

American University in Cairo

AUC Knowledge Fountain

Theses and Dissertations

Student Research

6-10-2019

Identification, expression and biochemical characterization of AMP phosphorylases from extreme environments

Noha Attallah

The American University in Cairo

Follow this and additional works at: <https://fount.aucegypt.edu/etds>

Recommended Citation

APA Citation

Attallah, N. (2019). *Identification, expression and biochemical characterization of AMP phosphorylases from extreme environments* [Master's Thesis, the American University in Cairo]. AUC Knowledge Fountain. <https://fount.aucegypt.edu/etds/1701>

MLA Citation

Attallah, Noha. *Identification, expression and biochemical characterization of AMP phosphorylases from extreme environments*. 2019. American University in Cairo, Master's Thesis. *AUC Knowledge Fountain*. <https://fount.aucegypt.edu/etds/1701>

This Master's Thesis is brought to you for free and open access by the Student Research at AUC Knowledge Fountain. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of AUC Knowledge Fountain. For more information, please contact thesisadmin@aucegypt.edu.



The American University in Cairo
School of Sciences and Engineering

**Identification, expression and biochemical
characterization of AMP phosphorylases from extreme
environments**

A Thesis Submitted to

The Graduate program of Biotechnology

in partial fulfillment of the requirements for
the degree of Master of Science

by Noha Ahmed Anwar Attallah

under the supervision of

Prof. Rania Siam
Prof. Peter Neubauer
Dr. Anke Wagner

May 2019

The American University in Cairo
School of Sciences and Engineering

Identification, expression and biochemical characterization of AMP phosphorylases from extreme environments

A Thesis Submitted by

Noha Ahmed Anwar Attallah

To the Biotechnology Graduate program
in partial fulfillment of the requirements for
the degree of Master of Science

May 2019

Has been approved by

Dr. Rania Siam _____

Thesis Committee Supervisor

Affiliation: Professor of Microbiology, Biology Department, The American university in Cairo

Dr. Peter Neubauer _____

Thesis Committee Co-Supervisor

Affiliation: Professor and Head, Department of Bioprocess Engineering, Institute of Biotechnology,
Technische Universität Berlin

Dr. Anke Wagner _____

Thesis Committee Co-Supervisor

Affiliation: Senior Scientist and Group leader, Department of Bioprocess Engineering, Institute of
Biotechnology, Technische Universität Berlin

Dr. Hassan Azzazy _____

Thesis Committee Reader/ Examiner

Affiliation: Professor and Chair, Department of Chemistry, The American university in Cairo

Dr. Ramy Karam Aziz _____

Thesis Committee Reader/ Examiner

Affiliation: Professor and Chair, Department of Microbiology and Immunology, Faculty of Pharmacy,
Cairo University

Dr. Andreas Kakaroukas _____

Thesis Committee Reader/ Examiner

Affiliation: Assistant Professor of Cell and molecular biology, The American university in Cairo

Dept. Chair/Director

Date

Dean

Date

DEDICATION

I dedicate this thesis

To my wonderful parents for their constant support, encouragement and unconditional love.

To my brother Mohamed who is always beside me and have faith in me.

To my Niece Layan and Nephew Yahia who bring laughter and happiness to my life.

To the soul of my amazing brother Mahmoud, who always believed in me, although no longer with us, my love to you will never pass away, I still remember your lovely and smiley face and you continue to inspire me every day in my life, rest in peace my beloved brother and soulmate until we meet again.

ACKNOWLEDGMENTS

Each and every day I am grateful to my professors, colleagues, family, and friends for their outstanding encouragement and support they provide to me, this research would not have been accomplished without their presence. I would like to express my deepest gratitude to my supervisor **Prof. Rania Siam** for accepting me in her team and giving me the opportunity to gain my master's degree under her supervision, guidance, and support, especially during my hard times.

In addition, thank you is a very small word to express my sincere appreciation to **Dr. Anke Wagner** for her valuable and continuous guidance, advice, suggestions, and discussions that greatly helped me personally in life and professionally through the period of my thesis work.

My sincere gratitude is also extended to **Prof. Peter Neubauer** for hosting me in his lab and for the valuable comments he provided me through my thesis work.

Special thanks and a lot of appreciation go to **Maryke Fehlau**, a Ph.D. student who supervised and helped me a lot during my thesis work. Thank you for the valuable remarks, for answering all my questions, for the support and encouragement you provided me. My research work wouldn't have been done unless your help and support.

I also owe **Amged Ouf**, a sincere and special thanks not only being our genomic specialist who helped me a lot professionally and continuously advise me but also as a friend who was always beside me in happy and hard times. Thanks to **Sarah Kamel** for supporting my thesis work.

Furthermore, I would like to acknowledge all the professors who taught me in the Biotechnology program: **Dr. Asma Amleh, Dr. Hamza El Dorry, Dr. Walid Fouad, Dr. Ahmed El Sayed, and Dr. Ahmed Moustafa.**

Special thanks to **Ahmed el Hosseiny** and **Mr. Osama Ali** for validating my bioinformatics results and for continuous advice in the Bioinformatics part.

Furthermore, my gratitude extends to all lab-mates and colleagues at the BioNukleo team in the Technical University in Berlin and at the Biotechnology graduate program in the American University in Cairo for support, remarks and for providing an inspiring environment through my study and thesis work.

Special thanks to my friends, I consider myself a lucky person surrounded by a lot of friends who always believe in me and push me forward to the best. They all mean a lot to me and I would never survive or accomplish the journey without them.

I especially thank my parents and brother who are my blessings in life. I love them so much and I wouldn't be able to do it without their support, sacrifices, unconditional love and care.

Special thanks to my source of happiness ever, my niece Layan and my nephew Yahia, their smile and laughter always makes me happy and give me hope. I would like also to thank my sisters in law, uncles, aunt and cousins for support and care.

To the wonderful soul of my brother Mahmoud, now you are not here to witness what you always encouraged me for, but I hold onto our memories and how you believed in me and support me. I will never forget you and I know that I am blessed having you as my brother because you were the best. Rest in peace until we meet again.

Finally, I would like to thank the American University in Cairo for granting me a laboratory instruction fellowship, University fellowship, study abroad grant and conference grant which supports my study and thesis work. I also thank the European Commission for granting me the Erasmus+ KA 107 student mobility scholarship that supports my mobility period in the Technical University in Berlin.

ABSTRACT

The American University in Cairo

Identification, expression and biochemical characterization of AMP phosphorylases from extreme environments

By Noha Ahmed Anwar Attallah

Under the supervision of Prof. Rania Siam, Prof. Peter Neubauer, and Dr. Anke Wagner

Nucleotide analogs are interesting pharmaceutical intermediates as they represent the active form of nucleoside analog drugs, that are used in the treatment of cancer or viral infections. They are used as precursors in preparation of artificial oligonucleotides for therapeutic or diagnostic use. Enzymes as active biocatalysts offer numerous advantages over traditional chemical processes with respect to high process selectivity and efficiency. Recently, adenosine-5'-monophosphate phosphorylase (AMP-P) was identified to catalyze the reversible phosphorolysis of nucleotides into nucleobase and ribose-1,5-bisphosphate. Therefore, it is an attractive and promising biocatalyst in the synthesis of nucleotides and their analogs. The availability of enzymes with wide substrate spectra is an important prerequisite to produce a variety of modified nucleotides in enzymatic processes. Consequently, interesting AMP-Ps were produced and characterized concerning their substrate spectra. Red sea metagenomic data and sequences of the National Center for Biotechnology Information database were screened for putative AMP-Ps. Phylogenetic analysis was performed to identify interesting candidates. Sixteen AMP-Ps were chosen from different phylogenetic clusters for gene synthesis based on the differences in the active residues, phylogenetic distance, and variability of their isolation extreme environment. AMP-Ps was successfully expressed in *E. coli* and purified. Expression conditions were optimized to reach higher amounts of soluble protein. Activity assays were performed with six substrates; AMP, CMP, GMP, UMP, adenosine, and uridine. For AMP-P of *Thermococcus khodakarensis* (Am01) published data were confirmed. Additionally, Am03 and Am15 showed similar substrate spectra (an activity with AMP, GMP, and UMP) and are putative AMP-P. Their activity increased with increasing temperature which is in good accordance with the temperature optimum of the donor organisms (*Thermosphaera aggregans*, *Thermofilum pendens* respectively). The other identified proteins could be putative pyrimidine/purine nucleoside phosphorylase as they show a phosphorylase activity against adenosine and uridine. Am12 from ATII-LCL is an interesting candidate to be further analyzed as it is derived from an extreme environment, has activity towards adenosine and uridine nucleosides and low activity against AMP. In this study, we confirmed the hypothesis that extreme environments can also provide cytoplasmic enzymes with novel characteristics.

Table of Contents

DEDICATION.....	iii
ACKNOWLEDGMENTS	iv
LIST OF FIGURES	ix
LIST OF TABLES.....	x
LIST OF ABBREVIATIONS.....	xi
Chapter One: Introduction and literature review	1
1.1 Introduction.....	1
1.2 Adenosine 5'-monophosphate phosphorylase.....	1
1.3 Substrate specificity	2
1.4 Classification and structure	3
1.5 Biotechnological applications of modified NMP analogs.....	5
1.6 Synthesis of modified nucleotide analogs.....	7
1.7 Extreme environment and extremophiles as source for novel industrial enzymes	9
1.8 Aim of the study.....	10
Chapter Two: Materials and Methods.....	11
2.1 Sample collection, DNA extraction, and Pyrosequencing.....	11
2.2 Computational Analysis	13
2.2.1 Screening the Red Sea metagenomic datasets for putative pyrimidine nucleoside phosphorylases.....	13
2.2.2 Phylogenetic analysis for Archaeal nucleoside/nucleotide phosphorylases to identify interesting candidates	13
2.3 Synthesis, Cloning & Expression of AMP phosphorylases	15
2.3.1 Synthesis of AMP phosphorylases and Transformation in <i>E. coli</i> competent Top10 cells.....	15
2.3.2 Re-Cloning of AMP phosphorylases in the expression vector pKS2-plasmid.....	17
2.3.3 Expression of recombinant AMP phosphorylases in <i>E. coli</i> BL21 cells using Enpresso®B medium and 2xYT Medium Broth.....	18
2.4 SDS-PAGE analysis.....	19
2.5 Purification of His-tag AMP phosphorylase by Ni-NTA spin column affinity chromatography.....	19

.....	20
2.6 Activity assay of the purified AMP phosphorylases	20
2.7 Analysis of Thermostability	21
2.8 High-Performance Liquid Chromatography (HPLC) analysis.....	22
Chapter Three: Results	22
3.1 Identification of Putative Thymidine /AMP phosphorylase sequences from Red Sea metagenomic dataset and NCBI protein database.....	22
3.2 Phylogenetic analysis of the identified ORFs from the Red Sea	23
3.3 Analysis of AMP-P structure	32
3.4 Biochemical characterization of the putative AMP-P.....	34
3.4.1 Gene synthesis and cloning	34
3.4.2 Protein expression and purification	35
3.5 Activity assays.....	39
3.6 Thermostability of the heterologous expressed AMP-P	41
3.6.1 Activity of heat-treated AMP-P.....	42
Chapter Four: Discussion	43
Chapter Five: Conclusion.....	46
Chapter six: Outlook	48
Appendix.....	49
Supplementary materials	49
Figure licenses.....	53
License for Figure 1, 3 and 4.....	53
License for Figure 5, 6, and 7	54
License for Figure 8.....	55
References	56

LIST OF FIGURES

Figure 1. Schematic illustration of the novel archaeal AMP metabolic pathway and the role of AMP-P in catalyzing the initial step	2
Figure 2. Schematic illustration of AMP-P reversible phosphorolysis reaction against AMP, CMP, UMP and dCMP substrates	2
Figure 3. Structure of AMP-P	4
Figure 4. Scheme for illustration of domain swapping interaction in AMP-P responsible for multimerization (>40mers).	5
Figure 5. Advantage of modified nucleotide analog against modified nucleoside analog.	6
Figure 6. Chemical structure of modified nucleotide analogs.	7
Figure 7. Examples of various modifications either on base, sugar or phosphate group to produce synthetic modified nucleotide analog	8
Figure 8 Enzymatic cascade for the production of purine nucleotide analogs using engineered hypoxanthine phosphoribosyl transferase.	9
Figure 9. Analysis pipeline diagram for searching AMP-P in the Red sea datasets using a reference database of AMP-P retrieved from NCBI.....	15
Figure 10 Cloning (pMA-T) and Expression (pKS2) vectors used in this study.....	16
Figure 11. Molecular phylogenetic analysis of Red Sea metagenomic identified ORFs by Maximum Likelihood method in Mega X.	25
Figure 12. Multiple sequence alignment for the chosen enzymes with the only characterized AMP-P from <i>Tk</i> archaeon.	29
Figure 13. Transmembrane helix prediction by Phyre2 web server using memsat-svm.	33
Figure 14. Test restriction after cloning of the sequences of interest into expression vector pKS2.	34
Figure 15. Expression of AMP-P in EnPresso [®] B medium with IPTG concentrations of 1 mM.....	36
Figure 16. Optimization of expression shown for Am07 as an example.....	37
Figure 17. SDS-PAGE after purification of His-tagged AMP-P by Ni-NTA spin column affinity chromatography.....	38
Figure 18. Relative phosphorylase activity of AMP-P	40
Figure 19. Analysis of heat resistance of the purified AMP-Ps on 12% SDS-PAGE gels.....	41
Figure 20. Relative phosphorylase activity of AMP phosphorylases after heat-treatment.....	42

LIST OF TABLES

Table 1. Metadata for Red Sea Metagenomic samples	11
Table 2. Characteristics of the Red Sea metagenomic sequence data.....	12
Table 3. Extinction coefficient of AMP phosphorylases calculated by Protparam web portal.	20
Table 4. Phosphorylase activity assay for purified proteins after heating at 70°C for 1hr.....	22
Table 5. BlastX and Interpro results for the 11 ORFs identified by HMMsearch against Red sea metagenomic data	24
Table 6. Characteristics of the 16 chosen enzymes for gene synthesis.....	27
Table 7. Size [bp] of the inserts and pKS2.....	35
Table 8. Parameters validated during expression and purification of AMP-P in EnPresso®B medium with 1mM IPTG conc.	39

LIST OF ABBREVIATIONS

3-PGA	3-phosphoglycerate
AMN	AMP nucleosidase
AMP	adenosine 5'-monophosphate
AMP-P	Adenosine 5'-monophosphate phosphorylase
ATII	Atlantis II Deep
BI	Brine influenced
Bp	base pair
BR	Brine
CMP	Cytidine 5'-monophosphate
dCMP	Deoxycytidine 5'-monophosphate
DD	Discovery Deep
DTT	Dithiothreitol
ESBRI	Evaluating the Salt BRIdges in Proteins
GMP	Guanosine 5'-monophosphate
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
HMM	Hidden Markov model
HPLC	High-Performance Liquid Chromatography
IN	Interface
IPTG	Isopropyl β -D-1-thiogalactopyranoside
iTOL	Interactive Tree of Life
KD	Kebrit Deep
LB	Luria-Bertani
LCL	Lower convective layer
LIN	Lower interface
ML	Maximum likelihood
NCBI	National Center for biotechnology information
NMP	Nucleoside 5'- monophosphate
NPT	Nucleoside phosphotransferases
NP-II	Nucleoside phosphorylase II family
nr	non-redundant
NS5B	Nonstructural protein 5B
ORFs	Open reading frames
PhnN	Ribose 1,5-bisphosphate phosphokinase
PMSF	Phenylmethylsulphonyl fluoride
PRPP	Phosphoribosyl pyrophosphate
PyNP	Pyrimidine nucleoside phosphorylase
R15P	Ribose-1,5-bisphosphate
Rubisco	Ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose-1,5-bisphosphate
S	Sediment

SDS- PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SRA	Sequence reads archive
<i>Tk</i>	<i>Thermococcus kodakarensis</i>
TP	Thymidine phosphorylase
UCL	Upper convective layer
UIN	Upper interface
UMP	Uridine 5'-monophosphate

Chapter One: Introduction and literature review

1.1 Introduction

Enzymes as an active biocatalyst are considered interesting especially in the fields of industrial biotechnology; that focuses on the production of small molecules. The traditional chemical process is usually a tedious multistep reaction that render the process, time and cost intensive. Enzymatic synthesis routes are considered a promising alternative to the chemical synthesis with respect to high process selectivity and efficiency. Despite the obvious advantages of enzymatic or chemo-enzymatic production routes, industrial biotechnology is still dominated by chemical production processes. One of the promising fields for the application of enzymatic routes is the production of nucleotides and their modified analogs. The latter are valuable drugs for the treatment of cancer and viral infections.

Recently, adenosine 5'-monophosphate phosphorylase (AMP-P) was identified to catalyze the reversible phosphorolysis of nucleotides into nucleobase and ribose-1,5-bisphosphate (R15P). Therefore, it is an attractive and promising biocatalyst in the synthesis of nucleotides and their analogs, however, the availability of enzymes with wide substrate spectrum is an important prerequisite to produce a wide variety of modified nucleotides in an enzymatic process.

Deep sea extreme environments are interesting habitats for microorganisms with specific characteristics. While it is evident that such habitats are interesting for secreted matter, they also can provide cytoplasmic enzymes with novel characteristics. Accordingly, exploring new industrial enzymes from an extreme environment is a crucial step for better biotechnological applications.

1.2 Adenosine 5'-monophosphate phosphorylase

AMP-P (EC 2.4.2.57) is a unique archaeal nucleotide phosphorylase, first identified in a hyper-thermophilic archaeon called *Thermococcus kodakarensis* (*Tk*) and it has a role in nucleic acid metabolism and regulation of the intracellular nucleotide level (Aono et al., 2012). AMP-P are involved in a novel metabolic pathway (Figure 1) of adenosine 5'-monophosphate (AMP) metabolism. This pathway is composed of three enzymatic reactions. The initial reaction is catalyzed by AMP-P which catalyze the reversible phosphorolysis of nucleotides into nucleobase and R15P. The final step is the carboxylase reaction of ribulose-

1,5-bisphosphate carboxylase/oxygenase (Rubisco; EC 4.1.1.39) which fixes carbon dioxide by converting the ribulose-1,5-bisphosphate (RuBP) into two molecules of 3-phosphoglycerate (3-PGA). This product is an important intermediate in central sugar metabolism (Nishitani et al., 2013; Sato, Atomi, & Imanaka, 2007)

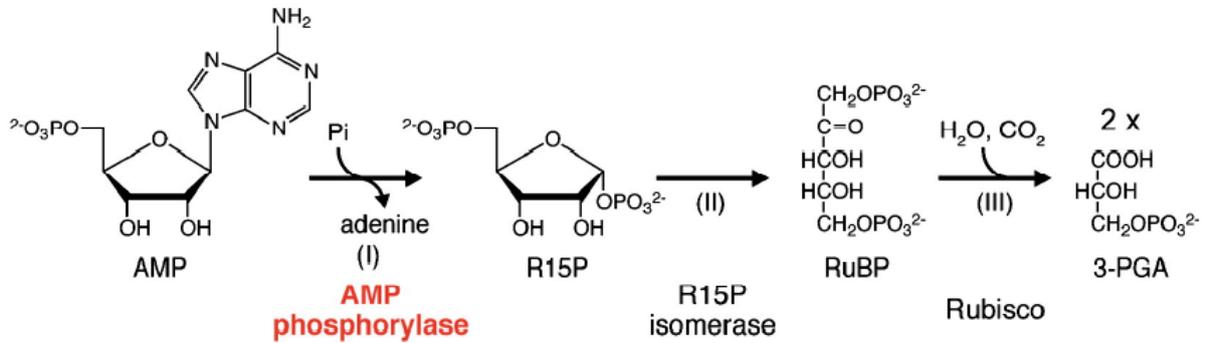


Figure 1. Schematic illustration of the novel archaeal AMP metabolic pathway and the role of AMP-P in catalyzing the initial step (Nishitani et al., 2013).

1.3 Substrate specificity

AMP-P is unique enzyme in its substrate specificity (Figure 2) as it is the first enzyme to be identified to catalyze phosphorolysis reaction on the nucleotide. It can accept not only AMP but cytidine 5'-monophosphate (CMP), uridine 5'-monophosphate (UMP) and deoxycytidine 5'-monophosphate (dCMP)(Aono et al., 2012).

Generally, the AMP phosphorylase reaction in equilibrium favors the reverse direction toward the formation of nucleoside 5'-monophosphate (NMP) but the presence of Rubisco enzyme drives the reaction forwarded toward the formation of 3-PGA (Aono et al., 2012).

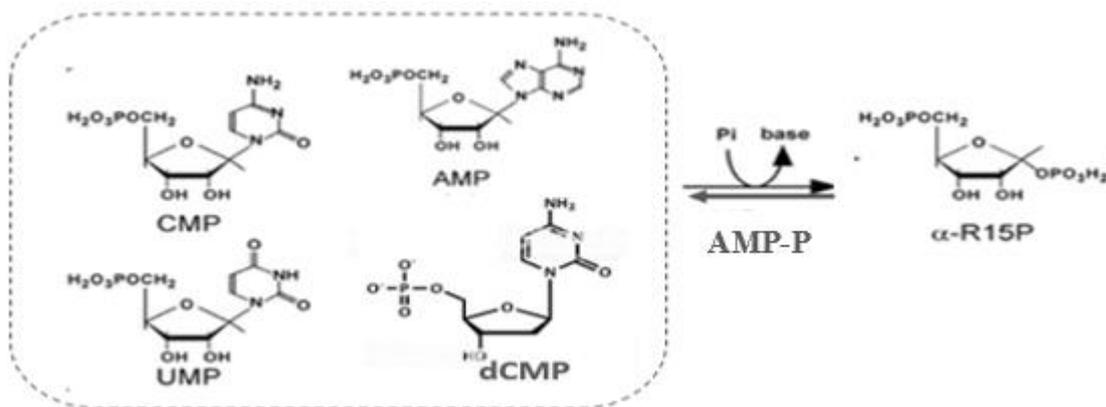


Figure 2. Schematic illustration of AMP-P reversible phosphorolysis reaction against AMP, CMP, UMP and dCMP substrates (Aono et al., 2012)

In addition, AMP-P is widely distributed among archaea (e.g. *Thermococcales*, *Archaeoglobales*, *Methanomicrobiales*, and *Methanosarcinales*).

1.4 Classification and structure

Based on AMP-P primary structure, AMP-P belongs to the nucleoside phosphorylase II family (NP-II), which includes pyrimidine nucleoside phosphorylase (PyNP) and thymidine phosphorylase (TP) (Nishitani et al., 2013). It is the only member of this family to catalyze the phosphorolysis of nucleotides, unlike all other members that act on pyrimidine nucleoside only forming pyrimidine base and ribose -1-phosphate (Pugmire & Ealick, 2002).

Previously, AMP-P was annotated as TP; however, its function and structure are distinct from the NP-II family members (Nishitani et al., 2013). Recently, most AMP-P members were re-annotated from homolog of TP to homolog of AMP-P. AMP-P structure consists of three essential domains (Figure 3a):

- a) **N-terminal domain** (82 residues; 2-84) is unique to AMP-P and not found in other members of NP-II family. It is essential for multimerization as AMP-P form large macromolecules (>40mers) in solution and this multimerization is crucial for the thermostability of the enzyme from a hyperthermophile (Nishitani et al., 2013). Moreover, it is essential to the enzymatic activity of AMP-P as it is stabilizing the closed conformation upon substrate binding unlike other members of NP-II family which has closed loop instead (Figure 3b & 3c) (Pugmire & Ealick, 1998).
- b) **Central domain** (96 residues; 84–149 and 241–272) whose structure is the same as NP-II family members and it contains residues essential for pyrimidine base binding (Glu250 & Lys268); however, it contains also residues essential for binding to purine base Asp256 and Ser264 unlike NP-II family members (Nishitani et al., 2013).

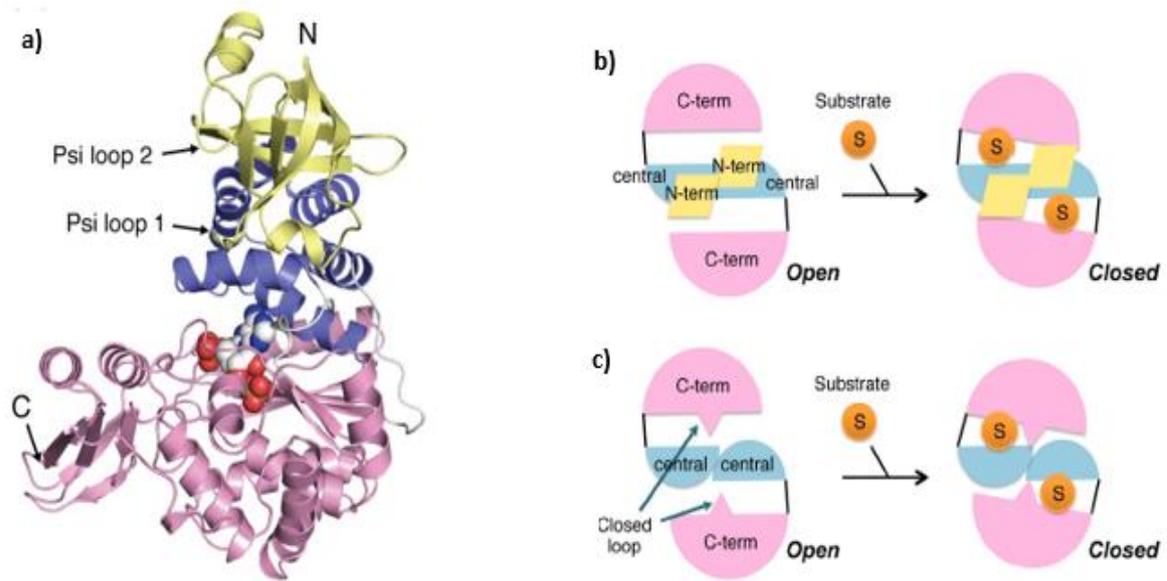


Figure 3. Structure of AMP-P a) AMP-P subunit showing the three essential domains.

The N-terminal, central, and C-terminal domains are shown in yellow, blue, and pink, respectively. b & c) Schematic drawing show open and closed conformation in AMP-P (b) and PyNP (c) (Nishitani et al., 2013)

c) C-terminal domain (299 residues; residues 161–233 and 276–503) its structure is the same as NP-II family members as it contains residues essential for phosphate binding (Ser165, Asn175, Lys191, Ser193, Ser194 & Thr203). In addition, it has residues which are unique to AMP-P called 5'-phosphate binding residues (Gly168, Ile197, Thr198, Ser199 & Lys288) which is responsible for binding to 5'-phosphate of nucleotide. Moreover, AMP-P contains an exceptional β 18 strand (10 residues) in the C-terminal domain which is responsible for unique domain swapping interaction (Figure 4) that contributes to the multimerization of AMP-P (Nishitani et al., 2013), unlike NP-II family members which form dimers (Pugmire & Ealick, 1998).

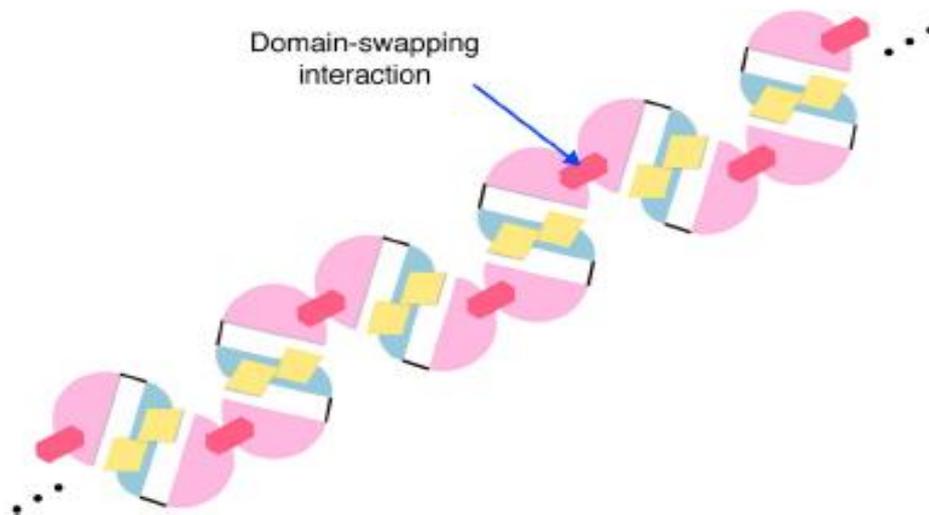


Figure 4. Scheme for illustration of domain swapping interaction in AMP-P responsible for multimerization (>40mers).

The N-terminal, central, and C-terminal domains are shown in yellow, blue, and pink, respectively. The C-terminal 10 residues, which participate in the domain swapping interaction with the neighboring molecule, are shown by red bars (Nishitani et al., 2013)

1.5 Biotechnological applications of modified NMP analogs

The emergence of drug resistance against nucleoside analogs together with their low effectiveness and increased side effects gave rise for new drugs which are prodrug modified NMP analogs (Jordheim, Durantel, Zoulim, & Dumontet, 2013). These modified nucleotide analogs are important as pharmacological ingredients in the pharmaceutical industry as it represents the active form of nucleoside analogs drugs that are used in the treatment of cancer and viral infections.

Furthermore, they have an advantage over the previously known modified nucleoside analogs (Figure 5) as:

- 1) They don't require the initial phosphorylation step which converts nucleoside to the corresponding NMP and this step is considered the rate-limiting step and one of the resistance mechanisms enforced by tumor or viral cells against nucleoside analogs drugs
- 2) They are insensitive to deamination either intracellular or extracellular and therefore increase the activity of the drug.

3) They don't depend on membrane transporters rather they enter the cell independently based on the nature of the chemical moieties that mask the phosphate group (Jordheim et al., 2013)

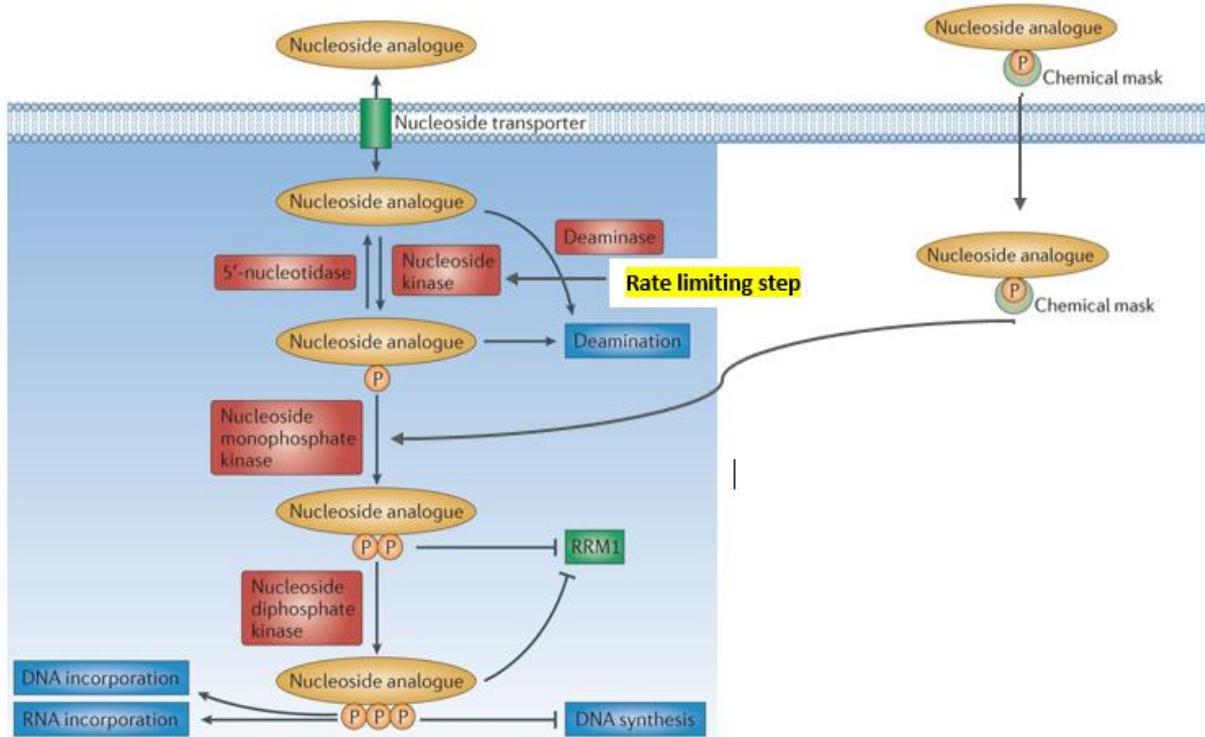


Figure 5. Advantage of modified nucleotide analog against modified nucleoside analog. (Jordheim et al., 2013)

Several NMPs analogs prodrugs are approved by FDA (Clercq & E., 2016) for the treatment of cancer and viral infections for example:

- **Adefovir dipivoxil** (Hepsera); acyclic nucleotide analog of AMP in the treatment of human hepatitis B virus (HBV) (Shah et al., 2015) (Figure 6a)
- **Tenofovir disoproxil fumarate** (Viread); acyclic nucleotide analog of deoxyadenosine 5'-monophosphate belongs to reverse transcriptase inhibitor family and used in the treatment of human immunodeficiency virus (HIV) and HBV (Lyseng-Williamson, Reynolds, & Plosker, 2006) (Figure 6b)
- **Sofosbuvir** (Sovaldi); nucleotide analog of UMP used in the treatment of Hepatitis C virus (HCV) by inhibiting the nonstructural protein 5B (NS5B) polymerase. It is highly potent over

other nucleoside analogs and characterized by low side effects and high resistance against all genotypes of HCV (Bhatia, Singh, Grewal, & Natt, 2014) (Figure 6c)

• **Fludarabine Phosphate** (Fludara); phosphate salt of a fluorinated nucleotide antimetabolite analog of the antiviral agent vidarabine (ara-A) with antineoplastic activity used in the treatment of leukemia and lymphoma by inhibiting DNA synthesis (Boogaerts et al., 2001) (Figure 6d)

• **Stampidine** (experimental); dideoxy thymidine 5'-monophosphate nucleotide analogue which is derivative of stavudine (nucleoside analogue) used in the treatment of multidrug resistance HIV infections and overcome the major problem of stavudine which the first phosphorylation step (rate limiting step) (Uckun, Pendergrass, Venkatachalam, Qazi, & Richman, 2002) (Figure 6e)

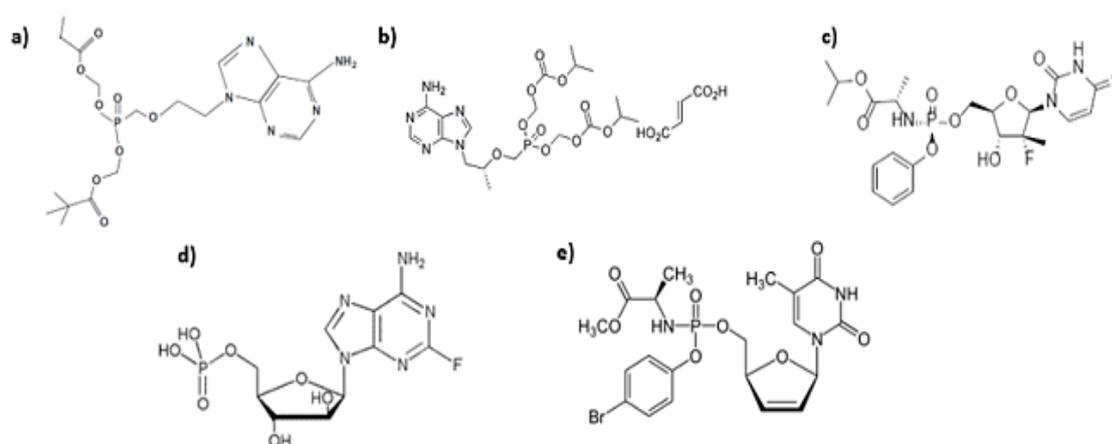


Figure 6. Chemical structure of modified nucleotide analogs.

a) Adefovir dipivoxil, b) Tenofovir disoproxil fumarate, c) Sofosbuvir, d) Fludarabine Phosphate and e) Stampidine

1.6 Synthesis of modified nucleotide analogs

Modified NMPs are synthesized either chemically (Burgess & Cook, 2000) or enzymatically (Barai, Kvach, Zinchenko, & Mikhailopulo, 2004; Caton-Williams, Smith, Carrasco, & Huang, 2011) through phosphorylation of a nucleoside. These modified analogs have various modification either on the sugar or the nucleobase part which renders their chemical synthesis tedious, labor-intensive and time-consuming. (Figure 7)

Moreover, chemical synthesis has different challenges as difficulty in control the regioselectivity of ribose 5'-phosphorylation and production of various byproduct impurities which are unstable and difficult to remove. Also, multiple protection and deprotection of functional groups (e.g. hydroxyl) of nucleoside are required. In addition, every nucleoside analog has different challenges which make a general chemical strategy impossible. All these potentials lead to low yield, difficulty in the purification of water-soluble products and high cost. (Burgess & Cook, 2000; Caton-Williams et al., 2011)

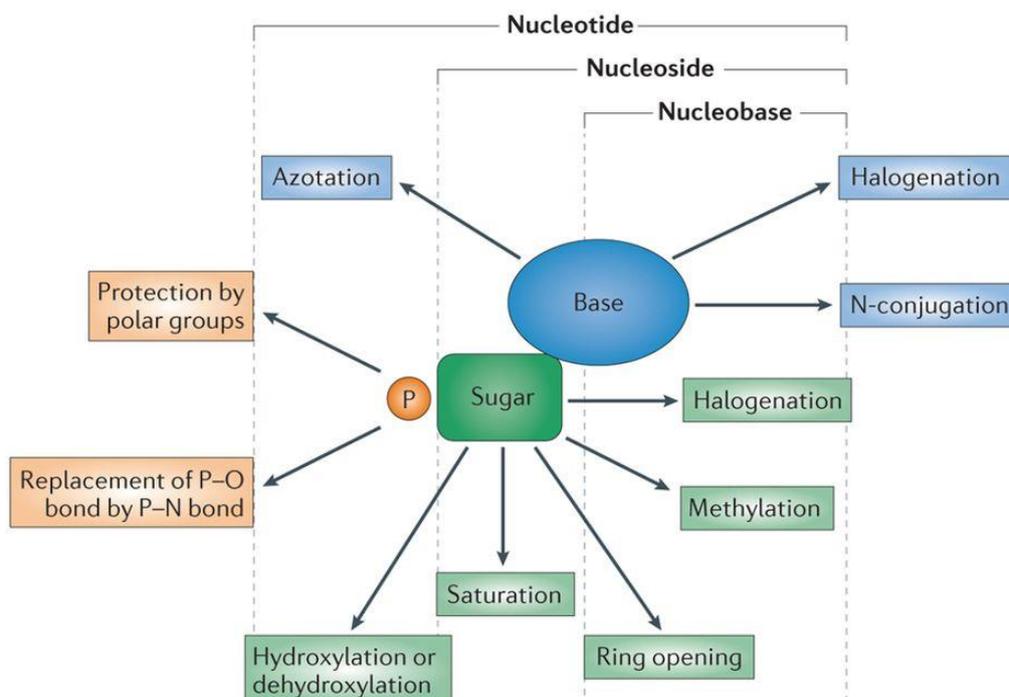


Figure 7. Examples of various modifications either on base, sugar or phosphate group to produce synthetic modified nucleotide analog

Emerging from the importance of the NMPs modified analogs and the difficulty of their chemical synthesis; an enzymatic alternative is of great demand. Several attempts with different enzymes were used such as nucleoside phosphotransferases (EC 2.7.1.77; NPT) that were used to catalyze 5'-mono-phosphorylation of nucleoside (Barai et al., 2004), engineered hypoxanthine phosphoribosyl transferase which has specificity for purine base analogues only; facilitating their addition to phosphoribosyl pyrophosphate (PRPP) to produce monophosphorylated nucleotides. This latter method has some limitations as PRPP has poor solubility, decomposes rapidly in solution and has limited substrate specificity (Scism & Bachmann, 2010). (Figure 8)

One of these environments is the Red Sea brine pools which are characterized by multiple abiotic stressors like high temperature, high salinity, high pressure and decreased oxygen levels (Siam et al., 2012). Potential thermostable enzymes were discovered in such environments as mercuric reductase (Sayed et al., 2014), nitrilase (Sonbol, Ferreira, & Siam, 2016), esterase (Mohamed et al., 2013), antibiotic resistance enzymes (Elbehery, Leak, & Siam, 2017) and anticancer and anti-bacterial orphan biosynthetic gene clusters (Ziko, Ouf, Aziz, & Siam, 2019).

Moreover, extremozymes from extremophile show potentials as an active biocatalyst in catalyzing various chemical conversion that was dominated by traditional chemistry with greater specificity and more stability (L. Kumar, Awasthi, & Singh, 2011). Extremophiles from an archaeal origin provide the majority of extremozymes and characterized by novel metabolic pathways (Aono et al., 2012; Nishitani et al., 2013; Sato et al., 2007) which make it a source for several enzymes with novel biotechnological application.

Owing to the advantages of extremophiles and extreme environment in providing novel enzymes and the industrial demand for active biocatalysts to be used in various biotechnological applications we screened the Red Sea brine pools and searched various extremozymes from archaeal extremophiles for putative AMP phosphorylases enzymes. It would thus be of interest to determine the currently unknown mechanism by which AMP phosphorylase alone recognizes various bases, sugars and 5'-phosphate groups of nucleotides.

1.8 Aim of the study

The aim of the study was to express and characterize new AMP-P. To identify interesting enzyme candidates, phylogenetic analysis was performed with archaeal AMP-P sequences identified either in the Red Sea or in the NCBI database. Sequences were chosen for further studies based on differences in the amino acid sequence or the structure. 16 AMP-P were expressed in *E. coli*, and their substrate spectra were studied.

Chapter Two: Materials and Methods

2.1 Sample collection, DNA extraction, and Pyrosequencing

Water and Core sediments samples were collected during the KAUST/WHOI/HCMR Red Sea expeditions, 2010 onboard the R/V Aegeo in collaboration with AUC. Samples were taken from Atlantis II Deep (**ATII**, Temperature 68°C), Discovery deep (**DD**, Temperature 45°C), Kebrit deep (**KD**, Temperature 23.4°C) and Brine influenced site (**BI**, Temperature 22°C) (Table 1)

Table 1. Metadata for Red Sea Metagenomic samples

Database	Location	Depth (m)	Description	Bio-project Accession no.
ATII50	21°21'N, 38°04'E	50	Sea Water	PRJNA193416
ATII200		200	Sea Water	PRJNA193416
ATII700		700	Sea Water	PRJNA193416
ATII1500		1500	Sea Water	PRJNA193416
ATII-IN		1997	Interface layer	PRJNA219363
ATII-UCL		1995	Brine	PRJNA193416
ATII-LCL		2100	Brine	PRJNA193416
ATII-S		2168	Brine Sediment	PRJNA299097
KD-UIN	24°71'N, 36°28'E	1468	Upper Interface layer	PRJNA219363
KD-LIN		1469	Lower Interface layer	PRJNA219363
KD-BR		1496	Brine	PRJNA193416
DD-IN	21°28'N 38°04'E	2030	Interface layer	PRJNA219363
DD-BR		2030	Brine	PRJNA193416
DD-S		2180	Brine Sediment	PRJNA299097
BI-S	21°40'N 38°09'E	1856	Brine influenced sediment	PRJNA299097

ATII: Atlantis II, **KD:** Kebrit deep, **DD:** Discovery deep, **IN:** interface, **UCL:** upper convective layer, **LCL:** lower convective layer, **S:** Sediment, **UIN:** upper interface, **LIN:** lower interface, **BR:** brine, **BI:** brine influenced

Water samples were collected by Niskin bottles mounted on rosette that is supplied by CTD (Conductivity, Temperature, and Depth) meter and immediately processed through serial filtration using mixed cellulose ester filters (nitrocellulose/ cellulose acetate) of various pore sizes 3.0, 0.8 and 0.1 µm (Millipore). Collected samples stored in sucrose buffer at -20°C for further analysis. Twelve water samples were taken from different sites and depths as follows:

From ATII (50m, 200m, 700m, 1500m, Brine water interface, Upper convective layer and Lower convective layer), from DD (Brine and Brine water interphase), from KD (Brine, lower interface and upper interphase) (Ferreira et al., 2014).

Sediment core samples were collected from ATII and DD by gravity cores between 3.5 and 4 m long (Benthos Instruments, KC Denmark), while from BI by short multi-coring (Siam et al., 2012).

DNA extraction from 0.1 μm filter is done as described (Rusch et al., 2007) and from sediment by PowerSoil[®] DNA Isolation Kit (MO-BIO, Calsbad, CA). The DNA concentration was measured by a Nanodrop 3300 Fluorospectrometer (Thermo Scientific, USA) and a Quant-iT[™] PicoGreen[®] dsDNA Kit (Invitrogen, USA). Shotgun sequencing of the extracted DNA was done on a GS FLX pyrosequencer using the Titanium pyrosequencing kit (454 Life Sciences) creating a dataset of all reads (Table 2). Assembly was done by GS assembler (The GS Data Analysis software package, 454 Life Sciences) forming contigs where ORFs were identified and annotated using Artemis (Rutherford et al., 2000).

Table 2. Characteristics of the Red Sea metagenomic sequence data

Database	Number of reads	Total reads size (bp)	Number of Contigs >500bp	Average contig size (bp)	Largest contig size (bp)
ATII50	1,461,910	588,584,375	36262	1149	21887
ATII200	1,260,578	492,013,769	34640	1131	25392
ATII700	1,128,514	489,155,258	32860	1285	33783
ATII1500	833,739	355,048,289	20374	1331	51927
ATII-IN	832,138	268,756,307	9933	1214	25150
ATII-UCL	886,030	330,511,319	11994	1454	103389
ATII-LCL	4,104,966	1,546,475,362	19165	2084	350936
ATII-S	1,138,406	444,647,496	30352	1194	33674
KD-UIN	1,562,521	470,950,183	24517	1495	58542
KD-LIN	1,510,272	501,870,390	31983	1241	38825
KD-BR	1,379,832	367,495,793	22280	945	14864
DD-IN	1,095,181	374,686,960	14144	1201	28080
DD-BR	1,111,044	393,207,432	15306	1216	22118
DD-S	1,258,290	509,882,594	38529	1233	38081
BI-S	253,568	122,268,823	7292	1177	1315

ATII: Atlantis II, KD: Kebrit deep, DD: Discovery deep, IN: interface, UCL: upper convective layer, LCL: lower convective layer, S: Sediment, UIN: upper interface, LIN: lower interface, BR: brine, BI: brine influenced. A highlight indicates the highest number.

2.2 Computational Analysis

2.2.1 Screening the Red Sea metagenomic datasets for putative pyrimidine nucleoside phosphorylases

A sequence-based metagenomic approach was used to identify putative PyNPs in the Red Sea metagenomic datasets using a probabilistic model called profile hidden Markov models (HMMs)(Eddy, 1998) against Red Sea metagenomic datasets:

- a- Pfam signatures (PF00591.20, PF02885.16, PF07831.12) representing the functional domains of PyNPs; α/β , α , and C-terminal domains respectively were retrieved from Pfam 31.0 (El-Gebali et al., 2019) and used to extract open reading frames (ORFs) containing these domains from the assembled Red sea datasets using HMMsearch.
- b- Another HMM profile was built using HMMbuild from a dataset of PyNPs which contain 119 PyNP enzymes of thermophilic origin that were retrieved from National center for biotechnology information (NCBI). This HMM profile is used to HMMsearch against assembled and annotated Red sea datasets to identify ORFs containing PyNP domains.

ORFs of interest that contain the three functional domains of PyNPs were selected and compared against non-redundant (nr) protein database using BLASTX algorithm of NCBI (Altschul et al., 1997) to identify a possible new PyNP sequences based on different parameters as E-value, query coverage and identity percentage to the nearest BLASTX hit and also against InterPro protein database to identify protein family (Finn et al., 2017).

2.2.2 Phylogenetic analysis for archaeal nucleoside/nucleotide phosphorylases to identify interesting candidates

Multiple sequence alignment of the computationally identified ORFs with customized database (96 sequences, Supplementary Table 1) of AMP-P reference sequences retrieved from NCBI from different phylum and classes of archaeal origin together with PyNP of *Geobacillus stearothermophilus* (most identified PyNP) (Pugmire & Ealick, 1998) and BioNukleo PyNP/TPs (identified by BioNukleo lab in Technische Universität Berlin) was performed using Mcoffee (Wallace, O'Sullivan, Higgins, & Notredame, 2006) (Moretti et al., 2007) which align sequences using multiple sequence aligners as Muscle, Mafft, ClustalW, Probcons and T-coffee and combine their output in one unique alignment.

Phylogenetic analysis was performed by the Maximum likelihood (ML) method and reliability was tested using bootstrapping of 1000 replicates in MEGA X software version 10.0.1 (S. Kumar, Stecher, Li, Knyaz, & Tamura, 2018). This step was done to analyze where the computationally identified proteins from Red Sea datasets will cluster and compare it to the BioNukleo PyNP/TPs. The tree was visualized with bootstrap using the online tool Interactive Tree of Life (iTOL) V4.3.3 (Letunic & Bork, 2019).

The phylogenetic tree was divided into clusters and enzymes were chosen for gene synthesis representative for each cluster together with the only AMP phosphorylase found in NCBI reference sequence of bacterial origin. Multiple sequence alignment of selected genes using Mcoffee were done to ensure the presence and the correct position of the nucleoside/nucleotide phosphorylase functional domains and to identify the active residues and compare it against *Tk*-AMP-P (only identified AMP-P).

Homology modeling for the chosen enzymes was done using Phyre2 tool (Kelley, Mezulis, Yates, Wass, & Sternberg, 2015). ESBRI online tool was used for the determination of salt bridges in the chosen enzymes (Costantini, Colonna, & Facchiano, 2008).

Validation of the result

This validation approach is designed to search for the characteristic domain architecture of AMP-P in the Red sea datasets and compare it with the results obtained before from HMMsearch. The analysis pipeline as shown in Figure 9 is applied on both Red Sea datasets; sequenced reads and assembled contigs using custom-made reference AMP-P protein database retrieved from NCBI (96 sequences, Supplementary Table 1).

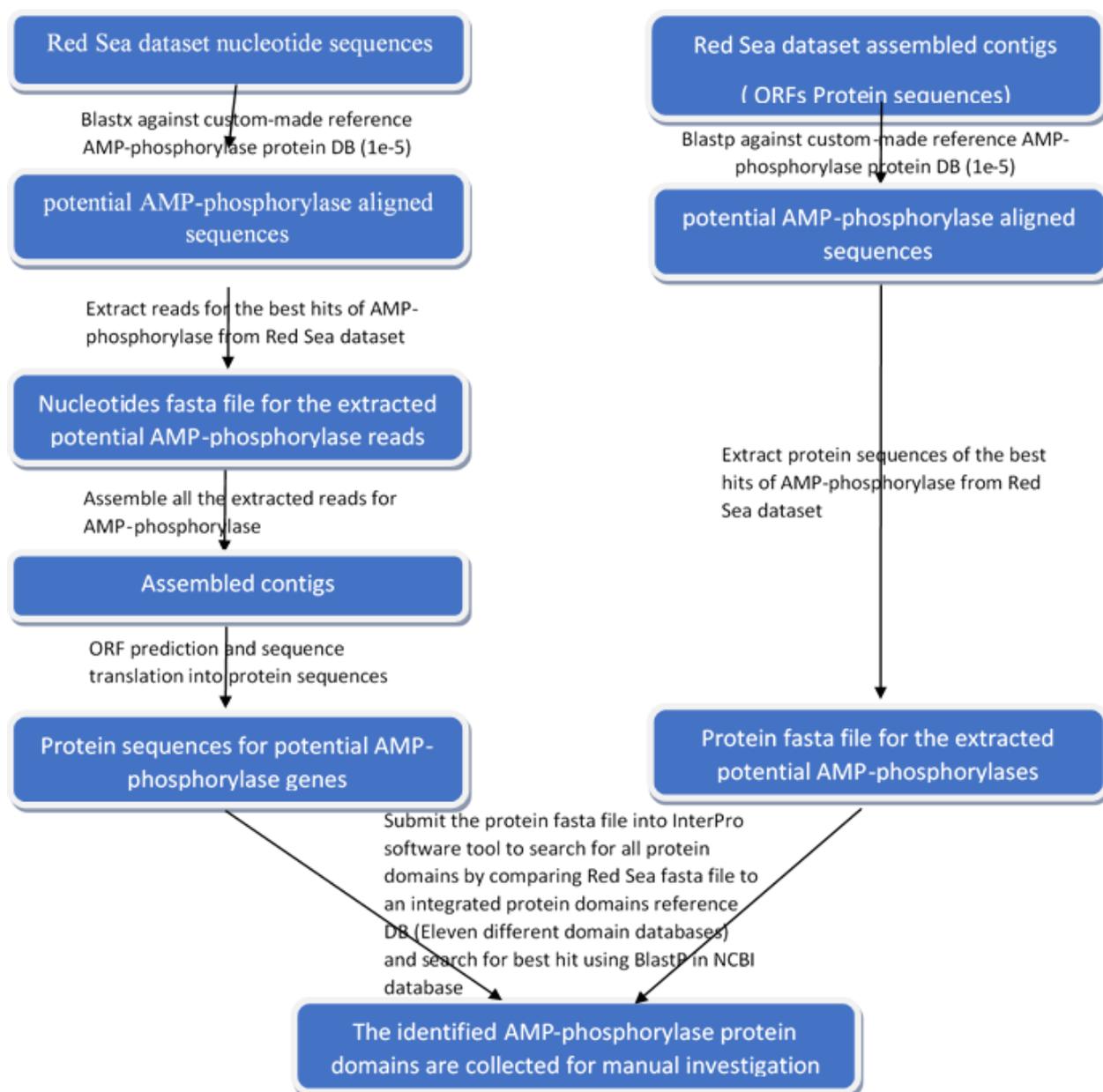


Figure 9. Analysis pipeline diagram for searching AMP-P in the Red sea datasets using a reference database of AMP-P retrieved from NCBI

2.3 Synthesis, Cloning & Expression of AMP phosphorylases

2.3.1 Synthesis of AMP phosphorylases and Transformation in *E. coli* competent Top10 cells

Genes coding for the chosen AMP-P (named Am01-16) enzymes were optimized for expression in the host organism *E. coli* using GeneOptimizer® software included in GeneArt™ in order to achieve the most efficient expression of AMP-P. Then they were synthesized by GeneArt™ Gene synthesis and designed to be flanked by BamHI and HindIII

at the 5' and 3' cloning sites respectively. The fragments were cloned in pMA-T cloning vector with Ampicillin resistance marker (Figure 10) forming cloning construct (pMAT-Am (01-16)).

The cloning constructs were transformed in *E. coli* competent Top10 cells by electroporation. Cells were grown in Luria-Bertani (LB) agar plates containing 100 $\mu\text{g. ml}^{-1}$ ampicillin and incubated overnight at 37°C. Transformed cells were used to propagate the cloning construct by incubating the cells in shake flask containing LB medium and 100 $\mu\text{g. ml}^{-1}$ ampicillin at 37°C, 200rpm everyday until turbid and glycerol stocks were prepared. Cloning construct was then extracted and purified from the transformed cells using the QIAprep® Spin Miniprep Kit (Qiagen, Germany) as mentioned by the manufacturer. The concentration of the purified recombinant plasmid was determined by a spectroscopic measurement at 260 nm by a Nanodrop™ 1000 spectrophotometer (Thermo Scientific, USA). The Purity of the recombinant plasmid is determined by measuring the ratio of spectrophotometric absorbance of the sample at 260 nm to that of 280 nm (A_{260}/A_{280}).

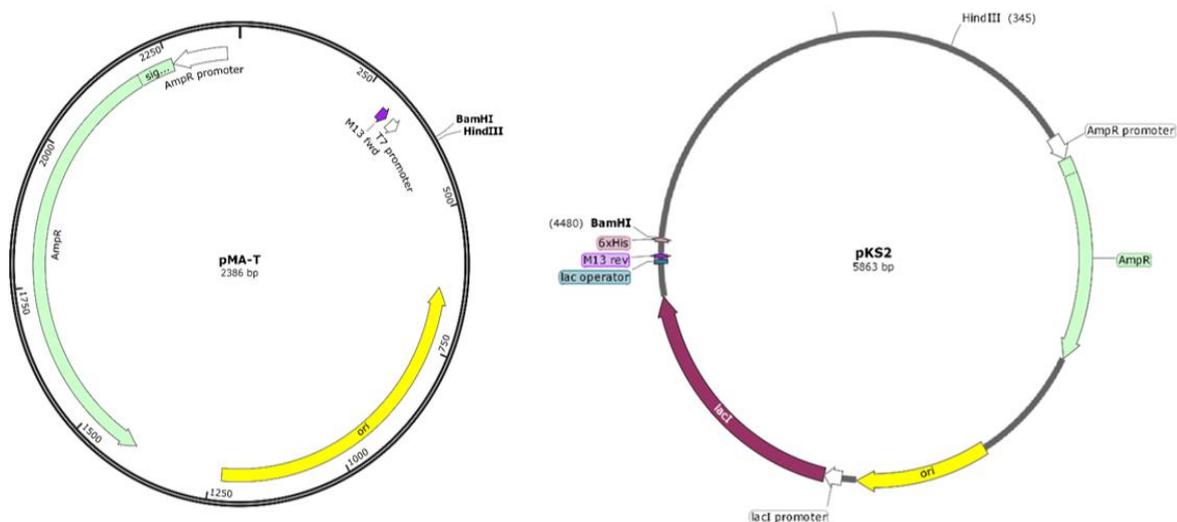


Figure 10 Cloning (pMA-T) and Expression (pKS2) vectors used in this study.

pMA-T plasmid shows the common features with ampicillin resistance gene and the two restriction enzymes used in this study BamHI and HindIII. pKS2 plasmid show also the lac operator responsible for expression with isopropyl β -D-1-thiogalactopyranoside (IPTG), two restriction enzymes BamHI and HindIII, ampicillin resistance gene, and 6X His tag to be added to the N-terminal of the expressed protein for further purification. Plasmid map is visualized using Snap gene viewer software (from GSL Biotech; available at snapgene.com).

2.3.2 Re-Cloning of AMP phosphorylases in the expression vector pKS2-plasmid

Restriction digestion of pMAT-Am recombinant plasmid was done using restriction enzymes BamHI and HindIII and the insert was separated by agarose gel electrophoresis. The gel piece containing the insert was cut and the insert was purified using NucleoSpin® Gel and PCR Clean-Up gel extraction kit (Macherey Nagel, Germany) as mentioned by the manufacturer. The concentration and the purity of the purified insert were determined as mentioned above.

The purified insert and the pKS2 expression vector (backbone, N-terminal 6xHis-Tag, Figure 10) were ligated using T4 DNA ligase. The ligation reaction containing the expression construct (pKS2-Am (01-16)) was transformed into *E. coli* competent Top10 cells by electroporation. Cells were grown in LB agar plates containing 100 µg. ml⁻¹ ampicillin and incubated overnight at 37°C. Transformed cells were used to propagate the expression construct by incubating the cells in shake flask containing LB medium and 100 µg. ml⁻¹ ampicillin at 37°C, 200rpm over day until turbid and glycerol stocks were prepared.

Expression construct was then extracted and purified from the transformed cells using the QIAprep® Spin Miniprep Kit as mentioned by the manufacturer. The concentration and the purity of the purified recombinant plasmid were determined as mentioned above. Each purified expression construct is checked by restriction digestion test using a FastDigest buffer (10X, Thermo Fisher Scientific) and restriction enzymes (BamHI and HindIII) and then analysis via agarose gel electrophoresis to check base pair (bp) number against DNA ladder marker.

The purified expression construct was transformed into *E. coli* competent BL21 cells by electroporation for expression. Cells were grown in LB agar plates containing 100 µg. ml⁻¹ ampicillin and incubated overnight at 37°C. Transformed cells were used to propagate the expression construct by incubating the cells in shake flask containing LB medium and 100 µg. ml⁻¹ ampicillin at 37°C, 200rpm over day until turbid. Glycerol stocks were prepared, and aliquots were frozen and stored at -80°C for further analysis.

2.3.3 Expression of recombinant AMP phosphorylases in *E. coli* BL21 cells using Enpresso®B medium and 2xYT Medium Broth

Enpresso®B medium

Recombinant *E. coli* strain BL21 was cultivated in EnPresso®B animal free medium 50ml (BioSilta, Finland) using Ultra Yield Flasks™ and AirOTop™ seals (Thomson Instrument Company, USA).

Preculture was prepared in animal-free LB broth containing 100 µg. ml⁻¹ ampicillin and incubated at 37⁰C (6-8hr, 250rpm). Preculture OD₆₀₀ was measured using UV/Vis spectrophotometer (Ultrospec 3000, Amersham Biosciences, Germany). The main culture was inoculated with fresh preculture to a final OD₆₀₀ equals 0.15 and incubate overnight at 30⁰C (200rpm).

Then, Protein expression was induced using 1 mM IPTG and incubated at 30⁰C (200 rpm). 24hr after induction, cells were harvested by centrifugation (8000 X g, 4⁰C, 10min) and the pellets were stored in -20⁰C for further analysis.

2xYT Medium Broth

The 2xYT medium was used for the expression of Am07 to study if cold expression will enhance the solubility of the expressed protein. Recombinant *E. coli* strain BL21 was cultivated in 2xYT medium 500ml (16g peptone, 10g yeast extract, and 5g NaCl) using Ultra Yield Flasks™ and AirOTop™ seals. Preculture was prepared in animal-free LB broth containing 100 µg. ml⁻¹ ampicillin and incubated overnight at 30⁰C (200rpm).

Preculture OD₆₀₀ was measured using UV/Vis spectrophotometer. The main culture was inoculated with fresh preculture to a final OD₆₀₀ equals 0.15 and incubate at 37⁰C (200rpm) till OD₆₀₀ reached 0.6.

Then, Protein expression was induced using both 50 µM and 1 mM IPTG and incubated at 22⁰C (200 rpm). Twenty two hours after induction, cells were harvested by centrifugation (8000 X g, 4⁰C, 10min) and the pellets were stored in -20⁰C for further analysis.

To analyze protein expression in both media, 1ml sample was taken before and after induction with IPTG. OD was measured using UV/Vis spectrophotometer with proper dilution to make sure that the absorbance value is in the range of 0.2-0.8 and then a sample of OD₆₀₀ equals 5 is prepared. Cells were harvested by centrifugation (16000 X g, 4⁰C, 5min)

and the pellets were stored in -20°C for analysis by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

2.4 SDS-PAGE analysis

Cell disruption of cell pellets obtained from $\text{OD}_{600}=5$ sample was done by suspending each pellet in 300 μl BugBusterTM Protein Extraction Reagent (Novagen, Germany) containing 0.3 $\mu\text{l. ml}^{-1}$ Benzonase® (Merck, Germany) and 1 $\mu\text{l. ml}^{-1}$ Lysozyme (Sigma Aldrich, Germany) followed by incubation for 30 min at 30°C . A sample (40 μl) was taken. Centrifugation (16000 x g, 4°C , 15min) for samples were done in order to separate the soluble and insoluble protein fractions followed by adding 1 volume of 2X SDS loading buffer (125mM Tris HCl pH6.8, 5% SDS, 0.25% bromophenol, 25% Glycerol, 200mM Dithiothreitol (DTT)) to soluble fractions and the same volume containing 8M Urea in ratio 1:1 was added to the insoluble fractions. All samples were heated at 95°C for 5 min and 10 μl was added for SDS-PAGE analysis on 12% gels and stained with Coomassie blue according to the Laemmli method (Laemmli, 1970).

2.5 Purification of His-tag AMP phosphorylase by Ni-NTA spin column affinity chromatography

Enzymatic lysis

Each cell pellet was disrupted by suspending in 3x Lysis buffer solution containing binding buffer (50mM sodium phosphate buffer, 300mM NaCl, 10 mM imidazole, pH 8.0), lysozyme (1mg. ml^{-1}), MgCl_2 (1mM), phenylmethylsulfonyl fluoride (PMSF, 0.1mM) and DNase powder followed by incubation at 30°C for 1hr at 300rpm in Thermo-shaker.

Mechanical lysis

Cells were subjected to sonication using Ultrasonicator UP200S (Hielscher Ultrasonics, Teltow) for 10 min with 30s on/off intervals and 30% amplitude. Samples (40 μl) were taken for further analysis by SDS-PAGE as mentioned before. Centrifugation (10000 x g, 4°C , 30min) to separate the supernatant from the cell debris was done.

Purification by Ni-NTA spin column affinity chromatography

The target protein was purified using Nickel (Ni^{2+}) affinity chromatography with Ni-NTA agarose resin (Jena BioScience, Germany) according to manufacture manual using spin columns at room temperature. The His-tagged AMP-P was eluted using Elution buffer (50mM sodium phosphate, 300mM NaCl, 250mM imidazole, pH 8.0). Purified proteins elution fractions were added for SDS-PAGE analysis on 12% gels and stained with Coomassie blue according to the Laemmli method (Laemmli, 1970).

Protein Concentration

The concentration of the purified protein was determined by a spectroscopic measurement at 280 nm by a Nanodrop™ 1000 spectrophotometer using theoretically extinction coefficient calculated by Protparam web tool (Gasteiger et al., n.d.) listed in Table 3.

Table 3. Extinction coefficient of AMP phosphorylases calculated by Protparam web portal.

AMP-P	Extinction coefficient	AMP-P	Extinction coefficient
Am01	4.83	Am09	5.63
Am02	3.94	Am10	5.19
Am03	4.56	Am11	3.96
Am04	4.38	Am12	6.25
Am05	5.78	Am13	3.88
Am06	7.35	Am14	3.82
Am07	7.11	Am15	5.89
Am08	6.07	Am16	9.62

2.6 Activity assay of the purified AMP phosphorylases

Phosphorylase activity assays were performed in 100 μ l reaction volume containing 1mM substrate (**NMP:** AMP or CMP or GMP (guanosine 5'-monophosphate) or UMP, **Nucleoside:** Adenosine or Uridine), 10mM potassium phosphate buffer (pH 7.5), 20 μ l purified protein (no heat) and incubated at different temperatures 37, 50 and 80 °C. After 16hr incubation, samples were withdrawn and stopped with 100 μ l methanol. Centrifugation (20000 x g, 4°C, 20 min) to remove coagulated protein and the samples were stored at 4°C for further analysis by high-performance liquid chromatography (HPLC).

Negative controls were done as mentioned above but by using 20µl water instead of purified protein and samples were withdrawn and stopped with 100µl methanol at 0hr and after 16hr incubation at 37, 50 and 80 °C.

From the HPLC results, the desired peak area under the curve of the substrate and their corresponding nucleosides and bases at specific retention times were calculated. The relative activity was calculated using the following formula:

$$\text{Relative phosphorylase activity (\% conversion)} = \frac{\text{Conc. of the product}}{\text{Conc. of (Product+substrate+Nucleoside)}} \times 100$$

2.7 Analysis of Thermostability

The purified AMP-P protein was diluted to 0.5mg.ml⁻¹ and heated at the corresponding temperature 60, 70 and 80°C. After 24hr samples were cooled at 4°C for 1hr followed by centrifugation (20000 x g, 4°C, 20 min). The soluble fraction was analyzed using 12 % SDS gels and concentration is measured using Nanodrop spectrophotometer as mentioned above. 20µl of the purified protein was heated at 70°C for 1hr followed by centrifugation (20000 x g, 4°C, 20 min). The soluble fraction was analyzed using 12 % SDS gels as mentioned above.

Phosphorylase activity assays were performed in 100µl reaction volume containing 1mM substrate (**NMP:** AMP or CMP **Nucleoside:** Uridine), 10mM potassium phosphate buffer (pH 7.5), 20µl purified protein (heat 70°C for 1hr) and incubated at different temperatures 37, 50 and 80 °C.

For Uridine: after 1hr 50µl of the sample was taken and stopped with 50µl methanol followed by centrifugation (20000 x g, 4°C, 20 min), complete to 200µl with ddH₂O and the samples were stored at 4°C for further HPLC analysis. After 16hr incubation, samples were withdrawn from all and stopped with methanol and negative control was used as mentioned above (Table 4).

Table 4. Phosphorylase activity assay for purified proteins after heating at 70°C for 1hr to test thermostability at a higher temperature

		AMP	CMP	Uridine
Negative controls	0hr (no incubation)			
	16hr (end point,37°C)	100µL reaction volume		
	16hr (end point,50°C)	1mM substrate, 10 mM KP, pH 7.5, 20 µl Water		
	16hr (end point,80°C)	100 µl Methanol, Centrifugation		
Phosphorylase Activity	AMPP	AMP	CMP	Uridine
	1hr (37,50,80°C)			50ul sample after 1 hr, 50 ul Methanol
	Uridine only			Centrifugation, add 100ul water
	16hr (end point,37°C)	100µL reaction volume		50µL reaction volume
	16hr (end point,50°C)	1mM substrate, 10 mM KP, pH 7.5, 1mM substrate, 10 mM KP, pH 7.5,		
		20 µl of enzyme solution (purified, without dialysis, heat treated 70 degree for 1 hr)		
	16hr (end point,80°C)	100 µl MeOH, Centrifugation		50 ul Methanol, Centrifugation, 100ul water

2.8 High-Performance Liquid Chromatography (HPLC) analysis

The amount of NMP and their corresponding nucleosides and bases was determined by following the absorption at 260 nm during HPLC analysis using a reversed phase C18 Evo column (Kinetex 250x 4.6 mm, Phenomenex) with the following gradient: from 97% 20 mM ammonium acetate and 3% acetonitrile to 60% 20 mM ammonium acetate and 40% acetonitrile in 10 min with flow rate 1ml. min⁻¹ at 25°C.

Chapter Three: Results

3.1 Identification of Putative Thymidine /AMP phosphorylase sequences from Red Sea metagenomic dataset and NCBI protein database

The original aim of the study was the identification of novel pyrimidine nucleoside phosphorylases from the Red Sea. Therefore, in the first approach Pfam domains of pyrimidine nucleoside phosphorylase (α/β domain, α domain, and C-terminal domain) were used to perform a HMMsearch against assembled data of the Red sea metagenomic database (ATII, KD, DD and BI; water and sediment data) searching for significantly similar sequence matches. As only a few hits were reached, a different strategy was used. Based on available sequences in NCBI, a database of thermophilic pyrimidine nucleoside phosphorylase enzymes (119 enzymes) was generated. HMMbuild was used to define the characteristic

domains from the multiple sequence alignment of the thermophilic PyNPs and HMM profile was built which was used to perform HMMsearch against assembled data of the Red Sea.

241 and 122 ORFs were identified from HMMsearch containing the domains of PyNPs using the thermophilic PyNPs customized database or the Pfam HMM profile, respectively. The identified ORFs were filtered by evaluating how many domains every ORF contained. Only ORFs containing the three essential functional domains were selected and the rest of the ORFs were either having a single domain or two domains only.

Eleven ORFs (assigned code ORF 01-11) were selected for further analysis (Table 5) based on the presence of the three essential functional domains of thymidine/pyrimidine nucleoside phosphorylase family. This family belongs to the glycosyltransferase superfamily which also includes rarely studied AMP-P or Ribose 1,5-bisphosphate phosphokinase (PhnN).

To confirm the preliminary screening, all ORFs were aligned against the nr protein database using BlastX and Interpro protein family database. Findings based on these databases confirmed that these ORFs belong to glycosyltransferase superfamily with significantly low E-values (Table 5). While most of the ORFs are related to bacteria, two ORFs (ORF 10,11; from Kebrit brine pool upper interface and brine) showed a high degree of identity to annotated AMP-P from the archaeal origin.

3.2 Phylogenetic analysis of the identified ORFs from the Red Sea

Phylogenetic trees were built to study PyNPs of bacteria and archaea in more detail. Representative sequences of each genus and all thermophilic, halophilic or psychrophilic sequences from NCBI were chosen to build the phylogenetic tree. Surprisingly, TP sequences of archaeal origin were longer than the expected 430 to 440 amino acids. A deeper search revealed that the sequences were more closely related to the rarely described AMP-P. Within the last year, many of the archaeal sequences were re-annotated from homolog of TPs to homologs of AMP-P. As only a few reports to AMP-P are available in literature the focus of the presented work went to study AMP-P of mainly archaeal origin.

Table 5. BlastX and Interpro results for the eleven ORFs identified by HMMsearch against Red sea metagenomic data. Ribose 1,5-bisphosphate phosphokinase is highlighted as it does not belong to thymidine/pyrimidine nucleoside phosphorylase family.

ORF name Database	ORF length DNA (bp) / Protein (aa)	BlastX (Non-redundant database)					Interpro (Protein database)
		Description (Best hit)	E- value	Identity %	Hit coverage	Taxonomy	Protein family membership
contig01155_gene2_641_1897_+ / ORF01 ATII1500m	1257 / 419	Thymidine phosphorylase [Candidatus Marinimicrobia bacterium]	0	67.83%	95%	Bacteria	Thymidine/pyrimidine nucleoside phosphorylase
contig00107_gene5_3412_4728_- / ORF02 ATII-IN	1317 / 439	Thymidine phosphorylase [Phyllobacterium myrsinacearum]	0	97.03%	99%	Bacteria (α - Proteobacteria)	Thymidine/pyrimidine nucleoside phosphorylase
contig00022_gene27_22317_23633_+ / ORF03 ATII-UCL	1317 / 439	Thymidine phosphorylase [Phyllobacterium myrsinacearum]	0	97.03%	99%		Thymidine/pyrimidine nucleoside phosphorylase
contig00011_gene175_179086_180402_- / ORF04 ATII-LCL	1317 / 439	Thymidine phosphorylase [Phyllobacterium myrsinacearum]	0	97.03%	99%		Thymidine/pyrimidine nucleoside phosphorylase
contig00130_gene9_8972_10225_- / ORF05 ATII-LCL	1254 / 418	Thymidine phosphorylase [Candidatus Marinimicrobia bacterium]	0	80.58%	99%	Bacteria	Thymidine/pyrimidine nucleoside phosphorylase
contig00132_gene16_14556_16028_+ / ORF06 ATII-LCL	1473 / 491	Putative Thymidine phosphorylase [Ralstonia pickettii]	0	100%	99%	Bacteria (β - Proteobacteria)	Thymidine/pyrimidine nucleoside phosphorylase
contig00136_gene6_5374_7536_- / ORF07 ATII-LCL	2163 / 721	Phosphonate metabolism protein/1,5- bisphosphokinase (PRPP-forming) PhnN [Cupriavidus basilensis]	0	94.31%	99%		Ribose 1,5-bisphosphate phosphokinase PhnN
contig04720_gene1_1_1548_+ / ORF08 ATII-LCL	1548 / 516	Thymidine phosphorylase family protein [Cupriavidus basilensis]	0	95.49%	98%		Thymidine/pyrimidine nucleoside phosphorylase
contig00230_gene3_1708_2988_+ / ORF09 KD-UIN	1281 / 427	Thymidine phosphorylase [Candidatus Marinimicrobia bacterium]	0	76.71%	99%	Bacteria	Thymidine/pyrimidine nucleoside phosphorylase
contig05364_gene2_226_1573_- / ORF10 KD-UIN	1348 / 449	AMP phosphorylase [uncultured bacterium]	0	81.47%	99%	Uncultured Bacteria	Thymidine nucleoside phosphorylase/ AMP phosphorylase
		AMP phosphorylase [Candidatus Woesearchaeota archaeon]	7E- 146	50.11%		Archaea	
contig00866_gene3_707_1948_- / ORF11 KD-BR	KD-BR 1242 / 414	AMP phosphorylase [Hadesarchaea archaeon]	3E- 179	61.71%	99%	Archaea	Thymidine nucleoside phosphorylase/ AMP phosphorylase

The phylogenetic tree included all ORFs identified in the Red Sea metagenomic data, representative sequences of each genus and all thermophilic, halophilic or psychrophilic sequences from NCBI. The sequences were divided into eight clusters and fifteen enzymes were selected for further characterization (Figure 11, Table 6).

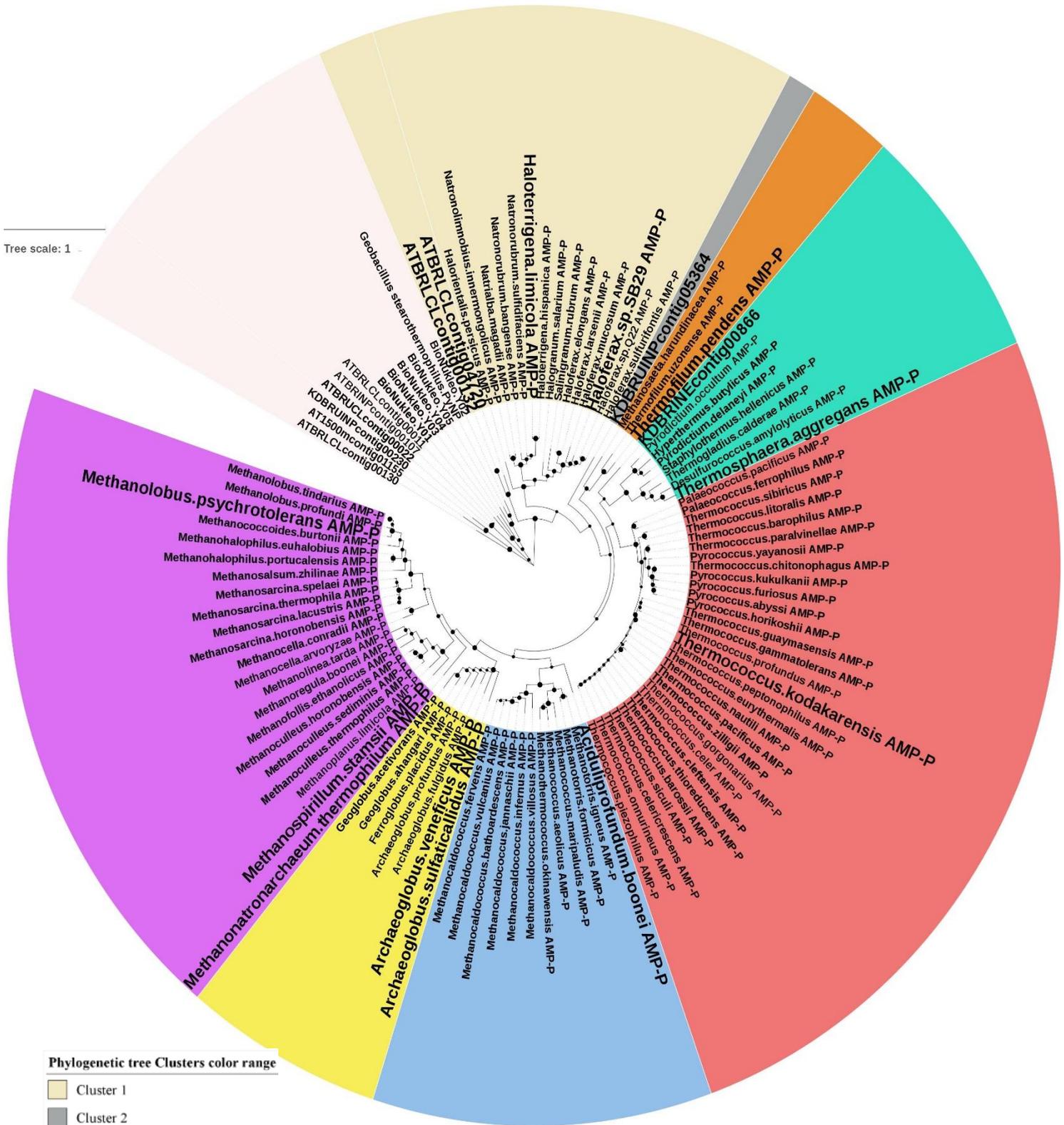
The AMP-P of *Tk* was used as a positive control for comparing structure and sequences with other selected candidates as it is the only characterized and crystalized member of this enzyme family (Aono et al., 2012a; Nishitani et al., 2013). Another enzyme was chosen for further characterization from the NCBI database the only RefSeq AMP-P from a bacterial origin. It was identified in a Parcubacteria group bacterium (Accession number WP_116594393.1). To guarantee that a wide range of enzymes with broad substrate spectrum is expressed the enzymes were chosen based on the following reasons:

- 1) The showed phylogenetic distance within the cluster.
- 2) The difference in the active residues compared to *Tk*-AMP-P were identified. It is the only identified and crystalized member of this enzyme family (Aono et al., 2012a; Nishitani et al., 2013).
- 3) The organisms harboring the enzymes were isolated from extreme environments.

Figure 11. Molecular phylogenetic analysis of Red Sea metagenomic identified ORFs by Maximum Likelihood method in Mega X.

ORFs from Red Sea metagenomic data were aligned together with archaeal AMP-Ps. PyNP from *Geobacillus stearothermophilus* and thermophilic TP/PyNP (BioNukleo) were used as non-AMP-P controls. Maximum likelihood method was used based on the Poisson correction model. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log-likelihood value. The tree with the highest log likelihood (-35246.78) is shown. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Circles indicate confident bootstrap values. The color range is assigned to each cluster shown in the box attached. Enzymes in bold indicate enzymes were chosen from each cluster. The phylogenetic tree is presented by Interactive tree of life (iTOL) online tool V.4.3.3.

Tree scale: 1



Phylogenetic tree Clusters color range

- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Cluster 5
- Cluster 6
- Cluster 7
- Cluster 8
- PyNP/TP Cluster

Table 6. Characteristics of the sixteen chosen enzymes for gene synthesis. The only bacterial AMP-P sequence is highlighted by a green background

16 chosen AMP-P enzymes	Cluster no.	Code name	Active residues differences (Reference: TK.AMP-P)	3D structure similarities (Phyre2 results, 100% confidence, best 2 hits)		Extremophilic signature (No. of Salt bridges)	Isolation environment &/or Archaea characteristics	Selection criteria from Phylogenetic clusters
				TK.AMP-P (PDB:4ga5)	Gs-PyNP (PDB:1brw)			
Thermococcus kodakarensis	5	Am01	100%	35%	182	Solfatara, Hyperthermophile	The only studied AMP-P, Positive control
ORF11	4	Am02	Ser175 instead of Asn (P04- binding)	60%	33%	126	Kebrtit deep brine (hypersaline and Sulfur-rich)	Clustered with AMP-P forming a separate branch between Crenarchaeote and Euryarchaeote (Halobacteria)
Thermosphaera aggregans		Am03	Ala175 instead of Asn (P04- binding)	49%	29%	104	Hot spring, Hyperthermophilic	A mostly branched enzyme in the clade (more speciation), Length of Enzyme 512aa (9 residues diff. to Tk.AMP-P)
ORF10	2	Am04	Cys165 instead of Ser (P04- binding) Thr175 instead of Asn (P04- binding) Arg250 instead of Glu (Pyrimidine binding)	41%	29%	151	Kebrtit deep upper brine interface	Clustered with AMP-P forming a separate branch between Crenarchaeote and Euryarchaeote (Thermococci)
Aciduliprofundum boonei	6	Am05	Ala175 instead of Asn (P04- binding) Ser198 instead of Thr (5'-P04- binding)	51%	31%	163	Hydrothermal vent, Thermoacidophile	Mostly branched enzyme in the clade (more speciation)
Archaeoglobus sulfaticallidus	7	Am06	Ala175 instead of Asn (P04- binding) Ser256 instead of Asp (Purine binding) Gln250 instead of Glu (Pyrimidine binding) Arg268 instead of Lys (Pyrimidine binding)	49%	34%	139	Black rust exposed to the hot ridge flank crustal fluids, Thermophilic	Branched alone forming separate branch (early speciation in the clade)
Archaeoglobus veneficus		Am07	Ser175 instead of Asn (P04- binding) Ile256 instead of Asp (Purine binding) Arg288 instead of Lys (5'-P04- binding)	51%	34%	173	Black smokers, Hyperthermophilic	
Methanotatronarchaeum thermophilum	8	Am08	Thr175 instead of Asn (P04- binding) Gly199 instead of Ser (5'-P04- binding)	48%	31%	142	Hypersaline soda lakes sediments, Extreme Halophilic, and moderate Thermophilic	Branched alone forming separate branch (early speciation in the clade)

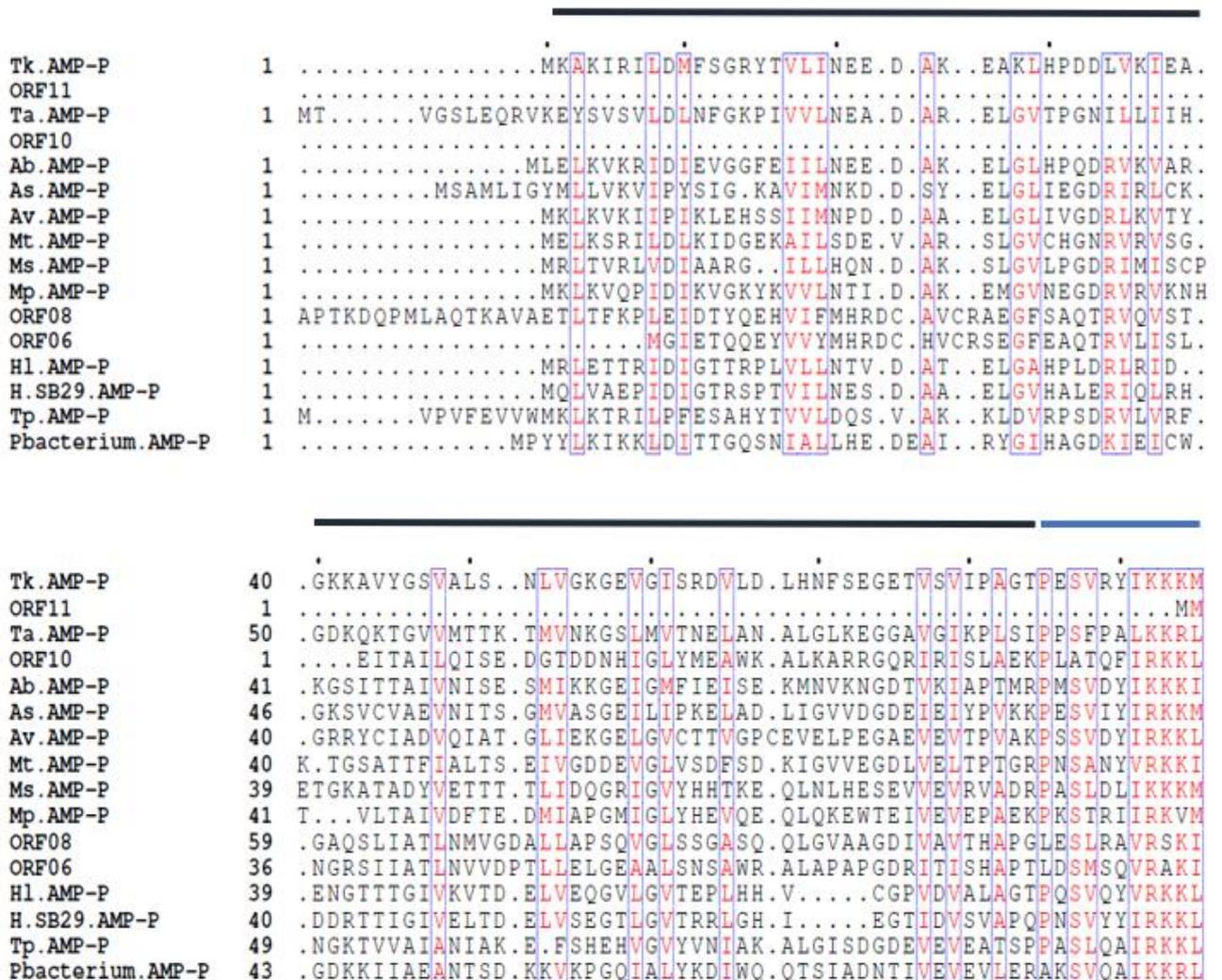
Continuation of **Table 6**

Methanospirillum stamsii	8	Am09	Ser175 instead of Asn (P04- binding) Gly199 instead of Ser (5'-P04- binding)	47%	30%	173	Anaerobic Bioreactor, Psychrotolerant	Branched alone forming separate branch (early speciation in the clade), Psychrotolerant, Length of Enzyme 516aa (13 residues diff. to <i>Th.kodakarensis</i>)
Methanolobus psychrotolerans		Am10		51%	31%	152	Saline meromictic Lake, Psychrotolerant	Last speciation in the clustered clade, Psychrotolerant
ORF08	1	Am11	Cys165 instead of Ser (P04- binding) Ser175 instead of Asn (P04- binding)	38%	29%	146	Atlantis II lower convective layer	Clustered together forming a separate branch between AMP phosphorylases and TP/PyNP phosphorylases.
ORF06		Am12	Cys165 instead of Ser (P04- binding) Thr175 instead of Asn (P04- binding)	35%	28%	186		
Haloterrigena limicola		Am13	Thr175 instead of Asn (P04- binding) Val197 instead of Ile (5'-P04- binding)	46%	29%	167	Aibi Salt Lake sediment, Extreme Halophilic and moderate Thermophilic	Last speciation in the clustered clade
Haloferax sp. SB29		Am14	Thr175 instead of Asn (P04- binding) Val197 instead of Ile (5'-P04- binding)	48%	32%	181	Discovery deep brine-seawater interface	
Thermofilum pendens	3	Am15	Ser175 instead of Asn (P04- binding)	49%	32%	130	Hot spring, Hyperthermophilic	Length of Enzyme 520aa (17 residues diff. to TK.AMP-P)
Parcubacteria group bacterium	bacteria	Am16	Thr175 instead of Asn (P04- binding) Ser250 instead of Glu (Pyrimidine binding)	41%	29%	149	Anaerobic Bioreactor	Only bacteria I found in NCBI RefSeq has AMP phosphorylase enzymes others are AMP nucleosidases

Multiple sequence alignment of the selected genes with *Tk*-AMP-P as positive control (Figure 12) revealed that all ORFs selected from the Red Sea showed all active residues responsible for activity as 5'-phosphate binding residues which is essential for binding with 5'-phosphate of nucleotide, purine and pyrimidine binding residues which are responsible for binding to purine and pyrimidine nucleosides respectively. However, there are some differences in the phosphate binding residues side from residues of *Tk*-AMP-P were identified (Table 6). It shows also ORF11 doesn't have the 82 residues of the N-terminal domain which is responsible for AMP-P thermostability, enzymatic activity, and multimerization.

Figure 12. Multiple sequence alignment for the chosen enzymes with the only characterized AMP-P from *Tk* archaeon.

Sequences were aligned using Mcoffee and the figure was displayed using ESript program (Robert & Gouet, 2014). The N-terminal (2-84), Central (84-149, 241-272), and C-terminal (161-233, 276-503) domains are shown in black, blue, and pink lines respectively above sequences based on *Tk*-AMP-P structure. Identical and similar residues are indicated by white letters on red backgrounds and red letters, respectively. A 5'-phosphate binding loop is indicated by a black box. Residues binding to a phosphate ion, a purine base, a pyrimidine base, and a 5'-phosphate group of NMP are indicated by pink circles, blue circles, green circles, and orange stars, respectively.



Tk. AMP-P	96	HGEKLRKV	EIEAIVR	DIVDRKLRD	IEISSFV	TALEINGL	LDMD	EIAAL	LTIA	MAET	GDML	LDI					
ORF11	3	KGSSLEDE	EIHSIV	EDIVSGK	LDTIQMT	ALLVTQE	IRGMSME	ETAAL	LRSM	IE	TGET	VDI					
Ta. AMP-P	107	KGERLEESH	YREIVK	DVVEGV	YGEAEIA	AFLVSQL	FYEP	TEEE	ELTYL	LIKAM	VE	TGSR	IVF				
ORF10	55	EGERLEPK	EIDEIK	DVTEDD	LSDIEL	TYFVSG	CVHGL	LNNP	ETAAL	LTKS	IVKH	GSKL	DF				
Ab. AMP-P	98	YGEKLRK	DEIYEI	KDIVDD	NLSTIEL	ASYVTA	IQIRG	MDMD	ETE	WMTK	AMVD	TGET	IDF				
As. AMP-P	103	EYKLSLD	EIRSII	NDTVSG	LSNIEL	TAFLSN	FVNG	MDFD	EIEW	MTRAM	IE	TGET	ITF				
Av. AMP-P	98	EGEKLSKE	ELYAII	RDIVVNM	LNEVEL	SFAVM	ANYLRD	VDDL	EVEW	MTRAM	ID	TGET	LTF				
Mt. AMP-P	97	YGEFLSRE	EMFELI	KDIANRN	LSNYEL	SALVT	AVSIRD	MSMD	EVQW	MTEAM	VEYGD	SLEF					
Ms. AMP-P	97	NGEKL	SRD	DIRGIV	ADIVQDT	LSPSEIT	AFVSS	YINQL	LDMD	EIES	LTRAM	VE	TGDQ	ITF			
Mp. AMP-P	96	DGRKLERD	EINELV	KDIVEEN	LDDIEL	AFLT	ATYIKD	LSED	ETE	WLT	KAMID	TGER	IEF				
ORF08	117	HGNPLDAP	QLCAI	MGDVA	AGRYSD	VHIA	AFLSAC	AGGR	MTTQ	ETVD	LTCAM	LD	TGDR	LDW			
ORF06	94	YGHRLD	SALDSII	SDVA	AGSYPG	THIA	AFLSAC	AGGR	MDLQ	ETID	LTRAM	LA	GKR	LDW			
Hl. AMP-P	91	NDIELEDD	ELEAIV	KDIHEN	RLSDIEL	SAYV	SAVYANG	LSLQ	ETKH	LTRAM	TDI	IGRQ	LSW				
H. SB29. AMP-P	92	NDIELE	ERHEL	GRIVR	DIYEER	LADVEL	GAYV	SATY	TNGL	LSME	ETLH	LTEC	MA	DV	GESITW		
Tp. AMP-P	105	QGLSLES	DEIYQ	VVKD	IVD	GKLE	LELAF	VAVHF	QGMTP	SEI	YSFT	LSM	VE	TGQR	LRL		
Pbacterium. AMP-P	100	LGKKLKYQ	DFQIF	S	DIAS	GVLR	TRTE	TTYF	VAS	GFMS	YTNQ	ELY	MTK	AMAE	S	GEM	LKF

Tk. AMP-P	156	DRKP	IMDVHS	IGGV	PGN	.KTN	ILVVP	IVAA	AGL	TIP	KTS	SR	AIT	SAA	G	FAD	VVE	FAD	V	S																										
ORF11	63	EKKP	VMDVHS	IGGV	PGN	.KYS	LITVP	IVTA	AGL	TIP	KTS	SR	AIT	S	PAG	FAD	IME	V	LAD	V	S																									
Ta. AMP-P	167	E. EP	ANDVHS	VGGV	PGN	S	KVAL	ITVP	IVAA	SGL	LIP	KTS	SR	AIT	S	PAG	FAD	TV	E	V	L	AR	V	D																						
ORF10	115	GKKL	VVDK	KHC	IGGV	PGN	.RTT	MIIPI	IVTA	AGL	LMP	KTS	SR	AIT	S	P	S	G	FAD	T	M	E	V	I	AN	V	K																			
Ab. AMP-P	158	E. Y	TVF	DVHS	IGGV	PGN	.KYA	PIAV	IVAA	AGL	KIP	KTS	SR	AI	S	S	AAG	FAD	L	M	E	V	L	T	N	V	R																			
As. AMP-P	163	ERGI	VVDK	H	SIGGV	PGN	.KTA	LLV	PIVA	SGL	LIP	KTA	SR	AI	S	S	A	G	FAD	T	F	E	V	L	A	D	V	N																		
Av. AMP-P	158	EKGI	VVDK	H	SIGGV	P	G	D	.KTS	L	V	V	P	I	I	A	A	G	L	L	I	P	K	T	A	S	R	A	I	T	S	A	S	G	FAD	T	M	E	V	L	A	N	V	S		
Mt. AMP-P	157	TEYP	IVDK	H	SIGGV	P	G	D	.KIT	L	L	I	V	P	I	I	A	A	G	L	I	P	K	T	S	SR	AI	T	G	A	A	G	FAD	I	M	E	V	L	A	P	V	K				
Ms. AMP-P	157	HAGP	IVDK	H	SIGGV	P	G	N	.KIS	L	I	V	V	P	I	I	A	A	S	G	L	F	I	P	K	T	S	SR	AI	T	G	A	G	FAD	L	M	E	V	L	C	P	V	E			
Mp. AMP-P	156	STSP	IMDK	H	SIGGV	P	G	N	.KIS	L	L	I	V	P	I	I	A	A	N	G	L	L	I	P	K	T	S	SR	AI	T	G	A	G	FAD	L	M	E	V	L	A	P	V	E			
ORF08	177	SRPV	VADK	H	CVG	GL	L	P	G	N	.RTS	P	I	V	V	A	I	C	A	A	G	L	L	I	P	K	T	S	SR	AI	T	S	P	A	G	FAD	T	M	E	V	L	T	R	V	T	
ORF06	154	GHAP	IADK	H	CVG	GL	L	P	G	N	.RTT	P	I	V	V	S	I	I	T	A	A	G	L	T	M	P	K	T	S	SR	AI	T	S	P	A	G	FAD	V	I	E	T	M	T	P	V	A
Hl. AMP-P	151	DAPV	VADK	H	SIGGV	A	G	N	.CVT	P	I	I	V	S	I	I	T	A	A	G	L	T	M	P	K	T	S	SR	AI	T	S	P	A	G	FAD	V	V	E	F	C	D	V	E			
H. SB29. AMP-P	152	DESI	IADK	H	SIGGV	A	G	N	.RV	T	P	I	L	V	P	I	I	A	A	G	L	K	I	P	K	T	S	SR	AI	T	S	A	A	G	FAD	T	M	E	V	L	C	D	V	E		
Tp. AMP-P	165	KRKP	ILDK	H	SIGGV	P	G	D	.KTS	L	L	V	V	P	I	I	A	S	L	G	F	T	I	P	K	T	S	SR	AI	T	S	A	A	G	FAD	R	M	E	V	L	A	P	V	N		
Pbacterium. AMP-P	160	P. GI	IADK	H	SVG	GL	A	G	N	.RTT	M	V	A	V	P	I	L	A	S	L	G	L	T	I	P	K	T	S	SR	AI	T	S	P	A	G	FAD	T	M	E	V	L	A	K	V	S	

● ★ ● ● ● ★★★ ●

5'-phosphate binding loop

Tk. AMP-P	215	FSLDEIKR	I	VEKV	GACLV	WGGAL	NL	LAP	ADD	ITIK	AER	ALS	IDP	TGL	ML	AS	S	I	M	S	K	K	Y	A	M	G	S																											
ORF11	122	FSLDEI	SEI	VSKV	GGVLA	WGGAV	NL	LAP	ADD	MLIN	I	ER	PLR	IDP	EC	QL	LA	S	V	L	S	K	K	L	A	M	S	S																										
Ta. AMP-P	226	LTIDEIKD	I	VKTK	KGVL	AWGGK	NL	LAP	ADD	IFVN	I	ERR	LA	IDP	I	YQ	M	V	A	S	I	L	S	K	K	L	A	M	G	I																								
ORF10	174	NDAKK	LKKI	IAEK	VG	GMTW	GGG	V	D	L	A	A	ADD	H	M	I	R	V	R	H	P	L	S	L	D	P	E	G	M	L	L	A	S	I	M	A	K	K	H	S	V	G	A											
Ab. AMP-P	216	LSVND	I	RRIV	DE	V	GATL	TW	GGAV	NL	LAP	ADD	K	I	V	H	V	E	Y	P	L	G	I	D	P	H	S	Q	V	L	A	S	V	M	A	K	K	A	V	G	A													
As. AMP-P	222	LTVDEI	KEI	I	TER	V	GGVIA	W	SAS	AN	LAP	ADD	R	L	I	E	I	Q	N	L	Q	I	S	P	I	P	H	F	I	T	S	I	M	S	R	K	S	A	V	G	A													
Av. AMP-P	217	LSVDEI	KEI	I	TER	V	GGVIA	W	GGAT	N	I	A	P	ADD	K	I	R	V	E	Y	P	L	A	I	P	K	P	H	L	L	S	V	M	A	K	K	G	A	L	G	A													
Mt. AMP-P	216	FTAEEI	YQIT	KQV	GAAI	VW	GGAT	H	I	A	P	ADD	E	I	R	A	E	Y	P	L	A	L	D	P	V	S	I	L	A	S	V	L	A	K	K	I	A	V	G	A														
Ms. AMP-P	216	FSATEV	QEM	TM	KTG	GVIV	W	GGAT	N	I	A	P	ADD	K	I	I	L	Q	E	Y	P	F	K	I	D	Q	I	G	O	M	I	A	S	V	M	A	K	K	F	A	V	G	A											
Mp. AMP-P	215	FTAQQI	KE	MTE	KV	GGVIV	W	GGAT	N	I	A	P	ADD	K	L	I	K	E	Y	P	L	S	I	D	P	H	C	Q	L	L	A	S	I	M	A	K	K	G	A	V	G	A												
ORF08	236	LSAAEM	R	R	V	VER	V	GAA	L	V	W	GG	L	T	L	S	P	ADD	V	L	I	R	V	E	R	A	L	E	I	D	S	D	A	Q	L	V	A	S	V	L	S	K	K	L	A	A	G	S						
ORF06	213	LDLEQ	M	R	R	V	V	Q	R	E	G	C	F	V	W	GG	L	A	L	S	P	ADD	M	L	I	R	V	E	R	P	L	D	S	D	A	Q	L	V	A	S	V	L	S	K	K	I	A	A	G	A				
Hl. AMP-P	210	FSMAE	I	E	S	I	V	A	E	A	N	G	C	L	V	W	GG	V	D	L	S	P	V	DD	A	I	R	A	E	N	P	L	S	I	D	P	G	Q	L	M	A	S	V	L	S	K	K	R	S	A	G	S		
H. SB29. AMP-P	211	FSVEE	I	R	I	V	G	E	T	G	G	C	L	V	W	GG	A	V	N	L	S	P	V	DD	R	I	R	A	E	T	P	L	S	I	D	P	R	G	Q	L	I	A	S	V	L	S	K	K	K	S	A	G	S	
Tp. AMP-P	224	LSIDEI	E	I	E	I	V	E	K	T	N	A	C	L	V	W	GG	A	N	L	A	P	ADD	I	I	R	V	E	Y	P	L	G	I	D	P	F	Y	.	I	P	S	I	L	A	K	K	L	A	V	G	S			
Pbacterium. AMP-P	218	FSAD	E	I	R	Q	I	V	K	K	I	G	A	C	M	V	W	GG	L	N	L	A	P	ADD	K	I	L	K	V	S	Y	P	L	S	L	E	F	Y	S	K	M	L	V	S	I	M	A	K	K	V	A	T	G	V

● ● ● ●

Tk. AMP-P	275	QYVLIDIPTGKGVKVEETVEEARSLARDFIELGKRLGQYVEVAITYGGQPIGHTVGPALAEA
ORF11	182	EKILIDIPTGAGAKIEDNEEARDLAHD FISLGHGLDVEIETATTYGGQPLGYHAGSGLEA
Ta. AMP-P	286	ERLVIDIPAGRGAKVQEVSKADEMAGIFIRQASKMNIALKVAITYGGQPIGLTTGPALEA
ORF10	234	KRCLIDIPVGPQVKVDLRRAKHLRTRFMAIGKMLGMKVKVIFSDGSEPIGYGIGP LLEV
Ab. AMP-P	276	NFMVLDIPMGPETKVPDEKTARKYAMDFIELGERLGITLEAAVTYGGQPIGRAIGPALEA
As. AMP-P	282	KHVVVDIPVGEYAKVTSLDVVKELAHKFSELGRRFGLNVTSVITNARQPVGRAIGPALEA
Av. AMP-P	277	KHVVVDIPVGEGRSRIATSEDRGLANDFSELGRRGLNVSCVLTYSQPVGRAIGPALEA
Mt. AMP-P	276	DKVVVDIPVGDGTKIEDRKEGSKLARDFIEIGERLDMETECAITYGGSPIGRAVGPALAEA
Ms. AMP-P	276	DVVAIDIPV GK KYC K V N S I E E G R K L A R Q F I D L G E R L N M R V E C A L T Y G A S P V G R T V G P A L E V
Mp. AMP-P	275	QKVVVDIPTGSGTKIPDIKAGRKLRDLISLGERLGMVDCAITYGASPVGRTVGPALAEV
ORF08	296	THVLIDVPLGPTAKVVRTSDLARLRLLLEEVARAFGMHVLVVHTDGSQPVGRGIGPALEA
ORF06	273	THSVIDVPLGPTAKVRSADADYECLKQMLEQVAQAFDLHLQVVRTDGTQPVGRGIGPSLEA
H1. AMP-P	270	THVVVDIPYGEGAKVESLADARELADDIKRVGAHLDVAVTCAITHGTEPIGHGVGPLLEA
H. SB29. AMP-P	271	THVVVDIPYGEGAKVESLAEARELAEDFNRVGSHLGMAVECAITNGVAVPVRGVGPVLEA
Tp. AMP-P	282	THVVLDVPTGRGTKVKTLLEAKRISQSFFEIARMFGMNLQAVATYAEPIGHAIGPALEA
Pbacterium. AMP-P	278	THLIIDMPVGPPTTKIPNMKIARSLLEAKFKYLARRFHIKIKVVMIRTEDPVGRGVGPALAEA

★

Tk. AMP-P	335	REALSALIMT.G.KGPGSLIEKATGLAGILLEMGGVAPAGTGKKMAKEILESCKAWEKMKKE
ORF11	242	REAIKTLQGG...NGP TSLIEKATS LAGILIEMSGSASYGTGKEKALDILKSGKAYEKLLD
Ta. AMP-P	346	REALSALIT.G.KGSRSLVDKALLLAGLVLELSGRVPAGSGEEVAREIFLSGKAYEKFKKE
ORF10	294	QDVMKVLNN.DPDAPADLREKALY MAGILLELGKAVRGTGKMKARRILKKG LALEQMNK
Ab. AMP-P	336	REAMQALLEG.K.RVSHSLIEKATDIAGMLLELGNIAERGEKDMAKDILKSGKAREKFLE
As. AMP-P	342	KEALKTLEE.R.KGSSSLIEKSLGLAGILLEMSSGKTSN..GYEYAKEIFNSGKAYEKLKE
Av. AMP-P	337	REALSALES.A.KGPKTLIEKSIGIAGVLFEMAGKATR..GADYARQILLESKALEKFRF
Mt. AMP-P	336	REALKALEG.K.ETPKSLIEKSCSLSGMLFEMAGTSSKDEGTEKARELLKNGKPLEKFKKE
Ms. AMP-P	336	KEALSVLEGN.DSPRSLIQKSCV IAGIAFELAGKSNRGE GAELALELLKSGKAHKKFLE
Mp. AMP-P	335	IEALKVLET.M.EGPNSLIEKSVV LAGMLLEMGGVAAKGGHRELALLETLKN GKALAKMKKE
ORF08	356	RDVLAVLQGAA.SAPADLRGRALLLSASLMFECGAVPAGQLALATRLLADGAAWAKFQA
ORF06	333	HDVLAVLQRAE.RAPEDLSVRAVALAGQLLEFCGHSEPGTGVLVAQRLLDSGAAWAKFQA
H1. AMP-P	330	RDVLAVLHG...DGPESLRLKSVRLAEMLLTHCGVDAS...AADLLES GAALDRFR
H. SB29. AMP-P	331	REVLATLAG...GGPNLRLVKAIRLADLLFESAGVDAD...AAEILD SGTARETFRE
Tp. AMP-P	342	REALIALRE.L.RPGDLVDKAASLAGTLLEMVGVEN...GYETAMEALRTGKAEKKLRE
Pbacterium. AMP-P	338	RDVLRVLQQRD.NYPTDLANKSIHLAGELLELTGKARKGQGAHAAWQALDSGAAWKMQE

Tk. AMP-P	393	IIEAQGGDPNIKPEEIPIGDKTYTFTAATS GYVTAIDNRAITAIARAAGAPEDKGAGI E L
ORF11	299	IIEAQGGDRNVKVDLPLGDKRET IKAPRDGYVQRIYNLSIRDIAVAAGAPQDKGAGLKL
Ta. AMP-P	404	IIEAQGGDPNVKPEEIPIGKHSYTLKSPLEGAVTHLDNAAITTLARACGAPYDKGAGVYL
ORF10	353	IIDAQGRKRMKPNARL...IHKIKSGKEGTVKHIDNKAIAKLARIAGAPLEAAAGVYL
Ab. AMP-P	394	IINAQGSRGIEKSEDVPIGKYKAD IHSPEEGYISIVSNKALVKIARTAGAPKDKGAGIIL
As. AMP-P	398	IILAQGGE.ITKADDVPIGDKVYTLKATKEGAVTHVRNDVIVKIARTAGAPKDKGAGVYL
Av. AMP-P	393	IIEAQGGNPDVKSDDIAVGDRTYTVTSNIDGAI SRLDNVMIKKVARAAGAPKDKGAGIML
Mt. AMP-P	394	IIEAQGGNPDITSE DIEI GQHKETLTAPSEGYVTSVNMKSIVNIARALGSPSDDKAGLEI
Ms. AMP-P	394	IIEVQGGSPDIKSDQINPGEHFYIVRADSNGYVIDLNNHSLISIARTAGAPADHGAGVYL
Mp. AMP-P	393	IIEIQGGNPNVTHK DIPV GKFTADLKAPTNGYILEFYNKRIVQIARLAGAPNDKGAGVMI
ORF08	415	ICEAQGGLHQPGSAPL...RREILAPADGIVTSDNRRLSRAAKLAGAPNRKEAGIDM
ORF06	392	ICKAQGGLREPQRARL...REPVLAERDGMVREIDSRRLARVAKLAGAPKAPAGLEL
H1. AMP-P	381	IVAAQDGDPIEPDLELPGSESTTVQADRAGVVTSVDNRQLCDLARRAGAPKDSRAGLVI
H. SB29. AMP-P	382	ILAAQNGDPDVVATDLVPGRHTHTARADRDGVITHINNRLVNEIARRAGAPRDPAGI E L
Tp. AMP-P	396	IIEAQGGDPDVTP E I P L G D K T Y T L Y S E E D G F V Y Y I D N S L L A N I G K I A G A P I D K G A G V Y I
Pbacterium. AMP-P	397	IIEAQGGNPNVTPS DIVIAAHKKYINAPRSGRIIFTDNKTIINV IARILGAPADKLAGVYL

```

Tk. AMP-P      453 YVKVGEKVKEGDPLFTIHAEHEARRLDQAIVLARRTEPTRIEG.....M...VLQRIGN..
ORF11         359 FHKEGQEVKKGDPLIHIYAEHEKKLDDALTIAKRNPPFRIEG.....M...LLNRFSGEP
Ta. AMP-P     464 HAKVGYRVNKGDPLITLYSNSASRLETAVNLSQTQPIIVEG.....M...LLKTLP...
ORF10        408 RKKLGEKVKKNEVLFEIHSNNDDRLGFAKTAKQNSGYTVVKX.....
Ab. AMP-P     454 NKKKSEKTDKGEVLYTIYADSKTKLDEAVKIAKMLKPLKIEG.....M...LLERIPSYK
As. AMP-P     457 HKKRGEVVKVGDPLITIYAEKEWKLDNAIDVALQEKPVEISG.....M...IFETYPSY.
Av. AMP-P     453 HKKGQHVKVGDPLFTTIYAEKEWKLDRAVELAMREPPVVVSG.....M...ILERWPSYK
Mt. AMP-P     454 CKKGDEVCKGDPLIHMYAEEKWKLKEGMKEAKKHFPLRVEG.....I...VLERIPE..
Ms. AMP-P     454 HAKHGTSLSKGDPLIFTIYADRKWRLEKAIEEARRLRPVMVEG.....M...LIDRVPNIR
Mp. AMP-P     453 HKKRGEAVEEGQPVLTIYAEKEEKLKDAIKSAREFLPIVVEG.....M...LLEKVADIT
ORF08        470 HVRLNDAVRAGQPLFTIHALAQGELAYSQDFLATHPAIHIG.....ATEQ.....
ORF06        447 HVKLGDRVERGMPLFTVHAEALGELDYAFDYLEAHPLIDVG.....DS.....
H1. AMP-P     441 HTSTADPVELGDDLYTIYAETGSKRAAAETLAADREPIRVRS.....KTDALVEQ.....
H. SB29. AMP-P 442 HHRTGEMAAEGESLFTIHAESPDKLADAVDLTERVEMVRVRH.....PDEALVDR.....
Tp. AMP-P     456 HVKLGEKVRKGDPLLLTVYSSSSSAKLQAVERILEDSKPVLVGRTAGRM...LLERIQYQP
Pbacterium. AMP-P 457 NKEYDDRVKKDERLFTLYARTPERIRLATKALEKMSIFKIGH.....

```

```

Tk. AMP-P      503 .....I
ORF11         411 RS.....K
Ta. AMP-P     .....
ORF10        .....
Ab. AMP-P     506 .....A
As. AMP-P     508 .I.....
Av. AMP-P     505 II.....
Mt. AMP-P     504 .....L
Ms. AMP-P     506 EWTPGRSRTIE
Mp. AMP-P     505 E.....V
ORF08        515 .....R
ORF06        490 .....R
H1. AMP-P     491 .....R
H. SB29. AMP-P 492 .....V
Tp. AMP-P     513 PRQLVLE...R
Pbacterium. AMP-P .....

```

Phylogenetic analysis (Figure 11) revealed that ORFs 06, 08, 10 and 11 probably share a common ancestor of AMP phosphorylases with bootstrapping values of 1, 1, 0.4 and 0.4, respectively. ORFs 06 and 08 derived from ATII-LCL are together forming a separate branch between AMP-P. ORFs 10 (KD-UIN) and 11 (KD-BR) formed separate branches between Crenarchaeote and Euryarchaeote. ORFs 10 and 11 were closely related to Halobacteria and Thermococci, respectively. These findings could assign these ORFs as separate lineage in the archaeal AMP-P but still, need further investigation.

3.3 Analysis of AMP-P structure

3D structure and transmembrane helices were studied for the chosen AMP-P sequences. Differences in structure would give a hint to changes in function or substrate spectrum. As a

transmembrane helix was identified in *Tk*-AMP-P the other sequences were studied for similar characteristics. Furthermore, the number of salt bridges was analyzed. The number of salt bridges increases for enzymes of thermophilic or hyperthermophilic origin (Chan, Yu, & Wong, 2011; Lam, Yeung, Yu, Sze, & Wong, 2011). Hence, the number of salt bridges might give a hint concerning the thermal stability of the enzymes.

Protein modeling and 3D structure prediction of the chosen enzymes were done by Phyre2 web portal (Kelley et al., 2015). 94-99% of enzymes sequences were modeled at >90% confidence with top-ranked template c4ga5H (*Tk*-AMP-PΔC10 apo-form). The best two hits are shown with % identity (Table 6).

Transmembrane helix prediction shows that only ORF11 has a transmembrane helix like *Tk*-AMP-P from residue 82 to 97. It is the equal position as observed for *Tk*-AMP-P (amino acids 175 to 190) (Figure 13). Amino acid numbering differs as ORF11 misses the N-terminal domain as stated before.

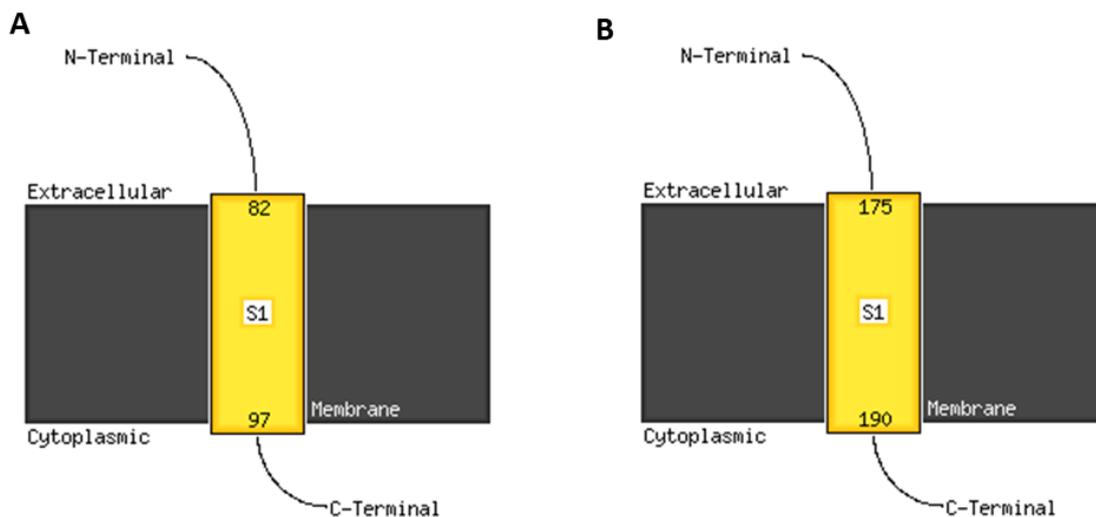


Figure 13. Transmembrane helix prediction by Phyre2 web server using memsat-svm. A. ORF11 from Kebrtit brine pool. B. AMP-P from *Thermococcus kodakarensis*.

The number of salt bridges calculated for each enzyme showed the highest number of 186 for ORF06 from ATII-LCL which confirms with the thermophilic characteristics of the brine pool. However, the lowest number (104 salt bridges) was determined for Am03 identified in *Thermosphaera aggregans*. It contrasts with the hyper-thermophilic characteristic described for this archaeon (Table 6).

Validation of results

To confirm previous analysis work, another approach was followed using the Red sea dataset reads and assembled contigs as mentioned in the materials and methods section. The identified ORFs from this approach was the same as the previous ones which confirmed the previous results.

3.4 Biochemical characterization of the putative AMP-P

Gene synthesis for the identified sequences and inserts were cloned into plasmids pMA-T. From this construct, the inserts were cloned into the expression vector pKS2. Expression conditions were optimized, and proteins were purified by affinity chromatography. The phosphorylase activity of the enzymes was studied against different substrates and analyzed by HPLC.

3.4.1 Gene synthesis and cloning

After the inserts were cloned in pKS2, cloning was validated by a test restriction. Test restriction with restriction enzymes BamHI and HindIII showed expected inserts size on agarose gels for all constructs. Concentrations of the purified recombinant expression plasmids were above 100ng/ μ l except for Am10 (92ng/ μ l) and the DNA purity $A_{260}/_{280}$ was >1.8 which indicates a good purity of the recombinant plasmid (Figure 14, Table 7).

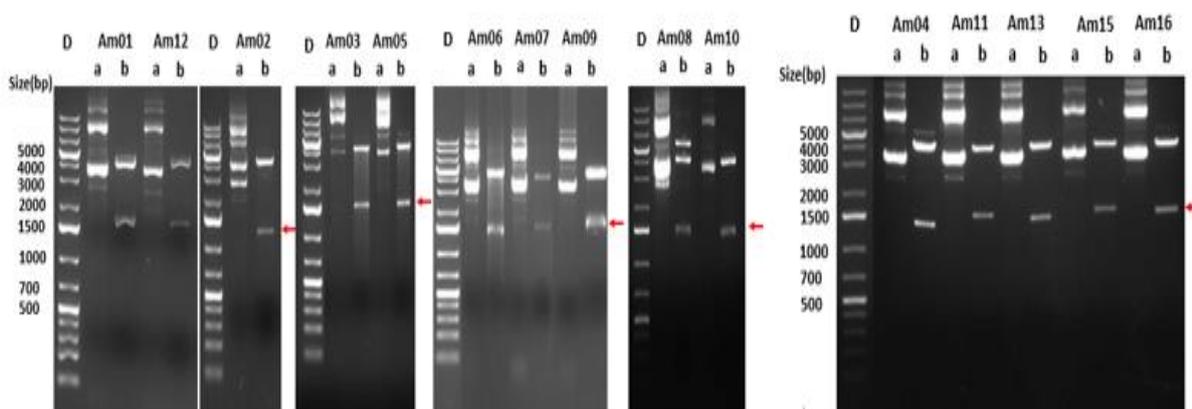


Figure 14. Test Restriction after cloning of the sequences of interest into expression vector pKS2. Lane D: GeneRulerTM 1Kb plus DNA ladder (Thermo Scientific). Agarose gel electrophoresis of the plasmids after test restriction. Inserts are marked with red arrows. **a:** expression constructs before the restriction, **b:** expression construct after restriction with BamHI and HindIII. Inserts are pointed with a red arrow.

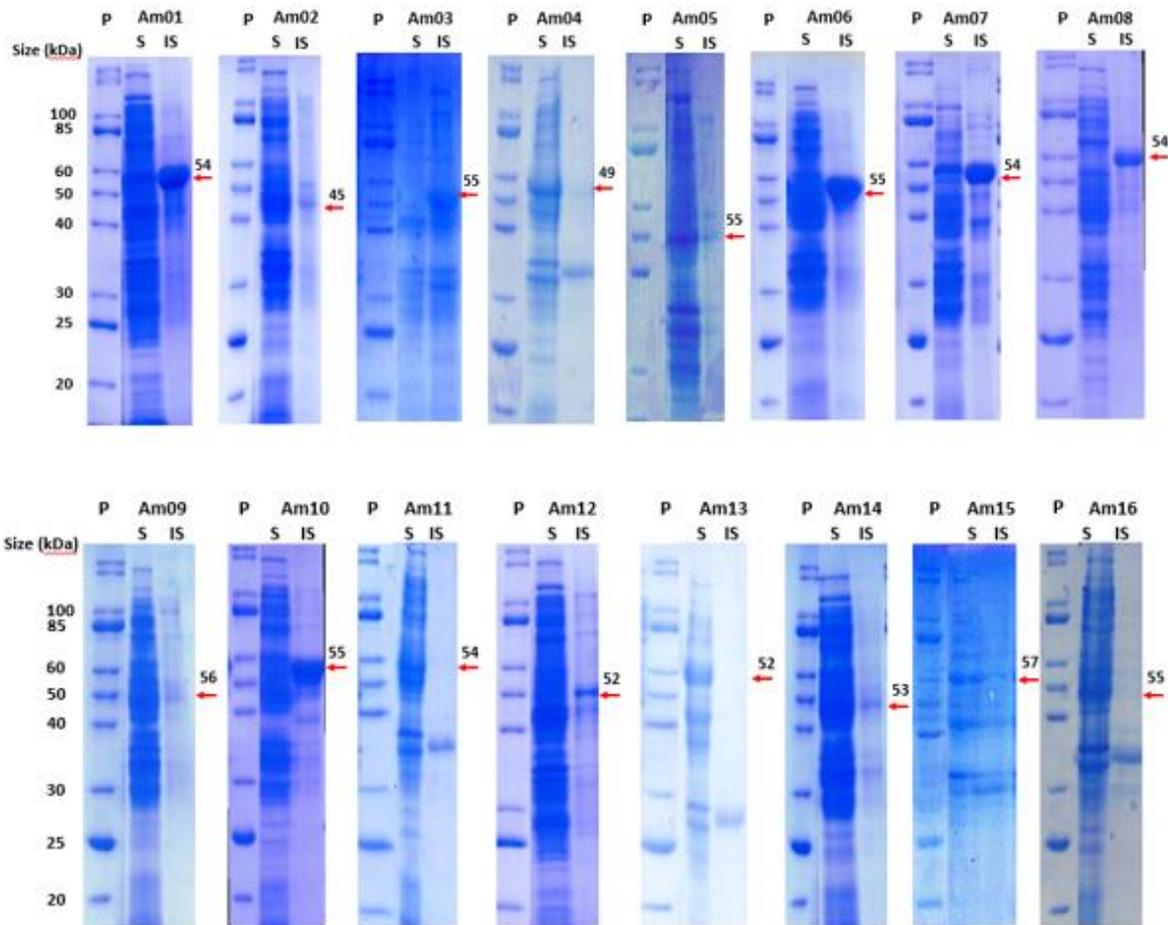
Table 7. Size [bp] of the inserts and pKS2. The concentration of purified recombinant expression plasmids was measured by a spectroscopic measurement at 260 nm by a Nanodrop™ 1000 spectrophotometer (Thermo scientific). Absorbance ratio $A_{260/280}$ was measured to determine DNA purity.

Enzyme	Size [bp]	Purified plasmid conc. (ng/μl)	A _{260/280}
Am01	1512	156.4	1.87
Am02	1242	152.7	1.87
Am03	1539	183.7	1.9
Am04	1348	236.9	1.91
Am05	1521	183	1.89
Am06	1527	222.1	1.93
Am07	1521	246.1	2.01
Am08	1515	570.8	1.85
Am09	1551	256.7	1.96
Am10	1521	92	1.85
Am11	1548	250.3	1.9
Am12	1473	237.5	1.84
Am13	1476	223.6	1.89
Am14	1479	138.3	1.94
Am15	1563	192.7	1.88
Am16	1497	235.9	1.88
pKS2	4136		

3.4.2 Protein expression and purification

N-terminal His-tagged proteins were expressed in EnPresso®B animal free medium. All target proteins were detected in both the soluble and the insoluble fractions. However, the majority of protein went to the insoluble fraction (Figure 15). Protein sizes observed in SDS-PAGE gels were slightly higher than the theoretically calculated molecular weight.

Different expression conditions were applied to optimize the expression and to enhance solubility. Different IPTG concentrations and varying expression temperatures were analyzed. Increased IPTG concentration led to an increased target protein expression. However, most of the protein remained in the insoluble fraction. Protein expression at low temperature (22°C) in the 2xYT medium did not enhance the expression of the target protein to the soluble fraction. In contrast, compared to the expression in EnPresso®B animal free medium the yield of the protein was less (Figure 16). Consequently, induction of protein expression with 1 mM IPTG in EnPresso®B animal free medium was used for further studies as mentioned in materials and methods.



l

Figure 15. Expression of AMP-P in EnPresso[®]B medium with IPTG concentrations of 1 mM.

12% SDS-PAGE gels were loaded with OD₆₀₀ = 5 samples of all AMP-P. Expected sizes of the target proteins are marked with a red arrow. The theoretical sizes of the AMP-P (above red arrow) were calculated using Protparam web tool (Gasteiger et al., n.d.) Lane P: Unstained protein ladder (NEB, range 10-250kDa), S: soluble fraction, IS: insoluble fraction.

N-terminal His-tagged proteins were purified by Ni-NTA spin column affinity chromatography (Figure 17) as mentioned in the materials and methods section. The size of the protein observed after purification in SDS-PAGE gels for Am01, 02, 03, 07 and 12 were slightly higher than the theoretically calculated molecular weight. Am07 show a very intense band for purified protein. In addition, Faint bands appeared at higher (\approx 85-100kDa) than expected molecular weights for most of the studied proteins. Protein concentrations and total protein mass produced from 50 ml *E. coli* BL21 culture together with other parameters validated during expression and purification of proteins are listed in Table 8. Adequate protein concentrations and total protein mass were obtained for all purified protein except

Am16 which show a very low concentration. All proteins were purified with a final protein concentration of 0.23 to 4.37mg/ml. Consequently, this was followed by enzymatic activity and thermal stability testing.

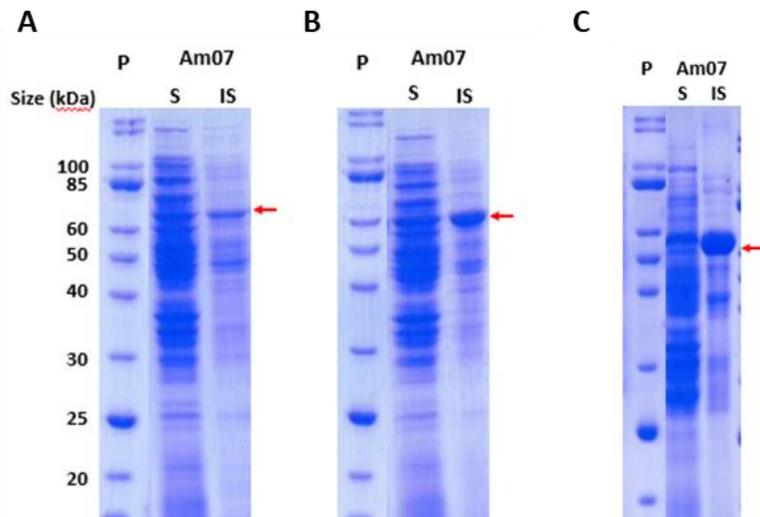


Figure 16. Optimization of expression shown for Am07 as an example. $OD_{600} = 5$ samples were analyzed on 12% SDS-PAGE gels.

A, B) Am07 after expression in the 2xYT medium at 22°C using IPTG conc. of 50µM (A) and 1 mM (B). C) Am07 after expression in EnPresso®B medium at 30°C using IPTG conc. of 1mM. Purified Am07 (54kDa) is marked by red arrows. P: Unstained protein ladder (NEB, range 10-250kDa), S: soluble fraction, IS: insoluble fraction

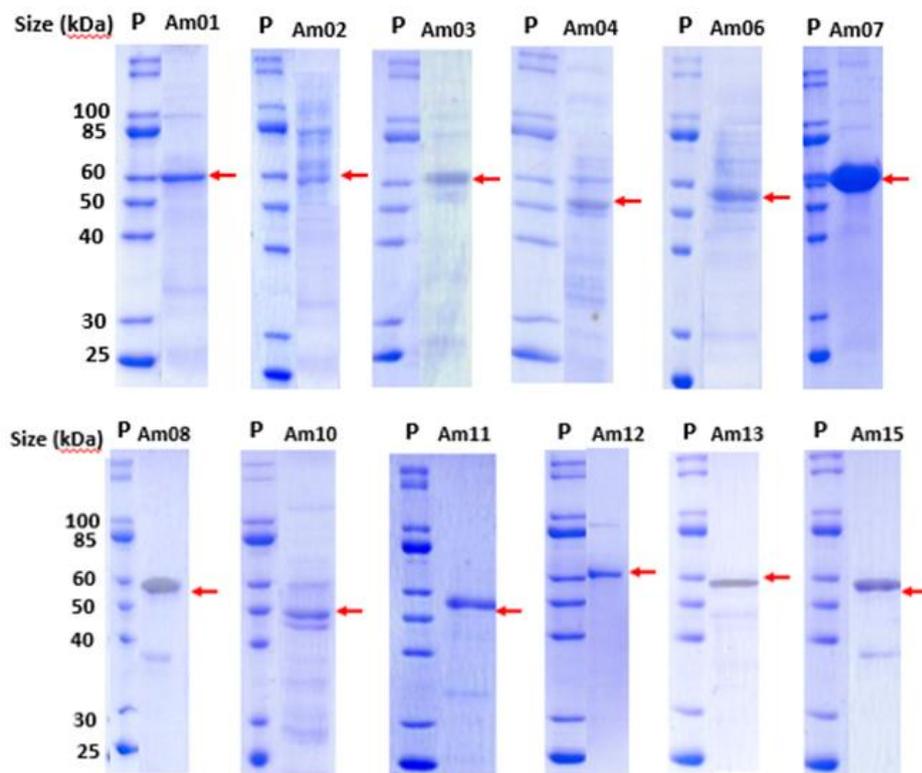


Figure 17. SDS-PAGE after purification of His-tagged AMP-P by Ni-NTA spin column affinity chromatography. Protein code is given above each lane and the size of the protein is marked with red arrows. **P**: Unstained protein ladder (NEB, range 10-250kDa).

Table 8. Parameters validated during expression and purification of AMP-P in EnPresso®B medium with 1mM IPTG conc. OD₆₀₀ measurements before and after the harvest of cells, protein concentration measured by Nanodrop and total protein mass in 50ml culture. Mean and the standard deviation is calculated for replicates.

Enzyme	OD ₆₀₀ -Ind.	OD ₆₀₀ -harvest	Protein concentration [mg/mL]	Total protein [mg]
Am01	11.01 ± 0.77	31.68 ± 11.24	3.42 ± 0.83	2.57
Am02	7.36	26.85	4.37 ± 1.48	3.28
Am03	9.12	29	1.81 ± 1.57	1.36
Am04	6.5	30.85	3.97 ± 2.88	2.98
Am05	11.4	37.75	2.92	2.19
Am06	10.54	31.4	3.34 ± 0.48	2.51
Am07	10.66 ± 0.49	25.97 ± 2.74	2.53 ± 0.72	1.90
Am08	8.46	34.2	2.13 ± 1.77	1.60
Am09	10.64	39.25	2.43	1.82
Am10	9.76	37.8	2.4 ± 1.77	1.80
Am11	9.6	26.95	3.49 ± 0.93	2.62
Am12	8.96 ± 1.63	16.37 ± 5.93	2.57 ± 1.23	1.93
Am13	10.24	32.9	3.42 ± 1.72	2.57
Am14	9.58	32.95	1.58	1.19
Am15	9.345	36.1	1.52 ± 0.85	1.14
Am16	8.72	12.5	0.23	0.17

3.5 Activity assays

The phosphorylase activity of the purified proteins was tested against six substrates including nucleoside monophosphates (AMP, GMP, CMP, and UMP), a purine nucleoside (Adenosine) and a pyrimidine nucleoside (Uridine). The relative activity [%] was determined for each protein and substrate at three different temperatures (37, 50, 80°C) (Figure 18). Only Am01, Am03, Am12 and Am15 showed activity using AMP as substrate. Activity increased with increasing temperature for Am01, Am03, and Am15. Am12 only showed low activity at 50°C. Percentage of conversion was higher for Am01 and Am03 compared to Am12 and Am15.

With GMP, Am01, Am03, and Am15 showed a temperature-dependent phosphorylase activity (18.1, 17.3, 17.1 %, respectively). Am04, Am06, Am07, Am08, Am11, Am12 showed phosphorylase activity against GMP at 37°C with 16.9, 18.8, 11.1, 19, 17.2, 14.8 %, respectively. Activity decreased dramatically at higher temperatures.

Activity against CMP was insignificant for almost all enzymes. Am01 and Am15 only showed very low activity at 50°C reaching percentages of conversion of 10.4 and 6.8 %, respectively.

With UMP only Am01, Am03, and Am15 showed phosphorylase activity which also increases with increasing temperature. Final conversion values of 29, 13.7 and 17 %, respectively, were reached at 80°C. All enzymes show good activity against Adenosine and Uridine.

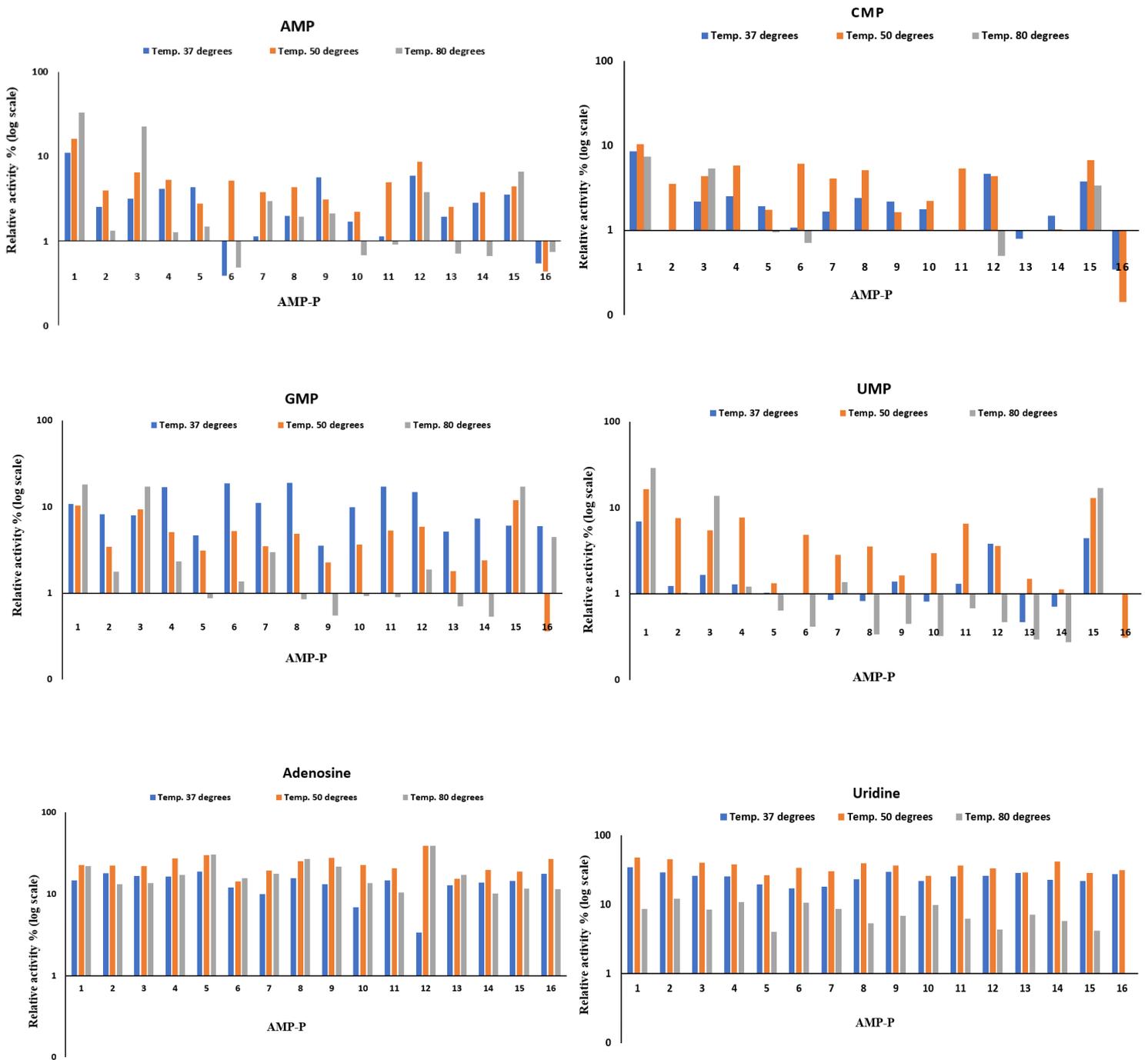


Figure 18. Relative phosphorylase activity of AMP-P. AMP-P activity at three different temperatures 37, 50 and 80°C was studied with six different substrates (AMP, CMP, GMP, UMP, Adenosine, and Uridine).

3.6 Thermostability of the heterologous expressed AMP-P

Analysis of heat resistance was performed for all purified proteins except for Am05, Am09, Am14 and Am16 as they showed only very low or no activity in the activity assays.

Incubation times of 1h (70°C) and 24h (60°C, 70°C, 80°C) were studied.

After incubation of the purified protein at increased temperatures for 24hr, Am07, Am08, Am10, and Am15 resisted denaturation until temperatures up to 70°C. Am12 and Am13 showed faint bands after incubation at 60°C. However, the protein bands possessed a higher molecular weight of approximately 100kDa. None of the proteins were resistant to 80°C (Figure 19).

Using incubation times of 1 h at 70°C all tested proteins were stable (Figure 19). The observed protein bands showed the expected sizes. An additional band appeared at a higher molecular weight (≈ 100 kDa), which might be corresponding to the dimer form of the protein.

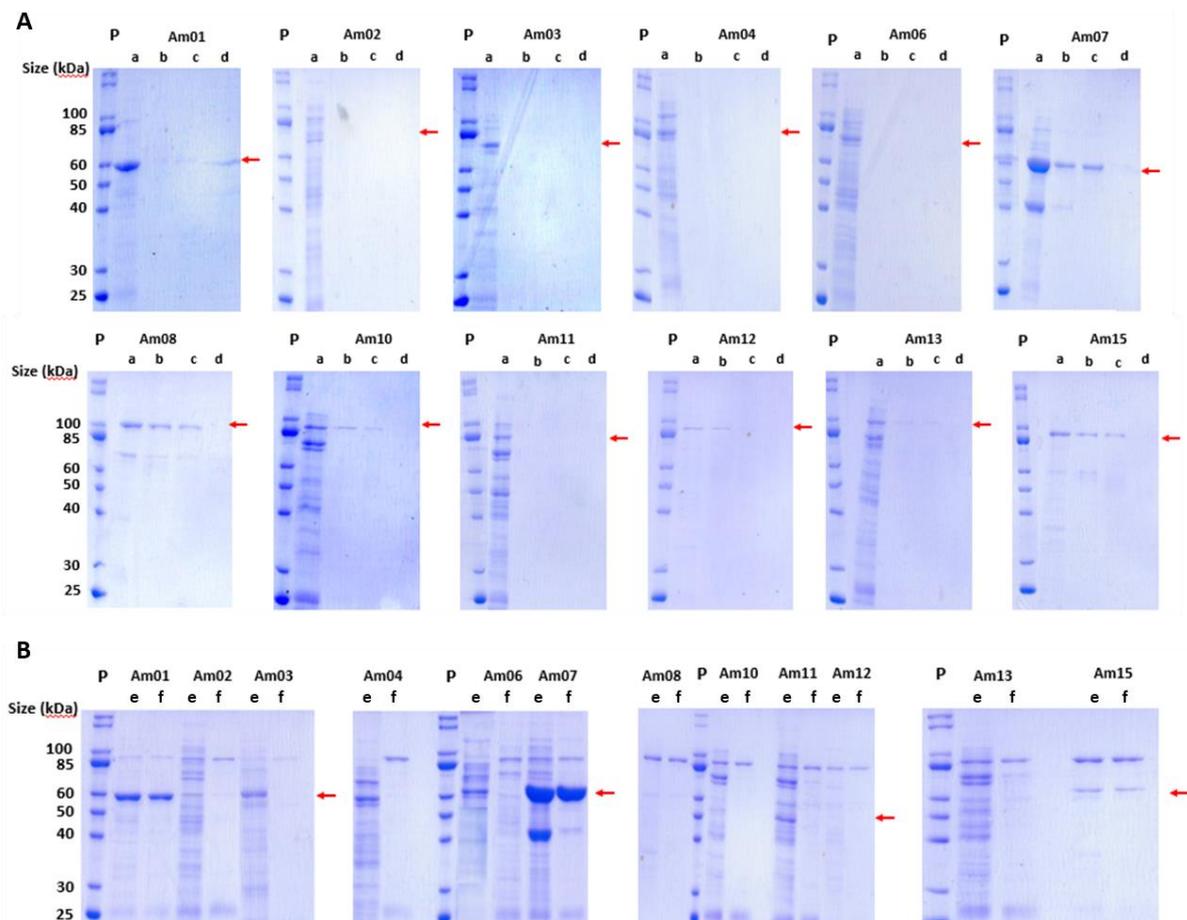


Figure 19. Analysis of heat resistance of the purified AMP-Ps on 12% SDS-PAGE gels. Stability was analyzed after incubation of 24hr (60, 70, 80°C, **A**) and 1hr (70°C, **B**). Target proteins are marked with a red arrow.

P: Unstained protein ladder (NEB, range 10-250kDa). **a:** purified protein with no heat, **b:** purified protein after incubating at 60°C, **c:** purified protein after incubating at 70°C and **d:** purified protein after incubating at 80°C. **e:** a purified protein with no heat, **f:** purified protein after incubating at 70°C for 1hr.

3.6.1 Activity of heat-treated AMP-P

The activity of the heat-treated AMP-P was analyzed. The activity assay was performed with three substrates AMP, CMP, and uridine. The activity was either analyzed after 1h and 16h (uridine) or at 16h only (AMP, CMP). Am01, Am03, and Am15 showed activity against AMP (Figure 20). Activity increased by increasing the temperature. Compared to non-heat-treated samples the relative activity was decreased. No activity was shown against CMP at any temperature. Using uridine as substrate, no activity was observed after 1h. However, after 16h of incubation activity for all tested proteins was measured. Highest activities were observed at 50°C. These results confirm with the previous observations that Am03 and Am15 are putative AMP phosphorylases.

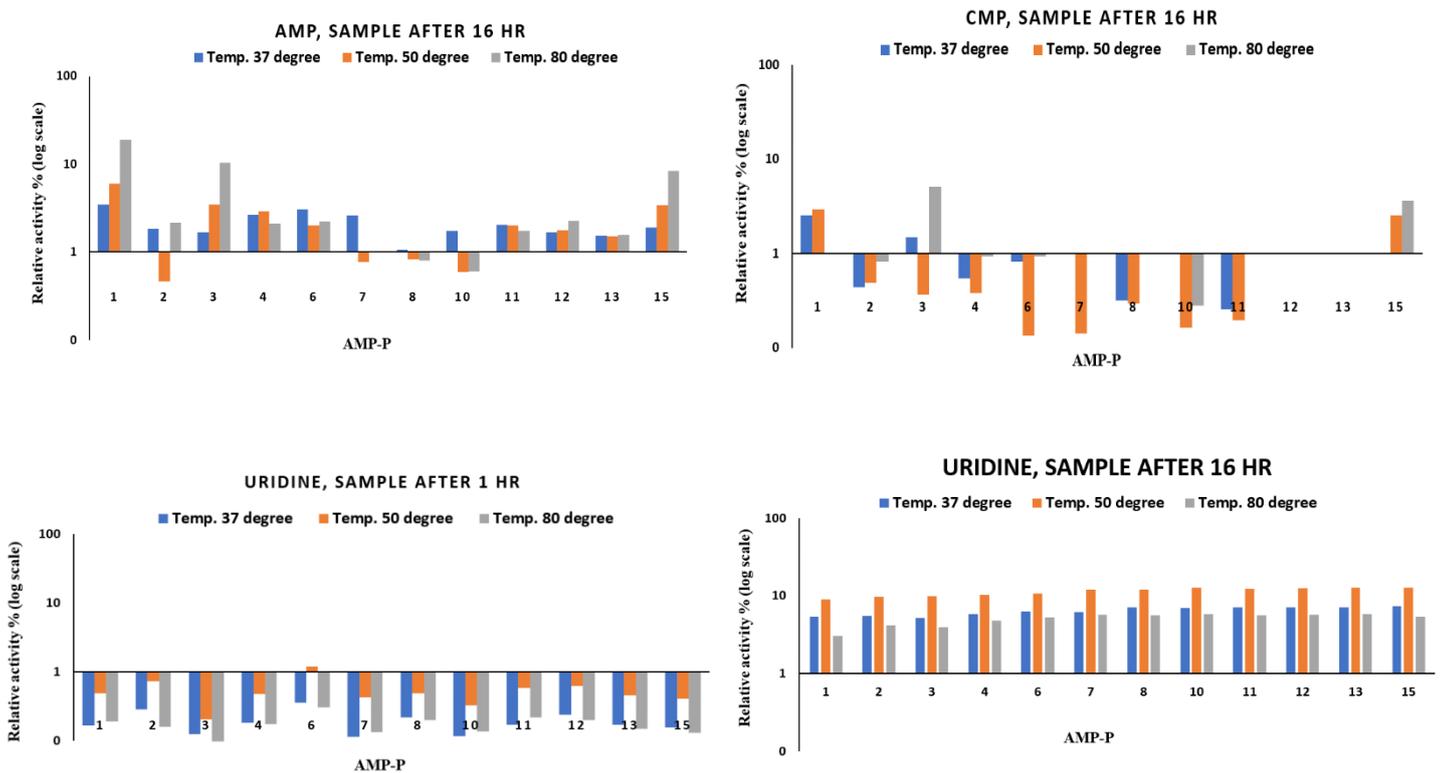


Figure 20. Relative phosphorylase activity of AMP phosphorylases after heat-treatment (1h, 70°C). AMP phosphorylase activity at three different temperatures 37, 50 and 80°C was studied. AMP, CMP, and uridine were used as substrates

Chapter Four: Discussion

The original aim of this study was the identification of novel PyNPs from the Red Sea metagenomic datasets. Therefore, a sequence based metagenomic approach was done using Pfam HMM profiles of PyNP functional domains (α/β domain, α domain, and C-terminal domain) against Red sea metagenomic assembled datasets. Surprisingly, few hits were reached so another strategy was used using a thermophilic PyNP custom-made database retrieved from NCBI. HMM profile was built from the multiple sequence alignment of the custom-made database and used to screen for novel PyNP in the Red Sea metagenomic dataset.

Eleven ORFs were identified to have the three essential functional domains of thymidine/pyrimidine nucleoside phosphorylase family. These ORFs belong to the glycosyltransferase superfamily. Interestingly, two ORFs (ORF 10,11; from Kebrit brine pool upper interface and brine) showed a high similarity to annotated AMP phosphorylase from the archaeal origin.

Moreover, the archaeal phylogenetic tree that was built to study PyNP showed that thymidine phosphorylases of archaeal origin are re-annotated as AMP-P, with sequences longer than the expected 430 to 440 amino acids. A deeper search in the literature revealed only a few reports that focus on AMP-P; consequently, this study focused on AMP-P of mainly archaeal origin.

Based on the literature, AMP-P is a unique enzyme as it is the first enzyme to be identified to catalyze a phosphorolysis reaction on nucleotides which differs from other identified phosphorylases that act on nucleosides. Moreover, AMP-P has a role in a novel pathway involved in nucleic acid metabolism and is widely distributed among archaea (Aono et al., 2012; Nishitani et al., 2013; Sato et al., 2007).

Phylogenetic analysis of the eleven ORFs in comparison to ninety-six archaeal AMP-P reference sequences of the NCBI protein database showed that only four ORFs of the Red Sea metagenomic data clustered within AMP-P. ORFs 06 and 08 from ATII-LCL formed a separate branch between AMP-P and TP/PyNP. ORF 10 from KD-UIN formed a separate branch between Crenarchaeote and Euryarchaeote (Halobacteria) and ORF 11 from KD-BR also formed a separate branch between Crenarchaeote and Euryarchaeote (Thermococci). These findings could assign these ORFs as separate lineage in archaeal AMP-P but still, need further investigation.

Previous studies showed that *Tk*-AMP-P has active residues responsible for its function. Phosphate-binding residues at the C-terminal domain involve amino acids Ser165, Asn175, Lys191, Ser193, Ser194, and Thr203. Purine and pyrimidine binding sites located in the central domain. Important residues for purine binding are Asp256 and Ser264. The pyrimidine-binding site is characterized by amino acids Glu250, Ser26, and Lys268. 5'-phosphate binding loop which is responsible for binding to 5'-phosphate of NMP is found in the C-terminal domain and important active residues are Gly168, Ile197, Thr198, Ser199 & Lys288 (Nishitani et al., 2013).

AMP-P has a unique sequence called N-terminal domain (82 residues) which is essential for multimerization as AMP-P form large macromolecules (>40mers) in solution and this multimerization is crucial for the thermostability of the enzyme from a hyperthermophile (Nishitani et al., 2013). Moreover, the N-terminal domain is essential to the enzymatic activity of AMP-P as it is stabilizing the closed conformation upon substrate binding.

Multiple sequence alignment of the selected 15 genes with *Tk*-AMP-P as positive control revealed that all ORFs selected from the Red Sea showed all active residues responsible for the activity. Other identified genes show some differences mainly in the phosphate-binding residues of their encoded proteins side from residues of *Tk*-AMP-P as mentioned in Table 6.

Moreover, Phyre2 results for 3D structure prediction revealed that 94-99% of enzymes sequences were modeled at >90% confidence with top-ranked template c4ga5H (*Tk*-AMP-P Δ C10 apo-form). Am02 from Kebrit brine has a transmembrane helix at same positions as AMP-P from *Tk* however, it lacks the 82 residues of the N-terminal domain which is responsible for AMP-P thermostability, enzymatic activity, and multimerization.

Purified expression recombinant plasmids show a good DNA purity ($A_{260/280}$ is >1.8) and high concentration above 100ng/ μ l except for Am10 was 92ng/ μ l, which indicates a good purity of the recombinant plasmid. Optimization of expression using 1 mM IPTG and incubation at 30°C in EnPresso[®]B animal free medium was done however, most of the protein was in the insoluble fraction. Despite that most of the protein retained in the insoluble fraction, adequate protein concentrations and total protein mass were obtained for all purified protein using Ni-NTA spin column affinity chromatography except for Am16 which show a very low concentration (0.27mg/ml).

Based on this preliminary study of phosphorylase activity against different substrates; Am03 and Am15 are putative AMP phosphorylases as they show significant activity toward AMP, GMP, and UMP; increases with increasing temperature which conforms with the hyper-thermophilic characteristics of their isolation Archaea. Therefore, further analysis against different substrates is needed to identify their substrate spectrum and confirms their function and structure. Although Am12 from ATII-LCL show low activity towards AMP and GMP with no activity toward UMP and CMP, it shows significant activity toward Adenosine and Uridine. It could be either pyrimidine or purine nucleoside phosphorylase, but further analysis is needed to confirm.

Furthermore, all enzymes show good activity against Adenosine and Uridine. Despite what mentioned before (Aono et al., 2012) that AMP-P from *Tk* (Am01, positive control) doesn't show any activity against Adenosine and Uridine. Relative activity percentage showed that there is a phosphorylase activity against these two substrates with the highest activity at 50°C reaching 22.6% with Adenosine and 47.4 % with Uridine.

In addition, Am02 was lacking activity against NMPs substrate. This could be due to it lacks the N-terminal domain which is unique to AMP-P enzymatic activity and responsible for its thermostability. However, this might be also a problem with the sequencing or assembly of reads. Further analysis should be done as it is a promising ORF and if it turns to be a problem with assembly or sequences and the whole ORF is identified that give rise to exceptional AMP phosphorylase from extreme hypersaline environment which characterized by high H₂S content.

The activity assay of the heat-treated AMP-P with the three substrates AMP, CMP and uridine revealed that Am03 and Am15 showed activity against AMP which increased by increasing the temperature; however, the relative activity is decreased compared to non-heat treated samples. No activity was shown against CMP at any temperature. These results confirm with the previous observations that Am03 and Am15 are putative AMP phosphorylases

Chapter Five: Conclusion

AMP phosphorylase is a unique enzyme that demonstrated activity towards nucleoside monophosphates. It is highly similar to nucleoside phosphorylase which acts on nucleosides. It is involved in a novel metabolic pathway for nucleic acid metabolism that is widely distributed among archaea. Being an attractive and promising biocatalyst for the synthesis of nucleoside monophosphates and their analogs, it is of high interest to search for novel AMP-P with broad substrate specificity.

In this study, interesting AMP-P were identified from both metagenomic data of Red sea brine pools and the NCBI database. HMMsearch of the metagenomic data revealed eleven ORFs that belong to TP or AMP-P with significantly low E-values based on BlastX and Interpro analysis. Two ORFs (ORF 10,11) from Kebrit brine pool showed a high similarity to annotated AMP-P from the archaeal origin.

Phylogenetic analysis of the eleven ORFs with ninety-six archaeal AMP-P reference sequences of the NCBI database showed that only four ORFs of the Red Sea metagenomic data clustered within AMP-P. ORFs 06 and 08 from ATII-LCL formed a separate branch between AMP-P and TP/PyNP. ORF 10 from KD-UIN formed a separate branch between Crenarchaeote and Euryarchaeote (Halobacteria) and ORF 11 from KD-BR also formed a separate branch between Crenarchaeote and Euryarchaeote (Thermococci). These findings could assign these ORFs as separate lineage in archaeal AMP-P but still, need further investigation.

Fifteen candidates were chosen for further studies as they showed distinct differences compared to the only characterized AMP-P of *Tk*. AMP-P of *Tk* was cloned as a positive control. Phyre2 results for 3D structure prediction revealed that 94-99% of enzymes sequences were modeled at >90% confidence with top-ranked template c4ga5H (*Tk*-AMP-P Δ C10 apo-form). Am02 from kebrit brine has a transmembrane helix at same positions as AMP-P from *Tk*.

All 16 putative AMP-P were successfully cloned in the expression vector pKS2 after gene synthesis. After optimization of expression, all enzymes were successfully expressed in *E. coli* BL21. However, most of the protein went to the insoluble fraction. Nevertheless, all proteins were purified with a final protein concentration of 0.23 to 4.37mg/ml.

Activity assays were performed with six substrates (4 NMPs, 2 nucleosides). For AMP-P of *Tk* (Am01) published data were confirmed. Additionally, Am03 and Am15 showed similar substrate spectra (an activity with AMP, GMP, and UMP) and are putative AMP-P. Their activity increased with increasing temperature which is in good accordance with the temperature optimum of the donor organisms (*Thermosphaera aggregans*, *Thermofilum pendens*).

The other identified proteins could be putative pyrimidine/purine nucleoside phosphorylase as they show a phosphorylase activity against adenosine and uridine. Am12 from ATII-LCL is an interesting candidate to be further analyzed as it is derived from an extreme environment, has activity towards adenosine and uridine nucleosides and low activity against AMP.

Thermal stability was determined for all purified proteins except for Am05, Am09, Am14 and Am16 as they showed only very low or no phosphorylase activity in the activity assays. After incubation of the purified proteins at increased temperatures (60°C, 70°C, 80°C) for 24hr; only Am07, Am08, Am10 and Am15 did not denature until temperatures up to 70°C. Am12 and Am13 showed faint bands after incubation at 60°C. None of the proteins were resistant to 80°C

After incubation for 1h at 70°C, all tested proteins were stable and can withstand the heat treatment. However, an additional band appeared at a higher molecular weight ($\approx 100\text{kDa}$), which might be corresponding to the dimer form of the protein.

The activity assay of the heat-treated AMP-P with the three substrates AMP, CMP and uridine were evaluated. Am01, Am03, and Am15 showed activity against AMP which increased by increasing the temperature however, the relative activity is decreased compared to non-heat treated samples. No activity was shown against CMP at any temperature. These results confirm with the previous observations that Am03 and Am15 are putative AMP-P.

Chapter six: Outlook

The availability of AMP-P enzymes with a wide substrate spectrum is an important prerequisite to produce a variety of modified nucleotides in enzymatic processes. Therefore, screening for novel AMP-Ps from metagenomic database and extremophiles is an interesting area of research to be studied.

In this preliminary study we identified two putative AMP phosphorylases from a hyperthermophilic archaeon; Am03 from *Thermosphaera aggregans* and Am15 from *Thermofilum pendens*. These promising AMP-Ps need further analysis against different substrates, especially substrates with modification on the sugar or base moiety to identify its broad spectrum. It would thus be of interest to determine the currently unknown mechanism by which AMP phosphorylase alone recognizes various bases, sugars and 5'-phosphate groups of nucleotides. In addition, further investigation to assess pH optimum, salt tolerance and kinetic parameters should be done. Native-PAGE and size exclusion chromatography also should be assessed to identify the number of subunits of the active protein.

Finally, other identified proteins could be a putative pyrimidine/purine nucleoside phosphorylase as they show a phosphorylase activity against Adenosine and Uridine. Am12 from ATII-LCL is an interesting candidate to be further analyzed as it is derived from an extreme environment, has activity towards Adenosine and Uridine and low activity against AMP.

Appendix

Supplementary materials

Supplementary Table 1. AMP phosphorylases retrieved from NCBI RefSeq. used for generating the Archaeal phylogenetic tree

AMP phosphorylase organism	Protein length (aa)	Accession no.	Archaeal taxonomy
<i>Thermococcus gammatolerans</i>	503	WP_015859397.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Pyrococcus abyssi</i>	503	WP_048146577.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Pyrococcus
<i>Thermococcus sibiricus</i>	503	WP_015848965.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Thermococcus onnurineus</i>	503	WP_012572022.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Methanococcus maripaludis</i>	505	WP_012193341.1	Archaea; Euryarchaeota; Methanomada group; Methanococci; Methanococcales; Methanococcaceae; Methanococcus.
<i>Methanococcoides burtonii</i>	506	WP_011498424.1	Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales; Methanosarcinaceae; Methanococcoides.
<i>Thermococcus kodakarensis</i>	503	WP_011249307.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Pyrococcus furiosus</i>	503	WP_011012753.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Pyrococcus
<i>Pyrococcus horikoshii</i>	503	WP_010885672.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Pyrococcus
<i>Archaeoglobus fulgidus</i>	504	WP_010878838.1	Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales; Archaeoglobaceae; Archaeoglobus.
<i>Methanocaldococcus vulcanius</i>	503	WP_012819702.1	Archaea; Euryarchaeota; Methanococci; Methanococcales; Methanocaldococcaceae; Methanocaldococcus.
<i>Methanocaldococcus bathoardescens</i>	503	WP_048202190.1	Archaea; Euryarchaeota; Methanococci; Methanococcales; Methanocaldococcaceae; Methanocaldococcus.
<i>Methanocaldococcus fervens</i>	503	WP_015791943.1	Archaea; Euryarchaeota; Methanococci; Methanococcales; Methanocaldococcaceae; Methanocaldococcus.
<i>Thermococcus peptonophilus</i>	503	WP_062387581.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Thermococcus guaymasensis</i>	503	WP_062370636.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Thermococcus celericrescens</i>	503	WP_058939554.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Thermococcus eurythermalis</i>	503	WP_050003764.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Methanosarcina thermophila</i>	506	WP_048166173.1	Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales; Methanosarcinaceae; Methanosarcina.
<i>Palaeococcus pacificus</i>	503	WP_048164909.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Palaeococcus.
<i>Thermococcus barophilus</i>	503	WP_048159765.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Palaeococcus ferrophilus</i>	503	WP_048151637.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Palaeococcus.
<i>Methanosarcina horonobensis</i>	506	WP_048139654.1	Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales; Methanosarcinaceae; Methanosarcina.
<i>Methanolobus tindarius</i>	507	WP_048135496.1	Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales; Methanosarcinaceae; Methanolobus.
<i>Methanosarcina lacustris</i>	506	WP_048126007.1	Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales; Methanosarcinaceae; Methanosarcina.
<i>Thermococcus nautili</i>	503	WP_042690675.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Thermococcus paralvinellae</i>	503	WP_042681022.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.

Continuation of **Supplementary Table 1.**

<i>Thermococcus cleftensis</i>	503	WP_014789454.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Pyrococcus yayanosii</i>	503	WP_013905230.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Pyrococcus
<i>Methanosalsum zhilinae</i>	506	WP_013897981.1	Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales; Methanosarcinaceae; Methanosalsum
<i>Archaeoglobus veneficus</i>	506	WP_013682953.1	Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales; Archaeoglobaceae; Archaeoglobus.
<i>Thermococcus zilligii</i>	503	WP_010477092.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Thermococcus litoralis</i>	503	WP_004069717.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Methanobolus psychrotolerans</i>	506	WP_094228514.1	Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales; Methanosarcinaceae; Methanobolus.
<i>Methanohalophilus portucalensis</i>	506	WP_072359909.1	Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales; Methanosarcinaceae; Methanohalophilus.
<i>Methanohalophilus euhalobius</i>	506	WP_096711719.1	Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales; Methanosarcinaceae; Methanohalophilus.
<i>Methanosarcina spelaei</i>	506	WP_095642796.1	Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales; Methanosarcinaceae; Methanosarcina.
<i>Methanobolus profundus</i>	506	WP_091932766.1	Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales; Methanosarcinaceae; Methanobolus.
<i>Thermococcus gorgonarius</i>	503	WP_088884543.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Thermococcus barossii</i>	503	WP_088864953.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Thermococcus celer</i>	503	WP_088862927.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Thermococcus profundus</i>	503	WP_088858967.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Thermococcus siculi</i>	503	WP_088856786.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Thermococcus pacificus</i>	503	WP_088854723.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Thermococcus chitonophagus</i>	503	WP_068578781.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Thermococcus thio reducens</i>	503	WP_055429706.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.
<i>Methanoregula boonei</i>	508	WP_012106062.1	Archaea; Euryarchaeota; Methanomicrobia; Methanomicrobiales; Methanoregulaceae; Methanoregula.
<i>Methanocella arvoryzae</i>	503	WP_012036091.1	Archaea; Euryarchaeota; Methanomicrobia; Methanocellales; Methanocellaceae; Methanocella.
<i>Methanococcus aeolicus</i>	504	WP_011973334.1	Archaea; Euryarchaeota; Methanococci; Methanococcales; Methanococcaceae; Methanococcus.
<i>Hyperthermus butylicus</i>	502	WP_011821569.1	Archaea; Euryarchaeota; Methanococci; Methanococcales; Methanocaldococcaceae; Methanocaldococcus.
<i>Methanocaldococcus jannaschii</i>	503	WP_010870172.1	Archaea; Euryarchaeota; Methanococci; Methanococcales; Methanocaldococcaceae; Methanocaldococcus.
<i>Methanothermococcus okinawensis</i>	504	WP_013867394.1	Archaea; Euryarchaeota; Methanococci; Methanococcales; Methanococcaceae; Methanothermococcus.
<i>Haloferax</i> sp. Q22	492	WP_058828152.1	Archaea; Euryarchaeota; Halobacteria; Haloferacales; Haloferacaceae; Haloferax.
<i>Haloferax</i> sp. SB29	492	WP_058573114.1	Archaea; Euryarchaeota; Halobacteria; Haloferacales; Haloferacaceae; Haloferax.
<i>Methanoculleus sediminis</i>	507	WP_048180143.1	Archaea; Euryarchaeota; Methanomicrobia; Methanomicrobiales; Methanomicrobiaceae; Methanoculleus.
<i>Methanoplanus limicola</i>	512	WP_048146198.1	Archaea; Euryarchaeota; Methanomicrobia; Methanomicrobiales; Methanomicrobiaceae; Methanoplanus.
<i>Geoglobus ahangari</i>	505	WP_048094337.1	Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales; Archaeoglobaceae; Geoglobus.

Continuation of **Supplementary Table 1.**

<i>Geoglobus acetivorans</i>	505	WP_048090314.1	Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales; Archaeoglobaceae; Geoglobus.
<i>Methanosaeta harundinacea</i>	514	WP_014587167.1	Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales; Methanosaetaceae; Methanosaeta.
<i>Methanocella conradii</i>	503	WP_014404868.1	Archaea; Euryarchaeota; Methanomicrobia; Methanocellales; Methanocellaceae; Methanocella.
<i>Methanotorrus igneus</i>	505	WP_013798979.1	Archaea; Euryarchaeota; Methanococci; Methanococcales; Methanocaldococcaceae; Methanotorrus.
<i>Methanocaldococcus infernus</i>	501	WP_013100164.1	Archaea; Euryarchaeota; Methanococci; Methanococcales; Methanocaldococcaceae; Methanocaldococcus.
<i>Ferroglobus placidus</i>	503	WP_012965322.1	Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales; Archaeoglobaceae; Ferroglobus.
<i>Archaeoglobus profundus</i>	506	WP_012940517.1	Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales; Archaeoglobaceae; Archaeoglobus.
<i>Halogranum salarium</i>	492	WP_009365914.1	Archaea; Euryarchaeota; Halobacteria; Haloferacales; Haloferacaceae.
<i>Haloferax elongans</i>	492	WP_008323542.1	Archaea; Euryarchaeota; Halobacteria; Haloferacales; Haloferacaceae; Haloferax.
<i>Haloferax mucosum</i>	492	WP_008321385.1	Archaea; Euryarchaeota; Halobacteria; Haloferacales; Haloferacaceae; Haloferax.
<i>Natronorubrum sulfidifaciens</i>	492	WP_008163149.1	Archaea; Euryarchaeota; Halobacteria; Natrialbales; Natrialbaceae; Natronorubrum.
<i>Aciduliprofundum boonei</i>	506	WP_008084769.1	Archaea; Euryarchaeota; DHVE2 group; Aciduliprofundum
<i>Haloterrigena limicola</i>	491	WP_008012656.1	Archaea; Euryarchaeota; Halobacteria; Natrialbales; Natrialbaceae; Haloterrigena.
<i>Haloferax larsenii</i>	492	WP_007539008.1	Archaea; Euryarchaeota; Halobacteria; Haloferacales; Haloferacaceae; Haloferax.
<i>Methanolinea tarda</i>	508	WP_007314415.1	Archaea; Euryarchaeota; Methanomicrobia; Methanomicrobiales; Methanoregulaceae; Methanolinea.
<i>Haloferax sulfurifontis</i>	492	WP_007274769.1	Archaea; Euryarchaeota; Halobacteria; Haloferacales; Haloferacaceae; Haloferax
<i>Natronolimnobius innermongolicus</i>	492	WP_007260948.1	Archaea; Euryarchaeota; Halobacteria; Natrialbales; Natrialbaceae; Natronolimnobius.
<i>Methanotorrus formicicus</i>	505	WP_007044211.1	Archaea; Euryarchaeota; Methanococci; Methanococcales; Methanocaldococcaceae; Methanotorrus.
<i>Natronorubrum bangense</i>	492	WP_006066791.1	Archaea; Euryarchaeota; Halobacteria; Natrialbales; Natrialbaceae; Natronorubrum.
<i>Methanocaldococcus villosus</i>	501	WP_004593508.1	Archaea; Euryarchaeota; Methanococci; Methanococcales; Methanocaldococcaceae; Methanocaldococcus
<i>Natrialba magadii</i>	492	WP_004216326.1	Archaea; Euryarchaeota; Halobacteria; Natrialbales; Natrialbaceae; Natrialba
<i>Archaeoglobus sulfatcallidus</i>	508	WP_015591340.1	Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales; Archaeoglobaceae; Archaeoglobus.
<i>Haloterrigena hispanica</i>	491	WP_092930949.1	Archaea; Euryarchaeota; Halobacteria; Natrialbales; Natrialbaceae; Haloterrigena
<i>Halorientalis persicus</i>	493	WP_092657549.1	Archaea; Euryarchaeota; Halobacteria; Halobacteriales; Haloarculaceae; Halorientalis
<i>Methanonatronarchaeum thermophilum</i>	504	WP_086637497.1	Archaea; Euryarchaeota; Methanonatronarchaeia; Methanonatronarchaeum.
<i>Methanospirillum stamsii</i>	516	WP_109940965.1	Archaea; Euryarchaeota; Methanomicrobia; Methanomicrobiales; Methanospirillaceae; Methanospirillum
<i>Salinigranum rubrum</i>	492	WP_103424990.1	Archaea; Euryarchaeota; Halobacteria; Haloferacales; Halorubraceae; Salinigranum.
<i>Pyrococcus kukulkanii</i>	503	WP_068323649.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Pyrococcus
<i>Methanoculleus horonobensis</i>	507	WP_067078709.1	Archaea; Euryarchaeota; Methanomicrobia; Methanomicrobiales; Methanomicrobiaceae; Methanoculleus
<i>Methanoculleus thermophilus</i>	507	WP_066957334.1	Archaea; Euryarchaeota; Methanomicrobia; Methanomicrobiales; Methanomicrobiaceae; Methanoculleus
<i>Thermococcus piezophilus</i>	503	WP_068664075.1	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae; Thermococcus.

Continuation of **Supplementary Table 1.**

Methanofollis ethanolicus	515	WP_067048152.1	Archaea; Euryarchaeota; Methanomicrobia; Methanomicrobiales; Methanomicrobiaceae; Methanofollis.
Desulfurococcus amylolyticus	509	WP_014767048.1	Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales; Desulfurococcaceae; Desulfurococcus
Pyrodictium delaneyi	509	WP_055410511.1	Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales; Pyrodictiaceae; Pyrodictium.
Thermofilum pendens	520	WP_011751768.1	Archaea; Crenarchaeota; Thermoprotei; Thermoproteales; Thermofilaceae; Thermofilum.
Thermogladius calderae	509	WP_014737920.1	Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales; Desulfurococcaceae; Thermogladius.
Pyrodictium occultum	504	WP_058370070.1	Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales; Pyrodictiaceae; Pyrodictium.
Thermosphaera aggregans	512	WP_052891646.1	Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales; Desulfurococcaceae; Thermosphaera.
Thermofilum uzonense	510	WP_052883489.1	Archaea; Crenarchaeota; Thermoprotei; Thermoproteales; Thermofilaceae; Thermofilum
Staphylothermus hellenicus	509	WP_013143293.1	Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales; Desulfurococcaceae; Staphylothermus.

Figure licenses

License for Figure 1, 3 and 4

License Details

This Agreement between The American University in Cairo – Noha Attallah ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Print Copy

License Number	4600970216472
License date	Jun 02, 2019
Licensed Content Publisher	Elsevier
Licensed Content Publication	Journal of Molecular Biology
Licensed Content Title	Structure Analysis of Archaeal AMP Phosphorylase Reveals Two Unique Modes of Dimerization
Licensed Content Author	Yuichi Nishitani,Riku Aono,Akira Nakamura,Takaaki Sato,Haruyuki Atomi,Tadayuki Imanaka,Kunio Miki
Licensed Content Date	Aug 9, 2013
Licensed Content Volume	425
Licensed Content Issue	15
Licensed Content Pages	13
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	5
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	figures 1, 5(a), 6, 7(c,d), 9
Title of your thesis/dissertation	Identification, expression and biochemical characterization of AMP phosphorylases from extreme environments
Expected completion date	Jun 2019
Estimated size (number of pages)	60
Requestor Location	The American University in Cairo New Cairo Campus, AUC Avenue P.O. Box 74 New Cairo Cairo, 11835 Egypt Attn: The American University in Cairo
Publisher Tax ID	GB 494 6272 12
Total	0.00 USD

License for Figure 5, 6, and 7

License Details

This Agreement between The American University in Cairo – Noha Attallah ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

[Print](#) [Copy](#)

License Number	4600971052596
License date	Jun 02, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Drug Discovery
Licensed Content Title	Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases
Licensed Content Author	Lars Petter Jordheim, David Durantel, Fabien Zoulim, Charles Dumontet
Licensed Content Date	May 31, 2013
Licensed Content Volume	12
Licensed Content Issue	6
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	4
High-res required	no
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	Identification, expression and biochemical characterization of AMP phosphorylases from extreme environments
Institution name	The American University in Cairo
Expected presentation date	Jun 2019
Portions	Figures 1, 2, 3, 4
Requestor Location	The American University in Cairo New Cairo Campus, AUC Avenue P.O. Box 74 New Cairo Cairo, 11835 Egypt Attn: The American University in Cairo
Total	0.00 USD

License for Figure 8

License Details

This Agreement between The American University in Cairo – Noha Attallah ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

[Print](#) [Copy](#)

License Number	4600980867133
License date	Jun 02, 2019
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	ChemBioChem
Licensed Content Title	Five-Component Cascade Synthesis of Nucleotide Analogues in an Engineered Self-Immobilized Enzyme Aggregate
Licensed Content Author	Robert A. Scism, Brian O. Bachmann
Licensed Content Date	Dec 23, 2009
Licensed Content Volume	11
Licensed Content Issue	1
Licensed Content Pages	4
Type of Use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 1b
Will you be translating?	No
Title of your thesis / dissertation	Identification, expression and biochemical characterization of AMP phosphorylases from extreme environments
Expected completion date	Jun 2019
Expected size (number of pages)	60
Requestor Location	The American University in Cairo New Cairo Campus, AUC Avenue P.O. Box 74 New Cairo Cairo, 11835 Egypt Attn: The American University in Cairo
Publisher Tax ID	EU826007151
Total	0.00 USD

References

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST : a new generation of protein database search programs, *25*(17), 3389–3402.
- Aono, R., Sato, T., Yano, A., Yoshida, S., Nishitani, Y., Miki, K., ... Atomi, H. (2012). Enzymatic characterization of amp phosphorylase and ribose-1,5-bisphosphate isomerase functioning in an archaeal amp metabolic pathway. *Journal of Bacteriology*, *194*(24), 6847–6855. <https://doi.org/10.1128/JB.01335-12>
- Barai, V. N., Kvach, S. V., Zinchenko, A. I., & Mikhailopulo, I. A. (2004). An improved method for the enzymatic transformation of nucleosides into 5'-monophosphates. *Biotechnology Letters*, *26*(24), 1847–1850. <https://doi.org/10.1007/s10529-004-5311-4>
- Bhatia, H. K., Singh, H., Grewal, N., & Natt, N. K. (2014). Sofosbuvir: A novel treatment option for chronic hepatitis C infection. *Journal of Pharmacology & Pharmacotherapeutics*, *5*(4), 278–284. <https://doi.org/10.4103/0976-500X.142464>
- Boogaerts, M. A., Van Hoof, A., Catovsky, D., Kovacs, M., Montillo, M., Zinzani, P. L., ... Klein, M. (2001). Activity of Oral Fludarabine Phosphate in Previously Treated Chronic Lymphocytic Leukemia. *Journal of Clinical Oncology*, *19*(22), 4252–4258. <https://doi.org/10.1200/JCO.2001.19.22.4252>
- Burgess, K., & Cook, D. (2000). Syntheses of nucleoside triphosphates. *Chemical Reviews*, *100*(6), 2047–2059. <https://doi.org/10.1021/cr990045m>
- Caton-Williams, J., Smith, M., Carrasco, N., & Huang, Z. (2011). Protection-free one-pot synthesis of 2'-deoxynucleoside 5'-triphosphates and DNA polymerization. *Organic Letters*, *13*(16), 4156–4159. <https://doi.org/10.1021/ol201073e>
- Chan, C., Yu, T., & Wong, K. (2011). Stabilizing Salt-Bridge Enhances Protein Thermostability by Reducing the Heat Capacity Change of Unfolding, *6*(6). <https://doi.org/10.1371/journal.pone.0021624>
- Clercq, E. de, & E., D. C. (2016). Approved antiviral drugs over the past 50 years. *Clinical Microbiology Reviews*, *29*(3), 695–747. <https://doi.org/10.1128/CMR.00102-15.Address>
- Costantini, S., Colonna, G., & Facchiano, A. M. (2008). Bioinformatics ESBRI : A web server for evaluating salt bridges in proteins Bioinformatics. *Biomedical Informatics*, *3*(2), 137–138.
- Eddy, S. R. (1998). Bioinformatics-1998-Eddy-755-63, 755–763.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., ... Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, *47*(D1), D427–D432. <https://doi.org/10.1093/nar/gky995>
- Elbehery, A. H. A., Leak, D. J., & Siam, R. (2017). Novel thermostable antibiotic resistance enzymes from the Atlantis II Deep Red Sea brine pool. *Microbial Biotechnology*, *10*(1), 189–202. <https://doi.org/10.1111/1751-7915.12468>
- Ferreira, A. J. S., Siam, R., Setubal, J. C., Moustafa, A., Sayed, A., Chambergo, F. S., ... El-Dorry, H. (2014). Core microbial functional activities in ocean environments revealed by global metagenomic profiling analyses. *PLoS ONE*, *9*(6). <https://doi.org/10.1371/journal.pone.0097338>

- Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, J., ... Mitchell, A. L. (2017). InterPro in 2017 — beyond protein family and domain annotations, *45*(November 2016), 190–199. <https://doi.org/10.1093/nar/gkw1107>
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (n.d.). Protein Identification and Analysis Tools on the ExpASY Server, 571–608.
- Jordheim, L. P., Durantel, D., Zoulim, F., & Dumontet, C. (2013). Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. *Nature Reviews Drug Discovery*, *12*(6), 447–464. <https://doi.org/10.1038/nrd4010>
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*, 845. Retrieved from <https://doi.org/10.1038/nprot.2015.053>
- Kumar, L., Awasthi, G., & Singh, B. (2011). Extremophiles: A novel source of industrially important enzymes. *Biotechnology*. <https://doi.org/10.3923/biotech.2011.121.135>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4, *227*, 680–685.
- Lam, S. Y., Yeung, R. C. Y., Yu, T. H., Sze, K. H., & Wong, K. B. (2011). A rigidifying salt-bridge favors the activity of thermophilic enzyme at high temperatures at the expense of low-temperature activity. *PLoS Biology*, *9*(3). <https://doi.org/10.1371/journal.pbio.1001027>
- Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4 : recent updates and, 2–5. <https://doi.org/10.1093/nar/gkz239>
- Littlechild, J. A. (2015). Enzymes from Extreme Environments and Their Industrial Applications. *Frontiers in Bioengineering and Biotechnology*, *3*(October), 1–9. <https://doi.org/10.3389/fbioe.2015.00161>
- Lyseng-Williamson, K. A., Reynolds, N. A., & Plosker, G. L. (2006). Tenofovir Disoproxil Fumarate. *Drugs*, *65*(3), 413–432. <https://doi.org/10.2165/00003495-200565030-00006>
- Mohamed, Y. M., Ghazy, M. A., Sayed, A., Ouf, A., El-Dorry, H., & Siam, R. (2013). Isolation and characterization of a heavy metal-resistant, thermophilic esterase from a Red Sea Brine Pool. *Scientific Reports*, *3*, 1–8. <https://doi.org/10.1038/srep03358>
- Moretti, S., Armougom, F., Wallace, I. M., Higgins, D. G., Jongeneel, C. V., & Notredame, C. (2007). The M-Coffee web server: A meta-method for computing multiple sequence alignments by combining alternative alignment methods. *Nucleic Acids Research*, *35*(SUPPL.2), 645–648. <https://doi.org/10.1093/nar/gkm333>
- Nishitani, Y., Aono, R., Nakamura, A., Sato, T., Atomi, H., Imanaka, T., & Miki, K. (2013). Structure analysis of archaeal AMP phosphorylase reveals two unique modes of dimerization. *Journal of Molecular Biology*, *425*(15), 2709–2721. <https://doi.org/10.1016/j.jmb.2013.04.026>
- Pugmire, M. J., & Ealick, S. E. (1998). The crystal structure of pyrimidine nucleoside

- phosphorylase in a closed conformation. *Structure*, 6(11), 1467–1479.
[https://doi.org/10.1016/S0969-2126\(98\)00145-2](https://doi.org/10.1016/S0969-2126(98)00145-2)
- Pugmire, M. J., & Ealick, S. E. (2002). Phosphorylases. *Society*, 25, 1–25.
- Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., ... Venter, J. C. (2007). The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biology*, 5(3), 0398–0431.
<https://doi.org/10.1371/journal.pbio.0050077>
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A., & Barrell, B. (2000). Artemis: Sequence visualization and annotation. *Bioinformatics*, 16(10), 944–945. <https://doi.org/10.1093/bioinformatics/16.10.944>
- Sato, T., Atomi, H., & Imanaka, T. (2007). Archaeal type {III} {RuBisCOs} function in a pathway for {AMP} metabolism. *Science*, 315(5814), 1003–1006.
- Sayed, A., Ghazy, M. A., Ferreira, A. J. S., Setubal, J. C., Chambergo, F. S., Ouf, A., ... El-dorry, H. (2014). A Novel Mercuric Reductase from the Unique Deep Brine Environment of Atlantis II in the Red Sea □, 289(3), 1675–1687.
<https://doi.org/10.1074/jbc.M113.493429>
- Scism, R. A., & Bachmann, B. O. (2010). Five-component cascade synthesis of nucleotide analogues in an engineered self-immobilized enzyme aggregate. *ChemBioChem*, 11(1), 67–70. <https://doi.org/10.1002/cbic.200900620>
- Shah, A., Priyadarshini, M., Khan, M. S., Aatif, M., Amin, F., Tabrez, S., ... Bano, B. (2015). Differential effects of anti-cancer and anti-hepatitis drugs on liver cystatin. *Saudi Journal of Biological Sciences*, 22(1), 69–74.
<https://doi.org/10.1016/j.sjbs.2014.06.006>
- Siam, R., Mustafa, G. A., Sharaf, H., Moustafa, A., Ramadan, A. R., Antunes, A., ... Dorry, H. El. (2012). Unique prokaryotic consortia in geochemically distinct sediments from red sea Atlantis II and discovery deep brine pools. *PLoS ONE*, 7(8).
<https://doi.org/10.1371/journal.pone.0042872>
- Sonbol, S. A., Ferreira, A. J. S., & Siam, R. (2016). Red Sea Atlantis II brine pool nitrilase with unique thermostability profile and heavy metal tolerance. *BMC Biotechnology*, 16(1), 6–8. <https://doi.org/10.1186/s12896-016-0244-2>
- Uckun, F. M., Pendergrass, S., Venkatachalam, T. K., Qazi, S., & Richman, D. (2002). Stampidine is a potent inhibitor of zidovudine- and nucleoside analog reverse transcriptase inhibitor-resistant primary clinical human immunodeficiency virus type 1 isolates with thymidine analog mutations. *Antimicrobial Agents and Chemotherapy*, 46(11), 3613–3616. <https://doi.org/10.1128/AAC.46.11.3613-3616.2002>
- Wallace, I. M., O’Sullivan, O., Higgins, D. G., & Notredame, C. (2006). M-Coffee: Combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Research*, 34(6), 1692–1699. <https://doi.org/10.1093/nar/gkl091>
- Zhang, Y., Cottet, S. E., & Ealick, S. E. (2004). Structure of Escherichia coli AMP nucleosidase reveals similarity to nucleoside phosphorylases. *Structure*, 12(8), 1383–1394. <https://doi.org/10.1016/j.str.2004.05.015>
- Ziko, L., Ouf, A., Aziz, R. K., & Siam, R. (2019). Antibacterial and anticancer activities of orphan biosynthetic gene clusters from Atlantis II Red Sea brine pool. *Microbial Cell*

Factories, 1–16. <https://doi.org/10.1186/s12934-019-1103-3>