1-1-2014

Hiv-1 Rna Dimerization At Single Molecule Level

Hansini R. Mundigala
Wayne State University,

HIV- 1 RNA DIMERIZATION AT THE SINGLE-MOLECULE LEVEL

by

HANSINI R. MUNDIGALA

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2014

MAJOR: CHEMISTRY (Analytical)

Approved by:

_____

Advisor Date

_____

Co-Advisor Date

_____

_____

_____

DEDICATION

To my parents Ranjani Mundigala, Nandasena Mundigala

and all my teachers who paved me the path to earn this….

ACKNOWLEDGEMENTS

I had amazing, life-lasting experiences at the Wayne State University and in the Rueda Lab. I am very grateful to a number of people for giving me the opportunity to study in this environment and helping me throughout the process.

First and foremost, I would like to express my sincere gratitude to my advisor, Prof. David Rueda for continuous support during my studies and research. His patience, motivation, enthusiasm, immense knowledge and admiration of the lab work and my project have been instrumental in my development. I am grateful for his systematic guidance and diligent effort to train me to be an independent scientist and a critical thinker. I am indebted for your trust and assistance provided to me during my last two years at Wayne state university. I cannot imagine learning under a more capable advisor.

Words cannot express my gratitude well enough for having Dr Andrew Feig as my co-supervisor. I am thankful for the patience, the guidance and the time dedicated to me. I am grateful for the amazing experience and the knowledge shared being a part of the Feig lab during last two years at Wayne state university. It is a wonder to realize the heaps of knowledge you have and I am truly grateful to be working under your guidance and I am amazed at the knowledge I have gathered during the short time I spent as a member of Feig lab. Thank you very much for the encouragement and motivation during the hardest time of my PhD career.

My special thanks go to Professor Christine Chow for all her help rendered to me during my time at Rueda lab. Thank you for your time dedicated to listen to me at

anytime I needed to talk and thank you for all the wonderful advice that encouraged me to not give up at hard times. Thank you for taking care of the Rueda lab for past two years. Thank you very much for making me feel as a part of Chow lab and for taking care of the financial situation. Thank you for setting up an ideal role model that I can look up to.

I feel very fortunate to graduate under the guidance of three wonderful advisors.

I'm also thankful to my thesis committee, Prof. Colin Poole and Prof. Arun Anantharam for their encouragement and insights on my research.

Dr Amanda Solem and Dr Alfonso Brenlla have always provided me support in every aspect. Thank you very much Dr Brenlla for your support during development of this document. My past labmates including Chandani, Bishnu, Imali, Pramodha, Dr Rajan Lamichane, Dr Krishanthi Karunathilake, Dr Rui Zhao, Dr Sharla Wood, Dr May Daher, Dr Eric Patrick, Dr Guo, Dr Marcus Wood and Dr Gayan Seneviratna have always provided me with stimulating discussions and also light-hearted moments during the last two years. Present lab mates in Feig lab and Chow lab have been always helpful and very supportive. I appreciate the help given by Ashley Floyd for SPR experiments.

I am also thankful for a supportive administrative departmental staff including Melissa, Debbie, Diane, Bernadette and Nestor. Numerous other people have helped me in this project along the way. Prof. Eric Ennifar has been extremely helpful providing me with antibiotics and feed back for my work. I also thank the Sri

Lankan Student Association (SLSA) for helping me and so many other Sri Lankan students adjust to life in the U.S.

Finally, I am extremely thankful to my beloved family for their encouragement. My parents deserve thanks for too many things to name, from the time of my birth throughout my life. Thank you for building a self belief attitude. I have always seen you both with a never giving up attitude and thank you for implanting that in me. I am who I am today because of you. I am indebted to both of you giving up most of your dreams to make our dreams colorful. My two brothers, Kalana and Supun gave me incredible strength, courage and more importantly a wonderful childhood memories to cherish. I always believe that we have the best family and thank you for that. I am also grateful for extremely supportive in-laws. And lastly, I express my greatest gratitude to my beloved husband, Prabath Wijesekara for all his support, motivation, guidance and most of all for his love in every step of my life with him. I cannot imagine how I would have survived through all the difficult times during the past years without him. One cannot wish for a better husband than you. Thank you.

TABLE OF CONTENTS

Chapter 3

Chapter 4

Chapter 5

LIST OF FIGURES

LIST OF TABLES

Chapter 1

1.1 Ribonucleic Acid

Ribonucleic Acid (RNA) is one of the four key biological macromolecules together with deoxyribonucleic acid (DNA), carbohydrates, lipids and proteins. The central dogma of molecular biology states that the information in biological systems flows from DNA to proteins through RNA (Figure 1). This process is known as gene expression and consists of two important steps, namely, transcription and translation. During transcription, information encoded in DNA is transferred into messenger RNA (mRNA) [1]. Thus a messenger RNA is basically an RNA copy of the genetic information encoded in a DNA template and is required for the synthesis of proteins. Transcription is one of the highly conserved steps in gene expression. Non-eukaryotic mRNA can be directly involved in translation to synthesize the required proteins. But eukaryotic mRNA undergo further processing before they are directed to translation in cells. The mRNA processing steps performed in eukaryotic cells are 5' cap addition, splicing of non-coding intron regions, polyadenylation at the 3' end, and in some instances RNA sequence editing [2, 3]. In addition to genes that code for functional proteins, some genes also code for functional RNA such as ribosomal RNA (rRNA) or transfer RNA (tRNA). Transfer RNA binds to both mRNA and amino acid units and synthesizes the correct amino acid correct amino acid sequence for the polypeptide chain according to the information encoded in the mRNA. This process of protein synthesis is termed translation. Further processing of mRNA occurs in a macromolecular complex called the ribosome, which is composed of protein and rRNA.

During the last few decades, a novel set of RNA functions besides protein synthesis have been discovered [4]. For example, RNAs have been found to catalyze biochemical reactions (ribozymes) [5] or store genetic information in retroviruses. RNA can act as a genetic information carrier similar to DNA. RNA genomes are commonly observed in viruses [6]. They are frequently single stranded, but there are a few double-stranded RNA genomes.

RNA also performs catalytic roles, in addition to the well characterized regulatory functions. The catalytic nature of RNA that was similar to enzymes in function was first discovered in 1983 by Cech and Altman, and RNA with catalytic roles were named “ribozymes” [7, 8].

Gene regulation is another important function performed by RNA. Micro RNAs are a subset of non-coding RNA of about 20-25 nucleotides in length and involved in gene regulation in a sequence-specific procedure. They are responsible for negative regulation in gene expression at the post-transcriptional level [9]. Non-protein coding RNAs are referred to as non-coding RNA [4].

In summary, RNA in cells inherits a functional versatility that results from its structural variation. RNA mainly performs gene regulation, catalysis and also can act as a genetic information carrier. Thus, the importance of RNA to the cell leads to the important “RNA world” hypothesis, which suggests that in the early development of life, RNA assured the genetic continuity before the DNA-protein based life [10].

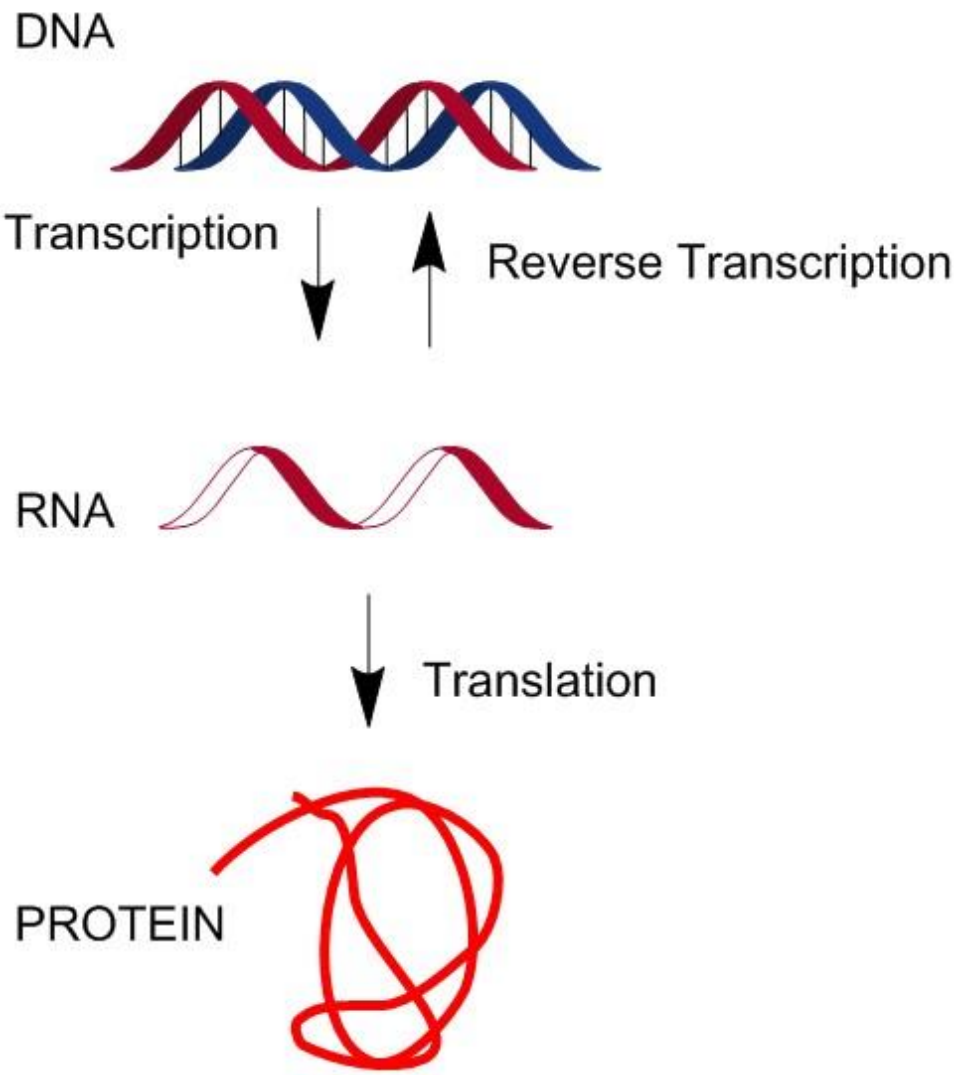

Figure 1: The central dogma of molecular biology. The genetic information in DNA is transcribed into RNA, which is then translated into protein. RNA can be reverse transcribed to form DNA.

1.1.1 Structure and Function of RNA

Although RNA has a wide range of functions in biological systems, the basic structure of RNA is similar between RNA with variable functions. The RNA primary structure is composed of four varying nucleotides bound to each other by

phosphodiester bonds. The nucleotides are composed of three main components as listed below.

1. Nitrogenous base (adenine, cytosine, guanine and uracil)

Nitrogenous bases in biological macromolecules are broadly divided into two main classes, namely purines and pyrimidines. Adenine and guanine are purine nucleotides where cytosine, uracil and thymine are pyrimidines. Thymine bases are only present in DNA structure and similarly uracil is mostly limited to RNA structure. Due to the aromatic nature of these bases, they are planar molecules. But oxygen and nitrogen atoms situated outside the aromatic ring structures might lie in a slightly different plane based on the conformation. These nitrogenous bases are attached to the five carbon sugar through an N-glycosidic bond.

2. A five carbon sugar (Ribose)

Ribose is a five-membered ring structure which is similar to the deoxyribose sugar present in DNA. The ring structure constitutes four C atoms and one oxygen atom, with additional side groups attached to the carbons. The 2'-hydroxyl group present in the ribose sugar makes RNA more polar due to its asymmetric charge distribution, resulting in an increased chemical reactivity. A N - glycosidic bond is formed between the 1'-carbon of the ribose sugar and the N atom of the base. In pyrimidine bases, the N1 involves in bond formation, and in purine bases, N9 is involved.

3. Additional phosphate group

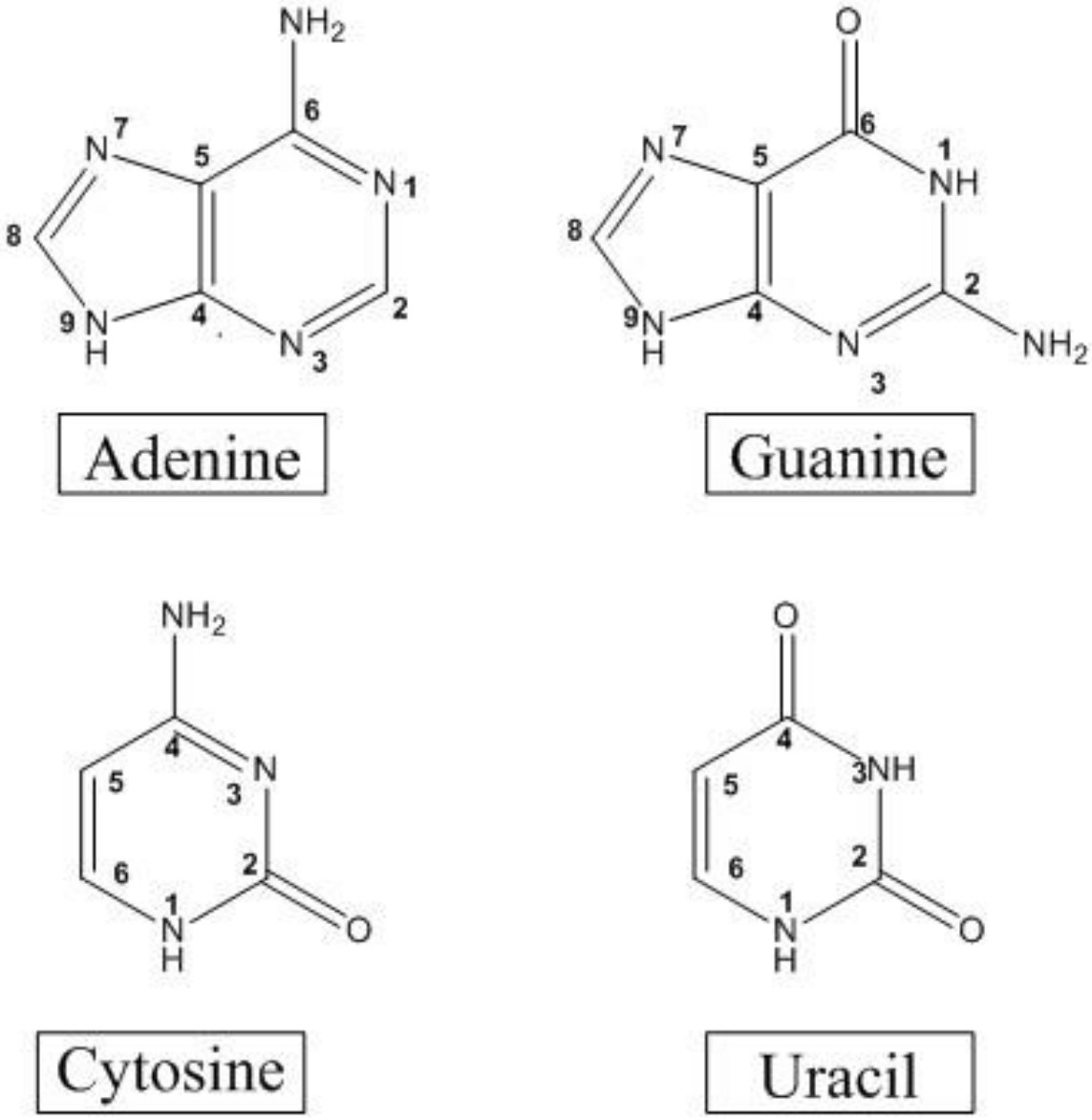

Figure 2: Four basic types of RNA bases

Individual nucleotides in RNA are linked to each other via a phophodiester bond. This phophodiester bond is identical in chemistry in both RNA and DNA. This ester bond links the phosphorous group of a nucleotide to the oxygen atom on the 3' position of the upstream ribose sugar. Similarly, the same phosphorous atom is linked to the 5' carbon of the downstream ribose sugar in the neighboring nucleotide via an ester

linkage. Thus the phosphorous atom participating in the phophodiester linkage is connected to two non-bridging oxygen atoms and two bridging oxygen atoms, as illustrated in Figure 3.

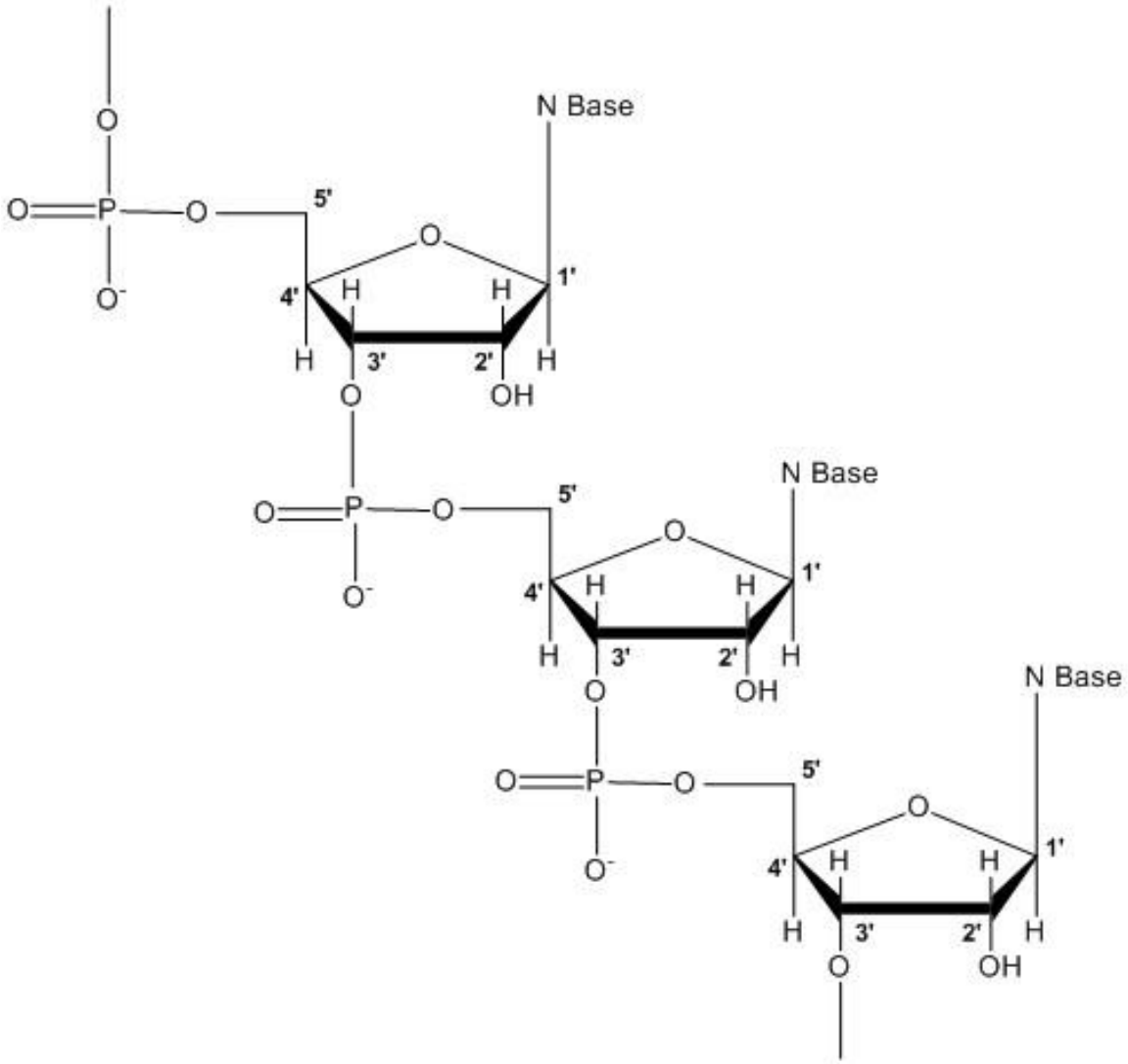

Figure 3: Individual nucleotides in RNA are linked together by 3'-5' phosphodiester bonds. Alternative sugar and phosphate components linked to each other in RNA or DNA is termed as the phophodiester backbone. The nitrogenous bases project outside this phosphodiester backbone.

The linear sequence of the nucleotides within the RNA polymer makes up the primary structure of RNA. There are two major biochemical differences in the constituents of RNA and DNA. They are 1) the presence of a 2'-hydroxyl group on the ribose sugar in RNA, and 2) the uracil base present in RNA is the non-methylated form of thymine which is primarily found in DNA. The presence of the 2' OH group provides additional flexibility to the RNA molecule. RNA molecules can further adopt secondary, tertiary, and quaternary structures to perform various functions within the cellular environment as a result of the single-stranded nature.

1.1.2 RNA Folding

For RNA to perform all of the above stated biological functions, folding into a precise three-dimensional structure is a pre-requisite. RNA folds in a precise sequential and hierarchical manner. Thus, RNA structure folding can be simplified into three distinct levels as illustrated in Figure 4.

- The primary structure of RNA, which is the sequence in which UGAC bases are connected.
- The secondary structure of RNA is formed by the interactions of the linear sequence and folding into different two-dimensional structures such as loops or hairpins as a result of RNA becoming double stranded at certain regions.
- The tertiary structure of RNA due to long-range interactions and stacking to form the complete three-dimensional conformations.

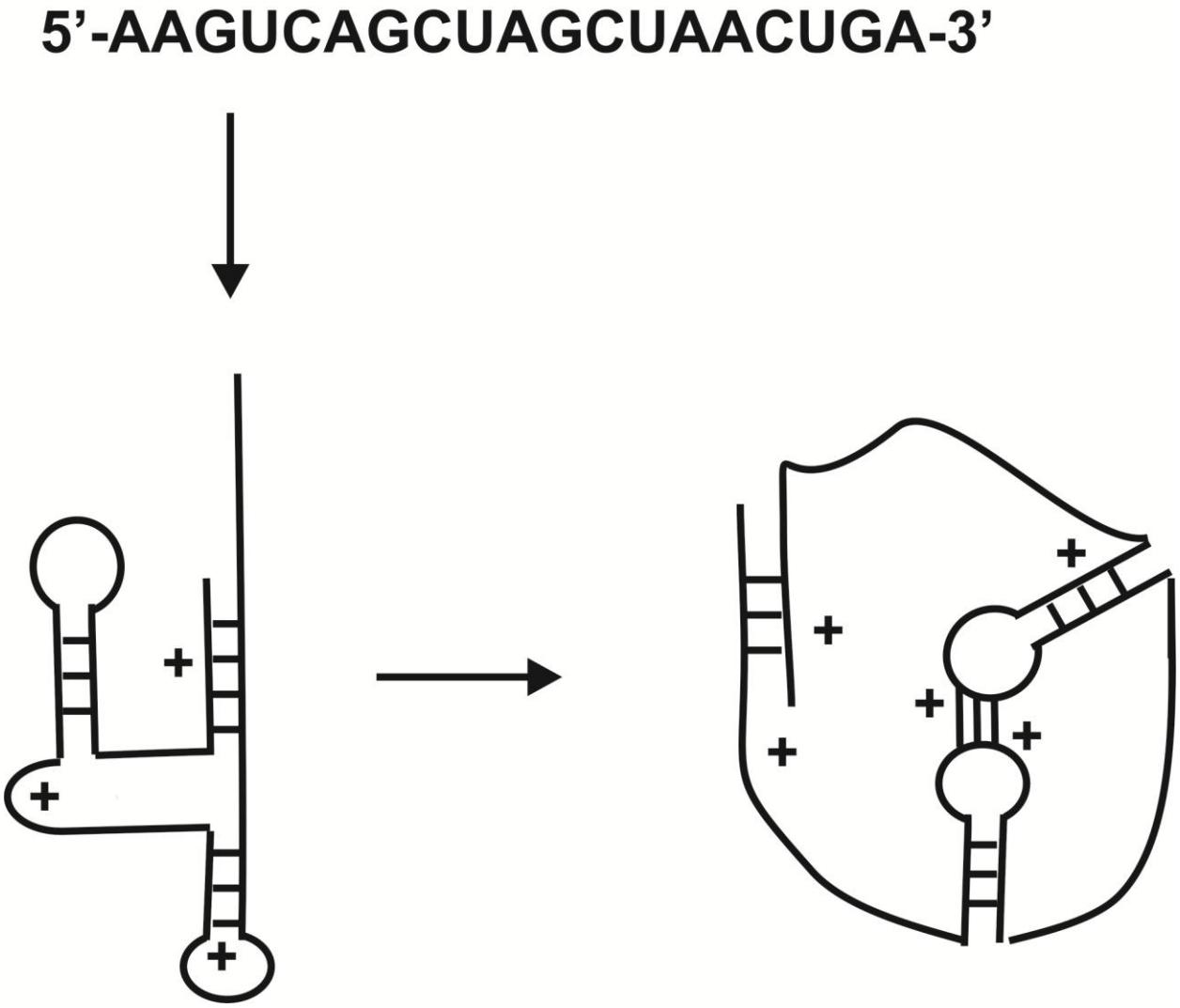

Figure 4: Hierarchy of RNA folding. A primary RNA structure is the sequence of constituting nucleic acid. Secondary structure of RNA is a result of formation of short discontinuous RNA duplexes. The formation of secondary structure is stabilized by the positive ions present in the media. Tertiary RNA structure is a result of the formation of more compact globular form with interactions between secondary structured regions.

The initial step of this folding pathway is the formation of the secondary structure, which is stimulated by any positive charge that can shield the charge repulsion originated by the close proximity of the negatively charged backbones. This charge

shielding can be accomplished by the presence of mono/di/trivalent metal ions, cationic polyamines and basic proteins [11]. Secondary structure consists of paired RNA regions forming hairpins, bulges, internal loops, and junctions (Figure 5) [12]. The next step in folding is the formation of tertiary structure, which involves interactions of the various secondary structure elements in space [12, 13]. Unlike the secondary structure, tertiary structure formation often involves specific interactions with metal ions, mostly divalent metal cations like Mg^{2+} [13]. In general potassium ions are sufficient for secondary structure formation, but the presence of Mg^{2+} or unnaturally high concentrations of potassium ions are required for tertiary structure formation [14]. In rare cases, Na^{+}, Li^{+} can replace with K^{+} and Ca^{2+}, Mn^{2+}, Cd^{2+} can replace the presence of Mg^{2+}.

[13]

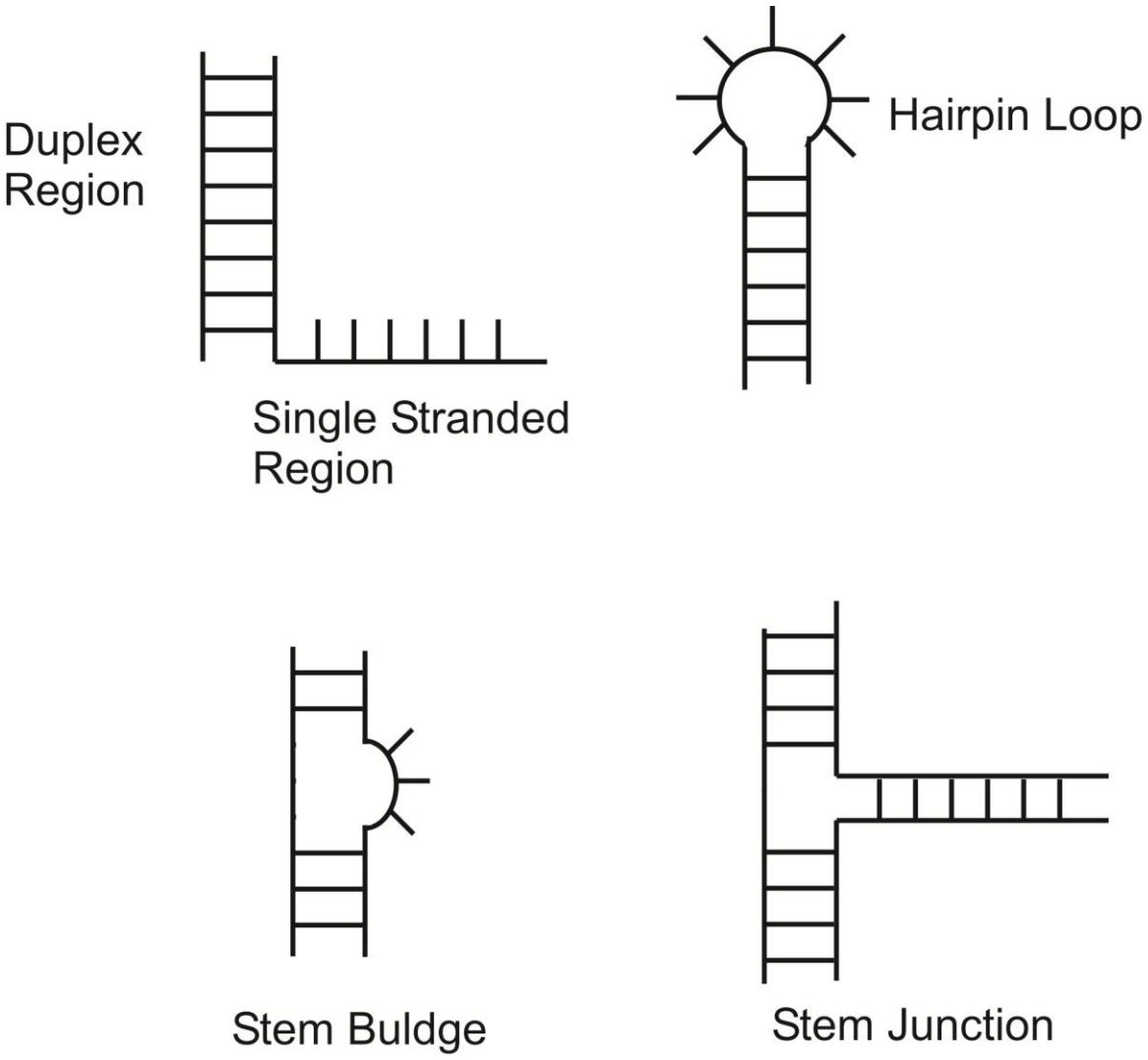

Figure 5: Commonly observed RNA secondary structure motifs. In addition to these structures, internal loops and pseudoknots are also frequently observed in RNA in physiological conditions [12].

1.1.3 Role of Metal Ions in RNA Folding

The ribose phosphate backbone in RNA strands provides a poly-anionic condition to the RNA due the presence of negatively charged phosphate groups at

physiological conditions. Thus, in biological conditions, positively charged ions act as effective charge screening groups and promote folding of RNA into complex hierarchical structures. Monovalent and divalent metal cations play a major role in RNA folding [15-17].

In general, monovalent cations are involved in non-specific electrostatic stabilization of RNA backbone negative charge [14, 15]. RNA-cation interactions that involve pure electrostatic interactions are termed diffuse binding. These interactions provide charge screening when RNA backbone segments are in close proximity [13, 15]. Magnesium ions often show specific binding to ligands without direct interaction of the Mg^{2+} with the ligand [13, 16]. These types of interactions are termed “outer-sphere interactions”, as they involve the mediation of water molecules in the hydration sphere. Magnesium involved in forming such binding can be replaced with a cobalt (III) or osmium (III)-hexamine [13, 14, 18]. In addition to specific binding, there can be non-specific interactions through water molecules in the hydration sphere; in such a situation, magnesium can be replaced with organic polyamines (spermine) [13].

The most important RNA-metal interactions for maintaining RNA structure and catalytic activity are found in the inner sphere [13, 16, 19]. The ligands involved in forming these interactions are commonly phosphoryl oxygens, heteroatoms attached to purine N7 position, keto groups in guanosine (O6) and uracil bases (O4), and the ribose 2’- OH group [13]. Generally three water groups are displaced in forming each inner sphere contact [13]. Interactions between the phosphoryl sulfur and Cd^{2+}/Mn^{2+} (thiophyllic metals) can be used to replace the phosphoryl oxygen and magnesium

interaction [13]. These metal-ion rescue experiments enable the identification of the specific metal-ligand interactions. The reason behind this is that hard metals (Mg^{2+}) from inner-sphere contacts with groups such as oxygen which are hard ligands, but with soft ligands such as sulfur inner sphere coordination is avoided [14, 18, 20]. Potassium is also a hard metal that preferentially binds oxygen [14, 18]. This specific interaction can be investigated by replacing K^+ (hard metal) with thallium (I) (soft metal), and oxygen (hard ligand) with sulfur (soft ligand) [13, 14].

In the double-stranded RNA structure metal ions prefer to bind in deep, narrow major groves [13, 16]. Tandem GU pairs (outer-sphere metal ion coordination), loop E binding motifs (a magnesium binuclear cluster), and A platform, (dehydrated K+ ion) are common specific metal-binding sites [13].

RNA needs a precise tertiary conformation to perform its specialized functions. Metal ion-RNA interactions are crucial in achieving the specific tertiary RNA conformation. Specific interactions that involve “outer-sphere interactions” and “site bound inner-sphere interactions” are mostly dominated by divalent ions (mostly Mg^{2+}) and non-specific interactions such as diffuse binding often involve monovalent ions (mostly K^+) [16]. Metal-ion rescue experiments are commonly utilized to identify the specific metal –ligand interactions in RNA [16].

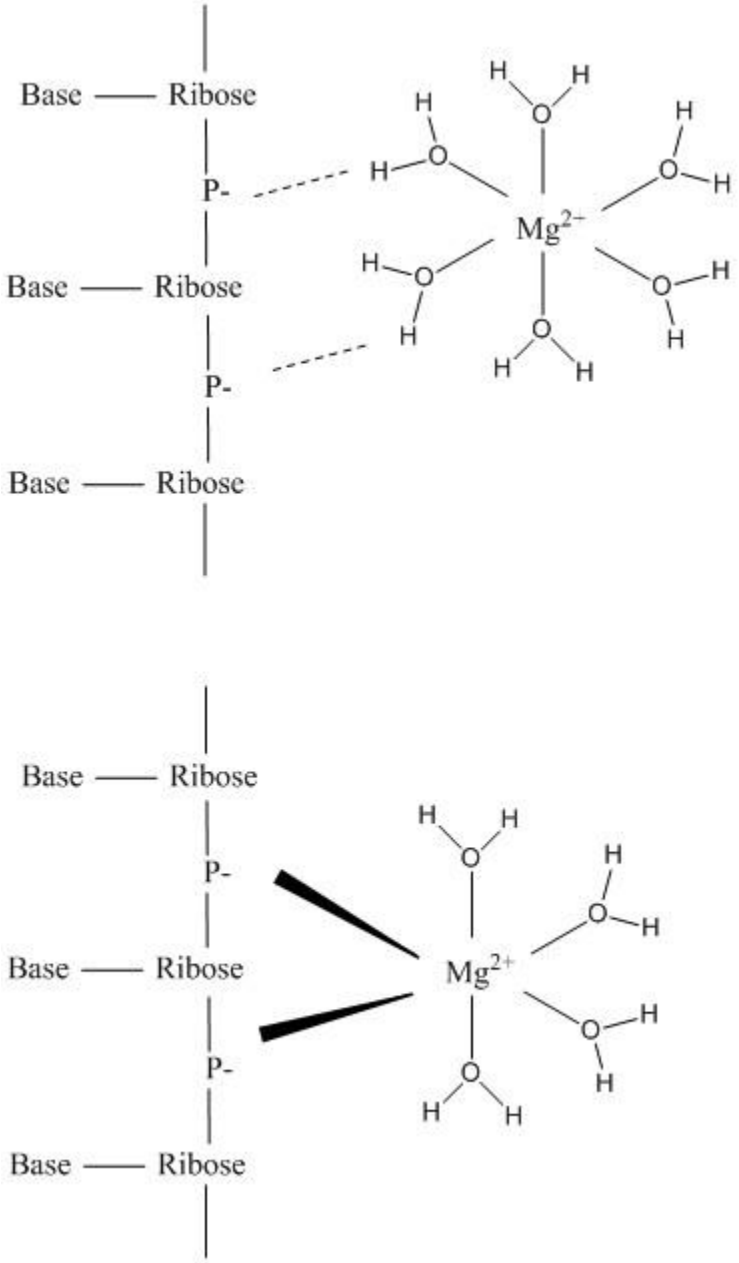

Figure 6: Modes of metal ion-RNA interactions. Outer-sphere interactions with RNA (top) and metal bound inner-sphere interactions are the most common types of interactions.

1.1.4 RNA-Protein Interactions Affect RNA Folding

In addition to the type of metal ions, ion density and concentration of metal ions, proteins also play an important role in RNA folding. Protein-RNA interactions can be specific or non-specific. These interactions can direct the RNA folding path towards a functional conformation [21, 22]. Based on the type of RNA-protein interactions and the folding mechanisms, RNA binding proteins can be divided into four main categories. They are RnA chaperones, specific RNA binding proteins, RNA helicases and RNA annealers.

RNA chaperones are proteins that unfold RNA non-specifically or avoid the formation of misfolded RNA structures [23, 24]. The nucleocapsid protein of the human immune deficiency virus (HIV) -1, NC, is one of the well-known RNA chaperones [25]. Specific RNA-binding proteins mediate RNA folding through specific RNA-protein interactions. RNA helicases use the energy from ATP hydrolysis to drive protein folding. It is known that about 1% of eukaryotic genes code for RNA helicases [26, 27]. Proteins belonging to the RNA annealer category direct the folding of unstructured RNA folding by molecular crowding [28]. These proteins bind to RNA, increasing the local concentration. This in turn increases the probability of RNA-RNA interactions facilitating RNA annealing**.**

1.1.5 RNA Folding Problem

The RNA folding problem was first described by D. Herschlag and explained the possibility of RNA misfolding in attempting to fold into the correct folding state [23]. In

his explanation, RNA folding presents two main problems. They can be listed as follows. First, RNA has the tendency to fold in to misfolded states that can get kinetically trapped. Secondly, difficulty in recognizing a single tertiary RNA structure that is thermodynamically stable [23]. RNA chaperones are known to solve the issue of a misfolded states kinetic traps through interacting with RNA non-specifically [23]. Specific RNA binding proteins aid to get through the difficulty in forming the thermodynamically stable state by interacting specifically to facilitate the formation of a specific tertiary folded state [23].

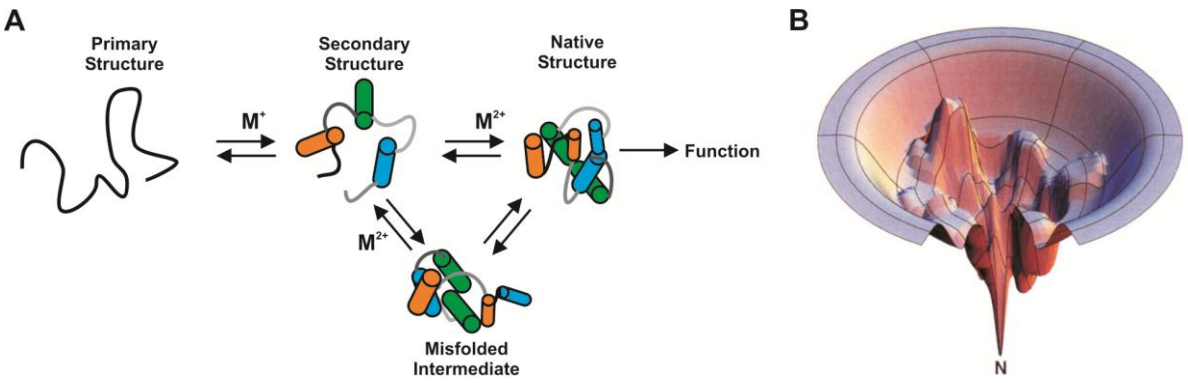

Figure 7: RNA folding pathway. (a) Primary sequence of RNA folds into more organized secondary structure with the aid of monovalent metal ions. This secondary structure further folds into the tertiary structure with the help of divalent metal ions. During the folding procedure, RNA may get trapped into misfolded intermediates. (b) The folding funnel representing the RNA folding landscape. This clearly shows the rugged energy surface and the multiple local energy minima with the global energy minima. This figure is modified from A.M. Pyle, (2002), J Biol Inorg Chem, 7, 679-90, and K. A Dill, H.S. Chan, 1997, NSMB, 4, 10-19, with permission.

1.1.6 RNA Genomes

Although all of the higher level organism bear DNA as the genetic information carrier, RNA genomes exist within viruses that are parasitic to organisms bearing a DNA genome [29]. These RNA viruses mostly present a single-stranded RNA genome, but there are a few instances where double-stranded RNA genomes have been observed. RNA viruses can be categorized in two main categories as follows: RNA viruses that replicate through a DNA provirus phase of the RNA (Retroviruses) and RNA viruses lacking a DNA phase in life cycle.

RNA in these retrovitruses acts as a template for reverse transcription upon infecting the host cell. Therefore retroviruses carry viral reverse transcriptase unlike other RNA viruses [29]. RNA viruses without a DNA phase on life cycle employ RNA replicase enzymes [29]. These viruses have a negative sense RNA, thus they utilize RNA-dependent RNA polymerases that make messenger RNA with genomic RNA as a template. A subset of RNA viruses lacking the DNA phase directly use their genomic RNA as the messenger RNA (picornavirus and togaviruses) [30].

1.2 Human Immunodeficiency Virus

Human immunodeficiency virus (HIV) belongs to the family of a retrovirus, which is an RNA genome bearing virus. It attacks the human immune system, which can result in acquired immune deficiency syndrome (AIDS) [31-33]. HIV is currently recognized as the main cause of death in Africa and the one of the leading causes of death worldwide [34, 35]. Current therapies against HIV target mainly two viral enzymes: reverse transcriptase [36] and protease [37]. Due to the rapid evolution of strains resistant to enzymatic inhibitors, new targets must be identified.

HIV belongs to the lentivirus family of retroviruses. There are two main types of HIV, HIV-1 and HIV-2. HIV-1 is identified as the most pathogenic and the predominant viral infection among both types. Virulence and the transmissibility of HIV-2 is significantly lower than HIV-1. Thus, HIV-1 and its subtypes are responsible for AIDS pandemic. HIV-1 is further divided into multiple subtypes as shown in Figure 8.

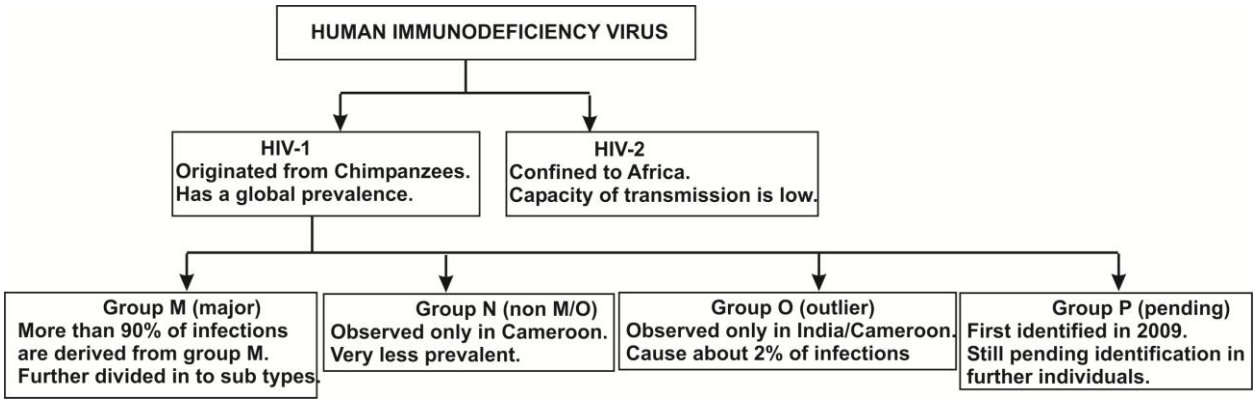

Figure 8: HIV virus subtypes

The genome of HIV consists of ribonucleic acid, which is a characteristic feature in retrovirus. There are nine open reading frames encoded by HIV-1 as illustrated in Figure 9 [38]. Three of these ORF's encode for gag, pol and env, which are essential structural protein components. These undergo proteolysis releasing the polyprotein complexes that are common to all known retroviruses. The gag polyprotein complex contain four proteins, namely matrix (MA), capsid (CA), nucleocapsid (NC) and p6. The two env proteins are surface/gp120 (SU) and transmembranal/gp41 (TM), which constitute the structural protein components that are required for the viral core and the outer viral membrane [38, 39]. The three proteins released from the pol polyprotein

complex, namely protease (PR), reverse transcriptase (RT) and intergrase (IN), are the three most important functional enzymes for the continuation of viral life cycle [38]. The proteins *tat, rev, vpu,* together with *vif, vpr* and *nef,* are known as accessory protein components [40, 41]. Protein tat and rev function as essential gene regulatory proteins. Vpu provides indirect assistance to HIV-1 viral assembly. The complete HIV genome is encoded in 9749 nucleotides (9.7 kb). Each virus contains two copies of genomic RNA.

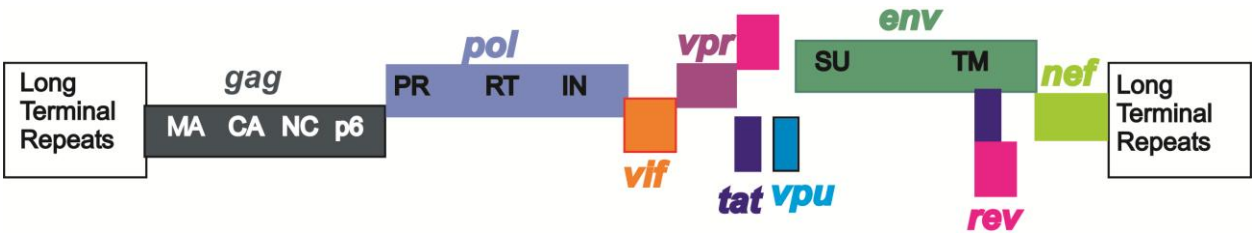

Figure 9: Open reading frame organization in HIV-1 genome

Non-covalently linked two copies of genomic RNA are encapsulated in the viral capsid, composed of capsid protein p24. The conical/bullet shaped capsid bears approximately 200 copies of the p24 protein. Viral proteins intergrase and reverse-transcriptase reside within the p24 coated viral capsid [38]. The viral envelope is the outer most core of the virus. It is composed of a lipid bi-layer. Viral structural protein *env* components (gp 41 and gp 120) protrude through the viral envelope [42, 43]. Later in the life cycle, these protrusions aid in attachment to new host cells during infection.

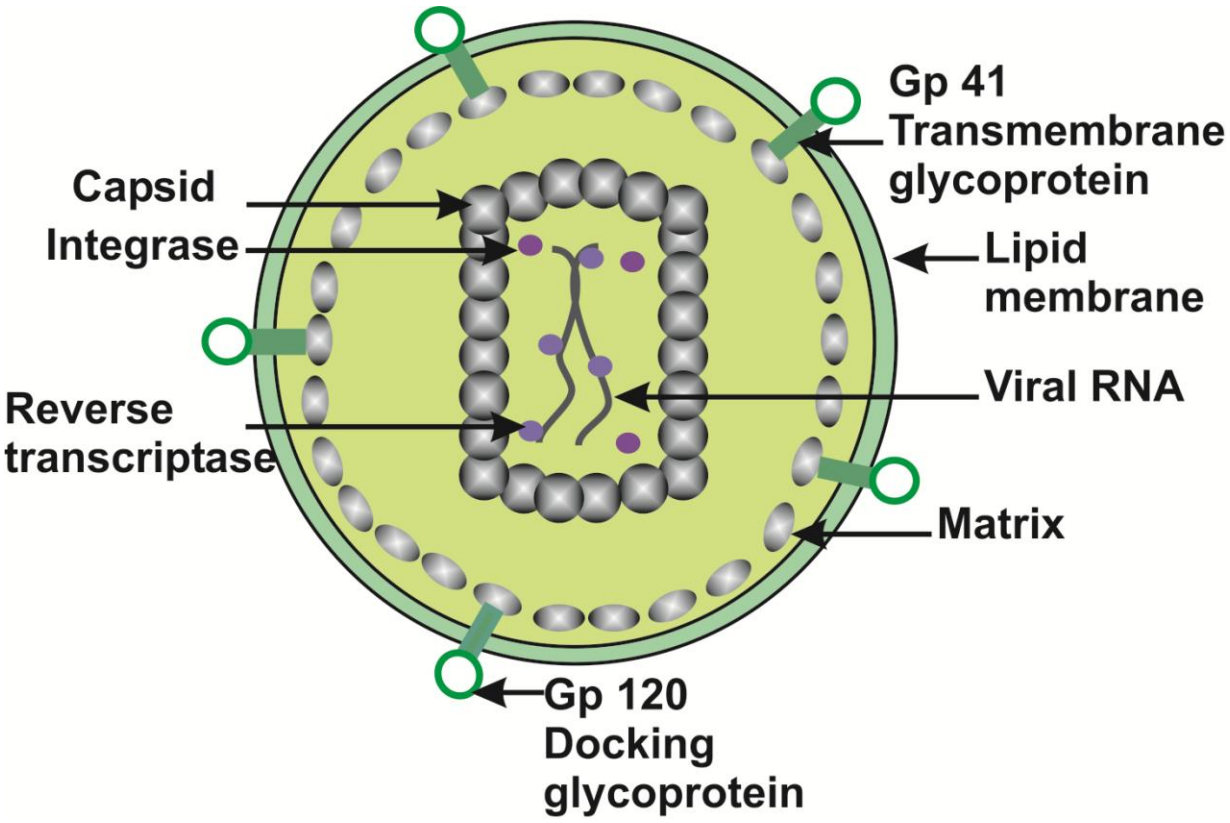

Figure 10: Structure of Human immunodeficiency virus

During HIV-1 replication, the genetic information encoded in the genome is reverse transcribed to the viral deoxyribonucleic acid (DNA), which is then transcribed into messenger RNA. During the reverse transcription, the two genomic RNA copies in the virus act as the template to produce the viral DNA through a reaction catalyzed by viral reverse transcriptase [38]. The nascent DNA is then incorporated into the host cell genome and a viral enzyme integrase is required for viral DNA integration into the host genome [44]. Multiple copies of viral genomic RNA are then produced during the host cell DNA replication. In order to complete the viral replication process, two full length

genomic RNA copies are co-packaged in a single virion [45]. Thus, the dimerization of the gRNA is a vital step in HIV-1 life cycle.

1.2.1 Genomic RNA Dimerization in HIV-1

HIV-1 genomic RNA dimerization is an essential step to gain the full inflectional capacity. Thus, dimerization step is an excellent candidate to be developed as a new drug target against HIV. Currently possibility of dimerization to be identified as a drug target is investigated under two approaches. First approach involves utilizing antisense nucleotides to target HIV-1 RNA dimerization [46]. The second approach is to target HIV-1 dimerization with aminoglycoside antibiotics based on the similarity of the HIV-1 RNA and the original aminoglycoside target which is bacterial 16S ribosomal A-site [47]. Thus detailed dissection of the RNA dimerization mechanism and identifying the role of each component in every steps of dimerization, as well as complete kinetic and thermodynamic characterization facilitate the development of new drugs.

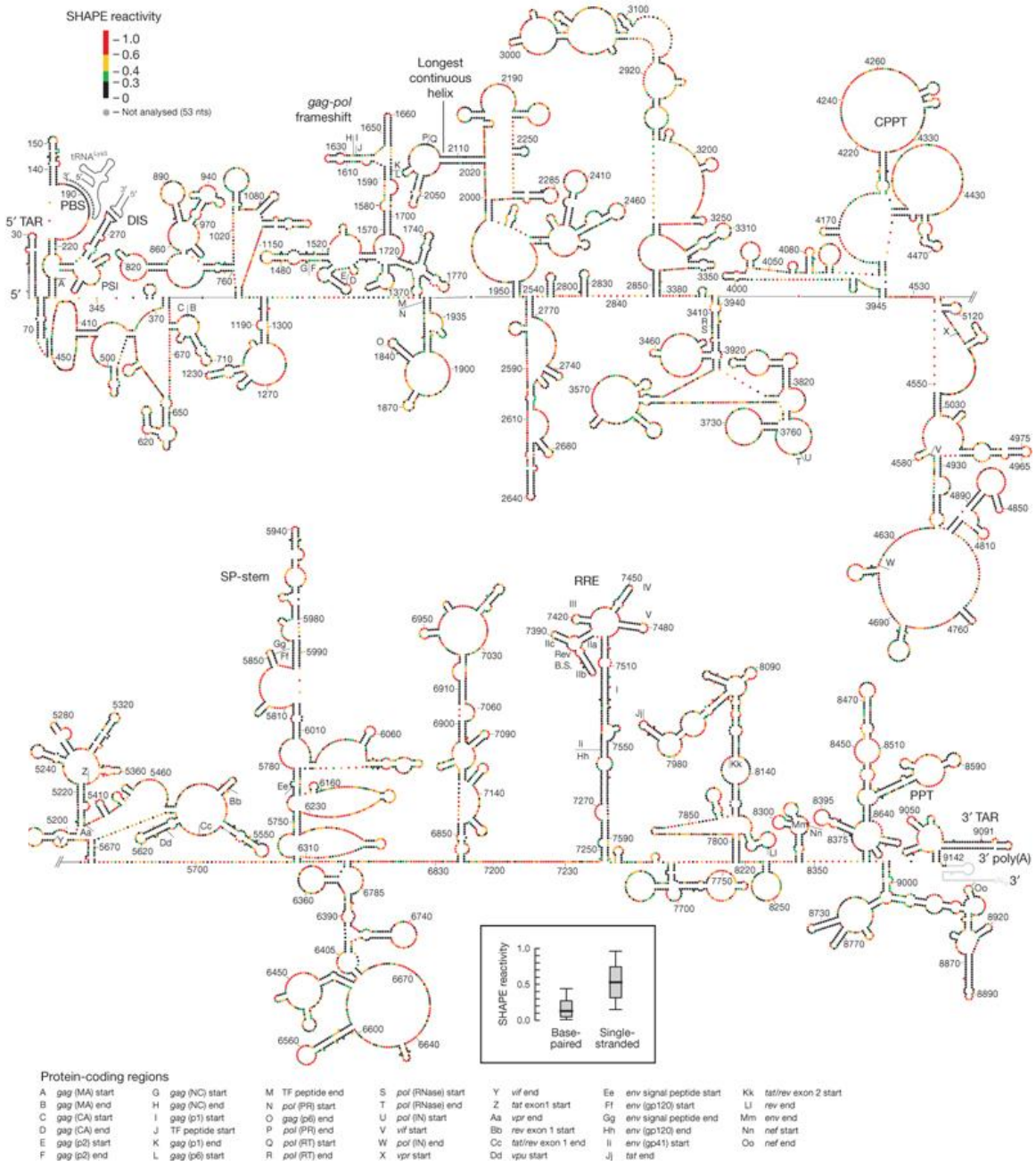

Figure 11: The 5’ and 3’ halves of the HIV-1 NL4-3 genome. The color code depicts the SHAPE reactivity of the nucleotides. This figure is used with the permission JM Watts *et al. Nature* 460, 711-716 (2009)

1.2.2 Dimerization Initiation Sequence

Mature HIV-1 viral particles (and all other retroviruses) bear two genomic RNA copies which are non-covalently linked through their 5' UTR. The region forming the interactions is known as dimer linkage structure (DLS) [48-52]. Studies on DLS have shown that a smaller portion forming a hairpin loop is responsible for dimerization initiation [31, 52]. A six-nucleotide palindromic sequence located at the apical loop region of this hairpin (dimerization initiation site, DIS) is highly conserved within HIV strains and is believed to play a key role in the initiation of RNA dimerization [53, 54].

In the loop, the DIS sequence is flanked by three purine bases. The sequence of the DIS and the three flanking purines presents small variations among different viral strains. Common DIS loop sequences in HIV-1 subtypes are shown in the following table [55].

Table 1: Common DIS sequences in different HIV-1 subtypes

DIS sequence	Flanking purines	Subtype
GCGCGC	5' - AA and 3'- A	Subtypes B, HIV-1 LAI, D and M
GUGCAC	5' - AG and 3'- A	Subtype A
	5' - AA and 3'- A	Subtype F, H
	5' - AA and 3'- U	Subtype C
	5' - GA and 3'- A	Subtype O

Mutations or alterations of the DIS prevents RNA dimerization and severely reduces the viral infectivity [48, 56-58]. Experiments performed with synthetic RNA fragments with DIS sequence have clearly illustrated that the initial kissing-loop dimer is subsequently stabilized by extension of intermolecular Watson-Crick base pairs as an extended duplex [59, 60]. This kissing-loop to extended duplex isomerization is strongly facilitated *in vitro* by incubation at high temperature (55°C), or at physiological temperature by the nucleocapsid protein (NC), a small, basic protein with two zinc finger domains [60-66]. Structural insights of the DIS kissing-loop and extended duplex forms have been provided by X-ray crystallography [67-69] and NMR [70-72]. This structural work led to the discovery of high-affinity ligands of both the kissing-loop and extended duplex forms [47, 73-75], thus highlighting the DIS as a potentially interesting new viral drug target. However, despite these progresses, little is known about the exact mechanism of isomerization.

According to the current model, two RNA hairpins first interact through the DIS, forming a kissing complex dimer. Activity of NC leads to the formation of the extended duplex dimer [31, 76] (Figure 12). The efficiency of extension in to the extended duplex dimer form depends on various factors including loop-loop interaction, temperature, ionic strength and viral nucleocapsid protein concentration [77-79]. A weak loop-loop interaction is deleterious to the formation of the KC dimer and therefore hinders the formation of the extended duplex. On the other hand, an extremely strong loop-loop interaction prevents formation of the extended duplex due to high stabilization of the kissing complex [47]. Thus the presence of moderate interactions at the KC is necessary for a successful progression towards the extended duplex.

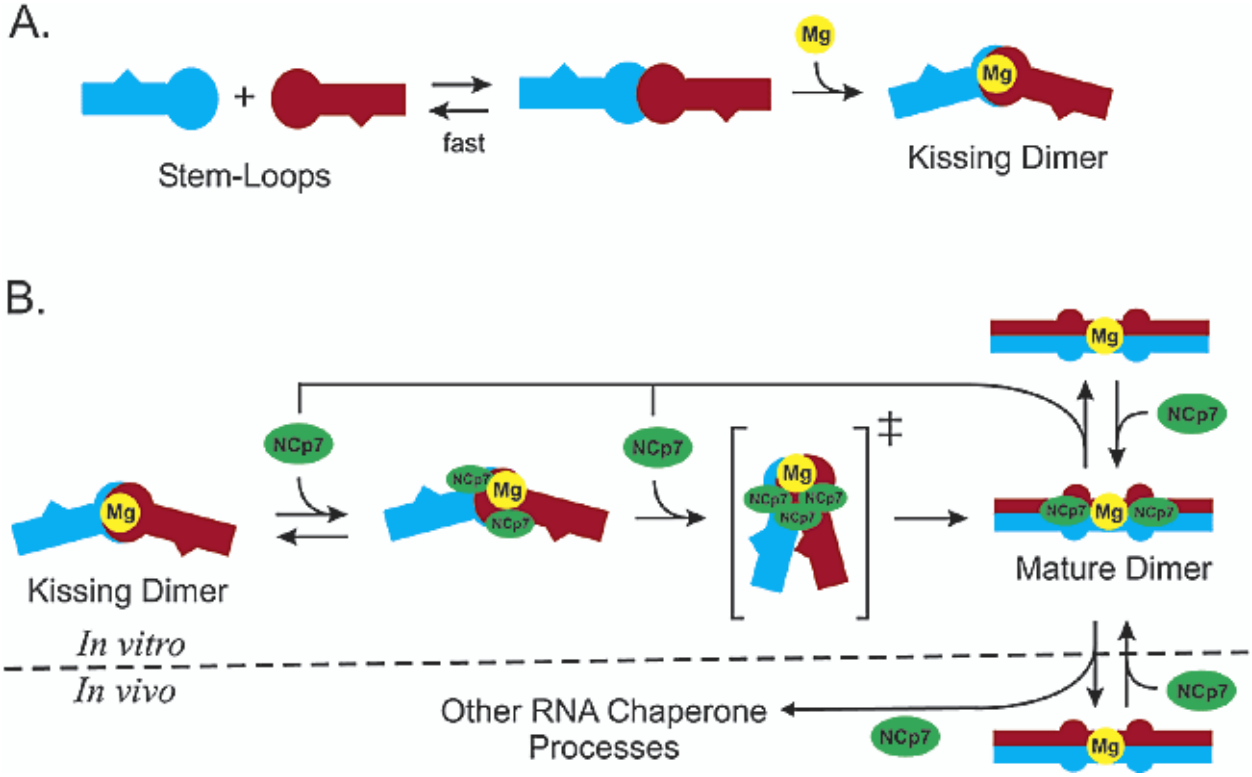

Figure 12: Model for DIS RNA dimerization proposed by. *Rist, M.J. and J.P. Marino,Biochemistry, 2002. 41(50): p. 14762-70*. (A) A collision complex initially forms via a loop−loop interaction between two individual DIS stem−loops that is stabilized by divalent metal ions. The DIS stem−loops are shown schematically in blue and red to allow them to be distinguishable from each other in the dimer complexes. The Intermediate loop−loop kissing complex is kinetically trapped in the presence of Mg^{2+} (represented by a yellow sphere) and does not spontaneously convert to the mature extended duplex form. (B) NCp7 (represented by a green sphere) acts by binding to the intermediate loop−loop kissing complex in an excess of the stoichiometry of two NCp7 molecules to one DIS dimer. Binding of NCp7 decreases the energy difference between the kissing intermediate and the transition-state in the refolding pathway, thereby lowering the activation energy and increasing the rate of structural isomerization. Dissociation of NCp7 from the mature extended duplex form of DIS allows it to act catalytically by rebinding and converting another intermediate kissing dimer. Since there is only one DIS interaction to act on in vivo, the biological significance of the observed recycling of NCp7 activity in vitro could be related to chaperone activities of NCp7 at different sites on the genomic RNA. This figure and the caption is used with the permission *Rist, M.J et al Biochemistry, 2002.*

1.2.3 Dimerization Dependence on Salt Concentration

It is been proposed that the negative charge shielding and specific metal-RNA interaction by positive ions stabilize the loop-loop structure and thus facilitate the formation of the extended duplex [55, 78]. According to the published KC crystal structure [69], a network of eight magnesium ions were identified and most importantly three of the Mg^{2+} closely associate with the purine bulges. It has been shown that in the presence of divalent metal ions, DIS KC is kinetically stabilized [64].

1.2.4 Role of HIV-1 Nucleocapsid (NC) Protein

HIV-1 nucleocapsid protein (NC) is a 55 amino acid small viral protein with two N terminal CCHC type zinc finger domains [80]. This protein plays few vital roles in viral life cycle. NC allows RNA assembly and efficient packaging into the new budding virion [38],chaperones the primer tRNA annealing to the primer binding site (PBS) and also aids vital nucleic acid rearrangements in reverse transcription [81, 82]. N on specific binding to single stranded nucleic acid is commonly observed with basic NC protein [61, 80, 83]. When bound to NC, RNA secondary structure becomes more compact and protected from nucleases activity. [38, 80]. In HIV-1, NC has the ability to show variable binding to HIV-1 RNA with different stoichiometric ratios at different functioning levels.

1.3 Aminoglycoside – Ribosomal A-site interaction

Ribosome is the target of aminoglycoside family antibiotics. These aminoglycosides inhibit protein synthesis by the introduction of codon misreading during

translation [84, 85].These effects on protein synthesis are a direct results of aminoglycoside binding to ribosomal RNA in aminoacyl-tRNA site (A-site) [86, 87] [88].

Aminoglycoside are a class of broad spectrum antibiotics with the basic amino sugars connected to a deoxystreptamine ring structure. All the aminoglycosides bear common functional groups first two ring moieties as shown in Figure 13. The interaction between aminoglycosides and the 16s ribosomal RNA A-site is well characterized with number of different methods such as nuclear magnetic resonance, surface Plasmon resonance, isothermal calorimetry and mass spectrometry [89-93]. This interaction is extensively studied due to its versatility in therapeutics.

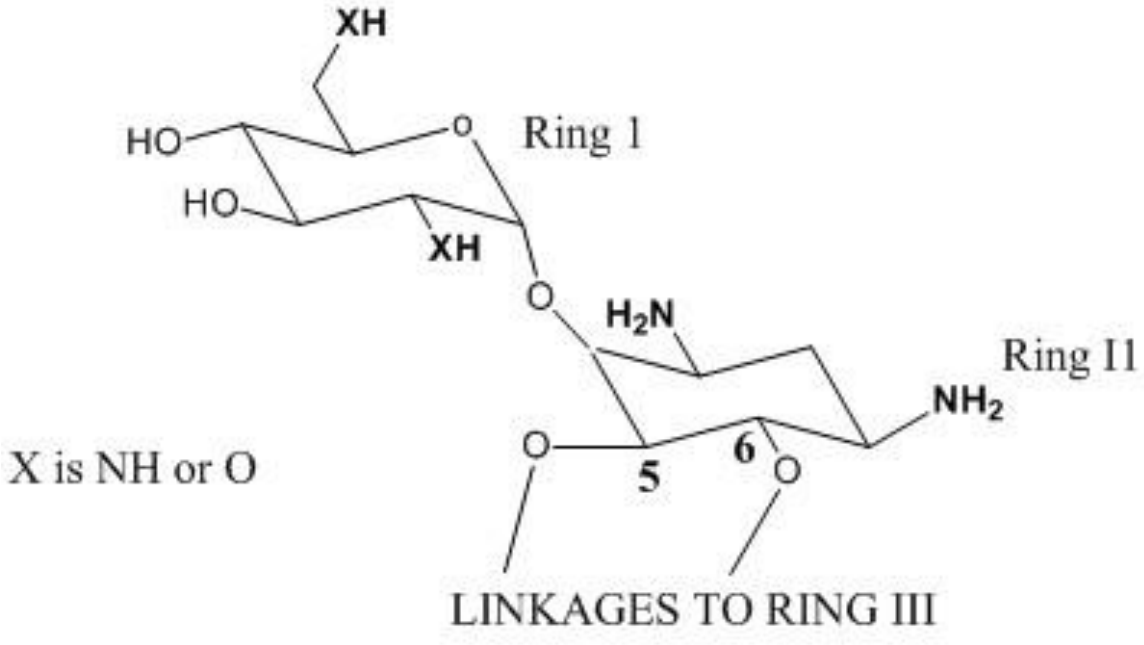

Figure 13: Common functional moiety in aminoglycosides. The different type sof aminoglycosides differ from substitutions at 5 or 6 position in ring 2. All A-site binding aminoglycosides bear this bicyclic moiety. Bold atoms identify the A-site binding positions common to all the aminoglycosides bearing the deoxystreptamine structure as shown above [86].

Neomycin family antibiotics (Ribostamycin, neomycin and paromomycin) bear substitutions at the 5th position (Figure 13 where as an additional ring is positioned at the 6 C in ring II in gentamycin and kanamycins.

Solution NMR and crystal structure information shows that neomycin family antibiotics bind to a major grove of A1493 - A1408 base pair pocket and also to a bulged out single adenosine residue 1492 [86]. The common functional groups found in all aminoglycosides (Figure 13 bold letters) have shown to interact with ribosomal A-site in a sequence specific manner [86]. It has been further observed that the further extensions to the common bicyclic moiety (at 5 or 6 th positions of ring II) interact with A-site non specifically. This facts were proven by neamine, the smallest aminoglycoside with only ring I and II also cause miscoding during protein sysnthesis upon binding to ribosomal A-site [94].

1.3.1 Ribosomal A-site and HIV-1 Dimerization Initiation Site

Comparison of primary and secondary structures of ribosomal RNA A-site and HIV- RNA has shown that prokaryotic ribosomal A-site has a sequence and structural resemblance to HIV-1 dimerization initiation sequence [50]. This similarity has been proven also in the kissing-loop and extended duplex conformations. The conformations formed by HIV-1 subtype F has been identified to show the maximum resemblance. Aminoglycoside antibiotics that bind to bacterial ribosomal A-site with a high-affinity have been reported to bind kissing complex and stabilize the KC form [50, 74]. Binding parameters of common aminoglycosides to HIV-1 DIS is summarized in table 2.

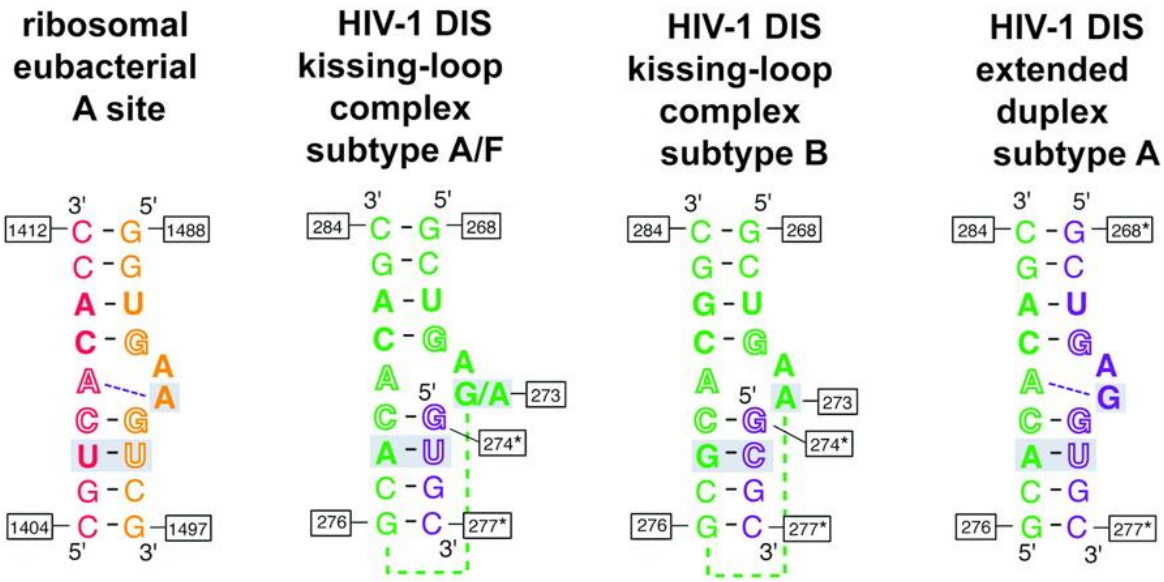

Figure 14: Illustration of the similarities between Ribosomal A-site and the HIV-1 dimerization initiation sequence as a kissing-loop complex or an extended duplex. Hollow letters stand for the conserved residues and bold letters stand for the residues defining the minimal A-site motif. Used with permission from Ennifar E, et al. J. Biol. Chem. 2003;278:2723-2730.

Table 2: Dissociation constants for aminoglycoside-DIS interaction [47]

Antibiotic	K_d (μM)
Paromomycin	1.3
Neamine	8.7
Apramycin	5.8
Neomycin	0.034
Lividomycin	0.032

1.4 Aims of This Dissertation Work

a. Characterization of HIV-1 genomic RNA dimerization at the single-molecule level.

The HIV-1 dimerization initiation sequence (DIS) is a conserved hairpin motif in the 5'-UTR of its RNA genome. DIS plays an important role in genome dimerization by formation of a "kissing complex" between two homologous hairpins. Here, we present a single-molecule fluorescence resonance energy transfer (smFRET) study of this kinetics. Our data show the real-time formation and dissociation dynamics of individual kissing complexes, as well as an extended duplex formation. Metal ion analysis showed that the dimerization process in driven by Mg^{2+} ions whose binding to the first stable intermediate (kissing complex), results in a stabilization of 5 ± 3 kcal/mol. More importantly, the single-molecule trajectories reveal the presence of a previously unobserved bent intermediate required for extended duplex formation. High Mg2+ ion concentrations stabilize this intermediate, whereas K+ ions destabilize it. DIS mutagenesis indicates that intermediate formation requires a Mg^{2+}-dependent conformational change of the purines flanking the loop nucleotides.

b. Unraveling the role of HIV-1 nucleocapsid protein on dimerization at the single-molecule level.

In the second main section of this research work, the effect of HIV-1 nucleocapsid protein (NC) on the kinetics and thermodynamic properties of the genomic RNA dimerization is studied. I have identified that binding of NC to individual monomer RNA hairpins favors flipping out of key adenine residues within the hairpin loop. These local

changes increase the rate of KC formation. Additionally, the results also show that binding of NC hinders dissociation of the KC back to monomeric RNA. Furthermore, experiments in presence of NC support the conclusion that the newly observed bent dimer conformation is an on path obligatory intermediate.

c. Exploring HIV-1 dimerization initiation sequence as a potential drug target.

Following the complete characterization of the dimerization pathway, the next goal of the project was to explore the potential of this DIS as a drug target. I investigated the effect of the Neomycin family of aminoglycosides on the proposed HIV dimerization pathway. This work on the DIS-aminoglycoside interaction shows that the aminoglycosides binding to initial monomeric RNA mimics the effect of Mg^{2+} with a high specificity. Upon binding of aminoglycosides with extended structures (ring 3, 4 and 5), the unpaired adenines in the kissing-loop interaction are locked in an inter-helical stacked conformation reducing the flexibility of the monomers at the kissing interface through non-specific interaction. As a result, the bulky, neomycin family aminoglycosides stall the dimerization at the kissing complex form. The mechanistic insights gained from these experiments represent significant progress towards understanding the HIV-1 dimerization process as w ell as the identification and development of DIS-targeting molecules.

Chapter 2

Materials and Methods

2.1 Detection of RNA Dynamics at Single-Molecule Level

Single-molecule techniques allow the visualization of single biomolecules, their dynamics and structural transitions under given conditions. Studies at the single-molecule level reveal population heterogeneity that is generally hidden in bulk ensemble averaged experiments [95]. Additionally, studies at the single-molecule level can also be used to observe folding intermediates that do not accumulate over time, explore multiple folding pathways and to obtain kinetic information from equilibrium experiments. Optical and magnetic tweezers [96, 97] fluorescence correlation spectroscopy (FCS) [98], confocal microscopy [99] atomic force microscopy, and total internal reflection fluorescence microscopy (TIRFM) [100] are some of the most widely used single-molecule approaches in biophysics. In this dissertation work fluorescence resonance energy transfer is utilized for single-molecule detection and TIRFM is employed as the methods of fluorescence excitation.

2.1.1 Fluorescence Resonance Energy Transfer (FRET)

FRET was first observed by Perrin at the beginning of the century [101]. But the theories underlining fluorescence resonance energy transfer were illustrated by Theodor Forster (1910-1974) [101]. Basics of FRET can be illustrated with a Jablonski diagram (Figure 15 and Figure 16). Excitation of a ground state fluorophore will result in the electrons in the ground state to change to excited states. Then the excited state electrons will release its energy to settle back into the ground state through different mechanisms. The relaxation back to ground state can occur as radiative fluorescent emission, internal conversions or inter system crossing resulting in phosphorescence (Figure 15).

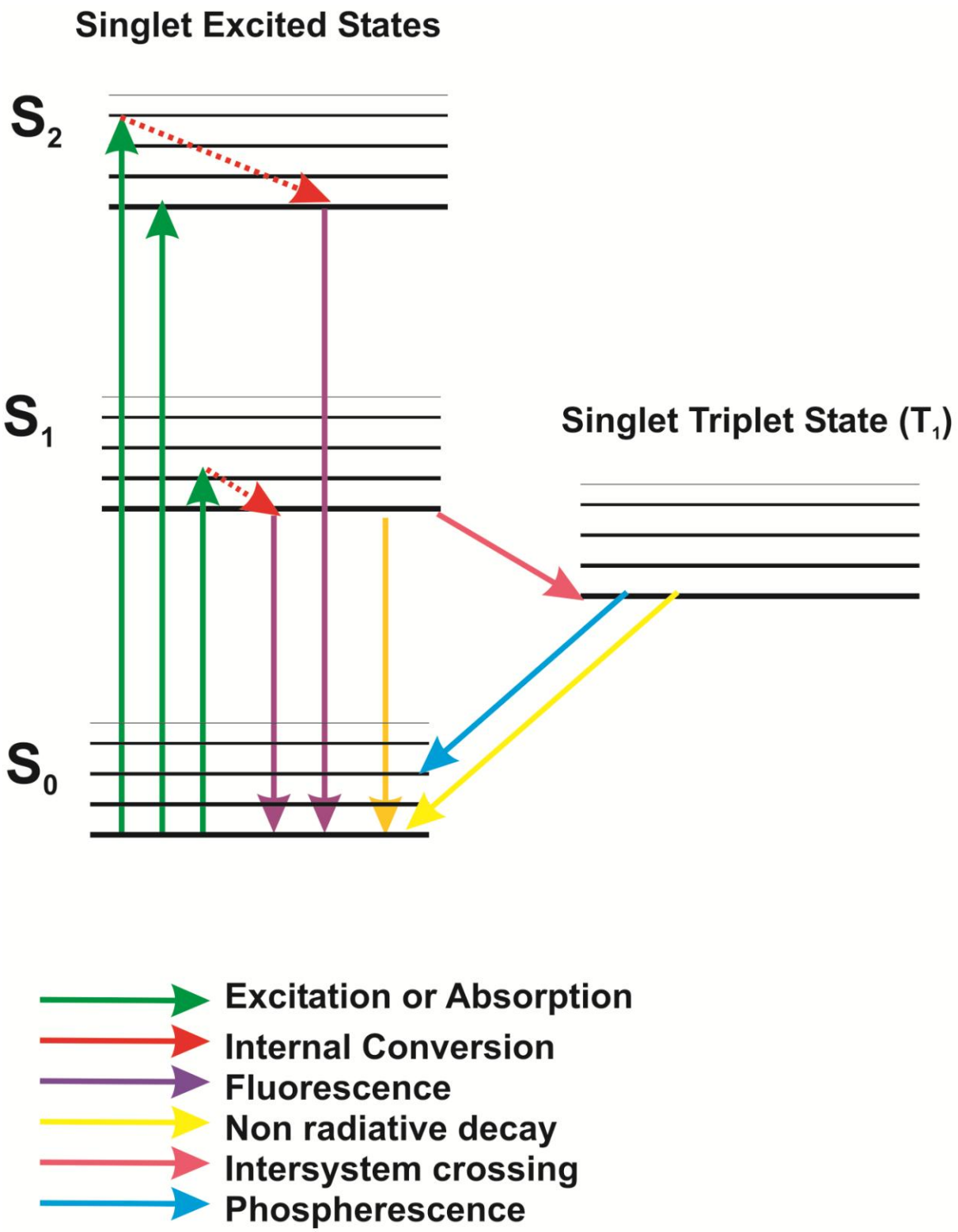

Figure 15: Jablonski Diagram illustrating the basics of fluorescence

FRET is a result of interaction between the excited electronic states of two fluorophores in a distance-dependent manner. During the interaction between the fluorophores, excitation is transferred between two fluorophores without emission of photos. Thus, fluorescence resonance energy transfer is defined as a non-radiative energy transfer between two chromophores identified as an excited donor fluorophore and acceptor fluorophore. Energy transfer takes place as a result of intermolecular long range dipole-dipole coupling. Observed energy transfer during FRET depends on the inverse of the sixth power of the distance between two participating fluorophores. Therefore FRET has become a versatile technique to measure distances in low nanometer range (<10nm) enabling the use of FRET in biological system studies [101]. Multiple requirements have to be satisfied in order to observe successful FRET. First, the emission spectrum of the donor fluorophore should be overlapped with the excitation spectrum of the acceptor fluorophore. Additionally, the transition dipoles of the donor and the acceptor fluorophores should posses rapid rotational freedom to a certain extent. The efficiency of the energy transfer during FRET mainly depends on the distance between the two fluorophores.

The distance dependence of the FRET efficiency is given by the Föster equation,

$$E_{FRET} = R^6/ (R^6 + R_o^6), \qquad \text{Equation 1}$$

where R is the distance between the donor and the acceptor and R_0 is the Fosters radius, the distance at which the efficiency of energy transfer is 50%. R_0 depends on several factors such as the quantum yield of the donor, overlap between the emission spectrum of the donor and the absorption spectrum of the acceptor, and the orientation

factors, which accounts for the relative orientation between the dipole moment of the two dyes. Efficiency of energy transfer is a direct measure of the amount of energy quanta transferred from donor fluorophore to the acceptor fluorophore [102].

In other words, E_{FRET} can be defined as the quantum yield of the fluorescence energy transfer from donor fluorophore to the acceptor.

E_{FRET} = (Amount of quanta of energy transferred from Donor to Acceptor)/

(Amount of quanta absorbed by donor)

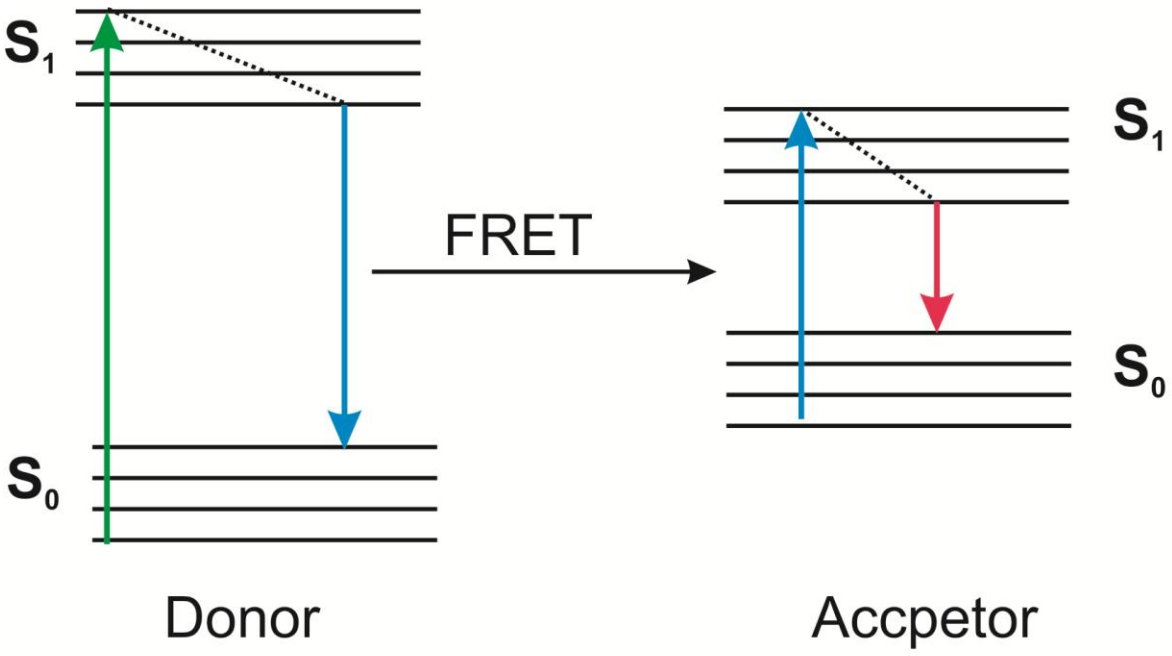

Figure 16: Jablonski diagram illustrating the principle of FRET between a donor fluorophore and an acceptor fluorophore

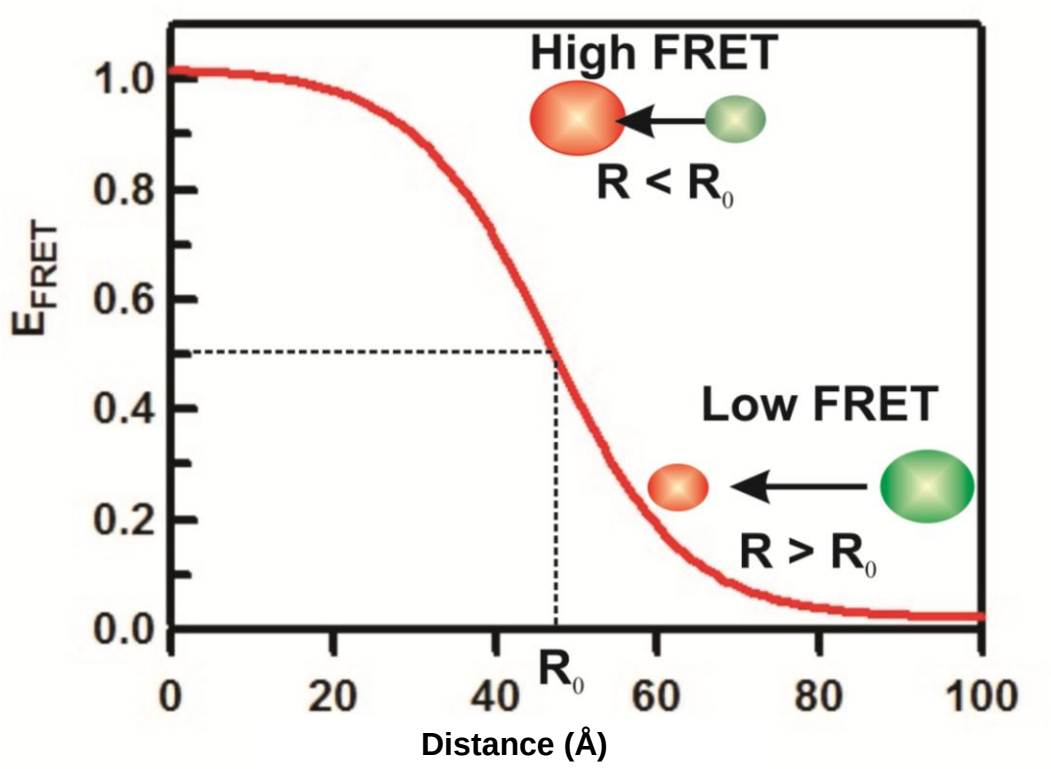

Figure 17: Relationship between FRET efficiency and inter fluorophore distance. R_0 or the Foster radius is the distance at which 50% energy transfer from donor fluorophore to the acceptor fluorophore is observed.

Value of R0 depends on several factors as shown by the following equation [101].

$$R_0^6 = 8.78\times10^{-5}\frac{\kappa^2\phi_D J}{n^4}A^0 6 \qquad \text{Equation 2}$$

According to the equation 2, value of R_0^6 depends on κ^2 (Orientation factor), ϕ_D (Quantum yield of the donor), J (Overlapping integral), n (Refractive index of the medium). Based on the orientation of the dipoles of the donor and acceptor fluorophores, the value of the orientation factor will vary between 0-4. In biophysical FRET experiments where fluorophores are attached to RNA/DNA or protein with long

linkers, a random orientation of fluorophore dipoles is expected due to fast tumbling. Therefore, in these experiments, the value of the orientation factor is 2/3 assuming the random orientation of the donor and the acceptor fluorophore dipole over all the possible dipole orientations. ϕ_D or the quantum yield of the donor is variable based on the type of the donor fluorophore used in the experiments. For the cy3 fluorophore that is used for the work discussed in this thesis, quantum yield is 0.28. Refractive index (n) for aqueous medium is ~ 1.3. Overlapping integral (J) is the amount of overlap observed between the emission spectrum of the donor fluorophore.

$$J = \int F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda \qquad \text{Equation 3}$$

In the above expression, normalized emission of donor is denoted with $F_D(\lambda)$ and normalized excitation spectrum of the acceptor is denoted with $\varepsilon_A(\lambda)$.

2.1.2 Total Internal Reflection Microscopy Based Single-Molecule FRET

Two main TIRF microscopy approaches are widely used in the field. Namely they are, Prism-based TIRF and Objective TIRF [95]. In the work discussed in this thesis, a prism-based TIRF setup mounted on an inverted fluorescent microscope (IX-71, Olympus, Center Valley, PA)) is used for obtaining smFRET data. In the TIRFM setup, the excitation volume is limited to a thin layer located between the sample solution and the microscope slide. The RNA solution to be studied is placed between a well cleaned quartz slide (Finkenbeiner, Waltham, MA) and a cover slip (VWR, West Chester, PA). The slide is mounted on top of the inverted fluorescent microscope. To achieve this minimized excitation volume, in the prism-based TIRF microscopy set up, incident laser beam (532 nm, 3 mW, CrystaLaser GCL-532-L, Reno, NV) is set to be incident on the

glass surface at an angle greater than the critical angle (θ_c). The experimental critical angle is obtained using Snells law as shown in equation 4.

$$\theta_C = \sin\text{-}1\frac{n_{sol}}{n_q} \qquad \textbf{Equation 4}$$

In Snells law, n_{sol} is the refractive index of the experimental solution and n_q is the refractive index of the quartz.

When the incident angle of the laser beam is more than the critical angle, the laser beam is subjected to total internal reflection on the quartz surface avoiding the penetration into the experimental sample. This arrangement will let only an evanescent wave to penetrate to the solution contacting biomolecules with fluorophores. The generated evanescent wave is capable of exciting donor fluorophores that is within 100-150 nm range. Thus, using TIRF, background fluorescence is limited, allowing the detection and tracking of single biomolecules over long periods of time. Resulting emission from acceptor fluorophores are collected by the objective lens in the TIRFM set up and is directed towards the ultra high quantum yield charged coupled device (CCD) camera (Ixon+, DV-897E, Andor, South Windsor, CT). Scattered light from the excitation of fluorophores is avoided by a bypass filter (E550LP, Chroma, Rockingham, VT). The CCD camera chip set is cooled to -80° C. In order to minimize the background fluorescent noise, the signal is amplified by the maximum possible electron multiplication value. Both these settings give rise to a higher signal to noise ratio. The maximum achievable frame resolution in the smFRET TIRF setup used for this

dissertation work is 30ms per frame. The CCD detected signal is converted into a digital signal and is then transferred to the computer for the next steps in the analysis.

2.1.3 Separation of Colors for smFRET Analysis

Unlike bulk FRET experiments, smFRET experiments require the separation of donor and acceptor emission. Thus, in smFRET experiments the emission from the two types of fluorophores should be measures separately. The separation of the emission channels is achieved by using a set of dichroic mirrors and lenses in the microscope set up.

The fluorescent emission from both acceptor and donor fluorophores is collected by the objective and is directed to the microscope side port. The width of the image of this emission is controlled with a variable width slit (home built). The slit is positioned correctly in the image plane, which is about 75 mm distance from the side port. The correct positioning of the slit avoids image distortions. The incoming emissions from both fluorophores are then sent through the first dichroic mirror (DM1, 635DCXR, Chroma, Rockingham, VT). DM1 transmits only the acceptor emission and reflects the donor emission. Next, the two resulting images are recollected together side-by side, avoiding the overlap with the second dichroic mirror (identical to DM1). The resulting emission light beams are then received by the two different halves of the CCD chip, allowing the separate and independent measurement of both donor and acceptor emission.

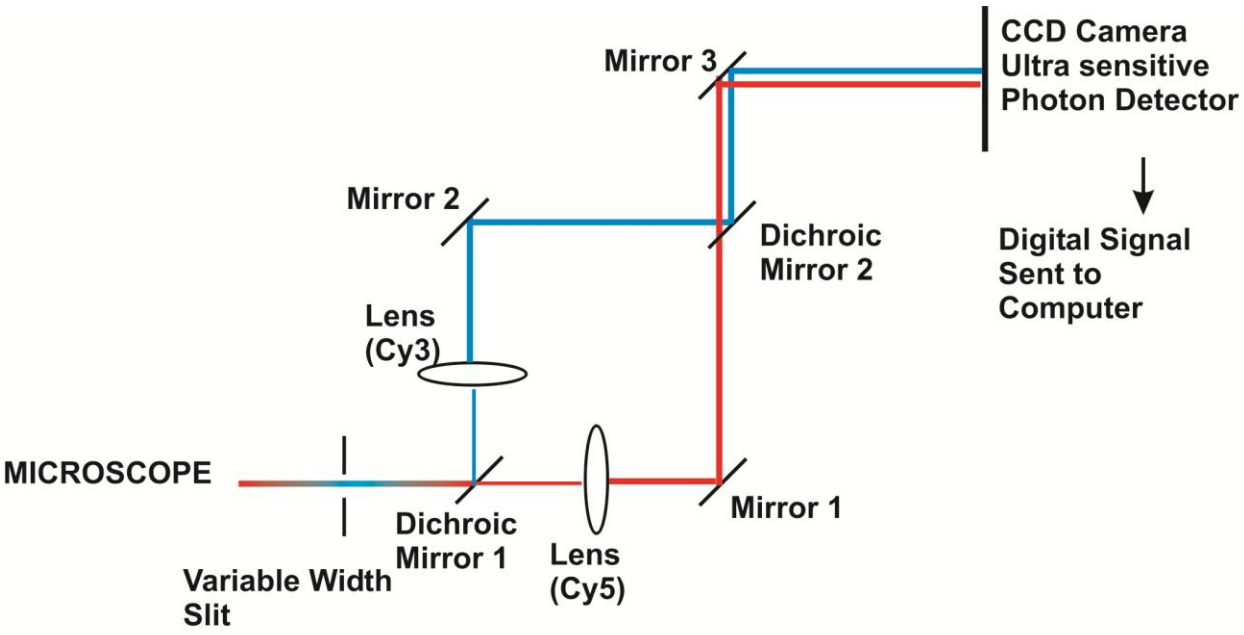

Figure 18: Schematic plot of the optical set up used to transmit the intensity signal to the CCD camera

2.1.4 Quartz Slide Preparation for SmFRET

Commercially purchased quartz slides have to be modified in order to build a microfluidic channel that contains the sample during smFRET experiments. The first step in the slide manipulation is drilling two holes in the quartz slides (Figure 19). For the slides utilized for the work discussed in this thesis, a hand drill (Dremel 300-N, Racine, WI) which was held by a workstation (Dremel 220-01, Racine, WI) was used for slide drilling process. Diamond drill bits (1.0 mm diameter, Kingsley North, Norway, MI) were used as the drilling tip.

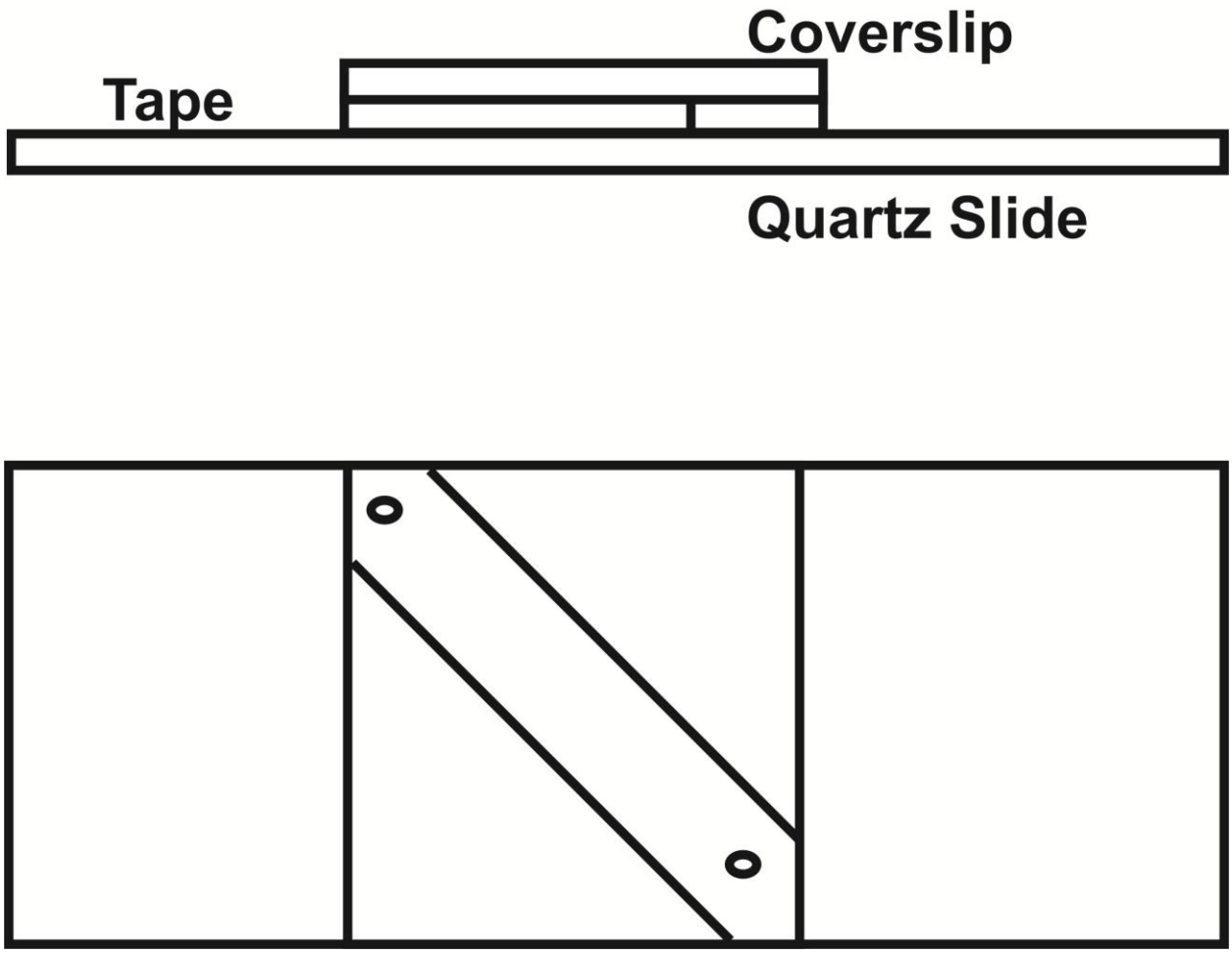

Figure 19: Quartz slide with the assembled microfluidic channel. The channel is set to be approximately 20 mm long, 150-200 µm thick. The cover slip is attached by using two layers of double sided tape and sealed with epoxy [103].

The quartz slides with the drilled holes are then subjected to an extensive cleaning procedure before sample injection. Initially the slides are soaked overnight in Contrad® soap solution for the removal of surface impurities. After completion of overnight soaking, slides are then cleaned with Alconox® soap paste. Rubbing with Alconox soap paste allows the removal of surface attached impurities. After multiple rounds of cleaning with Alconox soap paste, slides are washed with distilled water for multiple rounds to get rid of any remaining Alconox residues on the slides. The slides are next

subjected to boiling with distilled water for further removal of remaining non-organic impurities. Upon boiling, the slides are rinsed with ethanol to remove remaining organic residues. Slides are next boiled with acid piranha (conc. H_2SO_4 (> 51% acid): 30% H_2O_2, 3:1 ratio, respectively) or the basic piranha (100mL distilled water, 20 mL 30% NH_4OH and 20 mL 30% H_2O_2) for approximately 20 minutes. Finally the slides are washed thoroughly with double distilled H_2O and dried with a bunson burner flame. Special care is taken to avoid the formation of dry depositions on the slide surface.

2.2 RNA Purification and Labeling

RNA strands should bear a pair of FRET fluorophores as well as a biotin tag for surface immobilization in order to be successfully used in smFRET experiments. RNA oligos were purchased from the Keck Foundation Resource Laboratory at the Yale University School of Medicine with a 5'- biotin (B) and 3'-Cy3 (DIS 1 WT and DIS 1; Table S1) or with a 5'-C7 amino linker (DIS 2 WT and DIS 2; Table 1). Cy5-NHS esters for RNA labeling were purchased from GE healthcare. DIS 1 and DIS 2 RNA contains single base mutation in the loop which retains the base paring and does not affect the RNA binding affinities [104].

Table 3: Sequences of RNA constructs used in this study

Name	Description	Sequence
DIS 1 WT	Wild type RNA with the donor fluorophore	5' Cy3 –UU GCU GAA GUG GAC ACA GCA AUA U - Biotin 3'
DIS 1	Mutant type RNA with the	5' Cy3 –UU GCU GAA GUC CAC ACA GCA AUA U -

	donor fluorophore	Biotin 3’
DIS 2 WT	Wild type RNA with the acceptor fluorophore	5’ UU GCU GAA GUG CAC ACA GCA –Cy5 3’
DIS 2	Mutant type RNA with the acceptor fluorophore	5’ UU GCU GAA GUG CAC ACA GCAA –Cy5 3’
DIS 1 C	Mutant type RNA with the additional A272C donor fluorophore	5’ Cy3 –UU GCU GCA GUC CAC ACA GCA AUA U - Biotin 3’
DIS 2 C	Mutant type RNA with the additional A272C acceptor fluorophore	5’ UU GCU GCA GUG CAC ACA GCAA –Cy5 3’
DIS 2 S	Mutant type RNA with the stem terminal three base pairs flipped with acceptor fluorophore	5’ AA CCU GAA GUG CAC ACA GGU U –Cy5 3’

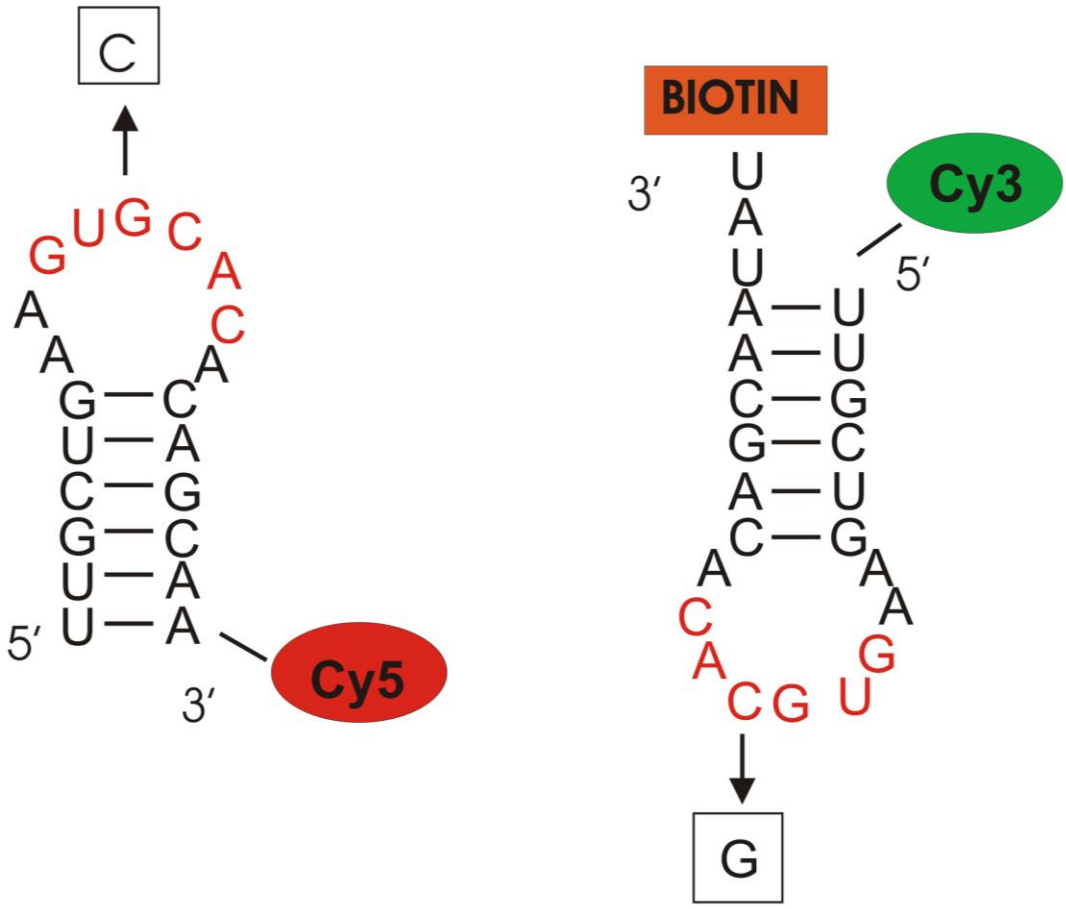

Figure 20: HIV-1 subtype F DIS sequence illustrating the single base mutation in the loop which retains the base paring and does not affect the RNA binding affinities.

RNA Deprotection

The 2'-hydroxyl protection groups were deprotected according to the manufacturer's protocol. Silyl ethers TBDMS (*tert*-butyldimethylsilyl) or TOM (tri-*iso*-propylsilyloxymethyl) are used for the protection of 2'-hydroxy function in RNA strands during RNA synthesis. Selection of TBDMS or TOM as the protecting agent is based on

the length of the RNA. TOM has a high coupling efficiency with RNA 2' groups compared to TBDMS. Thus, for RNA with at least 55 nucleotides in length or longer, TOM is preferentially used as the protecting agent. For RNAs with less than 55 nucleotides, TBDMS is used as the protecting agent due to appreciable cost and performance.

Initially, RNA sample is suspended in 100 µL of DMSO (Dimethylsulfoxide, anhydrous, 99.9% (Aldrich 27,685-5)). Once RNA oligo is fully dissolved, 125 µL of TEA.3HF (Triethylamine trihydrofluoride, 98% (Aldrich 34,464-8)) was added to the RNA vial, mixed well and heated to 65°C for 2.5 hours. Once the deprotection reaction is completed, RNA sample was cooled briefly in the freezer (-20 C) and continued for desalting by precipitation.

Desalting by Precipitation

25 µL of 3M sodium acetate (pH 5.2) in RNase free water was added to the RNA vial and mixed by vortexing thoroughly. In order to facilitate precipitation, 1 mL butanol and thorough vortexing was repeated. RNA sample was kept in -80℃ for 3 hours or -20℃ overnight to enhance precipitating of desalted RNA. Once the refrigeration is completed, RNA containing vial was centrifuged for 15 minutes at 13400 rpm to further facilitate suspending RNA. Remaining supernatant was removed and the precipitate was rinsed thrice with 100% ethanol, while repeating the centrifugation in between. Finally the precipitated RNA is dried under high vacuum speed vacuum to remove remaining butanol and ethanol traces. The dried RNA precipitate is resuspened in water and was directed for purification by PAGE.

The RNAs were purified by denaturing gel electrophoresis (20% polyacrylamide and 8 M urea) and the gel bands containing RNA was subjected to diffusion elution against elution buffer (0.5 M NH_4OAc and 0.1 mM EDTA) overnight at 4 °C, followed by chloroform extraction, ethanol precipitation and C8 reverse-phase HPLC as described.

Chloroform Extraction

A liquid content from the elution column was equilibrated with 4 mL of chloroform. The two liquid phases were shaken well to enhance the extraction of remaining organic contents in to chloroform and was centrifuged to facilitate separation of the two liquid phases. The aqueous layer containing the RNA sample was carefully extracted.

Ethanol Precipitation

Sodium acetate equivalent to $1/10^{th}$ of sample volume and 100% ethanol equivalent to 2-2.5 times of sample volume were added to the aqueous RNA sample. The vial containing the solution was kept in -80℃ for 3 hours or -20℃ overnight to enhance precipitating of desalted RNA. The precipitate was obtained by centrifugation and multiple ethanol washing steps were repeated, followed by drying in the speed vac.

Fluorophore Labeling

The C7 amino linker in DIS 2 WT and DIS 2 were labeled with Cy5-NHS in labeling buffer (100 mM Na_2CO_3, pH 8.5) overnight at room temperature. Labeled RNA was purified by ethanol precipitation and reverse-phase HPLC as described [105]. RNA concentrations and labeling efficiencies of all samples were calculated by UV-Vis absorbance at 260 nm, 550 nm and 650 nm.

2.3 Proteins used for the Study

2.3.1 HIV-1 Nucleocapsid Protein

The purified NC protein is received from Professor Karin Musier-Forsyth lab, Department of Chemistry and Biochemistry, The Ohio State University, Columbus. The purification protocol is published in their recent publications [106].

2.3.2 HPVHHYQ peptide

The synthesized HPVHHYQ peptide is received as a generous gift from Professor Christine Chow lab, Department of Chemistry, Wayne State University, Detroit, MI. The purification protocol is published in their recent publication [107].

2.4 Single-Molecule FRET

Single-molecule experiments were performed as described [105, 108]. Briefly, RNA strands DIS 1 and DIS 2 were diluted to 25 pM and 30 nM, respectively in standard buffer (20 mM Tris-HCl, pH 7.5, 150 mM KCl, and 5 mM $MgCl_2$ in saturating trolox). DIS 1, (25 pM) was heated at 90 °C for 45 s before flash cooling on ice. This flash cooling at low concentration was done to ensure formation of duplex structures. The DIS 1 RNA was immobilized to a streptavidin-coated quartz slide surface via a biotin-streptavidin bridge yielding a surface density of ~0.1 molecules/μm^2. Next, the DIS 2 RNA was introduced in solution onto a channel in the quartz slide. Measurements were obtained under variable [Mg^{2+}] and [K^+] (0.001 – 20 mM and 0-1 M respectively) at room temperature with an oxygen-scavenging system (OSS) consisting of 5 mM 3,4-dihydroxybenzoic acid (PCA) and 0.1 µM protocatechaute 3,4-dioxygenase(PCD) to reduce photo-bleaching. We then constructed histograms and calculated dwell times for

each dimerization event to determine the rate constants, as described [103]. A cutoff FRET value of 0.25 was used to distinguish between monomer and kissing complex (KC) dimer state. A cutoff of 0.75 FRET was used to distinguish between KC dimer state and extended duplex. Dwell time histograms were fit to single or double exponential functions depending on the distribution of dwell time data. Dissociation constants for Mg^{2+} binding were obtained by fitting the fraction of time spent in the KC form to the modified Hill equation

$$f(x) = f_0 + (f_{max} - f_0)\left(\frac{x^n}{K_D^n + x^n}\right)$$

Equation 5

Where f_0 is the initial fraction molecules in KC dimer state, f_{max} is the final fraction of molecules in the KC state, and x is the concentration of K^+.

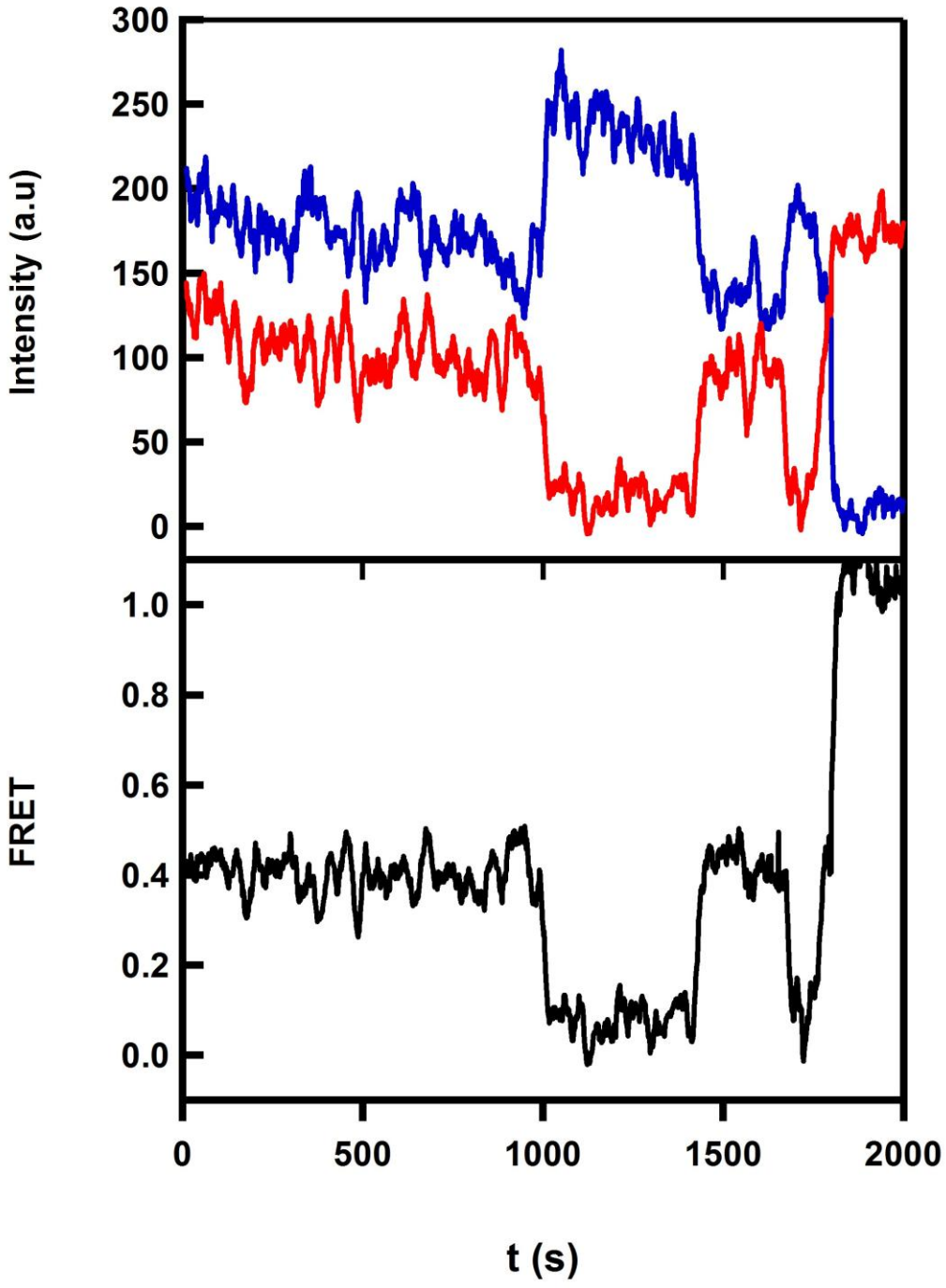

Figure 21: Variation of donor (green) and acceptor (red) intensities of two RNA hairpins dwelling between kissing complex and monomer forms and ultimately forming extended RNA duplex. FRET trajectory corresponding to the observed intensity variation is also shown.

2.4.1 Kinetic Analysis with SmFRET

smFRET data can be directly used to analyze the kinetics of a system of interest. Initially the dwell times in the bound (τ_{on}) and unbound (τ_{off}) states should be measured. Dwell time measurement should be obtained from a few hundred of molecule trajectories and the collected dwell times are represented in dwell time histograms. In a simple kinetic distribution with only two states, the pseudo first order kinetic rate constants (k'_{on}) can be directly obtained by fitting with a single exponential fit. The actual second-order binding rate constants (k_{on}) can be obtained from following calculation.

$$k_{on} = k'_{on} / [RNA] \qquad \text{Equation 6}$$

For systems with complex multiple exponential kinetics, poly exponential fitting can be used to reveal the kinetic rate constants. The dissociation constant can be obtained as follows and can be compared with the dissociation constants obtained from the thermodynamic data analysis.

$$K_D = k_{off} [RNA] / k_{on} \qquad \text{Equation 7}$$

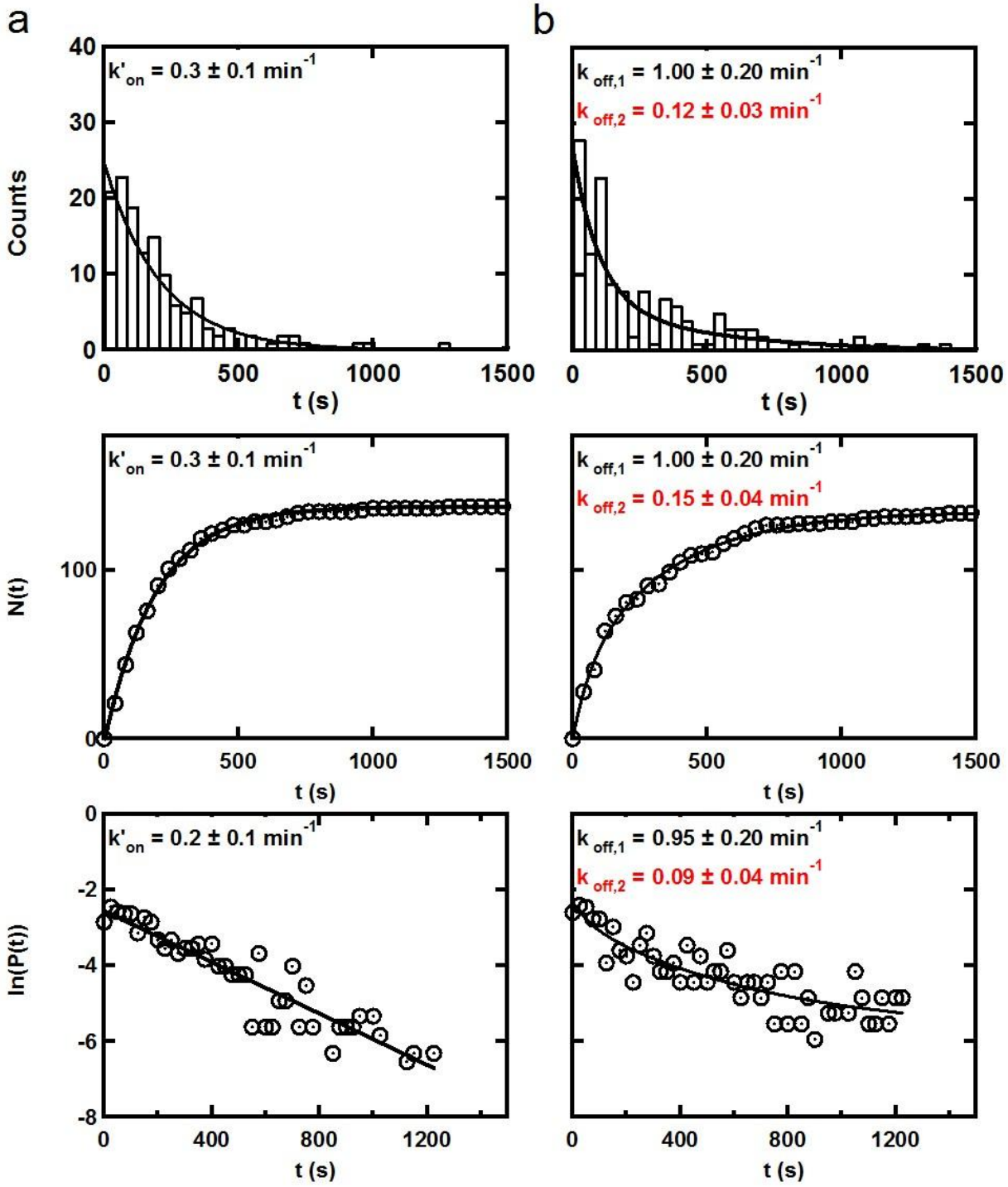

Figure 22: Single-molecule data analysis. Dwell time distributions in the kissing-loop complex (τ_{on}) and dissociated state (τ_{off}) are fit to single /double exponential decays to obtain the pseudo-fist order binding and dissociation rate constants k_{on} and k_{off}. Three times of graphical methods are used to

obtain more reliable rate calculations (Top: dwell time histograms, Middle: Integral dwell times and Bottom: log scale line fits).

Chapter 3

HIV-1 DIS Forms an Obligatory Bent Kissing Intermediate in the Dimerization Pathway

This work has been submitted for publication. (Mundigala, H.R., Michaux, J., Feig, A.L., Ennifar, E., and Rueda, D." HIV-1 DIS forms an obligatory bent, kissing intermediate in the dimerization pathway"

3.1 Introduction

Human immunodeficiency virus (HIV), a retrovirus, attacks the human immune system, which can result in acquired immune deficiency syndrome (AIDS)[31-33]. HIV is the leading cause of death in Africa and the fourth leading cause of death worldwide [34, 35]. Current therapies against HIV target mainly two viral enzymes: reverse transcriptase [36] and protease [37]. Due to the rapid evolution of strains resistant to enzymatic inhibitors, new targets must be identified.

HIV contains two similar copies of its genomic RNA, which share numerous intermolecular interactions [109]. The most prominent one is the dimer linkage structure (DLS) in the 5'-untranslated region (UTR) of the viral genome [53, 110], which has been shown to control translation, RNA packaging and recombination during proviral DNA synthesis. It has also been shown that multiple structural transitions in the 5' UTR can regulate genomic RNA packaging [111, 112].

Within the DLS, a highly conserved, nine-nucleotide apical loop, including a six-nucleotide, palindromic, dimerization initiation sequence (DIS), has been shown to be

important in the dimerization process. The palindromic sequence is flanked by three conserved purines (A272, A/G273 and A280) that are essential for the dimer stability [104, 113, 114]. The interaction between the DLS of the two genomic RNA is initiated by formation of a kissing-loop at the DIS [59, 115]. Mutations or alterations of the DIS prevents RNA dimerization and severely reduces the viral infectivity [48, 56-58]. It has been shown that the 35 nucleotide (nt) DLS with the lower stem bulge is required for the two-step dimerization in presence of NC whereas the 23 nt construct can achieve the two-step dimerization in absence of this protein [64, 65, 116]**.** Experiments on synthetic RNA fragments have shown that the initial kissing-loop dimer is subsequently stabilized by extension of intermolecular Watson-Crick base pairs as an extended duplex [59, 60]. This kissing-loop to extended duplex isomerization is strongly facilitated *in vitro* by incubation at high temperature (55°C), or at physiological temperature by the nucleocapsid protein (NC), a small, basic protein with two zinc finger domains [60-66]. Structural insights of the DIS kissing-loop and extended duplex forms have been provided by X-ray crystallography [67-69] and NMR [70-72]. This structural work has led to the discovery of high-affinity ligands of both the kissing-loop and extended duplex forms [47, 73-75], thus highlighting the DIS as a potentially interesting new viral RNA target. However, despite these progresses, little is known about the exact mechanism of isomerization.

To further investigate the isomerization mechanism, we have employed single-molecule fluorescence resonance energy transfer (smFRET) to dissect, in real-time, the *in vitro* dimerization reaction of 23 nt RNAs containing HIV-1 DIS in absence of NC. The data show that the dimerization pathway proceeds *via* three steps with two obligatory

intermediate conformations: the kissing complex and a bent intermediate. Progression along the pathway largely requires the presence of Mg^{2+} ions, which stabilize the bent intermediate. Conversely, monovalent ions, such as K^{+}, impede extended duplex formation by stabilizing the kissing complex over the bent intermediate. Mutations of the highly conserved purines in the loop indicate that the kissing complex stability and formation of the bent intermediate require a Mg^{2+}-dependent conformational change in the purines.

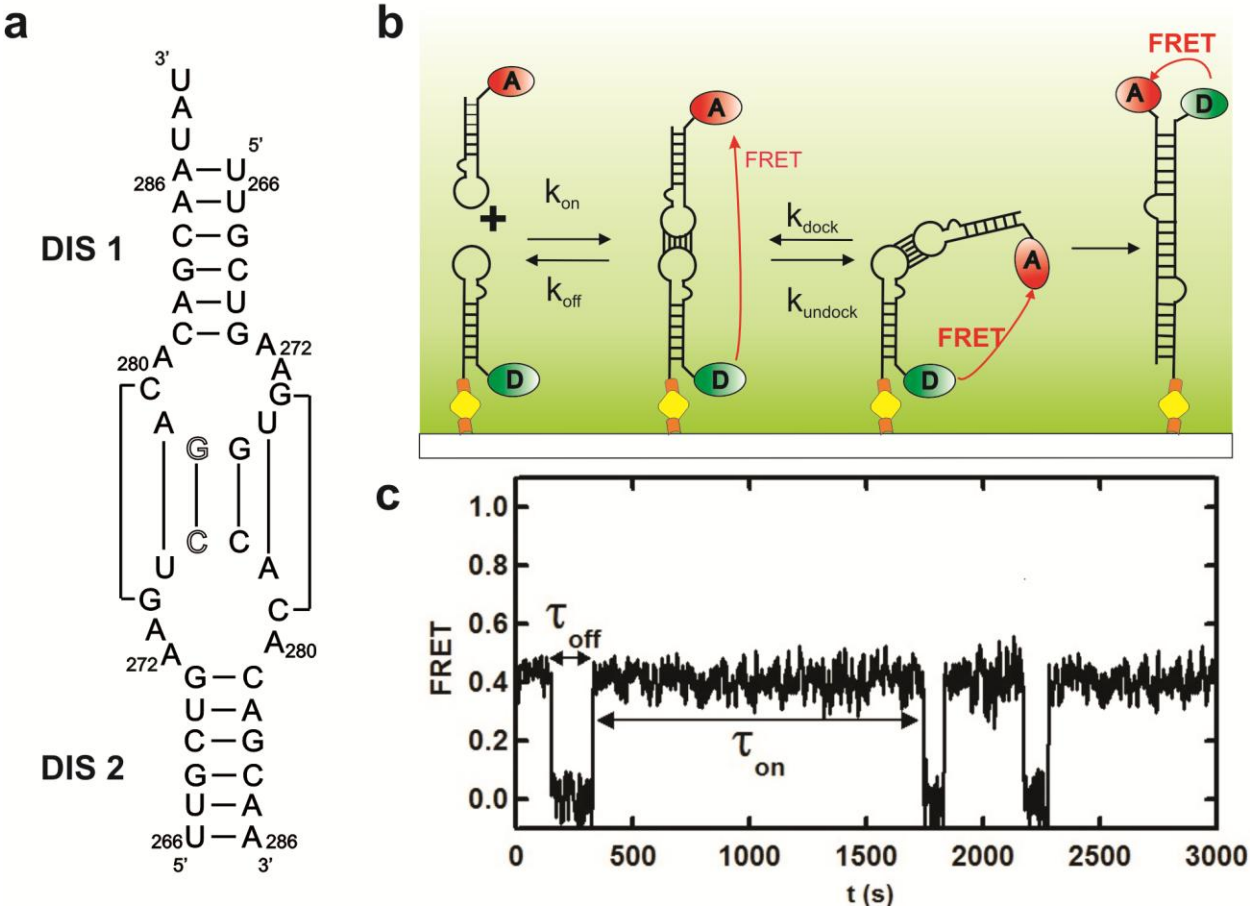

Figure 23: smFRET detection of the bimolecular HIV-1 RNA dimerization. (a) The secondary structure of the RNA strands used in the study with the fluorophores. The single base mutation in the center base pair of the loop is done in order to promote dimerization of cy3-cy5 labeled hairpins (b)

Schematic diagram of single-molecule experiments. The RNA hairpin labeled with Cy3 is immobilized onto the quartz slide surface via a biotin-streptavidin bridge. Cy5 labeled hairpin RNA is introduced to the slide and allowed to interact freely with the immobilized RNA. The Cy3 fluorophores are excited with 532 nm laser in a prism-based total internal reflection microscope. Fluorescence is collected through the objective lens and detected with a CCD camera. The dimerization pathway consists of three possible FRET states. The non interacting state (0.0 FRET), kissing complex dimer state (0.4 FRET) and the extended duplex state (1.0 FRET). (c) Calculated FRET time trajectory, showing the formation (FRET ~ 0.4) and dissociation (FRET ~ 0.0) of the kissing-loop complex. The dwell times in each FRET state (τ_{on} and τ_{off}) are then collected and used to build the dwell time distribution (see Figure 22).

3.2 Results and Discussion

3.2.1 smFRET Reveals that Kissing Complex Formation and Dissociation are Extraordinarily Slow Process

To dissect the *in vitro* dimerization reaction of DIS RNA (Figure 23a), we have used a smFRET assay previously developed to monitor the kissing complex formation of RNA hairpins in general [103, 117, 118]. The FRET donor-labeled hairpin (DIS1) is surfaced immobilized via a biotin-streptavidin bridge while the FRET acceptor-labeled hairpin (DIS2) diffuses in standard buffer, under near-physiological conditions (Figure 23b). A typical single-molecule FRET time trajectory (Figure 23c) reveals the presence of random excursions between two distinct states at ~0.0 and ~0.4 FRET, corresponding to DIS1 monomer and the DIS1:DIS2 kissing complex, respectively. Using the latter FRET value and typical approximations (κ^2 = 2/3 and R_0 = 65 Å for Cy3-Cy5) [119], we estimate the distance between the two fluorophores to be ~59 Å, in excellent agreement with the kissing complex crystal structure [68]. The time trajectories reveal that the hairpins can undergo many cycles of association and dissociation

without progressing to extended duplex formation and that the kissing complex can be very long-lived (>10 min). A dwell time analysis in the monomer (t_{off}) and kissing complex (t_{on}) states yields the pseudo-first order association (k'_{on}) and dissociation (k_{off}) rate constants, respectively (Figure 24a). The resulting histograms confirm the slow association and dissociation kinetics, as well as heterogeneous dissociation kinetics (double exponential), indicating that two kissing complex populations with different stabilities (*i.e.*, loop-loop interactions) may be formed. As expected for a binary reaction, k_{off} is independent of the RNA concentration, while k'_{on} increases linearly with [DIS 2] (Figure 24b). A linear fit to the latter yields the second-order binding rate constant k_{on} = 10^5 $M^{-1}s^{-1}$. This extremely slow formation rate constant may be the result of structural rearrangement of the hairpin prior to forming the kissing complex, which makes most diffusion-controlled collisions between the two hairpins unsuccessful. To further investigate the origin of the two kissing complex populations, we conducted metal ion titrations.

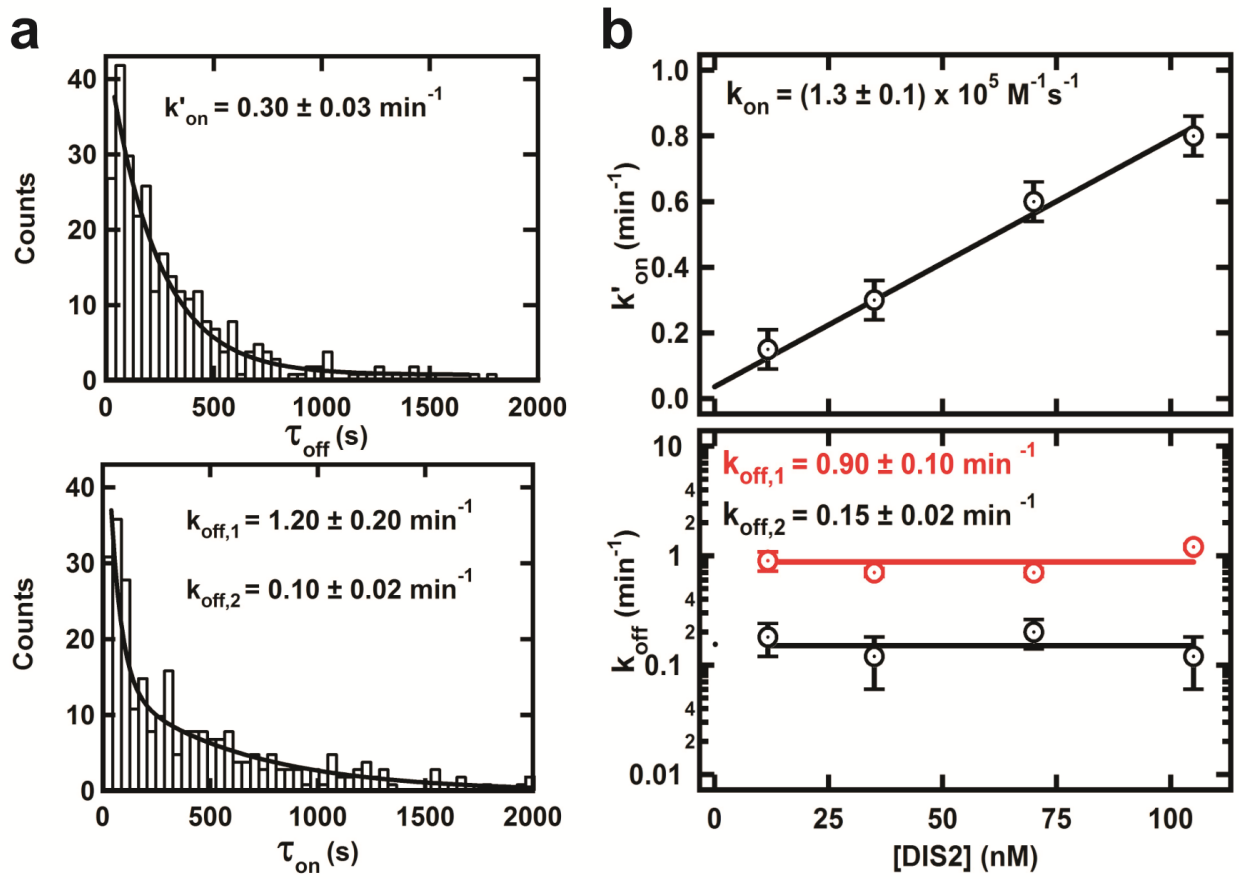

Figure 24: (a) Dwell time distributions in the monomer RNA form (τ_{off}) are fit to a single exponential decay to obtain the pseudo-first order binding rate constant k'_{on} and the dwell time distributions in the kissing-loop complex (τ_{on}) are fit to double exponential decays to obtain the dissociation rate constants, $k_{off,1}$ and $k_{off,2}$. (b) Variation of the pseudo-first order binding rate constant k'_{on} as a function of DIS2.The variation trend was used to calculate the second-order rate constant k_{on} and variation of the binding rate constant k_{off} as a function of DIS2.

3.2.2 Mg^{2+} Ions Stabilize the Kissing Complex

To test whether the observed heterogeneity results from different metal ion interactions, we measured k'_{on} and k_{off} as a function of [Mg^{2+}] and [K^{+}] (Figure 25). The observed k'_{on} is independent on both [Mg^{2+}] and [K^{+}] (Figure 25, top). Interestingly, the long-lived KC population with a slow dissociation constant ($k_{off,2}$) is only observed at

$[Mg^{2+}] \geq 0.5$ mM, whereas the short-lived KC population ($k_{off,1}$) is observed across the entire $[Mg^{2+}]$ range (Figure 25a, middle). The data show that the magnitude of both $k_{off,1}$ and $k_{off,2}$ remains invariant with $[Mg^{2+}]$, but the fraction of the short-lived KC population (f_1) decreases sharply near 0.5 mM $[Mg^{2+}]$ in favor of the long-lived population (Figure 25a, bottom). This result implies that the slow population arises from binding of a specific magnesium ion to the hairpin with a dissociation constant of 0.5 mM and stabilizing the kissing complex by ~2 $kcalmol^{-1}$. This result is in agreement with the presence of a Mg^{2+} ion in the crystal structure of the DIS kissing complex and with prior biochemical and biophysical bulk experiments [55, 120].

To test for the specificity of this magnesium ion interaction, we titrated potassium ions in a background of 5 mM Mg^{2+}. The data show that, above 300 mM, K^+ ions can partially recover the fast dissociating population (f_1) indicating that monovalent ions can compete against the divalent ion for the binding site. However, the large amount of K^+ ions required for this competition supports the notion of a tightly and specifically bound Mg^{2+} ion is primarily responsible for the observed high stability of the kissing complex.

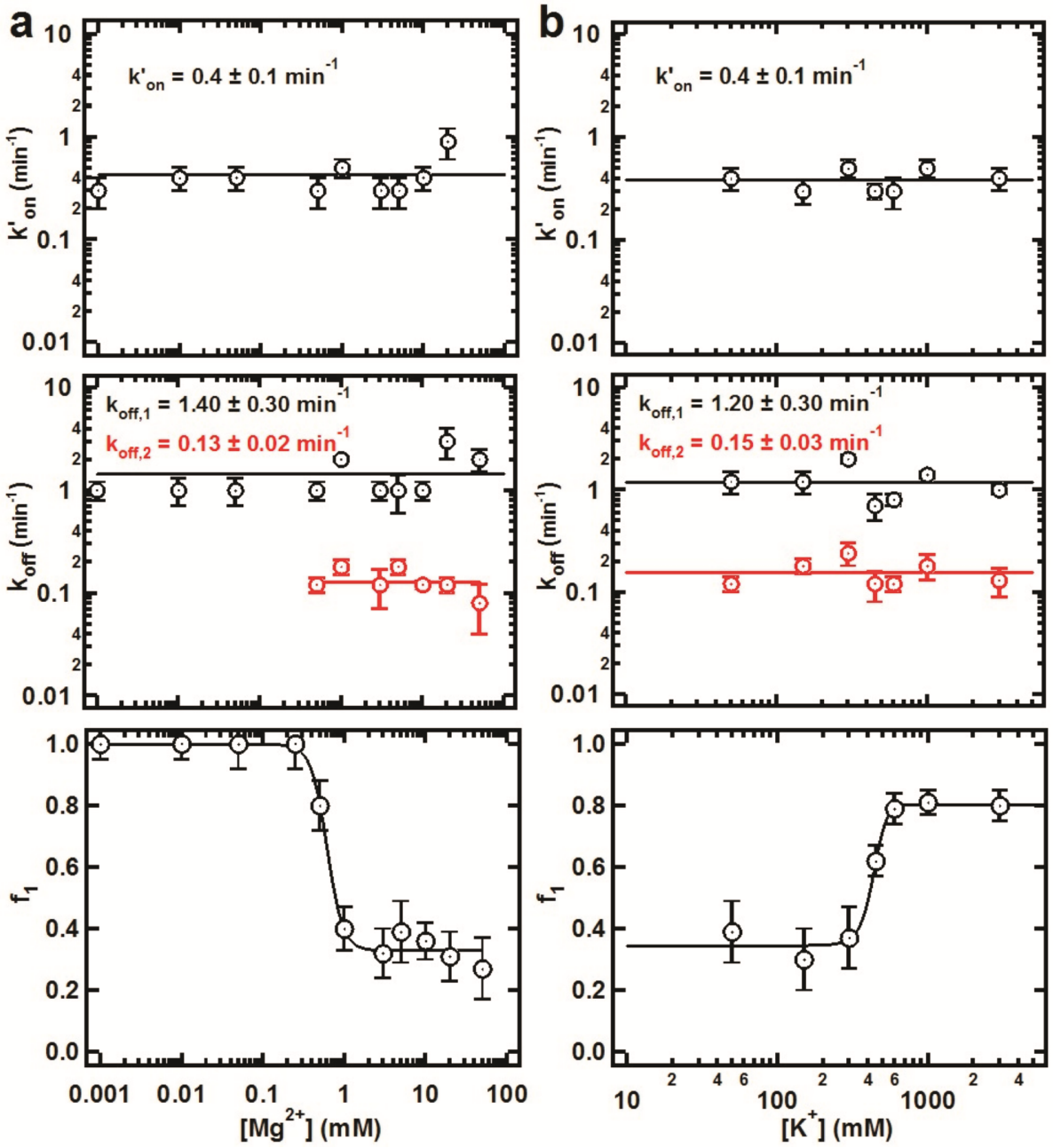

Figure 25: Effect of ionic strength on the dimerization dynamics. (a) Dependence of the pseudo-first order binding rate constant k'_{on} as a function of $[Mg^{2+}]$ and the dependence of the first order dissociation rate constant $k_{off,1}$ and $k_{off,2}$ as a function of $[Mg^{2+}]$. (b) The dependence of the pseudo-first order binding rate constant k'_{on} as a function of $[K^{+}]$ and Dependence of the first order dissociation rate constant $k_{off,1}$ and $k_{off,2}$ as a function of $[K^{+}]$. (c) Variation of population of transitions with $k_{off,2}$ rapidly increase with increasing Mg^{2+}.(d) Population of transitions with $k_{off,1}$ rapidly increase with increasing K^{+}.

3.2.3 smFRET Reveals an Obligatory Bent Intermediate

Our labeling strategy enables us to study the dynamic behavior of the kissing complex as well as its progression towards the extended RNA duplex (ED, Figure 23). With this labeling scheme, ED is expected to result in a high static FRET state, while the kissing complex (KC) is expected to yield a mid FRET state. A smFRET time trajectory (Figure 25a, left) shows the progression to the ED conformation (static FRET = 1.0) following multiple associations and dissociations of monomer hairpins to KC. Under standard conditions (5 mM Mg^{2+}, 150 mM K^{+}, 20 mM TRIS, pH 7.5, 22℃), only 13% of trajectories reach the ED state. However, under these conditions, we also observe 32% of trajectories exhibiting dynamic excursions between FRET 1.0 and 0.4 (Figure 26, Figure 27a, right). Control experiments confirm that this behavior is never observed with pre-annealed ED complexes, ruling out possible photophysical artifacts caused by the local environment in the ED conformation (Figure 27). Therefore, this dynamic population is likely to represent a different folded state that brings the fluorophores in close proximity, such as a bent kissing complex (similar to the TAR-TAR* complex [121-123]) or a cruciform intermediate initiated by fraying of the hairpins' stems [124].

The fraction of trajectories exhibiting this intermediate conformation increases with Mg^{2+} ion concentration, indicating that Mg^{2+} ions stabilize this population (Figure 28b left). Measuring the intermediate docking and undocking rate constants as a function of RNA concentration (Figure 28c and Figure 29) show that docking and undocking is a unimolecular process with respect to RNA and Mg^{2+}.

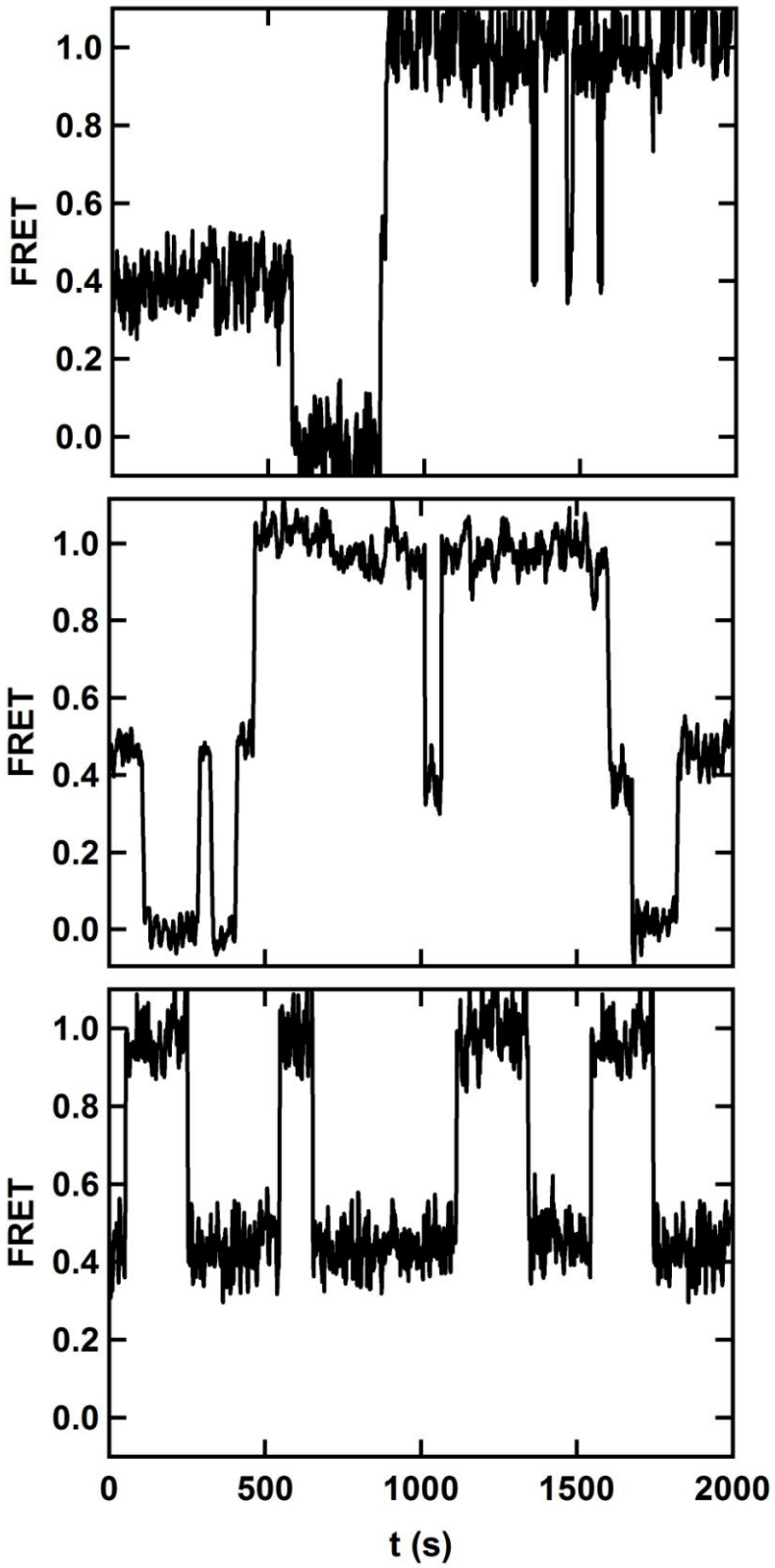

Figure 26: Types of the dynamic high FRET traces observed during the experiments. Each of these corresponds to the bent, kissing complex intermediate. Top trace corresponds to the conversion of the kissing complex to the extended RNA duplex via a bent kissing intermediate. Trace in the middle shows the formation of the bent, kissing intermediate from the kissing complex that comes back to the monomer-kissing complex equilibrium. The bottom trace shows a highly dynamic bent, kissing intermediate that dwells between the docked and undocked forms for the duration of the experiment.

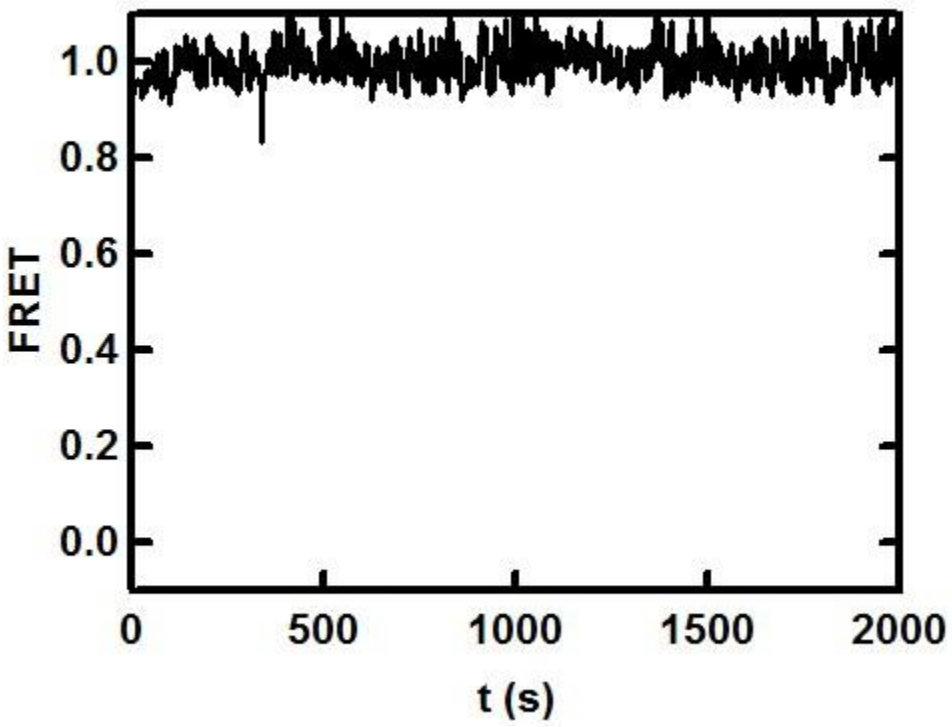

Figure 27: A Characteristic smFRET trace of a pre-annealed extended RNA duplex. DIS 1 and DIS 2 hairpin RNA were mixed, heated to 94℃, and allowed to cool gradually. This process allows the two RNA strands to be annealed and favors the formation of extended RNA duplex.

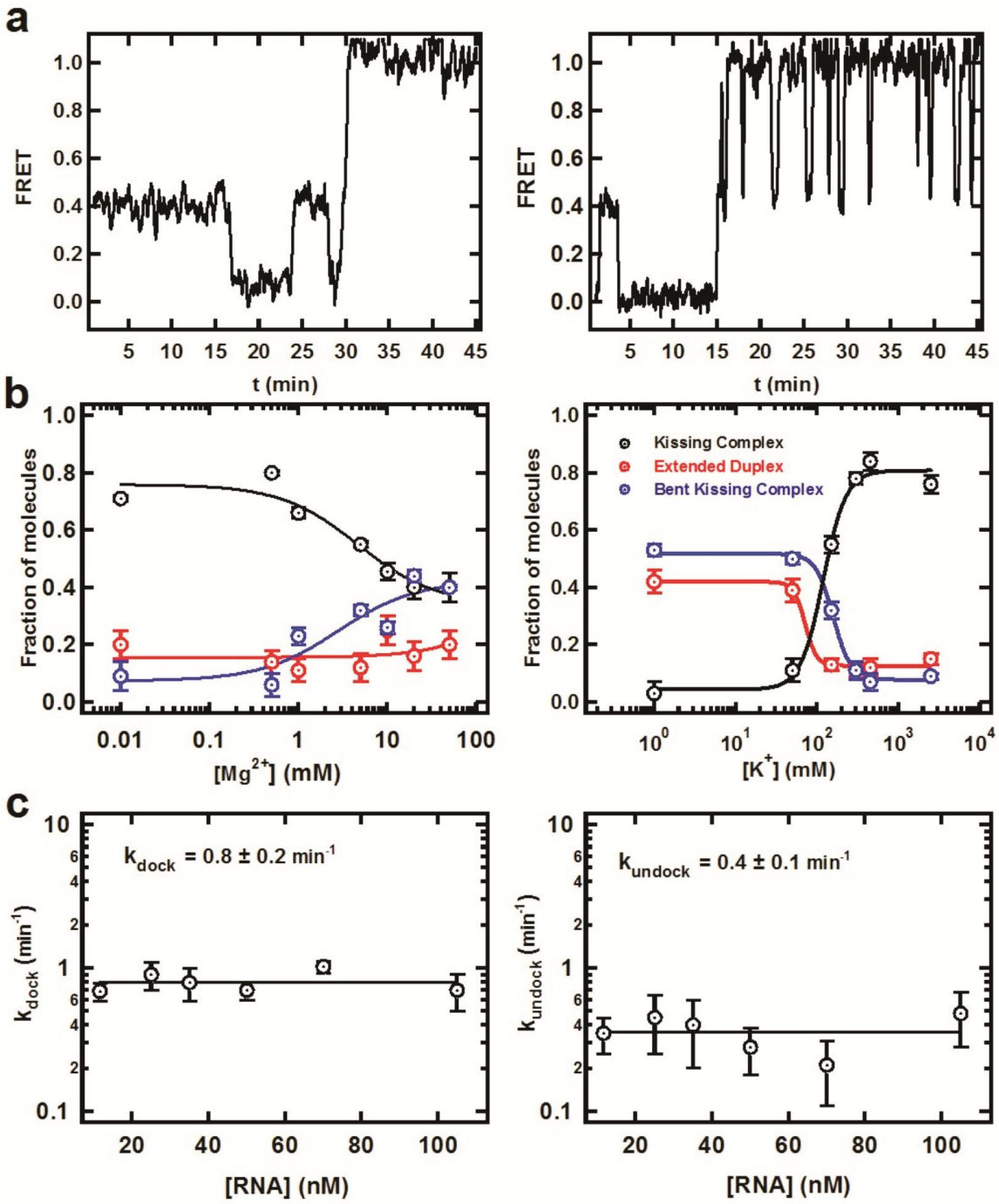

Figure 28: Presence of an additional intermediate dimer in the dimerization pathway. Representative FRET time trajectory showing (a) the static FRET 1.0 indicating the formation of the stable extended duplex and different types of dynamic FRET 1.0 state (Intermediate dimer) (b) Variation of the fraction of molecules exhibiting the intermediate dimer with varying [Mg^{2+}]

(c) Variation of the fraction of molecules exhibiting the intermediate dimer with varying [K^+] (d) & (e) Kinetics of the intermediate dimer formation and dissociation at varying [RNA].

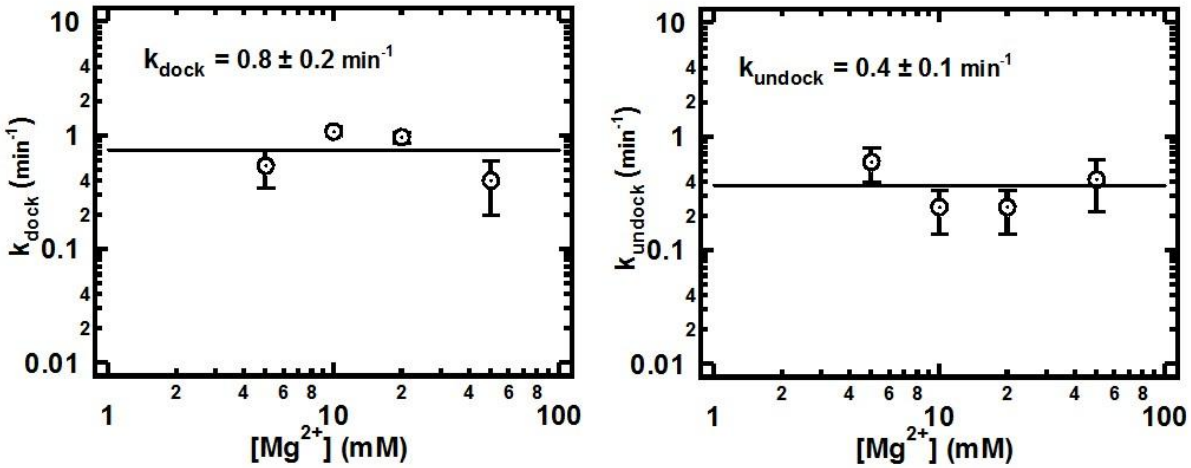

Figure 29: Kinetics of the intermediate dimer formation and dissociation at varying $[Mg]^{2+}$. Docking and Undocking rates are independent of the $[Mg]^{2+}$ indicating that docking and undocking is unimolecular with respect to Mg^{2+}.

3.2.4 Magnesium Ions are Required to Form the Bent Intermediate

To assess the effect of monovalent and divalent ions on the folding pathway, we determined the fraction of each observed state (kissing complex, intermediate bent conformation and extended) with varying [K^+] and [Mg^{2+}] in the presence saturating concentrations of free DIS 2. The Mg^{2+} titration was performed at 150 mM background K^+ ions, while the K^+ titration was performed in a 5 mM background of Mg^{2+} ions. At low [Mg^{2+}], ~75% of molecules adopt the KC conformation, and only 10-20% form the intermediate or the ED conformation. Increasing the [Mg^{2+}] to 50 mM (Figure 28b, left panel) shifts the KC population to the intermediate state, making both fractions equally populated (40%). This data suggests that Mg^{2+} ions promote conversion of KC into the bent intermediate. At high monovalent concentrations (>500 mM), the amount of bent

intermediate is highly reduced (~15%), confirming that Mg^{2+} is required for the transition of into the intermediate form.

3.2.5 A Stem Mutant Rules Out Possible Cruciform Intermediate

To characterize the structure of the intermediate and to distinguish between the possible conformations that bring the fluorophores in close proximity, we designed a hairpin mutant where the three terminal base pairs were flipped (Figure 30a), which prevents the formation of the extended duplex through a cruciform [124] or any other intermediate. Experiments with this mutant reveal that 35% of molecules still exhibit dynamic excursions between FRET 1.0 and FRET 0.4. This result clearly rules out the formation of a cruciform intermediate in our experimental conditions and supports the bent kissing complex conformation (Figure 30a). Additional experiments with the mutant also reveal that the excursion times in the FRET 1.0 state are reduced relative to the wild type (Table 4), suggesting that long range interactions between the complementary stem region of the RNA hairpins in kissing complex may contribute to the stability of the bent kissing structure (Figure 30). To further test this model, we introduced both donor and acceptor fluorophores on a single hairpin (DIS-2) and observed the extended duplex formation (Figure 31). In this scheme, the model predicts that both the kissing complex and the bent complex are characterized by high FRET values (the stem remains base paired), whereas the extended duplex should yield intermediate FRET. The data supports the kinetic model and shows that that the stem regions of the hairpin RNA remain base paired until the irreversible formation of the extended duplex.

In summary, these data suggest that the dynamic intermediate transiently adopts a bent kissing complex conformation (FRET 1.0) in equilibrium with the coaxially-stacked conformation (FRET 0.4, Figure 23b).

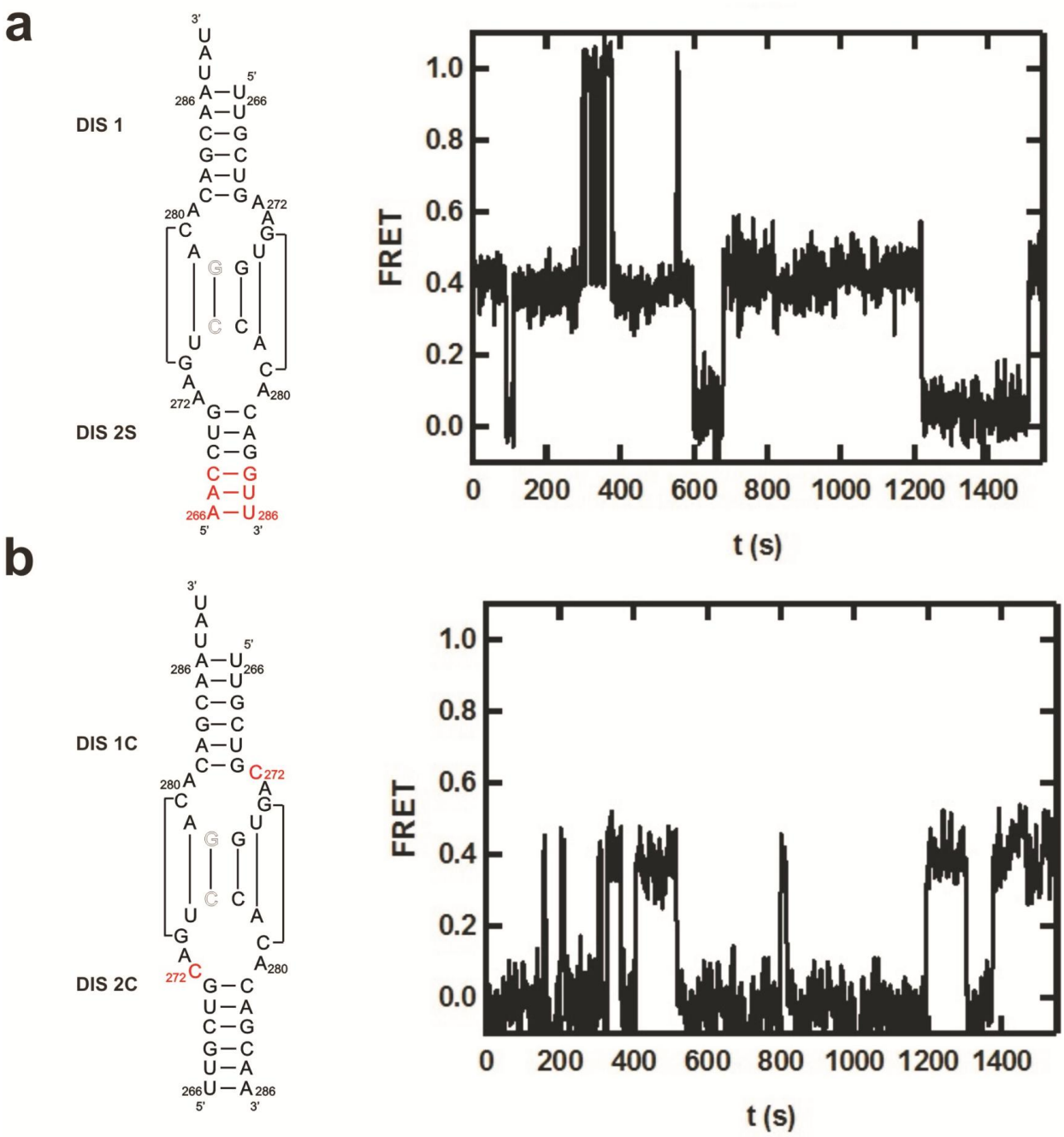

Figure 30: Loop and stem mutant studies to observe the effect on the bent KC events (a) Sequence of the stem mutation that stalls the dimerization at KC form and a characteristic trace with the stem mutant RNA showing very short-lived bent KC (b) sequence of the A272C mutation and a characteristic trace with the A272C mutant RNA showing short binding.

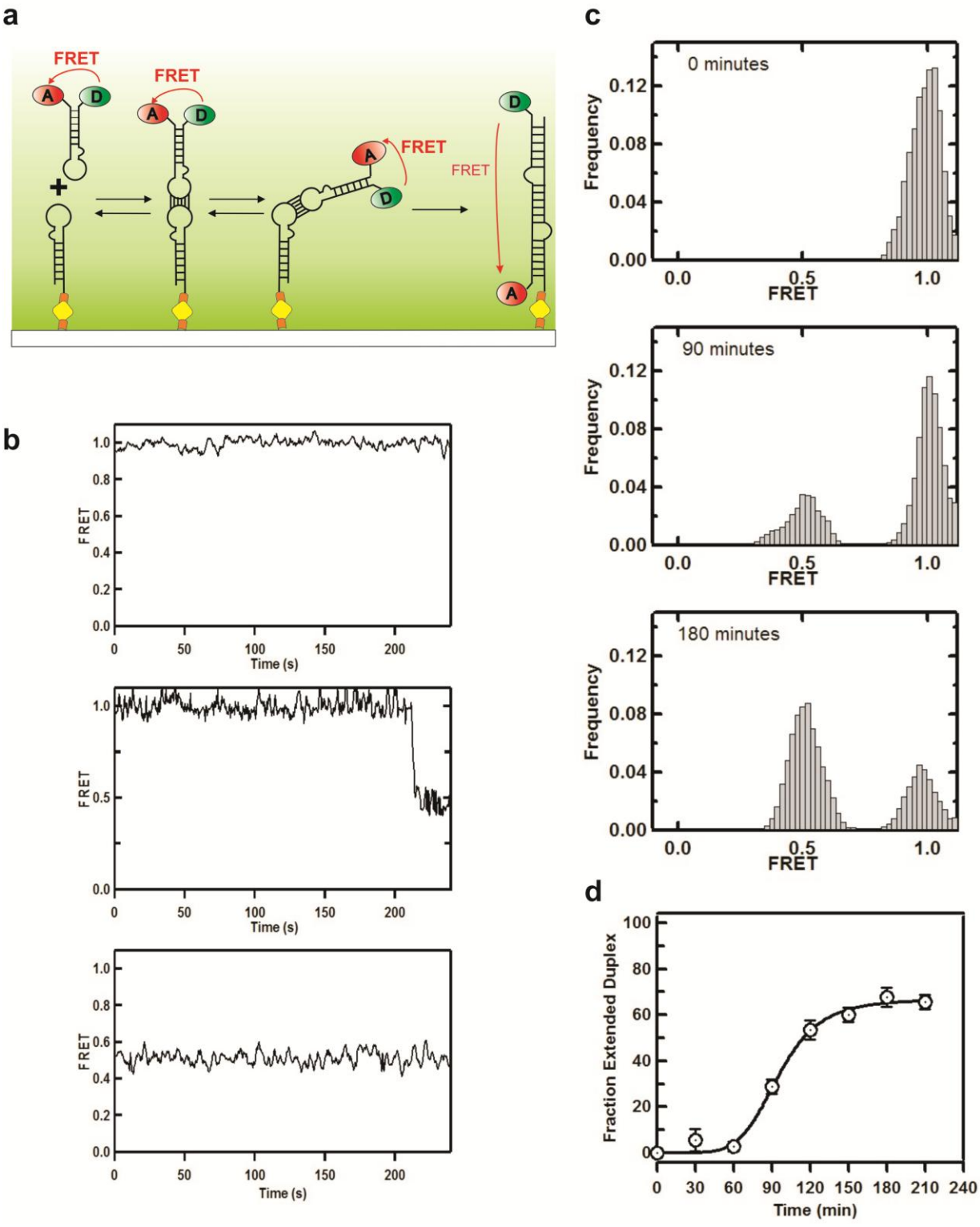

Figure 31: (a) Schematic diagram of smFRET experiments with both donor and acceptor fluorophores placed on DIS-2. According to the proposed model both the kissing complex and the bent kissing complex should exhibit high

FRET (~1.0), whereas the extended duplex formation should result in intermediate FRET (~0.5). (b) Under conditions that promote formation of all species (20 mM TRIS pH 7.4, 5 mM Mg^{2+} and 150 mM KCl, see Figure 3), three trajectories types are observed: static high FRET (top), single transitions from high (~1.0) to intermediate (~0.5) FRET indicating extended duplex formation (middle), and static intermediate FRET indicating pre-formed extended duplex (bottom). The lack of dynamics in these trajectories indicates that the stem of DIS-2 remains stably base paired until the extended duplex is formed. Kinetics of extended duplex formation: FRET histograms reveal the formation extended duplex (0.5 FRET) as a function of time (c). The observed kinetics (d) is consistent with previous measurements and the 60-minute lag period is in agreement with the presence of obligatory intermediates along the pathway.

3.2.6 Model of a DIS Kissing-Loop Bent Intermediate

Modeling of a possible bent DIS kissing complex was carried out starting from the kinked TAR-TAR* kissing-loop NMR structure and the coaxially-stacked DIS kissing-loop complex crystal structure. Structure modeling showed that only limited accommodations are required to induce a significant bending of the coaxially-stacked kissing complex. In particular, flanking adenines 272 and 273 5' of the self-complementary sequence can maintain an extra-helical conformation in the bent kissing complex similar to the one observed in the coaxial one.

3.2.7 A272 is Essential for Stability and Bending of the KC

It has been established that the stability of the KC is strongly dependent on the three highly conserved purines flanking the self-complementary sequence [113, 114]. In particular, it was suggested that A272 has a large impact on the dynamics and local conformational changes of the KC [114, 116]. We investigated the effect of A272 by mutating this base to a cytidine (A272C, Figure 30b). Under standard buffer conditions, the A272C mutant exhibits no long-lived KC population (Supplementary Table 1), similar

to the wild-type hairpins in the absence of Mg^{2+} ions. This result is consistent with the binding of a Mg^{2+} ion in the vicinity of A272, thus stabilizing the flipped out conformation of the base. This, in turn, can stabilize the KC through inter-helical adenine stacking interactions as observed in crystal structures [68].

Furthermore, the introduction of A272C mutation to one of the RNA hairpins significantly decreases (<10%) the population of molecules in the bent intermediate conformation. This asserts the fact that the metastable kissing complex is an obligatory step that is required for the transition into the bent kissing complex.

Table 4: Effect of base mutations on DIS and the stem on dimerization[1]

RNA		Rate constants[2]					Fraction of[3]		
HP1[4]	HP2	k'_{on}	$k_{off,1}$	$k_{off,2}$	k_{dock}	k_{undock}	KC	BKC	ED
DIS1	DIS2	0.3 ± 0.1	1.0 ± 0.4	0.2 ± 0.1	0.8 ± 0.2	0.4 ± 0.1	0.55	0.32	0.12
DIS1C	DIS2	2.1 ± 0.5	3.0 ± 0.8	-	-	-	0.79	0.03	0.18
DIS1	DIS2C	2.2 ± 0.7	3.0 ± 0.4	-	-	-	0.73	0.06	0.20
DIS1C	DIS2C	1.9 ± 0.4	1.3 ± 0.3	-	-	-	0.79	0.08	0.13
DIS1	DIS2S	0.8 ± 0.1	1.0 ± 0.2	0.3 ± 0.1	3.5 ± 0.2	53 ± 6	0.65	0.35	-

[1] Experimental condition is: 20 mM TRIS (pH 7.5), 150 mM KCl, 5 mM Mg^{2+}

[2] Rate constant units are all in min^{-1}

[3] KC (Kissing complex), BKC (Bent kissing complex) and ED (Extended duplex)

[4] HP1 is immobilized on quartz slide

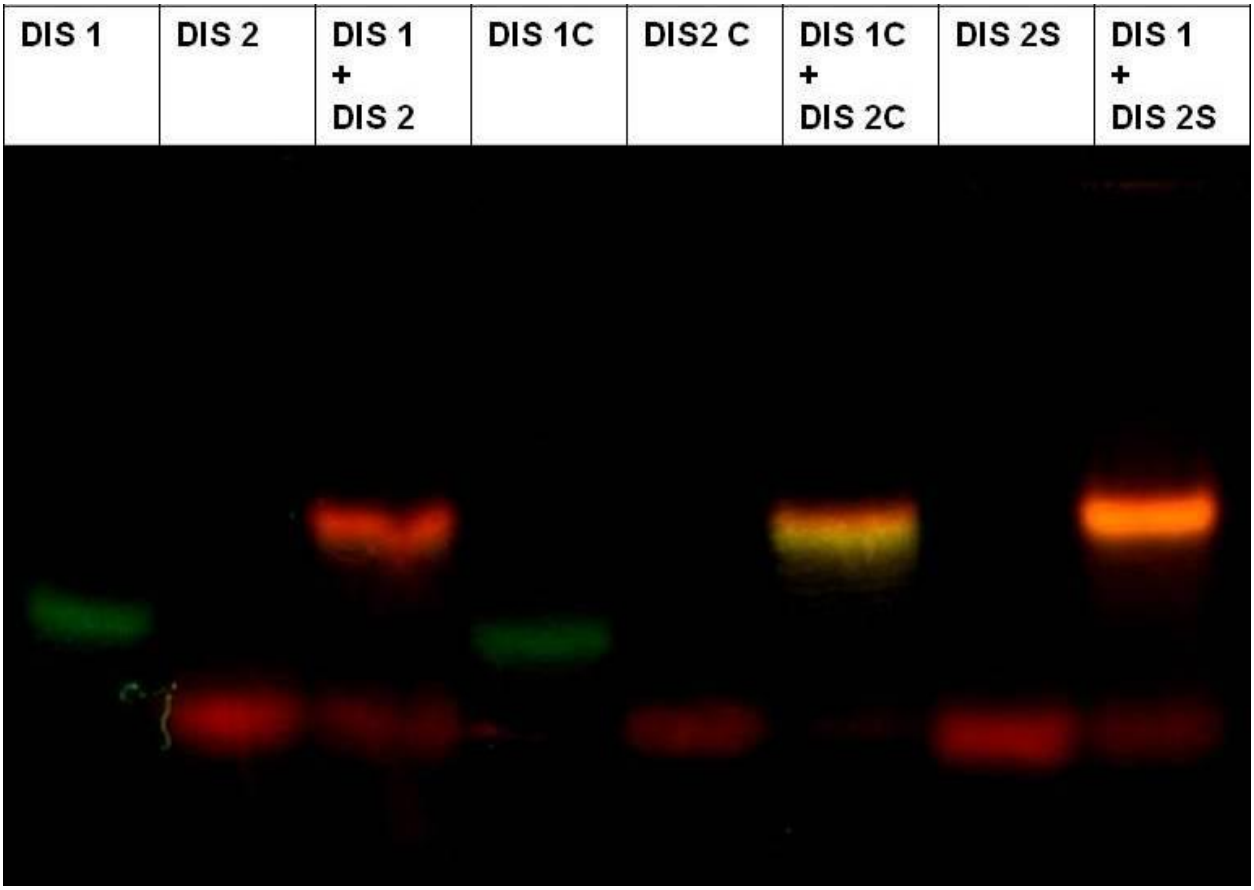

Figure 32: Native gel analysis of RNA sequences used in this study. In the presence of wild type HIV-1 sub type F sequence with modified palindromic sequence, monomer hairpin RNA is observed (Lane 1, Lane 2). When DIS1 and DIS2 RNA are allowed to interact, a mix of kissing complex (KC) and bent, kissing complex (bent KC) is observed as a dark orange FRET band in lane 3. Lane 4 and 5, respectively, show the monomer RNA of DIS1C and DIS2C with the A272C mutation. In the presence of both adenine mutated hairpins (Lane 6), only KC is formed and can be distinguished by a prominent yellow FRET band. Lane 7 contains the Cy5 labeled stem mutated RNA DIS2S.

3.3 Discussion

RNA kissing complexes/loop-loop interactions play a major role in multistep RNA folding pathways. Kinetics and thermodynamics of RNA loop-loop interactions have been previously evaluated using multiple techniques such as surface plasma resonance (SPR), electro spray ionization mass spectrometry (ESI-MS), ITC, nuclear magnetic resonance, bulk fluorescence measurements, electrospray ionization-Fourier transform mass spectrometry and UV melting [64, 116, 118, 124-127]. These studies concluded that isomerization proceeds without complete disruption of the loop-loop helix. However, details of the exact mechanism remained unclear. Studies using model RNA hairpins have shown that there is a large kinetic barrier for the conversion of KC to the ED [118].

In this study, we have used smFRET to characterize the dimerization mechanism of HIV-1 in real-time using the minimal RNA sequence responsible for the viral genome dimerization. Our smFRET kinetic data reveal that the formation and dissociation of HIV-1 kissing complex is an extraordinarily slow process under near-physiological conditions *in vitro*. We observed multiple association and dissociation steps from the monomer to the kissing complex dimer. The dissociation kinetics is highly dependent on salt conditions. In the presence of Mg^{2+}, highly stable kissing complex dimers were observed, which agrees with previous reports showing that metastable kissing dimers are formed in the presence of divalent metal ions [64]. It has been also shown that, in the presence of Mg^{2+}, conversion of kissing dimer to the extended duplex requires nucleocapsid protein [64]. Our results clearly show that, the conversion of KC to extended duplex can be achieved slowly in the absence of nucleocapsid protein after multiple dissociation and re-association steps. Our results with the A272C mutation

shows that, in the absence of A272, kissing dimers are unable to gain high stability even in the presence of high concentrations of [Mg^{2+}]. This confirms the essential role played by this universally-conserved adenine in the formation of a stable kissing-loop complex [68, 104, 116]. We propose that adenine base flipping and stabilization of flipped adenines upon Mg^{2+} binding might be the cause of the observed high stability of the kissing dimer in the presence of Mg^{2+}. In support of this hypothesis, it has been shown that A272 can be protonated and this might promotes the loop dynamics and the conversion to ED [116].

The smFRET trajectories also reveal the presence of an intermediate in the transition from kissing complex to the extended RNA duplex. Experiments performed with inverted DIS stem sequences clearly eliminate the possibility of cruciform intermediates as observed with monomeric DIS hairpins mutants [124]. Based on the smFRET data, we propose that the folding intermediate corresponds to a bent kissing complex (Figure 33, Figure 34), similar to a TAR complex structure [128] or ColEl plasmid specific RNA I and RNA II transcripts [129, 130]. A similar concept of a bent kissing complex retaining the WC base pairing at the loop interaction and the C2 symmetry of the loop-loop interface that facilitates the duplex formation via KC has been suggested previously based on NMR data [131]. A similarly bent KC transition-state has also been proposed in the NC-chaperoned dimerization pathway [64]. Alternatively, the bent intermediate may resemble a structure suggested in a recent NMR study, where the DIS hairpin base pairing remains intact and inter-stem interactions are facilitated as a result of KC bending, which brings the two stems in close proximity [132]. In addition, there is *in vivo* evidence for an intermediate HIV-1 genomic RNA (gRNA) dimer on the

path from immature gRNA dimer to mature gRNA dimer inside the HIV-1 particle [133]. We built a molecular model of a DIS kissing complex intermediate based on the TAR complex [128] by keeping the angle between both hairpins constant (Figure 33). In this model A272 and A273 are in a flipped out configuration and stacked onto each other as observed in KC crystal structures, showing the feasibility of bending the DIS KC with perfectly coaxially-stacked hairpins thanks to the plasticity provided by unpaired adenines. This is in line with previous studies highlighting the dynamics of these purines within the KC [67, 116, 134]. This molecular model yields an inter-stem distance of ~49 Å. However, based on the measured FRET efficiency (0.9 – 1.0), we estimate the distance between the ends of the two hairpin stems to be ≤ 38 Å (assuming $\kappa^2 = ⅔$). Therefore, we propose that the observed bent KC adopts a conformation in between the model in Figure 33 and the completely bent KC conformation proposed by Dethoff and coworkers [132].

Monovalent and divalent ion titrations clearly show that Mg^{2+} ions are required for the formation of the proposed bent kissing complex. Our data also highlights the important role of the universally-conserved adenine 272, 5' of the loop. Experiments performed with the DIS A272C mutant show that the KC formation is strongly hindered, thus transition to bent KC or the extended duplex cannot be readily achieved even at high Mg^{2+} concentrations.

.

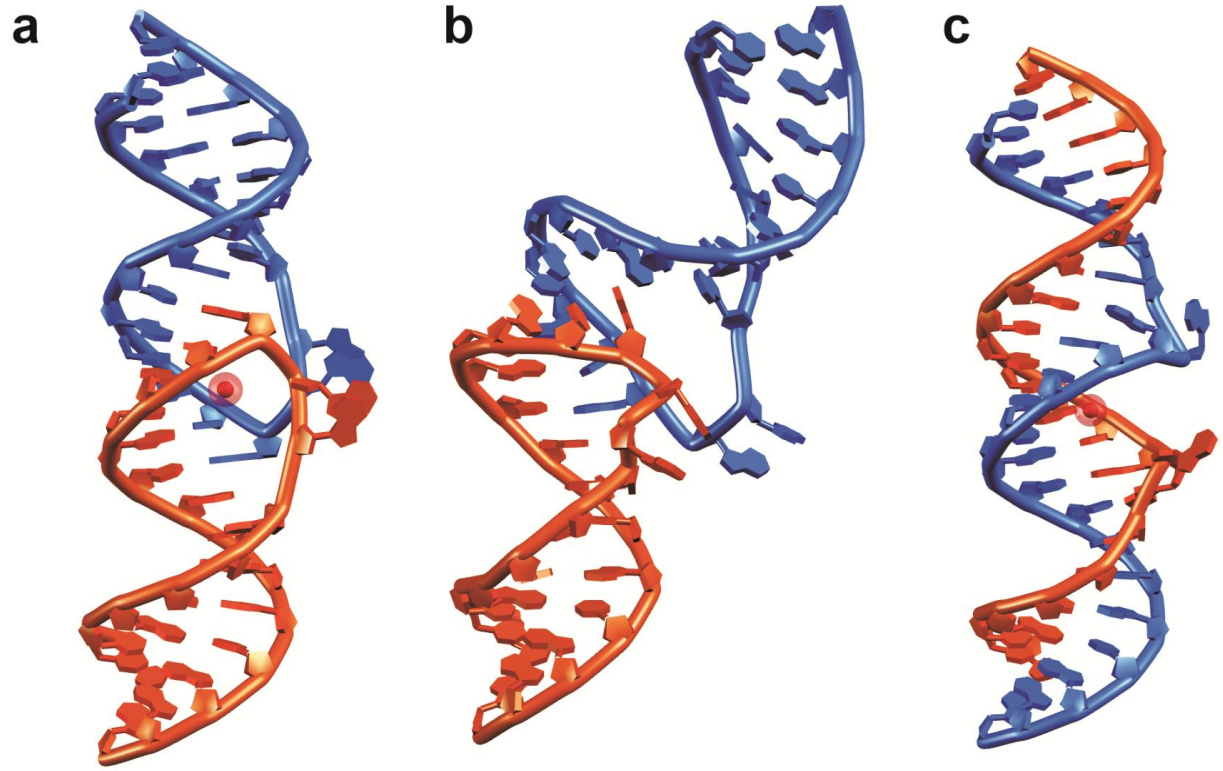

Figure 33: 3D architecture of HIV-1 DIS RNA dimers along the isomerization pathway. The two strands are shown in orange and blue. a) The DIS kissing complex as observed in crystal structures (PDB 1XPF). b) Putative bent kissing complex intermediate. c) DIS Extended duplex as observed in crystal structure (PDB 462D). Hexahydrated magnesium cations observed in X-ray structures are shown as red spheres.

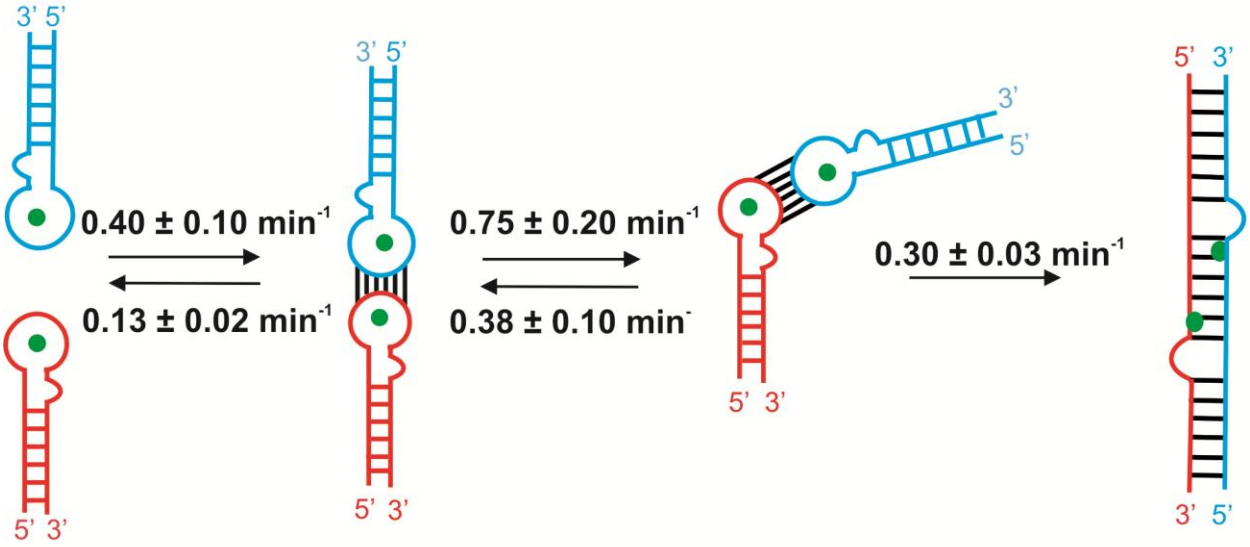

Figure 34: Proposed model for the HIV-1 RNA dimerization with the observed bent dimer. Proposed model involves the conversion of the KC to a bent form of a kissing complex. Then it converts to a stable extended duplex. The intermediate bent KC dimer is favorable under high [Mg^{2+}].

3.4 Conclusion

The ability to characterize the folding pathway of HIV-1 RNA *in vitro* is critical in understanding the viral RNA dimerization during viral assembly. We have demonstrated a single-molecule fluorescence resonance energy transfer assay to monitor the dimerization of minimal HIV-1 RNA sequence containing DIS. Our smFRET data revealed the bimolecular association rate constant of the two hairpin RNAs is 1.3×10^5 $M^{-1}s^{-1}$. This formation rate is independent of the Mg^{2+} and increases linearly with RNA concentration which confirms that the formation is a diffusion-controlled reaction. The important role of Mg^{2+} for the DIS dimerization has been established by various bulk-analysis methods [55, 64, 65, 135]. Our study supports the hypothesis that, in solution, Magnesium binds the DIS with an equilibrium dissociation constant near 5 mM. Magnesium binding stabilizes the kissing interaction and the kissing complex

dissociation rates significantly increase (1.5 ± 0.3 min^{-1}) in the absence of magnesium. We observe magnesium bound metastable KC population with extreme slow dissociation rate (0.12 ± 0.02 min^{-1}). Our smFRET analysis reveals that HIV-1 RNA dimerization occurs through a three-step folding pathway in which the RNA kissing complex shifts to a bent kissing conformation which leads to the formation of the extended RNA duplex *via* interaction through stems. We propose a 3D model of a possible bent DIS kissing complex intermediate as expected from smFRET data. Our data also confirms that docking of the KC to form a bent conformation is independent of Mg^{2+} and RNA concentrations, although it requires the presence of Mg^{2+}, indicating that bending is a unimolecular process. During maturation of the viral particle, the nucleocapsid protein (NC) chaperones the dimerization pathway. Thus, ongoing experiments are currently focused at studying the role of NC in the dimerization pathway at single-molecule level. The mechanistic insights gained from these experiments represent significant progress towards understanding the HIV-1 dimerization mechanism and might help the rationale development of new ligands targeting the HIV-1 DIS RNA [136].

.

Chapter 4

Role of the Nucleocapsid Protein in HIV-1 RNA Dimerization

4.1 Introduction

HIV-1 nucleocapsid protein (NC) is one of the structural proteins encoded by the gag precursor along with the matrix protein and the capsid protein in HIV-1 [38]. The main function of NC is binding to the packaging signal and then to deliver the full length viral RNA in to the assembling virion [61]. Therefore, in vivo, the dimerization of HIV is facilitated in the presence of NC as illustrated by previous studies [64]. In principle, NC remodels the RNA to its most thermodynamically stable form acting as an RNA chaperone. Nucleocapsid protein is fifty five residues long and includes two zinc finger domains (of the CCHC type) that are flanked by basic amino acids [38, 137].

Similar to other retroviral counterparts, HIV-1 NC specifically selects the viral genomic RNA from the pool of cellular RNA and allows RNA assembly and efficient packaging into the new budding virion [38]. Nucleocapsid protein is known to bind single stranded nucleic acid structures nonspecifically [61, 80, 83]. When bound to NC, RNA secondary structure becomes more compact and thus leads to protection from nucleases activity. NC has the ability to show variable binding to HIV-1 RNA with different stoichiometric ratios at different functioning levels. HIV-1 NC is involved in multiple functions throughout the viral life cycle, which are listed here.

1. Non-specific binding of HIV-1 NC to viral genomic RNA compacts the RNA within the viral capsid core and also protects from nuclease activity [138].
2. Non-specific binding of NC promotes annealing of the tRNA primer to the primer binding site [139] and DNA strand exchange during RT of the viral RNA.

3. NC stimulates the activation of viral reverse transcriptase [140] and improves the efficiency of intergrase activity [141].

4.2 HIV-1 Nucleocapsid Structure

HIV-1 NC is a 55 amino acid, basic protein in which 15 residues are highly basic with only four acidic residues. The protein contains two Zn finger domains (CCHC type) flanked by basic amino acid residues. It has been shown that specific binding to the dimerization initiation sequence of HIV-1 NC requires the presence of both Zn finger domains intact as well as the basic amino acid residues. The first Zn finger domain is well conserved throughout retroviral species when compared to the second Zn finger domain. Although most of the functions of HIV-1 NC involve only non-specific binding, which does not require the use of Zn finger domains, its role in RNA dimerization requires specific RNA-NC interactions and Zn finger domain interaction with RNA [25].

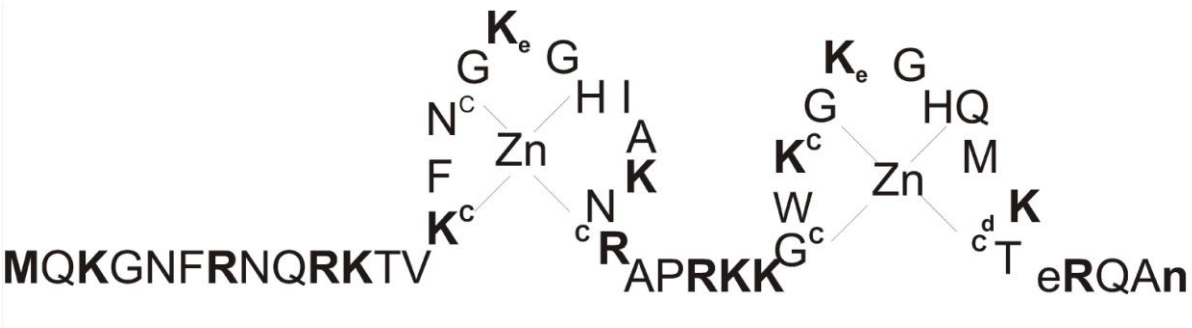

Figure 35: HIV-1 NC structure with the amino acid sequence

4.3 Role of NC in HIV-1 RNA dimerization

The role of the NC in HIV-1 RNA dimerization has been extensively studied by multiple research groups. It is currently believed that NC aids in conversion of loose

RNA dimer (KC) to a tight RNA dimer (extended RNA duplex) [61, 64, 142]. The initial formation of a metastable loose RNA dimer (KC) is believed to be independent of NC mediation. Work by *Rist et al*, 2002 suggests that the first step of RNA dimerization results in the formation of two classes of kissing complex structures with different isomerization potentials. They have identified the two different classes of KC as divalent-stabilized and monovalent-stabilized dimers. Furthermore, they state that the divalent-stabilized KC requires the involvement of NC for the further conversion into an extended RNA duplex, but monovalent-stabilized KC can be readily converted to ED at a slow rate without mediation of NC. In contrast to this work, the results discussed in the previous chapter of this dissertation have identified that only divalent-stabilized KC can proceed through the dimerization folding pathway. Therefore, the research work discussed in this chapter of the dissertation is focused on studying the role of NC protein in the dimerization pathway of HIV-1 RNA. We have extended our study to the HIV-1 nucleocapsid protein effects on the kinetics and thermodynamic properties of the genomic RNA dimerization.

4.4 Results and Discussion

All the experiments with NC protein were conducted at near physiological conditions (5 mM Mg^{2+}, 150 mM KCl, 20 mM TRIS, pH 7.4) unless otherwise specified.

4.4.1 Nucleocapsid Protein Facilitates the Extended RNA Duplex Formation

Initial experiments were conducted to observe if NC protein facilitates the extended RNA duplex formation from RNA monomers. smFRET experiments were performed as explained in chapter 2. At every NC concentration, population of kissing

complex (KC), bent kissing complex (BKC) and extended duplex were measured and plotted as a function of NC concentration (Figure 35).

It is observed that at low NC concentrations, amount of KC is around 55%, 35% RNA exist as BKC and only 10% RNA have reached the RNA extended duplex form. Upon increasing NC concentration, at 100nM NC, fraction of KC is significantly dropped to ~ 25% and ~45% of BKC and 30% of the extended RNA duplex is observed. Further increase in NC to 01 µM results in a further drop in KC level to ~15% and leads to increased amount of BKC (~45%) and extended RNA duplex (~40%). These results suggest that as NC concentration is increased, conversion of KC to BKC and BKC to extended duplex is enhanced. At saturating NC concentrations, ~50% of the RNA are in a duplex conformation, confirming that the NC used here is active and can therefore conduct the nucleic acid chaperoning in the HIV-1 RNA dimerization in vitro with the 23 nucleotide RNA construct used in this study.

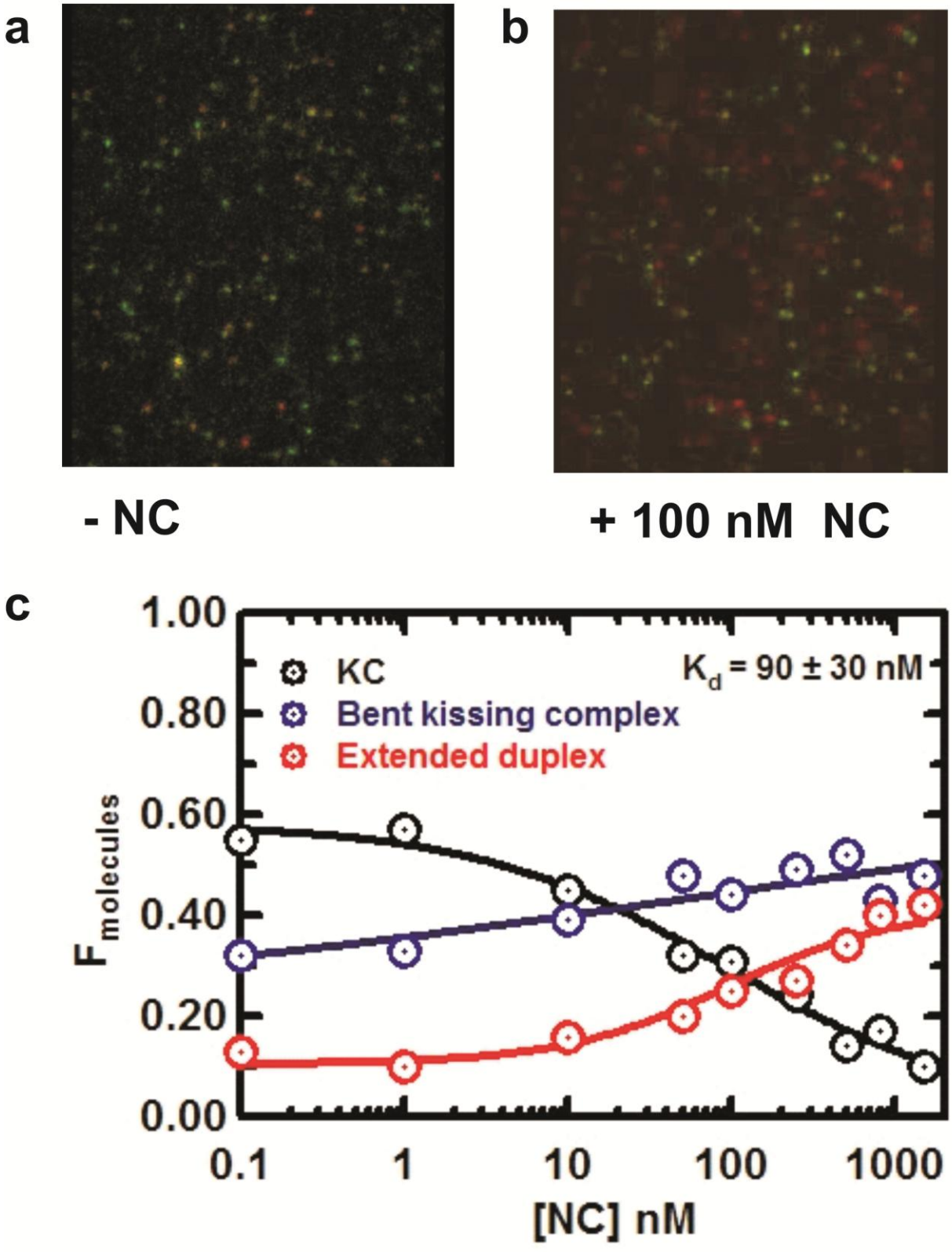

Figure 36: Effect of NC on the progression of HIV-1 in vitroerization pathway invitro. (a) and (b) show screen shots of single-molecule movies in the absence and presence of NC respectively. The dissociation constant for NC binding to RNA is 90 nM.

This experiment further provides proof that the proposed bent KC is present in the nucleocapsid protein catalyzed pathway, indicating that it is an obligatory intermediate in the dimerization of DIS containing HIV-1 23 nucleotide RNA sequence in vitro.

4.4.2 Nucleocapsid Protein Binds to Monomer RNA

The effect of KC formation and dissociation back to RNA monomers in the presence of NC protein was analyzed by observing the KC formation and dissociation rates at a range of NC concentrations. Experiments were conducted at near physiological conditions. Determining the effect of NC on the kinetics of KC formation and dissociation will pave the path for identifying the role of NC HIV-1 RNA dimerization in vitro. According to the currently accepted theory, NC does not mediate in the formation of KC, instead NC binds to the preformed KC and catalyze the formation of extended RNA duplex (Figure 12) [64].

At low NC concentrations, no significant effect on kissing complex formation rates was observed. Even though at low NC concentrations KC formation rates were unaffected by the presence of NC, at higher NC concentrations (>50 nM), a faster KC formation rate is observed (Figure 37). This suggests that in the presence of NC bound RNA, kissing complex formation is kinetically favored. Thus, initial NC binding is to RNA hairpin monomers. Therefore, in contradiction to the current knowledge, our smFRET

data suggest that NC binds to individual hairpins and alters the conformation in a way that facilitates KC formation.

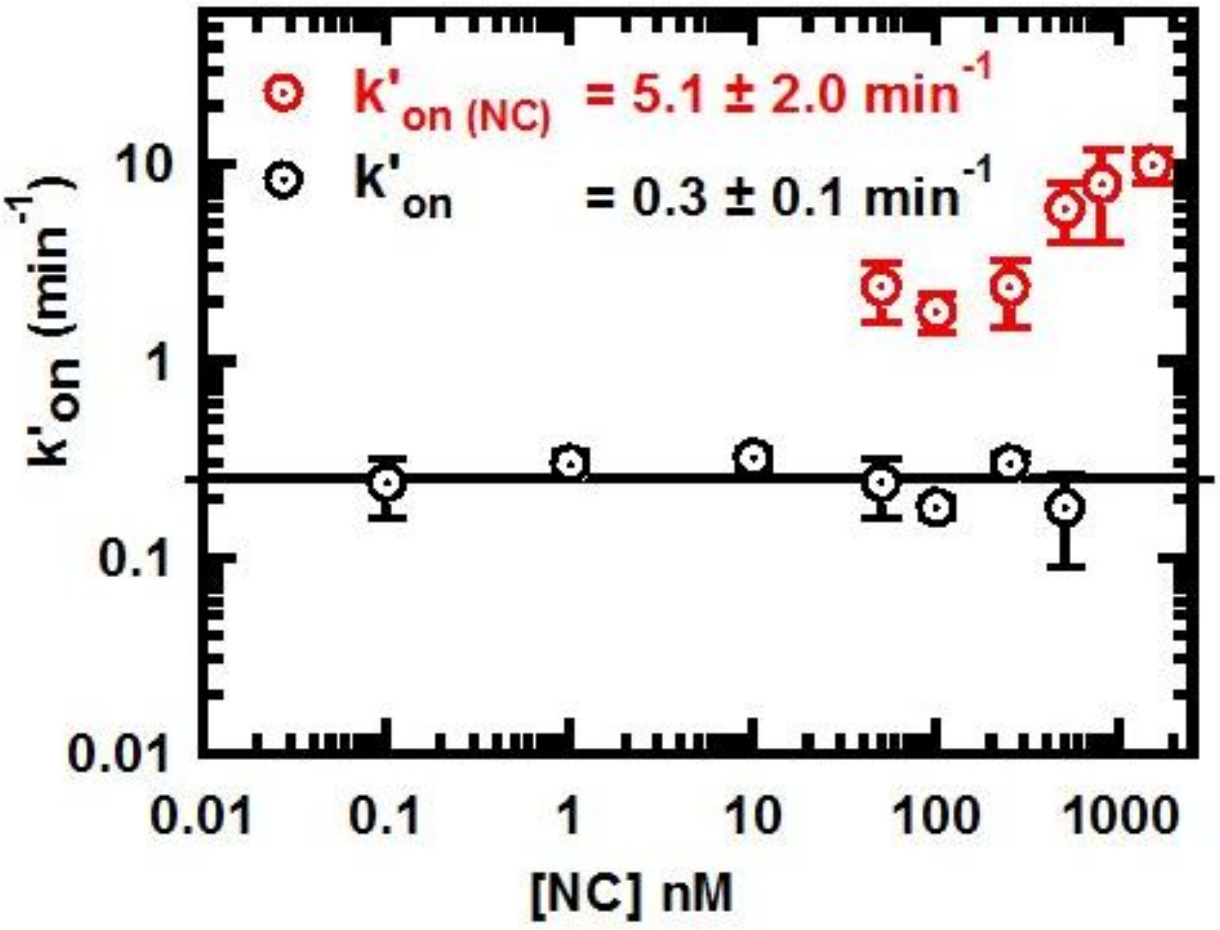

Figure 37: Effect of NC on kissing complex formation rates. Although no significant change in KC formation was observed at low NC concentrations, at higher NC concentrations, a faster KC formation rate is observed, indicating that NC binds to individual RNA monomers and alters the hairpin conformation to facilitate KC formation**.**

Rate calculations in the presence of NC were extended to KC dissociation rates to observe the NC impact on KC stability. Similar to KC formation rates, KC dissociation rates were unaltered at low NC concentrations (<50 nM). Interestingly, at NC concentrations higher than 50 nM, faster dissociating KC population was no longer

visible. Previously this faster dissociating KC population was assigned to a KC mediated by monovalent ions. In summary, when the effect of NC on KC stability and dissociation rates was studied, it was observed that at significant NC concentrations (>50nM), only metastable KC were observed. Therefore, we hypothesize that in the presence of NC, KC reaches a thermodynamically stable form hindering the possibility of dissociation back to monomer RNA. Additionally, binding of NC increases the dwell time in KC form which results in a higher probability of progression towards BKC and extended RNA duplex.

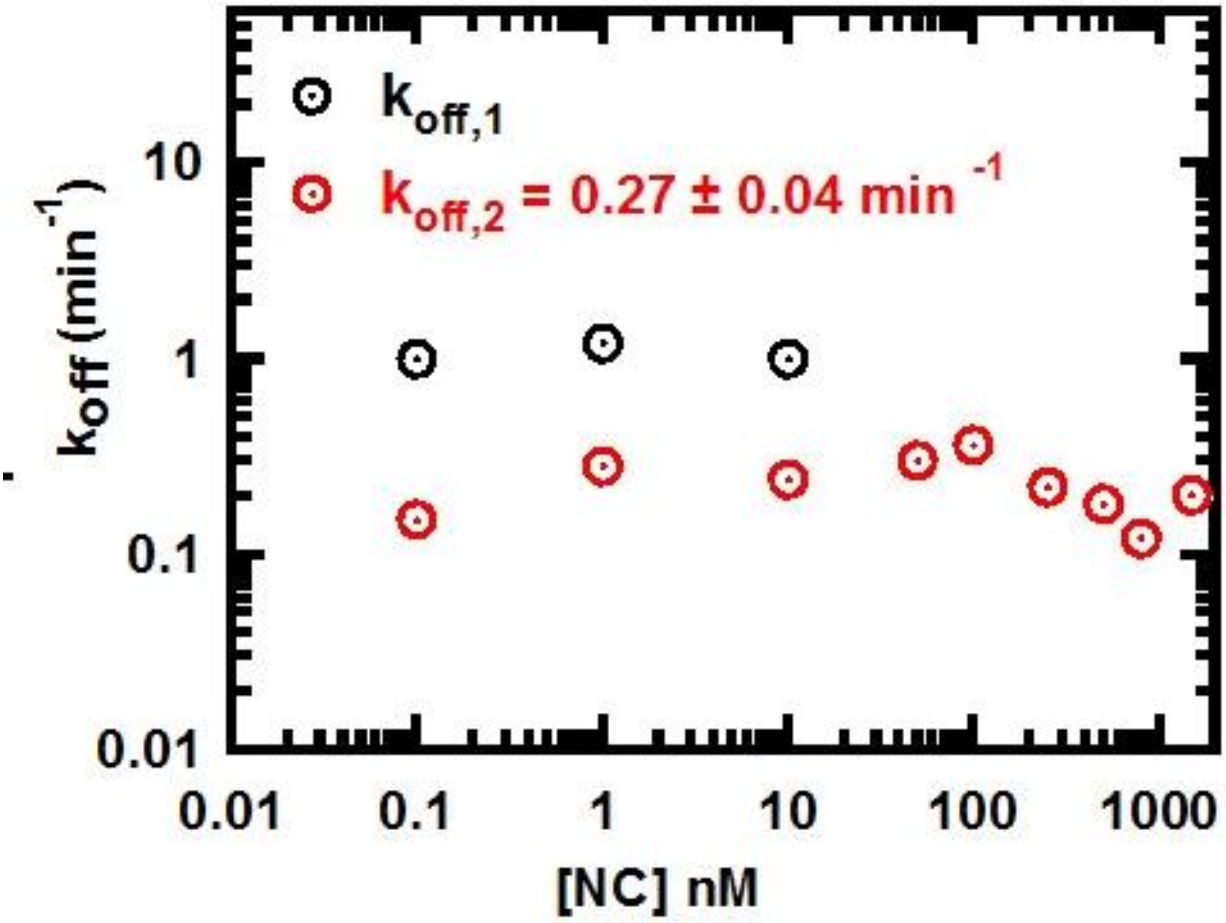

Figure 38: Effect of NC on the kissing complex dissociation rates. Although no significant change in KC dissociation rates was observed at low NC concentrations, at significantly higher NC concentrations, the fast dissociating KC population disappears. This indicates that in the presence of NC, KC reaches a thermodynamically stable form which hinders the possibility to dissociate back to monomer RNA. Additionally, binding of NC increases the dwell time in KC form, increasing the probability of progression into BKC and extended RNA duplex.

4.4.3 A272C mutation is Not Responsive to the Presence of NC

As previously explained, flanking adenine bases in the HIV DIS stem is important for characteristic features in the dimerization pathway [64, 67, 69, 143]. We have previously shown the specific contribution of A272 on the stability of RNA KC and also

on the successful progression of the dimerization pathway (Figure 27). As discussed in chapter 3, we have successfully identified A272 as a key requirement for specific Mg^{2+} binding to DIS loop and also in the conversion of KC to BKC. Thus, in the absence of A272 KC stability was greatly reduced and no further progression of dimerization pathway was observed. Due to the importance of A272 position, we extended our studies to evaluate the role of A272 on NC protein binding. In order to observe the role of A272C mutation in nucleocapsid protein binding, experiments were conducted with DIS 1C and DIS 2C RNA in the presence of NC protein (Figure 27 illustrates the structure of RNA mutants). DIS 1C and DIS 2C RNA exhibited a faster KC formation rate that lead to a short-lived KC in the absence of NC protein, indicating the need of A272 KC stabilization. Further A272C mutants were unable to proceed from kissing dimer form in the dimerization pathway. By conducting experiments with A272C mutated RNA, in the presence of NC, the goal was to identify the role of A272 in the suggested NC induced RNA conformational change.

Table 5: Effect of A272C mutation on NC chaperoned dimerization.

RNA	Rate constants (min^{-1})	
	k'_{on}	$k_{off,1}$
DIS1 DIS2C	2.2 ± 0.7	3.0 ± 0.4
DIS 1 DIS 2C +NC (150 nM)	1.7± 0.2	1.5 ± 0.4
DIS1C DIS2C	1.9 ± 0.4	1.3 ± 0.3
DIS 1C DIS 2C+NC (150 nM)	1.4 ± 0.5	2.1 ± 0.7

Experimental conditions are: 20 mM TRIS (pH 7.5), 150 mM KCl, 5 mM Mg^{2+}

We measured the kissing complex formation rates and the dissociation rates for the A272C mutated RNA in the presence and absence of NC. The calculated rate constant data are shown in table 5. Previous experiments clearly indicated that the addition of NC with concentrations above 50 nM gave rise to a population of KC formed with a faster association rate of ~5.1 min^{-1} (Figure 37). But for the A272C mutant, KC formation rates are independent of NC presence. This implies that the A272C mutation hinders HIV DIS RNA ability to complex with NC. The data clearly emphasize that, in the absence of at least one A272 nucleotide, NC fails to stabilize the formed KC. This result suggests that NC binding requires A272, indicating that the conformational change previously observed in the NC chaperoned RNA dimerization pathway could be the bulging out of unpaired adenines in the loop, orienting the hairpin for successful collisions between each other.

4.4.4 Bent Kissing Complex Formation is Favored in the Presence of NC

In order to characterize the dimerization pathway in the presence of NC, studies were extended to the formation of bent KC and dissociation back to kissing complex.

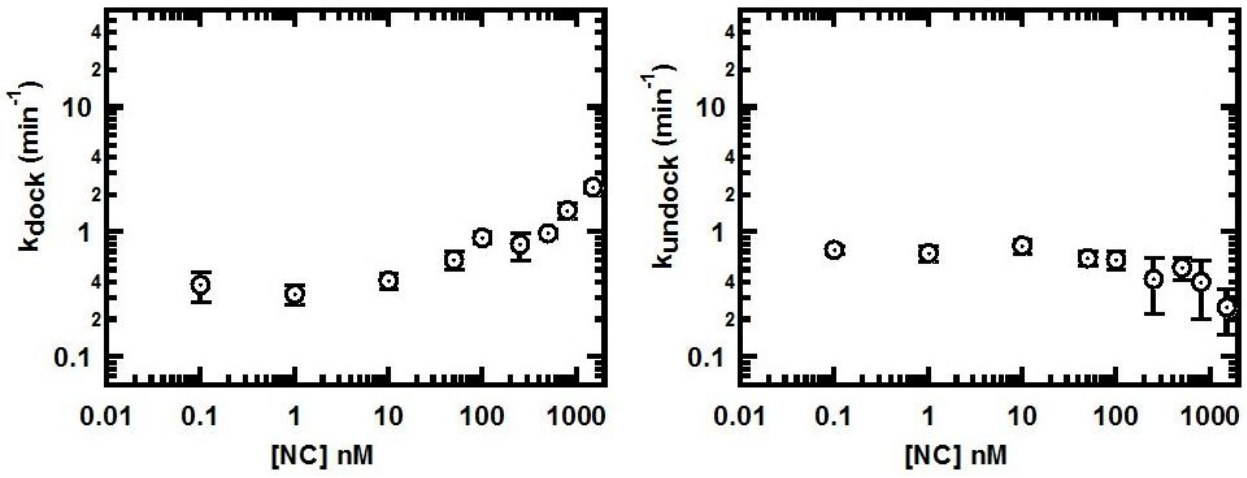

Figure 39: Variation of docking and undocking rates of the bent KC at different NC concentrations. The data unambiguously show that the kinetics of bent KC formation and disappearance is independent of NC.

Bent KC formation rates (k_{dock}) and bent KC undocking rates (k_{undock}) back to KC were studied at varying NC concentrations. Kinetics calculations indicate that at high NC concentrations, KC docking rates are increased and undocking is disfavored. This data unambiguously illustrates that NC binding occurs prior to the bent KC formation and NC is still intact with DIS RNA during KC docking and undocking. Thus dissociation of NC occur latter steps of the dimerization pathway after the bent KC is formed.

4.5 Discussion

The dimeric HIV-1 genome undergoes dimerization during viral replication. We have previously characterized this viral RNA dimerization mechanism with 23 nucleotide long RNA sequences containing HIV-1 dimerization initiation sequence. In addition, we have identified a novel obligatory intermediate in the HIV-1 RNA dimerization pathway in the absence of nucleocapsid protein. Experiments in this chapter have successfully characterized the nucleocapsid-chaperoned dimerization pathway.

Our experiments in the presence of the nucleocapsid protein (Figure 36) unambiguously identify the previously discovered bent kissing complex conformation as an obligatory step in the nucleocapsid catalyzed RNA dimerization pathway. Nucleocapsid protein-RNA dissociation constant was calculated to be 94 nM with smFRET, indicating a strong interaction between hairpin RNA and NC. At high nucleocapsid concentrations, a population with a faster KC formation rate (k'_{on} = 5.1 min^{-1}) appears. This population is dependent on NC indicating that faster KC formation rate appears as a result of NC binding to monomer RNA. Studies on the KC dissociation rates indicate that upon nucleocapsid protein binding; dissociation back to RNA monomer is hindered. Thus, in the presence of nucleocapsid protein, the dwell time in the KC form is increased, favoring KC conversion to bent KC and allowing progression trough the dimerization pathway. Kinetic analysis of docking of the KC into the bent KC and undocking back clearly shows that this process is independent of nucleocapsid protein. Therefore, the removal of NC from RNA might occur during KC bending.

It is evident that the effect of the nucleocapsid protein on HIV-1 RNA dimerization with the 23 nucleotide DIS containing RNA is solely a result of NC binding to initial RNA monomers, inducing a conformational change that fasten the KC formation rates. We have successfully illustrated the NC binding to HIV-1 DIS monomer RNA with surface plasma resonance methodology. SPR experiments suggest that NC binds to HIV-1 DIS monomer RNA with a K_d of 148 ± 50 nM which is comparable with the K_d for NC-DIS interaction obtained from smFRET studies.

We hypothesize that the conformational change is the bulging out of the unpaired adenines, as this faster formation rate effect is not observed with A272C mutated RNA. With the observations from RNA dimerization pathway parameters in the presence of nucleocapsid protein, the proposed NC catalyzed dimerization pathway is illustrated in the following Figure.

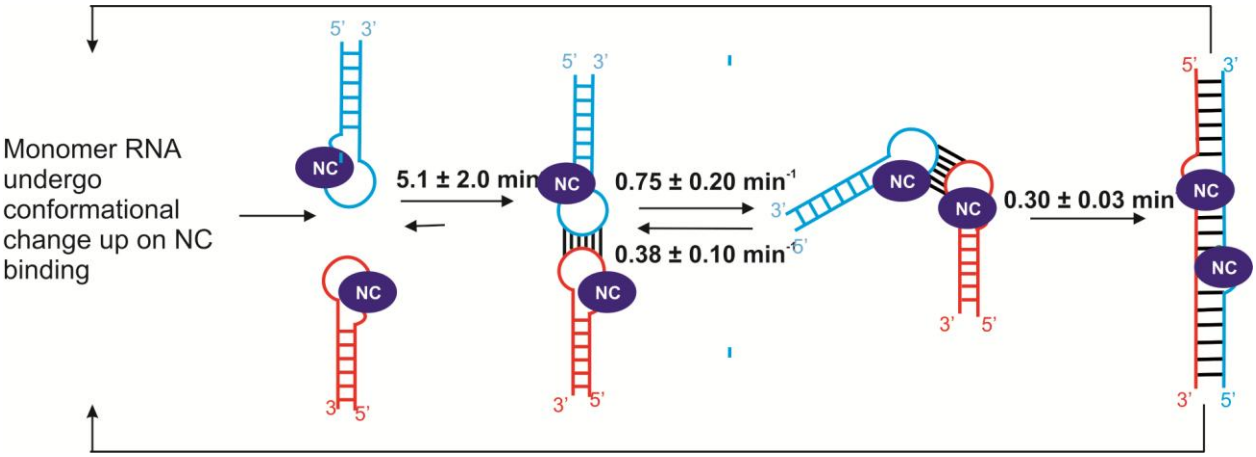

Figure 40: Proposed nucleocapsid-chaperoned HIV-1 RNA dimerization pathway. The initial step is identified as NC binding to RNA monomers inducing a conformational change that leads to faster KC formation rate. In the presence of NC, dissociation back to RNA monomers is hindered. During docking of the KC to form bent KC, NC gets removed from the NC-RNA complex and is available to bind a new monomer RNA.

4.6 Conclusion

Our smFRET data reveal that nucleocapsid binding to HIV-1 RNA occurs at the monomer RNA level. NC binding induces a conformational change in RNA monomer that leads to successful collisions between RNA hairpins, increasing the KC formation rate to 5.1 ± 2.0 min^{-1}. This binding of NC to DIS RNA increases the rate of KC formation ~10 times compared to the non NC mediated KC formation which is 0.3 ± 0.1 min^{-1}. Furthermore, our data suggest that NC bound KC does not readily dissociate

back to monomer RNA hairpins. Instead the dwelling time in the KC form is increased. Our A272C mutant RNA experiments prove that A272 is important for NC chaperoned KC formation.

Studies on the progression of the dimerization pathway from KC onwards reveal that, kinetics of bent KC formation are independent on NC. This leads to the fact that NC dissociation from RNA may occur during KC bending to form bent KC. The dissociated NC can readily bind a new RNA monomer hairpin and continue the chaperone activity.

Chapter 5

Exploring HIV-1 Dimerization Initiation Sequence as a Potential Drug Target

5.1 Aminoglycosides and ribosomal A-site binding

Antibiotics can be classified as broad spectrum and narrow spectrum antibiotics based on their applications [144, 145]. Broad spectrum antibiotics are used to target a wide range of bacteria, from gram positive to gram negative, whereas narrow spectrum antibiotics are used only for a specific phylogenetic group of organisms [144, 145]. Most of the broad spectrum antibiotics target ribosomal RNA. Based on the mode of action and binding sites on the ribosome, these antibiotics are classified as follows. Namely peptidyl-transferase center and exit-tunnel targeting antibiotics, translation relating protein factor targeting antibiotics and decoding site binding antibiotics. Amninoglycoside antibiotics are commonly used broad spectrum antibiotics that bind to the ribosomal decoding site. It has been shown that aminoglycoside binding disrupts tRNA selection mechanism, inhibits mRNA/tRNA translocation and also minimizes ribosome recycling process [146-149].

Aminoglycosides are composed of a 2-deoxystreptoamine (2-DOS) core bearing 1,3 di-amino functional residues and multiple hydroxyl residues [86, 150]. In biological systems, the amino groups are protonated, converting the aminoglycosides into polycations. The degree of protonation varies for different aminoglycosides. Due to protonation, aminoglycosides exhibit a high binding affinity for negatively charged RNA.

Their action mode involves binding to the bacterial ribosomal A-site, disturbing the tRNA selection fidelity during bacterial protein synthesis [88, 146].

Interaction of aminoglycosides neomycin, paromomycin, gentamycin and kanamycin were the first to be studied and it was shown that there are specific interactions with ribosomal decoding center 30s subunit [88]. Neamine, a bicyclic moiety, is a common core to all aminoglycosides [47]. Different aminoglycosides arise from various substitutions to ring 2 of neamine. Neomycin family of aminoglycosides (Paramomycin, neomycin B, ribostamycin and lividomycin A) bear 4,5 disubstitutions in ring 2 and gentamycin and kanamycin family aminoglycosides bear 4,5 disubstitutions in ring 2.

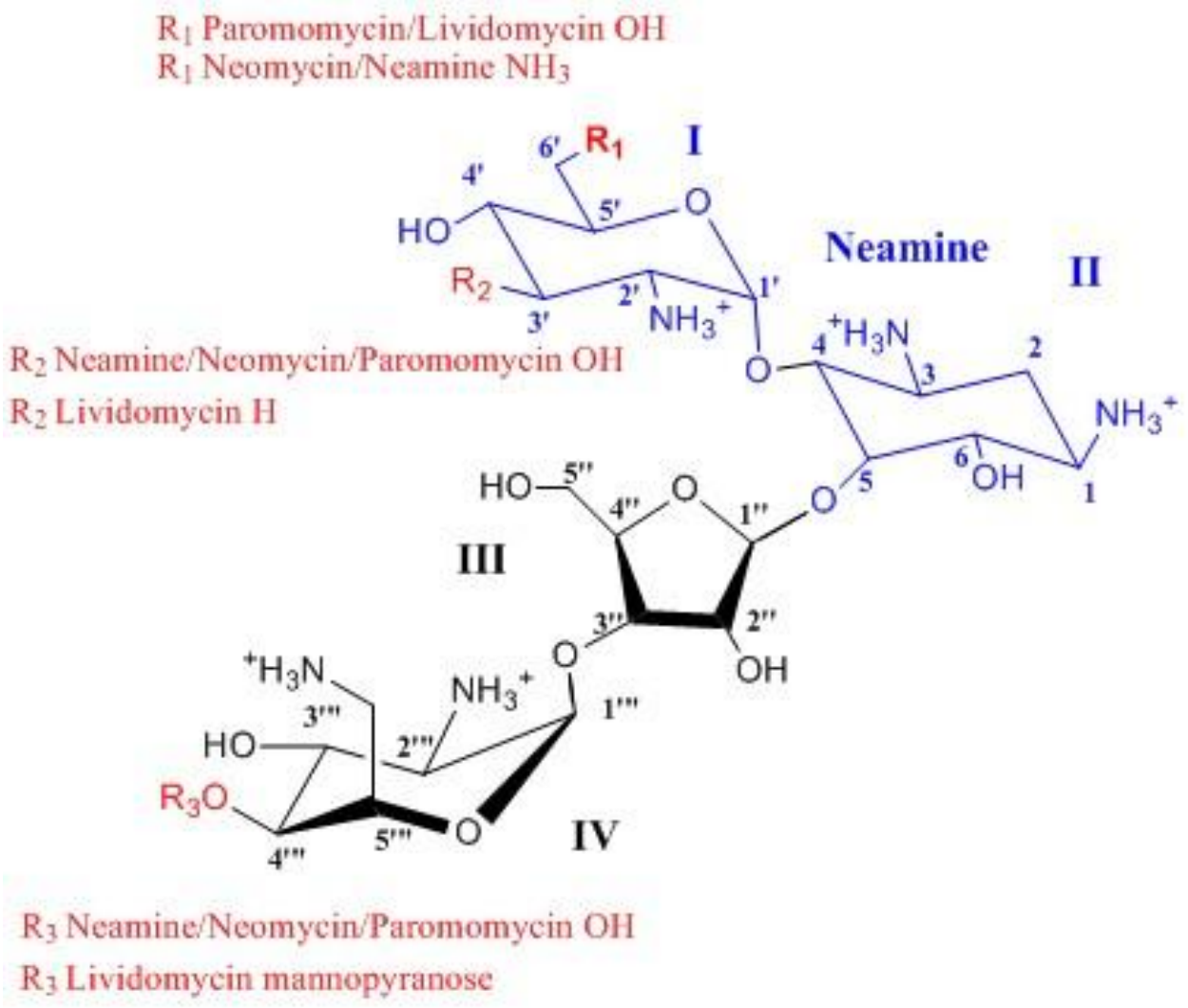

Figure 41: Structure of the common aminoglycoside antibiotics. Neamine core with ring I and II are common and the variations arise from the substitutions [50].

Aminoglycoside interaction with 16s rRNA is well studied due to strong binding potential. It has been revealed from chemical foot printing experiments that several bases in the ribosomal A-site are protected from chemical modifications in the presence of bound aminoglycosides indicating that aminoglycosides bear specific affinities to certain sites during binding [151]. The neamine ring alone exhibits eight direct contacts with ribosomal A-site according to the crystal structure information [151]. The number of

interactions between the A-site and extended aminoglycosides varies from 21-30 depending on the side groups present [151].

5.2 HIV-1 DIS is Structurally and Sequentially Similar to 16s Ribosomal A-site

It has been previously revealed that the nucleotide sequence and the secondary structure of the bacterial ribosomal A-site (base sequence from 1406-1410 and 1490-1495) are explicitly similar to the HIV-1 dimerization initiation site dimeric sequence (base sequence 278-282 and 270-275). Both the HIV-1 DIS kissing complex and the extended RNA duplex resemble the ribosomal A-site structure and sequence as shown in figure 14. Based on the HIV-1 subtype being compared, minor differences between A-site and the DIS dimer arise. HIV-1 DIS subtype F with A 273 and A275 residues has the closest resemblance to ribosomal A-site. In addition to the similarities exhibited at primary and secondary structures, HIV-1 DIS kissing complex has a three dimensional structure similarity to the ribosomal A-site bound to paromomycin which is a neomycin family aminoglycoside [50]. Comparison of the two crystal structures shows that the two bulged adenine nucleotides stacked on each other across the single unpaired adenine residue which is stacked inside the duplex helix is a common feature in both complexes [68, 152]. Molecular modeling studies have shown that the two bulged out adenine residues A272 and A273 in the HIV-1 DIS sequence are required to adopt a complete ribosomal A-site like structure [50].

5.3 Aminoglycosides - HIV-1 DIS Interaction

Based on the similarities observed between HIV-1 DIS and ribosomal A-site crystal structures, it is expected to observe a high binding affinity of aminoglycosides to DIS. Recently, multiple aminoglycosides complexed with HIV-1 DIS kissing complex have been crystallized [47, 50]. It has been shown that most of the aminoglycosides bind to DIS with several orders of magnitude higher affinities than their original target.

We selected four antibiotics with varying affinity to DIS, namely neomycin, Paromomycin, Lividomycin and neamine (synthesized neamine was provided by Jun Jiang from Santalucia Lab). The published dissociation constants for these aminoglycosides are listed in table 2 (chapter 1).

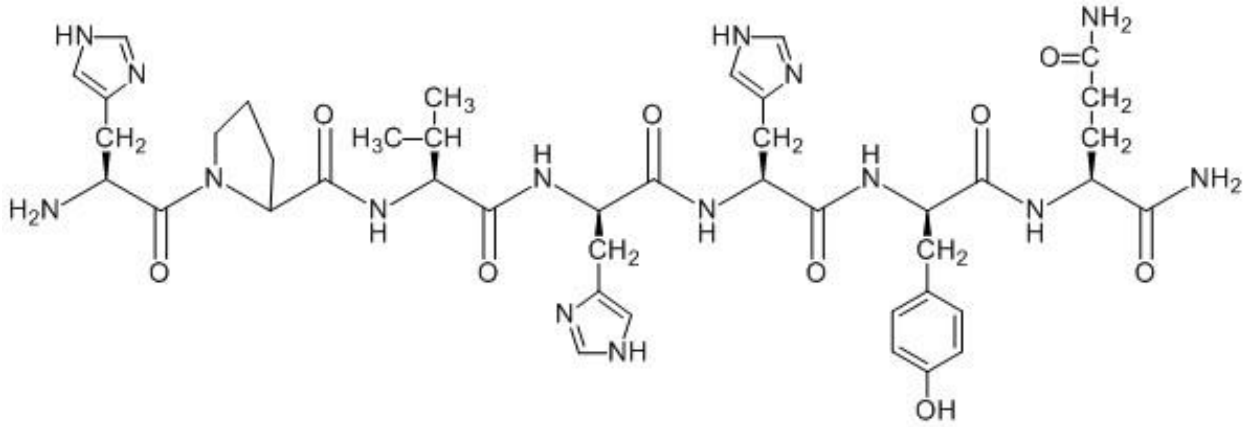

Figure 42: Structure of the HPVHHYQ peptide used for the study

We have also studied the effect of the heptapeptide HPVHHYQ-NH_2, which exhibits moderate affinity for the ribosomal A-site [55]. HPVHHYQ peptide was synthesized by Danielle Dremann (Chow lab, Department of Chemistry, Wayne State University). Binding studies with A-site of 16s rRNA via foot printing, ESI-MS, circular

dichroism (CD) and ITC have proven a binding affinity of 16 µM [107]. Thus, based on the previously discussed similarities between ribosomal A-site and HIV-1 DIS RNA dimer, we have extended our studies to investigate the affinity of HPVHHYQ for DIS RNA. Studies on how these antibiotics and peptides interfere with the dimerization pathway for HIV-1 DIS will be useful for selecting and developing target molecules with significantly high specificity and affinity to HIV-1 RNA *in vivo*. Further insights gained from these experiments will be useful in understanding DIS as a potential drug target.

5.5 Results

5.5.1 Neamine

5.5.1.1 Neamine Binds to DIS Monomer RNA

smFRET experiments were conducted at different neamine concentrations. The kinetic parameters of dimerization pathway were calculated at each neamine concentration to investigate the effect of neamine on HIV-1 RNA dimerization with minimal RNA constructs in vitro. Thus, kissing complex formation rate, dissociation rate, rate of KC docking to form bent KC and undocking of bent KC to form KC was studied. Additionally, K_d of neamine binding to DIS kissing dimer was also determined by analyzing the variation of the relative population for all the complexes.

Experiments in the presence of neamine show that neamine binds to HIV-1 DIS with a K_d of 10± 2 µM. Analysis of the kissing complex formation rates clearly indentifies the neamine binding to initial DIS RNA monomers. Neamine interaction with DIS RNA gives rise to a population with a faster kissing complex formation rate of 1.95 ± 0.25

min^{-1}, which is significantly higher than the kissing complex formation rate in the absence of neamine. The population distribution analysis clearly indicates that the KC formed with $k'_{on,2}$ is dependent on the concentration of neamine. Therefore, this observation suggests that neamine binding induces a rearrangement of the DIS monomer RNA conformation in order to facilitate the kissing complex formation.

5.5.1.2 Neamine Facilitates the ED Formation by Shifting Monomer KC Equilibrium

Although neamine binding facilitates kissing complex formation, it does not show an effect on the kissing complex stability. KC dissociation rates remain unchanged even at high neamine concentrations. But reaction speciation results suggest that in the presence of neamine, ED formation is facilitated. A gradual decrease in KC fraction, followed by an increase in bent KC and ED fraction is observed as the neamine concentration is raised. We propose that the extended RNA duplex formation enhancement is a result of the equilibrium shift due to the fast formation of kissing complex arising from neamine binding.

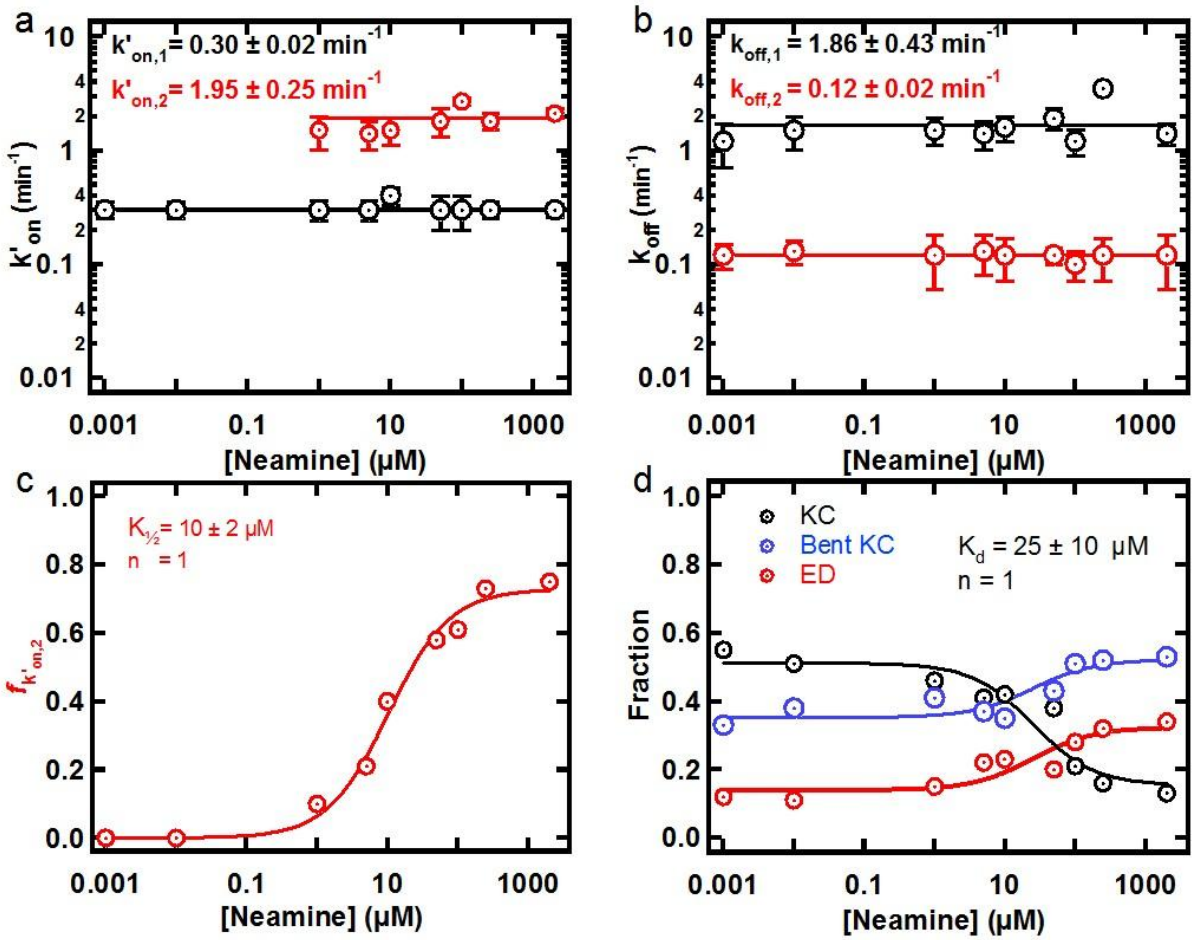

Figure 43: Effect of Neamine on the variation of (a) Formation rates (b) dissociation rates and (c) population of KC formed with $k'_{on,2}$ (d) Effect on the molecule populations in the presence of Neamine

5.5.2 Neomycin

5.5.2.1 Neomycin Binds to Individual DIS RNA Monomers

Experiments were conducted under increasing concentrations of neomycin to analyze the effect of neomycin on dimerization pathway kinetics. In the presence of neomycin, a faster kissing complex formation rate is observed, similar to previously described neamine. This result is not surprising as the neamine ring is common to all the neomycin family antibiotics. Reaction speciation studies suggest that the population

of kissing complexes formed with $k'_{on,2}$ is a direct result of neomycin binding to RNA monomer.

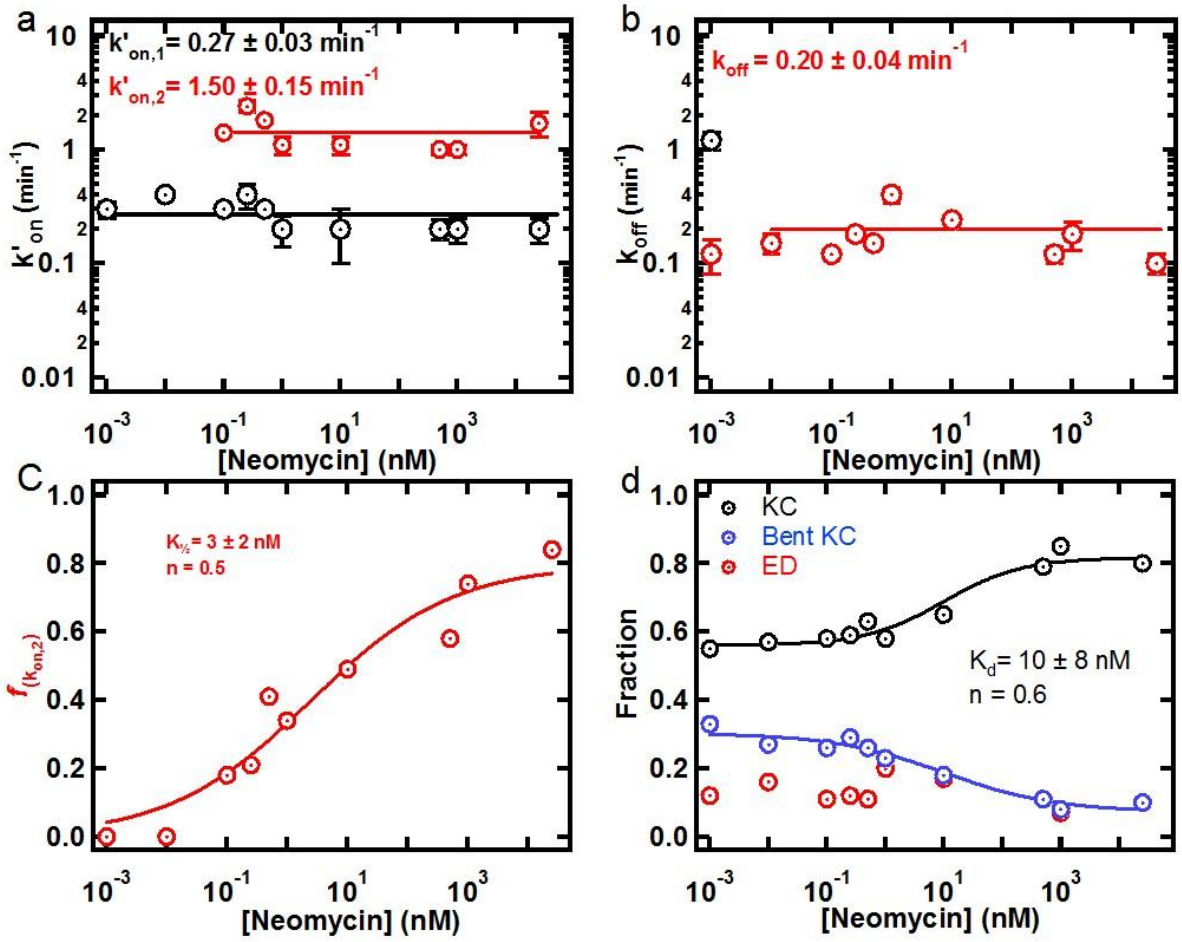

Figure 44: Effect of Neomycin on the variation of (a) Formation rates (b) dissociation rates and (c) population of KC formed with $k'_{on,2}$ (d) Effect on the molecule populations in the presence of Neomycin

5.5.2.2 Neomycin Hinders HIV-1 DIS Dimerization by Stabilizing KC Form

In contrast to neamine, neomycin exhibits a stabilizing effect on the HIV-1 DIS kissing complex. In addition, neomycin binds very tightly with a K_d of 10 ± 8 nM. Studies of KC dissociation rates indicate that neomycin-bound KC is extra stabilized due to neomycin mimicking the function of Mg^{2+}. Relative population fractions at various

neomycin concentrations show that in the presence of neomycin, kissing complex to bent kissing complex transition is hindered. At neomycin concentrations above 100 nM, 80% of RNA are in the kissing complex form with only 10% of RNA in the bent KC form. Compared to the near physiological conditions in the absence of antibiotics, KC population is doubled and the bent KC population is 4 times less populated (table 4, chapter 3). Thus, the progression of HIV-1 RNA dimerization is effectively stalled at the kissing complex stage. We propose that binding of neomycin might sterically hinder the conformational changes in the loop-loop interface required for kissing complex bending due to the bulky extended structure of neomycin, thus hindering the conversion of kissing complex to bent kissing complex conformation.

5.5.3 Lividomycin

5.5.3.1 Lividomycin Binds to DIS Monomer RNA and Stabilize the KC

Lividomycin exhibits the strongest binding affinity to HIV-1 DIS complex of the four antibiotics studied with a K_d of 0.2 nM. Similar to previously studied neomycin and neamine, lividomycin also increases the rate of KC formation as expected due the presence of neamine ring structure. In the presence of lividomycin, the strength of kissing complex interaction is greatly enhanced and only the meta-stabilized kissing complexes with a dissociation constant of 0.12 ± 0.04 min^{-1} are observed. The molecule population analysis clearly shows that the binding of lividomycin completely hinders the conversion of kissing complex dimers to the bent kissing complex conformation. Lividomycin effectively reduces KC to bent KC conversion at concentrations above 10

nM. At high concentrations of lividomycin (10 nM), about 80% of kissing complex dimers are observed with 1-5% of bent kissing complex dimers.

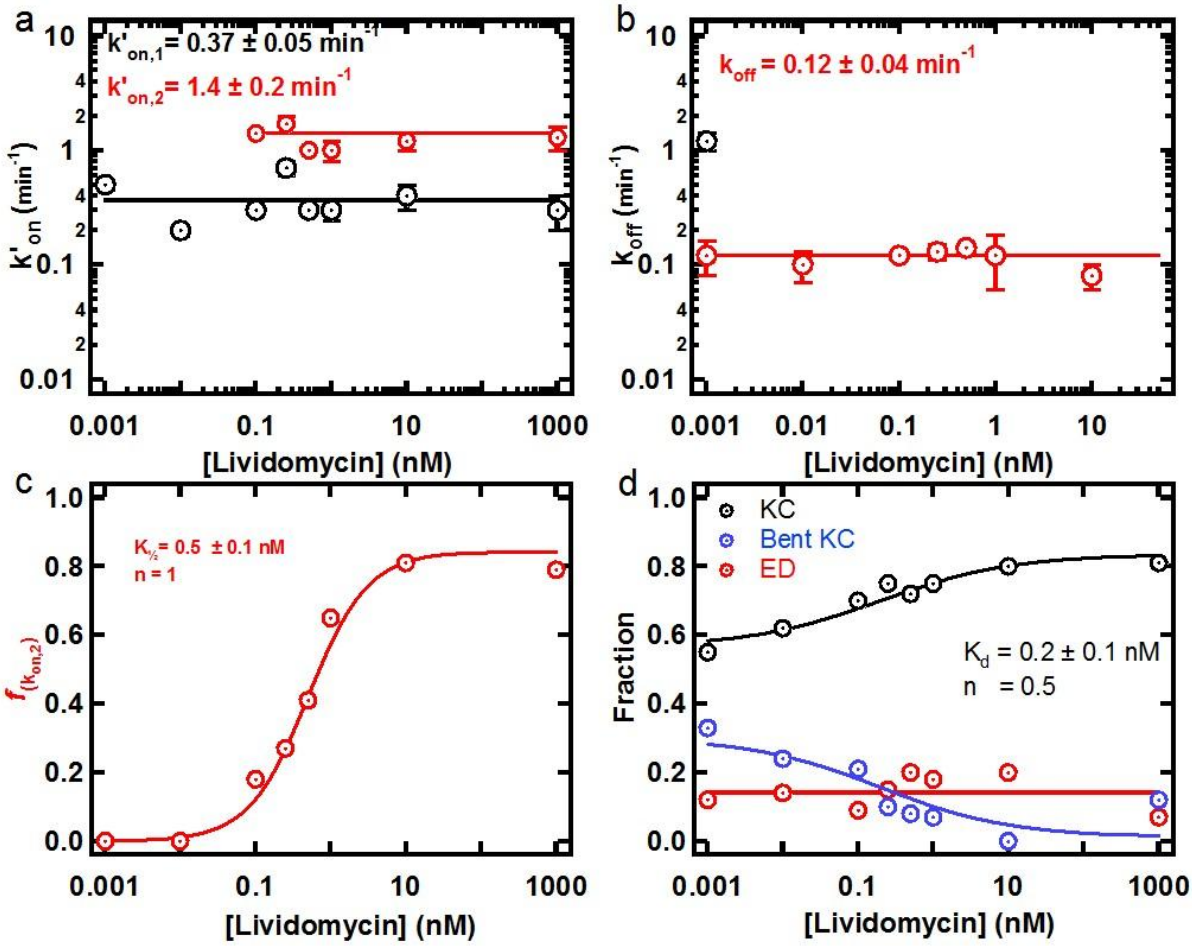

Figure 45: Effect of Lividomycin on the (a) Kissing complex formation rates (b) Kissing complex dissociation rates and (c) population of KC formed with $k'_{on,2}$ (d) Effect on the molecule populations in the presence of Lividomycin.

5.5.4 Paromomycin

5.5.4.1 Paromomycin Binds to Individual DIS Hairpin RNA and Stalls Dimerization at KC Form

Paromomycin exhibits similar effects as neomycin and lividomycin although binding affinities differ. The K_d of Paromomycin binding to HIV-1 DIS RNA is 0.25 µM. Paromomycin binding results in the appearance of a kissing complex population with faster formation rates (1.3 min^{-1}). Kissing complex to bent kissing complex equilibrium is disturbed upon binding to paromomycin and forward reaction is hindered.

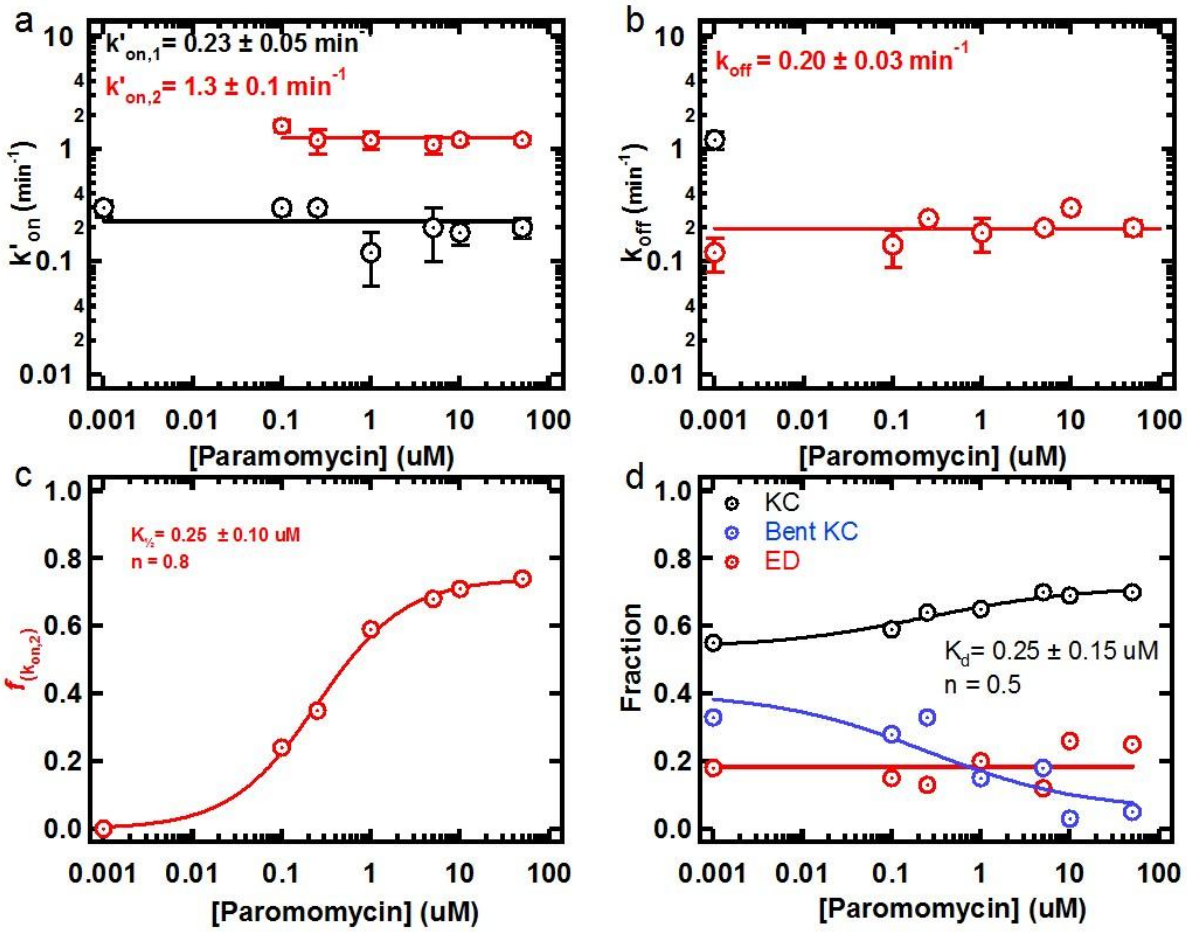

Figure 46: Effect of Paromomycin on the variation of (a) Formation rates (b) dissociation rates and (c) population of KC formed with $k'_{on,2}$ (d) Effect on the molecule populations in the presence of Paromomycin

We have observed that in the presence of neamine, neomycin, Paromomycin and Lividomycin antibiotics, a population with a faster KC formation rate appears (Figures 39-41). However, studies on dissociation rates show that neamine has no effect on the kissing complex stability while neomycin, Paromomycin and Lividomycin significantly stabilize the KC population. Analysis of differential conformational populations (Kissing complex, bent kissing complex and extended RNA duplex) clearly distinguish neamine as a dimerization accelerator whereas neomycin, paromomycin and lividomycin can be categorized as dimerization inhibitors.

5.5.5 HPVHHYQ Binds HIV-1 DIS RNA

HPVHHYQ-NH2 exhibits a moderate affinity for the bacterial A-site RNA, with an average K_d value of 16 μM [107]. From electro spray ionization, mass spectrometry (ESI-MS) data a 1:1 stoichiometric ratio for A-site binding is observed suggesting a quite specific binding. The peptide has shown binding to internal bulge in A-site RNA Studies were done to investigate whether HPVHHYQ binds with a significant affinity to DIS RNA.

We extended the studies on HPVHHYQ peptide to study the binding of peptide to HIV-1 DIS RNA based on the structural similarities observed between HIV-1 DIS dimer and ribosomal A-site. We studied the effect of HPVHHYQ on kissing complex formation rates, dissociation rates, docking rates of KC to form bent KC and undocking rates of KC to characterize the interaction between the peptide and DIS RNA.

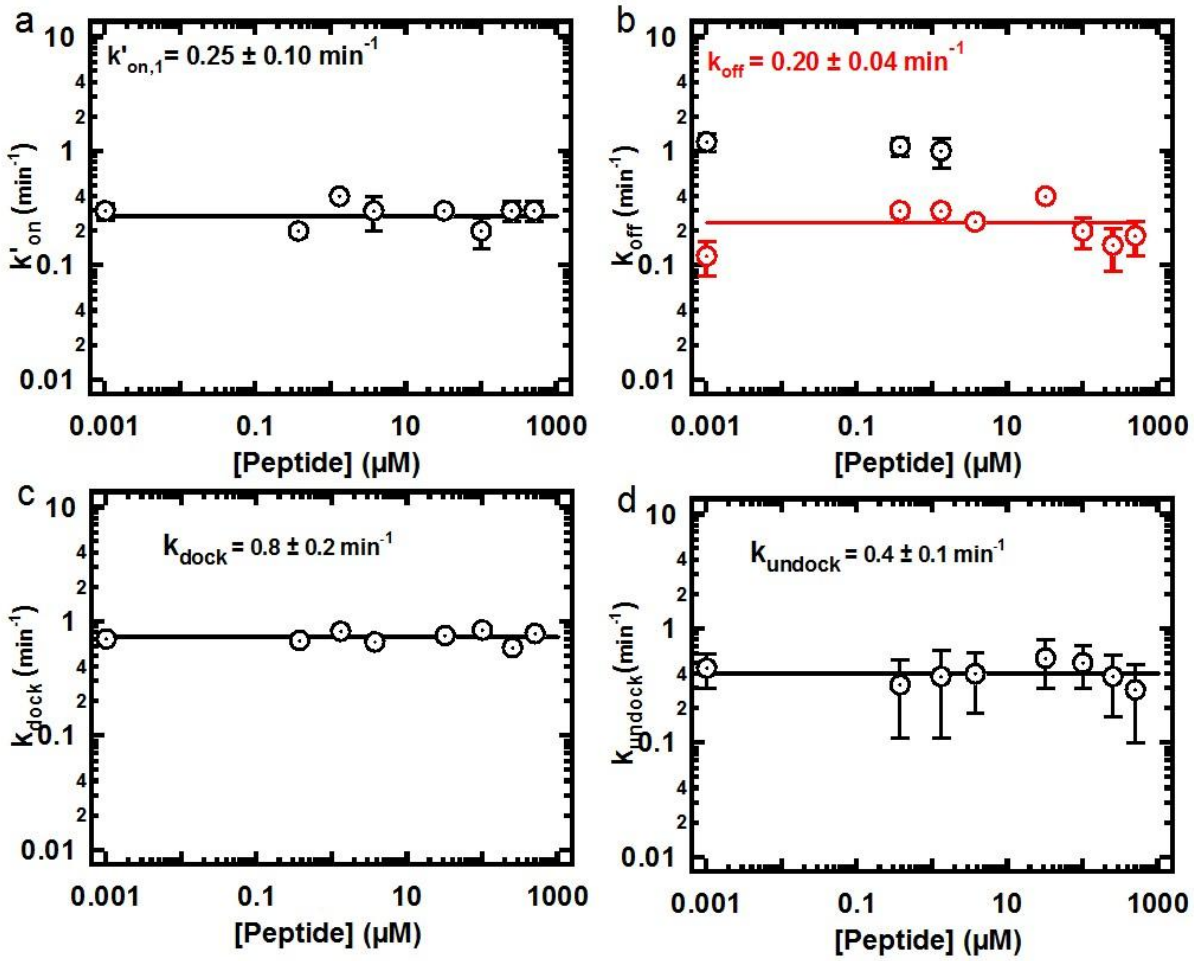

Figure 47: Effect of HPVHHYQ on dimerization pathway. (a) KC formation rates are unaffected by the presence of peptide indicating that peptide does not interact with monomer RNA. (b) A stabilization of KC was observed at high concentrations of HPVHHYQ. (c) and (d) KC docking in to BKC and undocking back to KC is not significantly affected by the presence of peptide. Altogether, the data suggests that the interaction of HPVHHYQ stabilize the KC, but does not significantly affect any the steps in the dimerization pathway.

smFRET studies on KC formation rates shows that, KC formation was not significantly affected by the presence of HPVHHYQ. But in the presence of the peptide, increased stability of the KC is observed similar to antibiotics. This suggests that HPVHHYQ did not bind to monomer hairpins and instead binds to the already formed KC, stabilizing the KC form. Our SPR experiments on HPVHHYQ binding to DIS RNA

monomer supported the smFRET data indicating no significant binding of HPVHHYQ to DIS monomer RNA. In contrast to data obtained from aminoglycosides, binding of HPVHHYQ does not significantly block the dimerization pathway. The rates of KC to BKC conversion is not significantly affected by HPVHHYQ.

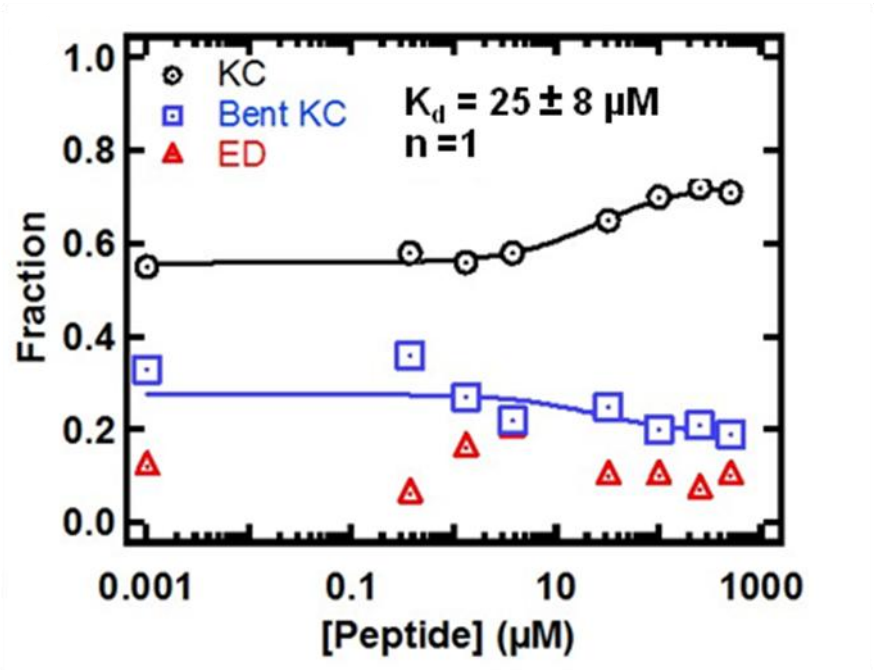

Figure 48: Effect of HPVHYQ on reaction speciation. Although a slight increase in KC fraction is observed with increased amounts of peptide, the progression of dimerization is not significantly affected with HPVHHYQ

Specificity of HPVHHYQ-DIS interaction was confirmed by conducting control experiments with helix 69 binding heptapeptides. Helix 69 binding peptides do not have any effect on the DIS dimerization pathway, according to the preliminary ESI experiments and smFRET experiments indicating that HPVHHYQ interaction is specific.

5.6 Discussion

It has been previously suggested that binding of neomycin and paromomycin displace a Mg^{2+} that is already bound to the loop hex nucleotide and results in extra stabilization of the kissing complex [68]. There has been a debate on whether the aminoglycosides bind to monomer RNA or to the already formed RNA dimer [50, 153]. Our data support that aminoglycosides bind to individual hairpins and induce the conformation changes (loop rearrangement for successful KC formation, including adenine flipping) that are required for successful collisions between hairpins. When the dissociation rates are considered, our data indicate that aminoglycosides with high ionic charge (neomycin, paromomycin and lividomycin) mimic the function of Mg^{2+} by displacing the loop bound Mg^{2+}. This displacement results in a meta-stabilized KC and a stalling trough the dimerization reaction pathway at the KC stage. Neomycin and lividomycin exhibit the highest affinity for DIS RNA and display significant dimerization inhibition. Based on the concept of binding to DIS RNA monomer, it can be stipulated that aminoglycoside-DIS complex bear specific interactions in addition to the adenine bulge interactions expected as a result of DIS and rRNA A-site similarity.

An increase in RNA KC formation rates was observed with all the four aminoglycosides studied, including the bicyclic neamine. Considering all four aminoglycosides present a common neamine core, we identify neamine as the primary interaction site with HIV-1 DIS RNA.

In contrast to the KC formation rates, stabilization effects on KC are exhibited only in the presence of neomycin, lividomycin and paromomycin. This clearly suggests

that interactions responsible for KC stabilization arise from the functional groups attached to the neamine core.

Considering the cooperativitiy values obtained from smFRET data analysis, it is evident that there is a relationship between the observed cooperativities and the aminoglycoside structure. Neamine being the smallest antibiotic, it exhibits a cooperativity of 1, whereas more bulky neomycin, paromomycin and lividomycin all exhibit cooperatively values close to 0.5. This result can be explained as follows. At saturating antibiotic concentrations, each RNA hairpin is bound to one antibiotic molecule at all times. Therefore, a newly formed KC will initially be bound to two antibiotic molecules. We propose that the formation of the KC can result in the displacement of one of the antibiotic molecules. This effect is only present with bulky aminoglycosides as the cooperativity value of one observed for the small neamine suggests that the KC can accommodate two neamines, but has only enough space for one of the other more bulky antibiotics.

The exceptional dimerization acceleration induced by neamine can be explained as follows. With a relatively weak RNA-neamine interaction (K_d = 25 μM), neamine-KC can be expected to have a fast dissociation rate, which increases [KC] at a given time compared to other antibiotics. This will allow the KC to proceed to bent KC and ED, thus shifting the KC and neamine-KC equilibrium to the bound form.

The calculated dissociation constants for Aminoglycoside- HIV-1 DIS RNA show string binding with lower binding constants when compared with previously documented values derived from bulk experiments [73]. Previously documented binding constants

were a result of ITC experimental methods which is measured at a relatively high RNA concentration in order to obtain a significant amount of heat exchange during binding experiments. On the other hand experimental conditions involved in our smFRET study utilized very low concentrations of RNA to achieve single-molecule immobilization. This is the major reason for the discrepancy among previously observed binding constants and the K_d values obtained from this research. Additionally, our binding constant calculations were done by considering the bent kissing complex population as a novel reaction pathway intermediate unlike previous bulk experimental techniques. Thus, differences may arise from the fact that for calculating the K_d in this work the population fraction of the bent kissing complex was also considered.

Chapter 6

Conclusions and Future Directions

Single-molecule fluorescence resonance energy transfer has a unique combination of advantages over conventional bulk experimental techniques. They are as follows: no loss of information due to ensemble averaging, capability to gain kinetic information from equilibrium experiments, ability to observe intermediates that do not accumulate over time, and the possibility of observing multiple folding pathways [154]. Additionally, based on the fluorophores used in this dissertation work (Cy3-5), the distance dependency of FRET bimolecular changes occurring between 10-100 Å can be monitored using smFRET methods. Based on these advantages, smFRET in combination with prism-based TIRF is employed in this dissertation work. During the later stages of the HIV-1 life cycle, the gag polyprotein complex chaperones a series of steps leading to the generation of infectious new viral particles from the host cell plasma membrane. One important step carried out during this stage is the packaging of two copies of the HIV-1 genomic RNA into newly budding viral particles. In order for this mechanism to be controlled effectively, HIV should possess a high packaging efficiency while maintaining the RNA copy number in each virion. Thus, based on the importance of the RNA packaging step on the HIV-1 life cycle, the purpose this dissertation work was to shed light on the mechanism of the dimerization pathway, the role of nucleocapsid protein, and the development of DIS-targeting molecules. Thermodynamics and kinetics of the folding pathway of the HIV-1 dimerization initiation site RNA in the presence and absence of HIV-1 nucleocapsid protein were studied with TIRF-based single-molecule microscopy. Further, the studies were extended to

investigate the binding of aminoglycoside antibiotics and small peptides on the HIV-1 RNA dimerization pathway. These ligands are known to bind the ribosomal A-site RNA, but were shown to impact the DIS folding pathway.

Characterization of HIV-1 genomic RNA dimerization *in vitro* with minimal RNA sequences is the first step in understanding viral RNA dimerization during viral assembly. Thus, for this study, 23-nucleotide sequences that mimic HIV-1 subtype F DIS were used. The results show that the initial hairpin RNA interaction is a diffusion-controlled event that is independent of Mg^{2+} and increases linearly upon increasing RNA concentration. The importance of Mg^{2+} in HIV-1 DIS dimerization has been illustrated by a number of bulk-analysis techniques [55, 64, 65, 135]. Supporting these studies, work in this dissertation reveals that, in solution, magnesium binds DIS with an equilibrium dissociation constant near 5 mM. Calculations based on the kinetic data show that Mg^{2+} binding to the kissing complex results in a stabilization of 5 ± 3 kcal/mol. More importantly, smFRET analysis reveals that HIV-1 RNA dimerization occurs through a three-step folding pathway in which the RNA kissing complex shifts to a bent, kissing conformation and leads to the formation of the extended RNA duplex via interaction through the stems. We propose a 3D model of a possible bent DIS kissing complex intermediate as expected from smFRET data. Alternatively, the bent intermediate may resemble a structure suggested in a recent NMR study, in which the DIS hairpin base pairing remains intact and inter-stem interactions are facilitated as a result of KC bending, which brings the two stems in proximity [132]. In addition, there is *in vivo* evidence for an intermediate HIV-1 genomic RNA (gRNA) dimer on the path from immature gRNA dimer to mature gRNA dimer inside the HIV-1 particle [133].

During viral maturation process, nucleocapsid protein is involved in the dimerization process. Thus, the next goal of this dissertation work was to investigate alterations to the dimerization pathway in the presence of nucleocapsid protein. The calculated kinetics of kissing complex formation in the initial step of the dimerization were extremely slow, and should be expected to be enhanced under *in vivo* conditions. The rate enhancement is a result of protein counterparts chaperoning the folding pathway and also intracellular crowding conditions [155, 156]. In support of this idea, the addition of nucleocapsid does show a ten-fold higher rate enhancement of kissing complex formation at concentrations above 10 nM. A bent KC transition-state similar to the bent KC intermediate observed in the first part of this study has also been proposed in the nucleocapsid-chaperoned dimerization pathway [64]. Thus, the studies were extended to look for the behavior of the bent KC intermediate in the presence of nucleocapsid protein. Contradictory to the transition-state hypothesis [64], this study shows that a bent KC intermediate exists in the presence of nucleocapsid protein, although the docking and undocking kinetics are slightly altered. It shows that, with increasing nucleocapsid protein conditions,the rate of KC docking to a bent KC is increased gradually, while the undocking process is disfavored. Additionally, this dissertation work highlights the important role of the universally-conserved adenine 272, in the loop for nucleocapsid protein binding. Experiments performed with the DIS A272C mutant show that the mutant RNA is unable to interact with nucleocapsid, indicating that A272 is a probable nucleocapsid binding site. Nucleocapsid protein stabilizes the KC form, indicating that facilitation of the extended RNA duplex formation is achieved by shifting the KC and monomer RNA equilibrium towards the KC. Comparison of KC

stability in the presence of nucleocapsid protein further suggests that Mg^{2+} mimics the role of protein under *in vitro* conditions.

In the secondary structure of the HIV-1 DIS neighboring region, a stem bulge located downstream DIS is suggested by a few studies to be involved in nucleocapsid protein binding, despite the fact that most of the *in vitro* bulk studies on HIV-1 have proved that the minimal RNA construct required to mimic the dimerization of HIV-1 RNA consists of a 23-nucleotide upper stem-loop structure [48, 69, 70, 157]. In addition,a few studies have suggested that a 35-39 nucleotide region containing a four-nucleotide bulge region and DIS stem-loop structure is required to enable the two-step dimerization of HIV-1 genomic RNA [65, 158, 159]. On the other hand, *in vivo* studies performed on this system have suggested that both the upper stem and the lower stem of the stem bulge-stem-loop structure have partial contributions to the RNA-RNA interaction [160, 161], although the exact role played by the lower stem bulge is poorly understood. Thus, as a future direction of this dissertation work, we propose to investigate the kinetic and thermodynamic parameters of dimerization on the introduction of the downstream stem bulge both in the presence and absence of NC protein.

The next parts of this dissertation were aimed at identifying potential DIS-targeting molecules and evaluating their roles in the dimerization pathway. Initially, aminoglycoside antibiotics, which have been thoroughly studied for their interactions with RNA, were selected for the study [47, 162]. Four neomycin family aminoglycosides, namely neamine, neomycin, paromomycin, and lividomycin, were selected based on their previously published affinities for DIS KC and extended RNA duplex [47, 162].

Studies with these selected aminoglycosides at the single-molecule level reveals that they bind to HIV-1 DIS with strong binding affinities (neomycin K_d= 10 nM, paromomycin K_d = 0.25 µM, lividomycin K_d = 0.2 nM). The data clearly show that all four selected aminoglycosides initially bind to DIS RNA monomers. Considering the common neamine core in all four species, it is likely that the neamine core binds to individual hairpins and induces a conformational change, which is required for successful collisions between hairpins. However, in contrast to this interaction with hairpin RNA, a KC stabilization effect is exhibited only by neomycin, paromomycin and lividomycin antibiotics. Thus, extensions to the neamine core at the fifth position of ring II must be required for the KC stabilization. Additionally, analysis of cooperativity values suggests that although initial binding of aminoglycosides is at a 1:1 stoichiometry with RNA hairpins, steric effects on the KC loop interface upon KC formation result in dissociation of one aminoglycoside moiety. Further studies on aminoglycoside binding reveal that the strong KC stabilization effect hinders dimerization progression in the presence of neomycin, paromomycin and lividomycin antibiotics at concentrations equal or greater than the K_d values. At aminoglycoside concentrations that are greater than the K_d values, a complete stalling of dimerization is observed and specifically, KC to bent KC conversion is hindered. This observation can be considered as a preliminary step towards the development of DIS-targeting molecules. We hypothesize that in the presence of aminoglycosides, kissing complex (KC) is kinetically trapped due to the high energy barrier between the KC and the bent KC. Thus, as a future goal in continuation of this work, we propose to use laser assisted single molecular refolding

(LASR) to determine the energy barrier between the two folded states of the RNA dimer [117].

The drawback of using aminoglycoside antibiotics is that they interact with multiple RNA domains in the cell such as ribosome. Thus, these molecules cannot be considered as specific DIS-targeting molecules. Therefore, studies should be focused on increasing the specificity of aminoglycosides to DIS, and evaluating the effect *in vivo*. Multiple strategies could be used to increase the specificity of aminoglycosides to DIS including conjugation with specific DIS-binding molecules. Nucleobase-aminoglycoside conjugates and antisense RNA-aminoglycoside conjugates are ideal candidates with high specificity to a unique RNA target. Thus, as an important future direction of this study is an evaluation of the aminoglycoside effect *in vitro* and *in vivo* with DIS-specific aminoglycoside conjugates. The specificity to DIS could be verified by observing the binding affinities of aminoglycoside-conjugates with the ribosomal A-site. Upon the development of specific DIS targets, *in vivo* effect of these conjugates should be evaluated. Further, work in this dissertation identified aminoglycosides that posses HIV-1 RNA dimerization pathway hindering ability. As a next step towards the development of DIS-targeting molecules, identification of aminoglycoside residues that interact with DIS leading to a strong KC stabilization could be performed via crystal structure analysis.

The mechanistic insights gained from these experiments represent significant progress towards understanding the HIV-1 dimerization mechanism, and unveils the presence of a key obligatory intermediate in the dimerization reaction of kissing

complexes in RNA. Further step-by-step analysis of the kinetic and thermodynamic aspects of nucleocapsid protein catalyzed dimerization pathway provides new directions to the field. This work addresses the importance of the fundamental issues in RNA folding, as well as the potential to exploit DIS as a possible HIV drug target. This work represents significant progress towards understanding the HIV-1 dimerization process, as well as the identification and development of DIS-targeting molecules.

REFERENCES

1. Crick, F., *Central dogma of molecular biology.* Nature, 1970. 227(5258): p. 561-3.
2. Kapp, L.D. and J.R. Lorsch, *The molecular mechanics of eukaryotic translation.* Annu Rev Biochem, 2004. 73: p. 657-704.
3. Kramer, A., et al., *Components involved in nuclear pre-mRNA splicing.* Mol Biol Rep, 1990. 14(2-3): p. 199-200.
4. Eddy, S.R., *Non-coding RNA genes and the modern RNA world.* Nat Rev Genet, 2001. 2(12): p. 919-29.
5. Pyle, A.M., *Ribozymes: a distinct class of metalloenzymes.* Science, 1993. 261(5122): p. 709-14.
6. Millward, S. and A.F. Graham, *Structural studies on reovirus: discontinuities in the genome.* Proc Natl Acad Sci U S A, 1970. 65(2): p. 422-9.
7. Cech, T.R., et al., *Secondary structure of the Tetrahymena ribosomal RNA intervening sequence: structural homology with fungal mitochondrial intervening sequences.* Proc Natl Acad Sci U S A, 1983. 80(13): p. 3903-7.
8. Guerrier-Takada, C., et al., *The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme.* Cell, 1983. 35(3 Pt 2): p. 849-57.
9. Lai, E.C., *microRNAs: runts of the genome assert themselves.* Curr Biol, 2003. 13(23): p. R925-36.
10. Robertson, M.P. and G.F. Joyce, *The origins of the RNA world.* Cold Spring Harb Perspect Biol, 2012. 4(5).

11. Cruz, J.A. and E. Westhof, *The dynamic landscapes of RNA architecture.* Cell, 2009. 136(4): p. 604-9.

12. Batey, R.T., R.P. Rambo, and J.A. Doudna, *Tertiary Motifs in RNA Structure and Folding.* Angew Chem Int Ed Engl, 1999. 38(16): p. 2326-2343.

13. Pyle, A.M., *Metal ions in the structure and function of RNA.* J Biol Inorg Chem, 2002. 7(7-8): p. 679-90.

14. DeRose, V.J., *Metal ion binding to catalytic RNA molecules.* Curr Opin Struct Biol, 2003. 13(3): p. 317-24.

15. Shiman, R. and D.E. Draper, *Stabilization of RNA tertiary structure by monovalent cations.* J Mol Biol, 2000. 302(1): p. 79-91.

16. Woodson, S.A., *Metal ions and RNA folding: a highly charged topic with a dynamic future.* Curr Opin Chem Biol, 2005. 9(2): p. 104-9.

17. Das, R., et al., *The fastest global events in RNA folding: electrostatic relaxation and tertiary collapse of the Tetrahymena ribozyme.* J Mol Biol, 2003. 332(2): p. 311-9.

18. Hougland, J.L., et al., *Functional identification of catalytic metal ion binding sites within RNA.* PLoS Biol, 2005. 3(9): p. e277.

19. Draper, D.E., *RNA folding: thermodynamic and molecular descriptions of the roles of ions.* Biophys J, 2008. 95(12): p. 5489-95.

20. Sontheimer, E.J., *The spliceosome shows its metal.* Nat Struct Biol, 2001. 8(1): p. 11-3.

21. Rajkowitsch, L., et al., *RNA chaperones, RNA annealers and RNA helicases.* RNA Biol, 2007. 4(3): p. 118-30.

22. Schroeder, R., A. Barta, and K. Semrad, *Strategies for RNA folding and assembly.* Nat Rev Mol Cell Biol, 2004. 5(11): p. 908-19.

23. Herschlag, D., *RNA chaperones and the RNA folding problem.* J Biol Chem, 1995. 270(36): p. 20871-4.

24. Lorsch, J.R., *RNA chaperones exist and DEAD box proteins get a life.* Cell, 2002. 109(7): p. 797-800.

25. Dannull, J., et al., *Specific binding of HIV-1 nucleocapsid protein to PSI RNA in vitro requires N-terminal zinc finger and flanking basic amino acid residues.* EMBO J, 1994. 13(7): p. 1525-33.

26. Wu, Y., *Unwinding and rewinding: double faces of helicase?* J Nucleic Acids, 2012. 2012: p. 140601.

27. Jankowsky, E. and M.E. Fairman, *RNA helicases--one fold for many functions.* Curr Opin Struct Biol, 2007. 17(3): p. 316-24.

28. Yang, Q. and E. Jankowsky, *ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1.* Biochemistry, 2005. 44(41): p. 13591-601.

29. Holland, J., et al., *Rapid evolution of RNA genomes.* Science, 1982. 215(4540): p. 1577-85.

30. Stanway, G., *Structure, function and evolution of picornaviruses.* J Gen Virol, 1990. 71 (Pt 11): p. 2483-501.

31. Bender, W. and N. Davidson, *Mapping of poly(A) sequences in the electron microscope reveals unusual structure of type C oncornavirus RNA molecules.* Cell, 1976. 7(4): p. 595-607.

32. Goto, T., M. Nakai, and K. Ikuta, *The life-cycle of human immunodeficiency virus type 1.* Micron, 1998. 29(2-3): p. 123-38.

33. Craveiro, M., et al., *Metabolic pathways as regulators of HIV infection.* Curr Opin HIV AIDS, 2013. 8(3): p. 182-9.

34. Mokili, J. and B. Korber, *The spread of HIV in Africa.* J Neurovirol, 2005. 11 Suppl 1: p. 66-75.

35. Cherutich, P., R. Bunnell, and J. Mermin, *HIV Testing: Current Practice and Future Directions.* Curr HIV/AIDS Rep, 2013. 10(2): p. 134-41.

36. Sarafianos, S.G., et al., *Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition.* J Mol Biol, 2009. 385(3): p. 693-713.

37. Huff, J.R., *HIV protease: a novel chemotherapeutic target for AIDS.* J Med Chem, 1991. 34(8): p. 2305-14.

38. Frankel, A.D. and J.A. Young, *HIV-1: fifteen proteins and an RNA.* Annu Rev Biochem, 1998. 67: p. 1-25.

39. Freed, E.O., *HIV-1 gag proteins: diverse functions in the virus life cycle.* Virology, 1998. 251(1): p. 1-15.

40. Pollard, V.W. and M.H. Malim, *The HIV-1 Rev protein.* Annu Rev Microbiol, 1998. 52: p. 491-532.

41. Schuler, W., et al., *NMR structure of the (52-96) C-terminal domain of the HIV-1 regulatory protein Vpr: molecular insights into its biological functions.* J Mol Biol, 1999. 285(5): p. 2105-17.

42. Schwartz, S., et al., *Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs.* J Virol, 1990. 64(11): p. 5448-56.

43. Ciminale, V., et al., *A bioassay for HIV-1 based on Env-CD4 interaction.* AIDS Res Hum Retroviruses, 1990. 6(11): p. 1281-7.

44. Kitamura, Y., et al., *Inhibition of replication of HIV-1 at both early and late stages of the viral life cycle by single-chain antibody against viral integrase.* J Acquir Immune Defic Syndr Hum Retrovirol, 1999. 20(2): p. 105-14.

45. Campbell, E.M. and T.J. Hope, *Live cell imaging of the HIV-1 life cycle.* Trends Microbiol, 2008. 16(12): p. 580-7.

46. Moore, M.D. and W.S. Hu, *HIV-1 RNA dimerization: It takes two to tango.* AIDS Rev, 2009. 11(2): p. 91-102.

47. Ennifar, E., et al., *Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: from crystal to cell.* Nucleic Acids Res, 2006. 34(8): p. 2328-39.

48. Berkhout, B. and J.L. van Wamel, *Role of the DIS hairpin in replication of human immunodeficiency virus type 1.* J Virol, 1996. 70(10): p. 6723-32.

49. Paillart, J.C., et al., *Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic RNA.* J Biol Chem, 1994. 269(44): p. 27486-93.

50. Ennifar, E., et al., *HIV-1 RNA dimerization initiation site is structurally similar to the ribosomal A site and binds aminoglycoside antibiotics.* J Biol Chem, 2003. 278(4): p. 2723-30.

51. Paillart, J.C., et al., *A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA.* Proc Natl Acad Sci U S A, 1996. 93(11): p. 5572-7.

52. Paillart, J.C., et al., *Dimerization of retroviral genomic RNAs: structural and functional implications.* Biochimie, 1996. 78(7): p. 639-53.

53. Skripkin, E., et al., *Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro.* Proc Natl Acad Sci U S A, 1994. 91(11): p. 4945-9.

54. Berkowitz, R., J. Fisher, and S.P. Goff, *RNA packaging.* Curr Top Microbiol Immunol, 1996. 214: p. 177-218.

55. Jossinet, F., et al., *Dimerization of HIV-1 genomic RNA of subtypes A and B: RNA loop structure and magnesium binding.* RNA, 1999. 5(9): p. 1222-34.

56. Clever, J.L. and T.G. Parslow, *Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation.* J Virol, 1997. 71(5): p. 3407-14.

57. Laughrea, M., et al., *Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization.* J Virol, 1997. 71(5): p. 3397-406.

58. Paillart, J.C., et al., *A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis.* J Virol, 1996. 70(12): p. 8348-54.

59. Laughrea, M. and L. Jette, *Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms, and all sequences upstream*

or downstream of hairpin 248-271 are dispensable for dimer formation. Biochemistry, 1996. 35(5): p. 1589-98.

60. Muriaux, D., P. Fossé, and J. Paoletti, *A kissing complex together with a stable dimer is involved in the HIV-1$_{Lai}$ RNA dimerization process in vitro.* Biochemistry, 1996. 35: p. 5075-5082.

61. Feng, Y.X., et al., *HIV-1 nucleocapsid protein induces "maturation" of dimeric retroviral RNA in vitro.* Proc Natl Acad Sci U S A, 1996. 93(15): p. 7577-81.

62. Laughrea, M. and L. Jette, *HIV-1 genome dimerization: formation kinetics and thermal stability of dimeric HIV-1Lai RNAs are not improved by the 1-232 and 296-790 regions flanking the kissing-loop domain.* Biochemistry, 1996. 35(29): p. 9366-74.

63. Muriaux, D., et al., *NCP7 activates HIV-1 Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer.* J Biol Chem, 1996. 271(52): p. 33686-33692.

64. Rist, M.J. and J.P. Marino, *Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1.* Biochemistry, 2002. 41(50): p. 14762-70.

65. Takahashi, K.I., et al., *Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome.* RNA, 2000. 6(1): p. 96-102.

66. Takahashi, K.I., et al., *Two basic regions of NCp7 are sufficient for conformational conversion of HIV-1 dimerization initiation site from kissing-loop dimer to extended-duplex dimer.* J Biol Chem, 2001. 276(33): p. 31274-8.

67. Ennifar, E. and P. Dumas, *Polymorphism of bulged-out residues in HIV-1 RNA DIS kissing complex and structure comparison with solution studies.* J Mol Biol, 2006. 356(3): p. 771-82.

68. Ennifar, E., et al., *Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site.* Nat Struct Biol, 2001. 8(12): p. 1064-8.

69. Ennifar, E., et al., *The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges.* Structure, 1999. 7(11): p. 1439-49.

70. Baba, S., et al., *Solution RNA structures of the HIV-1 dimerization initiation site in the kissing-loop and extended-duplex dimers.* J Biochem, 2005. 138(5): p. 583-92.

71. Kieken, F., et al., *A new NMR solution structure of the SL1 HIV-1Lai loop-loop dimer.* Nucleic Acids Res, 2006. 34(1): p. 343-52.

72. Ulyanov, N.B., et al., *NMR structure of the full-length linear dimer of stem-loop-1 RNA in the HIV-1 dimer initiation site.* J Biol Chem, 2006. 281(23): p. 16168-77.

73. Bernacchi, S., et al., *Aminoglycoside binding to the HIV-1 RNA dimerization initiation site: thermodynamics and effect on the kissing-loop to duplex conversion.* Nucleic Acids Res, 2007. 35(21): p. 7128-39.

74. Ennifar, E., et al., *A structure-based approach for targeting the HIV-1 genomic RNA dimerization initiation site.* Biochimie, 2007. 89(10): p. 1195-203.

75. Freisz, S., et al., *Binding of aminoglycoside antibiotics to the duplex form of the HIV-1 genomic RNA dimerization initiation site.* Angew Chem Int Ed Engl, 2008. 47(22): p. 4110-3.

76. Zhao, C. and J.P. Marino, *Synthesis of HIV-1 Psi-site RNA sequences with site specific incorporation of the fluorescent base analog 2-aminopurine.* Tetrahedron, 2007. 63(17): p. 3575-3584.

77. Brunel, C., et al., *RNA loop-loop interactions as dynamic functional motifs.* Biochimie, 2002. 84(9): p. 925-44.

78. Horiya, S., et al., *RNA LEGO: magnesium-dependent assembly of RNA building blocks through loop-loop interactions.* Nucleic Acids Res Suppl, 2002(2): p. 41-2.

79. Huthoff, H. and B. Berkhout, *Multiple secondary structure rearrangements during HIV-1 RNA dimerization.* Biochemistry, 2002. 41(33): p. 10439-45.

80. Darlix, J.L., et al., *Flexible Nature and Specific Functions of the HIV-1 Nucleocapsid Protein.* J Mol Biol, 2011. 410(4): p. 565-81.

81. Levin, J.G., et al., *Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism.* Prog Nucleic Acid Res Mol Biol, 2005. 80: p. 217-86.

82. Darlix, J.L., et al., *First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses.* J Mol Biol, 1995. 254(4): p. 523-37.

83. Berkowitz, R.D., et al., *Retroviral nucleocapsid domains mediate the specific recognition of genomic viral RNAs by chimeric Gag polyproteins during RNA packaging in vivo.* J Virol, 1995. 69(10): p. 6445-56.

84. Davies, J. and B.D. Davis, *Misreading of ribonucleic acid code words induced by aminoglycoside antibiotics. The effect of drug concentration.* J Biol Chem, 1968. 243(12): p. 3312-6.

85. Davies, J., *Effects of streptomycin and related antibiotics on protein synthesis.* Antimicrob Agents Chemother (Bethesda), 1965. 5: p. 1001-5.

86. Fourmy, D., M.I. Recht, and J.D. Puglisi, *Binding of neomycin-class aminoglycoside antibiotics to the A-site of 16 S rRNA.* J Mol Biol, 1998. 277(2): p. 347-62.

87. Blanchard, S.C., et al., *rRNA chemical groups required for aminoglycoside binding.* Biochemistry, 1998. 37(21): p. 7716-24.

88. Moazed, D. and H.F. Noller, *Interaction of antibiotics with functional sites in 16S ribosomal RNA.* Nature, 1987. 327(6121): p. 389-94.

89. Fourmy, D., et al., *Structure of the A site of Escherichia coli 16S ribosomal RNA complexed with an aminoglycoside antibiotic.* Science, 1996. 274(5291): p. 1367-71.

90. Wong, C.H., et al., *Specificity of aminoglycoside antibiotics for the A-site of the decoding region of ribosomal RNA.* Chem Biol, 1998. 5(7): p. 397-406.

91. Griffey, R.H., et al., *Determinants of aminoglycoside-binding specificity for rRNA by using mass spectrometry.* Proc Natl Acad Sci U S A, 1999. 96(18): p. 10129-33.

92. Ecker, D.J. and R.H. Griffey, *RNA as a small-molecule drug target: doubling the value of genomics.* Drug Discov Today, 1999. 4(9): p. 420-429.

93. Kaul, M. and D.S. Pilch, *Thermodynamics of aminoglycoside-rRNA recognition: the binding of neomycin-class aminoglycosides to the A site of 16S rRNA.* Biochemistry, 2002. 41(24): p. 7695-706.

94. Woodcock, J., et al., *Interaction of antibiotics with A- and P-site-specific bases in 16S ribosomal RNA.* EMBO J, 1991. 10(10): p. 3099-103.

95. Aleman, E.A., R. Lamichhane, and D. Rueda, *Exploring RNA folding one molecule at a time.* Curr Opin Chem Biol, 2008. 12(6): p. 647-54.

96. Wang, M.D., et al., *Stretching DNA with optical tweezers.* Biophys J, 1997. 72(3): p. 1335-46.

97. Neuman, K.C. and A. Nagy, *Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy.* Nat Methods, 2008. 5(6): p. 491-505.

98. Digman, M.A., et al., *Fluctuation correlation spectroscopy with a laser-scanning microscope: exploiting the hidden time structure.* Biophys J, 2005. 88(5): p. L33-6.

99. Jaiswal, J.K. and S.M. Simon, *Imaging single events at the cell membrane.* Nat Chem Biol, 2007. 3(2): p. 92-8.

100. Ha, T., *Single-molecule fluorescence resonance energy transfer.* Methods, 2001. 25(1): p. 78-86.

101. Sun, Y., et al., *FRET microscopy in 2010: the legacy of Theodor Forster on the 100th anniversary of his birth.* Chemphyschem, 2011. 12(3): p. 462-74.

102. Clegg, R.M., *Fluorescence resonance energy transfer.* Curr Opin Biotechnol, 1995. 6(1): p. 103-10.

103. Zhao, R. and D. Rueda, *RNA folding dynamics by single-molecule fluorescence resonance energy transfer.* Methods, 2009. 49(2): p. 112-7.

104. Lodmell, J.S., et al., *Convergence of natural and artificial evolution on an RNA loop-loop interaction: the HIV-1 dimerization initiation site.* RNA, 2000. 6(9): p. 1267-76.

105. Rueda, D. and N.G. Walter, *Fluorescent energy transfer readout of an aptazyme-based biosensor.* Methods Mol Biol, 2006. 335: p. 289-310.

106. Vo, M.N., et al., *Effect of Mg(2+) and Na(+) on the nucleic acid chaperone activity of HIV-1 nucleocapsid protein: implications for reverse transcription.* J Mol Biol, 2009. 386(3): p. 773-88.

107. Li, M., et al., *Selection of peptides that target the aminoacyl-tRNA site of bacterial 16S ribosomal RNA.* Biochemistry, 2009. 48(35): p. 8299-311.

108. Roy, R., S. Hohng, and T. Ha, *A practical guide to single-molecule FRET.* Nat Methods, 2008. 5(6): p. 507-16.

109. Cann, A.J. and J. Karn, *Molecular biology of HIV: new insights into the virus life-cycle.* AIDS, 1989. 3 Suppl 1: p. S19-34.

110. Laughrea, M. and L. Jetté, *A 19-nucleotide sequence upstream of the 5' major splice donor site is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA.* Biochemistry, 1994. 33: p. 13464-13474.

111. Lu, K., et al., *NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging.* Science, 2011. 334(6053): p. 242-5.

112. Kenyon, J.C., et al., *In-gel probing of individual RNA conformers within a mixed population reveals a dimerization structural switch in the HIV-1 leader.* Nucleic Acids Res, 2013. 41(18): p. e174.

113. Clever, J.L., M.L. Wong, and T.G. Parslow, *Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA.* J Virol, 1996. 70(9): p. 5902-8.

114. Paillart, J.C., et al., *Non-canonical interactions in a kissing loop complex: the dimerization initiation site of HIV-1 genomic RNA.* J Mol Biol, 1997. 270(1): p. 36-49.

115. Paillart, J.C., et al., *A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA.* Proc Natl Acad Sci U S A, 1996. 93: p. 5572-5577.

116. Mihailescu, M.R. and J.P. Marino, *A proton-coupled dynamic conformational switch in the HIV-1 dimerization initiation site kissing complex.* Proc Natl Acad Sci U S A, 2004. 101(5): p. 1189-94.

117. Zhao, R., et al., *Laser-assisted single-molecule refolding (LASR).* Biophys J, 2010. 99(6): p. 1925-31.

118. Salim, N., et al., *Thermodynamic and kinetic analysis of an RNA kissing interaction and its resolution into an extended duplex.* Biophys J, 2012. 102(5): p. 1097-107.

119. Murphy, M.C., et al., *Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy.* Biophys J, 2004. 86(4): p. 2530-7.

120. Weixlbaumer, A., et al., *Determination of thermodynamic parameters for HIV DIS type loop-loop kissing complexes.* Nucleic Acids Res, 2004. 32(17): p. 5126-33.

121. Ramos, A., C.C. Gubser, and G. Varani, *Recent solution structures of RNA and its complexes with drugs, peptides and proteins.* Curr Opin Struct Biol, 1997. 7(3): p. 317-23.

122. Beaurain, F., et al., *Molecular dynamics reveals the stabilizing role of loop closing residues in kissing interactions: comparison between TAR-TAR* and TAR-aptamer.* Nucleic Acids Res, 2003. 31(14): p. 4275-84.

123. Lebars, I., et al., *Exploring TAR-RNA aptamer loop-loop interaction by X-ray crystallography, UV spectroscopy and surface plasmon resonance.* Nucleic Acids Res, 2008. 36(22): p. 7146-56.

124. Bernacchi, S., et al., *Mechanism of hairpin-duplex conversion for the HIV-1 dimerization initiation site.* J Biol Chem, 2005. 280(48): p. 40112-21.

125. Turner, K.B., N.A. Hagan, and D. Fabris, *Understanding the isomerization of the HIV-1 dimerization initiation domain by the nucleocapsid protein.* J Mol Biol, 2007. 369(3): p. 812-28.

126. Hagan, N.A. and D. Fabris, *Dissecting the protein-RNA and RNA-RNA interactions in the nucleocapsid-mediated dimerization and isomerization of HIV-1 stemloop 1.* J Mol Biol, 2007. 365(2): p. 396-410.

127. Mujeeb, A., et al., *Nucleocapsid protein-mediated maturation of dimer initiation complex of full-length SL1 stemloop of HIV-1: sequence effects and mechanism of RNA refolding.* Nucleic Acids Res, 2007. 35(6): p. 2026-34.

128. Van Melckebeke, H., et al., *Liquid-crystal NMR structure of HIV TAR RNA bound to its SELEX RNA aptamer reveals the origins of the high stability of the complex.* Proc Natl Acad Sci U S A, 2008. 105(27): p. 9210-5.

129. Marino, J.P., et al., *Bent helix formation between RNA hairpins with complementary loops.* Science, 1995. 268(5216): p. 1448-54.

130. Lee, A.J. and D.M. Crothers, *The solution structure of an RNA loop-loop complex: the ColE1 inverted loop sequence.* Structure, 1998. 6(8): p. 993-1005.

131. Sun, X., Q. Zhang, and H.M. Al-Hashimi, *Resolving fast and slow motions in the internal loop containing stem-loop 1 of HIV-1 that are modulated by Mg2+ binding: role in the kissing-duplex structural transition.* Nucleic Acids Res, 2007. 35(5): p. 1698-713.

132. Dethoff, E.A., et al., *Visualizing transient low-populated structures of RNA.* Nature, 2012. 491(7426): p. 724-8.

133. Song, R., et al., *HIV-1 viral RNA is selected in the form of monomers that dimerize in a three-step protease-dependent process; the DIS of stem-loop 1 initiates viral RNA dimerization.* J Mol Biol, 2007. 371(4): p. 1084-98.

134. Reblova, K., et al., *Conformations of flanking bases in HIV-1 RNA DIS kissing complexes studied by molecular dynamics.* Biophys J, 2007. 93(11): p. 3932-49.

135. Li, P.T. and I. Tinoco, Jr., *Mechanical unfolding of two DIS RNA kissing complexes from HIV-1.* J Mol Biol, 2009. 386(5): p. 1343-56.

136. Ennifar, E., et al., *Structure-guided discovery of a novel aminoglycoside conjugate targeting HIV-1 RNA viral genome.* ACS Chem Biol, 2013. 8(11): p. 2509-17.

137. Tanchou, V., et al., *Role of the N-terminal zinc finger of human immunodeficiency virus type 1 nucleocapsid protein in virus structure and replication.* J Virol, 1998. 72(5): p. 4442-7.

138. Karpel, R.L., L.E. Henderson, and S. Oroszlan, *Interactions of retroviral structural proteins with single-stranded nucleic acids.* J Biol Chem, 1987. 262(11): p. 4961-7.

139. Barat, C., et al., *HIV-1 reverse transcriptase specifically interacts with the anticodon domain of its cognate primer tRNA.* EMBO J, 1989. 8(11): p. 3279-85.

140. Volkmann, S., et al., *Enzymatic analysis of two HIV-1 reverse transcriptase mutants with mutations in carboxyl-terminal amino acid residues conserved among retroviral ribonucleases H.* J Biol Chem, 1993. 268(4): p. 2674-83.

141. Lapadat-Tapolsky, M., et al., *Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle.* Nucleic Acids Res, 1993. 21(4): p. 831-9.

142. Muriaux, D., et al., *NCp7 activates HIV-1Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer.* J Biol Chem, 1996. 271(52): p. 33686-92.

143. Lodmell, J.S., et al., *Structure and dimerization of HIV-1 kissing loop aptamers.* J Mol Biol, 2001. 311(3): p. 475-90.

144. Dryden, M., K. Hand, and P. Davey, *Antibiotics for community-acquired pneumonia.* J Antimicrob Chemother, 2009. 64(6): p. 1123-5.

145. Chopra, I., *Chemotherapy of mycobacterial infections: molecular action of established drugs and investigational agents.* Expert Opin Investig Drugs, 1998. 7(2): p. 253-5.

146. Davies, J., L. Gorini, and B.D. Davis, *Misreading of RNA codewords induced by aminoglycoside antibiotics.* Mol Pharmacol, 1965. 1(1): p. 93-106.

147. Edelmann, P. and J. Gallant, *Mistranslation in E. coli.* Cell, 1977. 10(1): p. 131-7.

148. Hirokawa, G., et al., *Binding of ribosome recycling factor to ribosomes, comparison with tRNA.* J Biol Chem, 2002. 277(39): p. 35847-52.

149. Hirokawa, G., et al., *Post-termination complex disassembly by ribosome recycling factor, a functional tRNA mimic.* EMBO J, 2002. 21(9): p. 2272-81.

150. Chittapragada, M., S. Roberts, and Y.W. Ham, *Aminoglycosides: molecular insights on the recognition of RNA and aminoglycoside mimics.* Perspect Medicin Chem, 2009. 3: p. 21-37.

151. Francois, B., et al., *Crystal structures of complexes between aminoglycosides and decoding A site oligonucleotides: role of the number of rings and positive charges in the specific binding leading to miscoding.* Nucleic Acids Res, 2005. 33(17): p. 5677-90.

152. Carter, A.P., et al., *Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics.* Nature, 2000. 407(6802): p. 340-8.

153. McPike, M.P., et al., *Footprinting, circular dichroism and UV melting studies on neomycin B binding to the packaging region of human immunodeficiency virus type-1 RNA.* Nucleic Acids Res, 2002. 30(13): p. 2825-31.

154. Karunatilaka, K.S. and D. Rueda, *Single-Molecule Fluorescence Studies of RNA: A Decade's Progress.* Chem Phys Lett, 2009. 476(1): p. 1-10.

155. Ellis, R.J., *Macromolecular crowding: an important but neglected aspect of the intracellular environment.* Curr Opin Struct Biol, 2001. 11(1): p. 114-9.

156. Ellis, R.J., *Macromolecular crowding: obvious but underappreciated.* Trends Biochem Sci, 2001. 26(10): p. 597-604.

157. Lee, H.W., K.T. Briggs, and J.P. Marino, *Dissecting structural transitions in the HIV-1 dimerization initiation site RNA using 2-aminopurine fluorescence.* Methods, 2009. 49(2): p. 118-27.

158. Yuan, Y., et al., *Stem of SL1 RNA in HIV-1: structure and nucleocapsid protein binding for a 1 x 3 internal loop.* Biochemistry, 2003. 42(18): p. 5259-69.

159. Shen, N., et al., *Role of stem B, loop B, and nucleotides next to the primer binding site and the kissing-loop domain in human immunodeficiency virus type 1 replication and genomic-RNA dimerization.* J Virol, 2001. 75(21): p. 10543-9.

160. Sakuragi, J., et al., *Possible role of dimerization in human immunodeficiency virus type 1 genome RNA packaging.* J Virol, 2003. 77(7): p. 4060-9.

161. Sakuragi, J., S. Sakuragi, and T. Shioda, *Minimal region sufficient for genome dimerization in the human immunodeficiency virus type 1 virion and its potential roles in the early stages of viral replication.* J Virol, 2007. 81(15): p. 7985-92.

162. Tok, J.B., L.J. Dunn, and R.C. Des Jean, *Binding of dimeric aminoglycosides to the HIV-1 rev responsive element (RRE) RNA construct.* Bioorg Med Chem Lett, 2001. 11(9): p. 1127-31.

ABSTRACT

HIV-1 RNA DIMERIZATION AT THE SINGLE-MOLECUEL LEVEL

by

HANSINI RASANGA MUNDIGALA

May 2014

Advisors: Dr. David Rueda & Dr. Andrew Feig

Major: Chemistry (Analytical)

Degree: Doctor of Philosophy

The HIV-1 dimerization initiation sequence (DIS) is a conserved palindrome in the apical loop of a conserved hairpin motif in the 5'-UTR of its RNA genome. The DIS hairpin plays an important role in genome dimerization by forming a "kissing complex" (KC) between two complementary hairpins. Understanding the kinetics of this interaction is key to exploiting DIS as a possible HIV drug target. Here, we present a single-molecule Förster resonance energy transfer (smFRET) study of the dimerization reaction kinetics. Our data show the real-time formation and dissociation dynamics of individual kissing complexes, as well as, the formation of the mature extended duplex complex that is ultimately required for virion packaging. Interestingly, the single-molecule trajectories reveal the presence of a previously unobserved bent intermediate required for extended duplex formation. The universally conserved A272 is essential for the formation of this intermediate, which is stabilized by Mg^{2+}, but not by K^{+}, cations. We propose a 3D model of a possible bent intermediate and a minimal dimerization pathway consisting of three steps with two obligatory intermediates (kissing complex and bent intermediate) and driven by Mg^{2+} ions. Additionally, it was confirmed that in the

HIV-1 nucleocapsid protein (NC) chaperoned dimerization pathway, NC initially binds to monomer hairpin RNA and facitilates the KC formation at a faster rate. Thus, HIV-1 NC facilitates the formation of the extended RNA duplex by shifting the KC and monomer RNA equilibrium toward the KC. Further, it was observed that aminoglycoside antibiotics bind to the HIV-1 DIS monomer RNA through its neamine core with a high affinity. In contrast, aminoglycosides with extensions at the 5^{th} position of the neamine ring II are required to stabilize the kissing interactions and hinder the progression of dimerization. The mechanistic insights gained from these experiments represent significant progress towards understanding the HIV-1 dimerization process, as well as identification and development of DIS-targeting molecules.

AUTOBIOGRAPHIC STATEMENT

HANSINI RASANGA MUNDIGALA

I was born in Kandy, Sri Lanka and was raised in Homagama, a city about 25 km from Colombo, Sri Lankan capital. After graduation from high school, I entered the Faculty of Science at the University of Colombo, from where I graduated with a Bachelor of Science degree with Chemistry honors in 2008. I joined Wayne State University in August 2009 for my doctoral studies.

PUBLICATIONS

1 Mundigala, H.R., Michaux, J., Feig, A.L., Ennifar, E., and Rueda, D." HIV-1 DIS forms an obligatory bent, kissing intermediate in the dimerization pathway" (Submitted for publication)
2 Mundigala, H.R., Chow, C., Feig, A.L., Ennifar, E., and Rueda, D. "HIV-1 RNA Aminoglycoside binding to HIV-1 DIS: thermodynamics and kinetics of dimerization at single molecule level" (in preparation).
3 Mundigala, H.R., Feig, A.L., Musier-Forsyth, K., and Rueda, D. "Role of HIV-1 Nucleocapsid protein in the HIV-1 RNA dimerization at the single-molecule level " (in preparation).

ORAL PRESENTATIONS

1 HIV-1 DIS dimerizes through a bent kissing intermediate: Presented at the Rustbel RNA meeting, 2012 October.
2 Single-molecule FRET studies on dimerization of HIV: Presented at the Association of Analytical chemists , Detroit chapter. 2012 November.

POSTER PRESENTATIONS

1 "The HIV-1 DIS RNA Dimerization Pathway Involves a Bent Intermediate." 57th Biophysical Society Annual Meeting, February 01 - 06, 2012.
2 "The HIV-1 DIS RNA Dimerizes Through a Bent Kissing Complex Intermediate." 2nd Midwest Single Molecule Workshop, July 26 - 27, 2012.
3 Single-Molecule Studies of HIV-1 Dimerization Initiation Sequence Kissing Interaction and its Resolution to a Stable Extended Duplex" International RNA Society meeting, May 31, 2012.
4 "Single-Molecule Studies of HIV-1 Dimerization Initiation Sequence Kissing Interaction and its Resolution to a Stable Extended Duplex" Rustbelt RNA meeting Oct 21-22, abstract # 59.
5 "Study of HIV-1 RNA dimerization Mechanism by using FRET", Tenth Annual Chemistry Graduate Student Research Symposium, Department of Chemistry, Wayne State University.