# Structural and functional characterization of putative translation initiation factors and their homologs in archaea

*A Thesis Submitted in Partial Fulfillment of the*

*Requirement for the Degree of*

**Doctor of Philosophy**

By

*Prerana Gogoi*


**Department of Biosciences and Bioengineering**

**Indian Institute of Technology Guwahati**

**Guwahati, Assam-781039, India**

**April 2019**

# DECLARATION

I hereby declare that the work reported in this thesis entitled "***Structural and functional characterization of putative translation initiation factors and their homologs in archaea***" is entirely original and was carried out by me under the supervision of Dr. Shankar Prasad Kanaujia, Indian Institute of Technology Guwahati, India.

I also declare that the contents of this thesis have not been the basis for the award of any degree, diploma, fellowship or any other similar title of any University or Institution.

In keeping with the general practice of reporting scientific observations, due acknowledgements have been made wherever the work of other investigators are referred.

Prerana Gogoi

Place: IIT GUWAHATI

Prerana Gogoi

Date: 02.04.2019

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Guwahati-781039, Assam, India

# Indian Institute of Technology Guwahati
## Department of Biosciences and Bioengineering
## Guwahati-781 039

Shankar Prasad Kanaujia, Ph.D.
Associate Professor

Phone: 0361 258 2228
Email: spkanaujia@iitg.ac.in


## CERTIFICATE

This is to certify that the work described in the thesis entitled "***Structural and functional characterization of putative translation initiation factors and their homologs in archaea***" is the result of investigations carried out by Prerana Gogoi at the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Assam, India under my supervision and the results presented in this thesis have not previously formed the basis for the award of any other diploma, degree or fellowship.

Place: IIT GUWAHATI
Date: 02-04-2019

Shankar Prasad Kanaujia
Dr. Shankar Prasad Kanaujia

# ACKNOWLEDGEMENTS

As I am about to end another chapter of my career, I would like to express my gratitude towards the people who have helped me construct it into a successful one.

First and foremost, I am extremely grateful and indebted to my thesis supervisor ***Dr. Shankar Prasad Kanaujia*** for introducing and guiding me through the wonderful world of crystallography. I am thankful to him for always making himself available for scientific discussions and responding eagerly to my problems throughout my Ph.D. tenure. He always encouraged me to do better and bigger in science. It is his mentorship that has moulded me into the person that I am today.

I would like thank my doctoral committee members ***Prof. Vikash Kumar Dubey***, ***Dr. Anil Mukund Limaye***, ***Dr. Vibin Ramakrishnan*** and ***Dr. Manish Kumar*** for their valuable suggestions, motivation and scientific guidance which always helped me to make my work better.

I would like convey my gratitude to the ***Department of Biosciences and Bioengineering***, IIT Guwahati for providing me all the necessary facilities and a favorable academic environment to pursue my work. I also owe my gratitude to the ***Central Instruments Facility (CIF)***, IITG for providing me access to sophisticated instruments including the X-ray diffractometer.

I sincerely acknowledge the financial support from ***Ministry of Human Resource Development (MHRD)***, Government of India for providing me fellowship and the ***Department of Biotechnology (DBT)***, Government of India for providing research grants to our laboratory.

I would take this opportunity to thank my past and present laboratory mates. I owe special thanks to ***Monika*** and ***Suraj*** for their help and support during my Ph.D. tenure. It was a privilege to work and learn together with them. I am also thankful to ***Angshu***, ***Sayan***, ***Pratik***, ***Preeti***, ***Arunabh***, ***Akash***, ***Smriti***, ***Reshama***, ***Shalinee***, ***Ambuj***, ***Kanchan***, ***Tarini***, ***Arpita*** and ***Pratap*** for their cooperation and timely help. I would like to express

my gratitude to ***Prerana Mordina*** who made significant contributions to my work by helping me during the purification and crystallization experiments of two of the proteins.

Although ‘thank you’ would be too small a word for the following people, I would still like to express my deep gratitude to my ***family*** and my friend ***Hasnahana Chetia*** for their unconditional love and support. Despite the fact that sometimes work would become my priority, they were always understanding and encouraging. I am extremely grateful to Hasnahana for always standing by my side through my ugly and beautiful. Her trust in my abilities always encouraged me to do better both scientifically and personally.

# TABLE OF CONTENTS


**List of abbreviations** i-iii

**Abstract** v-xiv

**List of figures** xv-xviii

**List of tables** xix

**Chapter 1 Protein translation initiation in archaea: a chimera of bacterial and eukaryal attributes**

1.1 Protein translation 1

1.1.1 Protein translation initiation in bacteria 2

1.1.2 Protein translation initiation in eukarya 4

1.1.3 Protein translation initiation in archaea 6

1.1.3.1 Eukaryotic translation initiation factor 1 6

1.1.3.2 Eukaryotic translation initiation factor 2B 9

1.1.3.3 Ribose-1,5-bisphosphate isomerase 12

1.1.3.4 5-methylthioribose 1-phosphate isomerase 13

1.1.3.5 NDP-sugar pyrophosphorylase 15

1.2 Rationale of the study 16

1.3 Objectives 17

**Chapter 2 Materials and methods**

2.1 Procurement of materials 19

2.2 Construction of expression plasmids 20

2.3 Protein over expression and purification 22

2.4 Protein crystallization 23

2.5 X-ray intensity data collection and processing 24

2.6 Structure solution 26

2.7 Structure refinement 26

2.8 Model building 26

2.9 Cross-validation 27

2.10 Structure validation 27

2.11 Isothermal titration calorimetry 28

2.12 Analysis of sequences and structures 28
2.12.1 Sequence data and analysis 29
2.12.2 Sequence alignment 29
2.12.3 Phylogenetic tree 29
2.12.4 Secondary structure elements 30
2.12.5 Structural comparison 30
2.12.6 Electrostatic potentials and oligomeric state 30
2.12.7 Structure visualization and analysis 30
2.13 Molecular dynamics simulation 31
2.14 Molecular docking 32
**Chapter 3 *In silico* characterization of putative translation initiation factors in *Pyrococcus horikoshii* OT3**
Abstract 34
3.1 Introduction 35
3.2 Materials and methods 36
3.3 Results 37
3.3.1 *In silico* characterization of homologs of eukaryotic translation initiation factors in archaea 37
3.3.2 *In silico* characterization of PH1771.1 38
3.3.2.1 Evolutionary relationship 38
3.3.3 *In silico* characterization of PH0208 and PH0702 40
3.3.3.1 Evolutionary relationship 40
3.3.3.2 Comparison of the tertiary structures 42
3.3.3.3 Repertoire of residues at the active site 44
3.3.4 *In silico* characterization of PH1022 and PH1697 47
3.3.4.1 PhNSPases belonging to the promiscuous subgroup IC 51
3.3.4.2 Evolutionary relationship 52
3.3.4.3 The nucleotidyltransferase domain 55
3.3.4.4 Repertoire of residues at the active site 57

3.4 Discussion 67
3.5 Conclusion 68
**Chapter 4 Structural insights into the function of archaeal translation initiation factor 1**
Abstract 70
4.1 Introduction 71
4.2 Materials and methods 72
4.2.1 Cloning, over expression and protein purification 72
4.2.2 Crystallization, X-ray intensity data collection and data processing 74
4.2.3 Structure solution and model building 75
4.2.4 Sequence and structure analysis 77
4.3 Results 78
4.3.1 The overall structure 78
4.3.2 The acidic and basic loops of aIF1 79
4.3.3 The charge distribution of aIF1 81
4.4 Discussion 82
4.5 Conclusion 84
**Chapter 5 Structural implications of the enzyme ribose-1,5-bisphosphate isomerase**
Abstract 86
5.1 Introduction 87
5.2 Materials and methods 89
5.2.1 Cloning, over expression and protein purification 89
5.2.2 Protein crystallization 93
5.2.3 Data collection, processing and structure determination 95
5.2.4 Isothermal titration calorimetry 104
5.2.5 Multiple sequence alignment 104
5.3 Results 104
5.3.1 The overall structure 104

5.3.2 The active-site pocket 107
5.3.3 The binding affinity of substrate/product 109
5.3.4 The AMP and GMP binding sites 111
5.3.5 Comparison between PhR15Pi and eIF2Bα 115
5.4 Discussion 118
5.5 Conclusion 123
**Chapter 6 Structure-function relationship of ribose-1,5-bisphosphate isomerase**
Abstract 125
6.1 Introduction 126
6.2 Materials and methods 128
6.2.1 Molecular dynamics simulation 128
6.2.2 Identification of invariant water molecules 130
6.2.3 Molecular docking studies 132
6.3 Results 132
6.3.1 The hexameric state of R15Pi 132
6.3.2 Role of catalytic residue in structural stability 134
6.3.3 Invariant water molecules 137
6.3.4 The 'initial binding site' of substrate 148
6.3.5 Sliding of substrate from 'initial binding site' to the active site 154
6.3.6 Essential amino acid residues at the active-site pocket 155
6.4 Discussion 156
6.4.1 Proposed mechanism of 'substrate sliding' 160
6.5 Conclusion 162
**Chapter 7 Structural insights into the catalytic mechanism of 5-methylthioribose 1 phosphate isomerase**
Abstract 164
7.1 Introduction 165
7.2 Materials and methods 167
7.2.1 Cloning, over expression and protein purification 167

| | | | |
|---|---|---|---|
| | 7.2.2 | Crystallization, X-ray intensity data collection and processing | 169 |
| | 7.2.3 | Structure solution and model building | 173 |
| | 7.2.4 | Sequence and structure analysis | 174 |
| | 7.2.5 | Molecular docking | 174 |
| 7.3 | Results | | 175 |
| | 7.3.1 | The overall structure | 175 |
| | 7.3.2 | The open and closed conformations | 178 |
| | 7.3.3 | The N-terminal domain | 179 |
| | 7.3.4 | The C-terminal domain | 181 |
| | 7.3.5 | The hydrophobic active-site pocket | 183 |
| | 7.3.6 | The reaction mechanism | 184 |
| 7.4 | Discussion | | 188 |
| 7.5 | Conclusion | | 190 |
| **Summary** | | | 193 |
| **References** | | | 196 |

# LIST OF ABBREVIATIONS

| | |
|---|---|
| 3-PGA | 3-Phosphoglyceric acid |
| ADPGlc | ADP-glucose |
| ADP-R1P | ADP-dependent ribose-1-phosphate |
| aIF | Archaeal initiation factor |
| AMP | Adenosine 5′-monophosphate |
| ASL | Acceptor stem loop |
| ASU | Asymmetric unit |
| β-ME | β-mercaptoethanol |
| CDPIns | CDP-inositol |
| CMP | Cytosine 5′-monophosphate |
| CTD | C-terminal domain |
| DIPPS | CDP-alcohol phosphatidyltransferase |
| dTDPGlc | dTDP-glucose |
| EFBE | Estimated free energy of binding |
| eIF | Eukaryotic initiation factor |
| GALT | Galactose-1-phosphate uridylyltransferase |
| GAP | GTPase-activating protein |
| GCN2 | General control nonderepressible 2 |
| GDPMan | GDP-mannose |
| GEF | Guanine nucleotide exchange factor |
| GlcN 1-P | Glucosamine 1-phosphate |
| GMP | Guanosine 5′-monophosphate |
| HMM | Hidden Markov model |
| HRI | Heme-regulated inhibitor |
| IF | Initiation factor |
| IMAC | Immobilized metal affinity chromatography |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |
| ITC | Isothermal titration calorimetry |
| LB | Luria Bertani |
| LGA | Lamarckian genetic algorithm |
| LM | Low melting |

| | |
|---|---|
| M1Pi | 5-methylthioribose 1-phosphate isomerase |
| MAD | Multiple-wavelength anomalous dispersion |
| MD | Molecular dynamics |
| MIR | Multiple isomorphous replacement |
| ML | Maximum likelihood |
| MPD | 2-methyl-2,4-pentanediol |
| MR | Molecular replacement |
| MSA | Multiple sequence alignment |
| MSP | Methionine salvage pathway |
| MTA | 5-methylthioadenosine |
| MTR-1-P | 5-methylthioribose 1-phosphate |
| MTRu-1-P | 5-methylthioribulose 1-phosphate |
| NEB | New England biolabs |
| NJ | Neighbor joining |
| NMP | Nucleoside 5'-monophosphate |
| NNI | Nearest neighbor interchange |
| NSPases | NDP-sugar pyrophosphorylases |
| NTA | Nitrilotriacetic acid |
| NTD | N-terminal domain |
| NTP | Nucleotide triphosphate |
| NTT | N-terminal tail |
| ORF | Open reading frame |
| PABP | Poly(A)-binding protein |
| PBC | Periodic boundary condition |
| PCR | Polymerase chain reaction |
| PDB | Protein data bank |
| PEG | Poly(ethylene glycol) |
| PERK | PKR-like endoplasmic reticulum-resident kinase |
| PIC | Pre-initiation complex |
| PKR | Protein kinase R |
| PME | Particle mesh Ewald |
| PMI | Phophomannose isomerase |
| PMSF | Phenylmethylsulfonyl fluoride |

| | |
|---|---|
| R15P | Ribose-1,5-bisphosphate |
| R15Pi | Ribose-1,5-bisphosphate isomerase |
| RF | Residence frequency |
| RMSD | Root mean square deviation |
| RMSF | Root mean square fluctuation |
| RPI | Ribose-5-phosphate isomerase |
| RuBisCO | Ribulose-1,5-bisphosphate carboxylase/oxygenase |
| RuBP | Ribulose-1,5-bisphosphate |
| SAH | S-adenosyl homocysteine |
| SAM | S-adenosyl methionine |
| SASA | Solvent accessible surface area |
| SD | Shine-Dalgarno |
| Slg | Slow growth |
| SNT | Sugar nucleotide transferases |
| SPC | Simple point charge |
| SRL | Sisco research laboratories |
| Sui | Suppressor of initiation |
| TB | Terrific broth |
| TC | Ternary complex |
| TMP | Thymidine 5′-monophosphate |
| TPI | Triosephosphate isomerase |
| UDPGal | UDP-galactose |
| UDPGlc | UDP-glucose |
| UDPGlcNAc | UDP-N-acetylglucoseamine |
| UTR | Untranslated region |
| WT | Wild type |
| XI | Xylose isomerase |

# ABSTRACT

# ABSTRACT

The process of protein translation involves decoding of a nucleic acid into a polypeptide chain in three steps viz. initiation, elongation and termination. Although the basic mechanism and components of protein translation are highly conserved, the translation initiation stage has undergone extensive divergence in bacteria and eukaryotes. The most distinct difference lies in the number of accessory proteins called translation initiation factors (IFs). Bacteria utilize only three translation IFs while eukaryotes employ about a dozen of them during protein translation initiation. In case of archaea, which is a prokaryote, an unexpectedly large set of translation IFs are reported. The availability of a higher number of translation IFs in archaea provokes the idea that archaea imitate the eukaryal mechanism of protein translation initiation.

In eukaryotes, two translation IFs, eIF1 and eIF2B play key central role during the process of protein translation initiation. During start codon selection, eIF1 is involved in discriminating against non-AUG codons or AUG codons that have non-optimal context and also promotes the dissociation of aberrantly assembled ribosomal complexes. On the other hand, eIF2B functions as a guanine exchange factor (GEF) for eIF2 and is comprised of two copies each of α, β, δ, γ and ε subunits. The α, β and δ subunits form the regulatory subcomplex while γ and ε make up the catalytic subcomplex.

eIF1 is considered to be universally conserved in all the three domains of life and thus, it might be possible that archaea too possess an open reading frame (ORF) for aIF1. On the other hand, the homologs of only the regulatory subunits (α, β and δ) of eIF2B have been reported to be present in archaea and not the catalytic subunits (γ and ε). However, the regulatory subunits alone would not confer full functionality to eIF2B, as the catalytic subunits accomplish the crucial function of GTP exchange.

Thus, the presence of functional homologs of eukaryotic translation IFs, eIF1 and eIF2B in archaea remains unclear and there is a lack of experimental evidences describing the (non)availability of a functional aIF1 and aIF2B in archaea. Thus, in our study, we aimed to investigate the presence of aIF1 and aIF2B in a hyperthermophilic archaeon, *Pyrococcus horikoshii* OT3 which would further enhance our understanding of protein biosynthesis mechanism in archaea.

Initially, five ORFs, PH1771.1 homologous to eIF1, PH0208 & PH0702 homologous to the regulatory subunits (α, β and δ) and PH1022 & PH1697 homologous to the catalytic subunits (γ and ε) of eIF2B have been identified in *P. horikoshii* OT3. Intriguingly, each of these ORFs also share sequence similarity with certain functionally unrelated proteins involved in various cellular processes. As for instance, PH1771.1 is homologous to a bacterial translation initiation inhibitor, YciH. PH0208 and PH0702 are homologous to enzymes ribose-1,5-bisphosphate isomerase (R15Pi) and 5-methylthioribose 1-phosphate isomerase (M1Pi), respectively, while PH1022 and PH1697 are similar to enzyme NDP-sugar pyrophosphorylases (NSPases).

A preliminary *in silico* analysis was performed to obtain further insight into their function. The sequence analyses were performed using programs such as ProtParam, BLAST, Clustal Omega, MUSTANG and PROMALS3D. For deciphering the evolutionary relationship both at sequence as well as structure level, the program MEGA was used. Phylogenetic analysis reveals that PH1771.1 is evolutionarily related to both eIF1 and YciH while PH0208 and PH0702 reveal clear evolutionary closeness to R15Pi and M1Pi, respectively. Similarly, PH1022 and PH1697 are evolutionarily closer to NSPases. Furthermore, PH0208, PH0702 and PH1022 & PH1697 possess all the amino acid residues essential for fulfilling the enzymatic activity of R15Pi, M1Pi and NSPase, respectively. This prompted us to further confirm the functions of PH1771.1, PH0208 and PH0702 by performing structural studies.

Towards, attaining this goal, the gene encoding PH1771.1, PH0208 and PH0702 were cloned into either pET-22b(+) or pET-28a(+) vectors. Protein expression was achieved using IPTG and further purified using Ni-NTA affinity chromatography. The purified proteins were further concentrated and used for crystallization. The crystallization experiments were performed using micro-batch-under oil, hanging- and sitting-drop vapor-diffusion methods. X-ray intensity data were collected at -173°C using the home source Rigaku MicroMax-007 HF diffractometer (operated at 40 kV and 30 mA) and R-Axis IV++ imaging-plate detector. The diffraction data were processed and scaled using the programs HKL-3000 or iMosflm and Aimless, respectively. All the structures were solved using molecular replacement technique using the program Phaser and the refinements were carried out using Refmac5. Model building was performed using the tools COOT, PSAP and PDBGoodies and the structures were validated using the

programs PROCHECK and MolProbity. The structural analyses which include structure comparison, prediction of oligomeric states, analysis of interacting oligomeric interfaces, calculation of electrostatic potential charge distribution were performed using various tools such as 3dSS, DALI, PISA, PSAP and ABPS embedded in PyMOL.

Determination of the three-dimensional crystal structure of PH1771.1 reveals its high structural similarity to both aIF1 and YciH. However, an in-depth comparison of the structures reveals that PH1771.1 is structurally more similar to eIF1 and thus might function as aIF1 during archaeal translation initiation. Nonetheless, certain notable critical differences between aIF1 and eIF1 could still be perceived at functionally important loop regions that are involved in making important interactions with the ribosomal subunits during protein translation initiation. Apart from the loop regions, differences also occur at the solvent-exposed region between eIF1 and aIF1. Unlike eIF1, aIF1 possesses a negatively charged solvent-exposed region and this attribute might be utilized by archaea to compensate for the absence of the homolog of eukaryotic translation IF, eIF3. Thus, the differences identified in this study provoke the idea of a slightly divergent mode of action of aIF1 during archaeal translation initiation.

The three-dimensional crystal structure of PH0208 reveals its overall structure to be highly similar to the regulatory subunits (α, β and δ) of eIF2B. However, its ability to bind to the ribose-1,5-bisphosphate (R15P, substrate) and ribulose-1,5-bisphosphate (RuBP, product) of the enzyme R15Pi during crystallographic as well as ITC experiments suggests that PH0208 would function as R15Pi. This finding is further supported by the ability of PH0208 to convert R15P (substrate) to RuBP (product) during the process of crystallization. Moreover, mutation of the key catalytic residues Cys135 and Asp204 of the enzyme R15Pi led to the inability of PH0208 to convert the substrate into product. The enzymatic activity of R15Pi is known to be elevated in the presence of AMP and to a lesser extent by GMP also. However, due to the unavailability of an AMP/GMP-bound R15Pi structure, the exact role of purine nucleotides during enzyme catalysis remained unknown. Our study for the first time reports the binding sites of AMP and GMP in R15Pi. The revelation of the AMP binding site clarified the role of AMP in providing structural stability to R15Pi. The binding of GMP to the ‘AMP binding site’ in addition to its own binding site indicates that GMP might also execute a similar function, though with less specificity. Unlike AMP and GMP, there is no binding

observed for the pyrimidine nucleotides (CMP or TMP) explaining their inability to elevate the catalytic activity of the enzyme. Furthermore, we have also utilized the structural resemblance between R15Pi and the regulatory subunits of eIF2B to propose a model for the regulatory mechanism of eIF2B in eukaryotes.

Albeit the availability of the three-dimensional crystal structure of R15Pi, several structural aspects of the enzyme such as accountability of the enzyme to function as a hexamer, recognition of the essential catalytic residue for structural stability, identification of essential substrate- and product-binding residues and structural as well as functional role(s) of invariant water molecules remains undescribed. Molecular dynamics (MD) simulation using the package GROMACS was performed to understand the obligation of the enzyme R15Pi to oligomerize and also in deciphering the role of catalytic residue(s) in structural stability. Results indicate that the hexameric state of R15Pi provides stability to the otherwise highly fluctuating C-terminal domain (CTD) and also enables a more stable ligand interaction at the active-site pocket. MD simulations also led to the identification of Cys135 as the structurally essential catalytic residue which is required for the structural stability of the enzyme. Our study also identified the presence of invariant water molecules at the active-site pocket, its vicinity and at structurally important regions of R15Pi. Identification of the invariant water molecules was done using the 13 crystal structures of R15Pi from *P. horikoshii* OT3 solved during this study. Few invariant water molecules are directly involved in assisting the binding of the ligand at the active site while others interact with amino acid residues and impart structural integrity to the active-site pocket. Apart from the active site, the presence of invariant water molecules at flexible loop and in the inter-subunit regions also contributes towards the formation of a compact hexameric R15Pi structure. Finally molecular docking was performed using AutoDock to identify the amino acid residues essential for holding the substrate (R15P) or product (RuBP) in the active site of the enzyme R15Pi. During docking, the binding of the substrate R15P at an ‘alternate binding site’ (or ‘initial binding site’) hints toward a probable early binding event which could not be captured during the process of crystallization. Based on this observation of substrate binding to the ‘initial binding site’ instead of directly binding at the active site led us to propose a mechanism of ‘substrate sliding’ for the enzyme R15Pi.

The three-dimensional structure determination of PH0702 reveals a tertiary structure highly similar to M1Pi, R15Pi and the regulatory subunits (α, β and δ) of eIF2B. An in-depth structural investigation of PH0702 reveals the occurrence of certain structural attributes such as the presence of an N-terminal extension and a hydrophobic patch at the C-terminal which are exclusive to M1Pi and absent in both R15Pi and the regulatory subunits of eIF2B. This indicates that PH0702 would function as M1Pi instead of functioning as the regulatory subunits of eIF2B. These distinct structural attributes identified in M1Pi also indicates towards the presence of a hydrophobic active-site pocket. The hydrophobic nature of the active-site pocket is further augmented by the presence of a 'closed state-like' open conformation of M1Pi rendered by forward shift in a loop region which covers the active site. All these structural attributes contribute towards a hydrophobic microenvironment in the vicinity of the active site of the enzyme which in turn aids in commencement of the reaction mechanism of M1Pi. A hydrophobic active-site microenvironment in addition to the availability of an optimal repertoire of amino acid residues surrounding the catalytic residues in M1Pi led us to propose its probable reaction mechanism via the formation of a *cis*-phosphoenolate intermediate.

Thus, from the work presented in this thesis, it can be concluded that a functional and structural homolog of eIF1 is available in archaea while no functional homologs of the regulatory and catalytic subunits of eIF2B are available. Thus, the long-borne notion of archaeal protein translation initiation process being similar to eukaryal demands a re-evaluation.

The work presented in the thesis has been reported in the following publications.

## PUBLICATIONS FROM THESIS

1. **Gogoi, P**., Mordina, P. and Kanaujia, S.P., 2019. Deciphering the rationale behind promiscuity of thermophilic NDP-sugar pyrophosphorylases. (Manuscript under review)

2. **Gogoi, P**. and Kanaujia, S.P., 2019. Role of structural features in oligomerization, active-site integrity and ligand binding of ribose-1,5-bisphosphate isomerase. *Computational and Structural Biotechnology Journal*, *17*, 333-344.

3. **Gogoi, P**., Mordina, P. and Kanaujia, S.P., 2019. Structural insights into the catalytic mechanism of 5-methylthioribose 1-phosphate isomerase. *Journal of Structural Biology*, *205*(1), 67-77.

4. **Gogoi, P**. and Kanaujia, S.P., 2018. Archaeal and eukaryal translation initiation factor 1 differ in their RNA interacting loops. *FEBS Letters*, *592*(9), 1602-1610.

5. **Gogoi, P**. and Kanaujia, S.P., 2018. A presumed homologue of the regulatory subunits of eIF2B functions as ribose-1,5-bisphosphate isomerase in *Pyrococcus horikoshii* OT3. *Scientific Reports*, *8*(1), 1891.

6. **Gogoi, P**., Srivastava, A., Jayaprakash, P., Jeyakanthan, J. and Kanaujia, S.P., 2016. *In silico* analysis suggests that PH0702 and PH0208 encode for methylthioribose-1-phosphate isomerase and ribose-1,5-bisphosphate isomerase, respectively, rather than aIF2Bβ and aIF2Bδ. *Gene*, *575*(1), 118-126.

## PUBLICATIONS FROM OTHER WORK

7. Chandravanshi, M., **Gogoi, P**. and Kanaujia, S.P., 2019. Structural and thermodynamic correlation illuminates the selective transport mechanism of disaccharide α-glycosides through ABC transporter. (Manuscript under review)

8. Mandal, S.K., Adhikari, R., Sharma, A., Chandravanshi, M., **Gogoi, P**. and Kanaujia, S.P., 2019. Designating ligand specificities to metal uptake ABC transporters in *Thermus thermophilus* HB8. *Metallomics*, *11*, 597-612.

9. Mandal, S.K., Chandravanshi, M., **Gogoi, P**. and Kanaujia, S.P., 2017. *In silico* characterization of TTHA0596: A potential Zn2+ binding protein of ATP-binding cassette transporter. *Gene Reports*, *6*, 132-141.

10. Chandravanshi, M., **Gogoi, P**. and Kanaujia, S.P., 2016. Computational characterization of TTHA0379: A potential glycerophosphocholine binding protein of Ugp ATP-binding cassette transporter. *Gene, 592*(2), 260-268.

11. Srivastava, A., **Gogoi, P.**, Deka, B., Goswami, S. and Kanaujia, S.P., 2016. *In silico* analysis of 5′-UTRs highlights the prevalence of Shine–Dalgarno and leaderless-dependent mechanisms of translation initiation in bacteria and archaea, respectively. *Journal of Theoretical Biology, 402*, 54-61.

12. Dixit, B., Ghosh, K.K., Fernandes, G., Kumar, P., **Gogoi, P.** and Kumar, M., 2016. Dual nuclease activity of a Cas2 protein in CRISPR–Cas subtype I-B of Leptospira interrogans. *FEBS Letters, 590*(7), 1002-1016.

13. **Gogoi, P.**, Chandravanshi, M., Mandal, S.K., Srivastava, A. and Kanaujia, S.P., 2016. Heterogeneous behavior of metalloproteins toward metal ion binding and selectivity: insights from molecular dynamics studies. *Journal of Biomolecular Structure and Dynamics, 34*(7), 1470-1485.

## CONFERENCES ATTENDED

1. **Gogoi, P.** and Kanaujia, S.P. Poster presentation on "Structure-based functional characterization of archaeal translation initiation factors". National Seminar on Crystallography-46 [NSC-46], June 27$^{th}$ -29$^{th}$, 2018 NIMHANS, Bengaluru.

2. **Gogoi, P.** and Kanaujia, S.P. Oral presentation on the topic "Structural and functional characterization of a presumed homologue of the regulatory subunits of eIF2B". Research Conclave 2018, March 8$^{th}$-11$^{th}$, 2018, IIT Guwahati.

3. **Gogoi, P.** and Kanaujia, S.P. Poster presentation on "Architecture of ribose-1,5-bisphosphate isomerase, an enzyme unique to archaea". National Seminar on Crystallography-45 (NSC-45), July 9$^{th}$ -12$^{th}$, 2017, IIT BHU, Varanasi.

4. **Gogoi, P.** and Kanaujia, S.P. Poster presentation on "Structural and functional characterization of ribose-1,5-bisphosphate isomerase in archaea". 24th Congress and General Assembly of the International Union of Crystallography, August 21st-28th, 2017, Hyderabad.

5. **Gogoi, P.** and Kanaujia, S.P. Poster presentation on "*In silico* analysis suggests that PH0702 and PH0208 encode for methylthioribose-1-phosphate isomerase and ribose-1,5-bisphosphate isomerase, respectively, rather than aIF2Bβ and aIF2B". National Seminar on Crystallography-44 (NSC-44), July 10th -13th, 2016, IISER Pune.

6. **Gogoi, P.** and Kanaujia, S.P. (2014). Poster presentation on "*In silico* analysis suggests that PH0702 and PH0208 encode for methylthioribose-1-phosphate isomerase and ribose-1,5-bisphosphate isomerase, respectively, rather than aIF2Bβ and aIF2B". Recent Developments in Medical Biotechnology and Structure-Based Drug Designing, December 6th-7th, 2014, IIT Guwahati.

## WORKSHOPS ATTENDED

1. Workshop on "CCP4 Crystallography School and Workshop: From data processing to structure refinement and beyond", October 22nd-26th, 2018, IMTECH, Chandigarh.

2. Workshop on "Latest Methods in X-Ray Crystallography", November 14th-25th, 2016, JNU, New Delhi.

## PROTEIN STRUCTURES DEPOSITED TO PROTEIN DATA BANK

| S. No. | PDB id | Title of the PDB | Publication |
|---|---|---|---|
| 1. | 6A34 | Crystal structure of 5-methylthioribose 1-phosphate isomerase from *Pyrococcus horikoshii* OT3 - Form I | Gogoi et al., 2019 |
| 2. | 6A35 | Crystal structure of 5-methylthioribose 1-phosphate isomerase from *Pyrococcus horikoshii* OT3 - Form II | |
| 3. | 5ZCY | Crystal structure of archaeal translation initiation | Gogoi and |

| | | | |
|---|---|---|---|
| | | factor 1 at 1.5 Angstroms resolution | Kanaujia, 2018 |
| 4. | 5YFJ | Crystal structure of ribose-1,5-bisphosphate isomerase from *Pyrococcus horikoshii* OT3 in complex with ribulose-1,5-bisphosphate | Gogoi and Kanaujia, 2018 |
| 5. | 5YFS | Crystal structure of ribose-1,5-bisphosphate isomerase mutant C135S from *Pyrococcus horikoshii* OT3 in complex with ribose-1,5-bisphosphate | |
| 6. | 5YFT | Crystal structure of ribose-1,5-bisphosphate isomerase mutant D204N from *Pyrococcus horikoshii* OT3 in complex with ribose-1,5-bisphosphate | |
| 7. | 5YFU | Crystal structure of ribose-1,5-bisphosphate isomerase from *Pyrococcus horikoshii* OT3 in complex with ribulose-1,5-bisphosphate and AMP | |
| 8. | 5YFV | Crystal structure of ribose-1,5-bisphosphate isomerase mutant C135S from *Pyrococcus horikoshii* OT3 in complex with ribose-1,5-bisphosphate and AMP | |
| 9. | 5YFW | Crystal structure of ribose-1,5-bisphosphate isomerase mutant D204N from *Pyrococcus horikoshii* OT3 in complex with ribose-1,5-bisphosphate and AMP | |
| 10. | 5YFX | Crystal structure of ribose-1,5-bisphosphate isomerase mutant D204N from *Pyrococcus horikoshii* OT3 in complex with ribose-1,5-bisphosphate and AMP | |
| 11. | 5YG5 | Crystal structure of ribose-1,5-bisphosphate isomerase from *Pyrococcus horikoshii* OT3 in complex with ribulose-1,5-bisphosphate and GMP | |
| 12. | 5YG6 | Crystal structure of ribose-1,5-bisphosphate isomerase mutant C135S from *Pyrococcus horikoshii* OT3 in complex with ribose-1,5-bisphosphate and GMP | |
| 13. | 5YG7 | Crystal structure of ribose-1,5-bisphosphate isomerase mutant D204N from *Pyrococcus horikoshii* OT3 in complex with ribose-1,5-bisphosphate and GMP | |
| 14. | 5YG8 | Crystal structure of ribose-1,5-bisphosphate isomerase from *Pyrococcus horikoshii* OT3 in complex with | |

| | | | |
|---|---|---|---|
| | | ribulose-1,5-bisphosphate, AMP and GMP | |
| 15. | 5YG9 | Crystal structure of ribose-1,5-bisphosphate isomerase mutant C135S from *Pyrococcus horikoshii* OT3 in complex with ribose-1,5-bisphosphate, AMP and GMP | |
| 16. | 5YGA | Crystal structure of ribose-1,5-bisphosphate isomerase mutant D204N from *Pyrococcus horikoshii* OT3 in complex with ribose-1,5-bisphosphate, AMP and GMP | |

# LIST OF FIGURES

**Figure 1.1.** The three stages of protein translation. 1

**Figure 1.2.** A pictorial representation of the different stages of bacterial protein translation initiation. 3

**Figure 1.3.** A pictorial representation of the different stages of eukaryal protein translation initiation. 5

**Figure 1.4.** A pictorial representation of eIF1 rendering a 'scanning competent' to 'scanning arrested' state by binding to the 40S small ribosomal subunit. 7

**Figure 1.5.** Three-dimensional structure of eIF1. 8

**Figure 1.6.** A pictorial representation of the process of guanine nucleotide exchange by eIF2B during eukaryotic translation initiation. 9

**Figure 1.7.** Three-dimensional structure of eIF2B. 10

**Figure 1.8.** A pictorial representation of the stages involved in regulation of eIF2B during translation initiation in eukaryotes. 11

**Figure 1.9.** A schematic representation of pentose bisphosphate pathway in archaea. 12

**Figure 1.10.** Three-dimensional structure of enzyme R15Pi. 13

**Figure 1.11.** A schematic representation of methionine salvage pathway (MSP). 14

**Figure 1.12.** Three-dimensional structure of enzyme M1Pi. 15

**Figure 1.13.** Structure and function of NSPase. 16

**Figure 2.1.** A general representation of the PCR cycle. 21

**Figure 3.1.** Evolutionary relationships among eIF1, YciH and aIF1 from eukaryotes, bacteria and archaea respectively. 39

**Figure 3.2.** Phylogenetic tree showing relationship among M1Pi, R15Pi and eIF2B regulatory subunits. 41

**Figure 3.3.** Structural (dis)similarity of R15Pi, M1Pi and eIF2Bα. 43

**Figure 3.4.** Conserved active-site residues of M1Pi. 44

**Figure 3.5.** Conserved active-site residues of R15Pi. 46

**Figure 3.6.** Evolutionary relationship of NSPases and catalytic subunits of eIF2B. 48

**Figure 3.7.** Heat map illustrating the pairwise sequence similarity of all the identified NSPases from *P. horikoshii*. 49

**Figure 3.8.** Structure-based sequence alignment of PhNSPases along with representative NSPases from each group classified on the basis of their metal coordinating motifs. 52

**Figure 3.9.** Evolutionary relationship and domain organization of NSPases. 54

**Figure 3.10.** Evolutionary relationship of NSPases on the basis of the nucleotidyltransferase domain. 56

**Figure 3.11.** Multiple sequence alignment (MSA) of PH0365 with UDPGal pyrophosphorylases. 57

**Figure 3.12.** Multiple sequence alignment (MSA) of PH1219 with CDPIns pyrophosphorylases. 59

**Figure 3.13.** Comparison of active-site pockets of NSPases. 61

**Figure 3.14.** Multiple sequence alignment (MSA) of promiscuous NSPases. 64

**Figure 3.15.** Multiple sequence alignment (MSA) of promiscuous NSPases to identify the conserved amino acid residues of the acetyltransferase domain of UDPGlcNAc pyrophosphorylase. 66

**Figure 4.1.** Cloning, over expression and purification of aIF1. 74

**Figure 4.2.** Crystallization of aIF1. 75

**Figure 4.3.** Crystal structure of aIF1 represented as cartoon model. 79

**Figure 4.4.** Structural comparison of aIF1, eIF1 and YciH. 80

**Figure 4.5.** Surface charge distribution of eIF1 and aIF1. 82

**Figure 5.1.** Cloning and site-directed mutagenesis of PhR15Pi. 90

**Figure 5.2.** Over expression of PhR15Pi-WT and mutant proteins PhR15Pi-C135S, PhR15Pi-D204N & PhR15Pi-C135A. 91

**Figure 5.3.** Solubility test for PhR15Pi-WT and mutant proteins PhR15Pi-C135S, PhR15Pi-D204N & PhR15Pi-C135A. 92

**Figure 5.4.** Purification of PhR15Pi-WT and mutant proteins PhR15Pi-C135S, PhR15Pi-D204N & PhR15Pi-C135A. 93

**Figure 5.5.** Crystal optimization of PhR15Pi. 94

**Figure 5.6.** Crystallization of PhR15Pi-WT and mutant proteins PhR15Pi-C135S, PhR15Pi-D204N & PhR15Pi-C135A in complex with purine and pyrimidine nucleotides. 95

**Figure 5.7.** The overall monomeric structure of PH0208 protein from *P. horikoshii* represented as cartoon model. 105

**Figure 5.8.** Quarternary structure of PH0208 protein. 106
**Figure 5.9.** The active-site pocket of PhR15Pi. 108
**Figure 5.10.** Biophysical characterization of PhR15Pi protein. 110
**Figure 5.11.** AMP molecule binds at the interface of three protomers. 112
**Figure 5.12.** GMP binds at two different sites in PhR15Pi. 114
**Figure 5.13.** AMP is preferred over GMP at the 'AMP binding site'. 115
**Figure 5.14.** Crucial differences at the active site of PhR15Pi and eIF2Bα. 117
**Figure 5.15.** Structural similarity between PhR15Pi and the regulatory subunits of eIF2B. 122
**Figure 6.1.** Dynamics of R15Pi-monomer and R15Pi-hexamer. 134
**Figure 6.2.** Dynamics of the mutants R15Pi-C135S and R15Pi-D204N. 136
**Figure 6.3.** Invariant water molecules at the active site and in its vicinity. 140
**Figure 6.4.** Invariant water molecules at structurally important regions. 146
**Figure 6.5.** Multiple sequence alignment (MSA) of enzyme R15Pi. 147
**Figure 6.6.** Docking of R15P and RuBP to R15Pi-Form I and R15Pi-Form III. 149
**Figure 6.7.** The substrate (R15P) binding to R15Pi mutant protein. 154
**Figure 6.8.** The movement of the substrate (R15P) during MD simulation. 155
**Figure 6.9.** The product (RuBP) binding to the R15Pi mutants. 156
**Figure 6.10.** Proposed mechanism of the 'substrate sliding' in the enzyme R15Pi depicted in a three-dimensional model and chemical scheme. 161
**Figure 7.1.** Cloning, over expression and purification of PhM1Pi 168
**Figure 7.2.** Crystal hits of PhM1Pi. 169
**Figure 7.3.** Protein crystals of PhM1Pi at different stages of optimization. 171
**Figure 7.4.** Crystallization of PhM1Pi 172
**Figure 7.5.** Tertiary structure of PhM1Pi 176
**Figure 7.6.** Quarternary structure of PhM1Pi 178
**Figure 7.7.** Domain movement of PhM1Pi. 179
**Figure 7.8.** Structural evolution of the NTD of M1Pi. 180
**Figure 7.9.** Structural evolution of the CTD of M1Pi. 182
**Figure 7.10.** Water molecules in the open and closed states of M1Pi. 183
**Figure 7.11.** Microenvironment of the active-site pocket of M1Pi. 186
**Figure 7.12.** Hydrogen bonding network of polar residues at the active-site pocket 187

**Figure 7.13.** Proposed reaction mechanism of M1Pi via a *cis*-phosphoenolate intermediate. 188

# LIST OF TABLES

**Table 3.1.** Homologues of probable translation initiation factors in archaea. 38

**Table 3.2.** List of homologs of NSPases from *P. horikoshii* OT3. 50

**Table 4.1.** Data collection, processing and refinement statistics of aIF1. 76

**Table 5.1.** Data collection, processing and refinement statistics for PhR15Pi-WT bound to RuBP and PhR15Pi-C135S & PhR15Pi-D204N mutant proteins bound to R15P. 96

**Table 5.2.** Data collection, processing and refinement statistics for PhR15Pi-WT bound to RuBP & AMP, RuBP & GMP and RuBP, AMP & GMP. 98

**Table 5.3.** Data collection, processing and refinement statistics for PhR15Pi-C135S mutant protein bound to R15P & AMP, R15P & GMP and R15P, AMP & GMP. 100

**Table 5.4.** Data collection, processing and refinement statistics for PhR15Pi-D204N mutant protein bound to R15P & AMP, R15P & GMP and R15P, AMP & GMP. 102

**Table 5.5.** ITC data for the binding of R15P/RuBP to PhR15Pi-C135S, PhR15Pi-C135A and PhR15Pi-D204N mutant proteins. 111

**Table 6.1.** Details of three-dimensional crystal structures of R15Pi used during MD simulations. 129

**Table 6.2.** List of three-dimensional crystal structures of R15Pi and their details used for the identification of invariant water molecules. 131

**Table 6.3.** List of hydrogen bond interaction(s) between invariant water molecule(s) and amino acid residue(s) at the active-site pocket as well as in its vicinity. 138

**Table 6.4.** Invariant water molecules corresponding to the reference structure (PDB id: 5YFT, chain B) along with their B factor, solvent accessible surface area (SASA) and percentage residence frequency calculated from the ensembles generated using molecular dynamics simulation. 141

**Table 6.5.** Details of molecular docking interaction between the different forms of R15Pi with R15P and RuBP. 150

**Table 7.1.** Crystal optimization approaches for PhM1Pi 170

**Table 7.2.** Data collection, processing and refinement statistics of PhM1Pi. 172

# CHAPTER 1

## Protein translation initiation in archaea: a chimera of bacterial and eukaryal attributes

## 1.1 PROTEIN TRANSLATION

Protein translation, an indispensable process in the pathway of gene expression, involves decoding of the information available in the nucleic acids into a polypeptide with the assistance of a highly efficient machinery. Conceptually, the process of protein translation is divided into three steps: initiation, elongation and termination (Figure 1.1). During the initiation step, an initiator tRNA carrying the amino acid methionine comes together with the mRNA and the ribosome to initiate the process of protein translation (Pain, 1996). During elongation, additional amino acids are brought to the ribosome with the help of tRNAs which are then linked to form a polypeptide chain (Proud, 1994). Finally, during the termination stage, the ribosome releases the finished polypeptide chain upon recognition of the stop codon (Zhouravleva et al., 1995). Out of the three steps of protein translation, the first step i.e. translation initiation is considered to be the key step as the efficiency for decoding a given mRNA lies with the efficiency of initiation. This also makes translation initiation to be the 'rate-limiting step' of protein translation (Benelli and Londei, 2009; Chu and Haar, 2012). Although the basic mechanism of translation initiation is alike in all the three domains of life (bacteria, eukarya and archaea), the components involved during the process are highly divergent.

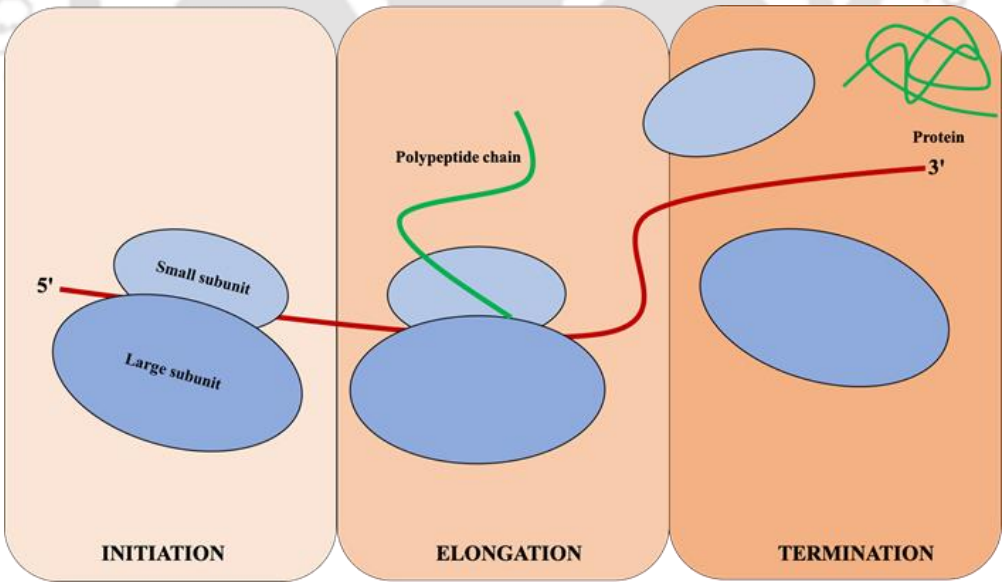



**Figure 1.1. The three stages of protein translation.** The small and large ribosomal subunits are shown as small and large blue ovals, respectively. The mRNA stand and the polypeptide chain (or protein) are depicted as red and green curves, respectively.

### 1.1.1 PROTEIN TRANSLATION INITIATION IN BACTERIA

In bacterial translation initiation, the key components that play the pivotal role include ribosomes, the mRNA, tRNA and a set of proteins known as translation initiation factors (IFs). Bacteria possess 70S ribosome composed of a large 50S and a small 30S subunit. The 70S ribosome has three tRNA binding sites namely, the aminoacyl (A), peptidyl (P) and exit (E) sites. Before mRNA is introduced to the 30S ribosome, IF3 promotes the dissociation of the two ribosomal subunits by strongly binding to the 30S ribosomal subunit. In addition, the initiation factor IF1 directly binds to the A-site of the 30S ribosome and in turn directs the initiator tRNA (fMet-tRNA$_f^{Met}$) to the ribosomal A-site (Dahlquist and Puglisi, 2000; Carter et al., 2001). The mRNA to be decoded is then brought to the 30S ribosomal subunit in complex with IF2 and fMet-tRNA$_f^{Met}$. The initiation codon is accommodated at the P-site of the 30S ribosome with the help of interaction between the Shine-Dalgarno (SD) sequence of the mRNA and the anti-SD sequence of the 16S rRNA (Yusupova et al, 2001). The 30S pre-initiation complex (PIC) consisting of 30S ribosomal subunit, the three initiation factors (IF1, IF2 and IF3), mRNA and fMet-tRNA$_f^{Met}$ undergoes a conformational change to attain a correct codon-anticodon base pairing. Once the 30S PIC is stabilised with the correct codon-anticodon pairing between mRNA and fMet-tRNA$_f^{Met}$, IF1 and IF3 are ejected while IF2 stimulates association of the 50S ribosomal subunit to the complex (Gualerzi et al., 1977; Pon and Gualerzi, 1990). Association of the 50S ribosomal subunit leads to the formation of the 70S initiation complex which subsequently enters the elongation phase of translation (Figure 1.2).

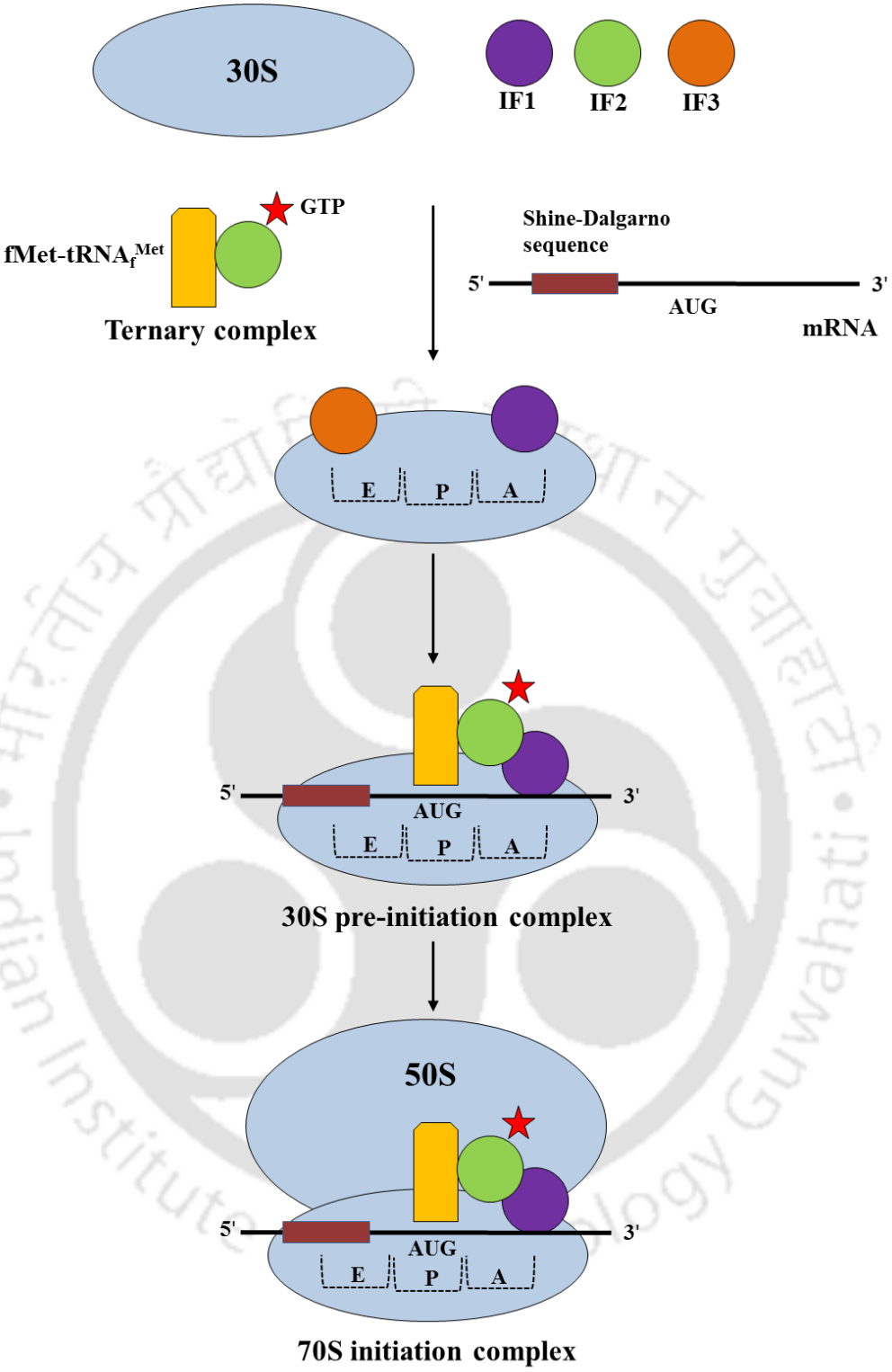



**Figure 1.2. A pictorial representation of the different stages of bacterial protein translation initiation.** The small and large ribosomal subunits are shown as small and large blue ovals, respectively. The translation initiation factors are distinguished as spheres of different colors. The fMet-tRNA$_f^{Met}$ is shown as yellow rectangle and the GTP is shown as red star. The mRNA strand and Shine-Dalgarno (SD) sequence are shown as black line and brown rectangle, respectively.

### 1.1.2 PROTEIN TRANSLATION INITIATION IN EUKARYA

Similar to bacteria, the key players of eukaryotic translation initiation include the ribosome, mRNA, tRNA and a set of translation IFs. In eukaryotes, a large 60S and a small 40S ribosomal subunit associate together to form the 80S ribosome. Intriguingly, unlike bacteria, the entire process of protein translation initiation in eukaryotes is coordinated by at least a dozen of eukaryotic translation IFs (Jackson et al., 2010; Hinnebusch, 2011). During eukaryotic translation initiation, a ternary complex (TC) comprising of initiator methionyl-tRNA (Met-$tRNA_i^{Met}$), GTP-bound eukaryotic initiation factor 2 (eIF2) binds to the 40S small ribosomal subunit to form the 43S PIC. The binding of the TC to the 40S small ribosomal subunit is promoted by other eukaryotic translation initiation factors such as eIF1, eIF1A, eIF5 and eIF3 complex. The 43S PIC binds to the 5′ end of the mRNA marked by a 7-methylguanosine cap with the help of eIF3, poly(A)-binding protein (PABP), eIF4B and the eIF4F complex consisting of eIF4E, eIF4A and eIF4G. Once attached to the cap, the 43S PIC scans along the mRNA to locate the start codon. A correct base pairing between anticodon of Met-$tRNA_i^{Met}$ and the start codon at the P-site of the 40S small ribosomal subunit causes the arrest of the 43S PIC (Lomakin et al., 2006; Kolitz et al., 2009; Hinnebusch, 2011). This triggers the conversion of eIF2•GTP of the TC into eIF2•GDP via gated phosphate ($P_i$) release and the action of the GTPase-activating protein (GAP) eIF5. The inactive eIF2•GDP along with eIF5 and several other IFs dissociates from the PIC at this stage. eIF5B, a second GTPase, along with eIF1A facilities the joining of the 60S large ribosomal subunit to the 40S small ribosomal subunit leading to the assembly of the 80S initiation complex (Aitken and Lorsch, 2012; Hinnebusch and Lorsch, 2012) (Figure 1.3).

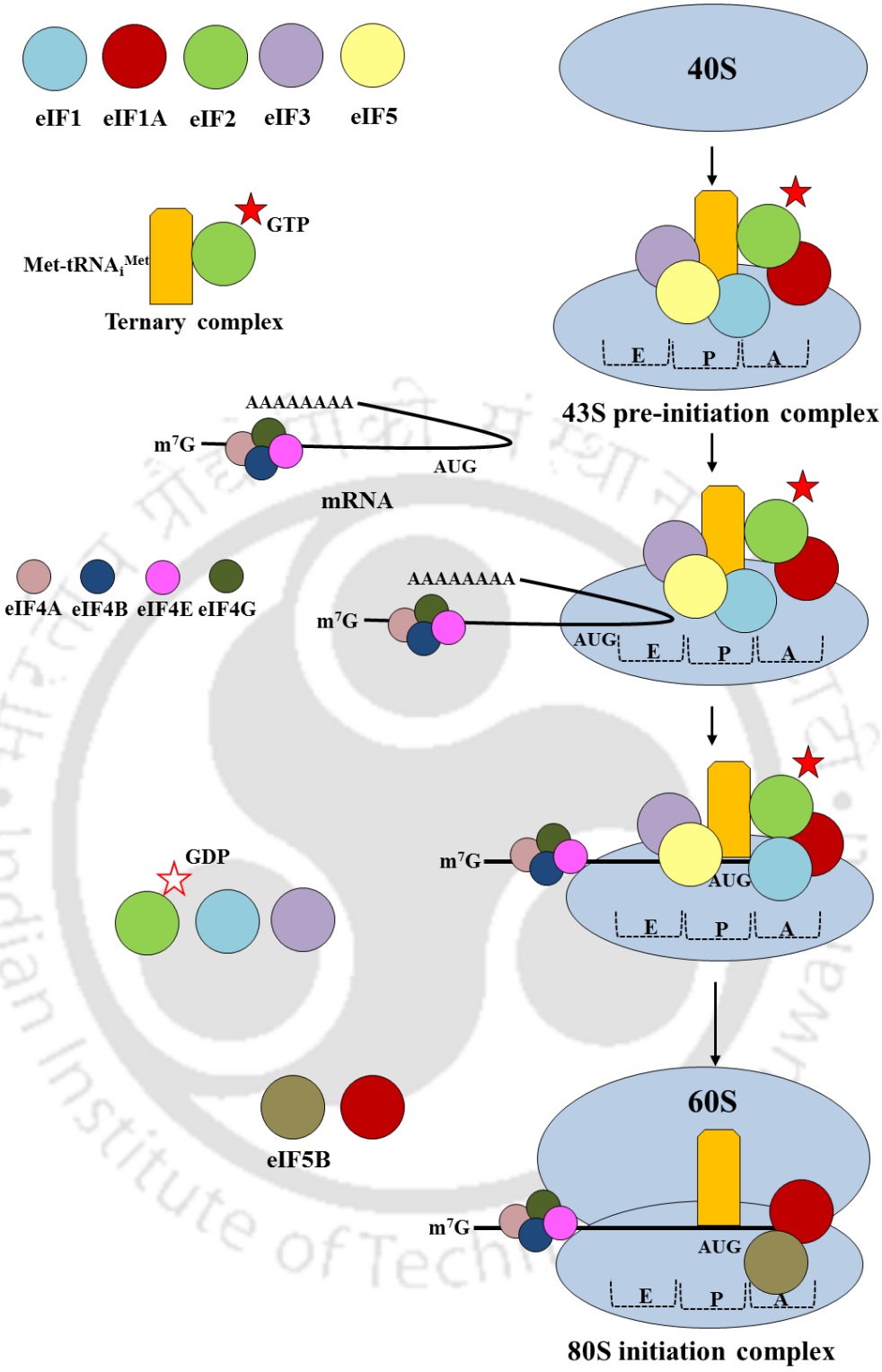



**Figure 1.3. A pictorial representation of the different stages of eukaryal protein translation initiation.** The small and large ribosomal subunits are shown as small and large blue ovals. The translation initiation factors are distinguished in spheres of different colors. The Met-tRNA$_i^{Met}$ and mRNA are shown as yellow rectangle and black line, respectively. GTP and GDP are shown as red star and white star with red border, respectively.

### 1.1.3 PROTEIN TRANSLATION INITIATION IN ARCHAEA

The nature of protein translation initiation in archaea has remained a matter of speculation ever since the discovery of this evolutionarily distinct group of organisms in 1977 (Woese and Fox, 1977). The mechanism of protein translation initiation in archaea seems to be a chimera of bacterial and eukaryal translation initiation processes. The archaeal mRNA is polycistronic, uncapped and lack poly-A tail, showing complete resemblance to the bacterial mRNA (Bult et al., 1996). However, the absence of an intact SD sequence at the 5′-untranslated region (UTR) in majority of the archaeal mRNAs further complicates the translation initiation mechanism (Londei, 2005). Moreover, unlike bacteria, archaea possess more than three IFs involved during the process of translation initiation. With the advent of the genome sequences of *Methanococcus jannaschii* and *Archaeoglobus fulgidus* (Bult et al., 1996; Klenk et al., 1997), it also became apparent that all the archaeal IFs have their closest homologs in eukaryotes rather than in bacteria (Benelli and Londei, 2009, 2011). Except those IFs which are involved in recognition of the 5′ cap in eukaryotes, the homologs of all the other eukaryotic IFs are assumed to be available in archaea. Out of all the archaeal translation IFs, availability of the homolog of two eukaryotic translation IFs viz. eIF1 and eIF2B, however, remains controversial.

#### 1.1.3.1 EUKARYOTIC TRANSLATION INITIATION FACTOR 1

eIF1, also known as the ‘gatekeeper’ of translation initiation, is involved in the process of start codon scanning and recognition. Along with eIF1A, eIF1 binds to the 40S ribosomal subunit to confer a conformational change which transforms the closed or ‘scanning arrested’ state of the ribosome into the open or ‘scanning competent’ state. In the open state, the ‘latch’ of the mRNA entry pore is opened allowing the scanning of the mRNA. eIF1 also plays a pivotal role in inhibiting the Met-tRNA$_i^{Met}$ anticodon interactions with near-cognate triplets or with AUG in the suboptimal initiation context (Pestova and Kolupaeva, 2002; Passmore et al., 2007). This inhibitory effect of eIF1 can only be overcome by the perfect AUG-anticodon duplex in an optimal AUG context as well as in the presence of a long 5′-UTR which binds to the RNA channel present in the 40S ribosomal subunit (Pestova and Kolupaeva, 2002; Martin-Marcos et al., 2011). Furthermore, eIF1 also impedes the function of other translation IFs until the recognition of the start codon. It inhibits the release of a $P_i$ from eIF2•GTP as well as restrains the

GAP function of eIF5 at non-AUG codons (Unbehaun et al., 2004; Maag et al., 2005). Upon recognition of the correct start codon, eIF1 is displaced or ejected from the 40S small ribosomal subunit as a result of the steric clash with the acceptor and D stems of initiator tRNA interacting with the start codon which ultimately leads to a 'scanning arrested' state (Nanda et al., 2009; Lomakin and Steiz, 2013) (Figure 1.4).

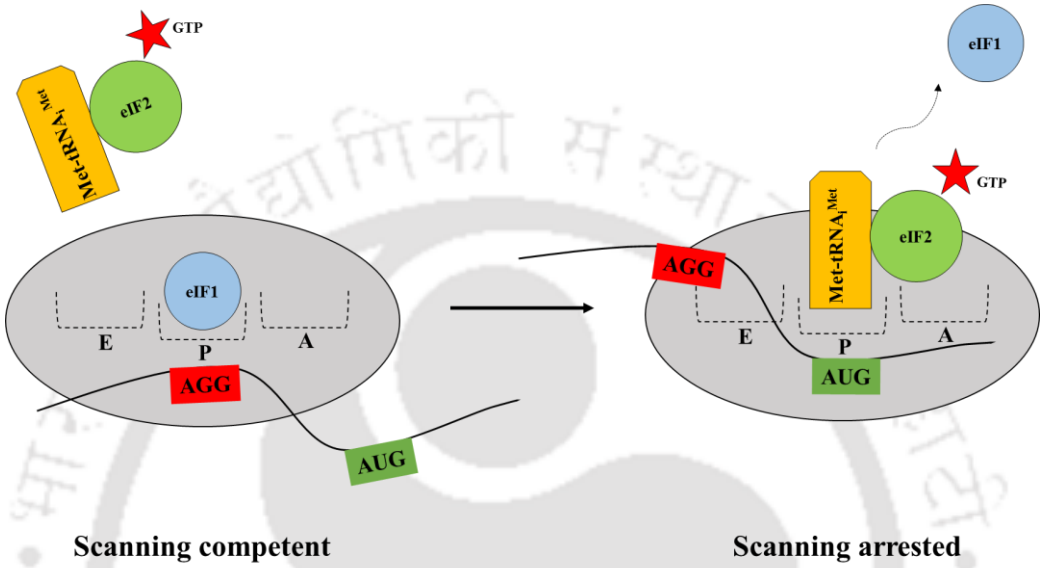



**Figure 1.4. A pictorial representation of eIF1 rendering a 'scanning competent' to 'scanning arrested' state by binding to the 40S small ribosomal subunit.** The small ribosomal subunit and the mRNA strand are shown as grey oval and black curve, respectively. eIF1 and eIF2 are shown as blue and green spheres, respectively, while Met-tRNA$_i^{Met}$ is shown as yellow rectangle.

Structurally, eIF1 is a single domain protein with two α-helices on one side of five parallel and antiparallel β-strands and an unstructured N-terminal tail (NTT) (Figure 1.5A). It shares a general structural similarity with mostly RNA-interacting proteins (Fletcher et al., 1999). One of the α- helices along with a loop region of eIF1 is involved in interacting with the 40S small ribosomal subunit while the NTT binds to the N-terminal domain (NTD) of eIF2 and C-terminal domain (CTD) of eIF5 (Cheung et al., 2007; Rabl et al., 2011).

Although the functional homolog of eIF1 in bacteria is IF3, they lack any mutual sequence or structural similarity (Lomakin et al., 2006). Rather, eIF1 displays significant

sequence and structural similarity with bacterial protein YciH (Figure 1.5B). Despite, a highly similar structure as well as a similar binding site on the ribosome, YciH does not play any role in maintaining the fidelity of translation initiation in bacteria (Fletcher et al., 1999; Cort et al., 1999; Lomakin et al., 2006). Major structural variations between eIF1 and YciH lie in the functionally important loop regions (Figure 1.5B). YciH lacks loop 2 of eIF1 that is involved in a functional interaction with the initiator tRNA. On the other hand, presence of a bulkier loop in YciH, which occupies the P-site of the ribosome, prevents the binding of the tRNA and thus allows it to function as a translation initiation inhibitor (Osterman et al., 2015).

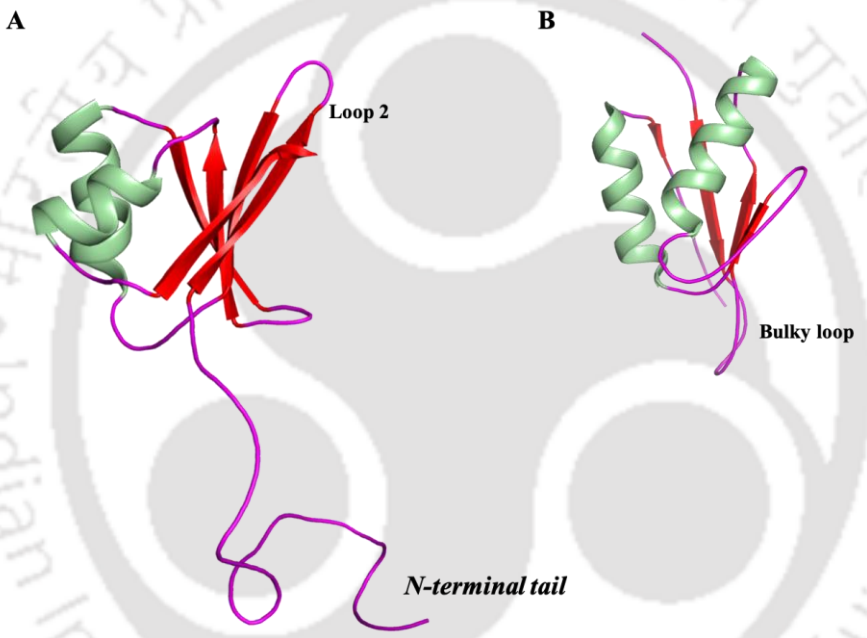



**Figure 1.5. Three-dimensional structure of eIF1.** (A) eIF1 from *Homo sapiens* (PDB id: 2IF1) and (B) YciH from *Escherichia coli* (PDB id: 1D1R). The α-helices, β-sheets and loops are colored in pale green, red and magenta, respectively.

### 1.1.3.2 EUKARYOTIC TRANSLATION INITIATION FACTOR 2B

Upon start codon recognition, the eIF2-bound GTP is hydrolysed to GDP in a reaction triggered by eIF5. The inactive eIF2•GDP disengages from the 40S ribosomal subunit and can be reutilised in the subsequent cycles of translation initiation only when the GDP is replaced by a GTP. This essential conversion is accomplished by a eukaryotic

translation IF, eIF2B which functions to promote the guanine nucleotide exchange (Erickson and Hannig, 1996; Kapp and Lorsch, 2004) (Figure 1.6).

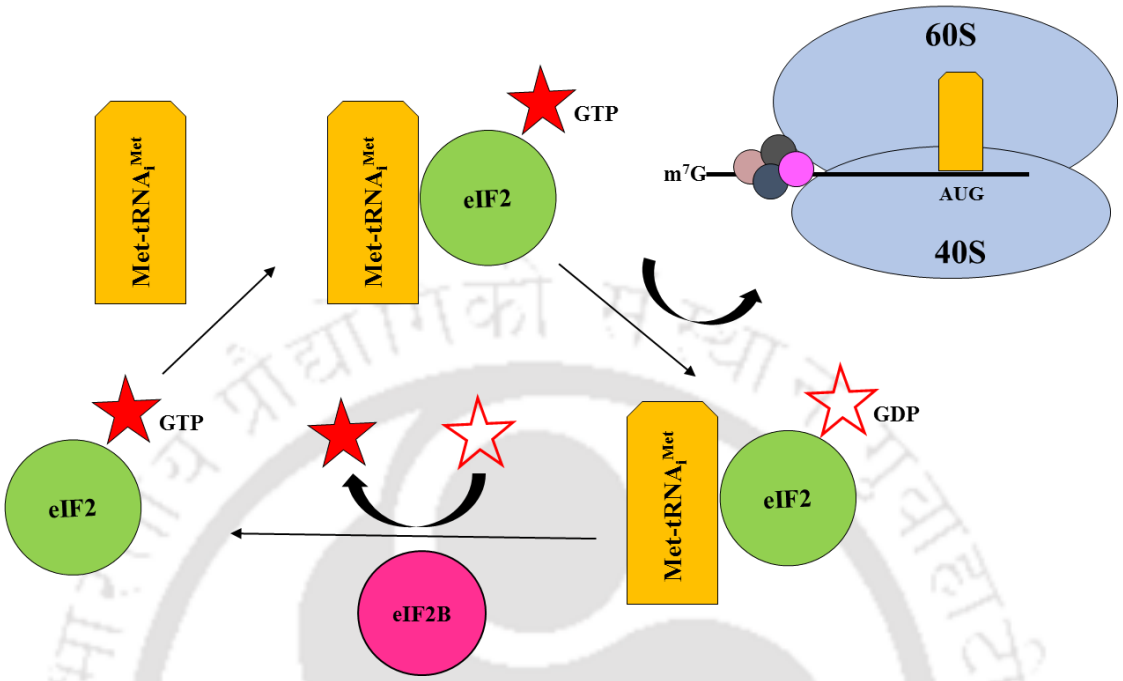



**Figure 1.6. A pictorial representation of the process of guanine nucleotide exchange by eIF2B during eukaryotic translation initiation.** The small and large ribosomal subunits are shown as small and large blue ovals, respectively. eIF2 and eIF2B are shown as green and pink spheres, respectively, while Met-tRNA$_i^{Met}$ is shown as yellow rectangle. GTP and GDP are shown as red star and white star with red border, respectively. The mRNA strand is shown as a black line.

eIF2B is a heterodecameric protein with two copies each of five subunits (α, β, γ, δ and ε) in which α, β and δ form the regulatory subcomplex while γ and ε form the catalytic subcomplex. High mutual similarity, both at the primary as well as tertiary structures, exists among the regulatory subunit as well as between the catalytic subunits (Wang et al., 2012; Wortham et al., 2014). In the three-dimensional crystal structure of eIF2B, two copies each of α, β, and δ engage together to form a hexameric subcomplex which resides at the centre with the two heterodimeric catalytic subcomplexes (γ and ε) adhered on the opposite sides (Figure 1.7A). The entire assembly is majorly mediated by the interaction occurring between the β and ε as well as δ and γ subunits (Wang et al., 2012; Bogorad et al., 2014; Gordiyenko et al., 2014; Wortham et al., 2014; Kashiwagi et al., 2016).

Each of the regulatory subunits consists of an all α-helical NTD and a Rossmann-fold-like CTD (Figure 1.7B-D). The regulatory subcomplex is assembled via interactions majorly occurring at the CTD of α, β and δ subunits. The arrangement of the regulatory subcomplex closely resembles the homohexameric structure of an enzyme ribose-1,5-bisphosphate isomerase (R15Pi) (Kuhle et al., 2015; Kashiwagi et al., 2016). On the other hand, the catalytic subunit comprises of an NTD consisting of mostly parallel β-strands surrounded by α-helices and a left-handed β-helix domain containing CTD (Figure 1.7E, F) (Kashiwagi et al., 2016). The structure of the catalytic subunits of eIF2B are highly comparable to the subunit structure of the homotetrameric ADP-glucose (ADPGlc) pyrophosphorylase (Jin et al., 2005).

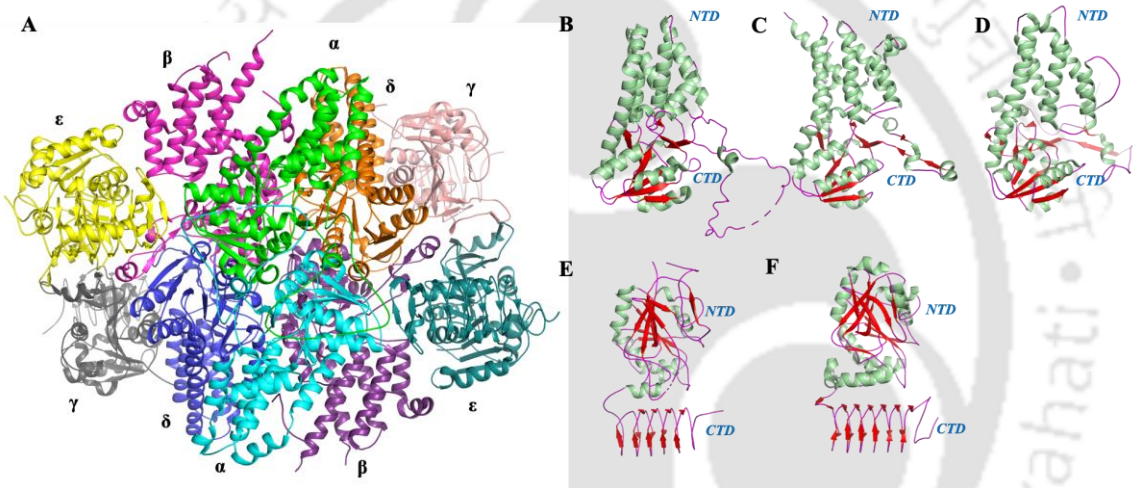



**Figure 1.7. Three-dimensional structure of eIF2B.** (A) Three-dimensional heterodecameric structure of eIF2B (PDB id: 5B04). All the subunits of the heterodecamer are distinguished with different colors. Tertiary structure of the (B) α, (C) β, (D) δ, (E) γ and (F) ε subunits of eIF2B with the α-helices, β-sheets and loops highlighted in pale green, red and magenta, respectively.

The importance of eIF2B during eukaryotic translation initiation is accentuated by the fact that it acts as a key limiting factor that imposes translation regulation in response to cellular stresses. Under various stress conditions, different kinases such as general control nonderepressible 2 (GCN2), heme-regulated inhibitor kinase (HRI), double-stranded RNA-activated protein kinase or protein kinase R (PKR) and PKR-like endoplasmic reticulum-resident kinase (PERK) phosphorylates eIF2 at an evolutionarily conserved serine residue (Ser51) in the NTD of the α-subunit yielding an eIF2αP.

Phosphorylation of eIF2α at Ser51 triggers a stronger binding affinity with the regulatory subcomplex of eIF2B resulting in the further unavailability of Met-tRNA$_i^{Met}$ to the ribosome leading to a global translation repression (Yang and Hinnebusch, 1996; Krishnamoorthy et al., 2001; Pavitt, 2005; Wek et al., 2006; Jackson et al., 2010) (Figure 1.8). The interaction between eIF2 and eIF2B occurs at the cavity formed around the centre of one set of regulatory subunits, thus, giving rise to two eIF2αP binding sites on the top and bottom ends of hexameric regulatory subcomplex (Kashiwagi et al., 2016).

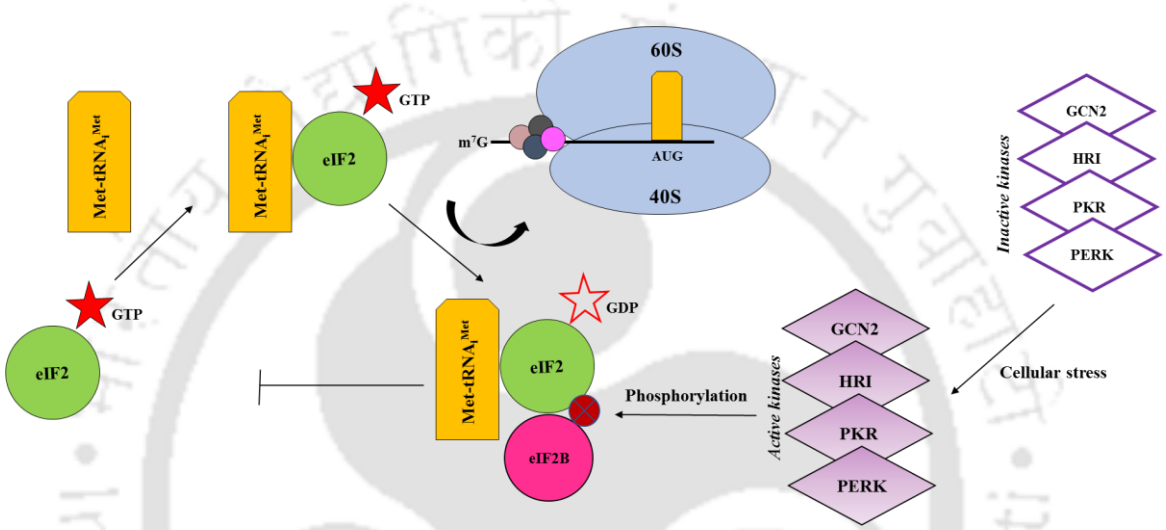



**Figure 1.8. A pictorial representation of the stages involved in regulation of eIF2B during translation initiation in eukaryotes.** The small and large ribosomal subunits are shown as small and large blue ovals, respectively. eIF2 and eIF2B are shown as green and pink spheres, respectively, while Met-tRNA$_i^{Met}$ is shown as yellow rectangle. The inactive and active kinases are shown as white diamond with violet border and violet diamond, respectively. The phosphate group added to eIF2B is shown as a red sphere with a cross within. GTP and GDP are shown as red star and white star with red border, respectively. The mRNA strand is shown as a black line.

### 1.1.3.3 RIBOSE-1,5-BISPHOSPHATE ISOMERASE

The enzyme ribose-1,5-bisphosphate isomerase (R15Pi), which shares high sequence as well as structure similarity with the regulatory subunits (α, β, and δ) of eIF2B, is involved in a metabolic pathway termed as pentose bisphosphate pathway (Nakamura et al., 2012, Kashiwagi et al., 2016). This pathway compensates for the lack of a canonical pentose phosphate pathway in archaea and is involved in the production of trioses and hexoses from pentoses (Grochowski et al., 2005; Soderberg, 2005; Orita et al., 2006;

Aono et al., 2015). In this pathway, nucleosides (the main source of pentoses) are converted into trioses and hexoses via the formation of ribose-1,5-bisphosphate (R15P). The molecule R15P can either be generated through the formation of ribose-1-phosphate (R1P) with the help of the enzymes, nucleoside phosphorylase and ADP-dependent ribose-1-phosphate (ADP-R1P) kinase or via the nucleoside 5′-monophosphate (NMP) degradation pathway (Aono et al., 2015). During the NMP dergradation pathway, R15Pi catalyses the isomerisation of R15P (substrate) to ribulose-1,5-bisphosphate (RuBP, product) via an enzymatic mechanism involving the formation of a *cis*-phosphoenolate intermediate. The RuBP is further utilised by ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) to produce 3-phosphoglycerate (3-PGA) which is ultimately directed towards the glycolysis or gluconeogenesis pathway (Sato et al., 2007; Aono et al., 2012, 2015; Nakamura et al., 2012) (Figure 1.9).

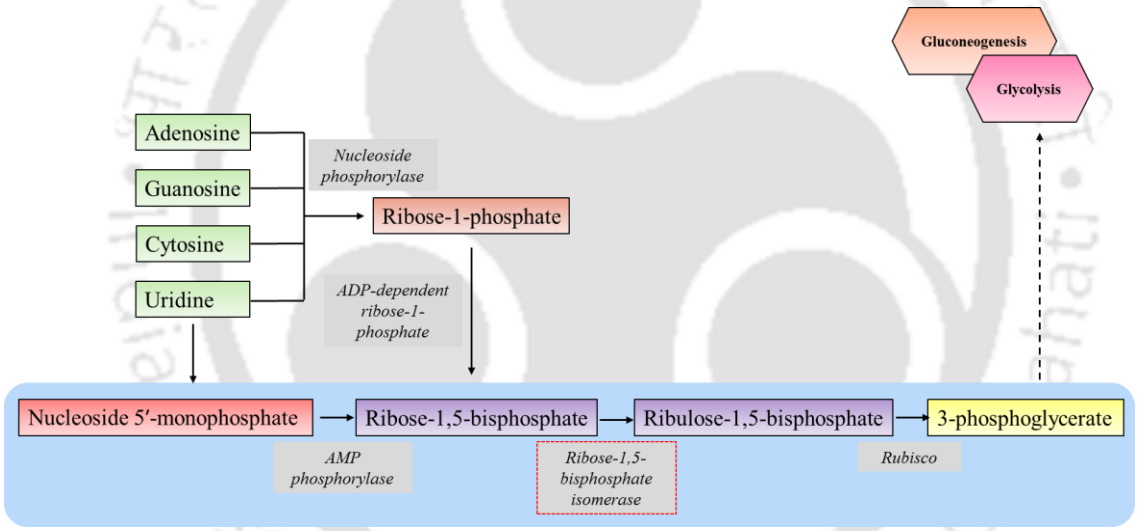



**Figure 1.9. A schematic representation of pentose bisphosphate pathway in archaea.** The enzymes involved in the pathway are shown in grey background while the metabolites formed during process are shown in enclosed boxes of different colors. The NMP degradation pathway is highlighted with blue background.

Similar to the regulatory subunits of eIF2B, R15Pi monomer comprises of a α-helical N-terminal domain (NTD) and a αβα-sandwich C-terminal domain (CTD) with the active-site pocket residing in between the two domains (Figure 1.10A). Two such monomers of R15Pi interact through their CTD to form a dimer and three dimers assemble to form the homohexameric structure (Figure 1.10B, C). In the homohexameric state, all the

monomers are held together via their CTD while the NTD remains completely free (Figure 1.10C).

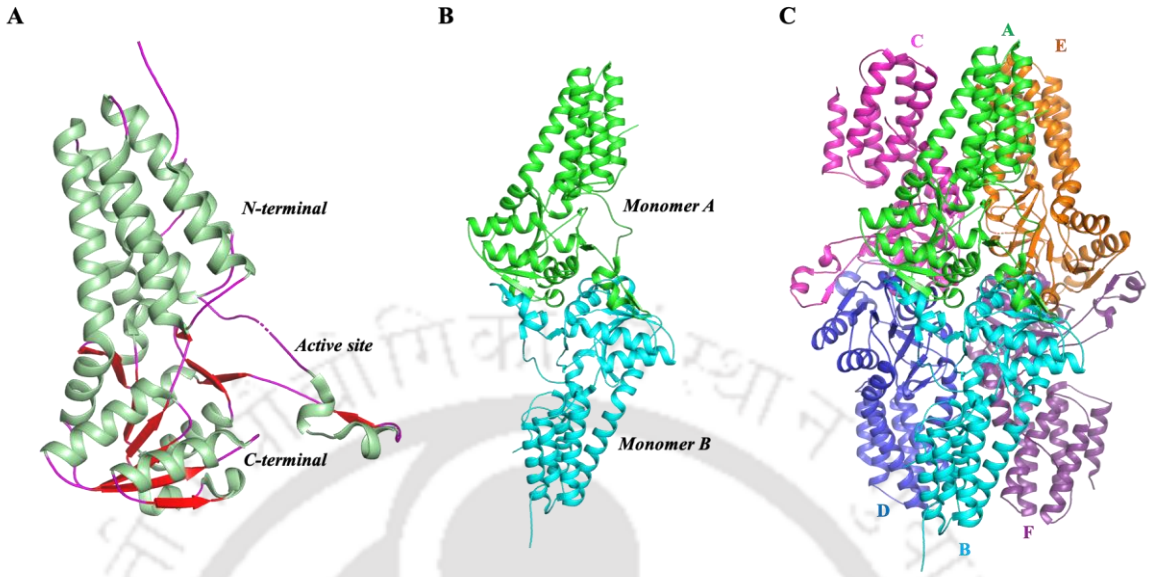



**Figure 1.10. Three-dimensional structure of enzyme R15Pi.** (A) Monomer, (B) dimer and (C) hexamer of enzyme R15Pi (PDB id: 3A11). The α-helices, β-sheets and loops in the monomer are highlighted in pale green, red and magenta, respectively. The subunits (A-F) are shown in different colors.

### 1.1.3.4 5-METHYLTHIORIBOSE 1-PHOSPHATE ISOMERASE

The enzyme 5-methylthioribose 1-phosphate isomerase (M1Pi) is involved in the universal methionine salvage pathway (MSP) which reprocesses sulphur-containing cellular metabolites to the amino acid methionine (Sekowska et al., 2004; Albers, 2009). During MSP, 5-methylthioadenosine (MTA) is cleaved and phosphorylated to 5-methylthioribose 1-phosphate (MTR-1-P). The production of MTR-1-P occurs either in one step where a phosphorylase is involved or two steps using a nucleosidase and a kinase. M1Pi then catalyses the conversion of MTR-1-P to 5-methylthioribulose 1-phosphate (MTRu-1-P) during MSP (Ashida et al., 2003). MTRu-1-P is finally converted to methionine via a series of actions involving ~11 enzymes including a dehydratase, an enolase, a phosphatase, a deoxygenate and transaminases (Sekowska and Danchin, 2002; Sekowska et al., 2004; Pirkov et al., 2008) (Figure 1.11).

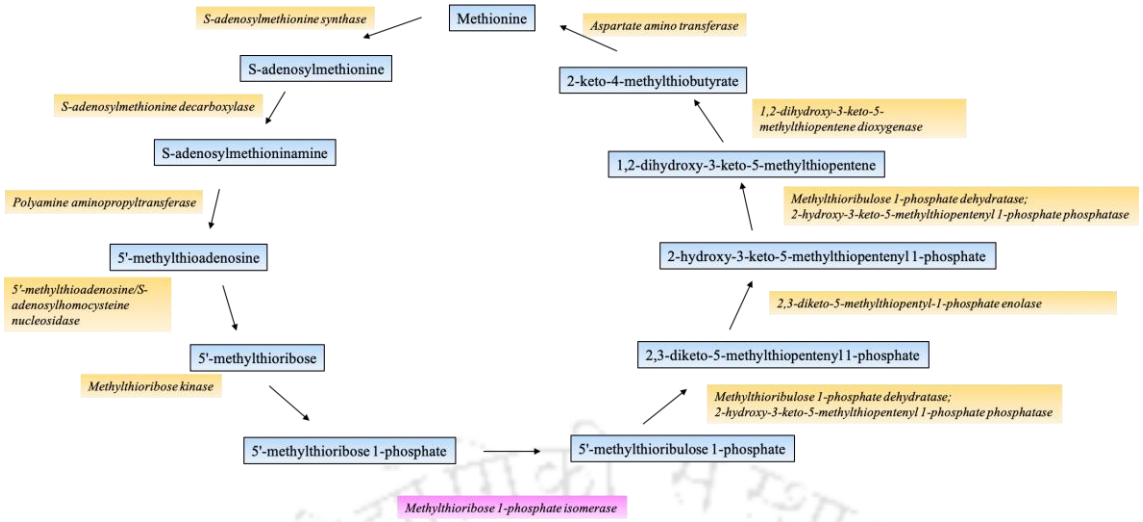



**Figure 1.11. A schematic representation of methionine salvage pathway (MSP).** The enzymes involved in the pathway are shown in yellow background while the metabolites formed during process are shown in blue enclosed boxes. The enzyme M1Pi is highlighted in magenta background.

Structurally, the monomer of M1Pi highly resembles the three-dimensional structure of the regulatory subunits of eIF2B as well as that of R15Pi. The monomer of M1Pi consists of a mostly α-helical NTD and a αβα-sandwich domain as the CTD. The active-site pocket is formed in a cavity in between the NTD and CTD (Figure 1.12A). Although, the monomeric arrangement of M1Pi highly resembles that of the regulatory subunits of eIF2B and R15Pi, their quaternary structure differs. Unlike the regulatory subcomplex of eIF2B and R15Pi which assembles into hexameric state, M1Pi is present only as a dimer (Figure 1.12B).

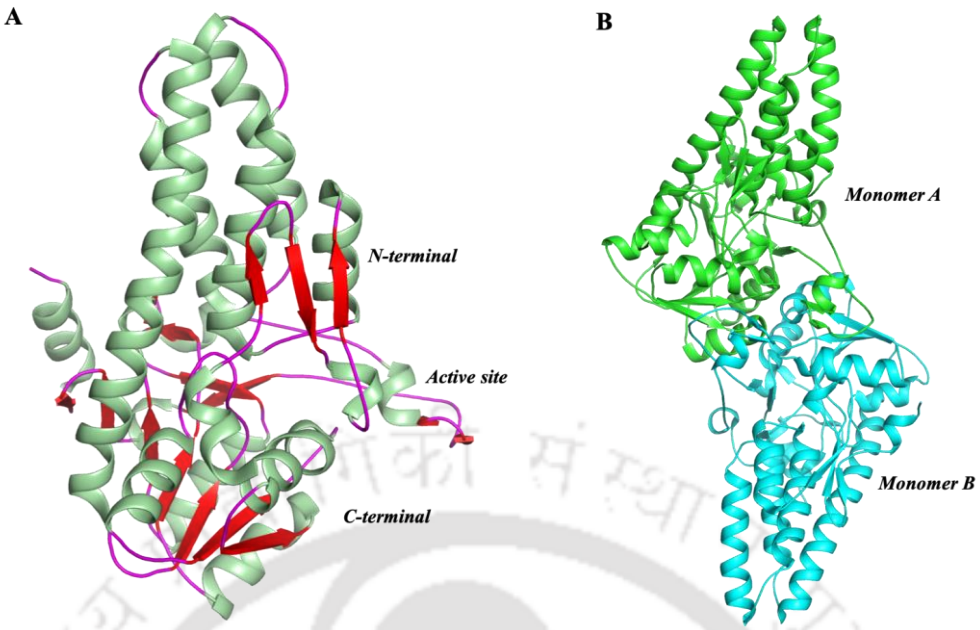



**Figure 1.12. Three-dimensional structure of enzyme M1Pi.** (A) Monomer and (B) dimer of the enzyme M1Pi (PDB id: 2YVK). The α-helices, β-sheets and loops in the monomer are highlighted in pale green, red and magenta, respectively.

### 1.1.3.5 NDP-SUGAR PYROPHOSPHORYLASE

NDP-sugar pyrophosphorylases (NSPases), also known as sugar nucleotide transferases (SNTs), are a group of enzymes that synthesizes nucleotide diphospho-sugars (NDP-sugars) or 'activated sugars' utilizing sugar-phosphate (sugar 1-phosphate) and nucleotide triphosphate (NTP) as their precursors (Figure 1.13A). A diverse repertoire of NSPases exists to accept an array of substrates to give rise to a myriad of NDP sugars from the various possible combinations of nucleotides and sugars. Nevertheless, all NSPases follow a common catalytic mechanism and employs either one or two magnesium ions. Based on the metal coordinating motifs, the diverse family of NSPases is divided into two major groups- Group I and Group II, which are further subdivided into IA, IB, IC, IIA and IIB (Jagtap et al., 2013). Interestingly, the specificity or promiscuity of an NSPase has also been associated to the different subgroups. The members of subgroup IC are found to be unusually promiscuous while those belonging to subgroups IA and IIB exhibit moderate promiscuity. Moreover, NSPases from thermophilic organisms are generally considered to be highly promiscuous while those from mesophillic exhibit limited promiscuity (Bais et al., 2018).

The three-dimensional structure of NSPase is highly similar to the catalytic (γ and ε) subunits of eIF2B. The NTD resembles a dinucleotide-binding Rossmann fold which contains mostly parallel and mixed seven-stranded β-sheet covered by α-helices while the CTD adopts a left-handed parallel β-sheets. The active-site pocket is present in the NTD (Figure 1.13B) (Jin et al., 2005).

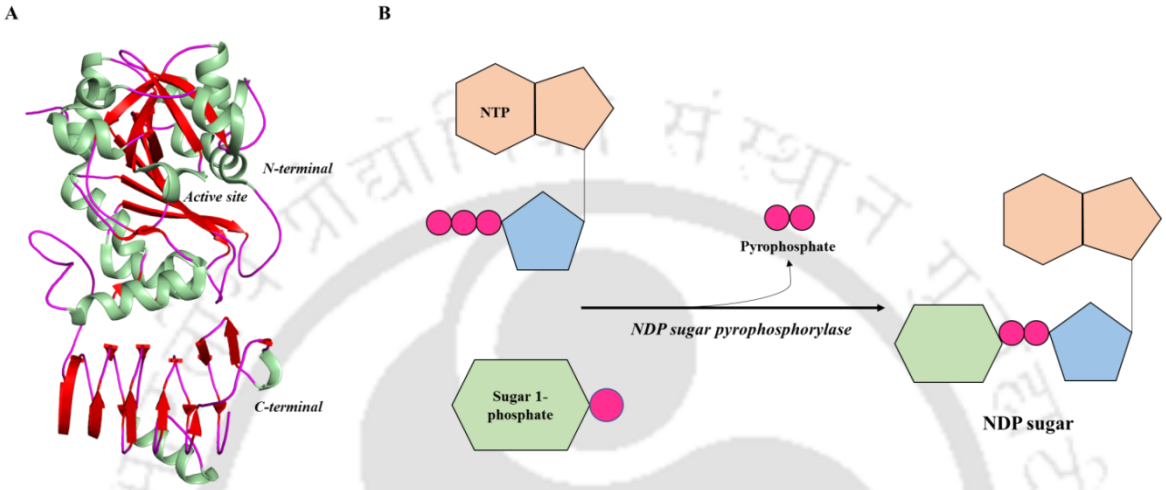



**Figure 1.13. Structure and function of NSPase.** (A) Three-dimensional structure of a NSPase ADPGlc pyrophosphorylase (PDB id: 1YP4). The α-helices, β-sheets and loops in the monomer are highlighted in pale green, red and magenta, respectively. (B) A pictorial representation of the reaction catalyzed by NSPases.

## 1.2 RATIONALE OF THE STUDY

The relative simplicity of the archaeal system owing to the presence of lesser number of translation IFs which are homologous to eukaryotes along with their inherent stability often makes it a desirable system to understand the fundamental mechanisms of the otherwise highly complicated eukaryotic system. However, the uncertainty regarding the presence of the functional homologs of eukaryotic translation IFs, eIF2B and eIF1 in archaea questions the analogy between the archaeal and eukaryal translation initiation mechanism. Initially, the availability of these translation IFs in archaea were merely presumed based on the sequence homology between the proteins (Kyrpides and Woese, 1998a, 1998b; Kakuta et al., 2004). However, eIF1 and eIF2B also shares sequence similarity with another set of functionally unrelated proteins viz. YciH, R15Pi, M1Pi and NSPases, most of which participate in various other cellular processes. Even though, the

three-dimensional structures of all the proteins are available, no in-depth study has been performed till date to understand the notable differences in the structure that might confer functional specificity to these proteins. Thus, in this study, we aim to investigate the (non)availability of a functional aIF1 and aIF2B in a hyperthermophillic archaeon, *Pyrococcus horikoshii* OT3. This would further enhance our understanding of the protein biosynthesis mechanism in archaea and also allow us to answer the long-borne question of the analogy between the mechanism of translation initiation in archaea and eukarya.

## 1.3 OBJECTIVES

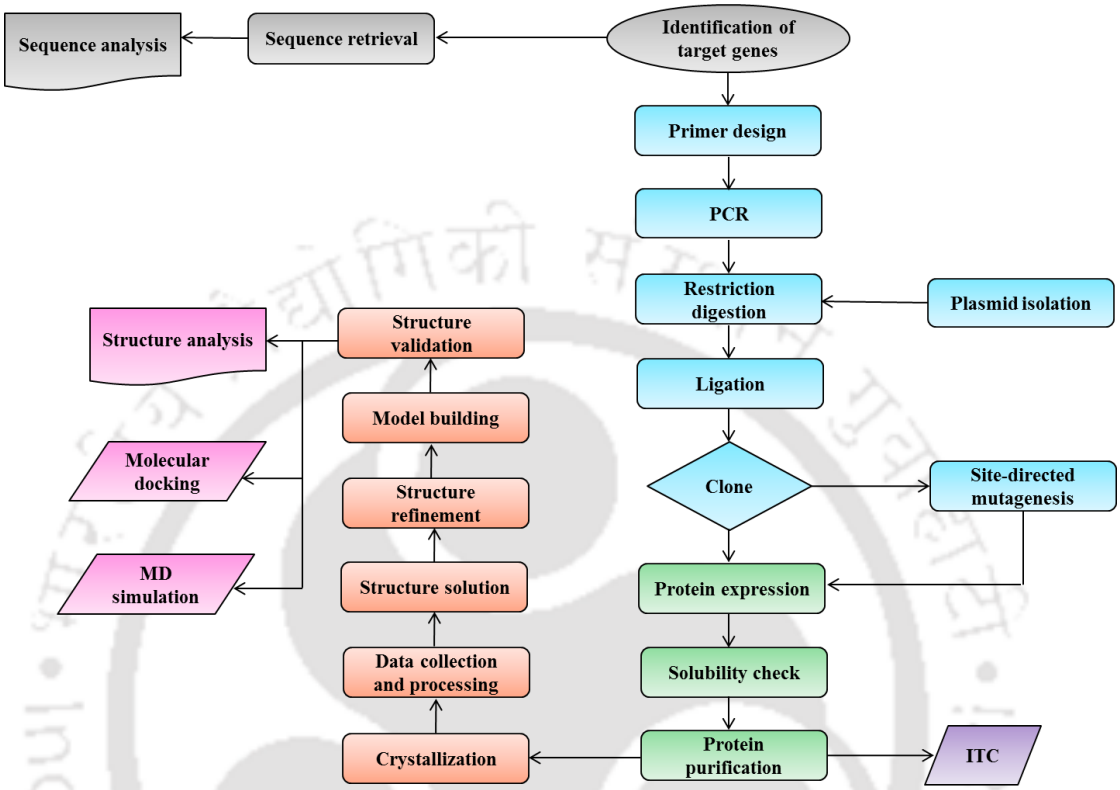
To illuminate the (dis)similarities present between the mechanism of archaeal and eukaryal protein translation initiation, the following objectives were formulated-

- ✓ *In silico* characterization of putative translation initiation factors (PH1771.1, PH0208, PH0702, PH1022, and PH1697) in archaea, particularly, *P. horikoshii* OT3.
- ✓ Cloning, over expression and purification of PH1771.1, PH0208, PH0702, PH1022 and PH1697.
- ✓ Crystallization and structural determination of the purified proteins.
- ✓ Biochemical and biophysical characterization of the above mentioned proteins and their complexes.

# CHAPTER 2

## Materials and methods

This chapter deals with the methodology involved while performing the experiments of the work reported in this thesis. The major experimental approaches are construction of expression plasmids, protein over expression and purification, protein crystallization, isothermal titration calorimetry, molecular dynamics simulation and molecular docking.




## 2.1 PROCUREMENT OF MATERIALS

All chemicals used during this study were obtained from Sigma-Aldrich, HiMedia, Merck Millipore and Sisco Research Laboratories Pvt. Ltd. (SRL). The enzymes used during molecular cloning were purchased from New England Biolabs (NEB) and/or Thermo Fisher Scientific. The kits used during molecular cloning which includes plasmid isolation, PCR clean up, gel extraction, etc. were obtained from QIAGEN and/or HiMedia. For site-directed mutagenesis, Q5 site-specific mutagenesis kit was obtained from NEB. The Ni-Nitrilotriacetic acid (Ni-NTA) beads used during purification were purchased from QIAGEN. For concentration of the purified proteins, concentrators were used from Amicon Ultra centrifugal filter unit (Milipore) and/or Vivaspin turbo 15 (Sartorius). The crystallization buffers and plates were purchased from Hampton Research for crystallization experiments. The ligands used during crystallization and isothermal titration calorimetry (ITC) experiments were obtained from Sigma-Aldrich.

## 2.2 CONSTRUCTION OF EXPRESSION PLASMIDS

Molecular cloning is an essential technique that can be implemented to obtain recombinant DNA molecules. The method includes inserting recombinant DNA molecule into a vector that will replicate the DNA molecule inside the host cell. To obtain the clones used in the present study, the following steps were performed.

### PREPARATION OF COMPETENT CELLS

The competent cells used during this study were prepared using the calcium chloride method or the chemical method. This method includes exposing the cells grown till log phase to $CaCl_2$ for a time period of 30 min at 4°C. The $CaCl_2$ treatment renders the cells competent of taking up foreign DNA efficiently.

### DESIGNING OF GENE-SPECIFIC PRIMERS

Primers specific to the genes were designed manually and analysed in the online tool OligoAnalyzer for calculating their GC content, melting temperature ($T_m$), dimerization tendencies, etc. Restriction digestion sites for restriction enzymes were added at the 5′ and 3′ ends of the primer to obtain cohesive or sticky ends. A polyhistidine (6xHis) tag was added to the 5′ end of the primer in all the cases.

### POLYMERASE CHAIN REACTION

The targeted genes were amplified using the genome of *P. horikoshii* OT3 with suitable primers using polymerase chain reaction (PCR). The PCR reaction was broadly performed in three steps viz, denaturation, annealing and extension. The temperature of denaturation and extension steps were kept 95 and 72°C, respectively, for all the cases while the annealing temperature varied according to the desired $T_m$ to the genes. The initial and final denaturation were kept for a time period of 4 min and 30 sec, respectively, while the initial and final extension were kept 35 sec and 10 min, respectively (Figure 2.1). The PCR product obtained were confirmed by running in a 0.8% agarose gel along with a DNA ladder.

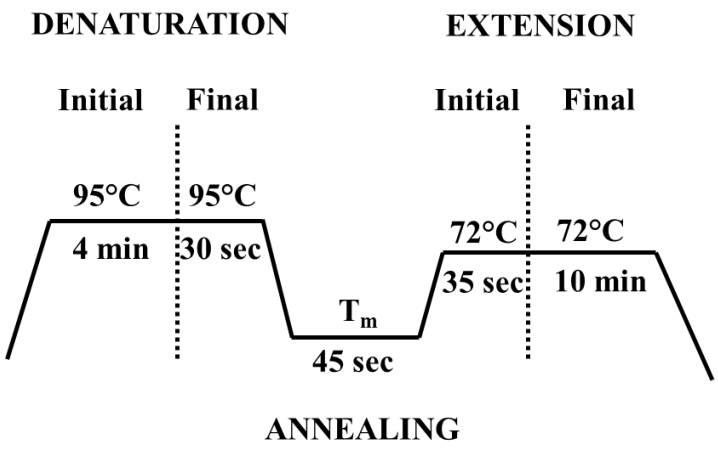



**Figure 2.1. A general representation of the PCR cycle.** The three stages (denaturation, annealing and extension) of a PCR cycle are depicted along with their temperatures and time periods. $T_m$ stands for melting temperature.

## RESTRICTION DIGESTION

The amplified gene products were subjected to restriction enzyme-specific double digestion for the formation of sticky ends at the 5′ and 3′ ends of the genes. The double digestion was performed at 37°C for a time period of 4 hours. Along with the PCR amplified products, the vector or plasmid into which the gene has to be incorporated was also subjected to double digestion with the same set of restriction enzymes. The double digestion of the vector was also carried out at 37°C for 4 hours. However, after the third hour, alkaline phosphatase was added to the vector in order to remove the 5′-phosphate group of the vector which in turn would prevent self-ligation of the vector.

## LIGATION

The gene as well the vector both of which contains sticky ends were ligated using T4 DNA ligase during the ligation reaction. The ligation reactions were performed either at 22°C for 4 hours or at 16°C for 16 hours. The ligation mixture was then transfered into *Escherichi coli* DH5α competent cells and the cells were allowed to grow in Luria Bertani (LB) agar plates supplemented with antibiotic at 37°C for 12 hours. The colonies obtained were then further subjected to screening for identification of positive clones. For confirmation of positive clones, the recombinant plasmids (or vectors) were isolated from the *E. coli* DH5α cells using either alkaline lysis method or commercially available plasmid isolation kits. The recombinant plasmids were subjected to double digestion at

37°C for two hours and the presence of the desired gene was detected in 0.8 % agarose gel in the presence of a DNA ladder. The positive clones were also confirmed by plasmid sequencing.

### SITE-DIRECTED MUTAGENESIS

The site-directed mutagenesis was performed using the Q5 site-specific mutagenesis kit. Initially, specific primers were designed to incorporate mutations at the desired site of the gene. The primers were further used to amplify the vector containing the gene with the desired mutation. The amplified products were then subjected to kinase, ligase and DpnI treatment and finally competent *E. coli* DH5α cells were transformed with the entire reaction mixture. The mutation at the desired site of the gene was further confirmed by sequencing.

## 2.3 PROTEIN OVER EXPRESSION AND PURIFICATION

### OVER EXPRESSION

For over expression of the desired proteins, expression hosts were transformed with the recombinant vectors. A variety of expression host systems were used during this study which includes *E. coli* BL21 (DE3), *E. coli* Rosetta (DE3), *E. coli* BL21 (DE3) RIL, etc. The expressions of the desired proteins were induced using different concentrations (0.1-1.0 mM) of isopropyl β-D-1-thiogalactopyranoside (IPTG). IPTG triggers the transcription of the lac operon which in turn induces the expression of the gene which is under the control of the lac operator (Donovan et al., 1996). The cells were grown at 37°C and the expressions of the proteins were monitored at each hour. For subsequent steps of protein purification, the concentration of IPTG and cell growth time was fixed at which maximum expression of the proteins was obtained.

### SOLUBILITY OF THE PROTEIN

For solubility check, the IPTG induced cells were recovered by centrifugation and lysed using appropriate lysis buffer. The cells were then mechanically disrupted using a sonicator with on and off time of 3 and 5 sec, respectively with a frequency of 25%. The cell lysate obtained was further centrifuged which yielded a pellet and a supernatant

fraction. Both the fractions were run in a 12% SDS-PAGE along with a protein marker. If the protein was present in the supernatant fraction, it was further used for protein purification. However, further optimization steps were performed if the protein was present in the pellet fraction. The different optimization steps included varying the IPTG concentration, temperatures, incubation time, culture media, use of buffers containing triton X-100, glycerol, ethanol, etc.

**PROTEIN PURIFICATION**

Purification of the desired proteins was obtained using immobilized metal affinity chromatography (IMAC). The proteins of interest were purified using Ni-NTA affinity chromatography. The proteins containing 6xHis-tag binds to the Ni-NTA groups via interaction between the imidazole ring of histidine residues. During purification, the proteins were allowed to bind to the Ni-NTA for at least 2 hours. Binding of the non-specific proteins were minimized by washing the column with wash buffers containing low concentrations of imidazole (5-20 mM). For eluting the proteins of interest from the Ni-NTA beads, a higher concentration of imidazole (200-400 mM) was used which causes the dissociation of the bound proteins. The eluted proteins were dialyzed to remove imidazole and finally concentrated using Amicon Ultra centrifugal filter unit (Milipore) and/or Vivaspin turbo 15 (Sartorius). The final concentration of the proteins were measured at a wavelength of 280 nm using ultraviolet-visible (UV) spectrophotometer.

## 2.4 PROTEIN CRYSTALLIZATION

Determination of the structure of biological molecules by X-ray crystallography requires the formation of crystals. Crystal formation requires supersaturation of macromolecules in the presence of precipitating agent(s) which promote(s) nucleation of protein crystals in the solution (McPherson, 1999, 2004). The two major techniques utilised in the present study include (a) the microbatch-under-oil and (b) the vapor-diffusion crystallization. In microbatch-under-oil crystallization, an appropriate amount of protein are mixed directly with precipitant and the mixture are covered with different ratios of paraffin and silicone oil to regulate the rate of diffusion. On the other hand, in vapor-diffusion method a drop containing appropriate amount of protein and precipitant

solution are equilibrated against pure precipitant in a sealed chamber. In due course of time, diffusion of water vapor occurs until the osmolarity of the drop and precipitant becomes equal. Vapor-diffusion are majorly carried out using two different techniques, (a) hanging drop and (b) sitting drop. In hanging-drop, a siliconised cover slip containing the protein and precipitant drop are placed over the precipitant and sealed with silicone grease. While in sitting-drop, no siliconised cover slip are used and the drop of protein and precipitant was kept seated with precipitant surrounding the drop. In both the methods i.e. microbatch-under-oil and vapor diffusion, the set-ups were placed under optimum temperatures such as 4 and/or 20°C for optimal crystal growth. In cases where optimal crystal growth was not obtained, further optimization was performed which are mentioned in details in the respective chapters.

## 2.5 X-RAY INTENSITY DATA COLLECTION AND PROCESSING

### 2.5.1 X-RAY INTENSITY DATA COLLECTION

After attaining appropriate crystals, the diffraction pattern of the crystals were obtained using monochromatic X-ray beam. A number of options such as crystal-to-detector distance, exposure time, oscillation angle, redundancy, resolution, etc. (Dauter, 1999) are available during data collection. The crystal-to-detector distance was adjusted so as to record and resolve the diffraction maxima on the image plate to the resolution limit of the crystal. In all the cases, the distance between crystal and detector was kept at a range of 80-200 mm. The exposure time was set to provide reasonable statistics at the highest resolution. Care was taken not to overload the detector with strong low-angle spots. In almost all the cases, an exposure time of 300-600 sec was maintained with an oscillation angle of 1.0° which seemed to be a good compromise between speed and the data quality.

### 2.5.2 DATA PROCESSING

The data sets obtained were initially processed using the program HKL-3000 (Minor et al., 2006) and finally by iMosflm (Battye et al., 2011) while the scaling was performed using Aimless (Evans and Murshudov, 2013) embedded in the CCP4 package (Winn et al., 2011). iMosflm is involved in performing a series of actions which include indexing

images (either single or together), estimate the mosaic spread, accurate refinement of the crystal parameters, calculating data collection strategy, integrating a series of images and providing a statistics on the processing. Following integration, scaling of the data set was performed using Aimless which determines the likely point group and possible space group. The structure factors were calculated using the program CTRUNCATE available in CCP4 package (Winn et al., 2011).

The assessment of the high-resolution limit of the diffraction pattern is done in two ways: the first is the mean ratio of the intensity to the error [I/σ(I)] and the second, $R_{merge}$, is the agreement between the symmetry related reflections (Drenth, 2007; Powell, 2017) and is given by

$$R_{merge} = \frac{\sum_{hkl}\sum_{j=1}^{N}|I_{hkl}-I_{hkl}(j)|}{\sum_{hkl}\sum_{j=1}^{N}I_{hkl}(j)}$$

where $I_{hkl}$ is the intensity of an individual reflection with indices *hkl* and $I_{hkl}$ *(j)* is the mean value of the intensity for all reflections with indices *hkl*.

*Rmeas* is proposed as a substitute for the conventional $R_{merge.}$ It describes the precision of the individual measurements, independent of how often a given reflection has been measured (Drenth, 2007; Powell, 2017).

$$R_{meas} = \frac{\sum_{hkl}\left[\frac{N}{(N-1)}\right]^{1/2}\sum_{i}|I_i(hkl)-\overline{I(hkl)}|}{\sum_{hkl}\sum_{i}I_i(hkl)}$$

where *N* is the redundancy of the data and $\overline{I(hkl)}$ the average intensity.

$R_{p.i.m}$ describes the precision of the averaged measurement. It is useful for predicting the performance of the data set in structure determination (phasing, refinement), because of the use of averaged measurements during phasing and refinement (Drenth, 2007; Powell, 2017).

$$R_{p.i.m.} = \frac{\sum_{hkl}\frac{1}{(N-1)^{1/2}}\sum_{i}|I_i(hkl)-\overline{I(hkl)}|}{\sum_{hkl}\sum_{i}I(hkl)}$$

where *N* is the redundancy of the data and $\overline{I(hkl)}$ the average intensity.

In cases where a common scale cannot be established between model data and observed data, the linear correlation coefficient is used. The correlation coefficient is calculated between two random half data sets (Rupp, 2009; Powell, 2017) and is given by

$$CC_{1/2} = \frac{\sum_{h}(|F_h(obs)|-|\overline{F_h(obs)}|\times(|F_h(calc)|-|\overline{F_h(calc)}|)}{\left[\sum_{h}(|F_h(obs)|-|\overline{F_h}(obs)|)^2\times\sum_{h}(|F_h(calc)|-(|\overline{F_h(calc)}|)^2\right]^{1/2}}.$$

## 2.6 STRUCTURE SOLUTION

During structure solution, the initial phases are normally obtained using one of the following mehods: (i) Multiple-wavelength anomalous dispersion (MAD), (ii) Multiple isomorphous replacement (MIR) and (iii) Molecular replacement (MR). All the crystal structures in this thesis were solved by MR method (Rossmann and Blow, 1962; Rossmann, 1972) using the program Phaser (McCoy et al., 2007). During MR, a known molecular model was used to solve the unknown structure of a homologous protein. In this method, the previously known structure helps in providing initial estimates of the phases of the new structure. A preliminary model was generated by trying all the possible orientations and positions of the known model in the unit cell of the target crystal. The phases calculated from the model were then utilised as phases for the reflections of the target protein. MR was performed using the program Phaser which is based on maximum likelihood probability theory and multivariate statistics (McCoy et al., 2007).

## 2.7 STRUCTURE REFINEMENT

Refinement involves bringing the agreement between the structural model and the experimental data. During refinement, optimisation of a posterior conditional probability of model parameters which include atomic coordinates, B factors, scale factors, parameters of solvent model and observed data such as amplitude and intensities of structure factors were performed. In this study, all refinements were carried out using the program Refmac5 (Vagin et al., 2004) embedded in the CCP4 suite. Refmac5 utilizes maximum likelihood and some elements of Bayesian statistics.

## 2.8 MODEL BUILDING

Model building was performed using the molecular-modelling package COOT (Crystallographic Object-Oriented Toolkit, Emsley and Cowtan, 2004). The online available tools PSAP (Protein Structure Analysis Package, Balamurugan et al., 2007) and PDB Goodies (Hussain et al., 2002) were also used during the different stages of model building. After every cycle of refinement, the model was inspected and manual model rebuilding was performed by inspecting the $2F_O$-$F_C$ and $F_O$-$F_C$ electron-density maps

contoured at 1σ and 3.0σ, respectively. The areas of poor electron density were examined with the maps contoured at a lower contour level. For examination and interpretation of the model against the electron-density maps, COOT was used.

## 2.9 CROSS-VALIDATION

The reliability of the fit of a model to the diffraction data is given by the R-factor, which measures the discrepancy between the observed structure factor amplitudes $F_{obs}$ and calculated structure factor amplitudes $|F_{calc}|$:

$$R=\frac{\sum_{hkl}||F_{obs}|-k|F_{calc}||}{\sum_{hkl}|F_{obs}|}$$

where *k* is a scale factor.

The method of statistical cross-validation by using R-factor is a more accurate indicator of model quality (Brünger, 1992). For cross-validation, the diffraction data are divided into a large working set (usually comprising of 90-95% of the data) and a small complementary test set (comprising the remaining of 10-5% of the data). The diffraction data present in the working set is used for refinement. It provides a more objective guide during model building and refinement process than the conventional R-factor.

$$R_{free}=\frac{\sum_{hkl\subset T}||F_{obs}|-k|F_{calc}||}{\sum_{hkl\subset T}|F_{obs}|}$$

where $hkl\subset T$ means all reflections belonging to test set *T* of unique reflections.

## 2.10 STRUCTURE VALIDATION

The structure validation programs analyse the geometrical parameters and list the root mean square deviation (RMSD) in bond lengths, bond angles and dihedral angles, etc. of the refined structure. The tools used for structure validation are described below.

### PROCHECK

The stereo chemical quality of the protein structures was analysed using the program PROCHECK (Laskowski et al., 1993). It provides a comprehensive residue-by-residue listing and gives an estimate of the overall quality of the structure. One of the few outputs of PROCHECK includes the Ramachandran plot which helps in analysing the

stereochemistry of the protein molecule. The quality of the protein model was assessed after every refinement cycle.

**MOLPROBITY**

The assessment of the quality of the three-dimensional structures of the proteins was also performed using MolProbity (Davis et al., 2007). The program adds both polar and nonpolar hydrogen atoms to the structure. It also detects the presence of any short contact which is already present in the structure or has appeared upon addition of hydrogen. The program also analyses the local chemical environment and suggests possible flips for certain amino acid residues if required. Furthermore, the program detects the Ramachandran map outliers, torsion-angles outliers, etc.

## 2.11 ISOTHERMAL TITRATION CALORIMETRY

Isothermal titration calorimetry (ITC) was used to obtain the thermodynamic parameters during protein-ligand interaction. ITC determines the heat change that occurs upon bond formation between two biological molecules. Heat can either be absorbed (endothermic reaction) or released (exothermic reaction) during an ITC experiment. It provides valuable information about the interaction such as the binding stoichiometry (N), the thermodynamic parameters of the binding reaction (enthalpy, $\Delta H$; entropy, $\Delta S$ and Gibbs free energy, $\Delta G$), the strength of the interaction (the equilibrium association constant $K_a$). Along with structural information, knowledge of binding energetics can enhance the amount of information related to interaction between two biological molecules. During the ITC experiments performed during this study, the cell temperature was kept at 25°C and the reference power was set as 6 µcal $sec^{-1}$. The initial injection was delayed by 60 sec and during each subsequent injections 1.2 µl of ligand was added to the ITC cell containing the protein with a spacing of 120 sec in between the injections. The duration of injection and filter period were kept 2 and 5 sec, respectively.

## 2.12 ANALYSIS OF SEQUENCES AND STRUCTURES

Various programs used for sequence and structure analysis are described below.

### 2.12.1 SEQUENCE DATA AND ANALYSIS

The nucleotide sequences of all the genes were obtained from GenBank database of National Centre for Biotechnology (NCBI) (Benson et al., 2000). The amino acid sequences of proteins used in this study were downloaded from UniProtKB database (UniProt Consortium, 2018). The program ProtParam was used to compute various physical and chemical parameters such as molecular weight, theoretical pI, molar extinction coefficient, etc. for a given protein sequence (Gasteiger et al., 2005).

### 2.12.2 SEQUENCE ALIGNMENT

The programs BLAST (Basic Local Alignment Search Tool, Altschul et al., 1990) and Clustal Omega (Sievers and Higgins, 2014) were used for pairwise and multiple sequence alignments (MSA), respectively. During BLAST, the expect threshold was kept at 10 with a word size of 6. BLOSUM62 was used as the scoring matrix with conditional compositional score matrix adjustment. In Clustal Omega, default parameters were used for the number of combined and hidden Markov model (HMM) iterations. The aligned sequences were further decorated using the online tool ESPript (Easy Sequencing in PostScript, Gouet et al. 2003). The programs MUSTANG (MUltiple STructural AligNment AlGorithm, Konagurthu et al., 2006) and PROMALS3D (PROfile Multiple Alignment with predicted Local Structures and 3D constraints, Pei and Grishin, 2014) were used to align the sequences based on the three-dimensional structure. During PROMALS3D, the three-dimensional structures were aligned using the programs FAST (FAST Alignment and Search Tool, Zhu and Weng, 2005) and TM-align (Zhang and Skolnick, 2005).

### 2.12.3 PHYLOGENETIC TREE

To deduce the evolutionary relationship among proteins, phylogenetic trees were built. For sequence-based phylogenetic tree, MSA was initially performed using ClustalW embedded in the program MEGA (Molecular Evolutionary Genetic Analysis) version 6.0 and 7.0 (Tamura et al., 2013; Kumar at al., 2016). During MSA, the gap opening and gap extension penalties were kept 10 and 0.2, respectively. Gonnet was used as the protein weight matrix with gap separation distance of 4. The alignment file was then used to construct the phylogenetic tree by maximum likelihood (ML) or neighbor joining (NJ) method. The reliability of the tree was evaluated by bootstrapping and the branch lengths (0.2 or 0.5) were automatically assigned by the program. During tree building by ML

method, the substitution model used was Jones-Taylor-Thornton (JTT) model and the heuristic method utilized was nearest neighbor interchange (NNI). While for tree building using NJ method, Poisson model was kept as the substitution model. For the structure-based phylogenetic tree, a distance matrix was generated for all the proteins by calculating their RMSD in PyMOL (PyMOL Molecular Graphics System, Schrodinger, LLC). Subsequently, a dendrogram was generated in Newick (.nwk) file format from the distance matrix using the program dendroUPGMA (GarciaVallvé et al., 1999). The Newick file was futher used to generate the tree in MEGA.

### 2.12.4 SECONDARY STRUCTURE ELEMENTS

The program DSSP (Dictionary of Protein Secondary Structure, Kabsch and Sander, 1983) was used to assign the secondary structure of the three-dimensional protein structures.

### 2.12.5 STRUCTURAL COMPARISON

Comparison of three-dimensional structures was performed using the programs PyMOL and 3dSS (3-dimensional Structural Superposition, Sumathi et al., 2006). The web server DALI (Holm and Rosenström, 2010) was used to search the structural homologs in the Protein Data Bank (PDB) (Berman et al., 2000).

### 2.12.6 ELECTROSTATIC POTENTIALS AND OLIGOMERIC STATE

The program APBS (Adaptive Poisson-Boltzmann Solver, Baker et al., 2001) embedded in PyMOL was used to calculate the electrostatic potential charge distribution of the solvent accessible region of the protein structures. The web-server PISA (Proteins, Interfaces, Structures and Assemblies, Krissinel and Henrick, 2007) was used to explore the macromolecular interfaces and to predict the probable quaternary structures (or assemblies). The program PSAP (Protein Structure Analysis Package, Balamurugan et al., 2007) was used to identify the amino acid residues involved in interacting at the oligomeric interfaces.

### 2.12.7 STRUCTURE VISUALISATION AND ANALYSIS

The program PyMOL was used to visualise all the three-dimensional structures and also to generate the figures. Analysis of the three-dimensional structures and inter- and intra-molecular interactions were performed in the program COOT (Emsley and Cowtan,

2004). The volume and surface area of e active site of proteins were computed using the program CASTp (Binkowski et al., 2003).

## 2.13 MOLECULAR DYNAMICS SIMULATION

Biological molecules do not exist in static state in solutions and can be represented by an ensemble of conformations in equilibrium in solution. These different conformations can be understood at the atomic level using molecular dynamics (MD) simulation (Dodson et al., 2008). MD simulation was performed using the package GROMACS v.5.1.4 (Abraham et al., 2015). GROMACAS utilizes Newtonian equations of motions for simulating biochemical systems such as proteins, lipids and nucleic acids. In almost all the simulations, the explicit water molecules present in the three-dimensional crystal structures of the proteins were removed and the implicit flexible simple point charge (SPC/E) water model (unless otherwise specified) was used. For almost all the simulations performed in this study, AMBER03 force field embedded in the package GROMACS was utilized (Ponder and Case, 2003). In general, a cubic box keeping the minimum distance between the solute and the edge of the box of 1.0 nm was used to simulate the protein molecules. The sodium and chloride ions, wherever needed, were used to neutralize the system. To avoid the edge effects of the boxes during simulation, periodic boundary condition (PBC) was applied in all the three spatial directions. In most of the cases, the energy of the system was minimized using the steepest descent method with a maximum force cutoff of 1000 kJ $mol^{-1}$ $nm^{-1}$. For almost all the MD simulations, the solvent and ions were equilibrated in two phases viz. NVT (canonical or isothermal-isochoric) and NPT (isothermal-isobaric). In NVT, ensembles were generated for 100-1000 ps with reference temperature of 310 K. The reference temperature was controlled using a velocity rescaling thermostat (Bussi et al., 2007) with coupling constant of 0.1 ps. In NPT, ensembles were generated for 500-1000 ps with a reference pressure of 1 bar (or 100 kPa), which was controlled using the Parrinello-Rahman barostat (Parrinello and Rahman, 1981) with a coupling constant of 2 ps. The the particle mesh Ewald (PME) was used to compute the long-range electrostatic interactions were computed method (Darden et al., 1993; Essmann et al., 1995) while the short-range van der Waals interactions were computed using Verlet neighbor list calculation with a cutoff of 0.8 nm. The bond lengths were constrained using the P-LINCS algorithm (Hess, 2008). The

Newtonian motion of equations were integrated with a time step of 2 fs. In most of the cases, MD simulations werer performed for a minimum time period of 100 ns. The analyses were performed using the programs available in GROMACS and home-built shell scripts. The program Xmgrace (Turner, 2005) was used to prepare the graphs.

## 2.14 MOLECULAR DOCKING

Molecular docking experiments in this study were carried out using the program Autodock version 4.0 (Morris et al., 2009) which is based on an empirical free energy force field and rapid Lamarckian genetic algorithm (LGA) search method (Goodsell and Olson, 1990). Two types of docking viz. rigid and flexible were performed during this study. In rigid docking, both the ligand and receptor (protein) molecules are kept rigid where the molecules are rotated and translated in 6 degrees of freedom. On the other hand, in flexible docking, the ligand molecule was kept flexible around the rotatable bonds while keeping the receptor molecule rigid. In the case of flexible ligand, all the torsional angles were considered for rotation. In almost all cases, unless otherwise specified in the particular chapter, the water molecules present in the protein structures were removed before the molecular docking experiments. The hydrogen atoms to both the receptor and ligand molecules were added before each docking experiment. The partial charges to the atoms of the receptor molecules were added using the Kollman charges while for the ligand molecules, Gasteiger charges were computed. In most of the cases, the blind molecular docking was performed where the ligand molecule is allowed to search all the accessible surface of the receptor molecule. Thus, grids were generated taking the protein center of mass and covering all the whole receptor with a grid points of 126x126x126 and spacing of 0.375 Å. In almost all cases, LGA with a total of 2000 runs was used to search the best possible conformation of the ligand. The ligand conformations after docking were clustered with a RMSD of 2.0 Å. Although in most of the cases, the ligand conformation having the lowest free energy of bindng were considered for further analysis, in few cases, the conformations exhibiting the biological relevance were also analysed. In almost all the cases, the default set of parameters such a number of individuals in population (150), maximum number of generations (27000), maximum number of energy evaluations (2500000), rate of mutation (0.02), rate of crossover (0.8), etc. were used for performing molecular docking experiments.

# CHAPTER 3

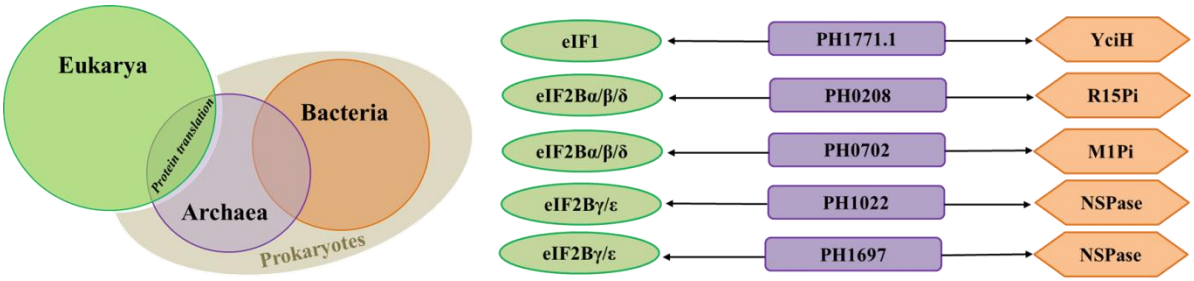
## *In silico* characterization of putative translation initiation factors in *Pyrococcus horikoshii* OT3

Parts of this chapter has been published as:

- Gogoi, P. and Kanaujia, S.P., 2018. Archaeal and eukaryal translation initiation factor 1 differ in their RNA interacting loops. *FEBS Letters*, *592*(9), 1602-1610.
- Gogoi, P., Srivastava, A., Jayaprakash, P., Jeyakanthan, J. and Kanaujia, S.P., 2016. *In silico* analysis suggests that PH0702 and PH0208 encode for methylthioribose-1-phosphate isomerase and ribose-1,5-bisphosphate isomerase, respectively, rather than aIF2Bβ and aIF2Bδ. *Gene*, *575*(1), 118-126.

# ABSTRACT

The overall process of protein biosynthesis across all domains of life is similar; however, detailed insights reveal a range of differences in the proteins involved. For decades, the process of protein translation in archaea has been considered to be closer to eukaryotes than to bacteria. In archaea, however, several homologs of eukaryotic translation initiation factors (IFs) such as eIF1 and eIF2B (α, β, γ, δ and ε subunits) have not yet been identified. Genomes of archaea such as *P. horikoshii* have been reported to consist of an open reading frame (ORF) PH1771.1 encoding a functional homolog of eIF1. Similarly, three ORFs PH0440, PH0702 and PH0208 in *P. horikoshii* display similarity to the regulatory subunits (α, β and δ) of eIF2B. Furthermore, two ORFs PH1022 and PH1697 exhibit similarity with the catalytic subunits (γ and ε) of eIF2B. However, all these ORFs also exhibit sequence similarity to another set of functionally unrelated proteins. For instance, PH1771.1 shows sequence similarity to a bacterial translation initiation inhibitor YciH. The crystal structure of PH0440 has already been elucidated and reported to be highly similar to the α-subunit of eIF2B. However, PH0208 and PH0702 exhibit similarity to the enzymes ribose-1,5-bisphoshate isomerase (R15Pi) and 5-methylthioribose 1-phosphate isomerase (M1Pi), respectively. Other two ORFs PH1022 and PH1697 are highly similar to NDP-sugar pyrophosphorylase (NSPase). Thus, to overcome this dilemma, *in silico* functional annotation of the five ORFs viz. PH1771.1, PH0208, PH0702, PH1022 and PH1697 has been performed. Preliminary *in silico* analysis suggests that PH1771.1 exhibit homology to both eIF1 and YciH owing to their evolutionary relatedness. In contrast, results indicate that PH0208 and PH0702 might encode the enzymes R15Pi and M1Pi, respectively, and PH1022 and PH1697 might encode NSPase instead of encoding the regulatory and catalytic subunits of eIF2B.




## 3.1 INTRODUCTION

Archaea, although being prokaryote, possess an unexpectedly large set of translation initiation factors (IFs) (Benelli and Londei, 2011). Furthermore, the archaeal translation IFs exhibit higher homology towards the eukaryal counterpart as compared to bacteria. With the advent of the genomic DNA sequences of hyperthermophilic archaea, a considerably substantial number of homologs of eukaryotic translation IFs has been identified. The availability of these translation IFs in archaea led towards a supposition that archaea imitate the eukaryal mechanism of translation initiation (Bult et al., 1996; Dennis, 1997; Bell and Jackson, 1998; Kakuta et al., 2004; Londei, 2005; Benelli and Londei, 2011).

In eukaryotes, the process of translation initiation commences with the binding of a ternary complex (eIF2•GTP•Met-tRNA$_i^{Met}$) composed of eIF2, GTP and initiator methionyl-tRNA to the 40S small ribosomal subunit along with several other factors such as eIF1, eIF1A, eIF3 and eIF5 giving rise to the formation of a 43S preinitiation complex (PIC) which scans along the mRNA from the 5' end to the 3' end until the recognition of a start codon (Lorsch and Dever, 2010). During the start codon selection, the protein eIF1 plays a central role by discriminating against non-AUG codons or AUG codons that have non-optimal context and also by promoting the dissociation of aberrantly assembled ribosomal complexes (Pestova et al., 1998; Pestova and Kolupaeva, 2002). Once an AUG is located, eIF2•GTP gets converted to an inactive state, eIF2•GDP. A nucleotide exchange factor, eIF2B is required for recycling eIF2•GDP to eIF2•GTP for the subsequent rounds of translation initiation (Cigan et al., 1993; Price and Proud, 1994). eIF2B comprises of two copies each of α, β, δ, γ and ε subunits where

α-, β- and δ-subunits form the regulatory subcomplex while γ and ε make up the catalytic subcomplex (Price and Proud, 1994).

Among the other translation IFs, eIF1 is considered to be universally conserved in all the three domains of life; however, the homolog of eIF1, known as YciH, is not present in all bacteria (Kyrpides and Woese, 1998a). Furthermore, recent studies have revealed that instead of functioning as a translation IF, YciH acts as an inhibitor of translation initiation during stress conditions (Osterman et al., 2015). The fidelity of translation initiation in bacteria is alternatively established by IF3; which, however, lacks sequence or structural homology with the protein eIF1 (Lomakin et al., 2006).

On the other hand, the homologs of only the regulatory subunits, and not the catalytic subunits, of eIF2B have been reported to be present in archaea (Dennis, 1997; Kyrpides and Woese, 1998b). However, the regulatory subunits alone would not confer full functionality to eIF2B, as the catalytic subunits accomplish the crucial function of GTP exchange. The regulatory subunits of eIF2B also share homology with two functionally non-related proteins ribose-1,5-bisphosphate isomerase (R15Pi) and 5-methylthioribose 1-phosphate isomerase (M1Pi). The γ and ε subunits share sequence similarity with each other as well with another group of functionally unrelated proteins known as NDP-sugar pyrophosphorylase (NSPase).

Thus, to identify the true function of these putative translation IFs in archaea, we took up the task of analyzing the available amino acid sequences and structures of these proteins using *in silico* approaches.

## 3.2 MATERIALS AND METHODS

The amino acid sequences of proteins used in this study were downloaded from UniProtKB database (UniProt Consortium, 2018). The homology searches were performed using the web tool BLAST (Altschul, 1990). Searches were also performed using PSI-BLAST (e-value cutoff: $10^{-5}$) to identify the distant homologs (Altschul et al., 1997). The three-dimensional atomic coordinates of the protein structures analyzed in the current study were downloaded from Protein Data Bank (Berman et al., 2000). The

multiple sequence alignment (MSA) of different sets of protein sequences were performed using the program Clustal Omega with default set of parameters viz. HHAlign algorithm and Gonnet transition matrix with gap open penalty of 6 bits and gap extension of 1 bit (Sievers and Higgins, 2014). The aligned sequences were further rendered using the online web server ESPript (Gouet et al., 2003). The phylogenetic trees were generated employing the neighbor joining (NJ) and maximum likelihood (ML) method in MEGA6 and MEGA7 (Tamura et al., 2013; Kumar et al., 2016). The reliability of the trees were evaluated by bootstrapping for 1000 replicates. The domain organization of proteins were identified using the Pfam database (El-Gebali et al., 2018). The electrostatic potentials of protein structures were generated using the APBS plugin (Baker et al., 2001) of the molecular visualization program PyMOL (The PyMOL Molecular Graphics System, Schrodinger, LLC). The volumes of the active site of proteins were computed using the program CASTp (Binkowski et al., 2003).

## 3.3 RESULTS

### 3.3.1 *IN SILICO* CHARACTERIZATION OF HOMOLOGS OF EUKARYOTIC TRANSLATION INITIATION FACTORS IN ARCHAEA

A homology search using the amino acid sequences of eukaryotic translation IFs eIF1 and eIF2B, in hyperthermophilic archaeon, *P. horikoshii* OT3 was performed to identify the presence of their archaeal homologs. An ORF PH1771.1 shows highest sequence similarity to eIF1 while four ORFs (PH0208, PH0702, PH1022 and PH1697) show similarity to the different subunits of eIF2B. PH0208 and PH0702 are homologous to the regulatory subunits (α, β and δ) of eIF2B while PH1022 and PH1697 exhibit homology towards the catalytic subunits (γ and ε) of eIF2B. Intriguingly, the identified ORFs also exhibit sequence similarity to another set of functionally unrelated proteins. PH1771.1 shows similarity to a translation initiation inhibitor YciH apart from eIF1. Similarly, PH0702 and PH0208 show similarity to enzymes M1Pi and R15Pi, respectively while PH1022 and PH1697 show similarity to NSPases (Table 3.1).

**Table 3.1.** Homologs of probable translation initiation factors in archaea.

| ORF id | Homologous proteins | | | |
|---|---|---|---|---|
| | Protein name | Query coverage, Sequence Identity (%) | Protein name | Query coverage, Sequence Identity (%) |
| PH1771.1 | eIF1 | 71, 33 | YciH | 70, 42 |
| PH0208 | eIF2Bα | 86, 25 | R15Pi | 98, 86 |
| | eIF2Bβ | 62, 31 | | |
| | eIF2Bδ | 80, 29 | | |
| PH0702 | eIF2Bα | 45, 28 | M1Pi | 93, 55 |
| | eIF2Bβ | 40, 30 | | |
| | eIF2Bδ | 39, 24 | | |
| PH1022 | eIF2Bγ | 83, 22 | NSPase | 88, 25 |
| | eIF2Bε | 80, 22 | | |
| PH1697 | eIF2Bγ | 95, 23 | | 98, 29 |
| | eIF2Bε | 92, 19 | | |

### 3.3.2 *IN SILICO* CHARACTERIZATION OF PH1771.1

#### 3.3.2.1 EVOLUTIONARY RELATIONSHIP

The ORF PH1771.1 shares sequence similarity with eukaryotic translation IF eIF1 as well as with bacterial translation initiation inhibitor YciH. A sequence-based rooted phylogeny tree including amino acid sequences of archaeal initiation factor 1 (aIF1) from archaea, eIF1 from eukaryotes and YciH from bacteria was built to understand the evolutionary relationship of PH1771.1 with these proteins. To trace the root of the tree, the functionally homologous IF3, which does not share sequence similarity to any of these proteins was also included during the tree building. Results reveal that PH1771.1

along with the other aIF1 forms a distinct clade (Figure 3.1). The three clades belonging to aIF1, YciH and eIF1, however, share a common ancestor with a closer evolutionary relationship between aIF1 and YciH (Figure 3.1). Furthermore, the phylogram also depicts that aIF1, in comparison to eIF1 and YciH, has undergone the least amount of changes during the course of evolution and thus is closest to the ancestral protein (Figure 3.1).

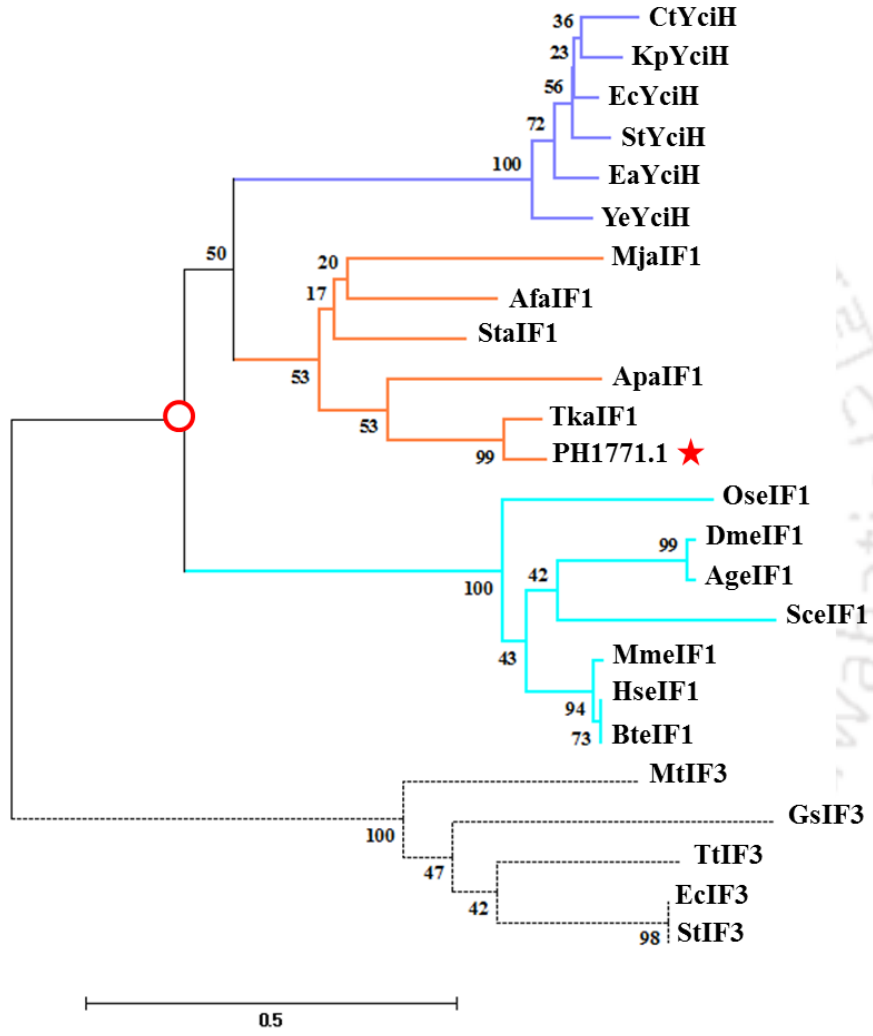



**Figure 3.1. Evolutionary relationships among eIF1, YciH and aIF1 from eukaryotes, bacteria and archaea respectively**. The three distinct clades formed by eukaryotes, bacteria and archaea are highlighted in cyan, purple and pink, respectively, while the outgroup is shown in dotted lines. The protein aIF1 encoded by PH1771.1 from *P. horikoshii* is marked with red star. The node representing the common ancestor is highlighted with red circle. Bootstrap values are shown at the branch points. The scale bar indicates 0.5 amino acid substitutions per single site. The name of the species (UniProt id) used to generate the phylogeny tree are (1) for YciH: CtYciH, *Cronobacter turicensis* (C9XTM4); KpYciH, *Klebsiella pneumoniae* (A0A0W8APJ4); EcYciH, *E.*

*coli* (P08245); StYciH, *Salmonella typhimurium* (P0A247); EaYciH, *Erwinia amylovora* (D1I3C6) and YeYciH, *Yersinia enterocolitica* (A0A2J9S9M9), (2) for aIF1: MjaIF1, *M. jannaschii* (Q57902); AfaIF1, *A. fulgidus* (O29348); StaIF1, *Sulfolobus tokodai* (Q975S0); ApaIF1, *Aeropyrum pernix* (Q9YBG9); TkaIF1, *Thermococcus kodarensis* (Q5JDH7) and PH1771.1, *P. horikoshii* (P58193) and (3) for eIF1: OseIF1, *Oryza sativa* (A6MZM2); DmeIF1, *Drosophila melanogaster* (Q9VEA1); AgeIF1, *Anopheles gambiae* (P42678); SceIF1, *Saccharomyces cerevisiae* (P32911); MmeIF1, *Mus musculus* (P48024); HseIF1, *Homo sapiens* (P41567); BseIF1, *Bos taurus* (Q5E938) (4) for IF3: MtIF3, *Mycobacterium tuberculosis* (P9WKJ9); GsIF3, *Geobacillus stearothermophilus* (P03000); TteIF1, *Thermus thermophilus* (Q5SKU2); EcIF3, *E. coli* (P0A707) and StIF3, *S. typhimurium* (P33321).

### 3.3.3 *IN SILICO* CHARACTERIZATION OF PH0208 AND PH0702

#### 3.3.3.1 EVOLUTIONARY RELATIONSHIP

The amino acid sequences of PH0702 and PH0208 share sequence similarity with M1Pi and R15Pi, respectively, as well as with the regulatory subunits of eIF2B (Table 3.1). Thus, to trace the evolutionary relationship of PH0702 and PH0208 with M1Pi, R15Pi as well as the regulatory subunits of eIF2B, a phylogeny analysis was performed. The phylogenetic tree clearly reveals that PH0702 belongs to the M1Pi cluster of proteins while PH0208 clusters with R15Pi proteins instead of the regulatory subunits of eIF2B (Figure 3.2).

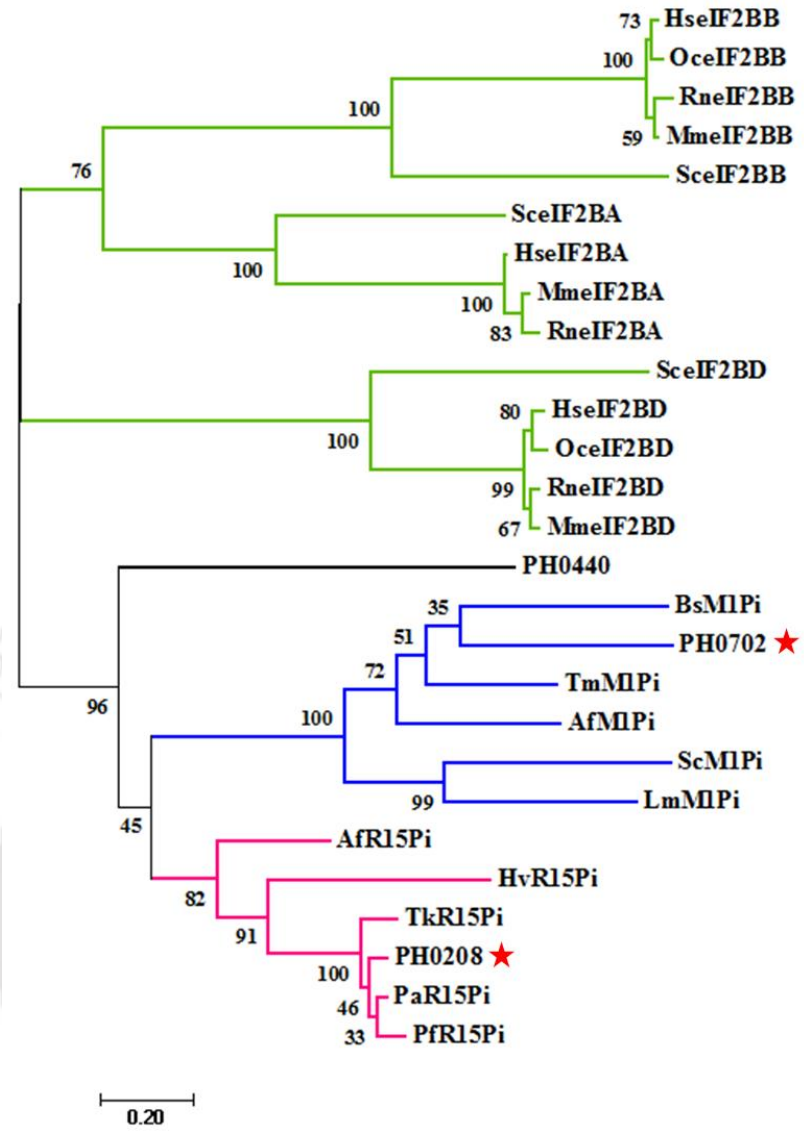



**Figure 3.2. Phylogenetic tree showing relationship among M1Pi, R15Pi and eIF2B regulatory subunits.** The phylogenetic tree was generated by performing MSA using the amino acid sequences of M1Pi, R15Pi and the regulatory subunits of eIF2B. The clusters belonging to eIF2B regulatory subunits (α, β and δ), M1Pi and R15Pi are highlighted by green, blue and pink, respectively. The protein sequences of PH0702 and PH0208 are marked with red stars. The name of the species (UniProt id) used to generate the phylogeny are (1) for M1Pi: BsM1Pi, *Bacillus subtilis* (O31662); ScM1Pi, *S. cerevisiae* (Q06489); TmM1Pi, *Thermotoga maritima* (Q9X013); AfM1Pi, A. *fulgidus* (O29877) and LmM1Pi, *Leishmania major* (Q4Q0R9), (2) for R15Pi: TkR15Pi, *T. kodakaraensis* (Q5JFM9); PaR15Pi, *Pyrococcus abyssi* (Q9V281); PfR15Pi, *Pyrococcus furiosus* (Q8U4G6); AfR15Pi, *A. fulgidus* (O28242) and HvR15Pi, *Haloferax volcanii* (D4GV73), (3) for eIF2Bα: HseIF2BA, *H. sapiens* (Q14232); SceIF2BA, *S. cerevisiae* (P14741); MmeIF2BA, *M. musculus* (Q99LC8) and RneIF2BA, *Rattus norvegicus* (Q64270), (4) for eIF2Bβ: HseIF2BB, *H. sapiens* (P49770); SceIF2BB, *S. cerevisiae* (P32502); MmeIF2B, *M. musculus* (Q99LD9); RneIF2BB, *R. norvegicus* (Q62818) and OceIF2BB, *Oryctolagus cuniculus* (Q28690), (5) for eIF2Bδ: HseIF2BD, *H. sapiens* (Q9UI10); SceIF2BD, *S. cerevisiae* (P12754); MmeIF2BD, *M. musculus* (Q61749);

RneIF2BD, *R. norvegicus* (Q63186) and OceIF2BD, *O. cuniculus* (P41111) and for PH0440, PH0702 and PH0208, *P. horikoshii* (O58185, O58433 and O57947, respectively). The α-, β- and δ- subunits of eIF2B are denoted by A, B and D, respectively in the figure.

### 3.3.3.2 COMPARISON OF THE TERTIARY STRUCTURES

As the protein sequences of eIF2B regulatory subunits, M1Pi and R15Pi share a significant level of sequence similarity among themselves, we compared their structures to find out the differences. Since eIF2Bα, eIF2Bβ and eIF2Bδ are homologous, we have used the tertiary structure of eIF2Bα as a prototype for the other two subunits (β and δ). A comparison of the crystal structures of eIF2Bα, M1Pi and R15Pi reveals a considerable degree of similarity among themselves with an average RMSD of 3.7 Å (Figure 3.3A, B). To further unriddle the differences among eIF2Bα, M1Pi and R15Pi, we compared the volumes of the active site of these proteins and found that the active-site pocket of the three proteins are also comparable to each other (eIF2Bα: 777.9 $Å^3$, M1Pi: 817.9 $Å^3$ and R15Pi: 718.5 $Å^3$) (Figure 3.3C). Furthermore, we tried to probe the composition of charged residues at the active site of the proteins eIF2Bα, M1Pi and R15Pi by computing the electrostatic potential charge distribution. Results reveal that eIF2Bα and R15Pi has comparatively more positively charged residues at the active-site pocket than M1Pi (Figure 3.3D).

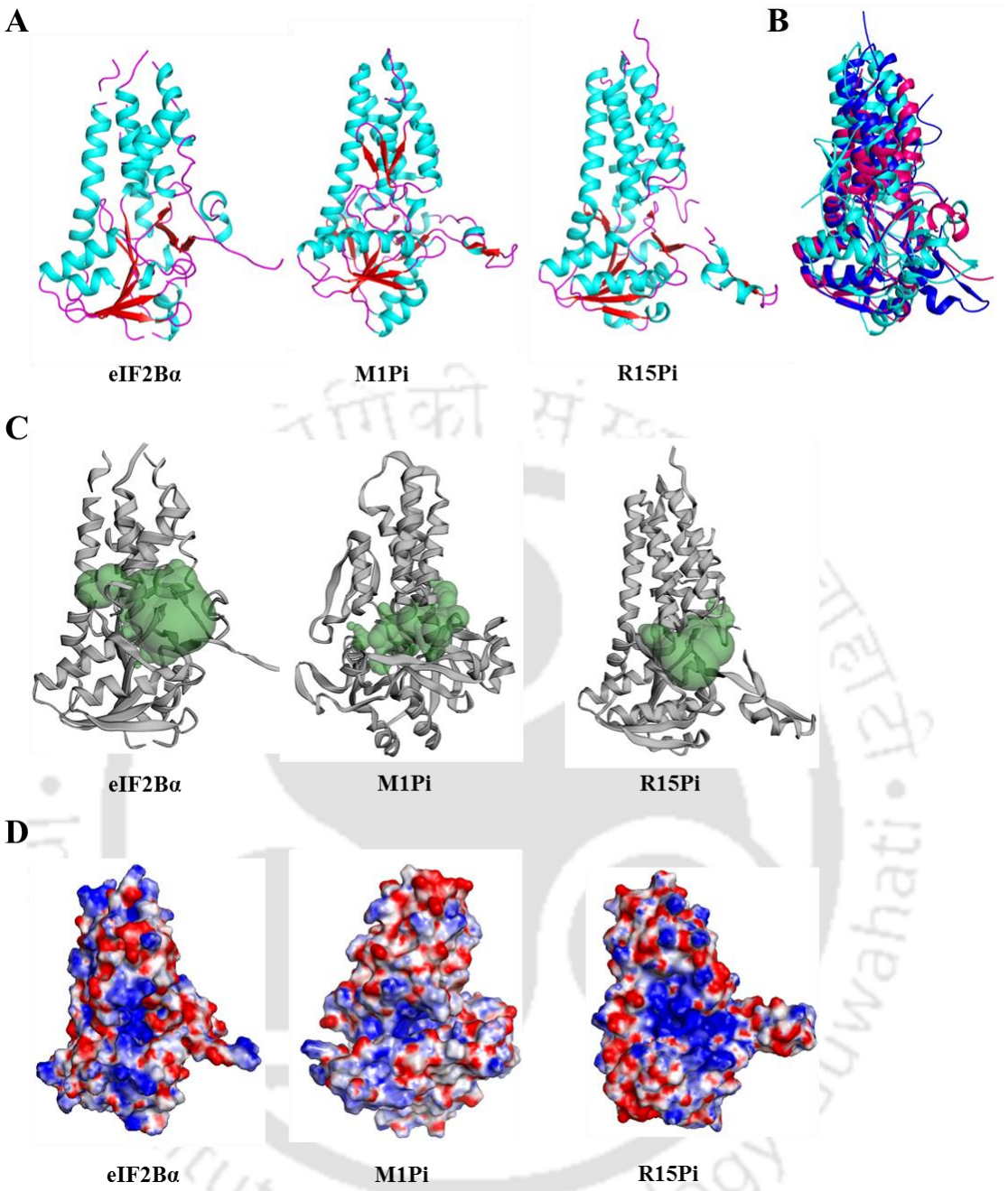



**Figure 3.3. Structural (dis)similarity of R15Pi, M1Pi and eIF2Bα.** (A) Comparison of the tertiary structures of R15Pi (PDB id: 3A11), M1Pi (PDB id: 2YVK) and eIF2Bα (PDB id: 3ECS). (B) Superimposition of R15Pi (blue), M1Pi (cyan) and eIF2Bα (pink). (C) Comparison of active-site volume of R15Pi, M1Pi and eIF2Bα. The active-site pockets of the proteins are shown as green surface. (D) Electrostatic potential charge distribution at the active-site pocket of R15Pi, M1Pi and eIF2Bα. The map was drawn on a solvent-accessible molecular surface (red, − 10 kT/e to blue, + 10 kT/e) and calculated with a probe radius of 1.4 Å.

### 3.3.3.3 REPERTOIRE OF RESDIUES AT THE ACTIVE SITE

To further differentiate among the three groups of proteins, the amino acid composition of the active-site pockets of M1Pi and R15Pi were analyzed and compared with the regulatory subunits of eIF2B. Firstly, we listed out the residues involved in the substrate and/or product stabilization in the active site of the M1Pi protein. The crystal structure of M1Pi from *Bacillus subtilis* (BsM1Pi) reveals that the residues Arg51, Gly52, Ala53, Arg94, Gln199, Ala239, Asp240, Asn250 and Lys251 anchor the polar (phosphate and ribulose) groups of the substrate (MTR-1-P)/product (MTRu-1-P) (Figure 3.4A). In addition, the residues Pro54, Ala162, Ala166 and Thr167 lend hydrophobic interactions to stabilize the non-polar (methylthio) group of the substrate/product (Figure 3.4A). Furthermore, the residues Cys160 and Asp240 are known to play a critical role in the catalytic activity of the enzyme (Tamura et al., 2008). To investigate whether the protein sequence of PH0702 contains the residues crucial for the M1Pi activity, we performed the MSA of M1Pi protein sequences from *B. subtilis*, *Thermotoga maritima*, *A. fulgidus*, *Saccharomyces cerevisiae*, *Leishmania major* and the protein sequence of PH0702 from *P. horikoshii*. The result shows that the residues of the enzyme M1Pi, which interact with different groups of the substrate, were found to be absolutely conserved in PH0702 also (Figure 3.4B). Furthermore, it is to be noted that the residues essential for the hydrophobic (methylthio group) stabilization of the substrate/product in the active site of the enzyme are conserved as well (Figure 3.4B). On the other hand, these residues are absent in the protein sequence of the regulatory subunits (α, β and δ) of eIF2B (Figure 3.4B). This indicates that the protein PH0702 function as the enzyme M1Pi rather than the regulatory subunits of eIF2B.

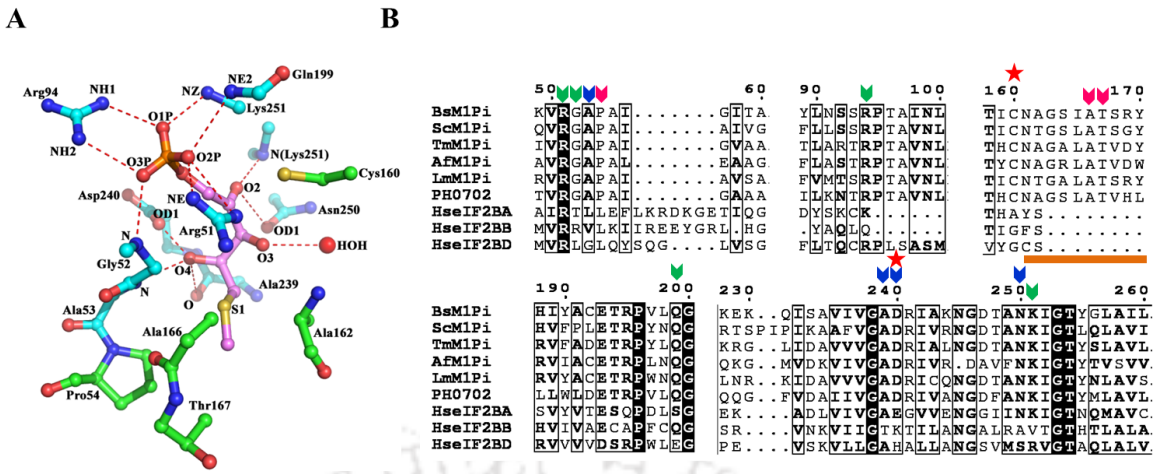



**Figure 3.4. Conserved active-site residues of M1Pi.** (A) The active-site residues of M1Pi bound with the product MTRu-1-P (PDB id: 2YVK). All the amino acids are shown as ball-and-stick model and the water molecules are shown as red sphere. The product (MTRu-1-P) is shown with carbons in pink. The atoms of the product involved in the protein interactions are labeled. The amino acids with carbon atoms in cyan are directly involved in the interaction with the polar (phosphate and ribulose) groups of the product. The residues shown with carbon atoms in green stabilize the hydrophobic (methylthio) group of the substrate and/or product. In many of the cases, only the interacting atoms of the residues are displayed here for the clarity of the figure. (B) Multiple sequence alignment (MSA) of protein sequences of M1Pi from *B. subtilis* (O31662), *S. cerevisiae* (Q06489), *T. maritima* (Q9X013), *A. fulgidus* (O29877), *L. major* (Q4Q0R9), PH0702 from *P. horikoshii* (O58433) and α, β and δ subunit of eIF2B from *H. sapiens* (Q14232, P49770, Q9UI10). The α-, β- and δ- subunits of eIF2B are denoted by A, B and D, respectively, in the figure. The UniProt id of each protein sequences are provided in parenthesis. Only the partial MSA has been shown for clarity of the figure. The catalytic residues are highlighted with red stars. The residues marked with green downward arrowheads are involved in the interaction with the phosphate group of the substrate/product. Those labeled with blue downward arrowheads depict the residues interacting with the ribulose group. The residues involved in the hydrophobic stabilization are differentiated with pink downward arrowheads and are highlighted with an orange bar.

Similarly, the residues essential for the catalytic activity of the enzyme R15Pi were identified by analyzing the available crystal structures bound to substrate and product as well as the protein sequences from various organisms. The structural analysis of R15Pi from *Thermococcus kodakarensis* (TkR15Pi) reveals that the residues Arg20, Gly21, Arg63, Ser135, Lys136, Ala137, Asn212, Lys213, Lys238 and Arg254 are involved in stabilizing either substrate or product or both in the active site of the enzyme R15Pi (Figure 3.5A). In addition, it is known that the residues Cys133 and Asp202 are crucial to initiate the catalytic activity of the enzyme (Nakamura et al., 2012). To further

examine whether the protein sequence of PH0208 contains the residues crucial for the R15Pi activity, we performed the MSA of various protein sequences of R15Pi from *T. kodakaraensis*, *Pyrococcus abyssi*, *Pyrococcus furiosus*, *A. fulgidus* and *Haloferax volcanii* and the protein sequence of PH0208 from *P. horikoshii*. All these residues show high conservation at the level of primary structure of the enzyme R15Pi from various organisms (Figure 3.5B). The analysis confirms the conservation of the residues in the protein sequence of PH0208 required for the enzymatic activity of R15Pi (Figure 3.5B). Moreover, the protein PH0208 harbors positively charged active-site residues, as in the enzyme R15Pi, required to coordinate with the two negatively charged phosphate groups of the substrate/product (Figure 3.5B). Thus, it can be assumed that the protein PH0208 performs the function of the enzyme R15Pi instead of the regulatory subunits of eIF2B.

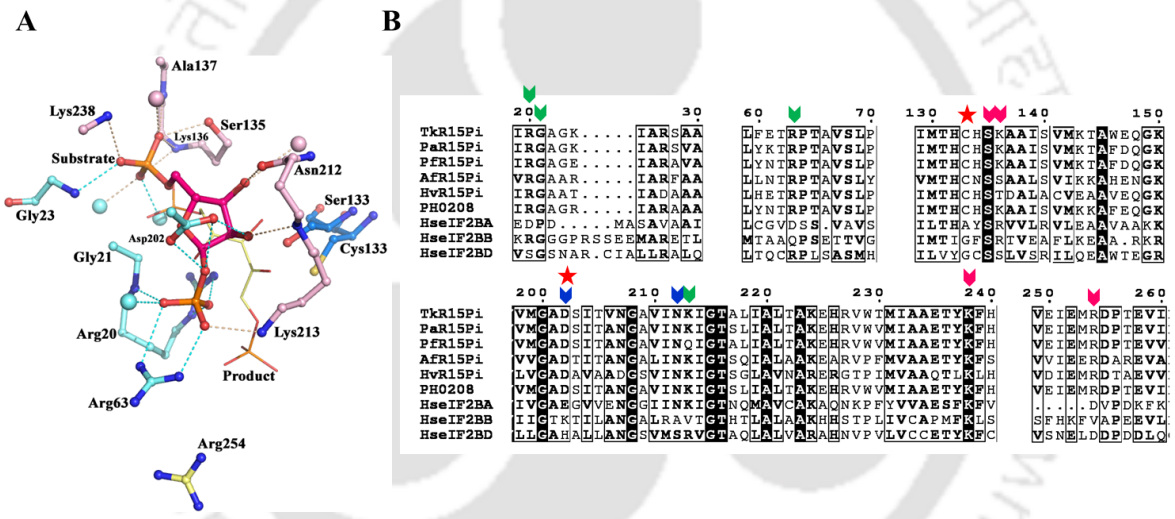



**Figure 3.5. Conserved active-site residues of R15Pi.** (A) The active-site residues of R15Pi bound to the substrate (R15P) and the product (RuBP) (PDB ids: 3VM6 and 3A9C). All the residues are shown as ball-and-stick model and water molecules are shown as cyan sphere. The substrate (R15P) is shown with carbon atoms in red. The product (RuBP) is shown with carbon atoms in yellow. The amino acids with carbon atoms in cyan are involved in interaction with various (phosphate and sugar) groups of only the substrate. The amino acids with carbon atoms in pink color are involved in interaction with various polar groups of both the substrate as well as the product. The residues shown in yellow carbon atoms coordinate only with the product. In many of the cases, only the interacting atoms of the residues are displayed here for the clarity of the figure. (B) Multiple sequence alignment (MSA) of protein sequences of R15Pi from *T. kodakarensis* (Q5JFM9), *P. abyssi* (Q9V281), *P. furiosus* (Q8U4G6), *A. fulgidus* (O28242), *H. volcanii* (D4GV73), PH0208 from *P. horikoshii* (O57947) and α, β and δ subunit of eIF2B from *H. sapiens* (Q14232, P49770, Q9UI10). The α-, β- and δ- subunits of eIF2B are denoted by A, B and D, respectively, in the figure. The UniProt id of each protein sequences are provided in parenthesis. Only the partial MSA has been shown for the clarity of the figure. The catalytic residues are highlighted with red stars. The

residues marked with green downward arrowheads are involved in the interaction with the 1-phosphate group of the substrate/product. Those labeled with blue downward arrowheads depict the residues interacting with the ribose or ribulose group of the substrate or the product, respectively. The amino acids marked with pink downward arrowheads draws the residues interacting with the 5-phosphate group of the substrate/product.

### 3.3.4 *IN SILICO* CHARACTERIZATION OF PH1022 AND PH1697

As the amino acid sequences of PH1022 and PH1697show significant sequence similarity to two different group of proteins which includes the catalytic subunits (γ and ε) of eIF2B as well as NSPases, a phylogeny analysis of PH1022, PH1697, catalytic subunits of eIF2B along with various NSPases was performed. The phylogenetic tree clearly depicts that both PH1022 and PH1697 cluster together with NSPases and not with the catalytic subunits of eIF2B (Figure 3.6A). To investigate the presence of signature motifs of NSPases in PH1022 and PH1697, MSA of PH1022, PH1697 along with various NSPases was performed. The signature motif of NSPase, G-X-G-T-R-$X_{\{5,10\}}$-K is found to be present in both PH1022 and PH1697 (Figure 3.6B). The presence of these signature motifs evidently suggests that PH1022 and PH1697 might function as NSPases rather than as the catalytic subunits of eIF2B.

In the process of assigning the cognate ligand(s) of PH1022 and PH1697, a total of six (PH0365, PH0380, PH0413, PH0925, PH1219 and PH1925) other NSPases were identified in *P. horikoshii* (hereafter referred to as PhNSPase). Except, PH0365 and PH1219, all the NSPases exhibit significant sequence similarity to one another (Figure 3.7).

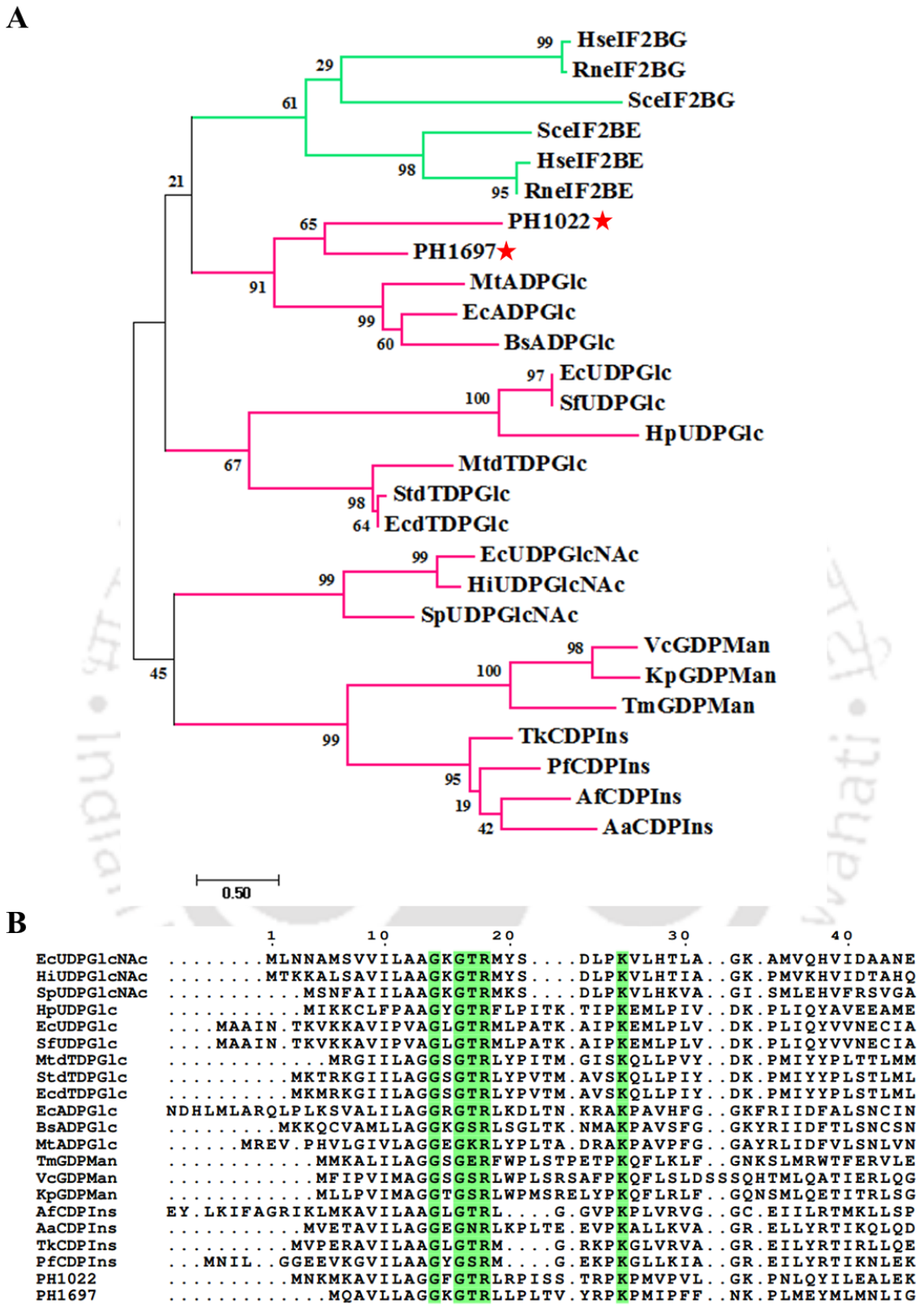



**Figure 3.6. Evolutionary relationship of NSPases and catalytic subunits of eIF2B.** (A) Phylogenetic tree showing relationship among NSPases, eIF2B catalytic subunits, PH1022 and PH1697. (B) Multiple sequence alignment (MSA) of NSPases, PH1022 and PH1697. The NSPases used to build the phylogenetic tree and MSA are UDPGlcNAc

pyrophosphorylase from *E. coli* (EcUDPGlcNAc, P0ACC7), *Haemophilus influenzae* (HiUDPGlcNAc, P43889) and *Streptococcus pneumonia* (SpUDPGlcNAc, Q97R46); ADPGlc pyrophosphorylase from *B. subtilis* (BsADPGlc, P39122), *M. tuberculosis* (MtADPGlc, P9WN43) and *E. coli* (EcADPGlc, P0A6V1); UDPGlc pyrophosphorylase from *E. coli* (EcUDPGlc, P0AEP3), *Shigella flexneri* (SfUDPGlc, P0AEP6) and *Helicobacter pylori* (HpUDPGlc, D0ISI7); dTDPGlc pyrophosphorylase from *M. tuberculosis* (MtdTDPGlc, P9WH13), *Salmonella typhimurium* (StdTDPGlc, P26393) and *E. coli* (EcdTDPGlc, P37744); CDPIns pyrophosphorylase from *Aquifex aeolicus* (AaCDPIns, O67379), *A. fulgidus* (AfCDPIns, O29976), *T. kodakarensis* (TkCDPIns, Q5JDA9) and *P. furiosus* (PfCDPIns, Q8U1Z6); GDPMan pyrophosphorylase from *T. maritima* (TmGDPMan, Q9X0C3), *Vibrio cholera* (VcGDPMan, Q07024) and *Klebsiella pneumoniae* (KpGDPMan, Q48462); PH1022 (O58775) and PH1697 (O59364) from *P. horikoshii*; ε subunit of eIF2B from *H. sapiens* (Q13144), *R. norvegicus* (Q64350) and *S. cerevisiae* (P32501); γ subunit of eIF2B from *H. sapiens* (Q9NR50), *R. norvegicus* (P70541) and *S. cerevisiae* (P09032). The γ- and ε- subunits of eIF2B are denoted by G and E, respectively, in the figure.

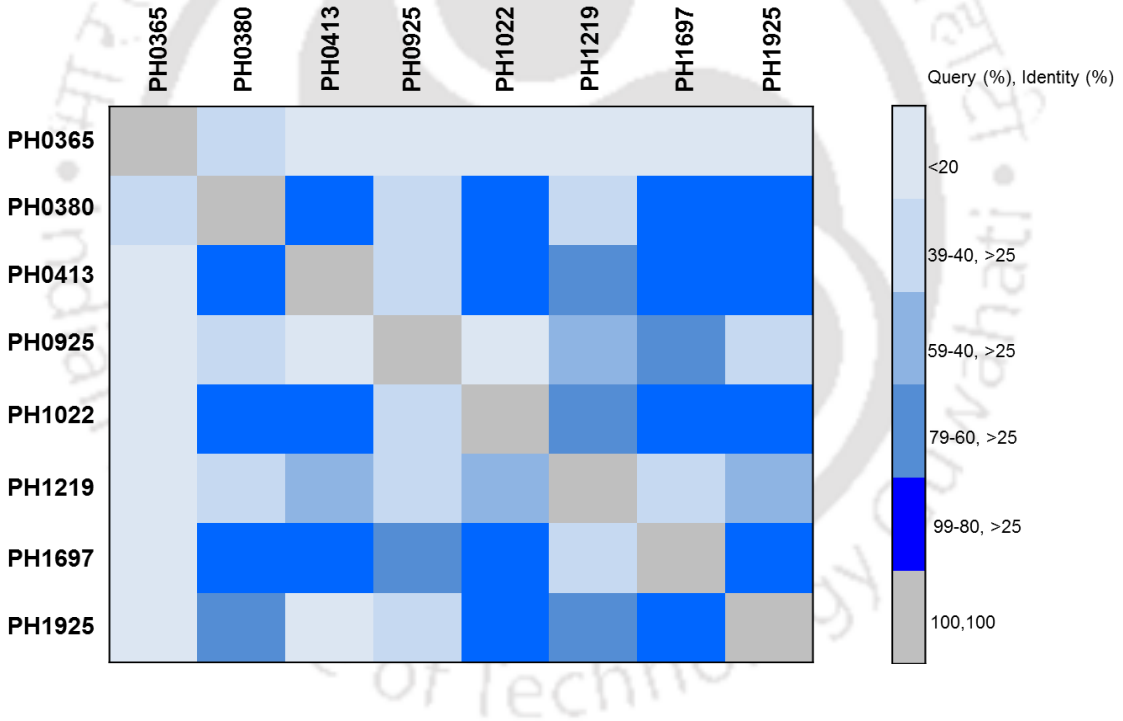



**Figure 3.7. Heat map illustrating the pairwise sequence similarity of all the identified NSPases from *P. horikoshii*.** Protein sequences with increasing similarity are depicted in an increasing gradient of blue.

To identify the closest homologs of the identified PhNSPases (PH0365, PH0925, PH1219 and PH1925) among prokaryotes, a homology search was performed, which produced a single homologous protein having a sequence identity (query coverage) in the

range of 55-27% (100-95%) (Table 3.2). On the other hand, the remaining four PhNSPases (PH0380, PH0413, PH1022 and PH1697) exhibited sequence homology to at least two types of NSPases with sequence identity (query coverage) in the range of 50-25% (99-88%) (Table 3.2).

**Table 3.2.** List of homologs of NSPases from *P. horikoshii* OT3. Abbreviations: QC, query coverage; SI, sequence identity.

| ORF id | First closest homolog | | Second closest homolog | |
|---|---|---|---|---|
| | Protein name | QC, SI (%) | Protein name | QC, SI (%) |
| PH0365 | Galactose 1-phosphate uridylyltransferase (UDP-galactose pyrophosphorylase ) | 95, 40 | - | - |
| PH0380 | Glucose 1-phosphate thymidylyltransferase (dTDP-glucose pyrophosphorylase) | 99, 28 | Glucose 1-phosphate adenylyltransferase (ADP-glucose pyrophosphorylase) | 99, 27 |
| PH0413 | Glucose 1-phosphate thymidylyltransferase (dTDP-glucose pyrophosphorylase) | 98, 50 | Glucose 1-phosphate adenylyltransferase (ADP-glucose pyrophosphorylase) | 99, 26 |
| PH0925 | Mannose 1-phosphate guanylyltransferase (GDP-mannose pyrophosphorylase) | 100, 43 | - | - |
| PH1022 | Glucose 1-phosphate adenylyltransferase (ADP-glucose pyrophosphorylase) | 88, 25 | N-acetylglucosamine 1-phosphate uridyltransferase (UDP-N-acetylglucoseamine pyrophosphorylase) | 87, 26 |
| PH1219 | 1L-myo-inositol 1-phosphate cytidylyltransferase (CDP- | 97, 55 | - | - |

| | | | | |
|---|---|---|---|---|
| | inositol pyrophosphorylase) | | | |
| PH1697 | Glucose 1-phosphate uridylyltransferase (UDP-glucose pyrophosphorylase) | 98, 39 | Glucose 1-phosphate adenylyltransferase (ADP-glucose pyrophosphorylase) | 98, 29 |
| PH1925 | N-acetylglucosamine 1-phosphate uridyltransferase (UDP-N-acetylglucoseamine pyrophosphorylase) | 100, 27 | - | - |

### 3.3.4.1 PhNSPases BELONGING TO THE PROMISCUOUS SUBGROUP IC

Homology search indicates that PH0365, PH0925 and PH1219 might function as UDP-galactose (UDPGal) pyrophosphorylase, GDP-mannose (GDPMan) pyrophosphorylase and CDP-inositol (CDPIns) pyrophosphorylase, respectively. However, no clear conclusion could be obtained for the remaining five PhNSPases (PH0380, PH0413, PH1022, PH1697 and PH1925). Thus, to identify whether these PhNSPases exhibit promiscuity, a structure-based sequence alignment was performed to categorize the identified PhNSPases based on the conservation of the amino acid residues at the two signature motifs I and II. The structure-based sequence alignment reveals that four of the PhNSPases (PH0380, PH0413, PH1022 and PH1925) possess ‘G-D-X’ and ‘D-X-G’ (where X denotes any amino acid) as motifs I and II, respectively, indicating that they belong to the subgroup IC (Figure 3.8). On the other hand, although PH0925 and PH1697 possess the motif II (D-X-G), they do not conserve the motif I (G-D-X). This kind of discrepancy, however, is also observed in other members of the subgroup IC such as GDPMan pyrophosphorylase from *T. maritima* (TmGDPMan, PDB id: 2X65) (Figure 3.8). Thus, six of the PhNSPases can be categorized into the subgroup IC, which contains NSPases, highly promiscuous in nature. From the preliminary analysis, it was obtained that PH0925 might be a probable GDPMan pyrophosphorylase, however, placement of PH0925 in the subgroup IC indicates it to be promiscuous in nature. The motif I of PH1219 has a serine residue instead of glycine while its motif II is substituted by D-X-D placing it in the subgroup IB, members of which do not exhibit promiscuity.

PH0365, however, does not possess either of these conserved motifs precluding its placement in any of the known subgroups (Figure 3.8).

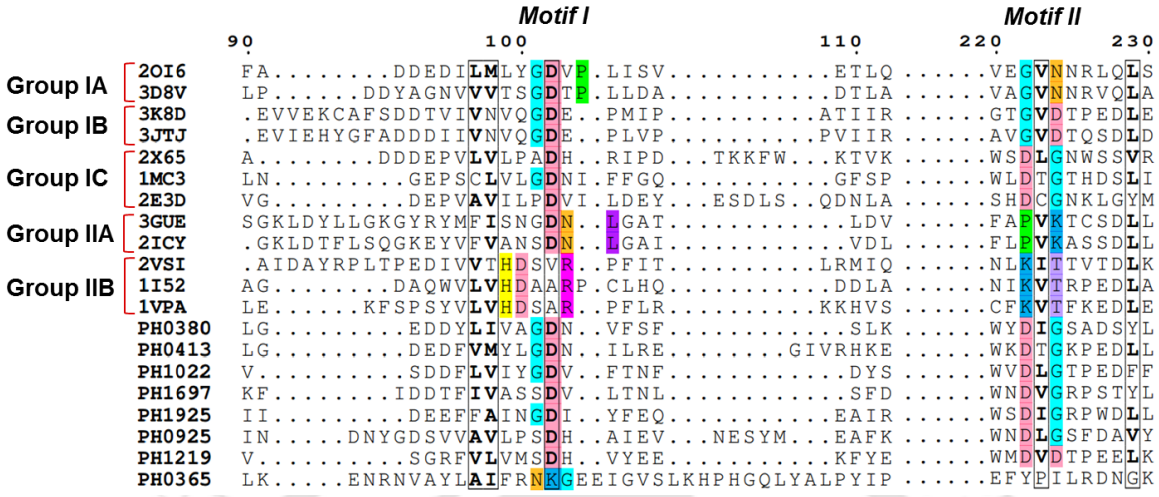



**Figure 3.8.** Structure-based sequence alignment of PhNSPases along with representative NSPases from each group classified on the basis of their metal coordinating motifs. Conservation at motif I and II are highlighted by color-coded amino acid residues. The structures used for the analysis are denoted by their PDB ids.

### 3.3.4.2 EVOLUTIONARY RELATIONSHIP

An evolutionary analysis of all the PhNSPases was performed to trace their relationship with their homologous NSPases. The evolutionary tree clearly depicts the formation of four major clades viz. I-IV (Figure 3.9). PH0365, PH0925 and PH1219 cluster with clade IV, III and II which contains UDPGal, GDPMan and CDPIns pyrophosphorylases, respectively. On the other hand, the remaining PhNSPases, which are also supposed to be promiscuous, cluster with clade I comprising of four types of NSPases, namely UDP-N-acetylglucoseamine (UDPGlcNAc), ADP-glucose (ADPGlc), UDP-glucose (UDPGlc), dTDP-glucose (dTDPGlc) pyrophosphorylases. In clade I, PH0380, PH1022 and PH1697 are evolutionarily closest to UDPGlcNAc pyrophosphorylase while both PH0413 and PH1925 are related to dTDPGlc pyrophosphorylase. Notably, the presence of PH0380, PH1022 and PH1697 as well as PH0413 and PH1925 to exist in the same cluster indicates their origin from a common ancestor through a duplication event (Figure 3.9).

An analysis of the domain organization indicates that all the PhNSPases, except PH0365, consist of a nucleotidyltransferase domain as the N-terminal domain (NTD). PH0365, instead of nucleotidyltransferase domain, consists of two galactose-1-phosphate uridylyltransferase (GALT) domains at the N- and C-termini. On the other hand, apart from containing the nucleotidyltransferase domain, PH0925 and PH1219 contain an additional phosphomannose isomerase (PMI) and CDP-alcohol phosphatidyltransferase (DIPPS) domain, respectively, at the C-terminal. Interestingly, the PhNSPases which contain an additional domain are clustered as distinct clades (II and III) in the phylogeny tree. The PhNSPases belonging to clade I also contain hexapeptide repeats at the C-terminal domain (CTD). These hexapeptide repeats are often found in the bifunctional UDPGlcNAc pyrophosphorylase (or GlmU) and are responsible for the CoA-dependent acetylation of glucosamine 1-phosphate (GlcN 1-P). However, out of the six PhNSPases, only PH1925 contains three hexapeptide repeats similar to UDPGlcNAc pyrophosphorylase (Figure 3.9).

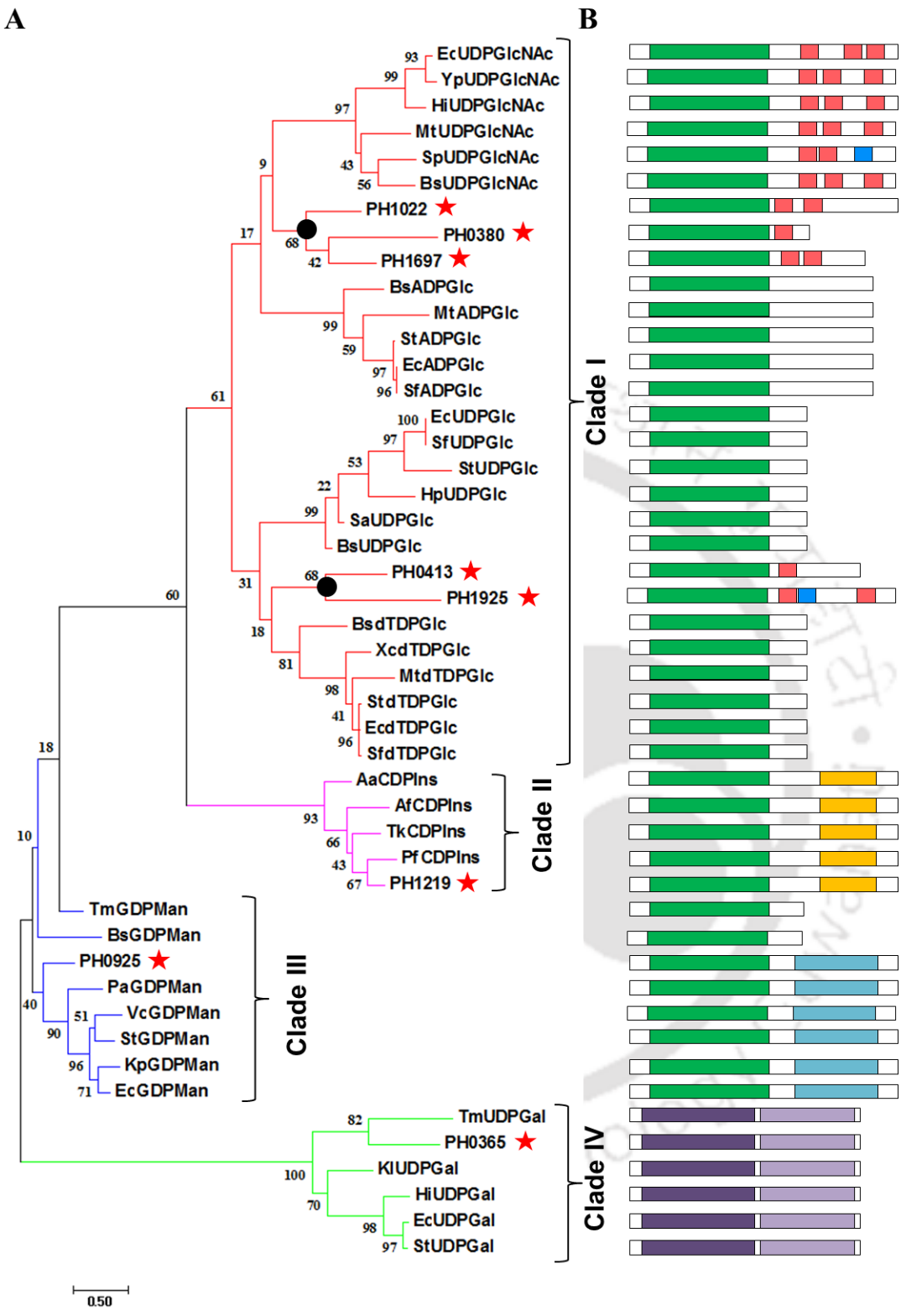



**Figure 3.9. Evolutionary relationship and domain organization of NSPases.** (A) Phylogeny tree depicting the formation of a monophyletic clade by the promiscuous PhNSPases (Clade I, red) while the specific PhNSPases form clear distinct clades (Clade II, magenta; Clade III, blue and Clade IV, green). The PhNSPases are highlighted with red star and the common ancestors of promiscuous PhNSPases are highlighted with a

black dot. Bootstrap values are indicated at the internal nodes. The scale bar indicates 0.5 amino-acid substitutions per single site. The NSPases used to build the phylogenetic tree are UDPGlcNAc pyrophosphorylase from *E. coli* (EcUDPGlcNAc, P0ACC7), *Yersinia pestis* (YpUDPGlcNAc, Q8Z957), *H. influenzae* (HiUDPGlcNAc, P43889), *M. tuberculosis* (MtUDPGlcNAc, P9WMN3), *S. pneumonia* (SpUDPGlcNAc, Q97R46) and *B. subtilis* (BsUDPGlcNAc, P14192); ADPGlc pyrophosphorylase from *B. subtilis* (BsADPGlc, P39122), *M. tuberculosis* (MtADPGlc, P9WN43), *S. typhimurium* (StADPGlc, P05415), *E. coli* (EcADPGlc, P0A6V1) and *S. flexneri* (SfADPGlc, P0A6V4); UDPGlc pyrophosphorylase from *E. coli* (EcUDPGlc, P0AEP3), *S. flexneri* (SfUDPGlc, P0AEP6), *S. typhimurium* (StUDPGlc, P0A2K7), *H. pylori* (HpUDPGlc, D0ISI7), *Staphylococcus aureus* (SaUDPGlc, Q2FE05) and *B. subtilis* (BsUDPGlc, Q05852); dTDPGlc pyrophosphorylase from *B. subtilis* (BsdTDPGlc, P39629), *Xanthomonas campestris* (XcdTDPGlc, P0C7J4), *M. tuberculosis* (MtdTDPGlc, P9WH13), *S. typhimurium* (StdTDPGlc, P26393), *E. coli* (EcdTDPGlc, P37744) and *S. flexneri* (SfdTDPGlc, P37779); CDPIns pyrophosphorylase from *A. aeolicus* (AaCDPIns, O67379), *A. fulgidus* (AfCDPIns, O29976), *T. kodakarensis* (TkCDPIns, Q5JDA9) and *P. furiosus* (PfCDPIns, Q8U1Z6); GDPMan pyrophosphorylase from *T. maritima* (TmGDPMan, Q9X0C3), *B. subtilis* (BsGDPMan, L8AQI9), *P. aeruginosa* (PaGDPMan, O87383), *V. cholera* (VcGDPMan, Q07024), *S. typhimurium* (StGDPMan, P26404), *K. pneumoniae* (KpGDPMan, Q48462) and *E. coli* (EcGDPMan, P24174); UDPGal pyrophosphorylase from *T. maritima* (TmUDPGal, O33836), *Kluyveromyces lactis* (KlUDPGal, P09580), *H. influenza* (HiUDPGal, P31764), *E. coli* (EcUDPGal, P09148) and *S. typhimurium* (StUDPGal, P22714); probable NSPases from *P. horikoshii* (PH0365, O58103; PH0380, O58117; PH0413, O58150; PH0925, O58649; PH1022, O58775; PH1219, O58964; PH1697, O59364 and PH1925, O59588). The UniProt id of each protein is mentioned in the parenthesis. (B) The domain organization of NSPases. The different domains in the NSPases are depicted in different colors (e.g. nucleotidyltransferase domain, green; CDP-alcohol phosphatidyltransferase domain, yellow; mannose-6-phosphate isomerase domain, blue; N- and C-terminal galactose-1-phosphate uridylyltransferase domains, dark and light purple; hexapeptide repeats, red).

### 3.3.4.3 THE NUCLEOTIDYLTRANSFERASE DOMAIN

Since the formation of a NDP-sugar from NTP and sugar occurs in the nucleotidyltransferase domain, it is expected that the phylogeny tree built using the NTD would be a true representation of the evolutionary relationship. In spite of the absence of the additional domains such as GALT, PMI and DIPPS, the proteins PH0365, PH0925 and PH1219 form distinct clusters with UDPGal, GDPMan and CDPIns pyrophosphorylases, respectively (Figure 3.10). On the other hand, PH0413, which clustered with the other promiscuous PhNSPases in case of the full-length protein, displays a closer evolutionary relationship with dTDPGlc pyrophosphorylase (Figure 3.10). However, no clear picture could be obtained for the remaining four promiscuous PhNSPases (i.e. PH0380, PH1022, PH1697 and PH1925) as all the four NSPases cluster with ADPGlc pyrophosphorylase (Figure 3.10).

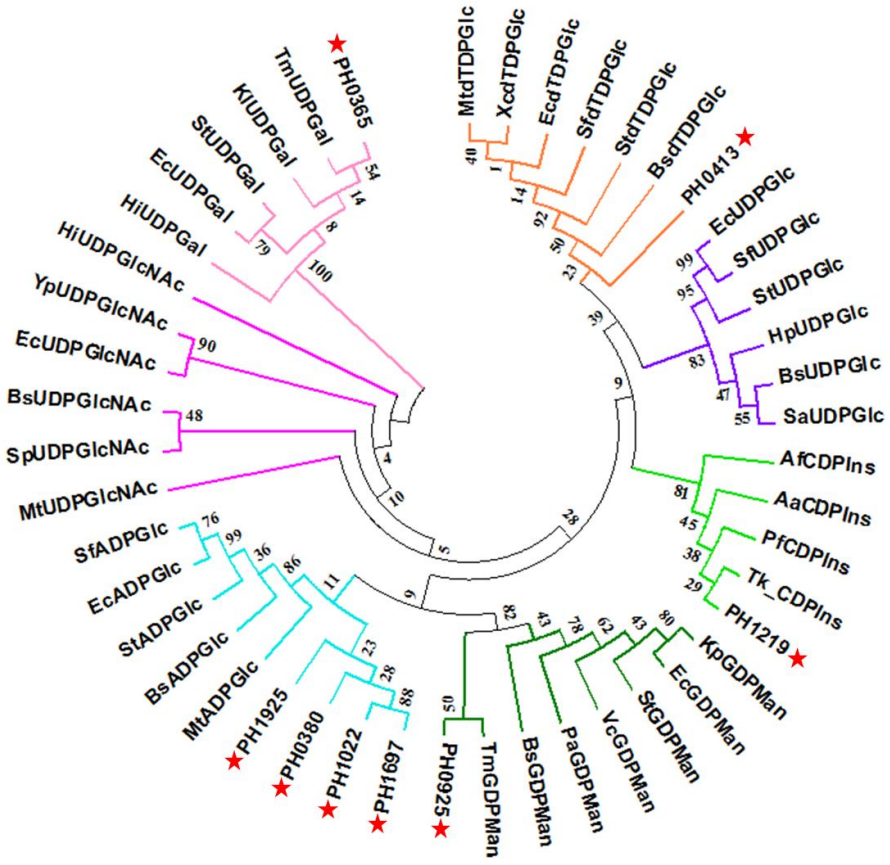



**Figure 3.10. Evolutionary relationship of NSPases on the basis of the nucleotidyltransferase domain.** Four of the PhNSPases (PH0365, PH0413, PH0925, PH1219) form distinct clades with UDPGal (pink), dTDPGlc (orange), GDPMan (green) and CDPins (light green) pyrophosphorylases, respectively. While, PH1022 and PH1697 cluster with ADPGlc pyrophosphorylase (cyan), PH0380 and PH1925 clade with UDPGlc (purple) as well as dTDPGlc (orange) pyrophosphorylase. The PhNSPases are highlighted with red stars. The NTD of NSPases used to build the phylogenetic tree are GDPMan pyrophosphorylase from *V. cholera* (VcGDPMan, Q07024), *S. typhimurium* (StGDPMan, P26404), *E. coli* (EcGDPMan, P24174), *K. pneumoniae* (KpGDPMan, Q48462), *T. maritima* (TmGDPMan, Q9X0C3), *B. subtilis* (BsGDPMan, L8AQI9) and *P. aeruginosa* (PaGDPMan, O87383); UDPGlcNAc pyrophosphorylase from *M. tuberculosis* (MtUDPGlcNAc, P9WMN3), *S. pneumonia* (SpUDPGlcNAc, Q97R46), *B. subtilis* (BsUDPGlcNAc, P14192), *H. influenzae* (HiUDPGlcNAc, P43889), *E. coli* (EcUDPGlcNAc, P0ACC7) and *Y. pestis* (YpUDPGlcNAc, Q8Z957); ADPGlc pyrophosphorylase from *M. tuberculosis* (MtADPGlc, P9WN43), *B. subtilis* (BsADPGlc, P39122), *S. typhimurium* (StADPGlc, P05415), *E. coli* (EcADPGlc, P0A6V1) and *S. flexneri* (SfADPGlc, P0A6V4); dTDPGlc pyrophosphorylase from *B. subtilis* (BsdTDPGlc, P39629), *E. coli* (Ec_dTDPGlc, P37744), *S. flexneri* (SfdTDPGlc, P37779), *S. typhimurium* (StdTDPGlc, P26393), *M. tuberculosis* (MtdTDPGlc, P9WH13) and *X. campestris* (XcdTDPGlc, P0C7J4); UDPGlc pyrophosphorylase from *E. coli* (EcUDPGlc, P0AEP3), *S. flexneri* (SfUDPGlc, P0AEP6), *S. typhimurium* (StUDPGlc, P0A2K7), *B. subtilis* (BsUDPGlc, Q05852), *S. aureus* (SaUDPGlc, Q2FE05) and *H. pylori* (HpUDPGlc, D0ISI7); CDPIns pyrophosphorylase from *A. fulgidus* (AfCDPIns, O29976), *A. aeolicus* (AaCDPIns, O67379), *P. furiosus* (PfCDPIns,

Q8U1Z6) and *T. kodakarensis* (TkCDPIns, Q5JDA9); UDPGal pyrophosphorylase from *E. coli* (EcUDPGal, P09148), *S. typhimurium* (StUDPGal, P22714), *H. influenza* (HiUDPGal, P31764), *K. lactis* (KlUDPGal, P09580) and *T. maritima* (TmUDPGal, O33836); probable NSPases from *P. horikoshii* (PH0925, O58649; PH1697, O59364; PH1022, O58775; PH0380, O58117; PH1925, O59588; PH0413, O58150; PH1219, O58964 and PH0365, O58103). The UniProt id of each protein is mentioned in the parenthesis.

### 3.3.4.4 REPERTOIRE OF RESIDUES AT THE ACTIVE SITE

To confirm the fact that PH0365 is specific towards UDPGal pyrophosphorylase, the amino acid residues at the active-site pocket were analyzed. The active-site pocket of UDPGal pyrophosphorylases is formed by amino acid residues contributed by two subunits. The nucleotide moiety is held by the side chain atoms of residues Arg28, Arg31, Asn77, Asn153, Ser161 and Gln168 while the sugar moiety is coordinated by Asn153, Lys311 and Glu317. Interestingly, the residue Gln168 is known to be critical for the stability of the intermediate in UDPGal pyrophosphorylases. Notably, PH0365 possesses all these residues. Most importantly, PH0365 also possesses His166 which serves as the nucleophillic catalyst (Wedekind et al., 1995) during the catalytic reaction (Figure 3.11).

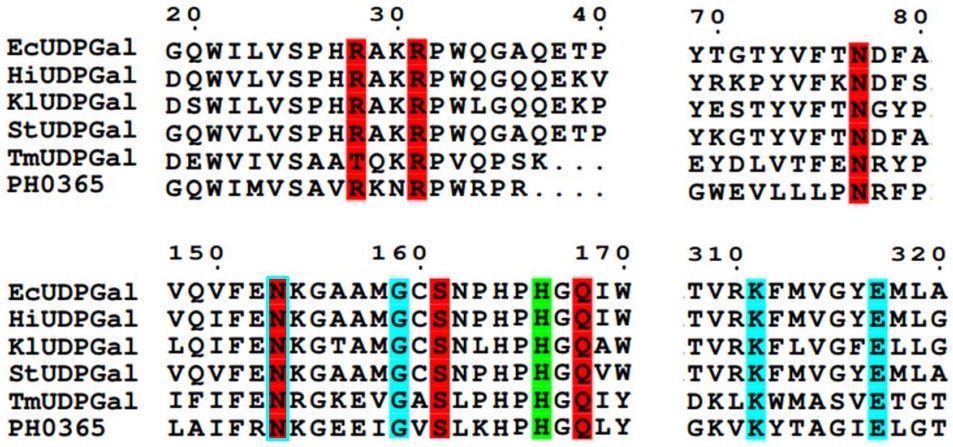



**Figure 3.11. Multiple sequence alignment (MSA) of PH0365 with UDPGal pyrophosphorylases.** The amino acid residues involved in interacting with uridine and galactose are highlighted in red and cyan, respectively. The residue His166, which serves as the nucleophillic catalyst, is highlighted in green. Only the partial MSA has been shown here for the clarity of the figure. The UDPGal pyrophosphorylases used in the MSA are from *E. coli* (EcUDPGal, P09148), *H. influenza* (HiUDPGal, P31764), *K. lactis* (KlUDPGal, P09580), *S. typhimurium* (StUDPGal, P22714) and *T. maritima*

(TmUDPGal, O33836) and PH0365 (O58103). The UniProt id of each protein is mentioned in the parenthesis.

On the other hand, PH1219 showed closest evolutionary relationship as well as sequence homology to CDPIns pyrophosphorylases. PH1219 along with the other CDPIns pyrophosphorylases also possess the signature consensus motif G-X-G-X-R-$X_{\{5,10\}}$-K of NSPases. A three-dimensional structure of CDPIns pyrophosphorylase bound to CDP-inositol is unavailable. However, CDP-inositol docked at the active-site pocket of the nucleotidyltransferase domain of CDPIns pyrophosphorylase have helped in identifying the amino acid residues essential for substrate binding (Brito et al., 2011). All the amino acid residues (Lys91, His143, Asn150, Asp171, His172, Asp226, Trp270, Thr203 and Lys217) required for coordinating the CDP-inositol are absolutely conserved in PH1219 as well (Figure 3.12). Furthermore, a consensus sequence pattern, D-G-$X_2$-A-R-$X_{7,8}$-G-$X_3$-D-$X_3$-D at DIPPS domain is also conserved in PH1219. Apart from these, amino acid residues Asn304, Arg305, Ser308, Ser354, Asp357 and Arg443, which are strictly conserved in DIPPS protein and might be involved in interacting with the substrate, are well conserved (Figure 3.12).

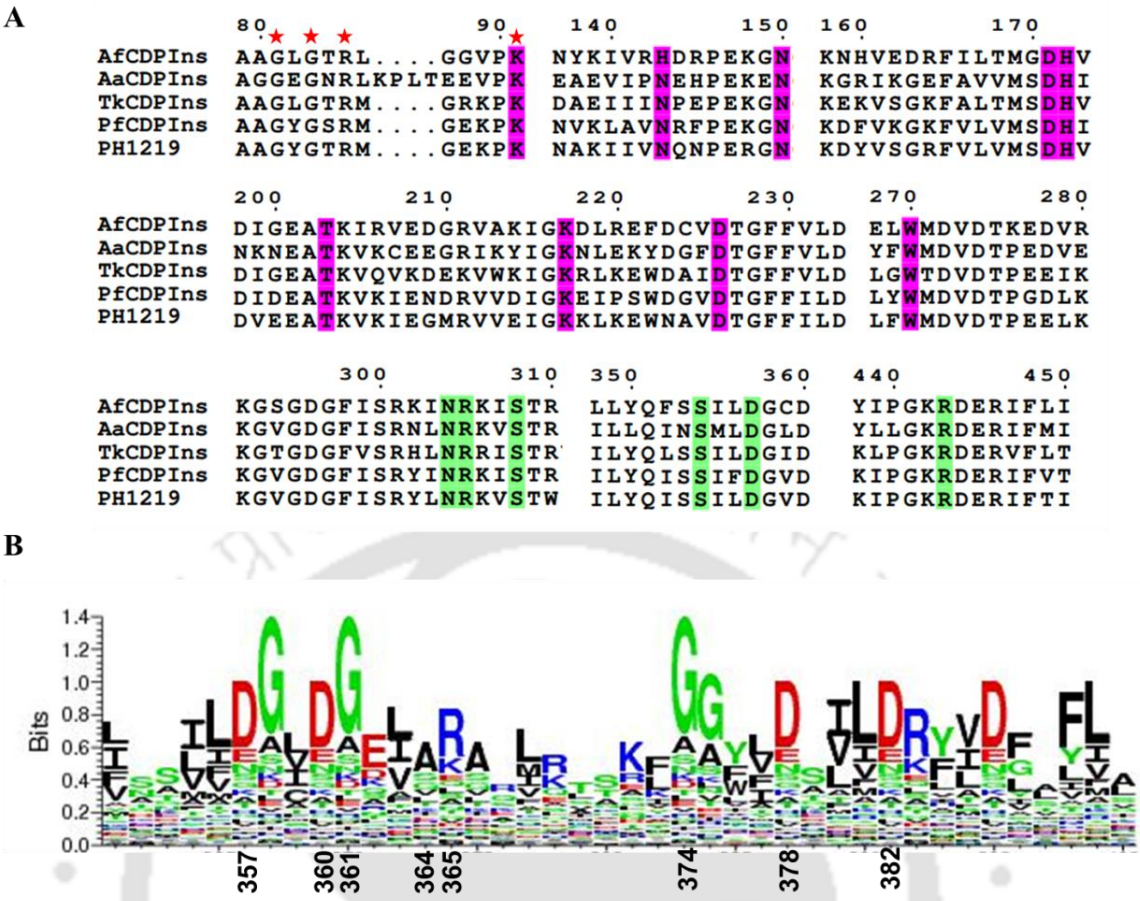



**Figure 3.12. Multiple sequence alignment (MSA) of PH1219 with CDPIns pyrophosphorylases.** The amino acid residues involved in interacting with CDP-inositol are highlighted in pink. The conserved amino acid residues of CDP-alcohol phosphotransferase domain (DIPPS) are highlighted in green. The consensus sequence motif G-X-G-X-R-$X_{\{5,10\}}$-K of NDP sugar pyrophosphorylase is highlighted with red stars. Only the partial MSA has been shown here for the clarity of the figure. The CDPIns pyrophosphorylases used in the MSA are from *A. fulgidus* (AfCDPIns, O29976), *A. aeolicus* (AaCDPIns, O67379), *T. kodakarensis* (TkCDPIns, Q5JDA9), *P. furiosus* (PfCDPIns, Q8U1Z6) and PH1219 (O58964). The UniProt id of each protein is mentioned in the parenthesis. (B) A graphical representation of the consensus sequence motif D-G-$X_2$-A-R-$X_{7,8}$-G-$X_3$-D-$X_3$-D at the DIPPS domain. The residue number (according to CDPIns from *A. fulgidus*) of the consensus motif is mentioned along the X-axis. The height of each residue represents the degree of conservation (Y-axis) and the size of each residue is proportional to the conservation frequency at the particular position (X-axis).

Although, a clear substrate preference could be predicted for PH0365 and PH1219, the remaining PhNSPases (PH0380, PH0413, PH0925, PH1022, PH1697 and PH1925) belonging to the subgroup IC exhibited a promiscuous behavior. This can be expected owing to the high degree of similarity of the atomic structures of the substrates.

Henceforth, we tried to focus on the minute differences among the substrates that would in turn demand the presence of an exclusive set of amino acid residues in each case. Those amino acid residues which interacted with the substrate via either the N or O atom of the main chain were not considered as they would provide only trivial information.

In UDPGlcNAc pyrophosphorylase, the amino acid residues interacting via their side chain atom(s) include Gln76, Thr82, Tyr103, Asp105 and Glu154 (Figure 3.13). Out of these five residues, Tyr103, which interacts with the O5' of GlcNAc moiety, itself is not conserved within UDPGlcNAc pyrophosphorylases and hence has not been discussed further. The N3 and O4 atoms of the uridine moiety are held by the $O^{\varepsilon 1}$ and $N^{\varepsilon 2}$ of Gln76. Along with UDPGlcNAc pyrophosphorylase, the glutamine is conserved in UDPGlc and dTDPGlc pyrophosphorylases as well. However, in ADPGlc and GDPMan pyrophosphorylases, a glutamate residue is present instead of glutamine. Among PhNSPases, glutamine is present in PH0413 and PH1925 while either glutamate or aspartate is present in PH0380, PH0925, PH1022 and PH1697 (Figure 3.14). This indicates that the cognate substrate for PH0413 and PH1925 might bind either UDP or dTDP while PH0380, PH0925, PH1022 and PH1697 might interact with ADP or GDP.

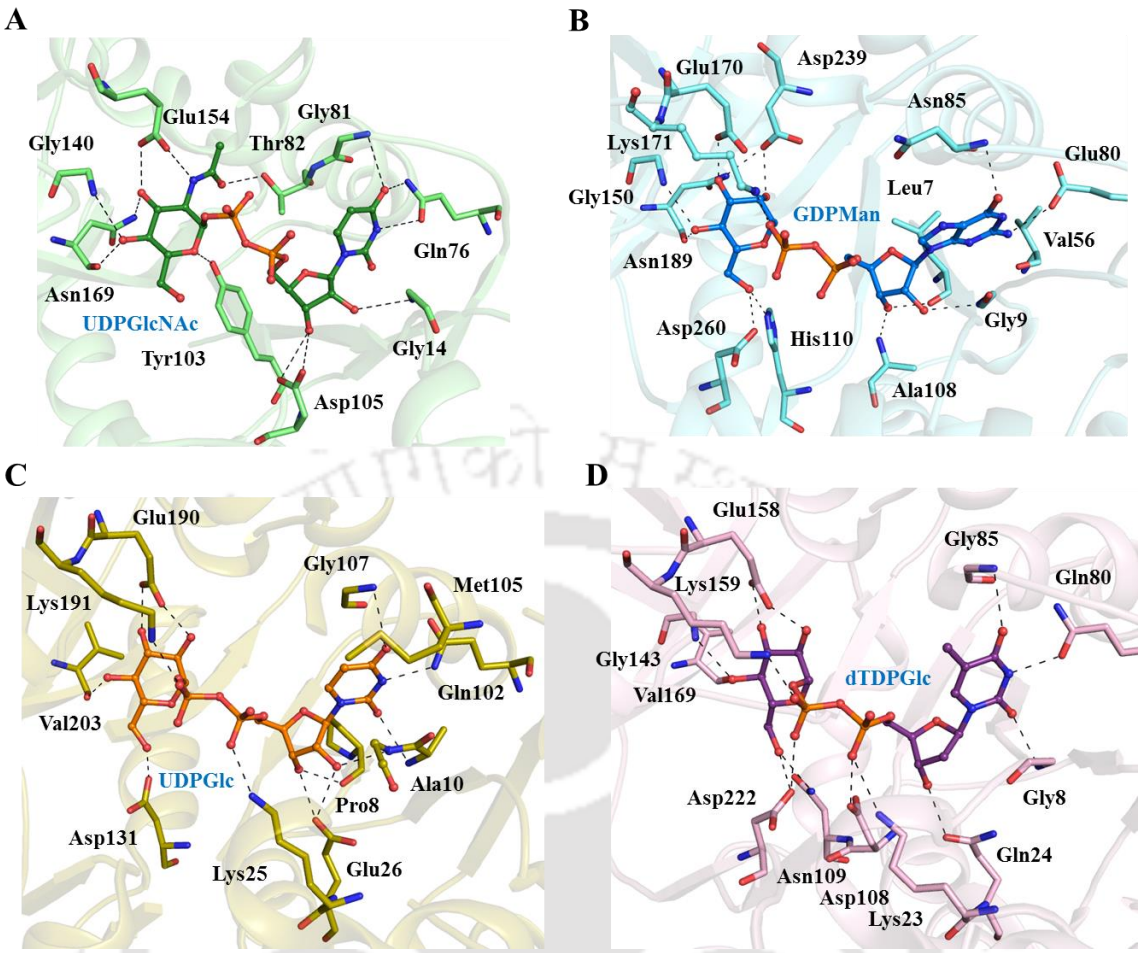



**Figure 3.13. Comparison of active-site pockets of NSPases.** Active-site pocket of (A) UDPGlcNAc (PDB id: 2IO6), (B) GDPMan (PDB id: 2X5Z), (C) UDPGlc (PDB id: 3JUK) and (D) dTDPGlc (PDB id: 6B5E) pyrophosphorylases. The amino acid residues involved in interacting with the substrate (shown in ball-and-stick model) are shown as lines. The interactions are depicted as black dashed lines.

As in UDPGlcNAc pyrophosphorylase, a similar conserved set of amino acid residues are involved in interacting with the nucleotide moiety in UDPGlc and dTDPGlc pyrophosphorylase. The only amino acid residue that differs is the presence of Gln24 (26$^{th}$ position in EcUDPGlcNAc pyrophosphorylase), which interacts with O3 atom of the deoxyribose moiety of dTDP-glucose pyrophosphorylase. In UDPGlc pyrophosphorylase, instead of glutamine, glutamate is present which interacts with both the O2 and O3 atoms of the ribose moiety (Figure 3.13). Among PhNSPases, PH0413 possesses glutamine and thus can be anticipated to interact with dTDP instead of UDP (Figure 3.14). Apart from PH0413, PH0925 like other GDPMan pyrophosphorylases also

possesses a glutamine, however, the oxygen atom of the ribose does not exhibit any interaction in the three-dimensional structure of GDPMan pyrophosphorylase from *T. maritima* (PDB id: 2X5Z) (Figure 3.13).

Next, we tried to probe the presence of amino acid residues specifically exclusive to the sugar moieties. The $O^{\gamma 1}$ of Thr82 interacts with the carbonyl oxygen atom of the GlcNAc moiety and thus can be anticipated that it would be exclusive to UDPGlcNAc pyrophosphorylases. However, ADPGlc as well GDPMan pyrophosphorylases also possess a threonine while UDPGlc and dTDPGlc possess leucine instead. In case of PhNSPases, threonine is conserved in PH0380, PH0925, PH1022, PH1697 and PH1925 while leucine is present in PH0413 (Figure 3.14). This indicates that PH0413 might not be able to interact with GlcNAc moiety. Another fact that needs to be considered is that the GlcNAc moiety is naturally present only in association with UDP. Apart from PH0413, only PH1925 has been found to interact with a pyrimidine (UDP) and presence of Thr82 in PH1925 further accentuates the fact of its being a probable UDPGlcNAc pyrophosphorylase. Furthermore, the amino acid residues involved in interacting with the GlcN 1-P at the acetyltransferase active site of UDPGlcNAc pyrophosphorylases (GlmU) are also conserved in PH1925 (Figure 3.15).

Residues Asp105 and Glu154, which coordinates O3 and O3' oxygen atoms of the ribose and GlcNAc are conserved in all the pyrophosphorylases including the PhNSPases. Although, Asp105 is not conserved in ADPGlc and GDPMan pyrophosphorylase, the presence of another aspartate adjacent to Asp105 might be involved in interacting with the oxygen atoms. The presence of Asp105, Glu154 and more importantly Thr82 in PH0380, PH0925, PH1022 and PH1697 indicates that in absence of their cognate (or any other) substrate, these PhNSPases might have the capability to accommodate a GlcNAc moiety.

In GDPMan pyrophosphorylase, apart from Glu170 (Glu154 in EcUDPGlcNAc pyrophosphorylase), the other amino acid residues interacting with their side chain atoms with mannose include His110, Asp239 and Asp260. Both His110 and Asp260 ($107^{th}$ and $223^{rd}$ position in EcUDPGlcNAc pyrophosphorylase) interact with the O6' of the mannose moiety in GDPMan pyrophosphorylase (Figure 3.13). Apart from GDPMan pyrophosphorylase, His110 is also conserved in ADPGlc pyrophosphorylase and among

the PhNSPases it is exclusive to PH0925 while Asp260 is conserved in all the NSPases except UDPGlcNAc pyrophosphorylase. Interestingly, Asp239 which forms hydrogen bond with the O2' of the mannose is exclusive to GDPMan pyrophosphorylase and is present in PH0925 alone (Figure 3.14). This explains the ability of PH0925 to accommodate GDP-mannose at the active-site pocket.

The amino acid residues, Glu158 and Glu190 in dTDPGlc and UDPGlc pyrophosphorylases, respectively which are responsible for holding the O2' & O3' of the glucose moiety are well conserved in all the promiscuous PhNSPases. In UDPGlc pyrophosphorylase, the O6' of the glucose moiety is coordinated by Asp131 ($106^{th}$ position in EcUDPGlcNAc) while in case of dTDPGlc pyrophosphorylase, it is held by Asp222 (Asp260 in TmGDPMan) (Figure 3.13). Unlike Asp131, Asp222 is absolutely conserved in all the promiscuous PhNSPases indicating that all of them might have the capability to accommodate a glucose moiety (Figure 3.14).

Although a very specific substrate for PH0380, PH1022 and PH1697 could not be obtained, the presence of glutamate (Gln76 in EcUDPGlcNAc pyrophosphorylase) which might interact with purine nucleotide as well as amino acid residues interacting with glucose indicate that PH0380 might be a probable ADPGlc pyrophosphorylase. On the other hand, the repertoire of amino acid residues at the active-site pocket hints toward the fact that PH1022 and PH1697 might function as highly promiscuous NSPases.

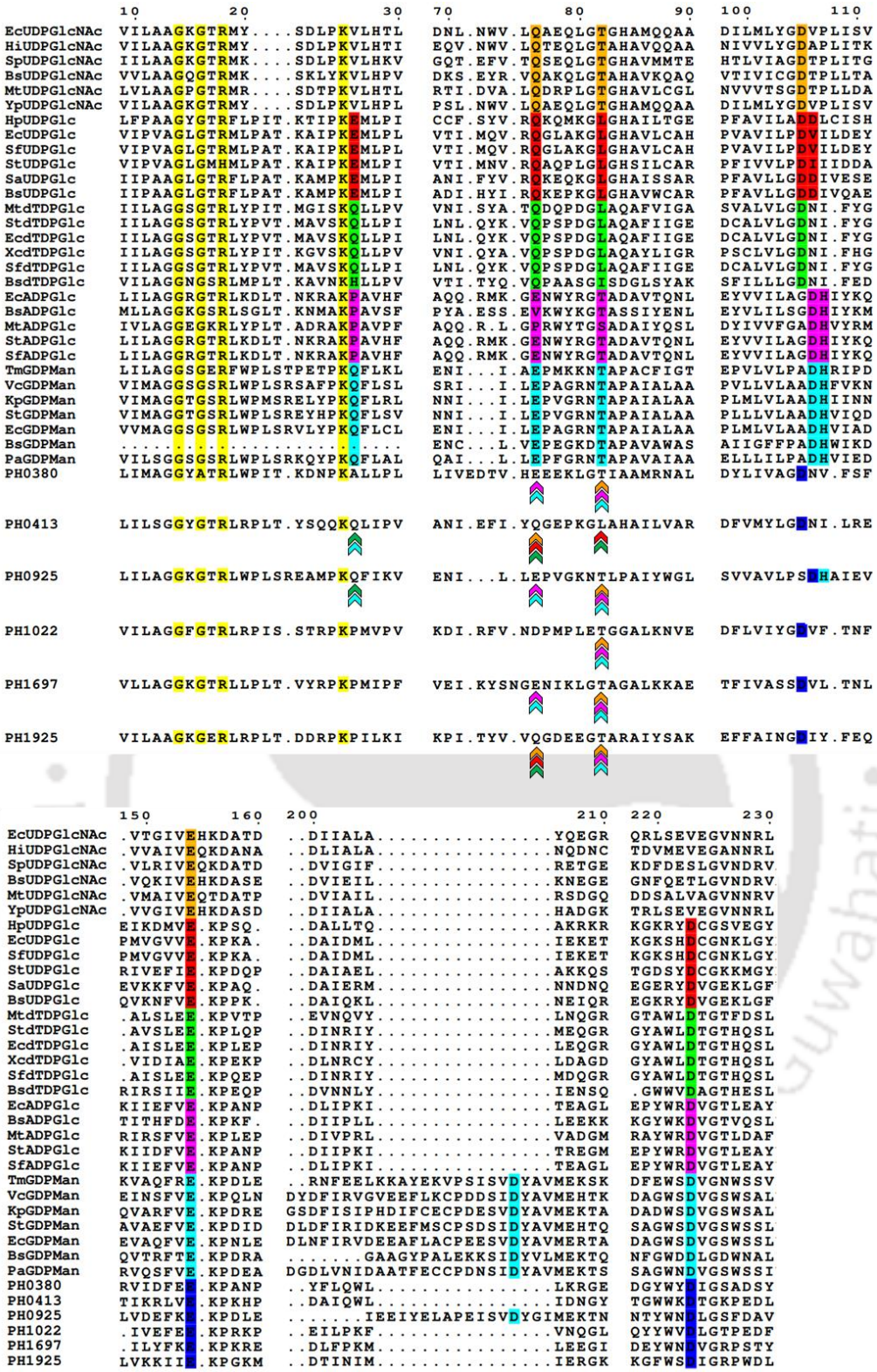


**Figure 3.14. Multiple sequence alignment (MSA) of promiscuous NSPases.** The amino acid residues involved in interacting with the substrate (NDP-sugars) via their side chain atoms are highlighted with different colors (such as UDPGlcNAc pyrophosphorylase, orange; UDPGlc pyrophosphorylase, red; dTDPGlc pyrophosphorylase, green; ADPGlc pyrophosphorylase, magenta and GDPMan

pyrophosphorylase, cyan). The amino acid residues showing conservation with the respective NSPase are highlighted with arrowheads of the corresponding color while those which are conserved in all the PhNSPases are highlighted in blue. The consensus sequence motif of NSPases is highlighted in yellow. Only the partial MSA has been shown for the clarity of the figure. The NSPases used in the MSA are UDPGlcNAc pyrophosphorylase from *E. coli* (EcUDPGlcNAc, P0ACC7), *H. influenzae* (HiUDPGlcNAc, P43889), *S. pneumonia* (SpUDPGlcNAc, Q97R46), *B. subtilis* (BsUDPGlcNAc, P14192), *M. tuberculosis* (MtUDPGlcNAc, P9WMN3) and *Y. pestis* (YpUDPGlcNAc, Q8Z957); UDPGlc pyrophosphorylase from *H. pylori* (HpUDPGlc, D0ISI7), *E. coli* (EcUDPGlc, P0AEP3), *S. flexneri* (SfUDPGlc, P0AEP6), *S. typhimurium* (StUDPGlc, P0A2K7), *S. aureus* (SaUDPGlc, Q2FE05) and *B. subtilis* (BsUDPGlc, Q05852); dTDPGlc pyrophosphorylase from *M. tuberculosis* (MtdTDPGlc, P9WH13), *S. typhimurium* (StdTDPGlc, P26393), *E. coli* (EcdTDPGlc, P37744), *X. campestris* (XcdTDPGlc, P0C7J4), *S. flexneri* (SfdTDPGlc, P37779) and *B. subtilis* (BsdTDPGlc, P39629); ADPGlc pyrophosphorylase from *E. coli* (EcADPGlc, P0A6V1), *B. subtilis* (BsADPGlc, P39122), *M. tuberculosis* (MtADPGlc, P9WN43), *S. typhimurium* (StADPGlc, P05415) and *S. flexneri* (SfADPGlc, P0A6V4); GDPMan pyrophosphorylase from *T. maritima* (TmGDPMan, Q9X0C3), *V. cholera* (VcGDPMan, Q07024), *K. pneumoniae* (KpGDPMan, Q48462), *S. typhimurium* (StGDPMan, P26404), *E. coli* (EcGDPMan, P24174), *B. subtilis* (BsGDPMan, L8AQI9) and *P. aeruginosa* (PaGDPMan, O87383); probable NSPases from *P. horikoshii* (PH0925, O58649; PH1697, O59364; PH1022, O58775; PH0380, O58117; PH1925, O59588; PH0413, O58150; PH1219, O58964 and PH0365, O58103). The UniProt id of each protein is mentioned in the parenthesis.

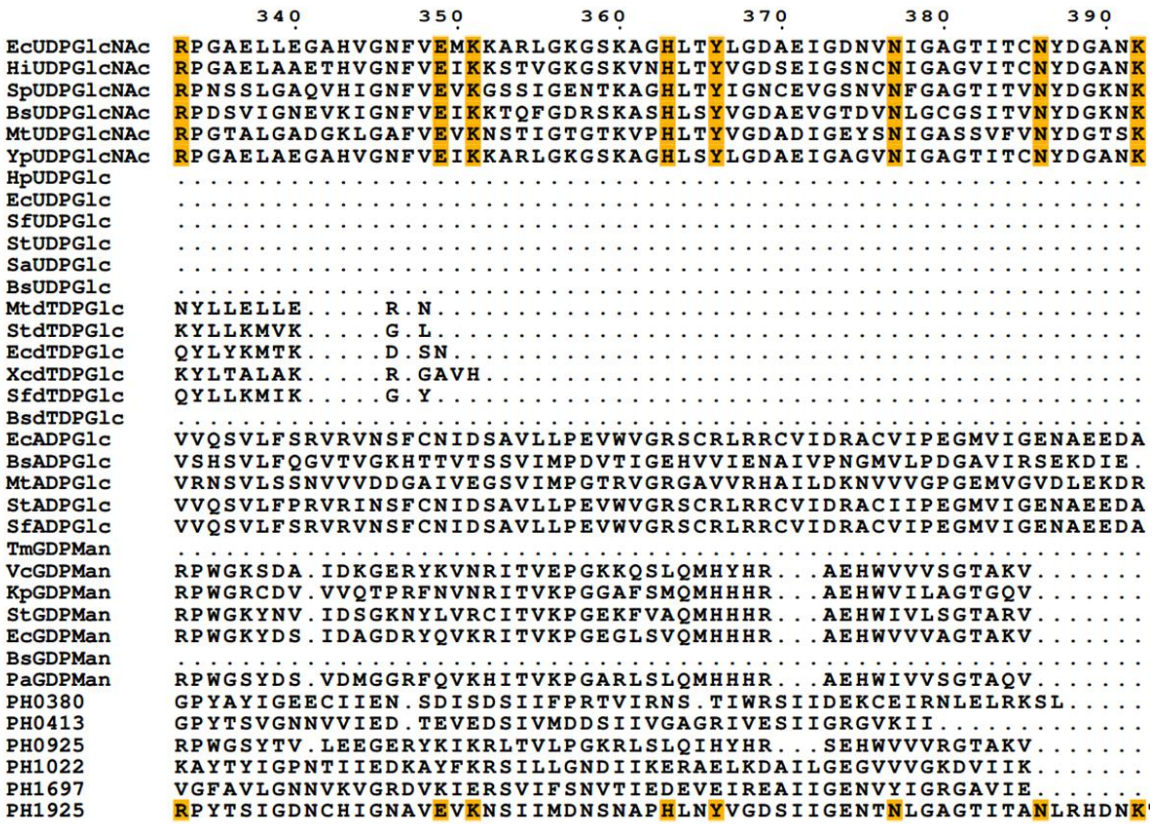


**Figure 3.15. Multiple sequence alignment (MSA) of promiscuous NSPases to identify the conserved amino acid residues of the acetyltransferase domain of UDPGlcNAc pyrophosphorylase.** The amino acid residues exclusively conserved in PH1925 are highlighted in orange. The NSPases used the MSA are UDPGlcNAc pyrophosphorylase from *E. coli* (EcUDPGlcNAc, P0ACC7), *H. influenzae* (HiUDPGlcNAc, P43889), *S. pneumonia* (SpUDPGlcNAc, Q97R46), *B. subtilis* (BsUDPGlcNAc, P14192), *M. tuberculosis* (MtUDPGlcNAc, P9WMN3) and *Y. pestis* (YpUDPGlcNAc, Q8Z957); UDPGlc pyrophosphorylase from *H. pylori* (HpUDPGlc, D0ISI7), *E. coli* (EcUDPGlc, P0AEP3), *S. flexneri* (SfUDPGlc, P0AEP6), *S. typhimurium* (StUDPGlc, P0A2K7), *S. aureus* (SaUDPGlc, Q2FE05) and *B. subtilis* (BsUDPGlc, Q05852); dTDPGlc pyrophosphorylase from *M. tuberculosis* (MtdTDPGlc, P9WH13), *S. typhimurium* (StdTDPGlc, P26393), *E. coli* (EcdTDPGlc, P37744), *X. campestris* (XcdTDPGlc, P0C7J4), *S. flexneri* (SfdTDPGlc, P37779) and *B. subtilis* (BsdTDPGlc, P39629); ADPGlc pyrophosphorylase from *E. coli* (EcADPGlc, P0A6V1), *B. subtilis* (BsADPGlc, P39122), *M. tuberculosis* (MtADPGlc, P9WN43), *S. typhimurium* (StADPGlc, P05415) and *S. flexneri* (SfADPGlc, P0A6V4); GDPMan pyrophosphorylase from *T. maritima* (TmGDPMan, Q9X0C3), *V. cholera* (VcGDPMan, Q07024), *K. pneumoniae* (KpGDPMan, Q48462), *S. typhimurium* (StGDPMan, P26404), *E. coli* (EcGDPMan, P24174), *B. subtilis* (BsGDPMan, L8AQI9 ) and *P. aeruginosa* (PaGDPMan, O87383); probable NSPases from *P. horikoshii* (PH0925, O58649; PH1697, O59364; PH1022, O58775; PH0380, O58117; PH1925, O59588; PH0413, O58150; PH1219, O58964 and PH0365, O58103). The UniProt id of each protein is mentioned in the parenthesis.

## 3.4 DISCUSSION

Since early, the question whether archaea actually possess the functional homologs of eukaryotic translation IFs remains debatable. Certain archaeal proteins are often annotated as archaeal translation IFs owing to the significant similarity with eukaryotic translation IFs at the primary structure level. This often creates incertitude and leads to misannotation of newly discovered proteins. In this chapter, we have taken up five ORFs (PH1771.1, PH0208, PH0702, PH1022 and PH1697) which shares significant sequence similarity with eukaryotic translation IFs such as eIF1 and eIF2B (α, β, δ, γ and ε subunits).

In *P. horikoshii* OT3, an ORF PH1771.1 has been found to be the closest homolog of eIF1 indicating that it might function as aIF1. However, a comparable sequence similarity between PH1771.1 and bacterial translation initiation inhibitor YciH as well leads to an uncertainty regarding the true function of PH1771.1. A close evolutionary relationship among aIF1 (including PH1771.1), eIF1 and YciH suggests a common ancestor among the three groups of proteins. The evolutionary relatedness also explains

their high resemblance and it might be possible that over the years, these proteins might have evolved further to perform 'domain-specific' functions.

Two genes PH0702 and PH0208, which were initially thought to be homologs of the regulatory subunits (α, β and δ) of eIF2B (Kakuta et al., 2004), were analyzed to annotate their function. Our results reveal that the proteins PH0702 and PH0208 encode the enzymes M1Pi and R15Pi, respectively. We found that all the active-site residues required for the catalytic activity of the enzymes M1Pi and R15Pi are absolutely conserved in the protein sequence of PH0702 and PH0208, respectively. Most importantly, the catalytic residues Cys160 and Asp240 of M1Pi (Cys133 and Asp202 in R15Pi) are also found to be conserved in PH0702 and PH0208, respectively. Furthermore, in the case of PH0702, the presence of a hydrophobic patch which is known to stabilize the thiol group of MTR-1-P/MTRu-1-P affirms the claim that the protein performs a function similar to that of the enzyme M1Pi. Similarly, the protein PH0208 harbors positively charged active-site residues, as in the enzyme R15Pi, required to coordinate with the two negatively charged phosphate groups of the R15P/RuBP.

Similarly, PH1022 and PH1697 which show homology with the catalytic subunits (γ and ε) of eIF2B are found to be similar to NSPase enzymes. The evolutionary relationship of PH1022 and PH1697 with NSPases instead of the catalytic subunits of eIF2B and presence of the consensus signature motif of NSPases in both the proteins hints toward the fact that they might function as NSPase. While identifying the cognate ligand of PH1022 and PH1697, another set of NSPases were identified which were also included in the study. The identified NSPases could be divided into two discrete sets of NSPases, one set exhibit substrate promiscuity while the other exhibits substrate specificity. Evolutionary analysis as well as in depth investigation of amino acid composition at the active-site pocket suggests that PH1022 and PH1697 are highly promiscuous NSPases with the propensity of interacting with various combination of nucleotide and sugar moieties to form NDP sugar.

## 3.5 CONCLUSION

From this chapter, it can be concluded that the ORFs which were assumed to encode the homologs of eukaryotic translation IFs might instead function as enzymes involved in various cellular processes. PH1771.1 might function as either aIF1 or YciH while PH0208 and PH0702 might function as R15Pi andM1Pi, respectively. Similarly, PH1022 and PH1697 might function as NSPase enzymes. In the following chapters, experimental validation has been performed to confirm the function of PH1771.1, PH0208 and PH0702 as aIF1, R15Pi and M1Pi, respectively.

# CHAPTER 4

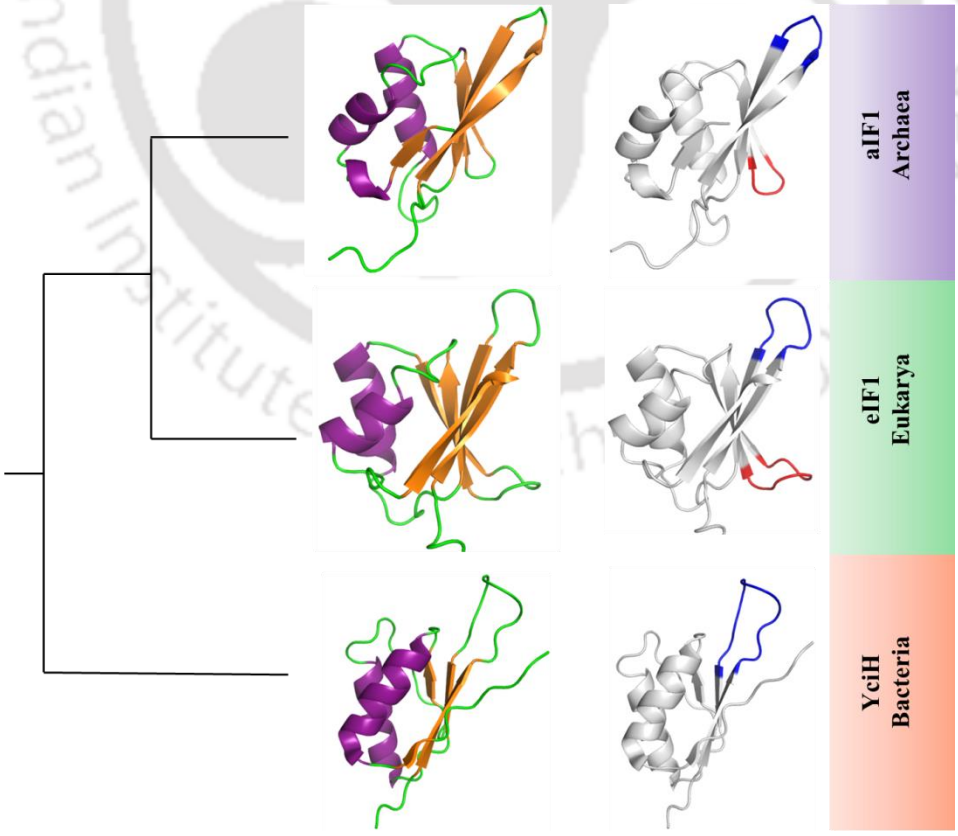
## Structural insights into the function of archaeal translation initiation factor 1

This chapter has been published as:

Gogoi, P. and Kanaujia, S.P., 2018. Archaeal and eukaryal translation initiation factor 1 differ in their RNA interacting loops. *FEBS Letters*, *592*(9), 1602-1610.

## ABSTRACT

The archaeal translation initiation factor 1 (aIF1) is reported to be functionally homologous to eIF1. However, lack of a structural comparison between aIF1 and eIF1 has limited our understanding of the structural (dis)similarities. In this chapter, we have determined the three-dimensional crystal structure of an open reading frame (ORF) PH1771.1 encoding aIF1 in *P. horikoshii* OT3. Results reveal that although aIF1 has low sequence similarity with eIF1, high structural homology exists between the two proteins. Nonetheless, notable critical differences between aIF1 and eIF1 could still be perceived at the $\beta_1$–$\beta_2$ basic loop, the acidic loop and the solvent-exposed surface. These differences might lead to a slightly divergent mode of action of aIF1 during archaeal translation initiation.



## 4.1 INTRODUCTION

Protein translation initiation mechanism and apparatus in the three domains of life viz. bacteria, archaea and eukarya has diverged significantly over time. This process has evolved from being a simple process in bacteria to a very complex one in eukaryotes. Bacteria harbor only three translation initiation factors (IFs) namely IF1, IF2 and IF3; each of them performing the crucial functions of preventing the premature binding of elongator tRNA to A-site, promoting the binding of fMet-tRNA$^{fMet}$ to the P-site and controlling the precision of codon-anticodon recognition, respectively (Gualerzi and Pon, 1990; La Teana et al., 1996; Dahlquist and Puglisi, 2000; Brandi et al., 2004). On the other hand, eukaryotes utilize almost a dozen of translation IFs to execute the various steps of protein biosynthesis. During the start codon selection, the protein eIF1 (referred to as SUI1 in yeast) plays a central role by discriminating against non-AUG codons or AUG codons that have non-optimal context and also by promoting the dissociation of aberrantly assembled ribosomal complexes (Pestova et al., 1998; Pestova and Kolupaeva, 2002). While executing its function, eIF1 is placed close to the P-site of the 40S small ribosomal subunit which in turn stabilizes a 'scanning competent' form obligatory for loading of the ternary complex (TC) eIF2•GTP•Met-tRNA$_i^{Met}$ during the scanning process (Martin-Marcos et al., 2013). In the P-site, eIF1 interacts with 18S rRNA through the basic residues present in the $\beta_1$–$\beta_2$ hairpin loop and helix $\alpha_1$. These interactions are considered to be pivotal for the functioning of eIF1 (Fletcher et al., 1999; Lomakin et al., 2003; Rabl et al., 2011; Martin-Marcos et al., 2013; Weisser et al., 2013). Until the recognition of an optimal start codon, the $\beta_1$–$\beta_2$ basic loop hinders the binding of the Met-tRNA$_i^{Met}$ at the P-site by clashing with its acceptor stem loop (ASL) (Rabl et al., 2011). This hindrance can only be overcome by the energy provided by correct start codon-anticodon base pairing which also promotes the dissociation of eIF1 (Nanda et al., 2009). Another loop comprising mostly of acidic residues remains in close proximity to the D-stem loop of Met-tRNA$_i^{Met}$ prior to the start codon recognition. Once the canonical AUG codon is encountered, the acidic loop is brought to a close contact with negatively charged phosphate backbone of Met-tRNA$_i^{Met}$ resulting in an electrostatic repulsion eventually leading to the eIF1 dissociation (Rabl et al., 2011; Hinnebusch, 2017). Among a few other translation initiation factors, eIF1 is considered to be universally conserved in all the three domains of life; however, the homolog of eIF1, known as YciH, is not present in all bacteria (Kyrpides and Woese, 1998a). Furthermore, recent studies have

revealed that instead of functioning as a translation initiation factor, YciH acts as an inhibitor of translation initiation during stress conditions by binding firmly to the P-site with the aid of a longer $\beta_1$–$\beta_2$ basic loop (Osterman et al., 2015). The fidelity of translation initiation in bacteria is alternatively established by IF3; which, however, lacks sequence or structural homology with the protein eIF1 (Lomakin et al., 2006). Unlike bacteria, archaea possess a relatively higher number of translation IFs; most of which are homologous to eukaryotic IFs (Dennis, 1997; Bell and Jackson, 1998). The only factors evidently absent in archaea include eIF3, eIF4F and eIF5 (Londei, 2005). Among others, archaea possess an ORF encoding aIF1. The protein aIF1 has been shown to share a similar topographical localization as of eIF1 by binding to a position adjacent to the P-site of 30S small ribosomal subunit (Coureux et al., 2016). Furthermore, similar to the functions established by eIF1, aIF1 also aids in the binding of aIF2 to the ribosome as well as retains the ability to discriminate against non-canonical start codons (Hasenöhrl et al., 2009). Thus, aIF1 can be considered to be functionally similar to eIF1 (Hasenöhrl et al., 2006). However, the structural analogy between aIF1 and eIF1 has yet to be comprehended. In this chapter, the three-dimensional crystal structure of the protein PH1771.1 (containing 99 amino acid residues), a homolog of eIF1, from *P. horikoshii* OT3 has been reported at 1.5 Å resolution. A detailed structural comparison has been made between aIF1 and its eukaryal (eIF1) as well as bacterial (YciH) counterparts to obtain the minute yet noteworthy differences that might evoke a slightly divergent mode of action of aIF1 during the process of translation initiation.

## 4.2 MATERIALS AND METHODS

### 4.2.1 CLONING, OVER EXPRESSION AND PROTEIN PURIFICATION

The gene fragment encoding the initiation factor aIF1 was amplified (Figure 4.1A) from the genome of *P. horikoshii* OT3 using the forward primer 5′-GCCG**CATATG**CACCATCATCATCACCACATGGTGCCTAGGATAG-3′ containing NdeI restriction site (bold) followed by a 6xHis-tag (underlined) and a reverse primer, 5′-GCG**CTCGAG**TTACTCTACCTCTATGAGCTCCTCTG-3′ containing XhoI restriction site (bold). The amplified gene was inserted into the plasmid pET-22b(+) excised with the same restriction enzymes. Subsequently, the clones were confirmed by

double digestion using the restriction enzymes, NdeI and XhoI by incubating the vectors at 37°C for two hours (Figure 4.1B). In addition, the clone was confirmed by DNA sequencing. The over expression of the gene encoding aIF1 was optimized using varying concentrations (0.1, 0.5 and 1.0 mM) of isopropyl β-D-1-thiogalactopyranoside (IPTG) in expression systems *Escherichia coli* BL21-CodonPlus (DE3)-RIL and *E. coli* Rosetta (DE3) competent cells grown at 37°C in Luria Bertani (LB) media supplemented with 100 µg $ml^{-1}$ ampicillin and 34 µg $ml^{-1}$ chloramphenicol. Similarly, the induction time of the over expression was optimized and it was found that the maximum expression of aIF1 (11.4 kDa) was obtained after fourth hour of induction (Figure 4.1C). Subsequently, the solubility of the protein was checked and the results showed that the protein remains in the soluble form (Figure 4.1D). For purification of aIF1, the protein expression was induced by the addition of 1 mM IPTG and the culture was further grown for five hours at 37°C. The cells were harvested by centrifugation and resuspended in lysis buffer composed of 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM imidazole, 1 mM phenylmethylsulfonyl fluoride (PMSF), 3 mM β-mercaptoethanol (β-ME) and 0.2 mg $ml^{-1}$ lysozyme. The cells were incubated in ice for one hour and then disrupted using sonicator. The cell lysate was incubated at 90°C for 13 min to remove the thermolabile proteins deriving from the host cells. The insoluble debris were removed by centrifugation at 12,000 rpm for 40 min at 4°C. The supernatant fraction was applied to Ni-NTA affinity column pre-equilibrated with binding buffer A (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM imidazole, 1 mM PMSF and 3 mM β-ME) and incubated for two hours. The column was washed with five column volume of each of wash buffer B (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM imidazole, 1 mM PMSF and 3 mM β-ME) and wash buffer C (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 20 mM imidazole, 1 mM PMSF and 3 mM β-ME). The proteins were eluted with 250 mM imidazole in 20 mM Tris-HCl pH 8.0, 150 mM NaCl and 1 mM PMSF. The eluted fractions (Figure 4.1E) were collected and step-wise dialysis was performed against 20 mM Tris-HCl pH 8.0 and 150 mM NaCl to remove imidazole. The dialysed proteins were concentrated (Figure 4.1F) using Amicon Ultra centrifugal filter unit (Millipore, Burlington, MA, USA) with a molecular weight cut-off of 3 kDa.

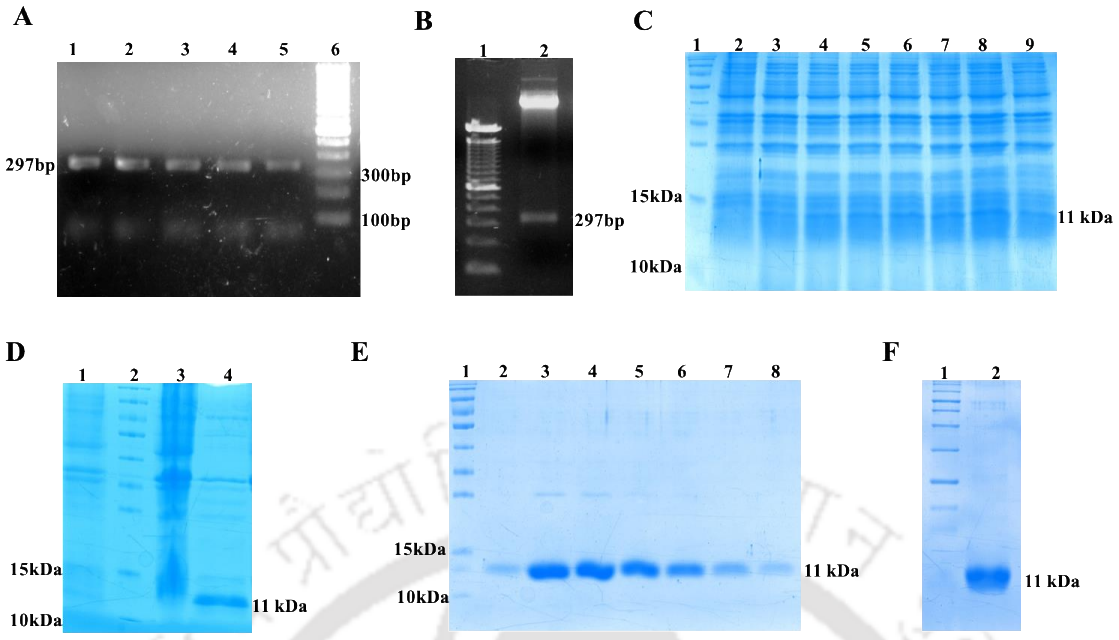



**Figure 4.1. Cloning, over expression and purification of aIF1.** (A) PCR amplification of aIF1 analysed in 0.8% agarose gel (Lane 1-5: PCR amplified aIF1, Lane 6: DNA ladder) (B) Clone confirmation using the double digestion of the plasmid pET22b(+) containing aIF1 analysed in 0.8% agarose gel (Lane 1: DNA ladder, Lane 2: positive clone). (C) Over expression of aIF1 at 1 mM IPTG at different time intervals analysed in 12% SDS-PAGE (Lane 1: protein marker, Lane 2: uninduced, Lane 3: 1$^{st}$ hr, Lane 4: 2$^{nd}$ hr, Lane 5: 3$^{rd}$ hr, Lane 6: 4$^{th}$ hr, Lane 7: 5$^{th}$ hr, Lane 8: 6$^{th}$ hr, Lane 9: 7$^{th}$ hr). (D) Solubility test for the over expressed aIF1 analysed in 12% SDS-PAGE (Lane 1: uninduced, Lane 2: protein marker, Lane 3: pellet fraction, Lane 4: supernatant fraction). (E) 12% SDS-PAGE of purified aIF1 (Lane 1: protein marker, Lane 2-8: purified aIF1 protein). (F) Concentrated aIF1 (Lane 1: protein marker, Lane 2: concentrated protein).

### 4.2.2 CRYSTALLIZATION, X-RAY INTENSITY DATA COLLECTION AND DATA PROCESSING

An initial screening of the purified aIF1 protein was performed using microbatch-under-oil technique by mixing 2 µl of protein and 2 µl of crystallization buffers available in Crystal Screen, Crystal Screen 2 and Poly(ethylene glycol)/Ion kits from Hampton Research at 4 and 20°C temperatures. Diffraction quality crystals of aIF1 (Figure 4.2A) were obtained in 0.2 M lithium nitrate and 20% poly(ethylene glycol) (PEG) 3350 at 20°C. X-ray intensity diffraction data (Figure 4.2B) was collected at -173°C using the home source Rigaku MicroMax-007 HF diffractometer (operated at 40 kV and 30 mA) and R-Axis IV++ imaging-plate detector available at the Central Instrument Facility (CIF) of Indian Institute of Technology Guwahati, India. The crystal to detector distance

was maintained at 80 mm. The crystals diffracted to a resolution of 1.5 Å. The diffraction data were processed and scaled using the programs iMosflm (Battye et al., 2011; Evans and Murshudov, 2013) and Aimless, respectively, embedded in the CCP4 package (Winn et al., 2011). The intensities were converted to structure factors using the program CTRUNCATE available in CCP4 package.

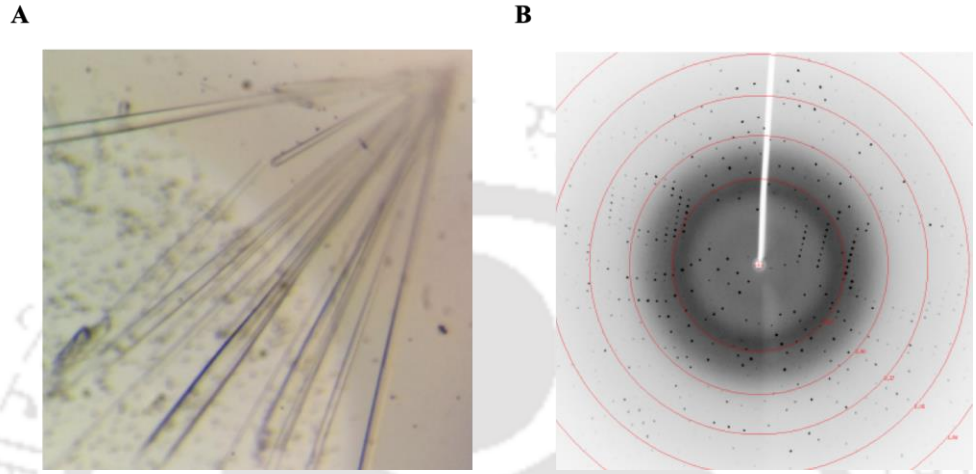



**Figure 4.2. Crystallization of aIF1.** (A) Protein crystals and (B) X-ray diffraction pattern of aIF1.

### 4.2.3 STRUCTURE SOLUTION AND MODEL BUILDING

The structure of aIF1 from *P. horikoshii* OT3 was solved by molecular replacement (MR) method using the program Phaser (McCoy et al., 2007). The three-dimensional atomic coordinates of aIF1 from *M. jannaschii* (PDB id: 4MO0) was used as the search model (sequence identity: 48%). A total of 5% of the reflections were set aside for the calculation of $R_{free}$ (Brunger, 1992). All the residues were first built using the program COOT (Emsley and Cowtan, 2004). Subsequently, water molecules were located and added from the difference electron density maps. All refinements were carried out using the program Refmac5 (Vagin et al, 2004) embedded in the program CCP4. The programs PROCHECK (Laskowski et al., 1993) and MolProbity (Chen et al., 2010) were used to check and validate the quality of the refined models. The three-dimensional atomic coordinates and the structure factors of the structure have been deposited in the RCSB Protein Data Bank (Berman et al., 2000) with a PDB id 5ZCY. The refinement statistics and other details of the refined models are provided in Table 4.1.

**Table 4.1.** Data collection, processing and refinement statistics of aIF1. The values provided in parenthesis are for the last resolution shell.

| | **aIF1** |
|---|---|
| Wavelength (Å) | 1.5418 |
| Temperature (K) | 100 |
| Space group | $P6_1$ |
| Unit-cell parameter (Å, °) | $a=b=53.58$, $c=56.94$, $\alpha=\beta=90$, $\gamma=120$ |
| Resolution (Å) | 35.97-1.50 (1.52-1.50) |
| No. of observed reflections | 162563 (7117) |
| No. of unique reflections | 14801 (691) |
| Mn(I) CC(1/2) | 1.000 (0.915) |
| Completeness (%) | 98.1 (95.9) |
| $V_M$ ($Å^3$ $Da^{-1}$) | 2.06 |
| Solvent content (%) | 40.33 |
| Mosaicity (°) | 0.373 |
| $I/\sigma(I)$ | 23.4 (3.4) |
| $R_{merge}$ (%) | 5.4 (51.4) |
| $R_{pim}$ (%) | 1.7 (16.5) |
| $R_{meas}$ (%) | 5.6 (54) |
| Multiplicity | 11.0 (10.3) |
| $R_{work}/R_{free}$ (%) | 17.38/20.46 |
| **Protein model** | |
| No. of subunits in ASU | 1 |

| | |
|---|---|
| Protein atoms | 672 |
| Water molecules | 82 |
| Others | 2 |
| **Deviation from ideal geometry** | |
| Bond length (Å) | 0.028 |
| Bond angles (°) | 2.578 |
| **Average *B*-factor (Å$^2$)** | |
| Protein atoms | 18.34 |
| Water molecules | 30.99 |
| Others | 20.17 |
| **Ramachandran plot (%)** | |
| Favored | 98.77 |
| Allowed | 1.23 |
| **PDB id** | 5ZCY |

### 4.2.4 SEQUENCE AND STRUCTURE ANALYSIS

The amino acid sequences of proteins used in this study were downloaded from UniProtKB database (UniProt Consortium, 2018). The homology searches were performed using the web tool BLAST (Altschul, 1990). For the structure-based phylogenetic tree, a distance matrix was generated for all the proteins by calculating their RMSD in PyMOL. Subsequently, a dendrogram was generated in Newick file format from the distance matrix using the program dendroUPGMA (Garcia-Vallvé et al., 1999). The phylogenetic tree was generated using the Newick file in the program MEGA7 (Kumar et al., 2016). The structure-based sequence alignment was performed using the program PROMALS3D (Pei et al., 2008) and the aligned sequences were further decorated using the online tool ESPript (Gouet et al., 2003). The electrostatic charge distribution of each protein was generated using the APBS plugin of the molecular

visualization tool PyMOL (Baker et al., 2001). All the figures were generated using the program PyMOL.

## 4.3 RESULTS

### 4.3.1 THE OVERALL STRUCTURE

The three-dimensional structure of the monomeric aIF1 encoded by PH1771.1 from *P. horikoshii* OT3 was solved by MR method using the atomic coordinates of aIF1 from *M. jannaschii* (PDB id: 4MO0). The protein aIF1 contains an α+β structure and is composed of a five-stranded antiparallel β-sheets ($\beta_1$–$\beta_5$) with two α-helices ($\alpha_1$ and $\alpha_2$) located on same side of the β-sheet (Figure 4.3). The four stranded β-sheets connected by flexible loops forms a 'RNA-binding platform' similar to other RNA-binding proteins. The N-terminal residues (1–16) of the protein, however, are unstructured which is evident from the lack of electron density corresponding to these amino acid residues. Structural homology search using the web server DALI (Holm and Rosenström, 2010) reveals its closest homologs to be aIF1 from *M. jannaschii* (PDB id: 4MO0, Z-score: 14.3, RMSD: 1.8 Å) and *P. abyssi* (PDB id: 5JB3, Z-score: 13.6, RMSD: 1.1 Å) followed by eIF1 from *T. thermophila* (PDB id: 4V5O, Z-score: 12.6, RMSD: 2.2 Å), *Lachancea kluyveri* (PDB id: 4UER, Z-score: 12.2, RMSD: 1.6 Å), *S. cerevisiae* (PDB id: 3J80, Z-score: 11.8, RMSD: 1.8 Å) and *H. sapiens* (PDB id: 4KZY, Z-score: 10.3, RMSD: 1.8 Å) and YciH from *E. coli* (PDB id: 1D1R, Z-score: 8.3, RMSD: 3.0 Å).

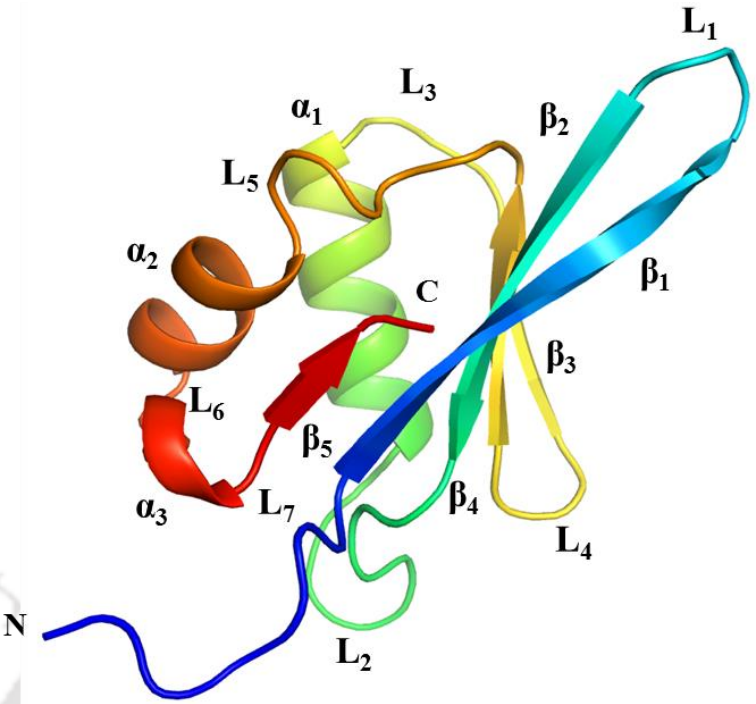



**Figure 4.3. Crystal structure of aIF1 represented as cartoon model.** The secondary structure elements have been labeled.

### 4.3.2 THE ACIDIC AND BASIC LOOPS OF aIF1

Similar to eIF1 and YciH, aIF1 also attains an overall core topology of ββαββαβ. Nonetheless, there are certain striking differences in the structure of the three proteins. The two major differences lie in the loop structure connecting the two pairs of antiparallel β-strands. The loop, $L_1$ connecting $\beta_1$ and $\beta_2$ in aIF1 is equivalent to the corresponding basic loop in eIF1 while considerably shorter than that of YciH. Another notable difference lies in the acidic hairpin loop connecting the second pair of β-strands. Although functionally crucial in eIF1, this acidic loop is completely absent in aIF1 (Figure 4.4A). Owing to the fact that these three structurally similar proteins share low sequence similarity among themselves, a structure-based phylogeny tree was generated. The functionally analogous bacterial translation initiation factor IF3, which lacks sequence and structural similarity with aIF1, eIF1 and YciH was used as an outgroup. In the structure-based rooted phylogeny tree, aIF1 clusters with eIF1 from lower eukaryotes indicating its structural closeness to eIF1. The phylogeny tree further indicates that aIF1, eIF1 and YciH have evolved from an ancestral protein having a common core structure (Figure 4.4B). In spite of possessing a significant structural similarity, aIF1 and eIF1

share a low sequence similarity. Thus, to figure out the differences at the functionally important sites, a structure-based sequence alignment was performed. Of the six basic residues, which are crucial for the functioning of eIF1 and for interacting with 18S rRNA, the protein aIF1 possesses only three residues viz. Lys37 in the loop $L_1$ and Lys56 and Lys59 in the helix $\alpha_1$ (numbering according to eIF1 from *S. cerevisiae*, SceIF1). Notably, the amino acid residues Lys37 and Lys59 reported to be critical for eIF1 interaction to 18S rRNA are conserved in aIF1 as well. However, the remaining three basic amino acid residues Arg33, Arg36 and Lys60 of eIF1 from *S. cerevisiae* are replaced by Ala31, Gly34 and Ala58, respectively, in aIF1. On the other hand, the three amino acid residues, Asp71, Glu73 and Glu76 making up the acidic loop in eIF1 are absolutely absent in aIF1 (Figure 4.4C).

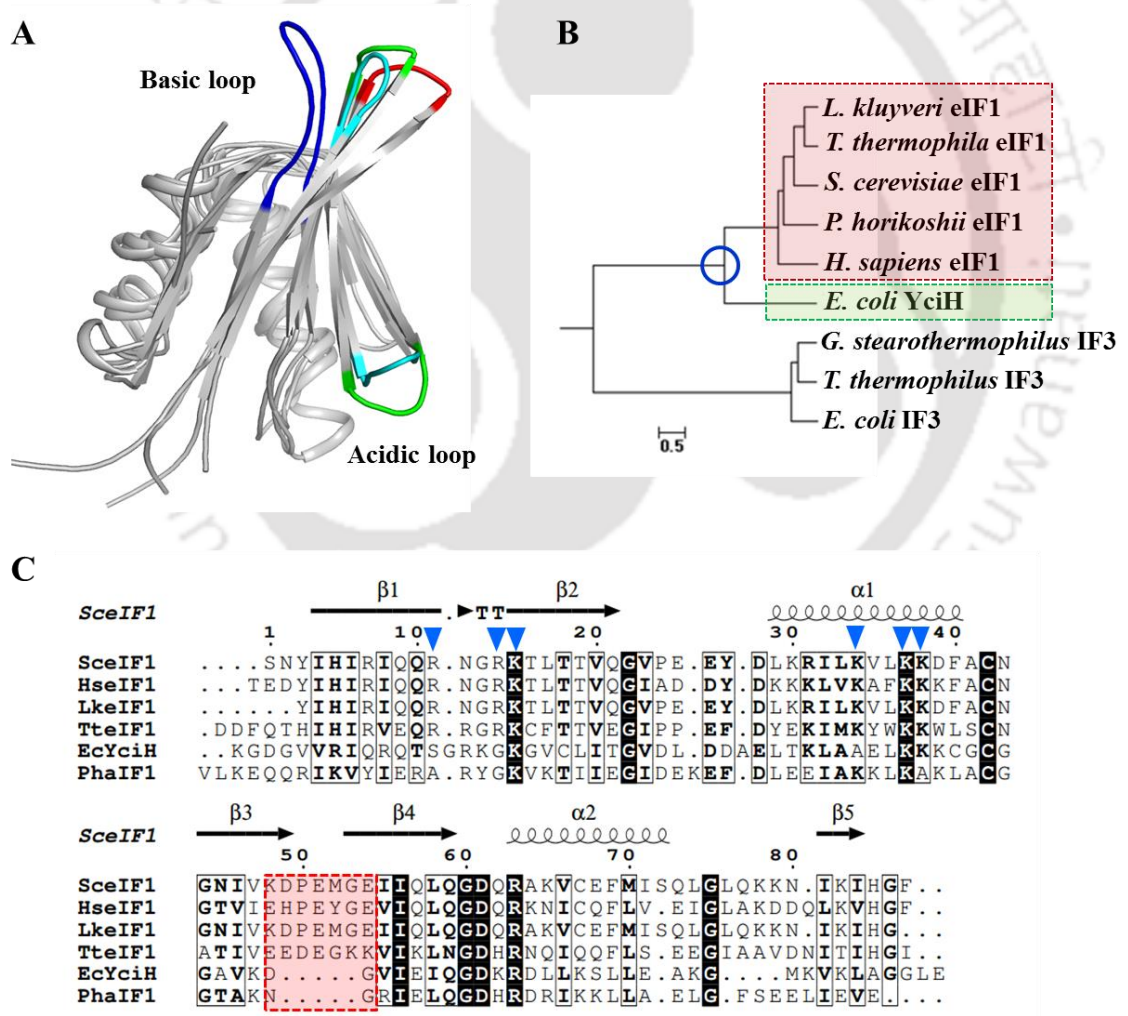



**Figure 4.4. Structural comparison of aIF1, eIF1 and YciH.** (A) Comparison of the basic and acidic loop in aIF1 from *P. horikoshii* (red), eIF1 from *S. cerevisiae* (green,

PDB id: 3J80) and *H. sapiens* (cyan, PDB id: 2IF1) and YciH from *E. coli* (blue, PDB id: 1D1R). (B) Structure-based phylogenetic tree depicts the clustering of aIF1 with eIF1 (pink box) while YciH forms a separate clade (green box). The ancestral core structure is highlighted with blue circle. The scale bar indicates 0.5 amino acid substitutions per single site. (C) A structure-based sequence alignment of eIF1 from *S. cerevisiae* (SceIF1, PDB id: 3J80), *H. sapiens* (HseIF1, PDB id: 2IF1), *L. kluyveri* (LkeIF1, PDB id: 4UER) and *T. thermophila* (TteIF1, PDB id: 4V5O), YciH from *E. coli* (EcYciH, PDB id: 1D1R) and aIF1 from *M. jannaschii* (MjaIF1, PDB id: 4MO0) & *P. horikoshii* (PhaIF1). The basic residues involved in eIF1 and 18S rRNA interaction are highlighted with blue downward arrowheads while the acidic residues conserved in eIF1 protein are highlighted with a pink box. The fully and partially conserved residues are enclosed in black box and shown in bold respectively.

### 4.3.3 THE CHARGE DISTRIBUTION OF aIF1

The electrostatic charge distribution of aIF1 from *P. horikoshii* along with eIF1 from *S. cerevisiae* (PDB id: 3J80) was calculated to identify the differences in the surface charge of the functionally important areas. As anticipated, the regions of eIF1 from *S. cerevisiae* which are involved in interacting with 18S rRNA and in close proximity to tRNA are positively charged. These positively charged regions, thus, provide a platform for the binding of the negatively charged phosphate backbone of the RNA molecules. Similarly, the corresponding regions in aIF1 are also observed to be positively charged indicating that aIF1 would also interact with rRNA and tRNA by means of the same binding surface (Figure 4.5). On the other hand, the solvent-exposed surface of aIF1, which is not involved in interaction either with rRNA or tRNA, is highly negatively charged. Quite interestingly, the corresponding region in eIF1 from *S. cerevisiae* is completely positively charged (Figure 4.5).

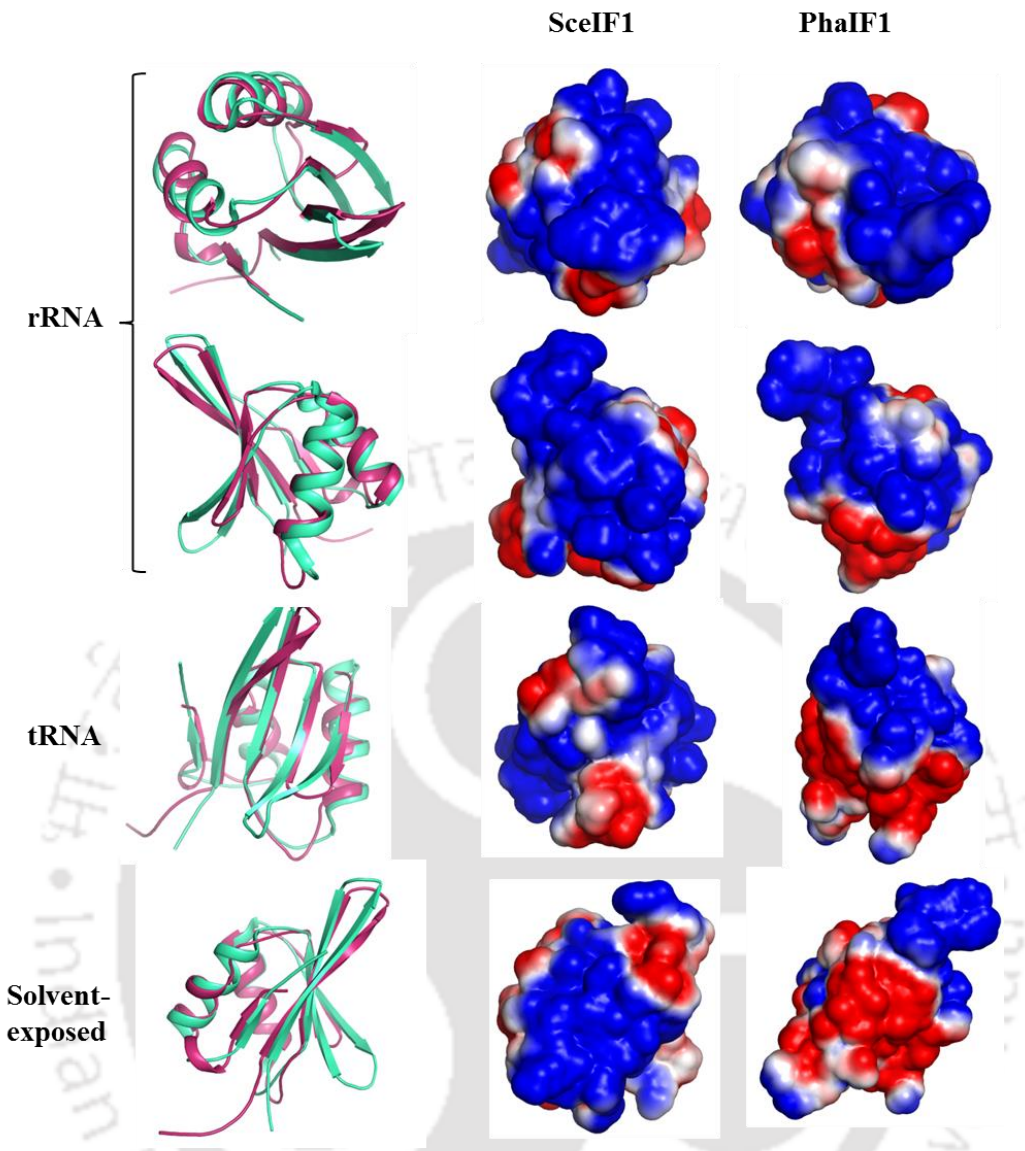



**Figure 4.5. Surface charge distribution of eIF1 and aIF1.** Comparison of the surface electrostatic charge distribution of the protein eIF1 from *S. cerevisiae* (SceIF1, PDB id: 3J80) and aIF1 from *P. horikoshii* (PhaIF1) based on their positions on the small ribosomal subunit. Superimposition of eIF1 (green) and aIF1 (warm pink) based on their rRNA binding, tRNA-contact surface and the solvent-exposed surface. The positively and negatively charged surfaces are shown in blue (+1 kT/e) and red (-1 kT/e), respectively.

## 4.4 DISCUSSION

The fact that aIF1 is functionally analogous to eIF1 has been well established (Hasenöhrl et al., 2006, 2009). A recently solved cryo-EM structure of an archaeal 30S initiation complex with aIF1 bound in a manner similar to eIF1 in 40S initiation complex further accentuates the fact (Coureux et al., 2016). In *P. horikoshii* OT3, an ORF PH1771.1,

encoding aIF1, has been annotated to be a homolog of eIF1 (also known as SUI1 in yeast). Although functionally homologous, aIF1 and eIF1 do not display very high sequence similarity. Moreover, aIF1 also displays an equivalent sequence similarity with bacterial protein, YciH. The resemblance among aIF1, eIF1 and YciH might be due to the fact that each of these proteins has evolved from a common ancestor. Thus, over the years, these proteins might have evolved further to perform 'domain-specific' functions. The elucidation of the three-dimensional crystal structure of aIF1 aided in implementing a structural comparison with eIF1 as well as YciH. Although, aIF1 shares equivalent sequence similarity with both eIF1 and YciH, more structural homology could be perceived between aIF1 and eIF1. In addition, a structure-based phylogenetic analysis demonstrates that the structure of aIF1 is structurally more related to eIF1 than to YciH. This can be anticipated as archaeal and eukaryal translation initiation apparatus are usually deemed to be homologous (Kyrpides and Woese, 1998a; Bell and Jackson, 1998; Londei, 2005; Benelli and Londei, 2011). The proteins aIF1, eIF1 and YciH share an overall similar topology owing to the fact that all of them have evolved from a common ancestral protein. However, prominent differences exist at the functionally relevant loop regions connecting the antiparallel β-strands. Unlike YciH, aIF1 contains a short $\beta_1$-$\beta_2$ loop similar to eIF1, although with fewer conserved basic residues. Even though aIF1 does not possess all the three basic residues present in the $\beta_1$-$\beta_2$ loop of eIF1, the residue Lys37, which makes the most critical interaction with 18S rRNA, is well conserved. Besides the basic residues in $\beta_1$-$\beta_2$ loop, three basic residues Lys56, Lys59 and Lys60 in helix $\alpha_1$ makes critical contact with the 18S rRNA (Martin-Marcos et al., 2013). Of these three residues, Lys56 and Lys59 (which makes an important contact between eIF1 and 40S small ribosomal subunit) are well conserved in the protein aIF1. On the other hand, Arg33 and Arg36 in $\beta_1$–$\beta_2$ loop and Lys60 in helix $\alpha_1$ are replaced by either alanine or glycine. Substitution of these residues to alanine or glycine is known to confer slow growth (Slg) and/or suppressor of initiation codon mutation (Sui) phenotypes. None of the substitutions are capable of completely shutting down the translation initiation process and could be mitigated by over expressing the eIF1 mutant proteins (Martin-Marcos et al., 2013). It implies that aIF1 possesses only those residues essential for anchoring of aIF1 to the 30S small ribosomal subunit. Those substitutions which lead to an increased initiation rate at non-AUG codons might be counteracted by the Shine-Dalgarno (SD) sequence or 'cryptic' SD sequences present in mRNA of archaea which would promote recognition of correct start codon (Betlach et al., 1984; Condò et al.,

1999; Dennis, 1997; Coureux et al., 2016). The acidic loop plays a pivotal role in the dissociation of eIF1 from the 40S small ribosomal subunit by creating repulsive forces between the loop and the D-stem loop of Met-$tRNA_i^{Met}$ (Hussain et al., 2014). Complete absence of the acidic loop in aIF1 hints towards the lack of this repulsive force that would aid in its dissociation from 30S small ribosomal subunit. This might be the reason for the presence of a fewer conserved basic residues in helix $\alpha_1$ and $\beta_1$-$\beta_2$ loop which would result in a weaker interaction between aIF1 and 30S ribosomal subunit, thus eliciting the requirement of an acidic loop. Similar to eIF1, the regions of aIF1 involved in interaction with the rRNA as well as the region which remains in close proximity to tRNA are positively charged. The presence of positively charged residues in the corresponding region of aIF1 confirms that it would bind to 30S small ribosomal subunit in a manner similar to that of eIF1 (Coureux et al., 2016). Binding of aIF1 to the P-site of 30S small ribosomal subunit would inhibit the interaction between the large and small subunits of ribosome. Furthermore, the negatively charged solvent-accessible region of aIF1 might further repel the negatively charged phosphate backbone of 23S rRNA of 50S large ribosomal subunit. Since, archaea lack a homolog of the ribosome anti-association factor eIF3 (Lomakin et al., 2003), the initiation factor aIF1 might itself inhibit the premature association of the ribosomal subunits.

## 4.5 CONCLUSION

From this chapter, it can be concluded that aIF1 is structurally homologous to eIF1. A thorough structural comparison with eIF1 has led towards the identification of structural features distinct from eIF1 which allows aIF1 protein to function slightly differently from eIF1 as well as compensate for the lack of an eIF3 homolog in archaea. Thus, aIFs might possess unique structural attributes that permit the accomplishment of a ‘eukaryote-like’ translation initiation process in a prokaryotic cell.

# CHAPTER 5

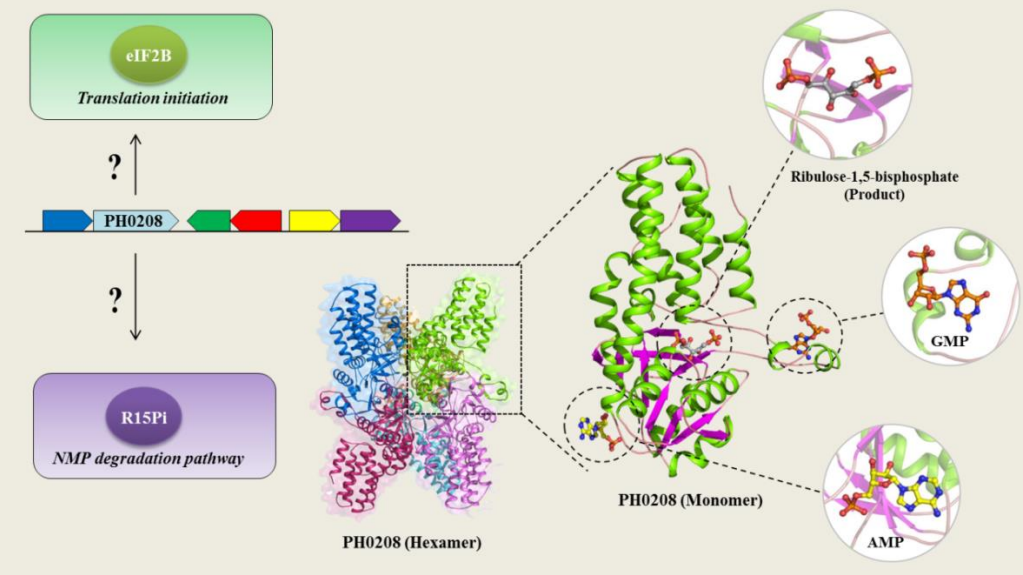
## Structural implications of the enzyme ribose-1,5-bisphosphate isomerase

This chapter has been published as:

Gogoi, P. and Kanaujia, S.P., 2018. A presumed homologue of the regulatory subunits of eIF2B functions as ribose-1,5-bisphosphate isomerase in *Pyrococcus horikoshii* OT3. *Scientific Reports*, *8*(1), 1891.

# ABSTRACT

The homologs of the regulatory subunits of eIF2B are assumed to be present in archaea. Likewise, an open reading frame (ORF) PH0208 in *P. horikoshii* OT3 have been proposed to encode one of the homologs of regulatory subunits of eIF2B. However, PH0208 protein also shares sequence similarity with a functionally non-related enzyme, ribose-1,5-bisphosphate isomerase (R15Pi), involved in conversion of ribose-1,5-bisphosphate (R15P) to ribulose-1,5-bisphosphate (RuBP) in an AMP-dependent manner. In this chapter, we have determined the crystal structure of PH0208 protein in order to decipher its true function. Although structurally similar to the regulatory subunits of eIF2B, the ability to bind R15P and RuBP suggests that PH0208 would function as R15Pi. Additionally, this study for the first time reports the binding sites of AMP and GMP in R15Pi. The AMP binding site in PH0208 protein clarified the role of AMP in providing structural stability to R15Pi. The binding of GMP to the 'AMP binding site' in addition to its own binding site indicates that GMP might also execute a similar function, though with less specificity. Furthermore, we have utilized the resemblance between PH0208 and the regulatory subunits of eIF2B to propose a model for the regulatory mechanism of eIF2B in eukaryotes.




## 5.1 INTRODUCTION

The resemblance of the protein translation initiation processes between eukaryotes and archaea has long been postulated (Kyrpides and Woese, 1998a). This notion has been well supported by the availability of homologous components, especially the translation initiation factors (IFs) in both the domains of life (Bell and Jackson, 1998; Londei, 2005; Benelli and Londei, 2011). With the advent of the genomic DNA sequence of hyperthermophillic archaeon, *M. jannaschii*, a considerably substantial number of homologs of eukaryotic translation IFs (viz. eIF1, eIF1A, eIF2, eIF4A, eIF5A, eIF5B and eIF6) have been identified (Bult et al., 1996; Kyrpides and Woese, 1998a, 1998b; Benelli and Londei, 2011; Gäbel et al., 2013). The availability of these translation IFs in archaea reinforces the fact that archaea imitates the eukaryal mechanism of translation initiation.

This apparent understanding of the archaeal translation initiation, however, remains questionable owing to the uncertainty concerning the presence of a functional homolog of eIF2B in archaea. In eukaryotes, eIF2B functions as a guanine exchange factor (GEF) for eukaryotic translation initiation factor 2 (eIF2) and is comprised of two copies each of α, β, δ, γ and ε subunits (Price and Proud, 1994). The α-, β- and δ-subunits form the regulatory subcomplex while γ- and ε-subunits make up the catalytic subcomplex (Gordiyenko et al., 2014; Wortham et al., 2014). The regulatory subunits of eIF2B execute its function by interacting with the α-subunit of eIF2 via the N-terminal domains

(NTD) (Kuhle et al., 2015; Kashiwagi et al., 2016). However, the actual mechanism underlying the regulatory mechanism of eIF2B remains unspecified. Earlier studies have suggested that under stressed conditions, eIF2α is phosphorylated to form eIF2αP which shows a higher affinity towards eIF2B. Thus, the substrate, eIF2α is converted into a competitive inhibitor (eIF2αP) of eIF2B which ultimately results in a global downregulation of protein synthesis (Hinnebusch, 1993; Sudhakar et al., 2000; Krishnamoorthy et al., 2001; Hinnebusch, 2005).

In archaea, the homologs of only the regulatory subunits of eIF2B have been reported to be present and not the catalytic subunits (Dennis, 1997; Kyrpides and Woese, 1998b). However, the regulatory subcomplex alone would not confer full functionality to eIF2B, as the catalytic subcomplex accomplishes the crucial function of GTP exchange (Pavitt et al., 1998; Gomez et al., 2002). In *P. horikoshii* OT3, a hyperthermophillic archaeon, three open reading frames (ORFs) (PH0440, PH0702 and PH0208) were postulated to encode aIF2Bα, aIF2Bβ and aIF2Bδ, respectively (Kakuta et al., 2004). This assumption was merely based on sequence similarity to the α-, β- and δ-subunits of eIF2B and lacked convincing evidence. Later, based on an in-depth *in silico* analysis, we have demonstrated that the protein PH0208 encodes an enzyme, ribose-1,5-bisphosphate isomerase (R15Pi) with all the essential amino acid residues conserved (Gogoi et al., 2016). The enzyme, R15Pi was recently identified in *T. kodakarensis* (TkR15Pi) and is known to play an indispensable role during the nucleoside 5′-monophosphate (NMP) degradation pathway available only in archaea (Nakamura et al., 2012). The NMP degradation pathway partially compensates for the lack of a pentose phosphate pathway in archaea and ultimately leads to the formation of 3-phosphoglyceric acid (3-PGA), a crucial intermediate in glycolysis and Calvin cycle (Aono et al., 2012, 2015). In NMP degradation pathway, R15Pi catalyzes the conversion of ribose-1,5-bisphosphate (R15P, substrate) into ribulose-1,5-bisphosphate (RuBP, product) in an adenosine 5′-monophosphate (AMP)-dependent manner. AMP acts as an activator of R15Pi and in its absence, the enzymatic activity has been found to decline dramatically. Apart from AMP, guanosine 5′-monophosphate (GMP) also elevates the activity of R15Pi, though to a lesser extent as compared to AMP. While, neither cytosine 5′-monophosphate (CMP) nor thymidine 5′-monophosphate (TMP) possesses the ability to increase the enzymatic activity of R15Pi (Aono et al., 2012). The three-dimensional structure of R15Pi from *T. kodakarensis* has been solved, both in the presence of the substrate and the product

(Nakamura et al., 2012). However, due to the unavailability of an AMP/GMP-bound structure, the exact role of purine nucleotides still remains obscure.

In this chapter, we attempted to abolish the uncertainty regarding the (un)availability of a homolog of the regulatory subunits of eIF2B in archaea by performing a meticulous characterization of PH0208 protein. We have determined the three-dimensional structure of PH0208 protein in complex with the cognate substrate as well as product of the enzyme, R15Pi. To understand the role of purine nucleotides in structure and activity, crystal structures of PH0208 protein (wild type and mutants) in complex with only AMP, only GMP and both AMP & GMP were also deciphered. Furthermore, a comparison between the PH0208 protein and the regulatory subunits of eIF2B has been drawn to obtain an in-depth understanding of the (dis)similarities that lies between them. Finally, a model depicting the regulatory mechanism of eIF2B in eukaryotes has been proposed based on the similitude perceived between PH0208 protein and the regulatory subunits of eIF2B.

## 5.2 MATERIALS AND METHODS

### 5.2.1 CLONING, OVER EXPRESSION AND PROTEIN PURIFICATION

The gene encoding PH0208 (R15Pi from *P. horikoshii*, PhR15Pi) was amplified (Figure 5.1A) using PCR from the genome of *P. horikoshii* OT3 using the forward primer 5′-CGGT **CATATG**CACCATCATCATCATCATGTGGGAGCCATGATAG-3′ containing NdeI restriction site (bold) followed by a 6xHis-tag (underlined) and a reverse primer, 5′-CGA**CTCGAG**CTAATCTTCCCATGGTTCTTTATACTTTAATGCCC-3′ containing XhoI restriction site (bold). The amplified gene was inserted into pET-22b(+) excised with the same restriction enzymes. Subsequently, the clones were confirmed by double digestion using the restriction enzymes, NdeI and XhoI by incubating the vectors at 37°C for two hours (Figure 5.1B). In addition, the clone was confirmed by DNA sequencing. The resulting vector was further used as a template to construct the mutants, C135S, C135A and D204N using Q5 Site-Directed Mutagenesis Kit (New England Biolabs). The primers used for the construction of the mutants were as follows: C135SF/R: 5′-

ATGACACAT**TCT**CACAGCAAAGC-3′/5′-TATTATATCCCCATCCTCAATTC-3′, C135A-F/R: 5′-AATGACACAT**GCT**CACAGCAAAGCC-3′/5′-ATTATATCCCCATCCTCAATTC-3′, D204N-F/R: 5′-TATGGGAGCT**AAT**TCTATAACAGC-3′/5′-ACAACCTTATCAGTCATCTTC-3′ (point mutations are indicated in bold). The plasmids of all the derived mutants were isolated (Figure 5.1C) and confirmed by sequencing before purification of the induced recombinant proteins.

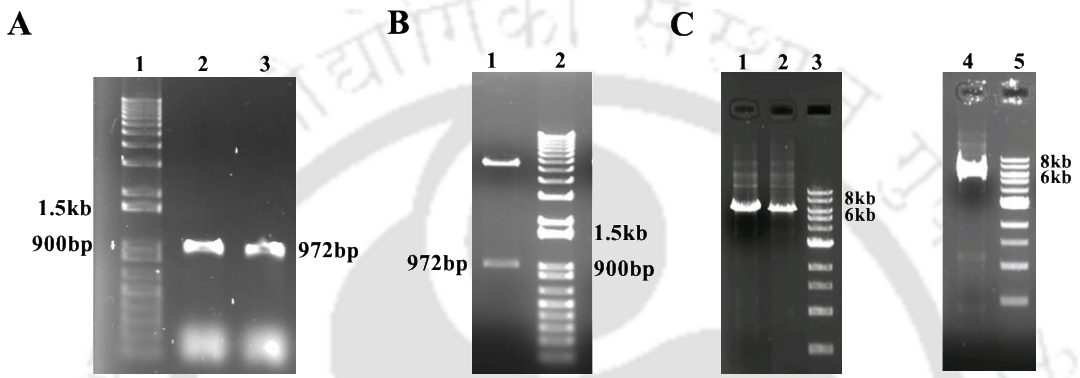



**Figure 5.1. Cloning and site-directed mutagenesis of PhR15Pi analysed in 0.8% agarose gel.** (A) PCR amplification of PhR15Pi (Lane 1: DNA ladder, Lane 2 & 3: PCR amplified PhR15Pi). (B) Clone confirmation using the double digestion of the plasmid pET-22b(+) containing PhR15Pi (Lane 1: positive clone, Lane 2: DNA ladder). (C) PCR amplification of the PhR15Pi clone with mutant specific primers (Lane 1: pET-22b(+) containing the mutant gene encoding PhR15Pi-C135S, Lane 2: pET-22b(+) containing the mutant gene encoding PhR15Pi-D204N, Lane 3: DNA ladder, Lane 4: pET-22b(+) containing the mutant gene encoding PhR15Pi-C135A, Lane 5: DNA ladder).

For over expression, the recombinant plasmid constructs PhR15Pi-wild type (WT), PhR15Pi-C135S, PhR15Pi-C135A and PhR15Pi-D204N were transformed into expression systems such as *E. coli* BL21-CodonPlus (DE3)-RIL and *E. coli* Rosetta (DE3) competent cells. The transformed competent *E. coli* cells were then grown at 37°C in LB supplemented with 100 μg $ml^{-1}$ ampicillin and 34 μg $ml^{-1}$ chloramphenicol. The over expression of PhR15Pi-WT and the mutant proteins PhR15Pi-C135S, PhR15Pi-D204N and PhR15Pi-C135A was optimized using varying concentrations (0.1, 0.5 and 1.0 mM) of isopropyl β-D-1-thiogalactopyranoside (IPTG). Optimum expression of the proteins was obtained by inducing the cells with 1 mM IPTG (Figure 5.2A-D).

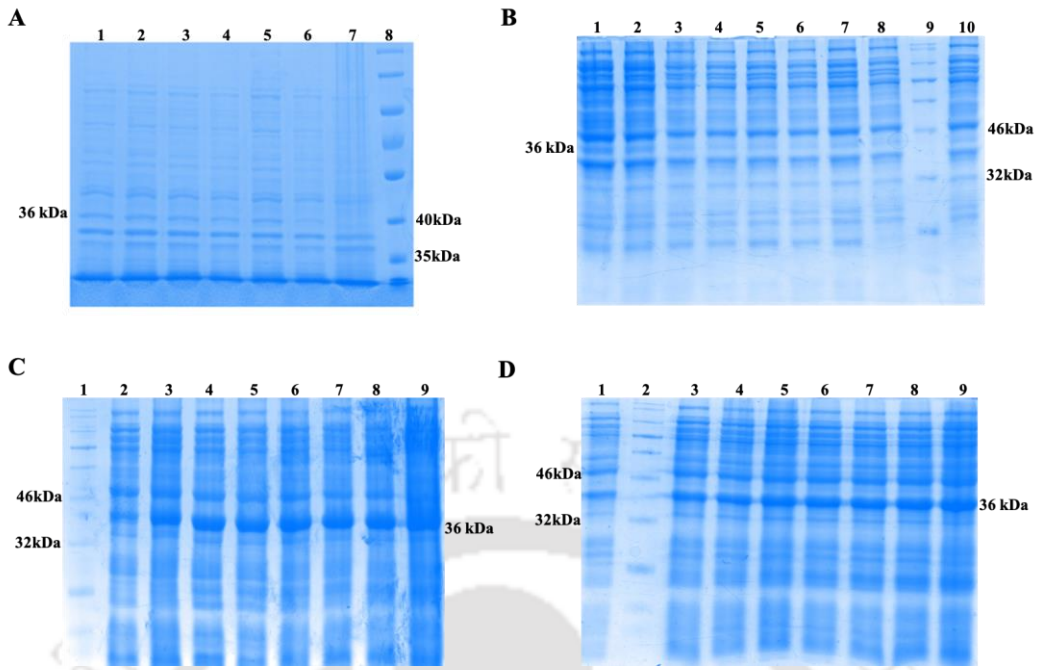



**Figure 5.2. Over expression of PhR15Pi-WT and mutant proteins PhR15Pi-C135S, PhR15Pi-D204N & PhR15Pi-C135A analysed in 12% SDS-PAGE.** (A) Over expression of PhR15Pi at 1 mM IPTG at different time intervals (Lane 1: 6$^{th}$ hr, Lane 2: 5$^{th}$ hr, Lane 3: 4$^{th}$ hr, Lane 4: 3$^{rd}$ hr, Lane 5: 2$^{nd}$ hr, Lane 6: 1$^{st}$ hr, Lane 7: uninduced, Lane 8: protein marker). (B) Over expression of PhR15Pi-C135S mutant at 1 mM IPTG at different time intervals (Lane 1: 8$^{th}$ hr, Lane 2: 7$^{th}$ hr, Lane 3: 6$^{th}$ hr, Lane 4: 5$^{th}$ hr, Lane 5: 4$^{th}$ hr, Lane 6: 3$^{rd}$ hr, Lane 7: 2$^{nd}$ hr, Lane 8: 1$^{st}$ hr, Lane 9: protein marker, Lane 10: uninduced). (C) Over expression of PhR15Pi-D204N mutant at 1 mM IPTG at different time intervals (Lane 1: protein marker, Lane 2: uninduced, Lane 3: 1$^{st}$ hr, Lane 4: 2$^{nd}$ hr, Lane 5: 3$^{rd}$ hr, Lane 6: 4$^{th}$ hr, Lane 7: 5$^{th}$ hr, Lane 8: 6$^{th}$ hr). (D) Over expression of PhR15Pi-C135A mutant at 1 mM IPTG at different time intervals (Lane 1: uninduecd, Lane 2: protein marker, Lane 3: 1$^{st}$ hr, Lane 4: 2$^{nd}$ hr, Lane 5: 3$^{rd}$ hr, Lane 6: 4$^{th}$ hr, Lane 7: 5$^{th}$ hr, Lane 8: 6$^{th}$ hr).

The protein was obtained in pellet fraction when over expression of the protein was induced with 1 mM IPTG at 37°C. Thus, to obtain the protein in the soluble fraction, various optimizations such a varying IPTG concentrations, temperatures, incubation time, culture media, use of buffers containing triton X-100, glycerol etc. were performed. The soluble form of the protein was finally obtained by inducing the protein with 1 mM IPTG at 30°C and adding 3% ethanol at the time of induction and the cells were grown for 12 hrs (Figure 5.3A-D). For enhancing the yield of the recombinant protein, Terrific broth (TB) instead of LB was used.

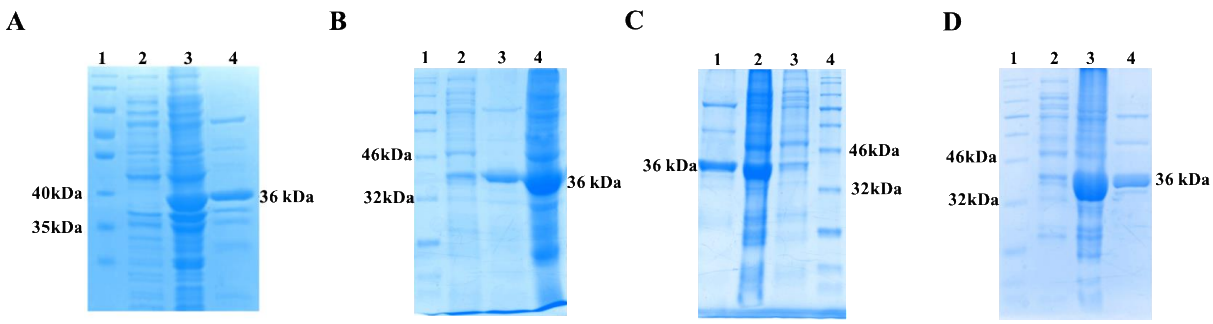



**Figure 5.3. Solubility test for PhR15Pi-WT and mutant proteins PhR15Pi-C135S, PhR15Pi-D204N & PhR15Pi-C135A analysed in 12% SDS-PAGE.** (A) Solubility test for the over expressed PhR15Pi-WT (Lane 1: protein marker, Lane 2: uninduced, Lane 3: pellet fraction, Lane 4: supernatant fraction). (B) Solubility test for the over expressed PhR15Pi-C135S protein (Lane 1: protein marker, Lane 2: uninduced, Lane 3: supernatant fraction, Lane 4: pellet fraction). (C) Solubility test for the over expressed PhR15Pi-D204N protein (Lane 1: supernatant fraction, Lane 2: pellet fraction, Lane 3: uninduced, Lane 4: protein marker).

For purifying the protein, the cells were harvested by centrifugation and resuspended in binding buffer containing 20 mM Tris-HCl pH 8.0, 150 mM KCl, 5 mM imidazole, 1 mM phenylmethylsulfonyl fluoride (PMSF), 3 mM β-mercaptoethanol (β-ME) and 0.2 mg $ml^{-1}$ lysozyme. The cells were then disrupted using sonicator and the cell lysate was incubated at 90°C for 13 min to remove the thermolabile proteins deriving from the host cells. The cell lysate was then centrifuged at 12,000 rpm for 40 min at 4°C. The supernatant fraction was applied to Ni-NTA affinity column containing Ni-NTA agarose pre-equilibrated with the binding buffer without lysozyme and incubated for two hours. The column was washed with five column-volume each of wash buffer A (20 mM Tris-HCl pH 8.0, 150 mM KCl, 10 mM imidazole, 1 mM PMSF and 3 mM β-ME) and wash buffer B (20 mM Tris-HCl pH 8.0, 150 mM KCl, 20 mM imidazole, 1 mM PMSF and 3 mM β-ME). The protein was eluted with 250 mM imidazole in 20 mM Tris-HCl pH 8.0, 150 mM KCl and 1 mM PMSF (Figure 5.4A-D). The eluted fractions were collected and step-wise dialysis was performed against 20 mM Tris-HCl pH 8.0 and 150 mM KCl to remove imidazole. The dialyzed proteins were concentrated (PhR15Pi-WT, PhR15Pi-C135S, PhR15Pi-C135A & PhR15Pi-D204N- 9.5, 9.0, 7.0 & 7.0 mg $ml^{-1}$, respectively) using Amicon Ultra centrifugal filter unit (Milipore) and Vivaspin turbo 15 (Sartorius) with a molecular weight cutoff of 10 kDa (Figure 5.4E-H).

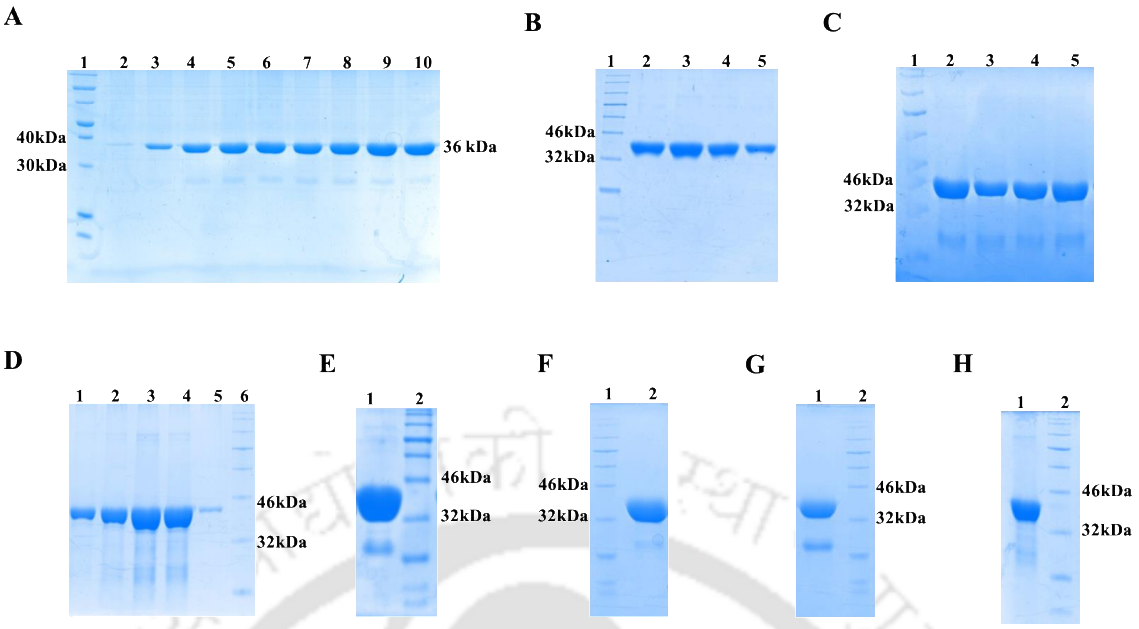



**Figure 5.4. Purification of PhR15Pi-WT and mutant proteins PhR15Pi-C135S, PhR15Pi-D204N & PhR15Pi-C135A.** 12% SDS-PAGE of purified (A) PhR15Pi (Lane 1: protein marker, Lane 2-10: purified PhR15Pi-WT), (B) PhR15Pi-C135S (Lane 1: protein marker, Lane 2-5: purified protein), (C) PhR15Pi-D204N (Lane 1: protein marker, Lane 2-5: purified protein) and (D) PhR15Pi-C135A (Lane 1-5: purified protein, Lane 6: protein marker) and concentrated (E) PhR15Pi (Lane 1: concentrated protein, Lane 2: protein marker), (F) PhR15Pi-C135S (lane 1: protein marker, lane 2: concentrated protein), (G) PhR15Pi-D204N (Lane 1: concentrated protein, Lane 2: protein marker) and (H) PhR15Pi-C135A (Lane 1: concentrated protein, Lane 2: protein marker).

### 5.2.2 PROTEIN CRYSTALLIZATION

To obtain crystals of PhR15Pi-WT, an initial screening was performed by mixing 2 µl of protein and 2 µl of crystallization condition available in Crystal Screen, Crystal Screen 2 and PEG/Ion kits from Hampton Research at 4 and 20°C by the microbatch-under-oil technique. Although, crystal hits were obtained in many conditions, most of the crystals formed were very small in size which could not be utilized further (Figure 5.5A). Further refinement of one of the conditions containing 2.0 M NaCl and 10% (w/v) PEG 6000 using hanging-drop vapor-diffusion technique in 24-well plate was performed to obtain better crystals. However, no improvement could be obtained in the crystals formed. To improve the crystals, the substrate (R15P) was added to the crystallization drop, which was equilibrated against 500 µl of 0.7 M NaCl and 3% PEG 6000 (Figure 5.5B, C). Yet, the crystals formed diffracted to a resolution of ~7 Å only (Figure 5.5D).

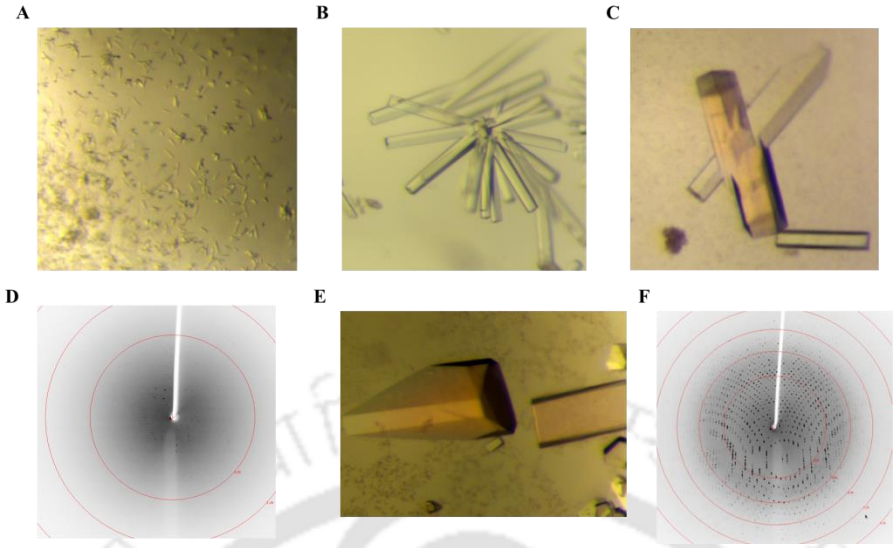



**Figure 5.5. Crystal optimization of PhR15Pi.** Protein crystals of PhR15Pi in the (A) absence and (B, C) presence of R15P. (D) X-ray diffraction pattern of PhR15Pi. Optimized (E) protein crystals and (F) diffraction pattern of PhR15Pi.

Even though, further optimization of the buffer condition resulted in better diffraction, the diffraction data obtained had high mosaicity. In order to get rid of mosaicity, various additives were used in the crystallization conditions. Addition of 20–30% of 2-methyl-2,4-pentanediol (MPD) to the reservoir buffer finally resulted in better crystals which were further used for data collection (Figure 5.5E, F). All the crystals of PhR15Pi-WT and mutant proteins PhR15Pi-C135S & PhR15Pi-D204N in the presence of substrates as well as nucleotides were grown in the same condition (Figure 5.6A-R). For co-crystallization of PhR15Pi-WT, PhR15Pi-C135S and PhR15Pi-D204N with different ligands, the final protein concentration was kept at 9.5, 9.0 and 7.0 mg $ml^{-1}$, respectively, in 20 mM Tris-HCl pH 8.0 and 150 mM KCl. The final concentration of R15P in the crystallization drops were 8.33–9 mM while that of AMP, GMP, CMP and TMP were 1.66, 4.5 & 9, 1.66 and 9 mM, respectively. In case of ligand soaking, the concentration of AMP was kept at 5 mM. For obtaining the crystals of ligand bound PhR15Pi-C135S and PhR15Pi-D204N complexes, the reservoir buffer was further supplemented with 0.1–0.3% of low melting (LM) agarose. All the crystals were obtained in hanging-drop vapor-diffusion method at 20°C within a period of 2-7 days.

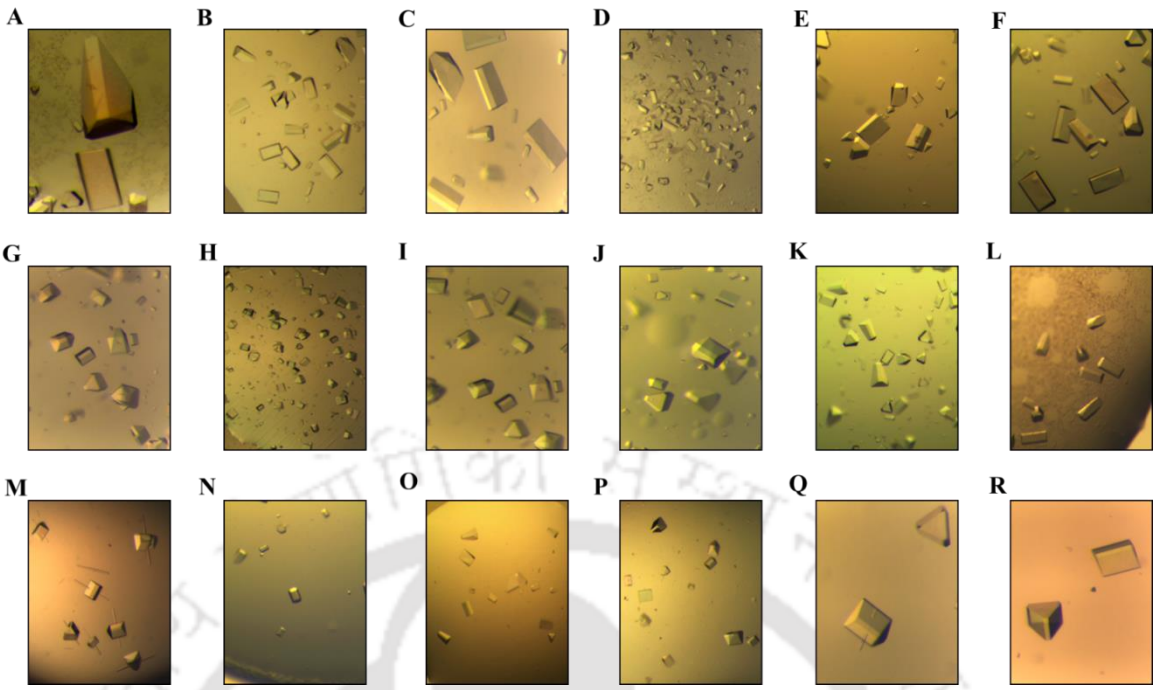



**Figure 5.6. Crystallization of PhR15Pi-WT and mutant proteins PhR15Pi-C135S, PhR15Pi-D204N & PhR15Pi-C135A in complex with purine and pyrimidine nucleotides.** Protein crystal of PhR15Pi-WT in complex with (A) RuBP, (B) AMP, (C) GMP, (D) AMP & GMP, (E) CMP and (F) TMP. Protein crystal of PhR15Pi-C135S in complex with (G) R15P, (H) AMP, (I) GMP, (J) AMP & GMP, (K) CMP and (L) TMP. Protein crystal of PhR15Pi-D204N in complex with (M) R15P, (N) AMP, (O) GMP, (P) AMP & GMP, (Q) CMP and (R) TMP.

### 5.2.3 DATA COLLECTION, PROCESSING AND STRUCTURE DETERMINATION

X-ray intensity diffraction data for all the proteins were collected at -173°C using the home source Rigaku MicroMax-007 HF diffractometer (operated at 40 kV and 30 mA) and R-Axis IV++ imaging-plate detector available at the Central Instrument Facility (CIF) of Indian Institute of Technology Guwahati, India. The crystal to detector distance was maintained at a range of 150-200 mm. The crystals diffracted to a resolution range of 2.2-2.8 Å. The diffraction data were processed and scaled using the programs iMosflm and Aimless, respectively, embedded in the CCP4 package (Battye et al., 2011; Evans and Murshudov, 2013). The intensities were converted to structure factors using the program CTRUNCATE available in CCP4 package. Details of the diffraction data statistics are given in Tables 5.1-5.4. The structure of the PhR15Pi-WT was solved by molecular replacement method (MR) using the program Phaser (McCoy et al., 2007).

The three-dimensional atomic coordinates of R15Pi from *T. kodakarensis* (PDB id: 3VM6) was used as the search model (sequence identity: 86%). A total of 5% of the reflections were set aside for the calculation of $R_{free}$ (Brunger, 1992). The electron density for the product was observed in the active site of the protein molecule, however, water molecules were first located and added from the difference electron density maps. Subsequently, the product molecule was also modeled and refined. All the refinements were carried out using the program Refmac5 embedded in the program CCP4 (Winn et al, 2011; Vagin, 2004). The program COOT was used to build the model (Emsley and Cowtan, 2004). A similar approach to that described above was used to solve and refine the structures of mutant proteins (PhR15Pi-C135S and PhR15Pi-D204N) as well as the ribonucleotide bound complexes of PhR15Pi-WT and mutant proteins. The structure solutions were obtained using the atomic coordinates of the refined model of PhR15Pi-WT for all the subsequent structures both in apo and holo forms. The programs PROCHECK (Laskowski et al., 1993) and MolProbity (Chen et al., 2010) were used to check and validate the quality of the refined models. The web server PSAP (Balamurugan et al., 2007) was extensively used during the modeling building and analyses of the proteins. The three-dimensional atomic coordinates and the structure factors of all the structures (PDB ids: 5YFJ, 5YFS, 5YFT, 5YFU, 5YFV, 5YFW, 5YFX, 5YG5, 5YG6, 5YG7, 5YG8, 5YG9 and 5YGA) have been deposited in the RCSB PDB (Berman et al., 2000).

**Table 5.1.** Data collection, processing and refinement statistics for PhR15Pi-WT bound to RuBP and PhR15Pi-C135S & PhR15Pi-D204N mutant proteins bound to R15P. The values provided in parenthesis are for the last resolution shell.

| | **PhR15Pi-WT•RuBP** | **PhR15Pi-C135S•R15P** | **PhR15Pi-D204N•R15P** |
|---|---|---|---|
| Wavelength (Å) | 1.5418 | 1.5418 | 1.5418 |
| Temperature (K) | 100 | 100 | 100 |
| Space group | $P3_112$ | $P3_112$ | $P3_112$ |
| Unit-cell parameter (Å, °) | $a=b$=98.63, $c$=257.06, $\alpha=\beta$=90, $\gamma$=120 | $a=b$=98.65, $c$=256.73, $\alpha=\beta$=90, $\gamma$=120 | $a=b$=98.81, $c$=257.08, $\alpha=\beta$=90, $\gamma$=120 |

| | | | |
|---|---|---|---|
| Resolution (Å) | 71.14-2.2 (2.25-2.2) | 85.58-2.3 (2.36-2.3) | 85.69-2.21 (2.26-2.21) |
| No. of observed reflections | 717401 | 765831 | 662589 |
| No. of unique reflections | 72651 (4463) | 64003 (4486) | 72313 (4460) |
| Mn(I) CC(1/2) | 0.998 (0.923) | 0.997 (0.942) | 0.998 (0.918) |
| Completeness (%) | 99.5 (98.7) | 100 (100) | 100 (99.8) |
| $V_M$ ($Å^3$ $Da^{-1}$) | 3.30 | 3.30 | 3.32 |
| Solvent content (%) | 62.80 | 62.77 | 62.94 |
| Mosaicity (°) | 0.568 | 0.513 | 0.451 |
| $I/\sigma(I)$ | 17.8 (4.5) | 16.2 (5.7) | 13.7 (3.9) |
| $R_{merge}$ (%) | 8.8 (50.8) | 12.4 (47.8) | 11.7 (49.6) |
| $R_{pim}$ (%) | 4.3 (25.2) | 5.4 (19.9) | 5.6 (25.5) |
| $R_{meas}$ (%) | 9.8 (56.7) | 13.5 (51.8) | 13.9 (59.9) |
| Multiplicity | 9.9 (9.7) | 12 (13.2) | 9.2 (8.1) |
| $R_{work}/R_{free}$ (%) | 14.97/18.97 | 14.39/19.82 | 15.12/19.34 |
| **Protein model** | | | |
| No. of subunits in ASU | 3 | 3 | 3 |
| Protein atoms | 7646 | 7646 | 7646 |
| Water molecules | 507 | 615 | 669 |
| RuBP | 3 | - | - |
| R15P | - | 3 | 3 |
| Others | 9 | 7 | 10 |
| **Deviation from ideal geometry** | | | |
| Bond length (Å) | 0.018 | 0.018 | 0.02 |
| Bond angles (°) | 1.903 | 1.889 | 1.956 |
| **Average *B*-factor ($Å^2$)** | | | |
| Protein atoms | 36.57 | 29.31 | 30.62 |
| Water molecules | 39.69 | 31.96 | 34.20 |
| RuBP | 29.65 | - | - |

| | | | |
|---|---|---|---|
| R15P | - | 21.7 | 19.34 |
| **Ramachandran plot** | | | |
| Favored (%) | 97.91 | 97.81 | 97.91 |
| Allowed (%) | 1.77 | 1.77 | 1.77 |
| Remaining (%) | 0.31 | 0.42 | 0.31 |
| **PDB id** | 5YFJ | 5YFS | 5YFT |

**Table 5.2.** Data collection, processing and refinement statistics for PhR15Pi-WT bound to RuBP & AMP, RuBP & GMP and RuBP, AMP & GMP. The values provided in parenthesis are for the last resolution shell.

| | **PhR15Pi-WT•RuBP•AMP** | **PhR15Pi-WT•RuBP•GMP** | **PhR15Pi-WT•RuBP•AMP•GMP** |
|---|---|---|---|
| Wavelength (Å) | 1.5418 | 1.5418 | 1.5418 |
| Temperature (K) | 100 | 100 | 100 |
| Space group | $P3_112$ | $P3_112$ | $P3_112$ |
| Unit-cell parameter (Å, °) | $a=b$=98.98, $c$=256.61, $\alpha=\beta$=90, $\gamma$=120 | $a=b$=98.99, $c$=257.09, $\alpha=\beta$=90, $\gamma$=120 | $a=b$=98.59, $c$=256.15, $\alpha=\beta$=90, $\gamma$=120 |
| Resolution (Å) | 85.72-2.35 (2.41-2.35) | 85.73-2.75 (2.87-2.75) | 85.38-2.8 (2.94-2.8) |
| No. of observed reflections | 666605 | 546738 | 352039 |
| No. of unique reflections | 60380 (4425) | 37954 (4564) | 35596 (4669) |
| Mn(I) CC (1/2) | 0.998 (0.972) | 0.998 (0.962) | 0.996 (0.956) |
| Completeness (%) | 100 (100) | 100 (100) | 100 (100) |
| $V_M$ ($Å^3$ $Da^{-1}$) | 3.32 | 3.33 | 3.29 |
| Solvent content (%) | 63.00 | 63.08 | 62.64 |
| Mosaicity (°) | 0.55 | 0.518 | 0.55 |
| $I/\sigma(I)$ | 14.4 (3.7) | 22.1 (5.9) | 13.6 (4.9) |

| | | | |
|---|---|---|---|
| $R_{merge}$ (%) | 10.9 (55.9) | 11.3 (49.8) | 13.8 (48) |
| $R_{pim}$ (%) | 5.1 (25.7) | 4.4 (19.5) | 6.7 (23.4) |
| $R_{meas}$ (%) | 12.1 (61.6) | 12.2 (53.5) | 15.4 (53.5) |
| Multiplicity | 11.0 (11.3) | 14.4 (14.8) | 9.9 (9.9) |
| $R_{work}/R_{free}$ (%) | 18.01/24.78 | 16.24/23.22 | 15.52/22.66 |
| **Protein model** | | | |
| No. of subunits in ASU | 3 | 3 | 3 |
| Protein atoms | 7652 | 7646 | 7646 |
| Water molecules | 321 | 193 | 241 |
| RuBP | 3 | 3 | 3 |
| AMP | 3 | - | 3 |
| GMP | - | 2 | 2 |
| GMP[a] | - | 3 | - |
| Others | 4 | 6 | 4 |
| **Deviation from ideal geometry** | | | |
| Bond length (Å) | 0.015 | 0.013 | 0.013 |
| Bond angles (°) | 1.845 | 1.679 | 1.750 |
| **Average *B*-factor (Å$^2$)** | | | |
| Protein atoms | 48.32 | 43.24 | 41.19 |
| Water molecules | 42.10 | 31.28 | 30.53 |
| RuBP | 37.96 | 32.61 | 33.39 |
| AMP | 43.57 | - | 39.55 |
| GMP | - | 102.9 | 89.01 |
| GMP[b] | - | 102.35 | - |
| **Ramachandran plot** | | | |
| Favored (%) | 96.66 | 96.35 | 96.98 |
| Allowed (%) | 2.71 | 3.13 | 2.4 |
| Remaining (%) | 0.63 | 0.52 | 0.63 |
| **PDB id** | 5YFU | 5YG5 | 5YG8 |

[a] GMP bound at the 'AMP binding site'.

**Table 5.3.** Data collection, processing and refinement statistics for PhR15Pi-C135S mutant protein bound to R15P & AMP, R15P & GMP and R15P, AMP & GMP. The values provided in parenthesis are for the last resolution shell.

| | **PhR15Pi-C135S•R15P•AMP** | **PhR15Pi-C135S•R15P•GMP** | **PhR15Pi-C135S•R15P•AMP•GMP** |
|---|---|---|---|
| Wavelength (Å) | 1.5418 | 1.5418 | 1.5418 |
| Temperature (K) | 100 | 100 | 100 |
| Space group | $P3_112$ | $P3_112$ | $P3_112$ |
| Unit-cell parameter (Å, °) | $a=b$=98.14, $c$=255.68, $\alpha=\beta$=90, $\gamma$=120 | $a=b$=98.57, $c$=256.62, $\alpha=\beta$=90, $\gamma$=120 | $a=b$=98.84, $c$=256.28, $\alpha=\beta$=90, $\gamma$=120 |
| Resolution (Å) | 85.23-2.75 (2.87-2.75) | 85.54-2.35 (2.41-2.35) | 85.60-2.8 (2.94-2.80) |
| No. of observed reflections | 292248 | 467995 | 417409 |
| No. of unique reflections | 37157 (4510) | 59895 (4593) | 35799 (4709) |
| Mn(I) CC (1/2) | 0.995 (0.910) | 0.995 (0.907) | 0.997 (0.943) |
| Completeness (%) | 100 (100) | 100 (100) | 100 (100) |
| $V_M$ (Å$^3$ Da$^{-1}$) | 3.25 | 3.30 | 3.31 |
| Solvent content (%) | 62.23 | 62.69 | 62.85 |
| Mosaicity (°) | 0.55 | 0.5 | 0.524 |
| $I/\sigma(I)$ | 12.0 (3.9) | 11.2 (3.9) | 15.6 (5.1) |
| $R_{merge}$ (%) | 14.4 (50.2) | 13.3 (49.7) | 14.2 (50.9) |
| $R_{pim}$ (%) | 8.1 (28.5) | 7.6 (28.7) | 6.3 (22) |
| $R_{meas}$ (%) | 16.6 (57.8) | 15.4 (57.4) | 15.5 (55.5) |
| Multiplicity | 7.9 (7.9) | 7.8 (7.7) | 11.7 (12.3) |
| $R_{work}/R_{free}$ (%) | 15.18/22.40 | 15.46/20.84 | 15.57/22.73 |
| **Protein model** | | | |
| No. of subunits in ASU | 3 | 3 | 3 |

| | | | |
|---|---|---|---|
| Protein atoms | 7641 | 7646 | 7646 |
| Water molecules | 254 | 499 | 210 |
| R15P | 3 | 3 | 3 |
| AMP | 3 | - | 3 |
| GMP | - | 2 | 2 |
| GMP[b] | - | 3 | - |
| Others | 4 | 4 | 4 |
| **Deviation from ideal geometry** | | | |
| Bond length (Å) | 0.013 | 0.017 | 0.013 |
| Bond angles (°) | 1.744 | 1.840 | 1.747 |
| **Average *B*-factor (Å$^2$)** | | | |
| Protein atoms | 36.60 | 32.34 | 37.74 |
| Water molecules | 27.77 | 32.96 | 28.68 |
| R15P | 31.99 | 26.26 | 35.17 |
| AMP | 34.96 | - | 44.32 |
| GMP | - | 78.47 | 117.78 |
| GMP[a] | - | 105.49 | - |
| **Ramachandran plot** | | | |
| Favored (%) | 97.5 | 97.91 | 96.87 |
| Allowed (%) | 1.98 | 1.67 | 2.5 |
| Remaining (%) | 0.52 | 0.42 | 0.63 |
| **PDB id** | 5YFV | 5YG6 | 5YG9 |

[a] GMP bound at the 'AMP binding site'.

**Table 5.4.** Data collection, processing and refinement statistics for PhR15Pi-D204N mutant protein bound to R15P & AMP, R15P & GMP and R15P, AMP & GMP. The values provided in parenthesis are for the last resolution shell.

| | **PhR15Pi-D204N•R15P•AMP (Co-crystal)** | **PhR15Pi-D204N•R15P•AMP (Soaking)** | **PhR15Pi-D204N•R15P•GMP** | **PhR15Pi-D204N•R15P•AMP•GMP** |
|---|---|---|---|---|
| Wavelength (Å) | 1.5418 | 1.5418 | 1.5418 | 1.5418 |
| Temperature (K) | 100 | 100 | 100 | 100 |
| Space group | *P*321 | $P3_112$ | $P3_112$ | $P3_112$ |
| Unit-cell parameter (Å, °) | *a*=*b*=81.13, *c*=100.18, *α*=*β*=90, *γ*=120 | *a*=*b*=98.7, *c*=256.55, *α*=*β*=90, *γ*=120 | *a*=*b*=99.160, *c*=257.04, *α*=*β*=90, *γ*=120 | *a*=*b*=98.8, *c*=256.43, *α*=*β*=90, *γ*=120 |
| Resolution (Å) | 70.26-2.7 (2.83-2.7) | 85.52-2.35 (2.41-2.35) | 85.87-2.5 (2.58-2.5) | 85.56-2.45 (2.53-2.45) |
| No. of observed reflections | 112204 | 626998 | 409894 | 671257 |
| No. of unique reflections | 10900 (1408) | 60020 (4384) | 46864 (4020) | 53172 (4603) |
| Mn(I) CC (1/2) | 0.998 (0.912) | 0.998 (0.954) | 0.999 (0.927) | 0.999 (0.967) |
| Completeness (%) | 100 (100) | 100 (100) | 92.7 (87.9) | 100 (100) |
| $V_M$ ($Å^3$ $Da^{-1}$) | 2.61 | 3.30 | 3.34 | 3.31 |
| Solvent content (%) | 52.98 | 62.78 | 63.20 | 62.84 |
| Mosaicity (°) | 0.764 | 0.683 | 0.567 | 0.527 |
| $I/\sigma(I)$ | 23.0 (4.4) | 16.3 (5.1) | 17.6 (3.7) | 19.6 (5.2) |
| $R_{merge}$ (%) | 8.5 (51.3) | 10.5 (46.1) | 8.1 (50.4) | 10.5 (55.2) |
| $R_{pim}$ (%) | 4 (25.2) | 4.9 (22.2) | 3.9 (23.3) | 4.5 (21.2) |
| $R_{meas}$ (%) | 9.4 (57.2) | 11.6 (51.2) | 9.1 (55.9) | 11.4 (59.2) |

| | | | | |
|---|---|---|---|---|
| Multiplicity | 10.3 (9.8) | 10.4 (10.3) | 8.7 (9.1) | 12.6 (15.1) |
| $R_{work}/R_{free}$ (%) | 17.65/26.44 | 15.58/20.27 | 17.34/22.61 | 16.75/22.53 |
| **Protein model** | | | | |
| No. of subunits in ASU | 1 | 3 | 3 | 3 |
| Protein atoms | 2539 | 7661 | 7646 | 7646 |
| Water molecules | 45 | 464 | 211 | 301 |
| R15P | 1 | 3 | 3 | 3 |
| AMP | 1 | 3 | - | 3 |
| GMP | - | - | 2 | 2 |
| GMP[b] | - | - | 3 | - |
| Others | - | 4 | 4 | 4 |
| **Deviation from ideal geometry** | | | | |
| Bond length (Å) | 0.011 | 0.017 | 0.015 | 0.016 |
| Bond angles (°) | 1.651 | 1.900 | 1.799 | 1.822 |
| **Average *B*-factor (Å$^2$)** | | | | |
| Protein atoms | 49.08 | 35.28 | 48.65 | 41.90 |
| Water molecules | 32.41 | 35.09 | 39.06 | 38.33 |
| R15P | 34.06 | 23.35 | 34.31 | 29.15 |
| AMP | 40.87 | 81.07 | - | 46.83 |
| GMP | - | - | 99.72 | 124.25 |
| GMP[b] | - | - | 86.60 | - |
| **Ramachandran plot** | | | | |
| Favored (%) | 96.24 | 98.01 | 96.87 | 96.77 |
| Allowed (%) | 3.13 | 1.57 | 2.29 | 2.61 |
| Remaining (%) | 0.63 | 0.42 | 0.83 | 0.63 |
| **PDB id** | 5YFX | 5YFW | 5YG7 | 5YGA |

[a] GMP bound at the 'AMP binding site'.

### 5.2.4 ISOTHERMAL TITRATION CALORIMETRY

Isothernal titration calorimetry (ITC) experiments of mutant proteins (PhR15Pi-C135S, PhR15Pi-C135A and PhR15Pi-D204N) with the substrate (R15P) and the product (RuBP) molecules were measured at 25°C using MicroCal iTC200. Before ITC experiments, the proteins were dialyzed extensively in buffer (20 mM Tris-HCl pH 8.0 and 150 mM KCl) and the equilibrated buffer was subsequently used to prepare ligand solutions to minimize the heat of dilution effects. Each titration consisted of a preliminary injection of 0.4 µl of 7.5–8.0 mM of R15P and RuBP followed by 24 injections of 1.2 µl (spacing 120 sec) into the sample cell filled with protein at a final concentration of 75–80 µM. During titration, the syringe rotated at a speed of 450 rpm to facilitate adequate mixing of the reactants. Data were analyzed using ORIGIN software version 7.0 and baselines were subtracted from data to obtain accurate heat exchanges.

### 5.2.5 MULTIPLE SEQUENCE ALIGNMENT

The amino acid sequences of proteins used in multiple sequence alignment (MSA) were downloaded from UniProtKB database (The UniProt Consortium, 2017). Homology searches were performed using the program BLAST (Altschul et al., 1990). The MSA was performed using the program Clustal Omega (Sievers and Higgins, 2014) and the aligned sequences were further decorated using the online tool ESPript (Gouet et al., 2003).

## 5.3 RESULTS

### 5.3.1 THE OVERALL STRUCTURE

The crystal structure of PH0208 protein was solved by MR method using the atomic coordinates of TkR15Pi (PDB id: 3VM6), which has a sequence identity of 86% with PH0208. One asymmetric unit (ASU) consists of three identical (root mean square deviation, RMSD: 0.16 Å for 321 $C_\alpha$ atoms of 324 residues) subunits of the protein. Monomer of PH0208 protein consists of an N-terminal α-helical domain (residues 1–122) and a C-terminal α-β-α sandwich domain (residues 123–324). The NTD consists of five α-helices ($\alpha_1$-$\alpha_5$) while the CTD adopts a Rossmann-like fold with eight α-helices ($\alpha_6$-$\alpha_{13}$), two $3_{10}$-helices ($\eta_1$ & $\eta_2$) and ten, mostly parallel β-strands ($\beta_1$-$\beta_{10}$). The NTD

and CTD are connected by the longest helix, $\alpha_5$ (residues 90–120), of the protein (Figure 5.7). Structural homology search using the web server DALI (Holm and Rosenström, 2010) reveals that PH0208 protein is closely related to proteins belonging to PF01008 family of Pfam database (Finn et al., 2014) with RMSD ranging from 0.5–2.5 Å. This family, also known as eIF2B-related, includes the regulatory subunits (α, β and δ) of the translation IFs, eIF2B and 5-methylthioribose 1-phosphate (M1Pi), an enzyme involved in methionine salvage pathway (MSP). The closest homologs of PH0208 include R15Pi from *T. kodakarensis* (PDB id: 3VM6, Z-score: 51.3, RMSD: 0.5 Å), M1Pi from *B. subtilis* (PDB id: 2YVK, Z-score: 37.5, RMSD: 1.9 Å), aIF2Bα from *P. horikoshii* (PDB id: 1VB5, Z-score: 35.6, RMSD: 1.8 Å), eIF2B regulatory subunits from *Schizosaccharomyces pombe* (PDB id: 5B04, Z-score: 33–30, RMSD: 3.8-2.0 Å) and *Chaetomium thermophilum* (PDB id: 5DBO, Z-score: 28.5–26.5, RMSD: 3.1 Å).

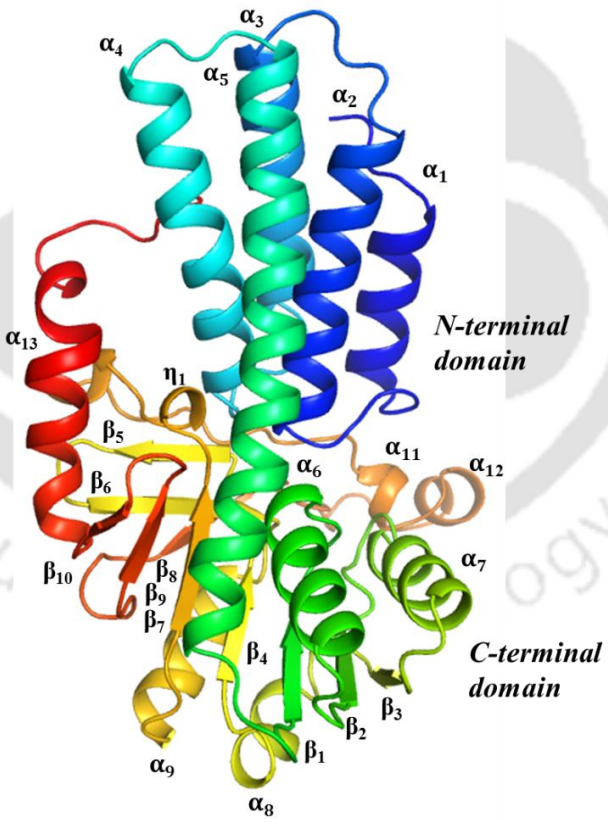



**Figure 5.7. The overall monomeric structure of PH0208 protein from *P. horikoshii* represented as cartoon model.** The secondary structure elements have been labeled.

The enzyme R15Pi from *T. kodakarensis* has been reported to be present as a hexamer in aqueous solution (Nakamura et al., 2012). Thus, the hexameric oligomer of PH0208 was generated using the program PISA (Krissinel and Henrick, 2007). PH0208 forms hexamer with a buried surface area of ~14866 $Å^2$. The interactions among the subunits of the hexamer are largely mediated by the CTD while the NTD remains completely dissociated (Figure 5.8A). The interaction between amino acid residues at the CTD of two monomers lead to the formation of a dimer with a buried surface area of ~4109 $Å^2$. Out of the many amino acid residues involved in dimer formation, Lys155, Arg162 and Lys168 form salt bridges with Glu275, Glu160 and Glu266 of the associating monomer, respectively (Figure 5.8B). Three dimers together constitute the hexamer through interactions involving the residue Lys226 of one monomer and Glu287 of the neighboring monomer while the residue Arg229 participates in hydrogen bonding with the residues Glu287 and Tyr288 of the adjacent monomer (Figure 5.8C).

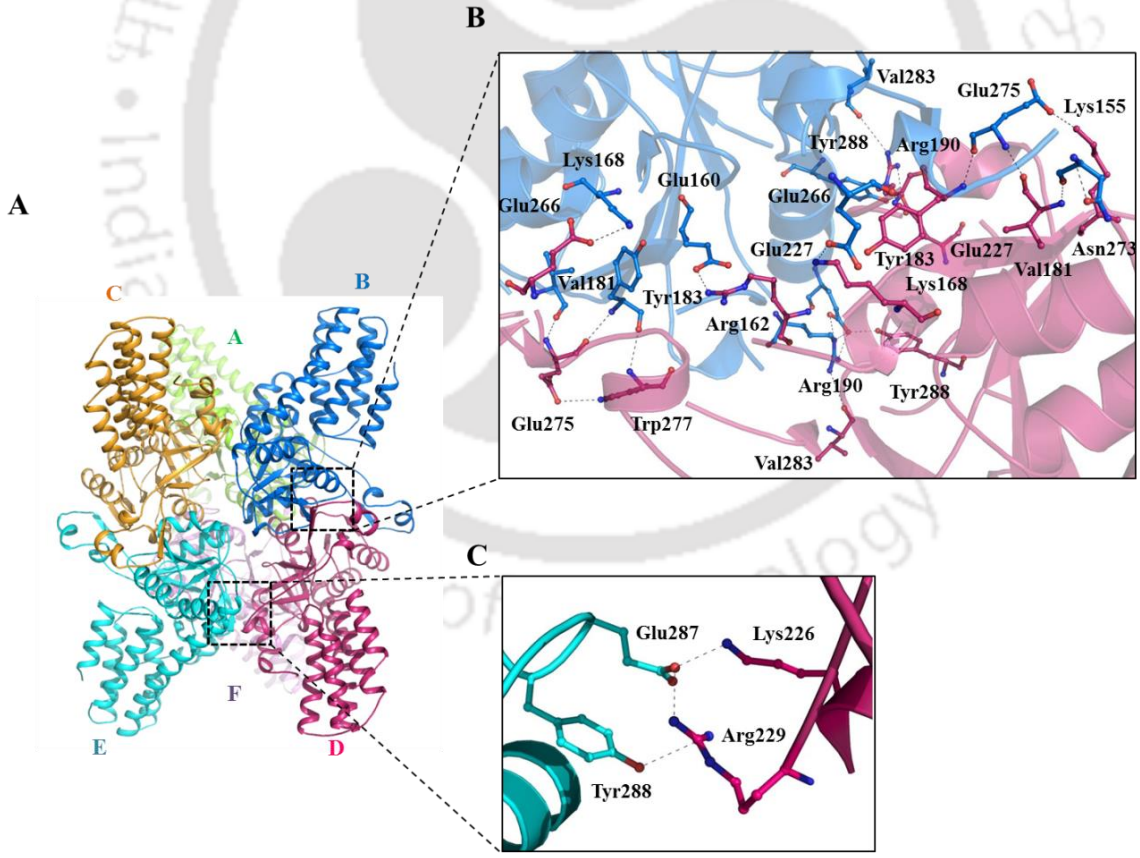



**Figure 5.8. Quaternary structure of PH0208 protein.** (A) Hexamer (chain A-F) of PH0208 protein. Each chain are highlighted with different colors. Amino acid residues involved in inter-subunit interaction between (B) dimer and (C) hexamer of PH0208

protein. The amino acid residues involved in interaction are shown in ball-and-stick model.

### 5.3.2 THE ACTIVE-SITE POCKET

Based on the hypothesis that PH0208 might encode the enzyme R15Pi (Gogoi et al., 2016), crystallization of the protein was attempted in the presence of the substrate (R15P). The active-site pocket of PH0208 protein resides in the cavity formed between the NTD and CTD and was found to hold the product (RuBP) indicating that the substrate was converted into product during the process of crystallization. However, unlike the homologous proteins, PH0208 (hereafter referred to as PhR15Pi) remains in a closed conformation even when the product is bound to the active site. This closed conformation is reflected by the bend angle of the $\alpha_5$ helix, which is ~10° for the closed conformation (Figure 5.9A).

In the active-site pocket, the 1-phosphate group of RuBP forms hydrogen bond to the side chains of the residues Arg22, Arg65, Gln166, Lys215 and the backbone amide group of Gly23. The 5-phosphate group forms hydrogen bond to the side chains of the residues Ser137, Lys240 and the backbone amide group of Ala24, Gly25, Lys138 and Ala139. The side chain atoms of Asp204, Asn214 and Lys215 interact with the oxygen atoms of the ribulose sugar (Figure 5.9B). The amino acid residues of PhR15Pi involved in interacting with the 1-phosphate, sugar and 5-phosphate group of RuBP are well conserved in all R15Pi enzymes. The two catalytic residues, Cys135 and Asp204 (Cys133 and Asp202 in TkR15Pi) essential for the enzymatic activity of R15Pi are also conserved in PhR15Pi protein. In order to verify the role of the catalytic residues in PhR15Pi protein, the residue Cys135 was mutated to serine. The mutation led to the formation of a catalytically inactive protein which lost the ability to convert the substrate into product and thus, held the substrate in the active-site pocket (Figure 5.9C). The substrate molecule (R15P) binds to the active site of PhR15Pi -C135S in a manner similar to that of the wild type (WT) protein (Figure 5.9C). The other catalytic residue, Asp204, reported to be vital for ligand binding in R15Pi enzyme was mutated to asparagine. However, quite surprisingly, a clear electron density of the substrate was still observed at the active-site pocket of the PhR15Pi-D204N mutant protein. The interactions involved in holding R15P in the active-site pocket of PhR15Pi-D204N

mutant protein are equivalent to those of the PhR15Pi-WT and PhR15Pi-C135S proteins (Figure 5.9D).

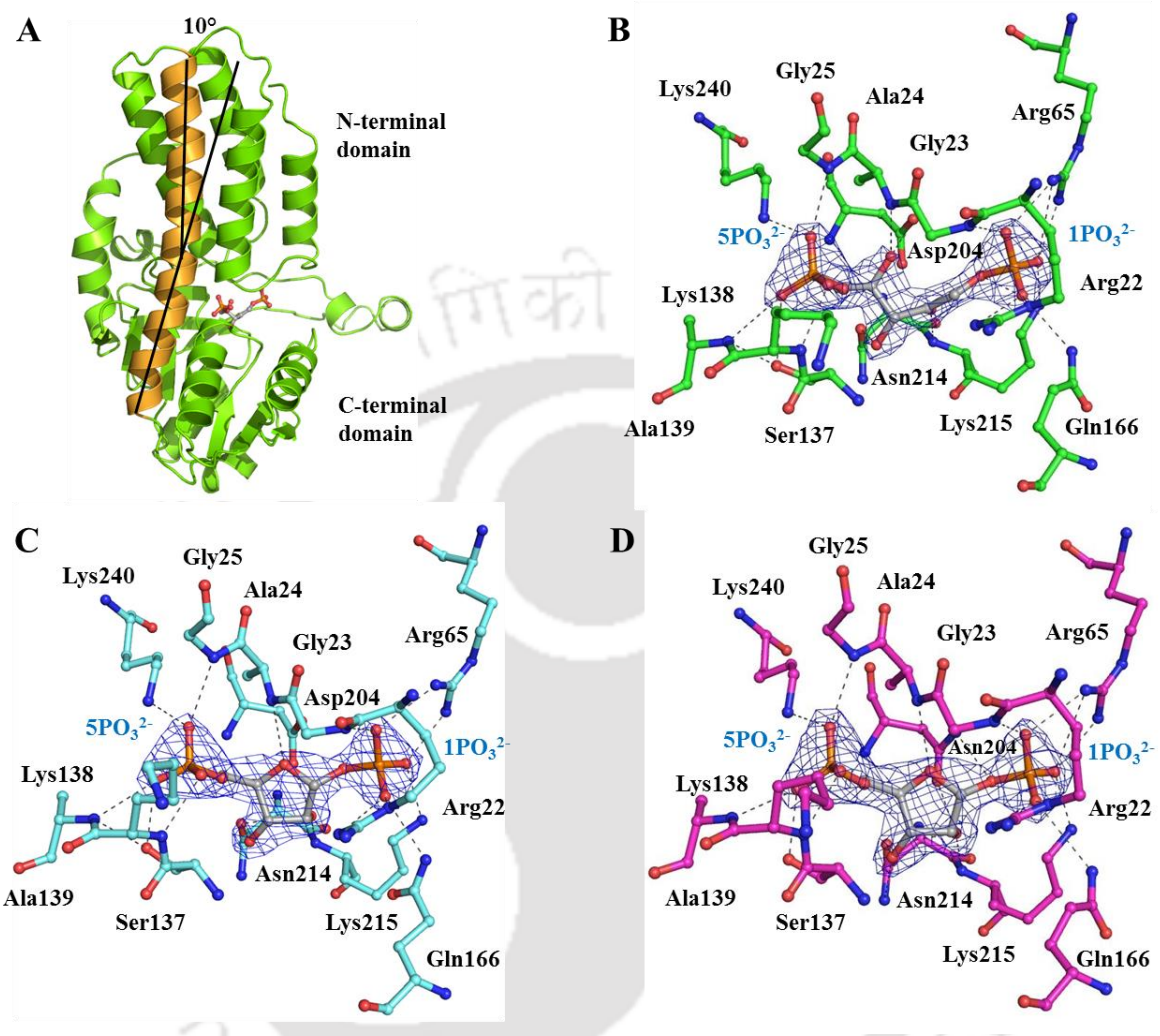



**Figure 5.9. The active-site pocket of PhR15Pi.** (A) Closed conformation of PhR15Pi bound to RuBP. The bend angle (~10°) of $\alpha_5$ helix highlighted in orange signifies the closed conformation. Active-site residues of (B) PhR15Pi-WT protein bound to the product, RuBP, (C) PhR15Pi -C135S and (D) PhR15Pi-D204N mutant protein bound to the substrate, R15P. The substrate (R15P) and the product (RuBP) are shown in grey using ball-and-stick model and the phosphate groups are labeled in blue. Blue meshes indicate the $2F_O$-$F_C$ maps for RuBP and R15P contoured at 1.0σ. All the amino acid residues involved in interaction are labeled and shown in ball-and-stick model.

### 5.3.3 THE BINDING AFFINITY OF SUBSTRATE/PRODUCT

To determine the binding affinity of the substrate (R15P) and the product (RuBP) to the protein PhR15Pi, ITC experiments were performed. Since binding of the substrate to the PhR15Pi-WT protein triggers a conformational change along with the formation of the product, the heat change measured during the ITC experiment would reflect the sum of heat generated and absorbed during each event. As a result, the binding affinity of PhR15Pi and R15P/RuBP obtained would be delusional. Therefore, binding affinity was studied using the mutant proteins, PhR15Pi-C135S and PhR15Pi-D204N in which case the heat change would be solely due to ligand binding. The analysis shows that both PhR15Pi-C135S and PhR15Pi-D204N proteins bind to R15P with high affinity with equilibrium dissociation constants ($K_d$) of 13.8 and 10 μM, respectively (Table 5.5). The raw data for both the mutants display an endothermic binding to R15P based on the positive values observed for the peak (Figure 5.10A, B). Positive ΔH (enthalpy) value indicates disruption of energetically favorable interactions between the atoms. However, from the crystal structure it is evident that R15P binds to the active-site pocket. Since the ΔH reflects the energy change in the entire system upon ligand binding, positive value of ΔH upon R15P binding might be contributed due to the disruption of interactions formed between the amino acid residues and water molecules in the active-site pocket during the shift from an open to closed conformation, which dominates the heat released upon ligand binding. On the contrary, the PhR15Pi-C135S and PhR15Pi-D204N bind the product (RuBP) with negative ΔH value for the binding peaks indicating an exothermic reaction (Figure 5.10C, D). The mutants formed strong interactions with RuBP with $K_d$ values of 7.5 and 7.2 μM, respectively, which corresponds to the binding affinity of the substrate (Table 5.5).

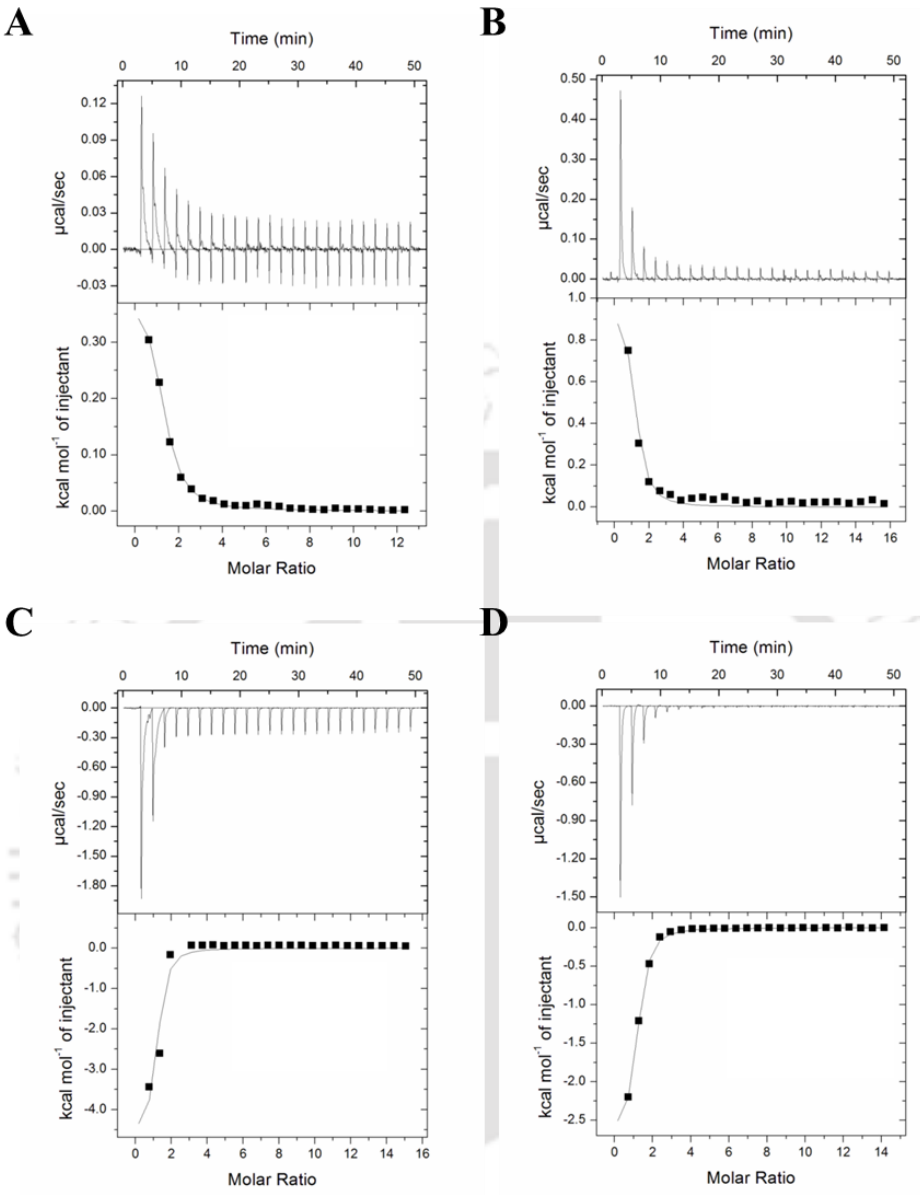



**Figure 5.10. Biophysical characterization of PhR15Pi protein.** ITC for the binding of (A) R15P to PhR15Pi-C135S, (B) R15P to PhR15Pi-D204N, (C) RuBP to PhR15Pi-C135S and (D) RuBP to PhR15Pi-D204N. The upper panel shows the heat change elicited upon successive injections of ligand into the protein. The lower panel shows the binding isotherm as a function of the molar ratio of ligand to protein. A theoretical curve was fitted to the integrated data using a single-site model.

**Table 5.5.** ITC data for the binding of R15P/RuBP to PhR15Pi-C135S, PhR15Pi-C135A and PhR15Pi-D204N mutant proteins.

| **Protein & ligand** | **N** | **ΔH (kcal mol$^{-1}$)** | **TΔS (kcal mol$^{-1}$)** | **$K_a$ ($M^{-1}$)** | **$K_d$ (μM)** | **ΔG (kcal mol$^{-1}$)** |
|---|---|---|---|---|---|---|
| PhR15Pi-C135S & R15P | 1.17±0.03 | 0.403±0.013 | 7.03 | $7.20X10^4$± $7.13X10^3$ | 13.8 | -6.6 |
| PhR15Pi-D204N & R15P | 1(fixed) | 0.981±0.049 | 7.77 | $9.91X10^4$± $2.63X10^3$ | 10 | -6.8 |
| PhR15Pi-C135S & RuBP | 1 (fixed) | -4.74±0.39 | 2.24 | $1.33X10^5$± $7.21X10^4$ | 7.5 | -6.98 |
| PhR15Pi-D204N & RuBP | 1.02 | -2.95±0.06 | 4.05 | $1.39X10^5$± $1.20X10^4$ | 7.2 | -7.0 |
| PhR15Pi-C135A & RuBP | 1 (fixed) | -8.94±0.40 | -4.14 | $3.23X10^3$± 299 | 309 | -4.8 |

N.B. N: binding stoichiometry, ΔH: enthalpy change, ΔS: entropy change, T: Temperature, $K_a$: association constant, $K_d$: dissociation constant, ΔG: Gibbs free energy

### 5.3.4 THE AMP AND GMP BINDING SITES

To identify the binding site of ribonucleotides, either co-crystallization or soaking of PhR15Pi-WT, PhR15Pi-C135S and PhR15Pi-D204N proteins with AMP, GMP, CMP and TMP was performed. Complex of AMP and GMP bound to the proteins were obtained, however, no electron density for CMP or TMP was observed. The binding site of AMP lies at the interface of three monomers (chain A, B and D) of the hexameric protein (Figure 5.11A). The side chains of Arg124 and Asp290 of chain A undergoes a shift towards AMP to form hydrogen bond with the O2′ and O3′ oxygen atoms of the ribose sugar (Figure 5.11B). Another two residues of chain A, Arg229 and Trp231 are involved in interacting with the phosphate groups of the AMP molecule. The side chain atoms of Asn209 and the backbone atoms of Val251 of chain B forms hydrogen bond with the oxygen (O2′ and O3′) atoms of the ribose sugar and the nitrogen (N1 and N6) atom of the adenosine base ring, respectively. Along with chain A and chain B, AMP

also binds to a third protomer of the hexameric protein through interactions involving the side chain atoms of Arg190 and Lys194 of chain D (Figure 5.11C). Thus, one AMP molecule positions itself to hold three protomers of the hexamer. All the amino acid residues of PhR15Pi forming hydrogen bond with AMP are found to be absolutely conserved in R15Pi enzymes. Mutation of the two catalytic residues (Cys135 and Asp204) did not alter the binding behavior of AMP to the protein. In case of both the mutants, PhR15Pi-C135S and PhR15Pi-D204N, AMP occupies a similar position to that of PhR15Pi-WT protein with the major interactions arising from the three subunits of the hexamer. Similar to PhR15Pi-WT, the adenosine base ring interacts with Val251 of chain B, the oxygen atoms of ribose sugar interacts with Arg124 & Asp290 of chain A and Asn209 of chain B while the phosphate group forms hydrogen bond with Arg229 & Trp231 of chain A and Arg190 & Lys194 of chain D of the mutant proteins (Figure 5.11D–F).

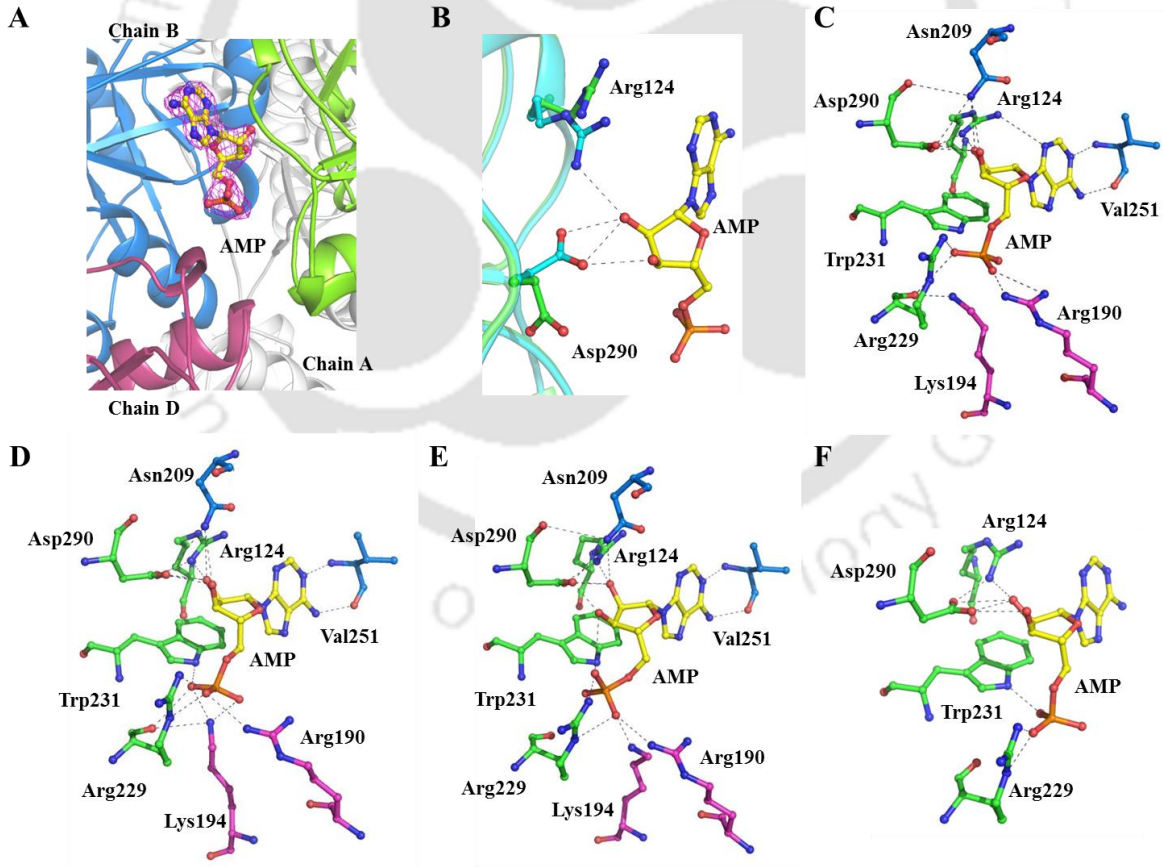



**Figure 5.11. AMP molecule binds at the interface of three protomers.** (A) The AMP molecule (shown in yellow ball-and-stick model) bound at the interface of three monomers of PhR15Pi. Magenta meshes indicate the $2F_O$-$F_C$ map for AMP contoured at

1.0σ. Chain A, B and D of PhR15Pi interacting with AMP are shown in green, blue and magenta, respectively. (B) Close up view of the shift in the side chain orientation of Arg124 and Asp290 upon AMP binding. Arg124 and Asp290 of PhR15Pi-WT and PhR15Pi-WT•AMP complex are shown in green and cyan ball-and-stick model, respectively. Close up view of AMP interacting with the amino acid residues from three different monomers of (C) PhR15Pi-WT, (D) PhR15Pi-C135S, (E) PhR15Pi-D204N (obtained through soaking) and (F) PhR15Pi-D204N (obtained through co-crystallization). The amino acid residues involved in hydrogen bond interaction are shown as ball-and-stick models in different colors for each subunit.

Interestingly, the complex structure of PhR15Pi-WT bound to GMP reveals the binding of two GMP molecules; one GMP at the 'AMP binding site' and another GMP molecule at a site present in the loop region connecting $\alpha_{12}$ and $\beta_8$ (hereafter referred to as 'GMP binding loop') (Figure 5.12A). Although, GMP at the 'AMP binding site' is held by the same set of amino acid of PhR15Pi-WT as in the case of AMP, there exists certain differences in the binding mode. The most pronounced difference lies in the lack of conformational change in the side chains of Arg124 and Asp290 upon GMP binding. This causes Arg124 to form hydrogen bond with N7 of the guanosine base ring instead of O2′ and Asp290 to form weak hydrogen bond with only O3′ of ribose sugar of the GMP molecule (Figure 5.12B). Furthermore, unlike AMP, the amide group of the residue Val251 interacts with O6 oxygen atom of the guanosine base ring of GMP. On the other hand, similar to AMP at the 'AMP binding site', Arg229 & Trp231 of chain A and Arg190 & Lys194 of chain D interact with the phosphate group of GMP (Figure 5.12C). Similar to PhR15Pi-WT protein, the complex structures of PhR15Pi-C135S and PhR15Pi-D204N binding to GMP revealed the presence of two molecules of GMP, with one GMP at the 'AMP binding site' and the other at the 'GMP binding loop'. All the interactions between the amino acid residues at the 'AMP binding site' of the mutant proteins and GMP molecule are equivalent to that of PhR15Pi-WT protein (Figure 5.12D, E). At the 'GMP binding loop', the backbone atoms of the conserved residues Pro271, Ile274 and Val276 interacts with N2, N1 and O6 atoms of the guanosine base ring of GMP molecule, respectively. The $O^{\varepsilon 1}$ of Glu264 further interacts with O5′ while the phosphate group remain absolutely free. These interactions remain consistent in case of all the three proteins viz., PhR15Pi-WT, PhR15Pi-C135S and PhR15Pi-D204N (Figure 5.12F).

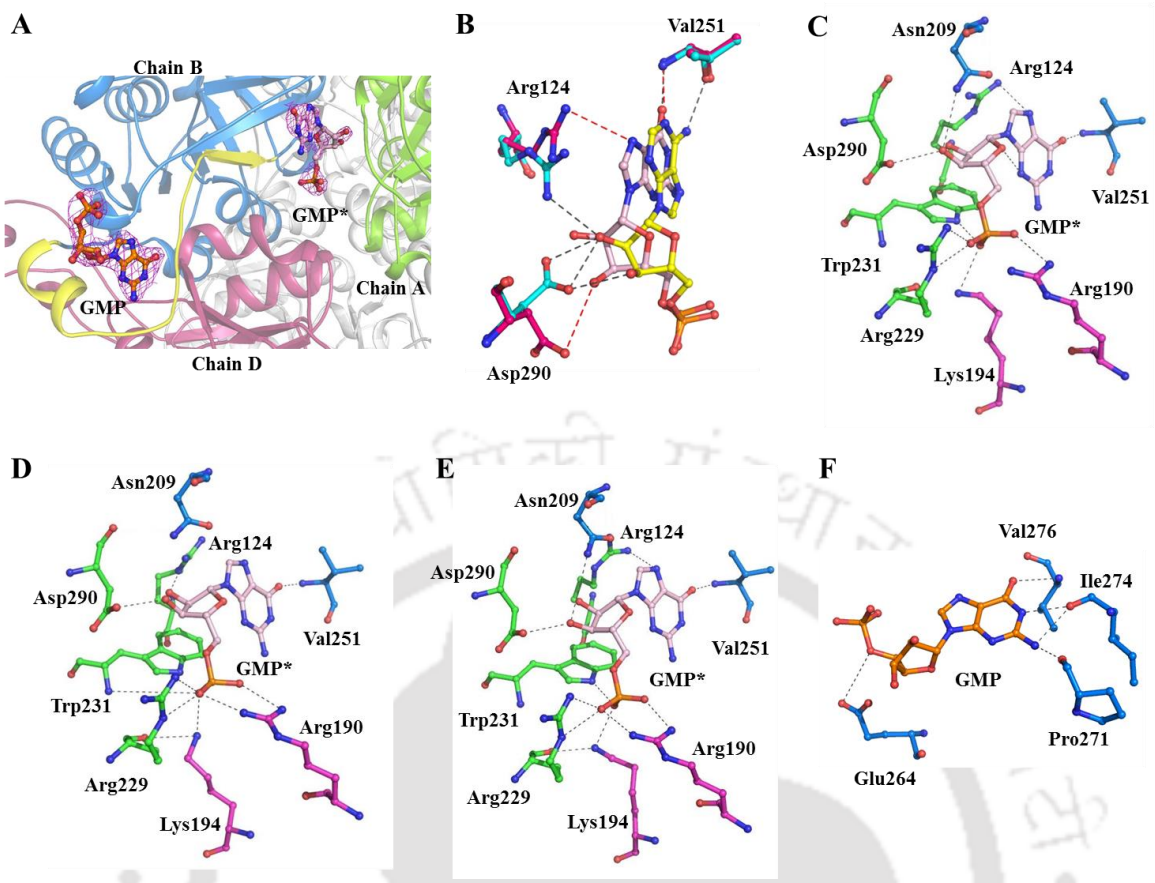



**Figure 5.12. GMP binds at two different sites in PhR15Pi.** (A) GMP bound at two different sites in PhR15Pi. GMP bound at the 'AMP binding site' and 'GMP binding loop' are shown in pink and orange colored ball-and-stick model, respectively. Magenta meshes indicate the $2F_O$-$F_C$ map for GMP contoured at 1.0σ. Chain A, B and D of PhR15Pi interacting with GMP at the 'AMP binding site' are shown in green, blue and magenta, respectively and the 'GMP binding loop' between $\alpha_{12}$ and $\beta_8$ is highlighted in yellow. (B) Close up view of the different binding modes of AMP and GMP at the 'AMP binding site'. Arg124, Val251 and Asp290 interacting with AMP and GMP are shown as cyan and dark pink ball-and-stick model, respectively. The hydrogen bonds between AMP & GMP and Arg124, Val251 and Asp290 are distinguished with grey and red dashed lines, respectively. Close up view of GMP at the 'AMP binding site' & the 'GMP binding loop' interacting with the amino acid residues from three different monomers of (C) PhR15Pi-WT (D) PhR15Pi-C135S and (E) PhR15Pi-D204N. The amino acid residues involved in hydrogen bond interaction are shown as ball-and-stick model in different colors for each subunit. (F) Close up view of GMP interacting with amino acid residues of the 'GMP binding loop'.

To verify the preferred molecule at the 'AMP binding site', PhR15Pi-WT, PhR15Pi-C135S and PhR15Pi-D204N were crystallized in the presence of both AMP and GMP. In the presence of both the potential ligands, AMP outcompete GMP and binds to the 'AMP binding site' utilizing the similar set of amino acid while GMP binds to the 'GMP

binding loop' (Figure 5.13). During crystallization, even a higher concentration of GMP (4.5 mM) as compared to AMP (1.5 mM) could not aid GMP to occupy the 'AMP binding site'. This further indicates a higher affinity of AMP at the 'AMP binding site'.

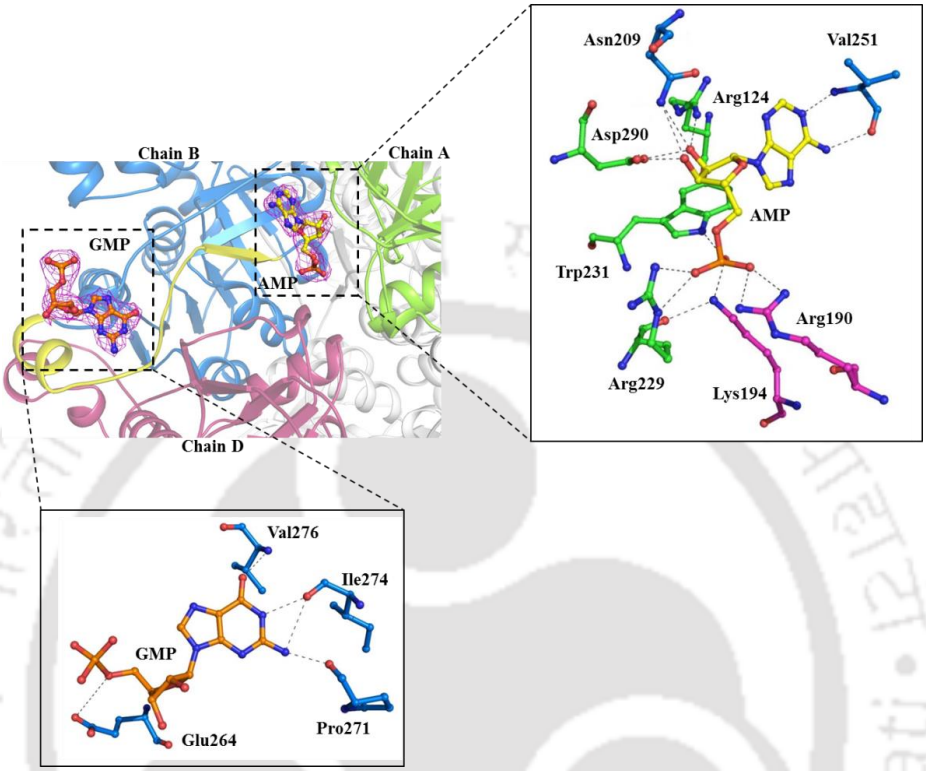



**Figure 5.13. AMP is preferred over GMP at the 'AMP binding site'.** The AMP and GMP molecules bound to PhR15Pi are shown in yellow and orange ball-and-stick model, respectively. Magenta meshes indicate the $2F_O$-$F_C$ map for AMP and GMP contoured at 1.0σ. The 'GMP binding loop' between $\alpha_{12}$ and $\beta_8$ is highlighted in yellow. The insets show close up view of AMP (upper panel) and GMP (lower panel) interacting with amino acid residues at the 'AMP binding site' and 'GMP binding loop', respectively.

### 5.3.5 COMPARISON BETWEEN PhR15Pi AND eIF2Bα

The PhR15Pi protein and the regulatory subunits (α, β and δ) of eIF2B share a pronounced resemblance at the tertiary level (RMSD, eIF2Bα- 2.0 Å, eIF2Bβ- 2.44 Å and eIFBδ- 1.48 Å) with a highly similar all α-helical NTD and Rossmann-like fold CTD. Similar to PhR15Pi, the regulatory subunits of eIF2B assemble together by making extensive interactions via the CTD of each subunit (Kuhle et al., 2015; Kashiwagi et al., 2016). Among the three regulatory subunits of eIF2B from *H. sapiens* and *S. pombe*, PhR15Pi shares highest sequence homology with the α-subunit (query coverage: 86 &

83%, identity: 25 & 30%, respectively). A comparison of the active-site pocket of PhR15Pi and eIF2Bα from *H. sapiens* (HseIF2Bα) & *S. pombe* (SpeIF2Bα), reveals that the amino acid residues required for interacting with the 5-phosphate group and the ribose/ribulose sugar are well conserved in both HseIF2Bα and SpeIF2Bα (Figure 5.14A). However, the amino acid residues interacting with the 1-phosphate group of the substrate/product are absent in eIF2Bα. Out of the two catalytic residues, Cys135 is replaced by the residues alanine and glycine in HseIF2Bα and SpeIF2Bα, respectively (Figure 5.14A). Henceforth, Cys135 was mutated to alanine (PhR15Pi-C135A) to understand its role in ligand (R15P/RuBP) binding. Interestingly, the mutant protein, PhR15Pi-C135A did not show any interaction with the substrate (Figure 5.14B). However, PhR15Pi-C135A mutant protein interacted with the product albeit with a lower binding afnity ($K_d$: 309 μM) as compared to the other two mutant proteins (PhR15Pi-C135S and PhR15Pi-D204N) (Figure 5.14C, Table 5.5).

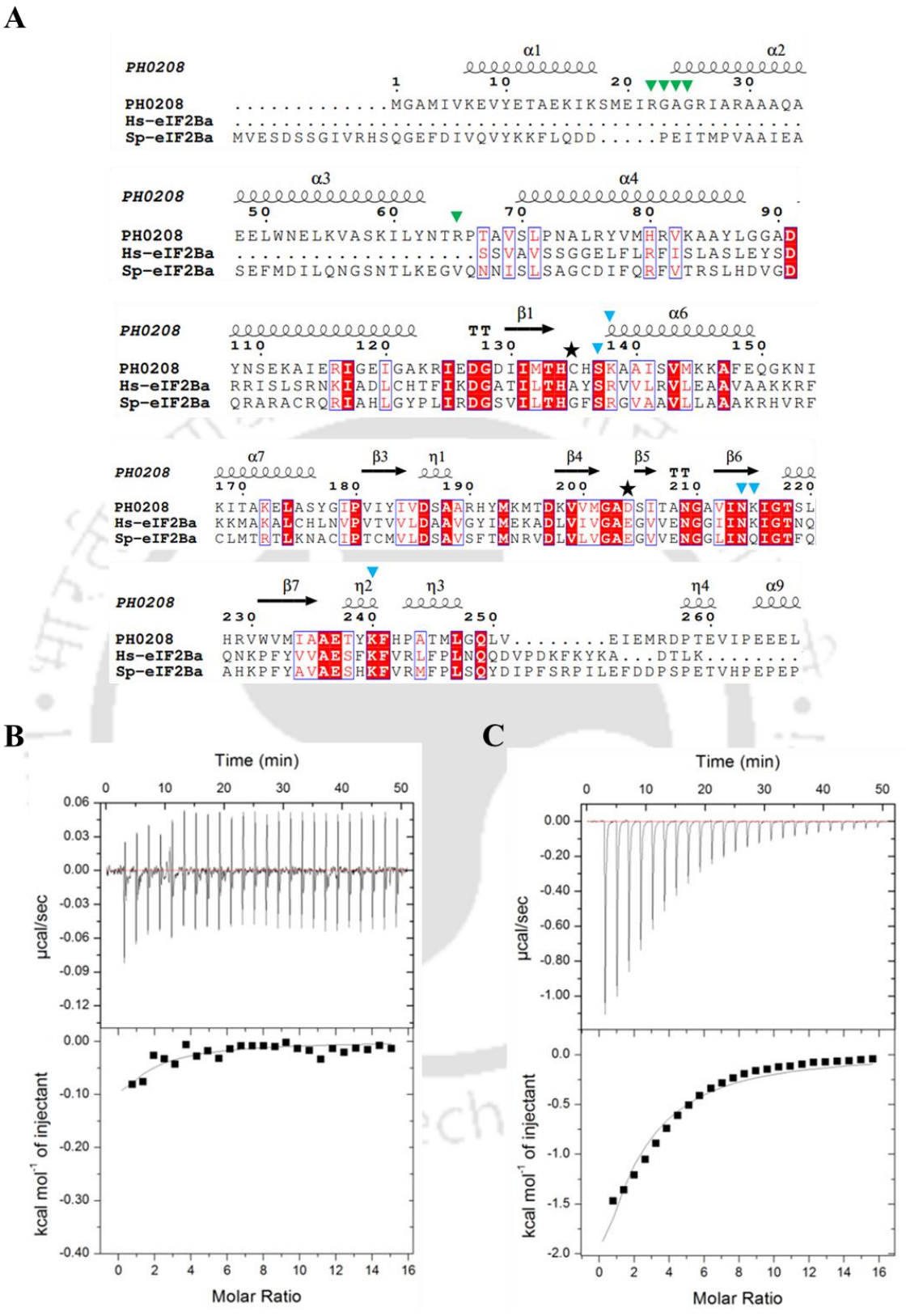



**Figure 5.14. Crucial differences at the active site of PhR15Pi and eIF2Bα.** (A) Multiple sequence alignment (MSA) of amino acid sequences of PhR15Pi (O57947) and eIF2Bα from *H. sapiens* (HseIF2Ba, Q14232) and *S. pombe* (SpeIF2Ba, Q9USP0). The UniProt id for each protein sequences are provided in parenthesis. The residues marked

with green downward arrowheads are involved in interaction with the 1-phosphate group of the R15P/RuBP. The residues marked with blue downward arrowheads are involved in interaction with the ribose or ribulose as well as the 5-phosphate group of R15P/RuBP. The catalytic residues are highlighted with black star. ITC for the binding of (B) R15P and (C) RuBP to PhR15Pi-C135A mutant protein. The upper panel shows the heat change elicited upon successive injections of ligand into the protein. The lower panel shows the binding isotherm as a function of the molar ratio of ligand to protein. A theoretical curve was fitted to the integrated data using a single-site model.

## 5.4 DISCUSSION

An analogy has often been drawn between the process of translation initiation in eukaryotes and archaea, primarily based on the sequence similarity of the components, especially the translation IFs. This belief is further strengthened by the structural similarity exhibited by most of these translation IFs of archaea and eukaryotes. Nevertheless, this equivalence should be drawn only after a comprehensive study of these proteins, both at the primary and tertiary structure levels. The homologs of the three regulatory subunits (α, β and δ) of eIF2B has been reported to be available in *P. horikoshii* OT3; one of them being PH0208 (Kakuta et al., 2004). This lacked direct evidence and required further in-depth analysis before assigning PH0208 with the function of one of the regulatory subunits of eIF2B.

The crystal structure reveals that the overall structure of PH0208 protein is similar to that of the regulatory subcomplex of eIF2B. However, the strong binding of the substrate/product of R15Pi enzyme to the active-site pocket of PH0208 protein by forming interactions with all the crucial and conserved amino acid residues clearly indicates that PH0208 protein would function as R15Pi (hereafter referred to as PhR15Pi). The residues Cys135 and Asp204, known to be absolutely conserved in R15Pi enzyme and essential for the catalysis of R15P (substrate) to RuBP (product) are also conserved in PhR15Pi. Mutation of these two residues in PhR15Pi led to the inhibition of the isomerization of the substrate into product. The absence of electron density of R15P in the active-site pocket of D202N (D204 in PhR15Pi) mutant of R15Pi from *T. kodakarensis* led to the proposition that this residue might play an essential role in ligand binding (Nakamura et al., 2012). Interestingly, the PhR15Pi-D204N mutant held the substrate at the active-site pocket indicating that the residue aspartate alone is not responsible for ligand binding and there might be other criteria that govern the binding of the ligand.

The ability of the AMP molecule to elevate the catalytic efficiency of R15Pi has been well documented (Aono et al., 2012). However, the lack of information regarding the AMP binding site in R15Pi protein has limited the knowledge towards understanding the exact role of AMP during enzyme catalysis. Thus, we determined the structure of PhR15Pi in complex with AMP molecule. The positioning of the AMP at the interface of three monomers implies that AMP might play an important role in providing conformational stability to the hexameric form of PhR15Pi. The change in the side chain orientation of Arg124 and Asp290 to interact with AMP portrays as the key interactions required for this stabilization. Aono and his co-workers demonstrated the inability of compounds such as ATP, S-adenosyl methionine (SAM) or S-adenosyl homocysteine (SAH) in increasing the catalytic efficiency of R15P (Aono et al., 2012). This might be due to the steric hindrance posed by the bulky group at the C5 position of these compounds which would inhibit interaction with the residues Arg190 and Lys194. Thus, it can be inferred that Arg190 and Lys194 provide further specificity to AMP at the 'AMP binding site'.

In the absence of AMP, binding of the GMP molecule at the 'AMP binding site' indicates that during AMP scarcity, GMP might occupy the same site and render similar structural stability to the protein. Although, GMP binds to the 'AMP binding site', the absence of the essential interactions of Arg124 and Asp290 with the ribose sugar of GMP causes the binding to be less specific. This explains the lower degree of activation of R15Pi enzyme in the presence of GMP (Aono et al., 2012). Evidently, when both AMP and GMP were made available, AMP binds to the 'AMP binding site' with a higher preference than GMP while GMP occupies the 'GMP binding loop' only. Binding of GMP to the 'GMP binding loop' involves very few interactions which predominantly includes only the guanosine base ring of the GMP molecule. Thus, it might also be possible that the binding of GMP to the 'GMP binding loop' is an outcome of crystallization artifact. Both, AMP and GMP bind to PhR15Pi even after the mutation of the catalytic residues (PhR15Pi-C135S and PhR15Pi-D204N) at the active-site pocket. This implies that purine nucleotides would interact at their respective sites irrespective of whether the substrate or product is bound at the active-site pocket. Interaction of the nitrogen (N1 and N6) and oxygen (O6) of adenine and guanosine base rings, respectively, with Val251 of chain B manifests the requirement of double-ring structure

(purine nucleotides) at the 'AMP binding site'. This explains the inability of pyrimidine nucleotides, CMP and TMP to bind to the 'AMP binding site'.

Although functionally divergent, the similarity between the regulatory subcomplex of eIF2B and PhR15Pi at the structural level remains indisputable. The overall arrangement of the six subunits of PhR15Pi with the major interactions occurring at the CTD is comparable to the hexamer assembly of the regulatory subunits of eIF2B. The high similarity between the two proteins might originate from the fact that during the course of evolution, R15Pi underwent refunctionalization from a metabolic enzyme to form eIF2B, a translation initiation factor (Kuhle et al., 2015). A detailed comparison of the active-site pocket of PhR15Pi and eIF2Bα reveals the prime differences that dictate the ligand specificity to each protein. The amino acid residues of R15Pi involved in interacting with the 5-phosphate group and the ribose/ribulose sugar of R15P are also conserved in eIF2Bα. However, eIF2Bα lacks those amino acid residues required for interacting with the 1-phosphate group. Thus, it can be hypothesized that the active-site pocket of eIF2Bα might accommodate a sugar-5-phosphate molecule such as ribose/ribulose-5-phosphate. Mutation of one of the catalytic residues, cysteine to alanine (PhR15Pi-C135A) inhibits the binding of R15P to PhR15Pi. Similarly, in eIF2Bα, the cysteine residue is replaced by alanine and thus might impart further ligand specificity. From this analysis, it can be postulated that during the course of refunctionalization, eIF2Bα underwent minimum amino acid substitution at the active-site pocket to change the ligand specificity.

The similarity in the overall structure, significant conservedness at the active-site pocket and a common evolutionary descent of R15Pi and the regulatory subunits of eIF2B indicates that the regulatory mechanism of eIF2B might utilize certain rudimentary features of R15Pi enzyme. Kuhle et al. (2015) proposed that similar to R15Pi, either all the three or fewer regulatory subunits individually undergo a closed-to-open conformation, which alters the size of the cavity formed by the NTD of eIF2B regulatory subunits (Figure 5.15A). The closing and opening of the cavity formed by the NTD, in turn modulates the binding affinity of eIF2α towards eIF2B (Kuhle et al., 2015). Based on our study, we hypothesized that the conformational change of the α-subunit of eIF2B alone would be sufficient to modulate the binding affinity of eIF2α. Out of the three regulatory subunits, the α-subunit possesses all the essential amino acid residues required

for interacting with a sugar-5-phosphate molecule. Furthermore, from the crystal structure of eIF2B (Kashiwagi et al., 2016), it can be deduced that the binding pocket of the α-subunit only remains accessible for ligand binding whereas the binding pockets of both β- and δ-subunits are completely engulfed by the catalytic subunits (γ and ε) (Figure 5.15B). Similar to PhR15Pi, the binding of a sugar-5-phosphate at the cavity formed between the NTD and CTD of the α-subunit would trigger a closed conformation (Figure 5.15C). The closed conformation is facilitated by the transition of the NTD which in turn would lead to an open cavity formed by the tips of the NTD of the regulatory subunits. This wider cavity would render a flexible interaction of eIF2α with eIF2B which might also aid in the process of GTP exchange by allowing more effective interaction between the catalytic subunits, eIF2γ and eIF2Bε (Figure 5.15C). However, during a low energy state such as glucose deprivation, the production of sugar-5-phosphates would be less. Unavailability of sugar-5-phosphate would impede the domain movement of α-subunit of eIF2B resulting in a narrow cavity (Figure 5.15C). The narrow cavity would promote a stronger interaction between eIF2α and eIF2B, thus rendering eIF2 unavailable for the subsequent rounds of translation initiation. Thus, the rate of translation initiation would be regulated in accordance to the energy state of the cell (Kuhle et al., 2015).

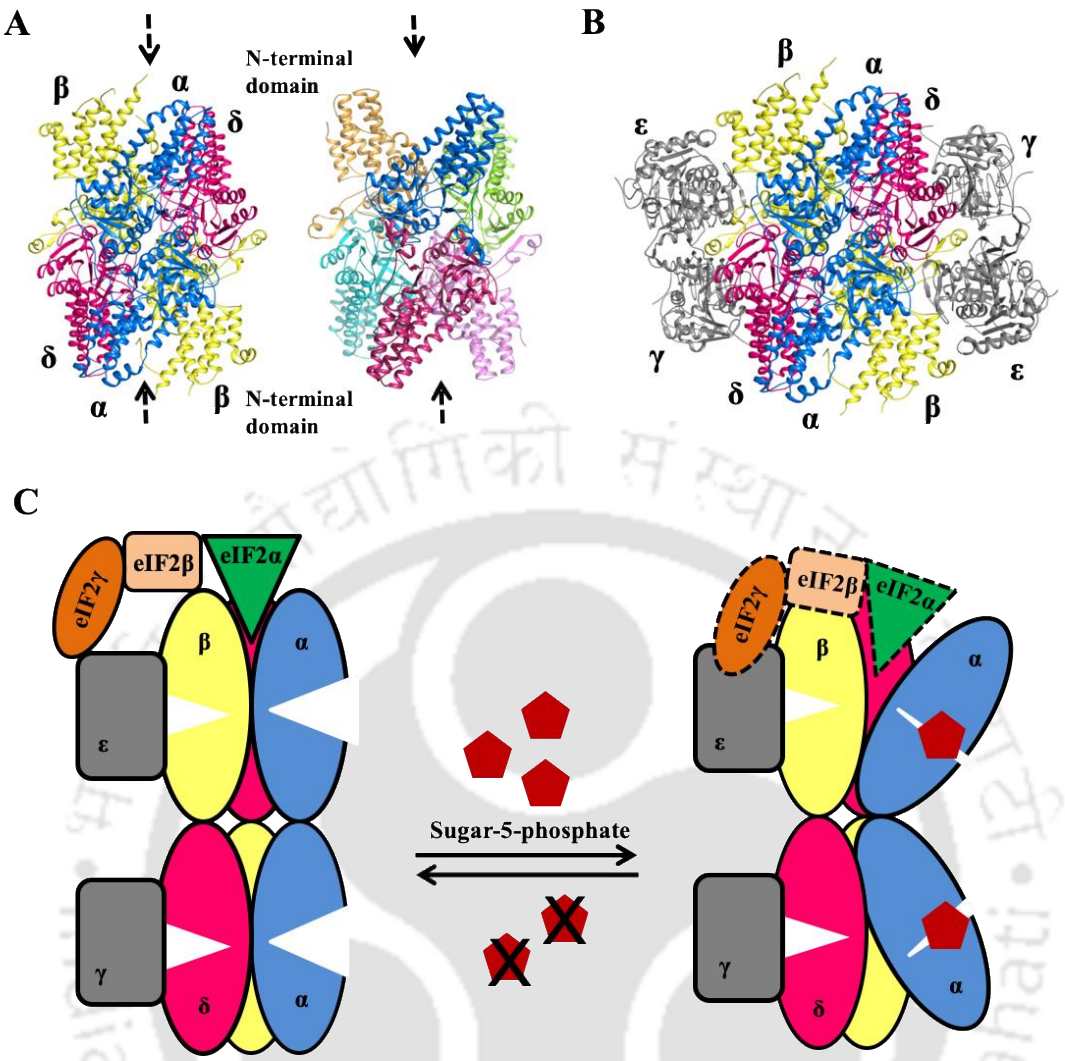



**Figure 5.15. Structural similarity between PhR15Pi and the regulatory subunits of eIF2B.** (A) Similar hexameric arrangement of the regulatory subunits of eIF2B (left) and PhR15Pi (right). The cavity formed by the NTD is shown with black dashed arrows. (B) Crystal structure of eIF2B from *S. pombe* (PDB id: 5B04). The regulatory subunits, α, β and δ subunits are shown in blue, yellow and magenta, respectively and the catalytic subunits, γ and ε are shown in grey. (C) Model for the regulatory mechanism of eIF2B. The regulatory subunits, α, β and δ are represented as blue, yellow and magenta ovals, respectively, while the catalytic subunits, γ and ε are shown as grey rounded rectangle. The α-, β- and γ-subunits of eIF2 are represented as green arrowhead, beige rounded rectangle and orange oval, respectively. Binding of sugar-5-phosphate (brick red pentagon) to the α-subunit (blue oval) of eIF2B triggers a conformational change which creates a wider binding cavity for eIF2α (green arrowhead) at the tip of the NTD of eIF2B. The dotted lines of the eIF2 complex indicate flexible binding.

## 5.5 CONCLUSION

In summary, this chapter demonstrates that PH0208 protein would function as a purine nucleotide dependent R15Pi enzyme and not as a homolog of the regulatory subunit of eIF2B. The outcome of this study hints towards the fact that a functional homolog of the regulatory subunits of eIF2B might not be available in archaea and similar to bacteria, archaea perhaps do not utilize a guanine exchange factor (GEF) during translation initiation (Pedulla et al., 2005). On the contrary, it is also likely that another unidentified ORF might accomplish the function of aIF2B in archaea. Thus, this dilemma of (un)availability of a homolog of the regulatory subunit of eIF2B impels the re-evaluation of the process of translation initiation in archaea.

# CHAPTER 6

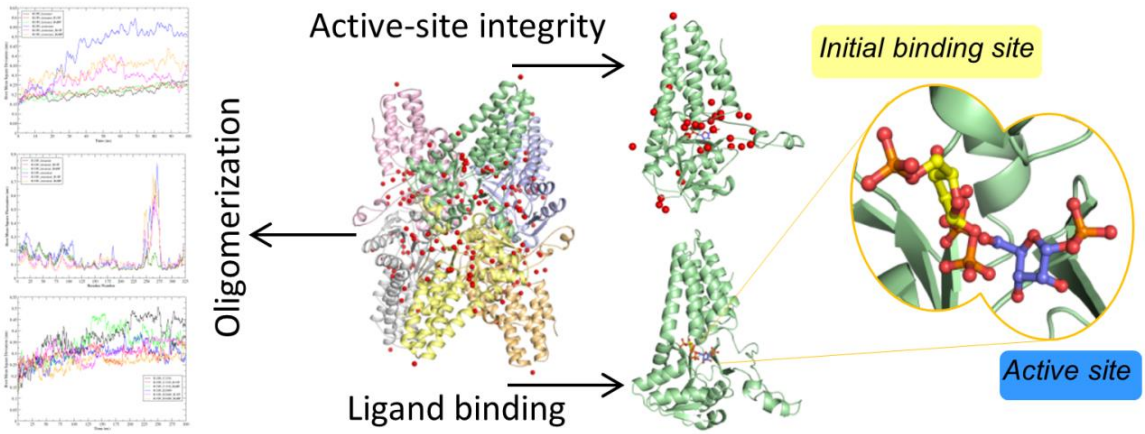
## Structure-function relationship of ribose-1,5-bisphosphate isomerase

This chapter has been published as:

Gogoi, P. and Kanaujia, S.P., 2019. Role of structural features in oligomerization, active-site integrity and ligand binding of ribose-1,5-bisphosphate isomerase. *Computational and Structural Biotechnology Journal, 17*, 333-344.

# ABSTRACT

Pentose bisphosphate pathway, exclusively found in archaea, is similar to the pentose phosphate pathway present in bacteria and eukarya. In pentose bisphosphate pathway, the conversion of ribose moieties of nucleosides into 3-phosphoglycerate (3-PGA) involves multiple steps; one of them being the conversion of ribose-1,5-bisphosphate (R15P) to ribulose-1,5-bisphosphate (RuBP) catalyzed by an enzyme ribose-1,5-bisphosphate isomerase (R15Pi). The availability of the three-dimensional structure of R15Pi had facilitated the understanding of various structural and functional aspects of the enzyme. Nevertheless, the structure of R15Pi also left several significant questions unanswered that would aid in understanding the structure-function relationship of the enzyme. Thus, we have taken up a computational approach to further understand the role of various structural features of the enzyme R15Pi. Results obtained from molecular dynamics (MD) simulations aided in understanding the obligation of the enzyme R15Pi to oligomerize and also in deciphering the role of catalytic residue(s) in structural stability. Identification of invariant water molecules of the enzyme R15Pi helped in discerning their significance at the active-site pocket and structurally important regions. Further, molecular docking studies allowed the identification of the amino acid residues essential for holding the substrate (R15P) or product (RuBP) in the vicinity of the active site of the enzyme R15Pi. Interestingly, results of the molecular docking studies assisted in the identification of an ‘alternate binding site’ (or ‘initial binding site’) for the substrate, R15P. Finally, based on these results, we propose a mechanism of ‘substrate sliding’ for the enzyme R15Pi.




## 6.1 INTRODUCTION

The pentose phosphate pathway present in bacteria and eukarya is accountable for the interconversion of pentoses and hexoses required for the nucleic acid biosynthesis and during the glycolysis as well as gluconeogenesis, respectively (Lunt et al., 2011; Stanton et al., 2012; Stincone et al., 2015). However, almost all archaea, are known to be devoid of a canonical pentose phosphate pathway (Grochowski et al., 2005; Soderberg, 2005; Orita et al., 2006). In archaea, the production of trioses and hexoses from pentoses occur via a metabolic pathway termed as pentose bisphosphate pathway (Aono et al., 2015). In this pathway, nucleosides (the main source of pentoses) are converted into trioses and hexoses via the formation of ribose-1,5-bisphosphate (R15P). The molecule R15P can either be generated through the formation of ribose-1-phosphate (R1P) with the help of the enzymes, nucleoside phosphorylase and ADP-dependent ribose-1-phosphate (ADP-R1P) kinase or via the nucleoside 5'-monophosphate (NMP) degradation pathway (Aono et al., 2015). Apart from the final step of NMP degradation pathway, where ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) produces 3-phosphoglycerate (3-PGA) to be directed towards the glycolysis and gluconeogenesis pathways, another crucial step includes the isomerization of R15P to ribulose-1,5-bisphosphate (RuBP) by an enzyme ribose-1,5-bisphosphate isomerase (R15Pi) (Finn et al., 2004; Sato et al., 2007; Aono et al., 2012, 2015; Nakamura et al., 2012).

Till date, the three-dimensional crystal structure of the enzyme R15Pi have been elucidated from only two archaea *T. kodakarensis* and *P. horikoshii* (Nakamura et al., 2012; Gogoi and Kanaujia, 2018). Structurally, the protomer of the enzyme R15Pi

comprises of a α-helical N-terminal domain (NTD) and a αβα-sandwich C-terminal domain (CTD). The active site of the enzyme R15Pi is formed at the interface of the NTD and CTD. Functionally, the enzyme R15Pi is made up of six identical protomers with interaction(s) occurring at the CTD while the NTD remains completely isolated. The available three-dimensional crystal structures of the enzyme R15Pi can be categorized in three different forms, (i) the unliganded open form (R15Pi-Form I), (ii) the substrate-bound closed form (R15Pi-Form II) and (iii) the product-bound open form (R15Pi-Form III). However, the substrate-bound structure of the wild type R15Pi (R15Pi-WT) is not available owing to the ability of the enzyme to spontaneously convert R15P to RuBP during the crystallization process. Thus, to obtain substrate-bound enzyme (R15Pi-Form II), mutation of one of the two crucial catalytic residues was carried out (Nakamura et al., 2012; Gogoi and Kanaujia, 2018).

In the absence of the substrate (R15P), the enzyme R15Pi retains an open conformation with a completely accessible active site. Upon R15P binding, the NTD move towards the CTD, resulting in the shielding of the active-site cleft from the aqueous environment (Nakamura et al., 2012). Once the conversion of the substrate (R15P) to product (RuBP) is accomplished, the NTD reverts back to its initial position propelling the enzyme into its open conformation, which also allows the product (RuBP) to exit the active site. This conversion process of R15P to RuBP requires the catalytic activity of two residues Cys133 and Asp202 (numbering according to R15Pi from *T. kodakarensis*, TkR15Pi) via the formation of a *cis*-phosphoenolate intermediate. Although, point mutation of either of these residues abolishes the enzymatic activity completely, it does not hinder the binding of R15P or RuBP to the enzyme (Nakamura et al., 2012; Gogoi and Kanaujia, 2018).

Although the three-dimensional structure of the enzyme R15Pi has been elucidated (Nakamura et al., 2012; Gogoi and Kanaujia, 2018), the accountability of the enzyme to function as a hexamer, recognition of the essential catalytic residue for structural stability, identification of essential substrate- and/or product-binding residues and structural as well as functional role(s) of invariant water molecules remains undescribed. Thus, in this chapter, we employed computational approaches such as molecular dynamics (MD) simulation, molecular docking and data mining to address these aspects and consequently to obtain an insight into the role of the structural features of the enzyme R15Pi in rendering adequate stability and optimal ligand binding environment.

## 6.2 MATERIALS AND METHODS

### 6.2.1 MOLECULAR DYNAMICS SIMULATION

MD simulation was performed using the package GROMACS v.5.1.4 (Abraham et al., 2015). For all the simulations (Table 6.1) AMBER03 force field embedded in GROMACS was utilized (Ponder and Case, 2003). The module *editconf* of GROMACS was used to generate a cubic box keeping the minimum distance between the solute and the edge of the box at least 1.0 nm. The module *genbox* present in the GROMACS suite was used to solvate the protein models with the flexible simple point charge (SPC/E) water model. Chloride and sodium ions were utilized to neutralize the overall charge of the system. To inhibit edge effects, periodic boundary condition (PBC) was applied along the three spatial directions. Energy minimization was performed using the steepest descent method with a maximum force cutoff of 1000 kJ $mol^{-1}$ $nm^{-1}$. The solvent and ions around the protein were equilibrated in two phases. Firstly, equilibrations were performed under an NVT (canonical or isothermal-isochoric) ensemble for 100 ps with a reference temperature of 37°C. Subsequently, second phase of equilibration was performed under an NPT (isothermal-isobaric) ensemble for 500 ps with a reference pressure of 1 bar. A velocity rescaling thermostat (Bussi et al., 2007) with coupling constant of 0.1 ps was used to control the reference temperature at 310 K. The reference pressure of 100 kPa (1 bar) with coupling constant 2 ps was controlled using the Parrinello-Rahman barostat (Parnello and Rahman, 1981). The long-range electrostatic interactions were computed using the particle mesh Ewald (PME) method (Darden et al., 1993; Essmann et al., 1995) while the short-range van der Waals interactions were computed with using Verlet neighbor list calculation with a cutoff of 0.8 nm. P-LINCS algorithm (Hess, 2008) was used to constrain the bond lengths and a time step of 2 fs was used to integrate the equations of motion. Each MD simulation was performed for a time period of 100 ns and 300 ns for R15Pi from *T. kodakarensis* (TkR15Pi) and *P. horikoshii* (PhR15Pi), respectively, resulting in a total of 2.6 μs of simulation. The analyses were performed using the programs available in GROMACS and home-built shell scripts. The program Xmgrace (Turner, 2005) was used to prepare the graphs.

**Table 6.1.** Details of three-dimensional crystal structures of R15Pi used during MD simulations.

| Name of simulated systems | PDB id | Wild/ Mutant | [a]Original bound ligand | Bound ligand during simulation | Oligomeric state | Time (ns) |
|---|---|---|---|---|---|---|
| R15Pi_hexamer | 3A9C | WT | RuBP | - | Hexamer | 100 |
| R15Pi_hexamer_R15P | | | | R15P | | |
| R15Pi_hexamer_RuBP | | | | RuBP | | |
| R15Pi_monomer | | | | - | Monomer | |
| R15Pi_monomer_R15P | | | | R15P | | |
| R15Pi_monomer_RuBP | | | | RuBP | | |
| Lys24 | 3A11 | | - | - | | 300 |
| Lys213_Arg254 | | | | - | | |
| R15Pi_C135S | 5YFS | C135S | R15P | - | | |
| R15Pi_C135S_R15P | | | | R15P | | |
| R15Pi_C135S_RuBP | | | | RuBP | | |
| R15Pi_D204N | 5YFT | D204N | | - | | |
| R15Pi_D204N_R15P | | | | R15P | | |
| R15Pi_D204N_RuBP | | | | RuBP | | |

[a]R15P: ribose-1,5-bisphosphate; RuBP: ribulose-1,5-bisphosphate

### 6.2.2 IDENTIFICATION OF INVARIANT WATER MOLECULES

The three-dimensional atomic coordinates of 13 crystal structures of PhR15Pi are available in closed conformation. These structures include R15Pi-WT, R15Pi-C135S and R15Pi-D204N proteins bound to either R15P or RuBP (Table 6.2). However, only those structures which harbor greater than or equal to 100 water molecules were considered for the identification of invariant water molecules. For identification of invariant water molecules in TkR15Pi, all the three available structures were considered even though they contain less than 100 water molecules. All the structures of PhR15Pi belong to the space group $P3_112$ and consist of three subunits in the asymmetric unit (ASU). The three monomeric units in each of the eight structures were considered as an independent molecule. Thus, a total of 18 monomers of PhR15Pi were taken as data set for the analysis. The crystal structure deposited with PDB id 5YFT (chain B) was taken as the reference or fixed molecule as it contains the highest number (236) of water molecules among the considered structures. All the remaining structures were considered as mobile molecules and superimposed on 5YFT (chain B) to identify the invariant water molecule(s). The superimposition of the structures was manually performed in the program COOT (Emsley amnd Cowtan, 2004) as well as using a home-built shell script. The cutoff distance between a pair of superposed water molecules was kept as 1.8 Å. A water molecule was considered to be invariant only when it was present in all the structures considered. An invariant water molecule was considered to be buried when it had a solvent accessible surface area (SASA) of less than equal to 2.5 $Å^2$ (Kanaujia and Sekar, 2009). For calculation of the residence frequency of the invariant water molecules, MD simulation was performed for a time period of 300 ns. The structures generated during MD simulation at every 1 ns were used for comparison and to compute the residence frequency of each invariant water molecule. The interaction(s) between protein atoms and solvent molecules were calculated with a hydrogen-bond distance of 3.5 Å. All the other parameters of MD simulation were kept same as mentioned in the previous section.

**Table 6.2.** List of three-dimensional crystal structures of R15Pi and their details used for the identification of invariant water molecules.

| PDB id | Chain id | Wild / Mutant | Space group | Resolution (Å) | $R_{work}$/ $R_{free}$ (%) | No. of waters | [a]Bound ligand |
|---|---|---|---|---|---|---|---|
| 5YFT | A | D204N | $P3_112$ | 2.21 | 15.1/ 19.3 | 230 | R15P |
| | B | | | | | 236 | |
| | C | | | | | 203 | |
| 5YFS | A | C135S | | 2.30 | 14.4/ 19.8 | 210 | R15P |
| | B | | | | | 196 | |
| | C | | | | | 209 | |
| 5YFJ | A | WT | | 2.20 | 15.0/ 19.0 | 183 | RuBP |
| | B | | | | | 179 | |
| | C | | | | | 145 | |
| 5YG6 | A | C135S | | 2.35 | 15.5/ 20.8 | 172 | R15P, GMP |
| | B | | | | | 149 | |
| | C | | | | | 178 | |
| 5YFW | A | D204N | | 2.35 | 15.6/ 21.3 | 163 | R15P, AMP |
| | B | | | | | 165 | |
| | C | | | | | 136 | |
| 5YFU | A | WT | | 2.35 | 18.0/ 24.8 | 135 | RuBP, AMP |
| 5YGA | A | D204N | | 2.45 | 16.8/ 22.5 | 114 | R15P, AMP, GMP |
| | B | | | | | 110 | |
| 5YG8 | A | WT | | 2.80 | 15.5/ 22.7 | 102 | RuBP, AMP, GMP |

[a]R15P: ribose-1,5-bisphosphate; RuBP: ribulose-1,5-bisphosphate; GMP: guanosine 5′-monophosphate; AMP: adenosine 5′-monophosphate.

### 6.2.3 MOLECULAR DOCKING STUDIES

To perform molecular docking experiments, the three-dimensional atomic coordinates of the two open forms of R15Pi i.e. R15Pi-Form I (PDB id: 3A11) and R15Pi-Form III (PDB id: 3A9C) from *T. kodakarensis* were extracted from the RCSB Protein Data Bank (PDB) (Berman et al., 2000). Since R15Pi-Form II (PDB id: 3VM6) has a closed active site, only the open forms of the protein were used for docking experiments. The ligands (R15P and RuBP) used for the docking experiments were extracted from their bound states of the enzyme TkR15Pi. Molecular docking experiments were carried out using the program Autodock version 4.0 (Morris et al., 2009). During each docking experiment, the receptor (i.e. protein) molecule was kept as rigid while the ligand (i.e. R15P and RuBP) molecules were considered both as rigid and flexible along the rotatable bonds. As both the rigid and flexible ligands produced similar outcomes, results of only the rigid-ligand docking were considered for analysis. The size of the grid box was kept to 126x126x126 with a spacing of 0.375 Å between the grid points and taking the center of mass of the protein as the grid center. To search for the best conformational space of the ligand, the Lamarckian genetic algorithm (LGA) with a total of 2000 runs was performed. The docked conformations of ligands in each molecular docking experiment were clustered with a RMSD cutoff of 2.0 Å. The ligand conformation having the lowest estimated free energy of binding (EFBE) was selected as the final docked ligand to the protein. Identification of the amino acid residues involved in the interaction(s) with the docked ligand was performed using the program COOT (Emsley and Cowtan, 2004). All the figures of the docked ligands were generated using the program PyMOL (PyMOL Molecular Graphics System, Schrodinger, LLC).

## 6.3 RESULTS

### 6.3.1 THE HEXAMERIC STATE OF R15Pi

To understand the significance of the oligomeric state of R15Pi, MD simulations of R15Pi-monomer and R15Pi-hexamer were performed. The R15Pi-hexamer is found to be more stable than the R15Pi-monomer during the simulation (Figure 6.1A). Nonetheless, binding of either of the ligands i.e. R15P or RuBP to the R15Pi-monomer reasonably increases the stability of the enzyme (Figure 6.1A). On the other hand, R15Pi-hexamer

remains stable throughout the simulation irrespective of the presence or absence of ligand (Figure 6.1A). The stability of the hexameric protein is further supported by the comparison of the potential energies of the R15Pi-monomer and R15Pi-hexamer (Figure 6.1B).

An analysis of the root mean square fluctuations (RMSFs) exhibits that the NTD of R15Pi-hexamer displays a slightly higher fluctuation than that of the R15Pi-monomer. In contrast, the RMSFs of the CTD are notably lower for R15Pi-hexamer than that of the R15Pi-monomer. The most significant fluctuation of the CTD in the R15Pi-monomer occurs at a long loop region (residues 245-283) of the protein. Interestingly, this fluctuation is somewhat decreased in the presence of either the substrate (R15P) or the product (RuBP) bound at the active site of the protein (Figure 6.1C).

To find whether the oligomeric state of R15Pi had any effect on the binding of the substrate or product, the distance between protein and the ligand (R15P or RuBP) was computed during the 100 ns simulation. Results show a lesser distance between them in the case of R15Pi-hexamer compared to that of the R15Pi-monomer suggesting a better ligand binding to the hexameric protein (Figure 6.1D).

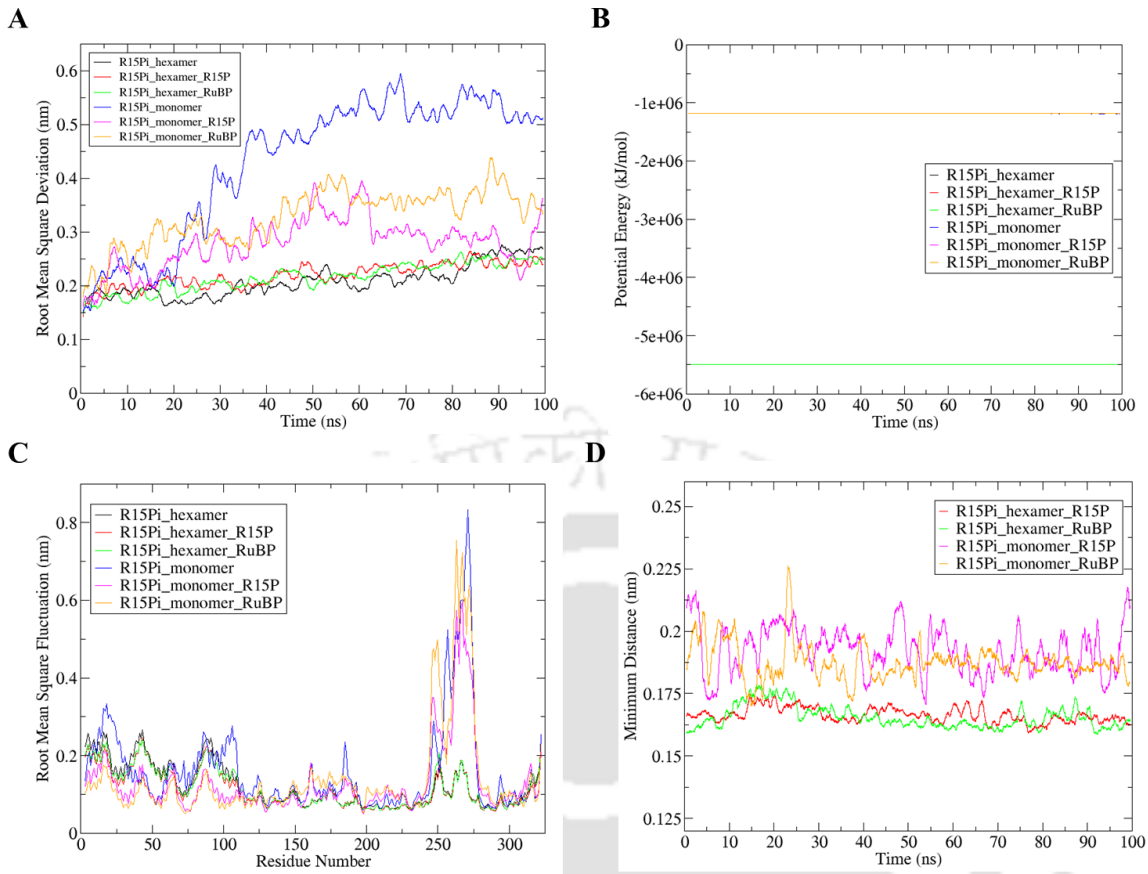



**Figure 6.1. Dynamics of R15Pi-monomer and R15Pi-hexamer.** The plot for the (A) root mean square deviation (RMSD), (B) potential energy, (C) root mean square fluctuation (RMSF) and (D) the minimum distance between the protein and ligand. The different forms of the enzyme R15Pi for which simulations were performed include R15Pi-monomer (blue), R15P-bound R15Pi-monomer (magenta), RuBP-bound R15Pi-monomer (orange), R15Pi-hexamer (black), R15P-bound R15Pi-hexamer (red) and RuBP-bound R15Pi-hexamer (green).

### 6.3.2 ROLE OF CATALYTIC RESIDUE IN STRUCTURAL STABILITY

To identify the preferred catalytic residue between Cys135 and Asp204 (Cys133 and Asp202 in TkR15Pi) for the structural stability of enzyme PhR15Pi, MD simulations of the mutant proteins R15Pi-C135S and R15Pi-D204N, both in bound and unbound forms, were performed. The RMSD plot reveals that the R15Pi-C135S mutant protein displays a relatively higher instability in the structure as compared to R15Pi-D204N mutant protein (Figure 6.2A). However, in the presence of either R15P or RuBP, the mutant gains structural stability and exhibits a lower RMSD (Figure 6.2A). On the other hand, in case of R15Pi-D204N mutant, the presence or absence of ligand (R15P or RuBP) shows no effect on the stability of the protein (Figure 6.2A). This is further supported by

comparison of the potential energy of two mutants, where the mutant R15Pi-C135S exhibits a higher potential energy than that of the mutant R15Pi-D204N (Figure 6.2B), indicating the importance of the cysteine residue in rendering structural stability.

In the three-dimensional crystal structure of R15Pi-C135S mutant protein, Ser135 interacts with Gln166 (bond distance: 2.8 Å) while in R15Pi-D204N, no such interactions occur between Cys135 and Gln166 (bond distance: 4.2 Å) (Figure 6.2C). Hence, in R15Pi-D204N mutant protein, Gln166 is available to interact with Lys164 which in turn coordinates with Glu260 present in the longest loop region. Glu260 further interacts with another amino acid residue Arg256 which is also present in the longest loop region. Thus, Cys135 is involved in indirectly rendering stability to the highly flexible longest loop region by adhering it to the core CTD. However, hydrogen bond formation between Ser135 and Gln166 in R15Pi-C135S mutant protein makes Gln166 unavailable for further interactions (Figure 6.2C). The higher fluctuation in the longest loop region in case of R15Pi-C135S mutant protein due to the lack of stabilizing interactions is also evident from the RMSF plot (Figure 6.2D). The fluctuation also decreases in the presence of either the substrate or product owing to their interaction with Arg256 which resides in the longest loop region (Figure 6.2D, E).

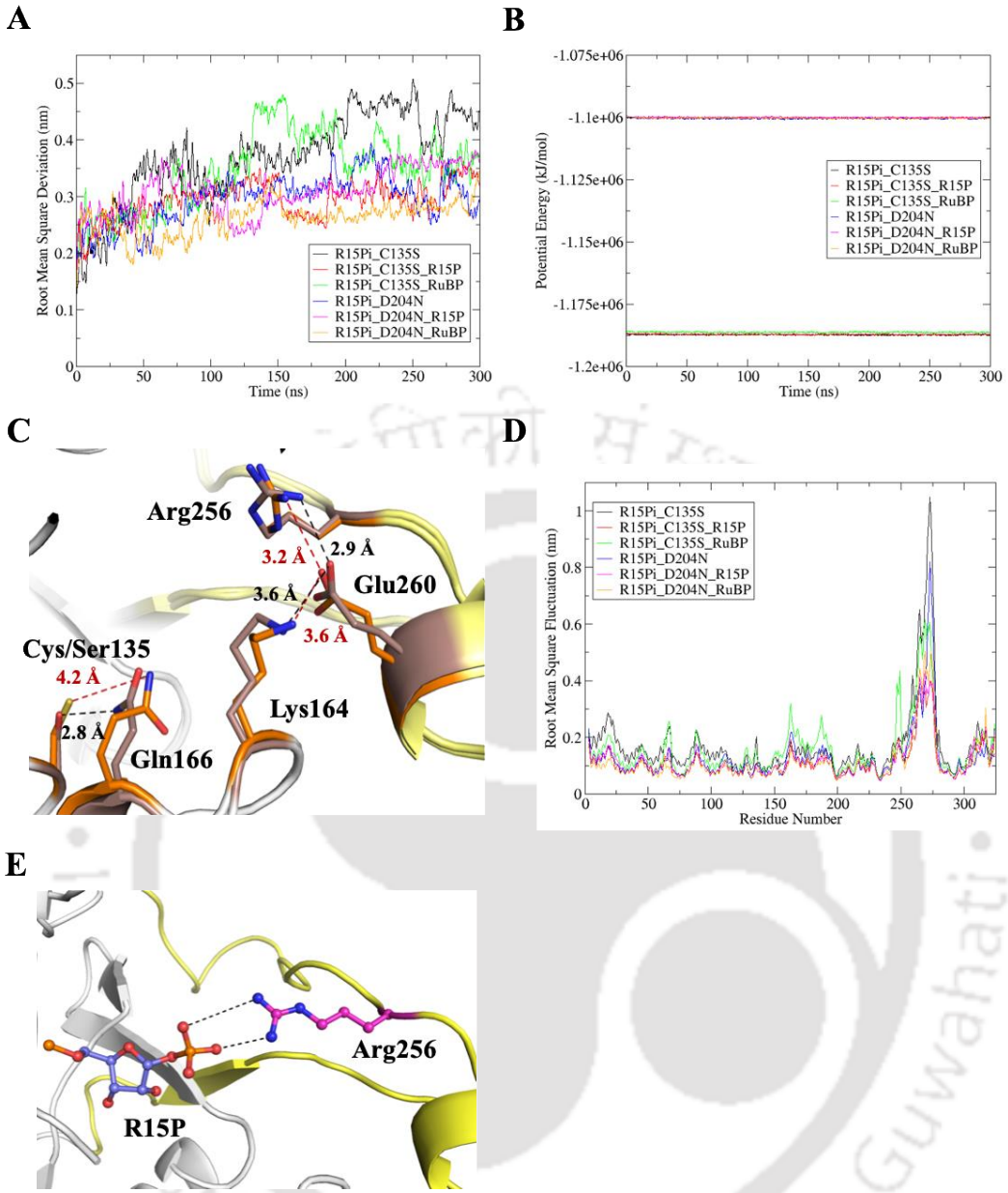



**Figure 6.2. Dynamics of the mutants R15Pi-C135S and R15Pi-D204N.** The plot of (A) root mean square deviation (RMSD) and (B) potential energy of the mutant proteins R15Pi-C135S and R15Pi-D204N during the MD simulation. (C) Absence (presence) of interaction between Gln166 and Cys135 (Ser135) in R15Pi-C135S (R15Pi-D204N) mutant protein(s). The amino acid residues involved in conferring structural stability in R15Pi-C135S (orange) and R15Pi-D204N (light brown) mutant proteins are shown as lines and their interactions are depicted as red and black dashed lines. The distances between the atoms are indicated adjacent to their corresponding dashed lines. The flexible longest loop region is highlighted in yellow. The amino acid residues are numbered according to the enzyme PhR15Pi. (D) The plot of root mean square fluctuation (RMSF) of the mutant proteins R15Pi-C135S (black) and R15Pi-D204N (blue) during the MD simulation. (E) Interaction between Arg256 (magenta ball-and-stick model) residing in the longest loop region (yellow) and R15P (blue ball-and-stick model) are shown with black dashed lines. The amino acid residue is numbered

according to the enzyme PhR15Pi. The different forms of the enzyme R15Pi for which simulations were performed include R15Pi-C135S (black), R15P-bound R15Pi-C135S (red), RuBP-bound R15Pi-C135S (green), R15Pi-D204N (blue), R15P-bound R15Pi-D204N (magenta) and RuBP-bound R15Pi-D204N (orange).

### 6.3.3 INVARIANT WATER MOLECULES

A total of 29 water molecules were identified to be invariant in a protomer of the closed conformation of PhR15Pi. Out of these, 12 invariant water molecules are present either in the active-site pocket or in its vicinity (Table 6.3), five (IW1, IW2, IW3, IW4 and IW5) of which make direct interaction with both R15P and RuBP in a similar manner. For instance, IW2 is hydrogen bonded to the 1-phosphate group of the ligand while the 5-phosphate group is held by three other invariant water molecules IW1, IW3 and IW5. The invariant water molecule IW4 forms hydrogen bond with the O3 oxygen atom of the ribose sugar (Figure 6.3A and Table 6.3). All these water molecules are highly stable as indicated by their low *B* factors (Table 6.4). Moreover, these invariant water molecules stabilize residues in the vicinity of the active site of the enzyme. One of the invariant water molecules IW2 forms hydrogen bond with $O^{\delta 1}$ atom of the catalytic residue Asp204 of R15Pi-WT. Similarly, invariant water molecules IW1, IW3, IW4, IW5 and IW6 form hydrogen bonds with the residues Arg26, His134, Gly202, Ala203, Asn214, Lys215, Ala235, Glu237 and Lys240 (Figure 6.3B and Table 6.3).

A total of another five invariant water molecules (IW7, IW8, IW9, IW10 and IW11) do not directly interact with the ligand, however, make contact with amino acid residues which form the ‘active-site roof’ of the enzyme. The water molecule IW8 is hydrogen bonded to backbone O atoms of Ile16 and Ile21 as well as to $N^{\eta 2}$ atom of Arg65. The invariant water molecule IW9 hydrogen bonds with the backbone O atom of Pro66 and networks three other water molecules (W585, W666 and W701 of PhR15Pi). Two invariant water molecules IW7 and IW11 are hydrogen bonded to the backbone O and $N^{\eta 2}$ atoms of Arg22 while IW10 holds the residue Arg29 through $N^{\eta 1}$ and $N^{\eta 2}$ atoms (Figure 6.3C and Table 6.3). Interestingly, only one invariant water molecule (IW12) interacts with amino acid residues forming the ‘active-site floor’ of the enzyme. More precisely, IW12 interacts with the N atom of Pro163 & Lys164, O atom of Thr161 & Arg162 and $O^{\varepsilon 1}$ atom of Gln166. These five amino acid residues are part of a loop region which shield the active-site pocket from underneath (Figure 6.3D and Table 6.3).

Notably, three invariant water molecules IW4, IW5 and IW12 are present in both open and closed conformations of the enzyme TkR15Pi as well, signifying their importance.

**Table 6.3.** List of hydrogen bond interaction(s) between invariant water molecule(s) and amino acid residue(s) at the active-site pocket as well as in its vicinity.

| [a]Water number | Invariant water | Location | [b]List of interaction (distance in Å) |
|---|---|---|---|
| 526 | IW1 | Active-site pocket | R15P 5-$PO_4^{2-}$ (2.6), W630 (2.7), W520 (2.9), Arg26 N (3.0) |
| 549 | IW2 | | R15P 1-$PO_4^{2-}$ (2.7), Asn204 $O^{\delta1}$ (3.0), W687 (3.1) |
| 553 | IW3 | | Glu237 $O^{\varepsilon2}$ (2.7), R15P 5-$PO_4^{2-}$ (2.8), W630 (2.9) |
| 554 | IW4 | | R15P O3 (2.7), His134 O (2.9), Gly202 N (3.0), Asn214 $N^{\delta2}$ (3.0) |
| 565 | IW5 | | Ala235 O (2.7), R15P 5-$PO_4^{2-}$ (2.8), Lys240 $N^{\zeta}$ (2.8), Ala203 N (3.0) |
| 599 | IW6 | | W586 (2.7), Lys215 $N^{\zeta}$ (2.8), W580 (2.9) |
| 520 | IW7 | 'Active-site roof' | Arg22 O (2.6), W568 (2.7), W526 (2.9) |
| 522 | IW8 | | Ile16 O (2.6), Ile21 O (2.7), Arg65 $N^{\eta2}$ (2.7) |
| 539 | IW9 | | Pro66 O (2.6), W585 (2.7), W701 (2.7), W666 (3.0) |
| 630 | IW10 | | W526 (2.7), Arg29 $N^{\eta2}$ (2.9), W553 (2.9), Arg29 $N^{\eta1}$ (3.1), W682 (3.1) |
| 653 | IW11 | | W727 (2.7), W649 (2.9), Arg22 $N^{\eta2}$ (3.0), W726 (3.1) |
| 580 | IW12 | 'Active-site floor' | Gln166 $O^{\varepsilon1}$ (2.7), W599 (2.9), Arg162 O (3.0), Pro163 N (3.1), Thr161 O (3.2), Lys164 N (3.2) |
| 527 | IW13 | Between CTD and the long loop | Pro163 O (2.6), W671 (2.6), Glu260 O (2.7), W648 (3.2) |
| 603 | IW14 | Between | W687 (2.5), Glu254 $O^{\varepsilon1}$ (2.8), Thr67 N (2.9) |

| | | | |
|---|---|---|---|
| | | NTD and the long loop | |
| 541 | IW15 | Long loop | Pro279 O (2.6), Asn278 $O^{\delta1}$ (2.8), W576 (2.9) |
| 618 | IW16 | | Arg256 O (2.9), Asn278 N (3.0) |
| 632 | IW17 | | W535 (2.6), Met255 N (2.9) |
| 665 | IW18 | | Asp257 N (3.2) |
| 537 | IW19 | Longest helix $\alpha_5$ | Lys123 $N^{\zeta}$ (2.6), Arg116 $N^{\eta1}$ (3.0) |
| 624 | IW20 | | W697 (2.7), Asn109 O (2.9), W664 (3.0) |
| 641 | IW21 | | W625 (2.4), Leu92 N (3.0) |
| 530 | IW22 | Inter-subunit region | Asp290 $O^{\delta1}$ (2.6), W551 (2.8), Glu287 O (2.9), Lys226 $N^{\zeta}$ (3.0) |
| 567 | IW23 | | W688 (2.7), Asp290 $O^{\delta2}$ (2.7), Val230 O (2.8) |
| 595 | IW24 | | W715 (2.4), Glu126 $O^{\varepsilon1}$ (2.8), W692 (2.8), Asp129 $O^{\delta1}$ (2.9), W730 (3.2) |
| 708 | IW25 | | W525 (2.1), W623 (2.5), W616 (2.6), W733 (2.6) |
| 715 | IW26 | | W674 (2.3), W730 (2.3), W595 (2.4), W733 (2.4), W616 (2.5), W623 (2.5) |
| 628 | IW27 | Helix $\alpha_7$ | W569 (2.5), W695 (2.8), Ile169 O (2.9), W706 (3.0) |
| 680 | IW28 | Helix $\alpha_4$ | W596 (3.2), Arg81 $N^{\eta1}$ (3.3) |
| 687 | IW29 | Between loops connecting helices $\alpha_1$-$\alpha_2$ & $\alpha_3$-$\alpha_4$ | W603 (2.5), W559 (2.8), W549 (3.1) |

[a] The numbering scheme is taken from the reference structure (PDB id: 5YFT and chain B).
[b]R15P: ribose-1,5-bisphosphate.

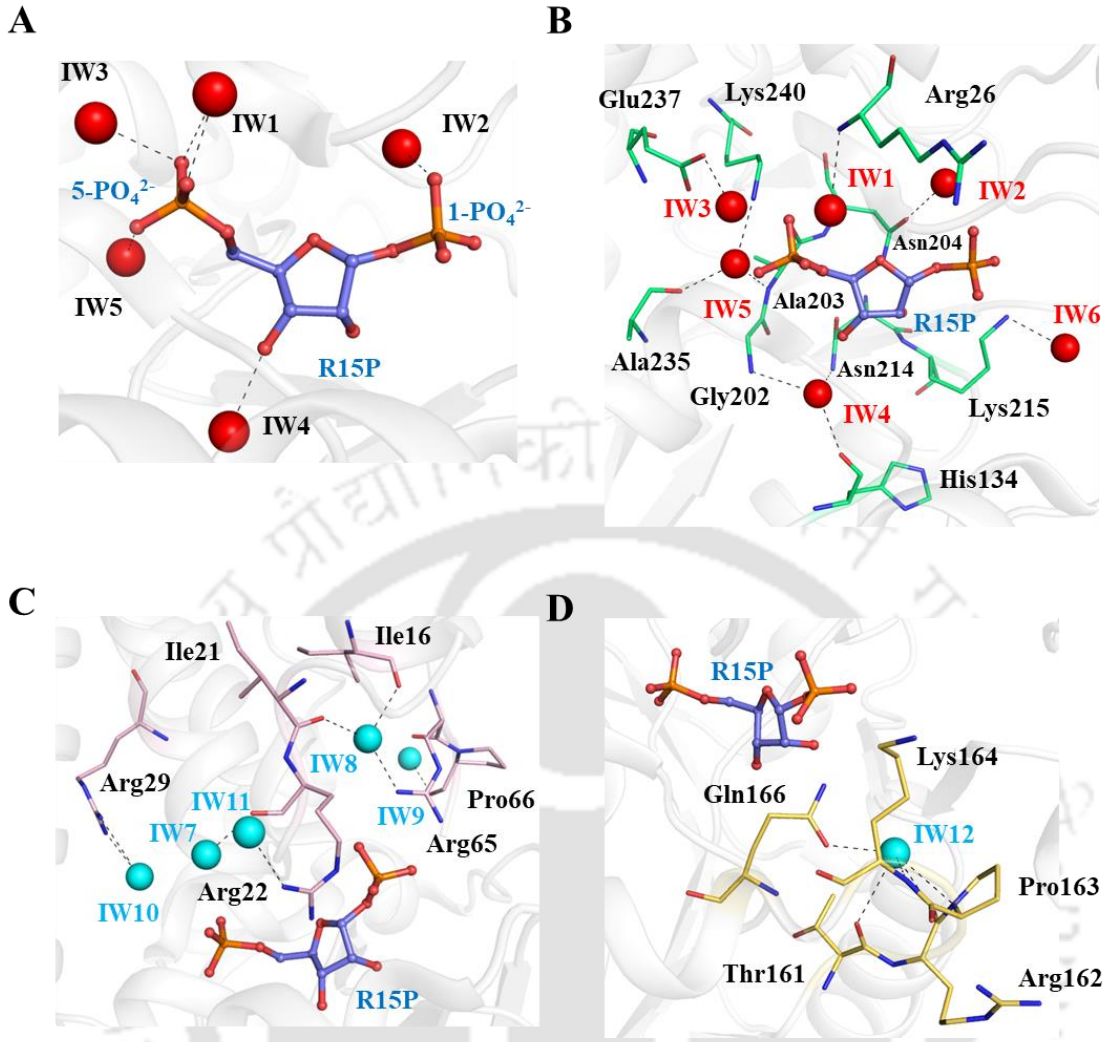



**Figure 6.3. Invariant water molecules at the active site and in its vicinity.** (A) Invariant water molecules (red sphere) interacting directly with R15P (blue ball-and-stick model) and (B) amino acid residues (green lines) at the active-site pocket. (C, D) Invariant water molecules (cyan sphere) interacting with the amino acid residues forming the 'active-site roof' (pink lines) and 'active-site floor' (yellow lines), respectively, of the protein. The interaction(s) between invariant water molecule and amino acid residues are shown in black dashed lines. All the amino acid residues are numbered according to the enzyme PhR15Pi.

**Table 6.4.** Invariant water molecules corresponding to the reference structure (PDB id: 5YFT, chain B) along with their B factor, solvent accessible surface area (SASA) and percentage residence frequency calculated from the ensembles generated using MD simulation.

| | Chain id | IW1 | IW2 | IW3 | IW4 | IW5 | IW6 |
|---|---|---|---|---|---|---|---|
| **5YFT** | A | 542 | 595 | 589 | 594 | 548 | 622 |
| | C | 535 | 548 | 514 | 568 | 553 | 556 |
| **5YFS** | A | 515 | 548 | 545 | 575 | 570 | 581 |
| | B | 521 | 530 | 508 | 580 | 546 | 617 |
| | C | 515 | 589 | 571 | 568 | 573 | 655 |
| **5YFJ** | A | 546 | 559 | 555 | 583 | 552 | 617 |
| | B | 506 | 520 | 541 | 554 | 549 | 558 |
| | C | 523 | 517 | 511 | 563 | 542 | 567 |
| **5YG6** | A | 509 | 565 | 564 | 555 | 533 | 570 |
| | B | 514 | 523 | 512 | 564 | 530 | 601 |
| | C | 530 | 593 | 594 | 575 | 569 | 629 |
| **5YFW** | A | 528 | 583 | 550 | 566 | 542 | 623 |
| | B | 506 | 563 | 554 | 528 | 543 | 605 |
| | C | 549 | 523 | 508 | 556 | 538 | 559 |
| **5YFU** | A | 557 | 571 | 518 | 545 | 526 | 587 |
| **5YGA** | A | 525 | 559 | 534 | 542 | 524 | 568 |
| | B | 501 | 542 | 548 | 519 | 517 | 563 |
| **5YG8** | A | 535 | 527 | 514 | 519 | 516 | 546 |
| ***B* factor (Å$^2$)** | | 14.65 | 19.40 | 16.50 | 13.99 | 13.54 | 18.77 |
| **[a]SASA (Å$^2$)** | | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **[b]RF (%)** | | 92.0 | 67.0 | 7.0 | 65.0 | 79.0 | 26.0 |

**Table 6.4. (Continued).**

| | Chain id | IW7 | IW8 | IW9 | IW10 | IW11 | IW12 |
|---|---|---|---|---|---|---|---|
| **5YFT** | A | 562 | 508 | 585 | 643 | 670 | 582 |
| | C | 573 | 509 | 572 | 611 | 626 | 596 |
| **5YFS** | A | 563 | 505 | 528 | 614 | 627 | 639 |
| | B | 535 | 509 | 565 | 639 | 638 | 626 |
| | C | 529 | 513 | 565 | 622 | 646 | 630 |
| **5YFJ** | A | 577 | 511 | 551 | 604 | 637 | 560 |
| | B | 536 | 515 | 509 | 619 | 635 | 540 |
| | C | 539 | 508 | 553 | 601 | 606 | 541 |
| **5YG6** | A | 538 | 523 | 547 | 598 | 626 | 545 |
| | B | 548 | 505 | 540 | 590 | 604 | 567 |
| | C | 546 | 508 | 539 | 579 | 619 | 571 |
| **5YFW** | A | 554 | 508 | 522 | 576 | 617 | 575 |
| | B | 541 | 526 | 518 | 611 | 617 | 562 |
| | C | 565 | 507 | 552 | 582 | 603 | 575 |
| **5YFU** | A | 546 | 522 | 542 | 563 | 577 | 527 |
| **5YGA** | A | 535 | 515 | 531 | 570 | 583 | 537 |
| | B | 512 | 507 | 504 | 580 | 573 | 540 |
| **5YG8** | A | 534 | 507 | 568 | 551 | 564 | 533 |
| ***B* factor (Å$^2$)** | | 17.24 | 14.31 | 27.66 | 19.15 | 21.15 | 15.86 |
| **[a]SASA (Å$^2$)** | | 2.34 | 0.00 | 0.00 | 0.00 | 1.34 | 0.00 |
| **[b]RF (%)** | | 67.0 | 16.0 | 9.0 | 43.0 | 71.0 | 31.0 |

**Table 6.4. (Continued).**

| | Chain id | IW13 | IW14 | IW15 | IW16 | IW17 | IW18 |
|---|---|---|---|---|---|---|---|
| **5YFT** | A | 509 | 533 | 566 | 601 | 637 | 664 |
| | C | 515 | 554 | 512 | 590 | 615 | 644 |
| **5YFS** | A | 520 | 573 | 561 | 568 | 620 | 656 |
| | B | 527 | 577 | 540 | 582 | 610 | 613 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | C | 506 | 564 | 557 | 598 | 599 | 637 |
| **5YFJ** | A | 513 | 528 | 562 | 586 | 582 | 634 |
| | B | 519 | 578 | 571 | 580 | 621 | 640 |
| | C | 516 | 546 | 536 | 576 | 585 | 616 |
| **5YG6** | A | 529 | 558 | 585 | 527 | 591 | 622 |
| | B | 522 | 554 | 571 | 542 | 608 | 582 |
| | C | 514 | 535 | 541 | 591 | 551 | 617 |
| **5YFW** | A | 516 | 537 | 557 | 580 | 581 | 596 |
| | B | 533 | 549 | 572 | 557 | 596 | 630 |
| | C | 518 | 540 | 542 | 547 | 590 | 577 |
| **5YFU** | A | 509 | 524 | 558 | 578 | 575 | 586 |
| **5YGA** | A | 512 | 510 | 533 | 541 | 572 | 573 |
| | B | 511 | 527 | 549 | 543 | 566 | 576 |
| **5YG8** | A | 510 | 518 | 517 | 521 | 538 | 539 |
| ***B* factor (Å$^2$)** | | 18.95 | 20.78 | 21.42 | 22.02 | 28.23 | 19.77 |
| **[a]SASA (Å$^2$)** | | 6.28 | 0.00 | 2.09 | 7.71 | 20.51 | 23.26 |
| **[b]RF (%)** | | 28.0 | 73.0 | 0.00 | 6.0 | 25.0 | 19.0 |

**Table 6.4. (Continued).**

| | **Chain id** | **IW19** | **IW20** | **IW21** | **IW22** | **IW23** | **IW24** |
|---|---|---|---|---|---|---|---|
| **5YFT** | A | 655 | 518 | 662 | 544 | 534 | 581 |
| | C | 501 | 537 | 648 | 539 | 525 | 508 |
| **5YFS** | A | 571 | 578 | 653 | 507 | 555 | 522 |
| | B | 502 | 554 | 636 | 506 | 516 | 504 |
| | C | 629 | 562 | 656 | 512 | 532 | 590 |
| **5YFJ** | A | 610 | 512 | 623 | 558 | 533 | 564 |
| | B | 505 | 610 | 638 | 560 | 548 | 550 |
| | C | 580 | 525 | 602 | 543 | 509 | 505 |
| **5YG6** | A | 519 | 586 | 692 | 540 | 537 | 531 |
| | B | 574 | 551 | 609 | 526 | 509 | 502 |
| | C | 599 | 526 | 626 | 555 | 511 | 524 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **5YFW** | A | 502 | 511 | 598 | 517 | 504 | 531 |
| | B | 511 | 599 | 625 | 508 | 538 | 560 |
| | C | 599 | 513 | 602 | 544 | 506 | 503 |
| **5YFU** | A | 566 | 516 | 596 | 574 | 502 | 534 |
| **5YGA** | A | 558 | 514 | 585 | 565 | 508 | 513 |
| | B | 515 | 544 | 587 | 551 | 528 | 538 |
| **5YG8** | A | 560 | 554 | 550 | 544 | 502 | 511 |
| ***B* factor (Å$^2$)** | | 31.65 | 25.44 | 37.73 | 25.63 | 24.16 | 21.19 |
| **[a]SASA (Å$^2$)** | | 25.43 | 7.25 | 18.61 | 0.93 | 0.00 | 0.00 |
| **[b]RF (%)** | | 43.0 | 56.0 | 7.0 | 46.0 | 37.0 | 12.0 |

**Table 6.4. (Continued).**

| | **Chain id** | **IW25** | **IW26** | **IW27** | **IW28** | **IW29** |
|---|---|---|---|---|---|---|
| **5YFT** | A | 698 | 705 | 588 | 688 | 687 |
| | C | 672 | 677 | 565 | 647 | 655 |
| **5YFS** | A | 689 | 688 | 606 | 661 | 668 |
| | B | 676 | 674 | 556 | 655 | 652 |
| | C | 689 | 687 | 578 | 667 | 666 |
| **5YFJ** | A | 663 | 665 | 608 | 650 | 657 |
| | B | 659 | 662 | 605 | 646 | 648 |
| | C | 627 | 636 | 551 | 623 | 620 |
| **5YG6** | A | 644 | 659 | 588 | 638 | 641 |
| | B | 638 | 636 | 519 | 623 | 611 |
| | C | 651 | 653 | 602 | 642 | 643 |
| **5YFW** | A | 647 | 649 | 585 | 635 | 639 |
| | B | 652 | 657 | 597 | 639 | 634 |
| | C | 609, 633 | 627 | 525 | 615 | 608 |
| **5YFU** | A | 620 | 624 | 564 | 608 | 610 |
| **5YGA** | A | 603 | 602 | 548 | 599 | 590 |
| | B | 605 | 603 | 565 | 583 | 592 |

| **5YG8** | A | 588 | 594 | 515 | 575 | 570 |
|---|---|---|---|---|---|---|
| ***B* factor (Å$^2$)** | | 25.83 | 23.57 | 18.50 | 23.56 | 21.31 |
| **[a]SASA (Å$^2$)** | | 16.42 | 0.00 | 1.45 | 14.10 | 0.00 |
| **[b]RF (%)** | | 32.0 | 45.0 | 64.0 | 71.0 | 86.0 |

[a]SASA: solvent accessible surface area; [b]RF: residence frequency.

Apart from the invariant water molecules at the active-site pocket and its vicinity, 17 invariant water molecules are present at various structurally relevant regions (Figure 6.4A). Four invariant water molecules (IW15, IW16, IW17 and IW18) are involved in holding the long loop region. Three invariant water molecules (IW19, IW20 and IW21) interact with the longest helix $\alpha_5$ while two other invariant water molecules (IW27 and IW28) interact with helices $\alpha_7$ and $\alpha_4$, respectively. On the other hand, IW29 forms hydrogen bond with three water molecules (W603, W559 and W549 of PhR15Pi) which in turn interacts with the loop region connecting helices $\alpha_3$ and $\alpha_4$. Five invariant water molecules (IW22, IW23, IW24, IW25 and IW26) interact with amino acid residues at the inter-subunit region. Although these invariant water molecules do not directly interact with the amino acid residues of the associating monomer, they interact with water molecules present at the void created at the hexameric interface (Figure 6.4A and Table 6.3). Two invariant water molecules (IW13 and IW14) are involved in holding together the three regions NTD, CTD and the long loop region of R15Pi by anchoring the residues Thr67, Pro163, Glu254 and Glu260 (Figure 6.4B, C and Table 6.3). All the amino acid residues involved in interacting with the invariant water molecules are well conserved in the enzyme R15Pi (Figure 6.5).

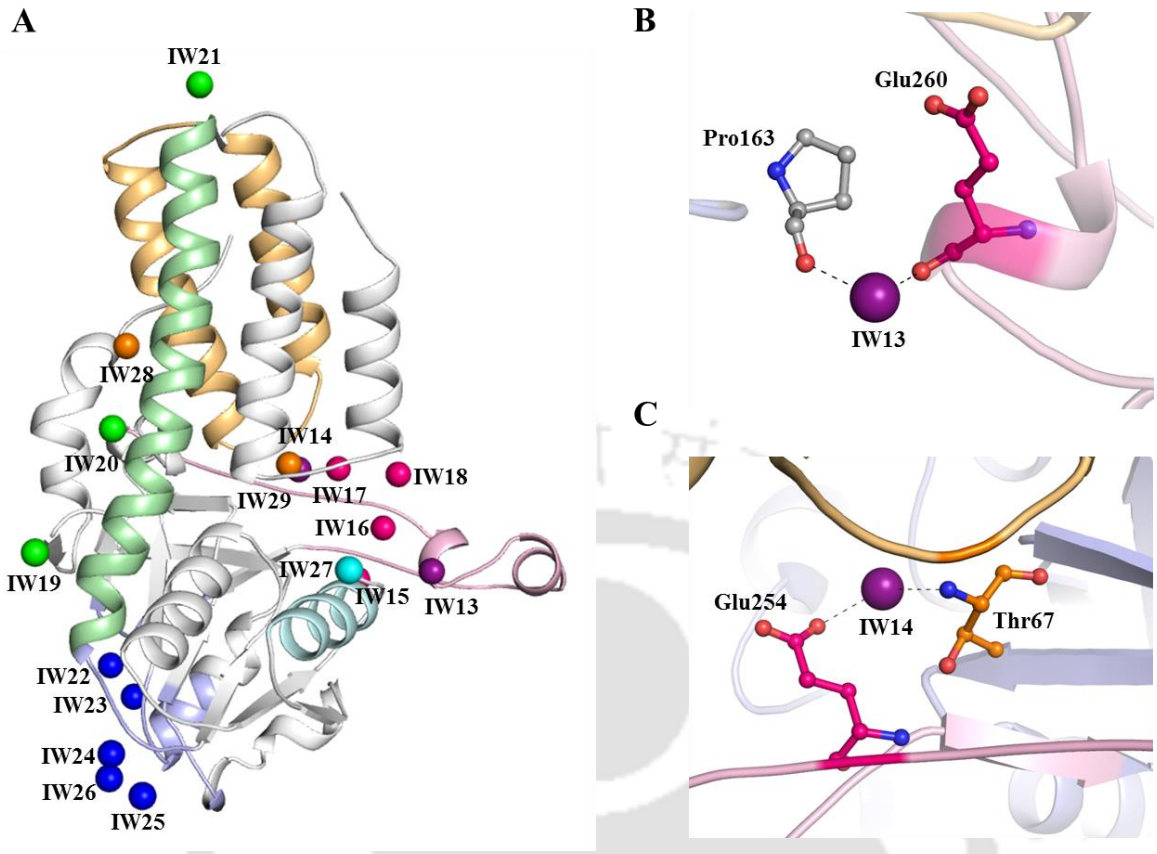



**Figure 6.4. Invariant water molecules at structurally important regions.** (A) Invariant water molecules interacting with amino acid residues at the helices $\alpha_4$ (light orange), $\alpha_5$ (pale green) & $\alpha_7$ (pale cyan), the long loop region (light pink), the loop connecting $\alpha_3$ and $\alpha_4$ (light orange) and at the inter-subunit region (light blue). The invariant water molecules (sphere) interacting with the corresponding structural elements are shown in a similar color. (B, C) Invariant water molecules (purple sphere) involved in holding together the CTD, NTD and the long loop region. The residues interacting with the invariant water molecules are shown as ball-and-stick model and the interaction(s) are shown as black-dashed lines. The amino acid residues are numbered according to the enzyme PhR15Pi.

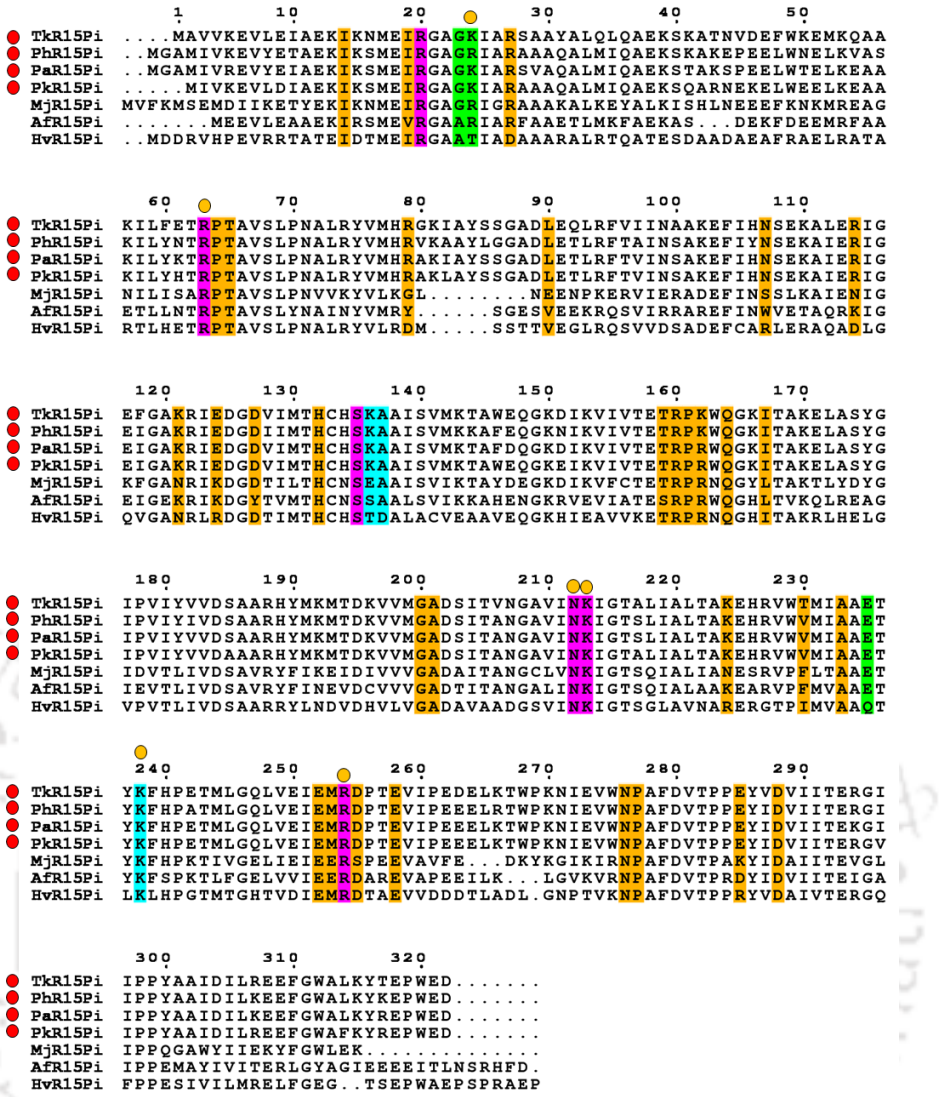



**Figure 6.5. Multiple sequence alignment (MSA) of enzyme R15Pi.** The amino acid residues involved in interacting with the ligand at the 'initial binding site', overlapping region and active-site pocket are highlighted in green, cyan and magenta, respectively. The amino acid residues involved in interacting with invariant water molecules are highlighted in orange. Those residues interacting with invariant water molecules and are also a part of the 'initial binding site' or the active site are highlighted with orange sphere. The enzyme R15Pi belonging to the order thermococcales of archaea are highlighted with red spheres. The enzyme R15Pi used in the MSA are from *T. kodakarensis* (TkR15Pi), *P. horikoshii* (PhR15Pi), *P. abyssi* (PaR15Pi), *P. kukulkanii* (PkR15Pi), *M. jannaschii* (MjR15Pi), *A. fulgidus* (AfR15Pi) and *H. volcanii* (HvR15Pi).

### 6.3.4 THE 'INITIAL BINDING SITE' OF SUBSTRATE

Among the available three-dimensional structures of R15Pi, only two (R15Pi-Form I, PDB id: 3A11 and R15Pi-Form III, PDB id: 3A9C) are present in an open conformation.

To gain insights into the early binding events of the substrate to the protein R15Pi, docking of the R15P with the R15Pi-Form I was performed. Interestingly, the result reveals that the substrate R15P is docked at a site (hereafter referred to as 'initial binding site'), which overlaps the active site (Figure 6.6A), with a binding energy of -10.18 kcal $mol^{-1}$ (Table 6.5). At this site, the 1-phosphate group of the substrate (R15P) forms hydrogen bond with the main chain N atoms of Lys136 and Ala137 while the 5-phosphate group is held by the $N^{\zeta}$ atoms of Lys24 and Lys136. Although, Lys24 of TkR15Pi is substituted by arginine in PhR15Pi, the availability of a positively charged amino acid residue will favor similar interaction with the negatively charged 5-phosphate group of the ligand (Figure 6.5). The main chain N atom of Gly23, $O^{\varepsilon 2}$ of Glu235 and $N^{\zeta}$ of Lys238 stabilize the ribose sugar oxygen atoms (Figure 6.6B and Table 6.5). Notably, the residues Lys136, Ala137 and Lys238 bridge the 'initial binding site' and active site of the protein.

However, the docking of R15P to R15Pi-Form III reveals its binding directly to the active site instead of the 'initial binding site', although with an identical binding energy (-10.18 kcal $mol^{-1}$), to that of the latter site. Notably, the docked conformation of the substrate is identical to that of the crystal structure. At the active site, the 1-phosphate group of the substrate (R15P) interacts with $N^{\zeta}$ of Lys213 and $N^{\eta 1}$ of Arg254 while the 5-phosphate group is coordinated by $O^{\gamma}$ of Ser135, main chain N atoms of Lys136 & Ala137 and $N^{\zeta}$ atoms of Lys213 & Lys238 (Table 6.5). The oxygen atoms (O2 and O3) of the ribose sugar are anchored by the $O^{\delta 1}$ & $N^{\delta 2}$ of Asn212 and N of Lys213 (Figure 6.6C and Table 6.5).

However, docking of the product (RuBP) to both the forms i.e. R15Pi-Form I as well as R15Pi-Form III establishes binding directly at the active site with an estimated free binding energies of -9.46 and -11.88 kcal $mol^{-1}$, respectively. Expectedly, in the R15Pi-form III, the coordinating residues of the protein to the docked RuBP at the active-site pocket are almost identical to that of the substrate, except for the interaction between $N^{\zeta}$ of Lys136 and 5-phosphate (Figure 6.6D, E and Table 6.5). The amino acid residues interacting with the substrate or product at the 'initial binding site' and active-site pocket are well conserved in the members of the order thermococcales while less conserved in the members from other archaeal orders (Figure 6.5).

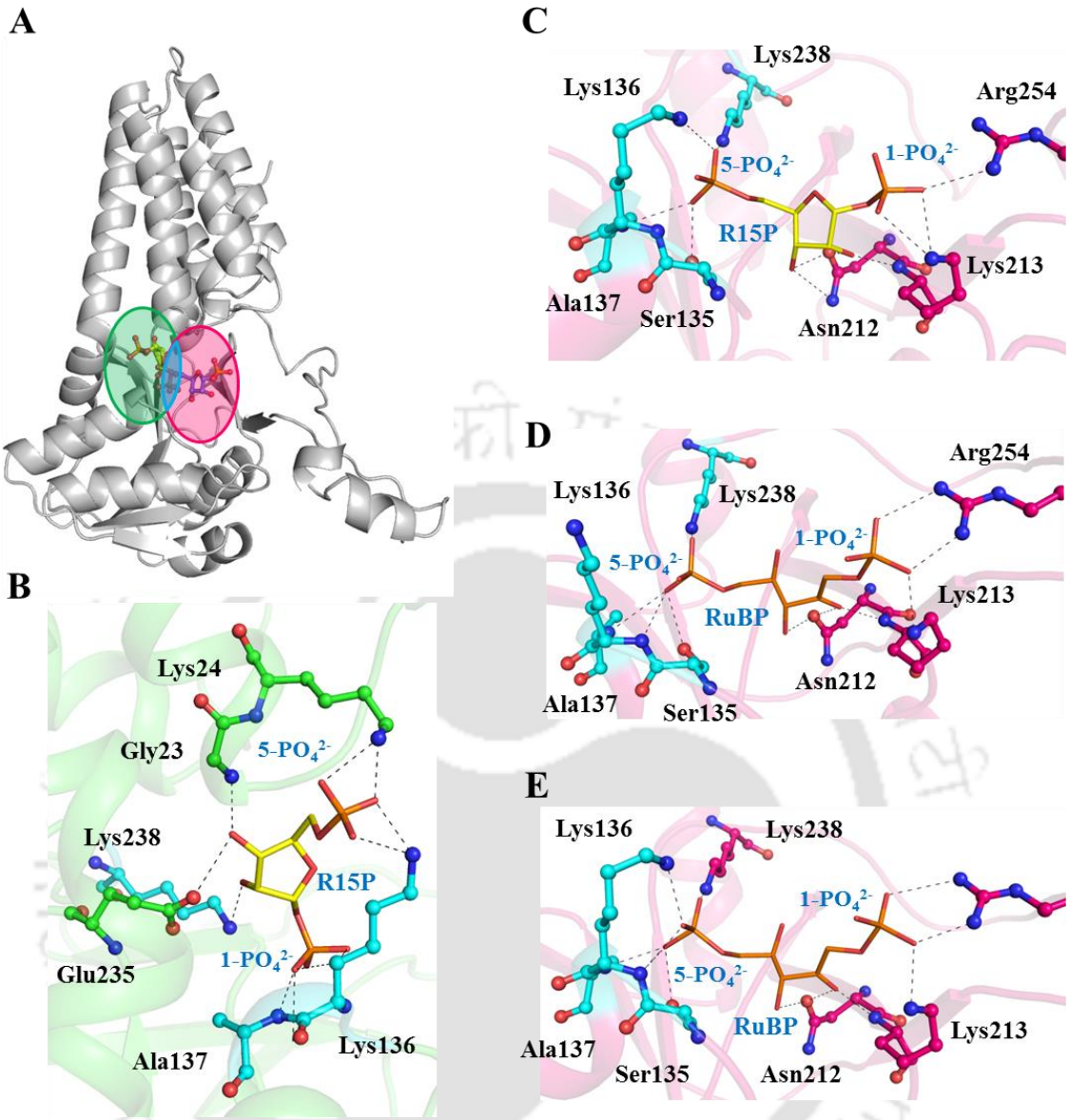



**Figure 6.6. Docking of R15P and RuBP to R15Pi-Form I and R15Pi-Form III.** (A) The two proposed substrate-binding sites viz. 'initial binding site' (green) and active site (magenta) in the enzyme R15Pi. The overlapping region between these two sites is highlighted by cyan oval. (B, C) The substrate (R15P, yellow lines) bound at the 'initial binding site' in R15Pi-Form I and at the active site in R15Pi-Form III, respectively. (D, E) The product (RuBP, orange lines) bound at the active sites of R15Pi-Form I and R15Pi-Form III. The amino acid residues involved in the interaction at the 'initial binding site', the overlapping region and the active site are shown in green, cyan and magenta, respectively. The interaction between ligand and amino acid residues are shown in black dashed lines. All the amino acid residues are numbered according to the enzyme TkR15Pi.

**Table 6.5.** Details of molecular docking interaction between the different forms of R15Pi with R15P and RuBP.

| Protein Form | Wild/ Mutant | Ligand | [a]EFBE (kcal mol$^{-1}$) | Ligand atom | List of interacting residues (distances in Å) | Binding Site |
|---|---|---|---|---|---|---|
| R15Pi-Form I | WT | R15P | -10.18 | 1-$PO_4^{2-}$ | Lys136 N (2.7, 3.0), Ala137 N (2.9) | 'Initial binding site' |
| | | | | O2 | Glu235 $O^{\varepsilon 2}$ (2.7), Lys238 $N^{\zeta}$ (2.7) | |
| | | | | O3 | Gly23 N (2.7), Glu235 $O^{\varepsilon 2}$ (3.0) | |
| | | | | 5-$PO_4^{2-}$ | Lys24 $N^{\zeta}$ (2.4, 2.8), Lys136 $N^{\zeta}$ (2.5, 2.9) | |
| R15Pi-Form III | | | -10.18 | 1-$PO_4^{2-}$ | Lys213 $N^{\zeta}$ (2.9, 2.9), Arg254 $N^{\eta 1}$ (3.1) | Active-site pocket |
| | | | | O2 | Asn212 $O^{\delta 1}$ (2.6), Lys213 N (2.9) | |
| | | | | O3 | Asn212 $O^{\delta 1}$ (2.5), Asn212 $N^{\delta 2}$ (3.2) | |
| | | | | 5-$PO_4^{2-}$ | Ser135 $O^{\gamma}$ (2.6), Lys136 $N^{\zeta}$ (2.6), Lys238 $N^{\zeta}$ (2.6), Lys136 N (2.7, 3.1), Ala137 N (3.1) | |
| R15Pi-Form I | WT | RuBP | -9.46 | 1-$PO_4^{2-}$ | Lys213 $N^{\zeta}$ (2.6), Arg254 $N^{\eta 1}$ (2.6), Arg254 $N^{\eta 2}$ (3.0) | Active-site pocket |
| | | | | O2 | Asn212 $O^{\delta 1}$ (3.1) | |
| | | | | O3 | Asn212 $O^{\delta 1}$ (2.8) | |
| | | | | 5-$PO_4^{2-}$ | Lys238 $N^{\zeta}$ (2.7), Ser135 $O^{\gamma}$ (2.9), Lys136 N (3.0), | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | Ala137 N (3.1) | |
| R15Pi-Form III | | | -11.88 | 1-$PO_4^{2-}$ | Lys213 $N^{\zeta}$ (2.6), Arg254 $N^{\eta1}$ (2.6), Arg254 $N^{\eta2}$ (3.2) | Active-site pocket |
| | | | | O2 | Asn212 $O^{\delta1}$ (2.8), Lys213 N (3.1) | |
| | | | | O3 | Asn212 $O^{\delta1}$ (2.7) | |
| | | | | 5-$PO_4^{2-}$ | Lys238 $N^{\zeta}$ (2.5), Ser135 $O^{\gamma}$ (2.6), Lys136 N (2.6), Lys136 $N^{\zeta}$ (2.7), Ala137 N (2.7) | |
| R15Pi-Form I | K24A | R15P | -7.89 | 1-$PO_4^{2-}$ | Lys213 $N^{\zeta}$ (2.9), Arg254 $N^{\eta1}$ (3.1) | Active-site pocket |
| | | | | O2 | Asn212 $O^{\delta1}$ (2.5), Lys213 N (3.1) | |
| | | | | O3 | Asn212 $O^{\delta1}$ (2.5) | |
| | | | | 5-$PO_4^{2-}$ | Ser135 $O^{\gamma}$ (2.8), Lys136 N (2.8, 3.0), Lys238 $N^{\zeta}$ (3.0), Ala137 N (3.2) | |
| | [b]K24 and K136 | R15P | -9.75 | 1-$PO_4^{2-}$ | Lys213 $N^{\zeta}$ (2.7), Arg254 $N^{\eta1}$ (3.2) | Active-site pocket |
| | | | | O2 | Asn212 $O^{\delta1}$ (2.4), Lys213 N (2.9) | |
| | | | | O3 | Asn212 $O^{\delta1}$ (2.5), Asn212 $N^{\delta2}$ (3.2) | |
| | | | | 5-$PO_4^{2-}$ | Ser135 $O^{\gamma}$ (2.5), Lys136 N (2.7, 2.8), Lys136 $N^{\zeta}$ (2.6, 3.0), Lys238 $N^{\zeta}$ (3.1), Ala137 N (3.2) | |
| | R20A | RuBP | -9.27 | 1-$PO_4^{2-}$ | Lys24 $N^{\zeta}$ (2.4), | 'Initial |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | Lys136 $N^{\zeta}$ (2.6) | binding site' |
| | | | | O2 | Gly23 N (3.0), Lys24 N (3.0) | |
| | | | | O3 | Ala22 N (2.5), Gly23 N (2.7) | |
| | | | | 5-$PO_4^{2-}$ | Ala137 N (2.7), Lys136 N (2.8, 3.0), Lys238 $N^{\zeta}$ (3.0) | |
| | R63A | RuBP | -9.3 | 1-$PO_4^{2-}$ | Lys24 $N^{\zeta}$ (2.4), Lys136 $N^{\zeta}$ (2.6) | 'Initial binding site' |
| | | | | O2 | Gly23 N (3.0), Lys24 N (3.0) | |
| | | | | O3 | Ala22 N (2.5), Gly23 N (2.7) | |
| | | | | 5-$PO_4^{2-}$ | Ala137 N (2.7), Lys136 N (2.8, 3.0), Lys238 $N^{\zeta}$ (3.1) | |
| | S135A | RuBP | -9.57 | 1-$PO_4^{2-}$ | Arg254 $N^{\eta 1}$ (2.6), Arg254 $N^{\eta 2}$ (2.6), Lys213 $N^{\zeta}$ (3.0) | Active-site pocket |
| | | | | O2 | Asn212 $O^{\delta 1}$ (3.1) | |
| | | | | O3 | Asn212 $O^{\delta 1}$ (2.8) | |
| | | | | 5-$PO_4^{2-}$ | Lys238 $N^{\zeta}$ (2.7), Lys136 N (3.0, 3.1), Ala137 N (3.1) | |
| | K213A | RuBP | -9.25 | 1-$PO_4^{2-}$ | Lys24 $N^{\zeta}$ (2.4), Lys136 $N^{\zeta}$ (2.6) | 'Initial binding site' |
| | | | | O2 | Gly23 N (3.0), Lys24 N (3.0) | |
| | | | | O3 | Ala22 N (2.5), Gly23 N (2.7) | |
| | | | | 5-$PO_4^{2-}$ | Ala137 N (2.7), Lys136 N (2.9, 3.0), | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | Lys238 $N^{\zeta}$ (3.0) | |
| | K238A | RuBP | -8.38 | 1-$PO_4^{2-}$ | Arg63 $N^{\eta1}$ (2.7), Arg63 $N^{\eta2}$ (3.1), Gly21 N (3.2) | 'Initial binding site' |
| | | | | 5-$PO_4^{2-}$ | Lys136 $N^{\zeta}$ (3.0) | |
| | R254A | RuBP | -9.31 | 1-$PO_4^{2-}$ | Lys24 $N^{\zeta}$ (2.4), Lys136 $N^{\zeta}$ (2.6) | 'Initial binding site' |
| | | | | O2 | Gly23 N (3.0), Lys24 N (3.0) | |
| | | | | O3 | Ala22 N (2.5), Gly23 N (2.7) | |
| | | | | 5-$PO_4^{2-}$ | Ala137 N (2.7), Lys136 N (2.9, 3.0), Lys238 $N^{\zeta}$ (3.0) | |

[a]EFBE: estimated free energy of binding.

[b]In this case, the conformation of Lys24 and Lys136 were kept similar to that of R15Pi-Form III.

The 1- and 5-phosphate groups of the docked substrate (R15P) at the 'initial binding site' interact with the residues Lys136 & Ala137 and Lys24, respectively. Among these, the residue Lys24 is exclusive to the 'initial binding site' while Lys136 and Ala137 lie at the overlapping region. Thus, to estimate the significance of Lys24 in the substrate binding at the 'initial binding site', it was mutated to alanine. Interestingly, the substrate docked at the active site of R15Pi-K24A mutant, instead of the 'initial binding site', with an EFBE of -7.89 kcal $mol^{-1}$ (Figure 6.7A and Table 6.5).

Docking of the R15P to R15Pi-Form III resulted in its binding to the active site instead of the 'initial binding site'. Thus, to identify the factors responsible for the R15P binding at the 'initial binding site', the structures of both the forms, R15Pi-Form I and R15Pi-Form III were compared. Notably, among the residues interacting with the R15P at the 'initial binding site', two residues Lys24 and Lys136 adapt different conformations in the two forms i.e. R15Pi-Form I and R15Pi-Form III (Figure 6.7B). To verify this, the

orientation of the residues Lys24 and Lys136 in R15Pi-Form I were altered according to that of R15Pi-Form III. The docking result reveals the binding of the substrate R15P at the active site rather than the 'initial binding site' with the side chain $N^{\zeta}$ atom of Lys136 interacting with 5-phosphate group of R15P (Figure 6.7C and Table 6.5).

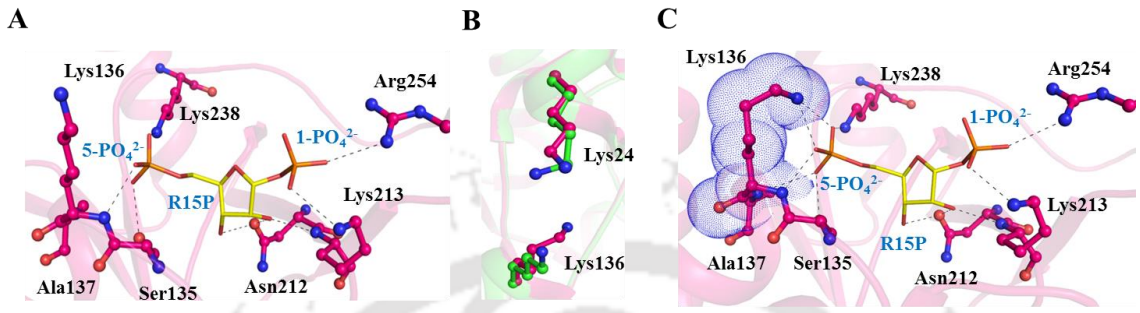



**Figure 6.7. The substrate (R15P) binding to R15Pi mutant protein.** (A) R15P (yellow lines) bound at the active site of R15Pi-K24A mutant; the amino acid residues involved in interaction are shown as ball-and-stick model in magenta. (B) Superimposition of R15Pi-Form I (green) and R15Pi-Form III (magenta) proteins. The residues Lys24 and Lys136, which show difference in their orientation, are highlighted. (C) R15P (yellow lines) bound at the active site after the alteration in the orientation of the residues Lys24 and Lys136. The altered orientation of Lys136 which interact with R15P via the 5-phosphate group is covered with blue dots. The amino acid residues interacting with the R15P at the active site are shown in magenta. All the amino acid residues are numbered according to the enzyme TkR15Pi.

### 6.3.5 SLIDING OF SUBSTRATE FROM 'INITIAL BINDING SITE' TO THE ACTIVE SITE

The docking results suggest that the substrate (R15P) binds firstly at the 'initial binding site' rather than the active site of the enzyme R15Pi. Thus, we speculated that the substrate may slide into the active site after binding briefly at the 'initial binding site'. To test this hypothesis, we calculated the distance between R15P and the residues Lys24 (interacts with the 5-phosphate group at the 'initial binding site') and Lys213 & Arg254 (interacts with the 1-phosphate group at the active site) during the MD simulation. Quite interestingly, the distance between R15P and Lys24 increases from 4.4 Å to 6.8 Å while that between R15P and Lys213 & Arg254 decreases from 6.8 Å to 2.1 Å indicating the movement of the substrate (R15P) from the 'initial binding site' to the active site of the protein (Figure 6.8A). This is further supported by the decrease (from -9.3 to 0 kJ $mol^{-1}$)

and increase (from 0 to -47 kJ $mol^{-1}$) in interaction energies between the R15P and Lys24 and Lys213 & Arg254, respectively (Figure 6.8B).

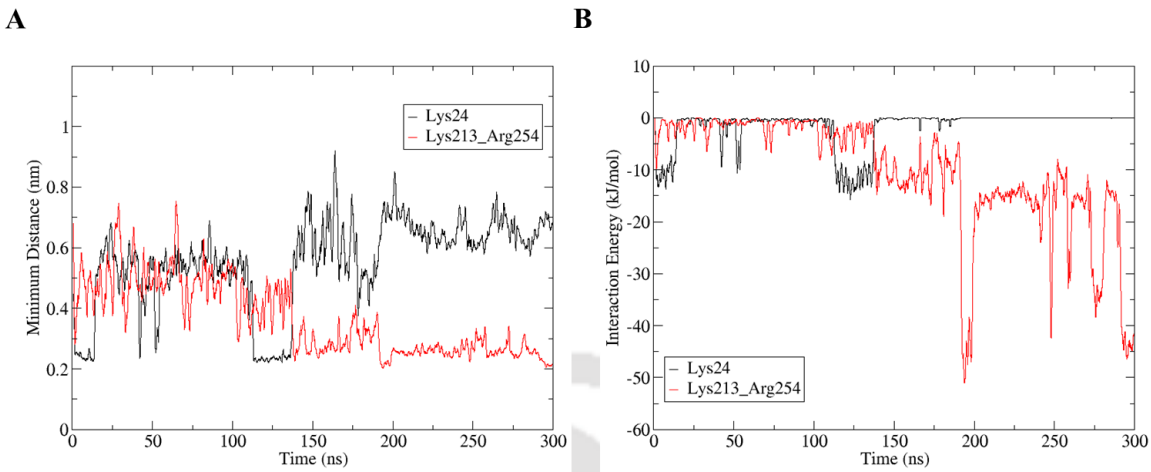



**Figure 6.8. The movement of the substrate (R15P) during MD simulation.** (A) Distance and (B) interaction energy between the substrate (R15P) and the residues Lys24 (black) and Lys213 & Arg254 (red) during the MD simulation.

### 6.3.6 ESSENTIAL AMINO ACID RESIDUES AT THE ACTIVE-SITE POCKET

The molecular docking results demonstrate that, unlike the substrate, the product (RuBP) binds at the active site instead of the 'initial binding site' irrespective of the protein forms i.e. R15Pi-Form I and R15Pi-Form III. Thus, to identify the residues aiding the binding of the product (RuBP) at the active site, we mutated the conserved residues Ser135, Lys213, Lys238 and Arg254 (Figure 6.5), which holds the 1- and 5-phosphate groups of the RuBP, individually to alanine. The docking results reveal that except S135A, the other point mutations K213A, K238A and R254A affect the binding of the product at the active site (Figure 6.9A, B, C, D and Table 6.5). Similarly, point mutation of conserved residues Arg20 and Arg63 (Figure 6.5), which interact with the 5-phosphate group of the substrate (R15P) in the R15Pi-Form II (i.e. closed conformation), to alanine causes the binding of the product (RuBP) to the 'initial binding site' instead of the active site of the protein (Figure 6.9E, F and Table 6.5).

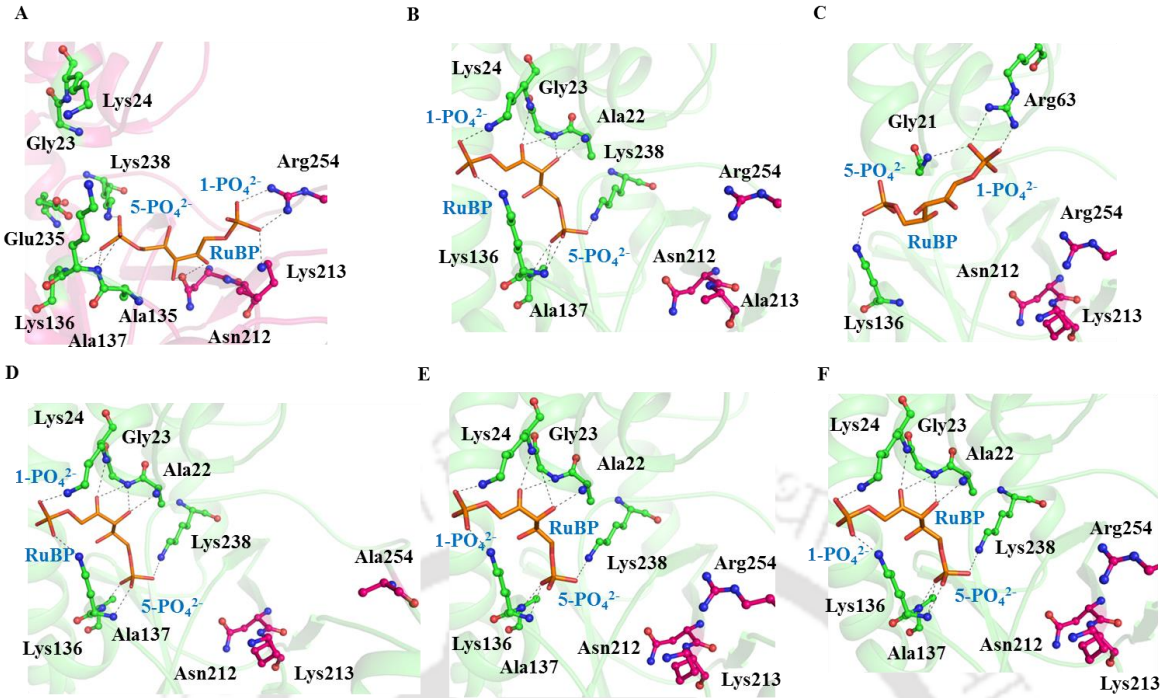



**Figure 6.9. The product (RuBP) binding to the R15Pi mutants.** The binding of RuBP (orange lines) at the active site of (A) R15Pi-S135A mutant and 'initial binding site' of (B) R15Pi-K213A, (C) R15Pi-K238A, (D) R15Pi-R254A, (E) R15Pi-R20A and (F) R15Pi-R63A mutant proteins. The amino acid residues interacting with RuBP at the active site and the 'initial binding site' are shown in magenta and green, respectively. The numbering of the residues shown in the figure is according to the enzyme TkR15Pi.

## 6.4 DISCUSSION

The enzyme R15Pi has been well studied in terms of its three-dimensional structure, domain motions, catalytic residues and enzyme reaction mechanism (Aono et al., 2012; Nakamura et al., 2012; Gogoi and Kanaujia, 2018). However, some of the questions that still need to be addressed include an understanding of the obligation of R15Pi to be present in a hexameric state, role of the catalytic residue in structural stability, role of invariant water molecules and identification of amino acid residues essential for ligand binding at the active site of the protein. Furthermore, acquisition of structural information related to the initial stages of substrate binding might provide new insight into the mechanism of enzyme-substrate interaction. Overall, understanding these aspects of the enzyme R15Pi would aid in further comprehending the structure-function relationship of the enzyme. From MD simulation results, it is evident that R15Pi-hexamer is more stable than R15Pi-monomer. In the R15Pi-monomer, higher

fluctuations could be perceived at the CTD including the long loop region. However, these fluctuations are stabilized to a large extent in the R15Pi-hexamer owing to the strong interaction present between the CTD of each monomer. Since the CTD remains rigid, only the NTD of R15Pi accomplishes the domain motion essential for ligand binding. R15Pi-hexamer also forms more stable interaction with the ligand which might also be due to the fact that most of the amino acid residues of the active-site pocket which interacts with the ligand are present at the CTD. A flexible CTD, as in the case of R15Pi-monomer, might hinder a stable binding with the ligand moiety, thus necessitating the requirement of a hexameric enzyme.

The two catalytic residues (cysteine and aspartate) required for the enzymatic activity are well identified and studied. Initially, the residue aspartate (202/204 in TkR15Pi/PhR15Pi) was considered to be more vital than the residue cysteine (133/135 in TkR15Pi/PhR15Pi) owing to the inability of the ligand to bind to the D202N mutant of TkR15Pi (Nakamura et al., 2012). However, elucidation of R15P-bound structure of R15Pi-D204N mutant protein from *P. horikoshii* later negated this observation (Gogoi and Kanaujia, 2018). Thus, to recognize the structurally essential catalytic residue, MD simulation was performed for the two mutant proteins. A higher instability of R15Pi-C135S mutant protein owing to the absence of amino acid interactions essential for stabilizing the longest loop region indicated that the cysteine residue is more crucial for the structural stability of enzyme R15Pi.

Although the enzyme R15Pi is known to undergo an open-to-closed transition during ligand binding, a set of invariant water molecules still populate the active-site pocket. These water molecules are not displaced even after ligand binding indicating their importance in the functioning of the enzyme. Few of the water molecules assist in the binding of the ligand at the active-site pocket by directly interacting with R15P and RuBP in a similar manner while the others interact with the conserved amino acid residues in the vicinity of the active-site pocket. The water molecules concentrated in the vicinity of the active-site pocket mostly interacts with those amino acid residues which are part of the flexible loop regions around the active-site pocket. By interacting with these amino acid residues, the water molecules permit sufficient movement of the loops around the active-site pocket for accommodating the ligand, without rendering the whole area too flaccid. Furthermore, the invariant water molecules at the active-site pocket and

in its vicinity are also observed to be completely buried. This implies that these invariant water molecules might have formed interaction with the amino acid residues at the active-site pocket or its vicinity prior to the folding of the protein and thus contributed towards the formation of the active-site pocket. Even though, the crystal structures of TkR15Pi contain low water molecules, three of the invariant water molecules are found to be present in 87% of the structures. Out of these, one of the invariant water molecules (IW4) interacts with the O3 oxygen atom while the other (IW5) interacts with the 5-phosphate group of the ligand. This indicates that among the five invariant water molecules that interact with the ligand, two (IW4 and IW5) are most essential for ligand binding at the active-site pocket. Another invariant water molecule IW12 is involved in holding a loop region that forms the 'active-site floor'. The presence of this water molecule in TkR15Pi protein hints toward the necessity of the water molecule for the stability of the loop.

Apart from the active-site pocket, invariant water molecules are also involved in providing structural integrity to the protein. They are present along the longest helix $\alpha_5$ which exhibits the most pronounced structural transition during the domain movement. The availability of water at this region might contribute towards the stability of the helix $\alpha_5$ during the domain movement. The long loop region, which harbors one of the conserved active-site residues Arg254, is also held by invariant water molecules. Presence of invariant water molecules provides rigidity to the otherwise flexible loop and in turn helps in maintaining the structural integrity of the active site of the enzyme. The structural integrity of the active-site pocket is further maintained by two invariant water molecules which couples the long loop region with the NTD and CTD. This interaction also results in the formation of a more compact protein molecule. The invariant water molecule at the inter-subunit region forms hydrogen bond(s) with other water molecule(s) concentrated in the void area created at the hexameric interface of R15Pi. These water molecules, in turn, interact with amino acid residues of the neighboring protomer. In this manner, a water mediated network is established that is involved in maintaining an intact hexamer.

An attempt to obtain a substrate-bound structure of R15Pi-WT resulted in the attainment of a product-bound structure (Nakamura et al., 2012; Gogoi and Kanaujia, 2018). This inhibited in achieving a clear picture of the initial stages of substrate (R15P) binding to

the enzyme R15Pi in its native form. Molecular docking of substrate to R15Pi-Form I interestingly led to the identification of an 'initial binding site' which overlaps with the active-site pocket. However, docking of the substrate to R15Pi-Form III did not result in binding of the R15P to the 'initial binding site'. This occurs due to the difference in the orientation of the two amino acid residues Lys24 and Lys136 in R15Pi-Form III, even though both forms are present in an open conformation. This difference in orientation of amino acid residues in R15Pi-Form III might be to ensure the binding of RuBP at the active-site pocket and inhibit its access to the 'initial binding site'.

The binding of the substrate R15P at the 'initial binding site' hints toward a probable early binding event which could not be captured during the process of crystallization. Binding of R15P to the 'initial binding site' might resemble the 'Two-step model' proposed by Stank et al., (2016) where a ligand first binds to an adjacent site transiently which, in turn, increases the possibility of its binding to its desired site. Similarly, in R15Pi, it might be possible that R15P initially binds at the 'initial binding site' with one of the phosphate groups interacting with residues, which are part of the active-site pocket, thus, guiding R15P to the actual catalytic site. Another possibility might be to render binding specificity to the substrate over the product. As observed from the docking studies, RuBP shows the tendency of binding to the active-site pocket rather than the 'initial binding site'. Although, the conversion of R15P to RuBP is a reversible process, the subsequent step i.e. the conversion of RuBP to 3-PGA by the enzyme RuBisCO is irreversible and thus would constantly require the availability of RuBP. Henceforth, the binding of R15P at the 'initial binding site' might help in the displacement of RuBP from the active-site pocket. However, it is very preliminary to arrive at such a conclusion based only on the docking studies. It would require a more detailed investigation supported by experimental evidences to prove the hypothesis.

To identify the amino acid residues which are crucial for ligand binding at the 'initial binding site' and the active-site pocket, point mutations were performed. Mutation of Lys24 led to the displacement of R15P from the 'initial binding site' to the active-site pocket indicating it to be the most crucial amino acid residue at the 'initial binding site'. At the active-site pocket, all the positively-charged conserved amino acid residues e.g. Arg20, Arg63, Lys213, Lys238 and Arg254 interacting with the 1-phosphate group are more crucial than the conserved amino acid residues Ser135, Lys136 and Ala137

interacting with the 5-phosphate group. The amino acid residues interacting with the 5-phosphate group act as an 'anchor' to bring the substrate R15P into a close proximity of the active-site pocket while the actual binding site is determined by the amino acid residues which interacts with the 1-phosphate group of the ligand.

### 6.4.1 PROPOSED MECHANISM OF 'SUBSTRATE SLIDING'

Binding of the substrate R15P to the 'initial binding site' instead of directly to the active-site pocket have led us to propose a probable 'substrate sliding' mechanism for the enzyme R15Pi. Initially, the substrate R15P binds to the 'initial binding site' with the 1-phosphate group interacting with amino acid residues at the overlapping region (Figure 6.10A). The binding of R15P to the 'initial binding site' might compel the NTD to move towards the CTD. Owing to this conformational change, R15P can no longer maintain its interaction with the amino acid residues at the 'initial binding site' and in turn, is forced to slide into the active-site pocket. The most significant bond breakage due to the forward movement of the NTD occurs in between Lys24 and the 5-phosphate group of R15P. This also causes Lys24 to come into close proximity of Lys136 and as a result, the positive-positive charge repulsion compels Lys136 to change its orientation. The change in the orientation of Lys136 changes the interacting atoms of 5-phosphate group from side chain $N^{\zeta}$ atom to the main chain N atom (Figure 6.10B). Simultaneously, forward movement of the NTD impels the other essential amino acid residues to come into close proximity to R15P. Once R15P forms a stable interaction with amino acid residues at the active-site pocket, Lys136 again changes its orientation in order to interact with the 5-phosphate group via its side chain atom $N^{\zeta}$ (Figure 6.10C). This movement of R15P from the 'initial binding site' to the active-site pocket can also not be ignored based on the change in orientation of Lys136 in the three different available forms of R15Pi.

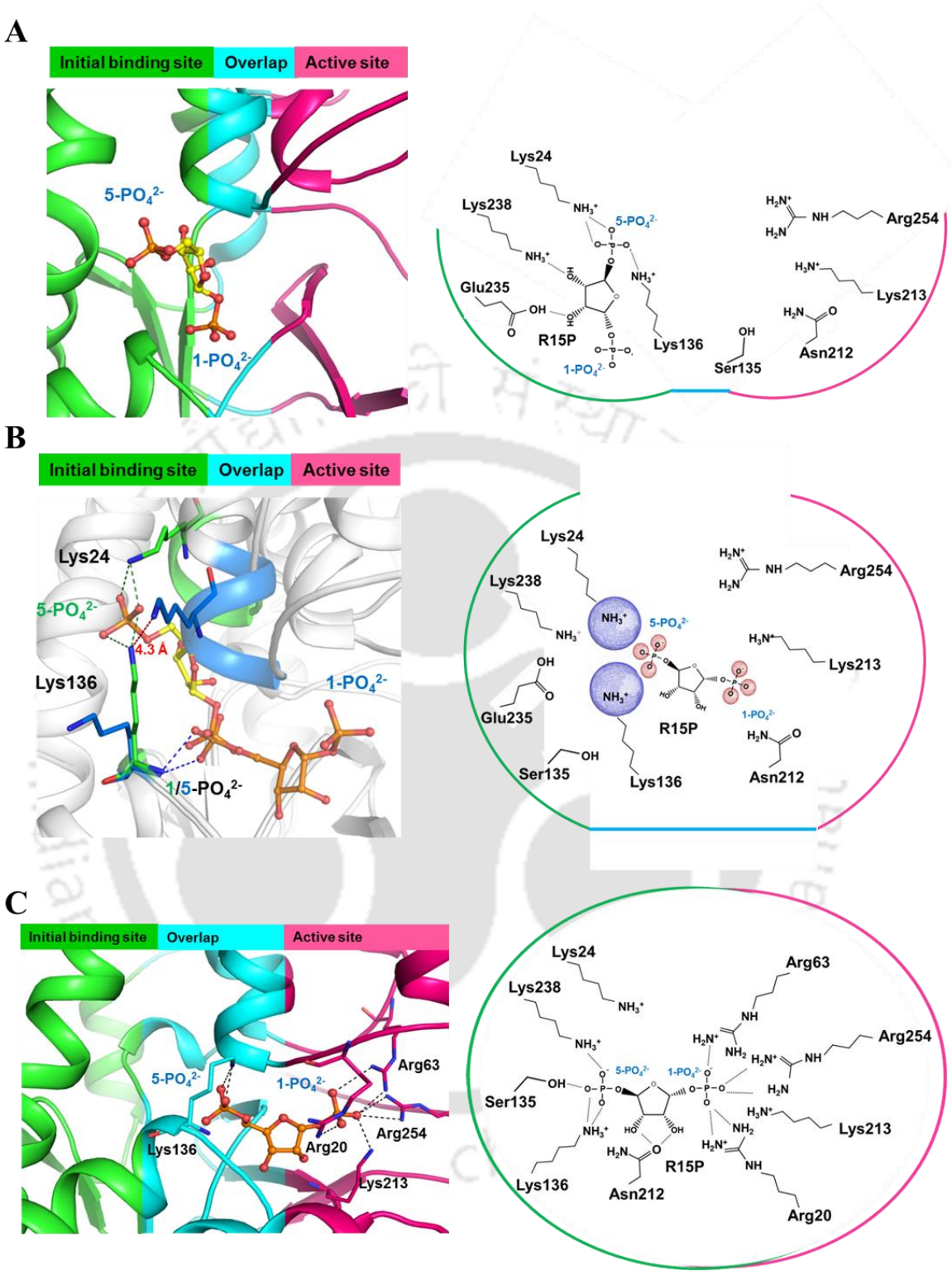



**Figure 6.10. Proposed mechanism of the 'substrate sliding' in the enzyme R15Pi depicted in a three-dimensional model (left) and chemical scheme (right).** (A) *Initial step*, R15P (yellow) binds at the 'initial binding site' of R15Pi present in an open conformation. (B) *Intermediate step*, transition from open to closed conformation causes Lys24 to move forward resulting in the disruption of the bond between Lys24 and the 5-phosphate group of R15P at the 'initial binding site'. The forward movement of Lys24

also causes Lys136 to change its orientation. This changes in interaction and orientation of amino acid residues as well as spatial constraints in the closed state of R15Pi obliges R15P to slide into the active-site pocket. In the three-dimensional model, the residues Lys24 and Lys136 in open and closed conformation of R15Pi are highlighted in green and blue lines, respectively. The substrate R15P at the 'initial binding site' and the active-site pocket are shown in yellow and orange, respectively. The interactions of R15P with Lys24 and Lys136 at the 'initial binding site' and the active-site pocket are shown in green and blue dashes, respectively. The repulsion of Lys24 and Lys136 in the closed and open conformation of R15Pi, respectively, is shown in red dash with the distance indicated in red. The phosphate groups of R15P at the 'initial binding site' and the active-site pocket are labeled in green and blue, respectively. In the chemical scheme, the repulsion between Lys24 and Lys136 as well as the spatial constraint between R15Pi and amino acid residues at the 'initial binding site' is highlighted with blue and red spheres, respectively. (C) *Final step,* the closed conformation of the enzyme helps R15P to interact with the amino acid residues at the active-site pocket. The 'initial binding site', overlapping region and the active-site pocket are shown in green, cyan and magenta, respectively. The open and closed conformation of the active-site pocket are represented in open and closed circles, respectively. All the amino acid residues are numbered according to the enzyme TkR15Pi.

## 6.5 CONCLUSION

In summary, this chapter provides an understanding of the different structural aspects of the enzyme R15Pi that are indispensable for its effective functioning. Furthermore, identification of an 'initial binding site' in R15Pi and the proposition of a 'substrate sliding' movement towards the active-site pocket provide a new perspective to the mechanism of action of the enzyme R15Pi.

# CHAPTER 7

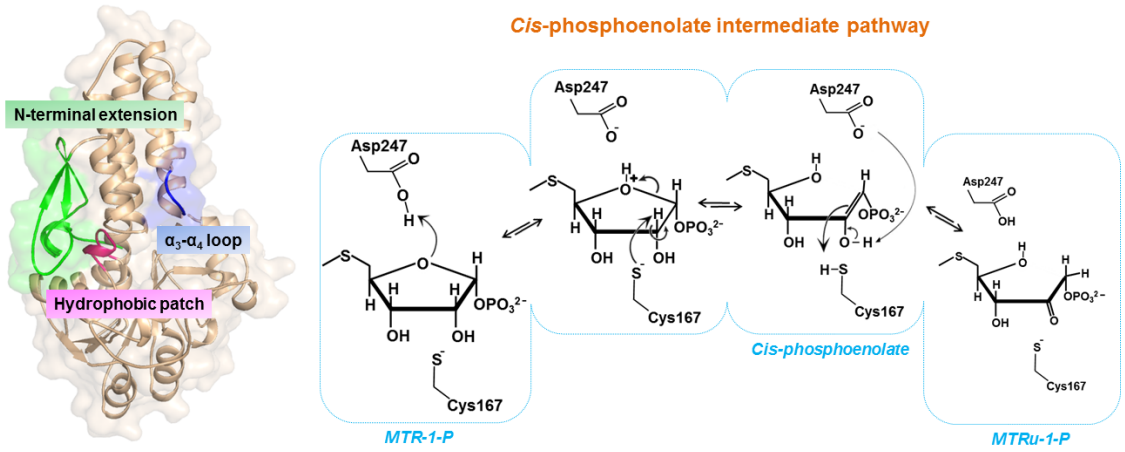
## Structural insights into the catalytic mechanism of 5-methylthioribose 1-phosphate isomerase

This chapter has been published as:

Gogoi, P., Mordina, P. and Kanaujia, S.P., 2019. Structural insights into the catalytic mechanism of 5-methylthioribose 1-phosphate isomerase. *Journal of Structural Biology*, *205*(1), 67-77.

# ABSTRACT

5-methylthioribose 1-phosphate isomerase (M1Pi) is a crucial enzyme involved in the universally conserved methionine salvage pathway (MSP) where it is known to catalyze the conversion of 5-methylthioribose 1-phosphate (MTR-1-P) to 5-methylthioribulose 1-phosphate (MTRu-1-P) via a mechanism which remains unspecified till date. Furthermore, although M1Pi has a discrete function, it surprisingly shares high structural similarity with two functionally non-related proteins such as ribose-1,5-bisphosphate isomerase (R15Pi) and the regulatory subunits of eIF2B. To identify the distinct structural features that lead to divergent functional obligations of M1Pi as well as to understand the mechanism of enzyme catalysis, the crystal structure of M1Pi from a hyperthermophilic archaeon *P. horikoshii* OT3 was determined. A meticulous structural investigation of the dimeric M1Pi revealed the presence of an N-terminal extension and a hydrophobic patch absent in R15Pi and the regulatory α-subunit of eIF2B. Furthermore, unlike R15Pi in which a kink formation is observed in one of the helices, the domain movement of M1Pi is distinguished by a forward shift in a loop covering the active-site pocket. All these structural attributes contribute towards a hydrophobic microenvironment in the vicinity of the active site of the enzyme making it favorable for the reaction mechanism to commence. Thus, a hydrophobic active-site microenvironment in addition to the availability of optimal amino acid residues surrounding the catalytic residues in M1Pi led us to propose its probable reaction mechanism via a *cis*-phosphoenolate intermediate formation.




## 7.1 INTRODUCTION

Methionine salvage pathway (MSP) is a universal pathway involved in the reprocessing of sulfur-containing cellular metabolites to the amino acid methionine (Sekowska et al., 2004; Albers, 2009). This pathway holds utmost importance owing to the fact that the amount of methionine in the cell is limiting and its *de novo* synthesis is energetically expensive (Winans and Bassler, 2002; Sekowska et al., 2004). In the MSP, 5-methylthioadenosine (MTA) undergoes a series of steps catalyzed by ~11 enzymes to form methionine (Sekowska and Danchin, 2002; Ashida et al., 2003). One of these steps involves the isomerization of 5-methylthioribose 1-phosphate (MTR-1-P) to 5-methylthioribulose 1-phosphate (MTRu-1-P) via the enzyme 5-methylthioribose 1-phosphate isomerase (M1Pi) (Ashida et al., 2003).

The M1Pi enzymes form a distinct family (No.: PF01008, known as IF-2B family) in the Pfam database (Finn et al., 2015). This family includes two other groups of proteins viz. ribose-1,5-bisphosphate isomerase (R15Pi) and the regulatory subunits (α, β and δ) of eIF2B as well. While, the enzyme R15Pi is involved in the nucleoside 5′-monophosphate (NMP) degradation pathway exclusively present in archaea, the eIF2B acts as a guanine nucleotide exchange factor (GEF) during the process of protein translation initiation in eukaryotes; not identified in prokaryotes till date (Price and Proud, 1994; Aono et al., 2012; Nakamura et al., 2012). Although functionally non-related, these three groups of proteins share significant similarity in their primary and tertiary structures (Dev et al., 2009; Gogoi et al., 2016; Kashiwagi et al., 2016). Even though the protomers of these

proteins (i.e. members of the PF01008 family) share a common overall topology comprising of N- (α-helical) and C-terminal (αβα-sandwich) domains, their quaternary structures differ. The enzyme R15Pi and the regulatory subunits (α, β and δ) of eIF2B form hexamer (trimer of dimers), whereas the enzyme M1Pi functions as dimer in solution (Bumann et al., 2004; Nakamura et al., 2012; Kashiwagi et al., 2016).

Both the isomerases, M1Pi and R15Pi are known to undergo transition from an open to closed state upon substrate binding mediated by the movement of the N-terminal domain (NTD) towards the C-terminal domain (CTD) (Tamura et al., 2008; Nakamura et al., 2012). The NTD movement upon substrate binding provides a favorable active-site environment for the enzyme reaction to occur (Nakamura et al., 2012). The transition from an open to closed state in these proteins are marked by the formation of a kink in the longest helix of the NTD (Nakamura et al., 2012; Kuhle et al., 2015). Although the domain movement has been proposed for the regulatory subunits (especially for the α-subunit) of eIF2B also (Kuhle et al., 2015; Gogoi and Kanaujia, 2018), no clear evidence has been reported till date.

The enzyme R15Pi catalyzes the conversion of ribose-1,5-bisphosphate (R15P) to ribulose-1,5-bisphosphate (RuBP) via the formation of a *cis*-phosphoenolate intermediate (Nakamura et al., 2012). However, the catalytic mechanism of the enzyme M1Pi to produce MTRu-1-P from MTR-1-P is not clear. In literature, two different mechanisms of its catalytic action have been suggested (Tamura et al., 2008). One mechanism proceeds via the formation of a *cis*-phosphoenolate intermediate and is followed by several other isomerases such as triosephosphate isomerase (TPI), ribose-5-phosphate isomerase (RPI) and R15Pi (Rose and O'Connell, 1960; Rose, 1975; Nakamura et al., 2012). The other mechanism involves a direct hydride transfer assisted by two divalent metal cations as observed in the case of xylose isomerase (XI) (Blow et al., 1992; Asboth and Naray-Szabo, 2000; Fenn et al., 2004). The direct hydride transfer has been suggested as the preferred mechanism of enzyme catalysis for M1Pi due to the lack of incorporation of proton (deuterium) in the product from the medium, a phenomenon common in isomerases following the *cis*-enediol mechanism (O'Donoghue et al., 2005a, 2005b; Berrisford et al., 2006; Saito et al., 2007; Tamura et al., 2008). However, in the case of the enzyme M1Pi, the reaction completes in a metal-independent manner (Saito et al., 2007). Thus, the exact catalytic mechanism of the enzyme M1Pi remains unsettled.

Herein, the three-dimensional crystal structure of M1Pi from *P. horikoshii* OT3 encoded by the open reading frame (ORF) PH0702 has been elucidated. A thorough structural comparison of M1Pi with R15Pi and the regulatory α-subunit of eIF2B to identify and to understand the structural (re)arrangements or the occurrence of structural evolution responsible for their functional specificities have been drawn. In-depth assessment of the microenvironment of the active-site pocket has also been performed to identify the probable preferred mechanism of enzyme reaction followed by the M1Pi enzyme.

## 7.2 MATERIALS AND METHODS

### 7.2.1 CLONING, OVER EXPRESSION AND PROTEIN PURIFICATION

The ORF PH0702 encoding PhM1Pi was PCR amplified (Figure 7.1A) using the forward primer 5′-GCGC**GCTAGC**ATGGAGATAAGATACACTCCAAAGG-3′ and reverse primer 5′-CGA**CTCGAG**TCAAACTGCTTCGTCTCC-3′ containing NheI and XhoI restriction sites (bold), respectively. The amplified gene was inserted into the plasmid pET-28a(+) excised with the same set of restriction enzymes. Subsequently, the clones were confirmed by double digestion using the restriction enzymes, NdeI and XhoI by incubating the vectors at 37°C for two hours (Figure 7.1B). In addition, the clone was confirmed by DNA sequencing. The over expression PhM1Pi was optimized using varying concentrations (0.1, 0.5 and 1.0 mM) of isopropyl β-D-1-thiogalactopyranoside (IPTG) in the expression systems such as *E. coli* BL21-CodonPlus (DE3)-RIL and *E. coli* Rosetta (DE3) *E. coli* competent cells. Cells were grown at 37°C to an $A_{600}$ of 0.6 in LB media supplemented with 50 μg $ml^{-1}$ kanamycin and 34 μg $ml^{-1}$ chloramphenicol. Optimum expression of the proteins was obtained by inducing the cells with 1 mM IPTG (Figure 7.1C). Subsequently, the solubility of the protein was checked and the results showed that the protein remains in the soluble form (Figure 7.1D). For protein purification, the cells were harvested and resuspended in lysis buffer composed of 20 mM Tris-HCl pH 8.0, 150 mM KCl, 5 mM imidazole, 5% glycerol, 1 mM phenylmethylsulfonyl fluoride (PMSF) and 5 mM β-mercaptoethanol (β-ME). The cells were lysed using sonicator and the cell lysate was incubated at 90°C for 13 min to remove the thermolabile proteins deriving from the host cells. The insoluble debris were removed by centrifugation at 12,000 rpm for 40 min at 4°C. The supernatant fraction was

applied to the Ni-NTA affinity column preequilibrated with binding buffer (20 mM Tris-HCl pH 8.0, 150 mM KCl, 5 mM imidazole, 5% glycerol, 1 mM PMSF and 5 mM β-ME) and incubated for two hours. The column was washed with five column volume of each of wash buffer A (20 mM Tris-HCl pH 8.0, 500 mM KCl, 5 mM imidazole, 5% glycerol, 1 mM PMSF and 5 mM β-ME), wash buffer B (20 mM Tris-HCl pH 8.0, 500 mM KCl, 20 mM imidazole, 5% glycerol, 1 mM PMSF and 5 mM β-ME), wash buffer C (20 mM Tris-HCl pH 8.0, 500 mM KCl, 50 mM imidazole, 5% glycerol, 1 mM PMSF and 5 mM β-ME) and wash buffer D (20 mM Tris-HCl pH 8.0, 500 mM KCl, 50 mM imidazole and 1 mM PMSF). The proteins were eluted with 250 mM imidazole in 20 mM Tris-HCl pH 8.0, 150 mM KCl and 1 mM PMSF (Figure 7.1E). The eluted fractions were collected and step-wise dialysis was performed against 20 mM Tris-HCl pH 8.0 and 150 mM KCl to remove imidazole. The dialyzed proteins were concentrated to a final concentration of ~15 mg $ml^{-1}$ using Vivaspin turbo 15 (Sartorius) centrifugal filter unit with a molecular weight cutoff of 10 kDa (Figure 7.1F).

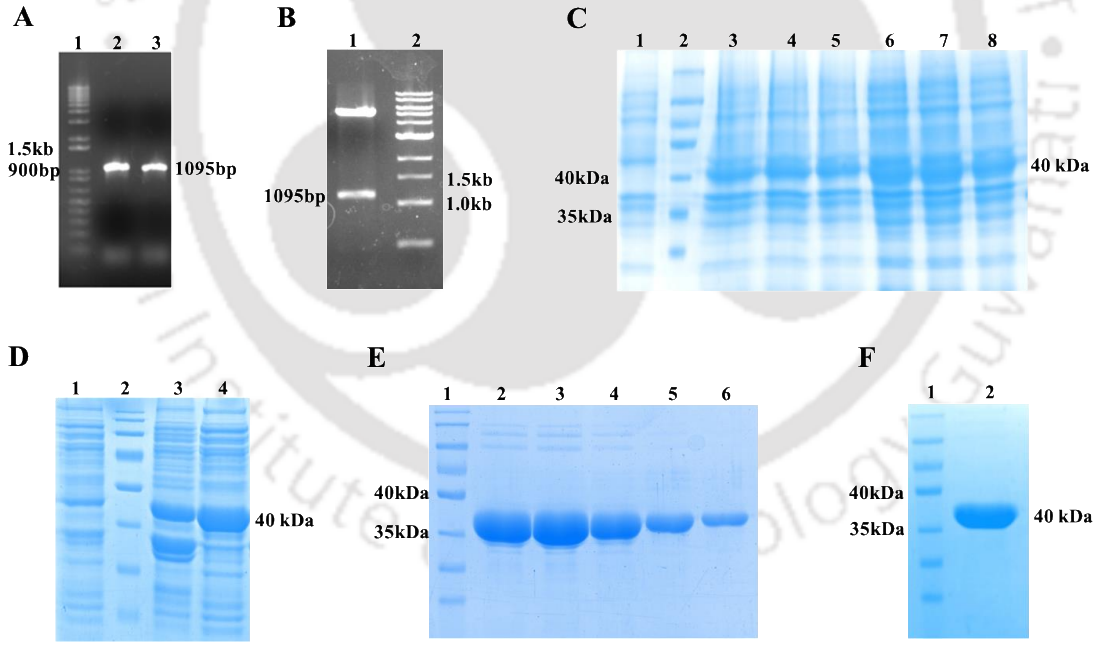



**Figure 7.1. Cloning, over expression and purification of PhM1Pi.** (A) PCR amplification of PhM1Pi analysed in 0.8% agarose gel (Lane 1: DNA ladder, Lane 2 & 3: PCR amplified PhM1Pi). (B) Clone confirmation using the double digestion of the plasmid pET-28a(+) containing PhM1Pi analysed in 0.8% agarose gel (Lane 1: positive clone, Lane 3: wild type pET-28a(+), Lane 4: DNA ladder). (C) Over expression of PhM1Pi at 1 mM IPTG at different time intervals analysed in 12% SDS-PAGE (Lane 1: uninduced, Lane 2: protein marker, Lane 3: 1$^{st}$ hr, Lane 4: 2$^{nd}$ hr, Lane 5: 3$^{rd}$ hr, Lane 6: 4$^{th}$ hr, Lane 7: 5$^{th}$ hr, Lane 8: 6$^{th}$ hr). (D) Solubility test for the over expressed PhM1Pi

analysed in 12% SDS-PAGE (Lane 1: uninduced, Lane 2: protein marker, Lane 3: pellet fraction, Lane 4: supernatant fraction). (E) 12 % SDS-PAGE of purified PhM1Pi (Lane 1: protein marker, Lane 2-6: purified PhM1Pi). (F) Concentrated PhM1Pi (Lane 1: protein marker, Lane 2: concentrated PhM1Pi).

### 7.2.2 CRYSTALLIZATION, X-RAY INTENSITY DATA COLLECTION AND PROCESSING

A preliminary screening of the purified PhM1Pi was performed using microbatch-under-oil technique by mixing 2 µl of protein and 2 µl of crystallization buffers available in crystallization screens from Hampton Research at 4 and 20°C temperatures. Although, protein crystals were obtained in buffer conditions including 0.2 M calcium acetate hydrate & 20% PEG 3350, 0.2 M calcium chloride dihydrate & 20% PEG 3350, 0.2 M ammonium sulfate, 0.1 M sodium acetate trihydrate & 20% PEG 3350 and 0.2 M sodium formate & 20% PEG 3350 (Figure 7.2A-D), none of the crystals diffracted better than ~4 Å (Figure 7.2E).

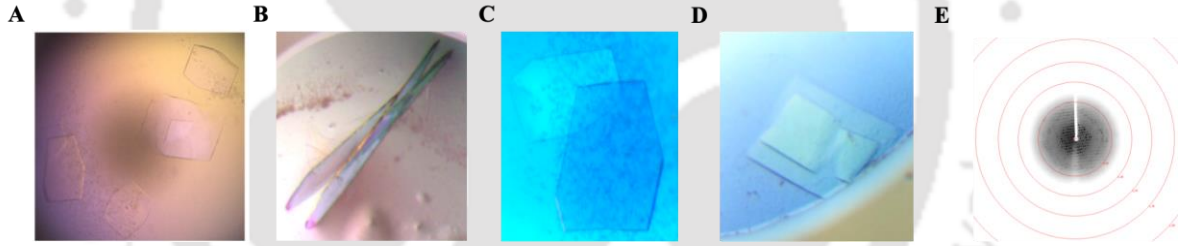



**Figure 7.2.** Crystal hits of PhM1Pi. (A-D) Protein crystals obtained in different buffer conditions during crystal screening. (E) X-ray diffraction pattern of PhM1Pi protein crystals diffracting up to 4 Å.

To obtain diffraction quality crystals, at least five different buffer conditions were optimized initially in hanging-drop and sitting-drop vapor-diffusion as well as microbatch-under-oil technique. The various approaches of crystal optimization are summarized in Table 7.1. However, in none of the conditions, optimal diffraction quality crystals could be obtained (Figure 7.3).

**Table 7.1.** Crystal optimization approaches for PhM1Pi.

| Optimization approach | Condition | Method |
|---|---|---|
| Protein concentration | 2.5-26 mg $ml^{-1}$ | Microbatch-under-oil, hanging-drop and sitting-drop vapor-diffusion |
| Salt concentration | Increasing and decreasing concentrations | |
| Precipitant concentration | | |
| pH of the buffer | Increasing and decreasing pH | Microbatch-under-oil and hanging-drop vapor-diffusion |
| Protein:Buffer ratio | 1:2, 2:1, 2:2, 2:3, 3:1, 3:2, 3:3, 4:1, 4:2, 4:3, 5:2, 5:3 | Microbatch-under-oil |
| | 1:2, 2:1, 2:2, 2:3, 3:1, 3:2, 3:3, 4:1, 4:2, 4:3, 5:2, 5:3 | Hanging-drop vapor-diffusion |
| | 5:5, 6:2, 9:3, 10:5, 10:6, 10:7, 10:8, 10:10, 12:4, 15:5 | Sitting-drop vapor-diffusion |
| Oil | 250-500 µl of different combinations of paraffin and silicone oil | Hanging-drop vapor-diffusion and sitting-drop vapor-diffusion |
| Microseeding | Using a seed stock without dilution to a seed stock of $10^{-6}$ dilution | Hanging-drop vapor-diffusion |
| Additives | Additive screen from Hampton research | Microbatch-under-oil and hanging-drop vapor-diffusion |
| Low melting agarose (LMA) | 0.1-0.3% | Microbatch-under-oil and hanging-drop vapor-diffusion |
| Glycerol | 5-30% | Hanging-drop vapor-diffusion and sitting-drop vapor-diffusion |
| 2-methyl-2,4-pentanediol (MPD) | 5-30% | Hanging-drop vapor-diffusion and sitting-drop vapor-diffusion |

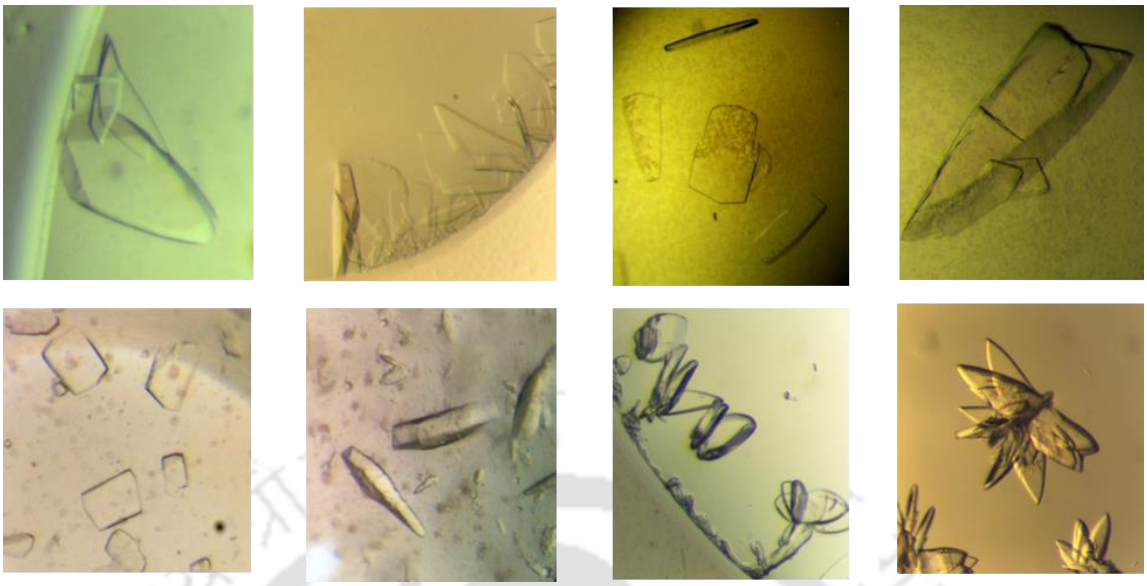


**Figure 7.3. Protein crystals of PhM1Pi at different stages of optimization.** The crystals shown in the figure were grown using microbatch-under-oil, hanging-drop and sitting-drop vapor-diffusion methods.

Further crystal screening in new buffer conditions were performed and two forms of crystals belonging to space groups $P3_221$ and $P1$ were obtained at 4 and 20°C, respectively. The crystal belonging to space group $P3_221$ grew in 0.1 M sodium citrate tribasic dihydrate pH 5.6 and 2.5 M 1,6-hexanediol in a microbatch-under-oil set up by mixing 2 µl of protein and 2 µl of the buffer (Figure 7.4A). Initial crystals belonging to space group *P*1 were obtained in 0.2 M ammonium acetate, 0.1 M Tris pH 8.0 and 16% PEG 10,000. However, the crystals obtained diffracted very poorly. To obtain diffraction quality crystals, vigorous optimization was performed. Finally, diffraction quality crystals were obtained in 0.05 M ammonium acetate, 0.1 M Tris pH 8.0, 16% PEG 10,000 and 0.05 M sodium acetate using hanging-drop vapor-diffusion method by mixing 3 µl of protein and 1 µl of the crystallization buffer solutions (Figure 7.4B). X-ray intensity diffraction data was collected at -173°C using the home source Rigaku MicroMax-007 HF diffractometer (operated at 40 kV and 30 mA) and RAxis IV++ imaging-plate detector available at the Central Instruments Facility (CIF) of Indian Institute of Technology Guwahati, India. The crystal to detector distance was maintained at 180 mm. The crystals belonging to space groups $P3_221$ and $P1$ diffracted to a resolution of 2.3 and 2.65 Å, respectively (Figure 7.4C, D). The diffraction data were

processed and scaled using the programs iMosflm (Battye et al., 2011) and Aimless (Evans and Murshudov, 2013), respectively, embedded in the CCP4 package (Winn et al., 2011). The intensities were converted to structure factors using the program CTRUNCATE available in CCP4 package. Details of the diffraction data collection and processing statistics are given in Table 7.2.

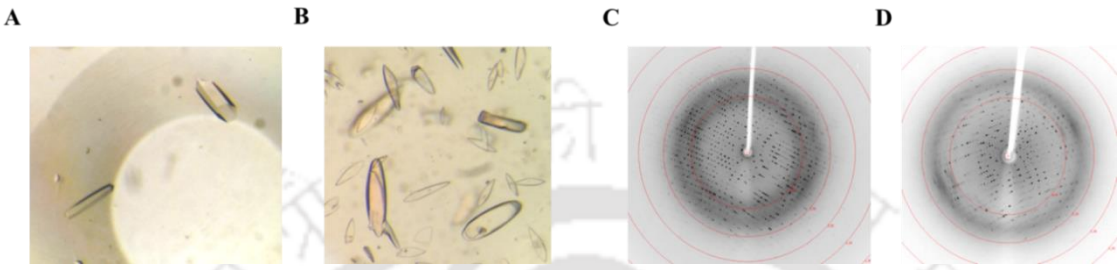



**Figure 7.4. Crystallization of PhM1Pi.** Protein crystals of (A) Form I and (B) Form II of PhM1Pi. X-ray diffraction pattern of (C) Form I and (D) Form II of PhM1Pi.

**Table 7.2.** Data collection, processing and refinement statistics of PhM1Pi. The values provided in parenthesis are for the last resolution shell.

| | **Form I** | **Form II** |
|---|---|---|
| Wavelength (Å) | 1.5418 | 1.5418 |
| Temperature (K) | 100 | 100 |
| Space group | $P3_221$ | $P1$ |
| Unit-cell parameter (Å, °) | $a$=$b$=169.8, $c$=53.7, $\alpha$=$\beta$=90, $\gamma$=120 | $a$= 46.2, $b$=91.2, $c$=102.1, $\alpha$=108.6, $\beta$=98.1, $\gamma$=104.6 |
| Resolution (Å) | 84.87-2.30 (2.38-2.30) | 82.02-2.65 (2.76-2.65) |
| No. of observed reflections | 316498 (26080) | 148751 (16542) |
| No. of unique reflections | 39641 (3881) | 41560 (4648) |
| CC(1/2) | 0.999 (0.900) | 0.992 (0.819) |
| Completeness (%) | 100 (100) | 96.5 (94.9) |
| $V_M$ ($Å^3$ $Da^{-1}$) | 2.76 | 2.37 |
| Solvent content (%) | 55.54 | 48.12 |
| Mosaicity (°) | 0.4 | 0.6 |

| | | |
|---|---|---|
| $I/\sigma(I)$ | 18.9 (3.3) | 9.4 (2.2) |
| $R_{merge}$ (%) | 6.5 (53.8) | 9.6 (55.7) |
| $R_{pim}$ (%) | 2.4 (21.9) | 6 (34.6) |
| $R_{meas}$ (%) | 6.9 (58.3) | 11.3 (65.7) |
| Multiplicity | 8.0 (6.7) | 3.6 (3.6) |
| $R_{work}/R_{free}$ (%) | 17.75/23.72 | 21.05/29.42 |
| **Protein model** | | |
| No. of subunits in ASU | 2 | 4 |
| Protein atoms | 5586 | 11172 |
| Water molecules | 260 | 96 |
| Others | 4 | - |
| **Deviation from ideal geometry** | | |
| Bond length (Å) | 0.010 | 0.012 |
| Bond angles (°) | 1.453 | 1.520 |
| **Average *B*-factor (Å$^2$)** | | |
| Protein atoms | 48.19 | 57.98 |
| Water molecules | 42.91 | 38.74 |
| Others | 62.23 | - |
| **Ramachandran plot (%)** | | |
| Favored | 95.48 | 90.96 |
| Allowed | 3.53 | 7.06 |
| Rotamer outliers (%) | 5.4 | 5.3 |
| Clashscore | 2 | 6 |
| **PDB id** | 6A34 | 6A35 |

.

### 7.2.3 STRUCTURE SOLUTION AND MODEL BUILDING

Molecular replacement (MR) method using the program Phaser (McCoy et al., 2007) was employed for solving the structure of PhM1Pi. The three-dimensional atomic coordinates of M1Pi from *A. fulgidus* (AfM1Pi, PDB id: 1T5O) having a sequence identity of 46% was used as the search model. For calculation of $R_{free}$ (Brünger, 1992), a total of 5% of the reflections were set aside. The program COOT (Emsley and Cowtan, 2004) was used for building all the residues. Subsequently, water molecules were located

and added from the difference electron density maps. All refinements were carried out using the program Refmac5 (Vagin et al., 2004) embedded in the CCP4 suite. The programs PROCHECK (Laskowski et al., 1993) and MolProbity (Chen et al., 2010) were used to check and to validate the quality of the refined models. The three-dimensional atomic coordinates and the structure factors of both the structures have been deposited in the RCSB Protein Data Bank (Berman et al., 2000) with PDB ids 6A34 and 6A35.

### 7.2.4 SEQUENCE AND STRUCTURE ANALYSIS

All the amino acid sequences of proteins used in this study were retrieved from UniProtKB database (UniProt Consortium, 2018). Sequence similarity searches were performed using the web tool BLAST (Altschul et al., 1990). Multiple sequence alignment (MSA) and structure-based sequence alignment was performed using the programs Clustal Omega (Sievers and Higgins, 2014) and PROMALS3D (Pei et al., 2008), respectively. The output was further rendered using the online web tool ESPript (Gouet et al., 2003). The buried-surface area and solvation-energy gain during dimer formation was calculated using the web server PISA (Krissinel and Henrick, 2007). For the structure-based phylogenetic tree, a distance matrix was generated for all the proteins by calculating their RMSD in PyMOL (PyMOL Molecular Graphics System, Schrodinger, LLC). Subsequently, a dendrogram was generated in Newick (.nwk) file format from the distance matrix using the program dendroUPGMA (GarciaVallvé et al., 1999). The phylogenetic tree was generated in the program MEGA7 (Kumar et al., 2016). The theoretical p*K*a values were calculated using the online web tool H++ (Gordon et al., 2005). The hydrophobicity of amino acid residues at the active-site pocket was calculated using the normalized consensus hydrophobicity scale (Eisenberg et al., 1984) embedded in the program PyMOL. All the three dimensional structures were analyzed and generated using the program PyMOL.

### 7.2.5 MOLECULAR DOCKING

Molecular docking experiments of PhM1Pi and the substrate (MTR-1-P) was performed using the program Autodock version 4.0 (Morris et al., 2009). The atomic coordinates of the substrate used for the docking experiments were retrieved from the database PubChem (Kim et al., 2015). During each docking experiment, the receptor (i.e. protein) molecule was kept rigid while the substrate was kept both as rigid and flexible along the rotatable bonds. Since both the rigid and flexible ligands produced similar docking

outcomes, results of only the rigid ligand docking are considered in this chapter. The size of the grid box was kept to 126x126x126 with a spacing of 0.375 Å between the grid points and taking the center of mass of the protein as the grid center. The Lamarckian genetic algorithm (LGA) with a total of 2000 runs was performed to search for the best conformational space of the substrate. The docked conformations of the substrate in each molecular docking experiment were clustered with an RMSD cutoff of 2.0 Å. The substrate conformation having the lowest estimated free energy of binding (EFBE) was selected as the final docked substrate to the protein.

## 7.3. RESULTS

### 7.3.1 THE OVERALL STRUCTURE

The crystal structure was solved by MR method using the atomic coordinates of *A. fulgidus* (AfM1Pi, PDB id: 1T5O), which shares a sequence identity of 46% with M1Pi from *P. horikoshii* (hereafter referred to as PhM1Pi). The crystal structure of PhM1Pi was solved in two different forms viz. Form I ($P3_221$) and Form II ($P1$) which diffracted to a resolution of 2.3 and 2.65 Å, respectively. The asymmetric unit (ASU) of the two forms i.e. $P3_221$ and $P1$ contain two and four protomers, respectively. The overall three-dimensional structure of PhM1Pi belonging to the two different space groups have a strikingly similar structure and superimposes with an RMSD of 0.287 Å. The NTD (residues 1–157) of PhM1Pi protein is made up of five α-helices ($\alpha_1$- $\alpha_5$), a $3_{10}$-helix ($\eta_1$) and three antiparallel β-strands ($\beta_1$-$\beta_3$). On the other hand, the CTD (residues 158–364) adopts a Rossmann-like fold and comprises of five α-helices ($\alpha_6$-$\alpha_{10}$), four $3_{10}$-helices ($\eta_2$-$\eta_5$) and nine, mostly parallel, β-strands ($\beta_4$-$\beta_{12}$). The NTD and CTD are bridged by the longest helix $\alpha_5$ (residues 126–157) (Figure 7.5A). A three-dimensional structural similarity search analysis using DALI server (Holm and Rosenström, 2010) reveals that PhM1Pi is structurally closest to AfM1Pi (PDB id: 1T5O, Z-score: 46.5, RMSD: 1.4 Å) followed by M1Pi from *H. sapiens* (PDB id: 4LDQ, Z-score: 46.0, RMSD: 1.6 Å). It also shares high structural similarity with M1Pi enzymes from *B. subtilis* (PDB id: 2YVK, Z-score: 44.2, RMSD: 3.0 Å), *T. maritima* (PDB id: 1T9K, Z-score: 44.0, RMSD: 3.0 Å) and *L. major* (PDB id: 2A0U, Z-score: 42.2, RMSD: 2.9 Å). Among M1Pi from different organisms, PhM1Pi is closest to AfM1Pi and M1Pi from *H. sapiens* (HsM1Pi) owing to

their similar conformations. PhM1Pi, AfM1Pi and HsM1Pi are present in an open conformation while M1Pi from the other organisms are available in a closed conformation. In addition, PhM1Pi shares a similar structural conformation with that of R15Pi from *T. kodakarensis* (PDB id: 3VM6, Zscore: 30.2, RMSD: 3.1 Å) and *P. horikoshii* (PDB id: 5YFJ, Z-score: 30.0, RMSD: 3.1 Å) as well as eIF2Bα from *H. sapiens* (PDB id: 3ECS, Z-score: 24.3, RMSD: 4.0 Å), aIF2Bα from *P. horikoshii* (PDB id: 1VB5, Z-score: 26.1, RMSD: 3.0 Å) and the regulatory subunits of eIF2B from *S. pombe* (PDB id: 5B04, Z-score: 26.2–23.5, RMSD: 3.1–4.2 Å) and *C. thermophilum* (PDB id: 5DBO, Z-score: 22.8–19.8, RMSD: 3.5–5.0 Å). The overall monomeric structure of PhM1Pi is also undistinguishable from the other two members of PF01008 family. Similar to the secondary structure arrangements of PhM1Pi, the NTD comprises mostly of α-helices, the CTD exhibits a Rossmann-like fold and are connected by a long helix in R15Pi as well as in the regulatory subunits of eIF2B (Figure 7.5B).

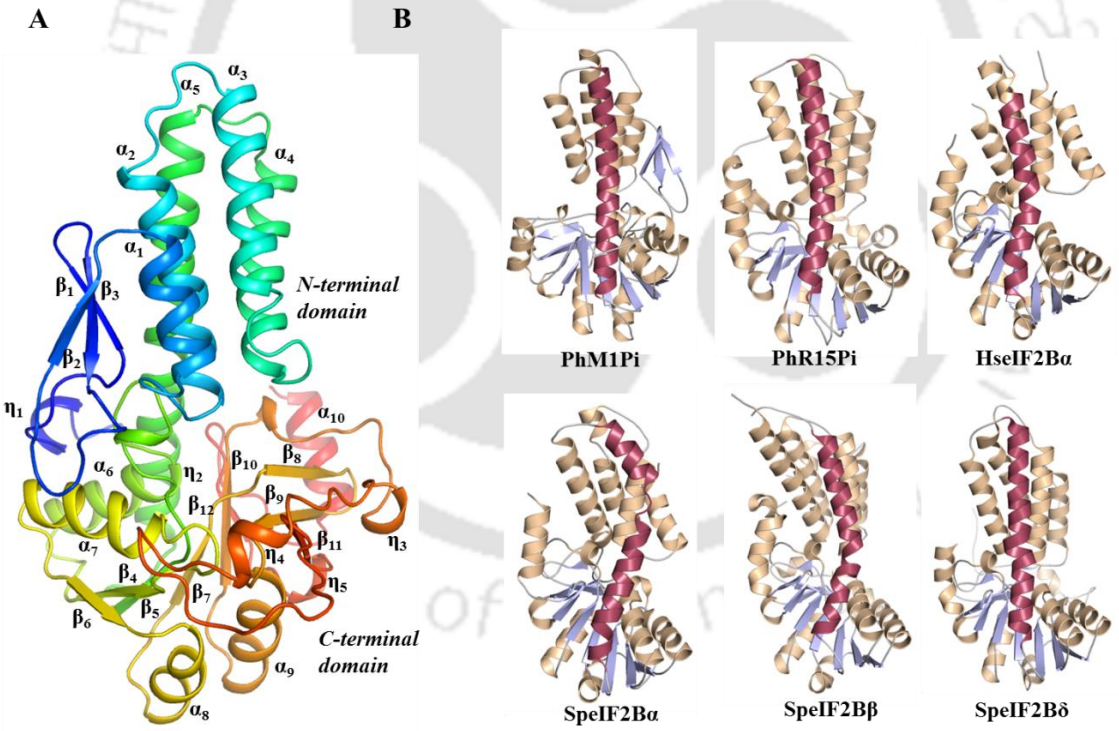



**Figure 7.5. Tertiary structure of PhM1Pi.** (A) Structure of monomer of PhM1Pi represented as cartoon model. (B) Structural resemblance of members of PF01008 family. Secondary structures of M1Pi from *P. horikoshii* (PhM1Pi); R15Pi from *P. horikoshii* (PhR15Pi, PDB id: 5YFJ); eIF2Bα from *S. pombe* (SpeIF2Bα, PDB id: 5B04), *H. sapiens* (HseIF2Bα, PDB id: 3ECS); eIF2Bβ from *S. pombe* (SpeIF2Bβ, PDB id: 5B04) and eIF2Bδ from *S. pombe* (SpeIF2BD, PDB id: 5B04) corresponding to α-

helices, β-strands and loops are colored in wheat, light blue and grey, respectively. The longest helix connecting the NTD and CTD is highlighted in maroon.

The enzyme PhM1Pi forms a homodimer with the help of interactions mediated largely by the CTD while the NTD remains completely free (Figure 7.6A). The buried-surface area and the solvation-energy gain upon dimerization of the two forms of PhM1Pi ($P3_221$ and $P1$) are 5230 & 4240 $Å^2$ and -23.6 & -37.1 kcal $mol^{-1}$, respectively. The hydrogen-bond formation among amino acid residues present in $\alpha_8$, $\alpha_9$, $\beta_{11}$, the loop regions between $\alpha_7$-$\beta_4$ and $\eta_3$-$\beta_{11}$ in the two participating monomers predominantly anchor the two protomers. Apart from the hydrogen bonds, two disulfide bonds involving Cys307A & Cys310B and Cys310A & Cys307B also participate in the dimer stabilization of PhM1Pi (Figure 7.6A). The role of these cysteine residues in dimer formation was further confirmed by performing SDS-PAGE of the PhM1Pi under reducing and non-reducing conditions which shows two protein bands corresponding to monomeric (~40 kDa) and dimeric (~80 kDa) forms, respectively (Figure 7.6B). The manner of dimerization in PhM1Pi is highly comparable to the dimer formation of R15Pi and the regulatory subunits of eIF2B (Figure 7.6C).

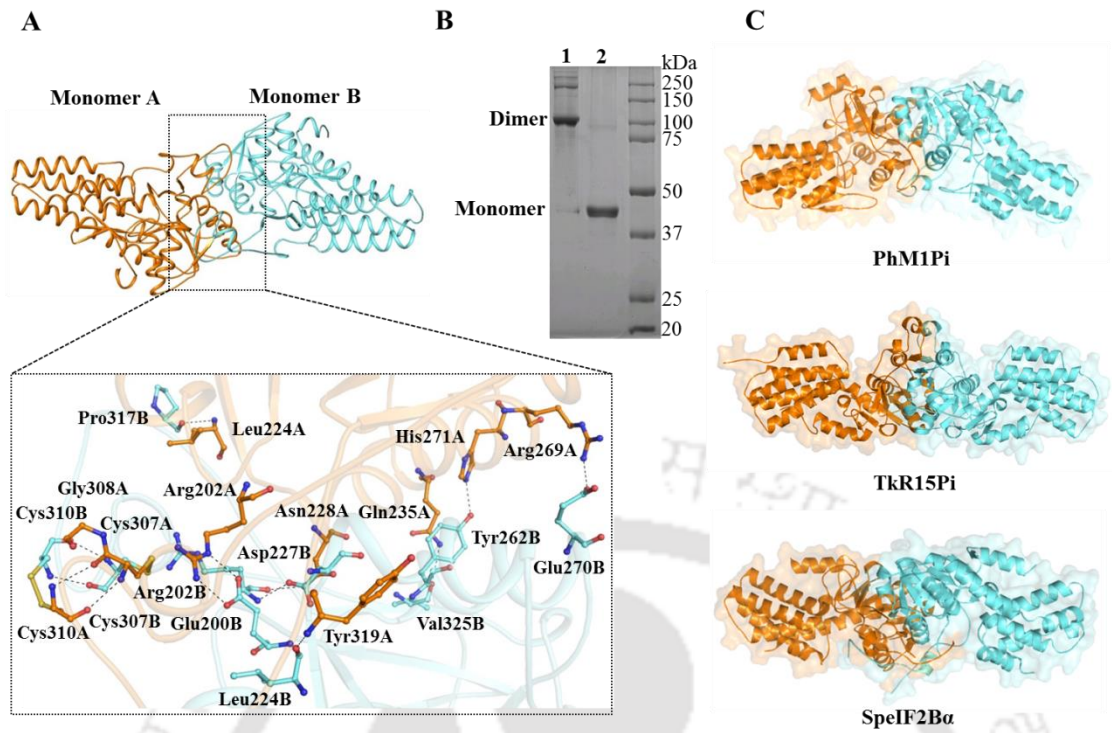



**Figure 7.6. Quarternary structure of PhM1Pi.** (A) Dimeric structure of PhM1Pi. The two subunits are shown in orange and cyan. The amino acid residues belonging to monomers A and B involved in the dimer formation are labeled and shown in ball-and-stick model. (B) SDS-PAGE of PhM1Pi in (1) absence and (2) presence of β-ME. The last lane represents the protein markers. (C) Structural comparison of the dimers of the members of PF01008 family. Dimer formation is shown for M1Pi from *P. horikoshii* (PhM1Pi), R15Pi from *T. kodakarensis* (TkR15Pi, PDB id: 3A11) and eIF2Bα from *S. pombe* (SpeIF2Bα, PDB id: 5B04). The monomers are shown in cyan and orange.

### 7.3.2 THE OPEN AND CLOSED CONFORMATIONS

The open and closed states of the proteins belonging to the family PF01008 are usually demarcated by the degree of angle formed owing to the bend in the longest helix $\alpha_5$. An open conformation of the protein R15Pi from *T. kodakarensis* (TkR15Pi) is kinked at the helix $\alpha_5$ by ~45° while in the closed conformation the bend angle reduces to ~25° (Nakamura et al., 2012). To obtain the bend angle formed in the open conformation of M1Pi, the helix $\alpha_5$ of PhM1Pi was analyzed; interestingly, no visible kink formation was observed. Similar to the closed conformation of M1Pi from *B. subtilis* (BsM1Pi, MTRu-1-P bound), the bend angle of helix $\alpha_5$ in PhM1Pi remains ~10° (Figure 7.7A). Thus, another structural feature that could be responsible for distinguishing the open and closed conformations was further explored. A careful comparison between PhM1Pi (open) and

BsM1Pi (closed) reveals that the entire NTD of M1Pi moves by ~15 Å towards the CTD while transforming from an open to closed state (Figure 7.7B). The most visible difference between the open and closed conformation could be perceived in the position of the loop connecting the helices $\alpha_3$ and $\alpha_4$ of the NTD. In the open state, the loop is away from the active-site pocket, whereas in the closed state the loop moves toward the active-site pocket with a forward shift of ~24 Å. (Figure 7.7C).

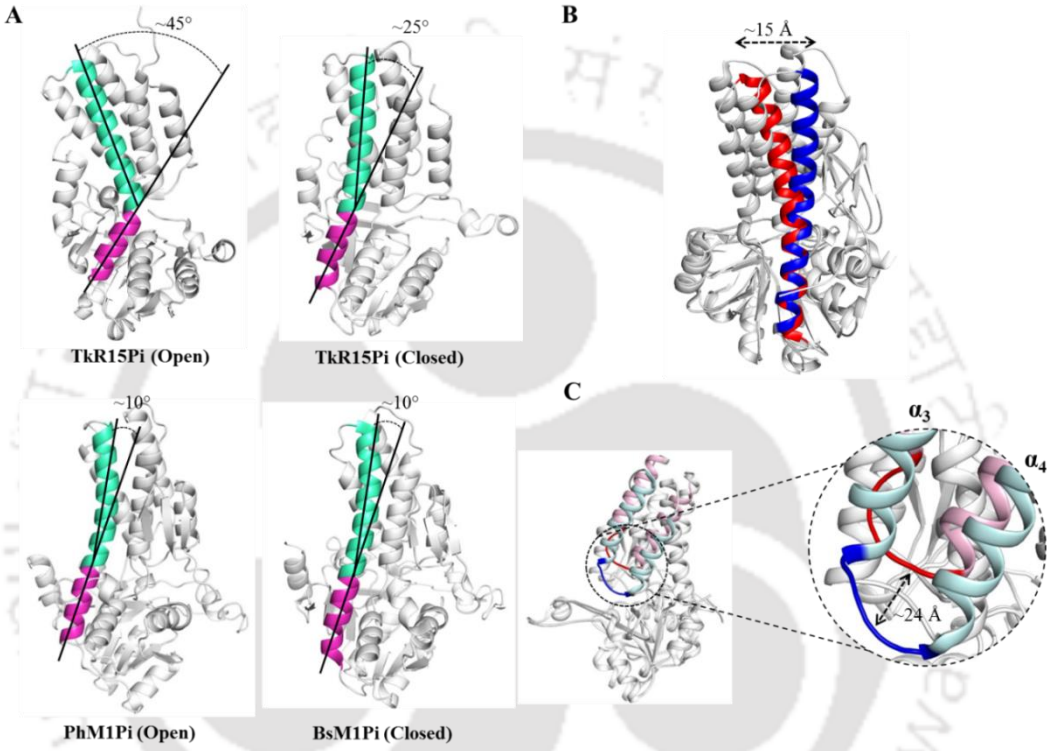



**Figure 7.7. Domain movement of PhM1Pi.** (A) Comparison between the bend angles formed due to the kink formation in the helix $\alpha_5$ (shown in green and magenta) of R15Pi (PDB ids: 3A11 and 3VM6) and M1Pi during open and closed conformations. The region of kink formation lies at the junction of the two colors. (B) Movement of NTD towards CTD with a shift of ~15 Å. (C) forward shift (~24 Å) of the loop connecting helices $\alpha_3$ and $\alpha_4$ during transition from an open (PhM1Pi) to closed (BsM1Pi, PDB id: 2YVK) conformation. The helix $\alpha_5$ and the loop between helices $\alpha_3$ and $\alpha_4$ in the open and closed states are highlighted in red and blue, respectively.

### 7.3.3 THE N-TERMINAL DOMAIN

To understand the structural evolution of the N-terminal domain (NTD) of PhM1Pi, a structure-based phylogeny tree including the protein members from the family PF01008 was deduced. Among the regulatory subunits of eIF2B, only the α-subunit has been

considered for further analyses as it exhibits the maximum similarity (query coverage: 72% & identity: 27%) compared to β (query coverage: 38% and identity: 23%) and δ (not significant) subunits with M1Pi. The phylogenetic analysis reveals that the NTD of PhM1Pi along with other M1Pi proteins is closer to R15Pi in comparison to eIF2Bα (Figure 7.8A). The NTD of all the three proteins, however, have evolved from a common ancestor having a common core structure. To validate the structural phylogeny, a sequence-based phylogenetic analysis using the structure-based MSA was also performed. A similar structural evolutionary relationship could be obtained among the proteins M1Pi, R15Pi and eIF2Bα. Quite interestingly, the phylogram also reveals that the NTD of M1Pi is more evolved than that of R15Pi (Figure 7.8B). Thus, to trace the occurrence of evolution at the tertiary structure level, a structure-based MSA was performed. Results reveal the presence of an N-terminal extension in M1Pi, absent in both R15Pi and eIF2Bα. This N-terminal extension (residues 1-40) comprises of three antiparallel β-strands ($\beta_1$-$\beta_3$) and a $3_{10}$-helix ($\eta_1$) (Figure 7.8C). Absence of this substructure in the other two members of the family PF01008 substantiates the fact that the tertiary structure of M1Pi has evolved to a greater extent.

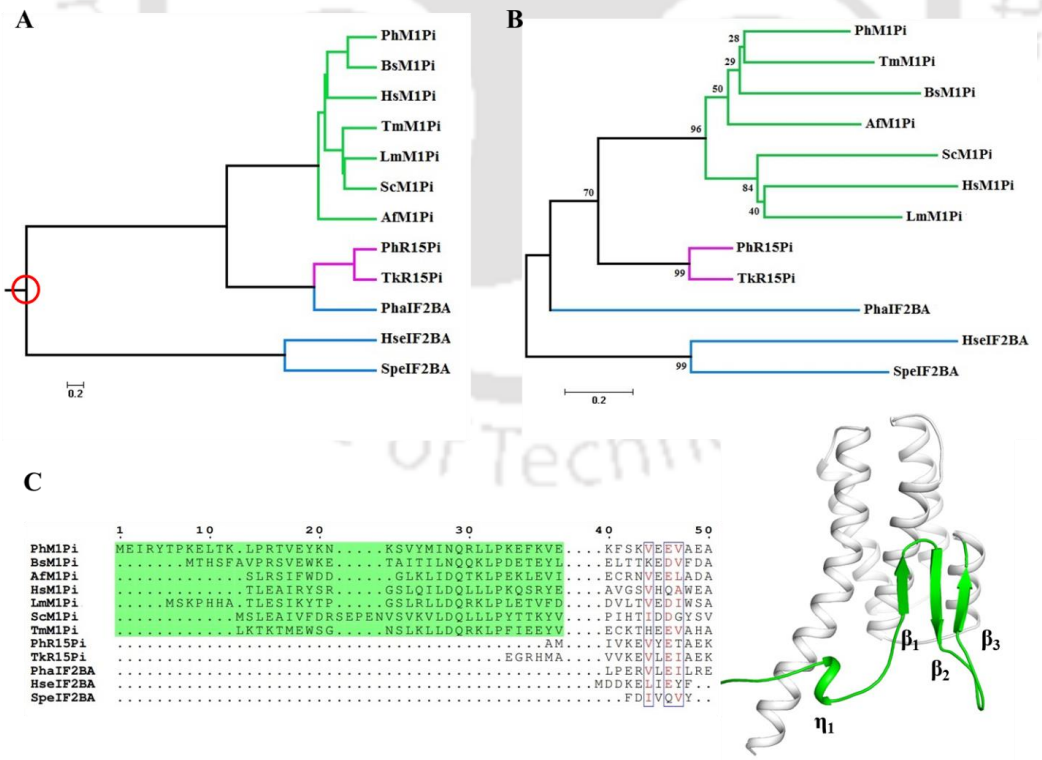



**Figure 7.8. Structural evolution of the NTD of M1Pi.** (A) Structure- and (B) sequence-based phylogenetic tree of the NTDs of M1Pi, R15Pi and eIF2Bα. Both the

trees depict a clear clustering of M1Pi (green) with R15Pi (magenta) while eIF2Bα (blue) forms a separate clade. The ancestral core structure is highlighted with red circle. The scale bar indicates 0.2 amino acid substitutions per single site. (C) Structure-based multiple sequence alignment (MSA) of NTDs of M1Pi, R15Pi and eIF2Bα. The N-terminal extension unique to M1Pi is highlighted in green. The members of the family PF01008 used for analysis are as follows: M1Pi from *B. subtilis* (BsM1Pi, PDB id: 2YVK), *A. fulgidus* (AfM1Pi, PDB id: 1T5O), *H. sapiens* (HsM1Pi, PDB id: 4LDQ), *L. major* (LmM1Pi, PDB id: 2A0U), *S. cerevisiae* (ScM1Pi, PDB id: 1W2W) and *T. maritima* (TmM1Pi, PDB id: 1T9K); R15Pi from *P. horikoshii* (PhR15Pi, PDB id: 5YFJ) and *T. kodakarensis* (TkR15Pi, PDB id: 3A11); aIF2Bα from *P. horikoshii* (PhaIF2BA, PDB id: 1VB5), eIF2Bα from *H. sapiens* (HseIF2BA, PDB id: 3ECS) and *S. pombe* (SpeIF2BA, PDB id: 5B04).

### 7.3.4 THE C-TERMINAL DOMAIN

Structure-based phylogeny tree of the C-terminal domain (CTD) of PhM1Pi along with M1Pi from other organisms as well as R15Pi and eIF2Bα was built to trace their structural (un)relatedness. Unlike in the case of NTD, the CTD of R15Pi and eIF2Bα are similar while PhM1Pi along with M1Pi proteins from other organisms form a distinct separate clade (Figure 7.9A). The difference in the CTD of M1Pi is rendered by the presence of an eight amino acid long insert (residue 170-177) unavailable in both R15Pi and eIF2Bα proteins. In M1Pi, this insert corresponds to a $3_{10}$-helix ($\eta_2$) and is made up of hydrophobic amino acid residues (Figure 7.9B). Another difference between the CTD of M1Pi and R15Pi & eIF2Bα lies in the orientation of the hindmost helix $\alpha_{10}$. When compared to the corresponding helices in R15Pi and eIF2Bα, helix $\alpha_{10}$ exhibits a visible outward shift of ~9.7 and ~14 Å, respectively (Figure 7.9C). This outward conformation of helix $\alpha_{10}$ would result in steric clashes during hexamer formation of M1Pi, thus compelling it to be present in a dimeric state (Figure 7.9D).

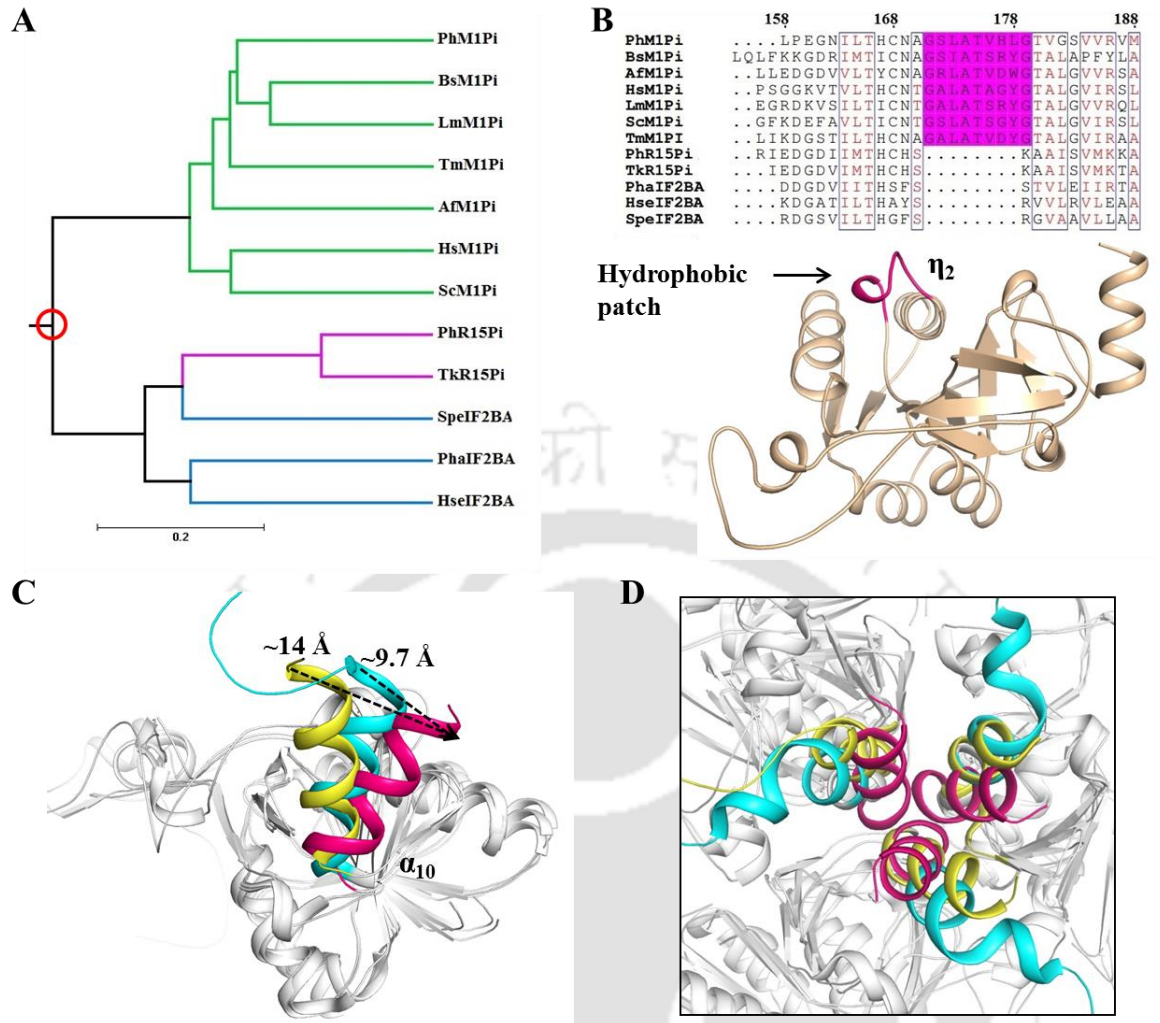



**Figure 7.9. Structural evolution of the CTD of M1Pi.** (A) Structure-based phylogenetic tree of the CTDs of M1Pi, R15Pi and eIF2Bα. M1Pi (green) forms a single separate clade while R15Pi (magenta) and eIF2Bα (blue) cluster together. The ancestral core structure is highlighted with red circle. The scale bar indicates 0.2 amino acid substitutions per single site. (B) Structure-based multiple sequence alignment (MSA) of CTDs of M1Pi, R15Pi and eIF2Bα. The hydrophobic patch unique to M1Pi is highlighted in magenta. The members of the family PF01008 used for phylogenetic tree analysis and MSA are as follows: M1Pi from *P. horikoshii* (PhM1Pi), *B. subtilis* (BsM1Pi, PDB id: 2YVK), *A. fulgidus* (AfM1Pi, PDB id: 1T5O), *H. sapiens* (HsM1Pi, PDB id: 4LDQ), *L. major* (LmM1Pi, PDB id: 2A0U), *S. cerevisiae* (ScM1Pi, PDB id: 1W2W) and *T. maritima* (TmM1Pi, PDB id: 1T9K); R15Pi from *P. horikoshii* (PhR15Pi, PDB id: 5YFJ) and *T. kodakarensis* (TkR15Pi, PDB id: 3A11); aIF2Bα from *P. horikoshii* (PhaIF2BA, PDB id: 1VB5), eIF2Bα from *H. sapiens* (HseIF2BA, PDB id: 3ECS) and *S. pombe* (SpeIF2BA, PDB id: 5B04). (C) Superimposition of CTDs of PhM1Pi (magenta), TkR15Pi (cyan, PDB id: 3A11) and SpeIF2Bα (yellow, PDB id: 5B04) depicts the outward shift of helix $\alpha_{10}$ in PhM1Pi. (D) The steric hindrance caused by the helix $\alpha_{10}$ of PhM1Pi (magenta) during hexamer formation. The corresponding helices in R15Pi and eIF2Bα, which are situated considerably apart in the hexameric state, are highlighted in cyan and yellow, respectively.

### 7.3.5 THE HYDROPHOBIC ACTIVE-SITE POCKET

A comparison of the water molecules in PhM1Pi (open) and BsM1Pi (closed, MTRu-1-P bound) was carried out to identify the conserved water molecules in the active-site pocket of M1Pi enzymes. The active-site pocket of PhM1Pi, which is solvent accessible, contains around 20 water molecules (Figure 7.10A). Upon substrate binding, these water molecules are replaced by the substrate itself as well as other structural elements of the NTD such as the loop connecting helices $\alpha_3$ and $\alpha_4$. Three water molecules (referred to as W1, W2 and W3), however, are retained via hydrogen-bond interaction with amino acid residues at the active-site pocket. In PhM1Pi, the water molecule W1 makes hydrogen bond interaction with the main chain carbonyl oxygen of His166, amide nitrogen of Gly245 and side chain $N^{\delta 2}$ of Asn257. Another water molecule W2 forms hydrogen bond with the side chain $O^{\gamma}$ and backbone carbonyl oxygen atom of Thr165 and His166, respectively. In the closed state (BsM1Pi), it (W505) makes further hydrogen-bond interaction with the main chain carbonyl oxygen atom of Ile159 & the catalytic residue Cys160 and the amide nitrogen of Ala162. The third water molecule W3, in the open conformation, does not make any contact with the protein, however, in the closed state (BsM1Pi), it (W516) interacts with the main chain carbonyl oxygen of Ala239 and the side chain oxygen $O^{\gamma}$ of Thr276 (Figure 7.10B). Thus, the transition from an open to closed state expels out majority of the water molecules from the active-site pocket.

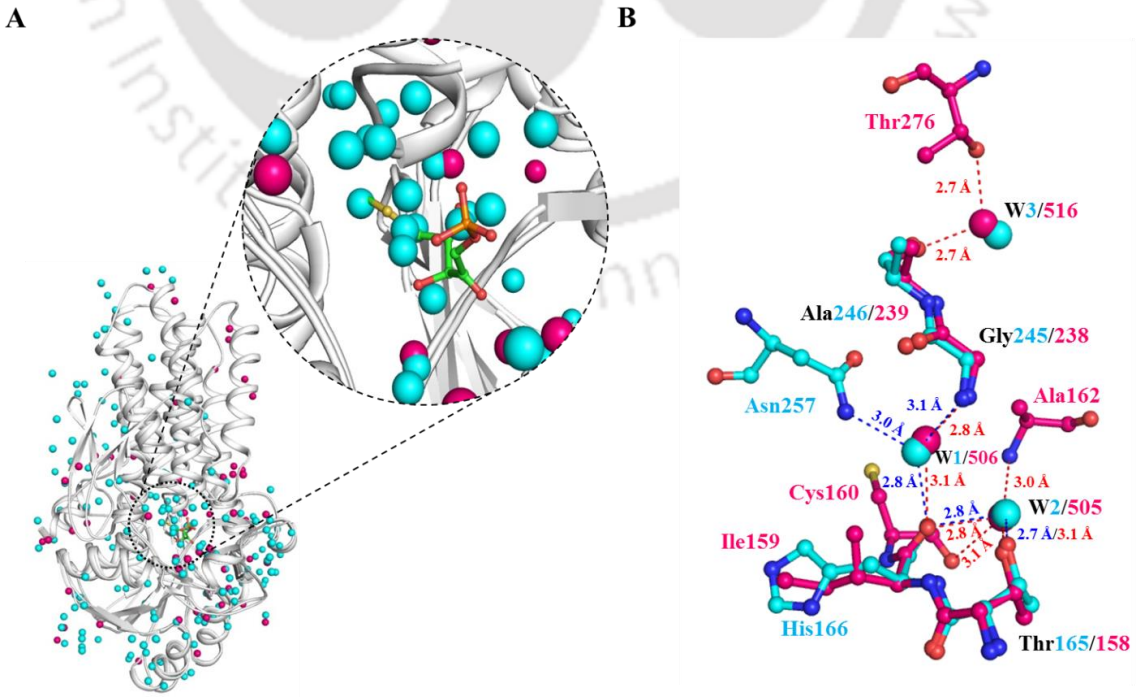


**Figure 7.10. Water molecules in the open and closed states of M1Pi.** (A) Repertoire of water molecules at the active-site pocket of PhM1Pi (open conformation) and BsM1Pi (closed conformation). The water molecules bound to the open and closed conformations of M1Pi are shown in cyan and magenta spheres, respectively. (B) Conserved water molecules in the active-site pocket interacting with the amino acid residues of PhM1Pi and BsM1Pi (PDB id: 2YVK) are shown as ball-and-stick model in cyan and pink, respectively. The hydrogen-bond interactions of amino acid residues of PhM1Pi and BsM1Pi with the water molecules are shown with blue and red dashes, respectively. The distances between the atoms are indicated adjacent to their corresponding dashed line.

### 7.3.6 THE REACTION MECHANISM

The two catalytic residues Cys133 and Asp202 crucial for TkR15Pi to accomplish the reaction mechanism via the formation of a *cis*-phosphoenolate intermediate are well conserved (Cys167 and Asp247) in PhM1Pi (Figure 7.11A). Furthermore, the fine-tuning of the electrostatic states of the two catalytic residues, Cys167 in a deprotonated and Asp247 in a protonated state, becomes mandatory for the enzyme to enter a reaction-ready state. This modulation of the electrostatic states is accomplished by the repertoire of amino acid residues surrounding the catalytic residues. PhM1Pi, like R15Pi, possesses a similar set of conserved amino acid residues surrounding the two catalytic residues Cys167 and Asp247 (Figure 7.11A). The catalytic cysteine residue is surrounded by four amino acid residues Arg57, Asn168, Gln206 and Lys258. The main chain amide group of the adjacent Asn168 (Asn161 in BsM1Pi) interacts with the thiol group of Cys167 at the active-site pocket. Moreover, a transition from an open to closed conformation causes the guanidinyl group of Arg57 (Arg51 in BsM1Pi) and the side chain amine group of Gln206 (Gln199 in BsM1Pi) & Lys258 (Lys251 in BsM1Pi) to approach the catalytic cysteine residue (Figure 7.11B). Out of these four residues, the positive charge of the side chain amine group of Arg57 and Lys258 might get saturated owing to their hydrogen bond interaction to the phosphate moiety of the substrate or product (in case of BsM1Pi, PDB id: 2YVK) as well as to the side chain amine group of Asn168 (Figure 7.12). However, the availability of the main chain amide of Asn168 and side chain amine of Gln206 could still lower the p*K*a of the thiol group of the residue Cys167 allowing it to remain in a deprotonated state (Kortemme et al., 1996; Naor and Jensen, 2004; Jao et al., 2006, Nakamura et al., 2012). Furthermore, the theoretical calculation of p*K*a of Cys167 (Cys160 in BsM1Pi) in the open conformation is found to be ~7.8 which decreases to ~6.4 in the closed conformation. On the other hand, the other catalytic residue Asp247 is surrounded by hydrophobic amino acid residues Ala59, Thr102,

Ala103, Val104, Ile259 and Phe323 (Figure 7.11C). The closing up of the active-site residues causes the water molecules to be eliminated in its vicinity further creating a hydrophobic environment around the residue Asp247. This would facilitate an increase in the p*K*a of Asp247 suggesting it to be present in a protonated form (Figure 7.11C). A highly hydrophobic microenvironment as well as a conserved optimal repertoire of amino acid residues in the active-site pocket might favor the reaction mechanism of M1Pi to occur via the formation of a *cis*-phosphoenolate intermediate.

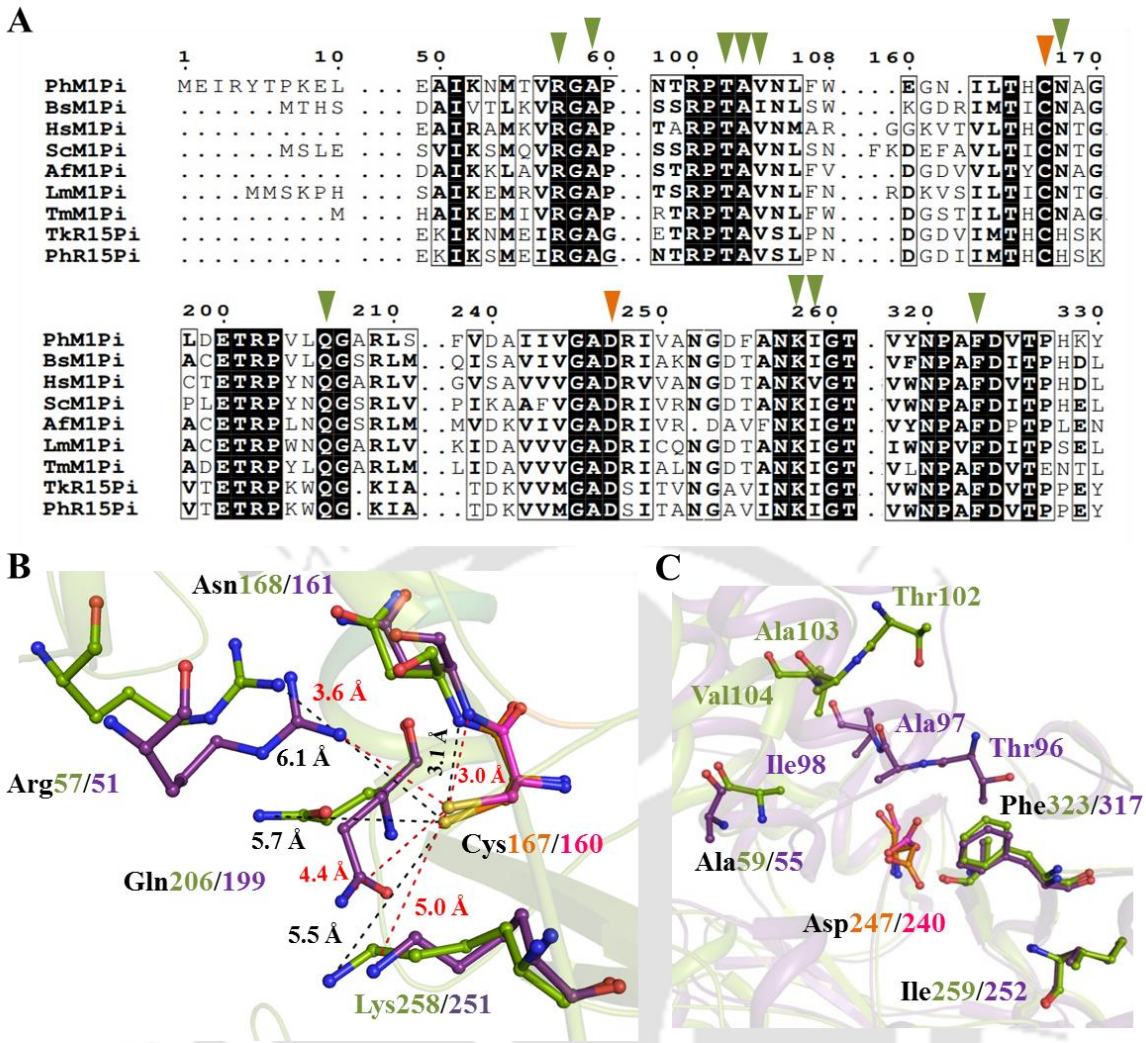



**Figure 7.11. Microenvironment of the active-site pocket of M1Pi.** (A) Multiple sequence alignment (MSA) of M1Pi and R15Pi depicting the conserved amino acid residues making the microenvironment of the two catalytic residues Cys167 and Asp247. These residues along with the conserved amino acid residues surrounding them are highlighted with orange and green downward arrowheads, respectively. The fully- and partially-conserved residues are enclosed in black box and shown in bold, respectively. (B) and (C) Change in the conformation of the amino acid residues around the catalytic residues shown as ball-and-stick model in orange/magenta during transition from an open (green) to closed (purple) conformation enables M1Pi to enter a reaction-ready state. The difference in the distance between atoms in open (black) and closed (red) state is indicated along the dashed lines.

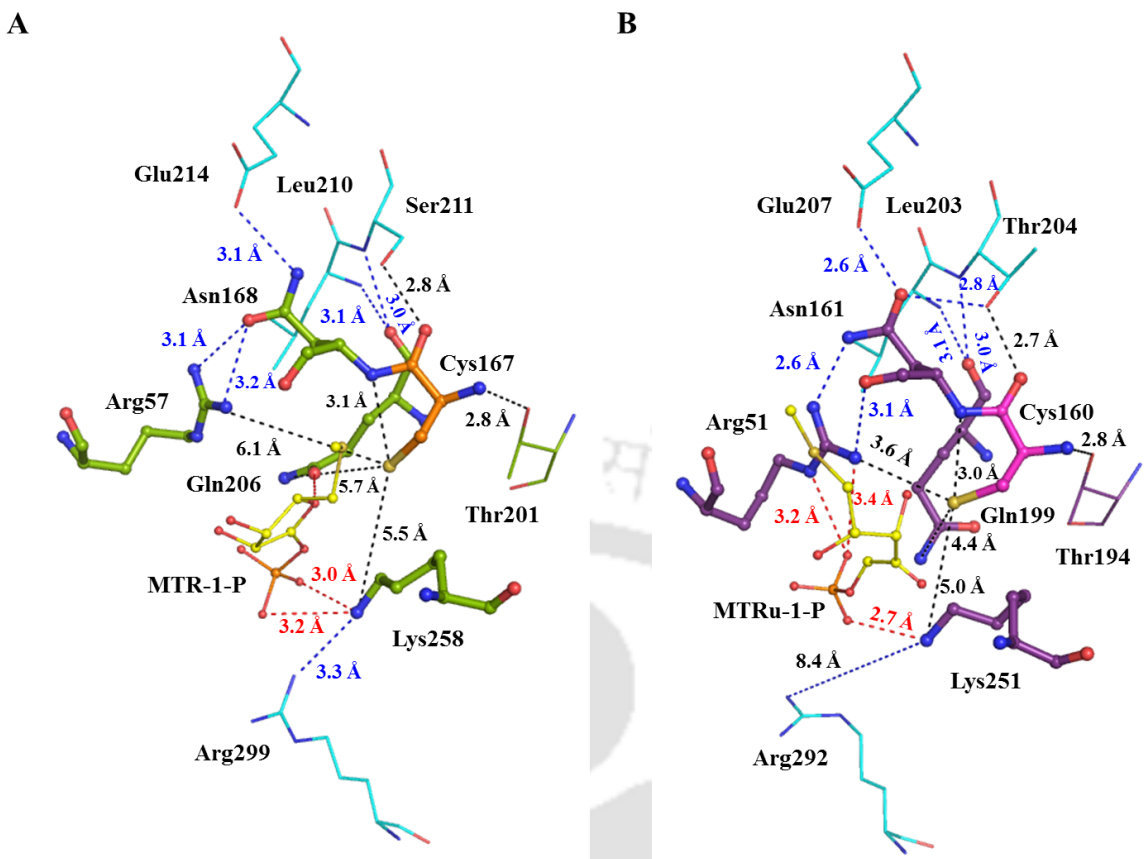



**Figure 7.12. Hydrogen bonding network of polar residues at the active-site pocket.** (A) Open and (B) closed conformation of M1Pi. The polar amino acid residues responsible for lowering the p$K_a$ of the thiol group of Cys167/160 (PhM1Pi/BsM1Pi in orange/magenta) are shown as green and purple ball-and-stick model in the open and closed conformations, respectively. The amino acid residues forming hydrogen bond interaction (blue dashed lines) with the polar amino acid residues are shown as cyan lines. The hydrogen bond interactions between the polar amino acid residues and the substrate (in open conformation) & product (in closed conformation) are shown as red dashed lines. The distances between atoms are mentioned along the dashed line. The substrate and product are shown as yellow ball-and-stick model.

During the reaction mechanism, the first step involves the donation of a proton from the protonated Asp247 to the O4 oxygen of the substrate promoting the opening of the ring structure of MTR-1-P. In the second step, the thiolate anion formed by the deprotonation of Cys167 abstracts a proton from C2 position of the substrate. The abstraction is accompanied by the formation of a double bond between C1 and C2, thus giving rise to the *cis*-phosphoenolate intermediate. The residue Asp247 is placed in the active-site pocket in such a way that it can access O2 in addition to O4. Thus, in the next step, Asp247, which is now in the deprotonated state, abstracts a proton from O2.

Simultaneously, the proton of the thiol group of Cys167 is transferred to the C1=C2 double bond, resulting in addition of a proton to the atom C1 and leading to the generation of the product, MTRu-1-P (Figure 7.13).

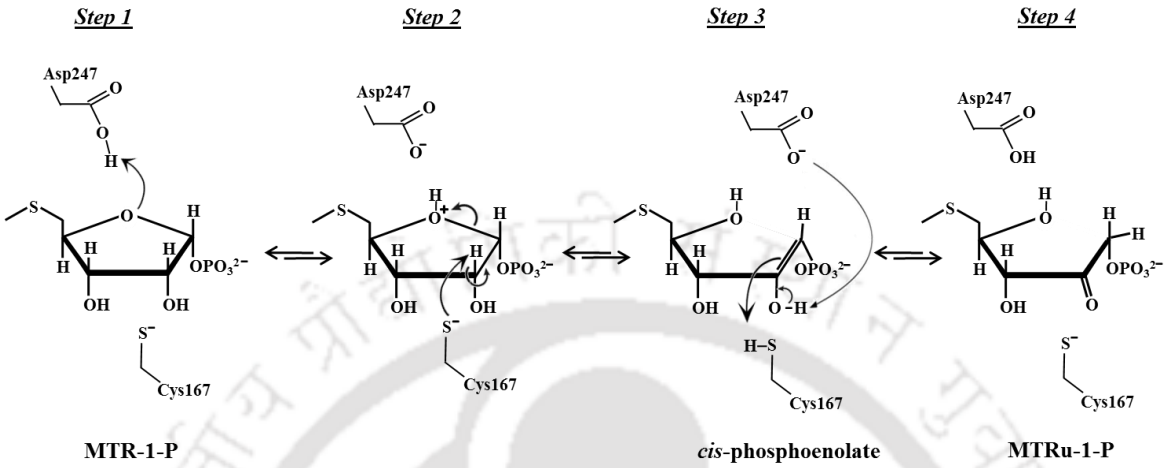



**Figure 7.13. Proposed reaction mechanism of M1Pi via a *cis*-phosphoenolate intermediate.** The catalytic mechanism of M1Pi proceeds through the following steps: *Step 1*, transfer of proton from Asp247 to the O4 atom of MTR-1-P; *Step 2*, proton abstraction from C2 atom of the substrate by the thiolate anion of Cys167 as well as cleavage of the C1-O4 bond; *Step 3*, simultaneous proton abstraction from O2 atom by the deprotonated Asp247 and transfer of proton from the thiol group of Cys167 to C1 atom; *Step 4*, generation of the product, MTRu-1-P.

## 7.4 DISCUSSION

The enzyme M1Pi involved in the universally conserved methionine salvage pathway (MSP) shares high structural similarity with R15Pi and the regulatory subunits (α, β and δ) of eIF2B. Despite this, the proteins M1Pi, R15Pi and eIF2Bα function in completely unrelated cellular processes and show high specificity towards their functions. To perceive the structural features that render functional specificity, a comparison has been drawn among M1Pi, R15Pi and eIF2Bα.

During the MSP, the hydrophobic ligand MTR-1-P occupies the active-site pocket of M1Pi and is later isomerized into MTRu-1-P. Thus, binding of MTR-1-P demands the availability of a hydrophobic active-site in M1Pi isolated from the solvent. The N-terminal extension acquired by M1Pi during evolution shields the active-site pocket from the external solvent. The emergence of this substructure has occurred at the periphery of

the conserved core domain and hence the overall folding of the core structure, which is similar to R15Pi and eIF2Bα, is not affected. The N-terminal extension is assisted by a hydrophobic patch at the CTD. The hydrophobic patch further covers the active-site pocket from the bulk solvent and aids in the accommodation of the hydrophobic methylthio group of the substrate or product. On the other hand, the substrate/product of R15Pi does not have a hydrophobic part and thus the requirement of a strict hydrophobic active-site environment is not essential.

The hydrophobic environment of the active-site pocket of M1Pi is further maintained by the lack of a pronounced kink at the helix $\alpha_5$ which minimizes the difference between the open and the closed conformations. Lack of a bend angle in this helix also aids in covering the hydrophobic patch from the solvent. The loop connecting helices $\alpha_3$ and $\alpha_4$, a structural feature which marks the open/closed conformation of M1Pi, further closes upon the active-site pocket during substrate binding. Thus, the availability of the open state of M1Pi in a 'closed state-like' conformation makes it favorable for the binding of a hydrophobic substrate. Thus, it can be suggested that the three-dimensional structure of the members of family PF01008 must have emerged from a common core structure where M1Pi have further evolved to adapt to its functional obligations.

Initially, two catalytic mechanisms of M1Pi were proposed wherein one involves the formation of a *cis*-phosphoenolate intermediate and the other involves direct hydride transfer. However, the mechanism involving *cis*-phosphoenolate intermediate was later eliminated based on an observation that no solvent deuterium was incorporated onto the atom C1 of the product during the isomerization reaction (Saito et al., 2007). A transition from an open to closed state forces all the water molecules, except three which are held by the active-site residues in both the states, out from the active site of the protein. This indicates that it is highly likely that no solvent molecule can enter the active-site pocket from the external bulk solvent once the protein has attained its closed conformation. Thus, lack of incorporation of deuterium might be due to the highly hydrophobic microenvironment of the active-site pocket which is well shielded from the bulk solvent. On the other hand, those isomerases which proceeds via the *cis*-phosphoenolate intermediate formation have their active-site pocket exposed to the solvent even in the substrate bound form making the exchange of hydrogen and deuterium possible

(O'Donoghue et al., 2005a,b). Henceforth, the *cis*-phosphoenolate mechanism for M1Pi cannot be eliminated based on the inability to exchange protons from the medium.

Apart from maintaining a strict hydrophobic interior, the microenvironment created by the amino acid residues around the two catalytic residues Cys167 and Asp247 of M1Pi seems to be well tuned for the reaction mechanism to proceed via the formation of a *cis*-phosphoenolate intermediate. The presence of a similar microenvironment at the active-site pocket of R15Pi further accentuates the proposition that M1Pi might also follow a similar reaction mechanism. Furthermore, the non-utilization of divalent metal ions such as $Mg^{2+}$ or $Mn^{2+}$ by M1Pi from *B. subtilis* during enzyme catalysis (Saito et al., 2007) discourages the hypothesis that M1Pi would follow a direct hydride transfer mechanism of enzyme catalysis.

## 7.5 CONCLUSION

From this chapter, it can be concluded that M1Pi, although structurally very similar to R15Pi and the regulatory subunits of eIF2B, owns distinct structural attributes that permit M1Pi to establish its functional specificities. A keen analysis of the active-site microenvironment also allowed us to favor the mechanism of enzyme catalysis via the formation of a *cis*-phosphoenolate intermediate. However, to establish the definite mechanism of enzyme catalysis followed by M1Pi, further experimental evidences would be necessary.

# SUMMARY

# SUMMARY

The work presented in this thesis involves the structural understanding of the putative translation initiation factors (IFs) in a hyperthermophilic organism *P. horikoshii* OT3. At the onset, the homologs of eukaryotic translation IFs, eIF1 and eIF2B were identified in *P. horikoshii* OT3. Apart from showing homology towards the eukaryotic translation IFs, the identified open reading frames (ORFs) also reveal homology towards enzymes involved in various metabolic processes such as nucleoside 5′-monophosphate (NMP) degradation pathway, methionine salvage pathway (MSP), etc. An in-depth *in silico* analysis suggests that out of the five ORFs, one (PH1771.1) shows high similairy to both eIF1 as well as to a bacterial translation inhibitor YciH. On the other hand, two ORFs (PH0208 and PH0702) possess all the essential amino acid residues required for the functioning of the enzymes ribose-1,5-bisphosphate isomerase (R15Pi) and 5-methylthioribose 1-phosphate isomerase (M1Pi), respectively. While, PH1022 and PH1697 possess amino acid residues required for the functioning as NDP-sugar pyrophosphorylase (NSPase). To futher confirm their functions, the three-dimensional crystal structures of three proteins (PH1771.1, PH0208 and PH0702) were elucidated.

A careful structural comparison among PH1771.1, eIF1 and YciH led to the identification of functionally important loop regions in PH1771.1 that would assist it to function as archaeal initiation factor 1 (aIF1). Although similar to eIF1, the presence of certain noteworthy differences between the three-dimensional structures of eIF1 and aIF1 further hints towards a slightly divergent mode of action of aIF1 during the archaeal protein translation initiation.

The elucidation of the three-dimensional structure of PH0208 reveals an overall structure highly similar to both R15Pi as well as the regulatory subunits (α, β and δ) of eIF2B. However, the ability to convert the ribose-1,5-bisphosphate (R15P, substrate) to ribulose-1,5-bisphosphate (RuBP, product) during co-crystallization indicates that PH0208 would function as R15Pi. Furthermore, mutation of the residues Cys135 and Asp204, which are the conserved catalytic residues of the enzyme R15Pi, led to the inability of PH0208 to catalyse the conversion of R15P to RuBP. Our study also identified the previously unknown binding site of AMP which is known to be absolutely crucial for the enzymatic activity of the enzyme R15Pi. Along with AMP, two GMP binding sites have also been

identified in the enzyme R15Pi, one of which is common between AMP and GMP. Furthermore, we have exploited the resemblance between R15Pi and the regulatory subunits of eIF2B to propose a model for the regulatory mechanism of eIF2B in eukaryotes.

The three-dimensional structure of enzyme R15Pi was further utilized to understand its structure-function relationship using computational approaches. Molecular dynamics (MD) simulation studies led to the understanding of the importance of the hexameric state and catalytic residues in the structural stability and ligand binding. Our study also identified the invariant water molecules and deciphered their role in maintaining the integrity of the active site and other functionally important regions. Furthermore, molecular docking strudies led to identification of amino acid residue(s) essential for the substrate or product binding at the active-site pocket. Molecular docking studies also led to the identification of an ‘alternate binding site’ (or ‘initial binding site’) for the substrate, R15P and proposal of a ‘substrate sliding’ mechanism for the enzyme R15Pi.

Elucidation of the three-dimensional crystal structure of PH0702 allowed the identification of the structural attributes that allows it to function as the enzyme M1Pi. The presence of an N-terminal extension, a hydrophobic patch, a loop covering the active site give rise to a highly hydrophobic active-site pocket. All these structural attributes along with the presence of a hydrophobic active-site microenvironment and the availability of an optimal set of amino acid residues surrounding the catalytic residues in M1Pi led us to propose its probable enzyme reaction mechanism.

Thus, it can be concluded that eventhough archaea possess the homolog of eIF1, they lack the functional homolog of eIF2B. This raises a concern for the long-held notion that the process of archaeal and eukaryal translation initiation are homologous. Thus, there is a demand for the re-evaluation of the mechanism of archaeal protein translation initiation which might be closer to bacteria or an exclusive mechanism of translation initiation might be persisting in archaea.

# REFERENCES

# REFERENCES


- Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. and Lindahl, E., 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1*, 19-25.
- Aitken, C.E. and Lorsch, J.R., 2012. A mechanistic overview of translation initiation in eukaryotes. *Nature Structural & Molecular Biology*, *19*(6), 568.
- Albers, E., 2009. Metabolic characteristics and importance of the universal methionine salvage pathway recycling methionine from 5′-methylthioadenosine. *IUBMB life*, *61*(12), 1132-1142.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403-410.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389-3402.
- Aono, R., Sato, T., Imanaka, T. and Atomi, H., 2015. A pentose bisphosphate pathway for nucleoside degradation in Archaea. *Nature Chemical Biology*, *11*(5), 355.
- Aono, R., Sato, T., Yano, A., Yoshida, S., Nishitani, Y., Miki, K., Imanaka, T. and 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.
- Asboth, B. and Naray-Szabo, G., 2000. Mechanism of action of D-xylose isomerase. *Current Protein and Peptide Ccience*, *1*(3), 237-254.
- Ashida, H., Saito, Y., Kojima, C., Kobayashi, K., Ogasawara, N. and Yokota, A., 2003. A functional link between RuBisCO-like protein of Bacillus and photosynthetic RuBisCO. *Science*, *302*(5643), 286-290.
- Bais, V.S., Batra, S. and Prakash, B., 2018. Identification of two highly promiscuous thermostable sugar nucleotidylyltransferases for glycorandomization. *The FEBS Journal*, *285*(15), 2840-2855.
- Baker, N.A., Sept, D., Joseph, S., Holst, M.J. and McCammon, J.A., 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, *98*(18), 10037-10041.

- Balamurugan, B., Roshan, M.N.A., Shaahul Hameed, B., Sumathi, K., Senthilkumar, R., Udayakumar, A., Venkatesh Babu, K.H., Kalaivani, M., Sowmiya, G., Sivasankari, P. and Saravanan, S., 2007. PSAP: protein structure analysis package. *Journal of Applied Crystallography*, *40*(4), 773-777.
- Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R. and Leslie, A.G., 2011. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography*, *67*(4), 271-281.
- Bell, S.D. and Jackson, S.P., 1998. Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends in Microbiology*, *6*(6), 222-228.
- Benelli, D. and Londei, P., 2009. Begin at the beginning: evolution of translational initiation. *Research in Microbiology*, *160*(7), 493-501.
- Benelli, D. and Londei, P., 2011. Translation initiation in Archaea: conserved and domain-specific features. *Biochemical Society Transactions*, *39*(1), 89-93.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A. and Wheeler, D.L., 2000. GenBank. *Nucleic Acids Research*, *28*(1), 15-18.
- Berman, H.M., Bhat, T.N., Bourne, P.E., Feng, Z., Gilliland, G., Weissig, H. and Westbrook, J., 2000. The Protein Data Bank and the challenge of structural genomics. *Nature Structural & Molecular Biology*, *7*(11s), 957.
- Berrisford, J.M., Hounslow, A.M., Akerboom, J., Hagen, W.R., Brouns, S.J., van der Oost, J., Murray, I.A., Blackburn, G.M., Waltho, J.P., Rice, D.W. and Baker, P.J., 2006. Evidence supporting a cis-enediol-based mechanism for *Pyrococcus furiosus* phosphoglucose isomerase. *Journal of Molecular Biology*, *358*(5), 1353-1366.
- Betlach, M., Friedman, J., Boyer, H.W. and Pfeifer, F., 1984. Characterization of a halobacterial gene affecting bacterio-opsin gene expression. *Nucleic Acids Research*, *12*(20), 7949-7959.
- Binkowski, T.A., Naghibzadeh, S. and Liang, J., 2003. CASTp: computed atlas of surface topography of proteins. *Nucleic Acids Research*, *31*(13), 3352-3355.
- Blow, D.M., Collyer, C.A., Goldberg, J.D. and Smart, O.S., 1992. Structure and mechanism of D-xylose isomerase. *Faraday Discussions*, *93*, 67-73.
- Bogorad, A.M., Xia, B., Sandor, D.G., Mamonov, A.B., Cafarella, T.R., Jehle, S., Vajda, S., Kozakov, D. and Marintchev, A., 2014. Insights into the architecture of the eIF2Bα/β/δ regulatory subcomplex. *Biochemistry*, *53*(21), 3432-3445.

- Brandi, L., Marzi, S., Fabbretti, A., Fleischer, C., Hill, W.E., Gualerzi, C.O. and Lodmell, J.S., 2004. The translation initiation functions of IF2: targets for thiostrepton inhibition. *Journal of Molecular Biology*, *335*(4), 881-894.
- Brito, J.A., Borges, N., Vonrhein, C., Santos, H. and Archer, M., 2011. Crystal structure of Archaeoglobus fulgidus CTP: inositol-1-phosphate cytidylyltransferase, a key enzyme for di-myo-inositol-phosphate synthesis in (hyper) thermophiles. *Journal of Bacteriology*, *193*(9), 2177-2185.
- Brünger, A.T., 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, *355*(6359), 472.
- Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D. and Kerlavage, A.R., 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*, *273*(5278), 1058-1073.
- Bumann, M., Djafarzadeh, S., Oberholzer, A.E., Bigler, P., Altmann, M., Trachsel, H. and Baumann, U., 2004. Crystal structure of yeast Ypr118w, a methylthioribose-1-phosphate isomerase related to regulatory eIF2B subunits. *Journal of Biological Chemistry*, *279*(35), 37087-37094.
- Bussi, G., Donadio, D. and Parrinello, M., 2007. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, *126*(1), 014101.
- Carter, A.P., Clemons, W.M., Brodersen, D.E., Morgan-Warren, R.J., Hartsch, T., Wimberly, B.T. and Ramakrishnan, V., 2001. Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science*, *291*(5503), 498-501.
- Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, *66*(1), 12-21.
- Cheung, Y.N., Maag, D., Mitchell, S.F., Fekete, C.A., Algire, M.A., Takacs, J.E., Shirokikh, N., Pestova, T., Lorsch, J.R. and Hinnebusch, A.G., 2007. Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start codon selection in vivo. *Genes & development*, *21*(10), 1217-1230.
- Chu, D. and von der Haar, T., 2012. The architecture of eukaryotic translation. *Nucleic Acids Research*, *40*(20), 10098-10106.

- Cigan, A.M., Bushman, J.L., Boal, T.R. and Hinnebusch, A.G., 1993. A protein complex of translational regulators of GCN4 mRNA is the guanine nucleotide-exchange factor for translation initiation factor 2 in yeast. *Proceedings of the National Academy of Sciences*, *90*(11), 5350-5354.
- Condò, I., Ciammaruconi, A., Benelli, D., Ruggero, D. and Londei, P., 1999. Cis-acting signals controlling translational initiation in the thermophilic archaeon *Sulfolobus solfataricus*. *Molecular Microbiology*, *34*(2), 377-384.
- Cort, J.R., Koonin, E.V., Bash, P.A. and Kennedy, M.A., 1999. A phylogenetic approach to target selection for structural genomics: solution structure of YciH. *Nucleic Acids Research*, *27*(20), 4018-4027.
- Coureux, P.D., Lazennec-Schurdevin, C., Monestier, A., Larquet, E., Cladiere, L., Klaholz, B.P., Schmitt, E. and Mechulam, Y., 2016. Cryo-EM study of start codon selection during archaeal translation initiation. *Nature Communications*, *7*, 13366.
- Dahlquist, K.D. and Puglisi, J.D., 2000. Interaction of translation initiation factor IF1 with the E. coli ribosomal A site. *Journal of Molecular Biology*, *299*(1), 1-15.
- Darden, T., York, D. and Pedersen, L., 1993. Particle mesh Ewald: An N·log (N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*(12), 10089-10092.
- Dauter, Z., 1999. Data-collection strategies. *Acta Crystallographica Section D: Biological Crystallography*, *55*(10), 1703-1717.
- Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall III, W.B., Snoeyink, J., Richardson, J.S. and Richardson, D.C., 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research*, *35*(suppl_2), W375-W383.
- Dennis, P.P., 1997. Ancient ciphers: translation in Archaea. *Cell*, *89*(7), 1007-1010.
- Dev, K., Santangelo, T.J., Rothenburg, S., Neculai, D., Dey, M., Sicheri, F., Dever, T.E., Reeve, J.N. and Hinnebusch, A.G., 2009. Archaeal aIF2B interacts with eukaryotic translation initiation factors eIF2α and eIF2Bα: implications for aIF2B function and eIF2B regulation. *Journal of Molecular Biology*, *392*(3), 701-722.
- Dodson, G.G., Lane, D.P. and Verma, C.S., 2008. Molecular simulations of protein dynamics: new windows on mechanisms in biology. *EMBO Reports*, *9*(2), 144-150.

- Donovan, R.S., Robinson, C.W. and Glick, B.R., 1996. Optimizing inducer and culture conditions for expression of foreign proteins under the control of thelac promoter. *Journal of Industrial Microbiology*, *16*(3), 145-154.
- Drenth, J., 2007. *Principles of protein X-ray crystallography*. Springer Science & Business Media.
- Eisenberg, D., Schwarz, E., Komarony, M. and Wall, R., 1984. Amino acid scale: Normalized consensus hydrophobicity scale. *Journal of Molecular Biology*, *179*, 125-142.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A. and Sonnhammer, E.L.L., 2018. The Pfam protein families database in 2019. *Nucleic Acids Research*, *47*(D1), D427-D432.
- Emsley, P. and Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, *60*(12), 2126-2132.
- Erickson, F. and Hannig, E.M., 1996. Ligand interactions with eukaryotic translation initiation factor 2: role of the gamma-subunit. *The EMBO Journal*, *15*(22), 6311-6320.
- Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. and Pedersen, L.G., 1995. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, *103*(19), 8577-8593.
- Evans, P.R. and Murshudov, G.N., 2013. How good are my data and what is the resolution?. *Acta Crystallographica Section D: Biological Crystallography*, *69*(7), 1204-1214.
- Fenn, T.D., Ringe, D. and Petsko, G.A., 2004. Xylose isomerase in substrate and inhibitor michaelis states: atomic resolution studies of a metal-mediated hydride shift. *Biochemistry*, *43*(21), 6464-6474.
- Finn, M.W. and Tabita, F.R., 2004. Modified pathway to synthesize ribulose 1, 5-bisphosphate in methanogenic archaea. *Journal of Bacteriology*, *186*(19), 6360-6366.
- Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A. and Salazar, G.A., 2015. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279-D285.

- Fletcher, C.M., Pestova, T.V., Hellen, C.U. and Wagner, G., 1999. Structure and interactions of the translation initiation factor eIF1. *The EMBO Journal*, *18*(9), 2631-2637.
- Gäbel, K., Schmitt, J., Schulz, S., Näther, D.J. and Soppa, J., 2013. A comprehensive analysis of the importance of translation initiation factors for *Haloferax volcanii* applying deletion and conditional depletion mutants. *PLoS ONE*, *8*(11), e77188.
- Garcia-Vallvé, S., Palau, J. and Romeu, A., 1999. Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in *Escherichia coli* and *Bacillus subtilis*. *Molecular Biology and Evolution*, *16*(9), 1125-1134.
- Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D. and Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook* (571-607). Humana press.
- Gogoi, P. and Kanaujia, S.P., 2018. A presumed homologue of the regulatory subunits of eIF2B functions as ribose-1,5-bisphosphate isomerase in *Pyrococcus horikoshii* OT3. *Scientific Reports*, *8*(1), 1891.
- Gogoi, P., Srivastava, A., Jayaprakash, P., Jeyakanthan, J. and Kanaujia, S.P., 2016. In silico analysis suggests that PH0702 and PH0208 encode for methylthioribose-1-phosphate isomerase and ribose-1,5-bisphosphate isomerase, respectively, rather than aIF2Bβ and aIF2Bδ. *Gene*, *575*(1), 118-126.
- Gomez, E., Mohammad, S.S. and Pavitt, G.D., 2002. Characterization of the minimal catalytic domain within eIF2B: the guanine-nucleotide exchange factor for translation initiation. *The EMBO Journal*, *21*(19), 5292-5301.
- Goodsell, D.S. and Olson, A.J., 1990. Automated docking of substrates to proteins by simulated annealing. *Proteins: Structure, Function, and Bioinformatics*, *8*(3), 195-202.
- Gordiyenko, Y., Schmidt, C., Jennings, M.D., Matak-Vinkovic, D., Pavitt, G.D. and Robinson, C.V., 2014. eIF2B is a decameric guanine nucleotide exchange factor with a $\gamma_2\varepsilon_2$ tetrameric core. *Nature Communications*, *5*, 3902.
- Gordon, J.C., Myers, J.B., Folta, T., Shoja, V., Heath, L.S. and Onufriev, A., 2005. H++: a server for estimating p K as and adding missing hydrogens to macromolecules. *Nucleic Acids Research*, *33*(suppl_2), W368-W371.

- Gouet, P., Robert, X. and Courcelle, E., 2003. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Research*, *31*(13), 3320-3323.
- Grochowski, L.L., Xu, H. and White, R.H., 2005. Ribose-5-phosphate biosynthesis in Methanocaldococcus jannaschii occurs in the absence of a pentose-phosphate pathway. *Journal of Bacteriology*, *187*(21), 7382-7389.
- Gualerzi, C., Risuleo, G. and Pon, C.L., 1977. Initial rate kinetic analysis of the mechanism of initiation complex formation and the role of initiation factor IF-3. *Biochemistry*, *16*(8), 1684-1689.
- Gualerzi, C.O. and Pon, C.L., 1990. Initiation of mRNA translation in prokaryotes. *Biochemistry*, *29*(25), 5881-5889.
- Hasenöhrl, D., Benelli, D., Barbazza, A., Londei, P. and Blaesi, U., 2006. Sulfolobus solfataricus translation initiation factor 1 stimulates translation initiation complex formation. *Rna*, *12*(4), 674-682.
- Hasenöhrl, D., Fabbretti, A., Londei, P., Gualerzi, C.O. and Bläsi, U., 2009. Translation initiation complex formation in the crenarchaeon *Sulfolobus solfataricus*. *Rna*, *15*(12), 2288-2298.
- Powell, H.R., 2017. X-ray data processing. *Bioscience reports*, 37(5), SR20170227.
- Hess, B., 2008. P-LINCS: A parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation*, *4*(1), 116-122.
- Hinnebusch, A.G. and Lorsch, J.R., 2012. The mechanism of eukaryotic translation initiation: new insights and challenges. *Cold Spring Harbor Perspectives in Biology*, *4*(10), a011544.
- Hinnebusch, A.G., 1993. Gene-specific translational control of the yeast GCN4 gene by phosphorylation of eukaryotic initiation factor 2. *Molecular Microbiology*, *10*(2), 215-223.
- Hinnebusch, A.G., 2005. Translational regulation of GCN4 and the general amino acid control of yeast. *Annual Review of Microbiology* , *59*, 407-450.
- Hinnebusch, A.G., 2011. Molecular mechanism of scanning and start codon selection in eukaryotes. *Microbiology and Molecular Biology Reviews* , *75*(3), 434-467.
- Hinnebusch, A.G., 2017. Structural insights into the mechanism of scanning and start codon recognition in eukaryotic translation initiation. *Trends in Biochemical Sciences*, *42*(8), 589-611.

- Holm, L. and Rosenström 2010. Dali server: conservation mapping in 3D. *Nucleic Acids Research*, *38*(suppl_2), W545-W549.
- Hussain, A.S.Z., Shanthi, V., Sheik, S.S., Jeyakanthan, J., Selvarani, P. and Sekar, K., 2002. PDB Goodies–a web-based GUI to manipulate the Protein Data Bank file. *Acta Crystallographica Section D: Biological Crystallography*, *58*(8), 1385-1386.
- Hussain, T., Llácer, J.L., Fernández, I.S., Munoz, A., Martin-Marcos, P., Savva, C.G., Lorsch, J.R., Hinnebusch, A.G. and Ramakrishnan, V., 2014. Structural changes enable start codon recognition by the eukaryotic translation initiation complex. *Cell*, *159*(3), 597-607.
- Jackson, R.J., Hellen, C.U. and Pestova, T.V., 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*, *11*(2), 113.
- Jagtap, P.K.A., Verma, S.K., Vithani, N., Bais, V.S. and Prakash, B., 2013. Crystal structures identify an atypical two-metal-ion mechanism for uridyltransfer in GlmU: its significance to sugar nucleotidyl transferases. *Journal of Molecular Biology*, *425*(10), 1745-1759.
- Jao, S.C., English Ospina, S.M., Berdis, A.J., Starke, D.W., Post, C.B. and Mieyal, J.J., 2006. Computational and mutational analysis of human glutaredoxin (thioltransferase): probing the molecular basis of the low pKa of cysteine 22 and its role in catalysis. *Biochemistry*, *45*(15), 4785-4796.
- Jin, X., Ballicora, M.A., Preiss, J. and Geiger, J.H., 2005. Crystal structure of potato tuber ADP- glucose pyrophosphorylase. *The EMBO Journal*, *24*(4), 694-704.
- Kabsch, W. and Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules*, *22*(12), 2577-2637.
- Kakuta, Y., Tahara, M., Maetani, S., Yao, M., Tanaka, I. and Kimura, M., 2004. Crystal structure of the regulatory subunit of archaeal initiation factor 2B (aIF2B) from hyperthermophilic archaeon *Pyrococcus horikoshii* OT3: a proposed structure of the regulatory subcomplex of eukaryotic IF2B. *Biochemical and Biophysical Research communications*, *319*(3), 725-732.
- Kanaujia, S.P. and Sekar, K., 2009. Structural and functional role of water molecules in bovine pancreatic phospholipase A2: a data-mining approach. *Acta Crystallographica Section D: Biological Crystallography*, *65*(1), 74-84.

- Kapp, L.D. and Lorsch, J.R., 2004. The molecular mechanics of eukaryotic translation. *Annual Review of Biochemistry*, *73*(1), 657-704.
- Kashiwagi, K., Takahashi, M., Nishimoto, M., Hiyama, T.B., Higo, T., Umehara, T., Sakamoto, K., Ito, T. and Yokoyama, S., 2016. Crystal structure of eukaryotic translation initiation factor 2B. *Nature*, *531*(7592), 122.
- Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A. and Wang, J., 2015. PubChem substance and compound databases. *Nucleic Acids Research*, *44*(D1), D1202-D1213.
- Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B., Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L., Utterback, T., Cotton, M.D., Spriggs, T., Artiach, P., Kaine, B.P., Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C., Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O., Woese, C.R. and Venter J.C., 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature*, *390*(6658), 364-370.
- Kolitz, S.E., Takacs, J.E. and Lorsch, J.R., 2009. Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation. *Rna*, *15*(1), 138-152.
- Konagurthu, A.S., Whisstock, J.C., Stuckey, P.J. and Lesk, A.M., 2006. MUSTANG: a multiple structural alignment algorithm. *Proteins: Structure, Function, and Bioinformatics*, *64*(3), 559-574.
- Kortemme, T., Darby, N.J. and Creighton, T.E., 1996. Electrostatic interactions in the active site of the N-terminal thioredoxin-like domain of protein disulfide isomerase. *Biochemistry*, *35*(46), 14503-14511.
- Krishnamoorthy, T., Pavitt, G.D., Zhang, F., Dever, T.E. and Hinnebusch, A.G., 2001. Tight binding of the phosphorylated α subunit of initiation factor 2 (eIF2α) to the regulatory subunits of guanine nucleotide exchange factor eIF2B is required for inhibition of translation initiation. *Molecular and Cellular Biology*, *21*(15), 5018-5030.

- Krissinel, E. and Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology*, *372*(3), 774-797.
- Kuhle, B., Eulig, N.K. and Ficner, R., 2015. Architecture of the eIF2B regulatory subcomplex and its implications for the regulation of guanine nucleotide exchange on eIF2. *Nucleic Acids Research*, *43*(20), 9994-10014.
- Kumar, S., Stecher, G. and Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*(7), 1870-1874.
- Kyrpides, N.C. and Woese, C.R., 1998a. Universally conserved translation initiation factors. *Proceedings of the National Academy of Sciences*, *95*(1), 224-228.
- Kyrpides, N.C. and Woese, C.R., 1998b. Archaeal translation initiation revisited: the initiation factor 2 and eukaryotic initiation factor 2B α-β-δ subunit families. *Proceedings of the National Academy of Sciences*, *95*(7), 3726-3730.
- La Teana, A., Pon, C.L. and Gualerzi, C.O., 1996. Late events in translation initiation. Adjustment of fMet-tRNA in the ribosomal P-site. *Journal of Molecular Biology*, *256*(4), 667-675.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283-291.
- Lomakin, I.B. and Steitz, T.A., 2013. The initiation of mammalian protein synthesis and mRNA scanning mechanism. *Nature*, *500*(7462), 307.
- Lomakin, I.B., Kolupaeva, V.G., Marintchev, A., Wagner, G. and Pestova, T.V., 2003. Position of eukaryotic initiation factor eIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing. *Genes & Development*, *17*(22), 2786-2797.
- Lomakin, I.B., Shirokikh, N.E., Yusupov, M.M., Hellen, C.U. and Pestova, T.V., 2006. The fidelity of translation initiation: reciprocal activities of eIF1, IF3 and YciH. *The EMBO Journal*, *25*(1), 196-210.
- Londei, P., 2005. Evolution of translational initiation: new insights from the archaea. *FEMS Microbiology Reviews*, *29*(2), 185-200.
- Lorsch, J.R. and Dever, T.E., 2010. Molecular view of 43 S complex formation and start site selection in eukaryotic translation initiation. *Journal of Biological Chemistry*, *285*(28), 21203-21207.

- Lunt, S.Y. and Vander Heiden, M.G., 2011. Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. *Annual Review of Cell and Developmental Biology*, *27*, 441-464.
- Maag, D., Fekete, C.A., Gryczynski, Z. and Lorsch, J.R., 2005. A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Molecular Cell*, *17*(2), 265-275.
- Martin-Marcos, P., Cheung, Y.N. and Hinnebusch, A.G., 2011. Functional elements in initiation factors 1, 1A, and 2β discriminate against poor AUG context and non-AUG start codons. *Molecular and Cellular Biology*, *31*(23), 4814-4831.
- Martin-Marcos, P., Nanda, J., Luna, R.E., Wagner, G., Lorsch, J.R. and Hinnebusch, A.G., 2013. β-Hairpin loop of eukaryotic initiation factor 1 (eIF1) mediates 40 S ribosome binding to regulate initiator tRNAMet recruitment and accuracy of AUG selection in vivo. *Journal of Biological Chemistry*, *288*(38), 27546-27562.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J., 2007. Phaser crystallographic software. *Journal of applied crystallography*, *40*(4), 658-674.
- McPherson, A., 1999. *Crystallization of Biological Macromolecules* (Vol. 586). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- McPherson, A., 2004. Introduction to protein crystallization. *Methods*, *34*(3), 254-265.
- Minor, W., Cymborowski, M., Otwinowski, Z. and Chruszcz, M., 2006. HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallographica Section D: Biological Crystallography*, *62*(8), 859-866.
- Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S. and Olson, A.J., 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785-2791.
- Nakamura, A., Fujihashi, M., Aono, R., Sato, T., Nishiba, Y., Yoshida, S., Yano, A., Atomi, H., Imanaka, T. and Miki, K., 2012. Dynamic, ligand-dependent conformational change triggers reaction of ribose-1, 5-bisphosphate isomerase from *Thermococcus kodakarensis* KOD1. *Journal of Biological Chemistry*, *287*(25), 20784-20796.

- Nanda, J.S., Cheung, Y.N., Takacs, J.E., Martin-Marcos, P., Saini, A.K., Hinnebusch, A.G. and Lorsch, J.R., 2009. eIF1 controls multiple steps in start codon recognition during eukaryotic translation initiation. *Journal of Molecular Biology*, *394*(2), 268-285.
- Naor, M.M. and Jensen, J.H., 2004. Determinants of cysteine pKa values in creatine kinase and α1- antitrypsin. *Proteins: Structure, Function, and Bioinformatics*, *57*(4), 799-803.
- O'Donoghue, A.C., Amyes, T.L. and Richard, J.P., 2005a. Hydron transfer catalyzed by triosephosphate isomerase. Products of isomerization of dihydroxyacetone phosphate in D2O. *Biochemistry*, *44*(7), 2622-2631.
- O'Donoghue, A.C., Amyes, T.L. and Richard, J.P., 2005b. Hydron transfer catalyzed by triosephosphate isomerase. Products of isomerization of (R)-glyceraldehyde 3-phosphate in D2O. *Biochemistry*, *44*(7), 2610-2621.
- Orita, I., Sato, T., Yurimoto, H., Kato, N., Atomi, H., Imanaka, T. and Sakai, Y., 2006. The ribulose monophosphate pathway substitutes for the missing pentose phosphate pathway in the archaeon *Thermococcus kodakaraensis*. *Journal of Bacteriology*, *188*(13), 4698-4704.
- Osterman, I.A., Evfratov, S.A., Dzama, M.M., Pletnev, P.I., Kovalchuk, S.I., Butenko, I.O., Pobeguts, O.V., Golovina, A.Y., Govorun, V.M., Bogdanov, A.A. and Sergiev, P.V., 2015. A bacterial homolog YciH of eukaryotic translation initiation factor eIF1 regulates stress-related gene expression and is unlikely to be involved in translation initiation fidelity. *RNA Biology*, *12*(9), 966-971.
- Pain, V.M., 1996. Initiation of protein synthesis in eukaryotic cells. *European Journal of Biochemistry*, *236*(3), 747-771.
- Parrinello, M. and Rahman, A., 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, *52*(12), 7182-7190.
- Passmore, L.A., Schmeing, T.M., Maag, D., Applefield, D.J., Acker, M.G., Algire, M.A., Lorsch, J.R. and Ramakrishnan, V., 2007. The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Molecular Cell*, *26*(1), 41-50.
- Pavitt, G.D., 2005. eIF2B, a mediator of general and gene-specific translational control. *Biochemical Society Transactions*, *33*(6), 1487-1492.

- Pavitt, G.D., Ramaiah, K.V., Kimball, S.R. and Hinnebusch, A.G., 1998. eIF2 independently binds two distinct eIF2B subcomplexes that catalyze and regulate guanine–nucleotide exchange. *Genes & Development*, *12*(4), 514-526.
- Pedulla, N., Palermo, R., Hasenöhrl, D., Bläsi, U., Cammarano, P. and Londei, P., 2005. The archaeal eIF2 homologue: functional properties of an ancient translation initiation factor. *Nucleic Acids Research*, *33*(6), 1804-1812.
- Pei, J., Kim, B.H. and Grishin, N.V., 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Research*, *36*(7), 2295-2300.
- Pestova, T.V. and Kolupaeva, V.G., 2002. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes & Development*, *16*(22), 2906-2922.
- Pestova, T.V., Borukhov, S.I. and Hellen, C.U., 1998. Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature*, *394*(6696), 854.
- Pirkov, I., Norbeck, J., Gustafsson, L. and Albers, E., 2008. A complete inventory of all enzymes in the eukaryotic methionine salvage pathway. *The FEBS Journal*, *275*(16), 4111-4120.
- Ponder, J.W. and Case, D.A., 2003. Force fields for protein simulations. In *Advances in protein chemistry* (Vol. 66, 27-85). Academic Press.
- Price, N. and Proud, C., 1994. The guanine nucleotide-exchange factor, eIF-2B. *Biochimie*, *76*(8), 748-760.
- Proud, C.G., 1994. Peptide-chain elongation in eukaryotes. *Molecular Biology Reports*, *19*(3), 161-170.
- Rabl, J., Leibundgut, M., Ataide, S.F., Haag, A. and Ban, N., 2011. Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science*, *331*(6018), 730-736.
- Reid, P.J., Mohammad-Qureshi, S.S. and Pavitt, G.D., 2012. Identification of intersubunit domain interactions within eukaryotic initiation factor (eIF)2B, the nucleotide exchange factor for translation initiation. *Journal of Biological Chemistry*, *287*(11), 8275-8285.
- Rose, I.A. and O'Connell, E.L., 1960. Stereospecificity of the sugarphosphate isomerase reactions; a uniformity. *Biochimica et Biophysica Acta*, *42*, 159.
- Rose, I.A., 1975. Mechanism of the aldose-ketose isomerase reactions. *Advances in Enzymology and Related Areas of Molecular Biology*, *43*, 491-517.

- Rossmann, M.G. and Blow, D.M., 1962. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica*, *15*(1), 24-31.
- Rossmann, M.G., 1972. The locked rotation function. *Journal of Molecular Biology*, *64*(1), 246-249.
- Rupp, B., 2009. *Biomolecular crystallography: principles, practice, and application to structural biology*. Garland Science.
- 
- Saito, Y., Ashida, H., Kojima, C., Tamura, H., Matsumura, H., Kai, Y. and Yokota, A., 2007. Enzymatic characterization of 5-methylthioribose 1-phosphate isomerase from *Bacillus subtilis*. *Bioscience, Biotechnology, and Biochemistry*, 0707060478-0707060478.
- Sato, T., Atomi, H. and Imanaka, T., 2007. Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science*, *315*(5814), 1003-1006.
- Sekowska, A. and Danchin, A., 2002. The methionine salvage pathway in *Bacillus subtilis*. *BMC Microbiology*, *2*(1), 8.
- Sekowska, A., Dénervaud, V., Ashida, H., Michoud, K., Haas, D., Yokota, A. and Danchin, A., 2004. Bacterial variations on the methionine salvage pathway. *BMC Microbiology*, *4*(1), 9.
- Sievers, F. and Higgins, D.G., 2014. Clustal Omega, accurate alignment of very large numbers of sequences. In *Multiple sequence alignment methods* (105-116). Humana Press, Totowa, NJ.
- Soderberg, T., 2005. Biosynthesis of ribose-5-phosphate and erythrose-4-phosphate in archaea: a phylogenetic analysis of archaeal genomes. *Archaea*, *1*(5), 347-352.
- Stank, A., Kokh, D.B., Fuller, J.C. and Wade, R.C., 2016. Protein binding pocket dynamics. *Accounts of Chemical Research*, *49*(5), 809-815.
- Stanton, R.C., 2012. Glucose-6-phosphate dehydrogenase, NADPH, and cell survival. *IUBMB life*, *64*(5), 362-369.
- Stincone, A., Prigione, A., Cramer, T., Wamelink, M.M., Campbell, K., Cheung, E., Olin- Sandoval, V., Grüning, N.M., Krüger, A., Tauqeer Alam, M. and Keller, M.A., 2015. The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. *Biological Reviews*, *90*(3), 927-963.
- Sudhakar, A., Ramachandran, A., Ghosh, S., Hasnain, S.E., Kaufman, R.J. and Ramaiah, K.V., 2000. Phosphorylation of serine 51 in initiation factor 2α (eIF2α) promotes complex formation between eIF2α (P) and eIF2B and causes inhibition in

the guanine nucleotide exchange activity of eIF2B. *Biochemistry*, *39*(42), 12929-12938.

- Sumathi, K., Ananthalakshmi, P., Roshan, M.M. and Sekar, K., 2006. 3dSS: 3D structural superposition. *Nucleic Acids Research*, *34*(suppl_2), W128-W132.
- Tamura, H., Saito, Y., Ashida, H., Inoue, T., Kai, Y., Yokota, A. and Matsumura, H., 2008. Crystal structure of 5-methylthioribose 1-phosphate isomerase product complex from *Bacillus subtilis*: implications for catalytic mechanism. *Protein Science*, *17*(1), 126-135.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725-2729.
- The UniProt Consortium, 2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, *45*, D158–D169.
- Trun, N. and Trempy, J., 2003. Bacteriophage. *Fundamental Bacterial Genetics*, 105-126.
- Turner, P.J., 2005. XMGRACE, Version 5.1. 19. *Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR*.
- Unbehaun, A., Borukhov, S.I., Hellen, C.U. and Pestova, T.V., 2004. Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes & Development*, *18*(24), 3078-3093.
- UniProt Consortium, 2018. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506-D515.
- Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F. and Murshudov, G.N., 2004. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallographica Section D: Biological Crystallography*, *60*(12), 2184-2195.
- Wang, X., Wortham, N.C., Liu, R. and Proud, C.G., 2012. Identification of residues that underpin interactions within the eukaryotic initiation factor (eIF2) 2B complex. *Journal of Biological Chemistry*, *287*(11), 8263-8274.

- Wedekind, J.E., Frey, P.A. and Rayment, I., 1995. Three-Dimensional Structure of Galactose-1-phosphate Uridylyltransferase from *Escherichia coli* at 1.8. ANG. Resolution. *Biochemistry*, *34*(35), 11049-11061.
- Weisser, M., Voigts-Hoffmann, F., Rabl, J., Leibundgut, M. and Ban, N., 2013. The crystal structure of the eukaryotic 40S ribosomal subunit in complex with eIF1 and eIF1A. *Nature Structural & Molecular Biology*, *20*(8), 1015.
- Wek, R.C., Jiang, H.Y. and Anthony, T.G., 2006. Coping with stress: eIF2 kinases and translational control. *Biochemical Society Transactions*, *34*(1), 7-11
- Winans, S.C. and Bassler, B.L., 2002. Mob psychology. *Journal of Bacteriology*, *184*(4), 873-883.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A. and McNicholas, S.J., 2011. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*, *67*(4), 235-242.
- Woese, C.R. and Fox, G.E., 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences*, *74*(11), 5088-5090.
- Wortham, N.C., Martinez, M., Gordiyenko, Y., Robinson, C.V. and Proud, C.G., 2014. Analysis of the subunit organization of the eIF2B complex reveals new insights into its structure and regulation. *The FASEB Journal*, *28*(5), 2225-2237.
- Yang, W. and Hinnebusch, A.G., 1996. Identification of a regulatory subcomplex in the guanine nucleotide exchange factor eIF2B that mediates inhibition by phosphorylated eIF2. *Molecular and Cellular Biology*, *16*(11), 6603-6616.
- Yusupova, G.Z., Yusupov, M.M., Cate, J.H.D. and Noller, H.F., 2001. The path of messenger RNA through the ribosome. *Cell*, *106*(2), 233-241.
- Zhang, Y. and Skolnick, J., 2005. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, *33*(7), 2302-2309.
- Zhouravleva, G., Frolova, L., Le Goff, X., Le Guellec, R., Inge-Vechtomov, S., Kisselev, L. and Philippe, M., 1995. Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. *The EMBO Journal*, *14*(16), 4065-4072.
- Zhu, J. and Weng, Z., 2005. FAST: a novel protein structure alignment algorithm. *PROTEINS: Structure, Function, and Bioinformatics*, *58*(3), 618-627.