Protein-RNA recognition: Insight from molecular dynamics simulations

A Thesis

Submitted in partial Fullfilment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

by

AMIT KUMAR

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Guwahati-781039, Assam, India

Protein-RNA recognition: Insight from molecular dynamics simulations

A Thesis

Submitted in partial Fullfilment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

by

AMIT KUMAR

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Guwahati-781039, Assam, India

Indian Institute of Technology Guwahati

Department of Biosciences and Bioengineering

DECLARATION

I do hereby declare that the thesis entitled "**Protein-RNA recognition: Insight from molecular dynamics simulations**", submitted by me to the Indian Institute of Technology Guwahati, for the award of the Doctor of Philosophy, is a record of original research carried out by me under the supervision of **Dr. Priyadarshi Satpati**. I declare that this work contains no material that has been accepted for the award of any other degree or diploma in any other University or Institute. I also certify that to the best of my knowledge and understanding nothing in this report amounts to plagiarism.

Date:

Amit kumar

(146106013)

Indian Institute of Technology Guwahati

Department of Biosciences and Bioengineering

CERTIFICATE

This is to certify that the thesis entitled "**Protein-RNA recognition: Insight from molecular dynamics simulations**", submitted by Amit Kumar (Roll No. 146106013), a PhD student in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, for the award of the Doctor of Philosophy, is a record of an authentic research work carried out by him under my supervision and guidance. The thesis has fulfilled all requirements as per the regulations of the institute and in my opinion, has reached the standard needed for submission. The results embodied in this thesis have not been submitted elsewhere for the award of any degree or diploma.

Supervisor: Dr. Priyadarshi Satpati
Department of Biosciences and Bioengineering
Indian Institute of Technology, Guwahati
Guwahati-781039, Assam, INDIA

Date:

ACKNOWLEDGEMENT

First and foremost, I want to thank my PhD supervisor, Dr. Priyadarshi Satpati, for believing in me and providing this opportunity. I would like to express my sincere gratitude and appreciation to him. He has taught me, both consciously and unconsciously, the importance of hard work and patience. I appreciate all his contributions of time, ideas, and suggestions to make my research experience productive and stimulating. I could not have imagined having a better mentor and supervisor for my PhD study.

In addition to my advisor, I would like to thank the rest of my Doctoral committee members: Dr. Shankar Prasad Kanaujia (Chairman), Dr. Shirisha Nagotu, and Dr. Katha Anki Reddy, for their valuable comments, suggestions, and encouragement for shaping my thesis work.

I would like to acknowledge Professor Latha Rangan, Head of the Department of Biosciences and Bioengineering (BSBE), IIT Guwahati and former Head of the Department Professor K. Pakshirajan, and Professor V. V. Dasu for providing me the necessary facilities, which helped me to perform my research work at IIT Guwahati.

I am also grateful to all faculty members and staff of BSBE for their help and support. I would like to acknowledge Bioinformatics Infrastructure Facility (BIF), IIT Guwahati for computing facility. I am also highly grateful to the PARAM ISHAN supercomputing facility, IIT Guwahtai for providing resources to execute my experiment. I am grateful to the Ministry of Human Resource Development (MHRD) and IIT Guwahati for providing research fellowship.

My time at Guwahati was made enjoyable in large part due to the many friends and groups that became a part of my life. I am grateful for their support and love through all my good and bad times. Adarsh, Anil, Satendra, Vimal, Neha, Dharitri, Purnima, and Bhaswati hold special places in my life for being friends-cum-family.

I would also like to acknowledge Dr. Sunanda Chatterjee and the December Trip group for all the great memories.

I thank my fellow labmates: Abhishek, Suvankar, Gaurav, Saransh, Debadrita, Vignesh and Tiasha for the stimulating discussions and for all the fun we have had throughout my PhD tenure.

Last but not least; I would like to thank my family: my mom, dad, brother and sister for almost unbelievable support and their trust in me throughout this journey and my life in general. They are the most important people in my world and I dedicate this thesis to them. Thanks to the Almighty for showering his grace to complete the research work successfully.

Amit Kumar

Synopsis

Protein-RNA recognition plays a major role in biology, viz., charging tRNA by synthetase, structural stability of ribosome, stop-codon recognition by translation factors in the ribosome, viral-RNA recognition by innate immune proteins, etc. The objective of this thesis was to explore the application of classical molecular dynamics free energy calculations to understand protein-RNA recognition events involved in gene expression and viral-RNA recognition by host protein. Energetics of discrimination (i.e., free energy differences between correct and incorrect protein-RNA complex) is linked to molecular recognition. Using X-ray structures of protein-RNA complexes as a template, we performed molecular dynamics simulations, and using appropriate thermodynamic cycle, we computed the strength of discrimination (cognate from non-cognate) employing popular statistical methods (e.g., Thermodynamic integration, Free energy perturbation, etc.). In this thesis, protein-RNA interactions were studied for three cases: (1) Alanine tRNA recognition by alanine-tRNA synthetase (AlaRS), (2) Stop-codon recognition by eukaryotic release factor 1 (eRF1), and (3) Viral RNA recognition by retinoic inducible gene I (RIG-I).

The entire thesis has been divided into **5 chapters**. **Chapter 1** deals with a brief review of the protein-RNA complexes; and the general methodology for molecular dynamics simulations, and for estimating free energies. **Chapter 2** discusses the principle of alanine tRNA selection by alanine tRNA synthetase based on the critical G3·U70 base pair. Presence of a single G3•U70 mismatch in the acceptor stem of $tRNA^{Ala}$ is essential for aminoacylation with alanine by alanyl-tRNA synthetase (AlaRS) in archaea, bacteria, and eukarya. In this chapter, we discussed quantitative estimation of relative binding free energies associated with AlaRS binding to 3•70 mutant $tRNA^{Ala}$ with respect wild-type $tRNA^{Ala}$/ G3•U70. Based on our computed relative binding free energetics, we have proposed a simple three-state kinetic scheme for aminoacylation which could explain the experimentally measured kinetics. The quantitative estimation of $tRNA^{Ala}$ selectivity by AlaRS offers a simple view of how accuracy in the aminoacylation process is achieved, thereby establishing the link between 3D structures, thermodynamics, and experimentally measured kinetics.

In **Chapter 3**, energetics of stop codon (UAA, UGA, and UAG) recognition by eukaryotic release factor 1 (eRF1) have been reported. Results suggest that eRF1 imposes a very high energetic penalty for binding to sense codon (CAA, UGG) programmed mRNA in the ribosomal A-site. The discriminatory power of eRF1 (favouring stop-codons relative to sense codons) is larger than its bacterial analog (RF1 and RF2). eRF1 ensures high selectivity by (a) introducing strain in the mRNA, (b) disrupting protein-mRNA interactions, and (c) placing sense-codons in the dry desolvated pocket with unsatisfied h-bonds in the near-cognate complexes. This chapter provides a clue to how eRF1 selectivity between the stop and sense codons could control error during translation termination.

In **Chapter 4**, viral RNA recognition by RIG-I has been reported. RIG-I protein recognizes viral RNAs and initiates an antiviral response, but unresponsive to similar host RNA's. Viral RNA's are usually double-stranded bearing 5'-tri/diphosphates (5'-ppp/pp-dsRNA), differing from host RNA's (bearing 5'-monophosphate/hydroxyl: 5'-P/OH-dsRNA). Experiments confirmed that RIG-I binds to viral RNA's with a very high affinity relative to host RNA's. Interestingly, 5'-p-dsRNA binds to RIG-I with surprisingly low affinity, and binding affinity is much weaker than 5'-OH-dsRNA. The mechanism of RNA discrimination by RIG-I is unclear and poorly understood. This chapter consists of two parts. The first part discusses viral (5'-ppp/pp-dsRNA) and host (5'-p) dsRNA discrimination by RIG-I. The second part discusses RIG-I discrimination between host RNA's (5'-p-dsRNA vs. 5'-OH-dsRNA). Based on our calculations, we provide a possible clue of how a simple counter ion dissociation from the 5' terminal of dsRNA during RIG-I binding could be used for efficient recognition.

In **Chapter 5** and the final section of this thesis, the overall conclusion from this thesis work is presented and the future scope of this current thesis work is discussed. We have shown that it is possible to bridge the microscopic structures of biomolecules and free energy differences. We believe that the methodology adopted in this thesis could be useful for studying other protein-RNA recognition events in general.

Contents

List of Figures xiii

List of Tables xvii

Abbreviations xviii

Chapter 1

Literature Survey and Methodology 1

1.1 Introduction 1

1.2 Alanine tRNA charging by Alanine-tRNA Synthetase 2

1.2.1 Background 2

1.2.2 Aminoacylation Reaction 2

1.2.3 tRNA selection by Aminoacyl-tRNA Synthetase 4

1.2.4 Structural and Biological Studies Related to $tRNA^{Ala}$ Recognition by AlaRS 5

1.3 Stop Codon Recognition by Eukaryotic Release Factor 1 8

1.3.1 Background 8

1.3.2 Translation Termination 9

1.3.3 Structural and Biochemical Studies of eRF1 10

1.4 Viral RNA Recognition by RIG-I 15

1.4.1 Background 15

1.4.2 RIG-I and its Different Domains 16

1.4.3 Features of RNA Recognized by RIG-I 16

1.4.4 Current View of Antiviral Response Triggered by RIG-I 17

1.4.5 Structural and Biochemical Studies Related to Viral RNA Recognition by RIG-I 18

1.5 Methods 21

1.5.1 Brief History of Molecular Dynamics and Free energy Simulations 21

1.5.2 Theory of Molecular Dynamics Simulations 22

1.5.3 Force Fields 25

1.5.4 Water Models 27
1.5.5 Boundary Conditions 27
1.5.6 Long-range Electrostatic Interaction 28
1.5.7 Short-range Van der Waals Interactions 29
1.5.8 Model of Biomolecules 30
1.5.9 Energy Minimization 31
1.5.10 Protocol of Molecular Dynamics Simulation of Biomolecules 32
1.5.11 Thermodynamic Cycle and Relative Binding 34
1.5.12 Relative Free Energy Calculations 35
1.5.12.1 Free Energy Perturbation (FEP) 36
1.5.12.2 Thermodynamic Integration (TI) 36
1.5.13 Software Used in This Thesis 37
1.6 Objectives of the Thesis 37

Chapter 2

Alanine tRNA Recognition by Alanine tRNA Synthetase 38
2.1 Background 39
2.2 Methods 41
2.2.1 Molecular Dynamics Procedure 41
2.2.2 Free Enenrgy Calculations 44
2.3 Results 47
2.3.1 Structure-based Energetics of $tRNA^{Ala}$ Selectivity in AlaRS 47
2.3.2 X-ray vs MD Structures of AlaRS-$tRNA^{Ala}$ Complexes 49
2.3.3 MD Structures of Non-cognate AlaRS-$tRNA^{Ala}$ Complexes 53
2.3.4 Structure-based Energetics and its Connection to Kinetics 55
2.4 Discussion 58
2.5 Conclusion 61

Chapter 3

Stop Codon Recognition by Eukaryotic Release Factor 1 62
3.1 Background 63
3.2 Methods 65
3.2.1 Molecular Dynamics Setup 65
3.2.2 Free Energy Calculations 68
3.3 Results 70
3.3.1 Structure-based Energetics for Stop Codon Recognition 70
3.3.2 Insights from Average MD Structures 71
3.3.3 Structural Insights for First Position Discrimination 74
3.3.4 Structural Insights for Second and Third Positions Discrimination 77
3.4 Discussion 80
3.5 Conclusion 82
Chapter 4
Viral RNA Recognition by Retinoic Acid-inducible Gene-I 83
4.1 Energetics of Preferential Binding of RIG-I to Double-stranded Viral RNAs with 5’ tri/diphosphate Over 5’ monophosphate 84
4.1.1 Background 85
4.1.2 Methods 88
4.1.2.1 Molecular Dynamics Setup 88
4.1.2.2 Protocol for Binding Free Energy Calculation 90
4.1.3 Results 92
4.1.3.1 Biochemical and Structural Views of RIG-I Binding to dsRNA 92
4.1.3.2 dsRNA with 5’ ppp, 5’ pp, and 5’ p in solution 95
4.1.3.3 Structure-based Enrgetics of dsRNA Binding to RIG-I 98
4.1.3.4 Comparison Between MD and X-ray Structures 99
4.1.3.5 Structure and Dynamics of RIG-I: 5’ ppp-dsRNA Complex 100
4.1.3.6 Structure and Dynamics of RIG-I: 5’ pp-dsRNA Complex 103

4.1.3.7 Structure and Dynamics of RIG-I: 5’ p-dsRNA Complex 104
4.1.4 Discussion 106
4.1.5 Conclusion 109
4.2 Why Double-stranded RNA with 5’ Monophosphate is a Poor Binder to RIG-I with Respect to 5’ Hydroxyl Analogue? 110
4.2.1 Background 111
4.2.2 Methods 114
4.2.2.1 Molecular Dynamics Setup 114
4.2.2.2 Alchemical Transformation and Relative Binding Free Energies 116
4.2.3 Results 118
4.2.3.1 Energetics of 5’ p-dsRNA vs 5’ OH-dsRNA Binding to RIG-I 118
4.2.3.2 Structure and Dynamics of RIG-I: 5’ OH-dsRNA Complex 122
4.2.3.3 Insight into RIG-I: 5’ p-dsRNA from MD 124
4.2.4 Discussion 128
4.2.5 Conclusion 129

Chapter 5
Concluding Remarks and Future Perspective 131

Appendices
Appendix 2A 133
Appendix 3A 138
Appendix 4A 140
Appendix 5 147

References 151
List of Publications 168

List of Figures

Figure 1.1	Mechanism of aminoacylation reaction.	3
Figure 1.2	The secondary and tertiary structure of *A. fulgidus* $tRNA^{Ala}$ molecule.	4
Figure 1.3	Three-dimensional structure of AlaRS dimer in complex with $tRNA^{Ala}$.	5
Figure 1.4	Differences between structures of A. fulgidus AlaRS in complex with $tRNA^{Ala}$ /G•U and $tRNA^{Ala}$ /A•U.	6
Figure 1.5	Schematic representation of translation termination process.	9
Figure 1.6	eRF1 and P-site tRNA bound to mRNA in the ribosome.	11
Figure 1.7	Zoomed in view of the ribosomal A-site of the eukaryotic translation termination complex .	12
Figure 1.8	Interaction network involving stop-codon, eRF1, and ribosome in the translation termination complex.	13
Figure 1.9	A color-coded schematic representation of the domain architecture of RIG-I primary structure.	16
Figure 1.10	Secondary and tertiary structure of panhandle-like RNA molecule.	17
Figure 1.11	Schematic representing the stages of RIG-I activation.	18
Figure 1.12	(a) RIG-I:5’ ppp-dsRNA bound complex (PDB 3LRR). (b) RIG-I: 5’ ppp-dsRNA bound complex (PDB 4AY2). (c) RIG-I: 5’ pp-dsRNA bound complex (PDB 3NCU).	20
Figure 1.13	Schematic representation of bond, angle, dihedral, and non-bonded terms.	26

Figure 1.14	Schematic of periodic boundary conditions in two-dimension.	28
Figure 1.15	Spherically truncated biomolecular model (from PDB: 3WQY) considered in our simulation.	30
Figure 1.16	Thermodynamic cycle used to calculate relative binding free energies of ligand-receptor binding.	35
Figure 2.1	AlaRS-tRNA^{Ala} complex, identity base pair and its variants.	40
Figure 2.2	MD setup.	42
Figure 2.3	Thermodynamic cycle for AlaRS:tRNA^{Ala} binding.	45
Figure 2.4	MD structure of reactive AlaRS:tRNA^{Ala}/G3•U70 complex and after reverse alchemical transformation.	46
Figure 2.5	Calculated energetics of tRNA^{Ala} reading in different states.	48
Figure 2.6	Thermodynamic cycle for alchemical transformation of G3•U70 into A3•U70 in free tRNA in water.	49
Figure 2.7	Structural insights from X-ray and MD simulations of AlaRS-tRNA^{Ala} complex.	50
Figure 2.8	Root-mean-square deviation (RMSD) of the heavy atoms of AlaRS.tRNA complexes G3•U70 and A3•U70.	51
Figure 2.9	Root-mean-square fluctuation of the heavy atoms of the loop region is highligted with rectangular box.	52
Figure 2.10	Loop comparison between X-ray (black) and MD structures of AlaRS.tRNA complex with respect to G3•U70 and A3•U70.	53
Figure 2.11	MD structures of reactive and non-reactive complexes with tRNA containing G3•C70 and A3•C70 base pairs.	54
Figure 2.12	Schematic free energy diagram for tRNA^{Ala} selection by AlaRS.	56

Figure 2.13	X-ray structure of the reactive and non-reactive complex.	60
Figure 3.1	Stop codon recognition by eRF1.	63
Figure 3.2	Watson-Crick (WC) and Hoogsteen edge of adenosine.	66
Figure 3.3	Thermodynamic cycle for Stop codon:eRF1 binding on ribosome.	69
Figure 3.4	Computed binding free energy difference of eRF1 upon stop codon mutation.	71
Figure 3.5	Interaction pattern between stop codon (UAA) and Release factor in eukaryotes (left) and bacteria in the ribosomal A Site.	72
Figure 3.6	Zoom into ribosomal A site.	73
Figure 3.7	Water density around the N9 of G3.	74
Figure 3.8	First position specificity, UAA vs CAA.	75
Figure 3.9	Distance between the heavy atoms of U1 and Lys63/Asn61 along the 5 ns MD trajectory.	76
Figure 3.10	MD structure of eRF1 free four nucleotide CAA(G) codon in the ribosomal A site.	77
Figure 3.11	Second and Third position specificity.	78
Figure 3.12	Different interaction network UAA, UGA and UGG in the termination complex.	79
Figure 4.1	Structural overview of blunt-ended 5’ ppp-dsRNA bound RIG-I (containing helicase and C-terminal domains) complex.	86
Figure 4.2	Root-mean-square deviation of heavy atoms and flexible loop.	89
Figure 4.3	Thermodynamic cycle (TC) for dsRNA: RIG-I binding.	91
Figure 4.4	X-ray structure of RIG-I:5’ ppp-dsRNA complexes.	93
Figure 4.5	Thermodynamic cycle for Mg^{2+} : dsRNA binding.	96
Figure 4.6	MD structures highlighting the interaction between Mg^{2+} and 5’ ppp/pp/p-dsRNA in water.	98
Figure 4.7	Binding pocket of 5' ppp-dsRNA: RIG-I complex: MD snapshot (cyan) compared to the crystal structure.	100
Figure 4.8	RIG-I:dsRNA complexes from MD: Flexibility of the loop region.	102
Figure 4.9	Water density around the terminal phosphates.	103
Figure 4.10	Binding pocket of 5' pp-dsRNA: RIG-I complex: MD snapshot	104

(cyan) compared to the crystal structure.

Figure 4.11 MD snapshot of 5' p-dsRNA: RIG-I complex binding pocket. 105

Figure 4.12 Binding pocket: MD snapshot (cyan) vs X-ray structure. 108

Figure 4.13 Structural overview of RIG-I:dsRNA complexes. 112

Figure 4.14 Root-mean-square deviation of heavy atoms within 27Å of unrestrained simulation sphere with respect to x-ray structure. 115

Figure 4.15 Thermodynamic cycle used to estimate relative binding free energies between 5’ OH-dsRNA and 5’ p-dsRNA to RIG-I. 117

Figure 4.16 Calculated relative binding free energies (in kcal/mol) in the active, inactive complex (blue bars). 119

Figure 4.17 Thermodynamic cycle for Mg^{2+} :5’ p-dsRNA binding. 120

Figure 4.18 Structural comparison around the 5’-terminus of dsRNA in complex with RIG-I. 122

Figure 4.19 Zoomed in view of 5’ OH binding pocket from MD simulations: (a) Inactive complex (b) active complex. 123

Figure 4.20 Zoomed in view of MD structures of 5’ p binding pocket: (a) Inactive complex (b) active complex. 124

Figure 4.21 (a) X-ray structure of “NA” conformation of RIG-I:5’ OH-dsRNA complex (PDB code: 5F9F). 125

Figure 4.22 Root-mean-square fluctuation (RMSF) of the Ca atom with respect to x-ray structure. 126

Figure 4.23 Key residues around the 5’ OH terminal of RIG-I: 5’ OH-dsRNA complex in the x-ray structure. 127

List of Tables

Table 2.1	Partial charges of RNA phosphates.	44
Table 2.2	Numerical analysis of aminoacylation without proof-reading.	58
Table 3.1	Computed free energy change for alchemical transformation of two different simulation models.	67
Table 4.1	Structural Comparison of free RIG-I and dsRNA bound complexes.	94
Table 4.2	Relative binding free energies of Mg 2+ binding to dsRNA (with 5’ ppp/5’ pp/5’ p).	97
Table 4.3	Relative binding free energies: 5' ppp-dsRNA/5' pp-dsRNA/ 5' p-dsRNA binding to RIG-I.	99
Table 4.4	Selected interatomic distances averaged over the MD trajectories.	101
Table 4.5	Relative binding free energies of Mg 2+ binding to dsRNA (with 5’ p /5’ OH).	120
Table 4.6	MD averaged selected interatomic distances (in angstrom) and standard deviations (after ±).	123

Abbreviations

A	Active
aaRS	Aminoacyl tRNA Synthetase
ABNR	Adopted basis Newton-Raphson
Ala-SA	Alanine Sulfamoyl Adenosine
CARDs	Caspase activation and recruitment domains
CG	Conjugate Gradient
CTD	C-terminal domain
DAMP	Danger Associated Molecular Pattern
dsRNA	Double-stranded RNA
eRF	Eukaryotic Release Factor
eEF	Eukaryotic Elongation Factor
FEP	Free Energy Perturbation
FWD	Forward
HEL1	Helicase1
HEL2	Helicase2
HEL2i	Helicase2 insertion
IF	Initiation Factor
LGP2	Laboratory of Genetics and Physiology 2
LysRS	Lysl tRNA Synthetase
MDA5	Melanoma Differentation Associated 5
MD	Molecular dynamics
MDFE	Molecular Dynamics Free Energy
mRNA	Messenger RNA
NA	Inactive
NR	Non-reactive
P	Pincer
PAMP	Pathogen Associated Molecular Pattern
PBC	Periodic Boundary Condition
PDB	Protein data bank
PME	Particle mesh Ewald
PRR	Pattern Recognition Receptor

R	Reactive
Rev	Reverse
Rg	Radius of Gyration
RIG-I	Retinoic acid-inducible gene-I
RLR	RIG-I Like Receptors
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
SASA	Solvent Accessible Surface Area
SD	Steepest descent
TC	Themodynamic Cycle
TI	Thermodynamic Integration
TLR	Toll Like Receptors
tRNA	Transfer RNA
$tRNA^{Ala}$	Alanine tRNA
vdw	van der Waals
WC	Watson-Crick

Chapter 1

Literature Survey and Methodology

1.1 Introduction

Cellular functions often rely on biomolecular recognition involving protein and RNA molecules. Accuracy of gene expression (viz, Transcription by RNA polymerase, translation termination by release factors, amino-acid charging of tRNA's, DNA damage repair etc.) (**Whitelaw & Proudfoot, 1986, Wan et al., 2007, Hawley et al., 2017**), structural stability of ribosome, recognition of foreign pathogenic genetic material (**Dyer et al., 2008**), etc. are few examples that rely on protein-RNA recognition. Strength of Protein-RNA recognition relies on its ability to discriminate right (cognate) from wrong (non-cognate) interactions. The discriminatory strength is linked with the frequency of errors and is thus directly related to accuracy in key cellular processes (viz., transcription, tRNA charging, translation, host defense, etc).

During the last 2-3 decades, enormous experimental structural investigations (X-ray, cryo-EM, NMR, etc) have been carried out to understand the 3D structures of protein-RNA complexes in atomic details. Whereas biochemical studies have provided the details of kinetics and thermodynamics associated with the function of those biomolecules. But, the link between the structural and biochemical studies is lacking. Molecular dynamics (MD) free energy simulations (**Lind et al., 2019**) of protein-RNA complexes are suitable for bridging the gap to some extent.

In this thesis, we performed MD simulations of protein-RNA complexes (using experimentally determined structures as a template), in an attempt to stitch thermodynamics, kinetics and, 3D-structures. The aim was to first estimate the strength of protein-RNA recognition by combining molecular dynamics and statistical mechanics methods, then relate the discriminatory power to the 3D structures of cognate and non-cognate protein-RNA complexes, and finally, propose a kinetic scheme that could explain experimentally measured kinetics.

We focused on three different cases where protein-RNA interactions were known to essential cellular function : (1) Alanine tRNA recognition by alanine-tRNA synthetase, (2) Stop-codon recognition by eukaryotic release factor 1, and (3) Viral RNA recognition by retinoic inducible gene I (RIG-I).

1.2 Alanine tRNA Charging by Alanine-tRNA Synthetase

1.2.1 Background

Translation is the process whereby messenger RNA (mRNA) or genetic information is decoded in the ribosome decoding center to synthesize a specific polypeptide. Ribosome contains mRNA and transfer RNA (tRNA) binding sites in the large ribosomal subunit. tRNA serves as an adaptor between the mRNA and amino acid sequences of proteins. The ribosome promotes the binding of mRNA and the complementary tRNA carrying a specific amino acid. The accuracy of protein synthesis depends on the codon-anticodon recognition and tRNA aminoacylation by aminoacyl-tRNA synthetase (aaRS) enzymes. These enzymes (aaRS) catalyze the covalent attachment of an amino acid to the adenosine in the conserved 3'-terminal CCA sequence of its cognate tRNA (**Delagoutte et al., 2000**). The biological importance of correct aminoacylation for the accurate translation of mRNA was first established by the Chapeville group in 1962 (**Chapeville et al., 1962**).

1.2.2 Aminoacylation Reaction

Aminoacyl-tRNA synthetases (aaRS) are protein enzymes that attach amino acids to the tRNA. The diverse set of these enzymes in each cell is united by a common two-step aminoacylation reaction (**Figure 1**) in which amino acids are ligated to the 3' end of their cognate tRNA (**Szymanski & Barciszewski, 1999**). In the first step of the reaction, the aaRS utilizes an ATP to form an aminoacyl-adenylate (AMP-amino acid) bound aaRS complex. In the second step, the activated amino acid is transferred from aminoacyl-adenylate (AMP-amino acid) to either the 2'- or 3'-hydroxyl group of the ribose at the 3' end of tRNA forming an aminoacyl ester bond, releasing AMP (**Szymanski & Barciszewski, 1999**). Amino acid bound tRNA is known as aminoacyl-tRNA or charged tRNA.

These charged tRNAs are then used as building blocks in the translation process for protein synthesis. For each standard amino acid, there is one aaRS residing in the cell. Thus, the job of the tRNA synthetase is two-fold: (1) selection of correct tRNA from a pool of tRNA's (2) selection of correct amino acid for charging. Aminoacyl-tRNA synthetases fall into either class I or class II, divided based on their structures and functions. Lysyl-tRNA synthetase (LysRS) is an exception that represents both classes. The catalytic domain of Class I synthetases contain characteristic Rossman fold and are monomeric (**Sugiura et al., 2000**). Class II aminoacyl-tRNA synthetases are mostly dimeric or multimeric containing anti-parallel beta-sheet fold flanked by alpha-helices (**Perona et al., 1993**). Occasional errors of amino acid mischarging or misactivation occur because of amino acid structural similarity resulting in mistranslation. Over time, aminoacyl tRNA synthetases have developed an editing/proof-reading function that enables these enzymes to remove mischarged amino acids and prevent mistranslation. In some cases, editing function increases the amino acid selectivity by 100 to 1000-fold, but the reaction involved in editing may differ for different aaRS enzymes (**Preorna & Gruic-Sovulj, 2014**).

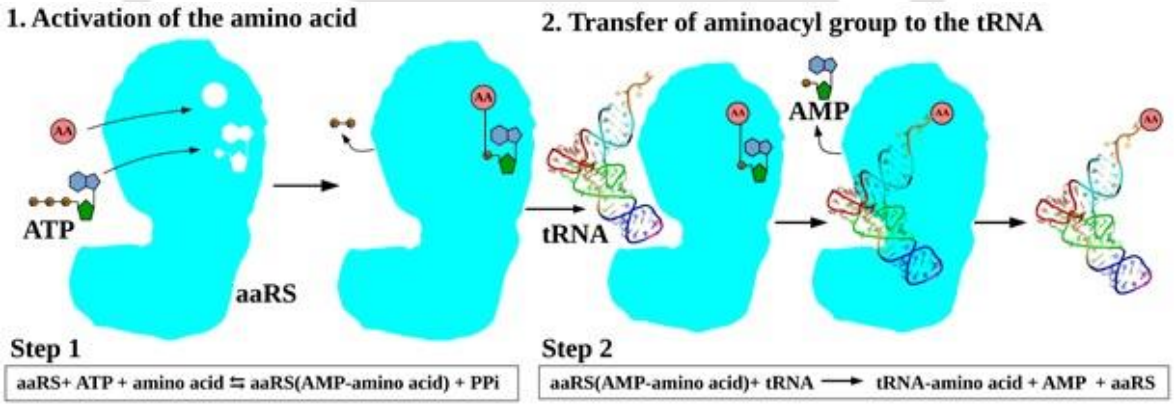


Figure 1.1: Mechanism of aminoacylation reaction. Step1: The aminoacyl tRNA synthetase (aaRS) first binds ATP and the corresponding amino acid (AA) to form an aminoacyl-adenylate complex, releasing inorganic pyrophosphate (PPi). Step2: The aminoacyl-adenylate complex binds the appropriate tRNA molecule, and then the amino acid is transferred to CCA end by the release of AMP.

1.2.3 tRNA Selection by Aminoacyl-tRNA Synthetase

Generally, tRNAs are 70 to 80 nucleotides long and have similar secondary and tertiary structures. Thus, the aaRS differentiates cognate tRNA from the pool of tRNAs in the cell for charging. aaRS recognizes unique identity elements present within the tRNA and ensure cognate tRNA binding. The tRNA structure consists of five main domains (**Figure 1.2**) known as the acceptor stem, TψC stem-loop, variable arm, anticodon stem-loop, and D stem-loop (**Sprinzl et al., 1998**). The tRNA identity is determined by a particular set of structural features called identity set, which consists of a limited number of nucleotides such as anticodon (**Giege et al., 1998**). Recognition of the appropriate tRNA by the aaRS is not mediated solely by the anticodon, but the variable pocket and acceptor stem also often play a prominent role (**Schimmel et al., 1993**).

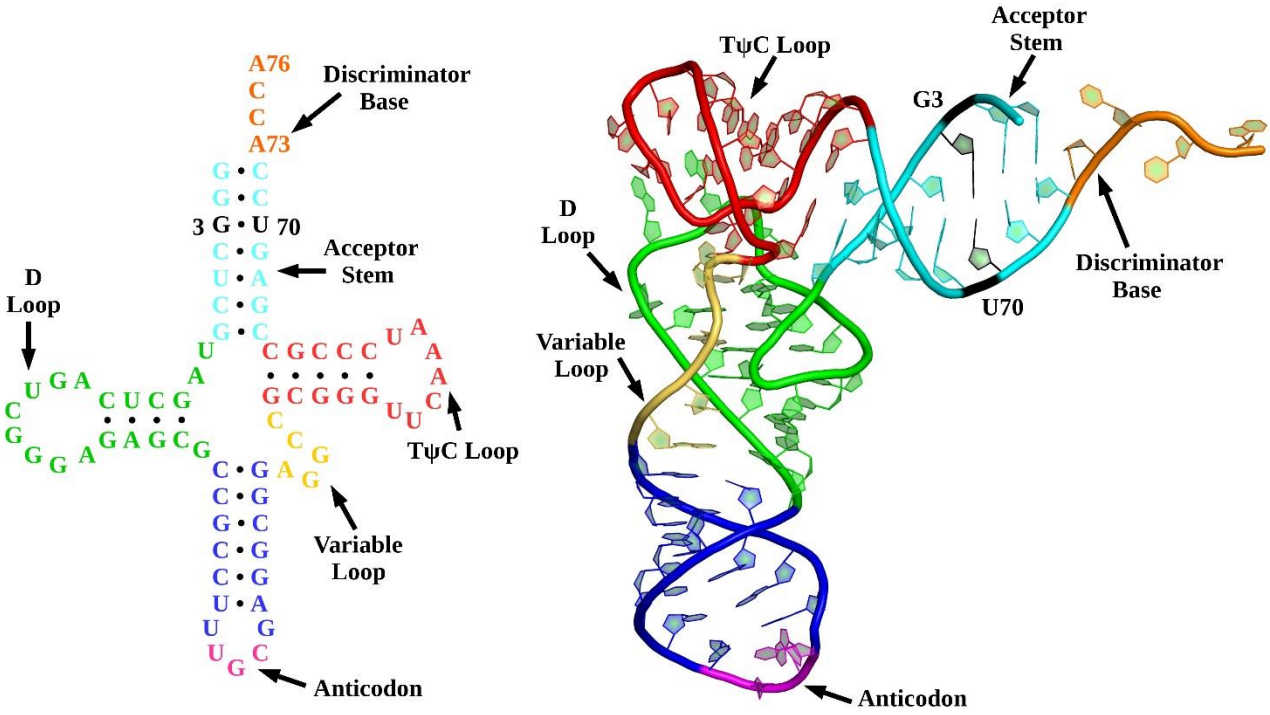


Figure 1.2: The secondary and tertiary structure of *A. fulgidus* $tRNA^{Ala}$ molecule. Sequence and secondary structure cloverleaf model (left) and tertiary structure (right) (PDB:3WQY). The particular features anticodon triplet (magenta), variable loop (yellow), D loop (green), TψC loop (red), acceptor stem (cyan), 3' acceptor stem (orange), and wobble base pair G3•U70 (black) are highlighted accordingly in both secondary and tertiary structures. The wobble base pair (black) in acceptor stem is highly conserved in $tRNA^{Ala}$ of all species.

It has been discovered that in *Escherichia coli* the identity set for alanine tRNA ($tRNA^{Ala}$) is just the single G•U wobble base pair (3•70) in the middle of acceptor stem (**Hou & Schimmel, 1988; McClain & Foss, 1988**) away from the triplet anticodon (**Figure 1.2**), which further gave rise to the concept of second genetic code (**de Duve, 1988**). This G•U wobble base pair is conserved in $tRNA^{Ala}$ of all kingdoms of life (**Hou & Schimmel, 1988**).

1.2.4 Structural and Biochemical Studies Related to $tRNA^{Ala}$ Recognition by AlaRS

AlaRS are dimeric class II synthetase, composed of four functional domains: N-terminal aminoacylation domain, tRNA-recognition domain, editing or proofreading domain, and C-terminal domain.

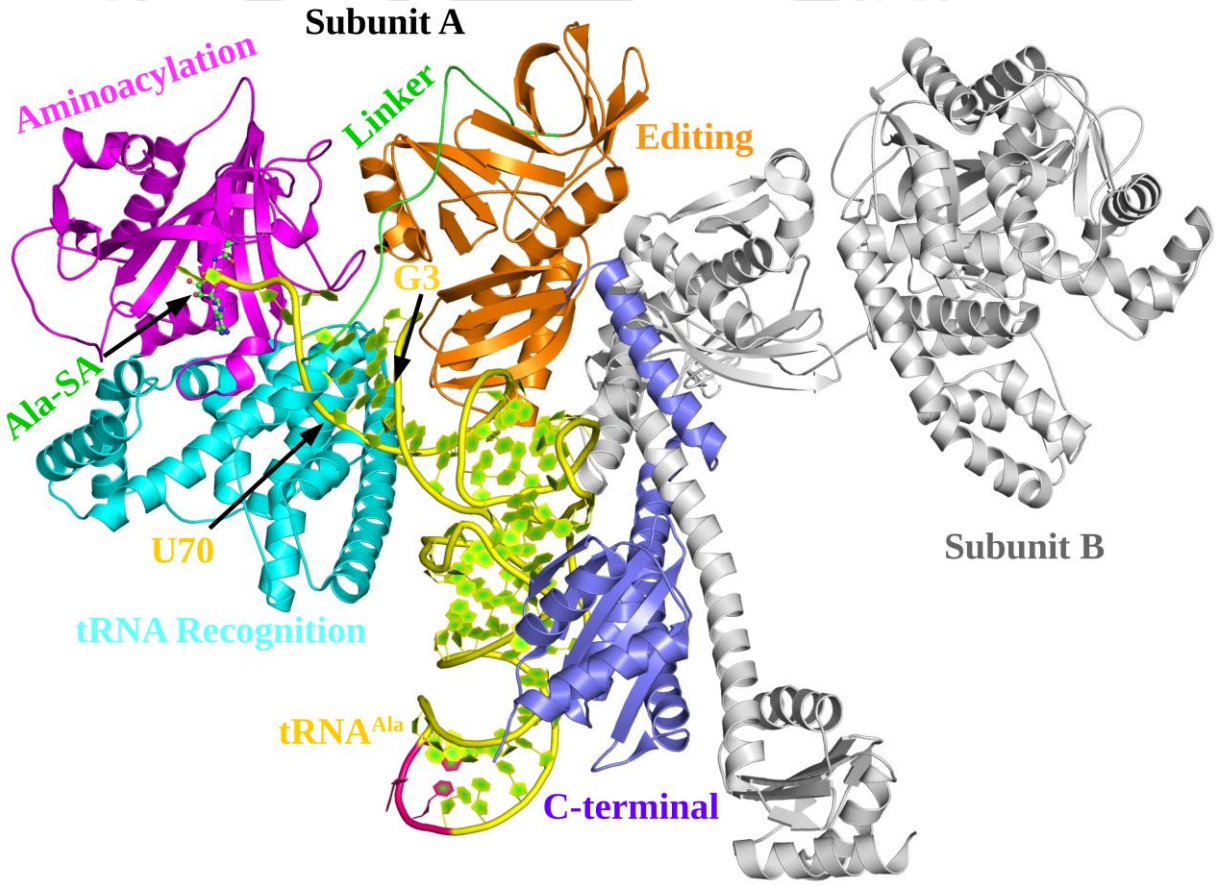


Figure 1.3: Three-dimensional structure of AlaRS dimer (subunit A (colored) and B (grey)) in complex with alanine tRNA (yellow, PDB: 3WQY). Different domains of AlaRS subunit A are shown in distinct colors: aminoacylation (magenta), tRNA recognition (cyan), editing (orange) and C-terminal domain (blue).

The N-terminal aminoacylation domain (**Figure 1.3**) of AlaRS adopts a typical class-II aminoacylation form (**Ruff et al., 1991; Biou et al., 1994**), and synthesizes alanine-adenylate for tRNA aminoacylation. It is directly involved in the interaction with the acceptor arm of tRNA and aminoacyl-adenylate. The aminoacylation domain forms a concave pocket where the amino acid is ligated to the CCA end of tRNA (**Smith & Hartman, 2015**). The tRNA recognition domain of AlaRS widens the major groove of $tRNA^{Ala}$ acceptor stem and recognizes the G3•U70 wobble base pair (**Naganuma et al., 2014**). The editing domain is involved in proof-reading and removes the incorrectly acylated tRNA. AlaRS functions as a dimer and the dimerization is mediated by C-terminal and editing domains.

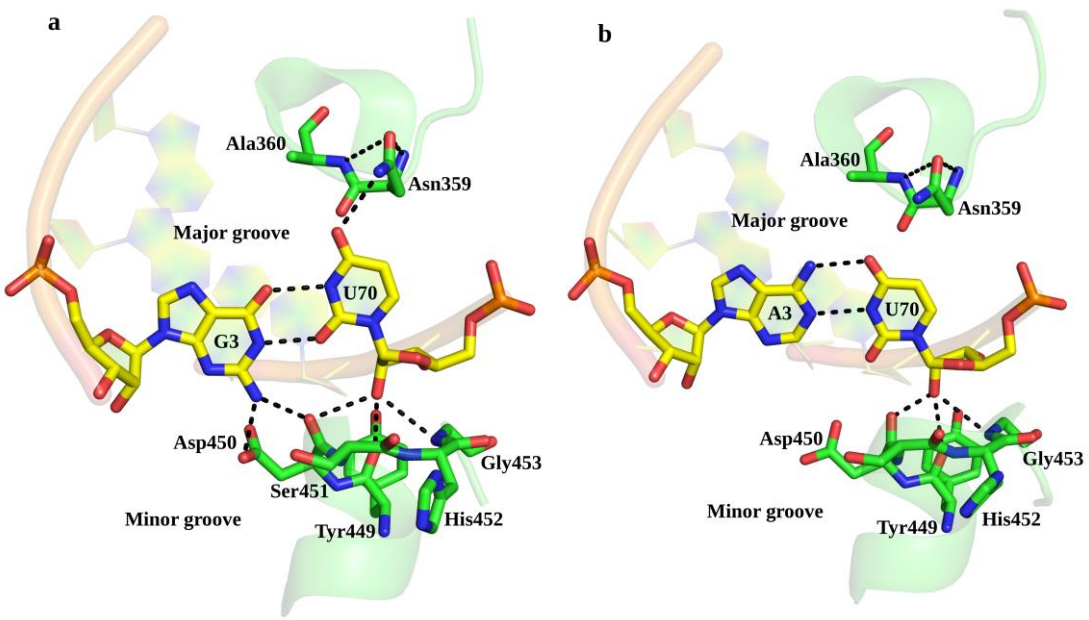


Figure 1.4: Differences between structures of *A. fulgidus* AlaRS in complex with $tRNA^{Ala}$/G•U (PDB:3WQY) and $tRNA^{Ala}$/A•U (PDB:3WQZ). The G3•U70 base pairs differ from A3•U70 in the type of functional groups projected into major and minor grooves. Hydrogen bonds are shown with black dashed lines.

Biochemical study (**Francklyn & Schimmel, 1989, Varani et al., 2000**) have shown that mutation of critical G3•U70 base pair in $tRNA^{Ala}$ by G3•C70/A3•U70 in the acceptor stem diminishes or prevents the aminoacylation with alanine by AlaRS. Thus suggests the wobble G3•U70 base pair is critical for function.

Further, experiments confirmed that upon transfer of G3•U70 base pair to a synthetic RNA hairpin minihelix (composed only of tRNA acceptor stem), enabled alanine aminoacylation. Experimental studies established two key points related to $tRNA^{Ala}$ charging by AlaRS, (1) Wobble G3•U70 base-pair close to the acceptor stem is the signature of $tRNA^{Ala}$, which is recognized by AlaRS (2) Anticodon triplet in not important for alanine aminoacylation by AlaRS. McClain group (**Chang et al., 1999**) suggested that C•C is the best mimic of G•U, but G•U base pair in $tRNA^{Ala}$ is conserved in all species, and no $tRNA^{Ala}$ sequence has been found to contain the C•C pair throughout the course of evolution.

Recently (**Naganuma et al., 2014**), structures of AlaRS of *A. fulgidus* in complex with $tRNA^{Ala}$/G3•U70 and the mutated variant with $tRNA^{Ala}$/A3•U70 in complex with alanyl-adenylate analogue has been crystalized; which explain the mystery of the strict AlaRS specificity for $tRNA^{Ala}$ consisting G3•U70 base pair. These structures suggest that geometrical differences between G3•U70 and A3•U70 propagate toward the 3'-CCA region leading A76 of $tRNA^{Ala}$/G3•U70 to the catalytic site while keeping A76 of tRNA/A3•U70 away from it. Further, AlaRS establishes specific interaction network in the major and minor groove of G3•U70 which is missing in other variants (A3•U70, G3•C70, and A3•C70) (**Figure 1.4**). It has been suggested (**Naganuma et al., 2014**) that $tRNA^{Ala}$ charging by AlaRS occurs by preferential binding (smaller k_m for $tRNA^{Ala}$/G3•U70 relative to other 3•70 variants) and facile aminoacylation (higher k_{cat} for $tRNA^{Ala}$/G3•U70 relative to other 3•70 variants). AlaRS active site conformation was proposed to be affected by the nature of 3•70 base-pair of the bound tRNA (**Park & Schimmel, 1988**).

Structural and biochemical information has undoubtedly enriched our understanding of $tRNA^{Ala}$ aminoacylation by AlaRS. Moreover, 3D structures of AlaRS complexes provide invaluable information; but a clear atomic insight into dynamics and detailed energetics of molecular recognition of $tRNA^{Ala}$ aminoacylation by AlaRS is still lacking. The following issues related to $tRNA^{Ala}$ recognition by AlaRS are unclear:

(a) Energetic basis of AlaRS selectivity in favour of cognate $tRNA^{Ala}$ (with G3•U70), disfavouring near-cognate tRNA's (with A3•U70, G3•C70, and A3•C70).

(b) The connection between calculated energetics, structures, and experimental kinetic measurements.

1.3 Stop Codon Recognition by Eukaryotic Release Factor1

1.3.1 Background

The premise of molecular biology states that genetic information in a cell is stored in the DNA. The information in DNA is transferred to RNA by a process known as transcription. Finally, RNA is processed by the ribosome to synthesize proteins (**Crick, 1970**) by a process known as translation. The synthesis of correct proteins in the cell is directly related to the fitness of all living organisms. To understand how muscles grow, how our bodies function, how genes function, it is important to comprehend the protein synthesis process. Translation machinery requires the following major components:

(a) *mRNA*- mRNA carries the genetic information (as three-nucleotide codes called "codon") for protein synthesis. Prokaryotic mRNA is ready for protein-synthesis immediately after transcription while eukaryotic mRNA undergoes post-transcriptional processing (**Haugland and Cline, 1980**). Codons that code for amino-acids are called as sense-codons.

(b) *tRNA*- tRNA acts as an adaptor molecule linking the nucleotide sequence to the amino acid; it is composed of RNA, generally 70 to 80 nucleotides in length. An amino acid is attached to the CCA 3'-terminal group in the acceptor arm. The anticodon arm contains the triplet nucleotide bases called "anticodon" required to base-pair with the mRNA codon in the ribosome (**Ibba and Soll, 2000**).

(c) *Ribosome*- ribosomes is the key machine which hosts the key components of translation and ensures correct mRNA-tRNA base pairing or mRNA-protein interactions. The ribosome is known to contain three sites for tRNA binding: aminoacyl (A), peptidyl (P), and exit (E). The A-site is a binding site for the amino-acyl tRNA (amino-acid bound tRNA), while the P-site contains the tRNA that holds the nascent poly-peptide chain. During elongation, the polypetide chain is transferred from P-site to A-site tRNA. The P-site tRNA after transferring the polypeptide chain is released from the ribosome through the E-site and the A-site tRNA containing the polypeptide chain is moved to the P-site (**Rheinberger, 1991**). The cycle continues during elongation process.

(d) *Translation Factors*- proteins that bind to the ribosome and regulate the translation process during the different phases. In prokaryotes, initiation is regulated by 3 factors (IF1, IF2, and IF3) as compared to 10 initiation factors in eukaryotes (**Merrick and Hershey, 1996**).

In bacteria, during elongation eEF1 (EF-Tu) delivers aminoacyl tRNA to the A site while eEF2 (EF-G) moves tRNA from A to P site. During the termination phase, in prokaryotes, all three stop codons are decoded by two different release factors (RF1 and RF2) while eukaryotes rely on an evolutionarily unrelated omnipotent release factor eRF1 (**Kisselev and Frolova, 1995**).

1.3.2 Translation Termination

The elongation cycle continues until the A-site of the ribosome encounters UAA/UAG/UGA mRNA codon (Stop codon or non-sense codon). The termination of protein synthesis occurs when a stop codon in the mRNA encounters the ribosomal A-site and signals the end of the mRNA open reading frame (**Figure 1.5**). In most of the species, three codons (UAA, UAG, and UGA) out of 64 are used for translational termination. Stop codons are not recognized by tRNAs, unlike sense codons. Stop codons are recognized by specialized proteins called class I release factors (RFs), which terminates the nascent peptide chain (**Freistroffer et al., 2000**).

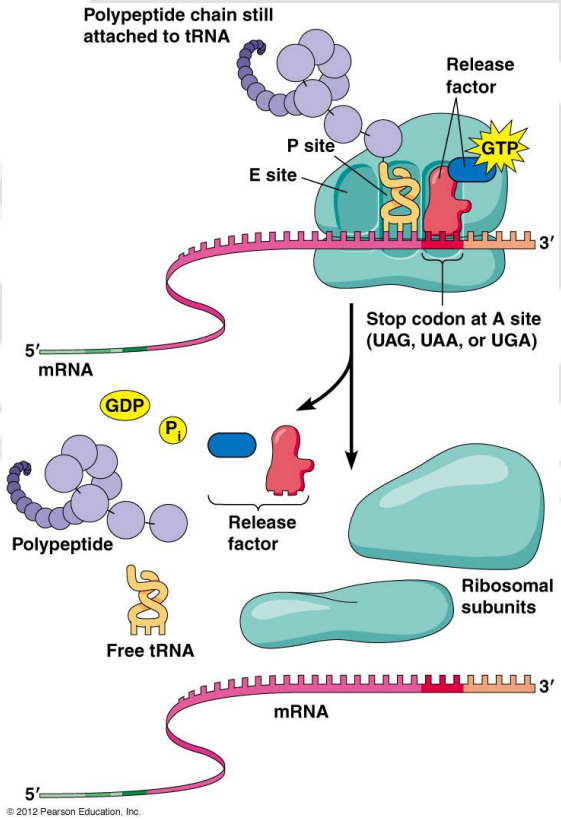


Figure 1.5: Schematic representation of translation termination process (https://www.ncbi.nlm.nih.gov/).

Release factors recognize the stop-codon and initiate the termination of translation by cleaving the bond between the polypeptide chain and P-tRNA. In eubacteria, stop codons are recognized by two types of release factors RF1 and RF2 (RF1 recognizes UAG and UAA, while RF2 recognizes UGA and UAA) whereas, in archaea and eukaryotes, all three stop codons are decoded by a single release factor, aRF1, and eRF1, respectively (**Kisselev, 2002**). Besides, class I release factors are being assisted by class II release factors (RF3) in a GTP-dependent manner. As a response to release factor binding to the ribosomal A site in the small ribosomal subunit, they cleave the ester bond bridging the polypeptide moiety and the A76 of P site tRNA through hydrolysis in the large subunit of ribosome (**Frolova et al., 199**9) (**see figure 1.5**). After translation termination, ribosomal subunits are dissociated and make themselves available for the next translational cycle.

Accurate translational termination is essential for life as it defines the lengths of cellular proteins, and it must be achieved with high fidelity. Premature translation termination at sense codons, results in dysfunctional truncated proteins, is rare and occurs less frequently. Premature translational termination leads to many genetic diseases, highlighting the toxicity of early termination (**Frischmeyer and Dietz, 1999**).

1.3.3 Structural and Biochemical Studies of eRF1

The translation termination process in eukaryotes significantly differs from that in bacteria. In this thesis, we have focused on the eukaryotic release factor (**Figure 1.6**). In eukaryotes, a single omnipotent release factor eRF1 recognizes all three stop codons. The eRF1 contains three functionally distinct domains (**Frolova et al., 2000**): the N-terminal domain mimics the tRNA anticodon arm and consists highly conserved GTS (residue no. 31-33), YxCxxxF (residue no. 125-131), and TAS-NIKS (residue no. 58-64) (numbering corresponds to human eRF1) motifs which ensure stop-codon recognition, the M domain mimics the tRNA acceptor stem, and it consists universal GGQ motif (residue no. 183-185) which is essential for peptidyl-tRNA hydrolysis at the peptidyl transferase centre, and the C-terminal domain recruits translational GTPase eRF3 to facilitate translation termination (**Figure 1.6**).

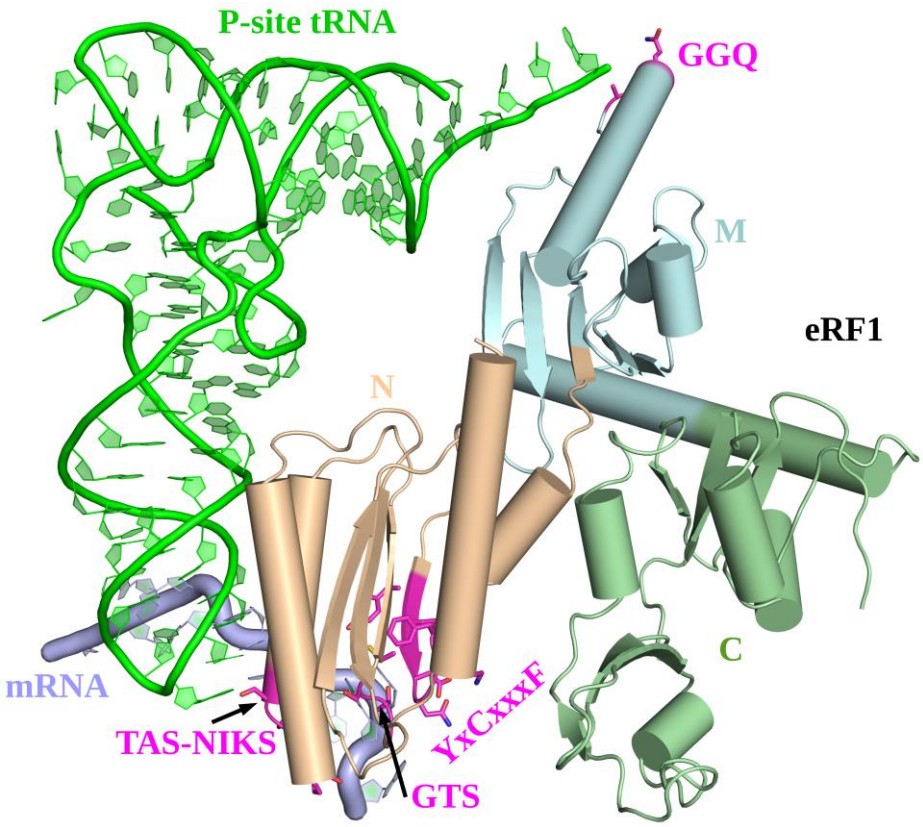


Figure 1.6: eRF1 and P-site tRNA (green) bound to mRNA (blue) in the ribosome (not shown) (taken from PDB: 3JAG). Different domains of eRF1 (N: brown, M: cyan, C: green) are indicated by distinct colors. Functional motifs GGQ (residue no. 183-185), TAS-NIKS (residue no. 58-64), YxCxxxF (residue no. 125-131), and GTS (residue no. 31-33) are shown in magenta.

Biochemical studies (**Ito et al., 2000; Kervestein et al., 2001**) suggest that stop codon recognition is associated with release factors rather than the tRNA. In vitro mutations (**Frolova et al., 1999; Seit-Nebi et al., 2000**) in the M-domain GGQ motif demonstrated complete or partial loss of the release factor activity, thereby suggesting its universality and functional importance in peptidyl-tRNA hydrolysis. The heptapeptide TAS-NIKS motif in N-domain has been found to be involved both in the stop codon decoding (**Knight and Landweber, 2000**) and the ribosome binding (**Kisselev et al., 2000**). A reduction in eRF1-ribosome binding ability upon TAS-NIKS mutations has been confirmed by the site-directed mutagenesis study (**Frolova et al., 2002**). The in vitro mutation study of YxCxxxF motif suggested it's reduced ability to discriminate the purine bases at the second and third positions of the stop codon (**Kolosov et al., 2005; Conard et al., 2012**).

Structures of eukaryotic translational termination complexes (**Bhushan at al., 2010; Brown et al., 2015; Matheisl et al., 2015**) suggest a compact U-turn like conformation of mRNA that accommodates four nucleotides (including three stop codon nucleotides) in the A site of the ribosome (**Figure 1.7**). This compact U-turn conformation in A site is stabilized by stacking of ribosomal bases A1825 and G626 of 18S rRNA with the second and fourth bases of mRNA, respectively (**Bhushan at al., 2010; Brown et al., 2015; Matheisl et al., 2015**) (**Figure 1.7**). Structural studies (**Brown et al., 2015**) along with photochemical cross-linking experiments (**Chavatte et al., 2002**) suggest that residues of conserved TAS-NIKS motif (Asn61 and Lys63) form hydrogen bonding with both carbonyl groups of invariant uridine and also disfavour purines at the first position of the stop codon due to steric hindrance (**Chavatte et al., 2002**) (**Figure 1.8**). Therefore, imposing a requirement for uridine at first position of the stop codon.

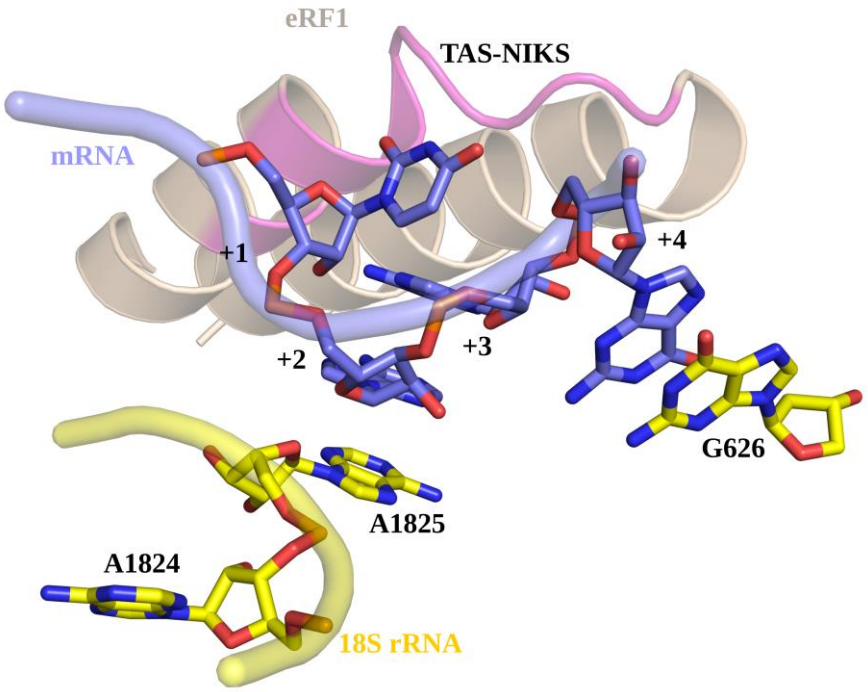


Figure 1.7: Zoomed in view of the ribosomal A-site of the eukaryotic translation termination complex (from PDB: 3JAG). 4mRNA bases (+1 to +4) including the stop codon UAA (blue sticks), ribosomal bases (yellow sticks), eRF1 (brown, TAS-NIKS motif in magenta).

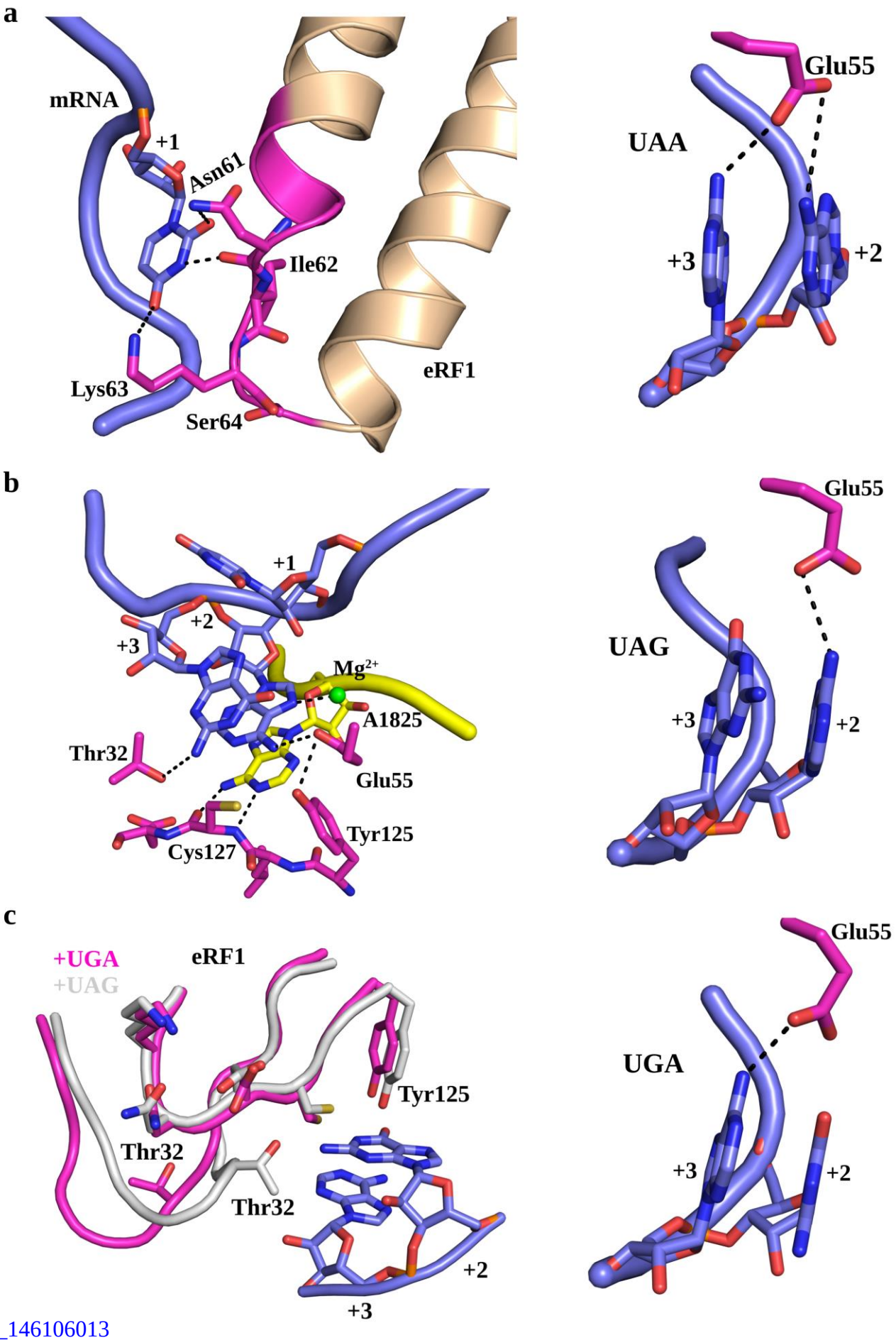

a
mRNA
+1
Asn61
Ile62
Lys63
Ser64
eRF1
Glu55
UAA
+3
+2
b
+1
+2
+3
Mg2+
A1825
Thr32
Glu55
Tyr125
Cys127
Glu55
UAG
+3
+2
c
+UGA
+UAG
eRF1
Tyr125
Thr32
Thr32
+2
+3
Glu55
UGA
+3
+2

Figure 1.8: Interaction network involving stop-codon, eRF1, and ribosome in the translation termination complex. (a) UAA bound structure: PDB 3JAG. TAS-NIKS motif (magenta, residue no. 58-64) hydrogen bonds with the +1 uridine (purple). Two adenines AA (+2,+3 bases) forms electrostatic contact with Glu55. (b) UAG bound structure: PDB 3JAH. The first base (+1 U) interaction network is the same as (a). A1825 of 18S rRNA (yellow) stack with the 2nd base +2 A. YxCxxxF motif (residue no. 125-131) forms h-bonds with +2 A and position A1825 for efficient stacking. Thr32 of the GTS loop forms a hydrogen bond with the exocyclic amine of +3 G. Glu55 interacts with 2nd base (+2 A). (c) UGA bound structure: PDB 3JAI (magenta). 1st base interaction network and the ribosomal bases are the same as in (a), (b). Thr32 is away from 3rd base +3 A (UAG-bound complex in grey for comparison). Glu55 of eRF1 (magenta) forms hydrogen bond with the +3 A base.

Further, it has been proposed that interactions of the YxCxxxF motif, GTS motif, and highly conserved Glu55 of eRF1 decode second and third position of stop codons (**Brown et al., 2015**). The GTS loop is suggested to be flexible to adopt two distinct conformations depending on the basis of purines (adenosine and guanosine) at the second and third position of stop codons (**Brown et al., 2015**) (**Figure 1.8**).

A wealth of biochemical and invaluable structural information of ribosome complexes have surely enhanced our understanding of stop codon recognition by eRF1. However, the following issues related to stop codon decoding were unknown:

(a) To what extent eRF1 favours stop codon binding relative to sense codons (e.g, CAA, CGA, CAG, and UGG).

(b) The relationship between relative binding energy and the 3D structures of cognate and near-cognate complexes.

(c) Role of water in the stop codon recognition by eRF1.

(d) Strength of stop-codon selectivity: Eukaryotic vs Prokaryotic release factor.

1.4 Viral RNA Recognition by RIG-I

1.4.1 Background

Viruses have always been a constant threat to mankind. Recent Ebola and Zika virus outbreak shows the scale at which viruses pose threats to a growing and increasingly mobile population (**Depoux et al., 2018; Gebretadik et al., 2015**). Many pathogenic viruses (e.g, Influenza, Dengue, yellow fever, Ebola, Zika, Hepatitis C, etc.) consist of the RNA genome. They infect animal cells by injecting their own RNA into the host cells and give rise to the epidemic and rarely pandemic diseases. The capacity of selectivity, bind to viral RNA's and induce antiviral response while leaving self-RNA's, is the goal of the sophisticated mammalian immune system. The immune system consists of two major sub-systems (a) an innate immune system and (b) an adaptive immune system. The innate immune response is less selective and triggered as a result of microbe recognition by pattern recognition receptors (PRRs) (**Medzhitov, 2007**). On the other hand, the adaptive immune system is a stronger and more selective immune response, where pathogen is recognized by immunological history or memory from the past (**Pancer & Cooper, 2006**). Pattern recognition receptors (PRRs) recognize conserved patterns among a broad group of pathogenic microorganisms. PRRs are proteins that sense the pathogen or danger-associated molecular patterns (PAMPs, DAMPs) (**Mogensen, 2009**) and activate the antiviral response in cells. PRR proteins include transmembrane Toll-like receptors (TLR) that sense viral RNA in endolysosomal compartment and extracellular space, and RIG-I like receptors (RLR) that sense the intracellular cytosol space (**Mogensen, 2009**).

Three types of proteins, RIG-I (retinoic acid inducible gene I), MDA5 (melanoma differentiation-associated 5), and LGP2 (laboratory of genetics and physiology 2) together forms RLR family. RIG-I and MDA5 stimulate innate immune response; while LGP2 is known to regulate RIG-I mediated signaling. In this thesis, we have focused to understand the principle of viral-RNA selection by RIG-I. Panhandle-like double-stranded RNA with blunt end 5' tri/di-phosphate is believed (**Schlee et al., 2009**) to be a characteristic feature of viral genome. RIG-I recognizes viral dsRNA based on the 5'-terminal moiety (presence of tri/diphosphate at the 5'-terminal). A wide variety of positive- and negative-strand RNA viruses are recognized by RIG-I, such as influenza, Ebola, Rabies, hepatitis C, and Respiratory syncytial viruses (**Kawai & Akira, 2007**).

1.4.2 RIG-I and its Different Domains

Human RIG-I is a 925 residue long highly conserved protein within vertebrates (**Loo & Gale, 2011**) (**Figure 1.9**). RIG-I is composed of a C-terminal domain (CTD), helicase domain (Two RecA-like helicase domains: Hel1, Hel2, and an insertion domain Hel2i), Pincer domain (P), and N-terminal caspase activation and recruitment domains (CARDs). CTD binds to the 5' terminal of dsRNA and the information is transmitted to the Hel2 domain via pincer (P) domain. Hel1 domain contains an ATP binding site. Helicase domain binds to the backbone of dsRNA. N-terminal tandem caspase activation and recruitment domains (CARDs) is responsible for downstream antiviral signalling (**Yoneyama et al., 2004; Yoneyama et al., 2005; Saito et al., 2007**). Although RIG-I is built from the parts of helicase proteins (functions of helicases are unwinding, widening and unpackaging) but the function of RIG-I is more like a switch rather than unwinding.

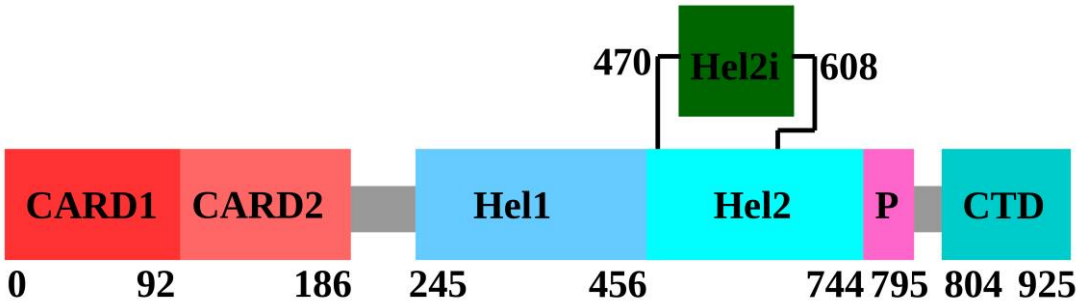


Figure 1.9: A color-coded schematic representation of the domain architecture of RIG-I primary structure. Flexible linkers are shown in grey lines.

1.4.3 Features of RNA Recognized by RIG-I

Most of the studies agree that RIG-I is activated upon binding of short (<300 bp) double-stranded RNA (dsRNA) having blunt end with 5'-triphosphate (5'-ppp) moiety (**Hornung et al., 2006; Kato et al., 2008; Kato et al., 2006; Pichlmair et al., 2006; Schmidt et al., 2009**). Earlier reports suggest a strict requirement of 5'-triphosphate, although the optimal in vitro RIG-I ligand is defined as panhandle-like Watson-Crick base-paired blunt-ended short 5' tri- or di-phosphorylated dsRNA with at least 10 base pairs (**Luo et al., 2012**), see **figure 1.10**. Thus panhandle like double-stranded RNA with 5'di-phosphate is a requirement for RIG-I activation (**Uzri and Gehrke, 2009**).

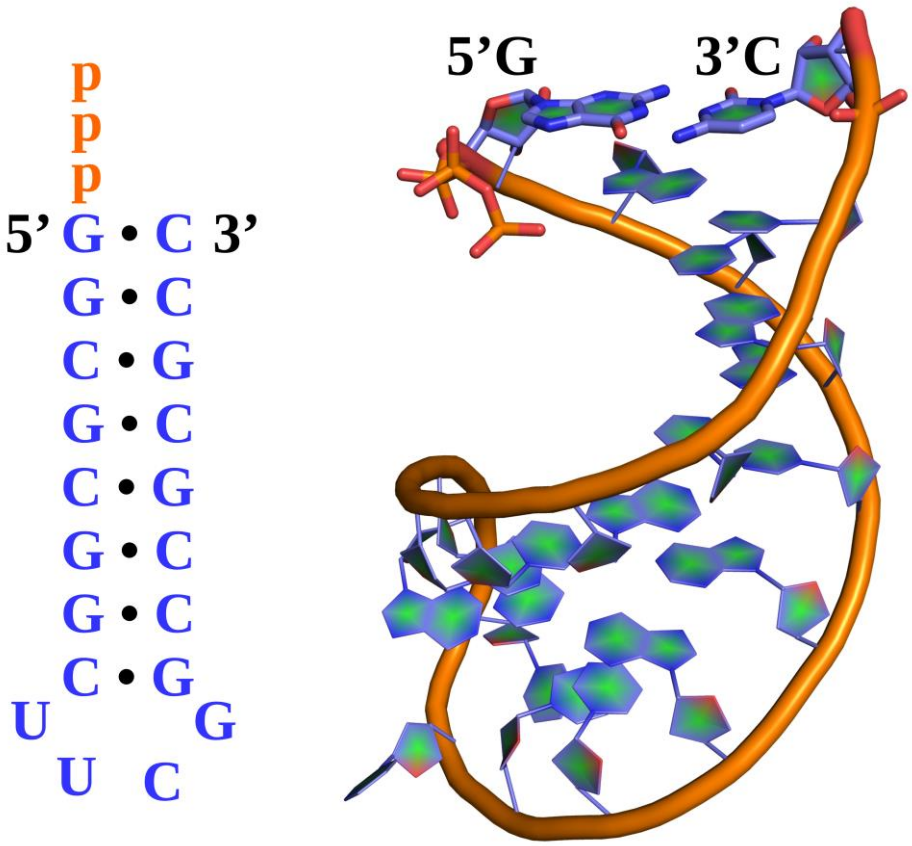


Figure 1.10: Secondary and tertiary structure of panhandle-like RNA molecule (PDB: 4AY2).

1.4.4 Current View of Antiviral Response Triggered by RIG-I

Crystal structures of dsRNA bound or unbound RIG-I from several labs have provided detailed structural insights into the mode of action of this receptor protein (**Wang et al., 2010; Lu et al., 2010; Kowalinski et al., 2011; Jiang et al., 2011; Luo et al., 2012; Kohlway et al., 2013; Devarkar et al., 2016; Ren et al., 2019**). In the absence of viral dsRNA, RIG-I exists as an extended monomer in an auto-repressed “signal off” conformation. CTD access wider conformational space in the cytosol in the absence of substrate RNA. The tandem CARD domains are tightly packed against HEL2i (**Figure 1.11, a**); thereby preventing downstream signaling (**Luo et al., 2012**). Upon dsRNA binding, RIG-I is proposed to undergo a dramatic conformational rearrangement (F**igure 1.11, b**) adopting a compact configuration, thereby allowing ATP to bind in the ATP binding pocket formed by the helicase core (**Kowalinski et al., 2011; Luo et al., 2012; Rawling & Pyle, 2014**).

ATP binding/hydrolysis (as shown in **Figure 1.11, c**) further squeezes/compacts the RIG-I, potentially dislodging the CARDs and exposing them for further downstream signaling.

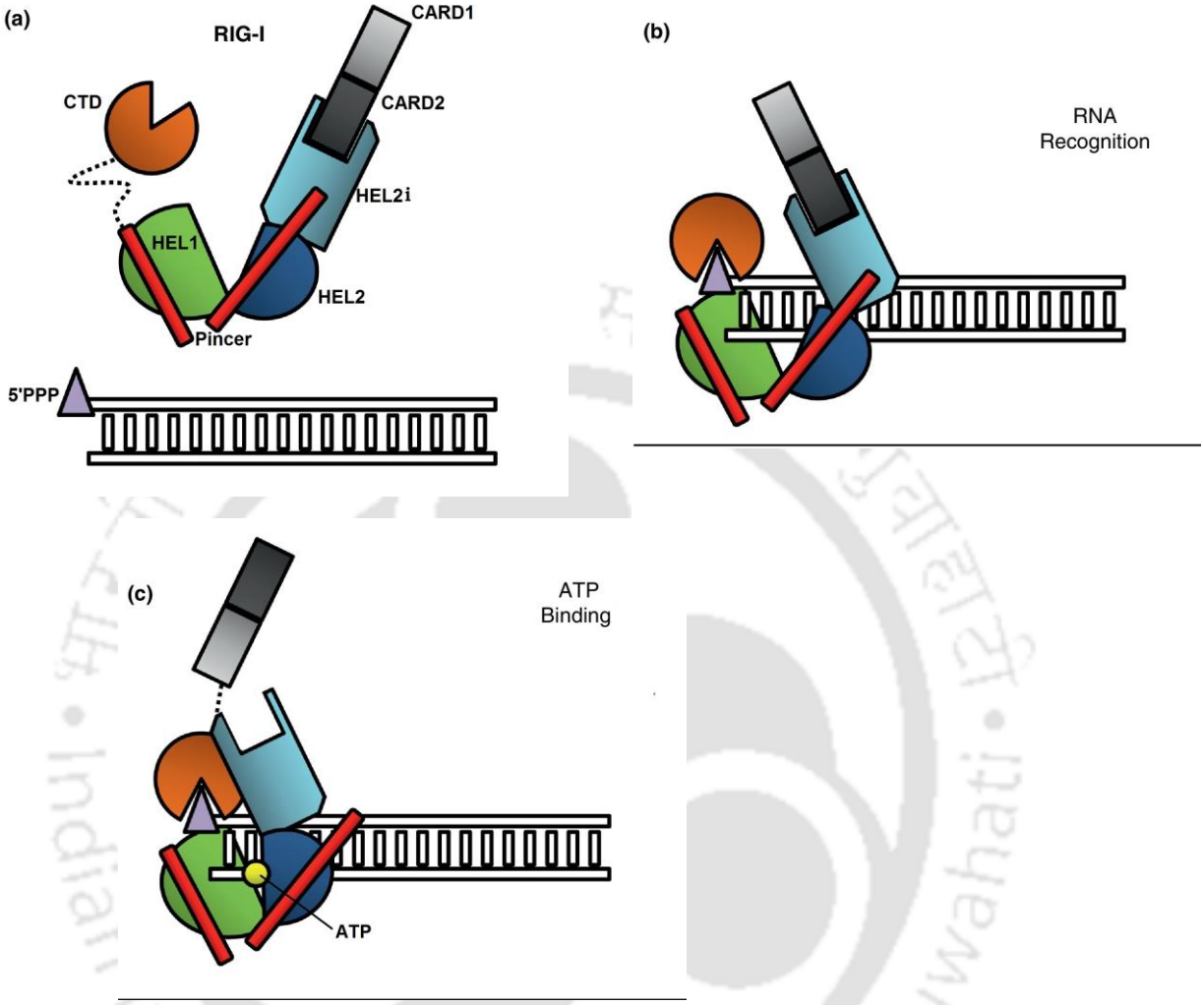


Figure 1.11: Schematic representing the stages of RIG-I activation. (a) RIG-I surveys the cell in an auto-repressed conformation. (b) The CTD of RIG-I binds terminal phosphates of dsRNA very strongly. (c) RIG-I undergoes conformational changes and ejects CARDs for downstream signaling (adapated from **Rawling & Pyle, 2014**).

1.4.5 Structural and Biochemical Studies Related to Viral RNA Recognition by RIG-I

RIG-I studies can broadly be classified into two groups: structural and biochemical studies. Structural studies (**Wang et al., 2010; Lu et al., 2010; Kowalinski et al., 2011; Jiang et al., 2011; Luo et al., 2012; Kohlway et al., 2013; Devarkar et al., 2016; Ren et al., 2019**) of viral RNA targets bound to RIG-I from several groups have captured different

conformational states and enriched our understanding of the interaction network between RIG-I and dsRNA containing terminal 5'-ppp, 5'-pp or 5'-OH. **Lu et al., 2010;** suggest a structural similarity between CTD of RIG-I in the bound and free state, excluding the loop region (residue 847-853) which is highly flexible in the free state in comparison of dsRNA bound structures (**Cui et al., 2008**). Structures (PDB 3LRR, 3LRN) crystalized by **Lu et al., 2010**; suggest that CTD of RIG-I forms extensive and conserved network of electrostatic interactions with α, β, γ-phosphates of 5' ppp-dsRNA, whereas structure (4AY2) crystalized by **Luo et al., 2012**; showed no direct interaction of CTD (see **Figure 1.12, b**) with γ-phosphate similar to the structure from the group **Lu et al., 2010** (see **Figure 1.12, a**). Clearly, the interaction between RIG-I and γ-phosphate of dsRNA appears to be conflicting. 5' OH-dsRNA bound RIG-I complex suggest the absence of interaction between protein and 5' terminal (**Ren et al., 2019**).

Structural studies confirmed that α-phosphates of dsRNA forms direct interaction with RIG-I, thus, one might assume that 5' p-dsRNA would bind RIG-I with an affinity that is somewhere between values determined for 5' ppp/pp-dsRNA and 5'-OH complexes. Interestingly, the structure of 5' p-dsRNA bound RIG-I could not be isolated despite several attempts.

Biochemical studies estimated the binding affinity of RIG-I for dsRNA's with 5' ppp, 5' pp, 5' p and 5' OH moiety at 5'-terminus. The results (**Goubau et al., 2014; Linehan et al., 2018, Ren et al., 2019**) suggest that (1) both 5' ppp-dsRNA and 5' pp-dsRNA binds to RIG-I with more or less similar affinity and induce the antiviral response. Thus, 5' pp-dsRNA is the minimum requirement for RIG-I binding and antiviral response. Further, it has been suggested that separated RIG-I domains bind more strongly to viral RNA than the full RIG-I because covalently linked domains undergo conformational changes to enhance the specificity for efficient discrimination of viral RNAs from host RNAs (**Vela et al., 2012**). (2) 5' ppp/pp-dsRNA binds to RIG-I very strongly (~ pM range) than 5' p/OH analogue. It is known that there is a high concentration of 5' p/OH-dsRNA in cytosol because of RNA processing and nuclease degradation, but RIG-I weak binding affinity might have prevented antiviral response of RIG-I for self-RNAs. (3) RIG-I binds 5' p-dsRNA very weakly (~3.5 fold weaker) than 5' OH-dsRNA (**Ren et al., 2019**), despite the presence of α-phosphate atoms in former.

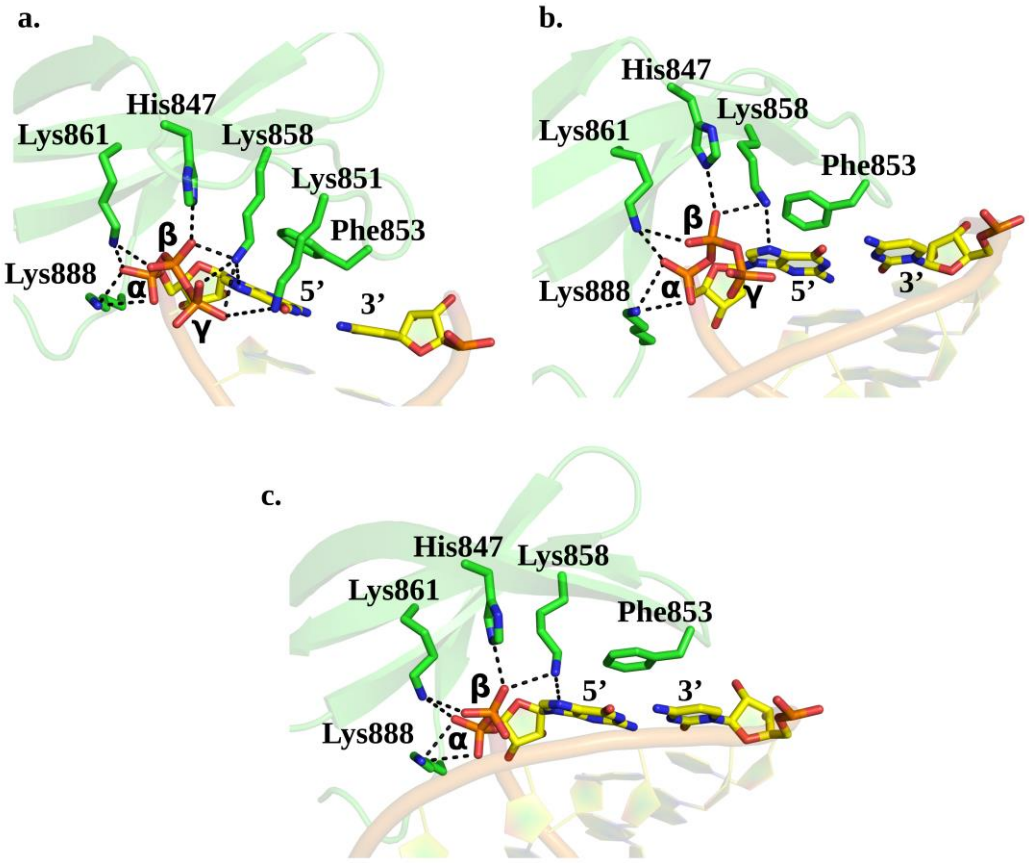


Figure 1.12: (a) RIG-I: 5' ppp-dsRNA bound complex (PDB 3LRR). α, β and γ phosphates of 5' terminal of dsRNA forms electrostatic contact with RIG-I. (b) RIG-I: 5' ppp-dsRNA bound complex (PDB 4AY2). γ phosphate of 5' ppp-dsRNA does not show electrostatic interaction with RIG-I. (c) RIG-I:5'-pp-dsRNA bound complex (PDB 3NCU). Direct interaction between α and β phosphates with RIG-I was evident as shown in (a) and (b). The network of electrostatic interactions as a dashed line. Key residues of RIG-I CTD involved in RNA binding are shown as stick models.

Despite the wealth of structural and biochemical studies, the link between microscopic structure and thermodynamics is missing. In this thesis, we have focused to compute the free energy differences of RIG binding to various RNA's differing at 5'-terminal (5' ppp, 5'pp, 5' p, and 5' OH) and established a link between the computed free energy differences with the 3D structures. The effort was to understand (1) why 5' pp-dsRNA binding to RIG-I is comparable with 5' ppp-dsRNA? (2) Why 5' p-dsRNA is a better binder to RIG-I than 5' OH-dsRNA?

1.5 Methods

Classical Molecular dynamics (MD) simulations were used to study protein-RNA interactions. The free energy differences between cognate (correct) and non-cognate (incorrect) protein-RNA complexes were computed from molecular dynamics trajectories using an appropriate thermodynamic cycle. Molecular dynamics free energy calculations provide a direct link between atomic structure and dynamics of biomolecules to its macroscopic thermodynamic property of prime interest, i.e, free energy. The general theoretical frameworks and principles involved in MD simulations and its application to quantify free energy differences are discussed in this chapter. The methodology specific to the biological problem is addressed in the corresponding chapters (2-4).

1.5.1 Brief History of Molecular Dynamics and Free energy Simulations

Molecular dynamics is a computer simulation method first introduced by Alder and Wainwright in 1957 to study phase transition in hard-sphere model liquids (**Alder and Wainwright, 1957**). The application of the same methodology to study protein dynamics (BPTI protein distributed in ~ 8 ps time-scale) appeared in 1977 by the McCammon group. (**McCammon et al, 1977**). Thereafter the need for longer simulation time-scale and consideration of larger size models lead to rapid development in this area. Access of supercomputers with efficient algorithms certainly allow us to perform molecular dynamics simulations for longer time (protein NTL9 for 1.073 milliseconds) (**Lindorff-Larsen et al., 2011**) and larger size (Full satellite tobacco mosaic virus (STMV): 100000 atoms, Simulation time: 50 nanoseconds) (**Freddolino et al., 2006**). The struggle for bridging the gap between time-scale and size-scale using molecular dynamics simulations is always of great concern for computational chemists/biologists. Free energy calculations are a choice for selecting relevant degrees of freedom of biomolecules and bridge the gap between time-scale and size-scale in an efficient way. Free energy calculations are commonly used in studying recognition in a biomolecular association, transport phenomenon, conformational transitions, etc. The calculations of estimating free energy differences could be classified into four different route: (1) Histogram based (**Torrie & Valleau, 1974; Widom, 1963**) (2) Non-equilibrium simulations based (**Jarzynski, 1997**) (3) Perturbation theory-based (**Zwanzig, 1954**) (4) Based on the force calculations (**Kirkwood, 1935; Carter, 1989**). In this thesis, the free energy calculations were performed adopting perturbation theory.

X-ray or Cryo-EM structures are unable to give insight into the biomolecular dynamics. The dynamics of protein-nucleic acid complexes were explored by MD simulations. A plentiful of information concerning structural and dynamic properties (such as breakage/formation of hydrogen bonds, solute-water interaction, substrate-ligand binding, and free energies) were obtained by analyzing the molecular dynamics trajectories of the bio-macromolecules. MD simulations allow us to simulate the natural dynamic motion of the biomolecules and help in explaining the wide range of thermally-accessible states, thereby providing a possible clue on structure, function relationship. Furthermore, calculations of free energy differences could estimate biomolecular recognition and establish a link with the atomic structures.

1.5.2 Theory of Molecular Dynamics Simulations

The molecular dynamics simulation method is based on Newtonian mechanics where the successive system configurations are generated by numerically solving Newton's second law (**equation 1**). The solution of the equation results in a trajectory that describes the positions, velocities, and accelerations of the interacting particles in the system as a function of time. This trajectory can be used to extract the average structural and thermodynamic properties by employing statistical mechanics. The method is deterministic, from the knowledge of the positions and velocities of each atom in the system the future state can be determined. Newton's second law of motion in its most simplistic form is given by:

$$F_i = m_i a_i = m_i \ddot{r}_i \qquad (1)$$

where F_i is the force acting on particle i, m_i is the mass of particle i, a_i is its acceleration, and $\ddot{r}_i$ is the second derivative of the particle position r with respect to time. The force can also be determined by the gradient of the potential energy (U(r)) function

$$F_i = -\nabla_i U(r) \qquad (2)$$

Combining equation (1) and equation (2), it can be written as

$$\frac{-dU(r)}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \qquad (3)$$

where U(r) is the potential energy function of the system that, in general, is a function of the position of all the atoms. Force field defines the value of U(r) and allows estimation of F_i. Thus the accuracy of force estimation depends on the accuracy of the force fields.

Newton's equation of motion, as shown in **equation 3** relates the derivative of the potential energy to the changes in the coordinates of atoms as a function of time. To generate molecular dynamics trajectories, we need to solve Newton's equation of motion, which involves numerical integrations. A standard method for solving these ordinary differential equations is the finite difference approach. In this approach, the molecular coordinates and dynamic properties (such as velocities and accelerations) are integrated using Taylor series expansions:

$$r(t+\Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + \frac{1}{6}b(t)\Delta t^3 + \cdots \quad (4)$$

$$v(t+\Delta t) = v(t) + a(t)\Delta t + \frac{1}{2}b(t)\Delta t^2 + \cdots \quad (5)$$

$$a(t+\Delta t) = a(t) + b(t)\Delta t + \cdots \quad (6)$$

where r is the position, v is the velocity (the first derivative of the position with respect to time), and a is the acceleration (the second derivative), b is the third derivative, and so on. In molecular dynamics simulations, the Verlet algorithm (**Verlet, 1967**) is the most commonly used algorithm. It is based on two Taylor expansions, a forward expansion (t+Δt) and a backward expansion (t-Δt). It uses positions and accelerations at time t, and the positions from previous time t-Δt, to calculate new positions r(t-Δt). To derive the Verlet algorithm, we can write the following relationships:

$$r(t+\Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + \cdots \quad (7)$$

$$r(t-\Delta t) = r(t) - v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 - \cdots \quad (8)$$

Summing equation (7) and (8) gives,

$$r(t+\Delta t) = 2r(t) - r(t-\Delta t) + a(t)\Delta t^2 \quad (9)$$

The Verlet algorithm has advantages as well as disadvantages, like any other algorithm. The advantages of the Verlet algorithm are, 1) position integration is quite accurate, and 2) it is independent of velocity term that makes it straightforward and reduces storage requirements. The disadvantage of the Verlet algorithm is that it lacks an explicit velocity term in the equation which makes it challenging to obtain the velocities until the positions are computed at the next step. Higher-order terms of the Taylor series could be neglected by choosing a small value for Δt. In practice Δt of about 1fs-2fs is used for performing molecular dynamics simulations of biomolecules.

To tackle the above velocity evaluation deficiencies, a variation to the Verlet algorithm have been developed. One of these variant is the leap-frog algorithm (**Hockney, 1970**), which uses the following relationships:

$$r(t+\Delta t) = r(t) + v\left(t+\tfrac{1}{2}\Delta t\right)\Delta t \tag{10}$$

$$v\left(t+\tfrac{1}{2}\Delta t\right) = v\left(t-\tfrac{1}{2}\Delta t\right) + a(t)\Delta t \tag{11}$$

The velocities at time t can be approximated from:

$$v(t) = \tfrac{1}{2}\left[v\left(t+\tfrac{1}{2}\Delta t\right) + v\left(t-\tfrac{1}{2}\Delta t\right)\right] \tag{12}$$

In this algorithm method, the velocities are calculated first, and then the positions are calculated using these velocities. Thus, the velocities leap over the positions to give their values. The advantage of the leap-frog algorithm is that it improves velocity evaluation; however, the disadvantage is that positions and velocities are not synchronized, thus it is still not satisfactory.

Another better variation of the Verlet algorithm is the velocity Verlet algorithm (**Swope, 1982**) which stores positions, velocities, and acceleration all at the same time without compromising precision.

$$r(t+\Delta t) = r(t) + v(t)\Delta t + \tfrac{1}{2}a(t)\Delta t^2 \tag{13}$$

$$v(t+\Delta t) = v(t) + \frac{1}{2}[a(t) + a(t+\Delta t)]\Delta t \quad (14)$$

The velocities at time t+Δt are approximated using:

$$v(t+\Delta t) = v\left(t + \frac{1}{2}\Delta t\right) + \frac{1}{2}a(t+\Delta t)\Delta t \quad (15)$$

The advantages of the velocity Verlet algorithm are, it provides an evaluation of velocities and kinetic energy with more accuracy, and it is numerically stable.

1.5.3 Force Fields

A force field allows potential energy (U(r)) of a system of particles to be calculated as a function of its three-dimensional structure. Force fields describe molecules are as a series of balls of variable sizes and point charges linked by bonds, angles, dihedrals, etc. Force field parameter describes the bonds, angles, and dihedrals, van der Waals and electrostatic using simple mathematical functions.

$$U_{total} = \sum_{bonds} K_b\,(b - b_0)^2 + \sum_{angles} K_\theta\,(\theta - \theta_0)^2 + \sum_{dihedrals} K_\chi\,(1 + cos(n\chi - \sigma)^2) \quad (16)$$
$$+ \sum_{nonbondedpairs,ij}\left(\varepsilon_{ij}\left[\left(\frac{R_{min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{min,ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_o \varepsilon r_{ij}}\right)$$

The first three terms in equation (16) represent bonded energy terms: bond, angle, and dihedrals, respectively (**Figure 1.13**). The first and second terms in equation (16) are modeled using the harmonic potential that gives an increase in the energy as bond and angle values change from the equilibrium values b_0 and θ_0. K_b, K_θ denote the force constants associated with the bond and angle, respectively. The third term in the equation (16) represents a torsional potential that models the rotation of dihedral where, K_χ = force constant, χ = dihedral angle, n = periodicity, σ = phase. Force field parameters corresponding to bond, angle and dihedral terms are K_b, K_θ K_χ, χ, n, σ. The fourth term represents the contribution of non-bonded interactions that is the sum of the three parts: a repulsive part preventing atoms from pervading at very short distances that varies as r^{-12}; an attractive dipole interaction part (van der Waals) that varies as r^{-6}; and a coulomb electrostatic part (repulsive/attractive depending on whether the sign of partial charges q_i

and q_j). Force field parameters, $R_{min,ij}$, ε_{ij} of the Lennard-Jones potential are the equilibrium distance and the corresponding interaction potential between atom i and j. This fourth term is calculated between all pairs of atoms (i and j, where $i \neq j$ and $i > j$) that are separated by at least three atoms or bonds. Force field combines two main components: the potential energy function and the parameters. The force field parameters are derived from experiments or quantum mechanics calculations. The total potential energy of a system is calculated by summing up the bonded terms and non-bonded terms.

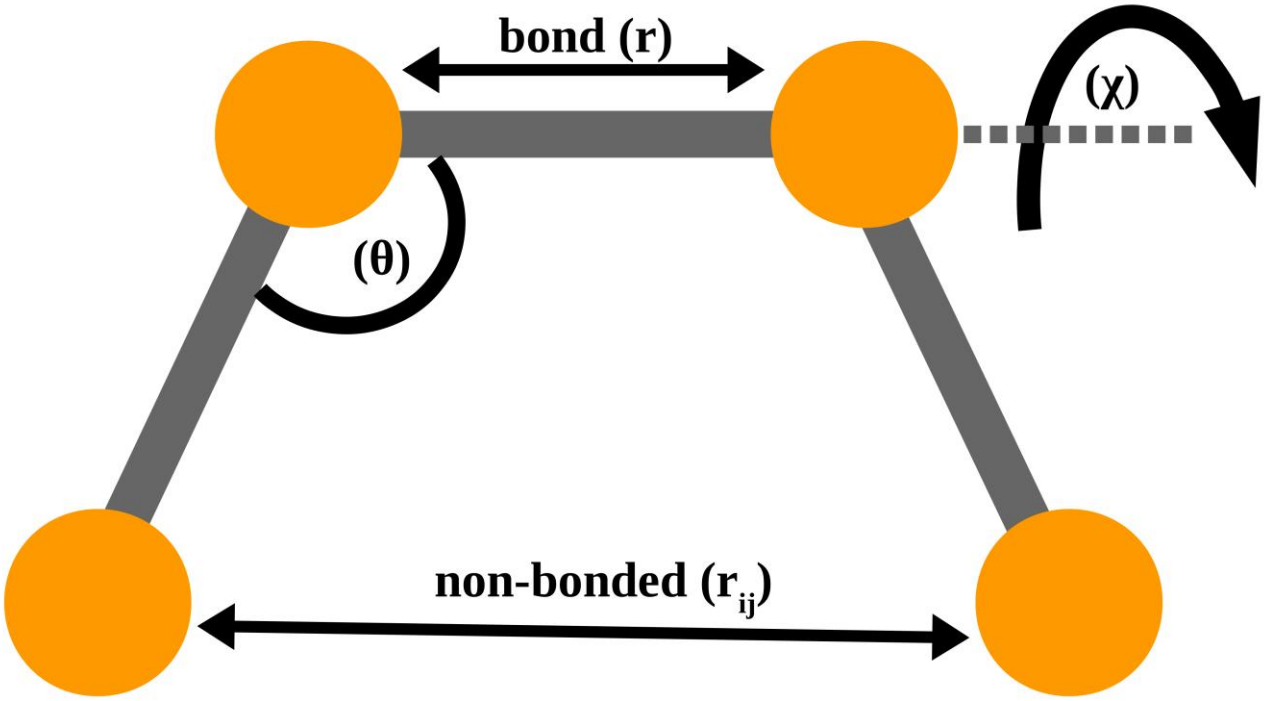


Figure 1.13: Schematic representation of bond, angle, dihedral, and non-bonded terms.

There are currently a number of empirical force fields available for the simulation of biomolecular systems such as CHARMM (**Brooks et al., 1983**), OPLS (**Jorgensen et al., 1996**), AMBER (**Ren & Ponder, 2002**) and GROMOS (**Scott et al., 1999**). In this thesis work, CHARMM22 (**MacKerell et al., 1998, MackKerell & Banavali, 2000**) and CHARMM36 (**Huang & MacKerell, 2013**) force fields have been used.

1.5.4 Water Models

Water was considered as a solvent in our MD simulations. Many different types of models (**Jorgensen, 1981; Jorgensen et al., 1983; Berendsen et al., 1987; Mahoney & Jorgensen, 2000**) have been developed; classified by the number of interactions points, whether the model is flexible or rigid, and whether the model includes polarization effects. Onufriev and Izadi have recently provided a comprehensive review with the emphasis on the conceptual basis on different classes of water models widely used in atomistic simulations, including popular implicit and explicit water models and accuracy challenges faced by them (**Onufriev & Izadi, 2017**). The simplest water models treat the water molecule as rigid by constraining the bond lengths using the SHAKE algorithm (**Armstrong, 1998; Ryckaert et al., 1977**). These models rely only on nonbonded interactions, and these interactions are modeled using Coulomb's law and Lennard-Jones potential term. The TIP3P model uses three sites for electrostatic interactions which correspond to the three atoms of the water molecule. The point charges on the hydrogen atoms are balanced by point charge assigned to an oxygen atom, the intermolecular interaction between two water molecules are approximated using Lennard-Jones parameters assigned to only oxygen atom per molecule. The TIP3P model used in this thesis is the most widely used model in MD simulations because of its simplicity.

1.5.5 Boundary Conditions

Molecular dynamics simulations are usually performed in a spherical droplet or in a box of water. In the case of a water box simulation, periodic boundary condition (PBC) is used. Keeping the biomolecule at the centre a finite water-box is overlaid. The box is then replicated in three dimensions. Periodic boundary conditions in two dimensions are shown in **figure 1.14**, where each box is surrounded by eight neighbours. Periodic boundary condition also ensures that if a particle moves outside the central cell during the simulation then it is compensated by introducing an imaged particle on the opposite side; keeping the number of particles constant in the central cell, as illustrated in **figure 1.14**. The three-dimensional cell shape is filled by translating five possible cell shapes- the cube, the hexagonal prism, the truncated octahedron, the orthorhombic and the rhombic dodecahedron. The simplest is a cubic cell.

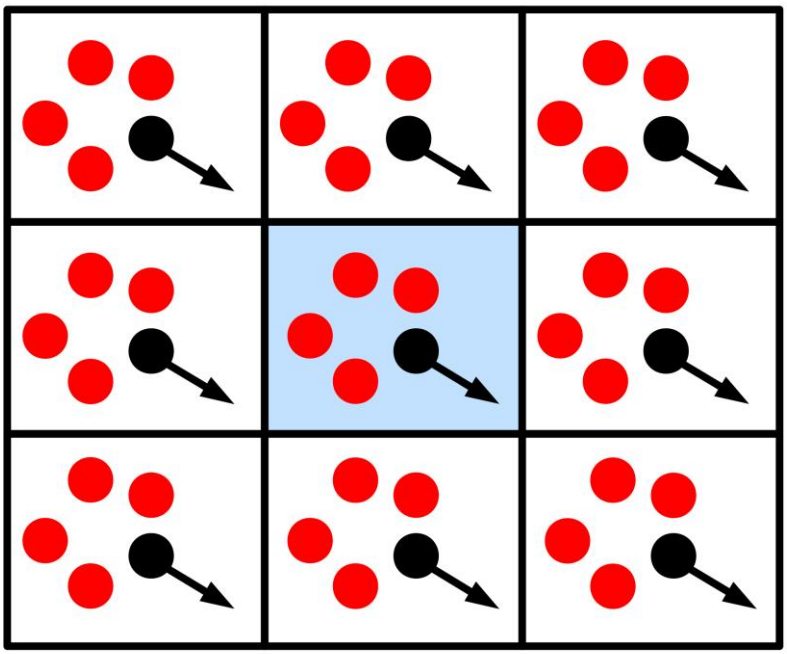


Figure 1.14: Schematic of periodic boundary conditions in two-dimension. The central simulation box (blue) is copied in every direction.

Another popular approach is to perform MD simulations of the biomolecule in a droplet of water (spherical boundary conditions). This is advantageous as no periodicity is imposed on the system and all interactions within the spherical droplet can be accurately calculated, using multipole expansions for long-range electrostatics (**Lee &Warshel, 1992**). Water molecules present at the spherical droplet boundary were subjected to radial and polarization restraints as described in the popular SCAAS model (**Marelius et al., 1998; King & Warshel, 1996**). In this thesis, we have primarily used PBC but simulations were also done with spherical droplets for ensuring convergence.

1.5.6 Long-range Electrostatic Interaction

Coulomb interactions decay as r^{-1}, thus called as long-range interactions. Calculation of Coulomb interaction energies is the most expensive part of a simulation. Different techniques have been developed in order to reduce the computational cost. In this work, we consider two different approaches; (1) Particle Mesh Ewald (PME) methods (**Darden,**

York, and Pedersen, 1993) for simulations with periodic boundary conditions and (2) Local reaction field multipole expansion method (**Lee &Warshel, 1992**) for spherical droplet simulations. PME method is a numerical approximation of the Ewald summation method (**Ewald, 1921**). The potential is divided into two parts. A real part computed in the real space. Another computed in a Fourier space, called Fourier part. The real part sum is short-ranged, it can be truncated (only if sufficiently decayed) or calculated fully without truncation. On the other hand, the Fourier part is long-ranged but smooth and periodic. The sum is calculated by utilizing a fast Fourier transforms (FFT) algorithm. The PME is a faster algorithm than Ewald summation on medium to large size systems, and it scales as *Nlog(N)*. This algorithm substantially reduces the computational time; however, the speed and accuracy depend on the grid size and interpolation scheme.

For spherical droplet simulation setup, all interactions within the droplet are calculated accurately, either directly or using multipole expansions for long-range electrostatics method (**Lee &Warshel, 1992**).

1.5.7 Short-range Van der Waals Interactions

Van der Waals interactions are short-range (r^{-6} dependence), thus decay at large distances. Thus, to reduce the computational cost, the cut-off for LJ potential is the usual choice for MD simulations. A cutoff function abruptly truncates the potential energy of non-bonded interaction. In order to truncate the nonbonded interactions gradually as a function of distance, MD simulations often use switching function which truncates the potential energy smoothly. It can be expressed as:

$$U_{switch}(r) = U(r), r < r_i \tag{17}$$

$$U_{switch} = U(r)\left(C_0 + C_1\frac{r-r_i}{r-r_c} + C_2\left(\frac{r-r_i}{r-r_c}\right)^2 + \cdots\right), r_i \le r \le r_c \tag{18}$$

$$U_{switch}(r) = 0, r > r_c \tag{19}$$

Where r_c is the cutoff distance and r_i is the distance at which the switching function is activated. The polynomial energy term in equation (18) gradually vanishes as $r_i \rightarrow r_c$. At

distances $r > r_c$, non-bonded energy is taken as zero. The value of switch off distance must always be less than that of cutoff, which is usually taken about 2Å less than the cutoff.

1.5.8 Model of Biomolecules

The size of the simulation sphere or box can be made sufficiently large to take into account the full biomolecule of our interest. But, it is disadvantageous due to the large computational cost. Thus, we first spherically cut out a particular region of interest, typically about 50-80 Å in diameter (**Figure 1.15**) and subject the model to MD simulations after overlaying water box/droplet. Spherically truncated models of biomacromolecules are advantageous not only in terms of computational cost but also improves the convergence of the calculated free energies. This approach is useful if the goal is not to sample conformational motions away from the region of interest, but to estimate converged free energies. Considering truncated biomolecular models, our target is to correctly sample the structures around an experimentally determined minima (X-ray structure) of the active site and avoid sampling distal larger conformational motions. Previously, a reasonable convergence of free energy estimates have been achieved with this strategy (**Trobro & Åqvist, 2005; Almlof et al., 2007; Allaner & Nilsson, 2011; Keranen et al., 2014; Satpati et al., 2014**). Truncated biomolecule is divided into two regions. The first region, as called reaction region, contains all atoms or groups of the active site which are subjected to MD simulation without any restraint (see **Figure 1.15**). Another region often called the buffer region, contains all atoms between the truncated surface and reaction region. The atoms in the buffer region are restrained to their x-ray determined positions using a harmonic spring.

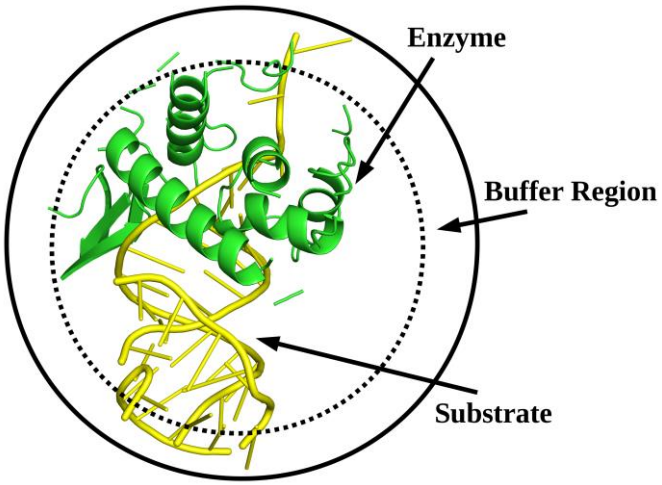


Figure 1.15: Spherically truncated biomolecular model (from PDB: 3WQY) considered in our simulation.

1.5.9 Energy Minimization

A ball placed on a curved surface will minimize its potential energy by rolling down along the curvature of the surface until it reaches the valley. Minimization algorithms follow the same procedure to find a minima of a macromolecular system for a given potential energy function. Experimentally determined (X-ray, Cryo-EM, NMR, etc) structures might be associated with local strains (e.g, van der Waals overlap) due to poor resolution. Thus, biomolecules subjected to minimization will relieve the strain in the structure by lowering the potential energy. The potential energy hypersurface usually has multiple local minima's due to the large degrees of freedom associated with the biomolecules. Minimization algorithms used in studying biomolecules usually rely on the "downhill" approach, thus they do not cross the energy barriers and end up in the local minima close to the initial point where the minimization process started. Different minimization algorithms have been developed to find a local minima, viz, using first derivative (the slope) or the second derivative (the curvature) of the energy function. Three popular minimization methods are briefly described below.

(a) Steepest Descent. Steepest descent (SD), is certainly not the most efficient algorithm, but it is easy to implement, follows the direction of the net force calculated at each iteration. This means that once the net force is calculated at the current position, the minimization step moves in the direction of the gradient,

$$\overrightarrow{s_k} = -\overrightarrow{g_k} = -\nabla_i U(r) \qquad (20)$$

To compensate lack of curvature information, this algorithm adjusts the step size (λ_k) at each iteration depending on the energy of the new conformation. Step size increases if the energy of the final conformation is lower than the previous conformation, otherwise it decreases the step-size. SD is useful in removing bad contacts in an initial structure quickly.

(b) Conjugate gradients. The conjugate gradients (CG) method uses gradients of previous as well as current iteration. The first step is taken along the direction of force as shown in SD. But from the second step onwards (k >1, k is iteration), the direction of minimization step is calculated as a weighted average of current gradient and the direction taken in the previous iteration, as

$$\overrightarrow{s_k} = -\overrightarrow{g_k} + b_k\overrightarrow{s_{k-1}} \quad (21)$$

Where b_k is weight factor, which is calculated as,

$$b_k = |g_k|^2/|g_{k-1}|^2 \quad (22)$$

(c) Newton-Raphson. The Newton-Raphson method uses both first and second-order derivative to locate a local minima. For a one-dimensional condition, assuming that

$$F(X) = a + bX + cX^2 \quad (23)$$

, the Newton-Raphson minimization can be calculated as

$$S = X - \{F'(X)/F''(X)\} \quad (24)$$

where S is the minimum, X is the current position, and F' and F" are the first and second derivatives at the current position. For a purely quadratic function, this algorithm finds the minimum in a single step from any point on the surface. For a multidimensional function:

$$\overrightarrow{s_k} = -{H_k}^{-1}\overrightarrow{g_k} \quad (25)$$

To eliminate the need to calculate the full Hessian matrix to save computational resources, a number of variations on the Newton-Raphson method have been developed. Adopted basis Newton-Raphson (ABNR) algorithm is a popular method of this class and used in this thesis.

1.5.10 Protocol of Molecular Dynamics Simulation of Biomolecules

There are many steps involved in performing a molecular dynamics simulation, and these steps altogether called- simulation protocol. The commonly used protocol for performing MD simulations is as follows:

(a) Preparation. Experimentally determined structures (obtained from Protein Databank) of biomolecules are often considered as a model for setting up MD simulations. Then hydrogen

atoms are added and solvated with water. The resulting structure undergoes minimization to remove the bad contacts.

(b) Heating up. Velocities are gradually scaled to obtain the desired temperature through a series of short MD simulations. Initial velocities are randomly assigned from Maxwell velocity distribution (correspond to a low temperature) to the atoms of the simulation system and the system is allowed to undergo dynamic relaxation. Velocities were then scaled to higher temperatures followed by molecular dynamics. The slow heating in a combination with MD continues until the desired temperature is reached.

(c) Equilibration. After attaining the desired temperature, a long dynamic simulation is performed to ensure that the different properties such as structure, pressure, temperature and the energy of the system are stable with respect to time and free of inconsistent fluctuations.

(d) Production phase. After equilibration, the MD trajectory is stored for analysis (production phase). The production run can be from several hundred picoseconds to nanoseconds and more.

(e) Analysis. When carrying out an MD simulation, it is better to plan the analysis procedure before starting the simulation. Quantities that are commonly analyzed from an MD trajectory include:

(i) Average structural properties (viz, distances, angles, dihedrals, water-distribution, etc.) from MD trajectories.

(ii) Radius of gyration (Rg)

The radius of gyration of a biomolecule is defined as the root mean square distance of the biomolecules atoms/residues from its centre of mass. It describes the overall spread or compactness of a biomolecule. For a n particle biomolecule R_g can be calculated as :

$$R_g = \sqrt{\sum_{i=1}^{n} m_i r_i^2 / \sum_{i=1}^{n} m_i} \qquad (26)$$

where m_i is the mass of i^{th} particle and r_i is the distance of i^{th} particle from center of mass.

(iii) Root mean square deviation (RMSD):

RMSD is the measure of the difference between a target structure and a reference structure (or initial structure) over time. It is generally used as a quantitative measure of structural deviation from the reference.

$$\left\langle \left(r_i^{\alpha} - r_i^{\beta} \right) \right\rangle^{1/2} = \sqrt{\frac{1}{N} \sum_i \left(r_i^{\alpha} - r_i^{\beta} \right)^2} \quad (27)$$

(iv) Root mean square fluctuation (RMSF):

RMSF is the measure of fluctuation of an atom or residue about a well-defined average position over time, it is a numerical measurement of individual residue flexibility.

$$\sqrt{\frac{1}{N_f} \sum_f \left(r_i^f - r_i^{ave} \right)^2} \quad (28)$$

(v) SASA (Solvent accessible surface area):

SASA is the surface area of a biomolecule that can be accessed by the solvent. The accessible surface is calculated by tracing the center of a probe sphere (typically 1.4 Å radius) rolling over a molecule's van der Waals surface (or vdw radii of atoms).

1.5.11 Thermodynamic Cycle and Relative Binding

Binding of two different ligands to a protein (receptor) could be described by an appropriate thermodynamic cycle (**Figure 1.16**). Vertical legs correspond to the ligand-protein binding while horizontal legs correspond to the alchemical transformation of A into B either in the complex (above) or in the free state (below). Horizontal legs cannot be realized experimentally, however, the free energy changes along these unphysical legs can be computed with reasonable accuracy using computational simulations with enormous flexibility (**van Gunsteren et al., 2002; Steinbrecher & Labahn, 2010**). Calculation of free energy changes (**Jorgensen et al., 1988**) along the vertical legs (absolute binding free energy) is limited by sampling and convergence issues and thus comparatively more difficult to estimate computationally. The free energy of ligand binding (vertical legs) often provided by experimental studies (solvation, binding). Thus, the calculation of free energies associated with the horizontal alchemical pathway is a choice of this thesis for estimating relative binding free energies. Relative binding free energy is calculated by taking the advantage that free energy is a state function, and so its value around a closed thermodynamic cycle path must be zero.

$$\Delta\Delta G = \Delta G_{bind}^{B} - \Delta G_{bind}^{A} = \Delta G^{comp} - \Delta G^{free} \quad (29)$$

The use of an appropriate thermodynamic cycle (**Tembe & McCammon, 1984**) can also describe the binding of a single ligand to two related proteins (Wild-type vs mutant). In this thesis thermodynamic cycle was used extensively to estimate relative binding free energies (biomolecular recognition) and establish its link to biomolecule structures.

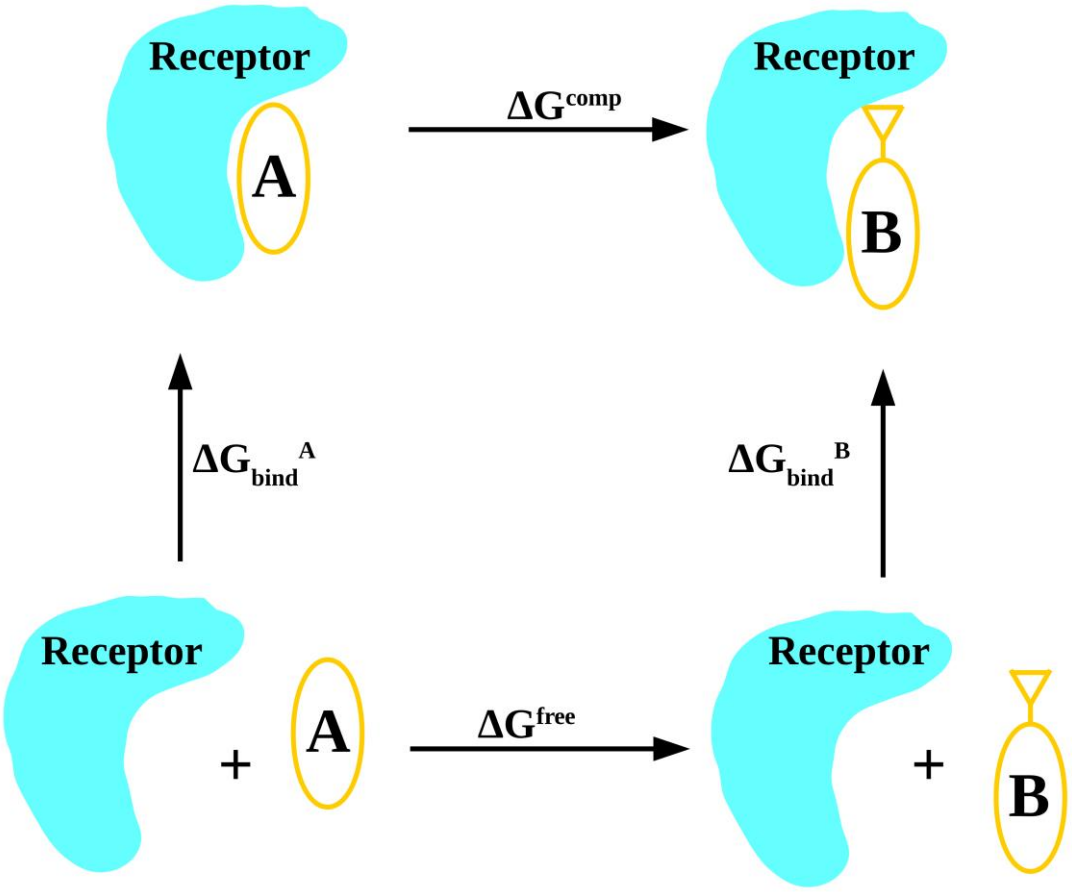


Figure 1.16: Thermodynamic cycle used to calculate relative binding free energies of ligand-receptor binding. Here, vertical legs correspond to the binding, while horizontal legs correspond to the alchemical transformation either in the complex (upper arm) or in the free state (lower arm).

1.5.12 Relative Free Energy Calculations

The free energy difference between states A and B is related to the ratio of the partition function of state A and B (**equation 29**). Here, $\beta = 1/K_BT$, where K_B is the Boltzmann constant, T is the temperature. The partition function is an integral over all the conformation available for a particular state. A reliable estimate of the partitional function

of a state (say A or B) is limited by (1) finite sampling, and (2) Low Boltzmann weight associated with the conformations sampled by MD simulation. Thus practical calculation always calculates the ratio of the partition function of free energy differences (**equation 29**).

$$\Delta G_{A\to B} = G_B - G_A = \frac{-1}{\beta} ln\left(\frac{Q_B}{Q_A}\right) \quad (29)$$

In this thesis, free energy changes along the horizontal leg of the thermodynamic cycle (**Figure 1.16**) was estimated by two statistical mechanical approaches: Free energy perturbation (FEP) and Thermodynamic integration (TI).

1.5.12.1 Free Energy Perturbation (FEP)

Equation 29 can be rearranged and expressed as an ensemble average of MD trajectory of A:

$$\Delta G_{A\to B} = \frac{-1}{\beta} ln\langle exp\{-\beta[H_B - H_A]\}\rangle_A \quad (30)$$

Equation 30 is known as Zwangzig's exponential formula (**Zwangzig, 1954**). The convergence and accuracy of estimated free energy ($\Delta G_{A\to B}$) strictly depend on the conformational overlap of A and B simulation. Good overlap results in accurate estimation of $\Delta G_{A\to B}$. Thus, to obtain appropriate phase space overlapping between two adjacent states (A and B), intermediate strata (windows) are inserted between A and B. Hamiltonian is scaled in each intermediate state by the general coupling parameter λ_i, where i is number of intermediate states in between A and B. Scaling parameter λ is such that ($0 \leq \lambda \geq 1$) where 0 is an initial state (A) and 1 is the final state (B). The energy change for A→B is estimated by summing over the finite changes as described in **equation 31**:

$$\Delta G_{A\to B} = \frac{-1}{\beta} \sum_{i=1}^{N} ln\,\langle exp\{-\beta[H(\lambda_{i+1}) - H(\lambda_i)]\}\rangle_i \quad (31)$$

1.5.12.2 Thermodynamic Integration (TI)

Thermodynamic integration (**Kirkwood, 1935**) is another approach, where the free energy derivative with respect to a coupling coordinate is calculated. The free energy derivates

$(\frac{\partial G(\lambda)}{\partial \lambda})$ is equal to the ensemble average of the energy derivative $\left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda}$. Simulation is carried out at discrete λ values, and the free energy derivative by numerically integrating **equation 32**.

$$\Delta G = \int_0^1 \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \qquad (32)$$

1.5.13 Software Used in This Thesis

CHARMM (**Brooks et al., 1983; Brooks et al., 2009**), NAMD (**Phillips et al., 200**5) and Q (**Marelius et al., 1998**) was used to perform MD simulation. Molecular graphics programs (VMD (**Humphrey et al., 1996**), PyMol (**The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.**), and CHIMERA (**Pettersen et al., 2004**) were used to visualize and analyse MD trajectory. Grace plotting tool (**ftp://plasma-gate.weizmann.ac.il/pub/grace/**) was used for plotting energies and structural parameters. Unix shell scripting was used extensively for post-processing the MD trajectories to extract free energies and structural parameters.

1.6 Objectives of the Thesis

1. How AlaRS recognizes $tRNA^{Ala}$ based on 3•70 base pair away from the anti-codon stem-loop of RNA (Cognate: G3•U70, Near-cognate: A3•U70, G3•C70, A3•C70)? Propose a kinetic scheme of tRNA selectivity consistent with the experiment.
2. Understand the principle of stop codons recognition by eukaryotic release factor (eRF1) relative to sense codons in the ribosome.
3. Estimate the energetics of viral RNA recognition by RIG-I. (a) How RIG-I selectively binds to dsRNA containing viral features (5'-tri/diphosphate) relative to host dsRNA (containing 5' monophosphate/hydroxyl group). How the estimated selectivity is linked to the structures of dsRNA bound RIG-I. (b) Why RIG-I discriminate between host dsRNAs (5' p is worst binder than 5' OH), a feature of RIG-I like family proteins.

Chapter 2

Alanine tRNA Recognition by Alanine tRNA Synthetase

Throughout evolution the presence of a single G3•U70 mismatch in the acceptor stem of $tRNA^{Ala}$ is the major determinant for aminoacylation with alanine by alanyl-tRNA synthetase (AlaRS). Recently reported crystal structures of the complexes AlaRS-$tRNA^{Ala}$/G3•U70 and AlaRS-$tRNA^{Ala}$/A3•U70 suggest two very different conformations, representing a reactive and a non-reactive state, respectively. Based on these structures, it has been proposed that the G3•U70 base pair guides the –CCA end of the tRNA acceptor stem into the active site of AlaRS, thereby enabling aminoacylation. The crystal structures open up the possibility of directly computing the energetics of tRNA specificity by AlaRS. We have carried out MD free energy simulations to quantitatively estimate tRNA discrimination by AlaRS, focusing on mutations of the single critical base pair G3•U70 to uncover the energetics underlying the accuracy of tRNA selection. The calculations show that reactive complex is highly selective in favor of cognate $tRNA^{Ala}$/G3•U70 over its non-cognate analogues (A3•U70/G3•C70/A3•C70). In contrast, the non-reactive complex is predicted to be unselective between $tRNA^{Ala}$/G3•U70 and $tRNA^{Ala}$/A3•U70. Utilizing our calculated relative binding free energies, we show how a simple three-step kinetic scheme for aminoacylation, involving both an initial non-specific binding step and a subsequent transition to a selective reactive complex, accounts for the observed kinetics of the process.

==========

The work published in ACS Omega, DOI:10.1021/acsomega.9b01827

2.1 Background

Correct amino acid attachment to its cognate tRNA followed by correct mRNA-tRNA interaction during mRNA decoding on the ribosome ensures the accuracy of genetic code translation. An aminoacyl-tRNA synthetase (aaRS) selects its cognate tRNA and ligates the correct amino acid at the 3'-CCA end, as described in **Chapter 1** (**Figure 1.1**). For alanine tRNA ($tRNA^{Ala}$) a single wobble G3•U70 base pair in the acceptor stem away from the triplet anticodon (**Chapter 1, Figure 1.2**) is the major determinant of the tRNA identity (**Hou & Schimmel, 1988; McClain & Foss, 1988; deDuve, 1988**) and is recognized by AlaRS in all kingdoms of life (**Hou & Schimmel, 1989**). As stated in **Chapter 1**, mutation of the G3•U70 pair thus diminishes the aminoacylation with alanine by AlaRS (**Park et al., 1989**).

As discussed in **Chapter 1**, crystal structures of AlaRS from *Arachaeoglobus fulgidus* in complex with native $tRNA^{Ala}$ ($tRNA^{Ala}$/G•U) and the mutated variant with A3•U70 ($tRNA^{Ala}$/A•U) (**Naganuma et al., 2014**), in complex with the alanyl-adenylate analogue, 5'-O-[N-(L-alanyl)sulfamoyl] adenosine (Ala-SA) (**Ueda et al., 1991**), have recently been reported (**Figure 2.1, a**). These structures suggested an explanation to the long-standing mystery of the strict specificity of AlaRS for $tRNA^{Ala}$ containing the conserved G3•U70 wobble pair (**Naganuma et al., 2014**). In the AlaRS-$tRNA^{Ala}$/G•U complex the 3'-CCA end of $tRNA^{Ala}$ has reached into the active site (reactive complex: R) where aminoacylation can occur, whereas in the AlaRS-$tRNA^{Ala}$/A•U complex, 3'-CCA end of $tRNA^{Ala}$ folds back into a different route away from the active site (non-reactive complex: NR) (**Figure 2.1, a**). Flexible domains allow proteins to access multiple conformational states and the highly mobile loop region of AlaRS that spans the minor groove of the acceptor stem (residues 470-490 with an average backbone B-factor ~ 195 $Å^2$) (**Naganuma et al., 2014**) might have a prominent role in selectivity (**Figure 2.1, a**). An interesting feature of the G3•U70 wobble geometry is the availability of the exocyclic $–NH_2$ group of G3 in the minor groove, and O4 of U70 in the major groove, for hydrogen bonding interaction with the protein (see **Figure 1.4**). As discussed in **Chapter 1**, AlaRS is thus able to establish a specific interaction network with G3•U70 both in the major and minor grooves to recognize the tRNA for alanine acylation. Other base pairs such as A3•U70, G3•C70 and A3•C70, change the shape of major/minor grooves and alter the mode of stacking and H-bonding (**Figure 2.1, b**).

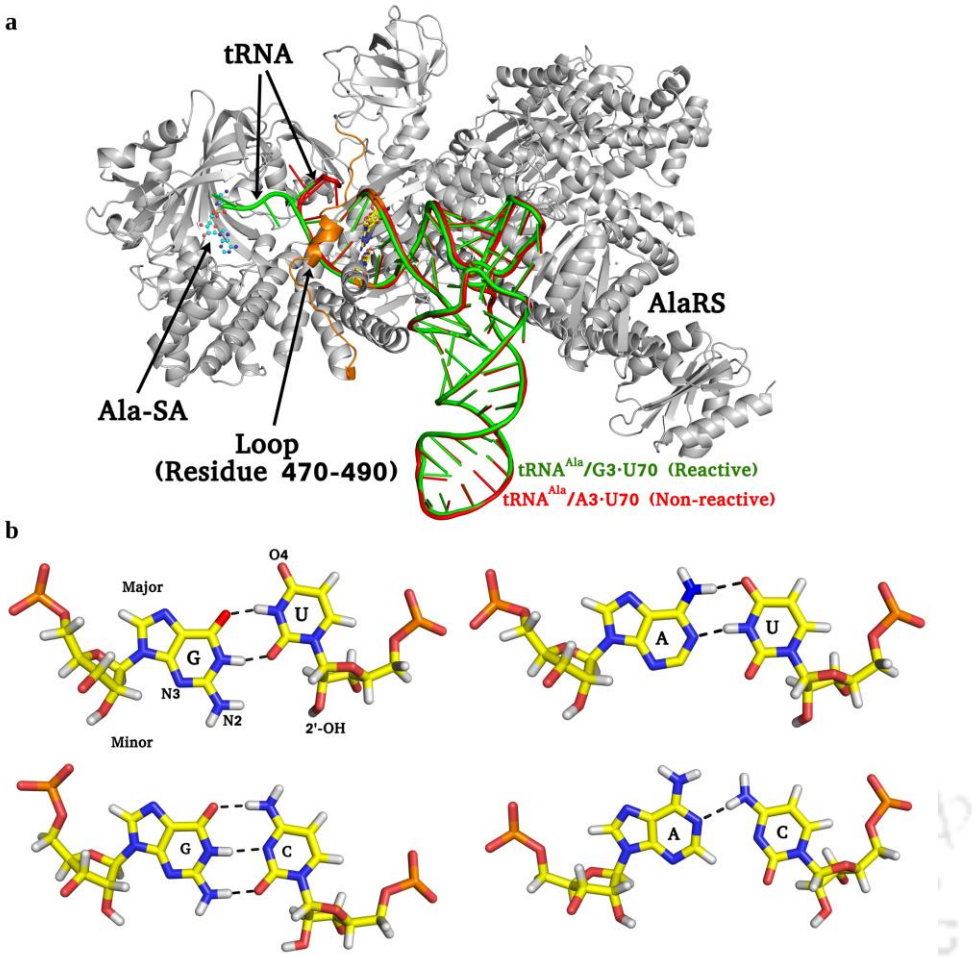


Figure 2.1. AlaRS-tRNA^{Ala} complex, identity base pair and its variants. (a) Reactive (green) and nonreactive (red) conformation of tRNA^{Ala} in complex with AlaRS (gray) and the identity base G3•U70 (yellow stick) at the acceptor stem; Ala-SA (ball and stick) and flexible loop (residue 470-490) shown in orange color. (b) close-up view of G•U (major, minor groove highlighted), A•U, G•C and A•C base pairing.

The crystal structures show that the interaction between AlaRS and the major and minor grooves of the G3•U70 base pair is disrupted in the A3•U70 variant (**Naganuma et al., 2014**). It has also been suggested that the geometrical difference between G3•U70 and A3•U70 propagates to the 3'-CCA end of tRNA^{Ala}, resulting in distinctly different orientations of the acceptor stem, to yield the reactive conformation for tRNA^{Ala}/G•U and the non-reactive one for tRNA^{Ala}/A•U. The proposed mechanism gives a clue to how a small difference in sequence away from the anticodon may result in tRNA^{Ala} specificity at the k_{cat} level (**Naganuma et al., 2014**). The ratio of k_{cat} for aminoacylation between the G3•U70

and A3•U70 variants of $tRNA^{Ala}$ is approximately 100-fold (**Naganuma et al., 2014**). Interestingly, many other proteins similar to the editing domain of class II aaRSs have also been isolated, which can edit misacylated tRNA's even after they have been released from the synthetase (**Schimmel & Ribas De Pouplana, 2000; Ahel, et al., 2003; Korencic et al., 2004; Ling et al., 2009**).

The difference between the R and NR complexes is mainly characterized by the alternate orientations of the 3'-CCA end of $tRNA^{Ala}$ since the AlaRS structure is very similar in the two medium resolution (~3.4 Å) X-ray structures (**Naganuma et al., 2014**). However, several key issues regarding the $tRNA^{Ala}$ specificity of AlaRS remain unclear. For example, replacement of G3•U70 by A3•U70 could perhaps already change the conformational preference of the CCA end (NR over R) of free $tRNA^{Ala}$ in water, leading to the NR complex upon AlaRS binding. Another question is how strongly the cognate $tRNA^{Ala}$ G3•U70 wobble pair is preferred with respect to its variants (A3•U70, G3•C70, A3•C70) by AlaRS in the R conformation and whether it is also preferred in NR. Note, the structures of non-cognate complexes such as those containing G3•C70 or A3•C70 have not yet been resolved experimentally. The near-atomic resolution crystallographic complexes (**Naganuma et al., 2014**), however, now provide sufficiently good models for structure-based computational evaluation of the energetics associated with the states they are trapped in (R and NR). Here, in this chapter, we report molecular dynamics (MD) free energy calculations of cognate and near-cognate AlaRS-$tRNA^{Ala}$ complexes to examine how the energetics of tRNA binding is affected by different variants of the 3•70 base pair in the reactive and nonreactive conformations. The quantitative estimation of $tRNA^{Ala}$ specificity by AlaRS offers a simple view of how fidelity in the aminoacylation process is achieved, thereby establishing the link between 3D structures, thermodynamics, and experimentally measured kinetics.

2.2 Methods

2.2.1 Molecular Dynamics Procedure

The MD protocol adapted for this work is described in **Figure 2.2**. The structures of $tRNA^{Ala}$ bound to AlaRS, in the reactive and nonreactive conformations, were taken from the Protein Data Bank (pdb codes 3WQY and 3WQZ, with crystallographic resolution of

3.3 Å and 3.49 Å, respectively) (**Naganuma et al., 2014**). Spherical systems with radius 25Å, centered on the N9 atom of the G3/A3 nucleotide of $tRNA^{Ala}$, were cut out from the crystallographic structures and used for MD simulations. Harmonic positional restraints to experimental coordinates were then applied to all heavy atoms in the "buffer region" between 22 and 25 Å from the sphere center. The force constants of restraints in the buffer region were increased linearly between 3.0 to 5.0 kcal/mol/$Å^2$ in the direction towards the outer boundary. The inner 22 Å radius sphere was treated as fully flexible in the production MD simulations. This system was solvated by overlaying a cubic water box with edge length 80 Å, where waters which overlapped with AlaRS-tRNA were removed. This yielded a total of about 49000 atoms in the simulations, which included ~14900 water molecules.

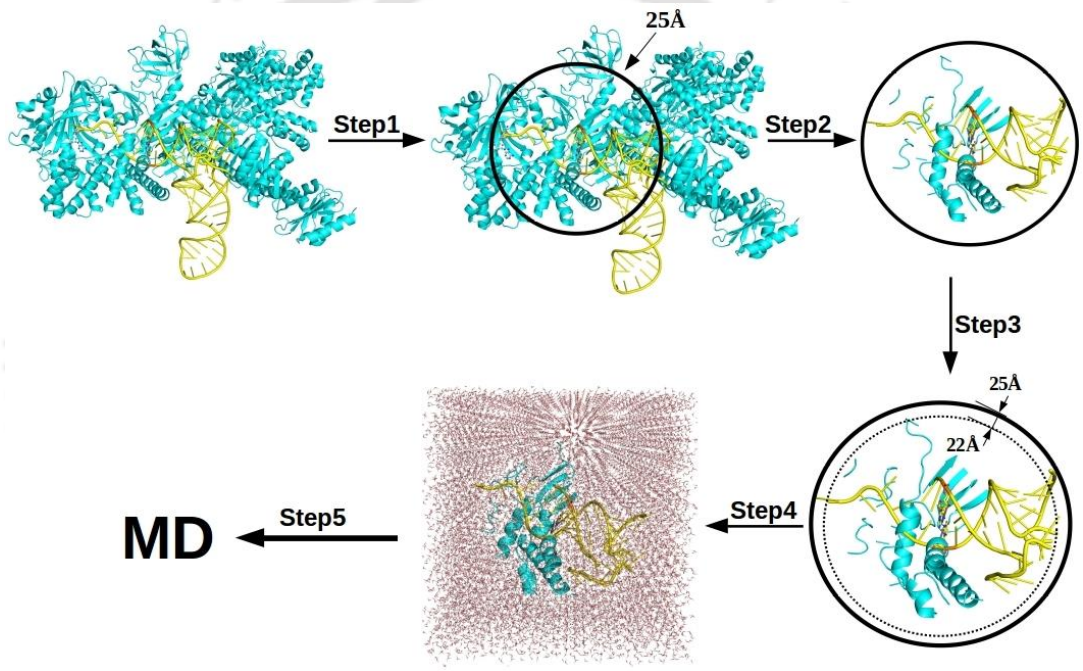


Figure 2.2: MD setup. Step1: Selection of 25Å radii sphere centred at N9 of G3/A3 (resid 1603) nucelotide of $tRNA^{Ala}$ bound Ala-Synthetase complex (PDB:3wqy, 3wqz). Step 2: Truncated model (25Å sphere) for molecular dynamics simulations. Step 3: Heavy atoms of the "buffer region" (22Å-25Å) are harmonically restrained to their experimentally determined positions. Step4: Overlay water box of edge length 80Å, equilibrate and proceed for MD.

Periodic boundary conditions (using the Particle Mesh Ewald method (**Darden et al., 1993**)) for electrostatics with tinfoil boundary conditions (**Hummer at al., 1996; Bogusz et al., 1998**) were used for running MD. Van der Waals interactions were truncated with a 16 Å

cutoff. Temperature and pressure were kept at 310K and 1 bar. Langevin dynamics (**Feller et al., 1995**) for non-hydrogen atoms with a coupling coefficient of 5 ps^{-1} were used to control temperature. Pressure was controlled by a Langevin piston using Nose-Hoover method (**Martyna et al., 1994**). The CHARMM36 force field (**Best et al., 2012; MacKerell & Banavali, 2000**) and TIP3P water model (**Jorgensen et al., 1983**) was used for running MD. Simulations were performed using the CHARMM (**Brooks et al., 1983; Brooks et al., 2009**) and NAMD (**Phillips et al., 2005**) programs. For each simulation model, we ran 5-10 replicas starting with different initial velocity distribution. Each replica involved 340 ps of equilibration (where system was heated up to 310K in the initial stages and then kept fixed throughout the simulations) followed by at least 5ns production dynamics. Overall, 240-360 ns of production dynamics distributed over 5-10 replicas were considered for structural analysis for the AlaRS-tRNA complex and free tRNA in water.

In the initial stage of equilibration, harmonic restraint was applied for the inner region (within 22Å) with force constant 4.0 kcal/mol/$Å^2$, and at the final stage of equilibration, the restraint (within 22Å) was completely removed. The simulation models were overall neutral, which was achieved by scaling partial charges of the phosphate backbone of tRNA. X-ray structures AlaRS-tRNA complexes does not contain counter ions within reduced 25Å truncated model. Hence, instead of placing the counter ions at random locations we mimicked the backbone neutralization by scaling down the partial charges of the RNA phosphates which are away from 3•70 pair. The charges of RNA base pair undergoing mutation (i.e, 3•70) and its nearby base pairs (2•71, 4•69 and 5•68) were not modified. The truncated 25Å model centered at the N9 of A3/G3 was neutralized by adding a +1 charge distributed over four atoms of the phosphate group of RNA residues (**Table 2.1**) present outside 13.4 Å from the center. Bond length between hydrogen and heavy atoms were constrained by ShakeH algorithm implemented in NAMD (**Phillips et al., 2005**) with allowable bond-length deviation of 1.0e-8 Å. 2fs time step was used for performing MD simulations. RMSD and RMSF of heavy atoms within 22Å of unrestrained simulation sphere were calculated with respect to x-ray structure and the average RMSD and RMSF was calculated for 7ns MD trajectory with a 1ps interval.

Table 2.1: Partial charges of RNA phosphates.

Atom Name	Charges in CHARMM36 FF	Scaled down partial charges
P	1.50	1.75
O1P	-0.78	-0.53
O2P	-0.78	-0.53
O5'	-0.57	-0.32

2.2.2 Free Energy Calculations

Relative binding free energies ($\Delta\Delta G_{bind}$) for tRNA mutations in the AlaRS-tRNA complexes (R and NR) were calculated by alchemically transforming G3/A3/U70 into A3/G3/C70 (horizontal legs of the thermodynamic cycle in **Figure 2.3**). The vertical legs of the thermodynamic cycle correspond to tRNA binding and the horizontal legs correspond to the alchemical transformation of tRNA (**Figure 2.3**). It should be noted that the horizontal paths (alchemical transformation) cannot be realized experimentally. We computed the free energy changes along the alchemical transformations (ΔG^{comp} and ΔG^{free}) and calculated the relative binding free energy as $\Delta\Delta G^{bind} = \Delta G^{comp} - \Delta G^{free} = \Delta G^{bind}_{A3\bullet U70} - \Delta G^{bind}_{G3\bullet U70}$. We used a hybrid energy function (U) which represent a mixture of two endpoint states of the horizontal leg (**Figure 2.3**), as applied in previous studies (**Simonson & Satpati, 2012; Satpati & Simonson, 2012; Satpati et al., 2011; Satapti et al., 2014**). The coupling parameter λ connects the end states by modifying the electrostatic, van der Waals energy and bonded terms. Changing λ from 1 to 0 thus transforms AlaRS-$tRNA^{Ala}$/G3•U70 into AlaRS-$tRNA^{Ala}$/A3•U70, by means of the mapping energy function: $U = \lambda\, U(G3) + (1-\lambda)\, U(A3)$. The free energy derivative was calculated as $\partial G/\partial\lambda = \langle \partial U/\partial\lambda \rangle_\lambda = \langle U(\lambda=1) - U(\lambda=0) \rangle_\lambda$ where the brackets "< >" represents averaging over MD trajectory for a particular value of λ. For the alchemical transformation of G3 into A3, we used 17 λ values between 1 and 0 (1.0, 0.999, 0.99, 0.95, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.05, 0.01, 0.001 and 0.0), while for the transformation of U70 into C70, we used 11 equally spaced λ values between 1 and 0 (1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1 and 0.0). The free energy derivative at each λ was calculated by computing the difference $\langle U(\lambda=1) - U(\lambda=0) \rangle_\lambda$. Each λ window simulation lasted for 1-2 nanosecond and the data from the last 600-1600 ps was used for averaging.

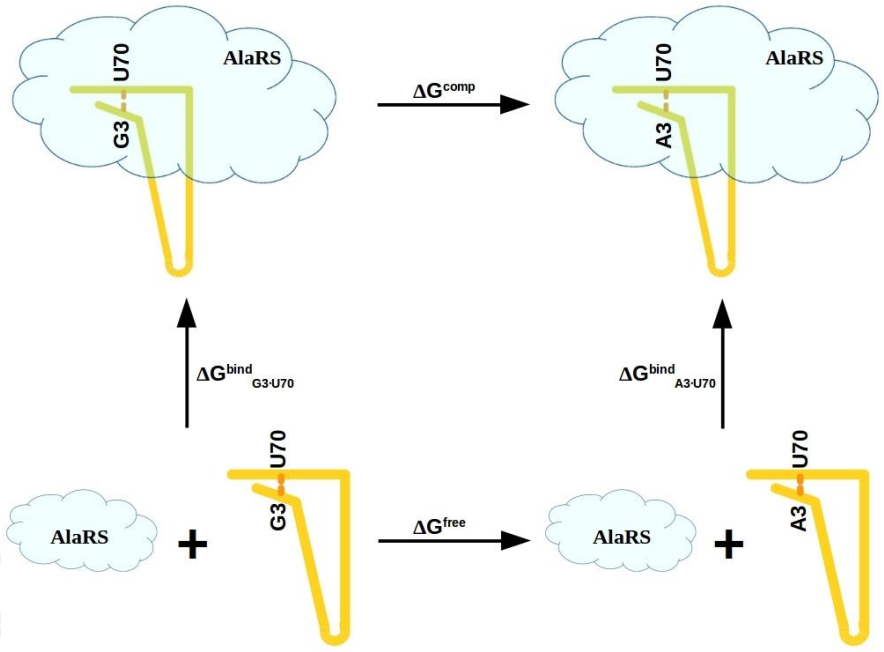


Figure 2.3. Thermodynamic cycle for AlaRS:tRNA^{Ala} binding. Vertical legs correspond to binding; horizontal legs correspond to the alchemical transformation of the identity base pair G3•U70 into A3•U70, either in the solvated protein (upper leg) or in solution (bottom leg).

Numerical integration (with standard trapezoidal method) was used for calculating the free energy change. The last 600-1600 ps of the trajectory at each λ was divided into two batches and the deviation of the batch averages was reported as an uncertainty associated with the free energy derivatives. The same is reported as the statistical error associated with calculated ΔG^{comp} or ΔG^{free} for each run. The average result obtained from 5-10 trajectories are used for computing the ΔΔG. The uncertainty in the averaged ΔG, is reported as the standard error of the mean (from 5-10 replicas) and the error in the final ΔΔG in the manuscript is calculated by propagating the standard error of mean associated with the averaged ΔG. The uncertainty in the averaged ΔG (~1.0 kcal/mol) is well within the acceptable statistical uncertainty (**Lind et al., 2019**).

Free energies were calculated for forward and reverse alchemical transformation (say G3→A3 and A3→G3) in tRNA, either in complex with AlaRS or free in water (**refer appendix Table 2A-1**). The results were averaged over the forward and backward runs,

except for reactive AlaRS-tRNA^{Ala}/G3•U70 complex (**refer appendix Table 2A-1**). A relatively large hysteresis (difference between G3→A3 and A3→G3 runs) was observed (**refer appendix Table 2A-1, a**) only in the reactive complex (~ 5 kcal/mol) and hence averaging was done based on the forward runs. Reverse alchemical transformation (AlaRS-tRNA^{Ala}/A3•U70→AlaRS-tRNA^{Ala}/G3•U70) in the reactive complex could not disrupt the salt bridge Arg483-Asp450 interaction and get back the G3-Asp450 interaction (**Figure 2.4**), which causes the hysteresis.

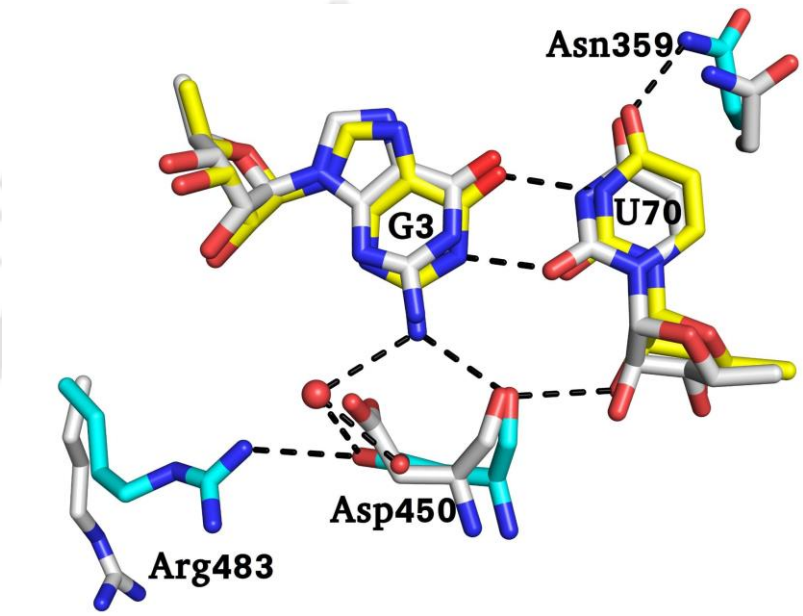


Figure 2.4. MD structure of reactive AlaRS:tRNA^{Ala}/G3•U70 complex (Gray sticks) and after reverse alchemical transformation, i.e, A3•U70→G3•U70 (yellow sticks for G3•U70 and protein residues in cyan sticks). Arg483-Asp450 salt bridge is present at the end point of the reverse alchemical transformation.

A reasonable estimate of the free energy of G→A transformation in the reactive complex is only possible to obtain from the forward runs. The reverse runs (A→G) in the reactive conformation do not end up in the x-ray structure (**Figure 2.4**), which makes it necessary to discarded these for reliable free energy estimates. In order to minimize hysteresis, we have extended MD simulations at 8 λ windows (which showed significant deviation between forward and reverse free energy derivatives) of the reverse run by at 10-20 ns each. The hysteresis was reduced from ~ 5 kcal/mol to ~2 kcal/mol, as a result of long MD simulation (distributed over 117 ns alchemical free energy simulations); however, the hysteresis was not completely eliminated. Disruption of salt bridge interaction between Asp450-Arg483

seems to be remedy in solving the hysteresis issue. Hence, we mutated Arg483→Ala483 in the reactive complex (Model R#) and performed G3•U70→A3•U70 (forward) and A3•U70→G3•U70 (reverse) and found almost no hysteresis (~ 0.6 kcal/mol) and ΔΔG= 3.3 kcal/mol. Thus, it can be argued that R# does not have hysteresis problem but still yields strong discrimination.

The free energy calculations for each system were based on 340-880 ns of MD data averaged over 5-10 replicas with different initial velocities. Overall, a total of about 2μs of molecular dynamics free energy simulations have been done to get good convergence and reasonable standard error of the mean (< 1 kcal/mol). In order to compute the binding free energy difference between G3•U70 and A3•C70, we performed direct calculations for two different paths, G3•U70→A3•U70 and A3•U70→A3•C70, and the result was then obtained indirectly from the free energies obtained from these two direct paths. As an additional check, the free energy calculations for G→A in the reactive complex (ΔG^{comp} of **Figure 2.3**) were also performed with a smaller water box with 70 Å side. The computed ΔG^{comp} are virtually identical for the two different water box sizes. The robustness of calculated energetics was examined by repeating calculations for G3•U70→A3•U70, G3•U70→G3•C70 transformations (in reactive complex, free tRNA) with different MD setup, viz, (a) spherical droplet instead of water box (b) considering explicit Na^+ ions for neutralization instead of scaled-down phosphate charges in a water box. The calculated energetics are found to be robust and independent of the box size, solvation protocol, and statistical methods employed in extracting free energies.

2.3 Results

2.3.1 Structure-based Energetics of $tRNA^{Ala}$ Selectivity in AlaRS

In order to compute the energetics of AlaRS binding specificity for $tRNA^{Ala}$ and elucidate the roles of the different conformational states (R and NR) in tRNA selection, we carried out MD simulations of $tRNA^{Ala}$/G•U and its variants ($tRNA^{Ala}$/A•U, $tRNA^{Ala}$/G•C, $tRNA^{Ala}$/A•C) in complex with AlaRS and free in water (**Figure 2.3**), using both reactive and non-reactive crystallographic structures (**Naganuma et al., 2014**) as starting points. These calculations involve computing the change in binding affinity of AlaRS for $tRNA^{Ala}$ upon G3•U70→A3•U70, G3•U70→G3•C70, A3•U70→A3•C70 mutations in both reactive

and non-reactive complex (**Figure 2.5**). The results reveal several remarkable features. First, the reactive complex strongly favors $tRNA^{Ala}$/G•U with respect to the A•U variant ($\Delta\Delta G_{bind}$ ~5 kcal/mol).

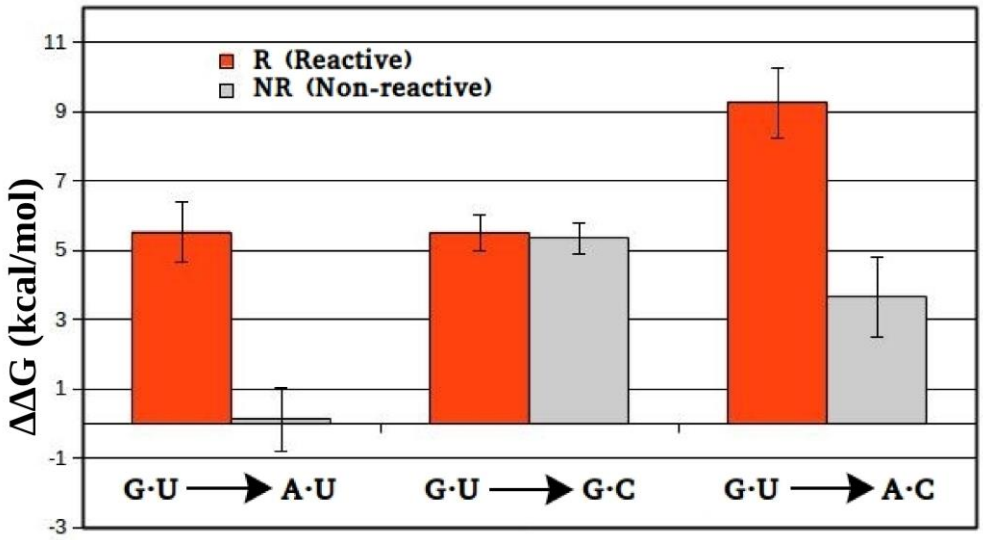


Figure 2.5. Calculated energetics of $tRNA^{Ala}$ reading in different states. Graph shows the calculated $tRNA^{Ala}$ binding free energy differences (kcal/mol) between identity base pair G3•U70 and its variants (A3•U70, G3•C70, A3•C70). Red and gray bars denote reactive and non-reactive states respectively. Binding free energies in kcal/mol and apparent statistical uncertainties shown as vertical bars (black line) in parentheses.

The non-reactive state is almost non-selective between $tRNA^{Ala}$/G•U and $tRNA^{Ala}$/A•U ($\Delta\Delta G_{bind}$ ~ 0.1 kcal/mol). Hence, the conformational change from NR to R can boost the discrimination strength (by $\Delta\Delta G_{bind}$ ~ 5 kcal/mol) between cognate $tRNA^{Ala}$/G•U and near-cognate $tRNA^{Ala}$/A•U, favoring the former. Second, both the reactive and non-reactive conformations disfavor non-cognate $tRNA^{Ala}$/G•C with equal strength of ~ 5 kcal/mol compared to the cognate $tRNA^{Ala}$/G•U. Third, the reactive complex shows the highest discriminatory power for $tRNA^{Ala}$/A•C rejection. Fourth, the difference in the discrimination strength between reactive and non-reactive complex can thus be linked with the purine (A3/G3)-AlaRS interaction.

Consideration of free unbound tRNA in water is essential for understanding tRNA binding to AlaRS. Local geometrical differences between G3•U70 and A3•U70 in the free $tRNA^{Ala}$ might alter the conformational preference (i.e, the orientation of 3' –CCA arm of tRNA

leading to reactive or non-reactive conformation) of free $tRNA^{Ala}$ in water. To understand the energetics of the conformational change in the free $tRNA^{Ala}$ we have also compared the mutation free energies for G3•U70→A3•U70 in the reactive and non-reactive conformation of free $tRNA^{Ala}$ in solution using the appropriate thermodynamic cycle (**Figure 2.6**). The very small computed $\Delta\Delta G^{NR\rightarrow R}_{free} \sim -0.2\pm0.9$ kcal/mol suggests that the geometrical difference between G3•U70 and A3•U70 does not play any significant role in driving the conformational change (NR ⇆ R) in free $tRNA^{Ala}$ in water.

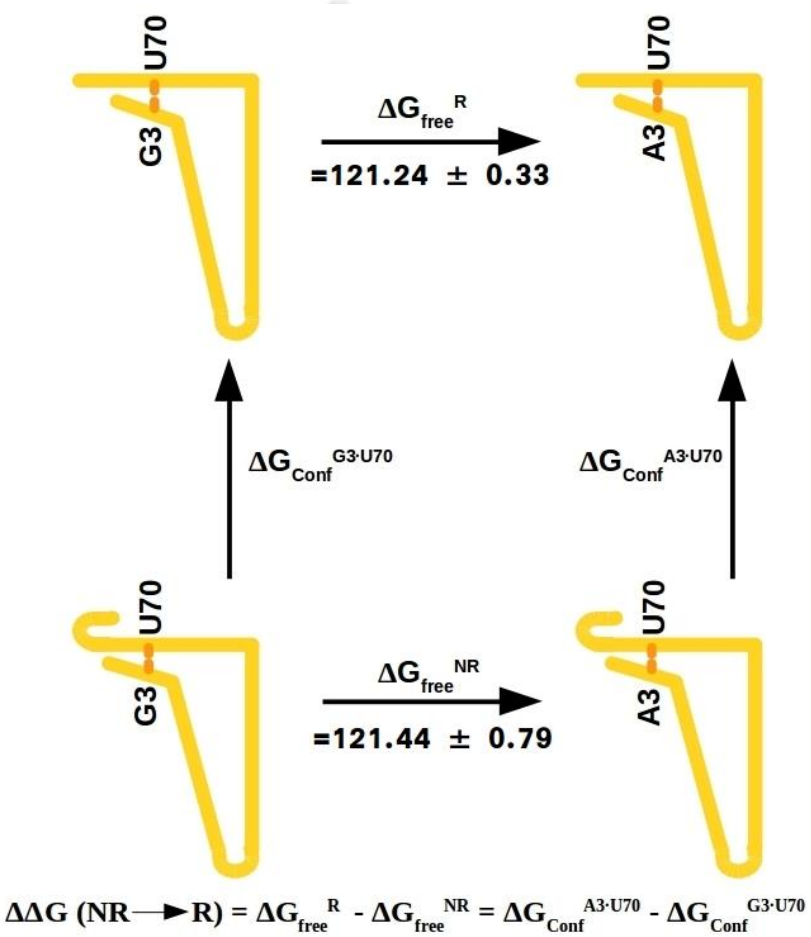


Figure 2.6. Thermodynamic cycle for alchemical transformation of G3•U70 into A3•U70 in free tRNA in water. Vertical legs correspond to the conformation transformation from non-reactive to reactive; horizontal legs correspond to the alchemical transformation of the identity base pair, either in the reactive conformation (upper leg) or in non-reactive conformation (bottom leg).

2.3.2 X-ray vs MD Structures of AlaRS-$tRNA^{Ala}$ Complexes

The crystal structure (**Naganuma et al., 2014**) of the reactive AlaRS-$tRNA^{Ala}$/G3•U70 (**Figure 2.7, a**) complex shows a double hydrogen bonded G3•U70 wobble base pair. This

wobble pair forms H-bonds with AlaRS involving its major (U70-O4 with Asn359) and minor (G3-N2 with Asp450) groove side. The 2'-OH of C4 forms a H-bond with the sidechain of Ser451 in the minor groove and C71-N4 with the mainchain carbonyl of Asn359 in the major groove.

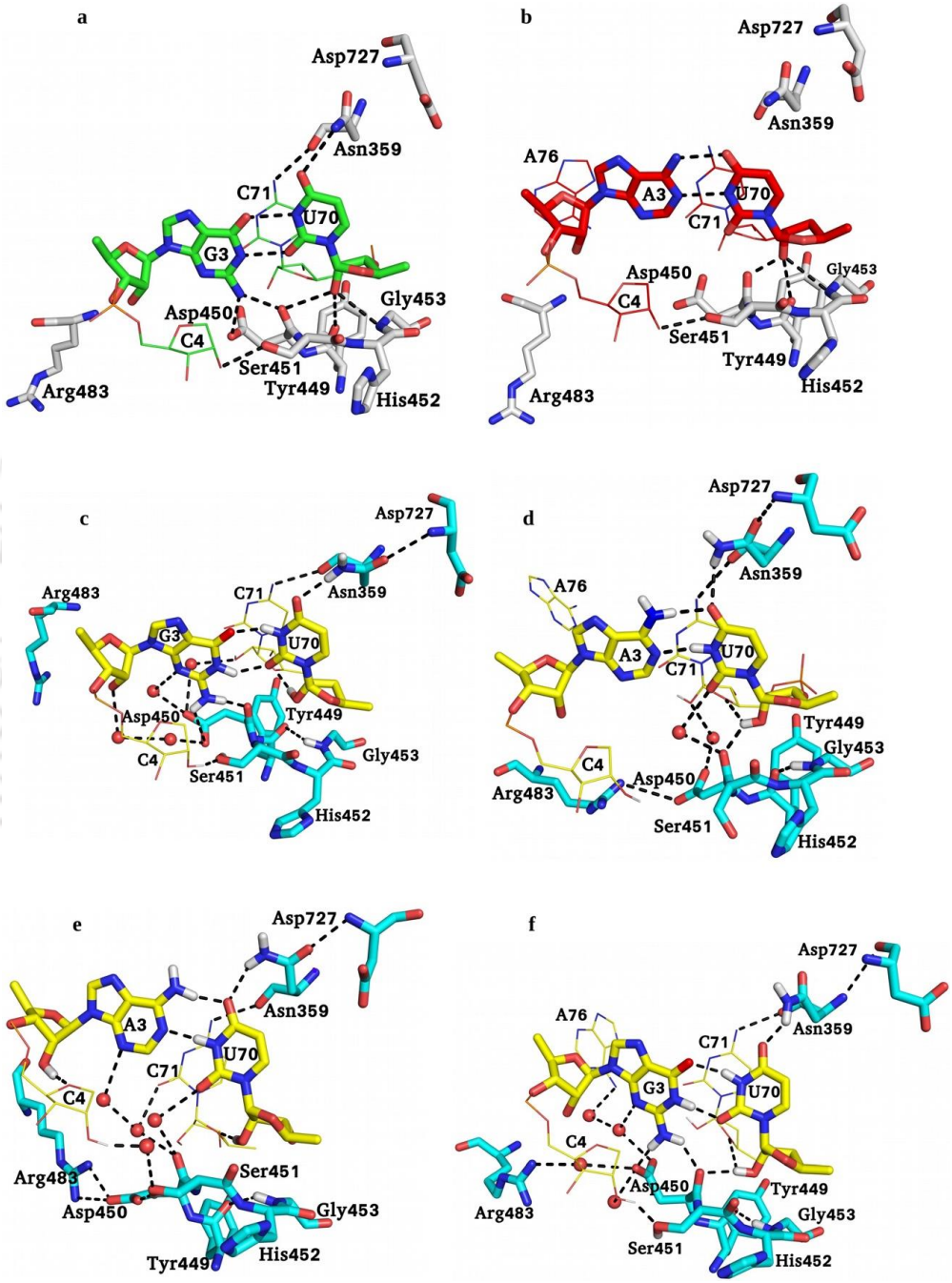
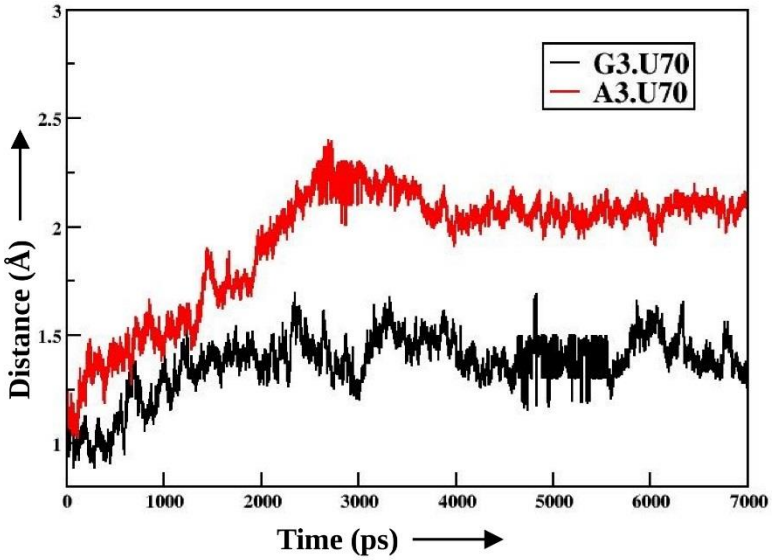


Figure 2.7. Structural insights from X-ray and MD simulations of AlaRS-tRNA^{Ala} complex. The 3•70 base pair is represented by sticks at the center, C4, C71 and A76 are the tRNA nucleotides shown in lines. Base of C4 has not been shown for clarity, few hydrogens are shown to clarify h-bonds. Amino acid residues interacting with the tRNA nucleotides are

shown in sticks (a) G3•U70-AlaRS interaction proposed in the X-ray structure (**Naganuma et al., 2014**). (b) A3•U70-AlaRS interaction proposed in the X-ray structure (**Naganuma et al., 2014**), color code is same as Figure 2.1, a. MD structures of reactive (left side) and non-reactive (right side) complexes are shown in (c, d, e, f) where tRNA is shown in yellow color and AlaRS in cyan color. Waters are indicated by red spheres.

The possibility of H-bonding between the 2'-OH of C71 and the side chain of Asp450 has also been suggested (**Naganuma et al., 2014**). In the non-reactive AlaRS-tRNA^{Ala}/A3•U70 complex (**Figure 2.7, b**), A3 shifted upward away from the minor groove and U70 shifted downward away from the major groove, disrupting tRNA-enzyme major-minor groove interactions. In the non-reactive complex, A76 at the 3'-CCA terminal of tRNA^{Ala} is placed near A3 (~9Å), whereas in the reactive complex (**Figure 2.7, a**) A76 is over 30 Å away from G3. The equilibrated MD structures of G3·U70 and A3·U70 in complex with AlaRS (**Figure 2.7, c,d**) are very similar to their corresponding template x-ray (**Naganuma et al., 2014**) structures (**Figure 2.7, a,b**), with a small MD averaged RMSD of ~ 1.4 Å and ~ 1.9 Å, respectively (**Figure 2.8**). The loop region (**Figure 2.1, a**) is highly flexible and its averaged RMSF is considerably large (**Figure 2.9, 2.10**).



	G3.U70	A3.U70
Average RMSD	1.37±0.15	1.93±0.28

Figure 2.8: Root-mean-square deviation (RMSD) of the heavy atoms of AlaRS.tRNA complexes G3•U70 (black) and A3•U70 (red). RMSD is plotted for 7ns MD trajectory with 1 ps interval. Average RMSD is shown in table (right) with standard deviation.

Robust structural features observed in our MD structures (**Figure 2.7, c,d**) are: (i) U70-O4 always makes an H-bond with the sidechain Asn359 in both the reactive G3•U70 and non-reactive A3•U70 complexes, (ii) the sidechain of Asn359 forms a stable H-bond with the main chain NH-group of Asp727. This interaction locks the conformation of the Asn359 sidechain such that its sidechain $–NH_2$ is pointing towards O4 of U70. (iii) The 2'-OH of U70 forms an H-bond with the ribose ring oxygen of C71 and/ backbone of Asp450. (iv) The side chain of Arg483 of the highly flexible loop region (**Figure 2.1, a**) forms a direct salt bridge interaction with the side chain of Asp450 in the non-reactive AlaRS-$tRNA^{Ala}$/A3•U70 complex (**Figure 2.7, d**) but not in the reactive AlaRS-$tRNA^{Ala}$/G3•U70 case (**Figure 2.7, c**). The 2'-OH of C71 and C4 of tRNA form water-mediated/direct interactions with Asp450 and Ser451, respectively, in the minor groove of reactive AlaRS-$tRNA^{Ala}$/G3•U70 complex (**Figure 2.7, c**). This might indicate the importance of the 2'-OH group of nearby nucleotides in the overall structural stability leading to facile alanylation (**Musier-Forsyth & Schimmel, 1992**).

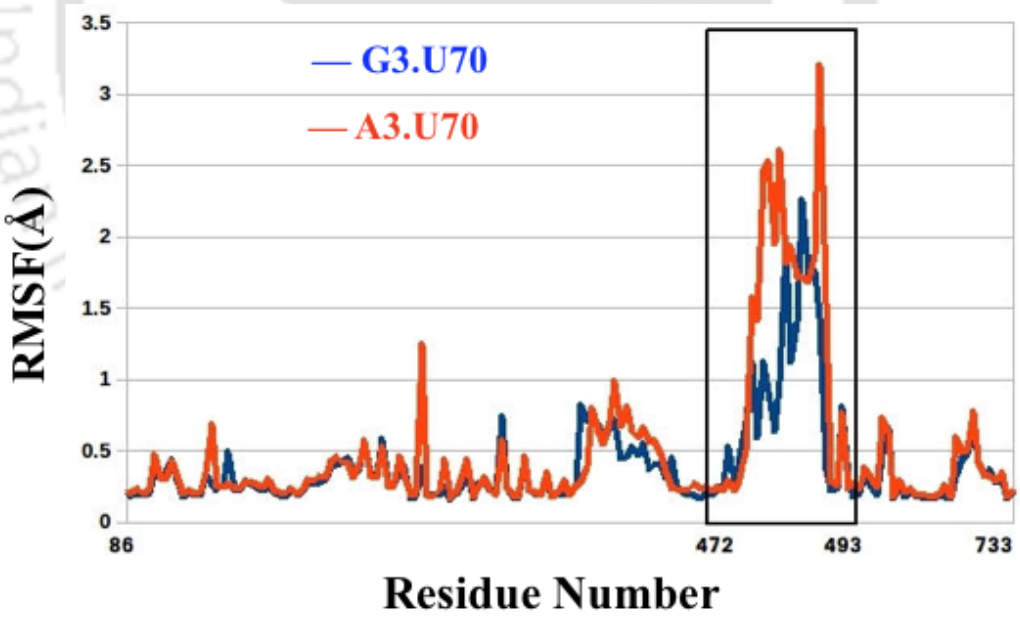


Figure 2.9: Root-mean-square fluctuation of the heavy atoms of the loop region is highlighted with rectangular box. RMSF plotted was averaged over 7ns MD trajectory with 1 ps interval.

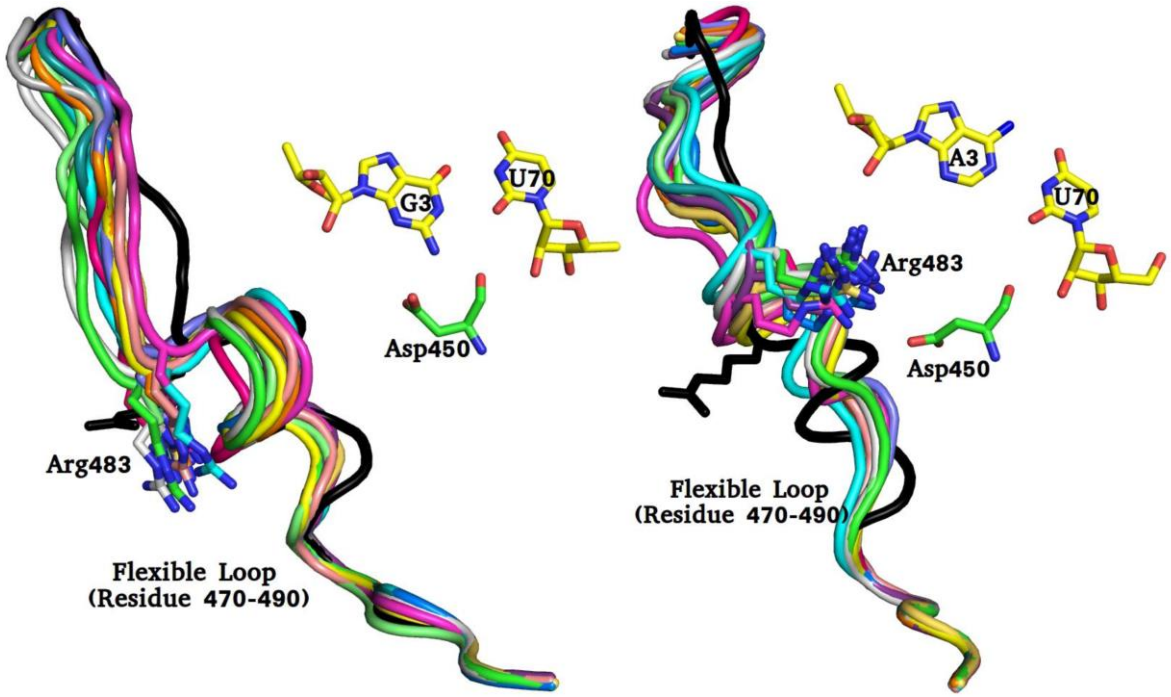


Figure 2.10. Loop comparison between X-ray (black) and MD structures of AlaRS.tRNA complex with respect to G3•U70 (left) and A3•U70 (right). High flexibility of the loop region (residue 470−490) is visible: overlaid 15 snapshots with a 125 ps spacing from a 2 ns MD trajectory. The distantly different orientation of Arg483 is visible. 3•70 base pair, Asp450 and Arg483 in sticks.

2.3.3 MD Structures of Non-cognate AlaRS-tRNA^{Ala} Complexes.

The reactive AlaRS-tRNA^{Ala}/A3•U70 (**Figure 2.7, e**) and non-reactive AlaRS-tRNA^{Ala}/G3•U70 (**Figure 2.7, f**) complexes were modeled by alchemically mutating G3/A3 into A3/G3 in the experimentally resolved reactive/non-reactive complex. Alchemical transformation of AlaRS-tRNA^{Ala}/G3•U70→AlaRS-tRNA^{Ala}/A3•U70 in the reactive complex (**Figure 2.7, e**) results in disruption of the G3-Asp450 interaction and formation of an Arg483-Asp450 salt bridge. Note that the Arg483-Asp450 interaction in the reactive AlaRS-tRNA^{Ala}/A3•U70 complex (**Figure 2.7, e**) is same as in the non-reactive complex (**Figure 2.7, d**), suggesting that the G3→A3 mutation in the reactive complex might reorient the Arg483 side chain resulting Arg483-Asp450 salt bridge.

The non-reactive AlaRS-tRNA^{Ala}/G3•U70 complex (**Figure 2.7, f**) shows almost the same interaction pattern as seen in the reactive complex (**Figure 2.7, d**), except for the formation

of a water mediated interaction between Asp450 and Arg483. Loss of the interaction between C70 and Asn359 in the major groove is observed in the G3•C70 and A3•C70 complexes (**Figure 2.11**) due to electrostatic repulsion. In complexes with the A3•C70 mismatch (**Figure 2.11, c,d**), A3 is shifted upward towards the major groove while C70 is shifted downward towards the minor groove, thereby disrupting the interactions in the major and minor groove. It should be noted that the Asp450-Arg483 salt bridge interaction in the non-reactive state is consistently observed in the MD trajectories (**Figure 2.7, d,f, Figure 2.11, b,d**).

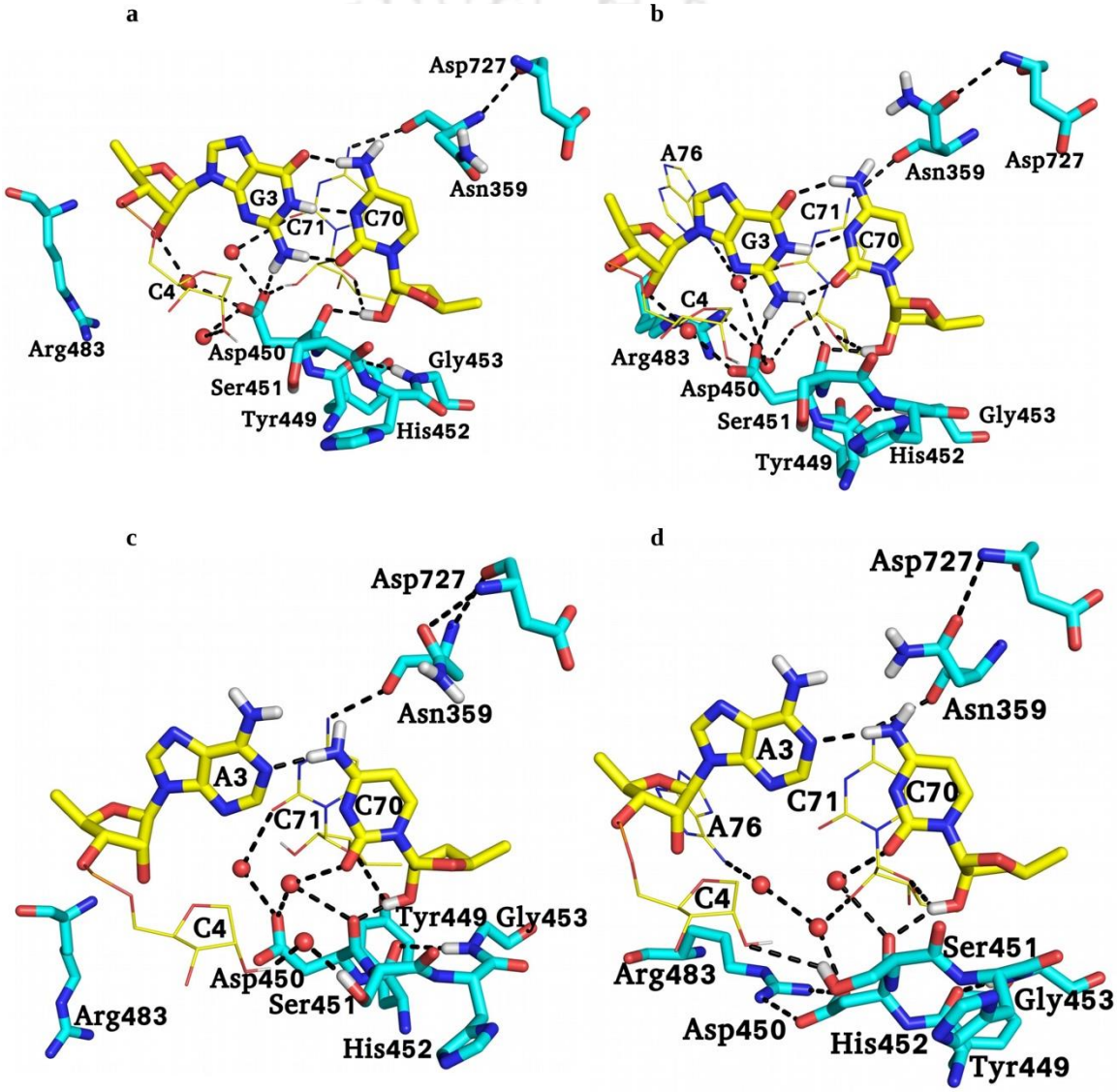


Figure 2.11. MD structures of reactive (left side: a, c) and non-reactive (right side: b, d) complexes with tRNA containing G3•C70 and A3•C70 base pairs. Representation and color coding is the same as Figure 2.7, c-f.

2.3.4 Structure-based Energetics and its Connection to Kinetics

A key question is how the calculated structure-based energetics of **Figure 2.3** is related to the observed kinetics. Our calculations suggest that the reactive complex strongly prefers $tRNA^{Ala}$/G3•U70 with respect to its A3•U70 variant, while the non-reactive complex is non-selective or only very weakly prefers the latter. It should be noted that alanine charging by AlaRS is also possible for G3•U70 containing mini- or micro-helices (**Shi et al., 1990; Francklyn & Schimmel, 1989**) with a substantially increased Michaelis constant (K_M) but an almost unaffected k_{cat}. This suggests that all the helices containing G3•U70 undergo aminoacylation at a comparable rate once they are bound to the enzyme. A3•U70 in mini-/micro-helices completely failed to demonstrate any aminoacylation by alanine, even after prolonged incubation with substrate level enzyme concentration (**Park et al., 1989; Shi et al., 1990; Francklyn & Schimmel, 1989**). Recently, an approximately 100-fold difference in k_{cat} was reported between wild-type $tRNA^{Ala}$/G3•U70 and $tRNA^{Ala}$/A3•U70 while the K_M values were relatively similar. Moreover, single turn-over measurements (**Naganuma et al., 2014**) revealed that both the rate of the chemical step and the binding affinity were similar for the two cases, with differences being much smaller than the two-orders of magnitude effect on k_{cat}. To account for this observation a kinetic model was proposed where the substrate can bind reversibly in both reactive and non-reactive conformations, which can also interconvert on the enzyme. After an irreversible aminoacyl transfer step occurring from the reactive state, the system ends up in the reactive bound product state where the product can either dissociate directly or via the non-reactive conformation, all these later steps being reversible. This complex model, with no less than 13 adjustable rate parameters, was fitted to the pre-steady state kinetics curves (**Naganuma et al., 2014**). However, although this model correctly predicts the 100-fold change in k_{cat}, closer examination of the resulting rate constants reveal that the k_{cat}/K_M and K_M values clearly disagree with the experiments. Thus, the predicted K_M for $tRNA^{Ala}$/A3•U70 is markedly lower (0.03 µM) than for $tRNA^{Ala}$/G3•U70 (0.9 µM) due to its strong binding to the NR state implied by the model. Moreover, the fitted forward and reverse rate constants for substrate binding in the reactive and non-reactive conformations would also imply that the non-reactive state is more selective for $tRNA^{Ala}$/A3•U70 (> 3 kcal/mol) than the reactive state is for $tRNA^{Ala}$/G3•U70 (~0.7 kcal/mol) (**Naganuma et al., 2014**). This is clearly at variance with our free energy calculations based on the experimental crystal structures, which thus are not consistent with

the proposed kinetic model. Furthermore, both the cognate and non-cognate thermodynamic cycles connecting the free enzyme to its substrate-bound states in the reactive and non-reactive conformations are subject to a ~2 kcal/mol closure error, if the fitted rate constants are used.

Applying Occam's razor to the present problem suggests that a much simpler kinetic model, based on our calculated selectivities of the reactive and non-reactive states, can better account for the observed data. Such a sequential three-step model is shown in **Figure 2.12**, where an initial unselective binding step yields a non-reactive complex ($E{\cdot}tRNA^{NR}$), which can subsequently reversibly convert to the reactive complex ($E{\cdot}tRNA^{R}$). Once formed,

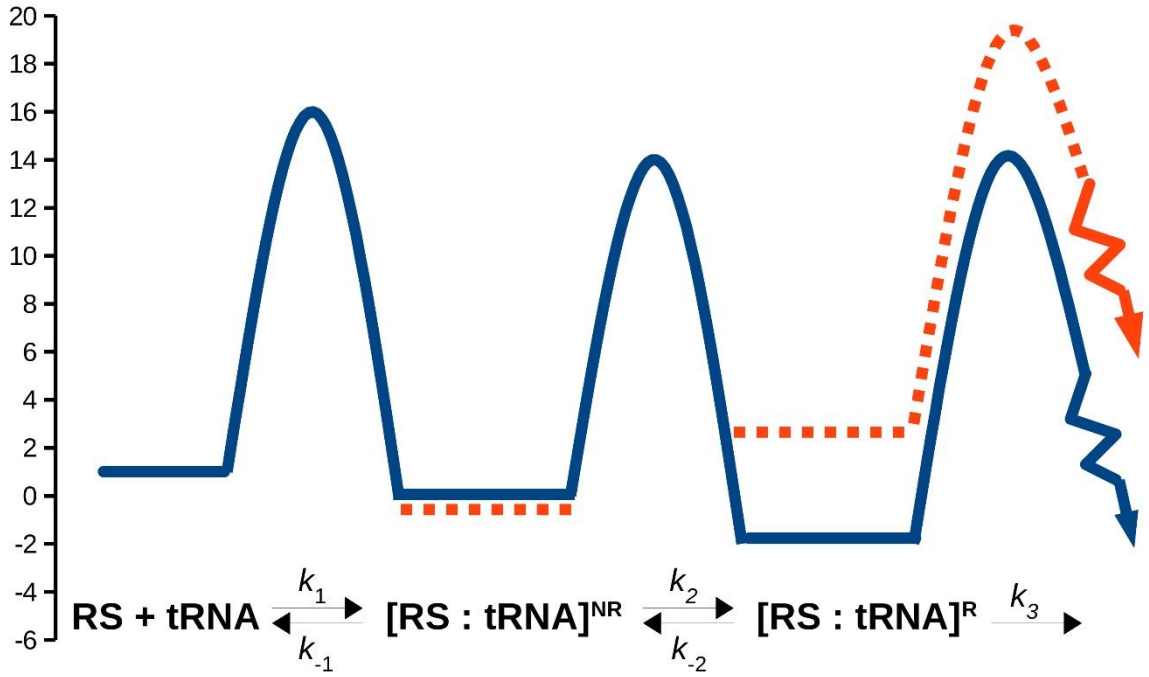


Figure 2.12. Schematic free energy diagram for $tRNA^{Ala}$ selection by AlaRS (without proofreading). Illustration of how the non-selective (non-reactive complex: NR) and highly selective (reactive complex: R) states can be used to drive aminoacylation favoring cognate tRNA/G3•U70. Binding free energy differences between cognate tRNA/G3·U70 and non-cognate tRNA/A3•U70 is approximately 0.1 and 5 kcal/mol for NR and R complexes, respectively. The reactive complex (R) is downhill for cognate tRNA/G3•U70 and uphill for non-cognate tRNA/A3•U70 (see text). Biochemically plausible activation barriers for the elementary steps are depicted as ~14-16 kcal/mol and are assumed to be the same for both cognate and near-cognate tRNA.

E·tRNA^{R} undergoes aminoacylation with the same rate irrespective of the whether the tRNA is cognate or non-cognate. According to our free energy calculations E·tRNA^{R}, is strongly selective for cognate tRNA^{Ala}/G3•U70 and it is basically the different (Boltzmann) probabilities of reaching this state that are responsible for the observed difference in k_{cat}. A key feature is also that the initial E·tRNA^{NR} state does not show any significant kinetic or thermodynamic discrimination between cognate and non-cognate tRNAs. It thus essentially has the same properties as those obtained here for the crystallographic non-reactive state, which could be considered as one example of an unspecific binding state. This scheme is thus similar to that proposed for the initial selection of aminoacyl-tRNA on the mRNA programmed ribosome (**Satpati et al., 2014**). Further, the conformational change from the weakly selective non-reactive complex to the highly selective reactive complex is proposed to be downhill for cognate tRNA^{Ala}/G3•U70. The calculated energetics then automatically places the reactive AlaRS-tRNA^{Ala}/A3•U70 complex uphill from its non-reactive state. This would explain why the crystal structures only trap AlaRS-tRNA^{Ala}/G3•U70 in the reactive state and AlaRS-tRNA^{Ala}/A3•U70 in the non-reactive conformation. The kinetic parameters (see **Appendix 5** for full derivation) of the proposed scheme (**Figure 2.12**) are thus

$$k_{cat} = \frac{k_2 k_3}{(k_2 + k_{-2} + k_3)} \qquad \text{(Eq. 1)}$$

and

$$K_M = \frac{(k_{-1}k_{-2} + k_{-1}k_3 + k_2 k_3)}{k_1(k_2 + k_{-2} + k_3)} \qquad \text{(Eq. 2)}$$

Here, the absolute magnitude of K_M (and k_{cat}/K_M) is of course dependent on the effective association rate constant k_1, but the accuracy $A = (k_{cat}/K_M)^{cognate}/(k_{cat}/K_M)^{non\text{-}cognate}$ is independent of k_1. By inserting values of k_3 (aminoacylation step) in correct experimental range (**Naganuma et al., 2014; Zhang et al., 2006**) and adjusting k_3 so as to get K_M also in the correct range, this model accounts well for the experimental kinetic data when we insert our calculated values for the selectivities of the NR and R states. A numerical example is shown in **Table 2.2**.

Table 2.2. Numerical analysis of aminoacylation without proof-reading.

rate constant (s^{-1})	**Correct (G3•U70)**	**Incorrect (A3•U70)**	**Correct (kcal/mol)**	**Incorrect (kcal/mol)**
k_1 ($\mu M^{-1}s^{-1}$)	10	10	ΔG(NR)	ΔG(NR)
k_{-1}	30	30	0	0
k_2	500	500	ΔG(R)	ΔG(R)
k_{-2}	20	50000	−2.1	+3.1
k_3	15	15		
k_{cat}	14.0(14.4*)	0.15(0.14*)		
K_M (μM)	1.6	3.0		
k_{cat}/K_M ($\mu M^{-1}s^{-1}$)	8.8	0.05		

Rate constants for the non-specific AlaRS-tRNA association (k_1), dissociation from the non-specific complex (k_{-1}), the non-reactive (NR) ↔ reactive (R) conformational change (k_2 and k_{-2}) and for aminoacylation from reactive complex (k_3). Calculated relative binding free energy values at the experimental temperature of 333 K (**Naganuma et al., 2014**) are given in the two rightmost columns. *Pre-steady state kinetic constants from (**Naganuma et al., 2014**).

2.4 Discussion

Kinetic and structural studies (**Park et al., 1989; Naganuma et al., 2014**) have certainly enriched our understanding of tRNA selection by AlaRS. We are, nevertheless, lacking a detailed structure-based free energy landscape of tRNA selection by the enzyme, as the link between the kinetics and 3D structure is still missing. However, MD simulations should in principle be able to fill this gap. Here, we have performed molecular dynamics free energy calculations and computed the relative $tRNA^{Ala}$ binding free energies to AlaRS in the reactive and non-reactive complex for various cases, G3•U70, A3•U70, G3•C70 and A3•C70. It was found that the reactive complex consistently disfavors A3•U70, G3•C70 and A3•C70 with respect to the cognate $tRNA^{Ala}$/G3•U70 by a ΔΔG over 5 kcal/mol. In contrast, the non-reactive complex does not discriminate between A3•U70 and G3•U70, although discrimination against G3•C70 and A3•C70 is clearly evident. The magnitudes of the relative binding free energies (ΔΔG) are experimentally unknown. Hence, our calculated

relative binding strengths cannot at present be compared to experimental data, but their magnitudes appear biochemically realistic.

Molecular dynamics free energy calculations adopted here have proven to be sufficiently accurate for quantitative prediction of base pairing energetics related to protein synthesis on ribosome (including mRNA decoding during initiation, elongation, termination, and role of tRNA modification in decoding, etc) (**Lind et al., 2019; Bock et al., 2018**). Application of MD simulations in studying RNA and DNA has been discussed extensively in the literature (**Sponer et al., 2018; Perez et al., 2012**). Truncated spherical region (radii ~25-40 Å) cut out of the large molecular assembly and solvated by a spherical droplet/box of explicit water is often preferred model for performing MD free energy simulations, as they reduce the computational time and increase the convergence. (**Bock et al., 2018; Sponer et al., 2018; Perez et al., 2012**). If the energetics is controlled by localized interactions, then the reduced truncated model is a very good choice, which reduces the computational cost and improves the convergence by not sampling irrelevant large scale conformational motion (**Satpati & Simonson, 2012; Satpati et al., 2011; Satpati et al., 2014; Lind et al., 2019; Kumar & Satpati, 2018; Kumar et al., 2017**). Large scale conformational change certainly requires consideration of much larger truncated/complete biomolecular system and exhaustive sampling. Accurate free energy calculations are certainly limited by the accuracy of force-fields, finite sampling, resolution of the experimental structures, inherent complexity of the biomolecular system (e.g, protonation states of ionizable groups, the position of ions and waters, long-range interactions etc.).

The reactive and non-reactive complexes are distinctly different in two aspects, the orientation of the tRNA 3'-CCA arm (as described previously) (**Naganuma et al., 2014**) and the RNA-protein interactions in the minor groove. The 3'-CCA arm is placed in the aminoacylation site only in the reactive complex and folded back in the non-reactive complex (**Figure 2.1, a**). The MD simulations suggest that the Arg483-Asp450 interaction in the minor groove is a unique feature of non-reactive complex, as it is evident from most of the MD runs. It should be noted that the loop region (**Figure 2.1, a**) containing Arg483 was poorly resolved (with a B-factor ~ 195 $Å^2$) in the x-ray structure (**Naganuma et al., 2014**). A different orientation of Arg483 in the reactive and non-reactive complex is also evident from the X-ray structures (**Figure 2.13**). The large RMSF (**Figure 2.9**) obtained from MD trajectories further suggests a high flexibility of this loop region. On the other

hand, the Arg483-Asp450 salt bridge seems to be the characteristic feature of the non-reactive AlaRS-tRNA^{Ala}/A3•U70 complex (**Figure 2.10**) and is stable throughout the trajectories of the non-reactive complex.

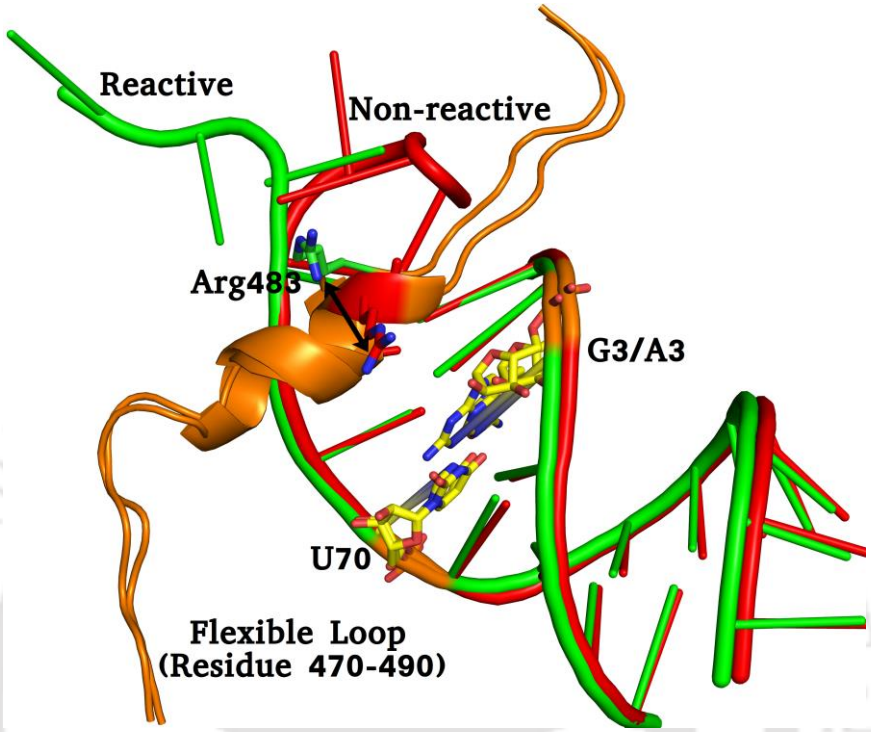


Figure 2.13. X-ray structure of the reactive (green) and non-reactive (red) complex revealed that Arg483 of highly flexible loop region (orange) is in different orientation (represented by double headed arrow) with respect to 3•70 base pair (yellow sticks).

Interestingly, the interaction between the negatively charged side chain of Asp450 and the exocyclic –NH_2 of G3 is crucial for stability of the native complex (**Figure 2.7, c**) and a major reason for why mutation of G3 to A3 leads to large discrimination in the reactive complex.

It should be noted that the forward alchemical transformation of AlaRS-tRNA^{Ala}/G3•U70 into AlaRS-tRNA^{Ala}/A3•U70 in the reactive complex (**Figure 2.7, e**) spontaneously produces the Arg483-Asp450 salt bridge, which appears as a characteristic structural feature of the non-reactive complex (**Figure 2.7, d**). The NR conformation need not be unique and different conformations of -CCA end, could probably yield non-reactive states with little selectivity ($\Delta\Delta G \sim 0$ kcal/mol). Hence, the observed X-ray conformation of the NR complex

could be seen as one of those possibilities and it seem clear that it is the difference in the local environment around 3•70 base pair that is responsible for the differential selectivity.

2.5 Conclusion

In conclusion, the molecular dynamics free energy simulations have revealed that the reactive complex strongly favors $tRNA^{Ala}$/G•U with respect to the A•U variant and the geometrical difference between G3•U70 and A3•U70 does not play any significant role in driving the conformational change (NR ⇆ R) in free $tRNA^{Ala}$ in water. We have also shown here how a simple three state kinetic model can explain how the single wobble pair ensures tRNA specificity by altering k_{cat} (favoring correct tRNA by ~ 100 fold), while keeping K_M more or less similar. The activation barriers for the forward processes are considered to be identical for the correct and incorrect tRNA's in this model. Hence, irreversible amino acid attachment will take place once the "R" state is reached, regardless of the nature of tRNA (correct or incorrect). It is rather the differential probability of reaching the highly selective "R" complex that gives different k_{cat} values for correct and incorrect tRNA, thereby ensuring accuracy in tRNA selection by AlaRS.

Chapter 3

Stop Codon Recognition by Eukaryotic Release Factor 1

In translation termination, eukaryotic release factor (eRF1) recognizes mRNA stop codons (UAA, UAG or UGA) in ribosomal A-site and triggers release of nascent polypeptide chain from P-site tRNA. eRF1 is highly selective for U in the first position and combinations of purines (except two consecutive guanine: GG) in the second and third positions. Eukaryotes decode all the three stop codons with a single release factor eRF1, instead of two (RF1 and RF2), in bacteria. Furthermore, unlike bacterial RF1/RF2, eRF1 stabilizes compact U-turn mRNA configuration in the ribosomal A-site by accommodating four nucleotides instead of three. Despite the available cryo-EM structures (resolution ~3.5-3.8Å), the energetic principle for eRF1 selectivity towards a stop codon remains a fundamentally unsolved problem. Using cryo-EM structures of eukaryotic translation termination complexes as templates, we carried out molecular dynamics free energy simulations of cognate and near cognate complexes to quantitatively address the energetics of stop codon recognition by eRF1. Our results suggest that eRF1 has a higher discriminatory power against sense codons, compared to that reported earlier for RF1/RF2. The compact mRNA formed specific intra-mRNA interactions, which itself contribute to stop codon specificity. Furthermore, the specificity is enhanced by the loss of protein-mRNA interactions, and most importantly, by desolvation of the incorrect codons in the near-cognate complexes. Our work provides a clue to how eRF1 discriminates between cognate and near cognate codons during protein synthesis.

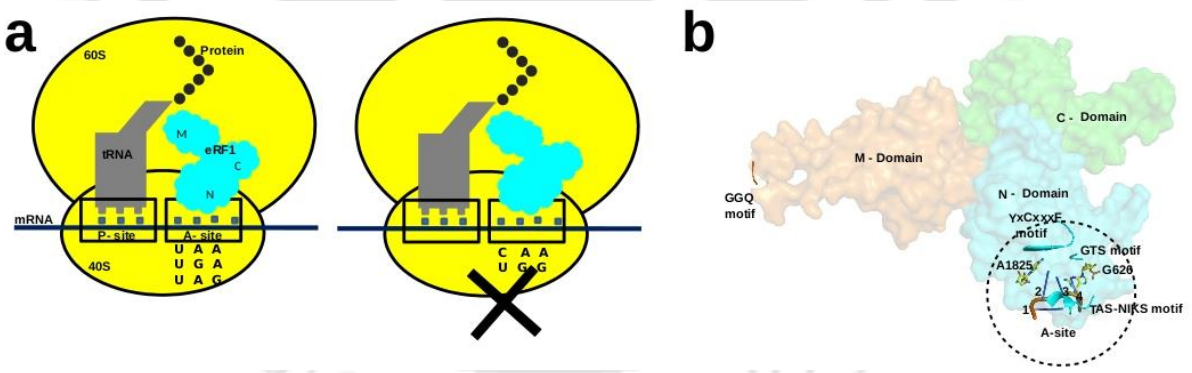
==========

The work published in Journal of Chemical Information and Modeling,

DOI:10.1021/acs.jcim.7b00340

3.1 Background

Triplet mRNA codes could be classified into two groups: sense and stop codons. Sense codons are sequentially recognized by aminoacyl-tRNA's in the ribosome during translation, whereas stop codons are recognized by proteins called release factors which terminate the growing polypetide chain, as shown in **Chapter 1, Figure 1.5**. Recognition of stop codon by release factor and release of polypetide chain from translating ribosome is the key for correct translation termination. Premature termination leads to disease (**Mort et al., 2008; Keeling et al., 2014**). The high accuracy of translation termination (**Freistroffer et al., 2000**) by release factor could be achieved from two aspects: (a) binding affinity difference between cognate and near-cognate codon on the ribosome, K_M (b) facile eRF1-catalyzed hydrolysis of polypeptide chain from P-site tRNA for the cognate codon, K_{cat}. However binding has been reported to be more crucial for bacteria (**Freistroffer et al., 2000; Trappl et al., 2014**). In bacteria, two release factors RF1 (recognizes UAA and UAG) and RF2 (recognizes UAA and UGA) ensure discrimination against the sense codons (**Zhou et al., 2012; Korostelev, 2011; Rodnina, 2013; Korostelev et al., 2008; Laurberg et al., 2008; Weixlbaumer et al., 2008; Jin et al., 2010**).



e 3.1. Stop codon recognition by eRF1. (a) Schematic representation of eRF1 (cyan) binding to cognate and near cognate codon (blue) in the A site of ribosome with P site tRNA (gray) holding the nascent polypetide chain (gray beads). E site of ribosome is not shown. (b) Three eRF1 domains N (cyan), M (orange) and C (green). Conserved motifs (TAS-NIKS, GTS, YxCxxxF) of N domain and GGQ motif of M domain are shown as cartoon representation. The first-fourth mRNA bases in the A site of ribosome is marked as 1-4. Ribosomal bases A1825 and G626 are represented by sticks (yellow). Decoding site is highlighted by dashed circle.

Medium resolution bacterial termination complexes (3-3.5Å) (**Laurberg et al., 2008; Weixlbaumer et al., 2008**) had shown extensive protein-RNA interactions and provided the basis for computational analysis leading to quantification of the discriminatory power of release factor against sense codons (**Trobro & Åqvist, 2007; Sund et al., 2010; Korkmaz et al., 2014**). Unlike sense codons the third base of the bacterial stop codon stack with the ribosomal G530 of 16S rRNA instead of stacking with the first two bases.

The versatile eukaryotic release factor (eRF1) can decode all three stop codons (UAA, UAG, and UGA) (**Figure 3.1, a**) and consists of three domains (N, M and C) (**Figure 1.6**; and **Figure 3.1, b**). As described in **Chapter 1**, the N-terminal domain is involved in stop codon recognition in the A site, the M domain triggers the hydrolysis and releases the polypetide, while the C domain mediates the interaction between eRF1 and other termination factors (eRF3 and ABCE1) (**Bertram et al., 2000; Song et al., 2000; Cheng et al., 2009; Preis et al., 2014; Pisarev et al., 2010; Shoemaker & Green, 2011**) (**Figure 3.1, b**). eRF1 is delivered to the ribosome by GTP bound eRF3 as a ternary complex. In response to stop codon binding, GTP is hydrolyzed to GDP in eRF3. GDP bound eRF3 loses its affinity for eRF1 and dissociates, which leads to accommodation of eRF1 in the A-site with correct positioning of the GGQ motif of the M domain for hydrolysis and peptide release (**Alkalaeva et al., 2006; Frolova et al., 2002**). ABCE1 can further bind to the accommodated eRF1 and assist in polypeptide release and trigger ribosome recycling factors for dissociation of post termination complex (**Pisarev et al., 2010**). Structural, biochemical, and mutational studies have suggested that GGQ motif of the M domain (**Figure 3.1, a, b**) is crucial for hydrolysis (**Preis et al., 2014; Frolova et al., 1996; Muhs et al., 2015**), whereas TAS-NIKS motif (residue 58-64), GTS loop (residue 31-33) and YxCxxxF motif (residue 125-131) of N doamin (**Frolova et al., 2002; Behrmann et al., 2015; Wong et al., 2012; Conard et al., 2012; Seit-Nebi et al., 2002**) are important for stop codon recognition (**Figure 3.1, b**).

Recent advancement in structural characterization of few eukaryotic translational termination complexes (**Brown et al., 2015; Bhushan et al., 2010; Matheisl et al., 2015**) have shown that stop codons adopt a compact U turn like conformation in the A-site of ribosome-eRF1 complex. As discussed in **Chapter 1**, unlike sense (**Jenner et al., 2010;**

Ogle et al., 2002; Kryuchkova et al., 2013) or bacterial stop codons (**Laurberg et al., 2008; Weixlbaumer et al., 2008**), the fourth base following the stop codon is also compacted into the ribosomal A-site (**Figure 1.7, and Figure 3.1, b**). It has been suggested (**Brown et al., 2015; Matheisl et al., 2015**) that a compacted four nucleotide U turn conformation in the A-site is stabilized by (1) stacking of ribosomal bases A1825 and G626 of 18S rRNA (A1493 and G530 of 16S rRNA in bacteria) with the second and fourth bases of mRNA respectively (**Chapter 1**, **Figure 1.7**) and (2) extensive eRF1-mRNA interactions networks (**Chapter 1**, **Figure 1.8**). It should be noted that no near-cognate complexes (e.g, eRF1-ribsome complex with sense codons) have been isolated experimentally. Those near-cognate complexes might be thermodynamically hidden due to very low Boltzmann weight. The following key questions related to stop codon decoding remain unanswered (a) How strongly are the sense codons discriminated by eRF1? (b) How is discrimination strength related to the 3D structures of the cognate and near-cognate complexes? Sufficiently good models for structure-based MD simulations are now possible using these medium resolution cryo-EM structures (**Brown et al., 2015**).

Here, in this chapter, we report structure-based computation for deciphering the energetics of stop codon decoding, thereby providing the key missing link between 3D structures and energetics. Starting with Cryo-EM structures (**Brown et al., 2015**) (resolution ~3.45-3.65Å) as our initial model, we have performed molecular dynamics free energy simulations of cognate and near-cognate translational termination complexes. These calculations involve computing the change in binding affinity for eRF1 in the ribosomal A-site upon stop codon mutations in all three codon positions.

3.2 Methods

3.2.1 Molecular Dynamics Setup

We have used the molecular dynamics setup as already described in **Chapter 2**, **Figure 2.2**. 25Å radii sphere centered on the N9 atom (**Figure 3.2**) of the second codon position were cut from the cryo-EM structures. Heavy atoms of Protein and RNA between 22 and 25Å from the sphere's center (buffer region) were harmonically restrained to their experimentally determined positions, leaving the inner 22 Å radius shell fully flexible. Simulations were performed with two different solvation models: (A) overlaying by a 37Å

radius water droplet and (B) overlaying with a cubic water box of 80Å edge. Free energy calculations were done for model A.

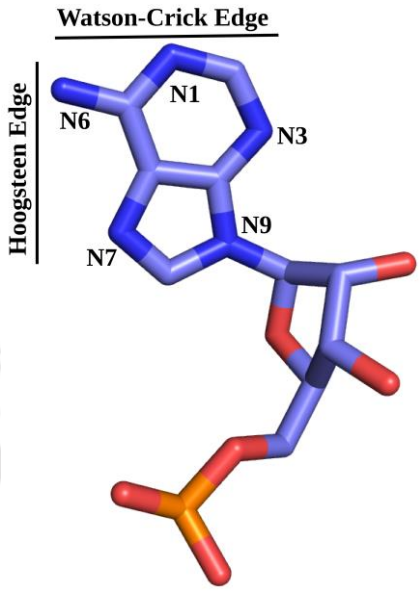


Figure 3.2. Watson-Crick (WC) and Hoogsteen edge of adenosine. Hydrogen atoms are not shown for clarity.

Model A. All MD simulations were performed with the CHARMM22 force field (**MacKerell et al., 1998**) implemented in Q (**Marelius et al., 1998**) program. Solute atoms of outer 3Å sphere were harmonically restrained to their experimentally reported structure with a 10 kcal.mol^{-1}.Å^{-2} force constant. Water molecules near the simulation sphere boundary were subjected to radial and polarization restraints as described in the SCAAS model (**Marelius et al., 1998; King & Warshel, 1989**). 1 and 2 fs time step have been used for MD with a direct cutoff of 10Å for nonbonded interactions. The local reaction field multipole expansion method (**Lee & Warshel, 1992**) has been used to incorporate the long-range electrostatic interactions beyond the cutoff. No cutoff was applied for the mutated base. The simulations involves 600 ps of equilibration through a series of short MD runs. In the first 400 ps of equilibration the system was heated upto 310K and kept fixed through out the production MD trajectories. For both the free and eRF1-bound complex, each free energy calculation (with 51 discrete λ points) was based on 54-82 ns of data collection averaged over 17-24 replicates with different initial velocities. Convergence has been checked by computing the standard errors of the mean from these replicas. A low standard error of mean (less than ±1 kcal/mol) was an indications of good convergence. Computed

free energy for forward and reverse process for individual replicate always differed by less than 1 kcal/mol (**refer appendix Figure 3A-1**).

Furthermore, each replica has been divided into two equal halves, and the difference between the computed free energy from the two halves is always less than 1 kcal/mol. The above results indicated good convergence of our simulations. About 600 ns of MD from the complete set of FEP runs have been used to compute relative binding affinities $\Delta\Delta G_{bind}$ (**refer appendix Figure 3A-1**), leading to an apparently small statistical error. The calculations were repeated for 30Å truncated sphere (with 27Å - 30Å buffer region) into a 40Å radius water droplet. The robustness of the free energy simulations were examined with a two different simulation sphere (25Å and 30Å radius). The computed $\Delta\Delta G_{bind}$ differed by less than 1.4 kcal/mol for two different truncated models (**Table 3.1**). Hence, we reported the averaged result obtained from the two different simulation models.

Table 3.1 Computed free energy change for alchemical transformation of two different simulation models. [a]

Mutation Type (PDB used)	**Truncated model radius**	$\Delta G^{Complex}$	ΔG^{free}	$\Delta\Delta G = \Delta G^{Complex} - \Delta G^{free}$
UAA → CAA (3JAG)	25Å	21.6 (1.3)	13.0 (0.9)	8.6 (1.6)
	30Å	20.3 (1.1)	12.6 (1.1)	7.7 (1.5)
UAA → UGA (3JAG)	25Å	-118.2 (1.4)	-120.5 (0.6)	2.3 (1.5)
	30Å	-118.9 (1.2)	-120.5 (0.8)	1.6 (1.4)
UAA → UAG (3JAG)	25Å	-117.2 (1.3)	-118.6 (0.4)	1.4 (1.4)
	30Å	-116.9 (1.4)	-119.3 (0.4)	2.4 (1.4)
UGA → UGG (3JAI)	25Å	-113.2 (2.6)	-119.9 (0.9)	6.7 (2.7)
	30Å	-112.7 (2.7)	-119.2 (1.4)	6.5 (3.0)

[a]Energies are in kcal/mol. Standard deviations are given in parenthesis for 10-14 independent simulations with different randomized starting velocities. Computed ΔG for 25Å and 30Å truncated system are given in the 2nd and 3rd column of the table. The uncertainty in ΔΔG is estimated by propagation of individual uncertainties of the $\Delta G^{Complex}$ and ΔG^{free}. Positive sign of ΔΔG indicate the stop codon preference of eRF1 in the ribosome.

Model B. Periodic boundary conditions with particle mesh Ewald method (**Darden et al., 1993**) for long-range electrostatics have been used for MD simulations. Here, a 16Å cutoff was used for calculating van der Waals interaction. Langevin dynamics for the heavy atoms of protein-RNA and solvent, with a coupling coefficient of 5 ps^{-1}, was used to control the temperature. The pressure was controlled by a Langevin piston, Nose-Hoover method. CHARMM36 (**Huang & MacKerell, 2013**) force field has been used. Calculations were done with CHARMM (**Brooks et al., 2009**) and NAMD (**Phillips et al., 2005**) programs.

Simulations for each complex has been repeated 6-12 times with different initial velocities and minimization steps. Each simulation model was minimized by steepest descent method (100-200 steps) followed by adopted basis Newton-Raphson algorithm (100-200 steps) implemented in CHARMM (**Brooks et al., 2009**) program. Equilibration has been done for 150-200 ps. Structural analysis has been done with 7ns production dynamics for each simulation models. About 300 ns of MD has been performed for structural analysis using model B. The temperature and pressure were maintained at 310 K and 1 bar, respectively. TIP3P model (**Jorgensen et al., 1989**) for water was used to run simulations. The rms deviation between 3JAG and 3JAI is 0.31 Å. The final complex (after alchemical transformation) gives an average rmsd of 1-1.15 Å with respect to the starting cryo-EM structure. Ions from the PDB were included for MD simulations. The position of internal Mg^{2+} ion in the A site were moderately stable, with shifts about 2-2.5Å observed. The overall charge of the simulation sphere/box was neutralized by scaling down the partial charges of the biomolecule starting from the edge of the sphere.

3.2.2 Free Energy Calculations

Free energy perturbation method (FEP) has been used to compute the relative binding free energies of mRNA mutations in the ribosome-eRF1 complexes (**Sund et al., 2010; Satpati et al., 2014**). Initial coordinates for MD were taken from the cryo-EM structures of the cognate complex of Protein Data Bank (PDB) entries 3JAG, 3JAI and 3JAH (**Brown et al., 2015**). U→C, A→G mutations were introduced at various sites of mRNA stop codon (UAA → CAA, UAA → UGA, UAA → UAG and UGA → UGG) for the eRF1-ribsome complex in A site. Corresponding simulations were also conducted for the stop codon-programmed ribosome with an empty A-site to be able to calculate $\Delta\Delta G_{bind}$ through the thermodynamic

cycle, see **Figure 3.3**. Ditect calculations for UAA → CAA/UGA/UAG and UGA → UGG have been done, starting from 3JAG and 3JAI, repectively. UAA → UGG were obtained indirectly from the two direct calculation. The equation for relative binding free energy becomes, $\Delta\Delta G_{bind} = \Delta G_{\text{bind}}(\text{CAA}) - \Delta G_{\text{bind}}(\text{UAA}) = \Delta G_{comp} - \Delta G_{free}$.

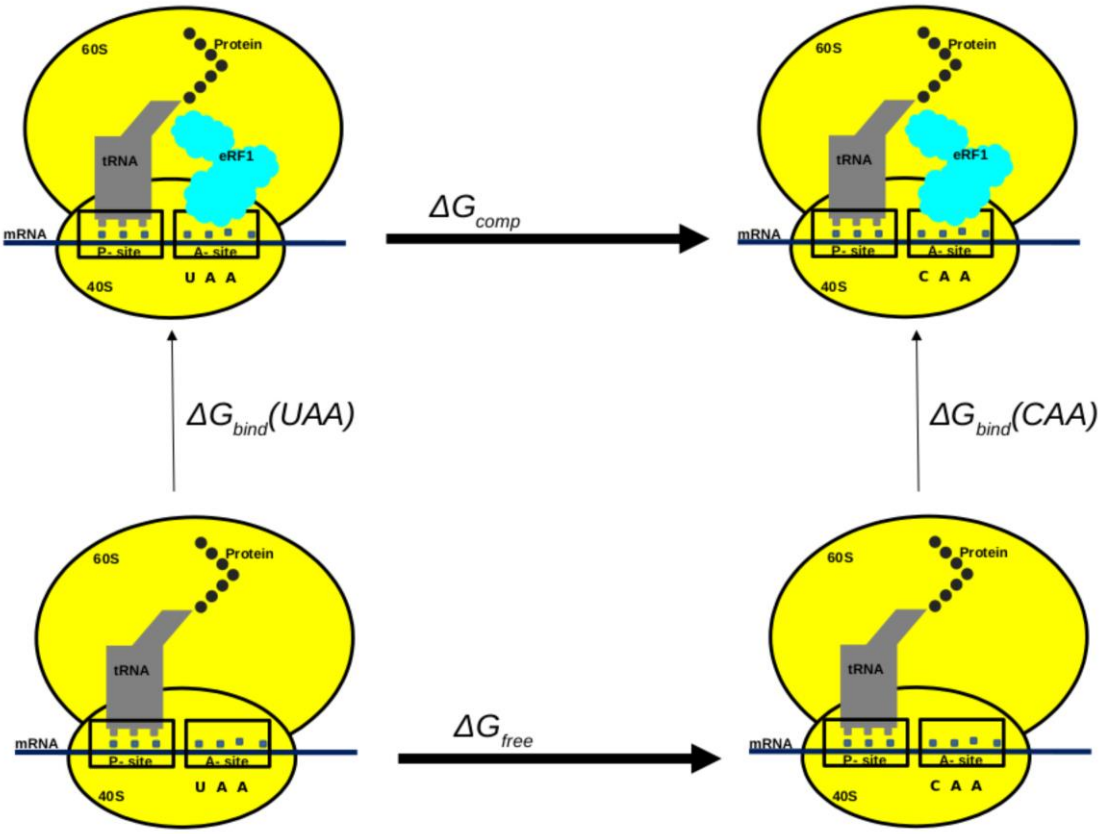


Figure 3.3. Thermodynamic cycle for Stop codon:eRF1 binding on ribosome. Vertical legs correspond to binding; horizontal legs correspond to the alchemical transformation/mutation of stop codon (UAA) into sense codon (CAA), either in the eRF1 bound (above) or eRF1 free (below) ribosome in solution . The horizontal legs of the thermodynamic cycle has been computed by MD simulations. The binding free energy difference is $\Delta\Delta G_{bind} = \Delta G_{comp} - \Delta G_{free} = \Delta G_{\text{bind}}(\text{UAA}) - \Delta G_{\text{bind}}(\text{CAA})$.

The free energies of mutation (U→C/A→G) in solution (ΔG_{free}) and complex (ΔG_{comp}) for a given state are used to calculate the relative binding free energies (**Figure 3.3**). The sign and magnitude of $\Delta\Delta G_{bind}$ indicate preference and magnitude of selecitvity.

$\Delta G(I \rightarrow F) = G_F - G_I = -\beta^{-1} \sum_{m=1}^{n-1} ln \langle exp[-\beta(V_{m+1} - V_m)] \rangle_m$, where $V_m = (1 - \lambda_m)V_I + \lambda_m V_F$ with λ_m varying from 0 to 1 and the number of points along the coupling coordinate is $m = 1, \ldots, (n-1)$. β is $k_B T$, where k_B is the Boltzmann constants and T is the temperature. Forward and backward runs are averaged for computing the total free energy change.

The previously described free energy perturbation (FEP) approach (**Sund et al., 2010; Satpati et al., 2014; Satpati & Åqvist, 2014**) has been used to calculate the corresponding free energies (ΔG_{free} and ΔG_{comp}). Using a coupling parameter λ, one nucleotide has been alchemically transformed into another (U→C/A→G); λ connects the initial (I) and final (F) state by a series of intermediate states. The total free energy change was obtained by summing over the intermediate states along the λ variable according to the following:

3.3 Results

3.3.1 Structure-based Energetics for Stop Codon Recognition

Relative binding free energies of eRF1 for the different codons in the ribosomal A-site are summarized in **Figure 3.4**. The results clearly indicate that eRF1 imposes a very high energetic penalty of about 8.5 kcal/mol for misreading CAA (codes for Gln), which corresponds to a probability of 10^{-6} relative to reading UAA. A uniformly low discrimination of about 2 kcal/mol preferring UAA relative to UAG, UGA has been obtained. This is expected due to the ability of eRF1 to read all three UAA, UAG, UGA stop codons. Furthermore, eRF1 discriminates UGG (codes for Trp) against UAA (stop) codon by 8.5 kcal/mol. It is interesting to note that eRF1 discriminates CAA (Gln) and UGG (Trp) with equal strength with respect to UAA (Stop). The discriminatory power of eRF1 (ΔΔG ~ 8.5 kcal/mol) is much larger compared to prokaryotic RF1, RF2 (< 5 kcal/mol) (**Sund et al., 2010**). The discrimination strength could be utilized for preferential dissociation of the higher energetic near-cognate complex. The large energetic penalty for misreading sense codon could further be boosted by higher k_{cat} for the stop codons.

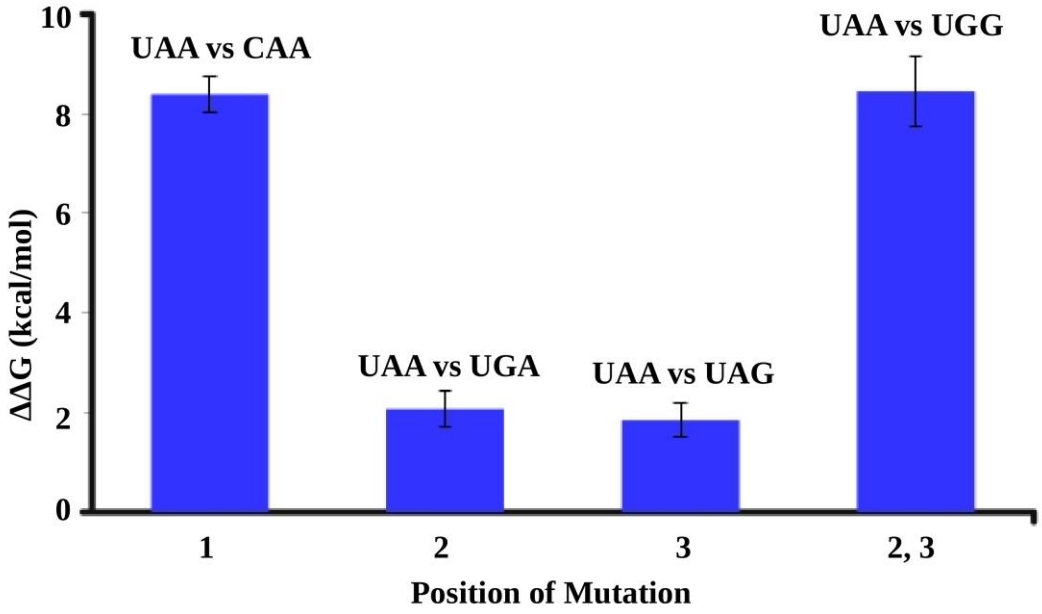


Figure 3.4. Computed binding free energy difference of eRF1 upon stop codon mutation. Error bars are one standard error of the mean.

3.3.2 Insights from Average MD Structures

Our MD averaged structures (**Figure 3.5, a**) are very similar to their corresponding cryo-EM structures and the RMSD of heavy atoms are between 1-1.4 Å (**Figure 3.5, b**). The eukaryotic translation complex is distinctly different from its bacterial analogue (**Figure 3.5, c**). The following features have been seen in our MD simulations (**Figure 3.6**): (a) There is a ~ 90^0 rotation of U1 with respect to N1-C1' bond (**Figure 3.6**), resulting a strong hydrogen bonding interaction between N3 of U1 and the phosphate backbone of fourth mRNA base, G4 (**Figure 3.8, a, b**). (b) Mg^{2+} forms a direct (mostly monodentate) and water-mediated interaction with the Glu55 side chain and Hoogsteen edge of second base, respectively (**Figure 3.6**). (c) Compacted stop codons are in a desolvated dry pocket of eRF1-ribosome complex with only two water molecules (w1, w2) at very specific locations, where w1 is absolutely conserved in all the simulations (**Figure 3.5** and **Figure 3.6** for structural context and **Figure 3.7** for solvation). The above structural features are insensitive to the initial structural models, force fields, and different solvation models used in these MD simulations (See **Methods**).

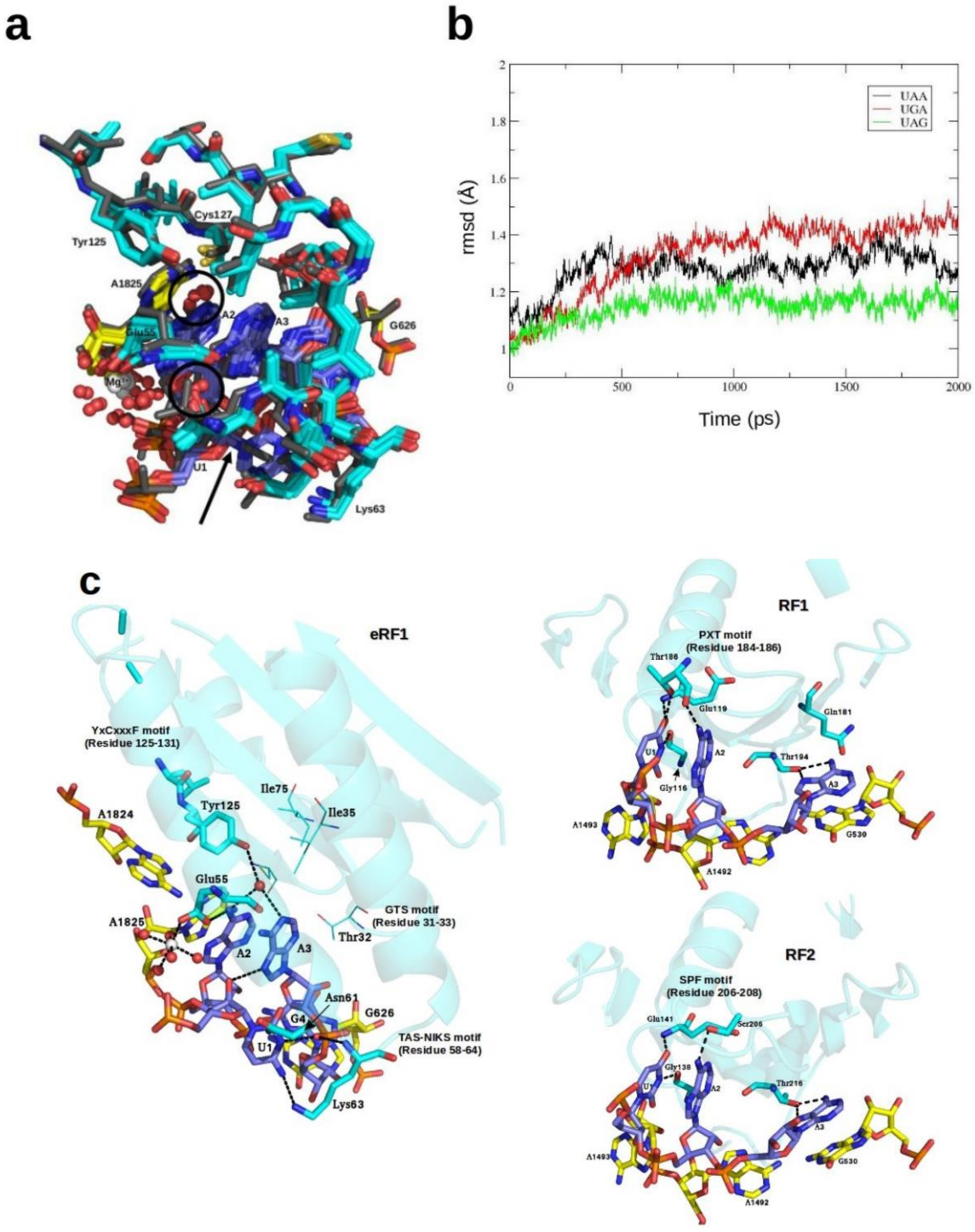


Figure 3.5. (a) Overlay of 6 independent MD averaged snaps (Color code same as Figure 3.1) with the cryo-EM structure (Gray) of a termination complex with UAA stop codon. The location of specific water molecules are highlighted by dashed circles. The rotation of U1 around C1'-N1 bond is highlighted with an arrow. (b) Root-mean-square deviation (RMSD) of the heavy atoms (within 22Å of simulation sphere; see METHODS for details) with

respect to the cryo-EM structure during dynamics. (c) Interaction pattern between stop codon (UAA) and Release factor in eukaryotes (left) and bacteria (right). In the ribosomal A-site, eukaryotes form U turn geometry with 4 nucleotides UAA-G and bacteria adopts a triplet stop codon where stacking of the third base on first two is disrupted. For eukaryotes, three motifs (YxCxxxF, GTS and TAS-NIKS) along with Glu, hydrophobic Ile's are involved in recognizing stop codons. In prokaryotes, PXT, SPF motif of RF1, RF2 along with Thr, Ser and Glu are involved in stop codon recognition. Ribosomal residues form very different interaction patterns for eukaryotes (left) and bacterial (right) termination complex.

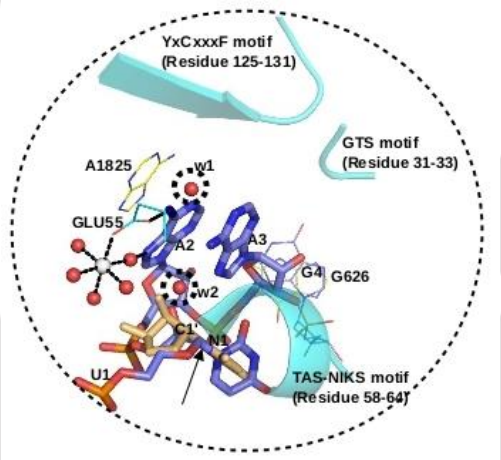


Figure 3.6. Zoom into ribosomal A site. Robust feature of ribosomal A-site in translation termination: Rotation of U1 about C1'-N1 bonds (highlighted by an arrow) with respect to the cryo-EM geometry (orange sticks), dry desolvated binding pocket with two waters (w1,w2) in very specific locations, w1 is absolutely conserved in all MD trajectories.

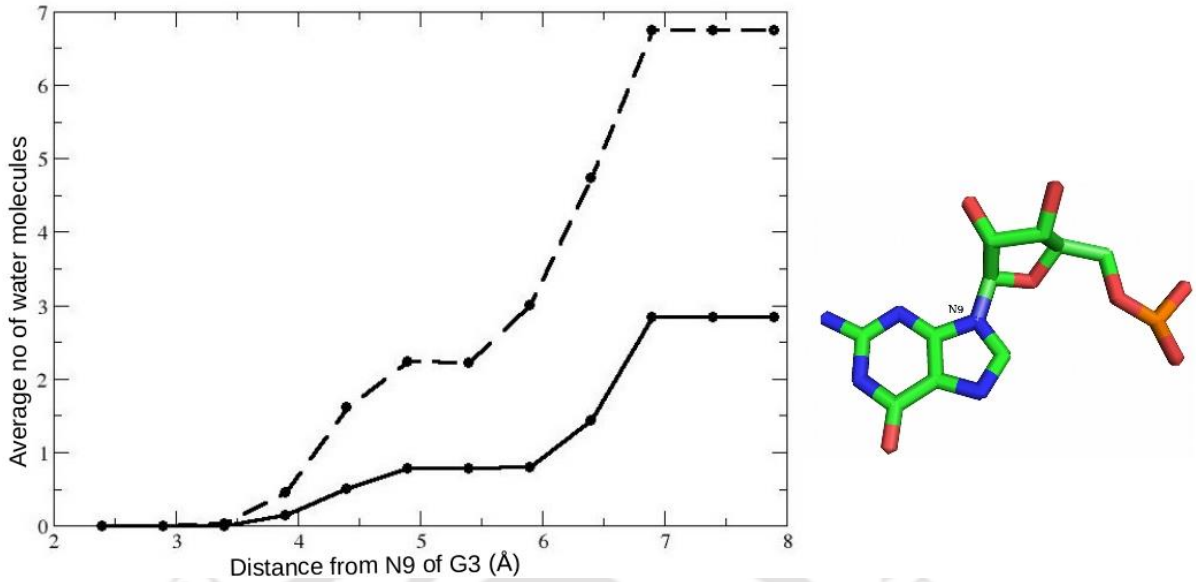


Figure 3.7. Water density around the N9 of G3. Results are shown for the UAG programmed ribosome with eRF1 (solid line) and without eRF1 (dashed line). The water numbers are averaged in radial shells over the corresponding MD trajectories.

3.3.3 Structural Insights for First Position Discrimination

In our MD-averaged structure with the UAA stop codon, flipping of the first U1 base results in strong hydrogen bond between N3(U1) and phosphate backbone (O1P) of foutrh mRNA base G4 (**Figure 3.8, b, Figure 3.9**), losing the proposed (**Brown et al., 2015**) interaction with Asn61 (**Figure 3.8, a, b**). The observed interaction (**Figure 3.8, b**) between N3(U1) and O1P(G4), has also been suggested based on a slightly lower resolution (resolution ~3.8Å) cryo-EM structure (**Matheisl et al., 2015**). Interaction between side chain of Lys63 and U1 is present as suggested (**Brown et al., 2015; Matheisl et al., 2015**). Direct interaction between U1 and Lys63 has also been suggested based on the chemical cross linking experiment (**Conard et al., 2012**). We have observed that backbone -NH of Lys63 forms a strong hydrogen bond with the phosphate of the fourth mRNA base in the A-site (**Figure 3.8, b, Figure 3.9**). This interaction is important in locking the phosphate backbone of fourth mRNA base, positioning the negatively charged oxygen towards the N3 of U1 (**Figure 3.8, b, Figure 3.9**).

Upon U1/C1 mutation, the amine group of C1 interacts with negatively charged phosphate backbone of the fourth base, positioning the WC edge into completely dry pocket and losing the interaction with TAS-NIKS motif of eRF1 (**Figure 3.8, c**).

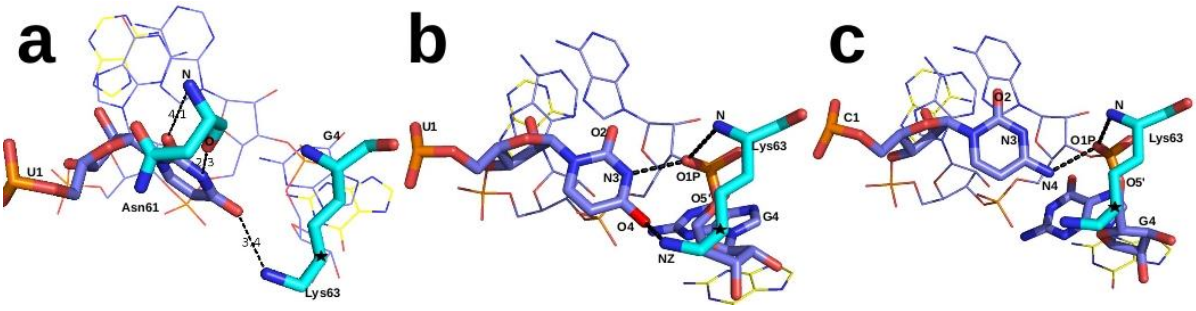


Figure 3.8. First position specificity, UAA vs CAA. Comparison between cryo-EM template structure, with averaged molecular dynamics structure of cognate and near-cognate codons. Color coding is the same as Figure 3.1, and hydrogen atoms are not shown for clarity. Distances are in Ångström (Å) and represented by dotted lines. Star represents the post-translational modification at the C4 position of Lys63. (a) Proposed interaction between U1 and TAS-NIKS motif (Lys63, Asn61) based on the cryo-EM structure with UAA. (b) MD-averaged structure of termination complex with UAA. (c) MD-averaged structure of termination complex with near-cognate CAA codon.

This is energetically unfavorable, leading to high discrimination. eRF1 free MD simulations suggest that, C1 can relax from its compacted U-turn conformation and form hydrogen bonds with the solvent molecules (**Figure 3.10**). Hence, eRF1 penalizes C1 by placing it in a tight hydrophobic pocket with unsatisfied hydrogen bonds. Interestingly, MD trajectories shows that the distance between N3(U1)-O1P(G4) is smaller (also less fluctuating) than O4(U1)-NZ(Lys63) (**Figure 3.9**). Hence, intra mRNA interactions N3(U1)-O1P(G4) may be the key factor for U1 specificity. It appears that the role of Lys63 is two fold: (a) optimal positioning of the phosphate oxygen (O1P) of 4th mRNA base to facilitate the interaction with first base and (b) locking the first base in the compacted U turn conformation. Substituted C1 is locked with a single hydrogen bond in a dry desolvated pocket. This is the origin of high energetic penalty for first position discrimination. It has also been suggested from experimental studies that hydroxylation at C4 position of Lys63 (**Figure 3.8**) can improve translation termination efficiency (**Feng et al., 2014**). Averaging over MD trajectories gives ~4.7Å distance between C4 (Lys63) and O5' (G4). It appears from our MD averaged structures that hydroxylation at C4 of Lys63 will lead to hydrogen bonding between hydroxyl group and O5' of G4. Hence, post translational hydroxylation of G4

would most likely be involved in orienting O1P of G4 for facile interaction with U1. Desolvation and intra mRNA interaction provide the key elements for first position discrimination along with loss of protein-mRNA interactions.

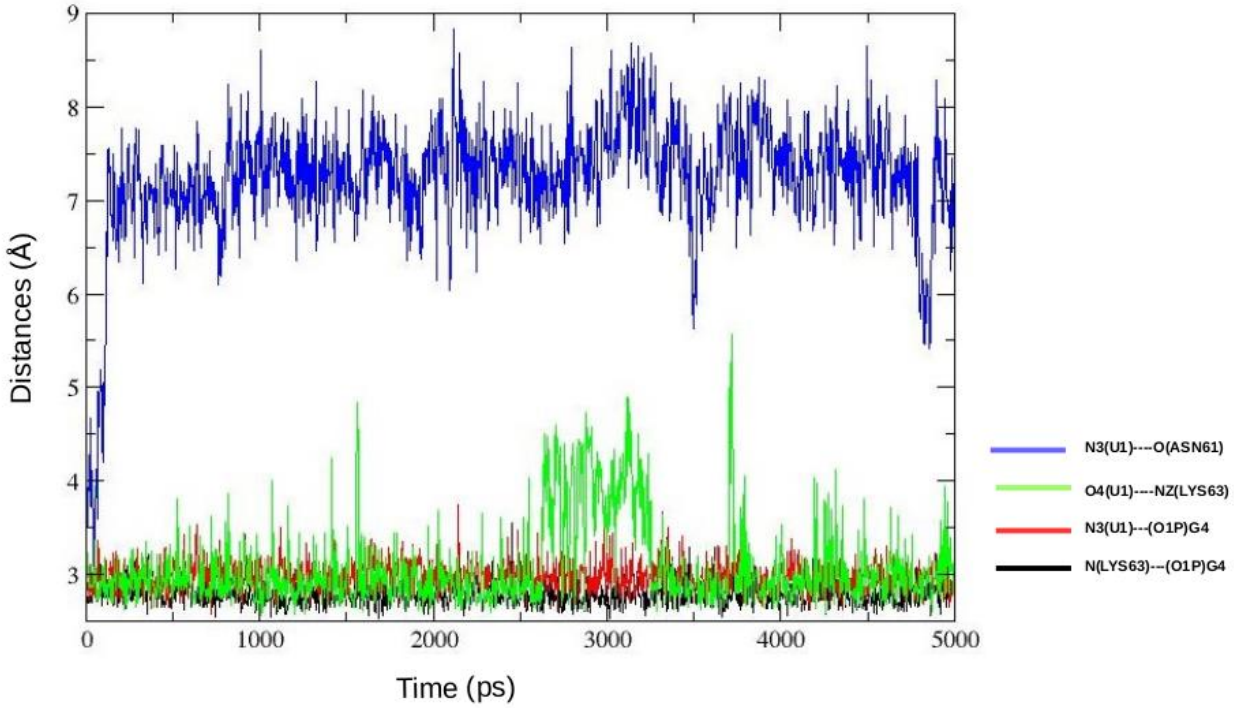


Figure 3.9. Distance between the heavy atoms of U1 and Lys63/Asn61 along the 5 ns MD trajectory. Frames are analyzed in 4 ps interval. Strong intra mRNA interaction between U1 and G4 (red), Lys63 and G4 interaction (black), Lys63 and U1 interaction (green), loss of interaction between Asn61 and U1 (Blue line).

In case of prokaryotic translation termination, U1 forms hydrogen bonds with the backbone atoms of RF1/RF2, which is reported to be the key factors for fidelity (**Sund et al., 2014**). Involvement of charge residues in eRF1-U1 interaction contributes towards higher discriminatory strength. It can be argued that purines in this compact geometry will be highly disfavored due to steric reasons as previously suggested (**Brown et al., 2015; Matheisl et al., 2015**).

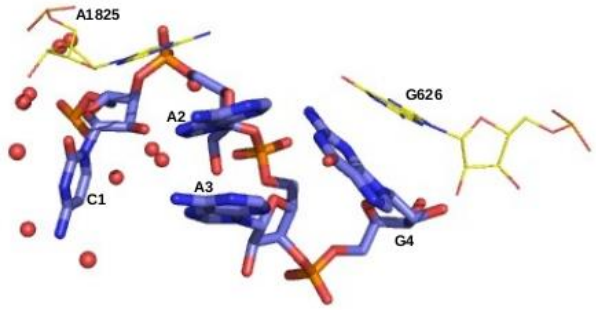


Figure 3.10. MD structure of eRF1 free four nucleotide CAA(G) codon in the ribosomal A site. Waters around 3Å of C1 are shown as red sphere. Color coding is same as Figure 3.1.

3.3.4 Structural Insights for Second and Third Positions Discrimination

The averaged MD structure of an eRF1-bound UAA codon in a ribosome suggests a single water molecule trapping by Tyr125 of the YxCxxxF motif (for structural context, see **Figure 3.3** and **Figure 3.6, c**). This water is very crucial for solvating the WC edge of A2A3 (**Figure 3.7, a**). It appears that the role of Tyr125 is to satisfy the hydrogen-bonding requirement of the WC edge of A2 and A3 by recruiting a single water molecule and contributing to the overall hydrophobicity of the binding pocket. The backbone of Cys127 forms two hydrogen bonds with A1825 as suggested (not shown for simplicity) (**Brown et al., 2015**). This stabilizes A1825's flipped orientation for a favorable stacking interaction with a +2 base. The side chain of Cys127 of YxCxxxF appears to be stabilizing the hydrophobic pocket rather than forming hydrogen bonds with the stop codon.

It has been reported that Cys side chain has stronger hydrophobic stabilization in micelles than Gly/Ser (**Heitmann, 1968**). The Hoogsteen edge of A2 forms water-mediated interaction with the side chain of Glu55/Mg^{2+} (**Figure 3.11, a**). The Hoogsteen edge of A3 forms direct/water-mediated interaction with ribsome -OH of U1 and side chain of Thr58 of TAS-NIKS motif (**Figure 3.11, a**). Hydrophobic residues (Tyr125, Cys127, Ile35, Ile75) of eRF1 creates the hydrophobic environment for WC edge of A2 and A3 (**Figure 3.11, a**). It should be noted that according to our simulations the side chain of Glu55 is involved in direct interaction only with A2, not with both A2 and A3 as suggested earlier (**Brown et al., 2015**). In UGA, the Hoogsteen edge of G2 (i.e., O6/N7) forms a water-mediated interaction with Mg^{2+}, while the WC edge of G2 forms a direct interaction with the backbone oxygen

of Cys127 of the YxCxxxF motif (**Figure 3.11, b**). Tyr125 together with Glu55 holds the water molecule to solvate A3 (**Figure 3.11, b**). The Hoogsteen edge of A3 satisfies the hydrogen bonding requirement by interaction with ribose -OH of U1 and water-mediated interaction with Thr58. Direct interaction of A3 with U1 and Thr58 has also been observed in MD simulations of UAA and UGA complexes (**Figure 3.12, a ,b**) without W2 (**Figure 3.3**).

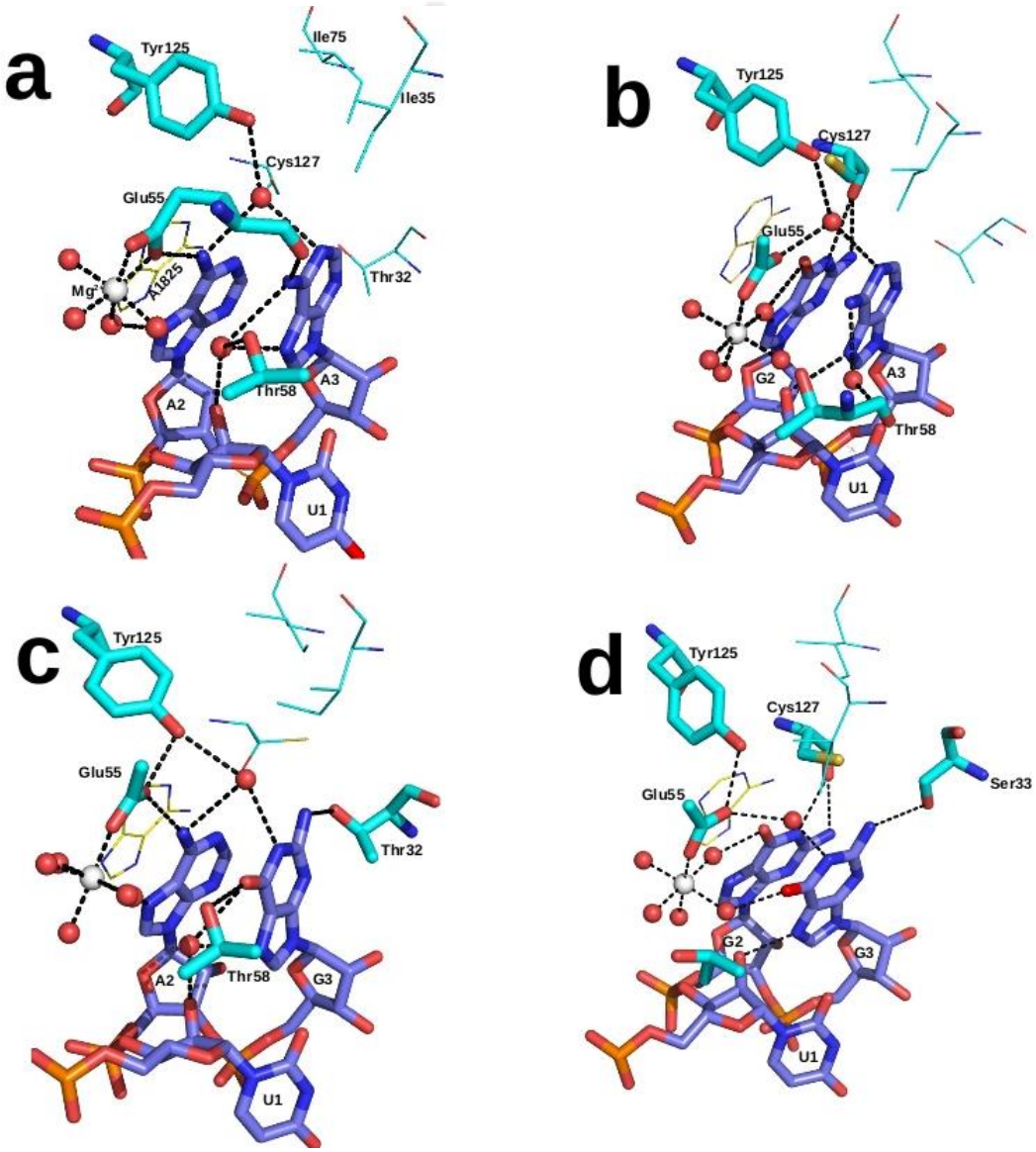


Figure 3.11. Second and Third position specificity. MD averaged structures of eukaryotic translation termination complex with different codons: (a) UAA, (b)UGA, (c)UAG vs (d) UGG.Color coding is the same as Figure 3.1. Residues involved in direct/indirect interactions with the second and third bases are in sticks. Hydrogen bondings are

represented by dotted lines, and hydrogen atoms are not shown for clarity. (a) MD-averaged structure of UAA stop codon-bound termination complex. Direct interaction between Glu55 and A2. Single water molecule (recruited by Tyr125) interacting with the WC edge of A2 and A3. Hoogsten edge of A2 and A3 forms hydrogen bonds with water molecules. (b,c,d) Presence of relatively high polar G in the second or third position opens up the binding pocket allowing water molecules to interact with the Hoogsteen edge of the base. However, the WC edge in the binding pocket is rather dry with a single water molecule. In UGG, the interaction between the WC edge of G2G3 with eRF1 is not enough for compensating the desolvation penalty.

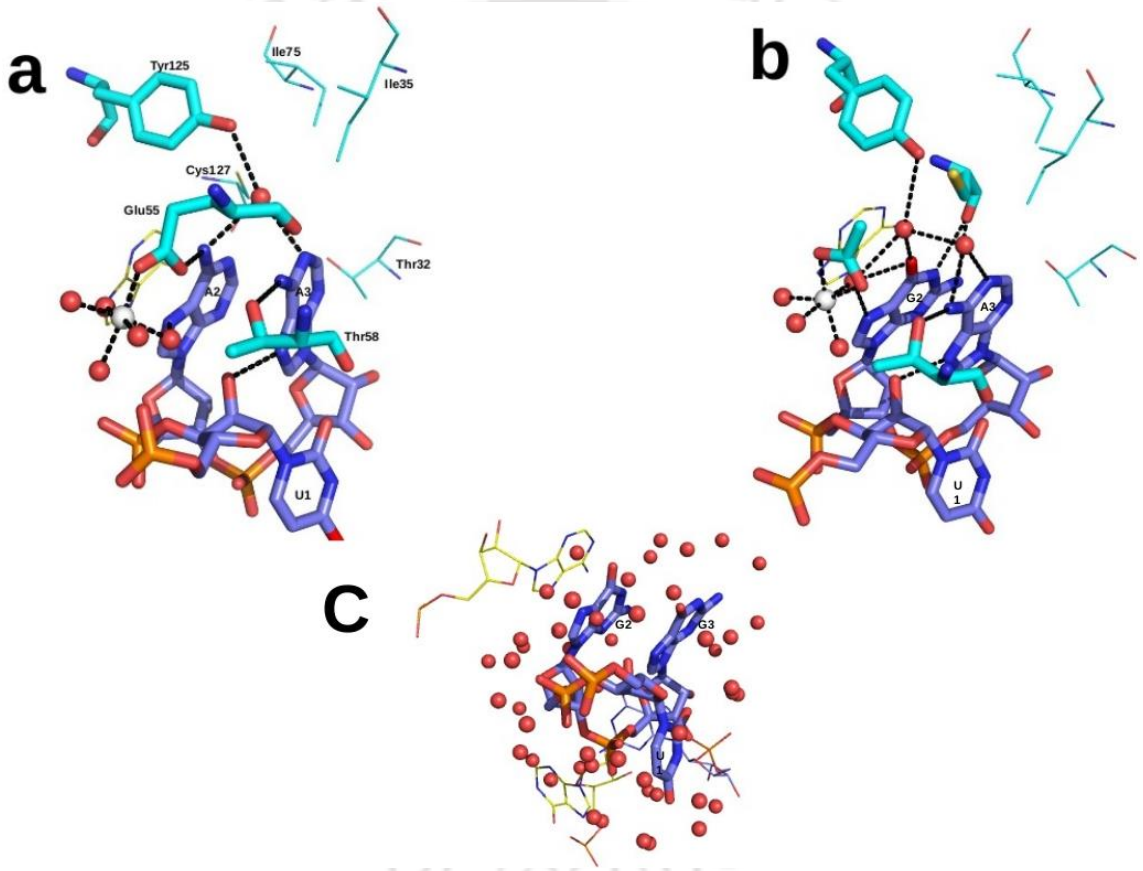


Figure 3.12. Different interaction network (a) UAA (b) UGA codons in the termination complex (c) eRF1 free UGG in ribosomal A site. (a,b) Direct interaction between hoogsteen edge of A3 with ribose -OH of U1 and side chain of Thr58. (c) eRF1 free four nucleotide UGG(G) codon in the ribosomal A site with many water molecules solvating the WC and hoogsteen edges of G2G3. Waters around 3Å of UGG are shown as red sphere. Color coding is same as Figure 3.1.

A2 in UAA and UAG shows very similar interaction pattern (**Figure 3.11, a,c**). In UAG, the Hoogsteen edge of G3 compensates its desolvation by interacting with Thr58 and U1, whereas the exocyclic amime forms a hydrogen bond with Thr32 of GTS motif (**Figure 3.11, c**). It is interesting to note that presence of G in the second or third position opens up the binding pocket, and the loss of protein-mRNA interaction is compensated by interacting with water molecule in the Hoogsteen edge. However, the WC edge is dry irrespective of the nature of purines in the second and third position of the stop codon. It appears that Tyr125 and Glu55 acts as a pair in recruiting a single water molecule in the WC edge and stabilizing the binding pocket. The direct/water-mediated interaction between ribose U1 and N7 of third base is very specific to purines and exclude pyrimidines at the third position.

In UGG, the direct interaction of the A site mRNA codon with Glu55 is lost (**Figure 3.11, d**) as predicted (**Brown et al., 2015; Matheisl et al., 2015**). The key factor for discrimination appears to be the desolvation of WC edge of G2 and G3. The direct/water interaction between guanines (at second and third stop codon position) and eRF1 is not enough for compensating the desolvation penalty (**Figure 3.12, c**). The large energetic penalty for UGG binding explains the strong discrimination of eRF1 against Trp sense codon.

3.4 Discussion

The codon reading strategy of eRF1 is different from bacterial RF1/RF2. The discriminatory power of eRF1 is greater than RF1/RF2 (**Sund et al., 2010**). The key effects of the eRF1 binding to a stop codon in the A-site of ribosome are three fold. First, as hypothesized from structural studies, loss of protein-mRNA interaction occurs in the near-cognate complexes. Second, compacted mRNA in U-turn conformation with four nucleotides is stabilized and locked into the eRF1-ribosome complex. This compact conformation of mRNA forms a specific intra-mRNA interaction, which can contribute to self recognition. Lastly, stop codons in the ribosomal A-site are embedded into a hydrophobic environment created by the hydrophobic residues of eRF1. Guanine is more polar than adenine. Therefore, placing guanine in the dry binding pocket is energetically unfavorable. Unsatisfied hydrogen bonding of guanine is known to have higher energetic penalty (**Satpati et al., 2014**) in a desolvated pocket. Stronger interaction with eRF1 is

required for compensating the desolvation penalty. MD simulations show that the WC edge of the second and third bases is relatively dry. GG forms direct/water mediated interaction with eRF1, however apparently not sufficient to compensate for the desolvation penalty. The large discrimination strength could be utilized efficiently for controlling the error frequency in translation termination. Near-cognate substrates would preferentially dissociate because of higher energy. Cognate substrates, on the other hand, would be stabilized and lead to extremely low rate of dissociation.

Base flipping within DNA and RNA is a common event in the cell. These flipping events occur on a millisecond time-scale and may be involved in replication, transcription, and translation processes. To better understand base-flipping/breathing, a lot of studies have been carried out. A recent study (**Fuxreiter et al., 2002**) suggested that the increased flexibility of DNA lowers the energy barrier for flipping. In another study, MacKerell group has calculated free energy profiles and molecular mechanisms of base flipping in double-stranded DNA (**Banavali & MacKerell, 2002**). In our MD simulations, we have observed a ~ 90^0 rotation of U1 with respect to N1-C1' bond (**Figure 3.6**). However, the estimation of free energy associated with U1 rotation/flipping is out of the scope of this study. This can be a separate project altogether.

Protein posttranslational modifications play a pivotal role in modulating proteins' structural and functional properties for the regulation of numerous cellular processes (**Ardito et al., 2017**). Experimental studies suggested that hydroxylation at the C4 position of Lys63 (**Figure 3.8**) can improve translation termination efficiency (**Feng et al., 2014**). A recent study using MD simulations have also shown that posttranslational phosphorylation of initiation factor can regulate the frequency of mRNA translation (**Lama & Verma, 2019**).

Almost all biological mechanisms in the cells occur in water, and these polar water molecules contribute to the formation of hydrogen bonds, thus considering the contribution of the water molecules is essential in stop codon recognition by eRF1. In this study, we have shown that the eRF1-stop codon complex is embedded into a rather dry desolvated ribosomal A site, which is key for boosting the discriminatory power of stop codon selection relative to sense codon. The importance of water in protein-DNA recognition was also highlighted previously (**Ho et al., 2008**).

3.5 Conclusion

In conclusion, MD simulations of cognate and near cognate eukaryotic termination complexes provide crucial insights into the energetics of stop codon selection. The computed strength of discrimination between sense (CAA and UGG) and the stop codons (UAA,UAG, and UGA) by eRF1 is ~ 8 kcal/mol. This corresponds to a low probability of misreading (~ 10^{-6}). Although the experimental relative binding free energy is not available, but the sign obtained here is correct since eRF1 preferentially binds stop codons. The magnitude of discrimination strength appears to be biochemically plausible. The eRF1-stop codon complex is embedded into a rather dry desolvated ribosomal A site. Sense codons are discriminated strongly due to loss of intra-mRNA and protein-mRNA interactions and desolvation. The results not only give insights into the near-cognate complexes but also provide the link between structures and computed energetics, illustrating the versatility of eRF1 in recognizing all three stop codons.

Chapter 4

Viral RNA Recognition by Retinoic Acid inducible Gene-I

Retinoic acid-inducible gene-I (RIG-I) receptor protein recognizes viral RNAs in the cell cytosol and triggers an innate immune response against viral infection, but it is unresponsive to host RNA's. Panhandle like base-paired blunt-ended double-stranded RNA with 5' tri/di-phosphate (5' ppp/pp-dsRNA) is a characteristic feature of viral RNA's which is different from host RNA's which are usually post-transcriptionally modified and contain 5' monophosphate/hydroxyl group (5' p/OH-dsRNA). Biochemical studies suggest that 5' pp-dsRNA is the minimal requirement for RIG-I mediated antiviral response. Experiments further confirmed that indeed RIG-I binds to viral RNA's (5' ppp/pp-dsRNA) with a very high affinity relative to host RNA's (5' p/OH-dsRNA). Surprisingly, 5' p-dsRNA binds to RIG-I with strikingly low affinity, and the binding affinity is much weaker than 5' OH-dsRNA. Preference of 5' OH-dsRNA binding over 5' p-dsRNA by RIG-I is a characteristic feature of RIG-I receptor. The energetic origin of RNA discrimination based on the variation of 5' terminal (5' ppp/pp/p/OH) by RIG-I is still unclear and poorly understood. This chapter has been divided into the following two parts which attempt to answer the following questions:

Part 1: Why 5' pp-dsRNA is the minimum requirement for RIG-I mediated antiviral response?

Part 2: Why 5' p-dsRNA is a poor binder to RIG-I with respect to 5' OH-dsRNA: characteristic feature of RIG-I receptor?

4.1 Energetics of Preferential Binding of RIG-I to Double-stranded Viral RNAs with 5' tri/diphosphate Over 5' Monophosphate

Retinoic acid-inducible gene-I (RIG-I) is a cytosolic sensor protein that recognizes viral RNAs and triggers an innate immune response in cells. Panhandle like base paired blunt ended 5' ppp/pp-dsRNA is a characteristic feature of viral RNAs. Structural studies of RIG-I C-terminal domain (CTD) bound 5' ppp/pp-dsRNA complexes show the direct interaction between all the 5' terminal phosphates (α, β, and γ) and protein, suggesting γ phosphate might be a major recognition determinant for RIG-I binding. Biochemical studies, however, suggest that 5' pp-dsRNA is the minimal determinant for RIG-I binding and antiviral response. Despite biochemical and structural studies, the origin of viral RNA recognition by RIG-I is an unsolved problem. X-ray structures of RIG-I bound dsRNA not only provide atomic insight into the interaction network but also provide sufficiently good models for computational studies. We report structure-based molecular dynamics free energy calculations to quantitatively estimate the energetics of RIG-I binding to dsRNA containing 5' ppp, 5' pp and 5' p. The results suggest that RIG-I weakly discriminates between 5' ppp-dsRNA and 5' pp-dsRNA (favoring former) and strongly disfavors 5' p-dsRNA with respect to the rest. Interestingly direct interaction between γ phosphate of 5’ ppp-dsRNA and RIG-I is a robust feature of the MD simulations. dsRNA binding to RIG-I is associated with Mg^{2+} dissociation from the 5’ phosphate/s of dsRNA. The higher Mg^{2+} dissociation penalty from 5’ ppp-dsRNA with respect to 5’ pp-dsRNA offsets most of the favorable interaction between RIG-I and γ phosphate of 5’ ppp-dsRNA. This leads to weak discrimination between 5’ ppp-dsRNA and 5’ pp-dsRNA. 5’ p-dsRNA is discriminated strongly due to the loss of interaction with RIG-I.

==========

The work published in ACS Omega, DOI:10.1021/acsomega.7b02019

4.1.1 Background

RIG-I is a major pathogen recognition receptor that recognizes a broad range of viruses (e.g, influenza, rabies, dengue, hepatitis C etc) and triggers an antiviral response in the cytoplasm (**Kato et al., 2011; Yoneyama et al., 2004; Kato et al., 2006; Saito et al., 2008; Kawai & Akira, 2006; Ramos & Gale, 2011**). As discussed in **Chapter 1**, Panhandle like blunt-ended dsRNA with 5' triphosphate is a signature of viral RNA (see **Figure 1.10**), which is preferentially recognized by RIG-I (**Hornung et al., 2006; Pichlmair et al., 2006; Schlee et al., 2008**). As already described in **Chapter 1**, RIG-I is composed of a C-terminal domain (CTD), helicase domains (HEL1, HEL2, and HEL2i), Pincer domain (P) and N-terminal caspase activation and recruitment domains (CARD) (see **Figure 1.9** and **Figure 4.1** CARD not shown). CTD of RIG-I is a beta sheet bundle stabilized by Zn^{2+} ion (**Figure 4.1**); that recognizes the viral RNA by interacting with the 5' tri/di phosphates of dsRNA (**Cui et al., 2008; Lu et al., 2010; Wang et al., 2010**).

Structural studies (**Lu et al., 2010**) suggest that the structure of CTD of RIG-I is very similar in complex and in isolation, except for the loop region (residue 847-853, **Figure 4.1**). The free CTD structures in solution (**Cui et al., 2008; Takahasi et al., 2008**) have revealed that the loop is very flexible and comparison with the dsRNA bound structures (**Lu et al., 2010**) suggest that dsRNA binding might stabilize the specific conformation of the loop.

X-ray structures (pdb 3LRR, 3LRN) (**Lu et al., 2010**) of RIG-I CTD bound 5' ppp-dsRNA have revealed extensive interactions between α, β, γ-phosphates of dsRNA and RIG-I, whereas a full RIG-I(ΔCARDs) bound 5' ppp-dsRNA structure (pdb 4AY2) suggests (**Luo et al., 2012**) no direct interaction between γ-phosphate and RIG-I. The conflicting γ-phosphate - RIG-I interaction is also limited by the resolution of the x-ray structures. It can be concluded that out of 3 X-ray structures (pdb 3LRR, 3LRN, 4AY2), first two structures suggest a direct interaction between γ phosphate and RIG-I. Thus, γ phosphate appears to be a major recognition factor for RIG-I binding and RIG-I should discriminate strongly between 5' ppp-dsRNA and 5' pp-dsRNA, preferring the former. It is worth mentioning that none of the resolved structures of the complexes contain Mg^{2+} bound to the terminal 5' tri/di phosphates of dsRNA.

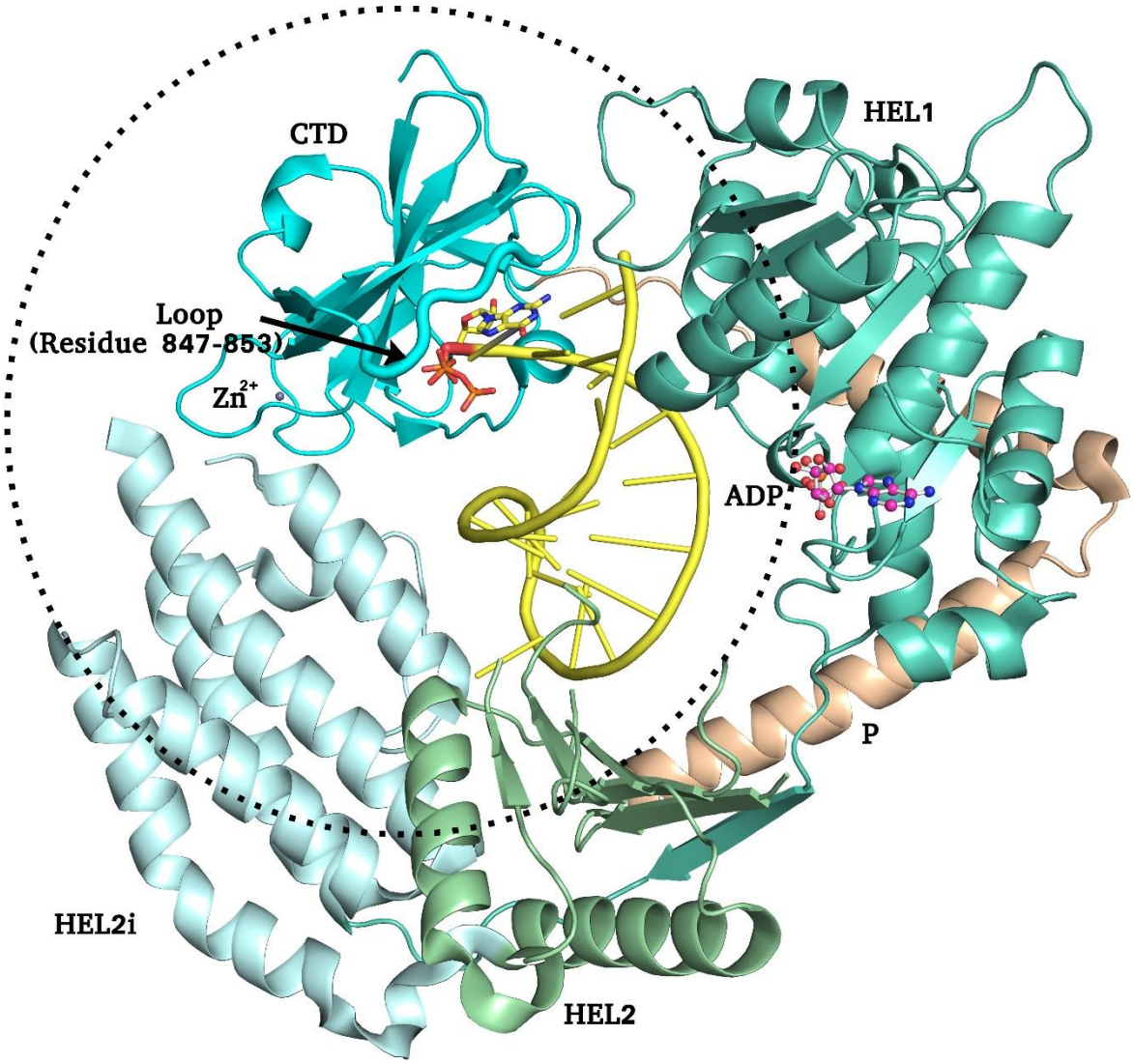


Figure 4.1: Structural overview of blunt-ended 5' ppp-dsRNA bound RIG-I (containing helicase and C-terminal domains) complex (pdb: 4AY2) (**Luo et al., 2012**). Individual domains of RIG are color coded: CTD: cyan, HEL1: Dark Green, HEL2: Light Green, HEL2i: Light Blue, Pincer (P): orange. Loop spanning position 847-853 of CTD is shown as thick tube and highlighted with an arrow. Zn^{2+} of the zinc-binding domain of CTD is represented as a cyan sphere and ADP-Mg^{2+} bound to the HEL-1 domain is in the magenta sticks. Blunt ended dsRNA is shown in yellow and the guanine base at the 5′ end with triphosphate is shown as sticks. Same color coding is used throughout this paper. The spherical region selected for molecular dynamics simulation is indicated in the black broken circle.

Recent biochemical studies (**Goubau et al., 2014**) have revealed that blunt-ended dsRNA with 5' pp-dsRNA is the minimum requirement for RIG-I binding and antiviral response. Both 5' ppp-dsRNA and 5' pp-dsRNA can induce antiviral response weakly preferring the former; hence, γ phosphate is certainly not a recognition determinant. Biochemical studies further reveal that isolated RIG-I domains bind more strongly to viral RNA with respect to the full RIG-I (**Vela et al., 2012**). It has been suggested (**Vela et al., 2012**) that covalently linked RIG-I domains reduce the overall binding affinity but increase the specificity for efficient discrimination of viral RNAs from host RNAs.

Despite the advancement of structural and biochemical studies, the atomic insight into the dynamics of these complexes are unknown and, the following key questions on dsRNA binding to RIG-I remain unanswered: (a) How strongly RIG-I discriminates between 5' ppp-dsRNA, 5' pp-dsRNA and 5' p-dsRNA (i.e. relative binding affinity)? (b) What is the relationship between relative binding affinity and the 3D structures? (c) Why γ phosphate is not a major recognition factor for RIG-I binding?

Medium resolution x-ray structures (**Cui et al., 2008; Luo et al., 2012**) now provide sufficiently good models for structure-based computer simulations for addressing the above questions. We report structure-based molecular dynamics free energy simulations for deciphering the energetics of dsRNA binding to RIG-I, thereby linking 3D structures and energetics. Starting with x-ray (**Luo et al., 2012**) structure (pdb: 4AY2, Full RIG-I(ΔCARDs): 5’ ppp-dsRNA, resolution 2.8Å) as our initial model, we have performed molecular dynamics free energy simulations of RIG-I: 5’ ppp-dsRNA, RIG-I: 5’ pp-dsRNA and RIG-I: 5’ p-dsRNA complexes. Calculations involve the change in binding affinity upon “mutation” of the 5’ terminal of dsRNA using an appropriate thermodynamic cycle shown in **Figure 4.3**. The calculations quantitatively estimated the binding affinity difference between 5’ ppp-dsRNA, 5’ pp-dsRNA and 5’ p-dsRNA to RIG-I binding. Our estimated binding free energy difference between 5’ ppp-dsRNA/5’ pp-dsRNA is small (~ 2 kcal/mol in favor of 5’ ppp-dsRNA) and 5’ ppp-dsRNA/5’ p-dsRNA is large (~ 9 kcal/mol favoring 5’ ppp-dsRNA). The signs are consistent with the experimental observations and the magnitudes seem biochemically sensible. dsRNA binding to RIG-I is associated with Mg^{2+} dissociation from 5’ terminal of dsRNA, desolvation and protein-RNA interactions.

Our simulations suggest that γ phosphate establish direct electrostatic contact with RIG-I. However, higher Mg^{2+} dissociation penalty from 5' ppp-dsRNA with respect to 5' pp-dsRNA offsets most of the γ phosphate-RIG-I interaction, leading to the loss of binding affinity for RIG-I. This seems to be responsible for weak discrimination between 5' ppp-dsRNA and 5' pp-dsRNA, with RIG-I weakly preferring the former. The binding of 5' p-dsRNA is strongly disfavored due to the substantial loss of interaction between 5' terminal of dsRNA and RIG-I, leading to strong discrimination.

4.1.2 Methods

4.1.2.1 Molecular Dynamics Setup

Molecular dynamics setup is given in **Figure 2.1**, **Chapter 2**. Structure of human RIG-I bound to 5' ppp-dsRNA was taken from the Protein Data Bank (entry 4AY2 (**Luo et al., 2012**); crystallographic resolution 2.8 Å). We have selected 4AY2 as a template for MD for two reasons (1) The panhandle like bound RNA with 5' triphosphate is an excellent mimic of viral RNAs (2) The RIG-I in this complex contains CTD as well as the helicase domains. 30Å radii sphere, centered at the terminal phosphate of dsRNA was cut from the selected structure considered for MD simulations. We retained residues that had at least one non-hydrogen atom within the 30Å sphere. Non-hydrogen atoms of RIG-I and dsRNA in the outer region between 27Å-30Å from the sphere's center ("buffer region") were harmonically restrained to their experimentally determined positions. The restraints in the buffer region were increased gradually from 3.0 to 5.0 kcal/mol/$Å^2$ as one moves closer to the outer boundary, leaving the inner 27Å radius shell fully flexible. A cubic water box (edge length= 80Å) was overlaid for solvation and waters that overlapped with RIG-I/dsRNA were removed. We deleted the terminal phosphate/s from 4AY2 and considered the resulting complex as the initial model of RIG-I bound 5' pp-dsRNA/5' p-dsRNA. The MD structures were compared with the x-ray structures of RIG-I CTD bound 5' pp-dsRNA and found them to be essentially the same. We have modelled RIG-I bound 5' p-dsRNA complex and performed MD simulations. The total number of atoms in our simulation model is about ~50100. The number of water molecules present in our MD is about 15000. Root-mean-square deviation (RMSD) of the heavy atoms (within 27Å of simulation sphere) of the complex with respect to the x-ray structure (pdb 4AY2) is given in **Figure 4.2**. Average

RMSD was calculated for the heavy atoms within 27Å of the simulation sphere, averaging over the 5 ns MD trajectory with 2ps interval.

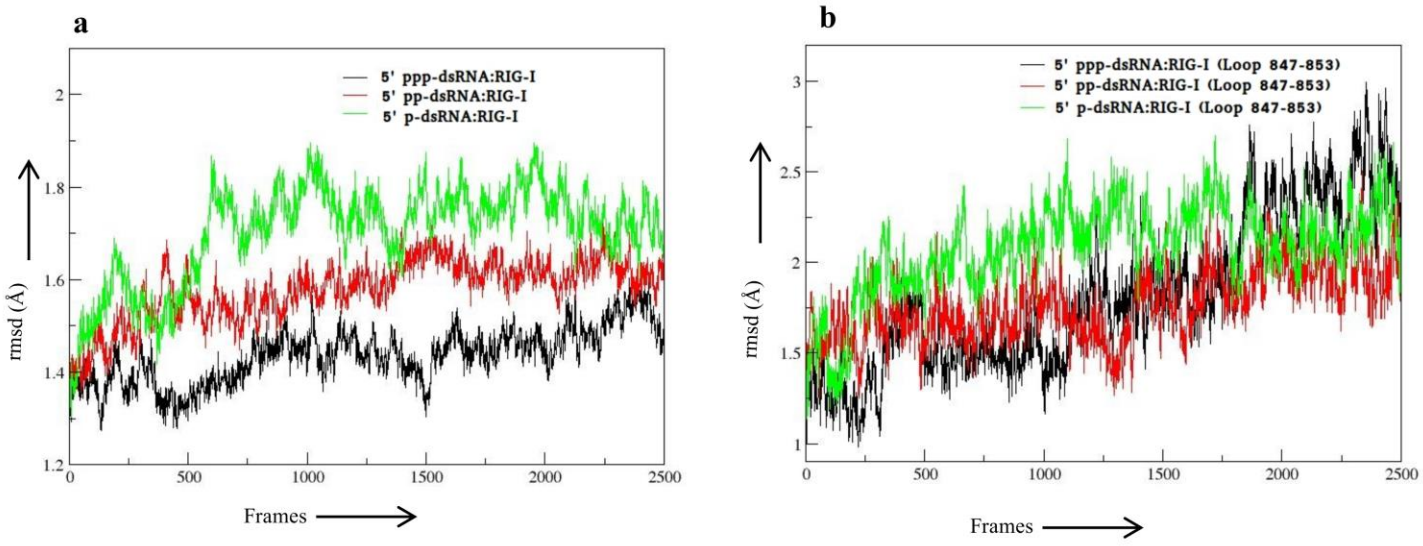


Figure 4.2: (a) Root-mean-square deviation of heavy atoms within 27Å of unrestrained simulation sphere with respect to x-ray structure. Graphs are shown for RIG-I bound 5' ppp-dsRNA(black)/5' pp-dsRNA(red)/5' p-dsRNA(green) complexes. RMSD is shown for 5ns MD trajectory averaging over 2ps interval. (b) RMSD of flexible loop region (residue 847-853).

The RMSD data suggested that the MD structures are very similar to the x-ray structure 4AY2. Largest RMSD for 5' p-dsRNA bound complex is expected as the models are generated by deleting the γ and β phosphates from 5' ppp-dsRNA bound RIG-I (pdb: 4AY2).

Simulation details are same as discussed in **section 2.2.1**. We performed 275-300ns of production dynamics for each simulation model. The simulations involve 250-300 ps of equilibration through a series of short MD runs. In the first 40 ps of equilibration, the system was heated up to 310K and then kept fixed throughout the production MD trajectories. At the initial stage of equilibration, heavy atoms of inner region (within 27Å) were harmonically retrained at their experimentally resolved position with a force constant of 4.0 kcal/mol/Å^2, and at the final stage of equilibration, the restraint from the inner region was completely removed. The overall charge of the simulation was neutralized by scaling down the partial charges of the phosphate backbone of dsRNA. It should be noted that 5' ppp-dsRNA → 5' pp-dsRNA, 5' pp-dsRNA → 5' p-dsRNA transformation alters the overall

charge of the complex. The use of tinfoil boundary conditions implemented in NAMD (**Phillips et al., 2005**) ensures that, as the overall charge changes, a compensating charge density is spread uniformly throughout the simulation box and does not contribute to the forces (**Hummer et al., 1996; Bogusz et al., 1998**).

4.1.2.2 Protocol for Binding Free Energy Calculation

Relative binding free energies (ΔΔG) of 5’ ppp-dsRNA/5’ pp-dsRNA/5’ p-dsRNA binding to RIG-I were calculated by alchemically transforming a terminal phosphate into a ghost following the horizontal legs of the thermodynamic cycle in **Figure 4.3**. The vertical legs correspond to RNA binding. On the other hand, horizontal legs correspond to the alchemical transformation of dsRNA which cannot be realized experimentally. We computed the free energy change associated with the phosphate deletion (horizontal arms of **Figure 4.3**) and calculated the relative binding free energy as, $\Delta\Delta G_{bind} = \Delta G^{comp} - \Delta G^{free} = \Delta G_{bind}$(5’ pp-dsRNA) – ΔG_{bind}(5’ ppp-dsRNA). A hybrid energy function (**Simonson & Satpati, 2012; Satpati & Simonson, 2012**) (U) was used to represent a mixture of two endpoint states for a particular horizontal leg (**Figure 4.3**). Coupling coordinate λ was used to connect two endpoints. Two coupling coordinates, λ_{elec}, and λ_{vdw} were used to modify the electrostatic and van der Waals energy terms, where $U = U(\lambda_{elec}, \lambda_{vdw})$.

MD simulations gradually remove the atomic charges of the γ-phosphate of 5' ppp-dsRNA with simultaneous modification of the β-phosphate charges to model 5' pp-dsRNA according to the following equation: $U = \lambda_{elec}\ U_{elec}$(5’ ppp-dsRNA) + $(1-\lambda_{elec})\ U_{elec}$(5’ pp-dsRNA), where, U_{elec}(5’ ppp-dsRNA) and U_{elec}(5' pp-dsRNA) represents the Coulomb interactions involving 5' ppp-dsRNA and 5' pp-dsRNA charges respectively. Next, we varied the van der Waals interaction parameter of γ-phosphate of 5' ppp-dsRNA into those of 5' pp-dsRNA by modifying the coupling coordinate λ_{vdw} from 1 to 0, and leaving a ghost phosphate, using: $U = \lambda_{vdw}\ U_{vdw}$(5’ ppp-dsRNA)+ $(1-\lambda_{vdw})\ U_{vdw}$(5' pp-dsRNA). Free energy derivative from Boltzmann statistics was calculated as $\partial G/\partial\lambda = \langle \partial U/\partial\lambda \rangle_{\lambda}$, where $\lambda = \lambda_{elec}$ or λ_{vdw} and the brackets "< >" represents averaging over MD trajectory for a particular value of λ. For the alchemical transformation of the charges, we used 11 equally spaced λ_{elec} values between 1 and 0 (1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, and 0.0).

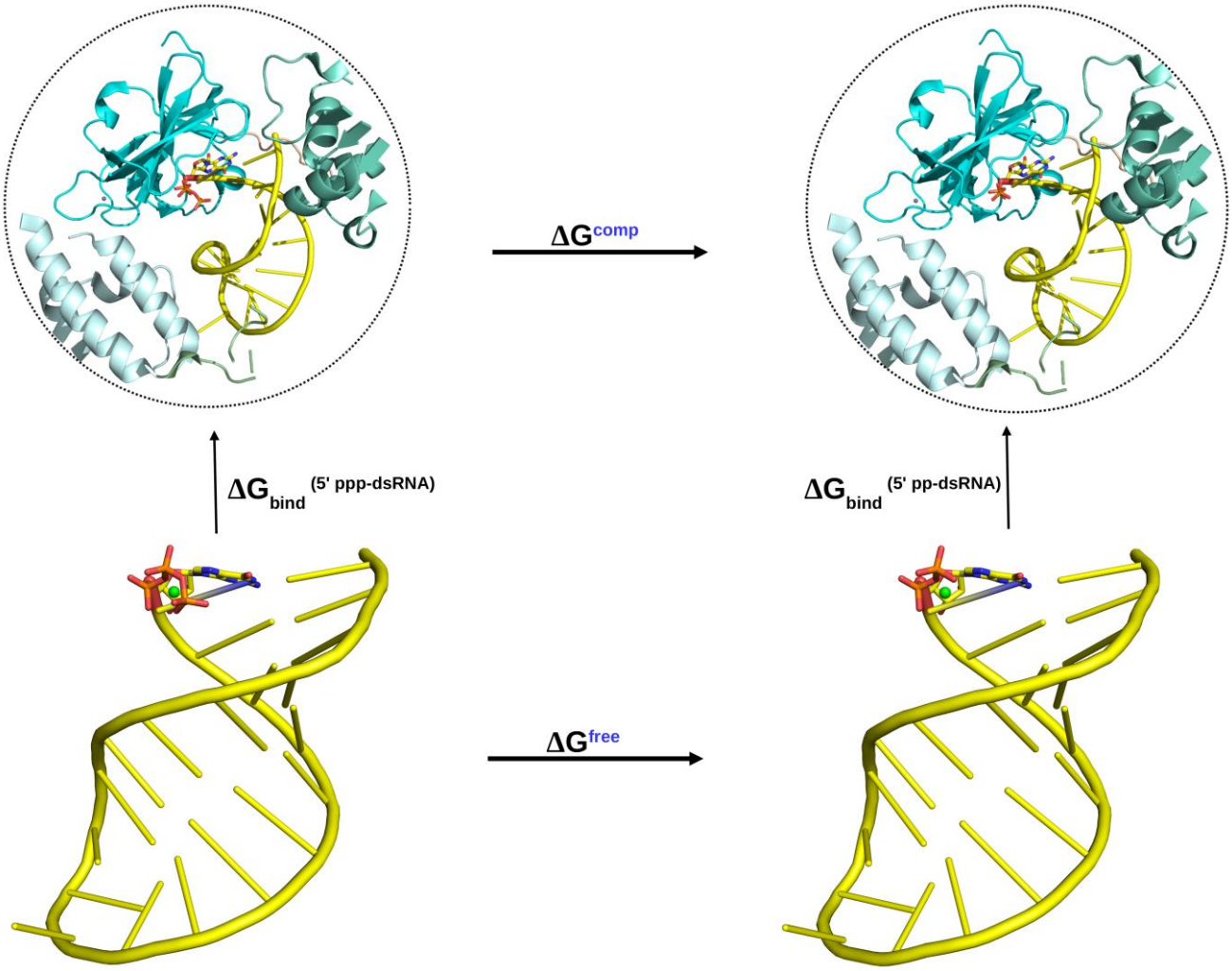


Figure 4.3. Thermodynamic cycle (TC) for dsRNA: RIG-I binding. Vertical legs correspond to binding; horizontal legs correspond to the alchemical transformation of 5' ppp-dsRNA into 5' pp-dsRNA, either in the solvated RIG-I complex (above) or in the free state in water (below). The 5' ppp-dsRNA/ 5' pp-dsRNA binding free energy difference is $\Delta\Delta G = \Delta G^{comp} - \Delta G^{free} = \Delta G_{bind}$(5' pp-dsRNA) - ΔG_{bind}(5' ppp-dsRNA). Horizontal legs of the thermodynamic cycle (ΔG^{comp}, ΔG^{free}) were computed by molecular dynamics free energy simulations. Mg^{2+} bound dsRNA is considered as the prevalent species in the lower horizontal leg; Mg^{2+} is shown as a green sphere.

The free energy derivative $\langle \partial U/ \partial \lambda \rangle_\lambda$ at each window was computed from a finite-difference estimate. Similarly, the van der Waals interactions were also transformed; the successive λ_{vdw} values were 1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.05, 0.01, and 0.001. Each λ window simulation lasted for 1-2 nanoseconds and the data from last 600-1600ps of each simulation was used for averaging. Free energy change was calculated using

numerical integration method. Standard trapezoidal method for both the complete electrostatic stage and for the van der Waals stage, down to a λ_{vdw} of 0.05, was used to integrate the derivatives. Van der Waals free energy derivative between λ_{vdw} =0.05 to 0, was fitted to the function $A_0\ \lambda_{vdw}^{-A1}$ and then analytically integrated, where A_0 and A_1 are adjustable parameters. Uncertainties of the free energy derivatives at each λ value were estimated by dividing the trajectory segment (which was used for averaging) into two batches and taking the deviation of the batch averages. The same is reported in the manuscript as statistical error. Multiple runs were performed for each state and the average result obtained from those different runs are reported in the main text of the manuscript. Each free energy calculation was based on 24-48 ns of data collection averaged over 6-12 replicas with different initial velocities (**refer appendix Table 4A-1**). Overall a total of about 1.6 μs of molecular dynamics free energy simulations have been done to get good convergence and reasonable statistical error (1-2 kcal/mol), comparable to the earlier reported force field uncertainty (**Satpati et al., 2011**). The λ vs free energy derivatives are shown in **appendix Figure 4A-1**. Different runs are in excellent agreement with each other (**refer appendix Figure 4A-1 and Table 4A-1**). The uncertainty of overall free energy change (**refer appendix Figure 4A-1**) and free energy derivatives at each of the λ points (**refer appendix Table 4A-1**) is well within the acceptable statistical uncertainty.

4.1.3 Results

4.1.3.1 Biochemical and Structural Views of RIG-I Binding to dsRNA

RIG-I recognizes 5' ppp-dsRNA or 5' pp-dsRNA and discriminates between host and viral RNA. Biochemical studies have showed that RIG-I has similar affinity for both 5’ ppp-dsRNA and 5’ pp-dsRNA, but binding is weakest with 5’ p-dsRNA (**Goubau et al., 2014**). Earlier studies indicated that 5' p-dsRNA fails to induce an antiviral response, which suggests that ligand binding is necessary but not sufficient for RIG-I activation (**Lu et al., 2010; Vela et al., 2012**). Structural studies (**Lu et al., 2010; Wang et al., 2010; Takahasi et al., 2008; Luo et al., 2012**) have revealed that 5' terminal of dsRNA is recognized by CTD of RIG-I, primarily by forming a network of electrostatic interactions. Direct interaction between α and β phosphates of dsRNA with Lys and His residues of RIG-I has been confirmed in all the resolved structures (**Lu et al., 2010; Wang et al., 2010; Luo et al., 2012**) (**Figure 4.4, a, b**).

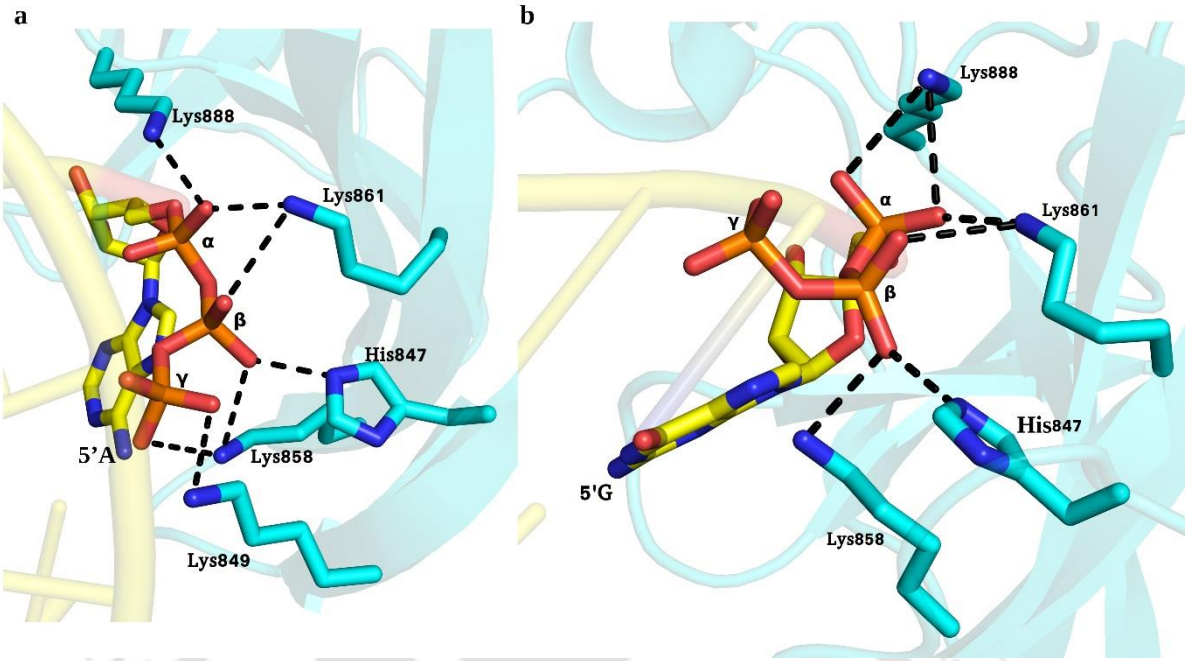


Figure 4.4. X-ray structure of RIG-I:5'ppp-dsRNA complex: (c) Representative structure (pdb: 3LRR) (**Lu et al., 2010**) revealed that RIG-I CTD of dsRNA interacts with α, β and γ phosphates of 5' adenine terminal of dsRNA through extensive electrostatic interactions. (d) pdb: 4AY2 (**Luo et al., 2012**) suggests no direct interaction between γ phosphate of 5' ppp-dsRNA and RIG-I. The network of electrostatic interactions as dashed lines. Key residues of RIG-I CTD involved in RNA binding are shown as stick models.

Conflicting literature however, exists regarding the interaction between γ phosphate of dsRNA and RIG-I. While crystal structures of RIG-I CTD bound 5' ppp-dsRNA complexes (pdb codes: 3LRN (resolution 2.6Å) (**Lu et al., 2010**), 3LRR (resolution 2.15Å) (**Lu et al., 2010**) reveal extensive electrostatic contact between all three phosphates of dsRNA with RIG-I (**Figure 4.4, a**), another x-ray structure (pdb: 4AY2 (resolution 2.8Å)) (**Luo et al., 2012**) of full RIG-I (without CARD domain) bound 5'ppp-dsRNA suggests no interaction between γ phosphate and RIG-I (**Figure 4.4, b**), suggesting γ phosphate may not be a major determinant for viral RNA detection. Still, the possibility of interaction between γ phosphate and RIG-I has not been ruled out (**Luo et al., 2012**). The controversy related to γ phosphate interaction with RIG-I is based on the slightly different orientation of γ phosphate in the complex (**Figure 4.4, a, b**).

Table 4.1: Structural Comparison of free RIG-I and dsRNA bound complexes.

	PDB Comparison	**rmsd of Full RIG-I**	**rmsd of Flexible Loop (residue no. 847-853)**
1.	4AY2 (2.8Å) vs 3LRN (2.6Å)	1.12Å	1.01Å
2.	4AY2 (2.8Å) vs 3LRR (2.15Å)	1.10Å	1.08Å
3.	3NCU (2.55Å) vs 3LRN (2.6Å)	0.96Å	1.13Å
4.	3NCU (2.55Å) vs 3LRR (2.15Å)	0.92Å	1.10Å
5.	3NCU (2.55Å) vs 4AY2 (2.8Å)	0.99Å	1.06Å
6.	3LRN (2.6Å) vs 3LRR (2.15Å)	0.81Å	0.56Å
7.	2QFB (3.0Å) vs 3LRN (2.6Å)	1.37Å	3.13Å
8.	2QFB (3.0Å) vs 3LRR (2.15Å)	1.39Å	3.24Å
9.	2QFB (3.0Å) vs 4AY2 (2.8Å)	1.52Å	3.26Å
10.	2RMJ (NMR) vs 3LRN (2.6Å)	3.79Å	9.61Å
11.	2RMJ (NMR) vs 3LRR (2.15Å)	3.89Å	9.87Å
12.	2RMJ (NMR) vs 4AY2 (2.8Å)	6.00Å	9.59Å

Two structures (pdb codes in column 2) were compared. RMSD for the CTD and the loop region (847-853) of aligned PDB's are given in Column 3 and 4 respectively. Resolution of the complexes are given in the parenthesis, NMR in the parenthesis denotes NMR structure averaged over 20 conformers. 4AY2 : X-ray structure of full RIG-I complex (without CARD's) bound to 5’ ppp-dsRNA (with 5’ terminal base Guanine); 3LRN : X-ray structure of CTD of RIG-I bound to 5’ ppp-dsRNA (with 5’ terminal base Guanine); 3LRR : X-ray structure of CTD of RIG-I bound to 5’ ppp-dsRNA (with 5’ terminal base Adinine); 3NCU: X-ray structure of CTD of RIG-I bound to 5’ pp-dsRNA (with 5’ terminal base Guanine); 2QFB : x-ray structure of free CTD of RIG-I; 2RMJ : NMR structure of free CTD of RIG-I.

We aligned some of the high resolution x-ray structures of RIG-I bound 5’ ppp/pp-dsRNA and the results are given in **Table 4.1**. The CTD of RIG-I in all the structures is very similar within an RMSD of 0.8 Å-1.1Å, including the loop region. Comparison of RIG-I CTD in the free and 5' ppp/pp-dsRNA bound structures (**Table 4.1**) does show noticeable conformational change for the loop region (RMSD 3.2Å-9.9Å). The loop region of unbound RIG-I CTD (dsRNA-free) is known to be highly flexible in solution (**Cui et al., 2008**) and dsRNA binding might stabilize specific conformation of the loop. 5' p-dsRNA bound RIG-I structure has not yet been resolved experimentally. It is worth mentioning that none of the

resolved structures of the RIG-I: dsRNA report the presence of divalent metal ion e.g. Mg^{2+} bound to the terminal phosphate/s of dsRNA. It should be noted that water molecules deposited in the PDB files may result from misinterpretation of the electron-density maps of Mg^{2+}.

Thus, we analyzed the water molecules present in the x-ray structures 3LRR (**Lu et al., 2010**) (RIG-I CTD: 5' ppp-dsRNA) and 3NCU (**Wang et al., 2010**) (RIG-I CTD: 5' pp-dsRNA). Interestingly a water molecule (with an average beta factor > 45 $Å^2$) is present in close proximity to an oxygen of the α phosphate (2.7Å in 3LRN, 3.4Å in 3NCU) in both the structures. The large beta factor of the reported water and its monodentate coordination with the α phosphate suggest that the binding site is most likely to be Mg^{2+} free. It is worth mentioning that both 4AY2 (**Luo et al., 2012**) and 3NCU (**Wang et al., 2010**) were crystallized in the buffer containing $MgCl_2$. Mg^{2+} has been isolated in the ATPase domain (**Figure 4.1**) along with ADP in 4AY2 (**Luo et al., 2012**), however, none of the reported structures resolved Mg^{2+} in the 5' terminal of dsRNA in its RIG-I bound complex. Mutagenesis of key residues involved in interaction with the 5'-phosphates has shown to affect RNA binding and signaling by RIG-I (**Lu et al., 2010**). Structural studies (**Lu et al., 2010; Wang et al., 2010; Takahasi et al., 2008; Luo et al., 2012**) further revealed that CTD of RIG-I primarily interacts with the 5' terminal and the backbone of the dsRNA, demonstrating that RIG-I can bind to dsRNA in a sequence-independent manner.

4.1.3.2 dsRNA with 5' ppp, 5' pp and 5' p in Solution

Consideration of unbound state is essential for understanding the dsRNA binding to RIG-I. Experiments (**Storer & Cornish-Bowden, 1976**) suggest that ATP^{4-} prefers to exist mainly in $[ATP:Mg]^{2-}$ form with fully deprotonated γ-phosphate along with trace amounts of $[ATPH:Mg]^{-}$, $[ATP:Mg_2]$ and $[ATP]^{4-}$. The same is also true for ADP^{3-} and AMP^{2-}. We may assume the same situation i.e, Mg^{2+} bound dsRNA is the prevalent species when free in water. The X-ray structures of RIG-I:dsRNA however, have ruled out the possibility of Mg^{2+} in the complex. This suggests that binding of 5' ppp/pp-dsRNA involves Mg^{2+} dissociation from dsRNA. It is worth mentioning that the calculations of ΔG^{comp} (without bound Mg^{2+}) and ΔG^{free} (with bound Mg^{2+}) may introduce force field errors in relative binding free energies (ΔΔG's, **Figure 4.3**). Alchemical transformation of 5' ppp-dsRNA to

5’ pp-dsRNA with bound Mg^{2+} involves the removal of a single Mg^{2+}:phosphate interaction. Mg^{2+} polarizes the electronic cloud (**Satpati et al., 2011**) and the fixed charge force fields are inaccurate as polarization effects are only implicitly included.

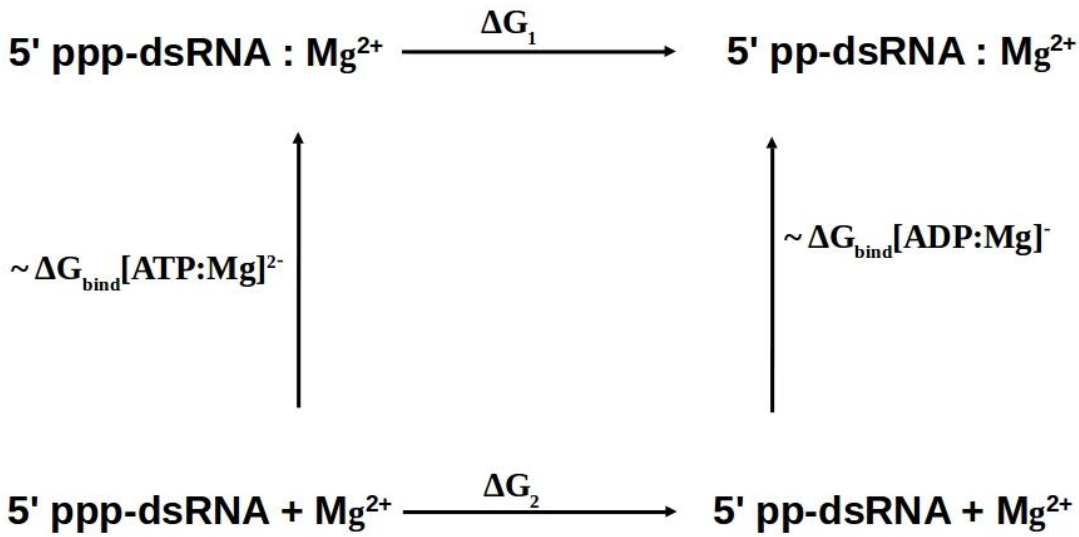


Figure 4.5: Thermodynamic cycle for Mg^{2+}: dsRNA binding. Vertical legs correspond to binding. Horizontal legs correspond to the alchemical transformation of 5' ppp-dsRNA to 5' pp-dsRNA in the presence (upper horizontal leg) and absence (lower horizontal leg) of Mg^{2+}. Free energy change for the horizontal legs is computed using molecular dynamics free energy simulations. Experimentally reported standard Mg^{2+} binding free energies to ATP and ADP have been considered (expected to be very close to Mg^{2+} binding to 5’ ppp-dsRNA/5’ pp-dsRNA) to quantify the force field error. $\Delta G_1 - \Delta G_2 = \Delta G_{bind}$[5’ pp-dsRNA:$Mg^{2+}$] - ΔG_{bind}[5’ ppp-dsRNA:Mg^{2+}] ~ ΔG_{bind}[ADP:Mg^{2+}] - Δg_{bind}[ATP:Mg^{2+}].

The effect of electronic polarization by Mg^{2+} is known (**Satpati et al., 2011**) to cancel out substantially if Mg^{2+} is present in both the horizontal legs of the thermodynamic cycle. In this study, however, force field error might be significant as the Mg^{2+} bound 5’ ppp/pp/p-dsRNA is expected to be the prevalent form in the RIG-I free form. Thus, we first computed relative binding of Mg^{2+} to dsRNA using an appropriate thermodynamic cycle (**Figure 4.5**) and compared with experimental binding data (**Alberty & Goldberg, 1992**). The results are summarized in **Table 4.2**. The results clearly indicate that the calculated relative binding free energies (column 4 of **Table 4.2**) are larger compared to that of free adenosine nucleotide in water (column 5 of **Table 4.2**).

Table 4.2: Relative binding free energies of Mg^{2+} binding to dsRNA (with 5' ppp/5' pp/5' p).

Alchemical transformation	ΔG_1	ΔG_2	$\Delta\Delta G^{calc} = \Delta G_1 - \Delta G_2$	$\Delta\Delta G^{expt}$	$\Delta G^{free} = \Delta G_2 + \Delta\Delta G^{expt}$
5' ppp-dsRNA → 5' pp-dsRNA	440.2 (1.7)	436.6 (1.6)	3.6 (2.3)	$\Delta G^0_{bind}[ADP:Mg]^-$ - $\Delta G^0_{bind}[ATP:Mg]^{2-}$ = 2.1	438.7 (1.6)
5' pp-dsRNA → 5' p-dsRNA	397.6 (1.4)	387.7 (1.2)	9.9 (1.8)	$\Delta G^0_{bind}[AMP:Mg]$ - $\Delta G^0_{bind}[ADP:Mg]^-$ = 2.5	390.2 (1.2)

Free energies are in kcal/mol. Statistical uncertainty is given in the parentheses. The MD trajectories were divided into two equal halves and the difference between the computed ΔG's from the two halves is reported as uncertainty in the parenthesis. The uncertainties for ΔΔG's were calculated by propagating the uncertainties of individual ΔG's.

Experimentally reported binding constants (**Alberty & Goldberg, 1992**) for Mg^{2+} binding to ATP and ADP (ADP and AMP) gives a standard binding free energy difference of +2.1 kcal/mol favoring ATP (+2.5 kcal/mol favoring ADP). These binding free energies are expected to be close to the calculated ΔΔG's (column 4, **Table 4.2**). Thus, the force field gives large error by overestimating Mg^{2+}:phosphate binding due to the lack of explicit electronic polarizability (**Ponder & Case, 2003; Warshel et al., 2007; Jiao et al., 2008**). Fixed charge force field is also known to make errors in accurately capturing GTP and GDP binding when electrostatic interactions with a divalent ion are involved (**Satpati et al., 2011**). Therefore, in order to compute the ΔG^{free} (**Figure 4.3**) we have taken the experimental Mg^{2+} binding data (**Table 4.3**) such that, $\Delta G^{free} = \Delta G_2 + \Delta\Delta G^{expt}$. Mg^{2+} forms bidentate coordination with the phosphate group of 5' p-dsRNA and the coordination sphere of Mg^{2+} is completed by coordination with 4 water molecules (**Figure 4.6**). The molecular dynamics simulations show that 3 water coordinated Mg^{2+} forms tridentate coordination with β and γ phosphates of 5' ppp-dsRNA and α and β phosphates of 5'pp-dsRNA (**Figure 4.6**).

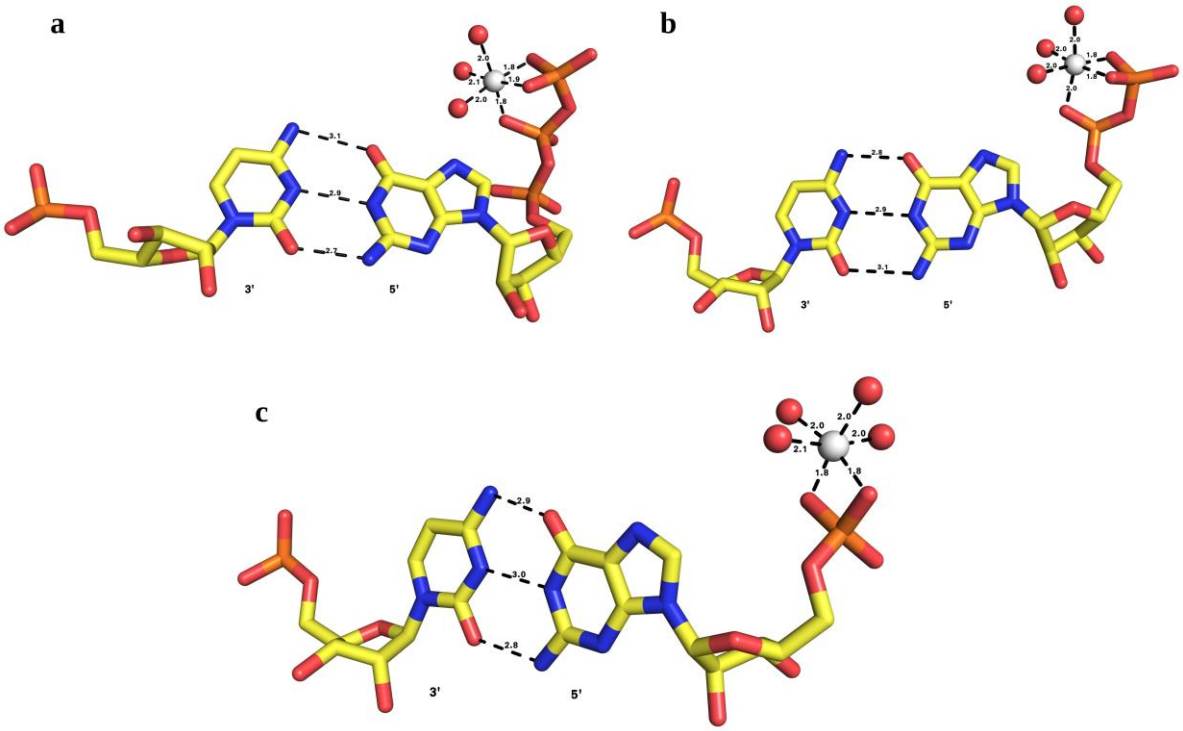


Figure 4.6: MD structures highlighting the interaction between Mg^{2+} and 5' ppp/pp/p-dsRNA in water: (a) 5' ppp-dsRNA:Mg^{2+}. (b) 5' pp-dsRNA:Mg^{2+}. (c) 5' p-dsRNA:Mg^{2+}. Mg^{2+} (white sphere), water (red sphere) and 5' terminal nucleotide (sticks). Hexa coordination of Mg^{2+} is satisfied by coordinating with water molecules. Distances in Å.

4.1.3.3 Structure-based Energetics of dsRNA Binding to RIG-I

To compute the dsRNA discrimination by RIG-I and elucidate the effect of different dsRNAs (i.e, 5'-ppp, 5'-pp and 5'-p) on the selectivity, we carried out extensive molecular dynamics free energy (MDFE) simulations of RIG-I:dsRNA complexes (5'-ppp or 5'-pp or 5'-p) using x-ray structure (pdb 4AY2) (**Luo et al., 2012**) as template. We computed the change in binding affinity for dsRNA upon 5'-ppp → 5'-pp → 5'-p mutations in the RIG-I bound complex. Relative binding free energies are summarized in **Table 4.3**. The results suggest that RIG-I imposes a very high energetic penalty of about 9 kcal/mol for binding 5' p-dsRNA with respect to 5' ppp-dsRNA, which corresponds to a probability of 10^{-7} relative to binding 5' ppp-dsRNA. Interestingly RIG-I selectivity between 5' ppp-dsRNA and 5' pp-dsRNA was predicted to be weak with a binding free energy difference of 2 kcal/mol, favoring 5' ppp-dsRNA. This corresponds to a read-through frequency as high as 10^{-2} for

dsRNA(5’-pp) with respect to dsRNA(5’-ppp). Biochemical studies indeed suggested RIG-I could be activated by 5’ ppp-dsRNA or 5’ pp-dsRNA but not by 5’ p-dsRNA (**Goubau et al., 2014**).

Table 4.3: Relative binding free energies: 5' ppp-dsRNA/5' pp-dsRNA/5' p-dsRNA binding to RIG-I.

Alchemical transformation	ΔG^{Comp}	ΔG^{free}	$\Delta\Delta G$
5' ppp-dsRNA → 5' pp-dsRNA	441.0 ± 1.8	438.7 ± 1.6	2.3 ± 2.4
5' pp-dsRNA → 5' p-dsRNA	397.4 ± 2.0	390.2 ± 1.2	7.2 ± 2.3
5' ppp-dsRNA → 5' p-dsRNA	—	—	9.5 ± 3.3

Free energies are in kcal/mol. Uncertainties are calculated in the same way described in Table 4.1.

4.1.3.4 Comparison Between MD and X-ray Structures

RMSD of the heavy atoms of about 1.3Å- 1.8Å with respect to the x-ray structure suggests that the simulated structures agree well with the x-ray structure (**Figure 4.3**). Robust features of our MD simulations are:

(a) γ-phosphate of 5’ ppp-dsRNA forms direct electrostatic interaction with the positively charged lysine side chains of RIG-I (**Figure 4.7**). The MD structures of the binding pocket are almost identical to their corresponding x-ray structures (**Figure 4.7, 4.10, 4.11; and Table 4.4**).

(b) The loop region (residue 847-853) is highly flexible (**Figure 4.8**), however, the interactions between loop residues and 5' terminal phosphates of dsRNA remain intact throughout the simulations (**Figure 4.7, 4.10, 4.11**). Phe853 stabilizes the complex by hydrophobic interaction with either G1 or C20, moving parallel to the 5’ terminal base pair G1-C20.

(c) 5’ ppp/pp/p-dsRNA binding to RIG-I is associated with desolvation of 5’ terminal of dsRNA (**Figure 4.9**).

The above structural features are insensitive to the initial structural models used in these MD simulations (**see Methods**).

4.1.3.5 Structure and Dynamics of RIG-I: 5' ppp-dsRNA Complex

The structures from MD simulations agree well with the corresponding 4AY2 crystal structure (**Figure 4.7 and Table 4.4**).

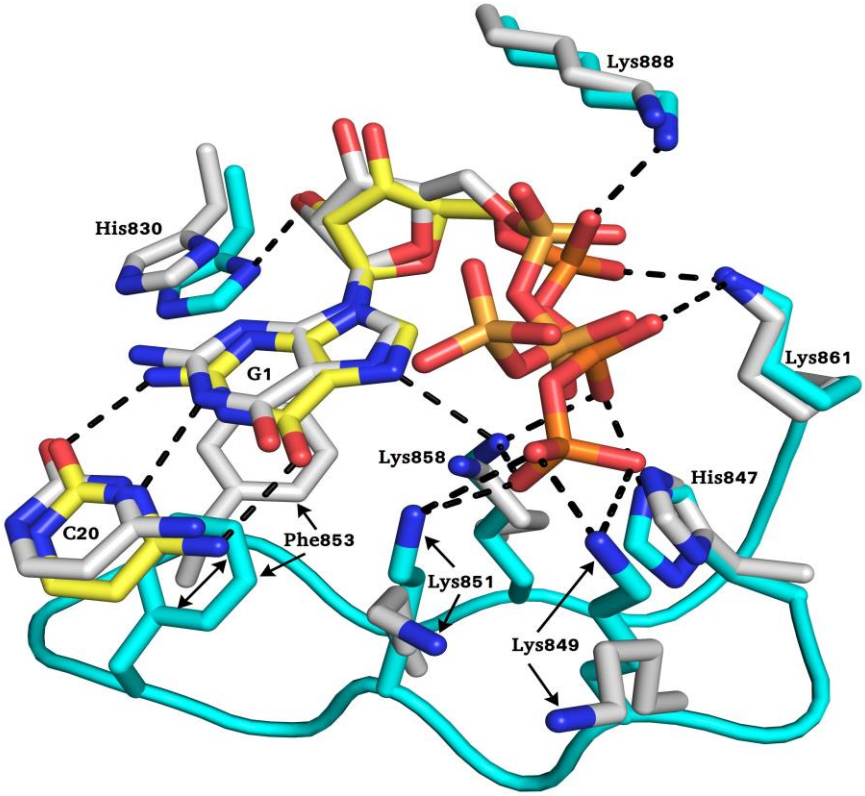


Figure 4.7: Binding pocket of 5' ppp-dsRNA: RIG-I complex: MD snapshot (cyan) compared to the crystal structure (gray; pdbcode: 4AY2). MD structure is very similar to x-ray structure, except the γ phosphate of dsRNA which forms direct interaction with Lys858 and Lys849 throughout the MD trajectory. Aromatic loop residue Phe853 stack on the terminal blunt ended base pair and show considerable movement parallel to G1-C20 base pair (indicated by a double-headed arrow). Key residues involved in RNA terminal recognition are shown as stick models and loop (residue 847-853) is shown as a cartoon. Color code same as Figure 4.1.

The rms deviation of main chain and side chain heavy atoms are 0.82±0.1Å and 1.68±0.15Å, respectively, with respect to 4AY2. The 5' triphosphate binding pocket is located at the positively charged patch of RIG-I (**Figure 4.7**), and the structural details are summarized in **Table 4.4**.

Table 4.4: Selected interatomic distances averaged over the MD trajectories.

Complex Interacting Pair		5' ppp-dsRNA Complex			5' pp-dsRNA Complex		5' p-dsRNA
		X-ray		MD	X-ray	MD	MD
		4AY2	3LRN		3NCU		
Lys849:NZ	RNA:O2G	8.30	X	2.95 (0.67)	X	X	X
Lys849:NZ	RNA:O3G	X	8.60	3.10 (0.20)	X	X	X
Lys851:NZ	RNA:O3G	5.70	X	X	X	X	X
Lys851:NZ	RNA:O1G	X	X	2.70 (0.23)	X	X	X
Lys851:NZ	RNA:O2G	X	3.70	3.03 (0.40)	X	X	X
Lys858:NZ	RNA:O2B	2.70	2.60	2.68 (0.10)	2.90	2.63(0.09)	X
Lys858:NZ	RNA:O1G	X	3.00	2.60 (0.40)	X	X	X
Lys858:NZ	RNA:N7	3.00	3.00	2.89 (0.18)	3.20	2.92 (0.13)	3.69(0.71)
His847:NE2	RNA:O2B	2.90	2.50	2.75 (0.21)	2.80	2.73 (0.19)	X
Lys861:NZ	RNA:O1B	3.00	2.90	2.77 (0.13)	X	2.62 (0.32)	X
Lys861:NZ	RNA:O3B	X	X	X	2.90	X	X
Lys861:NZ	RNA:O2A	2.60	2.30	2.72 (0.10)	X	2.75 (0.17)	X
Lys861:NZ	RNA:O1A	X	X	X	2.90	X	X
Lys888:NZ	RNA:O1A	2.90	3.30	2.67 (0.10)	3.10	2.96 (0.66)	4.03(0.69)
Lys888:NZ	RNA:O2A	3.30	2.50	4.06 (0.30)	3.20	3.28 (0.49)	3.40(0.38)
His830:ND1	RNA:O2`	2.70	2.40	2.80 (0.18)	2.70	2.80 (0.12)	2.82(0.19)

Standard deviations are in the parentheses. Distances are in angstrom. Residues/groups absent in the x-ray/MD structures are indicated with a cross.

Five lysines, two histidines, and one phenylalanine form the binding pocket for 5' triphosphate (**Figure 4.7**). Direct interaction between α phosphate and Lys888 was observed with a mean O-NZ distance of 2.67±0.1Å. Lys861 interacts simultaneously with α and β phosphate with a mean O-NZ distance of 2.72±0.1Å and 2.77±0.13Å, respectively. B phosphate is further stabilized by interacting with the side chains of Lys858 (mean O-NZ distance of 2.68±0.1Å) and His847 (mean O-NE2 distance of 2.75±0.21Å), respectively. Lys858 interacts with 5' G1 with an average distance of 2.89±0.18Å. Deviation of γ phosphate of 5' ppp-dsRNA with respect to the x-ray structure is shown in **Figure 4.7**; it is forming direct electrostatic contact with the side chains of Lys849, Lys 851 and Lys858. γ phosphate of dsRNA forms bidentate coordination with Lys849 (an average O-NZ distance of 3.1±0.2Å and 2.95±0.67Å) and Lys851 (an average O-NZ distance of 2.7±0.23Å and 3.03±0.4Å), and a monodentate coordination with Lys858 (an average O-NZ distance of 2.6±0.4Å).

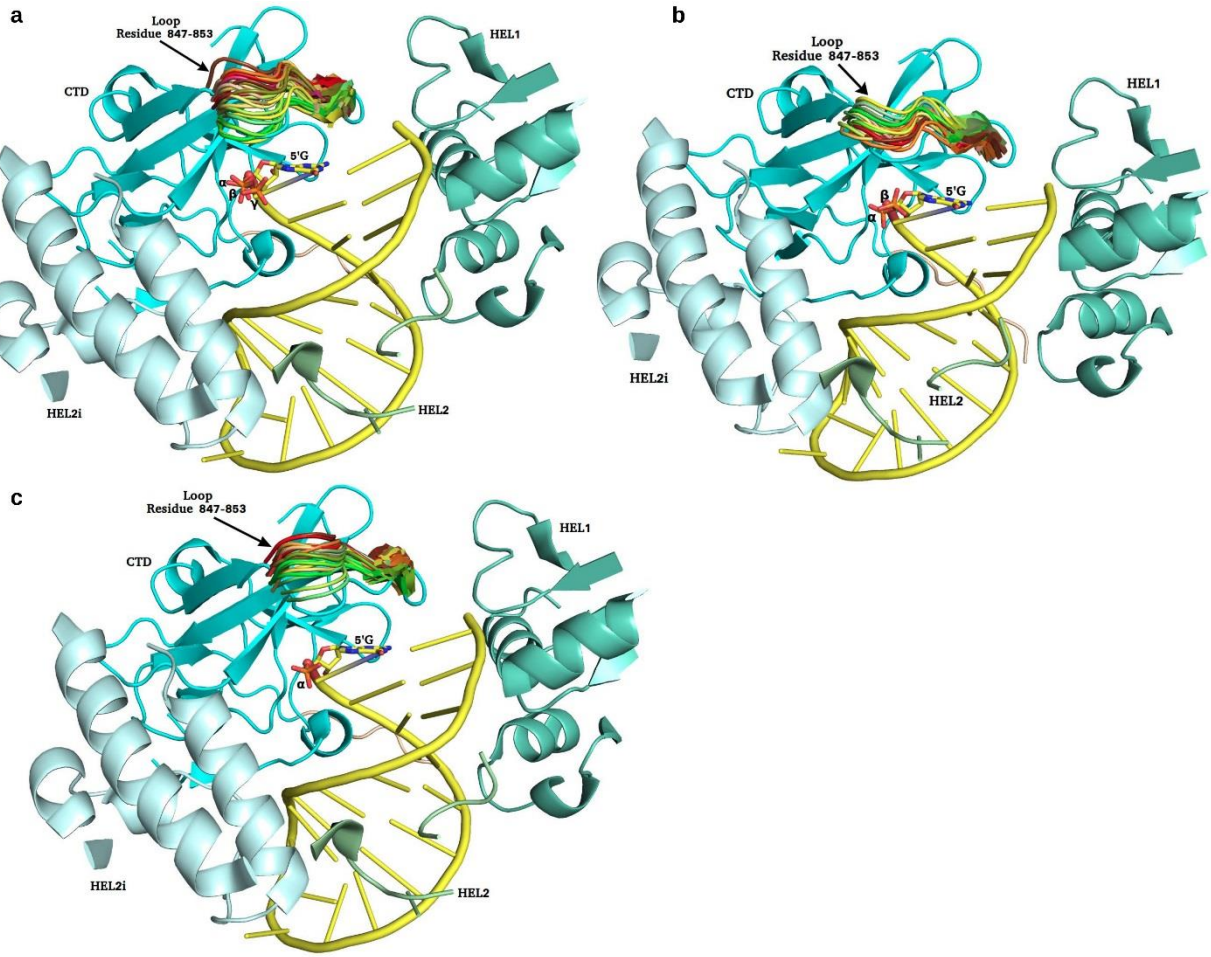


Figure 4.8: RIG-I:dsRNA complexes from MD: Flexibility of the loop region (residue 847-853): overlaid 25 snapshots with a 200ps spacing from a 5ns MD trajectory. The loop is highly flexible for RIG-I complexes with (a) 5' pp-dsRNA, (b) 5' pp-dsRNA, and (c) 5' p-dsRNA. Color code same as Figure 4.1.

Phe853 stacks over the terminal G1-C20 base pair, and His830 forms hydrogen bond with the ribose -OH of G1 with an average distance of 2.8±0.18Å. MD trajectory shows the movement of Phe853 parallel to the G1-C20 base pair and stacking with both G1 and C20 base (**Figure 4.7, Figure 4.8**). The water density around the γ phosphates of free and RIG-I bound dsRNA is shown in **Figure 4.9**. On average, the γ phosphate of 5' ppp-dsRNA is solvated with 6 and 8 water molecules in RIG-I bound and free forms, respectively. Thus, in the binding pocket, the availability of water is reduced by 2 water molecules.

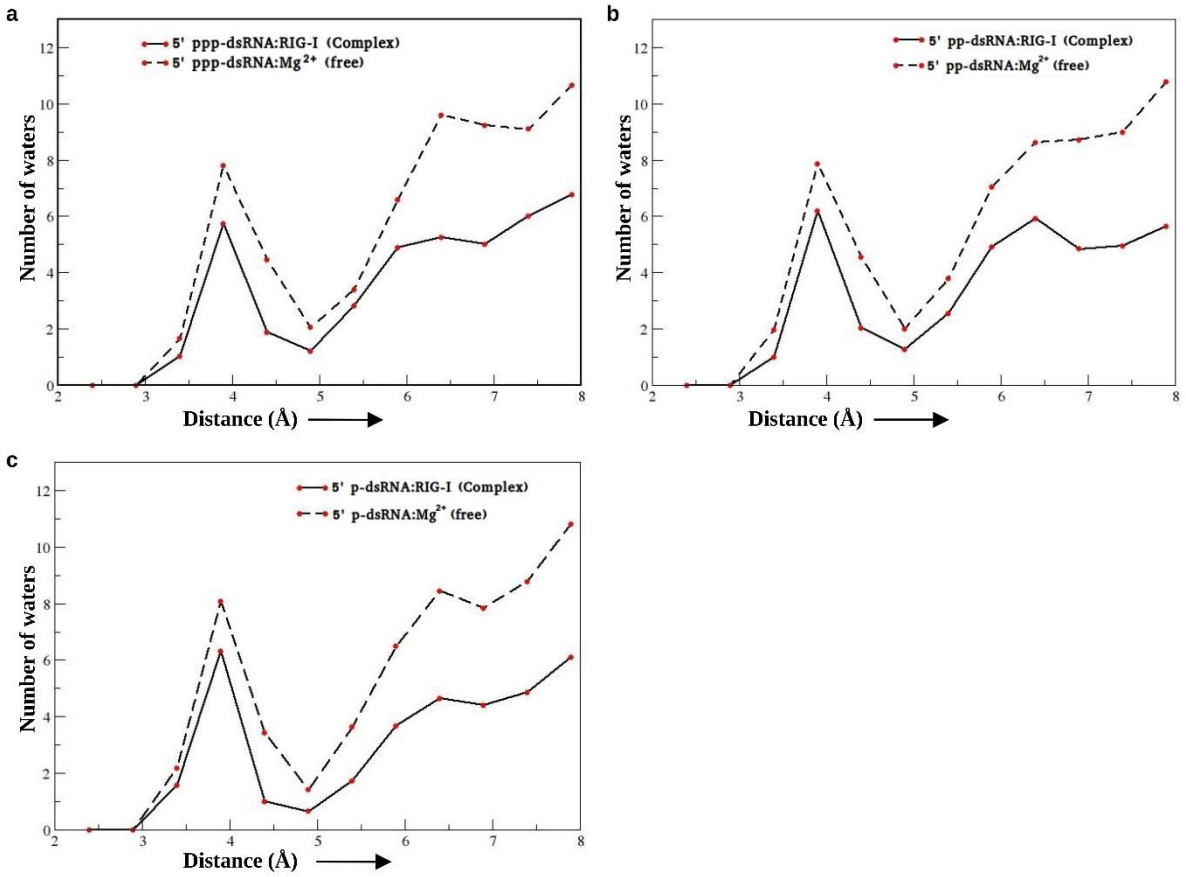


Figure 4.9: Water density around the terminal phosphate of (a) 5' ppp-dsRNA (b) 5' pp-dsRNA, and (c) 5' p-dsRNA. Results are shown for the RIG-I bound complex (thick line) and free dsRNA with bound Mg^{2+} (dash line). The water numbers are averaged in radial shells over the MD trajectories.

4.1.3.6 Structure and Dynamics of RIG-I : 5' pp-dsRNA Complex

We deleted the γ phosphate of RIG-I: 5' ppp-dsRNA complex (4AY2) and considered that as our model for 5' pp-dsRNA bound RIG-I. In our model, the helicase domain is present which is absent in the x-ray structure of 5' pp-dsRNA bound RIG-I complex (pdb 3NCU) (**Wang et al., 2010**). MD structures show that G1-RIG-I interactions are almost identical to the crystal structure (pdb: 3NCU) (**Wang et al., 2010**). Structural details are summarized in **Figure 4.10 and Table 4.4**. With respect to the 4AY2, the deviations are 1.06±0.14Å for the main chain and 1.76±0.08Å for side chains. Direct interactions of α and β phosphates of dsRNA with Lys888, Lys861, His847 and Lys858 have been observed (with an average distance ranging between 2.62Å-3.28Å; standard deviation of less than 0.7Å). The overall charge of -3 of the terminal G1 is neutralized by side chains of 3 lysine residues. His830 and

Lys858 form hydrogen bonds with the ribose -OH and N7 of G1 with an average distance of 2.8±0.12Å and 2.92±0.13Å, respectively. Movement of Phe853 parallel to the 5' terminal base pair G1-C20 has also been observed in the MD trajectory. Desolvation of 5' pp-dsRNA (similar to 5' ppp-dsRNA) has been observed upon binding to RIG-I (**Figure 4.9**).

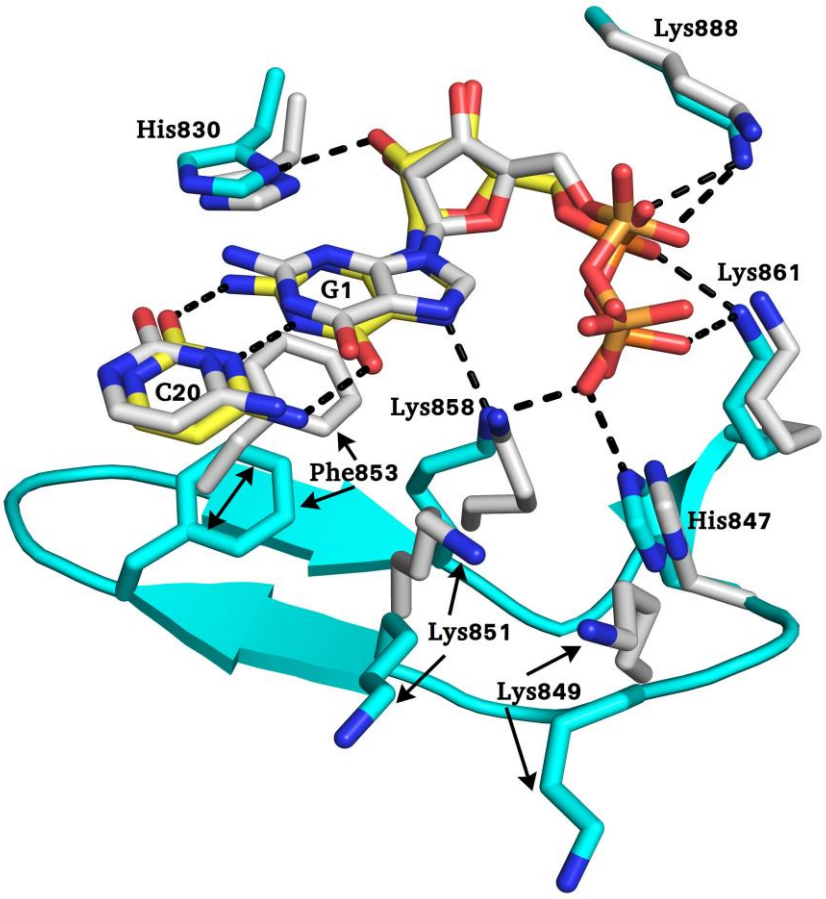


Figure 4.10: Binding pocket of 5' pp-dsRNA: RIG-I complex: MD snapshot (cyan) compared to the crystal structure (gray; 3NCU). MD structure is similar to the x-ray structure. Key residues involved in RNA terminal recognition are shown as stick models and loop (residue 847-853) is shown as a cartoon. Color code same as Figure 4.1.

4.1.3.7 Structure and Dynamics of RIG-I: 5' p-dsRNA Complex

The RIG-I: 5' p-dsRNA structure is not known experimentally. We have deleted β and γ phosphates from 4AY2 and considered that as the MD model. The average rms deviations for main chain and side chain heavy atoms are 1.1±0.28Å and 1.97±0.16Å, respectively, with respect to the starting MD model. Structural parameters of RIG-I: 5' p-dsRNA complex are given in **Figure 4.11** and **Table 4.4**.

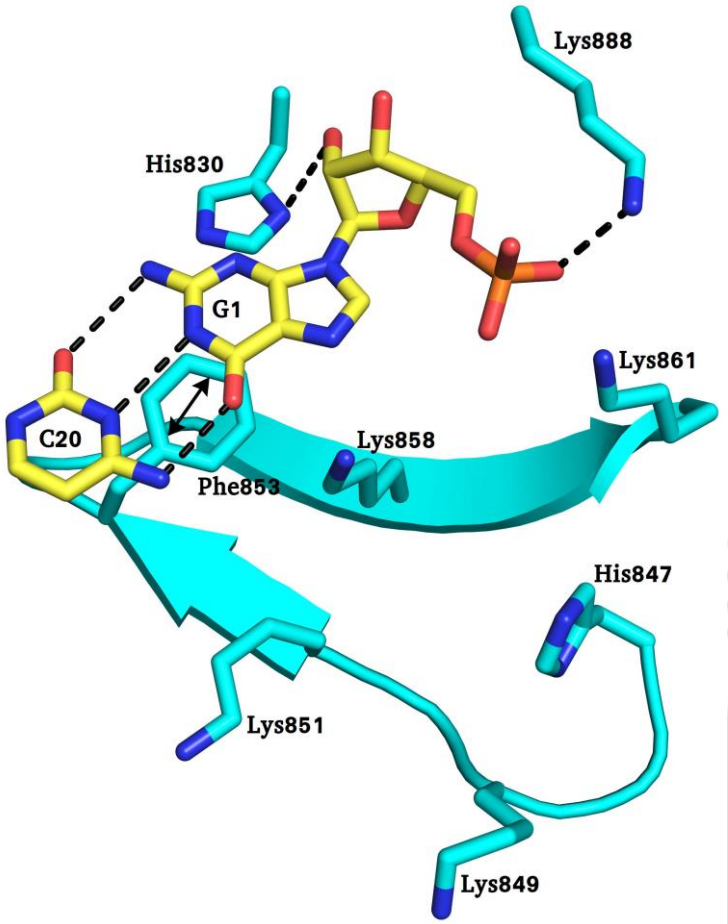


Figure 4.11: MD snapshot of 5' p-dsRNA: RIG-I complex binding pocket. Key residues involved in RNA terminal recognition are shown as stick models and loop (residue 847-853) is shown as a cartoon. Color code same as Figure 4.1.

The absence of β and γ phosphates leads to loss of interaction with RIG-I. The only interaction between His830 and ribose –OH of G1 remains intact as seen in RIG-I: 5' ppp/pp-dsRNA complexes. The average distance of 3.4±0.38Å between Lys888 and α phosphate of 5' p-dsRNA suggests weak interaction. Similar loop flexibility (Figure S5) as seen in RIG-I bound 5' ppp/pp-dsRNA has been observed. Binding of 5' p-dsRNA to RIG-I is associated with desolvation (**Figure 4.9**) as also seen in 5' ppp/pp-dsRNA. MD suggests that binding of 5' p-dsRNA to RIG-I is unfavorable due to the loss of direct interactions between 5' p-dsRNA and RIG-I.

4.1.4 Discussion

Understanding the RIG-I: dsRNA binding is a difficult challenge. Multiple RIG-I conformations (bound/unbound states) and multiple dsRNA species (e.g, different protonation states of dsRNA with bound/unbound ions and its RIG-I bound/unbound conformations) contribute to the overall binding process. Despite recent biochemical studies and medium resolution structures, we are far away from understanding the structure based detailed energy landscape associated with viral RNA recognition. Measuring the binding free energy between a specific nucleotide and a particular protein conformation is an extremely difficult task. MD simulations can only fill the gap to some extent. Using MDFE simulations we calculated 5' ppp-dsRNA/5' pp-dsRNA/5' p-dsRNA binding free energy differences. Calculated strength of discrimination, $\Delta\Delta G$ ~9 kcal/mol suggests that 5' ppp-dsRNA binding to RIG-I is strongly favored with respect to 5' p-dsRNA. On the other hand, small $\Delta\Delta G$ ~2 kcal/mol suggests that 5' ppp-dsRNA binding to RIG-I is weakly favored with respect to 5' pp-dsRNA. Since the magnitude of the relative preferences ($\Delta\Delta G$) is not known experimentally, the calculated MDFE values cannot be confirmed or disproved. Certainly, the signs are correct, and corroborates the experiment (**Goubau et al., 2014**). X-ray and NMR structures suggest that the loop region (residue 847-853) of free RIG-I is highly flexible (**Cui et al., 2008**) and dsRNA binding might lead to stabilization of specific conformation of the RIG-I loop (**Lu et al., 2010**). The average value of B-factor of the loop region (residue 847-853) of our template pdb 4AY2 is 85$Å^2$, indicating considerable movement. MD simulations suggest that the loop region (residue 847-853) is flexible and can have multiple conformations (**Figure 4.8**) even in the dsRNA bound state.

The key interaction between the loop side chains and 5’ terminal of dsRNA are however, retained throughout the MD trajectory (**Figure 4.7, 4.10;** and **Table 4.4**). The direct interaction between Lys858 and N7 of 5’ terminal base of dsRNA (**Figure 4.7, 4.10, 4.11;** and **Table 4.4**) is very specific to purines and might disfavor pyrimidines at the 5’ end. Almost all the interaction distances in the MD structures agree with their corresponding x-ray structures (**Table 4.4**), except for the fact that γ phosphate of 5' ppp-dsRNA establishes a direct contact with the RIG-I (absent in template pdb 4AY2). MD results are in agreement with the x-ray structures of RIG-I CTD bound 5' ppp-dsRNA, revealing direct electrostatic interactions between RIG-I and γ phosphate of dsRNA (**Lu et al., 2010; Devarkar et al.,**

2016). It might appear that γ phosphate might be a crucial recognition determinant and RIG-I should strongly favor 5' ppp-dsRNA binding with respect to 5' pp-dsRNA. Experiments (**Goubau et al., 2014**) however, have clearly shown that RIG-I discriminates weakly between 5' ppp-dsRNA and 5' pp-dsRNA. 5' pp-dsRNA is the minimum requirement for RIG-I binding and antiviral response. In order to understand dsRNA binding, consideration of unbound state is essential. Free 5' ppp/pp-dsRNA with bound Mg^{2+} is expected to be the prevalent species. Structural studies suggest that 5' ppp/pp-dsRNA in the RIG-I bound form do not have Mg^{2+}, indicating that Mg^{2+} detachment might be required for dsRNA binding to RIG-I. It is evident from the structures that the negatively charged 5' terminal of dsRNA is neutralized in the RIG-I binding pocket through extensive electrostatic interaction, justifying the absence of Mg^{2+} in the binding pocket. Structures of the binding pocket from MD simulations are very similar to their x-ray structures (**Figure 4.2, Table 4.4, Figure 4.8, and Figure 4.9**), suggesting that the starting structure of the complex (without Mg^{2+}) is a true minimum in the potential energy hypersurface. MD further revealed that dsRNA binding to RIG-I is associated with desolvation. The RIG-I binding site is water exposed and the extent of desolvation seems to be similar for 5' ppp/pp/p-dsRNA binding (**Figure 4.7, Figure 4.9**). The results suggest that desolvation is crucial for binding but may not be significant towards relative binding preference. It should be noted that the dissociation constants for the Mg^{2+}:AMP^{2-}, Mg^{2+}:$ADP^{3-}$and Mg^{2+}:ATP^{4-}complexes are K_d = 1.622x10^{-3} M, 2.239x10^{-5} M and 6.607x10^{-7}M at room temperature (298.15 K) and pressure (1 atm); the standard binding free energies (ΔG^0_{bind}) are −3.8 kcal/mol, -6.3 kcal/mol and -8.4 kcal/mol respectively (**Alberty & Goldberg, 1992**). Clearly the energetic cost for Mg^{2+} dissociation is highest from ATP^{4-} > ADP^{3-} > AMP^{2-} for electrostatic reason. We may expect the same is true for dsRNA containing 5' phosphate/s, hence, interaction between dsRNA and RIG-I in the binding pocket needs to offset the Mg^{2+} dissociation penalty from the dsRNA. Both 5' ppp-dsRNA and 5' pp-dsRNA form an extensive interaction network in the RIG-I binding pocket. The specificity between 5' ppp-dsRNA and 5' pp-dsRNA however, was predicted to be weak with a binding free energy difference of ~ 2 kcal/mol, favoring 5' ppp-dsRNA. Interactions between the γ phosphate of 5' ppp-dsRNA and lysine side chains in the RIG-I binding pocket offset the largest Mg^{2+} dissociation penalty (ΔG^0_{bind}= −8.4 kcal/mol); this might be the reason for weak specificity between 5' ppp-dsRNA and 5' pp-dsRNA. Strong RIG-I binding specificity of ~ 9 kcal/mol in favor of

5’ ppp-dsRNA with respect to 5’ p-dsRNA is due to the loss of protein-RNA interaction in the latter.

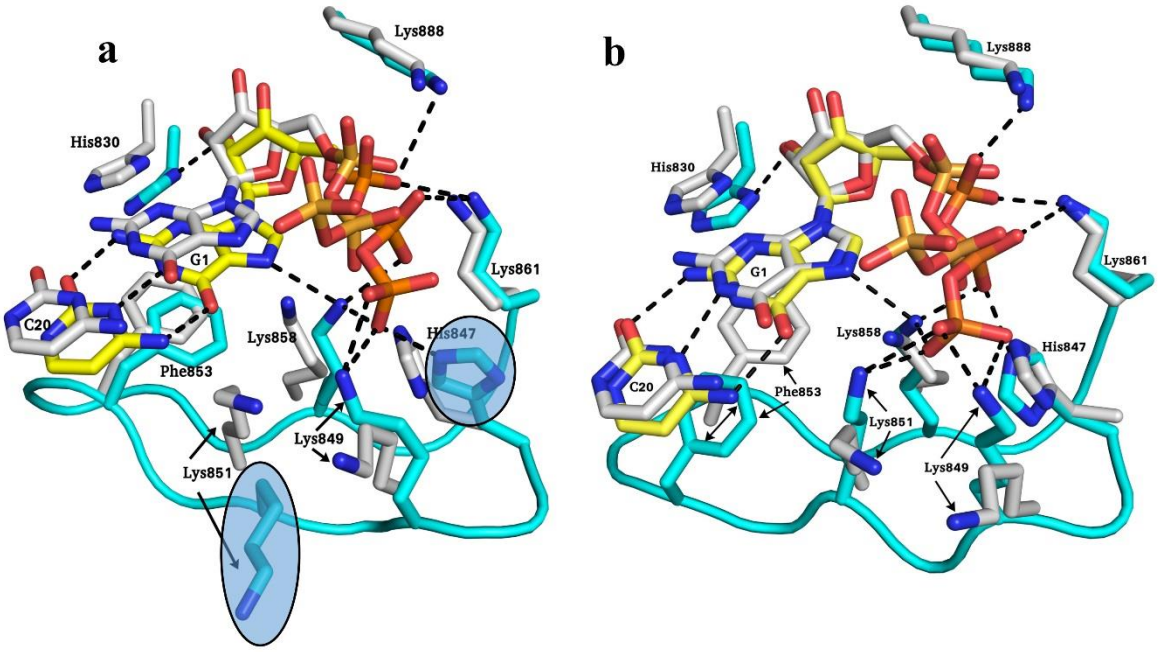


Figure 4.12: Binding pocket: MD snapshot (cyan) vs X-ray structure (gray; pdbcode: 4AY2 for 5’ppp-dsRNA:RIG-I and pdbcode: 3NCU for 5’pp-dsRNA:RIG-I). Hydrogens are omitted for clarity. (a) Complex with protonated His847 (b) Complex with neutral His847. Main difference between MD structures of (a) and (b) is highlighted with blue circle.

The close proximity of His847 to the negatively charged β phosphates of dsRNA might stabilize the protonated state of His847. MD simulations with RIG-I : 5’ ppp-dsRNA complex (with protonated His847) suggests alternate conformation of His847 (**See Figure 4.12**), in which the interaction with β phosphate is lost and interaction with γ phosphate is formed. It should be noted that in all the x-ray structures (pdb 3LRR, 3LRN, 4AY2) (**Lu et al., 2010; Luo et al., 2012**), the position of His847 is very much similar and His847 interacts only with the β phosphate. The alternate conformation seen in MD simulation with protonated His847 is not supported by x-ray structures. To the best of our estimate, RIG-I with protonated His847 imposes higher energetic penalty of about 3.7 ± 3.2 kcal/mol (relative to 2.3 ± 2.4 kcal/mol for deprotonated His847) for binding to 5' pp-dsRNA with respect to 5' ppp-dsRNA. Biochemical studies (**Goubau et al., 2014**) have shown that both 5' ppp-dsRNA and 5' pp-dsRNA can induce antiviral response. Hence the magnitude of

relative binding affinity ΔΔG (RIG-I binding to 5' ppp-dsRNA vs 5' pp-dsRNA) is expected to be small. The deviation of protonated-His847 (in MD) with respect to its experimentally determined position, along with the larger ΔΔG (computed) probably suggest the deprotonated state of His847.

4.1.5 Conclusion

MD simulations of RIG-I: 5’ ppp-dsRNA, RIG-I: 5’ pp-dsRNA and RIG-I: 5’ p-dsRNA complexes provide crucial insights into the energetics of viral RNA binding to RIG-I. Extensive interaction between 5’ terminal of dsRNA and RIG-I is favorable for binding, whereas desolvation and dissociation of Mg^{2+} from dsRNA are unfavorable. The extent of desolvation is very similar for 5' ppp-dsRNA, 5' pp-dsRNA, and 5' p-dsRNA, but the Mg^{2+} binding free energy is largest for 5’ ppp-dsRNA and smallest for 5’ p-dsRNA. RIG-I binding discriminates weakly between 5’ ppp-dsRNA and 5’ pp-dsRNA with a binding free energy difference of ~2 kcal/mol, favoring 5’ ppp-dsRNA. The favorable interaction between RIG-I and γ phosphate of 5’ ppp-dsRNA in the binding pocket is mostly offset by the higher Mg^{2+} dissociation penalty, leading to weak discrimination between 5' ppp-dsRNA and 5' pp-dsRNA. 5' p-dsRNA is discriminated strongly by ~ 9 kcal/mol with respect to 5' ppp-dsRNA due to the loss of interaction between 5' terminal of 5' p-dsRNA and RIG-I. In spite of all the limitations (sampling, convergence, force fields etc.) associated with MD simulations, the signs obtained for relative binding free energies (ΔΔG) are correct since RIG-I preferentially binds 5' ppp-dsRNA/5' pp-dsRNA with respect to 5' p-dsRNA. The strength of discrimination (ΔΔG) appears to be biochemically plausible. The precise relative binding free energies are not known experimentally and we hope that our study will encourage experimental verification.

4.2 Why Double-stranded RNA with 5’ Monophosphate is a Poor Binder to RIG-I with Respect to 5’ Hydroxyl Analogue?

5’ p-dsRNA binds to RIG-I with strikingly low affinity and the binding affinity is weaker relative to 5’ OH-dsRNA; a characteristic feature of RIG-I in general. The mechanism of RIG-I discrimination between host RNA’s bearing 5’ p and 5’ OH, preferring the latter has not yet been understood. Using X-ray structures as a template we performed molecular dynamics free energy calculations and quantitatively estimated the host RNA discrimination (5’ p-dsRNA vs 5’ OH-dsRNA) by RIG-I. We show that RIG-I prefers 5’ OH-dsRNA over 5’ p-dsRNA and the estimated strength of preference is in very good agreement with the experiment. The underlying thermodynamics for the relative preference could be explained in terms of counter ion dissociation from the 5’ p-terminus of dsRNA. We propose that Mg^{2+} dissociation form 5’ terminus of 5’ p-dsRNA might be a requirement for RIG-I binding and the energetic cost of Mg^{2+} dissociation form 5’ terminus of 5’ p-dsRNA offsets its binding preference over 5’ OH-dsRNA analogue.

==========

The work published in Journal of Biomolecular Structure and Dynamics,
DOI:10.1080/07391102.2019.1671225

4.2.1 Background

RIG-I (retinoic acid-inducible gene-I) is a mammalian protein that recognizes viral RNA and initiates the antiviral innate immune response, as described in **Section 4.1.1**. Hence, RIG-I is crucial for human health. Double-stranded RNA (dsRNA) molecules bearing 5' (tri/di)phosphates are believed to be the viral feature which is recognized by RIG-I (**Hornung et al., 2006; Schmidt et al., 2009; Davis et al., 2012; Goubau et al., 2014; Linehan et al., 2018; Ren et al., 2019**). Eukaryotic mRNA containing 5' triphosphate prevent RIG-I recognition by post-transcriptional modifications and its single-stranded nature (**Schuberth-wagner et al., 2015**). However, the eukaryotic cell contains a pool of RNA's containing 5' p/-OH (**van Dijk et al., 2002; Jonas, & Izaurralde, 2013; Braun et al., 2012; Nagarajan et al., 2013; Park et al., 2011; Peach, York, & Hesselberth, 2015**). RIG-I selectively bind 5' ppp/pp-dsRNA and activate antiviral signalling from the pool of cellular RNA's. Hence, understanding the mechanism of RIG-I discrimination between pathogenic 5' ppp/pp-RNAs and host 5' p/-OH-RNAs is relevant in terms of fundamental science and the welfare of the society. Structural studies (**Wang et al., 2010; Lu et al., 2010; Kowalinski et al., 2011; Jiang et al., 2011; Luo et al., 2012; Kohlway et al., 2013; Devarkar et al., 2016**) suggest that the C-terminal domain (CTD) of RIG-I establish extensive electrostatic interaction network with 5' ppp/pp-RNA which is lacking in 5' OH-RNA bound complex.

Biochemical studies (**Ren et al., 2019; Vela et al., 2012; Schlee et al., 2009; Linehan et al., 2018**) confirm high-affinity binding for 5' ppp/pp-dsRNA and lower affinity binding of 5' p/OH-RNA, corroborating structural studies (**Wang et al., 2010; Lu et al., 2010; Kowalinski et al., 2011; Jiang et al., 2011; Luo et al., 2012; Kohlway et al., 2013; Devarkar et al., 2016**). RIG-I binds to 5' ppp-dsRNA and 5' pp-dsRNA with very high affinity (~ pM range) and binding affinities are almost identical (**Ren et al., 2019**). Interestingly RIG-I binds 5' p-dsRNA very weakly and the binding affinity is weaker than 5' OH-dsRNA (**Ren et al., 201**9). Key points emerges from the existing literature are (1) γ-phosphate is certainly not a recognition determinant; 5' pp is the minimum requirement for RIG-I recognition and inducing antiviral response (**Goubau et al., 2014; Ren et al., 2019; Linehan et al., 2018; Kumar & Satpati, 2018**) (2) 5' p-dsRNA binding is disfavoured most by RIG-I with respect to RNA's containing 5' ppp/pp/-OH; which could be considered

as a signature of RIG-I selectivity in general (**Ren et al., 2019**). RIG-I binding affinity to 5' p-dsRNA is ~3.5 fold weaker (**Ren et al., 2019**) than it's 5' OH analogue, despite the fact that direct interaction between α-phosphate of 5' ppp/pp-dsRNA and RIG-I was confirmed by all the x-ray structures (**Wang et al., 2010; Lu et al., 2010; Kowalinski et al., 2011; Jiang et al., 2011; Luo et al., 2012; Kohlway et al., 2013; Devarkar et al., 2016**). One might expect that 5' p-dsRNA would bind RIG-I weakly with respect to 5' ppp/-pp-dsRNA, but certainly stronger than 5' OH-dsRNA, due to the absence of favourable phosphate-protein electrostatic interaction in the later. It should be noted that despite several attempts x-ray structure of RIG-I: 5' p-dsRNA could not be isolated. Whereas, RIG-I: 5' OH-dsRNA structures were characterized experimentally.

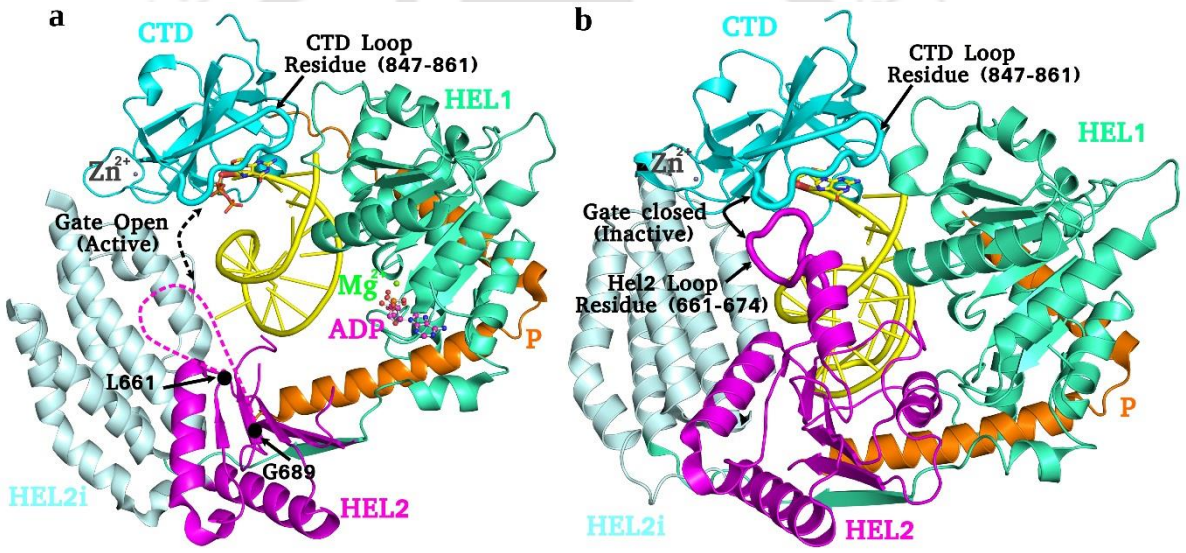


Figure 4.13: Structural overview of RIG-I:dsRNA complexes: (a) Active RIG-I: 5' ppp-dsRNA complex (PDB: 4AY2): CTD loop (residue 847-861) as thick tube and Hel2 loop (residue 661-674) was unresolved, thus indicated as broken line. (b) Inactive RIG-I: 5' OH-dsRNA complex (PDB: 5F9H) : CTD and Hel2 loop shown as thick tube.

RIG-I consists of 4 domains (**Figure 4.13, a, b**) which has already been discussed in **Section 4.1.1**. It was proposed that RIG-I could adopt two functionally different conformations (**Figure 4.13, a,b**), i.e, signalling active conformation (A) in complex with 5' ppp/pp-dsRNA and signalling inactive conformation (NA) in complex with 5' OH-dsRNA bound form (**Ren et al., 2019**). The difference between two complexes ("A" and "NA") is primarily due to the conformation of an evolutionarily conserved protein loop (the

Hel2 loop) (**Figure 4.13, b**) and the helicase domain. CTD is more or less identical for "A" and "NA". The relative position of Hel2 loop (residue 661-674) and CTD loop (847-861) around the 5'-terminal of dsRNA could result in gate open (active or "A") and the gate closed (inactive or "NA") conformation. Interestingly, Hel2 loop was disordered in the x-ray structure of the active complex and highlighted with a broken line (**Figure 4.13, a**). It can be argued that Hel2 loop is away from the 5'-terminal and the conserved amino acids of CTD forms direct interactions with the 5' terminal phosphates of dsRNA in the "A" conformation (Gate open conformation, **Figure 4.13, a**). In the "NA" complex the CTD-5' terminal interaction is altered by the insertion of Hel2 loop with the formation of a new set of RNA-protein interactions (Gate closed conformation, **Figure 4.13, b**). The "NA" conformation is supported by the x-ray structure of RIG-I: 5' OH-dsRNA (PDB 5F9F; resolution 2.6Å, **Figure 4.13, b**). RIG-I: RNA interactions in "NA" conformation are expected to be disrupted further by the presence of 5' p. It was proposed (**Ren et al., 2019**) that "A" conformation of RIG-I activates downstream antiviral signalling, whereas "NA" complex is inactive and speculated to be the preferred conformation in RIG-I: 5' p-dsRNA complex. It is worth mentioning that the same Hel2 loop was disordered in other x-ray structures of RIG-I: 5' OH-dsRNA complexes (PDB 2YKG; resolution 2.5Å, PDB 3ZD7; resolution 2.5Å, PDB 4BPB; resolution 2.58Å, PDB 5E3H; resolution 2.7Å, PDB 3ZD6; resolution 2.8Å, PDB 6GPG, resolution 2.89Å), (**Jiang et al., 2011; Kohlway et al., 2013; Luo et al., 2011; Lassig et al., 2018**) which demands explanation.

Despite the wealth of structural and biochemical studies, the atomic insight into the structure and dynamics of RIG-I: 5' p-dsRNA complex is unknown. Hence, limiting our understanding of the accuracy of dsRNA selection by RIG-I based on 5' terminal differ by few atoms. Available X-ray structures provide a template for structure-based computer simulations (**Kumar & Satpati, 2018**) of RIG-I complexes with dsRNA. Using X-ray structures as template (**Ren et al., 2019; Luo et al., 2012**) we modelled "A" and "NA" conformations of RIG-I: 5' p-dsRNA complex and performed MD simulations. Using molecular dynamics (MD) free energy calculations the relative RIG-I binding strength between 5' p-dsRNA and 5' OH-dsRNA was calculated for both "A" and "NA" complexes. In the paper we addressed the following issues, (a) how strongly different conformations ("A" and "NA") of RIG-I discriminates between 5' p-dsRNA and 5' OH-dsRNA (i.e.,

relative binding affinity)? (b) What is the relationship between relative binding affinity and the 3D structures of RIGI: 5' p-dsRNA and RIGI: 5' OH-dsRNA complexes?

Despite very different conformations of Hel2 loop ("A" vs "NA"), the calculations suggest that both the "A" and "NA" complex prefer 5' OH-dsRNA weakly with respect to 5' p-dsRNA. The estimated strength of relative preference is very close to experiment. We propose that Mg^{2+} dissociation from the terminal 5'-phosphate of dsRNA is a characteristic feature of RIG-I binding, which could fine-tune the energetics leading to the general and conserved feature i.e, preferring 5' OH-dsRNA binding to RIG-I over 5' p-dsRNA.

4.2.2 Methods

4.2.2.1 Molecular Dynamics Setup

X-ray structures of human RIG-I complex with panhandle 5' ppp-dsRNA (PDB 4AY2; crystallographic resolution 2.8 Å) (**Luo et al., 2012**) was used as a template to model active conformation. Terminal 5' β and γ phosphates from PDB 4AY2 (**Luo et al., 2012**) was deleted to model RIG-I bound 5' p-dsRNA in the active conformation. In this structure, the Hel2 loop was fully disordered, with 27 amino acids (residue 662-688) unresolved in the electron density. Adjacent amino acids connected to the missing loop, Leu661 and Gly689 (**Figure 4.13, a**), were rather distant (Leu661 and Gly689 are 31.5Å and 42.9Å away) from the 5'-phosphate. The overall orientation of the visible portion Hel2 loop suggests that the missing Hel2 loop would be even further away from the 5' phosphate. Therefore, initially, we have not considered the missing loop in our MD model and the missing Hel2 loop was effectively replaced by water (Model: $A^{\#}$). In the next step, we build the missing Hel2 loop using Phyre2 web portal (**Kelley et al., 2015**) and subjected the resulting model "$A^{\#}$" to MD simulations.

X-ray structure of RIG-I: 5' OH-dsRNA with intact Hel2 loop (PDB 5F9F (**Ren et al., 2019**), crystallographic resolution 2.6 Å) was used as a template for modelling "NA" complex for MD simulations. Similar to the active complex, we also considered the $NA^{\#}$ model by deleting Hel2 loop from "NA". Conformational change that defines the "A" and "NA" conformations are primarily localized in the Hel2 loop and helicase domains. Hel2 loop is close to the 5' terminal only in the "NA" complex. The "NA" conformation of RIG-I bound 5' p-dsRNA was modelled by slowly mutating 5' OH→5' p in the PDB 5F9F (**Ren**

et al., 2019) during a series of MD simulations (Alchemical transformation, described in details later).

In our MD models we considered protein-RNA residues having at least one nonhydrogen atom within a 30 Å radius sphere, centred at the 5' terminal of ribose C5' atom (connected to p/OH) of dsRNA. Smaller spherically truncated model of 25Å radius was successful in understanding the accuracy of protein synthesis in the ribosome (**Satpati et al., 2014**), which is very similar to this project. Thus, models of slightly larger sphere of 30Å in this project is a reasonable choice. The simulation protocol and details has already been discussed in **Section 4.2**.

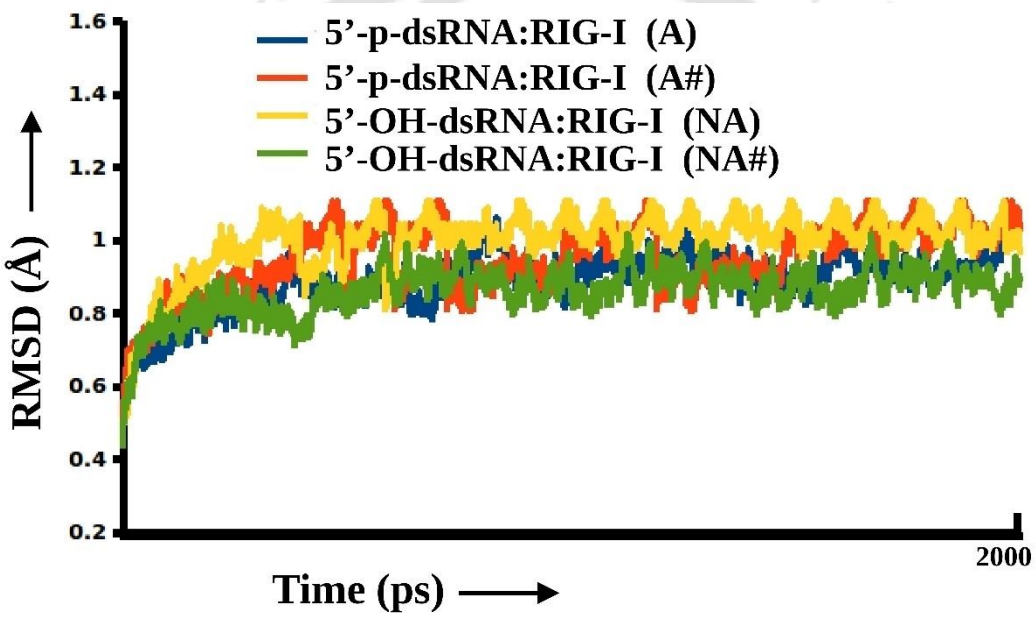


Figure 4.14: Root-mean-square deviation of heavy atoms within 27Å of unrestrained simulation sphere with respect to x-ray structure.

The overall simulation box was neutralized by scaling down the partial changes of the distant backbone phosphates of dsRNA (Distributing +0.25 charge to four atoms: 1 Phosphorous and 3 terminal oxygen atoms). We have studied 4 simulation models (Models: A, A#, NA, NA#) for each of RIG-I: 5' OH-dsRNA and RIG-I:5' p-dsRNA complex. For each of the models, several independent MD replicas were performed. The number of the replicas varied from minimum 2 to maximum of 7. The same approach was adopted for free dsRNA in water with different models (viz, 5'-p, 5'-OH terminal with bound/unbound

Mg^{2+}). Individual replicas were run for various length of time, verifying from a minimum of 18ns to maximum of 50ns. The convergence of the simulation was confirmed by calculating RMSD of the heavy atoms (excluding buffer region and solvent) relative to the template x-ray structure (**Figure 4.14**). Following production dynamics, alchemical transformation of 5' p-dsRNA→5' OH-dsRNA and 5' OH-dsRNA→5' p-dsRNA was performed through a series of alchemical transformation.

4.2.2.2 Alchemical Transformation and Relative Binding Free Energies

Relative binding free energies (ΔΔG) of 5' p-dsRNA and 5' OH-dsRNA binding to RIG-I was calculated using single topology thermodynamic integration (TI) and a standard thermodynamic cycle where 5' P were alchemically transformed into a 5' OH (in case of "A" complex) or vice versa (in case of "NA" complex) following the horizontal arms of the cycle as shown in **Figure 4.15**. The vertical legs correspond to the binding of RNA to RIG-I which has been measured experimentally (**Ren at al., 2019**; **Vela et al., 2012**). However, alchemical/unphysical transformation of dsRNA along the horizontal legs cannot be realized experimentally. Free energy change associated with the alchemical transformation (horizontal legs) was computed in the RIG-I bound complex (ΔG^{comp}) and free dsRNA (ΔG^{free}) in water (**Figure 4.15**); which estimates the relative binding free energy as described in **Section 4.1.2.2**. Calculations of ΔG^{comp} or ΔG^{free} involve series of intermediate MD simulations along the alchemical coordinate. A hybrid energy function U(λ) and coupling/scaling coordinate λ were defined along the alchemical coordinate/horizontal path (**Figure 4.15**), such that λ=0 represent the one end-point (say 5' OH-dsRNA) and λ=1 represent another end-point (5' p-dsRNA). Van der Waals and electrostatic interactions were modified separately with two different coupling parameter λ_{vdw} and λ_{elec} respectively, as discussed in **Section 4.1.2.2**.

The free energy change is averaged over multiple independent replicas with different initial velocities and reported in the main text along with the standard error of the mean (**refer appendix Table 4A-2**). Using appropriate thermodynamic cycle (**Figure 4.15**), we estimated the relative binding free energies (ΔΔG) and compared between active and inactive complex in the presence and absence of Hel2 loop (**refer appendix Table 4A-2, g**). It turns out that Hel2 loop has no significant contribution in 5' p vs 5' OH selectivity. In

order to achieve good convergence and reasonable statistical error (~1-2 kcal/mol), (**satpati et al., 2011**) we performed a total of >1 μs of alchemical free energy MD simulations. The same methodology adopted for 5' p-dsRNA→5' OH-dsRNA in the "A#" complex (**Kumar & Satpati, 2018**) has been used for "NA" and "NA#" complexes. Apparently, the charge of the simulation system is altered during 5' OH-dsRNA→5' p-dsRNA alchemical transformation and vice versa. But the Tinfoil boundary condition (**Hummer, Pratt, & Gracía, 1996; Bogusz, Cheatham, & Brooks, 1998**) implemented in NAMD (**Phillips et al., 2005**) ensures overall charge neutrality of the simulation system by spreading uniform charge density throughout the simulation box and do not contribute to forces.

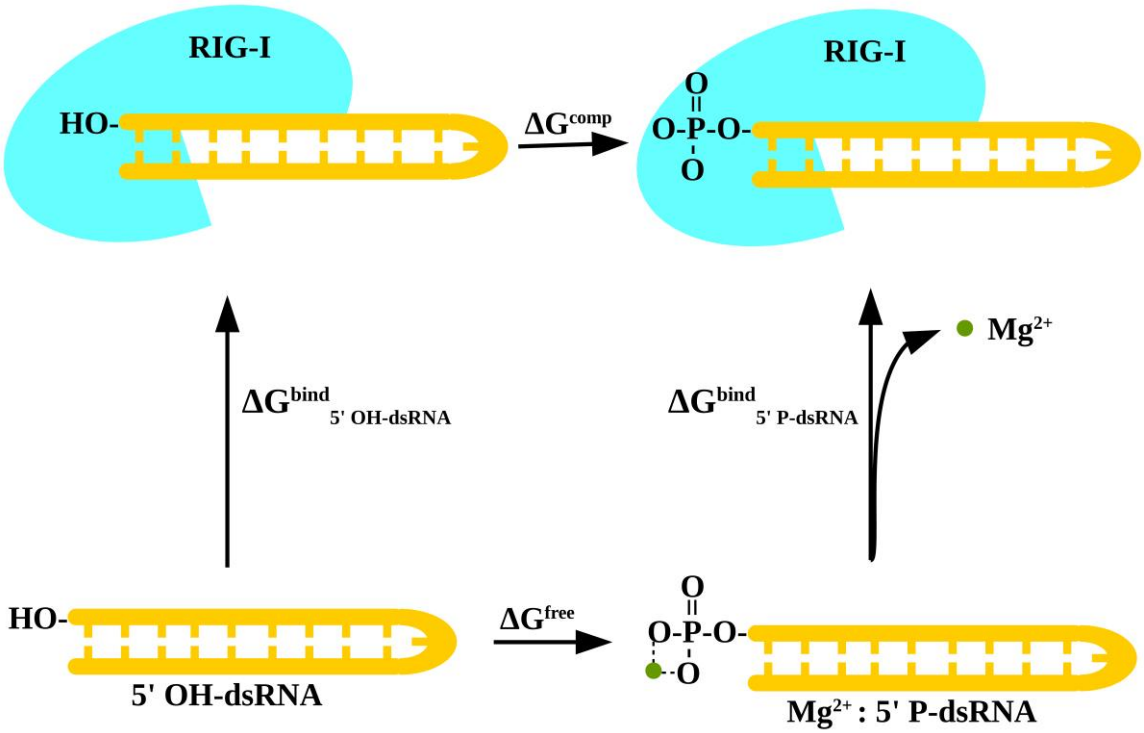


Figure 4.15. Thermodynamic cycle used to estimate relative binding free energies between 5' OH-dsRNA and 5' p-dsRNA to RIG-I. Horizontal legs of the thermodynamic cycle (ΔG^{comp}, ΔG^{free}) were computed by molecular dynamics free energy calculations. The 5' OH-dsRNA/5' p-dsRNA binding free energy difference ($\Delta\Delta G$) is computed using equation (1). Mg^{2+} bound 5' p-dsRNA is considered as the prevalent species which undergo dissociation during RIG-I binding (right vertical arm).

Accurate calculation of absolute binding affinities, i.e, $\Delta G^{bind}_{5'\text{-p-dsRNA}}$ and $\Delta G^{bind}_{5'\text{-OH-dsRNA}}$ (**Figure 4.15**) might be challenging due to significant conformational change associated with RNA:RIG-I binding. Hence, absolute binding affinity calculations will be limited by sampling and convergence issues. Thus, calculation of relative binding free energies using appropriate thermodynamic cycle is a practical route for understanding RIG-I selectivity towards dsRNA's differing at 5'-terminal (5' p vs 5' OH).

4.2.3 Results

4.2.3.1 Energetics of 5' p-dsRNA vs 5' OH-dsRNA Binding to RIG-I

We address dsRNA binding to RIG-I using the thermodynamic cycle in **Figure 4.15**. We compared the binding free energies of 5' OH-dsRNA and 5' p-dsRNA, for both the conformational states, "A" and "NA". We performed MD simulations to follow the horizontal legs of the cycle (**Figure 4.15**) by reversibly and alchemically (**Tembe & Mccammon, 1984**) transforming 5' OH-dsRNA into 5' p-dsRNA. The computed free energy change along alchemical transformation of 5'-terminal (5' OH→5' p) of RNA in complex with RIG-I and free in water was used to estimate the relative binding affinity i.e, ΔΔG. Considering X-ray structures as template, we estimated the relative binding free energy (ΔΔG) for "A" complex (using PDB 4AY2 (**Luo et al., 2012**)) and "NA" (using PDB 5F9F (**Devarkar et al., 2016**)) complex (**Figure 4.16**).

The calculations suggest that both "A" and "NA" complex prefer 5' OH-dsRNA with respect to 5' p-dsRNA by 1.9±0.8 kcal/mol and 1.2±0.6 kcal/mol respectively. ΔΔG for "A" and "NA" complex in the absence of Hel2 loop (model "$A^{\#}$" and "$NA^{\#}$") were estimated to be 1.1±0.4 and 1.4±0.3 kcal/mol respectively. Calculation with different models estimates ΔΔG ~ 1-2 kcal/mol favouring 5' OH-dsRNA binding (**refer appendix Table 4A-2, g**). The sign of the calculated relative binding free energy is correct (**Ren et al., 2019**) and the magnitude is in good agreement with the experiment ($\Delta\Delta G^{exp}$ ~ 0.73 kcal/mol) (**Ren et al., 2019**). Structural studies of RIGI:RNA complexes could not resolve any Mg^{2+} in the close vicinity of 5'-terminus of dsRNA. Interestingly, correct sign of calculated ΔΔG (i.e, RIG-I preference for 5' OH-dsRNA relative to 5' p-dsRNA) only emerges if one considers counter ion (e.g, Mg^{2+}) dissociation from the free 5' p-dsRNA during RIG-I binding, else calculated ΔΔG contradict experiments by incorrectly predicting

RIG-I preference for 5' P-dsRNA over its 5' OH analogue by ~ 2-3 kcal/mol (**refer appendix Table 4A-2**). Calculated relative binding free energy (ΔΔG) is certainly inaccurate if ΔG^{comp} is calculated in the absence of Mg^{2+} and ΔG^{free} is calculated with Mg^{2+} bound state.

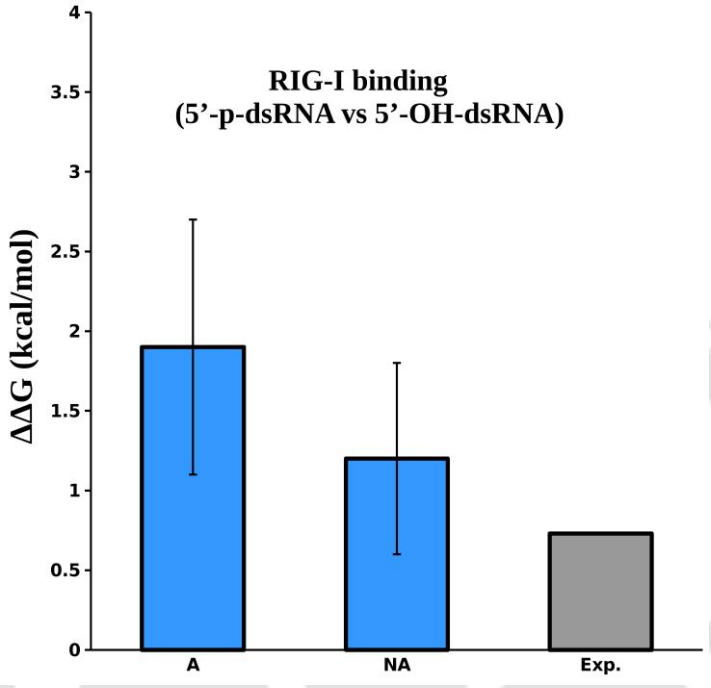


Figure 4.16. Calculated relative binding free energies (in kcal/mol) in the active, inactive complex (blue bars) is denoted A and NA, respectively (Error bars, 1 s.e.m.). Grey bar denote experimentally measured free energy difference.

The reliability of classical fixed charge force field for quantifying biomolecular selectivity is always of great concern. dsRNA with 5' p terminus is expected to be associated with bound counter ion (e.g, Mg^{2+}) which show bidentate coordination with 5'- phosphate in our MD trajectory. Replacing 5' p-dsRNA by 5' OH-dsRNA involves the Mg^{2+} dissociation, hence, removal of Mg^{2+}: phosphate interaction. Now the question is how reliable a fixed charge force field in capturing the 5' p-dsRNA and 5' OH-dsRNA binding to RIG-I when Mg^{2+}:5' P dissociation is expected to take place for the former. Mg^{2+} is expected to polarize the 5' p electronic cloud. Electronic polarization is only implicitly taken into account in the fixed charge force field and known to give large errors (**Ponder & Case, 2003; Warshel, Kato, & Pisliakov, 2007**). Mg^{2+} polarization effect is known (**Satpati et al., 2011**) to cancel out

substantially if Mg^{2+} is present in both the horizontal legs of the Thermodynamic cycle ("additivity" hypothesis).

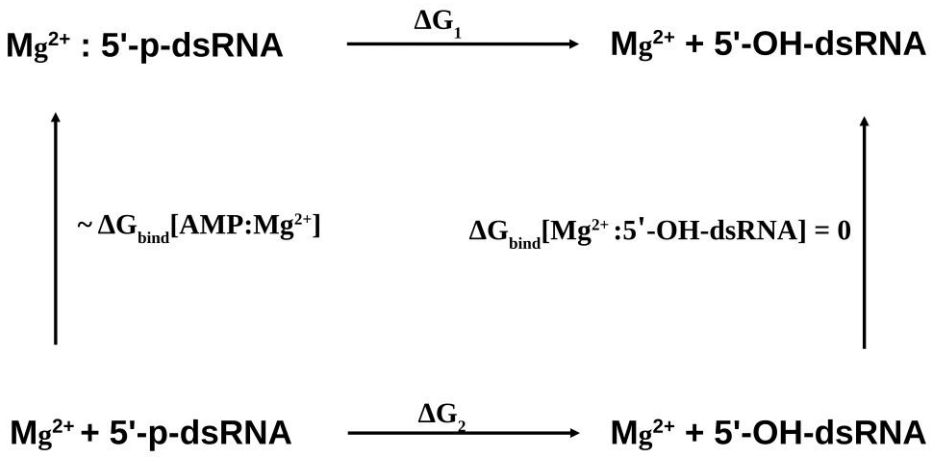


Figure 4.17: Thermodynamic cycle for Mg^{2+}:5' p-dsRNA binding. Mg^{2+} binding is represented by the vertical legs; horizontal legs correspond to the 5' p-dsRNA→5' OH-dsRNA alchemical transformation, either in the Mg^{2+} bound (above) or Mg^{2+} unbound form (below) in water. The 5' p/5' OH binding free energy difference is $\Delta\Delta G = \Delta G_1 - \Delta G_2 = -\Delta G_{bind}(5'\text{ p-dsRNA:}Mg^{2+}) \sim -\Delta G_{bind}(\text{AMP:}Mg^{2+})$. Free energy changes ($\Delta G_1$, ΔG_2) are computed analytically or approximated to be zero (right vertical leg, MD simulations suggest Mg^{2+} is always dissociated from 5' OH terminus of dsRNA).

Table 4.5: Relative binding free energies of Mg^{2+} binding to dsRNA (with 5' p /5' OH).

Alchemical transformation	*With bound* Mg^{2+} **ΔG_1 (refer appendix Table 4A-2 , f)**	*Without* Mg^{2+} **ΔG_2 (refer appendix Table 4A-2, b)**	**$\Delta\Delta G^{calc} = \Delta G_1 - \Delta G_2$**	**$\Delta\Delta G^{expt} = \Delta G_{bind}[Mg^{2+}$:5' OH-dsRNA] - $\Delta G_{bind}[AMP:Mg^{2+}]$**	**$\Delta G^{free} = \Delta G_2 + \Delta\Delta G^{expt}$**
5' -p-dsRNA → 5'-OH-dsRNA	392.5 ± 1.2	342.1 ± 0.2	50.4 ± 1.3	3.8	345.9 ± 0.2

Free energies are in kcal/mol. Statistical uncertainty after ± is given as standard error of the mean. The uncertainties for ΔΔG's were calculated by propagating the uncertainties of individual ΔG's.

In this study, however, error associated with classical fixed charge force field might be significant as the Mg^{2+} bound 5' p-dsRNA and Mg^{2+} free 5' OH-dsRNA is expected to be the prevalent species of free RNA in water. In order to check the force filed accuracy, we computed Mg^{2+} binding to 5' p-dsRNA using appropriate thermodynamic cycle (**Figure 4.17**) and compared the same with the experiment (**Table 4.5**). Mg^{2+} binding to 5' p-dsRNA was computed by alchemically transforming 5' p-dsRNA into 5' OH-dsRNA, either in complex with Mg^{2+} or alone in solution (**Figure 4.17**). MD structures show direct bidentate coordination of Mg^{2+} with the 5' terminus of 5' p-dsRNA. Mg^{2+} was dissociated from the 5' OH-dsRNA terminal. In the presence of Mg^{2+}, we estimated ΔG_1(5' p-dsRNA → 5' OH-dsRNA) = 392.5 (1.2) kcal/mol (electrostatic contribution= 385.4 (1.2) kcal/mol; van der Waals step contribution = 7.1 (0.1) kcal/mol). In the absence of Mg^{2+}, we obtain ΔG_2(5' p-dsRNA → 5' OH-dsRNA) = 342.1 (0.2) kcal/mol (electrostatic contribution= 337.7 (0.2) kcal/mol; van der Waals step contribution = 4.4 (0.1) kcal/mol). Calculated $\Delta\Delta G$ = 50.4 (1.3) kcal/mol, favoring 5' p-dsRNA. Experimentally, the Mg^{2+} dissociation constants from AMP and HPO42- are available (**Verbeeck et al., 1984; Alberty & Goldberg, 1992**), giving a standard binding free energy difference of 3.8 kcal/mol, which is expected to be very close to the computed $\Delta\Delta G$. Thus, in this context, the force field gives a large error (~ 47 kcal/mol) for Mg^{2+}: phosphate binding. Hence, we can conclude that we cannot use ΔG_1 (computed in the presence of Mg^{2+}) and subtract from ΔG^{Comp} (Thermodynamic cycle in the main text) for estimating relative binding of RNA to RIG-I. Reliable ΔG^{free} (**Figure 4.15**) can only be estimated if we include the experimental Mg^{2+} dissociation data and estimate $\Delta G^{free} = \Delta G_2 + \Delta G^{exp}_{bind}$. Thus, ΔG^{free} was computed by including experimental AMP:Mg^{2+} dissociation free energy [ΔG^{expt}(AMP:Mg^{2+}→AMP + Mg^{2+}) = 3.8 kcal/mol] (**Alberty & Goldberg, 1992**), such that ΔG^{free} = ΔG(5' OH→5' p in absence of Mg^{2+})+ΔG^{expt}.

Knowledge of experimental free energy of dissociation of other counter ions (say K^+, Na^+) could also be included (instead of Mg^{2+}) easily in the thermodynamic cycle for $\Delta\Delta G$ estimation. It can be argued that counter-ion dissociation (irrespective of its nature) will certainly weaken the binding affinity of 5' p-dsRNA to RIG-I. Recently MD simulations was performed to understanding structure, dynamics, free energy landscape of nucleic acids in presence of monovalent and divalent ions and details can be found in several excellent

papers (**Burak et al., 2003; Chakraborty et al., 2013; Fischer et al., 2018; Ghosh et al., 2015; Luan et al., 2008; Sarkar et al., 2019**).

4.2.3.2 Structure and Dynamics of RIG-I:5' OH-dsRNA Complex

MD structures of "NA" RIG-I:5' OH-dsRNA complex is very similar to its x-ray template structure (**Figure 4.14, Figure 4.18, and Table 4.6**). CTD-loop and Hel2 loop creates the binding

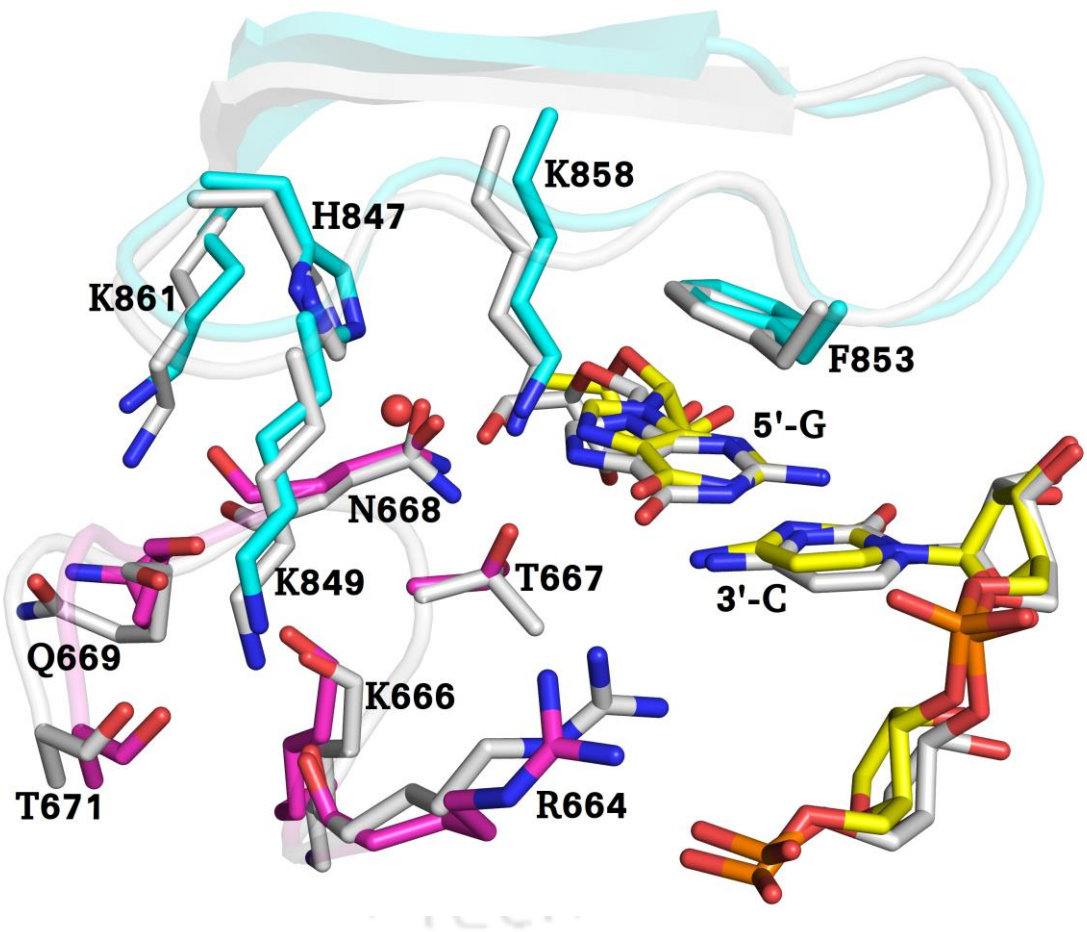


Figure 4.18: Structural comparison around the 5'-terminus of dsRNA in complex with RIG-I. MD snapshot (colored, after 50ns) and its template x-ray structure (gray, PDB code: 5F9F) of RIG-I:5' OH-dsRNA are superimposed. Hydrogens are omitted for clarity and key residues involved in 5'-terminus of dsRNA recognition are shown as sticks.

pocket for 5' OH of dsRNA. CTD-loop stabilizes the terminal base pair by stacking (Phe853) and positioning the Hel2 loop resulting in facile RNA-Hel2 loop interactions

(Arg664 sidechain-RNA backbone, Asn668 sidechain-5'purine G base) (**Figure 4.19**). Throughout the MD trajectory, Phe853 movement parallel to the plane of the terminal G-C pair was observed without disrupting the stacking interaction.

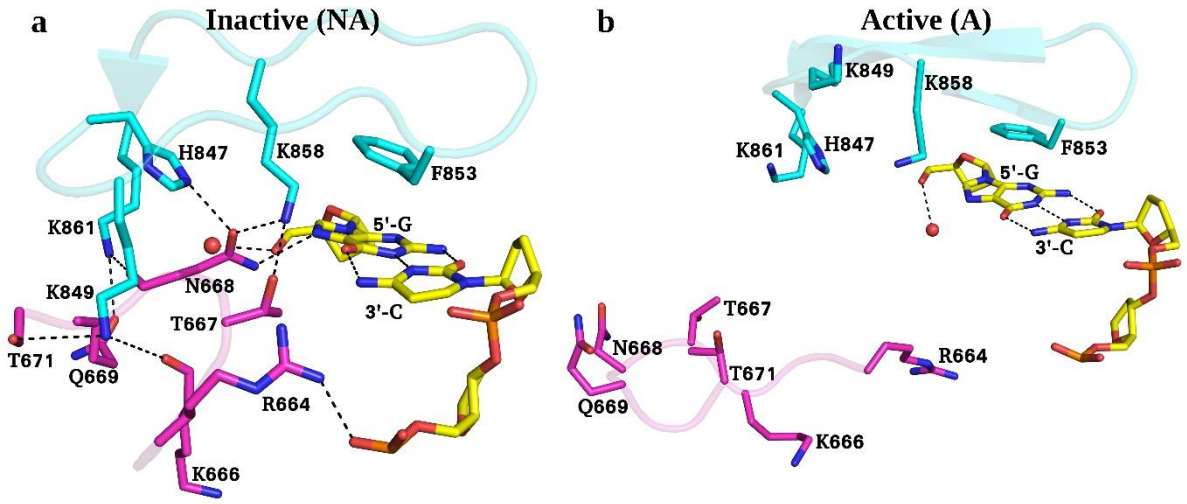


Figure 4.19: Zoomed in view of 5' OH binding pocket from MD simulations: (a) Inactive complex (b) active complex. Water coordinated with the 5' OH is shown as a red sphere. Hydrogens are omitted for clarity. Key residues involved in 5' OH terminus recognition are shown as a stick model. CTD and Hel2 loops are in cyan and magenta respectively.

Table 4.6: MD averaged selected interatomic distances (in angstrom) and standard deviations (after ±).

Interacting Pair		RIG-I:5' OH-dsRNA X-ray (PDB 5F9F)	NA	A	RIG-I:5' P-dsRNA NA	A
Lys858:NZ	Thr667:OG1	3.1	3.1± 0.3	X	X	X
Lys858:NZ	Asn668:OD1	2.9	2.9± 0.3	X	X	X
His847:NE2	Asn668:OD1	2.7	3.4± 0.5	X	X	X
Asn668:O	Lys861:NZ	3.0	2.8± 0.2	X	X	X
Asp872:OD1	Lys861:NZ	3.1	3.0± 1.0	X	2.7± 0.4	2.6± 0.1
Gln669:OE1	Lys861:NZ	4.4	3.0± 0.3	X	X	X
Lys888:NZ	Asp872:OD2	3.7	3.3± 0.8	X	2.8± 0.2	3.0± 0.2
Lys666:O	Lys849:NZ	2.9	3.1± 0.3	X	X	X
Thr671:O	Lys849:NZ	3.0	3.3± 0.8	X	X	X
Arg664:O	Lys849:NZ	2.5	X	X	X	X
5'-G:O2A	Lys888:NZ	X	X	X	2.8± 0.3	2.7± 0.2
5'-G:O2A	Lys861:NZ	X	X	X	X	2.8± 0.2
5'-G:O3A	Lys858:NZ	X	X	X	X	2.7± 0.1
5'-G:N7	Lys858:NZ	X	X	X	X	3.0± 0.2

Absence of Residues/groups in the X-ray/MD structures is represented with a cross.

MD structures also show a direct interaction between water and terminal 5' OH of dsRNA as seen in the x-ray structure (**Ren et al., 2019**). However, none of our MD replicas show a direct interaction between RIG-I and 5' OH terminal of dsRNA. MD trajectories confirm that the gate around 5' OH terminal created by side-chains of CTD-loop (of residue His847, Lys849, Lys858, Lys861) and Hel2 loop (involving side chains of Thr667, Asn668, Gln669, and backbone of Lys666,
Thr671), remains in close conformation in the "NA" complex (**Figure 4.19, a**). Both CTD and Hel2 loops are highly flexible and the interaction network (between CTD and Hel2 loop) was intact and stable throughout the simulations (**Table 4.6**). In the "A" complex, the gate was open and the interaction network was disrupted, except the Phe853 stacking above G-C base pair (**Figure 4.19, b**).

4.2.3.3 Insight into RIG-I:5' p-dsRNA from MD

Despite several attempts, experimental characterization of the structure of RIG-I:5' p-dsRNA complex has not been successful till date (**Ren et al., 2019**). Slow transformation (alchemical transformation, See Methods) of 5' OH into 5' P in the "NA" conformation result in partial gate opening by disrupting the overall interaction network (between CTD and Hel2 loop) observed in the template RIG-I:5' OH-dsRNA complex (**Figure 4.20, a**).

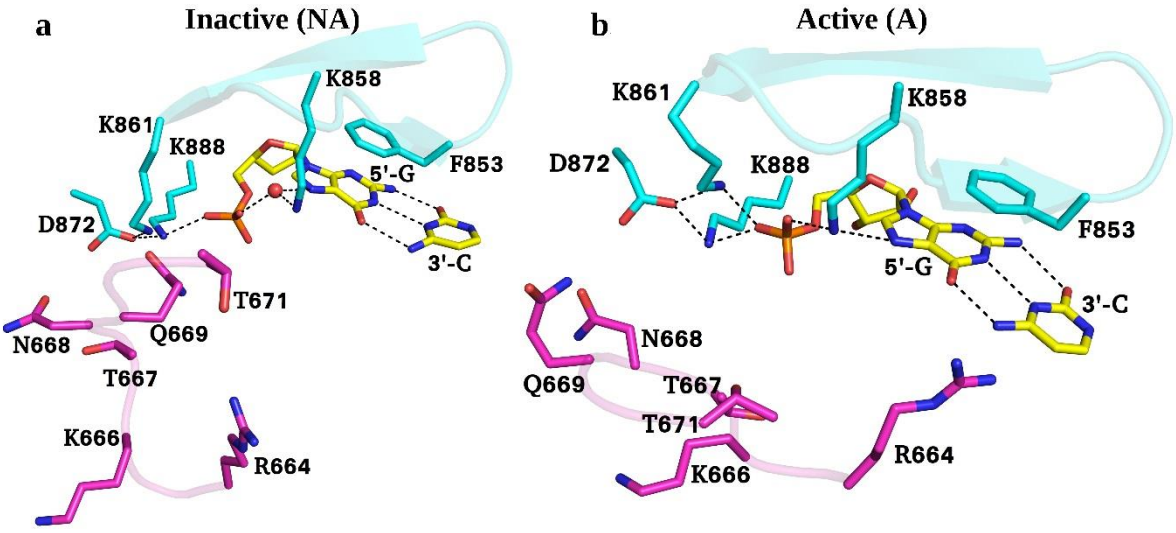


Figure 4.20: Zoomed in view of MD structures of 5' p binding pocket: (a) Inactive complex (b) active complex. Hydrogens are omitted for clarity. Key residues of CTD and Hel2 loops are in cyan and magenta sticks respectively, water as a red sphere, RNA in yellow sticks.

Clearly the space was not enough to place a 5’ p and the steric repulsion was responsible for disrupting the CTD-Hel2 cross talk (**Figure 4.21**, cavity size around 5’ OH), without disrupting the Phe853 stacking above terminal G-C base pair and forming new interactions between lysine side chains (Lys858, Lys 888) and 5’-phosphate terminus of dsRNA (**Figure 4.20, a**).

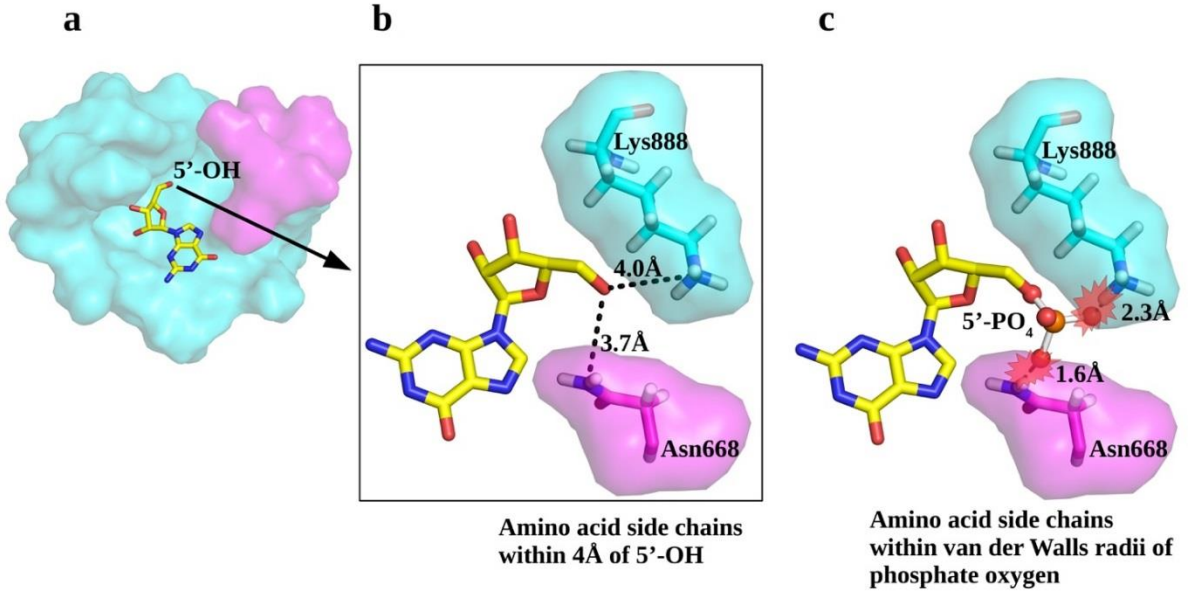


Figure 4.21: (a) X-ray structure of “NA” conformation of RIG-I:5’ OH-dsRNA complex (PDB code: 5F9F). Terminal RNA (yellow), CTD (cyan) and Hel2 loop (Magenta). Distances between heavy atoms are shown as broken black lines. (b) zoon-in view around 5’ OH. Closest amino acid side chains within ≤ 4 Å around 5’ OH is shown explicitly. (c) Replacement of 5’ OH by phosphate (5’ p) indicate steric clash (shown in red transparent star) between 5’ p of Guanine (yellow) and Asn668, Lys888 side chain of RIG-I.

In the “A” complex the Hel2 loop was away from the 5’-p terminal, however Phe853 stacking and direct interactions between lysine’s (Lys858, Lys861, Lys888) and 5’-terminal phosphate were present (**Figure 4.20, b**). The increased flexibility of the CTD and Hel2 loop was suggested from the larger RMSF values with respect to its template (**Figure 4.22**). It can be concluded from the MD structures that the modelled “NA” complex of RIG-I:5’ p-dsRNA is distinctly different from its RIG-I:5’ OH-dsRNA template (**Figure 4.20, a**, **Figure 4.19, a**). Instead of gradually growing 5’ phosphate we also consider “NA” model of RIG-I:5’ p-dsRNA complex by another approach; replacing 5’ OH with 5’ p in the x-ray

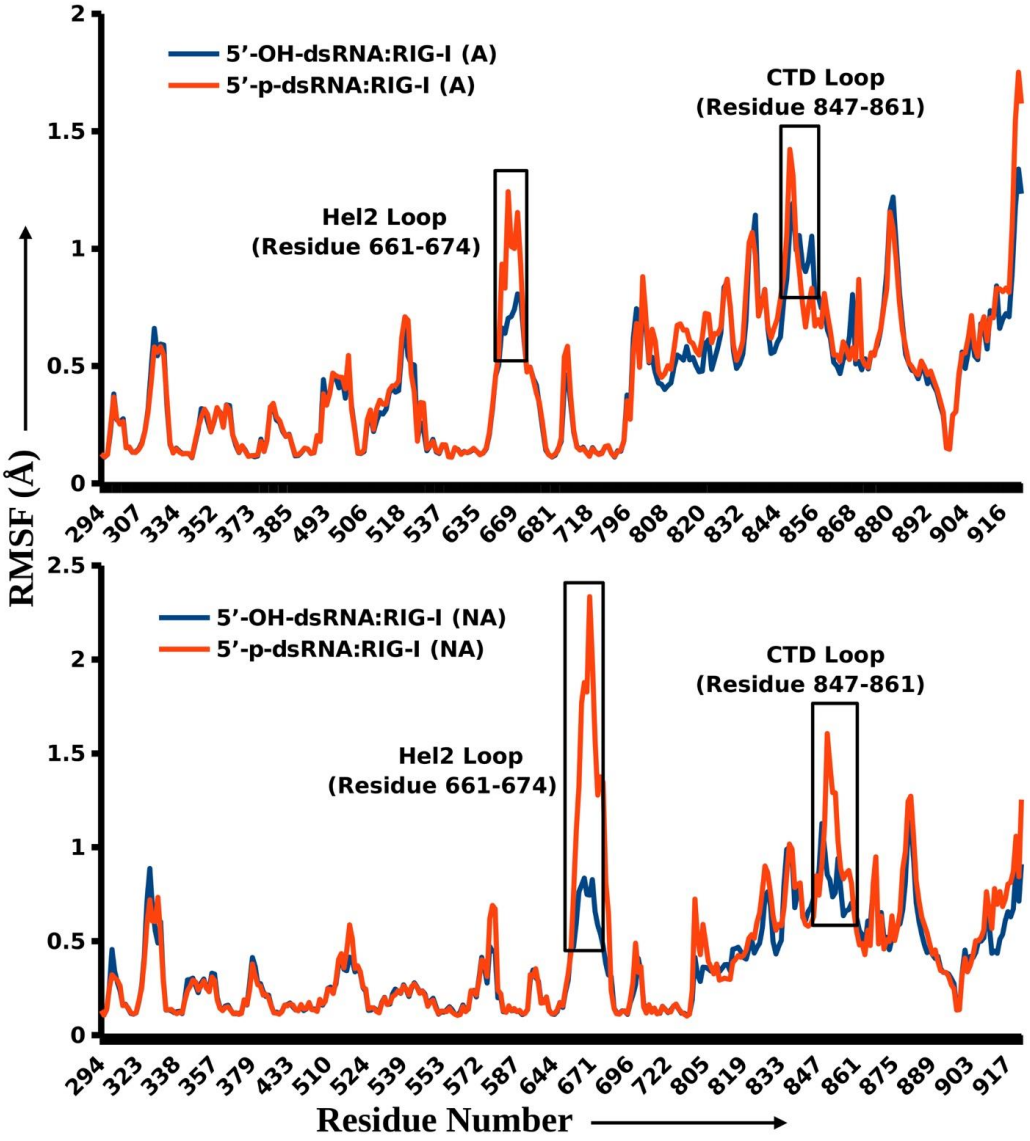


Figure 4.22: (a) Root-mean-square fluctuation (RMSF) of the C_α atom with respect to x-ray structure. Graphs are shown for RIG-I bound 5' p-dsRNA (red) and 5' OH-dsRNA (blue) complexes in the active (upper panel) and inactive (lower panel) complexes. RMSF is shown for last 2ns of production MD trajectory.

structure. MD simulations with the later model as initial geometry result in severe structural distortion both in the CTD (key Phe853 stacking is lost) and Hel2 loop (**Figure 4.23**). Stacking of phenyl side chain of Phe853 to the terminal GC base pair of dsRNA in complex with RIG-I is a characteristic feature observed in all the experimentally resolved RIG-I:dsRNA complexes. Biochemical studies (**Lu et al., 2010**) also confirmed that substitution

of Phe853 by serine reduce dsRNA binding to RIG-I, suggesting Phe853 is important for effective RNA binding. Based on the overall structural integrity, it may be speculated that a slow grow approach might a better choice for modelling RIG-I:5' p-dsRNA "NA" complex.

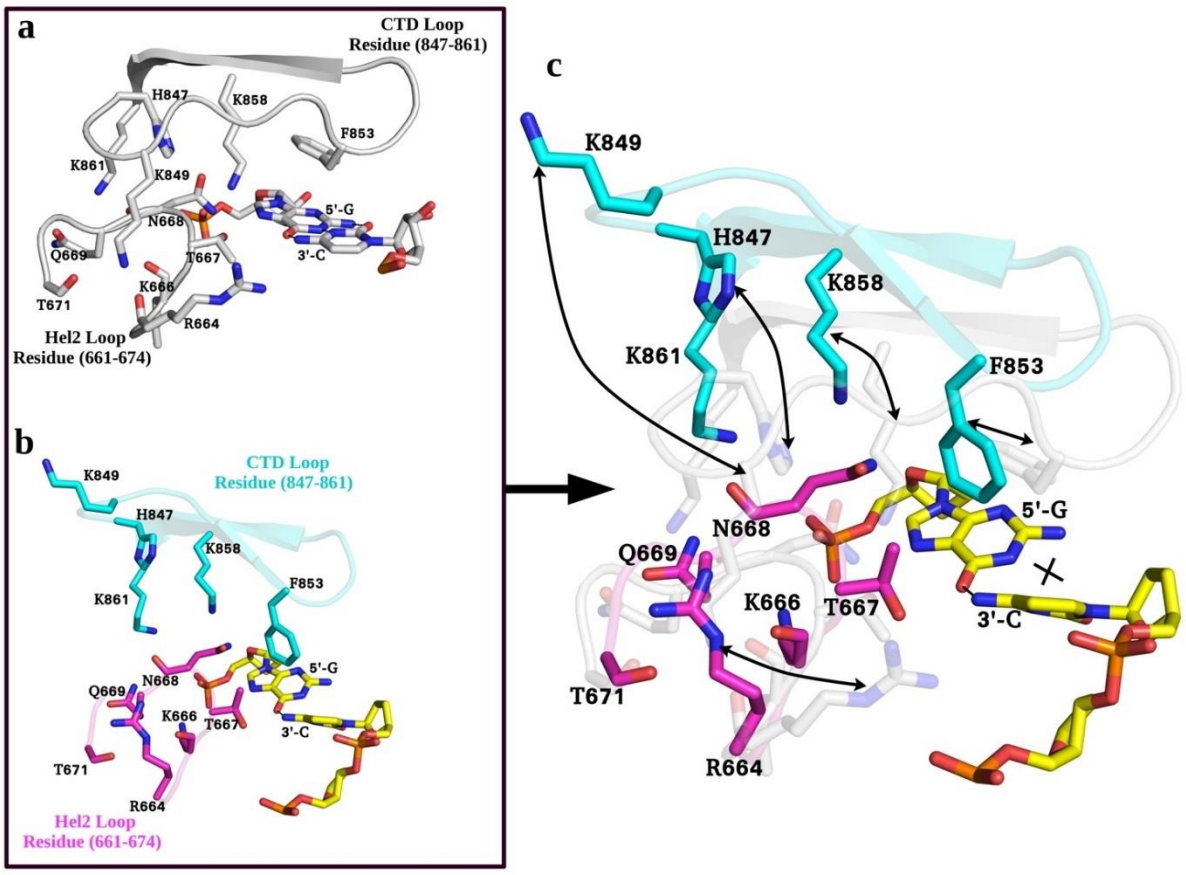


Figure 4.23: (a) Key residues around the 5' OH terminal of RIG-I:5' OH-dsRNA complex in the x-ray structure (PDB 5F9F, Gray). (b) 5' OH was replaced with 5' p in the x-ray structure of RIG-I:5' OH-dsRNA (PDB 5F9F) and the resulting model was subjected to MD simulations. Representative MD snapshot of modelled RIG-I:5' p-dsRNA complex (colored) after 18 ns. (c) Overlay of MD snapshot (colored) and template x-ray structure (grey). Deviation of amino-acid side chains of CTD and Hel2 loop (between Colored/MD structure and Gray/X-ray template) is indicated by double headed arrow. Hydrogen not shown for clarity.

4.2.4 Discussion

The connection between the computed structure-based energetics of Figure 1c and experimental kinetics (**Ren et al., 2019**) is of utmost interest. X-ray structures of RIG-I:RNA complexes did not report the presence of counterion in the vicinity of 5'-terminal phosphates/OH of dsRNA. However the ATPase site away from the 5'-terminal do have bound Mg^{2+}(**Luo et al., 2012**, see **Figure 4.13, a**: ADP (magenta ball and stick)-Mg^{2+}(green sphere)). A plausible reason could be the dissociation of counter ion from the 5'-terminal phosphate(s) of dsRNA during binding. A cluster of positively charged sidechains involving 5 lysine residue of RIG-I in the binding pocket of 5'-terminal of dsRNA, could be a possible reason for Mg^{2+} absence. ATP^{4-} in water is known (**Storer & Cornish-Bowden, 1976**) to exist prevalently in $[ATP:Mg]^{2-}$ form with traces of $[ATPH:Mg]^{-}$, $[ATP:Mg_2]$, and unbound $[ATP]^{4-}$. The distribution is the same for ADP^{3-} and AMP^{2-}. Hence, dsRNA with 5'-phosphate(s) may be assumed to be bound to Mg^{2+} with a fully deprotonated state in water. On the contrary, neutral 5' OH terminal of free dsRNA in water does not bind Mg^{2+}. The calculated relative binding free energies ($\Delta\Delta G$) suggest that both "A" and "NA" complex weakly prefer 5' OH-dsRNA relative to 5' p-dsRNA, which is in line with the experiments (**Ren et al., 2019**). Calculations considering "A" and "NA" complex estimated $\Delta\Delta G$ of 1.9±0.8 kcal/mol and 1.2±0.6 kcal/mol respectively, later being closer to the experiments ($\Delta\Delta G^{exp} \sim 0.73$ kcal/mol) (**Ren et al., 2019**). Distinctly different orientation of the Hel2 loop (between "A" and "NA") relative to CTD does not alter the relative binding affinity significantly (< 1 kcal/mol). The results do not exclude the fact that Hel2 loop could alter the absolute binding affinity, but predict that Hel2 loop may be insignificant towards relative binding affinity (5' OH vs 5' p dsRNA). Counter-ion dissociation from 5' p terminal of dsRNA turns out to be a key factor, without which $\Delta\Delta G$ estimation gives the wrong sign, preferring 5' p-dsRNA over 5' OH-dsRNA. 5' p terminal of dsRNA establish favourable interactions with side chains of lys858, lys861, and lys888 of RIG-I (**Figure 4.20, a,b**). No direct interaction between 5' OH terminal of dsRNA and RIG-I was observed in x-ray structures (**Ren et al., 2019; Jiang et al., 2011; Kohlway et al., 2013; Luo et al., 2011; Lassig et al., 2018**) as well as in MD (**Figure 4.18, a**,**b**). Apparently, 5' p-dsRNA binding to RIG-I should be preferred relative to 5' OH-dsRNA. However, favourable RIG-I:5' p interaction observed in the MD structures (**Figure 4.20**) could be offset by the Mg^{2+} dissociation penalty from free 5' p-dsRNA during binding, disfavouring 5' p-dsRNA

binding relative to 5' OH-dsRNA. We predict that Mg^{2+} dissociation from 5'-terminal of dsRNA during binding, in fact, provides a very efficient way of fine-tuning the dsRNA preference based on the difference of few atoms at the 5'-terminal.

Based on the relative orientation of Hel2 loop with respect to CTD two alternate conformation "A" and "NA" of RIG-I was proposed (**Figure 4.13**). Hence, conformational selection driven by 5'-terminal of dsRNA ("NA" selected only by 5' OH/p-dsRNA, "A" is selected by 5'ppp/pp-dsRNA) might be linked to function. It should be noted that "NA" conformation is only resolved in one x-ray structure (PDB 5F9F (**Ren et al., 2019**)), whereas the same Hel2 loop is not visible in other higher resolution structures of 5' OH-dsRNA bound RIG-I complexes **(Jiang et al., 2011; Kohlway et al., 2013; Luo et al., 2011; Lassig et al., 2018**). It is unclear whether the "NA" complex is a high energy minor state which is stabilized by crystal packing and/or crystallization procedure. MD studies show that modelling phosphate in the "NA" complex alters the overall architecture of the "NA" state by disrupting interaction between CTD and Hel2 loop (partial gate opening in **Figure 4.23**), as speculated previously (**Ren et al., 2019**).

4.2.5 Conclusion

RIG-I weakly favours 5' OH-dsRNA binding relative to 5' P-dsRNA, which is a characteristic feature of RIG-I (**Ren et al., 2019**). The principle for 5' OH-dsRNA vs 5' P-dsRNA selectivity by RIG-I, favouring the former is not fully understood. Structural studies proposed two different conformations ("active: A" and "inactive: NA") of RIG-I in complex with dsRNA (**Figure 4.13, a, b**). Using those two conformations as a template we estimated the relative binding free energies for RIG-I with 5' p-dsRNA and 5' OH-dsRNA (**Figure 4.16**). We propose a plausible mechanism in which the Mg^{2+} dissociation from the 5' p terminal of dsRNA is a requirement for RIG-I binding (**Figure 4.15**), which offsets the favourable 5' p:RIG-I interaction resulting weaker binding affinity relative to 5' OH-dsRNA. The proposed mechanism is successful in explaining the experimentally observed RIG-I selectivity ($\Delta\Delta G$) between 5' p-dsRNA and 5' OH-dsRNA (**Figure 4.16**). We predict that the Mg^{2+} dissociation from the negatively charged 5' p terminal of dsRNA, provides a very efficient way of controlling the RIG-I selectivity favouring 5' OH over 5' p dsRNA.

Calculation of absolute binding free energies (ΔG_{bind}) is challenging due to large scale conformational changes, the coupling between binding and ionization states of titratable groups, changes in ionic environments, limited force field accuracies, etc. Calculation of relative binding free energies ($\Delta\Delta G$) by combining MD simulations and statistical mechanics has been proved to be reasonably accurate in predicting protein-RNA binding affinity differences between congeneric RNA molecules (**Lind et al., 2019**). Our structure base energetics study provide a possible clue of how a simple counter ion dissociation from the 5' terminal of dsRNA during RIG-I binding could be used for efficient recognition. Our simulation strategy can be used to study how mutation in RIG-I (e.g, F853A, K858A, H847A, etc.) could alter the dsRNA binding ability.

Chapter 5

Concluding Remarks and Future Perspective

Protein-RNA interactions play critical roles in several cellular processes and serve numerous purposes within their respective biological processes. Advancement in structural characterization of protein-RNA complexes at different stages of the biochemical pathway along with increasing computational power enable Molecular Dynamics simulations as an effective method to estimate the structure-based energetics of protein-RNA interaction recognition. In this thesis structure-based classical molecular dynamics free energy calculations were performed to estimate thermodynamics of protein-RNA recognition events involved in gene expression and viral-RNA recognition by host protein. An attempt has been made to link the computed energetics with 3D structures and experimental kinetics. In **Chapter 1**, a brief overview of the protein-RNA events ($tRNA^{Ala}$ recognition by AlaRS: **Chapter 2**, Stop codons recognition by eukaryotic release factor: **Chapter 3**, and Viral RNA recognition by RIG-I protein: **Chapter 4**) studied in the thesis was presented. The methodology adopted in the thesis was presented next, followed by objectives of the thesis.

Chapter 2 discussed the principle of alanine tRNA selection by alanine tRNA synthetase based on the critical G3•U70 base pair and reported the quantitative estimation of relative binding free energies associated with AlaRS binding to 3•70 mutant $tRNA^{Ala}$ with respect to wild-type $tRNA^{Ala}$/G3•U70. It was found that the reactive complex strongly favors $tRNA^{Ala}$/G3•U70 with respect to the A3•U70 variant and the geometrical difference between G3•U70 and A3•U70 does not play any significant role in driving the conformational change (NR ⇆ R) in free $tRNA^{Ala}$ in water. Based on our computed relative binding free energetics, we have also shown how a simple three-state kinetic model can explain how the single wobble pair ensures tRNA specificity by altering k_{cat} (favoring correct tRNA by ~ 100 fold) while keeping K_M more or less similar.

Chapter 3 provides crucial insights into the energetics of stop codon (UAA, UGA, and UAG) recognition by eukaryotic release factor 1 (eRF1). Computed results suggest that eRF1 imposes a very high energetic penalty for binding to sense codon (CAA, UGG) programmed mRNA in the ribosomal A-site. eRF1 ensures high selectivity by (a)

introducing strain in the mRNA, (b) disrupting protein-mRNA interactions, and (c) placing sense-codons in the dry desolvated pocket with unsatisfied h-bonds in the near-cognate complexes. Results also suggest that eRF1 has a higher discriminatory power against sense codons than its bacterial analog (RF1/RF2) reported previously.

Chapter 4 has explored the energetics of viral RNA recognition by RIG-I. RIG-I receptor recognizes viral RNAs (bearing 5' tri/diphosphates: 5' ppp/pp-dsRNA) in the cell cytosol and initiates an antiviral response, but unresponsive to similar host RNA's (bearing 5'-monophosphate/hydroxyl: 5' P/OH-dsRNA). Experiments confirmed (1) 5' pp-dsRNA is the minimum requirement for RIG-I induced antiviral response, (2) RIG-I disfavour 5' p-dsRNA binding and the binding affinity is much weaker than 5' OH-dsRNA. We have divided the results of this chapter in two parts as following: the first part discusses viral (5'-ppp/pp-dsRNA) and host (5'-p) dsRNA discrimination by RIG-I; the second part discusses RIG-I discrimination between host RNA's (5'-p-dsRNA vs. 5'-OH-dsRNA). Based on our computed results, we have proposed a possible mechanism of RNA selection by RIG-I. We proposed that dsRNA binding to RIG-I requires Mg^{2+} dissociation from its 5' terminus dsRNA. The free energy cost associated with Mg^{2+} dissociation from RIG-I could tune the energetics and results in (1) weak discrimination by RIG-I between 5'-ppp vs. 5'-pp dsRNA, and (2) preferential binding of 5'-OH-dsRNA over 5'-p-dsRNA to RIG-I. We have also provided a possible clue of how a simple counter ion dissociation from the 5' terminal of dsRNA during RIG-I binding could be used for efficient recognition.

In this thesis, we have described that structure-based molecular dynamics free energy calculations is a practical and sufficiently accurate method to answer some of the most fundamental questions regarding specificity and accuracy of biological processes involving proteins and RNAs. The adopted methodology is successful in bridging the 3D structures (cognate or non-cognate complexes) of biomolecules and free energy differences. The simulation strategy could be used to answer the following issues in the future:

(1) Does mutation in eRF1 motifs (TASNIKS, YxCxxxF, and GTS loop) alter its binding affinity to stop codon?

(2) Mutation in AlaRS (e.g., D45A/N359A/R483A) and its ability to recognize $tRNA^{Ala}$.

(3) Energetics of dsRNA binding in response to RIG-I mutation (viz., F853A/K858A/H847A, etc.).

Appendices

Appendix 2A

Table 2A-1:

(a) Free energy change for the alchemical transformation of G3·U70→A3·U70 in reactive complex of AlaRS:$tRNA^{Ala}$.

System	**No of runs**	**Simulation Length (Fwd+Rev)**	**Forward ΔG^{comp}** (G3·U70) → (A3·U70)	**Reverse ΔG^{comp}** (A3·U70) → (G3·U70)
Reactive State (R): AlaRS:$tRNA^{Ala}$ /(G3·U70)	Run1	34 ns	125.23 (1.91)	-121.06 (1.84)
	Run2	34 ns	126.39 (1.33)	-122.37 (1.27)
Forward G-->A	Run3	34 ns	128.55 (1.52)	-121.46 (2.01)
Reverse A-->G	Run4	34 ns	127.90 (1.56)	-120.65 (1.43)
	Run5	34 ns	126.39 (1.91)	-121.33 (2.37)
Average			**126.89 (0.59)**	**-121.37 (0.28)**

(b) Free energy change for the alchemical transformation of G3/A3→A3/G3 in reactive/non-reactive conformation of $tRNA^{Ala}$ free in water. G3 in non-reactive conformation of free $tRNA^{Ala}$ was modelled by replacing A3 by G3.

System	**No of runs**	**Simulation Length (Fwd+Rev)**	**Forward ΔG_{free}^{R}** (G3·U70) → (A3·U70)	**Reverse ΔG_{free}^{R}** (A3·U70) → (G3·U70)
tRNA free in water (in Reactive conformation): $tRNA^{Ala}$ /(G3·U70)	Run1	34 ns	120.97 (1.49)	-121.63 (0.35)
	Run2	34 ns	121.25 (1.14)	-121.13 (2.13)
Forward G-->A	Run3	34 ns	120.14 (1.49)	-122.46 (1.58)
Reverse A-->G	Run4	34 ns	122.20 (1.45)	-119.39 (1.76)
	Run5	34 ns	121.40 (1.03)	-121.85 (1.10)
Average			121.19 (0.33)	-121.29 (0.52)
Forward+Reverse Average: ΔG_{free}^{R}				**121.24 (0.43)**

System	**No of runs**	**Simulation Length (Fwd+Rev)**	**Forward ΔG_{free}^{NR}** (A3·U70) → (G3·U70)	**Reverse ΔG_{free}^{NR}** (G3·U70) → (A3·U70)
tRNA free in water (Non-reactive conformation): $tRNA^{Ala}$ /(A·U) Forward A-->G Reverse G-->A	Run1	34 ns	-121.65 (1.57)	121.21 (1.57)
	Run2	34 ns	-121.72 (1.76)	122.03 (1.02)
	Run3	34 ns	-121.67 (1.20)	121.71 (1.38)
	Run4	34 ns	-121.57 (0.99)	121.33 (0.93)
	Run5	34 ns	-120.57 (1.17)	121.43 (0.91)
Average			-121.43 (0.22)	121.54 (0.15)
	No of runs	**Simulation Length (Fwd+Rev)**	**Forward ΔG_{free}^{NR}** (G3·U70) → (A3·U70)	**Reverse ΔG_{free}^{NR}** (A3·U70) → (G3·U70)
tRNA free in water (Non-reactive conformation): $tRNA^{Ala}$ /(G·U) - MODELEED Forward G-->A Reverse A-->G	Run1	34 ns	121.41 (1.15)	-121.36 (1.62)
Average			121.41 (1.15)	-121.36 (1.62)
Forward + Reverse Average of G3→A3, ΔG_{free}^{NR}				**121.44 (0.79)**
$\Delta\Delta G_{free}^{NR\rightarrow R}$ (G3.U70 vs A3.U70) = ΔG_{free}^{R} - ΔG_{free}^{NR}				**-0.2 (0.9)**

(c) Free energy change for the alchemical transformation of G3/A3·U→A3/G3·U in AlaRS:$tRNA^{Ala}$ non-reactive complex.

System	**No of runs**	**Simulation Length (Fwd+Rev)**	**Forward ΔG** (A·U) → (G·U)	**Reverse ΔG** (G·U) → (A·U)
Non-reactive State: AlaRS:$tRNA^{Ala}$ /(A3·U70) Forward A-->G Reverse G-->A	Run1	34 ns	-122.54 (1.62)	122.37 (1.63)
	Run2	34 ns	-119.96 (1.59)	118.35 (1.81)
	Run3	34 ns	-121.25 (1.64)	121.85 (2.14)
	Run4	34 ns	-123.98 (1.43)	118.48 (2.57)
	Run5	34 ns	-120.60 (1.98)	120.84 (1.08)
Average			**-121.67 (0.72)**	**120.38 (0.84)**

System	No of runs	Simulation Length (Fwd+Rev)	Forward ΔG (G·U)→(A·U)	Reverse ΔG (A·U)→(G·U)
Non-reactive State: AlaRS:tRNA^{Ala} /(G3·U70) - [A3 replaced by reactive state G3.] Forward G-->A Reverse A-->G	Run1	34 ns	123.78 (1.84)	-119.11 (0.66)
	Run2	34 ns	121.73 (1.70)	-119.10 (2.32)
	Run3	34 ns	123.66 (1.37)	-119.40 (1.34)
	Run4	34 ns	122.82 (1.80)	-120.64 (2.58)
	Run5	34 ns	122.17 (1.53)	-122.32 (1.27)
	Average		**122.83 (0.40)**	**-120.11 (0.62)**

(d) Free energy change for the alchemical transformation of G3·U70→G3·C70 in AlaRS:tRNA^{Ala} reactive complex

System	No of runs	Simulation Length (Fwd+Rev)	Forward ΔG (G·U)→(G·C)	Reverse ΔG (G·C) →(G·U)
Reactive State: AlaRS:tRNA^{Ala} /(G·U) Forward U-->C Reverse C-->U	Run1	22 ns	14.55 (0.19)	-12.96 (1.00)
	Run2	22 ns	12.70 (1.33)	-12.02 (1.00)
	Run3	22 ns	14.01 (1.08)	-14.50 (0.99)
	Run4	22 ns	13.87 (0.54)	-14.40 (0.78)
	Run5	22 ns	13.06 (0.93)	-11.04 (0.98)
	Average		**13.64 (0.33)**	**-12.98 (0.67)**

(e) Free energy change for the alchemical transformation of G3·U70→G3·C70 in AlaRS:tRNA^{Ala} non-reactive complex

System	No of runs	Simulation Length (F+R)	Forward ΔG (G·U)→(G·C)	Reverse ΔG (G·C)→(G·U)
Non-reactive State: AlaRS:tRNA^{Ala} /(G·U) Forward U-->C Reverse C-->U	Run1	22 ns	14.03 (1.14)	-13.23 (1.10)
	Run2	22 ns	15.95 (0.89)	-12.82 (0.96)
	Run3	22 ns	13.36 (0.66)	-11.38 (0.88)
	Run4	22 ns	13.82 (0.94)	-11.71 (0.75)
	Run5	22 ns	13.27 (0.51)	-11.93 (0.75)
	Average		**14.09 (0.49)**	**-12.21 (0.35)**

(f) Free energy change for the alchemical transformation of G3·U70→G3·C70 in reactive and non-reactive $tRNA^{Ala}$ free in water.

System	**No of runs**	**Simulation Length (F+R)**	**Forward ΔG** (G·U)→(G·C)	**Reverse ΔG** (G·C)→(G·U)
tRNA free in water (Reactive conformation): $tRNA^{Ala}$ /(G·U) Forward U-->C Reverse C-->U	Run1	22 ns	7.65 (0.11)	-7.37 (0.22)
	Run2	22 ns	7.86 (0.78)	-8.01 (0.51)
	Run3	22 ns	7.64 (0.76)	-8.07 (0.72)
	Run4	22 ns	7.74 (0.65)	-7.97 (0.93)
	Run5	22 ns	6.62 (0.94)	-8.64 (0.84)
	Average		**7.50 (0.22)**	**-8.01 (0.20)**

System	**No of runs**	**Simulation Length (F+R)**	**Forward ΔG** (G·U)→(G·C)	**Reverse ΔG** (G·C)→(G·U)
tRNA free in water (Non-reactive conformation): $tRNA^{Ala}$ /(G·U) Forward U-->C Reverse C-->U	Run1	22 ns	8.02 (0.71)	-8.01 (0.66)
	Run2	22 ns	7.92 (0.45)	-7.74 (0.75)
	Run3	22 ns	7.68 (0.69)	-7.59 (0.64)
	Run4	22 ns	7.70 (0.72)	-8.14 (0.51)
	Run5	22 ns	7.78 (0.72)	-8.08 (0.64)
	Average		**7.82 (0.06)**	**-7.91 (0.11)**

(g) Free energy change for the alchemical transformation of A3·U70→A3·C70 in AlaRS:$tRNA^{Ala}$ reactive complex where A3 taken from non-reactive complex.

System	**No of runs**	**Simulation Length (F+R)**	**Forward ΔG** (A·U)→(A·C)	**Reverse ΔG** (A·C)→(A·U)
Reactive State: AlaRS:$tRNA^{Ala}$ /(A3·U70)-A3 modelled from Non-Reactive state Forward U-->C Reverse C-->U	Run1	22 ns	18.35 (1.23)	-18.12 (0.60)
	Run2	22 ns	18.55 (1.44)	-17.69 (0.70)
	Run3	22 ns	17.25 (0.66)	-17.45 (1.59)
	Run4	22 ns	18.60 (0.91)	-17.49 (0.87)
	Run5	22 ns	17.74 (1.39)	-18.00 (1.42)
	Average		**18.10 (0.26)**	**-17.75 (0.13)**

(h) Free energy change for the alchemical transformation of A3·U70→A3·C70 in AlaRS:tRNA^{Ala} non-reactive complex

System	No of runs	Simulation Length (F+R)	Forward ΔG (A·U)→(A·C)	Reverse ΔG (A·C)→(A·U)
Non-reactive State: AlaRS:tRNA^{Ala} /(A·U)	Run1	22 ns	17.68 (0.38)	-18.59 (1.16)
	Run2	22 ns	17.72 (1.16)	-15.56 (0.55)
Forward U-->C Reverse C-->U	Run3	22 ns	18.23 (1.52)	-16.6 (1.00)
	Run4	22 ns	18.25 (1.35)	-17.77 (1.27)
	Run5	22 ns	18.11 (1.44)	-18.71 (1.60)
	Average		**18.00 (0.12)**	**-17.45 (0.60)**

(i) tRNA free in water Free energy change for the alchemical transformation of A3·U70→A3·C70 in reactive and non-reactive tRNA^{Ala} free in water where A in reactive tRNA^{Ala} taken from non-reactive complex.

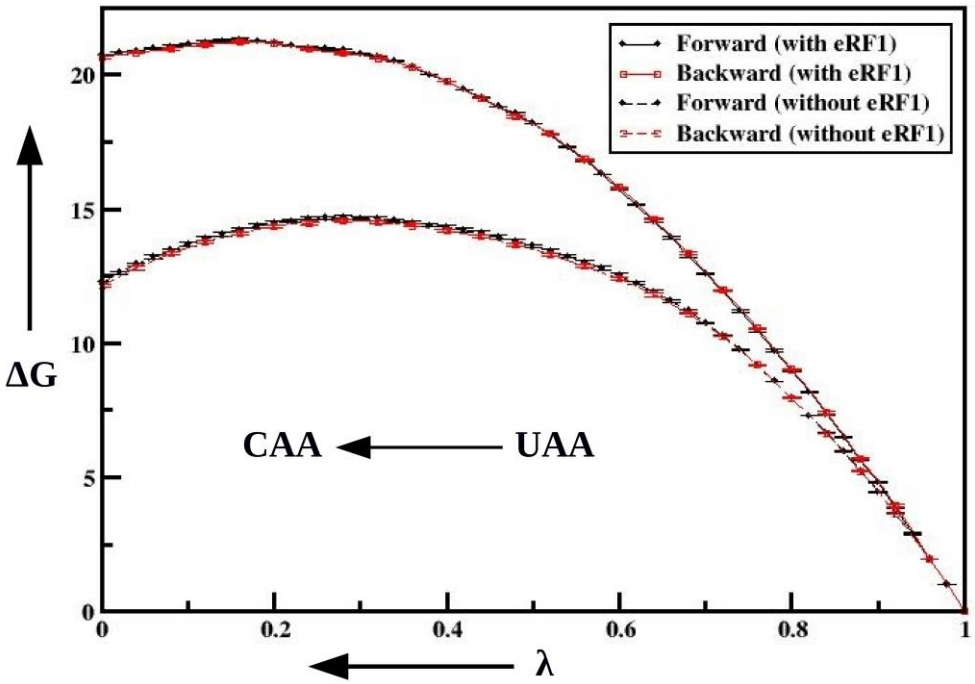
System	No of runs	Simulation Length (F+R)	Forward ΔG (A·U)→(A·C)	Reverse ΔG (A·C)→(A·U)
tRNA free in water (Reactive conformation): tRNA^{Ala} /(A·U)-"A" Modelled from Non-reactive state	Run1	22 ns	15.20 (0.87)	-14.88 (0.47)
	Run2	22 ns	14.24 (1.09)	-14.25 (0.74)
	Run3	22 ns	14.43 (1.04)	-14.02 (0.72)
	Run4	22 ns	13.41 (0.59)	-14.87 (0.78)
Forward U-->C Reverse C-->U	Run5	22 ns	14.18 (0.97)	-13.95 (0.87)
	Average		**14.29 (0.29)**	**-14.39 (0.20)**
System	**No of runs**	**Simulation Length (F+R)**	**Forward ΔG** (A·U)→(A·C)	**Reverse ΔG** (A·C)→(A·U)
Non-reactive conformation of tRNA free in water: tRNA^{Ala} /(A·U)	Run1	22 ns	14.60 (0.38)	-13.26 (0.45)
	Run2	22 ns	14.02 (0.91)	-11.50 (0.97)
	Run3	22 ns	14.45 (0.84)	-14.68 (0.89)
Forward U-->C Reverse C-->U	Run4	22 ns	14.92 (0.76)	-14.11 (1.15)
	Run5	22 ns	14.82 (1.14)	-14.41 (0.73)
	Average		**14.56 (0.16)**	**-13.59 (0.57)**

MD trajectories were divided into two equal halves and the difference between the computed ΔG's from the two-halves is reported as uncertainty in the parenthesis for

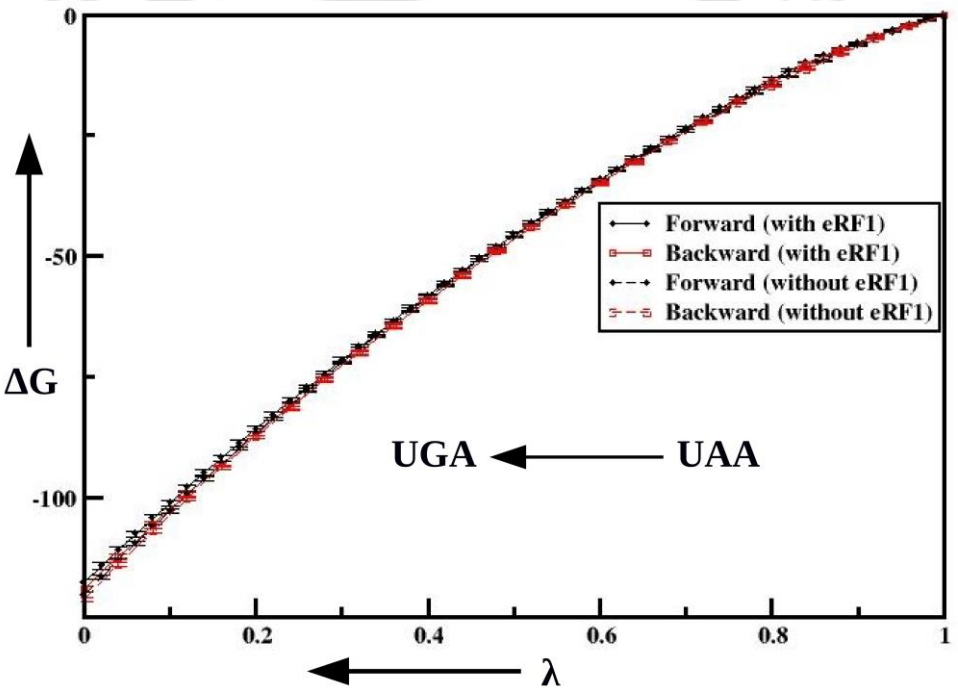
individual replicas. Uncertainty in the average result is reported as standard error (ΔG's from different replicas). Hysteresis of > 5 kcal/mol observed.

Appendix 3A

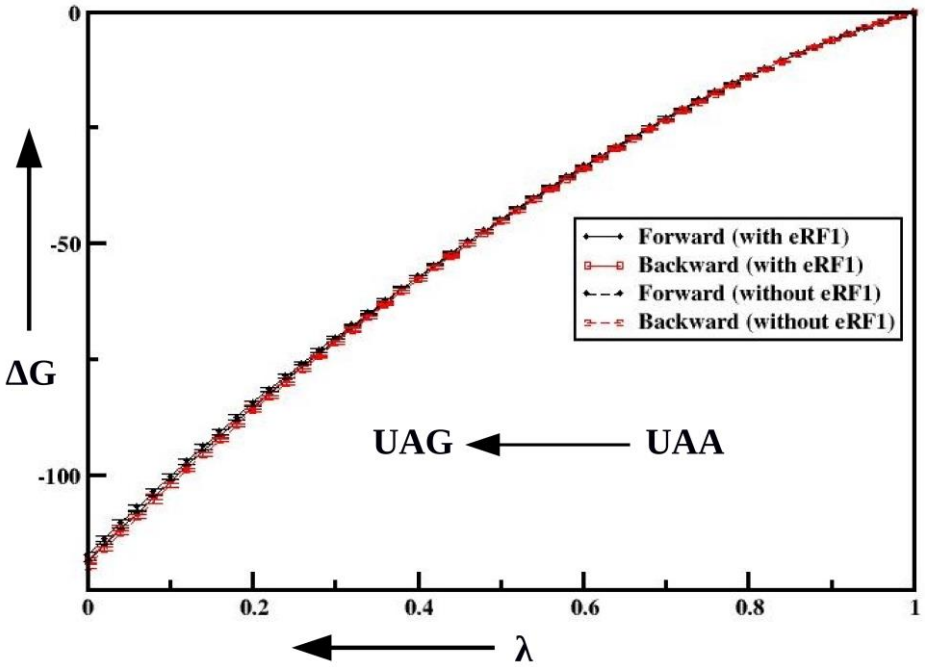
(a) Free energy vs λ (coupling coordinate) plot for alchemical transformation U→C in stop codon programmed ribosome with and without eRF1.



(b) Free energy vs λ (coupling coordinate) plot for alchemical transformation A→G in stop codon programmed ribosome with and without eRF1.



(c) Free energy vs λ (coupling coordinate) plot for alchemical transformation A→G in stop codon programmed ribosome with and without eRF1.



(d) Free energy vs λ (coupling coordinate) plot for alchemical transformation A→G in stop codon programmed ribosome with and without eRF1.

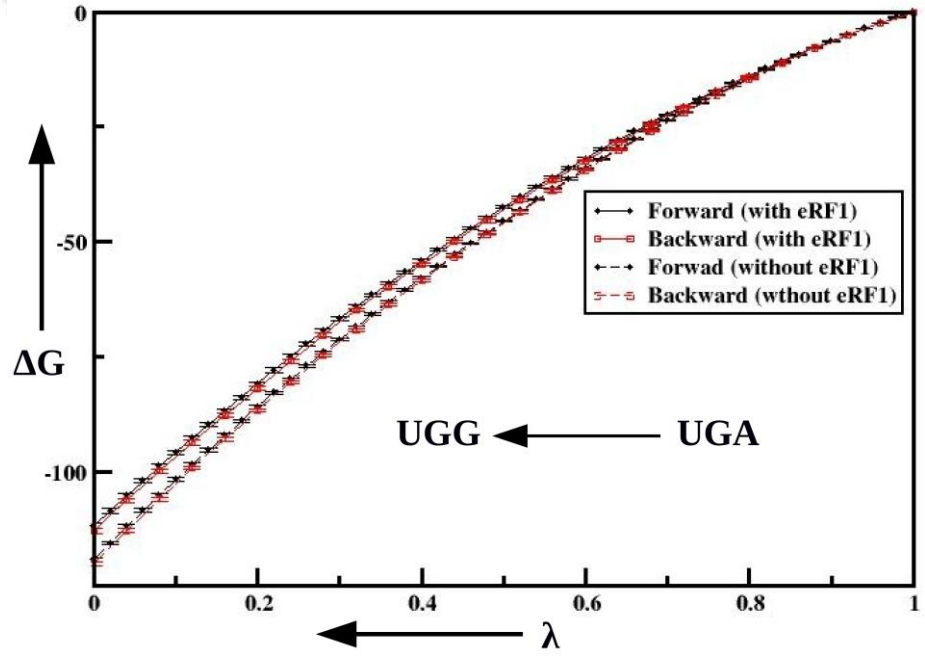


Figure 3A-1: Free energy vs λ (coupling coordinate) for the alchemical transformation of U/A → C/G in the stop codon programmed ribosome with/without eRF1. Data for forward (black) and backward (red) runs are shown. Uncertainties are shown as vertical bars (often too small to be seen). Total number of λ points = 51. Y axis is plotted in different scales in various subparts.

Appendix 4A

Table 4A-1: Calculated free energy change for alchemical transformation (a-f).

(a) Free energy change for the alchemical transformation 5' ppp-dsRNA:RIG-I→5' pp-dsRNA:RIG-I.

System	Run direction	Simulation length elec vdw	ΔG^{elec}	ΔG^{vdw}	$\Delta G^{elec}+\Delta G^{vdw}$
Complex	forward	11 + 13 ns	435.1 (2.0)	6.3 (0.6)	441.4 (2.1)
Complex	forward	22 + 26 ns	435.0 (1.5)	6.7 (0.7)	441.7 (1.7)
Complex	forward	22 + 26 ns	435.0 (1.7)	6.5 (0.9)	441.5 (1.9)
Complex	forward	22 + 26 ns	434.2 (1.4)	7.3 (0.4)	441.5 (1.7)
Complex	forward	22 + 26 ns	431.0 (2.1)	8.7 (0.5)	439.7 (2.2)
Complex	forward	22 + 26 ns	433.0 (1.4)	7.3 (0.5)	440.3 (1.5)

(b) Free energy change for the alchemical transformation 5' ppp-dsRNA→5' pp-dsRNA.

System	Run direction	Simulation length elec vdw	ΔG^{elec}	ΔG^{vdw}	$\Delta G^{elec}+\Delta G^{vdw}$
free dsRNA	forward	11 + 13 ns	431.2 (1.5)	6.5 (1.2)	437.7 (1.9)
free dsRNA	forward	22 + 26 ns	431.2 (0.6)	5.3 (1.8)	436.5 (1.9)
free dsRNA	forward	22 + 26 ns	431.1 (0.9)	5.7 (1.2)	436.8 (1.5)
free dsRNA	forward	22 + 26 ns	431.2 (1.2)	5.2 (1.8)	436.4 (2.2)
free dsRNA	forward	22 + 26 ns	430.7 (0.9)	5.5 (0.6)	436.2 (1.1)
free dsRNA	forward	22 + 26 ns	430.6 (0.5)	5.7 (1.1)	436.3 (1.2)

(c) Free energy change for the alchemical transformation 5' ppp-dsRNA:Mg^{2+}→5' pp-dsRNA:Mg^{2+}.

System	Run direction	Simulation length elec vdw	ΔG^{elec}	ΔG^{vdw}	$\Delta G^{elec}+\Delta G^{vdw}$
free dsRNA with MG	forward	11 + 13 ns	437.8 (0.9)	2.9 (1.5)	440.7 (1.7)
free dsRNA with MG	forward	22 + 26 ns	438.0 (1.6)	2.5 (1.4)	440.5 (2.1)
free dsRNA with MG	forward	22 + 26 ns	437.8 (0.7)	2.7 (1.3)	440.5 (1.5)
free dsRNA with MG	forward	22 + 26 ns	437.8 (0.8)	2.7 (1.2)	440.5 (1.4)
free dsRNA with MG	forward	22 + 26 ns	436.3 (0.7)	2.6 (1.7)	438.9 (1.8)
free dsRNA with MG	forward	22 + 26 ns	437.7 (0.8)	2.2 (1.2)	439.9 (1.8)

(d) Free energy change for the alchemical transformation 5' pp-dsRNA:RIG-I→5' p-dsRNA:RIG-I.

System	Run direction	Simulation length elec vdw	ΔG^{elec}	ΔG^{vdw}	$\Delta G^{elec}+\Delta G^{vdw}$
Complex	forward	11 + 13 ns	388.1 (2.6)	8.3 (1.2)	396.4 (2.9)
Complex	forward	22 + 26 ns	387.9 (1.8)	10.2 (1.3)	398.1 (2.2)
Complex	forward	22 + 26 ns	388.1 (1.8)	9.8 (1.1)	397.9 (2.1)
Complex	forward	22 + 26 ns	389.8 (2.0)	8.5 (1.0)	398.3 (2.2)
Complex	forward	22 + 26 ns	386.8 (1.1)	9.9 (0.7)	396.7 (1.3)
Complex	forward	22 + 26 ns	386.9 (0.9)	9.9 (0.9)	396.8 (1.3)

(e) Free energy change for the alchemical transformation 5' pp-dsRNA→5' p-dsRNA.

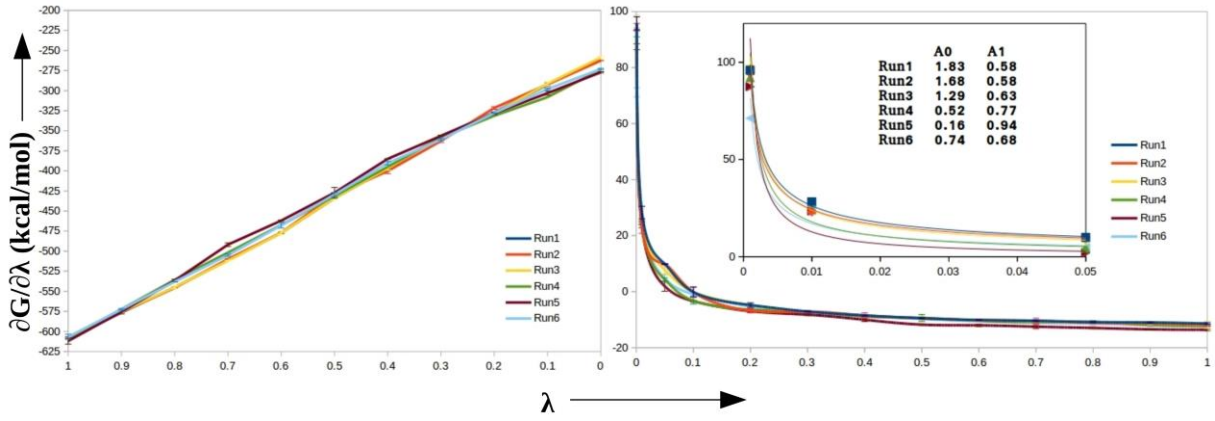
System	Run direction	Simulation length elec vdw	ΔG^{elec}	ΔG^{vdw}	$\Delta G^{elec}+\Delta G^{vdw}$
free dsRNA	forward	11 + 13 ns	380.0 (1.0)	7.6 (1.8)	387.6 (2.0)
free dsRNA	forward	22 + 26 ns	379.9 (0.9)	7.7 (0.2)	387.6 (0.9)
free dsRNA	forward	22 + 26 ns	379.9 (1.0)	7.7 (1.0)	387.6 (1.4)
free dsRNA	forward	22 + 26 ns	379.7 (0.6)	7.7 (0.5)	387.4 (0.8)
free dsRNA	forward	22 + 26 ns	380.2 (0.9)	7.6 (0.4)	387.8 (1.0)
free dsRNA	forward	22 + 26 ns	380.0 (0.8)	8.2 (0.5)	388.2 (0.9)

(f) Free energy change for the alchemical transformation 5' pp-dsRNA:Mg^{2+}→5' p-dsRNA:Mg^{2+}.

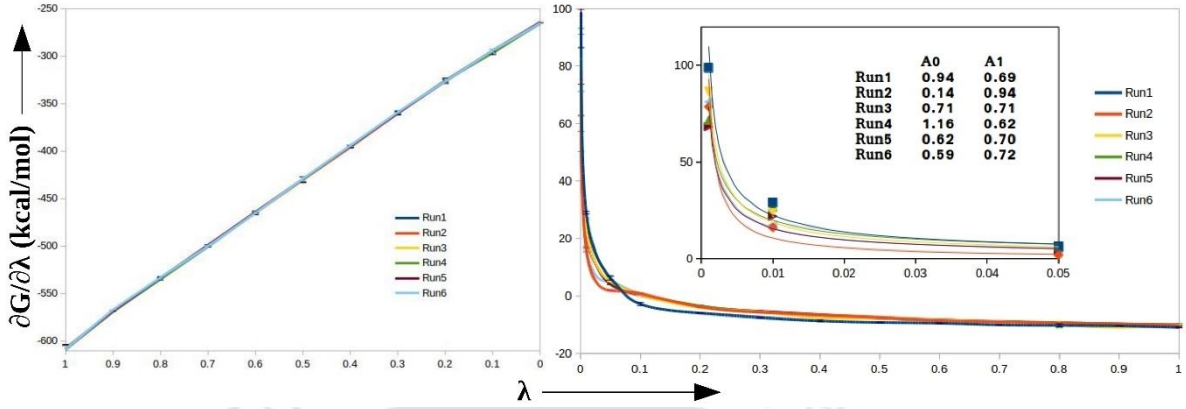
System	Run direction	Simulation length elec vdw	ΔG^{elec}	ΔG^{vdw}	$\Delta G^{elec}+\Delta G^{vdw}$
free dsRNA with MG	forward	11 + 13 ns	389.3 (2.6)	8.2 (1.2)	397.5 (2.9)
free dsRNA with MG	forward	22 + 26 ns	390.0 (1.0)	8.2 (1.0)	398.2 (1.4)
free dsRNA with MG	forward	22 + 26 ns	389.9 (0.4)	8.7 (0.6)	398.6 (0.7)
free dsRNA with MG	forward	22 + 26 ns	389.7 (0.7)	8.9 (0.9)	398.6 (1.1)
free dsRNA with MG	forward	22 + 26 ns	386.6 (0.4)	9.4 (1.0)	396.0 (1.0)
free dsRNA with MG	forward	22 + 26 ns	387.3 (0.9)	9.6 (0.9)	396.9 (1.3)

Free energies in kcal/mol and apparent statistical uncertainties in parentheses. Simulation lengths correspond to the sum of all λ points for the electrostatic and van der Waals steps, respectively.

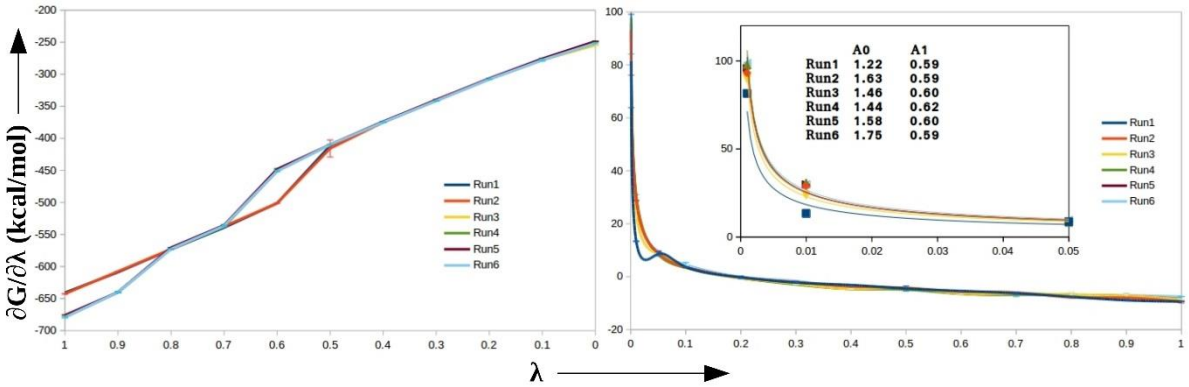
(a) Free energy derivative ∂G/∂λ vs λ plot for alchemical transformation 5' ppp-dsRNA:RIG-I→ 5' pp-dsRNA:RIG-I. The electrostatic part is on the left and van der waals part is on the right.



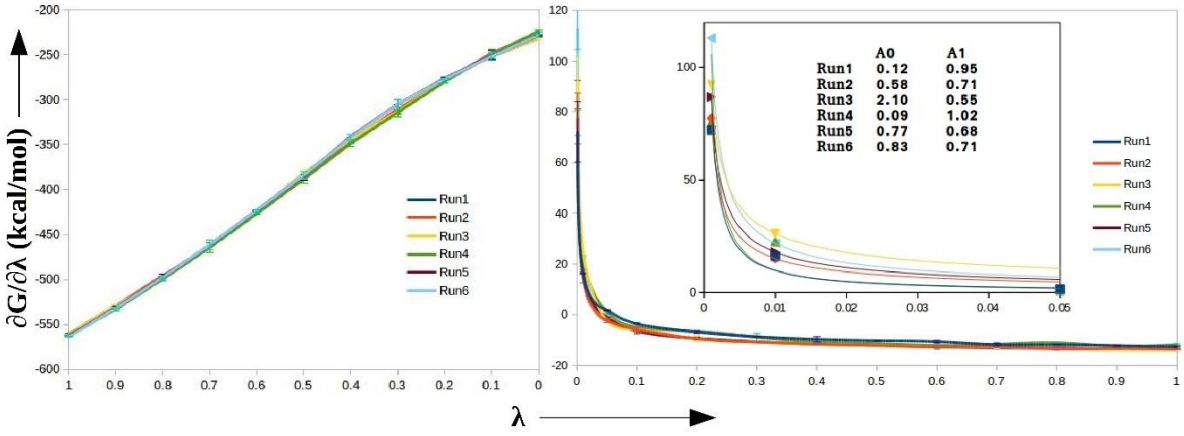
(b) Free energy derivative ∂G/∂λ vs λ plot for alchemical transformation 5' ppp-dsRNA→ 5' pp-dsRNA. The electrostatic part is on the left and van der waals part is on the right.



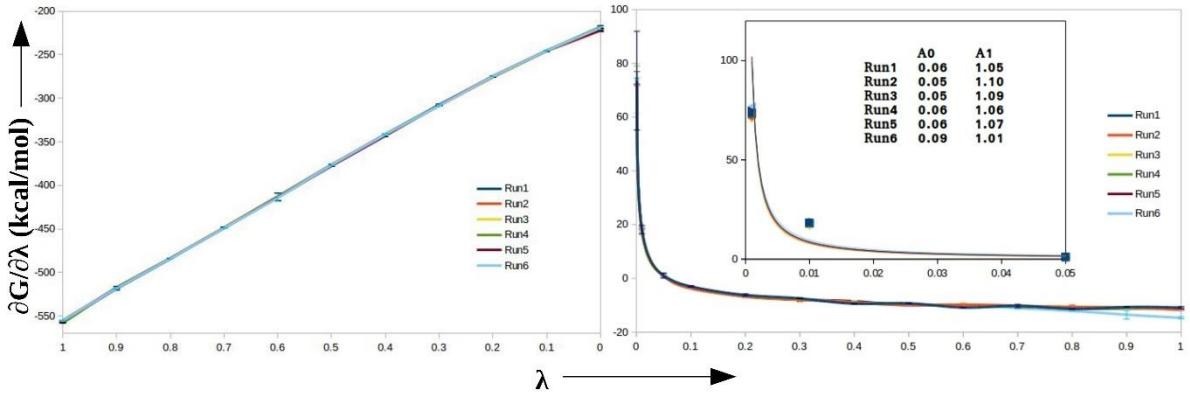
(c) Free energy derivative ∂G/∂λ vs λ plot for alchemical transformation 5' ppp-dsRNA:Mg^{2+}→ 5' pp-dsRNA:Mg^{2+}. The electrostatic part is on the left and van der waals part is on the right.



(d) Free energy derivative ∂G/∂λ vs λ plot for alchemical transformation 5' pp-dsRNA:RIG-I→ 5' p-dsRNA:RIG-I. The electrostatic part is on the left and van der waals part is on the right.



(e) Free energy derivative ∂G/∂λ vs λ plot for alchemical transformation 5' pp-dsRNA→ 5' p-dsRNA. The electrostatic part is on the left and van der waals part is on the right.



(f) Free energy derivative ∂G/∂λ vs λ plot for alchemical transformation 5' pp-dsRNA:Mg^{2+}→ 5' p-dsRNA:Mg^{2+}. The electrostatic part is on the left and van der waals part is on the right.

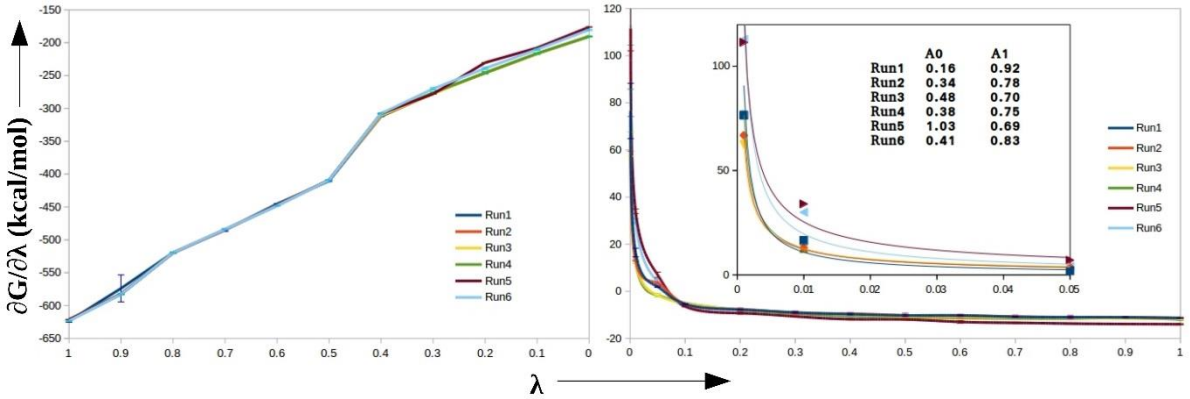


Figure 4A-1: (a-f) Free energy derivative ∂G/∂λ for multiple runs as a function of λ (coupling factor). Derivatives are plotted separately for Electrostatic (left) and van der Waals (right) steps. The inset (right) enlarges the region near λ=0 (for vdw derivatives); the smooth curves are fits to the numerical derivatives with the function $A_0\lambda^{-A1}$, where A_0 and A_1 are adjustable parameters. Uncertainties shown as vertical bars and often too small to be seen.

Table 4A-2: Estimated free energy (kcal/mol).

(**a**) Estimated free energy change along the alchemical transformation 5' p-dsRNA→5' OH-dsRNA in the **active complex without Hel2 loop (Model: A$^{\#}$)**.

System	Run direction	Simulation length elec vdw	ΔG^{elec}	ΔG^{vdw}	$\Delta G^{elec}+\Delta G^{vdw}$
Complex	forward	11 + 13 ns	339.5 ± 0.8	5.6 ± 0.1	345.1 ± 0.8
Complex	forward	22 + 26 ns	337.6 ± 1.1	6.4 ± 0.1	344.0 ± 1.1
Complex	forward	22 + 26 ns	337.6 ± 0.9	6.7 ± 0.3	344.3 ± 1.0
Complex	forward	22 + 26 ns	338.1 ± 0.6	6.7 ± 0.1	344.8 ± 0.6
Complex	forward	22 + 26 ns	339.3 ± 0.5	6.5 ± 0.2	345.8 ± 0.5
		Average	**338.4 ± 0.4**	**6.4 ± 0.2**	**344.8 ± 0.4**

(**b**) Estimated free energy change along the alchemical transformation 5' p-dsRNA→5' OH-dsRNA in **unbound/free** state. See **Figure S7** (Lower horizontal arm).

System	Run direction	Simulation length elec vdw	ΔG^{elec}	ΔG^{vdw}	$\Delta G^{elec}+\Delta G^{vdw}$
free dsRNA	forward	11 + 13 ns	338.1 ± 0.6	4.5 ± 0.1	342.6 ± 0.6
free dsRNA	forward	22 + 26 ns	337.9 ± 0.4	4.5 ± 0.1	342.4 ± 0.4
free dsRNA	forward	22 + 26 ns	337.9 ± 0.7	4.4 ± 0.1	342.3 ± 0.7
free dsRNA	forward	22 + 26 ns	338.0 ± 0.6	4.5 ± 0.2	342.5 ± 0.6
free dsRNA	forward	22 + 26 ns	338.2 ± 0.9	4.4 ± 0.1	342.6 ± 0.9
free dsRNA	forward	22 + 26 ns	336.5 ± 0.6	4.5 ± 0.1	341.0 ± 0.6
free dsRNA	forward	22 + 26 ns	337.0 ± 0.4	4.0 ± 0.2	341.0 ± 0.4
		Average	**337.7 ± 0.2**	**4.4 ± 0.1**	**342.1 ± 0.2**

(**c**) Estimated free energy change along the alchemical transformation 5' p-dsRNA→5' OH-dsRNA in the **active complex with build in Hel2 loop (Model: A)**.

System	Run direction	Simulation length elec vdw	ΔG^{elec}	ΔG^{vdw}	$\Delta G^{elec}+\Delta G^{vdw}$
Complex	forward	22 + 26 ns	338.1 ± 0.6	6.0 ± 0.1	344.1 ± 0.6
Complex	forward	22 + 26 ns	336.9 ± 1.3	7.0 ± 0.2	343.9 ± 1.3
		Average	**337.5 ± 0.6**	**6.5 ± 0.5**	**344.0 ± 0.8**

(**d**) Estimated free energy change along the alchemical transformation 5' OH-dsRNA →5' p-dsRNA in the **inactive complex (NA)**.

System	Run direction	Simulation length elec vdw	ΔG^{elec}	ΔG^{vdw}	$\Delta G^{elec}+\Delta G^{vdw}$
Complex	forward	22 + 13 ns	-338.1 ± 0.2	-5.9 ± 0.1	-344.0 ± 0.2
Complex	forward	22 + 26 ns	-339.9 ± 0.5	-5.7 ± 0.2	-345.6 ± 0.5
Complex	forward	22 + 26 ns	-338.1 ± 0.3	-6.4 ± 0.1	-344.5 ± 0.3
		Average	**-338.7 ± 0.6**	**-6.0 ± 0.2**	**-344.7 ± 0.6**

(**e**) Estimated free energy change along the alchemical transformation 5' OH-dsRNA→5' p-dsRNA in the **inactive complex without Hel2 loop (NA$^{\#}$)**.

System	Run direction	Simulation length elec vdw	ΔG^{elec}	ΔG^{vdw}	$\Delta G^{elec}+\Delta G^{vdw}$
Complex	forward	22 + 26 ns	-338.3 ± 0.3	-6.4 ± 0.1	-344.7 ± 0.3
Complex	forward	22 + 26 ns	-337.7 ± 0.5	-6.6 ± 0.1	-344.3 ± 0.5
		Average	**-338.0 ± 0.3**	**-6.5 ± 0.1**	**-344.5 ± 0.3**

(**f**) Estimated free energy change along the alchemical transformation Mg^{2+}:5' p-dsRNA→Mg^{2+}:5' OH-dsRNA in **unbound/free** form. See Figure S7 (upper horizontal arm).

System	Run direction	Simulation length elec vdw	ΔG^{elec}	ΔG^{vdw}	$\Delta G^{elec}+\Delta G^{vdw}$
free dsRNA with MG	forward	11 + 13 ns	386.0 ± 1.2	7.3 ± 0.4	393.3 ± 1.3
free dsRNA with MG	forward	22 + 26 ns	384.4 ± 1.1	6.8 ± 0.7	391.2 ± 1.3
free dsRNA with MG	forward	22 + 26 ns	385.3 ± 1.1	7.0 ± 0.9	392.3 ± 1.4
free dsRNA with MG	forward	22 + 26 ns	380.3 ± 1.0	7.3 ± 0.7	387.3 ± 1.2
free dsRNA with MG	forward	22 + 26 ns	388.5 ± 1.0	7.0 ± 0.8	395.5 ± 1.3
free dsRNA with MG	forward	22 + 26 ns	388.0 ± 1.2	7.0 ± 0.3	395.0 ± 1.2
		Average	**385.4 ± 1.2**	**7.1 ± 0.1**	**392.5 ± 1.2**

(g) Calculated RIG-I binding free energy of 5' OH-dsRNA (in kcal/mol) relative to the 5' p-dsRNA. Active (A, A#), inactive complex (NA, NA#) complexes.

	$\Delta G^{Complex}$	ΔG^{free}	$\Delta\Delta G$
A	344.0 ± 0.8	345.9 ± 0.2	-1.9 ± 0.8
A#	344.8 ± 0.4		-1.1 ± 0.4
NA	344.7 ± 0.6		-1.2 ± 0.6
NA#	344.5 ± 0.3		-1.4 ± 0.3
Experiment (Ren et al., 2019)	-	-	-0.73

Statistical uncertainty in the parentheses. The MD trajectories at each lambda window were divided into two equal halves and the difference between the computed ΔG's from the two-halves is reported as statistical uncertainty after ±, of ΔG^{elec}, ΔG^{vdw} for each replica. Uncertainty in ΔG^{comp}, ΔG^{free} for each replica is calculated by propagating the errors from ΔG^{elec} and ΔG^{vdw}. Uncertainty in the averaged result is reported as standard error (ΔG's from different replicas). The uncertainties for ΔΔG's were calculated by propagating the uncertainties of ΔG^{comp}, ΔG^{free}.

Appendix 5

Derivation for kinetic parameter (k_{cat} and K_M) calculation

$$RS + tRNA \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} [RS:tRNA]^{NR} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} [RS:tRNA]^{R} \xrightarrow{k_3}$$

$$\text{Rate} = k_3[RS:tRNA]^R \qquad \text{--- (1)}$$

Steady state approximation $\frac{d}{dt}[RS:tRNA]^R = 0$

$$\Rightarrow k_2[RS:tRNA]^{NR} - k_{-2}[RS:tRNA]^R - k_3[RS:tRNA]^R = 0$$

$$[RS:tRNA]^R = \frac{k_2\,[RS:tRNA]^{NR}}{(k_{-2} + k_3)} \qquad \text{-- (2)}$$

Steady state approximation $\frac{d}{dt}[RS:tRNA]^{NR} = 0$

$$\Rightarrow k_1[RS][tRNA] - k_{-1}[RS:tRNA]^{NR} - k_2[RS:tRNA]^{NR} + k_{-2}[RS:tRNA]^R = 0$$

From equation 2

$$\Rightarrow k_1[RS][tRNA] = (k_{-1} + k_2)[RS{:}\,tRNA]^{NR} - \frac{k_2\, k_{-2}[RS{:}\,tRNA]^{NR}}{(k_{-2} + k_3)}$$

--- (3)

Total tRNA concentration

$$[tRNA]_{total} = [tRNA]_{free} + [RS{:}\,tRNA]^{NR} + [RS{:}\,tRNA]^{R}$$

$$[tRNA]_{free} = [tRNA]_{total} - [RS{:}\,tRNA]^{NR} - [RS{:}\,tRNA]^{R} \quad \text{--- (4)}$$

From equation 3 and 4

$$\Rightarrow k_1[RS]\{[tRNA]_{total} - [RS{:}\,tRNA]^{NR} - [RS{:}\,tRNA]^{R}\} = (k_{-1} + k_2)[RS{:}\,tRNA]^{NR} - \frac{k_2\, k_{-2}[RS{:}\,tRNA]^{NR}}{(k_{-2} + k_3)}$$

$$\Rightarrow k_1[RS][tRNA]_{total} = (k_{-1} + k_2)[RS{:}\,tRNA]^{NR} + k_1[RS][RS{:}\,tRNA]^{NR} + k_1[RS][RS{:}\,tRNA]^{R} - \frac{k_2\, k_{-2}[RS{:}\,tRNA]^{NR}}{(k_{-2} + k_3)}$$

From equation 2 substituting $[RS{:}tRNA]^{R}$

$$\Rightarrow k_1[RS][tRNA]_{total} = (k_{-1} + k_2)[RS{:}\,tRNA]^{NR} + k_1[RS][RS{:}\,tRNA]^{NR} + k_1[RS]\frac{k_2[RS{:}\,tRNA]^{NR}}{(k_{-2} + k_3)} - \frac{k_2\, k_{-2}[RS{:}\,tRNA]^{NR}}{(k_{-2} + k_3)}$$

$$\Rightarrow k_1[RS][tRNA]_{total} = [RS{:}\,tRNA]^{NR}\{(k_{-1} + k_2)\cdot + k_1[RS] + k_1[RS]\frac{k_2}{(k_{-2} + k_3)} - \frac{k_2 k_{-2}}{(k_{-2} + k_3)}\}$$

$$\Rightarrow k_1[RS][tRNA]_{total} = [RS{:}\,tRNA]^{NR}\{\frac{(k_{-1} + k_2)(k_{-2} + k_3) - k_2 k_{-2}}{k_{-2} + k_3} + k_1[RS](1 + \frac{k_2}{(k_{-2} + k_3)})\}$$

$$\Rightarrow k_1[RS][tRNA]_{total} = [RS:tRNA]^{NR}\{\frac{k_{-1}k_{-2} + k_{-1}k_3 + k_{-2}\,k_2 + \,k_2k_3 - k_{-2}k_2}{k_{-2} + k_3}$$

$$+k_1[RS](1 + \frac{k_2}{(k_{-2} + \,k_3)})\}$$

$$\Rightarrow k_1[RS][tRNA]_{total} = [RS:tRNA]^{NR}\{\frac{k_{-1}k_{-2} + k_{-1}k_3 + \,k_2k_3}{k_{-2} + k_3}$$

$$+k_1[RS](1 + \frac{k_2}{(k_{-2} + \,k_3)})\}$$

$$\Rightarrow k_1[RS][tRNA]_{total}$$

$$= [RS:tRNA]^{NR}\{\frac{k_{-1}k_{-2} + k_{-1}k_3 + \,k_2k_3 + \,k_1(k_{-2} + k_3)[RS] + k_1k_2[RS]}{k_{-2} + k_3}$$

$$\Rightarrow [RS:tRNA]^{NR} = \frac{k_1[RS][tRNA]_{total}(k_{-2} + k_3)}{k_{-1}k_{-2} + k_{-1}k_3 + \,k_2k_3 + \{k_1(k_{-2} + k_3) + k_1k_2\}[RS]}$$

--- (5)

From equation 1 and 2

$$Rate = k_3[RS:tRNA]^R = \frac{k_2\,k_3[RS:tRNA]^{NR}}{(k_{-2} + \,k_3)}$$

Substituting equation 5 into the above equation

$$Rate = \frac{k_2\,k_3}{(k_{-2} + \,k_3)}\ \frac{k_1[RS][tRNA]_{total}(k_{-2} + k_3)}{k_{-1}k_{-2} + k_{-1}k_3 + \,k_2k_3 + \{k_1(k_{-2} + k_3) + k_1k_2\}[RS]}$$

$$Rate = \frac{k_1k_2\,k_3[RS][tRNA]_{total}}{(k_{-1}k_{-2} + k_{-1}k_3 + \,k_2k_3) + (k_1k_{-2} + k_1k_3 + k_1k_2)[RS]}$$

$$Rate = \frac{k_1k_2\,k_3[RS][tRNA]_{total}}{(k_1k_{-2} + k_1k_3 + k_1k_2)[\frac{k_{-1}k_{-2} + k_{-1}k_3 + \,k_2k_3}{k_1k_{-2} + k_1k_3 + k_1k_2} + [RS]]}$$

$$Rate = \frac{k_2\, k_3[RS][tRNA]_{total}}{(k_2 + k_{-2} + k_3)[\frac{k_{-1}k_{-2} + k_{-1}k_3 + k_2k_3}{k_1(k_2 + k_{-2} + k_3)} + [RS]]}$$

$$Rate = \frac{k_{cat}[RS][tRNA]_{total}}{K_M + [RS]}$$

Where,

$$k_{cat} = \frac{k_2\, k_3}{(k_2 + k_{-2} + k_3)}$$

$$K_M = \frac{k_{-1}k_{-2} + k_{-1}k_3 + k_2k_3}{k_1(k_2 + k_{-2} + k_3)}$$

References

Ahel, I.; Korencic, D.; Ibba, M.; & Soll, D. (2003). Trans-editing of mischarged tRNAs. Proc. Natl. Acad. Sci. USA 100, 15422-15427.

Alberty, R. A.; Goldberg, & R. N. (1992). Standard thermodynamic formation properties for the adenosine 5'-triphosphate series. Biochemistry 31, 10610-10615.

Alder, B. J., & Wainwright, T. E. (1957). Phase Transition for a hard sphere system. J. Chem. Phys. 27, 1208-1209.

Alkalaeva, E.Z.; Pisarev, A.V.; Frolova, L.Y.; Kisselev, L.L.; & Pestova, T.V. (2006). In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3. Cell 125, 1125-1136.

Ardito, F., Giuliani, M., Perrone, D., Troiano, G., & Lo Muzio, L. (2017). The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy. International journal of molecular medicine, 40(2), 271-280.

Armstrong, R. N. (1998). Mechanistic imperatives for the evolution of glutathione transferases. Curr. Opin. Chem. Biol. 2, 618-623.

Banavali, N. K., & MacKerell Jr, A. D. (2002). Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. Journal of molecular biology, 319(1), 141-160.

Behrmann, E.; Loerke, J.; Budkevich, T. V.; Yamamoto, K.; Schmidt, A.; Penczek, P. A.; Vos, M. R.; Bürger, J.; Mielke, T.; Scheerer, P.; & Spahn, C. M. (2015). Structural snapshots of actively translating human ribosomes. Cell 161, 845-857.

Berendsen, H. J. C.; Grigera, J. R.; & Straatsma, T. P. (1987). The missing term in effective pair potentials. J. Phys. Chem. 91, 6269-6271.

Bertram, G.; Bell, H.A.; Ritchie, D.W.; Fullerton, G.; & Stansfield, I. (2000). Terminating eukaryote translation: domain 1 of release factor eRF1 functions in stop codon recognition. RNA 6, 1236-1247.

Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; & Mackerell, A. D. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. J. Chem. Theory Comput. 8, 3257-3273.

Beuning, P. J.; Gulotta, M.; & Musier-Forsyth, K. (1997). Atomic group ‘‘mutagenesis’’ reveals major groove fine interactions of a tRNA synthetase with an RNA helix. J. Am. Chem. Soc. 119, 8397-8402.

Beuning, P. J.; & Musier-Forsyth, K. (1999). Transfer RNA recognition by aminoacyl-tRNA synthetases. Biopolymers 52, 1-28.

Beuning, P. J.; Nagan, M. C.; Cramer, C. J.; Musier-Forsyth, K.; Gelpi, J. J.; & Bashford, D. (2002). Efficient aminoacylation of the $tRNA^{Ala}$ acceptor stem: dependence on the 2:71 base pair. RNA 8, 659-670.

Bhushan, S.; Meyer, H.; Starosta, A. L.; Becker,T.; Mielke,T.; Berninghausen,O.; Sattler,M.; Wilson, D. N.; & Beckmann,R. (2010). Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. Mol. Cell 40, 138-146.
Biou, V.; Yaremchuk, A.; Tukalo, M.; & Cusack, S. (1994). The 2.9 A crystal structure of T. Thermophilus seryl-TRNA synthetase complexed with tRNA (Ser). Science 263, 1404-1410.

Bock, L.V.; Kolář, M.H.; & Grubmüller, H. (2018). Molecular simulations of the ribosome and associated translation factors. Curr. Opi. Str. Bio. 49, 27-35.

Bogusz, S.; Cheatham, T.; & Brooks, B. (1998). Removal of pressure and free energy artifacts in charged periodic systems via net charge corrections to the Ewald potential. J. Chem. Phys. 108, 7070-7084.

Braun, J. E.; Truffault, V.; Boland, A.; Huntzinger, E.; .Chang, C. T.; Haas, G.; Weichenrieder, O.; Coles, M.; & Izaurralde, E. (2012). A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5′ exonucleolytic degradation. Nat. Struct. Mol. Biol. 19, 1324-1331.
Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; & Caflisch, A. (2009). CHARMM: the biomolecular simulation program J. Comput. Chem. 30, 1545-1614.

Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; & Karplus, M. (1983). Charmm: a program for macro-molecular energy, minimization, and molecular dynamics calculations. J. Comput. Chem. 4, 187−217.

Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; & Karplus, M. (2009). CHARMM: the biomolecular simulation program. J. Comput. Chem. 30, 1545−1614.

Brown, A.; Shao, S.; Murray, J.; Hegde, R. S; & Ramakrishnan, V. (2015). Structural basis for stop codon recognition in eukaryotes. Nature 524, 493-496.

Burak, Y.; Ariel, G.; & Andelman, D. (2003). Onset of DNA aggregation in presence of monovalent and multivalent counterions. J. Biophys. 85, 2100-2110.

Carter, E.; Ciccotti, G.; Hynes, J. T.; & Kapral, R. (1989). Constrained reaction coordinate dynamics for the simulation of rare events. Chem. Phys. Lett. 156, 472-477.

Carter, R. W. (1989). Simple groups of Lie type (Vol. 22). John Wiley & Sons.

Cui, S.; Eisenacher, K.; Kirchhofer, A.; Brzozka, K.; Lammens, K.; Fujita, T.; Conzelmann, K. K.; Krug, A.; & Hopfner, K. P. (2008). The C-terminal regulatory domain is the RNA 5'-triphosphate sensor of RIG-I. Mol. Cell. 29, 169-179.

Chakraborty, K.; Mantha, S.; & Bandyopadhyay, S. (2013). Molecular dynamics simulation of a single-stranded DNA with heterogeneous distribution of nucleobases in aqueous medium. J. Chem. Phys. 139, 075103.

Chang, K. Y.; Varani, G.; Bhattacharya, S.; Choi, H.; & McClain, W. H. (1999). Correlation of deformatibility at a tRNA recognition site and aminoacylation specificity. Proc. Natl. Sci. USA 96, 11764-11769.

Chapeville, F.; Lipmann, F.; Von Ehrenstein, G.; Weisblum, B.; Ray, W. J. Jr.; & Benzer, S. (1962). On the role of soluble ribonucleic acid in coding for amino acids. Proc. Natl. Acad. Sci. USA 48, 1086-1092.

Chavatte, L.; Seit-Nebi, A.; Dubovaya, V.; & Favre, A. (2002). The invariant uridine od stop codons contacts the conserved NIKSR loop of human eRF1 in the ribosome. EMBO J. 21, 5302-5311.

Cheng, Z.; Saito, K.; Pisarev, A.V.; Wada, M.; Pisareva, V.P.; Pestova, T.V.; Gajda, M.; Round, A.; Kong, C.; Lim, M.; & Nakamura,Y. (2009). Structural insights into eRF3 and stop codon recognition by eRF1. Genes. Dev. 23, 1106-1118.

Cheng, G.; Zhong, J.; Chung, J.; & Chisari, F. V. (2007). Double-stranded DNA and double-stranded RNA induce a common antiviral signaling pathway in human cells. Proc. Natl. Acad. Sci. U.S.A. 104, 9035-9040.

Crick, F. (1970). Central dogma of molecular biology. Nature 227, 561-563.

Haugland, R. A.; & Cline, M. G. (1980). Post-transcriptional modifications of oat coleoptile ribonucleic acids. 5′-Terminal capping and methylation of internal nucleosides in poly(A)-rich RNA. Eur. J. Biochem. 104, 271-277.

Conard, S. E.; Buckley, J.; Dang, M.; Bedwell, G. J.; Carter, R. L.; Khass, M.; & Bedwell, D. M. (2012). Identification of eRF1 residues that play critical and complementary roles in stop codon recognition. RNA 18, 1210-1221.

Cui, S.; Eisenächer, K.; Kirchhofer, A.; Brzózka, K.; Lammens, A.; Lammens, K.; Fujita, T.; Conzelmann, K.; Krug, A.; & Hopfner, K. (2008). The C-terminal regulatory domain is the RNA 5′-triphosphate sensor of RIG-I. Mol. Cell 29, 169-179.

Darden, T.; York, D.; & Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089-10092.

Davis, W. G.; Bowzard, J. B.; Sharma, S. D.; Wiens, M. E.; Ranjan, P.; Gangappa, S.; Stuchlik, O.; Pohl, J.; Donis, R. O.; Katz, J. M.; Cameron, C. E.; Fujita, T.; & Sambhara, S. (2012). The 3′ untranslated regions of influenza genomic sequences are 5'PPP-independent ligands for RIG-I. PLoS ONE 7, p.e32661.

de Duve, C. (1988). Transfer RNAs: the second genetic code. Nature 333, 117-118.

Delagoutte, B.; Moras, D.; & Cavarelli, J. (2000). tRNA aminoacylation by arginyl-tRNA synthetase: induced conformations during substrates binding. EMBO J. 19, 5599-5610.

Depoux, A.; Philibert, A.; Rabier, S.; Philippe, H. J.; Fontanet, A.; & Flahault, A. (2018). A multi-faceted pandemic: a review of the state of knowledge on the Zika virus. Public Health Reviews 39 (10).

Devarkar, S. C.; Wang, C.; Miller, M. T.; Ramanathan, A.; Jiang, F.; Khan, A. G.; Patel, S. S.; & Marcotrigiano, J. (2016). Structural basis for m7G recognition and 2′-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I. Proc. Natl. Acad. Sci. USA 113, 596-601.

Dyer, D. M.; Murali, T. M.; & Sobral, B. W. (2008). The landscape of human proteins interacting with viruses and other pathogens. PLos Pathogen. 4, e32.

Ewald, P. (1921). Evaluation of optical and electrostatic lattice potentials. Ann. Phys. 64, 253-287.

Feller, S. E.; Zhang, Y.; Pastor, R. W.; & Brooks, B. R. (1995). Constant pressure molecular dynamics simulation: The Langevin piston method. J. Chem. Phys. 103, 4613-4621.

Feng, T.; Yamamoto, A.; Wilkins, S. E.; Sokolova, E.; Yates, L. A.; Münzel, M.; Singh, P.; Hopkinson, R. J.; Fischer, R.; Cockman, M. E.; & Shelley, J. (2014). Optimal translational termination requires C4 lysyl hydroxylation of eRF1. Mol. Cell 53, 645-654.

Fischer, N. M.; Polêto, M. D.; Steuer, J.; & van der Spoel, D. (2018). Influence of Na^{+} and Mg^{2+} ions on RNA structures studied with molecular dynamics simulations. Nucleic Acids Res. 46, 4872-4882.

Francklyn, C.; & Schimmel, P. (1989). Aminoacylation of RNA minihelices with alanine. Nature 337, 478-481.

Francklyn, C.; Shi, J. P.; & Schimmel, P. (1992). Overlapping nucleotide determinants for specific aminoacylation of RNA microhelices. Science 255, 1121-1125.

Freddolino, P.; Arkhipov, A.; Larson, S. B.; McPherson, A.; & Schulten, K. (2006). Molecular dynamics simulation of the Satellite Tobacco Mosaic Virus (STMV). Structure 14, 437-449.

Freistroffer, D. V.; Kwiatkowaski, M.; Buckingham, R. H.; & Ehernberg, M. (2000). The accuracy of codon recognition by polypeptide release factors. Proc. Natl. Acad. Sci. USA 97, 2046-2051.

Frischmeyer, P. A., & Dietz, H. C. (1999). Nonsense-mediated mRNA decay in health and disease. Hum. Mol. Genet. 8, 1893-1900.

Frolova, L.; Le Goff, X.; Zhouravleva, G.; Davydova, E.; Philippe, M.; & Kisselev, L. (1996). Eukaryotic polypeptide chain release factor eRF3 is an eRF1-and ribosome-dependent guanosine triphosphatase. RNA 2, 334-341.

Frolova, L.; Seit-Nebi, A; & Kisselev, L. (2002). Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1. RNA 8, 129-136.

Frolova, L. Y.; Tsivkovski, R. Y.; Sivolobova, G. F.; Oparina, N. Y.; Serpinsky, O. I.; Blinov, V. M.; Tatkov, S. I.; & Kisselev, L. L. (1999). Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-TRNA hydrolysis. RNA 5, 1014-1020.

Frolova, L. Y.; Merkulova, T. I.; & Kisselev, L. L. (2000). Translation termination in eukaryotes: polypeptide release factor eRF1 is composed of functionally and structurally distinct domains. RNA 6, 381-390.

Fuxreiter, M., Luo, N., Jedlovszky, P., Simon, I., & Osman, R. (2002). Role of base flipping in specific recognition of damaged DNA by repair enzymes. Journal of molecular biology, 323(5), 823-834.

Gebretadik, F. A.; Seifu, M. F.; and Gelaw, B. K. (2015). Review on Ebola Virus Disease: Its Outbreak and Current Status. Epidemiology (sunnyvale) 5 (4).

Ghosh, S.; Dixit, H.; & Chakrabarti, R. (2015) Ion assisted structural collapse of a single stranded DNA: A molecular dynamics approach. Chem. Phys. 459, 137-147.

Giegé, R.; Sissler, M.; & Florentz, C. (1998). Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Research 26, 5017-5035.

Goubau, D.; Schlee, M.; Deddouche, S.; Pruijssers, A. J.; Zillinger, T.; Goldeck, M.; Schuberth, C.; Van der Veen, A.G.; Fujimura, T.; Rehwinkel, J.; Iskarpatyoti, J. A.; Barchet, W.; Ludwig, J.; Dermody, T. S.; Hartmann, G.; & Reis e Sousa, C. (2014). Antiviral immunity via RIG-I-mediated recognition of RNA bearing 5′-diphosphates. Nature 514, 372-375.

Hawley, B. R.; Lu, W. T.; Wilczynska, A.; & Bushell, M. (2017). The emerging role of RNAs in DNA damage repair. Cell Death & Differentiation 24, 580-587.

Heitmann, P. (1968). A model for sulfhydryl groups in proteins. Hydrophobic interactions of the cysteine side chain in micelles. Eur. J. Bio. 3, 346-350.

Ho, K. L., McNae, I. W., Schmiedeberg, L., Klose, R. J., Bird, A. P., & Walkinshaw, M. D. (2008). MeCP2 binding to DNA depends upon hydration at methyl-CpG. Molecular cell, 29(4), 525-531.

Hockney, R. W. (1970). The potential calculation and some applications. Methods in Computational Physics 9, 136-211.

Hopfield, J.J.; Yamane, T; & Coutts, M. (1976) Direct experimental evidence for kinetic proofreading in amino acylation of $tRNA^{Ile}$. Proc. Natl. Acad. Sci. USA 73, 1164-1168.

Hornung, V.; Ellegast, J.; Kim, S.; Brzozka, K.; Jung, A.; Kato, H.; Poeck, H.; Akira S.; Conzelmann, K. K.; Schlee, M.; Endreas, S.; & Hartmann, G. (2006). 5'-Triphosphate RNA is the ligand for RIG-I. Science 314, 994-997.

Hou, Y. M.; & Schimmel, P. (1988). A simple structural feature is a major determinant of the identity of a transfer RNA. Nature 336, 140-145.

Hou, Y. M.; & Schimmel, P. (1989). Evidence that a major determinant for the identity of a transfer RNA is conserved in evolution. Biochemistry 28, 6800-6804.

Huang, J.; & MacKerell, A. D. (2013). CHARMM36: All-atom additive protein force field: validation based on comparison to NMR data. J. Comput. Chem. 34, 2135-2145.

Hummer, G.; Pratt, L.; & García, A. (1996). Free Energy of Ionic Hydration. J. Phys. Chem. 100, 1206-1215.

Humphrey, W.; Dalke, A.; & Schulten, K. (1996). VMD-Visual Molecular Dynamics. J. Molec. Graphics 14, 3-38.

Ibba, M.; & Soll, D. (2000). Aminoacyl-tRNA synthesis. Ann. Rev. Biochem. 69, 617-650.

Ishii, K. J.; Coban, C.; Kato, H.; Takahashi, K.; Torii, Y.; Takeshita, F.; Ludwig, H.; Sutter, G.; Suzuki, K.; Hemmi, H.; & Sato, S. (2006). A Toll-like receptor–independent antiviral response induced by double-stranded B-form DNA. Nature Immun. 7, 40-48.

Ito, K.; Uno, M.; & Nakamura, Y. (2000). A tripeptide 'anticodon' deciphers stop codons in messenger RNA. Nature 403, 680-684.

Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. Phys. Rev. Lett. 78, 2690.

Jenner, L.B.; Demeshkina, N.; Yusupova, G.; & Yusupov, M. (2010). Structural aspects of messenger RNA reading frame maintenance by the ribosome. Nat. Str. Mol. Bio. 17, 555-560.

Jiang, Q.; & Chen, Z. (2011). Structural insights into the activation of RIG-I, a nanosensor for viral RNAs. EMBO Reports 13, 7-8.

Jiang, F.; Ramanathan, A.; Miller, M. T.; Tang, G. Q.; Gale, M. Jr.; Patel, S. S.; & Marcotrigiano, J. (2011). Structural basis of RNA recognition and activation by innate immune receptor RIG-I. Nature 479, 423-427.

Jiao, D.; Golubkov, P. A.; Darden, T. A.; & Ren, P. (2008). Calculation of protein–ligand binding free energy by using a polarizable potential. Proc. Natl. Acad. Sci. USA 105, 6290-6295.

Jin, H.; Kelley, A.C.; Loakes, D.; & Ramakrishnan, V. (2010). Structure of the 70S ribosome bound to release factor 2 and a substrate analog provides insights into catalysis of peptide release Proc. Natl. Acad. Sci. USA 107, 8593-8598.

Jonas, S.; & Izaurralde, E. (2013). The role of disordered protein regions in the assembly of decapping complexes and RNP granules. Genes Dev. 27, 2628-2641.

Jorgensen, W. L. (1981). Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. J. Am. Chem. Soc. 103, 335-340.

Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; & Klein, M. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926-935.

Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; & Tirado- Rives, J. (1988). Efficient computation of absolute free energies of binding by computer simulations. Application to the methane dimer in water. J. Chem. Phys. 89, 3742-3746.

Kato, H.; Takahasi, K.; & Fujita, T. (2011). RIG-I-like receptors: cytoplasmic sensors for non-self RNA. Immunol. Rev. 243, 91-98.

Kato, H.; Takeuchi, O.; Mikamo-Satoh, E.; Hirai, R.; Kawai, T.; Matsushita, K.; Hiragi, A.; Dermody, T. S.; Fujita, T.; & Akira, S. (2008). Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5. J. Exp. Med. 205, 1601-1610.

Kato, H.; Takeuchi, O.; Sato, S.; Yoneyama, M.; Yamamoto, M.; Matsui, K.; Uematsu, S.; Jung, A.; Kawai, T.; Ishii, K. J.; Yamaguchi, O.; Otsu, K.; Tsujimura, T.; Koh, C. S.; Reis e Sousa, C.; Matsuura, Y.; Fujita, T.; & Akira, S. (2006). Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. Nature 441, 101-105.

Kawai, T.; & Akira, S. (2007). Antiviral signaling through pattern recognition receptors. J. Biochem. 141, 137-145.

Keeling, K. M.; Xue, X.; Gunn, G.; & Bedwell, D. M. (2014). Therapeutics based on stop codon readthrough Annu Rev Genomics Hum. Genet. 15, 371-394.

Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; & Sternberg, M. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols 10, 845-858.

kervestin, S.; Frolova, L.; Kisselev, L.; & Jean-Jean, O. (2001). Stop codon recognition in ciliates: Euplotes release factor does not respond to reassigned UGA codon. EMBO Rep. 2, 680-684.

King, G.; & Warshel, A. (1989). A surface constrained all-atom solvent model for effective simulations of polar solutions. J. Chem. Phys. 91, 3647-3661.

Kirkwood, J. G. (1935). Statistical Mechanics of Fluid Mixtures. J. Chem. Phys. 3, 300-313.

Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; & Shaw, D. E. (2011). How Fast-Folding Proteins Fold. Science 334, 517-520.

Kisselev, L. (2002). Polypeptide release factors in prokaryotes and eukaryotes: same function, different structure. Structure 10, 8-9.

Kisselev, L. L.; & Frolova, L. (1995). Termination of translation in eukaryotes. Biochem. Cell Bio. 73, 1079-1086.

Knight, R. D.; & Landwebeer, L. F. (2000). The early evolution of the genetic code. Cell 101, 569-572.

Kohlway, A.; Luo, D.; Rawling, D. C.; Ding, S. C.; & Pyle, A. M. (2013). Defining the functional determinants for RNA surveillance by RIG-I. EMBO Rep. 14, 772-779.

Kryuchkova, P.; Grishin, A.; Eliseev, B.; Karyagina, A.; Frolova, L.; & Alkalaeva, E. (2013). Two-step model of stop codon recognition by eukaryotic release factor eRF1 Nucleic Acids Res. 2013, 41, 4573-4586.

Kolakofsky, D.; Kowalinski, E.; & Cusack, S. (2012). A structure-based model of RIG-I activation. RNA 18, 2118-2127.

Kolosov, P.; Frolova, L.; Seit-Nebi, A.; Dubovaya, V.; Kononenko, A.; Oparina, N.; Justesen, J.; Efimov, A.; & Kisselev, L. (2005). Invariant amino acids essential for decoding function of polypeptide release factor eRF1. Nucleic Acids Res. 33, 6418-6425.

Korencic, D.; Ahel, I.; Suchelert, J.; Sacher, M.; Ruan, B.; Stathopoulos, C.; Blum, P.; Ibba, M.; & Soll, D. (2004). A freestanding proofreading domain is required for protein synthesis quality control in Archaea. Proc. Natl. Acad. Sci. USA 101, 10260-10265.

Korkmaz, G.; Lind, C.; Åqvist, J.; & Sanyal, S. (2014). Characterizing an engineered release factor capable of reading all three stop codons J. FASEB 28, 569-572.

Korostelev, A.A. (2011). Structural aspects of translation termination on the ribosome RNA 17, 1409-1421.

Korostelev, A.; Asahara, H.; Lancaster, L.; Laurberg, M.; Hirschi, A.; Zhu, J.; Trakhanov, S.; Scott, W. G.; & Noller, H. F. (2008). Crystal structure of a translation termination complex formed with release factor RF2 Proc. Natl. Acad. Sci. USA105, 19684-19689.

Kowalinski, E.; Lunardi, T.; McCarthy, A.; Louber, J.; Brunel, J.; Grigorov, B.; Gerlier, D.; & Cusack, S. (2011). Structural Basis for the Activation of Innate Immune Pattern-Recognition Receptor RIG-I by Viral RNA. Cell 147, 423-435.

Kumar, A.; Basu, D.; & Satpati, P. (2017). Structure-Based Energetics of Stop Codon Recognition by Eukaryotic Release Factor. J. Chem. Inf. Model. 57, 2321-2328.

Kumar, A.; & Satpati, P. (2018). Energetics of Preferential Binding of Retinoic Acid-Inducible Gene-I to Double-Stranded Viral RNAs with 5′ Tri-/Diphosphate over 5′ Monophosphate. ACS omega 3, 3786-3795.

Lama, D., & Verma, C. S. (2020). Deciphering the mechanistic effects of eIF4E phosphorylation on mRNA-cap recognition. Protein Science, 29(6), 1373-1386.

Lässig, C.; Lammens, K.; Lopez, J. L. G.; Michalski, S.; Fettscher, O.; & Hopfner, K.P. (2018). Unified mechanisms for self-RNA recognition by RIG-I Singleton-Merten syndrome variants. Elife 7, e38958.

Laurberg, M.; Asahara, H.; Korostelev, A.; Zhu, J.; Trakhanov, S.; & Noller, H.F. (2008). Structural basis for translation termination on the 70S ribosome. Nature 454, 852-857.

Lee, F. S.; & Warshel, A. (1992). A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations. J. Chem. Phys. 97, 3100-3107.

Lind, C.; Esguerra, M.; Jespers, W.; Satpati, P.; Gutierrez-de-Terán, H.; & Åqvist, J. (2019). Free energy calculations of RNA interactions. Methods 162-163, 85-95.

Ling, J.; So, B. R.; Yadavalli, S. S.; Roy, H.; Shoji, S.; Fredrick, K.; Musier-Forsyth, K.; & Ibba, M. (2009). Resampling and editing of mischarged tRNA prior to translation elongation. Mol. Cell. 33, 654-660.

Linehan, M. M.; Dickey, T. H.; Molinari, E. S.; Fitzerald, M. E.; Potapova, O.; Iwasaki, A.; & Pyle, A. M. (2018). A minimal RNA ligand for potent RIG-I activation in living mice. Sci. Adv., 4, e1701854.

Loo, Y. M.; & Gale, M., Jr. (2011). Immune signaling by RIG-I-like receptors. Immunity 34, 680-692.

Lu, C.; Xu, H.; Ranjith-Kumar, C. T.; Brooks, M. T.; Hou, T. Y.; Hu, F.; Herr, A. B.; Strong, R. K.; Kao, C. C.; & Li, P. (2010). The structural basis of 5′ triphosphate double-stranded RNA recognition by RIG-I C-terminal domain. Structure 18, 1032-1043.

Luan, B.; & Aksimentiev, A. (2008). DNA attraction in monovalent and divalent electrolytes. J. Am. Chem. Soc. 130, 15754-15755.

Luo, D.; Ding, S.; Vela, A.; Kohlway, A.; Lindenbach, B.; & Pyle, A. (2011). Structural Insights into RNA Recognition by RIG-I. Cell 147, 409-422.

Luo, D.; Kohlway, A.; Vela, A.; & Pyle, A. M. (2012). Visualizing the determinants of viral RNA recognition by innate immune sensor RIG-I. Structure 20, 1983-1988.

MacKerell, Jr. A.D.; & Banavali, N. (2000). All-atom empirical force field for nucleic acids: 2) Application to molecular dynamics simulations of DNA and RNA in solution. J. Comp. Chem. 21, 105-120.

MacKerell, A. D., Jr; Bashford, D.; Bellott, M. L. D. R.; Dunbrack, R. L., J; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; & Joseph-McCarthy, D. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B 102, 3586-3616.

Mackerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J.; Field, M. J. Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W.E.; Roux, B.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; & Karplus, M. (1998). An all-atom empirical potential for molecular modelling and dynamics study of proteins. J. Phys. Chem. B 102, 3586-3616.

MacKerell, A.; Wiorkiewicz-Kuczera, J.; & Karplus, M. (1995). An all-atom empirical energy function for the simulation of nucleic acids. J. Am. Chem. Soc. 117, 11946-11975.

Marelius, J.; Kolmodin, K.; Feierberg, I.; & Åqvist, J. (1998). Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems J. Mol. Graphics Modell. 16, 213-225.

Martyna, G. J.; Tobias, D. J.; & Klein, M. L. (1994). Constant pressure molecular dynamics algorithms. J. Chem. Phys. 101(5), 4177-4189.

Matheisl, S.; Berninghausen, O.; Becker, T.; & Beckmann, R. (2015). Structure of a human translation termination complex. Nucleic Acids Res. 43, 8615-8626.

McCammon, J. A.; Gelin, B. R.; & Karplus, M. (1977). Dynamics of folded proteins. Nature 267, 585-590.

Widom, B.J. (1963). Some Topics in the Theory of Liquids. J. Chem. Phys. 39, 2808.

McClain, W. H.; & Foss, K. (1988). Changing the acceptor identity of a transfer RNA by altering nucleotides in a “variable pocket”. Science 241, 1804-1807.

McClain, W. H.; & Foss, K. (1988). Changing the identity of a tRNA by introducing a G-U wobble pair near the 39 acceptor end. Science 240, 793-796.

Medzhitov, R. (2007). Recognition of microorganisms and activation of the immune response. Nature 449, 819-826.

Merrick W. C.; & Hershey, J. W. B. (1996). The pathway and mechanism of eukaryotic protein synthesis. Cold Spring Harbor Laboratory Press, pp. 31-69.

Mogensen, T. H. (2009). Pathogen recognition and inflammatory signaling in innate immune defenses. Am. Soc. Microbio. 22, 240-273.

Mort, M.; Ivanov, D.; Cooper, D.N.; & Chuzhanova, N.A. (2008). A meta- analysis of nonsense mutations causing human genetic disease. Hum. Mutat. 29, 1037-1047.

Muhs, M.; Hilal, T.; Mielke, T.; Skabkin, M. A.; Sanbonmatsu, K.Y.; Pestova, T. V.; & Spahn, C. M. (2015). Cryo-EM of ribosomal 80S complexes with termination factors reveals the translocated cricket paralysis virus IRES. Mol. Cell. 57, 422-432.

Musier-Forsyth, K.; & Schimmel, P. (1992). Functional contacts of a transferRNAsynthetase with 2'-hydroxyl groups in the RNA minor groove. Nature 357, 513-515.

Musier-Forsyth, K.; Usman, N.; Scaringe, S.; Doudna, J.; Freen, R.; & Schimmel, P. (1991). Specificity for aminoacylation of an RNA helix: an unpaired, exocyclic amino group in the minor groove. Science 253, 784-786.

Naganuma, M.; Sekine, S.; Chong, Y. E.; Guo, M.; Yang, X. L.; Gamper, H.; Hou, Y. M.; Schimmel, P.; & Yokoyama, S. (2014). The selective tRNA aminoacylation mechanism based on a single G•U pair. Nature 510, 507-511.

Nagarajan, V. K.; Jones, C. I.; Newbury, S. F.; & Green, P. J. (2013). XRN 5′→3′ exoribonucleases: structure, mechanisms and functions. Biochim. Biophys. Acta. 1829, 590-603.

Ogle, J. M.; Murphy, F. V.; Tarry, M. J.; & Ramakrishnan, V. (2002). Selection of tRNA by the ribosome requires a transition from an open to a closed form. Cell 111, 721-732.

Onufriev, A. V., & Izadi, S. (2018). Water models for biomolecular simulations. Wiley Interdisciplinary Reviews: Computational Molecular Science, 8(2), e1347.

Pérez, A.; Luque, F. J.; & Orozco, M. (2011). Frontiers in molecular dynamics simulations of DNA. Acc. Chem. Res. 45, 196-205.

Pancer, Z.; & Cooper, M. D. (2006). The evolution of adaptive immunity. Annu. Rev. Immunol. 24, 497-518.

Park, J. E.; Heo, I.; Tian, Y.; Simanshu, D. K.; Chang, H.; Jee, D.; Patel, D. J.; & Kim, V. N. (2011). Dicer recognizes the 5′ end of RNA for efficient and accurate processing. Nature, 475, 201-205.

Park, S. J.; Hou, Y. M.; & Schimmel, P. (1989). A single base pair affects binding and catalytic parameters in the molecular recognition of a transfer RNA. Biochemistry 28, 2740-2746.

Park, S. J.; & Schimmel, P. (1988). Evidence of an aminoacyl transfer RNA synthetase with a region important for the identity of its cognate transfer RNA. J. Biol. Chem. 263, 16527-16530.

Peach, S. E.; York, K.; & Hesselberth, J. R. (2015). Global analysis of RNA cleavage by 5′-hydroxyl RNA sequencing. Nucleic Acids Res. 43, e108.

Perorna, J. J.; & Gruic-Sovulj, I. (2014). Synthetic and editing mechanisms of aminoacyl-tRNA synthetases. Top. Curr. Chem. 344, 1-41.

Perona, J. J.; Rould, M. A.; & Steitz, T. A. (1993). Structural basis for transfer RNA aminoacylation by Escherichia coli glutaminyl-tRNA synthetase. Biochemistry 32, 8758-8771.

Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; & Ferrin, T. E. (2004). UCSF: Chimera-a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605-1612.

Tembe, B.; & McCammon, J. A. (1984). Ligand-receptor interactions. Comput. Chem. 8, 281-283.

Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kale, L.; & Schulten, K. (2005). Scalable molecular dynamics with NAMD. J. Comput. Chem. 26, 1781-1802.

Pichlmair, A.; Schulz, O.; Tan, C. P.; Naslund, T. I.; Lijestrim, P.; Weber, F.; & Reis e SousA, C. (2006). RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. Science 314, 997-1001.

Pillay, S.; Li, Y.; Wong, L. E.; & Pervushin, K. (2016). Structural characterization of eRF1 mutants indicate a complex mechanism of stop codon recognition. Scientific Rep. 6, 18644.

Pisarev, A. V.; Skabkin, M. A.; Pisareva, V. P.; Skabkina, O. V.; Rakotondrafara, A. M.; Hentze, M. W.; Hellen, C. U.; & Pestova, T. V. (2010). The role of ABCE1 in eukaryotic posttermination ribosomal recycling. Mol. Cell 37, 196-210.

Ponder, J.; & Case, D. A. (2003). Force fields for protein simulations. Adv. Protein Chem. 66, 27-85.

Preis, A.; Heuer, A.; Barrio-Garcia, C.; Hauser, A.; Eyler, D. E.; Berninghausen, O.; Green, R.; Becker, T.; & Beckmann, R. (2014). Cryoelectron microscopic structures of eukaryotic translation termination complexes containing eRF1-eRF3 or eRF1-ABCE1. Cell Rep. 8, 59-65.

Pütz, J.; Puglisi, J. D.; Florentz, C.; & Giegé, R. (1993).Additive, cooperative and anticooperative effects between identity nucleotides of a tRNA. EMBO J. 12, 2949-2957.

Ramos, H.; & Gale, M. (2011). RIG-I like receptors and their signaling crosstalk in the regulation of antiviral immunity. Curr. Opin. Virol. 1, 167-176.

Rawling, D.; & Pyle, A. (2014). Parts, assembly and operation of the RIG-I family of motors. Curr. Opin. Struct. Biol. 25, 25-33.

Ren, X.; Linehan, M. M.; Iwasaki, A.; & Pyle, A. M. (2019). RIG-I Selectively Discriminates against 5′-Monophosphate. Cell Rep. 26, 2019-2027.

Rheinberger, H. J. (1991). The function of the translating ribosome: allosteric three-site model of elongation. Biochomie 73, 1067-1088.

Rodnina, M.V. (2013). The ribosome as a versatile catalyst: reactions at the peptidyl transferase center. Curr. Opin. Struct. Bio. 23, 595-602.

Rogers, M. J.; Adachi, T.; Inokuchi, H.; & Söll, D. (1994). Functional communication in the recognition of tRNA by Escherichia coli glutaminyl-tRNA synthetase. Proc. Natl. Acad. Sci. USA 91, 291-295.

Ruff, M.; Krishnaswamy, S.; Boeglin, M.; Poterszman, A.; Mitschler, A.; Podjarny, A.; Rees, B.; Thierry, J. C.; & Moras, D. (1991). Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-TRNA synthetase complexed with tRNA (Asp). Science 252, 1682-1689.

Ryckaert, J. P.; Ciccotti, G.; & Berendsen, H. J. C. (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular Dynamics of n-Alkanes. J. Comput. Phys. 23, 327-341

Saito, T.; Hirai, R.; Loo, Y. M.; Owen, D.; Johnson, C. L.; Sinha S. C.; Akira, S.; Fujita, T.; & Gale, M., Jr. (2007). Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2. Proc. Natl. Acad. Sci. USA 104, 582-587.

Saito, T.; Owen, D.; Jiang, F.; Marcotrigiano, J.; & Gale Jr., M. (2008). Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. Nature 454, 523-527.

Sarkar, S.; Maity, A.; Phukon, A. S.; Ghosh, S.; & Chakrabarti, R. (2019). Salt Induced Structural Collapse, Swelling, and Signature of Aggregation of Two ssDNA Strands: Insights from Molecular Dynamics Simulation. J. Phys. Chem. B 123, 47-56.

Satpati, P; & Åqvist, J. (2014) Why base tautomerization does not cause errors in mRNA decoding on the ribosome. Nucleic Acids Res 42, 12876-12884.

Satpati, P.; Bauer, P.; & Åqvist, J. (2014). Energetic tuning by tRNA modifications ensures correct decoding of isoleucine and methionine on the ribosome. Chem. Eur. J. 20, 10271-10275.

Satpati, P.; Clavaguera, C.; Ohanessian, G.; & Simonson, T. (2011). Free Energy Simulations of a GTPase: GTP and GDP Binding to Archaeal Initiation Factor 2. J. Phys. Chem. B 115, 6749-6763.

Satpati, P.; & Simonson, T. Conformational selection by the aIF2 GTPase: a molecular dynamics study of functional pathways. Biochemistry 2012, 51, 353-361.

Satpati, P.; & Simonson, T. (2012). Conformational selection through electrostatics: free energy simulations of GTP and GDP binding to archaeal Initiation Factor 2. Proteins 80, 1264-1282.

Satpati, P.; Sund, J.; & Åqvist, J. (2014). Structure-based energetics of mRNA decoding on the ribosome. Biochemistry 53, 1714-1722.

Schimmel, P.; Giege, R.; Moras, D., & Yokoyama, S. (1993). An operational RNA code for amino acids and possible relationship to genetic code. Proc. Natl. Acad. Sci. USA 90, 8763-8768.

Schimmel, P.; & Ribas De Pouplana, L. (2000). Footprints of aminoacyl-tRNA synthetases are everywhere. Trends Biochem. Sci. 25, 207-209.

Schlee, M. (2015). A Conserved Histidine in the RNA Sensor RIG-I Controls Immune Tolerance to N1-2'O-Methylated Self RNA. Immunity 43, 41-51.

Schlee, M.; Roth, A.; Hornung, V.; Hagmann, C. A.; Wimmenauer, V.; Barchet, W.; Coch, C.; Janke, M.; Mihailovic, A.; Wardle, G.; Juranek, S.; Kato, H.; Kawai, T.; Poeck, H.; Fitzgerald, K. A.; Takeuchi, O.; Akira, S.; Tuschi, T.; Latz, E.; Ludwig, J.; & Hartmann, G. (2009). Recognition of 5’ triphosphate by RIG-I helicase requires short blunt double-stranded RNA as contained in panhandle of negative-strand virus. Immunity 31, 25-34.

Schmidt, A.; Schwerd, T.; Hamm, W.; Wellmuth, J. C.; Cui, S.; Wenzel, M.; Hoffmann, F. S.; Michallet, M. C.; Besch, R.; Hopfner, K. P.; Endres, S.; & Rothenfusser, S. (2009). 5’-triphosphate RNA requires base-paired structures to activate signaling via RIG-I. Proc. Natl. Acad. Sci. USA 106, 12067-12072.

Schuberth-wagner, C.; Ludwig, J.; Bruder, A. K.; Herzner, A. M.; Zillinger, M.; Goldeck, M.; Schmidt, T.; Schmid-Burgk, J. L.; Kerber, R.; Wolter, S.; Stumpel, J. P., Roth, A.; Bartok, E.; Drosten, C.; Coch, C.; Hornung, V.; Barchet, W.; Kummerer, B. M.; Harthmann, G., & Schlee, M. (2015). A Conserved Histidine in the RNA Sensor RIG-I Controls Immune Tolerance to N1-2'O-Methylated Self RNA. Immunity 43, 41-51.

Simonson, T.; & Satpati, P. (2012). Nucleotide recognition by the initiation factor aIF5B: free energy simulations of a neo-classical GTPase. Proteins 80, 2742-2757.

Schmied, W. H.; Elsässer, S. J.; Uttamapinant, C.; & Chin, J. W. (2014). Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1. J. Am. Chem. Soc. 136, 15577-15583.

Seit-Nebi, A.; Frolova, L.; Justesen, J.; & Kisselev, L. (2001). Class-1 translation termination factors: invariant GGQ minidomain is essential for release activity and ribosome binding but not for stop codon recognition. Nucleic Acid Res. 29, 3982-3987.

Seit- Nebi, A.; Frolova, L.; & Kisselev, L. (2002). Conversion of omnipotent translation termination factor eRF1 into ciliate- like UGA- only unipotent eRF1. EMBO Rep. 3, 881-886.

Shi, J. P.; Francklyn, C.; Hill, K.; & Schimmel, P. (1990). A nucleotide that enhances the charging of RNA minihelix sequence variants with alanine. Biochemistry 29, 3621-3626.

Shoemaker, C. J.; & Green, R. (2011). Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. Proc. Natl. Acad. Sci. USA 108, 1392-1398.

Steinbrecher, T.; & Labahn, A. (2010). Towards accurate free energy calculations in ligand protein-binding studies. Curr. Med. Chem. 17, 767-785.

Song, H.; Mugnier, P.; Das, A. K.; Webb, H. M.; Evans, D. R.; Tuite, M. F.; Hemmings, B. A.; & Barford, D. (2000). The crystal structure of human eukaryotic release factor eRF1—mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. Cell 100, 311-321.

Sponer, J.; Bussi, G.,; Krepl, M.; Banáš, P.; Bottaro, S.; Cunha, R.A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurečka, P.; & Walter, N.G. (2018). RNA structural dynamics as captured by molecular simulations: a comprehensive overview. Chem. Rev. 118, 4177-4338.

Sprinzl, M.; Horn, C.; Brown, M.; Ioudovitch, A.; & Steinberg, S. (1998). Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acid Res. 26, 148-153.

Storer, A. C.; & Cornish-Bowden, A. (1976). Concentration of Mg:ATP and other ions in solution. Biochem. J. 159, 1-5.

Sugiura, I.; Nureki, O.; Ugaji-Yoshikawa, Y.; Kuwabara, S.; Shimada, A.; Tateno, M.; Lorber, B.; Giege, R.; Moras, D.; Yokoyama, S.; & Konno, M. (1993). The 2.0 A crystal structure of Thermus thermophilus methionyl-tRNA synthetase reveals two RNA-binding modules. Structure 8, 197-208.

Sund, J.; Andér, M.; & Åqvist, J. (2010). Principles of stop-codon reading on the ribosome. Nature 465, 947-950.

Swope, W. C.; Hans, C.; Andersen, H. C.; Berens, P. H.; & Wilson, K. R. (1982). A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. J. Chem. Phys. 76, 637-649

Szymański, M.; Barciszewska, M. Z.; Erdmann, V. A.; & Barciszewski, J. (2000). An Analysis of G-U Base Pair Occurrence in Eukaryotic 5S rRNAs. Mol. Biol. Evol. 17, 1194-1198.

Szymanski, M.; & Barciszewski, J. (1999). The aminoacyl-tRNA synthetase data bank (AARSDB). Nucleic Acid Res. 27, 332-335.

Takahasi, K.; Yoneyama, M.; Nishihori, T.; Hirai, R.; Kumeta, H.; Narita, R.; Gale, M.; Inagaki, F.; & Fujita, T. (2008). Nonself RNA-Sensing Mechanism of RIG-I Helicase and Activation of Antiviral Immune Responses. Mol. Cell 29, 428-440.

Trappl, K.; Mathew, M.A.; & Joseph, S. (2014). Thermodynamic and kinetic insights into stop codon recognition by release factor 1. PLoS ONE 9, e94058.

Tembe, B. L.; & Mccammon, J. A. (1984). Ligand receptor interactions. Comput. Chem. 8, 281-283.

Torrie, G. M.; & Valleau, J. P. (1974). Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. Chem. Phys. Lett. 28, 578-581.

Trobro, S.; & Åqvist, J. (2007). A model for how ribosomal release factors induce peptidyl-tRNA cleavage in termination of protein synthesis. Mol. Cell. 27, 758-766.

Ueda, H.; Shouku, Y.; Hayashi, N.; Mitsunaga, J.; In, Y.; Doi, M.; Inoune, M.; & Ishida, T. (1991). X-ray crystallographic conformational study of 59-O-[N-(L-alanyl)-sulfamoyl] adenosine, a substrate analogue for alanyl-tRNA synthetase. Biochim. Biophys. Acta. 1080, 126-134.

Uter, N. T.; & Perona, J. J. (2004). Long-range intramolecular signaling in a tRNAsynthetase complex revealed by pre-steady-state kinetics. Proc. Natl. Acad. Sci. USA. 101, 14396-14401.

Uzri, D.; & Gehrke, L. (2009). Nucleotide sequences and modifications that determine RIG-I/RNA binding and signaling activities. J. Virol. 83, 4174-4184.

van Dijk, E.; Cougot, N.; Meyer, S.; Babajko, S.; Wahle, E.; & Séraphin, B. (2002). Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. EMBO J. 21, 6915-6924.

van Gunsteren, W. F.; Daura, X.; & Mark, A. E. (2002). Computation of free energy. Helvetica Chimica. Acta. 85, 3113-3129.

Varani, G.; & McClain, W. H. (2000). The G.U wobble base pair. A fundamental building block of RNAstructure crucial to RNAfunction in diverse biological systems. EMBO Rep. 1, 18-23.

Vela, A.; Fedorova, O.; Ding, S. C.; & Pyle, A. M. (2012). The thermodynamic basis for viral RNA detection by the RIG-I innate immune sensor. J. Biol. Chem. 287, 42564-42573.

Verlet, L. (1967). Computer "Experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Physical Rev. 159, 98-103.

Verbeeck, R. M. H.; Bruyne, P.; Driessens, F. C. M.; & Verbeek, F. (1984). Solubility of Magnesium Hydrogen Phoaphate Trihydrate and Ion-Pair Formation in the System $Mg(OH)_2$-H_3PO_4-H_2O at 25 C. Inorg. Chem. 23, 1922-1926.

Wan, F.; Anderson, D. E.; Barnitz, R. A.; Snow, A.; Bidere, N.; Zheng, L.; Hegde, V.; Lam, L. T.; Staudt, L. M.; Levens, D.; Deutsch, W. A.; & Lenardo, M. J. (2007). Ribosomal protein S3: a KH domain subunit in NF-kappaB complexes that mediates selective gene regulation. Cell 131, 927-939.

Wang, Y.; Ludwig, J.; Schuberth, C.; Goldeck, M.; Schlee, M.; Li, H.; Juranek, S.; Sheng, G.; Micura, R.; Tuschl, T.; Hartmann, G., & Patel, D. J. (2010). Structural and functional insights into 5′-ppp RNA pattern recognition by the innate immune receptor RIG-I. Nat. Struct. Mol. Biol. 17, 781-787.

Warshel, A.; Kato, M.; & Pisliakov, A. V. (2007). Polarizable Force Fields: History, Test Cases, and Prospects. J. Chem. Theory Comput. 3, 2034-2045.

Weixlbaumer, A.; Jin, H.; Neubauer, C.; Voorhees, R. M.; Petry, S.; Kelley, A. C.; & Ramakrishnan, V. (2008). Insights into translational termination from the structure of RF2 bound to the ribosome. Science 322, 953-956.

Weygand-Durasević, I.; Rogers, M. J.; & Söll, D. (1994). Connecting anticodon recognition with the active site of Escherichia coli glutaminyl tRNAsynthetase. J. Mol. Biol. 240, 111-118.

Whitelaw, E.; & Proudfoot, N. (1986). Alpha-thalassaemia caused by a poly(a) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. EMBO J. 5, 2915-2922.

Widom, B. (1963). Some topics in the theory of fluids. J. Chem. Phys. 39, 2808-2812.

Wong, L.E.; Li, Y.; Pillay, S.; Frolova, L.; & Pervushin, K. (2012). Selectivity of stop codon recognition in translation termination is modulated by multiple conformations of GTS loop in eRF1. Nucleic Acids Res. 40, 5751-5765.

Yoneyama, M.; Kikuchi, M.; Matsumoto, K.; Imaizumi, T.; Miyagishi, M.; Taira, K.; Foy, E.; Loo, Y. M.; Gale, M., Jr; Akira, S.; Yonehara, S.; Kato, A.; & Fujita, T. (2005). Shared

and unique functions of the DexD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. J. Immunol. 175, 2851-2858.

Yoneyama, M.; Kikuchi, M.; Natsukawa, T.; Shinobu, N.; Imaizumi, T.; Miyagishi, M.; Taira, K.; & Fujita, T. (2004). The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. Nat. Immunol. 5, 730-737.

Zhang, C. M.; Perona, J. J.; Ryu, K.; Francklyn, C.; & Hou, Y. M. (2006). Distinct Kinetic Mechanisms of the Two Classes of Aminoacyl-tRNA Synthetases. J. Mol. Biol. 361, 300-311.

Zhou, .J; Korostelev, A.; Lancaster, L.; & Noller, H.F. (2012). Crystal structures of 70S ribosomes bound to release factors RF1, RF2 and RF3. Curr. Opin. Struct. Biol. 22, 733-742.

Zwanzig, R. W. (1954). High- temperature equation of state by a perturbation method. I. Nonpolar gases. J. Chem. Phys. 22, 1420-1426.

List of Publications:

(1) **Kumar, A.**; Basu, D.; Satpati, P.* Structure based energetics of stop codon recognition by eukaryotic release factor. *Journal of Chemical Information and Modeling 2017,* 57, 2321-2328.

(2) **Kumar, A.**; Satpati, P.* Energetic of preferential binding of RIG-I to double-stranded viral RNAs with 5' tri/di phosphate over 5' monophosphate. *ACS omega 2018*, 3, 3786-3795.

(3) **Kumar, A.**; Åqvist, J.;* Satpati, P.* Principles of $tRNA^{Ala}$ selection by alanine tRNA synthetase based on the critical G3·U70 base pair. *ACS omega 2019*, 4, 15539-15548.

(4) **Kumar, A.**; Satpati, P.* Why dsRNA with 5'-monophosphate is a poor binder to RIG-I with respect to 5'-OH analogue? *Journal of Biomolecular Structure and Dynamics 2019*, 1-11.

(5) Shukla, S.; **Kumar, A.**; Das, D.; Satpati, P. Principle of DNA reocgnition by sporulation regulator protein Sp0A in B. Subtilis. *Journal of Biomolecular Structure and Dynamics 2019,* 1-9.

(6) **Kumar, A.**; Satpati, P.* Does X-ray structure of RIG-I bound viral 5'-OH-dsRNA is a model for understanding 5'-p-dsRNA binding? (*Manuscript under preparation*).

(7) **Kumar, A.**; Satpati, P.* Mutations in AlaRS and its ability to recognize $tRNA^{Ala}$ (*Manuscript under preparation*).

Conferences and Workshops attended:

(1) Participated in Symposium cum Workshop on "Advanced Mathematical Modeling for Bioengineering Applications" organized by Bioinformatics Infrastructure Facility, Supported by DBT, Govt. Of India, Department of Biosciences and Bioengineering, (BSBE), IIT Guwahati. 17^{th}-18^{th} May 2019.

(2) Participated in International Seminar cum Workshop on "Computer Aided Drug Design for Human Pathogens (CADDHP)" held by Tezpur University, Assam, India. 12^{th}-17^{th} February 2018.

(3) Poster presentation in Research conclave organized by Student's Academic Board (SAB), IIT, Guwahati. Presented a poster on "Breast Cancer Metastases to Bone". 17^{th}- 20^{th} March 2015.