Dissertation

submitted to the
Combined Faculties for the Natural Sciences and Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Diplom-biologist Aytac Dikfidan
born in Bremen, Germany

Oral examination: _____

Structural and functional characterization of Clp1, a eukaryotic RNA-specific polynucleotide kinase

Thesis Reviewers:
Prof. Dr. Ilme Schlichting
Dr. Anton Meinhart

Table of contents

Summary

Polynucleotide kinases (PNKs) are crucial enzymes involved in DNA and RNA repair, RNA maturation, as well as in RNA degradation processes. These enzymes have a conserved PNK domain, showing structural homology to the classical fold of P-loop kinases. PNKs catalyse the transfer of the γ-phosphate group of ATP molecule to the 5'-hydroxyl group of polynucleotide substrates. Depending on their *in vivo* function, PNKs show different substrate specificity. Interestingly, eukaryotic PNKs have recently been identified that specifically phosphorylate RNA substrates. These novel RNA-specific PNKs constitute the Clp1 subfamily of PNKs named after their first identified member. Human Clp1 was shown to participate in various RNA maturation pathways: (i) cleavage and polyadenylation of RNA polymerase II pre-mRNA transcripts, (ii) tRNA-splicing, and (iii) phosphorylation of synthetic siRNAs. Despite extensive studies on Clp1, the structural elements involved in RNA-specificity and the mechanism of the phosphoryl transfer reaction have remained elusive so far.

This thesis, therefore, aims for the structural and functional characterization of the RNA-specific Clp1 PNK. During this work, the three-dimensional structure of Clp1 from *Caenorhabditis elegans* was described for the first time at atomic resolution. Clp1 is a multi-domain protein that consists of a central PNK domain sandwiched by additional N- and C-terminal domains. Clp1 was crystallized with various RNA substrates that differ in length and sequence. Based on these results, the structural features of Clp1's RNA-specificity were elucidated. Clp1 uses an "RNA-sensor" that recognizes the 2'-hydroxyl group of the RNA at the ultimate position. Additionally, Clp1 was also crystallized in enzymatically relevant states such as an inhibited substrate bound state, a transition state analog and a product bound state. A general model for enzyme catalysis of PNKs was derived from these structures. In contrast to other described PNKs, Clp1's ATP-binding site within the PNK domain is obstructed by the N-terminal domain. The crystal structures as well as activity assays with truncated variants of Clp1 showed a contribution of the N-terminal domain to ATP-binding by interactions with the nucleobase. The phosphate groups of the ATP molecule are anchored in the active site tunnel, which is formed by the common P-loop motif, a divalent metal cofactor (Mg^{2+}), and an α-helical LID module. Structure-guided mutational analysis identified the essential role of the Walker A lysine for enzyme catalysis. Moreover, Clp1 crystal structures revealed a non-canonical Walker A lysine in an "arrested" conformation that acts as a molecular switch. Activation of the switch is only achieved in the transition state complex. In

contrast to other nucleotide kinases, Clp1 seems to apply a substrate-gating mechanism that prevents futile ATP hydrolysis. In this context, the classical Walker A lysine seems to have a so far underestimated regulatory function. Such a molecular switch mechanism of the Walker A lysine is not restricted to Clp1 exclusively, since the PDB database provides a significant number of crystal structures showing a similar “arrested“ conformation of the Walker A lysine. Thus, an additional function of the Walker A lysine as a molecular switch in enzyme catalysis is suggested. In conclusion, this thesis provides the first crystal structures of Clp1, elucidating its RNA-specificity as well as the phosphoryl transfer reaction mechanism.

Zusammenfassung

Polynukleotidkinasen (PNKasen) sind essentielle Enzyme, die in DNA- und RNA-Reparatur-, RNA-Reifungs- und RNA-Abbau-Prozessen beteiligt sind. Diese Enzyme besitzen eine konservierte PNKase Domäne, die strukturelle Homologie zu den klassischen P-loop Kinasen aufweist. Der Reaktionsmechanismus der PNKasen beinhaltet den Transfer der γ-Phosphatgruppe von ATP zu der 5'-Hydroxylgruppe eines Polynukleotid-Substrats. Abhängig von ihrer Funktion können PNKasen unterschiedliche Substratspezifitäten besitzen. Interessanterweise wurden erst kürzlich eukaryotische PNKasen identifiziert, die spezifisch RNA phosphorylieren. Diese einzigartigen RNA-spezifischen PNKasen sind Teil einer Subfamilie der PNKasen, der sogenannten Clp1 Proteinfamilie. Clp1 war die erste bekannte RNA-spezifische PNKase in Eukaryoten und wurde zum Namensgeber dieser neuen Proteinfamilie. Es wurde gezeigt, dass humanes Clp1 an verschiedenen RNA-Reifungsprozessen beteiligt ist: (i) dem Schneiden und die Polyadenylierung von RNA Polymerase II unreifen mRNA Transkripten, (ii) dem tRNA Splicing, und (iii) der Phosphorylierung synthetischer siRNAs. Trotz intensiver Studien an Clp1 ist nur wenig Mechanistisches über die RNA-Spezifität und den Reaktionsmechanismus des Phosphoryl-Transfers von Clp1 bekannt.

Es war daher das Ziel dieser Doktorarbeit, Clp1 als RNA-spezifische PNKase sowohl strukturell als auch funktionell detailliert zu untersuchen. Während dieser Arbeit wurde die erste dreidimensionale Kristallstruktur von Clp1 aus *Caenorhabditis elegans* mit atomarer Auflösung bestimmt. Clp1 ist ein Multidomänenenzym mit einer zentralen PNKase Domäne, flankiert von einer N- und C-terminalen Domäne. Auf der Grundlage diverser RNA substratetratgebundener Strukturen, die sich in Länge und ihrer Seqeuenz der RNA unterschieden, wurde ein „RNA-Sensor" identifiziert, welcher spezifisch die 2'-Hydroxylgruppe des endständigen Nukleotids erkennt. Zusätzlich konnten in dieser Arbeit die enzymatisch relevanten Konformationen des substratgebundenen Zustandes, des Übergangszustandes und des produktgebundenen Zustandes beschrieben werden. Mit Hilfe dieser Strukturen wurde ein allgemeingültiges Modell des Phosphoryl-Transfer Reaktionsmechanismus von PNKasen erarbeitet. Interessanterweise ist im Unterschied zu bereits bekannten PNKasen die ATP-Bindungsstelle innerhalb der PNKase-Domäne durch die zusätzliche N-terminale Domäne abgeschirmt. Sowohl die strukturellen Ergebnisse als auch biochemische Analysen in dieser Doktorarbeit konnten zeigen, dass die N-terminale Domäne entscheidend für die ATP-Bindung in Clp1 ist. Die Phosphatgruppen des ATP Moleküls sind

im katalytischen Zentrum verankert. Das katalytische Zentrum wird von dem bekannten P-loop Motiv, einem divalenten Metall-Cofaktor (Mg^{2+}) und durch ein sogenanntes α-helikales „LID-Modul“ geformt. Mittels ortsgerichteter Mutagenesestudien konnte die essenzielle Rolle des Walker A Lysins für die Enzymkatalyse gezeigt werden. Darüber hinaus zeigt die Kristallstruktur von Clp1 eine „arretierte“ Konformation des Walker A Lysins. Das Lysin fungiert in Clp1 als ein molekularer Schalter, der erst im Übergangszustand vollständig aktiviert wird. Im Gegensatz zu anderen Nukleotidkinasen scheint Clp1 eine Art „Gating-Mechanismus“ zu besitzen, welcher nicht-prozessive ATP-Hydrolyse verhindern kann. Diese „arretierte“ Konformation des Walker A Lysins übernimmt damit eine zusätzliche regulatorische Aufgabe in der Enzymkatalyse. Eine detaillierte Analyse der PDB Datenbank identifizierte eine signifikante Anzahl an Kristallstrukturen die ebenfalls ein Walker A Lysin in „arretierter“ Konformation besitzen. Mit den Ergebnissen dieser Doktorarbeit wird die Funktion des klassischen Walker A Lysins auf die eines molekularen Schalters in der Enzymkatalyse erweitert. Insgesamt führten die strukturellen und biochemischen Analysen über Clp1 zu neuen Erkenntnissen hinsichtlich der RNA-Spezifität als auch des Phosphoryl-Transfer-Reaktionsmechanismus.

Publications

The results of this dissertation are provided in the following manuscript:

Dikfidan A, Loll L, Clausen T, and Meinhart A (2013). Mechanism of RNA 5'-phosphorylation by eukaryotic RNA Polynucleotide Kinases. *Manuscript in progress.*

Additional publications:

Lüddeke F, **Dikfidan A**, Harder J (2012). Physiology of deletion mutants in the anaerobic β-myrcene degradation pathway in Castellaniella defragrans. *BMC Microbiol.* 4;12:192.

Lüddeke F, Wülfing A, Timke M, Germer F, Weber J, **Dikfidan A**, Rahnfeld T, Linder D, Meyerdierks A, Harder J (2012). Geraniol and geranial dehydrogenases induced in anaerobic monoterpene degradation by Castellaniella defragrans. *Appl Environ Microbiol.* Apr;78(7):2128-36.

1 Introduction

In nature, the chemistry of the 5'-terminus of nucleotides has a strong impact on the fate of DNA and RNA molecules. DNA and RNA synthesis reactions rely on nucleoside triphosphates, the building block of the nucleotide metabolism[1,2]. Nucleoside triphosphates originate from precursor nucleosides that are activated in a cascade of phosphorylation reactions[3]. Thus, newly synthesized polynucleotides typically bear a triphosphate at their 5'-termini. The 5'-triphosphate is chemically stable and persists until its removal or enzymatic modifications[4,5]. Interestingly, the phosphorylation state of the 5'-terminus is important in various cellular processes. In RNA degradation, the 5'-terminus is specifically recognized by both prokaryotic and eukaryotic exonucleases[6,7]. In both cases mRNA decay was initiated after conversion of the 5′-triphosphate (or m^7GpppN (7-methylguanosine) cap in eukaryotes) to a 5'-monophosphate (5'-phosphate)[8,9], whereas in case of mRNA with a 5'-triphosphate or a 5'-hydroxyl group, mRNA decay was impaired[10-16]. Furthermore, the 5'-terminus is also important for enzymes of DNA and RNA repair pathways[17,18]. These repair enzymes require a 5'-phosphate group for an efficient enzymatic activity. Damaged nucleotides, however, are characterized by a 5'-hydroxyl terminus that needs to be enzymatically processed for efficient repair. Because of the biological relevance of a signature 5'-phophate for the nucleotide metabolism, a protein family involved in the processing of the 5'-hydroxyl groups is emphasized in this thesis. Polynucleotide kinases (PNKs) enzymatically alter the 5'-phosphorylate state of polynucleotides to a 5'-phosphate. In contrast to the knowledge about PNKs in the DNA metabolism, our understanding of their function in RNA maturation and degradation pathways still needs to be improved considerably. However, the recent identification of novel RNA-specific PNKs (RPNKs) has led to new insights into the field of RNA research. Structural and functional characterizations of these RPNKs are very limited and need to be the subject of future work.

1.1 Importance of Polynucleotide kinases in the nucleotide metabolism

In general, PNKs have an evolutionary conserved reaction mechanism that catalyzes one of the most frequent reactions in biological systems[19]: a phosphoryl transfer reaction by

which the γ-phosphate group of a nucleoside triphosphate is transferred to a plethora of acceptor molecules[20]. In case of PNKs, the γ-phosphoryl group of an ATP molecule is transferred to DNA and RNA polynucleotide substrates (Figure 1.1.1). As mentioned previously, the phosphorylation state of nucleotides is important in various cellular processes and the following sections will describe in detail the role of PNKs in the nucleotide metabolisms and the importance of the signature 5'-phosphate in RNA degradation.

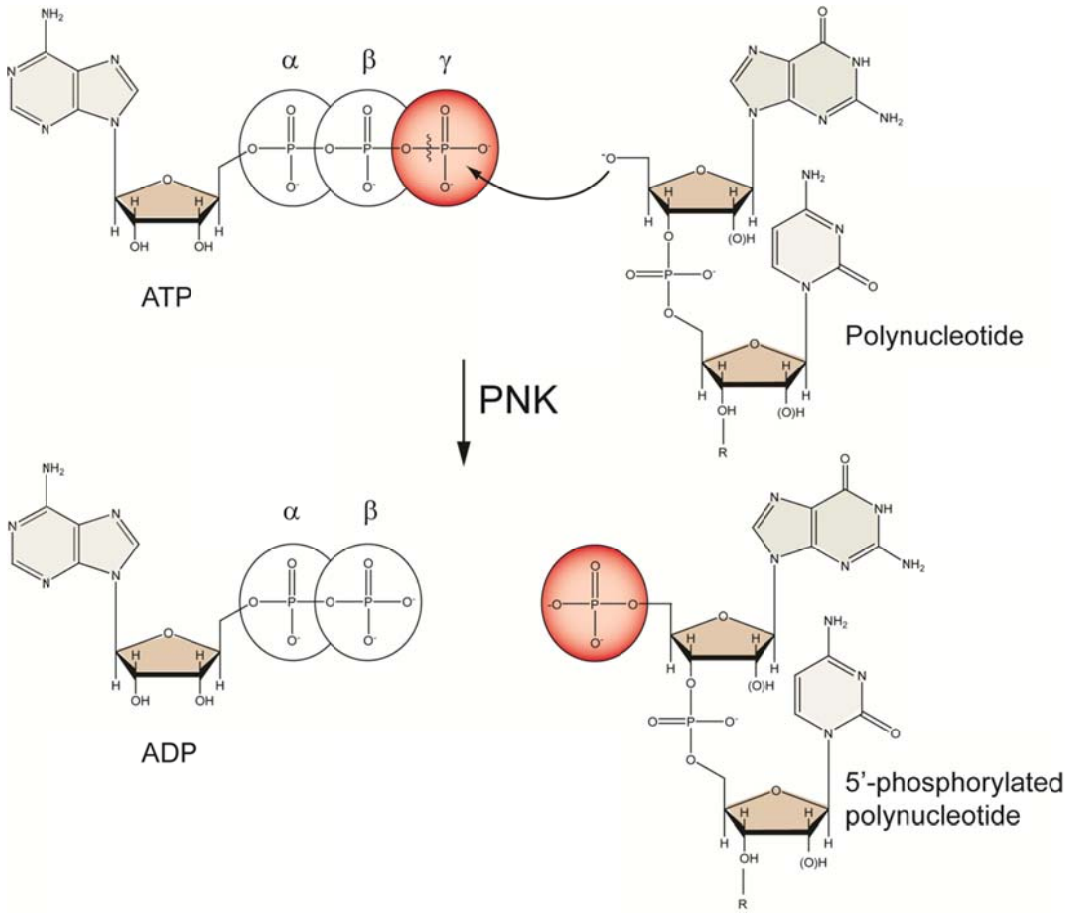


Figure 1.1.1 Schematic representation of the 5'-phosphoryl transfer reaction catalyzed by PNKs. The γ-phosphate of an ATP molecule is transferred to deprotonated 5'-hydroxyl termini of DNA or RNA polynucleotide substrates.

1.1.1 PNKs in DNA repair pathways

Accurate maintenance and errorless reproduction of the genome is a prerequisite for cell survival[21]. However, damage of cellular DNA is an unavoidable threat to genomic integrity that is implicated in the etiology of many diseases[22-25] and represents a major factor in aging[26]. DNA damage is caused by various physical and chemical stress conditions such as

ionizing radiation[27] or reactive oxygen species[28]. However, DNA damage is not necessarily linked to stress conditions, since it can also arise during regular cellular processes such as DNA replication, recombination, or differentiation[29-31].

Damaged DNA is characterized by a heterogeneous class of lesions including base modifications, base excision, and strand breaks[29]. Exposure of cells to ionizing radiation for instance induces elevated levels of hydroxyl radicals, which are chemicals suspected to generate DNA single strand breaks (SSB; Figure 1.1.2 A)[32]. Without end-healing, SSB can turn into the most lethal form of DNA damage, the double strand break (DSB; Figure 1.1.2 A)[33]. Other sources for DNA lesions are a stalled topoisomerase I after camptothecin treatment (Figure 1.1.2 A)[34] or nucleases, such as DNase II[35,36]. Many of these lesions leave DNA termini that are unsuitable for subsequent extension or ligation reactions mediated by DNA polymerases and ligases. Incompatible DNA-termini are composed of 3′-phosphate and free 5′-hydroxyl termini[17]. Repair enzymes however, require a free 3′-hydroxyl group for a subsequent elongation reaction and, in case of a ligation reaction, additional 5′-phosphate termini[17].

To ensure genetic stability, cells rely on a battery of different repair mechanisms that are counteracting DNA lesions (Figure 1.1.2 A). The mammalian PNK (mPNK) is a multi-functional enzyme that possesses the capacity to both phosphorylate 5′-hydroxyl and dephosphorylate 3′-phosphate termini[37,38]. The mPNK is considered as a key component for DNA repair contributing to three different pathways: Base excision repair[39], DNA SSB repair[40] and non-homologous end-joining[41].

Structurally, mPNK is a multi-domain enzyme consisting of an N-terminal FHA (forkhead-associated) domain and two spatially separated catalytic domains; a 3'-phosphatase and a PNK domain[42]. The FHA domain recognizes phospho-threonine residues on specific target proteins[42-44] and thereby recruits mPNK to distinctive repair complexes[45]. This active recruitment of mPNK and other DNA repair enzymes protects cells against the release of premature intermediates. Interaction partners of the FHA domain are the two scaffold proteins, XRCC1 and XRCC4. The SSB repair complex (mPNK, DNA polymerase β, and DNA ligase III) is organized by XRCC1[43], whereas DSB repair complex (mPNK, DNA end-binding protein Ku70/80, protein kinase DNA-PK and the XRCC4/ligase IV heterodimer) depends on XRCC4[44]. As a key factor for DNA repair, the mPNK appears as the recurrent linker in the distinctive DNA repair pathways.

1.1.2 PNKs in RNA repair pathways

The first PNK was discovered 1965 in bacteriophage-infected *Escherichia coli* [46-48]. Since then the T4 PNK has become one of the best-characterized PNKs and it is widely used for applications in molecular biology[49,50] Similar to mPNK, the T4 phage-encoded PNK is a multifunctional enzyme composed of an N-terminal PNK domain and a C-terminal phosphatase domain[51,52]. Together with the T4 ligase[53], it is shown to be involved in RNA repair, enabling the T-even phage to overcome the host's suicide defense mechanism[18] (Figure 1.1.2 B). In detail, if bacteria encounter a T4 phage infection, the Stp DNA restriction inhibitory peptide is activated, which subsequently activates a latent tRNA anti-codon nuclease[54-56]. The active nuclease triggers the lesion of the bacterial lysine tRNA, inhibits protein synthesis and thereby phage propagation[18]. The cleaved tRNA molecules have a characteristic free 5'-hydroxyl terminus and, in contrast to DNA repair processes, a 2',3'-cyclic phosphate group[18]. In order to provide acceptable substrates for T4 ligase, T4 PNK phosphorylates the 5'-hydroxyl terminus[46] and removes the cyclic phosphate group at the 2',3'-terminus in a two-stage process[18,57,58]. Subsequently, T4 ligase can seal the broken 5'- and 3'-termini of the tRNA[59]. Notably, T4 PNK and mPNK are functionally similar, but they have different polynucleotide specificity. The enzyme displays the broadest substrate specificity on DNA and RNA molecules compared to other described PNKs[60-63].

Recently, another RNA repair system was also discovered in bacteria repairing ribotoxin-cleaved RNAs[64]. Ribotoxins are proteins responsible for the site-specific cleavage of RNA molecules involved in transcription and transition[65-67]. In contrast to classical RNA repair pathways relaying on 5'-phosphorylation and 3'-dephosphorylation, this novel system additionally methylates the 2'-hydroxyl group of RNA[64] (Figure 1.1.2 B). Methylated RNA sites are thereby protected against repeated cleavage. The enzymatic reaction involves the bacterial Pnkp/Hen1 complex[68]. Bacterial PNK, Pnkp, is composed of three catalytic domains, an N-terminal PNK domain, a central phosphatase domain, and a C-terminal ligase domain[69], all being required for RNA repair[70]. The presence of RPNKs in bacteria emphasizes the relevance of RNA repair in the prokaryotic nucleotide metabolism. This fact might also be true for a eukaryotic pathway. This hypothesis became evident just recently with the discovery of several RPNKs, which are presented in section 1.2.

1.1.3 PNKs in RNA maturation and RNA turnover

Besides their functions in DNA and RNA repair mechanism, RPNKs are also involved in tRNA maturation[71]. Intron sequences in tRNA precursor molecules (pre-tRNAs) are found in all three kingdoms of life[72]. To obtain a mature tRNA these intron sequences have to be removed by splicing from the pre-tRNAs[73]. After cleavage by the tRNA-splicing endonuclease complex, the intron-cleaved tRNA are composed of a free 5'-hydoxyl group at the 3'-exon and a 2',3'-cyclic phosphate group at the 5'-exon[71]. To become rejoined into a mature tRNA molecule, the 5'- and 2',3'-termini have to be processed by a multi-functional tRNA ligase, named Trl1[74] that consists of three enzymatic activities in functionally independent domains, a cyclic phosphodiesterase domain, a PNK domain, and an RNA ligase domain[75,76]. Trl1 is able to first open the 2'-3'-cyclic phosphate[74], than to phosphorylate the 5'-hydroxyl terminus[74], and finally to rejoin both exon halves to provide a mature tRNA molecule[74]. Interestingly, no homolog Trl1 tRNA ligase was identified in higher eukaryotes, suggesting that Trl1 is a distinctive enzyme of yeast tRNA maturation pathway (Figure 1.1.2 C).

However, the signature 5'-phosphate is not only important for sealing reactions by ligases. RNA turnover is also controlled by the phosphorylation state of nucleotides at the 5'-terminus[8-16]. In contrast to prokaryotes, the mRNA of eukaryotes is modified at the 5'-terminus, where methylated GMP is incorporated to form the m^7GpppN cap[5]. Despite these differences in 5'-capping, the processes in mRNA decay are remarkably similar between eukaryotes and prokaryotes. In both cases, the 5′-phosphate represents the preferential substrate for nucleases in RNA turnover by exonucleases[10-16]. In eukaryotes, prior to the RNA decay the 5'-cap is hydrolyzed to a 5'-phosphate[77,78] Once the 3′-poly(A) tail is removed, the mRNA becomes degraded from the 5′-terminus by the 5′-3′ exonuclease Xrn1[79,80] (Figure 1.1.2 C). Furthermore, the signature 5'-phosphate couples the specificity of Xrn1 to its processivity[7]. In prokaryotes, the 5'-triphosphate of the mRNA is hydrolyzed to 5'-phoshate[8] and the mRNA is subsequently committed to degradation by RNase E, an enzyme that recognized the signature 5'-phosphate[6] (Figure 1.1.2 C). In addition, nucleases involved in RNA interference[81] (RNAi) also depend on the signature 5'-phosphate. Two prominent examples of enzymes that recognize the 5'-phosphate are the RNA induced silencing complex (RISC)[82] (Figure 1.1.2 C) and the DICER[83] nuclease. Strinkingly, a signature 5'-phosphate is also generated by enzymatic

cleavage of RISC[84]. Thus, the molecular basis of the signature 5'-phosphate has a central role in the nucleotide metabolism that ensures controlled decay of RNA molecules.

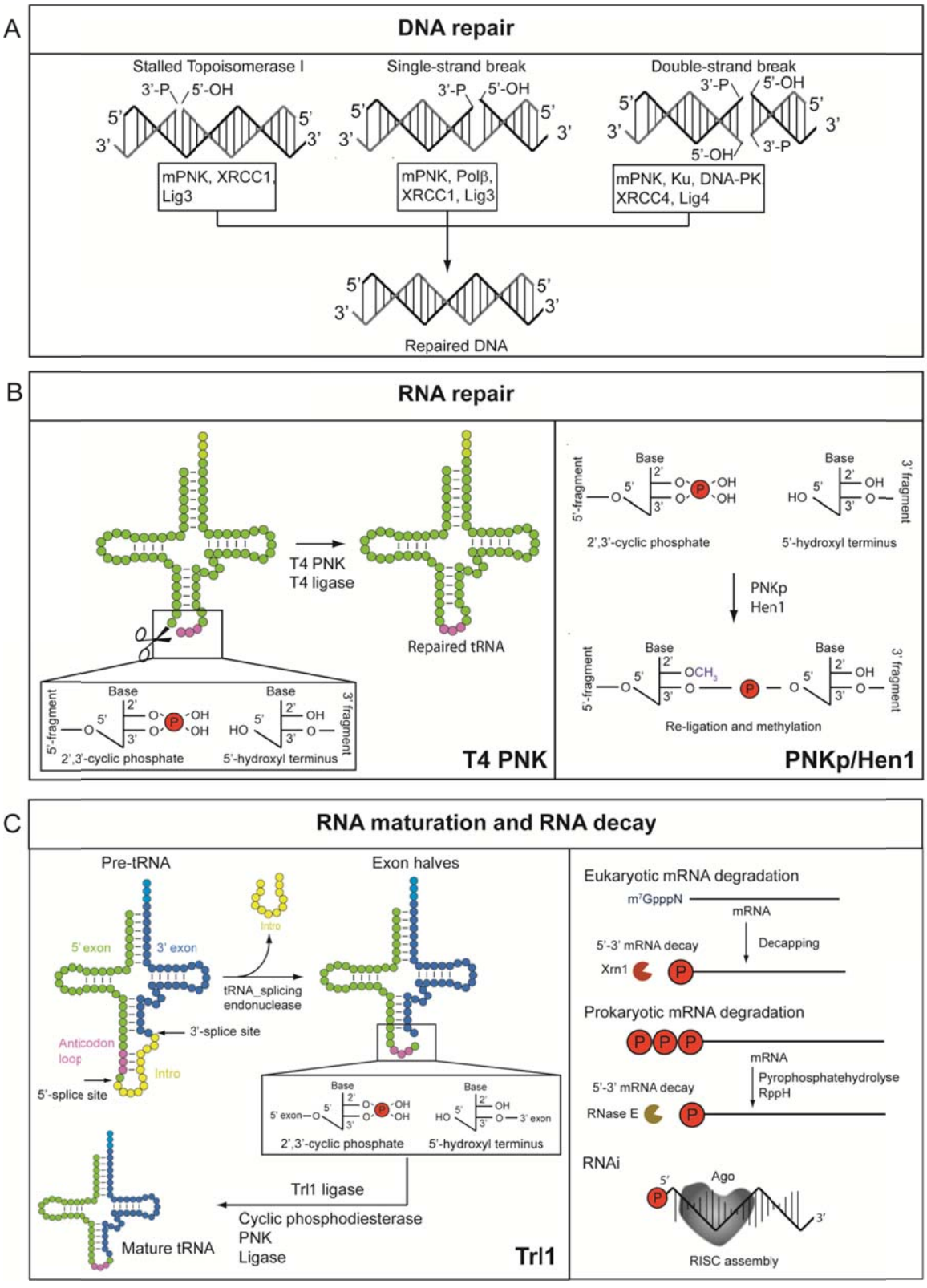


Figure 1.1.2. PNK involved in the nucleotide metabolism. A, mPNKs are a recurrent linker for repair of different lesion types. **B,** T4 PNK together with T4 ligase heals lesions of tRNA to overcome the host defense mechanism. PNKp-Hen1 complex is involved in a special RNA repair pathway that methylates the RNA after ligation to confer protective immunity. **C,** Trl1 is a tri-functional RNA ligase associated with tRNA maturation. The 5'-kinase activity of Trl1 is important to re-ligation of the 5'- and 3'-exons to produce a mature tRNA. Efficient mRNA decay relies on the presence of a signature 5'-phosphate. The exonucleases Xrn1 and RNase E only show efficient enzymatic activity on mRNA substrates with 5'-phosphate group. mRNAs are committed to degradation by decapping or removing the triphosphate group (RppH). In addition, the signature 5'-phosphate group is also required for siRNAs and miRNAs to become incorporated into RISC. Adapted from[17,71].

1.2 Clp1, a novel eukaryotic RPNK

Since eukaryotic RNA-maturation and RNA-degradation processes strongly rely on the presence of 5'-phosphorylated nucleotides, the existence of RPNKs in eukaryotic organisms was conceivable. Although this particular enzymatic reaction was already described in 1979 in HeLa cell extracts[85], characterization of the responsible enzymes started only in 2007[86-88]. Human Clp1 (hsClp1) is the first identified eukaryotic RPNK and is involved in different RNA maturation pathways; (i) in cleavage/polyadenylation of RNA polymerase II transcripts[89,90], (ii) in tRNA maturation[52,90-92], and (iii) in 5'-phosphorylation of synthetic siRNA molecules[88,93].

Eukaryotic Clp1 orthologs display a high degree of evolutionary conservation[94]. Interestingly, Clp1-like PNKs were also identified in plants, archaea and are predicted in prokaryotes[86,95,96], thereby covering all three kingdoms of life. Based on their sequence conservation, these homologous proteins can be grouped into a novel subfamily of PNKs, the Clp1 protein family. Despite their structural and sequence similarities, not all eukaryotic Clp1 orthologs show 5'-kinase activity. Paradoxically, in contrast to hsClp1, Clp1 from *Saccharomyces cerevisiae* (scClp1) appears to be enzymatically inactive[94] and functionally different from hsClp1. Consistent with this observation is the fact that hsClp1 is unable to compensate for the loss of scClp1[97]. Although enzymatically inactive, scClp1 is still able to bind ATP molecules[94,98-100]. Thus, it was suggested that ATP-binding might be required for structural stability of the protein[98-100].

Since archaeal Clp1 from *Pyrococcus horikoshii* (phClp1) is enzymatically active[95], it is conceivable that 5'-kinase activity is ancestral to the Clp1 protein family. scClp1, therefore, represents a variant that has lost its enzymatic activity due to distinctive features of the tRNA maturation pathway in yeast (see 1.1.3). Furthermore, it is anticipated that

scClp1 solely functions in pre-mRNA 3′-terminus processing as an integral subunit of the 3'-cleavage polyadenylation machinery[98-100], whereas hsClp1 represents an RPNK that is involved in multiple RNA maturation pathways as a recurrent linker[88,90,92,97].

1.2.1 hsClp1's involvement in RNA metabolism

(i) mRNA processing. Clp1 was originally identified as a member of the eukaryotic mRNA 3'-cleavage and polyadenylation machinery[89,101,102] (Figure 1.2.1 A). Components of this machinery are structurally conserved in eukaryotes[103]. scClp1 is part of the CF IA complex, and has been shown to interact Cleavage–Polyadenylation Factor (CPF)[98]. In mammals, Clp1 can be found as a subunit of the CF IIm, suggesting a function as a bridging factor between CF Im and CPSF[89]. The interaction between Clp1 and Pcf11 another subunit of the cleavage factor, is evolutionary conserved between scClp1 and hsClp1 orthologs[103].

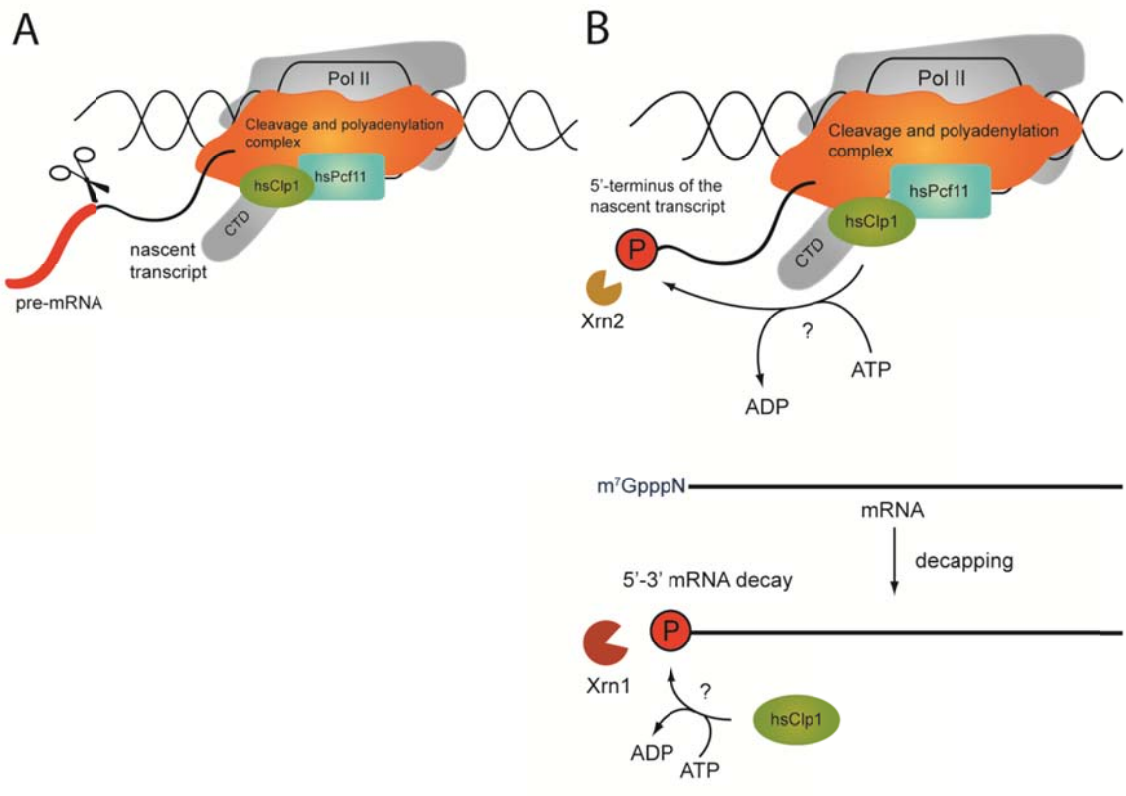


Figure 1.2.1: Proposed function of hsClp1 in mRNA 3'-end processing. A, hsClp1 is part of the 3'-end processing machinery and required for cleavage and polyadenylation of pre-mRNA. There it acts as a bridging factor and interactions Pcf11. **B,** hsClp1 might has a function in transcription termination or mRNA decay by maintaining the signature 5'-phosphate group on the RNA. The 5'-nascent transcript already contains a terminal 5'-phosphate after cleavage. This phosphate is required for dissociation of the Pol II from the DNA by Xrn2 "torpedo model"[104]. hsClp1 is suggested to maintain this phosphate. Furthermore, hsClp1 might have an additional function in maintenance of the 5'-phosphate of mRNAs committed to degradation by Xrn1. Adapted from[92].

After identification of hsClp1's RNA-specific 5'-kinase activity, Martinez and Weitzer suggested the involvement of hsClp1 in mRNA degradation[88,92] (Figure 1.2.1 B). Although the 5'-phosphate is important for efficient degradation by the exonucleases Xrn1[7] and Xrn2[15], both mRNA and its nascent transcript show no free 5'-hydroxyl termini[5,105,106]. Therefore, it was assumed that Clp1 is involved in maintaining the phosphorylation state of the 5'-terminus to counteract putative phosphatases[92]. This hypothesis, however, is in contrast to recent results showing that the 5'-kinase activity of hsClp1 is dispensable for mRNA 3′-end cleavage[90]. Similar to scClp1, hsClp1 seems to function only as a scaffold protein in the mRNA 3'- cleavage and polyadenylation machinery. Consistent with this assumption is the oberservation that the generation of a *Clp1*-knockout produced a lethal phenotype in mice, but a kinase deficient mutant (K127A) was viable[90].

(ii) tRNA processing. The endonuclease complex for tRNA-splicing is associated with hsClp1[91]. Since hsClp1 was additionally shown to 5'-phosphorylate intron cleaved tRNAs[88], it is likely that this enzyme is involved in tRNA-splicing (Figure 1.2.2). Currently, there are several postulated tRNA-splicing pathways[71]. However, all of them are based on the same underlying principle that splicing of pre-tRNA occurs in two steps: cleavage and rejoining of exon halves[71]. Characteristic ends for this cleavage reaction are a 5'-hydroxyl terminus and a 2',3'-cyclic phosphate, respectively[71]. Apart from the yeast ligation pathway, mammals harbor a distinctive pathway that relies on a 2',3'-cyclic phosphate ligase named HSPC117[107]. This enzyme rejoins the splice site junction by ligation of the 2',3'-cyclic phosphate with a free 5'-hydroxyl group and consequently a 5'-kinase activity is obsolete for this mechanism[71]. However, in addition to the HSPC117-dependent pathway, a second "yeast-like" pathway is hypothesized, involving a PNK (hsClp1[88,92]), a cyclic phosphodesterase[108], and phosphotransferease[109]. However, 5',3'-RNA ligase is still missing[110]. The yeast-like ligation pathway also explains the

functional difference between scClp1[94] and hsClp1[88] since yeast Trl1 already contains an additional RNA-specific 5'-kinase activity[76].

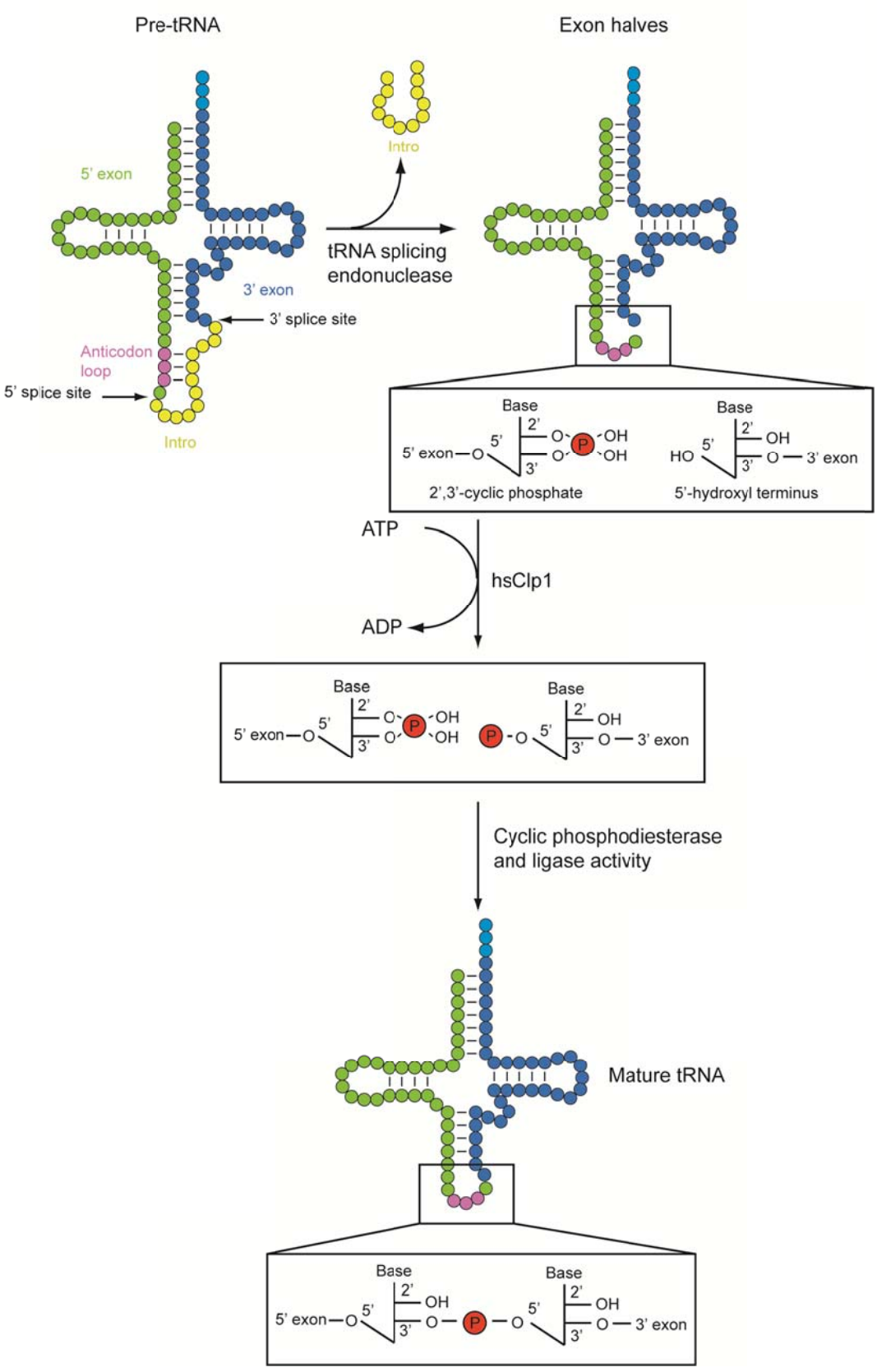


Figure 1.2.2: Proposed function of hsClp1 in tRNA-splicing. Intron sequences of pre-tRNAs are removed by enzymatic activity of the tRNA-splicing endonuclease complex. hsClp1 was shown to be associated to this complex[91]. The cleavage reaction leads to a 2',3' cyclic phosphate group at the 5'-exon and a 5'-hydroxyl group at the 3'-exon. Similar to the yeast ligation pathway (Trl1-like), hsClp1 phosphorylates the 5'-terminus of intron-cleaved 3'-exons. In combination with a cyclic phosphodiesterase and RNA ligase activity, the 3'- and 5'-exon halves can be rejoined to obtained a mature tRNA. The remaining 2'-phosphate can be removed by the activity of a 2′-phosphotransferase. Adapted from[92].

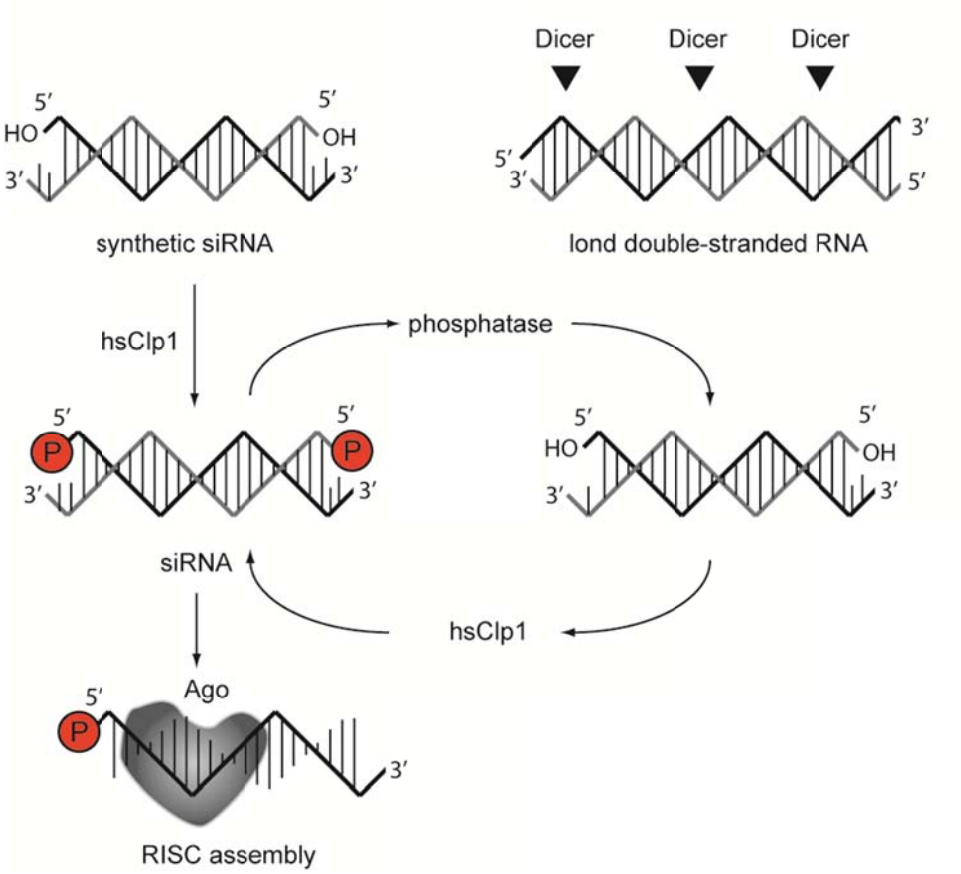


Figure 1.2.3: Proposed function of hsClp1 in RNAi. hsClp1 was shown to 5'-phosphorylate synthetic siRNAs. This signature 5'-phosphate group on the RNA is required for RISC assembly and subsequent gene silencing. Although Dicer-cleaved siRNAs are generated with a 5'-phosphate, hsClp1 may have a function in the maintenance the phosphate to counteract putative RNA phosphatases. The protection of the signature 5'-phosphate ensures efficient RISC assembly and stability of an Ago2-siRNA complex. Adapted from[92].

Consistent with this hypothesis are results from Remirez and colleagues showing that expression of hsClp1 in budding yeast can complement conditional and lethal mutations in the kinase module of yeast or plant tRNA ligases[97].

In contrast, *trans*-complementation was impossible with overexpressed scClp1 or kinase-defective hsClp1[97]. These results support a putative role of hsClp1 in

5'-terminus-healing similar to the yeast-like tRNA-splicing pathway[92]. Furthermore, recent experiments with transgenic kinase-defective Clp1 mice identified a novel tyrosine pre-tRNA as a substrate of Clp1's 5'-kinase activity[90]. Thus, Clp1 from higher eukaryotes seems to be involved in the phosphorylation of tRNA 3'-exon halves.

(iii) RNAi. Besides its function in the 3'-terminus processing machinery[89] and its role in tRNA processing[88,90,92], hsClp1 was shown to 5'-phosphorylate synthetic siRNAs *in vitro*[88,93] (Figure 1.2.3). The signature 5'-phosphate is required for the incorporation of siRNA and miRNAs into RISC[111]. Although siRNAs and microRNAs already contain a 5'-phosphate group[111,112], Weitzer and Martinez speculated whether hsClp1 or other PNKs (Nol9) are involved in the maintenance of this phosphate groups and thereby counteract putative phosphatases[92]. Recent results showed, however that Clp1's 5'-kinase activity is not essential for gene silencing by miRNAs[90].

In conclusion, hsClp1 participates in multiple RNA maturation pathways[88,90,92]. However, the 5'-kinase activity seems to be negligible for the function in the 3'-end processing of pre-mRNA[90]. Both scClp1 and hsClp1 are rather functioning as a structural scaffold for mRNA maturation. But in contrast to scClp1, hsClp1 is additionally part of the tRNA-splicing machinery[91] and involved in tRNA maturation[88,90,92].

1.2.2 Structural characterization of scClp1

The three-dimensional structure of the enzymatically inactive scClp1 in association with a fragment of its CF IA interaction partner Pcf11 was determined by X-ray crystallography at medium resolution (Figure 1.2.4 A and B)[94]. This analysis revealed a multi-domain protein composed of a central PNK domain, sandwiched between additional N- and C-terminal domains (NtD and CtD). The NtD (amino-acids 1 to 100) is composed of a β-sheet sandwich, whereas the PNK domain (amino-acids 101 to 341) contains the active site region of the protein. The PNK domain adopts a classical α,β fold that is similar to the previously described T4 PNK[51,52] and the DNA-specific mammalian PNK[42] (Figure 1.2.4 C). Characteristic for the PNK domain are two evolutionarily conserved motifs, the P-loop and the Walker B motif[113,114]. The P-loop is typically found in adenine or guanine triphosphate binding proteins[19] and it corroborates scClp1's ability to bind ATP molecules. In contrast to T4 PNK and mPNK, the ATP binding site of scClp1 is obstructed by the

additional NtD[94] (Figure 1.2.4 C). Furthermore, the amino acid sequences as well as the structural architecture of the proteins suggest that scClp1 and T4 PNK or mPNK have not evolved from a common ancestor. The CtD (amino-acids 342 to 446) is folded in a mixture of α-helices, β-strands and random coils.

1.2.3 The RNA-specific Clp1 protein family

Based on recent bioinformatical characterizations of hsClp1, the Clp1 protein family can be extended by two newly identified homologs, Nol9 and Grc3[86,87]. These closely related proteins also contain a highly conserved P-loop motif, a Walker B motif and were shown to be RPNKs[86,87]. Nol9 and Grc3 are widely conserved eukaryotic proteins and display an extensive sequence similarity to the PNK domain of other members of the Clp1 protein family, whereas the additional NtD and CtD are less conserved. Interestingly, in the archaeal Clp1 homolog the NtD is even missing[95]. All biochemically characterized members (hsClp1, phClp1, Nol9 and Grc3) of the Clp1 protein family were shown to phosphorylate a variety of different polynucleotide substrates *in vitro*, having preference for single-stranded RNA as well as double-stranded RNA and DNA substrates[86-88,95]. Therefore, the structural features of the PNK domain seem to define the *in vitro* specificity of the polynucleotide substrate. Despite of this promiscuous RNA-phosphorylation, members of the Clp1 protein family contribute specifically to different RNA maturation pathways. Exemplarily, Nol9 and Grc3 were shown to be involved in rRNA maturation[86,87,115,116], whereas hsClp1 is described as a recurrent linker in mRNA and tRNA processing[88,90,92]. This observation might be explained by regulatory functions of NtD and CtD. As known for mPNK, the N-terminal FHA domain serves as a protein-protein interaction platform that recruits mPNK to its molecular target[42]. This domain-dependent recruitment would explain the distinguishable *in vivo* substrate specificities of hsClp1 compared to Nol9 and Grc3. Furthermore, scClp1 was shown to interact with a variety of proteins involved in 3'-end processing, specifically with its central domain and the additional NtD and CtD[98]. However, it remains unclear if these additional domains also play a role in enzyme specificity for the enzymatically active members of the Clp1 protein family. Therefore, the function of additional domains and the identification of putative interaction partners will be of future interest to understand the *in vivo* function of RPNKs in more detail.

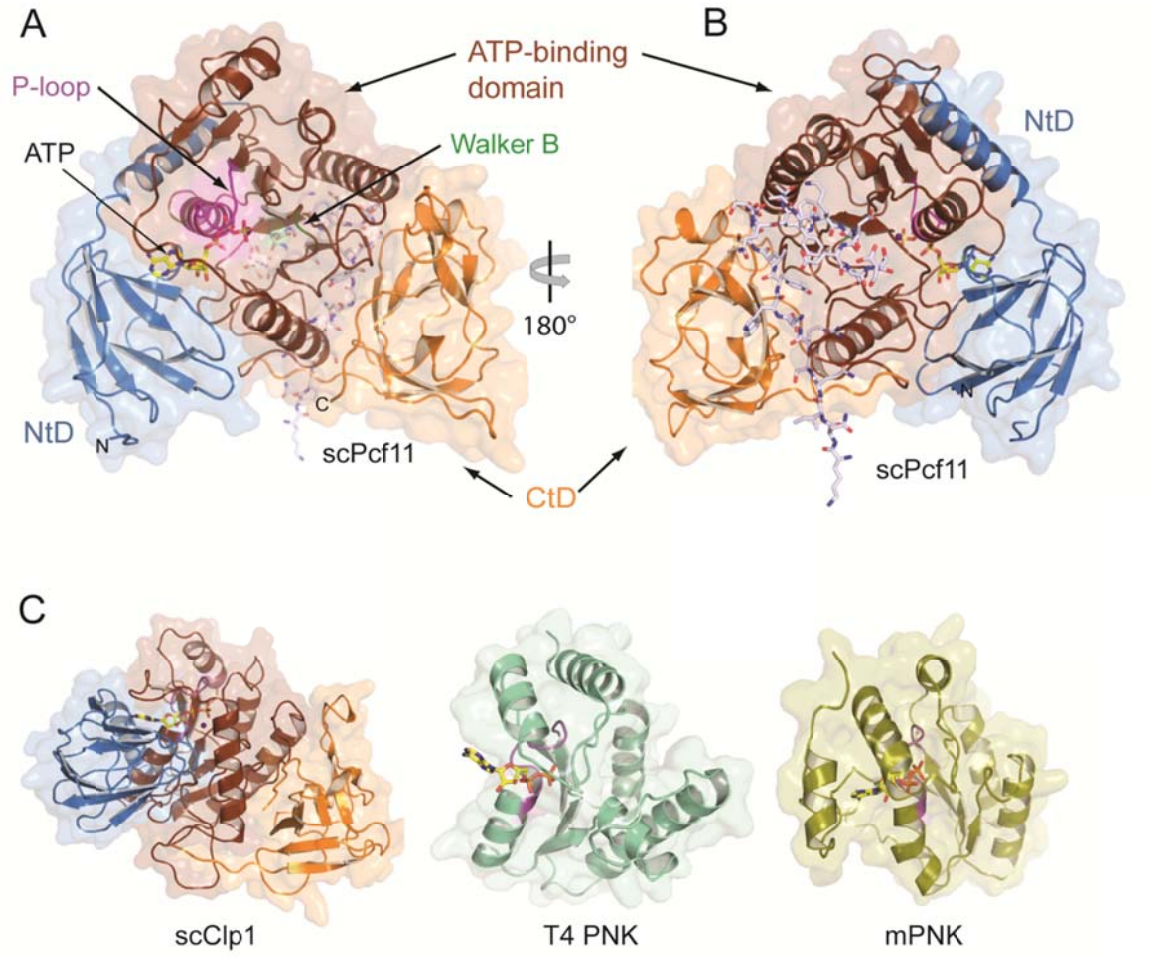


Figure 1.2.4. Structure of scClp1 in comparison with the structurally homologous proteins T4 PNK and mPNK. A, A ribbon representation of scClp1 showing the N-terminal domain (NtD), ATP-binding domain (AbD) and the C-terminal domain (CtD). Structural motifs important for ligand-binding are highlighted in purple (P-loop) and green (Walker B). scClp1 was crystallized in complex with scPcf11 (PDB 2NPI). ATP and scPcf11 are shown as a stick model. **B,** The orientation is rotated by 180 degree. **C,** The homologous structures of the kinase domains of T4 PNK (PDB 1RRC) and mPNK (PDB 3ZVN) were superimposed to scClp1 based on their P-loop motif. The bound ADP molecules of T4 PNK and mPNK are shown as a sick model (yellow).. The ATP-binding site of scClp1 is covered by the NtD, whereas the T4 PNK and the mPNK show an "open" ATP-binding site.

1.3 The phosphoryl transfer reaction mechanism

Since hsClp1 belongs to the P-loop kinase, the reaction mechanism of the phosphoryl transfer is evolutionary conserved and should resemble the one of previously described P-loop kinases[117]. Currently, our knowledge about this reaction mechanism mainly relies on experiments obtained from mononucleotide kinases (NMP-kinases)[118] and GTPases[119].

However, to fully understand enzyme catalysis of PNKs the knowledge about the nature of the transition state is mandatory. So far, structural details of the PNK-catalyzed phosphoryl transfer reaction mechanism are missing and need to be further investigated. The transition state theory distinguishes between two extreme cases that are characterized by the sequence of bond formation and bond breakage of the leaving group[120]. A dissociative transition state is described as a metaphosphate, the bond to the leaving group is fully broken and the bond to the nucleophile absent[120] (Figure 1.3.1). In contrast to the dissociative transition state, the associative transition state shows a penta-coordinated leaving group[120] (Figure 1.3.1). Depending on the nature of the transition state, charges are differently distributed and need to be specifically compensated (Figure 1.3.1). Theoretically, the transition state is inaccessible for structural characterization, due to its marginal lifetime. Therefore, a special experimental setup is necessary to obtain structural information. Instead of capturing the real transition state, compounds are used that mimic the transition state geometry and charge distribution[121]. In several crystal structures aluminum fluoride, beryllium fluoride, and vanadate were shown to function as transition state analogs for the transferred phosphoryl group[122,123]. Currently, no structural data of a transition state analog in PNKs are available. This information, however, would greatly facilitate our understanding of the reaction mechanism and amongst other questions is addressed in this work.

A characteristic feature shared by kinases is an evolutionary conserved P-loop motif with the consensus sequence GxxxxGK[S/T][19]. The phosphates of the nucleoside triphosphate are bound by a hydrogen-bonding network formed by the main-chain amides of the P-loop[113]. The Walker A lysine of the P-loop interacts electrostatically with the β- and γ-phosphates in a bifurcated manner[113]. The metal cofactor (Mg^{2+}) is octahedrally coordinated by the conserved Ser/Thr residue, the oxygen atoms of the β- and γ-phosphates, and water molecules[19]. The Walker B is classically composed of a catalytic aspartate that activates the acceptor oxygen by deprotonation, a prerequisite for the nucleophylic attack on the γ-phosphate group of the ATP molecule[52].

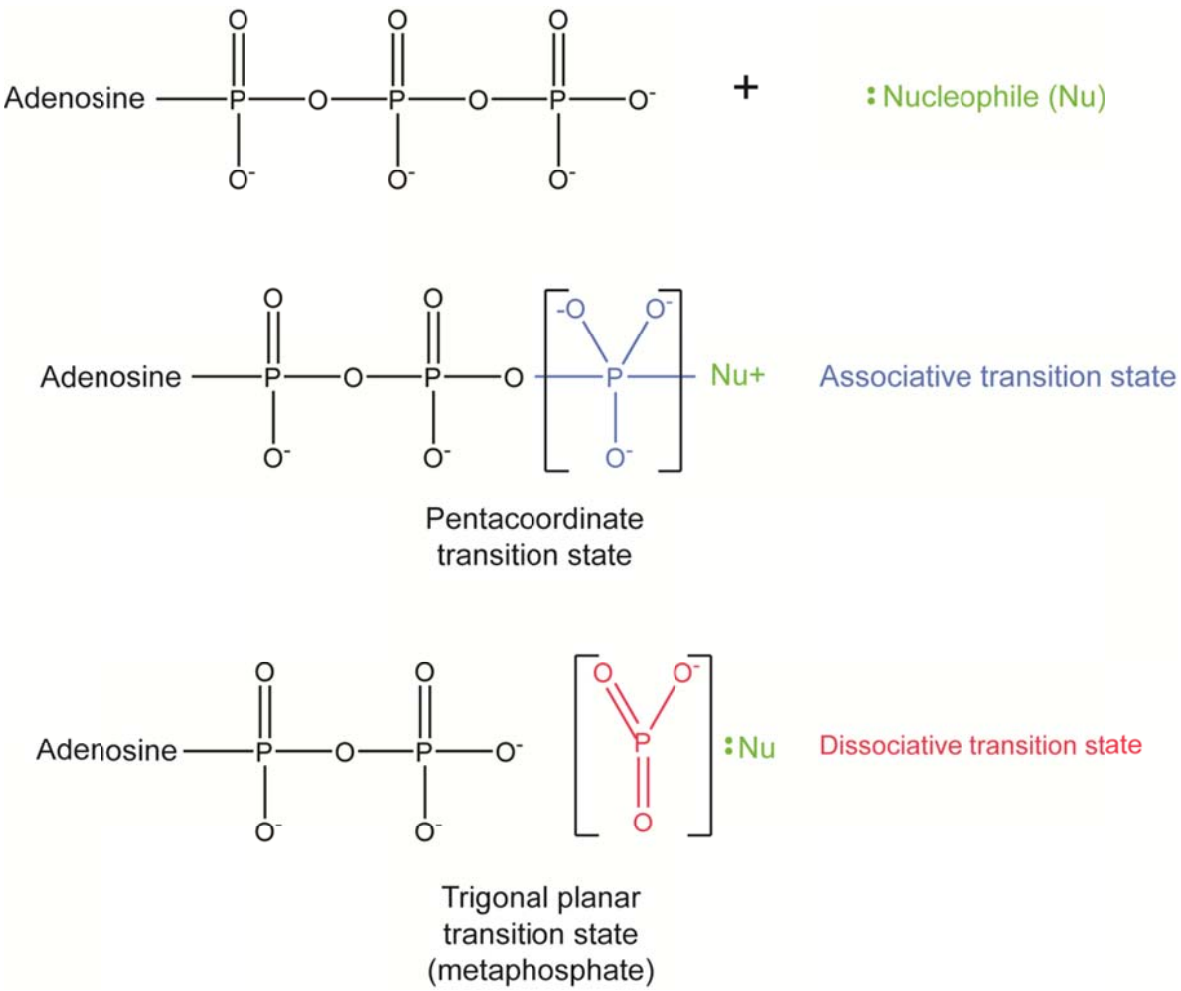


Figure 1.3.1. Nature of the transition state for phosphoryl transfer reactions. The transition state theory describes the phosphoryl transfer reaction either with an associative (SN2-like) or a dissociative (SN1-like) mechanism. The differences between both transition states are the sequence of bond formation and bond breakage between the attacking nucleophile and the bridging oxygen atom of the β- and γ-phosphate group. In case of an associative mechanism a pentagonal bipyrimidal structure is formed, whereas dissociative transition state involves a planar, trigonal, metaphosphate (PO_3^-) structure. Besides their different geometry, the two transitions state structures also differ in charge distribution. In contrast to an associative transitions state with a net charge of –3, the dissociative case has a net charge of –1. Adapted from[124].

1.4 Scope of this thesis

Even though the newly identified RPNKs are functionally well characterized[86,87,95,115,116], our structural knowledge about the protein family is limited. hsClp1 was shown to be involved in mRNA as well as tRNA processing and RNAi as a recurrent linker[88,92]. However, the molecular target of its 5'-kinase activity remains elusive. To date, only few crystal structures of PNKs with bound nucleotides[125,126] are available and no structure of an enzymatically active variant of Clp1 has been provided.

Therefore, the first aim of this PhD thesis is to gain structural insights into Clp1's substrate specificity by X-ray crystallography. The structural data presented in this thesis together with the comprehensive biochemical characterization provide an experimental and conceptual framework to understand Clp1's mode of RNA-recognition and RNA-specificity.

The second aim of this thesis addresses Clp1's mechanism of the phosphoryl transfer reaction. Our current knowledge of the phosphoryl transfer reaction and the transition state architecture is mainly based on observations made for NMP-kinases. The characterization of crystal structures in catalytically important states of Clp1 provides a model for the phosphoryl transfer reaction mechanism of PNKs. This model can be applied in general since their active sites are highly conserved. Furthermore, based on the crystal structure of an active Clp1 ortholog it was possible to identify enzymatically important residues that are missing in scClp1.

The third aim was to decipher the function of NtD and CtD in an enzymatically active Clp1. As indicated by the crystal structure of the inactive scClp1 ortholog, the central domain containing the P-loop motif is sandwiched between additional NtD and CtD with unknown functions[94]. In the case of scClp1, it was shown that these domains are involved in protein-protein interactions[98]. To elucidate the function of these additional domains in catalysis, truncated variants of Clp1 were designed and biochemically characterized. Based on these experiments, it was possible to show that the NtD is essential for ATP binding, whereas the CtD seems to be important in the correct positioning of enzymatically relevant residues.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

If not otherwise stated, all chemicals were purchased from Carl Roth (Karlsruhe, Germany), Fluka (Deisenhofen, Germany), GERBU Biotechnik (Gaiberg, Germany), Honeywell/Riedel de Haën (Seelze, Germany), Invitrogen (Paisley, UK), Merck (Darmstadt, Germany), NEB (Schwalbach, Germany), Serva (Heidelberg, Germany), and Sigma-Aldrich (Deisenhofen, Germany) in analytical grade purity.

2.1.2 Crystallization screens

Crystallization screen	Company
Classics Suite	Qiagen, Hilden, Germany
JCSG Core Suites	Qiagen, Hilden, Germany
PEGs and PEGs II Suites	Qiagen, Hilden, Germany
Ammoniumsulfate Suite	Qiagen, Hilden, Germany
Wizard I and II	Emerald, Bainbridge Island, WA, USA
Additive Screen	Hampton Research, Alison Viejo, CA, USA

2.1.3 Buffers

Buffer name	Buffer composition
Lysis buffer 1	20 mM HEPES-NaOH pH 7.0 100 mM NaCl 10 mM β-mercaptoethanol
Lysis buffer 2	20 mM HEPES-NaOH pH 7.0 100 mM NaCl 5 mM β-mercaptoethanol
Anion-exchange buffer	20 mM HEPES-NaOH pH 7.0 1 mM $MgCl_2$ 2 mM dithioerythritol (DTE)
Size exclusion buffer	10 mM HEPES-NaOH pH 7.0 100 mM NaCl 1 mM $MgCl_2$ 2 mM DTE
Crystallization buffer	10 mM HEPES-NaOH pH 7.0 100 mM NaCl 1 mM $MgCl_2$ 4 mM tris(2-carboxyethyl)phosphine (TCEP)
Dialysis buffer	20 mM HEPES-NaOH pH 7.3 100 mM NaCl 2 mM $MgCl_2$ 4 mM $CaCl_2$ 2 mM DTE
2x Phosphorylation buffer	100 mM HEPES-NaOH pH 8.0 200 mM NaCl 30 mM $MgCl_2$

2x ATPase buffer	20 mM HEPES-NaOH pH 7.0 200 mM NaCl 28 mM KCl 20 mM $MgCl_2$ 1 mM ethylenediaminetetraacetic acid (EDTA) 0.5 mM nicotinamide adenine dinucleotide (NADH) 0.4 mM phosphoenol pyruvat (PEP)
SDS-PAGE running buffer	25 mM Tris-HCl pH 8.0 200 mM glycine 0.1 % (*w/v*) sodium dodecyl sulfate (SDS)
SDS-PAGE separation gel buffer	1.5 M Tris-HCl pH 8.8 0.4 % (*w/v*) SDS
SDS-PAGE stacking gel buffer	1.5 M Tris-HCl pH 6.8 0.8 % (*w/v*) SDS
SDS-PAGE stacking gel (5 %)	125 mM Tris-HCl pH 6.8 15 % (*v/v*) Rotiphorese® Gel 30 (37.5:1) 0.1 % (*w/v*) ammonium persulfate (APS) 0.1 % (*v/v*) tetramethylethylenediamine (TEMED)
SDS-PAGE separation gel (15 %)	375 mM Tris-HCl pH 8.8 0.1 % (*w/v*) SDS 50 % (*v/v*) Rotiphorese® Gel 30 (37.5:1) 0.08 % (*w/v*) APS 0.08 % (*v/v*) TEMED
10x TBE buffer	890 mM Tris-HCl pH 8.0 890 mM boric acid 20 mM EDTA

RNA-urea gel (9 %)	38 % *(v/v)* Rotiphorese® sequencing gel concentrate (19:1) 52 % *(v/v)* Rotiphorese® sequencing gel diluent 1x TBE buffer 0.08 % (*w/v*) APS 0.08 % (*v/v*) TEMED
6x DNA loading dye	30 % (*v/v*) glycerol 0.25 % (*w/v*) bromphenol blue 0.25 % (*w/v*) xylene cyanole FF
5x SDS loading dye	50 mM Tris-HCl pH 7.0 10 % (*v/v*) glycerol 10 % (*w/v*) SDS 0.1 % (*w/v*) bromophenol blue 0.9 M DTE 0.1 M β-mercaptoethanol
RNA-urea loading dye	95 % (*v/v*) formamide 5 mM EDTA 0.025 % (*w/v*) SDS 0.25 % (*w/v*) bromophenol blue 0.25 % (*w/v*) xylene cyanol FF
Methylene blue staining solution	0.4 M sodium acetate 2.28 % (*v/v*) acetic acid 0.01 % (*w/v*) methylene blue
TFB-1	30 mM potassium acetate pH 5.8 50 mM $MnCl_2$ 100 mM RbCl 10 mM $CaCl_2$ 15% (v/v) glycerol
TFB-2	10 mM MOPS-NaOH pH 7.0 10 mM RbCl 75 mM $CaCl_2$ 15% (v/v) glycerol

2.1.4 Growth media

All bacterial cultures were grown in lysogeny broth (LB) medium. LB-agar plates for single colony growth were solidified by addition of bacto-agar. Both LB medium and LB-agar plates were routinely provided by the media kitchen of the Max Planck Institute for Medical Research (Heidelberg, Germany). Depending on the used *E. coli* strain and the plasmid constructs, different antibiotics were used as selection markers. For growth of *E. coli* strain BL21-CodonPlus(DE3)-RIL, the media was supplemented with 34 µg/ml of chloramphenicol. In case of expression experiments, the media additionally contained either 50 µg/ml of kanamycin when cells were transformed with pET28b vector or 100 µg/ml of ampicillin after transformation with pET21a vector.

Media name	Components
LB	10 g/l bacto tryptone 5 g/l bacto-yeast extract 10 g/l NaCl Adjusted to pH 7.0 with NaOH
LB-agar plates	25 g/l Standard-I-nutrient-broth 15 g/l bacto agar

For selenomethionine labeling of protein, cell cultures of *E. coli* strain BL21-CodonPlus(DE3)-RIL were grown in minimal media supplemented with selenomethionine instead of methionine. If not elsewhere stated, cells were treated similar to native expression experiments. The media were prepared according to[127].

2.1.5 Bacterial strains

Cells of *E. coli* strain DH5α were used for plasmid DNA propagation, whereas cells of strain BL21-CodonPlus(DE3)-RIL were used for protein overexpression.

***E. coli* strains**	**Genotype**	**Reference**
DH5α	F- φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(rk-, mk +) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ-	[128]
	E. coli B F– *ompT hsdS*(r_B– m_B–) *dcm*+ Tet^r *gal* λ(DE3) *endA* The [*argU ileY leuW* Cam^r]	Novagen, Wisconsin, USA

2.1.6 Plasmids

Protein expression of the different construct variants was performed with two plasmids of the pET expression vector series (Novagene, Madison, USA). The full-length gene and truncation variants thereof were cloned into the multiple cloning sites with the indicated restriction sites.

The pET28b vector carries both an N-terminal hexahistidine-tag fused to a thrombin cleavage site and an optional C-terminal hexahistidine-tag sequence. The expression construct was designed with an N-terminal hexahistidine-tag fusion. To prevent an additional C-terminal hexahistidine-tag, a stop-codon was introduced. In contrast to pET28b, the pET21a vector carries only a C-terminal hexahistidine-tag. Both plasmids are isopropyl-β-D-thiogalactopyranosid (IPTG) inducible expression systems.

Name	Description	Reference
pET28b	T7-based expression system coding for a recombinant protein of N-terminal hexahistidine-tag, a thrombin protease cleavage site fused to the protein of interest; Kan^R	Novagene
pET21a	T7-based expression system coding for a recombinant protein with C-terminal hexahistidine-tag; Amp^R	Novagene
pET28b-ceClp1	Cloned via *NcoI*/*NotI*	This thesis
pET28b-ceClp1-K127A	Cloned via *NcoI*/*NotI*	This thesis
pET28b-ceClp1-K127R	Cloned via *NcoI*/*NotI*	This thesis
pET28b-ceClp1-W233A	Cloned via *NcoI*/*NotI*	This thesis
pET28b-ceClp1ΔC104	Cloned via *NcoI*/*NotI*	This thesis
pET21a-ceClp1ΔN107	Cloned via *NdeI*/*NotI*	This thesis
pET28b-ceClp1ΔC310	Cloned via *NcoI*/*NotI*	This thesis
pET21a-ceClp1ΔNC315	Cloned via *NdeI*/*NotI*	This thesis

2.2 Methods

2.2.1 Molecular Biology

2.2.1.1 Competent E. coli cells and transformation

Chemically competent E. *coli* DH5α and *E. coli* BL21-CodonPlus(DE3)-RIL cells were used for cloning and expression experiments, respectively. Chemically competent cells were provided by Maike Gebhard (Max Planck Institute for Medical Research) and produced using a modified protocol of Hanahan[129]. Cells from a glycerol stock were transferred onto an LB-agar plate and incubated overnight. A single colony was used to inoculate an overnight pre-culture with antibiotics if required. This pre-culture was diluted 1:100 into 0.5 l of LB medium. The cell suspension was incubated at 37 °C until an OD_{600} of 0.5 was reached. Afterwards, cells were harvested by centrifugation (6000 x g, 4 °C, 10 min). The supernatant

was discarded and the cell pellet was washed with 50 ml of TFB-1. A final centrifugation step was performed (6000 x g, 4 °C, 10 min) and the cell pellet was resuspended in 4 ml of TFB-2. Aliquots of 50 µl were immediately flash-frozen in liquid nitrogen and stored at -80 °C until further use.

For transformation of *E. coli* cells with plasmid-DNA, aliquots of chemically competent cells were thawed on ice. After thawing, 1-2 µl of plasmid DNA (20-200 ng/µl) was added to the aliquot and incubated on ice for additional 15-30 minutes. Cells were heat shocked at 42 °C for 45 sec and immediately transferred onto ice for 2-3 minutes. Subsequently, 450 µl LB medium was added to the aliquot and incubated at 37 °C for 60 minutes under constant shaking at 750 rpm in a thermomixer comfort (Eppendorf, Hamburg, Germany). Prior to plating, the transformed cells were harvested by centrifugation at 4500 rpm in a tabletop centrifuge and resuspended in approximately 50 µl of LB medium. This cell suspension was plated on a pre-warmed LB-agar plate supplied with appropriate antibiotics. Plates were incubated at 37 °C overnight for single colony growth.

2.2.1.2 Polymerase chain reaction (PCR)

PCR was used to amplify genetically modified DNA-fragments, taking advantage of build in restriction sites within the PCR-primer sequences. A standard PCR protocol was used in a TPersonal thermocycler (Biometra, Göttingen, Germany). The protocol was adapted from Saiki and colleagues using a thermostable DNA-dependent DNA polymerase[130]. Primers were synthesized by Eurofins MWG Operon (Eurofins, Ebersberg, Germany) and the components for the PCR reaction were purchased at New England Biolabs (NEB, Schwalbach, Germany). Annealing temperature and elongation time were adapted according to primer characteristics and to the particular prduct size, respectively. PCR products were separated by agarose gel electrophoresis. These separated DNA fragments were excised and purified by the use of the gel extraction kit according to the manufacturer's recommendations (Qiagen, Hilden, Germany).

Component	Concentration		Volume [µl]
PCR template	50	ng/µl	0.5
Primer	100	pmol/µl	0.5
Formamide	> 99.5	%	3.5
$MgSO_4$	100	mM	2-4
dNTPs[a]	10	mM	5
10 x Thermopol buffer[b]			5
Vent polymerase	2	U/µl	1 µl
ddH2O			ad 50 µl

[a]Mixture of dNTPs containing dCTP, dATP, dGTP and dTTP

[b]200 mM Tris-HCl pH 8.8, 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 20 mM $MgSO_4$, 1 % Triton X-100

Standard PCR thermocyler protocol		
Program	Temperature [°C]	Time [min]
Initial denaturation	95	5
Denaturation	95	0.5
Annealing	55	1
Elongation	68	1 min per 1000 bp
Final elongation	68	10
Storage	4	Until further use

Standard PCR experiments were performed with 30 cycles of denaturation, annealing, and elongation.

2.2.1.3 Site-directed mutagenesis

The mutagenesis procedure was carried out according to the QuikChange protocol (Strategene, La Jolla, USA). For primer design, an online browser software tool of Agilent Technologies (Agilent, Santa Clara, USA) was used and resulting primers were ordered from Eurofins MWG Operon (Eurofins Ebersberg, Germany). The QuikChange PCR reaction (50 µl) contained 1x Pfu reaction buffer (Stratagene La Jolla, USA), 2.5 µM of each sense and antisense primer, 0.2 mM dNTP mix, 50 ng template DNA and 2.5 U Pfu Turbo polymerase (Stratagene La Jolla, USA).

In all QuikChange reactions, the following PCR program was used on a TPersonal thermocycler (Biometra, Göttingen, Germany):

Site-directed mutagenesis PCR reaction		
Reaction	Temperature [°C]	Time [min]
Initial Denaturation	95	5
Denaturation	95	0.5
Annealing	55	1
Elongation	68	1 min per 1000 bp
Final Elongation	68	10
Storage	4	Until further use

Denaturation, annealing and elongation were repeated in 16 cycles.

Annealing temperature and elongation time were adapted according to primer melting temperature and the particular plasmid size, respectively. Finally, the reaction sample was incubated with DpnI (10 U) at 37°C for 1-2 hours to remove methylated template DNA. Subsequently, the nicked PCR product was transformed into chemically competent *E. coli* DH5α cells. Success of site directed mutagenesis was verified by sequencing of the respective plasmid DNA from single colonies.

2.2.1.4 Amplification of plasmid DNA

A single colony of *E. coli* strain DH5α was picked after transformation and used to inoculate a 4 ml LB overnight culture at 37 °C. Appropriate antibiotics were supplemented according to the selection markers of the used plasmids. Plasmid DNA was isolated according to the manufacturer's instruction using a QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany). Concentration and purity of DNA was determined spectroscopically measuring absorbance at 260 nm and purity was verified by measuring the 260/280 nm ratio using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, USA).

2.2.1.5 Restriction digest and ligation of PCR products

DNA from PCR products or from plasmid DNA was incubated with restriction enzymes (NEB, Schwalbach, Germany) for at least 3 hours at 37 °C. To prevent religation, plasmid DNA was additionally incubated with 10 U of calf intestine phosphatase (NEB, Schwalbach, Germany) at 37 °C for 1 hour.

Standard digestion protocol Component	Concentration	Volume [µl]
DNA	~50 ng/µl	xx[b]
Restriction enzyme 1	10-20 U/µl	1
Restriction enzyme 2	10-20 U/µl	1
BSA[b]	10 mg/µl	0.5
10 x NEB buffer		5
ddH_2O		ad 50 µl

[a] If necessary, bovine serum albumine (BSA) was added

[b] The volume was adjusted to the concentration of the stock solution

After restriction digest, samples were either purified with a PCR purification kit (Qiagen, Hilden, Germany) or by agarose gel electrophoresis (see. 2.2.1.8) Separated DNA-fragments were then purified with a gel extraction kit (see 2.2.1.6; Qiagen, Hilden, Germany).

2.2.1.6 DNA extraction from agarose gels

If necessary, DNA bands of interest were excised from the gel with a sterile razor blade and the DNA was subsequently extracted using a QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

2.2.1.7 Ligation reaction

Ligation reactions were performed according to the manufacturer's instruction using a Rapid DNA ligation kit (Fermentas, St. Leon-Rot, Germany).

Standard ligation protocol		
Component	Concentration	Amount
Vector		50-100 ng
Insert		100-300 ng
5 x Ligation buffer		4 µl
T4 DNA ligase	5 U/µl	1 µl
ddH_2O		Ad 20 µl

2.2.1.8 Agarose gel electrophoresis

Agarose gel electrophoresis was used for analysis of DNA molecules that were obtained after restriction digest experiments or PCR reactions. Desired DNA molecules were purified in a gel containing 1-2 % agarose (*w/v*). For this, agarose was heated in 1x TBE buffer and cast before polymerization. To visualize the DNA under UV light irradiation, ethidium bromide

was added prior to solidification to a final concentration of 0.5 µg/ml. Electrophoresis was performed in 1x TBE buffer at 70-100 V for 1-2 hours. Samples were supplemented with DNA loading dye. As size markers, 1-10 µl of a NEB ruler 1 kb DNA Ladder were used.

2.2.1.9 Sequencing

To verify successful cloning of the different plasmid constructs, each plasmid DNA was sequenced at Eurofins/MWG (Ebersberg, Germany). Sequencing samples contained 20-200 ng/µl template DNA mixed with sequencing primer in a concentration of 10 µM. The sequencing primers used for the respective constructs are listed in the appendix.

2.2.1.10 Cloning of full-length Clp1

For generation of the expression construct pET28b-ceClp1, the coding sequence of ceClp1 was amplified by PCR from Caenorhabditis elegans cDNA (see. 2.2.1.2) According to the descriptions for a standard PCR reaction, the coding sequence was amplified using the primers ceClp1-pET28b-For and ceClp1-pET28b-Rev. The sequencing primers used for the respective constructs are listed in the appendix (see 6.1.1). The sequence of the forward primer contained an NcoI restriction site, whereas the reverse primer was designed with a NotI restriction site (see 2.2.1.5). The PCR product was purified with a QIAquick PCR Purification Kit and subsequently digested with the two restriction enzymes *Nco*I and *Not*I. The pET28b vector was digested with identical restriction enzymes. After restriction digest, both the linearized plasmid DNA and the PCR product were analyzed by agarose gel electrophoresis (see 2.2.1.8). The separated DNA-fragments were extracted from the excised gel slices using a QIAquick Gel Extraction Kit (see 2.2.1.6; Qiagen, Hilden, Germany). The PCR product was cloned into the linearized pET28b vector using a Rapid DNA Ligation Kit (Fermentas, St. Leon-Rot, Germany). The ligation leads to an open reading frame (ORF) coding for an N-terminal hexa-histidine fusion. For plasmid propagation, plasmid constructs were transformed into *E. coli* DH5α cells (see 2.2.1.1 and 2.2.1.4).

2.2.1.11 Cloning of C-terminal truncations

The expression construct pET28b-ceClp1ΔC104 was designed to express a C-terminally truncated variant of ceClp1. A stop codon leading to a truncated ORF was introduced by site-directed mutagenesis into the pET28b-ceClp1 construct using the QuikChange protocol (see 2.2.1.3; Strategene, La Jolla, California, USA). Thereby, the encoding codon triplet of residue 105 was mutated to a stop codon using the primer pair ceClp1ΔN104-pET28b-For and ceClp1ΔN104-pET28b_Rev. The sequencing primers used for the respective constructs are listed in the appendix (see 6.1.1).

The expression construct pET28-ceClp1ΔC310 was designed as a C-terminal truncation of the full-length gene obtained from pET28b-ceClp1. A stop codon was introduced into the coding sequence of ceClp1 by site directed mutagenesis using the QuikChange protocol (Strategene, La Jolla, California, USA). QuikChange primers were designed to replace the codon of residue 310 with a stop codon (ceClp1ΔC310-pET28b-For and ceClp1ΔC310-pET28b-Rev). The sequencing primers used for the respective constructs are listed in the appendix (see 6.1.1).

2.2.1.12 Cloning of N-terminal truncations

The ORF of an N-terminally truncated ceClp1 variant was amplified from the vector pET28b_ceClp1 using PCR primers designed with an *Nde*I and a *Not*I restriction site. Similar to the description for the cloning of the pET28b-ceClp1-Clp1 vector, the truncated coding sequence spanning from residue 107 to residue 428 was cloned into a pre-linearized pET21a vector (pET21a-ceClp1ΔN107). This ligation leads to an ORF encoding a C-terminally hexa-histidine fused protein tagged and was kingly provided from Bernhard Dichtl (Zürich, Switzerland).

The expression construct pET21-ceClp1NC314 was designed as an N-terminal truncation. Primers for amplification contained an *Nde*I (ceClp1ΔN314-pET28b-For) and a *Not*I (ceClp1ΔN314-pET28b-Rev) restriction site. This PCR product was digested and cloned into a pre-linearized pET21a vector. The ORF encodes for a recombinant protein of ceClp1ΔN314 fused to C-terminal hexahistidine tag. The sequencing primers used for the respective constructs are listed in the appendix (see 6.1.1).

2.2.1.13 Cloning of single amino acid mutations by site directed mutagenesis

The expression constructs pET28b_Clp1_K127A, pET28b_Clp1_K127R, and pET28b_Clp1_W233A were obtained by site directed mutagenesis based on full-length ceClp1. The codon-triplet of residue 127 was replaced with a codon coding either for alanine or arginine using the QuikChange protocol (Strategene, La Jolla, California, USA). Furthermore, the codon-triplet of residue 233 was also replaced with a codon coding for alanine. Primers for site directed mutagenesis were applied on the full-length expression construct pET28b_ceClp1 (pET28b_Clp1_K127A_for and pET28b_Clp1_K127Arev; pET28b_Clp1_K127R_for and pET28b_Clp1_K127R_rev; pET28b_Clp1_W233A_for and pET28b_Clp1_W233A_rev). The sequencing primers used for the respective constructs are listed in the appendix (see 6.1.1).

2.2.2 Biochemistry

2.2.2.1 SDS-polyacrylamide gel electrophoresis

Proteins were analyzed according to their molecular weight by discontinuous sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), initially described by Lämmli[131]. For gel casting and electrophoresis, a Protean 4 system (Bio-Rad, Munich, Germany) and a PowerPack 200 (BioRad, Munich, Germany) power source were used, respectively. Mini-gels were casted in parallel, starting with the separation gel. For the separation gel, 7.3 ml ddH_2O was mixed with 7.5 ml separation gel buffer and 15 ml aqueous 30 % acrylamide-, bisacrylamide (37.5:1, Roth, Karlsruhe, Germany) to obtain a 15 % gel. The polymerization reaction was started by addition of 250 µL of 10 % (*w/v*) APS (Merck, Darmstadt, Germany) and 25 µL TEMED (Serva, Heidelberg, Germany). To remove air bubbles, the gel was overlaid with isopropanol. To prepare the stacking gel, which was added on top of the separation gel, 8.9 ml ddH_2O were mixed with 3.8 ml stacking gel buffer and 2.3 ml 30 % acrylamide-, bisacrylamide (37.5:1). Polymerization was initiated by addition of 150 µl 10 % (*w/v*) APS and 15 µl TEMED. Gels were stored for up to two weeks at 4 °C until further use. Prior to loading, the protein samples were supplemented with 5x SDS PAGE sample buffer and boiled at 95 °C for 5-10 minutes. Electrophoresis was performed in SDS PAGE running buffer at 35 mA per gel. Gels were stained using Page Blue protein staining solution (Fermentas St. Leon-Rot, Germany) and destained with ddH_2O.

2.2.2.2 Expression of recombinant proteins

Large scale protein expression was performed in *E. coli* BL21-CodonPlus(DE3)-RIL cells harboring the respective expression plasmid constructs. LB medium (2 l) was supplemented with appropriate antibiotics and inoculated with a 1:100 dilution of an overnight culture. Cells were grown until they reached an OD_{600} of 1.0. Once cells were in their exponential growth phase, cultures were cooled down to 18 °C and protein expression was induced by addition of 1 mM IPTG. Cells were then incubated at 18 °C overnight. After cell harvesting, the pellets were resuspended in the appropriate buffers (see 2.1.3) and used either directly for protein purification or shock frozen with liquid nitrogen and stored at -80 °C. The bacterial cell wall was broken by pulsed-sonication for 20 minutes on ice and supernatants were cleared by centrifugation.

2.2.2.3 Purification of full-length ceClp1

For purification of wild-type ceClp1 and variants obtained by site directed mutagenesis, pellets were resuspended in lysis buffer 1. The full-length protein was expressed using pET28b vector (pET28b_ceClp1, pET28b_ceClp1_K127A pET28b_ceClp1_K127R, and pET28b_ceClp1W233A) and initially purified by metal ion chromatography using Ni^{2+}-NTA material (Qiagen, Hilden, Germany). The column (~ 1 ml column volume (CV)) was equilibrated with lysis buffer 1 prior to applying the supernatant and subsequently washed with 3 CV of lysis buffer 1. Bound proteins were eluted with 6 ml of lysis buffer 1 containing additional 300 mM imidazole. To remove the N-terminally fused hexahistidine-tag, the protein was cleaved with thrombin (Sigma-Aldrich, Deisenhofen, Germany). For this, eluted fractions were dialysed against dialysis buffer overnight and incubated with 80 U of thrombin. Cleaved ceClp1 protein was diluted with anion-exchange buffer until a final conductivity of 5 mS/cm was reached and loaded onto a MonoQ column (GE Healthcare, Freiburg, Germany), equilibrated with anion-exchange buffer. Bound proteins were eluted over a linear gradient of 30 CV to the anion-exchange buffer containing additional 600 mM NaCl. Fractions containing pure ceClp1 were pooled and concentrated using Amicon-15 centrifugal filter units (Millipore) with a molecular weight cut-off of 10 kDa. Finally, the protein solution was applied onto a Superdex 75 (GE Healthcare, Freiburg, Germany) column equilibrated in size exclusion buffer and eluted protein fractions were concentrated as described above. All variants generated by site directed mutagenesis were purified accordingly. Protein batches

used for crystallization experiments were purified similarly, except that DTE was used in a concentration of 4 mM instead of 2 mM, and that it was replaced with 4 mM TCEP in the final purification step.

2.2.2.4 Purification of truncation variants of ceClp1

For purification of the N- and C-terminal truncation variants, pellets were resuspended in lysis buffer 2. The truncation variants were expressed using pET28b-ceClp1ΔC104, pET21a-ceClp1ΔN107, pET28b-ceClp1ΔC310, pET21a-ceClp1ΔN314, and initially purified by metal ion chromatography using TALON (Clontech).

The untagged N-terminal truncated pET21a-ceClp1ΔN107 variant was co-expressed with pET28b-ceClp1ΔC104 to improve solubility. Furthermore, pET28b-ceClp1ΔC310 also was co-expressed with pET21a-ceClp1ΔN314. The supernatant containing the co-expressed truncation variants ceClp1ΔN107 and ceClp1ΔC104 as well as ceClp1ΔC310 and ceClp1ΔN314 was loaded on to a TALON Clontech (Saint-Germain-en-Laye, France) metal affinity column (1 ml cv) equilibrated with buffer 2.

After a washing step with buffer 2 the untagged ceClp1ΔN107 was eluted from the immobilized ceClp1ΔC104 in a high salt wash of buffer 2 containing additional 1 M NaCl. The pooled wash fractions were concentrated using Amicon-15 centrifugal filter units with a MWCO of 10 kDa and in a re-chromatography step using TALON Clontech (Saint-Germain-en-Laye, France) resin, minor traces of ceClp1ΔC104 were removed. A buffer exchange to buffer 2 was performed by concentration of the flow-through and re-dilution with buffer 2 in a concentration step as described above. The N-terminal domain of ceClp1, the ceClp1ΔC104 variant was expressed alone and purified over a TALON Clontech (Saint-Germain-en-Laye, France) metal affinity column (1 ml cv) equilibrated with buffer 2. After washing with 3 cv of buffer 1, the bound protein was eluted with 6 cv of buffer 2 containing additional 150 mM imidazole.

The supernatant of the co-expressed truncation variants ceClp1ΔC310 and ceClp1ΔN314 was loaded on to a TALON Clontech (Saint-Germain-en-Laye, France) metal affinity column (1 ml cv) equilibrated with buffer 2. After washing with 3 cv of buffer 2, the bound proteins were eluted with 6 cv of buffer 2 containing additional 150 mM imidazole.

2.2.2.5 Phosphorylation assay

5'-kinase activity of ceClp1 towards different RNA and DNA substrates was analyzed using a phosphorylation assay. ceClp1 was tested either with a dinucleotide RNA substrate guanylyl(3′→5′)cytidine (G_1C_2) or with RNA and DNA oligonucleotides (20 nt, see Appendix 6.1.2). Each reaction was performed under limiting enzyme concentrations (1 µM) in 1x phosphorylation buffer. In case of the G_1C_2, ceClp1 was incubated with 250 µM ATP and 250 µM G_1C_2 for 2 hours and subsequently the reaction was analyzed by chromatographic separation of the products ADP and 5'P- G_1C_2. In addition, G_1C_2 (500 µM) was also used in a two fold excess relative to ATP to prove that ceClp1 catalysis the phosphorylation of G_1C_2 in a one-to-one molar ratio. Phosphorylation was determined by chromatographic purification (see 2.2.2.6).

For the phosphorylation assay with oligonucleotides, ceClp1 was incubated with 1 mM ATP and three different RNA and DNA substrates. Oligonucleotides were provided either as single-stranded, double-stranded (blunt-end), or double-stranded (3'-overhang) substrates. To produce a double-stranded substrate, two complementary strands were mixed in equimolar amounts (0.1 mM). For annealing, nucleotide solution was heated to 95°C for 5 minutes and gradually cooled to room temperature. The resulting double-stranded substrate was stored at -20°C until further use. Phosphorylation reactions were incubated at 30 ºC and quenched upon addition of 8 M urea. Control experiments were performed with T4 PNK (20 U) in an identical experimental setup. Phosphorylation efficiency was determined by denaturing PAGE (see 2.2.2.9).

2.2.2.6 Chromatographic purification of the phosphorylation products

This approach is designed for baseline separation of reaction products based on changes of the ionic net charge after phosphoryl transfer reaction. After the incubation time of 2 hours, each reaction (200 µl) was diluted 1:2 with ddH_2O, filtered, and finally applied to a MonoQ column (5/50, GE Healthcare, Freiburg, Germany) equilibrated with ddH_2O at 4 °C. The baseline separation was achieved using a linear gradient to 80 mM NaCl. Under these conditions, G_1C_2 eluted at 5.8 ml, 5'P-G_1C_2 at 8.7 ml, ADP at 9.2 ml, and ATP at 10.8 ml. During chromatographic separation, absorbance at 260 nm and 280 nm and conductivity (mS/cm) were monitored.

2.2.2.7 Quantitative production of 5'P -G_1C_2

5'P-G_1C_2, which is used for a single turnover reverse reaction, was produced enzymatically using ceClp1 (see 2.2.2.8). The reaction sample containing 2 µM of recombinant ceClp1 supplemented with 500 µM of G_1C_2 and 500 µM of ATP in phosphorylation buffer was incubated at 25 °C for 8 hours. After incubation the reaction mixture was diluted 1:2 with ddH_2O and loaded onto a MonoQ anion exchange column (10/100, GE Healthcare, Freiburg, Germany) equilibrated with ddH_2O. G_1C_2 was baseline separated from the phosphorylated 5'P-G_1C_2 molecule using a linear gradient of 80 mM NaCl. To prevent contamination from residual traces of ADP, an ATP recovery system was used containing pyruvat kinase (PK) and 500 µM PEP. Eluted fractions containing 5'P-G_1C_2 were collected and concentrated using a vacuum concentrator. 5'P-G_1C_2 was resuspended in ddH_2O and residual salt was removed by gel filtration chromatography. The sample (500 µl) was applied onto a Superdex 75 column (5/50, GE Healthcare, Freiburg, Germany) and equilibrated with ddH_2O. Desalted fractions of 5'P-G_1C_2 were concentrated as described above by a vacuum concentrator and stored at -20 °C.

2.2.2.8 Reverse reaction

The reverse reaction was analyzed chromatographically under single turnover conditions (see 2.2.2.6). The reaction sample contained 60 µM of recombinant ceClp1 supplemented with 50 µM of 5'P-G_1C_2 and 50 µM of ADP in the phosphorylation buffer incubated for 20 and 40 minutes. Identical to the analysis of the forward reaction, the sample was applied to a MonoQ column (GE Healthcare, Freiburg, Germany). To inhibit enzyme activity of possible adenylate kinase contaminations, all samples of single turnover reactions were treated with 100 µM of diadenosine pentaphosphate *(*Ap5A*)*. Ap5A is known as a potent inhibitor of adenylate kinases functioning as a bisubstrate inhibitor[132]. Based on results obtained from the forward reaction, it was shown that ceClp1 5'-kinase activity is not inhibited by Ap5A. Additional control experiments were performed to show that ceClp1 protein batches do not posses any basal adenylate kinase activity on their own. ceClp1 was incubated with 50 µM of ADP. The forward reaction was also performed under single turnover conditions. The reaction consisted of 60 µM of ceClp1 supplement with 100 µM of Ap5A, 50 µM of G_1C_2 and 50 µM of ATP incubated for 20 minutes.

2.2.2.9 Denaturing polyacrylamide gel electrophoresis (Urea-PAGE)

Phosphorylation efficiency of ceClp1 towards single-stranded and double-stranded RNA and DNA oligonucleotides (20 nt) was analyzed by denaturing Urea-PAGE. For this, 9 % (*v/v*) polyacrylamide gels (19:1 acrylamide to bisacrylamide ratio) containing 8.3 M urea were used. Prior to electrophoresis, RNA samples were denatured with urea (4 M) and RNA and supplemented with sample buffer. Gels were cast in glass plates (40 cm x 33 cm) with 1 mm spacers and were run in 0.5x TBE buffer. Polymerization of the polyacrylamide gel solution was initiated by addition of 0.08 % (*w/v*) APS and 0.08 % (*v/v*) TEMED. To remove excess of APS and TEMED after polymerization, a pre-run step was applied at 25 W for 1 hour. After loading, the gel was run at 25 W until G_1C_2 and 5'P-G_1C_2 were separated, based on their different electrophoretic mobility. RNA and DNA oligonucleotides were visualized by methylene blue staining.

2.2.2.10 Steady-state kinase assay

A coupled colorimetric assay was used to determine consumption of ATP during RNA phosphorylation by ceClp1. This assay enables quantification of ceClp1's kinase activity. ATP consumption of Clp1 is coupled to the oxidation of NADH using a recovery system. The recovery system consists of two components. The first is PK that converts PEP and ADP to ATP and pyruvate, thus responsible for ATP-regeneration. The second is lactate dehydrogenase (LDH) that uses pyruvate to produce lactate by using NADH as reducing agent. This decrease in the amount of NADH was followed by measuring absorbance change at 340 nm and allowed spectroscopical quantification of Clp1's ATP hydrolysis rates. All experiments were performed at constant ATP concentrations of 1 mM. To achieve this, 0.2 µM of ceClp1 was incubated in ATPase buffer supplemented with 7 U/ml pyruvate kinase (Sigma-Adrich, Deisenhofen, Germany) and 10 U/ml lactate dehydrogenase (Sigma-Adrich, Deisenhofen, Germany). Prior to the kinetic measurements, the reaction mixture was incubated at room temperature for 20 minutes to obtain a stable baseline and subsequently the reaction was started by addition of ceClp1. To determine the Michaelis-Menten kinetics for the RNA dinucleotide G_1C_2 and urylyl-(3'-5')guanosine (U_1G_2), varying RNA concentration were used.

Changes in absorbance at 340 nm were monitored in a 1 cm quarz cuvette (Hellma, Mühlheim, Germany) using a JASCO V-650 spectrophotometer. The initial rates of the reaction were calculated using equation 1.1.

$$v = \frac{\Delta Abs_{340} / \Delta t}{\varepsilon N_{NADH} \cdot d} \qquad \text{[equation 1.1]}$$

ΔAbs_{340}: Difference of absorption at 340 nm

Δt: Time-difference (min or s)

ε_{NADH}: Absorption coefficient of NADH (6220 M^{-1} cm^{-1})

c_p: Protein concentration (M)

d: Path length of cuvette (cm)

Initial rates were plotted against the substrate concentration and fitted using the standard Michaelis-Menten using equation 1.2:

$$v = \frac{v_{max} \cdot S}{S + K_m} \qquad \text{[equation 1.2]}$$

S: Substrate concentration

v_{max}: Maximal apparent reaction velocity

K_m: Apparent Michaelis-Menten constant

Turnover numbers were calculated according to the following equation using equation 1.13:

$$K_{cat} = \frac{v}{c_p} \qquad \text{[equation 1.3]}$$

c_p: Protein concentration (M)

2.2.3 Biophysical Methods

2.2.3.1 CD spectroscopy of Clp1 mutant variants

Protein stability of wild type ceClp1 and mutant variants was compared using CD spectroscopy. Melting curves were determined using change in ellipticity at 222 nm from 10°C to 60°C with a heat-rate of 0.5 °C/min. Protein samples (15 µM) in size exclusion buffer were analyzed using a Jasco J-810 spectropolarimeter (Jasco, Groß-Umstadt, Germany) with 1 mm CD cuvettes (Hellma, Müllheim, Germany).

2.2.4 Crystallographic Methods

To identify suitable crystallization conditions, protein solution was pre-mixed with nucleotide ligands and applied to a variety of crystallization screens in a 96-well plate sitting drop vapor diffusion setup with the Mosquito nanolitre pipetter (TTP LabTech LTD, Melbourn, UK). Manual optimization in Linbro plates using the hanging drop vapor diffusion method was unsuccessful. For this, crystallization conditions were optimized in 96-well plates. Prior to flash cooling in liquid nitrogen, crystals were transferred with a cryo-loop into a cryo-protectant solution.

2.2.4.1 Crystallization and cryo-protection

Initial crystallization trails were performed with full-length ceClp1 (450 µM) crystallized in the presence of 1.5 mM adenyl-5'-(β,γ-imido)triphosphate (AppNHp, Jena Bioscience, Jena, Germany) and 1.5 mM G_1C_2 or ADP and 1.5 mM G_1C_2 or 1.5 mM of ATP in a three-fold excess compared to the protein concentration. Crystals were obtained at 20 °C using the sitting drop vapor diffusion method with a reservoir solution composed of 100 mM Na_2HPO_4/KH_2PO_4 between pH 5.0 and 6.0, 200 mM NaCl, 15 mM $MgCl_2$, 90 mM sarcosine and 25 % (*w/v*) polyethylene glycol (PEG) 1000. All crystals that were obtained under the described condition were suitable for soaking experiments. To ensure full occupancy of the soaked ligands, ligand concentration was increased to 10 mM in the cryo-protectant. An artificial apo structure of ceClp1 was obtained by soaking of a ternary complex (ceClp1•ATP•Mg^{2+}) with mother-liquor containing alkaline phosphatase solution. The

inhibited substrate bound state was obtained after incubation of ceClp1 with 10 mM G_1C_2 or U_1G_2 and 10 mM AppNHp (ceClp1-AppNHp-G_1C_2 and ceClp1-AppNHp-U_1G_2). The transition state analog (ceClp1-ADP-AlF_4^--G_1C_2) was obtained after soaking with 10 mM ADP and G_1C_2 dinucleotide and aluminum fluoride. AlF_X was generated *in situ* by addition of 20 mM NaF and 5 mM $AlCl_3$. In order to obtain crystals bound to ATP (ceClp1-ATP), crystals grown in the presence of ADP and subsequently incubated with 10 mM ATP. Residual traces of ADP were removed upon addition of a recycling system that converts ADP to ATP (PK and PEP). Crystals bound to the $G_1A_2A_3A_4$ tetra-nucleotide (ceClp1 RNA 4mer) were obtained by soaking with 10 mM ADP and 5 mM tetra-nucleotide. Finally, crystals bound to their educts were obtained by incubation with 10 mM ATP and G_1C_2 dinucleotide (ceClp1-ADP). The crystallized protein was still enzymatically active. However, those crystals appeared to have slightly suffered from incubation (max. resolution of 2.6 Å) but were virtually identical to crystals soaked with 10 mM ADP and 100 mM $MgCl_2$ (max. resolution of 2.1 Å). For flash cooling in liquid nitrogen, crystals were cryo-protected by addition of 5 % (*v/v*) PEG 600 and 2.5 % (*w/v*) sucrose to their respective soaking solutions. In order to identify the position of the catalytic divalent ion by, Mg^{2+} was replaced by Mn^{2+} in the ATP bound structure.

2.2.4.2 Data collection

Diffraction data were collected at 100 K either at the synchrotron beamline X10SA at the Swiss Light Source (SLS) in Villingen, Switzerland or in-house at the Max Planck Institute for Medical Research. All data sets were processed with the XDS package [133]. Diffraction data were indexed and integrated with XDS and scaled and merged with XSCALE. XDSCONVERT was used to convert files into a format compatible with the programs used for phasing, model building, and refinement. For calculation of a free R-factor, 5 % of reflections from diffraction data sets were randomly assigned and excluded from refinement.

2.2.4.3 Structure determination

Experimental phases for ceClp1 were obtained from a single anomalous diffraction experiment on selenomethionine-labelled protein crystals and experimental phases were obtained using SHELX[134]. An initial model for ceClp1 was built manually using COOT[135] and refined with CNS[136]. Phase extension to a high-resolution native ceClp1 data set was

performed following a rigid-body refinement protocol in CNS[136]. In later stages of ceClp1 structure refinement, iterative cycles of refinement using REFMAC[137] including TLS refinement[138] were performed, followed by manual model improvement using COOT[135]. Water molecules were assigned manually. Phases for different nucleotide-bound forms of ceClp1 were obtained by rigid-body refinement[137] with the refined structure of selenomethionine-labelled ceClp1, followed by cycles of manual model improvement and refinement using REFMAC[137] including TLS refinement[138]. Intermediate and final structures were evaluated with MOLPROBITY[139] and PROCHECK[140]. All figures were drawn using PYMOL[141]. Surface-potentials have been calculated using the APBS plug-in application[142].

2.2.5 Bioinformatic and -computational methods

2.2.5.1 Sequence alignment

Multiple sequence alignments were generated with ClustalW software (http://www.ebi.ac.uk/Tools/msa/clustalw2) and secondary protein structure predictions were obtained by the PSIPRED tool (http://bioinf.cs.ucl.ac.uk/psipred/). Sequence alignments were illustrated using the program ALSCRIPT[143] and the sequence conservation was determined by the AMAS-server[144].

2.2.5.2 Structural modeling of the Pcf11 interaction interface

The interaction interface between ceClp1 and cePcf11 was deduced by superposition of the previous provided crystal structure of scClp1 bound to scPcf11[94] (PDB 2NPI) and the crystal structure of ceClp1 (ceClp1-AppNHp-G_1C_2). Structures of both orthologs were superimposed by secondary-structure matching and the peptide representing scPcf11 was mutated according to a sequence alignment between scPcf11 and cePcf11. Side-chain geometry of cePcf11 was optimized to avoid steric clashes with the structure model of ceClp1. All modeling steps were performed by COOT software[135]. The superposition of individual polypeptide chains was performed using LSQKAB[145].

3 Results

3.1 Structural characterization of ceClp1

This chapter presents crystallization and subsequent structure determination of Clp1 by X-ray crystallography, providing the first three-dimensional structures of a eukaryotic, RPNK. ceClp1 was crystallized in complex with various RNA substrates that differ in length and sequence. The molecular basis of RNA-recognition and RNA-specificity were investigated using these RNA bound crystal structures. In addition, Clp1 was crystallized at important states of its catalytic cycle providing new insights into the phosphoryl transfer reaction mechanism of PNKs.

3.1.1 Bioinformatical characterization of the Clp1 protein family

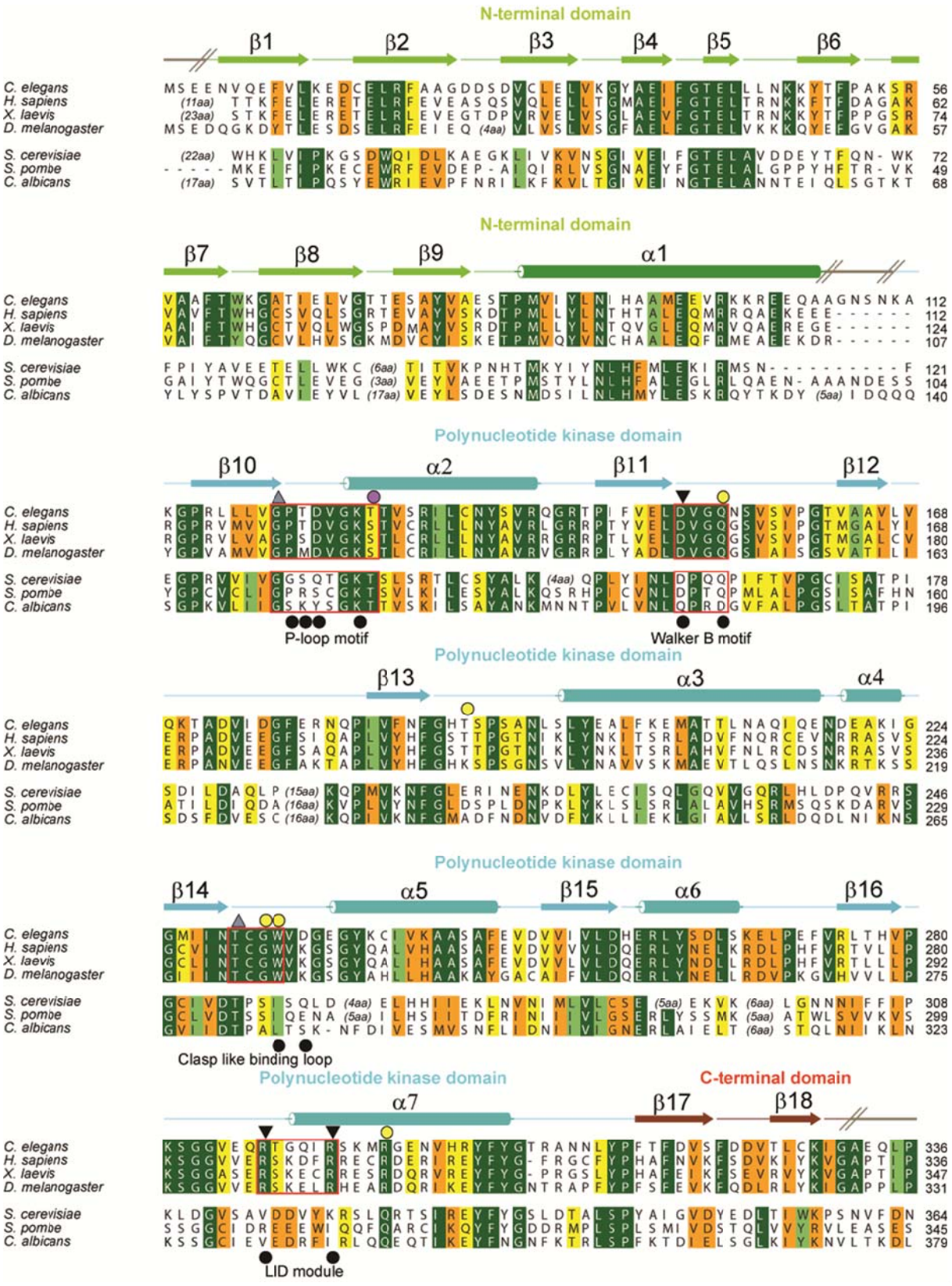
In contrast to enzymatically inactive scClp1[94], the human ortholog functions as a 5'-PNK[88]. Furthermore, hsClp1 is the first identified eukaryotic PNK that is specific for single stranded RNA rather than DNA substrates[85,92]. Unfortunately, hsClp1 is insoluble (Bernhard Loll, personal communication) and thus unsuitable for crystallization experiments. This necessitated the search for other enzymatically active orthologs that could be used for crystallization experiments. The amino acid sequences of Clp1 orthologs from a representative selection of eukaryotic organisms were analyzed by multiple structure-based sequence alignment (Figure 3.1.1). The aim of this bioinformatical characterization was to identify conserved structural features that might be specific to enzymatically active variants of Clp1.

Characteristic for the entire eukaryotic Clp1 protein family is the common P-loop motif (phosphate-binding loop) and the Walker B motif[113,114]. The P-loop motif consists of a flexible loop between strand β10 and helix α2 and shows the consensus sequence GxxxxGK[S/T] (Figure 3.1.1). The distal Walker B motif consists of a conserved aspartate residue located after strand β11. The P-loop and the Walker B are involved in substrate binding and enzyme catalysis[19,113]. Inspection of multiple sequence alignment within higher eukaryotes revealed two additional conserved motifs.

The first constitutes the LID module, a mostly helical structure located within α7, while the second is a clasp-like binding loop (referred to as clasp) located after strand β14.

The LID module contains the conserved RxxxxR sequence and was suggested to stabilize the transition state by neutralizing emerging negative charges of the transferred phosphoryl group as known for other P-loop kinases[118]. A comparison of orthologs from higher eukaryotes with Clp1 from fungi revealed that both or at least one of the conserved arginine residues is replaced by hydrophobic amino acids (Fig. 3.1.1). The second motif, the clasp, has the consensus sequence TxGW and plays a major role in RNA-binding. In contrast to higher eukaryotes, yeast orthologs showed no conservation of the clasp motif.

The low sequence conservation of Clp1 from fungi compared to higher eukaryotes suggests that scClp1 has lost its enzymatic activity during evolution by distinct mutations. ATP-binding by the P-loop motif, however, remained unaffected[94,98-100]. Orthologs from higher eukaryotes, in contrast, show high sequence conserved within the four critical motifs (Figure 3.1.1). Thus, it is most likely that orthologs from higher eukaryotes function in a similar way as shown for hsClp1. Interestingly, ceClp1 carries a unique insertion in the loop region connecting the NtD with the PNK domain. Since this insertion might have a positive affect on Clp1's protein solubility, the protein was purified and subsequently applied to structure determination experiments. ceClp1 has the highest homology to Clp1 from *Xenopus laevis* with a sequence identity of 78.7 %. However, it also shows a significant homology to both, hsClp1 and Clp1 from *Drosophila melanogaster*, with a sequence identity of 48.0 % and 42.0 %, respectively.


N-terminal domain
β1
β2
β3
β4
β5
β6
C. elegans
H. sapiens
X. laevis
D. melanogaster
S. cerevisiae
S. pombe
C. albicans
N-terminal domain
β7
β8
β9
α1
Polynucleotide kinase domain
β10
α2
β11
β12
P-loop motif
Walker B motif
Polynucleotide kinase domain
β13
α3
α4
Polynucleotide kinase domain
β14
α5
β15
α6
β16
Clasp like binding loop
Polynucleotide kinase domain
C-terminal domain
α7
β17
β18
LID module

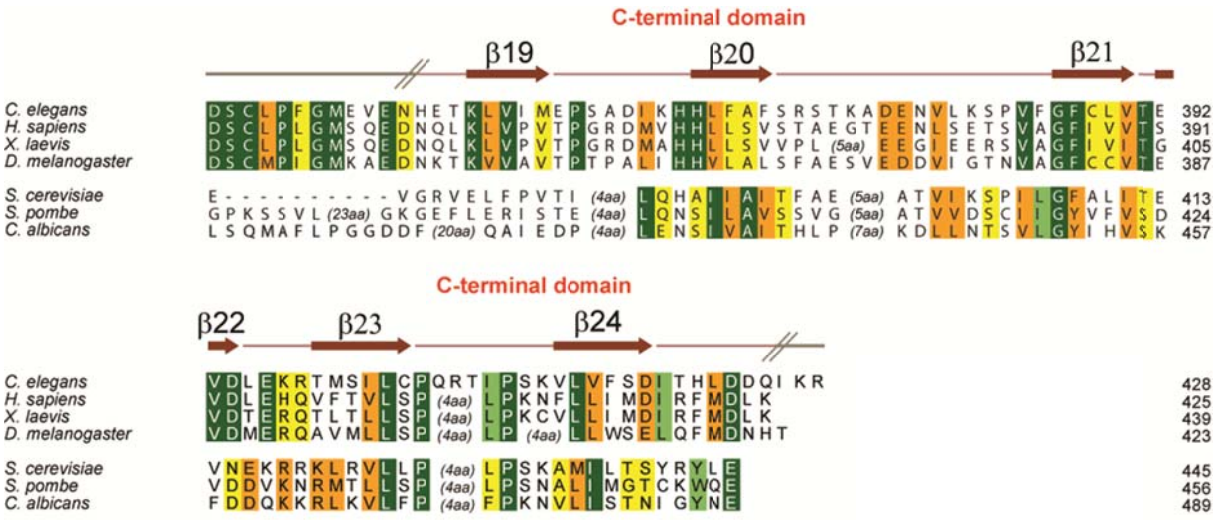


Figure 3.1.1 Sequence comparison of Clp1 orthologs. Sequence alignment of Clp1 in both higher eukaryotes (*Caenorhabditis elegans*: NP_001040858; *Homo sapiens*: NP_001136069; *Xenopus laevis*: NP_001084787, and *Drosophila melanogaster*: NP_610876) as well as fungi (*Saccharomyces cerevisiae*: NP_014893, *Schizosaccharomyces pombe*: NP_593741, and *Candida albicans*: XP_712695). The secondary structure elements are shown above the alignment, with α-helices depicted as cylinders and β-strands as arrows, assigned according to the ceClp1 crystal structure. The coloring of the secondary structure corresponds to the domain color from Figure 3.1.4 (NtD: green, PNK domain: cyan, and CtD: red). The P-loop motif, the Walker B motif, the LID module, and the clasp are highlighted by red boxes. Residues involved in the switch on/off of Lys127 are marked with a gray triangle, whereas residues, which interact with the RNA, are marked with a yellow circle. Catalytically important residues are indicated by black triangles. The Walker A threonine is involved in the coordination of the metal cofactor (purple circle). Residues subjected to site-directed mutagenesis in ceClp1 are marked with a black circle below the alignment. Conserved residues are colored, according to the degree of conservation decreasing from dark green to yellow.

3.1.2 Purification of ceClp1

To express and purify recombinant ceClp1, the cDNA sequence was cloned via standard methods into the bacterial expression vector pET28b. The sequence-verified expression plasmid was transformed into *E. coli* BL21-CodonPlus(DE3)-RIL cells allowing for an IPTG-inducible overexpression. The protein construct was designed to encode for a fusion variant of ceClp1 with an N-terminal hexahistidine tag. The resulting recombinant protein was overexpressed, cells were lysed using sonication, and the protein was then purified from the supernatant. Overexpressed ceClp1 was found in the soluble fraction of *E. coli* lysates. Further purification included three steps using Ni^{2+} affinity, anion exchange and size exclusion chromatography. The protein solution was applied to Ni^{2+}-resin. After a washing step, the protein was eluted with imidazole. The N-terminal hexahistidine tag was removed by a proteolytic digest using thrombin and the wild type ceClp1 protein was further purified by anion exchange (MonoQ) and size exclusion chromatography. A typical purification procedure yielded three milligrams of protein per liter of cell culture. Homogeneity was

verified by SDS-PAGE of peak fractions, showing the presence of a single protein band migrating to a molecular weight of around 54 kDa, which is in agreement with the calculated mass (Figure 3.2.1).

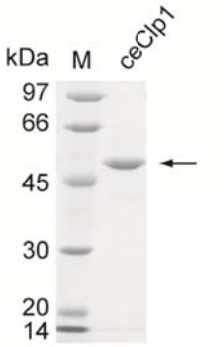


Figure 3.1.2 Expression and purification of ceClp1 in *E. coli*. Aliquot of ceClp1 obtained after Ni^{2+}-NTA, anion exchange, and size exclusion chromatography were analyzed by SDS-PAGE (15 %). The gel was stained with Coomassie Blue confirming purification of ceClp1 to high homogeneity. M: low molecular weight marker.

3.1.3 Co-crystallization of ceClp1

During the last years, significant progress has been made in analyzing the biochemical properties of hsClp1[88,92]. Furthermore, the crystal structure of inactive scClp1 was published in 2001 at medium resolution[94]. Nevertheless, the molecular mechanism of Clp1's RNA-specificity and the understanding of the phosphoryl transfer reaction on a structural level remained elusive. This thesis, therefore, aimed for crystallization of ceClp1 in the ligand-free form and in complex with its ligands. ceClp1 was purified (see 3.1.2) and used for initial crystallization experiments at a concentration of 650 µM. Crystallization trials of a ligand-free ceClp1 remained unsuccessful, possibly due to domain flexibility of *apo*-ceClp1 preventing crystal formation. However, co-crystallization of 450 µM ceClp1 together with a non-hydrolysable ATP analog (AppNHp) and an RNA dinucleotide (G_1C_2), both in a three-fold excess (~1.4 µM) relative to protein, finally led to spontaneous crystal formation. Screening was performed in a sitting drop vapor diffusion setup in a 96-well plate format. Initial crystals of ceClp1, co-crystallized with AppNHp and G_1C_2, were obtained after 12-48 hours in a buffer containing 20 % *(w/v)* PEG 1000, 100 µM Na/K phosphate buffer pH 6.2 and 200 µM NaCl (Figure 3.1.3 A). These conditions were further optimized by screening with varying PEG 1000 concentrations (20-25 % (*w/v*)) in a pH range from 5.5 to 7.5. Crystallization was further improved using an additive screen. The additive screen identified two components that supported crystal growth towards single crystal morphology. The first was $MgCl_2$ at a concentration of 15 mM, the second was sarcosine at a concentration of 90 mM. The optimization procedure finally led to needle-shaped single crystals of a maximal size of about 100-200 µm growing in a precipitant solution containing 22 % *(w/v)* PEG 1000,

100 mM Na/K phosphate buffer pH 6-7, 200 mM NaCl, 15 mM $MgCl_2$, and 90 mM sarcosine (Figure 3.1.3 B). Crystal growth was achieved only with a 96-well plate, while up-scaling into a 24-well plate format (Linbro plates) remained unsuccessful. During crystal harvesting, a cryo-protection solution was used preventing ice formation during flash-cooling in liquid nitrogen. This cryo-protectant consisted of mother-liquor with additional 5 % (*v/v*) PEG 600 and 2.5 % (*w/v*) sucrose. The cryo-protectant solution was supplemented with AppNHp and G_1C_2 at the same concentration as used for co-crystallization in order to avoid substrate diffusion from the active site of the protein crystals. Based on diffraction data obtained from the crystals, an electron density map was generated showing both a bound AppNHp molecule and a Mg^{2+} ion in the active site (ceClp1-AppNHp-Mg^{2+}), whereas no electron density for the RNA molecule was observed.

The crystal structure of ceClp1-AppNHp-Mg^{2+} revealed that crystallization of ceClp1 depends on the presence of a nucleoside triphosphate, whereas presence of G_1C_2 during crystallization experiments was not required. Thus, ceClp1 was also co-crystallized with a three-fold excess of ATP (1.3 µM) or ADP (1.3 µM) using the same precipitant conditions as described for growth of ceClp1-AppNHp-Mg^{2+} crystals. In both cases crystals started to grow after one day. However, no electron density for the Mg^{2+} ion was observable for the ceClp1-ADP complex. Increasing the $MgCl_2$ concentration in the cryo-protectant up to 100 mM finally enabled Mg^{2+} incorporation, leading to a ternary complex of ceClp1 (ceClp1-ADP). Since hsClp1 was shown to use different divalent metal cofactors[85], Mg^{2+} and Mn^{2+} were tested for crystallization experiments. Crystals co-crystallized with ATP were transferred to a cryo-protectant solution in which $MgCl_2$ was replaced with 15 mM of $MnCl_2$ (ceClp1-ATP). Furthermore, in these experiments a ATP recycling system was added to the cryo-protectant to remove residual traces of ADP arising from spontaneous ATP hydrolysis. This recycling system consisted of PK that converts ADP to ATP using PEP as substrate. The crystals were transferred to the cryo-protectant solution and incubated for 30 minutes. In contrast to soaking experiments, in which no recycling-system was added, a mixture of ADP/ATP was observed in the electron density map. To identify ceClp1's *in vivo* substrate, attempts to crystallize ceClp1 with double stranded DNA duplexes or with an RNA anti-codon loop analog were made, but co-crystallization was unsuccessful.

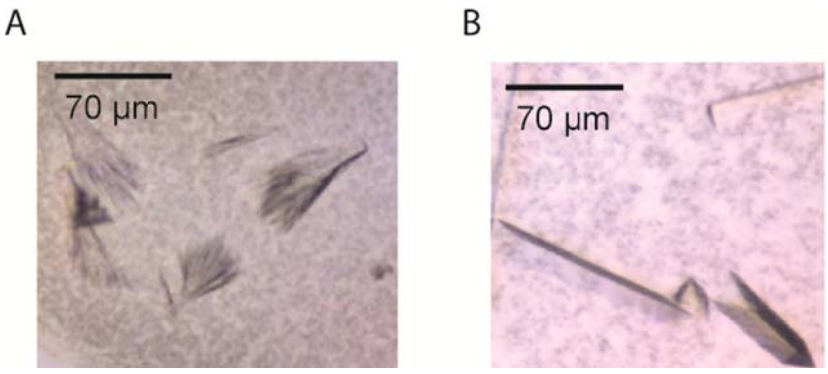


Figure 3.1.3 Crystals of ceClp1. A, Intergrown crystals appear after one day by using a precipitant solution of 20 % (*w/v*) PEG 1000, 100 mM Na/K phosphate buffer pH 6.2, 200 mM NaCl. **B,** Crystallization conditions were optimized to 22 % (*w/v*) PEG 1000, 100 mM Na/K phosphate buffer pH 6-7, 200 mM NaCl, 15 mM $MgCl_2$ and 90 mM sarcosine, resulting in growth of single needle-shaped crystals.

To obtain a crystal structure of *apo*-ceClp1, crystals of the ceClp1-ADP complex were soaked in mother-liquor supplemented with alkaline phosphatase. After incubation for 2 hours, the crystals were harvested and transferred into cryo-protected solution (*apo*-ceClp1).

To characterize ceClp1's RNA-binding site, the crystal structure of an RNA bound form was required. Whereas co-crystallization was unsuccessful, soaking experiments with ceClp1 crystals resulted in crystals with bound RNA substrates. Crystals were transferred into a cryo-protectant containing a ~25-fold excess of AppNHp (10 mM) and of an RNA dinucleotides (G_1C_2 and U_1G_2; 10 mM) relative to the protein concentration, leading to the crystal structure of ceClp1-AppNHp-G_1C_2 and ceClp1-AppNHp-U_1G_2. To ensure full occupancy of the ligand-binding site, these crystals were incubated for 2 hours in the cryo-protectant solution and subsequently harvested for flash-cooling.

Based on soaking experiments, it was possible to capture the first transition state analog of a PNK. To obtain this transition state analog, crystals of ceClp1-ADP were soaked in a cryo-protectant solution containing a ~25-fold excess of ADP (10 mM) and of G_1C_2 (10 mM) relative to the protein concentration, 100 mM of $MgCl_2$ and aluminum fluoride (AlF_x) as transition state analog. This compound is generally used as a transition state analog[121] and was generated *in situ* by addition of 20 mM NaF with 5 mM $AlCl_3$. This resulted in a tetrafluoroaluminate (AlF^-_4) transition state analog. Since it was previously shown that the coordination number of aluminum depends on the pH at, which the enzyme is crystallized (pH 4.5 leads to AlF^-_4 and pH 8.5 leads to trifluoroaluminate (AlF_3))[146], the buffer system of the cryo-protectant was changed from Na/K phosphate buffer to a HEPES buffer system (22 % (*w/v*) PEG 1000, 50 mM HEPES-NaOH pH 7, 200 mM NaCl, 15 mM $MgCl_2$,

90 mM sarcosine). Furthermore, soaking experiments were performed with pH ranging from 8 to 9, in order to induce formation of an AlF_3 transition state analog. However, at high pH it was impossible to trap a mimic and no AlF_3 was bound.

Based on the electron density map, it is not possible to distinguish between an AlF^-_4 and a tetrafluoromagnesate (MgF_4^{2-}) transition state analog[147]. Since both analogs would have been possible, two control soaking experiments were performed without $AlCl_3$. Both experiments differed in the type of the divalent metal cofactor. The 100 mM $MgCl_2$ of the cryo-protectant solution of the first experiment were replaced with 100 mM of $MnCl_2$ for the second experiment. Capturing of the transition state analog without $AlCl_3$ would indicate the existence of an MgF_4^{2-} or tetrafluoromanganate (MnF_4) transition state analog in the active site of ceClp1. Furthermore, a transition state analog formed by MnF_4 could easily be identified by an anomalous scattering signal of the manganese atom. However, no electron density for a transition state analog for MgF_4^{2-} and MnF_4 was obtained, suggests that indeed AlF_4^- was bound. Additionally, another trial was performed aiming for capturing the transition state analog using $G_1A_2A_3A_4$. Previous observations with the transition state mimic suggested full occupancy of the RNA-binding site at much lower concentrations during a transition state as compared to a substrate bound state. The cryo-protectant was, therefore, supplemented with 5 mM of $G_1A_2A_3A_4$. Analysis of crystals soaked under these conditions showed no transition state analog, instead those crystals resemble a product bound state (ceClp1-4mer).

Additionally, soaking experiments showed that ceClp1 remains enzymatically active within the crystal. To obtain crystals captured with their educts, cryo-protectant was supplemented with a ~25 fold excess of ATP and G_1C_2. Analysis of the diffraction data from those crystals showed reaction product formation (ADP) in the electron density map, whereas phosphorylated RNA was missing. These structures were virtually identical to the ceClp1-ADP crystal. Since turnover crystals appeared to have slightly suffered from incubation (max. resolution of 2.6 Å), the ceClp1-ADP structure (max. resolution of 2.1 Å) is referred to as RNA released product bound state.

All diffraction data were collected either at the PXII beamline at the SLS in Villigen, Switzerland, or in-house using a Rigaku MicroMax 007 HF microfocus X-ray generator. Space group, unit cell parameters, and the overall Wilson B-factor are given in Table 3.1.1. All crystals belonged to a trigonal crystal symmetry system and the space group was $P3_1$.

3.1.4 *De novo* phasing of ceClp1 by single-wavelength anomalous diffraction experiments

Experimental phases of ceClp1 were obtained by *de novo* phasing using a single-wavelength anomalous diffraction (SAD) experiment[148]. To obtain selenomethionine labeled protein, Clp1 was overexpressed similar to native protein, except that cells were grown in minimal medium, in which the amino acid methionine was substituted with its seleno-derivate[127]. Protein with incorporated selenomethionine was purified (see 2.2.2.3) and crystallized similar to the native protein. Precipitant conditions were similar to the description for native protein (22 % *(w/v)* PEG 1000, 100 mM Na/K phosphate buffer pH 6-7, 200 mM NaCl, 15 mM $MgCl_2$, and 90 mM sarcosine). Flash-cooling was achieved with a cryo-protectant containing the mother liquor supplemented with 5 % (*v/v*) PEG 600, 2.5 % (*w/v*) sucrose, 1.3 mM AppNHp, and 1.3 mM G_1C_2 (ceClp1-SeMet). A highly redundant data set from selenomethionine crystals that diffracted up to 2.3 Å resolution was collected at the absorption edge of selenium. Collected diffraction images were indexed, integrated, and scaled using the XDS package[133].

The positions of the selenium atoms were localized and native phases were determined using SHELXD[149] and SHELXE[134], respectively. The SHELX programs are embedded in the graphical user interface HKL2MAP that facilitates usage[150]. SHELXD identified 11 selenium atom positions with a correlation coefficient of 31.2 % (all) and 17.9 % (weak) as the best solution. The pseudo-free correlation coefficient, which is calculated by randomly leaving out 10 % of the reflections, was 54 %.

Similar to the observations made for native crystals, the SeMet-Clp1 crystals belong to the trigonal crystal system and have the space group $P3_1$. The crystals contain one molecule per asymmetric unit. Seven selenomethionine positions could unambiguously be identified and were used to calculate an initial electron density map. An initial model for ceClp1 was manually built using COOT[135] and refined with CNS[136]. In later stages of ceClp1 structure refinement, iterative cycles of refinement were performed using REFMAC[137] including TLS refinement[138] with manual model improvement using COOT[135]. They led to a final R_{work} and an R_{free} of 19.2 % and 24.0 %, respectively, at a resolution of 2.3 Å.

3.1.5 Phasing and refinement of the different ceClp1 complexes

Diffraction data of the different nucleotide free-/bound-forms of ceClp1 were processed with the XDS package[133] and phases were obtained by rigid-body refinement[137] using the refined model of selenomethionine-labelled ceClp1 as a starting model. Once phases were calculated, models were manually improved by COOT[135] and subsequently refined using REFMAC[137] with restrained maximum likelihood refinement including TLS refinement[138]. Water molecules were assigned manually. Data collection and refinement statistics are summarized in Table 3.1.1.

3.1.6 Crystal structure of *apo*-ceClp1

The overall architecture of ceClp1 can be subdivided into three distinct domains, a central PNK domain flanked by additional N- and C-terminal domains (NtD and CtD, Figure 3.1.4 A and B). Diffraction data of the apo-structure had a maximum resolution of 2.1 Å. The structure model was refined to an R-factor of 18.8 % with an R_{free} of 23.2 % (Table 3.1.1; ceClp1-apo). Based on the electron density map, a continuous polypeptide chain starting at residue 4 and ending at residue 425 could be modeled. Only two unassigned loop regions from residue 105 to residue 111 and from residue 331 to residue 349 could not be modeled due to ambiguous electron density. Furthermore, the electron density map revealed the presence of two inorganic phosphate molecules (P_i). One P_i is bound to the P-loop motif, positioned at the expected location of β-phosphate of an ATP molecule[113]. The other P_i is bound to an arginine residue (Arg406) at the molecular surface of the protein. There was also additional density observed for a PEG molecule, which was at a van der Waals distance to a hydrophobic patch on the surface of ceClp1 (composed of residues Trp62, Leu88, His91, Try137, and Arg140).

The PNK domain is the central domain of the ceClp1 structure, spanning from residue 110 to residue 310. It contains the active site region, and the fold of the PNK domain is structurally related to T4 PNK (r.m.s.d.: 3.1 Å)[51,52]. The PNK domain of ceClp1 is composed of an α/β fold also called Rossmann fold[151]. Characteristic for the Rossmann fold is a central β-sheet flanked by α-helices. The central β-sheet has a topological arrangement of four parallel strands in a 4-1-3-2 order, similar to the previously described T4 PNK[51,52]. A diagram describing the topological arrangement of secondary structures within ceClp1 is presented in Figure 3.1.4 C. However, in contrast to T4 PNK, ceClp1 forms three additional strands

leading to a seven-stranded central β-sheet. Whereas the PNK domain is structurally conserved and related to members of the kinase protein family[19], the additional NtD and CtD constitute novel structural features. A structural comparison of this domain from ceClp1 with the PBD database using the DALI server[152] revealed the greatest structural similarities between ceClp1 and the yeast ortholog[94].

Table 3.1.1 X-ray diffraction data collection and refinement statistics of ceClp1.

Data collection Data set	ceClp1-SeMet	ceClp1-ATP	ceClp1-ADP	ceClp1-AppNHp-G_1C_2
PDB entry				
Space group	$P3_1$	$P3_1$	$P3_1$	$P3_1$
Unit cell (*a,b,c* [Å])	100.7, 100.7, 40.3	99.3, 99.3, 40.8	100.0, 100.0, 40.4	100.4, 100.4, 40.3
Wavelength [Å]	0.9786	0.9793	0.9792	0.9786
Resolution [Å] [a]	50-2.3 (2.4-2.3)	50-2.3 (2.4-2.3)	50-2.0 (2.1-2.0)	50-2.0 (2.1-2.0)
No. reflections [a]	40589 (3026)	19707 (2323)	30056 (4086)	30691 (7581)
Completeness [%] [a]	99.7 (99.5)	98.5 (97.0)	98.6 (99.1)	99.9 (99.8)
$<I/\sigma(I)>$ [a]	13.2 (2.9)	12.4 (4.8)	14.8 (2.4)	17.2 (3.8)
R_{meas} [a, b]	9.6 (57.6)	8.5 (33.4)	5.8 (59.9)	6.2 (46.7)
Redundancy	4.8 (4.7)	3.6 (3.6)	3.5 (3.5)	4.4 (4.3)
Refinement				
R_{work} [a, c]	19.2 (32.9)	18.2 (22.3)	19.1 (43.6)	18.2 (28.8)
R_{free} [a, d]	24.0 (42.4)	23.5 (35.4)	24.2 (47.6)	22.8 (31.1)
Overall B factor [$Å^2$]	31.5	39.8	38.1	36.2
Wilson [$Å^2$]	43.8	46.1	45.8	42.1
r.m.s.d. [e] from ideal geometry: bond length [Å]	0.012	0.011	0.009	0.011
bond angles [°]	1.312	1.275	1.135	1.290
Ramachandran statistics:				
allowed [%]	92.3	92.1	91.8	92.1
additionally allowed [%]	7.7	7.9	8.2	7.9

Table 3.1.1 X-ray diffraction data collection and refinement statistics of ceClp1. Continue.

Data collection Data set	ceClp1 RNA 4mer	ceClp1 ADP-AlF_4^--G_1C_2	ceClp1 AppNHp-U_1G_2	ceClp1-apo
PDB entry				
Space group	$P3_1$	$P3_1$	$P3_1$	$P3_1$
Unit cell (*a,b,c* [Å])	99.7, 99.7, 40.8	100.7, 100.7, 40.4	100.7, 100.7, 40.3	99.42, 99.42, 40.40
Wavelength [Å]	0.9786	1.5418	0.9782	0.9792
Resolution [Å] [a]	43-2.4 (2.5-2.4)	50-2.2 (2.4-2.2)	43.6-2.1 (2.2-2.1)	49.7-2.3 (2.4-2.3)
No. reflections [a]	17640 (2001)	22553 (2442)	26620 (3434)	19795 (2374)
Completeness [%] [a]	99.6 (99.1)	96.9 (83.7)	99.9 (99.8)	99.7 (99.7)
$\langle I/\sigma(I)\rangle$ [a]	17.9 (2.7)	18.9 (7.0)	12.62 (4.44)	17.34 (3.46)
R_{meas} [a, b]	6.8 (56.4)	5.9 (22.7)	14.0 (48.2)	5.5 (41.8)
Redundancy	3.9 (3.7)	3.1 (3.0)	6.0 (5.3)	3.51 (3.59)
Refinement				
R_{work} [a, c]	18.5 (34.7)	22.7 (56.2)	20.2 (41.3)	19.0
R_{free} [a, d]	21.9 (35.9)	28.0 (57.7)	25.0 (40.7)	23.3
Overall B factor [$Å^2$]	34.1	26.5	21.9	35.1
Wilson [$Å^2$]	58.1	36.8	42.0	40.0
r.m.s.d. [e] from ideal geometry: bond length [Å]	0.013	0.009	0.010	0.010
bond angles [°]	1.355	1.155	1.330	1.090
Ramachandran statistics:				
Allowed	92.4	91.7	92.4	96.3
additionally allowed	7.6	8.3	7.6	3.7

[a] values in parentheses refer to the highest resolution shell

[zb] $R_{meas} = \Sigma_h [n/(n-1)]^{1/2} \Sigma_i \mid I_h - I_{h,i} \mid / \Sigma_h \Sigma_i I_{h,i}$.where I_h is the mean intensity of symmetry-equivalent reflections and n is the redundancy.

[c] $R_{work} = \Sigma_h \mid F_o - F_c \mid / \Sigma F_o$ (working set, no σ cut-off applied)

[d] R_{free} is the same as R_{work}, but calculated on 5% of the data excluded from refinement.

[e] root-mean-square deviation (r.m.s.d.) from target geometries.

Besides this structural conservation, the PNK domain of ceClp1 also showed a significant structural homology to a signal recognition particle GTPase[153] (*Z*-score 10.8, r.m.s.d.: 2.6 Å, and a sequence identity of 16 %; PDB accession code: 2J7P, chain: B).

The NtD covers residues 4 to 105 and is composed of a sandwich of two β-sheets. The electron density map did not allow modeling of the first three amino acids, suggesting that they are disordered. In contrast to the PNK domain, a DALI-server search[152] in the PDB database revealed no significant similarity to any solved structure. The most similar fold identified was Mif2p, a conserved DNA-binding kinetochore protein[154] (*Z*-score 8.4, r.m.s.d.: 1.8 Å, and 12 % sequence identity, PDB; accession code: 2VPV, chain: B).

The CtD is slightly larger than the NtD and encompasses residues 311 to 425, folded in a mixture of β-strands and random coils. For the CtD, a PDB database search[152] revealed low structural homology to the α-subunit of an ATP synthase[155] (*Z*-score: 4.5, r.m.s.d.: 2.5 Å, and a sequence identity of 9 %; PDB accession code: 3OEH, chain: T). Based on previously presented structural data of yeast Clp1 (scClp1), it was shown that the interface between the central PNK domain and CtD is involved in binding of Pcf11[94]. Pcf11 together with scClp1 are subunits of the cleavage factor CF IA involved in the 3'-end processing machinery of eukaryotic Pol II transcripts[102].

Based on a sequence alignment, the active site region of ceClp1 at the PNK domain can be divided into an ATP- and an RNA-binding site that contains four recognizable structural motifs associated with substrate binding and enzyme catalysis (Figure 3.1.1). The most prominent motif for kinases is the P-loop sequence ^{121}GxxxxGKT^{128}. Besides the P-loop, ceClp1 contains a Walker B motif ^{151}DxxQ^{154}, and two motifs that are only conserved among Clp1 family members of higher eukaryotes. These motifs are a LID module ^{288}RxxxxR^{293} and a clasp ^{230}TxCGW^{233}. To structurally characterize the substrate binding properties of ceClp1, crystals of the complex were generated by soaking with AppNHp, Mg^{2+} and G_1C_2 (inhibited substrate bound state) or by co-crystallization with ATP-Mn^{2+} (substrate bound state), and then analyzed.

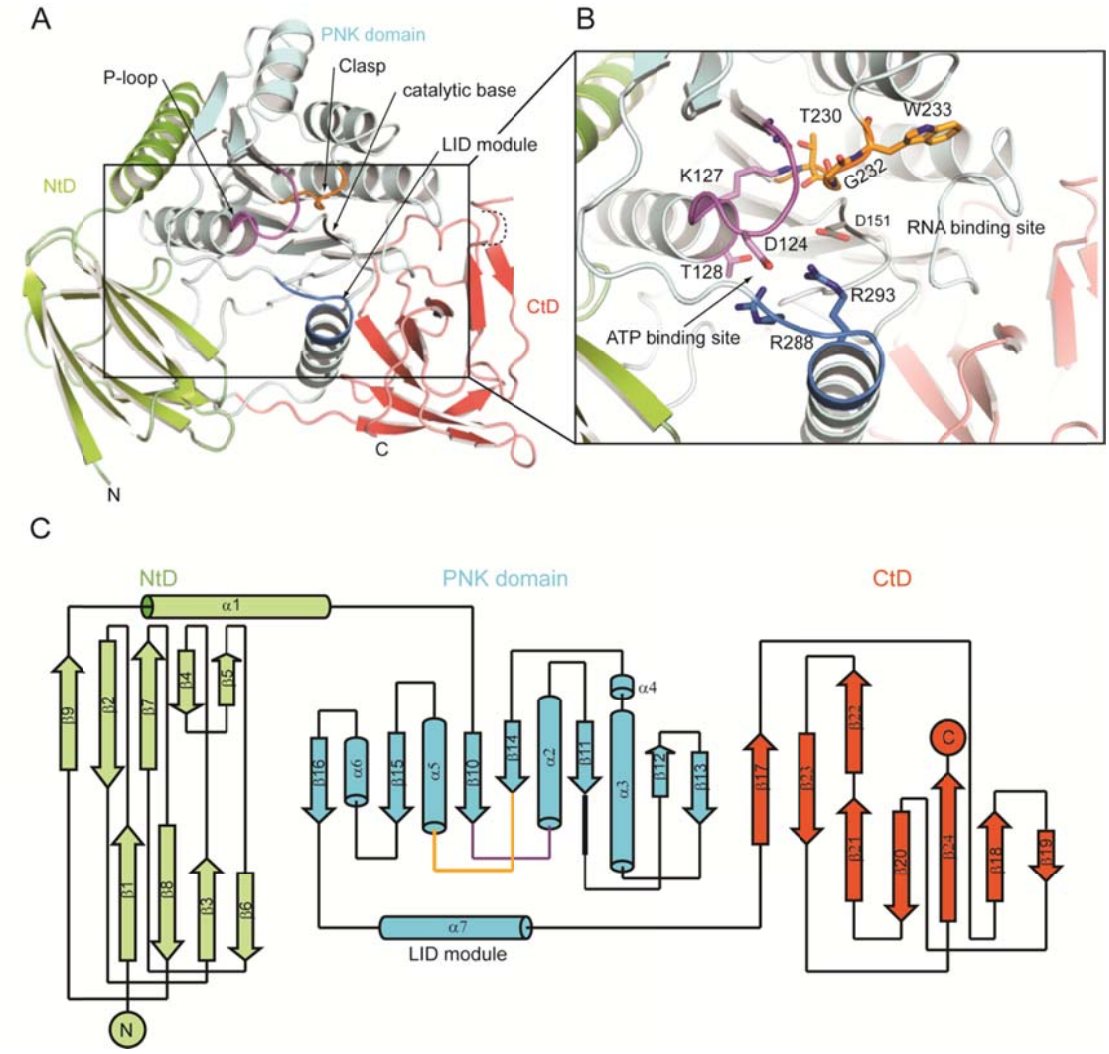


Figure 3.1.4 Structure of RNPK ceClp1. A, A ribbon representation of ceClp1 with the N-terminal domain (NtD), polynucleotide kinase domain (PNK domain) and the C-terminal domain (CtD). Structural motifs important for ligand-binding or catalysis are highlighted in color; P-loop (purple), clasp (orange), LID module (blue), catalytic base (black). **B,** Close-up view of the active site in an identical orientation and color code. Residues important for substrate binding and catalysis are shown as stick representations. **C,** Diagram showing the topological arrangement of secondary structure elements in ceClp1.

3.1.7 RNA-binding site

This thesis provides the first crystal structure of a eukaryotic RPNK bound to an RNA substrate. The crystal structure of ceClp1 bound to an RNA dinucleotide was solved and refined to an R-factor of 18.2 % and an R_{free} of 22.8 % at a resolution of 2.0 Å (Figure 3.1.5 A).

Crystals were soaked with AppNHp and G_1C_2 representing an inhibited substrate-bound state. The RNA-binding site could be localized in a cleft and showed a positively charged electrostatic surface potential (Figure 3.1.5 B). Both nucleotides of the

RNA were clearly visible and were found to bind to the conserved clasp comprising the consensus sequence TxGW. On the one hand, the indol function of the tryptophan residue (Trp233) mimics a base stacking interaction to the ultimate base (G_1) and provides an initial platform for RNA-binding (Figure 3.1.5 A). On the other hand, the penultimate base (C_2) is stacked against G_1, thereby perfectly sandwiching G_1 between the side chain of Trp233 and C_2 in a clasp-like binding mechanism. Apparently, G_1 gets locked by Trp233 and C_2. This observation suggests that ceClp1 requires two bases as minimal substrate length for efficient binding (see 3.2.2). G_1 gets locked to the Trp233 binding platform only in the presence of a second base.

Another interaction between clasp and RNA substrate is mediated by the main-chain amide of a glycine residue (Gly232) (Figure 3.1.5 A). Gly232 forms a hydrogen bond to the 5'-hydroxyl group of the ultimate ribose. The 5'-hydroxyl group is also recognized by catalytic aspartate of the Walker B motif (Asp151), forming an additional hydrogen bond. These interactions most likely are involved in the correct positioning of the RNA substrate. The free 5′-hydroxyl group of RNA points towards the γ-phosphate of the ATP molecule, as expected for an in-line mechanism[52].

ceClp1 also interacts with the oxygen atoms of the bridging phosphate backbone. A glutamine residue Gln154 forms a hydrogen bond to an oxygen atom of the phosphate group. Furthermore, guanidinium groups of Arg293 and Arg297 contribute to RNA-binding by salt-bridge formation to the phosphate groups of the RNA. Interestingly, ceClp1 was shown to phosphorylate G_1C_2 with the same efficiency as longer single-stranded RNA oligonucleotides (Bernhard Loll, personal communication). To explain this biochemical observation, the structure of ceClp1 bound to a longer RNA oligonucleotide ($G_1A_2A_3A_4$) was solved (Figure 3.1.5 C). The structure model was refined to an R-factor of 18.5 % and an R_{free} of 22 % at a resolution of 2.4 Å. An unambiguous electron density was visible for all nucleotides. Binding modes between the dinucleotide and the oligonucleotide were found to be virtually identical (r.m.s.d.: of 1.2 Å, Figure 3.1.5 D). After the first two nucleotides, no additional interaction contacts between RNA and ceClp1 were observed. Interestingly, a water molecule (W41) visible in the dinucleotide structure interacts with the 3'-hydroxyl group of C_2 and thereby indicates the putative position of the 3'-5' bridging phosphate group of a longer substrate (Figure 3.1.5 A). Indeed, the position of the bridging phosphate group found in the oligonucleotide crystal structure is identical to the position of W41. The second phosphate group forms a salt bridge to Arg297. The Watson-crick base pairing interfaces of the nucleotides are solvent exposed. In both structures, RNA models perfectly match an ideal

RNA in the A-form conformation and base moieties are in *anti*-conformation. Furthermore, the sugar moieties were in the favoured 3'-endo conformation.

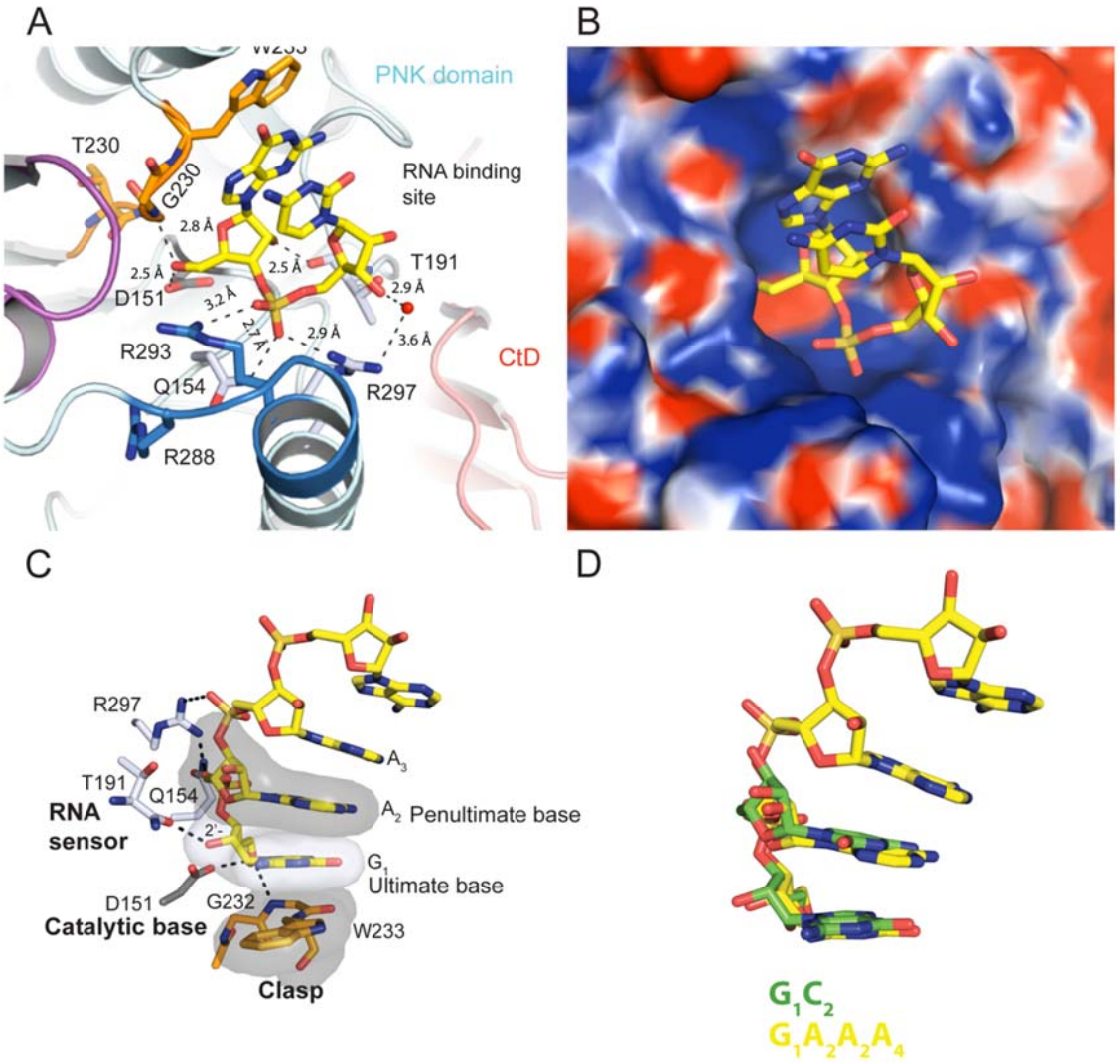


Figure 3.1.5 Structural basis of RNA-binding by ceClp1. A ribbon representation of ceClp1 with polynucleotide kinase domain (PNK domain) and C-terminal domain (CtD). Structural motifs important for ligand-binding or catalysis are highlighting in color; P-loop (purple), clasp (orange), LID module (blue), catalytic base (black). Dinucleotide (yellow), and water molecules (red) are shown in stick and sphere representations. The ultimate base (G_1) is sandwiched in the clasp between Trp233 and the penultimate base (C_2). Trp233 is part of the clasp. **B,** ceClp1 with bound G_1C_2 is shown as surface representation colored according to the electrostatic surface potential (contouring from +5 kT/e in blue to −5 kT/e in red). **C,** stick representation of the RNA-binding site bound to $G_1A_2A_3A_4$. The ultimate base (G_1) is sandwiched in the clasp between Trp233 and penultimate base (A_2). Trp233 is part of the clasp. **D,** Superposition of G_1C_2 and $G_1A_2A_3A_4$ showing a virtually identical conformation.

The 2‘-hydroxyl group of the ultimate base is recognized by a hydrogen bond formed with the main chain carbonyl of a threonine residue (Thr191, RNA-sensor). This hydrogen bond could be an explanation for hsClp1’s ability to discriminate single stranded RNA against single stranded DNA[88]. Interestingly, based on the oligonucleotide conformation, one could model an ideal double stranded A-form RNA bound to ceClp1 without any steric hindrance by superposition with the single stranded RNA (Figure 3.1.6 A). The modeled double-strand also displays no steric clashes for the model of longer double stranded RNA as a linear substrate. In contrast to the 5’-acceptor strand, the modeled complementary strand showed no contact with the protein except when substrates with a 3’-overhang were used (Figure 3.1.6 B). In conclusion, these two crystal structures explain Clp1’s ability to phosphorylate both single stranded and double stranded RNA substrates[88] (see 3.2.2).

Although the interactions of ceClp1 with the RNA backbone as well as the base-stacking interaction imposed no sequence specificity for the RNA substrate, the effect of sequence variation on RNA-binding was verified. For this, the dinucleotide U_1G_2 was soaked into crystals of ceClp1 and the third structural model of ceClp1 bound to RNA was obtained. This crystal structure was refined to an R-factor of 20.2 % and an R_{free} of 25 % at a resolution of 2.1 Å (Figure 3.1.7). The guanosine and the uracile nucleotides of both dinucleotide bound structures were found to be virtually identical positioned (Figure 3.1.7 A and B), mostly through the clasp-like binding mechanism and interactions with the sugar phosphate backbone. Furthermore, both crystal structures showed unspecific Watson Crick-like hydrogen bonds. However, the penultimate base of the U_1G_2 structure was flipped away into a *syn*-conformation and thus did not perfectly stack with the ultimate base (Figure 3.1.7 B). The observation of a *syn*- instead of an *anti*-conformation of the penultimate base of U_1G_2 is most likely an artifact due to the use of dinucleotides. For longer nucleotides, the penultimate base will most likely be in an *anti*-conformation due to stacking interactions with neighboring nucleotides.

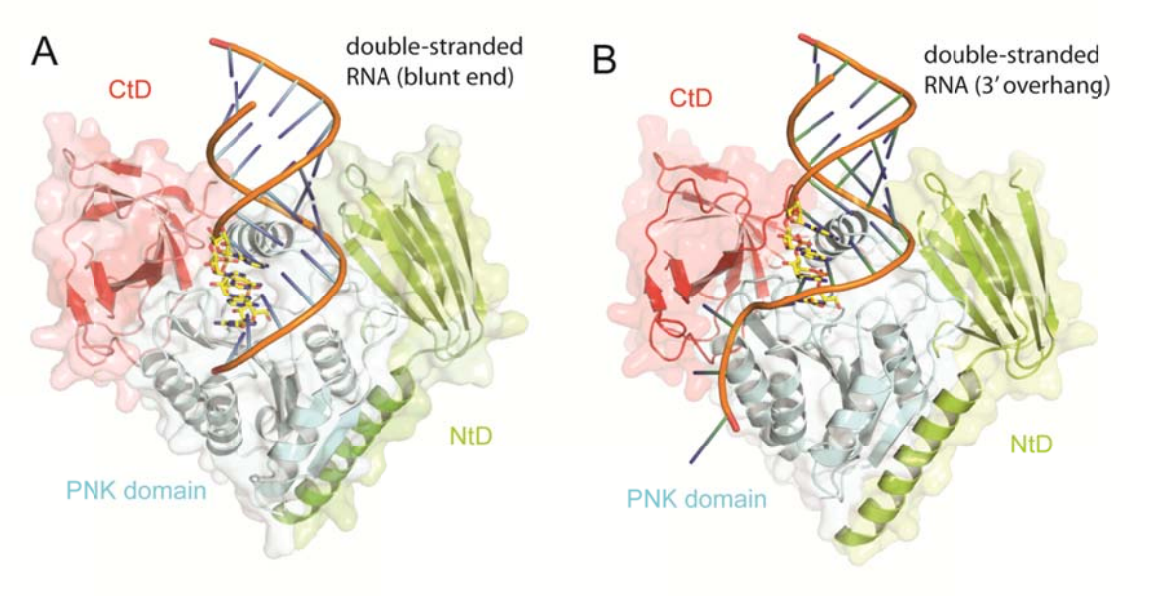


Figure 3.1.6 Structural features of RNA specificity and RNA recognition. A, Ribbon-stick representation of the RNA-binding site bound to an RNA oligonucleotide ($G_1A_2A_3A_4$). An ideal doubled stranded (blunt end) A-from RNA oligomer was superimposed onto the single stranded $G_1A_2A_3A_4$. The individual domains are color coded similar to Fig. 3.1.4. **B,** Furthermore an ideal doubled stranded (3'-overhang) A-from RNA oligomer was modeled.

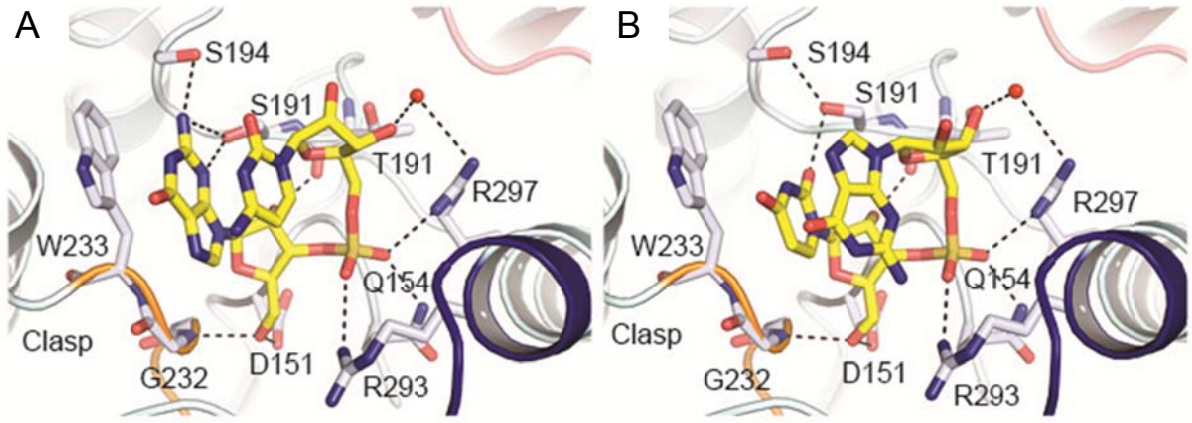


Figure 3.1.7 Comparison of structural details for RNA-binding to ceClp1. A, Ribbon-stick representation of the RNA-binding site bound to the dinucleotide G_1C_2. **B,** as well as the U_1G_2. Structural motifs important for RNA-binding highlighting in color; clasp (orange), LID module (blue), catalytic base (black). The dinucleotide (yellow), and water molecules (red) are shown in stick and sphere representation.

3.1.8 ATP binding site

The ATP binding site was characterized using the ternary complex structure of ceClp1 bound to an ATP molecule and Mn^{2+} ion. This structure was refined to an R-factor of 18.2 % and an R_{free} of 23.5 % at a resolution of 2.3 Å (Figure 3.1.8). Electron density for the ATP ligand was unambiguously identified within the active site region. In contrast to the RNA-binding site, the ATP molecule binds to the conserved P-loop motif (Figure 3.1.8) that is formed at the interface between NtD and PNK domain. Interestingly, ceClp1 showed substantial differences in the ATP-binding site when compared to all other previously described PNKs[51,52]. Whereas, in all other PNKs[42,51,52] the P-loop is part of a solvent-exposed, open-ended channel, the ATP binding site of ceClp1 is shielded by NtD. More precisely, the ATP binding pocket is not completely obstructed by NtD. Instead, a channel is created at the interdomain boundary of NtD and the PNK (Fig. 3.1.8; ATP/ADP exchange channel). This cleft might enable ATP/ADP exchange during catalysis. The triphosphate moiety of the ATP molecule is bound in a narrow channel that is traversing the core of the PNK domain. The wall of this mainly positively charged channel is formed by residues of the classical P-loop motif, the LID module that contains conserved arginine residues (Arg288 and Arg293) and a catalytic divalent metal cofactor that is octahedrally coordinated. Although the ATP-Mn^{2+} structure showed a distorted octahedral coordination sphere of the divalent metal cofactor, it was complete in crystal structures of ceClp1 in complex with an Mg^{2+} ion. The octahedral coordination sphere of the Mg^{2+} ion is formed by the hydroxyl group of Thr128, which is the only direct interaction with protein and metal cofactor. The remaining sites of the octahedron were occupied by two non-bridging β- and γ-phosphate oxygen atoms and three water molecules (W1, W2, and W3). The metal-coordinating water molecules are held in place by hydrogen bonds between W1 and the oxygen atom of side-chain Asn229, and W2, which forms a hydrogen bond to the oxygen atom of Asp151 (Figure 3.1.9).

The nucleobase and ribose of the ATP molecule are bound in the domain interface of NtD and PNK domain. Within this pocket, N6 of the adenine ring forms a hydrogen bond to the side-chain oxygen atom of Glu16. Furthermore, the nucleobase is at a perfect π-stacking orientation with the side chain of a phenylalanine (Phe39), which is part of NtD (Figure 3.1.8). The 3'-hydroxyl group of the ribose moiety forms a hydrogen bond to the main-chain carbonyl of Arg56. The tri-phosphate moiety of the ATP molecule is held in place by hydrogen bonds of the β- and γ- phosphate with the main chain amides of the P-loop motif as

well as coordination-bonds to the metal cofactor. The α-phosphate group is only bound by the main chain amide of the P-loop.

Interestingly, the Walker A lysine of the P-loop (Lys127) shows neither interaction with both ATP nor with AppNHp (Figure Figure 3.1.8 and 3.1.10). Instead, Lys127 is captured outside the active site and forms hydrogen bonds between its ε amino group and the mainchain of Gly121 and Thr230, residues which are part of the P-loop and the Clasp. This is in contrast to the prevailing structural and biochemical dogma, which suggests that the Walker A lysine is involved in the coordination of β- and γ-phosphate groups of nucleoside triphosphate[113,114]. In contrast to this non-canonical conformation of the "arrested" Lys127, the two arginine residues (Arg288 and Arg293) of the LID module interact in a classical manner with β- and γ-phosphate groups. These residues together with the divalent metal cofactor provide a positively charged environment that compensates the negatively charged phosphoryl groups. The opposite side of the active site channel is closed by the 5'-hydroxyl group of the bound RNA substrate.

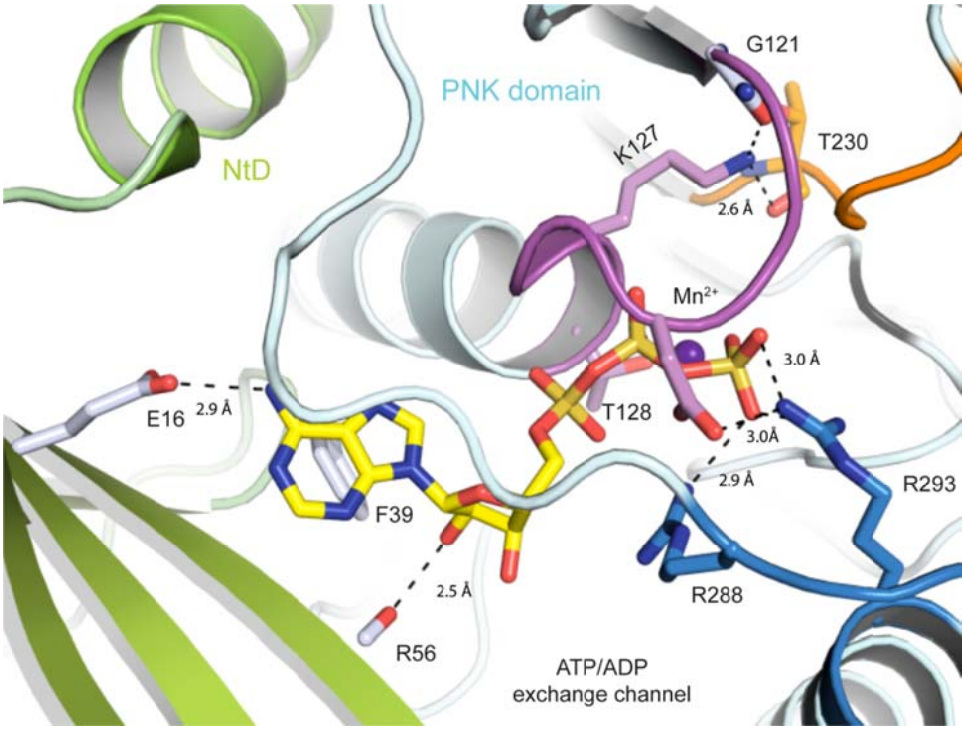


Figure 3.1.8 Characterization of the ATP binding site of ceClp1. A ribbon representation of ceClp1 with the N-terminal domain (NtD), polynucleotide kinase domain (PNK domain) and the C-terminal domain (CtD). Structural motifs important for ligand-binding or catalysis as highlighting in color (purple) clasp (orange), and LID module (blue). The ATP ligand (yellow, and the Mn^{2+} ion (purple) are shown as stick and sphere representation. Residues important for substrate binding and catalysis are shown as stick representation. The adenosine moiety of the ATP molecule interacts with the NtD, whereas the triphosphate group is bound to the P-loop motif within the active site. Furthermore, the ATP binding site is obstructed by the NtD, which forms a channel between the inter-domain boundary of the PNK domain and the NtD leading over to the triphosphate binding site.

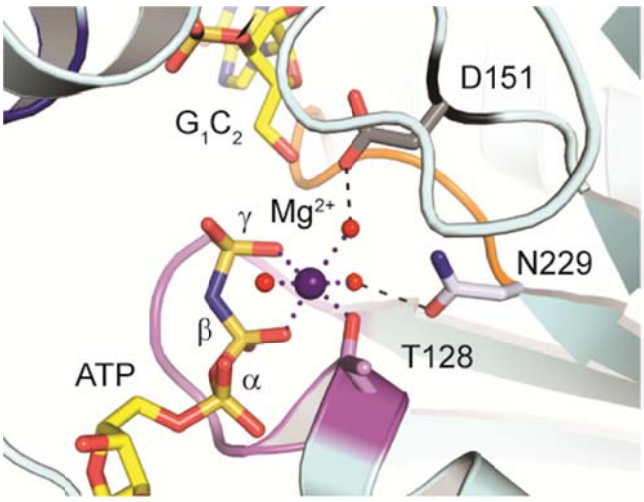


Figure 3.1.9. Octahedral coordination of the Mg^{2+} ion. Ribbon representation of ceClp1 with bound Mg^{2+}. Structural motifs important for ligand-binding or catalysis as highlighting in color; P-loop (purple) clasp (orange), LID module (blue), catalytic base (black). The ATP, G_1C_2 (yellow), the Mg^{2+} ion (purple), and water molecules (red) are shown in stick and sphere representation. Residues important for the coordination of the divalent metal cofactor are shown as stick representation..

3.1.9 Phosphoryl transfer reaction mechanism of ceClp1

Although extensive studies are available on the phosphoryl transfer reaction of mononucleotide kinases[118], relatively little progress has been made towards the structural understanding of the enzymatic mechanism of PNK. This thesis provides catalytically important crystal structures of the reaction trajectory of ceClp1. The crystal structures represent an inhibited substrate bound state, a transition state analog and a product bound state with released RNA.

3.1.9.1 Inhibited substrate bound state

Based on the structure model of the inhibited substrate bound state it was possible to describe the RNA-binding site and the binding of AppNHp to the P-loop motif (Figure 3.1.10). Presumably, the γ-phosphate group of the ATP molecule is oriented in-line with respect to the second substrate (G_1C_2), creating the correct geometry to enable the transfer of the γ-phosphoryl group to the 5'-hydroxyl group. Both the structural data and site directed mutagenesis studies suggested Asp151 to be the catalytic general base that activates the 5'-hydroxyl group. During enzyme catalysis, Asp151 initiates the reaction by deprotonation of the 5'-hydroxyl group of the ultimate base. As mentioned before, the lysine residue points

outwards from the active site prior to the phosphoryl transfer reaction. The LID arginines are involed in ATP-binding by interactions with their guanidinuim group and the oxygen atoms of the phosphate β- and γ-phosphate groups.

3.1.9.2 Transition state mimic with aluminum tetra fluoride

Our current understanding of phosphoryl transfer reactions is strongly supported by structural information on transition state analogs. It has been shown that aluminum, beryllium fluoride, and vanadate can be used as transition state analogs of the phosphoryl transfer reaction[122,123]. These analogs of a transferred phosphate group resemble transition state geometry and charge distribution[121]. This thesis provides the first crystal structure of a transition state analog obtained from a eukaryotic PNK (Figure 3.1.11). The crystal structure of a ternary complex (ADP-AlF_4^--Mg^{2+}-G_1C_2) was refined to an R-factor of 22 % and an R_{free} of 28 % at a resolution of 2.1 Å. Although bound AlF_4^- does not represent the trigonal bipyramidal transition state geometry, it is a good approximation and thus broadly accepted to represent a mimic of the transition state[121]. The octahedrally-coordinated aluminum atom interacts with four equatorial fluorine atoms, while the two axial positions are occupied by an oxygen atom of ADP's β-phosphate and an oxygen atom of the attacking nucleophile, respectively. The aluminum atom is positioned between the donor oxygen atom of ADP and the acceptor oxygen atom of the polynucleotide with a distance of 2.1 Å and 2.2 Å, respectively. This apical coordination of donor and acceptor oxygen atoms suggests an in-line associative reaction mechanism similar to that for UMP kinases[118]. Negative charges of the penta coordinated transition state seem to be neutralized by electrostatic interactions with positively charged side chains of Lys127, Arg288, Arg293 and the Mg^{2+} ion. The observation of a charge-neutralized transition state is consistent with an associative reaction mechanism. Moreover, RNA and ADP move towards each other under the concerted actions of Lys127, Arg288, and Arg293. Importantly, the Walker A lysine no longer interacts with Gly121 and Thr230 but instead becomes activated. After the transfer of the phosphoryl group, RNA is released from the active site, which leads to the product-bound state with bound ADP.

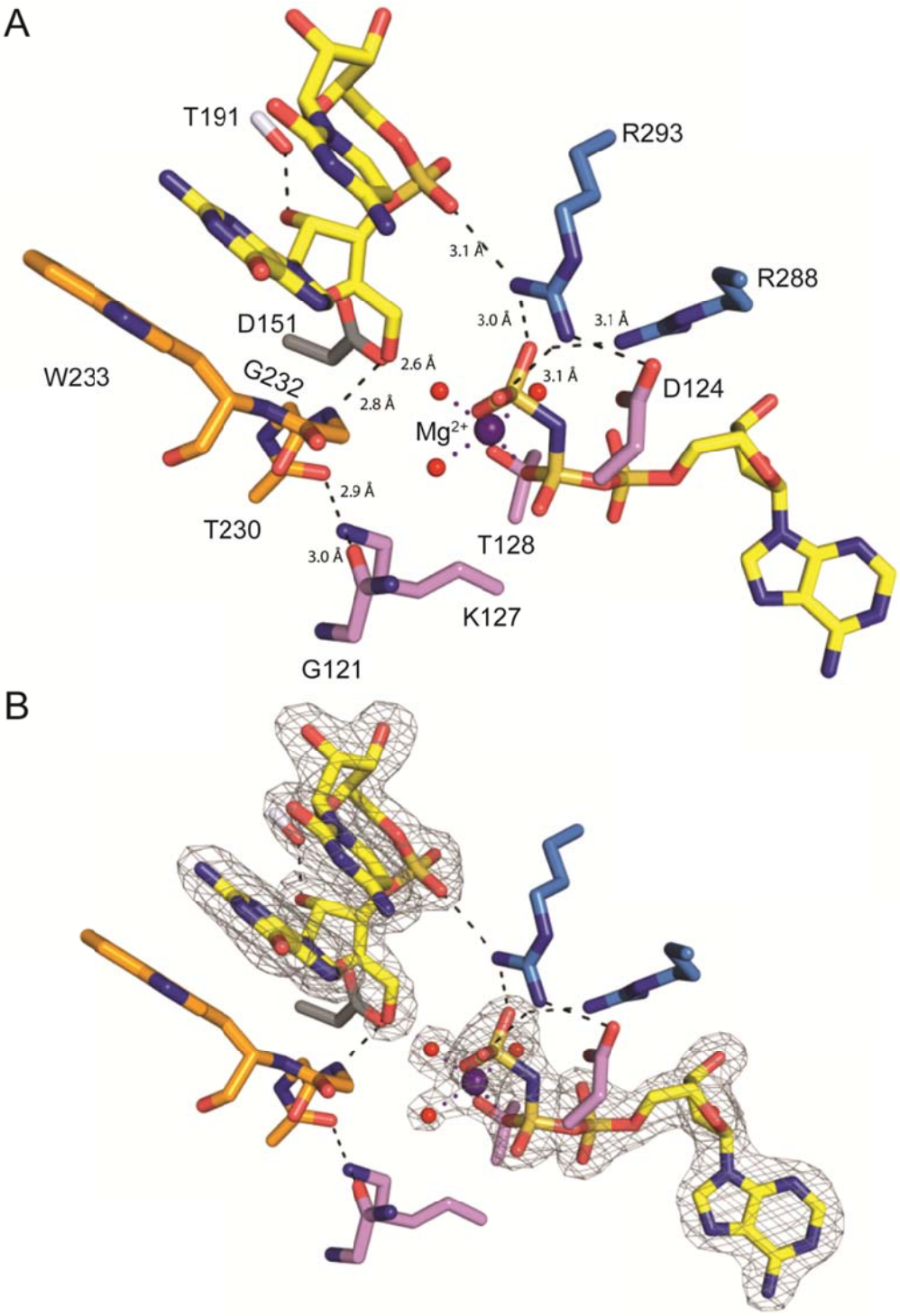


Figure 3.1.10. Inhibited substrate bound state. A, stick representation of a close-up view from the active site of ceClp1 crystallized in the inhibited substrate bound state AppNH**p**-Mg^{2+}-G_1C_2, prior to the transition state complex, the Walker A lysine (Lys127) is "arrested" outside the active site. **B,** Active site electron density of ATP-Mg^{2+}-G_1C_2. **T**he grey mesh depicts Fo - Fc difference density (contoured at 3.0 σ) calculated prior to inclusion of **t**he nucleotides or a water-metal complex in the model. The nucleotides are depicted as stick models. The orientation is the same as **i**n **A**. The color code of structural motifs important for ligand-binding or catalysis is similar to Figure. 3.1.4.

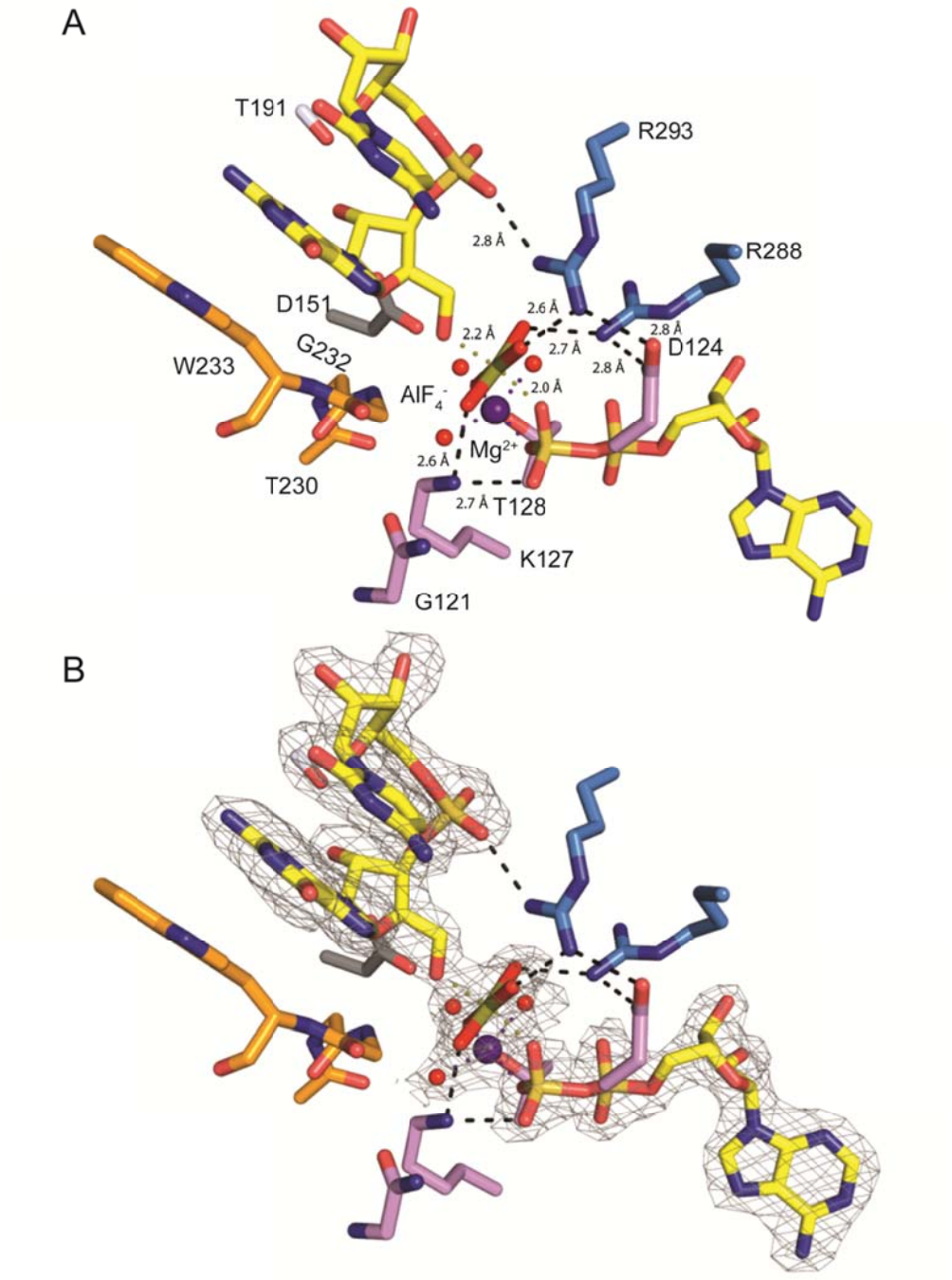


Figure 3.1.11 Transition state analog. stick representation of a close-up view from the active site of ceClp1 crystallized with the transition state analog ADP-AlF_4^--Mg^{2+}-G_1C_2. The Lys127 becomes activated "switched-on" in the transition state. **B,** Active site electron density of ADP-AlF_4^--Mg^{2+}-G_1C_2. The grey mesh depicts Fo - Fc difference density (contoured at 3.0 σ) calculated prior to the inclusion of the nucleotides or a metal complex in the model. The nucleotides are depicted as stick models. The orientation is the same as in **A**. The color code of structural motifs important for ligand-binding or catalysis is similar to Figure. 3.1.4.

3.1.9.3 Product bound state with released RNA

Since ceClp1 is active in the crystalline state, it was possible to obtain a product bound state by soaking crystals of ceClp1 with its substrates. After incubation of ceClp1 crystals with ATP and G_1C_2, crystal structures revealed electron density for an ADP molecule, while phosphorylated G_1C_2 was not present. This crystal structure represents the RNA released product bound state and was refined to an R-factor of 19 % and an R_{free} of 24 % at a resolution of 2.0 Å (Figure 3.1.12). It is likely that the lack of electron density for phosphorylated RNA results from electrostatic repulsion between β-phosphate of ADP and 5'-phosphorylated RNA. Interestingly, this crystal structure is characterized by a "switched off" Lys127 that points out of the active site, similar to the substrate-bound state. Lys127 switches into an "arrested" state prior to and after the transition state. Furthermore, the latch arginines, Arg288 and Arg293, are reoriented. In a latch-catch mechanism, Arg293 of the LID module forms a salt bridge with Asp124 of the P-loop motif. In conclusion, ceClp1 seems to use a molecular gating mechanism characterized by a reorganization of the active site such that Lys127, Arg288, and Arg293 stabilize negative charges of the transferred phosphoryl group in the transition state, while preventing ATP hydrolysis in the absence of an RNA substrate.

Interestingly, the arginines of the LID module seem to function in a latch-catch mechanism (Figure 3.1.11). These arginine latches are completely activated only in the transition state and form, together with Lys127, an environment that enables correct positioning and charge stabilization of the transition state (Figure 3.1.11).

3.1.10 Summary of the structural results

This thesis provides the first crystal structure of a eukaryotic PNK with a bound RNA oligonucleotide. Furthermore, high-resolution structures of enzymatically relevant states of ceClp1 were determined (AppNHp-Mg^{2+}-G_1C_2, ADP-Mg^{2+}-AlF_4^--G_1C_2, and ADP-Mg^{2+}). Clp1's RNA-specificity and RNA recognition are described based on these structural data. Additionally, it was possible to elucidate the phosphoryl transfer reaction mechanism. Interestingly, crystal structures showed a non-canonical Walker A lysine (Lys127) that seems to act as a molecular switch. Activation of this switch is achieved in the transition state complex while before and after the Walker A lysine is in an arrested conformation. In contrast to other nucleotide kinases, Clp1 seems to use a substrate-gating mechanism.

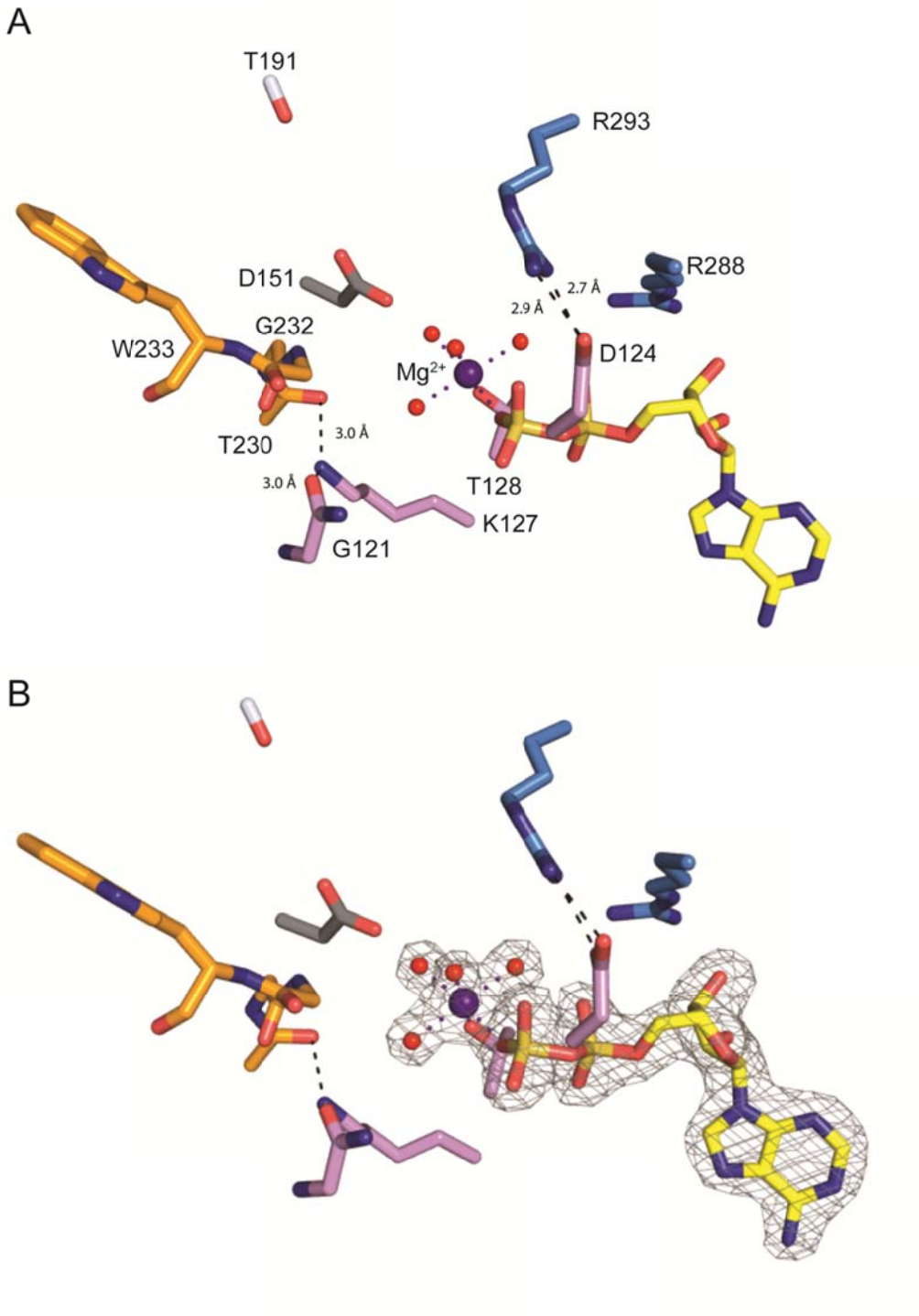


Figure 3.1.12 RNA released product bound state. stick representation of a close-up view from the active site of ceClp1 crystallized in the RNA released product bound state ADP-Mg^{2+}. Prior to and after the transition state complex, the Walker A lysine (Lys127) is "arrested" outside the active site. **B,** Active site electron density of ADP-Mg^{2+}. The grey mesh depicts Fo - Fc difference density (contoured at 3.0 σ) calculated prior to the inclusion of the nucleotides or a metal complex in the model. The nucleotides are depicted as stick models. The orientation is the same as in **A**. The color code of structural motifs important for ligand-binding or catalysis is similar to Figure. 3.1.4.

3.2 Biochemical characterization of Clp1 from *C. elegans*

This section describes the biochemical characterization of ceClp1, which was shown to be a *bona fide* RPNK enzyme similar to previously described human orthologs[88]. In addition to high sequence conservation between these orthologs, ceClp1 and hsClp1 have identical RNA and DNA substrate specificities. The enzymatic characterization of ceClp1 showed that the dinucleotide is the minimal substrate for an efficient phosphoryl transfer. Intriguingly, in case of short RNA substrates (dinucleotides), ceClp1 shows a clear preference for purine against pyrimidine bases at the ultimate position, which is in accordance with the crystal structure. Furthermore, no basal ATPase activity or product inhibition could be measured for ceClp1, which suggested a highly regulated 5'-kinase activity. Moreover, the 5'-kinase reaction is irreversible under single turnover conditions.

3.2.1 Substrate specificity of the Clp1 protein family

An RPNK activity of ceClp1 was demonstrated using an *in vitro* phosphorylation assay with different RNA and DNA substrates. A comprehensive biochemical characterization of hsClp1 revealed a strong preference for RNA substrates as the 5'-phosphate acceptor[88]. To compare the specificity of ceClp1 with hsClp1, the enzyme (2 µM) was incubated with different single stranded and double stranded oligonucleotides (0.1 mM), followed by a separation using denaturing PAGE. The control reactions with T4 PNK verified the differential electrophoretic mobility of phosphorylated oligonucleotide substrates (Fig. 3.2.1). In accordance with observations made for hsClp1, ceClp1 also discriminates single stranded RNA against single stranded DNA (Fig. 3.2.1 A). However, ceClp1 is able to phosphorylate double-strand nucleotide RNA and DNA substrates with blunt ends. But again, the efficiency is lower for double stranded DNA (Fig. 3.2.1 B). Similar results were obtained for double stranded RNA and double stranded DNA molecules with a 3'-overhang (Fig. 3.2.1 C). Interestingly, blunt ended double stranded RNA substrates showed higher phosphorylation efficiency when compared to double stranded RNA substrate with a 3'-overhang. Even though ceClp1 showed a clear preference for single stranded RNA and double stranded RNA, it also displayed a residual activity towards single stranded DNA during a longer incubation time. Complete

turnover of single stranded DNA substrates was observed after 2 hours of incubation. These results revealed conserved substrate specificity between ceClp1 and hsClp1.

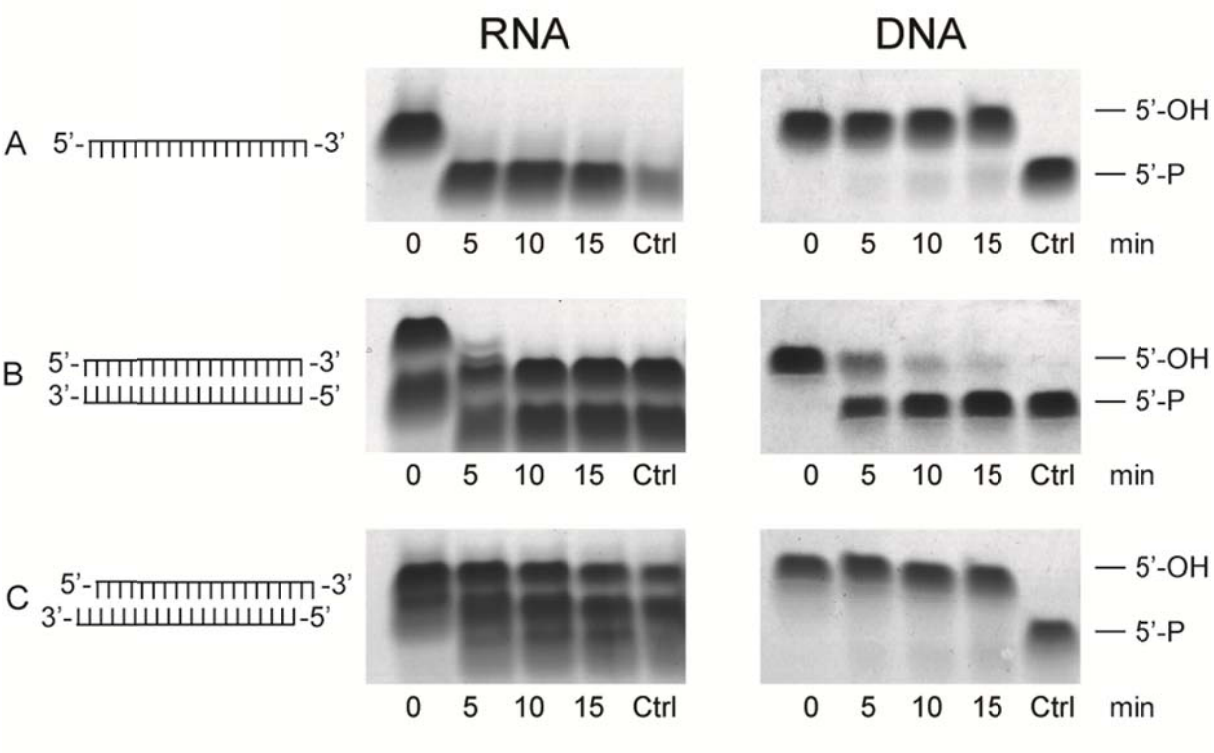


Figure 3.2.1 Substrate specificity of ceClp1 on RNA and DNA substrates. ceClp1 (1 µM) was incubated with various RNA and DNA oligonucleotide substrates (20 nt of length), at a concentration of 100 µM. Substrate molecules were either single-stranded oligonucleotide, double-stranded oligonucleotides containing a blunt end or double-stranded oligonucleotides with a 3'-overhang. T4 PNK reaction samples are shown as positive controls (Ctrl). **A,** ceClp1 has robust kinase activity upon incubation with single stranded RNA, whereas single stranded DNA was inefficiently phosphorylated. Notably, after 2 h incubation, basal phosphorylation of single stranded DNA was observed. **B,** ceClp1 displays enzymatic activity on doubled-stranded oligonucleotides with blunt ends, however, RNA-molecules are more efficiently phosphorylated. **C,** double stranded RNA oligonucleotides containing a 3'-overhang are less efficiently phosphorylated by ceClp1 and DNA molecules are not phosphorylated.

3.2.2 Minimal substrate requirements of ceClp1

The substrate specificity assay (see 3.2.1) as well as recent *in vitro* experiments by Weitzer and Martinez[88] indicated that Clp1 is able to phosphorylate a number of different RNA polynucleotides. However, information about ceClp1's requirements on the minimal substrate length of RNA is missing. To define these minimal substrate requirements, 5'-kinase activity of ceClp1 towards substrates with varying lengths was determined. Activity was analyzed using Michaelis-Menten kinetics. All measurements were performed under steady state conditions. Two dinucleotides, G_1C_2 and U_2G_2, identified in the crystal structure were used for 5'-PNK activity assays. ceClp1 was used at limiting enzyme concentration (0.2 µM). The

reaction mixture containing 1 mM ATP and 0.5 mM of either G_1C_2 or U_1G_2 and pre-incubated for 15 minutes at 20°C. After this pre-incubation step, the reaction was started by addition of ceClp1 and then analyzed using a coupled calorimetric ATPase assay (Figure 3.2.2). Although ceClp1 showed a significant ATP turnover with both dinucleotides, phosphorylation of G_1C_2 was much more efficient compared to U_1G_2 (Figure 3.2.2 A). After an initial lag phase, the ATP turnover of G_1C_2 and U_1G_2 reached its maximum velocity. The same experiment was performed with 0.5 mM of adenosine- and cytosine 3'-monophosphate (3'-AMP and 3'-CMP). Neither with 3'-AMP nor with 3'-CMP, was an ATP turnover measured. Results of these measurements, together with structural data (see 3.1.7), suggest that correct binding to the clasp motif of ceClp1 is impaired for 3'-nucleoside monophosphate substrates. These conclusions are further supported by results obtained for hsClp1 and the mammalian PNK, which also showed no 5' -kinase activity on 3' -nucleoside monophosphates[85,88]. Currently, only T4 PNK, has been shown to phosphorylate 3'-nucleoside monophosphates[46]. Therefore, a control reaction was performed with T4 PNK under identical conditions. As expected, in the presence of 3'-AMP and 3'-CMP, enzymatic activity could be detected for T4 PNK (Figure 3.2.2 B).

Michaelis-Menten kinetics of ceClp1 were measured at varying dinucleotide (G_1C_2 and U_1G_2) concentrations. The V_{max} and K_m values were obtained by fitting to a Michaelis-Menten equation (Equation 1.2). At saturating ATP concentration of 1 mM, ceClp1 (0.2 µM) showed an apparent k_{cat} of 2.5 s^{-1} and K_m of 103.0 µM for G_1C_2, and a k_{cat} of 2.3 s^{-1} and K_m of 1122.0 µM for U_1G_2 (Figure 3.2.3). These kinetic data suggest that under saturating dinucleotide concentrations, ceClp1 displays similar turnover rates for G_1C_2 and U_1G_2. However, a tenfold higher K_m value indicated a much lower substrate affinity for U_1G_2. This can be explained by the crystal structure of ceClp1 bound to the U_1G_2, which shows an imperfect base-stacking interaction mimic of the ultimate base (U_1) to Trp233. The penultimate base (G_2) is flipped into a *syn*-conformation. However, the position of the 5'-hydroxyl was in ideal hydrogen bonding distance to Gly232 and Asp151. These structural details are reflected in the increased K_m value and identical turnover rates at saturation. Interestingly, experiments with longer nucleotide substrate revealed similar Michaelis-Menten constants for G_1C_2. Michaelis-Menten values of a penta nucleotide $G_1A_2A_3A_4A_5$ determined under identical conditions as in this thesis exhibited a k_{cat} = 2.2 s^{-1} and K_m = 99.0 µM (Bernard Loll). As suggested by the protein-RNA interaction profile seen in the crystal structure of ceClp1 (Figure 3.1.5), substrate affinities for G_1C_2 and $G_1A_2A_3A_4A_5$ are almost identical.

Although ceClp1 showed a preference for purine bases at the ultimate position of dinucleotide substrates, it is questionable whether this is the case with oligonucleotide substrates as well. It is more likely that for longer substrates, all bases are in *anti*-conformation with continuous base-stacking, which may stabilize the imperfect interaction between a pyrimidine base and Trp233 at the ultimate position. To test whether ceClp1 discriminates purine and pyrimidine bases even for longer substrate, an experiment was performed comparing two complementary oligonucleotides (20mer) that differ at their ultimate and penultimate positions (G_1C_2-18mer and C_1C_2-18mer). Since it was shown that ceClp1 has similar turnover rates for longer oligonucleotide (18mer) compared to G_1C_2 (Bernard Loll), RNA concentrations (0.1 mM) were chosen in which the enzyme is under unsaturated RNA conditions. These measurements indicated that both single stranded RNA substrates are phosphorylated with the same efficiency (Figure 3.2.5).

In conclusion, ceClp1 is unable to phosphorylate 3'-nucleoside monophosphates, whereas RNA dinucleotides represent a suitable substrate for 5'-kinase activity. Furthermore, ceClp1 displayed no further preference with respect to length of the oligonucleotide. The dinucleotide G_1C_2 was identified as the minimal substrate requirements of ceClp1. Based on the structural data and the biochemical experiments, ceClp1 seems to have no sequence specificities for longer RNA substrates.

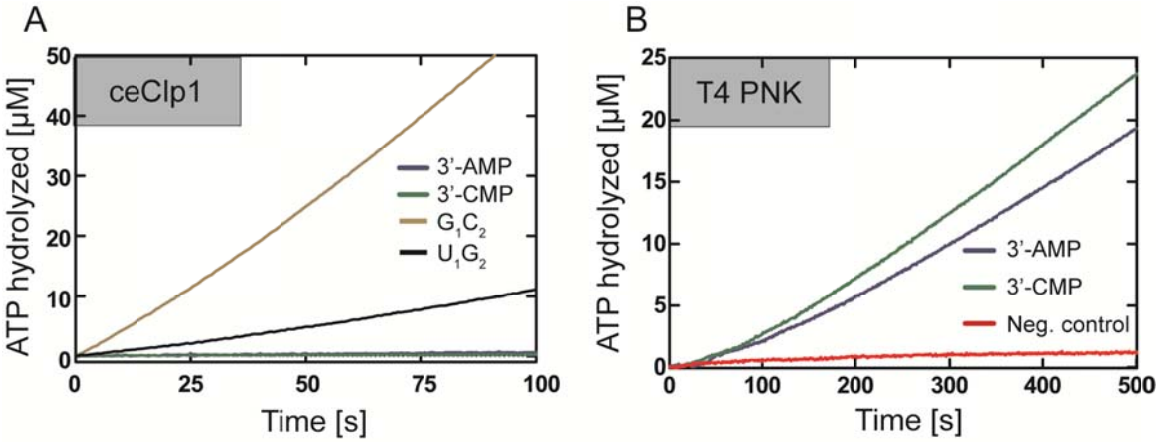


Figure 3.2.2 Nucleotide sequence length dependence of the 5'-kinase activity. The kinase activity of ceClp1 and T4 PNK was monitored spectroscopically based on a continuous coupled-ATPase activity assay. **A,** ceClp1 (0.2 µM) was incubated with 1 mM ATP, and 0.5 mM of 3'-AMP, 3'-CMP G_1C_2, and.U_1G_2. Once a dinucleotide was used as a substrate, ceClp1 showed a significant turnover of ATP, in contrast to mononucleotides. The dinucleotide appears to be ceClp1's minimal substrate requirement. **B,** A control experiment was conducted with the T4 PNK (5 U) supplemented with 1 mM ATP and 0.5 mM of two different mononucleotide substrates (3'-AMP and 3'-CMP). T4 PNK shows clear kinase activity towards both substrates, whereas no basal activity was detectable without acceptor.

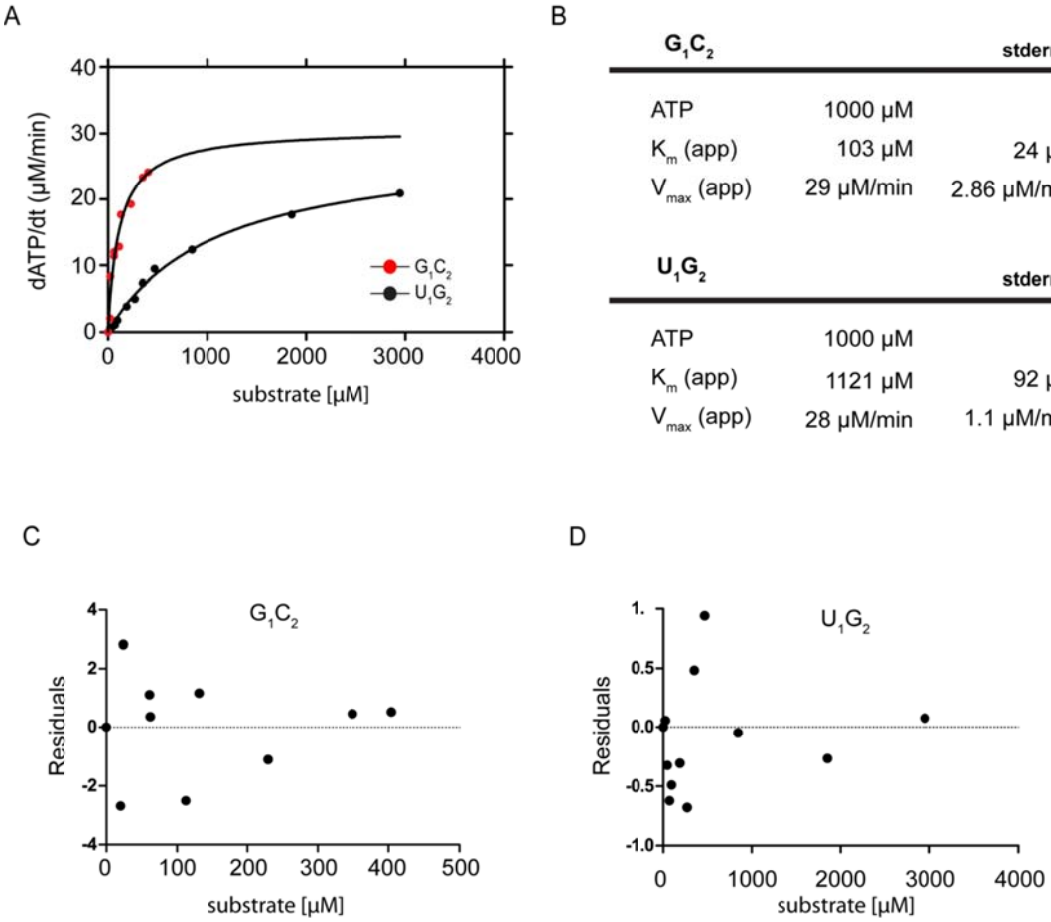


Figure 3.2.3 Minimal substrate requirements of ceClp1. A, The Michaelis-Menten kinetics of the initial enzyme velocities were plotted against varying dinucleotide concentrations. All measurements were performed at constant ATP concentrations (1 mM). The black line depicts the fit of the data using the standard Michaelis-Menten equation. **B,** Calculated constants resulting from the fit with the standard Michaelis-Menten model. **C,** and **D,** Residuals of the Michaelis-Menten plot.

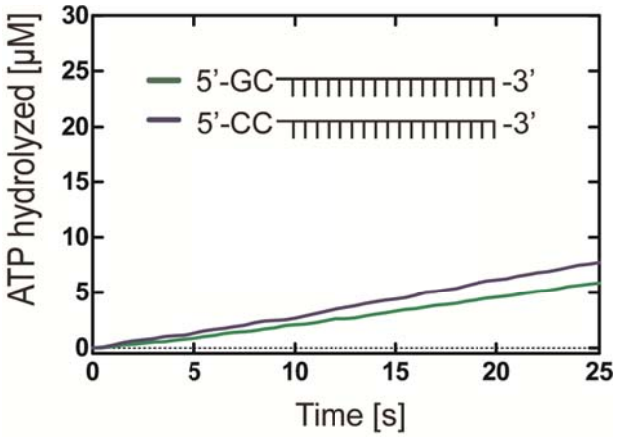


Figure 3.2.4 Sequence specificity of ceClp1 towards oligonucleotides. Steady state kinetics of ceClp1 measured with two different single stranded RNA oligonucleotides. Both oligonucleotides are complementary to each other.

3.2.3 Characterization of ceClp1's 5'-kinase activity

Using an enzymatically coupled spectroscopic ATPase assay, it was possible to characterize the substrate specificity of ceClp1. However, these experiments have the drawback that the quantification of enzymatic activity only relies on conversion of ATP to ADP. Therefore, a second experimental setup for an *in vitro* phosphorylation assay was used to support kinetic data obtained with the coupled-enzyme assay (Figure 3.2.5). The second assay is based on a MonoQ anion exchange chromatography protocol designed for binding and baseline separation of ceClp1's 5'-kinase reaction substrates from products.

Steady-state kinetics were measured under limiting enzyme concentrations (0.2 µM) in a reaction mixture containing saturating ATP (1 mM) and 0.5 mM G_1C_2 concentrations. For the second assay, ceClp1 (2µM) was incubated with ATP and G_1C_2, each in a concentration of 250 µM, at 25 °C for 2 hours. Due to a change in ionic net charge after the phosphoryl transfer reaction, substrates and products can be separated by anion exchange chromatography. To assess possible basal ATPase activity of ceClp1, both assays were performed in absence of G_1C_2, which showed no basal ATPase activity (Figure 3.2.5 A and B). However, after the addition of RNA a significant ATP turnover and an accumulation of the reaction products ADP and 5'P-G_1C_2 were detected. Using the phosphorylation assay, it was also tested if the net charge of the RNA is modified in an ATP-independent manner, which was not the case (Figure 3.2.6 C). Furthermore, it was possible to show that the kinase reaction proceeds in a one-to-one stoichiometry. Therefore, ceClp1 was incubated with ATP (500 µM) in two-fold excess compared to the G_1C_2 concentration. Observed ATP turnover was proportional to the conversion of G_1C_2 to 5'P-G_1C_2. ceClp1 shows neither unstimulated basal ATPase activity nor RNA-stimulated basal ATPase activity (Figure 3.2.5 D).

In addition, ceClp1 was tested for product inhibition using the coupled calorimetric ATPase assay at varying G_1C_2 concentrations. ATP turnover rates were followed by the decrease in absorbance at 340 nm due to oxidation of NADH (Figure 3.2.6). The reaction rate of each measurement obtained under varying G_1C_2 concentration was plotted and analyzed. Maximum velocity of the reaction was reached after an initial lag phase. Although a slight decrease of ATP turnover rates at the end of the reaction was measured, it is possible that this decrease is due to dilution effects. Consistent with this hypothesis the weaker decrease of ATP turnover rates at higher G_1C_2 concentration. Thus, ceClp1 seems to show no product inhibition by accumulation of 5'P-G_1C_2.

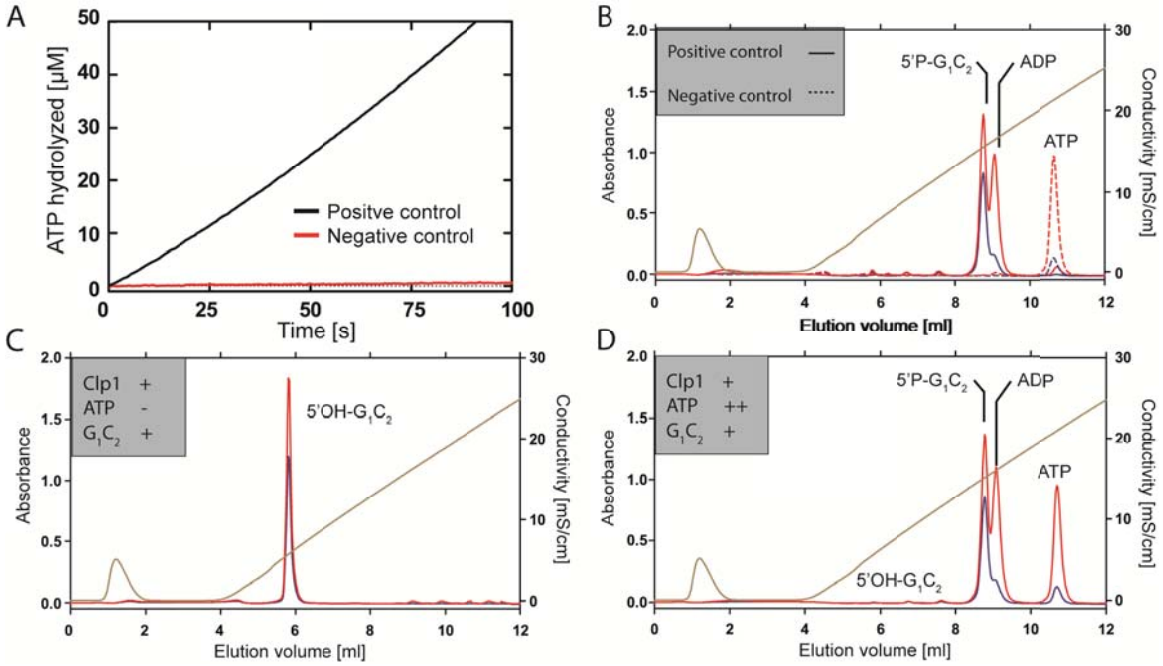


Figure 3.2.5 Biochemical characterization of ceClp1. A possible basal ATPase activity was tested by a spectroscopically based coupled-ATPase activity assay as well as a phosphorylation assay. **A,** ceClp1 (0.2 µM) incubated with 1 mM ATP, and with or without 0.5 mM of G_1C_2, showing no basal activity. **B,** 5'-subtrate phosphorylation was analyzed on a MonoQ anion exchange column. Elution of the nucleotide was achieved with an increasing linear NaCl gradient that determines the conductivity (brown). Depending on the numbers of phosphoryl groups, nucleotides bind with a different affinity to the column matrix (RNA-5'OH < RNA-5'P < ADP < ATP). The absorbance at 260 nm (red) and 280 nm (blue) is indicated in the plots. Note that the RNA-dinucleotide (G_1C_2) can be distinguished from adenine molecules (ATP/ADP) based on their different A_{260}/A_{280} ratio of ~2 and ~7, respectively. For the negative control, ceClp1 was incubated with 250 µM of ATP but without G_1C_2, showing no conversion of ATP to ADP. **C,** An ATP-indepented modification of the RNA was excluded by a phosphorylation assay without ATP, showing no net charge change of G_1C_2. **D,** The reaction was also conducted with an excess of 500µM of ATP. The observed turnover of ATP proceeds until the entire RNA substrate is consumed.

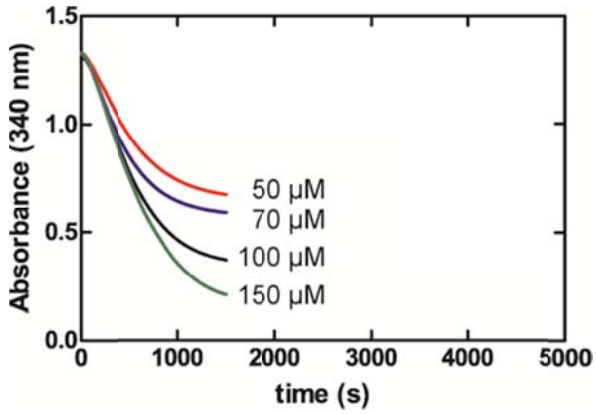


Figure 3.2.6 Characterization of production inhibition of ceClp1. Based on a spectroscopically enzyme-coupled ATPase activity assay, data were recorded to test for product inhibition. ceClp1 (0.2 µM) incubated at saturated ATP concentrations (1 mM) under varying G_1C_2 concentrations (50, 75, 100, 150 µM).

3.2.4 Reverse reaction of ceClp1

The exact function of Clp1's RNA 5'-kinase activity has remained elusive until now. However, the identification of a potential reversal of the phosphorylation reaction would give new insights into RNA maturation pathways. Experiments on the human ortholog suggested a quasi-irreversible 5'-kinase reaction[85]. This is in contrast to measurements with thermostable archaeal Clp1 homolog[95], which exhibited a reversible reaction. Furthermore, a reversal of the kinase reaction is also demonstrated for other PNKs like T4 PNK[156,157]. ceClp1 was tested for the putative reverse reaction under single-turnover condition. An ideal single-turnover experiment is characterized by the fact that every substrate molecule is initially bound by a single enzyme. If ceClp1 indeed is a quasi-unidirectional kinase, no equilibrium between forward and reverse reaction can be reached. Since the reaction proceeds into the direction of the energetically favorable forward reaction, it is impossible to detect the conversion of ADP to ATP. The reverse reaction was analyzed by anion exchange chromatography, based on a purification protocol that enables baseline separation of ADP and ATP (Figure 3.2.8). ceClp1 (60 µM) was incubated with a phosphorylated dinucleotide (5'P-G_1C_2) and ADP, each at a concentration of 50 µM. After an incubation time of 20-40 minutes at 20 °C, the reaction was quenched and conversion of ADP to ATP was measured. The two substrates ADP and 5'P-G_1C_2 were eluting from the anion exchange column as a single peak, but could be baseline separated from the potential products (ATP and G_1C_2). The reverse reaction experiment revealed that ceClp1 is functioning as a quasi-irreversible polynucleotide kinase, showing no

formation of ATP or G_1C_2 after 20 or even 40 minutes of incubation (Figure 3.2.8 A and B). A control experiment under identical condition was performed in the absence of the phosphate-donor (5'P-G_1C_2) to exclude any basal adenylate kinase activity (Figure 3.2.8 C). If a potential adenylate kinase contamination was present, a conversion of ADP to ATP and AMP would have been detected. However, no formation of ATP or AMP was measured after 20 minutes incubation time. Furthermore, to exclude that high protein concentration in the single turnover experiment affects Clp1's enzymatic activity, the enzyme was also tested in its forward direction. Therefore, ceClp1 (60 µM) was incubated with ATP and G_1C_2 each at a concentration of 50 µM. The analyzed reaction mixture revealed a complete conversion of ATP and G_1C_2 to ADP and 5'P-G_1C_2 (Figure 3.2.8 D) after an incubation time of 20 minutes. This result implies that, again, no equilibrium between forward and reverse reaction is reached and that the 5'-kinase reaction is the energetically favorable direction.

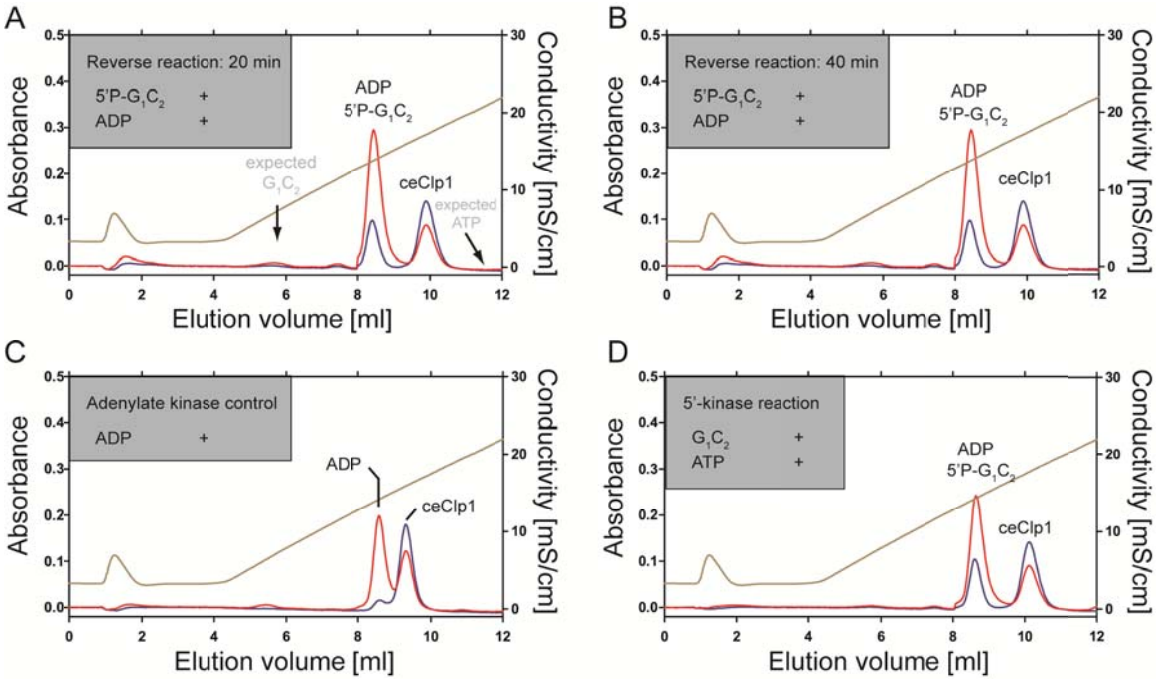


Figure 3.2.8 Reversal of the polynucleotide kinase reaction. A, and B, reaction mixtures (200 µl) containing 60 µM of ADP, 50 µM of 5'P-G_1C_2, and 50 µM of ceClp1 were incubated for 20 and 40 min, respectively. The products were analyzed on a MonoQ anion exchange column. Elution of the nucleotide was achieved with an increasing linear NaCl gradient similar to the description in figure 3.2.1. Even after 40 min, ceClp1 showed no reverse reaction which would lead to production of G_1C_2 and ATP in a single turnover experiment. **C,** to exclude any basal adenylate kinase activity, ceClp1 was incubated only with 50 µM of ADP for 40 min. Consistently, no accumulation of ATP and AMP were detectable. **D,** shows the forward reaction in a single turnover experiment. The reaction mixture (200 µl) containing 50 µM of ATP and 50 µM of G_1C_2 was incubated for 20 min. In contrast to the reverse reaction, the PNK forward reaction led to the accumulation of ADP and 5'P-G_1C_2.

3.2.5 Summary of the enzymology studies

Altogether, the biochemical characterization demonstrates that ceClp1 is an RPNK with identical substrate specificity as hsClp1. Although ceClp1 is able to phosphorylate RNA and DNA polynucleotides, it shows a much higher affinity for RNA substrates. Furthermore, steady-state kinetics revealed that ceClp1 requires at least a dinucleotide for an efficient phosphoryl transfer reaction. Intriguingly, the dinucleotide G_1C_2 represents the minimal substrate for ceClp1, showing similar Michaelis-Menten kinetics as measured for an oligonucleotide substrate. However, in case of the dinucleotide substrate, ceClp1 displays a preference for purine against pyrimidine bases at the ultimate position as indicated by the tenfold increased K_m of U_1G_2 compared to G_1C_2. Interestingly, the characterization of ceClp1 reveals no basal ATPase activity, and moreover, ceClp1 seems to show no product inhibition. In contrast to other PNKs[42,51,52], ceClp1 functions as a quasi-unidirectional enzyme with no reversal of the 5'-kinase reaction.

3.3 Mutagenesis studies on ceClp1

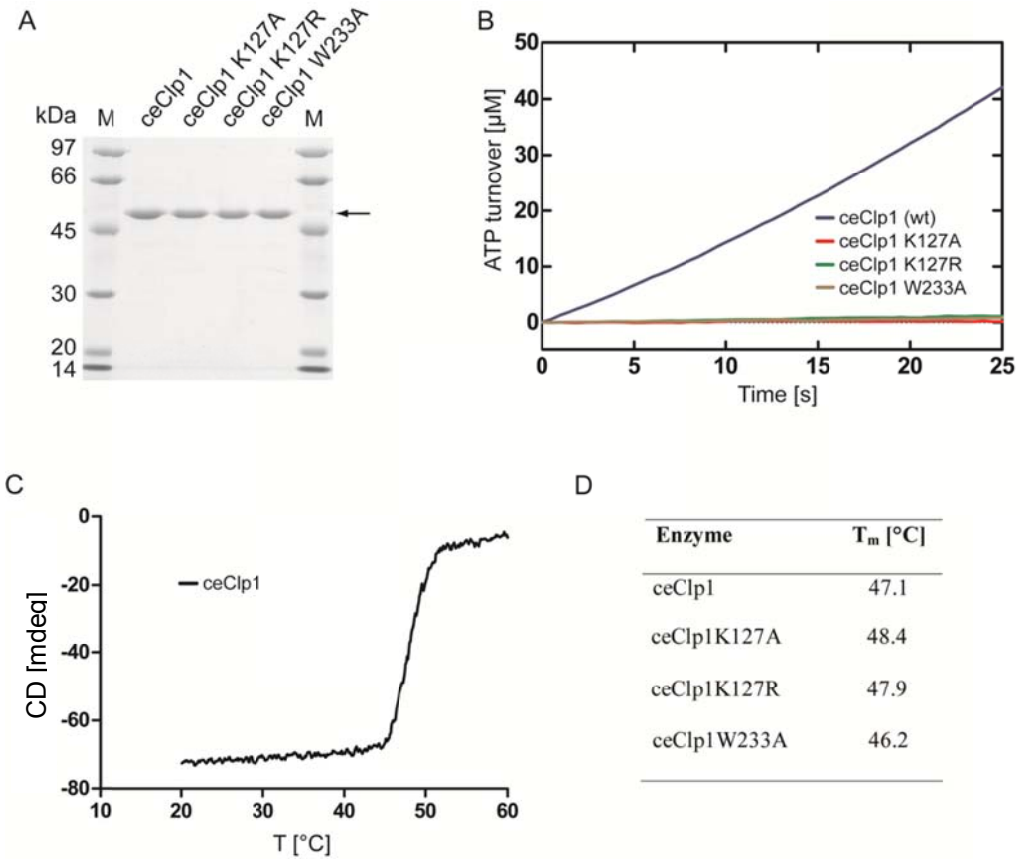
3.3.1 Site directed mutagenesis of Lys127 and Trp233

Previous mutagenesis studies in our lab allowed the identification of a number of catalytically important residues for ceClp1 enzyme activity (Table 3.3.1). Conservative and non-conservative mutations were introduced based on the sequence alignment of Clp1 from higher eukaryotes (Figure 3.1.1). As expected, hot spots for mutation sites were found in highly conserved regions like the P-loop motif, the Walker B motif, the LID module and the clasp. This characterization focused in particular on conserved residues of the PNK domain (Pro122, Thr123, Asp124, Asp151, Glu154, Asp235, Arg288, and Arg293) to show their potential involvement in enzyme catalysis. In addition, in this thesis structure-guided site-directed mutagenesis and biochemical methods were employed to supplement the previous mutational characterization of ceClp1 (Table 3.3.1). The wild-type protein and mutated protein-variants (K127A, K127R, and W233A) were expressed and purified as described for crystallization experiments (see 3.1.2, Figure 3.3.1 A). Subsequently, purified protein variants were characterized for their 5'-kinase activity using an enzymatically coupled spectroscopic ATPase assay (Figure 3.3.1 B). Steady state kinetics were measured under limiting protein concentration (2 µM) with excess of ATP (1 mM) and G_1C_2 (500 µM). To exclude any

destabilizing effects by the mutation, the thermodynamic stability of all mutant variants was verified by determining the apparent melting point (Tm) using circular dichroism spectroscopy. All mutant variants revealed no significant differences compared to the wild type (Figure 3.3.1 C and D).

Since the Walker A lysine (Lys127) of the P-loop showed a non-canonical "switched-off" conformation prior to and after the transition state, this residue was considered to be part of a molecular gating mechanism. Therefore, the structure-function relation was examined by a conservative substitution. Lys127 was replaced with an arginine residue and a non-conservative alanine substitution. Based on ATPase activity, Lys127 proved to be essential for enzyme catalysis since K127R was abolished in enzymatic activity. It is possible that K127R mutation might interfere with the bifurcated binding of the β- and γ- phosphates groups by Lys127 at the transition state structure (Figure 3.1.10). Non-conservative mutations of the Walker A lysine have previously been shown to abolish enzyme activity of P-loop kinases and therefore serve as a negative control[88]. As predicted, the 5'-kinase activity of the K127A variant was abolished (Figure 3.3.1 B).

The crystal structure of ceClp1 bound to an RNA substrate (Figure 3.1.5) showed a base stacking interaction mimic between the indol group of Trp233 and the ultimate base of the substrate. Alanine substitution of Trp233 (W233A) also abolished the 5′-kinase activity of ceClp1. Therefore, the π-stacking interaction between Trp233 and the ultimate nucleobase seems to be critical for RNA-binding of ceClp1.



Enzyme	T_m [°C]
ceClp1	47.1
ceClp1K127A	48.4
ceClp1K127R	47.9
ceClp1W233A	46.2

Figure 3.3.1 Site directed mutagenesis of key residues of ceClp1 involved in catalysis and RNA-binding interaction. A, SDS-PAGE of ceClp1 mutant variants. B, Kinase activity of wild type Clp1 and different mutant variants monitored spectroscopically based on a coupled ATPase activity assay. The experiment was conducted with 2 µM protein, 1 mM ATP, and 500 µM G_1C_2. Wild type (wt) protein shows 5'-kinase activity. The mutated variants K127R and W233A showed no enzymatic activity. A Walker A lysine mutant (K127A) serves as a control for an inactive enzyme. **C,** Thermal melting curve exemplified for native ceClp1. Thermal unfolding is an irreversible process due to precipitation of the protein. **D,** Determined apparent melting temperatures (T_m) of ceClp1 wild type and variants. All protein samples show comparable global stability.

3.3.2 Deletion of the N-terminal and C-terminal domain

ceClp1 is a multi-domain protein with a central PNK domain containing the active site region (Figure 3.1.4). The central domain is flanked by an additional NtD and CtD, but the function of these additional domains has remained elusive so far. To address the question to what extend the NtD and CtD contribute to 5'-kinase activity, various N- and C-terminal truncation mutant variants of ceClp1 were designed and tested for enzymatic activity.

Table 3.3.1 Michaelis-Menten steady-state constants of wild type ceClp1 and mutated variants using G_1C_2 as substrate. Titration experiment was performed at a constant concentration of 1 mM ATP and 0.2 µM ceClp1.

Enzyme	K_M [µM]	k_{cat} [s^{-1}]	Function	Reference
ceClp1wt	109	2.5		This thesis
ceClp1P122A	187	1.8	A P-loop residue involved in main chain interactions with the nucleoside triphosphate.	*
ceClp1P122S	691	3.1		*
ceClp1T123A	residual activity		A P-loop residue involved in main chain interactions with the nucleoside triphosphate.	*
ceClp1T123S	185	0.9		*
ceClp1D124A	1764	1.3	A P-loop residue that interacts via its side chain with both residue Arg288 and the ATP molecule.	*
ceClp1D124Q	no activity			*
ceClp1D124N	612	0.3		*
ceClp1K127A	no activity		The Walker A lysine is usually involved in coordination of the β- and γ-phosphate groups.	This thesis
ceClp1K127R	no activity			This thesis
ceClp1D151A	no activity		The catalytic base is involved in the deprotonation of the 5'-OH group.	*
ceClp1Q154A	no activity		The side chain of this residue interacts with 5'→3' phosphate group of the ultimate base.	*
ceClp1W233A	no activity		This residue is involved in a base stacking interaction mimic with the ultimate base.	This thesis
ceClp1D235A	244	3.2	The side chain of this residue forms a salt bridge interaction with Arg261.	*
ceClp1R288A	no activity		The side chain of this residue is involved in charge neutralization of phosphate groups.	*
ceClp1R288K	276	2.2		*
ceClp1R293L	no activity		The side chain of this residue is involved in charge neutralization of phosphate groups.	*
ceClp1R293K	residual activity			*

* Mutant variants were measured from Bernhard Loll.

3.3.2.1 Purification of truncated ceClp1 protein variants

To functionally and structurally characterize truncation variants of ceClp1, different constructs were designed and cloned via a standard cloning procedure into bacterial pET21a and pET28b expression vectors. Resulting sequence-verified clones were transformed into *E. coli* BL21-CodonPlus(DE3)-RIL cells allowing for an IPTG-inducible overexpression. The recombinant protein variants were overexpressed, followed by cell lysis via sonication, and

then further purified. In contrast to full-length protein, truncated variants appeared to be rather insoluble.

For NtD truncation variants, two constructs were designed in which the polypeptide chain was truncated in or after the N-terminal helix α1 (ceClp1ΔN85 and ceClp1ΔN107, Figure 3.1.1). Whereas purification of ceClp1ΔN85 was unsuccessful due to solubility problems, purification of ceClp1ΔN107 could be optimized by co-expression with a protein variant coding for NtD (ceClp1ΔC104). Both domains bound to each other and formed a complex that could be separated after immobilization on a Talon affinity column. Although the purification procedure included a single Talon affinity chromatography step, samples were highly homogeneous as confirmed by SDS-PAGE. Separated proteins were subsequently tested for 5'-kinase activity, which was measured by an enzymatically coupled spectroscopic ATPase assay.

Similar to the truncation experiments characterizing the functional importance of the NtD, three protein variants with a C-terminal deletion were designed. The truncations were located in helix α7 (ceClp1ΔC288, ceClp1ΔC304, and ceClp1ΔC310). Since only ceClp1ΔC310 was soluble, this mutated protein variant was purified and 5'-kinase activity was measured. Furthermore, ceClp1ΔC310 was also co-expressed with a protein variant coding for the CtD (ceClp1NC315) and the complex was tested for 5'-kinase activity.

3.3.2.2 Characterization of the truncated ceClp1 protein variants

Truncation of the NtD as well as CtD were tested enzymatically using two different IPTG inducible truncation variants ceClp1ΔN107 and ceClp1ΔC310 (Figure 3.3.2 A and B). Based on the ATP-bound crystal structure of ceClp1, specific interactions between the nucleobase and conserved residues of NtD were identified (Figure 3.1.8). The ATP molecule interacts, via hydrogen bonds, with the side chain of Glu16 and the main chain carbonyl of Arg56. Furthermore, the nucleobase forms a stacking interaction with Phe36 of the NtD. In addition, there is evidence for domain flexibility of the NtD. Sequence alignment of the eukaryotic Clp1 protein family shows a highly conserved putative hinge region within the loop connecting NtD with PNK domain Gly114 (Figure 3.1.1). This loop appears to be flexible, since a poorly defined electron density was observed for the crystal structure. Therefore, it seemed plausible that deletion of NtD should affect ceClp1 5'-kinase activity. Indeed, as a consequence of truncation, the ceClp1ΔN107 variant showed an abolished enzymatic activity

(Figure 3.3.2 C). Moreover, the mutated protein variant ceClp1ΔN107 was prone to aggregation, probably due to solvent-exposed hydrophobic patches. Interestingly, 5'-kinase activity of the ceClp1ΔN107 variant was restored by -complementation with the NtD (Figure 3.3.2 A).

In contrast to NtD, residues of CtD do not contribute to the active site of ceClp1 (Figure 3.1.4). However, the CtD and the PNK domain are linked downstream after helix α7. Helix α7 contains the LID module and a conserved arginine residue (Arg297, Figure 3.1.1). The LID module has an important role in charge stabilization of the transition state. Furthermore, Arg293 and Arg297 form hydrogen bonds to the backbone phosphates of the RNA substrate (Figure 3.1.5). Therefore, a truncation within helix α7 (ceClp1ΔC288) was expected to affect ceClp1 enzymatic activity, since three highly conserved arginine residues (Arg288, Arg293, and Arg297) are removed. Indeed, truncation at the end of helix α7 (ceClp1ΔC310) diminished ceClp1 5'-kinase activity (Figure 3.3.2 D). It is likely that the LID module becomes disordered upon removal of CtD. To restore 5'-kinase activity of ceClp1ΔC310, the CtD (ceClp1ΔN315) was *trans*-complemented. However, no kinase activity could be detected and thus *trans*-complementation is not sufficient to overcome a possible distortion of helix α7.

In conclusion, both additional domains affect ceClp1 5'-kinase activity. Whereas NtD is involved in ATP binding and shows domain flexibility, CtD could have a function in the correct positioning of the three conserved Arg288, Arg293, and Arg297 that are part of the LID module.

3.3.3 Summary

In conclusion, the site-directed mutagenesis experiments demonstrate the importance of the Walker A lysine for enzyme catalysis. Furthermore, the clasp-motif was identified, which is a novel feature of members of the Clp1 protein family.

To complement these mutagenesis experiments, ceClp1 was further characterized by deletion of the additional domains NtD and CtD. Based on these experiments, NtD seems highly flexible and this flexibility might be involved in a regulatory function. So far, the function of the CtD remained largely elusive. However, initial results suggest an important role in the positioning of the LID module.

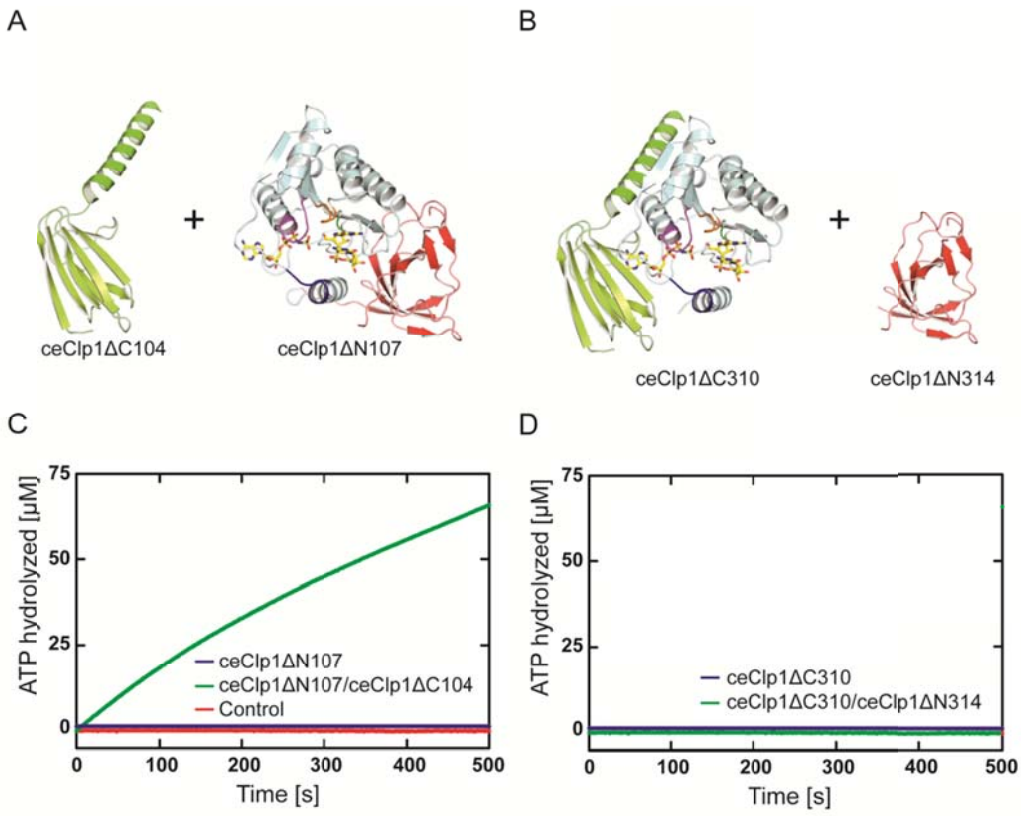


Figure 3.3.2 Characterization of N-terminal and C-terminal truncation variants of ceClp1. A and B, Ribbon diagram of the different truncation variants used for enzymatic activity. The kinase activity was monitored spectroscopically similar to the description for the coupled ATPase assay. All experiments were performed with 2 µM of protein, 1 mM ATP, and 500 µM G_1C_2. Reaction mixtures were equilibrated at 20 °C for 10 min prior to individual measurements. **C.** A truncation of ceClp1 to ceClp1ΔN107 abrogates enzymatic activity. However, after long-term incubation of ceClp1ΔN107 and ceClp1ΔC104 a nucleotide kinase activity was recovered. The control experiments without a nucleotide substrate (G_1C_2) showed no a basal ATPase activity for the complex of ceClp1ΔN107 and ceClp1ΔC104. **D,** Both ceClp1ΔC310 alone or co-expressed with ceClp1ΔN314 were abolished in enzymatic activity.

4 Discussion

4.1 The Clp1 protein family

In the course of this thesis it could be shown that Clp1 orthologs from higher eukaryotes represent novel types of RPNKs. Based on their structural conservation, Clp1 homologs define a subfamily of RPNKs, the Clp1 protein family. These RPNPs show broad phylogenetic distribution and can be found in all three kingdoms of life[86-88,95,96]. Although structurally conserved, members of the Clp1 protein family were shown to be involved in different RNA maturation pathways[86-88,90]. This functional difference might be due to additional NtD and CtD flanking the PNK domain. The NtD and CtD are novel structural elements that are unique to the Clp1 protein family. Truncation experiments with the NtD and CtD revealed their importance for ceClp1's enzymatic activity. Based on these observations, it is conceivable that either the *in vivo* function or the regulation of enzymatic activity in Clp1 depends on the structural dynamics of the additional domains, which will be discussed in more detail in the following sections.

4.1.1 Sequence conservation of the Clp1 protein family

ceClp1 is a multidomain protein composed of a central PNK domain flanked by additional NtD and CtD (Figure 3.1.1 and Figure 3.1.4). While hsClp1 contributes to mRNA processing, tRNA splicing and RNAi[88,90,92], the homologous proteins Nol9 and Grc3 were shown to participate in rRNA maturation[86,87,115,116]. Based on a sequence alignment, these enzymes define the Clp1 protein family, owing the highest degree of conservation within the PNK domain. The PNK domain is subdivided into an ATP- and an RNA-binding site. The sequence alignment also revealed four conserved structural motifs associated with substrate binding and enzyme catalysis (Figure 3.1.1 and Figure 4.1.1). The most prominent motifs for PNK are the P-loop motif GxxxxGK[T/S] and the Walker B motif DxxQ[113,114]. However, Clp1 revealed two additional motifs that are only conserved among members of the Clp1 protein family. These motifs are the LID module (RxxxxR) and the clasp motif (TxGW). The sequence of the central PNK domain seems to be conserved throughout all kingdoms of life[86,88,95,96] (Figure 4.1.1). Jain and Shuman[95] identified a Clp1 archaeal homolog from *Pyrococcus horikoshii* that displays similar RNA-specificity *in vitro* as shown for ceClp1 (Figure 3.2.1). Moreover, sequence alignments also suggested a bacterial candidate Clp1

homolog in *Nitrosococcus halophilus* (nhClp1) Figure (4.1.1). This "GTPase or GTP-binding protein-like protein" (CP001798.1) shows a high sequence homology to the PNK and the CtD domain. Interestingly, in both the bacterial (nhClp1) as well as the archaeal Clp1 (phClp1) homolog, no evidences for an NtD are found. Furthermore, Nol9 and Grc3 display sequence alterations in their NtD and CtD. On the basis of these observations it can be assumed that the additional NtD and CtD have a regulatory or a recruiting function important for the different *in vivo* functions of the enzymes. Since only Clp1 homologs in eukaryots consist of an NtD, it is likely that an NtD with a putative regulatory function was acquired during evolution to add an additional layer of regulation to its 5'-kinase activity.

4.1.2 Putative regulatory function of the additional NtD and CtD

With respect to the additional domains, ceClp1 is structurally different from all other previously described PNKs[51,52,69]. In contrast to the T4 PNK as well as the mPNK with an "open" ATP-binding site, the NtD of ceClp1 was covering its active site (Figure 4.1.2). The crystal structure of ceClp1 revealed interactions between the nucleobase and conserved residues of the NtD (Glu16, Phe38, and Arg56). Interestingly, the mPNK[125] also shows these kinds of interactions, however, in case of ceClp1, the ATP binding site is completely obstructed from the NtD (Figure 3.1.4). Truncation experiments of the NtD and the CtD showed the importance of both domains for Clp1's 5'-kinase activity (Figure 3.3.2). Truncation of both domains abolished the enzymatic activity, but only through *trans*-complementation of the NtD ceClp1's enzyme activity could be restored (Figure 3.3.2 C). These experiments suggest a function of the NtD in ATP-binding. This hypothesis is consistent with the conservation of the ATP-binding site within higher eukaryotes showing parts of the NtD that interact with ATP to be highly conserved (Figure 4.1.2). The CtD is distal to the α-helical LID module and truncation of the CtD might affect the alignment of the catalytically relevant arginine residues of the LID module. Thus, the CtD appeared to be important for the correct positioning of the LID module. A structural comparison of the NtD and CtD with structures of the PDB database revealed structural homology of the NtD to the FHA domain (r.m.s.d: 2 Å) of the mPNK. The FHA domain is involved in factor recruitment[42] by recognition of the phosphorylated forms of the two scaffold proteins XRCC1 and XRCC4. Although both the FHA domain and the NtD were composed of a two β-sheet sandwich, the topological arrangement of the strands is different (Figure 3.1.4). Moreover, a conserved loop of the FHA that recognizes the phosphorylated forms of XRCC1 and XRCC4

is missing in the NtD of ceClp1. This fact suggests that in contrast to the FHA of the mPNK, ceClp1 might be unable to recognize phosphorylated scaffold proteins for factor recruitment similar to the observations of the DNA repair[43,44]. However, interaction studies on scClp1 identified specific protein-protein interaction partners that are recognized by single domains or at the inter-domain boundary of two domains[94,98-100]. Strikingly, some of these interactions such as the one to scPcf11 are disturbed by mutations in the P-loop motif supporting a structural role of the P-loop in Clp1[98-100]. Similar interaction studies were performed for hsClp1 but not in a comparable quality[89,91]. These interaction studies showed that hsClp1 interacts with Pcf11, a factor of the mRNA 3'-end processing[89]. Moreover, it is also associated with a protein complex involved in tRNA-splicing, the TSEN complex[91]. This observation lead to the assumption that hsClp1 is a recurrent linker of different RNA maturation pathways[92]. Since the interaction interface of Pcf11 from *S. cerevisiea* and scClp1 is known and both proteins are conserved in eukaryotes[89], it was possible to model the interaction of ceClp1 and Pcf11 from *C. elegans* (cePcf11) (Figure 4.1.3).

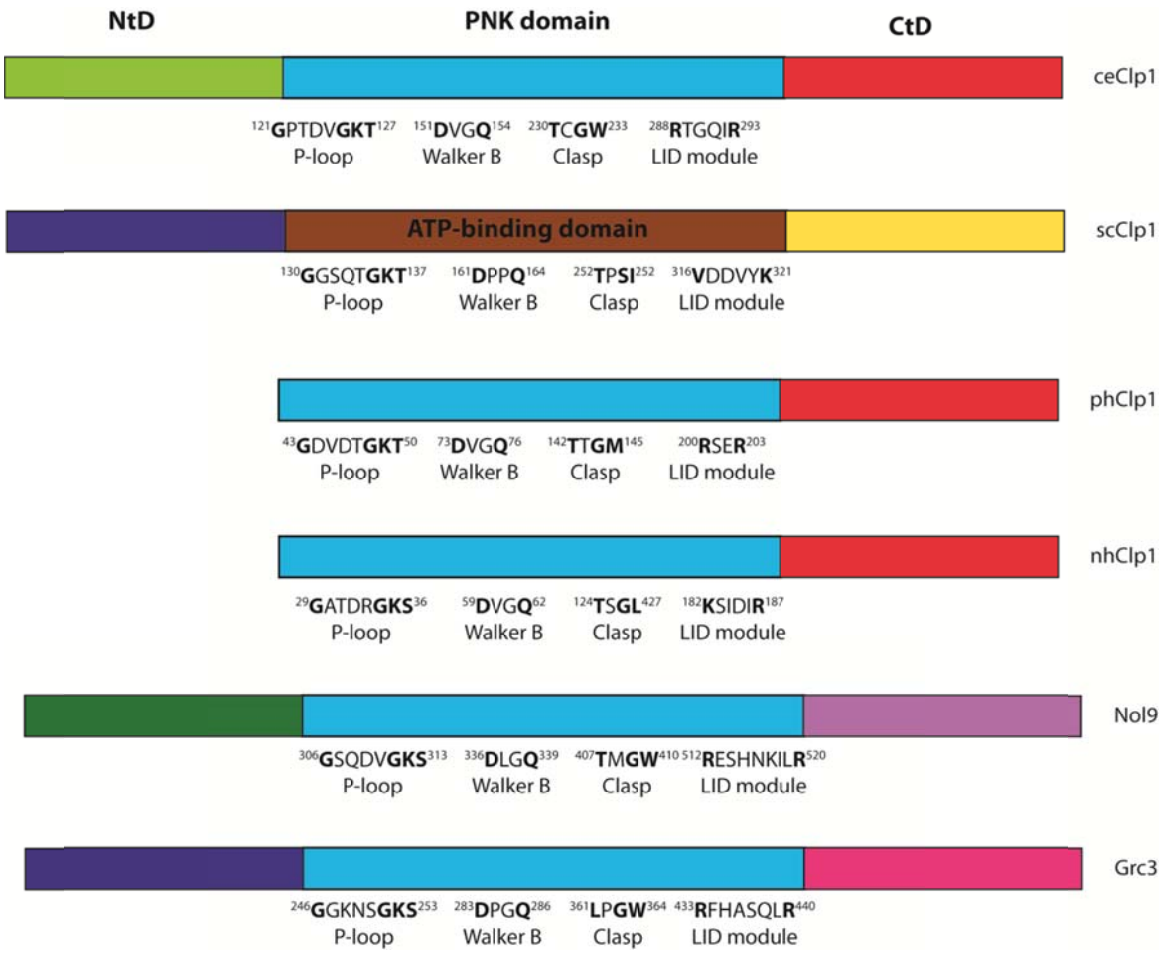


Figure 4.1.1. Sequence conservation and domain architecture of the Clp1 protein family. Members of the Clp1 protein family are present in all three kingdoms of life and show the highest homology in the PNK domain. Sequence alignment revealed 4 structural motifs being important for enzyme catalysis and/or substrate binding. The P-loop and Walker B motif show high sequence conservation, whereas the LID module and the clasp are less conserved throughout all members. Clp1 homologs are from *Caenorhabditis elegans* (ceClp1): NP_001040858; from *Saccharomyces cerevisiae* (scClp1): NP_014893, from *Pyrococcus horikoshii* (phClp1): NP_142196.1; from *Nitrosococcus halophilus* (nhClp1): CP001798.1; from *homo sapiens* (Nol9): NP_078930.3, and from *Saccharomyces cerevisiae* (Grc3): NP_013065.1.

Previously presented structural data of the inactive scClp1[94] showed that the interface between the central PNK domain and the CtD is involved in the binding of scPcf11[94]. scPcf11 together with scClp1 are subunits of the cleavage factor CF IA involved in the 3'-end processing machinery of eukaryotic Pol II transcripts[102]. Based on a structural model of the interaction interface between ceClp1 and cePcf11, it was shown that the protein-protein interactions are highly conserved in eukaryotes (Figure 4.1.3). This model provides the basis to discuss a CtD-dependent regulation of the enzymatic activity in Clp1 by affecting the correct position of the LID module.

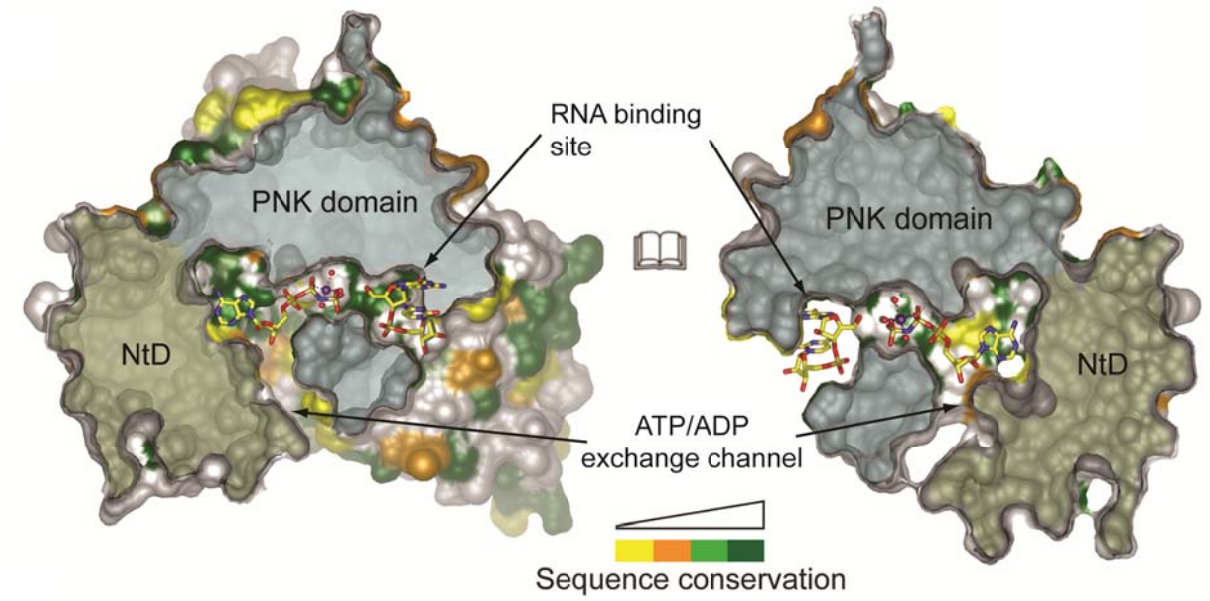


Figure 4.1.2 Conservation of the active site of ceClp1. An "Open-Book view" of ceClp1 showing the conserved ATP- and RNA-binding site. The amino acid sequence conservation (Figure 3.1.1) of Clp1 from higher eukaryotes is mapped onto the molecular surface representation and the molecule was cut into two halves shown side by side. The bound AppNHp and RNA are shown as stick representation and were mapped onto the active site of ceClp1 on each half of the molecule.

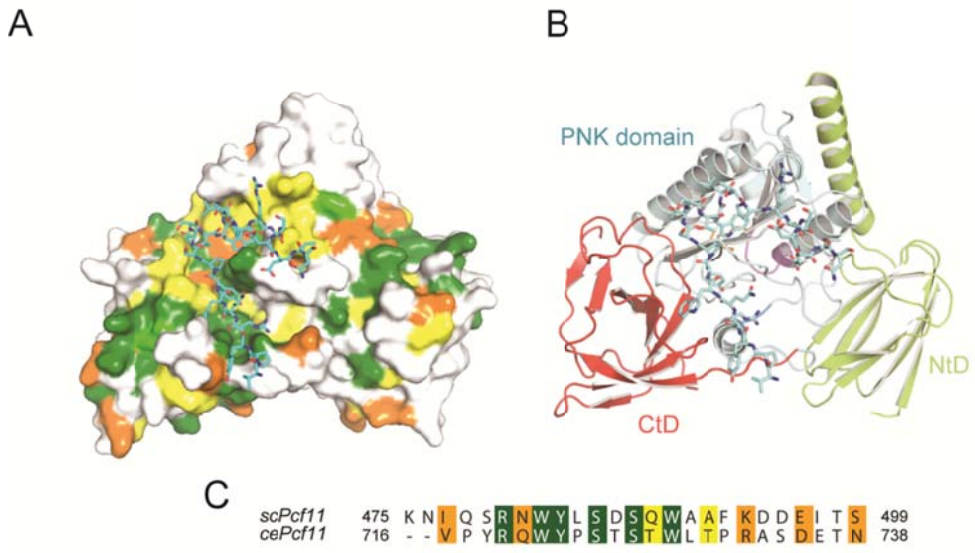


Figure 4.1.3 Protein-protein interaction interface between ceClp1 and cePcf11. A, Based on the crystal structure of scClp1 associated with scPcf11, cePcf11 was superimposed and modeled into the structure of ceClp1. cePcf11 is presented as stick model and the amino acid sequence conservation (Figure 3.1.1) of ceClp1 and seClp1 is mapped onto the molecular surface representation. The surface is colored according to the degree of conservation, decreasing from dark green to yellow. **B,** The individual domains are color coded similar to Fig. 3.1.4. **C,** Amino acid sequence alignment of the regions of cePcf11 and scPcf11 interacting with Clp1.

4.1.3 Structural difference between scClp1 and ceClp1

Structural comparison between scClp1 and ceClp1 revealed catalytically important residues that are required for ceClp1 5'-kinase activity. Whereas hsClp1 seems to be involved in mRNA and tRNA processing[88-90,92], scClp1 was shown to be inactive[94]. Even though the enzyme is inactive, the binding of ATP seems to be crucial for factor recruitment during the cleavage and polyadenylation reaction of RNA polymerase II transcripts[98-100]. The crystal structure of scClp1 (PDB 2NPI) reveals Gln133 (the corresponding residue in ceClp1 is Asp124) to tightly interact with the phosphate groups of ATP and to distort the nucleotide binding site (Figure 4.1.4). Site directed mutagenesis could show that a mutation of the Asp124 in ceClp1 to Gln abolishes enzymatic activity, whereas mutations to Asn or Ala showed much less dramatic effects (Figure 3.2.1). Furthermore, scClp1 lacks a residue that is important to neutralize the negative charges of the transition state as shown for ceClp1 (Figure 4.1.4 A and B). The arginine residues of the LID module (the corresponding residues in ceClp1 are Arg288 and Arg293) are non-conservatively mutated to Val316 and conservatively to Lys321. Therefore, it is suggested that scClp1 has lost its capability of

enzymatic activity during evolution, whereas ATP-binding by the P-loop motif remained important for structural integrity[98-100].

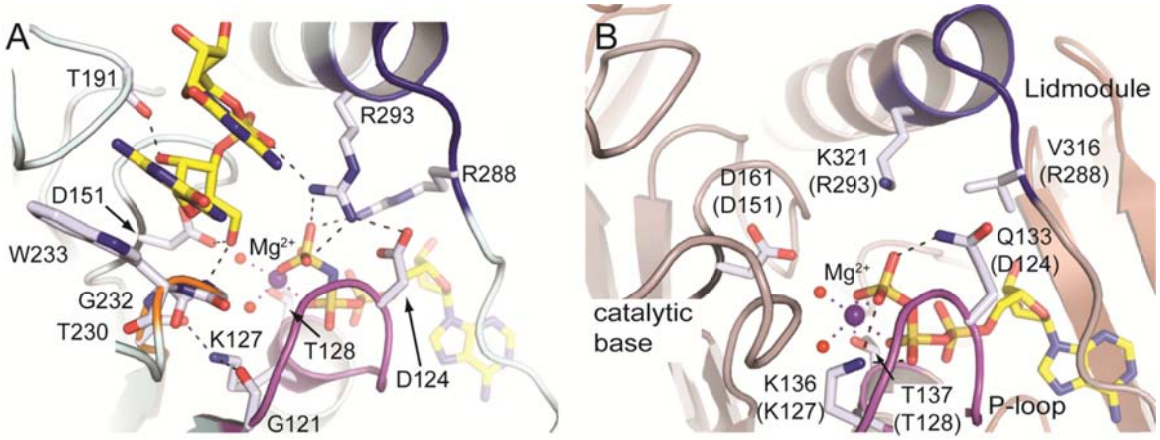


Figure 4.1.4. Comparison of the active site region of active and inactive Clp1 orthologs. A, The inhibited substrate bound state of ceClp1 **B,** scClp1 in complex with ATP-Mg^{2+}. Residues that render scClp1 incapable of ATP hydrolysis are shown in stick representation. For better comparison, scClp1 has an identical orientation as ceClp1. The corresponding amino acids in ceClp1 are indicated in brackets. In contrast to ceClp1, scClp1 shows no clasp sequence motif. Furthermore, the LID module has lost its function in scClp1 due to an exchange of Arg288 to Val316. Gln133 forms a hydrogen bond to the ATP, which interferes with a correct positioning of the ATP.

4.2 RNA-recognition and RNA-specificity in PNKs

PNKs are known to be important in DNA and RNA repair, RNA maturation as well as RNA degradation processes. Based on their specific function, these enzymes show different substrates specificity[60,86-88,158]. Crystals of ceClp1 bound to RNA provide first structural data of a PNK in complex with an RNA-substrate. A comparison of ceClp1 with DNA bound crystal structure of the T4 PNK (PDB 1RRC) and the mPNK (3ZVN) revealed a novel "RNA-sensor" and a clasp-like binding mechanism. These differences and the implications of the structural data for its substrate specificity are discussed in this section.

Since hsClp1 was shown to phosphorylate *in vivo* siRNAs but also tRNAs, a detailed characterization of its substrate specificity was provided[64,88,90]. However, the structural features involved in the RNA-specificity as well as in RNA recognition have remained elusive so far. For crystallization experiments, the uncharacterized Clp1 ortholog from *C. elegans* was used. ceClp1 was identified as a *bona fide* RPNK showing similar *in vitro* substrate specificity compared to the recently identified RPNKs (hsClp1[88], Nol9[87], Grc3[86], and archaeal Clp1[95]; Figure 3.2.1). Interestingly, in addition to their similar *in vitro* substrate specificity,

these homologs also show a high sequence homology of the PNK domain (Figure 4.1). Therefore, it is likely that structural features of the PNK domain are responsible for the specificity of the 5'-nucleotide substrate *in vitro*. These structural features mainly depend on the clasp motif and on the "RNA-sensor" recognizing the 2'-hydroxyl group at the ultimate position of the RNA.

Similar to previously described PNKs[42,51,52], ceClp1 interacts with its RNA substrate via a combination of several interactions including the backbone phosphates, sugars, and the bases themselves (Figure 3.1.6). In ceClp1, the ultimate nucleobase stacks on the hydrophobic side chains of Trp233. This residue is part of the clasp motif. Mutation of Trp233 to alanine effectively impaired enzyme catalysis (Figure 3.3.1). Even though the clasp motif represents a novel consensus sequence (TxGW) for RNA-binding, it is highly conserved throughout higher eukaryotes (Figure 3.1.1). This binding mechanism is similar to other PNKs[42,51,52]. The T4 PNK and the mPNK have similar hydrophobic interaction pockets. In contrast to Trp233, nucleobases are stacked to smaller hydrophobic residues (Val135 and Val477, T4 PNK and mPNK, respectively[42,51,52]). Another unique feature of Clp1's clasp motif is a hydrogen bond formed by Gly232 to the 5'-hydroxyl group of the RNA substrate. Thereby, the clasp ensures that the RNA is positioned in a favored 3'-endo form and in anti-conformation (Figure 3.1.5).

Interestingly, despite their similarities in binding of the ultimate nucleobase, ceClp1 shows a major difference in the orientation of the nucleotide, which is inverted compared to the other PNKs. This observation suggests convergent evolution of Clp1 in the protein family of PNKs. Despite these differences, the binding mechanism of the 3'-5' bridging phosphate groups seems to be conserved within the protein family of PNKs[42,51,52]. Again, Clp1, T4 PNK, and mPNK show a similar binding mechanism. In each case the first two 3'-5' bridging phosphate groups are recognized by conserved arginine residues (Arg34, Arg38 and Arg395, Arg432; T4 PNK and mPNK, respectively). The structure described in this thesis shows a hydrogen bond as well as salt bridges between three conserved residues (Gln154, Arg293, and Arg297) and the first two 3'-5' bridging phosphate groups being important for binding (Figure 3.1.5). The importance was also shown by comprehensive mutational analysis (Table 3.3.1). In this respect, ceClp1 is similar to the majority of RNA-binding enzymes, which are interacting more with the bound 3'-5' bridging phosphate groups and less with the nucleobase or the nucleosugar[159]. However, in contrast to the T4 PNK, which was previously described to phosphorylate 3'-nucleoside monophosphates[46], ceClp1 was unable to use this substrate (Figure 3.2.2). Although structurally similar, ceClp1 imposes different requirements on minimal substrate length compared to T4 PNK. Comparison of the active site of all three

PNKs reveals slight differences in binding of the first bridging phosphate. T4 PNK[126] shows the tightest interaction with its substrate compared to ceClp1 and mPNK[125]. Thus, ceClp1 and mPNK require dinucleotides or oligonucleotides for efficient binding[42,85].

As shown by experiments aiming at identification of the preferred substrate (Figure 3.2.1), Clp1 can discriminate single-stranded RNA against single-stranded DNA[88]. The possible explanation for this selectivity of ceClp1 is the main chain carbonyl of Thr191 that forms a hydrogen bond to the ribose 2'-hydroxyl group of the ultimate nucleotide (Figure 3.1.5). In contrast to published crystal structures of the other PNKs[42,51,52], Clp1 is the only PNK that recognizes the 2'-hydroxyl group. Interestingly, Clp1's RNA-sensor only recognizes the 2'-hydroxyl group at the ultimate position of the substrate. Phosphorylation assays showed that replacing the RNA molecule at the ultimate position by a desoxyribonucleotide abolishes hsClp1's ability to phosphorylate synthetic substrate in hsClp1[88]. Strikingly, it was possible to model an ideal double-stranded A-form RNA bound to ceClp1 without any steric hindrance by superposition with the single-stranded RNA observed in the crystal structure. This model explains why Clp1 is able to phosphorylate both, single as well as double-stranded RNAs[88]. The modeled double-stranded RNA has no steric clashes when modeled for longer double-stranded RNA substrates (Figure 3.1.6 A). In contrast to the 5'-acceptor strand, the complementary strand shows no contact with the protein, except for modeled substrates with a 3'-overhang (Figure 3.1.6 B). The continuous base stacking interaction of the 3'overhang of RNA has to be interrupted to become accommodated in the RNA-binding site of ceClp1. This observation might explain why double-stranded 3'-overhang RNA and DNA substrates were phosphorylated less efficiently compared to blunt end substrates. In conclusion, the crystal structures of Clp1 bound to RNA explain the selectivity for both single-stranded and double-stranded RNA substrates.

4.3 Phosphoryl transfer reaction

Even though extensive biochemical and structural studies are available on the phosphoryl transfer reactions of P-loop kinases, relatively little is known about the functional mechanism in PNKs. This thesis provides structural data about the first reaction trajectory of PNKs. Crystal structures of ceClp1 were trapped during enzyme catalysis such as an inhibited substrate bound state, a transition state analog and a product bound state. By comparing ceClp1's active site geometry between ground state and transition state, a general model for enzyme catalysis in PNKs is derived and will be discussed in detail in this section.

4.3.1 A molecular model of the phosphoryl transfer reaction in PNKs

Similar to other P-loop kinases, the active site of ceClp1 is formed by residues of the classical P-loop motif, a LID module, and a catalytic Mg^{2+} ion[19]. Interestingly, although the Walker A lysine of the P-loop is classically involved to interact in a bifurcated manner with the β- and γ-phosphate groups of ATP[113], ceClp1's Lys127 was captured outside the active site in both the inhibited substrate bound state (Figure 3.1.10) and the product bound state (Figure 3.1.12). In the ground state, Lys127 was "switched-off", whereas in the crystal structure of the transition state analog, Lys127 was switched into an active conformation (Figure 3.1.11). The "switched-on" Lys127 assists in enzyme catalysis by stabilizing the developing negative charges on the leaving group (3.1.11). The observation of the "arrested" Walker A lysine is in contrast to the prevailing structural and biochemical dogma underlying the Walker A lysine[113]. To exclude the possibility that ceClp1 belongs to the special group of Walker A lysine independent P-loop kinases[160-163], the functional role of Lys127 for enzyme catalysis was tested by non-conservative and conservative point mutations. Both mutant variants, K127A and K127R (Table 3.3.1), were abolished in their enzymatic activity similar to the results of Walker A lysine mutants in hsClp1[88]. This suggests that Clp1 uses a substrate gating mechanism composed of the catalytically important Lys127, which is switched on and off during enzyme catalysis. A highly regulated RPNK would have been in accordance with the biochemical characterization on ceClp1 revealing a quasi-unidirectional 5'-kinase (Figure 3.2.8) that suppresses futile side reactions (Figure 3.2.5 and 3.2.6).

Despite this exclusive substrate-gating mechanism, the overall architecture of ceClp1's active site resembles those of previously described P-loop kinases[19]. The proposed model of the phosphoryl transfer reaction follows a mechanism of a classical associative transition state[124] (Figure 3.1.11). The enzymatic reaction starts with the nucleophilic attack of the deprotonated 5'-hydroxyl group. Based on the crystal structure of ceClp1, the 5'-hydroxyl group of the bound RNA was shown to be in an ideal hydrogen bonding distance to the catalytic base (Asp151). Asp151 is proposed to abstract the proton of the attacking 5'-hydroxyl group prior to the nucleophilic attack on the γ-phosphate group of ATP. In fact, mutation of Asp151 to alanine abrogates the phosphoryl transfer reaction in ceClp1 (Table 3.3.1), supporting the function of Asp151 as general base. This result could also be shown for the equivalent residue in hsClp1[97]. The function of the general base is highly conserved in PNKs[42,51,52]. A comparison of ceClp1 with a selection of other PNKs[42,51,52] (Asp35 and

Asp396 in T4 PNK and mPNK, respectively) reveals a highly conserved aspartate residue in ideal hydrogen bonding distance to the respective position. Additionally, site-directed mutagenesis showed the catalytical importance of those general bases[164]. Also, the 5'-hydroxyl group of the incoming nucleophile is further recognized by a hydrogen bond from the main-chain amide of Gly232 (see 3.1.5), which is part of the RNA-binding clasp. Gly232 might contribute to the correct position of the attacking nucleophile.

On the opposite site of the attacking nucleophile, an ATP molecule is positioned enabling an efficient in-line phosphoryl transfer reaction. The phosphate groups of the ATP interact with the main-chain amide of the residues of the P-loop motif. In addition to the P-loop motif, an α-helical LID module together with a catalytic Mg^{2+} ion assists in charge neutralization of the ATP phosphate groups[19,113]. In the inhibited substrate bound state, two catalytically important arginine residues Arg288 and Arg293 form salt-bridges between their guanidinium groups and the oxygen atoms of the β- and γ-phosphate groups. Both arginine residues are essential for enzyme catalysis. Non-conservative mutations (R288A, R293L) led to an effectively abolished enzymatic activity, to an extent comparable with the Walker A lysine mutant K127A (Table 3.3.1). Interestingly, conservative mutation to lysine (R288K and R293K) revealed functional differences between Arg288 and Arg293. Whereas R288K appeared to be slightly impaired in kinase activity, R293K was inactivity (Table 3.3.1). Thus, Arg288 seems to be important for neutralizing the developing negative charges during the transition state. However, Arg293 additionally contributes to the RNA-binding site by its guanidinium group interacting in a bidendate mode forming salt-bridges to the phosphate groups of the ATP as well as the 5'-3' bridging phosphate of the RNA (Figure 3.1.5).

A LID module with conserved arginine residues is a common feature of the reaction mechanism in PNKs and P-loop kinases in general[19]. A structural comparison with other described PNKs identified similar arginine residues in T4 PNK[51,164] (Arg122 Arg126). Mutagenesis studies on T4 PNK showed the enzymatic relevance of those arginine residues as it was shown for ceClp1[165]. Based on the contribution of Arg293 in RNA-binding it is suggested that this residue also plays a role in RNA release after the phosphoryl transfer reaction. In the inhibited substrate bound state Arg293 interacts with the phosphate backbone of the RNA, but in the crystal structures of ceClp1 bound to ADP and $G_1A_2A_3A_4$, Arg293 is reoriented that breaks up the salt bridge to the RNA. Arg293 is captured by Asp124 as shown for the product bound state. Both residues function in a latch-catch mechanism[166]. Mutagenesis studies on Asp124 showed additionally the enzymatic relevance of the latch-catch by mutations of Asp124 and Arg293 (Table 3.3.1).

In the transition state, AlF_4^- is perfectly orientated apical to both the oxygen atoms of the ADP and the RNA necessary for the "in-line" transfer mechanism. Superposition of the substrate-bound and the transition state analog revealed that the donor and acceptor oxygen atoms approaches each other by 0.9 and 1.2 Å, respectively (Figure 4.3.1). Similarly to the "on-off switched" Lys127, also the LID module Arg288 and Arg293 residues are fully activated to neutralize negative charges of the transfer reaction. While the lysine switch has not been described before, arginine switches are reminiscent of the situation observed in NMP kinases[118]. Considering the positively charged vicinity that is necessary to stabilize charges of the transition state analog, ceClp1's phosphoryl transfer reaction mechanism is suggested to follow an associative reactions mechanism[124]. In conclusion, based on the structural conservation of the PNK domain, our proposed model for enzyme catalysis is generally applicable for PNKs from all three kingdoms of life.

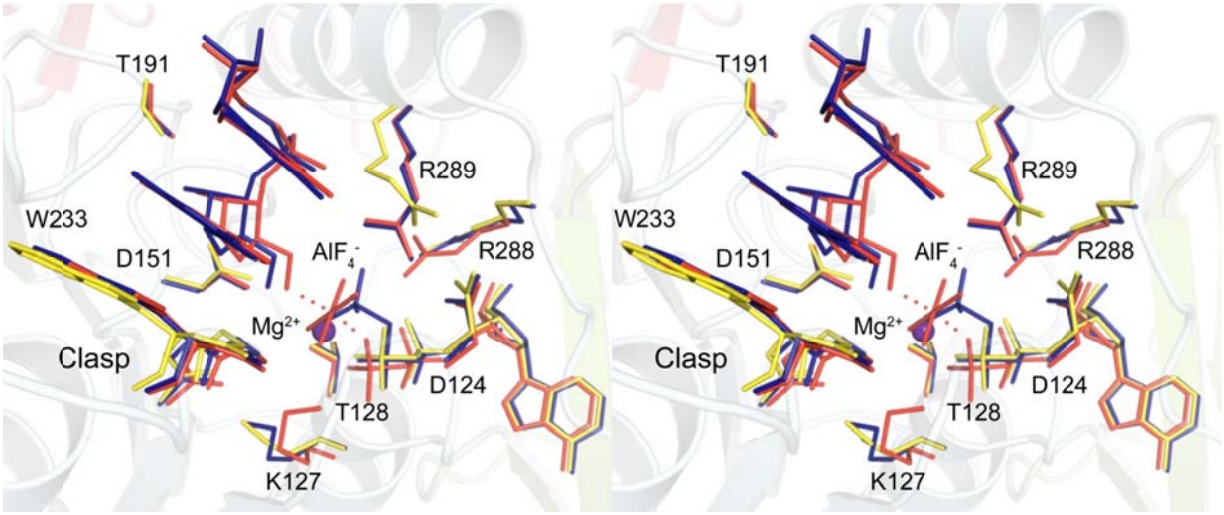


Figure 4.3.1: ceClp1 shows local conformational changes during the phosphoryl transfer reaction. Cross-eyed stereo-view of the inhibited substrate bound state (blue), the transition state analog (red), and the RNA released product bound state (yellow). In order to visualize conformational changes, crystal structures were superimposed and are shown as a stick models. A comparison of these crystal structures reveals only minor conformational changes, besides the movement of Lys127, Arg288, and Arg293. Furthermore, ATP and RNA move in closer proximity.

Table 4.3.1. Table of proteins owing a non-conventional Walker A lysine residue.

Nr.	PDB ID	Protein name	Resolution [Å]	P-loop Sequence	P-loop residue number and chain	Ligand	Walker A Lys interactions
Non-conventional Walker A lysine structures							
1	XXXX	ceClp1 polynucleotide kinase	2.10	GPTDVGKT	121-128 (A)	ANP	no nucleotide interactions
2	XXXX	ceClp1 polynucleotide kinase	2.10	GPTDVGKT	121-128 (A)	ADP	no nucleotide interactions
3	3A1S	FeoB iron iransporter	1.50	GCPNVGKT	23-30 (A)	GDP	no nucleotide interactions
4	4FCW	ClpB	2.35	GPTGVGKT	595-602 (A)	ADP	no nucleotide interactions
5	2XTN	Gimap2	1.90	GKTGTGKS	29-36 (A)	GTP	no nucleotide interactions
6	2XTO	Gimap2	2.80	GKTGTGKS	29-36 (A)	GDP	no nucleotide interactions
7	3PQC	Ribosome biogenesis GTP-binding protein EngB	1.90	GRSNVGKS	30-37 (A)	GDP	no nucleotide interactions
8	3QKT	Rad50 ABC-ATPase	1.90	GQNGSGKS	30-37 (A)	ANP	γ-phosphate
9	3PXN	Kinesin family member Kin10/NOD	2.60	GQTGTGKS	87-94 (A)	ADP	no nucleotide interactions
10	4AYT	ABC transporter ABCB10	2.85	GPSGSGKS	527-534 (A)	ACP	β-phosphate
11	3NH9	ABCB6	2.10	GPSGAGKS	623-630 (A)	ATP	β-phosphate
12	2JDI	F1-ATPase	1.90	GDRQTGKT	169-176 (A)	ANP	β-phosphate
13	2VHJ	P4 protein from bacteriophage	1.80	GKGNSGKT	130-137 (A)	ADP	no nucleotide interactions
14	4DCV	EngA	2.60	GRPNVGKS	182-189 (A)	GCP	no nucleotide interactions
15	3K0S	DNA mismatch repair protein MutS	2.20	GPNMGGKS	614-621 (A)	ADP	no nucleotide interactions
16	3RC3	Helicase Suv3	2.08	GPTNSGKT	10-17 (A)	ANP	no nucleotide interactions
17	1NKS	Adenylate kinase	2.57	GIPGVGKS	8-15 (F)	ADP	no nucleotide interactions
18	1R6B	ClpA	2.25	GPTGVGKT	495-502 (X)	ADP	no nucleotide interactions
19	2CK3	F1-ATPase	1.95	GDRQTGKT	169-176 (A)	ANP	β-phosphate

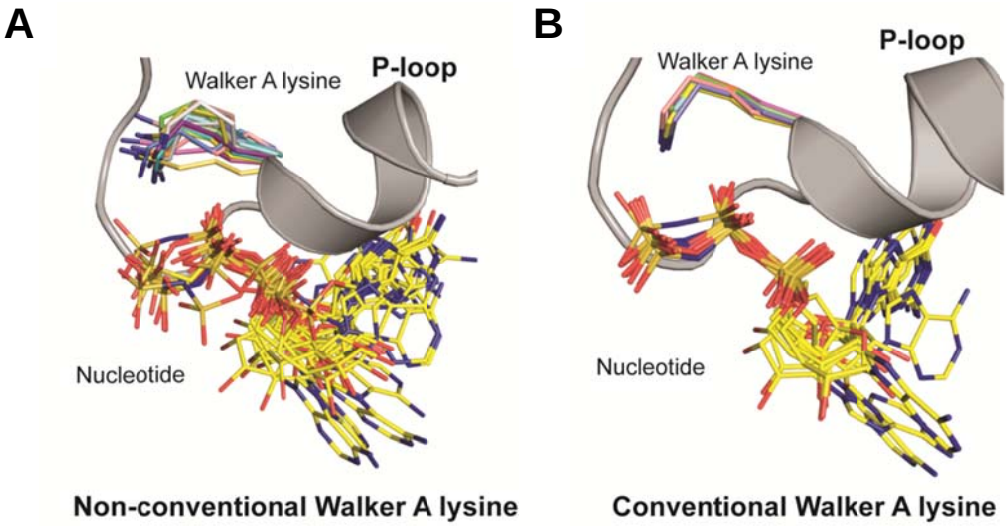


Figure 4.3.2: Crystal structures of non-conventional and conventional Walker A lysine residues. Close up view of the active site of different P-loop kinases showing either **A,** a **non-**conventional or **B,** a conventional conformation of the Walker A-lysine. The Walker A lysine and the nucleotide are represented as stick-model, and the P-loop motif as ribbon diagram. The conventional Walker A lysine variants are obtained from PDB entries from 6.2.1 (substrate bound state), whereas the non-conventional Walker A lysine residues belong to table 4.3.1.

4.3.2 Putative function of the non-canonical Walker A lysine Lys127

Conformational gating is already known since 1981[167], but a comprehensive characterization of gating mechanisms in general is still missing. The well known Walker A lysine is an unexpected residue for such a gating mechanism. Apparently, prior to the phosphoryl transfer reaction, Lys127 is in a "gated" conformation (Figure 3.1.10). The residue is arrested by the P-loop and more importantly by the clasp, which is involved in RNA-binding. ceClp1 seems to impose novel constraints on ligand binding, whereas other PNKs[42,51,164] seem to follow the classical reaction mechanism of phosphoryl transferases. In the transition state complex, however, the Walker A lysine has to re-orientate to compensate emerging negative charges during the phosphoryl transfer reaction. Indeed, in the structure with the transition state analog (ADP-AlF_4^--G_1C_2), Lys127 has moved into the active site and interacts in a bifurcated manner with an oxygen atom of the β-phosphate group and a fluorine atom of the planar AlF_4^-, representing a transition state geometry of a phosphoryl group being transferred (Figure 3.1.11). Crystallization artefacts by the use of the non-hydrolysable ATP analog (AppNHp) were excluded since the ATP structure also showed a Walker A lysine in an identical "arrested" conformation (Figure 3.1.8).

According to the Protein Data Bank, there are a significant number of other P-loop kinases with similar Walker-A lysine residues in an "arrested" conformation (Table 4.3.1). However, the relevance of this peculiar conformation of the Walker A lysine remained undiscussed in the literature until now. A superposition of these crystal structures shows that the triphosphate groups do not align and that their position is rather randomly distributed (Figure 4.3.2 A). In contrast, triphosphate groups of cases in which the Walker A lysine adopts the classical conformation, these phosphate groups are virtually identically oriented (Figure 4.3.2 B). The eukaryotic Clp1 seems to have evolved a sophisticated mechanism to repress basal site activities. One of these regulatory mechanisms is the newly identified Walker A lysine functioning as a molecular door step. Also the latch-catach (Arg293-Asp124) seems to play an important role in enzyme catalysis. These two active site regulation elements are in addition to the function of the NtD and CtD that were also schon to affect enzymatic activity of ceClp1.

4.4 Conclusions and Outlook

4.4.1 Clp1 a novel eukaryotic, RPNK

This work provides new structural, functional and mechanistic insights into the Clp1 protein family. RPNK represents a novel type of PNK that shows high sequence conservation within the PNK domain. Based on a sequence alignment, four different consensus sequences were identified that are characteristic for this protein family. The clasp motif together with the "RNA-sensor" enables Clp1 to contribute to different RNA maturation pathways as a recurrent linker. Both single-stranded RNA as well as double stranded DNA and RNA is accommodated in the RNA-binding site of this protein family[86-88,95]. Whereas hsClp1 is part of the mRNA, tRNA and RNAi maturation pathways[88,92], Nol9[87] and Grc3[86,115,116] are also involved in rRNA maturation. It could therefore be suggested that the different RNA maturation pathways are interconnected to a global RNA metabolism controlled by the Clp1 protein family. Currently, knowledge about *in vivo* substrates of Clp1 is limited[90]. However, the structural data of this thesis could provide the framework to understand RNA-recognition and RNA-specificity in more detail. In combination with a detailed biochemical characterization, future experiments will unveil novel *in vivo* substrates of the RNA-metabolism.

In contrast to previously described PNKs, Clp1 also consists of additional NtD and CtD which are flanking the PNK domain. The initial characterization of truncation variants presented in this work suggests that the enzymatic activity of ceClp1 is directly affected by these additional domains. Future experiments aiming towards a better understanding of the function of these additional domains will be reasonable. The additional NtD and CtD possibly serve as a protein-protein interaction interface. This interface may regulate Clp1's enzymatic activity by structural dynamics in domain flexibility. Possible strategies to test this protein-induced regulation include a phosphorylation assay measured in the presence of putative interaction partners. A possible candidate for initial screening experiments could be Pcf11, the interaction partner of the CFIIm[89].

4.4.2 Functional diversity within the Clp1 protein family

Although ceClp1 and scClp1 show a high sequence similarity, both enzymes are functionally different. scClp1 is enzymatically inactive[94], however, still able to bind ATP and ADP molecules[94,98-100]. ATP binding seems to be required for factor recruitment during the cleavage and polyadenylation reaction of RNA polymerase II transcripts[98-100]. The crystal structure of ceClp1 revealed structural features involved in enzyme catalysis. A comparison between scClp1 and ceClp1 shows that catalytically important residues are non-conservatively mutated, the P-loop motif, however, remains unaffected. Apparently, scClp1 has lost its PNK activity during evolution, whereas structural integrity of scClp1 still requires ATP-binding[98-100]. The bioinformatical and structural data suggest catalytically important residues that are required for active Clp1 homologs. These sequence profiles can be used to search for other active variants in all three kingdoms of life. A functional comparison of Clp1 homologs from all kingdoms of life may allow new insights into RNA metabolism such as underestimated cross-connection between mRNA, tRNA, rRNA, and RNAi.

4.4.3 The central dogma of the Walker A lysine has to be reconsidered

The presented structural work of ceClp1 provides a detailed model of the phosphoryl transfer reaction mechanism for a RPNK. Based on structural conservation of the PNK domain, this proposed model is generally applicable for PNKs from all three kingdoms of life. Interestingly, in ceClp1 the enzymatic activity is regulated by a substrate-gating

mechanism[168]. The Walker A lysine of ceClp1 functions as a door step that prevents futile ATP hydrolysis. This observation is consistent with experiments showing that ceClp1 has no basal ATPase activity and only functions in the forward reaction as a quasi unidirectional enzyme. Eukaryotic Clp1 seems to have evolved a sophisticated mechanism to repress basal site activities. Interestingly, a detailed search in the PDB database identified several other proteins showing a putative molecular switch for the Walker A lysine. In all cases, the Walker A lysine is in an arrested conformation, however, a corresponding activated structure is missing. Therefore, it will be an exciting task to further characterize a representative selection of these proteins. The identification of other Walker A lysine switches would argue for a novel function of the P-loop motif.

5 Acknowledgement

This work would have been impossible without the continuous support of so many people during the last years. These people contributed in various ways and I am very grateful for all the help. I would like to thank…

… my direct supervisor Dr. Anton Meinhart for giving me the opportunity to work on a challenging and exciting project in his laboratory, for his scientific guidance and for his help trying to bring me on the right track.

... Prof. Dr. Ilme Schlichting for her willingness of being my first appraiser and for the continuous support.

… Dr. Bernhard Loll for his advice and guidance in the beginning of the project.

… Dr Tatiana Domratcheva for her chemical expertise and her excellent suggestions.

… Dr. Thomas Barends and Dr. Andreas Winkler for help with crystallographic problems.

... Present and former members of the Meinhart lab for the scientific support and the social and enjoyable lab atmosphere throughout the years. Thanks to Andrea, Christina, Brad, Hannes, Iris, Julian, Juliane and Stefano. In particular, I would like to thank Maike Gebhardt for her great assistance and her kindness.

... Chris Roome for excellent IT service and for spreading good mood very morning.

… Present and former members of the BMM department for the great working atmosphere.

… Stephan and Anikó for their contiuous support with delicious and exotic smoothies.

… Important people outside the RNA-world. Verena, Stefan, Jan, Lena, Tine, and Jannik - thank you for providing a buffer to survive the rainy days of research.

At last, I would like to express my deepest gratitude towards my parents, my brothers

and my two nieces Lara and Sueda for their continuous support and their love.

My apologies to all others who are not mentioned by name but contributed in many ways.

References

1 Steitz, T. A. DNA polymerases: structural diversity and common mechanisms. *The Journal of biological chemistry* **274**, 17395-17398 (1999).

2 Brueckner, F., Ortiz, J. and Cramer, P. A movie of the RNA polymerase nucleotide addition cycle. *Current opinion in structural biology* **19**, 294-299, doi:10.1016/j.sbi.2009.04.005 (2009).

3 Van Rompay, A. R., Johansson, M. and Karlsson, A. Phosphorylation of nucleosides and nucleoside analogs by mammalian nucleoside monophosphate kinases. *Pharmacology & therapeutics* **87**, 189-198 (2000).

4 Richards, J., Liu, Q., Pellegrini, O., Celesnik, H., Yao, S., Bechhofer, D. H., Condon, C. and Belasco, J. G. An RNA pyrophosphohydrolase triggers 5'-exonucleolytic degradation of mRNA in Bacillus subtilis. *Molecular cell* **43**, 940-949, doi:10.1016/j.molcel.2011.07.023 (2011).

5 Shuman, S. Structure, mechanism, and evolution of the mRNA capping apparatus. *Progress in nucleic acid research and molecular biology* **66**, 1-40 (2001).

6 Dorleans, A., Li de la Sierra-Gallay, I., Piton, J., Zig, L., Gilet, L., Putzer, H. and Condon, C. Molecular basis for the recognition and cleavage of RNA by the bifunctional 5'-3' exo/endoribonuclease RNase J. *Structure* **19**, 1252-1261, doi:10.1016/j.str.2011.06.018 (2011).

7 Jinek, M., Coyle, S. M. and Doudna, J. A. Coupled 5' nucleotide recognition and processivity in Xrn1-mediated mRNA decay. *Molecular cell* **41**, 600-608, doi:10.1016/j.molcel.2011.02.004 (2011).

8 Deana, A., Celesnik, H. and Belasco, J. G. The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature* **451**, 355-358, doi:10.1038/nature06475 (2008).

9 Caponigro, G. and Parker, R. Mechanisms and control of mRNA turnover in Saccharomyces cerevisiae. *Microbiological reviews* **60**, 233-249 (1996).

10 Mackie, G. A. Ribonuclease E is a 5'-end-dependent endonuclease. *Nature* **395**, 720-723, doi:10.1038/27246 (1998).

11 Tock, M. R., Walsh, A. P., Carroll, G. and McDowall, K. J. The CafA protein required for the 5'-maturation of 16 S rRNA is a 5'-end-dependent ribonuclease that has context-dependent broad sequence specificity. *The Journal of biological chemistry* **275**, 8726-8732 (2000).

12 Walsh, A. P., Tock, M. R., Mallen, M. H., Kaberdin, V. R., von Gabain, A. and McDowall, K. J. Cleavage of poly(A) tails on the 3'-end of RNA by ribonuclease E of Escherichia coli. *Nucleic acids research* **29**, 1864-1871 (2001).

13 Jiang, X. and Belasco, J. G. Catalytic activation of multimeric RNase E and RNase G by 5'-monophosphorylated RNA. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9211-9216, doi:10.1073/pnas.0401382101 (2004).

14 Pellegrini, O., Mathy, N., Condon, C. and Benard, L. In vitro assays of 5' to 3'-exoribonuclease activity. *Methods in enzymology* **448**, 167-183, doi:10.1016/S0076-6879(08)02609-8 (2008).

15 Stevens, A. and Poole, T. L. 5'-exonuclease-2 of Saccharomyces cerevisiae. Purification and features of ribonuclease activity with comparison to 5'-exonuclease-1. *The Journal of biological chemistry* **270**, 16063-16069 (1995).

16 Stevens, A. 5'-exoribonuclease 1: Xrn1. *Methods in enzymology* **342**, 251-259 (2001).

17 Weinfeld, M., Mani, R. S., Abdou, I., Aceytuno, R. D. and Glover, J. N. Tidying up loose ends: the role of polynucleotide kinase/phosphatase in DNA strand break repair. *Trends in biochemical sciences* **36**, 262-271, doi:10.1016/j.tibs.2011.01.006 (2011).
18 Amitsur, M., Levitz, R. and Kaufmann, G. Bacteriophage T4 anticodon nuclease, polynucleotide kinase and RNA ligase reprocess the host lysine tRNA. *The EMBO journal* **6**, 2499-2503 (1987).
19 Leipe, D. D., Koonin, E. V. and Aravind, L. Evolution and classification of P-loop kinases and related proteins. *Journal of molecular biology* **333**, 781-815 (2003).
20 Morrison, J. F. and Heyde, E. Enzymic phosphoryl group transfer. *Annual review of biochemistry* **41**, 29-54, doi:10.1146/annurev.bi.41.070172.000333 (1972).
21 Wolters, S. and Schumacher, B. Genome maintenance and transcription integrity in aging and disease. *Frontiers in genetics* **4**, 19, doi:10.3389/fgene.2013.00019 (2013).
22 Loeb, L. A. and Harris, C. C. Advances in chemical carcinogenesis: a historical review and prospective. *Cancer research* **68**, 6863-6872, doi:10.1158/0008-5472.CAN-08-2852 (2008).
23 Paz-Elizur, T., Sevilya, Z., Leitner-Dagan, Y., Elinger, D., Roisman, L. C. and Livneh, Z. DNA repair of oxidative DNA damage in human carcinogenesis: potential application for cancer risk assessment and prevention. *Cancer letters* **266**, 60-72, doi:10.1016/j.canlet.2008.02.032 (2008).
24 Madhusudan, S. and Middleton, M. R. The emerging role of DNA repair proteins as predictive, prognostic and therapeutic targets in cancer. *Cancer treatment reviews* **31**, 603-617, doi:10.1016/j.ctrv.2005.09.006 (2005).
25 McKinnon, P. J. DNA repair deficiency and neurological disease. *Nature reviews. Neuroscience* **10**, 100-112, doi:10.1038/nrn2559 (2009).
26 Hoeijmakers, J. H. DNA damage, aging, and cancer. *The New England journal of medicine* **361**, 1475-1485, doi:10.1056/NEJMra0804615 (2009).
27 Burma, S., Chen, B. P. and Chen, D. J. Role of non-homologous end joining (NHEJ) in maintaining genomic integrity. *DNA repair* **5**, 1042-1048, doi:10.1016/j.dnarep.2006.05.026 (2006).
28 Valko, M., Izakovic, M., Mazur, M., Rhodes, C. J. and Telser, J. Role of oxygen radicals in DNA damage and cancer incidence. *Molecular and cellular biochemistry* **266**, 37-56 (2004).
29 Ciccia, A. and Elledge, S. J. The DNA damage response: making it safe to play with knives. *Molecular cell* **40**, 179-204, doi:10.1016/j.molcel.2010.09.019 (2010).
30 Sclafani, R. A. and Holzen, T. M. Cell cycle regulation of DNA replication. *Annual review of genetics* **41**, 237-280, doi:10.1146/annurev.genet.41.110306.130308 (2007).
31 Sherman, M. H., Bassing, C. H. and Teitell, M. A. Regulation of cell differentiation by the DNA damage response. *Trends in cell biology* **21**, 312-319, doi:10.1016/j.tcb.2011.01.004 (2011).
32 Buchko, G. W. and Weinfeld, M. Influence of nitrogen, oxygen, and nitroimidazole radiosensitizers on DNA damage induced by ionizing radiation. *Biochemistry* **32**, 2186-2193 (1993).
33 Mills, K. D., Ferguson, D. O. and Alt, F. W. The role of DNA breaks in genomic instability and tumorigenesis. *Immunological reviews* **194**, 77-95 (2003).
34 Alagoz, M., Gilbert, D. C., El-Khamisy, S. and Chalmers, A. J. DNA repair and resistance to topoisomerase I inhibitors: mechanisms, biomarkers and therapeutic targets. *Current medicinal chemistry* **19**, 3874-3885 (2012).
35 Shiokawa, D. and Tanuma, S. DLAD, a novel mammalian divalent cation-independent endonuclease with homology to DNase II. *Nucleic acids research* **27**, 4083-4089 (1999).

36 Harosh, I., Binninger, D. M., Harris, P. V., Mezzina, M. and Boyd, J. B. Mechanism of action of deoxyribonuclease II from human lymphoblasts. *European journal of biochemistry / FEBS* **202**, 479-484 (1991).

37 Pheiffer, B. H. and Zimmerman, S. B. 3'-Phosphatase activity of the DNA kinase from rat liver. *Biochemical and biophysical research communications* **109**, 1297-1302 (1982).

38 Habraken, Y. and Verly, W. G. The DNA 3'-phosphatase and 5'-hydroxyl kinase of rat liver chromatin. *FEBS letters* **160**, 46-50 (1983).

39 Wiederhold, L., Leppard, J. B., Kedar, P., Karimi-Busheri, F., Rasouli-Nia, A., Weinfeld, M., Tomkinson, A. E., Izumi, T., Prasad, R., Wilson, S. H., Mitra, S. and Hazra, T. K. AP endonuclease-independent DNA base excision repair in human cells. *Molecular cell* **15**, 209-220, doi:10.1016/j.molcel.2004.06.003 (2004).

40 Whitehouse, C. J., Taylor, R. M., Thistlethwaite, A., Zhang, H., Karimi-Busheri, F., Lasko, D. D., Weinfeld, M. and Caldecott, K. W. XRCC1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair. *Cell* **104**, 107-117 (2001).

41 Chappell, C., Hanakahi, L. A., Karimi-Busheri, F., Weinfeld, M. and West, S. C. Involvement of human polynucleotide kinase in double-strand break repair by non-homologous end joining. *The EMBO journal* **21**, 2827-2832, doi:10.1093/emboj/21.11.2827 (2002).

42 Bernstein, N. K., Williams, R. S., Rakovszky, M. L., Cui, D., Green, R., Karimi-Busheri, F., Mani, R. S., Galicia, S., Koch, C. A., Cass, C. E., Durocher, D., Weinfeld, M. and Glover, J. N. The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase. *Molecular cell* **17**, 657-670, doi:10.1016/j.molcel.2005.02.012 (2005).

43 Loizou, J. I., El-Khamisy, S. F., Zlatanou, A., Moore, D. J., Chan, D. W., Qin, J., Sarno, S., Meggio, F., Pinna, L. A. and Caldecott, K. W. The protein kinase CK2 facilitates repair of chromosomal DNA single-strand breaks. *Cell* **117**, 17-28 (2004).

44 Koch, C. A., Agyei, R., Galicia, S., Metalnikov, P., O'Donnell, P., Starostine, A., Weinfeld, M. and Durocher, D. Xrcc4 physically links DNA end processing by polynucleotide kinase to DNA ligation by DNA ligase IV. *The EMBO journal* **23**, 3874-3885, doi:10.1038/sj.emboj.7600375 (2004).

45 Ali, A. A., Jukes, R. M., Pearl, L. H. and Oliver, A. W. Specific recognition of a multiply phosphorylated motif in the DNA repair scaffold XRCC1 by the FHA domain of human PNK. *Nucleic acids research* **37**, 1701-1712, doi:10.1093/nar/gkn1086 (2009).

46 Richardson, C. C. Phosphorylation of nucleic acid by an enzyme from T4 bacteriophage-infected Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **54**, 158-165 (1965).

47 Novogrodsky, A., Tal, M., Traub, A. and Hurwitz, J. The enzymatic phosphorylation of ribonucleic acid and deoxyribonucleic acid. II. Further properties of the 5'-hydroxyl polynucleotide kinase. *The Journal of biological chemistry* **241**, 2933-2943 (1966).

48 Novogrodsky, A. and Hurwitz, J. The enzymatic phosphorylation of ribonucleic acid and deoxyribonucleic acid. I. Phosphorylation at 5'-hydroxyl termini. *The Journal of biological chemistry* **241**, 2923-2932 (1966).

49 Berkner, K. L. and Folk, W. R. Polynucleotide kinase exchange as an assay for class II restriction endonucleases. *Methods in enzymology* **65**, 28-36 (1980).

50 Chaconas, G. and van de Sande, J. H. 5'-32P labeling of RNA and DNA restriction fragments. *Methods in enzymology* **65**, 75-85 (1980).

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

52 Wang, L. K., Lima, C. D. and Shuman, S. Structure and mechanism of T4 polynucleotide kinase: an RNA repair enzyme. *The EMBO journal* **21**, 3873-3880, doi:10.1093/emboj/cdf397 (2002).

53 Silber, R., Malathi, V. G. and Hurwitz, J. Purification and properties of bacteriophage T4-induced RNA ligase. *Proceedings of the National Academy of Sciences of the United States of America* **69**, 3009-3013 (1972).

54 Tyndall, C., Meister, J. and Bickle, T. A. The Escherichia coli prr region encodes a functional type IC DNA restriction system closely integrated with an anticodon nuclease gene. *Journal of molecular biology* **237**, 266-274 (1994).

55 Levitz, R., Chapman, D., Amitsur, M., Green, R., Snyder, L. and Kaufmann, G. The optional E. coli prr locus encodes a latent form of phage T4-induced anticodon nuclease. *The EMBO journal* **9**, 1383-1389 (1990).

56 Penner, M., Morad, I., Snyder, L. and Kaufmann, G. Phage T4-coded Stp: double-edged effector of coupled DNA and tRNA-restriction systems. *Journal of molecular biology* **249**, 857-868, doi:10.1006/jmbi.1995.0343 (1995).

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

58 Sirotkin, K., Cooley, W., Runnels, J. and Snyder, L. R. A role in true-late gene expression for the T4 bacteriophage 5' polynucleotide kinase 3' phosphatase. *Journal of molecular biology* **123**, 221-233 (1978).

59 David, M., Borasio, G. D. and Kaufmann, G. T4 bacteriophage-coded polynucleotide kinase and RNA ligase are involved in host tRNA alteration or repair. *Virology* **123**, 480-483 (1982).

60 Lillehaug, J. R., Kleppe, R. K. and Kleppe, K. Phosphorylation of double-stranded DNAs by T4 polynucleotide kinase. *Biochemistry* **15**, 1858-1865 (1976).

61 Lillehaug, J. R. and Kleppe, K. Kinetics and specificity of T4 polynucleotide kinase. *Biochemistry* **14**, 1221-1225 (1975).

62 van de Sande, J. H. and Bilsker, M. Phosphorylation of N-protected deoxyoligonucleotides by T4 polynucleotide kinase. *Biochemistry* **12**, 5056-5062 (1973).

63 Fontanel, M. L., Bazin, H. and Teoule, R. Sterical recognition by T4 polynucleotide kinase of non-nucleosidic moieties 5'-attached to oligonucleotides. *Nucleic acids research* **22**, 2022-2027 (1994).

64 Chan, C. M., Zhou, C. and Huang, R. H. Reconstituting bacterial RNA repair and modification in vitro. *Science* **326**, 247, doi:10.1126/science.1179480 (2009).

65 Pedersen, K., Zavialov, A. V., Pavlov, M. Y., Elf, J., Gerdes, K. and Ehrenberg, M. The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. *Cell* **112**, 131-140 (2003).

66 Neubauer, C., Gao, Y. G., Andersen, K. R., Dunham, C. M., Kelley, A. C., Hentschel, J., Gerdes, K., Ramakrishnan, V. and Brodersen, D. E. The structural basis for mRNA recognition and cleavage by the ribosome-dependent endonuclease RelE. *Cell* **139**, 1084-1095, doi:10.1016/j.cell.2009.11.015 (2009).

67 Winther, K. S. and Gerdes, K. Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 7403-7407, doi:10.1073/pnas.1019587108 (2011).

68 Wang, P., Chan, C. M., Christensen, D., Zhang, C., Selvadurai, K. and Huang, R. H. Molecular basis of bacterial protein Hen1 activating the ligase activity of bacterial protein Pnkp for RNA repair. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 13248-13253, doi:10.1073/pnas.1209805109 (2012).

69 Wang, L. K., Das, U., Smith, P. and Shuman, S. Structure and mechanism of the polynucleotide kinase component of the bacterial Pnkp-Hen1 RNA repair system. *RNA* **18**, 2277-2286, doi:10.1261/rna.036061.112 (2012).

70 Martins, A. and Shuman, S. An end-healing enzyme from Clostridium thermocellum with 5' kinase, 2',3' phosphatase, and adenylyltransferase activities. *RNA* **11**, 1271-1280, doi:10.1261/rna.2690505 (2005).

71 Popow, J., Schleiffer, A. and Martinez, J. Diversity and roles of (t)RNA ligases. *Cellular and molecular life sciences : CMLS* **69**, 2657-2670, doi:10.1007/s00018-012-0944-2 (2012).

72 Chan, P. P. and Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic acids research* **37**, D93-97, doi:10.1093/nar/gkn787 (2009).

73 Abelson, J., Trotta, C. R. and Li, H. tRNA splicing. *The Journal of biological chemistry* **273**, 12685-12688 (1998).

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

75 Sawaya, R., Schwer, B. and Shuman, S. Genetic and biochemical analysis of the functional domains of yeast tRNA ligase. *The Journal of biological chemistry* **278**, 43928-43938, doi:10.1074/jbc.M307839200 (2003).

76 Wang, L. K., Schwer, B., Englert, M., Beier, H. and Shuman, S. 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 research* **34**, 517-527, doi:10.1093/nar/gkj441 (2006).

77 Coller, J. and Parker, R. Eukaryotic mRNA decapping. *Annual review of biochemistry* **73**, 861-890, doi:10.1146/annurev.biochem.73.011303.074032 (2004).

78 Franks, T. M. and Lykke-Andersen, J. The control of mRNA decapping and P-body formation. *Molecular cell* **32**, 605-615, doi:10.1016/j.molcel.2008.11.001 (2008).

79 Houseley, J. and Tollervey, D. The many pathways of RNA degradation. *Cell* **136**, 763-776, doi:10.1016/j.cell.2009.01.019 (2009).

80 Parker, R. and Song, H. The enzymes and control of eukaryotic mRNA turnover. *Nature structural & molecular biology* **11**, 121-127, doi:10.1038/nsmb724 (2004).

81 Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. and Mello, C. C. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811, doi:10.1038/35888 (1998).

82 Ma, J. B., Yuan, Y. R., Meister, G., Pei, Y., Tuschl, T. and Patel, D. J. Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature* **434**, 666-670, doi:10.1038/nature03514 (2005).

83 Park, J. E., Heo, I., Tian, Y., Simanshu, D. K., Chang, H., Jee, D., Patel, D. J. and Kim, V. N. Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature* **475**, 201-205, doi:10.1038/nature10198 (2011).

84 Martinez, J. and Tuschl, T. RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes & development* **18**, 975-980, doi:10.1101/gad.1187904 (2004).

85 Shuman, S. and Hurwitz, J. 5'-Hydroxyl polyribonucleotide kinase from HeLa cell nuclei. Purification and properties. *The Journal of biological chemistry* **254**, 10396-10404 (1979).

86 Braglia, P., Heindl, K., Schleiffer, A., Martinez, J. and Proudfoot, N. J. Role of the RNA/DNA kinase Grc3 in transcription termination by RNA polymerase I. *EMBO reports* **11**, 758-764, doi:10.1038/embor.2010.130 (2010).

87 Heindl, K. and Martinez, J. Nol9 is a novel polynucleotide 5'-kinase involved in ribosomal RNA processing. *The EMBO journal* **29**, 4161-4171, doi:10.1038/emboj.2010.275 (2010).

88 Weitzer, S. and Martinez, J. The human RNA kinase hClp1 is active on 3' transfer RNA exons and short interfering RNAs. *Nature* **447**, 222-226, doi:10.1038/nature05777 (2007).

89 de Vries, H., Ruegsegger, U., Hubner, W., Friedlein, A., Langen, H. and Keller, W. Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *The EMBO journal* **19**, 5895-5904, doi:10.1093/emboj/19.21.5895 (2000).

90 Hanada, T., Weitzer, S., Mair, B., Bernreuther, C., Wainger, B. J., Ichida, J., Hanada, R., Orthofer, M., Cronin, S. J., Komnenovic, V., Minis, A., Sato, F., Mimata, H., Yoshimura, A., Tamir, I., Rainer, J., Kofler, R., Yaron, A., Eggan, K. C., Woolf, C. J., Glatzel, M., Herbst, R., Martinez, J. and Penninger, J. M. CLP1 links tRNA metabolism to progressive motor-neuron loss. *Nature* **495**, 474-480, doi:10.1038/nature11923 (2013).

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

92 Weitzer, S. and Martinez, J. hClp1: a novel kinase revitalizes RNA metabolism. *Cell Cycle* **6**, 2133-2137 (2007).

93 Kenski, D. M., Cooper, A. J., Li, J. J., Willingham, A. T., Haringsma, H. J., Young, T. A., Kuklin, N. A., Jones, J. J., Cancilla, M. T., McMasters, D. R., Mathur, M., Sachs, A. B. and Flanagan, W. M. Analysis of acyclic nucleoside modifications in siRNAs finds sensitivity at position 1 that is restored by 5'-terminal phosphorylation both in vitro and in vivo. *Nucleic acids research* **38**, 660-671, doi:10.1093/nar/gkp913 (2010).

94 Noble, C. G., Beuth, B. and Taylor, I. A. Structure of a nucleotide-bound Clp1-Pcf11 polyadenylation factor. *Nucleic acids research* **35**, 87-99, doi:10.1093/nar/gkl1010 (2007).

95 Jain, R. and Shuman, S. Characterization of a thermostable archaeal polynucleotide kinase homologous to human Clp1. *RNA* **15**, 923-931, doi:10.1261/rna.1492809 (2009).

96 Xing, D., Zhao, H. and Li, Q. Q. Arabidopsis CLP1-SIMILAR PROTEIN3, an ortholog of human polyadenylation factor CLP1, functions in gametophyte, embryo, and postembryonic development. *Plant physiology* **148**, 2059-2069, doi:10.1104/pp.108.129817 (2008).

97 Ramirez, A., Shuman, S. and Schwer, B. Human RNA 5'-kinase (hClp1) can function as a tRNA splicing enzyme in vivo. *RNA* **14**, 1737-1745, doi:10.1261/rna.1142908 (2008).

98 Ghazy, M. A., Gordon, J. M., Lee, S. D., Singh, B. N., Bohm, A., Hampsey, M. and Moore, C. The interaction of Pcf11 and Clp1 is needed for mRNA 3'-end formation and is modulated by amino acids in the ATP-binding site. *Nucleic acids research* **40**, 1214-1225, doi:10.1093/nar/gkr801 (2012).

99 Haddad, R., Maurice, F., Viphakone, N., Voisinet-Hakil, F., Fribourg, S. and Minvielle-Sebastia, L. An essential role for Clp1 in assembly of polyadenylation complex CF IA and Pol II transcription termination. *Nucleic acids research* **40**, 1226-1239, doi:10.1093/nar/gkr800 (2012).

100 Holbein, S., Scola, S., Loll, B., Dichtl, B. S., Hubner, W., Meinhart, A. and Dichtl, B. The P-loop domain of yeast Clp1 mediates interactions between CF IA and CPF factors in pre-mRNA 3' end formation. *PloS one* **6**, e29139, doi:10.1371/journal.pone.0029139 (2011).

101 Gross, S. and Moore, C. Five subunits are required for reconstitution of the cleavage and polyadenylation activities of Saccharomyces cerevisiae cleavage factor I. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 6080-6085, doi:10.1073/pnas.101046598 (2001).

102 Minvielle-Sebastia, L., Preker, P. J., Wiederkehr, T., Strahm, Y. and Keller, W. The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 7897-7902 (1997).

103 Proudfoot, N. New perspectives on connecting messenger RNA 3' end formation to transcription. *Current opinion in cell biology* **16**, 272-278, doi:10.1016/j.ceb.2004.03.007 (2004).

104 Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedea, E., Greenblatt, J. F. and Buratowski, S. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**, 517-522, doi:10.1038/nature03041 (2004).

105 Moore, C. L., Skolnik-David, H. and Sharp, P. A. Analysis of RNA cleavage at the adenovirus-2 L3 polyadenylation site. *The EMBO journal* **5**, 1929-1938 (1986).

106 Sheets, M. D., Stephenson, P. and Wickens, M. P. Products of in vitro cleavage and polyadenylation of simian virus 40 late pre-mRNAs. *Molecular and cellular biology* **7**, 1518-1529 (1987).

107 Popow, J., Englert, M., Weitzer, S., Schleiffer, A., Mierzwa, B., Mechtler, K., Trowitzsch, S., Will, C. L., Luhrmann, R., Soll, D. and Martinez, J. HSPC117 is the essential subunit of a human tRNA splicing ligase complex. *Science* **331**, 760-764, doi:10.1126/science.1197847 (2011).

108 Kurihara, T., Fowler, A. V. and Takahashi, Y. cDNA cloning and amino acid sequence of bovine brain 2',3'-cyclic-nucleotide 3'-phosphodiesterase. *The Journal of biological chemistry* **262**, 3256-3261 (1987).

109 Spinelli, S. L., Malik, H. S., Consaul, S. A. and Phizicky, E. M. A functional homolog of a yeast tRNA splicing enzyme is conserved in higher eukaryotes and in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 14136-14141 (1998).

110 Englert, M., Sheppard, K., Gundllapalli, S., Beier, H. and Soll, D. Branchiostoma floridae has separate healing and sealing enzymes for 5'-phosphate RNA ligation. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 16834-16839, doi:10.1073/pnas.1011703107 (2010).

111 Hammond, S. M., Bernstein, E., Beach, D. and Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**, 293-296, doi:10.1038/35005107 (2000).

112 Nykanen, A., Haley, B. and Zamore, P. D. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **107**, 309-321 (2001).

113 Saraste, M., Sibbald, P. R. and Wittinghofer, A. The P-loop--a common motif in ATP- and GTP-binding proteins. *Trends in biochemical sciences* **15**, 430-434 (1990).

114 Walker, J. E., Saraste, M., Runswick, M. J. and Gay, N. J. 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. *The EMBO journal* **1**, 945-951 (1982).

115 Castle, C. D., Sardana, R., Dandekar, V., Borgianini, V., Johnson, A. W. and Denicourt, C. Las1 interacts with Grc3 polynucleotide kinase and is required for ribosome synthesis in Saccharomyces cerevisiae. *Nucleic acids research* **41**, 1135-1150, doi:10.1093/nar/gks1086 (2013).
116 Kitano, E., Hayashi, A., Kanai, D., Shinmyozu, K. and Nakayama, J. Roles of fission yeast Grc3 protein in ribosomal RNA processing and heterochromatic gene silencing. *The Journal of biological chemistry* **286**, 15391-15402, doi:10.1074/jbc.M110.201343 (2011).
117 Knowles, J. R. Enzyme-catalyzed phosphoryl transfer reactions. *Annual review of biochemistry* **49**, 877-919, doi:10.1146/annurev.bi.49.070180.004305 (1980).
118 Schlichting, I. and Reinstein, J. Structures of active conformations of UMP kinase from Dictyostelium discoideum suggest phosphoryl transfer is associative. *Biochemistry* **36**, 9290-9296, doi:10.1021/bi970974c (1997).
119 Wittinghofer, A. Phosphoryl transfer in Ras proteins, conclusive or elusive? *Trends in biochemical sciences* **31**, 20-23, doi:10.1016/j.tibs.2005.11.012 (2006).
120 Lassila, J. K., Zalatan, J. G. and Herschlag, D. Biological phosphoryl-transfer reactions: understanding mechanism and catalysis. *Annual review of biochemistry* **80**, 669-702, doi:10.1146/annurev-biochem-060409-092741 (2011).
121 Wittinghofer, A. Signaling mechanistics: aluminum fluoride for molecule of the year. *Current biology : CB* **7**, R682-685 (1997).
122 Antonny, B. and Chabre, M. Characterization of the aluminum and beryllium fluoride species which activate transducin. Analysis of the binding and dissociation kinetics. *The Journal of biological chemistry* **267**, 6710-6718 (1992).
123 Goodno, C. C. and Taylor, E. W. Inhibition of actomyosin ATPase by vanadate. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 21-25 (1982).
124 Matte, A., Tari, L. W. and Delbaere, L. T. How do kinases transfer phosphoryl groups? *Structure* **6**, 413-419 (1998).
125 Garces, F., Pearl, L. H. and Oliver, A. W. The structural basis for substrate recognition by mammalian polynucleotide kinase 3' phosphatase. *Molecular cell* **44**, 385-396, doi:10.1016/j.molcel.2011.08.036 (2011).
126 Eastberg, J. H., Pelletier, J. and Stoddard, B. L. Recognition of DNA substrates by T4 bacteriophage polynucleotide kinase. *Nucleic acids research* **32**, 653-660, doi:10.1093/nar/gkh212 (2004).
127 Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. and Clardy, J. Atomic structure of FKBP-FK506, an immunophilin-immunosuppressant complex. *Science* **252**, 839-842 (1991).
128 Taylor, R. G., Walker, D. C. and McInnes, R. R. E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic acids research* **21**, 1677-1678 (1993).
129 Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *Journal of molecular biology* **166**, 557-580 (1983).
130 Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491 (1988).
131 Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685 (1970).
132 Lienhard, G. E. and Secemski, II. P 1 ,P 5 -Di(adenosine-5')pentaphosphate, a potent multisubstrate inhibitor of adenylate kinase. *The Journal of biological chemistry* **248**, 1121-1123 (1973).

133 Kabsch, W. Xds. *Acta crystallographica. Section D, Biological crystallography* **66**, 125-132, doi:10.1107/S0907444909047337 (2010).
134 Sheldrick, G. M. A short history of SHELX. *Acta crystallographica. Section A, Foundations of crystallography* **64**, 112-122, doi:10.1107/S0108767307043930 (2008).
135 Emsley, P. and Cowtan, K. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
136 Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. and Warren, G. L. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta crystallographica. Section D, Biological crystallography* **54**, 905-921 (1998).
137 Murshudov, G. N., Vagin, A. A. and Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica. Section D, Biological crystallography* **53**, 240-255, doi:10.1107/S0907444996012255 (1997).
138 Winn, M. D., Isupov, M. N. and Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta crystallographica. Section D, Biological crystallography* **57**, 122-133 (2001).
139 Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. and Richardson, D. C. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **50**, 437-450, doi:10.1002/prot.10286 (2003).
140 Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, 283-291 (1993).
141 DeLano, W. L. The PyMOL Molecular Graphics System. *Palo Alto, CA, USA.* (2002).
142 Baker, N. A., Sept, D., Joseph, S., Holst, M. J. and McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 10037-10041, doi:10.1073/pnas.181342398 (2001).
143 Barton, G. J. ALSCRIPT: a tool to format multiple sequence alignments. *Protein engineering* **6**, 37-40 (1993).
144 Livingstone, C. D. and Barton, G. J. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Computer applications in the biosciences : CABIOS* **9**, 745-756 (1993).
145 Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta crystallographica. Section D, Biological crystallography* **A32**, 922–923 (1997).
146 Schlichting, I. and Reinstein, J. pH influences fluoride coordination number of the AlFx phosphoryl transfer transition state analog. *Nature structural biology* **6**, 721-723, doi:10.1038/11485 (1999).
147 Baxter, N. J., Blackburn, G. M., Marston, J. P., Hounslow, A. M., Cliff, M. J., Bermel, W., Williams, N. H., Hollfelder, F., Wemmer, D. E. and Waltho, J. P. Anionic charge is prioritized over geometry in aluminum and magnesium fluoride transition state analogs of phosphoryl transfer enzymes. *Journal of the American Chemical Society* **130**, 3952-3958, doi:10.1021/ja078000n (2008).
148 Hendrickson, W. A., Smith, J. L. and Sheriff, S. Direct phase determination based on anomalous scattering. *Methods in enzymology* **115**, 41-55 (1985).
149 Schneider, T. R. and Sheldrick, G. M. Substructure solution with SHELXD. *Acta crystallographica. Section D, Biological crystallography* **58**, 1772-1779 (2002).

150 Pape, T. and Schneider, T. R. HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Cryst.* **37**, 843-844 (2004).
151 Rao, S. T. and Rossmann, M. G. Comparison of super-secondary structures in proteins. *Journal of molecular biology* **76**, 241-256 (1973).
152 Holm, L. and Sander, C. Dali/FSSP classification of three-dimensional protein folds. *Nucleic acids research* **25**, 231-234 (1997).
153 Gawronski-Salerno, J. and Freymann, D. M. Structure of the GMPPNP-stabilized NG domain complex of the SRP GTPases Ffh and FtsY. *Journal of structural biology* **158**, 122-128, doi:10.1016/j.jsb.2006.10.025 (2007).
154 Cohen, R. L., Espelin, C. W., De Wulf, P., Sorger, P. K., Harrison, S. C. and Simons, K. T. Structural and functional dissection of Mif2p, a conserved DNA-binding kinetochore protein. *Molecular biology of the cell* **19**, 4480-4491, doi:10.1091/mbc.E08-03-0297 (2008).
155 Arsenieva, D., Symersky, J., Wang, Y., Pagadala, V. and Mueller, D. M. Crystal structures of mutant forms of the yeast F1 ATPase reveal two modes of uncoupling. *The Journal of biological chemistry* **285**, 36561-36569, doi:10.1074/jbc.M110.174383 (2010).
156 Berkner, K. L. and Folk, W. R. Polynucleotide kinase exchange reaction: quantitave assay for restriction endonuclease-generated 5'-phosphoroyl termini in DNA. *The Journal of biological chemistry* **252**, 3176-3184 (1977).
157 van de Sande, J. H., Kleppe, K. and Khorana, H. G. Reversal of bacteriophage T4 induced polynucleotide kinase action. *Biochemistry* **12**, 5050-5055 (1973).
158 Karimi-Busheri, F., Lee, J., Tomkinson, A. E. and Weinfeld, M. Repair of DNA strand gaps and nicks containing 3'-phosphate and 5'-hydroxyl termini by purified mammalian enzymes. *Nucleic acids research* **26**, 4395-4400 (1998).
159 Bahadur, R. P., Zacharias, M. and Janin, J. Dissecting protein-RNA recognition sites. *Nucleic acids research* **36**, 2705-2716, doi:10.1093/nar/gkn102 (2008).
160 Rusnak, P., Haney, P. and Konisky, J. The adenylate kinases from a mesophilic and three thermophilic methanogenic members of the Archaea. *Journal of bacteriology* **177**, 2977-2981 (1995).
161 Kath, T., Schmid, R. and Schafer, G. Identification, cloning, and expression of the gene for adenylate kinase from the thermoacidophilic archaebacterium Sulfolobus acidocaldarius. *Archives of biochemistry and biophysics* **307**, 405-410, doi:10.1006/abbi.1993.1607 (1993).
162 Lacher, K. and Schafer, G. Archaebacterial adenylate kinase from the thermoacidophile Sulfolobus acidocaldarius: purification, characterization, and partial sequence. *Archives of biochemistry and biophysics* **302**, 391-397, doi:10.1006/abbi.1993.1229 (1993).
163 Jha, S., Karnani, N., Dhar, S. K., Mukhopadhayay, K., Shukla, S., Saini, P., Mukhopadhayay, G. and Prasad, R. Purification and characterization of the N-terminal nucleotide binding domain of an ABC drug transporter of Candida albicans: uncommon cysteine 193 of Walker A is critical for ATP hydrolysis. *Biochemistry* **42**, 10822-10832, doi:10.1021/bi0345900 (2003).
164 Wang, L. K. and Shuman, S. Domain structure and mutational analysis of T4 polynucleotide kinase. *The Journal of biological chemistry* **276**, 26868-26874, doi:10.1074/jbc.M103663200 (2001).
165 Wang, L. K. and Shuman, S. Mutational analysis defines the 5'-kinase and 3'-phosphatase active sites of T4 polynucleotide kinase. *Nucleic acids research* **30**, 1073-1080 (2002).

166 Vale, R. D. Switches, latches, and amplifiers: common themes of G proteins and molecular motors. *The Journal of cell biology* **135**, 291-302 (1996).
167 McCammon, J. A. and Northrup, S. H. Gated binding of ligands to proteins. *Nature* **293**, 316-317 (1981).
168 Gora, A., Brezovsky, J. and Damborsky, J. Gates of Enzymes. *Chemical reviews*, doi:10.1021/cr300384w (2013).

6 Appendix

6.1 Appendix of the Materials and Methods section

6.1.1 List of primers

Cloning Primers	
Primer name	**Sequence [5'→3']**
ceClp1-pET21a-For	CGA AGC ATC ATA **TG**A GCG AGG AGA ATG TTC
ceClp1-pET21a-Rev	GAG TGC GGC CGC TCG TTT TAT TTG ATC AT
ceClp1-pET28b-For	CG AAG CAT CAT **ATG** AGC GAG GAG AAT GTT C
ceClp1-pET28b-Rev	TGC GGC CGC **TCA** TCG TTT TAT TTG ATC ATC
ceClp1ΔN314-pET21a_For	GCA TTA GCC ATA **TG**C CAT TCA CAT TTG ACG
Quikchange Primers	
Primer name	**Sequence [5'→3']**
ceClp1ΔN107-pET21a(ΔHis)-For	ATC TTG ATG ATC AAA TAA AAC GA**T AG**C TTG CGG CCG C
ceClp1ΔN107-pET21a(ΔHis)-Rev	GCG GCC GCA AG**C TA**T CGT TTT ATT TGA TCA TCA AGA T
ceClp1ΔC104-pET28b-For	GAA GAA GAG AGA GGA ACA G**TA GG**C TGG AAA CTC GAA TAA G
ceClp1ΔC104-pET28b-Rev	CTT ATT CGA GTT TCC AGC **CTA** CTG TTC CTC TCT CTT CTT C
ceClp1ΔC310-pET28b-For	TCT ACG GAA CCC GTG CC**T AG**A ATC TCT ACC CAT TCA C
ceClp1ΔC310-pET28b-Rev	GTG AAT GGG **TAG** AGA TTC TAG GCA CGG GTT CCG TAG A
ceClp1K127A-pET21a-For	ACC AAC GGA CGT CGG A**GC A**AC CAC AGT CTC C

ceClp1K127A-pET21a-Rev	GGA GAC TGT GGT **TGC** TCC GAC GTC CGT TGG T
ceClp1K127R-pET28b-For	ACC AAC GGA CGT CGG A**AG A**AC CAC AGT CTC C
ceClp1K127R-pET28b-Rev	GGA GAC TGT GGT **TCT** TCC GAC GTC CGT TGG T

6.1.2 Oligonucleotide sequences.

oligonucleotide	sequence
Single stranded RNA	5’-GCG AGA CAG UGU GAC UUU GG-3’
Double stranded RNA with blunt ends	5’-GCG AGA CAG UGU GAC UUU GG-3’ 5’-CCA AAG UCA CAC UGU CUC GC-3’
Double stranded RNA with a single stranded 3’-overhang	5’-GAG ACA GUG UGA CUU UGG AC-3’ 5’-CCA AAG UCA CAC UGU CUC GC-3’
Single stranded DNA	5’-GCG AGA CAG TGT GAC TTT GG-3’
Double stranded DNA with blunt ends	5’-GCG AGA CAG TGT GAC TTT GG-3’ 5’-CCA AAG TCA CAC TGT CTC GC-3’
Double stranded DNA with a single stranded 3’-overhang	5’-GAG ACA GTG TGA CTT TGG AC-3’ 5’-CCA AAG TCA CAC TGT CTC GC-3’

6.2 Additional Tables

6.2.1 Proteins showing a conventional Walker A lysine residue.

Nr.	PDB ID	Protein name	Resolution [Å]	P-loop Sequence	P-loop residue number and chain	Ligand	Walker A Lys interactions
Conventional Walker A lysine: Substrate bound state structures							
1	3RWM	Ypt32	2.00	GDSGVGKS	20-27 (A)	GNP	β- and γ-phosphate
2	4GP7	Pnkp-Hen1 RNA repair system	2.00	GSSGSGKS	15-22 (A)	ATP	β- and γ-phosphate
3	3QF7	Mre11:Rad50 complex	1.90	GPNGAGKS	30-37 (A)	ANP	β- and γ-phosphate
4	3A4L	O-phosphoseryl-tRNA(Sec) kinase	1.80	GLPGVGKS	11-18 (A)	ANP	β- and γ-phosphate
5	2IYW	Shikimate Kinase	1.85	GLPGSGKS	9-16 (A)	ATP	β- and γ-phosphate
6	2D7C	Rab11 in complex with FIP3 Rab-binding domain	1.75	GDSGVGKS	18-25 (A)	GTP	β- and γ-phosphate
7	2CBZ	Multidrug resistance protein 1 nucleotide binding domain 1	1.50	GQVGCGKS	678-685 (A)	ATP	β- and γ-phosphate
8	1YZL	Rab9 GTPase	1.85	GDGGVGKS	14-21 (A)	GNP	β- and γ-phosphate
9	3RLF	Maltose-binding protein/maltose transporter complex	2.20	GPSGCGKS	36-43 (A)	ANP	β- and γ-phosphate
10	3GPL	RecD2	2.50	GGPGTGKS	360-367 (A)	ANP	β- and γ-phosphate
11	4A6X	RadA C-terminal ATPase domain	1.55	GEFGSGKT	138-145 (A)	ATP	β- and γ-phosphate
12	2GCO	RhoC	1.40	GDGACGKT	12-129 (A)	GNP	β- and γ-phosphate
13	3MYK	Myosin	1.84	GESGAGKT	179-186 (X)	ANP	β- and γ-phosphate
14	3FVQ	Nucleotide binding domain of FbpC	1.90	GASGCGKT	37-44 (A)	ATP	β- and γ-phosphate
15	2QT0	Nicotinamide riboside kinase 1	1.92	GVTNSGKT	10-17 (A)	ANP	β- and γ-phosphate
16	2OLR	Phosphoenolpyruvate carboxykinase	1.60	GLSGTGKT	248-255 (A)	ATP	β- and γ-phosphate

17	1OXV	GlcV, the ABC-ATPase of the glucose ABC transporter	1.95	GPSGAGKT	38-45 (A)	ANP	β- and γ-phosphate
18	3Q9L	MinD	2.34	GKGGVGKT	10-17 (A)	ATP	β- and γ-phosphate
19	3HQD	Kinesin Eg5 motor domain	2.00	GQTGTGKT	105-112 (A)	ANP	β- and γ-phosphate
20	3EW9	RADA recombinase	2.40	GMFGSGKT	105-112 (A)	ANP	β- and γ-phosphate

Conventional Walker A lysine: Transition state analogue structures

1	3SR0	Adenylate kinase	1.57	GPPGAGKG	7-14 (A)	AlF_4^-	β- and γ-phosphate
2	5UKD	UMP/CMP kinase	1.90	GGPGSGKG	13-20 (A)	$Al\ F_3$	β- and γ-phosphate
3	1E2E	Thymidylate kinase	2.00	GVDRAGKS	13-20 (A)	$Al\ F_3$	β- and γ-phosphate
4	1WQ1	RAS-RASGAP complex	2.50	GAGGVGKS	10-17 (R)	$Al\ F_3$	β- and γ-phosphate
5	3T34	Dynamin-related protein 1A (AtDRP1A)	2.40	GGQSSGKS	41-48 (A)	AlF_4^-	β- and γ-phosphate
6	3PUW	MBP-Maltose transporter	2.30	GPSGCGKS	36-43 (A)	AlF_4^-	β- and γ-phosphate
7	2GJ9	MnmE G-domain	2.00	GRPNAGKS	223-230 (A)	AlF_4^-	β- and γ-phosphate
8	3MSX	RhoA with GAP domain of ArhGAP20	1.65	GDGACGKT	12-19 (A)	$Al\ F_3$	β- and γ-phosphate
9	2WOJ	GET3	2.00	GKGGVGKT	25-32 (A)	AlF_4^-	β- and γ-phosphate
10	3T12	MglA in complex with MglB	2.20	GPGLSGKT	19-26 (A)	AlF_4^-	β- and γ-phosphate
11	2XZO	UPF1 helicase	2.39	GPPGTGKT	492-499 (A)	AlF_4^-	β- and γ-phosphate
12	1IHU	Aresenite translocating ATPase	2.15	GKGGVGKT	324-341 (A)	$Al\ F_3$	β- and γ-phosphate
13	4ANJ	Myosin VI	2.60	GESGAGKT	151-158 (A)	AlF_4^-	β- and γ-phosphate
14	3BH7	RP2-Arl3 complex	1.90	GLDNAGKT	24-31 (A)	AlF_4^-	β- and γ-phosphate
15	1W9I	Myosin II	1.75	GESGAGKT	179-186 (A)	$Be\ F_3$	β- and γ-phosphate
16	2G77	Gyp1 TBC domain in complex with Rab33 GTPase	2.26	GDSNVGKT	40-47 (B)	$Al\ F_3$	β- and γ-phosphate

17	1W0J	F1-ATPASE	2.20	GGAGVGKT	155-162 (D)	Be F_3	β- and γ-phosphate
18	3SS8	NFeoB	2.51	GNPNSGKT	8-15 (A)	AlF_4^-	β- and γ-phosphate
19	2GTP	RGS1 in complex with the activated Gi alpha 1	2.55	GAGESGKS	40-47 (A)	AlF_4^-	β- and γ-phosphate
20	1GRN	CDC42-GAP complex	2.10	GDGAVGK	10-17 (A)	Al F_3	β- and γ-phosphate

Conventional Walker A lysine: Product bound state structures

1	4GP6	Pnkp-Hen1 RNA repair system	2.1	GSSGSGKS	15-22 (A)	ADP	β- and γ-phosphate
2	3ZVN	Polynucleotide kinase 3'-phosphatase	2.15	GFPGAGKS	371-378 (A)	APD	β- and γ-phosphate
3	4EDH	Thymidylate kinase	1.3	GPEGAGKS	10-17 (A)	APD	β- and γ-phosphate
4	3A4M	O-phosphoseryl-tRNA(Sec) kinase	1.8	GLPGVGKS	11-18 (A)	ADP	β- and γ-phosphate
5	3DC4	Kinesin family member NOD	1.9	GQTGTGKS	87-94 (A)	ADP	β- and γ-phosphate
6	2ZFI	Kif1A Motor Domain	1.55	GQTGAGKS	96-130 (A)	ADP	β- and γ-phosphate
7	2IYQ	Shikimate kinase	1.8	GLPGSGKS	9-16 (A)	ADP	β- and γ-phosphate
8	2DPX	Rad GTPase	1.8	GAPGVGKS	98-105 (A)	GDP	β- and γ-phosphate
9	2H58	KIFC3 motor domain	1.85	GSERVGKS	528-535 (A)	ADP	β- and γ-phosphate
10	1UJ2	Uridine-cytidine kinase 2	1.8	GGTASGKS	27-34 (A)	ADP	β- and γ-phosphate
11	1YQT	RNase-L Inhibitor	1.9	GPNGIGKT	371-378 (A)	ADP	β- and γ-phosphate
12	4A7O	RadA C-terminal ATPase domain	1.88	GEFGSGKT	138-145 (A)	ADP	β- and γ-phosphate
13	4A14	KIF7	1.6	GQTGSGKT	91-98 (A)	ADP	β- and γ-phosphate
14	3REF	EhRho1	1.95	GDGAVGKT	27-34 (A)	GDP	β- and γ-phosphate
15	3FYH	Recombinase	1.9	GVFGSGKT	105-112 (A)	ADP	β- and γ-phosphate
16	2QSY	Nicotinamide riboside kinase 1	1.95	GVTNSGKT	10-17 (A)	ADP	β- and γ-phosphate

17	1XMV	RecA	1.9	GPESSGKT	66-73 (A)	ADP	β- and γ-phosphate
18	1SVL	SV40 large T antigen helicase domain	1.95	GPIDSGKT	426-433 (A)	ADP	β- and γ-phosphate
19	1F6B	SAR1	1.7	GLDNAGKT	32-39 (A)	GDP	β- and γ-phosphate
20	2UKD	UMP/CMP kinase	2.2	GGPGSGKG	12-19 (A)	ADP	β- and γ-phosphate

6.3 Abbreviations

ADP	adenosine-5’-diphosphate
ATP	adenosine-5’-triphosphate
A.U.	arbitrary units
AppNHp	adenyl-5’-(β,γ-imido)triphosphate
BSA	bovine serum albumin
CD	circular dichroism
C-terminus	carboxy terminus
CtD	C-terminal domain
DNA	deoxyribonucleic acid
DTE	dithioerythritol
EDTA	ethylenediaminetetraacetic acid
G_1C_2	guanylyl(3′→5′)cytidine
GTP	guanosine-5’-triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	lysogeny broth medium
mRNA	messenger ribonucleic acid
NADH	nicotinamide adenine dinucleotide
N-terminus	amino terminus
NtD	N-terminal domain
ORF	open reading frame

PCR	polymerase chain reaction
PDB	protein database
PEP	phosphoenolpyruvate
PNK	Polynucleotide kinase
RNA	ribonucleic acid
RPNK	RNA-specific PNK
rRNA	ribosomal RNA
TCEP	(tris(2-carboxyethyl)phosphine)
TEMED	tetramethylethylenediamine
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
U_1G_2	urylyl-(3'→5')guanosine
UV	ultra violet light
% w/v	mass/volume percentage
% v/v	volume/volume percentage