



**UiT** The Arctic University of Norway

Faculty of Biosciences, Fisheries and Economics

## **Biodiscovery of Novel RNA Ligases with Biotechnological Potential**

New Insights into the RNA Ligase 3 Family

Hannah Ildikó Johnstad Gábor

BIO-3901 Master's thesis in Marine Biotechnology, August 2020

|

## Acknowledgements

The work presented in this thesis was performed as a collaboration between UiT and ArcticZymes AS, Norway. Being able to apply the knowledge gained from university courses in an industry setting has been an extremely meaningful experience and I am very grateful for this opportunity. I would like to thank my main supervisors Klara Stensvåg and Olav Lanes for making this possible. Furthermore, I want to thank my co-supervisors Bernd Ketelsen Striberny at ArcticZymes for his continuous encouragements and guidance, and Hans-Matti Blencke at UiT for his valuable feedback on my writing. For helping me in the lab I also want to thank Terese Solstad and Sigurd Eidem Gundesø. It has been a pleasure to pursue my thesis at ArcticZymes due to the inclusive work environment and great colleagues.

Finally, I want to thank my friends in Trondheim, Oslo and Tromsø for making the past six years memorable and sticking through.

Tromsø, August 2020

Hannah Ildikó Johnstad Gábor

## Abstract

Polynucleotide ligases constitute a class of enzymes which catalyse the formation of phosphodiester bonds between proximal 5'-PO<sub>4</sub> and 3'-OH groups in DNA, RNA and DNA/RNA-hybrid substrates. The reaction proceeds through three steps, where steps 1 and 2 prepare the substrate for ligation in step 3. The focus of this thesis is on ssRNA ligases of the RNA ligase (Rnl) 3 family. RNA ligases can repair breaks in oligonucleotides and create entirely new RNA molecules. They have biotechnological potential in applications ranging from diagnostics to preparation of sequencing libraries from RNA samples. One member of the Rnl3 family has already been commercialized for the latter, as well as for producing adenylated substrate for preparative purposes.

This thesis applies a biodiscovery pipeline to investigate the potential of 22 putative RNA ligases as new enzyme leads. Potential candidates were selected based on their expression in *Escherichia coli*, activity in the first two steps of the reaction, success of upscaling, and novelty. Selected candidates were purified using fast protein liquid chromatography (FPLC) and activity was further characterized. Activity was assessed after the first two reaction steps, after step 3 in isolation, and throughout the complete ligation reaction. For the latter two, new protocols were developed as part of this thesis. These did not demand additional equipment and were less laborious than the existing protocols on which they were based.

Four novel, thermotolerant candidates were found to adenylate ssDNA but surprisingly none exhibited ligation activity in the conditions tested here. Substrate adenylation could not be improved by increasing adenosine triphosphate (ATP) and enzyme concentrations. Indications of that adenylation by one candidate (X07) was increased at higher temperatures (80°C) were found but require further studies. Another hit (X06) consistently showed adenylation activity up to 90%, however due to poor reproducibility this needs further investigation. Knowledge gained from this thesis should be used to improve the purification protocol and thereby sample quality. Dimerization is thought to be important for ligation, and we believe means to study dimerization should be a priority in future work.

# Table of Contents

Abbreviations .....	v
1 Introduction .....	1
1.1 Polynucleotide Ligases .....	1
1.2 Bioprospecting of Ligases from Extremophiles .....	11
1.3 Recombinant DNA Technology and Protein Purification .....	13
2 Purpose and Specific Aims .....	18
3 Materials and Methods .....	19
3.1 Workflow of the Study .....	19
3.2 Media and Solutions .....	21
3.3 Protein Expression in <i>E. coli</i> .....	27
3.4 Chromatographical Purification of Candidates .....	28
3.5 Analytical Methods.....	31
3.6 Enzyme Characterization.....	34
3.7 Detection of Specific Contaminants .....	38
3.8 Similarity Matrix and Phylogenetic Analysis.....	39
4 Results .....	40
4.1 Initial screening of Candidate Polynucleotide Ligase Expression in <i>E. coli</i> .....	40
4.2 Optimizing the Soluble-to-Insoluble Ratio.....	43
4.3 Initial Activity Screening Using the Adenylyltransferase Assay .....	45
4.4 Purifying Active Candidates from Up-Scaled Cultures .....	48
4.5 Confirming Adenylyltransferase Activity in Upscaled and Purified Batches .....	60
4.6 Assessing Ligation Activity in Upscaled and Purified Batches .....	64
4.7 Testing the Usability of TSA with Rnl3 Family Ligases .....	70
4.8 Characterizing Adenylation of ssDNA.....	71
4.9 Testing Standard Assays for Phosphatase and Nuclease Analysis.....	77

4.10	Bioinformatics .....	81
5	Discussion .....	83
5.1	Screening Expression of Archaeal RNA Ligases in <i>E. coli</i> .....	83
5.2	Evaluation of Functional Screening of Ligase Candidates.....	85
5.3	Activities of the Four Most Promising Candidates .....	87
5.4	Chromatographical Purification of Rnl3-Family Ligases .....	91
5.5	Discussion of Analytical Methods: TSA, Nuclease and Phosphatase Assay .....	94
6	Conclusion.....	97
7	Further Work.....	98
8	References .....	99
9	Appendices.....	I
	Appendix A. Protein Data .....	I
	Appendix B. Complete Results of Initial Expression Screening.....	III
	Appendix C. Complete Results of Initial ssDNA Adenylation Screening.....	V
	Appendix D. Phosphatase Assay Graphs and Numbers .....	VII
	Appendix E. Thermal Shift Assay Graphs.....	X
	Appendix F. Identity Matrix of Rnl 3 Candidates .....	XI

## Abbreviations

---

<b>6FAM</b>	6-Carboxyfluorescein
<b>App-donor</b>	Adenylated donor
<b>ATP</b>	Adenosine triphosphate
<b>CV</b>	Column volume
<b>ddC</b>	Dideoxycytosine
<b>FPLC</b>	Fast protein liquid chromatography
<b>IEX</b>	Ion-Exchange Chromatography
<b>IMAC</b>	Immobilized Metal Affinity Chromatography
<b>kDa</b>	kilo Dalton
<b>MQ</b>	Milli-Q Ultrapure Water
<b>MthRnl</b>	<i>Methanobacterium thermoautotrophicum</i> RNA ligase
<b>NEB</b>	New England Biolabs
<b>nt</b>	Nucleotide
<b>OD<sub>600</sub></b>	Optical density at 600 nm
<b>ON</b>	Overnight
<b>Q</b>	Quaternary ammonium
<b>Rnl</b>	RNA ligase
<b>rSAP</b>	Recombinant shrimp alkaline phosphatase
<b>S/E</b>	Substrate-to-enzyme ratio

---

---

<b>SEC</b>	Size-Exclusion Chromatography
<b>SN</b>	Supernatant
<b>SP</b>	Sulphopropyl
<b>TSA</b>	Thermal shift assay
<b>WT</b>	Wild type

---

# 1 Introduction

Polynucleotide ligases are an essential class of enzymes crucial at important stages in the life cycle of cells and indispensable as tools in molecular biology research and diagnostics. Ligases of archaeal origin are still relatively unexplored and bioprospecting is expected to reveal enzymes with novel and useful characteristics (Chambers and Patrick, 2015). The following chapter will provide an introduction to polynucleotide ligases, bioprospecting, recombinant production of proteins, and the principles behind common chromatographical methods for protein purification.

## 1.1 Polynucleotide Ligases

Polynucleotide ligases are enzymes that catalyse the formation of intra- or intermolecular phosphodiester bonds between adjacent 3'-hydroxyl (OH) and 5'-phosphoryl (PO<sub>4</sub>) ends of oligodeoxynucleic (DNA) or ribonucleic (RNA) acid molecules (Pascal, 2008). These ligases make up the superfamily of covalent nucleotidyl transferases (Shuman and Lima, 2004).

All members of this family share a common three-step reaction mechanism and are dependent on nucleotide cofactors to activate the enzyme (Pascal, 2008). DNA ligases are found to depend either on adenosine triphosphate (ATP) or nicotinamide adenine dinucleotide, while only ATP-dependent RNA ligases have been described to date (Shuman and Lima, 2004). There exist also DNA ligases (*Aeropyrum pernix* DNA ligase) that can utilize adenosine diphosphate in addition to ATP, although this renders ligation less energetically favourable and hence is not sufficient to activate most polynucleotide ligases (Jeon and Ishikawa, 2003).

### ***In Vivo* Functions of Polynucleotide Ligases**

Polynucleotide ligases are found in organisms of all phylogenetic domains and function in a variety of pathways in DNA and RNA metabolism (Pascal, 2008). DNA ligases are essential for DNA replication and maintenance, as they fuse Okazaki fragments and are involved in repair pathways (Timson, Singleton and Wigley, 2000).

RNA ligases function in repair systems by fusing ssRNA breaks caused by noxious agents (Smith *et al.*, 2012). They also ligate purposeful breaks caused by endonucleases during tRNA and rRNA processing, and act as counter-defense mechanisms (Ho and Shuman, 2002;

Tanaka, Meineke and Shuman, 2011; Smith *et al.*, 2012). Circularization of various RNA substrates is another function that is gaining more attention recently partly through the ability of circular RNA (circRNA) to bind miRNAs (miRNA sponging) (Petkovic and Müller, 2015; Becker, L'Hermitte-Stead and Myllykallio, 2019). By being resistant to exonuclease degradation, circRNAs are more stable than their linear counterparts. Subject to circularization are excised introns from pre-tRNA transcripts, rRNA processing intermediates, and noncoding RNAs. RNA circularization is widespread in the archaeal transcriptomes sequenced to date and is explained by presence of members of an archaeal-specific class of RNA ligases. The biological functions of circRNAs are mostly not well understood, but may include guiding modifications of other RNAs (Becker, L'Hermitte-Stead and Myllykallio, 2019).

### **Reaction Mechanisms of Polynucleotide Ligases**

Unifying members of the nucleotidyl transferase superfamily is the three-step mechanism by which the nucleotide cofactor is used to generate the energy required to form phosphodiester bonds (Figure 1.1 *panel A*) (Pascal, 2008). In the first step, a conserved active site catalytic lysine (N $\xi$ ) attacks the  $\alpha$ -phosphate of the nucleotide cofactor, thereby cleaving the  $\alpha$ ,  $\beta$ -pyrophosphate bond and releasing pyrophosphate. In the second step, the ligase-adenosine monophosphate (AMP) intermediate contacts an oligonucleotide substrate, and its 5'-PO<sub>4</sub> terminus attacks the  $\alpha$ -phosphate of AMP. As a result, AMP is transferred to the 5'-terminus and the covalent bond to the enzyme is lost. Finally, in the third step a nearby 3'-OH performs a nucleophilic attack on the activated 5'-PO<sub>4</sub>, ligating the two nucleotide strands (Pascal, 2008). The side of the nick presenting the 5'-PO<sub>4</sub> is termed "donor", and the side supplying the 3'-OH is the "acceptor" (Figure 1.1 *panel B*).

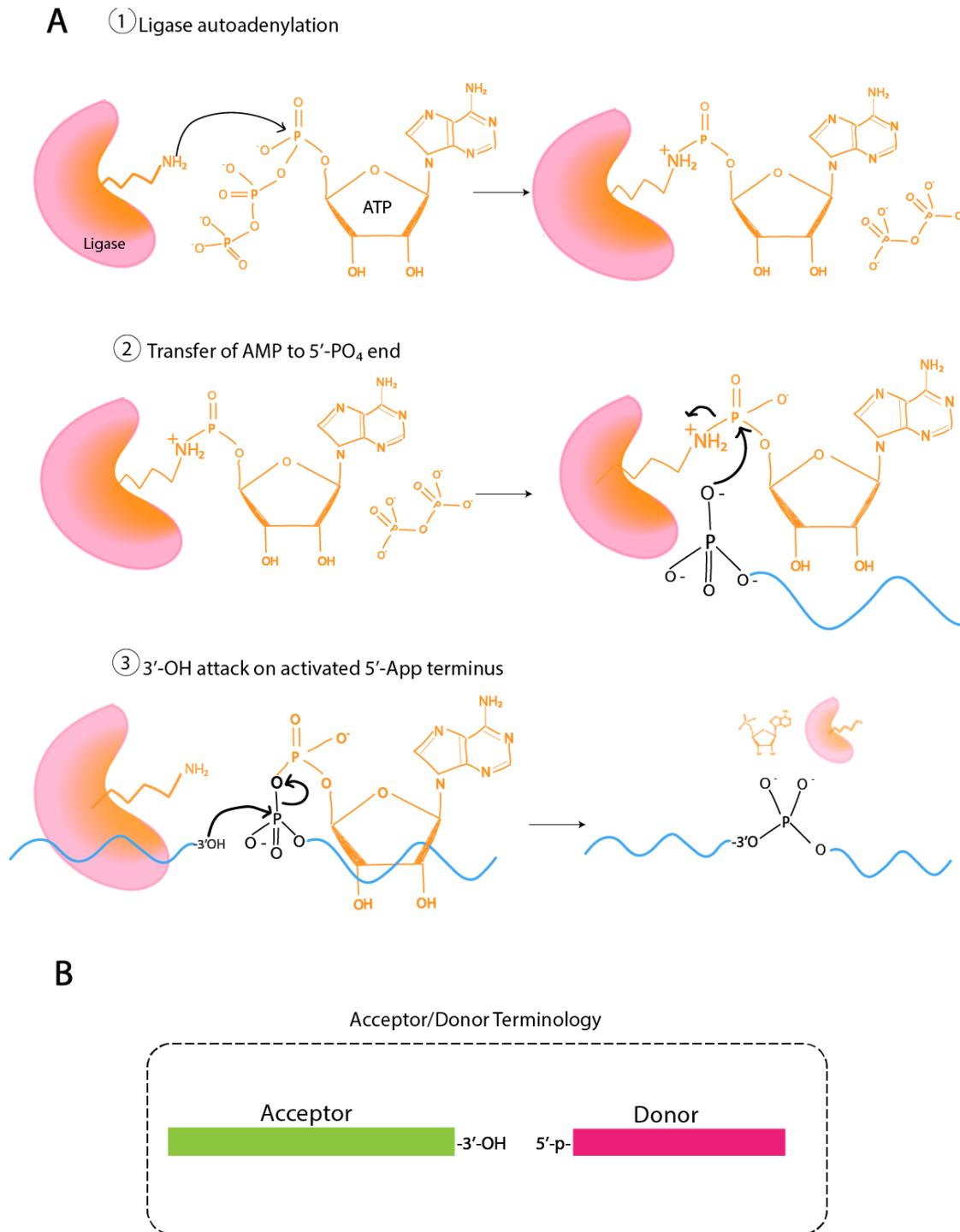


Figure 1.1: **Reaction mechanism and substrate terminology of polynucleotide ligases.** (A) General outline of the three-step reaction mechanism of polynucleotide ligases (here shown with ATP). RNA/DNA oligonucleotide shown in blue, ATP in orange, and enzyme in pink. Adapted from Nandakumar, Shuman and Lima, 2006. (B) Terminology used to specify the two sides of a nick.

Central to all DNA and RNA ligases is the nucleotide transferase (NTase) domain, which contains the catalytic elements responsible for transferal of AMP (Ho and Shuman, 2002). The nucleotide-binding pocket is made up of six conserved motifs, and many of the residues in these motifs make essential interactions with AMP, contribute to the oligonucleotide binding site, coordinate divalent metal ions, or stabilize the binding pocket (Nandakumar *et al.*, 2004; Wang, Schwer and Shuman, 2006; Banerjee *et al.*, 2019). Some residues alter their contacts during the different steps along the ligation process, while others are involved in only one of the three steps (Ho and Shuman, 2002; Nandakumar, Shuman and Lima, 2006; Wang, Schwer and Shuman, 2006). For some ligases, large conformational switches have been shown to help drive the reaction forward under optimal conditions, thereby preventing reverse reactions (Nandakumar, Shuman and Lima, 2006).

Flanking the common NTase domain are additional domains that confer substrate and pathway specificities (Timson, Singleton and Wigley, 2000). These domains vary between different polynucleotide ligases, and can be used to classify them (Becker, L'Hermitte-Stead and Myllykallio, 2019). Flanking domains may be needed to direct ligase activity to the appropriate substrate for its biological function, and contain additional enzymatic functions (Timson, Singleton and Wigley, 2000; Banerjee *et al.*, 2019). Minimal ligases only contain an additional C-terminal domain (Nandakumar, Shuman and Lima, 2006; Pascal, 2008), while members of the recently discovered Rnl3 family of RNA ligases have three flanking domains (Becker, L'Hermitte-Stead and Myllykallio, 2019; Banerjee *et al.* 2019). Through one of these domains, these RNA ligases are able to form homodimers. Dimerization has only been found in the Rnl3 family members *Pyrococcus abyssi* RNA ligase (Pab1020) and *Methanobacterium thermoautotrophicum* RNA ligase (MthRnl), and is surmised to be their biologically active form (Brooks *et al.*, 2008; Torchia, Takagi and Ho, 2008). This thesis explores members of the Rnl3 family.

### **Substrate Preferences of RNA/DNA Ligases**

For biotechnological applications of ligases their tolerance for the following substrate characteristics are interesting:

- 1) Double-strand versus single-strand substrates
- 2) Preference for certain secondary structures (related to point 1)

- 3) Nucleotide bias at the nick
- 4) Discrimination between RNA and DNA
- 5) Substrate length requirements (related to ligase footprint)
- 6) Tolerance for gaps (relevant for double-stranded ligases only)
- 7) Requirement for perfect complementarity (relevant for double-stranded ligases only)

Regarded as one of the main differences between polynucleotide ligases is whether they are able to seal nicks in single-stranded or double-stranded substrates. In double-stranded substrates a single-strand nick is templated with a complementary strand (also called a splint). Related to this preference is that ligases may be more inclined to ligate substrates with certain secondary structures (Cheng *et al.*, 2019). On the other hand, preferences on the primary sequence level are uncommon (Zhuang *et al.*, 2012). A slight nucleotide bias at the nick has been reported for some ligases e.g. the archaeal ssRNA ligase MthRnl (Zhelkovsky and McReynolds, 2011). However, detailed studies show that the primary sequence adjacent to the nick exerts less of an influence than the secondary structures formed within and between acceptors and donors (Fuchs *et al.*, 2015). Unsurprisingly, identified preferred substrates show similarity to natural substrates (Fuchs *et al.*, 2015).

Another important distinction is that many ligases show specificity for either DNA or RNA. This requirement has appeared to be mainly located at the acceptor side of the nick and is satisfied by two 3'-terminal ribonucleotides (Blondal *et al.*, 2003; Martins and Shuman, 2004; Yin *et al.*, 2004; Nandakumar, Shuman and Lima, 2006; Zhang and Tripathi, 2017). There are also ligases that do not discriminate between DNA and RNA substrates, such MthRnl (Zhelkovsky and McReynolds, 2012).

Substrate length affects ligase reactivity and product distribution, and is determined by the enzyme's footprint (number of nucleotides spanned when bound to substrate) (Yin *et al.*, 2004). Below a minimal length, the ligase cannot bind its substrate, which prevents ligation. Until this minimal length is reached, decreasing substrate length shifts product formation from circles (intramolecular ligation) to concatemers (intermolecular ligation) (Yin *et al.*, 2004).

Relevant to double-stranded ligases only is their tolerance for nucleotide mismatches between acceptor-donor-pair and template strand, and for gaps between acceptor and donor. Strict

requirements for complementarity and proximity make a ligase more useful for applications in DNA/RNA detection assays (introduced below) (Krzywkowski and Nilsson, 2017).

### **Applications of RNA ligases**

Polynucleotide ligases are important tools for a variety of applications in molecular biology, biotechnology and nanotechnology (Figure 1.2). The possible applications of a given ligase are mainly dictated by its substrate preference, efficiency, working conditions, and lack of side reactions or knowledge of how to avoid them. Based on a ligase's interaction with the substrate, there are four classes of applications:

- 1) Ligation of a nicked double-stranded (templated) substrate
- 2) Connection of two single-stranded ends of different molecules (intermolecular)
- 3) Circularization of a single stranded polynucleotide (intramolecular)
- 4) Modification of 5'- or 3'-ends

Double-strand polynucleotide ligases that require the nick to be templated are used to detect target RNAs in many types of complex samples. The critical property of ligases used in nucleic acid detection methods is their fidelity i.e. whether they can ligate despite a sequence mismatch (Krzywkowski and Nilsson, 2017). Specific detection of target micro RNAs (miRNA) among pools of miRNAs with similar sequences is becoming increasingly important as more insight into the functions of small, non-coding RNAs in health and disease is gained (Morris and Mattick, 2014; Krzywkowski and Nilsson, 2017). Fast and reliable RNA detection is also highly relevant in *in vitro* diagnostics for viral diseases e.g. Ebola (Zhang and Tripathi, 2017), and detection of clinically relevant single nucleotide polymorphisms e.g. causing drug-resistance in infectious diseases and cancers (Parikh *et al.*, 2017). Examples of such applications are in miRNA detection using padlock probe assays and RNA Annealing, Selection and Ligation assays (Jonstrup, 2006; Larman *et al.*, 2014).

In the second class of applications, ssRNA ligases are used to connect two ssDNA/RNA molecules. The combination of increasing interest in small RNAs (sRNA), and reduction of next generation sequencing (NGS) prices has made profiling and discovery of sRNAs by NGS more common. DNA and RNA ligases (typically bacteriophage T4 Rnl 1 and 2 and derivatives of these) are included in many protocols for NGS sequencing of transcriptomes, which is used to find biomarkers in cancer and stem cells, or understanding viral life cycles

and pathogenicity (e.g. recently of SARS-CoV-2) (Jayaprakash *et al.*, 2011; Kim, 2019). In these protocols, polynucleotide ligases serve to join 3' and 5' adapters to sample sRNAs. These adapters are required for downstream PCR amplification and may contain barcodes for multiplexing. Studies have shown that bias in sRNA representation arises mainly from the adapter ligation step (Hafner *et al.*, 2011; Jayaprakash *et al.*, 2011; Fuchs *et al.*, 2015). The source of this bias is the preference of ligases for certain adapter secondary structures (Zhuang *et al.*, 2012). For profiling of entire transcriptomes this can be solved by using pools of randomized adapters, and for detection of target sRNAs adapters can be optimized (Zhuang *et al.*, 2012). Another proposed option is the use of thermostable ligases for ligation at temperatures at which secondary structures are reduced (Zhelkovsky and McReynolds, 2012; Zhang and Tripathi, 2017).

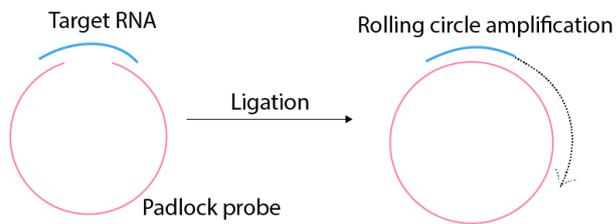
RNA/DNA ligase-mediated intermolecular ligation is also employed in mapping 3' and 5' ends of RNA (Liu and Gorovsky, 1993). In RNA ligase-mediated (RLM) Rapid amplification of cDNA ends (RACE), T4Rnl1 is used to add an ssDNA linker to the 3' side of transcripts. The 3'-end sequence is then amplified by PCR using a set of primers complementary to the 3'-linker and sequences within the transcript of interest. Similarly, mapping of 5'-ends proceeds by ligation of a 5'-ssRNA linker, and PCR-amplification with the appropriate set of primers (Liu and Gorovsky, 1993).

### (1) Double-strand nick ligation

RNA Annealing, Selection and Ligation assay:



Padlock probe assay:

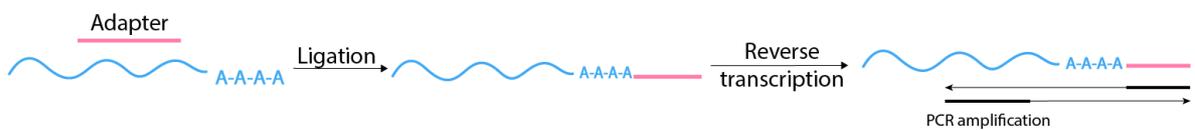


### (2) Intermolecular single-strand ligation

RNA-seq library preparation:



3'-RLM-RACE:



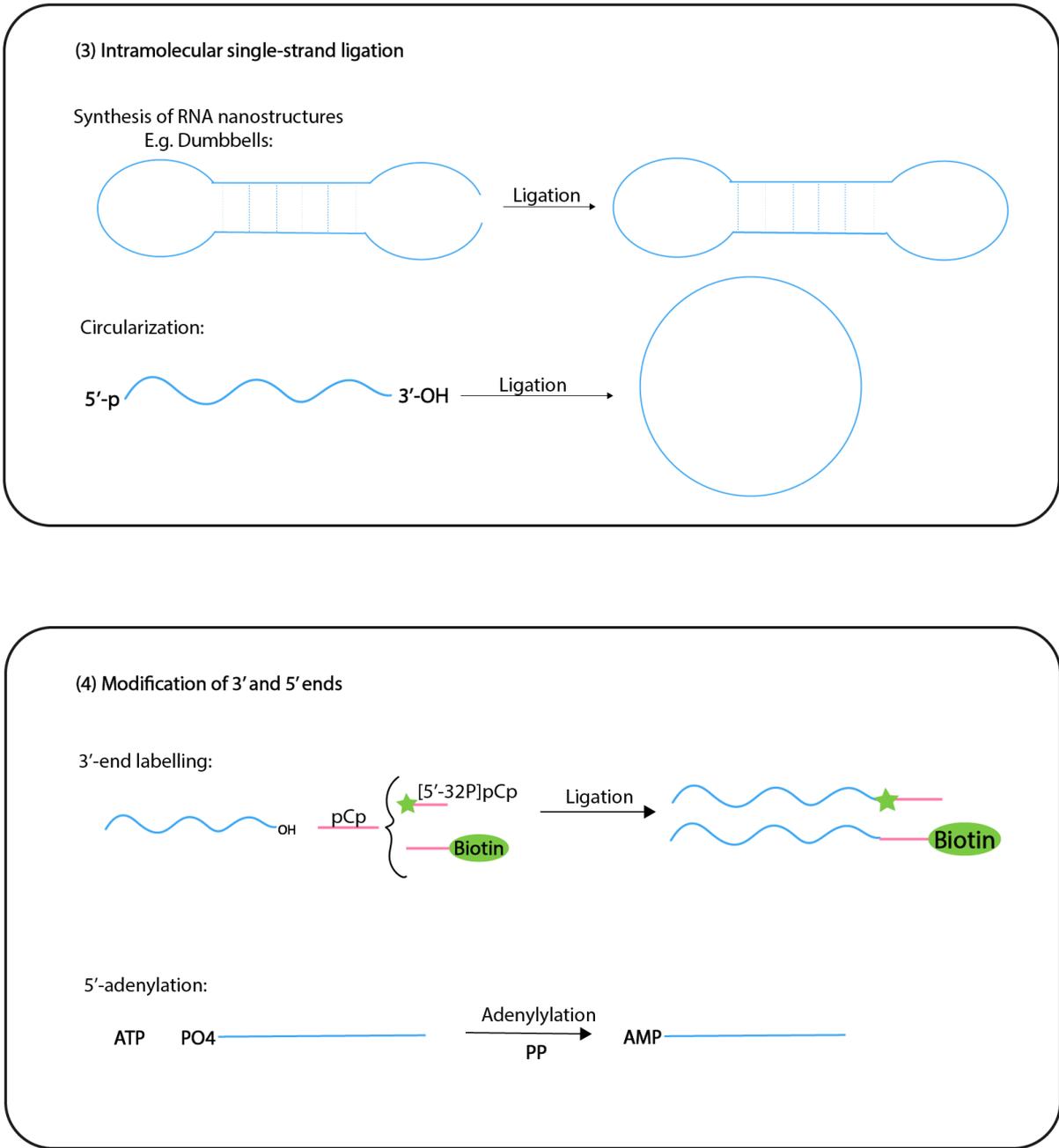


Figure 1.2: **Selection of applications for which polynucleotide ligases are used.** Application overviews are simplified. Adapted from: (Liu and Gorovsky, 1993; Jonstrup, 2006; Abe, Abe and Ito, 2007; Jayaprakash et al., 2011; Seidl and Ryan, 2011; Zhelkovsky and McReynolds, 2011; Larman et al., 2014; Cheng et al., 2019)

The third and second class of applications are similar in that they involve single-stranded

nucleic acid substrates, but while ligation is intermolecular in the second class it is intramolecular in the third. This difference changes the demands of substrate length (Yin *et al.*, 2004) and other factors such as cofactor concentrations (Cheng *et al.*, 2019). When supplying a suitable substrate, single-strand ligases can be used to form circRNA and DNA. CircRNA is used in nanotechnology e.g. for miRNA sponging (Jost *et al.*, 2018). DNA circles can be used as intracellular delivery vectors of therapeutic siRNA and miRNA by functioning as templates for rolling circle transcription (Seidl and Ryan, 2011). Producing nucleic acid circles without the need for a splint is desired for such applications, as the extra step of removing splints is circumvented (Seidl and Ryan, 2011).

As a fourth application area, RNA/DNA ligases are employed to introduce 3'- or 5'-end modifications into single-stranded substrates. 3'-OH termini of acceptors can be radioactively labelled with [5'-<sup>32</sup>P]pCp. It should be noted however, that ligase-mediated ssRNA labelling is prone to bias (England and Uhlenbeck, 1978). The reason for this is the same as for bias in adapter ligation for RNA-seq library preparation, as ligase activity is influenced by secondary structures (England and Uhlenbeck, 1978; Zhuang *et al.*, 2012).

A special type of 5'-end modification is catalyzed by thermostable MthRnl, which transfers an AMP group to a ssDNA/RNA end (step 2 of the ligation reaction) (Zhelkovsky and McReynolds, 2011). The products of this reaction (AppDNA; adenylated donors) can then be used in 3'-end labelling or adapter attachment.

## **Ligase Activity Assays**

Enzyme assays are used to confirm the presence of an enzyme in a sample (qualitative) and to calculate the amount of enzyme present (quantitative) (Bisswanger, 2014). Compared to many other enzymes, polynucleotide ligase activity is difficult to measure quantitatively. This is because increases in product and decreases in substrate are not easily measurable by common continuous, spectro-, fluoro-, or chemiluminometric assays. Although continuous fluorometric assays for measuring ligation activity have been published (Tang, 2003), product formation is still widely directly detected and quantified after fractionating samples by denaturing gel electrophoresis, hence product formation is measured in discontinuous assays.

An interesting feature of the three-step ligation reaction is that each step can be investigated in isolation by omitting necessary factors for subsequent steps or supplying pre-made

intermediates. For isolated step 1 (enzyme self-adenylation), the enzyme candidate can be incubated with radioactively labelled ATP, i.e.  $\alpha$ -[ $^{32}\text{P}$ ]-ATP with necessary co-factors and appropriate conditions (Zhelkovsky and McReynolds, 2012). Incorporation of  $^{32}\text{P}$  may then be detected in bands on an SDS-PAGE gel, by phosphor imaging and quantification (Zhelkovsky and McReynolds, 2012). The problem with using  $^{32}\text{P}$ -based methods are the safety concerns associated with radioactivity. Hence, in this thesis we refrain from using radioactivity. Step 2 (substrate-adenylation) can be isolated by omitting acceptors and using 3'-blocked donors that are unable to be circularized and concatemerized (Zhelkovsky and McReynolds, 2011). Resolution of 15-20% Urea-TBE denaturing polyacrylamide gels is sufficient to distinguish adenylated donors from non-adenylated ones, a difference of approximately one nucleotide (Zhelkovsky and McReynolds, 2011). Step 3 (formation of phosphodiester bonds) can be investigated by using pre-adenylated donors, which can be produced with commercially available kits. A number of polynucleotide ligases have been confirmed to use pre-adenylated substrates, even though this may not be the case for all polynucleotide ligases (Zhelkovsky and McReynolds, 2012). By using acceptors without 5'-phosphate and donors with blocked 3'-termini side-products (circles, concatemers, circularized concatemers) can be avoided, essentially making only intermolecular ligation possible.

The composite ligation reaction has commonly been investigated in circularization assays (e.g. Zhelkovsky and McReynolds, 2011). However, for many polynucleotide ligases and all known members of Rnl3, ATP (which is required for steps 1 and 2) inhibits step 3. Only single-turnover experiments where pre-adenylated ligases are used in absence of ATP may result in circularization (Torchia, Takagi and Ho, 2008; Zhelkovsky and McReynolds, 2011). When expressing ligases in *E. coli* a percentage of the product is often found to be pre-adenylated (Torchia, Takagi and Ho, 2008). However, this percentage shows batch-to-batch variation, and there is no guarantee that a ligase will exist in a pre-adenylated form after expression and purification.

## **1.2 Bioprospecting of Ligases from Extremophiles**

The majority of Rnl3 family ligases are found within the archaeal domain, and are especially prominent in thermophilic and hyperthermophilic archaea (Becker, L'Hermitte-Stead and Myllykallio, 2019). Bioprospecting of thermotolerant RNA ligases is interesting, because

activity at elevated temperatures reduces influences of secondary structures of ssRNA on ligation.

Bioprospecting is defined as “the systematic search for biochemical and genetic information in nature in order to develop commercially-valuable products for pharmaceutical, agricultural, cosmetic and other applications” (*Bioprospecting*, 2020). In general, bioprospecting based on gene sequences is done by either a culture-dependent or a metagenomic based approach (Vester, Glaring and Stougaard, 2015). Culture-dependent bioprospecting is severely hampered by the uncultivability of the majority of microorganisms under standard laboratory conditions, resulting in what is known as “The Great Plate Anomaly” (Amann, Ludwig and Schleifer, 1995). Decrease in sequencing price levels has made direct NGS of metagenomes feasible (Vester, Glaring and Stougaard, 2015). In this metagenomic approach extracted (g)DNA can either be directly sequenced and subjected to bioinformatic analyses (sequence-based screening), or alternatively be cloned into foreign hosts and screened for interesting functions (function-based screening) (Vester, Glaring and Stougaard, 2015). Sequence-based bioprospecting supports discovery of enzymes from extremophiles.

Extremophile is the collective term for organisms that thrive under environmental conditions such as: 55 to 121°C (thermophiles and hyperthermophiles), -2 to 20°C (psychrophiles), 2-5 M NaCl/KCl (halophiles), pH<4 (acidophiles), pH>8 (alkaliphiles) and >500 atm (barophiles) (Vieille and Zeikus, 2001; Egorova and Antranikian, 2005; Dumorne; *et al.*, 2017). Many extremophiles are unculturable by conventional methods and biomass densities in extreme environments are low and do not suffice for function-based metagenomics (Ferrer *et al.*, 2007).

The revolution in NGS is exponentially increasing the sizes of databases (Atkinson *et al.*, 2009). As a consequence of this, the new challenge lies in experimental characterization of the deposited *in silico* data (Ferrer *et al.*, 2007). Public DNA sequence databases can be searched with the help of bioinformatic tools such as sequence similarity networks (SSNs), that organize protein sequences according to their similarity (used to infer their relatedness) (Atkinson *et al.*, 2009). This was done in the work leading up to the identification of the candidates investigated in this thesis.

### **1.3 Recombinant DNA Technology and Protein Purification**

Since extremophiles often are difficult to culture in the laboratory, expression in a heterologous host using recombinant DNA technology (rDNA) is preferred. Studies have shown that mesophilic hosts such as *E. coli* are able to properly fold and express many thermophilic and hyperthermophilic enzymes (Vieille and Zeikus, 2001).

#### **Recombinant DNA Technology**

The term recombinant DNA describes laboratory-made DNA molecules, where sequences from different origins have been joined (Primrose, Sandy B. Twyman, 2009). rDNA technology can be used to create organisms with altered phenotypes such as mouse models for diseases, pest-resistant transgenic plants, and recombinant microorganisms for production of valuable molecules (Primrose, Sandy B. Twyman, 2009). Examples of the latter are therapeutic proteins, vaccines, small molecules, biopolymers and enzymes (Primrose, Sandy B. Twyman, 2009). By expressing such molecules in genetically engineered microorganisms, instead of harvesting them from natural sources, a more economical, controllable and steady supply of high quality products can be assured (Demain and Vaishnav, 2009).

#### **Recombinant Protein Expression**

The first step in producing a desired protein recombinantly is to decide on the host organism into which the gene sequence will be cloned (Rosano and Ceccarelli, 2014). Options include Gram-negative bacteria such as *E. coli* or the Gram-positive Bacilli, yeasts e.g. *Pichia pastoris*, filamentous fungi e.g. *Aspergillus* sp., mammalian cells, or even insect or plant cells (Demain and Vaishnav, 2009). The choice of host depends on the origin and characteristics of the desired protein e.g. requirements for post-translational modifications (Rosano and Ceccarelli, 2014). The coliform bacterium *E. coli* has become the most popular host organism due to its short generation time, requirement for relatively simple and inexpensive growth media, ability to attain high cell density, easy transformation, and vast amount of compatible expression systems and modified strains (Demain, 2001).

After deciding on a host organism, a suitable expression vector is chosen based on the size of the insert (gene sequence encoding the protein of interest), copy number, incompatibility, selectable marker, cloning sites, and other special features such as affinity tags (Rosano and Ceccarelli, 2014). For expressing a single protein (<15 kb), plasmids are commonly used

(Vester, Glaring and Stougaard, 2015). Essential components of plasmid expression vectors are the gene sequence encoding the protein of interest, a promoter, the 5' untranslated region, transcription terminator sequence, origin of replication, and selection markers (Figure 1.4) (Rosano and Ceccarelli, 2014; Gawin, Valla and Brautaset, 2017).

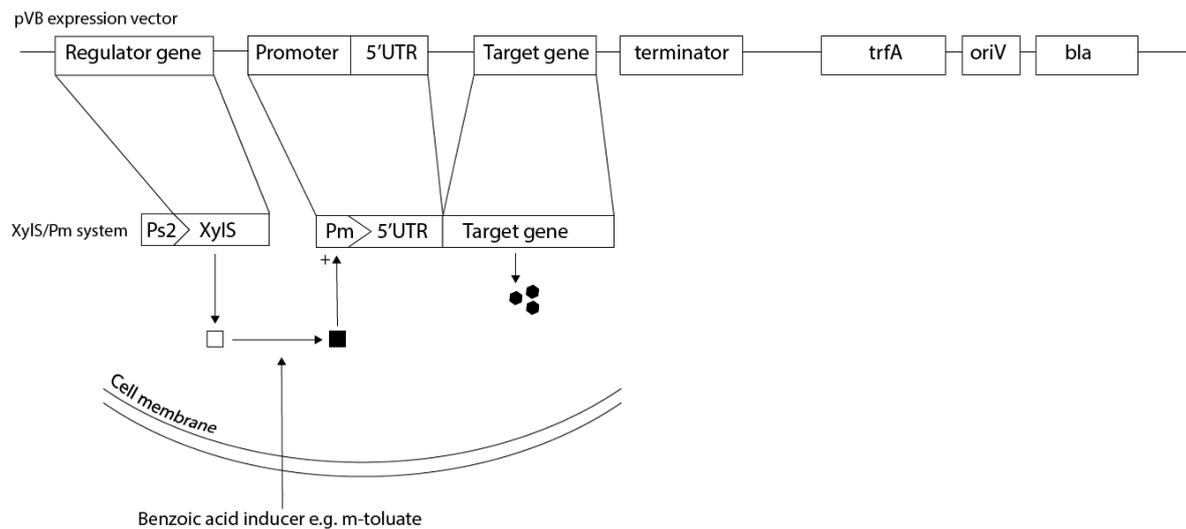


Figure 1.4: **Example of fundamental components of an expression vector.** Simple schematic of the expression vector and XylS/Pm regulator/promoter system used in this thesis. Adapted from Rosano and Ceccarelli, 2014; Gawin, Valla and Brautaset, 2017a. 5'UTR: 5' untranslated region; oriV: origin of vegetative replication; *trfA*: initiation protein TrfA-encoding gene; *bla*:  $\beta$ -lactamase-encoding gene.

Choice of host organism and expression vector influence one another, as the vector system must be compatible with the host (e.g. incompatibility groups, uptake of inducer, and inability to break down inducer) (Gawin, Valla and Brautaset, 2017). Flexibility in the host organism is offered by using broad-range expression vectors e.g. the XylS/*Pm* regulator/promoter system used in this thesis.

## The XylS/*Pm* Expression System

The expression system used herein consists of the RK2 plasmid harbouring the XylS/*Pm* regulator/promoter system. XylS/*Pm* originates from the TOL (toluene-degradative) plasmid pWW0 of *Pseudomonas putida* (Gawin, Valla and Brautaset, 2017). In *P. putida*, *Pm* controls transcription of enzymes involved in catabolism of aromatic hydrocarbons such as benzoates (Gawin, Valla and Brautaset, 2017). Transcription from the *Pm* promoter is positively regulated by XylS, whose transcriptionally active form is a dimer (Zwick, Lale and Valla,

2013). Dimerization of XylS is increased by a wide range of inducers, commonly derivatives of benzoic acid (Zwick, Lale and Valla, 2013).

Benefits of the XylS/*P<sub>m</sub>* regulator/promoter system are the many options for cheap and non-toxic inducers, the fact that inducers enter the cells passively without the need for specific uptake systems, that inducers in many cases are not metabolized and work in a dose-dependent manner, adjustable vector copy number, and compatibility with a wide variety of Gram negative and positive hosts (Gawin, Valla and Brautaset, 2017).

## **Chromatographical Protein Purification**

Chromatography denotes a class of methods for separation of components in a mixture, making up a mobile phase, by moving it through a stationary phase. Recombinantly expressed proteins directed to the cytoplasm are extracted by chemical, enzymatic, or mechanical cell disruption techniques. In this crude extract, the protein (enzyme) of interest cannot be reliably characterized, as other enzymes and substances might interfere with its activity (Kornberg, 2009). FPLC allows separation of the protein of interest (POI) and contaminating biomolecules with high-resolution. An FPLC system consists of a column packed with a resin (constituting the stationary phase) through which the cell extract-containing mobile phase is passed. Separation is achieved by exploiting differences in interactions between the sample components and the stationary phase in a general four-step process: equilibration, sample application, wash, elution. Commonly, the sample is applied under conditions favouring its binding to the stationary phase. Remaining non-binding components are removed by washing under similar conditions. The POI is eluted by changing conditions to become unfavourable for binding in either a stepwise or gradual manner.

FPLC methods are defined based on these interactions, where different properties of the POI are utilized: the isoelectric point (pI) and buffer pH (charge; ion exchange chromatography), hydrophobicity (hydrophobic interaction chromatography), size (and shape) (size exclusion chromatography), and affinity for a ligand embedded in the stationary phase (affinity chromatography) (GE Healthcare, 2010).

## **Affinity Chromatography**

Affinity chromatography is based on specific and reversible, biospecific or biononspecific interactions between a target protein and an affinity ligand (GE Healthcare, 2010).

Immobilized onto the stationary phase is an appropriate affinity ligand, which enables binding of the target under favourable conditions. Among the many possible-ligand target interactions that can be utilized, the strong affinity of transition metal ions for histidine and cysteine is often exploited in immobilized-metal affinity chromatography (IMAC). Polyhistidine-tagging of proteins allows relatively good purification with IMAC when little is known about the protein (Urh, Simpson and Zhao, 2009). IMAC stationary phase consists of an affinity matrix (e.g. agarose), to which a chelating agent (commonly nitrilotriacetic acid) is attached. Metal ions (e.g.  $\text{Ni}^{2+}$ ) are coordinated by the chelating agent in a way that leaves valencies available for binding to histidines via their imidazole rings (Urh, Simpson and Zhao, 2009). Captured proteins are eluted from the stationary phase by adding high concentrations of imidazole, which competes with the histidine-tagged protein for binding to metal ions.

### **Ion-Exchange Chromatography**

In ion-exchange chromatography (IEX) the interaction of a protein with the resin depends on its net charge, charge density, and surface charge distribution (GE Healthcare, 2016). Proteins can be differentiated by the pH at which their net charge is zero i.e. the isoelectric point (pI). For monomeric, non-posttranslationally modified proteins the theoretical isoelectric point (pI) can be calculated from their primary sequence (Burgess, 2009). Theoretical pI, however, does not take the charge distribution into account. The net surface charge of such a protein is determined by its pI and the pH of the surrounding medium. The protein has a positive net charge when the pH is below the pI, and negative net charge when the pH is above the pI. By exploiting the attraction of opposite charges, the protein in the mobile phase may be selectively retarded by an oppositely charged stationary phase. High resolution is attained by maximizing differences in surface net charge between components. Binding strength of a protein increases with the distance between its pI and the pH in solution (GE Healthcare, 2016).

Two types of ion-exchange chromatography can be distinguished based on the charge of their stationary phase: cation exchange (CEX; resin binds positively charged proteins) and anion exchange (AEEX; resin binds negatively charged proteins). A charged ligand is attached onto a base matrix composed of porous beads e.g. sepharose. It is the charge of this ligand which determines whether the column is an anion or a cation exchanger. In a cation exchanger, the

ligand is negatively charged e.g. sulphopropyl (SP), while in an anion exchanger it is positively charged e.g. quarternary ammonium (Q).

### **Analysis of Protein Purity**

Assessment of the quality of enzyme product after chromatographical purification is important, because subsequent characterization is affected by whether the product is pure and homogeneous, soluble and correctly folded, and precisely quantified (Raynal *et al.*, 2014). Protein quality is determined by the absence of unrelated proteins, non-functional POIs, and other macromolecules such as nucleic acids, lipids and carbohydrates (Rhodes and Laue, 2009; Raynal *et al.*, 2014) Depending on the nature of the impurities of interest, different analytical methods such as electrophoresis, chromatography, light scattering, and mass spectrometry can be used (Rhodes and Laue, 2009). In addition to these component-based analyses, purity assessment can also be activity-based. This is possible when the total protein concentration is known, and the enzyme activity is quantifiable. Contaminating enzymes can also be detected by their activity, if suitable activity assays exist.

A common first purity analysis is denaturing gel electrophoresis (sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), described in Section 3.5), due to its low cost and simplicity (Rhodes and Laue, 2009). Proteins are detected with Coomassie blue or other more sensitive stains such as silver staining (Rhodes and Laue, 2009). Sensitivity will however always depend on the amount and protein concentration of sample loaded, and small amounts of interfering contaminants may not be detectable. Contaminants of similar mass are also not distinguishable from the target protein on an SDS-PAGE but can be detected by including isoelectric focusing upstream. Analysis of oligomerization and aggregation can be performed by dynamic light scattering, which gives a qualitative indication of monodispersity. Analytical size-exclusion chromatography (SEC) is also useful for revealing oligomerization and reveals more about the nature of the oligomers than dynamic light scattering (Raynal *et al.*, 2014). A more detailed picture of contaminants and covalent modifications of the target protein can be obtained with mass spectrometry, which also can clarify whether an impurity is a true contaminant or self-associated form of the target protein (Rhodes and Laue, 2009). Mass spectrometry can also be coupled downstream of other methods such as reversed phase high performance liquid chromatography (RP-HPLC) and tryptic digestion (as in peptide mass finger-printing) (Raynal *et al.*, 2014).

## 2 Purpose and Specific Aims

The enzyme market is constantly demanding novel enzymes or novel characteristics of existing enzymes for use in emerging or existing technologies. To serve this demand, bioprospecting for new enzymes or enzyme activities, respectively, is an important part of the research and development efforts of biotechnology companies such as ArcticZymes AS, Norway (ArcticZymes, 2020). Previously, a set of 22 gene sequences encoding putative Rnl3 family ligases was established. Two Rnl3 family members have been characterized in the literature, of which one has found commercial use in preparing RNA sequencing libraries for NGS. The purpose of this thesis is to provide enzyme leads that can become alternatives to the existing commercial RNA ligases. The overall aim of the study was to identify new RNA ligases by screening this set, and to characterize hits as far as possible.

In order to fulfil the overall aim, this study is built upon the following sub-aims:

**I. Apply, provide feedback and suggest changes to a functional screening protocol for Rnl3 family ligases.**

Screening of the set involves testing their soluble expression from *E. coli*, initial activity tests with a protocol used for other Rnl3 ligases, and upscaled production. The workflow is used to select a smaller number of interesting candidates. Simultaneously, each step of the screening process is evaluated for improving future screening efforts.

**II. Scaling up production of interesting candidates and purifying them by FPLC.**

To provide sufficient pure and stable enzyme for in depth biochemical characterisation and to generate estimates about production yield and feasibility of the most interesting enzyme candidates.

**III. Develop improved methods to investigate activity of ssRNA ligases.**

In RNA ligases, step 3 of the ligation reaction is known to be inhibited by ATP, which is required for step and 1 and 2. Hence, to investigate the entire ligation process, protocols where ATP is depleted prior to step 3 are needed.

**IV. Investigate essential biochemical characteristics of ssRNA ligases and their usability in current applications.**

New and improved assays will be established or developed to determine properties such as temperature optima, substrate specificity and dependency on co-factors.

## 3 Materials and Methods

### 3.1 Workflow of the Study

The work in this thesis follows a pipeline in which the initial set of 22 putative Rnl3 family ligases is narrowed down to a smaller number of promising candidates. Starting with 22 *E. coli* BL21(DE3) clones, soluble expression of the candidates was screened in small-scale batches of 30 mL. The batches were FPLC-purified with a single IMAC step, and presence of target His-tagged RNA ligase candidates in eluates was confirmed by western blotting. Initial activity screening was performed using the standard substrate adenylation protocol for existing Rnl3 family ligase MthRnl. Candidates were selected and expressed in upscaled batches in *E. coli* (1 L) based on activity, novelty, and soluble expression. This batch was purified in a 2-step IMAC-IEX protocol and tested for phosphatase and endonuclease activity with standard assays. After confirming substrate adenylation activity in the new batch, ligation (step 3) activity was assessed in the purified samples. This was done with two new protocols, which were based on existing assays that were improved as part of this thesis. It was also investigated whether melting points of the candidates could be determined with thermal shift assay (TSA). The candidates with the best substrate adenylation efficiency were characterized in more detail with regard to effects of temperature and co-factor (ATP) on adenylation. The standard adenylyltransferase assays were adapted for this purpose. For an overview of the workflow see Figure 3.1.

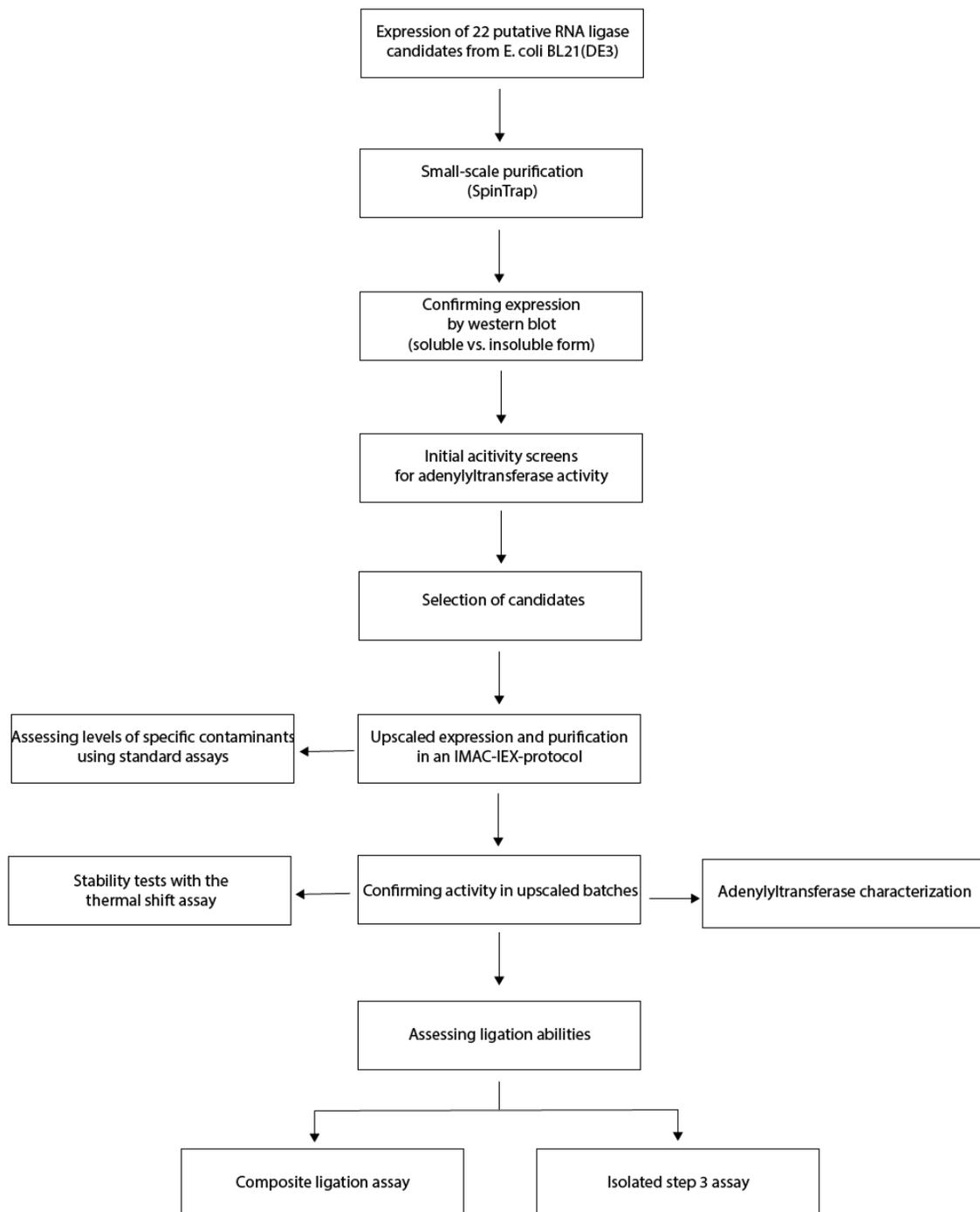


Figure 3.1: **Simplified overview of the workflow of this thesis.** IMAC: Immobilized Metal Affinity Chromatography. IEX: Ion-Exchange Chromatography

## 3.2 Media and Solutions

### Buffers and Solutions for Cell Culture and Lysis

Table 3.1: Components of buffers and solutions used for cell culture and lysis.

Solution	Content
<b>Growth media:</b>	
<b>Potassium phosphate buffer pH 7.2 @ 25°C</b>	717 mM K <sub>2</sub> HPO <sub>4</sub> 283 mM KH <sub>2</sub> PO <sub>4</sub> Phosphoric acid 85% or potassium hydroxide (saturated) for pH adjustments
<b>Lysis Broth (LB)</b>	10 g/L Bacto™ Tryptone 5 g/L Bacto™ Yeast Extract 5 g/L NaCl 1 piece of NaOH (pellets) (For agar plates 1.5% (w/v) agar was added) Milli-Q Ultrapure Water (MQ) to final volume
<b>Terrific Broth (TB)</b>	12 g/L Bacto™ Tryptone 24 g/L Bacto™ Yeast Extract 4 mL/L glycerol (85-87%)  100 mL/L autoclaved and filtered potassium-phosphate buffer pH 7.2@25°C. MQ to final volume
<b>Enzymatic cell lysis:</b>	
<b>Lysis buffer A</b>	1 mM MgCl <sub>2</sub> 1 mM PMSF Add small-scale IMAC binding buffer (Table 3.5) to final volume.

	<p>Add fresh before use:</p> <p>0.2 mg/mL lysozyme</p> <p>100 U/mL HL-SAN</p>
<b>Lysis buffer B</b>	<p>50 mM Tris-HCl pH 7.5</p> <p>500 mM NaCl</p> <p>0.5% Tween 20</p> <p>5% glycerol</p> <p>1 mM NaCl</p> <p>0.5% Tween 20</p> <p>Add fresh before use:</p> <p>1 mg/mL lysozyme</p> <p>400 U/mL HL-SAN</p>
<b>Lysis buffer C</b>	<p>50 mM Tris-HCl pH 8.5 @ 25°C</p> <p>10 mM imidazole</p> <p>500 mM NaCl</p> <p>5 mM MgCl<sub>2</sub></p> <p>5% glycerol (v/v)</p> <p>0.1% Triton X-100</p> <p>MQ to final</p>

## Solutions for Gel Electrophoresis

Table 3.2: Composition of solutions used for gel electrophoresis

<b>Solution</b>	<b>Content</b>
Tris/Borate/EDTA (TBE)	<p>89 mM Trizma base</p> <p>89 mM boric acid</p> <p>2 mM EDTA</p> <p>MQ to final volume</p>

Tris-acetate-EDTA (TAE)	40 mM Trizma base 5 mM Sodium acetate trihydrate 1 mM Titriplex® III, EDTA disodium salt dihydrate Acetic acid (100%) for adjusting the pH to 8.0 MQ to final
20% acrylamide/7M Urea mix	7 M Urea 1X TBE buffer (10X) 250 mL Acrylamide/Bis 19:1 40% (w/v) solution MQ to final
Denaturing Urea-PAGE gels	80 mL 20% acrylamide/7M Urea mix 800 µL 10% APS 24 µL TEMED
0.7% agarose gel	0.7% (w/v) analytical grade agarose to 300 mL with 20X TAE buffer Stain with 5 µL gel red per 50 mL

## Solutions for Western Blot

Table 3.3: Composition of solutions used for western blots.

Solution	Content
Western Blot Transfer buffer	25 mM Tris-HCl (pH 7.5 @ 25°C) 192 mM glycine 20% (v/v) methanol MQ to final
Tris-buffered saline with Tween 20 (TBST) buffer	20 mM Tris-HCl (pH 7.5 @ 25°C) 150 mM NaCl 0.1% Tween 20 MQ to final

3% Blocking buffer	3% (w/v) nonfat dry milk TBST to final
--------------------	---

## Solutions for Enzyme Assays

Table 3.4: Composition of solutions used in enzyme assays.

Solution	Content
<b>Adenylation assays:</b>	
<b>General dilution buffer</b>	25 mM Bis-Tris-Propane/HCl (pH 7.0 at 25°C) 100 mM NaCl
New England Biolabs (NEB) 1.1	10 mM Bis-Tris-Propane-HCl (pH 7.0 at 25°C) 10 mM MgCl <sub>2</sub> 100 µg/ml BSA
Ligation buffer (According to Zhelkovsky and McReynolds, 2012)	10 mM BisTris-Propane (pH 7.0 at 25°C) 10 mM MgCl <sub>2</sub> 5 mM DTT
NEB 1X 5' DNA Adenylation Reaction Buffer	50 mM Sodium Acetate (pH 6.0 at 25°C) 10 mM MgCl <sub>2</sub> 5 mM DTT 0.1 mM EDTA
<b>Endonuclease assay:</b>	
Endonuclease assay stop solution	0.25 % (w/v) Bromophenol blue pH 2.8-4.6 60 mM EDTA 0.6% (w/v) SDS 40 % (w/v) Sucrose MQ to final

---

**Phosphatase assay:**

---

Phosphatase buffer	200 mM glycine buffer (pH 9.5 at 25°C) 1 mM MgCl <sub>2</sub> MQ to final
rSAP dilution buffer	50 mM Tris-HCl (pH 7.6 at 4°C) 2 mM MgCl <sub>2</sub> 0.2 mM ZnCl <sub>2</sub> MQ to final

---

**Buffers for Chromatographical Protein Purification**

Table 3.5: Compositions of buffers used in chromatographical protein purification.

---

Solution	Content
<b>Small-scale IMAC (SpinTrap):</b>	
Binding buffer (BB)	20 mM Tris-HCl (pH 7.5 at 25°C) 500 mM NaCl 20 mM imidazole MQ to final
Elution buffer (EB)	20 mM Tris-HCl (pH 7.5 at 25°C) 500 mM NaCl 500 mM imidazole MQ to final
<b>IMAC (ÄKTA Avant):</b>	
Buffer A1 (Binding Buffer)	50 mM Tris-HCl (pH 7.5 at 25°C) 50 mM NaCl 20 mM imidazole

---

	1 mM MgCl <sub>2</sub> MQ to final
Buffer A2 (Wash Buffer)	50 mM Tris-HCl (pH 7.5 at 25°C) 2 M NaCl 20 mM imidazole 1 mM MgCl <sub>2</sub> MQ to final
Buffer B1 (Elution Buffer)	50 mM Tris-HCl (pH 7.5 at 25°C) 50 mM NaCl 500 mM imidazole 1 mM MgCl <sub>2</sub> MQ to final
<b>IEX:</b>	
IEX Buffer A	20 mM Tris-HCl (pH 7.5 at 25°C) 0.02% (v/v) Tween 20 5 mM MgCl 5% (v/v) glycerol MQ to final
IEX Buffer B	20 mM Tris-HCl (pH 7.5 at 25°C) 0.02% (v/v) Tween 20 5 mM MgCl 5% (v/v) glycerol 1 M KCl MQ to final

### **3.3 Protein Expression in *E. coli***

#### ***E. coli* Strain and Expression Vector**

In this thesis, expression of 22 RNA ligase candidates was induced in *E. coli* BL21(DE3) harbouring the pVB-1A0B1 expression vectors created by Vectron Biosolutions (Trondheim). Plasmid vectors were supplied with the candidate genes cloned downstream of a non-cleavable hexahistidine (6XHis) tag. Constructs are based on the RK2-plasmid and equipped with the XylS/*Pm* regulator/promoter system and antibiotic resistance gene *bla* (Further described in Section 1.3). This expression cassette is highly flexible in terms of host strain requirements but is supplied in *E. coli* BL21(DE3) here. Among its beneficial features is the deficiency of protease genes *Lon* and *OmpT* (Rosano, Morales and Ceccarelli, 2019). Candidate genes had been codon optimized by GenScript (Piscataway, NJ, USA), as archaeal proteins use rare *E. coli* codons (Kim *et al.*, 1998).

#### **Feasibility Testing of Expression**

All work with *E. coli* expression strains was done under sterile conditions, and growth media were autoclaved (at least 121°C for 20 min). For selection pressure, all growth media was supplemented with 100 µg/mL ampicillin.

In the initial screening round, expression from the 22 *E. coli* clones was done in 30-mL cultures. *E. coli* clones obtained from glycerol (20%) stocks stored at -80°C were cultivated on fresh LB-agar plates. A single colony was used to inoculate a 5 mL LB-medium pre-culture. Pre-cultures were incubated overnight (ON) at 37°C and with shaking (230 rpm). Pre-cultures were diluted 100-fold into 30 mL TB-medium and the optical density at 600 nm (OD<sub>600</sub>) was monitored with an Ultrospec® 10 Cell Density Meter (Amersham Biosciences) while cultures were grown at 30°C (230 rpm). At OD<sub>600</sub> approximately 1.5, expression was induced ON at 16°C (230 rpm) with 2 mM *m*-toluic acid. For expression analysis by SDS-PAGE, samples were withdrawn from the uninduced culture and the post-induction culture before harvesting. Cell pellets from these samples were collected by centrifugation at 11000 rpm for 1 min (Heraeus™ Fresco™ 21, 24 x 1.5/2.0mL rotor), and stored at -20°C.

#### **Optimization of Soluble Expression**

Effects of inducer concentration and cell density at the time of induction on soluble-to-insoluble-ratio was investigated for three candidates (X08, X13 and X14). For each parameter

a high and a low value was tested, resulting in four conditions for each clone. For inducer concentration a high value of 2 mM and a low value of 0.5 mM *m*-toluene was chosen. For cell density at induction a high value of OD<sub>600</sub>= 1.5-2.0 and a low value of OD<sub>600</sub>= 0.5 was chosen.

*E. coli* clones were grown as before during feasibility testing, except that incubation times were monitored more strictly. Precultures were grown for exactly 16 hours, as were induced cultures. Aliquots were taken immediately prior to induction (pre-induction sample) and after expression for 16 hours (post-induction sample) and prepared for SDS-PAGE analyses as follows. Cell pellets from both samples were obtained by centrifugation at 11000 rpm for 1 min (Heraeus™ Fresco™ 21, 24 x 1.5/2.0mL rotor). Pre-induction samples were resuspended in 50 mM Tris-HCl (pH 7.5 at 25°C) to an OD<sub>600</sub>=2. Viscosity caused by DNA was reduced by incubation with 100 U/mL HL-SAN (ArcticZymes) for 2 hours at 37°C. Post-induction samples were enzymatically lysed ON in lysis buffer C with shaking (80 rpm) at 15°C. For high OD-samples, volume of lysis buffer C resulted in an approximate OD<sub>600</sub>=20, while the OD<sub>600</sub> for low OD-samples was approx. 6. Samples were incubated with 500 U/mL HL-SAN for 3 hours at 37°C. Soluble and insoluble fractions were separated by centrifugation at 21000 g for 30 min (Heraeus™ Fresco™ 21, 24 x 1.5/2.0mL rotor). Supernatant was decanted and the insoluble fraction was resuspended in 50 mM Tris/HCl (pH 7.5 at 25°C). SDS-PAGE samples were prepared to contain OD<sub>600</sub>=1.5 of every sample and 1X LDS.

Coomassie-stained SDS-PAGE gels were digitalized and analyzed using the Gel Doc XR+ Gel Documentation System and Image Lab Software (Bio-Rad, Hercules, CA, USA). ImageLab (ver. 4.1) lane profile and analysis tools were used to obtain the intensity of the appropriate peaks.

### **3.4 Chromatographical Purification of Candidates**

A protocol for chromatographical protein purification can consist of one or more steps, depending on requirements for purity posed by use of the target protein. The recommended general purification scheme has three stages: capture, intermediate purification, and polishing (GE Healthcare, 2010). His-tagged proteins are preferably purified in an initial IMAC-step, which allows rapid capturing of the target protein from a relatively large volume of sample. Because IMAC-purification can result in sufficient purity, a single IMAC-step was used for

purification prior to initial activity screening. After selection of active candidates for further characterization a two-step protocol with an additional intermediate purification step was followed.

### **Small-Scale IMAC Purification on a His SpinTrap Column**

Ligase candidate-encoding genes had been cloned in frame with an N-terminal 6XHis-tag to allow fusion proteins to be expressed. His-tagged ligases could hence be IMAC purified by exploiting the affinity of polyhistidines for Ni<sup>2+</sup> cations. Soluble proteins were extracted from cell pellets of 30 mL-cultures by lysis in lysis buffer A. The volume of lysis buffer A was adjusted to result in cell densities (OD<sub>600</sub> = 30) based on the original densities of the sample. Lysis was performed for 30 min at 25°C with gentle shaking. Lysate was clarified by centrifugation at 21000 g for 30 min (4°C) (Fiberlite™ F15-8 x 50cy Fixed Angle Rotor; Thermo Scientific). The supernatant was loaded onto an IMAC binding buffer equilibrated His SpinTrap column (GE Healthcare, Little Chalfont, UK) and purified according to the manufacturer's instructions.

Pilot experiments (X03, X07 and X20) with varying amounts of lysate were performed to gauge the maximal amount of lysate that could be applied without oversaturating the Ni-Sepharose resin (binding capacity of 750 µg His-tagged protein per column). For moderately well-expressed samples (X01, X02, X04, X05, X06, X18) (based on initial Coomassie-stained SDS-PAGE analysis) 1.2 mL lysate was applied, for poorly expressed samples (X08-17, X19, X21, X22) 1.8 mL lysate was applied. Bound proteins were eluted with 500 mM imidazole into an elution buffer containing 500 mM NaCl. A high NaCl concentration was chosen in order to prevent ionic interactions between unrelated proteins and the medium, and because high salt is known to stabilize many proteins.

### **Up-Scaled IMAC Purification on a HisTrap Fast Flow Column**

Cell pellets of up-scaled 1-L *E. coli* cultures were lysed with shaking (80 rpm) at 15°C in lysis buffer C. The volume of lysis buffer C was adjusted to result in a cell density of OD<sub>600</sub>= 120 based on the original density of the culture. After incubation ON the lysate was clarified by centrifugation at 21000 g for 20 min (4°C) (Fiberlite™ F14-6 x 250LE Fixed Angle Rotor; Thermo Scientific). For X07, X06 and X03, the resulting supernatant was filtered (Nalgene™ Rapid-Flow™ 0.45 µm aPES) prior to purification. Purification was performed with an

ÄKTA Avant system (GE Healthcare, Little Chalfont, UK), which had been prepared by washing pumps/filling tubes with Buffer A1. Using a flow rate of 1 mL/min, the supernatants were applied onto a HisTrap FF Crude 5 mL column (GE Healthcare, Little Chalfont, UK) equilibrated with Buffer A1. The column was washed with five column volumes (CV) of Buffer A2 followed by 5 CV Buffer A1. Bound proteins were eluted with a gradient of 0-60% Buffer B1 over 10 CV. Remaining protein was eluted with a step to 100% Buffer B1 over 3 CV. Absorption at 280 nm was monitored constantly by the integral UV-detector, creating the chromatogram of the purification.

### **Ion-Exchange Chromatography**

IMAC eluates were further purified by ion exchange chromatography in either an Eppendorf tube format (X07), manually (X06) or using the ÄKTA Avant system (X03, X12, X19) through a 1 mL HiTrap column (Sigma-Aldrich, St. Louis, MO, USA). Depending on the theoretical pI of the putative polynucleotide ligases, anion (pI < 7.5; Q Sepharose HP) or cation (pI > 7.5; SP Sepharose Fast Flow) exchange resins were used.

For the Eppendorf tube format, 1 mL of the IMAC eluate (50 mM NaCl) was applied directly to 200 µL of IEX resin slurry in a 1.5 mL Eppendorf tube. To allow protein binding, IEX resin and eluate were mixed thoroughly and incubated at 15°C for 20 min with gentle agitation. The mixture was centrifuged at 1000 g (Heraeus™ Fresco™ 21, 24 x 1.5/2.0mL rotor) for 2 min, and the supernatant (constituting the flow through) was removed and stored at 4°C. The resin was washed three times with 1000 µL (5 CV) 5% IEX Buffer B, and FT was collected between every step by centrifugation as described earlier. Protein was eluted stepwise (10%, 20%, 30%, 50%, 100% IEX Buffer B) and each step, except the two last, was performed twice. Samples were taken at each step and analyzed by Nanodrop and Coomassie-stained SDS-PAGE. Care was taken to avoid IEX resin during analysis (by spinning down the sample).

Candidate X06 (theoretical pI = 8.29) was manually purified on a 1 mL HiTrap SP Sepharose column (Sigma-Aldrich, St. Louis, MO, USA) using a syringe. Throughout the purification the flow rate was kept at 1 mL/min. The column was equilibrated with 5% Buffer B (5CV), sample was applied, and the column was washed with 5% Buffer B (10CV). Elution was done

in steps as described above with 5CV of 10, 20, and 30% Buffer B and 10CV of 50 and 100% Buffer B.

Candidates X03, X12, and X19 were purified using a 1 mL HiTrap SP Sepharose (X03 and X12) or Q Sepharose (X19) column on an ÄKTA Avant system using the same buffers and protocol as above.

### **3.5 Analytical Methods**

#### **Dialysis**

For small-scale purified samples (<200 µL) salt concentrations were adjusted and imidazole was removed by dialysis with Pur-A-Lyzer™ Mini 12000 Dialysis Kit 12-14 kilo Dalton (kDa) MWCO (Sigma-Aldrich, St. Louis, MO, USA) or Mini Dialysis Kit, 8 kDa cut-off (GE Healthcare, Little Chalfont, UK) against a dialysis buffer (buffer-to-sample ratio > 1000). The standard composition of dialysis buffers was 20 mM Tris-HCl (pH 7.5 at 25°C), 100 mM KCl, 1 mM DTT, 1 mM MgCl<sub>2</sub> and variable concentrations of NaCl. Dialysis was completed ON at 4°C with stirring (50 rpm).

#### **SDS-PAGE**

Gel electrophoresis is a technique that allows separation of macromolecules based on differences in molecular weight, size and charge (Garfin, 1990). By applying an external, electrical field, polypeptides are pulled through a gel matrix with a velocity dependent on their size and charge (Garfin, 1990). In sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) proteins are treated with SDS, which is an anionic detergent that binds to and causes unfolding of the polypeptide chain. Because the negative charge imparted by SDS is proportional to the molecular weight of a protein, different species in a sample are separated based on their size (Garfin, 1990). In this thesis, proteins are separated by their molecular weight in order to confirm presence of desired proteins and for evaluating the homogeneity of a purified sample. For detection of the target protein in the cell lysate two methods were used. First, proteins were detected by staining with Coomassie Brilliant Blue. Second, for better sensitivity and specificity SDS-PAGE gels were used in Western Blotting.

Proteins were separated on precast gels, either SurePAGE™ (Genscript Biotech, Piscataway, NJ, USA) or NuPAGE™ (Thermo Fisher Scientific, Waltham, MA, USA). Protein gels were

run at 200 V for 45 min in Tris-MOPS-SDS running buffer (GenScript, Genscript Biotech, Piscataway, NJ, USA). Protein sizes could be approximated by running one well with 5  $\mu$ L of either Mark12™ Unstained Standard (Thermo Fisher Scientific, Waltham, MA, USA) or PageRuler™ Prestained Protein Ladder (Thermo Fisher Scientific, Waltham, MA, USA).

Different methods were used for preparation of samples, as described in Section 3.3. When assessing expression levels in soluble and insoluble fractions, the cell density was used to normalize samples. When resolving fractions from chromatographical purification, samples always contained 15  $\mu$ L fraction and 5  $\mu$ L 4X loading buffer. After addition of loading buffer samples were incubated at 95°C for 5 min, and 18  $\mu$ L were applied to the gel.

SDS-PAGE gels were either Coomassie-stained with SimplyBlue™ SafeStain (Invitrogen, Carlsbad, CA, USA) or used for Western Blotting. Coomassie-staining was done according to the microwave protocol suggested by the manufacturer. Gel images were digitalized on a Gel Doc XR+ Gel Documentation System (Bio-Rad, Hercules, CA, USA).

## **Western Blot**

Western blotting is a method by which antibodies are used to detect specific proteins among a mixture of different proteins (Alegria-Schaffer, Lodge and Vattem, 2009). Starting with a sample separated on an SDS-PAGE gel, proteins are transferred to a membrane, commonly nitrocellulose or polyvinylidene difluoride (PVDF), by electrophoresis. Spots on the membrane not bound by protein are blocked with a nonrelevant protein in order to avoid contaminants from binding non-specifically (Alegria-Schaffer, Lodge and Vattem, 2009). The membrane is then incubated with an antibody that has affinity for the protein of interest. In an indirect western blot, this primary antibody can then be bound by another, secondary antibody which is conjugated to a detection system e.g. an enzyme. Because the secondary antibodies are polyclonal they permit indirect labelling of the protein of interest and amplify the signal. For the sake of accurate detection and preventing false positives, the membrane is washed thoroughly after every incubation step.

Proteins were transferred from the SDS-PAGE gel to a PVDF membrane (Merck) using the The XCell II™ Blot Module (Invitrogen, Carlsbad, CA, USA). Fiber pads, filter paper, and PVDF membrane were equilibrated in Transfer buffer. The Immobilon-P PVDF membrane was pre-wet for 30 s in methanol, rinsed in MQ, and placed in Transfer buffer.

The blotting sandwich was assembled on the anode in the following order: fiber pad, absorbent paper, PVDF membrane, SDS-PAGE gel, filter paper, and fiber pad. Air bubbles were carefully removed by rolling a pipette over the stack. Proteins were transferred by electroblotting in Transfer buffer for approx. one hour at 30V.

For antibody-mediated detection, the PVDF membrane was blocked in 3% Blocking buffer for 40 min at RT. Blocking buffer was decanted and the PVDF membrane was incubated with diluted (1:250 in 3% Blocking buffer) primary antibodies (6x-His Tag Polyclonal Antibody, #PA1-983B, Thermo Fisher Scientific, Waltham, MA, USA) at RT for 2-3 hours. The membrane was washed three times for 5 min in TBST (Table 3.3) with gentle shaking, and incubated with diluted (1:5000) secondary antibodies (Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alkaline Phosphatase conjugate, #G-21079, Thermo Fisher Scientific, Waltham, MA, USA) for 1 hour at 25°C. By being conjugated to alkaline phosphatase, the secondary antibodies allow detection by colorimetry.

The PVDF membrane was washed three times in TBS-T for 5 min with shaking. For colorimetric detection, the membrane was incubated in 5-bromo-4-chloro-3-indolyl phosphate (BCIP)/nitro blue tetrazolium (NBT) Alkaline Phosphatase Substrate (Sigma-Aldrich, St. Louis, MO, USA) for 15-30 min, rinsed in MQ water and imaged with the Gel Doc XR+ Gel Documentation System (Bio-Rad, Hercules, CA, USA).

### **Determination and Increasing of Protein Concentration**

Protein concentrations were routinely assayed spectrophotometrically with a Nanodrop 2000c (Thermo Fisher Scientific, Waltham, MA, USA). Aromatic amino acids (Trp, Tyr) show absorbance at ultraviolet wavelengths of 280 nm. Protein concentration is derived from the Beer-Lambert law by calculating the molar extinction coefficient from the primary sequence (Noble and Bailey, 2009) (Equation 3.5.1).

Equation 3.5.1

$$A = a_m * c * l$$

Where A is the measured absorbance at a certain wavelength,  $a_m$  is the molar extinction coefficient, c is the protein concentration, and l is the path length (Noble and Bailey, 2009). Blank samples were measured in appropriate buffers or MQ water.

In cases where dialysis strongly reduced the protein concentration, samples (X12 100 mM and 300 mM) were concentrated with Amicon® Ultra 0.5 mL Centrifugal Filters (Sigma-Aldrich, St. Louis, MO, USA) at 14000 g (Heraeus™ Fresco™ 21, 24 x 1.5/2.0mL rotor) for 15 min.

## **3.6 Enzyme Characterization**

### **Adenylyltransferase Assay**

Polynucleotide ligase candidates' ability to transfer an AMP group to single-stranded DNA, equivalent to the product of the two first steps of the ligation process, was assessed and characterized in the adenylyltransferase assay. Substrates possessed 5'-phosphate groups and were either 3' blocked with dideoxycytosine (ddC) or 6-Carboxyfluorescein (6FAM)-labels, or open (3'OH)(Table 3.6). Activity was assessed under conditions described by Zhelkovsky and McReynolds (Zhelkovsky and McReynolds, 2011): assay mixture (20 µL total) consisted of 5 µM (100 pmol) 5'-phosphorylated 23-mer ssDNA, adenylation buffer/NEB's 5' DNA Adenylation Reaction Buffer (50 mM Bis-Tris buffer (pH 6.0 at 25°C), 10 mM MgCl<sub>2</sub>, 5 mM DTT, 0.1 mM EDTA), 100 µM ATP, and 100 pmol purified ligases. Commercial MthRnl (NEB#M2611A) was used in positive controls, and negative control reactions were run without added enzyme. Reaction mixtures were incubated at 65°C or 37°C for 1 hour. Incubation temperatures were chosen based on whether the ligase candidate derived from a mesophilic or thermophilic organism.

If not described otherwise, reactions were stopped by heat-inactivation at 85°C for 5 min, followed by addition of denaturing formamide stop solution. Samples were resolved on a 20% polyacrylamide/7M urea gel in TBE-buffer at 200V for approx. 80-90 min. Bands were detected by UV transillumination at 302 nm after SYBR Gold staining (Invitrogen) for 30 min. When using 6FAM-labelled substrates, bands were first detected directly in the fluorescein-channel on a Gel Doc XR+ Gel Documentation System (Bio-Rad, Hercules, CA, USA) and subsequently SYBR Gold-stained.

### **Composite Ligation Assay**

As part of this thesis the improved “composite ligation assay” was developed based on existing protocols. With this assay, the entire ligation process could be investigated in one tube. Existing protocols require both the WT Rnl3 family ligase MthRnl and its mutant

Thermostable 5' App DNA/RNA Ligase (NEB) which has reduced step 2 (de-adenylation) activity. This assay was not suitable for the work in this thesis, because the purpose of a composite ligation assay here was to screen for activity in new candidates. It was found that positive controls (MthRnl WT) could generate ligation products after a step of ATP depletion.

This assay was performed in a 40  $\mu$ L reaction mix containing 200 pmol of ligase candidate, 200 pmol of donor, 50 pmol of donor, 100  $\mu$ M ATP, 16  $\mu$ L 50% PEG 8000 (NEB), and 1X NEBuffer™ 1.1. The reaction was incubated at 65°C for 1 hour and cooled down on ice. rSAP was added to 0.07 U/ $\mu$ L, and the sample was incubated 30 min at 37°C. After the rSAP treatment the sample was again incubated at 65°C for 2 hours. Samples were taken throughout the process as described in Figure 4.15 *panel A*. Samples were resolved and detected as described above.

### **Isolated Step 3 Assay**

Step 3 of the ligation reaction was investigated in isolation by using the 5' DNA Adenylation Kit by NEB to prepare adenylated single stranded substrates as described above. MthRnl ligase was inactivated at 85°C for 5 min and degraded with Proteinase K (1.6 U/ $\mu$ L) for 15 min at 37°C). The sample was cooled down and rSAP (2 U/ $\mu$ L sample; ArcticZymes) was added in order to break down remaining ATP.

The ligation assay was set-up as described by Zhelkovsky and McReynolds, 2012: 8 pmol of pre-adenylated donors, 5 pmol of acceptor, 10 mM Bis-Tris-Propan-HCl pH 7.0 at 25°C, 10 mM MgCl<sub>2</sub>, 5 mM DTT, and 20 pmol ligase in a total volume of 10  $\mu$ L. The adenylation kit was assumed to be 90% effective. The reaction mix was incubated at 65°C for 2 hours, heated to 85°C for 5 min and then stopped with formamide-stop solution. For all incubation steps at above 37°C a thermocycler (Veriti™ 96-Well Thermal Cycler, Thermo Fisher Scientific, Waltham, MA, USA) was used.

Table 3.6: **The oligonucleotides used in this thesis.** 6FAM: 6-Carboxyfluorescein; ddC: dideoxycytosine; nt: nucleotide.

Name	Sequence (5'-3')	Features
<b>Acceptors</b>		
<b>6FAM-ssDNA23-OH</b>	ACTGATAGAAGTTCCCGAAACAG	<ul style="list-style-type: none"> <li>• 23 nt</li> <li>• 5'-label detectable by fluorescence</li> <li>• 3'-OH termini</li> </ul>
<b>ssRNA20-OH</b>	CGAGCUCGAAUUCACUGGCC	<ul style="list-style-type: none"> <li>• 20 nt</li> <li>• No 5'-phosphate</li> <li>• 3'-OH termini</li> </ul>
<b>ssRNA23-OH</b>	ACUGAUAGAAGUCCCCGAAACAG	<ul style="list-style-type: none"> <li>• 23 nt</li> <li>• No 5'-phosphate</li> <li>• 3'-OH termini</li> </ul>
<b>Donors</b>		
<b>p-ssDNA23-FAM</b>	AAAATTGTACTAGTCCTTAGATA	<ul style="list-style-type: none"> <li>• 23 nt</li> <li>• 3'-label detectable by fluorescence</li> <li>• 5'-phosphate</li> </ul>
<b>p-ssDNA23-ddC</b>	AAAATTGTACTAGTCCTTAGATA	<ul style="list-style-type: none"> <li>• 23 nt</li> <li>• 3'-termini blocked with ddC</li> <li>• 5'-phosphate</li> </ul>
<b>p-ssDNA23-OH</b>	AAAATTGTACTAGTCCTTAGATA	<ul style="list-style-type: none"> <li>• 23 nt</li> <li>• Open 3'-OH termini</li> <li>• 5'-phosphate</li> </ul>
<b>Other</b>		
<b>6FAM-ssDNA47</b> (47 nt ssDNA marker)	ACTGATAGAAGTTCCCGAAACAGAAA ATTGTACTAGTCCTTAGATAC	<ul style="list-style-type: none"> <li>• 5'-label detectable by fluorescence</li> <li>• 3'-OH termini</li> </ul>
<b>6FAM-ssDNA20</b> (20 nt ssDNA marker)	GGCCAGTGAATTCGAGCTCG	<ul style="list-style-type: none"> <li>• 5'-label detectable by fluorescence</li> <li>• 3'-OH termini</li> </ul>

## Thermal Shift Assay

The thermal shift assay is based on the concept that temperature-induced destabilization causes proteins to unfold and reveal hydrophobic environments (Ericsson *et al.*, 2006). Detection of fluorogenic dyes that increase their quantum yield when binding to such environments serves as an indirect measurement of unfolding. Unfolding can therefore be quantified by plotting the increase in fluorescence against increase in temperature, yielding the stability curve. From stability curves, the melting temperature can be derived. The effect of various buffer conditions and additives on protein stability can be assessed using the change in melting temperature. SYPRO Orange is a much-used dye for these assays, because its excitation and emission wavelengths are compatible with regular quantitative PCR (qPCR) instruments (Huynh and Partch, 2015).

Standard assay compositions were based on Ericsson *et al.*, 2006, consisting of 7.5  $\mu\text{L}$  of 6X SYPRO Orange (Thermo Fisher Scientific, Waltham, MA, USA), 12.5  $\mu\text{L}$  of 2X buffer, and 10  $\mu\text{M}$  of protein, MQ water was added to 25  $\mu\text{L}$ . Reactions were set up in MicroAmp<sup>TM</sup> Optical 96-Well Reaction Plates (Applied Biosystems), sealed with LightCycler<sup>®</sup> 480 Sealing Foil (Roche) and run in a QuantStudio<sup>TM</sup> 3 Real-Time PCR System (Applied Biosystems). The program was run from 20°C to 90°C in increments of 0.2°C/min. Data analysis was done using JTSA (Bond, 2017) where curves were fitted using the Sigmoid-5 model.

Pilot TSA experiments were conducted on purified X07 from upscaled cultures. Different buffers were tested: standard ligation buffer and NEB 1X 5' DNA Adenylation Reaction Buffer. MilliQ water was used for buffer controls, and positive enzyme controls were performed with another ligase resembling T4 DNA ligase (ArcticZymes).

For samples where clear melting curves were lacking, different dye and protein concentrations were tested. As recommended by Huynh and Partch (2015), final assay concentrations of 1, 5, 10, and 20  $\mu\text{M}$  of X07, and 1, 5, 10, and 20X SYPRO Orange were used (Huynh and Partch, 2015). Dilutions of X07 were prepared so that 12  $\mu\text{L}$  of diluted enzyme would result in the target protein concentration. Two sets of dilutions were prepared: high (1041 mM) and low (208 mM) NaCl, yielding final low (100 mM) and high (500 mM)

NaCl-concentrations. SYPRO Orange was diluted serially in MQ water from a 5000X stock. For the buffer, a 2X HEPES buffer (pH 7.5 at 25°C) was used.

### **3.7 Detection of Specific Contaminants**

Phosphatases are enzymes that catalyze hydrolysis of phosphate groups, and hence can render the donor unfit for adenylation. Nucleases catalyze cleavage of phosphodiester bonds between nucleic acids and can be specific for DNA or RNA and be active at ends of (exonucleases) or within (endonucleases) oligonucleotides. Single strand nucleases are detrimental to functional studies of the putative polynucleotide ligases in this study because they catalyze break down of substrate.

#### **Phosphatase Activity Assay**

Deleterious phosphatase activity in purified samples was analysed spectrophotometrically by addition of 4-*p*-Nitrophenyl Phosphate (pNPP) in excess. Conversion of colourless pNPP into the coloured *p*-nitrophenol is catalyzed by phosphatases.

Samples were prepared in lidded 96-well plates (Eppendorf Cell Culture Plates, 96-Well). To all wells, 100  $\mu$ L of 2X phosphatase buffer was added. For the enzyme samples, volumes equivalent to 50 and 100 pmol were added to the wells. A standard curve was prepared with serial dilutions (1 to  $10^{-6}$  U/ $\mu$ L in steps of  $10^{-1}$ ) of rSAP (ArcticZymes) in rSAP dilution buffer. Of each rSAP dilution, 5  $\mu$ L were added to the wells. To the samples and standard curve, 100  $\mu$ L of pre-warmed (37°C) 4-pNPP-substrate (5 mM) was added. Absorbance was measured every 30 s at 405 nm using a SpectraMax iD3 microplate reader (San Jose, California, United States).

Data was analysed using SoftMax Pro Software. rSAP concentrations for which curves could not be fitted were masked. For the remaining rSAP dilutions, absorbance was plotted against time. Only measurements from 3000 and 4000 s ( $\approx 27$  data points) were included in order to exclude the lag phase and measurements at saturation. Slopes from these graphs were then plotted against the assay concentrations of rSAP to create the standard curve. The original activity in the samples was derived multiplying activity calculations in the assay with the sample dilution factor.

## **Nuclease Activity Assay**

Presence of active single strand endonucleases in purified products was assessed qualitatively on circular, viral DNA from the phage vector M13mp18 (Tebu-Bio, Le Perray en Yvelines, France). A serial dilution (2X, 4X, 8X) of sample was prepared in 25 mM Tris-HCl. For every sample, 20  $\mu$ L of diluted sample was added to 30  $\mu$ L of assay mix (66 mM Tris-HCl (pH 8.0 at 25°C), 33 mM MgCl<sub>2</sub>, 0.5  $\mu$ g M13mp18 plasmid DNA), MQ water to final 30  $\mu$ L). Negative controls contained only 20  $\mu$ L dilution buffer and 30  $\mu$ L assay mix. Positive controls were made with 1  $\mu$ L of in-house reference batch with high nuclease contamination level. Samples were incubated at 37°C for 4 hours and stopped with 10  $\mu$ L of endonuclease assay stop solution.

Samples were analysed by running them through a 0.7% agarose gel stained with 1X GelRed (Lonza™ GelStar™ Nucleic Acid Gel Stain, Fisher Scientific) for 30 min at 90 V. Bands on the gel were detected in the GelRed channel of the GelDoc XR+ Gel Documentation System (Bio-Rad, Hercules, CA, USA).

## **3.8 Similarity Matrix and Phylogenetic Analysis**

To create similarity and identity matrices, amino acid sequences of the 22 candidates were aligned using ClustalW (with the settings gap open cost of 10 and gap extend cost of 0.2) inside the Geneious software (Version R10.2). From the alignment, similarity and identity matrices were created using the BLOSUM 62 matrix (with threshold 1) for the similarity calculations.

For the phylogenetic tree, an alignment was created again with ClustalW. The tree was built according to the Jukes and Cantor model of evolution using UPGMA clustering algorithm and bootstrapped 500 times to create the consensus tree. The tree was rooted with T4Rn12 as the outgroup.

## 4 Results

### 4.1 Initial screening of Candidate Polynucleotide Ligase Expression in *E. coli*

Expression of 22 putative RNA ligases (X01-X22) from pVB1A0B1 vectors in *E. coli* BL21(DE3) was analysed. The aim was to identify which candidates were expressed, and to what extent they were present in soluble form. SDS-PAGE analyses of crude lysates were difficult to interpret due to masking by an endogenous *E. coli* protein. Hence, small-scale SpinTrap purified samples were prepared and analysed by parallel SDS-PAGE and western blot.

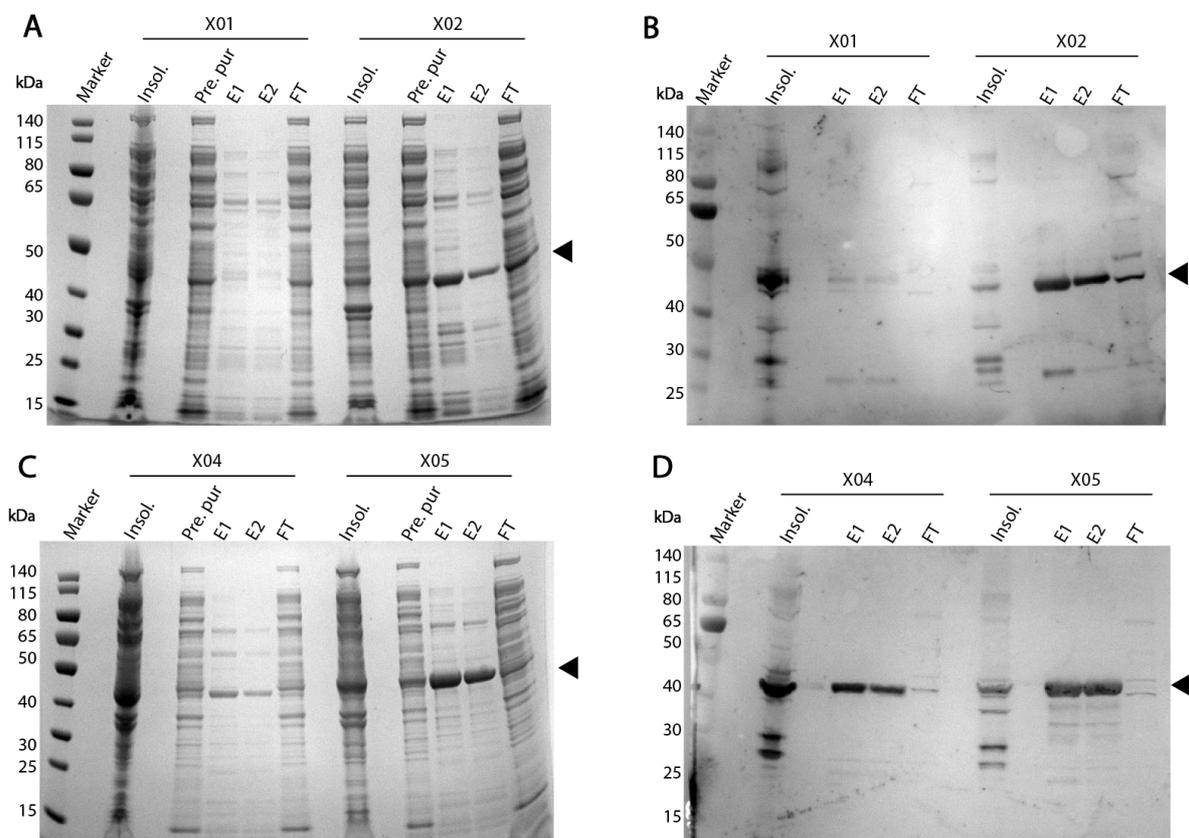


Figure 4.1: **Expression analysis of candidates X01, X02, X04 and X05.** Examples of the parallel SDS-PAGE and Western blot analyses of RNA/DNA ligase candidate expression in *E. coli*. **(A)** Coomassie-stained SDS-PAGE gel of X01 and X02. Samples are taken from the insoluble fraction (Insol.) after expression in *E. coli*, pre-purified (pre. pur.), SpinTrap purified eluates (first fraction: E1 and second fraction: E2), and flow through (FT). **(B)** Western blot detection of His-tagged target candidates X01 and X02 in the same samples as (A). Anti-His antibodies were bound by alkaline phosphatase-conjugated secondary antibodies. Coloured compound produced in an alkaline phosphatase catalyzed reaction was detected. **(C)** and **(D)** show the SDS-PAGE gel and western blot for X04 and X05. Marker is PageRuler™ Prestained Protein Ladder (ThermoFisher). Samples were run on a SurePAGE, Bis-Tris, 4-20% gel in SDS-MOPS.

The identity of the His-tagged RNA ligase candidates in eluates was confirmed by western blotting with anti-His antibodies. Purification was judged qualitatively by including a pre-purification sample in the SDS-PAGE analysis. Only a selection of the results is shown here, and the reader is directed to Appendix B for the remaining results.

Theoretical protein sizes had been calculated to be in the range of 42 to 47 kDa (Appendix A; Table A.1). Proteins detected by western blotting fell mainly within this size range.

Corresponding bands of similar intensities were observed by SDS-PAGE analysis. Additional smaller proteins of 10 or 15 kDa were also detected (Figure 4.1 *panel A and C*). Comparing pre-purification samples and eluates in the SDS-PAGE analysis, the extent of purification appears to be low. Distinctive contaminants of 65-80 kDa are present in most of the samples (Figure 4.1 *panel A and C*).

By comparing expression of insoluble and soluble protein in identical volumes of lysate, ligases could be categorized as having high, moderate or low soluble expression. Six candidates were deemed as being expressed in highly soluble form (X02, X05, X06, X07, X12, X18), all of these except X02 also had a high general expression level (Table 4.1). Five candidates were highly expressed in insoluble form (X08, X13, X15, X16, and X19).

All *E. coli* clones displayed similar growth during exponential phase. Note that there were no controls (empty vector and no vector controls) for comparison. The strongest growth reductions were noted in X01 and X02. X02 ceased to grow completely in the two biological replicates tested in this thesis (data not shown).

Candidates had been SpinTrap purified and stored in 50% (v/v) glycerol at -20°C for four weeks or more leaving them in a storage buffer containing 250 mM NaCl and 250 imidazole. A selection of purified samples had been dialyzed against different NaCl-concentrations prior to storage. Eluates of X02, X03, and X07 had been split and dialyzed against either a high (500 mM) or a low (100 mM) NaCl concentration and stored in a final 250 mM or 50 mM NaCl (respectively) in absence of imidazole for approximately one month before activity assessment.

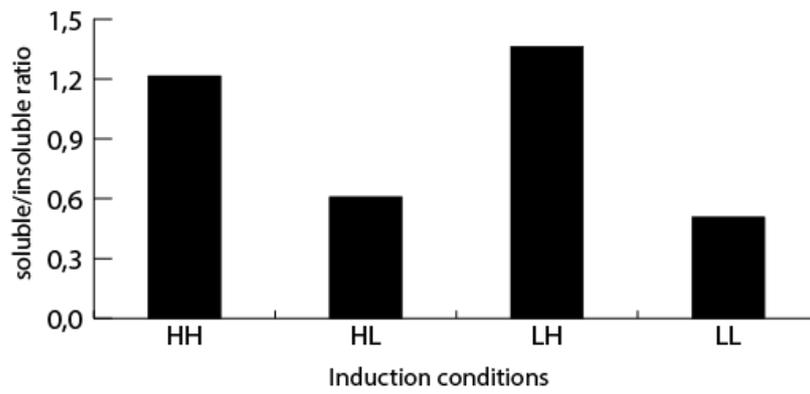
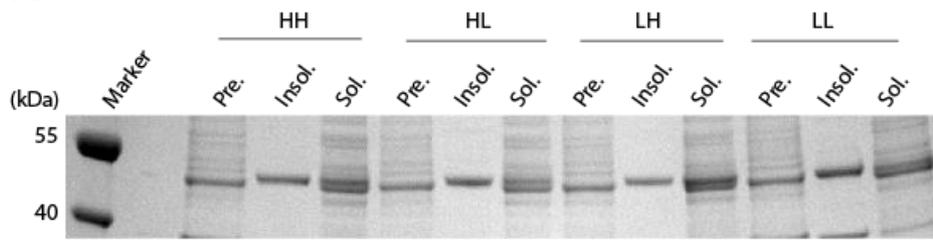
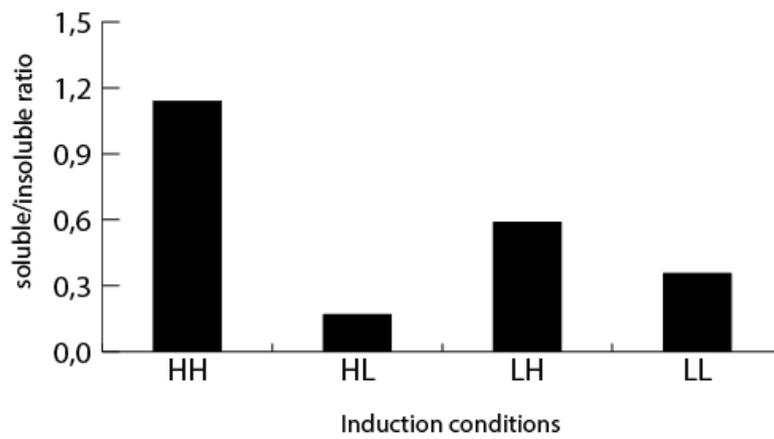
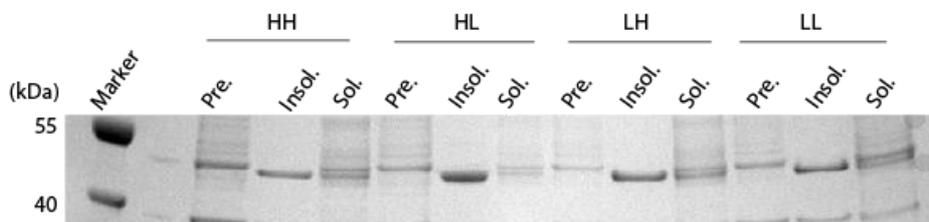
Table 4.1: **A summary of the general expression level and solubility of candidates, as determined in initial expression feasibility tests.** Qualitative judgements of general expression levels (+: low; ++: moderate; +++: high) are based on Coomassie-stained SDS-PAGE analysis. Solubility is defined by the soluble-to-insoluble ratio, where ratio<1 is low, ratio≈1 is moderate, and ratio>1 is high. Thermo-, halo-, and acido-/alkaliphilicity of the candidate's organism of origin are presented. ND: not detectable.

<b>Candidate</b>	<b>General expression</b>	<b>Solubility</b>	<b>Origin</b>
<b>X01</b>	+	Low	Thermophile, Halophile
<b>X02</b>	++	High	Thermophile
<b>X03</b>	+++	Moderate	Thermophile
<b>X04</b>	+++	Moderate	Mesophile
<b>X05</b>	+++	High	Mesophile
<b>X06</b>	+++	Moderate-High	Hyperthermophile, Halophile
<b>X07</b>	+++	High	Hyperthermophile
<b>X08</b>	+++	Low	Thermophile
<b>X09</b>	++	Moderate	Mesophile
<b>X10</b>	++	Low	Thermophile, Acidophile
<b>X11</b>	-	ND	Mesophile
<b>X12</b>	+++	Moderate-High	Thermophile
<b>X13</b>	+++	Low	Thermophile
<b>X14</b>	++	Low	Thermophile
<b>X15</b>	+++	Low	Mesophile, Halophile
<b>X16</b>	+++	Low	Mesophile, Halophile
<b>X17</b>	+	ND	Mesophile
<b>X18</b>	+++	High	Hyperthermophile
<b>X19</b>	+++	ND	Mesophile, extreme halophile
<b>X20</b>	++	Moderate	Mesophile, Halophile, Alkaliphile
<b>X21</b>	+	ND	Mesophile, Extreme halophile
<b>X22</b>	+	Low	Unknown

## 4.2 Optimizing the Soluble-to-Insoluble Ratio

Improving the soluble-to-insoluble ratio of expressed ligases was attempted. To this end, four different combinations of high ( $OD_{600}=1.5-2.0$ ) and low ( $OD_{600}=0.5$ ) cell density at induction and high (2.0 mM *m*-toluate) and low (0.5 mM *m*-toluate) inducer concentration were investigated for *E. coli* clones expressing either X08, X13, or X14. These candidates were, in initial expression screening experiments, deemed to be highly (moderately in the case of X14) expressed in insoluble form (Table 4.1). Mind that the experiment only was performed once without replicates, and without inducer and empty vector controls. For all three candidates high inducer concentrations resulted in lower cell densities after induction than the low inducer concentrations (data not shown). This was found to be true for both induction conditions tested.

Regarding changes in the soluble-to-insoluble ratio, none of the candidates showed clear improvements and at best the soluble and insoluble expression was similar (ratio=1). Initial screening experiments employed the high OD-high inducer ( $OD_{600}=1.5-2.0$  with 2 mM *m*-toluene; HH) regime. Soluble expression of X08 appeared to have benefited slightly from high inducer concentrations, both at low and high cell densities (Figure 4.2 *panel A*). Soluble levels of X13 were stable across the conditions, and only at the HH regime the fractions were equal in size i.e. insoluble expression dominated in HL (high OD-low inducer), LH (low OD-high inducer), and LL (low OD-low inducer) conditions) (Figure 4.2 *panel B*). Results showed that X14 barely was expressed in soluble form, while insoluble expression was relatively high (Figure 4.2 *panel C*).

**A****B**

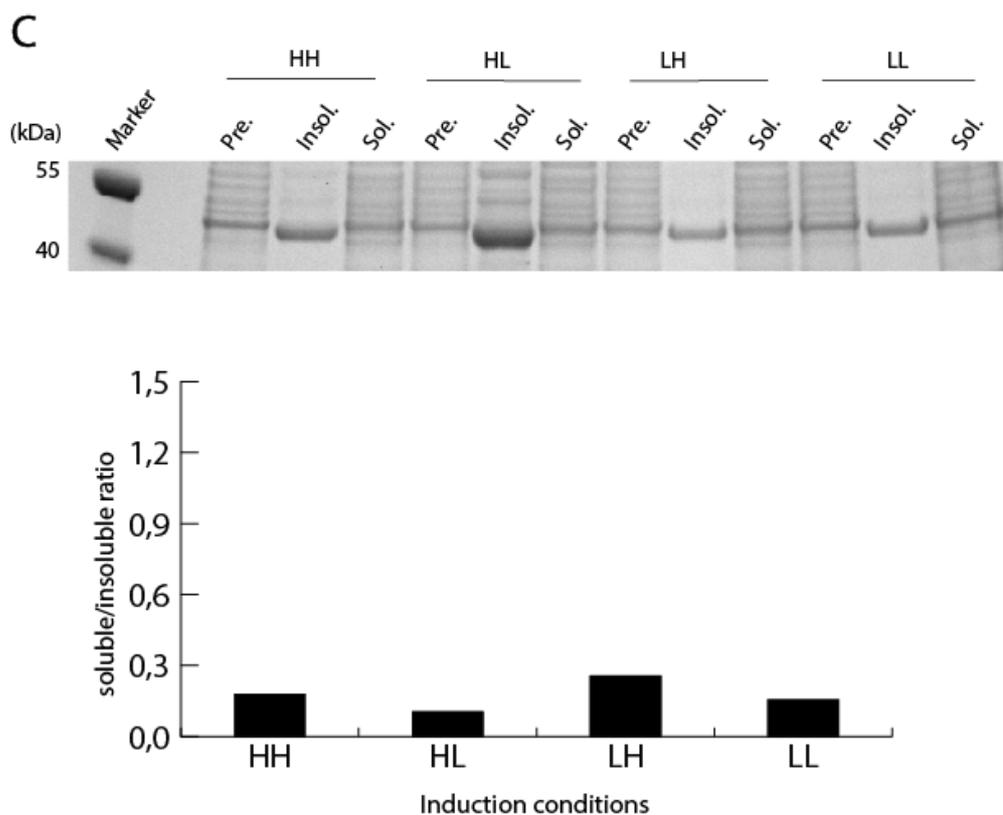


Figure 4.2: **Effects of cell densities at induction and inducer concentrations on soluble-to-insoluble ratio.** Combinations of high (OD600=1.5-2.0) and low (OD600=0.5) cell densities and high (2 mM) and low (0.5 mM) *m*-toluic acid concentrations were used with three candidates expressed mostly in insoluble form: **(A)** candidate X08, **(B)** candidate X13, and **(C)** candidate X14. HH: High OD/high inducer concentration; HL: High OD/low inducer concentration; LH: Low OD/high inducer concentration; LL: Low OD/low inducer concentration. Samples were resolved on a 10% Bis-Tris SDS-PAGE gel, stained with Coomassie blue, and detected in the Coomassie-blue channel on a Gel Doc XR+ Gel Documentation System.

### 4.3 Initial Activity Screening Using the Adenylyltransferase Assay

The two first steps of the ligation process, i.e. adenylyltransferase activities were assessed with 3'-6-carboxyfluorescein-labelled 23-mer pDNA donors (p-ssDNA23-6FAM) (Table 3.6) under standard conditions describes in the Methods. Replicates were performed using the same enzyme batch.

Six ligases were found to produce bands migrating at the same speed through denaturing PAGE gels as the positive control (MthRnl from NEB): X02 (in-house produced MthRnl),

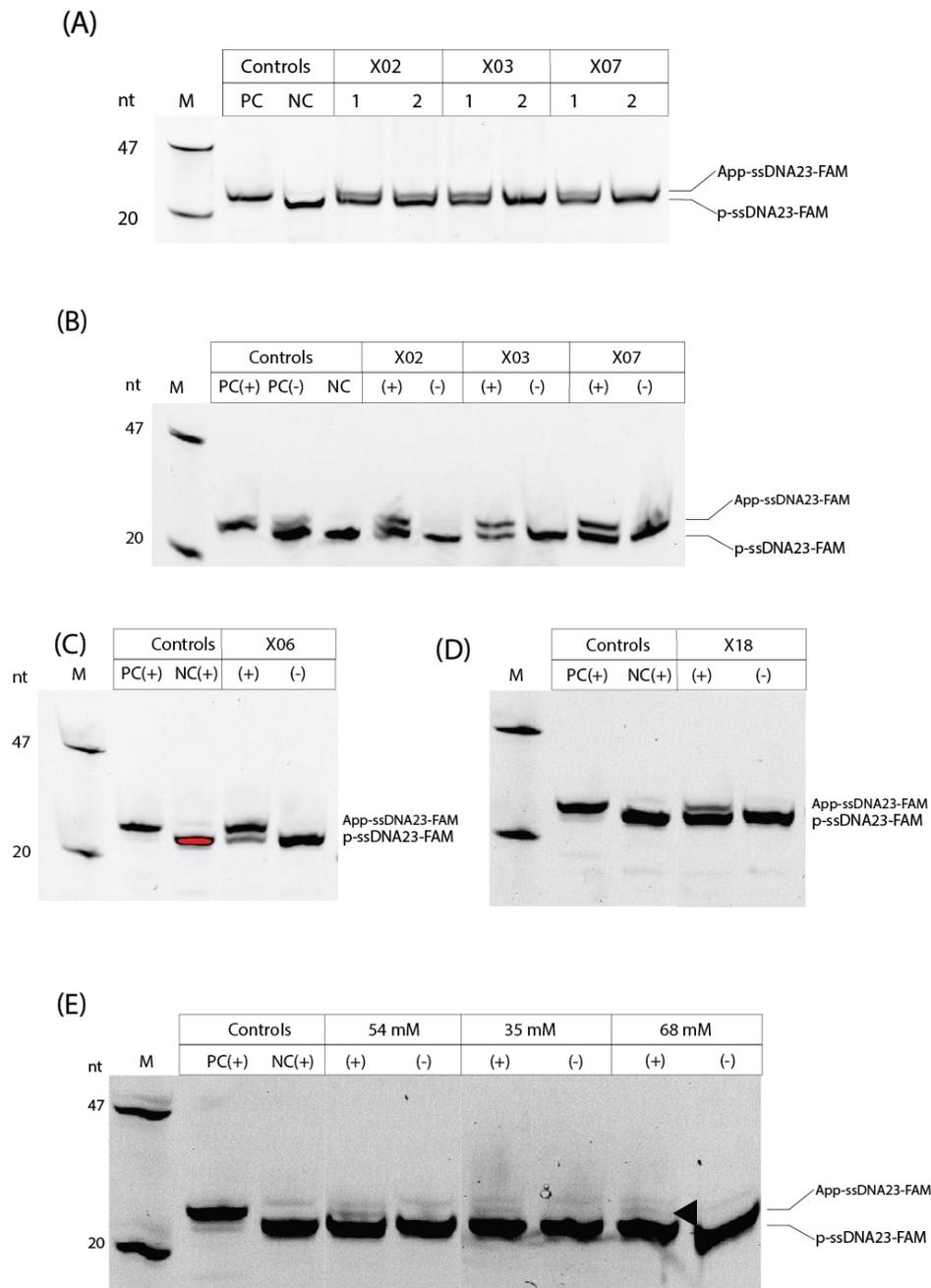
X03 from *Ferroglobus placidus*, X06 from *Methanocaldococcus infernus*, X07 from *Aquifex aeolicus*, X12 from *Aeropyrum pernix* and X18 from *P. abyssi*.

The ratio of storage buffer components (especially monovalent salts) to enzyme varied between the candidates. While trying to keep the enzyme concentration at a constant 100 pmol per assay, salt contribution to adenylyltransferase activity measurements therefore varied significantly. This could possibly have had interfering effects in some cases. The salt contributions to each assay can be found in Table 4.2. Other components of the storage buffer (e.g. MgCl<sub>2</sub>, DTT, and glycerol) varied less between experiments and are not expected to have interfering effects at these concentrations.

Table 4.2: **Initial adenylyltransferase activity screening:** Summary of storage and assay NaCl concentrations at which candidates were tested. Assays were performed with a 3'-6FAM-labelled ssDNA23-mer according to the standard protocol. Results of candidates that were tested with nuclease-degraded substrates are not included. Activity is indicated with yes (Y) or no (N).

Candidate	Storage [NaCl] (mM)	Assay [NaCl] (mM)	Activity (Y/N)
<b>X01</b>	50	25	N
<b>X02</b>	50	11	Y
	250	53	Y
<b>X03</b>	250	63	N
	50	11	Y
<b>X06</b>	50	20	Y
<b>X07</b>	250	50	N
	50	10	Y
<b>X08</b>	250	64	N
<b>X10</b>	250	163	N
<b>X12</b>	250	54	Y
	150	68	Y
	50	35	Y
<b>X13</b>	250	125	N
<b>X14</b>	250	38	N
<b>X18</b>	50	10	Y
<b>X22</b>	250	44	N

Assays with either a high or a low NaCl concentration were performed for candidates with good soluble expression. X02 was able to adenylylate ssDNA in the presence of both low (11 mM) and high (53 mM) salt. On the other hand, X03 and X07 were not active after storage (250 mM NaCl) in and activity assessment at high salt concentrations (X03: 63 mM; X07: 50 mM) (Table 4.2). X03 and X07 were active at 10-11 mM NaCl (Figure 4.3 panel A).



**Figure 4.3: Collection of positive results in initial adenylyltransferase assays.** Standard adenylyltransferase assays were performed with p-ssDNA23-FAM substrate. Positive controls (PC) were run with commercial MthRnl (NEB), while negative controls (NC) were run without added enzyme. Presence (+) and absence (-) of 100  $\mu$ M ATP is indicated. **(A)** Effects of salt on substrate adenylation by candidates X02, X03, and X07. Testing activity of samples stored in either a low (1: 50 mM) or a high (2: 250 mM) NaCl concentration, yielding final assay NaCl-concentrations of 1: 10-11 and 2: 50-60 mM. **(B)** Technical replicates of samples in (A) stored and tested in low NaCl concentration. The effect of ATP was assessed. **(C)** Assessment of candidate X06. **(D)** Assessment of candidate X18. **(E)** Assessment of candidate X12 at different assay NaCl-concentrations (indicated in mM). Product band is marked by an arrow. All samples were resolved on a 20% urea-PAGE gel and detected in the FAM-channel of a Gel Doc XR+ Gel Documentation System.

In another experiment with X02, X03, and X07 substrate adenylation in absence of ATP was assessed as a negative control. Activity was again detected, and because the gel was run longer, adenylated product bands were completely separated from the remaining substrate bands (Figure 4.3 *panel B*). An estimated 50% adenylation efficiency by X02, X03, and X07 was detected in these assays.

Adenylyltransferase activity of X06 was tested and detected once after storage in 50 mM NaCl. X06 was able to adenylate ssDNA at an assay salt concentration of 20 mM NaCl (Figure 4.3 *panel C*). Based on this single result, X06 appeared to have the highest efficiency among the six positive candidates as it adenylated approximately 80% of the donor.

Enzyme sample contamination with nuclease activity was a common problem connected to the expression host. As can be seen in Figure C.1 *panel B* (right) (Appendix C), and in a number of other samples, substrate degradation interfered with activity assays. More specifically, nuclease activity was observed in all samples that had been incubated at 37°C. E.g. X10 had been incubated at both 65°C and 37°C, but only displayed nuclease contamination when tested at 37°C (Figure C.1 *panel B and F*).

X12 was tested at three different assay salt concentrations (54 mM, 35 mM, and 68 mM NaCl) (Figure 4.3 *panel E*). In every assessment it showed the lowest activity among the positive candidates, and was barely detectable by fluorescence and below the sensitivity of SYBR Gold. Removal of imidazole by dialysis did not have an effect on adenylation by X12.

X18 was tested once and showed the second lowest efficiency when assessed in presence of 10 mM NaCl after storage in 50 mM NaCl (Figure 4.3 *panel D*).

None of the six positives (all tested at 65°C) showed signs of nuclease activity. Aside from candidates affected by nuclease-degraded substrates, the following candidates showed no adenylyltransferase activities under the conditions tested: X01, X08, X10, X13, X14, and X22. For details on the salt concentrations they had been exposed to during storage and testing see Table 4.2.

#### **4.4 Purifying Active Candidates from Up-Scaled Cultures**

Candidates X03, X06, X07, and X12 were chosen for upscaling based on their activity during initial screening and novelty. In addition, the mesophile-derived candidate X19 with high

nuclease levels i.e. no detectable activity during initial screening was chosen in order to evaluate the 2-step purification protocol for removing interfering nucleases.

*E. coli* BL21(DE3) harbouring pVB1A0B1-X03, -X06, -X07, X12 and X19 were grown in 1L cultures. Pellets were lysed and subjected to IMAC and IEX purification. In the first purification step (IMAC), proteins were eluted with a gradient of 0-300 mM imidazole and a step of 500 mM imidazole. In the second step (IEX) proteins were eluted in steps of 100, 200, 300, 500, and 1000 mM KCl. Expression and purification of the ligases was monitored by Coomassie-stained SDS-PAGE.

### **Purification of Candidate X03**

Analysis of soluble and insoluble fractions from upscaled cultures showed that X03 had been produced mostly in insoluble form, which was not observed in initial screening experiments (Figure 4.5 *panel E*). In an earlier attempt at up-scaling and purifying X03, the clarified lysate was filtered (0.45 µm), and high pressure-build up was noted. Due to this, the batch was discarded. In the second attempt (presented here) lysate was not filtered and instead applied directly onto a HisTrap FF (5 mL) crude column and purified with an ÄKTA Avant System.

The IMAC chromatogram showed increases in absorbance in fractions eluting from 75 mM and on throughout the gradient until 300 mM imidazole (Figure 4.5 *panel A*). Another increase in absorbance (peak) was noted when increasing the imidazole concentration to 500 mM.

Samples taken from 75 mM imidazole and on were resolved by SDS-PAGE and stained with Coomassie brilliant blue. The most prominent proteins in the first fractions eluting at 75 to 100 mM imidazole were one larger ( $\approx 70$  kDa) and one smaller ( $\approx 25$  kDa) contaminant (Figure 4.5 *panel B*). A protein of the expected size for candidate X03 (44.7 kDa) was observed in fraction eluting at 125 through 500 mM imidazole. Presumed X03 eluted in two peaks, the first one at 150 mM and the second at 500 mM (Figure 4.5 *panel B*). The UV absorption spectra of fractions constituting the first peak showed stronger absorbance at 230 nm than at 280 nm. While for the fractions of the second peak absorbance was stronger at 300-310 nm which is an effect that could indicate aggregation-induced scattering (data not shown).

A balance between target protein yield and major contaminants was found, and this selection of fractions (marked by a frame in Figure 4.5 *panel B*) was pooled and used for further IEX purification using a HiTrap SP Sepharose FF (1 mL) column on an ÄKTA avant system. In IEX, protein eluted in a relatively narrow peak (protein concentration of 0.5 mg/mL), with a shoulder at 300 mM KCl (Figure 4.5 *panel C*). The shoulder was observed to contain a protein of  $\approx 25$  kDa in addition to the presumed target candidate (Figure 4.5 *panel D*). Target candidate was the most predominant band in all fractions analysed from 200 to 500 mM KCl. When resolved by SDS-PAGE, the peak fraction contained proteins of various sizes, where the sizes appeared to be multiples of each other. This pattern was observed in the three fractions constituting the exact peak of the chromatogram, and intensities of the pattern bands coincided with the peak.

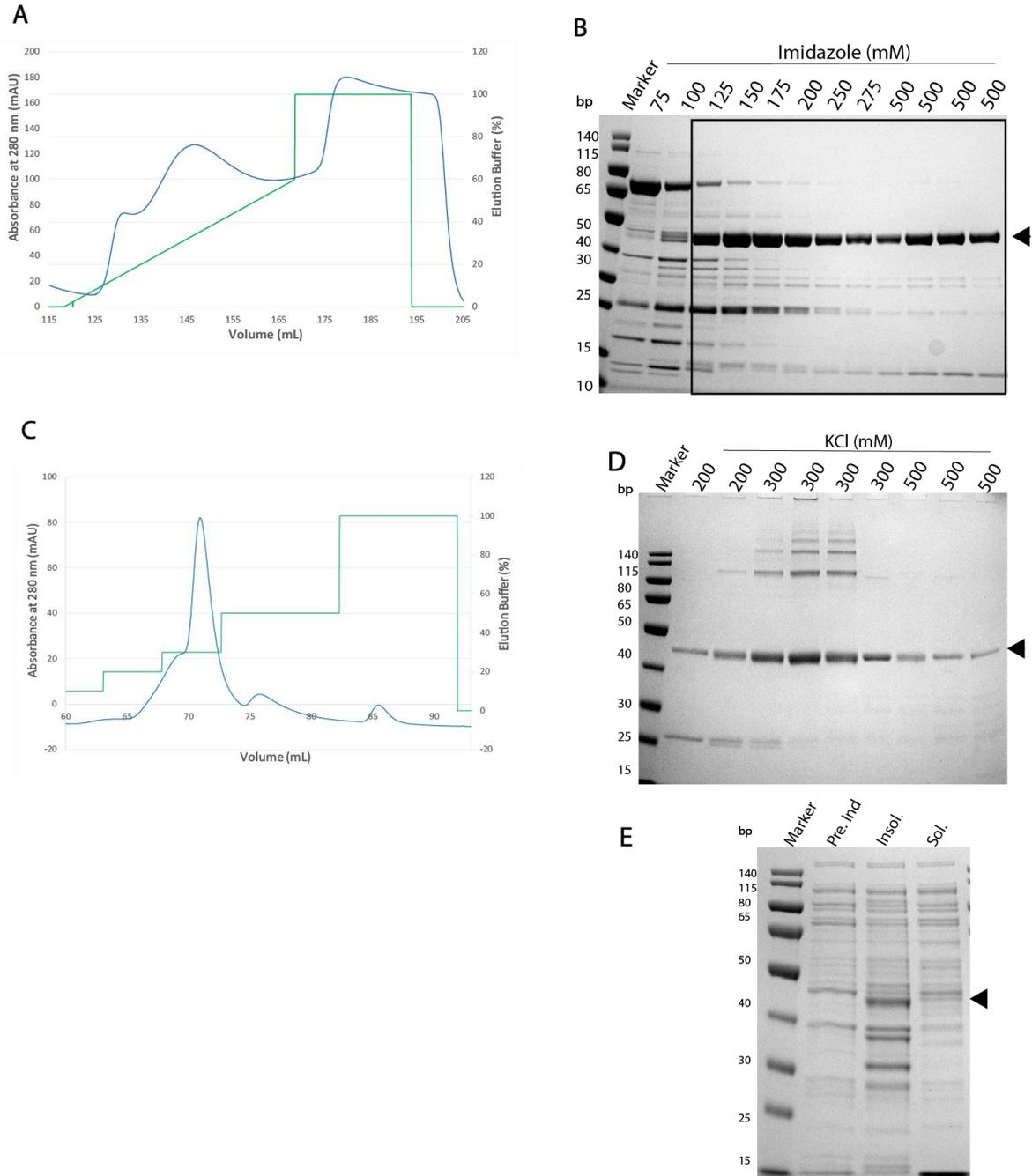


Figure 4.5: **Two-step chromatographical purification of candidate X03 using IMAC and CIEX.** For chromatograms, absorbance at 280 nm (primary x-axis; blue graph) and percentage elution buffer (secondary x-axis; green graph) is plotted against elution volume (mL). **(A)** Chromatogram from the IMAC purification. Lysate of 1 L of *E. coli* culture (sample volume 53 mL) was purified on a HisTrap 5 mL FF column and eluted with an imidazole gradient of 20 to 300 mM, with an additional step of 500 mM. **(B)** Coomassie-stained SDS-PAGE analysis of a selection of fractions from (A). Run on a SurePAGE 4-20% Bis-Tris gel. **(C)** Chromatogram from the IEX purification of IMAC fractions indicated by a square in (B) (sample volume 15 mL). Protein was eluted in steps of 100, 200, 300, 500, and 1000 mM KCl. **(D)** SDS-PAGE analysis of selected IEX fractions. Run on a NuPAGE 10% Bis-Tris gel. **(E)** Analysis of X03 expression in *E. coli* BL21(DE3). Samples were taken prior to induction (Pre. ind), and from the insoluble (Insol.) and soluble (Sol.) fraction 16 hours after induction. Run on a NuPage 10% Bis-Tris gel and stained with Coomassie blue. Marker used is PageRuler™ Prestained Protein Ladder, 10 to 180 kDa (Thermo Scientific). All gels were run in SDS-MOPS. Target protein is indicated by an arrow.

## Purification of Candidate X06

Analysis of expression in upscaled batches showed that X06 was produced primarily in soluble form, as observed during initial screening. Lysate of 100 mL from an upscaled culture was clarified by centrifugation and filtration through a 0.45  $\mu\text{m}$ -filter. Clogging of the filter was noted. The lysate was first purified by IMAC on a HisTrap FF crude (5 mL) column on an ÄKTA Avant System, and then purified manually through a SP Sepharose FF (1 mL) column.

The IMAC chromatogram did not show a clear peak above background, besides a slight increase in absorbance at 150 mM imidazole (Figure 4.6 *panel A*). A number of fractions eluted at 500 mM imidazole contained white precipitate (Figure 4.6 *panel E*). IMAC fractions eluting at 500 mM imidazole were pooled and centrifuged in order to remove precipitate, before being IEX purified. Based on Coomassie-stained SDS-PAGE analysis, the predominant protein in the IMAC fractions was approx. 40-50 kDa, coinciding with the expected molecular weight of candidate X06 (47.1 kDa) (Figure 4.6 *panel B*). An additional smaller ( $\approx 25$  kDa) and larger ( $\approx 100$  kDa) protein could also be observed.

During IEX purification, protein began eluting at 300 mM KCl (0.2 mg/mL protein on average), with complete elution during the first two column volumes of 500 mM KCl (corresponding to 500 mM-1 in Figure 4.6 *panel C*). These fractions contained 1.5 and 1.2 mg/mL protein, respectively. As for candidate X03, a ladder of suspected target protein multimers was observed in the peak fraction. Another observation from SDS-PAGE analysis of IEX fractions was that two bands of nearly the same size and intensity migrated together and match the expected size of X06.

Coomassie blue staining along the outside of wells was noted during SDS-PAGE analysis of IMAC and IEX fractions (Figure 4.6 *panel B and C*).

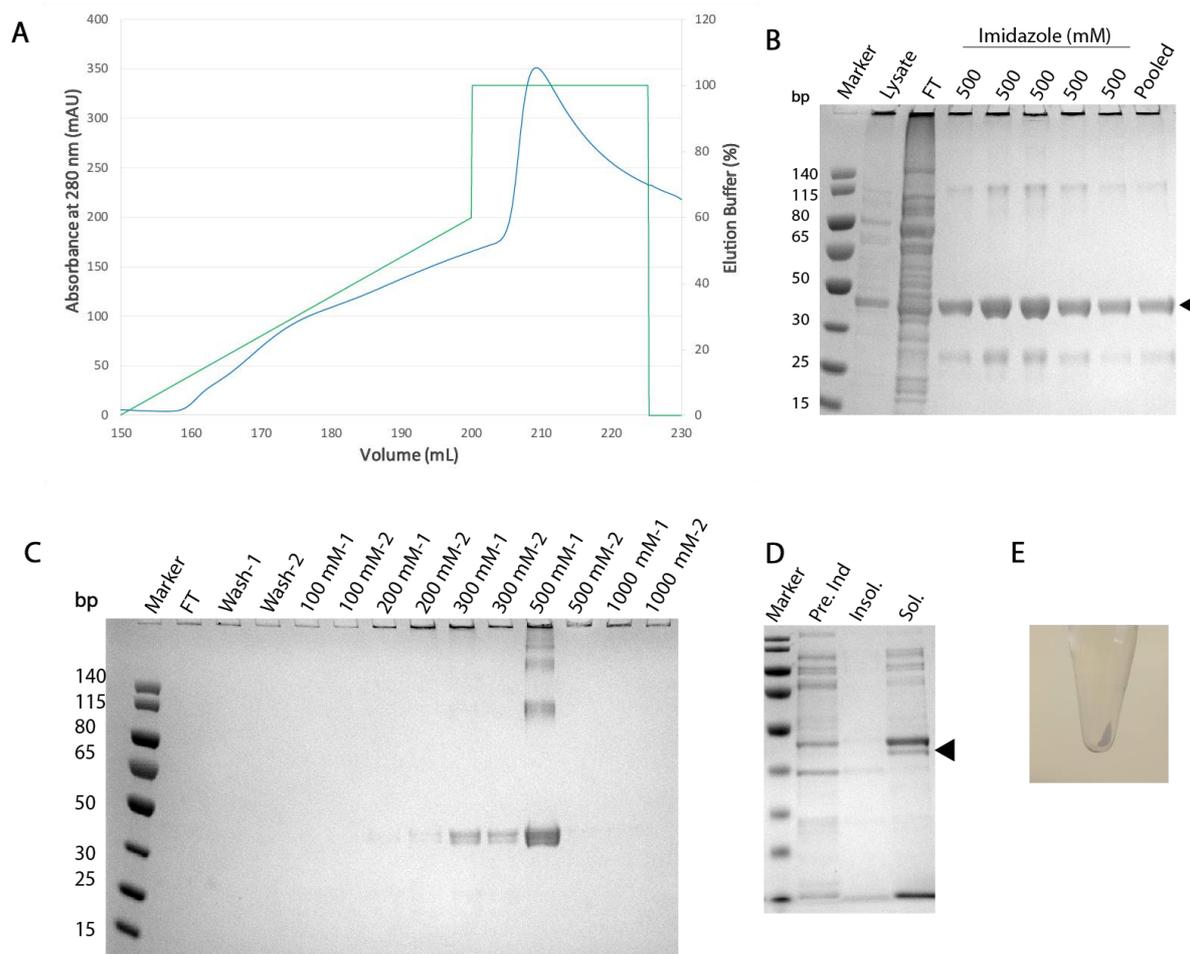


Figure 4.6: **Two-step chromatographical purification of candidate X06 using IMAC and CIEX.** For chromatograms, absorbance at 280 nm (primary x-axis; blue graph) and percentage elution buffer (secondary x-axis; green graph) is plotted against elution volume (mL). **(A)** Chromatogram from the IMAC purification of candidate X06. Lysate of 1 L of *E. coli* culture (sample volume 100 mL) was purified on a HisTrap 5 mL FF column. Protein was eluted with an imidazole gradient of 20 to 300 mM with an additional step at 500 mM. **(B)** Coomassie-stained SDS-PAGE analysis of flow through (FT) and fractions eluted at 500 mM imidazole during IMAC purification, and crude lysate. Ran on a 12-20% Bis-Tris SurePAGE gel. **(C)** Coomassie-stained SDS-PAGE analysis of FT, wash and fractions eluted at 100, 200, 300, 500 and 1000 mM KCl in CIEX purification (sample volume 11 mL). Ran on a 12-20% Bis-Tris SurePAGE gel. **(D)** Analysis of X06 expression in *E. coli* BL21(DE3) in samples taken prior to induction (Pre. Ind) and 16 hours after induction – separated into the soluble (Sol.) and insoluble (Insol.) fractions. Ran on a 10% Bis-Tris NuPage gel. **(E)** Precipitate observed in IMAC fractions with high protein concentrations immediately after elution. Marker used is PageRuler™ Prestained Protein Ladder (10 to 180 kDa). All gels were run in SDS-MOPS. Target protein is indicated by an arrow.

## Purification of Candidate X07

A volume of 120 mL of clarified and filtered lysate was first purified with a HisTrap FF crude (5 mL) column on an ÄKTA Avant System. A single 2.5 mL-IMAC eluate (eluted at 150 mM imidazole) was applied to 200  $\mu$ L Q Sepharose FF slurry and IEX purified in an Eppendorf-

tube format. Expression analysis confirmed that X07 was produced mainly in soluble form, in agreement with small-scale expression results (Figure 4.7 *panel B*). From the IMAC chromatogram, protein was observed to elute with a clear, broad peak from 50 to 500 mM imidazole, reaching a maximal absorbance at 150 mM imidazole (Figure 4.7 *panel A*). Fractions eluting during IMAC were not analysed by SDS-PAGE.

Unlike X03 and X06, no pressure build-up was noted during filtration of X07 lysate. The IMAC-fraction corresponding to the highest absorbance (eluting at 150 mM imidazole) was further purified by IEX. Absorbance measurements of IEX fractions showed moderate protein concentrations in the flow through (1.6 mg/mL protein) and wash (0.5-1.0 mg/mL) fractions. Bound protein began eluting at 100 mM KCl (0.5 mg/mL protein), peaked at 200 mM KCl (1.7 mg/mL protein) and declined until 500 mM KCl. No detectable protein was eluted at 1 M KCl (Figure 4.7 *panel C*).

Due to concentration differences in samples analysed by SDS-PAGE, it cannot be judged whether purity was improved after IEX-purification. A prominent 25-kDa contaminant persisted throughout the entire elution in IEX and appeared to have the same affinity for the SP Sepharose resin as the target protein. All protein concentration measurements agreed with the Coomassie-stained SDS-PAGE analysis of these fractions, i.e. the peak fraction had the highest protein concentration.

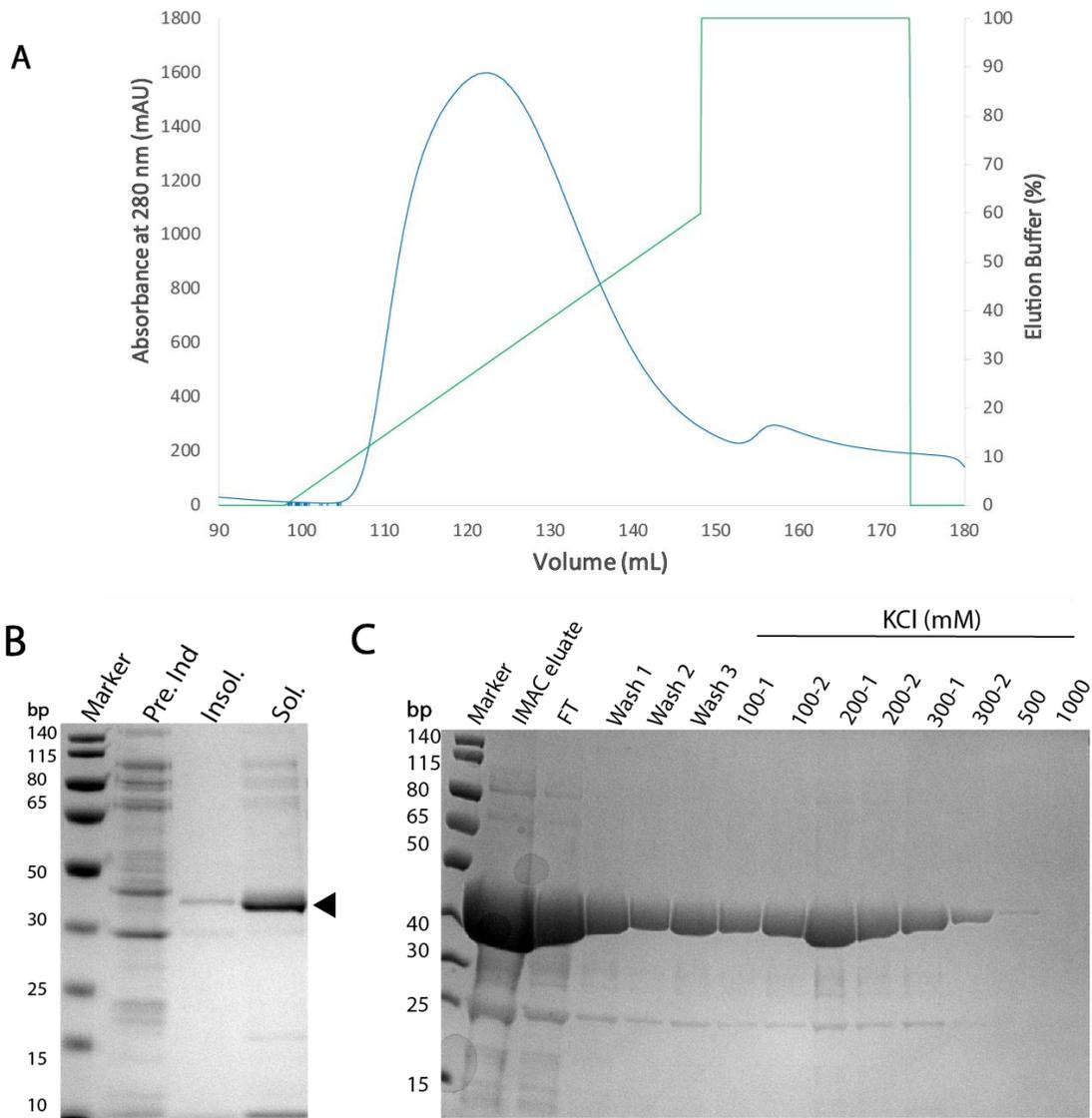


Figure 4.7: **Two-step chromatographical purification of candidate X07 using IMAC on ÄKTA Avant and AIEX in Eppendorf tube format.** For chromatograms, absorbance at 280 nm (primary x-axis; blue graph) and percentage elution buffer (secondary x-axis; green graph) is plotted against elution volume (mL). **(A)** Chromatogram from the IMAC purification of candidate X07. Lysate of 1 L of *E. coli* culture (sample volume 120 mL) was purified on a HisTrap 5 mL FF column. Protein was eluted with an imidazole gradient of 20 to 300 mM with an additional step at 500 mM. **(B)** Analysis of X07 expression in *E. coli* BL21(DE3) prior to induction (Pre. Ind) and 16 hours after induction, in the soluble (Sol.) and insoluble (Insol.) fractions. **(C)** Coomassie-stained SDS-PAGE analysis of IMAC eluate (a single peak fraction eluting at 150 mM imidazole), IEX flow through (FT), wash, and eluates at 100, 200, 300, 500, and 1000 mM KCl (sample volume 2.5 mL). Samples were run on a 12-20% Bis-Tris SurePAGE in SDS-MOPS. Markers are PageRuler™ Prestained Protein Ladder (10 to 180 kDa). Target protein is indicated by an arrow.

## Purification of Candidate X12

When expressed in the second, upscaled batch X12 was produced in soluble and insoluble form in equal parts, which is in agreement with its initial evaluation as moderately soluble

(Figure 4.8 *panel E*). Unfiltered lysate was first purified on a HisTrap FF column (5 mL) resulting in an elution profile with two peaks (Figure 4.8 *panel A*). The first eluting at approx. 125 mM, and the second at 500 mM imidazole. Between these peaks, baseline was not reached (Figure 4.8 *panel A and B*).

Coomassie-stained SDS-PAGE analysis of a selection of the IMAC fractions revealed that the shoulder of the chromatogram contained impurities. A major contaminant of  $\approx 70$  kDa was eluted with a peak at 75 mM imidazole, but persisted in fractions eluted at 500 mM imidazole (Figure 4.8 *panel B*). This contaminant was not completely eluted within 500 mM imidazole. Other contaminants were present in moderate to large amounts (e.g. a 25 kDa contaminant) in all eluted fractions containing the target protein.

From IMAC eluate, fractions eluted between 100 and 500 mM (indicated with a frame in Figure 4.8 *panel B*) were pooled and further IEX purified with an SP sepharose FF (1 mL) column and a step gradient of 100, 200, 300, 500 and 1000 mM KCl. Protein was eluted starting at 200 mM and continuing until 1 M KCl. Doublet peaks ( $\approx 1$  mg/mL protein) were detected in the chromatogram (Figure 4.8 *panel C*). It appeared that one eluted at 200 and the other at 300 mM KCl. Fractions were assessed by Coomassie-stained SDS-PAGE and revealed that the major protein matched the expected size of X12 (44.1 kDa) (Figure 4.8 *panel D*). One smaller contaminant (25 kDa) was clearly present in all fractions but was mostly removed by 300 mM KCl.

An interesting observation in the IEX fractions was that the target protein was joined by another protein of highly similar size, creating a double band. The intensity of these bands correlated strongly. As also observed in the IMAC purification, elution of target protein increased again at the maximal eluent concentration (1 M KCl), forming an additional, poorly resolved peak (Figure 4.8 *panel C*).

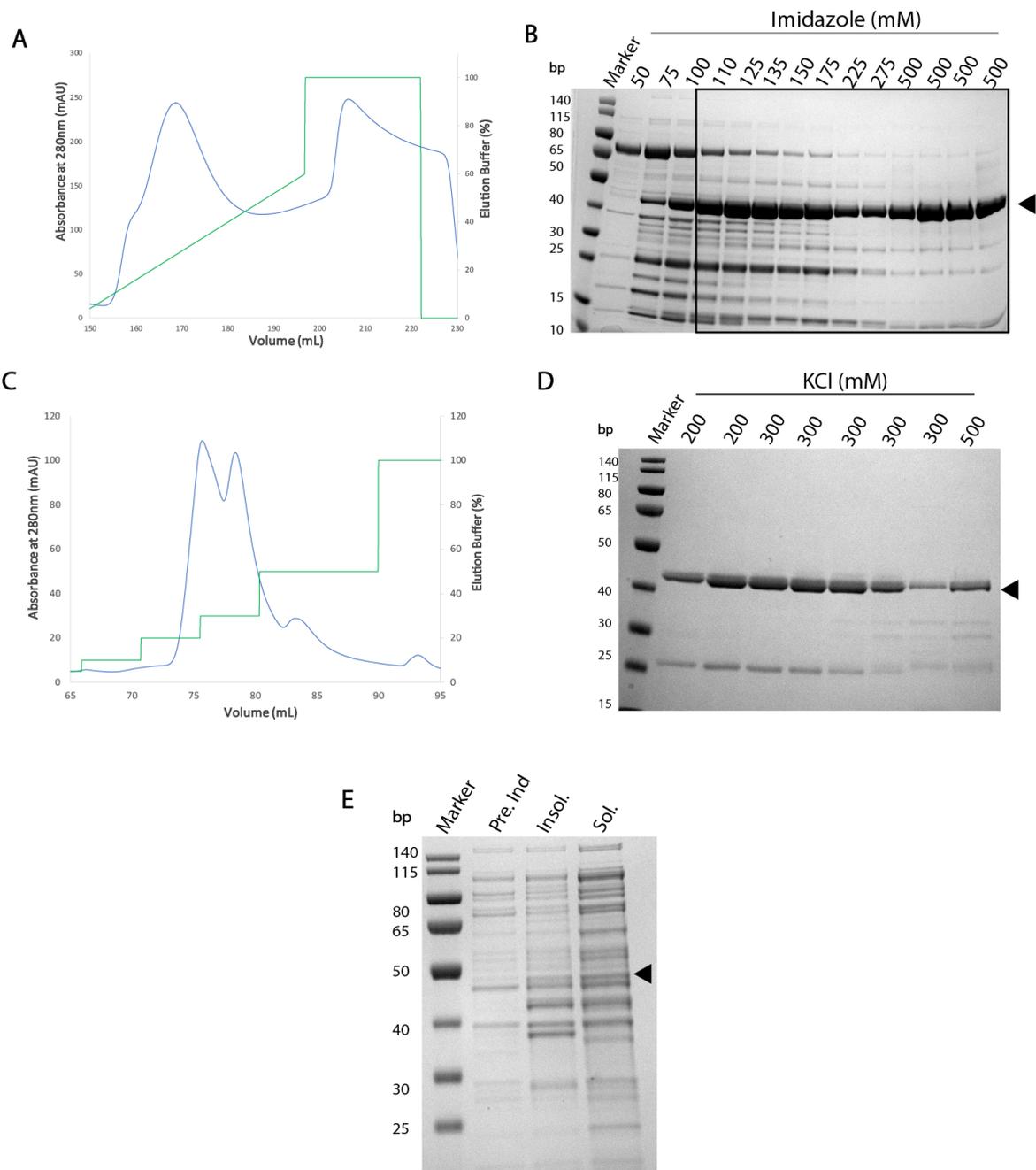


Figure 4.8: **Two-step chromatographical purification of candidate X12 using IMAC and CIEX.** For chromatograms, absorbance at 280 nm (primary x-axis; blue graph) and percentage elution buffer (secondary x-axis; green graph) is plotted against elution volume (mL). **(A)** Chromatogram from the IMAC purification of candidate X12. Lysate of 1 L of *E. coli* culture (sample volume 83 mL) was purified on a HisTrap 5 mL FF column. Protein was eluted with an imidazole gradient from 20 to 300 mM with an additional step at 500 mM. **(B)** Coomassie-stained SDS-PAGE analysis of a selection of IMAC fractions. Fractions pooled for further IEX purification are indicated by a frame. Ran on a SurePAGE 4-20% Bis-Tris gel. **(C)** Pooled IMAC fractions (sample volume 17.5 mL) were IEX purified using a SP Sepharose FF column. Protein was eluted in steps of 100, 200, 300, 500, and 1000 mM KCl. **(D)** Coomassie-stained SDS-PAGE analysis of selected IEX fractions. Target protein is marked with an arrow. Ran on a NuPAGE™ 10%, Bis-Tris gel. **(E)** Analysis of X12 expression in *E. coli* BL21(DE3). Samples were taken prior to induction (Pre. ind), and from the insoluble (Insol.) and soluble (Sol.) fraction 16 hours after induction. All gels were run in MOPS-SDS buffer. Marker used is PageRuler™ Prestained Protein Ladder (10 to 180 kDa). Target protein is indicated by an arrow.

## Purification of Candidate X19

X19 was not among the candidates expressed highly in soluble form or showing activity in initial activity screens. However, X19 was used to test the standard purification protocol. This candidate originates from a mesophile, hence initial adenylation screens were hindered by nuclease activity. Expression analysis of X19 in the first batch indicated poor solubility and was not analysed in the second batch.

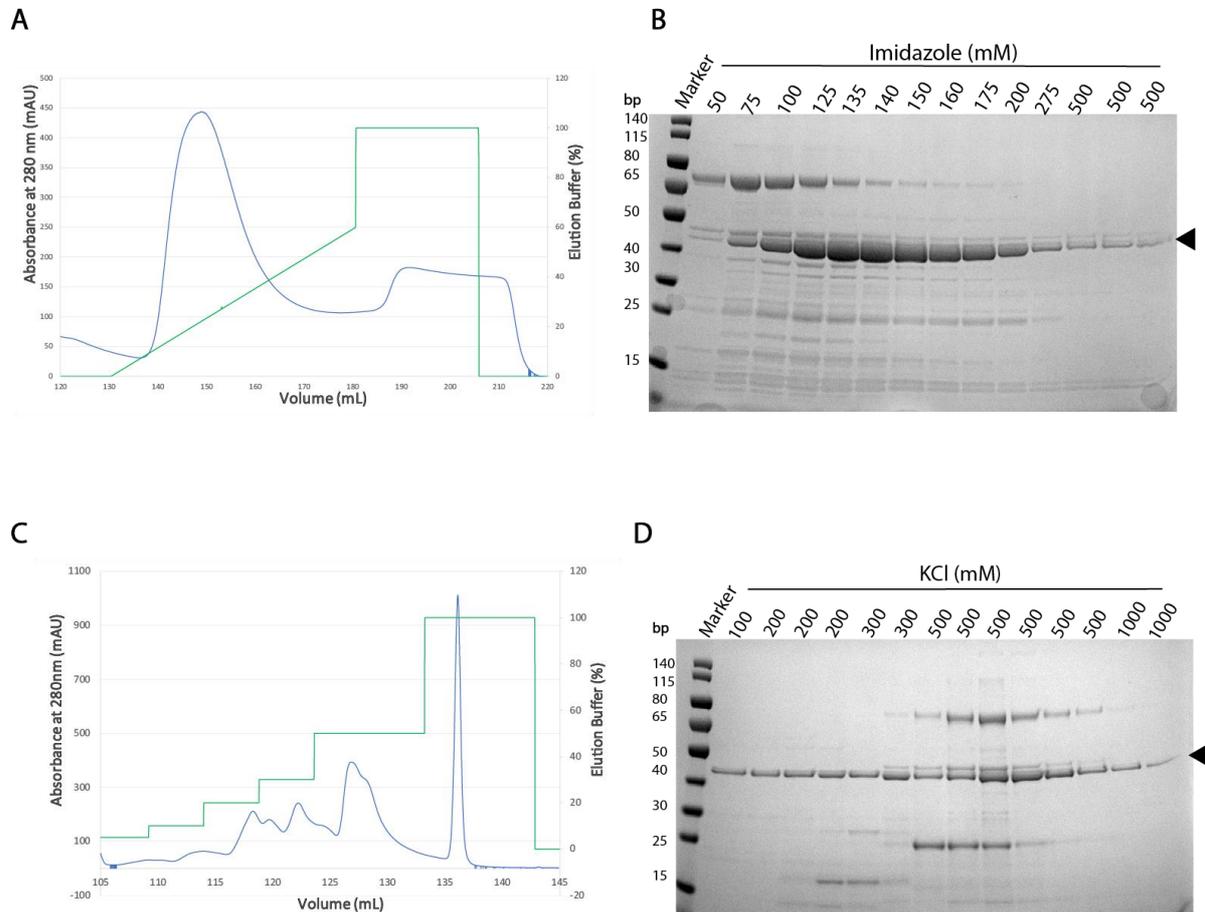
Lysate was purified using a HisTrap FF crude (5 mL) column without prior filtration. Protein began eluting from 75 mM throughout 200 mM imidazole (Figure 4.9 *panel A*). The major protein in fractions eluted at 125 mM and upwards was of the expected size for X19 (42.8 kDa) (Figure 4.9 *panel B*).

In early-eluting fractions, a protein of  $\approx 70$  kDa was observed by Coomassie-stained SDS-PAGE analyses (Figure 4.9 *panel B*). This presumed contaminant eluted with a peak at 75 mM imidazole, but slightly overlapped the elution peak of X19. Additional smaller proteins, e.g. one at 25 kDa eluted with similar profiles as X19 (Figure 4.9 *panel B*). It was noted that two proteins matching the size expected for X19 eluted together during IMAC. The slightly faster migrating protein appeared in intense bands, while the slower migrating one only appeared as a faint band. However, their intensities throughout the elution peak were similar.

All fractions eluted from IMAC were pooled and IEX purified on a HiTrap HP Q Sepharose (1 mL) column with a KCl step elution. The elution profile obtained from the chromatogram showed peaks at every KCl-step, except 100 mM KCl (Figure 4.9 *panel C*). Doublet peaks were observed at 200, 300, and 500 mM KCl. At 1 M KCl protein eluted with a narrow, clearly resolved peak.

Coomassie-stained SDS-PAGE analysis of these fractions revealed the presence of presumed X19 in all fractions (Figure 4.9 *panel D*). Intensity of the X19-band was maximal in fractions eluted at 500 mM KCl. However, as LDS is precipitated at higher KCl concentrations, the intensity of bands at higher KCl (especially 1 M) cannot be assumed to be accurate. Two main contaminants of 15 and 25 kDa eluted at 200 and 500 mM KCl, respectively. Hence, the elution profile of the 25 kDa protein overlapped with that of presumed X19.

Fractions constituting the X19 peak at 500 mM KCl also contained a slightly slower migrating species ( $\approx 40$ -50 kDa) and a 70 kDa protein. Elution of these two species is correlated, and follow the elution profile of the X19-band.



**Figure 4.9: Two-step purification of X19 using IMAC and AIEX** For chromatograms, absorbance at 280 nm (primary x-axis; blue graph) and percentage elution buffer (secondary x-axis; green graph) is plotted against elution volume (mL). **(A)** Chromatogram from the IMAC purification of candidate X19 (blue graph). Lysate of 1 L of *E. coli* culture (sample volume 62.5 mL) was purified on a HisTrap 5 mL FF column. Protein was eluted with an imidazole gradient of 20 to 300 mM with an additional step at 500 mM (green graph). **(B)** Coomassie-stained SDS-PAGE analysis of selected IMAC fractions. Ran on a 12-20% Bis-Tris gel. **(C)** Chromatogram from the AIEX purification of candidate X19 (sample volume 45 mL) on a HiTrap HP Q sepharose (1 mL). **(D)** Coomassie-stained SDS-PAGE analysis of selected AIEX fractions. Ran on a 12-20% Bis-Tris gel. All gels were run in SDS-MOPS. Marker used is PageRuler™ Prestained Protein Ladder (10 to 180 kDa). Target protein is indicated by an arrow.

## 4.5 Confirming Adenylyltransferase Activity in Upscaled and Purified Batches

A selection of *E. coli* clones expressing active candidates (X03, X06, X07 and X12) and X19 were grown in larger scale (1 L). Standard adenylyltransferase assays confirmed the ability of all four candidates (stored in 50% elution buffer and 50% glycerol) to produce slower migrating, adenylylated 23-mer ssDNA. No product was observed in the absence of ATP. X19 showed slight adenylation activity after purification by IMAC and IEX.

Candidate X03 showed 50% adenylation activity in the IEX purified sample (collected from the fraction eluting at 300 mM KCl). Only very little activity was detected in the sample obtained after IMAC purification (eluting at 500 mM imidazole) (Figure 4.10 *panel A*). Assay salt concentrations were 9 mM NaCl (and 90 mM imidazole) and 113 mM KCl after IMAC and IEX purification, respectively.

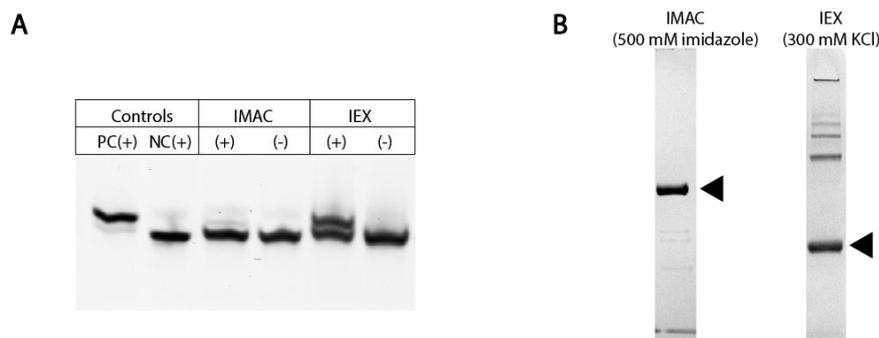


Figure 4.10: **Confirming adenylyltransferase activity in X03 produced in upscaled cultures and purified by IMAC and IEX.** **(A)** Activity on ssDNA was tested in glycerol stocks prepared from an IMAC and IEX fraction. Samples were resolved on a 20% urea-PAGE gel and stained with SYBR Gold. Commercial wild-type MthRnl is used in the positive control (PC), and enzyme is omitted from the negative control (NC). All reactions were run with (+) and without (-) ATP. **(B)** Fractions used for glycerol stocks resolved on an SDS-PAGE gel and stained with Coomassie blue. Concentration of eluent at which the sample was obtained is shown in parenthesis. Target protein is indicated by arrows. For full results see Section 4.4.

For X07, glycerol stocks were prepared from the peak fraction in IMAC (eluted at 150 mM imidazole) and the fraction with the highest protein concentration after IEX (eluted at 200 mM KCl) (Figure 4.11 *panel B*). When testing X07 after IMAC purification the NaCl-concentration was adjusted to 10 mM, as in the initial activity screens. The IEX purified sample was used as it was, yielding a final 27 mM KCl in the adenylyltransferase assay. Adenylation of more than 50% of the substrate was confirmed in both samples of X07.

Adenylation efficiency appeared to be slightly higher in the assessment of the IMAC fraction, compared with the IEX fraction. Adenylation efficiencies were unchanged from the initial activity screening in the first batch (Figure 4.3 *panel B*).

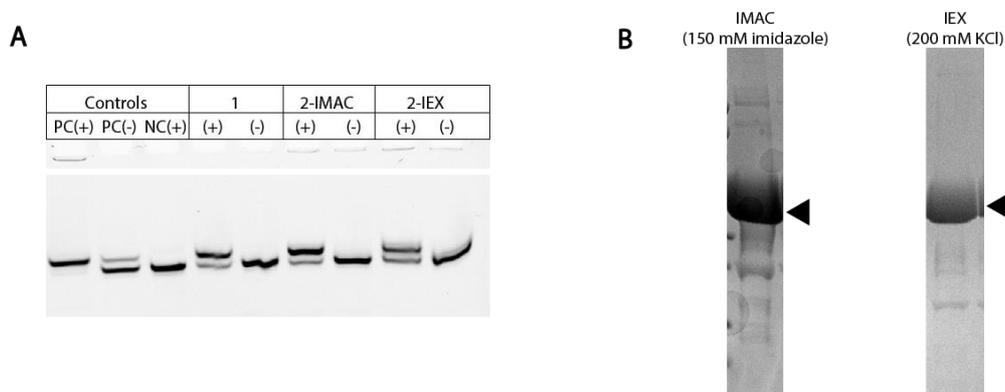
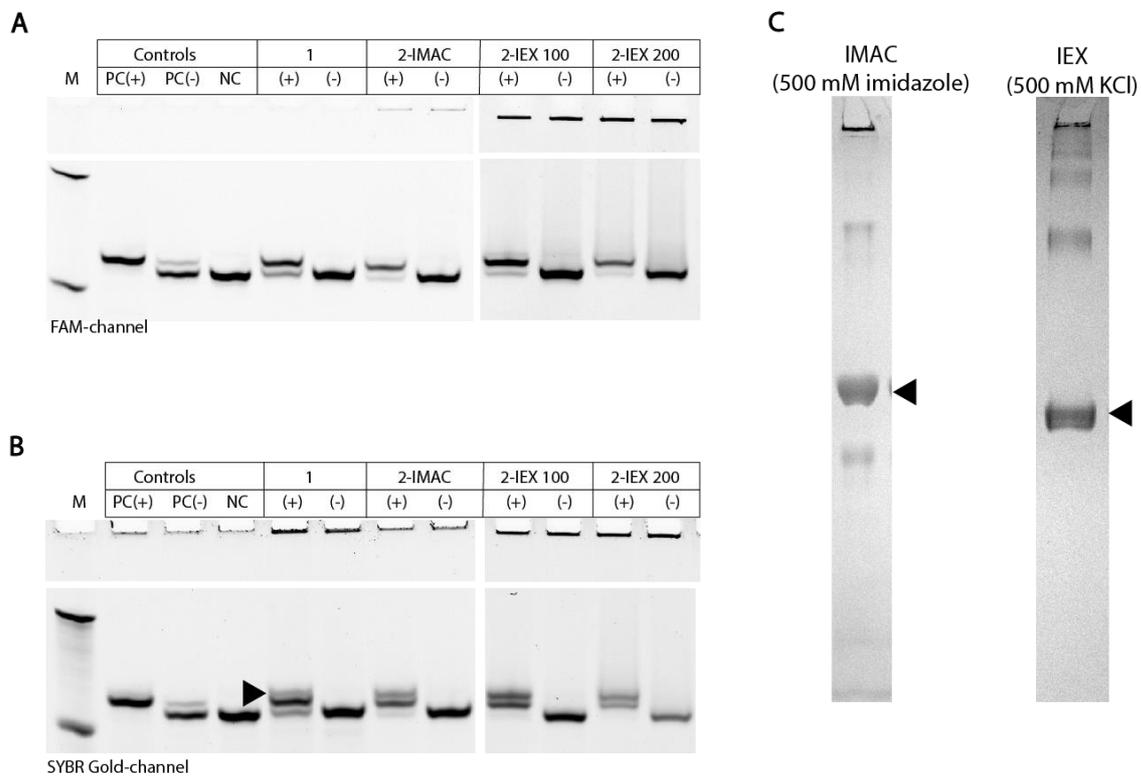


Figure 4.11: **Confirming adenylyltransferase activity of X07 produced in upscaled cultures and purified by IMAC and IEX.** (A) Detection of standard adenylyltransferase assay results (using a 3'-6FAM labelled ssDNA donor) in the FAM channel after fractionation on a 20% Urea-PAGE gel. The sample used in initial activity screening was included as a control (1=first batch). Commercial wild-type MthRnl is used in the positive control (PC), and enzyme is omitted from the negative control (NC). All reactions were run with (+) and without (-) ATP. (B) Coomassie-stained SDS-PAGE analysis of the fractions from which IMAC and IEX glycerol stocks had been prepared. Concentration of eluent at which the fractions were obtained in parenthesis. Target protein is indicated by arrows.

Glycerol stocks were prepared from X06 eluted at 500 mM imidazole in IMAC, and 500 mM KCl in IEX (Figure 4.12 *panel C*). The tested X06 IMAC fraction contained a final assay concentration of 62.5 mM imidazole and 6 mM NaCl. Product bands obtained from testing the first, small scale batch (denoted as 1 in Figure 4.12 *panels A and B*) and the IEX and IMAC-purified product of the second batch (denoted as 2) were similar in intensity. DNA was detected in wells of all reactions run with X06. Due to this, adenylation efficiency could not be estimated with confidence, but appeared to be between 50 and 90%. Here, two enzyme concentrations (100 and 200 pmol) were tested for the IEX-purified sample, resulting in final assay concentrations of 80 and 160 mM KCl. Band intensities were reduced when using 200 pmol, and an additional smear throughout the entire gel was observed for both IEX assays. These observations were made in both the FAM and SYBR Gold channel. An additional, unexpected observation of an extra band was made in X06 from the first and second batch after staining the gel with SYBR Gold (Figure 4.12 *panel B*; indicated with an arrow). This product was not present in the assay without ATP and migrated slightly slower than the

expected band seen in the FAM channel. Commercial MthRnl, used in the positive control, did not produce this band.



**Figure 4.12: Confirming adenylyltransferase activity of X06 produced in upscaled cultures and purified by IMAC and IEX.** Ability of X06 to adenylylate p-ssDNA23-6FAM was assessed with the standard adenylyltransferase assay. **(A)** Detection of adenylyltransferase assay results in the FAM-channel and **(B)** after SYBR Gold staining. The sample obtained from the first, small scale batch was included as a control. Numbers indicate the first, small scale batch (=1) and the second, upscaled batch (=2). Of the IEX-purified product 100 and 200 pmol was tested. Commercial wild-type MthRnl is used in the positive control (PC), and enzyme is omitted from the negative control (NC). All reactions were run with (+) and without (-) ATP. Arrow indicates the additional unexpected band. **(C)** Coomassie-stained SDS-PAGE analysis of the fractions from which IMAC and IEX glycerol stocks had been prepared. Concentration of eluent at which the fractions were obtained are in parenthesis. Target protein is indicated by arrows.

Regarding X12, 100 pmol IMAC and IEX purified enzyme resulted in final assay concentrations of 100 mM imidazole and 10 mM NaCl (IMAC purified sample) and 60 mM KCl (IEX purified sample). Adenylyltransferase activity was only confirmed in the IEX-purified fraction (Figure 4.13 panel A).

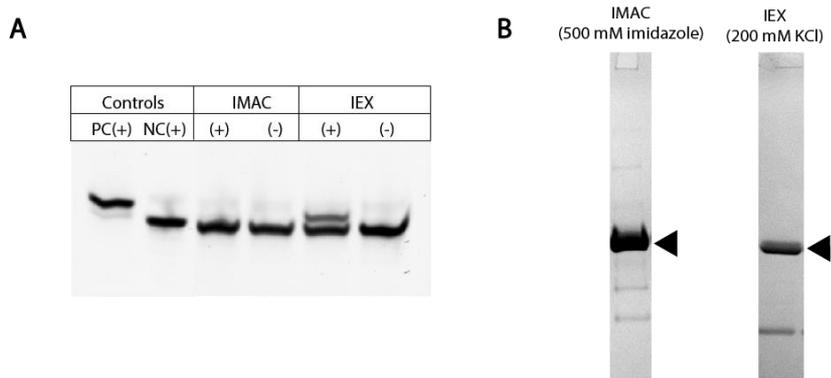


Figure 4.13: **Confirming adenylyltransferase activity of X12 produced in upscaled cultures and purified by IMAC and IEX.** **(A)** Resolving adenylyltransferase assay results on a denaturing 20% Urea-PAGE gel and detecting ssDNA with SYBR Gold staining. Commercial wild-type MthRnl is used in the positive control (PC), and enzyme is omitted from the negative control (NC). All reactions were run with (+) and without (-) ATP. **(B)** Coomassie-stained SDS-PAGE analysis of the fractions from which IMAC and IEX glycerol stocks had been prepared. Concentration of eluent at which the fractions were obtained are given in parenthesis. Target protein is indicated by arrows.

The mesophile-derived X19 was tested for adenylyltransferase activity only after IEX purification, because earlier tests (initial screening) had shown that IMAC could not remove enough of the contaminating nuclease activity. Glycerol stocks were prepared from each of the double peaks at 200, 300 and 500 mM and the narrow peak at 1000 mM KCl (Figure 4.14 *panel B*). Adenylyltransferase activity was barely detectable in fractions eluting at 200, 300 and 500 mM KCl (Figure 4.14 *panel A*). In these fractions, nuclease activity was noted by the presence of degradation products. In the sample of protein eluting at 500 mM KCl an additional slower migrating band was produced. Here, as opposed to samples from 200 and 300 mM KCl, presence of DNA was detected in the wells. The sample taken from 1000 mM KCl contained multiple slowly migrating bands, no detectable adenylylated product, and a strong smear. Little DNA was detected in the wells of this sample.

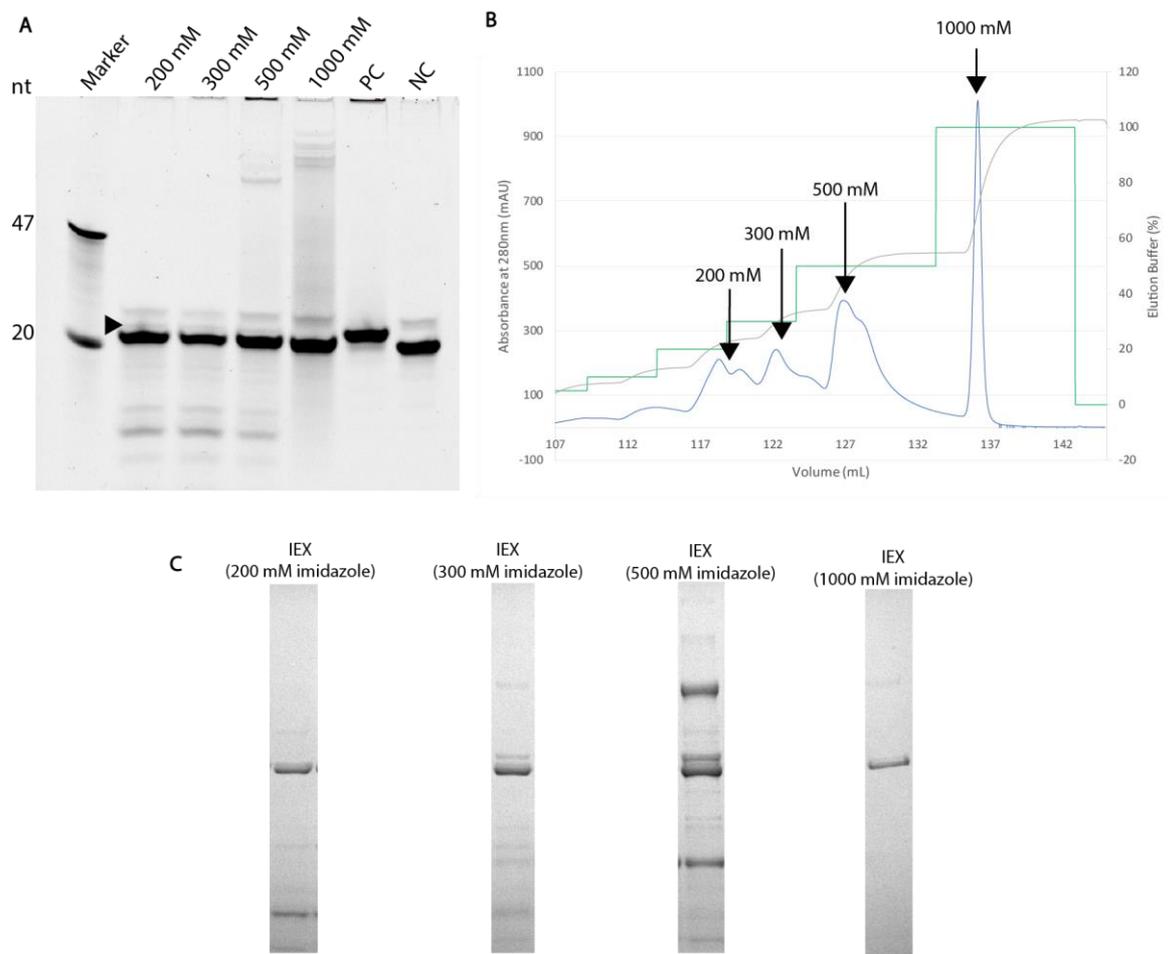


Figure 4.14: **Assessing adenylation activity in upscaled and IEX purified X19.** (A) Results from the standard adenylyltransferase assay were resolved on a 20% Urea-PAGE gel and detected with SYBR Gold. Glycerol stocks tested had been prepared from each of the peak fractions eluted during IEX. Presumed adenylylated donor is indicated by an arrow. (B) Overview of the IEX chromatogram with arrows indicating the fractions investigated for activity. (C) Coomassie-stained SDS-PAGE analysis of the fractions from which IEX glycerol stocks had been prepared. Concentration of eluent at which the fractions were obtained are given in parenthesis.

## 4.6 Assessing Ligation Activity in Upscaled and Purified Batches

After confirming that the candidates were active in the second batch their abilities to catalyze the final step of the ligation process was assessed. All ligation assays were conducted using the IMAC- or IEX-purified glycerol (50% v/v) stocks with confirmed step 1 and 2 activity as described in the previous section. Two approaches for testing ligation activity were developed. In one, amount of step 2 product was controlled by using a commercial kit for the first two steps (“isolated step 3 assay”). In the other, the candidates were tested for activity in the entire ligation process in a single-tube “composite ligation assay”. In both assays, ligation

(step 3) was assessed after depleting ATP with a phosphatase treatment, thereby avoiding the inhibitory effect of ATP on step 3. X03, X06, X07, and X12 were tested in the isolated step 3 assay. X06 and X07 were in addition tested in the composite ligation assay.

### Testing the Composite Ligation Assay

In the workflow of the “*composite ligation assay*”, recombinant shrimp alkaline phosphatase (rSAP; ArcticZymes) was used to degrade remaining ATP after the first two steps of the ligation process. rSAP is active at 37°C and is inactivated at 65°C, hence rSAP becomes inactivated concurrently with the ligation step when testing thermotolerant RNA ligases. Compared to the similar existing protocols, the ATP-degrading step makes it possible to perform the entire ligation reaction without the need for an ATP-independent RNA ligase such as mutant MthRnlK97A (#M0319; NEB). As a proof of concept this “*composite ligation assay*” showed that commercial WT MthRnl could ligate the 6FAM-ssDNA23-OH acceptor to the p-ssDNA23-ddC donor after rSAP treatment (Figure 4.15 *panel B right*). Without the rSAP treatment, ligation product was not detectable.

In Figure 4.15, detected by SYBR Gold staining, the fastest migrating band is the donor, the band migrating slightly slower is the 6FAM-labelled acceptor, and the slowest migrating band is an artefact (observed in the oligo manufacturer’s HPLC chromatogram).

Ligation activity of candidate X07 with a ssDNA acceptor and donor pair was investigated using the composite ligation assay (Figure 4.15 *panel B*). Only the IMAC sample was used, because it was more concentrated and the IEX step did not appear to visibly improve purity (Figure 4.7 *panel C*). For X06, glycerol stocks of IEX- purified product (See Section 4.5) were tested with both 23-mer ssDNA and ssRNA acceptors. Polyethylene glycol was added to the samples because crowding agents have improved step 3 activity of other ligases in the past (Hayashi *et al.*, 1986).

For both X06 and X07, no App-donor formation could be detected in an aliquot which had been withdrawn immediately after addition of enzyme on ice (“before adenylation”) (Figure 4.15 *panels B and C*). After incubation at 65°C for 1 hour, the positive control (WT MthRnl) had adenylated nearly 100% of the substrate in the X07 experiment, while X07 had adenylated ≈60%. In the X06 experiment MthRnl and X06 achieved the same adenylation efficiency of ≈80%. Phosphatase treatment caused a charge-related upward shift of the p-

donor band in all samples (Figure 4.15 *panel B middle and C*). rSAP and mock treated samples were incubated further for 2 hours at 65°C, after which a slower migrating band ( $\approx 47$  nt) was detected only in the positive controls containing MthRnl treated with rSAP (Figure 4.15 *panel B right and C both*). De-adenylation-related reappearance of the donor band was observed in both rSAP-treated samples of X07, and in MthRnl and X06 samples. De-adenylation is the reverse step 2 reaction and is seen as a downward shift of the App-donor band in the presence of ligase.



## Testing Ligation of Pre-Adenylated Donors

Ligation activities of X03, X06, X07, and X12 were also assessed in the “isolated step 3 assay”. First, a protocol for using pre-adenylated donors as starting points was developed. The commercially available 5'-DNA pre-adenylation kit with MthRnl was employed to prepare a 5'-phosphorylated 23-mer ssDNA donor for step 3. rSAP was used to degrade remaining ATP similar to the composite ligation assay (Figure 4.18). This protocol is less time consuming, does not require extra equipment, and does not influence yield in a negative way, compared to the methods recommended by the manufacturer of the 5'-DNA pre-adenylation kit.

An important question was whether MthRnl could be irreversibly inactivated in the preadenylated donor sample, which was to be directly used after rSAP treatment. In an initial test for residual activity MthRnl was found to be inactivated by heating at 85°C for 3 min, as recommended by the manufacturer (NEB). This was also confirmed in the negative controls of experiments (Figure 4.16). Later experiments also emphasized the importance of degrading MthRnl with Proteinase K in order to avoid any residual activity.

The positive control could produce bands of the expected size for ligated product ( $\approx 47$  nt). No such bands were detected in the reactions run with active candidates (Figure 4.16). Strong de-adenylation activity by X07 and X06 was noted when incubating them with preadenylated donors in absence of ATP (Figure 4.16 *panel A and B*). De-adenylation was also prominent during the ligation reactions. In reactions run with X06 and X07, all App-donors were mostly deadenylated (Figure 4.16 *panel A and B*). De-adenylation was not complete in MthRnl, X03, and X12 (Figure 4.16). When testing ligase activity on ssRNA acceptors alone, the intensity of the acceptor bands was reduced when using X06, X03, and X12 (Figure 4.16 *panel B and C*). In addition, RNA was deposited in the wells when running these acceptor+ligase controls. This was not observed when running the acceptor alone (data not shown).

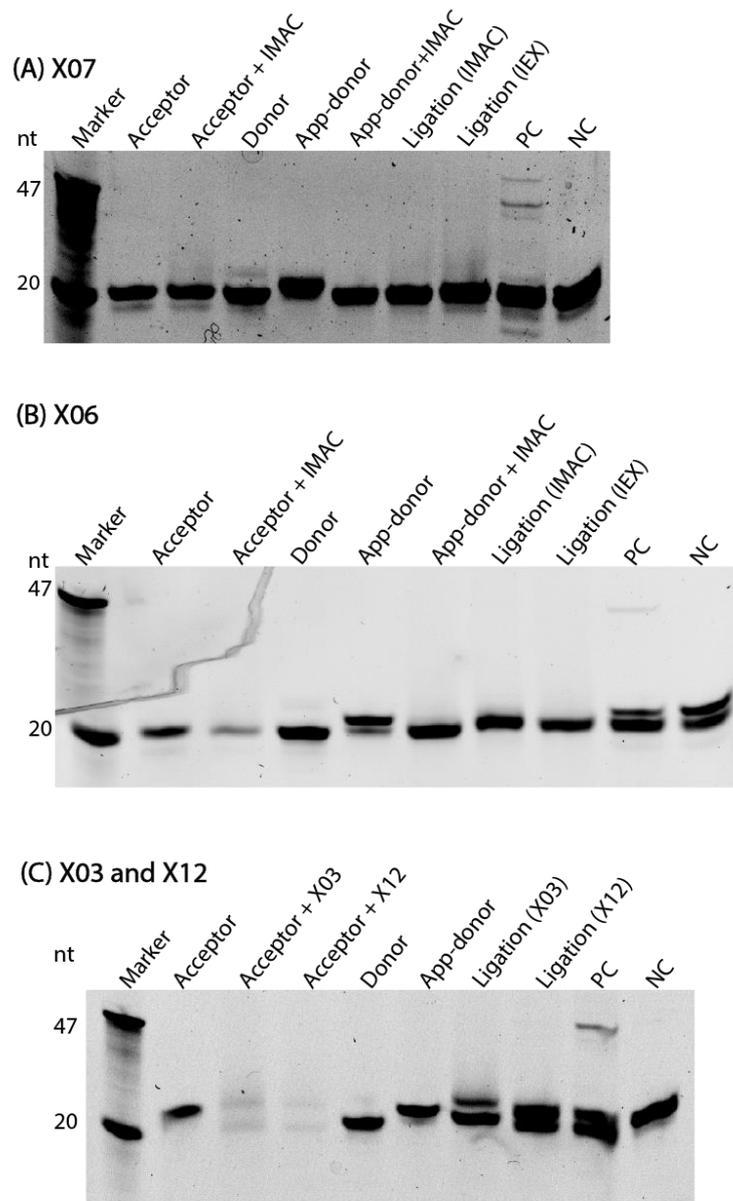


Figure 4.16: **Testing ligation activities of X03, X06, X07, and X12 in the isolated step 3 assay.** (A) Assessing ligation of and ssRNA20-OH acceptor and an App-ssDNA23-ddC donor by X07. (B) Assessing ligation of a ssRNA20-OH acceptor and an App-ssDNA23-ddC donor by X06. (C) Assessing ligation of a ssRNA23-OH acceptor and an App-ssDNA23-ddC donor with X03 and X12. Only IEX purified samples of X03 and X12 were used. Samples were resolved on a 20% Urea-PAGE gel and stained with SYBR Gold. Acceptor: acceptor only control; donor: non-adenylated donor; App-donor: donor pre-adenylated by MthRnl; PC: positive control (MthRnl); NC: negative control with no enzyme.

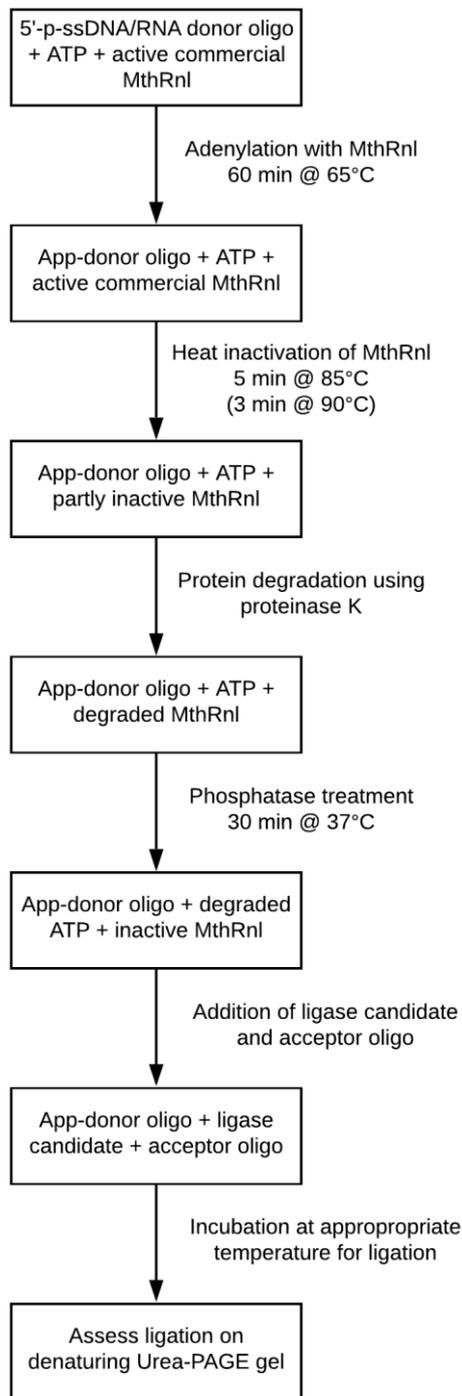


Figure 4.17: **The isolated step 3 assay.** Flowchart describing the steps during preparation of pre-adenylated substrate and its use in the isolated step 3 assay.

## 4.7 Testing the Usability of TSA with Rnl3 Family Ligases

Neither composite nor isolated step 3 assays yielded detectable ligation products for the four new RNA ligase candidates X03, X06, X07 and X12. One emerging question was whether the candidates were stable under the conditions tested for ligation. In an attempt at elucidating

stability in ligation buffer at elevated (65°C) temperatures, candidate X07 was tested using a TSA.

As before, glycerol stocks prepared from upscaled and purified candidate X07 were used. Pilot experiments with X07 in different solutions (IMAC elution buffer, standard ligation buffer, 5' DNA Adenylation Reaction Buffer (NEB), and MQ water) yielded no interpretable melting curves when using SYPRO Orange at a concentration of 6X.

In another attempt, X07 (IMAC), X06 (IMAC), MthRnl (commercial WT), and a DNA ligase (T4 DNA ligase related) were tested with both SYPRO Orange (6X) and GloMelt (1X). Neither of these dyes yielded detectable melting curves for X07 or X06. MthRnl was used as a control and did also not produce melting curves. Melting curves were obtained for the DNA ligase with both SYPRO Orange and GloMelt.

In a final attempt, X07 (IMAC) and DNA ligase (positive control) was adjusted to either high (500 mM) or low (100 mM) NaCl concentrations in MQ water. Four protein concentrations (1, 0.5, 0.2, and 0.05 mg/mL) were prepared for each NaCl concentration. No indications of melting curves were observed for either preparation of X07. The control T4 DNA ligase-like enzyme (ArcticZymes in-house) yielded curves at all SYPRO Orange concentrations in the in the 1 mg/mL protein sample (Figure E.1). A SYPRO Orange concentration of 1X yielded curves that were poorly discernable from background fluorescence. For X07, melting curve plots showed that initial fluorescence levels increased concomitantly with increased sample concentrations for every SYPRO Orange concentration tested (Figure E.2). In the DNA ligase samples that did not yield curves (protein concentrations below 1 mg/mL), very low levels of initial fluorescence were observed. In the T4 DNA samples yielding melting curves, the initial fluorescence was similar to that of X07 samples i.e. samples of X07 showed high initial fluorescence even when no curves were detectable (Figure E.3).

## **4.8 Characterizing Adenylation of ssDNA**

Despite inability to confirm ligation (step 3) activity in the four candidates and failed attempts to generate interpretable melting curves in TSA, their adenylation activities suggest that they have biotechnological potential. Adenylation without subsequent ligation is a desired feature for preparative adenylation, which is exploited in commercial MthRnl. Emerging knowledge

of adenylation preferences would also increase chances of successful ligation in the composite ligation assay.

A quantitative characterization requires that changes in product formation are measured in the linear range of the assay. Most enzymatic assays exhibit limitations regarding their linear range, that might be due to constraints in detection, substrate concentration or catalytic characteristics of the enzyme or a combination thereof. In order to evaluate effects of reaction conditions on adenylation efficiency we envisioned to use the standard adenylyltransferase assay followed by SYBR Gold-stained urea-PAGE analysis. In previous sections this assay was rather used to confirm activity than to compare catalytic activities. Certain drawbacks of this assay were obvious at first sight. To mention here are the proximity of the substrate and product bands after gel separation and staining, difficult quantitation of gel bands and the unfavorable enzyme-to-substrate ratio that is inferred by the stoichiometric rather than catalytic characteristics of this type of ligases (Lama and Ryan, 2016).

The first objective was to arrive at an assay time and enzyme concentration at which a change in adenylated product yield would reflect a semi-proportional change in enzymatic activity, and not be limited by substrate or enzyme depletion. Starting from the standard adenylyltransferase assay, product formation by X07 was monitored in a discontinuous assay. Both X07 glycerol stocks prepared from IMAC and IEX fractions (See Section 4.5) were tested, and yielded similar graphs when plotting band percentage against time (Figure 4.19).

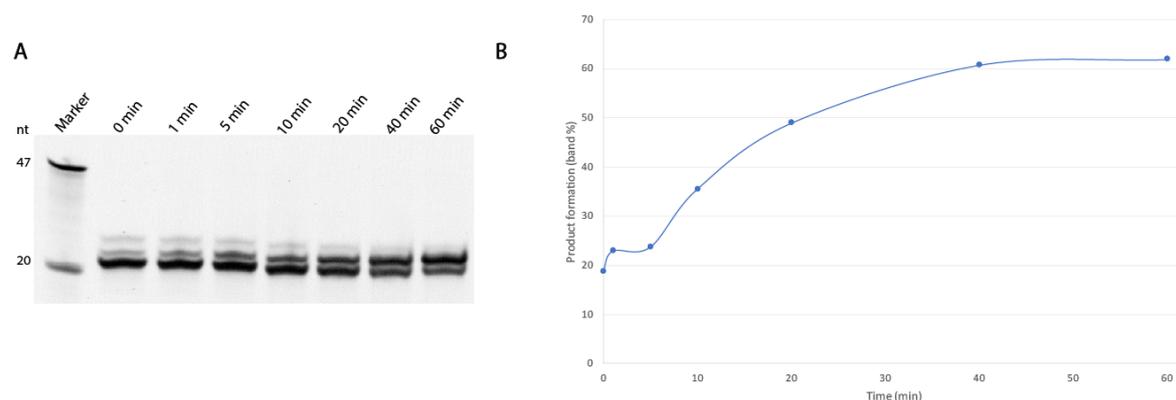


Figure 4.19: **Process curve of X07 adenylation activity.** Adenylation of p-ssDNA23-ddC by X07 (IEX-purified) was monitored in the standard adenylation assay (100 pmol ligase; S/E=1). **(A)** Samples from a discontinuous adenylation assay taken at different time points resolved on a 20% Urea-PAGE gel, stained with SYBR Gold and detected on a GelDoc XR+ Gel Documentation System. **(B)** Plotting product formation against time (min). Formation of App-donor was quantified from (A) as band % using the Image Lab lane and band analysis tool.

The progress curve showed that product formation increased with seeming linearity within the first 10 minutes (Figure 4.19 *panel B*). The rate of product formation declined after 10 minutes and reached saturation after 40 minutes. As observed in earlier activity testing, the maximal attained product formation was 50-60%.

Based on that all four candidates showed an adenylation efficiency of approximately 50% after 1 hour, this was taken as the maximal adenylation. The aim was then to develop a relatively fast assay, during which a maximal of 25% adenylation occurred (saturation is risked at adenylation >25%, and bands become difficult to differentiate from the background of the gel at adenylation <25%).

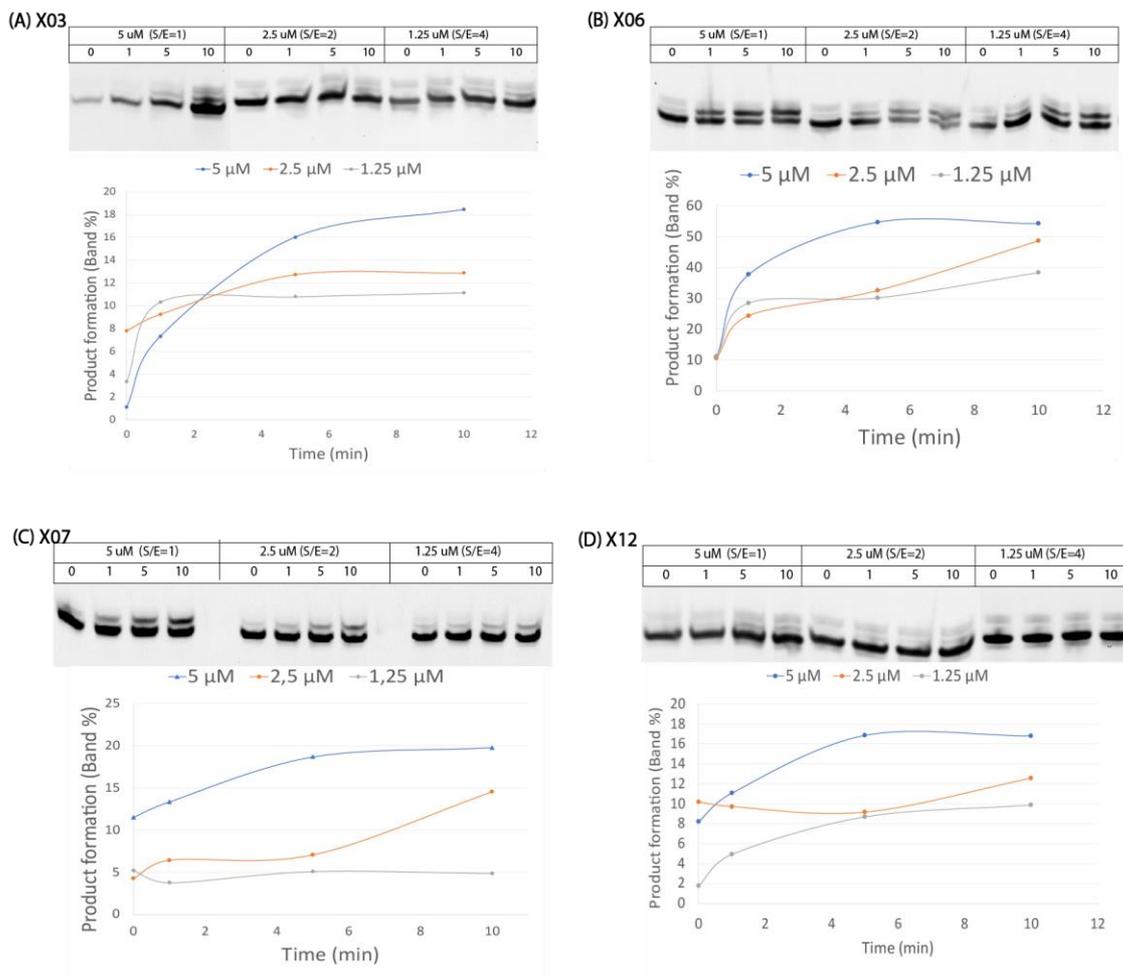


Figure 4.20: **Adenylation of p-ssDNA23-ddC in 10-minutes at different enzyme concentrations.** Candidates **(A)** X03, **(B)** X06, **(C)** X07, and **(D)** X12 were tested at 5 μM (S/E=1), 2.5 μM (S/E=2), and 1.25 μM (S/E=4) concentrations in the adenylyltransferase assay under standard conditions. Reactions were stopped after 0, 1, 5, and 10 minutes. Samples were resolved on a 20% urea-PAGE gel and stained with SYBR Gold. Using the lane and band analysis tool in Image Lab Software (Bio-Rad) the band percentage of the adenylylated donor was quantified and plotted against time (min).

Due to time required during assay preparation, it was not feasible to shorten the assay length to less than 5-10 minutes. Since the progress curve of X07 started to level out after 10-20 minutes, three different substrate-to-enzyme (S/E)-ratios were tested in a 10-minute assay for the four candidates: S/E=1, S/E=2, and S/E=4 (Figure 4.20).

X06 displayed the highest adenylation efficiency, reaching almost 40% adenylation within the first minute of the reaction using 5  $\mu\text{M}$  enzyme (S/E=1) (Figure 4.20 *panel B*). Hence, at the highest enzyme concentration measurements fell outside the linear range. At the lowest concentration tested, 1.25  $\mu\text{M}$  (S/E=4), X06 had adenylated almost 40% after 10 minutes.

X07 had the second highest adenylation efficiency. At a concentration of 5  $\mu\text{M}$  (S/E=1) 20% of the donor was adenylated after 10 minutes (Figure 4.20 *panel C*). It should be noted that the quantification of adenylation varies between gels, and that percentages between gels cannot be compared. Linear increases in adenylation were also seen at 2.5 (S/E=2) and 1.25 (S/E=4)  $\mu\text{M}$  enzyme.

Both X03 and X12 only adenylated less than 20% of donors within 10 minutes when using the highest enzyme concentration (S/E=1)(Figure 4.20 *panel A and D*). At lower concentrations of X12, adenylated donor bands fell below the detection limit of the SYBR Gold stain. For X03, adenylation products were only detectable after 5 minutes.

### **Temperature Effect on Adenylation**

X06 and X07 were deemed to be the most promising candidates based on the above results and feasibility of expression in *E. coli*. Hence, they were selected for further investigations into the effect of temperature on substrate adenylation. X07 was used at 2.5  $\mu\text{M}$  (S/E=2) concentration in the 10-minute assay. Due to its high activity, X06 was tested at a concentration of 2.5  $\mu\text{M}$  (S/E=2) and 1.25  $\mu\text{M}$  (S/E=4) and the assay time was decreased to 5 minutes in order to avoid measurements outside the linear range.

Adenylation by X07 was assessed at temperatures spanning 30 to 85°C. A clear trend of increase in adenylated product was observed with increase in temperature (Figure 4.21). Amount adenylated product reached 20% already at 30°C. This efficiency was constant in the range 30 to 60°C. From 65 to 80°C adenylation efficiency increased strongly (Figure 4.21

panel A and B). At these elevated temperatures, X07 adenylated at most (80°C) 50% of the ssDNA 23-mer donor.

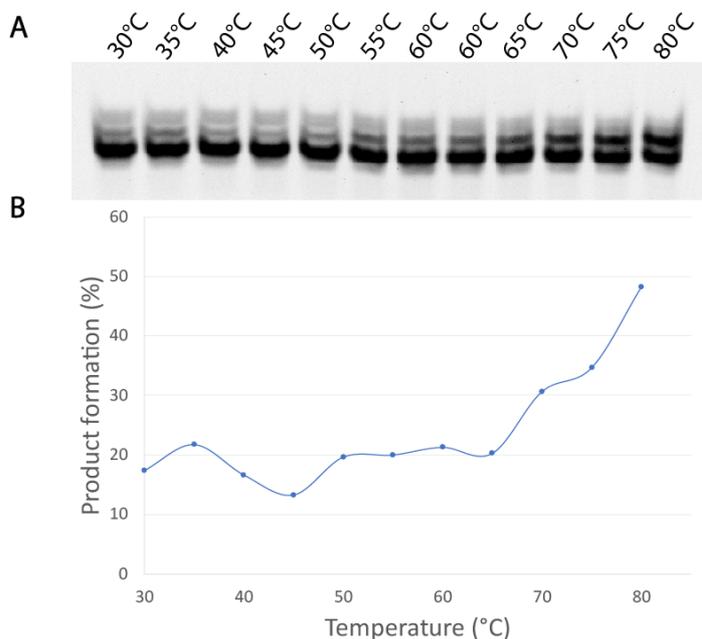


Figure 4.21: **Effect of temperature on adenylation efficiency of X07 tested on a p-ssDNA23-OH donor.** Candidate X07 was tested in the standard adenyllyltransferase assay with S/E=2 at 30 to 80°C in increments of 5°C on a PCR thermal block. **(A)** Samples were resolved on a 20% urea-PAGE gel, stained with SYBR Gold and detected in the SYBR Gold channel on a Gel Doc XR+ System. Double samples for 60°C serve as controls between separate incubations. **(B)** Product formation was quantified using the lane and band analysis tool in Image Lab Software (Bio-Rad). Band percentages were plotted against temperatures (°C).

Based on that candidate X06, just as X07, originates from a hyperthermophile organism and observations made in X07, temperature-dependence of X06 was tested in the range 60-85°C. The assay was performed with two different protein concentrations (1.25 and 2.5  $\mu$ M, S/E=4 and S/E=2, respectively), which both resulted in  $\approx$ 50% adenylation within 5 minutes at temperatures between 60-70°C (Figure 4.22). In both experiments, adenylation appeared to decrease from 75°C on and upwards. An interesting observation was made in the wells, where DNA accumulated with an increasing gradient towards the lower temperatures tested. Concomitant with the decrease in adenylation at 75°C and on, this accumulation also decreased. When running the sample incubated at 0°C, neither product band nor accumulation of DNA in wells was noted.

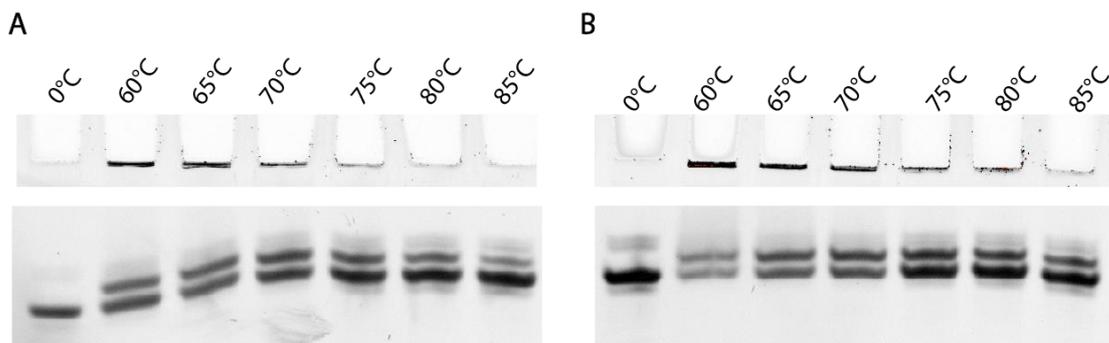


Figure 4.22: **Effect of temperature on adenylation efficiency of X06 tested on a p-ssDNA23-OH donor.** Samples were run in a 5-minute version of the standard adenylyltransferase assay. Two enzyme concentrations were tested: **(A)** 1.25  $\mu\text{M}$  (S/E=4) and **(B)** 2.5  $\mu\text{M}$  (S/E=2). Samples were resolved in a 20% urea-PAGE gel, stained with SYBR Gold, and detected in the SYBR Gold channel of a Gel Doc XR+ System.

By running this experiment again and incubating samples with Proteinase K (0.4 U per  $\mu\text{L}$  sample), it was tested whether digestion of protein would reduce accumulation of DNA in wells. Comparing sample aliquots treated with Proteinase K to those without treatment showed that protein degradation appeared to have little effect on well-deposition of DNA (not shown).

In a final experiment conducted on X06, ATP-dependence of adenylation was investigated. Adenylyltransferase assays with 2.5  $\mu\text{M}$  enzyme ( $S/E=2$ ) were incubated with 0.01, 0.1, 1, or 10 mM ATP or without ATP at 65°C.

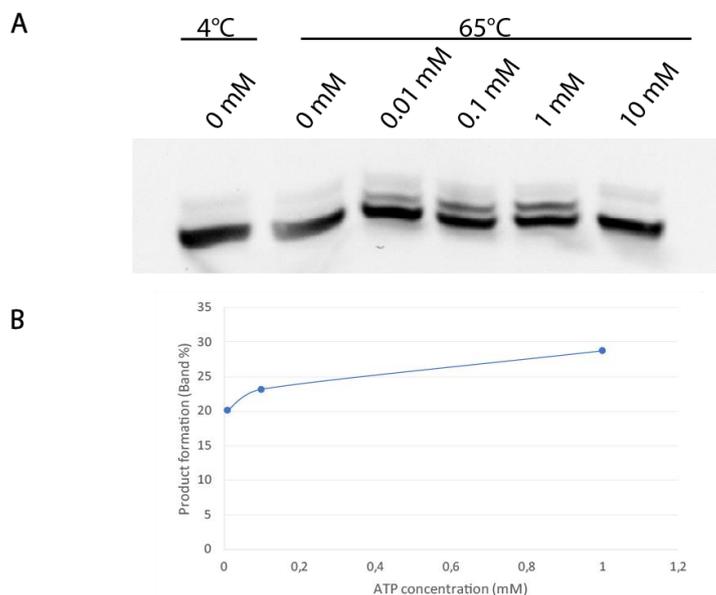


Figure 4.23: **ATP-dependence of adenylation by X06.** Ability of X06 to adenylate a p-ssDNA23-OH donor was assessed at ATP concentrations of 0 (negative controls), 0.01, 0.1, 1, and 10 mM ATP. Reactions were run for 5 min at 65°C with 2.5  $\mu\text{M}$  enzyme ( $S/E=2$ ). Samples had been treated with EDTA (20 mM) and Proteinase K (0.4 U/ $\mu\text{L}$  sample, 45°C, 10 min). **(A)** Samples were resolved on a 20% urea-PAGE gel, stained with SYBR Gold, and detected in the SYBR Gold channel of a Gel Doc XR+ System. No DNA was stained in the wells. **(B)** Product formation was quantified (when possible) using the lane and band analysis tool in Image Lab Software (Bio-Rad). Band percentage were plotted against ATP concentrations (mM).

The same level of adenylation ( $\approx 50\%$ ) as observed in Figure 4.22 was not obtained at the same ATP concentration (0.1 mM). Very little change in product formation ( $\approx 20\%$ ) was noted over the gradient 0.01 to 1 mM ATP. Product formation was absent when incubating with 10 mM ATP (Figure 4.23 *panel A*).

## 4.9 Testing Standard Assays for Phosphatase and Nuclease Analysis

Part of protein purification method development is to establish analysis protocols for problematic contaminants, which for polynucleotide ligases are oligonucleotide-modifying enzymes such as phosphatases and nucleases. Available samples (glycerol stocks) after each of the two purification steps (IMAC and IEX) and crude samples from purification experiments (Section 4.4) of the four active candidates and X19 were analyzed for phosphatase and single strand endonuclease activities using standard assays.

## Testing a Standard Phosphatase Assay

A dilution series of rSAP was used to generate a standard curve from  $2.5 * 10^{-2}$  to  $2.5 * 10^{-9}$  U/ $\mu$ L in steps of  $10^{-1}$ . Here, slopes of the colorigenic reaction catalyzed by phosphatases were calculated by Softmax Pro. Candidates X03 and X12 were measured on a separate 96-well plate from X06, X07, and X19. Standard curves are shown in Appendix D. Slopes were plotted against assay enzyme concentration, and a linear regression was fitted. Standard curves yielded non-zero intercepts. We defined the dynamic range of the assay to be located between  $2.5 * 10^{-7}$  U/ $\mu$ L and  $2.5 * 10^{-5}$  U/ $\mu$ L.

Using the standard curves shown in Appendix D, phosphatase activity in samples were calculated. Absorbance measurements were obtained from continuous measurements with both high and low enzyme concentration. Based on the standard curves described above, phosphatase activity of close to 50  $\mu$ U/ $\mu$ L was detected in unpurified (lysate) samples of X03 and X07, and 134  $\mu$ U/ $\mu$ L in X12 lysate (Table D.2). For all other purified samples, continuous absorbance measurements yielded in slopes outside the lower limit of the dynamic range of the assay. These results indicate very low phosphate activities implying high purity with regard to this contaminant.

Absorbances of two technical replicates were measured for each sample. Between these two replicates, coefficients of variation varied from 0.8 to 25.0%, but were on average 10% in the first (X03, X12) and 3.7% in the second (X06, X07, X19) experiment (data not shown). Using assay dilution factors, phosphatase activity in the original samples was back-calculated. This was done for both high and low enzyme concentration assays (two technical replicates each). The resulting four sample concentrations showed little agreement in case of measurements located outside the dynamic range of the assay (in many cases the coefficient of variation exceeded 100%) (Table D.2).

## Testing a Standard Endonuclease Assay

Single strand endonuclease activity was tested in dilution series of samples by supplying circular phage-vector ssDNA, similar to analytical assays used by RNA ligase manufacturers. Nuclease activity was detected by that degraded DNA migrated faster through the gel and created smears.

Where samples of lysate were included, DNA was degraded either completely (e.g. Figure 4.24 *panel E*) or to a strong degree (Figure 4.24 *panel A and C*). In IMAC-purified samples, plasmid DNA was mostly intact i.e. IMAC-purified samples were judged to be nuclease-free by the standards of this assay. However, after IEX all candidates (except X19) yielded smears and slower migrating bands. Smear patterns were different from those observed in lysate in that DNA products migrated slower than the intact substrate. Samples containing high levels of fragmented DNA were expected to show an accumulation of DNA below the band containing plasmid DNA (e.g. Figure 4.24 see X03-lysate and PC in *panel A* and X19 IEX in *panel F*).

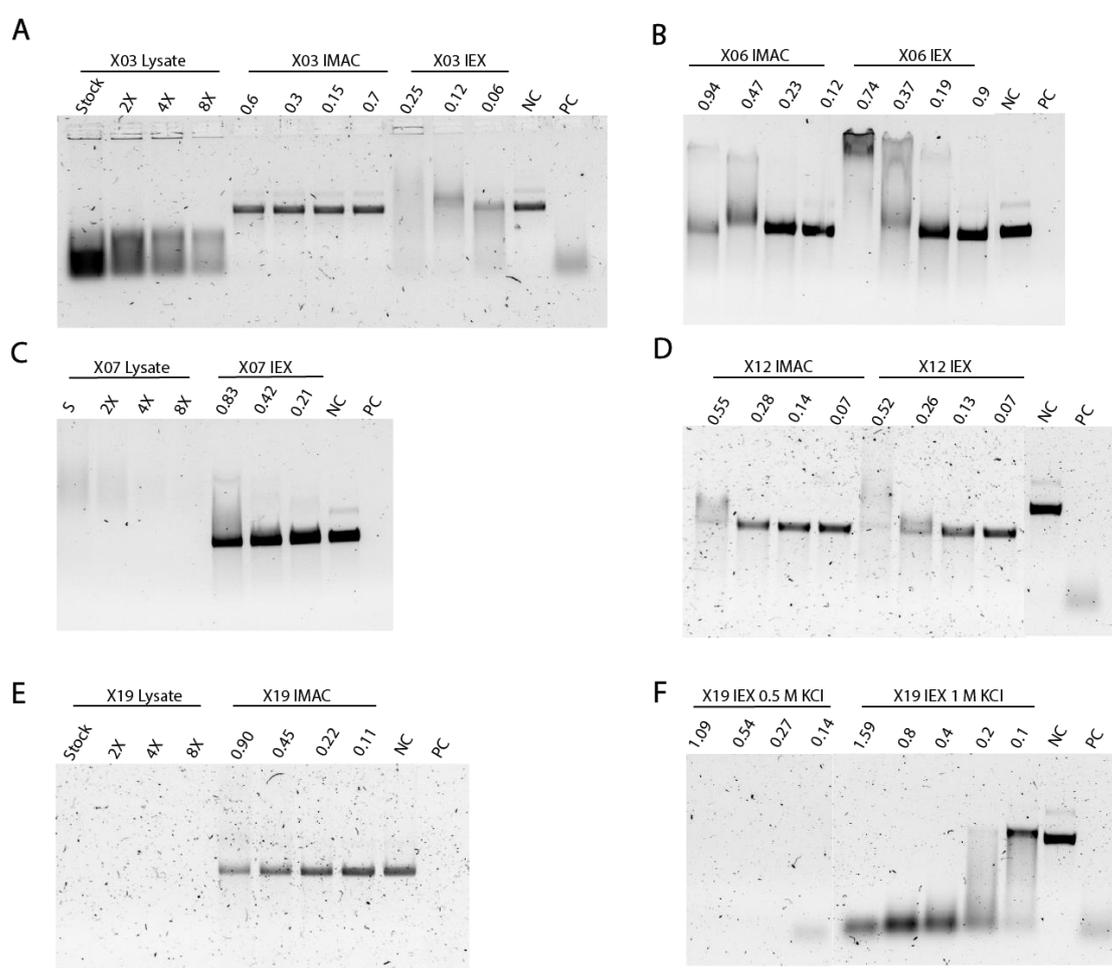


Figure 4.24: **Assessing presence of endonuclease activity in samples of (A) X03, (B) X06, (C) X07, (D) X12, and (E and F) X19.** Available samples from throughout the 2-step purification process were supplied with ssDNA plasmid substrate and incubated for 4 hours at 37°C. Each sample was prepared as a dilution series. Protein concentrations (mg/mL) are noted above each lane. Positive controls (PC) contain a reference sample known to be endonuclease-contaminated. Negative controls (NC) contain only dilution buffer (25 mM Tris) and substrate. Samples were resolved in a 0.7% agarose gel with GelRed and detected in the GelRed-channel on the GelDoc XR+ Gel Documentation System.

It should also be noted that slower migrating DNA bands were visible in IMAC samples of X06 and X12 at the highest sample concentration (Figure 4.24 *panel B and D*).

Unexpected fragmentation and band patterns were most pronounced for X06. It was hypothesized that these bands could be a gel shift resulting from a DNA-binding protein. Hence, these samples were incubated with Proteinase K (0.16 U/ $\mu$ L sample) and resolved on a 0.7% agarose gel anew. Smears were reduced after Proteinase K treatment and the bands observed before appeared more intense. Band migration was reduced concomitant with increasing sample concentration. In general, migration was slower in samples containing the IEX glycerol stock (Figure 4.25).

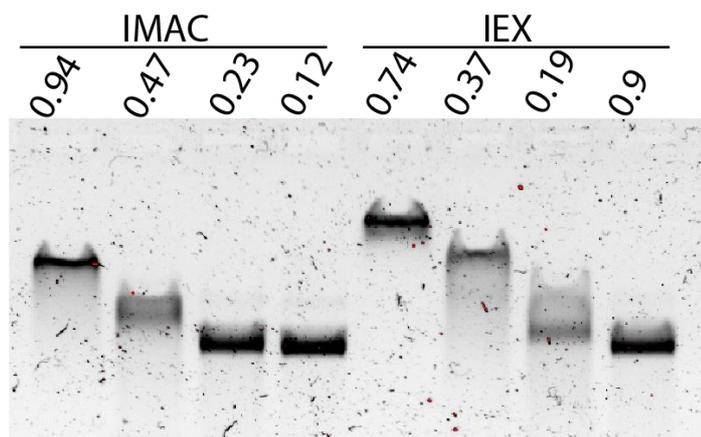


Figure 4.25: **Effects of treating X06 IMAC and IEX endonuclease tests with Proteinase K.** The samples analysed in Figure 4.23 were treated with Proteinase K (0.16U/ $\mu$ L sample) at 45°C for 10 min, resolved in a 0.7% agarose gel with GelRed and detected in the GelRed-channel on the GelDoc XR+ Gel Documentation System.

Detrimental effects of *E. coli*-derived nucleases were highly problematic when testing mesophile-derived candidates such as X19. After IMAC purification X19 was nearly nuclease-free, showing only a faint smear in addition to the intact plasmid DNA band at the highest concentration tested (Figure 4.24 *panel E*). However, the AIEX-purified X19 sample generated fragments that migrated faster than the intact plasmid DNA (Figure 4.24 *panel F*). The band containing intact plasmid DNA was only detected when diluting the X19 IEX sample to a protein concentration of 0.1 mg/mL.

## 4.10 Bioinformatics

DNA sequences were submitted to the Geneious Software and aligned using Clustal Omega algorithm. Similarity and identity matrices were created from this alignment using the BLOSUM algorithm. In Figure 4.27 the similarity matrix for all 22 candidates is presented. The identity matrix can be found in Appendix F.

Using this alignment in the Geneious tree builder the phylogenetic tree for all 22 candidates, rooted in T4Rnl2, was built (Figure 4.26).

Similarities among the 22 candidates ranged from 73.8 to 34.6%. The pair scoring the highest identity and similarity were X22 and X02, followed by X21 and X19. Several subclusters of candidates with higher similarities were present within the set (Figure 4.26 and 4.27).

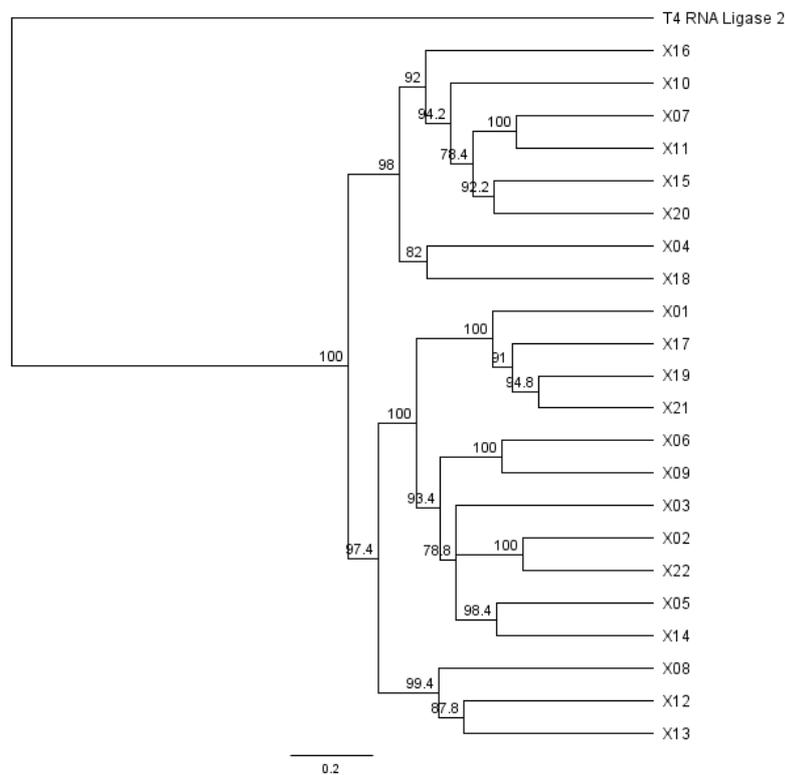


Figure 4.26: Rooted phylogenetic tree built from the 22 putative RNA ligases investigated in this thesis.

	X05	X14	X03	X02	X22	X06	X09	X19	X21	X17	X01	X12	X13	X08	X07	X11	X15	X20	X10	X16	X04	X18	T4 RNA Ligase 1	T4 RNA Ligase 2	
X05	66.9	66.9	58.3	57.9	59.5	54.7	57.2	51.8	54.7	50	52	42.6	45.4	48.1	41	41.8	36.4	39	42.3	40.1	43.4	48.7	25.6	24.7	
X14	58.3	60.1	60.1	56.7	55.8	55.5	52.6	49.2	50.7	47	48.3	45	47.2	47.2	41	40.4	39.8	39.5	39.1	38.4	42	46.8	23.3	20.6	
X03	57.9	60.1	55.9	55.9	59.8	58.7	54.4	45.7	49.5	49.9	47.8	46.4	46.5	48.2	41.8	39.9	38.6	39.7	38.7	36.7	42.1	47	26.7	20.1	
X02	59.5	56.7	55.9	73.8	73.8	59.6	57.3	50.7	54.3	51.2	52.7	44.5	45.2	45.9	40.9	39.5	39.2	38.7	40.6	42.2	43.2	46.4	26.2	21.9	
X22	54.7	55.5	58.7	59.8	62.2	62.2	61.3	53.6	53.3	53	51.1	46.3	47.3	47.7	43	42.4	40.7	38.1	43.2	40.9	45.6	47.4	25.9	20.8	
X06	54.7	55.5	58.7	59.8	62.2	62.2	61.3	53.6	53.3	53	51.1	46.3	47.3	47.7	43	42.4	40.7	38.1	43.2	40.9	45.6	47.4	25.4	25.4	
X09	57.2	52.6	54.4	57.3	61.3	69.3	49.5	50.8	50.8	46.5	47.5	46.3	47.6	50.3	42.4	40.1	39.5	37.2	43.8	40.8	45.5	50.9	27.7	26.4	
X19	51.8	49.2	45.7	50.7	53.6	48.6	49.5	50.8	71.8	66	60.5	42.9	40.6	44	40.9	39.8	37.9	38.2	38.9	41.6	44.4	45.9	24.0	21.6	
X21	54.7	50.7	49.5	54.3	53.3	51	50.8	71.8	66	66.6	65.1	44.5	43.5	47.4	43.2	42.2	41.9	41	42.4	43.2	45.5	47.7	24.3	23.9	
X19	54.7	50.7	49.5	54.3	53.3	51	50.8	71.8	66	66.6	65.1	44.5	43.5	47.4	43.2	42.2	41.9	41	42.4	43.2	45.5	47.7	24.3	23.9	
X17	50	47	49.9	51.2	53	47.6	46.5	66	66.6	66.6	61.6	41	41.6	40.9	37.9	39.5	39.2	40.2	39.9	39.9	42.2	43.7	25.5	22.6	
X01	52	48.3	47.8	52.7	51.1	50	47.5	60.5	65.1	61.6	61.6	42.4	39.3	44.5	37.9	39.5	39.7	40.6	41.5	40.8	45.9	42.4	24.6	21.1	
X12	42.6	45	46.4	44.5	46.3	47.9	46.3	42.9	44.5	41	42.4	56.3	56.3	55.7	38.6	36.5	34.6	36.2	36.5	35.4	39.7	38	23.0	20.6	
X13	45.4	47.2	46.5	45.2	47.3	49.2	47.6	40.6	43.5	41.6	39.3	56.3	56.3	56.6	41.2	37.8	37.4	37.5	36.6	35.6	39.7	45	26.0	19.5	
X08	48.1	47.2	48.2	45.9	47.7	48.6	50.3	44	47.4	40.9	44.5	55.7	55.7	56.6	41.9	41.6	40.2	40.3	39.3	38.9	49.1	28.3	20.6	20.6	
X07	41	41	41.8	40.9	43	43.7	42.4	40.9	43.2	37.9	37.9	38.6	41.2	41.9	68	68	58.8	60.6	60.2	57	53.2	58.2	27.5	23.5	
X11	41.8	40.4	39.9	39.5	42.4	38.3	40.1	39.8	42.2	39.5	39.5	36.5	37.8	41.6	68	68	57.5	60.6	61.2	53.8	53.2	55.5	26.2	23.2	
X15	38.4	39.8	38.6	39.2	40.7	38.4	39.5	37.9	41.9	39.2	39.7	34.6	37.4	40.2	58.8	57.5	61.2	60.3	55.1	52	46.3	49.1	23.1	21.6	
X20	39	39.5	39.7	38.7	38.1	37.4	37.2	38.2	41	40.2	40.6	36.2	37.5	40.3	60.6	61.2	60.3	39.3	59.3	50	48.6	52.5	22.0	22.4	
X10	42.3	39.1	41.2	40.6	43.2	41.6	43.8	38.9	42.4	39.9	41.5	36.5	36.6	39.3	60.2	58.9	55.1	52	49.6	49.6	48.3	52.3	23.7	24.2	
X16	40.1	38.4	38.7	42.2	40.9	39.6	40.8	41.6	43.2	39.9	40.8	35.4	35.6	38.9	57	53.8	52	50	49.6	49.6	51.6	52.1	24.7	21.7	
X04	43.4	42	42.1	43.2	45.6	43.7	45.5	44.4	45.5	42.2	45.9	38	39.7	38.9	53.2	53.2	46.3	48.6	48.3	51.6	57	57	27.5	24.4	
X18	48.7	46.8	47	46.4	47.4	48.9	50.9	45.9	47.7	43.7	42.4	42.4	45	49.1	58.2	55.5	49.1	52.5	52.3	52.1	57	57	26.4	24.7	
T4RnL2	25.6	23.3	26.7	26.2	25.9	26.4	27.7	24.0	24.3	25.5	24.6	23.0	26.0	28.3	27.5	26.2	23.1	22.0	23.7	24.7	27.5	26.4	20.4	20.4	20.4
T4RnL1	24.7	20.6	20.1	21.9	20.8	25.4	26.4	21.6	23.9	22.6	21.1	20.6	19.5	20.6	23.5	23.2	21.6	22.4	24.2	21.7	24.4	24.7	20.4	20.4	20.4

Figure 4.27 : Similarity matrix created from the 22 candidates of this thesis and the two bacteriophage T4 RNA ligases using the BLOSUM algorithm. Similarity is in percentage.

## 5 Discussion

In this project in collaboration with ArcticZymes, 22 RNA ligase candidates were analyzed for their ligase activity and expression levels with the aim to define lead candidates that might become future alternatives to existing commercial RNA ligases. Four of the most promising candidates, X03, X06, X07, and X12, were further analysed in detail for their activity, requirements and characteristics. In addition, procedures for optimized selections are suggested.

### 5.1 Screening Expression of Archaeal RNA Ligases in *E. coli*

Twelve out of the 22 RNA ligase candidates were highly expressed, and five of these showed good solubility. The attrition rate due to poor solubility ( $\approx 50\%$ ) is within what is expected from screening processes (Christendat *et al.*, 2000). Solubility is important since it indicates proper folding and hence biological activity of an enzyme. Eventually, solubility was chosen as a requirement for further investigation of candidates.

Reasons for failure to obtain these proteins in soluble form are manifold and could originate during expression or downstream sample treatment e.g. during lysis. Expression of recombinant proteins in *E. coli* is impacted at the gene, transcript, translation, and protein level (Singha *et al.*, 2017). As obtained plasmids were not sequence verified, lack of expression observed for X11 could be due to unsuccessful cloning (e.g. incorrect insert orientation) or deleterious mutations in the gene of interest or the plasmid backbone. For X11 the absence of expression was unexpected, based on its high similarity (68%) with X07 (50.9% identical), which was one of the best-expressed candidates. Other probable reasons for attrition are protein toxicity and folding issues. It is conceivable that RNA ligases could exert harmful effects on *E. coli* through their biological activity. The observation that growth ceases immediately after induction for X02 (MthRnl), which is highly expressed in soluble form supports this (data not shown).

Folding problems could occur when trying to express extreme halophile-derived candidates such as X19 and X21 (derived from members of the *Halobacteriaceae* family, 57.9% identical and 71.8% similar). While X19 showed high, and X21 showed low expression in initial studies, both were poorly soluble. Extreme halophilic archaea such as the *Halobacteriaceae* accumulate high intracellular KCl levels and may express haloadapted

enzymes (Oren, 2013). Such enzymes are known to be unstable at lower salt concentrations, and consequently difficult to express in soluble form in mesophilic hosts such as *E. coli* (Haque, Paradisi and Allers, 2020). Expression systems suitable for halophilic proteins exist e.g. *Haloferax volcanii* and have previously been used to express halophile derived enzymes (Haque, Paradisi and Allers, 2020). If desired, accessing polynucleotide ligases from these extreme halophiles should be tried by expression in an appropriate host.

It is also possible that solubility is decreased because the target genes have been codon optimized due to their archaeal origin (Kim *et al.*, 1998). By increasing the translational speed, codon optimization can in some cases increase aggregation, which is not necessarily offset by expression at lower temperatures such as in this thesis (Rosano and Ceccarelli, 2009).

Optimizing soluble expression with alternative expression vectors and host organisms was not attempted due to the throughput-based design of the approach described in this work. Instead, simple adjustments of induction conditions were tested, with the aim of improving soluble expression by regulating expression levels (Singha *et al.*, 2017). For three thermophile-derived candidates (X08, X13, X14) which initially were moderately/highly expressed in insoluble form, induction conditions were varied in hope of improving the soluble-to-insoluble ratio. Tunability of *XylS/Pm* allowed this to be tested here, however solubility was not improved by using a lower inducer concentration. The growth of all three clones in every condition was negatively affected by high inducer concentrations, which also had been observed in a previous study using the *XylS/Pm* system (Winther-Larsen *et al.*, 2000). Effects of growth phase were investigated by inducing at high and low OD, which also did not affect solubility. In log phase *E. coli* are actively growing and geared towards protein expression, while towards the end of log phase metabolism shifts in order to cope with an increasingly stressful extracellular and intracellular environment (Chou, 2007). Therefore, it is often advised against inducing expression at higher ODs. On the other side, by increasing biomass induction at high densities can improve low yields when the POI is toxic. This could be tested for promising, active candidates with low expression: X12 and X03. The experiment showed that the conditions used during screening gave the best soluble-to-insoluble ratio for these candidates.

## 5.2 Evaluation of Functional Screening of Ligase Candidates

Initial activity tests showed that six candidates could catalyse formation of 23-mer donor ssDNA products of slightly larger size than the substrate. Of these six, two (X02: MthRnl and X18: PAB1020) had already been described in the literature (Brooks *et al.*, 2008; Torchia, Takagi and Ho, 2008). The increase in size equivalent of approximately one nucleotide corresponded to the covalent attachment of an AMP moiety.

An important question to be discussed for the initial activity screening is how sensitive and specific the screening was i.e. how likely is the generation of false negative and false positive results, and what could cause such results?

False positives were minimised by also running the reaction without ATP, which is essential for adenylation by this class of RNA ligases (Brooks *et al.*, 2008; Torchia, Takagi and Ho, 2008). As expected, slower migrating products were not generated in absence of ATP. This indicated that X03, X06, X07, and X12 were new active thermophilic members of the Rnl3 family. No ligation-based side-products were observed, which is expected due to the blocking 3'-6FAM-label and presence of ATP, which inhibits ligation (Zhelkovsky and McReynolds, 2011). The finding that X02, which corresponds to the commercialized MthRnl, could adenylate ssDNA agrees with previous studies and hence serves as a control for the screening process (Torchia, Takagi and Ho, 2008). Ability of X18 (PAB1020) to adenylate ssDNA was a finding not previously reported but supports that RNA ligases belonging to the Rnl3 family do not discriminate against DNA on the donor side of the nick (Torchia, Takagi and Ho, 2008).

False negatives were difficult to control for, and it is conceivable that many true positives were overlooked for the following reasons connected to sample quality: unfavourably high NaCl (and KCl) concentrations, low concentration of active RNA ligase candidate, and interfering contaminants. All three reasons originate from the optimistic assumption that sample quality is high, and that protein concentration determined by absorbance at 280 nm was used directly to calculate the sample volume in assays. Therefore, when protein concentrations in sample stocks were low, components from the storage buffer made up a greater part of the final assay composition. E.g. adenylation assays of X10 and X13 (not originating from extreme halophiles) contained 170 mM and 140 mM NaCl, respectively.

Polynucleotide ligases are known to be inhibited at NaCl concentrations exceeding 100 mM NaCl, hence activity under these assay conditions is less likely (Blondal *et al.*, 2003; Torchia, Takagi and Ho, 2008). Contributing to this is also that solubility of DNA decreases at increasing salt concentrations (Creighton, 2010). Meanwhile, among the positive candidates, X02 had similar adenylation activity at both NaCl concentrations tested (11 and 53 mM NaCl), which is expected based on previous findings (Torchia, Takagi and Ho, 2008).

The true candidate concentration in assays is likely much lower than assumed, because purity after SpinTrap IMAC was deemed to be moderate-low by Coomassie-stained SDS-PAGE analysis. Hence, unrelated proteins make up a large part of the measured protein concentration. Overestimation of functional enzymes also results from the assumption that the candidate is completely present in active form, which is unlikely because no storage buffer optimization or removal of aggregated protein was performed. Basing protein concentration measurements purely on absorbance at 280 nm when the sample contains imidazole (after IMAC) also leads to inaccuracies, because imidazole absorbs UV radiation at this wavelength.

In addition to unrelated proteins and inactive forms of the RNA ligase candidate, other specific contaminants such as active enzymes (nucleases and phosphatases), and host oligonucleotides could interfere with the assays. This is especially likely in the initial activity screens because sample purity was low. Performing the assay at 37°C strongly indicated the presence of nucleases in the samples. As nuclease products only were observed in the SYBR Gold stained gels and not by fluorescence detection it was likely that ssDNA degradation was caused by 3'→5' exonucleases. 6FAM labels are known not to block exonucleases (Lohman, 2019). Exonucleases were however only active in samples incubated at 37°C, and heat-inactivated at 65°C. Nucleases are likely co-purified from *E. coli* because of metal-binding sites consisting of histidines (Bolanos-Garcia and Davies, 2006). Their degradation of substrates made it impossible to draw conclusions about ligase activity from the initial assays of mesophile-derived candidates.

Initial activity screening could have been improved by dialyzing all samples against a high and a low salt concentration, as was done for only part of the samples (the most promising in terms of expression). Nevertheless, through dialysis the protein concentration may be decreased and sample may be lost during handling. Additional steps of concentrating the

samples in spin concentrators adds laborious steps which reduce yield. Therefore, it will always be difficult to control the final reaction composition when protein concentration is held constant.

The impact of sample quality on functional screening outcomes is illustrated by that candidates showing activity coincide with candidates being highly expressed in soluble form. Based on the bias originating from soluble expression levels, one should consider limiting functional screening to candidates that score well in initial expression screening. In the first steps in an enzyme lead screening protocol one must find a balance between dismissing potentially interesting candidates, and spending time on candidates that are unlikely to show activity in screening. Furthermore, a candidate found to be active could later be discarded due to problems in its recombinant expression e.g. inefficient production for characterization and industrial-scale production. Time saved on initial activity screens could be used to optimize purification of well-expressed, active candidates (more in Section 5.4). This would improve the quality of further functional characterization, e.g. the search for step 3 activity.

### **5.3 Activities of the Four Most Promising Candidates**

Adenylation activity under standard conditions was confirmed in upscaled and purified batches of X03, X06, X07, and X12. These candidates were chosen based on their activity and novelty, and were all expressed in soluble form with X06 and X07 having the highest expression level. Compared to initial activity screening, adenylation efficiency was either equal or improved for all candidates in this batch. Yet only an average of 50% adenylation was attained at  $S/E=1$ , while commercial MthRnl (WT) almost completely adenylated donors. Ligation activity was tested using two different approaches: the isolated step 3 assay, and the composite ligation assay. Both ssRNA and ssDNA was tested as acceptors, but neither yielded detectable ligation products with the new candidates.

Low adenylation activity is likely the additive result of suboptimal reaction conditions. Complete lack of activity as in the composite assay could be explained the same way, because it relies on the efficiency in step 1 and 2. However, absence of ligation product in isolated step 3 assays, where sufficient step 3 substrate was ensured, implies a more fundamental issue with step 3.

It is expected that suboptimal performance arises from insufficient characterization of preferred substrate and enzyme concentrations, and their ratio (S/E-ratio), salt, pH, temperature, cofactors, and substrate characteristics. The factors related to sample quality, discussed for false negatives in initial screening (Section 5.2) are also expected to contribute depending on the success of purification.

Adenylation capabilities of these four candidates are similar to those of the commercial RNA ligases CircLigase, T4Rnl1, and T4Rnl2 in conditions not optimized for adenylation (Zhelkovsky and McReynolds, 2011). More thorough investigations into TS2126 Rnl1 (commercialized as CircLigase) show the same ATP-independence as X06 when the enzyme is used at a lower concentration (1  $\mu\text{M}$ ) (Lama and Ryan, 2016). Just as X06, TS2126 Rnl1 adenylated 50% of p-ssDNA donors at 1  $\mu\text{M}$  at both 50 and 500  $\mu\text{M}$  ATP, however when increasing the enzyme concentration to 5  $\mu\text{M}$  (and 10  $\mu\text{M}$ ), adenylation efficiency became ATP-dependent and attained nearly 100% at 500  $\mu\text{M}$  ATP (Lama and Ryan, 2016). This shows that complete adenylation may be achievable by optimizing the enzyme concentration used in adenylation assays. Supporting the notion that adenylation may attain 100% by assay optimization is that X02 (in-house produced wild type MthRnl) only adenylated 50% of the donor in initial activity assays. X02 was not selected for further studies due to its lack of novelty.

The adenylation process curve shows that the maximal adenylation is reached after 40 minutes (saturation) (S/E=1). Unfortunately, to our knowledge no such process curves have been published for other ligases (MthRnl and TS2126 Rnl1). No increases in adenylation efficiency were noted when increasing the ATP concentration beyond 10  $\mu\text{M}$  (100 pmol substrate; S/E=2). This implies that ATP is not the limiting factor in adenylation. When interpreting results from adenylyltransferase assays it is important to keep in mind that this assay only investigates formation of step 2 product. The assay cannot be used to make conclusions about performance level in isolated step 1 or 2. Under the assumption that ligation is stoichiometric as opposed to catalytic, i.e. RNA ligases are “used up” after catalyzing a single reaction, at least 50% of the ligases were adenylated at saturating ATP (no preadenylated enzymes existed after purification) (Lama and Ryan, 2016). This is surprising, because only 2% of MthRnl was adenylated at saturating ATP as reported by Torchia and co-workers (Torchia, Takagi and Ho, 2008). This could indicate that the candidates are

efficiently self-adenylated, and that step 2 limits adenylation. If these ligases function in a stoichiometric way, 100% substrate adenylation would require an equimolar amount of functional enzyme (actual  $S/E \leq 1$ ), which was likely not achieved during activity tests, due to sample quality. Yet, no difference in adenylation was observed when X06 was tested at both (apparent)  $S/E=0.5$  and 1. This indicates that adenylation is not limited due to lack of enzyme. The catalytic-vs-stoichiometric behavior of related MthRnl is poorly understood, and it could be that only a small number of active ligases were adenylating themselves and the substrate in a catalytical manner. Investigating step 1 and 2 in isolation would be useful regardless of this, and in the future, radioactive incorporation studies should be considered to characterize ligase-AMP formation (step 1).

No direct studies of salt effects on adenylation activities were conducted, but by comparing conditions under which candidates showed activity, indications of salt tolerance can be derived. X07 and X03 appeared to be sensitive to NaCl, as activity only was detected at lower (10 mM) and not at higher (50+ mM) NaCl concentrations. High KCl levels, on the other side, appear not to be detrimental to adenylation activity in any of the candidates (e.g. X06 with high adenylation activity at both 80 and 160 mM KCl). Nevertheless, these salt concentrations could reduce adenylation efficiency through direct effects on enzyme function and stability or indirectly by effects on the oligonucleotide substrate. It has been reported that RNA ligases can prefer certain monovalent salt over others (KCl over NaCl) in certain concentration ranges (Blondal *et al.*, 2003). The related RNA ligase MthRnl is supplied in a storage buffer containing KCl, resulting in 5 mM KCl in adenylation. In the future, RNA ligase candidates should be dialyzed against buffer with low concentrations of salt (e.g. KCl) and re-tested.

Little control was kept over pH in these assays. Candidates were eluted into Tris-HCl storage buffers of pH 7.5 at 25°C. In adenylation assays, the reaction mix by NEB, which contains sodium acetate buffer at pH 6.0 (no pH shift with temperature) (Zhelkovsky and McReynolds, 2011) was used. As the pH of Tris depends on the temperature, the real pH of the elution buffer at 65°C is calculated to be 6.38 pH ( $\Delta pK_a/^\circ C = -0.028$ ) (Stoll and Blanchard, 2009). MthRnl is known to have optimal adenylation activity in the 6.0-7.5 pH range (Zhelkovsky and McReynolds, 2012), hence the pH is not expected to be a major contribution to the suboptimal adenylation activity.

X03, X06, X07, and X12 were tested at 65°C because they originated from thermophilic organisms and the positive control (MthRnl) attains optimal activity at this temperature (Torchia, Takagi and Ho, 2008). Adenylation by X07 (10 min assay length) increased strongly from 65°C and onwards to the maximum tested temperature of 80°C. This could imply that this candidate is more thermostable than MthRnl, which in a comparable step 1 activity assay (5 min assay length) showed reduced product formation within this temperature range (65-85°C) (Torchia, Takagi and Ho, 2008). X06 was tested in the range 60-85°C and did not show increasing adenylation at higher temperatures. Rather, temperature sensitivity of X06 activity is more similar to that of MthRnl. Different from MthRnl is however that X06 could not be inactivated by heat-kill at 85°C for 5 minutes, because X06 retained activity at this temperature. A problem when interpreting X06 temperature-adenylation curves is that DNA was trapped in the wells and could not be released by proteinase K.

These insights into temperature tolerance are useful for developing candidates for preparative adenylation. However, the optimal temperature is not an intrinsic feature of the enzyme, but results from the balance between Arrhenius' law and denaturation (Creighton, 2010). For the composite ligation assay, the candidate is expected to be stable enough at 65°C to have activity during one hour of adenylation followed by two hours of ligation. MthRnl is known to be active at 65°C for two hours (Zhelkovsky and McReynolds, 2011). Buffer conditions could compromise the temperature stability of the candidates, and lead to inactivation after one hour at 65°C. In future experiments it would be interesting to test whether supplying more ligase after adenylation enables ligation.

In addition to the stability requirements, another drawback of the composite ligation assay is its inevitable dependence on adenylation efficiency as mentioned earlier. The isolated step 3 assay circumvents the negative impact of poor substrate adenylation on step 3, but is based on the assumption that the candidates can utilize preadenylated donors. This requires that ligases are non-adenylated, have affinity for, and are able to bind stably to the preadenylated donors (Zhelkovsky and McReynolds, 2011). Inability to use preadenylated donors would be unexpected, given that multiple closely and more distantly related RNA ligases can do this (MthRnl, T4Rnl1, T4Rnl2) (Wang, Schwer and Shuman, 2006; Zhelkovsky and McReynolds, 2011). There have however been indications of that ligation of exogenous preadenylated

donors poses different requirements than that of *in situ* formed step 2 product (Wang, Schwer and Shuman, 2006).

Problematic for both ligation assays are the competing reverse step 2 reactions (deadenylation) in which App-donors are hydrolyzed to non-adenylated donors. High levels of deadenylation were observed during composite and isolated step 3 ligation assays. It is known that deadenylation increases under non-optimal temperature conditions (Torchia, Takagi and Ho, 2008). To reduce deadenylation, a mutant MthRnl variant with diminished step 2 activity has already been developed (NEB, #M0319).

A more fundamental problem could be related to required substrate characteristics. The Rnl3 family of ligases is relatively poorly explored, and members' *in vivo* substrates and pathways are unknown. In this thesis, 20- and 23-mer ssDNA and ssRNA with no predicted secondary structure were tested. On multiple occasions accumulation of substrate and/or product was observed in the wells during gel electrophoresis. E.g. addition of X06 to ssRNA acceptors in the isolated step 3 assay induced deposition of ssRNA in wells, which can be interpreted as protein binding the ssRNA. For adenylation, knowledge of affinity is useful. However, substrate affinity does not necessarily imply dimerization, as MthRnl has been found to complete steps 1 and 2 without dimerization (Torchia, Takagi and Ho, 2008). The authors found that adenylation activity was compromising when dimerization was precluded, which could be part of the explanation for suboptimal adenylation activity.

## **5.4 Chromatographical Purification of Rnl3-Family Ligases**

Sample quality becomes a confounding variable in activity assays and interferes with melting curve analysis by TSA. Here, the performance of the 2-step IMAC-IEX purification protocol will be discussed along with proposed changes to the protocol.

Activity of all candidates was increased after purification with the 2-step protocol, compared to the 1-step IMAC (His SpinTrap) protocol, indicating increased specific activity. It is conceivable that removal of unrelated proteins contributing to protein measurements plays a role in this result. Connected to this, one should ask whether purity after IEX is sufficient for functional assays. In this thesis, purity is judged by Coomassie-stained SDS-PAGE analysis which reveals prominent contaminants such as the 25- and 70-kDa proteins, and more faint bands which are hardly discernable from the background.

IMAC elution profiles revealed by SDS-PAGE are characterized by wide peaks, i.e. target candidate is eluted with similar intensities throughout the gradient. For candidates with low expression (all except X07- only one fraction was chosen) most of the IMAC fractions were pooled for IEX in order to optimize recovery. Therefore, persisting contaminants in IMAC and IEX are a recurring problem e.g. the prominent 25-kDa contaminant (seen in X03, X07, X12, X19), which is poorly separated from the target in IEX. It is not known whether this contaminant interferes directly with the assay, but removing it would nevertheless improve the sample quality. This could be attained by improving resolution and stricter selection of fractions for the subsequent/following purification step.

Early eluting contaminants e.g. the prominent 70-kDa protein could be removed by increasing the imidazole concentration used in the sample (lysis), binding and wash buffer. Instead of 20 mM as used here (lysis buffer C; used for all candidates except X07), 40 mM imidazole or more could be tested in the sample and binding buffers. By increasing imidazole to 100 mM in the wash buffer the majority of the 70 kDa contaminant, which makes up most of the early eluting fractions, could be removed.

The prominent 25-kDa contaminant, however, would be more difficult to avoid by changing buffer compositions in IMAC, because it elutes later in the gradient together with the target protein (seen in X03, X07, X12, X19). This contaminant was also detected by western blotting, indicating the presence of solvent-exposed histidine patches. It cannot be ruled out that this could be a truncated form of the candidate ligase, because the His-tag is located at the N-terminus. In the future, samples of culture from non-transformed host cells and cells transformed with an empty (parent) vector should be run as controls. As the 25-kDa contaminant has similar physicochemical properties as the target protein one could either attempt to maximize charge differences in IEX by testing different buffer pH, or add a third purification method such as SEC.

The downside of an SDS-PAGE-based purity analysis is that even at above 95% purity (e.g. X19), the remaining 5% could be strongly interfering. In addition, when the gel is not overloaded (for all candidates except X07) the contaminant might be present at a concentration below the limit of detection (Raynal *et al.*, 2014). Therefore, the phosphatase

and nuclease assays should be improved (see Section 5.5) and used to evaluate the intermediate purification (and polishing) step.

Aside from purity, the purification protocol has to ensure the integrity of the candidate (Raynal *et al.*, 2014). Results showed that X06 was unstable in the IMAC buffer used in the 2-step protocol. In the future, stability tests could be performed in a His SpinTrap format. X06, X03 and X19 appeared to form multimers after IEX purification, which was surprising, as SDS is expected to disrupt aggregation. Multimers were most prominent in fractions with the highest protein concentrations and were hence believed to be artefacts due to running highly concentrated and pure protein on a gel. However, fractions of other candidates e.g. X07 with even higher protein concentrations did not display aggregates. Apparent multimers can also be hypothesized to be oxidization-related. This is supported by that X03, X06, and X19 have the highest number of cysteines among the four candidates (Kurien and Scofield, 2012). The true nature of the multimers could be elucidated by western blotting and additional purification by SEC. This interpretation also agrees with the presence of the double band (a slightly slower migrating band which co-eluted with the target candidate = likely it *is* the candidate). This is a slightly slower migrating species, which could be a modified version of the candidate ligase. Mass spectrometry could be employed to determine the identity of this species.

An important question is whether the purification strategy is useful for purifying these candidates. From Coomassie-stained SDS-PAGE analysis it becomes clear that the level of purity is low after the first purification step (IMAC). Does purification by IEX substantially increase purity, or would a different strategy such as IMAC-(IEX)-SEC be more useful? RNA ligases belonging to the Rnl3 family are expected to be homodimers in their functional form (Becker, L'Hermitte-Stead and Myllykallio, 2019). Inactivity in ligation assays could result from the lack of homodimerization. The oligomerization state in the samples could be examined with SEC. SEC would also separate aggregated protein, which is surmised to interfere in melting curve analysis by TSA. Benefits of SEC include the tolerance of a wide variety of buffer conditions. Therefore, the candidate could be eluted into a buffer optimized for stability. SEC would also be a very useful method for initial stability assays. If dimerization is assumed to be a prerequisite for activity, SEC could be a supporting assay for the stability of candidates in their functional form.

If SEC is to be implemented as a second step in the purification protocol, the previous step (IMAC) needs to be optimized, because SEC is a low-throughput method. With the current purification scheme, all candidates show strong peak tailing (wide peaks) when eluting with a gradient. More narrow peaks could be obtained by eluting in steps, which would result in a more concentrated sample (GE Healthcare, 2016). Based on IMAC elution/protein profiles of X03, X12, and X19 steps of 200, 300, and 500 mM imidazole appear to be a reasonable starting point. Fractions collected from these steps could in addition be used to investigate the (apparent) double peaks observed in IMAC purification of X03, X07, X12, and X19. These two peaks (eluting at approximately 200 mM and 500 mM imidazole) could represent different multimeric forms (improper folding) or be artefacts (sudden increase in imidazole concentration). By collecting the two peaks in separate fractions, they could be examined in SEC and activity assays.

## **5.5 Discussion of Analytical Methods: TSA, Nuclease and Phosphatase Assay**

Performance in assays could be negatively influenced by lack of thermostability and presence of specific DNA/RNA-modifying enzymes. Standard protocols for TSA, and nuclease and phosphatase assays were modified for testing the four active candidates X03, X06, X07, and X12, and X19.

Due to difficulties with confirming reaction step 3 (formation of phosphodiester bonds), where the assay involved incubation at 65°C for 2 hours (3 hours in the composite ligation assay), more information about temperature stability was sought. Thermal stability of candidates X06 and X07 was assessed by TSA with the aim of gaining a better understanding of thermostability and influence of buffer composition on stability. No interpretable curves were obtained, as clear temperature transitions were missing. It is known that 15-25% of recombinant proteins for various reasons do not show interpretable temperature curves (Huynh and Partch, 2015). Common reasons for this are intrinsic disorder (lack of globular fold), instability at room temperature, lack of hydrophobic core, and surface hydrophobic patches (Huynh and Partch, 2015). As X06 and X07 have been shown to possess adenylation activity at 65°C it would seem unlikely that they are highly unstable at room temperature. The related archaeal RNA ligase Pab1020 has also been found to be packed into a globular fold, and its structure is surmised to be highly similar to the one of X07 (Becker, L'Hermitte-Stead

and Myllykallio, 2019). Given that MthRnl also did not yield melting curves, the problem could be related to this specific class of RNA ligases. This is supported by that DNA and RNA ligases belonging to other families have a solvent-exposed and hydrophobic ATP-binding pocket (Banerjee *et al.*, 2019). It is conceivable that the binding pocket could accommodate SYPRO Orange, and hence make the dye unavailable.

Another possible explanation is the presence of aggregates, as the sample was not centrifuged or filtered prior to use in TSA. Positive controls in this TSA were commercial MthRnl and an in-house (ArcticZymes) produced DNA ligase, resembling T4 DNA ligase. Notably, as with X06 and X07, MthRnl also did not yield an interpretable temperature shift, while the mesophilic DNA ligase on the other hand showed clear temperature-induced denaturation at around 40°C. Lack of melting point for thermophile-derived X06 and X07 is not likely to result from instability of SYPRO Orange at higher temperatures (Dupeux *et al.*, 2011). Perhaps differences in the conformation of the ATP-binding site between Rnl3 family ligases and DNA ligases could explain this. Unspecific binding at room temperature to either aggregated proteins or hydrophobic patches is expected to result in a notably higher initial fluorescence signal, which was not detected here.

Phosphatase activity was assessed in available samples taken prior to purification, after step 1 (IMAC), and after step 2 (IEX). Activity at a high possibly interfering level (46  $\mu\text{U}/\mu\text{l}$  - 138  $\mu\text{U}/\mu\text{l}$ ) could only be measured in the non-purified sample with a coefficient of variance of 20%. The minimum detectable concentration was limited by that the x-intercept of the standard curve did not go through the origin. As blanks containing phosphate buffer and pNPP substrate had been subtracted, this could be due to instrumental error, noise and sample handling. Purified samples generated phosphatase measurements below the dynamic range with CVs above 100% showing that measurements are hardly discernible from background, hence it was not possible to detect significant amounts of phosphatase contamination. NEB accepts a contaminating phosphatase activity of below  $10^{-4}$  U/ $\mu\text{l}$  after 16 h in 50 pmol enzyme in their quality certificates connected to MthRnl (New England Biolabs, 2020). We tested 50 and 100 pmol of purified enzyme in the course of only 2 h and were not able to detect significant phosphatase activities down to  $2.5 \cdot 10^{-7}$  U/ $\mu\text{l}$ . Based on this, phosphatase concentrations falling outside the dynamic range of the present assay are assumed not to have a significant impact on RNA ligase performance.

Endonuclease activity was tested with bacteriophage ssDNA substrate, similar to the assay used by NEB (vendor). As for the phosphatase assay, samples at different stages of the purification process were tested. Degradation was only clearly observed in nonpurified samples. Other samples showed smears, DNA fragments running higher than the intact substrate, and gel shifts. Gel shift were prominent in samples of X06, where bands migrated slower at higher sample concentrations. Curiously, treatment with Proteinase K did not reduce the gel shifts. Exonuclease activity is clearly detectable in gel-based assays, however these tests of endonuclease activity did not yield interpretable results.

## 6 Conclusion

In this thesis we explore 22 putative members of the Rnl3 family with the aim of identifying promising enzyme leads. From the 22 gene sequences we could identify four novel thermophilic candidate RNA ligases (X03, X06, X07, X12) that were able to adenylate ssDNA with 50-90% efficiency and represent interesting enzyme leads.

Half of the candidates could be expressed in *E. coli* in soluble form, and four of these showed replicable adenylyltransferase activity. Activity in initial screening was strongly correlated with soluble expression levels from recombinant production in *E. coli*. Soluble-to-insoluble ratios of candidates with good expression could not be improved by lowering inducer concentration and cell density at induction. The functional screening protocol provided a fast method for initial screening of ssRNA ligases but was affected by contaminating nuclease and phosphatase activities and salt variation in reaction composition.

Purity after a one-step IMAC purification was not satisfactory. The 2-step IMAC-IEX protocol increased the specific activity of candidates, and reduced exonuclease and phosphatase contamination. Recombinant production of the four active candidates was successfully upscaled, and X06 and X07 were expressed mainly in soluble form. Candidate X06 appeared to be unstable in the FPLC buffers used here.

Two new protocols for assessing ligation activity were developed and validated with commercial MthRnl. rSAP treatment could successfully deplete ATP and enable catalysis of step 3 by MthRnl. None of the active candidates showed detectable ligation activity of ssDNA donor and ssDNA/ssRNA acceptors in these assays.

Melting curves of neither candidates X06 and X07, nor commercial MthRnl could be revealed by TSA most likely due to enzymatical features. X06 and X07 could efficiently adenylate 50% of the substrate within 5 and 10 minutes at 65°C, respectively. Activity of X07 increased with temperature between 65 and 80°C (maximal tested), while X06 was less active above 75°C. App-donor yield was not increased above 50% at ATP concentrations up to 1 mM, and at 10 mM ATP adenylation was absent.

## 7 Further Work

Known members of the Rnl3 family are characterized by that they are homodimers in their active form. This requirement for activity was not addressed in this thesis. In the future, the oligomerization state of active candidates identified here should be studied e.g. by analytical SEC. Related to this, the double peaks observed in chromatograms should be investigated further, because they may represent different oligomers of the candidates. SEC can also be used for buffer optimization to promote dimerization (if dimerization is observed to have a positive effect on activity) and increase stability of X06.

It could also be interesting to incorporate SEC in the purification protocol in order to remove aggregated protein and for the reasons mentioned above. An alternative 2-step protocol for candidates with poor expression could consist of an IMAC-step followed directly by SEC. For X07, which has good expression levels, a 3-step IMAC-IEX-SEC purification method could be considered.

In the future, gaining more insight into favourable reaction conditions will be important to control and reduce the competing deadenylation activity. With this knowledge substrate adenylation could also be optimized, which is desired for commercial applications in preparative adenylation and for improving the composite ligation assay. Based on adenylation by X02 (in-house MthRnl) and Lama and Ryan's optimizing of TS2126 Rnl1 activity we are optimistic about attaining complete substrate adenylation with the new candidates (Lama and Ryan, 2016).

Using rSAP to deplete ATP in the one-tube composite ligation assay and for clean-up after preparative adenylation circumvents the need for an adenylation-negative ligase and laborious precipitation- or chromatography-based clean-up methods. This approach has to our knowledge not been described previously and should be detailed further in the future as it may comprise patentable novelty.

## 8 References

- Alegria-Schaffer, A., Lodge, A. and Vattem, K. (2009) *Performing and Optimizing Western Blots with an Emphasis on Chemiluminescent Detection*. 1st edn, *Methods in Enzymology*. 1st edn. Elsevier Inc. doi: 10.1016/S0076-6879(09)63033-0.
- Amann, R. I., Ludwig, W. and Schleifer, K. H. (1995) 'Phylogenetic identification and in situ detection of individual microbial cells without cultivation', *Microbiological reviews*, 59(1), pp. 143–69. doi: 10.1128/membr.59.1.143-169.1995.
- ArcticZymes (2020) *Cutting-edge enzymes from Norway*. Available at: <https://arcticzymes.com/about-us/> (Accessed: 1 August 2020).
- Atkinson, H. J. *et al.* (2009) 'Using Sequence Similarity Networks for Visualization of Relationships Across Diverse Protein Superfamilies', *PLoS ONE*. Edited by I. K. Jordan, 4(2), p. e4345. doi: 10.1371/journal.pone.0004345.
- Banerjee, A. *et al.* (2019) 'Structure and two-metal mechanism of fungal tRNA ligase', *Nucleic Acids Research*. Oxford University Press, 47(3), pp. 1428–1439. doi: 10.1093/nar/gky1275.
- Becker, H. F., L'Hermitte-Stead, C. and Myllykallio, H. (2019) 'Diversity of circular RNAs and RNA ligases in archaeal cells', *Biochimie*. Elsevier Ltd, 164, pp. 37–44. doi: 10.1016/j.biochi.2019.06.011.
- Bioprospecting* (2020) *United Nations Development Programme*. Available at: <https://www.sdfinance.undp.org/content/sdfinance/en/home/solutions/bioprospecting.html> (Accessed: 21 June 2020).
- Bisswanger, H. (2014) 'Enzyme assays', *Perspectives in Science*. Elsevier, 1(1–6), pp. 41–55. doi: 10.1016/j.pisc.2014.02.005.
- Blondal, T. *et al.* (2003) 'Discovery and characterization of a thermostable bacteriophage RNA ligase homologous to T4 RNA ligase 1', *Nucleic Acids Research*, 31(24), pp. 7247–7254. doi: 10.1093/nar/gkg914.
- Bolanos-Garcia, V. M. and Davies, O. R. (2006) 'Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography', *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1760(9), pp. 1304–1313. doi: 10.1016/j.bbagen.2006.03.027.
- Bond, P. S. (2017) *JTSA*. Available at: <http://paulsbond.co.uk/jtsa> (Accessed: 19 July 2020).

- Brooks, M. A. *et al.* (2008) 'The structure of an archaeal homodimeric ligase which has RNA circularization activity', *Protein Science*, 17(8), pp. 1336–1345. doi: 10.1110/ps.035493.108.
- Burgess, R. R. (2009) 'Use of Bioinformatics in Planning a Protein Purification', in *Methods in Enzymology*. 1st edn. Elsevier Inc., pp. 21–28. doi: 10.1016/S0076-6879(09)63003-2.
- Chambers, C. R. and Patrick, W. M. (2015) 'Archaeal Nucleic Acid Ligases and Their Potential in Biotechnology', *Archaea*, 2015, pp. 1–10. doi: 10.1155/2015/170571.
- Cheng, K. *et al.* (2019) 'RNA ligation of very small pseudo nick structures by T4 RNA ligase 2, leading to efficient production of versatile RNA rings', *RSC Advances*. Royal Society of Chemistry, 9(15), pp. 8620–8627. doi: 10.1039/C9RA01513B.
- Chou, C. P. (2007) 'Engineering cell physiology to enhance recombinant protein production in Escherichia coli', *Applied Microbiology and Biotechnology*, 76(3), pp. 521–532. doi: 10.1007/s00253-007-1039-0.
- Christendat, D. *et al.* (2000) 'Structural proteomics of an archaeon.', *Nature structural biology*, 7(10), pp. 903–9. doi: 10.1038/82823.
- Creighton, T. E. (2010) *The Biophysical chemistry of nucleic acids & proteins*. Gardners Books.
- Demain, A. (2001) 'Genetics and microbiology of industrial microorganism: Molecular genetics and industrial microbiology—30 years of marriage', *Journal of Industrial Microbiology and Biotechnology*, 27, pp. 352–356. doi: 10.1038/sj.jim.7000187.
- Demain, A. L. and Vaishnav, P. (2009) 'Production of recombinant proteins by microbes and higher organisms', *Biotechnology Advances*. Elsevier Inc., 27(3), pp. 297–306. doi: 10.1016/j.biotechadv.2009.01.008.
- Dumorne, K. *et al.* (2017) 'Extremozymes: A Potential Source for Industrial Applications', *Journal of Microbiology and Biotechnology*, 27(4), pp. 649–659. doi: 10.4014/jmb.1611.11006.
- Dupeux, F. *et al.* (2011) 'A thermal stability assay can help to estimate the crystallization likelihood of biological samples', *Acta Crystallographica Section D Biological Crystallography*, 67(11), pp. 915–919. doi: 10.1107/S0907444911036225.
- Egorova, K. and Antranikian, G. (2005) 'Industrial relevance of thermophilic Archaea', *Current Opinion in Microbiology*, 8(6), pp. 649–655. doi: 10.1016/j.mib.2005.10.015.
- England, T. E. and Uhlenbeck, O. C. (1978) '3'-Terminal labelling of RNA with T4 RNA ligase', *Nature*, 275(October), pp. 560–561. doi: 10.1038/275560a0.
- Ericsson, U. B. *et al.* (2006) 'Thermofluor-based high-throughput stability optimization of proteins for structural studies', *Analytical Biochemistry*, 357(2), pp. 289–298. doi: 10.1016/j.ab.2006.07.027.

Ferrer, M. *et al.* (2007) 'Mining enzymes from extreme environments', *Current Opinion in Microbiology*, 10(3), pp. 207–214. doi: 10.1016/j.mib.2007.05.004.

Fuchs, R. T. *et al.* (2015) 'Bias in Ligation-Based Small RNA Sequencing Library Construction Is Determined by Adaptor and RNA Structure', *PLOS ONE*. Edited by B. Zhang, 10(5), p. e0126049. doi: 10.1371/journal.pone.0126049.

Garfin, D. E. (1990) 'One-dimensional gel electrophoresis', in *Guide to protein purification*, pp. 425–441. doi: 10.1016/0076-6879(90)82035-Z.

Gawin, A., Valla, S. and Brautaset, T. (2017) 'The XylS/Pm regulator/promoter system and its use in fundamental studies of bacterial gene expression, recombinant protein production and metabolic engineering', *Microbial Biotechnology*, 10(4), pp. 702–718. doi: 10.1111/1751-7915.12701.

GE Healthcare (2010) *Strategies for Protein Purification*. Uppsala. Available at: <https://www.cytivalifesciences.com/en/us/support/handbooks>.

GE Healthcare (2016) *Ion Exchange Chromatography*. Uppsala. Available at: <https://www.cytivalifesciences.com/en/us/support/handbooks>.

Hafner, M. *et al.* (2011) 'RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries', *Rna*, 17(9), pp. 1697–1712. doi: 10.1261/rna.2799511.

Haque, R. U., Paradisi, F. and Allers, T. (2020) 'Haloferax volcanii for biotechnology applications: challenges, current state and perspectives', *Applied Microbiology and Biotechnology*. Applied Microbiology and Biotechnology, 104(4), pp. 1371–1382. doi: 10.1007/s00253-019-10314-2.

Hayashi, K. *et al.* (1986) 'Regulation of inter- and intramolecular ligation with T4 DNA ligase in the presence of polyethylene glycol', *Nucleic Acids Research*, 14(19), pp. 7617–7631. doi: 10.1093/nar/14.19.7617.

Ho, C. K. and Shuman, S. (2002) 'Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains', *Proceedings of the National Academy of Sciences*, 99(20), pp. 12709–12714. doi: 10.1073/pnas.192184699.

Huynh, K. and Partch, C. L. (2015) 'Analysis of Protein Stability and Ligand Interactions by Thermal Shift Assay', *Current Protocols in Protein Science*, 79(1), pp. 28.9.1–28.9.14. doi: 10.1002/0471140864.ps2809s79.

Jayaprakash, A. D. *et al.* (2011) 'Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing', *Nucleic Acids Research*, 39(21), pp. e141–e141. doi: 10.1093/nar/gkr693.

Jeon, S. and Ishikawa, K. (2003) 'A novel ADP-dependent DNA ligase from *Aeropyrum pernix* K1', *FEBS Letters*, 550(1–3), pp. 69–73. doi: 10.1016/S0014-5793(03)00821-4.

Jonstrup, S. P. (2006) 'A microRNA detection system based on padlock probes and rolling circle amplification', *RNA*, 12(9), pp. 1747–1752. doi: 10.1261/rna.110706.

Jost, I. *et al.* (2018) 'Functional sequestration of microRNA-122 from Hepatitis C Virus by circular RNA sponges', *RNA Biology*. Taylor & Francis, 15(8), pp. 1–8. doi: 10.1080/15476286.2018.1435248.

Kim, D. (2019) 'The architecture of SARS-CoV-2 transcriptome', *Journal of Chemical Information and Modeling*, 53(9), pp. 1689–1699. doi: 10.1017/CBO9781107415324.004.

Kim, R. *et al.* (1998) 'Overexpression of archaeal proteins in Escherichia coli', *Biotechnology Letters*, 20(3), pp. 207–210. doi: 10.1023/A:1005305330517.

Kornberg, A. (2009) 'Why purify enzymes?', in *Guide to protein purification*, pp. 3–5.

Krzywkowski, T. and Nilsson, M. (2017) 'Fidelity of RNA templated end-joining by chlorella virus DNA ligase and a novel iLock assay with improved direct RNA detection accuracy', *Nucleic Acids Research*. Oxford University Press, 45(18), pp. e161–e161. doi: 10.1093/nar/gkx708.

Kurien, B. T. and Scofield, R. H. (2012) 'Common Artifacts and Mistakes Made in Electrophoresis', in, pp. 633–640. doi: 10.1007/978-1-61779-821-4\_58.

Lama, L. and Ryan, K. (2016) 'Adenylation of small RNA sequencing adapters using the TS2126 RNA ligase I', *RNA*, 22(1), pp. 155–161. doi: 10.1261/rna.054999.115.

Larman, H. B. *et al.* (2014) 'Sensitive, multiplex and direct quantification of RNA sequences using a modified RASL assay', *Nucleic Acids Research*, 42(14), pp. 9146–9157. doi: 10.1093/nar/gku636.

Liu, X. and Gorovsky, M. A. (1993) 'Mapping the 5' and 3' ends of Tetrahymena thermophila mRNAs using RNA ligase mediated amplification of cDNA ends (RLM-RACE)', *Nucleic Acids Research*, 21(21), pp. 4954–4960. doi: 10.1093/nar/21.21.4954.

Lohman, G. (2019) *The effect of nucleic acid modifications on digestion by DNA exonucleases*, New England Biolabs, Inc. Available at: <https://international.neb.com/tools-and-resources/feature-articles/the-effect-of-nucleic-acid-modifications-on-digestion-by-dna-exonucleases> (Accessed: 28 June 2020).

Martins, A. and Shuman, S. (2004) 'An RNA Ligase from Deinococcus radiodurans', *Journal of Biological Chemistry*, 279(49), pp. 50654–50661. doi: 10.1074/jbc.M407657200.

Morris, K. V. and Mattick, J. S. (2014) 'The rise of regulatory RNA', *Nature Reviews Genetics*, 15(6), pp. 423–437. doi: 10.1038/nrg3722.

Nandakumar, J. *et al.* (2004) 'RNA Substrate Specificity and Structure-guided Mutational Analysis of Bacteriophage T4 RNA Ligase 2', *Journal of Biological Chemistry*, 279(30), pp. 31337–31347. doi:

10.1074/jbc.M402394200.

Nandakumar, J., Shuman, S. and Lima, C. D. (2006) 'RNA Ligase Structures Reveal the Basis for RNA Specificity and Conformational Changes that Drive Ligation Forward', *Cell*, 127(1), pp. 71–84. doi: 10.1016/j.cell.2006.08.038.

New England Biolabs (2020) *5' DNA Adenylation Kit Certificate of Analysis*. Ipswich. Available at: [https://international.neb.com/products/e2610-5-dna-adenylation-kit#Quality, Safety & Legal\\_Certificate of Analysis](https://international.neb.com/products/e2610-5-dna-adenylation-kit#Quality, Safety & Legal_Certificate of Analysis).

Noble, J. E. and Bailey, M. J. A. (2009) 'Quantitation of Protein', in *Methods in Enzymology*. 1st edn. Elsevier Inc., pp. 73–95. doi: 10.1016/S0076-6879(09)63008-1.

Oren, A. (2013) 'Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes', *Frontiers in Microbiology*, 4(NOV), pp. 1–6. doi: 10.3389/fmicb.2013.00315.

Parikh, U. M. *et al.* (2017) 'Future technologies for monitoring HIV drug resistance and cure', *Current Opinion in HIV and AIDS*, 12(2), pp. 182–189. doi: 10.1097/COH.0000000000000344.

Pascal, J. M. (2008) 'DNA and RNA ligases: structural variations and shared mechanisms', *Current Opinion in Structural Biology*, 18(1), pp. 96–105. doi: 10.1016/j.sbi.2007.12.008.

Petkovic, S. and Müller, S. (2015) 'RNA circularization strategies in vivo and in vitro', *Nucleic Acids Research*, 43(4), pp. 2454–2465. doi: 10.1093/nar/gkv045.

Primrose, Sandy B. Twyman, R. (2009) *Principles of Gene Manipulation and Genomics*.

Raynal, B. *et al.* (2014) 'Quality assessment and optimization of purified protein samples: why and how?', *Microbial Cell Factories*, 13(1), p. 180. doi: 10.1186/s12934-014-0180-6.

Rhodes, D. G. and Laue, T. M. (2009) 'Determination of Protein Purity', in *Methods in Enzymology*. 1st edn. Elsevier Inc., pp. 677–689. doi: 10.1016/S0076-6879(09)63038-X.

Rosano, G. L. and Ceccarelli, E. A. (2009) 'Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted Escherichia coli strain', *Microbial Cell Factories*, 8, pp. 1–9. doi: 10.1186/1475-2859-8-41.

Rosano, G. L. and Ceccarelli, E. A. (2014) 'Recombinant protein expression in Escherichia coli: advances and challenges', *Frontiers in Microbiology*, 5(APR), pp. 1–17. doi: 10.3389/fmicb.2014.00172.

Rosano, G. L., Morales, E. S. and Ceccarelli, E. A. (2019) 'New tools for recombinant protein production in Escherichia coli : A 5-year update', *Protein Science*, 28(8), pp. 1412–1422. doi: 10.1002/pro.3668.

- Seidl, C. I. and Ryan, K. (2011) 'Circular Single-Stranded Synthetic DNA Delivery Vectors for MicroRNA', *PLoS ONE*. Edited by J.-L. Darlix, 6(2). doi: 10.1371/journal.pone.0016925.
- Shuman, S. and Lima, C. D. (2004) 'The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases', *Current Opinion in Structural Biology*, 14(6), pp. 757–764. doi: 10.1016/j.sbi.2004.10.006.
- Singha, T. K. *et al.* (2017) 'Efficient genetic approaches for improvement of plasmid based expression of recombinant protein in Escherichia coli : A review', *Process Biochemistry*. Elsevier Ltd, 55, pp. 17–31. doi: 10.1016/j.procbio.2017.01.026.
- Smith, P. *et al.* (2012) 'The adenylyltransferase domain of bacterial Pnkp defines a unique RNA ligase family', *Proceedings of the National Academy of Sciences*, 109(7), pp. 2296–2301. doi: 10.1073/pnas.1116827109.
- Stoll, V. S. and Blanchard, J. S. (2009) 'Buffers: Principles and Practice', in *Methods in Enzymology*. Academic Press, pp. 43–56. doi: 10.1016/S0076-6879(09)63006-8.
- Tanaka, N., Meineke, B. and Shuman, S. (2011) 'RtcB, a Novel RNA Ligase, Can Catalyze tRNA Splicing and HAC1 mRNA Splicing in Vivo', *Journal of Biological Chemistry*, 286(35), pp. 30253–30257. doi: 10.1074/jbc.C111.274597.
- Tang, Z. (2003) 'Real-time monitoring of nucleic acid ligation in homogenous solutions using molecular beacons', *Nucleic Acids Research*, 31(23), pp. 148e – 148. doi: 10.1093/nar/gng146.
- Timson, D. J., Singleton, M. R. and Wigley, D. B. (2000) 'DNA ligases in the repair and replication of DNA', *Mutation Research/DNA Repair*, 460(3–4), pp. 301–318. doi: 10.1016/S0921-8777(00)00033-1.
- Torchia, C., Takagi, Y. and Ho, C. K. (2008) 'Archaeal RNA ligase is a homodimeric protein that catalyzes intramolecular ligation of single-stranded RNA and DNA', *Nucleic Acids Research*, 36(19), pp. 6218–6227. doi: 10.1093/nar/gkn602.
- Urh, M., Simpson, D. and Zhao, K. (2009) 'Affinity Chromatography: General Methods', in *Guide to protein purification*, pp. 417–435.
- Vester, J. K., Glaring, M. A. and Stougaard, P. (2015) 'Improved cultivation and metagenomics as new tools for bioprospecting in cold environments', *Extremophiles*, 19(1), pp. 17–29. doi: 10.1007/s00792-014-0704-3.
- Vieille, C. and Zeikus, G. J. (2001) 'Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability', 65(1), pp. 1–43. doi: 10.1128/MMBR.65.1.1–43.2001.
- Wang, L. K., Schwer, B. and Shuman, S. (2006) 'Structure-guided mutational analysis of T4 RNA ligase 1', *RNA*, 12(12), pp. 2126–2134. doi: 10.1261/rna.271706.

Winther-Larsen, H. C. *et al.* (2000) 'Parameters Affecting Gene Expression from the Pm Promoter in Gram-Negative Bacteria', *Metabolic Engineering*, 2(2), pp. 79–91. doi: 10.1006/mben.1999.0142.

Yin, S. *et al.* (2004) 'Characterization of bacteriophage KVP40 and T4 RNA ligase 2', *Virology*, 319(1), pp. 141–151. doi: 10.1016/j.virol.2003.10.037.

Zhang, L. and Tripathi, A. (2017) 'Archaeal RNA ligase from thermococcus kodakarensis for template dependent ligation', *RNA Biology*. Taylor & Francis, 14(1), pp. 36–44. doi: 10.1080/15476286.2016.1239688.

Zhelkovsky, A. M. and McReynolds, L. A. (2011) 'Simple and efficient synthesis of 5' pre-adenylated DNA using thermostable RNA ligase', *Nucleic Acids Research*, 39(17), pp. e117–e117. doi: 10.1093/nar/gkr544.

Zhelkovsky, A. M. and McReynolds, L. A. (2012) 'Structure-function analysis of Methanobacterium thermoautotrophicum RNA ligase – engineering a thermostable ATP independent enzyme', *BMC Molecular Biology*. BMC Molecular Biology, 13(1), p. 24. doi: 10.1186/1471-2199-13-24.

Zhuang, F. *et al.* (2012) 'Structural bias in T4 RNA ligase-mediated 3'-adapter ligation', *Nucleic Acids Research*, 40(7), pp. e54–e54. doi: 10.1093/nar/gkr1263.

Zwick, F., Lale, R. and Valla, S. (2013) 'Regulation of the expression level of transcription factor XylS reveals new functional insight into its induction mechanism at the Pm promoter', *BMC Microbiology*, 13(1), p. 262. doi: 10.1186/1471-2180-13-262.

## 9 Appendices

### Appendix A. Protein Data

Table A.1: Basic information about the candidates. ND: no data.

<b>Candidate ID</b>	<b>Theoretical molecular weight (Da)</b>	<b>Theoretical pI</b>
X01	44331,89	4,61
X02	43146,02	5,36
X03	43555,37	8,47
X04	43864,9	6,3
X05	45544,93	5,53
X06	45961,53	8,29
X07	43272,76	5,57
X08	45315,9	9,38
X09	46599,32	8,62
X10	42508,15	8,07
X11	43579,93	6,78
X12	42981,48	8,79
X13	44766,79	8,44
X14	42873,03	5,92
X15	41336,82	5,72
X16	42747,17	6,71

X17	42528,78	4,75
X18	44403,44	5,64
X19	41607,04	4,77
X20	40974,55	6,22
X21	42042,76	4,77
X22	ND	ND

# Appendix B. Complete Results of Initial Expression Screening

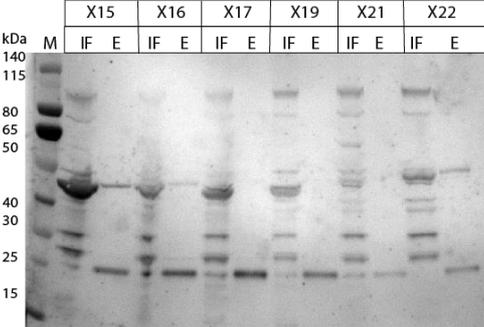
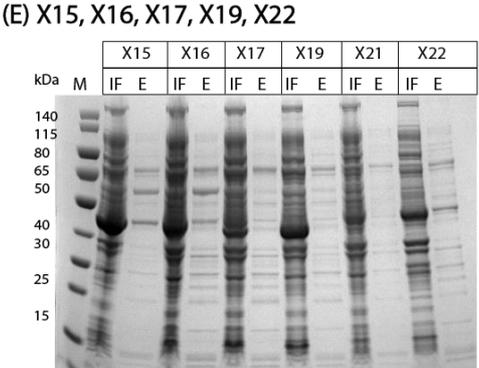
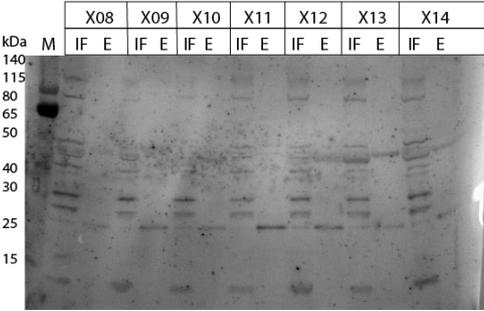
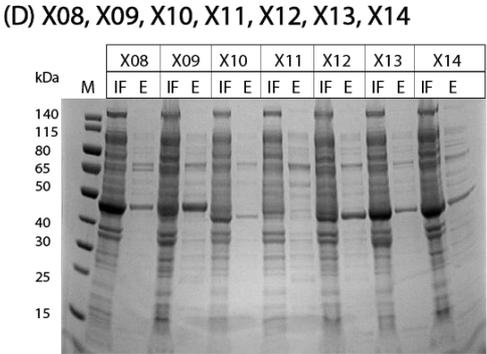
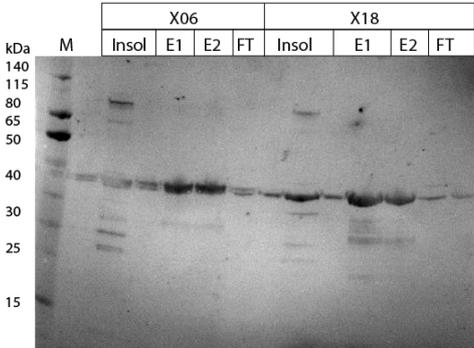
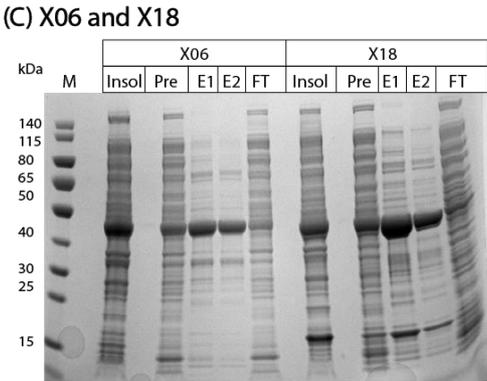
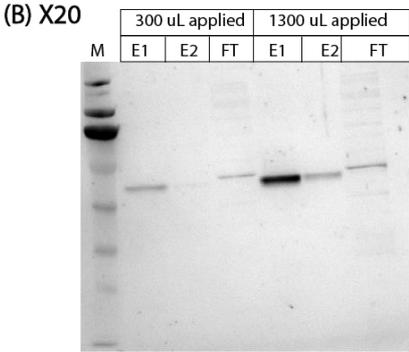
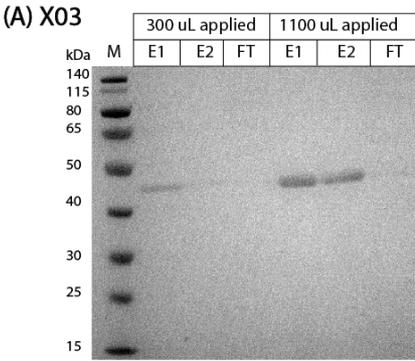
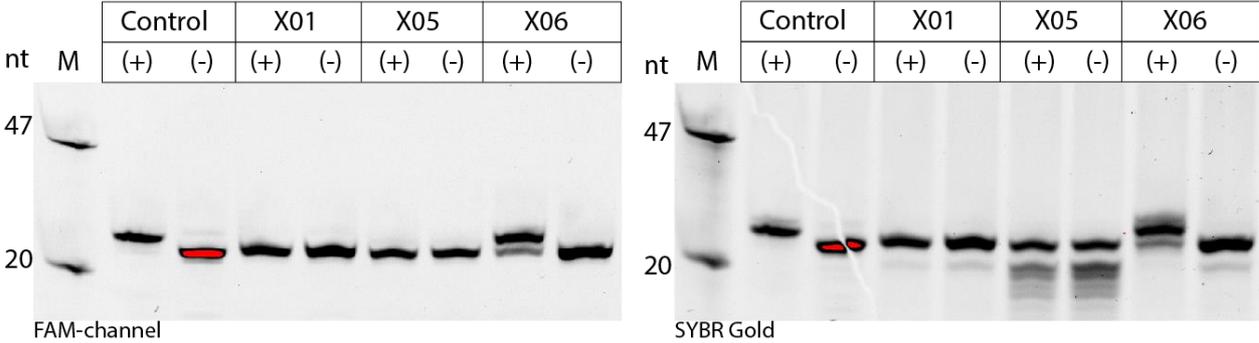


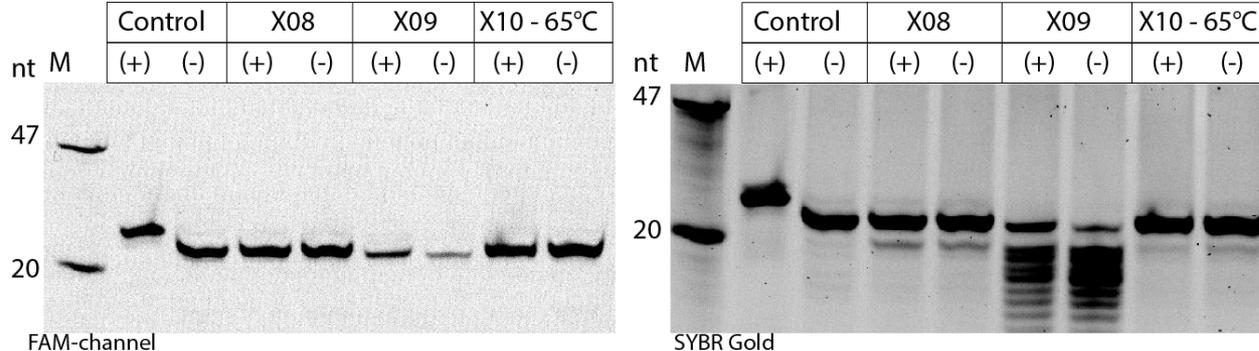
Figure B.1: **Remaining results of screening all 22 candidates for expression in *E. coli*.** Except for X03 and X20, analyses were run in parallel Coomassie-stained SDS-PAGE gels (left side) and western blots (right side). IF: insoluble fraction; E: eluate; Insol.: insoluble fraction; FT: flow through; Pre: pre-purification sample

# Appendix C. Complete Results of Initial ssDNA Adenylation Screening

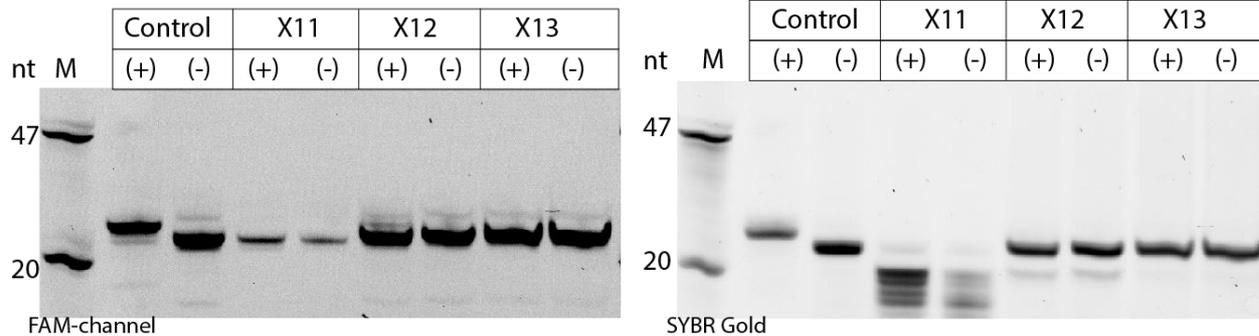
(A) X01, X05, X06



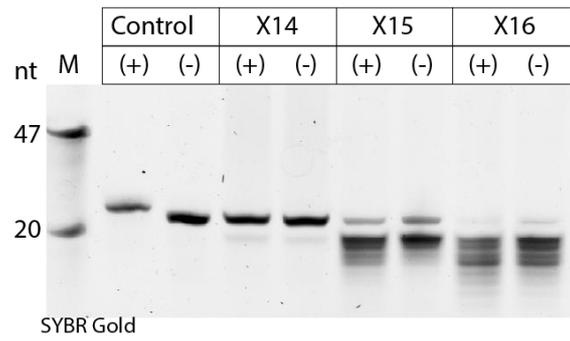
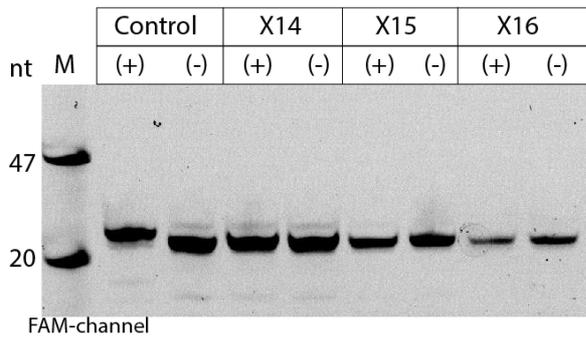
(B) X08, X09, X10-65°C



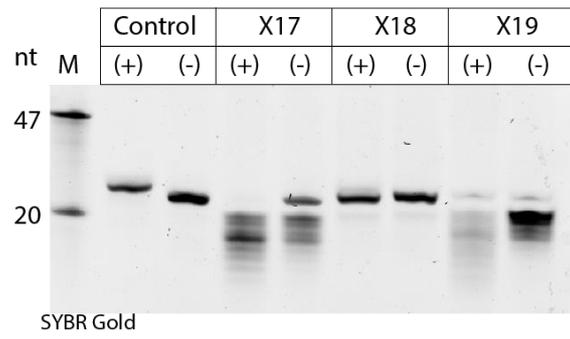
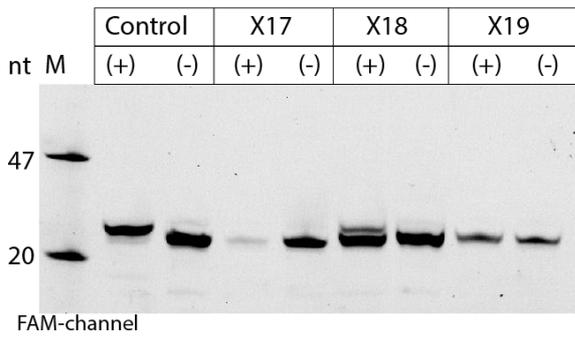
(C) X11, X12, X13



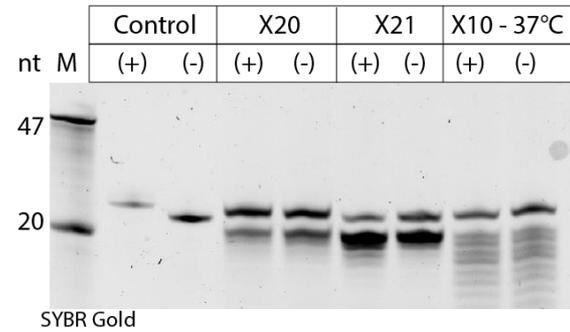
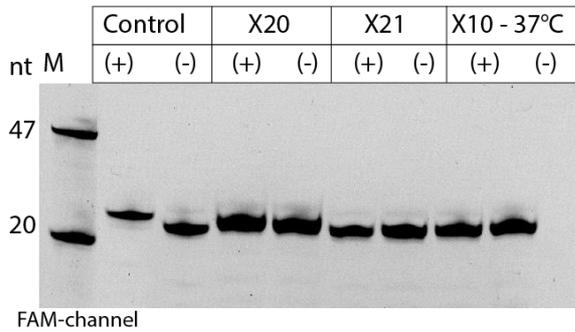
**(D) X14, X15, X16**



**(E) X17, X18, X19**



**(F) X20, X21, X10-37°C**



**(G) X22 65°C, X22 37°C**

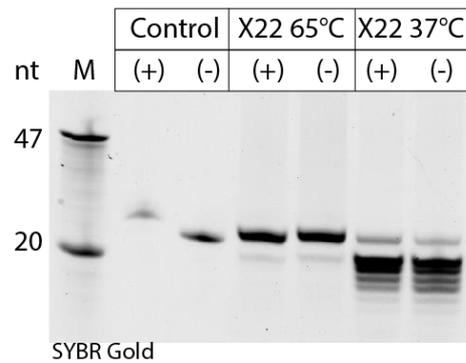
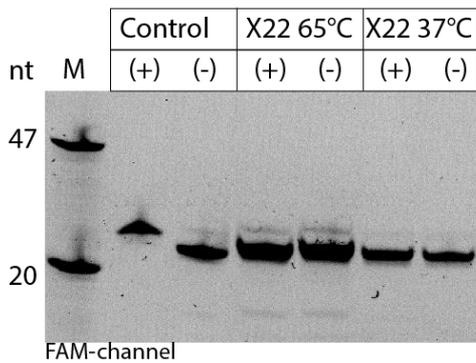


Figure C.1: **Full results of initial adenylyltransferase assays.** The 22 small-scale purified candidates (first batch) were assessed for ability to adenylylate p-ssDNA23-FAM donors in standard adenylyltransferase assays. Samples were resolved on 20% Urea-PAGE gels, detected in the FAM-channel on a X, stained with SYBR Gold, and detected again in the SYBR Gold-channel. Temperatures are included for candidates that were assessed at multiple temperatures.

## Appendix D. Phosphatase Assay Graphs and Numbers

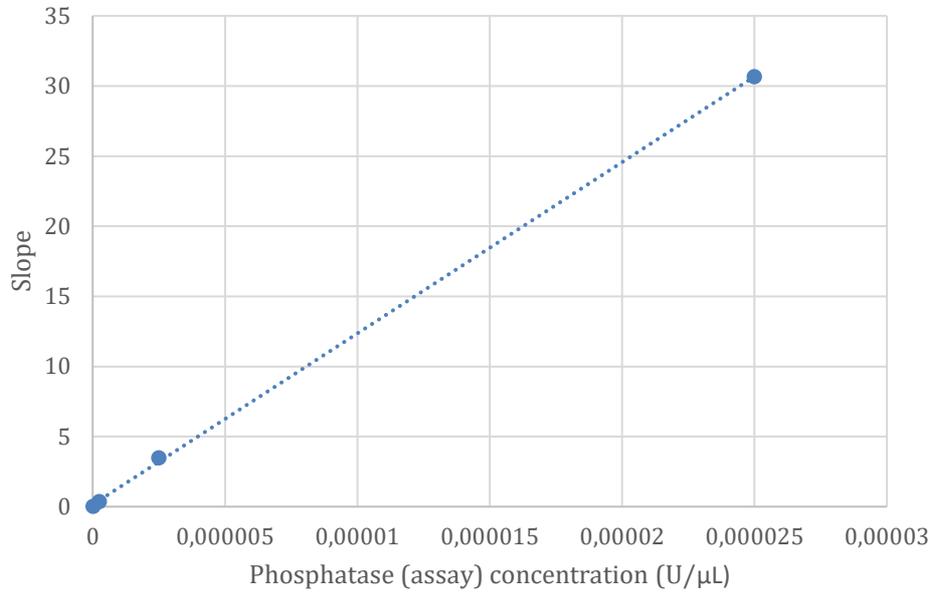


Figure D.1 **Standard curve used to calculate phosphatase activity in samples of X03 and X12.** Slopes derived from kinetic measurements were plotted against known assay phosphatase activity. The linear regression line was fitted using Excel. Intercept not forced through origin.

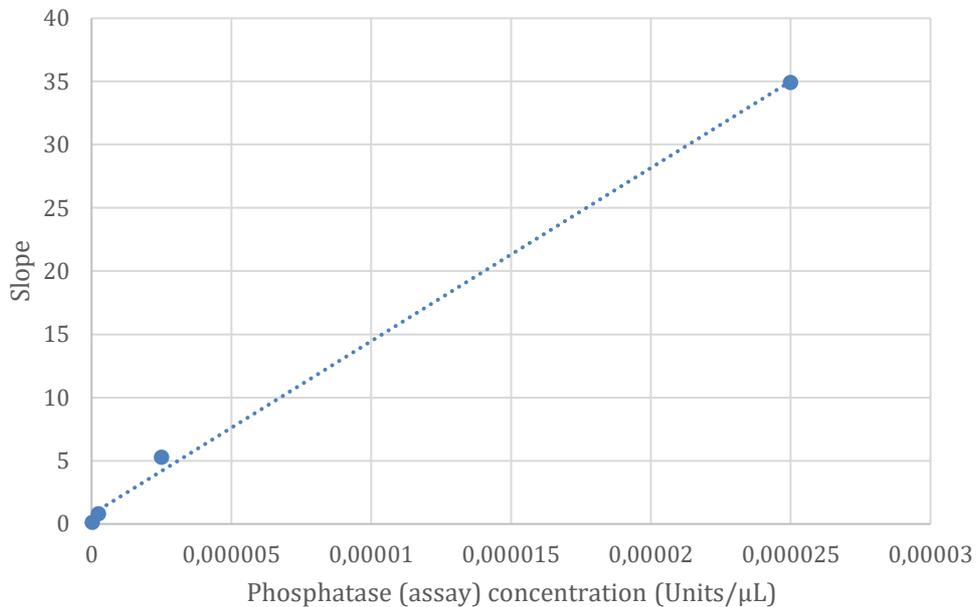


Figure D.2: **Standard curve used to calculate phosphatase activity in samples of X06, X07, and X19.** Slopes derived from kinetic measurements were plotted against known assay phosphatase activity. The linear regression line was fitted using Excel. Intercept not forced through origin.

Table D.1 :Overview of linear regressions made from standard curves of slope against phosphatase activity.

Description	Formula	R <sup>2</sup>
Standard curve X06, X07, X19 not forced	$Y = 0.759 + 1.37 * 10^6x$	0.998
Standard curve X06, X07, X19 forced	$Y = 1.40 * 10^6x$	0.996
Standard curve used for X03 and X12 (and MthRnl) (not forced)	$Y = 0.149 + 10^6x$	0.999
Standard curve X03 and X12 (and MthRnl) forced	$Y = 10^6x$	0.999

Table D.2: Overview of average phosphatase concentrations ( $\mu\text{U}/\mu\text{L}$ ) in samples taken at different stages of the purification protocol. Concentrations are calculated from standard curves not forced through origo. Sample phosphatase concentrations measured in two replicates of each low (50 pmol) and high (100 pmol) concentrations of candidate glycerol stock were used to calculate averages and coefficients of variation. Negative average values can be treated as containing undetectable amounts of phosphates activity.

Candidate	Sample	Average ( $\mu\text{U}/\mu\text{L}$ )	Coefficient of variation (%)
<b>X03</b>	Lysate	50.41	13.75
	IMAC	19.34	221.28
	IEX	-2.24	-25.85
<b>X06</b>	IMAC	-33.52	-34.15
	IEX	-26.01	-31.17

<b>X07</b>	Lysate	43.06	46.54
	IEX	-31.71	-36.20
<b>X12</b>	Lysate	133.21	18.33
	IMAC	-4.53	-27.09
	IEX	-4.93	-36.41
<b>X19</b>	IMAC	-43.81	-32.46
	IEX	-58.95	-32.13

## Appendix E. Thermal Shift Assay Graphs

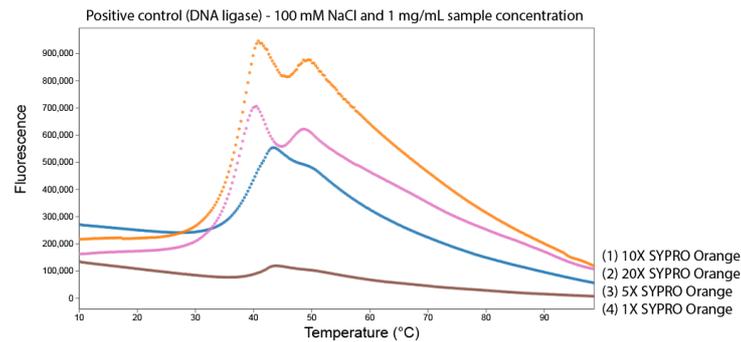


Figure E.1: **Melting curves of positive control DNA ligase obtained with TSA.** Plotting fluorescence against temperature (°C) for TSA with the positive control (1 mg/mL) at 1X, 5X, 10X, and 20X SYPRO Orange.

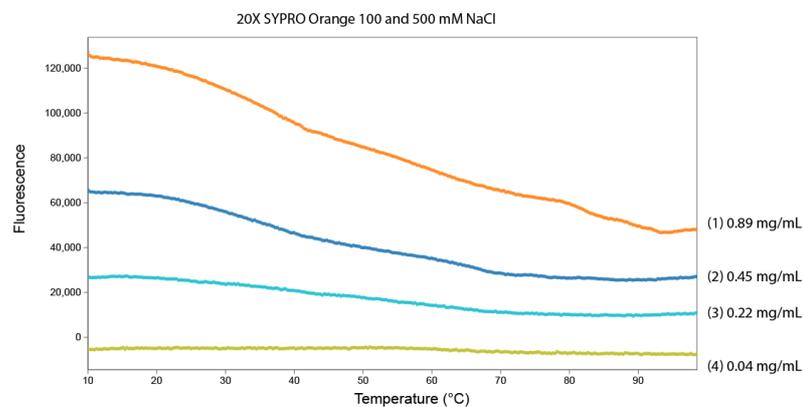


Figure E.2: **Initial fluorescence increases concomitantly with X07 concentration.** Melting curves of X07 at different concentrations with 20X SYPRO Orange obtained with TSA.

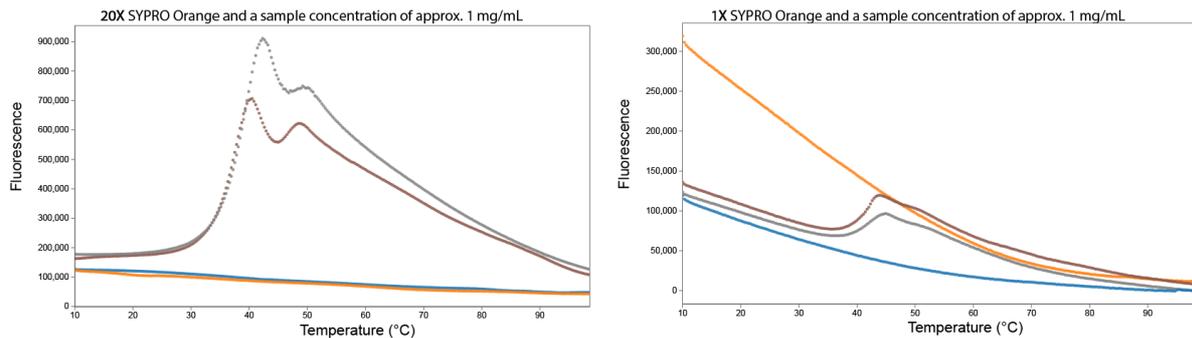


Figure E.3: **Initial fluorescence in samples of X07 and DNA ligase (PC) is similar.** Grey line: PC in 500 mM NaCl; Brown line: PC in 100 mM NaCl; Blue line: X07 in 100 mM NaCl; Orange line: X07 in 500 mM NaCl.

# Appendix F. Identity Matrix of Rnl 3 Candidates

	X05	X14	X03	X02	X22	X06	X09	X19	X21	X17	X01	X12	X13	X08	X07	X11	X15	X20	X10	X16	X04	X18	T4 RNA Ligase 1	T4 RNA Ligase 2	
X05	46.4	38.7	39.3	38.4	35.2	35.2	32.4	32.7	34.6	33.3	33.4	26.4	28.2	28	22.1	24.1	21.5	21.4	21.5	21.4	21.4	21.4	21.4	21.4	9.6
X14	46.4	40.4	39.1	37.6	36.5	36.5	35.2	35.2	35.9	31.8	31.6	27.2	29.6	24.7	23.9	24.6	24.7	23.1	24	23.3	23.2	27.1	26.4	9.6	
X03	38.7	40.4	38.6	37.6	38.5	38.5	35.2	31.1	32.3	30.5	29.1	31	28.8	29.8	23.7	24.7	23.1	23.2	24.2	19.5	23.4	30.1	27.1	8.6	
X02	39.3	39.1	37.6	53.6	39.1	34.8	34.8	34.6	38.6	37.1	35.4	29.3	27.2	28	25.6	24.3	26	23.7	24.8	23.9	25.8	27.8	27.8	7.2	
X22	38.4	37.8	38.6	53.6	36.8	36.8	37.5	35.6	36	33.4	31.7	27.7	27.5	26.8	25.4	25.1	22.2	21.6	24.8	20.6	23.7	25.5	25.5	10.7	
X06	35.2	36.5	38.5	39.1	36.8	48.2	48.2	48.2	35	30.1	30.5	27.3	29.2	29.5	24.3	24.3	22.2	19.8	20.7	19	23.9	25.2	25.2	10.9	
X09	35.2	32.4	35.2	34.8	37.5	48.2	48.2	29.7	29.9	25.8	26.1	26.7	28.6	26.2	23.9	21.9	24.4	19.7	24.6	18.7	24.7	29.2	29.2	9.8	
X19	32.7	35.2	31.1	34.6	35.6	35.6	33.2	29.7	57.9	52	46.1	29.3	26	25.5	24.1	24.4	26.3	23.3	23.1	24.9	23.9	26.2	27.5	8.3	
X21	34.6	35.9	32.3	38.6	36	35	29.9	57.9	57.9	48.9	48.3	30.6	27.3	28.4	25.3	26.7	28.8	27.1	25.6	25.7	26.2	27.5	27.5	8.3	
X17	33.3	31.8	30.5	37.1	33.4	30.1	25.8	52	48.9	46.8	46.8	25.4	24.2	22.4	21.7	25	22.4	25.1	26	24.7	24.9	25.2	25.2	10.4	
X01	33.4	31.6	29.1	35.4	31.7	30.5	26.1	46.1	46.3	46.8	46.8	26.1	22.4	25.7	21.7	23.5	23.3	27.2	24.9	23.5	23.3	27.2	27.2	8.8	
X12	26.4	27.2	31	29.3	27.7	27.3	26.7	29.3	30.6	25.4	26.1	29.6	39.6	36.9	23.9	21.5	21.7	22.3	24.9	21.3	18.6	21.4	27.3	27.3	8.9
X13	28.2	29.6	28.8	27.2	27.5	29.2	28.6	26	27.3	24.2	22.4	39.6	27.3	35.4	26.6	20.4	20.6	20.7	22.5	19.8	22.2	28.9	28.9	9	
X08	28	24.7	29.8	26	26.8	29.5	28.2	25.5	28.4	22.4	25.7	36.9	35.4	25.6	25.6	21.4	21.4	22.4	22.4	20.1	21.4	30.2	30.2	11.8	
X07	22.1	23.9	23.7	25.6	25.4	24.3	23.9	24.1	25.3	21.7	23.5	23.9	26.6	23.6	25.6	50.9	50.9	43.4	41.2	40.7	36.5	31.8	39.5	8.7	
X11	24.1	24.6	24.7	24.3	25.1	19.6	21.9	24.4	26.7	25	23.3	21.5	20.4	21.4	43.4	40.2	40.2	43.3	40.3	34.5	28.9	32.8	32.8	8.4	
X15	21.5	24	23.1	28	22.2	21.4	19.7	28.3	28.8	26	27.2	21.7	20.6	21.4	43.4	40.2	40.2	47	47	35.2	27.5	29.9	29.9	7.6	
X20	21.5	23.3	23.2	23.7	21.6	19.8	19	23.3	27.1	25.1	24.9	22.3	20.7	22.4	41.2	43.3	43.3	37.7	37.7	32.5	31.9	32.1	32.1	8.7	
X10	24.6	23.3	24.2	24.8	24.8	20.7	24.6	23.1	25.6	21.5	23.5	21.3	22.5	22.4	40.7	40.3	35.4	37.7	30.2	30.2	29.1	30.5	30.5	8.7	
X16	21.4	21.9	19.5	23.9	20.6	19	18.7	24.9	25.7	24.7	23.3	18.6	19.8	20.1	36.5	34.5	35.2	32.5	30.2	29.6	29.6	28	28	8.7	
X04	25.3	23.2	23.4	23.7	23.7	23.9	24.7	23.9	26.2	24.9	25.1	21.4	22.2	21.4	31.8	28.9	27.5	31.9	29.1	29.6	34.4	34.4	34.4	7.7	
X18	26.4	27.1	30.1	27.8	25.5	25.2	25.2	28.2	27.5	25.2	27.2	27.3	26.9	30.2	33.5	32.8	29.9	32.1	30.5	28	34.4	34.4	34.4	10.5	
T4 RNA Ligase 1	9.6	8.6	11.6	9.9	10.7	10.9	11.8	9.9	8.3	10.4	8.8	8.9	9	11.8	8.7	8.4	7.6	8.7	8.7	9.5	12.1	10.5	10.5	6.9	
T4 RNA Ligase 2	9.8	7.7	7.2	8.2	7.5	7.4	9.7	9.8	8.5	7.7	7.7	5.9	8.5	6.7	9.9	7.4	7.6	8.9	8.7	7.7	7.7	8.5	8.5	6.9	

Figure F.1: Identity matrix created from the 22 candidates of this thesis and the two bacteriophage T4 RNA ligases using the BLOSUM algorithm. Similarity is in percentage.

