rest in a position comfortable for breathing
P308+P313	IF exposed or concerned: Get medical advice/attention

P342+P311 IF experiencing respiratory symptoms: Call a POISON CENTER or doctor/physician

P403+P233 Store in a well-ventilated place. Keep container tightly closed

P301+P330+P331 IF SWALLOWED. Rinse mouth. Do NOT induce vomiting

P303+P361+P353 IF ON SKIN (or hair): Remove/take off immediately all contaminated clothing. Rinse skin with water/shower

P305+P351+P338 IF IN EYES: Rinse cautiously with water for several minutes. Remove contact lenses if present and easy to do – continue rinsing

9.1.3. GHS and hazard symbols

Hazard symbols for formulations and respective risk labels (according to Health and Safety Executive UK, http://www.hse.gov.uk)

GHS pictograms (Source: United Nations Economic Commission for Europe, http://www.unece.org)

9.2. Buffers used in the high-throughput thermal stability assay

	1	2	3	4	5	6	7	8	9	10	11	12
A	**Sodium acetate pH 3.5**				**MES pH 6.0**				**Imidazole pH 8.5**			
	0 mM NaCl	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**	**0 mM NaCl**	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**	**0 mM NaCl**	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**
B	**Sodium acetate pH 4.0**				**MES pH 6.5**				**HEPES pH 7.5**			
	0 mM NaCl	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**	**0 mM NaCl**	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**	**0 mM NaCl + 1 mM cNADP$^+$**	**50 mM NaCl + 1 mM cNADP$^+$**	**150 mM NaCl + 1 mM cNADP$^+$**	**500 mM NaCl + 1mM cNADP$^+$**
C	**Citric acid pH 4.0**				**Imidazole pH 6.5**							
	0 mM NaCl	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**	**0 mM NaCl**	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**				
D	**Sodium acetate pH 4.5**				**MOPS pH 6.5**				**HEPES pH 7.5**			
	0 mM NaCl	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**	**0 mM NaCl**	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**	**0 mM NaCl + 1 mM NADP$^+$**	**50 mM NaCl + 1 mM NADP$^+$**	**150 mM NaCl + 1 mM NADP$^+$**	**500 mM NaCl + 1 mM NADP$^+$**
E	**Sodium acetate pH 5.0**				**MES pH 7.0**							
	0 mM NaCl	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**	**0 mM NaCl**	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**				
F	**Citric acid pH 5.0**				**HEPES pH 7.0**				**HEPES pH 7.5**			
	0 mM NaCl	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**	**0 mM NaCl**	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**	**0 mM NaCl + 1 mM cCMP**	**50 mM NaCl + 1 mM cCMP**	**150 mM NaCl + 1 mM cCMP**	**500 mM NaCl + 1 mM cCMP**
G	**Citric acid pH 5.5**				**HEPES pH 7.5**							
	0 mM NaCl	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**	**0 mM NaCl**	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**				
H	**Sodium acetate pH 6.0**				**Imidazole pH 8.0**							
	0 mM NaCl	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**	**0 mM NaCl**	**50 mM NaCl**	**150 mM NaCl**	**500 mM NaCl**				

10. Scientific communication

Poster

Parts of the results of this study were presented as a poster at the following conference:

International Symposium on Structural and Infection Biology
Hamburg, July 1-2, 2011

Gopinath Muruganandam, Matti Myllykoski, Petri Kursula and Inari Kursula

"Characterization of invertebrate and vertebrate 2',3'-cyclic nucleotide 3'-phosphodiesterase (CNPase)"

Publication

The current study is part of a large collaborative project, and the results of this study will be published together with the results from other collaborating laboratories.

11. Curriculum vitae

- omitted for privacy reasons -

12. Acknowledgements

First and foremost, I wish to express my deepest gratitude to my supervisor Dr. Inari Kursula for her esteemed and valuable guidance over the course of my PhD. I am equally thankful to my co-supervisor Dr. Petri Kursula for constructive discussions and fruitful collaborations. Right from day one in Hamburg, the extent of support they both have offered me is really enormous, and throughout my PhD, I felt so fortunate to be their student and would always love to be one.

Our lab atmosphere truly resembles that of a family. Our daily-lunch discussions, frequent get-togethers, and lab retreats have made our lab life the healthiest possible. The freedom to disturb anyone at any time has allowed me to learn a lot of new things from almost every member of our lab. I thank my friends Dr. Juha Kallio, Moon Chatterjee, Esa-Pekka Kumpula, Saara Laulumaa, Dr. Huijong Han, Susanne Meier, Jani Salmivaara, Emilia Salmivaara, Nele Vervaet, Dr.Thorsten Mengesdorf, Benjamin Götte, Manuela Mirow and the trainees from vocational schools for providing an excellent working environment.

Our group members based in Oulu, Finland have always been willing to assist me wherever they can. I thank Dr. Alexander Ignatev, Bhargav Prabhakar, Juha Vahokoski, Markku Soronen, Dr. Salla Ruskamo and Maryna Chucklieb for their valuable inputs and kind support. Special thanks are due to Dr. Matti Myllykoski and Arne Raasakka for their close-collaboration during the course of this study.

I wish to thank Prof. Dieter Söll's research group in the Yale University for providing us the expression plasmids to begin with this study. I also thank the Protein Facility of the Netherlands Cancer Institute in Amsterdam for providing us cloning vectors.

I would like to thank the management of Centre for Structural Systems Biology (CSSB) - Helmholtz Centre for Infection Research (HZI) for state-of-the-art laboratory resources. The administrative support from the Personal Department of HZI (Braunschweig) is gratefully acknowledged. In this aspect, I also want to thank the International Office of Deutsches Elektronen-Synchrotron (Hamburg) for administrative assistance on-campus.

I sincerely acknowledge the financial support from the German Federal Ministry of Education and Research.

I thank all my wonderful friends for their constant encouragement. Last, but definitely not the least, I thank my parents for their love, affection, patience and support.

13. Erklärung

Erklärung über frühere Promotionsversuche

Hiermit erkläre ich, Gopinath Muruganandam, dass vorher keine weiteren Promotionsversuche unternommen worden sind, oder an einer anderen Stelle vorgelegt wurden.

Hamburg, 14.3.2014

Gopinath Muruganandam

Eidesstattliche Versicherung

Hiermit erkläre ich an Eides Statt, dass die vorliegende Dissertationsschrift selbstständig und allein von mir unter den angegebenen Hilfsmitteln angefertigt wurde.

Hamburg, 14.3.2014

Gopinath Muruganandam