AFRL-AFOSR-VA-TR-2023-0191

Search for a psychrotolerant RNA polymerase promoter-sequence for its use in riboswitch biosensors

Blamey, Jenny
FUNDACION CIENTIFICA Y CULTURAL BIOCIENCIA
JOSE DOMINGO CANAS 2280
SANTIAGO, ,
CHL

11/20/2022
Final Technical Report

Air Force Research Laboratory
Air Force Office of Scientific Research
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
20221120	Final	START DATE 20200930	END DATE 20210929

4. TITLE AND SUBTITLE
Search for a psychrotolerant RNA polymerase promoter-sequence for its use in riboswitch biosensors

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
	FA9550-20-1-0390	
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Jenny Blamey

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
FUNDACION CIENTIFICA Y CULTURAL BIOCIENCIA JOSE DOMINGO CANAS 2280 SANTIAGO CHL	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Air Force Office of Scientific Research 875 N. Randolph St. Room 3112 Arlington, VA 22203	AFRL/AFOSR IOS	AFRL-AFOSR-VA-TR-2023-0191

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The submission of this proposal was for the continuation of the AFOSR grant FA9550-16-1-0356: "Bioengineering Extremophile Enzymes to Develop a Rapid Diagnostic Test for Warfighters". This project was running in parallel to an existing grant with a unique potential application, directly feeding an initiative by the 711 HPW to create rapid diagnostic tests for Airmen. It possesses promising transition potential in a tech area aligned with the AFRL International Strategy, located in Chile, an emerging country, while leveraging co-funding by AFRL/RH.
2
Extremophiles are a source of new biocompounds and enzymatic activities, poorly described due to the complexity of obtaining and manipulating them. The majority of extremophiles have a low cell yield in cultures when they are cultivated. This can also be extended to viruses, since they need to be in the presence of their cellular hosts. These forms of extreme life are even more interesting when isolated from pristine extreme environments such as Union Glacier, Antarctica, which is located approximately 900 km from the South Pole, where very few researchers, mostly glaciologists have been. Therefore, the chances of obtaining novel strains containing unidentified biological compounds and metabolites that have not been described in the literature or found in available databases, from environmental samples collected from this extreme site are even higher.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	c. THIS PAGE		
U	U	U	SAR	29

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
STACY MANNI	0000 0000

Grant Award NO. FA9550-20-1-0390
Report: Final Technical Report
Period: 30/09/2020 to 29/09/2021
December 28th 2021

Final Report

Title:
Search for a psychrotolerant RNA polymerase Promoter-Sequence for its use in riboswitch biosensors.

Research Interest Category:
Chemistry and Life Sciences, Biophysical Mechanisms.

Corresponding Program Manager:
Dr. Stacy Manni, AFOSR/IOS, stacy.manni@us.af.mil

Technical Director (SOARD):
Dr. Stacy Manni

Main Objective
This study aims to search and find the specific promoter-sequence for a novel viral RNA polymerase enzyme(s) isolated from novel Antarctic viruses, for its use in biosensor systems.

Specific objectives. This project includes the following specific objectives:

a) To identify the Antarctic viruses found, using electron microscopy and additional microbiology techniques in order to isolate and identify them.
b) To obtain the gene sequence encoding for a novel RNA polymerase using metavirome analysis as well as bioinformatics tools.
c) To identify its specific promoter using a combination of molecular biology and biochemical techniques.

The main challenge of this project was to obtain the promoter sequence of the viral Antarctic RNA polymerase, so it could be useful as part of an AFRL designed biosensor: a riboswitch controlled cell-free system optimized to control gene expression, as a response to the presence of Air Force-related targets.

Project context.
The submission of this proposal was for the continuation of the AFOSR grant FA9550-16-1-0356: "Bioengineering Extremophile Enzymes to Develop a Rapid Diagnostic Test for Warfighters". This project was running in parallel to an existing grant with a unique potential application, directly feeding an initiative by the 711 HPW to create rapid diagnostic tests for Airmen. It possesses promising transition potential in a tech area aligned with the AFRL International Strategy, located in Chile, an emerging country, while leveraging co-funding by AFRL/RH.

Extremophiles are a source of new biocompounds and enzymatic activities, poorly described due to the complexity of obtaining and manipulating them. The majority of extremophiles have a low cell yield in cultures when they are cultivated. This can also be extended to viruses, since they need to be in the presence of their cellular hosts. These forms of extreme life are even more interesting when isolated from pristine extreme environments such as Union Glacier, Antarctica, which is located approximately 900 km from the South Pole, where very few researchers, mostly glaciologists have been. Therefore, the chances of obtaining novel strains containing unidentified biological compounds and metabolites that have not been described in the literature or found in available databases, from environmental samples collected from this extreme site are even higher.

As presented in the final report of the AFOSR grant FA9550-16-1-0356: "Bioengineering Extremophile Enzymes to Develop a Rapid Diagnostic Test for Warfighters". Several consortia containing psychrotolerant bacteria and viruses were obtained from environmental samples from Union Glacier, Antarctica. From one of them, named GGM3, it was possible to isolate a novel RNA polymerase from a psychrophilic origin in its native form. In this final report of the AFOSR grant FA9550-20-1-0390, the experimental work performed to identify the enzyme, develop its recombinant form, obtain its putative promoter sequence, along with its purification and characterization is described below.

- **Selection of Antarctic viruses for metagenomic sequencing**

The consortium GGM3 was found to be more active than other Antarctic consortia and was selected to further studies. From this, fifteen bacterial strains were isolated and several viruses were observed by electron microscopy. The viral particles present in each supernatant were not specific for a particular strain, as they were able to infect and produce bacterial lysis in all the different bacterial strains.

Experiments using the low temperature cell line CHSE-214, derived from *Oncorhynchus tshawytscha* embryo were performed in order to isolate and enrich the Antarctic viral cultures. This cell line was selected because it grows at low temperature (8 - 15°C), is susceptible to a wide range of viruses (Nakano et al., 1993) and is able to replicate high titers.

The filtered supernatants from the isolated strains were used for infection assays of the CHSE-214 cells. The infected cell cultures were incubated at 15º C for 48 h and then the cells monolayer infected were visualized by optic microscopy (Figure 1).

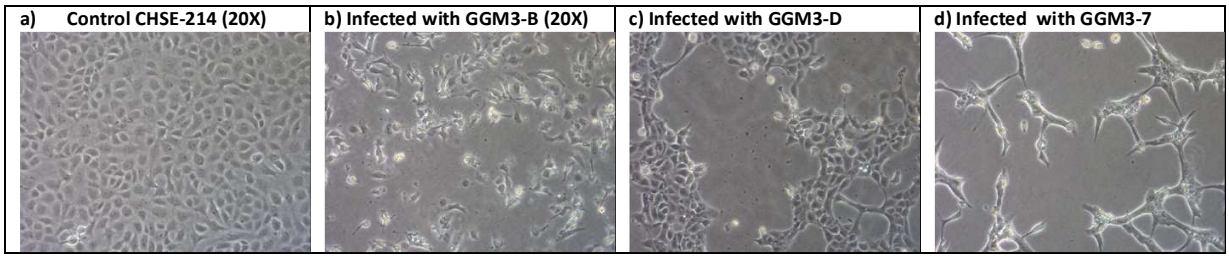


Figure 1. Optic microscopy images (20X) of CHSE-214 cells monolayer after 48h incubation at 15°C. *(a) Control (no infection), (b) infected with the filtered supernatant from the GGM3 isolated strain GGM3-B, (c) infected with the filtered supernatant from the GGM3 isolated strain GGM3-D, (d) infected with the filtered supernatant from the GGM3 isolated strain GGM3-7.*

As can be seen in Figure 1, the results showed a clear decrease in the number of cells when infected with the supernatant of GGM3-B, GGM3-D and GGM3-7 isolates, due to cell lysis induced by the viral particles.

The post infection supernatant containing the CHSE-214 cell lysates and the replicated viral particles of GGM3-B, GGM3-D and GGM3-7 were observed by means of transmission electron microscopy (Figure 2). For better comparison prior images from the supernatant of the preliminary consortium obtained from the environmental samples collected from Union Glacier, Antarctica are also shown.

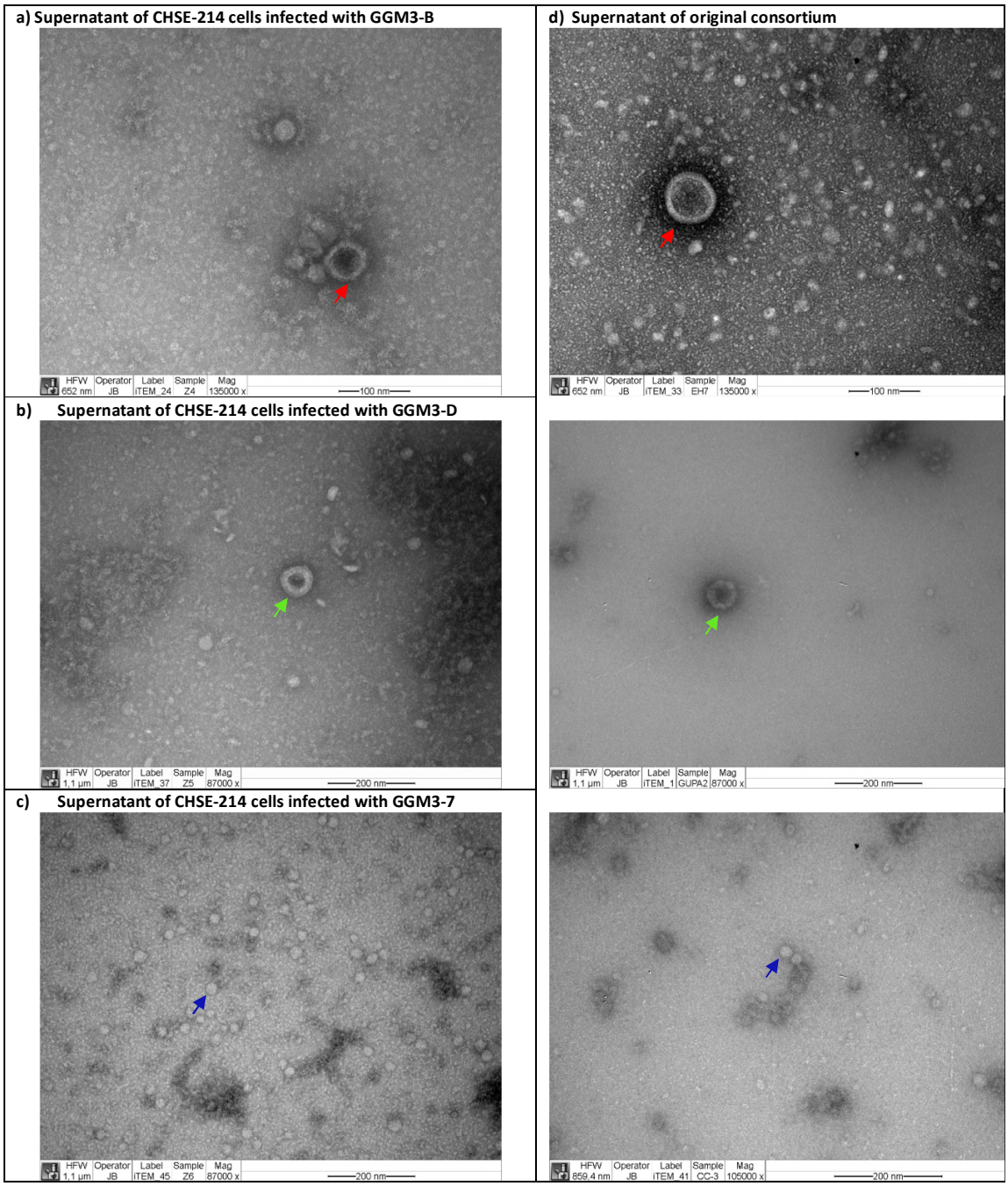


Figure 2. Transmission electron microscopy images confirming the presence and diversity of viral particles (highlighted in red arrows) found on environmental samples from Union Glacier, Antarctica. *a) Supernatant of CHSE-214 cells infected with GGM3-B, b) Supernatant of CHSE-214 cells infected with GGM3-D, c) Supernatant of CHSE-214 cells infected with GGM3-7, d) supernatant of original consortium. Highlighted in colored arrows are the viruses that are similar in the supernatants from the cell infected and the original consortium.*

These results indicate that viral particles present in the CHSE-214 cell lysates get enriched in one or just few types of virus, in contrast with the previously observed electron microscope images from the original consortium (Figure 3) where several different types of virus could be found.

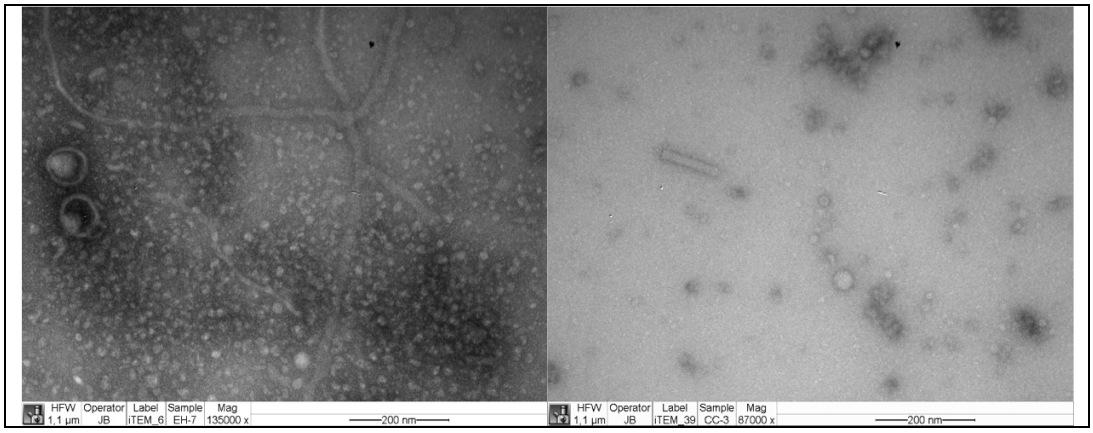


Figure 3. Transmission electron microscopy images of the original consortium.

- **<u>Bioinformatic analysis of four metagenomes in search of viral RNA polymerases (EC 2.7.7.6)</u>**

Due to GGM3-7 filtered supernatant showed to be more infective (Figure 1d) it was selected for further analysis. Total viral DNA was extracted from four different samples, detailed in table 1, and sent to the Duke Center for Genomic and Computational Biology (GCB) for metagenomic sequencing using MiSeq v2 250 bp PE, DNA-Seq (500 bp insert).

Table 1. Total Viral DNA samples for metagenomic analysis.

GGM3	Consortium generated from environmental sample
GGM3-isolate7	Strain isolated from GGM3-consortium
Z6o	Viral enrichment from CHSE-214 cell infection with GGM3-isolate 7 generated in October 2019.
Z6n	Viral enrichment from CHSE-214 cell infection with GGM3-isolate 7 generated in December 2019.

The sequencing results from each metagenome were subjected to standard pre-processing methods, taxonomic profiling, mapping against viral databases, assembling and mining in search for DNA-directed RNA polymerases (EC 2.7.7.6) of viral origin, particularly from Bacteriophages (bacterial viruses). The search strategy was focused particularly on phage or phage-type RNA polymerases, because of their advantageous properties for application in cell-free systems: high rate of RNA synthesis, high specificity towards their promoters and small size (monomeric ~100 kDa).

The methods for sequencing, preprocessing, taxonomic profiling, mapping, assembling and contigs annotation used during the bioinformatics analysis of the four metagenomes are described in the Progress report dated 20200620.

<u>Metagenomic mining of viral DNA-directed RNA polymerases</u>

The heatmap in Figure 4 shows the total counts of coding sequence (CDS) annotations containing the term “polymerase” or “primase” in each metagenome, including: DNA polymerases, RNA polymerases, RNA polymerases sigma factors and primases/helicases.

Upon inspection of the heatmap it is clear that GGM3 assembly gathers the highest amount of sequences annotated as polymerases, of which 29 correspond to DNA-directed RNA polymerases (2.7.7.6). Interestingly, only one of these enzymes was specifically annotated as **"Phage DNA-directed RNA polymerase"**, suggesting that it could be a single subunit phage-type polymerase, while the others were annotated as "DNA-directed RNA polymerase beta subunit" and "DNA-directed RNA polymerase beta' subunit", indicating that they are part of multi-subunit polymerases.

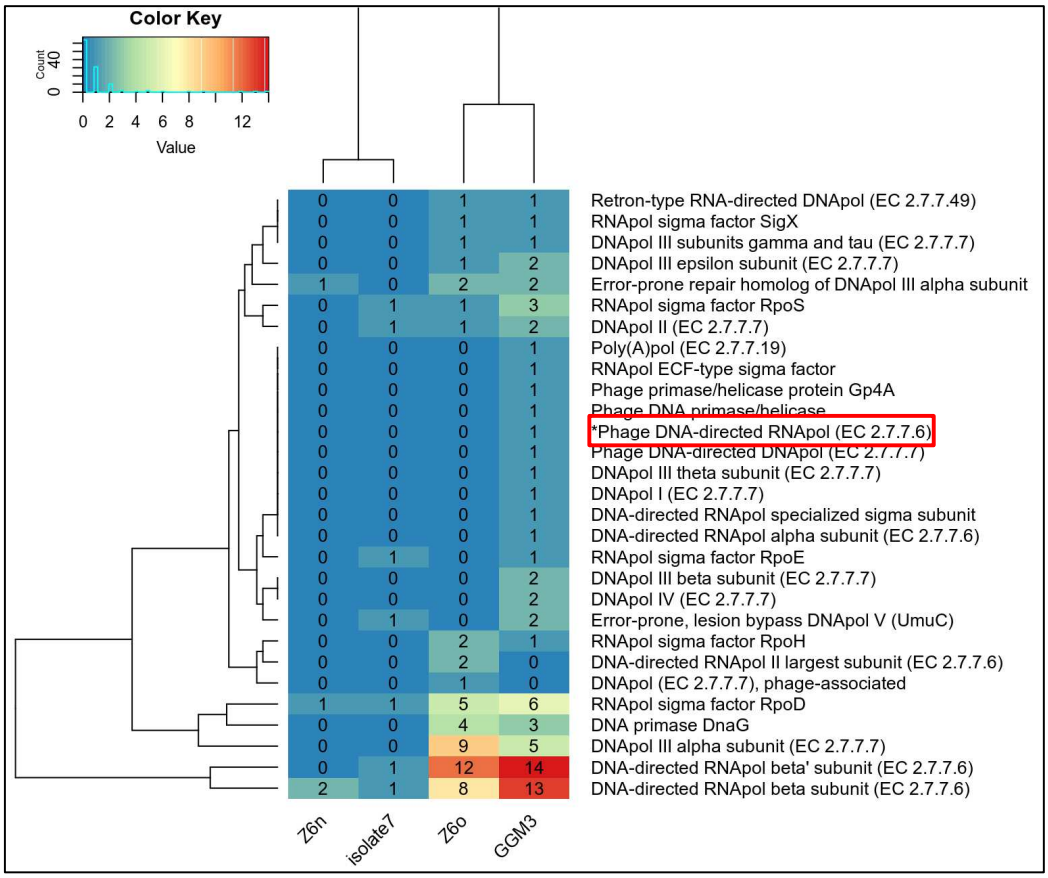


Figure 4. Heatmap showing the count of annotations containing the term "polymerase" per metagenome. *Highlighted in red is the Phage RNA polymerase.*

All the coding sequences annotated as "DNA-directed RNA polymerases" were subjected to BLASTp searches against the nr protein database for validating annotations and obtaining taxonomic information. Table 2 shows the top scored hits and the organism source for every RNA polymerase found in PATRIC annotations (Davis et al., 2019). Surprisingly, none of the query polymerase sequences matched with subjects annotated as viral RNA polymerases, but bacterial polymerases. Nevertheless, considering that all of the polymerases predicted come from assemblies of "viral reads" only (obtained by mapping), the BLAST results obtained could suggest that many of these sequences are actually prophage sequences that have been annotated as a part of host bacterial genomes, thus having bacterial taxonomy associated. This is why almost all of the best hits were obtained against entries from Bacteria.

Table 2. Blastp top scoring hits for the annotated RNA polymerases.

Sample	protein_ID	Blastp.best.match	Organism	Accession
GGM3	10239.10039.peg.2900	DNA-directed RNA polymerase subunit alpha	*Pseudomonas aeruginosa*	WP_061200583.1
	10239.10039.peg.1516	MULTISPECIES: DNA-directed RNA polymerase subunit beta	*Pseudomonas*	WP_043861935.1
	10239.10039.peg.1722	DNA-directed RNA polymerase subunit beta	*Stenotrophomonas maltophilia*	VUM75228.1
	10239.10039.peg.2236	DNA-directed RNA polymerase subunit beta	*Stenotrophomonas maltophilia*	VUJ43451.1
	10239.10039.peg.2288	MULTISPECIES: DNA-directed RNA polymerase subunit beta	*Pseudomonas*	WP_029241916.1
	10239.10039.peg.2295	MULTISPECIES: DNA-directed RNA polymerase subunit beta	*unclassified Enterococcus*	WP_070546922.1
	10239.10039.peg.2438	DNA-directed RNA polymerase subunit beta	*Enterococcus faecium EnGen0263*	EOH54421.1
	10239.10039.peg.3221	DNA-directed RNA polymerase subunit beta	*Enterococcus faecium*	WP_025479369.1
	10239.10039.peg.3991	DNA-directed RNA polymerase subunit beta	*Stenotrophomonas maltophilia*	WP_134951788.1
	10239.10039.peg.4038	MULTISPECIES: DNA-directed RNA polymerase subunit beta	*Pseudomonas*	WP_043861935.1
	10239.10039.peg.4438	DNA-directed RNA polymerase subunit beta	*Pseudomonas sp. PGPPP2*	OYT76074.1
	10239.10039.peg.4528	DNA-directed RNA polymerase subunit beta	*Microbacterium oxydans*	KTR75859.1
	10239.10039.peg.5358	DNA-directed RNA polymerase subunit beta	*Salmonella enterica subsp. enterica*	SUH11798.1
	10239.10039.peg.5553	DNA-directed RNA polymerase subunit beta	*Enterococcus sp. 5B7_DIV0075*	OTP22320.1
	10239.10039.peg.2206	DNA-directed RNA polymerase subunit beta'	*Pseudomonas putida*	WP_043861934.1
	10239.10039.peg.2235	DNA-directed RNA polymerase subunit beta'	*Stenotrophomonas sp. 364*	WP_159495851.1
	10239.10039.peg.2268	DNA-directed RNA polymerase subunit beta'	*Pseudomonas fluorescens*	VVN63204.1
	10239.10039.peg.2289	DNA-directed RNA polymerase subunit beta'	*Pseudomonas sp. RIT623*	WP_134693397.1
	10239.10039.peg.2329	DNA-directed RNA polymerase subunit beta'	*Stenotrophomonas maltophilia*	WP_135272431.1
	10239.10039.peg.2400	DNA-directed RNA polymerase subunit beta'	*Stenotrophomonas maltophilia*	WP_046988087.1
	10239.10039.peg.2460	DNA-directed RNA polymerase subunit beta'	*Enterococcus faecium*	WP_087626304.1
	10239.10039.peg.2479	DNA-directed RNA polymerase subunit beta'	*Pseudomonas putida*	WP_043861934.1
	10239.10039.peg.3430	DNA-directed RNA polymerase subunit beta	*Enterococcus hirae*	OJG50999.1
	10239.10039.peg.3696	DNA-directed RNA polymerase subunit beta'	*Xanthomonas euvesicatoria*	NEK91876.1
	10239.10039.peg.4373	DNA-directed RNA polymerase subunit beta'	*Escherichia coli*	WP_139499188.1
	10239.10039.peg.5728	DNA-directed RNA polymerase subunit beta'	*Enterobacter cloacae subsp. cloacae*	KPU01620.1
	10239.10039.peg.5924	TPA: DNA-directed RNA polymerase subunit beta'	*Citrobacter freundii*	HCM55921.1
	10239.10039.peg.5953	DNA-directed RNA polymerase subunit beta'	*Salsuginibacillus kocurii*	WP_018924679.1
	10239.10039.peg.2441	MULTISPECIES: DUF4880 domain-containing protein	*Pseudomonas*	WP_043863965.1
	10239.10039.peg.311	**T3/T7 RNA polymerase**	***Stenotrophomonas maltophilia***	**WP_164076363.1**
Isolate7	10239.10038.peg.211	DNA-directed RNA polymerase subunit beta'	*Hafnia alvei*	WP_063586007.1
	10239.10038.peg.377	DNA-directed RNA polymerase subunit beta'	*Yersinia pestis*	WP_016595455.1
Z6n	10239.10041.peg.172	RNA polymerase B subunit	*Brevundimonas bullata*	CBY93789.2
	10239.10041.peg.260	DNA-directed RNA polymerase subunit beta	*Devosia sp. 1566*	WP_127752899.1
Z6o	10239.10040.peg.990	RNA polymerase II largest subunit	*Drechmeria coniospora*	KYK54642.1
	10239.10040.peg.1650	RPB1	*Fusarium solani f. xanthoxyli*	AFV36993.1
	10239.10040.peg.367	DNA-directed RNA polymerase subunit beta'	*Pseudomonas sp. RIT623*	WP_134693397.1
	10239.10040.peg.633	DNA-directed RNA polymerase subunit beta'	*Pseudomonas putida*	WP_043861934.1
	10239.10040.peg.698	DNA-directed RNA polymerase subunit beta'	*Pseudomonas fluorescens*	VVN63204.1
	10239.10040.peg.795	DNA-directed RNA polymerase subunit beta'	*Rhizobiales bacterium*	WP_113484576.1
	10239.10040.peg.829	DNA-directed RNA polymerase subunit beta'	*Brevundimonas sp.*	PZU71558.1
	10239.10040.peg.967	DNA-directed RNA polymerase subunit beta'	*Pseudomonas sp.*	MPT01877.1
	10239.10040.peg.1444	DNA-directed RNA polymerase subunit beta'	*Devosia limi*	WP_046136730.1
	10239.10040.peg.1452	DNA-directed RNA polymerase subunit beta'	*Sphingomonas sp. AAP5*	WP_133190066.1
	10239.10040.peg.1623	DNA-directed RNA polymerase subunit beta'	*Iodobacter fluviatilis*	WP_115230160.1
	10239.10040.peg.1741	DNA-directed RNA polymerase subunit beta'	*Pseudomonas sp. NBRC 111142*	WP_054926105.1
	10239.10040.peg.1781	DNA-directed RNA polymerase subunit beta'	*Phyllobacterium myrsinacearum*	WP_105732678.1
	10239.10040.peg.2186	DNA-directed RNA polymerase subunit beta'	*Bradyrhizobium sp. AT1*	WP_063195100.1
	10239.10040.peg.222	DNA-directed RNA polymerase subunit beta	*Phyllobacterium myrsinacearum*	WP_105732677.1
	10239.10040.peg.368	MULTISPECIES: DNA-directed RNA polymerase subunit beta	*Pseudomonas*	WP_029241916.1
	10239.10040.peg.445	MULTISPECIES: DNA-directed RNA polymerase subunit beta	*Pseudomonas*	WP_043861935.1
	10239.10040.peg.809	DNA-directed RNA polymerase subunit beta	*Sphingomonas glacialis*	WP_140846858.1
	10239.10040.peg.1153	DNA-directed RNA polymerase subunit beta	*Phyllobacterium myrsinacearum*	WP_105732677.1
	10239.10040.peg.1491	DNA-directed RNA polymerase subunit beta	*Bradyrhizobiaceae bacterium PARB1*	OYU86059.1
	10239.10040.peg.1929	DNA-directed RNA polymerase subunit beta	*Sphingomonas echinoides*	WP_010407631.1
	10239.10040.peg.2008	RNA polymerase beta subunit	*Bartonella sp. FN17-1*	BAI47994.1

Highlighted in red in Table 2 is the only polymerase that matched a phage-type RNA polymerase was 10239.10039.peg.311 found in GGM3 assembly, as expected by the previous annotation result. BLASTp highest score match for this protein was a **"T3/T7 RNA polymerase" from *Stenotrophomonas maltophilia*** (NCBI accession WP_164076363.1).

The same search was conducted filtering by Viruses (taxid:10239): the highest identity was obtained with a DNA-dependent RNA polymerase from *Rhizobium phage RHEph01* (44%). Since this type of polymerase is closer to that required for the project's objectives, subsequent analyzes focused on it.

Phage-type RNA polymerase

The contig containing the phage-type RNA polymerase sequence was searched and identified in the GGM3 assembly and then annotated independently with PATRIC. The contig, named "NODE_17_length_8940_cov_6.186944", is 8490 bp long and has a sequence coverage of 6x (which is generally considered low) and encodes 15 CDS, of which only 3 has functional annotation (all of them being polymerases, as seen in Table 3. The remaining sequences were annotated as hypothetical proteins.

Table 3. Functional annotation of CDS from contig: NODE_17_length_8940_cov_6.186944. *The identifier of the sequences changed after independent annotation of the contig*

fig_id	Annotation
fig.10239.10042.peg.1	Phage DNA-directed DNA polymerase (EC 2.7.7.7)
fig.0239.10042.peg.5	Phage primase/helicase protein Gp4A
fig.10239.10042.peg.15	Phage DNA-directed RNA polymerase (EC 2.7.7.6)

Figure 5 shows the genomic context of the RNA polymerase "fig|10239.10042.peg.15".

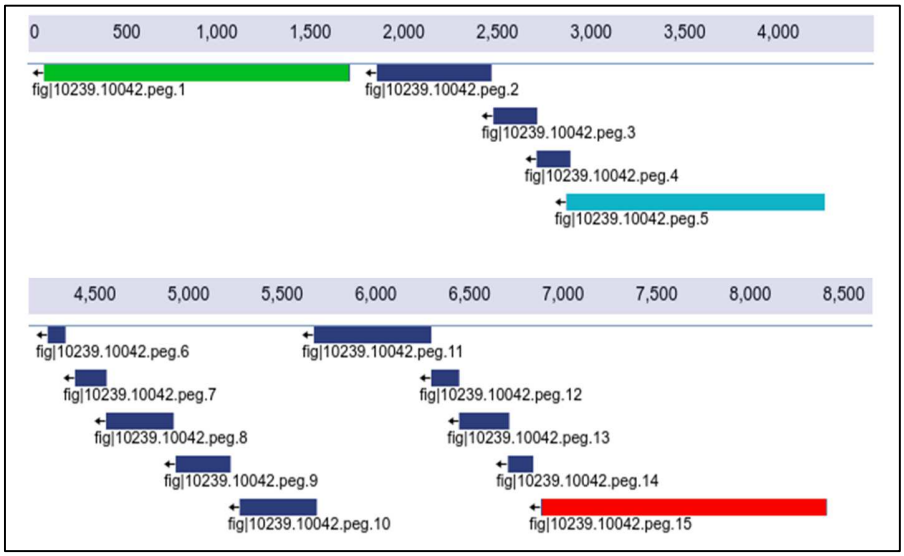


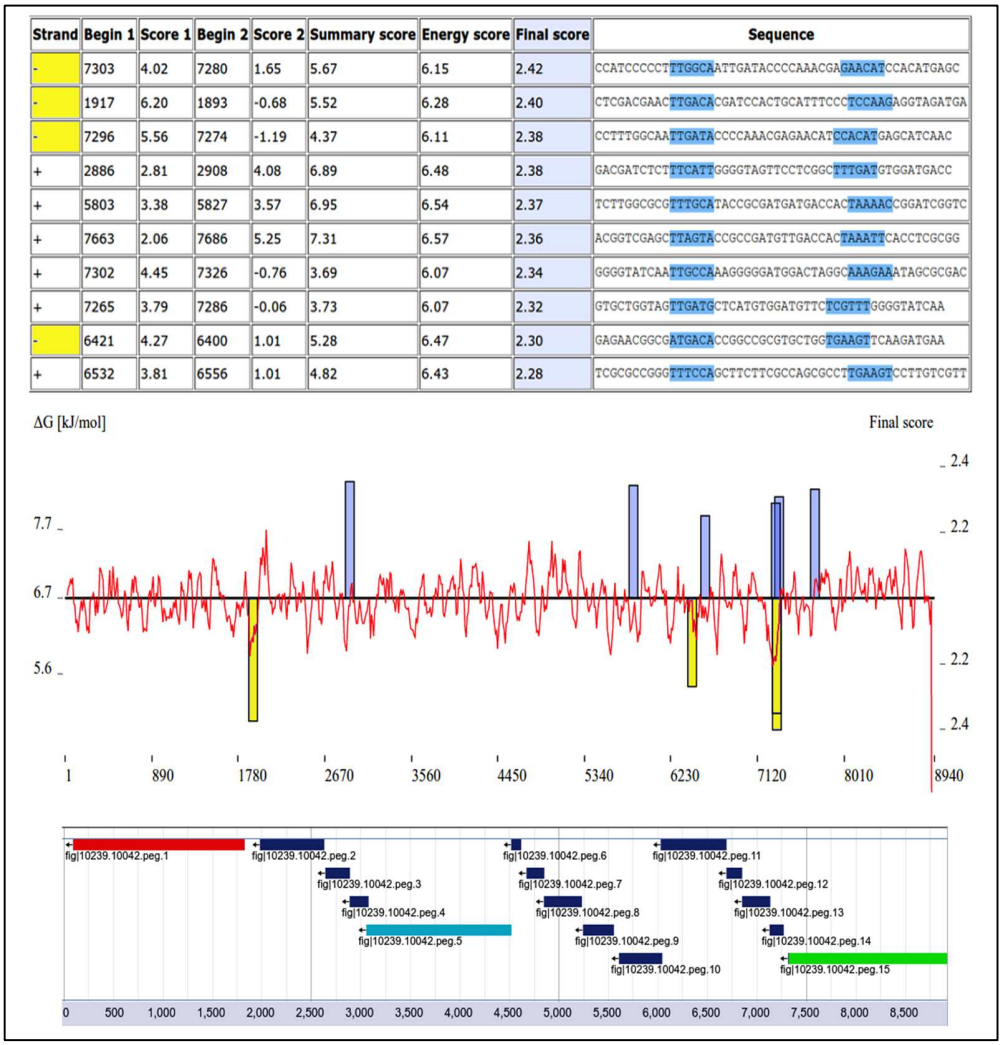
Figure 5. Genomic context of the Phage DNA-directed RNA polymerase found in the GGM3 assembly. *Phage RNA polymerase is shown in red, DNA polymerase in green and the primase/helicase in cyan.*

As can be seen in Figure 5, the sequence is encoded in the negative strand and seems to be part of a functional gene cluster that includes the DNA polymerase "fig.10239.10042.peg.1" and the primase "fig.0239.10042.peg.5".

RNA Polymerase Promoter Search

The search for promoters was performed using phiSITE's PromoterHunter tool with default scoring matrices (for -35 and -10 regions) and considering both strands. phiSITE is a database of gene regulation in bacteriophages that contains >700 experimentally confirmed and/or predicted regulatory elements (promoters, operators, terminators and attachment sites) from 32 bacteriophages belonging to Siphoviridae, Myoviridae and Podoviridae families (Klucar et al., 2010). The table in Figure 6 shows the

sequences and coordinates of the predicted promoters along with their respective scores. Additionally, the calculated distribution of Gibbs free energy (ΔG [kJ/mol]) is shown as a histogram. The free energy peaks calculated for the minus strand (yellow in the figure) matches the start position of some annotated proteins. For example, the energy peak near 1800 bp region matches the start of the DNA polymerase CDS (in red), indicating that the predicted promoter is well positioned. It is worth to mention that predicted promoters are theoretical and need further experimental validation.



Strand	Begin 1	Score 1	Begin 2	Score 2	Summary score	Energy score	Final score	Sequence
-	7303	4.02	7280	1.65	5.67	6.15	2.42	CCATCCCCCTTTGGCAATTGATACCCCAAACGAGAACATCCACATGAGC
-	1917	6.20	1893	-0.68	5.52	6.28	2.40	CTCGACGAACTTGACACGATCCACTGCATTTCCCTCCAAGAGGTAGATGA
-	7296	5.56	7274	-1.19	4.37	6.11	2.38	CCTTTGGCAATTGATACCCCAAACGAGAACATCCACATGAGCATCAAC
+	2886	2.81	2908	4.08	6.89	6.48	2.38	GACGATCTCTTTCATTGGGGTAGTTCCTCGGCTTTGATGTGGATGACC
+	5803	3.38	5827	3.57	6.95	6.54	2.37	TCTTGGCGCGTTTGCATACCGCGATGATGACCACTAAAACCGGATCGGTC
+	7663	2.06	7686	5.25	7.31	6.57	2.36	ACGGTCGAGCTTAGTACCGCCGATGTTGACCACTAAATTCACCTCGCGG
+	7302	4.45	7326	-0.76	3.69	6.07	2.34	GGGGTATCAATTGCCAAAGGGGGATGGACTAGGCAAAGAAATAGCGCGAC
+	7265	3.79	7286	-0.06	3.73	6.07	2.32	GTGCTGGTAGTTGATGCTCATGTGGATGTTCTCGTTTGGGGTATCAA
-	6421	4.27	6400	1.01	5.28	6.47	2.30	GAGAACGGCGATGACACCGGCCGCGTGCTGGTGAAGTTCAAGATGAA
+	6532	3.81	6556	1.01	4.82	6.43	2.28	TCGCGCCGGGTTTCCAGCTTCTTCGCCAGCGCCTTGAAGTCCTTGTCGTT

Figure 6. phiSITE PromoterHunter search result. *On top, predicted promoters are highlighted in blue in the Sequence column. Below, the distribution of free energy calculated is shown in red line and peaks in blue. Promoters in the minus strand are highlighted in yellow. The genomic context is also shown for comparative purposes.*

Phage RNA Polymerase model

A tentative structural model of the Phage RNA polymerase fig|10239.10042.peg.15 was built with the aim of finding evolutionary related structures matching the Phage RNA polymerase found.

A Swiss-Model Homology model is shown in Figure 7 and the Modelling report is show in Figure 8. The model was built using the crystal structure of the *Escherichia* bacteriophage T7 DNA-dependent RNA polymerase (T7 RNAP, pdb_id: 3E3J), which share 45% sequence identity. The oligo state of the model indicates a monomer.

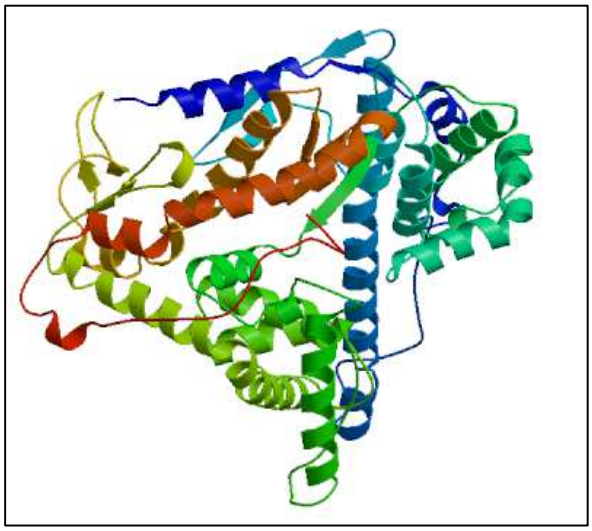


Figure 7. Swiss-Model homology model of the Phage RNA polymerase. Built using the crystal structure of the *Escherichia* bacteriophage T7 DNA-dependent RNA polymerase (T7 RNAP, pdb_id: 3E3J).

BIOZENTRUM
University of Basel
The Center for Molecular Life Sciences

SWISS-MODEL Homology Modelling Report

Model Building Report

This document lists the results for the homology modelling project "Phage DNA-directed RNA polymerase GGM3" submitted to SWISS-MODEL workspace on June 6, 2020, 8:55 p.m.

Results

The SWISS-MODEL template library (SMTL version 2020-06-03, PDB release 2020-05-29) was searched with BLAST (Camacho et al.) and HHBlits (Remmert et al.) for evolutionary related structures matching the target sequence in Table T1. For details on the template search, see Materials and Methods. Overall 48 templates were found (Table T2).

Models

The following model was built

Model #01	File	Built with	Oligo-State	Ligands	GMQE	QMEAN
	PDB	ProMod3 3.0.0	monomer	None	0.74	-5.27

QMEAN	-5.27
Cβ	-3.49
All Atom	-1.72
solvation	-0.18
torsion	-4.45

Residue Number

Protein Size (Residues)

Template	Seq Identity	Oligo-state	QSQE	Found by	Method	Resolution	Seq Similarity	Range	Coverage	Description
3e3j.1.A	45.02	monomer	0.00	BLAST	X-ray	6.70Å	0.41	7 - 539	0.97	DNA-directed RNA polymerase

Excluded ligands

Ligand Name.Number	Reason for Exclusion	Description
G-G-G-A-G-U-A-A.1	Binding site not conserved.	RNA (5'-R(*GP*GP*GP*AP*GP*UP*AP*A)-3')

Figure 8. Swiss-Model homology modelling report of the Phage RNA polymerase found in GGM3.

- **Cloning and Expression of Recombinant RNA polymerase**

As presented above, it was possible to identify 29 sequences annotated as polymerases (DNA-directed RNA polymerases (2.7.7.6)) from the bioinformatics analysis of GGM3 metagenome. From these, only one was specifically annotated as "Phage DNA-directed RNA polymerase", suggesting that it could be a single subunit phage-type polymerase, which is 100% identical to the single subunit T3/T7 RNA polymerase from *Strenotrophomonas maltophila* (WP_164076363.1) and shares 44% identity with a DNA-dependent RNA polymerase from *Rhizobium phage* RHEph01 (44%). The contig containing the phage-type RNA polymerase sequence is 8490 bp long and encodes 15 CDS, of which only 3 has functional annotation (all of them being polymerases) with the remaining sequences annotated as hypothetical proteins.

This Phage DNA-directed RNA polymerase encoding gene sequence was synthetized, and then cloned into the expression vector pJ444 (DNA 2.0). The resulting pJ444-RNAPol vector was sequenced to confirm correct cloning prior to transformation. *Escherichia coli* BL21 competent cells (New England Biolabs) were chemically transformed with pJ444 vector carrying the RNA polymerase encoding gene under the control of T5 promoter and a kanamycin resistance gene (Figure 9).

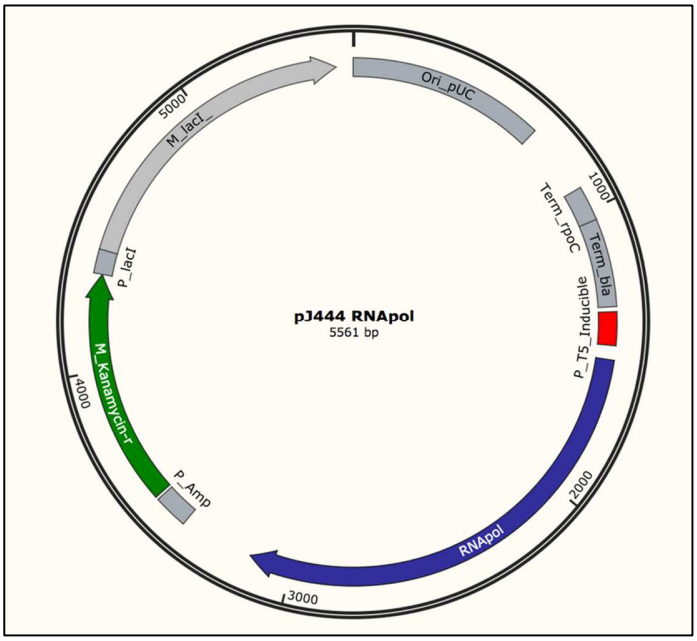


Figure 9. Expression vector pJ444-RNApolRec map with main features colored. *Red: T5 promoter, blue: RNA polymerase encoding gene, green: Kanamycin resistance gene.*

After correct cloning, it was necessary to test and optimize the soluble and functional heterologous expression of the protein. In order to do this, three preliminary sets of conditions for the recombinant expression of the RNA polymerase were selected (Table 4).

Table 4. Test/optimization of the RNA polymerase recombinant expression.

Expression Trial	Media	Temperature	Time	Agitation
#1	50 ml TBA	**30°C**	**20 hrs**	150 rev min−1
#2	50 ml TBA	**23°C**	**26 hrs**	150 rev min−1
# 3	50 ml TBA	**37°C**	**18 hrs**	150 rev min−1

For the first expression trial, two transformants, named as RNAP2 and RNAP4, were selected. Both clones, along with *E. coli* BL21 cells with no vector (control), were grown aerobically in 50 ml TBA autoinduction medium supplemented with 30 µg/ml kanamycin, at 30°C with shaking at 150 rev min−1 for 20 hrs. The cells were harvested by centrifugation at 9000g for 15 min at 4 °C and resuspended in 5 ml lysis buffer (50 mM Tris-HCl, pH 8.0, lysozyme 1 mg/ml). Cell disruption was carried out by ten 15 s bursts of sonication using a Digital Sonifier, and the cell lysate (Total Fraction) was centrifuged at 14,000g for 30 min at 4 °C, in order to obtain the soluble and Insoluble Fractions. Protein concentration was determined by the method of Bradford.

Recombinant expression of RNA polymerase was evaluated by SDS-PAGE and visualized by staining with Coomassie brilliant blue (Figure 10).

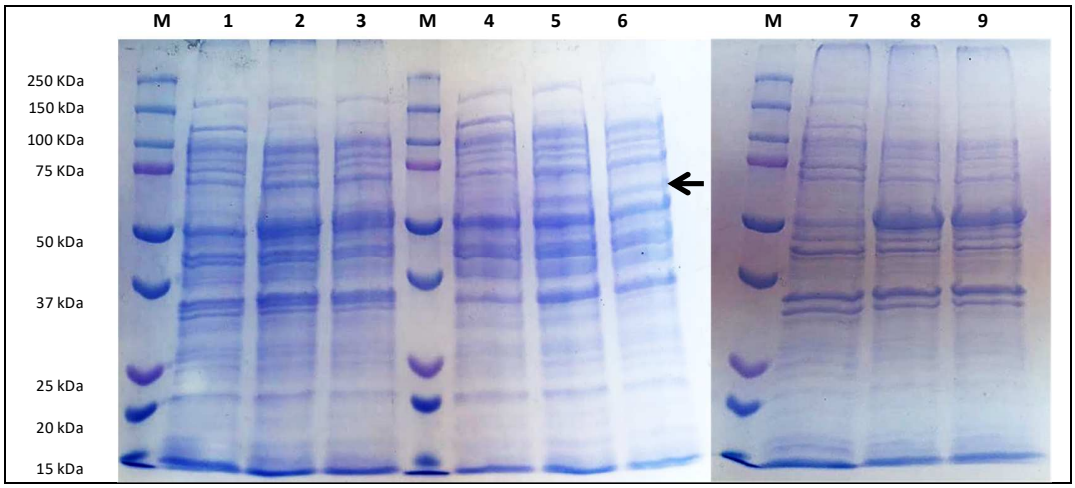


Figure 10. Coomassie Blue stained 12% SDS-PAGE electrophoresis. First RNA polymerase recombinant expression trial. *Lanes: (M) Protein Ladder (Precision Plus Protein Dual Color Standard Biorad). 1) Total Fraction Control (E. coli BL21 cells with no vector). 2) Total Fraction RNApol2 clone. 3) Total Fraction RNApol4 clone. 4) Soluble Fraction Control. 5) Soluble Fraction RNApol2 clone. 6) Soluble Fraction RNApol4 clone. 7) Insoluble Fraction Control. 8) Insoluble Fraction RNApol2 clone. 9) Insoluble Fraction RNApol4 clone. The protein concentration loaded in each well is 20 ug, the possible recombinant RNA polymerase is highlighted with an arrow.*

According to Expasy ProtParam Tools analysis, the recombinant Phage RNA polymerase has a theoretical molecular weight of 59.59 kDa predicted from its amino acid sequence. It is possible to observe in Figure 10, that in the soluble fraction of both clones, RNAP2 y RNAP4 (lanes 5 and 6 respectively) there is a light band (highlighted with an arrow) between the bands corresponding to 75 and 50 kDa in the Protein Ladder. This band is interesting as it seems does not appear in the control soluble fraction (lane 4) or in the insoluble fraction, suggesting recombinant expression of the RNA polymerase in soluble form under the conditions tested (TBA autoinduction medium, 30°C x 20 h). In order to check functionality of the recombinant enzyme, RNA polymerase activity assays needed to be perform.

For this, a non-radioactive RNA polymerase coupled assay using inorganic pyrophosphatase enzyme (based on Lee et al., 2009) has been implemented and used since the beginning of this project. During this enzyme assay, one pyrophosphate is released by each ribonucleoside that is added to nascent RNA chain, and pyrophosphatase enzyme is used to hydrolyze it and each orthophosphate released is bind to ammonium molybdate. The results are obtained with a calibration curve that show the linear correlation between the increasing concentrations of orthophosphate bound to the ammonium molybdate in the reaction measured at 820 nm. This coupled assay proved to be useful when trying to identify and purify the native RNA polymerase from the different consortia. Due to the promoter sequence is unknown, nanocircles (small, circular, single-stranded DNA (Mohsen and Kool 2016)) were used as non-promoter dependent DNA templates for the RNA polymerase assay. To generate them two different oligonucleotides of 34 bp, named Nano1 and Nano 2, were designed and synthetized with 5' phosphorylation modification in order to self-ligate with T4 ligase enzyme and correct circularization of the oligonucleotides was routinely checked by 15% Urea-PAGE gel (Summer et al., 2009).

RNA polymerase activity assays were performed using nanocircles DNA templates and 10 µL of the soluble fractions from both clones, RNAP2 y RNAP4 (Figure10, lane 5 and 6 respectively). The results are shown in Table 5:

Table 5. *RNA polymerase activity assays were performed using templates nanocircle 1 and nanocircle 2, 10 µl sample at 37°Cx3h.*

	Nano1			Nano2		
	Total Phosphate (nmol)	**Total Pyrophosphate (nmol)**	**Specific Activity**	**Total Phosphate (nmol)**	**Total Pyrophosphate (nmol)**	**Specific Activity**
Soluble Fraction Control	2240.9	1120.4	**179.9**	2012.2	1006.1	**161.5**
Soluble Fraction RNAP2	4226.3	2113.1	**427.8**	3695.9	1847.9	**374.1**
Soluble Fraction RNAP4	11330.9	5665.5	**762.5**	11399.0	5699.5	**767.1**

*One unit is defined as the amount of enzyme that will incorporate 1 nmol ATP (free 1 nmol of pyrophosphate) into acid-insoluble material in 3 hours at 37°C.

The results obtained indicate that both clones possess higher RNA polymerase activity than the control *E. coli* BL21 cells with no vector, with clone RNAP4 being more active than clone RNAP2. It is unclear if the activity measured in the control fraction corresponds to RNA polymerase activity or something unspecific.

The first expression trial performed showed that the recombinant enzyme seems to be expressed in soluble and active form, however its expression needs to be optimized. For the second and third expression trials (Table 4), RNAP4 clone and *E. coli* BL21 cells with no vector (control), were grown aerobically in 50 ml TBA autoinduction medium supplemented with 30 µg/ml kanamycin, at 23°C with shaking at 150 rev min−1 for 26 h (trial #2) and at 37°C with shaking at 150 rev min−1 for 18 h (trial #3).

The cells were harvested by centrifugation at 9000g for 15 min at 4 °C and resuspended in 5 ml lysis buffer (50 mM Tris-HCl, pH 8.0, lysozyme 1 mg/ml). Cell disruption was carried out by ten 15 s bursts of sonication using a Digital Sonifier, and the cell lysate (Total Fraction) was centrifuged at 14,000g for 30 min at 4 °C, in order to obtain the Soluble and Insoluble Fractions. Protein concentration was determined by the method of Bradford.

Recombinant expression of the RNA polymerase was evaluated by SDS-PAGE and visualized by staining with Coomassie brilliant blue (Figure 11).

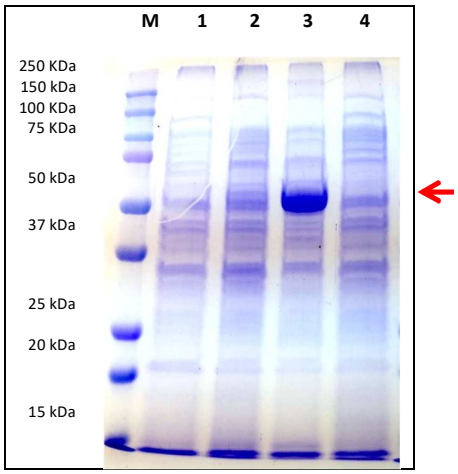


***Figure 11**. Coomassie Blue stained 12% SDS-PAGE electrophoresis. Second and third RNA polymerase recombinant expression trials. Lanes: (M) Protein Ladder (Precision Plus Protein Dual Color Standard Biorad). 1) Soluble Fraction Control (E. coli BL21 cells with no vector) grew at 23°C x 26 h. 2) Soluble Fraction RNAP4 grew at 23°C x 26 h. 3) Soluble Fraction RNAP4 grew at 37°C x 18 h. 4) Soluble Fraction Control (E. coli BL21 cells with no vector) grew at 37°C x 18 h. The protein concentration loaded in each well is 20 µg, the possible recombinant RNA polymerase is highlighted with an arrow.*

It is possible to observe in Figure 11, that in the soluble fractions from RNAP4 grew at 37°C x 18 h (lane 3) there is a strong band (highlighted with an arrow) between the bands corresponding to 75 and 50 kDa in the Protein Ladder. This prominent band does not appear in the control fractions (lane 1 and 4) and is fainter in the RNAP4 grew at 23°C x 26 h (lane 2). Therefore, the three different conditions trials tested resulted in the optimization of the recombinant expression of the RNA polymerase in soluble form under the conditions tested in the trial #3: TBA autoinduction medium, 37°C x 18 h).

In order to check functionality of the recombinant enzyme, RNA polymerase assays were perform using nanocircles DNA templates and only 1 µL of each soluble fractions. The results are shown in Table 6:

***Table 6.** RNA polymerase activity assays were performed using templates nanocircle 1 and nanocircle 2, 1 ul sample at 37°C x 3h.*

	Nanocircle 1			Nanocircle 2		
Trial #3, 37°C x 18 h	**Total Phosphate (nmol)**	**Total Pyrophosphate (nmol)**	**Specific Activity**	**Total Phosphate (nmol)**	**Total Pyrophosphate (nmol)**	**Specific Activity**
Soluble Fraction Control	44817.5	22408.8	**747.0**	55912.4	27956.2	**931.9**
Soluble Fraction RNAP4,	108321.2	54160.6	**1593.0**	122919.7	61459.9	**1807.6**

*One unit is defined as the amount of enzyme that will incorporate 1 nmol ATP (free 1 nmol of pyrophosphate) into acid-insoluble material in 3 hours at 37 °C.

The results obtained indicate that RNAP4 possess higher RNA polymerase activity than the control *E. coli* BL21 cells with no vector. Even though activity is also detected in the control fractions, the difference between the specific activities of soluble and heat shocked fraction from RNAP4 is at least 800 units above the specific activity of the control for both nanocircles.

It is worth to note that these RNA polymerase activity assays were performed using only 1 ul, instead of 10 µl sample that were normally used for the activity assays and nanocircles DNA templates. Furthermore, the activity assays were performed after 2 weeks stored at 4 °C suggesting that the recombinant enzyme is stable under those conditions

Partial Purification of Soluble Recombinant RNA Polymerase

In order to facilitate the purification process of the recombinant enzyme, Heat-shock tests were performed to denature some of the native *E. coli* proteins prior to purification. These were done in small volumes (1 ml) at 50**,** 60 y 70**°**C for 10, 20 y 30 min. After incubation, the samples were centrifuged at 14,000g for 30 min at 4 °C, and loaded on an SDS-PAGE gel (Figure 12).

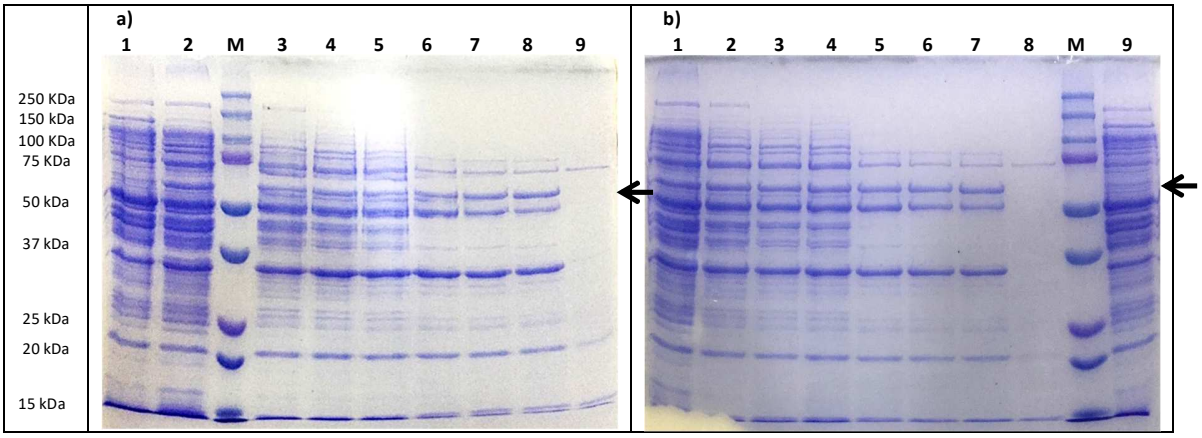


Figure 12. Coomassie Blue stained 12% SDS-PAGE electrophoresis. Heat shock trial*. Lanes: (M) Protein Ladder (Precision Plus Protein Dual Color Standard Biorad).* ***a. Soluble Fraction RNAP2.*** *1) Total Fraction Control. 2) Total Fraction Clone RNAP2. 3) HS 50°C x 10 min. 4) HS 50°C x 20 min. 5) HS 50°C x 30 min. 6) HS 60°C x 10 min. 7) HS 60°C x 20 min. 8) HS 60°C x 30 min. 9) HS 70°C x 10 min.* ***b) Soluble Fraction RNAP.*** *1) Total Fraction Clone RNAP4. 2) HS 50°C x 10 min. 3) HS 50°C x 20 min. 4) HS 50°C x 30 min. 5) HS 60°C x 10 min. 6) HS 60°C x 20 min. 7) HS 60°C x 30 min. 8) HS 70°C x 10 min. 9) Total Fraction Control. The gel was normalized by simple volume (10 ul each), the possible recombinant RNA polymerase is highlighted with an arrow.*

Figure 12, 3a and 3b showed that the band that possibly correspond to the recombinant RNA polymerase (highlighted with an arrow) is present in all conditions tested, excepting the controls (lanes a1 and b9) and HS at 70°C x 10 min (lanes a9 and b8).

As shown above in Table 4, the transformant named as RNAP4 possess higher specific activity than the other clone RNAP2, and for this reasons this clone was selected for further studies. Activity assays were done for the heat shock conditions: 50°C x 20 min, 60°C x 10 min, y 60°C x 20 min (Table 7). Heat shocks performed for 30 min were not considered as they did not show much difference in terms of purity, and longer incubations could affect enzyme activity (depending on its thermostability). It is worth to note that these RNA polymerase activity assays were performed using only 3 µl, instead of 10 µl sample that are normally used for the activity assays and nanocircles DNA templates, and as they started with the same protein concentration, enzyme activity is expressed as volumetric not specific.

Table 7. ***RNA polymerase activity assays were performed using templates nanocircle 1 and nanocircle 2, 3 µl sample at 37°C x 3h****.*

	Nano1		Nano2	
Trial #1, 30°C x 20 h	**Total Phosphate (nmol)**	**Total Pyrophosphate (nmol)**	**Total Phosphate (nmol)**	**Total Pyrophosphate (nmol)**
Soluble Fraction Control	8625.3	**4312.7**	9768.9	**4884.4**
Soluble Fraction Pol 4	15851.6	**7925.8**	17262.8	**8631.4**
Pol4 HS 50°C x 20 min	15973.2	**7986.6**	22007.3	**11003.7**
Pol4 HS 60°C x 10 min	20328.5	**10164.2**	23856.5	**11928.2**
Pol4 HS 60°C x 20 min	20036.5	**10018.3**	21496.4	**10748.2**
T7 RNA Pol	401.5	**200.7**	60.8	**30.4**

According to these results, heat shock at 60°C x 10 min was selected for further experiments.

Identification of RNA Polymerase Promoter

The use of nanocircles that are efficiently transcribed by RNA Polymerases without the requirement for promoter sequences has allowed to successfully assess the enzyme activity. However, this project aims to identify the specific promoter sequence used by this new viral RNA polymerase for the transcription process, particularly thinking in the potential applications of it, for the development of specific biosensors of high interest for the Air Force.

As presented above, the bioinformatics search for promoters was conducted using Promoter Hunter with default scoring matrices and considering both strands. The table in Figure 13 shows the sequences and coordinates of the predicted promoters along with their respective scores.

Predicted Promoter	Strand	Begin 1	Score 1	Begin 2	Score 2	Summary score	Energy score	Final score	Sequence
P1	-	7303	4.02	7280	1.65	5.67	6.15	2.42	CCATCCCCCTTTGGCAATTGATACCCCAAACGAGAACATCCACATGAGC
P2	-	1917	6.20	1893	-0.68	5.52	6.28	2.40	CTCGACGAACTTGACACGATCCACTGCATTTCCCTCCAAGAGGTAGATGA
P3	-	7296	5.56	7274	-1.19	4.37	6.11	2.38	CCTTTGGCAATTGATACCCCAAACGAGAACATCCACATGAGCATCAAC
P4	+	2886	2.81	2908	4.08	6.89	6.48	2.38	GACGATCTCTTTCATTGGGGTAGTTCCTCGGCTTTGATGTGGATGACC
P5	+	5803	3.38	5827	3.57	6.95	6.54	2.37	TCTTGGCGCGTTTGCATACCGCGATGATGACCACTAAAACCGGATCGGTC
P6	+	7663	2.06	7686	5.25	7.31	6.57	2.36	ACGGTCGAGCTTAGTACCGCCGATGTTGACCACTAAATTCACCTCGCGG
P7	+	7302	4.45	7326	-0.76	3.69	6.07	2.34	GGGGTATCAATTGCCAAAGGGGGATGGACTAGGCAAAGAATAGCGCGAC
P8	+	7265	3.79	7286	-0.06	3.73	6.07	2.32	GTGCTGGTAGTTGATGCTCATGTGGATGTTCTCGTTTGGGGTATCAA
P9	-	6421	4.27	6400	1.01	5.28	6.47	2.30	GAGAACGGCGATGACACCGGCCGCGTGCTGGTGAAGTTCAAGATGAA
P10	+	6532	3.81	6556	1.01	4.82	6.43	2.28	TCGCGCCGGGGTTTCCAGCTTCTTCGCCAGCGCCTTGAAGTCCTTGTCGTT

Figure 13. phiSITE PromoterHunter search result. *Predicted promoters are highlighted in blue in the Sequence column.*

In order to experimentally test and validate the functionality of the ten different predicted theoretical promoter sequences found, they were synthetized along with a ribosome-binding site and the nucleotide sequence of GFP as reporter gene. These constructs were based on pJL1-sfGFP vector Addgene #69496, Wang et al., 2018), and were synthetized and cloned in opposite direction to the ampicillin promoter in the cloning vector pUC57-mini that contains only the origin of replication and ampicillin resistance gene. An example of the constructs designed to test the putative promoter sequence is shown in Figure 14.

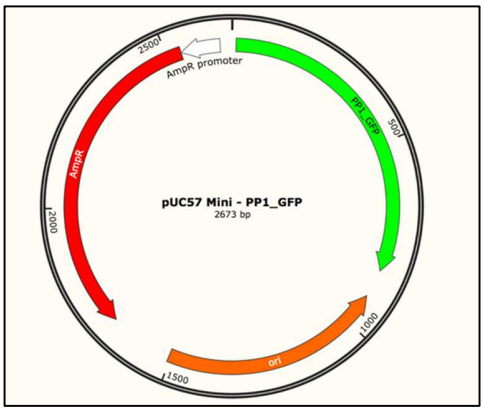


Figure 14. Example of the constructs designed to test the putative promoter sequence*, the insert (putative promoter sequence, a ribosome binding site and GFP reporter gene) is cloned in pUC57 Mini cloning vector. The 10 different he constructs were synthetized by Genscript (USA).*

As soon as the constructs arrived, they were transformed in *E. coli* chemically competent cells. Single colonies carrying each construct were grown and stored as glycerol stocks at -80°C. Minipreps were performed to obtain each plasmid. P0 was designed as a control construct, containing T7 promoter.

In order to check functionality of the recombinant enzyme, the same RNA polymerase activity assay coupled to a phosphatase enzyme was performed. This time using each plasmid containing the putative promoter sequence as DNA templates instead of nanocircles.

The reaction conditions were 5 µl of the different promoter constructs as templates and 1 µL of sample (heat shocked soluble fractions) instead of 10 ul sample that were normally used for the activity assays when using the native enzyme. The results obtained are shown in Table 8:

Table 8. *RNA polymerase activity assays were performed using 5 µl different promoter constructs as templates, 1 µl sample at 37°C x 3 h.*

	RNAP4			Control		
Promoter Constructs	**Total Phosphate (nmol)**	**Total Pyrophosphate (nmol)**	**Specific Activity**	**Total Phosphate (nmol)**	**Total Pyrophosphate (nmol)**	**Specific Activity**
P1_pUC57-Mini	17226.3	8613.1	**5236.4**	34598.5	17299.3	**8731.6**
P2_pUC57-Mini	54890.5	27445.3	**12059.7**	34890.5	17445.3	**8786.7**
P3_pUC57-Mini	25109.5	12554.7	**6664.6**	30948.9	15474.5	**8043.0**
P4_pUC57-Mini	24963.5	12481.8	**6638.1**	35036.5	17518.2	**8814.2**
P5_pUC57-Mini	39854.0	19927.0	**9335.7**	31970.8	15985.4	**8235.8**
P6_pUC57-Mini	30073.0	15036.5	**7563.7**	39854.0	19927.0	**9723.2**
P7_pUC57-Mini	38978.1	19489.1	**9177.0**	35328.5	17664.2	**8869.3**
P8_pUC57-Mini	18248.2	9124.1	**5421.6**	31970.8	15985.4	**8235.8**
P9_pUC57-Mini	26423.4	13211.7	**6902.6**	31824.8	15912.4	**8208.2**
P10_pUC57-Mini	28467.2	14233.6	**7272.8**	34452.6	17226.3	**8704.0**

*One unit is defined as the amount of enzyme that will incorporate 1 nmol ATP (free 1 nmol of pyrophosphate) into acid-insoluble material in 3 hours at 37 °C.

The results obtained in the first activity assay using promoters was quite promising as the conditions for the recombinant enzymes were not optimized. Using the putative promoter sequence annotated as P2 shows the highest activity and it is significantly different than the activity of the control *E. coli* BL21 cells with no vector. Even though activity is also detected in the control fractions using all the different constructs, the difference between the specific activities of Promoter P2 with RNAP4 is over 3000 units above the specific activity obtained with the control. Also, the activity measured this time is higher than the observed and using nanocircles 1 and 2 (Table 9) indicating that the 10 different constructs containing the putative promoters might be more specific templates for the recombinant RNA polymerase.

Table 9. RNA polymerase activity assays were performed using templates nanocircle 1 and nanocircle 2, 1 µl sample at 37°C x 3h.

	Nanocircle 1			Nanocircle 2		
Trial #3, 37°C x 18 h	**Total Phosphate (nmol)**	**Total Pyrophosphate (nmol)**	**Specific Activity**	**Total Phosphate (nmol)**	**Total Pyrophosphate (nmol)**	**Specific Activity**
Heat Shock 60 °C x 10 min Control	58394.2	29197.1	**2919.7**	71678.8	35839.4	**3583.9**
Heat Shock 60 °C x 10 min RNAP4	69051.1	34525.5	**4932.2**	61605.8	30802.9	**4400.4**

*One unit is defined as the amount of enzyme that will incorporate 1 nmol ATP (free 1 nmol of pyrophosphate) into acid-insoluble material in 3 hours at 37 °C.

Four different biological replicates (independent bacterial growth, expression, cell fractionation and heat-shock) were tested. The results obtained are shown in Tables 10 – 13 and Figures 15 - 18.

Table 10. *First RNA polymerase expression. RNA polymerase activity assays were performed using 5 μl of different promoter constructs as templates (P0-P10), 1 μl heat shocked enzyme sample at 37℃ x 3 h.*

#1	RNAP4			Control		
Promoter Constructs	Total Phosphate (nmol)	Total Pyrophosphate (nmol)	Specific Activity	Total Phosphate (nmol)	Total Pyrophosphate (nmol)	Specific Activity
P1_pUC57-Mini	21605.8	10802.9	**3914.1**	38978.1	19489.1	**7354.4**
P2_pUC57-Mini	59270.1	29635.0	**10737.3**	39270.1	19635.0	**7409.4**
P3_pUC57-Mini	29489.1	14744.5	**5342.2**	35328.5	17664.2	**6665.7**
P4_pUC57-Mini	29343.1	14671.5	**5315.8**	39416.1	19708.0	**7437.0**
P5_pUC57-Mini	44233.6	22116.8	**8013.3**	36350.4	18175.2	**6858.6**
P6_pUC57-Mini	34452.6	17226.3	**6241.4**	44233.6	22116.8	**8346.0**
P7_pUC57-Mini	43357.7	21678.8	**7854.6**	39708.0	19854.0	**7492.1**
P8_pUC57-Mini	22627.7	11313.9	**4099.2**	36350.4	18175.2	**6858.6**
P9_pUC57-Mini	30802.9	15401.5	**5580.2**	36204.4	18102.2	**6831.0**
P10_pUC57-Mini	32846.7	16423.4	**5950.5**	38832.1	19416.1	**7326.8**

*One unit is defined as the amount of enzyme that will incorporate 1 nmol ATP (free 1 nmol of pyrophosphate) into acid-insoluble material in 3 hours at 37 °C.

Table 11. *Second RNA polymerase expression. RNA polymerase activity assays were performed using 5 μl of different promoter constructs as templates (P0-P10), 1 μl heat shocked enzyme sample at 37℃ x 3 h.*

#2	RNAP4			Control		
Promoter Constructs	Total Phosphate (nmol)	Total Pyrophosphate (nmol)	Specific Activity	Total Phosphate (nmol)	Total Pyrophosphate (nmol)	Specific Activity
P0_pUC57-Mini	38978.1	19489.1	**9892.9**	33430.7	16715.3	**5571.8**
P1_pUC57-Mini	39854.0	19927.0	**10115.2**	35036.5	17518.2	**5839.4**
P2_pUC57-Mini	43065.7	21532.8	**10930.4**	36350.4	18175.2	**6058.4**
P3_pUC57-Mini	37518.2	18759.1	**9522.4**	35474.5	17737.2	**5912.4**
P4_pUC57-Mini	34890.5	17445.3	**8855.5**	38978.1	19489.1	**6496.4**
P5_pUC57-Mini	31678.8	15839.4	**8040.3**	45839.4	22919.7	**7639.9**
P6_pUC57-Mini	32992.7	16496.4	**8373.8**	40146.0	20073.0	**6691.0**
P7_pUC57-Mini	32846.7	16423.4	**8336.7**	39562.0	19781.0	**6593.7**
P8_pUC57-Mini	34744.5	17372.3	**8818.4**	35766.4	17883.2	**5961.1**
P9_pUC57-Mini	37226.3	18613.1	**9448.3**	18832.1	9416.1	**3138.7**
P10_pUC57-Mini	41751.8	20875.9	**10596.9**	48175.2	24087.6	**8029.2**

*One unit is defined as the amount of enzyme that will incorporate 1 nmol ATP (free 1 nmol of pyrophosphate) into acid-insoluble material in 3 hours at 37 °C.

Table 12. *Third RNA polymerase expression. RNA polymerase activity assays were performed using 5 μl different promoter constructs as templates (P0-P10), 1 μl heat shocked, 1 μl sample at 37℃ x 3 h.*

#3	RNAP4			Control		
Promoter Constructs	Total Phosphate (nmol)	Total Pyrophosphate (nmol)	Specific Activity	Total Phosphate (nmol)	Total Pyrophosphate (nmol)	Specific Activity
P0_pUC57-Mini	109781.0	54890.5	8483.9	90365.0	45182.5	5647.8
P1_pUC57-Mini	120292.0	60146.0	9296.1	92262.8	46131.4	5766.4
P2_pUC57-Mini	117518.2	58759.1	9081.8	79854.0	39927.0	4990.9
P3_pUC57-Mini	148029.2	74014.6	11439.7	98248.2	49124.1	6140.5
P4_pUC57-Mini	169343.1	84671.5	13086.8	123649.6	61824.8	7728.1
P5_pUC57-Mini	199708.0	99854.0	15433.4	160583.9	80292.0	10036.5
P6_pUC57-Mini	148905.1	74452.6	11507.4	118248.2	59124.1	7390.5
P7_pUC57-Mini	105109.5	52554.7	8122.8	108029.2	54014.6	6751.8
P8_pUC57-Mini	128613.1	64306.6	9939.2	114014.6	57007.3	7125.9
P9_pUC57-Mini	181021.9	90510.9	13989.3	147883.2	73941.6	9242.7
P10_pUC57-Mini	151240.9	75620.4	11687.9	129635.0	64817.5	8102.2

*One unit is defined as the amount of enzyme that will incorporate 1 nmol ATP (free 1 nmol of pyrophosphate) into acid-insoluble material in 3 hours at 37 °C.

Table 13. *Fourth RNA polymerase expression. RNA polymerase activity assays were performed using 5 µl different promoter constructs as templates (P0-P10), 1 µl heat shocked sample at 37°C x 3 h.*

#4	RNAP4			Control		
Promoter Constructs	Total Phosphate (nmol)	Total Pyrophosphate (nmol)	Specific Activity	Total Phosphate (nmol)	Total Pyrophosphate (nmol)	Specific Activity
P0_pUC57-Mini	88029.2	**44014.6**	7588.7	55912.4	27956.2	4904.6
P1_pUC57-Mini	117372.3	**58686.1**	10118.3	56788.3	28394.2	4981.4
P2_pUC57-Mini	84817.5	**42408.8**	7311.9	73722.6	36861.3	6466.9
P3_pUC57-Mini	59416.1	**29708.0**	5122.1	78540.1	39270.1	6889.5
P4_pUC57-Mini	159854.0	**79927.0**	13780.5	88321.2	44160.6	7747.5
P5_pUC57-Mini	120000.0	**60000.0**	10344.8	69489.1	34744.5	6095.5
P6_pUC57-Mini	122481.8	**61240.9**	10558.8	94744.5	47372.3	8310.9
P7_pUC57-Mini	153868.6	**76934.3**	13264.5	160875.9	80438.0	14111.9
P8_pUC57-Mini	156642.3	**78321.2**	13503.6	87883.2	43941.6	7709.1
P9_pUC57-Mini	153722.6	**76861.3**	13252.0	80438.0	40219.0	7056.0
P10_pUC57-Mini	176496.4	88248.2	15215.2	125547.4	62773.7	11012.9

*One unit is defined as the amount of enzyme that will incorporate 1 nmol ATP (free 1 nmol of pyrophosphate) into acid-insoluble material in 3 hours at 37 °C.

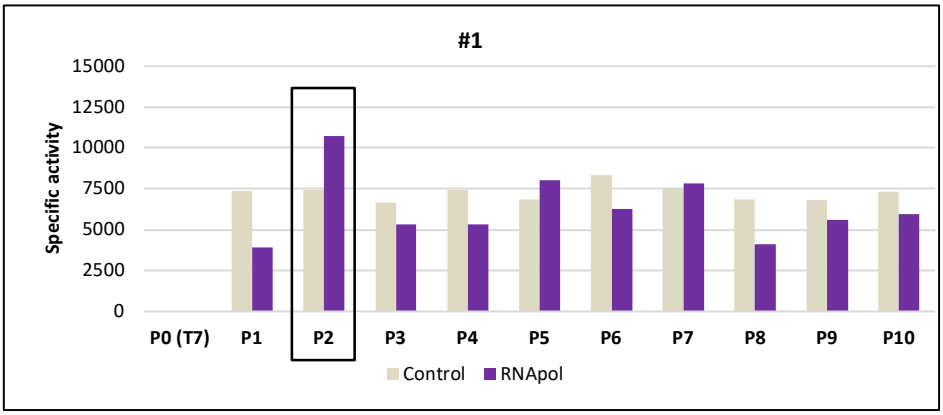


Figure 15. *First RNA polymerase expression. RNA polymerase activity assays were performed using 5 µl different promoter constructs as templates (P0-P10), 1 µl heat shocked sample at 37°C x 3 h.*

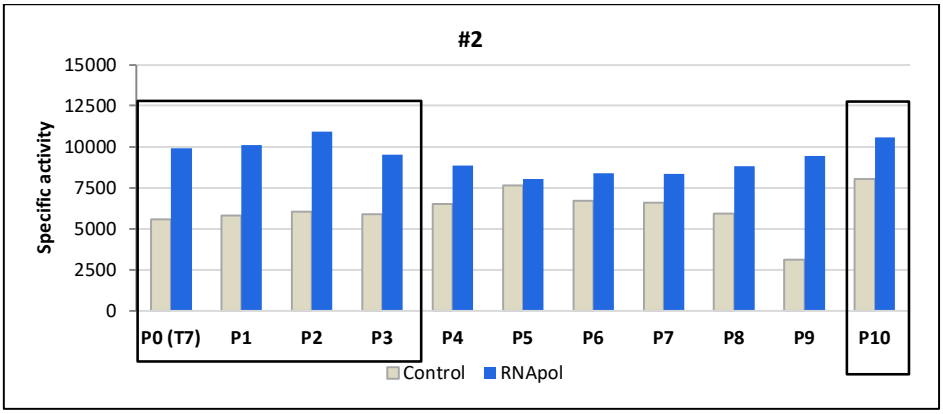


Figure 16. *Second RNA polymerase expression. RNA polymerase activity assays were performed using 5µl different promoter constructs as templates (P0-P10), 1 µl heat shocked sample at 37°C x 3 h.*

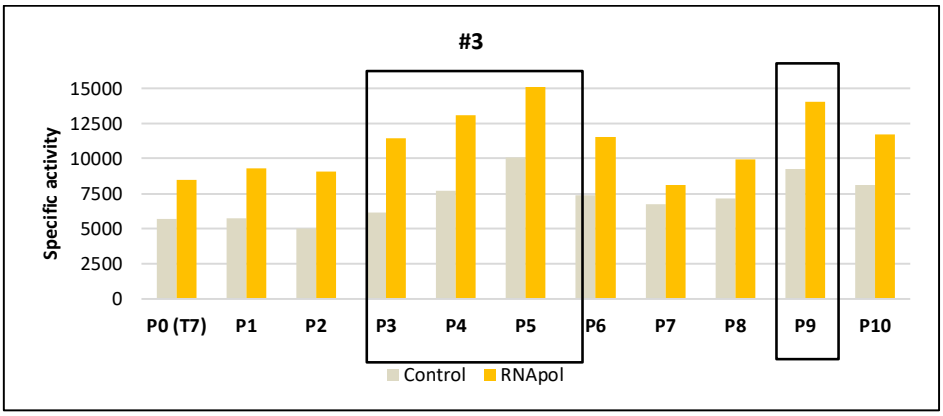


Figure 17. *Third RNA polymerase expression. RNA polymerase activity assays were performed using 5 μl different promoter constructs as templates (P0-P10), 1 μl heat shocked sample at 37°C x 3 h.*

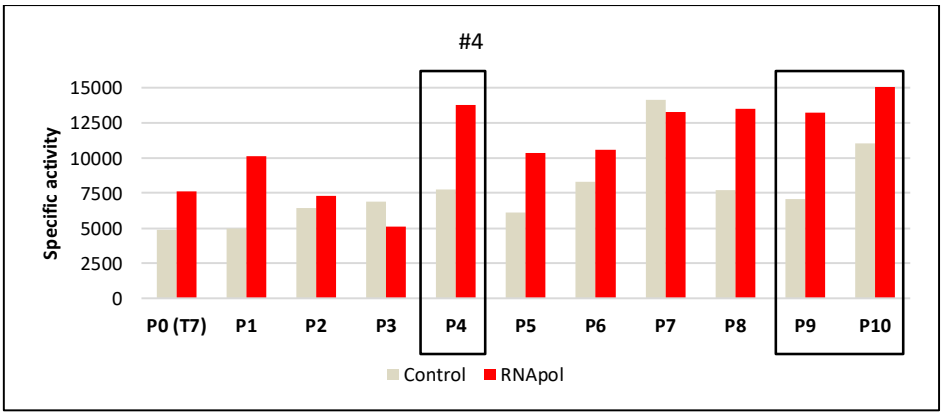


Figure 18. *Fourth RNA polymerase expression. RNA polymerase activity assays were performed using 5 μl different promoter constructs as templates (P0-P10), 1 μl heat shocked sample at 37°C x 3 h.*

The results obtained from these four independent replicates indicate that the recombinant RNA polymerase is active with several predicted putative promoters sequences, confirming that they are indeed promoters. Even though the main objective of this project is to identify **The specific promoter**, it seems that the enzyme is not highly selective towards just one specific promoter sequence, indicating the versatility of this enzyme. Another reason could be that this coupled enzyme assay is not very reproducible to allows us to identify a specific promoter sequence. However, results obtained at AFRL by Dr. Svetlana Harbaugh and cols. testing the same promoter sequences in different constructs in *E. coli* cell-free extract seems to also indicate the versatility of the enzyme (Figure 19).

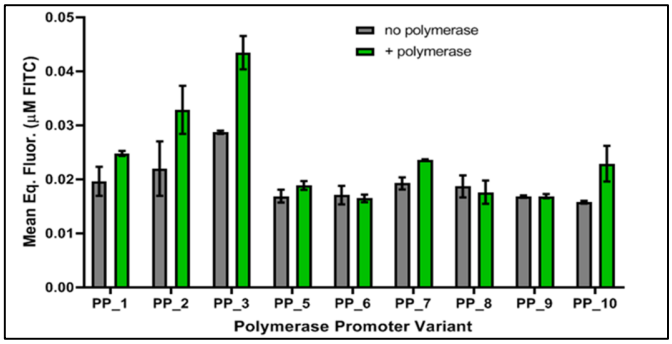


Figure 19. Testing functionality of predicted polymerase promoters in E. coli cell-free extract. Results obtained at AFRL by Dr. Svetlana Harbaugh and cols.

E. coli cells with no vector were used as a control, they also showed RNA polymerase activity, although it was significantly different than most of the activities obtained with *E. coli* cells harboring the RNApol4 expression vector. It is unclear if the activity measured in the control fractions corresponds to RNA polymerase activity, something unspecific or an interference, as the assay we used is based on the detection of phosphates. The enzyme assay based on sGFP fluorescence performed at AFRL (Figure 19) also showed activity with the no polymerase-control.

- **Scale up the recombinant expression of RNA polymerase**

Due to unknown problems and in spite of several trials, it was not possible to scale-up the expression of the recombinant RNA polymerase using a 5L bioreactor, presumably due to oxygen level differences between the different scale culture vessels, or equipment problem (e.g. electrodes or sensors), ended in very low soluble expression. The largest culture was obtained using a 400 ml using a 2 L Erlenmeyer Flask and the same growth conditions previously determined to be the best (trial #3): 37°C x 18 h, 150 rev min^{-1}, using LBA medium instead of TBA, followed by a 60 °C x 10 min heat shock of the crude extract.

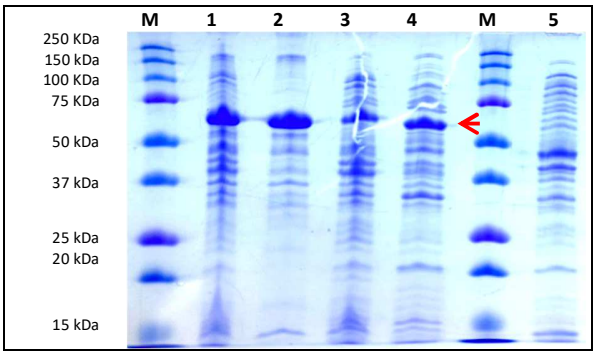


Figure 20. Pre-scale-up using 400 ml LBA culture grew at 37°C x 18 h. *Coomassie Blue stained 12% SDS-PAGE electrophoresis. Lanes: (M) Protein Ladder. 1) Total Fraction E. coli RNAP4. 2) Insoluble Fraction E. coli RNAP4. 3) Soluble Fraction E. coli RNAP4. 4) Heat Shock 60°C x 10 min Soluble Fraction RNAP4. 5) Soluble Fraction E. coli (control). The protein concentration loaded in each well is 20 µg, the recombinant RNA polymerase is highlighted with an arrow.*

The scaling up the expression from 50 ml to 400 ml cultures using a 2L Erlenmeyer flask was successful as it is possible to observe a similar band as the one previously obtained. Interestingly, there is a large amount of recombinant protein being overexpressed in the insoluble fraction as inclusion bodies (Figure 20, Lane 2) indicating there is still room for improvement of the soluble expression or another source to explore for enzyme purification. Even though the heat shocked fraction (Figure 20, Lane 4) does not look very different than soluble fraction (Figure 20, Lane 3), there is a reduction of over 50% of total proteins as measured by Bradford method, and the specific activity of this first step of purification increases accordingly (Table 14).

Table 14. RNA polymerase activity. *Enzyme assays were performed using 5 µl different promoter constructs as templates (P2 and P3), 1 µl heat shocked sample at 37°C x 3 h.*

	Promoter Construct P2_pUC57-Mini			Promoter Construct P3_pUC57-Mini		
Fraction	**Total Phosphate** (nmol)	**Total Pyrophosphate** (nmol)	**Specific Activity**	**Total Phosphate (nmol)**	**Total Pyrophosphate (nmol)**	**Specific Activity**
Soluble	31459.9	**15729.9**	**1014.83**	45547.4	**22773.7**	**1469.27**
Heat shocked 60 °C x 10 min	39343.1	**19671.5**	**2892.87**	41386.9	**20693.4**	**3043.15**

*One unit is defined as the amount of enzyme that will incorporate 1 nmol ATP (free 1 nmol of pyrophosphate) into acid-insoluble material in 3 hours at 37 °C.

- **Protein Purification of Insoluble RNA polymerase**

Unfortunately, even though the enzyme was initially successfully obtained in functional, soluble form (Figure 11 and 20). Due to unknown issues, we started to encounter difficulties during the recombinant expression even at small scale using the same expression conditions previously optimized, and the enzyme has been expressed mainly in the insoluble fraction as inclusion bodies. For this reason, several growth conditions were selected for further testing, aiming to obtain the soluble protein by means of small-scale (50 ml) culture optimization (Table 15).

Table 15. Selected growth conditions for small scale 50 ml cultures for soluble expression testing.

1	TBA x 23°C x 26 h x 150 rpm	**7**	LBA x 23°C x 26 h x 150 rpm	**13**	LB x 33°C x 0.1 mM IPTG x 6 h x 150rpm
2	TBA x 25°C x 26 h x 150 rpm	**8**	LBA x 25°C x 26 h x 150 rpm	**14**	LB x 33°C x 0.5 mM IPTG x 6 h x 150rpm
3	TBA x 27°C x 20 h x 150 rpm	**9**	LBA x 27°C x 20 h x 150 rpm	**15**	LB x 25°C x 0.1 mM IPTG x 6 h x 150rpm
4	TBA x 30°C x 20 h x 150 rpm	**10**	LBA x 30°C x 20 h x 150 rpm	**16**	LB x 25°C x 0.5 mM IPTG x 6 h x 150rpm
5	TBA x 33°C x 18 h x 150 rpm	**11**	LBA x 33°C x 18 h x 150 rpm	**17**	LB x 20°C x 0.1 mM IPTG x 6 h x 150rpm
6	TBA x 37°C x 18 h x 150rpm	**12**	LBA x 37°C x 18 h x 150 rpm		

The results of the recombinant expression obtained from each of these 17 different tests were visualized by SDS-PAGE gels (Figure 21).

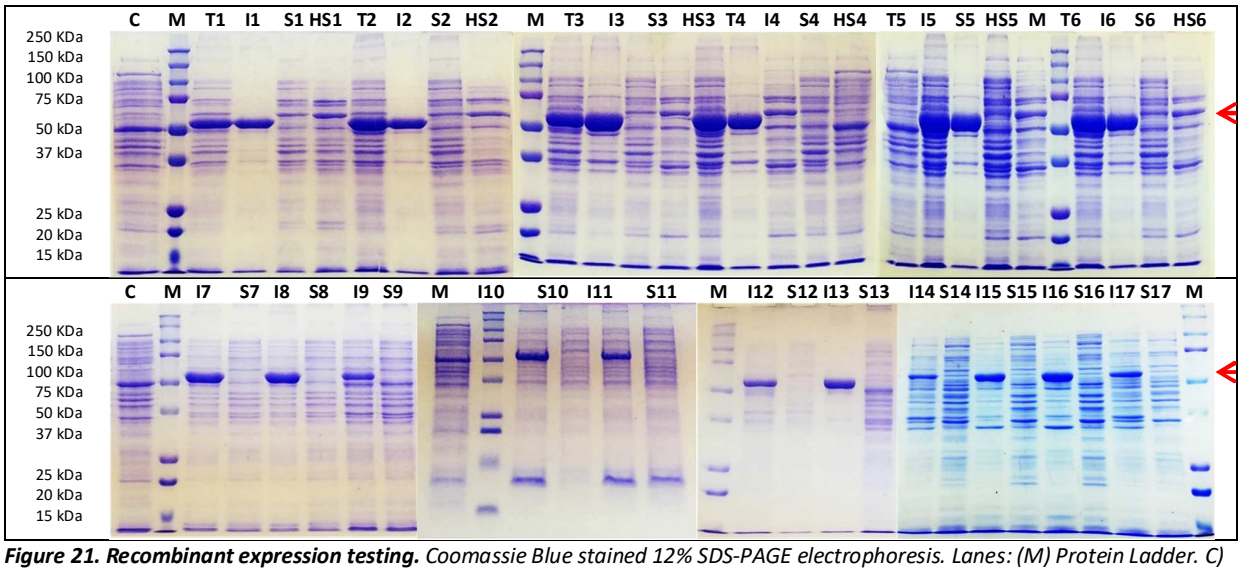


Figure 21. Recombinant expression testing. *Coomassie Blue stained 12% SDS-PAGE electrophoresis. Lanes: (M) Protein Ladder. C) Control (E. coli BL21 cells with no vector). T) Total fraction. I) Insoluble Fraction. S) Soluble Fraction. HS) Heat Shock 60°C x 10 min Soluble Fraction. The 17 tested conditions are detailed in table 1. The protein concentration loaded is not normalized, the recombinant RNA polymerase is highlighted with an arrow.*

Unfortunately, in spite of the several attempts, the recombinant expression of the RNA polymerase remains insoluble. For this reason, and considering that from the insoluble fraction could be possible to obtain a higher and purer amount of recombinant protein, it was decided to perform the protein purification from the inclusion bodies. In order to do this, it was necessary to test the ability of the insoluble enzyme to get solubilized and refolded.

A novel protocol was developed, and inclusion bodies were first mixed with urea 8M and then further solubilized using a water bath sonicator for 5 min. In order to remove the urea, the solution was filtered through 0.22 um and 10 ml solution were diluted 10X, and concentrated to 1 ml using an Amicon concentrator with a 30 kDa membrane. Then, it was further diluted 100X and concentrated again, this process (dilution and concentration) was repeated 3 times to ensure buffer exchange and removal of the excess urea. Protein purification was done with 1 mL of the resolubilized and washed insoluble fraction, which was loaded onto a Sephacryl S-200 column for size exclusion chromatography. The protein was eluted using buffer B1 (0.01 M Tris HCl, pH 8.0, 1 mM EDTA, 0.1 mM DTT, 15% v/v glycerol, 0.05M KCl) at 1 ml/min flow rate using a Pharmacia FPLC. The eluted fractions were run in a 12% SDS-PAGE gel (Figure 22).

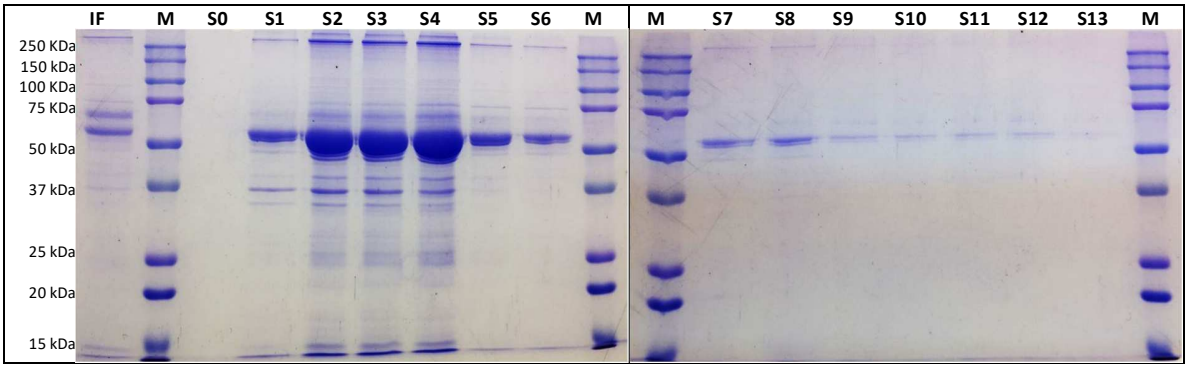


Figure 22. Purification of insoluble RNA polymerase using Size Exclusion Chromatography. *Coomassie Blue stained 12% SDS-PAGE electrophoresis. Lanes: (M) Protein Ladder (Precision Plus Protein Dual Color Standard Biorad). IF) Resolubilized Insoluble Fraction. S0-S13) Fractions eluted from the Sephacryl column. The protein concentration loaded in each well is 15 µg (with the exception of the insoluble fraction), the recombinant RNA polymerase is highlighted with an arrow.*

The eluted fractions with sufficient amount of proteins were used to assess the refold capability of the enzyme by measuring its functionality through the RNA polymerase activity assay using promoters P2 and P3 (Table 16 and 17).

Table 16. RNA polymerase activity. Enzyme assays were performed using 5 ul P2 promoter constructs as templates, 1 µl sample at 37°C x 3 h.

	P2					
	Total Phosphate (nmol)	**Total Pyrophosphate (nmol)**	**Protein Concentration (mg/ml)**	**Sample Volume (ml)**	**Total Protein (mg)**	**Specific Activity U/mg**
IF	13722.6	**6861.3**	33	1	33	**205**
S1	2919.7	**1459.9**	2.1	3.8	7.98	**695**
S2	4014.6	**2007.3**	5.2	1.8	9.36	**386**
S3	6277.4	**3138.7**	9.1	1.5	13.65	**345**
S4	5766.4	**2883.2**	6.2	1.6	9.92	**465**
S5	5328.5	**2664.2**	5.4	2.0	10.8	**493**
S6	4744.5	**2372.3**	2	2.0	4	**1186**
S7	3941.6	**1970.8**	1.3	2.1	2.73	**1516**
S9	6350.4	**3175.2**	1.3	2.3	2.99	**2442**
S10	10365.0	**5182.5**	1.3	2.3	2.99	**3987**

*One unit is defined as the amount of enzyme that will incorporate 1 nmol ATP (free 1 nmol of pyrophosphate) into acid-insoluble material in 3 hours at 37 °C.

Table 17. RNA polymerase activity. Enzyme assays were performed using 5 µl P3 promoter constructs as templates, 1 µl sample at 37°C x 3 h.

	P3					
	Total Phosphate (nmol)	**Total Pyrophosphate (nmol)**	**Protein Concentration (mg/ml)**	**Sample Volume (ml)**	**Total Protein (mg)**	**Specific Activity (U/mg)**
IF	14160.6	**7080.3**	33	1	33	**215**
S1	9562.0	**4781.0**	2.1	3.8	7.98	**2277**
S2	8832.1	**4416.1**	5.2	1.8	9.36	**849**
S3	8467.2	**4233.6**	9.1	1.5	13.65	**465**
S4	11532.8	**5766.4**	6.2	1.6	9.92	**930**
S5	12408.8	**6204.4**	5.4	2.0	10.8	**1149**
S6	10219.0	**5109.5**	2	2.0	4	**2555**
S7	10948.9	**5474.5**	1.3	2.1	2.73	**4211**
S9	36715.3	**18357.7**	1.3	2.3	2.99	**14121**
S10	13722.6	**6861.3**	1.3	2.3	2.99	**5278**

*One unit is defined as the amount of enzyme that will incorporate 1 nmol ATP (free 1 nmol of pyrophosphate) into acid-insoluble material in 3 hours at 37 °C.

Surprisingly, the refolded RNA polymerase was functional even after being expressed as inclusion bodies, and after treatment of the insoluble fraction with 8M urea, sonication and several dilution and concentration for buffer exchange. This finding emphasizes the robustness of extremophilic enzymes and their great potential.

A second purification step of size exclusion chromatography using a Sephacryl S-200 column was done, however, it was considered not necessary, as the level of purity obtained does not compensate for the activity loss.

The purification of the insoluble protein from the host inclusion bodies was repeated, and the refolded RNA polymerase was successfully purified using size exclusion chromatography following the same protocol described above, with few exceptions. The insoluble protein was solubilized using 8M urea. Then, the obtained solution was further solubilized using a water bath sonicator for 15 min and filtered. In order to remove the urea, 10 ml solution were diluted 35X, and concentrated to 1 ml using an ultrafiltration Amicon cell with a 30 kDa membrane to ensure buffer exchange and removal of the excess urea. Then, 1 mL of the resolubilized inclusion bodies was loaded onto a Sephacryl S-200 column for size exclusion chromatography. The protein was eluted using buffer B1 at 1 ml/min flow rate using a Pharmacia FPLC. The eluted fractions were run in a 12% SDS-PAGE gel (Figure 23).

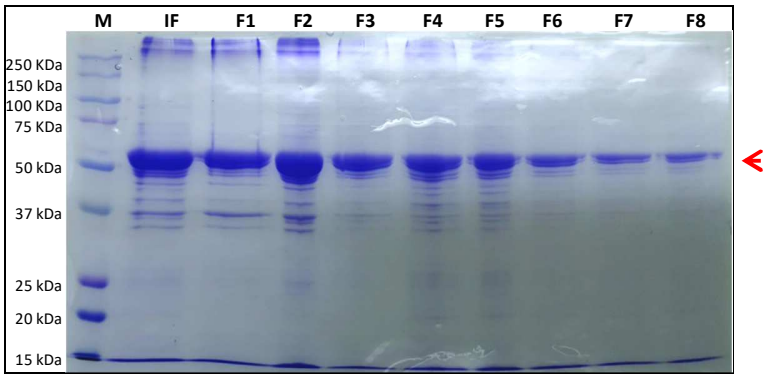


Figure 23. Purification of insoluble RNA polymerase using Size Exclusion Chromatography. *Coomassie Blue stained 12% SDS-PAGE electrophoresis. Lanes: (M) Protein Ladder (Precision Plus Protein Dual Color Standard Biorad). IF) Resolubilized Insoluble Fraction. F1-F8) Fractions eluted from the column. The protein concentration loaded in each well is 15 µg (with the exception of the insoluble fraction), the recombinant RNA polymerase is highlighted with an arrow.*

The eluted fractions with sufficient amount of proteins were used to assess the refold capability of the enzyme by measuring its functionality through the RNA polymerase activity assay using promoter construct P3 (Table 18).

Table 18. RNA polymerase activity. Enzyme assays were performed using 5 μl P3 promoter constructs as templates, 1 µl sample at 37°C x 3 h.

	Total Phosphate (nmol)	Total Pyrophosphate (nmol)	Protein Concentration (mg/ml)	Sample Volume (ml)	Total Proteins	Specific Activity (U/mg)
IF	18978.1	**9489.1**	9.3	1	9.3	**1020**
F1	5474.5	**2737.2**	0.6	4.1	2.46	**4562**
F2	9343.1	**4671.5**	4.6	3	13.8	**1016**
F3	12335.8	**6167.9**	3.3	4.2	13.86	**1869**
F4	10510.9	**5255.5**	1.4	3.3	4.62	**3754**
F5	10365.0	**5182.5**	0.7	2	1.4	**7404**
F6	8540.1	**4270.1**	0.7	2.9	2.03	**6100**

*One unit is defined as the amount of enzyme that will incorporate 1 nmol ATP (free 1 nmol of pyrophosphate) into acid-insoluble material in 3 hours at 37 °C.

The results obtained from this biological replicate were very similar to the ones previously reported, indicating reproducibility of the experiments. Fraction 3 (Figure 23, Lane 4) was selected based on its level of purity, activity and protein concentration for further characterization assays.

- **<u>Characterization of the solubilized RNA polymerase from the inclusion bodies.</u>**

Even though the recombinant enzyme was not expressed in soluble form, it was possible to obtain functional protein from the inclusion bodies in an active form, as they were re-solubilized and purified using size exclusion chromatography.

The sample selected for enzyme characterization was the eluted S2, S3 and S4 from the first size exclusion purification pooled together and concentrated (Fig 24b) as it has a high protein concentration, level of purity, and activity after solubilization.

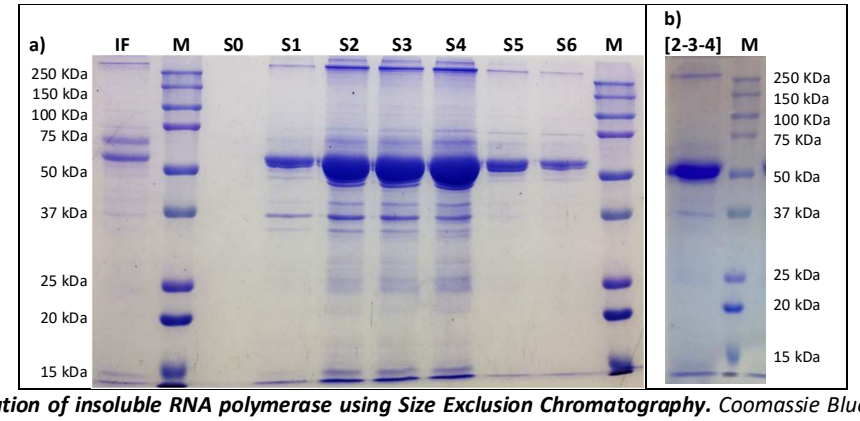


Figure 24. Purification of insoluble RNA polymerase using Size Exclusion Chromatography. *Coomassie Blue stained 12% SDS-PAGE electrophoresis. Lanes: (M) Protein Ladder (Precision Plus Protein Dual Color Standard Biorad). a) IF) Re-solubilized Insoluble Fraction. Fractions eluted from the column. b) Fractions S2, S3 and S4 pooled together and concentrated. The protein concentration loaded in each well is 15 µg (with the exception of the insoluble fraction)), the recombinant RNA polymerase is highlighted with an arrow.*

The effect of temperature was assessed in the range from 5 - 60 °C and the assay was done using the RNA polymerase buffer described for T7 RNA polymerase at pH 7.9 (Figure 25).

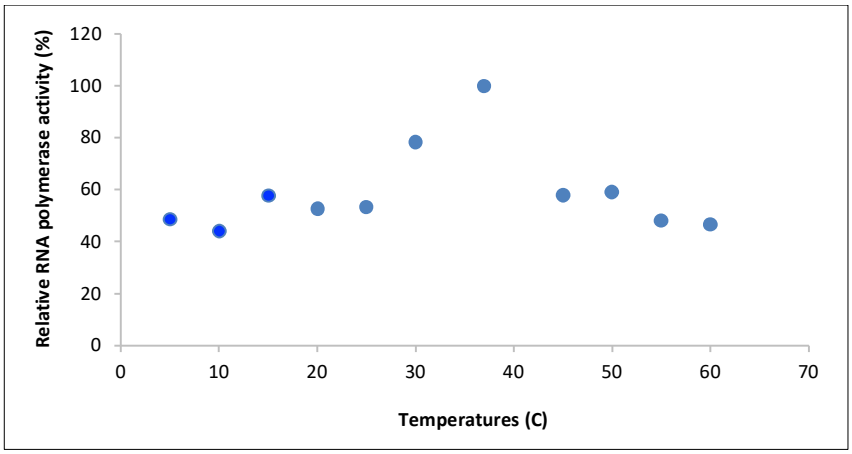


Figure 25. Influence of temperature on the catalytic activity of the resolubilized RNA polymerase. *Activity assays were performed using 5 µl P3 promoter constructs as template and 1 µl [2,3,4] sample at different temperatures (5°C, 10°C, 15°C, 20°C, 25°C, 37°C,45°C, 50°C, 55°C, 60°C), pH 7.9. 100% is the highest activity obtained at 37°C.*

The effect of pH was assessed in the range from 4.0 - 10.0., and the assay was done at 37 °C (Figure 26).

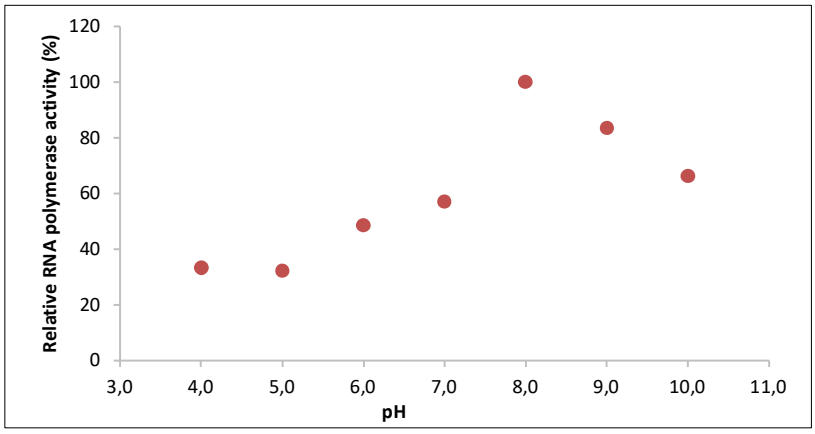


Figure 26. Influence of temperature and pH on the catalytic activity of the re-solubilized RNA polymerase. *Activity assays were performed using 5 ul P3 promoter constructs as template and 1 ul [2,3,4] sample at different pH (4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0) at 37 °C.*

These results indicate that the enzyme is robust, as it has been solubilized and refolded from the inclusion bodies, and is thermoactive in spite of being isolated from Antarctic samples, and seems to be active in the whole range of temperatures tested (5-60°C), with optimum activity at 37°C (Figure 25). Assessment of the effect of pH in the solubilized RNA polymerase activity (Figure 26) showed optimal pH 8.0.

Thermostability assays were performed incubating the enzyme at each temperature (5°C, 10°C, 15°C, 20°C, 25°C, 37°C) for 3 h and measuring the residual activity at 37°C (Figure 27).

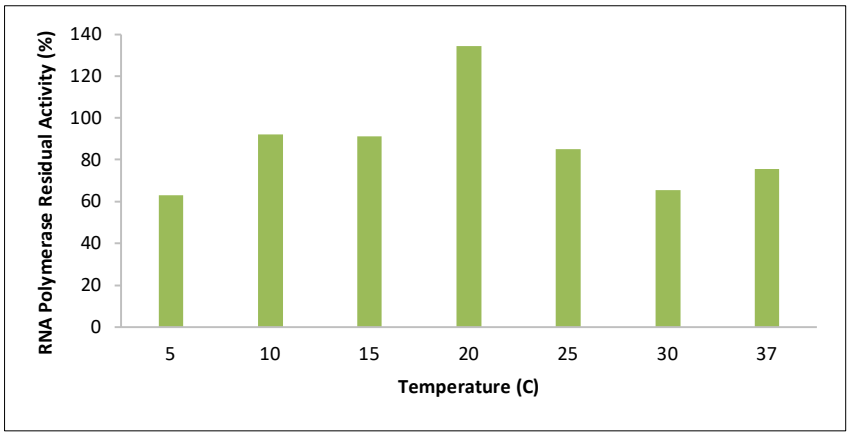


Figure 27. Residual RNA polymerase activity after 3 h incubation at different temperatures. *Activity assays were performed using 5 µl P3 promoter constructs as template and 1 µl F3 (Fig 3, lane 4) sample incubated for 3 h at different temperatures (5°C, 10°C, 15°C, 20°C, 25°C, 37°C), pH 7.9. 100% is the activity obtained without incubation.*

These results indicate that the enzyme maintains over 60% activity after 3 h incubation at the different temperatures tested. In the case of 20°C, the incubation seems to have activated the enzyme. However this experiment will need to be repeated in order to check reproducibility of these results and discard experimental error. Also, it will be interesting to assess the stability of the enzyme at different temperatures and times of incubation.

Stability of recombinant RNA polymerase stored at -20°C

Fraction F3 (Figure 23, lane 4), eluted from the second purification from insoluble inclusion bodies was selected based on its level of purity, activity and protein concentration for stability tests. Aliquots of F3 sample have been stored at -20°C and were measured at different times (Table 4).

Table 19. Stability of recombinant RNA polymerase stored at -20°C. Enzyme assays were performed using 5 ul P3 promoter constructs as templates, 1 µl sample at 37°C x 3 h.

F3 sample	Total Phosphate (nmol)	Total Pyrophosphate (nmol)	Specific Activity (U/mg) *
t= 0 (14/10)	12335.8	6167.9	**1869.1**
t= 26 days (09/11)	11386.9	5693.4	**1725.3**
t= 43 days (26/11)	11970.8	5985.4	**1813.8**

*One unit is defined as the amount of enzyme that will incorporate 1 nmol ATP (free 1 nmol of pyrophosphate) into acid-insoluble material in 3 hours at 37 °C.

These results emphasizes once again the robustness of extremophilic enzymes and their great biotechnological potential, as the recombinant RNA polymerase purified from the inclusion bodies is able to resist 8M urea solubilization, sonication, buffer exchange washes, and at least 6 weeks storage at -20°C maintaining its activity with no further addition of cryoprotectants.

- **Comparison with T7 RNA polymerase**

Samples from different insoluble RNA polymerase purifications were tested along with commercial T7 RNA polymerase (New England Biolabs) using the constructs with Promoter T7 (P0) and Promoter 3 (Table 20) The activity is expressed as volumetric (U/ml) as the manufacturer does not specify protein concentration for the commercial T7 RNApol.

Table 20. Comparison of the recombinant RNA polymerase with commercial T7 (NEB). Enzyme assays were performed using 5 µl DNA templates (promoter constructs T7(P0) and P3), 1 µl sample at 37°C x 3 h.

		PT7 (P0)		P3	
		Total Phosphate (nmol)	Total Pyrophosphate (nmol)	Total Phosphate (nmol)	Total Pyrophosphate (nmol)
T7 RNApol	old enzyme	146.0	**73.0**	124.1	**62.0**
		773.7	**386.9**	700.7	**350.4**
	new enzyme	7591.2	**3795.6**	6812.7	**3406.3**
		6350.4	**3175.2**	8102.2	**4051.1**
Recombinant RNApol	Fraction F3	11386.9	**5693.5**	13211.7	**6605.9**
	Fraction S2	8029.2	**4014.6**	8077.9	**4038.9**
	Fraction S3	8905.1	**4452.6**	12798.1	**6399.0**

These results indicate that the recombinant RNA polymerase developed during this project seems to possess higher activity than the commercial T7 RNApol enzyme, even when freshly opened new package of enzyme was used for the assay.

Conclusions

- Four different DNA samples were used for metagenomic sequencing for mining of viral DNA-directed RNA polymerases.
- Comprehensive bioinformatic analysis of GGM3 consortium metagenome assembly allows to identify the highest amount of sequences annotated as polymerases, of which 29 correspond to DNA-directed RNA polymerases (2.7.7.6).
- Only one was specifically annotated as “Phage DNA-directed RNA polymerase”, suggesting that it could be a single subunit phage-type polymerase, with a theoretical molecular weight of 59.59 kDa.
- This RNA polymerase sequence is 100% identical to the single subunit T3/T7 RNA polymerase from *Strenotrophomonas maltophila* (WP_164076363.1)
- The Phage RNA polymerase found shares 44% identity with a DNA-dependent RNA polymerase from *Rhizobium phage* RHEph01.
- Ten different potential promoter sequences were identified in the contig bearing the sequence.
- The RNA polymerase sequence found in GGM3 metagenome has been successfully cloned and expressed in *E. coli* BL21 cells.
- The best conditions for the successful expression of the recombinant RNA polymerase in soluble and active form were found after optimization trials (trial #3: TBA autoinduction medium, 37°C x 18 h, 150 rev min^{-1}). Heat shock at 60 °C for 10 min proved to be efficient as a first purification step.
- Soluble recombinant RNA polymerase was found to be active after 2 weeks storage in solution at 4 °C.
- 10 different constructs were designed to experimentally test the putative promoter sequences predicted by bioinformatic analysis, this includes the synthesis of the promoters along with a ribosome-binding site and GFP reporter gene cloned in the cloning vector pUC57-Mini.

- The recombinant RNA polymerase, identified from the bioinformatics analysis of GGM3 metagenome, seems to be Not preferentially selective towards one specific promoter sequence, indicating the versatility of this enzyme.
- This result was contrary to what was expected, as our aim was to identify The specific promoter. However, four different and independent results, plus the results obtained by Dr. Svetlana Harbaugh at AFRL indicate that the enzyme can recognize more than one putative promoter sequences for its functioning.
- There was an unknown problem with the scaling-up of the cultures and soluble recombinant expression of the RNA polymerase. To date, in spite of several trials, it has not been possible to re obtain the recombinant RNA polymerase expressed in soluble functional form.
- In order to continue with the experimental work, the recombinant protein obtained in the insoluble inclusion bodies was successfully re-solubilized using 8M urea and sonication. After buffer exchange for removal of the urea, the protein was found to be active indicating proper refold.
- A protocol for solubilization and refold of insoluble RNA polymerase was successfully developed and implemented and protein purification from the inclusion bodies using size exclusion chromatography has been standardized and showed to be reproducible.
- The results obtained emphasizes the robustness of extremophilic enzymes and their great biotechnological potential as the recombinant RNA polymerase purified from the inclusion bodies was able to resist 8M urea solubilization, sonication and several buffer exchange washes.
- Enzyme characterization shows that the recombinant enzyme is thermoactive in spite of its psychrophilic origin, and active in a wide range of temperatures (5°C, 10°C, 15°C, 20°C, 25°C, 30°C, 37°C,45°C, 50°C, 55°C, 60°C).
- The effect of temperature and pH was assessed and the optimum values were found to be 37°C, pH 8.0.
- The enzyme maintains over 60% activity after 3 h incubation at the different temperatures tested (5°C, 10°C, 15°C, 20°C, 25°C, 30°C, 37°C).
- A purified fraction from the inclusion bodies has kept its activity after 6 weeks stored at -20°C with no further addition of cryoprotectants.
- Results indicate that catalytically our RNA polymerase is superior and possess higher activity than commercial T7 RNA polymerase (New England Biolabs) using different DNA templates: promoter construct P3, and even T7 promoter.

a) Next Steps

- To evaluate the patentability of the novel RNA polymerase developed during this project.

b) Collaboration

In the frame of this project, we submitted monthly reports of the work we experimentally developed at Fundación Biociencia (information presented in this final report) and had monthly conference calls with scientists from the 711th Human Performance Wing, Air Force Research Laboratory, Wright Patterson Air Force Base:

Dr. Nancy Kelley-Loughnane,
Dr. Michael Goodson,
Dr. Jorge Chavez Benavides,
Dr. Svetlana Harbaugh,
Dr. Irina Drachuk.

c) Publications.

No publication has been made yet in order to protect possible patentable findings.

d) Personnel Involved in this work

Principal investigator: Dr Jenny Blamey.
Co-Researcher, Post-Doc.: Dr. Giannina Espina. Dr. Sebastian Márquez.
Professionals: Sebastián Muñoz, Litsy Martinez.

e) Presentations at meetings.

No presentation has been made in order to protect possible patentable findings.

f) References

- Davis, James J, Alice R Wattam, Ramy K Aziz, Thomas Brettin, Ralph Butler, Rory M Butler, Philippe Chlenski, et al. 2019. "The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities". *Nucleic Acids Research* 48 (D1): D606–D612.
- Klucar, Lubos, Matej Stano, and Matus Hajduk. 2010. "PhiSITE: Database of Gene Regulation in Bacteriophages". *Nucleic Acids Research* 38: D366–D370.
- Lee, B. H., Park, S. Y., Heo, Y. S., Yea, S. S., Kim, D. E. (2009). Efficient colorimetric assay of RNA polymerase activity using inorganic pyrophosphatase and ammonium molybdate. Bulletin of the Korean Chemical Society, 30(10), 2485-2488.
- Mohsen, M. G., & Kool, E. T. (2016). The discovery of rolling circle amplification and rolling circle transcription. Accounts of chemical research, 49(11), 2540-2550.
- Nakano H., Wada Y., Hasobe M., . Chinook Salmon Embryo (CHSE-sp) Cells for Fish Virus Research In Vitro Cellular & Developmental Biology. Animal. Vol. 29A, No. 4 (Apr., 1993), pp. 265-267.
- Summer H, Grämer R, Dröge P. (2009) Denaturing urea polyacrylamide gel electrophoresis (Urea PAGE). J Vis Exp.;(32). pii: 1485.
- Wang, H., Li, J., & Jewett, M. C. (2018). Development of a Pseudomonas putida cell-free protein synthesis platform for rapid screening of gene regulatory elements. Synthetic Biology, 3(1), ysy003.