Molecular dynamics simulation of RNA ribosomal frameshift stimulatory elements

Seah, Yi Ling

2017

Seah, Y. L. (2017). Molecular dynamics simulation of RNA ribosomal frameshift stimulatory elements. Master's thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/72394>

<https://doi.org/10.32657/10356/72394>

MD SIMULATION OF RNA RIBOSOMAL
FRAMESHIFT STIMULATORY ELEMENT

**MOLECULAR DYNAMICS SIMULATION
OF RNA RIBOSOMAL FRAMESHIFT
STIMULATORY ELEMENTS**

SEAH YI LING

**SEAH YI LING
SCHOOL OF BIOLOGICAL SCIENCES**

2017

2017

**MOLECULAR DYNAMICS SIMULATION
OF RNA RIBOSOMAL FRAMESHIFT
STIMULATORY ELEMENTS**

SEAH YI LING 2017

SEAH YI LING

School of Biological Sciences

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Masters of Science

2017

Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Dr. Lu Lanyuan of Nanyang Technological University, for his meticulous guidance and remarkable patience throughout the Masters project. I am extremely thankful to be given the opportunity to work in the lab.

I am grateful to Mr Zhang Haiping who started the HIV project, for his suggestions and invaluable advice for troubleshooting. I am also thankful to Dr. Nalaparaju Anjaiah, Mr Zhu Guanhua, Mr Bope Domilongo Christian, Mr Tong Dudu, Dr. Chen Wenwen, Ms Pang Shi Hui Esther and Mr Lim Wei Sheng Melvin for the advice and encouragement they have rendered to me throughout the course of the project.

I would also like to express my deepest gratitude to Dr. Chen Gang of Nanyang Technological University, School of Mathematical and Physical Sciences (SPMS) and his lab members for their collaborations on the optical tweezer experiments.

Last but not least, I am grateful to my family and friends for their never-ending encouragement and continual support.

Table of Contents

Chapter 1: Introduction	1
1.1 Background	1
1.1.1 Programmed Ribosomal Frameshifting	2
1.1.2 -1 Programmed Ribosomal Frameshifting Mechanism	4
1.2 Related Works	8
1.2.1 hTR Experimental Results	8
1.2.2 HIV RNA Experimental Results	14
1.3 Computational Approaches	17
1.3.1 Basic Principles of Molecular Dynamics	17
1.3.2 Force Fields	18
1.4 Motivation	19
Chapter 2: Materials and Methods	21
2.1 Preparation of Initial RNA and Ligand Structures	21
2.2 Generation of Topology and Energy Minimized Structures	23
2.3 Generation of Ligand Topology	25
2.4 Equilibrium Simulation	25
2.5 Steered Molecular Dynamics Simulation	26
2.6 Umbrella Sampling	27
2.7 Metadynamics Simulation	28
Chapter 3: Results	30
3.1 RNA Systems of Study	30
3.2 H-type Pseudoknot of human telomerase RNA (hTR)	31
3.2.1 Stable initial pseudoknot structures	31
3.2.2 Three possible pathways for mechanical unfolding ----	33
3.2.3 Two energy minima observed in metadynamics	39
3.2.4 Native structure destabilized by base triples disruption	41
3.2.5 Base triples disruption reduced potential mean force	47
3.2.6 Simulated force coincides with experimental results --	49

3.3 Stem-Loop of HIV RNA -----	51
3.3.1 Stable initial stem-loop structures -----	51
3.3.2 Unfolding force slightly reduced by anti-FS ligand -----	52
Chapter 4: Discussion -----	55
Chapter 5: Future Works -----	62
Chapter 6: Conclusion -----	64
References -----	65
Appendix -----	76

List of Figures

Introduction

Figure 1.1: Alternative products produced from the same mRNA ---	3
Figure 1.2: The three mRNA components required for -1 PRF -----	5
Figure 1.3: Structure of a classical RNA H-type pseudoknot -----	7
Figure 1.4: Structure of hTR pseudoknots used for MD simulation -	10
Figure 1.5: Proposed unfolding pathways of $\Delta U177$ -----	11
Figure 1.6: 1-step experimental unfolding force histograms -----	12
Figure 1.7: Correlation of FS efficiency and experimental force -----	13
Figure 1.8: Structure of HIV RNA stem-loop and ligand -----	16

Results

Figure 3.1: Equilibrium simulations of hTR pseudoknots -----	32
Figure 3.2: Snapshots of hTR $\Delta U177$ during unfolding -----	34
Figure 3.3: SMD of 1-step unfolding of $\Delta U177$ native -----	36
Figure 3.4: SMD of 1-step unfolding of $\Delta U177$ intermediate -----	37
Figure 3.5: SMD of 2-step unfolding of $\Delta U177$ intermediate -----	38
Figure 3.6: Free energy plots obtained from metadynamics -----	40
Figure 3.7: Gaussian-fitted distribution of the 1-step forces (I) -----	43
Figure 3.8: Gaussian-fitted distribution of the 1-step forces (II) -----	44
Figure 3.9: SMD of 1-step unfolding of CCC-GU native -----	46
Figure 3.10: Umbrella sampling of native 1-step $\Delta U177$ + CCCGU	48
Figure 3.11: Correlations between experiment and simulation -----	50
Figure 3.12: Equilibrium simulations of HIV RNA Stem-Loop -----	51
Figure 3.13: SMD results for unfolding of HIV Stem-Loop -----	53
Figure 3.14: SMD results for unfolding of Stem-Loop with ligand ---	54

List of Tables

Materials and Methods

Table 2.1: RNA sequences and chemical formula of ligand -----	22
Table 2.2: Dimensions of box for each RNA system -----	24
Table 2.3: RNA residues defined as CV1 and CV2 -----	29

Results

Table 3.1: List of RNA systems and its mutants investigated -----	30
Table 3.2: Total number of the type of unfolding pathways -----	45

Abbreviations

-1 PRF	Minus-one Programmed Ribosomal Frameshifting
CV	Collective Variable
FES	Free Energy Surface
ff	Force Field
FS	Frameshifting
H-type	Hairpin-type
HIV	Human Immunodeficiency Virus
hTR	Human Telomerase RNA
MD	Molecular Dynamics
PDB	Protein Data Bank
PMF	Potential of Mean Force
RMSD	Root-Mean-Square Deviation
SMD	Steered Molecular Dynamics
WHAM	Weighted Histogram Analysis Method

Abstract

The RNA hairpin-type pseudoknots and stem-loops structures are essential downstream elements for minus-one ribosomal frameshifting. While frameshifting efficiency and mechanical stability of these structures were investigated *in vivo* and *in vitro*, *in silico* studies were limited. Here, steered molecular dynamics simulations were performed to unfold hTR Δ U177 pseudoknot and its mutants. Microscopic molecular structures and the dynamic unfolding process, which are not observed under experimental conditions, were illustrated. Furthermore, less force was required to unfold the native pseudoknot mutants due to the destabilization of base triples. This suggests that pseudoknot stability is strongly dependent on stem-loop interactions. HIV RNA frameshifting stem-loop was also investigated. The presence of an anti-frameshifting ligand, DB213, was found to slightly stabilize the stem-loops, increasing the force required in our pulling simulations. Taken together, the link between mechanical stability, three-dimensional structure and frameshifting efficiency was examined, which may provide a better understanding of minus-one ribosomal frameshifting.

Chapter 1. Introduction

1.1 Background

To fulfill its diverse enzymatic and structural roles, simple ribonucleic acid (RNA) strands can fold onto itself to form complex, spatial structures. Functions of RNA can range from coding, decoding, regulation to gene expression (Elliott and Lodomery, 2016). The above concept applies to messenger RNAs (mRNAs) as well. The mRNAs transcribed function to encode proteins according to its codon sequences. Single-stranded mRNA molecules can fold back on itself due to the lack of complementary strand, forming complex secondary structures. Typical RNA secondary structures includes double-stranded helix and loops stabilized by Watson-crick (A•U or G•C) and wobble (G•U) base pairings (Ananth *et al.*, 2013). In the cytosol, ribosomes will unfold these mRNAs in order to read the codon sequence during translation. Ribosomes tend to maintain the reading frame at less than 5 errors for every 10^5 codons translated, ensuring translational fidelity (Brierley *et al.*, 1995).

1.1.1 Programmed Ribosomal Frameshifting

While the translational error rate is kept low by ribosomes, several viruses and eukaryotic cells developed specialized mechanisms to alter translational fidelity. One of these mechanisms is programmed ribosomal frameshifting (PRF). This molecular mechanism allows the synthesis of a desired ratio of structural and enzymatic proteins, by programming ribosomes to read a non-canonical frame (Dinman, 2012). This results in dual-coding genes, allowing the production of multiple functional products from a single mRNA with overlapping open reading frames (Figure 1.1). This expansion of information content is particularly important for viruses given that their genomic space is physically constrained.

The elongating ribosome can either be shifted one nucleotide position upstream or one nucleotide position downstream. Hence, PRF can be categorized into two groups, namely +1 PRF (downstream) and -1 PRF (upstream).

While mRNA recoding by +1 PRF are less common than -1 PRF, it was still observed in several organisms (Moon *et al.*, 2004). In +1 frameshift, the ribosome stalls for a long period of time at a rare codon on the slippery sequence. The ribosome can resolve this long pause simply by moving one nucleotide forward, resuming translation in the +1 frame.

On the other hand, the -1 frameshift is mechanistically different from the +1 frameshift. The ribosome is stalled at the slippery sequence due to the downstream pseudoknot structure (Moon *et al.*, 2004). To resume translation, the ribosome can backtrack one nucleotide, resolving translation in the -1 frame.

Figure 1.1. Alternative products produced from the same mRNA. The alternative products are a result of +1 frameshift and -1 frameshift (Modified from Ketteler, 2012).

1.1.2 -1 Programmed Ribosomal Frameshifting Mechanism

For -1 PRF to take place, three important components are required. These includes a 7-nt long slippery sequence, a single-stranded linker and a downstream pseudoknot (Figure 1.2). The downstream pseudoknot acts as a essential stimulatory element for -1 PRF (Huang *et al.*, 2013).

As the ribosome stalls at the mRNA slippery sequence during translation, the single-stranded linker helps to position the pseudoknot onto the ribosomal surface, inducing ribosomal stalling (Plant *et al.*, 2003). The long pause may thus stimulate the ribosome to backtrack one nucleotide, from the 0 frame (- X XXY YYZ -) to the -1 frame (- XXX YYY Z -). This movement allows the ribosome to resume translation by directing the ribosome to utilize an alternative start site. In Figure 1.2, XXX can be substituted by any three identical nucleotides, YYY can be substituted by three adenines or uracils and Z can be substituted to adenine, uracil or cytosine.

Figure 1.2. The three mRNA components required for -1 PRF: Slippery sequence, spacer (also known as linker) and pseudoknot. XXX can be substituted to any three identical nucleotides. YYY can be substituted to either three identical adenine or uracil and Z can be substituted to adenine, uracil or cytosine. (Adapted from Chen *et al.*, 2009).

The most common stimulatory pseudoknot is the H-type pseudoknot (Gledroc and Cornish, 2009). The H-type pseudoknot involves intramolecular base-pairings between nucleotides in a loop and nucleotides outside a stem (Figure 1.3A and 1.3B). This results in the formation of a pseudoknot structure with two stem-loops (Figure 1.3C). The single-stranded loops of the pseudoknot often form base-pairings with the neighboring stems, resulting in the formation of base triples which further stabilizes the triplex structure (Staple and Butcher, 2005). These H-type pseudoknots can be found in many RNA viruses such as severe acute respiratory syndrome coronavirus (SARS-CoV) (Figure 1.3D), mouse mammary tumour virus (MMTV) and beet western yellow

virus (BWYV) (Giedroc *et al.*, 2000; Staple and Butcher, 2005). This H-type pseudoknot can also be found in humans, such as the human telomerase RNA Δ U177 pseudoknot (Chen *et al.*, 2009)

While pseudoknot is the most common downstream frameshifting (FS) stimulant, other RNA secondary structures may also effectively substitute pseudoknots in -1 RFS (Bidou *et al.*, 1997). One such example is HIV-1, whereby the RNA stem-loop structure stimulates -1 RFS instead of the H-type pseudoknot.

For translation to occur, all mRNA secondary and tertiary structures have to be unfolded by elongating ribosomes. Hence, the downstream pseudoknot or stem-loop will be exposed to the mechanical force exerted by the ribosome (Green *et al.*, 2007). With this, it was proposed that -1 PRF is likely to be dependent on mechanical tension induced by the downstream pseudoknots or stem-loops (Namy *et al.*, 2006). It is therefore insightful to mechanically unfold these stimulatory structures involved in translational regulation, to investigate their mechanical properties.

Figure 1.3. Structure of a classical RNA H-type pseudoknot. (A) RNA pseudoknot that is completely unfolded. Dotted lines represent the base-pairings between different RNA segments. (B) Initial hairpin stem 1 structure. Base-pairings between the nucleotides of loop 1 and nucleotides outside stem 1 are formed. (C) Classical H-type pseudoknot structure. (D) Complex H-type pseudoknot of the SARS coronavirus (SARS-CoV) with three loops and three stems (Adapted from Staple and Butcher, 2005).

1.2 Related Works

Several optical tweezers experiments had been carried out to test the mechanical tension of -1 PRF stimulatory elements, such as RNA pseudoknots. This chapter highlights the experimental results of two RNA systems which we aim to simulate computationally, namely human telomerase RNA (hTR) and human immunodeficiency virus (HIV) RNA.

1.2.1 hTR Experimental Results

Single-molecule manipulation of the hTR P2BP3 pseudoknot was performed by Chen *et al.*, (2009). This pseudoknot, also known as $\Delta U177$, was selected for study as it is well-defined H-type pseudoknot with extensive minor- and major-groove interactions.

Optical tweezers were used to fold and unfold $\Delta U177$ and its mutants. The mutant pseudoknots were specifically designed to investigate the effect of major-groove and minor-groove base triples on mechanical stability and FS efficiency (Figure 1.4). By applying external forces onto the 5' and 3' end of the pseudoknots, the kinetics of mechanical unfolding of these RNA pseudoknots revealed their unfolding mechanisms. Figure 1.5 illustrates the 3 unfolding pathways of $\Delta U177$ proposed by Chen *et al.* (2009).

Figure 1.6 highlights the 1-step unfolding forces of wild-type $\Delta U177$ hTR pseudoknots and its three mutants, namely CCC-GU, CCC and GU pseudoknots. Mutational analysis revealed that the triplex structure highly aids in the stability of native pseudoknots, but not unfolding intermediate structures. Moreover, results of the experimental FS assay showed that FS efficiency was the highest in $\Delta U177$ and lowest in CCC-GU whereby all 5 base triples were broken (Chen *et al.*, 2009). Lastly, the unfolding forces were found to exponentially correlate with FS efficiency (Figure 1.7). Collectively, it was concluded that the triplex structure of hTR pseudoknot is essential for mechanical stability, and thus -1 PRF.

Figure 1.4. Structure of hTR pseudoknots used for MD simulation. Stem 1, stem 2, loop1 and loop 2 are colour-coded in red, blue, orange and green respectively. The mutant structures had UUU residues replaced with CCC, or AA residues replaced with GU, or both. These results in the disruption of hydrogen bonds of the base triples between stem 1 and loop 2 and/or stem 2 and loop 1, as indicated by the dotted lines (Modified from Chen *et al.*, 2009).

Figure 1.5. Proposed unfolding pathways of $\Delta U177$. Stems and loops were color-coded according to Figure 1.4. (A) Proposed stem 1 hairpin intermediate structure. (B) Proposed folding intermediate with stem 1 fully intact. (C) Native pseudoknot structure (Adapted from Chen *et al.*, 2009).

Figure 1.6. 1-step experimental unfolding force histograms. (A) $\Delta U177$ (B) CCC-GU mutant (C) CCC mutant (D) GU mutant. Values indicated in the graph highlights the peak values of Gaussian distributions with its standard deviations. The values for 1-step intermediate unfolding reactions were represented in green while the values for 1-step native pseudoknot unfolding reactions were represented in blue (Adapted from Chen *et al.*, 2009).

Figure 1.7. Correlation between frameshifting (FS) efficiency and experimental unfolding force. An exponential function with a correlation coefficient (R^2) of 0.84 fits the data well. Error bars indicates the standard deviation for the unfolding forces from multiple measurements and standard errors of FS efficiency (Adapted from Chen *et al.*, 2009).

1.2.2 HIV RNA Experimental Results

HIV, which is the primary cause of AIDS, continues to pose as a major health concern around the world. Current worldwide estimates as of 2015 indicate that 36.7 million people are infected with HIV, and there are about 35 million deaths that result from complications arising from AIDS (UNAIDS, 2016). While the AIDS-related deaths had fallen from 2 million in 2005 to 1.1 million in 2015 due to the use of anti-retroviral therapy, the numbers remain shockingly high (Hazuda *et al.*, 2000; Lefebvre & Schiffer, 2008; Schiller & Youssef-Bessler, 2009; UNAIDS, 2016). This may be due to the prevalence of multi-drug resistant mutants from the use of anti-retroviral therapy with drugs targeting integrase, reverse transcriptase and protease (Booth *et al.*, 2007; Gonzalez *et al.*, 2007).

To address this increasingly concerning problem, recent research has focused on -1 PRF of HIV, which is a crucial event involved in the HIV replication cycle (Bidou *et al.*, 1997, Hung *et al.*, 1998). In HIV, -1 PRF efficiency occurs at about 5% to encode structural and enzymatic proteins in a molar ratio of 20:1 (Brierley *et al.*, 1992). A change in this efficiency significantly reduces infectious virus production (Dulude *et al.*, 2006). Given that -1 PRF is essential for HIV replication, several researches were carried out on this putative drug target. Interestingly, a

small molecule, known as 1,4-bis[*N*-(3-*N,N*-dimethylpropyl)amidino]-benzene tetrahydrochloride, was identified in a screening test, out of 56 000 compounds, to stimulate -1 PRF in HIV (Hung *et al.*, 1998).

Further studies of this small molecule, later termed as DB213, was found to bind to the frameshift site of HIV-1 RNA (Marcheschi *et al.*, 2011). In the study, DB213 was found to stabilize HIV-1 RNA in the UV-monitored thermal denaturation experiment as binding of DB213 increases the melting temperature by 10°C. The NMR structure of HIV RNA-DB213 complex was obtained, whereby DB213 forms hydrogen bonds with the major groove of HIV FS stem-loop. This provides a starting point for the structure-based optimization of HIV FS-targeting compounds.

Figure 1.8. Structure of HIV RNA stem-loop and ligand used for MD simulation. (A) Stem 1, stem 2, loop 1 and loop 2 of the HIV RNA frameshift site are colour-coded in red, blue, green and turquoise respectively. (B) Structure of anti-FS ligand, DB213 (1,4-bis[*N*-(3-*N,N*-dimethylpropyl)amidino]-benzene tetrahydrochloride).

1.3 Computational Approaches

The investigation of real-time conformational changes at an atomistic resolution experimentally is highly challenging. Computational modeling can thus be used to study the frameshift stimulatory elements and reveal its theoretical intermediate structures. Yet, it is important to note that the large molecular size of these RNA structures also poses problems for computational biologists. Generally, molecular dynamics (MD) simulations are restricted to picosecond-nanosecond time scale. However, this simulation time may be insufficient as the complete understanding of relevant RNA structural dynamics may only be observed after hours of monitoring (Sponer *et al.*, 2012). In this paper, the computational methods are used to complement experimental techniques to provide insights that are inaccessible by experiment.

1.3.1 Basic Principles of Molecular Dynamics

MD allows for the simulation of physical movements of molecules and atoms, with atoms and bonds being represented by balls and springs. These trajectories can be determined by solving the classical equations of motions theoretically. The potential energy and forces between the particles are defined by the molecular mechanics force fields (ff). It is important to select the relevant ff for simulations as different ff have its own advantages and disadvantages.

1.3.2 Force Fields

The amber ff is widely used for MD simulations of nucleic acids as it has well-defined partial charges, bond angle, bond length and van der Waals parameters for all common nucleotides (Aduri *et al.*, 2007). Moreover, improvements had been made to accommodate to the non-canonical DNA and RNA structures (Case *et al.*, 2015).

The current recommended ff for nucleic acids is the amber ff14sb, whereby ff99bsc0 is used for DNA and ff99bsc0_chiOL3 is used for RNA (Case *et al.*, 2015). This ff improves the accuracy of nucleic acids structural fitting as different parameters are used for DNA and RNA. For instance, ff99bsc0 tends to lock RNA sugar puckers in a northern conformation, even if the nucleotides prefer southern orientations. Hence, the parameters between DNA and RNA are non-transferable.

1.4 Motivation

Viruses tend to utilize PRF to regulate the expression of multiple gene post-transcriptionally. Hence, -1 PFS is essential for viral replication, survival as well as viral infections (Brierley & Dos Ramos, 2006; Dulude *et al.*, 2006). It is therefore insightful to fully understand the FS processes. Previously, -1 PRF of several viruses was found to be correlated to mechanical tension (Hansen *et al.*, 2007; Cao & Chen, 2008). However, recent studies on four retroviruses, two luteoviruses, one coronavirus and one non-FS bacteriophage pseudoknots showed that conformational fluctuations played a more important role in -1 RFS than the resistance of pseudoknots (Ritchie *et al.*, 2012). Hence, this project aims to simulate the relevant experiments of hTR and HIV RNA to further understand the mechanisms of -1 PRF.

Computer-assisted mechanical unfolding of the RNA pseudoknots or stem-loops will reveal the three-dimensional intermediate structures that were not fully elucidated under the optical tweezer experiments. The simulated MD force-extension profiles will be used to validate the experimental results and proposed unfolding pathways. Furthermore, the microscopic intermediate structures obtained may facilitate the development of anti-viral therapeutics targeting FS. Mutation-induced destabilization will also be studied. The analysis of both experimental and computational data obtained can ultimately reveal the RNA unfolding reactions in nature.

Moreover, the feasibility of using SMD, umbrella sampling and metadynamics to derive conclusions on the RNA models will be evaluated. This may help in future applications on similar biological problems.

Chapter 2. Materials and Methods

This chapter describes the applications and its corresponding parameters used to investigate the unfolding of hTR H-type pseudoknot and HIV RNA stem-loop structures. Parameters for both RNA systems tested were included from Section 2.1 to Section 2.7.

2.1 Preparation of Initial RNA and Ligand Structures

Initial structure of wild-type hTR Δ U177 was obtained from the NMR structure downloaded from the Protein Data Bank (PDB) (Table 2.1). For Δ U177, the 13th structure was selected out of a total of 20 NMR structures in PDB, as it has the lowest root-mean-square deviation (RMSD) to the average structure. The final hTR sequence used consists of 47 nucleotides.

Similarly, the NMR structures of HIV frameshift site RNA and RNA bound with anti-FS ligand were retrieved from the Protein Data Bank (PDB) (Table 2.1). For HIV RNA, the 1st structure was chosen out of the 20 NMR structures in PDB. Similarly, for HIV RNA bound with anti-FS ligand, the 1st structure was chosen out of the 10 NMR structures in PDB. The final HIV RNA sequence used consists of 45 nucleotides.

The hTR mutant structures (CCC-GU, CCC and GU) were generated based on wild-type Δ U177 structure downloaded. tLEAP utility

from the AmberTools package (<http://www.ambermd.org/>) was used to mutate the RNA structures (Case *et al.*, 2014). The UUU major-groove residues of loop 1 were mutated to CCC by replacing the purine and pyrimidine rings (Table 2.1, Figure 1.4). Similarly, minor-groove AA residues of loop 2 were mutated to GU.

Table 2.1: RNA sequences and chemical formula of ligand used in MD simulations. Underlined text refers to the mutated site.

RNA or Ligand	PDB ID	Sequence or Chemical Formula
hTR ΔU177	2K96	5'-GGGCUGUUUUUCUCGCUGACUUUCAGC CCCAAACAAAAAAGUCAGCA -3'
CCCGU Mutant		5'- GGGCUGU <u>CCC</u> UCUCGCUGACUUUCAGC CCCAAAC <u>GU</u> AAAAGUCAGCA -3'
CCC Mutant		5'- GGGCUGU <u>CCC</u> UCUCGCUGACUUUCAGC CCCAAACAAAAAAGUCAGCA -3'
GU Mutant		5'- GGGCUGUUUUUCUCGCUGACUUUCAGC CCCAAAC <u>GU</u> AAAAGUCAGCA -3'
HIV FS Site	1Z2J	5'- GGGAAGAUCUGGCCUUCCCACAAGGGA AGGCCAGGGAAUCUUCCC -3'
HIV FS Site + DB213 Ligand	2L94	5'- GGGAAGAUCUGGCCUUCCCACAAGGGA AGGCCAGGGAAUCUUCCC -3' + C ₁₈ H ₃₆ N ₆

2.2 Generation of Topology and Energy Minimized Structures

For both hTR and HIV systems, O3' atom of the 3' end cytosine (C) and O5' atom of the 5' end guanine (G) were aligned to the principle x-axis with the use of VMD (<http://www.ks.uiuc.edu/Research/vmd/>) (Hsin *et al.*, 2010).

All simulations in this paper were carried out with Groningen Machine for Chemical Simulations (GROMACS) software version 4.6.7 (<http://www.gromacs.org/>) (Hess *et al.*, 2008). The amber14sb force-field was applied for the generation of initial topology. Table 2.2 presents the dimensions of the box used for each of the RNA system. The dimensions were set up to fully accommodate each of the stretched RNA. The fully-stretched RNA should not take up more than half of the box along the x-axis as GROMACS takes periodicity into account while calculating distances.

Table 2.2: Dimensions of box for each RNA system.

RNA System	RNA	Dimensions of box
hTR	Wild-type	70 nm x 6 nm x 6nm
	CCC-GU Mutant	
	CCC Mutant	
	GU Mutant	
HIV	Wild-type	60 nm x 6 nm x 6nm
	Wild-type + Anti-FS	
	Ligand	

Depending on each system, the pseudoknot was solvated with about 71 200 to 82 557 TIP3P water molecules. For hTR, sodium and chloride ions were added to neutralize the negatively-charged RNA phosphate groups. Besides that, an additional 200 mM of sodium chloride was added to simulate the experimental conditions. As for the HIV RNA, potassium ions were added to counteract the negative charges.

In all simulations, periodic boundary conditions were used and the particle mesh Ewald (PME) approach was applied for calculation of electrostatic interactions. A space cut-off of 1.4 nm was used. Similarly, van der Waals interactions were evaluated with 1.4 nm cut-off. All systems were subjected to energy minimization so as to obtain a relaxed structure with optimal geometries and proper solvent orientation. LINCS

algorithm was implemented to bond constraints and steepest descent algorithm was used for energy minimization.

2.3 Generation of Ligand Topology

Given that our HIV-ligand complex would be simulated with the amber14sb force field, the topology of the L94 ligand was generated based on General Amber Force Field (GAFF) using the ACPYPE software (<http://www.biomedcentral.com/1756-0500/5/367>) (Sousa da Silva & Vranken, 2012). This software is an automated tool used to generate topologies for chemical compounds. The generated topology was used to build the HIV-ligand complex, by combining the processed RNA topology and ligand topology.

2.4 Equilibrium Simulation

Prior to steered MD (SMD) pulling simulations, the RNAs were equilibrated for a short 100 ps at a temperature of 295 K. This is also known as the isothermal-isochoric equilibration. With the exception of water molecules and ions, constraints were applied to all atoms at their initial positions during the equilibration.

Next, a final equilibration was conducted in an isothermal-isobaric ensemble. Position restraints were only applied on the O3' atom of 3' end cytosine and O5' atoms of 5' end guanine. The systems were

equilibrated for 45 ns at fixed conditions of 295K and 1 atm. Similar to the work of Bergues-Pupo *et al.* (2015), the temperature and pressure were fixed using modified Berendsen thermostat, V-rescale and Parrinello-Rahman barostat respectively.

The equilibration was verified by computing the root-mean-square deviation (RMSD). By 45 ns, the RMSD of RNA stem-loop structures should stabilize. These results were analyzed with the use of Gnuplot (<http://www.gnuplot.info/>) (Racine, 2006).

2.5 Steered Molecular Dynamics Simulation

Standard SMD simulations were employed to completely unfold the RNA structures in each of the equilibrated system. From the last 10 ns of hTR equilibration, 20 frames, with a fixed interval of 500 ps were extracted. Each extracted frame was used as the initial configuration for SMD simulation. Hence, a total of 20 SMD pulling simulations were conducted on each hTR pseudoknot. Similar to the work of Li *et al.*, (2009), the rate of pulling and force constant were fixed at 1 nm/ns and 1000 kJ/mol·nm² respectively.

While the O5' atom of 5' end guanine was physically fixed, the O3' atom of 3' end cytosine was attached to an artificial harmonic spring. With this harmonic spring, the O3' atom was pulled along the x-axis.

Each trajectory of hTR and HIV RNA were monitored and recorded over 20 ns and 27 ns respectively.

For data analysis, pulling forces were calculated every 4 ps. The coordinates of the O5' and O3' atom were calculated every 100 ps. Energy values and number of hydrogen bonds were written every 4 ps. The forces curves were smoothened by averaging equally-spaced intervals of 20 ps. All the data obtained were studied with the use of Gnuplot and VMD.

2.6 Umbrella Sampling

A series of configurations were extracted from the SMD trajectory of hTR Δ U177 and CCC-GU native pseudoknots. These configurations were chosen at an interval of 0.2 nm, which allows the sampling windows to slightly overlap. Umbrella sampling simulations were carried out on 80 configurations selected per sample. Additional simulations were added in windows with insufficient overlap.

The box size of each configuration was similar to the SMD simulations (Table 2.2). The RNA systems were re-equilibrated for 100 ps and 10 ns under the isothermal-isochoric ensemble and isothermal-isobaric ensemble respectively, similar to Section 2.4. The selected RNA pseudoknots were subjected to umbrella sampling for 20 ns each.

Weighted Histogram Analysis Method (WHAM) was used to extract and reconstruct the potential of mean force (PMF) profiles. The free energy during the pulling simulation can be derived from the PMF extracted from these series of umbrella sampling simulations.

2.7 Metadynamics Simulation

hTR systems were re-equilibrated under the isothermal-isobaric ensemble for 45 ns according to Section 2.4, except that position restraints were removed. Since no pulling will be carried out, the box size was significantly reduced to 8 nm x 8 nm x 8 nm. The unrestrained systems were subjected to metadynamics simulation.

All metadynamics simulations were carried out using the PLUMED plugin version 2.2 (<http://www.plumed.org/>) (Bonomi *et al.*, 2009). Hydrogen bond number (N_{hb}) was computed and calculated by coordination number as shown in equation 1 below, similar to that of Vargiu *et al* (2008). This coordination number was in turn used to investigate interactions or contacts between stem 1 or stem 2.

$$N_{hb} = \sum_{ij} \frac{1-(r_{ij}/r_0)^a}{1-(r_{ij}/r_0)^b} \text{ ----- (Equation 1)}$$

where $a = 8$, $b = 12$ and $r_0 = 2.5 \text{ \AA}$.

Collective variables (CVs) of hTR were identified in Table 2.3. The whole of stem 1 and stem 2 were defined as CV1 and CV2 respectively. Base triples were included in the CVs.

Table 2.3: RNA residues defined as CV1 and CV2.

RNA System	Collective Variables (CVs)	Residues
hTR	CV1	1 – 10, 21 – 29
	CV2	15 – 25, 35 – 46

The potential built up with Gaussians in this standard implementation is history-dependent. Gaussians, with the height and width of 0.3 kJ/mol and 5 respectively, were added to explore higher free-energy structures. These parameters used were similar to that of Vargiu *et al.* (2008). Gaussians were deposited at an interval of 1 ps, and a grid size of 700 x 800 was set up to speed up computation.

For analysis, the free energy surface (FES) was plotted as a function of the coordination number. This FES landscape was plotted and analyzed using Gnuplot.

Chapter 3. Results

Section 3.1 lists the RNAs targeted in this study. Section 3.2 and 3.3 presents the results of hTR and HIV RNA stem-loop structures respectively.

3.1 RNA Systems of Study

Table 3.1 lists the two RNA systems and its mutants investigated. These RNAs range in length from 45 to 47 nucleotides with a stem-loop structure. These RNAs were selected from studies performed by other research groups, if any, so that our simulated results could be compared with the experimental results.

Table 3.1: List of RNA systems and its mutants investigated.

	RNA Systems	Wild-type or Mutant	Length/bp
1	Human telomerase RNA (hTR)	Wild-type	47
		CCC-GU Mutant	47
		CCC Mutant	47
		GU Mutant	47
2	Human immunodeficiency virus (HIV)	Wild-type	45
		Wild-type + Anti-FS Ligand	45

3.2 H-type pseudoknot of human telomerase RNA (hTR)

Chen *et al.* (2009) found that the presence of base triples enhanced mechanical stability and increased -1 RFS efficiency of hTR Δ U177. The CCC-GU mutation in RNA major-grooves and minor-grooves was found to greatly destabilize the native pseudoknot, lowering the unfolding force required in the optical tweezer experiment. Furthermore, this mutation was reported to greatly reduce the frameshifting efficiency of Δ U177 from 53% to 0.1% in the bulk frameshifting assays. To fully elucidate the unfolding mechanism in nature, the mechanical unfolding of Δ U177 and its mutants were simulated.

3.2.1 Stable initial pseudoknot structures

45 ns equilibration simulations were conducted to test the validity of force field and ensure the stability of RNA structures. Amber14sb force field was initially chosen as it was found to improve the accuracy of nucleic acids simulations, including both canonical and non-canonical RNA structures (Case *et al.*, 2015).

To analyze the stability of RNA pseudoknots, RMSD was used to compute the difference between the initial crystal structure and final simulated structure. Analysis of RMSD values revealed that Δ U177 was relatively more stable than all of its mutant (Figure 3.1). Overall,

equilibrium simulations ensured that all pseudoknots were relatively stable within the last 10 ns prior to MD simulations.

Figure 3.1. Equilibrium simulations of hTR pseudoknots. (A) RMSD of Δ U177 (B) RMSD of CCC-GU (C) RMSD of CCC (D) RMSD of GU.

3.2.2 Three possible pathways for mechanical unfolding

The 20 different frames of Δ U177 pseudoknot extracted from the last 10 ns of equilibration were subjected to constant-speed SMD simulation. These 20 frames were extracted with a fixed interval of 500 ps. These independent pulling simulations revealed the different types of unfolding reactions. A total of 3 different pathways were identified by observing the trajectories in VMD together with the force peaks. The 1-step unfolding of Δ U177 native pseudoknot whereby both stems were concurrently unfolded was illustrated in Figure 3.2A. On the other hand, the 1-step unfolding pathway involving the intermediate structure with stem 2 being partially unfolded was illustrated in Figure 3.2B. As for Figure 3.2C, it exemplifies the 2-step unfolding pathway whereby a stem 1 hairpin was partially formed.

Clearly, SMD simulations are informative and essential as the intermediates states of mechanical unfolding, which cannot be seen in optical tweezer experiments, can be revealed. The trajectories validate the unfolding pathway proposed by Chen *et al.*, (2009) shown in Figure 1.5, but with a few exceptions. Three bases at the 5' end of stem 2 was observed to be still base-paired in the stem 1 hairpin intermediate, as opposed to the suggested pathway which proposed that stem 2 was completely unfolded. Also, in the stem 1 hairpin intermediate, a few hydrogen bonds were broken in the middle of stem 1 due to triplex rotation (Figure 3.2C).

Figure 3.2: Snapshots of hTR Δ U177 during unfolding. (A) Native pseudoknot unfolding, whereby both stems of the native pseudoknot unfold simultaneously. (B) 1-step intermediate unfolding. (C) 2-step intermediate unfolding with the formation of a partial stem 1 hairpin intermediate. Stems and loops were indicated in the same color as Figure 1.4.

The representative force-extension curves and hydrogen bond profiles of the three different unfolding reactions were shown in Figure 3.3, 3.4 and 3.5. Given that hydrogen bonds are defined with a cut-off distance of less than 0.35 nm between acceptor and donor atoms, only hydrogen bonds within 0.35 nm were recorded in the simulations (Hess *et al.*, 2008). The native pseudoknot unfolding force was found to be higher than that of the intermediate unfolding. For the 1-step unfolding, the highest force peak corresponds to the unfolding of native pseudoknot or intermediate. For the $\Delta U177$ native pseudoknot unfolding, the force peak value at t_1 corresponds to the breakage of all the base triples (Figure 3.3). This results in the destabilization of the native pseudoknot structure, further disrupting the triplex. Although some base pairs are still intact within stem 1 and between loop 1 and stem 2 after time t_1 , the bases within stem 2 and between loop 2 and stem 1 were completely disrupted. Hence, no base triples remained after t_1 . As for the 2-step unfolding, the two force peaks correspond to the unfolding of stem 2 that was partially formed at t_1 , and the unfolding of stem 1 hairpin intermediate at t_2 (Figure 3.5). Interestingly, it was observed that the stem 2 unfolding permits triplex rotation, allowing the disruption of base-pairings at the 3' end and in the middle of stem 1 (Figure 3.5C). Hence, stability of stem 2 is crucial as its unwinding permits structural rotation, thereby breaking stem 1. Another interesting insight about the intermediates was that no base triples were present (Figure 3.4 and 3.5).

Figure 3.3: SMD results for 1-step unfolding of $\Delta U177$ native pseudoknot.

(A) Representative force-extension profiles showing unfolding forces in kJ/mol·nm (left Y-axis) and extended length in nm (right Y-axis). Force curves (red in colour) were smoothened with splines, with a highest standard error of 0.763 kJ/mol·nm (at 14280 ps). Distances between O5' of 5' guanine and O3' of 3' cytosine were indicated by the green curve throughout 20 ns. (B) Hydrogen bonds profile throughout 20 ns. (C) Snapshots of pseudoknot structures corresponding to the different time-points: at the start of simulation (0 ps), at the force peak of 868 kJ/mol·nm (9940 ps) and at the end of simulation (20000 ps).

Figure 3.4: SMD results for 1-step unfolding of $\Delta U177$ unfolding

intermediate. (A) Representative force-extension profiles showing unfolding forces in kJ/mol·nm (left Y-axis) and extended length in nm (right Y-axis). Force curves (red in colour) were smoothened with splines, with a highest standard error of 0.694 kJ/mol·nm (at 1920 ps). Distances between O5' of 5' guanine and O3' of 3' cytosine were indicated by the green curve throughout 20 ns. (B) Hydrogen bond profile throughout 20 ns. (C) Snapshots of pseudoknot structures corresponding to the different time-points: at the start of simulation (0 ps), at the force peak of 659 kJ/mol·nm (8920 ps) and at the end of simulation (20000 ps).

Figure 3.5: SMD results for 2-step unfolding of $\Delta U177$ unfolding intermediate.

(A) Representative force-extension profiles showing unfolding forces in kJ/mol·nm (left Y-axis) and extended length in nm (right Y-axis). Force curves (red in colour) were smoothened with splines, with a highest standard error of 0.681 kJ/mol·nm (at 13060 ps). Distances between O5' of 5' guanine and O3' of 3' cytosine were indicated by the green curve throughout 20 ns. (B) Hydrogen bond profile throughout 20 ns. (C) Snapshots of pseudoknot structures corresponding to the different time-points: at the start of simulation (0 ps), at the force peaks of 327 kJ/mol·nm (6590 ps) and 546 kJ/mol·nm (9500 ps) at the end of simulation (20000 ps).

3.2.3 Two energy minima observed in metadynamics

Metadynamics simulation was carried out on $\Delta U177$ of the hTR to validate the presence of its intermediate structures. The simulation was conducted using the standard MD that was accompanied by a history-dependent bias potential (Barducci *et al.*, 2011). This bias potential reduces the possibility of visiting conformations that were already sampled, thereby accelerating the sampling of rare occurrences. The simulation was carried out for 210 ns. From the profiles, the free energies were converged from 190 ns (Figure 3.6A). Figure 3.6B shows the FES landscape that was computed as a function of the CVs in Table 2.3. From the trajectory and CV values, the metadynamics simulation began with the native pseudoknot structure, which has the lowest free energy minima. Throughout the simulation, Gaussians were added to explore different configurations once the system overcomes the lowest saddle point to escape the local minima.

The second free energy minimum was examined. The conformation at this minimum had most of its hydrogen bonds in stem 2 disrupted. This explains why the alternative pathways involve this intermediate structure with partially-unfolded stem 2, due to its relatively high stability. The free energy difference between the two local minima was calculated to be 15.5 kJ/mol. The small difference in free energy between the native and intermediate structure allows $\Delta U177$ to unfold using alternative pathways.

Figure 3.6: Free energy plots obtained from metadynamics simulations. (A) FES profiles reconstructed as a function of CV1 within the last 20 ns, with a 5 ns bin width. **(B)** FES landscape of $\Delta U177$ pseudoknot with snapshots corresponding to the local minima identified.

3.2.4 Native structure destabilized by the base triples disruption

To further analyze the effect of base triples disruptions, mutant pseudoknots CCC-GU, CCC and GU were subjected to 20 SMD pulling simulations each. The 1-step force peak values were extracted and plotted with Gaussian distribution as shown in Figure 3.7 and 3.8. The reactions were grouped into two categories: direct native pseudoknot unfolding or intermediate unfolding. Given that CCC-GU, CCC and GU mutations disrupt 5, 3 and 2 base triples respectively, less force was needed to mechanically unfold these mutant native pseudoknots. As comparison, the force peak values of the 1-step intermediate reactions were not affected by these mutations, given that all the 5 base triples were already broken in the intermediates, as mentioned previously in Section 3.2.2.

Out of a total of 20 reactions, only 15% of $\Delta U177$ underwent the 1-step native pseudoknot unfolding, as compared to 25 – 30% for the three mutants (Table 3.2). This highlights that during force-induced unfolding, the wild-type hTR pseudoknots tends to favor intermediate formation. This may because that the energy barrier of the intermediates are lower.

To further investigate the mutation-induced destabilization, the force-extension profiles of the mutated native pseudoknots were analyzed. Figure 3.9 illustrates the force-extension curve and hydrogen bond profiles of the mutant CCC-GU 1-step native pseudoknot unfolding. As compared to the native pseudoknot profiles of $\Delta U177$ in Figure 3.3 with a force peak value of 868 kJ/mol·nm, the force peak value of CCC-GU native pseudoknot was significantly lower at 649 kJ/mol·nm. This is because the force peak value of $\Delta U177$ corresponds to the disruption of all 5 base triples, while no base triples were present in CCC-GU mutant.

Interestingly, during the unfolding of CCC-GU native pseudoknot, stem 1 was completely unfolded before stem 2. This was observed to be the exact opposite for $\Delta U177$, whereby stem 2 was the first to completely unfold. This indicates the UUUA-to-CCCGU mutations loosened the stem-loop interactions, destabilizing stem 1. Hence, these base mutations significantly weakened the pseudoknot triplex, which lowered its mechanical stability.

Figure 3.7: Gaussian-fitted distribution of the 1-step unfolding forces (I).

(A) $\Delta U177$ and (2) CCC-GU. Blue curves illustrate the force required to mechanically unfold native pseudoknots, while red curves illustrate the force required to mechanically unfold intermediates. Values shown were the Gaussian peak unfolding force with standard deviations.

Figure 3.8: Gaussian-fitted distribution of the 1-step unfolding forces (II).

(A) CCC and (2) GU. Blue curves illustrate the force required to mechanically unfold native pseudoknots, while red curves illustrate the force required to mechanically unfold intermediates. Values shown were the Gaussian peak unfolding force with standard deviations.

Table 3.2: Total number and percentage of the type of unfolding pathways.

The three pathways include one-step native pseudoknot unfolding, one-step intermediate unfolding and two-step intermediate unfolding reactions. A total of 20 unfolding reactions were carried out per pseudoknot.

Pseudoknot	Numbers and Percentages of the three Classes of Reactions		
	1-step Native	1-step Intermediate	2-step Intermediate
ΔU177	3 (15%)	4 (20%)	13 (65%)
CCC-GU	6 (30%)	7 (35%)	7 (35%)
CCC	5 (25%)	9 (45%)	6 (30%)
GU	6 (30%)	9 (45%)	5 (25%)

Figure 3.9: SMD results for 1-step unfolding of CCC-GU native pseudoknot.

(A) Representative force-extension profiles showing unfolding forces in kJ/mol·nm (left Y-axis) and extended length in nm (right Y-axis). Force curves red (in colour) were smoothened with splines, with a highest standard error of 0.632 kJ/mol·nm (at 14840 ps). Distances between O5' of 5' guanine and O3' of 3' cytosine were indicated by the green curve throughout 20 ns. (B) Number of hydrogen bonds throughout the 20 ns. (C) Snapshots of pseudoknot structures corresponding to the different time-points: at the start of simulation (0 ps), at the force peak of 649 kJ/mol·nm (8600 ps) and at the end of simulation (20000 ps).

3.2.5 Base triples disruption reduced potential of mean force

Umbrella sampling was employed to determine the thermodynamic changes of the $\Delta U177$ and CCC-GU native pseudoknots throughout the 1-step SMD unfolding simulations. Restraining potentials were applied onto the 90 different reaction coordinates and the free energy changes underwent by the native pseudoknots was recorded in PMF. The PMF was extracted with the use of WHAM analysis.

The reconstructed PMF profile was plotted in Figure 3.10A. As expected, the unfolding event of CCC-GU native pseudoknot resulted in a lower PMF as compared to $\Delta U177$. This is due to the disruption of base triples which lowered the mechanical stability of the pseudoknot.

Another output of umbrella sampling is the histogram graph. A series of histograms, containing the distribution of reaction coordinates, were unbiased and collated for the computation of PMF profiles (Mills & Andricioaei, 2008). It is important to ensure that there is sufficient overlap between each window for an accurate analysis. The histograms of the 80 configurations of $\Delta U177$ and CCC-GU within the sampling windows were showed in Figure 10B and Figure10C respectively. The histogram showed a reasonable overlap between the umbrella sampling windows providing reliability of the above analysis.

Figure 3.10: Umbrella sampling of 1-step Δ U177 and CCC-GU native unfolding. (A) The potential-of-mean-force (PMF) along the end-to-end distance of Δ U177 (purple) and CCC-GU (green) native pseudoknots. (B) Histograms of configurations of Δ U177 sampled. (C) Histograms of configurations of CCC-GU sampled.

3.2.6 Simulate force coincides with experimental results

To further confirm the experimental results by Chen *et al.*, (2009), the Gaussian peak experimental and simulated forces were analyzed. The unfolding forces were defined as the force peak value from the 1-step native pseudoknot of $\Delta U177$, CCC-GU, CCC and GU. The simulated forces were found to strongly correlate with the experimental forces, with a high correlation coefficient of 0.964 (Figure 3.11A).

Furthermore, the correlation between unfolding forces and FS efficiency was tested. Figure 3.9 (B) illustrates the exponential relationship between the simulated unfolding force of the 1-step native pseudoknot unfolding and -1 FS efficiency, with a high correlation coefficient of 0.960. This validates the experimental results in Figure 1.7. This shows that the mutant pseudoknots, which had a lower FS efficiency, require a smaller unfolding force experimentally and computationally. Overall, the correlation curves further strengthened our results and analysis as our MD simulations can reveal the relative stability of hTR pseudoknot structures accurately.

Figure 3.11: Correlations between experimental and simulated results. (A) Correlation observed between experimental (pN) and simulated unfolding force ((kJ/mol·nm). The error bars indicates the standard deviations obtained from force-ramp experiments and frameshifting assays. (B) Correlation observed between experimental FS efficiency (%) and simulated unfolding force (kJ/mol·nm).

3.3 Stem-Loop of HIV RNA

Previously, Bidou *et al.* (1997) reported that *in vivo* HIV-1 FS efficiency is directly dependent on the stability of FS site stem-loop. Furthermore, Marcheschi *et al.* (2011) obtained the NMR structure of the RNA stem-loop with an anti-FS ligand attached. To better understand how the ligand works in altering FS efficiency, the mechanical unfolding of HIV RNA FS-site were simulated.

3.3.1 Stable initial stem-loop structures

Similarly, 45 ns equilibration simulations were carried out to obtain stable RNA stem-loops as a starting structure for SMD simulations. Overall, the equilibrium simulations showed that the stem-loop structures were relatively stable within the last 10 ns prior to SMD simulations (Figure 3.12).

Figure 3.12: Equilibrium simulations of HIV RNA Stem-Loop. (A) RMSD of HIV RNA Stem-Loop alone (B) RMSD of HIV RNA Stem-Loop with Ligand.

3.3.2 Unfolding force slightly reduced by anti-FS ligand

DB213 was identified as a possible anti-FS ligand by Hung *et al.* (1998). To further investigate the effect of DB213 ligand, HIV FS site stem-loop was subjected to SMD pulling simulations with and without DB213. The DB213 ligand binds to the major groove of the stem-loop and is positioned at the 5' end of stem 2.

Figure 3.13 and 3.14 illustrates the force-extension curves and hydrogen bond profiles of the HIV stem-loop and HIV stem-loop with DB213 ligand respectively. The highest force peak corresponds to the complete unfolding of the last few bases at the 3' end of stem 1. The peak unfolding force of the stem-loop with DB213 was found to be slightly higher at 269 kJ/mol·nm, as compared to 196 kJ/mol·nm of the stem-loop itself. This may indicate that the presence of ligand further increase the mechanical stability of the HIV stem-loop structure, thereby requiring a greater force to unfold.

Figure 3.13: SMD results for unfolding of HIV RNA Stem-Loop. (A)

Representative force-extension profiles showing unfolding forces in kJ/mol·nm (left Y-axis) and extended length in nm (right Y-axis). Force curves (red in colour) were smoothened with splines, with a highest standard error of 0.687 kJ/mol·nm (at 37 ps). Distances between O5' of 5' guanine and O3' of 3' cytosine were indicated by the green curve throughout 27 ns. (B) Hydrogen bond profile throughout 27 ns. (C) Snapshots of stem-loop structures corresponding to the different time-points: at the start of simulation (0 ps), at the force peak of 196 kJ/mol·nm (5000 ps) and at the end of simulation (27000 ps).

Figure 3.14: SMD results for unfolding of HIV RNA Stem-Loop with anti-FS

ligand. (A) Representative force-extension profiles showing unfolding forces in kJ/mol·nm (left Y-axis) and extended length in nm (right Y-axis). Force curves (red in colour) were smoothened with splines, with a highest standard error of 0.552 kJ/mol·nm (at 79 ps). Distances between O5' of 5' guanine and O3' of 3' cytosine were indicated by the green curve throughout 27 ns. (B) Hydrogen bond profile throughout 27 ns. (C) Snapshots of stem-loop structures corresponding to the different time-points: at the start of simulation (0 ps), at the force peak of 269 kJ/mol·nm (4933 ps) and at the end of simulation (27000 ps).

Chapter 4. Discussion

Firstly, the $\Delta U177$ pseudoknot of hTR was used as a model for the investigation of force-induced unfolding reaction using MD simulation. The findings from our computer simulations will be discussed together with the experimental data from Chen *et al.* (2009) to fully reveal the relationship between the RNA pseudoknot structure, mechanical stability and FS efficiency.

The SMD trajectories of 20 $\Delta U177$ pseudoknot samples were examined to validate the unfolding pathways proposed by Chen *et al.* (2009). In accordance to the unfolding reactions proposed, three classes of unfolding pathways were identified in our computer simulations. The presence of the unfolding intermediate was further verified and tested in metadynamics simulation as shown in the FES landscape. However, the selection of metadynamics CVs is critical. An unsuitable CV chosen may significantly affect the free energies extracted and it may not converge (Ferrario *et al.*, 2007). Hence, the reliability of our findings may be further verified by replicating the simulation with different sets of optimal CV.

While the Gaussian distribution of the unfolding force of hTR 1-step reactions was in accordance to the optical tweezer experiment, the resolution was much lower as compared to the experimental Gaussian distributions. There were greater overlaps between the native pseudoknot and unfolding intermediate distributions in our simulated

results. This may be due to the different parameters used in computational and experimental conditions. For instance, the pulling speed used in SMD (1 nm/ns) was orders of magnitude higher than the experimental pulling speed (10^{-7} nm/ns). Due to limited computational power, 1 nm/ns is the standard pulling speed used.

Furthermore, as compared to wild-type $\Delta U177$, the percentages of the 2-step unfolding reactions were reduced by nearly half in all 3 mutants. This may be due to the destabilization of stem 1 by the major-groove CCC mutations and minor-groove GU mutations. Hence, the frequency of stem 1 formation during the 2-step unfolding would be expected to be much lower. To increase the accuracy of results, the SMD simulations can be replicated more than 20 times if time permits, given that the actual optical tweezer experiment was replicated more than 500 times per sample.

In addition, the PMF profile provides an alternative approach to investigate the relative mechanical stability of the hTR pseudoknots (Bergues-Pupo *et al.*, 2015). Umbrella sampling and WHAM were used to reconstruct the PMF profiles of the 1-step unfolding of $\Delta U177$ and CCC-GU pseudoknots, as it is one of the most established and commonly used methods for free energy calculations (Torrie and Valleau, 1977). Overall, our SMD simulations, metadynamics simulation and umbrella sampling findings coincide with results of Chen *et al.*, (2009),

which emphasized that the CCC-GU mutants require a lesser force to unfold mechanically.

The points and directions of the pulling force exerted by ribosomes in cells may differ from our computational and experimental conditions. However, these SMD simulations and optical tweezer experiments still helps us to understand how the triplex structure can affect mechanical stability. Ribosomes translate and exert its helicase activity from the 5' end to 3' end. Hence, mechanical force will first be applied on stem 1 so as to unwind the pseudoknot. Our collective findings showed that the disruption of base triples, which results in stem 1 destabilization, greatly decreased the mechanical stability of triplex. Therefore, a relationship between pseudoknot force resistance and FS efficiency was still observed under experimental and simulated conditions, despite the use of different parameters and forces. Consistently, previous findings reported that there is a strong correlation between FS efficiency and stability of the first 4 base pairs at the 5' end of HIV stem-loop (Mouzakis *et al.* 2012). Hence, the pseudoknot triplex conformation strengthened by coaxial stacking may further stabilize the 5' terminal, thereby increasing the frequency of -1 PRF.

Collectively, the experimental and simulation findings of hTR pseudoknots suggest that unfolding force, and thus mechanical stability, is strongly correlated with FS efficiency. While an exponential correlation was observed in the findings of Chen *et al.*, (2009) with the use of ten

mutants, the possible correlation was not extensively investigated in our simulation as only three mutants were tested. Due to computational limitations, and thus higher pulling speed applied, the two Gaussian distributions of the 1-step unfolding forces were closer to each other. Therefore, there is likely to be no clear distinction between the two distributions for the remaining mutants. Despite these, our simulation results still coincide with their experimental findings. Hence, both our results suggest that H-type FS pseudoknot was mechanically stabilized by base triples. These tertiary interactions may serve as blockades for translating ribosomes, thereby prolonging ribosomal stalling over the slippery sequence and increasing FS efficiency. Although the mechanism and determinants of FS were not fully elucidated till today, our collective findings were consistent with several other reports. Minor-groove stem 1 and loop 2 interactions of the simian retrovirus type-1 (SRV-1) pseudoknots were reported to increase FS efficiency *in vitro* and *in vivo* (Olsthoorn *et al.*, 2010). Likewise, in the case of beet western yellow virus (BWYV), the helix-loop interaction was found to be crucial for mechanical stability and high FS efficiency (Liphardt *et al.*, 1999). An interesting point to note is that the intermediate pseudoknots probably do not induce high FS efficiency. This is due to its lower mechanical stability and force resistance, thus a smaller unfolding force is required.

Other than mechanical stability, pseudoknot structures were found to stimulate FS by enhancing torsional resistance (Plant & Dinman,

2005). Consistently, the major-groove CCC mutations that disrupt base triples between stem 2 and loop 1 was reported to destabilize the triplex structure, thereby significantly reducing FS efficiency. A possible reason is that these mutations lowered the global torsional resistance, making it easier for stem 1 to unfold. This is further validated in our simulated trajectories as the unfolding of stem 2 allows for the rotation of stem 1, breaking some of the base pairings in the middle of stem 1. Besides, Chen *et al.*, (2009) also suggested that the minor-groove interactions directly increased the mechanical stability of stem 1. With these new insights, anti-frameshifting ligands may be designed to target viral frameshifting, given that the FS stimulatory element in most viruses is the H-type pseudoknot (Huang *et al.*, 2013; Mazauric *et al.*, 2009).

Yet, FS efficiency is not solely dependent on mechanical tension and torsional resistance. It can also be affected by several elements, including the specific 3-dimensional conformations and length of stem-loops (Chen *et al.*, 2009; Green *et al.*, 2008). To further analyze the FS simulants, the HIV RNA stem-loop was also studied.

The stability HIV RNA FS stem-loop was previously found to be directly related to its FS efficiency (Bidou *et al.*, 1997; Mouzakis *et al.* 2012). Moreover, a small molecule, DB213, was found to stimulate HIV FS and inhibits viral replication in a high-throughput screen (Marcheschi *et al.*, 2011). Hence, the HIV RNA FS stem-loop was subjected to SMD

simulations with and without the DB213 ligand to investigate if the ligand increases mechanical stability of the stimulatory stem-loop.

The SMD results showed that DB213 ligand slightly increased the stability of HIV RNA stem-loop. In particular, the peak unfolding force, which corresponds to the unfolding of the last few bases at the 3' end of stem 1, was found to be higher in the sample with ligand. Hence, the addition of DB213 further stabilized stem 1. The stability of stem 1 is important as it would be the first to be unwound by the translocating ribosome. Furthermore, optical tweezer experiments of the HIV stem-loop suggests that stem 1 is unstable and it may readily be unfolded by ribosomes (Mazauric *et al.*, 2009).

While pseudoknots like hTR have high mechanical strength, the stimulatory element of HIV is a simple stem-loop structure. Hence, the base pairings in the HIV stem-loop would be broken base-by-base during pulling. This explains the slight increase in the unfolding force in the presence of DB213, as well as the relatively constant unfolding force throughout the whole simulation. Also, other factors may play a more important role in manipulating the FS efficiency of HIV, such as mutations in the slippery sequence (Chang, 2012).

Taken together, the study of these pseudoknots and stem-loop structures remain as a hot spot for drug design. Given that frameshifting is crucial for infectivity and reproduction of viruses, the discovery of anti-FS ligands may pose as a solution for viral infections and drug resistance. These novel anti-HIV therapeutic ligands which stabilize the FS RNA stem-loop, such as DB213, may be more effective in treating HIV if used in combination with current treatments.

Chapter 5. Future Works

Future research may focus on the temperature-induced unfolding of the pseudoknots and stem-loops. The thermal stability of hTR and HIV RNA structures can be investigated in MD simulations. These thermal denaturation studies will be carried out at various temperatures. Given that the DB213 can only mechanically stabilize the HIV stem-loop slightly, it may have a more significant impact on thermal stability.

Furthermore, the binding affinity of the anti-frameshifting ligand, DB213, may also be tested. Energetics often provides useful link between structure and functions. All-atom MD simulations, coupled with free-energy sampling algorithms, can be used to predict of these binding free energies (Gilson & Zhou, 2007). The potential-of-mean force (PMF) and double-decoupling method (DDM) are the more commonly used free-energy algorithms. Thermodynamic integration (TI) and free-energy perturbation (FEP) enables the computation of free energy change in either PMF or DDM (Jiao *et al.*, 2008). Other alternatives include the linear interaction energy (LIE) method, whereby only the free and bound states of the ligand will be simulated in MD (Aqvist *et al.*, 1994; Aqvist *et al.*, 2002). The main concept is to consider the absolute binding free energy as the free energy change when the ligand is transferred from a free ligand state to a bound state.

The actual optical tweezer experiments can also be carried out on the HIV RNA stem-loop structure with and without DB213. The pulling speed used in experiment can be much smaller as compared to our simulations, allowing a clearer distinction between the unfolding forces measured.

Lastly, the insights derived from hTR studies may also offer insights on other RNA viruses with similar H-type frameshifting pseudoknots. The methodology used for this project can also be used to study other frameshift stimulatory elements. With these, future studies may focus on the design of potential ligands that target viral frameshifting. This anti-viral therapeutics may be rationally designed based on the three-dimensional structures of pseudoknots during mechanical unfolding. Moreover, a virtual library screen may be carried out to identify possible small molecules that can bind to the pseudoknots and alter frameshifting efficiency.

Chapter 6. Conclusion

Consistent with experimental findings, $\Delta U177$ of the hTR was found to mechanically unfold in three types of reactions. Mutational analysis by steered molecular dynamics revealed that the UUUAA-to-CCCGU mutations greatly destabilized the pseudoknot triplex structure by disrupting base triples interactions. Taken together, the experimental and simulated studies revealed that mechanical stability of H-type pseudoknot exponentially correlates with -1 ribosomal frameshifting efficiency. Another RNA frameshifting stimulatory element was also examined in this study. Previously, the frameshifting efficiency of the HIV RNA frameshifting stem-loop was found to be dependent on its stability. Steered molecular dynamics simulation of the HIV stem-loop was carried out in the presence and absence of an anti-frameshifting ligand, DB213. DB213 was found to interact with the major-groove of the stem-loop, mainly in stem 2. Results revealed that the stem-loop structure, particularly stem 1, was slightly stabilized. Overall, our findings provide important insights on the unfolding of frameshifting pseudoknots and stem-loops in nature.

References

- Aduri R., Psciuk B. T., Saro P., Taniga H., Schlegel H. B., & SantaLucia J., (2007). AMBER Force Field Parameters for the Naturally Occuring Modified Nucleosides in RNA. *J. Chem. Theory Comput.* 3(4): 1464 – 1475. doi: 10.1021/ct600329w
- Aqvist J., Medina C., & Samuelsson J. E., (1994). New method for predicting binding-affinity in computer-aided drug design. *Protein Eng* 7:385–391
- Aqvist J., Luzhkov V. B., & Brandsdal B. O., (2002). Ligand binding affinities from MD simulations. *Acc. Chem. Res.*, 35(6), 358 – 365. doi: 10.1021/ar010014p
- Ananth P., Goldsmith G., & Yathindra N. (2013). An innate twist between Crick's wobble and Watson-Crick base pairs. *RNA* 19(8): 1038 – 1053. doi: 10.1261/rna.036905.112
- Barducci A., Bonomi M., & Parrinello M. (2011). Metadynamics. *WIREs Comput Mol Sci*, 1: 826-843. doi: 10.1002/wcms.31
- Bergues-Pupo A. E., Arias-Gonzalez J. R., Morón M. C., Fiasconaro A., & Falo F. (2015). Role of the central cations in the mechanical

unfolding of DNA and RNA G-quadruplexes. *Nucleic Acids Research*, 43(15): 7638-747. doi: 10.1093/narlgkv690

Bidou L., Stahl G., Grima B., Liu H., Cassan M., & Rousset J. P. (1997). In vivo HIV-1 frameshifting efficiency is directly related to the stability of the stem-loop stimulatory signal. *RNA*, 3: 1153-1158.

Booth C. L., Garcia-Diaz A. M., Youle M. S., Johnson M. A., Phillips A., Geretti A. M., (2007). Prevalence and predictors of antiretroviral drug resistance in newly diagnosed HIV-1 infection. *J Antimicrob Chemother*, 59:517–524

Brierley I., Jenner A..J., & Inglis, S. C., (1992). Mutational analysis of the “slippery-sequence” component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.*, 227:463–479

Brierley Ian (1995). Ribosomal frameshifting on viral RNAs. *J. Gen. Virol.*, 76: 1885-1892. doi: 10.1099/0022-1317-76-8-1885

Brierley, I., & Dos Ramos, F. J. (2006). Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res.* 119, 29–42.

Cao S., & Chen S., (2007). Predicting ribosomal frameshifting efficiency.
Phys Biol, 5(1): 16002. doi: 10.1088/1478-3975/5/1/016002

Case D.A., Berryman J.T., Betz R.M., Cerutti D.S., Cheatham T.E., III,
Darden T.A., Duke R.E., Giese T.J., Gohlke H., Goetz A.W.,
Homeyer N., Izadi S., Janowski P., Kaus J., Kovalenko A., Lee
T.S., LeGrand S., Li P., Luchko T., Luo R., Madej B., Merz K.M.,
Monard G., Needham P., Nguyen H., Nguyen H.T., Omelyan I.,
Onufriev A., Roe D.R., Roitberg A., Salomon-Ferrer R.,
Simmerling C.L., Smith W., Swails J., Walker R.C., Wang J., Wolf
R.M., Wu X., York D.M. & Kollman P.A., (2015). AMBER 2015,
University of California, San Francisco.

Chang K., (2012) Revealing -1 Programmed Ribosomal Frameshifting
Mechanisms by Single-Molecule Techniques and Computational
Methods. *Comput Math Methods Med*, 569870.
doi: 10.1155/2012/569870

Chen G., Chang K., Chou M., Bustamante C., & Tinoco I. Jr. (2009).
Triplex structures in an RNA pseudoknot enhance mechanical
stability and increase efficiency of -1 ribosomal frameshifting.
PNAS 106(31): 12706 – 12711. doi: 10.1073/pnas.0905046106

- Dinman Jonathan D. (2012). Mechanisms and implications of programmed translational frameshifting. *WIREs RNA*, 3:661-673. doi: 10.1002/wrna.1126
- Dulude, D., Berchiche, Y. A., Gendron, K., Brakier-Gingras, L., and Heveker, N. (2006). Decreasing the frameshift efficiency translates into an equivalent reduction of the replication of the human immunodeficiency virus type 1. *Virology* 345, 127–136.
- Elliott D., & Lodomery M. (2016). *Molecular Biology of RNA*. United Kingdom: Oxford University Press.
- Ferrario M., Ciccotti G., & Binder K., (2007). *Computer Simulations in Condensed Matter: From Materials to Chemical Biology (Volume 1)*. Mainz, Germany: Springer.
- Giedroc D. P., Theimer C. A., & Nixon P. L. (2000). Structure, Stability and Function of RNA Pseudoknots Involved in Stimulating Ribosomal Frameshifting. *J. Mol. Biol.*, 298, 167 – 185. doi: 10.1006/jmbi.2000.3668

Giedroc D. P., & Cornish P. V. (2010). Frameshifting RNA Pseudoknots: Structure and Mechanism. *Virus Res.* 139(2): 193-208. doi: 10.1016/j.virusres.2008.06.008

Gilson M. K., & Zhou H. X., (2007). Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct* 36:21–42

Gonsalez C. R., Alcalde R., Nishiya A., Barreto C. C., Silva F. E., de Almeida A., Mendonca M., Ferreira F., Fernandes S. S., Casseb J., Duarte A. J., (2007). Drug resistance among chronic HIV-1-infected patients naive for use of anti-retroviral therapy in Sao Paulo city. *Virus Res*, 129:87–90

Green L., Kim C., Bustamante C., & Tinoco I. Jr. (2008). Characterization of the Mechanical Unfolding of RNA Pseudoknots. *Journal of Molecular Biology*, 375(2): 511-528. doi: doi:10.1016/j.jmb.2007.05.058

Hansen M. T., Reihani S. N. S., Oddershede L. B., & Serensen M. A., (2008). Correlation between mechanical strength of messenger RNA pseudoknots and ribosomal frameshifting. *PNAS*, 104(14), 5830-5835. doi: www.pnas.org/cgi/doi10.1073/pnas.0608668104

Hazuda D. J., Felock P., Witmer M., Wolfe A., Stillmock K., Grobler J. A.,
Espeseth A., Gabryelski L., Schleif W., Blau C., Miller M. D.,
(2000). Inhibitors of strand transfer that prevent integration and
inhibit HIV-1 replication in cells. *Science*, 287:646–65.

Hess B., Kutzner C., Van Der Spoel D., & Lindahl E. (2008). GROMACS
4: algorithms for highly efficient, load-balanced and scalable
molecular simulations. *J. Chem. Theory Comp.*, 4, 435-447. doi:
10.1021/ct700301q

Huang X., Cheng Q., & Du Z. (2013). A Genome-Wide Analysis of RNA
Pseudoknots That Stimulate Efficient -1 Ribosomal Frameshifting
or Readthrough in Animal Viruses. *BioMed Research International*,
Volume 2013, 984028. <http://dx.doi.org/10.1155/2013/984028>

Huang X., Du Z., Cheng J., & Cheng Q. (2013). PKscan: a program to
identify H-type RNA pseudoknots in any RNA sequence with
unlimited length. *Bioinformation* 9(9): 440-442. doi:
10.6026/97320630009440

Hung, M., Patel, P., Davis, S., & Green, S. R., (1998). Importance of
ribosomal frameshifting for human immunodeficiency virus type 1
particle assembly and replication. *J. Virol.* 72, 4819–4824.

Jiao D., Golubkov P. A., Darden T. A., & Ren P., (2008). Calculation of protein–ligand binding free energy by using a polarizable potential. *PNAS*, 105(17): 6290 - 6295

Ketteler R., (2012). On programmed ribosomal frameshifting: the alternative proteomes. *Front. Genet.*, 3 (242).
doi.org/10.3389/fgene.2012.00242

Laio A., & Parrinello M., (2002). Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20), p.12562-12566.

Laio A., & Gervasio, F. L., (2008). Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics*, 71(12), p.126601.

Lefebvre E., & Schiffer C. A., (2008). Resilience to resistance of HIV-1 protease inhibitors: profile of darunavir. *AIDS Rev*, 10:131-142.

- Liphardt J., Naphine S., Kontos H., & Brierley I. (1999) Evidence for an RNA pseudoknot loop-helix interaction essential for efficient -1 ribosomal frameshifting. *J Mol Bio* 288: 321-335. doi: 10.1006/jmbi.1999.2689
- Marcheschi R. J., Tonelli M., Kumar A., Butcher S. E., (2011). Structure of the HIV-1 frameshift site RNA bound to a small molecule inhibitor of viral replication. *ACS Chem Biol.*, 6(8), 857-864. doi: 10.1021/cb200082d
- Mazauric M., Seol Y., Yoshizawa S., Visscher K., & Fourmy D., (2009). Interaction of the HIV-1 frameshift signal with the ribosome. *Nucleic Acids Res*, 37(22): 7654-7664. doi: 10.1093/nar/gkp779
- Mills, M., & Andricioaei, I., (2008). An experimentally guided umbrella sampling protocol for biomolecules. *The Journal of Chemical Physics*, 129(11), 114101. <http://doi.org/10.1063/1.2976440>
- Moon S., Byun Y., Kim H., Jeong S., Han K., (2004). Predicting genes expressed via 1 and +1 frameshifts. *Nucleic Acids Research*, 32(16), 4884-4892. doi: 10.1093/nar/gkh829

Mouzakis K. D., Lang A. L., Meulen K. A. V., Easterday P. D. & Butcher S. E. (2012). HIV-1 frameshift efficiency is primarily determined by the stability of base pairs positioned at the mRNA entrance channel of the ribosome. *Nucl. Acids Res*, 41(3): 1901-1913. doi: 10.1093/nar/gks1254

Namy O., Moran S. J., Stuart D. I., Gilbert R. J., & Brierley I. (2006). A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature*, 441: 244-247. doi: 10.1038/nature04735

Olsthoorn R. C. L., Reumerman R., Hilbers C. W., Pleij C. W. A., & Heus H. A. (2010). Functional analysis of the SRV-1 RNA frameshifting pseudoknot. *Nucleic Acids Research*, 2010, Vol. 38, No. 21 7665–7672. doi:10.1093/nar/gkq629

Plant E.P., & Dinman J. D. (2005). Torsional restraint: a new twist on frameshifting pseudoknots. *Nucl. Acids Res.* (2005) 33 (6): 1825-1833. doi: 10.1093/nar/gki329

Plant E.P., Jacobs K. L. M., Harger J. W., Meskauskas A., Jacobs J. L., Baxter J. L., Petrov A. N., & Dinman J. D. (2003). The 9-Å solution: How mRNA pseudoknots promote efficient programmed –1

ribosomal frameshifting. *RNA* 9(2): 168-174. doi:

10.1261/rna.2132503

Ritchie D. B., Foster D. A. N., & Woodside M. T., (2012). Programmed -1

frameshifting efficiency correlates with RNA pseudoknot

conformational plasticity, not resistance to mechanical unfolding.

Proc Natl Acad Sci USA, 109(40): 16167 – 16172. doi:

10.1073/pnas.1204114109

Schiller D. S., & Youssef-Bessler M., E., (2009). A second-generation

nonnucleoside reverse transcriptase inhibitor (NNRTI) active

against NNRTI-resistant strains of HIV. *Clin Ther*, 31:692-704.

Sousa Da Silva, A. W., & Vranken, W. F., (2012). ACPYPE -

AnteChamber PYthon Parser interfacE. *BMC Research Notes*.

5:367, doi:10.1186/1756-0500-5-367

Sponer J., Otyepka M., Banas P., Reblonva K., & Walter N. G., (2012).

Molecular Dynamics Simulations of RNA Molecules. In

Innovations in Biomolecular Modeling and Simulations: Volume 2

(pp. 129 – 155). doi: 10.1039/9781849735056-00129

Staple D. W., & Butcher S. E. (2005). Pseudoknots: RNA Structures with Diverse Functions. *PLoS Biol* 3(6): e213.

doi:10.1371/journal.pbio.0030213

Torrie G. M., & Valleau J. P., (1977). Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys*, 23(2): 187-199.

doi: 10.1016/0021-9991-8.

UNAIDS (2016). Fact Sheet November 2016 on Global HIV Statistics.

UNAIDS; Geneva, Switzerland. Retrieved from

http://www.unaids.org/sites/default/files/media_asset

/UNAIDS_FactSheet_en.pdf

Vargiu A. V., Ruggerone P., Magistrato A., & Carloni P. (2008).

Dissociation of minor groove binders from DNA: insights from metadynamics simulations. *Nucleic Acids Research*, 36(18):

5910-5921. doi:10.1093/nar/gkn561

Appendix

Supplementary Figure 1: SMD results for 1-step unfolding of CCC-GU unfolding intermediate.

(A) Representative force-extension profiles showing unfolding forces in kJ/mol-nm (left Y-axis) and extended length in nm (right Y-axis). Force curves (red in colour) were smoothened with splines, with a highest standard error of 0.645 kJ/mol-nm (at 11320 ps). Distances between O5' of 5' guanine and O3' of 3' cytosine were indicated by the green curve throughout 20 ns. (B) Hydrogen bond profile throughout 20 ns. (C) Snapshots of pseudoknot structures corresponding to the different time-points: at the start of simulation (0 ps), at the force peak of 573 kJ/mol-nm (8680 ps) and at the end of simulation (20000 ps). Stems and loops were indicated in the same color as in Figure 1.4.

Supplementary Figure 2: SMD results for 2-step unfolding of CCC-GU unfolding intermediate.

(A) Representative force-extension profiles showing unfolding forces in kJ/mol·nm (left Y-axis) and extended length in nm (right Y-axis). Force curves (red in colour) were smoothened with splines, with a highest standard error of 0.601 kJ/mol·nm (at 11360 ps). Distances between O5' of 5' guanine and O3' of 3' cytosine were indicated by the green curve throughout 20 ns. (B) Hydrogen bond profile throughout 20 ns. (C) Snapshots of pseudoknot structures corresponding to the different time-points: at the start of simulation (0 ps), at the force peaks of 336 kJ/mol·nm (6420 ps) and 579 kJ/mol·nm (8420 ps), and at the end of simulation (20000 ps).

Supplementary Figure 3: SMD results for 1-step unfolding of CCC native pseudoknot.

(A) Representative force-extension profiles showing unfolding forces in kJ/mol·nm (left Y-axis) and extended length in nm (right Y-axis). Force curves (red in colour) were smoothened with splines, with a highest standard error of 0.645 kJ/mol·nm (at 3620 ps). Distances between O5' of 5' guanine and O3' of 3' cytosine were indicated by the green curve throughout 20 ns. (B) Hydrogen bond profile throughout 20 ns. (C) Snapshots of pseudoknot structures corresponding to the different time-points: at the start of simulation (0 ps), at the force peak of 734 kJ/mol·nm (9020 ps) and at the end of simulation (20000 ps). Stems and loops were indicated in the same color as in Figure 1.4.

Supplementary Figure 4: SMD results for 1-step unfolding of CCC unfolding intermediate.

(A) Representative force-extension profiles showing unfolding forces in kJ/mol-nm (left Y-axis) and extended length in nm (right Y-axis). Force curves (red in colour) were smoothened with splines, with a highest standard error of 0.698 kJ/mol-nm (at 11400 ps). Distances between O5' of 5' guanine and O3' of 3' cytosine were indicated by the green curve throughout 20 ns. (B) Hydrogen bond profile throughout 20 ns. (C) Snapshots of pseudoknot structures corresponding to the different time-points: at the start of simulation (0 ps), at the force peak of 689 kJ/mol-nm (9060 ps) and at the end of simulation (20000 ps). Stems and loops were indicated in the same color as in Figure 1.4.

Supplementary Figure 5: SMD results for 2-step unfolding of CCC unfolding intermediate.

(A) Representative force-extension profiles showing unfolding forces in kJ/mol.nm (left Y-axis) and extended length in nm (right Y-axis). Force curves (red in colour) were smoothened with splines, with a highest standard error of 0.615 kJ/mol.nm (at 14660 ps). Distances between O5' of 5' guanine and O3' of 3' cytosine were indicated by the green curve throughout 20 ns. (B) Hydrogen bond profile throughout 20 ns. (C) Snapshots of pseudoknot structures corresponding to the different time-points: at the start of simulation (0 ps), at the force peaks of 552 kJ/mol.nm (7113 ps) and 598 kJ/mol.nm (8200 ps), and at the end of simulation (20000 ps).

Supplementary Figure 6: SMD results for 1-step unfolding of GU native pseudoknot.

(A) Representative force-extension profiles showing unfolding forces in kJ/mol·nm (left Y-axis) and extended length in nm (right Y-axis). Force curves (red in colour) were smoothened with splines, with a highest standard error of 0.652 kJ/mol·nm (at 5640 ps). Distances between O5' of 5' guanine and O3' of 3' cytosine were indicated by the green curve throughout 20 ns. (B) Hydrogen bond profile throughout 20 ns. (C) Snapshots of pseudoknot structures corresponding to the different time-points: at the start of simulation (0 ps), at the force peak of 700 kJ/mol·nm (8560 ps) and at the end of simulation (20000 ps). Stems and loops were indicated in the same color as in Figure 1.4.

Supplementary Figure 7: SMD results for 1-step unfolding of GU unfolding intermediate.

(A) Representative force-extension profiles showing unfolding forces in kJ/mol·nm (left Y-axis) and extended length in nm (right Y-axis). Force curves (red in colour) were smoothened with splines, with a highest standard error of 0.623 kJ/mol·nm (at 260 ps). Distances between O5' of 5' guanine and O3' of 3' cytosine were indicated by the green curve throughout 20 ns. (B) Hydrogen bond profile throughout 20 ns. (C) Snapshots of pseudoknot structures corresponding to the different time-points: at the start of simulation (0 ps), at the force peak of 647 kJ/mol·nm (8260 ps) and at the end of simulation (20000 ps). Stems and loops were indicated in the same color as in Figure 1.4.

Supplementary Figure 8: SMD results for 2-step unfolding of GU unfolding intermediate. (A) Representative force-extension profiles showing unfolding forces in kJ/mol·nm (left Y-axis) and extended length in nm (right Y-axis). Force curves (red in colour) were smoothed with splines, with a highest standard error of 0.657 kJ/mol·nm (at 2000 ps). Distances between O5' of 5' guanine and O3' of 3' cytosine were indicated by the green curve throughout 20 ns. (B) Hydrogen bond profile throughout 20 ns. (C) Snapshots of pseudoknot structures corresponding to the different time-points: at the start of simulation (0 ps), at the force peaks of 400 kJ/mol·nm (5205 ps) and 606 kJ/mol·nm (8100 ps), and at the end of simulation (20000 ps). Stems and loops were indicated in the same color as in Figure 1.4.